Anastrozole Aromatase Inhibitor Plasma Drug Concentration Genome-Wide Association Study: Functional Epistatic Interaction Between *SLC38A7* and *ALPPL2*

Tanda M. Dudenkov[1,†], Duan Liu[1,†] (iD), Junmei Cairns[1,†] (iD), Sandhya Devarajan[1], Yongxian Zhuang[1], James N. Ingle[2] (iD), Aman U. Buzdar[3], Mark E. Robson[4], Michiaki Kubo[5], Anthony Batzler[6], Poulami Barman[6], Gregory D. Jenkins[6], Erin E. Carlson[6], Matthew P. Goetz[2], Donald W. Northfelt[7], Alvaro Moreno-Aspitia[8], Zeruesenay Desta[9], Joel M. Reid[1], Krishna R. Kalari[6], Liewei Wang[1,*] and Richard M. Weinshilboum[1,*] (iD)

Anastrozole is a widely prescribed aromatase inhibitor for the therapy of estrogen receptor positive (ER+) breast cancer. We performed a genome-wide association study (GWAS) for plasma anastrozole concentrations in 687 postmenopausal women with ER+ breast cancer. The top single-nucleotide polymorphism (SNP) signal mapped across *SLC38A7* (rs11648166, P = 2.3E-08), which we showed to encode an anastrozole influx transporter. The second most significant signal (rs28845026, P = 5.4E-08) mapped near *ALPPL2* and displayed epistasis with the *SLC38A7* signal. Both of these SNPs were cis expression quantitative trait loci (eQTL)s for these genes, and patients homozygous for variant genotypes for both SNPs had the highest drug concentrations, the highest *SLC38A7* expression, and the lowest *ALPPL2* expression. In summary, our GWAS identified a novel gene encoding an anastrozole transporter, *SLC38A7*, as well as epistatic interaction between SNPs in that gene and SNPs near *ALPPL2* that influenced both the expression of the transporter and anastrozole plasma concentrations.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Anastrozole is a highly prescribed aromatase inhibitor that plays an important role in the management of postmenopausal women with endocrine therapy–sensitive breast cancer. However, no biomarkers for variation in plasma anastrozole concentrations have been reported, and no genome-wide association study for plasma anastrozole concentrations has been performed.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study addressed the question of whether genetics might contribute to variation in plasma anastrozole concentrations in patients with hormone receptor positive breast cancer being treated with this drug.

WHAT DOES THIS STUDY ADD TO OUR KNOW-LEDGE?

☑ The rs11648166 SNP in *SLC38A7* and the rs28845026 SNP near *ALPPL2* were associated with plasma anastrozole concentrations in patients with estrogen receptor positive breast cancer who received anastrozole therapy. SLC38A7 is a transporter for anastrozole, and ALPPL2, by an epistatic mechanism, seems to repress SLC38A7 expression and, as a result, decreases anastrozole plasma concentrations.

HOW MIGHT THIS CHANGE CLINICAL PHARMA-COLOGY OR TRANSLATIONAL SCIENCE?

☑ This study provides genetic markers, single-nucleotide polymorphisms, for the possible individualization of plasma anastrozole dosing and, as a result, plasma drug concentrations. It also highlights the potential importance of epistasis as a pharmacogenomic mechanism.

[1]Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA; [2]Division of Medical Oncology, Department of Oncology, Mayo Clinic, Rochester, Minnesota, USA; [3]Department of Breast Oncology, MD Anderson Cancer Center, Houston, Texas, USA; [4]Breast Medicine Service, Memorial Sloan Kettering Cancer Center, New York, New York, USA; [5]RIKEN Center for Integrative Medical Sciences, Yokohama City, Japan; [6]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA; [7]Division of Hematology/Oncology, Department of Internal Medicine, Mayo Clinic, Scottsdale, Arizona, USA; [8]Division of Hematology/Oncology, Department of Internal Medicine, Mayo Clinic, Jacksonville, Florida, USA; [9]Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA. *Correspondence: Liewei Wang (wang.liewei@mayo.edu) and Richard M. Weinshilboum (weinshilboum.richard@mayo.edu)

[†]These three authors contributed equally to this paper.

Aromatase inhibitors (AIs) are first-line adjuvant endocrine therapy for postmenopausal women with estrogen receptor positive (ER+) breast cancer.[1,2] Anastrozole is the most widely prescribed AI in the United States,[3] and it was recently recommended in the United Kingdom for the prevention of breast cancer in high-risk women.[4] Anastrozole reversibly and competitively inhibits aromatase, the enzyme that catalyzes the biosynthesis of estrogens from androgens. The US Food and Drug Administration–approved dose of anastrozole is 1 mg/day, which results in average reductions of estradiol (E2) and estrone (E1) plasma concentrations of $\geq 83.5\%$ and $\geq 86.5\%$, respectively.[5]

Anastrozole undergoes both phase I and phase II metabolism.[6] A large proportion of plasma anastrozole, a hydrophilic compound, is not metabolized,[7,8] with mean anastrozole concentrations that are 30-fold higher than are those of its most abundant metabolite, anastrozole-*N*-glucuronide.[8] However, there is large interindividual variability in plasma anastrozole concentrations and in the occurrence of anastrozole treatment–related phenotypes. Specifically, not all patients experience estrogen suppression during anastrozole therapy and, surprisingly, ~ 10% of patients experience either no change or an increase in plasma estrogen concentrations despite being adherent to therapy.[9] Of greatest clinical importance, ~ 10% of patients receiving anastrozole experience breast cancer recurrence within 5 years, and there is also wide variation in the occurrence and severity of adverse drug reactions.[10] Finally, no clear mechanistic links have been established between anastrozole plasma concentrations and long-term clinical outcomes and/or the occurrence of adverse reactions to this important antineoplastic drug.

