

# Quantitative PCR Tissue Expression Profiling of the Human SGLT2 Gene and Related Family Members

Jian Chen · Sandy Williams · Samantha Ho · Howard Loraine · Deborah Hagan · Jean M. Whaley · John N. Feder

Received: September 29, 2010 / Published online: December 17, 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

## ABSTRACT

SGLT2 (for “Sodium GLucose coTransporter” protein 2) is the major protein responsible for glucose reabsorption in the kidney and its inhibition has been the focus of drug discovery efforts to treat type 2 diabetes. In order to better clarify the human tissue distribution of expression of SGLT2 and related members of this cotransporter class, we performed TaqMan™ (Applied Biosystems, Foster City, CA, USA) quantitative polymerase chain reaction (PCR) analysis of SGLT2 and other sodium/glucose transporter genes on RNAs from 72 normal tissues from three different individuals. We consistently observe that SGLT2 is highly kidney specific while SGLT5 is highly kidney abundant; SGLT1, sodium-dependent amino acid transporter

(SAAT1), and SGLT4 are highly abundant in small intestine and skeletal muscle; SGLT6 is expressed in the central nervous system; and sodium myo-inositol cotransporter is ubiquitously expressed across all human tissues.

**Keywords:** quantitative PCR; SGLT2; sodium-glucose cotransporter protein; tissue expression; type 2 diabetes

## INTRODUCTION

As a worldwide medical and economic problem type 2 diabetes is expanding internationally. The International Diabetes Federation estimates that in 2010 approximately 285 million individuals have type 2 diabetes across the world;<sup>1</sup> this number is expected to expand to 439 million individuals by 2030. Diabetes imposes a significant health and economic burden, and factoring in the additional costs of undiagnosed diabetes, prediabetes, and gestational diabetes, the total cost of diabetes in the US in 2007 amounted to \$218 billion.<sup>2</sup> Despite the availability of several oral and injectable therapies for type 2 diabetes, there remains significant unmet medical need in this disease, justifying the

---

Jian Chen · John N. Feder (✉)  
Applied Biotechnology and the Department of  
Applied Genomics, Bristol-Myers Squibb Company,  
311 Pennington-Rocky Hill Road, Pennington, NJ  
08534, USA. Email: john.feder@bms.com

Sandy Williams · Samantha Ho · Howard Loraine  
Asterand, Royston, Hertfordshire, UK

Deborah Hagan · Jean M. Whaley  
Department of Metabolic Diseases, Bristol-Myers Squibb  
R&D, Princeton, NJ, USA

search for more efficacious and safe treatments that can prevent disease progression and protect patients from microvascular and macrovascular complications. Among the types of therapies under development, inhibitors of SGLT2 (for “Sodium GLucose coTransporter” protein 2) represent a promising new class.<sup>3,4</sup>

One consideration for choosing a molecular target for the identification of a new treatment of a chronic disease such as type 2 diabetes is the spectrum of tissues in which the target of interest is expressed. A molecular target with

a ubiquitous pattern of expression could pose concerns related to the activities of agonists or antagonists to this target in a wide variety of tissues, whereas a molecular target expressed in a restricted number of tissues might suggest a more selective pharmacologic profile. We have evaluated the expression pattern of SGLT2 and related family members by quantitative reverse transcription real-time polymerase chain reaction (RT-PCR) methodology in order to better understand the potential impact of a selective SGLT2 inhibitor in vivo.

**Table 1.** SGLT (sodium glucose cotransporter protein) family members. **1A** and **1B**: The common names, system names, and human reference sequence numbers of the 12 SGLT family members are listed. The putative substrates of the transporters and the chromosomal locations of the transporter genes are given in the tables. **1C**: Percentage of protein sequence identity among SGLT family members as determined with VectorNTI AlignX software.

### 1A

Common Name	SGLT1	SGLT2	SAAT1	SMIT	SGLT4	SGLT5	SGLT6
System Name	SLC5A1	SLC5A2	SLC5A4	SLC5A3	SLC5A9	SLC5A10	SLC5A11
Reference Sequence	NM_000343	NM_003041	NM_014227	NM_006933	NM_001011547	NM_152351	NM_052944
Substrates	Glucose, Galactose	Glucose, Galactose	Glucose	Myo-inositol glucose	Glucose, Mannose	Unknown	Myo-inositol glucose
Chromosomal Location	22q12.3	16p12-p11	22q12.2-q12.3	21q22.12	1p33	10q23	16pter-p11

### 1B

Common Name	NIS	SMVT	CHT	SMCT1/AIT	SMCT2
System Name	SLC5A5	SLC5A6	SLC5A7	SLC5A8	SLC5A12
Reference Sequence	NM_000453	NM_021095	NM_021815	NM_145913	NM_178498
Substrates	Iodine	Multivitamins	Choline	Iodine, Lactate, Monocarboxylate, Short-chain fatty acids	Lactate
Chromosomal Location	19p13.2-p12	2p23	2q12	12q23.1	11p14.2

### 1C

	SGLT1	SAAT1	SGLT4	SGLT5	SGLT6	SMIT	NIS	SMCT2	SMVT	AIT	CHT
SGLT2	59	57	58	55	49	45	19	20	18	18	19
SGLT1		71	59	54	51	47	18	20	19	19	18
SAAT1			53	48	49	44	19	21	21	21	18
SGLT4				64	55	55	20	23	21	22	18
SGLT5					52	48	19	21	22	20	19
SGLT6						48	19	22	21	20	16
SMIT							19	22	18	20	17
NIS								45	38	47	16
SMCT2									42	53	14
SMVT										44	14
AIT											15

There are more than 200 SGLT family members, including 12 human orthologs.<sup>5</sup> Based on sequence homology, these 12 SGLT family members can be divided into two subfamilies, as shown in Table 1. SGLT1, SGLT2, sodium-dependent amino acid transporter (SAAT1; also known as SGLT3), sodium myo-inositol cotransporter (SMIT), SGLT4, SGLT5, and SGLT6 belong to one subfamily, sharing between 45% and 70% protein sequence identity amongst themselves. Most of the members of this subfamily transport or bind sugar molecules. The five other solute carrier family 5A (SLC5A) family members Na<sup>+</sup>/I<sup>-</sup> symporter (NIS), sodium-dependent multivitamin transporter (SMVT), choline transporter (CHT), apical iodide transporter/sodium monocarboxylate cotransporter 1 (AIT/SMCT1), and SMCT2 form another subfamily. They share between 40% and 50% protein sequence identity amongst themselves; members of this latter subfamily are involved in the cotransport of sodium with other physiologically important molecules such as iodide, ascorbate, biotin, pantothenate, lipoate, choline, and monocarboxylates such as lactate.<sup>5</sup> Since only 18% to 20% protein sequence identity exists between the two subfamilies, the focus of our studies was the sugar-binding class of SGLT cotransporters most closely related to SGLT2 (Table 1).

The first sugar-binding SGLT sequence to be cloned, by Wright and colleagues, was the high-affinity sodium-glucose cotransporter SGLT1, which was found to be expressed in the small intestinal mucosa<sup>6</sup> and associated with glucose and galactose transport at that site. SGLT1 was later found to be expressed in many tissues across the body,<sup>7</sup> and mutations in SGLT1 were associated with the human genetic syndrome glucose-galactose malabsorption.<sup>8</sup> SGLT2 was cloned subsequently, and was characterized as a low-affinity sodium-glucose cotransporter

expressed in the renal early proximal tubule.<sup>9,10</sup> SAAT1 was first cloned as a sodium-amino acid cotransporter<sup>11</sup> but was later found to have glucose cotransporter activity.<sup>12</sup> It was found in kidney, small intestine, and other tissues and is now suggested to be a sodium-dependent glucose sensor rather than a sodium-glucose cotransporter.<sup>13</sup> SMIT is an osmoregulatory sodium-inositol cotransporter found in many tissues including brain and cardiac myocytes.<sup>14,15</sup> SGLT4 is a low-affinity sodium-dependent transporter for mannose and fructose found in kidney and small intestine tissue.<sup>16</sup> SGLT5 was identified in the Mammalian Gene Collection (MGC) human cDNA sequencing project, by similarity to other SGLT family members.<sup>17</sup> SGLT6 (also known as KST1 or SMIT2) was identified as a novel sodium-glucose cotransporter<sup>18</sup> located within a genomic region associated with infantile convulsion and choreoathetosis as well as benign familial infantile convulsion diseases. It was found to be able to transport myo-inositol in a sodium-dependent manner.<sup>19</sup> Even though the functional activities of these SGLT family members have been described, there has not been a systematic study of the expression profiles of this family across the same broad set of normal human tissues.

There have been conflicting reports about the mRNA expression profile of SGLT2 in human tissues. It was initially reported to be expressed predominantly in the kidney using northern blot techniques;<sup>9,20</sup> however, Wright and colleagues have subsequently reported a broader pattern of tissue expression of SGLT2 beyond the kidney using RNase protection methods, although no detailed methods or data were presented.<sup>21</sup> In 2003, Zhou et al.<sup>22</sup> employed quantitative RT-PCR techniques to show that SGLT2 was ubiquitously expressed in most human tissues. In 2005, however, Tazawa et al.<sup>16</sup> used the

same methodological approach but reported contradictory findings: SGLT2 was primarily expressed in the kidney and to a smaller degree in the small intestine. It has also been reported that in mice, SGLT2 is specifically expressed in kidney proximal tubule.<sup>23</sup> Because SGLT2 inhibitor compounds are being developed by several pharmaceutical companies as antidiabetic agents which inhibit renal glucose reabsorption, it has become increasingly important to verify the tissue expression profile of SGLT2. To our knowledge, no antibodies shown to be specific for individual human SGLT family members have been developed successfully to enable expression profiling based on the protein level, although antibodies to one or more members of the protein family have been reported.<sup>6,24-29</sup> Therefore, we chose to verify the expression profile of seven SGLT family members in human tissues by quantitative PCR methods across a broad panel of human tissues.

## MATERIALS AND METHODS

### Reverse Transcription-Coupled Quantitative Real-Time PCR

AmpliTaq Gold™ DNA Polymerase and AmpErase® Uracil-N-Glycosylase were obtained from Applied Biosystems (Warrington, UK) and MMLV reverse transcriptase was purchased from Promega (Southampton, UK). TRIzol® and DNase I were obtained from Invitrogen (Paisley, UK). TaqMan™ probes and oligonucleotide primers were purchased from Sigma-Genosys (Haverhill, UK) and deoxynucleotide triphosphates were obtained from BioGene Limited (Kimbolton, UK). Purification of total RNA and quantitative RT-PCR was performed as previously described.<sup>30</sup> Briefly, total RNA was extracted from snap-frozen human tissues using TRIzol® according to the manufacturer's

instructions. Purified RNA samples were subject to a number of quality control criteria before being passed as suitable for quantitative RT-PCR.<sup>30</sup> RNA samples were treated with DNase I to remove any residual genomic DNA and were reverse transcribed using gene-specific primers.

Quantitative RT-PCR was performed with TaqMan primer/probe sets designed using PrimerExpress software (Applied Biosystems). To analyze the expression of the SGLT2 locus we used five primer/probe combinations: four new designs and one previously published.<sup>22</sup> These primer/probe sets were designed to survey the entire locus and to be specific with respect to other known expressed sequences, as well as specific for spliced transcripts versus nonspliced transcripts and genomic DNA. The primer/probe combinations were not matched for amplification efficiency, hence the absolute transcript number detected with each primer/probe set can be expected to vary. The sequences of the primer/probe sets utilized in these studies are available in the Appendix, Additional Table S1.

The amplification reaction was performed using AmpliTaq Gold™ polymerase in a standard PCR buffer containing cDNA prepared from 100 ng of total RNA. Uracil N-glycosylase was included in all reactions to prevent cross-contamination and amplification of previous PCR products. PCR reactions were done using an ABI 7900 Sequence Detection System and the thermocycling conditions were 50°C for 2 minutes followed by 95°C for 2 minutes, and then 40 cycles of 95°C for 15 seconds and 60°C for 30 seconds. Amplification of the target transcripts was performed as part of a multiplex reaction in which glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an internal control. Asterand's Global Standard curve was used to interpolate transcript copy numbers from the quantitative PCR C<sub>T</sub> values.<sup>30</sup>

## RESULTS

All tissues used in this study were obtained from ethically approved intermediaries. Tissue supply was governed by legal agreements and by stringent ethical review from local research ethics committees. In addition, and in all cases, the informed consent of the donor or the donor's next of kin was obtained for the use of the donated tissue for research. In all, 46% of the tissues were from female donors while 54% were from male donors. The age distributions of the tissue donors were: <20 years old: 5%; 20-39 years old: 17%; 40-60 years old: 32%; 60-80 years old: 31%; >80 years old: 14%. No ethnicity information relating to tissue donors was available.

The genes that were profiled by quantitative RT-PCR for their expression levels across a panel of 72 different human tissues are listed by both their common name and their systematic name in Table 1A. Also listed are the putative substrates for each of the transporters, as well as the chromosomal location of each transporter gene. The primer sequences used for PCR were designed to be general for all known isoforms and splice variants of each gene. The sequences of the

amplicons, the exon location of the amplicons, and their precise genomic coordinates are listed in Table 2.<sup>31</sup> Included in each PCR reaction was a primer/probe set for the GAPDH gene. This was done to control for the success of the first strand cDNA synthesis reaction and the eventual PCR. These data were not used for normalization purposes since GAPDH levels themselves vary considerably across the tissue type and individual donor.<sup>30</sup> The study included total RNA isolated from three different individuals for each tissue in a panel of 72 tissues. The 72 tissues in the study extend across all major human biological systems: cardiovascular, digestive, endocrine, male and female reproductive, hematopoietic and lymphatic, integumentary, musculoskeletal, nervous, respiratory, and urinary systems.<sup>30</sup> The individual RNA samples used in the study were prepared from normal tissues, although their donors may have had abnormal or diseased symptoms in other tissues or organs. The donors represent different genders and age groups, and every attempt was made to use the same samples for the analysis of all seven SGLT family members. A few exceptions did occur which bear no impact on our study conclusions.

**Table 2.** Genomic information of the quantitative polymerase chain reaction (PCR) amplicons of sodium glucose cotransporter protein (SGLT) family members.

	SGLT1	SGLT2	SAAT1	SMIT	SGLT4	SGLT5	SGLT6
<b>Amplicon Sequence</b>	CCCTGGTTTT GGTGGTTGTG CTGGGCTGGC TGTTTGTCCC CATCTATATTA AGGC TGGGGT GGTGA	GACACGGTAC AGACCTTCGT CATTCTGGGG GGCGCCTGCA TCCTCATGGG TTACGCCTTC CACGAGGTGG GCGGGTATTC GGGTCTCTTC GACAAATACC TGGGAG	TCCTCGCTGG TCGTGATATG GCCTGGTGGC CGATGGGCGC CTCTCTCTTTG CCAGTAACAT CGGC	ACCCAGCTT CAGTATGGTA CTGGTGTGCT GACCAAGTCA TCGTGCAGAG GGTCCTTGCA GCCAAAACA TTG	GCAGCAGCAC CCTGTTCAACC ATTGATGTGT GGCAGCGCTT CCGCAGGAAG TCAACAGAGC AGGAGC	GCTGATGATC GCAGTGATGC TGGCGGCGCT CATGTCTGTCG CTGACCTCCA TCTTCAACAG CAGCAGCACC CTC	CTCCCTCACC TCCATCTTAA CAGTGCCAGC ACCATCTTCA CCATGGACCT CTGGAATCAC CTCCGGCC
<b>Chromosomal Coordinates</b>	chr22:32,463,95 5-32,464,489	chr16:31,498,66 7-31,498,901	chr22:32,637,33 5-32,660,654	chr21:35,431,83 8-35,504,837	chr1:48,671,803 -48,731,202	chr17:18,808,90 8-19,027,907	chr16:24,843,90 6-24,937,055
<b>Exon Location of Amplicon</b>	Exon 4-5	Exon 6-7	Exon 2-3	Exon 2	Exon 11	Exon 11	Exon 12

The chromosomal coordinates and exon locations of the amplicons are determined by alignment of the amplicon sequences against human genome GRCh37 Assembly (hg19) on the UCSC Genome Browser.<sup>31</sup>

Tissue expression profiling of the seven SGLT family members reveals distinct patterns of tissue expression, summarized below.

