CYP2D6 Polymorphisms as Predictors of Outcome in Breast Cancer Patients Treated with Tamoxifen: Expanded Polymorphism Coverage Improves Risk Stratification

Werner Schroth¹, Ute Hamann², Peter A. Fasching³,⁴, Silke Dauser¹, Stefan Winter¹, Michel Eichelbaum¹, Matthias Schwab¹,⁵, and Hiltrud Brauch¹

Abstract

Purpose: This study aimed to validate matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS)/Taqman copy number assay (CNA) CYP2D6 genotyping by AmpliChip CYP450 Test for the prediction of tamoxifen metabolizer phenotypes in breast cancer, and to investigate the influence of CYP2D6 variant coverage on genotype-phenotype relationships and tamoxifen outcome.

Experimental Design: Hormone receptor–positive postmenopausal breast cancer patients (n = 492) treated with adjuvant tamoxifen, previously analyzed by MALDI-TOF MS/CNA, were reanalyzed by AmpliChip CYP450 Test and validated by independent methods. Cox proportional hazard ratios (HR) were calculated for recurrence of poor (PM) relative to extensive metabolizer (EM) phenotypes with increasing numbers of CYP2D6 variants. Kaplan-Meier distributions were calculated for different phenotype classifications.

Results: Concordance was 99.2% to 99.5% for CNA and 99.8% to 100% per CYP2D6 allele (*3, *4, *5, *9, *10, and *41). The prevalence of predicted phenotypes was 1.2% for ultrarapid metabolizer (UM), 37.2% for EM without variant, 43.5% for heterozygous EM, 9.7% for intermediate metabolizer (IM), and 8.3% for PM. Approximately, one third of patients were misclassified based on a *4 analysis only, but inclusion of all reduced-function alleles increased the PM-associated HR from 1.33 (P = 0.58) to 2.87 (P = 0.006). Kaplan-Meier analyses showed highest and lowest clinical benefit for UM and PM with respect to both the AmpliChip-based and a redefined phenotype assignment. The latter revealed significant allele–dose-dependent associations (P = 0.011) and largest effect size (HR_PM_EM = 2.77; 95% confidence interval, 1.31-5.89).

Conclusions: MALDI-TOF MS/CNA is suitable for accurate CYP2D6 genotyping. For tamoxifen pharmacogenetics, broad CYP2D6 allele coverage is recommended to reduce phenotype misclassification. Classification based on refined EM and reduced-function metabolizers is advisable.

Tamoxifen, which is prescribed worldwide for the treatment of estrogen receptor–positive breast cancer, is known to fail in 30% to 50% of patients (1, 2). Underlying mechanisms of tamoxifen resistance include tumor-associated and host genome–associated factors (3). Recent developments in the understanding of the pharmacogenetic relationship between cytochrome P450 2D6 (CYP2D6) polymorphisms and tamoxifen outcome in early breast cancer have shown a strong relationship between a patient’s capacity to metabolize tamoxifen and treatment outcome (4), assigning this capacity at least in part to the patient’s genetic makeup. Such a relationship has long been suspected based on clinical (5–12) and pharmacokinetic studies (13–15). However, negative and conflicting studies (16–18) point to the need for standardized and uniform study designs including comprehensive CYP2D6 genetic analyses for phenotypic assignment.

These complex pharmacogenetic relationships have been underpinned recently by a well-powered multicenter clinical study (4). The capacity to benefit from tamoxifen can be genetically evaluated with CYP2D6 being the rate-limiting enzyme for the formation of the active tamoxifen...
metabolite endoxifen (13, 19). CYP2D6 metabolizer phenotypes have been thoroughly described (20, 21). Patients with an extensive metabolizer (EM) phenotype carry two functional CYP2D6 copies. In contrast, patients with an intermediate metabolizer (IM) phenotype carry reduced function alleles (e.g., CYP2D6 *10 and *41) and may show less benefit from tamoxifen. This may also apply to heterozygous EM patients with a presumably reduced CYP2D6 enzyme function. Such personalized treatment strategies in favor of or against tamoxifen will require CYP2D6 genotype-guided decisions, which must rely on accurate and comprehensive genotyping technology. This also applies to the high-throughput CYP2D6 pharmacogenetic testing of large prospective endocrine treatment trials, tamoxifen chemoprevention trials, and detailed genotype-phenotype investigations in the retrospective setting.

