# SLC46A1 haplotype with predicted functional impact has prognostic value in breast carcinoma

# *SLC46A1* Haplotype with Predicted Functional Impact has Prognostic Value in Breast Carcinoma

**Viktor Hlavac[1,2], Radka Vaclavikova[1,2], Veronika Brynychova[1,2], Pavel Dvorak[1,3], Katerina Elsnerova[1], Renata Kozevnikovova[4], Karel Raus[5], Katerina Kopeckova[6], Sona Mestakova[7], David Vrana[8], Jiri Gatek[9] Pavel Soucek[12]**

*Viktor Hlavac viktor.hlavac@lfp.cuni.cz

[1]*Laboratory of Pharmacogenomics, Faculty of Medicine in Pilsen, Biomedical Center, Charles University, alej Svobody 76, 323 00 Pilsen, Czech Republic*
[2]*Toxicogenomics Unit, National Institute of Public Health, Prague, Czech Republic*
[3]*Department of Biology, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic*
[4]*Department of Oncosurgery, MEDICON, Prague, Czech Republic*
[5]*Institute for the Care for Mother and Child, Prague,Czech Republic*
[6]*Department of Oncology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic*
[7]*Department of Surgery, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic*
[8]*Department of Oncology, Palacky University Medical School and Teaching Hospital, Olomouc, Czech Republic*
[9]*Department of Surgery, EUC Hospital and University of Tomas Bata in Zlin, Zlin, Czech Republic*

**Abstract**

**Background and Objective** Membrane solute carrier transporters play an important role in the transport of a wide spectrum of substrates including anticancer drugs and cancer-related physiological substrates. This study aimed to assess the prognostic relevance of gene expression and genetic variability of selected solute carrier transporters in breast cancer.

**Methods** Gene expression was determined by quantitative real-time polymerase chain reaction. All *SLC46A1* and *SLCO1A2* exons and surrounding non-coding sequences in DNA extracted from the blood of patients with breast cancer (exploratory phase) were analyzed by next-generation sequencing technology. Common variants (minor allele frequency > 5%) with in silico-predicted functional relevance were further analyzed in a large cohort of patients with breast cancer (n = 815) and their prognostic and predictive potential was estimated (validation phase).

**Results** A gene expression and bioinformatics analysis suggested *SLC46A1* and *SLCO1A2* to play a putative role in the prognosis of patients with breast cancer. In total, 135 genetic variants (20 novel) were identified in both genes in the exploratory phase. Of these variants, 130 were non-coding, three missense, and two synonymous. One common variant in *SLCO1A2* and four variants in *SLC46A1* were predicted to be pathogenic by in silico programs and subsequently validated. A *SLC46A1* haplotype block composed of rs2239911-rs2239910-rs8079943 was significantly associated with ERBB2/HER2 status and disease-free survival of hormonally treated patients.

**Conclusions** This study revealed the prognostic value of a *SLC46A1* haplotype block for breast cancer that should be further studied.

## 1 Introduction

Breast cancer (OMIM no. 114480) remains the most frequent cancer in women globally [**1**]. Despite enormous progress in therapeutic approaches over the last decades, a number of cellular processes can lead to the development of tumor resistance [**2**].

Among tumor resistance pathways, drug uptake and efflux mostly mediated by membrane-anchored ATP-bind- ing cassette (ABC) and solute carrier (SLC) transporters [**3, 4**] play a pivotal role. It seems obvious that the equilibrium of these exporters/importers may be exploited for estimation of cellular drug resistance [**5, 6**]. Dysregula- tions of gene expression in a number of ABCs and SLCs in tumor tissues from patients with different solid tumors, including breast cancer, were demonstrated by several recent studies [**7-11**]. Additionally, these studies revealed associations between gene expression of transporters and response to preoperative or palliative cytotoxic therapy and/or survival of patients after adjuvant therapy. Therefore, initial implications for prognosis and personalized therapy of cancer were already suggested in this area and further study should open new horizons in the clinical use of this new knowledge.

In contrast with ABCs [8], the role of SLCs in breast cancer prognosis and therapy efficacy remains less explored, perhaps owing to a much higher complexity of the SLC superfamily comprising more than 450 genes compared with 48 human ABC genes [**12**]. The SLC superfamily consists of 65 families encompassing ion coupled, concentrative, equilibrative, or passive transporters transferring a wide range of substrates, including amino acids, lipids, inorganic ions, peptides, saccharides, metal ions, proteins, and xenobiotics, including drugs (**https:// www.bioparadigms.org/**). For cancer research, particularly nucleoside transporters (SLC28/hCNT and SLC29/hENT families) providing the influx of nucleoside analogs as 5-fluorouracil attract attention [**13**]. Several SLCs from families SLC19 and SLC22 transport folate and methotrexate [**14, 15**]. Another folate transporter is SCL46A1, which translocates folates across the apical membrane of entero- cytes to the proximal jejunum. Loss-of-function mutations in this gene were found in patients with the autosomal recessive disorder, hereditary folate malabsorption [**16**]. Organic cation/carnitine transporters from the SLC22/ OCT family transport doxorubicin in vitro [**17**] and their genetic variability may predict the prognosis of patients with gastrointestinal stromal tumors treated with a selective tyrosine kinase inhibitor imatinib [**18**]. Organic anion transporting polypeptides currently classified into the SLCO family transport numerous drugs and xenobiotics and their role in absorption, elimination, and distribution of anticancer drugs is awaiting characterization [**19**]. A number of SLCs, for example, those from the SLCO family, translocate hormones, including estrogen and its conjugates [**20**] and may be important for the etiology and progression of hormone-dependent breast cancers.

Taken together, available data in the literature suggest a potential role of SLC transporters in the prognosis and response of patients to anticancer therapy. However, studies in target tissues of patients are limited precluding the translation of knowledge into precision medicine.

We selected, by help of the genomic SLC database Bioparadigms (**http://slc.bioparadigms.org/**), 21 drug transport- and cancer etiology-related SLC transporters (SLC19A1, 22A1/2/3/6/7/8/11, 28A1/2/3, 29A1/2/3/, 31A1, 46A1,47A1/2, SLCO1A2, and 1B1/3; Table S1 of the Electronic Supplementary Material [ESM]). Subsequently, we explored their gene expression levels in a discovery set of post-treatment target tissues from patients with breast cancer treated with neoadjuvant cytotoxic therapy (NACT) and compared expression profiles with the response of the patients to NACT. An independent cohort of chemotherapy-naive patients was used for the comparison and assessment of prognostic relevance. We then evaluated germline genetic variability

of two prognostically most relevant genes by targeted sequencing of complete coding sequences in a small scale and subsequently predicted the functional consequences of identified variants. Common functionally prioritized variants were further validated in a large-scale follow-up study evaluating their prognostic and predictive value in a clinical context.

## 2 Materials and Methods

### 2.1 Patients

The post-treatment tissue samples of human carcinomas of the mammary gland were prospectively obtained from 33 patients with incident breast cancer diagnosed at the Department of Oncosurgery, Medicon, in Prague during 2006-2010. Patients were treated with NACT based on 5-fluorouracil/anthracyclines/cyclophosphamide (FAC or FEC) and/or taxanes (for patient characteristics, see Table S2 of the ESM). Paired adjacent tissue samples without morphological signs of carcinoma (non-neoplastic controls) were available from 22 patients. Collection and pathological processing of tissue samples and retrieval of data were described before [**8**].

