Functional characterization of glucose transporter SLC2A14/GLUT14 and Vitamin C transporter SLC23A1/SVCT1：SLC2A14 isoforms substrates and redundancy of the SLC23A1 N-terminus

By

Ruotong Liu

A Thesis Submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

Master of Science

Department of Food and Human Nutritional Sciences
University of Manitoba
Winnipeg

Overall Abstract

Glucose and Vitamin C are both essential nutrients for human survival and share some transmembrane transport pathways.

The solute carriers of the *SLC2A* (facilitated glucose transporters, GLUT) family mediate glucose/hexose/dehydroascorbic acid transport through facilitated diffusion. The last discovered member *SLC2A14* encodes GLUT14, a very poorly characterized protein. Five GLUT14 isoforms are identified, however, only one had been described to mediate multi-specific hexose and dehydroascorbic acid transport, while the others' substrates remain to be identified.

The apical ascorbic acid uptake in epithelia cells is mediated by *SLC23A1*, a very well characterized gene/protein.

SLC2A14 and *SLC23A1* share genetic association with inflammatory bowel disease, Alzheimer's disease, and Parkinson's diseases. However, it remains unclear if these are independent or related to the shared pathways. To inform on this, the characterizations of the functions of the GLUT14 isoform will be essential. It is hypothesized that all GLUT14 isoforms encode multi-specific hexose transporters.

Since SLC23A1 functions are well known, the identification of functional genetic variations will be contributing to decipher the genetic associations. Currently, there are 440 missense mutations recorded in the databases, precipitating the need to prioritize them for functional testing. The 5' region of the *SLC23A1* open reading frame, corresponding to the N-terminus of the protein, seems to harbor an elevated amount of missense mutation. It is therefore hypothesized that it might be poorly conserved and redundant.

These research gaps are addressed using bioinformatic and functional studies in the *Xenopus laevis* expression system. Alignments and conservations were analyzed using Sequencher 5.0, SnapGene 6.0.2, and Aminode software. All the functional studies were conducted using the *Xenopus laveis* oocyte system for determining the uptake of C^{14}-ascorbic acid or H^{3}-2-D-deoxyglucose upon expression of the respective transporters.

SLC23A1 30 N-terminal amino acids are redundant based on poor conservation and unaltered protein function. Moreover, the elimination of 77 N-terminal amino acids reduced

SLC23A1 function by only 20%. Therefore, the 74 missense mutations in this region can be deprioritized from functional testing and are unlikely to impact the clinical relevance.

All isoforms but GLUT14-C, which is non-functional, are multi-specific hexose transporters. The substrate spectrum includes glucose, galactose, mannose, and xylose, allowing further explorations into their role in disease associations.

Acknowledgement

I want to express my appreciation to a few people who helped and supported me through my master's project. I am sincerely grateful to my supervisor, Dr. Peter Eck, for providing me with a great opportunity to explore the nutrigenomic area. Thank you for your excellent supervision, encouragement, outstanding support, and trust. I met you when I was twenty, without knowing any molecular biology background. I knocked over a new pipette tip box and broke the 10 ul pipette the first time you showed me how to do a PCR reaction. You didn't blame me at all and encouraged me to continue the experiments. I will always remember your kindness and your encouragement. You taught me how to think independently and always keep curation on my research. It is an unforgettable journey in my life, and I truly thank you! I would also like to thank my committee members: Dr. Carla Taylor and Dr. Dylan MacKay. Dr. Talyor, you are always so helpful. Your questions were always nice and clear. They made me realize my weakness and gave me a chance to catch up. Dr. MacKay, your expertise in clinical trials open my view to the clinical study. Without your help, I may not realize the potential influence on my work. I would also like to sincerely thank my friend Haonan Zhouyao, a Ph.D. student in our group and one of my best friends here. Thank you for teaching me laboratory skills, sharing the experimental tricks with me and always supporting me when I fail my experiments. We share our happiness in and out of the lab. I am also grateful for the staff and the Food and Human Nutritional Science department at the University of Manitoba, especially Denise Aminot-Gilchrist, Emily Gregorchuk, Helena Marak, Jerry Jin and Shari Rey. They offered me great support in all paper works and order receiving and handling processes.

Dedication

I dedicate this thesis to my family whom raise me with full of love, trust and encouragnment.

To my mother Hui Ma.

To my father Jian Liu.

To my husband Longhao Li.

I could not do these by myself without their consistenly love, support, and trust.

Table of Content

List of Tables

List of Figures

Abbreviations:

- 3-(N-morpholino) propanesulfonic acid (MOPS)
- Alzheimer's disease (AD)
- Amino Acid Evolutionary Constrained Analysis's (Aminode)
- Amino Acids (AAs)
- Basic Local Alignment Search Tool (BLAST)
- Complementary ribonucleic acid (cRNA)
- Counts per minute (CPM)
- European Molecular Biology Laboratory (EMBL)
- Human Genome Project (HGP)
- Inflammatory bowel disease (IBD)
- Messenger ribonucleic acid (mRNA)
- Na+-coupled glucose transporter (SGLT)
- National Center for Biotechnology Information (NCBI)
- Open reading frame (ORF)
- Parkinson's disease (PD)
- Peripheral blood mononuclear cells (PBMCs).
- Phylogenetic Analysis with Space/Time models (PHAST)
- Protein Data Bank (PDB)
- Protein Homology/analogY Recognition Engine (Phyre)
- Recommended dietary allowance (RDA)
- Reference Sequences (RefSeq)
- Single cell RNA sequencing (scRNAseq)
- Single nucleotide polymorphisms (SNPs)
- Sodium dodecyl sulfate (SDS)
- Solute carrier family 2 (SLC2)
- Solute carrier family 2 member 14 (SLC2A14)
- Solute carrier family 23 member 1 (SLC23A1)
- Solute carrier family 23 member 2 (SLC23A2)

- Sorting intolerant from tolerant (SIFT)
- Transcript per million (nTPM).
- Translation starting site (TSS)
- Tris-acetate-EDTA (TAE)
- University of California Santa Cruz (UCSC)

-

Chapter 1: Overall Introduction

Membrane transporters are proteins located on the cell membranes involved in the movement of organic and inorganic molecules (Perland & Fredriksson, 2017). These proteins function as carriers and channels to bring the substrates in and out from cells (Alberts B, Johnson A, Lewis J, et al., 2002).

Glucose and Vitamin C are two essential nutrients for the human body. The solute carriers of the *SLC2A* (facilitated glucose transporters, GLUT) family mediate glucose transport through facilitated diffusion, and a substantial number of them (GLUT1, GLUT3, GLUT4, GLUT8, and GLUT14) also accept dehydroascorbic acid (reduced form of Vitamin C) as a substrate.

Cellular uptake of ascorbic acid is mediated by *SLC23A1* and *SLC23A2* in the *SLC23A* family, and they are sodium dependent and concentrative. Therefore the proteins of this gene family are often called sodium dependent Vitamin C transporters or SVCTs (Daruwala, Song, Koh, Rumsey, & Levine, 1999; Holman, 2020).

Before the discovery of the ascorbic acid transporter (SVCT1 and SVCT2) (Savini, Rossi, Pierro, Avigliano, & Catani, 2008; H. Wang et al., 1999), glucose transporters (GLUT1, GLUT3, GLUT8 and GLUT4) were thought to be the main membrane transporters for Vitamin C, by transport of dehydroascorbic acid into cells (Rumsey et al., 2000, 1997). Upon entry into the cell, dehydroascorbic acid is reduced to the active Vitamin form, ascorbic acid. However, it is currently accepted that the GLUTs are only occupying a minor role in the overall homeostasis of Vitamin C.

Genetic variations of *SLC23A1* and *SLC2A14* are both associated with inflammatory bowel disease (IBD), Alzheimer's disease (AD), and Parkinson's disease (PD) (Chen et al., 2021; Infante et al., 2015; Liu et al., 2021; Mandana Amir; Shaghaghi et al., 2017; Shulman et al., 2011; W. Wang et al., 2012).

The biology of the *SLC23A1* gene is well established and it can influence body homoeostasis of ascorbic acid through its role in the reabsorption of ascorbic acid in the proximal renal tubule (Corpe et al., 2010). Therefore, the investigations into the impact of variations on the gene will contribute to the establishment of their clinical relevance. On the contrary, there are very few reports on the *SLC2A14* gene, and its biology remains to be characterized.

The knowledge status of these two genes are so different and there is a need to fill the research gaps for both genes to help with the understanding their interaction in the disease development. The following chapter of this thesis is going to review the current knowledge status of *SLC23A1* and *SLC2A14*. The strengths and the limitations of using the *Xenopus laevis* oocytes as the expression model to study the substrates uptake will be discussed.

References

Chen, L., Sun, X., Wang, Z., Lu, Y., Chen, M., He, Y., … Zheng, L. (2021). The impact of plasma vitamin C levels on the risk of cardiovascular diseases and Alzheimer's disease: A Mendelian randomization study. *Clinical Nutrition*, *40*(10), 5327–5334. https://doi.org/10.1016/j.clnu.2021.08.020

Daruwala, R., Song, J., Koh, W. S., Rumsey, S. C., & Levine, M. (1999). Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2. *FEBS Letters*, *460*(3), 480–484. https://doi.org/10.1016/S0014-5793(99)01393-9

Holman, G. D. (2020). Structure, function and regulation of mammalian glucose transporters of the SLC2 family. *Pflugers Archiv European Journal of Physiology*, *472*(9), 1155–1175. https://doi.org/10.1007/s00424-020-02411-3

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., … Sainz, J. (2015). Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. *Neurobiology of Aging*, *36*(2), 1105–1109. https://doi.org/10.1016/j.neurobiolaging.2014.10.039

Liu, H., Zhang, Y., Zhang, H., Wang, L., Wang, T., Han, Z., … Liu, G. (2021). Effect of plasma vitamin C levels on Parkinson's disease and age at onset: a Mendelian randomization study. *Journal of Translational Medicine*, *19*(1), 1–9. https://doi.org/10.1186/s12967-021-02892-5

Perland, E., & Fredriksson, R. (2017). Classification Systems of Secondary Active Transporters. *Trends in Pharmacological Sciences*, *38*(3), 305–315. https://doi.org/10.1016/j.tips.2016.11.008

Rumsey, S. C., Daruwala, R., Al-Hasani, H., Zarnowski, M. J., Simpson, I. A., & Levine, M. (2000). Dehydroascorbic acid transport by GLUT4 xenopus oocytes and isolated rat adipocytes. *Journal of Biological Chemistry*, *275*(36), 28246–28253. https://doi.org/10.1074/jbc.M000988200

Rumsey, S. C., Kwon, O., Xu, G. W., Burant, C. F., Simpson, I., & Levine, M. (1997). Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid. *Journal of Biological Chemistry*, *272*(30), 18982–18989. https://doi.org/10.1074/jbc.272.30.18982

Savini, I., Rossi, A., Pierro, C., Avigliano, L., & Catani, M. V. (2008). SVCT1 and SVCT2: Key proteins for vitamin C uptake. *Amino Acids*, *34*(3), 347–355. https://doi.org/10.1007/s00726-007-0555-7

Shaghaghi, M. A., Zhouyao, H., Tu, H., El-Gabalawy, H., Crow, G. H., Levine, M., … Eck, P. (2017). The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease. *American Journal of Clinical Nutrition*, *106*(6), 1508–1513. https://doi.org/10.3945/ajcn.116.147603

Shulman, J. M., Chipendo, P., Chibnik, L. B., Aubin, C., Tran, D., Keenan, B. T., … De Jager, P. L. (2011). Functional screening of alzheimer pathology genome-wide association signals in drosophila. *American Journal of Human Genetics*, *88*(2), 232–238. https://doi.org/10.1016/j.ajhg.2011.01.006

Wang, H., Dutta, B., Huang, W., Devoe, L. D., Leibach, F. H., Ganapathy, V., & Prasad, P. D. (1999). Human Na+-dependent vitamin C transporter 1 (hSVCT1): Primary structure, functional characteristics and evidence for a non-functional splice variant. *Biochimica et Biophysica Acta - Biomembranes*, *1461*(1), 1–9. https://doi.org/10.1016/S0005-2736(99)00182-0

Wang, W., Yu, J. T., Zhang, W., Cui, W. Z., Wu, Z. C., Zhang, Q., & Tan, L. (2012). Genetic association of SLC2A14 polymorphism with Alzheimer's disease in a Han Chinese population. *Journal of Molecular Neuroscience*, *47*(3), 481–484. https://doi.org/10.1007/s12031-012-9748-y

Chapter 2: Literature Review

SLC2A14

Glucose, as a fundamental energy source for cells, plays an essential role in ATP generation from both oxidation and nonoxidative pathways. Due to the polar nature of the glucose molecule, it cannot cross the plasma membranes by diffusion. Thus, it requires specific carrier molecules to deliver glucose into cells. To date, 19 glucose transporters are identified (Holman, 2020; Stringer, Zahradka, & Taylor, 2015). These glucose transporters can be classified into two different groups based on their glucose uptake mechanism: active hexose transport and facilitative hexose transport (Stringer et al., 2015). Our candidate gene belongs to the *solute carrier family 2* (*SLC2*) family which encode the protein called GLUTs. Although the basic biology of most of the glucose transporters in *SLC2* family is established, the recently discovered *solute carrier family 2 member 14* (*SLC2A14*) has significant research gaps, as described in the following paragraphs.

The History of *SLC2A14* Discoveries

The *SLC2A14* gene encodes a protein named GLUT14, and it was first recorded in the literature in 2002. Due to the coding similarities, GLUT14 was recognised as a gene duplicon of the *SLC2A3* gene, which encodes GLUT3 (Wu & Freeze, 2002). Later research showed that GLUT14 recognises glucose and dehydroascorbic as substrates (Mandana Amir; Shaghaghi et al., 2017). This is consistent with the other hexose transporters in the *SLC2A* family, specifically class-I GLUT proteins, which are generally multi-specific monosaccharide carriers (Holman, 2020). Following this notion, it was very recently described that the GLUT14 protein isoform A is also accepting mannose, xylose, glucosamine, arabinose and galactose as substrates (Alhashim, 2022). However, there are five GLUT14 protein isoforms, and the biology of four of them remains to be determined (Alhashim, 2022).

Structural Identification of *SLC2A14*

The *SLC2A14* gene is located on chromosome 12 q31.2. In 2002, two isoforms were described and denominated as the GLUT14 long and short isoforms (Wu & Freeze, 2002). These two

isoforms are now denominated GLUT14 isoform A and GLUT14 isoform B through the US National Center for Biotechnology Information (NCBI) annotation process. With the recently increasing sequencing influx, there are two additional GLUT14 isoforms annotated (C and D). Moreover, by visual curation of the aligned expressed sequence tags one additional isoform was identified by us, and named GLUT14 isoform E (Alhashim, 2022) **(Figure 2.1)**.

Overall, five GLUT14 isoforms (A-E) are identified, derived from twenty exons. As indicated in the Figure 2.1, exon 12 to exon 20 are utilized in all GLUT14 isoforms except for isoform C, which skips exon 13. The main difference between these isoforms is exon utilization in the 5' of the gene, resulting in alternative translation start sites (TSS). Isoform A starts in exon 11 and isoform B starts on exon 10 and skips exon 11. Isoform C is the shortest isoform with its translation start site in exon 12, while isoform D has the longest coding sequence (1608bp) and starts on exon 4. Isoform C-E are the orphan isoforms for GLUT14.

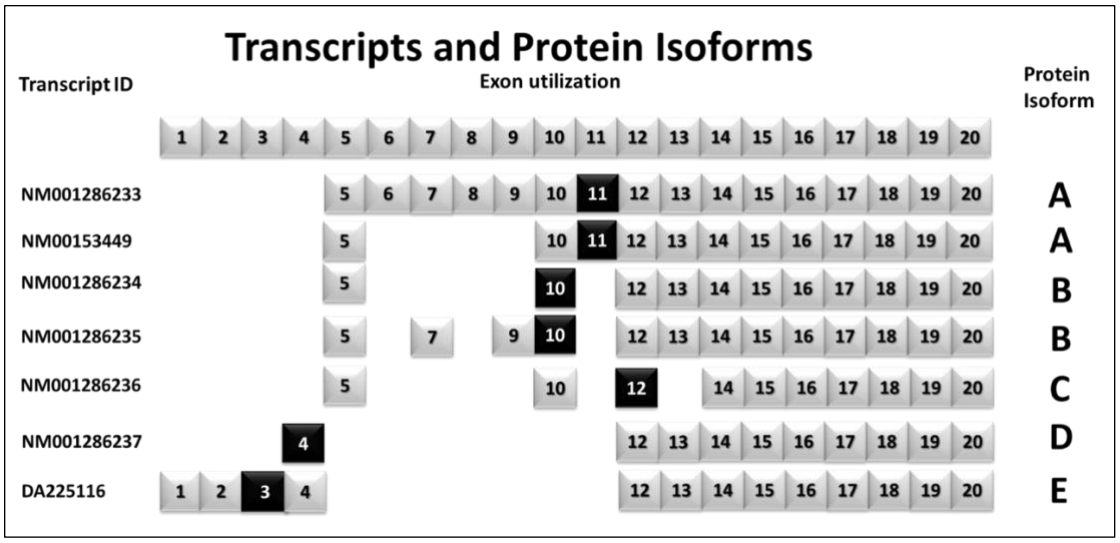


Figure 2.1. *GLUT14* exons utilization. The black boxes indicate the exon containing the translation start site. Note: this graph is adapted from Aqilah Alhashim's master's thesis: The homo-sapiens solute carrier family 2 member 14 (SLC2A14) – *further insight into the genomic organization, protein isoforms, substrates and kinetics* (2022). Copyright 2022 by Aqilah Alhashim. It is reproduced with permission.

SLC2A14 Gene Expression

SLC2A14 gene expression was first described to be exclusively in human testis (Wu & Freeze, 2002). They subcloned two *SLC2A14* transcripts from human testis cDNA and described the open

reading frames (ORF) of GLUT14 isoforms A and B. Subsequently, expression in the human small intestine, liver and kidney was reported, which partially explained existing disease associations (M. Shaghaghi, Murphy, & Eck, 2016). More recent single-cell RNA sequencing data has verified GLUT14 expression in adipose tissue, bone marrow, brain, breast, endometrium, heart muscle, lung, ovary, pancreas, placenta, prostate, skeletal muscle and skin (Karlsson et al., 2021). The expression in the brain and nervous system was also described in a report which indicated that the *SLC2A14* was associated with Alzheimer's disease and Parkinson's disease (Infante et al., 2015; Shulman et al., 2011; W. Wang et al., 2012). However, this was not further investigated.

The GLUT14 isoforms A and B are targeted to the plasmalemma membrane in CHO-K1 cells (M. Shaghaghi et al., 2016). This targeting was confirmed in CaCo-2 cells, where it was also observed for GLUT14 isoforms C, D and E. Thus, all isoforms of GLUT14 are targeted to the plasmalemma membrane (Alhashim, 2022).

SLC23A1

L-ascorbic acid, the reduced form of Vitamin C, is an essential nutrient for humans. It is a cofactor for eight different enzymes and also functions as a chemical reducing agent or antioxidant (Levine, Rumsey, Daruwala, Park, & Wang, 1999; Padayatty et al., 2003). Recommended dietary allowances for Vitamin C are used worldwide as ingestion guidelines. Although prior recommendations were based on intake preventing the deficiency disease, i.e. scurvy, new recommendations are based on dose-concentration relationships determined from pharmacokinetics studies in healthy individuals (Compounds, 2000; Levine et al., 2018; Levine, Wang, Padayatty, & Morrow, 2001). These pharmacokinetic studies served as a first step to determine systemic Vitamin C concentrations necessary to prevent diseases caused by marginal supply. Marginal Vitamin C intake has been linked to cardiovascular diseases, some cancers, metabolic syndrome, diabetes, inflammation, infections, and birth complications (Levine et al., 1999; Padayatty et al., 2003; Padayatty & Levine, 2008). The key role of the solute carrier 23 family member A1 (*SLC23A1*) in ascorbic acid homoeostasis is recognized.

The *SLC23A1* gene encodes a protein named SVCT1. Its key role in systemic maintenance of ascorbic acid levels was demonstrated by *slc23a1*$^{-/-}$ mice, which have disrupted systemic control of ascorbate concentrations caused by high renal excretion (Corpe et al. 2010).

Based on data obtained in 2012-2013, around 20% of Canadians, who did not take any Vitamin C supplements, were at moderate risk of Vitamin C deficiency (Langlois, Cooper, & Colapinto, 2016). This result could be explained by low dietary intake of Vitamin C, but also could be explained by the functional-damaging mutations' impacting SLC23A1, or a combination of both. A total of 9 studied SNPs were listed in the NCBI database and two of them are associated with lower plasma ascorbic acid levels (Corpe et al. 2010; Timpson et al., 2013). However, the functional impacts of variation in the *SLC23A1* gene remains undescribed, priming such investigations.

The Structure and the Tissue Expression of *SLC23A1*

The *SLC23A1* gene is located on chromosome 5q31.2, transcribing into three different transcripts encoding three in-frame-isoforms, of which two are annotated as NCBI refence transcripts.

SLC23A1 transcript NM_005847.5 contains a 1797 bp open reading frame, and it encodes for the canonical SLC23A1 isoform A (598 amino acids), which exhibits the highest transport activity (Mandana Amir Shaghaghi, 2015).

SLC23A1 transcript NM_152685.4 encodes the non-functional isoform B (602 amino acids), which utilizes an in-frame alternative donor splice site at one of the internal coding exons resulting in four additional amino acids compared to isoform A (H. Wang et al., 1999).

The third transcript is annotated as the predicted isoform X2 (XM_005272149.4), and we previously reported its utilization of an alternative 5' exon (**Figure 2.2**). The encoded protein exhibited only 7% of the ascorbic acid uptake compared to isoform A (Mandana Amir Shaghaghi, 2015). Moreover, the expression of the first exon of transcript XM_005272149.4 is only 1% compared to the canonical NM_005847.5. This isoform seems to be redundant.

Figure 2.2: Exons and introns utilization for three SLC23A1 transcripts. Three *SLC23A1* transcripts aligned to the genome as depicted in the NCBI genome browser, March 7th 2022. Boxes indicate exons, and lines indicate introns. Untranslated regions are light green, and coding regions are dark green.

SLC23A1 expression is mainly confined to absorptive epithelial tissues, such as the human gastrointestinal tract (mainly in the small intestine), kidney, liver, gall bladder and prostate (Karlsson et al., 2021). SLC23A1/SVCT1 is targeted to the apical membrane of polarised epithelial cells in culture (Boyer, Campbell, Sigurdson, & Kuo, 2005). Immunohistochemical confirmation of the apical location in tissues is depicted in **Figure 2.3**, available from the public database "The Human Protein Atlas" (https://www.proteinatlas.org/ENSG00000170482-SLC23A1/tissue).

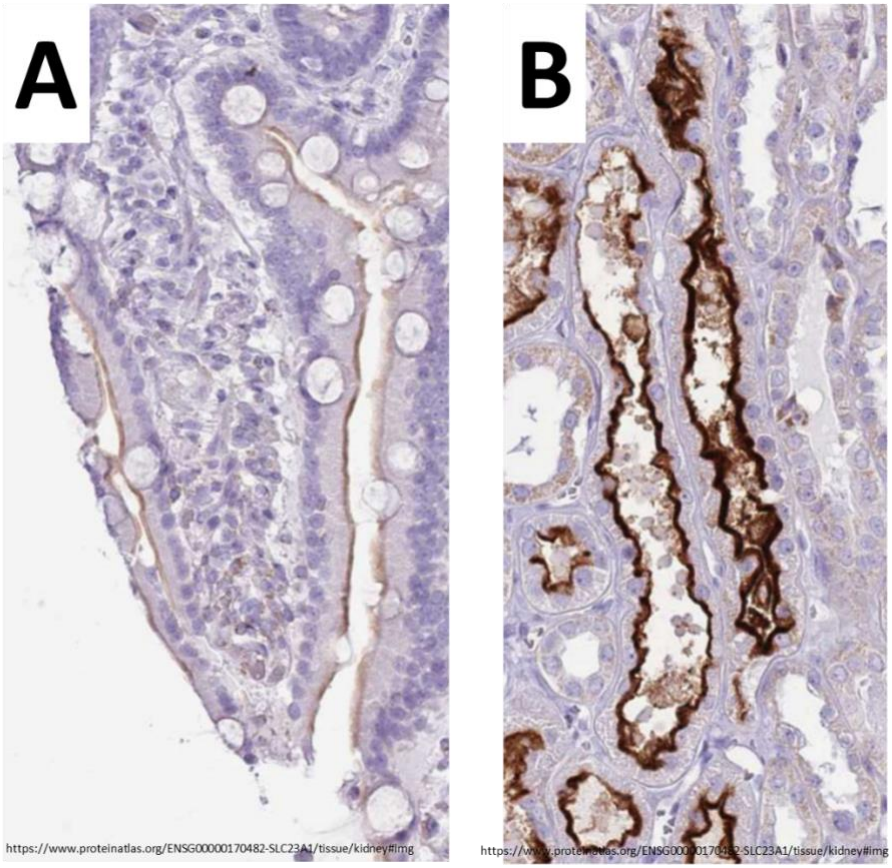


Figure 2.3: The immunohistochemical localization of *SLC23A1*. Immunohistochemical localisation of human SLC23A1/SVCT1 to the apical side of the enterocyte (villus depicted in panel A) and the renal proximal tubular cell (two longitudinally sliced proximal tubules are in the center and two transverse sliced proximal tubules are on the lower left side of panel B). Both pictures depicted from Human Protein Atlas website, March 7th (https://www.proteinatlas.org/ENSG00000170482-SLC23A1/tissue).

SLC23A1 and Kidney Ascorbic Acid Reabsorption

The ascorbic acid that humans consume from the diet is mainly absorbed through the small intestinal enterocyte, where SLC23A1 is expressed (**Figure 2.3**). SLC23A1/SVCT1 is a high capacity and low affinity ascorbic acid transporter (Timpson et al., 2013). However, in the *slc23a1* knock out mouse study, the intestinal ascorbic acid absorption of the mice is not impacted (Corpe et al., 2010). On the other hand, the *slc23a1*$^{-/-}$ mice served as a model to establish the gene's key role in systemic maintenance of ascorbic acid levels through the reabsorption in the kidney. The *slc23a1*$^{-/-}$ mice showed lower plasma ascorbic acid concentrations and they were viable if they survive birth (Corpe et al. 2010). These low plasma ascorbic acid levels observed are due to a lack of renal reabsorption. For example, female *Slc23a1*$^{-/-}$ mice had an 18-fold increase in the fractional excretion of ascorbate in the urine, and a 70% decrease in plasma ascorbate concentrations compared to wildtype counterparts (Corpe et al., 2010). This resulted in a perinatal mortality of about 50% of the offspring, which could be corrected by ascorbic acid supplementation.

With the development of next generation sequencing, a significant number of single nucleotide variations have been discovered for *SLC23A1*. However, none of the studies tested the functional impact of variations. However, existing disease associations and the proteins' key role in Vitamin C homeostasis warrants investigations into the functional impact and clinical significance of these variations.

Key Method for Functional Studies: *Xenopus laevis* oocytes System

Xenopus laevis has been used in many scientific laboratory studies, for instance, pregnancy tests, kidney and muscle function studies (Dawid & Sargent, 1988). Moreover, the oocytes of *Xenopus laevis* were found to be a great model to study the heterologous protein expression (Gurdon, Lane, Woodland, & Marbaix, 1971). The first protein expression experiment done in oocyte was with hemoglobin mRNA (Gurdon et al., 1971). *Xenopus laevis* oocytes had been used to characterized the function of many glucose transporters such as GLUT1, GLUT3, GLUT9 and GLUT12 (Long, O'Neill, & Cheeseman, 2018; Rogers, Chandler, Clarke, Petrou, & Best, 2003; Rumsey et al., 1997). Moreover, the functional characterization on *SLC23A1* was first done in the oocyte system (Daruwala, Song, Koh, Rumsey, & Levine, 1999). The functional and non-functional isoforms of *SLC23A1* were successfully determined (H. Wang et al., 1999). The

Xenopus laevis oocytes system also used in study the variations' impact on SLC23A1. Four missense variants on SLC23A1 were introduced into the oocytes system. All of them showed an significant decrease in ascorbic acid up take ability (Corpe et al., 2010).

In this thesis, all the functional studies on *SLC23A1* and *SLC2A14* were performed on the *Xenopus laevis* oocytes. The cRNA for each gene was injected into the oocytes and the oocytes were later incubated with C^{14}-ascorbic acid or H^3-2-D-dexoyglucose. The protein function was determined by measuring the internal radioactive activity for each oocyte.

Strength and limitations

As a mature model to study protein expression, the *Xenopus laevis* oocytes system does have significant advantages. First of all, the genomic structure of oocytes is very simple compared with other experimental animals such as mice or rats (Kloc, 2014). Except for a few ion transporters, not many transporters are originally expressed in the oocytes' membrane (Dawid & Sargent, 1988). Therefore, the oocytes provide us with a blank experimental background for the transporters we aimed to study. Second, the size of the oocytes and the number of the oocytes we can achieve from one female frog is remarkable. The size of the individual oocytes is around 1.1-1.3 mm, which is easy to handle and for performing the injections. Between 500 to 3000 oocytes can be retrieved from one female frog (Fortriede et al., 2020). Last but not least, the handling method of *Xenopus laevis* oocytes is easy. They can survive in an unfavorable environment for several days because they already store most of the essential nutrients for growth and the protein expression is in the yolk (Wagner, Friedrich, Setiawan, Lang, & Bröer, 2000).

Like all other methods, the oocytes system also has some disadvantages. *Xenopus laevis* is an poikilothermic animal, therefore its oocytes prefer an environment between 16 °C to 22 °C (Fortriede et al., 2020). But most of the experiments on the oocytes were done at room temperature (25 °C). This may increase the risk of protein expression due to the temperature changing (Wagner et al., 2000). Another limitation of this study is the oocysts do not distinguish between the apical and basolateral membrane sites. Therefore, the oocytes system can only help determine the function of the heterologous protein (Kloc, 2014).

Bioinformatic Resources

The Human Genome Project (HGP) determined the sequence and the location of the estimated 25,000 human genes (Lander et al., 2001). Expanding on the Human Genome Project public genome browsers and an ever-growing number of specific databases were developed to house and depict genomic, genetic and disease associations data. For the purpose of the current project, these databases serve as source for primary sequences, genetic variations, genomics/evolutionary comparison, expression and disease related data. In a blended approach the data would be used as depicted in browsers or downloaded to be further refined/analyzed.

In the current thesis a variety of bioinformatics tools and databases are applied to annotate genes, their encoded products (e.g., splice variants and protein isoforms), expression patterns, conservation, genetic variations and disease associations.

There are currently three leading public biomedical databases for the retrieval of genomic information, software tools for analyzing molecular and genomic data, and research in computational biology:

- The National Center for Biotechnology Information (NCBI) of the United States National Institutes (NIH) of Health at https://www.ncbi.nlm.nih.gov/ (Agarwala et al., 2016).
- The European Molecular Biology Laboratory (EMBL) Ensembl Genome Browser at https://uswest.ensembl.org/index.html (Madeira, Madhusoodanan, Lee, Tivey, & Lopez, 2019; Madeira, Park, et al., 2019).
- The University of California, Santa Cruz (UCSC) Genome Browser at http://genome.ucsc.edu (Kent et al., 2002; Raney et al., 2014)

These websites and genome browsers cross refence their content; however, their resources are distinct and unique and all three need to be queried and explored to cover the breadth of genomics and genetics information.

The NCBI creates and maintains over 40 integrated databases for the medical and scientific communities as well as the general public. It designs, develops, implements, and manages automated systems for the collection, storage, retrieval, analysis, and dissemination of knowledge concerning human molecular biology, biochemistry, and genetics. All content is freely available and not copyrighted. There are over 3 million visitors daily to its website, approximately 27

terabytes of data downloaded per day, and the number of users as well as downloads increases dramatically each year (Smith, 2008) . The NCBI served as a primary source for sequence downloads, and automatically names and annotates genes and genetic sequences. The sequences can subsequently be used and further analysed in local software, as for example Sequencher™ for alignment and visually curated annotation projects, as used here.

The NCBI is also a main protein sequence knowledgebase used in this thesis. Its protein database maintains the text record for individual protein sequences, derived from many different resources such as NCBI Reference Sequences (RefSeq), GenBank, Protein Data Bank (PDB), and UniProtKB/SWISS-Prot. The protein family- and domain-based classifications is annotated and can be interrogated through links from NCBI nucleotide Reference Sequences or through protein BLAST (Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov). Protein family sequences can be downloaded and aligned using the Clustal Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Madeira, Madhusoodanan, et al., 2019; Madeira, Park, et al., 2019; Sievers et al., 2011) and visually curated annotations of the alignments can be created for publications.

