Molecular Genetics and Metabolism Reports 22 (2020) 100568

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

journal homepage: www.elsevier.com/locate/ymgmr

# Molecular analysis of the human placental cysteine dioxygenase type 1 gene

Paul A. Dawson[a,b,*], Shalini J. Weerasekera[a], Ranita J. Atcheson[a], Sarah A. Twomey[b], David G. Simmons[b]

[a] Developmental Disorders Group, Mater Research Institute, The University of Queensland, Translational Research Institute, 37 Kent St, Woolloongabba, QLD 4102, Australia
[b] School of Biomedical Sciences, The University of Queensland, St. Lucia, Australia

ARTICLE INFO

*Keywords:*
Sulfate
Isoform
Splice variant
Placenta
Mutation

ABSTRACT

Sulfate is essential for healthy fetal growth and development. Cysteine dioxygenase type 1 (CDO1) plays an important role in the catabolism of cysteine to sulfate. *Cdo1* knockout mice exhibit severe and lethal fetal phenotypes but the involvement of *CDO1* gene variants in human development is unknown. We searched the NCBI and Ensembl gene databases and identified four alternatively spliced *CDO1* coding mRNA transcripts, as well as 148 validated *CDO1* gene variants, including 138 missense, 6 nonsense, 1 frameshift, 1 in-frame deletion, and 2 splice site variants. *In silico* analyses predicted 68 of the missense variants to be deleterious to CDO1 protein structure and function. We examined the relative abundance of the four *CDO1* coding mRNA transcripts in human term placentas using qRT-PCR. *CDO1* mRNA variant 2 was the most abundant transcript, with intermediate levels of variant 4 and lower levels of variants 1 and 3. Using *in situ* hybridization, we localised *CDO1* mRNA expression to the syncytiotrophoblast layer of human term placenta. To investigate the regulation of *CDO1* gene expression, we analysed the transcriptional activity of the human *CDO1* 5′-flanking region in the JEG-3 placental cell line using luciferase reporter assays. Transcriptional activities were identified in the regions −5 to −269 and − 269 to −1200 nucleotides upstream of the *CDO1* transcription initiation site. Mutational analyses of a single nucleotide polymorphism -289C > G that is common in the general population (allele frequency = 0.37) and a putative transcription factor binding motif (CCAAT enhancer binding protein beta) did not alter transcriptional activity of the *CDO1* 5′-flanking region. Collectively, this study provides an overview and analysis of human *CDO1* for future investigations of this gene in human health.

## 1. Introduction

Sulfate is an obligate nutrient for numerous metabolic and cellular processes during fetal growth and development [1]. During pregnancy in mice, sulfate is supplied from maternal circulation to the fetus *via* the placenta [2]. Our previous studies showed that either reduced levels of sulfate in maternal circulation during pregnancy or disruption of sulfate transport through the placental syncytiotrophoblast layer leads to fetal demise in mice [3,4]. Remarkably, little is known about the physiology of maintaining sulfate supply to the developing human fetus, or the clinical consequences of human fetal sulfate deficiency.

In adults and children, approximately one third of sulfate requirements are obtained from the diet and the remaining two thirds are obtained from the catabolism of the sulfur-containing amino acids methionine and cysteine [5]. A higher proportion of absorbed free inorganic sulfate ($SO_4^{2-}$) from the diet, together with increased renal reabsorption of filtered sulfate, helps to maintain body sulfate homeostasis when dietary methionine and cysteine intake is reduced (*i.e.* low protein) [5]. Methionine is converted to cysteine *via* the transsulfuration pathway, and cysteine is oxidised to sulfate *via* a major pathway involving cysteine dioxygenase type 1 (CDO1, EC 1.13.11.20) [6]. Human CDO1 is abundantly expressed in the adult liver, whereas negligible levels were found in first-trimester fetal liver [7]. The gestational age when CDO1 expression increases in the human fetus is not known but this most likely occurs in late gestation, which was found to be the case in fetal mice [8]. The expression of CDO1 from late gestation may be related to the absence of methionine to cysteine conversion in the human fetus because cystathionine γ-lyase, the last enzyme of the transsulfuration pathway, is not typically expressed until the early neonatal period [9]. Early studies also reported abundant expression of *CDO1* mRNA in human term placenta [10]. The potential role of placental CDO1 in sulfate generation and supply to the fetus has not been considered.

As an initial step towards understanding the potential role of CDO1

* Corresponding author at: Mater Research Institute, University of Queensland, Level 4 Translational Research Institute, Woolloongabba, QLD 4102, Australia.
*E-mail address:* paul.dawson@mater.uq.edu.au (P.A. Dawson).

https://doi.org/10.1016/j.ymgmr.2020.100568
Received 25 November 2019; Received in revised form 14 January 2020; Accepted 14 January 2020

in contributing to sulfate homeostasis in the human placenta and fetus, we provide an update on: *CDO1* gene structure and its mRNA variants; CDO1 protein isoforms; *CDO1* gene variants and their predicted impact on CDO1 protein function; and identify the cellular distribution of *CDO1* mRNA expression in human placenta, as well as the minimal promotor sequence of the human *CDO1* 5′-flanking region that confers transcriptional activity.

## 2. Materials and methods

### 2.1. CDO1 gene, cDNA, protein and 5′-flanking sequences

We searched the NCBI Gene, Nucleotide, Protein, UniGene and SNP databases (https://www.ncbi.nlm.nih.gov/) using the term "*CDO1*" and "*Homo sapiens*" within the date range 19 to 22 January 2019. For this study, we used the reference *CDO1* gene and 5′-flanking region (NC_000005.10), *CDO1* mRNA (NM_001323565.1, NM_001801.2, NM_001323566.1, NM_001323567.1) and CDO1 protein (NP_001310494.1, NP_001792.2, NP_001310495.1, NP_001310496.1) sequences. Validated *CDO1* gene variants were obtained from the Ensembl gene browser database (http://asia.ensembl.org/index.html) within the date range 24 to 30 July 2019. The predicted impact of each missense variant on CDO1 protein function was obtained using *Sorting Intolerant From Tolerant* (SIFT), *Polymorphism Phenotyping* (PolyPhen), *Combined Annotation Dependent Depletion* (CADD), *Rare Exome Variant Ensemble Learner* (REVEL), Method*for deleterious missense mutations using Logistic Regression* (MetaLR), and *Mutation Assessor* (MA) scores from the Ensembl Variant Effect Predictor (http://www.ensembl.org/Tools/VEP). Amino acid sequences of all four CDO1 protein isoforms were aligned using ClustalW software [11]. A phylogenetic tree of 20 species, with *CDO1* orthologues reported in the NCBI database, was generated using the Interactive Tree of Life program (http://itol.embl.de/). Putative transcription factor binding motifs within the first 1500 nucleotides of the 5′-flanking region of *CDO1* were identified using MatInspector software [12] and then a multiple species alignment of those motifs was generated using the DiAlign TF program (http://www.genomatix.de).

