

Reclassification of SLC22 Transporters: Analysis of OAT, OCT, OCTN, and other Family

Members Reveals 8 Functional Subgroups

Darcy Engelhart¹, Jeffrey C. Granados², Da Shi³, Milton Saier Jr.⁴, Michael Baker⁶, Ruben Abagyan³, Sanjay K. Nigam^{5,6}

¹Department of Biology, University of California San Diego, La Jolla 92093

²Department of Bioengineering, University of California San Diego, La Jolla 92093

³School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla 92093

⁴Department of Molecular Biology, Division of Biological Sciences, University of California at San Diego, San Diego, CA, USA

⁵Department of Pediatrics, University of California San Diego, La Jolla 92093

⁶Department of Medicine, University of California San Diego, La Jolla 92093

*To whom correspondence should be addressed: snigam@ucsd.edu

Running title: Functional subgroups for SLC22

Abstract

Among transporters, the SLC22 family is emerging as a central hub of endogenous physiology. The family consists of organic anion transporters (OATs), organic cation transporters (OCTs) and zwitterion transporters (OCTNs). Despite being known as “drug” transporters, these multi-specific, oligo-specific, and relatively mono-specific transporters facilitate the movement of metabolites and key signaling molecules. An in-depth reanalysis supports a reassignment of these proteins into eight functional subgroups with four new subgroups arising from the previously defined OAT subclade. These OAT subgroups are: OATS1 (SLC22A6, SLC22A8, and SLC22A20), OATS2 (SLC22A7), OATS3 (SLC22A11, SLC22A12, and Slc22a22), and OATS4 (SLC22A9, SLC22A10, SLC22A24, and SLC22A25). We propose merging the OCTN (SLC22A4, SLC22A5, and Slc22a21) and OCT-related (SLC22A15 and SLC22A16) subclades into the OCTN/OCTN-related subgroup. Functional support for the eight subgroups comes from network analysis of data from GWAS, in vivo models, and in vitro assays. These data emphasize shared substrate specificity of SLC22 transporters for characteristic metabolites such as prostaglandins, uric acid, carnitine, creatinine, and estrone sulfate. Some important subgroup associations include: OATS1 with metabolites, signaling molecules, uremic toxins and odorants, OATS2 with cyclic nucleotides, OATS3 with uric acid, OATS4 with conjugated sex hormones, particularly etiocholanolone glucuronide, OCT with monoamine neurotransmitters, and OCTN/OCTN-related with ergothioneine and carnitine derivatives. The OAT-like and OAT-related subgroups remain understudied and therefore do not have assigned functionality. Relatedness within subgroups is supported by multiple sequence alignments, evolutionarily conserved protein motifs, genomic localization, and tissue expression. We also highlight low level sequence similarity of SLC22 members with other non-transport proteins. Our data suggest that the SLC22 family can work among itself, as well as with other transporters and enzymes, to

optimize levels of numerous metabolites and signaling molecules, as proposed by the Remote Sensing and Signaling Theory.

Keywords: Transporters; endogenous metabolism; Functional subgroups; SLC22; Remote Sensing and Signaling; drug transporters

Introduction

The SLC (solute carrier) gene family includes 65 families with over 400 transporter genes. In humans, 52 of these families are expressed, encompassing more than 395 genes and it has been estimated that ~2000 (10% of the genome) human genes are transporter-related (1). Various solute carrier 22 (SLC22) members are expressed on both the apical and basolateral surfaces of epithelial cells where they direct small molecule transport between body fluids and vital organs, such as the kidney, liver, heart, and brain (2). SLC22 transporters are also found in circulating cell types such as erythrocytes (e.g. SLC22A7), monocytes, and macrophages (e.g. SLC22A3, SLC22A4, SLC22A15, and SLC22A16) (3,4). With recent calls for research on solute carriers, there has been a large influx of data over the past five years, including novel roles in remote sensing and signaling, leading to the need for a more comprehensive understanding of the functional importance of transporters (5).

The SLC22 family is comprised of at least 31 transporters and is found in species ranging from *Arabidopsis thaliana* of the plant kingdom to modern day humans (6,7). Knowledge surrounding this family of proteins has expanded greatly since its proposed formation in 1997, when SLC22A6 (OAT1, originally known as novel kidney transporter or NKT) was first cloned (8). Its homology to SLC22A1 (OCT1) and SLC22A7 (OAT2/NLT) led to the establishment of a new family (SLC22, TC# 2.A.1.19) of transport proteins within the Major Facilitator Superfamily (TC# 2.A.1, MFS) as classified by the IUBMB-approved Transporter Classification (TC) system (8,9). These proteins all share 12 α -helical transmembrane domains (TMD), a large extracellular domain (ECD) between TMD1 and TMD2, and a large intracellular domain (ICD) between TMD6 and TMD7 (10). Research has shown these transporters to be integral participants in the movement of drugs, toxins, and endogenous metabolites and signaling

molecules, such as prostaglandins, urate, α -ketoglutarate, carnitine, and cyclic nucleotides across the cell membrane (11).

As key players in small molecule transport, SLC22 members are hypothesized to play a role in the Remote Sensing and Signaling Theory (12–15). The Remote Sensing and Signaling Theory posits that several drug related genes, like transporters and enzymes, aid in maintaining homeostasis through remote communication between organs via small molecule substrates in the blood that may also serve as signaling molecules that regulate gene expression (16). This communication is supported by the example of serum uric acid levels. In the setting of compromised kidney function, the increase in serum uric acid seems to be partly mitigated through a compensatory increase in the expression and/or function of ABCG2 in the intestine, which allows the excretion of uric acid in the feces rather than the urine (17,18). Current research is focused on determining the ways in which these transporters collaborate to regulate metabolite levels throughout the body (19).

Rather than maintaining a simple division of SLC22 into organic anion transporters (OATs), organic cation transporters (OCTs) and organic zwitterion/cation transporters (OCTNs), previous evolutionary studies have identified six phylogenetic “subclades” – OAT, OAT-like, OAT-related, OCT, OCTN-related, and OCTN – within the OAT and OCT “major clades” (10). These subclades consist, on average, of three to four members with the exception of the OAT subclade which claims more than half of the 31 known members of SLC22 (10). Although these subclades are phylogenetically sound, the endogenous functions of many SLC22 members within the six subclades remain ill-defined or unknown. With the emergence of new data, we performed a re-analysis of the SLC22 family to better characterize the functional, endogenous grouping of these transporters. Our re-analysis shows eight apparent subgroups, with four of these subgroups

arising out of the previously defined OAT subclade. Because these groupings are more closely related to well-known OATs rather than OCTs, OCTNs or other subclades, we refer to these as OAT subgroups (OATS1, OATS2, OATS3, OATS4).

We considered many factors in our re-analysis of SLC22 and subsequent designation of functionally based subgroups. To better describe the subgroups while still highlighting the nuances of each individual transporter, we utilized data from genomic loci, tissue expression, sequence similarity searches, proteomic motif searches, and functional transporter-metabolite data from GWAS, in vitro assays, and in vivo models. In place of phylogenetic studies, we performed multiple sequence alignments (MSA) and generated guide-trees that are based on sequence similarity or homology and thus provide more insight into function than solely phylogenetic studies. While the SLC22 family is composed of putative transporters, some members, like Slc22a20 and Slc22a17, have proposed mechanisms that differ from those of classic transporters (20,21). To that effect, we explored the sequence similarities between SLC22 transporters and non-transport related proteins. This work elucidates the diversity of the endogenous functions of SLC22 transporters in various tissues and provides an updated functional framework for assigning each transporter to a subgroup. Considering the importance of SLC22 transporters, forming functional groups that incorporate endogenous substrates and tissue expression patterns can help better define their roles in intra-organ, inter-organ, and inter-organismal communication.

Materials and Methods

Data collection

SLC22 human and mouse sequences were collected from the National Center for Biotechnology Information (NCBI) protein database. Sequences were confirmed and genomic loci were recorded using the University of California Santa Cruz (UCSC) genome browser by searching within each available species on the online platform (<https://genome.ucsc.edu/cgi-bin/hgGateway>) (22). The NCBI BLASTp web-based program was used to find similar sequences to those from the NCBI protein database. BLASTp was run with default parameters using query SLC22 sequences from human or mouse. The database chosen was “non-redundant protein sequence” (nr) and no organisms were excluded (23). Tissue expression of all human SLC22 members was collected from the Human Protein Atlas, GTEx dataset, Illumina Body Map, ENCODE dataset, and RNA-seq datasets available on the EMBL-EBI Expression Atlas (<https://www.ebi.ac.uk/gxa/home>) (24). Tissue expression data were also collected via extensive literature search.

Sequence Alignment and Guide-trees

Sequences for SLC22 were aligned using Clustal-Omega with default parameters via the online platform provided by the European Bioinformatics Institute (EMBL-EBI) (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), as well as MAFFT (Multiple alignment using fast Fourier transform) and ICM-Pro v3.8-7c (25–28). Clustal-Omega, MAFFT and ICM-Pro v3.8-7c produced similar topologies. These alignments were then visualized using The Interactive Tree of Life (<http://itol.embl.de/>) (29). Topology was analyzed by branch length values, which are a result of the neighbor-joining method. This method calculates the number of amino acid changes

between each organism and the common ancestor from which it branched. It then adopts the minimum-evolution criteria (ME) by building a tree which minimizes the sum of all branch lengths to visually display relatedness (30). SSearch36 was utilized to compare representative sequences of all members of the Drug:H⁺ Antiporter-1 (12 Spanner) (DHA1) Family (2.A.1.2) and the the SLC22A family (2.A.1.19) from the Transporter Classification Database (<http://www.tcdb.org/>) with the Cyanate Porter (CP) Family (2.A.1.17) as an outgroup to further investigate the belongingness of SLC22A18 in either the SLC22 or DHA1 family (9,31). SSearch36 is an exhaustive comparison method that uses the Smith-Waterman (SW) algorithm to compare FASTA files find sequence similarities (31,32).

ICM-Pro v3.8-7c was used to align sequences in FASTA format as well as perform homology searches of all human and mouse SLC22 sequences against ICM-Pro's curated database of all known proteins (28). Threshold for homology significance was determined by the probability of structural insignificance (pP), defined as the negative log of the probability value of a homology comparison. Alignments were discarded if the pP value was less than 5.0, indicating that the homology shared between two sequences is likely not due to random sequence similarities.

Motif Analysis

Motif comparisons were performed on the subgroups of the OAT Subclade using the Multiple Expectation-maximum for Motif Elicitation (MEME) (<http://meme-suite.org/tools/meme>) suite (33). A threshold of 20 motifs containing a range of 6-20 amino acid length was set with the normal discovery mode. This detection method yielded a set of evolutionarily conserved motifs within all Oat Subclade sequences (n=57) as well as a set of

evolutionarily conserved motifs for each of the four proposed Oat subgroups. These motifs were then mapped onto 2D topologies of one member from each of the newly proposed Oat subgroups (SLC22A6 for OatS1, SLC22A7 for OatS2, SLC22A12 for OatS3, and SLC22A9 for OatS4). A separate motif analysis was also performed for the rodent expansion consisting of *slc22a19*, and *slc22a26-30* and was mapped onto mouse *slc22a27*. Transmembrane domains (TMDs) of these transporters were predicted by the Constrained Consensus TOPology prediction server (CCTOP) (<http://cctop.enzim.ttk.mta.hu/>) (34). TMD locations and the motif locations were entered into TOPO2 (<http://www.sacs.ucsf.edu/TOPO2/>) to visualize the 2D representation of the transporters with the OAT subclade's evolutionarily conserved motifs shown in blue and each subgroup's evolutionarily conserved motifs shown in red (35).

SNP, mutation, in-vitro, knockout, and drug transport data

To determine the diversity of substrate specificity, the number of drugs that list SLC22 members as a target on DrugBank were recorded (36). The Metabolomics GWAS server was utilized to determine SNPs within all SLC22 members. The dataset produced by Shin et al. (2014) with the cohort KORA+TwinsUK (blood) and the association of single metabolites was chosen. This dataset was searched by gene symbol (e.g. SLC22A6) (37,38). The EMBL GWAS Catalog and Metabolomix's table of all published GWAS with metabolomics (<http://www.metabolomix.com/list-of-all-published-gwas-with-metabolomics/>) were also utilized in searching for SNPs and their effect on metabolite transport by SLC22 members (39). Current literature available on the NCBI gene database under Gene References Into Functions (Gene RIFs) (<https://www.ncbi.nlm.nih.gov/gene/about-generif>) was used to search for non-synonymous mutations that did not affect protein expression yet affected transport of metabolites and/or drugs. These methods were accompanied by an extensive literature search for in-vitro

transport and knockout data. The import from table feature on Cytoscape 3.7.2 was used to generate functional networks for the entire SLC22 family and the subgroups (40).

Results

Emerging data continues to indicate the centrality of the SLC22 family (particularly OATs, OCTs, and OCTNs) in endogenous physiology (5,16). Our thorough reanalysis of the previous phylogenetic subclades revealed eight functional subgroups: OATS1, OATS2, OATS3, OATS4, OAT-like, OAT-related, OCT and OCTN/OCTN-related (Table 1). By grouping this large family of proteins, we highlighted differences in substrate selectivity, showing that each member has a unique profile of associated metabolites. Based on the number of different metabolites it interacts with, each SLC22 transporter can be classified as relatively mono-, oligo-, or multi-specific. In what follows, publicly available data from GWAS, in vitro, and in vivo datasets were used to build functional networks that support the subgroups (Figure 1). In addition to these functional data, subgroups were also supported by structural, genomic, and other analyses explained below. Because some SLC22 members remain understudied, we also investigated low level sequence identity with non-transport proteins to better characterize these “orphaned” transporters.

Analysis of Substrate Specificity and Selectivity Helps Categorize Mono-, Oligo-, and Multi-Specificity of SLC22 Members

The concept of multi-, oligo-, and mono-specific SLC22 transporters was supported in part based on the number of unique drugs that are known to interact with each SLC22 member (Table 2) (36). Several SLC22 members (e.g., OAT1, OCT1) are best known as “drug” transporters and due to this association, many have been tested as potential drug targets. While drugs are not the primary focus of this research, the number of drugs a transporter is linked to is indicative of how many structurally different substrates it can interact with. This may translate to endogenous compounds from different metabolic pathways. As interest in solute carriers has increased over the past decade, there has been a large influx of functional data (Supplementary

Table 1). We used these data to validate our initial specificity assignments, and found that, for the most part, the metabolite data were in agreement with the drug data. A transporter linked to many unique drugs was often linked to many unique metabolites. For example, OATS1 members SLC22A6 and SLC22A8 are linked to 99 and 126 drugs, respectively. This is reflected in the metabolite data, as each transporter was associated with at least 50 unique metabolites, confirming their multi-specific nature. OATS4 members SLC22A9, SLC22A10, SLC22A24, and SLC22A5 are understudied with respect to drugs. As a group, they are only associated with three drugs, making it difficult to predict their substrate selectivity. Endogenously, the group appears to have relatively mono-specific members that are dedicated to conjugated sex steroids, and oligo-specific members which are linked to conjugated sex hormones, short chain fatty acids, and bile acids.