### The Kidney-Specific and Kidney-Abundant SGLT Family Members

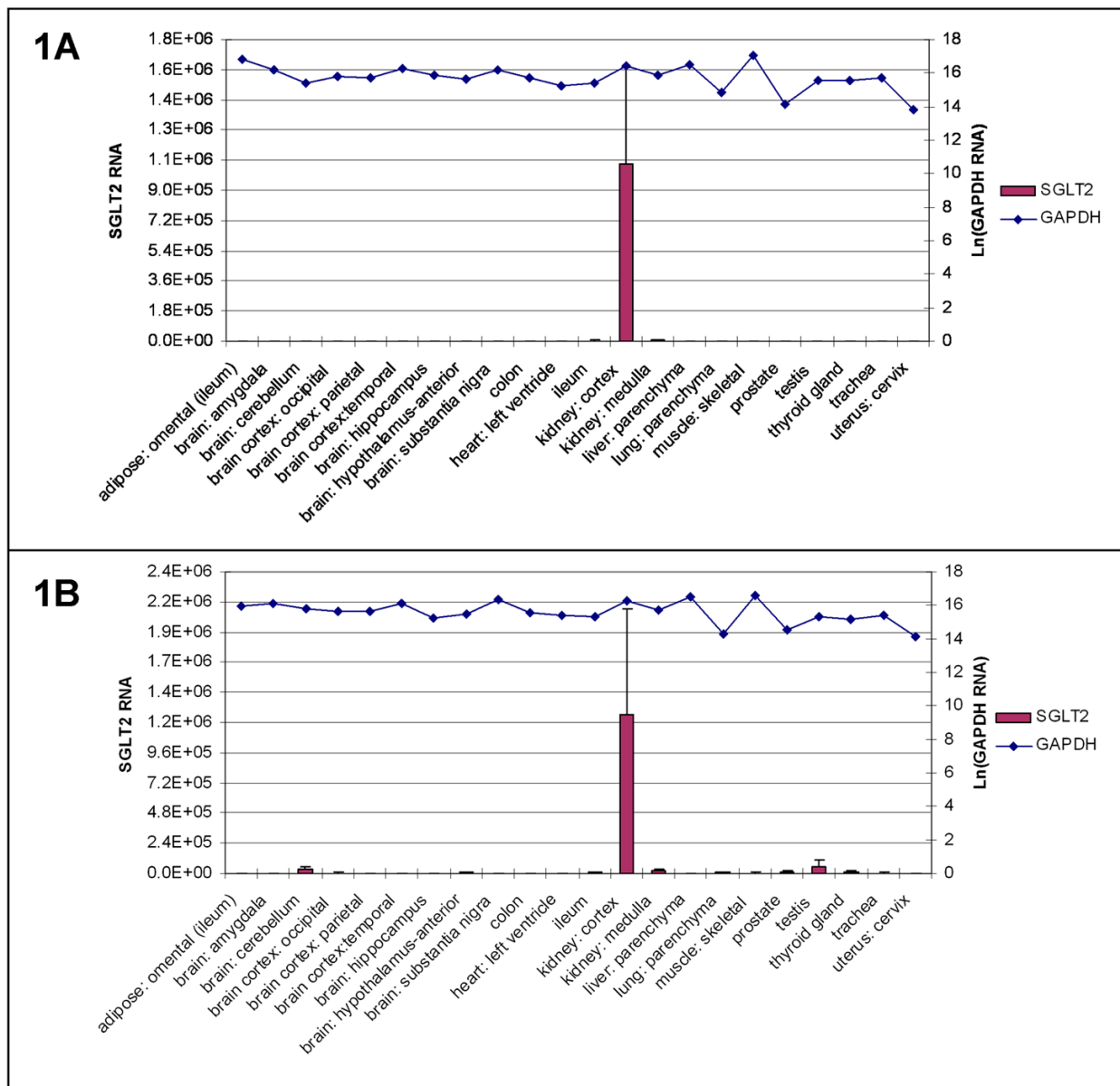
Figure 1A presents the results evaluating the expression of SGLT2 from a subset of the 72 tissues profiled (data from all the tissues are available in the Appendix, Additional Figure S1) with a primer/probe set (SGLT2-e6,7) designed to span the boundaries of exons 6 and 7 of SGLT2, as noted in Table 2). In all figures presented, the Y-axis represents the number of transcripts per  $\mu\text{g}$  of RNA. The tissue with the highest level of expression was the kidney cortex where the expression was approximately 300-fold higher than the tissue with the next highest level of expression, the kidney medulla. Although small numbers of putative transcripts could be observed in some of the other tissues, no evidence of SGLT2 expression was observed in the majority of other tissues, including the 20 brain subregion RNA samples tested (eight of these are shown in Figure 1A). These data are consistent with several other reports evaluating the pattern of expression of SGLT2 in human tissues.<sup>9,10,16</sup> The GAPDH data, plotted in  $\log_2$  format on the right-hand y-axis of the panel of Figure 1A, indicates that all tissue RNA samples had successful first strand synthesis and PCR reactions. The error bar associated with the tissue data indicates the variation observed in expression measurements using individual RNA samples from three individual donors and not that obtained with technical replicates. Our study methodology is thus different from other studies reported in the literature where pooled samples from commercial vendors were utilized.<sup>22</sup>

Similar data was obtained using three additional primer/probe combinations spanning

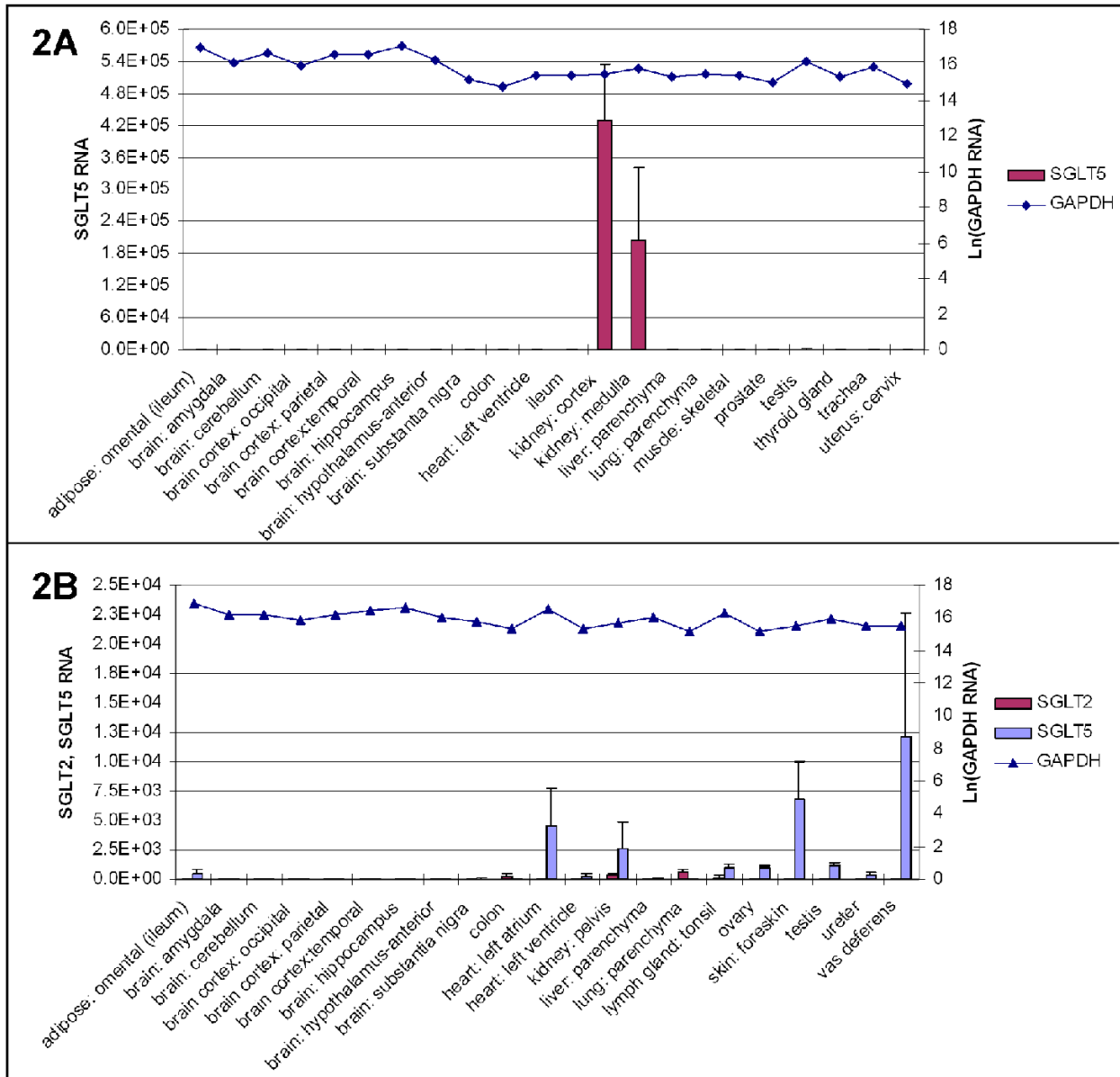
different exon regions (exons 1 and 2, exons 4 and 5, exons 9 and 10) of the SGLT2 gene (see the Appendix, Additional Figures S2-S4), further supporting that the expression of the SGLT2 gene is highly specific for the kidney cortex. Low transcript levels were detected across a range of tissues outside the kidney for the primer/probe set SGLT2-e4,5 (Table S2), possibly due to greater primer/probe set efficiency; however, even in this case, kidney expression was 100-fold higher than in the next highest tissue observed (the ileum). In addition, the primer/probe combination (SGLT2-e13), located in exon 13 close to the 3' end of the SGLT2 gene and identical to the one used by Zhou et al.,<sup>22</sup> shows that the SGLT2 gene is highly specifically expressed in the kidney (Figure 1B, Figure S5, and Table S2), unlike the result previously reported by these investigators. We observed that this primer/probe set essentially mirrors the data obtained with the primer/probe set spanning the junction of exons 6 and 7, as well as primer/probe sets spanning the exons 1-2, 4-5, and 9-10. We note that the primer/probe SGLT2-e13 lies in a region of overlap with another SGLT2-unrelated transcript, and thus is less specific to the SGLT2 sequence compared with other SGLT2 primer/probes we tested. The potential consequences of using this probe are further discussed in "the SGLT2 locus" section below. In summary, these data suggest that the steady state levels of SGLT2 transcripts are highly specific for the cortex of the kidney in humans.

SGLT5, a relatively uncharacterized SGLT family member, is also found to have a highly kidney-abundant tissue expression pattern. The expression of SGLT5 in a subset of the 72 human tissues is shown in Figure 2A (data from all 72 human tissues can be found in the Appendix, Additional Figure S6). The highest expression level of SGLT5 is found in the kidney cortex, while in the kidney medulla SGLT5 is expressed at about half of the level found in

**Figure 1.** Expression of sodium glucose cotransport protein SGLT2 in human tissues. Copies of SGLT2 transcript per 1  $\mu\text{g}$  of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu\text{g}$  of total RNA is shown on the right Y-axis. A subset of the 72 human tissues is shown on the X-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. (A) SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 6-7 of SGLT2 (SGLT2-e6,7); (B) SGLT2 tissue expression from TaqMan quantitative PCR reactions performed using primer/probe set designed in exon 13 of SGLT2 (SGLT2-e13).

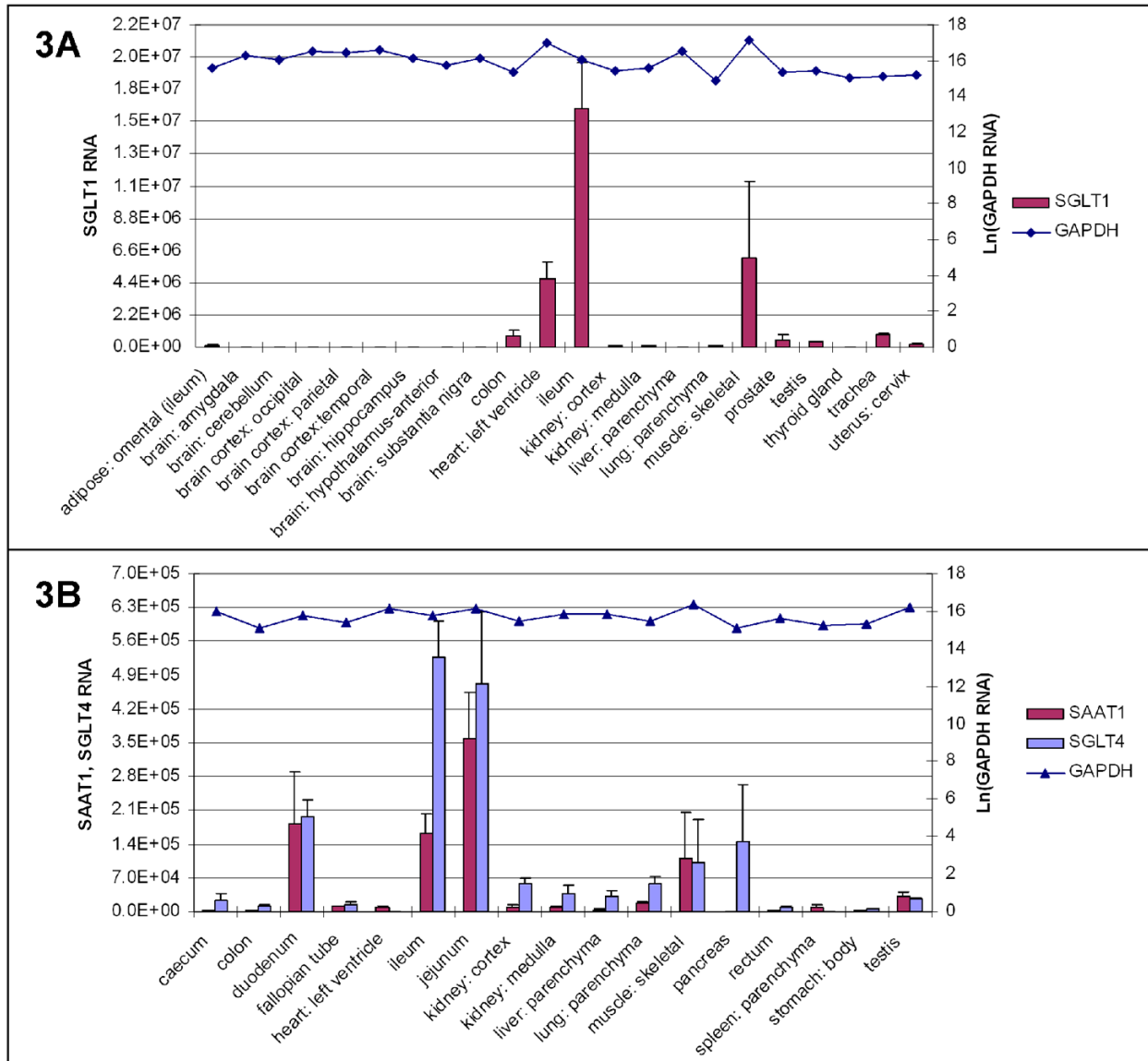


**Figure 2.** Expression of sodium glucose cotransport protein SGLT5 in human tissues. Copies of SGLT5 transcript per 1 µg of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the right Y-axis. A subset of the 72 human tissues is shown on the X-axis. The average number of transcripts from tissues of three human donors are shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. **(A)** SGLT5 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 11 of SGLT5. **(B)** Tissue expression of SGLT2 (primer/probe set in exon 6-7) and SGLT5 (primer/probe set in exon 11) plotted side-by-side in tissues absent of kidney cortex or medulla.





**Figure 3.** Expression of sodium glucose cotransport proteins SGLT1, sodium-dependent amino acid transporter (SAAT1), and SGLT4 in human tissues. Copies of SGLT1 transcript per 1  $\mu$ g of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the right Y-axis. A subset of the 72 human tissues are shown on the X-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. **(A)** SGLT1 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 4-5 of SGLT1. **(B)** Tissue expression of SAAT1 (primer/probe set in exon 2-3) and SGLT4 (primer/probe set in exon 11) plotted side-by-side.



the kidney cortex. Because the extremely high level of kidney expression may obscure the magnitude of expression observed in other tissues, we replotted the data without the kidney cortex and medulla data in Figure 2B. Unlike the profile observed for SGLT2, which has little or no detectable level in tissues other than the kidney cortex and medulla, SGLT5 exhibits a low level of expression in some tissues like the kidney pelvis, vas deferens, left atrium of the heart, skin, and testes. However, the expression of SGLT5 in the kidney cortex is still 35 times higher than that observed in the vas deferens. These data suggest that the expression of SGLT5 in human tissues is highly abundant in kidney, compared with other tissues.