The chemotherapy-naive tissue samples of human carcinomas of the mammary gland were prospectively obtained from 57 incident patients, independent to the post-treatment set, diagnosed at the Faculty Hospital in Motol, Prague during 2003-2007. Paired non-neoplastic controls were available from 31 patients. Patients were treated with adjuvant chemotherapy and/or hormonal therapy after surgery (characteristics in Table S2 of the ESM). Collection and pathological processing of tissue samples and retrieval of data were previously described [**21**].

The exploratory phase of targeted sequencing included blood DNA samples from 105 incident patients with breast cancer of Caucasian origin diagnosed at the Department of Oncosurgery, Medicon, in Prague during 2006-2012. Sixty- eight patients underwent NACT with regimens described above. The validation set was composed of 815 patients with incident breast cancer recruited at both above-listed surgery departments in Prague and at the Department of Surgery in Zlin during 2001-2013. The patient recruitment and sample and data processing were described before [**22**] and the characteristics are in Table S3 of the ESM. Briefly, the following data on patients were retrieved from medical records: age at diagnosis, menopausal status, personal medical history, family history (number of relatives affected by breast/ovar- ian carcinoma or other malignant diseases), stage, tumor size, presence of lymph node or distant metastasis, histological type and grade of the tumor, expression of estrogen, progesterone, and ERBB2 (v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2, OMIM:164870) receptors, expression of Ki67 (proliferation-related Ki-67 antigen, OMIM:176741), therapy (regimen, duration, response to NACT where applicable), and survival follow-up.

All patients after the primary chemotherapy and surgery were followed for local or distant relapse. Response to the NACT was evaluated according to RECIST [**23**] based on ultrasonography performed before and after the cytotoxic therapy. Disease-free survival (DFS) of patients was defined as the time elapsed between surgery and disease recurrence or death of any cause. The study flow diagram is presented in Fig. S1 of the ESM.

All patients were informed about the study and those who agreed and signed an informed consent participated in the study. Procedures performed in the present study were in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study protocol was approved by the Ethical Commission of the National Institute of Public Health in Prague (approvals no. 9799-4 and 15-25618A).

## 2.2 Isolation of Total RNA and Complementary DNA Preparation

Total RNA was isolated from snap-frozen tissues using Trizol reagent (Invitrogen, Carlsbad, CA, USA). RNA quantity and quality were assessed and complementary DNA was synthesized using 0.5 µg of total RNA as described before [**21**]. The complementary DNA was then pre-amplified according to the published procedure [**8**] using 25 µL of TaqMan PreAmp Master Mix (Life Technologies Corp., Carlsbad, CA, USA) and a pool of 24 specific TaqMan Gene Expression Assays (Life Technologies Corp.), listed in Table S4 of the ESM.

## 2.3 DNA Extraction

Blood samples were collected during the diagnostic procedures using tubes with $K_3$EDTA anticoagulant and genomic DNA was isolated from human peripheral blood lymphocytes by the standard phenol/chloroform extraction and ethanol precipitation method [**24**].

## 2.4 Quantitative Real-Time Polymerase Chain Reaction

Quantitative real-time polymerase chain reaction was performed and results evaluated as described previously [**8**]. The relative standard curve was generated from five log dilutions of one non-neoplastic control tissue sample (calibrator). Amplification efficiencies (E) for each reference and target gene were calculated applying the formula $E = 10^{-1/slope}-1$. *EIF2B1, MRPL19,* and *IPO8* were selected as the most stable reference genes for data normalization [**8**]. The quantitative real-time PCR study design adhered to the MIQE Guidelines (Minimum Information for Publication of Quantitative Real-Time PCR Experiments [**25**]).

## 2.5 Targeted Sequencing

Libraries encompassing all exons of *SLC46A1* (six exons) and *SLCO1A2* (14 exons) and corresponding untranslated regions were prepared following the previously published design [**22**]. Based on the character of probe design, i.e., tiling, the exons were surrounded by approximately 30-bp regions of intronic sequences that were also sequenced in both directions. Target enrichment was performed by the Nimblegen's SeqCap EZ Choice (Roche, Prague, Czech Republic) using a standard SeqCap protocol: libraries were prepared using the Kapa Library Preparation Kit (Roche) and sequenced on the MiSeq platform (Illumina Inc., San Diego, CA, USA).

## 2.6 Genotyping of the Validation Set

Five genetic variants in *SLC46A1* and *SLCO1A2* were analyzed in DNA from 815 patients with breast cancer using real-time PCR methods. Three variants were analyzed by TaqMan SNP Genotyping assays *(SLC46A1;* rs2239910 by assay C_2536568_20, rs2239911 by assay C_2536569_10 and *SLCO1A2;* rs11568563 by assay C_25605897_10) using ViiA7 Real-time PCR system with a 384-well block (ThermoFisher Scientific, Waltham, MA, USA) using standard conditions.

*SLC46A1* variants rs8079943 and rs2239908 were estimated by high-resolution melting analysis using a Light- Cycler 480 real-time PCR instrument (Roche). Primer sequences, composition of amplification mix, and cycling conditions are listed in Table S5 of the ESM. The non-template control contained water instead of genomic DNA as a negative control. Additionally, positive controls (DNA samples with known genotype estimated by direct sequencing) were employed in each analysis run. Collected data were analyzed using LightCycler 480 software (Roche) as published before [**26**].

## 2.7 Data Analysis

Raw cycle thresholds ($C_t$) from quantitative real-time PCR were analyzed by the REST2009 program (Qiagen, Hildesheim, Germany). Each sample was assayed in duplicate and the mean value was used for calculations. Samples with higher $C_t$ than 40 were treated as missing data. Data were normalized to reference genes using a formula $AC_t$ = mean $C_t$(reference) - $C_t$(target). For statistical analyses of associations of transcript levels with clinical data, non-parametric tests (Kruskal-Wallis, Mann-Whitney, and Spearman rank) were used. Response to NACT was classified into good (partial pathological response) vs poor (stable disease or progression). Disease-free survival was evaluated by the Kaplan-Meier method and the log-rank test was used for evaluation of the compared groups of patients. For the multiparametric analysis, the Cox proportional hazards model with tumor size and grade, lymph node metastasis, and estrogen receptor expression as covariates was used. The influence of *SLC46A1* and *SLCO1A2* gene expression on breast cancer progression was also evaluated on independent data obtained from freely available on-line databases. Searches in cBioPortal [**27, 28**], MEXPRESS [**29**], and KM-plotter [**30**] were performed to collect these bioinformatics results.

Raw data from the targeted sequencing were processed by the in-house bioinformatics pipeline in Linux [**22**]. Detection of single nucleotide polymorphisms (SNPs) and small indels was performed by the Genome Analysis Toolkit (Broad Institute, Cambridge, UK) 4.1.8.0 Haplotype Caller according to Genome Analysis Toolkit Best Practices [**31**].

The following genetic models were evaluated: ancestral homozygote or heterozygote vs non-ancestral homozygote (dominant model) and ancestral homozygote vs non-ancestral homozygote or heterozygote (recessive model) in the validation set. The additive model was also tested. Haplotypes were evaluated using HaploView software program version 4.2; phasing of haplotypes prior to a block selection was done using the E-M algorithm and the block selection was based on confidence intervals (CIs) as described by Gabriel et al. [**32**].

The functional relevance of the variants evaluated in the exploratory phase was analyzed in silico by RegulomeDB (**http://regulome.stanford.edu**) [**33**], PolyPhen-2 [**34**], SIFT [**35**], and HaploReg version 2 and version 3 [**36**] programs. The annotation of the variants was performed in Annovar (version 2019 October 24, Pennsylvania, PA, USA) [**37**].