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcription (Navarro Gonzalez et al., 2021). Ensembl annotates genes and links them to detailed information on genetic variations and associated disease data. For example, this information in the Ensembl (Navarro Gonzalez et al., 2021) variant table was used to collect and further analyse the coding sequence variations (of the *SLC23A1* gene) ("Gene: SLC23A1 (ENSG00000170482) - Variant table - Homo_sapiens - Ensembl genome browser 105," n.d.).

Ensembl also houses the EMBL-EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home), an open public repository of gene expression pattern data under basal and different biological conditions, one of the most user-friendly resources to retrieve and further analyse expression data (Navarro Gonzalez et al., 2021). Expression Atlas was one major resource to query gene and protein expression data and to visualise down-stream analysis results to explore co-expression (Papatheodorou et al., 2020).

The UCSC Genome Browser is an online and downloadable genome browser hosted by the University of California, Santa Cruz (UCSC) (Karolchik et al., 2004). The Browser is a graphical viewer presenting a diverse collection of annotation datasets which are called "tracks" and

presented graphically (Raney et al., 2014). Similar to the other browsers, it includes mRNA alignments, gene predictions, gene-expression data as well as disease-association data.

The UCSC Genome Browsers utilise a variety of tools to analyse and display different kinds of conservation data, which is integrated into its "conservation track". The track shows multiple alignments and measurements of evolutionary conservation using the methods phastCons and phyloP from the Phylogenetic Analysis with Space/Time models (PHAST) package (http://compgen.cshl.edu/phast/) (Zhou, Liang, Lynch, Dennis, & Wishart, 2011). This software package is freely available and consists of a collection of command-line programs and supporting libraries for comparative and evolutionary genomics. Multiple alignments are generated using multiz (https://bio.tools/multiz) (Blanchette et al., 2004) and other tools in the UCSC/Penn State Bioinformatics comparative genomics alignment pipeline. PhyloP was used for the conservation analysis performed here; it separately measures conservation at individual columns, ignoring the effects of neighbors. The phyloP plots do not have a smooth appearance, but the "texture" represents individual nucleotides sites.

The Human Protein Atlas (https://www.proteinatlas.org/) integrates of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics and systems biology (Uhlén et al., 2015). All the data in the knowledge resource is open access to allow scientists both in academia and industry to freely access the data for exploration of the human proteome. Significantly, The Human Protein Atlas integrates the latest single cell RNA sequencing (scRNAseq) data from 25 human tissues and peripheral blood mononuclear cells (PBMCs). The scRNAseq analysis is based on publicly available genome-wide expression data and comprises all protein-coding genes in 444 individual cell type clusters corresponding to 15 different cell type groups. These data were a valuable source utilised to differentiate the genes expression in distinct cell types.

The Amino Acid Evolutionary Constrained Analysis's (Aminode) web interphase (http://www.aminode.org) is a easy to use tool to identify a proteins regions that are under evolutionary constraints and download high-resolution alignments images.

SLC23A1 and SLC2A14 are membrane spanning transporter proteins, and multiple tools exist to predict their structures based on already identified analogues. The Protein Homology/analogY Recognition Engine (Phyre) is a web based up to date to predict and analyze protein structure, function and mutations. The current version of the updated Phyre2 web portal

(http://www.sbg.bio.ic.ac.uk/phyre2) uses advanced remote homology detection methods to build models and predict ligand binding sites for a user's protein sequence (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). Due to its versatility and accuracy it was chosen for the protein structure predictions.

Table 2.1: Summary of bioinformatic website recourses. All resources were used to assist the project.

Website Name	Abbreviation	Link
National Center for Biotechnology Information	NCBI	https://www.ncbi.nlm.nih.gov/
European Molecular Biology Laboratory	EMBL	https://uswest.ensembl.org/index.html
The University of California, Santa Cruz Genome Browser	USCS	http://genome.ucsc.edu
Basic Local Alignment Search Tool	BLAST	https://blast.ncbi.nlm.nih.gov
Clustal Omega	---	https://www.ebi.ac.uk/Tools/msa/clustalo/
EMBL-EBI Expression Atlas	---	https://www.ebi.ac.uk/gxa/home
Phylogenetic Analysis with Space/Time Models	PHAST	http://compgen.cshl.edu/phast/
Multiz	---	https://bio.tools/multiz
Human Protein Atlas	---	https://www.proteinatlas.org/
Amino Acid Evolutionary Constrained Analysis's	Aminode	http://www.aminode.org
Protein Homology/analogY Recognition Engine (Phyre)	Phyre	http://www.sbg.bio.ic.ac.uk/phyre2

Knowledge gaps addressed in this thesis

SLC2A14/GLUT14

Current knowledge indicated that GLUT14 isoforms A and B are glucose and dehydroascorbic acid transporters (Mandana Amir; Shaghaghi et al., 2017). Further study characterised GLUT14 isoform A as a high affinity and low capacity multi-specific hexose solute carrier on the plasmalemma membrane (Alhashim, 2022). However, isoforms B remains largely uncharacterised and isoforms C, D, and E are not characterised at all. Moreover, none of the studies have ever compared the uptake efficiency between all fives isoforms. Lastly, the full spectrum of substrate needs to be identified for all GLUT14 isoforms.

SLC23A1/SVCT1

The knowledge on SLC23A1 is much further advanced compared to GLUT14, with functions well characterized. However, the functional impact of *SLC23A1* variations and their clinical implications are unknown. Some disease associations are reported, but the causal variants are not identified. Significantly, the functional impact of *SLC23A1* coding variations can be determined using the *Xenopus laevis* oocyte expression system (Corpe et al., 2010). However, the recent influx of next generation sequencing has lead to the identification of large amounts of coding variations now deposited in databases. Testing large numbers of mutations in the *Xenopus laevis* oocyte expression system is time consuming and costly. The *SLC23A1* gene currently harbors 440 missense variations, and 249 mutations were predicted to be deleterious by the variant-prioritization software programs SIFT and PolyPhen (Howe et al., 2021). To reduce the amount of functional testing, redundant sections of a transcript/protein can be determined, de-prioritizing the variants harbored in the region. The 5' region of the *SLC23A1* gene seems to harbor an elevated amount of coding variations but this remain to be interrogated.

Hypotheses

The knowledges gaps outlined above lead to the formation of two hypotheses which will be tested:

Hypothesis 1: The orphan isoforms of GLUT14 are encoding for multi-specific hexose carriers.

Hypothesis 2: The N-terminus of the SLC23A1/SVCT1 protein is redundant, and truncated proteins utilising alternative translational start sites are therefore fully functional.

Objectives

The following objectives serve to address the two hypotheses:

Objective 1: determine if GLUT14 isoforms C, D and E mediate glucose transport when expressed in *Xenopus laevis* oocytes.

Objective 2: compare the glucose transport of all GLUT14 isoforms in *Xenopus laevis* oocytes.

Objective 3: determine the spectrum of substrates for GLUT14 isoforms except for isoform A.

Objective 4: estimate the degree of redundancy of the SLC23A1 N-terminus through conservation analyses.

Objective 5: compare the ascorbic acid uptake mediated by SLC23A1 isoforms representing alternative in-frame downstream N-terminal translation start sites.

References

Alberts B, Johnson A, Lewis J, et al. (2002) Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Principles of Membrane Transport. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26815/

Alhashim, A. (2022). *The homo-sapiens solute carrier family 2 member 14 (SLC2A14) – further insight into the genomic organization, protein isoforms, substrates and kinetics*

Boyer, J. C., Campbell, C. E., Sigurdson, W. J., & Kuo, S. M. (2005). Polarized localization of vitamin C transporters, SVCT1 and SVCT2, in epithelial cells. *Biochemical and Biophysical Research Communications*, *334*(1), 150–156. https://doi.org/10.1016/j.bbrc.2005.06.069

Chen, L., Sun, X., Wang, Z., Lu, Y., Chen, M., He, Y., … Zheng, L. (2021). The impact of plasma vitamin C levels on the risk of cardiovascular diseases and Alzheimer's disease: A Mendelian randomization study. *Clinical Nutrition*, *40*(10), 5327–5334. https://doi.org/10.1016/j.clnu.2021.08.020

Compounds, I. of M. (US) P. on D. A. and R. (2000). Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids. *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. https://doi.org/10.17226/9810

Corpe, C. P., Tu, H., Eck, P., Wang, J., Faulhaber-Walter, R., Schnermann, J., … Levine, M. (2010). Vitamin C transporter Slc23a1 links renal reabsorption, vitamin C tissue accumulation, and perinatal survival in mice. *Journal of Clinical Investigation*, *120*(4), 1069–1083. https://doi.org/10.1172/JCI39191

Daruwala, R., Song, J., Koh, W. S., Rumsey, S. C., & Levine, M. (1999). Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2. *FEBS Letters*, *460*(3), 480–484. https://doi.org/10.1016/S0014-5793(99)01393-9

Holman, G. D. (2020). Structure, function and regulation of mammalian glucose transporters of the SLC2 family. *Pflugers Archiv European Journal of Physiology*, *472*(9), 1155–1175. https://doi.org/10.1007/s00424-020-02411-3

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. https://doi.org/10.1093/nar/gkaa942

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., … Sainz, J. (2015). Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. *Neurobiology of Aging*, *36*(2), 1105–1109. https://doi.org/10.1016/j.neurobiolaging.2014.10.039

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., … Lindskog, C. (2021). A single–cell type transcriptomics map of human tissues. *Science Advances*, *7*(31). https://doi.org/10.1126/SCIADV.ABH2169 Human Protein Atlas proteinatlas.org

Langlois, K., Cooper, M., & Colapinto, C. K. (2016). Vitamin C status of Canadian adults: Findings from the 2012/2013 Canadian Health Measures Survey. Retrieved February 26, 2022, from https://www150.statcan.gc.ca/n1/pub/82-003-x/2016005/article/14612-eng.htm

Levine, M., Conry-cantilena, C., Wang, Y., Welch, R. W., Washko, W., Dhariwal, K. R., … Cantilena, L. R. (2018). *Vitamin C Pharmacokinetics in Healthy Volunteers : Evidence for a Recommended Dietary Allowance Jean King and Louis R . Cantilena Source : Proceedings of the National Academy of Sciences of the United States of America , Published by : National Academy of*.

Levine, M., Rumsey, S. C., Daruwala, R., Park, J. B., & Wang, Y. (1999). Criteria and recommendations for vitamin C intake. *Journal of the American Medical Association*, *281*(15), 1415–1423. https://doi.org/10.1001/jama.281.15.1415

Levine, M., Wang, Y., Padayatty, S. J., & Morrow, J. (2001). A new recommended dietary allowance of vitamin C for healthy young women. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 9842–9846. https://doi.org/10.1073/pnas.171318198

Liu, H., Zhang, Y., Zhang, H., Wang, L., Wang, T., Han, Z., … Liu, G. (2021). Effect of plasma vitamin C levels on Parkinson's disease and age at onset: a Mendelian randomization study. *Journal of Translational Medicine*, *19*(1), 1–9. https://doi.org/10.1186/s12967-021-02892-5

Padayatty, S. J., Katz, A., Wang, Y., Eck, P., Kwon, O., Lee, J. H., … Dutta, S. K. (2003). Vitamin C as an Antioxidant: Evaluation of Its Role in Disease Prevention. *Journal of the American College of Nutrition*, *22*(1), 18–35. https://doi.org/10.1080/07315724.2003.10719272

Padayatty, S. J., & Levine, M. (2008). Fruit and vegetables: Think variety, go ahead, eat! *American Journal of Clinical Nutrition*, *87*(1), 5–7. https://doi.org/10.1093/ajcn/87.1.5

Perland, E., & Fredriksson, R. (2017). Classification Systems of Secondary Active Transporters. *Trends in Pharmacological Sciences*, *38*(3), 305–315. https://doi.org/10.1016/j.tips.2016.11.008

Rumsey, S. C., Daruwala, R., Al-Hasani, H., Zarnowski, M. J., Simpson, I. A., & Levine, M. (2000). Dehydroascorbic acid transport by GLUT4 xenopus oocytes and isolated rat adipocytes. *Journal of Biological Chemistry*, *275*(36), 28246–28253. https://doi.org/10.1074/jbc.M000988200

Rumsey, S. C., Kwon, O., Xu, G. W., Burant, C. F., Simpson, I., & Levine, M. (1997). Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid. *Journal of Biological Chemistry*, *272*(30), 18982–18989. https://doi.org/10.1074/jbc.272.30.18982

Savini, I., Rossi, A., Pierro, C., Avigliano, L., & Catani, M. V. (2008). SVCT1 and SVCT2: Key proteins for vitamin C uptake. *Amino Acids*, *34*(3), 347–355. https://doi.org/10.1007/s00726-007-0555-7

Shaghaghi, M., Murphy, B., & Eck, P. (2016). The SLC2A14 gene: Genomic locus, tissue expression, splice variants, and subcellular localization of the protein. *Biochemistry and Cell Biology*, *94*(4), 331–335. https://doi.org/10.1139/bcb-2015-0089

Shaghaghi, Mandana Amir;, Zhouyao, H., Tu, H., El-Gabalawy, H., Crow, G. H., Levine, M., … Eck, P. (2017). The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease. *American Journal of Clinical Nutrition*, *106*(6), 1508–1513. https://doi.org/10.3945/ajcn.116.147603

Shaghaghi, Mandana Amir. (2015). Genetic and functional studies of two intestinal vitamin C transporters, SLC23A1 and GLUT14, associated with inflammatory bowel disease B. *MySpace*, 161. https://doi.org/10.1503/cmaj.180710

Shulman, J. M., Chipendo, P., Chibnik, L. B., Aubin, C., Tran, D., Keenan, B. T., … De Jager, P. L. (2011). Functional screening of alzheimer pathology genome-wide association signals in drosophila. *American Journal of Human Genetics*, *88*(2), 232–238. https://doi.org/10.1016/j.ajhg.2011.01.006

Stringer, D. M., Zahradka, P., & Taylor, C. G. (2015). Glucose transporters: Cellular links to hyperglycemia in insulin resistance and diabetes. *Nutrition Reviews*, *73*(3), 140–154. https://doi.org/10.1093/nutrit/nuu012

Timpson, N. J., Forouhi, N. G., Brion, M. J., Harbord, R. M., Cook, D. G., Johnson, P., … Luan, J. (2013). Genetic variation at the SLC23A1 locus is associated with circulating concentrations of L-ascorbic acid (vitamin C): Evidence from 5 independent studies with > 15,000 participant (American Journal of Clinical Nutrition (2010) 92 (375-382)). *American Journal of Clinical Nutrition*, *98*(1), 253–254. https://doi.org/10.3945/ajcn.113.065342

Wang, H., Dutta, B., Huang, W., Devoe, L. D., Leibach, F. H., Ganapathy, V., & Prasad, P. D. (1999). Human Na+-dependent vitamin C transporter 1 (hSVCT1): Primary structure, functional characteristics and evidence for a non-functional splice variant. *Biochimica et Biophysica Acta - Biomembranes*, *1461*(1), 1–9. https://doi.org/10.1016/S0005-2736(99)00182-0

Wang, W., Yu, J. T., Zhang, W., Cui, W. Z., Wu, Z. C., Zhang, Q., & Tan, L. (2012). Genetic association of SLC2A14 polymorphism with Alzheimer's disease in a Han Chinese population. *Journal of Molecular Neuroscience*, *47*(3), 481–484. https://doi.org/10.1007/s12031-012-9748-y

Wu, X., & Freeze, H. H. (2002). GLUT14, a duplicon of GLUT3, is specifically expressed in testis as alternative splice forms. *Genomics*, *80*(6), 553–557. https://doi.org/10.1006/geno.2002.7010

TRANSITION STATEMENT 1

Ascorbic acid, as an oxidized form of Vitamin C in the diet, is an essential nutrient and a strong antioxidant for humans. The animal study on mice proved that *SLC23A1* is a major ascorbic acid absorption transporter in the small intestine, and it is the key protein/gene in the kidney for ascorbic acid reabsorption.

To date, there are 440 missense mutations for the *SCL23A1* transcript ENST00000348729.8 recorded in the ENSEMBL database. These coding variations can be introduced into expression plasmids and tested in the *Xenopus laevis* expression system *in vitro*. However, this will be very time-consuming and expensive. The determination of nonredundant domains of the SLC23A1 protein will help to deprioritize variations located therein.

In the following manuscript, the redundancy of the open reading frames 5' region corresponding to the protein's N-terminus was interrogated for evolutionary conservation using bioinformatics approaches including the models (PHAST) package (http://compgen.cshl.edu/phast/) and the Amino Acid Evolutionary Constrained Analysis (Aminode, http://www.aminode.org) prediction tool.

Moreover, functional studies were performed on two N-terminal truncated protein of SLC23A1, which utilize alternative cryptic in frame translational start codons (ATGs) to determine the redundancy in of the N-terminus. This served to de-prioritize some of the coding variations.

Chapter 3: Manuscript 2

Redundancy of the proximal coding region of the ascorbic acid solute carrier *SLC23A1*

Ruotong Liu[1], Haonan Zhouyao[1], Peter Eck

Target journal: APNM; article type: a brief communication.

Contributions of Authors:

Ruotong Liu performed most of the molecular work, including the amplification of the T7–tagged PCR products for SLC23A1 proteins. She collaborated with Haonan Zhouyao on the mRNA preparations and injections into *Xenopus laevis* oocytes and ascorbic acid up take experiments. She performed the data and statistical analysis and wrote the initial draft of the manuscript. Under Peter Eck's supervision she edited the manuscript.
Haonan Zhouyao designed the T7 tagged forward primer to amplify the SLC23A1 coding sequences. She assisted in the PCR amplifications. She collaborated with Ruotong Liu on the mRNA preparations and injections into *Xenopus laevis* oocytes and ascorbic acid up take experiments.
Peter Eck designed the study and managed and supervised the whole project. He edited the drafts.

Abstract

Background: The human *SLC23A1* gene encodes a solute carrier that mediates cellular ascorbic acid uptake on the apical pole of epithelial cells. Some of its single nucleotide polymorphisms (SNPs) are associated with lower ascorbic acid status and common diseases. In recent years a sizable amount of coding variations have been identified in the *SLC23A1* gene but remain to be functionally characterized. The 5' of the coding region seems to have an elevated density of coding variations, leading to the hypothesis that this region is redundant for the protein's functioning.

Objective: To determine the degree of redundancy of the 5' coding region through the analysis of conservation and functional expression.

Methodology: The nucleotides and amino acids conservations were evaluated *in silico*. The coding regions for the three most proximal translational starts sites were subcloned and expressed in *Xenopus laevis* oocytes. Respective solute carrier activities were assessed through the uptake of radiolabeled ^{14}[C]-ascorbic acid.

Results: The *SLC23A1* 5' region shows low evolutionary conservation. Ascorbic acid transport was identical for the two most N-terminal proteins and was slightly diminished for the truncation using the third translational start site.

Conclusion: The very 5' of the *SLC23A1* coding region is redundant for the protein's functioning, which means that nonsynonymous variations in this region should not impact the solute carrier function.

Introduction

The solute carrier *SLC23A1* mediates ascorbic acid uptake into cells (Daruwala, Song, Koh, Rumsey, & Levine, 1999), with the main expression in epithelia of the small intestine, liver and kidney and the reproductive system (Wang et al., 1999). Its key role in renal reabsorption of ascorbic acid was demonstrated through the *slc23a1*$^{-/-}$ mouse (Corpe et al., 2010). The *slc23a1*$^{-/-}$ mouse exhibits very high urinary ascorbic acid excretion leading to lowered plasma and tissue levels, which caused high perinatal offspring loss. This serves as a model that genetic elimination or variations in *SLC23A1* could lead to similar phenotypes in humans.

Polymorphisms in the human *SLC23A1* gene had been associated with decreased circulating ascorbic acid concentrations (Kobylecki, Afzal, Smith, & Nordestgaard, 2015; Senthilkumari et al., 2014; Timpson et al., 2013) and various diseases (Ravindran et al., 2019); however, causal variation(s) have not been identified.

The development of high throughput genome sequencing enabled the rapid identification of genetic variants, which are subsequently deposited in databases. For the *SLC23A1* gene, there are currently 7540 entries in dbSNP, a public-domain archive for human single nucleotide variations, microsatellites, and small-scale insertions and deletions. Of these, 440 are predicted to impact the protein sequence. Most of these coding variations will change the amino acid sequence and might impact the proteins functioning to various degrees. Algorithms to predict the degree of deleteriousness for coding variants exist and are integrated into various genome browsers (Howe et al., 2021). However, the ultimate proof of a functional impact of a variation can only be achieved experimentally.

The degree of *SLC23A1s* functioning can be assessed in the *Xenopus laevis* expression system (Corpe et al., 2010). However, assessing the impact of each variation in the *SLC23A1* coding region using the *Xenopus laevis* expression system will be time and resource-consuming since each variation needs to be introduced into an existing expression plasmid either by site-directed mutagenesis or gene synthesis. In order to eliminate the number of variations to be investigated, redundant sections of the gene could be identified to de-prioritize variation in the coding region. Here we test the redundancy of *SLC23A1's* N-terminal region using a combination of bioinformatics and functional expression.

Material and Methods

Bioinformatics

SLC23A1 isoform A open reading frame (ORF) - NM_005847.5 sequence was downloaded from the NCBI website and loaded into the Sequencer 5.1 software (Gene code corporation, Ann Arbor, MI, USA). Search function was applied in the software with keyword "ATG". Only the in frame ATG was counted. A total three T7 tagged forward PCR primers were designed for the in frame ATGs. The sequence for 1st, 2nd and 3rd ATG forward primers are: GGAAATTAATACGACTCACTATAGGGCCACCACCACCATGAGGGCCCAGGAGGACC ;GAAATTAATACGACTCACTATAGGGCCACCACCACCATGTTGTACAAGATCGAGGA CG and GGAAATTAATACGACTCACTATAGGGCCACCACCACCATGGTTAGTCAGCTC ATCGG.

The evolutional conservation score of human *SLC23A1* compared to 62 vertebrate species was generated on an online software called Aminode (http://www.aminode.org/search). The graph was automatically generated and the red line on the graph indicated the conservation score. CLUSTAL Omega protein alignments were created through the web interphase (https://www.ebi.ac.uk/Tools/msa/clustalo/).

All alignments in this paper were performed on Sequencher 5.1 software (Gene code corporation, USA) and proportionally translation into the graph. The sequences of *SLC23A1* NM_005847.5 and NM_152685.4 and Nucleobase Cation Symporter 2 XM_035465162.1 were downloaded from the NCBI website. All of them plus two 5' truncated SLC23A1 sequence were aligned together by clean data in Sequencher 5.1.

ORF Amplification and cRNA Transcription

The ORF was amplified by polymerase chain reaction (PCR) with Phusion High-Fidelity DNA polymerase (New England Biolabs Inc., Ipswitch, MA, USA). The annealing temperature for 1st, 2nd and 3rd ATG was 55 °C, 57 °C and 58 °C respectively. The PCR was run for 35 cycles with final extension temperature at 72C for 10 mins. The size of the PCR products was checked on 1% Tris-acetate-EDTA (TAE) agarose gel. Then, the PCR products were purified with a Gel/PCR DNA Fragments Kit (Geneaid Biotech Ltd., New Taibei City, Taiwan, China). One ug

of purified PCR product was added to the mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA) according to the manufacturer's protocol to produce 5' capped mRNA. The purification of cRNA was done with the E.Z.N.A.® MicroElute RNA Clean-Up Kit (OMEGA Bio-TEK, Norcross, GA, USA). The quality of cRNA was checked on 3-(N-morpholino) propanesulfonic acid gel (MOPs gel) and Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.). All cRNAs were stored in -80 °C freezer until injection.

Functional Studies Using Xenopus Laevis Oocytes

A female adult *Xenopus laevis* ovary was collected and digested with Collagenase, Type IV, powder (Sigma-Aldrich Canada Co., Oakville, ON, Canada) in calcium free OR2 buffer (4.82 g/L NaCl; 1.3 g/L HEPES; 2.5 mmol/L KCl, 1 mmol/L $MgCl_2$; 1 mmol/L Na_2HPO_4; pH=7.6). The digested oocytes were washed with regular OR_2 buffer (4.82 g/L NaCl; 1.3 g/L HEPES; 2.5 mmol/L KCl, 1 mmol/L $MgCl_2$; 1 mmol/L Na_2HPO_4; 1 mmol/L $CaCl_2$; pH=7.6). The calcium in the OR2 buffer stops the function of collagenase. The good oocytes (**Supplemental Figure 4**) were selected and sorted into a sterilized 60 mm petri-dish (FroggaBio Inc., Concord, ON, Canada) with OR_2 buffer plus 55 mg/L sodium pyruvate, 100 mg/L gentamicin (Sigma Aldrich) and 10 mL/L Penicillin-Streptomycin (Sigma Aldrich). This buffer combination was named overnight buffer in our lab. 36.8 nl of cRNA or DNase and RNase free water were injected into each oocyte with a Nanoject II Auto-Nanoliter Injector (Drummond Scientific Company, Amityville, NY, USA). The injected oocytes were incubated in overnight buffer at 16 °C for 72 hours.

Radiolabeled ascorbic acid, L-[1-14C] (American Radiolabeled Chemicals Inc., Saint Louis, MO, USA) was diluted to 0.045 mmol/L with OR_2 buffer. Twenty oocytes were considered as one experimental group. Each group was incubated in the C14-ascorbic acid buffer for 30 mins. Incubated oocytes were then washed with ice-cold OR_2 buffer for 5 times and transfered into individual 6mL pony vials (Perkinelmer Inc., Woodbrideg, ON, Canada). Later, 200 ul of 10% Sodium dodecyl sulfate (SDS) solution was added into each vial to dissolve the oocyte. Finally, 2 mL of Ultima Gold scintillation cocktail (Perkinelmer Inc.) was added into each vial. The radioactivity of each oocyte was measured by a Tri-Carb 3110TR Liquid

Scintillation Analyzer (Perkinelmer Inc.) for 60 s. The original data were counted as CPM per oocyte. It was converted to fmol per oocytes based on the radiolabeled specific activity.

Statistical Analysis

Outliers for each group were removed following the 1.5 times inter-quartile range (IQR) rule. The remaining data were imported in Graphpad Prism 8 (version 8.4.2, Graphpad Software Inc, CA, USA). A One-Way ANOVA was performed for multiple comparisons. When determining the statistically significant difference, the mean of each column with the mean of every other column was compared with Bartlett's test and Brown-Forsythe test to confirm statistically significant differences.

Results

The *SLC23A1* gene utilizes two alternative 5' exons resulting in two translational start sites (**Supplemental Figure 1**), coding for two in-frame proteins of 634 and 598 amino acids (Amir Shaghaghi, Yurkova, Tu, Levine, & Eck, 2014) However, the expression of the longer transcript/protein isoform is less than 1% of the shorter one in all tissues except the small intestine, where it is about 18%, indicating evolutionary redundancy. Moreover, the ascorbic acid transport of the longer isoform is only about 10% of the shorter protein, concurring with redundant evolution. Therefore, the current investigations will focus on the canonical transcript (NM_005847.5, **Supplemental Sequence 1**), containing a coding sequence of 1794 nucleotides (**Supplemental Sequence 2**), translating into the 598 amino acid reference protein NP_005838.3 (**Supplemental Sequence 3**). Transcript NM_005847.5 contains two in frame translational start codons located 90 and 231 bases downstream of the officially annotated start codon, which translates in proteins truncated by 30 or 77 amino acids, respectively (**Figure 3.1**, lower panels).

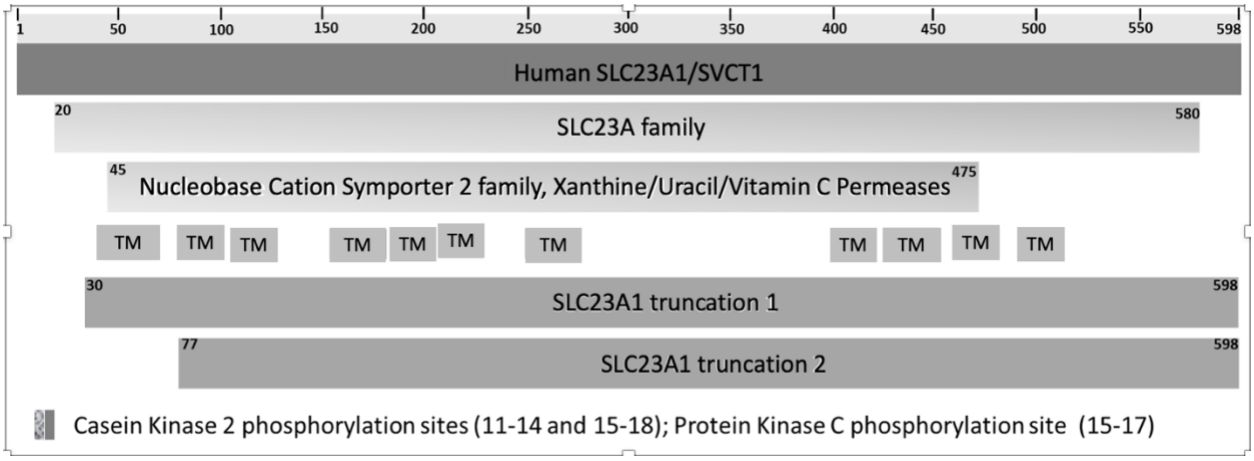


Figure 3.1: The *SLC23A1* protein alignment with *SLC23A1*, Nucleobase Cation Symporter 2 protein domains and two truncated protein of *SLC23A1*. The SLC23A1/SVCT1 protein aligned with recognized domains clustering with the SLC23A and the Nucleobase Cation Symporter 2 (Xanthine/Uracil/Vitamin C Permeases) protein families (three upper panels). The predicted location of Transmembrane domains (TM) are indicated by the dark squares; the truncated SLC23A1/SVCT1 proteins are depicted in the lower two panels (Diallinas, Gorfinkiel, Arst, Cecchetto, & Scazzocchio, 1995). Nomenclature and analysis followed the Interpro classification of protein families.

The SLC23A1 protein clusters within the SLC23A and the Nucleobase Cation Symporter 2 (Xanthine/Uracil/Vitamin C Permeases) protein families (**Figure 3.1**). This could indicate moderate protein conservation from amino acid 20 onwards (SLC23A family) and stronger conservation from amino acid 77 onwards (Nucleobase Cation Symporter 2 family, Xanthine/Uracil/Vitamin C Permeases family). One predicted transmembrane domain falls within the first 77 amino acids, on which this report focuses. The very N-terminus is predicted to be extracellular (**Supplemental Table 1**).

Two protein kinase sites are predicted in the N-terminal region (**Figure 3.1**), the first casein kinase II phosphorylation site for amino acids 11-14 followed by a casein kinase II/protein kinase C phosphorylation site at amino acids 15-18. This could indicate a regulative role of this region, which would somewhat counterargue against low evolutionary constraints.

To get a more precise understanding of the evolutionary conservation, the gene, the transcript as well as the protein was further analyzed.

The nucleotide conservation in the SLC23A1 (transcript NM_005847.5) first exon and the proximal second exon is very low compared to the distal parts of the gene (**Supplemental Figure 2**).

The amino acid conservation of the SLC23A1 proteins N-terminus is low (**Figure 3.2. A&B**), and only becomes more stringent after a second putative translation initiation site, corresponding to methionine 31 in the human protein. The three predicted phosphorylation sites fall within the region of low conservation, while the transmembrane domains fall into regions of higher conservation (**Figure 3.2.C**).

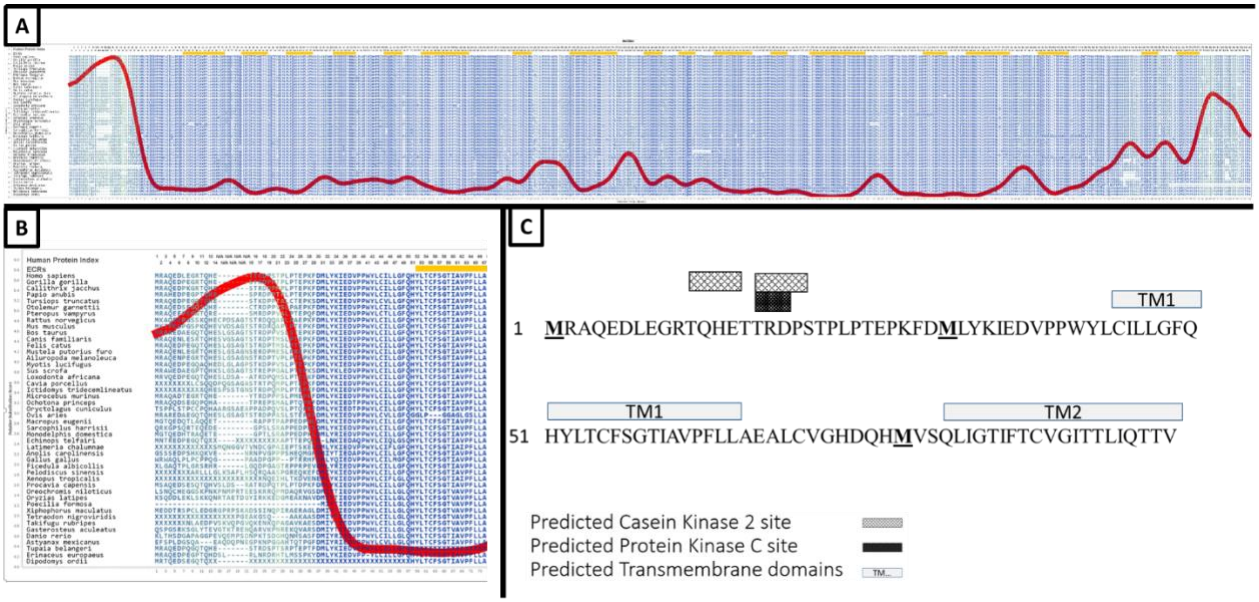


Figure 3.2: Amino Acid conservation in the SLC23A1 protein. Relative amino acid substitution scores for entire SLC23A1 protein (panel A) and its N-terminus (panel B), as calculated and aligned using the Aminode conservation browser. Scores are calculated from 62 vertebrae proteins (**Supplemental Table 2**); higher scores indicate lower conservation. The predication was done online by a software named Aminode, March 7th 2022 (http://www.aminode.org/search) Panel C: Features of the 100 N-terminal amino acids of the SLC23A1 protein.

