



Article

# Non-*CYP2D6* Variants Selected by a GWAS Improve the Prediction of Impaired Tamoxifen Metabolism in Patients with Breast Cancer

Ewa E. Hennig <sup>1,2,\*</sup> , Magdalena Piątkowska <sup>2</sup>, Krzysztof Goryca <sup>2</sup>, Ewelina Pośpiech <sup>3</sup>, Agnieszka Paziewska <sup>1</sup>, Jakub Karczmarski <sup>2</sup>, Anna Kluska <sup>2</sup>, Elżbieta Brewczyńska <sup>4</sup> and Jerzy Ostrowski <sup>1,2</sup>

<sup>1</sup> Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, 01-813 Warsaw, Poland

<sup>2</sup> Department of Genetics, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, 02-781 Warsaw, Poland

<sup>3</sup> Malopolska Centre of Biotechnology, Jagiellonian University, 30-387 Kraków, Poland

<sup>4</sup> Department of Breast Cancer and Reconstructive Surgery, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, 02-781 Warsaw, Poland

\* Correspondence: hennige@coi.waw.pl; Tel.: +48-22-5462593

Received: 31 May 2019; Accepted: 10 July 2019; Published: 24 July 2019



**Abstract:** A certain minimum plasma concentration of (Z)-endoxifen is presumably required for breast cancer patients to benefit from tamoxifen therapy. In this study, we searched for DNA variants that could aid in the prediction of risk for insufficient (Z)-endoxifen exposure. A metabolic ratio (MR) corresponding to the (Z)-endoxifen efficacy threshold level was adopted as a cutoff value for a genome-wide association study comprised of 287 breast cancer patients. Multivariate regression was used to preselect variables exhibiting an independent impact on the MR and develop models to predict below-threshold MR values. In total, 15 single-nucleotide polymorphisms (SNPs) were significantly associated with below-threshold MR values. The strongest association was with rs8138080 (*WBP2NL*). Two alternative models for MR prediction were developed. The predictive accuracy of Model 1, including rs7245, rs6950784, rs1320308, and the *CYP2D6* genotype, was considerably higher than that of the *CYP2D6* genotype alone (AUC 0.879 vs 0.758). Model 2, which was developed using the same three SNPs as for Model 1 plus rs8138080, appeared as an interesting alternative to the full *CYP2D6* genotype testing. In conclusion, the four novel SNPs, tested alone or in combination with the *CYP2D6* genotype, improved the prediction of impaired tamoxifen-to-endoxifen metabolism, potentially allowing for treatment optimization.

**Keywords:** Tamoxifen; breast cancer; endoxifen; genome-wide association study; *CYP2D6* genotype; *CYP2D6* enhancer; genotyping; SNP microarrays; prediction modeling; threshold level

## 1. Introduction

Tamoxifen is a selective estrogen receptor (ER) modulator that is highly effective for the treatment of ER-positive breast cancer. To elicit its therapeutic effects, tamoxifen requires internal metabolism. It is primarily biotransformed to *N*-desmethyl-tamoxifen (NDM-Tam) and 4-hydroxy-tamoxifen (4-OH-Tam), which are further converted to the secondary metabolite 4-hydroxy-*N*-desmethyl-tamoxifen (4-OH-NDM-Tam; endoxifen) [1]. Both 4-OH-Tam and endoxifen are considered to be principal active metabolites, exhibiting similar antiestrogenic potency up to 100-fold higher than that of other tamoxifen metabolites and the parent drug [2–4]: However, endoxifen is widely

recognized as the most potent metabolite in terms of its relative contribution to the overall therapeutic activity of the drug [2,5–8].

The rate-limiting step in tamoxifen metabolism directed to endoxifen production depends primarily on cytochrome P450 2D6 (CYP2D6) activity, which oxidizes tamoxifen and NDM-Tam to 4-OH-Tam and endoxifen, respectively [5,9]. This enzyme is encoded by a highly polymorphic gene comprising over 150 allelic variants, many of which are associated with absent or decreased activity of the generated enzyme [10]. Based on the *CYP2D6* genotype, individuals can be classified as an ultrarapid metabolizer (UM), a normal metabolizer (NM; wild-type (WT)), an intermediate metabolizer (IM), or a poor metabolizer (PM) in terms of CYP2D6 enzyme function [11,12]. Therefore, *CYP2D6* genotyping is considered to have great potential for predicting the efficacy of tamoxifen treatment. Unfortunately, the influence of impaired CYP2D6 enzyme activity (IM or PM) on patient clinical outcomes is inconsistent [13–22]. Some methodological issues have been suggested to be the main reasons for conflicting data in studies generating negative results [23–25], but not all of them have been finally confirmed [26].

Nevertheless, it is not disputed that patients with low-activity CYP2D6 phenotypes produce significantly less endoxifen than WT allele carriers, and systemic concentration of this metabolite decreases in proportion to enzyme deficiency [4,15,27–31]. Hence, the therapeutic failure of tamoxifen may be due to the functional impairment of the metabolism directed toward endoxifen production. In a large retrospective study, a minimal threshold level of endoxifen to achieve the desired treatment efficacy has been described: For patients with a metabolite concentration  $>5.97 \text{ ng mL}^{-1}$  (lowest quintile), a 26% lower breast cancer recurrence rate was observed relative to those with lower concentrations [28]. Similarly, threshold levels of  $5.3 \text{ ng mL}^{-1}$  [30] and  $3.36 \text{ ng mL}^{-1}$  [32] were later reported. However, a recent prospective CYPTAM (The Netherlands National Trial Register: NTR1509) study of 667 patients with early-stage breast cancer did not confirm an association between endoxifen concentration and clinical outcome, neither grouping endoxifen concentrations into quantiles nor using  $5.9 \text{ ng mL}^{-1}$  as a threshold [33]. Thus, further studies are needed to settle the existing controversy.

Although the CYP2D6 phenotype is currently considered to be the best predictor of tamoxifen metabolism directed toward endoxifen production [9], it explains approximately 40% of interpatient variability in endoxifen steady-state concentrations [4,27,34,35]. The use of concomitant medications, some of which are strong CYP2D6 inhibitors [5,29,36–38], the genetic variability of other phase I or II drug-metabolizing enzymes [4,21,39–42], weak compliance [43], or other unpredictable factors may also significantly influence endoxifen exposure. Despite many attempts to improve the predictive value of the *CYP2D6* genotype and expected phenotype, there remains insufficient evidence to justify the implementation of routine *CYP2D6* testing in clinical practice [21,31,44–49].

We previously reported that the *CYP2D6* genotype accounts for 51% of the variability in tamoxifen metabolism directed toward endoxifen production, expressed as the metabolic ratio (MR) of (Z)-endoxifen (the major isomer of endoxifen produced from tamoxifen [50]) plasma concentration divided by the sum of concentrations of tamoxifen and other measured metabolites [27]. The metabolic ratio is more useful than (Z)-endoxifen concentration alone, because it accounts for omitted drug doses or the concomitant use of CYP2D6 inhibitors. Using simple linear regression, we estimated that an MR value of 0.0146 corresponds to the predefined  $6 \text{ ng mL}^{-1}$  (Z)-endoxifen efficacy threshold level. Here, we adopted a genome-wide association study (GWAS) approach to search for new DNA variants that could assist in the prediction of impaired tamoxifen metabolism in Polish patients with breast cancer treated with the standard daily dose (20 mg). Of 15 GWAS-selected single-nucleotide polymorphisms (SNPs) associated with a below-cutoff (0.0146) MR value, four SNPs, tested alone or in combination with the *CYP2D6* genotype, significantly improved the prediction of impaired metabolism directed toward endoxifen, which could potentially inform decisions to change the drug or dosing regimen before treatment initiation.

## 2. Material and Methods

### 2.1. Ethics Statement

All patients were recruited at the Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology in Warsaw, Poland. The local Ethics Committee approved the study (project identification code 38/2011), and all subjects provided their informed consent for inclusion before they participated in the study. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

### 2.2. Clinical Cohort

Between May 2012 and June 2014, 293 patients were recruited for a study of *CYP2D6* genotype association with plasma concentrations of tamoxifen and its metabolites. Details of the inclusion criteria, study protocol, genotyping and mass spectrometry (MS) methodology, and primary results were reported previously [27]. Briefly, all patients were Polish Caucasians, unselected women with verified hormone receptor-positive breast cancer (median age at diagnosis: 55 years; range: 25–95 years) receiving the standard treatment of 20 mg of tamoxifen daily for at least 1 month to ensure the steady-state plasma concentration of tested compounds. The median time between the initiation of tamoxifen treatment and blood sample collection was 21.5 months. Six patients had plasma drug concentrations <10% of the mean tamoxifen level across all patients and were excluded from further analyses. The clinical characteristics of the remaining 287 patients are listed in Table S1.

### 2.3. Quantifying Tamoxifen and Its Metabolites in Plasma

As previously described [27], whole blood samples were collected at study enrollment, and plasma was immediately separated and stored at  $-80^{\circ}\text{C}$  until ultraperformance liquid chromatography tandem MS (UPLC-MS/MS) analysis. For all samples, plasma concentrations of tamoxifen and its 14 key metabolites were measured, including the following: NDM-Tam, (Z)-endoxifen, (E)-endoxifen, 3-OH-NDM-Tam, 4'-OH-NDM-Tam, (Z)-4-OH-Tam, 3-OH-Tam, 4'-OH-Tam, (Z)- $\alpha$ -OH-Tam, (E)- $\alpha$ -OH-Tam, Tam-N-oxide, Tam-N- $\beta$ -D-glucuronide (Tam-N-gluc), (E/Z)-4-OH-NDM-Tam- $\beta$ -D-glucuronide ((E/Z)-4-OH-NDM-Tam-gluc), and (E)-4-OH-Tam-O- $\beta$ -D-glucuronide ((E)-4-OH-Tam-O-gluc). All of the above compounds and deuterated internal standards (Tam-d5, NDM-Tam-d5, and (Z)-4-OH-Tam-d5) were from Toronto Research Chemicals (North York, ON, Canada). Details of the sample preparation method, MS/MS parameters, and linearity ranges of the metabolite standards are included in File S1.