Associations of SNPs and novel pathogenic variants with prognostic clinical data were evaluated by the twosided Pearson's chi-square or the one-way analysis of variance tests. Disease-free survival was evaluated by the Kaplan-Meier method and the log-rank test was used for evaluation of the compared groups of patients. For the multiparametric analysis, the Cox proportional hazards model with tumor size and grade, lymph node metastasis, and estrogen and progesterone receptor expression as covariates was used. A $p$ value of less than 0.05 was considered statistically significant. Analyses were conducted by the statistical program SPSS version 16.0 (SPSS Inc., Chicago, IL, USA). The correction for the false discovery rate was applied according to Benjamini and Hochberg [**38**]. The PharmGKB [**39**] database was searched for relevant pharmacogenomics information for data interpretation. The sequencing data that support the findings of this study are openly available in Sequence Read Archive (**https://www.ncbi.nlm.nih.gov/sra**) under accession no. PRJNA510917.

## 3 Results

### 3.1 Gene Expression Profiles in Target Tissues Compared to Non-Neoplastic Controls

*SLC19A1, SLC29A3,* and *SLC31A1* were significantly overexpressed in tumors compared with non-neoplastic controls in the posttreatment set and *SLC22A3, SLC22A11, SLC47A2,* and *SLCO1A2* were significantly downregulated in tumors. Transcript levels of *SLC22A6, SLC22A7, SLC22A8,* and *SLCO1B1* were below the limit of detection and the rest of SLCs did not significantly differ in the post-treatment set. Gene expression profiles of SLCs in the posttreatment set are graphically displayed in Fig. S2A of the ESM.

Congruently, the same genes were dysregulated in the same direction also in the pre-treatment set. Additionally, *SLC22A7* and *SLC29A2* were significantly overexpressed and *SLC22A2, SLC22A6, SLC22A8, SLC28A3, SLC47A1, SLCO1B1,* and *SLCO1B3* were significantly downregulated in tumors compared with non-neoplastic control tissues in the pre-treatment set. The rest of the genes was unchanged or below the detection limit. Gene expression profiles of SLCs in the pre-treatment set are graphically displayed in Fig. S2B of the ESM. The fold change between tumor and control tissues (mean expression values) with *p* values and 95% confidence intervals are listed in Table S6 of the ESM.

### 3.2 Associations of Gene Expression Levels with Clinical Data

Good responders to NACT in the post-treatment set had significantly higher intra-tumoral *SLCO1A2* transcript levels (p = 0.012) and higher non-neoplastic *SLC28A1* levels than poor responders (p = 0.025) [Fig. S3 of the ESM]. Patients with a higher intra-tumoral *SLC19A1* level than the median had significantly shorter DFS than patients with low expression (p = 0.007) and patients with a lower *SLC22A11* level than the median had shorter DFS than the rest of the patients (*p* = 0.023; Fig. S4 of the ESM) in the pre-treatment set. A multi-variate analysis using the Cox regression hazard model with tumor size and grade, lymph node metastasis, and estrogen receptor expression as covariates has confirmed the association of *SLC19A1 (p*= 0.023; hazard ratio = 7.14 for high expression; CI 1.31-3.33), but not that of *SLC22A11 (p* = 0.056; hazard ratio = 0.13 for high expression; CI 0.02-1.05).

A higher *SLC46A1* level in estrogen receptor-expressing tumors compared with estrogen receptor-negative tumors was observed in both sets (p < 0.001 for the pre-treatment set and *p* = 0.014 for the post-treatment set; Fig. S3 of the ESM). The rest of the evaluated associations (with tumor size, lymph node metastasis, grade, progesterone or ERBB2 receptor status, and Ki67 expression) were not significant or did not replicate in either of the studied sets.

### 3.3 In Silico Analysis of SLC46A1 and SLCO1A2 Gene Expression and Breast Cancer Progression

In the MEXPRESS tool, 871 The Cancer Genome Atlas (TCGA) breast cancer samples were available for evaluation. Congruently with our observations, patients with estrogen receptor (and progesterone receptor) expressing tumors had a significantly higher level of *SLC46A1* expression than receptor-negative patients (p = 2.2e-16). Comparison of PAM50 subtypes revealed that patients assigned to Luminal A and Luminal B subtypes displayed the highest level of *SLC46A1* expression and basal-like subtype patients had the lowest level (Fig. S5 of the ESM).

In contrast to *SLC46A1,* patients with estrogen receptorpositive tumors had significantly lower levels of *SLCO1A2* than those without receptor expression (p = 0.025). Basal- like subtype patients

displayed the highest level of *SLCO1A2* expression and patients belonging to Luminal A and Luminal B subtypes had the lowest level (Fig. S5 of the ESM).

Analysis of 994 TCGA patients in cBioPortal confirmed the above observed correlations for both genes. This analysis also stressed the significance of the mutual exclusivity between the two genes (Log2 odds ratio <-3, *p* value and *q* value equals 0.033).

The analysis of survival data of 1764 patients with breast cancer in KM-plotter showed a significant influence of the *SLC46A1* as well as *SLCO1A2* gene expression on relapse-free survival. In the case of the *SLC46A1* gene, the upper-quartile survival of the high expression cohort was 43 months vs 29 months of the low-expression cohort (the log-rank test $p = 0.008$; Fig. **1**a). In the case of the *SLCO1A2* gene, the upper-quartile survival of the high-expression cohort was 61 months vs 40 months of the low-expression cohort (log-rank test, $p = 2.3e\text{-}06$; Fig. **1**b).

Stratified analysis according to major histological characteristics of tumors and intrinsic subtypes showed that low expression of *SLC46A1* is a negative prognostic factor in patients with luminal A subtype ($p = 0.046$). Dividing patients by estrogen receptor status has demonstrated opposite prognostic roles of *SLC46A1.* Low *SLC46A1* expression was a non-significant poor prognosis factor in estrogen receptor-positive patients, but weakly significantly ($p = 0.041$) protective in patients without expression of the receptor (Table S7 and Fig. S6 of the ESM). Low expression of *SLCO1A2* is a poor prognosis factor in patients with any intrinsic subtype (significant for luminal A and B and basal, $p = 0.018$, 0.006, and 0.013, respectively) and also for estrogen receptor-negative or ERBB2-positive patients ($p = 0.018$ and 0.016, respectively; Table S7 and Fig. S7 of the ESM). In general, stratified analysis confirmed that prognostic value of low expression of SLCO1A2 as a poor prognosis factor is universal while that of SLC46A1 depends on expression of the estrogen receptor.

Based on the above associations of *SLCO1A2* with the response to NACT in the pre-treatment set and *SLC46A1* with the estrogen receptor status, a major predicting factor for hormonal therapy in breast cancer, replicated in both sets, and given the supportive in silico data suggesting a significant prognostic role, these two genes were selected for further study.

### 3.4 Genetic Variability of SLC46A1 and SLCO1A2 in Patients with Breast Cancer

All samples were sequenced with a mean coverage of 67. Ninety-two percent of bases were covered at least ten times. Of the total number of 27 regions sequenced, 100% base pairs were called in 20 regions. Three regions were covered by > 98%, and two by > 85%. Two regions in exon 1 of *SLCO1A2* were covered by 53% (ENST00000435179) and 59% (ENST00000307378). A total of 135 genetic variants were identified in both genes. Of these, 130 variants were identified in non-coding regions. In the coding regions, three were missense amino acid changes and two were synonymous variants.

