The SLC2A14 gene: genomic locus, tissue expression, splice variants, and subcellular localization of the protein

Journal:	<i>Biochemistry and Cell Biology</i>
Manuscript ID	bcb-2015-0089.R1
Manuscript Type:	Article
Date Submitted by the Author:	19-Aug-2015
Complete List of Authors:	Amir Shaghaghi, Mandana; University of Manitoba, Human Nutritional Sciences Murphy, Brent; University of Manitoba, Plant Sciences Eck, Peter; University of Manitoba, Human Nutritional Sciences
Keyword:	SLC2A14, GLUT14

The *SLC2A14* gene: genomic locus, tissue expression, splice variants, and subcellular localization of the protein

Mandana Amir Shaghaghi¹, Brent Murphy², Peter Eck¹

Affiliations: ¹Department of Human Nutritional Sciences, University of Manitoba, Winnipeg, Canada, R3T 2N2, ²Department of Plant Science, University of Manitoba, Winnipeg, Canada.

Author's family name: Amir Shaghaghi. Murphy. Eck.

Correspondence: Peter Eck, Department of Human Nutritional Sciences, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2.

E-mail: peter.eck@umanitoba.ca

Phone: +1-204- 474-1506; Fax: +1-204- 272-1506

Funding source: This research was supported Peter Eck's Canada Research Chair.

Abbreviated title: The SLC2A14 gene

Abstract

The *SLC2A14* gene encodes for GLUT14, an orphan member of the facilitated membrane glucose transporter family, which was originally described to be exclusively expressed in human testis. However, genetic variations in *SLC2A14* are associated to chronic diseases such as Alzheimer's disease and Inflammatory Bowel Disease, which cannot be explained by a strictly testicular expression. Therefore we analyzed available information on the *SLC2A14* gene to update the knowledge on the locus and its encoded products. This report represents an expanded *SLC2A14* gene locus and a more diverse tissue expression, concurring with existing evidence of disease associations. The exon utilization is tissue specific, with major expression in testis. When the two major testicular protein isoforms were expressed in mammalian cells, they located to the plasmalemma membrane, providing early evidence that GLUT14 could have a function as a membrane transporter.

Keywords: GLUT14, SLC2A14

Introduction

The *SLC2A14* gene encodes for GLUT14, which was describes as a specific to human and an orphan member of the SLC2/GLUT family of facilitated membrane hexose transporter (Wu and Freeze 2002). In 2002 Wu and Freeze described the *SLC2A14* gene structure, showing high homology on the transcript level to *SLC2A3* (GLUT3) (Wu and Freeze 2002). GLUT 3 has high affinity and high transport capacity for glucose and it also mediates cellular uptake of dehydroascorbic acid, the oxidized form of Vitamin C. Based on the high homology between GLUT14 with GLUT3, a function as facilitate glucose transporter is deduced for GLUT14 (Cura and Carruthers 2012), however, remains to be experimentally proven.

GLUT 14 was first described to be exclusively expressed in testis (Wu and Freeze 2002), however, further evidence indicates expression in other tissues and cell lines, such as the ovarian cancer cell line A2780, where expression is down-regulated in clonal cell lines resistant to the anticancer agents paclitaxel and cisplatin (Januchowski, Zawierucha et al. 2013, Januchowski, Zawierucha et al. 2014). Abundance of the GLUT 14 transcript is also reported in the acute lymphoblastic leukemia cell lines NALM-6 and Jurkat (Taylor, Pena-Hernandez et al. 2007), as well as in peripheral blood of Parkinson's disease patients, where it was upregulated compared to healthy controls (Infante, Prieto et al. 2015). Taken together, this indicates that GLUT14 is expressed in tissues other than testis and could play a role in the development of common disease such as leukemia and Parkinson's disease.

