Systematic *in silico* discovery of novel solute carrier-like proteins from proteomes

Gergely Gyimesi* and Matthias A. Hediger*

Membrane Transport Discovery Lab, Department of Nephrology and Hypertension and Department for BioMedical Research, Inselspital, University of Bern, Freiburgstrasse 15, CH-3010 Bern, Switzerland.

*Corresponding authors.
E-mail: gergely.gyimesi@dbmr.unibe.ch (GG), matthias.hediger@ibmm.unibe.ch (MH).

Abstract

Solute carrier (SLC) proteins represent the largest superfamily of transmembrane transporters. While many of them play key biological roles, their systematic analysis has been hampered by their functional and structural heterogeneity. Based on available nomenclature systems, we hypothesized that many as yet unidentified SLC transporters exist in the human genome, which await further systematic analysis. Here, we present criteria for defining "SLC-likeness" to curate a set of "SLC-like" protein families from the Transporter Classification Database (TCDB) and Protein families (Pfam) databases. Computational sequence similarity searches surprisingly identified ~120 more proteins in human with potential SLC-like properties compared to previous annotations. Interestingly, several of these have documented transport activity in the scientific literature. To complete the overview of the "SLC-ome", we present an algorithm to classify SLC-like proteins into protein families, investigating their known functions and evolutionary relationships to similar proteins from 6 other clinically relevant experimental organisms, and pinpoint structural orphans. We envision that our work will serve as a stepping stone for future studies of the biological function and the identification of the natural substrates of the many under-explored SLC transporters, as well as for the development of new therapeutic applications, including strategies for personalized medicine and drug delivery.

Introduction

Membrane transporters and channels are the main entry routes for nutrients, ions, xenobiotics and serve as major exit routes for waste products and metabolites. The solute carrier (SLC) protein superfamily accounts

31 for over 50% of all transport-related proteins and about 10% of all membrane proteins encoded by the
32 human genome. With more than 400 annotated members, it is the largest superfamily of membrane
33 transporter proteins [1,2]. The roles of SLC transporters as cellular gatekeepers, determinants of nutrient
34 homeostasis and facilitators of drug metabolism and drug targeting has recently been revisited [3]. Around
35 ~50% of currently annotated SLCs are predicted to be associated with human disease phenotypes and many
36 SLCs are considered to represent promising drug targets or drug delivery systems or to affect drug ADMET
37 (absorption, distribution, metabolism, extrusion, toxicity). It has recently become evident that the SLC
38 superfamily offers an enormous unexplored therapeutic treasure. But while the list of approved drugs that
39 target transporter proteins is increasing, many promising SLCs still remain unexplored, uncharacterized and
40 underrepresented in the literature.

41 The SLC nomenclature system has traditionally been used to classify mammalian secondary active and
42 facilitative transporters, including exchangers and antiporters, into families based on sequence identity per
43 enquiry by the Human Gene Nomenclature Committee (HGNC) starting in the early 1990s [1]. Originally, SLCs
44 assignments have been made for all membrane transport proteins that are not channels, ATP-driven pumps,
45 aquaporins, porins of the outer mitochondrial membrane or ATP-binding cassette (ABC) transporters, while
46 usually having multiple transmembrane spanning segments and exhibiting transmembrane transport of a
47 solute, or showing homology to membrane proteins having such features. Due to this construction, the SLC
48 superfamily is structurally and functionally highly heterogeneous and thus most likely evolutionarily
49 polyphyletic in origin. Because of these properties and the lack of common sequence patterns in the
50 different SLC transporters, it has been difficult to assess how many SLC transporters exist in the human
51 genome or which proteins could be predicted as SLC transporters. Proteins were typically added to the SLC
52 system on a case-by-case basis. However, in view of recent requests to add new members into the SLC
53 nomenclature, we suspected that the SLC system might be incomplete.

54 Despite their heterogeneity, SLC transporters seem to share common properties that probably were
55 evolutionarily selected based on their suitability as facilitative transporters, secondary active transporters or
56 exchangers. A remarkable property is that they generally have a symmetric inverted repeat architecture,

57 which can be observed based on the available structures [4] and can sometimes also be detected at the
58 sequence level [5,6]. Another important aspect is that the currently well-studied SLC transporters seem to
59 follow an alternating access mechanism, meaning that the substrate-binding site is exposed on either one or
60 the other side of the membrane, but not on both sides simultaneously [7,8]. Consequences of these
61 properties are that SLC transporters typically contain many transmembrane helices (TMHs), and functionally
62 exhibit saturable transport activity. In addition, most but not all SLC carriers transport water-soluble small
63 molecules. These properties could be used as criteria to identify additional SLC transporter proteins.

64 In fact, there have been earlier attempts to gather additional SLCs from the human genome [9–11], as well
65 as to classify them using automatic methods [12,13]. In this regard, one study [9] used BLAST searches to
66 find SLC transporters that have local sequence similarities and found that 15 of the known SLC families fall
67 into 4 phylogenetic clusters, which were termed α, β, γ, and δ groups. In addition, they have found 19
68 sequences that have previously not been described as SLCs. A later study [12] used a more sensitive HMM-
69 HMM (hidden Markov-model) comparison-based method to identify locally similar regions in known SLC
70 proteins. Visualization of the similarity network revealed visible protein clusters that correlated with existing
71 SLC families. In addition, they identified two unannotated protein sequences that showed similarity to
72 existing SLC proteins. A common limitation of these studies, however, is that they only searched for proteins
73 that were similar to proteins already annotated as human SLC transporters. Nevertheless, these efforts using
74 sequence similarity-based approaches have highlighted that there are additional as yet unannotated SLC
75 transporters in human protein databases.

76 In our current study, we aimed to identify missing SLC transporters that may differ from those currently
77 annotated in human. To do this, we turned to sequence databases and annotation systems that are
78 phylogenetically broader and not limited to human proteins, and we developed criteria to define “SLC-like”
79 proteins. Our method is thus more complete and general than previous approaches and tackles the task of
80 identifying SLC transporters starting from first principles. For this reason, we turned to the Transporter
81 Classification DataBase (TCDB) to enable the extraction of missing “SLC-like” proteins.

82 The Transporter Classification DataBase (TCDB) is an alternative classification system that was created in the
83 1990s in parallel to the SLC nomenclature series [14,15]. It collects transport-related membrane proteins,
84 including membrane receptors, transporters, ion channels, and membrane-anchored enzymes from all
85 kingdoms of life, with a particular focus on proteins from lower organisms. The proteins in the TCDB are
86 organized hierarchically into subfamilies, families and superfamilies based on phylogenetic and functional
87 considerations, and each member in the database is given a five-segmented TC# similar to the EC# that is
88 used for enzyme classification. In addition, a brief description is provided for each family that introduces
89 identified members and contains links to the most important relevant papers.

90 However, the TCDB dataset is not directly applicable for creating an overview of the collection of SLC
91 transporters encoded in the human genome (the human SLC-ome). One of the reasons is that the TCDB is set
92 up as a "representative database", which means that it only contains certain representative sequences from
93 each family. In addition, there is no particular focus on human proteins, and in fact several annotated human
94 SLC transporters are not present in the database.

95 It also seems problematic to consider all proteins in the TCDB that are annotated as part of the secondary
96 transporter superfamily TC# 2.A as being SLC-like. Indeed, many of the TCDB families annotated as part of
97 TC# 2.A exhibit structural or sequence features that do not match the characteristics of existing SLCs.
98 Examples of this are the Trk K^+ transporters (#2.A.38), which display an ion-channel like structural fold [16],
99 the GUP glycerol uptake proteins (#2.A.50), which exhibit enzymatic activity based on follow-up studies [17],
100 and the Twin Arginine Targeting (Tat) family (#2.A.64), which are actually protein secretion complexes
101 [18,19]. Since none of these families correspond to structural or functional characteristics of currently known
102 SLC transporters, it is likely that not all proteins annotated under the TC# 2.A superfamily are "SLC-like".
103 Thus, while the TCDB could be a rich source of information for finding new transporters, it is clear that the
104 perception of what a "transporter" should be according to TCDB does not always correspond to the typical
105 properties of well-characterized SLC proteins. Additional filtering of TCDB data is therefore necessary.

106 As part of the TransportDB project, there were parallel efforts to collect secondary transporters from human
107 and several other organisms [20–22]. Within the TransportDB project, the authors have built an automatic
108 transporter annotation pipeline (TransAAP), which relies on BLAST searches, the Clusters of Orthologous
109 Groups (COG) database [23], "selected HMMs for transporter protein families" [22] from the TIGRfam and
110 Pfam databases [24,25] and hydropathy predictions of TMHMM [26]. This pipeline was used as a
111 semiautomatic tool to annotate transporter-like proteins from the NCBI RefSeq database. However, based
112 on the currently available TransportDB website, the resulting protein hits are neither linked to protein
113 annotation databases, such as UniProt [27], nor are their official gene symbols or SLC names displayed.
114 Therefore, no correlation with the existing SLC nomenclature is provided, and it is not trivial to say whether
115 or not an existing SLC protein is included in the database.

116 We would also like to mention the Protein families (Pfam) database [28], which aims to maintain a curated
117 set of protein families, often represented by functional domains. Notably, Pfam provides curated HMMs for
118 each Pfam family to facilitate sequence similarity searches for the occurrence of those domains. In addition,
119 Pfam groups protein families into higher-order groups called clans. Pfam clans contain evolutionarily related
120 families whose relationships are supported either by sequence similarity, structural similarity or other
121 orthogonal biological evidence [29]. While many known Pfam models correspond to the functional regions of
122 known SLC transporters, Pfam neither attaches special importance to transport-related domains nor to
123 membrane-spanning domains. For this reason, while the Pfam database could be a rich source of
124 information about transporters, extracting Pfam families that encode transporter-like domains is non-trivial.

125 Therefore, there is a clear need in the field to define the criteria of "SLC-likeness" and to identify and classify
126 all proteins in humans and other species that exhibit "SLC-likeness". Thus, in our work, we interpret the term
127 "SLC-likeness" by defining the essential criteria for it, and we carry out an exhaustive search for proteins that
128 potentially meet these criteria, both with manual curation of datasets and with automatic sequence
129 similarity-based approaches.

Results

Elaboration of criteria for "SLC-likeness"

Since the TCDB takes a very inclusive approach to collecting membrane transport-related proteins from a broad range of biological organisms, we have selected it as the source database for our endeavors. However, as outlined in the introduction, selecting SLC-like proteins from the TCDB is non-trivial. Therefore, we have introduced a set of criteria based on current knowledge of SLC transporters in order to select SLC-like protein families from the TCDB. We believe that these criteria represent the most important properties of currently known SLC transporters. The criteria used to infer SLC-likeness were as follows.

(1) Structure of the protein should be α-helical, with at least three transmembrane helices (TMHs). Proteins with a β-barrel architecture, mostly β-structure proteins, membrane-anchored proteins, cyclic peptides and proteins consisting only of soluble domains (based on predictions or structural data) were excluded.

(2) The size of the transported substrate should fall within the small-molecule range (i.e. oligopeptides might be accepted as substrates but protein secretion systems are excluded). Also excluded are DNA-, RNA- and polysaccharide-transporting systems.

(3) Proteins with a channel-like mechanism were excluded, except in some rare cases. In particular, holins, toxins and other pore-forming proteins, and proteins bearing similarity to them, have been excluded.

(4) Nucleotide-driven transporters (e.g. ATP-binding cassette/ABC families) were excluded.

(5) Receptors that trigger endocytosis upon substrate binding were excluded. Only receptors were included where the receptor protein itself mediates the translocation of the substrate through the membrane, or the insertion of the substrate in the membrane, if that is its final location.

(6) Proteins with enzymatic activity, or similarity to known enzymes were excluded. In some cases, where the protein was believed to contain both a transport domain and a soluble enzyme domain, the proteins have been included.

155 (7) Proteins where transport activity was used as a synonym for trafficking (i.e. protein or vesicle
156 translocation within the cell) but otherwise seemingly having no small-molecule transmembrane
157 transport activity were excluded. On this basis, chaperones and other proteins helping the
158 insertion of nascent proteins into a cellular membrane were also excluded.
159 (8) For some (mostly putative) transporter families, TCDB does not give an explanation why the
160 proteins would be considered as transporters. Families with no resemblance to known
161 transporters and no indication or argument as to why they would be transporters were
162 excluded.

163 We must emphasize that for specific proteins or protein families, the verification of some of the criteria for
164 SLC-likeness as described above requires extensive and detailed information about the nature of the
165 transport phenomenon. For example, deciding whether the substrate is a small molecule requires the
166 identification of the transported solute, while ruling out a channel-like mechanism requires extensive
167 information about the transport mechanism itself. However, due to missing information, these criteria could
168 not always be verified while filtering the TCDB for "SLC-like" protein families. In certain cases, even the
169 identity of the protein that performs the actual membrane translocation within a transporter complex can
170 be unclear. For similar reasons, the historical discrepancy that the naming and classification of a
171 gene/protein generally precludes the detailed analysis of its structure and function has also caused the
172 currently known ("classic") SLC nomenclature to contain transport proteins with a channel-like mechanism
173 (e.g., the SLC41 family), single TMH (SLC27 family) as well as auxiliary proteins to actual transporters (SLC3
174 family). Some of these classic SLC families have been included in our selection in order to provide
175 consistency with the existing SLC nomenclature despite the fact that they do not meet some of our criteria
176 for SLC-likeness. Nevertheless, we believe that our criteria are broad enough to allow the identification of all
177 putative SLC-like transporters, while also being specific enough to distinguish them from other well-known,
178 non transport-related transmembrane protein families. Our criteria enabled us to initiate, for the first time,
179 an attempt to set up guidelines defining SLC-likeness, in the sense that explicit criteria are established to
180 disambiguate "transmembrane solute transport" and distinguish from other related membrane protein

181 activities such as channel-like transport, receptor-mediated endocytosis or protein secretion. In our current
182 study, we designate proteins and protein families that potentially meet the above criteria as "SLC-like".

183 Search for novel "SLC-like" proteins

184 The above-mentioned criteria were applied to manually select protein families in the TCDB that either fulfill
185 or can potentially fulfill the criteria for SLC-likeness based on the description of each third-level family from
186 the TCDB database. Throughout this manuscript, we use the term "TCDB family" to refer to third-level
187 groupings in the classification hierarchy (i.e. TC# x.y.z), while "subfamily" and "superfamily" refer to fourth-
188 level (TC# x.y.z.w) and second-level (TC# x.y) classes, respectively. In the work presented herein, we have
189 analyzed superfamilies TC# 1.A, 2.A, 9.A and 9.B, as well as the families and in certain cases the subfamilies
190 within them (see Table 1).

191 **Table 1. Families and subfamilies examined from the TCDB for SLC-like proteins.** Superfamilies marked in
192 boldface have been analyzed. The superfamily #1.A was examined since several existing SLC families are
193 classified here, while superfamily #2.A was expected to contain most known SLC proteins. Superfamilies #9.A
194 and #9.B were also examined. Total number of families and subfamilies in the first-level classes are indicated
195 in parentheses. Numbers show the total number of level 3 families and level 4 subfamilies in each
196 superfamily, as well as those found to be "SLC-like" (see text).

TCDB superfamilies	**Level 3 families (TC# x.y.z)**			**Level 4 subfamilies (TC# x.y.z.w)**		
	Total	**Total examined**	**Found to be "SLC-like"**	**Total**	**Total examined**	**Found to be "SLC-like"**
1 – Channels/Pores	(556)			(997)		
1.A – α-Type Channels		**112**	**7**		**321**	**17**
2 – Electrochemical Potential-driven Transporters	(178)			(561)		
2.A – Porters (uniporters, symporters, antiporters)		**129**	**114**		**556**	**501**
3 – Primary Active Transporters	(47)			(250)		
4 – Group Translocators	(18)			(36)		
5 – Transmembrane Electron Carriers	(17)			(42)		
8 – Accessory Factors Involved in Transport	(163)			(252)		
9 – Incompletely Characterized Transport Systems	(388)			(657)		

	9.A – Recognized Transporters of Unknown Biochemical Mechanism		**69**	**20**		**122**	**37**
	9.B – Putative Transport Proteins		**306**	**24**		**535**	**43**
	Total:	(1367)			(2795)		
			616	**166**		**1534**	**600**

197

198 It was a significant curation effort to manually assess the 1534 subfamilies within 616 families in the above-
199 mentioned superfamilies of the TCDB, which were expected to contain SLC transporter-like proteins (Table
200 1). To streamline our curation efforts, we have also expanded our analysis to Pfam protein families. To this
201 end, we used HMMER [30] to search for all Pfam models in the sequences of all TCDB members within the
202 four superfamilies analyzed, as shown in Table 1. The resulting Pfam families found in TCDB sequences were
203 manually analyzed in tandem with the TCDB families and subfamilies in which they occur whether they also
204 meet the criteria for SLC-likeness as defined above. The SLC-likeness criteria have been assessed according to
205 the following guidelines:

206 • As an inclusive approach, we mainly searched for information that can be used to exclude
207 (sub)families based on one or more criteria. Missing information (e.g. about transport mechanism,
208 structure, potential other functions, exact chemical identity of substrate) was not interpreted as a
209 reason for exclusion, unless it could be replaced by predictions (see below). Therefore, we strived to
210 include (sub)families in our list that potentially fulfill the criteria for being SLC-like.
211 • When a (sub)family could be excluded from being SLC-like based on one criterion, the remaining
212 criteria were not assessed.
213 • The functional criteria were assessed based on information presented on the TCDB web site in the
214 family description text and in the description of certain individual members, where this was present,
215 as well as in the textual descriptions of Pfam families. Families that claim their members to be
216 "channels" in their descriptions were assumed to function with a channel-like mechanism and were
217 therefore excluded, unless there was evidence for saturable transport or alternating access
218 mechanism for certain members. TCDB and Pfam families showing similarity to other transporter

219 families, either as mentioned in the family descriptions in text or *via* SCOOP-similarity [31], have
220 been chosen to potentially fulfill the functional criteria, unless specific functional annotations
221 indicated otherwise.

222 • The criterion about the size of the substrate was judged based on the known or putative substrate
223 according to TCDB family descriptions. The assumption that certain proteins are transporters,
224 whether based on experimental observations or purely on prediction, always entails the presence of
225 either a tested or an assumed class of substrates. Families where no indication was available why
226 they would function as transporters were thus excluded for the moment until further information on
227 their function becomes available.

228 • The structural criterion was assessed based on the "number of transmembrane segments"
229 annotations in the TCDB for each protein. (Sub)families where their TCDB descriptions mentioned a
230 β-barrel structure or relationship with such proteins were excluded. Care was taken to identify
231 outliers, i.e. proteins with less TMHs as the functional unit, which could potentially be fragments,
232 and proteins with extra TMHs in addition to the functional core. In this regard, we based our
233 decision on Pfam models conserved in the (sub)family and the number of THMs the model spans
234 according to predictions at the Pfam web site for the most characterized members of the
235 (sub)family. For certain Pfam families, structures and annotations from the Protein Data Bank (PDB)
236 linked with the family on the Pfam web site were used to decide whether or not the Pfam model
237 encodes soluble proteins, provided that the family model has sufficient coverage by the structure
238 available.

239 • In some TCDB families with multicomponent transport systems, the membrane-spanning core
240 components were identified by the presence of conserved Pfam models in all members of the
241 family. Whether Pfam domains were membrane-spanning was deduced by their descriptions at the
242 Pfam web site and Pfam prediction of TMHs. The TCDB description of the family also helped in
243 deciding which components were essential and membrane-spanning. If multiple such Pfam models
244 were found, the proteins were assumed to function as an obligate heteromultimer, and such

245 proteins were considered as a single entity to assess the overall number of TMHs for the structural
246 criterion.

247 Strikingly, of the 1534 subfamilies examined, only 600 in 166 families were found to be SLC-like according to
248 our criteria. In particular, of the 556 subfamilies within superfamily TC# 2.A ("porters"), only 501 appeared to
249 meet our selection criteria. This underlines once more that the term "porter" in relation to solute transport
250 is ambiguous in this field and a clearer definition of the perception of a solute transporter protein is needed.

251 Our curation efforts also yielded 209 Pfam families bearing SLC-like properties, which likely contain the
252 membrane-spanning regions of SLC-like proteins selected from the 336 Pfam families present in total in the
253 TCDB sequences analyzed. Special care has been taken to consistently include or exclude TCDB (sub)families
254 and their corresponding Pfam family, if applicable, from our selection. Notably, many of those Pfam families
255 that have been excluded represented soluble structural or regulatory domains. Following our initial round of
256 selection, we took advantage of clan groupings in Pfam and extended our selection efforts to analyze Pfam
257 families belonging to the same clan as the selected 209 SLC-like Pfam families. This was based on the
258 observation that many SLC-like transporter families in Pfam appear grouped into clans, so that members of
259 these clans might represent SLC-like transporters themselves. As noted before, Pfam clans represent
260 remotely related protein families, and while protein function is not always conserved across different
261 families within a Pfam clan, the inspection of these families whether they meet our SLC-likeness criteria is
262 validated by their relatedness or similarity to SLC-like families. Such a "clan expansion" procedure resulted in
263 12 additional Pfam families that are evolutionarily related to SLC-like protein families, of which 8 Pfam
264 families potentially meet our criteria as SLC-like, as evaluated according to the guidelines and criteria above.
265 Interestingly, these 8 SLC-like Pfam families currently have no representatives with a modest score (bit score
266 > 25) in the TCDB, while 4 of the 8 families are annotated in Pfam as domains of unknown function (DUF).
267 Thus, a total of 217 SLC-like Pfam families were identified in our search (see S1 Table).

268 We would like to note that our selection of SLC-like TCDB families and subfamilies as well as Pfam families is
269 limited by the availability of information on less characterized protein families and may therefore need to be

270 revised as more information becomes available. Our criteria are objective, however, and can easily be used
271 to revise the decision as to whether a particular family is SLC-like in the light of new information.

272 As a next step, we wanted to know whether the selected families either from the TCDB or from Pfam have
273 representatives in the proteomes of human and other clinically relevant organisms. For this analysis, we
274 selected 7 organisms due to their clinical relevance or scientific utility (*Homo sapiens*, *Rattus norvegicus*,
275 *Mus musculus*, *Gallus gallus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*), for which we
276 downloaded all sequences from the UniProt database [27], including Swiss-Prot (curated) and TrEMBL
277 (predicted) [32] entries. Sequences of proteins in the TCDB have been aligned within each SLC-like family and
278 subfamily and the alignments converted to HMMs for sensitive sequence similarity searches (see Methods).
279 In addition, HMMs of SLC-like Pfam families were downloaded from the Pfam database. HMM-based
280 similarity searches were then performed on the sequences downloaded for the 7 organisms to find proteins
281 similar to any of the SLC-like TCDB families or subfamilies, or SLC-like Pfam domains, followed by the
282 clustering of sequence fragments to arrive at one representative protein sequence per gene (see Methods).

283 The results of our search for SLC-like proteins are summarized in numbers in Table 2. Briefly, 59-67 of the
284 166 TCDB families have representatives in the 7 organisms (66 in human), and the organisms seem to
285 contain 434-673 SLC-like proteins (549 in human). In total, 3733 proteins have been found in the 7 organisms
286 studied. Notably, the number of SLC-like transporters found in human in our search is ~130 higher than
287 previously reported [2], indicating that the human SLC-ome may be significantly larger than previously
288 thought.