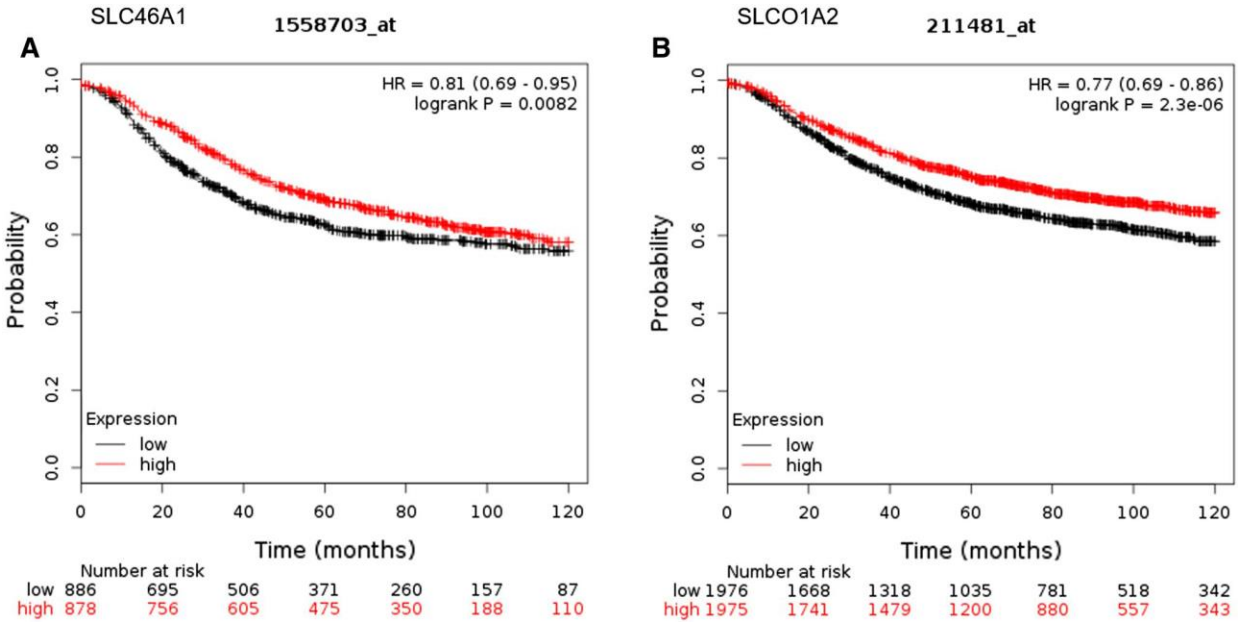


**Fig. 1** Associations between gene expression levels and survival of patients with breast cancer. Kaplan-Meier survival curves for patients with intra-tumoral expression levels above the median (red line) of **a** *SLC46A1 (p* = 0.008), and **b** *SLCO1A2 (p* = 2.3e-06) compared to patients with levels below or equal the median (black line) are displayed. Significance was evaluated by the log-rank test using data of 1764 patients with breast cancer in KM-plotter

**Table 1** Overview of identified variants in *SLC46A1* and *SLCO1A2* genes in the exploratory phase

| Variant | *SLCO1A2*[a] | *SLC46A1*[a] | Total[a] | Novel[b] |
|---|---|---|---|---|
| Intronic | 46 (2) | 7 (3) | 53 (5) | 6 |
| 3′UTR | 32 (0) | 21 (6) | 53 (6) | 7 |
| 5′UTR | 6 (0) | 0 | 6 (0) | 1 |
| Downstream | 2 (0) | 0 | 2 (0) | 0 |
| Upstream | 0 | 3 (0) | 3 (0) | 1 |
| Intergenic | 11 (0) | 2 (0) | 13 (0) | 5 |
| Missense | 2 (1) | 1 (0) | 3 (1) | 0 |
| Synonymous | 1 (NA) | 1 (NA) | 2 (NA) | 0 |
| All | 100 (3) | 35 (9) | 135 (12) | 20 |

*NA not applicable, UTR untranslated region [a]Numbers of identified variants with pathogenic variants predicted by SIFT, PolyPhen-2, or RegulomeDB in parentheses [b]Numbers of novel variants, i.e., not found in dbSNP*

Together, 20 (15%) novel variants (6 of 35, i.e., 17% in *SLC46A1* and 14 of 100, i.e., 14% in *SLCO1A2)* according to dbSNP Build 150 were discovered. The distribution of observed germline variants is summarized in Table 1.

### 3.5 Functional Predictions Classifying Germline Variants

Functional relevance of all 135 identified variants in the exploratory phase was predicted by the in silico tools described in methods. Three potentially functional variants in *SLCO1A2* and nine in *SLC46A1* were identified (Table 1). A potentially functional SNP rs11568563 in the coding region of *SLCO1A2* was evaluated by four different programs. This variant was predicted to be probably damaging by PolyPhen-2 and deleterious by SIFT. The analysis of rs11568563 by HaploReg v3 suggested that it changes binding sites for multiple transcription factors (Table S8 of the ESM). The position of *SLCO1A2* coding variant rs11568563 with the predicted pathogenic effect is depicted in Fig. 2. The other two potentially functional variants had minor allele frequencies (MAF) < 5%.

In *SLC46A1,* four common SNPs (rs8079943, rs2239911, rs2239910, and rs2239908) in non-coding regions were predicted by the RegulomeDB and HaploReg as likely to affect the binding of transcription factors (Table S8 of the ESM). Following the priorities of carrying functionally relevant genetic variants, SNPs rs8079943, rs2239911, rs2239910, and rs2239908 of *SLC46A1* and SNP rs11568563 in *SLCO1A2* were selected for the next validation phase of the present study in a larger cohort *(n* = 815) of patients with breast cancer. Furthermore, the MAF of those variants was ≥ 5% as shown in Table 2, which supports their potential as functionally relevant variations.

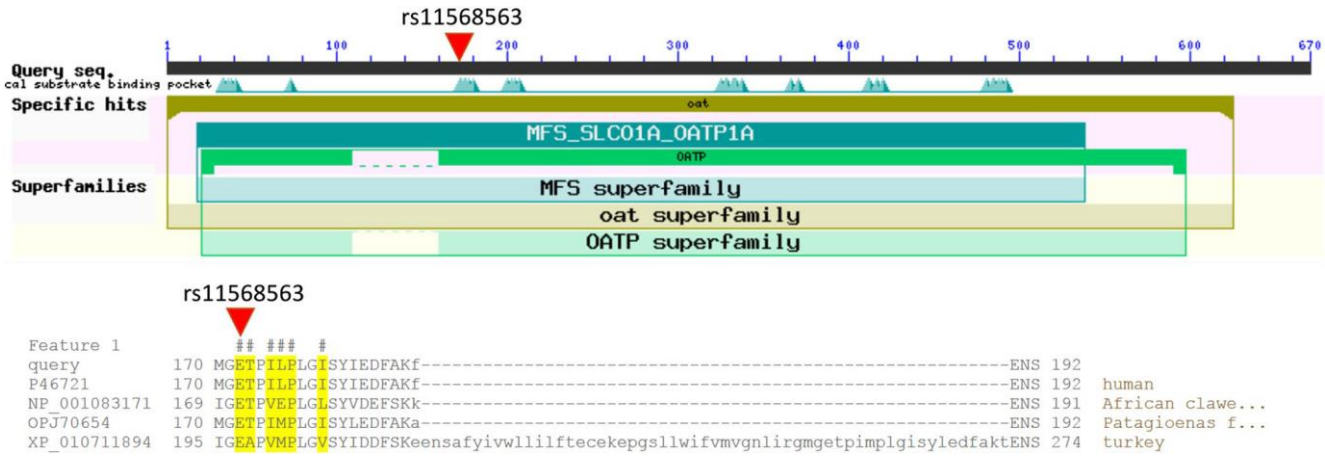


**Fig. 2** Position of the deleterious variant in *SLCO1A2* gene. The positions of coding single nucleotide polymorphisms with the predicted pathogenic effect in *SLCO1A2* are depicted by triangles. The schematic of solute carrier domains is adopted from the National Center for Biotechnology Information's Conserved Domain Database [**40**]. *OATP* organic anion transporting polypeptide

### 3.6 Validation Study

Three variants *(SLC46A1;* rs2239910, rs2239911 and *SLCO1A2;* rs11568563) were assessed by real-time PCR using TaqMan SNP genotyping assays. Two SNPs *(SLC46A1;* rs8079943 and rs2239908) were assessed by high-resolution melting analysis. Minor allele frequencies estimated in the validation set of patients (n = 815) did not substantially differ from those observed in the exploratory phase and in the general Caucasian European population (Table **3**).