Of the currently available CYP2D6 genotyping methods, the AmpliChip CYP450 Test represents a Food and Drug Administration (FDA)- and certified according to EU guidelines for in vitro diagnostics (CE-IVD)-cleared genotyping test system able to simultaneously identify 33 CYP2D6 alleles using the Affymetrix microarray platform. This system relies on the availability of high-molecular-weight genomic DNA for initial long-distance PCR and allows the comprehensive CYP2D6 genotyping of a patient sample within a single analytical step (22, 23). Although the AmpliChip CYP450 Test can be regarded as a gold standard in CYP2D6 genotyping, high-throughput analyses of large patient cohorts, e.g., in retrospective settings, usually cannot take advantage of its high-density allele coverage because fragmented DNA derived from formalin-fixed paraffin-embedded tissue is not a suitable DNA source.

This current work seeks to validate the high-throughput compatible matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) technology in combination with a Taqman copy number assay (CNA) for comprehensive CYP2D6 genotyping and phenotype assignments to be applied within the pharmacogenetic analysis of clinical trials for which only archived materials are available for genotyping. We compared the laboratory performance of the MALDI-TOF MS/Taqman CNA and the AmpliChip CYP450 Test platforms in a patient collection for which genomic DNA was available from blood. At the clinical level we carried out sensitivity analyses to assess the contribution of increasing CYP2D6 deficient-allele coverage and tested the influence of different CYP2D6 phenotype classifications on risk estimates for disease recurrence in tamoxifen-treated patients. These classifications were the AmpliChip-based algorithm and a redefined definition of those patients with partially decreased enzyme function.

Materials and Methods

Study design

The 492 German patients included in this validation study were derived from a larger retrospective investigation (n = 1,361) of CYP2D6 association with tamoxifen outcome in early-stage breast cancer (4). Clinical patient characteristics are given in Supplementary Table S1. Of the 601 blood-derived DNA samples of the original study (4) the quality of 492 samples was sufficient for re-genotyping by the AmpliChip CYP450 Test for a comparative assessment of genotype frequencies and phenotype assignments. For these patients reference genotypes at selected loci (CYP2D6 *3, *4, *5, *10, and *41) previously established by MALDI-TOF MS and Taqman CNA were available (4). The AmpliChip CYP450 Test-generated genotype/phenotype data were then subjected to clinical outcome analyses following the inclusion criteria described in Schrot et al., i.e., intended 5 years of adjuvant tamoxifen, no chemotherapy, hormone receptor positivity, and last follow-up between August and December 2008 (4). Ethical approval was obtained by written informed consent, or, in cases where consent was not available, by approval from local ethical committees.

Genotyping

Genomic DNA was previously extracted from whole blood using standard procedures and QIAamp DNA blood mini kit. DNA samples were genotyped by the AmpliChip CYP450 Test (Roche Diagnostics) according to the manufacturer’s instructions. Briefly, whole gene PCR amplifications were carried out for CYP2D6 and CYP2C19 separately, followed by pooling of the amplicons, DNase I fragmentation, and biotin labeling of fragments using terminal transferase. Samples were then hybridized to AmpliChip CYP450 microarrays, washed, and stained with streptavidin-conjugated phycoerythrin,
using the Affymetrix GeneChip fluids station. Fluorescence signals were scanned by Affymetrix Scanner 3000Dx, and genotype and predicted phenotype were assigned using AmpliChip Data Analysis software. The AmpliChip Test assigns 33 CYP2D6 alleles including allele-specific gene duplication and three CYP2C19 variants. Individuals with three or more functional alleles (e.g., *1, *2, and *35) due to gene duplication were assigned ultrarapid metabolizer (UM) status with high enzyme activity, individuals with two functional alleles predicted to have normal enzyme function were assigned EM, individuals with reduced enzyme function (e.g., *9, *10, and *41) either homozygous or in combination with a PM allele were assigned IM, and individuals lacking enzyme function by homozygous or compound heterozygous state for alleles *3, *4, *5 (gene deletion), *6, *7, and *8 were assigned PM. For the clinical outcome analyses, two approaches were employed with respect to the classification of a partially reduced function group. Accordingly, the AmpliChip algorithm assigned patients heterozygous for IM or PM alleles as EM phenotypes. The other approach as used in Schroth et al. (4) more clearly defined the EM phenotype in that genuine EM, i.e., carriers of two fully functional alleles, were secluded from heterozygous EM (IM/EM and EM/PM). The latter patient group was then combined with IM patients to constitute a group with gradually decreased CYP2D6 function. Standard definitions of CYP2D6 alleles are provided by the CYP Allele Nomenclature Committee (http://www.cypalleles.ki.se). Allele-phenotype relationships are given in Supplementary Table S2.

Genotypes and predicted phenotype assignments from the AmpliChip CYP450 Test were compared with those previously obtained by MALDI-TOF MS (Sequenom MassARRAY platform), i.e., CYP2D6 *3, *4, *9, *10, and *41, and CYP2D6 Taqman CNA (Applied Biosystems) for CYP2D6 *5 and gene duplication (ref. 4; *9 was genotyped within the context of this article). Primers were designed to prevent pseudogene (CYP2D7-, 2D8) amplification while limiting amplicon sizes to 150 to 313 nucleotides. PCR primers and primers used for homogenous MassExtend reactions are shown in Supplementary Table S3; assay-specific PCR conditions are available on request.