Variations in the *SLC2A14* locus are implicated in common diseases of the central nervous system (Shulman, Chipendo et al. 2011, Wang, Yu et al. 2012), lymphatic cancer (Taylor, Pena-

Hernandez et al. 2007), rheumatoid arthritis (Veal, Reekie et al. 2014), and intraocular pressure in primary open-angle glaucoma (Nag, Venturini et al. 2013). Specifically, the single nucleotide polymorphism (SNP) rs10845990 is found to be associated with Alzheimer's disease in a Caucasian cohort (Shulman, Chipendo et al. 2011), and carriers of the rs10845990-G allele showed a 41% higher risk for late-onset Alzheimer's disease in a Chinese cohort (Wang, Yu et al. 2012). Moreover, a deletion of 129 kilobytes in the chromosome 12p13.31 region, which includes the *SLC2A14* gene, confers a protective affect against rheumatoid arthritis across two Caucasian cohorts (Veal, Reekie et al. 2014). A copy number variation in the 12p13.3 locus, overlapping the genes *SLC2A14* and *SLC2A3*, associated with intraocular pressure in glaucoma (Nag, Venturini et al. 2013).

Despite emerging evidence that genetic variations in *SLC2A14* are associated with common and complex diseases, major gaps in our knowledge of the genomic locus, its transcripts, and encoded protein isoforms exist. In this paper we revisit the *SLC2A14* gene to present novel information on the exon structure and utilization, alternative splicing, tissue expression, and intracellular localization of the two canonical protein isoforms.

Material and Methods

Bioinformatics

Sequences. The National Center for Biotechnology Information (NCBI) database was used to obtain the genomic sequence (GRCh37.p13 (GCF_000001405.25); NC_000012.11 (7965110.8043792, complement) and the reference and non-reference RNAs for GLUT14. In addition, the Basic Local Alignment Search Tool (BLAST) was used to screen for additional

human non-reference RNAs as well as expressed sequence tags (ESTs). Based on the reference RNA sequence, 17 usable non-reference RNAs and 318 usable ESTs sequences were identified. To increase the accuracy of the search, the reference RNA sequence was broken into shorter sequences (exons 1-5, 6-8, 9-11, and 12, as called in earlier publications(Wu and Freeze 2002)) and all search results with <90% identity were included in the analyses. The DNA Sequence Analysis Software - Sequencher 4.8 (Gene Codes Corporation; Ann Arbor, USA) was used to align and hand curate the remaining sequences, allowing for large gaps, 20 minimum overlap, and minimum match of 80%. Finally, the resultants were compared to the consensus sequences and the variations analyzed. Exons were deduced from the alignments of all transcripts (RNAs and ESTs) in order of appearance on the RNA sense orientation in *SLC2A14* locus.

Tissue Expression

Tissue expression was determined using mixed models of available bioinformatics information, quantitative PCR and regular PCR.

The abundance of the human mRNA in specific tissues was deduced from RNA sequencing data, determined by the Reference Transcriptome Resource (<http://nhprtr.org/>) (Pipes, Li et al. 2013). The abundance of reads was assessed by the magic index against the intergenic background and expression values were normalized as significant Fragments Per Kilobase Of Exon Per Million Fragments Mapped (sFPKM) (Pipes, Li et al. 2013).

Quantitative real time PCR was performed on marathon ready cDNA from human small intestine, testis, kidney, and liver (Clontech Laboratories Inc., Mountain View, CA; cat# 639326,

cat#639314, cat#639305, cat#639307, respectively), on a StepOnePlus Real-Time PCR System (Life Technologies, Carlsbad, CA), using a custom made PrimeTime® Mini qPCR Assay (MiniPrb6-FAM/ZEN/IBFQ) (Integrated DNA Technologies, Coralville, Iowa, USA), with the following specifications: sense primer GAAGAGAAATTGGAGAGGGAGTC, antisense primer GTGTGCGGCTTCATTCTTG, probe CTGGAATTGGTGGGTCTTGGTCTCA.

PCR was performed on Marathon Ready cDNA (Clontech) from liver (cat#639307), small intestine (cat#639326), testis (cat#639314) and placenta (cat#639311) using the following primers: sense GAGATGGACAACAGACAG, antisense GACATTGGTGGTGGTCTCCTTA. Amplicons were separated by DNA-gel electrophoresis on a 1.5% agarose gel in TAE buffer, containing ethidium bromide and imaged in a ChemiDoc gel imaging systems (Bio-Rad Laboratories, Mississauga, Ontario, Canada).