### The Small Intestine and Muscle-Abundant SGLT Family Members

As shown in Figure 3A, the expression of SGLT1 (SLC5A1), the closest homolog to SGLT2, is essentially restricted to the small intestine, the skeletal muscle, and the heart (the data for all 72 human tissues are in the Appendix, Additional Figure S7 and Additional Table S2). Minor numbers of transcripts are observed in the trachea, prostate, cervix, and mesenteric adipose tissue, but like SGLT2, we see no evidence for any expression in the brain subregions tested. The GAPDH data indicates successful first strand synthesis and PCR in all samples. The expression pattern of SGLT1 across human tissues we observed above is generally consistent with reports from other studies using mRNA-based methods.<sup>6,7,9</sup> It is worth noting that brain expression of SGLT1 has been reported in other species (rats and pigs) using immunological, in situ hybridization and RT-PCR techniques.<sup>29,32,33</sup>

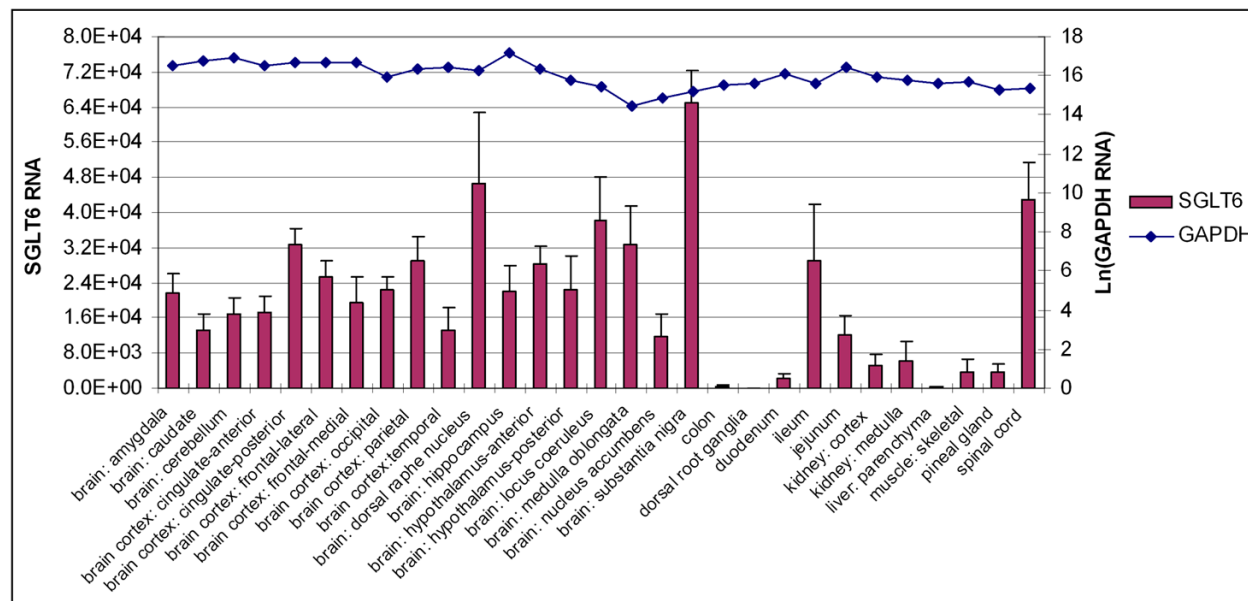
Interestingly, the glucose sensor SAAT1 (SLC5A4) and the low-affinity glucose/mannose cotransporter SGLT4 (SLC5A9) display a similar

high level of expression in the small intestine (duodenum, jejunum, and ileum) as well as in skeletal muscle (Figure 3B; data for all 72 human tissues are in the Appendix, Additional Figures S8 and S9). In the case of SAAT1, the highest expression could be found in the jejunum, and the steady state SAAT1 RNA level there is about 3.5-fold higher than in skeletal muscle. These data are consistent with other reports of SAAT1 expression in the intestine and skeletal muscle.<sup>11,13</sup> The steady state SGLT4 RNA level in the ileum is on the order of 5-fold higher than the next highest-expressing tissue outside of the gastrointestinal (GI) tract, the skeletal muscle. The expression of SGLT1, SAAT1, and SGLT4 in the GI tract seems to be enriched in the small intestinal region: duodenum, jejunum, and ileum; whereas these genes are expressed at a much lower level in the large intestine (cecum, colon, and rectum) as well as in other parts of the GI tract such as stomach and oesophagus (Figure 3). Unlike SGLT1, SAAT1 and SGLT4 display a much lower level of expression in the heart. On the other hand, SGLT4 has a uniquely moderate expression level in pancreas compared with SGLT1 and SAAT1. In all other tissues tested, SGLT1, SAAT1, and SGLT4 have a generally low level of expression (see the Appendix, Additional Figures S7, S8, and S9). Overall, the profile observed here for SGLT4 expression across human tissues is similar to that previously reported.<sup>16</sup>

### The Brain Expresses SGLT6

The solute carrier family 5A11 or SGLT6, a cotransporter with substrate specificity for myo-inositol and glucose, seems to be the only SGLT family member, aside from the ubiquitously-expressed SMIT (see next section), that has extensive expression in all the brain subregions tested. We observed high

**Figure 4.** Expression of sodium glucose cotransport protein SGLT6 in human tissues. Copies of SGLT6 transcript per 1  $\mu\text{g}$  of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu\text{g}$  of total RNA is shown on the right Y-axis. A subset of the 72 human tissues are shown on the X-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



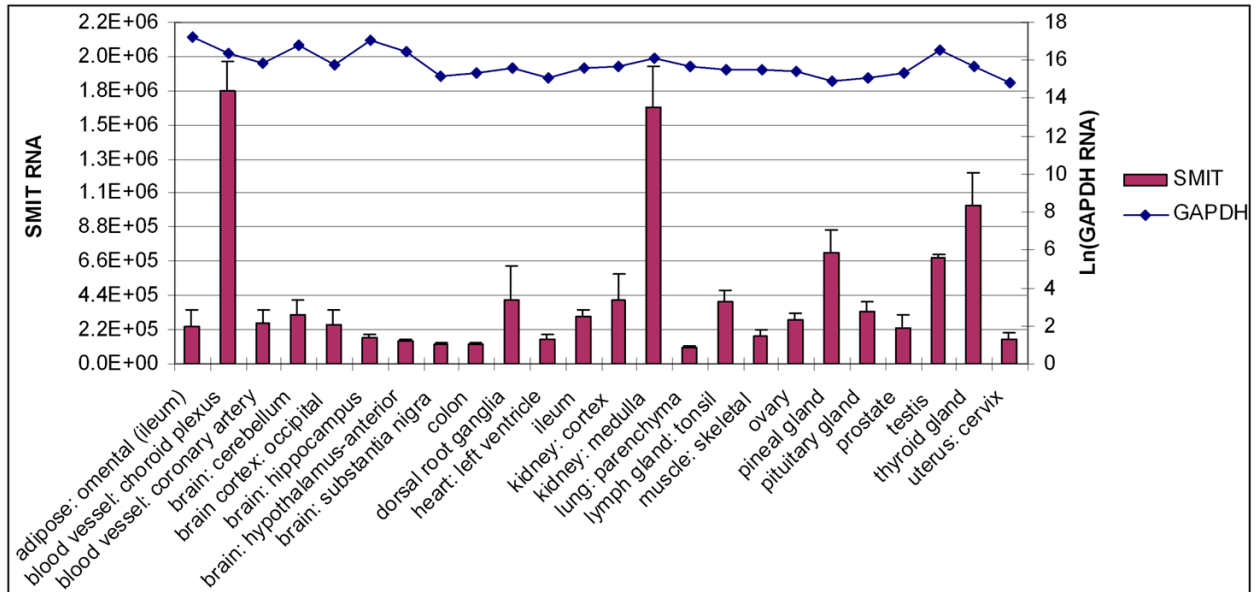
steady state RNA levels of SGLT6 collectively in the brain, with the highest subregion being the substantia nigra where it was found to be expressed at levels 2-5-fold higher than the other regions tested (Figure 4; data for all 72 human tissues are in the Appendix, Additional Figure S10). SGLT6 is also highly expressed in the spinal cord, at a level similar to that in substantia nigra. However, it is not detected in the dorsal root ganglion (DRG). Other tissue RNA samples with notable expression levels of SGLT6 are the small intestine (ileum and jejunum), kidney (cortex and medulla), as well as skeletal muscle. The observed pattern of SGLT6 in these studies appears to be unique among the SGLT family members tested, and if this pattern is similar in rodents, SGLT6 activity could account, at least in part, for the observation of functional SGLT expression in rat brain.<sup>29</sup> It is interesting

to note that the human genomic location of SGLT6 coincides with a locus associated with the nervous system disorders of infantile convulsions and choreoathetosis, though no disease-associated mutations were found in the exon or intron/exon boundary sequences of the SGLT6 gene.<sup>18</sup> The pattern of expression in human tissues described here differs from that initially described by northern blot,<sup>18</sup> although brain expression was detected in both studies.

### The Ubiquitously Expressed SMIT

The sodium myo-inositol cotransporter SMIT (SLC5A3) shows a ubiquitous pattern of expression with the highest expression level in the medulla of the kidney and the blood vessel of the choroid plexus (Figure 5; data for all 72 human tissues are in the Appendix,

**Figure 5.** Expression of sodium myo-inositol cotransporter (SMIT) in human tissues. Copies of SMIT transcript per 1  $\mu$ g of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the right Y-axis. A subset of the 72 human tissues are shown on the X-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



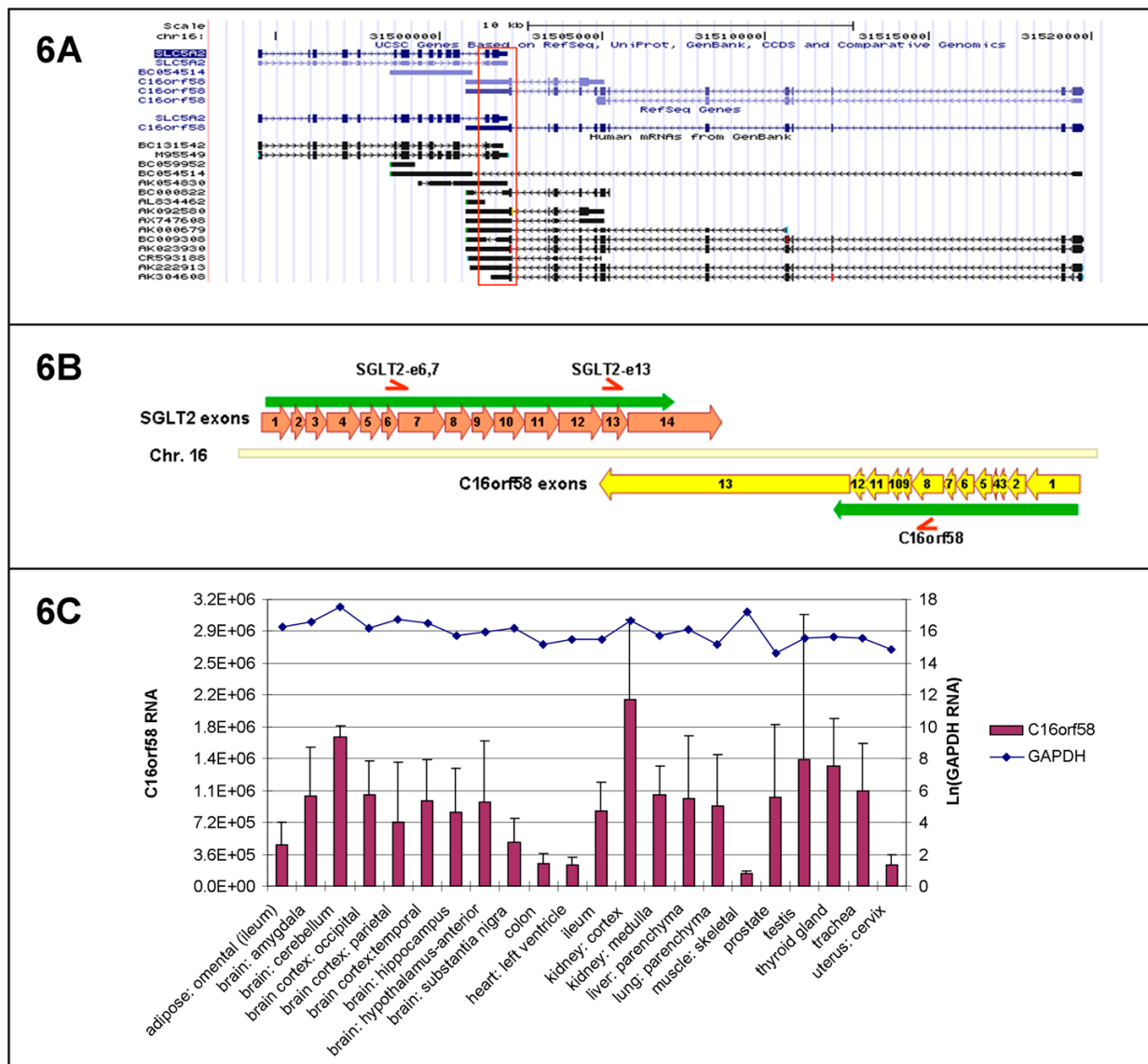
Additional Figure S11). The thyroid gland, pineal gland, dorsal root ganglion, and the testes also have high levels of SMIT expression. Overall, the pattern of expression of SMIT is consistent with what has been previously reported,<sup>14</sup> and its expression in all tissues examined highlights its potentially important role in the maintenance of osmotic balance within cells.

### The SGLT2 Locus

The discrepancy between our results and that reported by Zhou et al.<sup>22</sup> prompted us to examine the SGLT2 locus in more detail. The SGLT2 gene resides on chromosome 16 where 14 exons span approximately 7 kilobases of genomic DNA. The last two exons, exon 13 and 14, overlap with exon 13 of a gene that is encoded on the opposite strand, called C16orf58 (Figure 6A). This gene, conserved in plants, invertebrates, and vertebrates, contains 13 exons spanning 20 kb of genomic

DNA, and encodes a protein homologous to the Arabidopsis RUS1 gene.<sup>34,35</sup> Numerous sequence submissions to Genbank suggest C16orf58 is indeed expressed. Our internal cDNA cloning effort has obtained full-length cDNA clone of C16orf58. The C16orf58 cDNA clone has a long 3' untranslated region that contains the reverse-complement sequence of exon 13 and 14 of the SGLT2 gene (data not shown). Electronic northern blots using the region of overlap as a 'probe' indicates that expressed sequence tags (ESTs) for C16orf58 versus that of SGLT2 can be found in the NCBI database at the relative abundance of 10 transcripts to one (data not shown). We designed a primer/probe set located in exon 8 of C16orf58 and one that is specific for C16orf58 (Figure 6B). Expression profiling analysis with this primer/probe set indicates that this gene is expressed ubiquitously with the highest steady state RNA levels also located in the cortex of the kidney, the cerebellum (Figure 6C), and the

**Figure 6.** Analysis of the sodium glucose cotransport protein SGLT2 locus in human genome. (A) A genome browser view of the 22 kb region of the human SGLT2 locus in chromosome 16 obtained from the UCSC Genome Browser of the human genome GRCh37 Assembly (hg19). The red rectangular box highlights the overlapping of exon 13 and 14 of SGLT2 with exon 13 of C16orf58. (B) Schematic drawing of the exon structure of the SGLT2 gene and C16orf58 gene on chromosome 16. Green block arrows represent the open reading frames of the genes. Pink block arrows represent the exons of SGLT2 gene while yellow block arrows represent the exons of C16orf58 genes. Locations of amplicons of TaqMan quantitative polymerase chain reactions (PCR) are represented by red arrow heads. (C) Copies of C16orf58 transcript per 1 µg of total RNA are shown on the left Y-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the right Y-axis. A subset of the 72 human tissues are shown on the X-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



thyroid (see the Appendix, Additional Figure S12). The primer/probe set (SGLT2-e13) used by Zhou et al.<sup>22</sup> resides in exon 13 of SGLT2, in the area of overlap between SGLT2 and C16orf58 (Figure 6B), suggesting that measurement of both transcripts would be confounding in experiments where poly d(T) was used to prime first strand synthesis in a one-step PCR process. We have used each of the primers employed by these investigators in separate first strand reactions and have shown that the reverse primer is capable of synthesizing kidney cortex abundant SGLT2 cDNA (Figure 1B), however we have been unable to recapitulate the expression pattern seen with the C16orf58-specific primer (Figure 6C) using the Zhou et al.<sup>22</sup> forward primer (data not shown).