289 **Table 2. Results of the initial search for "SLC-like" proteins.** The search was performed using HMM-based
290 sequence similarity analysis based on selected families and subfamilies from the TCDB, as well as selected
291 Pfam models that likely encode transmembrane domains of "SLC-like" transporters (see text for details). The
292 table shows the number of families, subfamilies from the TCDB and the number of Pfam models that had
293 representative sequences in each organism. In addition, the initial number of "SLC-like" proteins found is
294 shown.

Species	Level 3 families (TC# x.y.z)	Level 4 subfamilies (TC# x.y.z.w)	Pfam models	Proteins
Human (*H. sapiens*)	66	308	79	549
Mouse (*M. musculus*)	67	314	79	532
Rat (*R. norvegicus*)	67	316	79	537
Zebrafish (*D. rerio*)	65	312	80	673
Chicken (*G. gallus*)	63	305	77	473
Fruitfly (*D. melanogaster*)	60	273	66	434
Roundworm (*C. elegans*)	59	283	69	535

After arriving at this initial set of SLC-like proteins, we proceeded by a performing a first round of sanitization based on human proteins. First, we manually investigated the UniProt records of all human hits that are non Swiss-Prot-validated sequences, which revealed that 19 out of 23 are likely non-human or fragment sequences (S2 Table). After excluding these, we investigated the number of helical transmembrane (TM) segments annotated in UniProt for 144 of all remaining 530 human hits that do not correspond to classic SLC transporters (families SLC1-SLC52). In total, 17 of these proteins, showing similarity to 8 different TCDB families, contained less than 3 TMHs based on UniProt annotations and thus did not satisfy our structural criterion (1). These proteins have been investigated individually based on their coverage of HMMs used in the search and the location of TMHs according to UniProt annotations in other SLC-like proteins covering the same HMMs, as well as available literature on oligomeric state and function. This way, we have retained 8 sequences where the functional unit contains 3 or more TMHs, or the protein spans at least 3 TMHs based on HMM coverage, or for other reasons (see S4 Table). These proteins have been included in Table 3 and S4 Table. In contrast, 10 of the 17 proteins seemed to cover regions of HMMs where other members did not seem to contain TMHs, while in addition containing less than 3 TMHs or known non-transporter TM domains. Notably, certain Pfam models seem to cover both TM and non-TM domains, and several TCDB families and their corresponding HMMs encode multi-domain proteins, leading to hits that show similarity to the model in the non-TM region, but not in the TM region that would be crucial for SLC-likeness. These 10 proteins have been excluded from further analyses and included in S2 Table. Furthermore, initial SLC-like hits

314 that only match with such non-TM regions of these HMMs (#2.A.19.3 positions 250-725, #9.B.64.1 positions
315 above 200, “OST3_OST6” positions below 200, “RBP_receptor” positions above 400) in all other organisms
316 have also been excluded from further analyses. Thus, in this initial sanitization step, 29 human proteins, as
317 well as 8, 10, 2, 9, and 6 proteins from *M. musculus*, *R. norvegicus*, *C. elegans*, *D. rerio*, and *G. gallus*,
318 respectively, have been excluded, leaving a remaining 3669 proteins for further analysis, including 520 from
319 human.

320 As a next step, we proceeded by attempting to classify the remaining proteins into protein families, followed
321 by manual database and literature searches focused on novel human proteins, as detailed below.

322 Classification into families

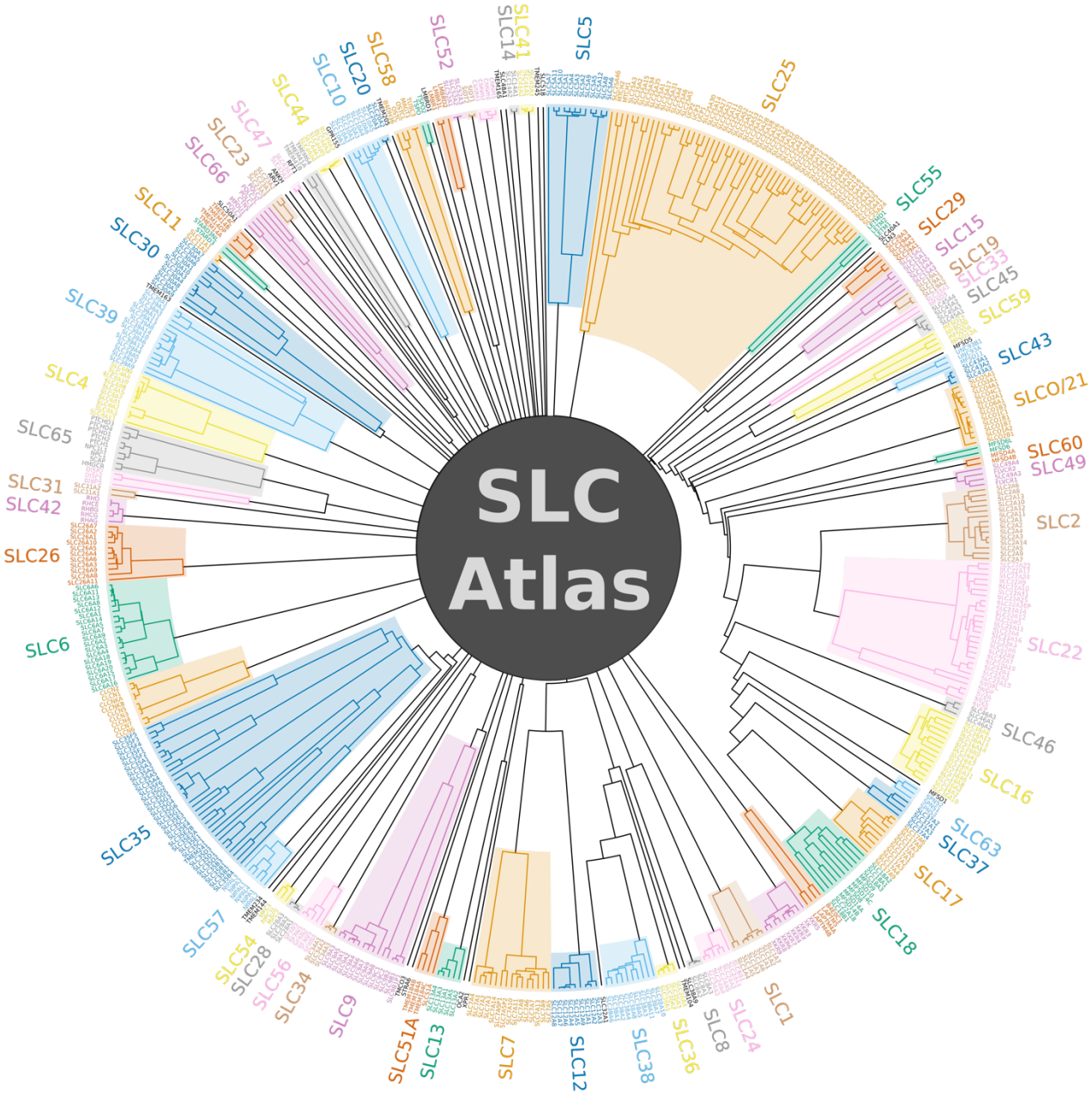
323 As mentioned in the introduction, SLC carriers are likely of polyphyletic origin, and individual families can be
324 so diverse that even sensitive sequence similarity-based methods may have difficulties grouping related SLCs
325 [12]. In our experience, multiple sequence alignment-based methods were not able to cluster the identified
326 SLC-like sequences and to reproduce known SLC families, so we devised a custom method for clustering
327 distantly related sequences into proteins families based on the introduction of “HMM fingerprints”. An
328 HMM fingerprint is a mathematical vector of numbers assigned to a protein sequence, where the numbers
329 represent the similarity scores of that protein sequence against each of the TCDB families, subfamilies and
330 Pfam families that we have selected to be SLC-like. Thus, two protein sequences that show a similar pattern
331 in their HMM fingerprints indicate their similarity. The usefulness of an HMM fingerprint depends on a
332 meaningful definition of HMMs used in the fingerprint, whereby we capitalize on the evolutionary principle
333 in the construction of TCDB families, subfamilies as well as Pfam families. Nevertheless, our goal was not to
334 reconstruct the evolutionary history of a set of proteins, but to group proteins that share similar sequence
335 features. However, due to the transitivity of homology and since similarity to a group of proteins suggests
336 homology, clusters derived using HMM fingerprinting are likely to contain homologous proteins.

337 The HMM fingerprint-based classification of the SLC-like proteins found in our search yielded 102 protein
338 families in total, 94 of which had representatives in human (Fig 1). For existing SLC transporters, the

339 generated families corresponded well to classical SLC families. Interestingly, outlier proteins were found in
340 several families, which did not cluster with their families at the threshold we used. Examples include SLC5A7,
341 SLC10A7, SLC25A46, SLC30A9, SLC39A9, MPDU1/SLC66A5 as well as the SLC9B family and SLC35 subfamilies.
342 This shows that the HMM fingerprints, and thus likely the sequences of these outlier proteins diverge from
343 those of other members of their families, and the sequences of subfamilies seem to diverge in certain cases.
344 Several classic SLC families also clustered together, such as SLC32-SLC36-SLC38, SLC2-SLC22, and SLC17-
345 SLC18-SLC37 proteins, likely due to their high sequence similarity. To ensure optimal correlation with the
346 preexisting classification of classic SLC proteins, we have introduced split and join constraints to keep 1)
347 closely related families separated, 2) outlier proteins merged with their families and 3) heterogenous
348 families merged (see Methods). These constraints did not affect the classification of novel SLC-like proteins.
349 The number of families proposed by our method is somewhat dependent on the threshold of clustering
350 used. Raising it to 0.8 lowers the number of families to 98 (90 in human), joining families SLC58-pSLC.OSTC,
351 SLC8-SLC24, SLC45-SLC59, and SLC65-pSLC.Dispatched, as well as pSLC.TSPO to a 2-membered subfamily of
352 unannotated proteins from *D. rerio*. Since our clustering method is based on shared HMMs, these families
353 are likely related. Indeed, SLC8 and SLC24 proteins constitute a superfamily of Na^+/Ca^{2+} antiporters [33],
354 while NPC1/SLC65A1 and Dispatched proteins share structural similarity [34]. Lowering the clustering
355 threshold to 0.6 generates 107 families (96 in human), splitting the pSLC.TMEM41-64 family to the TMEM41
356 and TMEM64 subfamilies, the MFSD3 proteins off the SLC33 family, an unannotated protein from *D.*
357 *melanogaster* off the XK protein family, as well as splitting an unannotated protein family from *D.*
358 *melanogaster* into two subfamilies.

359 **Figure 1. Dendrogram of human SLC-like protein families based on HMM fingerprint-based clustering.**
360 Classic and newly incorporated SLC families are shown with colors. Since SLC proteins are polyphyletic, which
361 manifests itself in mathematically orthogonal HMM fingerprints that cannot be further clustered, the
362 dendrogram does not join into a single branch. Branch lengths have been transformed for better visibility
363 (see Methods).

364



365

366 Our analysis revealed 42 new proteins families in total in human, containing proteins that have not yet been

367 annotated as SLCs. Interestingly, our search has also found new proteins that clustered into existing SLC

368 families (Table 3, S4 Table).

369 **Table 3. Novel SLC-like proteins.** The table shows SLC-like human proteins from our search that are not in

370 the classic group of SLCs (SLC1-52 families). The most similar (highest-scoring) TCDB family/subfamily and

371 Pfam family are shown, as well as the most similar (highest-scoring) protein with structural information

372 based on the pdb70 dataset (see Methods). Substrate information, where available, was retrieved from

373 available literature. Based on our search, SLC families SLC53-66 have recently been incorporated into the
374 nomenclature in collaboration with the HUGO/HGNC. Proteins marked with a single asterisk are included
375 due to sequence similarity but may have functions other than transporter. Proteins marked with double
376 asterisks might have a channel-like transport mechanism. For additional details, see S4 Table.

Gene symbol	TCDB family	Pfam family	Protein name	Structural fold	Substrates
SLC18 family					
MFSD8	2.A.1.2	MFS_1	Major facilitator superfamily domain-containing protein 8	MFS	orphan
MFSD9	2.A.1.2	MFS_1	Major facilitator superfamily domain-containing protein 9	MFS	orphan
MFSD10	2.A.1.2	MFS_1	Major facilitator superfamily domain-containing protein 10	MFS	indomethacin
MFSD14A	2.A.1.2	MFS_1	Hippocampus abundant transcript 1 protein	MFS	orphan
MFSD14B	2.A.1.2	MFS_1	Hippocampus abundant transcript-like protein 1	MFS	orphan
MFSD14C	2.A.1.2	MFS_1	Hippocampus abundant transcript-like protein 2	MFS	orphan
SLC22A18	2.A.1.2	MFS_1	Solute carrier family 22 member 18	MFS	orphan
SLC22 family					
SV2A (SLC22B1)	2.A.1.22	Sugar_tr	Synaptic vesicle glycoprotein 2A	MFS	galactose; also binds levetiracetam, seletracetam, brivaracetam.
SV2B (SLC22B2)	2.A.1.22	MFS_1	Synaptic vesicle glycoprotein 2B	MFS	orphan
SV2C (SLC22B3)	2.A.1.22	MFS_1	Synaptic vesicle glycoprotein 2C	MFS	orphan
SVOP (SLC22B4)	2.A.1.82	Sugar_tr	Synaptic vesicle 2-related protein	MFS	nicotinate; binds nucleotides in the TM region: 8-azido-ATP; ATP; GTP; TTP; CTP; NAD
SVOPL (SLC22B5)	2.A.1.82	MFS_1	Putative transporter SVOPL	MFS	orphan
SLC33 family					
MFSD3	2.A.1.57	Acatn	Major facilitator superfamily domain-containing protein 3	MFS	orphan
SLC35 family					
TMEM241	2.A.7.13	TPT	Transmembrane protein 241	NST	orphan
SLC42 family					
RHCE*	1.A.11.4	Ammonium_transp	Blood group Rh(CE) polypeptide	Amt	orphan
RHD*	1.A.11.4	Ammonium_transp	Blood group Rh(D) polypeptide	Amt	orphan
SLC51A family					
TMEM184A*	2.A.82	Solute_trans_a	Transmembrane protein 184A		orphan
TMEM184B	2.A.82	Solute_trans_a	Transmembrane protein 184B		orphan
TMEM184C	2.A.82	Solute_trans_a	Transmembrane protein 184C		orphan
SLC53 family					
XPR1 (SLC53A1)	2.A.94	EXS	Xenotropic and polytropic retrovirus receptor 1		phosphate
SLC54 family					
MPC1 (SLC54A1)	2.A.105	MPC	Mitochondrial pyruvate carrier 1		pyruvate
MPC2 (SLC54A2)	2.A.105	MPC	Mitochondrial pyruvate carrier 2		pyruvate
MPC1L (SLC54A3)	2.A.105	MPC	Mitochondrial pyruvate carrier 1-like protein		pyruvate
SLC55 family					
LETM1 (SLC55A1)	2.A.97	LETM1	Mitochondrial proton/calcium exchanger protein		K^+/H^+
LETM2 (SLC55A2)	2.A.97	LETM1	LETM1 domain-containing protein LETM2, mitochondrial		orphan
LETMD1 (SLC55A3)	2.A.97	LETM1	LETM1 domain-containing protein 1		orphan
SLC56 family					
SFXN1 (SLC56A1)	2.A.54	Mtc	Sideroflexin-1		L-serine

SFXN2 (SLC56A2)	2.A.54	Mtc	Sideroflexin-2		orphan
SFXN3 (SLC56A3)	2.A.54	Mtc	Sideroflexin-3		orphan
SFXN4 (SLC56A4)	2.A.54	Mtc	Sideroflexin-4		orphan
SFXN5 (SLC56A5)	2.A.54	Mtc	Sideroflexin-5		orphan
SLC57 family					
NIPA1 (SLC57A1)	2.A.7.25	Mg_trans_NIPA	Magnesium transporter NIPA1	NST	Mg^{2+}; to a lesser extent Sr^{2+}, Fe^{2+}, Co^{2+}
NIPA2 (SLC57A2)	2.A.7.25	Mg_trans_NIPA	Magnesium transporter NIPA2	NST	very specific to Mg^{2+}
NIPAL1 (SLC57A3)	2.A.7.25	Mg_trans_NIPA	Magnesium transporter NIPA3	NST	Mg^{2+}, Sr^{2+}, Ba^{2+}, Fe^{2+}, Cu^{2+}
NIPAL2 (SLC57A4)	2.A.7.25	Mg_trans_NIPA	NIPA-like protein 2	NST	Mg^{2+}, Sr^{2+}, Ba^{2+}
NIPAL3 (SLC57A5)	2.A.7.25	Mg_trans_NIPA	NIPA-like protein 3	NST	orphan
NIPAL4 (SLC57A6)	2.A.7.25	Mg_trans_NIPA	Magnesium transporter NIPA4	NST	orphan
SLC58 family					
MAGT1** (SLC58A1)	1.A.76.1	OST3_OST6	Magnesium transporter protein 1	MagT	Mg^{2+}
TUSC3 (SLC58A2)	1.A.76.1	OST3_OST6	Tumor suppressor candidate 3	MagT	Mg^{2+}, Fe^{2+}, Cu^{2+}, Mn^{2+}
SLC59 family					
MFSD2A (SLC59A1)	2.A.2.3	MFS_2	Sodium-dependent lysophosphatidylcholine symporter 1	MFS	LPC DHA (docosahexaenoic acid); LPC palmitate; also binds TopFluor LPE
MFSD2B (SLC59A2)	2.A.2.3	MFS_2	Major facilitator superfamily domain-containing protein 2B	MFS	sphingosine-1-phosphate (S1P)
MFSD12	2.A.2.7	MFS_2	Major facilitator superfamily domain-containing protein 12	MFS	cysteine
MFSD13A	2.A.2.3	MFS_2	Transmembrane protein 180	MFS	orphan
SLC60 family					
MFSD4A (SLC60A1)	2.A.1.7	MFS_1	Major facilitator superfamily domain-containing protein 4A	MFS	orphan
MFSD4B (SLC60A2)	2.A.1.7	MFS_1	Sodium-dependent glucose transporter 1	MFS	alpha-methyl-d-glucopyranoside, D-glucose, urea
SLC61 family					
MFSD5 (SLC61A1)	2.A.1.40	MFS_5	Molybdate-anion transporter	MFS	molybdate
SLC62 family					
ANKH (SLC62A1)	2.A.66.9	ANKH	Progressive ankylosis protein homolog	MATE	pyrophosphate
SLC63 family					
SPNS1 (SLC63A1)	2.A.1.49	MFS_1	Protein spinster homolog 1	MFS	orphan
SPNS2 (SLC63A2)	2.A.1.49	MFS_1	Protein spinster homolog 2	MFS	sphingosine-1-phosphate (S1P), phosphorylated Fingolimod (FTY720-P); dihydrosphingosine-1-phosphate (DH-S1P); phyto-S1P, C17-S1P
SPNS3 (SLC63A3)	2.A.1.49	MFS_1	Protein spinster homolog 3	MFS	orphan
SLC64 family					
TMEM165 (SLC64A1)	2.A.106.2	UPF0016	Transmembrane protein 165		Ca^{2+}/H^{+}, Mn^{2+}
SLC65 family					
NPC1 (SLC65A1)	2.A.6.6	Patched	NPC intracellular cholesterol transporter 1	RND	cholesterol
NPC1L1 (SLC65A2)	2.A.6.6	Patched	NPC1-like intracellular cholesterol transporter 1	RND	cholesterol
PTCH1	2.A.6.6	Patched	Protein patched homolog 1	RND	cholesterol
PTCH2	2.A.6.6	Patched	Protein patched homolog 2	RND	likely cholesterol based on functional similarity to PTCH1
PTCHD1	2.A.6.6	Patched	Patched domain-containing protein 1	RND	orphan
PTCHD3	2.A.6.6	Patched	Patched domain-containing protein 3	RND	orphan
PTCHD4	2.A.6.6	Patched	Patched domain-containing protein 4	RND	orphan

SCAP*	2.A.6.6	Patched	Sterol regulatory element-binding protein cleavage-activating protein	RND (half)	cholesterol, 25-hydroxycholesterol
HMGCR*	2.A.6.6	Patched	3-hydroxy-3-methylglutaryl-coenzyme A reductase	RND (half)	probably sterols
SLC66 family					
PQLC2 (SLC66A1)	2.A.43.2	PQ-loop	Lysosomal amino acid transporter 1 homolog	SWEET	L-Arg, L-Lys, L-His, L-canavanine, L-ornithine, cysteamine-cysteine mixed disulfide
PQLC1 (SLC66A2)	2.A.43	PQ-loop	PQ-loop repeat-containing protein 1	SWEET	orphan
PQLC3 (SLC66A3)	2.A.43.3	PQ-loop	PQ-loop repeat-containing protein 3	SWEET	orphan
CTNS (SLC66A4)	2.A.43.1	PQ-loop	Cystinosin	SWEET	L-cystine, L-selenocystine, L-cystathionine, H^+
MPDU1 (SLC66A5)	2.A.43.3	PQ-loop	Mannose-P-dolichol utilization defect 1 protein	SWEET	orphan
PQLC2L	2.A.43.2	PQ-loop	Putative uncharacterized protein PQLC2L	SWEET	orphan
pSLC.ARV1 family					
ARV1	9.A.19	Arv1	Protein ARV1		orphan
pSLC.Battenin					
CLN3	2.A.57.5	CLN3	Battenin	MFS	orphan
pSLC.CLCN					
CLCN1**	2.A.49.2	Voltage_CLC	Chloride channel protein 1	CLC	channel for chloride, thiocyanate, perchlorate, bromide, nitrate, chlorate, iodide
CLCN2**	2.A.49.2	Voltage_CLC	Chloride channel protein 2	CLC	channel for chloride, bromide, iodide
CLCN3	2.A.49.2	Voltage_CLC	H(+)/Cl(-) exchange transporter 3	CLC	antiport of chloride and H^+
CLCN4	2.A.49.2	Voltage_CLC	H(+)/Cl(-) exchange transporter 4	CLC	antiport of chloride and H^+, also nitrate, bromide, iodide
CLCN5	2.A.49.2	Voltage_CLC	H(+)/Cl(-) exchange transporter 5	CLC	antiport of chloride and H^+, also nitrate, bromide, iodide
CLCN6	2.A.49.3	Voltage_CLC	Chloride transport protein 6	CLC	antiport of chloride and H^+, also nitrate, iodide
CLCN7	2.A.49.3	Voltage_CLC	H(+)/Cl(-) exchange transporter 7	CLC	antiport of chloride and H^+
CLCNKA**	2.A.49.2	Voltage_CLC	Chloride channel protein ClC-Ka	CLC	channel for chloride, bromide, nitrate, iodide
CLCNKB**	2.A.49.2	Voltage_CLC	Chloride channel protein ClC-Kb	CLC	channel for chloride, bromide, nitrate, iodide
pSLC.CNNM					
CNNM1	1.A.112	DUF21	Metal transporter CNNM1	CorB	orphan
CNNM2	1.A.112	DUF21	Metal transporter CNNM2	CorB	Mg^{2+}, Co^{2+}, Mn^{2+}, Sr^{2+}, Ba^{2+}, Cu^{2+}, Fe^{2+}, Na^+
CNNM3	1.A.112	DUF21	Metal transporter CNNM3	CorB	orphan
CNNM4	1.A.112	DUF21	Metal transporter CNNM4	CorB	putatively Mg^{2+}, maybe antiport of Na^+
pSLC.CitMHS					
OCA2	2.A.45.2	CitMHS	P protein	AbgT	putatively chloride
pSLC.Dispatched					
DISP1	2.A.6.9	Patched	Protein dispatched homolog 1	RND	hedgehog protein
DISP2	2.A.6.9	Sterol-sensing	Protein dispatched homolog 2	RND	orphan
DISP3	2.A.6	Patched	Protein dispatched homolog 3	RND	putatively cholesterol
pSLC.GPR155					
GPR155	2.A.69.3	Mem_trans	Integral membrane protein GPR155	NhaA	orphan
pSLC.LAPTM					
	2.A.74.1	Mtp	cDNA FLJ61683, moderately similar to Lysosomal-associated multitransmembrane protein		orphan
LAPTM4A	2.A.74.1	Mtp	Lysosomal-associated transmembrane protein 4A	Tspn	thymidine, uridine, multi-drug resistance
LAPTM4B	2.A.74.1	Mtp	Lysosomal-associated transmembrane protein 4B		putatively ceramide, multi-drug resistance
LAPTM5	2.A.74.1	Mtp	Lysosomal-associated transmembrane protein 5	Tspn	orphan
pSLC.LMBR-A					
LMBRD1	9.A.54.1	LMBR1	Probable lysosomal cobalamin transporter		orphan
pSLC.LMBR-B					
LMBR1		LMBR1	Limb region 1 protein homolog		orphan
LMBR1L*	9.A.54.1	LMBR1	Protein LMBR1L		orphan
LMBRD2*	9.A.54.3	LMBR1	LMBR1 domain-containing protein 2		orphan
pSLC.MFSD1					
MFSD1	2.A.1.53	MFS_1	Major facilitator superfamily domain-containing protein 1	MFS	orphan
pSLC.MFSD6					
MFSD6	2.A.1.65	MFS_1_like	Major facilitator superfamily domain-containing protein 6	MFS	putatively sugars
MFSD6L	2.A.1.65	MFS_1_like	Major facilitator superfamily domain-containing protein 6-like	MFS	orphan
pSLC.OSTC					