Construction of Functional Networks from Metabolite-Transporter Interaction Data Support the Eight Subgroups

To visualize these transporter-metabolite interactions, which were acquired from a combination of GWAS, in vivo, and in vitro studies, we created networks using Cytoscape (40). These networks allowed us to see the extent of unique and overlapping substrate specificity between transporters in the SLC22 family and within the proposed subgroups (Supplementary Figures 1 and 2). In these networks, all edges are undirected and represent a statistically significant result linking an SLC22 member to a metabolite. To give an example, the OATS1 network uses the members (SLC22A6, SLC22A8, SLC22A20) as central nodes. Each associated metabolite is connected to the member, and the networks are then combined to represent the entire subgroup and demonstrate how a metabolite may be linked to multiple transporters (Supplementary Figure 1A). Functional data were available for 21 of 31 known SLC22 transporters. The trimmed SLC22 network is displayed in Figure 1, the individual subgroup

networks are in Supplementary Figures 1 and 2, and the total SLC22 network is in Supplementary Figure 3. The compiled data with transporter, metabolite, study, quantitative metric, and citation are present in Supplementary Table 1. While there is no single metabolite that is associated with all SLC22 transporters, some are linked to multiple family members, and thus may be characteristic for the family as a whole. These metabolites are prostaglandin E2, prostaglandin F2, estrone sulfate, uric acid, carnitine and creatinine, which are each linked to at least five different SLC22 members, respectively (Supplementary Figure 3). This result demonstrated that SLC22, as a group, is involved in regulating several metabolic processes, ranging from blood vessel dilation through prostaglandins to cellular energy production through carnitine (41,42). This also implies that the particular structural features of the SLC22 family in general (12 TMD, large ECD between TMD1 and TMD2, and large ICD between TMD6 and TMD7) lends itself well to interacting with these compounds. This is further supported by the subgroup-specific network analyses and motif analysis we performed (Figure 1)

OAT5(SLC22A6, SLC22A8, SLC22A20) Handles A Wide Variety of Metabolites, Signaling Molecules, Uremic Toxins, and Odorants

Several metabolites have been identified as substrates of SLC22A6 (OAT1) and SLC22A8 (OAT3). While many are unique, there is notable overlap. Both OAT1 and OAT3 interact with uremic toxins (indoxyl sulfate, p-cresol sulfate, uric acid) and gut microbiome derived products (indolelactate, 4-hydroxyphenylacetate), as well as many of the more general SLC22 metabolites, like prostaglandin E2, prostaglandin F2, uric acid, and creatinine (43–47). SLC22A20 (OAT6), while not as well-studied, has affinity for several odorants and short chain fatty acids that are also associated with OAT1 (48). OAT1 and OAT3 are clearly multi-specific, and OAT6 appears to be oligo-specific, as it handles both odorants and some short chain fatty

acids. With respect to remote signaling, the shared metabolites among these transporters (Supplementary Figure 1A) are noteworthy because of their tissue localization (Table 4). OAT1 and OAT3 are primarily expressed in the kidney proximal tubule, with some expression in other tissues, like the choroid plexus and retina (Table 4). OAT6, however, is expressed in the olfactory mucosa of mice, presumably reflecting its affinity for odorants (21,48). In the kidney, OAT1 and OAT3, along with many other SLC22 transport proteins, help regulate the urine levels of many metabolites and signaling molecules which may potentially facilitate inter-organismal communication. For example, a volatile compound in one organism may be excreted into the urine through OAT1 and then somehow sensed by another organism through a mechanism involving OAT6 in the olfactory mucosa (12).

OATS2 (SLC22A7) is a Systemically-Expressed Transporter of Organic Anions and Cyclic Nucleotides

SLC22A7 (OAT2) is the only member of the OATS2 subgroup and is associated with prototypical SLC22 substrates, such as prostaglandins, carnitine, creatinine, and uric acid (38,46,49,50). OAT2 is also linked to cyclic nucleotides and dicarboxylic acids, which when taken with the previous metabolites, creates a unique profile worthy of its own subgroup (Supplementary Figure 1B) (51). Another distinguishing feature of OAT2 is its tissue expression patterns (Table 4). While its expression in the liver and kidney are common to many SLC22 members, it has been localized to circulating red blood cells, where it may function in cyclic nucleotide transport (3). Its expression in a mobile cell type and transport of cyclic nucleotides raises the possibility that it may act as an avenue for signaling.

OATS3 (SLC22A11, SLC22A12, Slc22a22) Functions to Balance Uric Acid and Prostaglandins

In humans, SLC22A11 (OAT4) and SLC22A12 (URAT1) share only two substrates, uric acid and succinate (Supplementary Figure 1C) (52,53). Uric acid is a beneficial metabolite in the serum as it is thought to be responsible for more than half of human antioxidant activity in the blood (54). However, high levels of uric acid can be harmful and are associated with gout (55). URAT1 is associated with very few metabolites and is best understood for its role in uric acid reabsorption in the kidney proximal tubule, making it relatively mono-specific (52). OAT4, on the other hand, has been shown to transport prostaglandins and conjugated sex hormones in addition to uric acid, making it oligo-specific (56–58). URAT1 is almost exclusively expressed in the kidney, and OAT4 is expressed in the kidney, placenta, and epididymis (Table 4). The more diverse tissue expression of SLC22A11 seems consistent with its wider range of substrates. The subgroup differs in rodents because mice do not express Oat4. Instead, the rodent subgroup is composed of Slc22a12, known as renal-specific transporter (Rst) in mice, and Slc22a22, known as prostaglandin-specific organic anion transporter (Oat-pg). While Rst and Oat-pg do not share substrate specificity, together, they combine to play the role of URAT1 and OAT4 by handling uric acid and prostaglandins (59).

OATS4 (SLC22A9, SLC22A10, SLC22A24, SLC22A25) Members are Specifically Associated with Conjugated Sex Hormones

GWAS analyses support the association of all human members of this subgroup with one common metabolite, etiocholanolone glucuronide, a conjugated sex hormone, with a p-value of 4.12×10^{-27} or lower for all members (Table 3, Supplementary Table 1) (60). While this group shares at least one conjugated sex hormone, SLC22A24 and SLC22A9 appear to be more oligo-specific transporters, with SLC22A9 linked to short chain fatty acids and SLC22A24 linked to bile acids (61,62). SLC22A10 and SLC22A25 are only linked to conjugated sex hormones,

making them relatively mono-specific transporters (Supplementary Figure 1D) (60). In terms of tissue expression, there is a distinct correlation between patterns and shared function amongst human OATS4 members (Table 4). We predict that all four members are conjugated sex steroid transporters with SLC22A9, A10, and A25 showing high expression in the liver where conjugation of glucuronides and sulfates to androgens and other gonadal steroids occurs (61). SLC22A24 has low expression levels in the liver but is highly expressed in the proximal tubule, where it is predicted to reabsorb these conjugated steroids (61). This subgroup also includes a large rodent-specific expansion, consisting of Slc22a19 and Slc22a26-30. Although the rodent-specific expansion is greatly understudied, transport data for rat Slc22a9/a24 shows shared substrate specificity for estrone sulfate with SLC22A24, but not for bile acids or glucuronidated steroids, which is consistent with the lack of glucuronides in rat urine and serum (61). While sulfatases are extremely highly conserved amongst humans, rats, and mice, the separation of rodent- and nonrodent-specific OATS4 groups is likely due to the species differences in expression and function of glucuronidases (63). Despite their distinct differences from human OATS4 members in sequence similarity studies and minimal functional data, the rodent-specific transporters are also highly expressed in both liver and kidney (64).

OAT-Like (SLC22A13, SLC22A14) has Potentially Physiologically Important Roles

Very little functional data is available for the OAT-like subgroup. SLC22A13 (OAT10/ORCTL3) has been well characterized as a transporter of both urate and nicotinate, but SLC22A14 has no available transport data (65). However, N^α-methyl nicotinate is increased in the plasma levels of self-reported smokers, and GWAS studies have implicated SNPs in the SLC22A14 gene to be associated with success in smoking cessation (66,67). Although this data does not directly relate SLC22A14 to nicotinate, it suggests a possible route of investigation into

the functional role of this transporter, one that may, in some ways, overlap with that of OAT10. SLC22A13 is primarily expressed in the kidney, and although we found no human protein expression data for SLC22A14, transcripts for this gene are found at low levels in the kidney and notably high levels in the testis (Table 4), which is in concordance with its critical role in sperm motility and fertility in male mice (68). Future studies are required to determine the functional classification of this subgroup; however, our genomic localization and sequence-based analyses provide enough data to support the notion that these two belong in their own individual subgroup.

OAT-Related (SLC22A17, SLC22A18, SLC22A23, SLC22A31) is Anomalous Amongst SLC22 Members but has Interesting Functional Mechanisms and Disease Associations

The OAT-related subgroup is an outlier within the SLC22 family, consisting of the orphan transporters SLC22A17, SLC22A18, SLC22A23, and SLC22A31. SLC22A17 and SLC22A23 are strongly related, with greater than 30% shared amino acid identity. When these two transporters were initially identified together as BOCT1 (SLC22A17) and BOCT2 (SLC22A23), it was noted that they both show high expression levels in the brain, as well as a nonconserved amino terminus that may negate prototypical SLC22 function (69). SLC22A17 is known as LCN2-R (Lipocalin receptor 2) and mediates iron homeostasis through binding and endocytosis of iron-bound lipocalin, as well as exhaustive protein clearance from the urine as shown by high affinities for proteins such as calbindin (20,70). SLC22A23 has no confirmed substrates, but SNPs and mutations within this gene have medically-relevant phenotypic associations such as QT elongation, inflammatory bowel disease, endometriosis-related infertility, and the clearance of antipsychotic drugs (71–73). SLC22A31 is the most understudied transporter of the SLC22 family but has been associated with right-side colon cancer (74).

SLC22A18 remains an outlier and lacks the characteristic SLC22 large ECD. Its membership within the SLC22 family is arguable due to high sequence similarity with the DHA H⁺-antiporter family (Supplementary Figure 4) (10). Further study is required to confirm if the OAT-related members share substrates or if their unusual characteristics and associations are the reason for their phylogenetic association.

OCT (SLC22A1, SLC22A2, SLC22A3) Members are Characteristic Organic Cation Transporters with High Affinities for Monoamine Neurotransmitters and Other Biologically Important Metabolites and Signaling Molecules

The OCT subclade of SLC22A1 (OCT1), SLC22A2 (OCT2), and SLC22A3 (OCT3) has ample data to support its formation and has been widely accepted and utilized as the prototypical subgroup of organic cation transporters. All three members of this subgroup transport monoamine neurotransmitters, carnitine derivatives, creatinine and the characteristic OCT substrates, MPP⁺ and TEA (Supplementary Figure 2A) (38,75–79). All three members of this subgroup are expressed in the liver, kidney, and brain (Table 4). When taken with the transport of neurotransmitters, this subgroup serves as an example of inter-organ communication between the brain and the kidney-liver axis via transporters. The systemic levels of these neurotransmitters and thus, their availability to the brain can be regulated by the expression of OCT subgroup members in the liver, where the metabolites can be enzymatically modified, and expression in the kidney, which may serve as an excretory route (7).

OCTN/OCTN-Related (SLC22A4, SLC22A5, SLC22A15, SLC22A16) Subgroup Consists of Prototypical Carnitine and Ergothioneine Transporters

The OCTN/OCTN-Related subgroup is a combination of two previously established subclades, OCTN and OCTN-related (10). Previous studies have mistakenly named SLC22A15

as CT1 (carnitine transporter 1), but this name actually belongs to SLC22A5 (OCTN2) (13). GWAS data show that SLC22A4 (OCTN1), SLC22A5 (OCTN2/CT1), and SLC22A16 (CT2/FLIPT2) are heavily linked to carnitine and its derivatives (38). This is consistent with in vitro data showing that OCTN2 and CT2 are carnitine transporters (80,81). Although OCTN1 has lower affinity for carnitine than OCTN2 and CT2, it has high affinity for the endogenous antioxidant ergothioneine, which GWAS data suggests may be a shared metabolite with both SLC22A15 (FLIPT1) and CT2 (Supplementary Figure 2B) (38,82). SLC22A15 is associated with many complex lipids that are not characteristic of any other SLC22 transporter (60). This anomalous SLC22 member may only share one substrate with this subgroup, but its inclusion is supported by multiple sequence alignments focusing on the intracellular loop and tissue expression patterns. Most other subgroups in this family are limited to a few tissues, mainly the liver and kidney, but the members of the OCTN/OCTN-Related subgroup are all expressed in at least five tissues as well as circulating immune cells (Table 4) (4,7). This broad tissue expression pattern, in conjunction with our network analysis, supports the belief that these transporters' main task is transporting carnitine derivatives, as carnitine metabolism is an energy producing mechanism in nearly every cell. It may also play a role in regulating levels of the antioxidant ergothioneine, which is uniquely a substrate of this subgroup (41,83).

Multiple Sequence Alignment Further Supports the Classification of Subgroups

Our new subgroupings are primarily based on the endogenous function of the transporters, but they are also supported by additional analyses. These analyses are necessary, as structural and evolutionary similarities can predict functional traits that have yet to be discovered. Though the previously established phylogenetic subclades remain sound, our re-analysis includes new and updated amino acid sequences that support the proposed subgroups

with more confidence, especially when investigating similarities within functional regions (10). MSA programs were favored over phylogenetics because MSA searches are based upon structural similarities rather than evolutionary relatedness (84). These structural similarities, especially in the large ECD and large ICD of SLC22 proteins, may indicate shared function.

Full length sequence analysis via Clustal-Omega, MAFFT, and ICM-Pro v3.8-7 supports the division of SLC22 into eight subgroups (Figure 2A, Supplementary Figure 5, Supplementary Figure 6). While the OATS1, OATS2, OATS4, OAT-like, OAT-related, and OCT subgroups are supported by full-length sequence analyses, OATS3 and OCTN/OCTN-Related required a more rigorous investigation. To further clarify “borderline” subgroup assignments from the full-length sequence analysis, sequence similarity between the ECDs and ICDs of all human and mouse SLC22 members was determined using ICM-Pro v3.8-7, and the results were visualized via guide trees (Figure 2B, Figure 2C). ECD alignment preserved all eight subgroups, with the exception of SLC22A15 in the OCTN/OCTN-Related subgroup. In contrast, ICD alignment preserved only the OATS4, OATS2, and OCT subgroups.

The branching pattern of OATS3 member Oat-pg (*Slc22a22*) differs between tree variations. These analyses consistently indicate a similar relationship between Oat-pg and OATS3, as well as OATS4. However, in an analysis of the SLC22 ECDs, it is most closely associated with OATS3 over any other subgroup. This, in conjunction with shared substrate specificity with both SLC22A12 and SLC22A11, and not OATS4 members, supports its membership within the OATS3 subgroup (46,52,53,59).

In full-length sequence alignments, the grouping of SLC22A4, SLC22A5, and *Slc22a21* is consistently conserved, while the topology of both SLC22A15 and SLC22A16 is irregular. Despite this, analysis of the large ECD shows similarity between all OCTN/OCTN-related

members other than SLC22A15. Previous analyses have noted the large difference between the ECD of SLC22A15 and all other SLC22 members which is supported by our analysis in Figure 2B (10). Interestingly, there appears to be some similarity between the large intracellular domains of SLC22A16 and SLC22A15. Although much of the support for the establishment of the OCTN/OCTN-related subgroup comes from functional data (Supplementary Figure 2B), the described MSA analyses highlight shared structural, and possibly functional, regions.