Samples with discrepant genotype calls across both platforms were further investigated by the following independent methods: allelic discrimination assays (Applied Biosystems), DNA sequencing for single nucleotide polymorphisms (SNP), and long-distance PCR for the CYP2D6 *5 deletion (24) and gene duplication (25).

Statistical methods

The association between time to recurrence and CYP2D6 PM allelic contribution to clinical outcome was analyzed using Cox proportional hazard modeling to calculate nonadjusted hazard ratios (HR) of PM in comparison with EM patients with the phenotypes coded as categorical variables. Power calculations were used to investigate the sensitivity of the different Cox models as per variant inclusion. Phenotype definitions based on the biologically refined prediction of EM and partially deficient phenotype, or based on the AmpliChip data analysis software were applied to calculate Kaplan-Meier estimates. In Kaplan-Meier analyses, log-rank trend tests were applied assuming a linear allele–dose-dependent association with outcome by coding UM, EM, IM (hetEM/IM), and PM as 0, 1, 2, and 3, respectively. In all outcome analyses time to recurrence, defined by time to documentation of a local, locoregional, distant recurrence, or contralateral breast cancer, was used as an end point, because this end point (but not overall survival) showed a significant effect in previous studies (4, 12). All statistical tests were two-sided, and statistical significance was defined as \( P < 0.05 \). Statistical analyses were done using SPSS V16 (SPSS Inc.) and R-2.10.0 (http://www.r-project.org).

Results

Genotyping

CYP2D6 genotypes and summarized genotype classes are presented in Table 1. Table 2 gives the predicted phenotypes derived from the AmpliChip CYP450 Test classification as well as the adapted phenotype classification used in Schroth et al. (4) and throughout this article. The latter accounts for distinction between genuine EM and heterozygous EM, whereas UM and PM are uniformly defined in both classifications. There were 183 (37.2%) EM patients and 41 (8.3%) PM patients with two nonfunctional alleles. The remainder included 214 heterozygous EM (43.5%) and 48 (9.7%) IM patients who were combined into the group of 262 hetEM/IM patients (53.2%; Table 2). A gene duplication was observed in 15 cases (3.0%), of which 6 (1.2%) individuals with duplicated normal activity gene copies were assigned as UM. The prevalence of gene deletion (*5) was 3.6%, i.e., 18 patients, of whom 4 were assigned to PM.

Validation across platforms

For the 492 patients, i.e. for a maximum of 984 chromosomes, we compared genotypes for CYP2D6 variants *3, *4, *9, *10, *41, and *5 gene deletion, and gene duplication previously obtained by MALDI-TOF MS/Taqman CNA with the more comprehensive data set obtained by the AmpliChip CYP450 Test. There was a high degree of concordant genotype/allele calls between both methods ranging between 99.2% and 100% (Table 3). With respect to SNP genotyping, CYP2D6 *3 and *9 allele calls were each 100% identical, and those for alleles *4, *10, and *41 were each concordant in 99.8%. The concordance rates of CYP2D6 copy number variation was 99.5% for gene duplication and 99.2% for the *5 deletion allele.

At the SNP level, discordant allele calls were reported for CYP2D6 *4, *10, and *41 (Table 4). Verification by allelic
discrimination and DNA sequencing revealed one *4 allele (sample 1) being misclassified as *10 by the AmpliChip CYP450 Test, one *4 allele (sample 2) being misclassified as *10 by MALDI-TOF MS, and in two cases (samples 3 and 4) the *41 had been incorrectly assigned by the AmpliChip CYP450 Test. With regard to copy number variation, there were five and seven discrepant calls involving gene duplication and deletion allele *5, respectively. These cases were retested by long-distance PCR or CNA. Results showed in one case (sample 5) a duplicated *1 allele that had been incorrectly assigned as nonduplicated by Taqman CNA. In four cases (samples 6-9) long-distance PCR or CNA revealed duplicated alleles, which had been classified as nonduplicated by the AmpliChip CYP450 Test. Among the seven discrepant calls involving gene deletion, three were from Taqman CNA and four from the AmpliChip CYP450 Test, most likely as a consequence of compromised DNA quality. Evidence for this comes from the shifted UV absorbance ratios and poor PCR efficiency for large amplicon sizes in almost every sample with discrepant copy number data (data not shown).

