

Illuminating the Function of the Orphan Transporter, SLC22A10 in Humans and Other Primates

Kathleen Giacomini (✉ kathy.giacomini@ucsf.edu)

University of California, San Francisco <https://orcid.org/0000-0001-8041-5430>

Sook Wah Yee

University of California, San Francisco

Luis Ferrández-Peral

Instituto de Biología Evolutiva

Pol Alentorn

Institute of Evolutionary Biology (UPF-CSIC)

Claudia Fontserè

The Globe Institute, University of Copenhagen <https://orcid.org/0000-0003-2233-6026>

Merve Ceylan

Uppsala University <https://orcid.org/0000-0003-4252-6329>

Megan Koleske

University of California, San Francisco

Niklas Handin

Uppsala University

Virginia Artegoitia

United States Department of Agriculture

Giovanni Lara

University of California, San Francisco <https://orcid.org/0009-0004-6706-0945>

Huan-Chieh Chien

University of California, San Francisco

Xujia Zhou

University of California, San Francisco

Jacques Dainat

University of Montpellier <https://orcid.org/0000-0002-6629-0173>

Arthur Zalevsky

University of California, San Francisco <https://orcid.org/0000-0001-6987-8119>

Andrej Sali

University of California, San Francisco <https://orcid.org/0000-0003-0435-6197>

Colin Brand

University of California, San Francisco

John Capra

University of California, San Francisco <https://orcid.org/0000-0001-9743-1795>

Per Artursson

Uppsala University <https://orcid.org/0000-0002-3708-7395>

John Newman

USDA-ARS-WHNRC <https://orcid.org/0000-0001-9632-6571>

Tomas Marques-Bonet

Institut de Biologia Evolutiva (CSIC/Universitat Pompeu Fabra) / CNAG (Centro de Analisis Genomico)
PCB <https://orcid.org/0000-0002-5597-3075>

Article

Keywords:

Posted Date: September 14th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3263845/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

There is **NO** Competing Interest.

Table 1 is available in the Supplementary Files section.

1 **Title:** Illuminating the Function of the Orphan Transporter, SLC22A10 in Humans
2 and Other Primates
3

4 **Authors:** Sook Wah Yee¹, Luis Ferrández-Peral², Pol Alentorn², Claudia Fontserè³, Merve Ceylan⁴,
5 Megan L. Koleske¹, Niklas Handin⁴, Virginia M. Artegoitia⁵, Giovanni Lara¹, Huan-Chieh Chien¹, Xujia
6 Zhou¹, Jacques Dainat⁶, Arthur Zalevsky¹, Andrej Sali⁷, Colin M. Brand⁸, John A. Capra⁸, Per Artursson⁴,
7 John W. Newman⁹, Tomas Marques-Bonet¹⁰, Kathleen M. Giacomini^{1,11,*}
8

9 **Affiliations:**

10 ¹Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco,
11 California, USA.

12 ²Institute of Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain

13 ³Institute of Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain; Center for Evolutionary
14 Hologenomics, The Globe Institute, University of Copenhagen, Øster Farimagsgade 5A, 1352
15 Copenhagen, Denmark.

16 ⁴Department of Pharmacy and Science for Life Laboratory, Uppsala University, P.O. Box 580, 75123,
17 Uppsala, Sweden.

18 ⁵United States Department of Agriculture, Agricultural Research Service, Western Human Nutrition
19 Research Center, Davis, CA 95616, USA.

20 ⁶Joint Research Unit for Infectious Diseases and Vectors Ecology Genetics Evolution and Control
21 (MIVEGEC), University of Montpellier, French National Center for Scientific Research (CNRS 5290),
22 French National Research Institute for Sustainable Development (IRD 224), 911 Avenue Agropolis, BP
23 64501, 34394 Montpellier Cedex 5, France.

24 ⁷Department of Bioengineering and Therapeutic Sciences, UCSF Box 0775 1700 4th St, University of
25 California, San Francisco, San Francisco, CA 94158, United States; Department of Pharmaceutical
26 Chemistry, University of California, San Francisco, UCSF Box 2880 600 16th St, San Francisco, CA
27 94143, United States; Quantitative Biosciences Institute (QBI), University of California, San Francisco,
28 1700 4th St, San Francisco, CA, United States.

29 ⁸Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA;
30 Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA.

31 ⁹United States Department of Agriculture, Agricultural Research Service, Western Human Nutrition
32 Research Center, Davis, CA 95616, USA; Department of Nutrition, University of California, Davis, Davis,
33 CA 95616, USA; UC Davis West Coast Metabolomics Center, Davis, CA 95616, USA

34 ¹⁰Institute of Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain; Institute of Evolutionary
35 Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain; Catalan Institution of Research
36 and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain; CNAG,
37 Centro Nacional de Análisis Genómico, Barcelona Institute of Science and Technology (BIST), Baldiri i
38 Reixac 4, 08028 Barcelona, Spain; Institut Català de Paleontologia Miquel Crusafont, Universitat
39 Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona,
40 Spain

41 ¹¹Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco,
42 California, USA; Institute for Human Genetics, University of California San Francisco, San Francisco, CA,
43 USA.

44
45 *Corresponding author: kathy.giacomini@ucsf.edu
46

47 **Abstract**

48 SLC22A10 is classified as an orphan transporter with unknown substrates and function. Here we
49 describe the discovery of the substrate specificity and functional characteristics of SLC22A10. The
50 human SLC22A10 tagged with green fluorescent protein was found to be absent from the plasma
51 membrane, in contrast to the SLC22A10 orthologs found in great apes. Estradiol-17 β -glucuronide
52 accumulated in cells expressing great ape SLC22A10 orthologs (over 4-fold, $p < 0.001$). In contrast,
53 human SLC22A10 displayed no uptake function. Sequence alignments revealed two amino acid
54 differences including a proline at position 220 of the human SLC22A10 and a leucine at the same
55 position of great ape orthologs. Site-directed mutagenesis yielding the human SLC22A10-P220L
56 produced a protein with excellent plasma membrane localization and associated uptake function.
57 Neanderthal and Denisovan genomes show human-like sequences at proline 220 position, corroborating
58 that SLC22A10 were rendered nonfunctional during hominin evolution after the divergence from the pan
59 lineage (chimpanzees and bonobos). These findings demonstrate that human SLC22A10 is a unitary
60 pseudogene and was inactivated by a missense mutation that is fixed in humans, whereas orthologs in
61 great apes transport sex steroid conjugates.

62
63 **Introduction**

64 About 30% of the members of the large Solute Carrier Superfamily in the human genome have no known
65 substrate¹, representing a major gap in understanding human biology. Deorphaning is the process of
66 determining the function of a protein that has not yet been characterized. For deorphaning proteins in
67 the SLC superfamily, which includes multi-membrane spanning transporters, phylogenetic analysis
68 represents the first step for identifying the substrates of an orphan transporter. Other methods include
69 metabolomic methods in cells or in knockout mice²⁻⁴. For example, the substrate of mouse SLC16A6, a
70 transporter in the MonoCarboxylate Transporter Family, MCT7, was discovered through the analysis of
71 amino acid transport in cell lines that overexpressed MCT7⁵.

72
73 There has been increasing interest in deorphaning solute carrier transporters due to the significant
74 potential role of SLCs in human physiology^{1, 6}. While some orphan genes in the SLC superfamily encode
75 proteins which have evolved other functions and do not participate in transmembrane solute flux, it is
76 more probable that these multi-membrane spanning proteins primarily serve as transporters. Therefore,
77 efforts to deorphan SLC family members should include attempts to identify their endogenous substrates,
78 ligands of the transporter that are translocated across biological membranes. In certain cases, the roles
79 of transporters have been discovered, yet the substrate of the transporter remains unknown. This creates
80 gaps in our mechanistic understanding of the transporter's function in those processes. For example,
81 Ren *et al.* used untargeted metabolomics and found elevated levels of lipid diacylglycerol and altered
82 fatty acid metabolites in liver and plasma samples of Mct6 knockout mice⁷. This finding supports a role of
83 SLC16A5 in lipid and amino acid homeostasis, but does not reveal its substrates and as such, the

84 mechanism remains poorly understood. Similarly, the function of the orphan transporter SLC38A10
85 (SNAT10) was assessed by studying mice lacking the *Slc38a10* gene (Slc38a10-deficient mice). The
86 findings indicated that *Slc38a10*-deficient mice exhibited reduced body weight and lower plasma levels of
87 threonine and histidine. However, no studies have specifically investigated whether these amino acids
88 serve as substrates for SLC38A10⁸; therefore a gap in understanding the mechanism by which the
89 transporter affects body weight remains. Information regarding recently deorphaned transporters is
90 presented in a recent review⁹.

91
92 Orphan transporters can be found in over 20 families in the SLC superfamily¹. In the Solute Carrier 22
93 family A (SLC22A), there are 28 members that transport organic ions and another 10 that are orphans⁶.
94 Largely representing plasma membrane transporters, members of the SLC22A family are clustered
95 together based on their charge specificity for organic cations (OCTs), organic anions (OATs), and
96 organic zwitterion/cations (OCTNs). Solute carrier 22 family member 10 (SLC22A10) and its direct
97 species orthologs are orphan transporters whose substrates and transport mechanisms are yet to be
98 characterized. In humans, SLC22A10 has been given a protein name of OAT5. Based on Northern
99 blotting¹⁰ and RNA seq studies (<https://www.proteinatlas.org/ENSG00000184999-SLC22A10/tissue>)¹¹,
100 human SLC22A10 is expressed specifically in the liver.

101
102 Orthologs of human SLC22A10 are present in some primates including great apes
103 ([https://useast.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000184999;r=](https://useast.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000184999;r=11:63268022-63311783)
104 [11:63268022-63311783](https://useast.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000184999;r=11:63268022-63311783)). Intrigued by this observation, our study aimed to identify the substrates of
105 SLC22A10 and the transport mechanism by expressing primate orthologs of SLC22A10 in cell lines and
106 performing analytical procedures including cellular uptake studies, metabolomic analyses and proteomic
107 assays. We attempted to identify crucial amino acids that contribute to the differences in function
108 between direct species orthologs in humans and great apes. Kinetic parameters and transport
109 mechanisms of various predicted isoforms of SLC22A10 were determined, along with their ability to
110 accumulate different endogenous ligands. Proteomic studies in cell lines recombinantly expressing
111 human and chimpanzee SLC22A10 were conducted. To the best of our knowledge, this is the first study
112 to characterize the function of the orphan transporter SLC22A10. Our study shows that human
113 SLC22A10 was inactivated by a single missense mutation and is a unitary pseudogene. The ORF-
114 disrupting mutation in SLC22A10, which led to Pro220, is not observed in great apes and primates. This
115 particular amino acid is crucial for protein abundance and expression on the plasma membrane. Our
116 work provides a roadmap for how orthologous genes, along with sequence comparison, and proteomic
117 and transporter assays, can be used to deorphan the function of solute carrier proteins. These
118 discoveries have significant evolutionary implications.

119

120 **Results**

121 ***Human SLC22A10 showed no expression on the plasma membrane and no transporter activity of***
122 ***prototypical anionic substrates of SLC22 family members.***

123 Human SLC22A10 is in a cluster that includes known organic anion transporters: SLC22A24 is closest,
124 followed by SLC22A9, SLC22A11 and SLC22A12 (**Fig. 1A**). Phylogenetic analyses reveals that its
125 closest homolog is SLC22A24. The substrates of SLC22A24 are steroid conjugates, bile acids and
126 dicarboxylic acids, which our laboratory has successfully deorphaned³. Overexpression of human
127 SLC22A10 tagged with GFP in the N-terminal resulted in no detection of a GFP-tagged protein on the
128 plasma membrane (**Fig. 1B**). Furthermore, no uptake of prototypical organic anions was observed in
129 cells expressing SLC22A10 whereas significant uptake was observed in cells expressing known SLC22
130 organic anion transporters including SLC22A6, SLC22A8 and SLC22A24 (**Fig. 1C**).

131 ***The long isoform of chimpanzee and gorilla SLC22A10 was expressed on the plasma membrane***
132 ***whereas the short isoform was not.***

133 The organic anion transporters depicted in **Fig. 1A** consist of 536 to 563 amino acid proteins. Predictions
134 from reliable sources such as Uniprot, Ensembl, and NCBI Nucleotide databases confirmed that the
135 human SLC22A10 gene produces a 541 amino acid isoform. Conversely, according to reports from
136 Ensembl and UniProt, the orthologs of SLC22A10 found in great apes are predicted to have two
137 isoforms: a short isoform comprising 540 amino acids and a longer isoform containing 552 amino acids.
138 Because there was no detectable expression of the human SLC22A10 on the plasma membrane of cells
139 recombinantly expressing the transporter (**Fig. 1B**), we inquired whether the direct species orthologs in
140 great apes exhibited a similar lack of plasma membrane expression when expressed recombinantly in
141 cells. In fact, we observed that the long isoforms (552 amino acids) of both chimpanzee and gorilla
142 SLC22A10 were detected on the plasma membrane (**Fig. 2A**). The shorter isoforms (540 amino acids) of
143 chimpanzee, bonobo and gorilla SLC22A10 showed a similar lack of plasma membrane localization as
144 the human ortholog, which consists of 541 amino acids (**Fig. 2A and Fig. 1B**).

145 ***Chimpanzee and gorilla SLC22A10 expressing the long isoform transport estradiol glucuronide***
146 ***but not other anions that are canonical substrates of members in the SLC22A family.***

147 Because the long forms of the great ape SLC22A10 showed a plasma membrane localization, we
148 attempted to identify substrates of SLC22A10 using isotopic uptake assays in cells recombinantly
149 expressing the long isoforms of the great ape transporters. Typical anions that are canonical substrates
150 of members in the SLC22A family were screened for accumulation in human, chimpanzee, bonobo and
151 gorilla expressing the long as well as the short isoforms. Significant accumulation of [³H]-estradiol-17β-
152 glucuronide and [³H]-androstenediol-3α-glucuronide were observed in cells expressing chimpanzee and
153 gorilla SLC22A10 encoding the long but not the short isoforms (**Fig. 2B, Supplemental Fig. 1A**). No
154 significant uptake was detected for other anions that are canonical substrates of members of the
155 SLC22A family, such as estrone sulfate, taurocholic acid, cGMP, uric acid and succinic acid
156
157

158 (**Supplemental Fig. 1A to F**). However, there was a small but significant uptake of [³H]-methotrexate in
159 HEK293 cells expressing chimpanzee and gorilla SLC22A10 long isoforms (**Supplemental Fig. 1G**).