## DISCUSSION

We have used quantitative RT-PCR to examine the expression profile of SGLT2 and related family members across a panel of 72 tissues from three healthy individuals. This study is the most comprehensive analysis done to date on this important family of sodium-monosaccharide cotransporters, and was carried out to increase our understanding of where these transporters are expressed in human tissues. The data presented clearly demonstrates that SGLT2 is expressed primarily in the cortex of the kidney at steady state levels that are several hundred-fold greater than the tissue with the next highest abundance, the kidney medulla. We have confirmed these data with multiple primer/probe combinations representing essentially the entire SGLT2 locus.

These data are in agreement with some previous studies<sup>9,16,20</sup> but in conflict with others,<sup>21,22</sup> who reported that SGLT2 showed a more widespread pattern of expression. We have attempted to reconcile our results with those of Zhou et al.<sup>22</sup> by using their primer/

probe combination in conjunction with our PCR methodology, which uses orientation-specific first strand cDNA synthesis; however, we were unable to reproduce their results. Instead we demonstrate that like that obtained with all SGLT2 primer/probe combinations tested, SGLT2 is indeed restricted in its expression to the cortex of the kidney. We analyzed the SGLT2 locus and discovered that the primer/probe amplicon SGLT2-e13 overlaps with extreme 3' end of another gene encoded on the opposite strand, C16orf58. Our results indicate that this conserved gene with homology to the Arabidopsis *RUS1* gene<sup>34,35</sup> is indeed ubiquitously expressed at moderately abundant levels in all 72 tissues tested and is also expressed in the cortex of the kidney. We attempted to use the forward primer employed by Zhou et al.<sup>22</sup> to prime first strand synthesis of the 3' end of C16orf58 followed by PCR, to determine if that method could give rise to the expression profile they reported, but we were unable to recapitulate the published profile, nor could we replicate the profile obtained with the C16orf58-specific primer/probe set located in exon 8. Either the C16orf58 forward primer is inefficient in this role as a primer of first strand synthesis, or the 3' end of the C16orf58 transcript is particularly unstable and easily degraded. However, the large number of Express Sequence Tags that can be found in Genbank generated from the C16orf58 strand that overlaps the 3' end of SGLT2 as well as our own internal C16orf58 cloning efforts suggests that the latter is not the case.

To better understand the physiological consequences of agonizing and antagonizing protein function it is extremely important to know in which tissues a gene is expressed, particularly from the viewpoint of assessing potential liabilities. Included in the tissue panel was RNA isolated from 20 different brain subregions. We see no evidence that SGLT2 is

expressed in any of the brain subregions tested, even though the same RNA samples were used for other profiles that did return detectable transcript numbers, e.g. SGLT6, where 52 out of 60 brain subregion RNA samples used in both experiments were identical, and the GAPDH control indicated successful enzymatic reactions. However, this conclusion is limited by the detection level afforded by quantitative PCR methodologies, and we can not rule out the possibility that SGLT2 is expressed within small numbers of discrete brain cells, as well as other tissues, that are undetectable by the TaqMan protocol.

Another method to evaluate RNA expression is by Affymetrix RNA chip hybridization. The human SGLT2 gene is represented by probe set 207771\_at on Affymetrix RNA chips. The mouse SGLT2 gene is represented by probe sets 1419166\_at and 1455005\_a\_at, while rat SGLT2 is represented by probe set U29881\_at. The public domain gene expression database BioGPS<sup>36</sup> contains tissue expression profiling data for these probe sets (see the Appendix, Additional Figure S13). In the Affymetrix chip database, SGLT2 transcripts are found to be highly specific to the kidney tissue of human, mouse, and rat, consistent with our human tissue quantitative RT-PCR study results. These data are further supported by the recent publication of immunohistochemical data in mice, confirming the original localization of the rodent SGLT2 protein to the early proximal tubule,<sup>19</sup> which was clearly absent in the SGLT2 knock-out mouse.<sup>27</sup>

Certainly, the sodium/glucose cotransporter family member with the closest expression profile to that of SGLT2 is SGLT5, which is expressed almost exclusively in the cortex and medulla of the kidney. The only other tissues that show a much lower level of expression are the heart, skin, and vas deferens. This highly restricted gene expression suggests a kidney-specific function,

however, little is known about this gene and its gene product, including what its substrate specificity may be. Hence the putative role of this protein in regulating solute homeostasis in the kidney remains to be determined.

The SGLT1 expression profile agrees with the well-established role of this gene as encoding the intestinal high-affinity low capacity glucose reabsorption cotransporter<sup>6</sup> with the highest steady state levels being observed in the ileum and other regions of the small intestine.<sup>7</sup> The observed expression of SGLT1 in the ventricle of the heart agrees with that reported by Zhou et al.<sup>22</sup> as well as with our own unpublished data (Hagan D., data not shown). It is curious to note that SGLT1 expression in the heart is apparently restricted to humans, since no expression of SGLT1 is detectable in the heart tissue of rats,<sup>7</sup> dogs, pigs, or cynomolgus monkeys (Hagan D., unpublished results.) Our data also suggest that transcripts for SGLT1 can be found in skeletal muscle at a level similar to the heart. The relative high level of SGLT1 expression in heart and skeletal muscle suggests that it might play a significant but yet-to-be-defined role in glucose absorption in those tissues.

The glucose transporters SAAT1 and SGLT4, the latter also having affinity for mannose as a substrate, have a surprisingly similar tissue expression pattern in the small intestine and skeletal muscle compared to SGLT1. Interestingly, the expression of these three SGLT member genes seem to be limited to the small intestinal portion of the GI tract with the duodenum, jejunum, and ileum all having the most prominent levels of detectable RNA; other parts of the GI tract, such as the cecum, colon, rectum, and the stomach have much lower level of gene expression. However, unlike SGLT1, SAAT1 and SGLT4 do not have a high level of expression in the heart. SGLT4 also demonstrates high level of expression in the

pancreas, although at this time we do not know which specific cell types in the pancreas may contribute to the overall steady state level of RNA. The specific overlap in tissue expression pattern of SGLT1, SAAT1, and SGLT4 suggest that these three SGLT members could have similar and/or complementary functions in the body. Based on the localization of the SAAT1 protein, and the observation of ion transport decoupled from glucose transport activity, Diez-Sampedro and colleagues have proposed a role for SAAT1 as a glucose sensor in cholinergic neurons of the GI tract and at neuromuscular junctions, as a potential modulator of gastric motility and muscular activity.<sup>13</sup>

We show here that SGLT6, a gene recently implicated in a genetic association study to be a modifying locus of systemic lupus erythematosus (SLE), displays a neuronal expression pattern suggesting an additional putative role for this gene outside that of the immune system in regulating myo-inositol homeostasis in the brain and spinal cord.<sup>37</sup> Interestingly, our data does not support an immune cell expression pattern but instead indicates that SGLT6 is not significantly expressed (at the level of quantitative RT-PCR resolution) outside of the nervous system, the small intestine, and the kidney.

Given the interest in SGLT2 inhibitors as a potential therapeutic approach for diabetes, it is worth noting that not only is the tissue localization of SGLT2 expression relevant to the therapeutic profile, but also the selectivity profile for inhibitors currently in clinical development, since a lack of specificity for one or more other SGLT proteins could have physiological consequences beyond the kidney. Various groups have reported the *in vitro*, assay-specific 50% inhibitory concentration ( $IC_{50}$ )-based selectivity of SGLT2 inhibitors versus SGLT1, with dapagliflozin and BI 10773 showing the highest degree of selectivity: 1241-fold and

>2500-fold selectivity, respectively;<sup>38,39</sup> whereas the compounds canagliflozin, ASP1941, and LX4211 have reported selectivities of 414-, 255- and 20-fold, respectively.<sup>40,41,42</sup> The activity of these compounds vs. other members of the SGLT family have not been consistently reported to date. Once the inhibitory constant ( $K_i$ ) values, generated from data obtained across a range of substrate concentrations, are reported, the values should be more directly comparable across compounds. The impact of the overall selectivity profile of these agents should become clear as the therapeutic profile emerges during clinical development.

## CONCLUSION

In summary, in order to better understand the spectrum of human tissues in which the sodium-glucose cotransporter family of proteins is expressed, we have examined the mRNA expression profile of human SGLT2 and six of its homologs (SGLT1, SAAT1, SMIT, SGLT4, SGLT5, SGLT6) in a panel of 72 normal human tissues from three separate donors using quantitative RT-PCR methods. For SGLT2, using five different primer/probe sets spanning the entire locus, we confirm the kidney-specific expression of this gene. Thus we conclude that SGLT2 expression is highly restricted to the kidney in humans, and that the expression of related proteins of this family, with the exception of SGLT5, diverges from this kidney specificity. It is a highly desirable requirement that the expression of a protein drug target be restricted to the tissue where its therapeutic effects can be most efficacious. The work presented here demonstrating the kidney-specific expression of SGLT2 helps to satisfy this requirement and further augments the importance and value of this protein as a drug target for treatment of diabetes.



## ACKNOWLEDGMENTS

Jian Chen was the lead scientist designing experiments and interpreting results. Sandy Williams, Samantha Ho, and Howard Loraine performed the experimental work. John Feder contributed interpretation of results. Jian Chen and John Feder prepared the manuscript. Jean Wader provided critical reading and editing of the manuscript as well as helping to direct the experimental design. All authors read and approved the final manuscript.

We would like to thank M. Liu for her early contributions to this work, Gabriel Mintier for the molecular cloning and sequence analysis of the C16orf58 cDNA, and Ida Wall for her contribution to the quantitative RT-PCR experimental work. Bristol-Myers Squibb (BMS) and AstraZeneca (AZ) provided the funding for the studies described. BMS authors designed the experiments which were implemented by Asterand authors according to their technical standards. Asterand authors collected the data, and BMS authors analyzed and interpreted the data. BMS authors wrote the draft manuscript with review and comment by Asterand authors, and BMS authors made the decision to submit the manuscript for publication.

John Feder is the guarantor for this article, and takes responsibility for the integrity of the work as a whole.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## REFERENCES

1. International Diabetes Federation. Some 285 million people worldwide will live with diabetes in 2010. IDF Diabetes Atlas. Available at: <http://www.diabetesatlas.org/content/some-285-million-people-worldwide-will-live-diabetes-2010>. Last accessed February 3, 2010.
2. Diabetes Statistics - American Diabetes Association. Available at: <http://www.diabetes.org/diabetes-basics/diabetes-statistics>. Last accessed February 4, 2010.
3. Bakris GL, Fonseca VA, Sharma K, Wright EM. Renal sodium-glucose transport: role in diabetes mellitus and potential clinical implications. *Kidney Int.* 2009;75:1272-1277.
4. Bailey CJ, Gross JL, Pieters A, et al. Effect of dapagliflozin in patients with type 2 diabetes who have inadequate glycaemic control with metformin: a randomised, double-blind, placebo-controlled trial. *Lancet.* 2010;375:2223-2233.
5. Wright EM, Turk E. The sodium/glucose cotransport family SLC5. *Pflugers Arch.* 2004;447:510-518.
6. Hediger MA, Coady MJ, Ikeda TS, Wright EM. Expression cloning and cDNA sequencing of the Na<sup>+</sup>/glucose co-transporter. *Nature.* 1987;330:379-381.
7. Lee WS, Kanai Y, Wells RG, Hediger MA. The high affinity Na<sup>+</sup>/glucose cotransporter. Re-evaluation of function and distribution of expression. *J Biol Chem.* 1994;269:12032-12039.
8. Martín MG, Turk E, Lostao MP, et al. Defects in Na<sup>+</sup>/glucose cotransporter (SGLT1) trafficking and function cause glucose-galactose malabsorption. *Nat Genet.* 1996;12:216-220.
9. Wells RG, Pajor AM, Kanai Y, et al. Cloning of a human kidney cDNA with similarity to the sodium-glucose cotransporter. *Am J Physiol.* 1992;263:F459-465.
10. Kanai Y, Lee WS, You G, et al. The human kidney low affinity Na<sup>+</sup>/glucose cotransporter SGLT2. Delineation of the major renal reabsorptive mechanism for D-glucose. *J Clin Invest.* 1994;93:397-404.
11. Kong CT, Yet SF, Lever JE. Cloning and expression of a mammalian Na<sup>+</sup>/amino acid cotransporter with sequence similarity to Na<sup>+</sup>/glucose cotransporters. *J Biol Chem.* 1993;268:1509-1512.
12. Mackenzie B, Panayotova-Heiermann M, Loo DD, et al. SAAT1 is a low affinity Na<sup>+</sup>/glucose cotransporter and not an amino acid transporter. A reinterpretation. *J Biol Chem.* 1994;269:22488-22491.

13. Diez-Sampedro A, Hirayama BA, Osswald C, et al. A glucose sensor hiding in a family of transporters. *Proc Natl Acad Sci USA*. 2003;100:11753-11758.
14. Berry GT, Mallee JJ, Kwon HM, et al. The human osmoregulatory Na<sup>+</sup>/myo-inositol cotransporter gene (SLC5A3): molecular cloning and localization to chromosome 21. *Genomics*. 1995;25:507-513.
15. Hale CC, Rubin LJ. Ion specificity and stoichiometry of the cardiac inositol transporter. *J Mol Cell Cardiol*. 1995;27:1123-1130.
16. Tazawa S, Yamato T, Fujikura H, et al. SLC5A9/SGLT4, a new Na<sup>+</sup>-dependent glucose transporter, is an essential transporter for mannose, 1,5-anhydro-D-glucitol, and fructose. *Life Sci*. 2005;76:1039-1050.
17. Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA*. 2002;99:16899-16903.
18. Roll P, Massacrier A, Pereira S, et al. New human sodium/glucose cotransporter gene (KST1): identification, characterization, and mutation analysis in ICCA (infantile convulsions and choreoathetosis) and BFIC (benign familial infantile convulsions) families. *Gene*. 2002;285:141-148.
19. Coady MJ, Wallendorff B, Gagnon DG, Lapointe J. Identification of a novel Na<sup>+</sup>/myo-inositol cotransporter. *J Biol Chem*. 2002;277:35219-35224.
20. You G, Lee WS, Barros EJ, et al. Molecular characteristics of Na<sup>(+)</sup>-coupled glucose transporters in adult and embryonic rat kidney. *J Biol Chem*. 1995;270:29365-29371.
21. Wright EM, Hirayama BA, Loo DF. Active sugar transport in health and disease. *J Intern Med*. 2007;261:32-43.
22. Zhou L, Cryan EV, D'Andrea MR, et al. Human cardiomyocytes express high level of Na<sup>+</sup>/glucose cotransporter 1 (SGLT1). *J Cell Biochem*. 2003;90:339-346.
23. Rubera I, Poujeol C, Bertin G, et al. Specific Cre/Lox recombination in the mouse proximal tubule. *J Am Soc Nephrol*. 2004;15:2050-2056.
24. Silverman M, Speight P, Ho L. Identification of two unique polypeptides from dog kidney outer cortex and outer medulla that exhibit different Na<sup>+</sup>/D-glucose cotransport functional properties. *Biochim Biophys Acta*. 1993;1153:43-52.
25. Canani RB, De Marco G, Passariello A, et al. Inhibitory effect of HIV-1 Tat protein on the sodium-D-glucose symporter of human intestinal epithelial cells. *AIDS*. 2006;20:5-10.
26. Lahjouji K, Aouameur R, Bissonnette P, et al. Expression and functionality of the Na<sup>+</sup>/myo-inositol cotransporter SMT2 in rabbit kidney. *Biochim Biophys Acta*. 2007;1768:1154-1159.
27. Balen D, Ljubojevic M, Breljak D, et al. Revised immunolocalization of the Na<sup>+</sup>-D-glucose cotransporter SGLT1 in rat organs with an improved antibody. *Am J Physiol Cell Physiol*. 2008;295:C475-489.
28. Vallon V, Platt KA, Cunard R, et al. SGLT2 mediates glucose reabsorption in the early proximal tubule. *J Am Soc Nephrol*. 2010. In press.
29. Yu AS, Hirayama BA, Timbol G, et al. Functional expression of SGLTs in rat brain. *Am J Physiol Cell Physiol*. 2010. In press.
30. Barber RD, Harmer DW, Coleman RA, Clark BJ. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol Genomics*. 2005;21:389-395.
31. Human BLAT Search Genome. Available at: <http://genome.ucsc.edu/cgi-bin/hgBlat?hgsid=151562971>. Last accessed October 1, 2010.
32. Poppe R, Karbach U, Gambaryan S, et al. Expression of the Na<sup>+</sup>-D-glucose cotransporter SGLT1 in neurons. *J Neurochem*. 1997;69:84-94.
33. Elfeber K, Köhler A, Lutzenburg M, et al. Localization of the Na<sup>+</sup>-D-glucose cotransporter SGLT1 in the blood-brain barrier. *Histochem Cell Biol*. 2004;121:201-207.
34. Tong H, Leasure CD, Hou X, Yuen G, Briggs W, He Z. Role of root UV-B sensing in Arabidopsis early seedling development. *Proc Natl Acad Sci USA*. 2008;105:21039-21044.
35. Leasure CD, Tong H, Yuen G, Hou X, Sun X, He Z. ROOT UV-B SENSITIVE2 acts with ROOT UV-B SENSITIVE1 in a root ultraviolet B-sensing pathway. *Plant Physiol*. 2009;150:1902-1915.
36. Wu C, Orozco C, Boyer J, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology*. 2009;10:R130.
37. Tsai LJ, Hsiao SH, Tsai LM, et al. The sodium-dependent glucose cotransporter SLC5A11 as an

