Tamoxifen, the pioneering antiestrogenic medicine targeted to the tumor estrogen receptor (ER), is used successfully for long-term adjuvant therapy in breast cancer.[1,2] Extensive analyses of clinical trials demonstrate a major increase in patient survivorship in ER-positive patients. In this age of personalized medicine, any opportunity to improve response rates with tamoxifen should be rigorously investigated. Tamoxifen is considered a prodrug, given that hepatic cytochrome P450 2D6 (CYP2D6) metabolizes tamoxifen to metabolites (4-hydroxy tamoxifen and 4-hydroxy-*N*-desmethyl tamoxifen (endoxifen)) that exhibit significantly greater potency in terms of ER-binding affinity[3] and suppression of estradiol-stimulated cell proliferation.[4] CYP2D6-mediated metabolism is the rate-limiting enzymatic step for the formation of endoxifen, the most abundant active metabolite.

There has been great inconsistency among studies that have reported the association of known genetic and drug factors influencing CYP2D6 enzyme activity with tamoxifen efficacy. Therefore, the International Tamoxifen Pharmacogenomics Consortium (ITPC) was conceived, and researchers were invited to submit their data—both published and unpublished data sets regarding *CYP2D6* genetic variants and clinical outcomes in women treated with tamoxifen in the adjuvant breast cancer setting—to allow a meta-analysis of the potential associations between *CYP2D6* and clinical outcomes.

RESULTS

The ITPC comprises 12 research projects from nine countries and three continents that contributed clinical and genetic data for a total of 4,973 breast cancer patients treated with tamoxifen. In **Table 1**, we show the sample size by site and criteria. Further details for each site are shown in **S3c and S5 online**. We reported preliminary analyses of these collected cohorts before complete curation by pooling the data from each site.[5] For our meta-analyses, three detailed criteria, which ranged from the most restrictive (criterion 1) to the most inclusive (criterion 3), were defined before final curation (see **S4** online). In brief, criterion 1, derived from the NCCTG 89-30-52 clinical trial, consisted of postmenopausal women with surgically resected nonmetastatic invasive ER-positive breast cancers who received adjuvant tamoxifen monotherapy at a dose of 20 mg/day for an intended duration of 5 years, and were followed at least annually for recurrence. In addition, analysis of at least *CYP2D6*4* was required (detailed in **S4a** online). Criterion 2 included criterion 1 but allowed both pre- and postmenopausal patients who had received any duration of tamoxifen; moreover, annual follow-up was not required. Criterion 3 included all samples not excluded by any exclusion test for missing data or data inconsistencies (least restrictive). Patient characteristics according to each criterion are provided in **Table 2**.

The meta-analysis results combining the hazard ratio (HR) estimates (and the corresponding standard errors (SEs)) from each site are shown for all three criteria groups and both clinical outcomes in **Table 3**. For each of the six clinical outcome/criteria groups, we give the combined meta-analysis estimate across all 12 sites, its SE, and the results of two statistical tests: a test of the significance that the meta-HR differs from 1 and a test of "homogeneity of the estimates" across sites (a significant value for the latter test indicates that there is more variability than the derSimonian and Laird random-effects model can reasonably accommodate, suggesting that the meta-estimate and its associated *P* value are suspect). As can be seen for invasive disease–free survival (IDFS), the meta-analyses for criteria 2 and 3 are nearly significantly heterogeneous, whereas there was no indication of heterogeneity for criterion 1 ($P = 0.899$). For patients meeting criterion 1, the meta-HR for IDFS was 1.25 (95% confidence interval = 1.06, 1.47), and for breast cancer–free interval, it was 1.27 (95% confidence interval = 1.01, 1.61). These are both statistically significant, at $P = 0.009$ and $P = 0.04$, respectively. However, for the criterion 2 ($P = 0.25$) and criterion 3 ($P = 0.38$) subsets, the *CYP2D6* HR was not significant for either outcome.