Prediction of phenotypes and clinical association based on allele coverage

We compared the probabilities of correctly detecting a PM as a function of the number of analyzed PM alleles in the study cohort of 492 patients (Fig. 1). Analysis confined to the CYP2D6 *4 allele resulted in the detection of 65.8% of all PMs. This analysis resulted in 5.5% PMs and 61.8% homozygous EMs. The combined analysis of CYP2D6 *3, *4, and *5 increased the PM coverage to 90.2% and resulted in a total of 7.5% PMs and 55.9% EMs. Finally, the full PM coverage of 8.3%, as captured by the AmpliChip Test, included the *6 and *7 alleles. The inclusion of all available PM and IM allele assignments (Fig. 1, right) markedly reduced the numbers of homozygous EM, and as a consequence resulted in an increase of the numbers of hetEM/IM patients. Altogether, the gradual inclusion of variant alleles from a single *4 allele to the full coverage of null
alleles raised PM and hetEM/IM patient numbers by 34% and 40%, respectively, and lowered the EM group by 40%.

The AmpliChip CYP450 Test data revealed a strong positive correlation between the number of analyzed PM alleles and the associated risk of developing a breast cancer recurrence. Unadjusted HRs of PM relative to EM alleles and the associated risk of developing a breast cancer recurrence. HRs of PM relative to EM were 2.95 (95% CI, 1.10-4.27). In contrast, the AmpliChip CYP450 algorithm, which incorporates heterozygous IM and homozygous PM allele carriers combined with IM, did not reveal a significant allele–dose-dependent relationship with recurrence rates (P trend = 0.098; Fig. 2A). However, PM significantly differed from EM by HR of 2.17 (95% CI, 1.10-4.27).

Discussion

The AmpliChip CYP450 Test, due to its comprehensive coverage of all globally relevant CYP2D6 alleles, provides accurate phenotype prediction and is suitable for applications within the context of tamoxifen outcome prediction. At the technical level this study aimed to test whether alternative CYP2D6 genotyping methods perform comparably with the FDA-cleared, CE-IVD-approved AmpliChip CYP450 Test, i.e., the gold standard. A methodologic comparison with data generated by MALDI-TOF MS/Taqman CNA used in Schroth et al. (4), i.e., *3, *4, *5, *10, and *41, are shown in Fig. 2C. Differences in tamoxifen outcomes by both a linear trend test and a pairwise comparison between EM and PM were in line with the predicted functional effects. Accordingly, PM patients with null or very low CYP2D6 function had poor outcome [HR, 2.77; 95% confidence interval (95% CI), 1.31-5.89; Fig. 2B] and UM patients with increased CYP2D6 function had the highest benefit as shown by the Kaplan-Meier distributions. EM and hetEM/IM carriers were between UM and PM curves, with EM showing a benefit over hetEM/IM. Of note, limited allele coverage (Fig. 2C) was inferior to the full allele coverage (Fig. 2B) as suggested by the log rank test (P trend = 0.056 versus P trend = 0.011) as well as the PM associated HRs (2.12; 95% CI, 0.96-4.69 versus 2.77; 95% CI, 1.31-5.89). In contrast, the AmpliChip CYP450 algorithm, which incorporates heterozygous PM and IM allele carriers into EM (80.7%), did not reveal a significant allele–dose-dependent relationship with recurrence rates (P trend = 0.098; Fig. 2A). However, PM significantly differed from EM by HR of 2.17 (95% CI, 1.10-4.27).

Clinical association of CYP2D6 based on phenotype classification

In a next step we used Kaplan-Meier distributions and corresponding log-rank tests as a surrogate for a linear allele–dose-dependent clinical association between CYP2D6 and tamoxifen outcome as well as HRs for PM versus EM, which is the most relevant clinical comparison. Based on the AmpliChip CYP450 Test data we did outcome analyses according to different phenotype definitions. Kaplan-Meier curves were derived from phenotypes, i.e., UM, EM, IM, and PM as defined by the AmpliChip CYP450 Test algorithm (Fig. 2A), and compared with Kaplan-Meier curves derived from the redefined phenotypic classification as previously applied by Schroth et al. (4). In the latter, EM patients were defined as carriers of two functional alleles and a hetEM/IM group combining all patients heterozygous for a variant allele or IM genotype, whereas UM and PM definitions were identical (Fig. 2B and C). Accordingly, full allele coverage by AmpliChip CYP450 Test (Fig. 2B) resulted in the following phenotype frequencies: UM, 1.2%; EM, 37.2%; hetEM/IM, 53.2%; and PM, 8.3% (Table 2). For comparison, Kaplan-Meier distributions based on the more limited allele set generated by MALDI-TOF MS/Taqman CNA used in Schroth et al. (4), i.e., *3, *4, *5, *10, and *41, are shown in Fig. 2C. Differences in tamoxifen outcomes by both a linear trend test and a pairwise comparison between EM and PM were in line with the predicted functional effects. Accordingly, PM patients with null or very low CYP2D6 function had poor outcome [HR, 2.77; 95% confidence interval (95% CI), 1.31-5.89; Fig. 2B] and UM patients with increased CYP2D6 function had the highest benefit as shown by the Kaplan-Meier distributions. EM and hetEM/IM carriers were between UM and PM curves, with EM showing a benefit over hetEM/IM. Of note, limited allele coverage (Fig. 2C) was inferior to the full allele coverage (Fig. 2B) as suggested by the log rank test (P trend = 0.056 versus P trend = 0.011) as well as the PM associated HRs (2.12; 95% CI, 0.96-4.69 versus 2.77; 95% CI, 1.31-5.89). In contrast, the AmpliChip CYP450 algorithm, which incorporates heterozygous PM and IM allele carriers into EM (80.7%), did not reveal a significant allele–dose-dependent relationship with recurrence rates (P trend = 0.098; Fig. 2A). However, PM significantly differed from EM by HR of 2.17 (95% CI, 1.10-4.27).