In the present study, we set out to test the hypothesis that plasma anastrozole concentrations might be regulated, at least in part, by genetics. To test that hypothesis, we performed a genome-wide association study (GWAS) of plasma anastrozole concentrations in 687 postmenopausal patients with ER+ breast cancer enrolled in the Mayo Clinic-MD Anderson-Memorial Sloan Kettering (Mayo-MDA-MSK) Pharmacogenomics Research Network (PGRN) Breast Cancer Aromatase Inhibitor Pharmacogenomics study. As described subsequently, we observed two major single-nucleotide polymorphism (SNP) signals—a genome-wide significant SNP signal across a transporter gene, *SLC38A7*, and a near genome-wide significant SNP signal that mapped near *ALPPL2* as well as a novel epistatic relationship (i.e., the genotype of one locus influenced the phenotypic expression of the other locus) between these two SNP signals. These results provide novel insight into
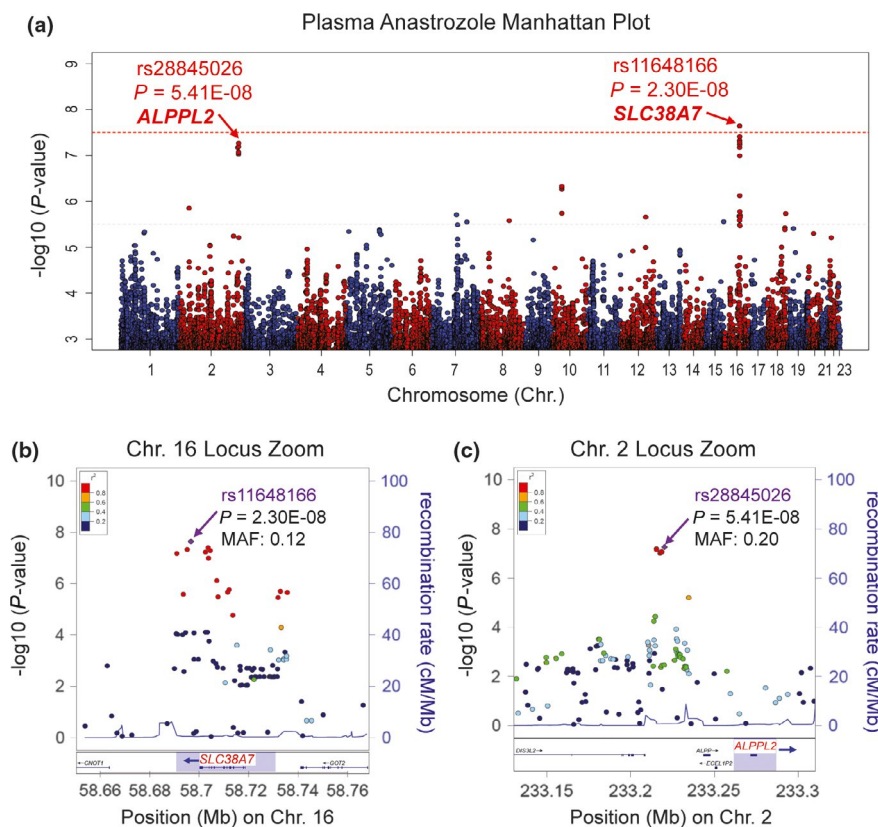

Figure 1 Genome-wide association study (GWAS) for plasma anastrozole concentrations. (**a**) Manhattan plot of plasma anastrozole concentration GWAS. The *P* values for "top single-nucleotide polymorphisms (SNP)s" in the chromosome 2 and 16 signals are indicated. The dashed line represents the threshold for genome-wide significance ($P \leq 5.0E\text{-}08$). Regional association plots (Locus Zooms) for the two top signals on (**b**) chromosome 16 and (**c**) chromosome 2 are also shown. The "top SNPs" for those two GWAS signals, the rs11648166 SNP with a genome-wide significant *P* value that mapped to 3′-end of *SCL38A7* on chromosome 16 in **b** and the rs28845026 SNP on chromosome 2 near a group of genes that included *ALPPL2* in **c**, are shown in purple. The red to dark blue scale of colors indicates the degree of linkage disequilibrium (LD) with the "top SNP." Red indicates perfect LD, and dark blue indicates no LD. Minor allele frequencies (MAFs) for the two "top SNPs" in the study cohort are also indicated.

Table 1 Top 10 SNPs from the GWAS for anastrozole plasma concentrations

SNP	Chr	Minor allele	MAF	Unadjusted model			Adjusted model			Gene	SNP location	Cis-eQTL gene(s)
				Effect size	Direction of effect	P value	Effect size	Direction of effect	P value			
rs11648166	**16**	**G**	**0.12**	**0.44**	**↑**	**2.30E-08**	**0.42**	**↑**	**1.80E-07**	**SLC38A7**	**3 kb Downstream**	**SLC38A7**
rs3784911	16	T	0.12	0.44	↑	3.90E-08	0.41	↑	3.10E-07	SLC38A7	Intronic	SLC38A7
rs60979655	16	T	0.11	0.44	↑	4.70E-08	0.42	↑	2.70E-07	SLC38A7	5 kb Downstream	SLC38A7
rs16960359	16	A	0.12	0.43	↑	5.20E-08	0.41	↑	4.00E-07	SLC38A7	Intronic	SLC38A7
rs28845026	**2**	**T**	**0.20**	**0.37**	**↑**	**5.41E-08**	**0.35**	**↑**	**8.90E-07**	**—**	**Intergenic**	**ALPPL2, ECEL1P2**
rs76733922	16	C	0.12	0.43	↑	5.80E-08	0.4	↑	4.90E-07	SLC38A7	Intronic	SLC38A7
rs883013	2	T	0.19	0.36	↑	6.40E-08	0.34	↑	1.10E-06	—	Intergenic	ALPPL2, ECEL1P2
rs12149056	16	A	0.12	0.43	↑	6.70E-08	0.42	↑	3.60E-07	SLC38A7	8 kb Downstream	SLC38A7
rs4996188	2	A	0.19	0.36	↑	6.80E-08	0.34	↑	1.20E-06	—	Intergenic	ALPPL2, ECEL1P2
rs6715083	2	T	0.18	0.36	↑	8.40E-08	0.34	↑	1.40E-06	—	Intergenic	ALPPL2, ECEL1P2

The bold SNPs were the "top" hits for each of the two signals.
Chr, chromosome number; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; MAF, minor allele frequency; SNPs, single-nucleotide polymorphisms.

genomic influences on individual variation in plasma anastrozole concentrations, and they should also heighten sensitivity to the possibility that two separate genetic loci might interact to influence the same pharmacogenomic phenotype, in this case, plasma drug concentration.