### 2.4. CYP2D6 Genotyping

Genomic DNA samples extracted from blood were genotyped for *CYP2D6* alleles, according to the nomenclature described at <https://www.parmvar.org/gene/CYP2D6> [10], as previously reported [27]. Briefly, individual TaqMan allelic discrimination assays were performed for 11 SNPs, and 3 TaqMan Gene Copy Number Assays were conducted to assess gene deletion or duplication (Thermo Fisher, Waltham, MA, USA). Each *CYP2D6* allele was assigned to one of four phenotypic categories, according to its associated enzyme function. PM (nonfunctional) alleles included *CYP2D6*\*3, \*4, \*5, \*6, and \*7; IM (reduced-function) alleles were *CYP2D6*\*9, \*10, \*17, and \*41; NM (WT; fully functional) alleles included *CYP2D6*\*1 and \*2; and UM (increased-function) alleles were duplicates of NM variants, such as *CYP2D6*\*1XN and \*2XN. For the current study, patients were categorized into two groups of *CYP2D6* functional activity: 1) Impaired metabolizers, who were presumed to have impaired metabolism of tamoxifen toward endoxifen (with *CYP2D6* diplotypes PM/PM, IM/PM, IM/IM, NM/PM, and NM/IM), and 2) normal metabolizers, with predicted therapeutically effective tamoxifen metabolism (carrying *CYP2D6* diplotypes NM/NM and NM/UM).

## 2.5. Genome-Wide Microarray Analysis

For the current GWAS, 192 DNA samples were selected from the cohort described above based on the ratio of the sum of two tamoxifen active metabolite ((Z)-endoxifen and (Z)-4-OH-Tam) plasma concentrations to the sum of the concentrations of tamoxifen and the remaining measured compounds: 96 samples were from patients with the highest ratio, and 96 were from those with the lowest ratio. Selected DNA samples were hybridized and analyzed individually on Illumina Human Omni2.5-Exome BeadChip microarrays by a commercial organization (AROS Appl. Biotech., Aarhus N, Denmark). The datasets from GWAS are available from the GEO database (Accession number: GSE129162).

## 2.6. Verification Genotyping

According to our previously described approach for the verification of GWAS findings [51], loci were chosen that were represented by blocks of at least three SNPs associated with low (Z)-endoxifen plasma concentrations (MR < 0.0146) at  $p < 10^{-3}$ , including at least two SNPs in a block associated at  $p < 10^{-4}$ , for which the interval between all pairs of adjacent SNPs was <30 kb. On chromosome 22, associated SNPs formed three blocks with a total length of almost 436 kb, of which eight SNPs were selected (range of  $p$ -values:  $1.11 \times 10^{-11}$  to  $3.27 \times 10^{-7}$ ) for further verification and multivariate regression analysis to investigate their independent impact on the MR. For all other chromosomes, the most strongly associated SNP (with the lowest  $p$ -value) at each identified locus was selected as an index SNP for further verification using 287 DNA samples from tamoxifen-treated patients with breast cancer, analyzed using TaqMan SNP Genotyping Assays (Thermo Fisher), a SensiMix™ II Probe Kit (Bioline Ltd., London, UK), and a 7900HT Real-Time PCR system (Thermo Fisher) in 384-well format.

## 2.7. Statistical Analyses

### 2.7.1. GWAS and Individual Genotyping

Probes with missing signal reads in more than four samples were discarded. The chi-square test was used to compare allele distribution between groups. A principal component analysis (PCA) was carried out using theta values (representing the proportion of the signal for one variant to the sum of signals for all variants at a given SNP) as input. Nine samples were discarded as outliers, and the remaining samples were homogeneously spread on the planes of the first two principal components (Figure S1). Distribution assumptions were further verified by visual inspection of a quantile–quantile (Q–Q) plot of  $p$ -values (Figure S2). The calculated lambda value was 1.076, and together with the Q–Q plot, this raised no concerns regarding the homogeneity of the final population. All calculations were performed using the statistical software package R, version 3.4.1 [52].

The Hardy–Weinberg equilibrium of GWAS-selected SNPs was checked using the chi-square test implemented in R and showed no significant deviation. The Cochran–Armitage trend test was used for comparisons of allele frequencies. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated by normal approximation and implemented in the EpiTools R package (CRAN-Package epitools) [53]. The  $p$ -value significance threshold was Bonferroni-corrected for multiple comparisons (0.05/18 variants = 0.00278). Power calculations were performed using R, assuming equal groups of 96 samples and allele frequency differences of 0.075–0.2 (Table S2).

### 2.7.2. Prediction Modeling

Prediction models were constructed using binary logistic regression implemented in IBM SPSS statistics software, v.24. For prediction modeling, samples with no missing data ( $N = 262$ ) were used. MR values were binary-classified as 1 if MR < 0.0146 and 0 if MR  $\geq$  0.0146. Multivariate regression analysis using the forward stepwise method (with an entry value of 0.05) was applied for the preliminary selection of variables showing independent effects on the phenotype. Two approaches for predictor preselection were applied. In Approach 1, predictors were chosen from a set of 13 SNPs verified by individual genotyping, in addition to the *CYP2D6* genotype. In Approach 2, predictors

from the same set of 13 SNPs were considered: However, two SNPs from the *CYP2D6* gene region were also used, rather than the entire *CYP2D6* genotype.

Preselected variables were ranked according to their importance using the  $-2$  log-likelihood reduced model statistic. The impact of each variable on MR predictive accuracy was then evaluated through calculation of the area under the receiver-operating characteristic (ROC) curve (AUC) parameter, designated after the implementation of subsequent SNPs into the model. The proportion of total variation in the MR explained by the tested variants was estimated using the Nagelkerke pseudo- $R^2$  statistic. Final prediction models (Models 1 and 2) were selected (DNA variants that increased the AUC value by  $\geq 0.005$  were retained in the model) and further evaluated through the calculation of additional parameters, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) [54]. All parameters describing the predictive performance of the models were assessed using a 10-fold cross-validation procedure, as described previously [55]. For this purpose, the entire dataset was split randomly into 10 equinumerous parts indexed by  $k$  ( $k = 1, 2, \dots, 10$ ). For each  $k$ , the  $k$ th part was excluded, and the model was built using data from the other  $k-1$  parts. Parameters describing the accuracy of prediction were then calculated on the excluded  $k$ th part of the data. Final AUC, sensitivity, specificity, PPV, and NPV values were estimated as the mean of errors of the 10 models developed during the cross-validation procedure.

### 3. Results

A GWAS approach was applied to search for new DNA variants that could assist in the prediction of impaired metabolism of tamoxifen toward endoxifen. DNA samples for both GWAS and verification analyses were from tamoxifen-treated patients with breast cancer recruited for our previous study of correlations between *CYP2D6* genotype and plasma concentrations of the drug and its 14 metabolites [27]: All patient *CYP2D6* genotype data and concentration measurements were derived from that study.

Based on the *CYP2D6* allele set, each patient was assigned to one of seven genotype-predicted functional groups: NM/UM, NM/NM, NM/IM, NM/PM, IM/IM, IM/PM, or PM/PM. Mean values of the MR coefficient were above the predefined therapeutically beneficial level of 0.0146 in only NM/UM and NM/NM patients, reaching  $0.0201 \pm 0.0076$  for NM/UM patients and  $0.0185 \pm 0.0057$  for NM/NM patients (Table 1). For prediction modeling, these two groups of patients were referred to as normal metabolizers, while others, including all those carrying at least one variant allele, were classified as impaired metabolizers.

**Table 1.** *CYP2D6* genotype frequency, steady-state plasma concentration, and metabolic ratio (MR) of (Z)-endoxifen.

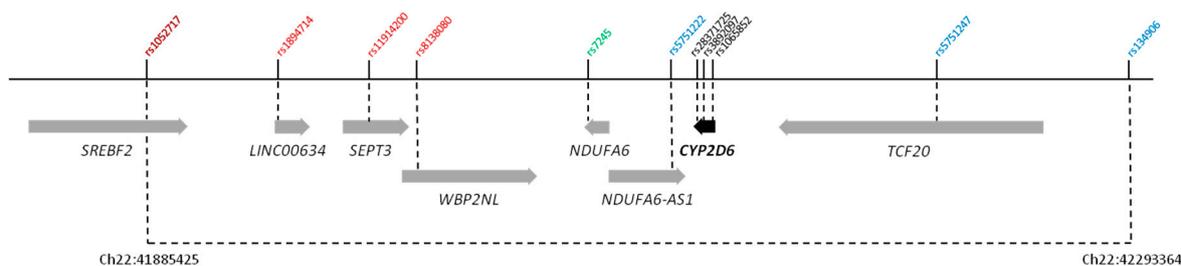
Total N = 287			
Genotype	Number of Patients (%)	(Z)-endoxifen (ng mL <sup>-1</sup> ) Mean $\pm$ SD	Metabolic Ratio <sup>a</sup> Mean $\pm$ SD
NM/UM	18 (6.3)	<b>6.96 <math>\pm</math> 3.75</b>	<b>0.0201 <math>\pm</math> 0.0076</b>
NM/NM	90 (31.4)	<b>7.26 <math>\pm</math> 3.18</b>	<b>0.0185 <math>\pm</math> 0.0057</b>
NM/IM	44 (15.3)	5.60 $\pm$ 3.26	0.0125 $\pm$ 0.0054
NM/PM	99 (34.5)	4.85 $\pm$ 2.46	0.0108 $\pm$ 0.0043
IM/IM	3 (1.0)	3.05 $\pm$ 1.91	0.0063 $\pm$ 0.0022
IM/PM	12 (4.2)	2.22 $\pm$ 0.87	0.0047 $\pm$ 0.0014
PM/PM	21 (7.3)	1.79 $\pm$ 0.66	0.0037 $\pm$ 0.0013

Mean values in bold font are higher than the thresholds of 6 ng mL<sup>-1</sup> and 0.0146 for the (Z)-endoxifen concentration and MR, respectively. NM, normal metabolizer; IM, intermediate metabolizer; PM, poor metabolizer; UM, ultrarapid metabolizer; <sup>a</sup>, the MR coefficient was estimated as the (Z)-endoxifen plasma concentration divided by the sum of the concentrations of the remaining measured compounds.