Analysis of Genomic Localization Highlights Evolutionary Relatedness of Subgroup Members and Suggests Basis of Coregulation

Genomic clustering within the SLC22 family has been previously described (10). Specifically, genes found in tandem on the chromosome, such as OAT-like members SLC22A13 and SLC22A14, are predicted to have arisen from duplication events, indicating a strong evolutionary relationship. Despite the majority of the OAT subclade being found on chromosome 11 in humans and chromosome 19 in mice, clustering within the chromosome supports the division of the OAT subclade into smaller subgroups. For example, OATS4 members SLC22A9, SLC22A10, SLC22A24, and SLC22A25 appear in tandem on human chromosome 11 within the UST (Unknown Substrate Transporter) region of the genome. This region is analogous to the UST region within the mouse genome on chromosome 19, where the mouse-specific OATS4 members Slc22a19, Slc22a26, Slc22a27, Slc22a28, Slc22a29, and Slc22a30 reside as well as the rat UST region on chromosome 1 that contains Slc22a9/a24, Slc22a9/a25, Ust4r and Ust5r (Table 4) (85,86). It has been proposed that genes within clusters are coordinately regulated and thus are predicted to have similar overall tissue expression patterns. Support for shared regulatory mechanisms of subgroup members within genomic clusters can be inferred from similar patterns of tissue expression or by expression of subgroup members along a common axis of metabolite transport such as the gut-kidney-liver axis.

Genomic localization from the UCSC Genome browser and resultant tissue expression patterns for all SLC22 members are shown in Table 4.

Analysis of OAT Subgroup Specific Motifs Highlight Patterns Potentially Involved in Specificity

Motif analyses revealed subgroup specific motifs within functionally important regions, such as the large ICD, large ECD, and the region spanning TMD9 and TMD10, for all novel OAT subgroups (10,87). However, the number of unique residues appears to be correlated to the range of substrate specificity.

Of the newly proposed OAT subgroups, OATS2 claims the smallest number of subgroup-specific amino acid motifs and is the only subgroup without a specific motif in TMD9 (Figure 3B). The lack of multiple subgroup-specific regions is interesting not only because this subgroup consists of a single transporter but also because this may be indicative of a more promiscuous transporter with a wide range of substrates, which is substantiated by the functional data. This pattern is also seen in OATS1, which consists of multi- and oligo-specific transporters OAT1, OAT3, and OAT6. In addition to having few subgroup-specific motifs, the multi/oligo-specific nature of this subgroup is reflected by the shared evolutionary conservation of the large extracellular domain with other OAT subclade members (Figure 3A).

To further clarify the membership of Oat-pg in OATS3, evolutionarily conserved motifs were determined between all three members, as well as just Slc22a11 and Slc22a12. This analysis revealed a total of ten evolutionarily conserved amino acid motifs between all three members, eight of which are present in the analysis of only OAT4 and URAT1. Specifically, both analyses exhibited a notably large motif in the large intracellular loop found at D313-Q332 on URAT1 and Q312-G331 on OAT-PG (Figure 3C, Figure 3D). This larger number of conserved regions correlates with a more limited range of substrates (eg. uric acid and prostaglandins) (56).

Motif analysis was performed separately on the OATS4 rodent and non-rodent specific subgroups and the entirety of the OATS4 subgroup members. In all analyses, OATS4 claims the largest number of evolutionarily conserved and subgroup-specific amino acid residues amongst the OAT subgroups, indicating selective substrate specificity, possibly for conjugated sex steroids (Figure 3E, Figure 3F). In the case of non-rodent transporters, a unique motif spans the sixth extracellular domain and TMD12. This region is predicted to govern substrate specificity of transporters of the MFS, to which the SLC22 family belongs (87). Recent publications defining the substrate specificity of SLC22A24 point to a more narrow range of substrates and conservation of this specific region amongst OATS4 members may explain the association of conjugated steroid hormones with SLC22A9, SLC22A10, SLC22A24, and SLC22A25 in GWAS studies (60,61). Although further analysis is required to fully understand the relationship between structure and substrate specificity in SLC22 transporters, we provided a basis for investigation into specific regions that may determine functional patterns. The sequences and p-values for each motif are in Supplementary Tables 2-7.

Sequence Similarity Study Suggests Novel Potential Functions and Possible Tertiary Structure of SLC22

Each SLC22 member is a putative transporter, but there is evidence that suggests some members may have alternative mechanisms of action ((48,70)). To further explore this possibility and to potentially find sequence similarity to other proteins, the specific amino acid sequences for the extracellular and intracellular loops of each SLC22 member were compared to all proteins in the ICM-Pro v3.8-7c database. The extracellular loop of mouse Slc22a16 shares 26% sequence identity (pP=5.4) with chicken beta-crystallin B3 (CRBB3). Beta-crystallin is a structural protein mainly comprised of beta sheets (88). The similarity between the ECD of

mouse Slc22a16 and CRBB3 could point to potential for a beta sheet like configuration. Because none of the SLC22 family members have been crystallized, any insight into tertiary structure is of interest.

SLC22A31, a member of the divergent OAT-Related subclade, is the most ambiguous member of the SLC22 family with no functional data available. An investigation of the human SLC22A31 large ECD shows at least 30% shared sequence identity with RNA-binding protein 42 (RBM42) in mouse, rat, cow and human. This analysis also showed a 37% sequence identity (pP=5.5) shared between the ECD of hSLC22A31 and human heterochromatinization factor BAHD1. These and other interesting sequence similarities are noted in Table 5.

Discussion

In the years following the establishment of the previous SLC22 subclades, there has been a notable increase in functional data concerning these transporters and their substrates (10). With these data, we are now in a position to better characterize these transporters, which play important physiological roles. However, our newly proposed subgroups are not entirely dependent on functional data, as we have considered multiple approaches including phylogenetics, multiple sequence alignments, evolutionarily conserved motifs, sequence homology, and both tissue and genomic localization. Each of these approaches has individual value in that they reveal unique characteristics of each transporter, but it is the combination of multiple approaches that ensures nearly all available data (though still incomplete) for these transporters is considered when forming functional subgroups. We support the subgroups with a thorough literature search of metabolites associated with SLC22 proteins. Although the functional data is inherently biased due to the high level of interest in some SLC22 members, particularly the “drug” transporters OAT1, OAT3, OCT1, and OCT2, for the majority of the transporters, there is enough data to create functional subgroups that play distinct and overlapping roles in metabolism (Figure 1, Supplementary Table 1). Genomic localization reveals evolutionary information and provides insight on how genes may arise from duplication events. Phylogenetic analysis determines the evolutionary relatedness of these proteins, while MSA, motif analysis, and sequence homology focus on structural similarities, which can be indicative of function. We often see that members of a subgroup are expressed in the same tissues or along functional axes. For example, substrates transported from the liver via SLC22 transporters (e.g., SLC22A1, OCT1) can be either excreted into or retrieved from the urine by other SLC22 members (SLC22A2, OCT2) of the same subgroup. Establishment of these

functional subgroups may also inform future virtual screenings for metabolites of understudied transporters.

Protein families are established based on shared ancestry and structural similarity, which is commonly considered grounds for shared functionality. This is exemplified amongst SLC22 members with the generally shared structural characteristics of 12 TMDs, a large extracellular loop between TMD1 and TMD2, and a smaller intracellular loop between TMD6 and TMD7. Despite these shared features, we show here that there are many functional differences between these transporters. Although our analyses mostly align with previous evolutionary studies when considering ancestry, here, we show that phylogenetic grouping is not always reflective of similar structure and function. For example, although the previously established OCTN subclade of SLC22A4, SLC22A5 and Slc22a21 does not share common ancestry with Slc22a16, the newly proposed group shares functional similarity and ECD homology. Thus, by expanding our investigation beyond phylogenetic relationships, we can now more appropriately group proteins from the same family and better understand their roles in endogenous physiology.

An important concept in the Remote Sensing and Signaling Network is that of multi-specific, oligo-specific, and relatively mono-specific transporters working in a coordinated function (16). Multi-specific transporters are able to interact with a wide variety of structurally different compounds, oligo-specific with a smaller variety, and relatively mono-specific transporters are thought to interact with only one or a few substrates. Existing functional data suggests that it is unlikely that any truly mono-specific transporters exist within the SLC22 family, yet the different subgroups we have formed imply that multi-specific, oligo-specific, and relatively mono-specific transporters are more likely to form subgroups with transporters that share substrate specificity. Multi-specific transporters, like those in the OATS1 and OCT

subgroups, handle a diverse set of drugs, toxins, endogenous metabolites, and signaling molecules (14,75). Conversely, the OATS4 subgroup appears to be a collection of relatively mono-specific transporters with an affinity for conjugated sex steroid hormones, specifically etiocholanolone glucuronide (61). Previous evolutionary studies have suggested that multi-specific transporters arose before the mono-specific transporters (10). As evolution has progressed, more specific transporters have developed to handle the burden of changing metabolism. The multi-specific transporters have been more extensively characterized because of their importance in pharmaceuticals, but in the case of endogenous metabolic diseases, the oligo and mono-specific transporters may be more appropriate targets for drugs or therapies.

One of the best examples of multi-specific transporters working in concert with oligo, and mono-specific transporters is the regulation of uric acid (17,18). Handling of uric acid mainly occurs in the kidney, but when renal function is compromised, multi-specific transporters regulate their expression to compensate. Two proteins, SLC22A12 and SLC2A9, are expressed in the proximal tubule and are nearly exclusively associated with uric acid. The multi-specific transporters SLC22A6 and SLC22A8 are also present in the proximal tubule and are able to transport uric acid. When the kidney is damaged, one would expect serum uric acid levels to increase because most of the proteins involved in its elimination are in the kidney. However, this is alleviated due to the increased expression of ABCG2 and/or functional activity in the intestine (17,18). SLC2A9 is a relatively mono-specific transporter and ABCG2 (BCRP) is a multi-specific ABC transporter. The example of uric acid serves to illustrate how, when certain mono-, oligo-, and multi-specific transporters are unable to perform their primary function, multi-specific transporters of the same or different function (even of the ABC superfamily) can use their shared substrate specificity to mitigate the consequences. It is generally assumed that all

SLC22 family members are transporters. However, Slc22a17, a member of the outlier Oat-related subclade, functions as an endocytosed iron-bound lipocalin receptor and some SLC22 members have been suggested to function as “transceptors” due to homology with GPCR odorant receptors and shared odorant substrates (20,21). Thus, to better understand the SLC22 family members’ individually unique functions and their placement into subgroups/subclades, we compared the full-length amino acid sequences, large ECDs, and large ICDs of all SLC22 family members to a database of known proteins.

When considering such a large number of proteins, the function on both local and systemic levels of metabolites is likely to be impacted. The SLC22 family is a central hub of coexpression for ADME (absorption, distribution, metabolism, excretion)-related genes in non-drug treated conditions, which underscores their importance in regulating endogenous metabolism through the transport of small molecules (16). In the context of the Remote Sensing and Signaling Theory (RSST), it is essential to understand substrate specificity of different SLC22 members and the 8 subgroups. The RSST would seem to imply that organisms are constantly solving a multi-objective optimization problem, where balancing each particular compound’s serum concentration represents a single objective. Each compound present in the blood has a range of healthy concentrations, and when the concentration is outside of that range, the body must address it, in part through the regulation of transporters and enzymes. Due to their wide range of tissue expression and diverse functional roles at body fluid interfaces, the particular combination of transporters and enzymes are critical variables necessary for solving this multi-objective optimization problem. Transporters regulate the entry and exit of substrates to and from cells, but enzymes are responsible for the altering of these compounds. To use a simple hypothetical example, if a metabolite’s serum concentration is too high, a transporter with

high affinity for that metabolite can move it into the cell, where an enzyme with high affinity for the substrate can change it so that it may re-enter the circulation or be more readily cleared from the body. The existence of abundant multi-specific and oligo-specific transporters and enzymes, in addition to relatively mono-specific ones, expressed differentially in tissues and at body fluid interfaces, allows for a highly flexible and responsive complex adaptive system that not only maintains homeostasis in blood, tissue, and body fluid compartments (e.g. CSF), but also helps restore it after acute or chronic perturbations. Together, transporters and enzymes have tremendous potential to manage levels of metabolites, signaling molecules, and antioxidants in the circulation and in specific tissues. By developing functional groupings for the SLC22 family, we can better understand the metabolic networks in which they function and how their expression is utilized to regulate concentrations of metabolites, signaling molecules (e.g. cyclic nucleotides, prostaglandins, short chain fatty acids, and sex steroids), antioxidants (ergothioneine, uric acid), and other molecules affecting diverse aspects of homeostasis (e.g. lipocalin). Although this analysis focuses on the SLC22 family, a similar approach can be applied to develop a deeper understanding of other families of transporters and enzymes.

In the past, the majority of functional data has come from hypothesis-driven transport assays using cells overexpressing a specific SLC22 transporter and a single metabolite of interest. These assays lack uniformity and, as the OAT knockouts have shown, are not necessarily reflective of endogenous physiology (43,45,47). Recently, GWAS studies have linked many metabolites to polymorphisms in SLC22 genes, and in vivo metabolomic studies using knockout models have also identified several metabolites that may be substrates of transporters (43,45,47,60). In upcoming years, the integration of multiple types of omics data related to SLC22 family members with functional studies of transporters and evolutionary

analyses will likely produce a more fine-grained picture of the roles of these and other transporters in inter-organ and inter-organismal Remote Sensing and Signaling.

Acknowledgements

This work was supported by the National Institutes of Health by grants R01DK109392 and R01GM132938 awarded to S.K.N. and the training grant T32EB009380 awarded to J.C.G.

References

1. Hediger MA, Clémenton B, Burrier RE, Bruford EA. The ABCs of membrane transporters in health and disease (SLC series): Introduction. Vol. 34, *Molecular Aspects of Medicine*. 2013. p. 95–107.
2. Lamhonwah AM, Hawkins CE, Tam C, Wong J, Mai L, Tein I. Expression patterns of the organic cation/carnitine transporter family in adult murine brain. *Brain Dev*. 2008 Jan;30(1):31–42.
3. Sager G, Smaglyukova N, Fuskevaag OM. The role of OAT2 (SLC22A7) in the cyclic nucleotide biokinetics of human erythrocytes. *J Cell Physiol*. 2018 Aug 1;233(8):5972–80.
4. Minuesa G, Purcet S, Erkizia I, Molina-Arcas M, Bofill M, Izquierdo-Useros N, et al. Expression and functionality of anti-human immunodeficiency virus and anticancer drug uptake transporters in immune cells. *J Pharmacol Exp Ther*. 2008 Feb;324(2):558–67.
5. César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, et al. A Call for Systematic Research on Solute Carriers [Internet]. Vol. 162, *Cell*. Austria: Elsevier; 2015 [cited 2019 Oct 9]. p. 478–87. Available from: <http://www.cell.com/article/S0092867415008995/fulltext>
6. Lelandais-Brière C, Jovanovic M, Torres GAM, Perrin Y, Lemoine R, Corre-Menguy F, et al. Disruption of AtOCT1, an organic cation transporter gene, affects root development and carnitine-related responses in Arabidopsis. *Plant J [Internet]*. 2007 Jul [cited 2019 Nov 18];51(2):154–64. Available from: <http://doi.wiley.com/10.1111/j.1365-313X.2007.03131.x>
7. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science (80-)*. 2015 Jan 23;347(6220).