### 2.2. Placental tissues, RNA isolation, PCR and in situ hybridization

The research protocol was approved by the Mater Human Research Ethics Committee. Human placentae ($n$ = 10 male and 6 female babies) were obtained from uncomplicated pregnancies ≥37 weeks gestation at elective caesarean section between 0900 and 1200 h, and sampled within 20 min of delivery. Placental weights and neonatal birth weights were within normal ranges [13,14]. Total RNA was isolated from placentae using previously described methods [15]. Four μg RNA was reverse transcribed by using random hexamers and an Omniscript RT kit (Qiagen) as recommended by the manufacturer. PCR was performed as previously described [16] using 200 nM forward and reverse primers (Table 1) in a Rotor-Gene 6000 thermal cycler (Corbett Research, Sydney, Australia). *CDO1* variant mRNA-specific primers were used to determine the relative abundance of the 4 coding *CDO1* mRNA variants V1-V4 (Table 1). The thermal cycling protocol was: 50 °C for 2 min; 94 °C for 2 min; 45 cycles of 94 °C for 1 s, 60 °C for 10 s, and 72 °C for 15 s. RNA expression levels and absolute threshold cycle values (Ct values) of each gene were normalized to those of *GAPDH* RNA with the Rotor-Gene 6000 series software (Corbett Research). Amplification specificity was confirmed by melting curve analysis. For *in situ* hybridization, we synthesized digoxigenin (DIG) labelled probes as previously described [17], using *CDO1* forward (P17) and reverse (P18) primers (Table 1), which can bind to all 4 coding *CDO1* variant mRNAs. Tissues were dissected into 4% paraformaldehyde in PBS and fixed overnight at 4 °C, embedded in paraffin, sectioned at 7 μm, probed, washed and then processed for antibody detection of DIG-labelled probes and colour development as previously described [18].

### 2.3. Sub-cloning and mutagenesis of the CDO1 5′-flanking region

Fragments of the *CDO1* 5′-flanking region with 5′-ends at −1200, −269 and −5, and with a common 3′-end at +311 (+1 is the transcription start site of *CDO1* variant 2 mRNA NM_001801.2), were amplified by PCR using 1 μM primers (Table 1), and 1.25 U LA TAQ polymerase (TakaRa). The thermal cycling protocol was: 95 °C for 1 min; 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 3 min, followed by 1 cycle of 72 °C for 10 min. Each fragment was cloned upstream of a luciferase reporter gene in the multiple cloning site of the pMetLuc2 expression vector (Clontech) using In-Fusion reagent as described by the manufacturer (Clontech). Sequence variants were introduced into the cloned −1200 *CDO1* 5′-flanking region by PCR using 1 μM forward and reverse primers (Table 1), 6% DMSO, 1 U Phusion DNA Polymerase and thermal cycling parameters: 95 °C for 2 min; 35 cycles of 95 °C for 20 s, 59 °C for 20 s, and 72 °C for 3 min, followed by 1 cycle of 72 °C for 5 min. These PCR products were then treated with In-Fusion regent (Clontech) to generate circularized vectors, as previously described [19]. Nucleotide sequences of all cloned *CDO1* sequences were verified by DNA sequence analysis using 9.6 pmol forward and reverse primers (Table 1) as previously described [19].

### 2.4. Cell culture and luciferase assays

The JEG-3 placental cell line was cultured in DMEM medium containing L-glutamine and glucose (Invitrogen) supplemented with 10% heat-inactivated fetal calf serum, 1% penicillin–streptomycin and 0.1% plasmocin (InvivoGen). Cultures were maintained at 37 °C in a humidified atmosphere of 5% $CO_2$ and 95% air. JEG-3 cells were transiently co-transfected using Lipofectamine$_{2000}$ (Invitrogen) with pMetLuc2 containing individual cloned *CDO1* fragments or pMetLuc2 that lacks cloned fragments (negative control), and pSEAP2 alkaline phosphatase expression vector used to normalise luciferase activity. Cells (60% confluent in 48 well plates) were incubated with 2 μl Lipofectamine$_{2000}$ and plasmids (3.2 ng each pMetLuc2 and pSEAP2) in DMEM containing 10% fetal calf serum (total volume 100 μl) for 7 h, and then the media was replaced with 800 μl OptiMEM® (Invitrogen) containing 10% fetal calf serum for an additional 70 h at 37 °C. Luciferase and alkaline phosphatase activities in the cell culture media were assayed using protocols and reagents (Ready-To-Glow™ Dual secreted reporter assay) purchased from Clontech, and measured using a PolarStar Omega plate reader (BMG Labtech).

### 2.5. Statistical analyses

Statistical significance of *CDO1* mRNA levels between placentae from male and female babies was evaluated using the unpaired 2-tailed Student's $t$-test. The statistical significance of the differences of luciferase activities between each clone and control vector was evaluated using a one-way ANOVA, followed by a Dunnett multiple comparisons test. $p < 0.05$ considered significant.

## 3. Results and discussion

### 3.1. CDO1 gene, mRNA, protein and variant sequences

The human *CDO1* gene contains 8 exons, spanning approximately 12 kb at chromosomal location 5q22.3 (Fig. 1A). Earlier investigations reported a *CDO1* mRNA of approximately 1.5 kb encoding a 200 amino acid CDO1 protein [20], corresponding to transcript variant 2 (NM_001801.2) and protein isoform 2 (NP_001792.2) sequences in the current NCBI database. More recently, additional mRNA variants have been added to the NCBI database, including a total of 4 protein coding variants V1 to V4 (Fig. 1B), and 4 non-coding variants (not shown): V5 (NR_136618.1), V6 (NR_136619.1), V7 (NR_136620.1) and V8 (NR_136621.1). These eight mRNA variants arise from alternative