### 3.1. Association Analyses

A MR value of 0.0146, previously described as corresponding to the (Z)-endoxifen threshold level of  $6 \text{ ng mL}^{-1}$  [27], was adopted as the cutoff level in a GWAS using 192 DNA samples. According to an approach successfully applied and discussed in detail in previous studies [51,56,57], 15 SNPs were selected for further verification of their association with below-threshold MR values. Of these, seven SNPs were at independent loci outside of chromosome 22, where there were blocks of at least three SNPs associated with  $p < 10^{-3}$  at a distance of  $<30 \text{ kb}$  from one another and at least two SNPs associated with  $p < 10^{-4}$ . Two of these blocks consisted of  $>10$  SNPs. The most strongly associated SNP (with the lowest  $p$ -value) at each locus was selected as a marker SNP for verification analyses.

On chromosome 22, three blocks of SNPs associated at  $p < 10^{-3}$ , located close to one another, were identified: Block 1, 28 SNPs; block 2, 6 SNPs; and block 3, 37 SNPs. In total, these three blocks covered almost 436 kb. Given the strong associations with numerous SNPs in this region, eight variants were selected (range of  $p$ -values:  $1.11 \times 10^{-11}$  to  $3.27 \times 10^{-7}$ ) for verification and further multivariate regression analysis to investigate their independent impact on MR: Four SNPs from block 1, one from block 2, and three from block 3 (Figure 1). Additionally, three *CYP2D6* variants (rs3892097, rs28371725, and rs1065852) indicative of the most frequently occurring functionally impaired *CYP2D6* alleles among Polish patients (\*4, \*41, and \*10; Table S3) [27] were included in verification genotyping, as all three SNPs were absent from the Illumina microarray used in this study.



**Figure 1.** Long-range polymorphisms on chromosome 22, including single-nucleotide polymorphisms (SNPs) located upstream and downstream of the *CYP2D6* gene. Variants selected for verification from blocks 1, 2, and 3 of associated SNPs are indicated in red, green, and blue, respectively.

The 18 SNPs finally selected (Table 2) were evaluated in 287 DNA samples using TaqMan-based genotyping (for detailed genotyping data, see Table S4). In total, 15 SNPs (13 GWAS-selected and two *CYP2D6* variants) exhibited significantly different allele frequencies after adjustment for multiple testing ( $p < 2.78 \times 10^{-3}$ ) between the group with an MR below the 0.0146 threshold and that with an MR above the threshold. Very strong associations (range of  $p$ -values:  $1.78 \times 10^{-15}$  to  $1.36 \times 10^{-7}$ ) were observed for chromosome 22 SNPs, with the strongest for the *WBP2NL* gene variant rs8138080 ( $p = 1.78 \times 10^{-15}$ ). Among loci outside chromosome 22, the most significant association was with rs11780345 in *TNFRSF10A* ( $p = 5.57 \times 10^{-5}$ ).

The minor alleles (MAs) of nine SNPs (including two from *CYP2D6*) were associated with an increased risk of impaired drug metabolism and a low MR value, while those of the remaining six SNPs showed a protective effect (Table 2). The strongest effects on the MR were observed for the two SNPs in *CYP2D6*, rs3892097, and rs1065852 (odds ratio (OR) = 6.93, 95% confidence interval (CI) = 3.92–12.24 and OR = 6.85, 95% CI = 3.98–11.77, respectively). Other than *CYP2D6*, the strongest impacts were observed for rs5751222 (OR = 5.96, 95% CI = 3.55–10.00,  $p = 2.31 \times 10^{-12}$ ) and rs8138080 (OR = 5.55, 95% CI = 3.52–8.75,  $p = 1.78 \times 10^{-15}$ ). Among the GWAS-selected SNPs, eight were located within protein-coding gene regions: One in an exon (rs1320308 in *S100Z*), one in a 3'-untranslated region (rs7245 in *NDUFA6*), and six in introns. Two SNPs were variants of noncoding RNA (ncRNA) genes, and three were intergenic.

**Table 2.** Verification analysis of genome-wide association study (GWAS)-selected SNPs and the most frequent *CYP2D6* variants.

dbSNP ID <sup>a</sup>	Region <sup>b</sup>	MA	MAF <sup>c</sup>	OR (95% CI)	<i>p</i> -Value
rs9844493	Chr3:67210593 intergenic	A	0.482	1.59 (1.13–2.23)	$3.74 \times 10^{-3}$
rs1320308	Chr5:76875427 <i>S100Z</i> exon 6 (E23A)	A	0.447	0.51 (0.35–0.73)	$1.59 \times 10^{-3}$
rs6950784	Chr7:155902980 intergenic	G	0.519	1.93 (1.37–2.72)	$4.42 \times 10^{-4}$
rs980729	Chr7:98697127 intergenic	A	0.361	0.83 (0.59–1.1)	$1.22 \times 10^{-3}$
rs11780345	Chr8:23194453 <i>TNFRSF10A</i> intron 9	C	0.196	0.69 (0.49–0.97)	$5.57 \times 10^{-5}$
rs11786748	Chr8:27924000 <i>SCARA5</i> intron 3	G	0.403	1.95 (1.35–2.82)	$6.18 \times 10^{-4}$
rs7988513	Chr13:92232033 <i>GPC5</i> intron 7	C	0.300	0.80 (0.57–1.14)	$7.9 \times 10^{-3}$
rs1052717	Chr22:41885425 <i>SREBF2</i> intron 13	A	0.304	0.36 (0.25–0.50)	$1.36 \times 10^{-7}$
rs1894714	Chr22:41953130 <i>LINC00634</i> ncRNA	T	0.188	4.31 (2.65–7.00)	$7.48 \times 10^{-9}$
rs11914200	Chr22:41982066 <i>SEPT3</i> intron 4	A	0.239	4.29 (2.82–6.55)	$4.73 \times 10^{-12}$
rs8138080	Chr22:42000367 <i>WBP2NL</i> intron 2	A	0.261	5.55 (3.52–8.75)	$1.78 \times 10^{-15}$
rs7245	Chr22:42085845 <i>NDUFA6</i> 3' UTR	G	0.326	0.28 (0.2–0.4)	$7.52 \times 10^{-13}$
rs5751222	Chr22:42121918 <i>NDUFA6-AS1</i> ncRNA	A	0.229	5.96 (3.55–10.00)	$2.31 \times 10^{-12}$
rs5751247	Chr22:42237048 <i>TCF20</i> intron 3	C	0.290	4.82 (3.09–7.52)	$2.81 \times 10^{-13}$
rs134906	Chr22:42293364 intergenic	T	0.333	0.37 (0.26–0.52)	$2.35 \times 10^{-8}$
rs28371725	Chr22:42127803 <i>CYP2D6</i> intron 6 (SSV)	T	0.064	2.24 (1.11–4.50)	$6.81 \times 10^{-2}$
rs3892097	Chr22:42128945 <i>CYP2D6</i> intron 3 (SSV)	T	0.093	6.93 (3.92–12.24)	$1.98 \times 10^{-12}$
rs1065852	Chr22:42130692 <i>CYP2D6</i> exon 1 (P34S)	A	0.238	6.85 (3.98–11.77)	$3.5 \times 10^{-13}$

Bold font denotes significant associations ( $p < 2.78 \times 10^{-3}$ ) in verification TaqMan-based genotyping (Cochran–Armitage trend test). A metabolic ratio (MR) of 0.0146 was used as the specific cutoff value for comparison groups:  $MR < 0.0146$  ( $N = 175$ ) and  $MR \geq 0.0146$  ( $N = 112$ ). MA, minor allele; MAF, MA frequency; OR, odds ratio (calculated for MA); CI, confidence interval; SSV, splice site variant; <sup>a</sup>, SNP identifier (ID) based on the National Center for Biotechnology Information (NCBI) SNP database (<https://www.ncbi.nlm.nih.gov/snp/>); <sup>b</sup>, chromosome position (GRCh38.p7) and NCBI ID of genes close to SNPs of interest (<https://www.ncbi.nlm.nih.gov/snp/>); <sup>c</sup>, MAF based on the NCBI SNP database 1000 genomes (<https://www.ncbi.nlm.nih.gov/snp/>).