8. Lopez-Nieto CE, You G, Bush KT, Barros EJG, Beier DR, Nigam SK. Molecular cloning and characterization of NKT, a gene product related to the organic cation transporter family that is almost exclusively expressed in the kidney. *J Biol Chem.* 1997 Mar 7;272(10):6471–8.
9. Saier MH, Reddy VS, Tamang DG, Västermark Å. The transporter classification database. *Nucleic Acids Res.* 2014 Jan 1;42(D1).
10. Zhu C, Nigam KB, Date RC, Bush KT, Springer SA, Saier MH, et al. Evolutionary analysis and classification of OATs, OCTs, OCTNs, and other SLC22 transporters: Structure-function implications and analysis of sequence motifs. *PLoS One* [Internet]. 2015 [cited 2019 Sep 27];10(11):140569. Available from: <http://www.ncbi.nlm.nih.gov/>
11. Nigam SK. The SLC22 Transporter Family: A Paradigm for the Impact of Drug Transporters on Metabolic Pathways, Signaling, and Disease. *Annu Rev Pharmacol Toxicol.* 2018 Jan 6;58(1):663–87.
12. Ahn SY, Nigam SK. Toward a systems level understanding of organic anion and other multispecific drug transporters: A remote sensing and signaling hypothesis. Vol. 76, *Molecular Pharmacology.* 2009. p. 481–90.
13. Wu W, Dnyanmote A V., Nigam SK. Remote communication through solute carriers and ATP binding cassette drug transporter pathways: An update on the Remote Sensing and Signaling Hypothesis. Vol. 79, *Molecular Pharmacology.* 2011. p. 795–805.
14. Nigam SK, Bush KT, Martovetsky G, Ahn SY, Liu HC, Richard E, et al. The organic anion transporter (OAT) family: A systems biology perspective. *Physiol Rev.* 2015 Jan 1;95(1):83–123.
15. Nigam SK. What do drug transporters really do? Vol. 14, *Nature Reviews Drug*

- Discovery. Nature Publishing Group; 2014. p. 29–44.
16. Rosenthal SB, Bush KT, Nigam SK. A Network of SLC and ABC Transporter and DME Genes Involved in Remote Sensing and Signaling in the Gut-Liver-Kidney Axis. *Sci Rep* [Internet]. 2019 Aug 15 [cited 2019 Oct 9];9(1):11879. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31417100>
 17. Bhatnagar V, Richard EL, Wu W, Nievergelt CM, Lipkowitz MS, Jeff J, et al. Analysis of ABCG2 and other urate transporters in uric acid homeostasis in chronic kidney disease: Potential role of remote sensing and signaling. *Clin Kidney J*. 2016;9(3):444–53.
 18. Yano H, Tamura Y, Kobayashi K, Tanemoto M, Uchida S. Uric acid transporter ABCG2 is increased in the intestine of the 5/6 nephrectomy rat model of chronic kidney disease. *Clin Exp Nephrol*. 2014;18(1):50–5.
 19. Nigam SK, Bhatnagar V. The systems biology of uric acid transporters: The role of remote sensing and signaling. Vol. 27, *Current Opinion in Nephrology and Hypertension*. Lippincott Williams and Wilkins; 2018. p. 305–13.
 20. Langelueddecke C, Roussa E, Fenton R a, Wolff N a, Lee WK, Thévenod F. Lipocalin-2 (24p3/neutrophil gelatinase-associated lipocalin (NGAL)) receptor is expressed in distal nephron and mediates protein endocytosis. *J Biol Chem* [Internet]. 2012 Jan 2 [cited 2019 Oct 4];287(1):159–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22084236>
 21. Kaler G, Truong DM, Sweeney DE, Logan DW, Nagle M, Wu W, et al. Olfactory mucosa-expressed organic anion transporter, Oat6, manifests high affinity interactions with odorant organic anions. *Biochem Biophys Res Commun* [Internet]. 2006 Dec 29 [cited 2019 Nov 18];351(4):872–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17094945>

22. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* [Internet]. 2018 [cited 2019 Oct 9];46(D1):D762–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29106570>
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
24. Madeira F, Park Y mi, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019 Jul 2;47(W1):W636–41.
25. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018 Jan 1;27(1):135–45.
26. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7.
27. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013 Apr;30(4):772–80.
28. Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* [Internet]. 1994 May [cited 2019 Nov 18];15(5):488–506. Available from: <http://doi.wiley.com/10.1002/jcc.540150503>
29. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019 Jul 2;47(W1):W256–9.
30. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987 Jul;

31. Reddy VS, Saier MH. BioV Suite - A collection of programs for the study of transport protein evolution. *FEBS J.* 2012 Jun;279(11):2036–46.
32. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981 Mar 25;147(1):195–7.
33. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2:28–36.
34. Dobson L, Reményi I, Tusnády GE. CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* 2015;43(W1):W408–12.
35. Johns SJ. TOPO2, Transmembrane protein display software [Internet]. Available from: <http://www.sacs.ucsf.edu/TOPO2/>
36. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res [Internet].* 2018 Jan 4 [cited 2019 Dec 17];46(D1):D1074–82. Available from: <http://academic.oup.com/nar/article/46/D1/D1074/4602867>
37. Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature.* 2011 Sep 1;477(7362):54–62.
38. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet.* 2014;46(6):543–50.
39. Kastenmüller G, Raffler J, Gieger C, Suhre K. Genetics of human metabolism: An update [Internet]. Vol. 24, *Human Molecular Genetics.* 2015 [cited 2019 Nov 18]. p. R93–101. Available from: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddv263>
40. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for

- data integration and network visualization. *Bioinformatics* [Internet]. 2011 Feb 1 [cited 2019 Dec 17];27(3):431–2. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq675>
41. Bremer J. Carnitine. Metabolism and functions. Vol. 63, *Physiological Reviews*. 1983. p. 1420–80.
 42. Blumberg AL, Denny SE, Marshall GR, Needleman P. Blood vessel hormone interactions: angiotensin, bradykinin, and prostaglandins. *Am J Physiol*. 1977;232(3).
 43. Wikoff WR, Nagle MA, Kouznetsova VL, Tsigelny IF, Nigam SK. Untargeted metabolomics identifies enterobiome metabolites and putative uremic toxins as substrates of organic anion transporter 1 (Oat1). *J Proteome Res*. 2011 Jun 3;10(6):2842–51.
 44. Vallon V, Eraly SA, Rao SR, Gerasimova M, Rose M, Nagle M, et al. A role for the organic anion transporter OAT3 in renal creatinine secretion in mice. *Am J Physiol - Ren Physiol*. 2012 May 15;302(10).
 45. Wu W, Bush KT, Nigam SK. Key Role for the Organic Anion Transporters, OAT1 and OAT3, in the in vivo Handling of Uremic Toxins and Solutes. *Sci Rep*. 2017 Dec 1;7(1).
 46. Kimura H, Takeda M, Narikawa S, Enomoto A, Ichida K, Endou H. Human organic anion transporters and human organic cation transporters mediate renal transport of prostaglandins. *J Pharmacol Exp Ther* [Internet]. 2002 Apr [cited 2019 Oct 4];301(1):293–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11907186>
 47. Bush KT, Wu W, Lun C, Nigam SK. The drug transporter OAT3 (SLC22A8) and endogenous metabolite communication via the gut–liver– kidney axis. *J Biol Chem*. 2017 Sep 22;292(38):15789–803.

48. Wu W, Bush KT, Liu HC, Zhu C, Abagyan R, Nigam SK. Shared ligands between organic anion transporters (OAT1 and OAT6) and odorant receptors. *Drug Metab Dispos*. 2015 Dec 1;43(12):1855–63.
49. Shen H, Liu T, Morse BL, Zhao Y, Zhang Y, Qiu X, et al. Characterization of Organic Anion Transporter 2 (SLC22A7): A Highly Efficient Transporter for Creatinine and Species-Dependent Renal Tubular Expression. *Drug Metab Dispos* [Internet]. 2015 Jul [cited 2019 Oct 3];43(7):984–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25904762>
50. Köttgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* [Internet]. 2013 Feb [cited 2019 Oct 4];45(2):145–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23263486>
51. Sun W, Wu RR, Van Poelje PD, Erion MD. Isolation of a family of organic anion transporters from human liver and kidney. *Biochem Biophys Res Commun*. 2001;283(2):417–22.
52. Enomoto A, Kimura H, Chairoungdua A, Shigeta Y, Jutabha P, Cha SH, et al. Molecular identification of a renal urate-anion exchanger that regulates blood urate levels. *Nature* [Internet]. 2002 May 23 [cited 2019 Oct 4];417(6887):447–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12024214>
53. Hagos Y, Stein D, Ugele B, Burckhardt G, Bahn A. Human renal organic anion transporter 4 operates as an asymmetric urate transporter. *J Am Soc Nephrol* [Internet]. 2007 Feb [cited 2019 Oct 4];18(2):430–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17229912>

54. Maxwell SRJ, Thomason H, Sandler D, Leguen C, Baxter MA, Thorpe GHG, et al. Antioxidant status in patients with uncomplicated insulin-dependent and non-insulin-dependent diabetes mellitus. *Eur J Clin Invest*. 1997;27(6):484–90.
55. Sakiyama M, Matsuo H, Shimizu S, Nakashima H, Nakayama A, Chiba T, et al. A common variant of organic anion transporter 4 (OAT4/SLC22A11) gene is associated with renal underexcretion type gout. *Drug Metab Pharmacokinet* [Internet]. 2014 [cited 2019 Oct 4];29(2):208–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24025986>
56. Cha SH, Sekine T, Kusuhara H, Yu E, Kim JY, Kim DK, et al. Molecular cloning and characterization of multispecific organic anion transporter 4 expressed in the placenta. *J Biol Chem*. 2000 Feb 11;275(6):4507–12.
57. Tomi M, Eguchi H, Ozaki M, Tawara T, Nishimura S, Higuchi K, et al. Role of OAT4 in uptake of estriol precursor 16 α -hydroxydehydroepiandrosterone sulfate into human placental syncytiotrophoblasts from fetus. *Endocrinology* [Internet]. 2015 Jul [cited 2019 Oct 4];156(7):2704–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25919187>
58. Skwara P, Schömig E, Gründemann D. A novel mode of operation of SLC22A11: Membrane insertion of estrone sulfate versus translocation of uric acid and glutamate. *Biochem Pharmacol*. 2017 Mar 15;128:74–82.
59. Shiraya K, Hirata T, Hatano R, Nagamori S, Wiriyasermkul P, Jutabha P, et al. A novel transporter of SLC22 family specifically transports prostaglandins and co-localizes with 15-hydroxyprostaglandin dehydrogenase in renal proximal tubules. *J Biol Chem* [Internet]. 2010 Jul 16 [cited 2019 Oct 4];285(29):22141–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20448048>

60. Long T, Hicks M, Yu HC, Biggs WH, Kirkness EF, Menni C, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet.* 2017 Mar 30;49(4):568–78.
61. Wah Yee S, Stecula A, Chien H-C, Zou L, Feofanova E V., Borselen V, et al. Unraveling the functional role of the orphan solute carrier, SLC22A24 in the transport of 2 steroid conjugates through metabolomic and genome-wide association studies 3 4. Evans WE, editor. *PLOS Genet* [Internet]. 2019 Sep 25 [cited 2019 Oct 2];15(9):e1008208. Available from: <http://dx.plos.org/10.1371/journal.pgen.1008208>
62. Ho JS, Anzai N, Enomoto A, He X, Do KK, Endou H, et al. Novel liver-specific organic anion transporter OAT7 that operates the exchange of sulfate conjugates for short chain fatty acid butyrate. *Hepatology.* 2007 Apr;45(4):1046–55.
63. Sardiello M, Annunziata I, Roma G, Ballabio A. Sulfatases and sulfatase modifying factors: An exclusive and promiscuous relationship. *Hum Mol Genet* [Internet]. 2005 Nov 1 [cited 2019 Dec 8];14(21):3203–17. Available from: <http://academic.oup.com/hmg/article/14/21/3203/2355820/Sulfatases-and-sulfatase-modifying-factors-an>
64. Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, et al. Expression Atlas: Gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* [Internet]. 2018 [cited 2019 Oct 2];46(D1):D246–51. Available from: <http://europepmc.org/abstract/med/29165655>
65. Bahn A, Hagos Y, Reuter S, Balen D, Brzica H, Krick W, et al. Identification of a new urate and high affinity nicotinate transporter - human organic anion transporter 10 (hOAT10, SLC22A131). *Journal of Biological Chemistry.* p.

66. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, et al. Molecular genetics of successful smoking cessation: Convergent genome-wide association study results. *Arch Gen Psychiatry*. 2008 Jun;65(6):683–93.
67. Uhl GR, Drgon T, Li CY, Johnson C, Liu QR. Smoking and smoking cessation in disadvantaged women: Assessing genetic contributions [Internet]. Vol. 104, *Drug and Alcohol Dependence*. 2009 [cited 2019 Oct 4]. p. S58-63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19442458>
68. Maruyama SY, Ito M, Ikami Y, Okitsu Y, Ito C, Toshimori K, et al. A critical role of solute carrier 22a14 in sperm motility and male fertility in mice. *Sci Rep*. 2016 Nov 4;6.
69. Bennett KM, Liu J, Hoelting C, Stoll J. Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23. *Mol Cell Biochem* [Internet]. 2011 Jun [cited 2019 Oct 4];352(1–2):143–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21359964>
70. Wan-Jie C, Wei-Yi O, Gavin D. Lipocalin 2 and lipocalin 2 receptor in neuron-glia interactions following injury. *Front Cell Neurosci*. 2016;10.
71. Zhao ZZ, Croft L, Nyholt DR, Chapman B, Treloar SA, Hull ML, et al. Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. *Mol Hum Reprod* [Internet]. 2011 Feb 1 [cited 2019 Dec 17];17(2):92–103. Available from: <https://academic.oup.com/molehr/article-lookup/doi/10.1093/molehr/gaq084>
72. Serrano León A, Amir Shaghghi M, Yurkova N, Bernstein CN, El-Gabalawy H, Eck P. Single-nucleotide polymorphisms in SLC22A23 are associated with ulcerative colitis in a Canadian white cohort. *Am J Clin Nutr* [Internet]. 2014 Jul 1 [cited 2019 Dec

- 17];100(1):289–94. Available from:
<https://academic.oup.com/ajcn/article/100/1/289/4576586>
73. Åberg K, Adkins DE, Liu Y, McClay JL, Bukszár J, Jia P, et al. Genome-wide association study of antipsychotic-induced QTc interval prolongation. *Pharmacogenomics J* [Internet]. 2012 Apr [cited 2019 Oct 4];12(2):165–72. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/20921969>
74. Bien SA, Su YR, Conti D V., Harrison TA, Qu C, Guo X, et al. Genetic variant predictors of gene expression provide new insight into risk of colorectal cancer. *Hum Genet*. 2019 Apr 1;138(4):307–26.
75. Amphoux A, Vialou V, Drescher E, Brüß M, La Cour CM, Rochat C, et al. Differential pharmacological in vitro properties of organic cation transporters and regional distribution in rat brain. *Neuropharmacology*. 2006 Jun;50(8):941–52.
76. Busch AE, Quester S, Ulzheimer JC, Gorboulev V, Akhoundova A, Waldegger S, et al. Monoamine neurotransmitter transport mediated by the polyspecific cation transport rOCT1. *FEBS Lett* [Internet]. 1996 Oct 21 [cited 2019 Nov 23];395(2–3):153–6. Available from: <http://doi.wiley.com/10.1016/0014-5793%2896%2901030-7>
77. Gründemann D, Liebich G, Kiefer N, Köster S, Schömig E. Selective substrates for non-neuronal monoamine transporters. *Mol Pharmacol* [Internet]. 1999 [cited 2019 Nov 24];56(1):1–10. Available from: <http://www.molpharm.org>
78. Urakami Y, Kimura N, Okuda M, Inui KI. Creatinine transport by basolateral organic cation transporter hOCT2 in the human kidney. *Pharm Res*. 2004 Jun;21(6):976–81.
79. Wu X, Kekuda R, Huang W, Fei YJ, Leibach FH, Chen J, et al. Identity of the organic cation transporter OCT3 as the extraneuronal monoamine transporter (uptake2) and