- autoimmune modifier gene in SLE. *Tissue Antigens*. 2008;71:114-126.
38. Han S, Hagan DL, Taylor JR, et al. Dapagliflozin, a selective SGLT2 inhibitor, improves glucose homeostasis in normal and diabetic rats. *Diabetes*. 2008;57:1723-1729.
  39. Grempler R, Thomas L, Eckhardt M, et al. In vitro properties and in vivo effect on urinary glucose excretion of BI 10773, a novel selective SGLT2 inhibitor. *Diabetes*. 2009;58(suppl. 1):A139.
  40. Nomura S, Sakamaki S, Hongu M, et al. Discovery of canagliflozin, a novel C-glucoside with thiophene ring, as sodium-dependent glucose cotransporter 2 inhibitor for the treatment of type 2 diabetes mellitus. *J Med Chem*. 2010;53:6355-6360.
  41. Kurosaki E, Tahara A, Yokono M, et al. In vitro and in vivo pharmacological properties of ASP1941: a novel, potent and selective SGLT2 inhibitor. *Diabetes*. 2010;59(suppl. 1):A156.
  42. Freiman J, Ruff DA, Frazier K, et al. LX4211, a dual SGLT2/SGLT1 inhibitor, shows rapid and significant improvement in glycemic control over 28 days in patients with type 2 diabetes (T2D). American Diabetes Association 2010; poster #17-LB. Available at: [www.lexicon-genetics.com](http://www.lexicon-genetics.com). Last accessed July 13, 2010.

## APPENDIX

Additional Table S1. Primer/probe sequences of TaqMan quantitative polymerase chain reactions (PCR).

	Forward	Reverse	TaqMan Probe
<b>SGLT1</b>	5'-CCCTGGTTTTGGTGGTTGTG-3'	5'-TCACCACCCCAGCCTTAATATAG-3'	FAM-5'-TGGGGACAAACAGCCAGCCCA-3'-TAMRA
<b>SGLT2-e6,7 (exon 6-7)</b>	5'-GACACGGTACAGACCTTCGTCAT-3'	5'-CTCCCAGGTATTTGTCGAAGAGA-3'	FAM-5'-CCTCATGGTTACGCCCT-3'-TAMRA
<b>SGLT2-e1,2 (exon 1-2)</b>	5'-AAGGCCCTGATTGACAATCCT-3'	5'-TGTTCTGCACATGGACCAC-3'	FAM-5'-ACATCCTAGTCATTGCTG-3'-TAMRA
<b>SGLT2-e4,5 (exon 4-5)</b>	5'-CCGCATCCGCCCTCTACCT-3'	5'-CATGTCCACTGAGATCTTGGTGAA-3'	FAM-5'-GTCGTGCTCTCCCTT-3'-TAMRA
<b>SGLT2-e9,10 (exon 9-10)</b>	5'-CATGCCCAACGGTCTGC-3'	5'-GCGTGTAGATGTCCATGGTGA-3'	FAM-5'-CCGCGCTCATGTCCCTCGCTG-3'-TAMRA
<b>SGLT2-e13 (exon 13)</b>	5'-TTCAGTCTCCGGCATAGCAAG-3'	5'-CATCTCCATGGCACTCTCTGG-3'	FAM-5'-AGGAACGGGAGGACCTGGATGCTG-3'-TAMRA
<b>SAAT1</b>	5'-TCCTCGCTGGTCGTGATATG-3'	5'-TCCTCGCTGGTCGTGATATG-3'	FAM-5'-CGCCCATCGGCCACCAGG-3'-TAMRA
<b>SMIT</b>	5'-ACCCCAGCTTCAGTATGGTACTG-3'	5'-CAATGTTTTGGCTGCAAGGA-3'	FAM-5'-CTCTGCACGATGACTTGGTCAGCACA-3'-TAMRA
<b>SGLT4</b>	5'-GCAGCAGCACCCCTGTTAC-3'	5'-GCTCCTGCTCTGTTGACTTCCT-3'	FAM-5'-CGGAAGCGCTGCCACACATCAA-3'-TAMRA
<b>SGLT5</b>	5'-GCTGATGATCGCAGTATGCT-3'	5'-GAGGGTCTGCTGCTGTTG-3'	FAM-5'-CGGCGCTCATGTCGTCGCTG-3'-TAMRA
<b>SGLT6</b>	5'-CTCCCTCACCTCCATCTTAAACA-3'	5'-GGCCGGAGGTGATTCCA-3'	FAM-5'-TGCCAGCACCATCTTCCACATGGA-3'-TAMRA
<b>C16orf58</b>	5'-GGTCATGGAGACCTGAACGA-3'	5'-GTTGGGTCGAGTACCTCTCCC-3'	FAM-5'-CTCCGGCTGGTCTGAAGCACTACC-3'-TAMRA

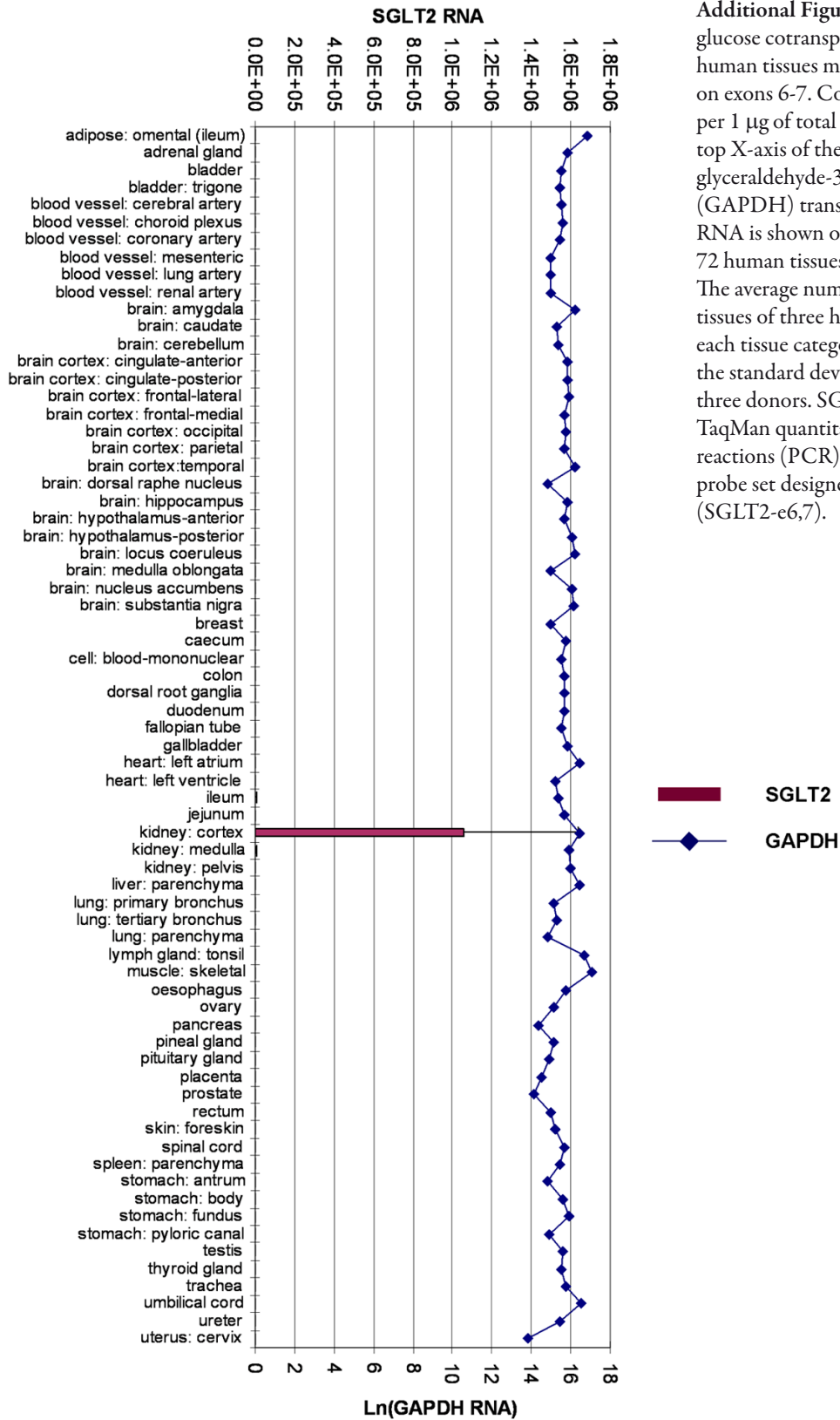
The forward, reverse primer sequences and the TaqMan Probe sequence of the PCR reactions are given in the table. Sodium glucose cotransporter protein SGLT2-e1,2: amplicon spans exon 1 and 2 of the SGLT2 gene; SGLT2-e4,5: amplicon spans exon 4 and 5 of the SGLT2 gene; SGLT2-e6,7: amplicon spans exon 6 and 7 of the SGLT2 gene; SGLT2-e9,10: amplicon spans exon 9 and 10 of the SGLT2 gene; SGLT2-e13: amplicon is located in exon 13 of the SGLT2 gene; SAAT=sodium-dependent amino acid transporter; SMIT=sodium myo-inositol cotransporter.

SAAT=sodium-dependent amino acid transporter; SMIT=sodium myo-inositol cotransporter.

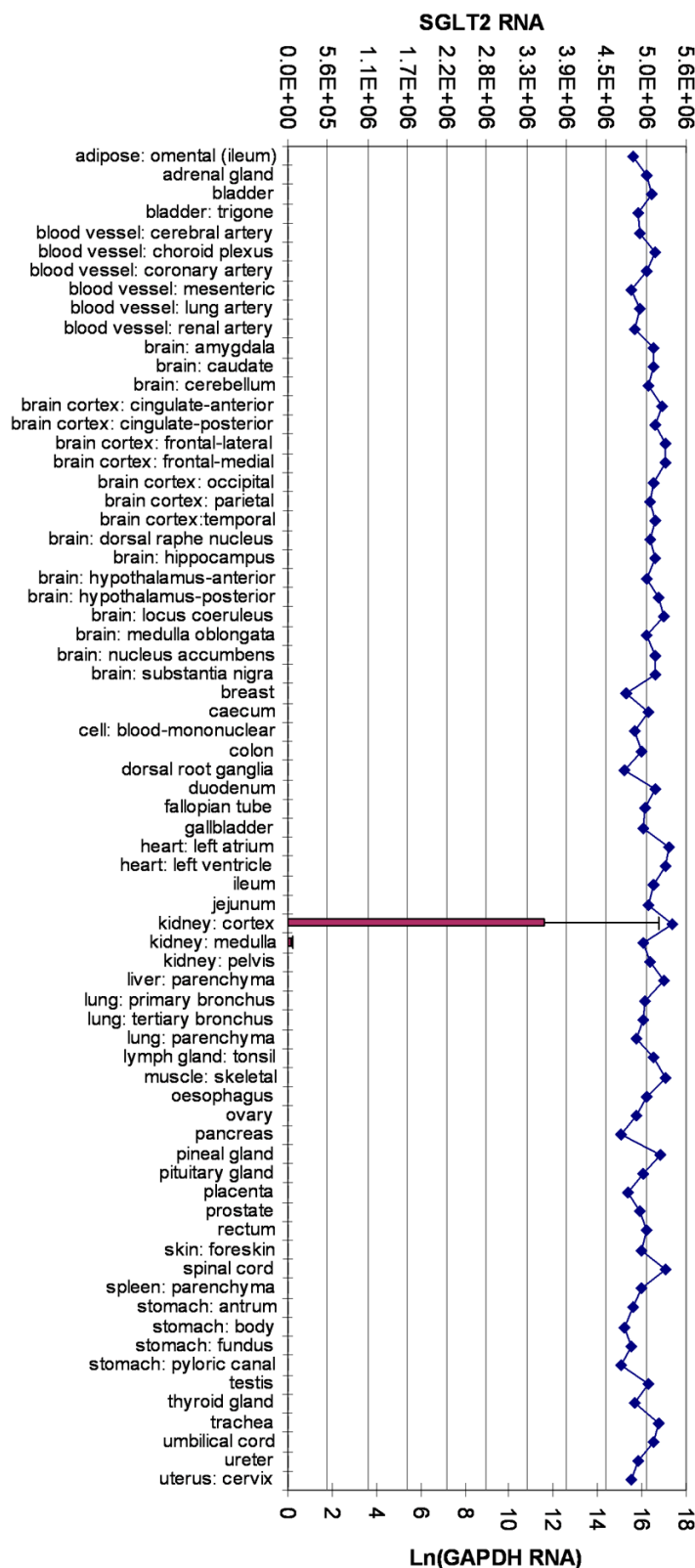
**Additional Table S2.** Expression of sodium glucose cotransporter proteins SGLT1 and SGLT2 in selected human tissues.