RESULTS

Plasma anastrozole GWAS

Plasma anastrozole concentrations in these patients with ER+ breast cancer were measured at a time when they had been on the once-a-day dose of the drug for several months, and patients were instructed not to take their dose of anastrozole for that day until after their blood sample had been drawn. Patient information for participants in the study is summarized in **Table S1**. The Manhattan plot for the GWAS for plasma anastrozole concentrations is shown in **Figure 1a**, and the QQ-plots are shown in **Figure S1**. The "top" 10 SNPs with low P values in the GWAS are listed in **Table 1**, and a list of SNPs with P value < 5.0E-06 is shown in **Table S2**. A genome-wide significant signal was observed across the *SLC38A7* transporter gene. The top SNP in that signal, rs11648166, had a minor allele frequency of 0.12 in our cohort, and the variant allele was associated with increased plasma concentrations of anastrozole (P = 2.3E-08; **Figure 1b, Table 1**). This SNP was located ~ 3 kb downstream (3′) of *SLC38A7* and was an expression quantitative trait locus (eQTL) for that gene in multiple tissues, with the minor allele being associated with increased gene expression in several tissues, including esophageal mucosa, ovaries, etc. (lowest P = 5.7E-06 in thyroid tissue, Genotype-Tissue Expression (GTEx)[11]). *SLC38A7* is known to encode a glutamine plasma membrane influx transporter,[12] but it has not been reported to transport drugs, including anastrozole. Anastrozole is a hydrophilic compound and is unable to diffuse across the cell membrane. However, until the present study, no transporter for anastrozole had been reported. These GWAS observations raised the question of how increased expression of an influx transporter that would be expected to move drug into cells, as we demonstrated to be the case in the functional genomic experiments described subsequently, might be associated with higher rather than lower plasma drug concentrations. One possible explanation would be the transport of this orally administered drug across the gut mucosa from the gastrointestinal tract into the circulation, a possibility that is addressed subsequently.

The next question involved the role of the near genome-wide significant signal that mapped to chromosome 2. The top SNP in that signal was rs28845026 (**Figure 1c**). That SNP had a minor allele frequency of 0.20 in our sample and the minor allele, like that for rs11648166 in *SLC38A7*, was associated with increased plasma anastrozole concentrations (P = 5.4E-08; **Table 1**) and was an eQTL for several of the genes that mapped near the SNP, including *ECEL1* and *ALPPL2*. According to GTEx, rs28845026 is an eQTL for *ALPPL2* in transverse colon tissue with a P value of 9.7E-05, with the variant allele being associated with decreased ALPPL2 expression[11] (**Figure S2**). The functional genomic studies described subsequently indicated that ALPPL2 had a significant effect on both SLC38A7 expression and anastrozole transport.
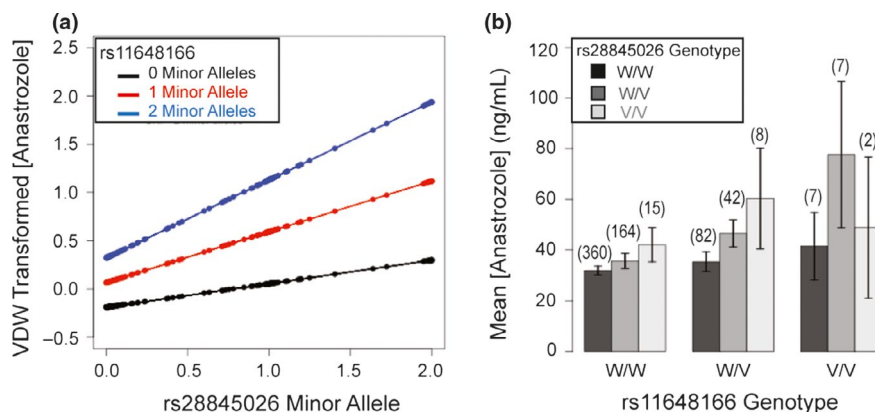
Figure 2 Plasma anastrozole concentration vs. genotypes for the chromosome 16 (rs11648166) and chromosome 2 (rs28845026) single-nucleotide polymorphisms (SNPs). (**a**) Plot of predicted anastrozole concentration by genotypes for the top 2 SNPs from the interaction model. The *y*-axis represents predicted Van der Waerden (VDW) transformed plasma anastrozole concentrations, a normally distributed quantitative phenotype that was used in the interaction model for the multivariate linear regression analysis. Negative values on the *y*-axis correspond to actual anastrozole concentrations that lie below the median. (**b**) Bar plots of observed plasma anastrozole concentrations by genotypes for the 2 SNPs. Values are mean ± SEM. The number of patient samples for each genotype combination is listed above the bars. [Colour figure can be viewed at wileyonlinelibrary.com]

Epistatic interaction between chromosome 16 and chromosome 2 SNPs

As a step toward addressing the questions raised in preceding paragraphs, we created a linear regression model for possible additive SNP effects to test the possibility that there might be interaction between the top SNPs for the chromosome 16 signal (rs11648166) and the chromosome 2 GWAS signal (rs28845026). This regression model tested the hypothesis that the effect of the rs11648166 SNP on plasma anastrozole concentrations might be modified by genotype for the rs28845026 SNP. The interaction *P* value was 0.0227 (**Table S3**). Therefore, in addition to the main effects of these SNPs, the effect of each SNP on anastrozole concentration also varied depending on the genotype of the other SNP. Based on the predictions of this model, subjects carrying minor alleles for both SNPs would have higher anastrozole concentrations than would be predicted based on the effect of each allele separately. The results of the predicted allele-dependent effects predicted by the interaction model are shown graphically in **Figure 2a**, and the actual phenotype-dependent relationship to genotypes for the two alleles are shown graphically in **Figure 2b**. The model for this SNP–SNP interaction explained 8.5% of the variation in plasma drug concentrations, whereas individually, rs11648166 explained 4.4% of the variation in anastrozole concentration and rs28845026 explained 4.2%.