To evaluate clinical associations in the validation set, variants were compared with clinical prognostic factors and responses of the subgroup of patients to NACT *(n* = 168). None of the individually analyzed variants were significantly associated with any of the clinical characteristics (tumor size, grade, lymph node metastasis, status of estrogen and progesterone and ERBB2 receptors, Ki67

expression, and response to NACT). However, haplotype analysis of *SLC46A1* suggested that rs2239911-rs2239910-rs8079943 forms a haplotype block (Fig. **3**). This block was significantly associated with expression of the ERBB2 (p = 0.007), see Table **4**. No association of this block with the rest of the clinical factors including estrogen receptor status was found (Table S9 of the EMS).
To further dissect the prognostic relevance of the followed variants, DFS of groups of patients stratified by therapy type were evaluated. Individual variants were not significantly associated with DFS of patients, but patients harboring the CGA haplotype had significantly shorter DFS than patients with other haplotypes in a subgroup of patients treated exclusively with adjuvant hormonal therapy (p = 0.025, mean DFS 95 ± 6 months for the patients with the CGA haplotype, *n* = 53 vs 108 ± 2 months for patients with *MAF* minor allele frequency [a]Including functional consequence and HGVS nomenclature (build GRCh38.p13) [b]Number of carriers with % in parentheses breast cancer. Furthermore, the MAF of those variants was > 5% as shown in Table **2**, which supports their potential as functionally relevant variations.

**Table 2** Distribution of genotypes for *SLC46A1* and *SLCO1A2* functional variants in the exploratory phase

| Genetic variants[a] | Patients, *N* (%)[b] | MAF |
|---|---|---|
| *SLCO1A2* | | |
| *rs11568563*, missense, NC_000012.12:g.21304500T>G, NP_602307.1:p.Glu172Asp | | |
| TT | 91 (87) | 0.07 |
| GT | 14 (13) | |
| *SLC46A1* | | |
| *rs8079943*, 3′-UTR, NC_000017.11:g.28395626G>A | | |
| GG | 20 (19) | 0.53 |
| GA | 59 (56) | |
| AA | 26 (25) | |
| *rs2239911*, 3′-UTR, NC_000017.11:g.28396594G>T | | |
| GG | 20 (19) | 0.53 |
| GT | 59 (56) | |
| TT | 26 (25) | |
| *rs2239910*, 3′-UTR, NC_000017.11:g.28396647A>C | | |
| AA | 15 (14) | |
| AC | 61 (58) | 0.57 |
| CC | 29 (28) | |
| *rs2239908*, 3′-UTR, NC_000017.11:g.28398249G>A | | |
| GG | 20 (19) | |
| GA | 59 (56) | 0.53 |
| AA | 26 (25) | |

*MAF minor allele frequency [a]Including functional consequence and HGVS nomenclature (build GRCh38.p13) [b]Number of carriers with % in parentheses*

### 3.7 Validation Study

Three variants *(SLC46A1;* rs2239910, rs2239911 and *SLCO1A2;* rs11568563) were assessed by real-time PCR using TaqMan SNP genotyping assays. Two SNPs *(SLC46A1;* rs8079943 and rs2239908) were assessed by high-resolution melting analysis. Minor allele frequencies estimated in the validation set of patients (n = 815) did not substantially differ from those observed in the exploratory phase and in the general Caucasian European population (Table **3**).

To evaluate clinical associations in the validation set, variants were compared with clinical prognostic factors and responses of the subgroup of patients to NACT *(n* = 168). None of the individually analyzed variants were significantly associated with any of the clinical characteristics (tumor size, grade, lymph node metastasis, status of estrogen and progesterone and ERBB2 receptors, Ki67 expression, and response to NACT). However, haplotype analysis of *SLC46A1* suggested that rs2239911-rs2239910-rs8079943 forms a haplotype block (Fig. **3**). This block was significantly associated with expression of the ERBB2 (p = 0.007), see Table **4**. No association of this block with the rest of the clinical factors including estrogen receptor status was found (Table S9 of the EMS).

To further dissect the prognostic relevance of the followed variants, DFS of groups of patients stratified by therapy type were evaluated. Individual variants were not significantly associated with DFS of patients, but patients harboring the CGA haplotype had significantly shorter DFS than patients with other haplotypes in a subgroup of patients treated exclusively with adjuvant hormonal therapy (p = 0.025, mean DFS 95 ± 6 months for the patients with the CGA haplotype, *n* = 53 vs 108 ± 2 months for patients with other haplotypes, *n* = 251).

**Table 3** Distribution of genotypes for *SLC46A1* and *SLCO1A2* functional variants in the validation phase

| Gene/SNP | Genotypes | | | | MAF | MAF in CEU |
|---|---|---|---|---|---|---|
| | Common homozygous | Heterozygous | Rare homozygous | Missing data[a] | | |
| *SLC46A1* | | | | | | |
| *rs2239908* | 225 | 404 | 190 | 0 | 0.52 | 0.54 |
| *rs2239910* | 251 | 416 | 148 | 4 | 0.56 | 0.57 |
| *rs2239911* | 214 | 424 | 178 | 3 | 0.52 | 0.54 |
| *rs8079943* | 211 | 422 | 184 | 2 | 0.51 | 0.50 |
| *SLCO1A2* | | | | | | |
| *rs11568563* | 735 | 76 | 4 | 4 | 0.05 | 0.10 (0.07[b]) |
| *SLC46A1* | TTC | CGA | Other | Missing data[a] | | |
| Haplotype | 210 | 145 | 454 | 6 | N/A | N/A |

*CEU Caucasian European population, MAF minor allele frequency in this study, N/A not applicable, SNP single nucleotide polymorphism [a]Missing data because of the low quantity or quality of DNA samples [b]MAF in the general Czech population. Data from the National Center for Medical Genomics (**https://www.ncmg.cz**)*