When expressed in *Xenopus laevis* oocytes, the intracellular accumulation of ascorbic acid was identical when the first two translation initiation codons were utilized, demonstrating redundancy of the first 30 amino acids (**Figure 3.3**). Further truncation of the protein to the third potential translational start site (the elimination of 77 amino acids) resulted in a 30% reduction of the transport activity.

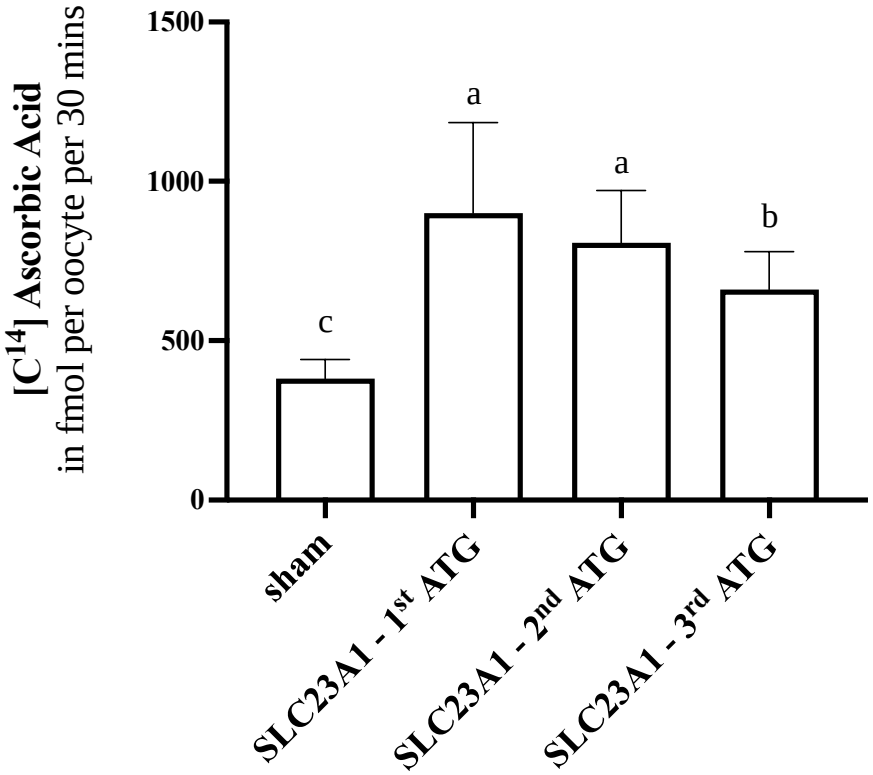


Figure 3.3: Intracellular accumulation of radiolabeled ascorbic acid in oocytes expressing the full length human SLC23A1 protein (1st ATG, currently annotated as canonical) and the two truncated forms utilizing the following downstream translation initiation codons. Oocyte were incubated in 0.045 mmol/L C^{14}-ascorbic acid for 30 mins at room temperature. Error bars represent the standard deviations; small letters indicate the statistical differences. The same small letter indicated there is no statistically significant difference between these groups and the different small letters indicate the significant difference between each group (each group contains > 15 oocytes, $p<0.0001$, ANOVA/ Bartlett's test/ Brown-Forsythe test to determine the statistically significant difference)

Discussion

The lack of evolutionary conservation of the proximal nucleotide coding sequence and the corresponding first 30 amino acids of the *SLC23A1* protein indicates a high degree of redundancy of this region. The experimental data validate this redundancy on a functional basis. There is one limitation to be considered for the *Xenopus laevis* oocytes expression system; compared to mammalian cells, the *Xenopus laevis* oocytes might not express the complement of kinases which was needed for the protein's signaling and regulation. The proximal human *SLC23A1* N-terminus contains three predicted kinase sites (**Figure 3.2 C**) and their presence or

elimination might not be impacting activation in the *Xenopus laevis* oocytes, but may in mammalian cells. However, the predicted kinase sites are not conserved, as demonstrated by a comparison to the mouse protein (**Figure 3.4**), indicating that they are nonfunctional and therefore not relevant for the human SLC23A1 protein activation.

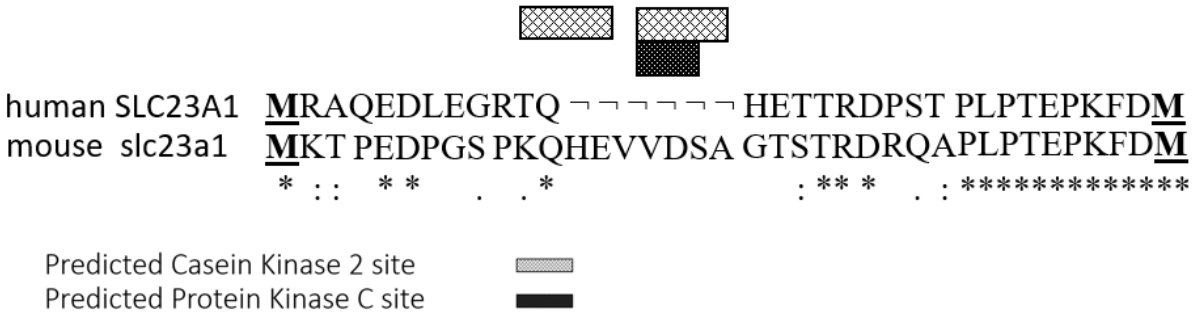


Figure 3.4: Alignment of the N-termini of the human and mouse SLC23A1 proteins and the predicted kinase sites. The CLUSTLW alignment is depicted. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties (roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix. A . (period) indicates conservation between groups of weakly similar properties (roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix). The absence of a sign indicates no conservation.

We propose that our analysis strongly indicates that variations changing the first 30 N-terminal amino acids can be deprioritized for functional studies. Since a termination of up to 77 amino acids resulted in a moderate reduction in activity, variations in that region might not have a very high priority for functional assessment. To date, there are 440 missense single nucleotide variations in the canonical *SLC23A1* transcript ENST00000348729.8 (which is equivalent to NCBI transcript NM_005847.5) reported in the Ensemble database (**Supplemental Table 3**). Among these variations, 28 are before the 2nd start codon, and 74 variations are located before the 3rd start codon. All of these can be considered of low priority for functional testing. Therefore, 16.8% of missense variations in *SLC23A1* ORF can be deprioritized for functional testing.

The presented results also suggest that the three first in-frame ATG codons can be utilized as alternative translation initiation sites for the *SLC23A1* transcript. In eukaryotes, the utilization of ribosomal translation-initiation sites which use downstream in-frame ATG codons, is documented and common (Yeom, Ju, Choi, Paek, & Lee, 2017). However, very little

information exists on the functionality of truncated proteins and the information on the *SLC23A1* truncations will therefore contribute valuable baseline information for future investigations.

The Kozak consensus sequence is a strong indicator on the likelihood of an ATG start codon to be utilized by the ribosomal pre-initiation complex (Kozak, 2002). All three ATG codons in *SLC23A1* contain some elements of the Kozak consensus sequences (**Figure 3.5**). However, the annotated canonical coding sequence contains a very strong recognition sequence, while the other two ATGs are still adequately recognized, containing at least one consensus element. Considering this, it is somewhat surprising that there was no functional difference between the first two sites when expressed in *Xenopus laevis* oocytes. We therefore suggest that additional investigations would be needed for the strength of the translational initiations. This may validate the contributions of individual nucleobases in the Kozak consensus sequence.

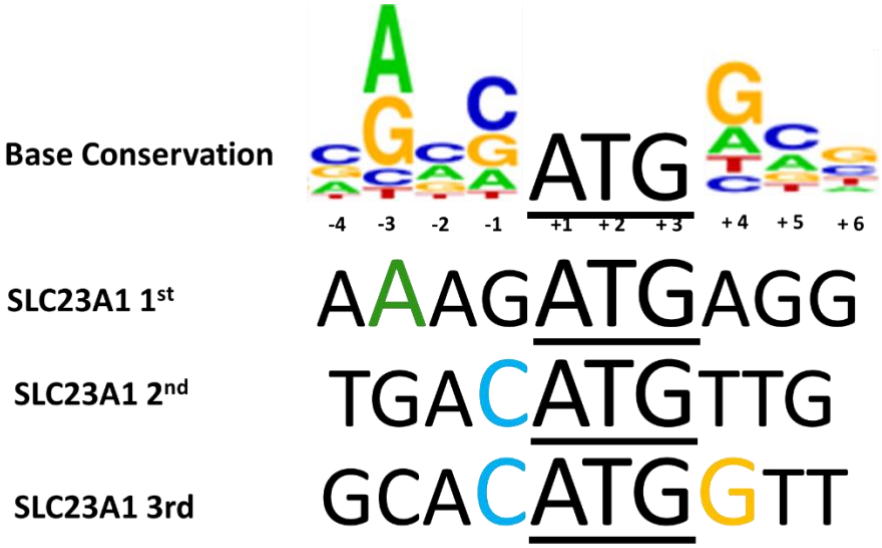


Figure 3.5: Nucleobases surrounding the first three *SLC23A1* ATG translation initiation codons and the base conservation in the Kozak consensus sequence. Variations within the Kozak sequence determine the degree of recognition by the ribosomal pre-initiation complex and the amount of protein translated (Kozak, 1984). For a 'strong' consensus, the nucleotides must contain a G in +4 and an A or G in −3. An 'adequate' consensus has only 1 of these sites, while a 'weak' consensus has neither. The cc at −1 and −2 are contribute to the overall strength (Kozak, 1986).

Variations affecting all ATG codons are reported in current databases (**Supplemental Table 3)**. The rs754063849 A/G change leads to a loss of the start codon of the canonical annotated coding sequence. The rs758202251 A/G change is leading to a change of methionine 31 to threonine, however, this would also change this cryptic ATG to a loss of the start codon. Similar, the rs760995469 T/C change changes methionine 78 to valine but would also change this ATG to a loss of the start codon. In practice, the evaluation of the clinical impact of start-loss mutations is very complex and must consider alternative possibilities. The presented functional data indicate that a loss of the first three putative translation initiation codons in *SLC23A1* can be compensated through the utilization of the alternatives. Although the population frequencies of the reported single nucleotide variations are very low, this should be considered in future clinical evaluations of *SLC23A1* variations.

Conclusion

In conclusion, the 30 N-terminal amino acids of the SLC23A1 protein are redundant, and the elimination of the first 77 N-terminal amino acids only lead to a moderate reduction in activity. Variations in these regions should therefore be of low penetration and can be deprioritized for functional studies and clinical considerations.

Reference

Amir Shaghaghi, M., Yurkova, N., Tu, H., Levine, M., & Eck, P. (2014). Identification and functional characterization of an alternative 5' exon of the sodium dependent ascorbic acid transporter SLC23A1 (818.5). *The FASEB Journal*, *28*(S1). https://doi.org/10.1096/fasebj.28.1_supplement.818.5

Corpe, C. P., Tu, H., Eck, P., Wang, J., Faulhaber-Walter, R., Schnermann, J., … Levine, M. (2010). Vitamin C transporter Slc23a1 links renal reabsorption, vitamin C tissue accumulation, and perinatal survival in mice. *Journal of Clinical Investigation*, *120*(4), 1069–1083. https://doi.org/10.1172/JCI39191

Daruwala, R., Song, J., Koh, W. S., Rumsey, S. C., & Levine, M. (1999). Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2. *FEBS Letters*, *460*(3), 480–484. https://doi.org/10.1016/S0014-5793(99)01393-9

Diallinas, G., Gorfinkiel, L., Arst, H. N., Cecchetto, G., & Scazzocchio, C. (1995). Genetic and molecular characterization of a gene encoding a wide specificity purine permease of Aspergillus nidulans reveals a novel family of transporters conserved in prokaryotes and eukaryotes. *Journal of Biological Chemistry*, *270*(15), 8610–8622. https://doi.org/10.1074/jbc.270.15.8610

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. https://doi.org/10.1093/nar/gkaa942

Kobylecki, C. J., Afzal, S., Smith, G. D., & Nordestgaard, B. G. (2015). Genetically high plasma vitamin C, intake of fruit and vegetables, and risk of ischemic heart disease and all-cause mortality: A Mendelian randomization study. *American Journal of Clinical Nutrition*, *101*(6), 1135–1143. https://doi.org/10.3945/ajcn.114.104497

Kozak, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature*, *308*(5956), 241–246. https://doi.org/10.1038/308241a0

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, *44*(2), 283–292. https://doi.org/10.1016/0092-8674(86)90762-2

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, *299*(1–2), 1–34. https://doi.org/10.1016/S0378-1119(02)01056-9

Ravindran, R. D., Sundaresan, P., Krishnan, T., Vashist, P., Maraini, G., Saravanan, V., … Fletcher, A. E. (2019). Genetic variants in a sodium-dependent Vitamin C transporter gene and age-related cataract. *British Journal of Ophthalmology*, *103*(9), 1223–1227. https://doi.org/10.1136/bjophthalmol-2018-312257

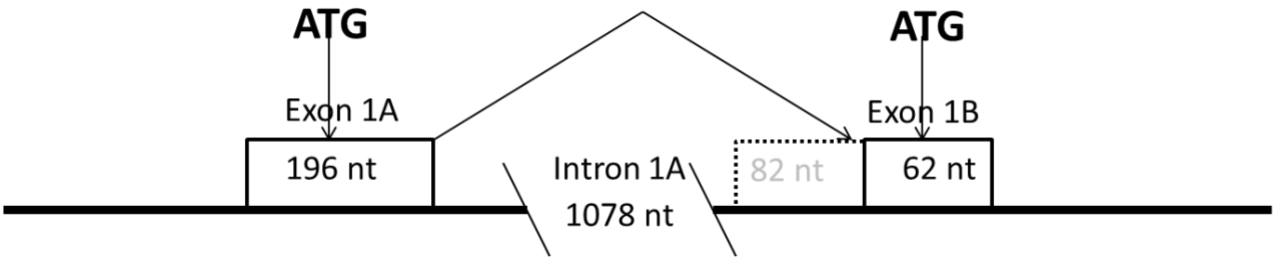
Senthilkumari, S., Talwar, B., Dharmalingam, K., Ravindran, R. D., Jayanthi, R., Sundaresan, P., … Fletcher, A. E. (2014). Polymorphisms in sodium-dependent vitamin C transporter genes and plasma, aqueous humor and lens nucleus ascorbate concentrations in an ascorbate depleted setting. *Experimental Eye Research*, *124*, 24–30. https://doi.org/10.1016/j.exer.2014.04.022

Timpson, N. J., Forouhi, N. G., Brion, M. J., Harbord, R. M., Cook, D. G., Johnson, P., … Luan, J. (2013). Genetic variation at the SLC23A1 locus is associated with circulating concentrations of L-ascorbic acid (vitamin C): Evidence from 5 independent studies with > 15,000 participant (American Journal of Clinical Nutrition (2010) 92 (375-382)). *American Journal of Clinical Nutrition*, *98*(1), 253–254. https://doi.org/10.3945/ajcn.113.065342

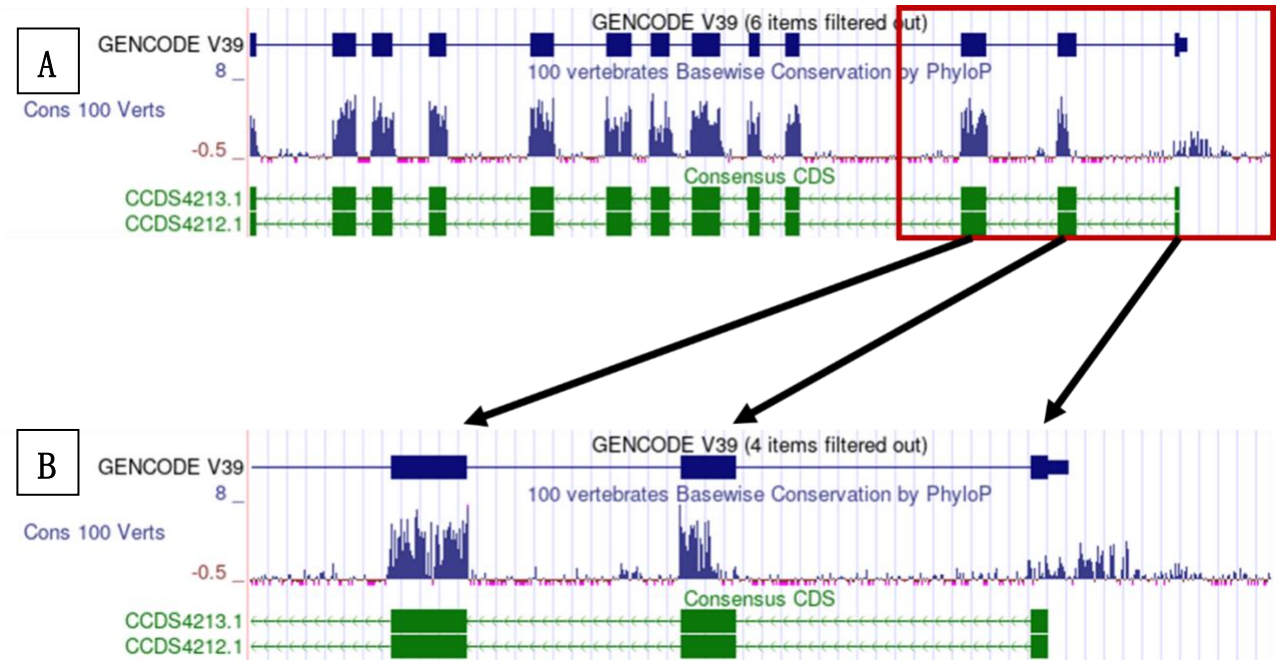
Wang, H., Dutta, B., Huang, W., Devoe, L. D., Leibach, F. H., Ganapathy, V., & Prasad, P. D. (1999). Human Na+-dependent vitamin C transporter 1 (hSVCT1): Primary structure, functional characteristics and evidence for a non-functional splice variant. *Biochimica et Biophysica Acta - Biomembranes*, *1461*(1), 1–9. https://doi.org/10.1016/S0005-2736(99)00182-0

Yeom, J., Ju, S., Choi, Y. J., Paek, E., & Lee, C. (2017). Comprehensive analysis of human protein N-termini enables assessment of various protein forms. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-06314-9

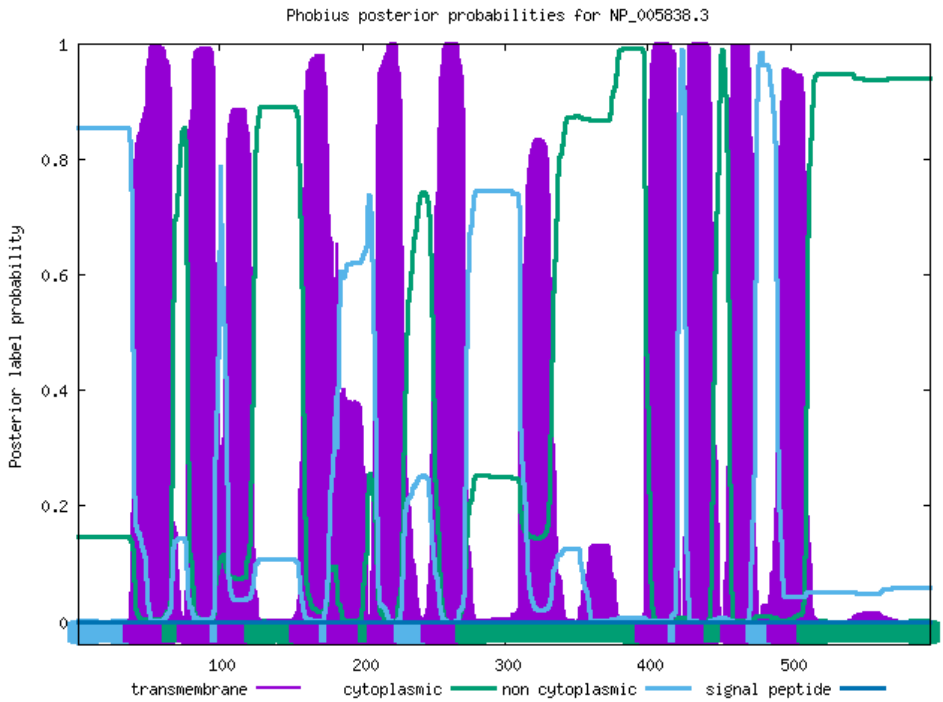
Supplemental Information



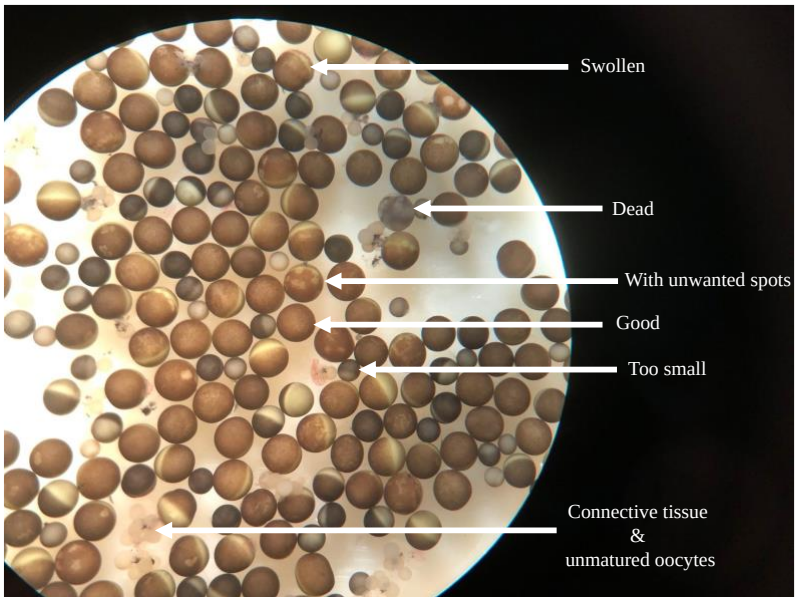
Supplemental Figure 1: the two alternative exons utilized in the human SLC23A1 locus.



Supplemental Figure 2: Nucleotide conservation in the 5' of the SLC23A1 coding region for the transcript NM_005847.5. PhyloP analysis is depicted through the UCSC genome browser conservation track calculated from 100 mammals (PhyloP Cons100 Verts as of 03-11-2022). The height of the bar indicates conservation levels of individual nucleotides. Panel A depicts the first 12 exons and panel B depicts the first three exons. Note: very low evolutionary conservation is observed for the proximal 80 bases of the coding sequence (CCDS).



Supplemental Figure 3: Transmembrane and signal peptide predicted for the SLC23A1 protein NP_005838.3. Analysis via the Phobius web interphase: https://phobius.sbc.su.se/cgi-bin/predict.pl (March 7th 2022).



Supplemental Figure 4: An image of the oocytes after digestion. The oocytes were transfered to a 60 mm sterilized petri-dish with OR_2 buffer. Only the good oocytes were selected for cRNA injection, and all the other oocytes or connective tissue in the plate were removed with 1.5 mL Pasteur pipette (Sigma Aldrich) or 5mL transfer pipette (Sigma Aldrich). The good oocytes as displayed in the picture were transfered into a new 60 mm petri-dish and incubated with OR_2 buffer with antibiotic at 16°C in the incubator.

Supplemental Sequence 1:

>NM_152685.4 Homo sapiens solute carrier family 23 member 1 (SLC23A1), transcript variant 2, mRNA

ATCCCCTCTTCTCCTCAGGAACTGCTCAAACCTGTGCCCCAAAGATGAGGGCCCAGG
AGGACCTCGAGGGCCGGACACAGCATGAAACCACCAGGGACCCCTCGACCCCGCTA
CCCACAGAGCCTAAGTTTGACATGTTGTACAAGATCGAGGACGTGCCACCTTGGTAC
CTGTGCATCCTGCTGGGCTTCCAGCACTACCTGACATGCTTCAGTGGTACCATCGCC
GTGCCCTTCCTGCTGGCTGAGGCGCTGTGTGTGGGCCACGACCAGCACATGGTTAGT
CAGCTCATCGGCACCATCTTCACGTGCGTGGGCATCACCACTCTCATCCAGACCACC
GTGGGCATCCGGCTGCCGCTGTTCCAGGCCAGTGCCTTTGCATTTCTGGTTCCAGCC
AAAGCCATACTGGCTCTGGAGAGATGGAAATGCCCCCCGGAAGAGGAGATCTACGG
TAACTGGAGTCTGCCCCTGAACACCTCTCATATTTGGCACCCACGGATACGGGAGGT
CCAGGGTGCAATCATGGTGTCCAGCGTGGTGGAGGTGGTGATTGGCCTGCTGGGGCT
GCCTGGGGCCCTGCTCAACTACATTGGGCCTCTCACAGTCACCCCCACTGTCTCCCT
CATTGGCCTTTCTGTCTTCCAAGCTGCTGGCGACCGAGCTGGCTCCCACTGGGGCAT
CTCAGCTTGCTCCATTCTCCTGATCATCCTCTTCTCCCAGTACCTGCGCAACCTCACC
TTCCTGCTGCCTGTCTACCGCTGGGGCAAGGGCCTCACTCTCCTCCGCATCCAGATCT
TCAAAATGTTTCCTATCATGCTGGCCATCATGACCGTGTGGCTGCTCTGCTATGTCCT
GACCTTGACAGACGTGCTGCCCACAGACCCAAAAGCCTATGGCTTCCAGGCACGAA
CCGATGCCCGTGGTGACATCATGGCTATTGCACCCTGGATCCGCATCCCCTACCCCT
GTCAGTGGGGCCTGCCCACGGTGACTGCGGCTGCTGTCCTGGGAATGTTCAGCGCCA
CTCTGGCAGGCATCATTGAGTCCATCGGAGATTACTACGCCTGTGCCCGCCTGGCTG
GTGCACCACCCCCTCCAGTACATGCTATCAACAGGGGCATCTTCACCGAAGGCATTT
GCTGCATCATCGCGGGGCTATTGGGCACGGGCAACGGGTCCACCTCGTCCAGTCCCA
ACATTGGCGTCCTGGGAATTACCAAGGTGGGCAGCCGGCGCGTGGTGCAGTATGGT
GCGGCTATCATGCTGGTCCTGGGCACCATCGGCAAGTTCACGGCCCTCTTCGCCTCG
CTCCCTGACCCCATCCTGGGGGGCATGTTCTGCACTCTCTTTGGCATGATTACAGCTG
TGGGGCTGTCCAACCTGCAATTTGTGGACATGAACTCCTCTCGCAACCTCTTCGTGCT
GGGATTTTCCATGTTCTTCGGGCTCACGCTGCCCAATTACCTGGAGTCCAACCCTGG
CGCCATCAATACAGGCATTCTTGAAGTGGATCAGATTCTGATTGTGCTGCTGACCAC
GGAGATGTTTGTGGGCGGGTGCCTTGCTTTCATACTTGACAACACAGTGCCAGGGAG
CCCAGAGGAGCGTGGTCTGATACAGTGGAAAGCTGGGGCTCATGCCAACAGTGACA
TGTCTTCCAGCCTCAAGAGCTACGATTTCCCCATTGGGATGGGCATAGTAAAAAGAA
TTACCTTTCTGAAATACATTCCTATCTGCCCAGTCTTCAAAGGATTTTCTTCAAGTTC
AAAAGATCAGATTGCAATTCCAGAAGACACTCCAGAAAATACAGAAACTGCATCTG
TGTGCACCAAGGTCTGAAAAATGACTTCCAGGAAAGGAAGCATGGTATATAACAGG
AAAAGAAAACTACATGGGGAACCAGAAGACCTAAGCCTGAAATCCCAGCCCTGCCC
CTAACTAACTTCTGTGTAAACTCAGATAAGTCACCTTTCTCTGGGATTCAAATTTTTG
CATCAGTTAAAAAAAAAGGGGTGGGGGGGAATGGGCCAAAGTCTGAGTCTTAGAGA
CTTGTACCAATGTTATGCTATGTCTCTAAATCTTTACTCTCCTAAGTAGACTTGTCAG
CATCTAGGAAGAACAGCTAGAAATTTTCCTCTGTGATATTTTAGACTGCAAGTTGAA
AAAAATAAAAAGAAATGAGGGCAGGTTCCAGGGCCTGAAATGTAGGTATGCTGCAA
GGCTTTTACATTGAATTTGACCCTACATCACTTCAAGACTAATGCATAATATTAAAC
ATCATGTTGAAGAAATAA

Supplemental Sequence 2:

>Coding sequence of NM_005847.5 Homo sapiens solute carrier family 23 member 1 (SLC23A1), transcript variant 1, mRNA

ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACACAGCATGAAACCACCAGGGACC
CCTCGACCCCGCTACCCACAGAGCCTAAGTTTGACATGTTGTACAAGATCGAGGACG
TGCCACCTTGGTACCTGTGCATCCTGCTGGGCTTCCAGCACTACCTGACATGCTTCA
GTGGTACCATCGCCGTGCCCTTCCTGCTGGCTGAGGCGCTGTGTGTGGGCCACGACC
AGCACATGGTTAGTCAGCTCATCGGCACCATCTTCACGTGCGTGGGCATCACCACTC
TCATCCAGACCACCGTGGGCATCCGGCTGCCGCTGTTCCAGGCCAGTGCCTTTGCAT
TTCTGGTTCCAGCCAAAGCCATACTGGCTCTGGAGAGATGGAAATGCCCCCCGGAA
GAGGAGATCTACGGTAACTGGAGTCTGCCCCTGAACACCTCTCATATTTGGCACCCA
CGGATACGGGAGGTCCAGGGTGCAATCATGGTGTCCAGCGTGGTGGAGGTGGTGAT
TGGCCTGCTGGGGCTGCCTGGGGCCCTGCTCAACTACATTGGGCCTCTCACAGTCAC
CCCCACTGTCTCCCTCATTGGCCTTTCTGTCTTCCAAGCTGCTGGCGACCGAGCTGGC
TCCCACTGGGGCATCTCAGCTTGCTCCATTCTCCTGATCATCCTCTTCTCCCAGTACC
TGCGCAACCTCACCTTCCTGCTGCCTGTCTACCGCTGGGGCAAGGGCCTCACTCTCC
TCCGCATCCAGATCTTCAAAATGTTTCCTATCATGCTGGCCATCATGACCGTGTGGCT
GCTCTGCTATGTCCTGACCTTGACAGACGTGCTGCCCACAGACCCAAAAGCCTATGG
CTTCCAGGCACGAACCGATGCCCGTGGTGACATCATGGCTATTGCACCCTGGATCCG
CATCCCCTACCCCTGTCAGTGGGGCCTGCCCACGGTGACTGCGGCTGCTGTCCTGGG
AATGTTCAGCGCCACTCTGGCAGGCATCATTGAGTCCATCGGAGATTACTACGCCTG
TGCCCGCCTGGCTGGTGCACCACCCCCTCCAGTACATGCTATCAACAGGGGCATCTT
CACCGAAGGCATTTGCTGCATCATCGCGGGGCTATTGGGCACGGGCAACGGGTCCA
CCTCGTCCAGTCCCAACATTGGCGTCCTGGGAATTACCAAGGTGGGCAGCCGGCGC
GTGGTGCAGTATGGTGCGGCTATCATGCTGGTCCTGGGCACCATCGGCAAGTTCACG
GCCCTCTTCGCCTCGCTCCCTGACCCCATCCTGGGGGGCATGTTCTGCACTCTCTTTG
GCATGATTACAGCTGTGGGGCTGTCCAACCTGCAATTTGTGGACATGAACTCCTCTC
GCAACCTCTTCGTGCTGGGATTTTCCATGTTCTTCGGGCTCACGCTGCCCAATTACCT
GGAGTCCAACCCTGGCGCCATCAATACAGGCATTCTTGAAGTGGATCAGATTCTGAT
TGTGCTGCTGACCACGGAGATGTTTGTGGGCGGGTGCCTTGCTTTCATACTTGACAA
CACAGTGCCAGGGAGCCCAGAGGAGCGTGGTCTGATACAGTGGAAAGCTGGGGCTC
ATGCCAACAGTGACATGTCTTCCAGCCTCAAGAGCTACGATTTCCCCATTGGGATGG
GCATAGTAAAAAGAATTACCTTTCTGAAATACATTCCTATCTGCCCAGTCTTCAAAG
GATTTTCTTCAAGTTCAAAAGATCAGATTGCAATTCCAGAAGACACTCCAGAAAATA
CAGAAACTGCA

Supplemental Sequence 3

>NP_005838.3 solute carrier family 23 member 1 isoform a [Homo sapiens]

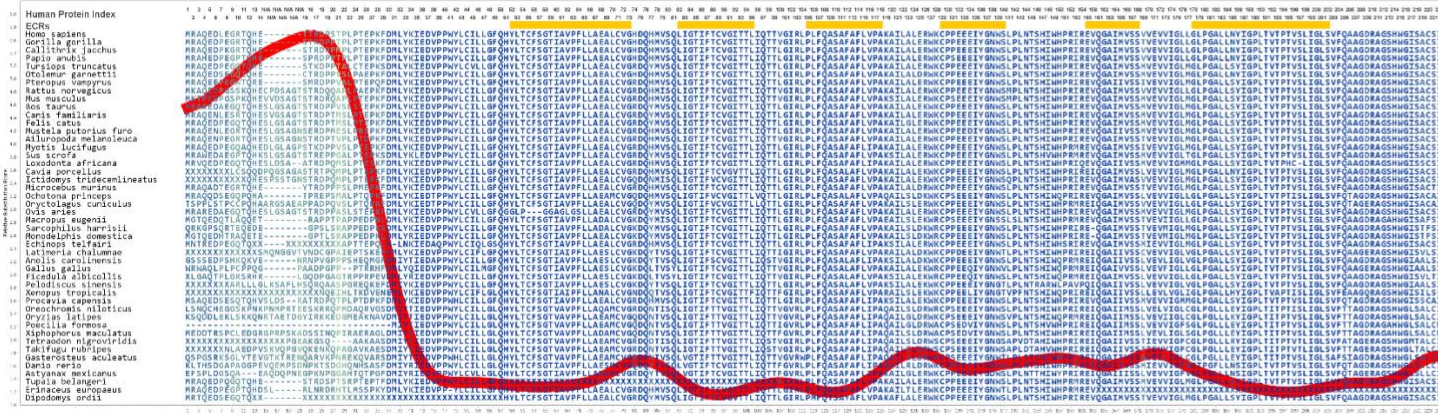
MRAQEDLEGRTQHETTRDPSTPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTI
AVPFLLAEALCVGHDQHMVSQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKAIL
ALERWKCPPEEEIYGNWSLPLNTSHIWHPRIREVQGAIMVSSVVEVVIGLLGLPGALLNY
IGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSILLIILFSQYLRNLTFLLPVYRWGKGL
TLLRIQIFKMFPIMLAIMTVWLLCYVLTLTDVLPTDPKAYGFQARTDARGDIMAIAPWIR
IPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGI
CCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLVLGTIGKFTALFASLPDPI
LGGMFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIL
EVDQILIVLLTTEMFVGGCLAFILDNTVPGSPEERGLIQWKAGAHANSDMSSSLKSYDFP
IGMGIVKRITFLKYIPICPVFKGFSSSSKDQIAIPEDTPENTETASVCTKV

Supplemental Table 1: Transmembrane and signal peptide predicted for the SLC23A1 protein NP_005838.3. Analysis via the Phobius web interphase: https://phobius.sbc.su.se/cgi-bin/predict.pl (March 7th 2022). Numbers indicate the amino acids positions.