Table 3. Comparison of AmpliChip and MALDI-TOF MS/CNA-based genotyping results: concordance rates

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Chromosomes tested</th>
<th>Identical (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2D6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*3</td>
<td>956</td>
<td>956 (100)</td>
</tr>
<tr>
<td>*4</td>
<td>984</td>
<td>982 (99.8)</td>
</tr>
<tr>
<td>*9</td>
<td>956</td>
<td>956 (100)</td>
</tr>
<tr>
<td>*10</td>
<td>930</td>
<td>928 (99.8)</td>
</tr>
<tr>
<td>*41</td>
<td>970</td>
<td>968 (99.8)</td>
</tr>
<tr>
<td>CYP2D6- duplication gene</td>
<td>958</td>
<td>953 (99.5)</td>
</tr>
<tr>
<td>CYP2D6-deletion</td>
<td>958</td>
<td>951 (99.2)</td>
</tr>
</tbody>
</table>

Table 2. CYP2D6 phenotypes classified by AmpliChip CYP450 Test and by a redefined algorithm

<table>
<thead>
<tr>
<th>Phenotype - AmpliChip</th>
<th>Phenotype - redefined algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>UM: 6 (1.2%)</td>
<td>UM: 6 (1.2%)</td>
</tr>
<tr>
<td>EM*: 397 (80.7%)</td>
<td>EM*: 183 (37.2%)</td>
</tr>
<tr>
<td>IM: 48 (9.7%)</td>
<td>hetEM/IM*: 262 (53.2%)</td>
</tr>
<tr>
<td>PM: 41 (8.3%)</td>
<td>PM: 41 (8.3%)</td>
</tr>
</tbody>
</table>

*EM assigned by AmpliChip algorithm including heterozygous IM and PM allele carriers.
†EM defined by presence of two functional alleles.
‡hetEM/IM: a group comprising partially reduced function phenotypes defined by heterozygous IM and PM allele carriers combined with IM.
included more than 1,300 patients (4). Although genotyping methods such as MALDI-TOF MS or fluorescence-based allelic discrimination are more robust with respect to lesser-quality genomic DNA, e.g., from formalin-fixed paraffin-embedded tissue (26), individual assays for the various types of genetic variants must be applied at the expense of time or lack of validation. Most of the previous high-throughput CYP2D6 genotyping strategies for the prediction of tamoxifen outcome have therefore been restricted to the more prevalent variants in individuals of European (*3, *4, *5, *6, *10, *41, gene duplication) and Asian (mainly *10) descent. To this end, there is a need to validate the CYP2D6 genotyping accuracy of MALDI-TOF MS and Taqman copy number analysis through comparison with corresponding genotype data obtained by the AmpliChip CYP450 Test.

The concordance rates for individual polymorphisms between MALDI-TOF MS/Taqman CNA and the AmpliChip CYP450 Test ranged between 99.2% and 100%, which is similar to other method comparisons (27). Of 492 samples, 16 showed discrepant results, of which 75% (12 samples) were misclassified for copy number variations. Despite the fact that samples for this validation had been selected on the basis of the genomic DNA extracted from blood, we conclude that the failure rate in copy number variation calls is likely attributed to the compromised quality of individual genomic DNA samples as indicated by low UV absorbance ratios and poor long-range PCR efficiency. This is possible because the DNA samples under debate have been isolated since up to 10 years ago, and some of them may have been subject to DNA degradation over time. Although this would be of concern in retrospective analyses of breast cancer patient cohorts with long-term clinical follow-up, this is not a real concern within the context of predictive testing for a personalized treatment decision. Standard clinical diagnostic centers routinely isolate high-molecular-weight DNA from the patient’s blood drawn at the time of first intervention, thus avoiding long-term storage and reducing the likelihood of DNA degradation.