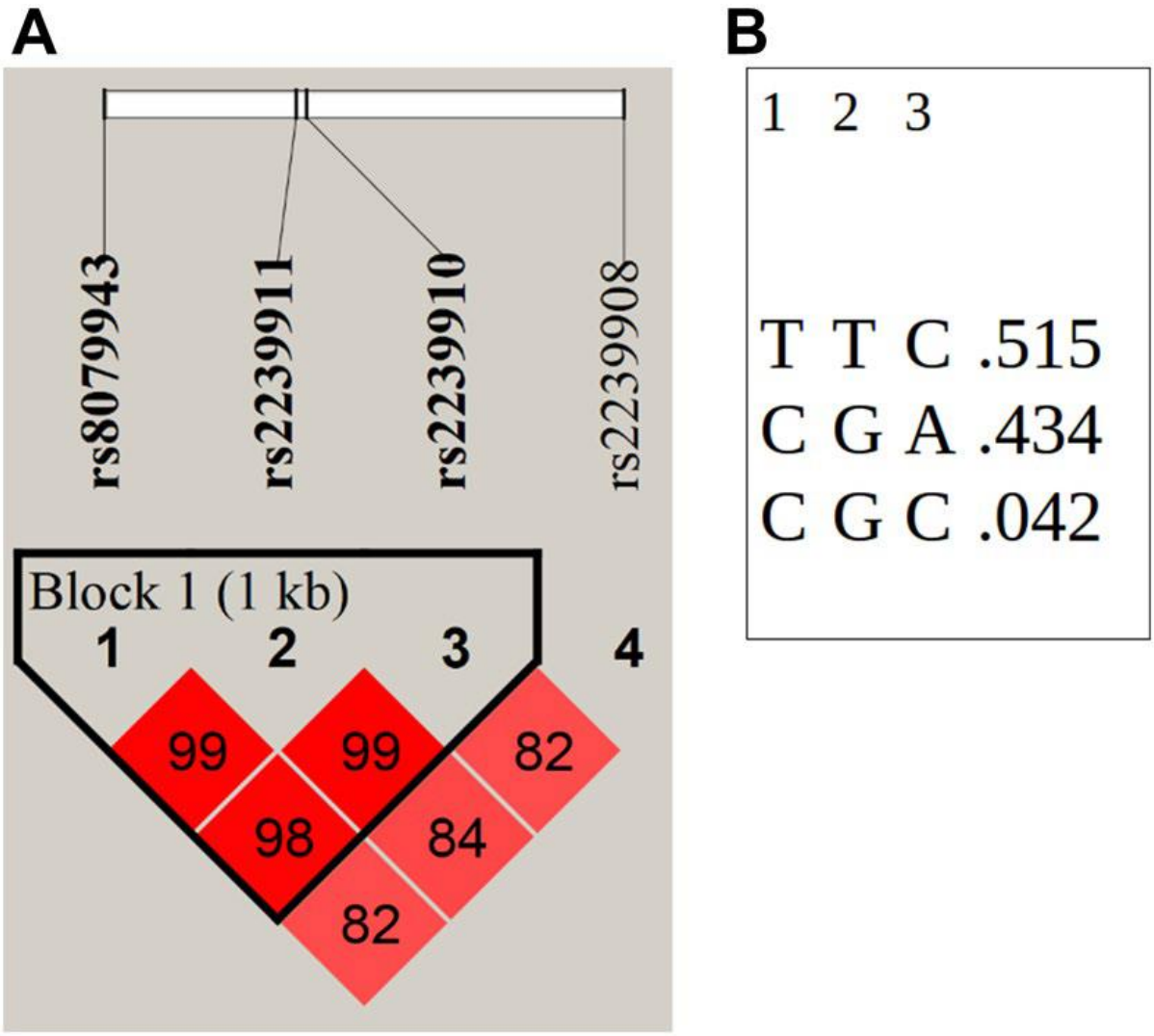


**Fig. 3** Haplotype analysis of the *SLC46A1* single nucleotide polymorphisms. The figure indicates the linkage disequilibrium plot (**a**) and one block comprising single nucleotide polymorphism haplo- types (**b**) predicted from experimental data obtained in the present study. The likelihood of linkage of two tested single nucleotide polymorphisms increases from a light to a dark red color (**a**). Frequencies of haplotypes in the Caucasian European population are shown (**b**). Analysis was performed by the HaploView v4.2 program

Kaplan-Meier curves of the DFS analysis are depicted in Fig. **4**. This association was confirmed also by a multi-parametric analysis with, tumor size and grade, lymph node metastasis, and estrogen and progesterone receptor expression status as covariates *(p* = 0.030; hazard ratio = 2.11 for the CGA haplotype; CI 1.08-4.13). No significant associations with DFS were observed in the unselected group of patients, in the subgroup of patients treated with different adjuvant cytotoxic therapy regimens, and in groups divided by the estrogen receptor status (Fig. S8 of the ESM).

**Table 4** Association of haplotype block of *SLC46A1* single nucleotide polymorphisms with ERBB2 receptor status assessed in the validation phase

| Haplotype[a] | ERBB2 positive[b] | ERBB2 negative[b] | HR | *P* value/*q* value[c] |
|---|---|---|---|---|
| CGA | 44 | 99 | 1.5 | 0.007/0.022 |
| TTC or CGC | 151 | 511 | – | – |

HR hazard ratio [a]Haplotype composed of SLC46A1 SNPs rs2239911-rs2239910-rs8079943
[b]Number of patients [c]q value adjusted by the false discovery rate test

## 4 Discussion

The role of SLC transporters in human cancer is underexplored. This study observed associations of intra-tumoral transcript levels of selected drug-relevant SLC transporters with clinical data of patients with breast cancer and further substantiated the potential prognostic role of two candidate genes using in silico tools. A genetic study was designed to address the potential use of candidate biomarkers in precision medicine, as a genotype based on germline SNPs is stable and can be determined on a time- and cost-effective basis.

The transcript level of *SLC46A1* and *SLCO1A2* in carcinomas associated with estrogen receptor expression status or with responses of patients to NACT in the discovery phase and their significance was confirmed by their association with survival in the in silico analysis based on the TCGA breast cancer dataset.

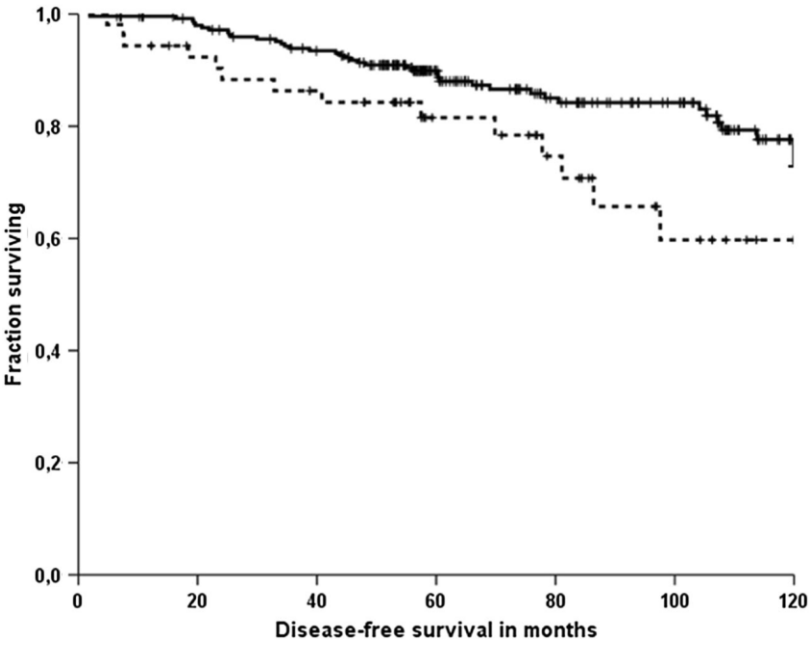


**Fig. 4** Kaplan-Meier plot of the association of haplotype block of three single nucleotide polymorphisms in *SLC46A1* with disease-free survival of patients treated only with hormonal therapy. The dashed line represents the CGA haplotype and the solid line represents the TTC, CGC, or other haplotypes. Significance was evaluated by the log-rank test (*p* = 0.025)

To explore the genetic basis of these associations, we further explored functional aspects of germline variability of complete coding sequences of both genes in the small-scale exploratory study. Finally, the identified in silico functional common variants were compared with clinical data in the large validation study.