Anastrozole concentrations in this group of patients were negatively associated with the probability of having detectable levels of either E1 (*P* = 8.5E-07) or E2 (*P* = 2.8E-06) in their plasma because the mechanism of anastrozole action involves inhibition of the enzyme that catalyzes the biosynthesis of these hormones. Specifically, patients with undetectable levels of E1 or E2 had significantly higher mean plasma anastrozole concentrations (37.1 and 36.8 ng/mL, respectively) than did patients who had detectable levels of plasma E1 or E2 (30.7 and 29.7 ng/mL, respectively; **Figure 3**). Scatter plots for plasma anastrozole and E1 or E2 concentrations for all patients are shown in **Figure S3**.

The Mayo-MDA-MSK study had been designed to focus on the study of endophenotypes (e.g., hormone and drug concentrations) but was not powered to study long-term clinical outcomes, such as recurrence-free survival. Therefore, we were unable to test the association of the SNPs that we identified during our GWAS with clinical outcomes. However, we were able to study mechanism(s) by which these SNPs might influence plasma anastrozole concentrations as well as the epistatic interaction between the two SNP signals, so we next set out to study possible mechanisms that might explain these observations.

Functional characterization of chromosome 16 and chromosome 2 SNP mechanisms

The rs11648166 SNP on chromosome 16 mapped to the *SLC38A7* gene, which is known to encode a glutamine influx transporter,[12,13] but SLC38A7 has not been reported to transport drugs, including anastrozole. In fact, no transporter for anastrozole had been reported previously. Therefore, our initial functional studies were designed to address the questions of whether *SLC38A7* encoded a protein capable of transporting anastrozole into cells and, if that were the case, why might this SNP—for which the variant genotype was an eQTL associated with increased transporter expression—be associated with increased rather than decreased plasma drug concentrations. The "other SNP signal," the chromosome 2 rs28845026 SNP, mapped to an intergenic region that was surrounded by the *DIS3L2*, *ALPP*, *ECEL1P2*, and *ALPPL2* genes (**Figure 1c**). That SNP was a cis eQTL for *ALPPL2* in the colon (*P* = 9.7E-05; **Figure S2**) and a cis eQTL for *ECEL1P2* in the thyroid but, based on GTEx data, it was not an eQTL for the *DIS3L2* or *ALPP* genes in any tissue.[14] *ALPP* and *ALPPL2* both encode membrane-bound alkaline phosphatases that are 98% identical in amino acid sequence. Both alkaline phosphatases catalyze the dephosphorylation of a wide variety of phosphate monoesters,[15,16] and they are expressed in the placenta, cervix, uterus, fallopian tubes, lungs, and a variety
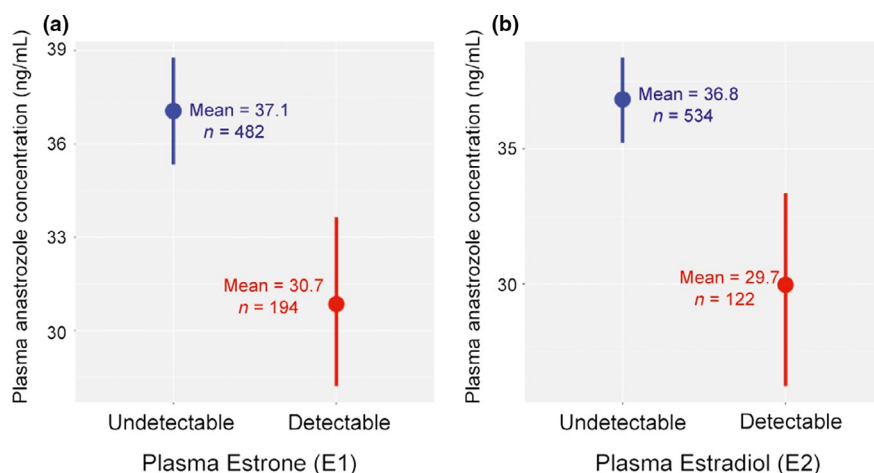
Figure 3 Associations of plasma anastrozole concentrations and the detectability of plasma estrogens. Higher anastrozole plasma concentrations were associated with undetectable concentrations of (**a**) estrone (E1; $P = 8.5E\text{-}07$) and (**b**) estradiol (E2; $P = 2.8E\text{-}06$) during anastrozole therapy. The points represent mean anastrozole concentrations, and the lines represent the upper and lower limits of the 95% confidence intervals. The lower limit of quantitation (LLQ) for E1 was 1.56 pg/mL and for E2 was 0.625 pg/mL. The mean LLQ in all samples combined was calculated as 1.64 pg/mL for E1 and 0.66 pg/mL for E2.
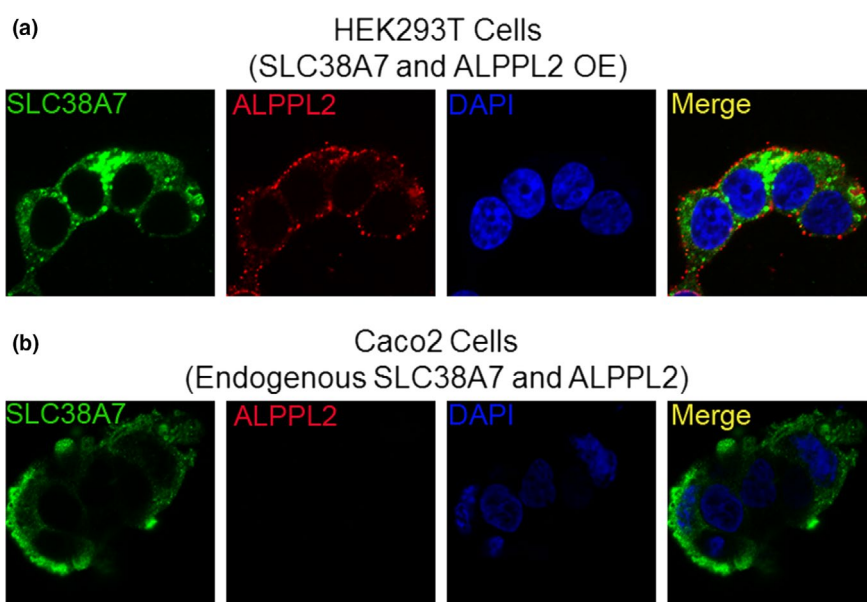
Figure 4 Immunofluorescence studies of SLC38A7 and ALPPL2 in (**a**) HEK293T cells transfected with cDNA plasmids and (**b**) Caco2 cells that had not been transfected. SLC38A7 and ALPPL2 antibodies were visualized by secondary antibodies that were labeled with green and red fluorescence, respectively. Nuclei were stained by 4′,6-diamidino-2-phenylindole, showing blue fluorescence. SLC38A7 protein, either overexpressed in HEK293T cells or endogenously expressed in Caco2 cells, displayed plasma membrane and cytosol localization. Overexpressed ALPPL2 localized to the plasma membrane in HEK293T cells. No endogenous ALPPL2 protein was detected in Caco2 cells. The cells were not permeabilized by Triton X-100.