**Table 1**
Primers used for PCR, mutagenesis, DNA sequencing, and *in situ* probes.

| Primer | [g]Direction | Sequence (5′ to 3′) | | [h]Primer location |
|---|---|---|---|---|
| [a]*Primers used for PCR-amplifying CDO1 5′-flanking regions* | | | | |
| P1 | F | CCCGGGATCCACCGGTACATCACCAAGGCTGGCTTC | | −1200 |
| P2 | F | CCCGGGATCCACCGGTTACGTCCCAGCGTCGCGAAC | | −269 |
| P3 | F | CCCGGGATCCACCGGTGGTACATTCCTAGTGACTCC | | −5 |
| P4 | R | CATGGTGGCGACCGGTCTCGTGGGGAGCTGGCTG | | +311 |
| [b]*Primers used for site directed mutagenesis* | | | | |
| P5 | F | TCAGTCCCCGCAGCCATGTCCTCCGACCCTTTTTG | | −306 |
| P6 | R | GGCTGCGGGGACTGACGCTGAGTAAAGGAGGAAAA | | −292 |
| P7 | F | GCATTTCTTCCACCTTTGGCATCTTTCTCATTTAAA | | −915 |
| P8 | R | AGGTGGAAGAAATGCAAGTGTAGATTTCTTCCTTC | | −901 |
| [c]*Primers used for DNA sequencing* | | | | |
| P9 | F | CTGTGGATAACCGTATTAC | | |
| P10 | R | CAGATGTCGATGTTGGGG | | |
| [d]*Primers used for quantitative PCR* | | | | |
| P11 | F | GTTCGACCAGTACAGTCGTG | V1 c.156 | |
| [e]P12 | F | GGTGAAGGACATGGCAGCAG | V2 c.232 | |
| P13 | F | GGGGTGAAGGACATGGCAGT | V3 c.230 | |
| P14 | F | TGCCAGGGGCCTGGGGGTAT | V4 c.-62 | |
| P15 | R | AGCATCTTCAGAAAGCAGTG | V1 c.350, | V4 c.77 |
| P16 | R | GTAAGCCAATGGAATCATTG | V2 c.418, | V3 c.415 |
| [f]*Primers used for PCR-amplifying in situ probes* | | | | |
| P17 | F | aattaaccctcactaaagggACCTGCTGTGAGCCTTCACT | V1 c.815, | V2 c.758 |
| | | | V3 c.755, | V4 c.351 |
| P18 | R | taatacgactcactatagggCCATGGGACACAGACATGAG | V1 c.*1605, | V2 c.*1548 |
| | | | V3 c.*1545, | V4 c.*1141 |

[a] Underlined sequence homologous to the pMetLuc2 vector to allow recombination cloning.
[b] Underlined nucleotides introduce the -289C > G SNP (P5) and mutate the core sequence of the CEBPB motif (P7).
[c] Primers located in the *pMetLuc2* vector upstream (P9) or downstream (P10) of the multiple cloning site.
[d] *CDO1* variant mRNA-specific primers enabled amplification of V1 (P11 + P15), V3 (P13 + P16) and V4 (P14 + P15).
[e] Since primer combination P12 + P16 amplified variants 1, 2 and 4, we calculated variant 2 mRNA abundance as [(amplicon abundance using P12 + P16) – (V1 + V4 amplicon abundance)].
[f] Lower case represents *T3* (P17) and *T7* (P18) RNA polymerase binding sites, and uppercase represents *CDO1* sequences in exon 7 (P17) and the 3′-UTR of exon 8 (P18).
[g] F, forward primer; R, reverse primer.
[h] Number denotes the position of the first 5′-nucleotide in the *CDO1* 5′-flanking region (P1-P8) or *CDO1* variant V1, V2, V3 or V4 mRNAs (P11-P18).

splicing of exons and differential use of exon-intron junction sequences, as well as transcription initiation from 2 sites. Variants 1, 2, 3, 5, 6 and 7 are transcribed from 311 nt upstream of the ATG start codon, whereas variants 4 and 8 have a transcription start site at 260 nt upstream of ATG which was previously identified in an early study that investigated human *CDO1* mRNA using primer extension analysis [10]. Variant 2 and its encoded 200 amino acid isoform 2 are the reference sequences in the current literature. However, the physiological roles of the other 3 coding *CDO1* mRNA variants (Fig. 1B) and CDO1 protein isoforms (Fig. 1C) awaits further investigation.

To date, 148 validated non-synonymous *CDO1* sequence variants have been added to the Ensembl variant database. These include 138 missense, 6 nonsense, 1 frameshift, 1 in-frame deletion and 2 splice site variants (Supplemental Tables 1 and 2). The nonsense and frameshift variants are most likely deleterious to CDO1 function, whereas the impact of missense variants on CDO1 function has not been extensively investigated. Using the SIFT, Poly-Phen, CADD, REVEL, MetaLR and MA tools, 68 missense variants are predicted to be deleterious to CDO1 structure and function (Supplemental Table 1, Fig. 1D). These variants are located throughout the CDO1 protein, with most of the predicted damaging variants clustering in 2 amino acid regions that contain β-sheets 1–5 and 8–13 (Fig. 1D). Substitution of amino acids (Q34P, Y58C, R60G and R60Q, W77C, H88N, C93R, C130Y, Y157S and Y157C, F161V) that directly interact with cysteine substrate, are predicted to be damaging to CDO1 protein function (Fig. 1D), suggesting the likely importance of these cysteine-binding residues for CDO1 function. In addition, the missense variant H88N that alters an $Fe^{2+}$ interacting amino acid is predicted to be detrimental, suggesting the importance of this combined cysteine and $Fe^{2+}$ ion-interacting amino acid for CDO1 function.

All of the non-synonymous *CDO1* variants are relatively rare: 3 have an allelic frequency $> 1 \times 10^{-4}$ (intron 3 splice donor variant rs201589147 $\sim 1.0 \times 10^{-3}$, and missense variants D64Y $\sim 2.6 \times 10^{-4}$ and S42R $\sim 1.0 \times 10^{-3}$), whereas the other 135 variants have allelic frequencies $< 1 \times 10^{-4}$ (Supplemental Tables 1 and 2). The low abundance of more common *CDO1* missense variants (allelic frequency > 0.001) in the general population may suggest a harmful physiological effect from disrupting the CDO1 protein sequence. This was found to be the case for the *Cdo1* knock-out mouse which has severe developmental defects and high postnatal mortality [6]. Taken together, these observations may explain why non-synonymous *CDO1* variants have yet to be linked to any human pathology, which has led us to investigate the *CDO1* gene and its mRNA expression as a step towards understanding its role in early development.