The remaining 25% (4 cases) of genotype discordance at CYP2D6 SNPs cannot be attributed to DNA quality but rather must be considered true discrepancies. In one case, the MALDI-TOF MS method missed a *4 call. Three misclassifications were with the AmpliChip CYP450 Test calling a *10 instead of *4 and falsely assigning *41 for two chromosomes. Altogether, the failure rate per variant did not exceed 2 in 1,000 chromosomes for both methods. Possible explanations include sample mix-up, contamination, or SNPs under PCR primers. Following validation by the AmpliChip CYP450 Test, we consider the MALDI-TOF MS/Taqman CNA platform suitable for accurate CYP2D6 genotyping. A slightly lower failure rate with respect to gene duplication and *5 deletion polymorphism makes

### Table 4. Discordant genotyping results between AmpliChip and MALDI-TOF MS/CNA

<table>
<thead>
<tr>
<th>Sample (DNA quality)</th>
<th>AmpliChip</th>
<th>MALDI-TOF*</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (moderate)</td>
<td>*35/*10</td>
<td>EM/*4</td>
<td>*1/*4 (S)</td>
</tr>
<tr>
<td>2 (moderate)</td>
<td>*3/*4</td>
<td>EM/*41</td>
<td>*1/*41 (AD), *1/*41 (S)</td>
</tr>
<tr>
<td>3 (good)</td>
<td>*41/*41</td>
<td>EM/*41</td>
<td>*1/*41 (AD), *1/*41 (S)</td>
</tr>
<tr>
<td>4 (moderate)</td>
<td>*1/*41xN</td>
<td>EM/EM</td>
<td>*1/*1 (AD), *1/*1 (S), (LD-Dup)</td>
</tr>
<tr>
<td>5 (poor)</td>
<td>*1xN/*4</td>
<td>2C</td>
<td>&gt;2C (LD-DUP)</td>
</tr>
<tr>
<td>6 (moderate)</td>
<td>*1/*1</td>
<td>&gt;2C</td>
<td>&gt;2C (LD-DUP)</td>
</tr>
<tr>
<td>7 (moderate)</td>
<td>*1/*2</td>
<td>&gt;2C</td>
<td>&gt;2C (LD-DUP)</td>
</tr>
<tr>
<td>8 (poor)</td>
<td>*2/*2</td>
<td>&gt;2C</td>
<td>&gt;2C (CNA)</td>
</tr>
<tr>
<td>9 (poor)</td>
<td>*2/*2</td>
<td>&gt;2C</td>
<td>&gt;2C (LD-DUP)</td>
</tr>
<tr>
<td>10 (poor)</td>
<td>*2/*2</td>
<td>1C</td>
<td>1C (CNA, LD-Del)</td>
</tr>
<tr>
<td>11 (poor)</td>
<td>*2/*2</td>
<td>1C</td>
<td>1C (CNA)</td>
</tr>
<tr>
<td>12 (poor)</td>
<td>*10/*10</td>
<td>1C</td>
<td>2C (CNA)</td>
</tr>
<tr>
<td>13 (poor)</td>
<td>*1/*1</td>
<td>1C</td>
<td>1C (CNA)</td>
</tr>
<tr>
<td>14 (moderate)</td>
<td>*4/*4</td>
<td>1C</td>
<td>2C (LD-Del)</td>
</tr>
<tr>
<td>15 (poor)</td>
<td>*1/*1</td>
<td>1C</td>
<td>1C (LD-Del)</td>
</tr>
<tr>
<td>16 (moderate)</td>
<td>*1/*5</td>
<td>2C</td>
<td>1C (LD-Del)</td>
</tr>
</tbody>
</table>

Abbreviations: AD, allelic discrimination, C, gene copy; LD-Del, long distance PCR for gene deletion; LD-DUP, long distance PCR for gene duplication; S, sequencing analysis.

*MALDI-TOF MS genotyping: EM allelic state was assigned by the absence of a variant allele and corresponds to *1, *2, or *35.
†1C, 2C, >2C: one, two, or more than two CYP2D6 copies measured by Taqman-based CNA.
the Taqman CNA particularly attractive for retrospective investigations of patient cohorts whose DNA samples have undergone long-term storage.