	1.A.76.2	OST3_OST6	cDNA FLJ52625	MagT	orphan
OSTC	1.A.76.2	OST3_OST6	Oligosaccharyltransferase complex subunit OSTC	MagT	orphan
OSTCL	1.A.76.2	OST3_OST6	Oligosaccharyltransferase complex subunit-like	MagT	orphan
pSLC.RFT1					
RFT1	2.A.66.3	Rft-1	Protein RFT1 homolog	MATE	orphan
pSLC.SIDT					
SIDT1	1.A.79.1	SID-1_RNA_chan	SID1 transmembrane family member 1		cholesterol
SIDT2	1.A.79.1	SID-1_RNA_chan	SID1 transmembrane family member 2		cholesterol
pSLC.STAR					
STARD3	9.B.64	MENTAL	StAR-related lipid transfer protein 3		probably cholesterol
STARD3NL	9.B.64	MENTAL	STARD3 N-terminal-like protein		binds cholesterol
pSLC.STRA					
STRA6	2.A.90.1	RBP_receptor	Receptor for retinol uptake STRA6	STRA6	retinol
pSLC.TMCO3					
TMCO3	2.A.37.1	Na_H_Exchanger	Transmembrane and coiled-coil domain-containing protein 3	NhaA	orphan
pSLC.TMEM14					
TMEM14A	2.A.126.1	Tmemb_14	Transmembrane protein 14A	TMEM14	orphan
TMEM14B	2.A.126.1	Tmemb_14	Transmembrane protein 14B	TMEM14	orphan
TMEM14C	2.A.126.1	Tmemb_14	Transmembrane protein 14C	TMEM14	protoporphyrinogen IX
TMEM14DP	2.A.126.1	Tmemb_14	Transmembrane protein 14DP	TMEM14	orphan
pSLC.TMEM41-64					
TMEM41A	9.B.27.1	SNARE_assoc	Transmembrane protein 41A		orphan
TMEM41B	9.B.27.1	SNARE_assoc	Transmembrane protein 41B		scramblase of cholesterol, phosphatidylserine
TMEM64*	9.B.27.5	SNARE_assoc	Transmembrane protein 64		orphan
pSLC.TMEM104					
TMEM104	2.A.18.10	Aa_trans	Transmembrane protein 104	APC	orphan
pSLC.TMEM144					
TMEM144	2.A.7.8	TMEM144	Transmembrane protein 144	NST	orphan
pSLC.TMEM163					
TMEM163	2.A.4.8	Cation_efflux	Transmembrane protein 163	CDF	Zn^{2+}
pSLC.TMEM205					
TMEM205	9.A.55.1	DUF4149	Transmembrane protein 205		orphan
pSLC.TMEM234					
TMEM234	2.A.7.32	TMEM234	Transmembrane protein 234	NST	orphan
pSLC.TMEM245					
TMEM245	2.A.86	AI-2E_transport	Transmembrane protein 245		orphan
pSLC.TSPO					
TSPO	9.A.24.1	TspO_MBR	Translocator protein	TSPO	protoporphyrin IX, cholesterol
TSPO2	9.A.24.1	TspO_MBR	Translocator protein 2	TSPO	5-aminolevulinic acid
pSLC.UNC93					
MFSD11	2.A.1.58	UNC-93	UNC93-like protein MFSD11	MFS	orphan
UNC93A	2.A.1.58	UNC-93	Protein unc-93 homolog A	MFS	orphan
UNC93B1*	2.A.1.58	UNC-93	Protein unc-93 homolog B1	MFS	orphan
pSLC.XK					
XK	2.A.112.1	XK-related	Membrane transport protein XK	XK	orphan
XKR3	2.A.112.1	XK-related	XK-related protein 3	XK	orphan
XKR4	2.A.112.1	XK-related	XK-related protein 4	XK	phosphatidylserine
XKR5	2.A.112.1	XK-related	XK-related protein 5	XK	orphan
XKR6	2.A.112.1	XK-related	XK-related protein 6	XK	orphan
XKR7	2.A.112.1	XK-related	XK-related protein 7	XK	orphan
XKR8	2.A.112.1	XK-related	XK-related protein 8	XK	phosphatidylserine
XKR9	2.A.112.1	XK-related	XK-related protein 9	XK	phosphatidylserine
XKRX	2.A.112.1	XK-related	XK-related protein 2	XK	orphan

Structural homologues

For polytopic transmembrane proteins, structural similarity can support evolutionary relatedness, while at the same time evolutionary relatedness can provide a basis for homology-based model building efforts

381 [35,36]. On the other hand, the lack of predicted similarity to any protein of known structure could pin down
382 interesting targets for structural biology efforts by highlighting proteins likely belonging to new fold families.

383 We performed HMM-based searches on the pdb70 database (sequences of the proteins represented in the
384 Protein Data Bank clustered to 70% sequence identity, see Methods) to assess whether structural
385 homologues are available for the proteins found. In total, for 79 of the 102 families, at least one similar
386 protein was found with a corresponding structure in the Protein Data Bank. Importantly, 477 human SLC-like
387 proteins likely have a homologue whose structure has been solved. On the other hand, 43 human SLC-like
388 proteins belonging to 19 different families do not seem to have homologues with a known structure and
389 thus are likely to constitute novel fold families. For the classical SLC proteins, their best-scoring similar
390 proteins from the pdb70 dataset and the corresponding structural fold families are summarized in S3 Table.
391 Based on this, it appears that classical SLCs from families SLC34, SLC44, SLC48 and SLC51 are still "structural
392 orphans".

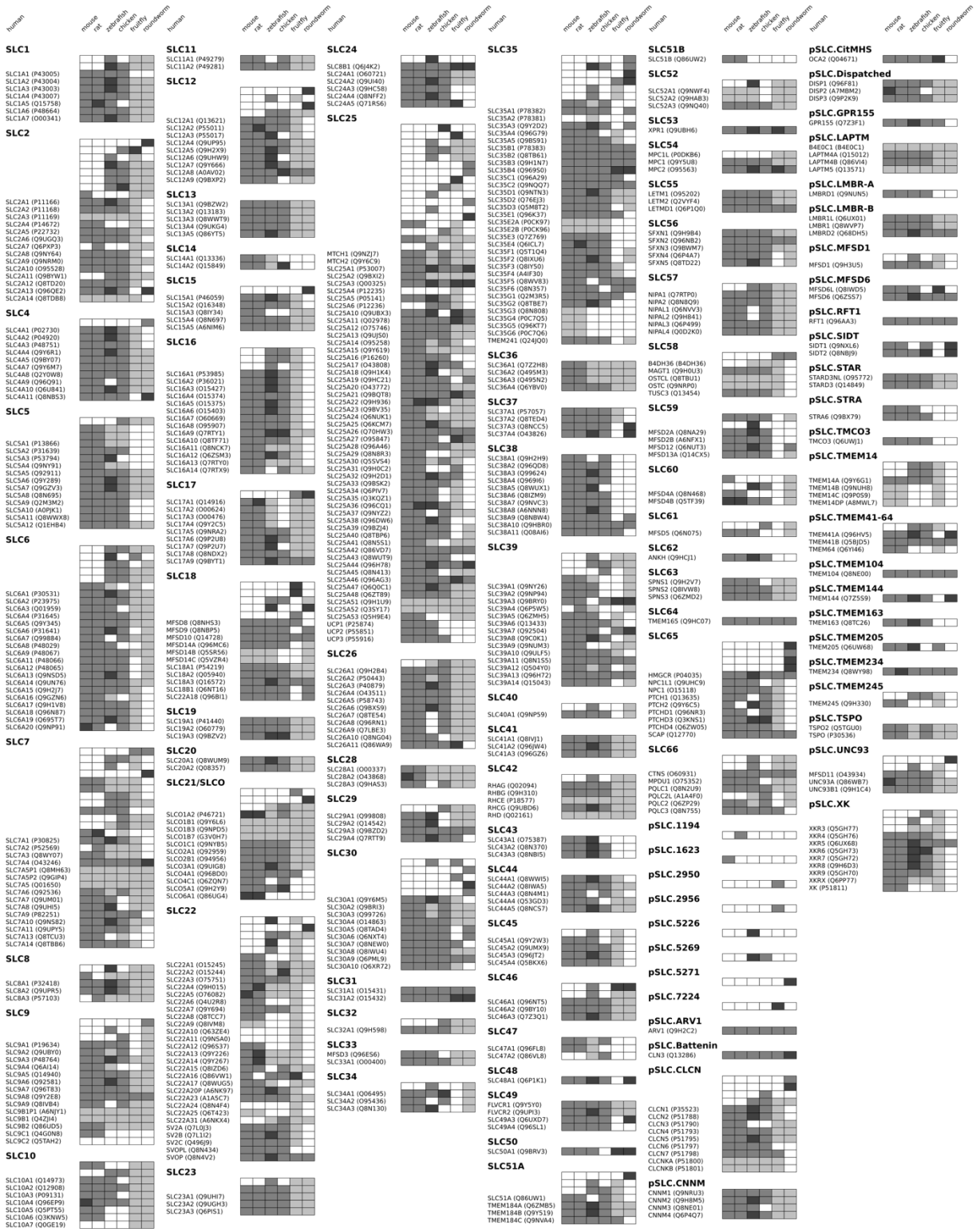
393 Detecting remote homology to proteins with a known transporter fold can also support their transporter
394 function as well as give clues about their transport mechanism. Fold families that are well-characterized and
395 typically host transporters with an alternating access mechanism encode potential transporter-like
396 structures. Out of the 18 fold families to which novel SLC-like proteins show similarity to (S4 Table),
397 representatives of 7 have been observed in various conformational states that hint at an alternating access
398 mechanism, while for a further 7 fold families, a transport mechanism has been suggested (S4 Table). These
399 fold families cover 87 of the novel SLC-like transporters, suggesting a possible transporter-like transport
400 mechanism for these proteins (S4 Table).

401 Phylogenetic analysis

402 Model organisms can be useful to study the biological function of various proteins, including solute
403 transporters. In order to relate the results to human, however, knowledge of orthologous gene pairs is
404 necessary. To this end, we performed phylogenetic analysis on each family of SLC-like proteins
405 corresponding to our clustering. In brief, unrooted phylogenetic trees for each SLC-like protein family from

406 all organisms have been generated and reconciled with the species tree of the 7 organisms in our study to
407 identify gene duplication and speciation events in their evolutionary histories (see Methods). The resulting
408 evolutionary trees are deposited in S1 File. Based on these trees, we carried out orthology analysis focusing
409 on human proteins and the human lineage. The resulting data is presented in Fig 2, showing relationships
410 between human genes and their orthologs in the other 6 organisms in our study. In addition, gene clusters
411 are presented that have arisen through gene duplication events in the evolutionary history of the human
412 lineage, but where the corresponding human genes have likely been lost.

413 **Figure 2. Orthology analysis of SLC-like proteins based on the human lineage.** Each line shows a gene
414 cluster that has evolved through gene duplication events along the human lineage. The human gene in the
415 cluster, if present, is noted in the text label. Normal grey boxes in each column indicate a 1:1 orthology
416 relationship between the human gene and a corresponding gene in the organism specified. Dark grey boxes
417 indicate that the human gene has several orthologs, and light grey boxes indicate that several human genes
418 share a common ortholog in that organism. White boxes denote no ortholog in the organism specified, while
419 lines without a human gene name correspond to genes that have been lost in human.