Prediction of NP_005838.3

FT	TOPO_DOM	1	39	NON CYTOPLASMIC.
FT	TRANSMEM	40	66	
FT	TOPO_DOM	67	77	CYTOPLASMIC.
FT	TRANSMEM	78	100	
FT	TOPO_DOM	101	105	NON CYTOPLASMIC.
FT	TRANSMEM	106	124	
FT	TOPO_DOM	125	155	CYTOPLASMIC.
FT	TRANSMEM	156	176	
FT	TOPO_DOM	177	181	NON CYTOPLASMIC.
FT	TRANSMEM	182	204	
FT	TOPO_DOM	205	210	CYTOPLASMIC.
FT	TRANSMEM	211	229	
FT	TOPO_DOM	230	248	NON CYTOPLASMIC.
FT	TRANSMEM	249	272	
FT	TOPO_DOM	273	398	CYTOPLASMIC.
FT	TRANSMEM	399	421	
FT	TOPO_DOM	422	426	NON CYTOPLASMIC.
FT	TRANSMEM	427	446	
FT	TOPO_DOM	447	457	CYTOPLASMIC.
FT	TRANSMEM	458	475	
FT	TOPO_DOM	476	490	NON CYTOPLASMIC.
FT	TRANSMEM	491	511	
FT	TOPO_DOM	512	598	CYTOPLASMIC.

Supplemental Table 2: Amino Acid conservation scores as calculated using the Aminode web browser (http://www.aminode.org/?gene=SLC23A1) and sequences form 47 vertebrae utilised for the calculations and alignments.



sequences aligned:

>Homo_sapiens ENSG00000170482|ENST00000353963|ENSP00000302851
MRAQEDLEGRTQHETTRDPSTPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GHDQHMVSQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSLPLNTS
HIWHPRIREVGLHVQGAIMVSSVVEVVIGLLGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIS
ACSILLIILFSQYLRNLTFLLPVYRWGKGLTLLRIQIFKMFPIMLAIMTVWLLCYVLTLTDVLPTDPKAYGFQ
ARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGI
FTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLVLGTIGKFTALFASLPDPILGGMFCT
LFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGILEVDQILIVLLTTEMFVGGCL
AFILDNTVPGSPEERGLIQWKAGAHANSDMSSSLKSYDFPIGMGIVKRITFLKYIPICPVFKGFSSSSKDQIAIP
EDTPENTETASVCTKV*

>Oreochromis_niloticus ENSONIG00000017809|ENSONIT00000022469|ENSONIP00000022449
LSNQCHEGRGSKPNKPNMPRTEESKRRQPMDAQRVGSDMIYTIEDVPPWYLCILLGLQHYLTCFSGTVAVP
FLLAQAMCVGRDQNTISQLIGTIFTTVGITTLIQTTVGVRLPLFQASAFAFLIPAQAILSLDRWRCPSEEEIYG
NWSLPLNTSHIWKPRIREIQGAIIMSSLVELVIGLCGLPGLLLEYIGPLTITPTISLIGLSVFTTAGDRAGSHWG
LSALCILLIVLFAQYLRATSLPVPVYSRKKGLRATRVQIFKMFPIILAIMLVWLVCYVLTLTDLLPRDPDRYG
HKGRTDARGDIMTLSPWFRMPYPCQWGLPVITVPGVLGMLSATMAGIVESIGDYYACARLSGATPPPVHAI
NRGIFIEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMFVLGTVGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQLVDLNSSRNLFVLGFSMFFGLTLPTYLDTHPNSIQTGVSELDQILTVLLSTEMFV
GGFLAFCLDNTIPGTREERGLVEWNFSSSSSSSSTYDLPFVMPVLRRTRCLWRFPISPTFTGFRASDDPPPVE
QEQEEEEDDDELPSTKV*

>Gasterosteus_aculeatus ENSGACG00000020918|ENSGACT00000027710|ENSGACP00000027658
MVQVPQSPGSRKSHGLYTEVGTKTRENQARVKPNREKQVARSDMIYTIEDVPPWHLCILLGLQHYLTCFSG
TVAVPFLLAEAMCVGRDPNTVSQLVGTIFTTVGITTLIQTTVGVRWPLFQASAFAFLIPAQAILSLDRWKCPS
EEEIYGNWTLPLSTSHIWHPRIREIQGAIIVSSAVEILIGLCGLPGLLLEYIGPLTITPTVSLIGLSAFSIAGDRAG
SHWGLSALCIFLIVLFAQYLRETSLPVPVYTRKKGLTSTSVQLFKMFPIILAIMLVWLICYVLTLNNLLPSNSE
SYGHKARTDARGDIVTSSPWFRVPYPCQWGLPVVTVAGVLGMLSATMTGIVESIGDYYACARLSGATPPP
VHAINRGIFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLIMGSVGKFTALFASLPD
PILGGMFCTLFGMITAVGLSNLQFVDLNSSRNLFVLGFSMFFGLTLPTYLDKHPESIKTGLPELDNILTVLLST
EMFVGGFLAFCLDNTIPGSKQERGLVEWGSSSEYDLPFGMGVVGRTRWLRRFPISPSFTGFKASAPSRLPEE
EEAPDMNPLSTKV*

>Tetraodon_nigroviridis ENSTNIG00000007658|ENSTNIT00000010654|ENSTNIP00000010473
PGEAKGSQAAKAASDMIYTIEDVPPWYLCILLGLQHYLTCFSGTVAVPFLLAEAMCVGRDQDTISQLIGTIF
TTVGITTLIQSTVGIRLPLFQASAFAFLIPAQAILSLDRWSCPSEEEIYGNGSAPVDTAHIWHPRIREIQGAIIVS
STIEVVIGFCGLPGLLLRYIGPLTITPTVTLIGLSVFATAGERAGSHWGMTALCIFLIVLFAQYLRETSIPLPYY
SRKKGLTSTRVQIFKMFPIIMAIMVVWLVCYIFTLTGLLPSDPNRYGYKARTDARGDIMTSAPWFRVPYPCQ
WGLPVVTVAGVLGMLSATMAGIVESIGDYYACARLAGAAPPPVHAINRGIFTEGVCCIIAGLLGTGNGSTS
SSPNIGVLGITKVGSRRVVQYGAGIMFLLGAVGKFTALFASLPDPILGGMFCTLFGMITAVGLSNLQLVDLN

SSRNLFVLGFSMFFGLTLPAYLDAHPKSINTVGVAELDQILTVLLSTEMFVGGFLAFCLDNTIPGTREERGLV
HWGTSSSSCSSSYDFPLGMSVVRRAGWLRRLPISPTFTGFRAPDRSASEDEKADTSVTLASTKV*
>Echinops_telfairi ENSETEG00000006082|ENSETET00000006082|ENSETEP00000004937
MNTREDPEGQTQXXXXXXXXXXXXAPTTEPQLNLNKIEDAQPWYLCIQLGSQHYLTCFSGTIAVPFLLAE
ALCVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLIPAKSILALERWKCPPEEEIYGNWSLP
LNTSHIWHSRIREVSLLVQGAIMVSSMIEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLSLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPTDPSA
YGYQARTDARGDIMAIAPWIRIPYPCQWGLFTVTAAAVGMFSALAGIIEIGDYYACALAGAPPPAVHAINR
GIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGTIGKFTALFASLPDPILGGMFC
TLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXSPEERGLIQWKAGAHANSETSASLKSYDFPIGMGMVKRTAFLKYIPICPVFKGFSKP
KAEFPVPENTPTNTETGSVCTKV*
>Papio_anubis ENSPANG00000024864|ENSPANT00000009008|ENSPANP00000000834
MRAHEDPEGPTQHESPRDPPTPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCVG
HDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPSEEEIYGNWSLPLNTSH
IWHPRIREVQGAIMVSSVVEVVIGLLGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSILLI
ILFSQYLRNLTFLLPVYRWGKGLTLLRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPTDPKAYGFQARTDAR
GDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGIC
CIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLVLGTIGKFTALFASLPDPILGGMFCTLFGMIT
AVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGILEVDQILTVLLTTEMFVGGCLAFILD
NTVPGSPEERGLIQWKAGAHANSDTSSSLKSYDFPIGMGIVKRIAFLKYIPICPVFKGFSSSSKDQFAIPEDTP
ENTETASVCTKV*
>Macropus_eugenii ENSMEUG00000007441|ENSMEUT00000007467|ENSMEUP00000006800
MGTQEDQTTLAQQETRAPPTPAPPEDPNLDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLADALCV
GKDQYMISQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAEAILSLEKWRCPSEEEIYGNSSLSLNTSHI
WHPRMREVSCRIQGAIMVSSTVEVIIGLLGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGAHWGISAFS
ILLIILFSQYLRNVTFCLPGYKWGKGFTLFRIQIFKMFPIVLAIMTVWLFCYILTLTDVLPSDPNAYGFKARTD
ARGEIMSISPWVRIPYPCQWGLPSVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGI
CCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGMIGKFTALFASLPDPILGGMFCTLFGMI
TAVGLSNLQFIDMNSSRNLFVLGFSMFFGLMLPNYLDSNPTAINTGVPEVDQILTVLLTTEMFVGGCLAFIL
DNTVPGSPEERGLIQWKAGAHANSETSASLKSYDFPIGMNTIQKMAFLKYIPICPVFKGFSPKAQEQLPAPE
DNPENTDILSLSTKV*
>Pteropus_vampyrus ENSPVAG00000015433|ENSPVAT00000015433|ENSPVAP00000014553
MRAQEEPEAGWTQRESMRDPPVALPTEPQFDMLYKIEDVPPWYLCTLLGFQHYLTCFSGTIAVPFLLAEAL
CVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSLPLN
TSHIWHPRIREVGLHVQGAIIVSSIVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIS
ACSILLIILFSQYLRNLNLLLPVYRWGKGLTLFRIQIFKMFPIVLAIMIVWLLCYLLTLTDVLPTDPTAYGFHA
RTDARGDIMGITPWIRISYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIF
IEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGACIMLVLGTIGKFTALFASLPDPILGGMFCTLF
GMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIPEVDQILTVLLTTEMFVGGCLA
FILDNTVPGSPKERGLIQWKAGAHANSEMSTSLKSYDFPIGMKMIKRIAFLKYIPICPVFKGFSSRSKDQLPG
PEDTPENKETRPVCTKV*
>Dipodomys_ordii ENSDORG00000002270|ENSDORT00000002270|ENSDORP00000002131
MRTQEDSEGQTQXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXHYLTCFSGTIAVPFLLA
EALCVGRDQYMVSQLIGTIFTCVGVTTLIQTTLGIRLPMFQTSAYAFLVPAKAFLALERWKCPPEEEIYGNW
SLPLNTSHIWHPRIREVVQGAIMVSSTVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPADPTA
YGFQARTDARGDIMSISPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAI
NRXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XIPEVDQILTVLLTTEMFVGGCLAFILDNTVPGSPEERGLIQWKAGAHASSATSASLKSYDFPIGMGIVKRIA
FLKYIPVCPVFRGFSTRSKSQPSVPEDTPENIEIGSVCTKV*
>Poecilia_formosa ENSPFOG00000018676|ENSPFOT00000018812|ENSPFOP00000018790
MIYTIEDVPPWYLCILLGLQHYLTCFSGTVAVPFLLAEAMCVGQDQNTVSQLIGTIFTTVGITTLIQTTIGVRL
PLFQASAFAFLIPAQAILGLDRWACPSEDVIYGNWSLPLNTSHVWHPRIREIQGAIIMSSLVEVVIGFSGLPGL
LLEYIGPLTITPTVSLIGLSVFTTAGDRAGAHWGLSALCMLLIVLFAQYLRTTSIPVPVYSRRKGLTSTRVQIF

KMFPIILAIMLVWLVCYILTLTDLLPNDPDRYGHKGRTDARGDIMASAPWFRMPYPCQWGLPVVTVAGVL
GMLSATMAGIVESIGDYYACARLSGATPPPVHAINRGIFTEGICCVIAGLLGTGNGSTSSSPNIGVLGITKVGS
RRVVQFGAGIMFILGTVGKFTALFASLPDPILGGMFCTLFGMITAVGLSNLQLVDLNSSRNLFVLGFSMFFG
LTLPAYLDAHPKVIKTGVVELDQILTVLLSTEMFVGGFLAFCLDNTIPGTREERGLVQWSSSSTAPTSGPARS
PGSTSYDFPVGMGVVRRTRWLRRFPICPTFTGFRVREEDLPPEGEEMAGQGEELPDASLPSTKV*
>Anolis_carolinensis ENSACAG00000016325|ENSACAT00000016427|ENSACAP00000016106
MGSSSEDPSHHKQKVENRNPVGPPPSHEQMGFDMIYTIEDAPPWYLCILLGFQHYLTCFSGTIAVPFLLAES
LCVGKDQYTVSQLIGTIFSCVGITTLIQSTVGIRLPLFQASALAFLIPAKSILALDKWKCPPEEEIYGNWSLPL
NTSHIWQPRMREIQGAIIVSSLVEVLIGLVGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIAAL
SIVLIILFAQYLRNVSFLLPGYKCGKGCTVFRIQIFKMFPIILAILVVWLLCYILTVTDVFPRDANAYGFKART
DARGEIISIAPWFRFPYPCQWGIPTVTAAAVLGMFSATLSGIIESIGDYYSCARLAGAPPPPVHAINRGIFTEGI
SCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRKVVQYGAGIMLILGTIGKFTALFASLPDPILGGMFCTLFGMI
TAVGLSNLQFVDMNSSRNLFVLGFAMFFGLTLPNYLDSHPDAIDTGIGEVDQILKVLLTTEMFVGGGIAFIL
DNTVPGTEKERGLIQWKAGAHANSDTSAKLKSYDFPFGMNVIRRTWWLKYVPVCPGFMGFNFKARRDLD
ISEHIPTVTTIPAVCTKV*
>Procavia_capensis ENSPCAG00000006953|ENSPCAT00000006954|ENSPCAP00000006505
MSAQEDSESQTQHVSLDSRATRDPQTPLPTDPKFDMLYKIEDVPPWHLCILLGFQHYLTCFSGTIAVPFLLA
EALCVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLIPAKSILALEKWKCPPEEEIYGNWSL
PLNTSHIWHPRIREVGLHVQGAIIVSSMVEVVIGMMGLPGALLSYIGPLTVTPTVSLIGLSVFQTAGDRAGSH
WGISSCAILLIILFSQYLRNLTFVLPVYRWGKGLTLFRIQIFKMFPIMLAIMTVLLCYXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLGMFSATLAGIIQSIGDYYACARLAGAPPP
PVHAINRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGSIGKFTALFASLPDP
ILGGMFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPDVINTXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXSPQERGLIQWKAGAHANSEMSTSLKSYDFPIGMGLVKRIAFLKYIPICPVF
KGFLSRSKGEIPVPEDTPENTETAFVSTKV*
>Rattus_norvegicus ENSRNOG00000061695|ENSRNOT00000078202|ENSRNOP00000073343
MKQLPSWKAMLPVGPWLSLCPCAPDGTRAQQLRMKELTHAAPEGRCDSAYTNTEISQTRVSEMKAQEDP
GSSKQHECPDSAGTSTRDQQAPLPAEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GRDQHMISQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSMPLNTS
HIWHPRIREVQGAIMVSSVVEVVIGLLGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSIL
LIVLFSQYLRNLTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPADPTVYGFQARTD
ARGDIMAISPWIRIPYPCQWGLPTVTVAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEG
VCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGAIGKFTALFASLPDPILGGMFCTLFG
MITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPGAINTGVPEVDQILTVLLTTEMFVGGCLAF
ILDNTVPGSPEERGLIQWKAGAHANSETLASLKSYDFPFGMGMVKRTTFFRYIPICPVFRGFSKTENQPAVL
EDAPDNTETGSVCTKV*
>Cavia_porcellus ENSCPOG00000014082|ENSCPOT00000014224|ENSCPOP00000012685
LCSQQDPQGSAGASTRTPQMPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCVG
RDQHMVSQLIGTIFTCVGVTTLIQTTLGIRLPLFQASALAFLVPAKAILALEKWKCPPEEEIYGNWSLPLNTS
HIWHPRIREIQGAIMVSSLVEVVIGLMGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSIL
LIVLFSQYLRNVTFLLPGYRWGKGLTFFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPVDPTDYGFQARTD
ARGDIITISPWVRIPYPCQWGVPTVTMAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGI
CCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAVIMLVLGAIGKFTALFASLPDPILGGMFCTLFGM
ITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLMLPNYLDSNPGAINTGILEVDQILTVLLTTEMFVGGCLAFIL
DNTVPGSPEERGLLQWKAGAHANSETSASLKSYDFPIGMDIVKKIAFLKYIPICPVFKGLSLKSKNKPPVLED
TPENTEAASVCTKV*
>Tursiops_truncatus ENSTTRG00000003747|ENSTTRT00000003746|ENSTTRP00000003519
MRAQEDPEGQTQHESTKDPPTALCTEPKSDMLYKIEDVPPWYLCVLLGFQHYLTCFSGTIAVPFLLAEALC
VGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSLPLNT
SHIWHPRIREVGLHVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGI
SACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYILTLTDVLPSDPTAYGFQ
ARTDARGDIVAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGI
FTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTIGKFTALFASLPDPILGGMFCT
LFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPDAINTGIPEVDQILTVLLTTEMFVGGC
LAFILDNTVPGSPEERGLIQWKAGAHANSEMSTSLKSYDFPIGMSMVKRIAFLKYIPICPVFKGFSSRSKDQL
PVPEDTPQNTETGSVCTKI*

>Bos_taurus ENSBTAG00000010798|ENSBTAT00000010823|ENSBTAP00000010823
MRAREDAEGQTQHESLGSAGTSTRDPPVSLSTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALEKWKCPPEEEIYGN
WSLPLNTSHIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDMLPSDPTA
YGFQARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHA
INRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGTIGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPDVINTGVPEVDQILTVLLTTEMF
VGGGCLAFILDNTVPGSPEERGLIQWKAGAHADSEMSSSLKSYDFPIGMSMVKRIAFLKYIPICPVFKGFPSRS
KNQLPVPEDIPENTETVYTKV*
>Ovis_aries ENSOARG00000017065|ENSOART00000018578|ENSOARP00000018317
MRAREDAEGQTQHESLGSAGTSTRDPPASLSTEPKFDMLYKIEDVPPWYLCVLLGFQGGLPGGAGLGSLLA
EALCVGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWS
LPLNTSHIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWG
ISACSILLIILFSQYLRNLTFLLPVYRWGKGLTLFRVQIFKMFPIVLAIMTVWLLCYVLTLTDVLPSDPTAYGF
QARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINR
GIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVGQYGVLGTIGKFTALFASLPDPILGGMFCTLFG
MITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPDVINTGVPEVDQILTVLLTTEMFVGGCLAF
ILDNTVPGSPEERGLIQWKAGAHADSEMSSSLKSYDFPIGMSMVKRTAFLKYIPICPVFKGFPSRSKNQLPVP
EDIPENAETVYTKV*
>Takifugu_rubripes ENSTRUG00000009428|ENSTRUT00000023789|ENSTRUP00000023691
NGLAEDPVSKVQPGVQKENKQPAGAVKAESDMIYTIEDVPPWYLCILLGLQHYLTCFSGTVAVPFLLAEA
MCVGRDQNTVSQLIGTIFTTVGITTLIQSTVGIRLPLFQASAFAFLIPAQAILSLDRWSCPSEEEIYGNWSAPL
DTAHVWHPRIREIQGAIIVSSTIEVVIGFCGLPGLLLEYIGPLTITPTVTLIGLSVFTTAGERAGSHWGLTALCI
FLIVLFAQYLRETSIPVPFYSREKGLTSTRVQIFKMFPIILAIMVVWLVCYIFTLTNLLPSDPSRYGYKARTDA
RGDIMTSAPWFRMPYPCQWGLPVVTVAGVLGMLSATMAGIVESIGDYYACARLSGAAAPPVHAINRGIFT
EGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMFLLGAVGKFTALFASLPDPILGGMFCTL
FGMITAVGLSNLQLVDLNSSRNLFVLGFSIFFGLTLPAYLDAHPKSINTGVAELDQILTVLLSTEMFVGGFLA
FCLDNTIPGTREERGLVHWRTSSSSSSSYDFPLGMSVVRRARWLRWFPISPTFTGFQASNNPPAVKEEEVE
GADITLASTKV*
>Erinaceus_europaeus ENSEEUG00000012467|ENSEEUT00000012483|ENSEEUP00000011385
MRAQEDPEGPTQHDSLRLNRDRHTLMSSPKYDMLYKIEDVPPYLCILLGFQHYLTCFSGTIAVPFLLAEALC
VGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILTLERWKCPPEEEIYGNWTLPLNT
SHIWHPRMREVGVSXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXIPYCQWGLPTVTAAAVLGMFSATLAGIIESVGDY
YACARLAGAPPPPVHAINRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGTI
GKFTALFASLPDPILGGMFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNRGAINTG
VPEVDQILTVLLTTEMFVGGCLAFILDNTVPGTPEERGLIQWKAGAHANSETSASLQSYDFPIGMSVVKETA
FLKYIPICPVFKGFSSKSKGQRPVPEDTAENTETNSVCTKV*
>Danio_rerio ENSDARG00000015033|ENSDART00000014223|ENSDARP00000009802
MNQSETAEELKLTHSDGAVPAGGPEVQEMPSDNPKTSDGHQNHSASFDMIYRIEDVPPWYLCILLGLQHYL
TCFSGTIAVPFLLAESMCVGQDQYTVSQLVGTIFTCVGITTLIQTTFGVRLPLFQASAFAFLIPAQAILRLDRW
KCPPEEEIYGDWSLPLNTSHIWHPRIREIQGAIIVSSMIEVVIGFAGIPGFLLNSIGPLTVTPTVSLIGLSVFQTA
GDRAGSHWGLSLLCIFLIVLFAQYLRNWACPLPSFSKEKGCHITHVQIFKMFPIIMAIMVVWLVCYILTLTN
VLPDDPDLYGYKARTDARGDIMTQAPWFRFPYPCQWGLPTVTVAGVLGMFSATLAGIVESIGDYYACARL
SGAPPPPVHAINRGIFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVIQYGAGIMLILGTIGKFTALF
ASLPDPILGGMFCTLFGMITAVGLSNLQSVDLNSSRNLFVLGFSMFSGLMLPNYLDAHPGSIKTGVAELDQII
TVLLTTEMFVGGFLAFVLDNTIPGTRKERGLVEWVDEGSSGAGTVKSDTYNFPIGMGLVRKTGCLRYLPIC
PTFRGFKSSCRKNKEEEDEEDIKIKEGVTDAPIEMVCTKI*
>Xiphophorus_maculatus ENSXMAG00000006227|ENSXMAT00000006272|ENSXMAP00000006265
MEDDTRSPDCLEDGRGPRPSKADSSINQPIRAERAGLDMIYTIEDVPPWYLCILLGLQHYLTCFSGTVAVPFL
LADAMCVGQDQNTVSQLIGTIFTTVGITTLIQTTFGVRLPLFQASAFAFLIPAQAILSLDRWACPSEDVIYGN
WSLPLNTSHIWHPRIREIQGAIIMSSLVEVLIGLSGLPGLLLEYIGPLTITPTVSLIGLSVFTTAGDRAGAHWGL
SALCMLLIVLFAQYLRTTSLPVPVYSRRKGLTSTRVQIFKMFPIILAIMLVWLVCYVLTLTDLLPNDPKSYGH
KGRTDARGDIMASAPWFRMPYPCQWGLPVMTVAGVLGMLSATMAGIVESIGDYYACARLSGATPPPVHA

INRGIFTEGICCVIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQFGAGIMFILGAVGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQLVDLNSSRNLFVLGFSMFFGLTLPAYLDTHPKAIRTGVAELDQILTVLLSTEMF
VGGFLAFCLDNTIPGTREERGLVQWSSSSTAPTSGHAPSPGSASYDFPVGMGVVRRTRWLRRFPICPTFTGF
RVCEEDLPLEGEELAGQGEELGDAPLPSTKV*
>Canis_familiaris ENSCAFG00000005686|ENSCAFT00000009175|ENSCAFP00000008511
MRAQENLESRTQHESVGSAGTSTRDPTMSLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGN
WSLPLNTSHIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRVQIFKMFPIVLAIMSVWLLCYILTLTNVLPSDPTA
YGFQARTDARGDIMTIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHA
INRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLVLGTIGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIPEVDQILTVLLTTEMF
VGGCLAFILDNTVPGSLEERGLIQWKAGAHANSEMSTSLKSYDFPIGMSIVKRTAFLKYIPICPIFKGFSSRSK
AQLPVPEDPPENIQTGSACTKV*
>Astyanax_mexicanus ENSAMXG00000005734|ENSAMXT00000005863|ENSAMXP00000005863
MGPAEPVHTSSLCEGPVKHSDKAAEPEFSPLDGSMQAEAQDQPNEGPKNPGGAHTQTPGFDMIYRIEDVPP
WYLCVLLGLQHYLTCFSGTIAVPFLLADAMCVGRDQYTVSQLVGTIFTCVGITTLIQTTFGVRLPLFQASAF
AFLIPAKAILGLERWKCPPEEEIYGDWSLPLNTTHIWYPRIREIQGAIIISSLVEVVIGFVGLPGLLLNSIGPLTV
TPTVSLIGLSVFQTAGERAGSHWGLSLLCIFLIVLFAQYLRNWAFPFPFYSRQKGCHITRFQIFKMFPIIMAIM
LVWLVCYILTLTDVLPNDPSQYGYKARTDARGEIMSQAPWFRMPYPCQWGLPTVTIAGVLGMFSATLAGI
VESIGDYYACARLSGAPPPPVHAINRGIFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGALI
MFLLGTIGKFTALFASLPDPVLGGMFCTLFGMITAVGLSNLQSVDLNSSRNLFVLGFSMFSGLMLPNYLETH
PGSIKTGVPELDQILTVLLTTEMFVGGFLGFILDNTIPGKYRIQHFTLKSFKPDSYIKIHESLGYSSIMCTYSSG
KK*
>Gallus_gallus ENSGALG00000002513|ENSGALT00000003962|ENSGALP00000003953
WRWAQLPLQPCPPQGPAADPGPPPTRRHFDMLYQIEDVPPWYLCILMGFQHYLTCFSGTIAVPFLLAESLC
VGKDQLTVSYLIGTIFTCVGITTLIQTTVGIRLPLFQASALAFLVPAKAILALEKWRCPPEAPEQIYGNWVLPL
NTSHIWQPRMREIQGAIVVSSLVEVFIGLLGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGERAGSHWGIAAL
SIVLIVLFAQYLRNVTVRLPGYRWGRGFVLLRVQIFKMFPIILAIMVMWLLCYVLTRAGVFPSHPEEYGYK
ARTDARGEILSVAPWFRFPYPCQWGIPTVTSAAVLGMFSATLAGIIESGDYYSCARLSGAPAPPVHAINRGIF
TEGISCIIAGLMGTGNGSTSSSPNIGVLGITKVGSRRVIQYGAGIMLILGTIGKFTALFASLPDPILGGMFCTLF
GMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDAHPKAINTGVPELDQILTVLLTTEMFVGGTL
AFILDNTIPGTREERGLVQWKAGAHADSTASADLRSYDFPFGMSAVRRSRWLRHVPICPLFTGFVWRAKKS
HAADDGTQDGTDGTAVCTKV*
>Latimeria_chalumnae ENSLACG00000010772|ENSLACT00000012331|ENSLACP00000012239
SMQNGGVTVNDCGPAIEPTSKEFDMLYKIEDVPPWYLCIFLGFQHYLTCFSGTIAIPFLLAESLCVGKDQYT
VSQLIGTIFTCVGITTLIQTTIGIRLPLFQASAFAFLVPARAILALDKWKCPPDEEIYGNWTLPLHTSHIWHPRI
REIQGAIIVSSCVEVVIGLVGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGERAGSHWGISVLSIVLIILFAQYL
RNLQIVLPFYKWGKGCTTIKIQIFKMFPIIMAILVAWLLCHILTITEVFPMDPSAYGFKARTDARGEIMSTAP
WFRMPYPCQWGIPTVTIAGVLGMFSATLAGIIESIGDYYACARLAGALPPPVHAINRGIFIEGISCIIAGLLGT
GNGSTSSSPNIGVLGITKVGSRRVVQYGAIIMLILGTVGKFTALFASLPDPILGGMFCTLFGMITAVGLSNLQ
FVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPNAIKTGVPEIDQILTVLLTTEMFVGGSIAFILDNTIPGTDEER
GLVQWKAGAHANSETVAGLKSYDFPVGMDLVRKVECMKYIPVCPAFKGFRVRRKEGSDETQKEDTENFT
ACTKI*
>Sus_scrofa ENSSSCG00000026423|ENSSSCT00000025910|ENSSSCP00000021850
MRAWEDAEGPTQHKSLGSAGTSTREPPGALPTEPKSDMLYKLEDVPPWYLCILLGFQHYLTCFSGTIAVPF
LLAEALCVGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKSILALERWKCPPEEEIYGN
WSLPLNTSHVWHPRMREVQGAIMVSSMVEVVIGLTGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGS
HWGISACSILLIILFSQYLRNLTFLLPAYRWGKGVTLFRVQIFKMFPIVLAIMTVWLLCYVLTLTDVLPPDPT
AYGFQARTDARGDIMALAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPV
HAINRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTIGKFTALFASIPDPIL
GGMFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPGAINTGIPELDQILTVLLTTE
MFVGGCLAFILDNTVPGSPEERGLTQWKAGAHAHSEMSASLRSYDLPVGMSVVKRTAFLKYVPICPVFKG
FSSRSKDQLPVPADAPENTDARSVCTKV*
>Mustela_putorius_furo ENSMPUG00000009270|ENSMPUT00000009348|ENSMPUP00000009195