of tumors.[17,18] In addition, ALPPL2, but not ALPP, is expressed in the gut, including the small intestine (both duodenum and terminal ileum),[14] an expression pattern that might influence drug absorption. Because ALPPL2 is expressed in the gut and because rs28845026 is an eQTL for *ALPPL2*, we focused our attention in subsequent experiments on the *ALPPL2* gene.

HEK293T cells were chosen to study the potential roles of SLC38A7 and ALPPL2 in anastrozole transport because those

cells did not express either protein. Because of their low background expression of drug transporters, HEK293T cells are often used to study specific drug-transporter pairs by overexpressing the transporter, followed by exposure to the drug.[19] Immunofluorescence (IF) studies performed with *SLC38A7* and *ALPPL2* cDNA transfected HEK293T cells showed that both SLC38A7 and ALPPL2 localized to the plasma membrane of these cells (**Figure 4a**). Cells transfected with empty vector displayed no staining for either
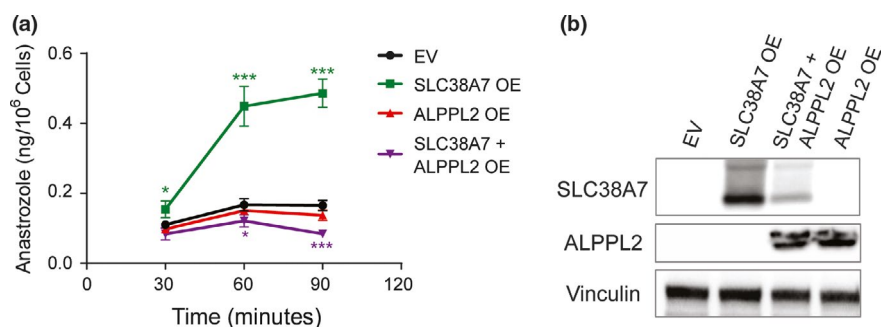
Figure 5 Effect of SLC38A7 and ALPPL2 on anastrozole uptake. (**a**) Results for anastrozole uptake assays in HEK293T cells after the overexpression (OE) of SLC38A7, ALPPL2, or both after treatment with 50 μM anastrozole. Values are mean ± SD; $N = 5$; * $P < 0.05$, ***$P < 0.001$. (**b**) Western blots after the OE of either SLC38A7, ALPPL2, or both together in HEK293T cells. EV, empty vector.

SLC38A7 or ALPPL2 (data not shown). IF studies were also performed with Caco2 cells, a cell line that has been used as a model for intestinal transport.[19] Caco2 cells displayed high levels of endogenous SLC38A7 expression (**Figure 4b**), a fact of which we took advantage, as described subsequently, to study interaction between ALPPL2 and endogenous SLC38A7. We did not use Caco2 cells to study anastrozole transport because they express several influx and efflux drug transporters, which would have complicated interpretation of transport assay results.

To test the hypothesis that SLC38A7 and/or ALPPL2 might influence anastrozole cellular uptake, both proteins were overexpressed, separately and together, in HEK293T cells. The cells were then exposed to anastrozole. Compared with cells transfected with empty vector, anastrozole concentrations in lysates from cells overexpressing SLC38A7 were significantly elevated after 60 and 90 minutes of incubation (**Figure 5a**), indicating anastrozole uptake by the cells. However, overexpression (OE) of ALPPL2 together with SLC38A7 prevented anastrozole uptake (**Figure 5a**), probably because of decreased SLC38A7 protein expression when SLC38A7 and ALPPL2 were co-overexpressed together when compared with SLC38A7 transfection alone (**Figure 5b**). These results were compatible with the conclusion that SLC38A7 is an inward transporter for anastrozole and that cotransfection with ALPPL2 decreased the expression of SLC38A7 and, thus, the

inward transport of anastrozole. These results were also compatible with the GWAS observation that minor alleles for both the *SLC38A7* and *ALPPL2* SNPs were associated with increased plasma anastrozole concentrations, as well as the directionality of expression associated with minor alleles for these two genes, which, as stated earlier, are opposite in direction. Specifically, the minor allele for the *SLC38A7* SNP was associated with increased SLC38A7 expression, but the minor allele for the *ALPPL2* SNP was associated with decreased ALPPL2 expression. We next turned our attention to possible mechanisms that might explain the epistatic interaction between the two loci identified by our GWAS.