Subcloning of SLC2A14/GLUT 14 isoforms

Subcloning was performed using PCR techniques with Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, Massachusetts), following the manufacturers guidelines. The two major testis specific transcripts (Wu and Freeze 2002), which encode for two protein isoforms with very high homology to GLUT3, were subcloned.

The shorter GLUT14 proteins isoforms open reading frame was amplified from cDNA clone 5297510 (plasmid: Open Biosystems catalogue # MHS1010-9204259; GENE BANK: BC060766). This major testis specific transcript skips exon eleven, creating a start codon in exon ten and encodes for the shorter N-terminus (MDNRQN) (Wu and Freeze 2002), which is after

these initiates six amino acids in frame with the longer isoform. The following primers were utilized: sense GAGATGGACAACAGACAG, antisense GACATTGGTGGTGGTCTCCTTA. This amplicon was C-terminally tagged with the red fluorescent protein mCherry (Clontech Laboratories) using a PCR based approach of overlapping primers: sense hybrid primer for GLUT14 and mCherry: TAAGGA-GACCACCACCAATGTC-ATGGCCAGCAAAGGAGAAGA. The mCherry was amplified from plasmid pcDNATM-DEST53 Vector (Life Technologies) using the following primers: sense ATGGCTAGCAAAGGAG-AAGA, antisense GACCACCACCAATGTCTAA. The GLUT 14 amplicon and the mCherry amplicon were ligated in a third round of PCR and tagged with the “Gateway attB1 and attB2” sequences to allow transfer of the amplicon into the Gateway[®] (Life Technologies) subcloning system. The final amplicon was subcloned into pDONRTM221 Vector and expression vector pcDNA3.2/V5-DEST (Life Technologies).

The longer GLUT14 protein isoforms open reading frame was amplified from testis marathon ready cDNA (catalogue # 639314, Clontech Laboratories, Mountain View, California) using the following primers: sense ATGGAGTTTCACAATGGTGG, antisense GACATTGGTGGTGGTCTCCTTA. The reference sequence RNA NM_153449.2, represents the longer version of the protein, utilizing a start codon in exon 11, encoding a protein in frame with the shorter isoform, but with a distinct N-terminus of 30 amino acids (Wu and Freeze 2002). This amplicon was C-terminally tagged with the green fluorescent protein tGFP (Clontech Laboratories) using a PCR based approach of overlapping primers: sense hybrid primer for GLUT14 and tGFP: TAAGGAGACCACCACCAATGTCATGGCCAGCAAAGGAGAAGA. The tGFP was amplified using the following primers: sense ATGGTGAGCAAGGGCGAGGA, antisense

TCACTTGT-ACAGCTCGTCCAT. The GLUT 14 amplicon and the tGFP amplicon were ligated in a third round of PCR and tagged to transfer them into the Gateway[®] (Life Technologies) subcloning system. The final amplicon was subcloned into pDONR[™]221 Vector and expression vector pcDNA3.2/V5-DEST (Life Technologies). The identity of the inserts was confirmed by sequencing. The following two plasmids are derived of this process: A) tGFP-SLC2A14-Long-pcDNA/V5-DEST; B) mCherry-SLC2A14-Short-pcDNA/V5-DEST.

Molecular imaging

Chinese Hamsters Ovary cells (CHO-K1) were purchased from American Type Culture Collection (ATCC, Manassas, Virginia, USA), and maintained on 24-well tissue culture plates (Thermo Fisher Scientific, BioLite) at 37°C in DMEM/F-12 HyClone media (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS). At 60-80% confluency the GLUT14 expression plasmids (mCherry-SLC2A14-Short pcDNA/V5-DEST or tGFP-SLC2A14-Long pcDNA/V5-DEST) were transfected into the cells using Effectene Transfection Reagent (Quiagen), following the manufacturer's protocols. After 24 to 72 hours, transfected cells were imaged at an inverted Axiovert 200M Fluorescence/Live cell Imaging Microscope (Carl Zeiss GMBH, Jena, Germany) using standard green and red filter sets. Images were collected with the Axiovision software package AxioVS40 v 4.5, and were processed using Zen Software (Carl Zeiss GMBH).