In **Figure 1**, we show the individual HRs for each site for subjects meeting criterion 1, assuming an additive genetic model for *CYP2D6* (coded 0 = extensive metabolizer (EM), 1 = intermediate metabolizer (IM), and 2 = poor metabolizer (PM)) as estimated from a Cox proportional-hazards model using additional risk covariates to predict clinical outcome. Corresponding figures for criteria 2 and 3 are provided in **S6** online. (Note that the list of covariates used in the Cox models included age at primary diagnosis, menopause status at diagnosis, metastatic disease at primary diagnosis, maximum tumor dimension, number of positive nodes, grade, smoking status, ER and progesterone receptor status, intended tamoxifen dose and duration, systemic therapy before surgery, chemotherapy, radiation treatment, adjuvant aromatase inhibitor therapy, and additional hormone therapy. The specific set of covariates used for each site was chosen from this list so as to retain at least 70% of the patients from that site; hence, the exact set of covariates used differs in each site's Cox model. Moreover, several of these covariates were used as inclusion/exclusion items in the basic definitions of the three basic criteria subset groups and thus became irrelevant for those analyses.)

Table 1 Sample size by site and criteria

	Site (*N*)												
Criterion	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	0	70	124	60	212	243	0	847	5	222	179	34	1,996
2	0	127	208	98	212	304	0	898	10	289	228	69	2,443
3	174	320	282	265	214	391	801	1,140	165	516	397	270	4,935
Total	174	320	282	267	214	423	801	1,140	165	519	398	279	4,973

Table 2 Baseline patient and tumor characteristics

Characteristic	Criterion 1 (1,996 patients)		Criterion 2 (2,443 patients)		Criterion 3 (4,935 patients)	
Age at diagnosis, years: data reported in binned ages						
Median	[65–69]		[60–64]		[60–64]	
Range	55 (ages 41–95)		75 (ages 21–95)		76 (ages 21–96)	
Menopausal status —no. (%)	Menopausal status	By age	Menopausal status	By age	Menopausal status	By age
Premenopausal (age ≤ 50)	0 (0.0%)	54 (2.7%)	241 (9.9%)	414 (16.9%)	607 (12.3%)	1,207 (24.5%)
Postmenopausal (>50)	1,688 (84.6%)	1,922 (96.3%)	1,714 (70.2%)	1,997 (81.7%)	3,267 (66.2%)	3,642 (73.8%)
Not available	308 (15.4%)	20 (1.0%)	488 (20.0%)	32 (1.3%)	1,061 (21.5%)	86 (1.7%)
Tumor size—no. (%): maximum dimension of tumor reported (if multiple tumors, largest one is ≤2 cm)						
≤2 cm	1,071 (53.7%)		1,327 (54.3%)		2,303 (46.7%)	
>2 cm	752 (37.7%)		882 (36.1%)		2,182 (44.2%)	
Unknown	173 (8.7%)		234 (9.6%)		450 (9.1%)	
Nodal status—no. (%): number of positive nodes						
Zero nodes	1,243 (62.3%)		1,531 (62.7%)		2,423 (49.1%)	
1–3 nodes	407 (20.4%)		461 (18.9%)		1,281 (26.0%)	
4–9 nodes	103 (5.2%)		111 (4.5%)		438 (8.9%)	
> 9 nodes	43 (2.2%)		45 (1.8%)		185 (3.7%)	
Not available	200 (10.0%)		295 (12.1%)		608 (12.3%)	
Grading—no. (%): 0.5 to 1.49 considered G1, 1.5 to 2.49 G2, etc.						
G1	249 (12.5%)		317 (13%)		456 (9.2%)	
G2	1,148 (57.5%)		1,324 (54.2%)		1,965 (39.8%)	
G3	330 (16.5%)		398 (16.3%)		838 (17.0%)	
Unknown	269 (13.5%)		295 (12.1%)		1,676 (34.0%)	
ER status—no. (%)						
ER-positive	1,996 (100.0%)		2,443 (100.0%)		4,675 (94.7%)	
ER-negative	0 (0.0%)		0 (0.0%)		158 (3.2%)	
Unknown	0 (0.0%)		0 (0.0%)		102 (2.1%)	
PgR status—no. (%)						
PgR-positive	1,479 (74.1%)		1,847 (75.6%)		3,634 (73.6%)	
PgR-negative	273 (13.7%)		302 (12.4%)		665 (13.5%)	
Unknown	244 (12.2%)		294 (12.0%)		102 (2.1%)	
Radiotherapy—no. (%): radiation therapy						
Yes	1,138 (57.0%)		1,412 (57.8%)		2,868 (58.1%)	
No	720 (36.1%)		842 (34.5%)		1,507 (30.5%)	
Unknown	244 (12.2%)		189 (7.7%)		560 (11.3%)	
CYP2D6 metabolizer status						
Extensive	893 (44.7%)		1,077 (44.1%)		2,286 (46.3%)	
Intermediate	985 (49.3%)		1,230 (50.3%)		2,311 (46.8%)	
Poor	118 (5.9%)		136 (5.6%)		244 (4.9%)	
Unknown	0 (0.0%)		0 (0.0%)		94 (1.9%)	
CYP2D6 metabolizer types						
EM/UM	17 (0.9%)		23 (0.9%)		49 (1.0%)	
IM/UM	2 (0.1%)		2 (0.1%)		4 (0.1%)	
EM/EM	874 (43.8%)		1,052 (43.1%)		2,233 (45.2%)	
PM/UM	7 (0.4%)		7 (0.3%)		12 (0.2%)	