human
mouse
rat
zebrafish
chicken
fruitfly
roundworm
SLC1
SLC1A1 (P43005)
SLC1A2 (P43004)
SLC1A3 (P43003)
SLC1A4 (P43007)
SLC1A5 (Q15758)
SLC1A6 (P48664)
SLC1A7 (O00341)
SLC2
SLC2A1 (P11166)
SLC2A2 (P11168)
SLC2A3 (P11169)
SLC2A4 (P14672)
SLC2A5 (P22732)
SLC2A6 (Q9UGQ3)
SLC2A7 (Q6PXP3)
SLC2A8 (Q9NY64)
SLC2A9 (Q9NRM0)
SLC2A10 (O95528)
SLC2A11 (Q9BYW1)
SLC2A12 (Q8TD20)
SLC2A13 (Q96QE2)
SLC2A14 (Q8TDB8)
SLC4
SLC4A1 (P02730)
SLC4A2 (P04920)
SLC4A3 (P48751)
SLC4A4 (Q9Y6R1)
SLC4A5 (Q9BY07)
SLC4A7 (Q9Y6M7)
SLC4A8 (Q2Y0W8)
SLC4A9 (Q96Q91)
SLC4A10 (Q6U841)
SLC4A11 (Q8NBS3)
SLC5
SLC5A1 (P13866)
SLC5A2 (P31639)
SLC5A3 (P53794)
SLC5A4 (Q9NY91)
SLC5A5 (Q92911)
SLC5A6 (Q9Y289)
SLC5A7 (Q9GZV3)
SLC5A8 (Q8N695)
SLC5A9 (Q2M3M2)
SLC5A10 (A0PJK1)
SLC5A11 (Q8WWX8)
SLC5A12 (Q1EHB4)
SLC6
SLC6A1 (P30531)
SLC6A2 (P23975)
SLC6A3 (Q01959)
SLC6A4 (P31645)
SLC6A5 (Q9Y345)
SLC6A6 (P31641)
SLC6A7 (Q99884)
SLC6A8 (P48029)
SLC6A9 (P48067)
SLC6A11 (P48066)
SLC6A12 (P48065)
SLC6A13 (Q9NSD5)
SLC6A14 (Q9UN76)
SLC6A15 (Q9H2J7)
SLC6A16 (Q96ZN6)
SLC6A17 (Q9NP91)
SLC6A18 (Q96N87)
SLC6A19 (Q695T7)
SLC6A20 (Q9NP91)
SLC7
SLC7A1 (P30825)
SLC7A2 (P52569)
SLC7A3 (Q8WY07)
SLC7A4 (O43246)
SLC7A5P1 (Q8MH63)
SLC7A5P2 (Q9GIP4)
SLC7A5 (Q01650)
SLC7A6 (Q92536)
SLC7A7 (Q9UM01)
SLC7A8 (Q9UHI5)
SLC7A9 (P82251)
SLC7A10 (Q9NS82)
SLC7A11 (Q9UPY5)
SLC7A13 (Q8TCU3)
SLC7A14 (Q8TBB6)
SLC8
SLC8A1 (P32418)
SLC8A2 (Q9UPR5)
SLC8A3 (P57103)
SLC9
SLC9A1 (P19634)
SLC9A2 (Q9UBY0)
SLC9A3 (P48764)
SLC9A4 (Q6AI14)
SLC9A5 (Q14940)
SLC9A6 (Q92581)
SLC9A7 (Q96T83)
SLC9A8 (Q9Y2E8)
SLC9A9 (Q8IVB4)
SLC9B1P1 (A6NJY1)
SLC9B1 (Q4ZJI4)
SLC9B2 (Q86UD5)
SLC9C1 (Q4G0N8)
SLC9C2 (Q5TAH2)
SLC10
SLC10A1 (Q14973)
SLC10A2 (Q12908)
SLC10A3 (P09131)
SLC10A4 (Q96EP9)
SLC10A5 (Q5PT55)
SLC10A6 (Q3KNW5)
SLC10A7 (Q0GE19)
SLC11
SLC11A1 (P49279)
SLC11A2 (P49281)
SLC12
SLC12A1 (Q13621)
SLC12A2 (P55011)
SLC12A3 (P55017)
SLC12A4 (Q9UP95)
SLC12A5 (Q9H2X9)
SLC12A6 (Q9UHW9)
SLC12A7 (Q9Y666)
SLC12A8 (A0AV02)
SLC12A9 (Q9BXP2)
SLC13
SLC13A1 (Q9BZW2)
SLC13A2 (Q13183)
SLC13A3 (Q8WWT9)
SLC13A4 (Q9UKG4)
SLC13A5 (Q86YT5)
SLC14
SLC14A1 (Q13336)
SLC14A2 (Q15849)
SLC15
SLC15A1 (P46059)
SLC15A2 (Q16348)
SLC15A3 (Q8YT34)
SLC15A4 (Q8N697)
SLC15A5 (A6NIM6)
SLC16
SLC16A1 (P53985)
SLC16A2 (P36021)
SLC16A3 (O15427)
SLC16A4 (O15374)
SLC16A5 (O15375)
SLC16A6 (O15403)
SLC16A7 (O60669)
SLC16A8 (O95907)
SLC16A9 (Q7RTY1)
SLC16A10 (Q8TF71)
SLC16A11 (Q8NCK7)
SLC16A12 (Q6ZSM3)
SLC16A13 (Q7RTY0)
SLC16A14 (Q7RTX9)
SLC17
SLC17A1 (Q14916)
SLC17A2 (O00624)
SLC17A3 (O00476)
SLC17A4 (Q9Y2C5)
SLC17A5 (Q9NRA2)
SLC17A6 (Q9P2U8)
SLC17A7 (Q9P2U7)
SLC17A8 (Q8NDX2)
SLC17A9 (Q9BYT1)
SLC18
MFSD8 (Q8NH53)
MFSD9 (Q8NBP5)
MFSD10 (Q14728)
MFSD14A (Q96MC6)
MFSD14B (Q5SR56)
MFSD14C (Q5VZR4)
SLC18A1 (P54219)
SLC18A2 (O05940)
SLC18A3 (Q16572)
SLC18B1 (Q6NT16)
SLC22A18 (Q96BI1)
SLC19
SLC19A1 (P41440)
SLC19A2 (O60779)
SLC19A3 (Q9BZV2)
SLC20
SLC20A1 (Q8WUM9)
SLC20A2 (Q08357)
SLC21/SLCO
SLCO1A2 (P46721)
SLCO1B1 (Q9Y6L6)
SLCO1B3 (Q9NPD5)
SLCO1B7 (G3V0H7)
SLCO1C1 (Q9NYB5)
SLCO2A1 (O92959)
SLCO2B1 (O94956)
SLCO3A1 (Q9UIG8)
SLCO4A1 (Q96BD0)
SLCO4C1 (Q6ZQN7)
SLCO5A1 (Q9H2Y9)
SLCO6A1 (Q86UG4)
SLC22
SLC22A1 (O15245)
SLC22A2 (O15244)
SLC22A3 (O75751)
SLC22A4 (O94015)
SLC22A5 (O76082)
SLC22A6 (O4U2R8)
SLC22A7 (Q9Y694)
SLC22A8 (Q8TCC7)
SLC22A9 (Q8IVM8)
SLC22A10 (Q63ZE4)
SLC22A11 (Q9NSA0)
SLC22A12 (Q96S37)
SLC22A13 (Q9Y226)
SLC22A14 (Q9Y267)
SLC22A15 (Q8IZD6)
SLC22A16 (Q86VW1)
SLC22A17 (Q9WUG5)
SLC22A20P (A6NK97)
SLC22A23 (A1A5C7)
SLC22A24 (Q8N4F4)
SLC22A25 (Q6T423)
SLC22A31 (A6NKX4)
SV2A (Q7LJ3)
SV2B (Q7L1I2)
SV2C (Q496J9)
SVOPL (Q8N434)
SVOP (Q8N4V2)
SLC23
SLC23A1 (Q9UHI7)
SLC23A2 (Q9UGH3)
SLC23A3 (Q6PIS1)
SLC24
SLC8B1 (Q6J4K2)
SLC24A1 (O60721)
SLC24A2 (Q9UJ40)
SLC24A3 (Q9HC58)
SLC24A4 (Q8NFF2)
SLC24A5 (Q71RS6)
SLC25
MTCH1 (Q9NZJ7)
MTCH2 (Q9Y6C9)
SLC25A1 (P53007)
SLC25A2 (Q9BXI2)
SLC25A3 (Q00325)
SLC25A4 (P12235)
SLC25A5 (P05141)
SLC25A6 (P12236)
SLC25A10 (Q9UBX3)
SLC25A11 (O02978)
SLC25A12 (O75746)
SLC25A13 (Q9UJS0)
SLC25A14 (O95258)
SLC25A15 (O9Y619)
SLC25A16 (P16260)
SLC25A17 (O43808)
SLC25A18 (O9H1K4)
SLC25A19 (O9HC21)
SLC25A20 (O43772)
SLC25A21 (O9BQT8)
SLC25A22 (O9H936)
SLC25A23 (O9BV35)
SLC25A24 (Q6NUK1)
SLC25A25 (Q6KCM7)
SLC25A26 (O70HW3)
SLC25A27 (O95847)
SLC25A28 (Q96A46)
SLC25A29 (Q8N8R3)
SLC25A30 (Q5SV54)
SLC25A31 (O9H0C2)
SLC25A32 (O9H2D1)
SLC25A33 (Q9BSK2)
SLC25A34 (Q6PIV7)
SLC25A35 (O3KQZ1)
SLC25A36 (Q96CQ1)
SLC25A37 (Q9NYZ2)
SLC25A38 (Q96DW6)
SLC25A39 (Q9BZJ4)
SLC25A40 (Q8TBP6)
SLC25A41 (Q8N5S1)
SLC25A42 (Q86VD7)
SLC25A43 (Q8WUT9)
SLC25A44 (Q96H78)
SLC25A45 (Q8N413)
SLC25A46 (Q96AG3)
SLC25A47 (Q6QC1)
SLC25A48 (Q6ZT89)
SLC25A51 (Q9H1U9)
SLC25A52 (Q3SY17)
SLC25A53 (Q5H9E4)
UCP1 (P25874)
UCP2 (P55851)
UCP3 (P55916)
SLC26
SLC26A1 (Q9H2B4)
SLC26A2 (P50443)
SLC26A3 (P40879)
SLC26A4 (O43511)
SLC26A5 (P58743)
SLC26A6 (Q9BX59)
SLC26A7 (Q8TE54)
SLC26A8 (Q96RN1)
SLC26A9 (Q7LBE3)
SLC26A10 (Q96RG04)
SLC26A11 (Q86WA9)
SLC28
SLC28A1 (O00337)
SLC28A2 (O43868)
SLC28A3 (Q9HA53)
SLC29
SLC29A1 (O99808)
SLC29A2 (O14542)
SLC29A3 (Q98ZD2)
SLC29A4 (Q7RTT9)
SLC30
SLC30A1 (Q9Y6M5)
SLC30A2 (Q9BRI3)
SLC30A3 (Q99726)
SLC30A4 (O14863)
SLC30A5 (Q8TAD4)
SLC30A6 (Q6NXT4)
SLC30A7 (Q8NEW0)
SLC30A8 (Q8IWU4)
SLC30A9 (Q6PML9)
SLC30A10 (Q6XR72)
SLC31
SLC31A1 (O15431)
SLC31A2 (O15432)
SLC32
SLC32A1 (Q9H598)
SLC33
MFSD3 (Q96E56)
SLC33A1 (O00400)
SLC34
SLC34A1 (Q06495)
SLC34A2 (O95436)
SLC34A3 (Q8N130)
SLC35
SLC35A1 (P78382)
SLC35A2 (P78381)
SLC35A3 (Q9Y2D2)
SLC35A4 (Q96G79)
SLC35A5 (Q9B591)
SLC35B1 (P78383)
SLC35B2 (Q8TB61)
SLC35B3 (Q9H1N7)
SLC35B4 (Q96950)
SLC35C1 (Q96A29)
SLC35C2 (Q9NQQ7)
SLC35D1 (Q9NTN3)
SLC35D2 (Q76EJ3)
SLC35D3 (Q5M8T2)
SLC35E1 (Q96K37)
SLC35E2A (P0CK97)
SLC35E2B (P0CK96)
SLC35E3 (Q7Z769)
SLC35E4 (Q6ICL7)
SLC35F1 (Q5T1Q4)
SLC35F2 (Q8IXU6)
SLC35F3 (Q8IY50)
SLC35F4 (A4IF30)
SLC35F5 (Q8WV83)
SLC35F6 (Q8N357)
SLC35G1 (Q2M3R5)
SLC35G2 (Q8TBE7)
SLC35G3 (Q8N808)
SLC35G4 (P0C7Q5)
SLC35G5 (Q96KT7)
SLC35G6 (P0C7Q6)
TMEM241 (Q24JQ0)
SLC36
SLC36A1 (Q7Z2H8)
SLC36A2 (Q495M3)
SLC36A3 (Q495N2)
SLC36A4 (Q6YBV0)
SLC37
SLC37A1 (P57057)
SLC37A2 (Q8TED4)
SLC37A3 (Q8NCC5)
SLC37A4 (O43826)
SLC38
SLC38A1 (Q9H2H9)
SLC38A2 (Q96QD8)
SLC38A3 (Q99624)
SLC38A4 (Q96J96)
SLC38A5 (Q8WUX1)
SLC38A6 (Q8IZM9)
SLC38A7 (Q9NVC3)
SLC38A8 (A6NNN8)
SLC38A9 (Q8NBW4)
SLC38A10 (Q96HBR0)
SLC38A11 (Q08AI6)
SLC39
SLC39A1 (Q9NY26)
SLC39A2 (Q9NP94)
SLC39A3 (Q9BRY0)
SLC39A4 (Q6P5W5)
SLC39A5 (Q6ZMH5)
SLC39A6 (Q13433)
SLC39A7 (Q92504)
SLC39A8 (Q9C0K1)
SLC39A9 (Q9NUM3)
SLC39A10 (Q9ULF5)
SLC39A11 (Q8N15S)
SLC39A12 (Q504Y0)
SLC39A13 (Q96H72)
SLC39A14 (Q15043)
SLC40
SLC40A1 (Q9NP59)
SLC41
SLC41A1 (Q8IVJ1)
SLC41A2 (Q96JW4)
SLC41A3 (Q96GZ6)
SLC42
RHAG (Q02094)
RHBG (Q9H310)
RHCE (P18577)
RHCG (Q9UBD6)
RHD (Q02161)
SLC43
SLC43A1 (O75387)
SLC43A2 (Q8N370)
SLC43A3 (Q8NBI5)
SLC44
SLC44A1 (Q8WWI5)
SLC44A2 (Q8IWA5)
SLC44A3 (Q8NAM1)
SLC44A4 (Q53GD3)
SLC44A5 (Q8NCS7)
SLC45
SLC45A1 (Q9Y2W3)
SLC45A2 (Q9UMX9)
SLC45A3 (Q96JT2)
SLC45A4 (Q5BKX6)
SLC46
SLC46A1 (Q96NT5)
SLC46A2 (Q9BY10)
SLC46A3 (Q7Z3Q1)
SLC47
SLC47A1 (Q96FL8)
SLC47A2 (Q86VL8)
SLC48
SLC48A1 (Q6P1K1)
SLC49
FLVCR1 (Q9Y5Y0)
FLVCR2 (Q9UPI3)
SLC49A3 (Q6UXD7)
SLC49A4 (Q96SL1)
SLC50
SLC50A1 (Q9BRV3)
SLC51A
SLC51A (Q86UW1)
TMEM184A (Q6ZMB5)
TMEM184B (Q9Y519)
TMEM184C (Q9NVA4)
SLC51B
SLC51B (Q86UW2)
SLC52
SLC52A1 (Q9NWF4)
SLC52A2 (Q9HAB3)
SLC52A3 (Q9NQ40)
SLC53
XPR1 (Q9UBH6)
SLC54
MPC1L (P0DKB6)
MPC1 (Q9Y5U8)
MPC2 (O95563)
SLC55
LETM1 (O95202)
LETM2 (Q2VYF4)
LETMD1 (Q6P1Q0)
SLC56
SFXN1 (Q9H9B4)
SFXN2 (Q96NB2)
SFXN3 (Q9BWM7)
SFXN4 (Q6P4A7)
SFXN5 (Q8TD22)
SLC57
NIPA1 (Q7RTP0)
NIPA2 (Q8NBQ9)
NIPAL1 (Q6NVV3)
NIPAL2 (Q9H8I4)
NIPAL3 (Q6P499)
NIPAL4 (Q0D2K0)
SLC58
B4DH36 (B4DH36)
MAGT1 (Q9H0U3)
OSTCL (Q8TBU1)
OSTC (Q9NRP0)
TUSC3 (Q13454)
SLC59
MFSD2A (Q8NA29)
MFSD2B (A6NFX1)
MFSD12 (Q6NUT3)
MFSD13A (Q14CX5)
SLC60
MFSD4A (Q8N468)
MFSD4B (Q5TF39)
SLC61
MFSD5 (Q6N075)
SLC62
ANKH (Q9HCJ1)
SLC63
SPNS1 (Q9H2V7)
SPNS2 (Q8IVW8)
SPNS3 (Q6ZMD2)
SLC64
TMEM165 (Q9HC07)
SLC65
HMGCR (P04035)
NPC1L1 (Q9UHC9)
NPC1 (O15118)
PTCH1 (Q13635)
PTCH2 (Q9Y6C5)
PTCHD1 (Q96NR3)
PTCHD3 (Q3KNS1)
PTCHD4 (Q6ZW05)
SCAP (Q12770)
SLC66
CTNS (O60931)
MPDU1 (O75352)
PQLC1 (Q8N2U9)
PQLC2L (A1A4F0)
PQLC2 (Q6ZP29)
PQLC3 (Q8N755)
pSLC.1194
pSLC.1623
pSLC.2950
pSLC.2956
pSLC.5226
pSLC.5269
pSLC.5271
pSLC.7224
pSLC.ARV1
ARV1 (Q9H2C2)
pSLC.Battenin
CLN3 (Q13286)
pSLC.CLCN
CLCN1 (P35523)
CLCN2 (P51788)
CLCN3 (P51790)
CLCN4 (P51793)
CLCN5 (P51795)
CLCN6 (P51797)
CLCN7 (P51798)
CLCNKA (P51800)
CLCNKB (P51801)
pSLC.CNNM
CNNM1 (Q9NRU3)
CNNM2 (Q9H8M5)
CNNM3 (Q8NE01)
CNNM4 (Q6P4Q7)
pSLC.CitMHS
OCA2 (Q04671)
pSLC.Dispatched
DISP1 (Q96F81)
DISP2 (A7MBM2)
DISP3 (Q9P2K9)
pSLC.GPR155
GPR155 (Q7Z3F1)
pSLC.LAPTM
B4E0C1 (B4E0C1)
LAPTM4A (Q15012)
LAPTM4B (Q86VI4)
LAPTM5 (Q13571)
pSLC.LMBR-A
LMBRD1 (Q9NUN5)
pSLC.LMBR-B
LMBR1L (Q6UX01)
LMBR1 (Q8WVP7)
LMBRD2 (Q68DH5)
pSLC.MFSD1
MFSD1 (Q9H3U5)
pSLC.MFSD6
MFSD6L (Q8IWD5)
MFSD6 (Q6Z5S7)
pSLC.RFT1
RFT1 (Q96AA3)
pSLC.SIDT
SIDT1 (Q9NXL6)
SIDT2 (Q8NBJ9)
pSLC.STAR
STARD3NL (O95772)
STARD3 (Q14849)
pSLC.STRA
STRA6 (Q9BX79)
pSLC.TMCO3
TMCO3 (Q6UWJ1)
pSLC.TMEM14
TMEM14A (Q9Y6G1)
TMEM14B (Q9NUH8)
TMEM14C (Q9P059)
TMEM14DP (A8MWL7)
pSLC.TMEM41-64
TMEM41A (Q96HV5)
TMEM41B (Q5BJD5)
TMEM64 (Q6Y146)
pSLC.TMEM104
TMEM104 (Q8NE00)
pSLC.TMEM144
TMEM144 (Q7Z5S9)
pSLC.TMEM163
TMEM163 (Q8TC26)
pSLC.TMEM205
TMEM205 (Q6UW68)
pSLC.TMEM234
TMEM234 (Q8WY98)
pSLC.TMEM245
TMEM245 (Q9H330)
pSLC.TSPO
TSPO2 (Q5TGU0)
TSPO (P30536)
pSLC.UNC93
MFSD11 (O43934)
UNC93A (Q86WB7)
UNC93B1 (Q9H1C4)
pSLC.XK
XKR3 (Q5GH77)
XKR4 (Q5GH76)
XKR5 (Q6UX68)
XKR6 (Q5GH73)
XKR7 (Q5GH72)
XKR8 (Q9H6D3)
XKR9 (Q5GH70)
XKRX (Q6PP77)
XK (P51811)

420

421

422 Literature search on newly found SLC-like proteins

423 As alluded to above, our search was followed by thorough investigation of the available literature to check
424 whether a description of transport activity for the newly found proteins is available. Surprisingly, our search
425 has revealed human proteins for which transported substrates are known. These data have been included in
426 Table 3. While examining the available literature of these proteins, we have also extracted information
427 related to our criteria for SLC-likeness, especially with a focus on structure, function and transport
428 mechanism. In total, 53 proteins could be assigned a small-molecule substrate and therefore fulfill our
429 criterion for substrate size (2), and for a further 6, putative small-molecule substrates have been suggested.
430 Nevertheless, 74 proteins, while showing sequence similarity to transporters, are still "orphans" with no
431 indication of a substrate. We have not found any indication in the literature that any of the proteins found in
432 our search would have nucleoside-driven transport activity, therefore all of these proteins likely also fulfill
433 our criterion 4. Interestingly, 9 proteins have been found that have described roles other than
434 transmembrane solute transport, while showing similarity to SLC-like transporter proteins. These proteins
435 potentially do not fulfill our functional criteria (5, 6, 7), they are marked with an asterisk in Table 3 and the
436 details are further evaluated in S4 Table. In total, 3669 potentially SLC-like proteins have been found, and
437 out of the 520 human proteins, 134 proteins that are not classic SLC transporters showed similarity to SLC-
438 like proteins. Ruling out proteins with suggested other functions and a channel-like mechanism leaves 119
439 proteins, out of which 79 show similarity to transporters with a proposed transport mechanism, while 47
440 could be assigned endogenous small-molecule substrates.

441 Discussion

442 Our curation and search efforts have revealed a surprising 134 human proteins that are SLC-like but have not
443 been officially part of the SLC nomenclature before the start of our study. Interestingly, around 30 of them
444 were addressed in an earlier study and referred to as "atypical SLC transporters" [10,11]. All of these atypical
445 transporters have also been identified in our study, together with many others, including several that are not

446 part of the major facilitator superfamily (MFS) or the amino acid-polyamine-cation (APC) transporter
447 superfamily.

448 Our HMM fingerprint-based classification method yielded 102 protein families. Interestingly, in order to
449 reproduce some of the classic SLC families, we had to introduce clustering constraints to artificially split or
450 merge predicted families. In turn, increasing or decreasing the clustering threshold by 0.1 changes the
451 number of families to 98 and 107, respectively. Researchers have reported similar problems with the
452 classification of the highly diverse protein set of the Pfam database [29]. In particular, Finn et al. note that
453 closely related Pfam families may have artificially high thresholds to prevent them from overlapping, while
454 divergent families cannot always be covered by a single model [29]. Thus, we believe that however protein
455 families are shaped, the discovery of evolutionary relationships between them will always be necessary to
456 define higher-order superfamilies, similarly to Pfam clans [29]. Similarly, the analysis of the evolutionary
457 relationship of proteins within each family is equally validated, and several classic SLC families indeed
458 contain established subfamilies based on sequence and/or functional comparisons [37–40]. In this regard,
459 we believe that our work provides correlation between transporter proteins in the studied organisms with
460 the hierarchy of families and subfamilies in the TCDB database on the one hand, and a scale-free framework
461 for protein grouping and classification through HMM fingerprints on the other hand. Still, more work is
462 needed to examine the evolutionary relatedness of different SLC-like families using more robust
463 phylogenetic approaches.

464 The phylogenetic trees of SLC-like protein families presented in S1 File can be instrumental for functional
465 annotation based on orthology analysis in lower organisms, as well for studying species-specific differences
466 in transport pathways. Of particular note, different orthologs of hepatic drug transporters are present in
467 humans and experimental mammals used in pre-clinical studies, contributing to imperfect prediction of drug
468 half-life and toxicity in animal models [41]. On the other hand, genetic ablation of transporters and
469 phenotype studies in lower organisms could shed light on the function of their human orthologs. However,
470 as can be seen in Fig 2 and the phylogenetic trees in S1 File, a significant number of SLC-like protein
471 sequences were found in *D. melanogaster* and *C. elegans* that do not appear to have orthologs in the higher

472 organisms studied here. Most of these protein sequences are poorly annotated and their expression and
473 function have yet to be confirmed.

474 The interesting cases of apparent outlier sequences in several families (SLC5A7, SLC10A7, SLC25A46,
475 SLC30A9, SLC39A9, MPDU1/SLC66A5 as well as the SLC9B and SLC35 families) have previously been partially
476 documented. These members show striking sequence and/or functional divergence from other members of
477 their SLC families. SLC5A7 diverges from other SLC5 proteins in phylogenetic trees [42,43] and shares only
478 20-25% sequence identity with them [44]. The sequence of SLC10A7 is more similar to its bacterial relatives
479 than to other SLC10 members, and its genetic structure, particularly its number of exons, is also different
480 from other proteins in the SLC10 family [45,46]. SLC10A7 also appears to have diverged at the functional
481 level, as it has been suggested to be a regulator of Ca^{2+} influx [47], while having no documented transport
482 activity. SLC25A46 turns out to be an outer mitochondrial membrane protein [48], in contrast to most other
483 members of the SLC25 family, which are generally located on the inner mitochondrial membrane. Consistent
484 with this, its yeast ortholog Ugo1, referred to as "a degenerate member of the mitochondrial metabolite
485 carrier family", was reported to be part of the mitochondrial outer membrane fusion machinery [49].
486 Similarly, MTCH1 and MTCH2, which also seem to be distantly related to other SLC25 members according to
487 our dendrogram (Fig 1), are also outer mitochondrial membrane proteins [50,51], and together with
488 SLC25A46, are collectively referred to as "peculiar" members of the family [52]. SLC30A9 is found in the
489 cytosol and nuclear fractions and functions as a coactivator of nuclear receptors after hormonal stimulation
490 [53], in contrast to other members of the family, which are Zn^{2+} transporters. SLC39A9, the only member of
491 "subfamily I" of Zrt/Irt-like proteins (ZIPs) [37] might function as an androgen hormone receptor [54], while
492 most other family members are primarily transporters of divalent metal ions. SLC9B subfamily proteins
493 NHA1 and NHA2 of the SLC9 Na^+/H^+ exchanger family are more similar to their bacterial homologs than to
494 SLC9A members [55].

495 In the following sections, we discuss the non-classic set of human SLC-like proteins resulting from our search,
496 with a particular focus on the identities of the transported substrates and, if known, the structural and
497 mechanistic aspects of transport. In certain cases, the proteins have already been officially included in the

498 SLC nomenclature as per our initiative and as a result of the precursor of this work, following approval by the
499 Human Gene Nomenclature Committee (HGNC). Where existing information about substrate and function is
500 not available, we have speculated on these aspects using available information in the literature and
501 considerations of sequence similarity. For several proteins, however, their classification requires further
502 consideration. We would also like to articulate open questions about what further work is needed in order to
503 identify additional transporters. We believe our approach has been useful to pinpoint proteins that have a
504 high probability of being novel transporters. Thus, specific biological, biochemical and structural efforts
505 could focus on these specific targets highlighted in our work, all of which would contribute to a complete
506 assessment of the SLC-ome in human cells.

507 Transporters with existing evidence for transport function

508 Importantly, our search highlighted several proteins for which our literature search uncovered previously
509 reported evidence of transporter activity. Several of the proteins in these families have been assigned SLC
510 family numbers and, in collaboration with the HGNC, included in the SLC nomenclature. Among these hits
511 are several mitochondrial transporters (MPC/SLC54, LETM/SLC55, Sideroflexins/SLC56), which have been
512 reviewed before [56]. In terms of transported substrate, many of the proteins with documented transport
513 activity appear to be ion transporters or exchangers (Table 3). In the next paragraph, we will highlight certain
514 proteins and families that have particularly caught our attention.

515 Interestingly, the **SLC60** family contains two MFS-like proteins (MFSD4A, MFSD4B), of which MFSD4B has
516 been shown to transport D-glucose and urea [57,58].

517 The **SLC61A1** protein (MFSD5) is the only protein in human that shows similarity to the #2.A.1.40 family of
518 molybdate transporters and contains the “MFS_5” Pfam model. It has been claimed to be the homolog of
519 similar transporters from algae and plants, and complementation assays suggested its ability to transport
520 molybdate [59]. While molybdenum is a biologically active trace element, not much is known about its
521 transport and homeostasis in human [60].

522 Interestingly, the **TMEM163** protein clustered together with SLC30 zinc transporter (ZnT) proteins, since the
523 "Cation_efflux" Pfam model, representative of the SLC30 family, was present in its sequence, although at a
524 low score and non-significant e-value (2.9e-3). In the TCDB, TMEM163 is also classified under subfamily
525 #2.A.4.8, sharing a common family with SLC30 transporters (#2.A.4.2). Multiple sequence alignment as well
526 as pairwise alignments with existing SLC30 members reveal very low sequence identity with SLC30 proteins
527 (4.2-14.4%), albeit these numbers are similar to those of SLC30A9 (6.5-13.4%). Given the marginal similarity
528 to the "Cation_efflux" domain, it is tempting to assume that SLC30 proteins and TMEM163 are distantly
529 related. Indeed, TMEM163 has been shown to bind [61] and transport Zn^{2+} [62–64], and substitution of its
530 proposed substrate-binding residues with alanine abolished Zn^{2+} efflux activity [64]. Transport has been
531 demonstrated to be H^{+}-coupled, and the protein functioning as a dimer [62], while extruding Zn^{2+} from the
532 cell [64]. Intracellularly, TMEM163 was originally shown to be expressed in synaptic vesicles [65]. In
533 overexpression systems, it is localized to both the plasma membrane and intracellular membrane
534 compartments [64]. TMEM163 has been linked to Parkinson's disease (PD) [66], even though the opposite
535 conclusion has also been drawn [67]. TMEM163 has also been reported to be upregulated by olanzapine, a
536 psychotropic drug prescribed for PD patients [68]. In addition, TMEM163 was also shown to be highly
537 expressed in insulin secretory vesicles in human pancreas [69], and has been identified as a risk factor in type
538 2 diabetes [70,71]. Disruption of TMEM163 expression might impair insulin secretion at high glucose stimuli
539 [69].

540 The **TMEM165** protein clustered into its own family and is the only protein in human containing the
541 "UPF0016" Pfam domain and showing similarity to TCDB family #2.A.106.2. TMEM165 is a member of a
542 highly conserved family of transmembrane proteins that is present in many species of eukaryotes and
543 bacteria [72]. Initially, TMEM165 and its yeast homolog, Gdt1p, have been hypothesized to be Ca^{2+}/H^{+}
544 exchangers [73,74]. However, recently, evidence has been mounting about its involvement in manganese
545 (Mn^{2+}) homeostasis [74], and both Ca^{2+} and Mn^{2+} transport activity has directly been shown [75]. TMEM165
546 is localized to the trans-Golgi in human cells [72], and is proposed to play a crucial role in regulating Mn^{2+}
547 uptake into the Golgi apparatus [72,74]. In line with this, its homologs in other organisms, also containing

548 the UPF0016 domain, are also annotated as Mn^{2+} transporters [74]. Manganese plays an important role as a
549 co-factor for enzymes involved in glycosylation, and impairment of TMEM165 function results in
550 glycosylation defects. Indeed, mutations of TMEM165 found in patients with congenital disorder of
551 glycosylation (CDG) type II hamper the transport function or localization of TMEM165 [75]. Due to the
552 importance of TMEM165 in lactate biosynthesis [76], it has also been suggested that TMEM165 could be a
553 transporter importing both Ca^{2+} [77] and Mn^{2+} into the Golgi in exchange for protons [74]. TMEM165
554 proteins contain two copies of the UPF0016 domain, and each domain contains a signature motif, E-φ-G-D-
555 (K/R)-(T/S), where φ denotes a hydrophobic amino acid. The glutamic acid of the second motif, E248, has
556 been shown to be crucial for affecting the glycosylation function of the Golgi but not the expression of the
557 protein [78], and so can be speculated to form part of a binding site for transporter function. However, in
558 the absence of an experimentally determined structure, further investigation will be required to understand
559 the transport mechanism of TMEM165.

560 Proteins with sequence similarity to existing transporters

561 Our search uncovered a large number of proteins that show sequence similarity and thus possible
562 relationships to existing transporters in the SLC nomenclature. Since transport activity has not been
563 demonstrated, these proteins are either orphan transporters or they could have transceptor functions. What
564 follows is a comprehensive discussion of these proteins, as their similarity to transporters makes them ideal
565 targets for further studies to elucidate their putative transporter activity.