### 3.2. Predictive Performance

To evaluate the predictive performance of the selected DNA variants, two multivariate binary logistic regression models were developed, and parameters describing the accuracy of prediction were calculated. A multivariate regression analysis allowed for the preselection of variables exhibiting an independent impact on the MR. Two approaches to the selection of predictors were applied. In Approach 1, the *CYP2D6* genotype and 13 additional SNPs that were selected in the current study (Table 2) were tested to determine whether any of the new variables could improve the predictive accuracy achieved by testing solely the *CYP2D6* genotype. In Approach 2, the same set of 13 SNPs was analyzed in addition to two SNPs from *CYP2D6*, rather than the full *CYP2D6* genotype (as tested in Approach 1). The goal of Approach 2 was to establish a simplified, alternative predictive model for MR, avoiding the need for an analysis of the full *CYP2D6* genotype.

The application of a multivariate regression analysis in Approach 1 revealed five significant variables, including the *CYP2D6* genotype ( $p = 6.28 \times 10^{-13}$ ), rs7245 ( $p = 1.38 \times 10^{-3}$ ), rs6950784 ( $p = 7.94 \times 10^{-4}$ ), rs1320308 ( $p = 6.26 \times 10^{-3}$ ), and rs11786748 ( $p = 0.047$ ) (Table 3). Nagelkerke pseudo- $R^2$  analysis indicated that the *CYP2D6* genotype could explain 42.7% of total observed variation in the MR, while the additional four SNPs explained a further 14.1% of the variability. The final prediction model (Model 1) was built using the *CYP2D6* genotype, rs7245, rs6950784, and rs1320308, each of which increased the AUC value by  $\geq 0.005$  (Table 3). The AUC value for MR prediction based on the *CYP2D6* genotype alone was 0.758 (AUC values range from 0.5, indicating random prediction, to 1.0, indicating perfect prediction). A noticeable increase in predictive performance was observed when an additional three SNPs were included in the model, reaching a final AUC value of 0.879.

**Table 3.** Multivariate regression analysis considering a set of 13 SNPs and the *CYP2D6* genotype (Approach 1).

Variable/ dbSNP ID <sup>a</sup>	Gene	-2 Log Likelihood	Rank	$R^2$ <sup>b</sup>	OR (95% CI)	$p$ -Value	AUC <sup>c</sup>
<b><i>CYP2D6</i> genotype</b>	<i>CYP2D6</i>	254.137	1	0.427	16.13 (7.58–34.48)	$6.28 \times 10^{-13}$	0.758
<b>rs7245</b>	<i>NDUFA6</i>	238.013	2	0.482	0.40 (0.22–0.70)	$1.38 \times 10^{-3}$	0.842
<b>rs6950784</b>	Intergenic	223.925	3	0.527	2.24 (1.40–3.59)	$7.94 \times 10^{-4}$	0.871
<b>rs1320308</b>	<i>S100Z</i>	214.671	4	0.556	0.48 (0.29–0.81)	$6.26 \times 10^{-3}$	0.879
rs11786748	<i>SCARA5</i>	210.654	5	0.568	1.68 (1.01–2.80)	0.047	0.880

Only variables significant ( $p < 0.05$ ) in the multivariate regression analysis are shown, ranked according to their importance and designated using the -2 log-likelihood reduced model statistic. Bold font denotes the variables that changed the area under the receiver-operating characteristic curve (AUC) value by  $\geq 0.005$ , included in the final prediction model (Model 1). Odds ratios (ORs) were calculated for the minor alleles categorized in an additive manner. The *CYP2D6* genotype indicates diplotypes of impaired metabolism (PM/PM, PM/IM, IM/IM, NM/IM, and NM/PM) versus normal metabolism (NM/UM and NM/NM). CI, confidence interval. <sup>a</sup>, SNP identifier (ID) based on the NCBI SNP database (<https://www.ncbi.nlm.nih.gov/snp/>); <sup>b</sup>, Nagelkerke pseudo- $R^2$  values calculated after sequential implementation of the ranked SNPs; <sup>c</sup>, AUC value calculated after sequential implementation of the ranked SNPs.

In Approach 2, a multivariate regression analysis revealed six SNPs significantly associated with the MR, including rs8138080 ( $p = 1.56 \times 10^{-3}$ ), rs1320308 ( $p = 3.7 \times 10^{-3}$ ), rs6950784 ( $p = 2.81 \times 10^{-3}$ ), rs7245 ( $p = 0.034$ ), rs1065852 ( $p = 0.020$ ), and rs11786748 ( $p = 0.028$ ) (Table 4). Analysis of the selected SNPs using Nagelkerke pseudo- $R^2$  revealed that they could explain 48.5% of total variation in the MR, which was more than the *CYP2D6* genotype alone. The strongest association was observed for SNP rs8138080 in *WBP2NL*, which explained the largest proportion of MR variation (33.7%). Prediction Model 2 was developed using four SNPs (rs8138080, rs1320308, rs6950784, rs7245) that improved the AUC by  $\geq 0.005$  (Table 4).

**Table 4.** Multivariate regression analysis considering a set of 15 SNPs, including two *CYP2D6* variants (Approach 2).

dbSNP ID <sup>a</sup>	Gene	-2 Log Likelihood	Rank	$R^2$ <sup>b</sup>	OR (95% CI)	$p$ -Value	AUC <sup>c</sup>
<b>rs8138080</b>	<i>WBP2NL</i>	278.432	1	0.337	3.73 (1.65–8.43)	$1.56 \times 10^{-3}$	0.718
<b>rs1320308</b>	<i>S100Z</i>	264.234	2	0.390	0.48 (0.29–0.79)	$3.7 \times 10^{-3}$	0.795
<b>rs6950784</b>	Intergenic	255.872	3	0.420	1.91 (1.25–2.92)	$2.81 \times 10^{-3}$	0.819
<b>rs7245</b>	<i>NDUFA6</i>	247.300	4	0.450	0.54 (0.30–0.95)	0.034	0.830
rs1065852	<i>CYP2D6</i>	242.012	5	0.468	2.78 (1.18–6.59)	0.020	0.817
rs11786748	<i>SCARA5</i>	237.058	6	0.485	1.70 (1.06–2.74)	0.028	0.813

Only SNPs significant ( $p < 0.05$ ) in the multivariate regression analysis are shown, ranked according to their importance and designated using the -2 log-likelihood reduced model statistic. Bold font denotes SNPs that changed the area under the receiver-operating characteristic curve (AUC) value by  $\geq 0.005$ , included in the final prediction model (Model 2). Odds ratios (ORs) were calculated for the minor alleles categorized in an additive manner. CI, confidence interval. <sup>a</sup>, SNP identifier (ID) based on the NCBI SNP database (<https://www.ncbi.nlm.nih.gov/snp/>); <sup>b</sup>, Nagelkerke pseudo- $R^2$  values calculated after sequential implementation of the ranked SNPs; <sup>c</sup>, AUC value calculated after sequential implementation of the ranked SNPs.

Next, prediction parameters were calculated for both models (Table 5). The overall AUC values for Models 1 and 2 were 0.879 and 0.830, respectively. Model 1 had a sensitivity for MR prediction of 87.8%, indicating that, of 156 individuals with MR levels  $<0.0146$ , 137 would be detected. The specificity for Model 1 was 70.8%, indicating that 31 of 106 patients would be falsely positively predicted as having an MR  $<0.0146$ . For Model 2, sensitivity and specificity values for MR prediction were only slightly lower, at 80.1% and 64.2%, respectively. Overall, Model 2 is a potentially interesting alternative to the entire *CYP2D6* genotype testing.

**Table 5.** Parameters describing the accuracy of MR prediction using two binary logistic prediction models.

Parameter	Prediction Model Type	
	Model 1	Model 2
AUC	0.879	0.830
Sensitivity (%)	87.8	80.1
Specificity (%)	70.8	64.2
PPV (%)	81.6	76.7
NPV (%)	79.8	68.7

Model 1 included the *CYP2D6* genotype, rs7245, rs6950784, and rs1320308. Model 2 included rs8138080, rs1320308, rs6950784, and rs7245. The *CYP2D6* genotype indicates diplotypes of impaired metabolism (PM/PM, PM/IM, IM/IM, NM/IM, and NM/PM) versus normal metabolism (NM/UM and NM/NM). AUC, area under the receiver-operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

#### 4. Discussion

It is in the best interest of patients with breast cancer to avoid suboptimal treatment and introduce the appropriate therapy as soon as possible. Hence, there is a need to predict which patients are at risk of not achieving a therapeutically beneficial level of (Z)-endoxifen during treatment with standard-dose tamoxifen (20 mg daily). Promisingly, increasing the daily dose of tamoxifen from 20 to 30–40 mg significantly raised endoxifen concentrations in *CYP2D6* IM and PM patients to levels above or near the efficacy threshold, without appreciable effects on quality of life [58–62]. Alternatively, the use of an aromatase inhibitor, either with or without ovarian function suppression, has recently been recommended for PM and IM patients [63], particularly since the switch from tamoxifen to anastrozole did not increase the risk of recurrence in PM patients [13]. Hence, tamoxifen treatment outcomes may be improved by ensuring the appropriate hormonal therapy regimen or dose.

Although the genetic polymorphism of *CYP2D6* clearly has a major impact on (Z)-endoxifen plasma level variability, there remains considerable controversy regarding its usefulness in clinical practice. In this study, we used a GWAS approach to search for DNA variants that, alone or in addition to the *CYP2D6* genotype, could improve the prediction of failure to achieve therapeutically beneficial (Z)-endoxifen exposure. We identified 13 novel variants outside *CYP2D6* that showed significant differences in allele frequency between patients with MR values below and above the 0.0146 threshold. This previously delineated MR threshold, which correlates with a 6-ng mL<sup>-1</sup> (Z)-endoxifen efficacy threshold, was used because it is a better predictor of impaired metabolism directed to endoxifen than plasma concentrations of this metabolite [27], distinguishing between the impact of concomitant use of *CYP2D6* inhibitors and poor compliance with treatment. The MR can account for 61% of variability in (Z)-endoxifen absolute plasma levels [27].