Tissues	SGLT1		SGLT2-e6,7		SGLT2-e1,2		SGLT2-e4,5		SGLT2-e9,10		SGLT2-e13	
	Molecules	Stdev	Molecules	Stdev	Molecules	Stdev	Molecules	Stdev	Molecules	Stdev	Molecules	Stdev
adrenal_gland	480	256	0	0	0	0	1,567	422	0	0	13,179	10,101
bladder	167	119	0	0	0	0	10,947	3,211	0	0	556	325
brain_amygdala	5,140	1,208	0	0	0	0	573	247	7	6	3,146	1,635
brain_cerebellum	1,453	722	13	22	0	0	2,447	803	1,553	580	31,177	16,932
brain_cortex_frontal_lateral	4,987	917	0	0	0	0	433	81	0	0	3,901	2,065
brain_hippocampus	2,840	714	0	0	0	0	33	21	0	0	2,617	1,964
brain_hypothalamus_anterior	3,260	1,023	0	0	0	0	300	132	0	0	6,586	7,225
brain_substantia_nigra	1,373	42	0	0	0	0	693	300	0	0	1,655	1,485
colon	735,400	442,859	201	224	0	0	12,807	4,049	80	53	1,258	1,343
duodenum	7,582,327	2,087,015	724	679	0	0	25,440	16,532	540	425	4,386	5,882
heart_left_atria	1,194,247	619,050	16	29	0	0	287	248	0	0	598	270
heart_left_ventricle	4,619,293	1,211,044	0	0	0	0	213	185	0	0	1,055	1,265
ileum	16,239,473	3,231,922	1,997	1,951	53	46	120,340	67,645	1,140	330	8,786	5,791
jejunum	11,689,540	3,950,853	504	436	0	0	19,293	7,478	313	74	11,757	18,227
kidney_cortex	48,233	12,179	1,060,193	576,118	3,603,953	1,603,929	12,674,780	5,048,294	510,427	313,667	1,266,094	840,276
kidney_medulla	70,787	32,861	5,708	4,635	39,640	20,568	237,793	116,510	2,387	1,465	16,587	17,180
liver_parenchyma	28,093	4,942	8	14	0	0	2,013	425	0	0	3,878	1,371
lung_parenchyma	102,007	18,553	582	215	0	0	23,853	3,009	360	174	8,253	6,988
muscle_skeletal	6,095,513	5,278,514	0	0	0	0	707	76	0	0	2,127	3,465
pancreas	32,707	9,275	268	69	253	219	12,693	4,550	233	51	8,240	8,222
placenta	2,733	1,248	30	52	0	0	2,980	286	40	17	1,666	915
prostate	484,867	351,243	0	0	0	0	2,907	1,740	380	174	9,384	10,968
spinal_cord	280	44	0	0	53	46	1,213	621	0	0	763	631
spleen_parenchyma	60	52	22	39	47	40	793	280	67	58	9,996	14,554
stomach_body	22,960	8,034	80	29	33	29	4,027	1,131	93	57	1,677	787
testis	392,287	25,196	15	27	40	35	2,707	846	15,340	7,588	55,533	50,601
thyroid_gland	967	452	119	176	453	393	4,527	2,319	673	261	9,333	7,245
trachea	830,613	148,191	17	30	0	0	1,667	102	240	121	3,837	2,448
uterus_cervix	204,773	108,308	13	22	0	0	1,293	338	7	6	624	503

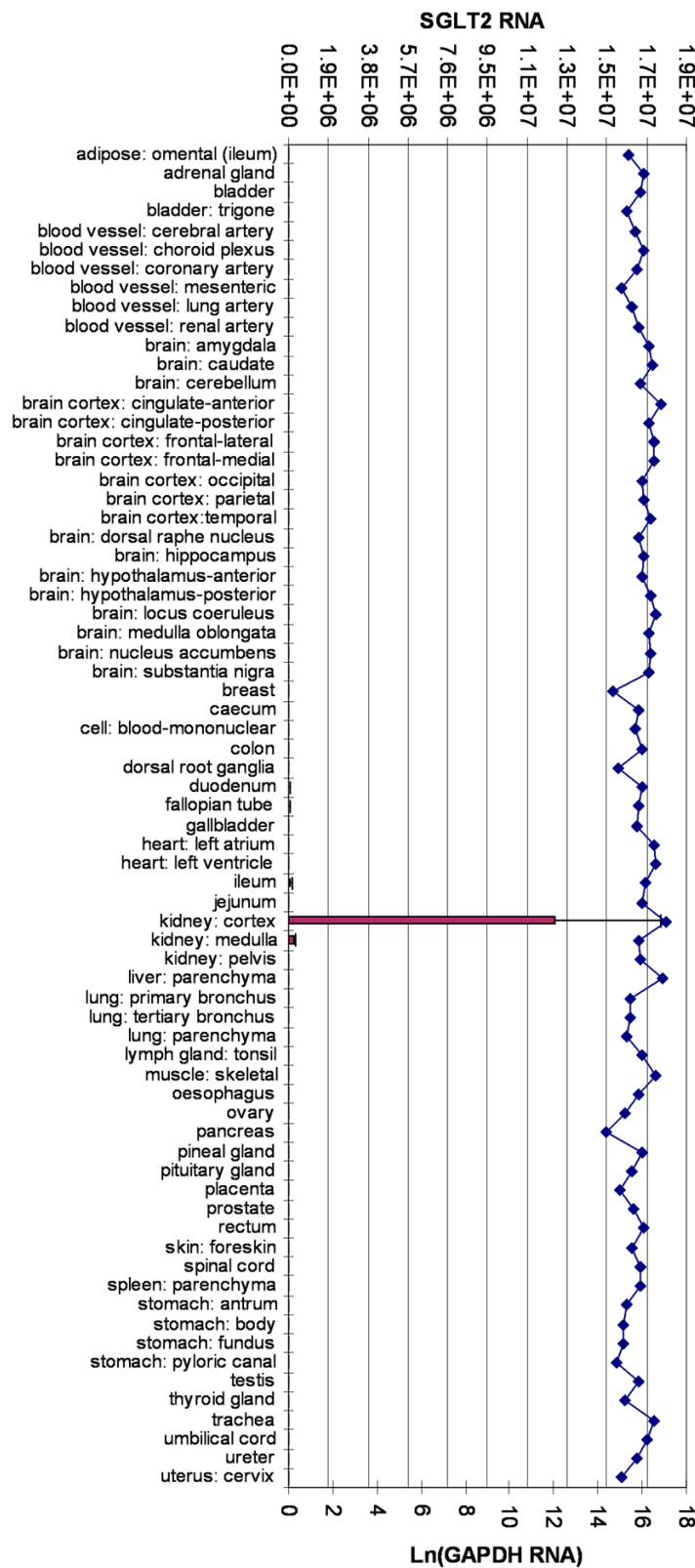
The numbers of SGLT1 and SGLT2 molecules per  $\mu\text{g}$  of RNA in selected human tissues are shown in the table. SGLT2-e1,2: amplicon spans exon 1 and 2 of the SGLT2 gene; SGLT2-e4,5: amplicon spans exon 4 and 5 of the SGLT2 gene; SGLT2-e6,7: amplicon spans exon 6 and 7 of the SGLT2 gene; SGLT2-e9,10: amplicon spans exon 9 and 10 of the SGLT2 gene; SGLT2-e13: amplicon is located in exon 13 of the SGLT2 gene.



**Additional Figure S1.** Expression of sodium glucose cotransporter protein SGLT2 in human tissues measured by primer/probe set on exons 6-7. Copies of SGLT2 transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 6-7 of SGLT2 (SGLT2-e6,7).

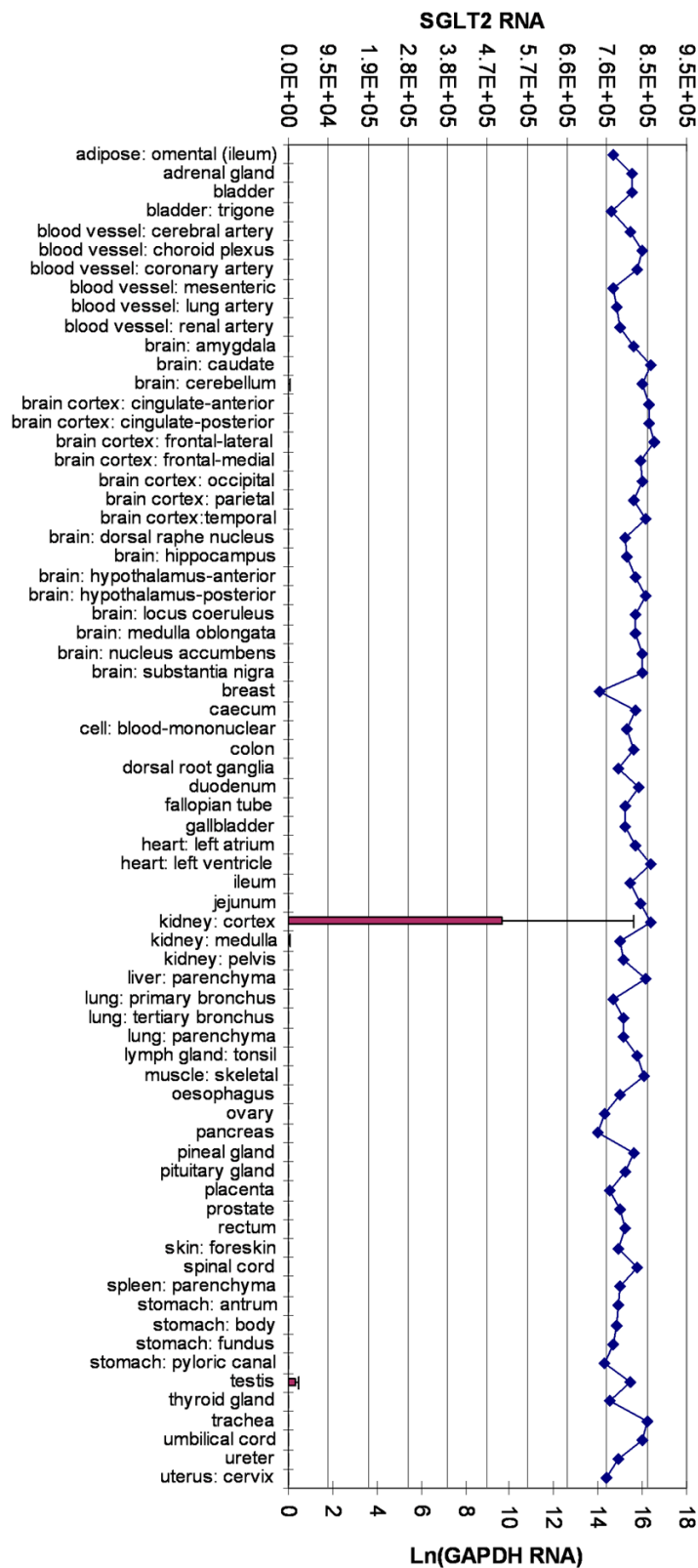


**Additional Figure S2.** Expression of sodium glucose cotransporter protein SGLT2 in human tissues measured by primer/probe set on exons 1-2. Copies of SGLT2 transcript per 1 µg of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 1-2 of SGLT2 (SGLT2-c1,2).

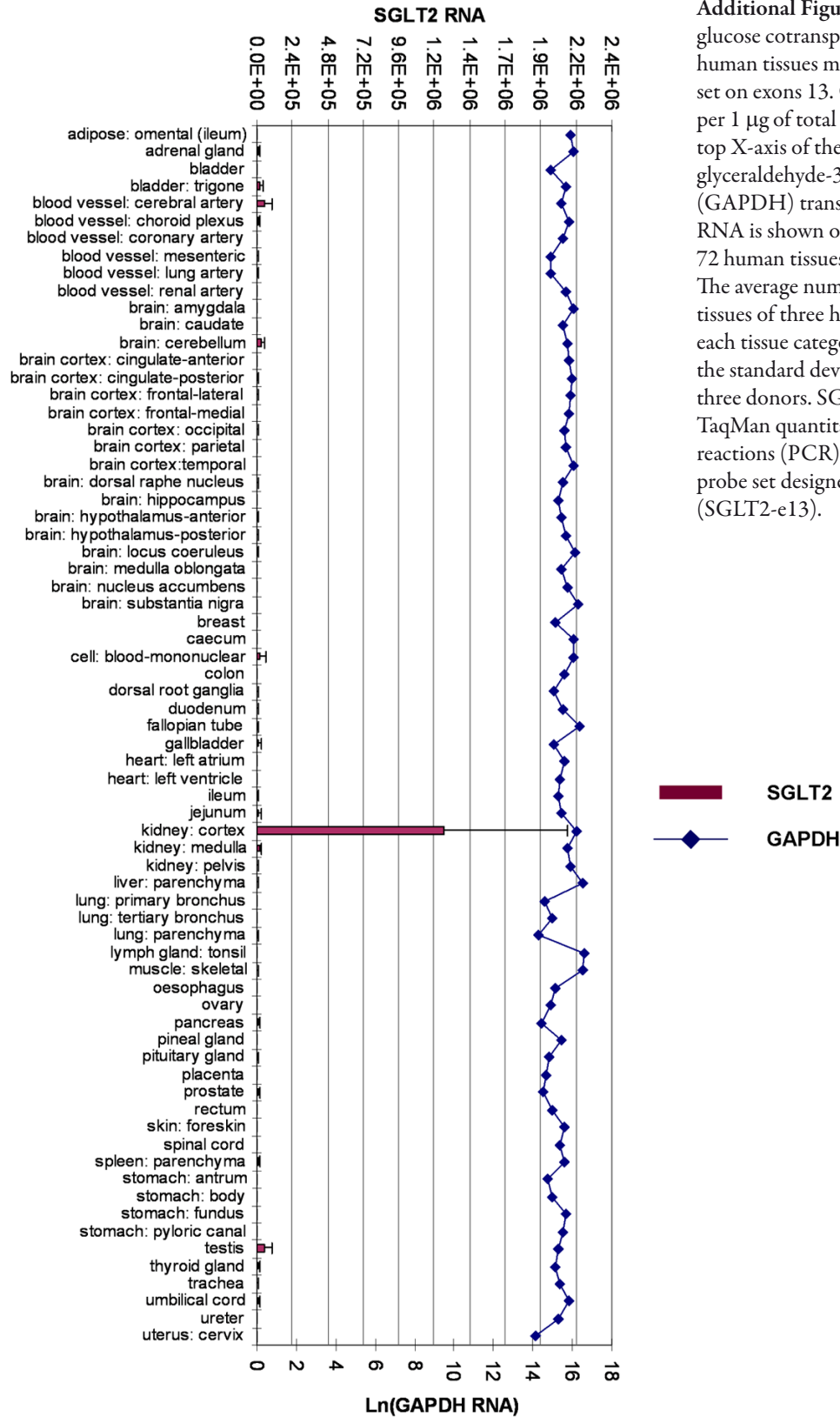


**Additional Figure S3.** Expression of sodium glucose cotransporter protein SGLT2 in human tissues measured by primer/probe set on exons 4-5. Copies of SGLT2 transcript per 1 µg of total RNA are shown on the top X-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 4-5 of SGLT2 (SGLT2-e4,5).