ALPPL2 influences SLC38A7 expression

We began our functional genomic studies of interaction between these two loci by testing the hypothesis that *SLC38A7* gene expression might be correlated with the expression of one or more of the genes for which the rs28845026 SNP on chromosome 2 was an eQTL—having already shown that OE of ALPPL2 could result in decreased expression of SLC38A7 protein based on the Western blot data shown in **Figure 5b**. As a first step, we knocked down (KD) SLC38A7 in Caco2 cells. KD of SLC38A7 had no effect on the expression of mRNA for genes in the gene cluster surrounding the chromosome 2 rs28845026 SNP (**Figure 6a**), whereas KD of ECEL1, NPPC, and ALPP all resulted in either no significant
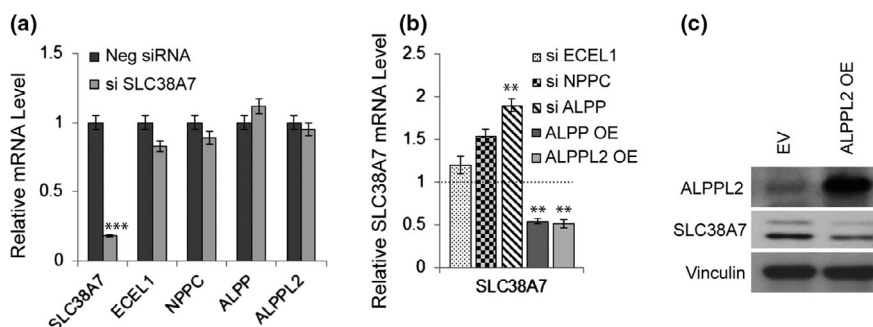

Figure 6 SLC38A7 and ALPPL2 interaction studies in Caco2 cells. (**a**) The mRNA levels for *SLC38A7* and genes near the chromosome 2 single-nucleotide polymorphism (SNP) signal (*ECEL1*, *NPPC*, *ALPP*, and *ALPPL2*) after the knockdown (KD) of SLC38A7 by siRNA transfection are shown. None of the chromosome 2 gene mRNA levels were affected after SLC38A7 KD in Caco2 cells. (**b**) Levels of SLC38A7 mRNA expression after the KD of the chromosome 2 genes and the overexpression (OE) of *ALPP* and ALPPL2 in Caco2 cells. Expression levels were normalized to controls for KD and OE, which are indicated by the dashed line. (**c**) SLC38A7 protein levels after ALPPL2 OE are shown by Western blot assay. **$P < 0.05$, ***$P < 0.001$ in triplicate experiments. EV, empty vector.

change or increased expression of SLC38A7 mRNA, and OE of ALPP and ALPPL2 resulted in significant decreases of SLC38A7 expression at the mRNA level (**Figure 6b**). Because of its low level of expression, it was not possible to KD ALPPL2 in these cells, but we were able to overexpress ALPPL2 in Caco2 cells, and we observed significant downregulation of the expression of both SLC38A7 mRNA and protein after the OE of ALPPL2 (**Figure 6b,c**). As mentioned previously, ALPPL2 and ALPP are 98% identical in their amino acid sequences, which may explain the decrease in SLC38A7 mRNA expression after OE of both ALPPL2 and ALPP in Caco2 cells (**Figure 6b**). However, only ALPPL2 is expressed in the gut, and the chromosome 2 SNP is an eQTL for ALPPL2 but not for ALPP. Therefore, ALPPL2 is the most likely candidate among the chromosome 2 genes to influence the variation in plasma anastrozole concentration that we studied with our GWAS. In summary, these experiments suggested that ALPPL2 might influence plasma anastrozole concentrations through an effect on SLC38A7 expression.

DISCUSSION

Our GWAS for plasma anastrozole concentrations in postmenopausal women with ER+ breast cancer yielded two major SNP signals. The first was genome-wide significant and mapped across *SLC38A7* on chromosome 16. The minor allele for the top SNP, rs11648166, was associated with increased plasma anastrozole concentrations and with increased SLC38A7 expression in several tissues.[11] *SLC38A7* had been reported to encode a polar amino acid influx transporter but, prior to the present studies, had not been reported to transport drugs. Anastrozole is a hydrophilic compound and, as a result, likely requires transport into cells. However, to date, no anastrozole transporter has been reported. We have demonstrated that SLC38A7 can mediate the transport of anastrozole across the plasma membrane (**Figure 5a**). In addition, we showed by IF assay that SLC38A7 is localized to the plasma membrane after OE in HEK293T cells (**Figure 4a**).

The second GWAS signal mapped to an intergenic region on chromosome 2 in the midst of a cluster of genes. Like the chromosome 16 signal, the minor allele for the top SNP, rs28845026, was associated with increased plasma anastrozole concentrations. This SNP mapped 5′ of the *ALPP* and *ALPPL2* genes (**Figure 1c**), genes that encode two alkaline phosphatase isozymes that are 98% identical in amino acid sequence. However, only ALPPL2 has been shown to be expressed in the gut where it might influence the absorption of an orally administered drug. The rs28845026 SNP is an eQTL for *ALPPL2* in the colon based on GTEx data. Unfortunately, the number of small intestinal samples that express ALPPL2 in GTEx was too small to perform eQTL analysis. The minor allele for the rs28845026 SNP is associated with increased plasma anastrozole concentration and with decreased ALPPL2 expression in the colon.[14] Our functional studies of ALPPL2 showed that OE of ALPPL2 decreased expression of SLC38A7 and anastrozole uptake by cells (**Figure 5**), a result consistent with our GWAS observations. In addition to *ALPPL2*, the *ALPI* gene, which encodes "alkaline phosphatase, intestinal," is highly expressed in the small intestine. The *ALPI* gene, which is not shown in **Figure 1c**, maps 45.4 kb downstream of the *ALPPL2* gene. However, the rs28845026 SNP is not an eQTL for *ALPI*

expression in any of the tissues included in the GTEx data. Therefore, ALPI is unlikely to influence the variation in plasma anastrozole concentrations that we studied in our GWAS.