MRAQENLEGRTQHESLGSAGNSERDPMESLPREPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEDIYGN
WSLPLNTSHIWHPRIREVQGAIIVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIILFSQYLRNLTFLLPIYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYILTLTNVLPSDPTAYG
FQARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAIN
RGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTVGKFTALFASLPDPILGGM
FCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGTINTGIPEVDQILTVLLTTEMFVG
GCLAFILDNTVPGSPEERGLIQWKAGAHANSEMSTSLKSYDFPIGMNIVKRIAFLKYIPICPVFKGFSSRSKTQ
PPVPEDTPENIQTGSACTKV*
>Sarcophilus_harrisii ENSSHAG00000010598|ENSSHAT00000012472|ENSSHAP00000012370
TTLQEQRKGPSQRRTEQEDEGPSLSRAPPEDPKSDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAD
ALCVGKDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILSLDKWRCPPEEEIYGNWSLP
LNTSHIWHPRIREIQGAIMVSSTVEVMIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIST
FSIFLIILFSQYLRNVTFRLPGYKWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYILTLTDVLPADPNTYGFRAR
TDARGEIMSISPWVRFPYPCQWGLPSVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFT
EGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAIIMLILGTIGKFTALFASLPDPILGGMFCTLFG
MITAVGLSNLQFIDMNSSRNLFVLGFSMFFGLTLPNYLDSNPTAINTGIPEIDQILTVLLTTEMFVGGCLAFIL
DNTVPGSPEERGLVQWKAGAHSNSETSASLKSYDFPIGMNTIQKMAFLKYIPICPVFKGFSSKPQEQPSAPL
DNPENTDILSLSTKV*
>Callithrix_jacchus ENSCJAG00000006997|ENSCJAT00000013785|ENSCJAP00000013076
MRAQEDPKGRTQHESTRDPPMPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GHDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSLPLNTS
HIWHPRIREVGLHVQGAIMVSSVVEVVIGLLGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIS
ACSILLIILFSQYLRNLTFLLPVYSWGKGLTVLRIQIFKMFPIMLAIMTVWLLCYVLTLTDVLPTDPKAYGFQ
ARTDARGDVMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINR
GIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAVIMLVLGTIGKFTALFASLPDPILGGMF
CTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGILEVDQILTVLLTTEMFVG
GCLAFILDNTVPGSPEERGLIQWKAGAHANSDTSSSLKSYDFPIGMGIVKRTAFLKYIPICPVFKGFSSSKDQ
FPIPEDTPENTETASVCTKV*
>Monodelphis_domestica ENSMODG00000011886|ENSMODT00000015171|ENSMODP00000014899
MGTQEDHTARAQETEGPTLSRAPPEDPKLDMLYRIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GKDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILSLDKWRCPPEEEIYGNWSLPLNTS
HIWHPRIREIQGAIMVSSTVEVMIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISTFSILL
IVLFSQYLRNVTFRLPGYKWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYILTLTDLLPADPNTYGFRARTDAR
GEIMSISPWFRFPYPCQWGLPSVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGICC
IIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGTIGKFTALFASLPDPILGGMFCTLFGMITA
VGLSNLQFIDMNSSRNLFVLGFSMFFGLTLPNYLDSNPTAINTGIPEVDQILTVLLTTEMFVGGCLAFILDNT
VPGSPEERGLVQWKAGAHANSETSASLRSYDFPIGMKTIQKMAFLKYIPVCPVFKGFSSKPQEQPPAPEDTP
ENTDNLSLSTKV*
>Loxodonta_africana ENSLAFG00000009958|ENSLAFT00000009957|ENSLAFP00000008338
MRVQEDPEGQTQHESLDSAATRDPQMSLPTESKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLA
EALCVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLIPAKAILALERWKCPPEEEIYGNWSL
PLNTSHIWHPRIQEVQGAIIVSSMVEVVIGMMGLPGALLSYIGPLTVTPHCLIGLSVFQAAGDRAGSHWGISS
CSILLIILFSQYLRNFTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMIVWLLCYVLTLTDVLPTDSTAYGFQAR
TDARGDIMAIAPWVRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIF
TEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAVIMLVLGSIGKFTALFASLPDPILGGMFCTL
FGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIPEVDQILTVLLTTEMFVGGFL
AFILDNTVPGSPQERGLIQWKAGAHANSEMSTSLKSYDFPIGMGMVKRIAFLKYIPICPVFKGFSSRSKQELP
VPDDTPENMETGSVSTKV*
>Xenopus_tropicalis ENSXETG00000008402|ENSXETT00000018416|ENSXETP00000018416
NQEIHLTKDVENEMLYKIEDVPPWYLCIFLGLQHYLTCFSGTIAIPFLLANALCVGNDQQTVSQLIGTIFTCV
GITTFIQTTFGIRLPLFQASAFAFLVPARAILSLEKWKCPPEELIYGNGTVPFNTSHIWQPRIREIQGAIIVSSLL
EVLVGLIGLPGALLHYIGPLTVAPTISLIGLSVFEAAGQRAGSHWGISILSLTLIIMFAQYLRNVTFSVPGYKY
GEGLKIYKIQIFKMFPIIMAIMVVWLLCYILTLSGIFPTEDKTYGYSARTDARGEIMTSSPWFRFPYPCQWGL
PTVTVAGVLGMFSATLAGIVESMGDYYACARLSGAPPPPVHAINRGIFIEGICCIIAGFLGTGNGSTSSSPNIG
VLGITKIGSRRVVQYGAGIMFILGTVGKFTALFASIPDPILGGMFCTLFGMITAIGLSNLQFVDMNSSRNLFV

LGFSLFFGLALPNFLDSHPNFIQTGLKELDQILTVLLTTEMFVGGCIAFFLDNTMPGTVEERGLVQWKQGAN
ANSETSEDLKSYDFPFGMSFIKNVRCFQKMPICPVFKGFRRLAGRVDADNANGDSVKQIKVCTRV*
>Oryzias_latipes ENSORLG00000012701|ENSORLT00000015917|ENSORLP00000015916
KSQDDLEKQLSKKQNRTAETDGYIRKKEDGMEAKNAVDMIYTIEDVPPWYLCILLGLQHYLTCFSGTVAV
PFLLAEAMCVGQDQNTVSQLIGTIFTTVGITTLIQTTVGVRLPLFQASAFAFLIPAQAILGLDRWKCPSEEEIY
GNWSVPLNTSHIWQPRMREIQGAIIMSSLVEVIIGLCGLPGLLLKYIGPLTITPTVSLIGLSVFTTAGDRAGSH
WGLSALCILCIVLFAQYLRTTSIPVPFYSRKKGLTSTKVQIFKMFPIILAIMLVWLVCYILTLTNLLPSNPSRY
GHKARTDARGDIMASAPWFRVPYPCQWGLPVVTVAGVLGMFSATMAGIVESIGDYYACARLSGATPPPV
HAINRGIFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLLLGSVGKFTALFASLPDPI
LGGMFCTLFGMITAVGLSNLQLVDLNSSRNLFVLGFSMFFGLTLPTYLDTHPKSISTGVPELDQILTVLLSTE
MFVGGFLAFCLDNTIPGIPKKKTSVLSWPGSRCCFCNADLFSSYSQYDFPVGMTVIRRTRWLKRFPISPSFTG
FRAHDDDAPPEQEEREEEDESQLPTTKL*
>Gorilla_gorilla ENSGGOG00000010690|ENSGGOT00000010736|ENSGGOP00000010429
MRAQEDPEGRTQHESTRDPPTPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GHDQHMVSQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGNWSLPLNTS
HIWHPRIREVGLHVQGAIMVSSTVEVVIGLLGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIS
ACSILLIILFSQYLRNLTFLLPVYRWGKGLTLLRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPTDPKAYGFQ
ARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGI
FTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAAIMLVLGTIGKFTALFASLPDPILGGMFCT
LFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGILEVDQILTVLLTTEMFVGGC
LAFILDNTVPGSPEERGLIQWKAGAHANSDMSSSLKSYDFPIGMGIVKRIAFLKYIPICPVFKGFSSSSKDQIA
IPEDTPENTETASVCTKV*
>Mus_musculus ENSMUSG00000024354|ENSMUST00000025212|ENSMUSP00000025212
MKTPEDPGSPKQHEVVDSAGTSTRDRQAPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQHMVSQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKSILALERWKCPSEEEIYGN
WSMPLNTSHIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRVQIFKMFPIVLAIMTVWLLCYVLTLTDVLPADPTV
YGFQARTDARGDIMAISPWIRIPYPCQWGLPTVTVAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHA
INRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGAIGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPGAINTGIPEVDQILTVLLTTEMF
VGGCLAFILDNTVPGSPEERGLIQWKAGAHANSETSASLKSYDFPFGMGMVKRTTFFRYIPICPVFRGFSKK
TQNQPPVLEDTPDNIETGSVCTKV*
>Pelodiscus_sinensis ENSPSIG00000014954|ENSPSIT00000017065|ENSPSIP00000016987
ARLLLGLKSAFLHSQRQAASPGREQKEFDMIYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAESLCVG
KDQYTVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASALAFLVPAKSILALEKWKCPPEEEIYGNGTLPLNTRA
RWLPAVPQIQGAIIVSSLVEVVIGLVGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIAALSIFLI
VLFAQYLRNVTFRLPGYKCGKGCVVFRLQIFKMFPIILAIMVVWLLCYVLTLTDVFPRNAEAYGYKARTDA
RGEIMSIAPWFRLPYPCQWGWMFLPTLAGIIESIGDYYACARLAGAPAPPIHAINRGIFTEGISCIIAGLLGTG
NGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTIGKFTALFASLPDPILGGMFCTLFGMITAVGLSNLQF
VDMNSSRNLFILGFAMFFGLTVPNYLDSHPNAINTGVPEVDQILTVLLTTEMFVGGSIAFVLDNTIPGTQEER
GLVQWKAGAHADRTTAANLKSYDFPFGMSVVKRIQWLKYVPVCPVFKGFSSRTRNCDTTENVEEHTDNL
FVCTKV*
>Ailuropoda_melanoleuca ENSAMEG00000011305|ENSAMET00000012421|ENSAMEP00000011917
MRAQENPEGRTQHESLGSAGNSTRDPTVPLPTDPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQHMVSQLIGTIFTCVGITTLIQTTVGIRLPLFQASAFAFLVPAKAILALERWKCPPEEEIYGN
WSLPLNTSHIWHPRIREVQGAIMVSSTVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYVLTLMNVLPSDPTA
YGFQARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHA
INRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTVGKFTALFASLPDPILG
GMFCTLFGMITAGVSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAISTGIPEVDQILTVLLTTEMF
VGGCLAFILDNTVPGSPEERGLIQWKAGAHANSEMSTNLKSYDFPIGMSMVKRTAFLKYIPICPVFKGFSPR
SKTQLPVPEDTPENIQIGSACTKV*
>Ochotona_princeps ENSOPRG00000000264|ENSOPRT00000000266|ENSOPRP00000000239
MRAQQDSEGQPQHATPREPSMALPTQPKFDMLYKIEDTPPWYLCILLGFQHYLTCFSGTIAVPFLLAEAMC
VGQDQHMVSQLIGTIFVCVGITTLVQTTLGIRLPLFQASALAFLVPAQAILSLDKWKCPSEEEIYGNWSLPLN
TSHIWQPRIREVGLYIQGAIMVSSLVEVVIGLLGLPGALLSYIGPLTVTPTISLIGLSVFQTAGDRAGSHWGIS

ACAILLIILFSQYLRNVSFLLPIYRWGKGITVYRIQIFKMFPIVLAIMTVWLLCYILTLTNVLPADSSAYGFQA
RTDARGDIMAIAPWIRVPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGI
FTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQFGAGIMLILGVVGKFTALFASLPDPILGGMFCT
LFGMITAVGLSNLQFVDLNSSRNLFVLGFPIFFGLMLPNYLDSHPGAINTGVPEVDQILTVLLTTEMFVGGC
LAFILDNTVPGSPEERGLIQWKAGAHDSSETSASLKSYDFPFGMGTLKRVTFLRYIPICPFFKGFSSKSKDKPS
LPEDTPENTESRSVCTKV*
>Oryctolagus_cuniculus ENSOCUG00000006934|ENSOCUT00000006936|ENSOCUP00000005997
LTSPPLSTPLCCPQHAARGSAEAPPADPQVSLPTQPKFDMLYKIEDTPPWYLCILLGFQHYLTCFSGTIAVPF
LLAEALCVGQDQYMVSQLIGTIFVCVGITTLLQTTLGIRLPLFQASAFAFLVPAKAILSLERWKCPPEEEIYG
NWSLPLNTSHIWHPRMREIQGAIMVSSVVEVVIGLLGLPGALLSYIGPLTVTPTVSLIGLSVFQTAGERAGSH
WGISACAILLIVLFSQYLRNFALLLPVYRWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYILTITNVLPRDSTAY
GFQARTDARGEIMAIAPWIRLPYPCQWGLPTVTVAAVLGMFSATLAGIIESIGDYYACRLAGAPPPPVHAIN
RGIFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQFGAGIMLILGIIGKFTALFASLPDPILGGMF
CTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSIFFGLTLPNYLDSNPGAINTGITEVDQILTVLLTTEMFVGG
SIAFILDNTVPGSPEERGLIQWKAGAHANSETSASLKSYDFPIGMGTVKRIAFLRYIPICPVFKGFSPRSKDQH
SVPEDTPRNTESGSVCTKV*
>Microcebus_murinus ENSMICG00000012457|ENSMICT00000012460|ENSMICP00000011344
MRAQADTEGRTQHEYTRDPPPSLPMEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALDRWKCPPEEEIYGNWSLPLNTS
HIWHPRIREVGLHVQGAIMVSSTVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGIS
ACTILLIVLFSQYLRNLTFLLPVYRWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYILTLTDVLPTDPKAYGFQ
ARTDARGDIMATSPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRG
IFTEGVCCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMIILGTIGKFTALASLXXXXXXXXXX
XXXXMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGVPEVDQILTVLLTTEMFVG
GCLAFILDNTVPGSPEERGLIQWKAGAHANSETSTSLKSYDFPFGMGMVKRIGFLKYIPICPVFKGFSSRSKD
QFPVPEDTPENIETVSTYTKV*
>Myotis_lucifugus ENSMLUG00000003008|ENSMLUT00000003010|ENSMLUP00000002735
MRAQEDPEGQAQHEDLGLAGPSTKDPPVSLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLIPAKAILALERWKCPPEEEIYGNW
SLPLNTSHIWHPRMREVQGAIMVSSIVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHW
GISACSILLIVLFSQYLRNLTFLLPVYRWGKGLSFFRIQIFKMFPIVLAIMTVWLLCYILTLTDVLPTDPTAYG
FQARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAIN
RGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTIGKFTALFASLPDPILGGM
FCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIPEVDQILTVLLTTEMFVG
GCLAFLLDNTVPGSPEERGLVQWKAGAHANSEMSTSLKSYDFPFGMNMIKRIAFLKYIPICPVFKGFSSRSK
DQRPVPEDISENRETGPVCTKV*
>Ictidomys_tridecemlineatus ENSSTOG00000014761|ENSSTOT00000014762|ENSSTOP00000013225
QHESPSSTGNSTRDPQMPLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCVGQDQ
NMISQLIGTIFVCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALEKWKCPPEEEIYGNWSQPLNTSHIWH
PRIREVQGAIMVASVVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSILLIILF
SQYLRNVTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPKDPTAYGYQARTDARGD
IMSISPWIRIPYPCQWGLPTVTVAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEGICCIIA
GLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGAIGKFTALFASLPDPILGGMFCTLFGMITAV
GLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLDSNPGAIDTGVAEIDQILTVLLTTEMFVGGCLAFILDNT
VPGSPEERGLIQWKAGAHANSETSASLKSYDFPIGMGTVKRIAFLKYIPVCPVFKGFSSRSKTQSRVPEDIPE
NIETGSGCTKV*
>Felis_catus ENSFCAG00000001723|ENSFCAT00000001723|ENSFCAP00000001597
MRAQEDPEGQTQHESLGSAGTSTRDPTMSLPTEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFL
LAEALCVGRDQYMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILGLERWKCPPEEEIYGN
WSLPLNTSHIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSH
WGISACSILLIVLFSQYLRNLTFLLPVYRWGKGLTLFRIQIFKMFPIVLAIMTVWLLCYILTLTNVLPSDPTAY
GFQARTDARGDIMAIAPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAI
NRGIFTEGICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLVLGTVGKFTALFASLPDPILGG
MFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGVINTGIPEVDQILTVLLTTEMF
VGGCLAFILDNTVPGSPEERGLIQWKAGAHANSEMSTSLKSYDFPIGMSMVKRTAFLKYIPICPVFRGFSRS
KAQHPVPEDTPDNIHTGSACTKV*

>Otolemur_garnettii ENSOGAG00000010614|ENSOGAT00000010618|ENSOGAP00000009500
MRAQEDSEGQIQHECTRDPPVSLPAEPKFDMLYKIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAEALCV
GHDQHMVSQLIGTIFTCVGITTLIQTTLGIRLPLFQASAFAFLVPAKAILALDRWKCPPEEEIYGNWSLPLNTS
HIWHPRIREVQGAIMVSSMVEVVIGLMGLPGALLNYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISACSI
LLIILFSQYLRDFTFLLPVYRWGKGFTLFRIQIFKMFPIVLAIMTVWLLCYVLTLTDVLPTDPTAYGFQARTD
ARGDIMATSPWIRIPYPCQWGLPTVTAAAVLGMFSATLAGIIESIGDYYACARLAGAPPPPVHAINRGIFTEG
ICCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVVQYGAGIMLILGTIGKFTALFASLPDPILGGMFCTLFGMI
TAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAINTGIPEVDQILTVLLTTEMFVGGCLAFILD
NTVPGSAKERGLIQWKAGAHANSEMSTTLRSYDFPIGMGTVKRISFLKYIPICPVFKGFSSRSKDQFPVPEDT
PENTETGSVNTKV*
>Tupaia_belangeri ENSTBEG00000004467|ENSTBET00000004472|ENSTBEP00000003859
MRAQEDPQGQTQHESTRDSPTSRPTEPTFDMLYKIEDVPPWYLCILLGFQXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLPLFQASAFAFLVPAKAILALERWKCPPEEEIYG
NWSLPLNTSHIWHPRMREVGLHVQGAIMVSSMVEVVIGLMGLPGALLSYIGPLTVTPTVSLIGLSVFQAAG
DRAGSHWGISACSILLIVLFSQYLRNLTFLLPVYRWGKGFTFFRIQIFKMFPXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXQWGLPTVTVAAVLGMFSATLAGIIESIGDYYAC
ARLAGAPPPPVHAINXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXVGRRRVVQYGAGIMLV
LGTIGKFTALFASLPDPILECMFCTLFGMITAVGLSNLQFVDMNSSRNLFVLGFSMFFGLTLPNYLESNPGAI
NTGIPEVDQILTVLLTTEMFVGGCLAFILDNTVPGSPEERGLIQWKAGAHANSETSTSLQSYDFPIGMGMVK
RIACLKYIPICPVFKGFPTKSKGQLPEPEDIPENMETGSVCTKV*
>Ficedula_albicollis ENSFALG00000003008|ENSFALT00000003173|ENSFALP00000003160
LGAQTPLGGRSRHRLGQDPGAGTRPPRPEVDMLYRIEDVPPWYLCILLGFQHYLTCFSGTIAVPFLLAESLC
VGKDQLTVSYLIGTIFTCVGITTLIQTTVGIRLPLFQASALAFLVPAKSILALEKWRCPPEEQIYGNWSLPLNT
SHIWQPRMREIQGAIMVSSLVEVVIGLLGLPGALLSYIGPLTVTPTVSLIGLSVFQAAGDRAGSHWGISVLTI
FLIVLFAQYLRQVAIYVPGYRRGHGFVLLRVQIFKMFPIILAIMLVWLICYVLTRTGVFPSRPEEYGYKARTD
ARGEILSVAPWFRVPYPCQWGLPTVTSAAVLGMFSATLAGIIESIGDYYSCARLAGAPPPPVHAINRGIFIEGI
SCIIAGLLGTGNGSTSSSPNIGVLGITKVGSRRVIQYGAGIMLLLGTIGKFTALFASLPDPVLGGMFCTLFGMI
TAVGLSNLQFVDMNSSRNLFVLGFAMFFGLTLPNYLDSHPGAISTGVPELDQILTVLLTTEMFVGGTIAFVL
DNTIPGTQEERGLVQWKAGAHSDSSSSASLRSYDFPVGMGAVRRSRWLRSVPICPVFTGFRGRAGGRGTA
AAAGPAGADGGSVCTKV*

Supplemental Table 3. *SLC23A1* dbSNP and ENSEMBL variation table. The original data was download from Ensemble database on March 10th, 2022 (https://uswest.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000170482;r=5:139367196-139384553). Only missense mutations, stop gained and start lost mutations on transcript ENST00000348729.8 for *SLC23A1* were selected.

Variant ID	Location	Chr: bp	vf_allele	Alleles	Evidence	Clin. Sig.	Conseq. T	AA	AA c	sift_sort	sift_class	cadd_class	Transcript
rs1151029	5:1393720	5:1393720	G	C/G	Frequency~1000Gen		missense v	V/L	598	21	deleterious	likely beni	ENST00000348729.8
rs7768248	5:1393720	5:1393720	C	T/C	Frequency~ExAC~gn		missense v	K/E	597	51	deleterious	likely beni	ENST00000348729.8
rs9649872	5:1393720	5:1393720	G	A/G	Frequency~TOPMed		missense v	V/A	594	171	tolerated -	likely beni	ENST00000348729.8
rs1474852	5:1393720	5:1393720	T	C/T	Frequency~gnomAD		missense v	V/M	594	121	tolerated -	likely beni	ENST00000348729.8
rs7702308	5:1393720	5:1393720	C	G/C	Frequency~ExAC~gn		missense v	S/C	593	151	tolerated	likely beni	ENST00000348729.8
rs1376685	5:1393720	5:1393720	G	T/G	Frequency~gnomAD		missense v	E/D	590	621	tolerated	likely beni	ENST00000348729.8
rs1435270	5:1393720	5:1393720	T	C/T	Frequency~gnomAD		missense v	E/K	590	241	tolerated	likely beni	ENST00000348729.8
rs7737145	5:1393720	5:1393720	T	G/T	Frequency~ExAC~gn		missense v	T/K	589	631	tolerated	likely beni	ENST00000348729.8
rs5699136	5:1393720	5:1393720	A	G/A/C	Frequency~1000Gen		missense v	T/I	585	201	tolerated	likely beni	ENST00000348729.8
rs5699136	5:1393720	5:1393720	C	G/A/C	Frequency~1000Gen		missense v	T/S	585	801	tolerated	likely beni	ENST00000348729.8
rs1462882	5:1393720	5:1393720	C	T/C	Frequency~TOPMed		missense v	T/A	585	761	tolerated	likely beni	ENST00000348729.8
rs5307520	5:1393720	5:1393720	G	C/G/T	Frequency~1000Gen		missense v	D/H	584	161	tolerated	likely beni	ENST00000348729.8
rs5307520	5:1393720	5:1393720	T	C/G/T	Frequency~1000Gen		missense v	D/N	584	201	tolerated	likely beni	ENST00000348729.8
rs1327281	5:1393720	5:1393720	G	A/G	Frequency~gnomAD		missense v	I/T	581	721	tolerated	likely beni	ENST00000348729.8
rs7602037	5:1393720	5:1393720	G	C/G	Frequency~ExAC~gn		missense v	A/P	580	371	tolerated	likely beni	ENST00000348729.8
rs7638283	5:1393720	5:1393720	G	A/G	Frequency~ExAC~TO		missense v	I/T	579	391	tolerated	likely beni	ENST00000348729.8
rs7571204	5:1393720	5:1393720	A	T/A	Frequency~ExAC~gn		missense v	D/V	577	231	tolerated	likely beni	ENST00000348729.8
rs7651881	5:1393720	5:1393720	T	C/T	Frequency~ExAC~gn		missense v	D/N	577	511	tolerated	likely beni	ENST00000348729.8
rs7584540	5:1393720	5:1393720	T	A/T	Frequency~ExAC~gn		missense v	S/T	575	581	tolerated	likely beni	ENST00000348729.8
rs7799709	5:1393720	5:1393720	A	T/A	Frequency~ExAC~gn		missense v	S/C	574	61	tolerated	likely beni	ENST00000348729.8
rs7471482	5:1393720	5:1393720	C	G/C	Frequency~ExAC~gn		stop gaine	S/*	573		-	-	ENST00000348729.8
rs1396435	5:1393720	5:1393720	G	A/G	Frequency~gnomAD		missense v	S/P	572	121	tolerated	likely beni	ENST00000348729.8
rs1454323	5:1393720	5:1393720	G	A/G	Frequency~gnomAD		missense v	F/L	571	561	tolerated	likely beni	ENST00000348729.8
rs1430292	5:1393720	5:1393720	T	C/T	Frequency~TOPMed		missense v	G/E	570	31	deleterious	likely beni	ENST00000348729.8
rs9766672	5:1393721	5:1393721	C	G/C	Frequency~TOPMed		missense v	P/A	566	1	deleterious	likely beni	ENST00000348729.8
rs7551098	5:1393721	5:1393721	G	A/G	Frequency~ExAC~gn		missense v	I/T	562	1	deleterious	likely beni	ENST00000348729.8
rs7484465	5:1393721	5:1393721	A	G/A	Frequency~ExAC~gn		missense v	T/I	557	91	tolerated	likely beni	ENST00000348729.8
rs9512561	5:1393721	5:1393721	A	C/A/G	Frequency~TOPMed		missense v	R/I	555	31	deleterious	likely beni	ENST00000348729.8
rs9512561	5:1393721	5:1393721	G	C/A/G	Frequency~TOPMed		missense v	R/T	555	201	tolerated	likely beni	ENST00000348729.8
rs7701409	5:1393721	5:1393721	C	T/C	Frequency~ExAC~gn		missense v	R/G	555	31	deleterious	likely beni	ENST00000348729.8
rs1504616	5:1393721	5:1393721	C	T/C	Frequency~ESP~ExA		missense v	K/R	554	1001	tolerated	likely beni	ENST00000348729.8
rs7497735	5:1393721	5:1393721	T	C/T	Frequency~ExAC~TO		missense v	V/I	553	771	tolerated	likely beni	ENST00000348729.8
rs1318768	5:1393721	5:1393721	G	A/G	Frequency~gnomAD		missense v	I/T	552	391	tolerated	likely beni	ENST00000348729.8
rs1373345	5:1393721	5:1393721	T	C/T	Frequency~TOPMed		missense v	G/D	551	321	tolerated	likely beni	ENST00000348729.8
rs7714992	5:1393721	5:1393721	G	A/G	Frequency~ExAC~gn		missense v	I/T	548	21	deleterious	likely beni	ENST00000348729.8
rs7748139	5:1393721	5:1393721	C	T/C	Frequency~ExAC~gn		missense v	I/V	548	121	tolerated	likely beni	ENST00000348729.8
rs2015657	5:1393721	5:1393721	C	G/C	Frequency~1000Gen		missense v	P/A	547	11	deleterious	likely beni	ENST00000348729.8
rs2002648	5:1393721	5:1393721	T	C/T	Frequency~1000Gen		missense v	D/N	545	181	tolerated	likely beni	ENST00000348729.8
rs9506272	5:1393721	5:1393721	C	T/C	Frequency~TOPMed		missense v	Y/C	544	1	deleterious	likely beni	ENST00000348729.8

rs1300174	5:1393782	5:1393782	T	G/T	Frequency~TOPMed	missense v	A/D	418	11	deleterious	likely benig	ENST00000348729.8
rs1268338	5:1393782	5:1393782	A	G/A	Frequency~gnomAD	missense v	T/M	417	1	deleterious	likely benig	ENST00000348729.8
rs1464131	5:1393782	5:1393782	G	C/G	Frequency~TOPMed	missense v	K/N	415	1	deleterious	likely benig	ENST00000348729.8
rs1325807	5:1393782	5:1393782	C	G/C	Frequency~gnomAD	missense v	I/M	413	41	deleterious	likely benig	ENST00000348729.8
rs1186150	5:1393783	5:1393783	T	C/T	Frequency~TOPMed	missense v	G/S	411	31	deleterious	likely benig	ENST00000348729.8
rs1581367	5:1393783	5:1393783	G	C/G	Frequency	missense v	M/I	407	41	deleterious	likely benig	ENST00000348729.8
rs1053398	5:1393783	5:1393783	G	A/G	Frequency~TOPMed	missense v	I/T	406	11	deleterious	likely benig	ENST00000348729.8
rs7747100	5:1393783	5:1393783	A	G/A/C	Frequency~ExAC~TO	missense v	A/V	404	1	deleterious	likely benig	ENST00000348729.8
rs7747100	5:1393783	5:1393783	C	G/A/C	Frequency~ExAC~TO	missense v	A/G	404	721	tolerated	likely benig	ENST00000348729.8
rs7461584	5:1393783	5:1393783	T	C/T	Frequency~ExAC~gn	missense v	G/S	403	231	tolerated	likely benig	ENST00000348729.8
rs1581367	5:1393783	5:1393783	C	T/C		missense v	Y/C	402	171	tolerated	likely benig	ENST00000348729.8
rs1389684	5:1393783	5:1393783	G	A/G	Frequency~gnomAD	missense v	Y/H	402	341	tolerated	likely benig	ENST00000348729.8
rs1456854	5:1393783	5:1393783	T	C/T	Frequency~gnomAD	missense v	V/M	400	51	deleterious	likely benig	ENST00000348729.8
rs7723018	5:1393783	5:1393783	C	A/C	Frequency~ExAC~gn	missense v	V/G	399	1	deleterious	likely delet	ENST00000348729.8
rs1373281	5:1393783	5:1393783	T	C/T	Frequency~TOPMed	missense v	V/M	399	1	deleterious	likely benig	ENST00000348729.8
rs7646790	5:1393783	5:1393783	A	G/A/T	Frequency~ExAC~TO	missense v	R/C	398	31	deleterious	likely benig	ENST00000348729.8
rs7646790	5:1393783	5:1393783	T	G/A/T	Frequency~ExAC~TO	missense v	R/S	398	21	deleterious	likely benig	ENST00000348729.8
rs1329750	5:1393783	5:1393783	C	A/C	Frequency~gnomAD	missense v	V/G	394	1	deleterious	likely delet	ENST00000348729.8
rs1164191	5:1393785	5:1393785	G	T/G	Frequency~gnomAD	missense v	K/Q	393	61	tolerated	likely benig	ENST00000348729.8
rs3772487	5:1393785	5:1393785	G	A/G	Frequency~ESP~ExA	missense v	I/T	391	1	deleterious	likely benig	ENST00000348729.8
rs1463693	5:1393785	5:1393785	T	G/T	Frequency~TOPMed	missense v	L/M	389	11	deleterious	likely benig	ENST00000348729.8
rs7771212	5:1393785	5:1393785	T	C/T	Frequency~ExAC~gn	missense v	V/I	388	11	deleterious	likely benig	ENST00000348729.8
rs1038790	5:1393785	5:1393785	G	C/G	Frequency~gnomAD	missense v	G/A	387	181	tolerated	likely benig	ENST00000348729.8
rs7623672	5:1393786	5:1393786	G	A/G	Frequency~ExAC~gn	missense v	I/T	386	21	deleterious	likely benig	ENST00000348729.8
rs1315132	5:1393786	5:1393786	T	C/T	Frequency~gnomAD	missense v	S/N	383	1	deleterious	likely benig	ENST00000348729.8
rs1379459	5:1393786	5:1393786	C	T/C	Frequency~gnomAD	missense v	S/G	383	111	tolerated	likely benig	ENST00000348729.8
rs7658575	5:1393786	5:1393786	T	A/T	Frequency~ExAC~TO	missense v	S/T	381	131	tolerated	likely benig	ENST00000348729.8
rs1379522	5:1393786	5:1393786	A	G/A	Frequency~ESP~ExA	missense v	T/I	380	1	deleterious	likely benig	ENST00000348729.8
rs7590708	5:1393786	5:1393786	G	T/G	Frequency~ExAC~gn	missense v	T/P	380	21	deleterious	likely benig	ENST00000348729.8
rs1339925	5:1393786	5:1393786	G	C/G	Frequency~gnomAD	missense v	G/R	378	1	deleterious	likely benig	ENST00000348729.8
rs7664640	5:1393786	5:1393786	A	G/A	Frequency	missense v	T/M	375	1	deleterious	likely benig	ENST00000348729.8
rs1483704	5:1393786	5:1393786	A	C/A	Frequency~gnomAD	missense v	G/C	374	1	deleterious	likely benig	ENST00000348729.8
rs1390139	5:1393786	5:1393786	A	G/A	Frequency~gnomAD	missense v	A/V	370	11	deleterious	likely benig	ENST00000348729.8
rs1428314	5:1393786	5:1393786	T	A/T	Frequency~gnomAD	missense v	I/N	368	1	deleterious	likely benig	ENST00000348729.8
rs7584872	5:1393786	5:1393786	G	C/G/T	Frequency~ExAC~gn	missense v	G/A	364	1	deleterious	likely benig	ENST00000348729.8
rs7584872	5:1393786	5:1393786	T	C/G/T	Frequency~ExAC~gn	missense v	G/D	364	1	deleterious	likely benig	ENST00000348729.8
rs1150231	5:1393786	5:1393786	T	C/T	Frequency~1000Gen	missense v	E/K	363	1	deleterious	likely benig	ENST00000348729.8
rs7803252	5:1393786	5:1393786	A	G/A	Frequency~ExAC~TO	missense v	T/I	362	1001	tolerated	likely benig	ENST00000348729.8
rs1162047	5:1393786	5:1393786	G	A/G	Frequency~TOPMed	missense v	F/L	361	211	tolerated	likely benig	ENST00000348729.8
rs7485345	5:1393792	5:1393792	T	G/T	Frequency~ExAC~gn	missense v	A/D	355	1	deleterious	likely benig	ENST00000348729.8