### 3.2. CDO1 mRNA expression and localization in placental tissue

Early studies used Northern blot analyses to show abundant *CDO1* mRNA expression in term placental and adult human liver tissue, lower levels in heart and whole brain, and undetectable levels in lung, skeletal muscle and kidney [10]. Using the NCBI database, we confirmed these findings and extend the *CDO1* mRNA expression profile to 11 additional adult tissues, with intermediate *CDO1* mRNA levels in cerebellum and lower levels in prostate, adrenal gland, thyroid, trachea, uterus, thymus, stomach, spleen, salivary gland and small intestine (Fig. 2A). In addition, the NCBI database shows moderate *CDO1* mRNA levels in fetal liver (22 to 40 weeks gestation) and fetal whole brain (20 to 33 weeks gestation). Since these data are derived from pooled fetal

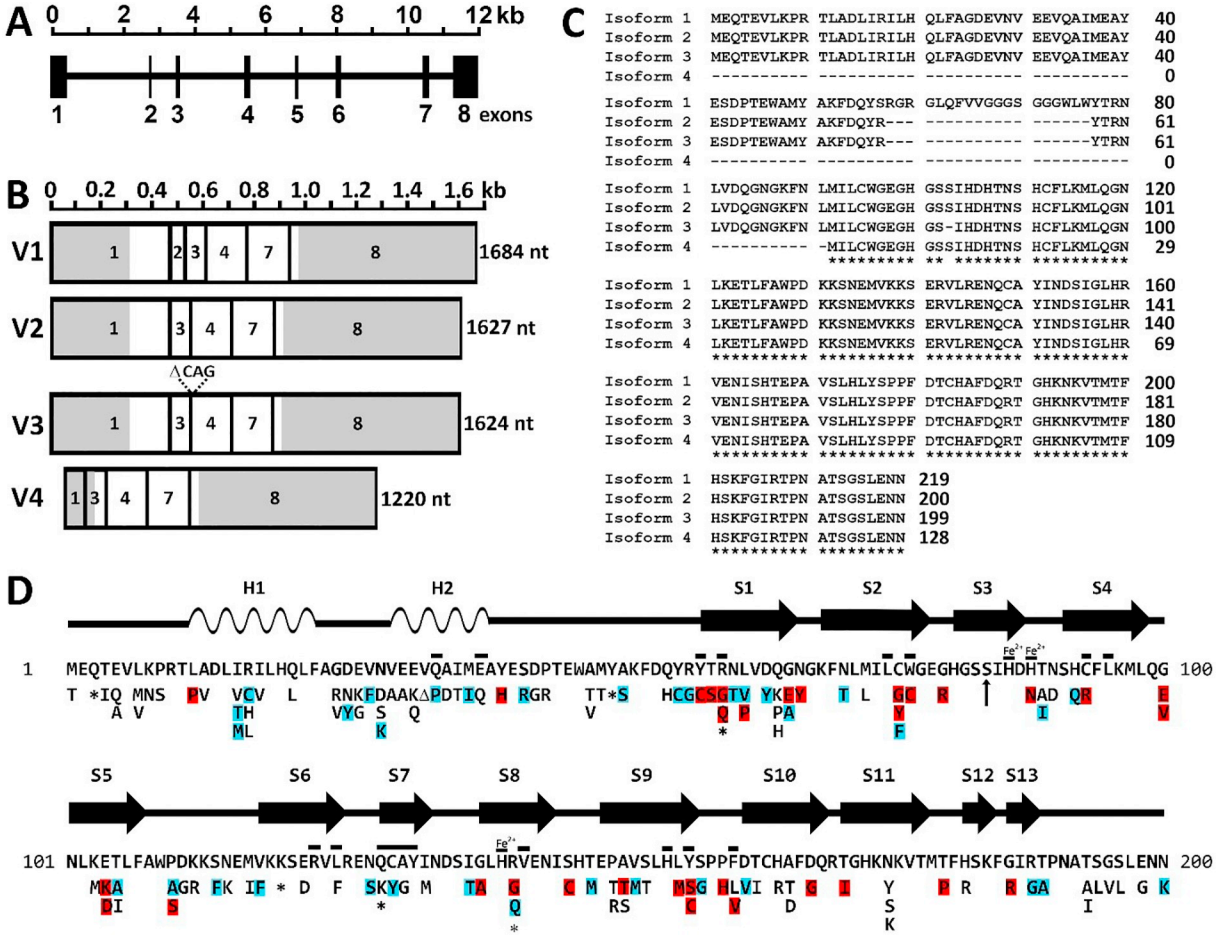


**Fig. 1.** Human *CDO1* gene structure, mRNA variants and protein isoforms. (A) Exon-intron organization showing exons (vertical lines) and introns (horizontal lines) spread over approximately 12 kb. Reference sequence NC_000005.10. (B) Schematic showing exons 1–8 (boxes) and protein coding sequences (white portions) for *CDO1* mRNA variants V1 (NM_001323565.1), V2 (NM_001801.2), V3 (NM_001323566.1) and V4 (NM_001323567.1). (C) Aligned human CDO1 protein isoform 1 (NP_001310494.1), 2 (NP_001792.2), 3 (NP_001310495.1) and 4 (NP_001310496.1) sequences. Alignments were generated using the Clustal W program [11]. *Amino acids present in all 4 isoforms. Sequences accessed from the NCBI from 19 to 22 January 2019. (D) Validated *CDO1* missense (*letters*), nonsense (*) and in-frame single amino acid deletion (Δ) variants (Ensembl database) are shown below the CDO1 isoform 2 protein sequence. Location of each α-helix (H1-H2, *waved lines*) and β-sheet (S1-S13, *horizontal arrows*), as well as the amino acids interacting with cysteine substrate (*bar above CDO1 sequence*) and/or $Fe^{2+}$, in CDO1 isoform 2 protein (PDB ID: 2IC1) from the NCBI structure database (https://www.ncbi.nlm.nih.gov/structure/). Highlighted missense variants are predicted to be detrimental (*red*) or moderately disruptive (*blue*) to the CDO1 protein structure as determined using a combination of tools (SIFT, PolyPhen, CADD, REVEL, MetaLR, MA). Serine 83 (*vertical arrow*) is absent in the alternatively spliced *CDO1* variant 3 mRNA transcript (as indicated by ΔCAG in panel B).

samples from a range of gestational ages, it is not possible to determine when fetal *CDO1* mRNA expression increases in gestation. Our studies in mice showed negligible *Cdo1* mRNA levels in whole fetal tissues at embryonic day 10.5 (E10.5) with levels increasing from E12.5 [8], suggesting an important role of CDO1 in the later gestational stages of fetal development. This is also relevant to the strong expression of *CDO1* mRNA in term placenta (Fig. 2A). In the present study, we localised *CDO1* mRNA to the syncytiotrophoblast layer (Fig. 2B), the site of nutrient and waste exchange between maternal and fetal circulation. However, this finding was unexpected as our previous study localised *Cdo1* mRNA to the decidua but not the syncytiotrophoblast layers of the mouse placenta [8]. The reason for these observed differences between human *CDO1* and mouse *Cdo1* expression in the placenta are unknown but could reflect species differences in the organization of the trophoblast cell types, as well as the much longer gestational length for humans. The next phase of our research will investigate *CDO1* mRNA expression in placentas obtained from preterm births to determine the gestational age when *CDO1* mRNA abundance increases and whether its cellular localization changes during gestation.