160
161 ***The SLC22A10 protein in humans consists of 541 amino acids, resulting from a single nucleotide***
162 ***insertion that causes a frameshift in the last exon***

163 The genetic mechanism that led to the formation of the 541 amino acid SLC22A10 protein in humans
164 was investigated. Sequence alignments of the last exon (exon 10) of the SLC22A10 gene was compared
165 between humans and great apes and revealed an insertion of one nucleotide leading to the expression
166 of different isoforms in each species (see **Fig. 2C**). In particular, humans exhibit an A nucleotide insertion
167 at the first base pair of exon 10, which is highly prevalent with an allele frequency of 98% in all
168 populations in gnomAD. In contrast, the adenosine insertion has a 2.5% allele frequency in chimpanzees
169 and is not present in other great apes. The adenosine insertion in the human SLC22A10 gene causes a
170 frameshift and results in a 541 amino acid protein, instead of the predicted 552 amino acids in the
171 SLC22A10 gene of great apes and in the humans who do not harbor the adenosine insertion (**Fig. 2C**).
172 We utilized our previously generated human and liver RNAseq data¹² alongside a large human RNA-seq
173 databases (recount3¹³) to validate the splicing event in human and chimpanzee SLC22A10. Our analysis
174 confirmed that splicing mostly occurs at the exact orthologous genomic region in both species, utilizing
175 the canonical splice sites in their corresponding genomes. Thus, we found no coordinated splicing
176 alterations that compensate for the A nucleotide insertion. Consequently, the additional A nucleotide
177 remains present in the final transcribed transcript in humans. This additional nucleotide provides
178 evidence that human SLC22A10 protein contains 541 amino acids, while the chimpanzee SLC22A10
179 protein comprises 552 amino acids.

180
181 ***Mutagenesis of a single amino acid, at position 220 of the human SLC22A10 ortholog, to the***
182 ***respective amino acid in great apes rescues the function of human SLC22A10.***

183 The alignment of human SLC22A10 and primate ortholog sequences revealed differences in amino acid
184 positions p.Met18Ile and p.Pro220Leu (**Fig. 3A**). Interestingly, site-directed mutagenesis experiments
185 demonstrated that the substitution of proline with leucine at position 220 (p.Pro220Leu) restored the
186 plasma membrane localization and function of human SLC22A10, but the substitution of methionine with
187 isoleucine at position 18 (p.Met18Ile) did not have any effect (see **Fig. 3B**). In contrast, the replacement
188 of leucine with proline at position 220 (p.Leu220Pro) abolished the localization and function of the
189 chimpanzee SLC22A10 (see **Fig. 3B and 3C**). Additionally, we observed that a human-chimpanzee
190 chimera protein, consisting of a fusion of human SLC22A10 (1-533) with chimpanzee SLC22A10 (534-
191 552), while retaining the proline residue at position 220, showed no function (**Fig. 3B and 3C**). However,
192 with a leucine substitution at position 220, SLC22A10 remained functional (**Fig. 3B**).

193
194 ***Human and chimpanzee SLC22A10 with Pro220 exhibit lower protein expression compared to***
195 ***orthologs with Leu220.***

196 The objective of this study was to analyze the protein expression of SLC22A10 in HEK293 Flp-In cells
197 that were transfected with either vector only, or the cDNA of human or chimpanzee SLC22A10. This was
198 achieved by quantifying the global proteomes of the cells, with a specific focus on amino acids at position
199 220 of SLC22A10. Comparable transcript levels of SLC22A10 in HEK293 cells that over-expressed
200 either human or chimpanzee SLC22A10, as well as the respective variants (p.P220L or p.L220P), were
201 observed (**Supplemental Fig. 2**). However, as illustrated in **Table 1**, lower protein expression levels for
202 the human SLC22A10 reference (Proline220) and chimpanzee SLC22A10-L220P were observed when
203 compared to SLC22A10 with leucine at the 220 amino acid position. The results showed that protein
204 levels of human SLC22A10 are approximately 10-fold lower in cells expressing human SLC22A10-
205 Pro220 compared to human SLC22A10-Leu220 (**Table 1**) suggesting that human SLC22A10 is
206 transcribed but the protein is unstable. The lower overall protein expression may explain the lack of
207 detectable expression of human SLC22A10 on the plasma membrane in contrast to orthologs (both
208 human and chimpanzee) of SLC22A10 with Leu220. In particular, human SLC22A10 was not detected
209 on the plasma membrane (**Fig. 1B**), whereas the mutant, SLC22A10-Leu220 exhibited expression on the
210 plasma membrane (**Fig. 3C**).

211
212 ***Human SLC22A10 with Pro220 is predicted to have poor stability compared to orthologs with***
213 ***Leu220.***

214 We tested the effect of the P220L mutation on the protein with the stability prediction pipeline tuned for
215 transmembrane proteins¹⁴. The Rosetta physics-based score showed that P220L is a stabilizing mutation
216 with a $\Delta\Delta G$ value of -9.84 Rosetta Energy Units (REU), a value more negative than less stable mutant.
217 Whereas the rank-normalized evolutionary-based $\Delta\Delta E$ scores are 0.0 for leucine and 0.16 for proline.
218 The combination of a higher $\Delta\Delta E$ value and a less negative $\Delta\Delta G$ value suggests a potential loss of
219 function protein, likely due to decreased stability and cellular abundance¹⁵.

220
221 ***Analyses of SLC22A10 isoforms in chimpanzee, bonobo, orangutan and gibbon***

222 Upon examining various databases (UniProt, Ensembl, and NCBI nucleotide) that predict SLC22A10
223 isoforms, we observed that chimpanzee, bonobo, orangutan, and gibbon are predicted to have different
224 isoforms, including (i) chimpanzee (XM_024347215.1, 533 amino acid); (ii) bonobo (XM_034932815.1,
225 538 amino acid); (iii) gibbon (XM_003274123.1, 533 amino acid); and (iv) orangutan (XM_024255628.1,
226 533 amino acid). The shorter isoforms are derived from alternative acceptor sites (chimpanzee and
227 bonobo) and exon extensions (orangutan and gibbon) resulting in 533-amino acid proteins (see
228 **Supplementary Fig. 3**). We conducted experiments using HEK293 Flp-In cells that were stably
229 transfected with GFP-conjugated chimpanzee, bonobo, and gorilla SLC22A10 to examine the localization
230 and function of the shorter isoforms. Our results indicated that the SLC22A10 533 amino acid isoform
231 from chimpanzee, orangutan and gibbon accumulated [³H]-estradiol-17 β -glucuronide and [³H]-folic acid,
232 but not [³H]-estrone sulfate and [³H]-taurocholic acid, similar to the SLC22A10 isoform with 552 amino
233 acids (**Fig. 4C**, **Supplementary Fig. 4**). Moreover, we observed that longer isoforms of SLC22A10

234 created for the bonobo and orangutan, which were not predicted to express these isoforms (552 amino
235 acids), accumulated [³H]-estradiol-17β-glucuronide, and were also expressed on the plasma membrane
236 (**Supplementary Fig. 4**).

237

238 ***Uptake of various steroid glucuronides by chimpanzee SLC22A10 reveals that 17β-glucuronides***
239 ***are preferred over 3α-glucuronide conjugates***

240 Our experiments revealed significant accumulation of estradiol-3β-glucuronide, estradiol-17β-
241 glucuronide, estrone-3β-glucuronide and testosterone-17β-glucuronide (**Fig. 5A**). We also observed
242 weak but significant accumulation of androstanediol-3α-glucuronide, but no significant accumulation of
243 androsterone-3-glucuronide, etiocholanolone-3α-glucuronide, progesterone, and testosterone (**Fig. 5A**).
244 Although the data are limited, it is intriguing to note the substrate preference of SLC22A10. Specifically,
245 17β-glucuronides appear to be favored over 3α-glucuronide conjugates.

246

247 ***Chimpanzee SLC22A10-mediated uptake is sodium independent, pH independent, chloride***
248 ***dependent and trans-stimulated by glutaric acid.***

249 Transporters in the SLC22 family may be secondary active, in which case they rely on various sources of
250 energy to mediate active flux of their substrates. Accordingly, we examined the transport mechanism for
251 two isoforms (533 and 552 amino acids) of chimpanzee SLC22A10. At the three pH levels evaluated, the
252 uptake of [³H]-estradiol-17β-glucuronide by both isoforms of chimpanzee SLC22A10 were similar (**Fig.**
253 **5B**). While the uptake was not dependent on sodium (**Fig. 5C**), it was partially reduced in the absence of
254 chloride in the buffer (**Fig. 5C**). Several organic anion transporters in the SLC22 family are *trans*-
255 stimulated by dicarboxylic acids^{3, 16}. Similarly, we observed that the uptake of [³H]-estradiol-17β-
256 glucuronide was *trans*-stimulated by glutaric acid but not significantly by other dicarboxylates (α-
257 ketoglutarate and succinic acid) or the monocarboxylic acid, butyrate (**Fig. 5D**). The uptake kinetics of
258 [³H]-estradiol-17β-glucuronide exhibited saturable characteristics with virtually identical Km values for the
259 two protein isoforms at $3.28 \pm 2.25 \mu\text{M}$ and $2.16 \pm 0.59 \mu\text{M}$ for the 533 aa and 552 aa isoforms,
260 respectively (**Fig. 5E**).

261

262 **Discussion**

263 Transporters play a major role in total body homeostasis as they function to regulate the levels of many
264 solutes including endogenous metabolites, essential nutrients, and environmental toxins. Over 30% of
265 genes in the Solute Carrier superfamily have no known function¹. Identifying the substrates or
266 deorphaning transporters presents significant challenges due to their diverse structural and functional
267 characteristics, as well as the intricate cellular and physiological context in which they operate.
268 Overcoming these challenges necessitates the implementation of innovative approaches that integrate
269 computational predictions, high-throughput screening, functional assays, and targeted experimental
270 investigations^{1, 17}. Recent advancements in transporter research in the past four years have led to the
271 identification of ligands for several transporters in the organic ion transporter family, SLC22. Notably,

272 SLC22A14, SLC22A15, and SLC22A24 have been successfully characterized^{2, 3, 18}. In this study, we
273 pursued a distinct approach to deorphaning a human SLC22 family member, SLC22A10 by investigating
274 the function of great ape orthologs. Our comprehensive investigations have yielded five major findings
275 that contribute to our understanding of the function and evolution of SLC22A10 in higher order primates.
276

277 First, SLC22A10 functions as a steroid glucuronide transporter in great apes (as shown in **Fig. 1** and **Fig.**
278 **2**). Unlike other organic anion transporters in the SLC22 family such as SLC22A6, SLC22A7, and
279 SLC22A8, which transport a variety of organic anions including uric acid, steroid glucuronides, bile acids,
280 steroid sulfates, dicarboxylic acid, and phosphate-containing nucleotides⁶, the ortholog of SLC22A10
281 from primates (chimpanzee, bonobo, gorilla, orangutan and gibbon) transported primarily estradiol-17 β -
282 glucuronide (**Fig. S1**) with significantly weaker uptake of other organic anions such as folic acid and
283 methotrexate (**Fig. S1** and **Fig. S4**). Interestingly, the SLC22A10 ortholog in the squirrel monkey, a New
284 World monkey, preferred estrone sulfate over estradiol-17 β - glucuronide (**Fig. S5**) whereas the gene
285 encoding SLC22A10 is absent in Old World monkeys (*Cercopithecidae* family) (ensemble release 110).
286 These results suggest an evolving role of the function of the transporter in non-human primates. In both
287 chimpanzee and squirrel monkey, SLC22A10 is expressed specifically in the liver (NHPRTR,
288 <http://nhprtr.org/>), consistent with a role in steroid metabolism. The closest paralog of SLC22A10 is
289 SLC22A24, a recently deorphaned transporter³. Both transporters take up steroid conjugates though
290 chimpanzee SLC22A22A10 has a narrower substrate specificity than the human SLC22A24.

291

292 Our second major finding is that the function of SLC22A10 is lost in humans. That is, while human
293 SLC22A10 is transcribed, the protein expression is undetectable in human cell lines transfected with
294 human SLC22A10 and in human liver tissue¹⁹⁻²³ (<https://www.proteinatlas.org/>), consistent with a loss of
295 function of the gene in humans. Indeed, proteomic analysis demonstrated that the reference proline
296 (Pro220) at the 220 amino acid position of human SLC22A10 displayed significantly lower protein
297 expression compared to the mutated form, Leu220 (see **Table 1**). Further, as Pro220 is fixed in humans
298 resulting in a loss of function, the gene has been under less selective pressure. Not surprisingly,
299 SLC22A10 harbors a prevalent nonsense variant, p.Trp96Ter (rs1790218), which is frequently observed
300 at high allele frequencies in human populations ranging from ~20% in African to 50% in European (**Fig.**
301 **S5**). Nonsense-mediated decay (NMD) is known to be triggered by premature stop codons. The
302 Genotype-Tissue Expression (GTEx) project uncovered a significant association between the rs1790218
303 variant, which encodes the A-allele and p.Trp96Ter, and a considerable decrease in the transcript level
304 of SLC22A10 samples (source: <https://gtexportal.org/home/snp/rs1790218>). This finding suggests that
305 individuals carrying one copy of the p.Trp96Ter are likely to exhibit an even lower protein signal in the
306 liver in comparison to individuals that harbor p.Trp96. This finding is relevant to how the protein itself
307 (though not a functioning protein) is being lost. During our efforts to clone the human SLC22A10 gene
308 using pooled human liver samples, we selected 27 individual colonies for further analysis and

309 subsequent sequencing. Only four of these colonies contained the complete transcript spanning 1626
310 base pairs. The remaining 23 out of 27 colonies selected had a identical 219 bp deletion, resulting in a
311 truncated sequence of 1407 bp, corresponding to 469 amino acids (**Fig. S6**). However, NCBI predicted a
312 different deletion of 324 bp, starting at the same GT splice donor site that we observed, but extending to
313 a further GT acceptor site (NCBI Reference Sequence: XM_047426921).

314
315 Our third major finding is that Leu220Pro alone caused the loss of function of the gene in hominins. That
316 is, a single proline substitution resulted in no expression of SLC22A10 on the plasma membrane and
317 significantly reduced protein abundance (**Table1**). When Pro220 in human SLC22A10 was mutated to
318 Leu220, it acquired the functional capacity observed in chimpanzee SLC22A10, resulting in substantial
319 accumulation of the substrate, estradiol-17 β -glucuronide. Conversely, when the amino acid at position
320 220 in chimpanzee SLC22A10 was mutated to proline, the uptake of estradiol-17 β -glucuronide was
321 completely abolished (**Fig. 3B**). Cordes F. et al. (2002) reported that distortions of transmembrane
322 helices can be induced by the presence of proline²⁴. Using AlphaFold2 model of human SLC22A10, we
323 observed that the proline 220 in humans introduces a kink in the alpha-helix, which in turn can affect the
324 conformation of the 225-230 loop and, consequently increasing the accessibility of the lysine 230 for
325 ubiquitination. In addition, based on the $\Delta\Delta G$ calculation the proline 220 is significantly less stable with a
326 less negative value of $\Delta\Delta G$. Based on comprehensive genomics datasets from large-scale populations
327 such as gnomAD²⁵, GDBIG (<http://www.bigcs.com.cn/>), TopMed Freeze 8²⁶, and GME Variome²⁷, there
328 have been no reported individuals with SLC22A10-Leu220. Additionally, four high-coverage archaic
329 hominin genomes – three Neanderthals and a Denisovan are homozygous for the C allele
330 (chr11:63297455 (hg38)), suggesting that this mutation emerged following Pan-Homo divergence and
331 before modern humans diverged from archaic hominins. While the SLC22A10-Trp96Ter variant has not
332 been assayed in archaic hominin or ancient human genomes, two SNPs in strong LD with this variant
333 occur in ancient human genomes at least 30,000 years ago (**Fig. S7**) and the variant itself is estimated to
334 emerge ~120,000 years ago (**Fig. S8**).