We developed multivariate binary logistic regression models to evaluate GWAS-selected and verified variants in terms of their independent impact on the MR and ability to predict impaired tamoxifen metabolism. In addition to the *CYP2D6* genotype, four novel SNPs were found to explain a further 14.1% of MR variation, representing a total of 56.8% (Table 3). Prediction Model 1, which included three SNPs (rs7245, rs6950784, and rs1320308) plus the *CYP2D6* genotype, clearly improved the accuracy of MR prediction relative to the solely *CYP2D6* genotype-based model: The AUC value increased by 0.121 to 0.879. The sensitivity of almost 88% indicates that only 12% of patients with an MR below the beneficial level would be overlooked using Model 1 (Table 5). In addition, 29% of individuals

would be falsely positively predicted, but this seems potentially less disadvantageous. It has been shown that increasing the daily dose of tamoxifen does not tend to increase treatment-related toxicity, regardless of the CYP2D6 functional group [59]. However, it should be taken into consideration that the majority of studies reporting the safety of tamoxifen-dose escalation have concerned patients with impaired drug metabolism and low endoxifen levels [61,62].

Several attempts have been made to improve CYP2D6 genotype-predicted phenotype-based testing for prediction of the impaired metabolism of tamoxifen to endoxifen. Primarily, the CYP2D6 enzyme activity score (AS) was introduced to optimize the calibration of each patient's metabolizer phenotype [10]. This involves the assignment of an activity value to each CYP2D6 allele carried, with values of 0, 0.5, and 1 for null, reduced-function, and fully functional alleles, respectively. The enzyme AS is the sum of activity values for the patient's particular allele combination and is used to classify them into a specific phenotype group (UM, NM, IM, or PM): However, methods of phenotype grouping have differed significantly among studies, strongly influencing the prediction of endoxifen exposure [34]. Hence, there is a need for further improvement and standardization of the CYP2D6 activity scoring system and phenotype grouping, particularly to avoid the collapse of distinct IM diplotypes into a composite phenotype group and take into account the reduced activity of the \*10 allele relative to other IM alleles or the real activity of the \*2 allele, which appears to be closer to an IM than an NM allele [31,63,64].

Recently, CYP2D6 diplotypes were reported to be the best predictors of plasma endoxifen variability, compared to various diplotype-based phenotypical groupings [34]: Prediction values ranged from 39% to 58%, depending on population ethnicity. In Caucasians, endoxifen plasma concentrations above the efficacy threshold of 5.9 ng mL<sup>-1</sup> could be predicted using CYP2D6 diplotypes, with AS ≥ 1 at 94% sensitivity and 59% specificity. Another approach indicated that besides CYP2D6, the CYP3A4\*22 genotype, seasonal variation of sample collection (associated with vitamin D variation), and concomitant use of CYP2D6-inhibiting drugs may be useful in the prediction of falling below beneficial levels of plasma endoxifen concentration ( $R^2 = 0.46$ ) [35]. The sensitivity and specificity achieved by this model at a probability threshold of 0.8 were 81% and 77%, respectively. Together with our results, these observations clearly indicate that CYP2D6 genotype-based predictions can be significantly improved by including other specific genetic or environmental variables into predictive models.

Of particular note, only six SNPs (five novel variants and the CYP2D6 variant rs1065852, which defines the impaired allele \*10) exhibited significant, independent impacts on the MR, accounting for 48.5% of MR variation (Table 4). The SNP with the strongest effect was rs8138080 in WBP2NL, which explained almost 34% of MR variation. None of the tested CYP2D6 variants were finally included in prediction Model 2, which was built using the same three SNPs as applied in Model 1 and WBP2NL rs8138080 rather than the CYP2D6 genotype. Although the accuracy of MR prediction using Model 2, measured by the AUC value, was slightly lower than that of Model 1 (by 0.049), it was clearly higher than that of the CYP2D6 genotype-based model alone (by 0.072), indicating that prediction Model 2 may provide a valuable, simpler alternative to the entire CYP2D6 genotype testing.

The most important finding of this study is that four newly discovered SNPs, tested alone or in addition to CYP2D6 genotype variants, appreciably improved the prediction of impaired metabolism of tamoxifen toward endoxifen, i.e., rs7245 (NDUFA6), rs1320308 (S100Z), rs8138080 (WBP2NL), and rs6950784 (intergenic). Although SNP associations generally do not imply direct functional relationships, intriguingly, the three genes containing three of these SNPs are involved in breast cancer. High levels of NDUFA6 expression were found to predict increased tamoxifen treatment failure and tumor recurrence in high-risk ER-positive breast cancer patients at diagnosis [65]. Elevated mRNA levels of S100Z have been associated with both shorter overall survival [66] and opposite, longer, relapse-free survival and distant metastasis-free survival [67] in breast cancer patients. In the current study, the MA of both NDUFA6 (rs7245) and S100Z (rs1320308) variants was associated with a decreased risk of impaired metabolism of tamoxifen to endoxifen (OR = 0.28 and 0.51, respectively). In turn, rs8138080 in WBP2NL exhibited the strongest association with a below-threshold MR value

(OR = 5.55;  $p = 1.78 \times 10^{-15}$ ), indicating an increased risk of impaired drug metabolism: However, its effect on the MR is unlikely to be independent from those of *CYP2D6* functional variants, since it was not selected as an independent variable by multivariate regression analysis in our Approach 1.

Why patients with the same *CYP2D6* genotype can differ significantly in the efficiency of tamoxifen metabolism toward endoxifen remains largely unexplained. As we reported previously, over 30% of Polish women with fully active *CYP2D6* (NM/NM) may not achieve beneficial threshold levels of plasma (Z)-endoxifen when receiving the standard dose of tamoxifen (20 mg daily) [27]. The use of *CYP2D6* inhibitors or a low degree of compliance with therapy are known critical factors [41,47,63], and increasing evidence suggests that differential regulation of *CYP2D6* transcription may also contribute to this interindividual variability [68]. Recent reports have revealed that long-range polymorphisms, located downstream and upstream of *CYP2D6*, may alter its expression and consequently its activity [69–71]. The SNP rs8138080 in *WBP2NL* (126 kb downstream of *CYP2D6*), identified in the present study as associated with a below-threshold MR value, has previously been reported as being associated with lower *CYP2D6* mRNA levels and enzyme activity [69]. In addition, two SNPs located ~127 and ~106 kb upstream of *CYP2D6* (rs17478227 and rs5751247, respectively) are associated with significantly decreased enzyme activity [69]. Conversely, rs5758550, located 115 kb downstream of *CYP2D6*, is associated with enhanced expression of this gene [70]. In agreement with these reports, rs5751247 in the transcription factor 20 (*TCF20*) gene was strongly associated with impaired tamoxifen metabolism and a below-threshold MR value in the current study (OR = 4.82,  $p = 2.81 \times 10^{-13}$ ). However, it was not included in any of the predictive models developed by multivariate regression, suggesting that its impact on the MR may not be independent.

Since long-range polymorphisms are in strong linkage disequilibrium (LD) with two functional *CYP2D6* SNPs, rs1065852 (100C>T) and rs3892097 (1846G>A), defining the decreased-activity allele \*10 (100C>T) and the nonfunctional allele \*4 (1846G>A, 100C>T), it has been suggested that the effect of long-range SNPs on *CYP2D6* activity is most likely conferred by alleles \*4 and \*10 [69]: However, the functional relationship between the variability of *CYP2D6* mRNA expression and impaired enzyme activity variants is unclear. Recently, two-fold higher levels of in vitro *CYP2D6* promoter activity were identified as associated with the MA (G) of the regulating enhancer rs5758550, compared to the major (A) allele [71], indicating that at least some long-range variants may directly influence *CYP2D6* expression. Therefore, it cannot be entirely ruled out that the rs7245 variant in *NDUFA6*, which exhibited a protective effect for the risk of impaired tamoxifen metabolism, may act as another *CYP2D6* expression enhancer. Although it remains in strong LD with impaired *CYP2D6* metabolism alleles, it was included by multivariate regression in both predictive models, either along with or without the *CYP2D6* genotype, suggesting that it has an independent impact on MR variability: Further functional analyses are needed to verify this hypothesis. Together, it seems likely that there are additional genetic variants that have roles in the transcriptional regulation of *CYP2D6* expression, whose incorporation into predictive models could significantly improve the prediction of failure to achieve beneficial (Z)-endoxifen plasma concentrations during tamoxifen treatment.