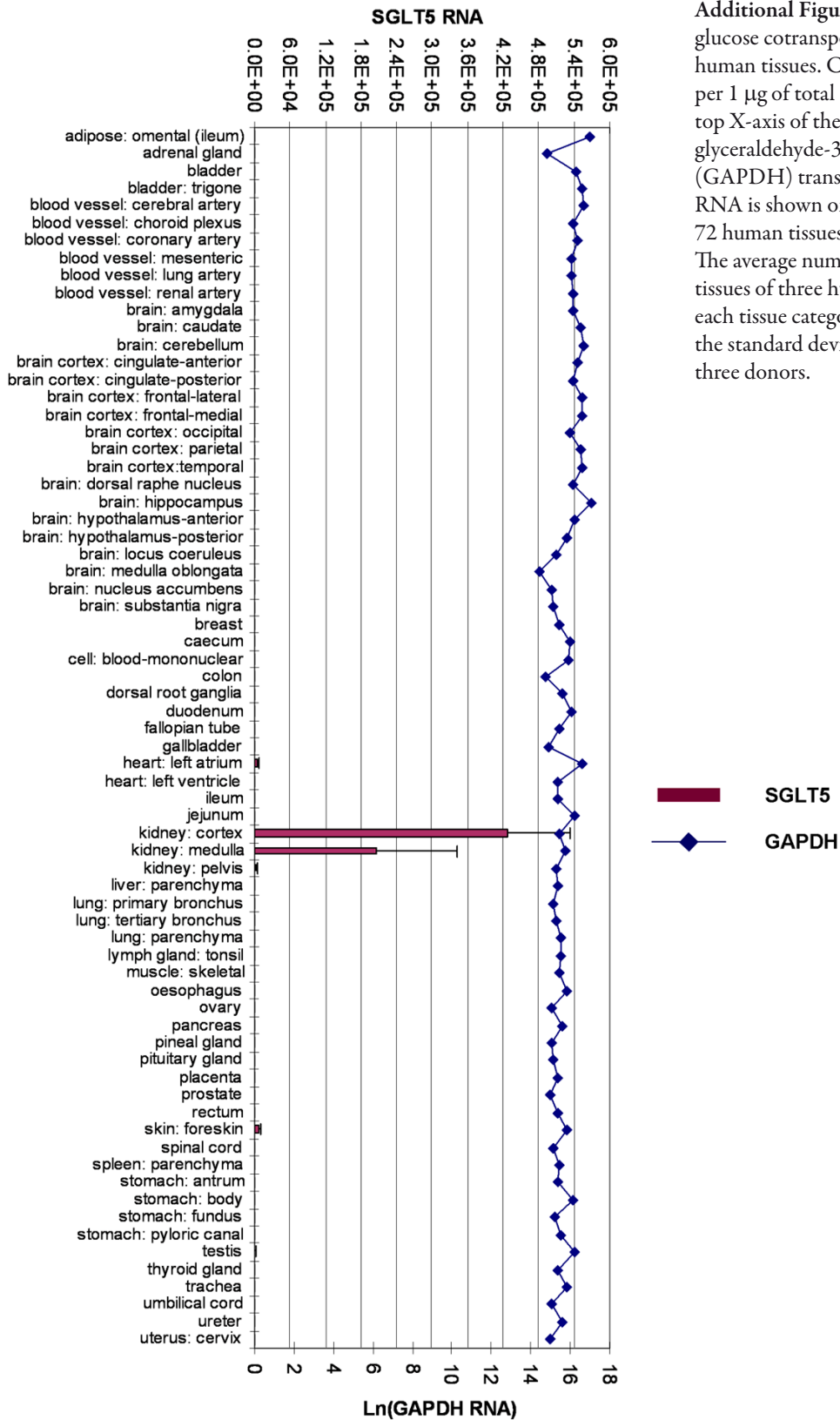




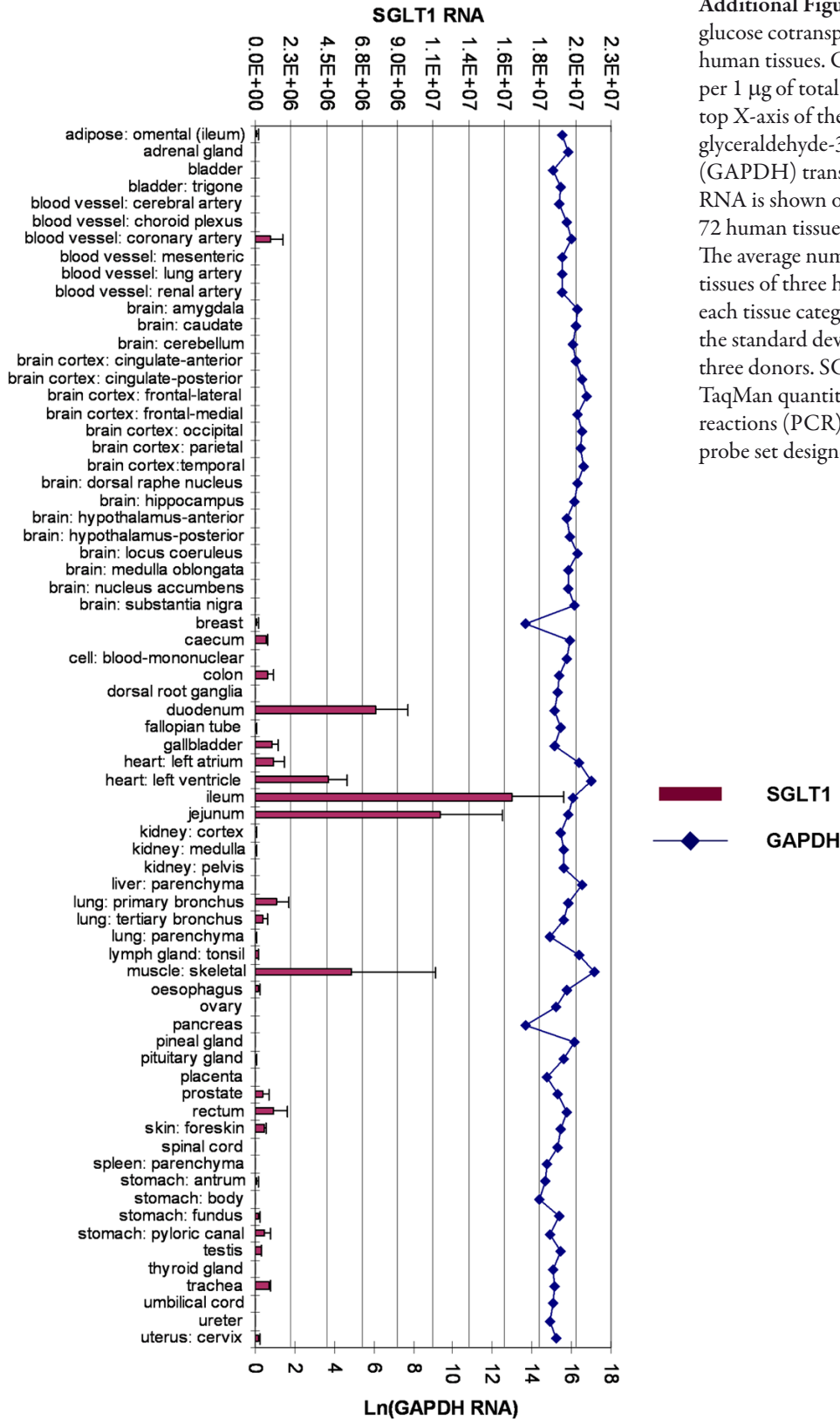
**Additional Figure S4.** Expression of sodium glucose cotransporter protein SGLT2 in human tissues measured by primer/probe set on exons 9-10. Copies of SGLT2 transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 9-10 of SGLT2 (SGLT2-e9,10).



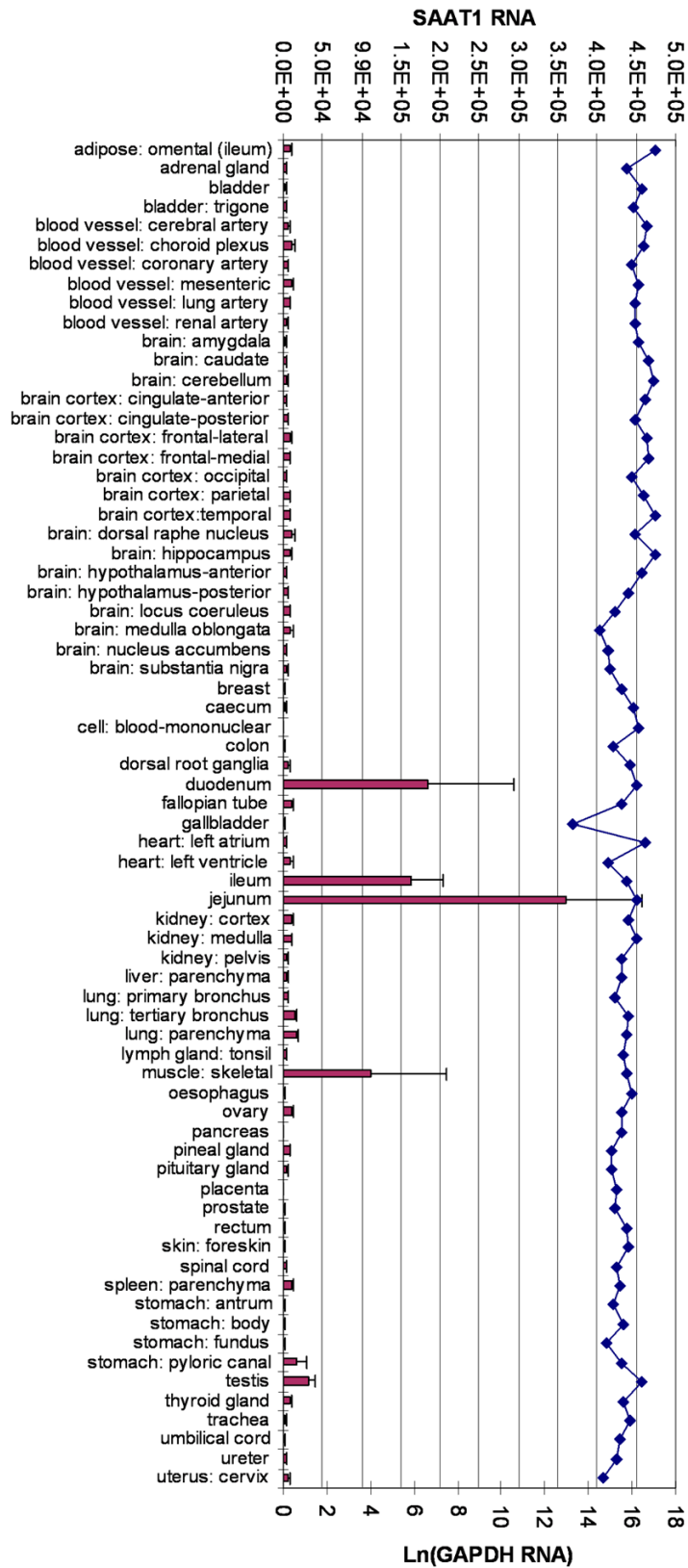
**Additional Figure S5.** Expression of sodium glucose cotransporter protein SGLT2 in human tissues measured by primer/probe set on exons 13. Copies of SGLT2 transcript per 1 µg of total RNA are shown on the top X-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT2 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 13 of SGLT2 (SGLT2-e13).



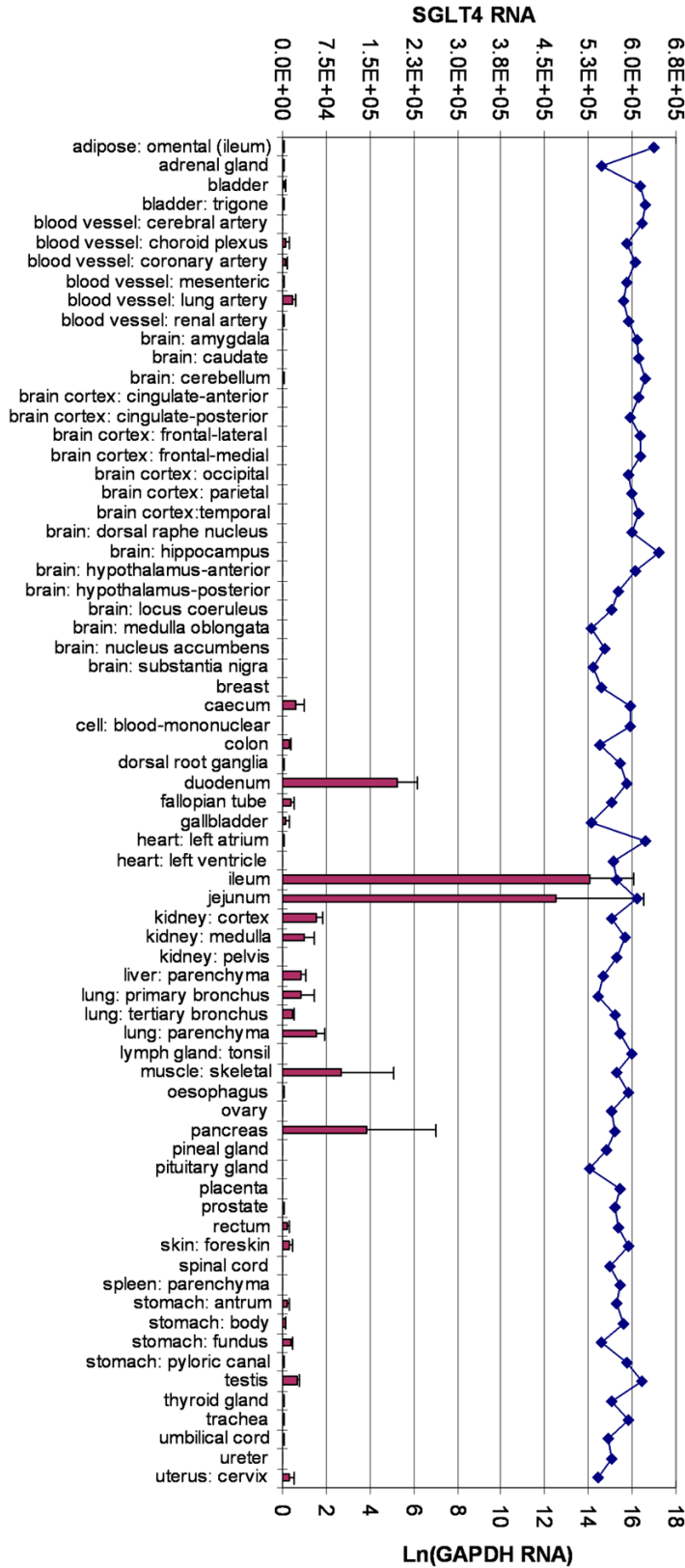
**Additional Figure S6.** Expression of sodium glucose cotransporter protein SGLT5 in human tissues. Copies of SGLT5 transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



**Additional Figure S7.** Expression of sodium glucose cotransporter protein SGLT1 in human tissues. Copies of SGLT transcript per 1 µg of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT1 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 4-5 of SGLT1.

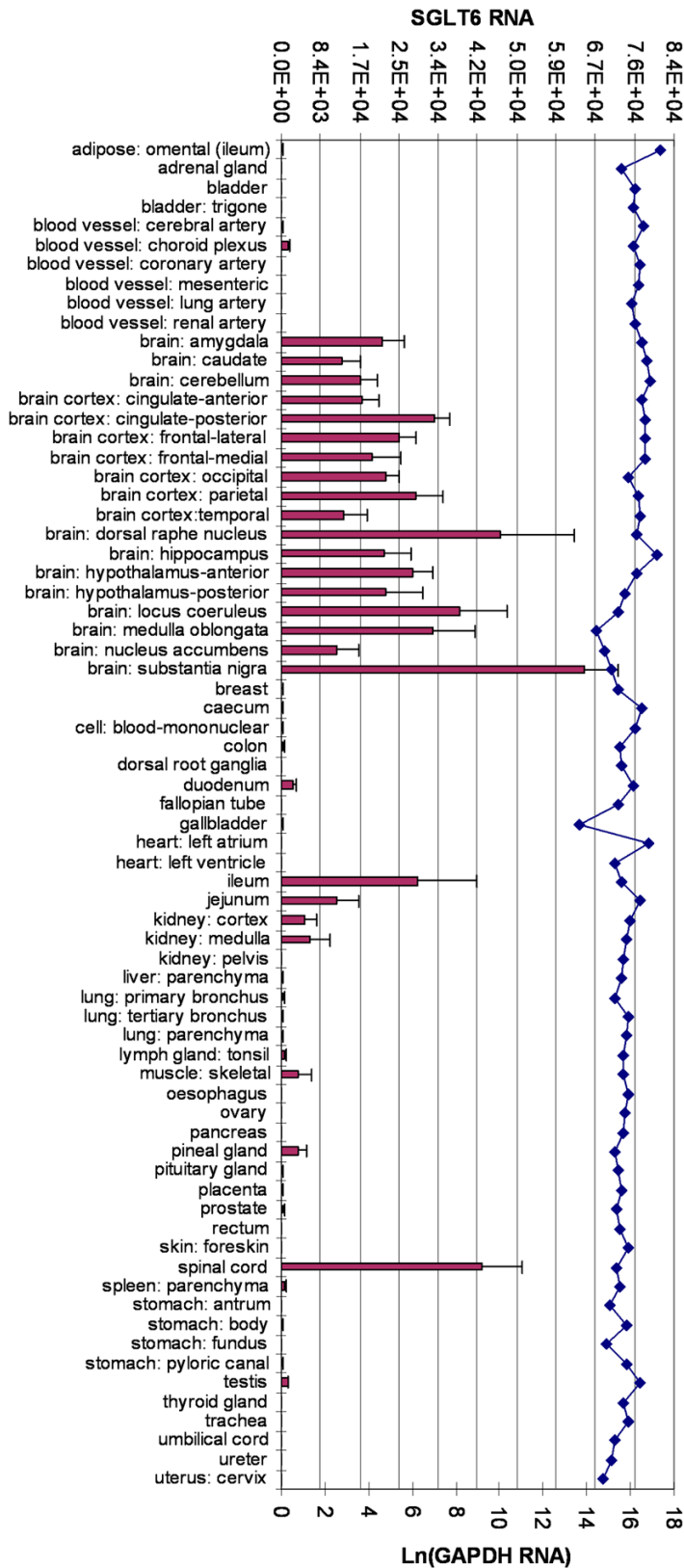


**Additional Figure S8.** Expression of sodium-dependent amino acid transporter (SAAT1) in human tissues. Copies of sodium glucose cotransporter protein (SGLT) transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SAAT1 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 2-3 of SAAT1.