335
336 Our fourth finding unveils the transporter activity and expression of the predicted isoforms of SLC22A10
337 on the plasma membrane in great apes (refer to **Fig. 2**, **Fig. 3**, and **Fig. 4**). Importantly, among the great
338 apes, three isoforms were initially predicted: 533, 540, and 552 amino acids in length. However, further
339 investigation revealed that only the isoforms with 533 and 552 amino acids were actually expressed on
340 the plasma membrane and exhibited transporter activity, as depicted in **Figure 4** and **Figure 5**. In **Figure**
341 **2C**, the last exons (exon number 10) for humans and other great apes is presented. It is worth noting that
342 human SLC22A10 is predicted to consist of 541 amino acids. The gene annotation of chimpanzees and
343 other primates in Ensembl is limited by an anthropocentric approach that heavily relies on human
344 annotation as a reference. In the current version 110 of Ensemble.org, the chimpanzee, gorilla, bonobo
345 gene are annotated with 540 amino acids. This annotation utilizes a non-canonical splice site at the end

346 of exon 9 ('CA') instead of the canonical donor splice site ('GT'), which is not supported by our
347 chimpanzee liver RNA-seq data¹² (**Fig. S3** and **Fig. 2C**). In contrast, NCBI predicted that chimpanzee
348 possess isoforms with 533 and 552 amino acids, which is consistent with our observation from
349 chimpanzee liver RNAseq data¹². Within the SLC22A family, several transporters, including SLC22A7²⁸
350 and SLC22A24³ exhibit distinct isoforms.

351
352 Overall, our studies revealed that the activity of SLC22A10 has evolved in primates, with Old World
353 monkeys lack SLC22A10 orthologs, and New World monkeys exhibit a different substrate preference
354 compared to great apes. In addition, the great apes SLC22A10 was rendered nonfunctional by a single
355 missense mutation during hominin evolution after our shared ancestor with the chimpanzee. This
356 missense mutation resulted in a complete loss of human SLC22A10 transporter activity, due to lack
357 of protein expression on plasma membrane and reduce protein abundance. With time, the gene has
358 accumulated additional mutations, including a stop codon (p.Trp96Ter), which has led to reductions in
359 the levels of the mRNA transcript and corresponding reductions in protein level. This gene exhibits
360 features that classify it as unitary pseudogene^{29, 30}. A pseudogene can be defined by the loss of the
361 original function due to errors during transcription or translation, or as a gene producing a protein that
362 does not have the same functional repertoire as the original gene³⁰. Consequently, a pseudogene will not
363 necessarily evolve under a neutral theory of molecular evolution. Pseudogenes can be categorized into 3
364 different types depending on their functional state. These include exapted pseudogenes, which have
365 gained a new biological function; "dying" pseudogenes, which still have some transcriptional activity; and
366 "dead" pseudogenes, which do not exhibit any signs of activity and evolve under the neutral theory²⁹.
367 Based on the evidence at hand, we cannot differentiate if the pseudogene is exapted or dying. However,
368 the SLC22A10 gene is a well-established gene that originated from the last common ancestor of
369 boreoeutheria, with no functioning counterparts in the human genome. Thus, we can classify human
370 SLC22A10 an unitary pseudogene. Examples of unitary pseudogene in human is MUP (major urinary
371 protein), whereas uricase (UOX) and GULO (L-Gulonolactone oxidase) are well established unitary
372 pseudogenes that were inactivated before the separation of human and chimpanzee³¹, There are known
373 pseudogenes in the SLC superfamily (SLC22A20, SLC35E2A, SLC6A10, SLC23A4P, SLC6A21P)³²,
374 however none of them are due to a single missense mutation. Future studies are needed to determine
375 whether the loss of function human SLC22A10-P220 is a favorable situation for humans and whether
376 similar mechanisms have led to the inactivation of other orphan genes in the human genome.

377
378

379 **Figure Legend**

380 **Figure 1.** Analysis of the phylogenetic tree, plasma membrane expression of SLC22A10, and uptake of
381 organic anion substrates of the human SLC22 family.

382 **A.** Multiple sequence alignments were performed with reference amino acid sequences for each anion
383 transporter from humans and rodents, using the Clustal Omega Multiple Sequence Alignment program

384 (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The dendrogram was generated from the output of the
385 Clustal Omega alignment.

386 **B.** Localization of human SLC22A10 conjugated to green fluorescent protein (GFP) was examined in
387 HEK293 cells using high-content imaging and cellular staining with the plasma membrane marker wheat
388 germ agglutinin (WGA). The results showed no colocalization of GFP-tagged SLC22A10 with WGA.

389 **C.** Uptake of various radiolabeled organic anions, which are typical substrates of organic anion
390 transporters in the SLC22A family, was assessed. Uptake was performed 48 hours after transient
391 transfection of plasmids encoding human SLC22A10, GFP expression vector, and one other member in
392 the SLC22A family as a positive control. Accumulation of substrates inside cells was determined after 15
393 minutes. Figure shows a representative plot from one experiment (mean \pm S.D. from three replicate
394 wells). The experiments were repeated at least one time and showed similar results. Multiple
395 comparisons using one-way analysis of variance followed by Dunnett's two-tailed test were performed.
396 HEK293 cells transiently transfected with the GFP vector served as the control. The fold uptake of the
397 substrate, relative to the control cells, was plotted based on one representative experiment conducted in
398 triplicate wells (mean \pm s.d.). The statistical significance for all the cells transfected with organic anion
399 transporters SLC22A6, SLC22A8, or SLC22A10 is $p < 0.001$.

400

401 **Figure 2.** Localization to the plasma membrane, uptake, and sequence comparison of human SLC22A10
402 were examined in comparison with SLC22A10 from great apes (chimpanzee, bonobo, gorilla and
403 orangutan).

404 **A.** This figure shows the plasma membrane localization of SLC22A10 orthologs from great apes, which
405 were conjugated to green fluorescent protein (GFP) in HEK293 cells. The GFP tag is located at the N-
406 terminus of SLC22A10. Confocal imaging revealed that the 552 amino acid isoforms of SLC22A10 from
407 chimpanzee, bonobo, gorilla, and orangutan primarily colocalized with wheat germ agglutinin (WGA) on
408 the plasma membrane of the cell. In contrast, the 540 amino acid isoform of SLC22A10 from bonobo,
409 chimpanzee, and gorilla showed no colocalization of GFP-tagged SLC22A10 with WGA on the plasma
410 membrane, suggesting intracellular localization in the cytoplasm.

411 **B.** The uptake of [³H]-estradiol-17 β -glucuronide was determined in HEK293 cells overexpressing either a
412 GFP expression vector or SLC22A10 expression vectors containing sequences from various primates
413 including human, chimpanzee, bonobo, gorilla, and orangutan. SLC22A10 orthologs from chimpanzee,
414 bonobo, gorilla, and orangutan expressing the longer isoform (552 amino acids) significantly
415 accumulated [³H]-estradiol-17 β -glucuronide. Please refer to the "Statistical Analysis" section for details
416 on the statistical methods used to determine the significance of each cell transfected with the different
417 SLC22A10 orthologs.

418 **C.** Sequence alignments of the last exon of SLC22A10 in human, chimpanzee, bonobo, gorilla, and
419 orangutan are shown. In humans, the frequency of the A-allele insertion is significantly greater (98%)
420 than in chimpanzees (2.5%) and is not present in available sequences from bonobos, gorillas, or
421 orangutans. The A-allele insertion results in the expression of human SLC22A10 with 541 amino acids,
422 while bonobo, gorilla, orangutan and the majority of chimpanzees are predicted to express isoforms of
423 SLC22A10 with 552 amino acids.

424

425 **Figure 3.** A single mutation of proline to leucine at amino acid position 220 of human SLC22A10
426 significantly enhances the accumulation of [³H]-estradiol-17 β -glucuronide in HEK293.

427 **A.** The amino acid sequence alignment of human SLC22A10 and SLC22A10 from other great apes
428 (chimpanzee, bonobo and gorilla) shows that only the amino acids at positions 18 and 220 differ
429 between the human ortholog and orthologs from great apes. Additionally, there are several amino acid
430 differences starting at position 533.

431 **B.** The uptake of [³H]-estradiol-17 β -glucuronide in HEK293 cells transiently transfected with plasmids
432 encoding human SLC22A10 with reference amino acids or amino acids that are similar to those found in
433 other great apes, namely SLC22A10-p.M18I and SLC22A10-p.P220L. A chimeric protein consisting of
434 the first 533 amino acids of human SLC22A10 and the last 19 amino acids of chimpanzee SLC22A10
435 (534-552) was also evaluated, but did not significantly accumulate [³H]-estradiol-17 β -glucuronide
436 compared to the chimeric protein with p.P220L. The fold uptake of the substrate, relative to the control
437 (GFP) cells, was plotted based on one representative experiment conducted in triplicate wells (mean \pm

438 s.d.). The statistical significance for cells transfected with SLC22A10 #4 (Human SLC22A10 p.P220L (541
439 aa)), #5 (Human SLC22A10 p.P220L (1-533) + Pt SLC22A10 (534 - 552)) and #6 (Chimp SLC22A10 (552
440 aa)) is $p < 0.001$.

441 **C.** This figure shows the plasma membrane localization of SLC22A10 conjugated to green fluorescent
442 protein (GFP) in HEK293 cells. The GFP tag is located at the N-terminus of SLC22A10. Confocal
443 imaging revealed that human SLC22A10-p.P220L localizes primarily to the plasma membrane of the cell,
444 while there was no localization to the plasma membrane in cells expressing a chimeric protein or
445 chimpanzee SLC22A10 with proline at the 220 amino acid position.
446

447 **Figure 4.** SLC22A10 of chimpanzees, bonobos, orangutans, and gibbons are predicted to have shorter
448 isoforms expressing 533 or 538 amino acids. **A.** A comparison of the SLC22A10 amino acid sequence of
449 humans, chimpanzees, bonobos, and orangutans, which express 533 (chimpanzee, orangutan, gibbon),
450 538 (bonobo), 540 (bonobo, chimpanzee), or 541 (human) amino acids, shows that the major differences
451 are at the end of the SLC22A10 sequence. **B.** Confocal imaging revealed that SLC22A10 from
452 chimpanzees and bonobos (isoforms expressing 533 or 538 amino acids) primarily localize to the plasma
453 membrane of the cell, whereas weaker localization was observed for orangutan SLC22A10 (533 amino
454 acids) to the plasma membrane of the cell. GFP conjugated to SLC22A10 was used for this experiment.
455 **C.** The uptake of [³H]-estradiol-17 β -glucuronide in HEK293 cells was observed after transient
456 transfection of plasmids encoding human SLC22A10 with reference amino acids or SLC22A10 with
457 reference amino acids of other great apes with different isoforms. The results showed that SLC22A10
458 isoforms expressing 533 and 552 amino acids significantly accumulate the substrate. However, weaker
459 substrate accumulation was observed in cells transfected with the bonobo SLC22A10 isoform expressing
460 538 amino acids.
461

462 **Figure 5.** This figure presents information about the transport mechanism and kinetics of chimpanzee
463 SLC22A10.

464 **A.** The uptake of seven steroid glucuronides and two steroids in HEK293 cells stably transfected with
465 chimpanzee SLC22A10 (552 amino acids) was measured using LC/MS-MS to determine the
466 accumulation of the compounds.

467 **B.** The effect of pH on accumulation of [³H]-estradiol-17 β -glucuronide in HEK293 cells stably transfected
468 with chimpanzee SLC22A10 isoforms expressing 533 and 552 amino acids was investigated.

469 **C.** The effect of sodium and chloride on accumulation of [³H]-estradiol-17 β -glucuronide in HEK293 cells
470 stably transfected with 533 and 552 amino acid isoforms of chimpanzee SLC22A10 was investigated.

471 **D.** The effects of *trans*-stimulation of [³H]-estradiol-17 β -glucuronide uptake by chimpanzee SLC22A10
472 was determined. Uptake was *trans*-stimulated by preloading the cells with 2 mM of butyrate, glutaric acid,
473 alpha-ketoglutarate, or succinic acid for 2 hours, and then measuring the uptake of [³H]-estradiol-17 β -
474 glucuronide after 15 minutes. The data are presented as mean \pm S.D. and were normalized by setting
475 the uptake of SLC22A10-expressing cells *trans*-stimulated by HBSS to 1.0. *Trans*-stimulation of [³H]-
476 estradiol-17 β -glucuronide by glutaric acid was observed for both isoforms of chimpanzee SLC22A10. **E.**

477 The kinetics of [³H]-estradiol-17 β -glucuronide uptake for chimpanzee SLC22A10 isoforms expressing
478 533 and 552 amino acids were analyzed. The uptake rate was evaluated at 5 minutes and the data were
479 fit to a Michaelis-Menten equation. To fit the kinetic curve to a Michaelis-Menten equation, the
480 concentration of estradiol-17 β -glucuronide is set up to 10 μ M. The figure shows a representative plot
481 from one experiment. All experiments were repeated once, in triplicate and showed similar results.
482
483
484
485

486 **Materials and Methods**

487

488 **Generation of SLC22A10 ortholog cDNA constructs**

489 Sequences of SLC22A10 orthologs from various species, along with their respective sequence IDs, were
490 obtained from ensembl.org (release version 104)³³ and the National Center for Biotechnology Information
491 (NCBI). The gene's cDNA was synthesized using GenScript Gene Synthesis and DNA Synthesis
492 Services and subsequently inserted into the multiple cloning sites on the expression vectors
493 pcDNA3.1(+), pcDNA3.1(+)-eGFP, or pcDNA3.1(+)-N-eGFP. The sequence of the constructs was
494 verified through sequencing conducted by GenScript.

495

496 Human SLC22A10, NM_001039752, NP_001034841.3 (541 aa)
497 Chimpanzee SLC22A10, XM_016921102.2, XP_016776591.1 (552 aa)
498 Chimpanzee SLC22A10, ENSPTRT00000087785.1, A0A2I3TWY5 (540 aa)
499 Chimpanzee SLC22A10, XM_024347215.1, XP_024202983 (533 aa)
500 Bonobo SLC22A10, XM_034932815.1, XP_034788706 (538 aa)
501 Bonobo SLC22A10, ENSPPAT00000058172.1, A0A2R9CA90 (540 aa)
502 Gorilla SLC22A10, XM_019037494.1, XP_018893039 (552 aa)
503 Gorilla SLC22A10, ENSGGOT00000014974.3, G3RG24 (540 aa)
504 Orangutan SLC22A10, XM_024255628.1, XP_024111396.1 (533 aa)
505 Orangutan SLC22A10, XM_054439153.1, XP_054295128.1 (552 aa)
506 Gibbon SLC22A10, XM_003274123.2, XP_003274171.2 (533 aa)
507 Squirrel monkey SLC22A10, ENSBOT00000021267.1, A0A2K6SBP5 (552 aa)

508

509 **Site-directed mutagenesis**

510 Site-directed mutagenesis to create Methionine18 to Isoleucine18 in human SLC22A10, Proline220 to
511 Leucine220 or vice versa in chimpanzee SLC22A10 and human SLC22A10, were performed using Q5
512 Site-Directed Mutagenesis Protocol from NEB (#E0554). NEBaseChanger,
513 <https://nebasechanger.neb.com/>, is used to assist in the design of primers for the site-directed
514 mutagenesis experiment.