rs7766121	5:1393779	5:1393779	A	G/A/T	Frequency~ExAC~TO	missense v	P/S	479	231	tolerated	likely beni	ENST00000348729.8
rs7766121	5:1393779	5:1393779	T	G/A/T	Frequency~ExAC~TO	missense v	P/T	479	41	deleterious	likely beni	ENST00000348729.8
rs7620504	5:1393779	5:1393779	T	A/T	Frequency~ExAC~gn	missense v	S/T	477	381	tolerated	likely beni	ENST00000348729.8
rs5321603	5:1393780	5:1393780	A	T/A/C	Frequency~1000Gen	missense v	N/I	473	31	deleterious	likely beni	ENST00000348729.8
rs5321603	5:1393780	5:1393780	C	T/A/C	Frequency~1000Gen	missense v	N/S	473	451	tolerated	likely beni	ENST00000348729.8
rs1026260	5:1393780	5:1393780	T	G/A/T	Frequency~TOPMed~	missense v	L/M	471	21	deleterious	likely beni	ENST00000348729.8
rs1166592	5:1393780	5:1393780	A	G/A	Frequency~1000Gen	missense v	T/M	470	121	tolerated	likely beni	ENST00000348729.8
rs5659618	5:1393780	5:1393780	C	A/C	Frequency~1000Gen	missense v	L/R	469	1	deleterious	likely beni	ENST00000348729.8
rs8684273	5:1393780	5:1393780	T	C/T	Frequency	missense v	G/R	468	1	deleterious	likely beni	ENST00000348729.8
rs1410695	5:1393780	5:1393780	G	A/G	Frequency~gnomAD	missense v	M/T	465	141	tolerated	likely beni	ENST00000348729.8
rs1401632	5:1393780	5:1393780	A	G/A	Frequency~gnomAD	missense v	S/F	464	1	deleterious	likely beni	ENST00000348729.8
rs7669205	5:1393780	5:1393780	G	C/G	Frequency~ExAC~TO	missense v	G/A	462	11	deleterious	likely beni	ENST00000348729.8
rs7692621	5:1393780	5:1393780	T	C/T	Frequency	missense v	G/R	462	1	deleterious	likely beni	ENST00000348729.8
rs7773086	5:1393780	5:1393780	T	C/T	Frequency~ExAC	missense v	V/M	460	1	deleterious	likely beni	ENST00000348729.8
rs9828401	5:1393780	5:1393780	A	G/A	Frequency~TOPMed	missense v	L/F	458	1	deleterious	likely beni	ENST00000348729.8
rs7533195	5:1393780	5:1393780	C	T/C	Frequency~ExAC~gn	missense v	N/S	457	1	deleterious	likely beni	ENST00000348729.8
rs1326013	5:1393780	5:1393780	A	C/A	Frequency~gnomAD	missense v	R/L	456	1	deleterious	likely beni	ENST00000348729.8
rs7568857	5:1393780	5:1393780	G	C/G	Frequency~ExAC~gn	missense v	M/I	452	11	deleterious	likely beni	ENST00000348729.8
rs1318853	5:1393780	5:1393780	T	A/T	Frequency~TOPMed	missense v	M/K	452	21	deleterious	likely beni	ENST00000348729.8
rs7784344	5:1393780	5:1393780	C	T/C	Frequency~ExAC~TO	missense v	M/V	452	1	deleterious	likely beni	ENST00000348729.8
rs7491951	5:1393780	5:1393780	C	A/C	Frequency~TOPMed~	missense v	F/C	449	1	deleterious	likely beni	ENST00000348729.8
rs1227590	5:1393780	5:1393780	G	A/G	Frequency~gnomAD	missense v	F/L	449	71	tolerated	likely beni	ENST00000348729.8
rs7455986	5:1393780	5:1393780	C	G/C	Frequency~ExAC~gn	missense v	L/V	447	1	deleterious	likely beni	ENST00000348729.8
rs7718394	5:1393780	5:1393780	C	T/C	Frequency~ExAC~TO	missense v	N/S	446	61	tolerated	likely beni	ENST00000348729.8
rs7798694	5:1393780	5:1393780	C	G/C	Frequency~ExAC~TO	missense v	L/V	444	221	tolerated	likely beni	ENST00000348729.8
rs7469154	5:1393781	5:1393781	G	A/G	Frequency~ExAC~gn	missense v	I/T	439	1	deleterious	likely beni	ENST00000348729.8
rs5484753	5:1393781	5:1393781	G	T/G	Frequency~1000Gen	missense v	I/L	439	1	deleterious	likely beni	ENST00000348729.8
rs7767196	5:1393781	5:1393781	C	T/C	Frequency~ExAC~gn	missense v	M/V	438	341	tolerated	likely beni	ENST00000348729.8
rs1302029	5:1393782	5:1393782	A	G/A	Frequency~TOPMed~	missense v	T/I	434	171	tolerated	likely beni	ENST00000348729.8
rs1399712	5:1393782	5:1393782	T	C/T	Frequency~gnomAD	missense v	C/Y	433	31	deleterious	likely beni	ENST00000348729.8
rs7512916	5:1393782	5:1393782	T	A/T	Frequency~ExAC~gn	missense v	F/Y	432	1	deleterious	likely beni	ENST00000348729.8
rs1320521	5:1393782	5:1393782	A	C/A	Frequency~gnomAD	missense v	M/I	431	191	tolerated	likely beni	ENST00000348729.8
rs1324004	5:1393782	5:1393782	G	A/G	Frequency~TOPMed	missense v	M/T	431	51	deleterious	likely beni	ENST00000348729.8
rs1434501	5:1393782	5:1393782	T	C/T	Frequency~gnomAD	missense v	G/S	430	1	deleterious	likely beni	ENST00000348729.8
rs1295432	5:1393782	5:1393782	A	G/A	Frequency~gnomAD	missense v	P/S	426	21	deleterious	likely beni	ENST00000348729.8
rs7811038	5:1393782	5:1393782	T	G/T	Frequency~ExAC~gn	missense v	P/T	424	1	deleterious	likely beni	ENST00000348729.8
rs8928429	5:1393782	5:1393782	C	G/C		missense v	L/V	423	81	tolerated	likely beni	ENST00000348729.8
rs7480176	5:1393782	5:1393782	A	G/A	Frequency~ExAC~TO	missense v	S/L	422	41	deleterious	likely delet	ENST00000348729.8
rs6596474	5:1393782	5:1393782	A	C/A/G	Frequency~1000Gen	missense v	A/S	421	51	deleterious	likely beni	ENST00000348729.8
rs6596474	5:1393782	5:1393782	G	C/A/G	Frequency~1000Gen	missense v	A/P	421	11	deleterious	likely beni	ENST00000348729.8

rs7502790	5:1393721	5:1393721	G	C/G	Frequency~ExAC~gn	missense v	M/I	537	231	tolerated	likely beni	ENST00000348729.8
rs7627634	5:1393721	5:1393721	G	A/G	Frequency~ExAC~TO	missense v	M/T	537	951	tolerated	likely beni	ENST00000348729.8
rs1192502	5:1393721	5:1393721	T	G/T	Frequency~TOPMed	missense v	D/E	536	941	tolerated	likely beni	ENST00000348729.8
rs7663669	5:1393721	5:1393721	C	T/C	Frequency~ExAC~TO	missense v	D/G	536	151	tolerated	likely beni	ENST00000348729.8
rs9593155	5:1393722	5:1393722	G	T/G		missense v	S/R	535	111	tolerated	likely beni	ENST00000348729.8
rs7515380	5:1393722	5:1393722	T	G/T	Frequency~ExAC~TO	missense v	A/D	533	541	tolerated	likely beni	ENST00000348729.8
rs1469375	5:1393722	5:1393722	G	C/G	Frequency~TOPMed	missense v	A/P	531	271	tolerated	likely beni	ENST00000348729.8
rs7550926	5:1393722	5:1393722	A	G/A/C	Frequency~ExAC~gn	missense v	A/V	529	261	tolerated	likely beni	ENST00000348729.8
rs7550926	5:1393722	5:1393722	C	G/A/C	Frequency~ExAC~gn	missense v	A/G	529	361	tolerated	likely beni	ENST00000348729.8
rs9159050	5:1393722	5:1393722	C	T/C		missense v	K/E	528	201	tolerated	likely beni	ENST00000348729.8
rs3737788	5:1393722	5:1393722	T	C/T	Frequency~ESP~ExA	stop gaine	W/*	527		-	-	ENST00000348729.8
rs7255225	5:1393722	5:1393722	T	G/T	Frequency~1000Gen	missense v	Q/K	526	651	tolerated	likely beni	ENST00000348729.8
rs7563785	5:1393722	5:1393722	G	A/G	Frequency~ExAC~TO	missense v	I/T	525	291	tolerated	likely beni	ENST00000348729.8
rs7613570	5:1393722	5:1393722	T	C/T	Frequency~ExAC~gn	missense v	R/H	522	1	deleterious	likely beni	ENST00000348729.8
rs1393720	5:1393722	5:1393722	A	G/A	Frequency~gnomAD	missense v	R/C	522	1	deleterious	likely beni	ENST00000348729.8
rs7712838	5:1393722	5:1393722	G	C/G	Frequency~ExAC~gn	missense v	E/D	520	11	deleterious	likely beni	ENST00000348729.8
rs1581350	5:1393722	5:1393722	C	G/C	Frequency	missense v	P/R	519	571	tolerated	likely beni	ENST00000348729.8
rs1180886	5:1393722	5:1393722	C	T/C	Frequency~TOPMed	missense v	S/G	518	1	deleterious	likely beni	ENST00000348729.8
rs1561975	5:1393774	5:1393774	A	G/A		missense v	P/S	516	1	deleterious	likely beni	ENST00000348729.8
rs5697247	5:1393774	5:1393774	G	C/G	Frequency~1000Gen	missense v	V/L	515	111	tolerated	likely beni	ENST00000348729.8
rs7695167	5:1393774	5:1393774	A	T/A	Frequency~ExAC	missense v	T/S	514	11	deleterious	likely beni	ENST00000348729.8
rs1246886	5:1393774	5:1393774	C	T/C	Frequency~TOPMed	missense v	N/D	513	1	deleterious	likely beni	ENST00000348729.8
rs7489639	5:1393774	5:1393774	C	A/C	Frequency~ExAC~gn	missense v	L/R	511	1	deleterious	likely delet	ENST00000348729.8
rs7707487	5:1393774	5:1393774	C	G/C/T	Frequency~ExAC~gn	missense v	L/V	511	1	deleterious	likely beni	ENST00000348729.8
rs7707487	5:1393774	5:1393774	T	G/C/T	Frequency~ExAC~gn	missense v	L/I	511	1	deleterious	likely beni	ENST00000348729.8
rs3704297	5:1393774	5:1393774	G	C/G/T	Frequency~ESP~ExA	missense v	A/P	508	21	deleterious	likely beni	ENST00000348729.8
rs3704297	5:1393774	5:1393774	T	C/G/T	Frequency~ESP~ExA	missense v	A/T	508	1	deleterious	likely beni	ENST00000348729.8
rs1315334	5:1393774	5:1393774	T	C/T	Frequency~gnomAD	missense v	G/R	505	1	deleterious	likely beni	ENST00000348729.8
rs1447719	5:1393774	5:1393774	T	C/T	Frequency~gnomAD	missense v	G/D	504	1	deleterious	likely beni	ENST00000348729.8
rs7674638	5:1393774	5:1393774	G	A/G	Frequency~ExAC~TO	missense v	M/T	501	1	deleterious	likely beni	ENST00000348729.8
rs1346990	5:1393774	5:1393774	T	C/T	Frequency	missense v	E/K	500	81	tolerated	likely beni	ENST00000348729.8
rs7607458	5:1393774	5:1393774	A	G/A	Frequency~ExAC~gn	missense v	T/M	499	1	deleterious	likely beni	ENST00000348729.8
rs1294234	5:1393774	5:1393774	G	A/G	Frequency~gnomAD	missense v	V/A	495	11	deleterious	likely beni	ENST00000348729.8
rs7642248	5:1393774	5:1393774	G	A/G	Frequency~ExAC~TO	missense v	I/T	494	1001	tolerated	likely beni	ENST00000348729.8
rs1236482	5:1393774	5:1393774	C	G/C	Frequency~gnomAD	missense v	Q/E	491	1	deleterious	likely beni	ENST00000348729.8
rs1280934	5:1393774	5:1393774	T	C/T	Frequency~gnomAD	missense v	D/N	490	61	tolerated	likely beni	ENST00000348729.8
rs1217350	5:1393779	5:1393779	A	G/A	Frequency~gnomAD	missense v	T/I	484	11	deleterious	likely delet	ENST00000348729.8
rs7606620	5:1393779	5:1393779	C	T/C	Frequency~ExAC~gn	missense v	T/A	484	1	deleterious	likely beni	ENST00000348729.8
rs9659240	5:1393779	5:1393779	T	C/T		missense v	A/T	481	421	tolerated	likely beni	ENST00000348729.8
rs1018094	5:1393779	5:1393779	T	G/T	Frequency~TOPMed	missense v	P/H	479	1	deleterious	likely beni	ENST00000348729.8

TRANSITION STATEMENT 2

Before the discovery of the sodium dependent ascorbic acid transporters *SLC23A1 and SLC23A2,* due to their ability to transport dehydroascorbic acid, glucose transporters were thought to be the primary Vitamin C transporter in the human body. The GLUT type transporters are now considered to occupy a minor nice role in Vitamin C transport but might have specific roles in certain tissues and physiological situations.

SLC23A1 and *SLC2A14* share several genetic disease associations. After reviewing the literature, both *SLC23A1* and *SLC2A14* are associated with Alzheimer's disease, Parkinson's disease, and inflammatory bowel disease. Unlike *SLC23A1*, the function of *SLC2A14* was not fully described in the literature. Thus, the disease association of *SLC2A14* and the underlying mechanisms can currently not be properly explained.

SLC2A14 encodes a protein named GLUT14, which is known as a dehydroascorbic acid and glucose transporter. A total of five isoforms were identified for GLUT14. However, the functional study and substrate identification was only done for GLUT14-isoform A. The function of the other four isoforms remains unclear. Thus, the functions and the substrates for the other isoforms were investigated.

In the following manuscript, functional and non-functional protein isoforms are described, as well as the substrate spectra.

Chapter 4: Manuscript 2

Spectrum of Substrates of the GLUT14 Protein Isoforms

Designed as a short report to one of the NRC journals

Ruotong Liu, Peter Eck

Contributions of Authors:

Ruotong Liu designed the T7 tagged forward primers for all isoforms of GLUT14. She amplified the PCR products for all isoforms and preformed the cRNA injection. She completed the uptake and the competitive inhibition study in this manuscript and did the statistical analysis. She wrote the original draft for abstract, method and result section. She helped with manuscript editing.

Peter Eck designed the concept and managed and supervised the whole project. He wrote the original draft for introduction and discussion section. He managed the editing for this manuscript.

Abstract

Background: The *SLC2A14* gene encodes five GLUT14 isoforms. There is very limited information published about the isoforms and their substrate spectrum. Recent studies indicated that GLUT14 isoforms A and B transport glucose and dehydroascorbic acid, and GLUT14 isoform A is multi-specific, accepting substrate in the know range for GLUT-type sugar transporters. Due to very high protein similarities, we hypothesize that GLUT14 isoform B, C, D and E are also multi-specific transporters.

Objective: This study aimed to characterize GLUT14 isoform B, C, D and E substrates.

Design: The *Xenopus laevis* oocyte expression system was used to evaluate the uptake of radiolabelled 2-D-deoxyglucose in all GLUT14 isoforms. Alternative substrates were determined based on the inhibition of radiolabelled 2-D-deoxyglucose.

Result: All isoforms, except GLUT14-C mediate glucose uptake into *Xenopus laevis* oocytes. The substrate spectrum for all isoforms is similar, accepting the monomeric substrates glucose, galactose, fructose, mannose, xylose, arabinose, glucosamine, and uric acid. The dimers or trimers maltose, lactose and raffinose were also inhibitors; however, that is likely due to a blocking of the carrier rather than substrate competition of transmembrane transport.

Conclusion: For the first time the full substrate spectrum for all GLUT14 isoforms is described. This will inform future study designs assessing GLUT14's disease associations.

Introduction:

GLUT14 is encoded by the *SLC2A14* gene and a member of the family of facilitated glucose transporters with very recent evolutionary origins (Wu & Freeze, 2002). It is reported that two protein isoforms, denominated A and B, mediate glucose and dehydroascorbic acid transmembrane transport (Shaghaghi et al., 2017). However, recently three more protein isoforms were identified (Alhashim, 2022), but their substrates remain to be described.

Single nucleotide polymorphisms in the *SLC2A14* gene had been associated with inflammatory bowel diseases and Alzheimer's disease (Shaghaghi et al., 2017; Shulman et al., 2011; Wang et al., 2012), while copy number variation has been associated with rheumatoid arthritis, congenital heart defects in Turner Syndrome, and intraocular pressure (Nag et al., 2013; Prakash et al., 2016; Veal et al., 2014)

Differential expression of the *SLC2A14* gene was associated with disease outcomes in several cancers, including papillary thyroid carcinoma, gastric adenocarcinoma, glioblastoma, ovarian cancer, lymphoblastic leukemia, and colon cancer (Berlth et al., 2015; Chai et al., 2017; Januchowski, Zawierucha, Andrzejewska, Ruciński, & Zabel, 2013; Januchowski et al., 2014; Sharpe et al., 2021; Taylor et al., 2007; Valli et al., 2019). Besides cancers, links to Parkinson's Disease (Infante et al., 2015), and a differential expression in blastocysts development (Adjaye et al., 2007) were reported.

To elucidate disease mechanisms, it is of the highest priority to gain more information about the basic biology of all GLUT14 isoforms. Based on a high similarity with the GLUT3 protein, which is a multi-specific transporter, GLUT14 is expected to recognize more substrates than reported. GLUT3 is multi-specific and mediates the transmembrane transport of the hexose/hexose like substrates D-glucose, D-galactose, D-mannose, D-xylose and D-fucose, L-arabinose and D-lyxose (Deng et al., 2015). GLUT3 also recognizes some disaccharides and oligosaccharides, most notably maltose and cellobiose (Deng et al., 2015); however, these are likely to bind to the active site but not be translocated. In a very recent report, GLUT14-A was also described as a multi-specific transporter, exhibiting a substrate spectrum very similar to GLUT3 (Alhashim, 2022).

This study aims to determine the substrate spectra of the novel GLUT14 protein isoforms C, D and E.

Material and Methods

Plasmid Construct

A total of five different plasmids were used in the experiments. SLC2A14/GLUT14 isoform-A and isoform B subcloning were previously described (Amir Shaghaghi et al., 2017, Amir Shaghaghi et al., 2016, Alhashim, 2022) (Supplemental Data C1 and C2). Both isoforms were subcloned into a pcDNA3.5-V5-DEST vector. GLUT14 isoform C, isoform D and isoform E were subcloned into a 362 pCS mcherry DEST vectors with Gateway cloning technology (Alhashim, 2022) (**Supplemental Data C3, C4 and C5**).

PCR Amplification Of The Open Reading Frames (ORF) As Templates For Reverse Transcription

To create templates for the reverse transcription into cRNA, which was injected into the *Xenopus laevis* oocytes, open reading frames corresponding to GLUT14 C, D and E were PCR amplified from the existing plasmids using the following T7-tagged sense oligonucleotides: Isoform C: AAAATAATACGACTCACTATAGGGACCATGCTCCTGA GACGGCGCAATT; Isoform D: AAATAATACGACTCACTATAGGGACCATGCAAAGACTCCAACTGTTGAG; Isoform E: AAAATAATACGACTCACTATAGGGACCATGGATGGTTTTCTTCAGGCCC.

An existing M13 reserve motif in the 362 pCS mcherry DEST plasmid was used as the antisense oligonucleotide.

Phusion High-Fidelity DNA polymerase (New England Biolabs Inc., Ipswitch, MA, USA) was used with the following amplification steps: The initial denaturation temperature was 98 ºC for 30s, followed by 35 PCR cycles with the following conditions: The annealing temperature for isoform C was 65 ºC for 30 s and the annealing temperature for isoform D and E was 55 ºC for 30 s. The extension temperature was at 72 ºC for 1min. The final extension temperature was at 72 ºC for 10 mins. Amplified PCR products were run on 1% agarose gel to confirm the correct size and the PCR purification was performed with a Gel/PCR DNA Fragments Kit (Geneaid Biotech Ltd., New Taibei City, Taiwan, China) following the manufacturer's protocol.

In Virto Transcriptions of GLUT14 Isoform A, B, C, D And E

Restriction enzyme SacII (New England Biolabs Inc) was used to linearize the GLUT14 A and GLUT14 B plasmids. 6 ug DNA with 6 units of SacII were incubated in the PCR machine for 2 hours at 37 °C, followed with 15 mins incubation at 80 °C to inactive the SacII enzyme. Linearized vectors were purified with a Gel/PCR DNA Fragments Kit (Geneaid Biotech Ltd.), and measured the DNA concentration with Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) by optical density at 260/280 nm.

A mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA) was utilised to produce 5' capped complimentary RNA (cRNA) for injections into the *Xenopus laevis* oocytes. Purified and linearized GLUT14 isoform A and B vectors, and purified PCR products of GLUT14 isoforms C, D and E were used as DNA templates in the *in vitro* cRNA transcription. Between a minimum of 600 ng and a maximum of 1000ng of DNA templates were added into each reaction and followed by the manufacturer's protocol. All samples were incubated at 37 °C with extended incubation hours between 6 to 12 hours. DNase treatments were performed after incubation at 37C for 15mins. cRNA was recoveredrecovery employed an E.Z.N.A.® MicroElute RNA Clean-Up Kit (OMEGA Bio-TEK, Norcross, GA, USA). A Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) was used to measure the concentration of cRNA. After measurement, the cRNA concentration was adjusted to 500 ng/uLwith DNase/RNase free water. All cRNA samples were stored in an -80 °C freezer if not immediately used.

Oocyte Preparation

The animal protocol for using *Xenopus laevis* was approved by University of Manitoba Fort Garry animal care committee. Ovaries were harvested from adult female *Xenopus laevis* by following the standard procedure approved in the animal protocol. The ovaries were washed with calcium-free OR_2 buffer (4.82 g/L NaCl; 1.3 g/L HEPES; 2.5 mmol/L KCl, 1 mmol/L $MgCl_2$; 1 mmol/L Na_2HPO_4; pH=7.6) until the blood was washed out. 0.5 mg/L of Collagenase, Type IV, powder (Sigma-Aldrich) was added into the tube to digest the connective tissue around the ovary. The tube was rapped with tinfoil and put on a shaker at the lowest speed at room temperature. The digestion took 3 to 6 hours depending on the size of ovary.

After the ovaries were fully digested, they were washed with regular OR_2 buffer (4.82 g/L NaCl; 1.3 g/L HEPES; 2.5 mmol/L KCl, 1 mmol/L $MgCl_2$; 1 mmol/L Na_2HPO_4; 1 mmol/L $CaCl_2$; pH=7.6). The calcium in the OR_2 buffer stopped the collagenase function and prevented over digestion. Then, oocytes were transfer into a 60 mm petri-dish (FroggaBio Inc., Concord, ON, Canada) with OR_2 buffer for selection. Suitable oocytes (**Supplemental Data C6**) were selected for the injection experiment. The selected oocytes were stored in a new 60 mm petri-dish with OR_2 buffer plus 55 mg/L sodium pyruvate, 100 mg/L gentamicin (Sigma Aldrich) and 10mL/L Penicillin-Streptomycin (Sigma Aldrich), and the pH was adjusted to 7.6. This OR_2 buffer and antibiotic combination was named overnight buffer in our lab.

Injections of cRNA into the Xenopus laevis Oocytes

Injection needles were made from 3.5' glass capillaries (Drummond Scientific Company, Broomall, PA, USA) by a PC-10 Puller (Narishige International USA, Inc., Amityville, NY, USA). Either 6ul of cRNA sample or sterilized DNase/RNase-free water was backloaded into the needle with Eppendorf™ Femtotips™ Microloader Tips for Femtojet Microinjector (Eppendorf Canada Ltd., Mississauga, ON, Canada). The loaded needle was put into the Nanoject II Auto-Nanoliter Injector (Drummond Scientific Company). The needle was trimmed right before the injection start. 36.8 nl of cRNA (500 ng/L) or water (sham) was injected into each oocyte. After injection, the oocytes were transferred into a new petri-dish and incubated with overnight buffer at 16 ºC for 72 hours. During the 72 hours incubation, the overnight buffer was changed every 24 hours and the dead oocytes were removed when observed.

Radio-Isotope Uptake by Oocytes

Radiolabelled 2-Deoxy-D-Glucose (2-DDG [1,2-3H(N)], American Radiolabeled Chemicals Inc. (ARC), Saint Louis, MO, USA) was adjusted to $3.36*10^{-4}$ mmol/L in the experimental incubation OR_2 buffer. Different (inhibitory) sugars were added into the basic incubation buffer with the concentration adjusted to 0.3 mol/L. Inhibition sugars included glucose, galactose, fructose, mannose, xylose, arabinose, glucosamine, uric acid, maltose, lactose, sucrose and raffinose. 20 oocytes were considered one experimental group. Oocytes were incubated with the radiolabelled 2-D-deoxyglucose incubation buffer for 30mins at room

temperature. Then, the oocytes were washed five times with ice-cold OR2 buffer. The washed oocytes were then transferred into a 6mL pony vial (Perkinelmer Inc.). Next, 200 ul of 10% Sodium dodecyl sulfate (SDS) solution was added into each vial to dissolve the oocyte. Last, 2mL of Ultima Gold liquid scintillation cocktail (Perkinelmer Inc.) was added to each vial. The internal activity of each oocyte was counted by a Tri-Carb 3110TR Liquid Scintillation Analyzer (Perkinelmer Inc.) for 60 s. The CPM of each oocyte was originally collected from the machine and converted into fmol per oocyte.

Statistical Analysis

The experimental outliers were identified and removed by employing the 1.5 times inter-quartile range (IQR) rule. After the outliers were removed, the data were analyzed and plotted using Graphpad Prism 8 (version 8.4.2, Graphpad Software Inc, CA, USA). One-Way ANOVA with Bartlett's test and Brown-Forsythe test were used to determine statistical differences ($P<0.001$) between each sample group.

Results

The GLUT14 Protein Isoforms

All GLUT14 isoforms have been recently described isoforms A, B, D, and E show discrepancies in the first twenty-five N-terminal amino acids (**Supplemental Data B: Clustal protein alignment**). However, isoform C lacks a minimum of 100 N-terminal amino acids compared to the other isoforms, altering the predicted the transmembrane architecture (**Figure 4.1**).

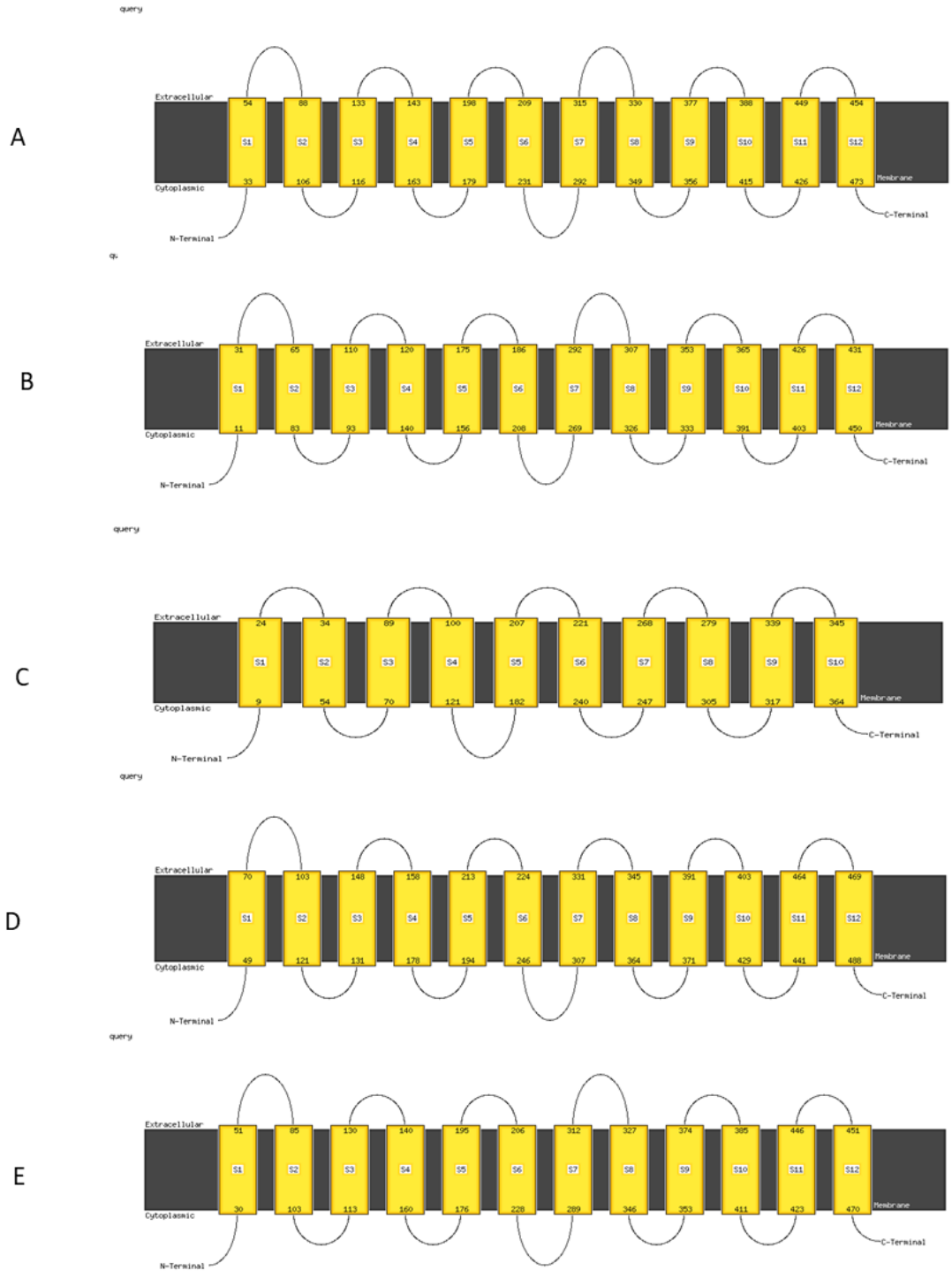


Figure 4.1. Predicted transmembrane architecture for all GLUT14 isoforms. Models were created with PHYRE2, http://www.sbg.bio.ic.ac.uk/phyre2/webscripts/jobmonitor.cgi?jobid=a779128089b48e8f.

GLUT14 Isoforms A, B, D, E Mediate 2-Deoxyglucose Uptake into Xenopus laevis Oocytes.

Under identical experimental conditions, all GLUT14 isoforms, except GLUT14-C, mediated the uptake of radiolabeled 2-deoxyglucose when expressed in *Xenopus laevis* oocytes. GLUT14 A showed highest uptake (127.8 fmol/oocyte/30 minutes), while isoforms B, D, and E had about 20% of this (21.66, 24.37 and 23.59 fmol/oocytes/30 minutes, respectively) (**Figure 4.2**).