- evidence for the expression of the transporter in the brain. *J Biol Chem.* 1998 Dec 4;273(49):32776–86.
80. Enomoto A, Wempe MF, Tsuchida H, Shin HJ, Cha SH, Anzai N, et al. Molecular identification of a novel carnitine transporter specific to human testis: Insights into the mechanism of carnitine recognition. *J Biol Chem.* 2002 Sep 27;277(39):36262–71.
81. Ohashi R, Tamai I, Yabuuchi H, Nezu JI, Oku A, Sai Y, et al. Na⁺-dependent carnitine transport by organic cation transporter (OCTN2): Its pharmacological and toxicological relevance. *J Pharmacol Exp Ther.* 1999 Nov;291(2):778–84.
82. Gründemann D, Harlfinger S, Golz S, Geerts A, Lazar A, Berkels R, et al. Discovery of the ergothioneine transporter. *Proc Natl Acad Sci U S A* [Internet]. 2005 [cited 2019 Nov 24];102(14):5256–61. Available from: www.pnas.org/cgi/doi/10.1073/pnas.0408624102
83. Cheah IK, Halliwell B. Ergothioneine; antioxidant potential, physiological function and role in disease [Internet]. Vol. 1822, *Biochimica et Biophysica Acta - Molecular Basis of Disease*. 2012 [cited 2019 Sep 24]. p. 784–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22001064>
84. Ogden TH, Rosenberg MS. Multiple sequence alignment accuracy and phylogenetic inference. Page R, editor. *Syst Biol* [Internet]. 2006 Apr 1 [cited 2019 Dec 17];55(2):314–28. Available from: <https://academic.oup.com/sysbio/article/55/2/314/1623777>
85. Wu W, Baker ME, Eraly SA, Bush KT, Nigam SK. Analysis of a large cluster of SLC22 transporter genes, including novel USTs, reveals species-specific amplification of subsets of family members. *Physiol Genomics.* 2009 Jul;38(2):116–24.
86. Yokoyama H, Anzai N, Ljubojevic M, Ohtsu N, Sakata T, Miyazaki H, et al. Functional and immunochemical characterization of a novel organic anion transporter Oat8 (Slc22a9)

- in rat renal collecting duct. *Cell Physiol Biochem*. 2008;21(4):269–78.
87. Quistgaard EM, Löw C, Guettou F, Nordlund P. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way [Internet]. Vol. 17, *Nature Reviews Molecular Cell Biology*. 2016 [cited 2019 Oct 4]. p. 123–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26758938>
88. Inana G, Piatigorsky J, Norman B, Slingsby C, Blundell T. Gene and protein structure of a β -crystallin polypeptide in murine lens: Relationship of exons and structural motifs. *Nature*. 1983;302(5906):310–5.

Table 1: Updated SLC22 Family Subgroups. The SLC22 family was previously separated into 6 phylogenetic subclades. We propose a reclassification into 8 subgroups based on functional data and supported by the methods described.

FORMER GROUPINGS		UPDATED GROUPINGS	
SUBCLADE	MEMBERS	SUBGROUP	MEMBERS
OAT	A6, A7, A8, A9, A10, A11, A12, A19, A20, A22, A24, A25, A26, A27, A28, A29, A30	OATS1	A6, A8, A20
		OATS2	A7
		OATS3	A11, A12, A22
		OATS4	A9, A10, A24, A25; A19, A26, A27, A28, A29, A30
OAT-like	A13, A14	OAT-like	A13, A14
OAT-related	A17, A18, A23, A31	OAT-related	A17, A18*, A23, A31
OCTN-related	A15, A16	OCTN/OCTN-related	A4, A5, A15, A16
OCTN	A4, A5		
OCT	A1, A2, A3	OCT	A1, A2, A3

Table 2: Number of SLC22 transporter associations with unique drugs from DrugBank and metabolites. SLC22 transporter substrate specificity (mono-, oligo-, or multi) was predicted from the number of drugs each was associated with. Metabolite data were then used to support the predicted assignment. In the absence of drug data, metabolites were used to determine specificity.

SLC22 Transporter	# of Unique Drugs	# of Metabolites	Sum	Specificity	Metabolic Pathways
A1	70	15	85	multi	Monoamines, carnitines, prostaglandins
A2	84	24	108	multi	Monoamines, carnitines, prostaglandins, creatinine
A3	40	12	52	oligo	Monoamines, carnitines, creatinine
A4	33	25	58	oligo	Carnitines, ergothioneine
A5	55	20	75	oligo	Carnitines
A6	99	52	151	multi	Uric acid, prostaglandins, gut microbiome derived products, TCA intermediates
A7	35	16	51	oligo	Cyclic nucleotides, prostaglandins, carnitine, creatinine, TCA intermediates
A8	126	88	214	multi	Uric acid, prostaglandins, creatinine, gut microbiome derived products, TCA intermediates, bile acids
A9	0	9	9	oligo	Conjugated sex steroids, short chain fatty acids
A10	3	2	5	relatively mono	Conjugated sex steroids
A11	42	9	51	oligo	Uric acid, prostaglandins, conjugated sex steroids
A12	4	7	15	relatively mono	Uric acid, TCA intermediates
A13	n/a	13	13	relatively mono	Uric acid, TCA intermediates
A14	n/a	n/a	n/a	n/a	Understudied
A15	n/a	7	7	relatively mono	Ergothioneine, complex lipids
A16	2	16	18	oligo	Carnitines, ergothioneine
A17	n/a	2	2	relatively mono	Lipocalin
A18	n/a	2	2	relatively mono	Understudied
A20	n/a	13	13	oligo	Odorants, short chain fatty acids
A22	n/a	12	12	relatively mono	Prostaglandins, conjugated sex steroids
A23	n/a	12	12	oligo	Fatty acids
A24	n/a	10	10	oligo	Conjugated sex steroids, bile acids
A25	n/a	1	1	relatively mono	Conjugated sex steroids
A31	n/a	n/a	n/a	n/a	Understudied

Table 3: Combined functional data for OATS4. These data were manually curated and collected from genome-wide association, in vitro, and in vivo studies. Only statistically significant results from each study are included. Column A is the SLC22 transporter, column B is the metabolite, column C is the source of the data (rsid for GWAS, cell line for in vitro, and the physiological measurement for in vivo), column D is the quantitative metric (p value for GWAS, Km, Ki, IC50, or inhibition percentage compared to control for in vitro, and p value for in vivo), and column E is the citation.

GENE	METABOLITE	Source (GWAS, in vitro, in vivo)	Metrics	Citation
SLC22A9	BUTYRATE	in vitro, Xenopus oocytes	trans-stimulates transport p<0.001	Shin et al., 2007
SLC22A9	DEHYDROEPIANDROSTERONE SULFATE	in vitro, Xenopus oocytes	Km: 2.2 uM	Shin et al., 2007
SLC22A9	ESTRONE SULFATE	in vitro, Xenopus oocytes	Km: 8.7 uM	Shin et al., 2007
SLC22A9	ETIOCHOLANOLONE GLUCURONIDE	GWAS, rs113747568	p=5.27E-28	Long et al., 2017
SLC22A9	NICOTINATE	in vitro, Xenopus oocytes	trans-stimulates transport p<0.01	Shin et al., 2007
SLC22A9	PROGESTERONE	GWAS, rs112295236	p=8.00E-12	Ruth et al., 2015
SLC22A9	PROPIONATE	in vitro, Xenopus oocytes	trans-stimulates transport p<0.01	Shin et al., 2007
SLC22A9	TYRAMINE O-SULFATE	GWAS, rs397740636	P=2.06E-06	Long et al., 2017
SLC22A9	VALERATE	in vitro, Xenopus oocytes	trans-stimulates transport p<0.001	Shin et al., 2007
SLC22A10	EPIANDROSTERONE SULFATE	GWAS, rs1939769	p=2.06E-07	Shin et al., 2014
SLC22A10	ETIOCHOLANOLONE GLUCURONIDE	GWAS, rs112753913	p=1.88E-27	Long et al., 2017
SLC22A24	ANDROSTANEDIOL GLUCURONIDE	in vitro, HEK293 Flp-In	IC50: 21±11 uM	Yee et al., 2019
SLC22A24	CHENODEOXYCHOLATE	in vitro, HEK293 Flp-In	IC50: 2.6±1.0 uM	Yee et al., 2019
SLC22A24	ESTRADIOL GLUCURONIDE	in vitro, HEK293 Flp-In	3-5 fold over vector control	Yee et al., 2019
SLC22A24	ESTRONE SULFATE	in vitro, HEK293 Flp-In	5-10 fold over vector control	Yee et al., 2019
SLC22A24	ETIOCHOLANOLONE GLUCURONIDE	in vitro, HEK293 Flp-In	IC50: 29±4.7 uM	Yee et al., 2019
SLC22A24	ETIOCHOLANOLONE GLUCURONIDE	GWAS, rs113532193	p=5.90E-37	Long et al., 2017
SLC22A24	PREGNANEDIOL-3-GLUCURONIDE	in vitro, HEK293 Flp-In	IC50: >200uM	Yee et al., 2019
SLC22A24	PREGNANEDIOL-3-GLUCURONIDE	GWAS, rs202187460	p=5.91E-07	Long et al., 2017
SLC22A24	PREGNENOLONE SULFATE	in vitro, HEK293 Flp-In	IC50: 1.4±0.1 uM	Yee et al., 2019
SLC22A24	PROGESTERONE	in vitro, HEK293 Flp-In	IC50: 7.4±3.0 uM	Yee et al., 2019
SLC22A24	TAUROCHOLIC ACID	in vitro, HEK293 Flp-In	10-20 fold over vector control	Yee et al., 2019
SLC22A24	URSODEOXYCHOLATE	in vitro, HEK293 Flp-In	IC50: 7.6±1.2 uM	Yee et al., 2019
SLC22A25	ETIOCHOLANOLONE GLUCURONIDE	GWAS, rs113950742	p=4.12E-27	Long et al., 2017

Table 4: Genomic localization and tissue expression of SLC22 family. The following table describes the genomic localization and tissue expression patterns of all SLC22 members excluding the mouse-specific Slc22a19, Slc22a26, Slc22a27, Slc22a28, Slc22a29, and Slc22a30. Slc22a22 and Slc22a21 expression patterns described are from mouse (59,64). **(m)** denotes expression patterns observed exclusively in mice. Tissue expression data in humans was collected from various sources and databases (4,24,64). Expression is assumed from mRNA expression analysis, unless confirmed experimentally. Bolded transporters within subgroups are found in tandem within the human genome.

Subgroup	SLC22 member	Genomic Localization		Tissue Localization										
		human chromosome	mouse chromosome	Liver	Kidney	Brain	Gut	Heart	Lung	Retina	Testis	Immune Cell	Bone Marrow	Placenta
OATS1	SLC22A6	11	19		✓	✓					✓			
	SLC22A8	11	19		✓	✓					✓			
	SLC22A20	11	19		✓	✓						✓ (m)		✓
OATS2	SLC22A7	6	17	✓	✓							✓		
OATS3	SLC22A11	11			✓									✓
	SLC22A12	11	19		✓									
	slc22a22 (m)		15		✓									
OATS4	SLC22A9	11		✓		✓								✓
	SLC22A10	11		✓		✓								
	SLC22A24	11		✓	✓	✓	✓					✓		
	SLC22A25	11		✓										
OAT-Like	SLC22A13	3	9		✓	✓	✓	✓						
	SLC22A14	3	9		✓	✓						✓		
OAT-Related	SLC22A17	14	14	✓	✓	✓	✓	✓				✓		
	SLC22A18	11	7	✓	✓	✓	✓							
	SLC22A23	6	13	✓	✓	✓	✓			✓				
	SLC22A31	16								✓		✓		
OCTN/ OCTN-related	SLC22A4	5	11	✓	✓	✓	✓		✓					✓
	SLC22A5	5	11	✓	✓	✓	✓		✓			✓		✓
	SLC22A16	6	10	✓	✓	✓					✓	✓		
	SLC22A15	1	3		✓	✓	✓	✓	✓	✓	✓	✓	✓	
	slc22a21 (m)		11		✓	✓	✓					✓		✓
OCT	SLC22A1	6	17	✓	✓	✓	✓	✓	✓	✓		✓		✓
	SLC22A2	6	17	✓	✓	✓	✓	✓	✓					
	SLC22A3	6	17	✓	✓	✓	✓	✓	✓			✓		✓

Table 5: ICM finds significant similarities with SLC22 members. The following table shows significant amino acid similarities found between full-length and the ECD sequences of SLC22 members and other known proteins from human (*homo sapiens*, h), cow (*bos taurus*, b), chicken (*gallus gallus*, g), mouse (*mus musculus*, m) and rat (*rattus norvegicus*, r). No significant similarities were found for SLC22 ICDs. pP value is the log of the p-value and is described in the methods.

Subclade	SLC22 Family Member	Non-SLC22 Protein	Identity Shared (%)	pP value
Oct	hSLC22A1 ECD	hSCO-spondin	28.97	5.47
		bSCO-spondin	30.84	5.35
		mSCO-spondin	24.3	6.24
		rSCO-spondin	24.3	6.16
	hSLC22A2 ECD	bSCO-spondin	30.84	5.29
		mSCO-spondin	25.23	5.92
		rSCO-spondin	24.3	5.61
	hSLC22A3 ECD	hSCO-spondin	27.43	5.33
		bSCO-spondin	22.12	5.49
		mSCO-spondin	27.43	5.89
rSCO-spondin		25.66	5.81	
Oat-related	hSLC22A31 ECD	hRBM42	30.95	5.95
		bRBM42	32.14	6.05
		mRBM42	30.95	5.95
		rRBM42	30.95	5.95
		hBAHD1	36.9	5.39
Octn	mSlc22a16 ECD	gCRBB3	26	5.4
	hSLC22A16	hTAS2R41	20	5.3
	hSLC22A5	GPR160	21	6.1

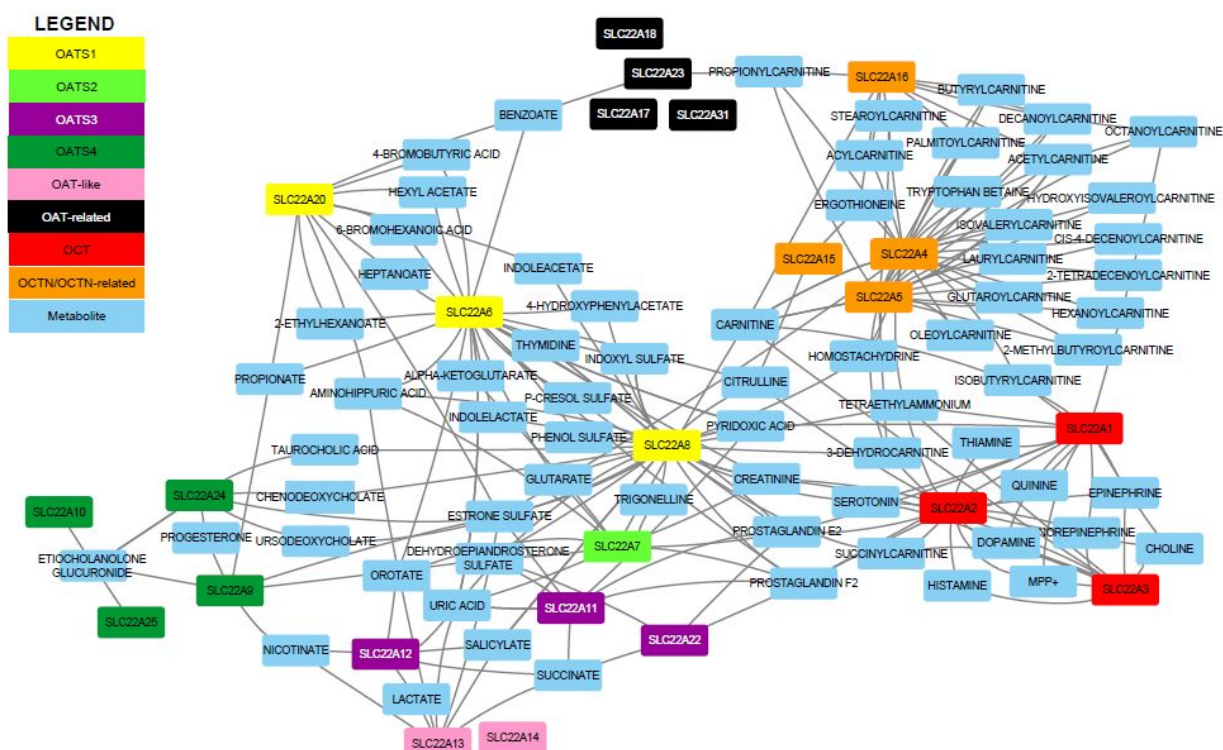


Figure 1: Pruned SLC22 network. All SLC22 transporters with functional data were initially included. Metabolites associated with only one transporter were removed for improved visualization. SLC22 transporters and metabolites are colored nodes. Each edge represents a significant transporter-metabolite association. Multiple edges connecting one metabolite to a specific transporter were bundled (e.g., *in vitro* and GWAS support).