As mentioned above, there are 4 protein coding *CDO1* mRNA variant sequences reported in the NCBI database. Using qPCR, we show that variant 2 is the most abundant *CDO1* mRNA transcript in the term placenta, with lower levels of variant 4 and negligible levels of variants 1 and 3 (Fig. 2C), and no sex-specific differences ($p > 0.05$) in the abundance of each variant *CDO1* mRNA (Fig. 2D). The higher abundance of variant 4 *CDO1* mRNA when compared to variants 1 and 3 was unexpected as variant 4 encodes a truncated CDO1 protein (Fig. 1C) that is most likely unstable and inactive. Overall, these findings suggest that the splicing of exon 2 in variant 1, the alternative splicing at the 5′-end of exon 4 (ΔCAG) in variant 3, and the differential splicing of exon 1 to exon 3 in variant 4, are either inefficient when compared to variant 2 and/or that variant 1, 3 and 4 mRNA transcripts are unstable leading to their rapid mRNA decay. Further studies, such as expressing each variant *CDO1* mRNA in cultured cells, are required to confirm whether mRNA stability contributes to the different level of each *CDO1* mRNA transcript. Nonetheless, variant 2 appears to be the predominant *CDO1* mRNA transcript, implying that isoform 2 is the major form of CDO1 protein in the placenta. To further understand the biological relevance of the *CDO1* mRNA transcripts, the next phase of our research will investigate the relative abundance and activity of the placental CDO1 proteins, which is relevant to the post-translational regulation of CDO1 [21,22].

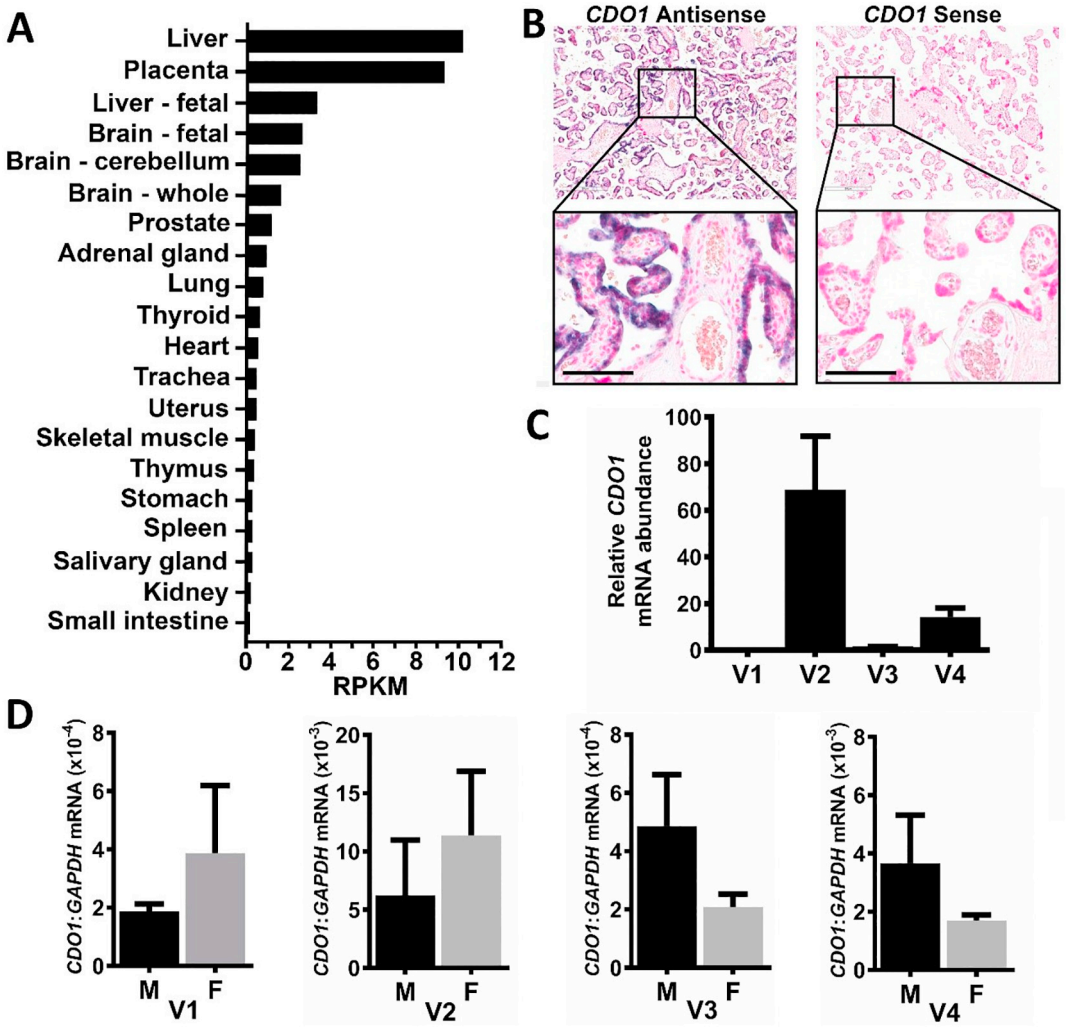


**Fig. 2.** Tissue expression of *CDO1* mRNA. (A) Relative *CDO1* mRNA levels in 20 human tissues using data obtained from the online NCBI database http://www.ncbi.nlm.nih.gov/gene, which was accessed 19 January 2019. RPKM, Reads Per Kilobase of transcript, per Million mapped reads. (B) Cellular localization of *CDO1* mRNA in human term placenta. Low (*top panels*) and high (*bottom panels*) magnification images of placental sections probed with antisense and sense (control) *CDO1* probes. Images show antisense probe staining of *CDO1* mRNA within the syncytiotrophoblast layer. Black scale bar = 100 μM. (C) Relative abundance of each placental *CDO1* mRNA variant (V1-V4) abundance. (D) Comparison of each *CDO1* variant (V1-V4) mRNA abundance in placentas of male (M) and female (F) babies ($p > 0.05$ for each variant). Data are mean ± SEM with $n = 4$–5 males and $n = 3$ females.