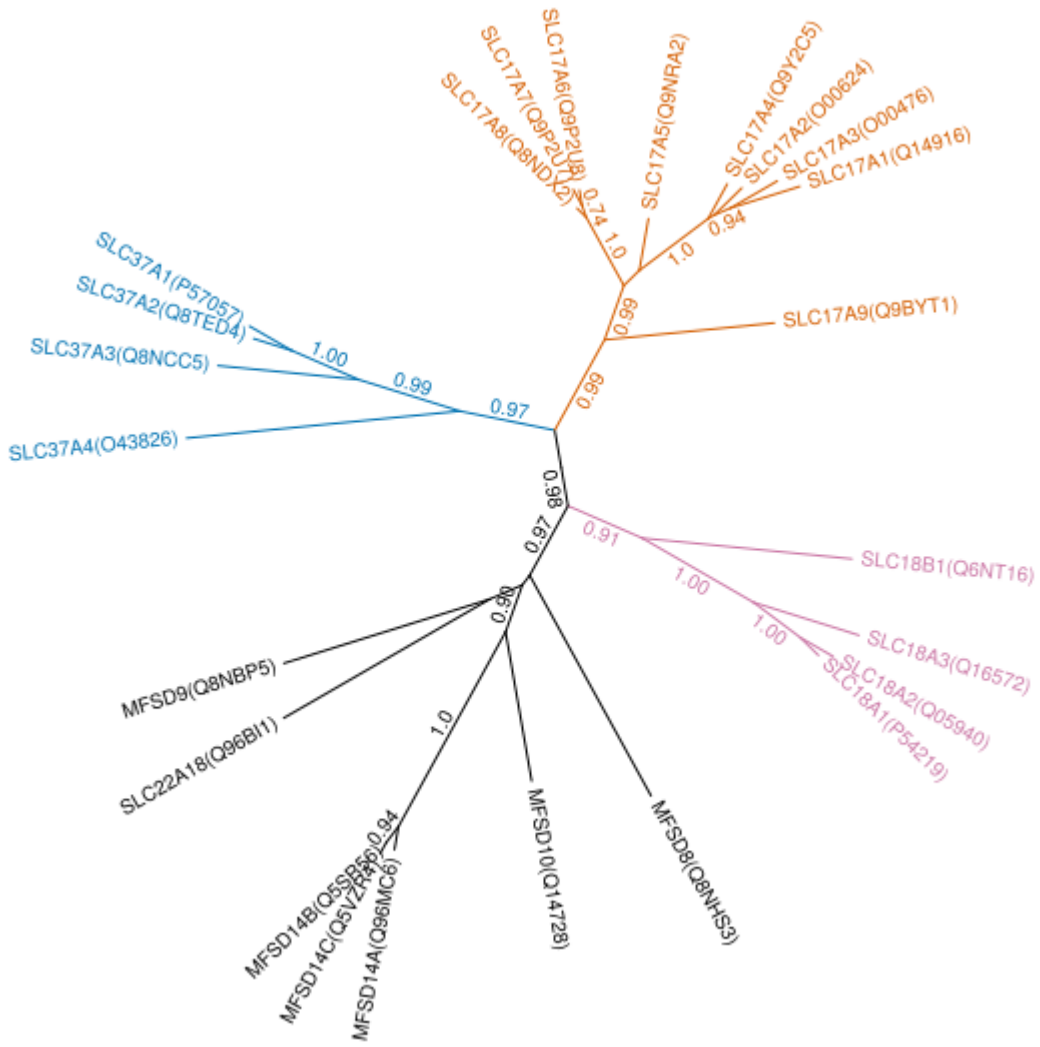
566 Atypical transporters

567 A previous effort by Perland and coworkers has uncovered novel transporter-like proteins mostly from the
568 MFS and APC superfamilies [11], which have also been recognized by our search. In general, the function of
569 these atypical transporters is not well known, but some have been reported to be expressed in the brain,
570 and their expression levels seem to be affected by nutrient availability [79–82]. For MFSD1, MFSD6 and
571 UNC93A, the study of the *D. melanogaster* and *C. elegans* orthologs have provided some information on the
572 loss-of-function phenotype [83–86]. MFSD8 and MFSD10 have been linked to the Wolf-Hirschhorn syndrome
573 and to LINCL (late-infantile-onset neuronal ceroid lipofuscinoses), respectively [87,88], and MFSD8 seems to

574 be localized in the lysosomes [89,90]. More studies about the biological function and transport activity of
575 these proteins is required to fully understand their physiological roles.

576 Some of the "atypical" SLC-like transporters (e.g. MFSD8, MFSD9, MFSD10 and MFSD14 proteins) clustered
577 together with members of the classical SLC18 family, which prompted us to examine the relationship of
578 these and neighboring proteins in more detail. We constructed multiple alignments and a phylogenetic tree
579 of the proteins one level above these proteins in our clustering dendrogram (i.e., members of the SLC17,
580 SLC18, SLC37 families as well as MFSD8, MFSD9, MFSD10, MFSD14A-C and SLC22A18 proteins, Fig 3). As
581 expected, the phylogenetic tree gives a better separation of these very similar proteins than the HMM
582 fingerprint-based dendrogram, and the branch support values suggest a clear separation of the SLC17, SLC18
583 and SLC37 families. In addition, the phylogenetic tree highlights that the atypical SLC proteins MFSD8,
584 MFSD9, MFSD10, MFSD14A-C as well as SLC22A18, while being more divergent, are likely to have evolved
585 from a single common ancestor. The relationship between the MFSD9, MFSD10 and MFSD14A-B proteins
586 also agrees with earlier studies [10,11]. Similarly, the evolutionary dendrogram created using all 7 organisms
587 in our study for the SLC18 family (S1 File) suggests that MFSD9, MFSD10, MFSD14A-B and SLC22A18 likely
588 share a common evolutionary origin and are thus more closely related to each other than to SLC18 proteins,
589 while MFSD8 is more distantly related. Further studies may be required to elucidate the particular
590 evolutionary relationship between these proteins.

591 **Figure 3. Phylogenetic tree of the SLC17, SLC18 and SLC37 families and proteins clustering in their**
592 **neighborhood.** Branch support values ≥ 0.7 are shown.



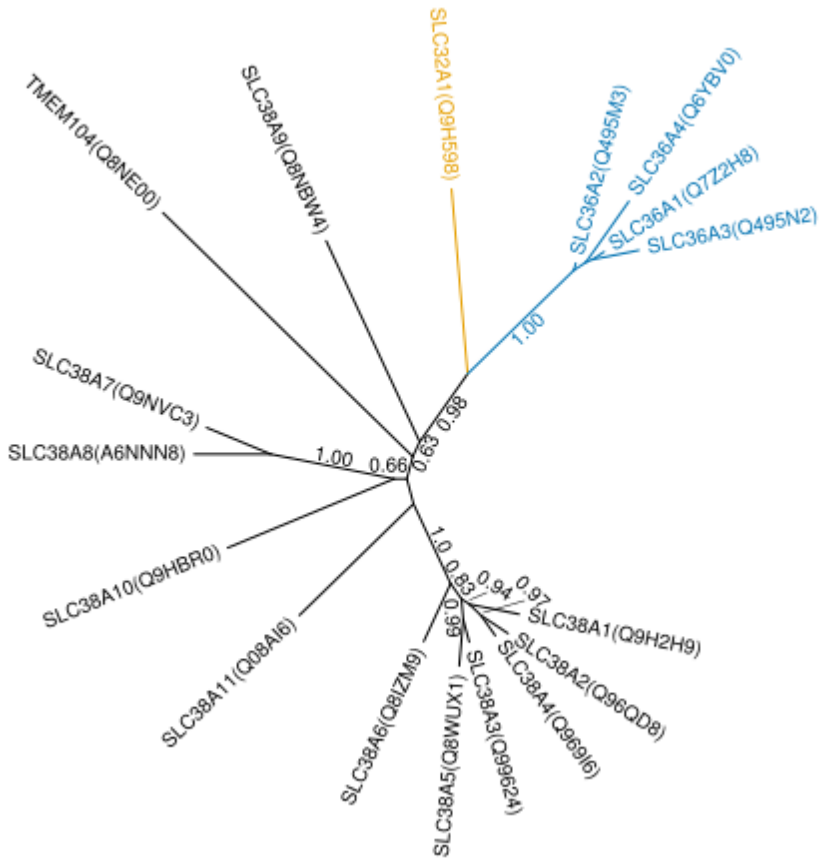
593
594

595 Interestingly, **MFSD3** has clustered together with the SLC33A1 protein in our HMM fingerprint-based
596 clustering analysis. Indeed, the "Acatn" (Acetyl-coenzyme A transporter 1) Pfam domain is present in MFSD3,
597 albeit with a relatively low score, but with significant e-value (5.6e-12). Sequence alignment between
598 SLC33A1 and MFSD3 gives 18.2% sequence identity. Even though the sequence identity between MFSD3 and
599 SLC33A1 is relatively low, the "Acatn" domain was found only in these proteins. The relatedness of MFSD3
600 and SLC33A1 is also corroborated by previous results of other groups [81]. The biological function of MFSD3
601 is still unclear [81,91,92].

602 The **TMEM104** protein in our analysis clustered together with amino acid transporter families SLC32, SLC36
603 and SLC38. Based on multiple and pairwise sequence alignments and sequence identity, TMEM104 was most

604 similar to SLC38A7 (13.3-15.1%), SLC38A8 (13.1-15.1%), and SLC36A1 (10.9-16.1%). Interestingly, TMEM104
605 also bears moderate similarity to the “Aa_trans” Pfam domain, which describes the transmembrane region
606 of SLC38 proteins. In our SLC classification dendrogram (Fig 1), TMEM104 clustered with SLC38 proteins,
607 even though it seems to be an outlier from the family, similarly to SLC38A9. In addition, TMEM104, SLC38A7
608 and SLC38A8 all show low similarity to the “Trp_Tyr_perm” Pfam domain, which describes bacterial tyrosine
609 and tryptophan permeases. Despite the low sequence similarity to SLC38 members, these data suggest that
610 TMEM104 might be an amino acid transporter distantly related to the SLC38 family. To get a more detailed
611 picture of the evolutionary relationship of TMEM104 and the SLC38, SLC36 and SLC32 families, we
612 constructed a multiple alignment and a phylogenetic tree of these proteins (Fig 4). While the tree
613 undoubtedly separates the SLC32 and SLC36 clades due to high branch support values, TMEM104 could not
614 be clearly separated from the SLC38 family, and it likely has a similar relationship to the rest of the family as
615 SLC38A9, which is playing a transceptor role in cells [93]. However, there is currently no experimental
616 evidence for this and the biological function of TMEM104 remains elusive.

617 **Figure 4. Phylogenetic tree of proteins clustering in the SLC38-36-32 families and TMEM104.** Branch
618 support values ≥ 0.5 are shown. SLC32 and SLC36 family members are colored gold and blue, respectively.



Proteins similar to SLC35 transporters

Interestingly, our search revealed several proteins that show sequence similarity to transporters of the SLC35 family. SLC35 proteins are currently classified into subfamilies A-G, which have relatively low sequence identity among them (4.0-22.0%). SLC35 transporters belong to the family of “DMT” (drug-metabolite transporters), which is classified in TCDB family #2.A.7, and corresponds to a clan of Pfam families, called DMT. Currently known substrates of human SLC35 members include nucleotide-sugar conjugates [40]. However, the substrate range of this superfamily is substantially larger [94].

The **TMEM144** proteins harbors its dedicated Pfam model called “TMEM144”, which itself is a member of the DMT clan of transporters. Its relatedness to the DMT family is further corroborated by high-scoring similarity of the TCDB subfamily #2.A.7.8 to the sequence of TMEM144. Otherwise, functionally, the protein is uncharacterized, although it might be related to sterol metabolism/transport, because its function has

632 been linked to bovine milk cholesterol levels [95], the hypothalamic-gonadal axis and testosterone response
633 [96]. It is also highly expressed in the hypothalamus [96].

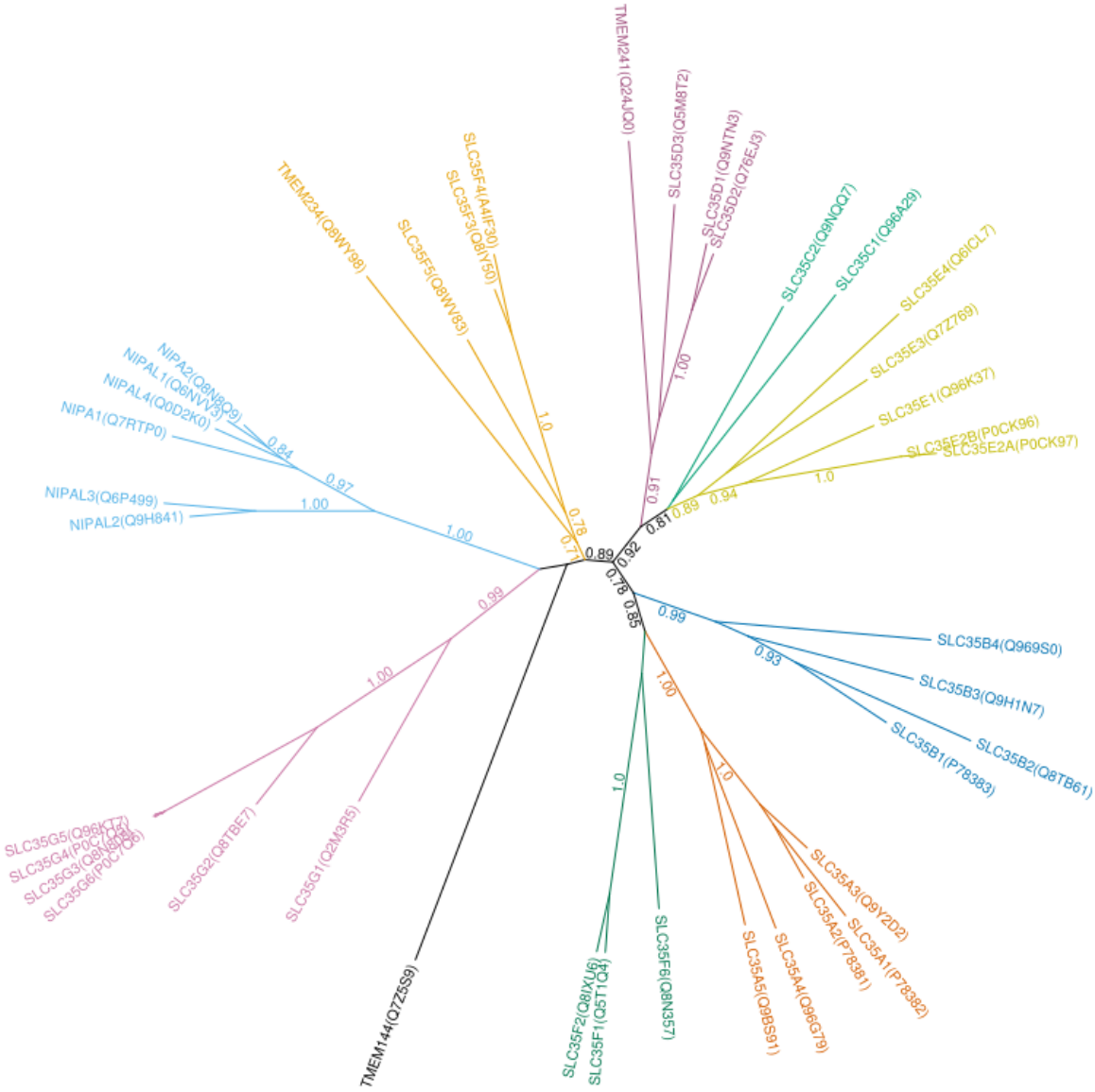
634 **TMEM234** is classified in the TCDB #2.A.7.32 family and also contains a corresponding "TMEM234" Pfam
635 domain, which is a member of the "DMT" clan of Pfam domains as well. The physiological role of TMEM234
636 is not known. However, in zebrafish, its homolog might play a role in the formation of the kidney filtration
637 barrier, as its knockdown causes proteinurea [97].

638 In our clustering analysis, **TMEM241** clustered with the SLC35 family very closely. Its HMM fingerprint shows
639 similarity to the TC# 2.A.7.13 subfamily, and weak similarity to the "TPT" Pfam model (which also belongs to
640 the "DMT" Pfam clan) over the whole length of the protein. The proteins in the #2.A.7.13 family are Golgi
641 GDP-mannose:GMP antiporters from plants, yeast and other organisms [98,99], but not from vertebrates.
642 Nevertheless, the protein seems to be present in many higher organisms according to the Swiss-Prot
643 database. However, these protein are not listed in the TCDB. The biological function of TMEM241 is still
644 unknown, but it has been suggested to affect serum triglyceride levels [100].

645 Due to the sequence diversity of the SLC35 family, we were interested in the relationships between
646 individual proteins. To this end, we have built a phylogenetic tree of human proteins that showed similarity
647 to existing SLC35 transporters (Fig 5). In the tree, most SLC35 subfamilies could be resolved as a single clade,
648 while TMEM241 and TMEM234 form clades with SLC35D and SLC35F3-5 proteins with a support value of
649 0.91 and 0.71, respectively. The relationship of TMEM241 with the SLC35D subfamily is also supported by
650 our HMM fingerprint-based clustering results. TMEM241 shows 12.0-21.4% sequence identity with SLC35D
651 proteins. In contrast, our phylogenetic tree with SLC35 proteins from all 7 organisms (S1 File) indicated that
652 TMEM241 is most closely related to SLC35E4. On the other hand, TMEM234 only weakly associated with
653 SLC35F proteins, with sequence identities 3.6-9.2%. TMEM144 appears to be only distantly related to SLC35
654 proteins. The elucidation of evolutionary relationships between proteins in the SLC35 thus likely requires
655 further investigation.

656 **Figure 5. Phylogenetic tree of proteins with HMM fingerprints overlapping with the SLC35 family.** Branch

657 support values ≥ 0.7 are shown. The different subfamilies are colored in various colors.



658

659

660 ## Others

661 **GPR155** is an enigmatic protein that seems to be a concatenation of a membrane transporter domain (Pfam:

662 "Mem_trans") and a G-protein coupled receptor (GPCR) domain, which might be the reason why it is

663 annotated as a GPCR. The membrane transporter part seems to be most similar to TCDB #2.A.69.3 subfamily

664 proteins, which are annotated as malate/malonate transporters in the Auxin Efflux Carrier (AEC) family

665 (#2.A.69). Gene structure analysis suggests that the concatenation is real [101], and both the human, mouse
666 and fruitfly proteins seem to contain 17 TMHs according to UniProt annotations.

667 The "Mem_trans" domain is only present in GPR155 from all human proteins analyzed, and matches the first
668 10 TMHs of the protein in a 5+5 arrangement. In our structural search, the second half of this transporter
669 domain (TMHs 6-10) of GPR155 exhibits similarity to the N-terminal half of sodium/bile transporters of the
670 AsbT fold (SLC10 family). On the other hand, the last 7 TMHs of GPR155 (TMHs 11-17) indeed show
671 similarity to GPCR-fold (7-TM) proteins with known structure, with highest similarity to structures of the
672 human Smoothened receptor homolog (PDB ID: 6OT0). Interestingly, in our search, the N-terminal half of the
673 transporter domain of GPR155 did not show any structural homologues.

674 The precise function of GPR155 still remains elusive. However, because highest expression levels were found
675 in the brain, especially in GABAergic neurons, it might play a role in GABAergic neurotransmission [101]. It
676 also has been suggested that GPR155 might play a role in neurons involved in motor brain function as well as
677 sensory information processing [101]. In *D. melanogaster*, knockdown of the homologous gene, "anchor",
678 resulted in increased wing size and thickened veins [102]. This phenotype was similar to what appeared in
679 bone morphogenetic protein (BMP) signaling gain-of-function experiments [102]. GPR155 has also been
680 linked to a number of different cancers [103,104].

681 The **RFT1** protein was originally thought to be a scramblase of lipid-linked origosaccharides [105]. However,
682 these molecules have at least 12-14 sugar moieties, so given their size, it is unlikely that a single transporter
683 could catalyze their flipping. Later studies refuted the scramblase concept and suggested instead that RFT1
684 could serve as an accessory protein to a flippase, but would not act as a flippase itself [106–108].

685 Nevertheless, the corresponding "Rft1" Pfam model shows similarity to multidrug and toxic compound
686 extrusion (MATE) transporters (SLC47 family) and belongs to a clan of Pfam domains ("MviN_MATE") that
687 contains transporters as well. In line with this, human RFT1 showed significant similarity to MATE
688 transporters in our search for structural homologs, indicating likely structural similarity. Some members of

689 the corresponding TC# 2.A.66.3 subfamily also contain weak hits of the "MatE" Pfam domain. Thus, while
690 RFT1 shows similarity to existing transporters, its biological function is still unclear.

691 The C-terminal half of the **TMEM245** protein shows weak similarity to TC# 2.A.86 proteins (Autoinducer-2
692 Exporter/AI-2E family), which contain both small-molecule exporters [109,110] as well as Na^+/H^+ antiporter
693 proteins [111–113]. Accordingly, TMEM245 also has weak similarity to the corresponding Pfam model ("AI-
694 2E_transport").

695 The HMMs of TC# 2.A.86 and #2.A.86.1 match residues 444-866 of human TMEM245, which are the last 6
696 TMHs according to UniProt predictions. The last 5 TMHs are separated from the previous ones by a slightly
697 larger loop. This architecture is similar to the 3+5 arrangement of the previously described bacterial
698 $Na^+(Li^+)/H^+$ antiporter TC# 2.A.86.1.14 according to UniProt predictions (accession code: NLHAP_HALAA). This
699 bacterial protein also matched the full-length "AI-2E_transport" domain from Pfam, while only the last 5
700 TMHs of TMEM245 match with C-terminal region of "AI-2E_transport". The human TMEM245 protein
701 contains 14 TMHs in total according to UniProt predictions. Thus, TMEM245 might have a transporter-like
702 domain at the C-terminus. In terms of the structure, we have not found any similarities to proteins with
703 known structure. Therefore both TMEM245 and bacterial exporters and antiporters in the TC# 2.A.86 family
704 are likely to have a yet uncharacterized tertiary structure. Functionally, the TMEM245 protein also remains
705 elusive.

706 The **TMEM41A**, **TMEM41B**, and **TMEM64** proteins clustered to the same family in our results. These are the
707 only proteins in human that show any similarity to the "SNARE_assoc" Pfam domain, as well as to the TCDB
708 family #9.B.27. While no protein with this domain or from this family has direct evidence for transport
709 activity, Pfam reports SCOOP-based similarity [31] of the "SNARE_assoc" domain with "Sm_multidrug_ex",
710 which is a domain encoding transporter proteins. Some members of the family in the TCDB have been
711 proposed to be "cation:proton importers" (#9.B.27.2.2) or "selenite transport proteins" (#9.B.27.2.3).

712 The most characterized member of the human protein family is TMEM41B. Interestingly, a recent study
713 reported a putative structure generated *ab initio* using evolutionary covariance-derived information [114].