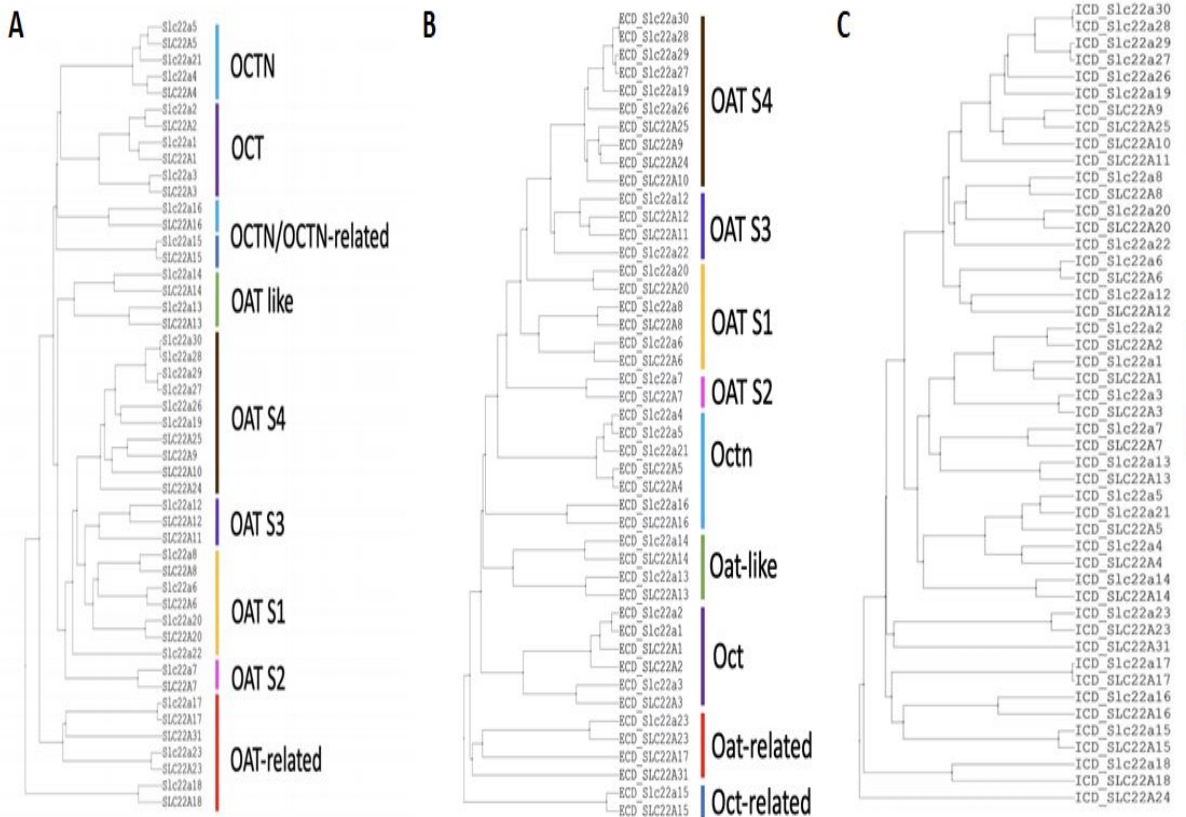


Figure 2: Multiple sequence alignment using ICM-Pro v3.8-7c tree of SLC22 members implies function. All known mouse and human SLC22 sequences, excluding Slc22a18, were aligned using ICM-Pro v3.8-7c sequence similarity-based alignment. A) Full sequence. B) Extracellular loop (not including Slc22a18, due to its lack of a characteristic large extracellular loop between TMD1 and TMD2). C) Intracellular loop.

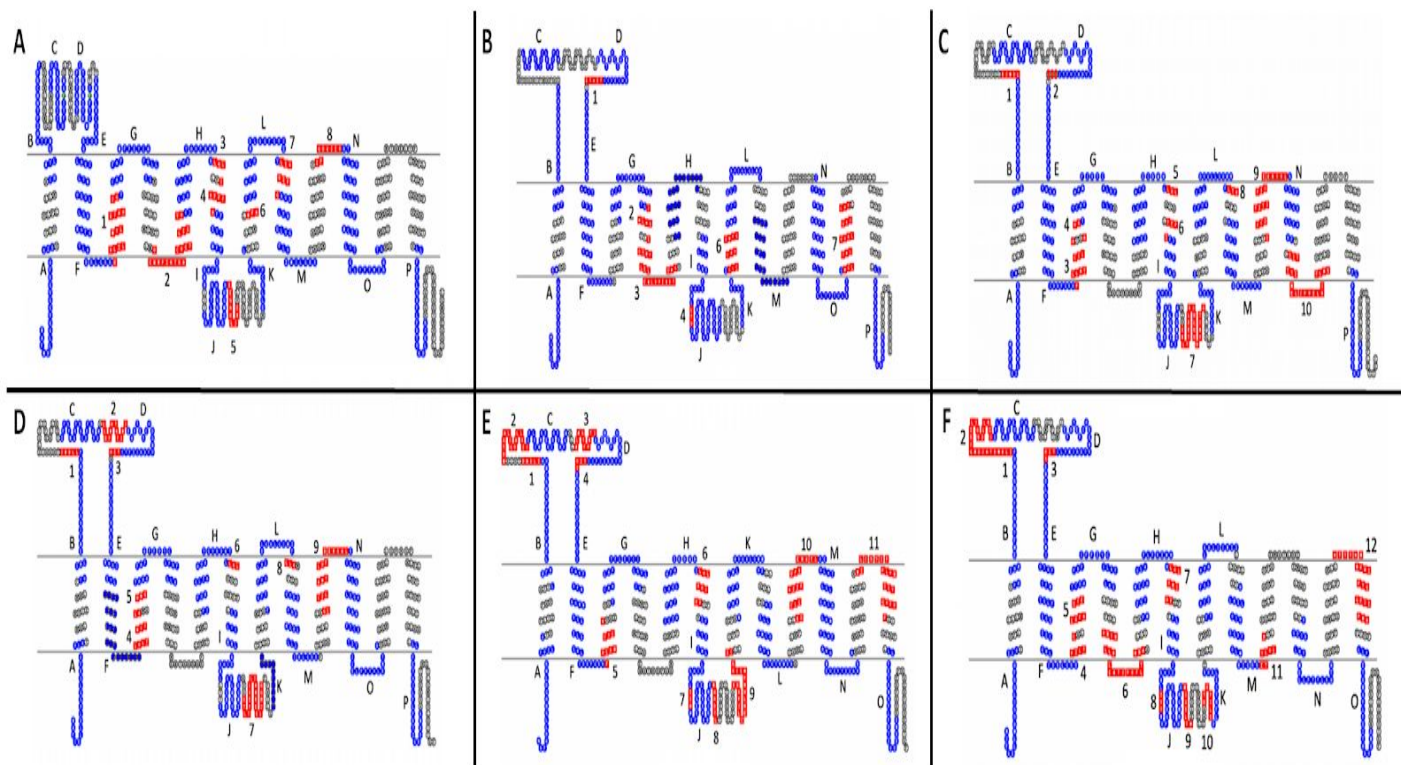
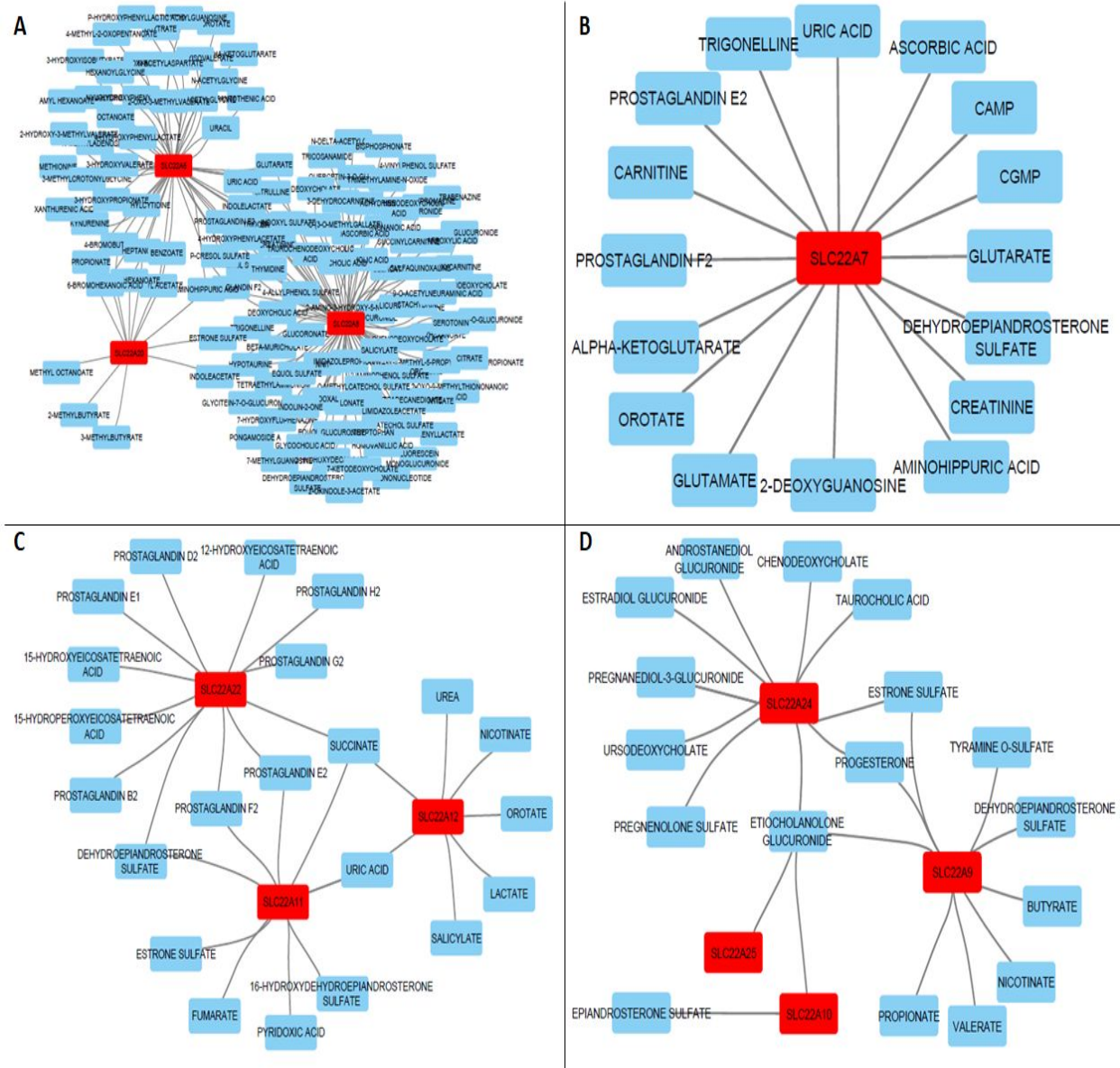


Figure 3: Evolutionarily conserved motifs for each subgroup within the OAT major subgroup mapped onto 2D topology of prototypical members. A) OATS1 mapped onto SLC22A6. B) OATS2 mapped onto hSLC22A7. C) OATS3 mapped onto hSLC22a12. D) OATS3 mapped onto mSlc22a22. E) OATS4 mapped onto hSLC22A9. F) OATS4 mapped onto mSlc22a7. In each panel, red sequences are subgroup specific motifs, blue sequences are OAT-major subgroup motifs, and green diamonds represent non-synonymous SNPs that affect serum metabolite levels. Conserved OAT-major subgroup motifs are assigned letters and specific, conserved OAT subgroup motifs are numbered. Data, including motif sequence identities, exact locations, and p-values can be found in Supplementary Tables 2-7.

Supplementary Figures



SFigure 1: OATS1-4 Subgroups. The previously defined OAT major subclade was reconstructed into smaller groups based on additional support. A) OATS1 (SLC22A6, SLC22A8, SLC22A20). B) OATS2 (SLC22A7). C) OATS3 (SLC22A11, SLC22A12, SLC22A22). D) OATS4 (SLC22A9, SLC22A10, SLC22A24, SLC22A25). In each panel, the SLC22 family members are red nodes, the metabolites are blue nodes, and the edges connecting them are evidence of interaction between the protein and the metabolite.