The performed in-depth exploratory analysis of coding and regulatory regions of *SLC46A1* and *SLCO1A2* genes enabled identification of the total number of 135 alterations, of which one coding variant in *SLCO1A2* and four variants in non-coding regions of *SLC46A1* were predicted as functionally relevant by RegulomeDB, PolyPhen-2, SIFT, and Hap- loReg [**33-36**] in silico tools. Most interestingly, the haplotype block composed of rs2239911-rs2239910-rs8079943 SNPs in *SLC46A1* was significantly associated with DFS of patients receiving adjuvant hormonal therapy in the validation set of patients with breast cancer. In addition, this block was also associated with the expression of the ERBB2 receptor (OMIM no. 164870). Expression of ERBB2 is known for its association with aggressive tumors and worse prognosis in patients with breast cancer [**41**]. Thus, the CGA haplotype of rs2239911-rs2239910-rs8079943 SNPs in *SLC46A1* seems to predict worse prognosis for the patients because it was associated with shorter DFS and positive expression of epidermal growth factor receptor ERBB2.

*SLC46A1* (OMIM no. 611672) encodes a protein identified as a proton-coupled folate transporter (PCFT, UniProtKB entry Q96NT5) [**16**] and together with *SLC19A1* (OMIM no. 600424) is responsible for folate malabsorption and hereditary folic acid deficiency anemia. *SLC46A1* SNPs rs37514694 and rs739439 were significantly associated with plasma levels of high-density lipoprotein in a previous study [**42**] and with an increased risk of methotrexate-related overall toxicity in patients with rheumatoid arthritis [**43**]. Additionally, *SLC46A1* SNP rs2239907 can regulate plasma homocysteine levels via interaction with a lifestyle, for example, smoking and alcohol intake [**44**]. SLC46A1 is not known to translocate tamoxifen or aromatase inhibitors used for the treatment of patients with breast cancer. Thus, the observed prognostic association may be attributed to its effect on physiological homeostasis, for example, dysregulation of folate uptake in cancer cells. Such dysregulation was described in an estrogen receptor-expressing breast cancer cell model [**45**]. A subsequent report demonstrated that folate can modulate cell type-specific gene dysregulation, consistent with the proliferative phenotype, potentially involving differential activities of folate transporters [**46**]. Regulatory mechanisms influencing expression and activity of these transporters could be theoretically modified by the interaction of the *SLC46A1* CGA haplotype with folate homeostasis. Another explanation is that the CGA haplotype may act just as a correlative type of biomarker, i.e., correlating with some other phenomenon of causal nature (modulating breast cancer progression differently in CGA carriers). This assumption is further supported by the lack of significant eQTL (Fig. 9 of the ESM). Taken together, associations of *SLC46A1* SNPs with the expression of ERBB2 and DFS of patients with breast carcinoma, suggested by the present study, should be further investigated in terms of function of rs2239911, rs2239910, and rs8079943 SNPs in breast tumor cells.

*SLCO1A2* (OMIM no. 602883) encoding organic anion transporting polypeptide 1A2 protein (UniProtKB entry P46721) is a sodium-independent transporter responsible for cellular uptake of organic ions in the liver including bile salts and steroid conjugates, xenobiotics, and pharmacological drugs, for example, imatinib, fexofenadine, methotrexate, human immunodeficiency virus protease inhibitors, and HMG-CoA reductase inhibitors [**19**] (**www.genec ards.org**). Alternative splicing results in the production of multiple transcript variants (data provided by the National Center for Biotechnology Information's RefSeq, Dec 2008). *SLCO1A2* is associated with diverse diseases, for example, aneurysmal bone cysts and intrahepatic cholestasis. Results of functional impacts of

different *SLCO1A2* SNPs, including transport of imatinib, methotrexate, and estrone-3-sulfate are summarized in Zhou et al. [**47**]. A pharmacogenetics study reported association of the *SLCO1A2* haplotype T-T (rs4762699-rs2857468) with febrile neutropenia in patients with breast cancer receiving docetaxel (n = 627, odds ratio = 6.8, *p* = 1.15e-4) [**48**] indirectly implicating relevance to the prognosis of patients.

According to in silico predictions performed by the present study, SNP rs11568563 in *SLCO1A2* may have functional consequences. The rs11568563 (E172D) was found in 14 of 105 patients in the heterozygous state, which is in line with the MAF found in the Caucasian European population sample (MAF = 0.1). Single nucleotide polymorphism rs11568563 is located in the transmembrane domain 4 of *SLCO1A2* [**49**]. Six *SLCO1A2* nonsynonymous SNPs were previously identified in 95 healthy volunteers and functionally characterized in Hela cells in vitro. Two of these variants (rs11568563 and rs45502302, N135I) markedly reduced transport activity toward all three substrates tested, i.e., estrone 3-sulfate, deltorphin II, and delta-opioid receptor agonist [**49**]. Moreover, the rs11568563 variant, evaluated by the present study, causes decreased methotrexate [**50**] and imatinib uptake [**51**] and is thus functionally relevant. However, the present study has not observed a predictive or prognostic role of *SLCO1A2* on a genetic basis.

Our bioinformatics analyses on publicly available, large patient datasets showed the connection between *SLC46A1* and *SLCO1A2* gene expression and breast cancer PAM50 subtypes and confirmed their relevance for patient prognosis. The involvement of both genes in the methotrexate pathway was also highlighted by analysis of the PharmGKB database [**39**], suggesting potential pharmacogenomics-related consequences for cancer therapy. Thus, experimental evaluation of relationships between clinically relevant genotypes/hap- lotypes and phenotypes in both genes should be a priority for further study.

A modest sample size in the exploratory phase may be seen as a major limitation of this study. Because of this fact, rare (MAF = 1-5%) and very rare (MAF <1%) variants in the studied genes could have been missed. Larger replication studies in ethnically diverse populations will be needed to confirm present observations. The lack of knowledge about consequences of the *SLC46A1* rs2239911-rs2239910- rs8079943 haplotype for function of the enzyme presents another limitation. Our ongoing research is now focused on the fulfillment of these gaps.

## 5 Conclusions

The present study provides new information within the SLC protein superfamily field justifying further testing of the importance of genetic variability in *SLC46A1* and *SLCO1A2* transporters for outcomes in patients with cancer. Functionally relevant variants in *SLC46A1* and *SLCO1A2* genes were identified and validated in a large cohort of patients with breast cancer. The *SLC46A1* hap- lotype (rs2239911-rs2239910-rs8079943) with the clinical significance for breast cancer prognosis was revealed and its clinical utility should be further explored.