714 Strikingly, this structural model shows features reminiscent of secondary transporters, such as a tandem
715 internal repeat with two-fold rotational symmetry, and the authors suggest a H^+ antiporter activity as a
716 mechanism of transport [114].

717 While the exact function of TMEM41B is still unclear, it forms a complex with vacuole membrane protein 1
718 (VMP1), also harboring the "SNARE_assoc" domain, and both are required for autophagosome formation
719 [115,116]. Tmem41b localized to mitochondria-associated ER membranes [117–119]. Interestingly,
720 TMEM41B seems to be an absolutely required factor for SARS-CoV-2 [120], and probably also flaviviral [121]
721 infection, possibly by facilitating a membrane curvature that is beneficial for viral replication [121].

722 The proteins **TMEM184A**, **TMEM184B**, **TMEM184C** clustered together with **SLC51A** (family of transporters of
723 steroid-derived molecules) in our analysis. While human TMEM184B and SLC51A are included in the TCDB as
724 members of family #2.A.82, TMEM184A and C are not. Independently, the "Solute_trans_a" Pfam model was
725 found in all four proteins with high scores and significance, but not in other human proteins. Therefore, it is
726 likely that the four proteins, TMEM184A-C and SLC51A, are homologous. In spite of this, sequence identity
727 between TMEM184 proteins and SLC51A is low (12.3%-13.6%), but moderate among TMEM184 proteins
728 (26.5%-62.0%). All four proteins are predicted to harbor 7 TM helices according to UniProt, yet our search
729 has found no similar proteins with known structures.

730 TMEM184A was identified as a heparin receptor in vascular cells [122], but no transport activity has been
731 reported. Interestingly, while SLC51A is known to function as a bile acid transporter [123–125], TMEM184B
732 has been proposed to be responsible for ibuprofen uptake [126]. This is interesting in view of the partial
733 chemical similarity between steroid acids and ibuprofen, both harboring a hydrophobic hydrocarbon part
734 and a carboxyl moiety. TMEM184C resides in a genetic locus that has been suggested to be responsible for
735 the pathogenicity of X-linked congenital hypertrichosis syndrome [127], but no transport activity has been
736 suggested.

737 ## Putative transporters

738 Our search also identified proteins whose transport activity is either controversial or not characterized, and
739 which do not show sequence similarity to transporters of known function. Thus, the proteins in these
740 families require further investigation to uncover their function.

741 The **CNNM1-4** proteins (also called ACDP1-4) are distant homologs of the cyclins, but have no documented
742 enzymatic activity. Instead, CNNM proteins belong to a highly conserved family of Mg^{2+} transport-related
743 proteins [128], and CNNM2 and CNNM4 have been proposed to be the long sought after basolateral
744 Na^{+}/Mg^{2+} exchangers in the kidney and intestine, respectively [129,130]. The function of these proteins is,
745 however, controversial [131–134], and there are hypotheses that CNNM proteins *per se* are not Mg^{2+}
746 transporters [135]. Most recently, however, the structure of a bacterial homolog, CorC, has been resolved,
747 revealing its membrane topology, as well as a conserved Mg^{2+}-binding site [136]. Strikingly, the Mg^{2+} ions in
748 the structure are fully dehydrated, in contrast to those in other known Mg^{2+} channel structures [136], which
749 makes it unlikely that the proteins function via a channel-like transport mechanism. In line with this, the
750 authors suggest an alternating-access exchange mechanism [136], however, further studies are required to
751 understand how and whether CNNM proteins might be able to mediate the translocation of Mg^{2+} ions across
752 the membrane.

753 A family of 4 lysosomal-associated transmembrane proteins (**LAPTM4A**, **LAPTM4B**, **LAPTM5** and sequence
754 **B4E0C1**) turned up in our search, corresponding to the TCDB subfamily #2.A.74.1 and Pfam model "Mtp"
755 (mouse transporter protein). The family also includes an uncharacterized transcript with the UniProt
756 accession "B4E0C1". Originally, the mouse transporter protein (Mtp, ortholog of LAPTM4A) was
757 characterized as a transporter mediating the transport of nucleosides and nucleobases between the
758 cytoplasm and intracellular compartments [137], and was later also associated with multidrug-resistance
759 (MDR) in yeast, where its expression changed the subcellular compartmentalization of a heterogenous group
760 of compounds [138,139]. LAPTM4A was shown to be involved in glycosylation and glycolipid regulation
761 [140,141]. All three proteins seem to be lysosomal [137,142,143]. Nevertheless, these proteins appear to
762 interact with other characterized transporters, such as SLC22A2 (hOCT2), SLC7A5/SLC3A2 (LAT1/4F2hc) and

763 MDR-related ABC transporters [144–146]. But it has been claimed that they are not *per se* transporters, but
764 rather regulatory factors, either assisting the localization and targeting or the function of other transporters
765 [146]. Interestingly, the transcript "B4E0C1" appears to have 4 TMHs at its N-terminus, which is identical to
766 human LAPTM5 apart from a ~40-amino acid insertion between TMH3 and TMH4. This region also shows
767 significant similarity to both the "Mtp" Pfam model and the TC# 2.A.74.1 subfamily. However, the C-terminal
768 region of the transcript is identical to the C-terminal segment of "actin filament-associated protein 1-like 1"
769 protein (UniProt accession Q8TED9). We did not find any similar fusion sequences in the other organisms we
770 analyzed. The "Mtp" Pfam domain, which is the hallmark of the family, belongs to the "Tetraspannin" Pfam
771 clan, which has no other domains with annotated transporter function and no similarity to existing
772 transporters. Structural information on "Tetraspannin" proteins is also not available. In summary, the
773 transport function of these proteins requires further investigation.

774 Our search identified four proteins in human (**LMBR1**, **LMBR1L**, **LMBD1/LMBRD1**, **LMBRD2**) bearing the
775 "LMBR1" Pfam domain, which clustered into two families in our analysis. These proteins correspond to TCDB
776 family #9.A.54. The LMBD1 protein, encoded by the LMBRD1 gene, was suggested to function as a vitamin
777 B12 (cobalamin) transporter, exporting vitamin B12 from the lysosomes into the cytoplasm [147]. However,
778 it was later shown that LMBD1 actually interacts with ABCD4 and assists in its lysosomal trafficking [148],
779 and that ABCD4 transports vitamin B12 even in the absence of LMBD1 [149]. Therefore, it is likely that
780 LMBD1 itself is not a vitamin B12 transporter. LMBD1 was originally coined as having "significant homology"
781 to lipocalin membrane receptors [147], and indeed the LIMR (lipocalin-1-interacting membrane receptor)
782 protein, encoded by the LMBR1L gene, is responsible for binding lipocalin 1 (LCN-1) with high affinity
783 [150,151]. LMBRD2 was proposed to be a regulator of β2-adrenoceptor signaling [152], while the first
784 protein identified in the family, LMBR1, was associated with polydactyly and limb malformations [153,154].
785 However, its physiological role is still elusive. The proteins seem to contain 9 TMHs in a 5+4 arrangement
786 according to UniProt predictions, but the tertiary structure of the proteins is still unknown, and no homologs
787 with a known structure were found in our search.

788 Two MagT1-like proteins (**MAGT1** and **TUSC3**), as well as **OSTC** and **OSTCL** (oligosaccharyltransferase
789 complex subunit) turned up in our search, showing similarity to TCDB family #1.A.76 members. MAGT1 and
790 TUSC3 also have high-scoring hits for the Pfam domain "OST3_OST6", which is characteristic of members of
791 the oligosaccharyltransferase (OST) complex. TUSC3 (also called N33) was first identified as a tumor
792 suppressor gene [155], and its presence, together with that of MAGT1 in the OST complex has been attested
793 later on [156–160]. Therefore, it was suggested that these proteins act as oxidoreductases [159]. Meanwhile,
794 MAGT1 and TUSC3 were also proposed to act as Mg^{2+} transporters [161,162]. On the other hand, recent
795 structural findings of human MAGT1 [160] indicated that this protein may not function as a transporter or
796 channel due to the lack of substrate-binding site or pore. OSTC (also called DC2) has similarly been shown to
797 be part of the OST complex and to have a structure similar to MAGT1 [160]. Whether MagT1-like proteins
798 still have a transport function remains to be clarified.

799 The **TMEM14A**, **TMEM14B** and **TMEM14C** proteins are the only ones in human containing the "Tmemb_14"
800 Pfam domain. While Pfam lists this domain as functionally uncharacterized, a plant protein (FAX1) containing
801 this domain was suggested to be involved in fatty acid export from chloroplasts [163]. However, the
802 physiological roles of TMEM14A and TMEM14B in human remain elusive. The third member of the family,
803 TMEM14C, was identified as a putative mitochondrial protein whose transcript is consistently coexpressed
804 with proteins from the core machinery of heme biosynthesis [164]. It was later shown that TMEM14C
805 mediates the import of protoporphyrinogen IX (PPgenIX) into the mitochondrial matrix [165,166]. While the
806 structure of TMEM14C was solved using nuclear magnetic resonance (NMR) [167], showing a bundle of three
807 TM helices and an amphipathic helix, the transport mechanism remains elusive. Interestingly, despite their
808 proposed function in the mitochondria, TargetP-2.0 [168] did not predict a mitochondrial targeting sequence
809 in the amino acid sequence of any of the human TMEM14 proteins in our hands.

810 **TMEM205** is a 4-TMH protein according to UniProt annotations, which was linked to cisplatin resistance
811 [169]. The protein is expressed mostly in liver, pancreas and adrenal glands, and is present on the plasma
812 membrane [169]. TMEM205-mediated resistance was shown to be selective towards platinum-based drugs,
813 such as cisplatin and oxaliplatin, but not carboplatin [170]. While structural information about the protein is

814 not available, mutagenesis studies of TMEM205 showed that mutating sulfur-containing residues, especially
815 in TMH2 and TMH4, diminishes the effect of cisplatin resistance [170]. Nevertheless, neither the biological
816 function nor the physiological substrates of TMEM205 are known.

817 Transporters with hydrophobic substrates

818 In addition to proteins that transport solutes or are similar to transporters that typically translocate water-
819 soluble small molecule compounds, our search has uncovered numerous proteins that have been reported
820 to take part in modulating the intracellular distribution of hydrophobic or amphipathic compounds, such as
821 cholesterol, fat-soluble vitamins, lipids and fatty acids (Table 3). Although these proteins do not, strictly
822 speaking, transport so-called "solutes", they translocate small hydrophobic molecules that have
823 fundamental biological functions. Thus one could argue that they belong to the SLC superfamily as well.
824 Accordingly, they have been integrated in our search and some of them have already been included in the
825 SLC nomenclature (SLC59, SLC63, SLC65). In view of the biological and pharmaceutical importance of the
826 transport mechanisms of hydrophobic substrates, our results on this topic will be discussed in a separate
827 paper.

828 Conclusions

829 Our study represents the first systematic correlation of the SLC and TCDB nomenclature schemes. Many of
830 the transport proteins discovered in our search are underexplored and there is limited information about
831 them, although they often have important physiological roles and/or potentially represent new therapeutic
832 targets. Even with proteins that have been studied for their physiological involvement, it was often not taken
833 into account that they could have a transport function. Numerous proteins uncovered in our search have
834 similarity to proteins with transport function in other organisms, but their physiological substrates remain
835 unknown. These proteins will be interesting targets for deorphanization studies in order to reveal their
836 natural substrates. In our work, we also highlight proteins for which transport activity has been
837 controversial, and more specific analyses are required to clarify their biological function. In addition, our
838 search reveals new SLC-like proteins that have no structural information. This hinders a deeper

839 understanding of their transport mechanism. Future structure determinations would be of crucial
840 importance to accelerate validation of the identified proteins. The combination of all these efforts would
841 greatly facilitate the completion of the SLC-ome in human cells. Thus, our study points out important
842 directions in which future studies could help resolve the lack of information about SLC transporters, which
843 will help unlock their therapeutic potential.

844 Acknowledgements

845 We acknowledge support by the Swiss National Science Foundation for the grants # CRSII5_180326 ("The
846 role of mitochondrial carriers in metabolic tuning and reprogramming by calcium flow across membrane
847 contact sites") and # 310030_182272 ("Intestinal absorption of transition metals in human health and
848 disease") as well as by the Swiss National Research Programme NRP 78, # 4078P0_198281 ("New insights
849 into the COVID-19 pandemic"). Calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the
850 HPC cluster at the University of Bern.

851 Methods

852 HMM building for TCDB families and subfamilies

853 Sequences of selected TCDB families and subfamilies were aligned using PSI-Coffee 11.00 [171] and NCBI
854 BLAST+ 2.6.0 [172] using the "nr" BLAST database of 2018-02-12. PSI-Coffee was run with BLAST mode
855 "LOCAL" and otherwise with default options. Altogether, 4221 different sequences from the TCDB were
856 present in the subfamily and family alignments, and the profiles used for alignments contained 1-1070
857 sequences. In total, 130 sequences from the TCDB returned zero hits from the "nr" database in the iterative
858 BLAST searches, meaning their profiles just contained the query sequence itself. PSI-Coffee uses the profiles
859 to guide the generation of multiple alignments of the query sequences within each subfamily and family. The
860 alignments were subsequently turned into HMMs using the "hmmbuild" command of HMMER 3.1b2 [30]
861 using default settings.

862 ## Sequence similarity search

863 All protein belonging to each of the 7 organisms studied were downloaded from UniProt on 2019-07-31 into
864 a FASTA file. Sequence similarity search was performed using the HMMs downloaded for selected Pfam
865 families on 2019-02-14 and those generated for selected TCDB families and subfamilies using "hmmsearch"
866 from HMMER 3.1b2. Hits with bit scores larger than 50 were used for further analysis. These hits presented a
867 maximal hit e-value of 7.9e-12, while hits with bit score larger than 25, used for HMM fingerprint-based
868 clustering (see below), had hit e-values below 6e-4.

869 ## Sequence clustering for fragment elimination

870 Since the downloaded protein set from UniProt contained fragments as well as predicted open reading
871 frames and sequences from genomic screening methods, we strived to retain one sequence per gene for
872 further analysis. In order to achieve this, hits yielded by the sequence similarity search were clustered using
873 the following method. First, all-against-all similarity searches were performed using NCBI BLAST+ 2.4.0 [172].
874 Sequences were assigned the same cluster if they either share common gene annotations according to their
875 UniProt records, or a high-scoring segment pair (HSP) with more than 95% sequence identity based on the
876 all-against-all BLAST search. For gene annotation, the fields Gene Symbol (GN), HGNC symbol, GeneID,
877 UniGene, FlyBase, KEGG identifiers were used from UniProt records. Ambiguous or conflicting annotations,
878 as well as annotations conflicting with BLAST-reported sequence similarity were detected and resolved
879 manually. Afterwards, clusters were reduced to representative sequences. For clusters containing a single
880 Swiss-Prot sequence, that sequence was taken as representative. For cluster with no Swiss-Prot sequence,
881 the longest sequence of the cluster was taken as representative. Clusters with more than one Swiss-Prot
882 sequence were manually analyzed and split if necessary.

883 ## HMM fingerprint-based sequence clustering

884 We introduce the concept of a "HMM fingerprint", which is a mathematical vector of numbers assigned to a
885 protein sequence, corresponding to the bit scores of similarity to each of the set of HMMs used in our
886 analysis, consisting of the HMMs of TCDB families and subfamilies, as well as Pfam HMMs. We have
887 restricted the number of HMMs to those that gave a hit with bit score > 25, in total 513 HMMs. Two proteins

888 sequences that are related are expected to show similarity to a similar subset of TCDB families, subfamilies,
889 or Pfam families, and therefore a similar pattern in their HMM fingerprints. In turn, if two protein sequences
890 show similarity to the same subset of TCDB families, subfamilies or Pfam families, as indicated by a similar
891 HMM fingerprint, then they can be expected to be related. Once the HMM fingerprint has been assigned to
892 each protein sequence found in our search, the unweighted pair group method with arithmetic mean
893 (UPGMA) method [173] was used using the cosine metric to arrive at a hierarchical clustering of the
894 sequences. The tree representing the clustering was cut at 0.7 cosine distance to arrive at branches that
895 formed the basis of protein families. Join and split constraints have been introduced to keep certain proteins
896 artificially grouped or separated. If a join constraint was acting between a pair of proteins and they were not
897 grouped into the same cluster by the default tree cut threshold, then the clustering tree was cut just above
898 the branch that leads to the smallest cluster containing both proteins. Similarly, split constraints between
899 two proteins caused the clustering tree to be cut just below the branch containing both proteins, unless the
900 two proteins were already in different clusters. Join constraints were introduced between the following
901 protein pairs: SLC5A1-SLC5A7, SLC9A1-SLC9B1, SLC10A1-SLC10A7, SLC25A1-SLC25A46, SLC30A1-SLC30A9,
902 SLC35C1-SLC35G1, SLC39A1-SLC39A9, and SLC66A1-SLC66A5. Split constraints were introduced between the
903 following protein pairs: SLC2A1-SLC22A1, SLC17A1-SLC18A1, SLC17A1-SLC37A1, SLC32A1-SLC36A1, and
904 SLC32A1-SLC38A1. The clustering tree with the resulting families was shown in Fig 1 in a polar coordinate
905 system, with the radial component ($d \in [0; 1]$) transformed according to $d' = 0.5 + \frac{\arcsin(2d-1)}{\pi}$ to magnify
906 tree details around $d = 0$ and $d = 1$ for better visual representation.

907 Structural homolog search

908 Sequences of SLC-like proteins were turned into HMMs using “hhblits” from the HH-suite3 package [174],
909 using the UniRef30 database of 2020-06 [35,175], 3 iterations, an E-value threshold of 1e-3 for inclusion, and
910 a probability threshold of 0.35 for MAC re-alignment. The HMMs were searched against the “pdb70”
911 database of 2021-08-04 [35] with no MAC realignment, “predicted vs predicted” secondary structure
912 scoring, and amino acid score of 1. The resulting hits were checked against PDB annotations of

913 transmembrane helices, and were accepted if at least 3 TM segments were contained within the aligned
914 region and the E-value of the hit was less than 1e-4.

915 Phylogenetic trees

916 Selected groups of SLC-like proteins were aligned using Clustal Omega 1.2.1 [176,177] with 5 iterations and
917 default settings. Smart Model Selection 1.8.4 [178] and PhyML 3.3.20190909 [179] were used to generate
918 the phylogenetic trees, with 10 random starting trees and using the approximate likelihood ratio test aLRT
919 method [180]. Trees in main figures were visualized using TreeViewer 1.2.2 (https://treeviewer.org/). Tree
920 rooting, rearrangement (with threshold 0.9) and reconciliation with the species tree was done using
921 NOTUNG 2.9.1.5 [181]. Reconciled phylogenetic trees were visualized using custom Python scripts in the
922 style used by NOTUNG. Orthologs were identified using the reconciled phylogenetic trees and custom
923 Python scripts.

924 Data Availability

925 Scripts and data generated during the preparation of the manuscript are available at
926 https://github.com/bioparadigms/SLCAtlas. In addition, the datasets resulting from this work will also be
927 made available in a browsable format at https://www.bioparadigms.org/.

928 Supplementary information

929 **S1 Table. "SLC-like" Pfam models and their clan memberships.** The dataset is a result of our manual
930 curation based on our "SLC-like" criteria (see text) and Pfam families present in protein sequences listed in
931 the TCDB. Family names and data are from the Pfam database.

932 **S2 Table. False positive hits or incorrectly annotated human sequences found in our search.** Gene symbol
933 and protein name shown based on UniProt data. Most similar (highest scoring) TCDB families/subfamilies
934 and Pfam families are shown, and reason for being false positive are indicated.

935 **S3 Table. Table of existing (classic) SLC transporters, indicating the correlation with TCDB**
936 **families/subfamilies, Pfam families, PDB structures and structural folds.** Gene symbol and protein name
937 based on UniProt information are shown. Most similar (highest-scoring) TCDB families and subfamilies, as
938 well as Pfam families are shown for each protein. The PDB structure from the pdb70 dataset that is most

939 similar (highest-scoring) to each protein is shown, along with its fold family assignment by us, partly based
940 on TCDB family names.

941 AbgT: p-Aminobenzoyl-glutamate Transporter family; Amt: Ammonium Transporter family; APC: Amino acid-
942 Polyamine-Cation family; CDF: Cation Diffusion Facilitator family; CNT: Concentrative Nucleoside Transporter
943 family; Ctr: Copper Transporter family; DAACS: Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter
944 family; MATE: Multidrug And Toxic compound Extrusion family; MCF: Mitochondrial Carrier Family; MFS:
945 Major Facilitator Superfamily; MgtE: Mg^{2+} transporter-E family; NAT: Nucleobase/Ascorbate Transporter or
946 Nucleobase:Cation Symporter-2 (NCS2) family; NCX: Sodium/Calcium exchanger family; NhaA:
947 Sodium/proton antiporter family; NST: Nucleoside-Sugar Transporter family; PiT: Type III Sodium/phosphate
948 cotransporter family; SWEET: Sugar Will Eventually be Transported family.; ZIP: Zrt/Irt-like Transporter
949 family.

950 **S4 Table. Non-classic "SLC-like" proteins of Table 3 found in our search, extended with additional**
951 **information.** Protein sequence identifiers, most similar (highest-scoring) PDB structure from the pdb70
952 dataset, chemical identifiers, PubMed links to references, and comments have been included. The table also
953 contains information about whether individual SLC-likeness criteria (numbered 1-8) have been satisfied or
954 violated.

955 **S1 File. Phylogenetic trees of SLC-like protein families with more than two members.** Trees were generated
956 using multiple alignment by ClustalO, maximum likelihood tree generation by PhyML, followed by tree
957 reconciliation with the species tree using NOTUNG (see Methods). The species tree with internal names of
958 putative ancestor taxa is shown on each page on the upper-right hand corner. The trees are shown as
959 dendrograms and branch lengths are not indicative of evolutionary distance. Each tree leaf corresponds to
960 an SLC-like protein sequence denoting a gene, labels show the gene symbol, UniProt accession and taxon
961 name. Leaves with labels ending with "*LOST" denote putative genes lost in the indicated ancestral species.
962 Red "D" denote gene duplication nodes, normal nodes correspond to speciation nodes. Light green numbers
963 denote branch support values as calculated by NOTUNG.