### 3.3. Phylogenetic comparison of CDO1 protein and the CDO1 5′-flanking region

Alignment of human CDO1 amino acid sequences with 19 orthologues showed the highest identity (88 to 93%) to hemochorial species, intermediate identity (64 to 90%) with endotheliochorial and epitheliochorial species, and lowest identity (49 to 59%) with non-placental species (Fig. 3A). To identify conserved DNA sequences that are potentially involved in the transcriptional control of human *CDO1*, the first 3500 nucleotides of the *CDO1* 5′-flanking region was compared among all 20 species. Of the 387 putative transcription factor binding motifs in the human *CDO1* 5′-flanking region (data not shown), 6 motifs exhibited similar spatial conservation among placental species as determined by a multiple species alignment (Fig. 3B): Kruppel like factor (KLF1) at position −63 to −81; myeloid zinc finger 1 (MZF1) at position −91 to −101; zinc finger protein 300 (ZNF300) at position −174 to −196; zinc finger protein 239 (ZNF239) at position −272 to −292; distal-less homeobox 1 (DLX1) at position −741 to −759; and CCAAT enhancer binding protein beta (CEBPB) at position −890 to −904 (Fig. 4A). The conserved location of these 6 putative transcription factor binding motifs in the 5′-flanking region of orthologous *CDO1* genes of placental species led us to further investigate these DNA sequences as potential regulators of placental *CDO1* mRNA expression.

### 3.4. Analysis of the human CDO1 5′-flanking region

For the 6 putative motifs in the *CDO1* 5′-flanking region (Fig. 4A) to be involved in regulating placental *CDO1* mRNA expression, the transcription factors that bind to these DNA sequences need to be expressed in the placenta. Indeed, we found this to be the case for CEBPB which is abundantly expressed in human placenta and the JEG-3 placental cell line (Fig. 4B). The other 5 transcription factors have either lower abundance (MZF1 and ZNF300) or undetectable (DLX1, ZNF239 and KLF1) levels in human term placenta (Fig. 4B), suggesting that these transcription factors may possibly not regulate *CDO1* mRNA expression. To further localise the DNA sequences which are important for *CDO1* promoter activity, a series of *CDO1* 5′-flanking regions were cloned upstream of the luciferase gene, transfected into JEG-3 cells and then assayed for luciferase activity (Fig. 4C). The highest luciferase expression was obtained with an upstream end of −1200, when compared to the construct with a 5′-end at −269 which had intermediate expression, and to the construct with a 5′-end at position −5 which had

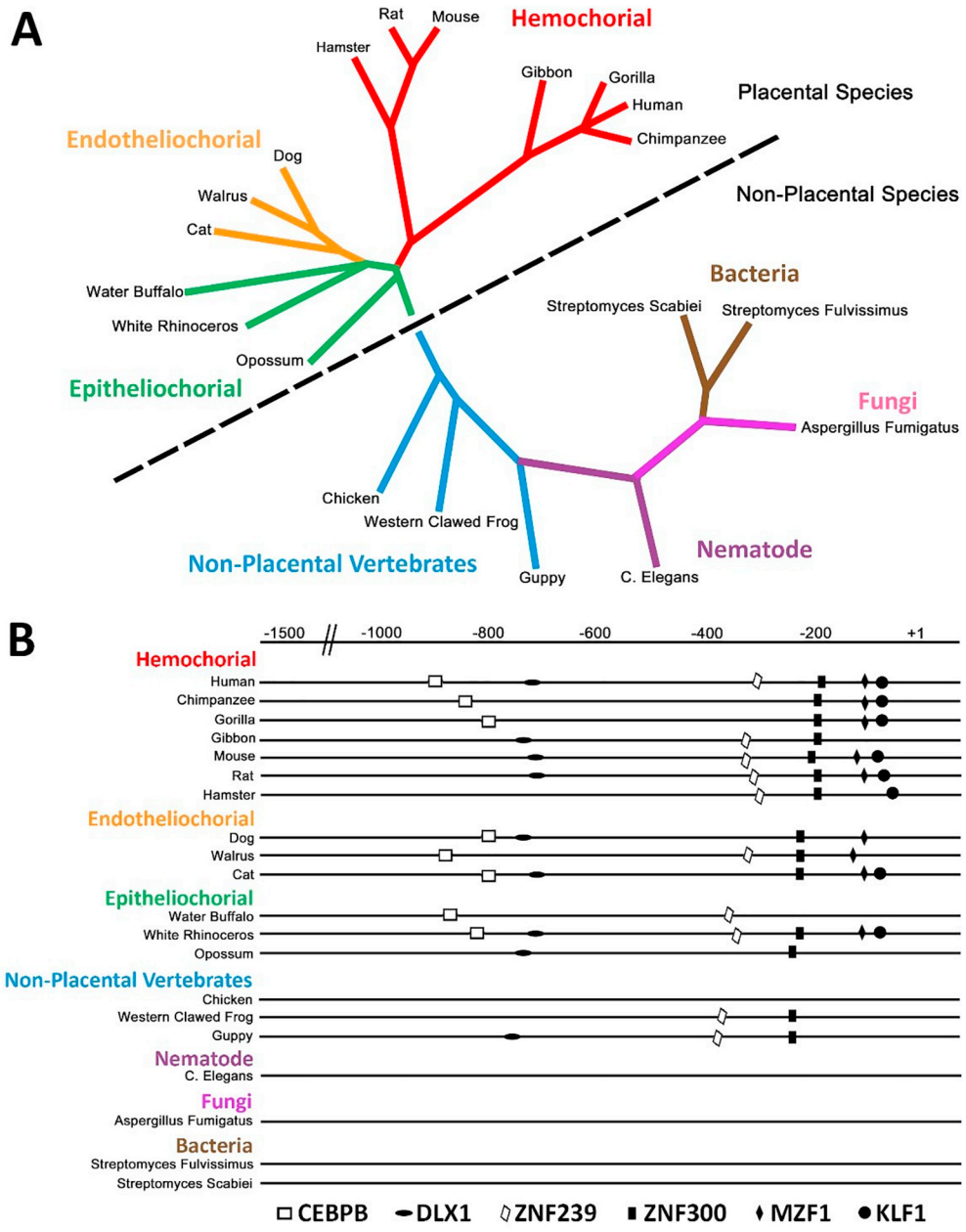


**Fig. 3.** Phylogenetic analysis of *CDO1* and its 5′-flanking region. (A) Phylogenetic tree with distances representing the evolutionary divergence of species that carry a *CDO1* gene. Species above the dotted line represent the placental types: hemochorial (*red*), endotheliochorial (*orange*) and epitheliochorial (*green*). Image adapted from a tree generated online using the Interactive Tree of Life (ITOL, http://itol.embl.de/). (B) Multiple species alignment of conserved putative transcription factor binding motifs in the *CDO1* 5′-flanking region. Thirteen placental species share six highly conserved domains with human *CDO1*: CEBPB, DLX1, ZNF239, ZNF300, MZF1 and KLF1. Scale bar at top is relative to the +1 of the transcription start site.