515

516 Primers for p.M18I:

517 mut_hsA10_c54C_F: GATTTTCAGATcCTTCATCTGGTTTTTATTCTTC

518 mut_hsA10_c54C_R: TCCCAAGGCCTCCAACCTT

519

520 Primers for p.P220L:

521 mut_hsA10_c659T_F: AATTCTTTGctCATTACTGAG

522 mut_hsA10_c659T_R: ATTTGATATAATGATCATGGAAG

523

524 Primers for p.L220P:

525 mut_ptA10_c659C_F: AATTCTTTGccCATTACTGAGTG

526 mut_ptA10_c659C_R: ATTTGATATAATGATCATGGAAG

527

528 Primers for deletion of A nucleotide in position c1607 and then inserting the tail at the last exon for
529 human to chimpanzee (consist of 34 bp):

530 mut_hsA10_c1607Adel_F: TCTCAAGGAAAAGGCATAAATG

531 mut_hsA10_c1607Adel_R: TTTTTTTTTGATTTTCCACATCC

532 mut_hsA10_c1622ins_F: caaaagtgaccaaattTAActCGAGTCTAGAGGG

533 mut_hsA10_c1622ins_R: tgtagcaatcattttaTGCCTTTTCCTTGAGATTTTTTTTTTG

534

535

536 **Cell culture**

537 Human embryonic kidney cell lines (HEK293) containing a Flp-In expression vector (HEK293 Flp-In)
538 obtained from ThermoFisher Scientific were utilized in this study. These cells have previously been
539 employed for the generation of stable cell lines for transporter assays^{2, 3, 16, 28}. The HEK293-FlpIn cells

540 from ATCC were cultured in DMEM, high glucose (#11965118, ThermoFisher Scientific) supplemented
541 with 10% fetal bovine serum (heat inactivate, #10438026, ThermoFisher Scientific). Penicillin-
542 Streptomycin (#15070063, ThermoFisher Scientific) was added to DMEM media (50 unit/500 mL
543 DMEM). During transfection and when cells were plated for transporter studies, media without
544 penicillin/streptomycin supplementation were used. The cells were regularly screened for mycoplasma
545 contamination (Mycoprobe Mycoplasma Detection Kit, #CUL001B, Fisher).
546

547 **Generation of cells transiently or stably expressing cDNAs**

548 Expression vectors of SLC22A10 orthologs were introduced into HEK293 Flp-In cells either through
549 transient transfection or stable transfection using Lipofectamine LTX (Thermo Fisher Scientific). For
550 transfections in a 48-well plate (seeding density: 1.0×10^5 cells/well), 200 ng of DNA and 0.4 μ L of
551 Lipofectamine LTX were utilized, while for transfections in a 100 mm tissue culture plate (seeding
552 density: 4×10^6 cells/well), 10 μ g of DNA and 44 μ L of Lipofectamine LTX were used. More
553 comprehensive methods for generating transiently or stably transfected cells have been described in our
554 previous work (see reference^{2,3}). In the case of transient transfection, cells were used for transporter
555 studies (refer to the section titled "Transporter uptake studies") after 36-48 hours or for protein
556 quantification after 72 hours. To establish stable cell lines, 3000 ng of DNA (SLC22A10 ortholog
557 expression vectors) and 10.5 μ L of Lipofectamine LTX were employed to transfect HEK293 Flp-In cells
558 seeded in a 6-well plate (seeding density: $7-8 \times 10^5$ cells/well). After 48 hours, cells were transferred to a
559 new 100 mm tissue culture plate and treated with 800 μ g/mL Geneticin. Fresh media containing 800
560 μ g/mL Geneticin was replenished every other day for 1 week. Stable cell lines were utilized for confocal
561 imaging to determine the plasma membrane localization of SLC22A10 orthologs and their various
562 isoforms. Unless specified otherwise, stable cell lines were used for transporter assays.
563

564 **Fluorescence microscopy**

565 For the immunostaining experiments, HEK293 Flp-In stable cell lines expressing different SLC22A10
566 orthologs were cultured on poly-D-lysine-treated 12-well plates with sterile coverslips at a density of
567 200,000 cells per well. After two days of seeding when the cells reached 90-100% confluency, the
568 staining procedure was conducted. On the day of staining, the cell culture media was carefully removed,
569 and the cells were washed using cold Hank's Balanced Salt Solution (HBSS, #14025092, ThermoFisher
570 Scientific). To initiate the staining process, the plasma membrane was first labeled using Wheat Germ
571 Agglutinin (WGA) Alexa Fluor 647 conjugate (Invitrogen Life Sciences Corporation) diluted at a ratio of
572 1:500 in HBSS, followed by a 15-minute incubation at room temperature. Following the staining step, the
573 WGA solution was aspirated, and the cells were washed three times with HBSS. Subsequently, the cells
574 were fixed with a solution of 3.7% formaldehyde in HBSS for 20 minutes. After the fixation step, the cells
575 were washed three times with HBSS. To stain the nucleus, Hoechst solution (ThermoFisher Scientific
576 Inc.) diluted at a ratio of 1:2000 in HBSS was applied to the cells and incubated for 20 minutes at RT, in
577 darkness. After the staining period, the Hoechst solution was aspirated, and the cells were washed twice
578 with HBSS. The coverslips were carefully mounted on Superfrost Plus Microscope Slides (ThermoFisher
579 Scientific) using a small amount of SlowFadeTM Gold Antifade Mountant (#S36940, ThermoFisher
580 Scientific). The mounted slides were left to dry overnight in darkness before being imaged using an
581 inverted Nikon Ti microscope equipped with a CSU-22 spinning disk confocal system available at the
582 Center for Advanced Light Microscopy (CALM) at University of California San Francisco. The image
583 acquisition settings were as follows: DAPI channel with a 300ms exposure time and 50% laser power,
584 FITC channel with a 300ms exposure time and 25% laser power, and CY5 channel with a 100ms
585 exposure time and 5% laser power. Image alignment and merging were performed using Fiji software.
586 This experimental protocol has been previously utilized and described in our published work^{34,35}.
587

588 **Transporter uptake studies**

589 HEK293 Flp-In cells expressing SLC22A10 were seeded at a density of 120,000 to 150,000 cells/0.3 mL
590 in poly-D-lysine-coated 48-well plates approximately 16 to 24 hours prior to conducting uptake studies.
591 The uptake studies for transporters, as detailed below, are methods we have previously described^{2,3}. For
592 transiently expressing SLC22A10, the methods outlined in the previous section pertaining to transient
593 expression in HEK293 Flp-In cells were followed prior to this step. Prior to uptake studies, the culture

594 medium (Dulbecco's modified Eagle's medium, DMEM) supplemented with 10% fetal bovine serum was
595 aspirated, and the cells were incubated in 0.8 mL of Hank's Balanced Salt Solution (HBSS) at 37 °C for
596 10-20 minutes. For screening radiolabeled compounds as SLC22A10 substrates, minute quantities of
597 radiolabeled compounds (³H or ¹⁴C) were diluted in HBSS (at ratios of 1:2000 or 1:3000) for uptake
598 experiments. Unlabeled compounds were added to obtain specific concentrations, which are described in
599 the Results section or figure legends along with the uptake times. Uptake reactions were terminated by
600 washing the cells twice with 0.8 mL of HBSS buffer, followed by incubation in 750 μL of lysis buffer (0.1
601 N NaOH, 0.1% v/v SDS). A 690 μL portion of the cell lysate was transferred to scintillation fluid for
602 scintillation counting. For pH dependence experiments, the HBSS buffer was adjusted to different pH
603 levels (5.5, 7.4, and 8.5) using hydrochloric acid or sodium hydroxide. For sodium and chloride
604 dependence studies, three distinct uptake buffers were employed: (1) chloride-free buffer (composed of
605 125 mM sodium gluconate, 4.8 mM potassium gluconate, 1.2 mM magnesium sulfate, 1.3 mM calcium
606 gluconate, and 5 mM HEPES; adjusted to pH 7.4 with sodium hydroxide); or (2) sodium buffer
607 (composed of 140 mM sodium chloride, 4.73 mM potassium chloride, 1.25 mM calcium chloride, 1.25
608 mM magnesium sulfate, and 5 mM HEPES, adjusted to pH 7.4 with sodium hydroxide); or (3) sodium-
609 free buffer (composed of 140 mM N-methyl-D-glucamine chloride, 1.25 mM magnesium sulfate, 4.73 mM
610 potassium chloride and 1.25 mM calcium chloride), adjusted to pH 7.4 with potassium hydroxide). For
611 trans-stimulation studies, the experimental conditions described in our previously published methods
612 were followed^{2, 3}. In brief, the SLC22A10 or EV stable cell lines were pre-incubated with either buffer or 2
613 mM succinic acid, 2 mM α-ketoglutaric acid, 2 mM butyric acid, or 2 mM glutaric acid for 2 hours.
614 Subsequently, the cells were washed twice with HBSS before commencing the uptake of the anions
615 (estradiol-17β-glucuronide).
616

617 **Kinetic studies of estradiol glucuronide**

618 Kinetic studies of estradiol-17β-glucuronide were conducted in HEK293 Flp-In cells expressing
619 chimpanzee SLC22A10 isoforms (533 amino acid and 552 amino acid) that were stably transfected. The
620 experimental conditions for the kinetic studies closely followed the methods previously published by our
621 research group (reference provided). Initially, we examined the time-dependent uptake of the substrates
622 using trace amounts of the radioactive compound. Concentrations of the non-labeled compounds were
623 varied up to 50 μM. For the kinetic studies, a duration of five minutes at 37°C was chosen as it fell within
624 the linear range observed in the uptake versus time plot for each substrate. Each data point represents
625 the mean ± standard deviation of uptake in the cells transfected with the transporter, subtracted by that in
626 empty vector cells. The obtained data were fitted to a Michaelis-Menten equation to estimate the kinetic
627 parameters. Plots were generated based on a representative experiment out of three independent
628 studies.
629

630 **Protein extraction and global proteomics of HEK293 cells expressing SLC22A10 orthologs**

631 HEK293 Flp-In cells were transfected transiently with various SLC22A10 orthologs, including human,
632 chimpanzee, and the mutations to proline or leucine at position 220. After 72 hours of transfection, cell
633 pellets were collected and shipped to Dr. Per Artursson's laboratory in Uppsala University for protein
634 quantification. The quantification was performed on both HEK293 cells and HEK293 cells expressing the
635 different SLC22A10 orthologs and mutations. HEK293 cell pellets (50–92 mg) were lysed in a lysis buffer
636 containing 50 mM dithiothreitol, 2% sodium dodecyl sulfate in 100 mM Tris/HCl pH 7.8. The lysates were
637 incubated at 95°C for 5 min and sonicated with 20 pulses of 1 second, 20% amplitude by using a
638 sonicator coupled with a microtip probe. The lysates were centrifuged at 14,000×g for 10 min and
639 supernatants were collected. Using LysC and trypsin, the multi-enzyme digestion filter-aided sample
640 preparation (MED-FASP) approach was performed³⁶. C18 stage tips were used to desalt the peptide
641 mixture^{37, 38} and samples were stored at -20°C until analysis. Protein and peptide content were
642 determined by using tryptophan fluorescence assay³⁹. The global proteomics analysis was performed on
643 a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) coupled to a nano-liquid chromatography
644 (nLC). EASY-spray C18-column (50 cm long, 75 μm inner diameter) was used to separate peptides on a
645 ACN/water gradient (with 0.1% formic acid) over 150 min. MS was set to data dependent acquisition with
646 a Top-N method (full MS followed by ddMS2 scans). Proteins were identified using MaxQuant software

647 (version 2.1.0.0)⁴⁰ with the human proteome reference from UniProtKB (October, 2022). Total protein
648 approach was used as the protein quantification method⁴¹.
649

650 **RNA isolation and quantitative RT-PCR**

651 HEK293 Flp-In cells were cultured in poly-D-lysine coated 24-well plates at a seeding density of 1.5-1.8 x
652 10⁵ cells per well, allowing them to reach 75-80% confluency. The RT-PCR method for transcript levels
653 determination as detailed below, are methods we have previously described³. Once the desired
654 confluency was achieved, the cells were transiently transfected with either the vector alone or the vector
655 containing different SLC22A10 orthologs (in the pcDNA3.1(+) expression vector). For the transfection
656 mixture, 500 ng of plasmid DNA, 1 μ L of Lipofectamine LTX (Thermo Fisher Scientific), and 100 μ L of
657 Opti-MEM I reduced serum media (Thermo Fisher Scientific) were used. After 36-48 hours of
658 transfection, the media was removed, and RNA Lysis buffer (350 μ L) was added to each well. Total RNA
659 was isolated from the cells using the Qiagen RNeasy kit (Qiagen). Subsequently, cDNA was synthesized
660 using the SuperScript VILO cDNA Synthesis Kit (ThermoFisher Scientific). For quantitative RT-PCR
661 (qRT-PCR), Taqman reagents and specific primer and probe sets were used, targeting human
662 SLC22A10 (Assay ID: Hs01397962_m1) and beta actin (Assay ID: Hs99999903_m1) (Applied
663 Biosystems, Foster City, CA). The qRT-PCR reactions were conducted in a 96-well plate, with a reaction
664 volume of 10 μ L, using the QuantStudio™ 6 Flex Real-Time PCR System and the default instrument
665 settings. The expression levels were determined using the Ct method, and the data were normalized to
666 the endogenous levels of beta actin. The results are presented as fold-increases in the SLC22A10
667 transcript levels relative to the cell lines expressing the vector control. The analysis was based on three
668 independent biological samples.
669

670 **Cloning of SLC22A10 in pooled human liver**

671 For the cloning process, pooled total RNA samples from human liver were obtained from Clontech. Each
672 sample (2 μ g) of total RNA was reverse transcribed into cDNA using the SuperScript VILO cDNA
673 Synthesis kit (Thermo Fisher Scientific) following the manufacturer's instructions. The primers specified
674 below were employed for PCR amplification of the NM_001039752 transcript: Forward primer:
675 ACCGAGCTCGGATCCATGGCCTTTGAGGAGCTC; and reverse primer:
676 CCCTCTAGACTCGAGTTATGCCTTTTCCTTGAGATT. The nucleotide underlined are open reading
677 frame of SLC22A10. The resulting PCR products were cloned into BamHI and XhoI multiple-cloning site
678 of the pcDNA5FRT vector and subsequently subjected to sequencing at MCLAB in South San Francisco
679 to determine the sequence of the transcript. For the cloning of human SLC22A10, the KOD Xtreme Hot
680 Start DNA polymerase kit (Takara) was utilized. The PCR cycling conditions were as follows: (i) initial
681 activation at 94°C for 2 minutes, (ii) denaturation at 98°C for 10 seconds, (iii) annealing at 57.5°C for 30
682 seconds, and (iv) extension at 68°C for 1 minute.
683