Results

Genomic Locus

The revised genomic organization of the *SLC2A14* gene shows a total of twenty exons covering 103,477 nucleotides from the first Transcriptional Start Site (TSS) to the termination of the longest transcript. *SLC2A14* is transcribed from the complement strand of chromosome 12p13.31 (**Figure 1**). It is in tandem with the paralogue *SLC2A3* gene, which is located 3232 nucleotides upstream but only has a size of one fourth of *SLC2A14* gene (**Figure 1**). The locus overlaps the LOC100130582, ROS20P29, and NANOGP1 genes, which are transcribed from the opposite strand (**Figure 1**). *SLC2A14* orthologues are only found in primates, whereas *SLC2A3* orthologues are found throughout the animal kingdom (Supplement Figure 1).

Tissue Expression

SLC2A14 is highly expressed in testis, but according to publicly available sources, significant levels of mRNA are also found in the colon, lung, ovaries and blood cells, brain, skeletal muscle, heart, kidney, and liver (**Figure 2A**). Real Time PCR concurs with these data, showing strongest expression in testicular tissue, and lesser but significant expression in small intestine, liver and kidney (**Figure 2B**). Presence in placenta is demonstrated by the fact that GLUT14 could be subcloned out of placental, liver and small intestinal cDNA (Supplement Figure 2).

Exon Utilization and alternative Splicing

Several differentially spliced transcripts show tissue specific exon utilization, specifically in the 5' portion of *SLC2A14*.

The first four exons are exclusively utilized in brain hippocampal neuronal cells, where two distinct transcriptional start sites (TSS) in exon one and exon four exist (**Figure 1**). Exon five is an alternative transcriptional start site in testicular tissue (**Figure1**) and is alternatively spliced at its 3' end (Supplement Figure 3). Exons five to nine are utilized in testicular tissue, but not in neuronal (**Figure 1**).

One third of the neuronal transcripts utilize exon ten, while all testis transcripts contain it. In addition, exon ten is a major alternative transcriptional start site for testicular transcripts. Exon eleven stands out in regard to splicing patterns, through showing tissue specific utilization as well as an intra-tissue splicing pattern. All neuronal transcripts skip exon eleven. All testis specific transcripts originating from the transcriptional start site in exon ten skip exon eleven, while it is contained in 47% of all transcripts originating from the earlier transcriptional start site in exon five (Supplement Figure 4).

Exons twelve through twenty are utilized in all tissues. Exon twelve also represents the last identifiable transcriptional start site (**Figure 1**), where neuronal as well as testicular transcripts originate. Exon 13 is skipped in about 3.5% of transcripts (Supplement Figure 5). The first 198 nucleotides of the last exon (exon 20) are skipped in 7% of the transcripts (Supplement Figure 6).

Transcripts Termination Sites

The last exon contains a maximum of 2014 nucleotides, where 3% of the transcripts support this longest extension. At least two earlier transcription termination sites are apparent, one resulting

in a 1003 nucleotide exon and the other in a 733 nucleotide exon, the latter is equivalent to the termination site in the paralogue *SLC2A3* gene. All of the alternative transcript termination sites do not impact on the open reading frame.

Transcripts

Three major transcripts encoding for two protein isoforms exist in testicular tissue. The shortest transcript originates in exon ten and skips exon eleven, encoding for a shorter protein of 497 amino acids, originally named by Wu and Freeze GLUT14S.

Two major transcripts originate in exon five, skip exons six to nine, and utilize exon ten. 53% of these transcripts utilize the start codon in exon ten by skipping exon eleven, encoding the GLUT14S isoform. In contrast, 47% incorporate exon 11, which disrupts the GLUT14S open reading frame. Instead, a translation start codon located at the immediate start of exon eleven creates a novel open reading frame encoding for a longer isoform of 520 amino acid, earlier named GLUT14L (Wu and Freeze 2002). The two protein isoform differ in their N-terminus; GLUT14L has twenty-nine distinct amino acids, where GLUT14S has six, consistent with the isoforms described earlier (Wu and Freeze 2002).