## 5. Conclusions

Based on currently available data, the efficacy of tamoxifen treatment depends on meeting a minimum threshold plasma level of (Z)-endoxifen. Accordingly, the identification of patients unlikely to attain clinically sufficient (Z)-endoxifen exposure is of great interest in applications in individual drug dose adjustments and therapy optimization. Since *CYP2D6* activity has a major impact on (Z)-endoxifen production, many attempts have been made to predict the risk of treatment failure with standard-dose tamoxifen based on patient *CYP2D6* genotype. In the current study, a GWAS analysis revealed several novel DNA variants, both in the long-range *CYP2D6* locus and outside chromosome 22, that were significantly associated with a below-efficacy threshold MR value. A multivariate regression analysis indicated that three of these SNPs exhibited effects on MR variability independent from the *CYP2D6* genotype, and their inclusion in a predictive model, in addition to the *CYP2D6*

genotype, increased the accuracy of prediction of impaired tamoxifen metabolism. Alternatively, a model consisting of the same three SNPs and rs8138080 in *WBP2NL*, rather than the *CYP2D6* genotype, was proposed, that had a clearly higher predictive performance than the *CYP2D6* genotype-based model alone did. In summary, our results clearly indicate that *CYP2D6* genotype-based predictions can be significantly improved by including other specific genetic variables into predictive models. Many lines of evidence suggest that variants modifying enzyme activity through transcriptional regulation are of particular interest in this context.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2077-0383/8/8/1087/s1>, File S1: Detailed information on ultraperformance liquid chromatography (UPLC) separation of tamoxifen metabolites, MS/MS measurement, and the linearity ranges of metabolite standards; Figure S1: Sample locations on the planes of the first two principal components. The results of the principal component analysis (PCA) conducted on the theta values (the proportion of the signal for one variant to the sum of the signals for all variants at a given SNP) obtained during microarray measurements; Figure S2: Quantile–quantile (Q–Q) plot of GWAS-obtained *p*-values from the whole exome microarray analysis; Table S1: Characteristics of patients; Table S2: Power of testing, assuming no alleles in question in one group and the proportion indicated in the top row in the second group; Table S3: Main SNPs indicative of the most frequently occurring functionally impaired *CYP2D6* variants in the Polish population; Table S4: Detailed information on the *CYP2D6* genotype, metabolic ratios (MRs), and genotyping using the TaqMan method in the verification analysis of 287 samples.

**Author Contributions:** Conceptualization, E.E.H. and J.O.; data curation, E.E.H. and K.G.; formal analysis, K.G. and E.P.; funding acquisition, E.E.H.; investigation, E.E.H., M.P., K.G., A.P., J.K., and A.K.; methodology, E.E.H., E.P., and J.O.; resources, A.P. and E.B.; software, K.G. and E.P.; supervision, E.E.H.; validation, M.P.; writing—original draft, E.E.H. and E.P.; writing—review and editing, E.E.H.

**Funding:** This work was supported by grant 2011/01/B/NZ2/05374 from the Polish National Center of Science and grant 501-109-12-14 from the Centre of Postgraduate Medical Education.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Desta, Z.; Ward, B.A.; Soukhova, N.V.; Flockhart, D.A. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: Prominent roles for CYP3A and CYP2D6. *J. Pharmacol. Exp. Ther.* **2004**, *310*, 1062–1075. [[CrossRef](#)] [[PubMed](#)]
2. Johnson, M.D.; Zuo, H.; Lee, K.-H.; Trebley, J.P.; Rae, J.M.; Weatherman, R.V.; Desta, Z.; Flockhart, D.A.; Skaar, T.C. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res. Treat.* **2004**, *85*, 151–159. [[CrossRef](#)] [[PubMed](#)]
3. Lim, Y.C.; Li, L.; Desta, Z.; Zhao, Q.; Rae, J.M.; Flockhart, D.A.; Skaar, T.C. Endoxifen, a secondary metabolite of tamoxifen, and 4-OH-tamoxifen induce similar changes in global gene expression patterns in MCF-7 breast cancer cells. *J. Pharmacol. Exp. Ther.* **2006**, *318*, 503–512. [[CrossRef](#)] [[PubMed](#)]
4. Mürdter, T.E.; Schroth, W.; Bacchus-Gerybadze, L.; Winter, S.; Heinkele, G.; Simon, W.; Fasching, P.A.; Fehm, T.; German Tamoxifen and AI Clinicians Group; Eichelbaum, M.; et al. Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. *Clin. Pharmacol. Ther.* **2011**, *89*, 708–717.
5. Stearns, V.; Johnson, M.D.; Rae, J.M.; Morocho, A.; Novielli, A.; Bhargava, P.; Hayes, D.F.; Desta, Z.; Flockhart, D.A. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J. Natl. Cancer Inst.* **2003**, *95*, 1758–1764. [[CrossRef](#)]
6. Hawse, J.R.; Subramaniam, M.; Cicek, M.; Wu, X.; Gingery, A.; Grygo, S.B.; Sun, Z.; Pitel, K.S.; Lingle, W.L.; Goetz, M.P.; et al. Endoxifen's molecular mechanisms of action are concentration dependent and different than that of other anti-estrogens. *PLoS ONE* **2013**, *8*, e54613. [[CrossRef](#)]
7. Gong, I.Y.; Teft, W.A.; Ly, J.; Chen, Y.-H.; Alicko, B.; Kim, R.B.; Choo, E.F. Determination of clinically therapeutic endoxifen concentrations based on efficacy from human MCF7 breast cancer xenografts. *Breast Cancer Res. Treat.* **2013**, *139*, 61–69. [[CrossRef](#)]
8. Wu, X.; Hawse, J.R.; Subramaniam, M.; Goetz, M.P.; Ingle, J.N.; Spelsberg, T.C. The tamoxifen metabolite, endoxifen, is a potent antiestrogen that targets estrogen receptor alpha for degradation in breast cancer cells. *Cancer Res.* **2009**, *69*, 1722–1727. [[CrossRef](#)]

9. Khan, B.A.; Robinson, R.; Fohner, A.E.; Muzquiz, L.I.; Schilling, B.D.; Beans, J.A.; Olnes, M.J.; Trawicki, L.; Frydenlund, H.; Laukes, C.; et al. Cytochrome P450 Genetic Variation Associated with Tamoxifen Biotransformation in American Indian and Alaska Native People. *Clin. Transl. Sci.* **2018**, *11*, 312–321. [[CrossRef](#)]
10. Gaedigk, A.; Ingelman-Sundberg, M.; Miller, N.A.; Leeder, J.S.; Whirl-Carrillo, M.; Klein, T.E. PharmVar Steering Committee the Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin. Pharmacol. Ther.* **2018**, *103*, 399–401. [[CrossRef](#)]
11. Rebsamen, M.C.; Desmeules, J.; Daali, Y.; Chiappe, A.; Diemand, A.; Rey, C.; Chabert, J.; Dayer, P.; Hochstrasser, D.; Rossier, M.F. The AmpliChip CYP450 test: Cytochrome P450 2D6 genotype assessment and phenotype prediction. *Pharmacogenomics J.* **2009**, *9*, 34–41. [[CrossRef](#)] [[PubMed](#)]
12. Caudle, K.E.; Dunnenberger, H.M.; Freimuth, R.R.; Peterson, J.F.; Burlison, J.D.; Whirl-Carrillo, M.; Scott, S.A.; Rehm, H.L.; Williams, M.S.; Klein, T.E.; et al. Standardizing terms for clinical pharmacogenetics test results: Consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genet. Med.* **2017**, *19*, 215–223. [[CrossRef](#)] [[PubMed](#)]
13. Goetz, M.P.; Suman, V.J.; Hoskin, T.L.; Gnant, M.; Filipits, M.; Safgren, S.L.; Kuffel, M.; Jakesz, R.; Rudas, M.; Greil, R.; et al. CYP2D6 metabolism and patient outcome in the Austrian Breast and Colorectal Cancer Study Group trial (ABCSCG) 8. *Clin. Cancer Res.* **2013**, *19*, 500–507. [[CrossRef](#)] [[PubMed](#)]
14. Hertz, D.L.; McLeod, H.L.; Irvin, W.J. Tamoxifen and CYP2D6: A contradiction of data. *Oncologist* **2012**, *17*, 620–630. [[CrossRef](#)] [[PubMed](#)]
15. Kiyotani, K.; Mushiroda, T.; Imamura, C.K.; Hosono, N.; Tsunoda, T.; Kubo, M.; Tanigawara, Y.; Flockhart, D.A.; Desta, Z.; Skaar, T.C.; et al. Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. *J. Clin. Oncol.* **2010**, *28*, 1287–1293. [[CrossRef](#)] [[PubMed](#)]
16. Jung, J.-A.; Lim, H.-S. Association between CYP2D6 genotypes and the clinical outcomes of adjuvant tamoxifen for breast cancer: A meta-analysis. *Pharmacogenomics* **2014**, *15*, 49–60. [[CrossRef](#)] [[PubMed](#)]
17. Park, I.H.; Ro, J.; Park, S.; Lim, H.-S.; Lee, K.S.; Kang, H.S.; Jung, S.-Y.; Lee, S. Lack of any association between functionally significant CYP2D6 polymorphisms and clinical outcomes in early breast cancer patients receiving adjuvant tamoxifen treatment. *Breast Cancer Res. Treat.* **2012**, *131*, 455–461. [[CrossRef](#)]
18. Rae, J.M.; Drury, S.; Hayes, D.F.; Stearns, V.; Thibert, J.N.; Haynes, B.P.; Salter, J.; Sestak, I.; Cuzick, J.; Dowsett, M.; et al. CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J. Natl. Cancer Inst.* **2012**, *104*, 452–460. [[CrossRef](#)]
19. Regan, M.M.; Leyland-Jones, B.; Bouzyk, M.; Pagani, O.; Tang, W.; Kammler, R.; Dell'orto, P.; Biasi, M.O.; Thürlimann, B.; Lyng, M.B.; et al. CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: The breast international group 1-98 trial. *J. Natl. Cancer Inst.* **2012**, *104*, 441–451. [[CrossRef](#)]
20. Schroth, W.; Goetz, M.P.; Hamann, U.; Fasching, P.A.; Schmidt, M.; Winter, S.; Fritz, P.; Simon, W.; Suman, V.J.; Ames, M.M.; et al. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* **2009**, *302*, 1429–1436. [[CrossRef](#)]
21. De Vries Schultink, A.H.M.; Zwart, W.; Linn, S.C.; Beijnen, J.H.; Huitema, A.D.R. Effects of Pharmacogenetics on the Pharmacokinetics and Pharmacodynamics of Tamoxifen. *Clin. Pharmacokinet.* **2015**, *54*, 797–810. [[CrossRef](#)] [[PubMed](#)]
22. Hertz, D.L.; Kidwell, K.M.; Hilsenbeck, S.G.; Oesterreich, S.; Osborne, C.K.; Philips, S.; Chenault, C.; Hartmaier, R.J.; Skaar, T.C.; Sikora, M.J.; et al. CYP2D6 genotype is not associated with survival in breast cancer patients treated with tamoxifen: Results from a population-based study. *Breast Cancer Res. Treat.* **2017**, *166*, 277–287. [[CrossRef](#)] [[PubMed](#)]
23. Goetz, M.P.; Sun, J.X.; Suman, V.J.; Silva, G.O.; Perou, C.M.; Nakamura, Y.; Cox, N.J.; Stephens, P.J.; Miller, V.A.; Ross, J.S.; et al. Loss of heterozygosity at the CYP2D6 locus in breast cancer: Implications for germline pharmacogenetic studies. *J. Natl. Cancer Inst.* **2014**, *107*. [[CrossRef](#)] [[PubMed](#)]
24. Province, M.A.; Goetz, M.P.; Brauch, H.; Flockhart, D.A.; Hebert, J.M.; Whaley, R.; Suman, V.J.; Schroth, W.; Winter, S.; Zembutsu, H.; et al. CYP2D6 genotype and adjuvant tamoxifen: Meta-analysis of heterogeneous study populations. *Clin. Pharmacol. Ther.* **2014**, *95*, 216–227. [[CrossRef](#)] [[PubMed](#)]