Table 2 Continued on next page

Table 2 Continued

Characteristic	Criterion 1 (1,996 patients)	Criterion 2 (2,443 patients)	Criterion 3 (4,935 patients)
EM/IM	327 (16.4%)	407 (16.7%)	693 (14.0%)
EM/PM	496 (24.8%)	616 (25.2%)	1,230 (25.1%)
IM/IM	64 (3.2%)	94 (3.8%)	174 (3.5%)
IM/PM	91 (4.6%)	106 (4.3%)	192 (3.9%)
PM/PM	118 (5.9%)	136 (5.6%)	244 (4.9%)
Unknown	0 (0.0%)	0 (0.0%)	94 (1.9%)
DNA source			
Blood	996 (49.9%)	1,344 (55.0%)	2,513 (50.9%)
Tumor—Frozen	431 (21.6%)	500 (20.5%)	1,575 (31.9%)
Tumor—FFPE	569 (28.5%)	598 (24.5%)	659 (13.4%)
Normal—FFPE	0 (0.0%)	0 (0.0%)	174 (3.5%)
Unknown	0 (0.0%)	1 (0.0%)	14 (0.3%)

CYP2D6, cytochrome P450 2D6; EM, extensive metabolizer; ER, estrogen receptor; FFPE, formalin-fixed–paraffin-embedded; IM, intermediate metabolizer; PgR, progesterone receptor; PM, poor metabolizer; UM, unknown metabolizer.

Table 3 Meta-analyses of CYP2D6 HRs on clinical outcome in inclusion/exclusion criteria subsets

	IDFS				BCFI			
	Meta-estimates		*P* value		Meta-estimates		*P* value	
	HR	95% CI	Homog[a]	Association[b]	HR	95% CI	Homog[a]	Association[a]
Criterion 1	1.25	(1.06,1.47)	0.899	0.009	1.27	(1.01,1.61)	0.858	0.041
Criterion 2	1.17	(0.90,1.52)	0.055	0.249	1.21	(0.889,1.65)	0.130	0.224
Criterion 3	1.07	(0.92,1.26)	0.099	0.382	1.10	(0.868,1.35)	0.114	0.352

BCFI, breast cancer–free interval; Homog, homogeneity; HR, hazard ratio; IDFS, invasive disease–free survival; ITPC, International Tamoxifen Pharmacogenomics Consortium.

[a]The homogeneity *P* value tests the hypothesis that the individual ITPC site estimates meet the statistical random-effects modeling assumptions of the meta-analysis. A significant value indicates that there is significant heterogeneity among the sites, which casts doubt on the "combinability" of the studies for that parameter and on the validity of the corresponding association test. [b]The association *P* value tests the hypothesis that the combined meta-analysis estimate of the HR is significantly different from the null hypothesis value of HR = 1.