Subcellular localization

Upon expression of both fluorescently labeled major GLUT14 isoforms in CHO-K1 cells a subcellular localization to the membrane can be observed. There is also a perinuclear intracellular localization, consistent with the sites of synthesis and processing on the Endoplasmic Reticulum and the Golgi Apparatus (**Figure 3**).

Discussion

In expansion of prior work (Wu and Freeze 2002) we describe ten additional exons originating from the *SLC2A14* gene locus; therefore, according to the current analysis it contains twenty exons. We named the exon in the order of appearance, consistent with latest recommendations (hgsv.org, <http://www.hgvs.org/mutnomen/refseq.html#exonnumber>). Only one prior publication (Wu and Freeze 2002) numbered *SLC2A14* exons; however, the locus was only partially described. Wu and Freezes (Wu and Freeze 2002) first exon (exon 1a in the Wu and Freeze publication) is represented by our tenth exon, where their exon 1b is our eleventh exon, and all following exons correspond to their description of their gene structure. Therefore Wu and Freeze (Wu and Freeze 2002) did not describe the first five exons for testicular transcripts which are located 5' of exon ten. They also failed to determine the neuronal splice variants and therefore did not describe the neuron specific exons one to four.

In expansion of the earlier record of gene expression (Wu and Freeze 2002), we find evidence for *SLC2A14* expression beyond testis. The expression in neuronal cell types could explain why genetic variations in *SLC2A14* are associated with Alzheimer's disease (Shulman, Chipendo et al. 2011, Wang, Yu et al. 2012) and why expression patterns vary in Parkinson's disease patients (Infante, Prieto et al. 2015). The presented data therefore contribute to the understanding of future disease associations relating to *SLC2A14* as a putative expression quantitative trait locus (eQTL).

The current analysis revealed distinct tissue specific exon utilization, which could be determined for neuronal tissues and testis. The existence of five distinct transcriptional start sites with distinct promoter regions might be the hallmark of an ongoing evolution in the *SLC2A14* gene, which was only recently duplicated during primate evolution. The specific functional differences between the protein isoform encoded by these splice variants remain to be determined. However, the major testicular protein isoforms are likely to be solute carriers on the plasmalemma membrane, similar to the highly conserved paralogue GLUT3 (Wu and Freeze 2002).

The existence of three transcriptional termination sites could also indicate an ongoing evolutionary process, which might distinguish GLUT14 from its paralogue GLUT3. Alternative termination in the 3' untranslated region of genes affects the scaffolding of the proteins, resulting in differential intracellular targeting (Berkovits and Mayr 2015). Although we have proven plasmalemma localization of the protein, we had used plasmids carrying the open reading frame only. In future studies it needs to be investigated how the 3' untranslated regions of the *SLC2A14* transcript determine the intracellular location of the GLUT14 isoforms or how these distinguish it from GLUT3.

In conclusion, we update the knowledge about the *SLC2A14* genomic locus, identifying ten additional exons, expression beyond testicular tissue, and tissue specific exon utilization. The major protein isoforms locate to the plasmalemma membrane, indicating a function as solute carrier.