negligible luciferase expression. This finding suggests the possible presence of positive regulatory elements between regions −269 to −1200, and −5 to −269. Given the abundant expression of CEBPB in human term placenta and the JEG-3 cell line (Fig. 4B), and the location of a putative CEBPB binding motif within the *CDO1* 5′-flanking region that produced the highest luciferase expression, we mutated the core sequence of the CEBPB motif and compared luciferase activities to the control sequence (Fig. 4D). However, luciferase activity from the mutated CEBPB sequence was similar to the control sequence, suggesting that this site is unlikely to be responsible for the transcriptional activity of *CDO1*. We also searched the NCBI SNP database for *CDO1* 5′-flanking region genetic variants that occur in the general population, and which may potentially disrupt *CDO1* mRNA expression. Of the 40 validated SNPs reported within 1200 nt of the *CDO1* 5′-flanking region (Supplemental Table 3), only 6 variants are located within a putative transcription binding motif reported in the present study (Fig. 4A): −70G > T (KLF1), −97G > A and -99G > T (MZF1), −193G > A (ZNF300) and -283C > T,G and -289C > G (ZNF239). The latter variant is relatively common (allelic frequency = 0.367) in the general population (Supplemental Table 3). Accordingly, we tested the luciferase activity from the mutated -289C > G sequence and found similar luciferase levels when compared to the control sequence (Fig. 4D) suggesting that this commonly occurring variant is unlikely to alter *CDO1* mRNA expression. Whilst this study did not identify the specific transcription factor binding motifs that regulate *CDO1* mRNA expression, our findings do suggest that regulatory sequences are most likely located within 1200 nucleotides of the *CDO1* 5′-flanking region.

### 3.5. Summary

The importance of sulfate in fetal growth and development cannot be overestimated. While interest in sulfate physiology continues to expand, there is still much to learn about the genes that maintain sulfate homeostasis in early development. It is remarkable that *CDO1* has

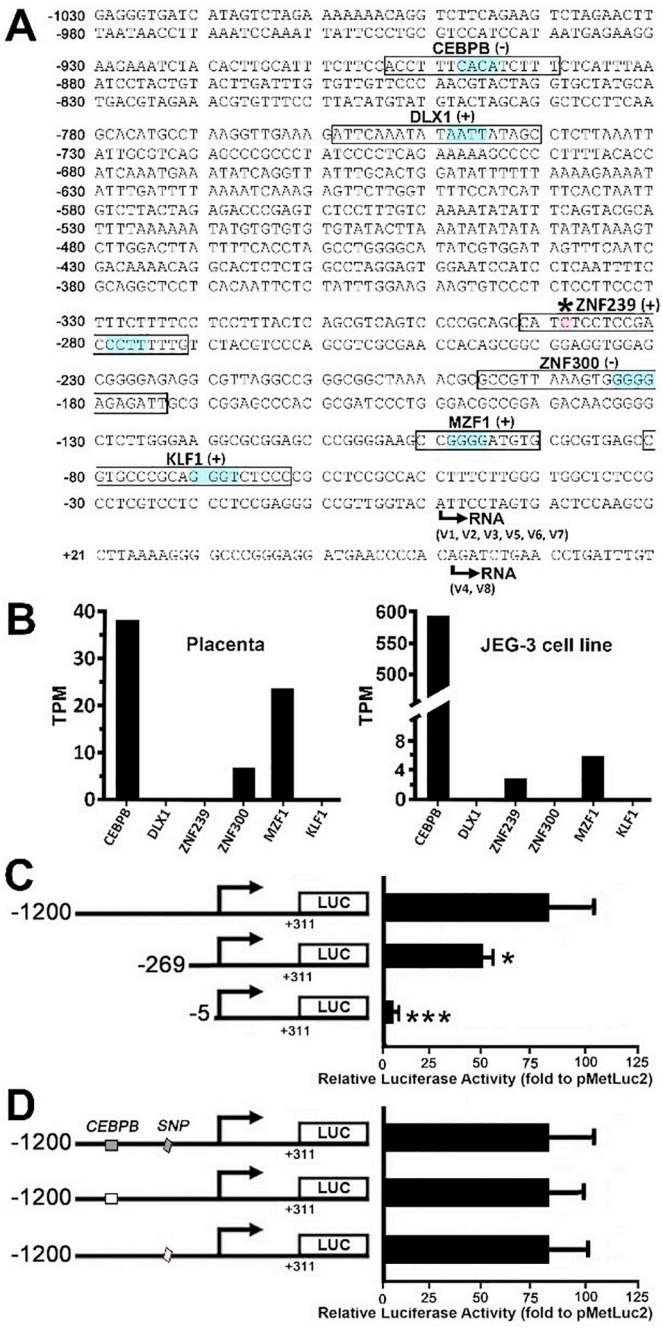


**Fig. 4.** Sequence and transcriptional activity of the human *CDO1* 5′-flanking region. (A) *CDO1* nucleotide sequence from −1030 to +70, with +1 and +52 (*arrows*) denoting the transcription initiation sites of *CDO1* mRNA variants 1, 2, 3, 5, 6 and 7, and variants 4 and 8, respectively. Putative transcription factor binding motifs are boxed, with core sequences shaded. *Location of the -289C > G single nucleotide polymorphism (rs34869) reported in the NCBI SNP database. (B) Approximate mRNA expression level of each transcription factor in human placenta and the JEG-3 cell line using data obtained from the online NCBI UniGene EST database http://www.ncbi.nlm.nih.gov/est and the Gene Expression Atlas database https://www.ebi.ac.uk/gxa/home, respectively, which were accessed between 30 and 31 January 2019. TPM, Transcripts Per Million. (C) Transcriptional activity of the 5′-truncated *CDO1* 5′-flanking region assessed by luciferase activity in JEG-3 cells. Data are shown as fold-induction to luciferase activity of the pMetLuc2 reporter, $n = 3$. $^{*}P < 0.05$ and $^{***}P < 0.001$ when compared to constructs containing the −1200 fragment. (D) Transcriptional activity of the pMetLuc2–1200 vector containing wild-type sequence (*filled symbols*) and individually mutated CEBPB motif and introduction of the SNP rs34869 (*unfilled symbols*). Data are shown as fold-induction to luciferase activity of the pMetLuc2 reporter, n = 3.