Site-specific product-limit estimates of the three CYP2D6 metabolizer status genotype groups (EM, IM, and PM) are shown in **Figures 2** and 3 for criterion 1 patients. Sites 1 and 7 had no subjects who met inclusion/exclusion for criterion 1. The corresponding figures for patients meeting criteria 2 and 3 are shown in **S6** online. As seen in **Figure 2**, for IDFS sites, 3, 5, and 8 show a strong significant effect in the direction expected by the known pharmacokinetic effects of *CYP2D6* on endoxifen exposure, namely, a poorer clinical response for the IM and/or PM genotype groups. Other sites show a trend in the expected direction between the IM and EM groups, but the much smaller PM group is often inconsistent with the expectation, and the separation in the three survival curves is not strong enough to reach statistical significance (e.g., sites 6 and 12). For some sites, there is no hint of any significant difference (e.g., sites 2, 4, 10, and 11), and for one of these, site 2, the direction of effect is exactly opposite than expected. There is a danger in overinterpreting such "trends" (either in favor or against expectation) when there is no statistically significant difference, because some level of site-to-site variation is to be expected. The key question is not whether such variation exists but whether it centers over the null hypothesis or over the alternative; this is the question that the meta-analysis is designed to answer. However, these simple product-limit survival curves show great study-to-study heterogeneity, which complicates both the analyses and the interpretation. We have similar heterogeneous results for the breast cancer–free interval outcome, shown in **Figure 3**. The corresponding figures in **S6** online show a similar pattern for the subsets of patients meeting criteria 2 and 3, although the heterogeneity seems to be even more pronounced as the exclusion criteria are loosened. This is not a surprising result, considering that the criteria themselves impose a certain level of homogeneity.

DISCUSSION

Prospective pharmacology studies consistently demonstrate that *CYP2D6* genetic variants are associated with variable plasma concentrations of endoxifen.[4,6] Endoxifen exposure is related to duration of tamoxifen use and dose, wherein an increase in the tamoxifen dose (from 20 to 40 mg daily) significantly increases endoxifen exposure in patients with reduced or null CYP2D6 metabolism but not in CYP2D6 EMs.[7] However, coadministration of CYP2D6-inhibiting drugs[4] reduces CYP2D6 enzyme activity, and nonadherence to tamoxifen is more commonly observed in patients with normal or increased CYP2D6 metabolism.[8]

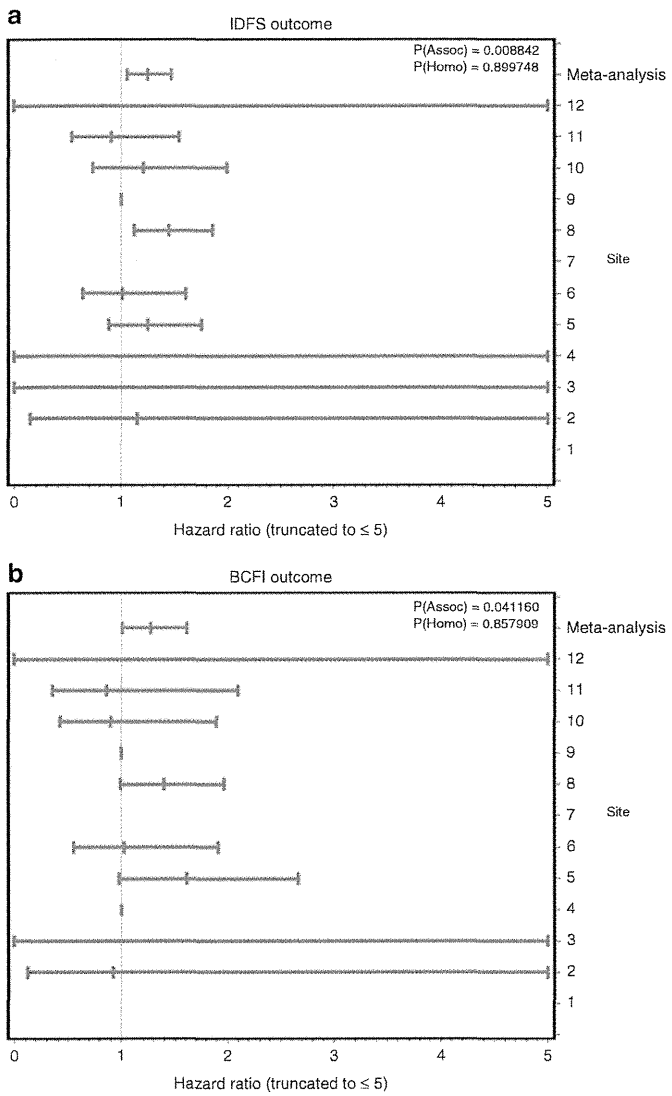

Figure 1 Individual site estimates of hazard ratios of *CYP2D6* genotype on clinical outcome, along with the meta-analyses for the criterion 1 subset. (**a**) Invasive disease–free survival (IDFS) outcome. (**b**) Breast cancer–free interval (BCFI) outcome.