References

- Berkovits, B. D. and C. Mayr 2015. "Alternative 3[prime] UTRs act as scaffolds to regulate membrane protein localization." Nature **advance online publication**.
- Cura, A. J. and A. Carruthers 2012. "Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis." Comprehensive Physiology **2**(2): 863-914.
- Infante, J., C. Prieto, M. Sierra, P. Sanchez-Juan, I. Gonzalez-Aramburu, C. Sanchez-Quintana, J. Berciano, O. Combarros and J. Sainz 2015. "Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls." Neurobiol Aging **36**(2): 1105-1109.
- Januchowski, R., P. Zawierucha, M. Andrzejewska, M. Ruciński and M. Zabel 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.
- Januchowski, R., P. Zawierucha, M. Ruciński, M. Andrzejewska, K. Wojtowicz, M. Nowicki and M. Zabel 2014. "Drug transporter expression profiling in chemoresistant variants of the A2780 ovarian cancer cell line." Biomedicine and Pharmacotherapy **68**(4): 447-453.
- Nag, A., C. Venturini, P. G. Hysi, M. Arno, E. Aldecoa-Otalora Astarloa, S. Macgregor, A. W. Hewitt, T. L. Young, P. Mitchell, A. C. Viswanathan, D. A. Mackey and C. J. Hammond 2013. "Copy number variation at chromosome 5q21.2 is associated with intraocular pressure." Investigative ophthalmology & visual science **54**(5): 3607-3612.
- Pipes, L., S. Li, M. Bozinoski, R. Palermo, X. Peng, P. Blood, S. Kelly, J. M. Weiss, J. Thierry-Mieg, D. Thierry-Mieg, P. Zumbo, R. Chen, G. P. Schroth, C. E. Mason and M. G. Katze 2013. "The non-human primate reference transcriptome resource (NHPRT) for comparative functional genomics." Nucleic Acids Res **41**(Database issue): D906-914.
- Shulman, J. M., P. Chipendo, L. B. Chibnik, C. Aubin, D. Tran, B. T. Keenan, P. L. Kramer, J. A. Schneider, D. A. Bennett, M. B. Feany and P. L. De Jager 2011. "Functional Screening of

Alzheimer Pathology Genome-wide Association Signals in Drosophila." The American Journal of Human Genetics **88**(2): 232-238.

Taylor, K. H., K. E. Pena-Hernandez, J. W. Davis, G. L. Arthur, D. J. Duff, H. Shi, F. B. Rahmatpanah, O. Sjahputera and C. W. Caldwell 2007. "Large-Scale CpG Methylation Analysis Identifies Novel Candidate Genes and Reveals Methylation Hotspots in Acute Lymphoblastic Leukemia." Cancer Research **67**(6): 2617-2625.

Veal, C. D., K. E. Reekie, J. C. Lorentzen, P. K. Gregersen, L. Padyukov and A. J. Brookes 2014. "A 129-kb Deletion on Chromosome 12 Confers Substantial Protection Against Rheumatoid Arthritis, Implicating the Gene SLC2A3." Human Mutation **35**(2): 248-256.

Wang, W., J. T. Yu, W. Zhang, W. Z. Cui, Z. C. Wu, Q. Zhang and L. Tan 2012. "Genetic Association of SLC2A14 Polymorphism with Alzheimer's Disease in a Han Chinese Population." Journal of Molecular Neuroscience **47**(3): 481-484.

Wu, X. and H. H. Freeze 2002. "GLUT14, a duplicon of GLUT3, is specifically expressed in testis as alternative splice forms." Genomics **80**(6): 553-557.

Wu, X. H. and H. H. Freeze 2002. "GLUT14, a duplicon of GLUT3, is specifically expressed in testis as alternative splice forms." Genomics **80**(6): 553-557.

Figure legends:

Figure 1: Chromosomal location, overlapping and neighboring genes, overall genomic organization and tissue specific exon utilization of the *SLC2A14* gene.

Figure 2: Tissues expression of *SLC2A14* (*GLUT14*). Panel A demonstrates the tissues of significant RNA expression as determined from publicly available RNA-sequencing datasets and expressed in significant Fragments Per Kilobase Of Exon Per Million Fragments Mapped (sFPKM) (Pipes, Li et al. 2013) (see material and methods). Panel B establishes expression in testis, small intestine, liver and kidney through quantitative Real Time PCR.

Figure 3: The two major testicular GLUT14 isoforms are synthesized and processed on the Endoplasmic Reticulum/Golgi Apparatus and locate to the plasmalemma membrane in CHO-K1 cells. Upon expression, fluorescent signals for the GLUT14 long isoform (green fluorescence, panel A) and the GLUT14 short isoforms (red fluorescence, panel B) can be detected perinuclear at the site of synthesis and processing (endoplasmic reticulum/Golgi Apparatus, blue arrows) as well as on the plasmalemma membrane (red arrows). The fluorescent images are matched with the correspondent light images to show the cells alignment in a tight monolayer.

254x190mm (96 x 96 DPI)

254x190mm (96 x 96 DPI)

1

254x190mm (96 x 96 DPI)