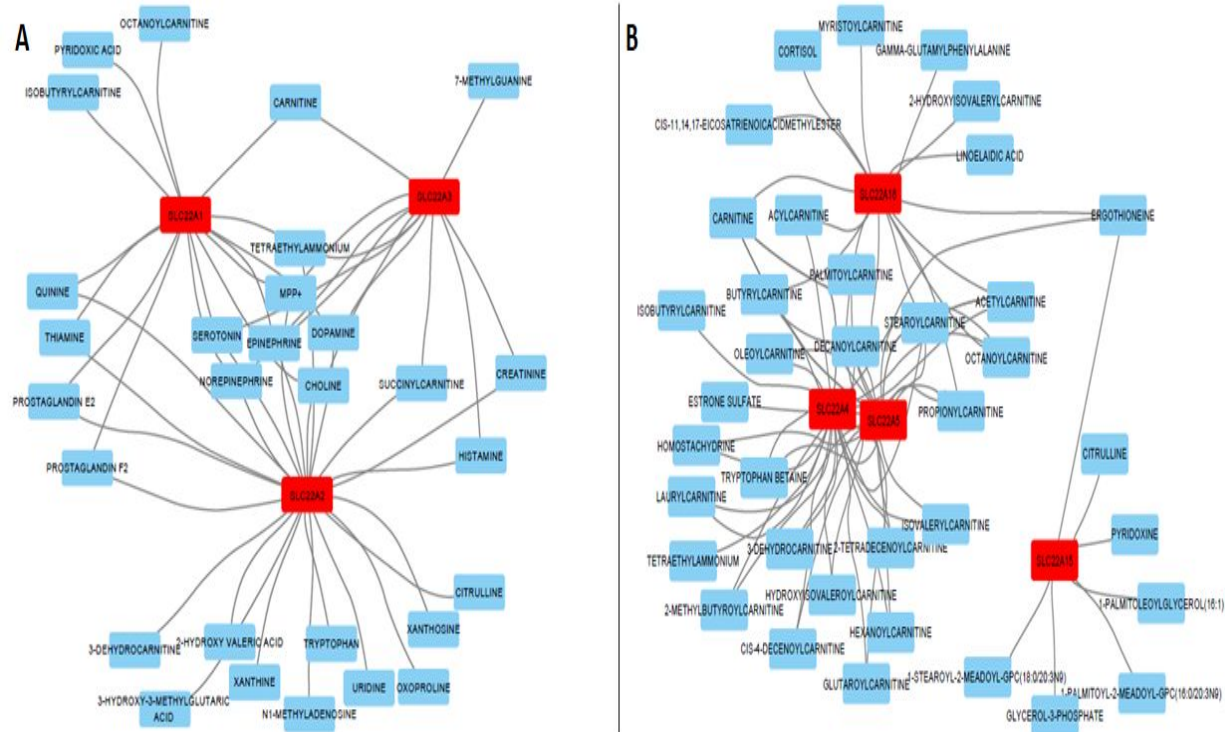
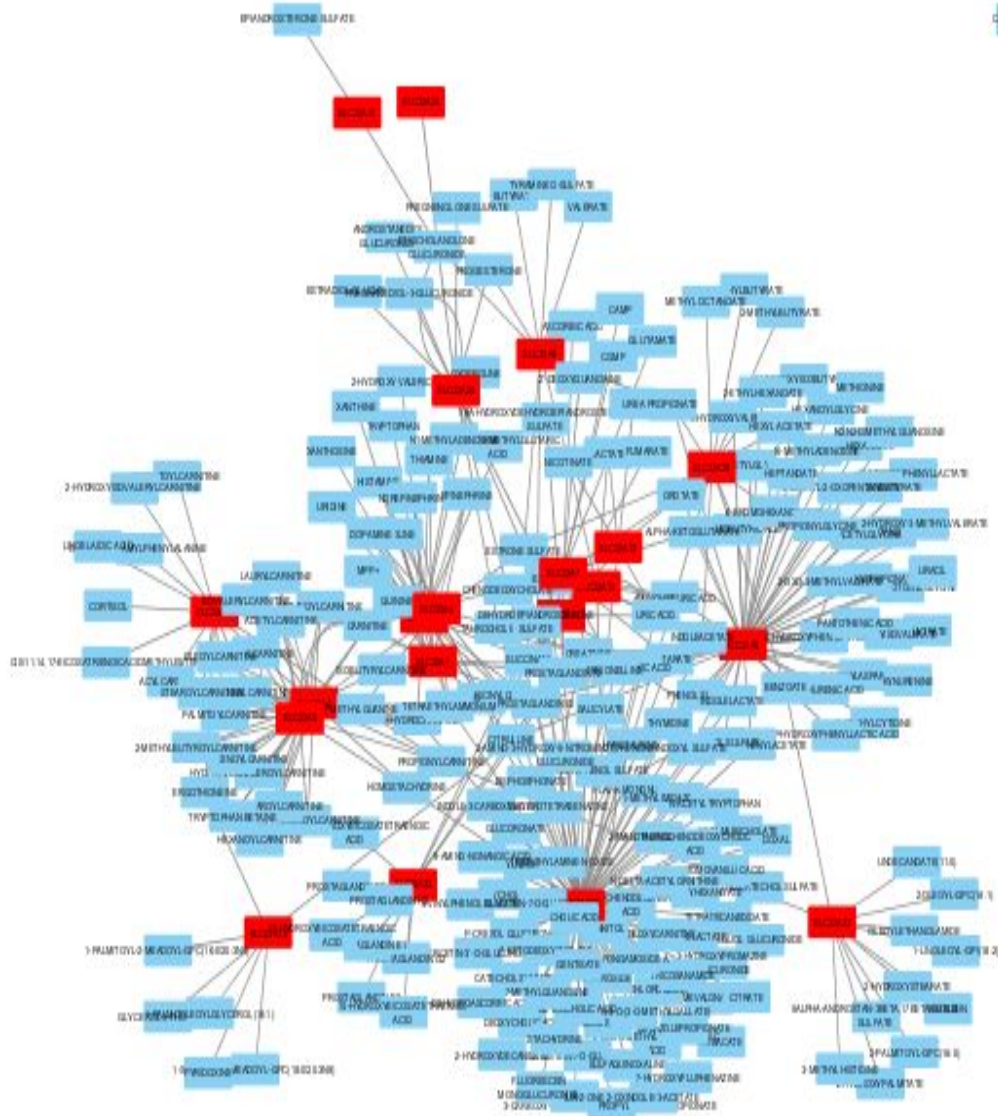
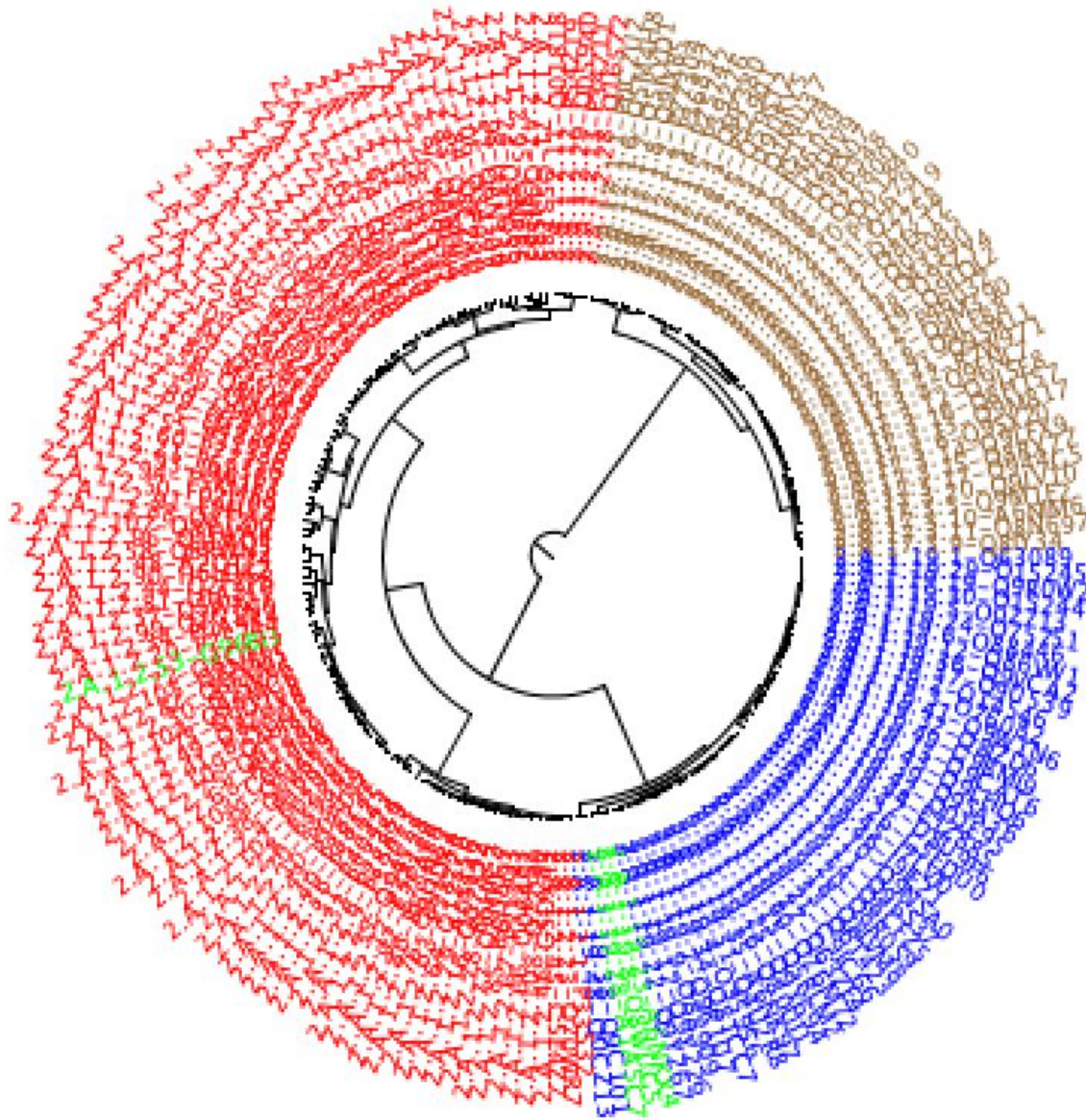


Figure 2: OCT and OCTN/OCTN-related subgroups. The previously defined OCT major subclade was reconstructed into new groups based on additional support. A) OCT (SLC22A1, SLC22A2, SLC22A3). B) OCTN/OCTN-Related (SLC22A4, SLC22A5, SLC22A15, SLC22A16). In each panel, the SLC22 family members are red nodes, the metabolites are blue nodes, and edges connecting them are evidence of interaction between the protein and the metabolite.



SFigure 3: Total SLC22 Network. Every found association is represented by an edge connecting a metabolite to a transporter. SLC22 transporters are red nodes, and metabolites are light blue nodes.



SFigure 4: SSearch36 homology results show Slc22a18 within the DHA family.

The DHA family is shown in red, with SLC22A18 highlighted in green. The SLC22 family is shown in blue, with the other Oat-related subclade members (SLC22A17, SLC22A23, and SLC22A31) highlighted in green. The POT transporter family outgroup is shown in brown.

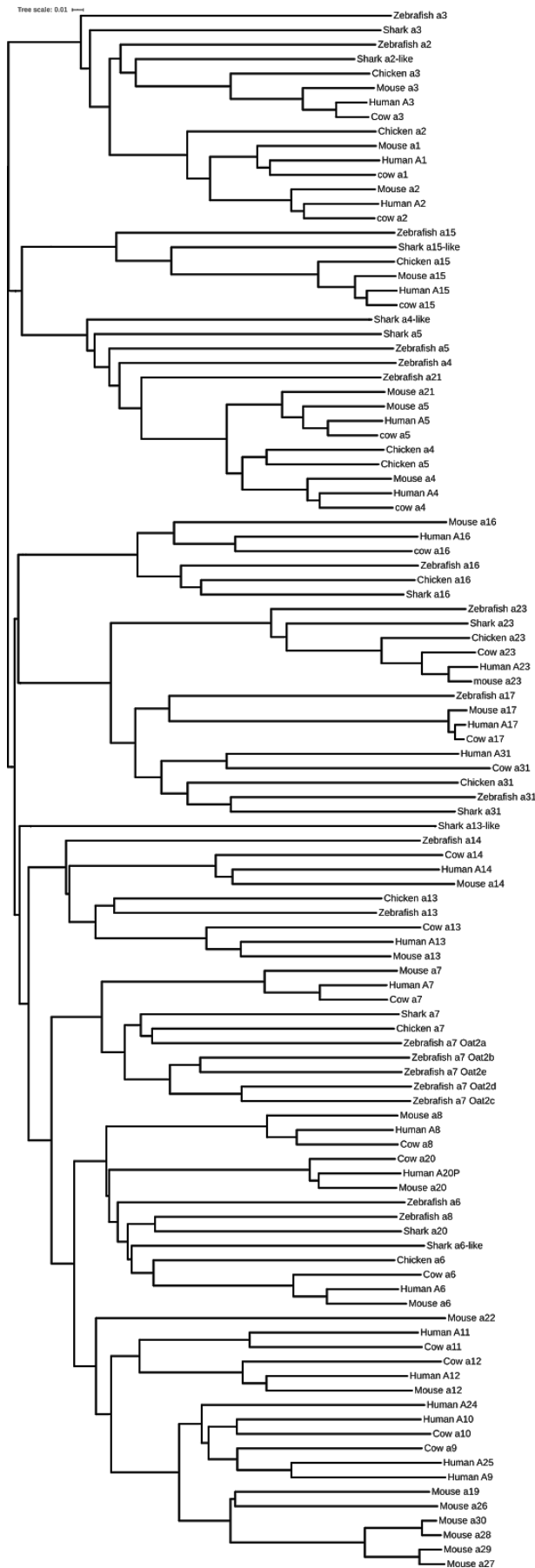
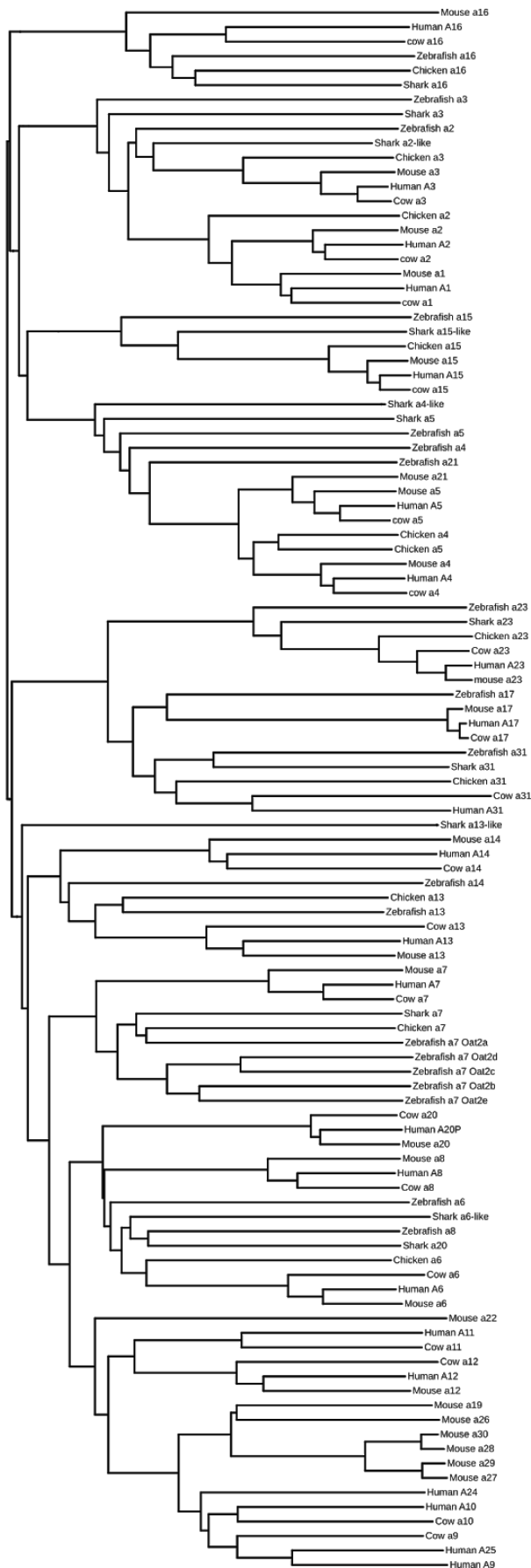


Figure 5: MAFFT multiple sequence alignment. Phylogenetic tree of previously defined OAT subclade. 90 sequences were aligned using MAFFT and viewed on the Interactive Tree of Life (iTOL). Branch lengths indicate the number of amino acid changes per residue between the sequences and the most recent ancestor. This analysis supports the formation of 4 functionally distinct subgroups from the OAT subclade.



SFigure 6: Clustal-Omega multiple sequence alignment. All SLC22 sequences from human, mouse, cow, chicken, shark, and zebrafish were aligned using Clustal-Omega. Tree was visualised using iTOL.

Supplementary Tables

This table is very large and is attached.

STable 1: Combined functional data for individual SLC22 transporters. These data were manually curated and collected from genome-wide association, in vitro, and in vivo studies. Only statistically significant results from each study are included. Column A is the SLC22 transporter, column B is the metabolite, column C is the source of the data (rsid for GWAS, cell line for in vitro, and the physiological measurement for in vivo), column D is the quantitative metric (p value for GWAS, Km, Ki, IC50, or inhibition percentage compared to control for in vitro, and p value for in vivo), and column E is the citation.