25. Schroth, W.; Hamann, U.; Fasching, P.A.; Dauser, S.; Winter, S.; Eichelbaum, M.; Schwab, M.; Brauch, H. CYP2D6 polymorphisms as predictors of outcome in breast cancer patients treated with tamoxifen: Expanded polymorphism coverage improves risk stratification. *Clin. Cancer Res.* **2010**, *16*, 4468–4477. [[CrossRef](#)] [[PubMed](#)]
26. Ahern, T.P.; Hertz, D.L.; Damkier, P.; Ejlersen, B.; Hamilton-Dutoit, S.J.; Rae, J.M.; Regan, M.M.; Thompson, A.M.; Lash, T.L.; Cronin-Fenton, D.P. Cytochrome P-450 2D6 (CYP2D6) Genotype and Breast Cancer Recurrence in Tamoxifen-Treated Patients: Evaluating the Importance of Loss of Heterozygosity. *Am. J. Epidemiol.* **2017**, *185*, 75–85. [[CrossRef](#)]
27. Hennig, E.E.; Piatkowska, M.; Karczmarski, J.; Goryca, K.; Brewczynska, E.; Jazwiec, R.; Kluska, A.; Omiotek, R.; Paziewska, A.; Dadlez, M.; et al. Limited predictive value of achieving beneficial plasma (Z)-endoxifen threshold level by CYP2D6 genotyping in tamoxifen-treated Polish women with breast cancer. *BMC Cancer* **2015**, *15*, 570. [[CrossRef](#)]
28. Madlensky, L.; Natarajan, L.; Tchu, S.; Pu, M.; Mortimer, J.; Flatt, S.W.; Nikoloff, D.M.; Hillman, G.; Fontecha, M.R.; Lawrence, H.J.; et al. Tamoxifen metabolite concentrations, CYP2D6 genotype, and breast cancer outcomes. *Clin. Pharmacol. Ther.* **2011**, *89*, 718–725. [[CrossRef](#)]
29. Borges, S.; Desta, Z.; Li, L.; Skaar, T.C.; Ward, B.A.; Nguyen, A.; Jin, Y.; Storniolo, A.M.; Nikoloff, D.M.; Wu, L.; et al. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: Implication for optimization of breast cancer treatment. *Clin. Pharmacol. Ther.* **2006**, *80*, 61–74. [[CrossRef](#)]
30. Saladores, P.; Mürdter, T.; Eccles, D.; Chowbay, B.; Zgheib, N.K.; Winter, S.; Ganchev, B.; Eccles, B.; Gerty, S.; Tfayli, A.; et al. Tamoxifen metabolism predicts drug concentrations and outcome in premenopausal patients with early breast cancer. *Pharmacogenomics J.* **2015**, *15*, 84–94. [[CrossRef](#)]
31. Hertz, D.L.; Snavely, A.C.; McLeod, H.L.; Walko, C.M.; Ibrahim, J.G.; Anderson, S.; Weck, K.E.; Magrinat, G.; Olajide, O.; Moore, S.; et al. In vivo assessment of the metabolic activity of CYP2D6 diplotypes and alleles. *Br. J. Clin. Pharmacol.* **2015**, *80*, 1122–1130. [[CrossRef](#)] [[PubMed](#)]
32. Helland, T.; Henne, N.; Bifulco, E.; Naume, B.; Borgen, E.; Kristensen, V.N.; Kvaløy, J.T.; Lash, T.L.; Alnæs, G.I.G.; van Schaik, R.H.; et al. Serum concentrations of active tamoxifen metabolites predict long-term survival in adjuvantly treated breast cancer patients. *Breast Cancer Res. BCR* **2017**, *19*, 125. [[CrossRef](#)] [[PubMed](#)]
33. Sanchez-Spitman, A.; Dezentje, V.; Swen, J.; Moes, D.J.A.R.; Bohringer, S.; Batman, E.; van Druten, E.; Smorenburg, C.; van Bochove, A.; Zeillemaker, A.; et al. Tamoxifen pharmacogenetics and metabolism: Results from the prospective CYPTAM study. *J. Clin. Oncol.* **2019**, *37*, 636–646. [[CrossRef](#)] [[PubMed](#)]
34. Schroth, W.; Winter, S.; Mürdter, T.; Schaeffeler, E.; Eccles, D.; Eccles, B.; Chowbay, B.; Khor, C.C.; Tfayli, A.; Zgheib, N.K.; et al. Improved Prediction of Endoxifen Metabolism by CYP2D6 Genotype in Breast Cancer Patients Treated with Tamoxifen. *Front. Pharmacol.* **2017**, *8*, 582. [[CrossRef](#)] [[PubMed](#)]
35. Teft, W.A.; Gong, I.Y.; Dingle, B.; Potvin, K.; Younus, J.; Vandenberg, T.A.; Brackstone, M.; Perera, F.E.; Choi, Y.-H.; Zou, G.; et al. CYP3A4 and seasonal variation in vitamin D status in addition to CYP2D6 contribute to therapeutic endoxifen level during tamoxifen therapy. *Breast Cancer Res. Treat.* **2013**, *139*, 95–105. [[CrossRef](#)]
36. Jin, Y.; Desta, Z.; Stearns, V.; Ward, B.; Ho, H.; Lee, K.-H.; Skaar, T.; Storniolo, A.M.; Li, L.; Araba, A.; et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J. Natl. Cancer Inst.* **2005**, *97*, 30–39. [[CrossRef](#)]
37. Binkhorst, L.; Bannink, M.; de Bruijn, P.; Ruit, J.; Droogendijk, H.; van Alphen, R.J.; den Boer, T.D.; Lam, M.H.; Jager, A.; van Gelder, T.; et al. Augmentation of Endoxifen Exposure in Tamoxifen-Treated Women Following SSRI Switch. *Clin. Pharmacokinet.* **2016**, *55*, 249–255. [[CrossRef](#)]
38. Zembutsu, H.; Sasa, M.; Kiyotani, K.; Mushiroda, T.; Nakamura, Y. Should CYP2D6 inhibitors be administered in conjunction with tamoxifen? *Expert Rev. Anticancer Ther.* **2011**, *11*, 185–193. [[CrossRef](#)]
39. Fernández-Santander, A.; Gaibar, M.; Novillo, A.; Romero-Lorca, A.; Rubio, M.; Chicharro, L.M.; Tejerina, A.; Bandrés, F. Relationship between genotypes Sult1a2 and Cyp2d6 and tamoxifen metabolism in breast cancer patients. *PLoS ONE* **2013**, *8*, e70183. [[CrossRef](#)]
40. Blevins-Primeau, A.S.; Sun, D.; Chen, G.; Sharma, A.K.; Gallagher, C.J.; Amin, S.; Lazarus, P. Functional significance of UDP-glucuronosyltransferase variants in the metabolism of active tamoxifen metabolites. *Cancer Res.* **2009**, *69*, 1892–1900. [[CrossRef](#)]