684 **Calculating $\Delta\Delta G$ with the PRISM's rosetta_ddG_pipeline v 0.2.4¹⁴ using Rosetta v 3.15.**

685 In brief, the full-length human SLC22A10-P220 (Uniprot ID Q63ZE4) structure predicted by the
686 AlphaFold⁴² from the AlphaFold DB⁴³ was oriented in the membrane with PPM 3.0 web server⁴⁴ with the
687 default settings and relaxed using the Rosetta's relax protocol⁴⁴. The best of 20 generated structures was
688 used for the $\Delta\Delta G$ calculations with the cartesian_ddG protocol⁴⁵ repeated in 5 replicas. $\Delta\Delta E$ scores were
689 calculated using GEMME⁴⁶ and rank-normalized as in the reference paper¹⁴. The relaxed
690 structure, $\Delta\Delta G$, and $\Delta\Delta E$ results are available at zenodo.
691

692 **Transporter uptake studies and LC/MS/MS analysis**

693 The list of steroid conjugates and their sources can be found in the **Table S1**. Each steroid conjugates
694 was dissolved in DMSO to obtain 20 mM stock solution. Compounds were stored in -20°C freezer.
695 HEK293 Flp-In cells stably transfected with GFP only, chimpanzee SLC22A10 (533 amino acid) and
696 chimpanzee SLC22A10 (552 amino acid) were plated in poly-D-lysine coated 48-well plates at a seeding
697 density of 1.5 x 10⁵ cells per well, allowing them to reach 90-95% confluency after 16-24 hours. Prior to
698 uptake studies, the culture medium (Dulbecco's modified Eagle's medium, DMEM) supplemented with
699 10% fetal bovine serum was aspirated, and the cells were incubated in 0.8 mL of Hank's Balanced Salt
700 Solution (HBSS) at 37 °C for 10-20 minutes. To screen various steroid and steroid conjugate compounds

701 as SLC22A10 substrates, HEK293 Flp-In cells stably expressing GFP or chimpanzee SLC22A10 were
702 incubated with HBSS buffer containing 10 μ M of the respective compounds for 20 minutes. The uptake
703 reactions were terminated by washing the cells twice with 0.8 mL of HBSS buffer, followed by incubation
704 in 400 μ L of methanol. After 30 minutes of shaking at room temperature, 300 μ L of methanol containing
705 the extracted steroid or steroid conjugates from each well were transferred to a 1.5 mL tube and stored
706 at -80°C before quantification using LC/MS/MS analysis.

707
708 Subaliquot of cellular extracts (90 μ L) were spiked with 10 μ L of deuterated 17 β -estradiol glucuronide,
709 mixed by vortexing, and filtered at 0.2 μ m through polyvinyl difluoride membranes (Agilent Technologies,
710 Santa Rosa, CA, USA) by centrifugation and 10,000g. After filtration, samples were enriched with 25nM
711 1-cyclohexyl-3-uriedo-decanoic acid (Sigma-Adrich, St. Lousi MO) as an internal standard. Metabolites
712 were measured using ultra-performance liquid chromatography–electrospray ionization tandem mass
713 spectrometry (UPLC-ESI-MS/MS) on a API 4500 QQQ (Sciex, Framingham, MA) with a scheduled
714 multiple reaction monitoring (MRM) using methods adapted from Ke et.al, 2015⁴⁷. Analytes were
715 separated on a Waters I-Class UPLC-FTN equipped with a 2.1 \times 100 mm i.d., 1.7 μ m Acquity BEH C₁₈
716 column (Waters Co; Milford, MA) held at 50°C. Analytes in 5 μ L injections were separated using water
717 (solvent A) and methanol (solvent B) both containing 2 mM ammonium formate at 400 μ L/min with the
718 following gradient: Initial 40% B to 70% B at 2 min, to 98%B at 3 min, held to 4 min, to 40%B at 5.1 min
719 held to 6 min. Mass spectrometer acquisition parameters and analyte retention times are described in
720 **Table S2**.

721 722 **Sequencing data processing and analyses of SLC22A10 in greater apes**

723 Orthologous genome regions of *SLC22A10* coding sequence in multiple primate species were obtained
724 using UCSC liftOver (default parameters) (Hinrichs et al. 2006) based on hg38 (human), panTro6
725 (chimpanzee), panPan3 (bonobo), gorGor6 (gorilla) and ponAbe3 (orangutan) assemblies. These regions
726 were aligned using MUSCLE (Edgar 2004) and visualized using MView (Brown, Leroy, and Sander 1998).
727 *SLC22A10* exons and coding sequences are based on RefSeq annotations (O’Leary et al. 2016) for human,
728 chimpanzee and gorilla. Despite the long *SLC22A10* isoform not being annotated in RefSeq for bonobo
729 and orangutan, we included these two species in the alignment considering the plausibility of the protein
730 model in the context of their genomes and the high genomic conservation in comparison to chimpanzee
731 and gorilla.

732
733 Allele frequencies from non-human great apes were obtained from available whole-genome sequencing
734 data including 59 chimpanzees, 10 bonobos, 49 gorillas and 16 orangutans⁴⁸⁻⁵¹. All samples were mapped
735 to Hg19. We extracted the genotyping information in position chr11:63078478 (Hg19 coordinates) to
736 calculate the allele frequencies per population. We manually curated the genotypes by checking the raw
737 reads overlapping this region in the BAM files. In chimpanzees we report a frequency of the insertion to be
738 3/108=2.5%; in bonobos is 0/20=0%; in gorillas is 0/98=0%; and in orangutans is 0/32=0%. The global
739 human allele frequencies were obtained from 1000 Genomes Project Phase 3⁵² database in Ensembl for
740 the rs562147200 SNP.

741 742 **SLC22A10 variants in archaic hominins and ancient humans**

743 Archaic hominin genotypes for three Neanderthals and a Denisovan were retrieved from
744 <http://ftp.eva.mpg.de/neandertal/Vindija/VCF/> and <http://ftp.eva.mpg.de/neandertal/Chagyrskaya/VCF/>⁵³⁻⁵⁶.
745 We used the LDproxy tool from LDlink⁵⁷ to identify variants in high LD with rs1790218 in all Thousand
746 Genomes populations and intersected these variants with those assayed in ancient humans from the
747 Allen Ancient DNA Resource (AADR)⁵⁸, retrieved from
748 [https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V54/V54.1.p1/SHARE/public.](https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V54/V54.1.p1/SHARE/public.dir/v54.1.p1_1240K_public.tar)
749 [dir/v54.1.p1_1240K_public.tar](https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V54/V54.1.p1/SHARE/public.dir/v54.1.p1_1240K_public.tar). We identified two such variants: rs1783634 ($D' = 1$, $r_2 = 0.9839$) and
750 rs1201559 ($D' = 0.9975$, $r_2 = 0.9775$). We filtered the AADR genotypes for these SNPs, excluding
751 samples from archaic hominins and individuals that were not genotyped at both loci. We calculated allele

752 frequency in 17 time periods, stratifying by sample location. We also retrieved the allele age estimate for
753 rs1790218 using the Human Genome Dating tool⁵⁹ with all default settings.
754

755 **Statistical Analysis**

756 When comparing the significant differences among HEK293 cells transfected with GFP only and various
757 SLC22A10 ortholog species or other transporters, we performed multiple comparisons using one-way
758 analysis of variance followed by Dunnett's two-tailed test. HEK293 cells transiently transfected with the
759 GFP vector served as the control. The fold uptake of the substrate, relative to the control cells, was
760 plotted based on one representative experiment conducted in triplicate wells (mean \pm s.d.). Statistical
761 significance was indicated as *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$. These findings were further confirmed
762 through at least one or two additional experiments. For specific differences and more detailed
763 information, please refer to the figure legend.
764

765 **Funding/Acknowledgements**

766 This research was supported by NIH GM117163 (SWY) and GM139875 (KMG, SWY). Additional support
767 was provided by USDA Intramural Projects 2032-51530-025-00D (JWN). The USDA is an equal
768 opportunity employer and provider. TMB is supported by funding from the European Research Council
769 (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No.
770 864203), PID2021-126004NB-100 (MICIIN/FEDER, UE) and Secretaria d'Universitats i Recerca and
771 CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC
772 2021 SGR 00177). We would like to acknowledge Esther Lizano, Alba Duch and Mina Azimi for their
773 invaluable assistance and support in the successful execution of our research.
774

775 **References**

- 776 1. Meixner E, Goldmann U, Sedlyarov V, Scorzoni S, Rebsamen M, Girardi E, Superti-Furga G. A
777 substrate-based ontology for human solute carriers. *Mol Syst Biol.* 2020;16(7):e9652. Epub 2020/07/23.
778 doi: 10.15252/msb.20209652. PubMed PMID: 32697042; PMCID: PMC7374931.
- 779 2. Yee SW, Buitrago D, Stecula A, Ngo HX, Chien HC, Zou L, Koleske ML, Giacomini KM.
780 Deorphaning a solute carrier 22 family member, SLC22A15, through functional genomic studies. *FASEB*
781 *J.* 2020;34(12):15734-52. Epub 2020/10/31. doi: 10.1096/fj.202001497R. PubMed PMID: 33124720;
782 PMCID: PMC7839234.
- 783 3. Yee SW, Stecula A, Chien HC, Zou L, Feofanova EV, van Borselen M, Cheung KWK, Yousri NA,
784 Suhre K, Kinchen JM, Boerwinkle E, Irannejad R, Yu B, Giacomini KM. Unraveling the functional role of
785 the orphan solute carrier, SLC22A24 in the transport of steroid conjugates through metabolomic and
786 genome-wide association studies. *PLoS Genet.* 2019;15(9):e1008208. Epub 2019/09/26. doi:
787 10.1371/journal.pgen.1008208. PubMed PMID: 31553721; PMCID: PMC6760779 competing interests
788 exist.
- 789 4. Girardi E, Agrimi G, Goldmann U, Fiume G, Lindinger S, Sedlyarov V, Srndic I, Gurtl B, Agerer B,
790 Kartnig F, Scarcia P, Di Noia MA, Lineiro E, Rebsamen M, Wiedmer T, Bergthaler A, Palmieri L, Superti-
791 Furga G. Epistasis-driven identification of SLC25A51 as a regulator of human mitochondrial NAD import.
792 *Nat Commun.* 2020;11(1):6145. Epub 2020/12/03. doi: 10.1038/s41467-020-19871-x. PubMed PMID:
793 33262325; PMCID: PMC7708531.
- 794 5. Higuchi K, Sugiyama K, Tomabechei R, Kishimoto H, Inoue K. Mammalian monocarboxylate
795 transporter 7 (MCT7/Slc16a6) is a novel facilitative taurine transporter. *J Biol Chem.*
796 2022;298(4):101800. Epub 2022/03/09. doi: 10.1016/j.jbc.2022.101800. PubMed PMID: 35257743;
797 PMCID: PMC8980330.
- 798 6. Yee SW, Giacomini KM. Emerging Roles of the Human Solute Carrier 22 Family. *Drug Metab*
799 *Dispos.* 2021;50(9):1193-210. Epub 2021/12/19. doi: 10.1124/dmd.121.000702. PubMed PMID:
800 34921098; PMCID: PMC9488978.
- 801 7. Ren T, Jones RS, Morris ME. Untargeted metabolomics identifies the potential role of
802 monocarboxylate transporter 6 (MCT6/SLC16A5) in lipid and amino acid metabolism pathways.
803 *Pharmacol Res Perspect.* 2022;10(3):e00944. Epub 2022/04/26. doi: 10.1002/prp2.944. PubMed PMID:
804 35466588; PMCID: PMC9035569.

- 805 8. Lindberg FA, Nordenankar K, Forsberg EC, Fredriksson R. SLC38A10 Deficiency in Mice Affects
806 Plasma Levels of Threonine and Histidine in Males but Not in Females: A Preliminary Characterization
807 Study of SLC38A10(-/-) Mice. *Genes (Basel)*. 2023;14(4). Epub 2023/04/28. doi:
808 10.3390/genes14040835. PubMed PMID: 37107593; PMCID: PMC10138244.
- 809 9. Giacomini KM, Yee SW, Koleske ML, Zou L, Matsson P, Chen EC, Kroetz DL, Miller MA,
810 Gozalpour E, Chu X. New and Emerging Research on Solute Carrier and ATP Binding Cassette
811 Transporters in Drug Discovery and Development: Outlook From the International Transporter
812 Consortium. *Clin Pharmacol Ther*. 2022;112(3):540-61. Epub 2022/05/01. doi: 10.1002/cpt.2627.
813 PubMed PMID: 35488474; PMCID: PMC9398938.
- 814 10. Sun W, Wu RR, van Poelje PD, Erion MD. Isolation of a family of organic anion transporters from
815 human liver and kidney. *Biochem Biophys Res Commun*. 2001;283(2):417-22. Epub 2001/05/01. doi:
816 10.1006/bbrc.2001.4774. PubMed PMID: 11327718.
- 817 11. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf
818 C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szgyarto CA, Odeberg J,
819 Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J,
820 Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M,
821 von Heijne G, Nielsen J, Ponten F. Proteomics. Tissue-based map of the human proteome. *Science*.
822 2015;347(6220):1260419. Epub 2015/01/24. doi: 10.1126/science.1260419. PubMed PMID: 25613900.
- 823 12. Ruiz-Orera J, Hernandez-Rodriguez J, Chiva C, Sabido E, Kondova I, Bontrop R, Marques-Bonet
824 T, Alba MM. Origins of De Novo Genes in Human and Chimpanzee. *PLoS Genet*.
825 2015;11(12):e1005721. Epub 2016/01/01. doi: 10.1371/journal.pgen.1005721. PubMed PMID:
826 26720152; PMCID: PMC4697840.
- 827 13. Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, Imada EL, Zhang D, Joseph L,
828 Leek JT, Jaffe AE, Nellore A, Collado-Torres L, Hansen KD, Langmead B. recount3: summaries and
829 queries for large-scale RNA-seq expression and splicing. *Genome Biol*. 2021;22(1):323. Epub
830 2021/12/01. doi: 10.1186/s13059-021-02533-6. PubMed PMID: 34844637; PMCID: PMC8628444.
- 831 14. Tiemann JKS, Zschach H, Lindorff-Larsen K, Stein A. Interpreting the molecular mechanisms of
832 disease variants in human transmembrane proteins. *Biophys J*. 2023;122(11):2176-91. Epub
833 2023/01/06. doi: 10.1016/j.bpj.2022.12.031. PubMed PMID: 36600598; PMCID: PMC10257119.
- 834 15. Cagiada M, Johansson KE, Valanciute A, Nielsen SV, Hartmann-Petersen R, Yang JJ, Fowler
835 DM, Stein A, Lindorff-Larsen K. Understanding the Origins of Loss of Protein Function by Analyzing the
836 Effects of Thousands of Variants on Activity and Abundance. *Mol Biol Evol*. 2021;38(8):3235-46. Epub
837 2021/03/30. doi: 10.1093/molbev/msab095. PubMed PMID: 33779753; PMCID: PMC8321532.
- 838 16. Zou L, Stecula A, Gupta A, Prasad B, Chien HC, Yee SW, Wang L, Unadkat JD, Stahl SH,
839 Fenner KS, Giacomini KM. Molecular Mechanisms for Species Differences in Organic Anion Transporter
840 1, OAT1: Implications for Renal Drug Toxicity. *Mol Pharmacol*. 2018;94(1):689-99. Epub 2018/05/04. doi:
841 10.1124/mol.117.111153. PubMed PMID: 29720497; PMCID: PMC5987998.
- 842 17. Dvorak V, Wiedmer T, Ingles-Prieto A, Altermatt P, Batoulis H, Barenz F, Bender E, Digles D,
843 Durrenberger F, Heitman LH, AP IJ, Kell DB, Kicking S, Korzo D, Leippe P, Licher T, Manolova V,
844 Rizzetto R, Sassone F, Scarabottolo L, Schlessinger A, Schneider V, Sijben HJ, Steck AL, Sundstrom H,
845 Tremolada S, Wilhelm M, Wright Muelas M, Zindel D, Steppan CM, Superti-Furga G. An Overview of
846 Cell-Based Assay Platforms for the Solute Carrier Family of Transporters. *Front Pharmacol*.
847 2021;12:722889. Epub 2021/08/28. doi: 10.3389/fphar.2021.722889. PubMed PMID: 34447313; PMCID:
848 PMC8383457.
- 849 18. Kuang W, Zhang J, Lan Z, Deepak R, Liu C, Ma Z, Cheng L, Zhao X, Meng X, Wang W, Wang X,
850 Xu L, Jiao Y, Luo Q, Meng Z, Kee K, Liu X, Deng H, Li W, Fan H, Chen L. SLC22A14 is a mitochondrial
851 riboflavin transporter required for sperm oxidative phosphorylation and male fertility. *Cell Rep*.
852 2021;35(3):109025. Epub 2021/04/22. doi: 10.1016/j.celrep.2021.109025. PubMed PMID: 33882315;
853 PMCID: PMC8065176.
- 854 19. Wegler C, Wisniewski JR, Robertsen I, Christensen H, Kristoffer Hertel J, Hjelmesaeth J,
855 Jansson-Lofmark R, Asberg A, Andersson TB, Artursson P. Drug Disposition Protein Quantification in
856 Matched Human Jejunum and Liver From Donors With Obesity. *Clin Pharmacol Ther*. 2022;111(5):1142-
857 54. Epub 2022/02/15. doi: 10.1002/cpt.2558. PubMed PMID: 35158408; PMCID: PMC9310776.