**References**


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424.
2. Videira M, Reis RL, Brito MA. Deconstructing breast cancer cell biology and the mechanisms of multidrug resistance. Biochim Biophys Acta. 2014;1846(2):312-25.
3. El-Gebali S, Bentz S, Hediger M, Anderle P. Solute carriers (SLCs) in cancer. Mol Aspects Med. 2013;34(2-3):719-34.
4. Szakacs G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottes- man MM. Targeting multidrug resistance in cancer. Nat Rev Drug Discov. 2006;5(3):219-34.
5. Lemstrova R, Soucek P, Melichar B, Mohelnikova-Duchonova B. Role of solute carrier transporters in pancreatic cancer: a review. Pharmacogenomics. 2014;15(8):1133-45.
6. Mohelnikova-Duchonova B, Melichar B, Soucek P. FOL- FOX/FOLFIRI pharmacogenetics: the call for a personalized approach in colorectal cancer therapy. World J Gastroenterol. 2014;20(30):10316-30.
7. Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, Liska V, Pitule P, Novak P, et al. The role of ABC transporters in progression and clinical outcome of colorectal cancer. Mutagenesis. 2012;27(2):187-96.
8. Hlavac V, Brynychova V, Vaclavikova R, Ehrlichova M, Vrana
   Pecha V, et al. The expression profile of ATP-binding cassette transporter genes in breast carcinoma. Pharmacogenomics. 2013;14(5):515-29.
9. Elsnerova K, Mohelnikova-Duchonova B, Cerovska E, Ehrlichova M, Gut I, Rob L, et al. Gene expression of membrane transporters: importance for prognosis and progression of ovarian carcinoma. Oncol Rep. 2016;35(4):2159-70.
10. Mohelnikova-Duchonova B, Brynychova V, Hlavac V, Kocik M, Oliverius M, Hlavsa J, et al. The association between the expression of solute carrier transporters and the prognosis of pancreatic cancer. Cancer Chemother Pharmacol. 2013;72(3):669-82.
11. Mohelnikova-Duchonova B, Brynychova V, Oliverius M, Honsova
    Kala Z, Muckova K, et al. Differences in transcript levels of ABC transporters between pancreatic adenocarcinoma and nonneoplastic tissues. Pancreas. 2013;42(4):707-16.
12. Cesar-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyi- mesi G, Bai X, et al. A call for systematic research on solute carriers. Cell. 2015;162(3):478-87.
13. Zemanek T, Melichar B, Lovecek M, Soucek P, Mohelnikova- Duchonova B. Biomarkers and pathways of chemoresistance and chemosensitivity for personalized treatment of pancreatic adenocarcinoma. Pharmacogenomics. 2019;20(2):113-27.
14. Baldwin SA, Beal PR, Yao SY, King AE, Cass CE, Young JD. The equilibrative nucleoside transporter family, SLC29. Pflug Arch. 2004;447:735-43.
15. Gray JH, Owen RP, Giacomini KM. The concentrative nucleoside transporter family, SLC28. Pflug Arch. 2004;447:728-34.
16. Qiu A, Jansen M, Sakaris A, Min SH, Chattopadhyay S, Tsai E, et al. Identification of an intestinal folate transporter and the molecular basis for hereditary folate malabsorption. Cell. 2006;127:917-28.
17. Sprowl JA, Mikkelsen TS, Giovinazzo H, Sparreboom A. Contribution of tumoral and host solute carriers to clinical drug response. Drug Resist Updates. 2012;15:5-20.
18. Angelini S, Pantaleo MA, Ravegnini G, Zenesini C, Cavrini G, Nannini M, et al. Polymorphisms in OCTN1 and OCTN2 transporters genes are associated with prolonged time to progression in unresectable gastrointestinal stromal tumours treated with imatinib therapy. Pharm Res. 2013;68:1-6.
19. Nakanishi T, Tamai I. Putative roles of organic anion transporting polypeptides (OATPs) in cell survival and progression of human cancers. Biopharm Drug Dispos. 2014;35(8):463-84.

20. Sutherland R, Meeson A, Lowes S. Solute transporters and malignancy: establishing the role of uptake transporters in breast cancer and breast cancer metastasis. Cancer Metast Rev. 2020;39(3):919-32.
21. Hubackova M, Vaclavikova R, Ehrlichova M, Mrhalova M, Kodet R, Kubackova K, et al. Association of superoxide dismutases and NAD(P)H quinone oxidoreductases with prognosis of patients with breast carcinomas. Int J Cancer. 2012;130:338-48.
22. Hlavac V, Kovacova M, Elsnerova K, Brynychova V, Kozevniko- vova R, Raus K, et al. Use of germline genetic variability for prediction of chemoresistance and prognosis of breast cancer patients. Cancers. 2018;10:511.
23. Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: a review of validation studies on tumour assessment. Eur J Cancer. 2006;42(8):1031-9.
24. Topic E, Gluhak J. Isolation of restrictible DNA. Eur J Clin Chem Clin Biochem. 1991;29(5):327-30.
25. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611-22.
26. Vaclavikova R, Klajic J, Brynychova V, Elsnerova K, Alnaes GIG, Tost J, et al. Development of high resolution melting analysis for ABCB1 promoter methylation: clinical consequences in breast and ovarian carcinoma. Oncol Rep. 2019;42(2):763-74.
27. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-4.
28. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):l1.
29. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. BMC Genom. 2015;16:636.
30. Nagy A, Lanczky A, Menyhart O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep. 2018;8:9227.
31. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinform. 2013;43(1110):11.10.1-33.
32. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blu- menstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002;296(5576):2225-9.
33. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790-7.
34. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
35. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11(1):1-9.
36. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930-4.
37. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289-300.
39. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharm Ther. 2012;92(4):414-7.

40. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chit- saz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43(Database issue):D222-6.
41. Jarman EJ, Ward C, Turnbull AK, Martinez-Perez C, Meehan J, Xintaropoulou C, et al. HER2 regulates HIF-2alpha and drives an increased hypoxic response in breast cancer. Breast Cancer Res. 2019;21(1):10.
42. Clifford AJ, Rincon G, Owens JE, Medrano JF, Moshfegh AJ, Baer DJ, et al. Single nucleotide polymorphisms in CETP, SLC46A1, SLC19A1, CD36, BCMO1, APOA5, and ABCA1 are significant predictors of plasma HDL in healthy adults. Lipids Health Dis. 2013;12:66.
43. Lima A, Bernardes M, Azevedo R, Monteiro J, Sousa H, Medeiros R, et al. SLC19A1, SLC46A1 and SLCO1B1 polymorphisms as predictors of methotrexate-related toxicity in Portuguese rheumatoid arthritis patients. Toxicol Sci. 2014;142(1):196-209.
44. Huang T, Tucker KL, Lee YC, Crott JW, Parnell LD, Shen J, et al. Interactions between genetic variants of folate metabolism genes and lifestyle affect plasma homocysteine concentrations in the Boston Puerto Rican population. Public Health Nutr. 2011;14(10):1805-12.
45. Kelley KM, Rowan BG, Ratnam M. Modulation of the folate receptor alpha gene by the estrogen receptor: mechanism and implications in tumor targeting. Cancer Res. 2003;63(11):2820-8.
46. Price RJ, Lillycrop KA, Burdge GC. Folic acid induces cell type- specific changes in the transcriptome of breast cancer cell lines: a proof-of-concept study. J Nutr Sci. 2016;5:e17.
47. Zhou F, Zhu L, Wang K, Murray M. Recent advance in the phar- macogenomics of human solute carrier transporters (SLCs) in drug disposition. Adv Drug Deliv Rev. 2017;116:21-36.
48. Callens C, Debled M, Delord M, Turbiez-Stalain I, Veyret C, Bieche I, Brain E. High-throughput pharmacogenetics identifies SLCO1A2 polymorphisms as candidates to elucidate the risk of febrile neutropenia in the breast cancer RAPP-01 trial. Breast Cancer Res Treat. 2015;153(2):383-9.
49. Lee W, Glaeser H, Smith LH, Roberts RL, Moeckel GW, Ger- vasini G, et al. Polymorphisms in human organic anion-transporting polypeptide 1A2 (OATP1A2): implications for altered drug disposition and central nervous system drug entry. J Biol Chem. 2005;280(10):9610-7.
50. Badagnani I, Castro RA, Taylor TR, Brett CM, Huang CC, Stryke D, et al. Interaction of methotrexate with organic-anion transporting polypeptide 1A2 and its genetic variants. J Pharmacol Exp Ther. 2006;318(2):521-9.
51. Eechoute K, Franke RM, Loos WJ, Scherkenbach LA, Boere I, Verweij J, et al. Environmental and genetic factors affecting transport of imatinib by OATP1A2. Clin Pharmacol Ther. 2011;89(6):816-20.