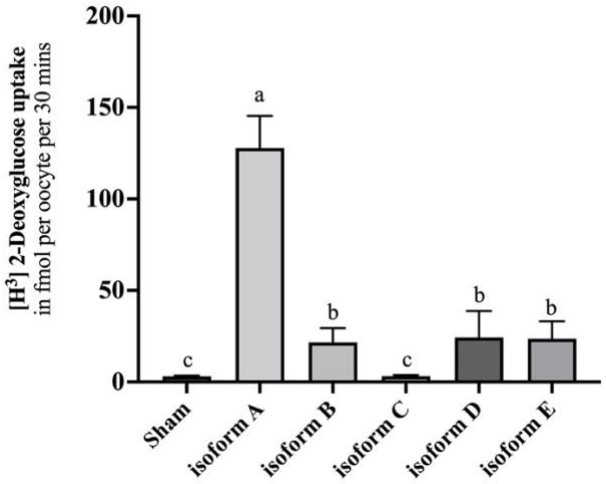


Figure 4.2. Uptake of radiolabeled 2-deoxyglucose into *Xenopus laevis* oocytes expressing GLUT14 isoforms. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose solution for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other ($n>15$, $p<0.0001$ in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

Monosaccharide(/Like) Substrates of GLUT14 Isoforms

Competitive inhibition of radiolabelled 2-deoxyglucose uptake was used to determine which previously reported GLUT substrates are recognised by GLUT14 isoforms B, D, and E.

For GLUT 14 isoform B, glucose, galactose, fructose, mannose, xylose, arabinose, and glucosamine strongly inhibited radiolabelled 2-deoxyglucose uptake, while uric acid showed a more moderate inhibition (**Figure 4.3**).

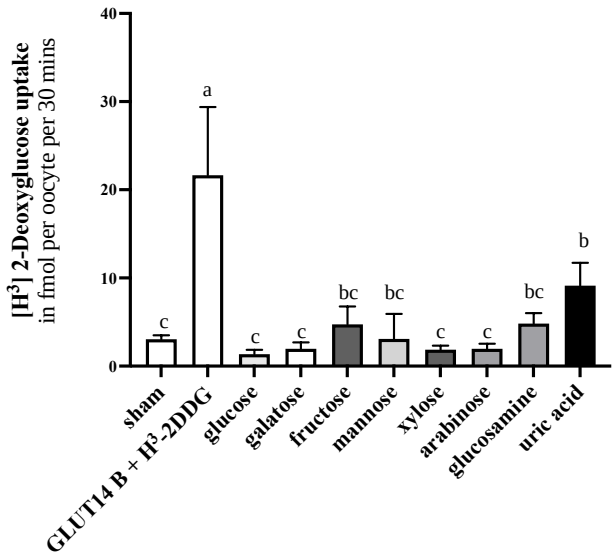


Figure 4.3. Competitive inhibition of radiolabelled 2-deoxyglucose uptake exhibited by selected monosaccharides or monosaccharide-like molecules for GLUT14-B. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose and 0.3 mol/L of inhibitor for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other ($n>15$, $p<0.0001$ in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

For GLUT 14 isoform D, glucose, galactose, mannose, xylose, strongly inhibited radiolabelled 2-deoxyglucose uptake, while fructose, arabinose and glucosamine showed moderate inhibition, while uric acid did not inhibit (**Figure 4.4**).

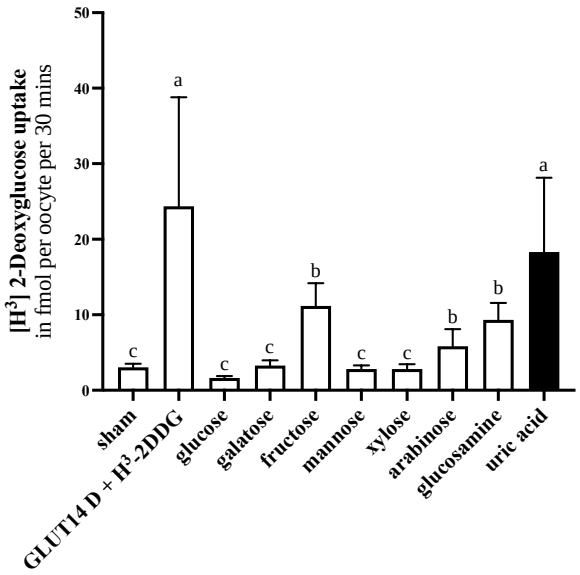


Figure 4.4. Competitive inhibition of radiolabelled 2-deoxyglucose uptake exhibited by selected monosaccharides or monosaccharide-like molecules for GLUT14-D. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose + 0.3 mol/L inhibitor for 30 mins. Error bars represent the standard deviations; Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other (n>15, p<0.0001 in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

For GLUT 14 isoform E, glucose, galactose, fructose, mannose, xylose, and arabinose strongly inhibited radiolabelled 2-deoxyglucose uptake, while glucosamine and uric acid showed moderate inhibition (**Figure 4.5**).

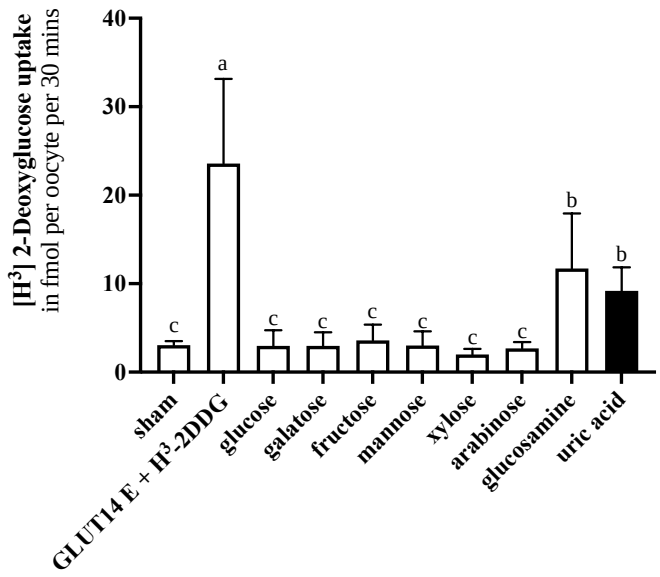


Figure 4.5 Competitive inhibition of radiolabelled 2-deoxyglucose uptake exhibited by selected monosaccharides or monosaccharide-like molecules for GLUT14-E. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose + 0.3 mol/L inhibition sugar solution for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other (n>15, p<0.0001 in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

Competitive Inhibition Experiment of GLUT14 Isoforms B, D, E With

Disaccharides/Oligosaccharides

Competitive inhibition of radiolabelled 2-deoxyglucose uptake was used to determine which disaccharides can bind to the GLUT14 isoforms B, D and E. However, in contrast to the monosaccharides (above), the disaccharides and oligosaccharides are not translocated through the membrane and therefore inhibit the protein by competitive binding on the active site of the solute carrier (Deng et al., 2015)

For GLUT 14 isoform B, maltose and lactose strongly inhibited radiolabelled 2-deoxyglucose uptake, while sucrose and raffinose weakly reduced the transport activity (**Figure 4.6**).

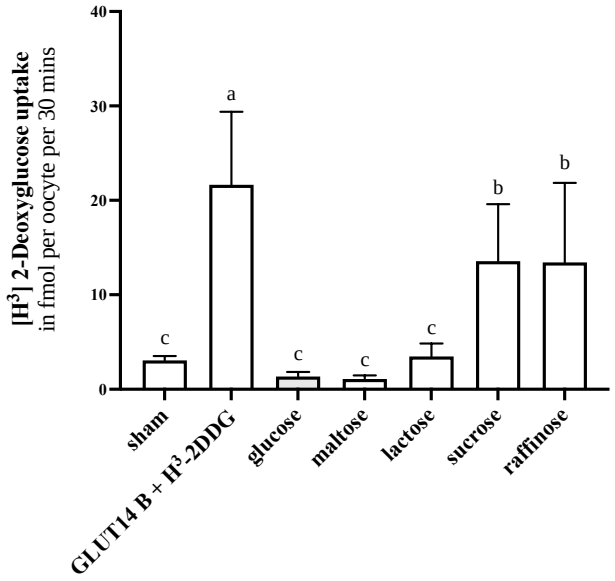


Figure 4.6. Competitive inhibition of radiolabelled 2-deoxyglucose uptake by selected disaccharides and oligosaccharides for GLUT14-B. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose + 0.3 mol/L inhibitor for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other ($n>15$, $p<0.0001$ in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

For GLUT 14 isoform D, only maltose strongly inhibited radiolabelled 2-deoxyglucose uptake, while lactose moderately reduced the transport activity, and sucrose and raffinose did not affect it (**Figure 4.7**).

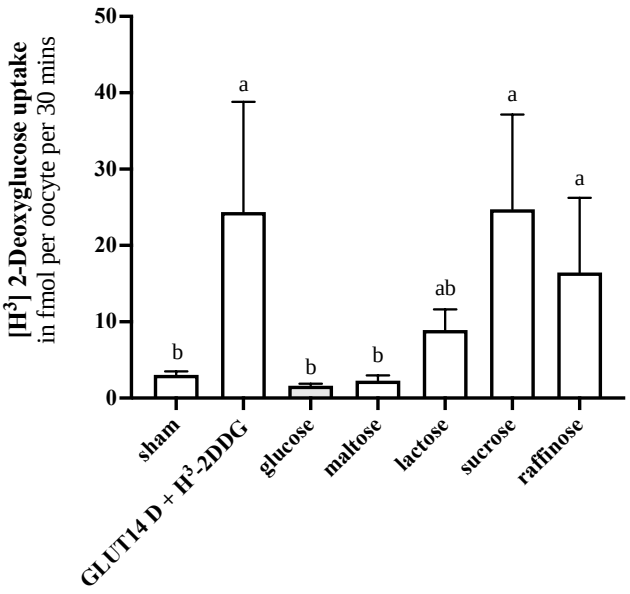


Figure 4.7. Competitive inhibition of radiolabelled 2-deoxyglucose uptake selected disaccharides and oligosaccharides for GLUT14-D. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose + 0.3 mol/L inhibitor for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other (n>15, p<0.0001 in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

For GLUT 14 isoform E, maltose, lactose, sucrose and raffinose all strongly inhibited radiolabelled 2-deoxyglucose uptake (**Figure 4.8**).

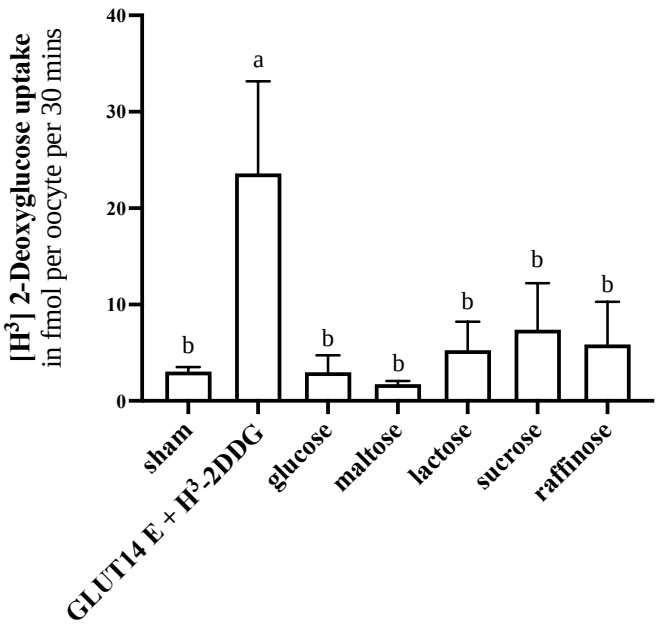


Figure 4.8 Competitive inhibition of radiolabelled 2-deoxyglucose uptake selected disaccharides and oligosaccharides for GLUT14-E. Oocytes were incubated in 3.36 x 10^{-4} mmol/L H^3-2-deoxyglucose or 3.36 x 10^{-4} mmol/L solution or H^3-2-deoxyglucose + 0.3 mol/L inhibitor for 30 mins. Error bars represent the standard deviations; the different small letters indicate the significant differences between groups. a, b and c are all significantly different from each other ($n>15$, $p<0.0001$ in ANOVA/ Bartlett's test/ Brown-Forsythe test to detect and confirm the statistically significant differences).

Discussion:

This report substantially expands on the knowledge of the basic biology of the GLUT14 protein isoforms.

First Functional Characterisations of GLUT14 Isoforms C, D And E

Utilizing the non-existing background in the *Xenopus laevis* oocyte expression system (Bentley, Shao, Misra, Morielli, & Zhao, 2012), GLUT14 isoforms D and E were identified as glucose/sugar transmembrane carriers. Previously, glucose and dehydroascorbic acid transport were reported for GLUT14-A and GLUT14-B (Shaghaghi et al., 2017). We also conclude that GLUT14 isoform C is not a functional glucose solute carrier. Future experiments need to elucidate if isoform C has an alternative substrate or is non-functional due to the elimination of more than 109 N-terminal amino acids. Both scenarios appear feasible considering that targeting to the plasmalemma membrane is reported for all GLUT14 isoforms, including isoform C

(Alhashim, 2022). Moreover, the interactive regulation between all *SLC2A14* transcripts should be investigated in human cells to determine possible antag- or synergisms to regulate GLUT14 functioning.

Substrates multi-specificity of GLUT14 isoforms B, D and E

This report describes GLUT14 isoforms B, D and E as multi-specific solute carriers, which had recently been reported for GLUT14-A (Alhashim, 2022). A substrate spectrum typical for the GLUT-type sugar transporters, including glucose, galactose, fructose, mannose, xylose, arabinose, glucosamine and uric acid is reported. The substrate recognition is specifically similar to the closest paralogue *SLC2A3*/GLUT3 (Custódio, Paulsen, Frain, & Pedersen, 2021). This functionally confirms their evolutionary closeness, which had been reported based on their transcripts and protein sequences similarities (Wu & Freeze, 2002). Therefore, it could be postulated that GLUT14's canonical substrates include the hexoses glucose, galactose, mannose and the pentose xylose, mirroring GLUT3.

However, a few exceptions in substrates utilization between the GLUT14 isoforms should be noted. Uric acid seems to be a low-affinity substrate of isoforms B and E, but not GLUT14-D. Fructose appeared to be a substantial substrate for isoforms B and E, while its transport is weak for GLUT14-D. Moreover, no recognition of fructose as a substrate was recently reported for GLUT14-A (Alhashim, 2022). These discrepancies likely originate from the sequence discrepancies in the N-termini, and how these cause the differences could be addressed by future experiments. Moreover, glucosamine and uric acid, which are not saccharides but are recognized by other members of the GLUT family (Holman, 2020), showed variable activity in the inhibition experiments and might be weakly recognized substrates for some GLUT14 isoforms.

Although dehydroascorbic acid was not evaluated in the present report, the substrate spectra allow for extrapolating that it is a substrate for the functional GLUT14 isoforms, as dehydroascorbic acid transport was already reported for GLUT14-A and -B (Shaghaghi et al., 2017) and GLUT3 (Rumsey et al., 1997).

Disaccharides as GLUT14 inhibitors

The disaccharide maltose, lactose, sucrose and the trisaccharide raffinose were recognized inhibitors to varying degrees and mostly weak. Maltose binding to GLUT3 and GLUT1 had been demonstrated and classified as exofacial inhibition, where maltose is recognized by the substrate-binding site but cannot be translocated through the membrane transporter protein (Deng et al., 2015). The same phenomenon appears on GLUT14 isoforms with the identified disaccharides and oligosaccharides. However, the physiologic relevance of such an allosteric inhibition remains low, since these higher-order saccharides are rapidly digested.

GLUT14 and Disease Association

The identification of the substrates establishes a baseline to investigate the underlying mechanism of reported disease associations. *SLC2A14* is differentially expressed in acute lymphoblastic leukemia (Taylor et al., 2007), developing blastocysts (Adjaye et al., 2007), drug-resistant ovarian cancer cells (Januchowski et al., 2013, 2014), Parkinson's Disease (Infante et al., 2015), gastric adenocarcinoma (Berlth et al., 2015), papillary thyroid carcinoma (Chai et al., 2017), human skin fibrosis (Song et al., 2002), hypoxic colon cancer cells (Valli et al., 2019), and Glioblastoma (Sharpe et al., 2021). Individual *SLC2A14* single nucleotide polymorphisms (SNPs) have been associated with Alzheimer's disease (Shulman et al., 2011; Wang et al., 2012), and inflammatory bowel diseases (Shaghaghi et al., 2017). In regard to the variety of associations reported, it should be noted that all but one of them are derived from one study only, and future replication would strengthen the evidence.

Five types of cancers are associated with the differential regulation of GLUT14. Generally, glucose consumption in cancer cells is upregulated, and the upregulation of closely related GLUT1 and GLUT3 is one mechanism to elevate the basal uptake and the flux of sugar into metabolic pathways (Holman, 2018). We speculate that GLUT14 upregulation in cancer cells serves the same purpose. Therefore, GLUT14 should also be considered to be a target for inhibitors to reduce glucose availability to cancer cells. Moreover, an interesting avenue of future research would be the investigations of GLUT14 as a biomarker for certain cancers and a possible correlation of differential expression with survival.

Conclusion

GLUT14 isoforms A, B, D, E are multi-specific solute carries with the canonical substrates of glucose, galactose, mannose, and xylose. The transport capacity is highest for GLUT14-A. Other substrates reported for the GLUT family transporters are weakly recognised. This establishes a baseline on substrate specificity, which can be utilised for future clinical studies.

References

Adjaye, J., Herwig, R., Brink, T. C., Herrmann, D., Greber, B., Sudheer, S., … Niemann, H. (2007). Conserved molecular portraits of bovine and human blastocysts as a consequence of the transition from maternal to embryonic control of gene expression. *Physiological Genomics*, *31*(2), 315–327. https://doi.org/10.1152/physiolgenomics.00041.2007

Alhashim, A. (2022). *The homo-sapiens solute carrier family 2 member 14 (SLC2A14) – further insight into the genomic organization, protein isoforms, substrates and kinetics* (p. 56). p. 56.

Bentley, P. A., Shao, Y., Misra, Y., Morielli, A. D., & Zhao, F. Q. (2012). Characterization of bovine glucose transporter 1 kinetics and substrate specificities in Xenopus oocytes. *Journal of Dairy Science*, *95*(3), 1188–1197. https://doi.org/10.3168/jds.2011-4430

Berlth, F., Mönig, S., Pinther, B., Grimminger, P., Maus, M., Schlösser, H., … Alakus, H. (2015). Both GLUT-1 and GLUT-14 are Independent Prognostic Factors in Gastric Adenocarcinoma. *Annals of Surgical Oncology*, *22*, 822–831. https://doi.org/10.1245/s10434-015-4730-x

Chai, Y. J., Yi, J. W., Oh, S. W., Kim, Y. A., Yi, K. H., Kim, J. H., & Lee, K. E. (2017). Upregulation of SLC2 (GLUT) family genes is related to poor survival outcomes in papillary thyroid carcinoma: Analysis of data from The Cancer Genome Atlas. *Surgery (United States)*, *161*(1), 188–194. https://doi.org/10.1016/j.surg.2016.04.050

Custódio, T. F., Paulsen, P. A., Frain, K. M., & Pedersen, B. P. (2021). Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family. *Life Science Alliance*, *4*(4), 1–12. https://doi.org/10.26508/LSA.202000858

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., … Yan, N. (2015). Molecular basis of ligand recognition and transport by glucose transporters. *Nature*, Vol. 526, pp. 391–396. https://doi.org/10.1038/nature14655

Duell, E. J., Lujan-Barroso, L., Llivina, C., Muñoz, X., Jenab, M., Boutron-Ruault, M. C., … González, C. A. (2013). Vitamin C transporter gene (SLC23A1 and SLC23A2) polymorphisms, plasma vitamin C levels, and gastric cancer risk in the EPIC cohort. *Genes and Nutrition*, *8*(6), 549–560. https://doi.org/10.1007/s12263-013-0346-6

Holman, G. D. (2018). Chemical biology probes of mammalian GLUT structure and function. *Biochemical Journal*, *475*(22), 3511–3534. https://doi.org/10.1042/BCJ20170677

Holman, G. D. (2020). Structure, function and regulation of mammalian glucose transporters of the SLC2 family. *Pflugers Archiv European Journal of Physiology*, *472*(9), 1155–1175. https://doi.org/10.1007/s00424-020-02411-3

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., … Sainz, J. (2015). Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. *Neurobiology of Aging*, *36*(2), 1105–1109. https://doi.org/10.1016/j.neurobiolaging.2014.10.039

Januchowski, R., Zawierucha, P., Andrzejewska, M., Ruciński, M., & Zabel, M. (2013). Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines. *Biomedicine and Pharmacotherapy*, *67*(3), 240–245. https://doi.org/10.1016/j.biopha.2012.11.011

Januchowski, R., Zawierucha, P., Ruciński, M., Andrzejewska, M., Wojtowicz, K., Nowicki, M., & Zabel, M. (2014). Drug transporter expression profiling in chemoresistant variants of the A2780 ovarian cancer cell line. *Biomedicine and Pharmacotherapy*, *68*(4), 447–453. https://doi.org/10.1016/j.biopha.2014.02.002

Liu, H., Zhang, Y., Zhang, H., Wang, L., Wang, T., Han, Z., … Liu, G. (2021). Effect of plasma vitamin C levels on Parkinson's disease and age at onset: a Mendelian randomization study. *Journal of Translational Medicine*, *19*(1), 1–9. https://doi.org/10.1186/s12967-021-02892-5

Nag, A., Venturini, C., Hysi, P. G., Arno, M., Aldecoa-Otalora Astarloa, E., Macgregor, S., … Hammond, C. J. (2013). Copy number variation at chromosome 5q21.2 is associated with intraocular pressure. *Investigative Ophthalmology & Visual Science*, *54*(5), 3607–3612. https://doi.org/10.1167/iovs.13-11952

Prakash, S. K., Bondy, C. A., Maslen, C. L., Silberbach, M., Lin, A. E., Perrone, L., … Milewicz, D. M. (2016). Autosomal and X chromosome structural variants are associated with congenital heart defects in Turner syndrome: The NHLBI GenTAC registry. *American Journal of Medical Genetics, Part A*, *170*(12), 3157–3164. https://doi.org/10.1002/ajmg.a.37953

Rumsey, S. C., Kwon, O., Xu, G. W., Burant, C. F., Simpson, I., & Levine, M. (1997). Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid. *Journal of Biological Chemistry*, *272*(30), 18982–18989. https://doi.org/10.1074/jbc.272.30.18982

Shaghaghi, M. A., Zhouyao, H., Tu, H., El-Gabalawy, H., Crow, G. H., Levine, M., … Eck, P. (2017). The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease. *American Journal of Clinical Nutrition*, *106*(6), 1508–1513. https://doi.org/10.3945/ajcn.116.147603

Sharpe, M. A., Ijare, O. B., Baskin, D. S., Baskin, A. M., Baskin, B. N., & Pichumani, K. (2021). The leloir cycle in glioblastoma: Galactose scavenging and metabolic remodeling. *Cancers*, *13*(8). https://doi.org/10.3390/cancers13081815

Shulman, J. M., Chipendo, P., Chibnik, L. B., Aubin, C., Tran, D., Keenan, B. T., … De Jager, P. L. (2011). Functional screening of alzheimer pathology genome-wide association signals in drosophila. *American Journal of Human Genetics*, *88*(2), 232–238. https://doi.org/10.1016/j.ajhg.2011.01.006

Song, J., Kwon, O., Chen, S., Daruwala, R., Eck, P., Park, J. B., & Levine, M. (2002). Flavonoid inhibition of sodium-dependent vitamin C transporter 1 (SVCT1) and glucose transporter isoform 2 (GLUT2), intestinal transporters for vitamin C and glucose. *Journal of Biological Chemistry*, *277*(18), 15252–15260. https://doi.org/10.1074/jbc.M110496200

Taylor, K. H., Pena-Hernandez, K. E., Davis, J. W., Arthur, G. L., Duff, D. J., Shi, H., … Caldwell, C. W. (2007). Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. *Cancer Research*, *67*(6), 2617–2625. https://doi.org/10.1158/0008-5472.CAN-06-3993

Valli, A., Morotti, M., Zois, C. E., Albers, P. K., Soga, T., Feldinger, K., … Harris, A. L. (2019). Adaptation to HIF1a deletion in hypoxic cancer cells by upregulation of GLUT14 and creatine metabolism. *Molecular Cancer Research*, *17*(7), 1531–1544. https://doi.org/10.1158/1541-7786.MCR-18-0315

Veal, C. D., Reekie, K. E., Lorentzen, J. C., Gregersen, P. K., Padyukov, L., & Brookes, A. J. (2014). A 129-kb deletion on chromosome 12 confers substantial protection against rheumatoid arthritis, implicating the gene SLC2A3. *Human Mutation*, *35*(2), 248–256. https://doi.org/10.1002/humu.22471

Wang, W., Yu, J. T., Zhang, W., Cui, W. Z., Wu, Z. C., Zhang, Q., & Tan, L. (2012). Genetic association of SLC2A14 polymorphism with Alzheimer's disease in a Han Chinese population. *Journal of Molecular Neuroscience*, *47*(3), 481–484. https://doi.org/10.1007/s12031-012-9748-y

Wu, X., & Freeze, H. H. (2002). GLUT14, a duplicon of GLUT3, is specifically expressed in testis as alternative splice forms. *Genomics*, *80*(6), 553–557. https://doi.org/10.1006/geno.2002.7010

Supplemental information

Supplemental Information

Supplemental Data A: Amino Acid Sequences for all GUT14 protein isoforms. Current sequences were created using Sequencher™ software earlier and cross referenced to the NCBI database, March 7th, 2022.

>GLUT14A
MEFHNGGHVSGIGGFLVSLTSRMKPHTLAVTPALIFAITVATIGSFQFGYNTGVINAPETIIKEFINKTLTDKA
NAPPSEVLLTNLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLIL
GRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAILQ
SAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPII
ISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNTIFTLLSLFLVERAGRRTLHMIGLGGMAFC
STLMTVSLLLKNHYNGMSFVCIGAILVFVACFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVG
LLFPSAAYYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAK
ETTTNV

>GLUT14B
MDNRQNVTPALIFAITVATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVG
GMIGSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTGFVPMYIGEIS
PTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAILQSAALPCCPESPRFLLINRKKEEN
ATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTG
IFKDAGVQQPIYATISAGVVNTIFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVC
IGAILVFVACFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFTGFLI
TFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPA
KETTTNV

>GLUT14C
MLLRRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTL
NQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAILQSAALPCCPESPRFLLINRKKEENATRILQRLWGT
QDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPI
YATISAGVVNTIFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVACF
EIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFTGFLITFLAFTFFKVP
ETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV

>GLUT14D

MQRLQLLRVEVLLGVKQGDEMRHFFFSSQTSTLEKSQNGGVGEEVTPALIFAITVATIGSFQFGYNTGVINA
PETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAATGGC
LMGLCKIAESVEMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSE
ELWPVLLGFTILPAILQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEK
QVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNTIFTLLSLFLVER
AGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVACFEIGPGPIPWFIVAELFSQGPRPA
AMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGA
DRSGKDGVMGMNSIEPAKETTTNV

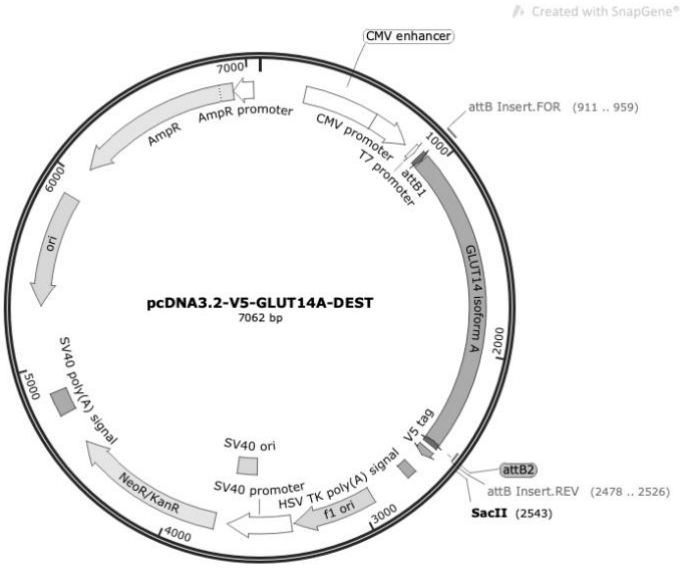
>GLUT14E
MDGFLQAHKSQTSTLEKSQNGGVGEEVTPALIFAITVATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANA
PPSEVLLTNLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGR
LVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAILQSA
ALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISI
VLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNTIFTLLSLFLVERAGRRTLHMIGLGGMAFCST
LMTVSLLLKNHYNGMSFVCIGAILVFVACFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLL
FPSAAYYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKET
TTNV

Supplemental Data B: Clustal protein alignment for all GLUT14 protein isoforms

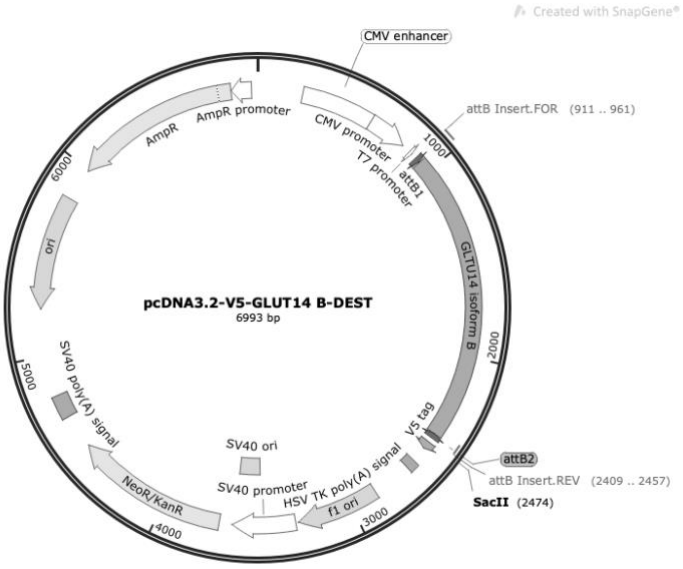
Obtained from EBI, 2022-01-19:
https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20220120-043755-0092-40322447-p2m

```
GLUT14E      ----------------------MDGFLQAHKSQTSTLEKSQNGGVGEEVTPALIFAITV       37
GLUT14D      MQRLQLLRVEVLLGVKQGDE---MRHF--FFSSQTSTLEKSQNGGVGEEVTPALIFAITV       55
GLUT14B      ------------------------------------------MDNRQNVTPALIFAITV       17
GLUT14A      ------------MEFHNGGHVSGIGGF--LVS-LT-----SRMKPHTLAVTPALIFAITV       40
GLUT14C      ------------------------------------------------------------       0


GLUT14E      ATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVGGMI       97
GLUT14D      ATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVGGMI       115
GLUT14B      ATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVGGMI       77
GLUT14A      ATIGSFQFGYNTGVINAPETIIKEFINKTLTDKANAPPSEVLLTNLWSLSVAIFSVGGMI       100
GLUT14C      ------------------------------------------------------------       0


GLUT14E      GSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTG       157
GLUT14D      GSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTG       175
GLUT14B      GSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTG       137
GLUT14A      GSFSVGLFVNRFGRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTG       160
GLUT14C      ---------MLLRRRNSMLIVNLLAATGGCLMGLCKIAESVEMLILGRLVIGLFCGLCTG       51
                      : ********************************************


GLUT14E      FVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAI       217
GLUT14D      FVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAI       235
GLUT14B      FVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAI       197
GLUT14A      FVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAI       220
GLUT14C      FVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLELILGSEELWPVLLGFTILPAI       111
             ************************************************************


GLUT14E      LQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVT       277
GLUT14D      LQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVT       295
GLUT14B      LQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVT       257
GLUT14A      LQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVT       280
GLUT14C      LQSAALPCCPESPRFLLINRKKEENATRILQRLWGTQDVSQDIQEMKDESARMSQEKQVT       171
             ************************************************************


GLUT14E      VLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNT       337
GLUT14D      VLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNT       355
GLUT14B      VLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNT       317
GLUT14A      VLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNT       340
GLUT14C      VLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQQPIYATISAGVVNT       231
             ************************************************************


GLUT14E      IFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVAC       397
GLUT14D      IFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVAC       415
GLUT14B      IFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVAC       377
GLUT14A      IFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVAC       400
GLUT14C      IFTLLSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKNHYNGMSFVCIGAILVFVAC       291
             ************************************************************


GLUT14E      FEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFT       457
GLUT14D      FEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFT       475
GLUT14B      FEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFT       437
GLUT14A      FEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFT       460
GLUT14C      FEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLFPSAAYYLGAYVFIIFT       351
             ************************************************************


GLUT14E      GFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV       517
GLUT14D      GFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV       535
GLUT14B      GFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV       497
GLUT14A      GFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV       520
GLUT14C      GFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMGMNSIEPAKETTTNV       411
             ************************************************************
```

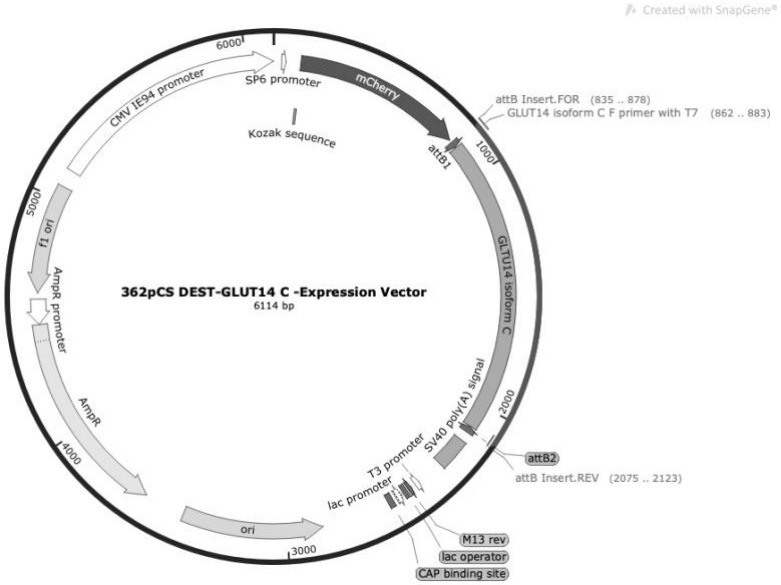
Supplemental Data C1: plasmid construct for GLUT14 isoform A in pcDNA3.2-V5-DEST vector.