- 858 20. Wegler C, Matsson P, Krogstad V, Urdzik J, Christensen H, Andersson TB, Artursson P.
859 Influence of Proteome Profiles and Intracellular Drug Exposure on Differences in CYP Activity in Donor-
860 Matched Human Liver Microsomes and Hepatocytes. *Mol Pharm*. 2021;18(4):1792-805. Epub
861 2021/03/20. doi: 10.1021/acs.molpharmaceut.1c00053. PubMed PMID: 33739838; PMCID:
862 PMC8041379.
- 863 21. Niu L, Geyer PE, Gupta R, Santos A, Meier F, Doll S, Wewer Albrechtsen NJ, Klein S, Ortiz C,
864 Uschner FE, Schierwagen R, Trebicka J, Mann M. Dynamic human liver proteome atlas reveals
865 functional insights into disease pathways. *Mol Syst Biol*. 2022;18(5):e10947. Epub 2022/05/18. doi:
866 10.15252/msb.202210947. PubMed PMID: 35579278; PMCID: PMC9112488.
- 867 22. Eide Kvitne K, Hole K, Krogstad V, Wollmann BM, Wegler C, Johnson LK, Hertel JK, Artursson P,
868 Karlsson C, Andersson S, Andersson TB, Sandbu R, Hjelmesaeth J, Skovlund E, Christensen H,
869 Jansson-Lofmark R, Asberg A, Molden E, Robertsen I. Correlations between 4beta-hydroxycholesterol
870 and hepatic and intestinal CYP3A4: protein expression, microsomal ex vivo activity, and in vivo activity in
871 patients with a wide body weight range. *Eur J Clin Pharmacol*. 2022;78(8):1289-99. Epub 2022/06/02.
872 doi: 10.1007/s00228-022-03336-9. PubMed PMID: 35648149; PMCID: PMC9283167.
- 873 23. El-Khateeb E, Al-Majdoub ZM, Rostami-Hodjegan A, Barber J, Achour B. Proteomic
874 Quantification of Changes in Abundance of Drug-Metabolizing Enzymes and Drug Transporters in
875 Human Liver Cirrhosis: Different Methods, Similar Outcomes. *Drug Metab Dispos*. 2021;49(8):610-8.
876 Epub 2021/05/29. doi: 10.1124/dmd.121.000484. PubMed PMID: 34045218.
- 877 24. Cordes FS, Bright JN, Sansom MS. Proline-induced distortions of transmembrane helices. *J Mol*
878 *Biol*. 2002;323(5):951-60. Epub 2002/11/06. doi: 10.1016/s0022-2836(02)01006-9. PubMed PMID:
879 12417206.
- 880 25. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM,
881 Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M,
882 England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A,
883 Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel
884 EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM,
885 Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C,
886 Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K,
887 Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur
888 DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*.
889 2020;581(7809):434-43. Epub 2020/05/29. doi: 10.1038/s41586-020-2308-7. PubMed PMID: 32461654;
890 PMCID: PMC7334197 from Takeda Pharmaceutical Company. A.H.O'D.-L. has received honoraria from
891 ARUP and Chan Zuckerberg Initiative. E.V.M. has received research support in the form of charitable
892 contributions from Charles River Laboratories and Ionis Pharmaceuticals, and has consulted for
893 Deerfield Management. J.S.W. is a consultant for MyoKardia. B.M.N. is a member of the scientific
894 advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical, and
895 Biogen. M.J.D. is a founder of Maze Therapeutics. D.G.M. is a founder with equity in Goldfinch Bio, and
896 has received research support from AbbVie, Astellas, Biogen, BioMarin, Eisai, Merck, Pfizer, and Sanofi-
897 Genzyme. The views expressed in this article are those of the author(s) and not necessarily those of the
898 NHS, the NIHR, or the Department of Health. M.I.M. has served on advisory panels for Pfizer,
899 NovoNordisk, Zoe Global; has received honoraria from Merck, Pfizer, NovoNordisk and Eli Lilly; has
900 stock options in Zoe Global and has received research funding from Abbvie, Astra Zeneca, Boehringer
901 Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier & Takeda. As of
902 June 2019, M.I.M. is an employee of Genentech, and holds stock in Roche. N.R. is a non-executive
903 director of AstraZeneca.
- 904 26. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A,
905 Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee SB, Tian X, Browning BL, Das S, Emde AK,
906 Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM,
907 Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick
908 L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA,
909 Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen YI, Cho MH, Choi SH, Chung MK,
910 Clish CB, Correa A, Curran JE, Custer B, Darbar D, Daya M, de Andrade M, DeMeo DL, Dutcher SK,
911 Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C,

912 Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR,
913 Irvin MR, Johnsen JM, Johnson AD, Kaplan R, Kardia SLR, Kelly T, Kelly S, Kenny EE, Kiel DP,
914 Klemmer R, Konkle BA, Kooperberg C, Kottgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin KH, Liu C,
915 Loos RJJ, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK,
916 Mathias RA, McManus DD, McGarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD,
917 Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson
918 SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS,
919 Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S,
920 Schwartz DA, Seo JS, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith JA,
921 Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ,
922 Vasani RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng LC, Willer CJ, Zhang Y,
923 Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS,
924 Silverman EK, Qasba P, Gan W, Consortium NT-OfPM, Papanicolaou GJ, Nickerson DA, Browning SR,
925 Zody MC, Zollner S, Wilson JG, Cupples LA, Laurie CC, Jaquish CE, Hernandez RD, O'Connor TD,
926 Abecasis GR. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*.
927 2021;590(7845):290-9. Epub 2021/02/12. doi: 10.1038/s41586-021-03205-y. PubMed PMID: 33568819;
928 PMID: PMC7875770.

929 27. Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, Gabriel SB, Belkadi A, Boisson B, Abel
930 L, Clark AG, Greater Middle East Variome C, Alkuraya FS, Casanova JL, Gleeson JG. Characterization
931 of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet*.
932 2016;48(9):1071-6. Epub 2016/07/19. doi: 10.1038/ng.3592. PubMed PMID: 27428751; PMID:
933 PMC5019950.

934 28. Cropp CD, Komori T, Shima JE, Urban TJ, Yee SW, More SS, Giacomini KM. Organic anion
935 transporter 2 (SLC22A7) is a facilitative transporter of cGMP. *Mol Pharmacol*. 2008;73(4):1151-8. Epub
936 2008/01/25. doi: 10.1124/mol.107.043117. PubMed PMID: 18216183; PMID: PMC2698938.

937 29. Zheng D, Gerstein MB. The ambiguous boundary between genes and pseudogenes: the dead
938 rise up, or do they? *Trends Genet*. 2007;23(5):219-24. Epub 2007/03/27. doi: 10.1016/j.tig.2007.03.003.
939 PubMed PMID: 17382428.

940 30. Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. *FEBS Lett*.
941 2000;468(2-3):109-14. Epub 2000/02/29. doi: 10.1016/s0014-5793(00)01199-6. PubMed PMID:
942 10692568.

943 31. Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. Identification and analysis of unitary
944 pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol*.
945 2010;11(3):R26. Epub 2010/03/10. doi: 10.1186/gb-2010-11-3-r26. PubMed PMID: 20210993; PMID:
946 PMC2864566.

947 32. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, Bruford EA.
948 Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res*. 2023;51(D1):D1003-D9. Epub
949 2022/10/17. doi: 10.1093/nar/gkac888. PubMed PMID: 36243972; PMID: PMC9825485.

950 33. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye
951 O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M,
952 Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T,
953 Fatima R, Giron CG, Genev T, Martinez JG, Gujjarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T,
954 Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugan JC, Mohanan S, Mushtaq
955 A, Naven M, Ogeh DN, Parker A, Parton A, Perry M, Pilizota I, Prosovetskaia I, Sakthivel MP, Salam
956 AIA, Schmitt BM, Schuilenburg H, Sheppard D, Perez-Silva JG, Stark W, Steed E, Sutinen K,
957 Sukumaran R, Sumathipala D, Suner MM, Szpak M, Thormann A, Tricoli FF, Urbina-Gomez D,
958 Veidenberg A, Walsh TA, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish
959 A, Giorgetti S, Haggerty L, Hunt SE, GR II, Loveland JE, Martin FJ, Moore B, Mudge JM, Muffato M,
960 Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Dyer S, Harrison PW, Howe KL, Yates AD, Zerbino
961 DR, Flicek P. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):D988-D95. Epub 2021/11/19. doi:
962 10.1093/nar/gkab1049. PubMed PMID: 34791404; PMID: PMC8728283.

963 34. Koleske ML, McInnes G, Brown JEH, Thomas N, Hutchinson K, Chin MY, Koehl A, Arkin MR,
964 Schlessinger A, Gallagher RC, Song YS, Altman RB, Giacomini KM. Functional genomics of OCTN2
965 variants informs protein-specific variant effect predictor for Carnitine Transporter Deficiency. *Proc Natl*

966 Acad Sci U S A. 2022;119(46):e2210247119. Epub 2022/11/08. doi: 10.1073/pnas.2210247119.
967 PubMed PMID: 36343260; PMCID: PMC9674959.

968 35. Azimi M, Yee SW, Riselli A, Silva DB, Giacomini CP, Giacomini KM, Brett CM. Characterization
969 of P-glycoprotein orthologs from human, sheep, pig, dog, and cat. *J Vet Pharmacol Ther.* 2023. Epub
970 2023/05/18. doi: 10.1111/jvp.13386. PubMed PMID: 37198956.

971 36. Wisniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth
972 of proteomic and phosphoproteomic analysis. *Anal Chem.* 2012;84(6):2631-7. Epub 2012/02/14. doi:
973 10.1021/ac300006b. PubMed PMID: 22324799.

974 37. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser
975 desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem.*
976 2003;75(3):663-70. Epub 2003/02/15. doi: 10.1021/ac026117i. PubMed PMID: 12585499.

977 38. Wiśniewski JR. Label-Free Quantitative Analysis of Mitochondrial Proteomes Using the
978 Multienzyme Digestion-Filter Aided Sample Preparation (MED-FASP) and "Total Protein Approach". In:
979 Mokranjac D, Perocchi F, editors. *Mitochondria Methods in Molecular Biology.* New York: Humana Press;
980 2017.

981 39. Wisniewski JR, Gaugaz FZ. Fast and sensitive total protein and Peptide assays for proteomic
982 analysis. *Anal Chem.* 2015;87(8):4110-6. Epub 2015/04/04. doi: 10.1021/ac504689z. PubMed PMID:
983 25837572.

984 40. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based
985 shotgun proteomics. *Nat Protoc.* 2016;11(12):2301-19. Epub 2016/11/04. doi: 10.1038/nprot.2016.136.
986 PubMed PMID: 27809316.

987 41. Wisniewski JR, Rakus D. Multi-enzyme digestion FASP and the 'Total Protein Approach'-based
988 absolute quantification of the *Escherichia coli* proteome. *J Proteomics.* 2014;109:322-31. Epub
989 2014/07/27. doi: 10.1016/j.jprot.2014.07.012. PubMed PMID: 25063446.

990 42. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates
991 R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B,
992 Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M,
993 Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P,
994 Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-9.
995 Epub 2021/07/16. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PMCID: PMC8371605
996 have filed non-provisional patent applications 16/701,070 and PCT/EP2020/084238, and provisional
997 patent applications 63/107,362, 63/118,917, 63/118,918, 63/118,921 and 63/118,919, each in the name
998 of DeepMind Technologies Limited, each pending, relating to machine learning for predicting protein
999 structures. The other authors declare no competing interests.

000 43. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood
001 G, Laydon A, Zidek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A,
002 Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S.
003 AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence
004 space with high-accuracy models. *Nucleic Acids Res.* 2022;50(D1):D439-D44. Epub 2021/11/19. doi:
005 10.1093/nar/gkab1061. PubMed PMID: 34791371; PMCID: PMC8728224.

006 44. Lomize AL, Todd SC, Pogozheva ID. Spatial arrangement of proteins in planar and curved
007 membranes by PPM 3.0. *Protein Sci.* 2022;31(1):209-20. Epub 2021/10/31. doi: 10.1002/pro.4219.
008 PubMed PMID: 34716622; PMCID: PMC8740824.

009 45. Frenz B, Lewis SM, King I, DiMaio F, Park H, Song Y. Prediction of Protein Mutational Free
010 Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. *Front Bioeng*
011 *Biotechnol.* 2020;8:558247. Epub 2020/11/03. doi: 10.3389/fbioe.2020.558247. PubMed PMID:
012 33134287; PMCID: PMC7579412.

013 46. Laine E, Karami Y, Carbone A. GEMME: A Simple and Fast Global Epistatic Model Predicting
014 Mutational Effects. *Mol Biol Evol.* 2019;36(11):2604-19. Epub 2019/08/14. doi: 10.1093/molbev/msz179.
015 PubMed PMID: 31406981; PMCID: PMC6805226.