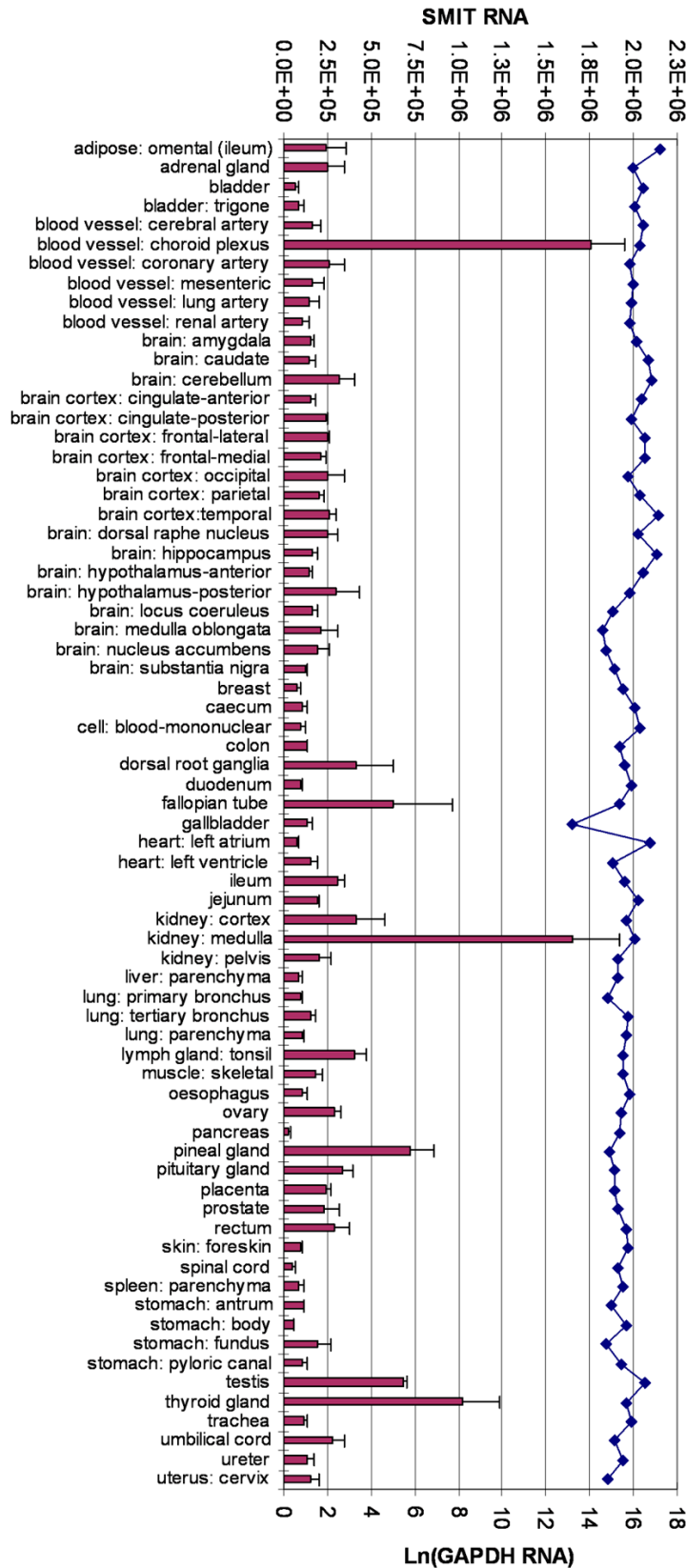


**Additional Figure S9.** Expression of sodium glucose cotransporter (SGLT4) in human tissues. Copies of SGLT transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors. SGLT4 tissue expression from TaqMan quantitative polymerase chain reactions (PCR) performed using primer/probe set designed in exon 11 of SGLT4.

■ SGLT4  
◆ GAPDH

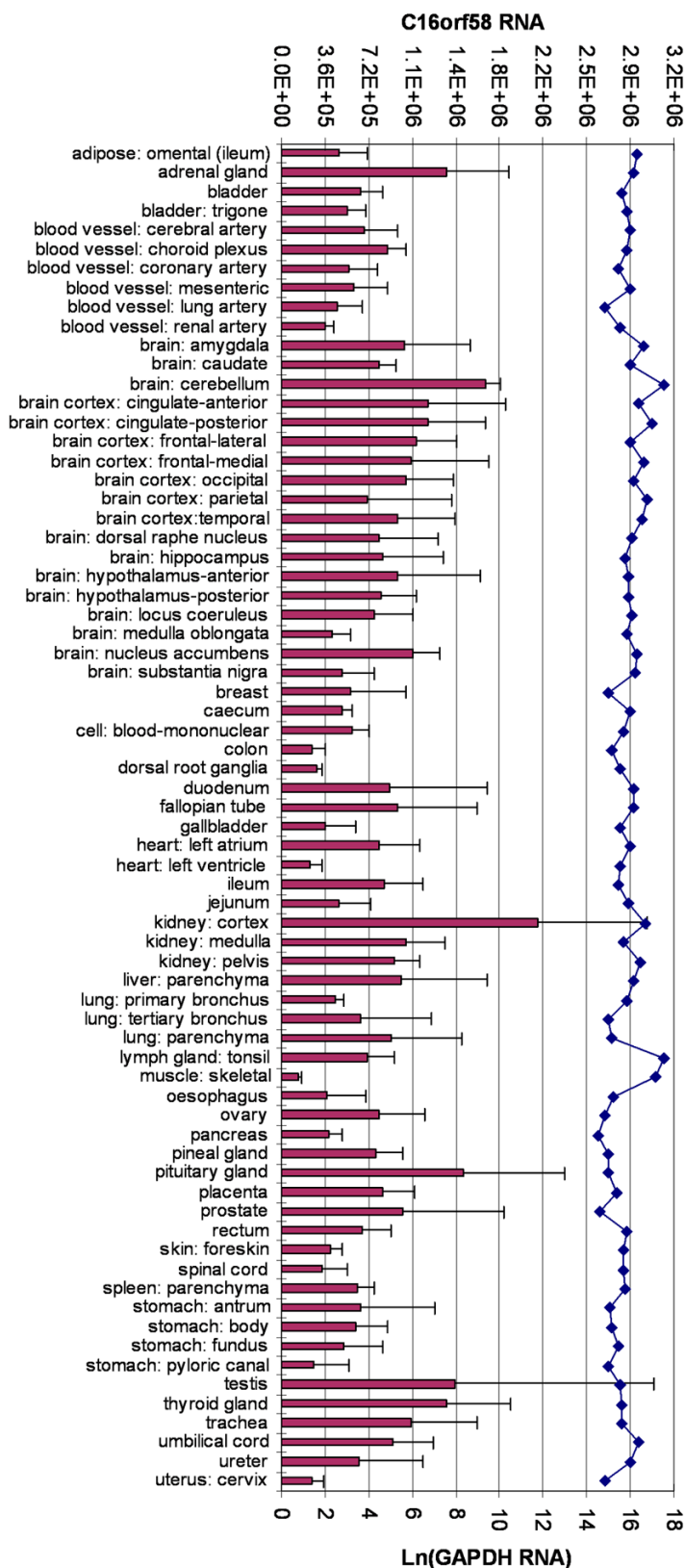


**Additional Figure S10.** Expression of sodium glucose cotransporter protein SGLT6 in human tissues. Copies of SGLT6 transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



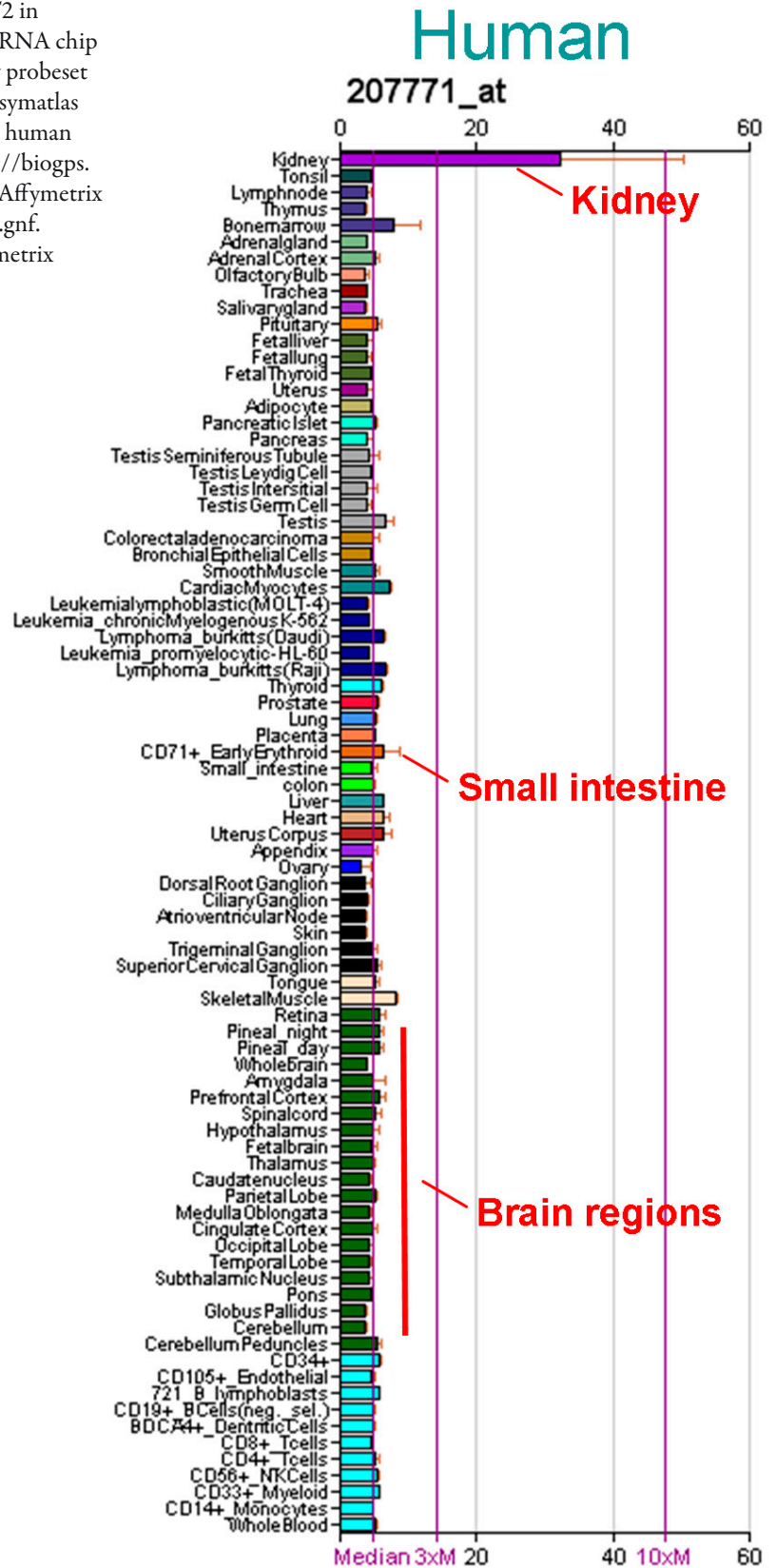
**Additional Figure S11.** Expression of sodium myo-inositol cotransporter (SMIT) in human tissues. Copies of SMIT transcript per 1  $\mu$ g of total RNA are shown on the top X-axis of the graph. Log<sub>2</sub> value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1  $\mu$ g of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.



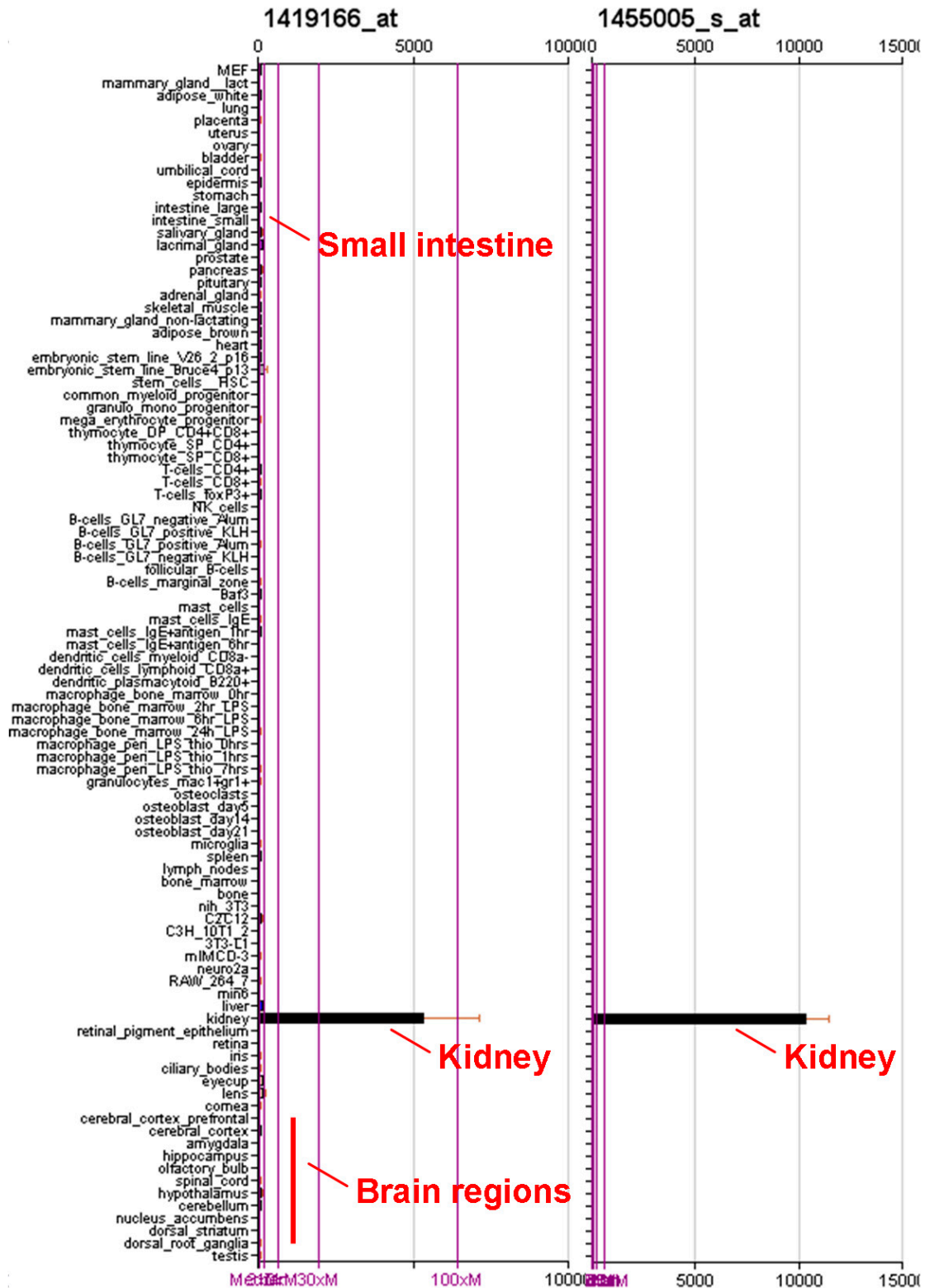


**Additional Figure S12.** Expression of C16orf58 in human tissues. Copies of C16orf58 transcript per 1 µg of total RNA are shown on the top X-axis of the graph. Log2 value of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript per 1 µg of total RNA is shown on the bottom X-axis. All 72 human tissues are shown on the Y-axis. The average number of transcripts from tissues of three human donors is shown in each tissue category. The error bar represents the standard deviation of transcripts of the three donors.

**Additional Figure S13.** Expression of SGLT2 in human, mouse and rat tissues by Affymetrix RNA chip hybridization. SGLT2 gene is represented by probe set 207771\_at (<http://biogps.gnf.org/?referer=symatlas#goto=genereport&cid=6524>) in Affymetrix human chips, 1419166\_at and 1455005\_s\_at (<http://biogps.gnf.org/#goto=genereport&cid=246787>) in Affymetrix mouse chips and U29881\_at (<http://biogps.gnf.org/#goto=genereport&cid=64522>) in Affymetrix rat chips.



# Mouse



# Rat

U29881\_at