The second goal of this study was to investigate whether prediction of tamoxifen outcome in early breast cancer can be improved either by comprehensive CYP2D6 allele coverage, as obtained by the AmpliChip CYP450 data set, and/or by the application of different algorithms of CYP2D6 phenotypic assignments, i.e., AmpliChip CYP450 Test versus an adapted phenotype definition (4). Previous clinical

---

**Fig. 1.** CYP2D6 phenotype predictability depending on the CYP2D6 variant inclusion. Bars, proportion of PM (black), hetEM/IM (grey), and EM (white) patients within the tamoxifen patient cohort as a function of increasing number of PM alleles *3, *4, *5, *6, and *7, or inclusion of IM alleles (right). Diamonds, probability of a PM detection based on allele inclusion (plotted on Log scale, right axis). Boxed numbers, nonadjusted HR and 95% CI for recurrence of PM versus EM patients together with corresponding P values from Cox regression, depending on selected CYP2D6 alleles. The associated power of these Cox models is given as a measure of sensitivity and has been calculated for a significance level of 5% and with HR, number of events, and proportion of phenotype groups depending on variant inclusion. EM was defined by patients without variant allele. For statistical reasons the 6 UM patients were combined with the EM group.

**Fig. 2.** Kaplan-Meier survival distributions for time to recurrence according to different CYP2D6 metabolizer phenotype definitions and variant inclusion. A, AmpliChip full allele coverage: patients' metabolizer phenotypes were predicted by definitions used in AmpliChip software for UM (ultrarapid metabolizer), EM (extensive metabolizer), IM (intermediate metabolizer), and PM (poor metabolizer). B, AmpliChip full allele coverage: patients' metabolizer phenotypes were classified using the predefined algorithm: UM, EM without any variant allele, hetEM/IM all patients heterozygous for a variant allele together with IM patients, and PM. C, MALDI TOF-MS/Taqman data for variants *3, *4, *5, *10, and *41: patients' metabolizer phenotypes were classified using the redefined algorithm. P values refer to a log-rank trend test assuming linear association between the number of functional alleles per phenotype class and outcome. HR and 95% CI for PM versus EM are given as the clinical most relevant pairwise comparison.
studies focused on major CYP2D6 variants prevalent in Europe and Asia, including *3, *4, *5, *10, and *41 (12). Although a recent well-powered investigation substantiated the pharmacogenetic relationship between CYP2D6 deficiency and tamoxifen outcome (4), the controversy in favor of or against CYP2D6 testing is still ongoing, mainly due to a few null or controversial results (16–18). Given that the globally available patient cohorts suitable to tackle this question are limited, the only other option to further address this issue is through comprehensive CYP2D6 allele coverage in available cohorts together with phenotype definitions based on a biological rationale.

With respect to a more comprehensive CYP2D6 allele coverage, the extended genotyping with *6 and *7 alleles resulted in an additional identification of 10% of PM patients as compared with PMs assigned through *3, *4, and *5 coverage in the original study (4). It follows that within a European population testing of CYP2D6 *6 and *7 should be considered as a critical part of standard CYP2D6 genotyping. To further support this view, we did a series of association subanalyses based on stepwise addition of PM and IM variants. The limitation to *4 only, an analysis reminiscent of some previously published clinical studies (16–18, 28), resulted in as many as one third of all PM patients being misclassified. We suggest that due to the stochastics of event numbers allocated to either phenotype, this highly restricted analysis resulted in a low nonsignificant risk of relapse (HR, 1.33) for PM relative to EM patients. In contrast, the stepwise inclusion of PM alleles led to an increase of associated risk estimates up to a statistically significant HR of 2.87 when all available impaired/null-function alleles were included. Moreover, the inclusion of all PM and IM variants caused a tremendous shift of phenotype class composition, in that PM and hetEM/IM patient numbers increased by 34% and 40%, respectively, and EM numbers dropped by 40%. This implies that more than one third of either phenotypic group would have been incorrectly classified based on a *4 analysis only. Of note, the increase in power for the detection of PM-associated recurrence risk with an increase of deficiency alleles further substantiates the requirement of comprehensive CYP2D6 allele coverage for clinical risk assessment.

Our data clearly indicate the importance of broad CYP2D6 variant coverage for accurate phenotype prediction and a maximum stratification between metabolizer phenotypes. Thus, it seems that any conclusions on CYP2D6-related tamoxifen outcome association should be regarded with caution unless a comprehensive allele variant coverage has been employed. A less comprehensive genotyping approach as given by retrospective studies where formalin-fixed paraffin-embedded tissue is the source material for genotyping must account for limited allelic information through increase of power by expanding the sample size.