016 47. Ke Y, Gonthier R, Isabelle M, Bertin J, Simard JN, Dury AY, Labrie F. A rapid and sensitive
017 UPLC-MS/MS method for the simultaneous quantification of serum androsterone glucuronide,
018 etiocholanolone glucuronide, and androstan-3alpha, 17beta diol 17-glucuronide in postmenopausal

019 women. *J Steroid Biochem Mol Biol.* 2015;149:146-52. Epub 2015/02/24. doi:
020 10.1016/j.jsbmb.2015.02.009. PubMed PMID: 25701608.

021 48. Kuderna LFK, Gao H, Janiak MC, Kuhlwilm M, Orkin JD, Bataillon T, Manu S, Valenzuela A,
022 Bergman J, Rousselle M, Silva FE, Agueda L, Blanc J, Gut M, de Vries D, Goodhead I, Harris RA,
023 Raveendran M, Jensen A, Chuma IS, Horvath JE, Hvilsom C, Juan D, Frandsen P, Schraiber JG, de
024 Melo FR, Bertuol F, Byrne H, Sampaio I, Farias I, Valsecchi J, Messias M, da Silva MNF, Trivedi M,
025 Rossi R, Hrbek T, Andriaholinirina N, Rabarivola CJ, Zaramody A, Jolly CJ, Phillips-Conroy J, Wilkerson
026 G, Abee C, Simmons JH, Fernandez-Duque E, Kanthaswamy S, Shiferaw F, Wu D, Zhou L, Shao Y,
027 Zhang G, Keyyu JD, Knauf S, Le MD, Lizano E, Merker S, Navarro A, Nadler T, Khor CC, Lee J, Tan P,
028 Lim WK, Kitchener AC, Zinner D, Gut I, Melin AD, Guschanski K, Schierup MH, Beck RMD, Umapathy
029 G, Roos C, Boubli JP, Rogers J, Farh KK, Marques Bonet T. A global catalog of whole-genome diversity
030 from 233 primate species. *Science.* 2023;380(6648):906-13. Epub 2023/06/01 23:42. doi:
031 10.1126/science.abn7829. PubMed PMID: 37262161.

032 49. de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-
033 Rodriguez J, Dupanloup I, Lao O, Hallast P, Schmidt JM, Heredia-Genestar JM, Benazzo A, Barbujani
034 G, Peter BM, Kuderna LF, Casals F, Angedakin S, Arandjelovic M, Boesch C, Kuhl H, Vigilant L,
035 Langergraber K, Novembre J, Gut M, Gut I, Navarro A, Carlsen F, Andres AM, Siegismund HR, Scally A,
036 Excoffier L, Tyler-Smith C, Castellano S, Xue Y, Hvilsom C, Marques-Bonet T. Chimpanzee genomic
037 diversity reveals ancient admixture with bonobos. *Science.* 2016;354(6311):477-81. Epub 2016/10/30.
038 doi: 10.1126/science.aag2602. PubMed PMID: 27789843; PMCID: PMC5546212.

039 50. Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T, Groves C,
040 Pybus M, Sonay TB, Roos C, Lameira AR, Wich SA, Askew J, Davila-Ross M, Fredriksson G, de Valles
041 G, Casals F, Prado-Martinez J, Goossens B, Verschoor EJ, Warren KS, Singleton I, Marques DA,
042 Pamungkas J, Perwitasari-Farajallah D, Rianti P, Tuuga A, Gut IG, Gut M, Orozco-terWengel P, van
043 Schaik CP, Bertranpetit J, Anisimova M, Scally A, Marques-Bonet T, Meijaard E, Krutzen M.
044 Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Curr Biol.*
045 2017;27(22):3576-7. Epub 2017/11/22. doi: 10.1016/j.cub.2017.11.020. PubMed PMID: 29161551.

046 51. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR,
047 Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth
048 A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prufer K, Pybus M, Johnstone L,
049 Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M,
050 Wilson ML, Stevison L, Camprubi C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I,
051 Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P,
052 Comas D, Siegismund H, Blanc J, Agueda-Calpena L, Gut M, Fulton L, Tishkoff SA, Mullikin JC, Wilson
053 RK, Gut IG, Gonder MK, Ryder OA, Hahn BH, Navarro A, Akey JM, Bertranpetit J, Reich D, Mailund T,
054 Schierup MH, Hvilsom C, Andres AM, Wall JD, Bustamante CD, Hammer MF, Eichler EE, Marques-
055 Bonet T. Great ape genetic diversity and population history. *Nature.* 2013;499(7459):471-5. Epub
056 2013/07/05. doi: 10.1038/nature12228. PubMed PMID: 23823723; PMCID: PMC3822165.

057 52. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO,
058 Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation.
059 *Nature.* 2015;526(7571):68-74. Epub 2015/10/04. doi: 10.1038/nature15393. PubMed PMID: 26432245;
060 PMCID: PMC4750478.

061 53. Mafessoni F, Grote S, de Filippo C, Slon V, Kolobova KA, Viola B, Markin SV, Chintalapati M,
062 Peyregne S, Skov L, Skoglund P, Krivoschapkin AI, Derevianko AP, Meyer M, Kelso J, Peter B, Prufer K,
063 Paabo S. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc Natl Acad Sci U S A.*
064 2020;117(26):15132-6. Epub 2020/06/18. doi: 10.1073/pnas.2004944117. PubMed PMID: 32546518;
065 PMCID: PMC7334501.

066 54. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de
067 Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K,
068 Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP,
069 Patterson N, Andres AM, Eichler EE, Slatkin M, Reich D, Kelso J, Paabo S. A high-coverage genome
070 sequence from an archaic Denisovan individual. *Science.* 2012;338(6104):222-6. Epub 2012/09/01. doi:
071 10.1126/science.1224344. PubMed PMID: 22936568; PMCID: PMC3617501.

072 55. Prufer K, de Filippo C, Grote S, Mafessoni F, Korlevic P, Hajdinjak M, Vernot B, Skov L, Hsieh P,
073 Peyregne S, Reher D, Hopfe C, Nagel S, Maricic T, Fu Q, Theunert C, Rogers R, Skoglund P,
074 Chintalapati M, Dannemann M, Nelson BJ, Key FM, Rudan P, Kucan Z, Gusic I, Golovanova LV,
075 Doronichev VB, Patterson N, Reich D, Eichler EE, Slatkin M, Schierup MH, Andres AM, Kelso J, Meyer
076 M, Paabo S. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*.
077 2017;358(6363):655-8. Epub 2017/10/07. doi: 10.1126/science.aao1887. PubMed PMID: 28982794;
078 PMCID: PMC6185897.

079 56. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G,
080 Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M,
081 Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH,
082 Green RE, Hellmann I, Johnson PL, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES,
083 Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D,
084 Kelso J, Paabo S. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*.
085 2014;505(7481):43-9. Epub 2013/12/20. doi: 10.1038/nature12886. PubMed PMID: 24352235; PMCID:
086 PMC4031459.

087 57. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific
088 haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*.
089 2015;31(21):3555-7. Epub 2015/07/04. doi: 10.1093/bioinformatics/btv402. PubMed PMID: 26139635;
090 PMCID: PMC4626747.

091 58. Mallick S, Micco A, Mah M, Ringbauer H, Lazaridis I, Olalde I, Patterson N, Reich D. The Allen
092 Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. *bioRxiv*. 2023.
093 Epub 2023/04/18. doi: 10.1101/2023.04.06.535797. PubMed PMID: 37066305; PMCID: PMC10104067.

094 59. Albers PK, McVean G. Dating genomic variants and shared ancestry in population-scale
095 sequencing data. *PLoS Biol*. 2020;18(1):e3000586. Epub 2020/01/18. doi:
096 10.1371/journal.pbio.3000586. PubMed PMID: 31951611; PMCID: PMC6992231 following competing
097 interests: GM is a shareholder in and non-executive director of Genomics PLC, and is a partner in
098 Peptide Groove LLP. PKA is a shareholder in and a director of BioMe Oxford Ltd.
099

Figures

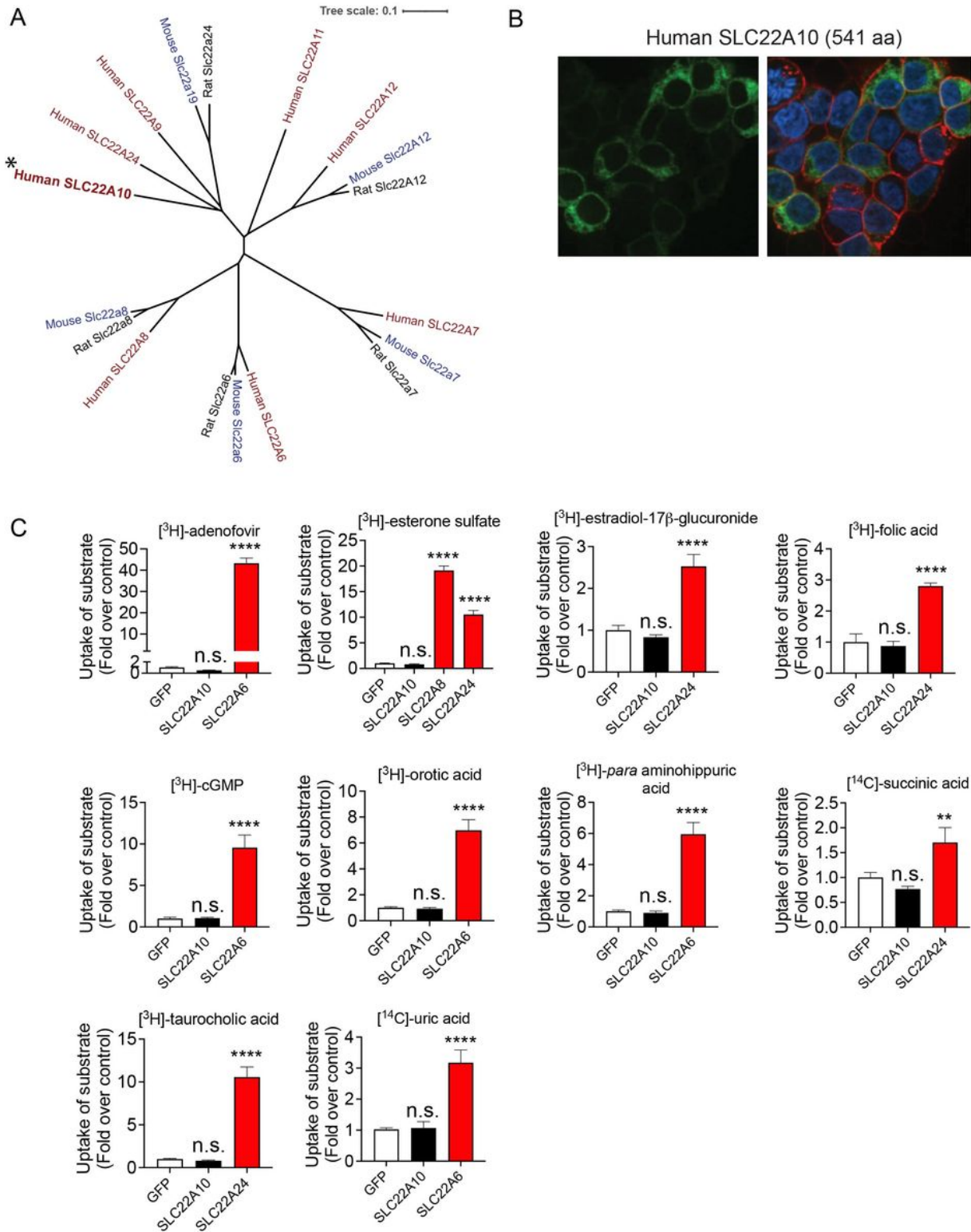


Figure 1

Analysis of the phylogenetic tree, plasma membrane expression of SLC22A10, and uptake of organic anion substrates of the human SLC22 family.

A. Multiple sequence alignments were performed with reference amino acid sequences for each anion transporter from humans and rodents, using the Clustal Omega Multiple Sequence Alignment program (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The dendrogram was generated from the output of the Clustal Omega alignment.

B. Localization of human SLC22A10 conjugated to green fluorescent protein (GFP) was examined in HEK293 cells using high-content imaging and cellular staining with the plasma membrane marker wheat germ agglutinin (WGA). The results showed no colocalization of GFP-tagged SLC22A10 with WGA.

C. Uptake of various radiolabeled organic anions, which are typical substrates of organic anion transporters in the SLC22A family, was assessed. Uptake was performed 48 hours after transient transfection of plasmids encoding human SLC22A10, GFP expression vector, and one other member in the SLC22A family as a positive control. Accumulation of substrates inside cells was determined after 15 minutes. Figure shows a representative plot from one experiment (mean \pm S.D. from three replicate wells). The experiments were repeated at least one time and showed similar results. Multiple comparisons using one-way analysis of variance followed by Dunnett's two-tailed test were performed. HEK293 cells transiently transfected with the GFP vector served as the control. The fold uptake of the substrate, relative to the control cells, was plotted based on one representative experiment conducted in triplicate wells (mean \pm s.d.). The statistical significance for all the cells transfected with organic anion transporters SLC22A6, SLC22A8, or SLC22A24 is $p < 0.001$.