Table 2: Evolutionarily conserved Oat subclade and Oat subgroup 1 motifs. Table describing the Oat Subclade-specific (blue) and OATS1-specific (Slc22a6, Slc22a8, and Slc22a20; red) conserved motifs that have been identified within the hOAT1 (hSLC22A6) sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of hOAT1 associated with the particular motif letter.

	Motif	Region	% Present	MEME#	p-value	Sequence
Oat subclade	A	1-20	0.97	12	2.27E-20	MAFNDLLQQVGGVGRFQQIQ
	B	33-52	0.99	2	1.23E-22	SHNTLQNFTA AIPTHHCRPP
	C	67-86	0.99	5	8.88E-19	LPRDRQGQPESCLRFTSPQW
	D	99-118	0.99	1	7.82E-21	TGATEPCTDGWIYDNSTFPS
	E	123-142	0.99	3	3.01E-20	EWDLVCSHRALRQLAQSLYM
	F	143-162	1	6	5.34E-18	VGVLLGAMVFGYLADRLGRR
	G	177-196	1	13	1.09E-20	TCAAFAPNFPIYCAFRLLSG
	H	230-249	0.99	10	5.99E-21	YSLGQFLLAGVAYAVPHWRH
	I	261-275	0.95	11	2.59E-16	FFIYSWFFIESARWH
	J	279-298	0.95	7	6.4E-17	GRLDLTLRALQRVARINGKR
	K	328-338	0.86	15	1.69E-09	ELLRCPTLRHL
	L	349-368	0.97	9	1.28E-21	TSFAYYGLVMDLQGFVSIY
	M	376-395	0.97	16	5.25E-18	AVDLPKLVGFLVINSLGRR
	N	423-437	0.86	14	2.1E-16	RTSLAVLGKGCCLAAS
	O	438-457	0.93	8	2.02E-23	FNCIFLYTGELYPTMIRQTG
P	503-522	0.97	4	8.46E-20	LLPETLGQPLPDTVQDLESR	
Oat subgroup 1	Motif	Region	% Present	MEME #	p-value	Sequence
	1	163-176	1	12	2.92E-22	KVLILNYLQTAVSG
	2	211-229	1	9	4.90E-25	VEWMPIHTRACVGTLLIGYV
	3	250-252	1	6	1.39E-23	LQL
	4	255-260	1	5	1.14E-23	SAPFFA
	5	299-307	1	7	1.65E-23	EEGAKLSME
	6	346-348	1	3	2.32E-24	WFA
	7	369-375	1	10	4.42E-22	LIQVIFGAVDLPKLVG
8	415-422	1	13	1.34E-21	IPQDQSIV	

STable 3: Evolutionarily conserved Oat subclade and Oat subgroup 2 motifs. Table describing the Oat Subclade-specific (blue) and OATS2-specific (Slc22a7; red) conserved motifs that have been identified within the hOAT2 (hSLC22A7) sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of hOAT2 associated with the particular motif letter.

	Motif	Region	% Present	MEME #	p-value	Sequence
Oat Subclade	A	43485	0.97	12	5.3E-18	MGFEELLEQVGGFGPFQLRN
	B	33-52	0.99	2	4.6E-17	LHFLLPIFLAAVPAHRCALP
	C	69-88	0.99	5	7.5E-16	LPREPDGTLSSCLRFAYPQA
	D	107-126	0.99	1	5.3E-20	EPATVPCSQGWEDHSEFSS
	E	133-152	0.99	3	1.1E-16	QWDLVCEQKGLNRAASTFFF
	F	153-172	1	6	8.6E-21	AGVLVGAVAFGYLSDRFGR
	G	187-206	1	13	1.2E-11	LASAASVSYVMFAITRTLGT
	H	240-259	0.99	10	4.4E-17	WTGGVMLLALVGYLIRDWRW
	I	271-285	0.95	11	1.3E-14	GILSLWWVPESARWL
	J	289-318	0.95	7	1.5E-14	GHVKEAHRYLLHCARLNGRP
	K	336-346	0.86	15	1.4E-12	DLFRTPRLRHI
	L	357-376	0.97	9	1.1E-17	VNFSYYGLSLDVSGGLNVY
	M	384-403	0.97	16	1.3E-15	AVELPSKLLVYLSVRYAGR
	N	431-445	0.86	14	9.3E-12	STVLAVMGKAFSEAA
	O	446-465	0.93	8	6.6E-21	FTTAYLFTSELYPTVLRQTG
	P	511-530	0.97	4	8.5E-20	LLPETRQAQLPETIQDVERK
Oat subgroup 2	Motif	Region	% Present	MEME #	p-value	Sequence
	1	127-131	1	4	4.61E-26	TIATE
	2	207-211	1	14	1.05E-23	SALAG
	3	214-233	1	9	3.18E-25	IIVMPLELEWLDVEHRTVAG
	4	286-288	1	2	2.51E-26	LTQ
	5	347-356	1	5	2.29E-26	SLCCVVVWFG
	6	377-383	1	10	4.05E-23	QTQLLFGAVELPSK
7	466-481	0.93	15	2.44E-22	MGLTALVGRLLGGSLAP	

STable 4: Evolutionarily conserved Oat subclade and Oat subgroup 3 motifs. Table describing the Oat Subclade-specific (blue) and OATS3-specific (Slc22a11 and Slc22a12; red) conserved motifs that have been identified within the hURAT1 (hSLC22A12) sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of hOAT3 associated with the particular motif letter.

	Motif	Region	% Present	MEME #	p-value	Sequence
Oat Subclade	A	1-20	0.97	12	6.05E-20	MAFSELLDLVGGLGRFQVLQ
	B	33-52	0.99	2	7.53E-19	TQSMLNFSAAVPSHRCWAP
	C	77-96	0.99	5	7.02E-23	IPPGPNQRPHQCRRFRQPQW
	D	110-129	0.99	1	4.49E-25	EADTEPCVDGWVYDRSIFTS
	E	134-153	0.99	3	3.36E-19	KWNLVCDSHALKPMAQSIYL
	F	154-173	1	6	2.27E-18	AGILVGAAACGPASDRFGRR
	G	188-207	1	13	1.50E-17	TAAAFAPAFPVYCLFRLLA
	H	241-260	0.99	10	1.64E-13	FSFGHGLTAAVAYGVRDWTL
	I	272-286	0.95	11	1.06E-17	CFLYSWWLAESARWL
	J	290-309	0.95	7	1.41E-16	GRLDWGLQELWRVAAINGKG
	K	339-349	0.86	15	1.34E-07	TLLRMPGLRFR
	L	360-379	0.97	9	2.17E-16	FGFTFFGLALDLQALGSNIF
	M	387-406	0.97	16	1.92E-17	VVDIPAKMGALLLLSHLGRR
	N	434-448	0.86	14	2.87E-11	RSALAVLGLGGVGAA
	O	449-468	0.93	8	4.70E-19	FTCITIYSSELFPTVLRMTA
P	514-533	0.97	4	2.33E-20	LLPETQSLPLPDTIQDVQNQ	
Oat subgroup 3	Motif	Region	% Present	MEME #	p-value	Sequence
	1	53-57	1	3	7.75E-25	LLDNS
	2	98-109	0.85	15	5.94E-19	LLDPNATATSW
	3	130-132	1	1	4.79E-27	TIV
	4	174-182	1	8	3.61E-22	LVLTWSYLQ
	5	185-187	1	12	2.53E-21	VMG
	6	261-263	1	13	7.90E-20	LQL
	7	313-332	0.95	14	3.34E-20	DTLTPEVLLSAMREELSMGQ
	8	380-382	0.95	10	1.65E-21	LLQ
9	417-433	0.9	16	2.86E-23	GLCILANTLVPHEMGAL	

Table 5: Evolutionarily conserved Oat subclade and Oat subgroup 3 motifs. Table describing the Oat Subclade-specific (blue) and OATS3-specific (Slc22a11, Slc22a12, and Slc22a22; red) conserved motifs that have been identified within the mOat-pg (mSlc22a22) sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of mOat-pg associated with the particular motif letter.

Oat Subclade	Motif	Region	% Present	MEME #	p-value	Sequence
	A	1-20	0.97	12	2.41E-15	MDFDEILHHVGDSDGRFQICM
	B	33-52	0.99	2	7.07E-18	PHDVLENFTAAIPAHHCSIN
	C	76-95	0.99	5	9.03E-21	IPMGPNQKPEQCRFRFYTQW
	D	109-128	0.99	1	1.34E-20	ELETEPCLDGWTYDHSVFTS
	E	133-152	0.99	3	1.64E-14	EWDLVCFDFQSFKYAQAATSL
	F	153-172	1	6	1.13E-13	AGHLVSCPLSGHISDRFGRK
	G	187-206	1	13	3.93E-16	TYCAFAPNFSVYCVLRFLLS
	H	240-259	0.99	10	7.52E-15	NSIGQGVLGGLAYVISDWHL
	I	271-285	0.95	11	3.33E-12	FFVLFCWVPESVRWL
	J	289-308	0.95	7	1.12E-17	GKTDQAWKELQRIASINGKK
	K	338-348	0.86	15	1.17E-07	DIFINPLIRKI
	L	359-378	0.97	9	4.39E-12	ELFSFVGLLLDVQLLGKNMF
	M	386-405	0.97	16	1.20E-12	AIDVPSKSLTYFTIRNVSR
	N	433-447	0.86	14	4.15E-10	RTIIFILGKGCFAAF
O	513-532	0.97	4	7.45E-16	FLPETFNLPHPTDIKDMEK	
Oat subgroup 3	Motif	Region	% Present	MEME #	E-value	Sequence
	1	53-57	1	3	2.67E-16	LDNSR
	2	129-131	1	2	1.62E-24	TIV
	3	173-181	1	8	1.57E-16	PLLMYCSLA
	4	184-186	1	12	4.47E-18	AVG
	5	260-262	1	13	2.32E-17	LQL
	6	267-270	1	4	1.73E-18	PFFI
	7	312-331	1	14	8.23E-14	QNLTTEDLRSKLKKDVNSTG
	8	379-381	1	11	3.17E-15	LTQ
	9	416-432	1	16	2.34E-12	GSCITITIFISEEMYVL
10	450-469	1	10	5.19E-13	ISTTYINELSPVELRSTLNG	

Table 6: Evolutionarily conserved Oat subclade and Oat subgroup 4 motifs. Table describing the Oat Subclade-specific (blue) and OATS4-specific (Slc22a9, Slc22a10, Slc22a24, Slc22a25, Slc22a19, Slc22a26, Slc22a27, Slc22a28, Slc22a29, and Slc22a30; red) conserved motifs that have been identified within the hOAT7 (hSLC22A9) sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of hOAT7 associated with the particular motif letter

Motif	Region	% Present	MEME #	p-value	Sequence
A	1-20	0.97	12	4.45E-17	MAFQDLLGHAGDLWRFQILQ
B	33-52	0.99	2	1.24E-21	LHFMLENFTAFIPGHRCWVH
C	77-96	0.99	5	1.62E-24	IPLDSNMRPEKCRRFVHPQW
D	110-129	0.99	1	5.65E-21	DADMEPCVDGWVYDRISFSS
E	134-153	0.99	3	5.02E-20	EWDLCVDSQSLTSVAKFVFM
F	154-173	1	6	1.32E-19	AGMMVGGILGGHLSDRFGRR
G	188-207	1	13	3.15E-19	TCAALAPTFLIYCSLRFSLG
H	241-260	0.99	10	3.35E-14	SGIAFMTLAGLAFAIRDWHI
I	272-286	0.95	11	1.16E-13	IFLTSSWLLESARWL
J	290-309	0.95	7	9.21E-19	NKPEEGLKELRKAHRSGMK
K	361-380	0.97	9	9.18E-17	NFMAYFGLNLHVQHLGNVVF
L	388-407	0.97	16	2.04E-13	AVILLANCVAPWALKYMNRR
M	435-449	0.86	14	1.11E-09	REVLATLGLGASALA
N	450-469	0.93	8	1.77E-13	NTLAFAHGNEVIPTIIRARA
O	515-534	0.97	4	2.09E-19	LLPETRKNPLFDTIQDEKNE
Motif	Region	% Present	MEME #	p-value	Sequence
1	53-57	1	3	6.06E-27	ILDND
2	62-76	0.96	14	1.06E-16	NDTGALSQDALLRIS
3	100-109	1	13	8.06E-17	HLNGTFPNTS
4	130-133	1	1	7.76E-25	TIVT
5	174-182	1	8	1.39E-23	FVLRWCYLQ
6	261-268	1	10	9.11E-23	LQLVVSVP
7	287-289	0.92	5	2.20E-25	IIN
8	310-315	0.92	15	1.28E-16	NARDTL
9	333-352	0.92	12	2.41E-18	KKKPSLCEMLHMPNICKRIS
10	418-434	0.88	16	2.51E-21	AICLLAIIFVPQEMQTL
11	488-507	0.92	11	1.70E-23	MILSVYSPPLPWIIYGVFPF

STable 7: Evolutionarily conserved Oat subclade and Oat subgroup 4 motifs. Table describing the Oat Subclade-specific (blue) and OATS4-specific (Slc22a19, Slc22a26, Slc22a27, Slc22a28, Slc22a29, and Slc22a30; red) conserved motifs that have been identified within the mouse Slc22a27 sequence. This table lists the motif letter, prevalence, MEME#, p-value and amino acid sequence of mSlc22a27 associated with the particular motif letter.

	Motif	Region	% Present	MEME #	p-value	Sequence
Oat Subclade	A	1-20	0.97	12	6.05E-20	MSFQELLNQVGS LGRFQILQ
	B	33-52	0.99	2	1.44E-21	PHIAMENFTA AIPNHRCWVP
	C	77-96	0.99	5	1.59E-23	IPLDSNLRPEK CRRFAQPQW
	D	110-129	0.99	1	8.62E-25	EPDTEPCVDG WVYDRSNFLS
	E	134-153	0.99	3	3.89E-20	EWDLVCESQAL NSVTKFSFM
	F	154-173	1	6	3.19E-18	IGLFIGGIICGH LSDRLGRK
	G	188-207	1	13	1.96E-20	TCVAFAPSFFI YCSLRFLAG
	H	241-260	0.99	10	5.73E-15	PNIGYMISAGLA FLRIWHH
	I	272-286	0.95	11	1.06E-13	FLILTRWLSESA RWL
	J	290-309	0.95	7	8.47E-24	NKPQKGLKELR KVAHMNGMK
	K	338-348	0.86	15	1.76E-11	DLFHTSILRKR
	L	359-378	0.97	9	1.08E-13	FTVSIFGLAVHL QHLSSNII
	M	386-405	0.97	16	2.05E-12	ALAILVSVIGPF VLNHIGRR
	N	448-467	0.93	8	1.10E-12	VGVSRLHTNELL PTLRATA
	O	513-532	0.97	4	2.90E-19	LLPETKNQPLPD STHDVGND
Oat subgroup 4	Motif	Region	% Present	MEME #	p-value	Sequence
	1	53-57	1	3	2.87E-27	ILDND
	2	58-76	1	8	1.88E-24	TASDNRSRILSQ DDLRIS
	3	130-133	1	6	7.81E-26	TIVT
	4	174-178	1	12	1.44E-23	FILTC
	5	181-187	1	7	2.59E-26	LQFAITE
	6	215-234	1	14	2.03E-24	VNSHLLMLEWTSP KFLTMM
	7	261-268	1	9	9.82E-25	LQLTMSVP
	8	287-289	1	5	2.26E-26	IVT
	9	310-316	1	10	1.41E-25	NSGDLTM
	10	331-337	1	11	2.08E-24	KTKPSLR
	11	406-412	1	15	2.00E-24	ITYLVLM
12	489-508	1	13	2.12E-24	ATYYANMPWIFYG GFSIFNA	