not yet been linked to any human pathology, despite its role in a major pathway of sulfate generation, as well as its link to severe developmental pathologies in laboratory mice. More than 2 decades have passed since the last description of the human *CDO1* gene and its expression in human tissues. Accordingly, it is timely that we now provide an update on its gene, mRNA, promoter region and protein isoform structures. Our finding of *CDO1* mRNA expression in the syncytiotrophoblast layer of the human term placenta warrants further studies to determine the physiological contribution of CDO1 to placental and fetal physiology. This study also collated a list of validated gene variants and assessed their predicted impact on CDO1 protein structure and function, providing valuable reference information for future genetic studies of *CDO1* in human health.

## Acknowledgements

This work was supported by Mater Research, the Queensland Perinatal Consortium (QPaCt) and the Mater Foundation. PAD is supported by a Mater Foundation Principal Research Fellowship and NHMRC funding from project grant APP1081911.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgmr.2020.100568.

## References

[1] P.A. Dawson, Sulfate in fetal development, Semin. Cell Dev. Biol. 22 (2011) 653–659.
[2] P.A. Dawson, K. Richard, A. Perkins, Z. Zhang, D.G. Simmons, Review: nutrient sulfate supply from mother to fetus: placental adaptive responses during human and animal gestation, Placenta 54 (2017) 45–51.
[3] P.A. Dawson, P. Sim, D.G. Simmons, D. Markovich, Fetal loss and hyposulfataemia in pregnant NaS1 transporter null mice, J. Reprod. Dev. 57 (2011) 444–449.
[4] J. Rakoczy, Z. Zhang, F.G. Bowling, P.A. Dawson, D.G. Simmons, Loss of the sulfate transporter Slc13a4 in placenta causes severe fetal abnormalities and death in mice, Cell Res. 25 (2015) 1273–1276.
[5] P.A. Dawson, A. Elliott, F.G. Bowling, Sulphate in pregnancy, Nutrients 7 (2015) 1594–1606.
[6] I. Ueki, H.B. Roman, A. Valli, K. Fieselmann, J. Lam, R. Peters, L. Hirschberger, M.H. Stipanuk, Knockout of the murine cysteine dioxygenase gene results in severe impairment in ability to synthesize taurine and an increased catabolism of cysteine to hydrogen sulfide, Am. J. Physiol. Endocrinol. Metab. 301 (2011) E668–E684.
[7] G. Gaull, J.A. Sturman, N.C. Raiha, Development of mammalian sulfur metabolism: absence of cystathionase in human fetal tissues, Pediatr. Res. 6 (1972) 538–547.
[8] J. Rakoczy, S. Lee, S.J. Weerasekera, D.G. Simmons, P.A. Dawson, Placental and fetal cysteine dioxygenase gene expression in mouse gestation, Placenta 36 (2015) 956–959.
[9] A.L. Levonen, R. Lapatto, M. Saksela, K.O. Raivio, Human cystathionine gamma-lyase: developmental and in vitro expression of two isoforms, Biochem. J. 347 (Pt 1) (2000) 291–295.
[10] N. Tsuboyama-Kasaoka, Y. Hosokawa, H. Kodama, A. Matsumoto, J. Oka, M. Totani, Human cysteine dioxygenase gene: structural organization, tissue-specific expression and downregulation by phorbol 12-myristate 13-acetate, Biosci. Biotechnol. Biochem. 63 (1999) 1017–1024.
[11] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994).
[12] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics 21 (2005) 2933–2942.
[13] F.T. Kraus, R.W. Redline, D.J. Gersell, D.M. Nelson, J.M. Dicke, Placental Pathology, Atlas of Nontumor Pathology, American Registry of Pathology, Washington DC, 2004.
[14] C.L. Roberts, P.A.L. Lancaster, Australian national birthweight percentiles by gestational age, Med.J.Aust. 170 (1999) 114–118.
[15] P.A. Dawson, L. Beck, D. Markovich, Hyposulfatemia, growth retardation, reduced fertility and seizures in mice lacking a functional $NaS_i$-1 gene, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 13704–13709.
[16] D.G. Simmons, J. Rakoczy, J. Jefferis, R. Lourie, H.D. McIntyre, P.A. Dawson, Human placental sulfate transporter mRNA profiling identifies abundant SLC13A4 in syncytiotrophoblasts and SLC26A2 in cytotrophoblasts, Placenta 34 (2013) 381–384.
[17] P.A. Dawson, J. Rakoczy, D.G. Simmons, Placental, Renal, and Ileal Sulfate transporter gene expression in mouse gestation, Biol. Reprod. 87 (2012) 1–9.
[18] D.G. Simmons, S. Rawn, A. Davies, M. Hughes, J.C. Cross, Spatial and temporal

expression of the 23 murine prolactin/placental Lactogen-related genes is not associated with their position in the locus, BMC Genomics 9 (2008) 352.
[19] J. Jefferis, J. Rakoczy, D.G. Simmons, P.A. Dawson, Molecular analysis of the human SLC13A4 sulfate transporter gene promoter, Biochem. Biophys. Res. Commun. 433 (2013) 79–83.
[20] K.P. McCann, M.T. Akbari, A.C. Williams, D.B. Ramsden, Human cysteine dioxygenase type I: primary structure derived from base sequencing of cDNA, Biochim. Biophys. Acta 1209 (1994) 107–110.
[21] J.E. Dominy Jr., J. Hwang, S. Guo, L.L. Hirschberger, S. Zhang, M.H. Stipanuk, Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity, J. Biol. Chem. 283 (2008) 12188–12201.
[22] M.H. Stipanuk, J.E. Dominy Jr., I. Ueki, L.L. Hirschberger, Measurement of cysteine dioxygenase activity and protein abundance, Curr. Protoc. Toxicol. 38 (2008) 6.15.11–16.15.25.