964 # References

965 1. Hediger MA, Romero MF, Peng J-B, Rolfs A, Takanaga H, Bruford EA. The ABCs of solute
966 carriers: physiological, pathological and therapeutic implications of human membrane transport
967 proteinsIntroduction. Pflugers Arch. 2004;447: 465–468. doi:10.1007/s00424-003-1192-y
968 2. Hediger MA, Clémençon B, Burrier RE, Bruford EA. The ABCs of membrane transporters in
969 health and disease (SLC series): introduction. Mol Aspects Med. 2013;34: 95–107.
970 doi:10.1016/j.mam.2012.12.009
971 3. César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, et al. A Call for
972 Systematic Research on Solute Carriers. Cell. 2015;162: 478–487. doi:10.1016/j.cell.2015.07.022
973 4. Bai X, Moraes TF, Reithmeier RAF. Structural biology of solute carrier (SLC) membrane
974 transport proteins. Mol Membr Biol. 2018; 1–32. doi:10.1080/09687688.2018.1448123
975 5. Choi S, Jeon J, Yang J-S, Kim S. Common occurrence of internal repeat symmetry in
976 membrane proteins. Proteins. 2008;71: 68–80. doi:10.1002/prot.21656
977 6. Forrest LR. Structural Symmetry in Membrane Proteins. Annu Rev Biophys. 2015;44: 311–
978 337. doi:10.1146/annurev-biophys-051013-023008
979 7. Jardetzky O. Simple allosteric model for membrane pumps. Nature. 1966;211: 969–970.
980 doi:10.1038/211969a0
981 8. Forrest LR, Krämer R, Ziegler C. The structural basis of secondary active transport
982 mechanisms. Biochim Biophys Acta. 2011;1807: 167–188. doi:10.1016/j.bbabio.2010.10.014
983 9. Fredriksson R, Nordström KJV, Stephansson O, Hägglund MGA, Schiöth HB. The solute
984 carrier (SLC) complement of the human genome: phylogenetic classification reveals four major
985 families. FEBS Lett. 2008;582: 3811–3816. doi:10.1016/j.febslet.2008.10.016
986 10. Perland E, Fredriksson R. Classification Systems of Secondary Active Transporters. Trends
987 Pharmacol Sci. 2017;38: 305–315. doi:10.1016/j.tips.2016.11.008
988 11. Perland E, Bagchi S, Klaesson A, Fredriksson R. Characteristics of 29 novel atypical solute
989 carriers of major facilitator superfamily type: evolutionary conservation, predicted structure and
990 neuronal co-expression. Open Biol. 2017;7. doi:10.1098/rsob.170142
991 12. Schlessinger A, Matsson P, Shima JE, Pieper U, Yee SW, Kelly L, et al. Comparison of human
992 solute carriers. Protein Science. 2010;19: 412–428. doi:10.1002/pro.320
993 13. Schlessinger A, Yee SW, Sali A, Giacomini KM. SLC classification: an update. Clin Pharmacol
994 Ther. 2013;94: 19–23. doi:10.1038/clpt.2013.73
995 14. Saier MH. Molecular phylogeny as a basis for the classification of transport proteins from
996 bacteria, archaea and eukarya. Adv Microb Physiol. 1998;40: 81–136.
997 15. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The
998 Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res. 2021;49: D461–D467.
999 doi:10.1093/nar/gkaa1004
1000 16. Cao Y, Jin X, Huang H, Derebe MG, Levin EJ, Kabaleeswaran V, et al. Crystal structure of a
1001 potassium ion transporter, TrkH. Nature. 2011;471: 336–340. doi:10.1038/nature09731
1002 17. Bosson R, Jaquenoud M, Conzelmann A. GUP1 of Saccharomyces cerevisiae encodes an O-
1003 acyltransferase involved in remodeling of the GPI anchor. Mol Biol Cell. 2006;17: 2636–2645.
1004 doi:10.1091/mbc.e06-02-0104
1005 18. Berks BC, Sargent F, Palmer T. The Tat protein export pathway. Mol Microbiol. 2000;35:
1006 260–274. doi:10.1046/j.1365-2958.2000.01719.x
1007 19. Berks BC, Sargent F, De Leeuw E, Hinsley AP, Stanley NR, Jack RL, et al. A novel protein
1008 transport system involved in the biogenesis of bacterial electron transfer chains. Biochim Biophys
1009 Acta. 2000;1459: 325–330. doi:10.1016/s0005-2728(00)00168-7

1010 20. Ren Q, Kang KH, Paulsen IT. TransportDB: a relational database of cellular membrane
1011 transport systems. Nucleic Acids Res. 2004;32: D284-288. doi:10.1093/nar/gkh016
1012 21. Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic
1013 membrane transport systems and outer membrane channels. Nucleic Acids Res. 2007;35: D274-
1014 279. doi:10.1093/nar/gkl925
1015 22. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring
1016 membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res. 2017;45:
1017 D320–D324. doi:10.1093/nar/gkw1068
1018 23. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG
1019 database: new developments in phylogenetic classification of proteins from complete genomes.
1020 Nucleic Acids Res. 2001;29: 22–28. doi:10.1093/nar/29.1.22
1021 24. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and Genome
1022 Properties in 2013. Nucleic Acids Res. 2013;41: D387-395. doi:10.1093/nar/gks1234
1023 25. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein
1024 families database: towards a more sustainable future. Nucleic Acids Res. 2016;44: D279-285.
1025 doi:10.1093/nar/gkv1344
1026 26. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein
1027 topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:
1028 567–580. doi:10.1006/jmbi.2000.4315
1029 27. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res.
1030 2019;47: D506–D515. doi:10.1093/nar/gky1049
1031 28. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein
1032 families database in 2019. Nucleic Acids Res. 2019;47: D427–D432. doi:10.1093/nar/gky995
1033 29. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. Pfam:
1034 clans, web tools and services. Nucleic Acids Res. 2006;34: D247-251. doi:10.1093/nar/gkj149
1035 30. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195.
1036 doi:10.1371/journal.pcbi.1002195
1037 31. Bateman A, Finn RD. SCOOP: a simple method for identification of novel protein superfamily
1038 relationships. Bioinformatics. 2007;23: 809–814. doi:10.1093/bioinformatics/btm034
1039 32. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. Methods
1040 Mol Biol. 2007;406: 89–112. doi:10.1007/978-1-59745-535-0_4
1041 33. Quednau BD, Nicoll DA, Philipson KD. The sodium/calcium exchanger family-SLC8. Pflugers
1042 Arch. 2004;447: 543–548. doi:10.1007/s00424-003-1065-4
1043 34. Chen H, Liu Y, Li X. Structure of human Dispatched-1 provides insights into Hedgehog ligand
1044 biogenesis. Life Sci Alliance. 2020;3: e202000776. doi:10.26508/lsa.202000776
1045 35. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence
1046 searching by HMM-HMM alignment. Nat Methods. 2011;9: 173–175. doi:10.1038/nmeth.1818
1047 36. Söding J, Remmert M. Protein sequence comparison and fold recognition: progress and
1048 good-practice benchmarking. Curr Opin Struct Biol. 2011;21: 404–411.
1049 doi:10.1016/j.sbi.2011.03.005
1050 37. Jeong J, Eide DJ. The SLC39 family of zinc transporters. Mol Aspects Med. 2013;34: 612–619.
1051 doi:10.1016/j.mam.2012.05.011
1052 38. Donowitz M, Ming Tse C, Fuster D. SLC9/NHE gene family, a plasma membrane and
1053 organellar family of Na^+/H^+ exchangers. Mol Aspects Med. 2013;34: 236–251.
1054 doi:10.1016/j.mam.2012.05.001

1055 39. Pramod AB, Foster J, Carvelli L, Henry LK. SLC6 transporters: structure, function, regulation,
1056 disease association and therapeutics. Mol Aspects Med. 2013;34: 197–219.
1057 doi:10.1016/j.mam.2012.07.002
1058 40. Song Z. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. Mol
1059 Aspects Med. 2013;34: 590–600. doi:10.1016/j.mam.2012.12.004
1060 41. Chu X, Bleasby K, Evers R. Species differences in drug transporters and implications for
1061 translating preclinical findings to humans. Expert Opin Drug Metab Toxicol. 2013;9: 237–252.
1062 doi:10.1517/17425255.2013.741589
1063 42. Wright EM. Glucose transport families SLC5 and SLC50. Mol Aspects Med. 2013;34: 183–
1064 196. doi:10.1016/j.mam.2012.11.002
1065 43. Gyimesi G, Pujol-Giménez J, Kanai Y, Hediger MA. Sodium-coupled glucose transport, the
1066 SLC5 family, and therapeutically relevant inhibitors: from molecular discovery to clinical application.
1067 Pflugers Arch. 2020;472: 1177–1206. doi:10.1007/s00424-020-02433-x
1068 44. Haga T. Molecular properties of the high-affinity choline transporter CHT1. J Biochem.
1069 2014;156: 181–194. doi:10.1093/jb/mvu047
1070 45. Claro da Silva T, Polli JE, Swaan PW. The solute carrier family 10 (SLC10): beyond bile acid
1071 transport. Mol Aspects Med. 2013;34: 252–269. doi:10.1016/j.mam.2012.07.004
1072 46. Godoy JR, Fernandes C, Döring B, Beuerlein K, Petzinger E, Geyer J. Molecular and
1073 phylogenetic characterization of a novel putative membrane transporter (SLC10A7), conserved in
1074 vertebrates and bacteria. Eur J Cell Biol. 2007;86: 445–460. doi:10.1016/j.ejcb.2007.06.001
1075 47. Karakus E, Wannowius M, Müller SF, Leiting S, Leidolf R, Noppes S, et al. The orphan solute
1076 carrier SLC10A7 is a novel negative regulator of intracellular calcium signaling. Sci Rep. 2020;10:
1077 7248. doi:10.1038/s41598-020-64006-3
1078 48. Janer A, Prudent J, Paupe V, Fahiminiya S, Majewski J, Sgarioto N, et al. SLC25A46 is required
1079 for mitochondrial lipid homeostasis and cristae maintenance and is responsible for Leigh syndrome.
1080 EMBO Mol Med. 2016;8: 1019–1038. doi:10.15252/emmm.201506159
1081 49. Sesaki H, Jensen RE. UGO1 encodes an outer membrane protein required for mitochondrial
1082 fusion. J Cell Biol. 2001;152: 1123–1134. doi:10.1083/jcb.152.6.1123
1083 50. Lamarca V, Sanz-Clemente A, Pérez-Pé R, Martínez-Lorenzo MJ, Halaihel N, Muniesa P, et al.
1084 Two isoforms of PSAP/MTCH1 share two proapoptotic domains and multiple internal signals for
1085 import into the mitochondrial outer membrane. Am J Physiol, Cell Physiol. 2007;293: C1347-1361.
1086 doi:10.1152/ajpcell.00431.2006
1087 51. Labbé K, Mookerjee S, Le Vasseur M, Gibbs E, Lerner C, Nunnari J. The modified
1088 mitochondrial outer membrane carrier MTCH2 links mitochondrial fusion to lipogenesis. J Cell Biol.
1089 2021;220: e202103122. doi:10.1083/jcb.202103122
1090 52. Ruprecht JJ, Kunji ERS. The SLC25 Mitochondrial Carrier Family: Structure and Mechanism.
1091 Trends Biochem Sci. 2020;45: 244–258. doi:10.1016/j.tibs.2019.11.001
1092 53. Chen Y-H, Kim JH, Stallcup MR. GAC63, a GRIP1-dependent nuclear receptor coactivator.
1093 Mol Cell Biol. 2005;25: 5965–5972. doi:10.1128/MCB.25.14.5965-5972.2005
1094 54. Thomas P, Pang Y, Dong J, Berg AH. Identification and characterization of membrane
1095 androgen receptors in the ZIP9 zinc transporter subfamily: II. Role of human ZIP9 in testosterone-
1096 induced prostate and breast cancer cell apoptosis. Endocrinology. 2014;155: 4250–4265.
1097 doi:10.1210/en.2014-1201
1098 55. Brett CL, Donowitz M, Rao R. Evolutionary origins of eukaryotic sodium/proton exchangers.
1099 Am J Physiol Cell Physiol. 2005;288: C223-239. doi:10.1152/ajpcell.00360.2004
1100 56. Gyimesi G, Hediger MA. Sequence Features of Mitochondrial Transporter Protein Families.
1101 Biomolecules. 2020;10. doi:10.3390/biom10121611

1102 57. Horiba N, Masuda S, Takeuchi A, Takeuchi D, Okuda M, Inui K. Cloning and characterization
1103 of a novel Na+-dependent glucose transporter (NaGLT1) in rat kidney. J Biol Chem. 2003;278:
1104 14669–14676. doi:10.1074/jbc.M212240200
1105 58. Nawata CM, Dantzler WH, Pannabecker TL. Alternative channels for urea in the inner
1106 medulla of the rat kidney. Am J Physiol Renal Physiol. 2015;309: F916-924.
1107 doi:10.1152/ajprenal.00392.2015
1108 59. Tejada-Jiménez M, Galván A, Fernández E. Algae and humans share a molybdate
1109 transporter. Proc Natl Acad Sci USA. 2011;108: 6420–6425. doi:10.1073/pnas.1100700108
1110 60. Zhu W, Spiga L, Winter S. Transition metals and host-microbe interactions in the inflamed
1111 intestine. Biometals. 2019;32: 369–384. doi:10.1007/s10534-019-00182-8
1112 61. Barth J, Zimmermann H, Volknandt W. SV31 is a Zn2+-binding synaptic vesicle protein. J
1113 Neurochem. 2011;118: 558–570. doi:10.1111/j.1471-4159.2011.07344.x
1114 62. Waberer L, Henrich E, Peetz O, Morgner N, Dötsch V, Bernhard F, et al. The synaptic vesicle
1115 protein SV31 assembles into a dimer and transports Zn2. J Neurochem. 2017;140: 280–293.
1116 doi:10.1111/jnc.13886
1117 63. Cuajungco MP, Kiselyov K. The mucolipin-1 (TRPML1) ion channel, transmembrane-163
1118 (TMEM163) protein, and lysosomal zinc handling. Front Biosci (Landmark Ed). 2017;22: 1330–1343.
1119 64. Sanchez VB, Ali S, Escobar A, Cuajungco MP. Transmembrane 163 (TMEM163) protein
1120 effluxes zinc. Arch Biochem Biophys. 2019;677: 108166. doi:10.1016/j.abb.2019.108166
1121 65. Burré J, Zimmermann H, Volknandt W. Identification and characterization of SV31, a novel
1122 synaptic vesicle membrane protein and potential transporter. J Neurochem. 2007;103: 276–287.
1123 doi:10.1111/j.1471-4159.2007.04758.x
1124 66. Wang L, Li N-N, Lu Z-J, Li J-Y, Peng J-X, Duan L-R, et al. Association of three candidate genetic
1125 variants in ACMSD/TMEM163, GPNMB and BCKDK /STX1B with sporadic Parkinson's disease in Han
1126 Chinese. Neurosci Lett. 2019;703: 45–48. doi:10.1016/j.neulet.2019.03.019
1127 67. Chang K-H, Chen C-M, Chen Y-C, Fung H-C, Wu Y-R. Polymorphisms of ACMSD-TMEM163,
1128 MCCC1, and BCKDK-STX1B Are Not Associated with Parkinson's Disease in Taiwan. Parkinsons Dis.
1129 2019;2019: 3489638. doi:10.1155/2019/3489638
1130 68. Lauterbach EC. Psychotropic drug effects on gene transcriptomics relevant to Parkinson's
1131 disease. Prog Neuropsychopharmacol Biol Psychiatry. 2012;38: 107–115.
1132 doi:10.1016/j.pnpbp.2012.03.011
1133 69. Chakraborty S, Vellarikkal SK, Sivasubbu S, Roy SS, Tandon N, Bharadwaj D. Role of
1134 Tmem163 in zinc-regulated insulin storage of MIN6 cells: Functional exploration of an Indian type 2
1135 diabetes GWAS associated gene. Biochem Biophys Res Commun. 2019.
1136 doi:10.1016/j.bbrc.2019.11.117
1137 70. Tabassum R, Chauhan G, Dwivedi OP, Mahajan A, Jaiswal A, Kaur I, et al. Genome-wide
1138 association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21.
1139 Diabetes. 2013;62: 977–986. doi:10.2337/db12-0406
1140 71. Bai H, Liu H, Suyalatu S, Guo X, Chu S, Chen Y, et al. Association Analysis of Genetic Variants
1141 with Type 2 Diabetes in a Mongolian Population in China. J Diabetes Res. 2015;2015: 613236.
1142 doi:10.1155/2015/613236
1143 72. Foulquier F, Amyere M, Jaeken J, Zeevaert R, Schollen E, Race V, et al. TMEM165 deficiency
1144 causes a congenital disorder of glycosylation. Am J Hum Genet. 2012;91: 15–26.
1145 doi:10.1016/j.ajhg.2012.05.002
1146 73. Demaegd D, Foulquier F, Colinet A-S, Gremillon L, Legrand D, Mariot P, et al. Newly
1147 characterized Golgi-localized family of proteins is involved in calcium and pH homeostasis in yeast
1148 and human cells. Proc Natl Acad Sci USA. 2013;110: 6859–6864. doi:10.1073/pnas.1219871110

1149 74. Foulquier F, Legrand D. Biometals and glycosylation in humans: Congenital disorders of
1150 glycosylation shed lights into the crucial role of Golgi manganese homeostasis. Biochim Biophys
1151 Acta Gen Subj. 2020;1864: 129674. doi:10.1016/j.bbagen.2020.129674
1152 75. Stribny J, Thines L, Deschamps A, Goffin P, Morsomme P. The human Golgi protein
1153 TMEM165 transports calcium and manganese in yeast and bacterial cells. J Biol Chem. 2020;295:
1154 3865–3874. doi:10.1074/jbc.RA119.012249
1155 76. Snyder NA, Palmer MV, Reinhardt TA, Cunningham KW. Milk biosynthesis requires the Golgi
1156 cation exchanger TMEM165. J Biol Chem. 2019;294: 3181–3191. doi:10.1074/jbc.RA118.006270
1157 77. Reinhardt TA, Lippolis JD, Sacco RE. The Ca(2+)/H(+) antiporter TMEM165 expression,
1158 localization in the developing, lactating and involuting mammary gland parallels the secretory
1159 pathway Ca(2+) ATPase (SPCA1). Biochem Biophys Res Commun. 2014;445: 417–421.
1160 doi:10.1016/j.bbrc.2014.02.020
1161 78. Lebredonchel E, Houdou M, Potelle S, de Bettignies G, Schulz C, Krzewinski Recchi M-A, et al.
1162 Dissection of TMEM165 function in Golgi glycosylation and its Mn2+ sensitivity. Biochimie.
1163 2019;165: 123–130. doi:10.1016/j.biochi.2019.07.016
1164 79. Perland E, Lekholm E, Eriksson MM, Bagchi S, Arapi V, Fredriksson R. The Putative SLC
1165 Transporters Mfsd5 and Mfsd11 Are Abundantly Expressed in the Mouse Brain and Have a Potential
1166 Role in Energy Homeostasis. PLoS ONE. 2016;11: e0156912. doi:10.1371/journal.pone.0156912
1167 80. Lekholm E, Perland E, Eriksson MM, Hellsten SV, Lindberg FA, Rostami J, et al. Putative
1168 Membrane-Bound Transporters MFSD14A and MFSD14B Are Neuronal and Affected by Nutrient
1169 Availability. Front Mol Neurosci. 2017;10: 11. doi:10.3389/fnmol.2017.00011
1170 81. Perland E, Hellsten SV, Lekholm E, Eriksson MM, Arapi V, Fredriksson R. The Novel
1171 Membrane-Bound Proteins MFSD1 and MFSD3 are Putative SLC Transporters Affected by Altered
1172 Nutrient Intake. J Mol Neurosci. 2017;61: 199–214. doi:10.1007/s12031-016-0867-8
1173 82. Bagchi S, Perland E, Hosseini K, Lundgren J, Al-Walai N, Kheder S, et al. Probable role for
1174 major facilitator superfamily domain containing 6 (MFSD6) in the brain during variable energy
1175 consumption. Int J Neurosci. 2020;130: 476–489. doi:10.1080/00207454.2019.1694020
1176 83. Valoskova K, Biebl J, Roblek M, Emtenani S, Gyoergy A, Misova M, et al. A conserved major
1177 facilitator superfamily member orchestrates a subset of O-glycosylation to aid macrophage tissue
1178 invasion. Elife. 2019;8. doi:10.7554/eLife.41801
1179 84. Landis GN, Bhole D, Tower J. A search for doxycycline-dependent mutations that increase
1180 Drosophila melanogaster life span identifies the VhaSFD, Sugar baby, filamin, fwd and Cctl genes.
1181 Genome Biol. 2003;4: R8. doi:10.1186/gb-2003-4-2-r8
1182 85. Kim KW, Tang NH, Piggott CA, Andrusiak MG, Park S, Zhu M, et al. Expanded genetic
1183 screening in Caenorhabditis elegans identifies new regulators and an inhibitory role for NAD+ in
1184 axon regeneration. Elife. 2018;7: e39756. doi:10.7554/eLife.39756
1185 86. Ceder MM, Aggarwal T, Hosseini K, Maturi V, Patil S, Perland E, et al. CG4928 Is Vital for
1186 Renal Function in Fruit Flies and Membrane Potential in Cells: A First In-Depth Characterization of
1187 the Putative Solute Carrier UNC93A. Front Cell Dev Biol. 2020;8: 580291.
1188 doi:10.3389/fcell.2020.580291
1189 87. Hannes F, Hammond P, Quarrell O, Fryns J-P, Devriendt K, Vermeesch JR. A microdeletion
1190 proximal of the critical deletion region is associated with mild Wolf-Hirschhorn syndrome. Am J
1191 Med Genet A. 2012;158A: 996–1004. doi:10.1002/ajmg.a.35299
1192 88. Damme M, Brandenstein L, Fehr S, Jankowiak W, Bartsch U, Schweizer M, et al. Gene
1193 disruption of Mfsd8 in mice provides the first animal model for CLN7 disease. Neurobiol Dis.
1194 2014;65: 12–24. doi:10.1016/j.nbd.2014.01.003

1195 89. Siintola E, Topcu M, Aula N, Lohi H, Minassian BA, Paterson AD, et al. The novel neuronal
1196 ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter. Am J Hum Genet.
1197 2007;81: 136–146. doi:10.1086/518902
1198 90. von Kleist L, Ariunbat K, Braren I, Stauber T, Storch S, Danyukova T. A newly generated
1199 neuronal cell model of CLN7 disease reveals aberrant lysosome motility and impaired cell survival.
1200 Mol Genet Metab. 2019;126: 196–205. doi:10.1016/j.ymgme.2018.09.009
1201 91. Li Y, Yang X, Yang J, Wang H, Wei W. An 11-gene-based prognostic signature for uveal
1202 melanoma metastasis based on gene expression and DNA methylation profile. J Cell Biochem. 2018.
1203 doi:10.1002/jcb.28151
1204 92. Nicoletti CF, Pinhel MS, Noronha NY, Jácome A, Crujeiras AB, Nonino CB. Association of
1205 MFSD3 promoter methylation level and weight regain after gastric bypass: Assessment for 3 y after
1206 surgery. Nutrition. 2020;70: 110499. doi:10.1016/j.nut.2019.04.010
1207 93. Rebsamen M, Pochini L, Stasyk T, de Araújo MEG, Galluccio M, Kandasamy RK, et al.
1208 SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1.
1209 Nature. 2015;519: 477–481. doi:10.1038/nature14107
1210 94. Västermark Å, Almén MS, Simmen MW, Fredriksson R, Schiöth HB. Functional specialization
1211 in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6)
1212 before the radiation of Viridiplantae. BMC Evol Biol. 2011;11: 123. doi:10.1186/1471-2148-11-123
1213 95. Do DN, Schenkel FS, Miglior F, Zhao X, Ibeagha-Awemu EM. Genome wide association study
1214 identifies novel potential candidate genes for bovine milk cholesterol content. Sci Rep. 2018;8:
1215 13239. doi:10.1038/s41598-018-31427-0
1216 96. Prentice LM, d'Anglemont de Tassigny X, McKinney S, Ruiz de Algara T, Yap D, Turashvili G,
1217 et al. The testosterone-dependent and independent transcriptional networks in the hypothalamus
1218 of Gpr54 and Kiss1 knockout male mice are not fully equivalent. BMC Genomics. 2011;12: 209.
1219 doi:10.1186/1471-2164-12-209
1220 97. Rodriguez PQ, Oddsson A, Ebarasi L, He B, Hultenby K, Wernerson A, et al. Knockdown of
1221 Tmem234 in zebrafish results in proteinuria. Am J Physiol Renal Physiol. 2015;309: F955-966.
1222 doi:10.1152/ajprenal.00525.2014
1223 98. Dean N, Zhang YB, Poster JB. The VRG4 gene is required for GDP-mannose transport into the
1224 lumen of the Golgi in the yeast, Saccharomyces cerevisiae. J Biol Chem. 1997;272: 31908–31914.
1225 doi:10.1074/jbc.272.50.31908
1226 99. Baldwin TC, Handford MG, Yuseff MI, Orellana A, Dupree P. Identification and
1227 characterization of GONST1, a golgi-localized GDP-mannose transporter in Arabidopsis. Plant Cell.
1228 2001;13: 2283–2295. doi:10.1105/tpc.010247
1229 100. Rodríguez A, Gonzalez L, Ko A, Alvarez M, Miao Z, Bhagat Y, et al. Molecular Characterization
1230 of the Lipid Genome-Wide Association Study Signal on Chromosome 18q11.2 Implicates HNF4A-
1231 Mediated Regulation of the TMEM241 Gene. Arterioscler Thromb Vasc Biol. 2016;36: 1350–1355.
1232 doi:10.1161/ATVBAHA.116.307182
1233 101. Trifonov S, Houtani T, Shimizu J-I, Hamada S, Kase M, Maruyama M, et al. GPR155: Gene
1234 organization, multiple mRNA splice variants and expression in mouse central nervous system.
1235 Biochem Biophys Res Commun. 2010;398: 19–25. doi:10.1016/j.bbrc.2010.05.162
1236 102. Wang XC, Liu Z, Jin LH. Anchor negatively regulates BMP signalling to control Drosophila
1237 wing development. Eur J Cell Biol. 2018;97: 308–317. doi:10.1016/j.ejcb.2018.04.007
1238 103. Shimizu D, Kanda M, Tanaka H, Kobayashi D, Tanaka C, Hayashi M, et al. GPR155 Serves as a
1239 Predictive Biomarker for Hematogenous Metastasis in Patients with Gastric Cancer. Sci Rep. 2017;7:
1240 42089. doi:10.1038/srep42089