Our analysis also revealed a significant interaction between the chromosome 16 and the chromosome 2 top SNPs. That is, the effect of one of these alleles on anastrozole concentration differed across genotypes for the other allele. That statistical observation raised the possibility of a functional relationship between *SLC38A7* and *ALPPL2*, a hypothesis that was supported by our demonstration that OE of ALPPL2 decreased the expression of SLC38A7 (**Figures 5 and 6**). According to our interaction model, individuals homozygous for the minor allele for both SNPs were expected to have the highest mean anastrozole concentrations, whereas those homozygous for the major alleles would have the lowest anastrozole concentrations. We concluded that subjects homozygous for both allele genotypes displayed the highest overall SLC38A7 expression due, at least in part, to low *ALPPL2* expression, and this fact explained why this group of women displayed the highest anastrozole plasma concentrations. Future studies will obviously be required to fully elucidate the mechanism(s) underlying the effect of ALPPL2 on SLC38A7 mRNA and protein expression.

In summary, we have performed a GWAS for plasma anastrozole concentrations in postmenopausal women with ER+ breast cancer. That GWAS resulted in the discovery of a novel anastrozole plasma membrane transporter, SLC38A7, as well as an epistatic relationship between the two top SNP signals—that near the *SLC38A7* gene on chromosome 16 and that near the *ALPPL2* gene on chromosome 2. Functional studies supported a chromosome 2 SNP by chromosome 16 SNP effect on SLC38A7 expression and showed that ALPPL2 influenced the expression and the anastrozole transport activity of SLC38A7. This series of observations, beyond their implications for the possible individualization of clinical dosing of anastrozole, an extremely important drug for the adjuvant treatment of ER+ breast cancer, also serve to emphasize the potentially important role of epistasis as a pharmacogenomic mechanism.

METHODS
Mayo-MDA-MSK PGRN AI pharmacogenomic trial patient samples

Postmenopausal patients with resected early stage ER+ breast cancer were recruited from the Mayo Clinic, MD Anderson Cancer Center, and Memorial Sloan Kettering Cancer Center as part of the National Institute of General Medical Sciences–funded PGRN Aromatase Inhibitor Pharmacogenomics Trial (Mayo-MDA-MSK). See previous publications[8,9,20] for a detailed description. All patients received the standard clinically approved dose of anastrozole of 1 mg/day. At least 4 weeks after the initiation of therapy, blood was drawn from each patient to assay anastrozole and anastrozole metabolite concentrations. Anastrozole reaches steady state in ~ 7 days.[21] Prior to the initiation of therapy, blood samples were drawn from each patient to extract DNA to genotype ~ 600K SNPs using the Ilumina 610 Quad and to quantitate pretreatment plasma E2 and E1 concentrations by gas chromatography–tandem mass spectroscopy. Details with regard to the liquid chromatography–tandem mass spectrometry quantification of plasma anastrozole concentrations and the gas chromatography–tandem mass spectroscopy quantification of plasma E1 and E2 concentrations can be found in Ingle *et al.*[8] The lower limit of quantitation for anastrozole was 100 pg/mL, for E1 was 1.56 pg/mL, and for E2 was 0.625 pg/mL. This trial was performed after approval by local institutional review boards in accordance

with assurances filed with and approved by the US Department of Health and Human Services. Written informed consent was provided by each patient before entry on study. See **Table S1** for patient characteristics.

Plasma anastrozole GWAS

Only patients known to be compliant to drug therapy were included in the GWAS. Therefore, subjects with undetectable levels of anastrozole as well as all assayed anastrozole metabolites (anastrozole glucuronide, hydroxyl-anastrozole, and hydroxyl-anastrozole glucuronide) were excluded from the analysis. An additional three extreme outliers subjects were excluded who had anastrozole metabolite concentrations greater than 6 SDs above the mean. A total of 687 women remained for the final analysis. The distribution of anastrozole concentrations was right skewed, so a Van der Waerden[22] transformation was applied to render the distribution normal. We used both simple and multiple linear regression models to identify clinical and demographic variables that were associated with anastrozole concentrations. For the GWAS analysis, we used a linear regression model with additive SNP effects, with European Ancestry captured by the first and second eigenvectors from a principal components analysis. The top two independent SNP signals from the GWAS were also assessed for a possible SNP–SNP interaction. Linear regression models and both line and bar plots were used to assess this relationship. All analyses were performed using R statistical computing software (version 3.0.2; https://www.r-project.org/) and Plink (version 1.07; http://zzz.bwh.harvard.edu/plink/).[23]

Anastrozole relationship with estrogen concentrations

To determine whether anastrozole plasma concentrations were associated with absolute E2 or E1 concentrations while patients were on treatment, we used data from the Mayo-MDA-MSK cohort. Specifically, we dichotomized E1 and E2 into subjects with undetectable levels (controls) of the hormone vs. those with detectable levels (cases) and used logistic regression models to test the association of anastrozole concentrations with this phenotype. We ran both crude models and those with baseline E1 or E2, body mass index, and recruitment site as covariates. In those models, we adjusted for baseline E1 or E2, body mass index, and recruitment site.

Transfection, IF, RNA quantification, and Western blot

HEK293T cells (ATCC, CRL-3216) were cultured and transfected with *SLC38A7* and *ALPPL2* cDNA plasmids (OriGene Technologies, Rockville, MD) separately and were also cotransfected with both cDNA plasmids using the Lipofectamine 3000 Reagent (Thermo Fisher Scientific, Waltham, MA). After 24 hours, cells for each cDNA transfection were split into several 60-mm plates and were incubated for an additional 24 hours for anastrozole transport assays. A small aliquot of *SLC38A7* and *ALPPL2* cDNA plasmid cotransfected cells were seeded into CultureWell Chambered Coverglass (Thermo Fisher Scientific) for IF study. To confirm the efficiency of OE, one plate of cells for each transfection was collected for protein analysis by Western blot (see **Supplementary Methods S1** for additional details).

Anastrozole transport studies

Anastrozole transport assays were modified based on previous studies.[24,25] Specifically, 48 hours after cDNA plasmid transfection, cells were washed once with warm Hank's Balanced Salt Solution (HBSS) buffer (HBSS buffered with 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid at pH 7.4) and were then preincubated with fresh HBSS buffer at 37°C for 30 minutes. After preincubation, cells were incubated in fresh HBSS buffer with 50 µM of anastrozole (Sigma-Aldrich, St. Louis, MO) for 30, 60, or 90 minutes. To terminate drug transport, cells were washed three times with ice-cold HBSS buffer and were then scrubbed and moved to a 15-mL centrifuge tube. Cell numbers were counted by using a TC10 Automated Cell Counter (Bio-Rad, Hercules, CA), and the cells were

collected by centrifugation at $100g$, 4°C for 10 minutes. Cell pellets were stored at −80°C prior to anastrozole analysis.