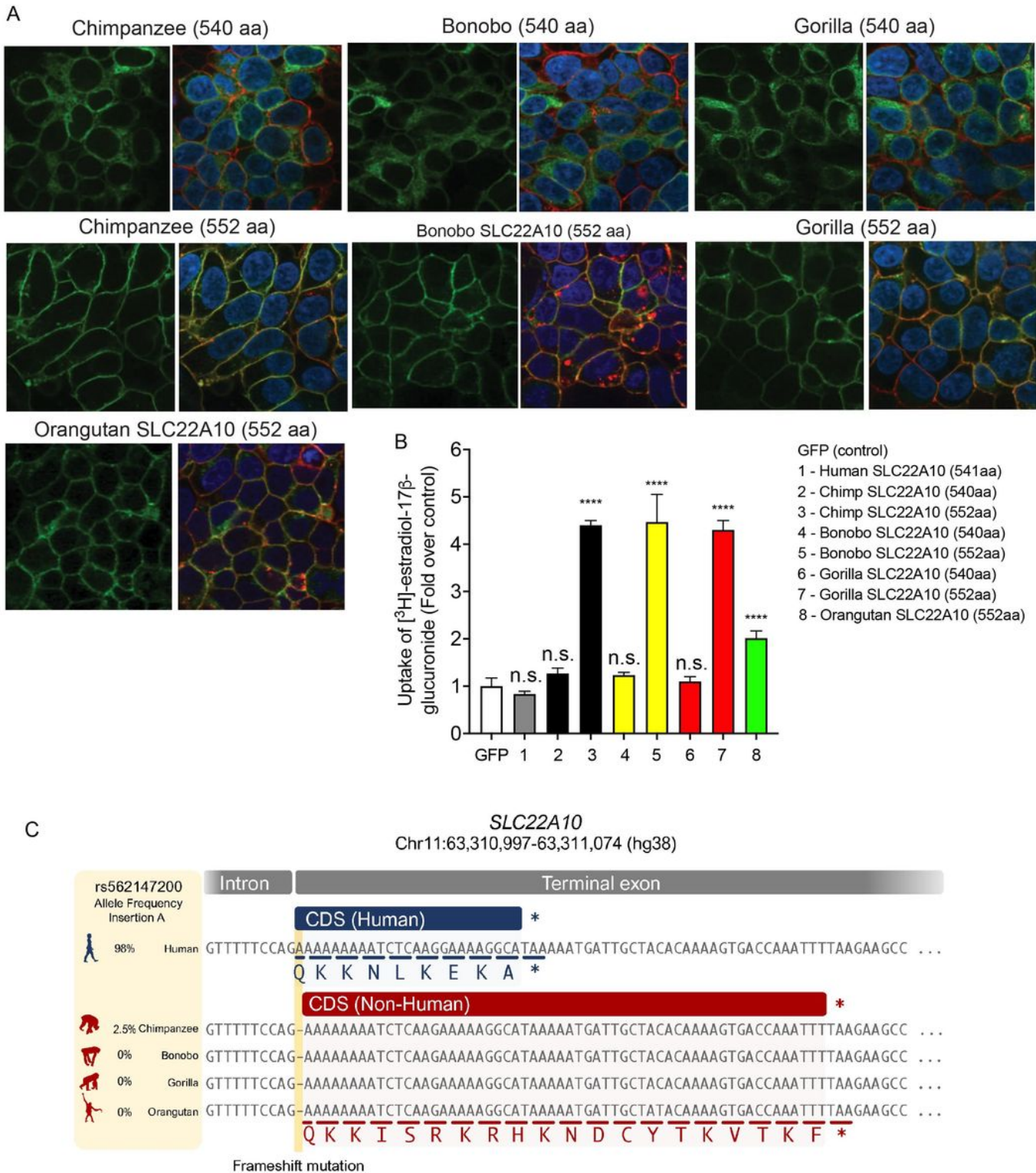


Figure 2

Localization to the plasma membrane, uptake, and sequence comparison of human SLC22A10 were examined in comparison with SLC22A10 from great apes (chimpanzee, bonobo, gorilla and orangutan).

A. This figure shows the plasma membrane localization of SLC22A10 orthologs from great apes, which were conjugated to green fluorescent protein (GFP) in HEK293 cells. The GFP tag is located at the N-

terminus of SLC22A10. Confocal imaging revealed that the 552 amino acid isoforms of SLC22A10 from chimpanzee, bonobo, gorilla, and orangutan primarily colocalized with wheat germ agglutinin (WGA) on the plasma membrane of the cell. In contrast, the 540 amino acid isoform of SLC22A10 from bonobo, chimpanzee, and gorilla showed no colocalization of GFP-tagged SLC22A10 with WGA on the plasma membrane, suggesting intracellular localization in the cytoplasm.

B. The uptake of [³H]-estradiol-17 β -glucuronide was determined in HEK293 cells overexpressing either a GFP expression vector or SLC22A10 expression vectors containing sequences from various primates including human, chimpanzee, bonobo, gorilla, and orangutan. SLC22A10 orthologs from chimpanzee, bonobo, gorilla, and orangutan expressing the longer isoform (552 amino acids) significantly accumulated [³H]-estradiol-17 β -glucuronide. Please refer to the "Statistical Analysis" section for details on the statistical methods used to determine the significance of each cell transfected with the different SLC22A10 orthologs.

C. Sequence alignments of the last exon of SLC22A10 in human, chimpanzee, bonobo, gorilla, and orangutan are shown. In humans, the frequency of the A-allele insertion is significantly greater (98%) than in chimpanzees (2.5%) and is not present in available sequences from bonobos, gorillas, or orangutans. The A-allele insertion results in the expression of human SLC22A10 with 541 amino acids, while bonobo, gorilla, orangutan and the majority of chimpanzees are predicted to express isoforms of SLC22A10 with 552 amino acids.

C. This figure shows the plasma membrane localization of SLC22A10 conjugated to green fluorescent protein (GFP) in HEK293 cells. The GFP tag is located at the N-terminus of SLC22A10. Confocal imaging revealed that human SLC22A10-p.P220L localizes primarily to the plasma membrane of the cell, while there was no localization to the plasma membrane in cells expressing a chimeric protein or chimpanzee SLC22A10 with proline at the 220 amino acid position.

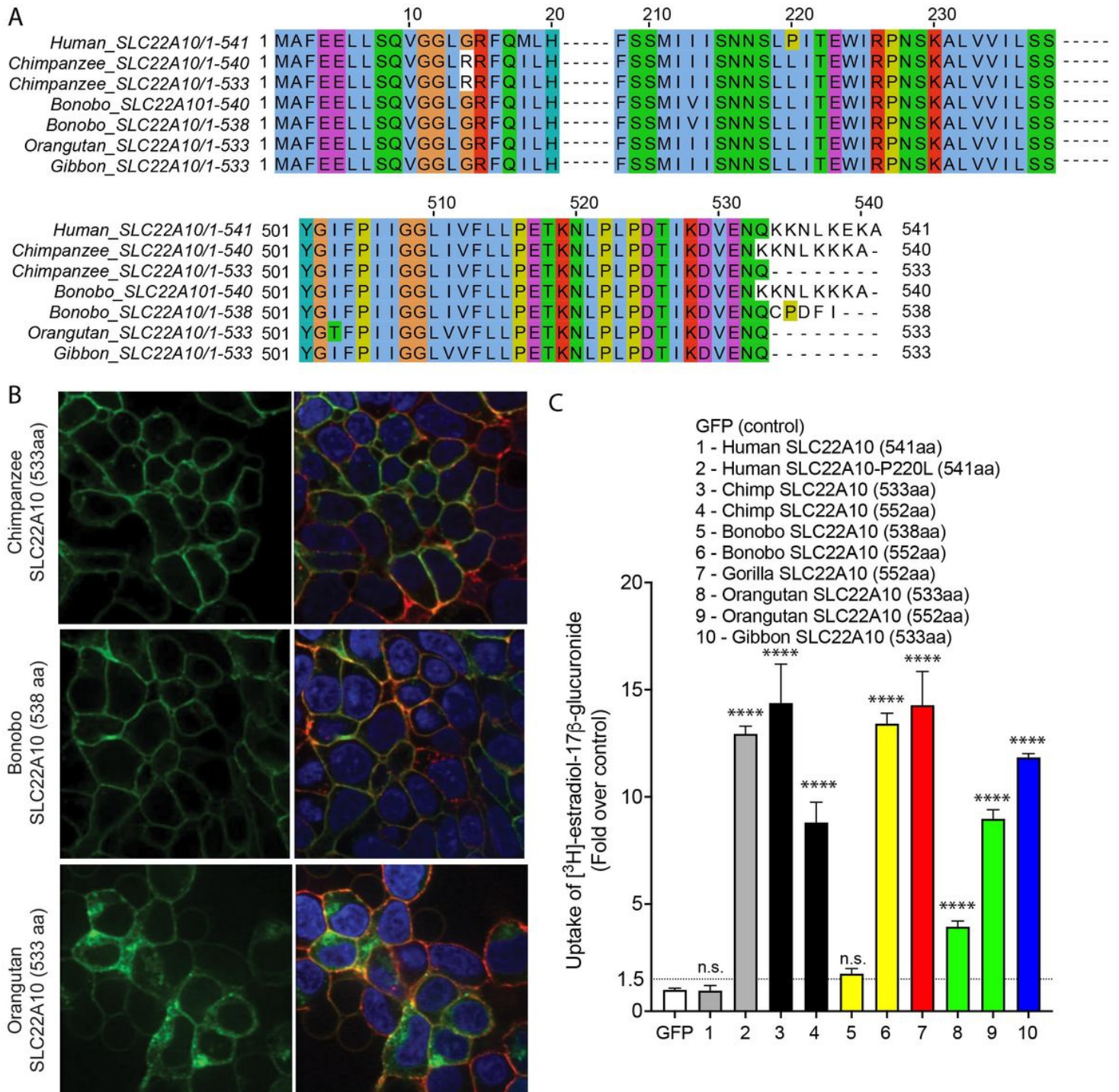


Figure 4

SLC22A10 of chimpanzees, bonobos, orangutans, and gibbons are predicted to have shorter isoforms expressing 533 or 538 amino acids. A. A comparison of the SLC22A10 amino acid sequence of humans, chimpanzees, bonobos, and orangutans, which express 533 (chimpanzee, orangutan, gibbon), 538 (bonobo), 540 (bonobo, chimpanzee), or 541 (human) amino acids, shows that the major differences are at the end of the SLC22A10 sequence. B. Confocal imaging revealed that SLC22A10 from chimpanzees and bonobos (isoforms expressing 533 or 538 amino acids) primarily localize to the plasma membrane of the cell, whereas weaker localization was observed for orangutan SLC22A10 (533 amino acids) to the plasma membrane of the cell. GFP conjugated to SLC22A10 was used for this experiment. C. The uptake of [³H]-estradiol-17 β -glucuronide in HEK293 cells was observed after transient transfection of plasmids encoding human SLC22A10 with reference amino acids or SLC22A10 with reference amino acids of other great apes with different isoforms. The results showed that SLC22A10 isoforms expressing 533 and 552 amino acids significantly accumulate the substrate. However, weaker substrate accumulation was observed in cells transfected with the bonobo SLC22A10 isoform expressing 538 amino acids.

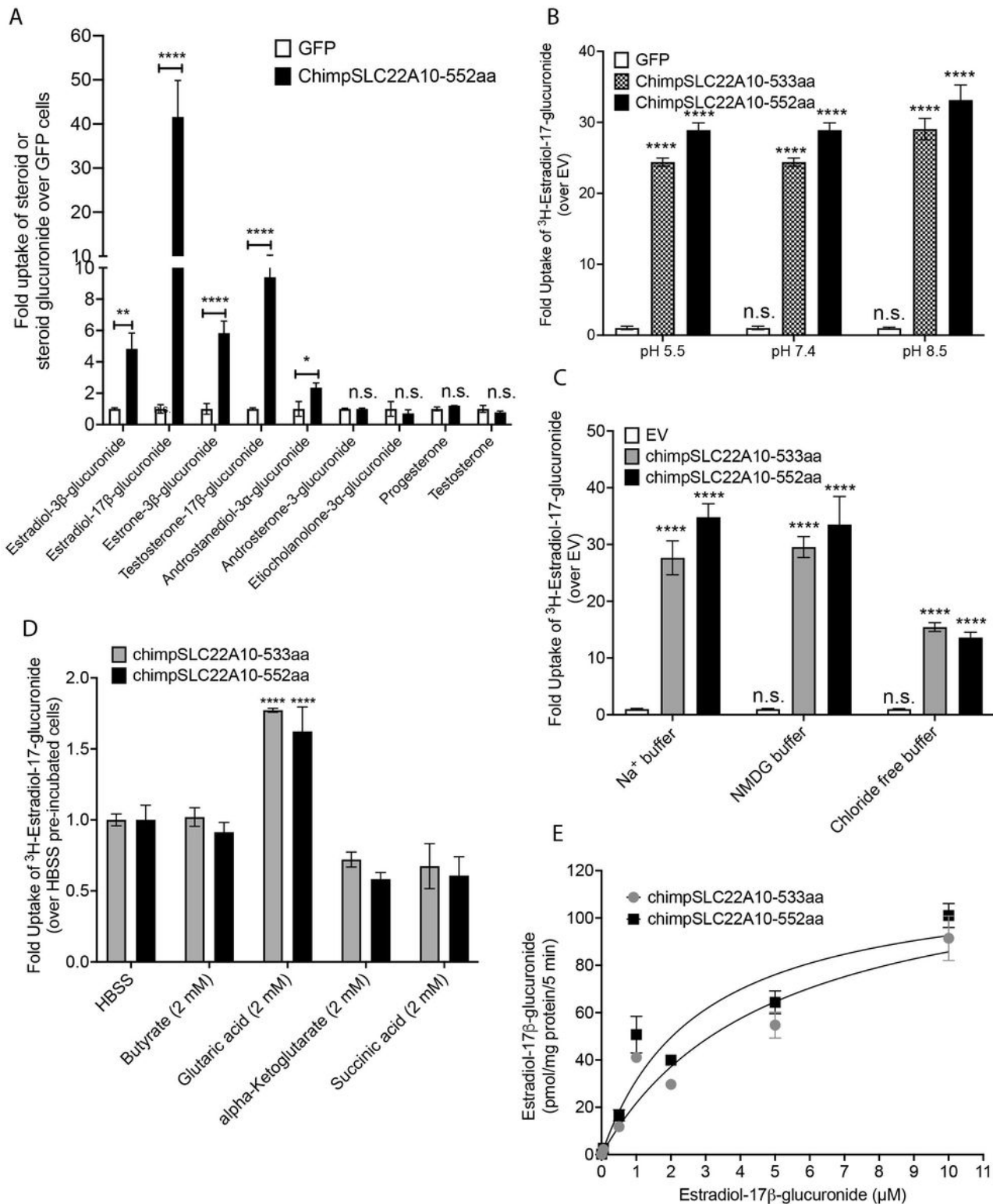


Figure 5

This figure presents information about the transport mechanism and kinetics of chimpanzee SLC22A10.

A. The uptake of seven steroid glucuronides and two steroids in HEK293 cells stably transfected with chimpanzee SLC22A10 (552 amino acids) was measured using LC/MS-MS to determine the accumulation of the compounds.

B. The effect of pH on accumulation of [³H]-estradiol-17b-glucuronide in HEK293 cells stably transfected with chimpanzee SLC22A10 isoforms expressing 533 and 552 amino acids was investigated.

C. The effect of sodium and chloride on accumulation of [³H]-estradiol-17b-glucuronide in HEK293 cells stably transfected with 533 and 552 amino acid isoforms of chimpanzee SLC22A10 was investigated.

D. The effects of trans-stimulation of [³H]-estradiol-17b-glucuronide uptake by chimpanzee SLC22A10 was determined. Uptake was trans-stimulated by preloading the cells with 2 mM of butyrate, glutaric acid, alpha-ketoglutarate, or succinic acid for 2 hours, and then measuring the uptake of [³H]-estradiol-17b-glucuronide after 15 minutes. The data are presented as mean ± S.D. and were normalized by setting the uptake of SLC22A10-expressing cells trans-stimulated by HBSS to 1.0. Trans-stimulation of [³H]-estradiol-17b-glucuronide by glutaric acid was observed for both isoforms of chimpanzee SLC22A10. E. The kinetics of [³H]-estradiol-17b-glucuronide uptake for chimpanzee SLC22A10 isoforms expressing 533 and 552 amino acids were analyzed. The uptake rate was evaluated at 5 minutes and the data were fit to a Michaelis-Menten equation. To fit the kinetic curve to a Michaelis-Menten equation, the concentration of estradiol-17b-glucuronide is set up to 10 μM. The figure shows a representative plot from one experiment. All experiments were repeated once, in triplicate and showed similar results.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableV2.pdf](#)
- [SupplementaryFiguresV3.pdf](#)
- [rs.pdf](#)
- [TableSLC22A10msV4.pdf](#)