1241 104. Umeda S, Kanda M, Sugimoto H, Tanaka H, Hayashi M, Yamada S, et al. Downregulation of
1242 GPR155 as a prognostic factor after curative resection of hepatocellular carcinoma. BMC Cancer.
1243 2017;17: 610. doi:10.1186/s12885-017-3629-2
1244 105. Helenius J, Ng DTW, Marolda CL, Walter P, Valvano MA, Aebi M. Translocation of lipid-linked
1245 oligosaccharides across the ER membrane requires Rft1 protein. Nature. 2002;415: 447–450.
1246 doi:10.1038/415447a
1247 106. Frank CG, Sanyal S, Rush JS, Waechter CJ, Menon AK. Does Rft1 flip an N-glycan lipid
1248 precursor? Nature. 2008;454: E3-4; discussion E4-5. doi:10.1038/nature07165
1249 107. Gottier P, Gonzalez-Salgado A, Menon AK, Liu Y-C, Acosta-Serrano A, Bütikofer P. RFT1
1250 Protein Affects Glycosylphosphatidylinositol (GPI) Anchor Glycosylation. J Biol Chem. 2017;292:
1251 1103–1111. doi:10.1074/jbc.M116.758367
1252 108. Verchère A, Cowton A, Jenni A, Rauch M, Häner R, Graumann J, et al. Complexity of the
1253 eukaryotic dolichol-linked oligosaccharide scramblase suggested by activity correlation profiling
1254 mass spectrometry. Sci Rep. 2021;11: 1411. doi:10.1038/s41598-020-80956-0
1255 109. Nobre LS, Al-Shahrour F, Dopazo J, Saraiva LM. Exploring the antimicrobial action of a
1256 carbon monoxide-releasing compound through whole-genome transcription profiling of Escherichia
1257 coli. Microbiology (Reading). 2009;155: 813–824. doi:10.1099/mic.0.023911-0
1258 110. Herzberg M, Kaye IK, Peti W, Wood TK. YdgG (TqsA) controls biofilm formation in
1259 Escherichia coli K-12 through autoinducer 2 transport. J Bacteriol. 2006;188: 587–598.
1260 doi:10.1128/JB.188.2.587-598.2006
1261 111. Dong P, Wang L, Song N, Yang L, Chen J, Yan M, et al. A UPF0118 family protein with
1262 uncharacterized function from the moderate halophile Halobacillus andaensis represents a novel
1263 class of Na+(Li+)/H+ antiporter. Sci Rep. 2017;7: 45936. doi:10.1038/srep45936
1264 112. Wang L, Zou Q, Yan M, Wang Y, Guo S, Zhang R, et al. Polar or Charged Residues Located in
1265 Four Highly Conserved Motifs Play a Vital Role in the Function or pH Response of a UPF0118 Family
1266 Na+(Li+)/H+ Antiporter. Front Microbiol. 2020;11: 841. doi:10.3389/fmicb.2020.00841
1267 113. Shao L, Xu T, Zheng X, Shao D, Zhang H, Chen H, et al. A novel three-TMH Na+/H+ antiporter
1268 and the functional role of its oligomerization. J Mol Biol. 2021;433: 166730.
1269 doi:10.1016/j.jmb.2020.166730
1270 114. Mesdaghi S, Murphy DL, Sánchez Rodríguez F, Burgos-Mármol JJ, Rigden DJ. In silico
1271 prediction of structure and function for a large family of transmembrane proteins that includes
1272 human Tmem41b. F1000Res. 2020;9: 1395. doi:10.12688/f1000research.27676.2
1273 115. Morita K, Hama Y, Izume T, Tamura N, Ueno T, Yamashita Y, et al. Genome-wide CRISPR
1274 screen identifies TMEM41B as a gene required for autophagosome formation. J Cell Biol. 2018;217:
1275 3817–3828. doi:10.1083/jcb.201804132
1276 116. Moretti F, Bergman P, Dodgson S, Marcellin D, Claerr I, Goodwin JM, et al. TMEM41B is a
1277 novel regulator of autophagy and lipid mobilization. EMBO Rep. 2018;19.
1278 doi:10.15252/embr.201845889
1279 117. Van Alstyne M, Lotti F, Dal Mas A, Area-Gomez E, Pellizzoni L. Stasimon/Tmem41b localizes
1280 to mitochondria-associated ER membranes and is essential for mouse embryonic development.
1281 Biochem Biophys Res Commun. 2018;506: 463–470. doi:10.1016/j.bbrc.2018.10.073
1282 118. Morita K, Hama Y, Mizushima N. TMEM41B functions with VMP1 in autophagosome
1283 formation. Autophagy. 2019;15: 922–923. doi:10.1080/15548627.2019.1582952
1284 119. Shoemaker CJ, Huang TQ, Weir NR, Polyakov NJ, Schultz SW, Denic V. CRISPR screening
1285 using an expanded toolkit of autophagy reporters identifies TMEM41B as a novel autophagy factor.
1286 PLoS Biol. 2019;17: e2007044. doi:10.1371/journal.pbio.2007044

1287 120. Schneider WM, Luna JM, Hoffmann H-H, Sánchez-Rivera FJ, Leal AA, Ashbrook AW, et al.
1288 Genome-scale identification of SARS-CoV-2 and pan-coronavirus host factor networks. bioRxiv.
1289 2020; 2020.10.07.326462. doi:10.1101/2020.10.07.326462
1290 121. Hoffmann H-H, Schneider WM, Rozen-Gagnon K, Miles LA, Schuster F, Razooky B, et al.
1291 TMEM41B Is a Pan-flavivirus Host Factor. Cell. 2021;184: 133-148.e20.
1292 doi:10.1016/j.cell.2020.12.005
1293 122. Farwell SLN, Kanyi D, Hamel M, Slee JB, Miller EA, Cipolle MD, et al. Heparin Decreases in
1294 Tumor Necrosis Factor α (TNFα)-induced Endothelial Stress Responses Require Transmembrane
1295 Protein 184A and Induction of Dual Specificity Phosphatase 1. J Biol Chem. 2016;291: 5342–5354.
1296 doi:10.1074/jbc.M115.681288
1297 123. Dawson PA, Hubbert M, Haywood J, Craddock AL, Zerangue N, Christian WV, et al. The
1298 heteromeric organic solute transporter alpha-beta, Ostalpha-Ostbeta, is an ileal basolateral bile
1299 acid transporter. J Biol Chem. 2005;280: 6960–6968. doi:10.1074/jbc.M412752200
1300 124. Seward DJ, Koh AS, Boyer JL, Ballatori N. Functional complementation between a novel
1301 mammalian polygenic transport complex and an evolutionarily ancient organic solute transporter,
1302 OSTalpha-OSTbeta. J Biol Chem. 2003;278: 27473–27482. doi:10.1074/jbc.M301106200
1303 125. Wang W, Seward DJ, Li L, Boyer JL, Ballatori N. Expression cloning of two genes that together
1304 mediate organic solute and steroid transport in the liver of a marine vertebrate. Proc Natl Acad Sci
1305 USA. 2001;98: 9431–9436. doi:10.1073/pnas.161099898
1306 126. Rasmussen RN, Christensen KV, Holm R, Nielsen CU. Nfat5 is involved in the hyperosmotic
1307 regulation of Tmem184b: a putative modulator of ibuprofen transport in renal MDCK I cells. FEBS
1308 Open Bio. 2019;9: 1071–1081. doi:10.1002/2211-5463.12630
1309 127. Zhu H, Shang D, Sun M, Choi S, Liu Q, Hao J, et al. X-linked congenital hypertrichosis
1310 syndrome is associated with interchromosomal insertions mediated by a human-specific
1311 palindrome near SOX3. Am J Hum Genet. 2011;88: 819–826. doi:10.1016/j.ajhg.2011.05.004
1312 128. Quamme GA. Molecular identification of ancient and modern mammalian magnesium
1313 transporters. Am J Physiol, Cell Physiol. 2010;298: C407-429. doi:10.1152/ajpcell.00124.2009
1314 129. Stuiver M, Lainez S, Will C, Terryn S, Günzel D, Debaix H, et al. CNNM2, encoding a
1315 basolateral protein required for renal Mg2+ handling, is mutated in dominant hypomagnesemia.
1316 Am J Hum Genet. 2011;88: 333–343. doi:10.1016/j.ajhg.2011.02.005
1317 130. Yamazaki D, Funato Y, Miura J, Sato S, Toyosawa S, Furutani K, et al. Basolateral Mg2+
1318 extrusion via CNNM4 mediates transcellular Mg2+ transport across epithelia: a mouse model. PLoS
1319 Genet. 2013;9: e1003983. doi:10.1371/journal.pgen.1003983
1320 131. Funato Y, Furutani K, Kurachi Y, Miki H. CrossTalk proposal: CNNM proteins are Na+ /Mg2+
1321 exchangers playing a central role in transepithelial Mg2+ (re)absorption. J Physiol. 2018;596: 743–
1322 746. doi:10.1113/JP275248
1323 132. Arjona FJ, de Baaij JHF. CrossTalk opposing view: CNNM proteins are not Na+ /Mg2+
1324 exchangers but Mg2+ transport regulators playing a central role in transepithelial Mg2+
1325 (re)absorption. J Physiol (Lond). 2018;596: 747–750. doi:10.1113/JP275249
1326 133. Funato Y, Furutani K, Kurachi Y, Miki H. Rebuttal from Yosuke Funato, Kazuharu Furutani,
1327 Yoshihisa Kurachi and Hiroaki Miki. J Physiol. 2018;596: 751. doi:10.1113/JP275706
1328 134. Arjona FJ, de Baaij JHF. Rebuttal from Francisco J. Arjona and Jeroen H. F. de Baaij. J Physiol.
1329 2018;596: 753–754. doi:10.1113/JP275705
1330 135. Sponder G, Mastrototaro L, Kurth K, Merolle L, Zhang Z, Abdulhanan N, et al. Human
1331 CNNM2 is not a Mg(2+) transporter per se. Pflugers Arch. 2016;468: 1223–1240.
1332 doi:10.1007/s00424-016-1816-7

1333 136. Huang Y, Jin F, Funato Y, Xu Z, Zhu W, Wang J, et al. Structural basis for the Mg2+
1334 recognition and regulation of the CorC Mg2+ transporter. Sci Adv. 2021;7: eabe6140.
1335 doi:10.1126/sciadv.abe6140
1336 137. Hogue DL, Ellison MJ, Young JD, Cass CE. Identification of a novel membrane transporter
1337 associated with intracellular membranes by phenotypic complementation in the yeast
1338 Saccharomyces cerevisiae. J Biol Chem. 1996;271: 9801–9808. doi:10.1074/jbc.271.16.9801
1339 138. Cabrita MA, Hobman TC, Hogue DL, King KM, Cass CE. Mouse transporter protein, a
1340 membrane protein that regulates cellular multidrug resistance, is localized to lysosomes. Cancer
1341 Res. 1999;59: 4890–4897.
1342 139. Hogue DL, Kerby L, Ling V. A mammalian lysosomal membrane protein confers multidrug
1343 resistance upon expression in Saccharomyces cerevisiae. J Biol Chem. 1999;274: 12877–12882.
1344 doi:10.1074/jbc.274.18.12877
1345 140. Tian S, Muneeruddin K, Choi MY, Tao L, Bhuiyan RH, Ohmi Y, et al. Genome-wide CRISPR
1346 screens for Shiga toxins and ricin reveal Golgi proteins critical for glycosylation. PLoS Biol. 2018;16:
1347 e2006951. doi:10.1371/journal.pbio.2006951
1348 141. Yamaji T, Sekizuka T, Tachida Y, Sakuma C, Morimoto K, Kuroda M, et al. A CRISPR Screen
1349 Identifies LAPTM4A and TM9SF Proteins as Glycolipid-Regulating Factors. iScience. 2019;11: 409–
1350 424. doi:10.1016/j.isci.2018.12.039
1351 142. Shao G-Z, Zhou R-L, Zhang Q-Y, Zhang Y, Liu J-J, Rui J-A, et al. Molecular cloning and
1352 characterization of LAPTM4B, a novel gene upregulated in hepatocellular carcinoma. Oncogene.
1353 2003;22: 5060–5069. doi:10.1038/sj.onc.1206832
1354 143. Adra CN, Zhu S, Ko JL, Guillemot JC, Cuervo AM, Kobayashi H, et al. LAPTM5: a novel
1355 lysosomal-associated multispanning membrane protein preferentially expressed in hematopoietic
1356 cells. Genomics. 1996;35: 328–337. doi:10.1006/geno.1996.0364
1357 144. Grabner A, Brast S, Sucic S, Bierer S, Hirsch B, Pavenstädt H, et al. LAPTM4A interacts with
1358 hOCT2 and regulates its endocytotic recruitment. Cell Mol Life Sci. 2011;68: 4079–4090.
1359 doi:10.1007/s00018-011-0694-6
1360 145. Milkereit R, Persaud A, Vanoaica L, Guetg A, Verrey F, Rotin D. LAPTM4b recruits the LAT1-
1361 4F2hc Leu transporter to lysosomes and promotes mTORC1 activation. Nat Commun. 2015;6: 7250.
1362 doi:10.1038/ncomms8250
1363 146. Li L, Wei XH, Pan YP, Li HC, Yang H, He QH, et al. LAPTM4B: a novel cancer-associated gene
1364 motivates multidrug resistance through efflux and activating PI3K/AKT signaling. Oncogene.
1365 2010;29: 5785–5795. doi:10.1038/onc.2010.303
1366 147. Rutsch F, Gailus S, Miousse IR, Suormala T, Sagné C, Toliat MR, et al. Identification of a
1367 putative lysosomal cobalamin exporter altered in the cblF defect of vitamin B12 metabolism. Nat
1368 Genet. 2009;41: 234–239. doi:10.1038/ng.294
1369 148. Kawaguchi K, Okamoto T, Morita M, Imanaka T. Translocation of the ABC transporter ABCD4
1370 from the endoplasmic reticulum to lysosomes requires the escort protein LMBD1. Sci Rep. 2016;6:
1371 30183. doi:10.1038/srep30183
1372 149. Kitai K, Kawaguchi K, Tomohiro T, Morita M, So T, Imanaka T. The lysosomal protein ABCD4
1373 can transport vitamin B12 across liposomal membranes in vitro. J Biol Chem. 2021;296: 100654.
1374 doi:10.1016/j.jbc.2021.100654
1375 150. Wojnar P, Lechner M, Merschak P, Redl B. Molecular cloning of a novel lipocalin-1
1376 interacting human cell membrane receptor using phage display. J Biol Chem. 2001;276: 20206–
1377 20212. doi:10.1074/jbc.M101762200

1378 151. Hesselink RW, Findlay JBC. Expression, characterization and ligand specificity of lipocalin-1
1379 interacting membrane receptor (LIMR). Mol Membr Biol. 2013;30: 327–337.
1380 doi:10.3109/09687688.2013.823018
1381 152. Paek J, Kalocsay M, Staus DP, Wingler L, Pascolutti R, Paulo JA, et al. Multidimensional
1382 Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. Cell. 2017;169: 338-
1383 349.e11. doi:10.1016/j.cell.2017.03.028
1384 153. Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh
1385 enhancer regulates expression in the developing limb and fin and is associated with preaxial
1386 polydactyly. Hum Mol Genet. 2003;12: 1725–1735. doi:10.1093/hmg/ddg180
1387 154. Ianakiev P, van Baren MJ null, Daly MJ, Toledo SP, Cavalcanti MG, Neto JC, et al.
1388 Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene.
1389 Am J Hum Genet. 2001;68: 38–45. doi:10.1086/316955
1390 155. MacGrogan D, Levy A, Bova GS, Isaacs WB, Bookstein R. Structure and methylation-
1391 associated silencing of a gene within a homozygously deleted region of human chromosome band
1392 8p22. Genomics. 1996;35: 55–65. doi:10.1006/geno.1996.0322
1393 156. Knauer R, Lehle L. The oligosaccharyltransferase complex from Saccharomyces cerevisiae.
1394 Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. J
1395 Biol Chem. 1999;274: 17249–17256. doi:10.1074/jbc.274.24.17249
1396 157. Knauer R, Lehle L. The oligosaccharyltransferase complex from yeast. Biochim Biophys Acta.
1397 1999;1426: 259–273. doi:10.1016/s0304-4165(98)00128-7
1398 158. Kelleher DJ, Karaoglu D, Mandon EC, Gilmore R. Oligosaccharyltransferase isoforms that
1399 contain different catalytic STT3 subunits have distinct enzymatic properties. Mol Cell. 2003;12: 101–
1400 111. doi:10.1016/s1097-2765(03)00243-0
1401 159. Cherepanova NA, Shrimal S, Gilmore R. Oxidoreductase activity is necessary for N-
1402 glycosylation of cysteine-proximal acceptor sites in glycoproteins. J Cell Biol. 2014;206: 525–539.
1403 doi:10.1083/jcb.201404083
1404 160. Ramírez AS, Kowal J, Locher KP. Cryo-electron microscopy structures of human
1405 oligosaccharyltransferase complexes OST-A and OST-B. Science. 2019;366: 1372–1375.
1406 doi:10.1126/science.aaz3505
1407 161. Goytain A, Quamme GA. Identification and characterization of a novel mammalian Mg2+
1408 transporter with channel-like properties. BMC Genomics. 2005;6: 48. doi:10.1186/1471-2164-6-48
1409 162. Zhou H, Clapham DE. Mammalian MagT1 and TUSC3 are required for cellular magnesium
1410 uptake and vertebrate embryonic development. Proc Natl Acad Sci USA. 2009;106: 15750–15755.
1411 doi:10.1073/pnas.0908332106
1412 163. Li N, Gügel IL, Giavalisco P, Zeisler V, Schreiber L, Soll J, et al. FAX1, a novel membrane
1413 protein mediating plastid fatty acid export. PLoS Biol. 2015;13: e1002053.
1414 doi:10.1371/journal.pbio.1002053
1415 164. Nilsson R, Schultz IJ, Pierce EL, Soltis KA, Naranuntarat A, Ward DM, et al. Discovery of genes
1416 essential for heme biosynthesis through large-scale gene expression analysis. Cell Metab. 2009;10:
1417 119–130. doi:10.1016/j.cmet.2009.06.012
1418 165. Yien YY, Robledo RF, Schultz IJ, Takahashi-Makise N, Gwynn B, Bauer DE, et al. TMEM14C is
1419 required for erythroid mitochondrial heme metabolism. J Clin Invest. 2014;124: 4294–4304.
1420 doi:10.1172/JCI76979
1421 166. Yien YY, Ringel AR, Paw BH. Mitochondrial transport of protoporphyrinogen IX in erythroid
1422 cells. Oncotarget. 2015;6: 20742–20743. doi:10.18632/oncotarget.5124

1423 167. Klammt C, Maslennikov I, Bayrhuber M, Eichmann C, Vajpai N, Chiu EJC, et al. Facile
1424 backbone structure determination of human membrane proteins by NMR spectroscopy. Nat
1425 Methods. 2012;9: 834–839. doi:10.1038/nmeth.2033
1426 168. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A,
1427 et al. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2.
1428 doi:10.26508/lsa.201900429
1429 169. Shen D-W, Ma J, Okabe M, Zhang G, Xia D, Gottesman MM. Elevated expression of
1430 TMEM205, a hypothetical membrane protein, is associated with cisplatin resistance. J Cell Physiol.
1431 2010;225: 822–828. doi:10.1002/jcp.22287
1432 170. Gallenito MJ, Qasim TS, Tutol JN, Prakash V, Dodani SC, Meloni G. A recombinant platform
1433 to characterize the role of transmembrane protein hTMEM205 in Pt(II)-drug resistance and
1434 extrusion. Metallomics. 2020;12: 1542–1554. doi:10.1039/d0mt00114g
1435 171. Chang J-M, Di Tommaso P, Taly J-F, Notredame C. Accurate multiple sequence alignment of
1436 transmembrane proteins with PSI-Coffee. BMC Bioinformatics. 2012;13 Suppl 4: S1.
1437 doi:10.1186/1471-2105-13-S4-S1
1438 172. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
1439 architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421
1440 173. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships.
1441 University of Kansas Science Bulletin. 1958;38: 1409–1438.
1442 174. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast
1443 remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20: 473.
1444 doi:10.1186/s12859-019-3019-7
1445 175. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust
1446 databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res.
1447 2017;45: D170–D176. doi:10.1093/nar/gkw1081
1448 176. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of
1449 high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539.
1450 doi:10.1038/msb.2011.75
1451 177. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein
1452 sequences. Protein Sci. 2018;27: 135–145. doi:10.1002/pro.3290
1453 178. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol.
1454 2017;34: 2422–2424. doi:10.1093/molbev/msx149
1455 179. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and
1456 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.
1457 Syst Biol. 2010;59: 307–321. doi:10.1093/sysbio/syq010
1458 180. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate,
1459 and powerful alternative. Syst Biol. 2006;55: 539–552. doi:10.1080/10635150600755453
1460 181. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and
1461 optimizing gene family trees. J Comput Biol. 2000;7: 429–447. doi:10.1089/106652700750050871
1462
1463
1464