Because the definition of metabolizer phenotype may also affect association results, we pursued two outcome analyses based on different phenotype assignments. The AmpliChip CYP450 Test employs a phenotype definition according to which EM patients carry at least one functional CYP2D6 allele. This definition includes heterozygotes with a putative decreased function, however, our redefined EM phenotype classification includes EM homozygotes only (4). CYP2D6 phenotyping is likely different with respect to various substrates (29). As shown by others, individuals with an EM phenotype based on AmpliChip CYP450 algorithm are characterized by genotypic heterogeneity associated with variation in CYP2D6 activity towards amitriptyline pointing to a functional gene dose effect (30). We did not adapt this concept of semiquantitative gene dose (30, 31) for two reasons. First, the gene dose score would define at least six score groups, which, given the moderate sample size in this study, would reduce power for each group, resulting in unreliable survival estimates in outcome analyses. Second, endoxifen plasma levels were not available for these patients to apply direct phenotyping. Rather, we accounted for the "EM heterogeneity" in such a way that EM patients were defined to comprise patients with two fully functional CYP2D6 alleles and all patients with a heterozygous variant genotype were grouped together with IM taking into account at least some degree of functional impairment. Instead of assigning the vast majority of patients to EM (80.7%, AmpliChip CYP450) the redefined genuine EM group merely comprises 37.2% while simultaneously increasing the group of patients with partially reduced function (53.2%, hetEM/IM). Assuming a linear allele–dose-dependent association, significant differences in tamoxifen outcome were observed among the redefined CYP2D6 phenotype groups, but not between those derived from the AmpliChip definition. In line with biological and pharmacokinetic hypotheses (15, 21, 32–34), PM patients with null or very low CYP2D6 function had the worst outcome, and UM patients with increased CYP2D6 function had the best outcome in terms of the reduction of breast cancer recurrence. Patients with no variant CYP2D6 alleles showed slightly better recurrence profiles than patients with at least one impaired CYP2D6 allele, depending on the extent of allele coverage which affected the level of statistical significance.

The application of a modified phenotype definition increased the size of the patient group with partial impairment to >50% as compared with <10% based on the AmpliChip CYP2D6 phenotype assignment. This shift in phenotype distribution increased the power of a stratified analysis and points to a decreasing benefit of tamoxifen from UM, EM, hetEM/IM to PM, which would be expected based on the underlying pharmacokinetic functional relationships. Future investigations towards the confirmation of these phenotype assignments will require the measurement of the patient's steady-state tamoxifen metabolite levels with respect to the suggested phenotypic grouping.

Of note, comprehensive phenotype assignments must also take into account nongenetic phenocopying effects of CYP2D6 inhibiting comedication (14, 15), some of which
render EM or hetEM/IM a further impaired metabolizer phenotype. Based on unpublished data from an ongoing prospective study the proportion of patients receiving co-treatment with strong CYP2D6 inhibitors (e.g., paroxetine) is 1% to 2% in this German cohort. From this it follows that we may have missed assigning nine PMs, thus underestimating the true effect size of a CYP2D6 PM-associated recurrence risk by approximately 16%.

In summary, this study supports the CYP2D6 tamoxifen pharmacogenetic relationship in early breast cancer based on comprehensive variant allele coverage and evidence-based phenotype assignments. Validation of the MALDI-TOF MS/Taqman CNA method for CYP2D6 genotyping has been achieved and therefore can be applied in those experimental settings in which the AmpliChip CYP450 Test for technical or financial reasons cannot be applied. Finally, this work has important implications for individual patient testing within the framework of predicting treatment response. It has become increasingly clear that such a diagnostic procedure requires complete information on allele status for the accurate identification of the individual patient’s CYP2D6 genotype and prediction of the patient’s true phenotype.

The latter will be critical in predicting clinical outcome and may lead to informed decisions for endocrine treatment with major implications on the patient’s recurrence-free survival and quality of life.

Disclosure of Potential Conflicts of Interest

H. Brauch and M. Schwab: commercial research grant, Roche Molecular Systems.

Acknowledgments

We thank D.M. Nikoloﬀ, G. Hillman, J. Kleiber, and H.J. Lawrence (Roche Molecular Systems, Inc., Pleasanton, CA) for technical support and discussion.

Grant Support

Bosch Foundation Stuttgart, Germany. P.A. Fasching is in part funded by Dr. Mildred Scheel Sﬁftung, Deutsche Krebshilfe, M. Schwab was supported by the BMBF grant 03IS0261C and H. Brauch by BMBF grant 01ZP0502.

We acknowledge support by Roche Molecular Systems for providing chips and reagents for the AmpliChip testing.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 02/24/2010; revised 05/05/2010; accepted 05/24/2010; published OnlineFirst 06/01/2010.

References

Polymorphism Coverage Improves Risk Stratification in Breast Cancer Patients Treated with Tamoxifen: Expanded Polymorphism Coverage Improves Risk Stratification

Werner Schroth, Ute Hamann, Peter A. Fasching, et al.

Clin Cancer Res 2010;16:4468-4477. Published OnlineFirst June 1, 2010.