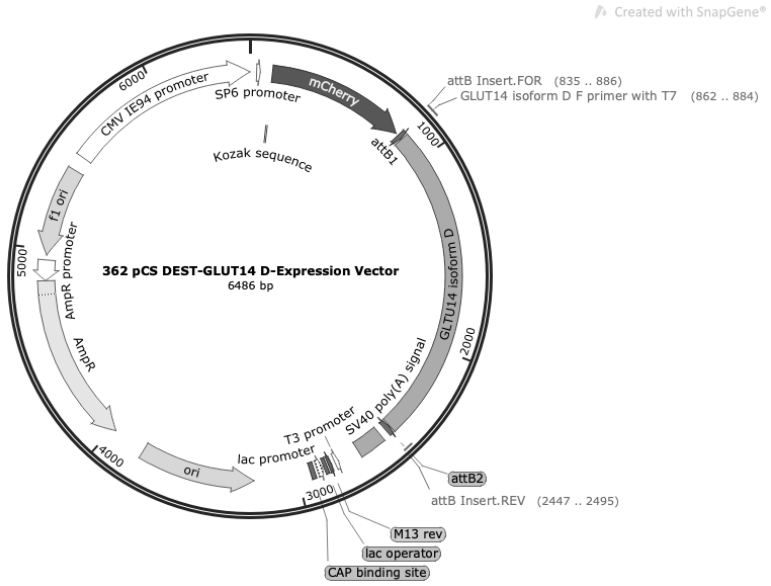
Supplemental Data C2: plasmid construct for GLUT14 isoform B in pcDNA3.2-V5-DEST vector.



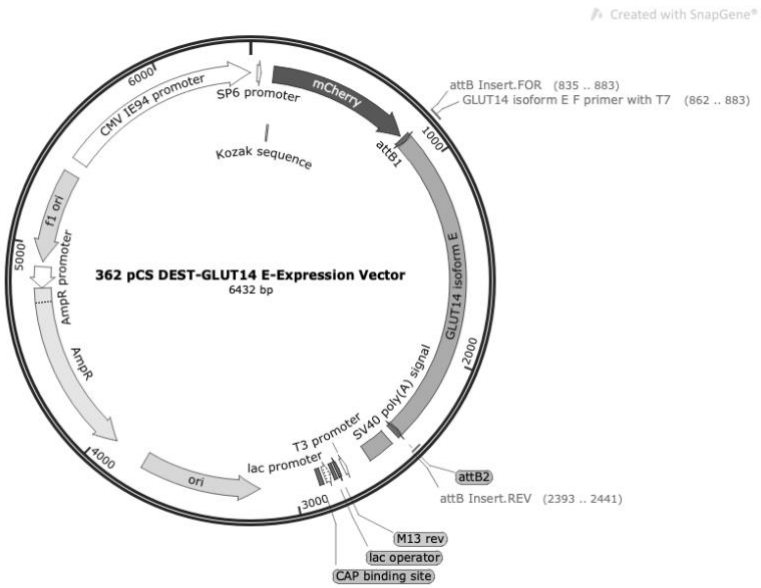
Supplemental Data C3: plasmid construct for GLUT14 isoform C in 362 pCS mCherry-DEST vector.



Supplemental Data C4: plasmid construct for GLUT14 isoform D in 362 pCS mCherry-DEST vector.

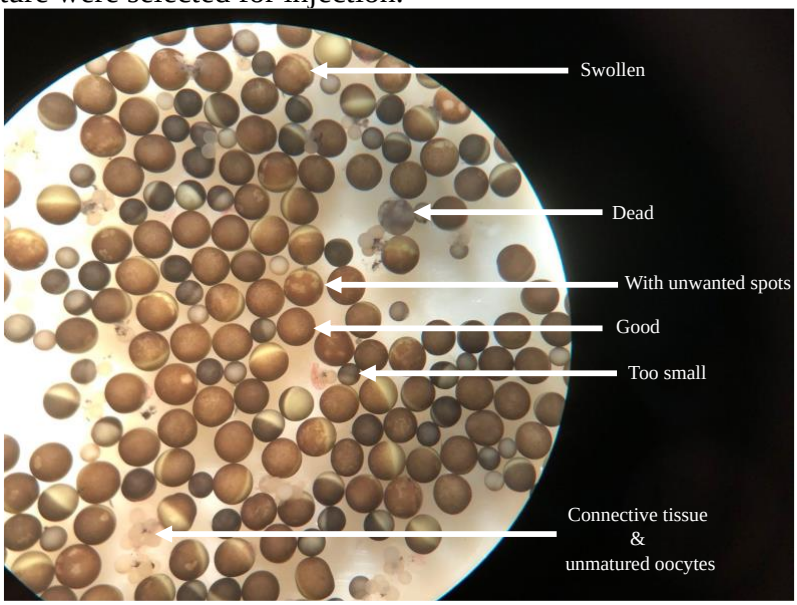


Supplemental Data C5: plasmid construct for GLUT14 isoform E in 362 pCS mCherry-DEST vector.



Supplemental Data C6: Oocyte selection during the injection experiment.

The following picture indicates the different stages of the oocytes. Only the good or suitable oocytes in the picture were selected for injection.



Chapter 5: General Discussion

The presented data add some pieces to the puzzle of possible genetic and disease associations with *SLC23A1* and *SLC2A14's*.

The first 30 N-terminal SLC23A1 amino acids seem to be completely redundant, and the elimination of the first 77 amino acids decreases the transport activity by only 20%. This confirms Hypothesis 2 (The N-terminus of the SLC23A1/SVCT1 protein is redundant, and truncated proteins utilizing alternative translational start sites are therefore fully functional). Therefore, the missense variations in this region can be de-prioritized for functional studies, since their presence will unlikely change the proteins functioning.

The function and substrates for all GLUT14 isoforms are now identified. Hypothesis 1 (The orphan isoforms of GLUT14 are encoding for multi-specific hexose carriers) was confirmed for GLUT14 B, D, and E, but not for GLUT14-C, which is nonfunctional.

Implications for investigations into genetic disease associations

SLC23A1 SNP rs33972313 and *SLC2A14* SNP rs10845990 are both associated with Alzheimer's disease (AD) (Chen et al., 2021; Wang et al., 2012). Similarly, at *SLC23A1* SNP rs10063949 and *SLC2A14* SNPs rs10846086, rs2889504, and rs12815313 are associated with Inflammatory bowel disease (IBD) (Mandana Amir; Shaghaghi et al., 2017; Mandana Amir Shaghaghi, Bernstein, León, El-Gabalawy, & Eck, 2014). Moreover, dysregulated of *SLC2A14* expression and *SLC23A1* SNP rs33972313 are associated with an increased risk of Parkinson's disease (PD) (Duell et al., 2013; Infante et al., 2015; Liu et al., 2021). Nevertheless, all the above variations are intronic and very likely serve as markers for causal variations in the locus, which still must be identified. However, the identifications of both transporters' substrate might allow some extrapolations if the common substrate Vitamin C or the distinct hexoses could play a role in disease development and severity.

Inflammatory Bowel Diseases

Inflammatory Bowel Diseases is a broad term that is used to describe chronic inflammation in the digestive tract, mainly Ulcerative colitis (UC) and Crohn's disease (CD), where the barrier

and immune functions are dysregulated. To date, the exact causes of IBD remain unknown, but genetic variations and environmental factors such as the diet play a part (Olsen, 2009).

As an antioxidant that inhibits oxidative stress in cells, Vitamin C shows a beneficial effect in the prevention and treatment of IBD (Jarmakiewicz-Czaja, Piątek, & Filip, 2020). This might be related to the transport activity of both SVCT1 and GLUT14, since the oxidized and reduced form of Vitamin C is present in the intestinal lumen.

Recent single cell RNA sequencing data show *SLC2A14* expression in Paneth cells and granulocytes of the small intestine and colon, but not in enterocytes (Alhashim, 2022). On the other hand, *SLC23A1* is highly expressed in small intestinal enterocytes and moderately expressed in colonic enterocytes, but not in significant amounts in granulocytes (**Figure 5.1**). This implicates two different mechanisms underlying both genes disease associations. *SLC2A14* acts in immune cells, while *SLC23A1* in epithelial barrier cells.

The *SLC23A1* expression in enterocytes implies a role of intracellular ascorbic acid to maintain the epithelial barrier functions, either through its antioxidant functions but more likely by the modulation of extracellular matrix protein hydroxylation and crosslinking (Padayatty et al., 2003). This also means that genetically reduced SLC23A1 activity could be compensated by oral dehydroascorbic acid, which provides an avenue for interventions of affected individuals, if they ever get identified.

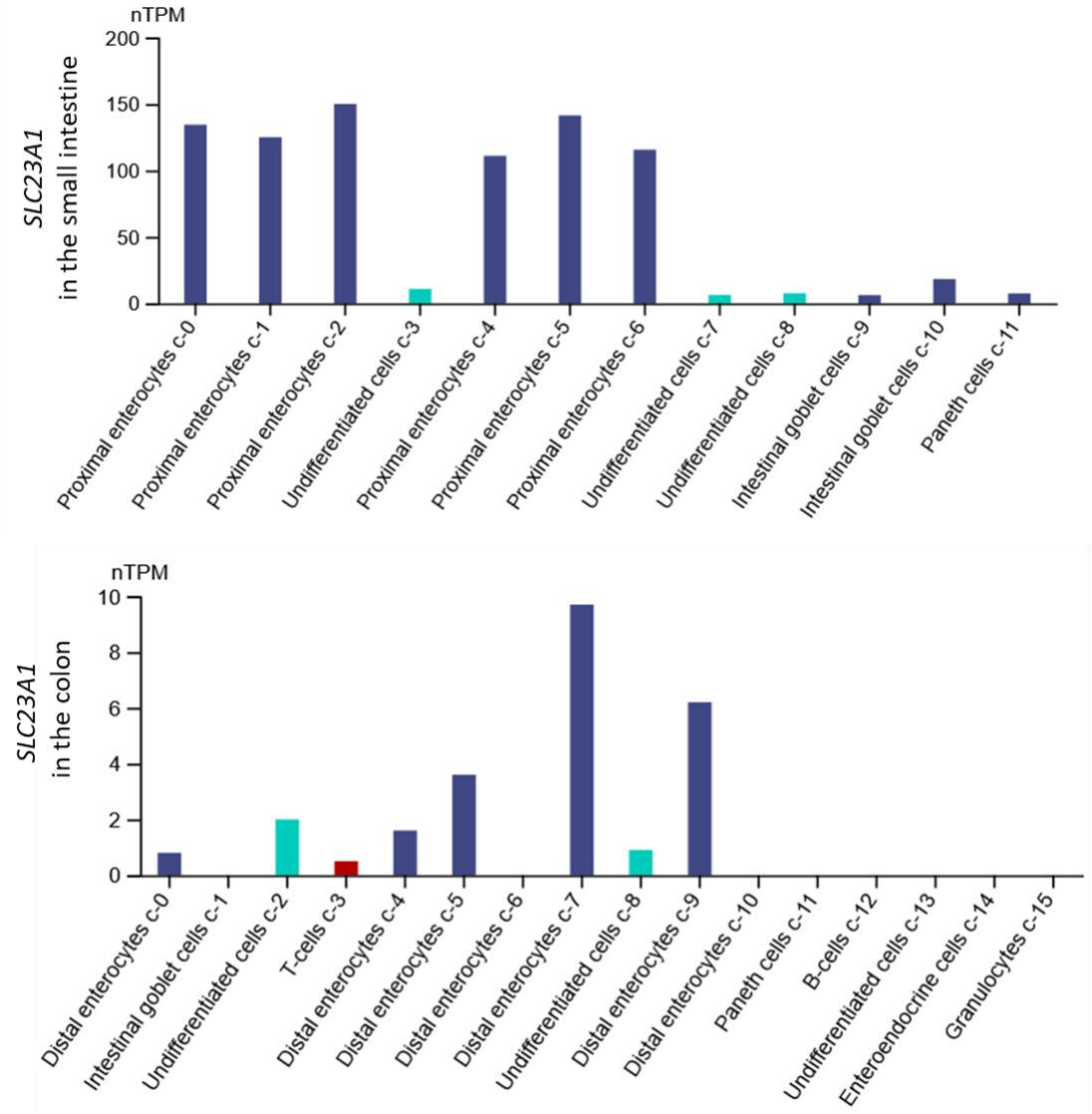


Figure 5.1: *SLC23A1* expression levels in intestinal cells in normalised transcript per million (nTPM). The graph is obtained from Human Protein Atlas: A single-cell type transcriptomics map of human tissues (Karlsson et al. 2021). Copyright 2022 by Human Protein Atlas, reprinted with permission.

The *SLC2A14* expression in intestinal innate immune cells strongly implies that disease mechanisms are mediated through a dysregulated immune response, supporting the notion that a dysregulated immune response against the intestinal microbiome is a key driver of intestinal inflammation (Sartor & Wu, 2017). However, due to GLUT14's multi-specific nature it is less clear which substrate dysregulation might cause the intestinal inflammation.

Innate immune cells, specifically neutrophiles undergo oxidative burst to kill invading microbes with reactive oxygen species (Mills et al., 2016). In the process, extracellular ascorbic

acid is oxidized to dehydroascorbic acid (Washko, Wang, & Levine, 1993), which is one of the substrates for GLUT14. The GLUT mediated transport of dehydroascorbic acid leads to a very high influx and conversion to ascorbic acid. If this is compromised the immune cell's function might be compromised, likely due to a lack of antioxidant equivalent which might lead to decreased survival (Washko et al., 1993). However, innate immune functioning, including the oxidative burst, also demands increased energy, which can be provided through elevated glucose influx. Reduced GLUT14 activity could result in reduced glucose flux and therefore aggravate inflammation. The same principle applies for mannose, which is proposed to be a protective factor against inflammation (Zhang et al., 2021). Contrary, high-fructose uptakes are associated with an increased risk of development IBD (Montrose et al., 2021), specifically with reduced thickness of the colonic mucosa (Montrose et al., 2021)., Knowing all the GLUT14 substrates should aid investigations to elucidate which of these possible disease mechanisms might be relevant.

Alzheimer's and Parkinson's Disease

Alzheimer's disease is a progressive neurological disease with mild memory loss at the beginning and dementia at the end-stage. Incidence in Canada exceeded 500,000 in 2021 and numbers are predicted to increase to 912,000 in 2030 (Chambers, Bancej, & McDowell, 2016).

Parkinson's Disease is another type of chronic neurological disease that will cause movement disorders, memory difficulties, sleeping problems and fatigue. There are about 100,000 people currently living with Parkinson's Disease in Canada, and this number keeps increasing ("Parkinsonism in Canada - Canada.ca," 2019). The risk of AD and PD are both increased with age, family history and certain genetic variations. The mechanisms of development of these two diseases are not fully understood. But emerging evidence indicated glucose and ascorbic acid metabolism could strongly affect development of these diseases (Aljuaid et al., 2019; Kocot, Luchowska-Kocot, Kiełczykowska, Musik, & Kurzepa, 2017; Kuehn, 2020).

Glucose metabolism is crucial for neurons, where anaerobic glycolysis and the pentose phosphate pathway (PPP) are used to metabolize glucose. PPP can generate antioxidants such as NADPH and glutathione which are the reducing equivalents regenerating intracellular antioxidant capacity (Yan, Hu, Wang, Wang, & Zhang, 2020). In AD and PD, the glucose

metabolism in neuron cells is dramatically changed. Currently, GLUT1 and GLUT3 are proposed to mediate bulk glucose flux into neurons. However, *SLC2A14* is also expressed in the neurons in the brain (**Figure 5.2**) and reduced GLUT14 functioning might therefore contribute to neurological disorders. Disease mechanism are more likely related to sugar rather than dehydroascorbic acid transport, since the main Vitamin C transporter in the brain is SLC23A2/SVCT2 (Sotiriou et al., 2002).

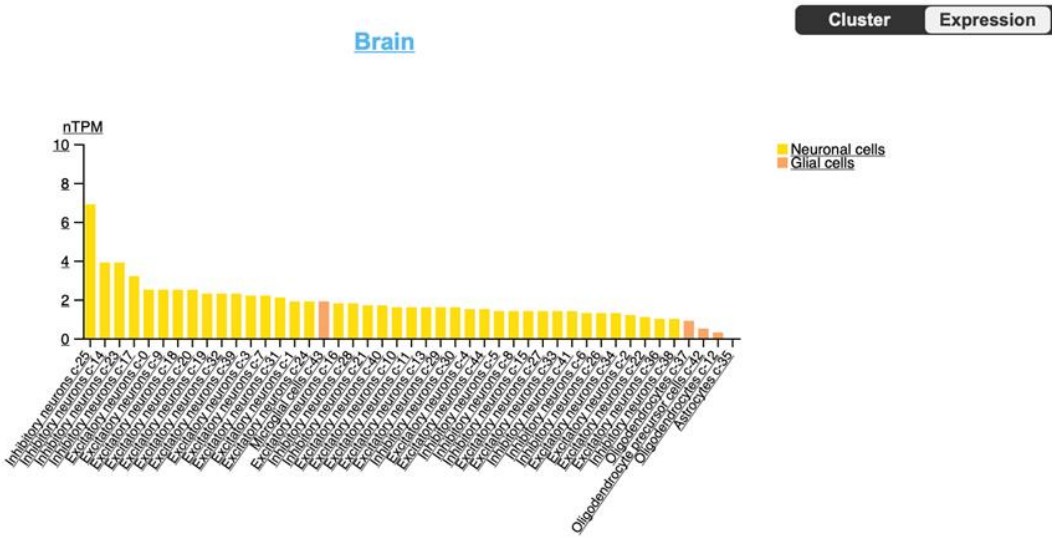


Figure 5.2: *SLC2A14* expression levels in neurons in the brain in normalised transcript per million (nTPM). The graph is obtained from Human Protein Atlas: A single-cell type transcriptomics map of human tissues (Karlsson et al. 2021). Copyright 2022 by Human Protein Atlas, reprinted with permission.

Although *SLC23A2*/SVCT2 is the main ascorbic acid transporter in the brain, *SLC23A1* is expressed in very modest amounts in neurons (**Figure 5.3**). Since ascorbic acid is a co-enzyme in the neurotransmitter synthesis, for instance, norepinephrine, the impact function of SLC23A1 is likelihood play a role in neurological disorders. Alternatively, neuronal antioxidant capacity might be reduced and contribute to the disease development. However, an intervention with dehydroascorbic acid to boost the neuron's Vitamin C status will be less straight forward than interventions in the enterocyte, since oral or even intravenous dehydroascorbic acid is unlikely to reach the neurons in the brain (Sotiriou et al., 2002).

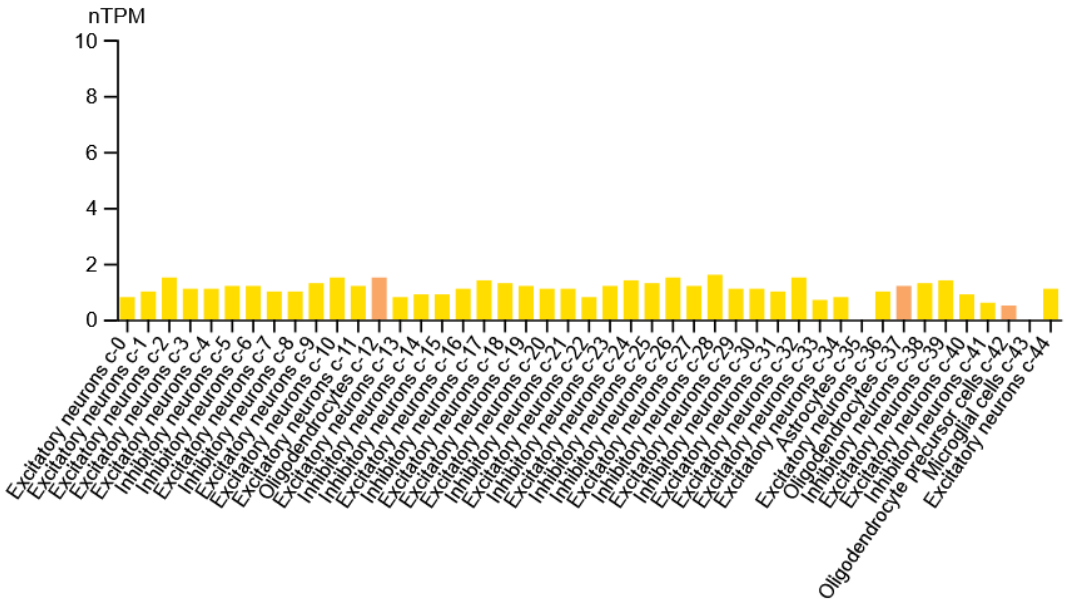


Figure 5.3: *SLC23A1* expression levels in neurons in the brain in normalised transcript per million (nTPM). The graph is obtained from Human Protein Atlas: A single-cell type transcriptomics map of human tissues (Karlsson et al. 2021). Copyright 2022 by Human Protein Atlas, reprinted with permission.

Strengths and Limitations

This study had several strengths. First, this thesis chose *Xenopus lavies* oocyte system as a mature and reliable model to study the protein function (Dawid & Sargent, 1988). *Xenopus* oocytes were recognized as a "live test tube" in the study of heterologous protein expression (Fortriede et al., 2020). Therefore, the function of orphan GLUT14 isoforms and two truncated *SLC23A1* proteins were well characterized. Second, one female *Xenopus* could provide thousands of oocytes, it provided a sufficient sample for all experiments. Therefore, each experimental group had equal or more than 15 oocytes to ensure the experimental and statistical accuracy. Last but not least, by tracking the H^3 and C^{14} labels from the radioactive incubation solution, the accurate uptake data for each oocyte could be calculated without the influence of the background.

This thesis also had some limitations. *Xenopus* oocytes system was great to study the function of an orphan membrane transporter, but it could not predict the cellular location of this orphan membrane transporter (Pike, Matthes, McSteen, & Gassmann, 2019). As oocytes were a simple living organisms, the cell signalling might also be different. Therefore, by using the

oocytes, I could only conclude the protein was functional but could not predict if the two truncated proteins for *SLC23A1* could express on the apical site of the membrane as it would normally. Another limitation of this thesis was all the experimental data were obtained *in vitro* and there is a lack of human data. It would be better if I could link the current findings and the available cohort study together and investigate the associations with the diseases.

Future Directions

One of the goals to investigate the orphan transporter in humans to have a better understanding of nutrient metabolism. Then, it will further help us to study the disease mechanism or treatment. The SGLT2 inhibitor is a perfect example of how scientists can transfer the *in vitro* study into medicine that treats type 2 diabetes (Hsia, Grove, & Cefaiu, 2017).

The data from this thesis will inform future clinical trial designs. First, common disease associations for *SLC23A1* and *SLC2A14* are unlikely caused by common disease mechanisms, and need to be investigated separately. Second, variations in the SLC23A1 N-terminus do not need to be considered due to their predicted low penetrance. Third, the impact of GLUT14 substrates in disease interventions can now be tested.

References

Alhashim, A. (2022). *The homo-sapiens solute carrier family 2 member 14 (SLC2A14) – further insight into the genomic organization, protein isoforms, substrates and kinetics* (p. 56). p. 56.

Aljuaid, M., Booth, S., Hobson, D. E., Borys, A., Williams, K., Katako, A., … Ko, J. H. (2019). Blood Flow and Glucose Metabolism Dissociation in the Putamen Is Predictive of Levodopa Induced Dyskinesia in Parkinson's Disease Patients. *Frontiers in Neurology*, *10*(November), 1–10. https://doi.org/10.3389/fneur.2019.01217

Chambers, L. W., Bancej, C., & McDowell, I. (2016). Prevalence and Monetary Costs of Dementia in Canada. *Alzheimer Society Canada*, 249–312. https://doi.org/10.1201/9780203025314

Chen, L., Sun, X., Wang, Z., Lu, Y., Chen, M., He, Y., … Zheng, L. (2021). The impact of plasma vitamin C levels on the risk of cardiovascular diseases and Alzheimer's disease: A Mendelian randomization study. *Clinical Nutrition*, *40*(10), 5327–5334. https://doi.org/10.1016/j.clnu.2021.08.020

Dawid, I. B., & Sargent, T. D. (1988). Xenopus laevis in developmental and molecular biology. *Science*, *240*(4858), 1443–1448. https://doi.org/10.1126/science.3287620

Duell, E. J., Lujan-Barroso, L., Llivina, C., Muñoz, X., Jenab, M., Boutron-Ruault, M. C., … González, C. A. (2013). Vitamin C transporter gene (SLC23A1 and SLC23A2) polymorphisms, plasma vitamin C levels, and gastric cancer risk in the EPIC cohort. *Genes and Nutrition*, *8*(6), 549–560. https://doi.org/10.1007/s12263-013-0346-6

Fortriede, J. D., Pells, T. J., Chu, S., Chaturvedi, P., Wang, D. Z., Fisher, M. E., … Vize, P. D. (2020). Xenbase: Deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. *Nucleic Acids Research*, *48*(D1), D776–D782. https://doi.org/10.1093/nar/gkz933

Hsia, D. S., Grove, O., & Cefaiu, W. T. (2017). An Update on SGLT2 Inhibitors for the Treatment of Diabetes Mellitus. *Physiology & Behavior*, *176*(10), 73–79. https://doi.org/10.1097/MED.0000000000000311.An

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., … Sainz, J. (2015). Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls.

Neurobiology of Aging, *36*(2), 1105–1109. https://doi.org/10.1016/j.neurobiolaging.2014.10.039

Jarmakiewicz-Czaja, S., Piątek, D., & Filip, R. (2020). The Influence of Nutrients on Inflammatory Bowel Diseases. *Journal of Nutrition and Metabolism*, *2020*. https://doi.org/10.1155/2020/2894169

Kocot, J., Luchowska-Kocot, D., Kiełczykowska, M., Musik, I., & Kurzepa, J. (2017). Does vitamin c influence neurodegenerative diseases and psychiatric disorders? *Nutrients*, *9*(7). https://doi.org/10.3390/nu9070659

Kuehn, B. M. (2020). In Alzheimer Research, Glucose Metabolism Moves to Center Stage. *JAMA - Journal of the American Medical Association*, *323*(4), 297–299. https://doi.org/10.1001/jama.2019.20939

Liu, H., Zhang, Y., Zhang, H., Wang, L., Wang, T., Han, Z., … Liu, G. (2021). Effect of plasma vitamin C levels on Parkinson's disease and age at onset: a Mendelian randomization study. *Journal of Translational Medicine*, *19*(1), 1–9. https://doi.org/10.1186/s12967-021-02892-5

Mills, E. L., Kelly, B., Logan, A., Costa, A. S. H., Varma, M., Bryant, C. E., … O'Neill, L. A. (2016). Succinate Dehydrogenase Supports Metabolic Repurposing of Mitochondria to Drive Inflammatory Macrophages. *Cell*, *167*(2), 457-470.e13. https://doi.org/10.1016/j.cell.2016.08.064

Montrose, D. C., Nishiguchi, R., Basu, S., Staab, H. A., Zhou, X. K., Wang, H., … Dannenberg, A. J. (2021). Dietary Fructose Alters the Composition, Localization, and Metabolism of Gut Microbiota in Association With Worsening Colitis. *Cellular and Molecular Gastroenterology and Hepatology*, *11*(2), 525–550. https://doi.org/10.1016/J.JCMGH.2020.09.008

Olsen, L. (2009). The Facts About Inflammatory bowel diseases. *Pediatric Surgery Digest*, 503–510. https://doi.org/10.1007/978-3-540-34033-1_25

Padayatty, S. J., Katz, A., Wang, Y., Eck, P., Kwon, O., Lee, J. H., … Dutta, S. K. (2003). Vitamin C as an Antioxidant: Evaluation of Its Role in Disease Prevention. *Journal of the American College of Nutrition*, *22*(1), 18–35. https://doi.org/10.1080/07315724.2003.10719272

Parkinsonism in Canada - Canada.ca. (2019, December 9). Retrieved March 5, 2022, from https://www.canada.ca/en/public-health/services/publications/diseases-conditions/parkinsonism.html

Pike, S., Matthes, M. S., McSteen, P., & Gassmann, W. (2019). Using Xenopus laevis Oocytes to Functionally Characterize Plant Transporters. *Current Protocols in Plant Biology*, *4*(1), e20087. https://doi.org/10.1002/cppb.20087

Sartor, R. B., & Wu, G. D. (2017). Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches. *Gastroenterology*, *152*(2), 327-339.e4. https://doi.org/10.1053/j.gastro.2016.10.012

Shaghaghi, Mandana Amir;, Zhouyao, H., Tu, H., El-Gabalawy, H., Crow, G. H., Levine, M., … Eck, P. (2017). The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease. *American Journal of Clinical Nutrition*, *106*(6), 1508–1513. https://doi.org/10.3945/ajcn.116.147603

Shaghaghi, Mandana Amir, Bernstein, C. N., León, A. S., El-Gabalawy, H., & Eck, P. (2014). Polymorphisms in the sodium-dependent ascorbate transporter gene SLC23A1 are associated with susceptibility to Crohn disease. *American Journal of Clinical Nutrition*, *99*(2), 378–383. https://doi.org/10.3945/ajcn.113.068015

Sotiriou, S., Gispert, S., Cheng, J., Wang, Y., Chen, A., Hoogstraten-Miller, S., … Nussbaum, R. L. (2002). Ascorbic-acid transporter Slc23a1 is essential for vitamin C transport into the brain and for perinatal survival. *Nature Medicine*, *8*(5), 514–517. https://doi.org/10.1038/0502-514

Wang, W., Yu, J. T., Zhang, W., Cui, W. Z., Wu, Z. C., Zhang, Q., & Tan, L. (2012). Genetic association of SLC2A14 polymorphism with Alzheimer's disease in a Han Chinese population. *Journal of Molecular Neuroscience*, *47*(3), 481–484. https://doi.org/10.1007/s12031-012-9748-y

Washko, P. W., Wang, Y., & Levine, M. (1993). Ascorbic acid recycling in human neutrophils. *Journal of Biological Chemistry*, *268*(21), 15531–15535. https://doi.org/10.1016/s0021-9258(18)82289-x

Yan, X., Hu, Y., Wang, B., Wang, S., & Zhang, X. (2020). Metabolic Dysregulation Contributes to the Progression of Alzheimer's Disease. *Frontiers in Neuroscience*, *14*(November), 1–20. https://doi.org/10.3389/fnins.2020.530219

Zhang, W., Cheng, H., Gui, Y., Zhan, Q., Li, S., Qiao, W., & Tong, A. (2021). Mannose Treatment: A Promising Novel Strategy to Suppress Inflammation. *Frontiers in Immunology*, *12*, 3954. https://doi.org/10.3389/FIMMU.2021.756920/BIBTEX

CHAPTER 6: Overall Conclusion

The presented data advance the knowledge on two genes/proteins encoding Vitamin C transporters, although of different pathways and at different research stages.

Three viable in frame translation initiation codons were identified for the *SLC23A1* gene. This is a rare example where the functionality of such cryptic sites is investigated and confirmed. It also implies that the N-terminal region is redundant and variations located in the region are unlikely to have a causal functions for disease development.

GLUT14 protein isoforms A, B, D, and E are multi-specific hexose transporters, while isoform C remains nonfunctional.

The created knowledge will inform on the design of future genomics, genetics and clinical studies.