41. Del Re, M.; Citi, V.; Crucitta, S.; Rofi, E.; Belcari, F.; van Schaik, R.H.; Danesi, R. Pharmacogenetics of CYP2D6 and tamoxifen therapy: Light at the end of the tunnel? *Pharmacol. Res.* **2016**, *107*, 398–406. [[CrossRef](#)] [[PubMed](#)]
42. Woo, H.I.; Lee, S.K.; Kim, J.; Kim, S.W.; Yu, J.; Bae, S.Y.; Lee, J.E.; Nam, S.J.; Lee, S.-Y. Variations in plasma concentrations of tamoxifen metabolites and the effects of genetic polymorphisms on tamoxifen metabolism in Korean patients with breast cancer. *Oncotarget* **2017**, *8*, 100296–100311. [[CrossRef](#)] [[PubMed](#)]
43. Thompson, A.M.; Johnson, A.; Quinlan, P.; Hillman, G.; Fontecha, M.; Bray, S.E.; Purdie, C.A.; Jordan, L.B.; Ferraldeschi, R.; Latif, A.; et al. Comprehensive CYP2D6 genotype and adherence affect outcome in breast cancer patients treated with tamoxifen monotherapy. *Breast Cancer Res. Treat.* **2011**, *125*, 279–287. [[CrossRef](#)] [[PubMed](#)]
44. Lum, D.W.K.; Perel, P.; Hingorani, A.D.; Holmes, M.V. CYP2D6 genotype and tamoxifen response for breast cancer: A systematic review and meta-analysis. *PLoS ONE* **2013**, *8*, e76648. [[CrossRef](#)] [[PubMed](#)]
45. Vianna-Jorge, R.; Festa-Vasconcellos, J.S.; Goulart-Citrangulo, S.M.T.; Leite, M.S. Functional polymorphisms in xenobiotic metabolizing enzymes and their impact on the therapy of breast cancer. *Front. Genet.* **2012**, *3*, 329. [[CrossRef](#)] [[PubMed](#)]
46. Zeng, Z.; Liu, Y.; Liu, Z.; You, J.; Chen, Z.; Wang, J.; Peng, Q.; Xie, L.; Li, R.; Li, S.; et al. CYP2D6 polymorphisms influence tamoxifen treatment outcomes in breast cancer patients: A meta-analysis. *Cancer Chemother. Pharmacol.* **2013**, *72*, 287–303. [[CrossRef](#)] [[PubMed](#)]
47. Binkhorst, L.; Mathijssen, R.H.J.; Jager, A.; van Gelder, T. Individualization of tamoxifen therapy: Much more than just CYP2D6 genotyping. *Cancer Treat. Rev.* **2015**, *41*, 289–299. [[CrossRef](#)]
48. Del Re, M.; Michelucci, A.; Simi, P.; Danesi, R. Pharmacogenetics of anti-estrogen treatment of breast cancer. *Cancer Treat. Rev.* **2012**, *38*, 442–450. [[CrossRef](#)]
49. Fleeman, N.; Martin Saborido, C.; Payne, K.; Boland, A.; Dickson, R.; Dundar, Y.; Fernández Santander, A.; Howell, S.; Newman, W.; Oyee, J.; et al. The clinical effectiveness and cost-effectiveness of genotyping for CYP2D6 for the management of women with breast cancer treated with tamoxifen: A systematic review. *Health Technol. Assess.* **2011**, *15*, 1–102. [[CrossRef](#)]
50. Robertson, D.W.; Katzenellenbogen, J.A.; Long, D.J.; Rorke, E.A.; Katzenellenbogen, B.S. Tamoxifen antiestrogens. A comparison of the activity, pharmacokinetics, and metabolic activation of the cis and trans isomers of tamoxifen. *J. Steroid Biochem.* **1982**, *16*, 1–13. [[CrossRef](#)]
51. Zagajewska, K.; Piątkowska, M.; Goryca, K.; Bałabas, A.; Kluska, A.; Paziewska, A.; Pośpiech, E.; Grabska-Liberek, I.; Hennig, E.E. GWAS links variants in neuronal development and actin remodeling related loci with pseudoexfoliation syndrome without glaucoma. *Exp. Eye Res.* **2018**, *168*, 138–148. [[CrossRef](#)] [[PubMed](#)]
52. R: The R Project for Statistical Computing. Available online: <http://www.r-project.org/> (accessed on 2 April 2015).
53. CRAN—Package Epitools. Available online: <https://cran.r-project.org/web/packages/epitools/index.html> (accessed on 12 August 2016).
54. Pośpiech, E.; Draus-Barini, J.; Kupiec, T.; Wojas-Pelc, A.; Branicki, W. Prediction of eye color from genetic data using Bayesian approach. *J. Forensic Sci.* **2012**, *57*, 880–886. [[CrossRef](#)] [[PubMed](#)]
55. Hastie, T.; Trevor, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series of Statistics: New York, NY, USA, 2009.
56. Ostrowski, J.; Paziewska, A.; Lazowska, I.; Ambrozkiwicz, F.; Goryca, K.; Kulecka, M.; Rawa, T.; Karczmariski, J.; Dabrowska, M.; Zeber-Lubecka, N.; et al. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Sci. Rep.* **2016**, *6*, 39831. [[CrossRef](#)] [[PubMed](#)]
57. Paziewska, A.; Habor, A.; Rogowska, A.; Zych, W.; Goryca, K.; Karczmariski, J.; Dabrowska, M.; Ambrozkiwicz, F.; Walewska-Zielecka, B.; Krawczyk, M.; et al. A novel approach to genome-wide association analysis identifies genetic associations with primary biliary cholangitis and primary sclerosing cholangitis in Polish patients. *BMC Med. Genom.* **2017**, *10*, 2. [[CrossRef](#)] [[PubMed](#)]
58. Jager, N.G.L.; Rosing, H.; Schellens, J.H.M.; Linn, S.C.; Beijnen, J.H. Tamoxifen dose and serum concentrations of tamoxifen and six of its metabolites in routine clinical outpatient care. *Breast Cancer Res. Treat.* **2014**, *143*, 477–483. [[CrossRef](#)] [[PubMed](#)]

59. Irvin, W.J.; Walko, C.M.; Weck, K.E.; Ibrahim, J.G.; Chiu, W.K.; Dees, E.C.; Moore, S.G.; Olajide, O.A.; Graham, M.L.; Canale, S.T.; et al. Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: A multicenter study. *J. Clin. Oncol.* **2011**, *29*, 3232–3239. [[CrossRef](#)] [[PubMed](#)]
60. Martínez de Dueñas, E.; Ochoa Aranda, E.; Blancas Lopez-Barajas, I.; Ferrer Magdalena, T.; Bandrés Moya, F.; Chicharro García, L.M.; Gómez Capilla, J.A.; Zafra Ceres, M.; de Haro, T.; Romero Llorens, R.; et al. Adjusting the dose of tamoxifen in patients with early breast cancer and CYP2D6 poor metabolizer phenotype. *Breast* **2014**, *23*, 400–406. [[CrossRef](#)] [[PubMed](#)]
61. Dezentjé, V.O.; Opdam, F.L.; Gelderblom, H.; Hartigh den, J.; Van der Straaten, T.; Vree, R.; Maartense, E.; Smorenburg, C.H.; Putter, H.; Dieudonné, A.S.; et al. CYP2D6 genotype- and endoxifen-guided tamoxifen dose escalation increases endoxifen serum concentrations without increasing side effects. *Breast Cancer Res. Treat.* **2015**, *153*, 583–590. [[CrossRef](#)] [[PubMed](#)]
62. Hertz, D.L.; Deal, A.; Ibrahim, J.G.; Walko, C.M.; Weck, K.E.; Anderson, S.; Magrinat, G.; Olajide, O.; Moore, S.; Raab, R.; et al. Tamoxifen Dose Escalation in Patients with Diminished CYP2D6 Activity Normalizes Endoxifen Concentrations Without Increasing Toxicity. *Oncologist* **2016**, *21*, 795–803. [[CrossRef](#)] [[PubMed](#)]
63. Goetz, M.P.; Sangkuhl, K.; Guchelaar, H.-J.; Schwab, M.; Province, M.; Whirl-Carrillo, M.; Symmans, W.F.; McLeod, H.L.; Ratain, M.J.; Zembutsu, H.; et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. *Clin. Pharmacol. Ther.* **2018**, *103*, 770–777. [[CrossRef](#)]
64. Hicks, J.K.; Swen, J.J.; Gaedigk, A. Challenges in CYP2D6 phenotype assignment from genotype data: A critical assessment and call for standardization. *Curr. Drug Metab.* **2014**, *15*, 218–232. [[CrossRef](#)] [[PubMed](#)]
65. Sotgia, F.; Fiorillo, M.; Lisanti, M.P. Mitochondrial markers predict recurrence, metastasis and tamoxifen-resistance in breast cancer patients: Early detection of treatment failure with companion diagnostics. *Oncotarget* **2017**, *8*, 68730–68745. [[CrossRef](#)] [[PubMed](#)]
66. Zhang, S.; Wang, Z.; Liu, W.; Lei, R.; Shan, J.; Li, L.; Wang, X. Distinct prognostic values of S100 mRNA expression in breast cancer. *Sci. Rep.* **2017**, *7*, 39786. [[CrossRef](#)] [[PubMed](#)]
67. Cancemi, P.; Buttacavoli, M.; Di Cara, G.; Albanese, N.N.; Bivona, S.; Pucci-Minafra, I.; Feo, S. A multiomics analysis of S100 protein family in breast cancer. *Oncotarget* **2018**, *9*, 29064–29081. [[CrossRef](#)] [[PubMed](#)]
68. Pan, X.; Ning, M.; Jeong, H. Transcriptional Regulation of CYP2D6 Expression. *Drug Metab. Dispos.* **2017**, *45*, 42–48. [[CrossRef](#)] [[PubMed](#)]
69. Yang, X.; Zhang, B.; Molony, C.; Chudin, E.; Hao, K.; Zhu, J.; Gaedigk, A.; Suver, C.; Zhong, H.; Leeder, J.S.; et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res.* **2010**, *20*, 1020–1036. [[CrossRef](#)]
70. Wang, D.; Poi, M.J.; Sun, X.; Gaedigk, A.; Leeder, J.S.; Sadee, W. Common CYP2D6 polymorphisms affecting alternative splicing and transcription: Long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum. Mol. Genet.* **2014**, *23*, 268–278. [[CrossRef](#)]
71. Wang, D.; Papp, A.C.; Sun, X. Functional characterization of CYP2D6 enhancer polymorphisms. *Hum. Mol. Genet.* **2015**, *24*, 1556–1562. [[CrossRef](#)]