Anastrozole uptake by cells was quantified by liquid chromatography–tandem mass spectrometry assay. Specifically, cell pellets were reconstituted with 25 µL of deionized/distilled water and lysed by water-bath sonication for 10 minutes and 5 freeze and thaw cycles. Anastrozole in the cell lysates was extracted by adding 50 µL of acetonitrile containing 100 ng/mL of anastrozole-d12 (deuterated; Toronto Research Chemicals, Ontario, Canada) as an internal standard. Samples were vortexed and centrifuged at $14,000g$ for 10 minutes. Five microliters of supernatant were injected into an ACQUITY UPLC H-Class System coupled with a Quattro Micro Mass Spectrometer fitted with an electrospray ionization probe (Waters Corporation, Milford, MA) for anastrozole quantification. Anastrozole and anastrozole-d12 were separated by a Waters X Select HSS T3 column (2.1 × 100 mm, XP, 2.5 µm; Waters Corporation) with isocratic elution utilizing the following profile: 55% water (0.1% formic acid), 45% acetonitrile (0.1% formic acid), and 5% methanol (0.1% formic acid) at 0.2 mL/minute flow rate. Anastrozole detection was accomplished by using multiple reaction monitoring in positive ESI MODE with a parent ion of 294.1 m/z, daughter ion of 225.3 m/z, and anastrozole-d12 was detected with a parent ion of 306.03 m/z and daughter ion of 237.36 m/z (see **Supplementary Methods S1** for details).

AUTHOR CONTRIBUTIONS

1. Buzdar, A.U. & ATAC Trialists' Group. 'Arimidex' (anastrozole) versus tamoxifen as adjuvant therapy in postmenopausal women with early breast cancer–efficacy overview. *J. Steroid Biochem. Mol. Biol.* **86**, 399–403 (2003).
2. Santen, R.J., Brodie, H., Simpson, E.R., Siiteri, P.K. & Brodie, A. History of aromatase: saga of an important biological mediator and therapeutic target. *Endocr. Rev.* **30**, 343–375 (2009).
3. Kane, S.P. Anastrozole, ClinCalc DrugStats Database, Version 17.0. ClinCalc (2017).
4. National Institute for Health and Care Excellence. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. Clinical guideline [CG164] <https://www.nice.org.uk/guidance/cg164> (2017).
5. Geisler, J. *et al.* Influence of anastrozole (Arimidex), a selective, non-steroidal aromatase inhibitor, on in vivo aromatisation and plasma oestrogen levels in postmenopausal women with breast cancer. *Br. J. Cancer* **74**, 1286–1291 (1996).
6. Kamdem, L.K. *et al.* In vitro and in vivo oxidative metabolism and glucuronidation of anastrozole. *Br. J. Clin. Pharmacol.* **70**, 854–869 (2010).
7. Hubalek, M. *et al.* Does obesity interfere with anastrozole treatment? Positive association between body mass index and anastrozole plasma levels. *Clin. Breast Cancer* **14**, 291–296 (2014).
8. Ingle, J.N. *et al.* Variation in anastrozole metabolism and pharmacodynamics in women with early breast cancer. *Cancer Res.* **70**, 3278–3286 (2010).
9. Ingle, J.N. *et al.* Estrogens and their precursors in postmenopausal women with early breast cancer receiving anastrozole. *Steroids* **99**, 32–38 (2015).
10. Baum, M. *et al.* Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. *Lancet* **359**, 2131–2139 (2002).
11. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
12. Hagglund, M.G. *et al.* Identification of SLC38A7 (SNAT7) protein as a glutamine transporter expressed in neurons. *J. Biol. Chem.* **286**, 20500–20511 (2011).
13. Verdon, Q., Boonen, M., Ribes, C., Jadot, M., Gasnier, B. & Sagne, C. SNAT7 is the primary lysosomal glutamine exporter required for extracellular protein-dependent growth of cancer cells. *Proc. Natl. Acad. Sci. USA* **114**, E3602–E3611 (2017).
14. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204 (2017).
15. Kim, E.E. & Wyckoff, H.W. Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J. Mol. Biol.* **218**, 449–464 (1991).
16. Stec, B., Holtz, K.M. & Kantrowitz, E.R. A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J. Mol. Biol.* **299**, 1303–1311 (2000).
17. Haarhaus, M., Brandenburg, V., Kalantar-Zadeh, K., Stenvinkel, P. & Magnusson, P. Alkaline phosphatase: a novel treatment target for cardiovascular disease in CKD. *Nat. Rev. Nephrol.* **13**, 429–442 (2017).
18. Millan, J.L. Alkaline phosphatases: structure, substrate specificity and functional relatedness to other members of a large superfamily of enzymes. *Purinergic Signal* **2**, 335–341 (2006).
19. Giacomini, K.M. *et al.* Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **9**, 215–236 (2010).
20. Liu, M. *et al.* TSPYL5 SNPs: association with plasma estradiol concentrations and aromatase expression. *Mol. Endocrinol.* **27**, 657–670 (2013).
21. Plourde, P.V., Dyroff, M. & Dukes, M. Arimidex: a potent and selective fourth-generation aromatase inhibitor. *Breast Cancer Res. Treat.* **30**, 103–111 (1994).
22. Van der Waerden, B.L. Order tests for the two-sample problem and their power. *Indagat. Math.* **55**, 453–458 (1952).
23. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
24. Hayeshi, R. *et al.* Comparison of drug transporter gene expression and functionality in Caco-2 cells from 10 different laboratories. *Eur. J. Pharm. Sci.* **35**, 383–396 (2008).
25. Lohmann, C. *et al.* Scintillation proximity assay for measuring uptake by the human drug transporters hOCT1, hOAT3, and hOATP1B1. *Anal. Biochem.* **366**, 117–125 (2007).