Despite the consistent pharmacogenetic effects of *CYP2D6* on endoxifen exposure, there is considerable controversy regarding the validity of *CYP2D6* as a predictor of tamoxifen outcome.[9,10] Although recent secondary analyses from the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial and the Breast International Group (BIG) 1-98 study[11,12] did not demonstrate an association between *CYP2D6* and tamoxifen outcome, these studies provoked criticism due to concerns regarding genotyping error and the analysis of small subsets of the main trials.[13–16]

By contrast, a secondary analysis from another large prospective adjuvant tamoxifen trial, the Austrian Breast and Colorectal Cancer Study Group 8 (ABCSG 8), demonstrated that for women treated with 5 years of adjuvant tamoxifen at a dose of 20 mg/day, CYP2D6 PMs had a statistically significant higher odds of recurrence or death as compared with CYP2D6 EMs, and CYP2D6 PMs/IMs and PMs/EMs tended to exhibit a higher odds of recurrence as compared with patients without the PM alleles. However, this effect was not observed for patients who had switched to anastrozole, a drug not metabolized by CYP2D6. These data suggest that the effects of *CYP2D6* genotype may be masked if patients receive a shorter duration of tamoxifen or other active drugs besides tamoxifen, which alter the hazard for recurrence.[17]

We approached the tamoxifen controversy by performing a global meta-analysis of available clinical and *CYP2D6* genetic data of tamoxifen-treated breast cancer patients. All groups from across the world with both published and unpublished *CYP2D6* data were invited to participate. We initially presented a pooled analysis of these data,[5] in which we found no association between *CYP2D6* and IDFS. Following this presentation, we developed a new analysis plan (not defined before the initial negative presentation), which included the following: (i) articulation of three criteria to analyze the data according to the quality of the genetic and clinical data, (ii) additional curation to obtain missing clinical and genetic data, and (iii) a new statistical analysis plan, which applied a random-effects meta-analysis strategy instead of a pooled analysis strategy. Notably, Criterion 1 is most stringent, requiring strict control for as many pharmacologic factors as possible known to affect endoxifen exposure, which include use of tamoxifen monotherapy, genotyping of multiple *CYP2D6* alleles for accurate CYP2D6 phenotype assignment, use of one tamoxifen dose (20 mg), and intended duration of tamoxifen use for 5 years. In addition, eligibility for this cohort was restricted to women with invasive ER-positive status, postmenopausal breast cancer, and the requirement for annual follow-up, parameters required in any prospective clinical trial and that were requirements of criterion 1 (patients who were knowingly not followed were excluded from criterion 1), but not from criteria 2 and 3. These factors may have contributed to the substantial increase in heterogeneity comparing criterion 1 with criteria 2 and 3. However, it should be noted that these criteria impose a certain bias because the majority of negative studies submitted to the ITPC were observed in criteria 2 and 3.

In general, a substantial number of subjects comprising criterion 3 had misclassification of the predicted drug metabolism phenotype due to the lack of a comprehensive coverage of loss-of-function alleles.[18,19] More than 20 loss-of-function alleles out of 100 known *CYP2D6* genetic variants contributed to a frequency of ~8% of PMs in a population of European descent. Limiting the analysis to the most common such allele, *CYP2D6*4*, as was frequently done in the older published literature, will result in misclassification of 35% of PMs, thereby falsely assigning the undetected PMs to the EM or IM groups. Notably, 871/1,996 patients comprising criterion 1 had optimal CYP2D6 phenotype assessment obtained by AmpliChip genotyping, and this may have contributed to the robustness of criterion 1 results, which demonstrated an association between *CYP2D6* and tamoxifen treatment outcome (breast cancer–free interval: HR = 1.27, 95% confidence interval = 1.01–1.61).

The ITPC intended to perform a global study including several thousand patient samples; however, the majority of the subjects were not comprehensively genotyped because DNA was not of sufficient quality. We performed a subgroup analysis using patient samples for which full coverage of alleles by the

AmpliChip genotyping platform was available using criterion 1 (871/1,635 AmpliChip-genotyped subjects met criterion 1). When confined to the Amplichip subjects, the estimates of the pharmacodynamic HRs for *CYP2D6* were similar to what they were for the entire set of subjects meeting criterion 1.

A major source of potential genotyping errors may be related to DNA source. *CYP2D6* is one of the most difficult genes to genotype because of the numerous polymorphisms and adjacent pseudogenes. Some platforms cannot detect the presence of the *5 deletion, particularly in DNA derived from

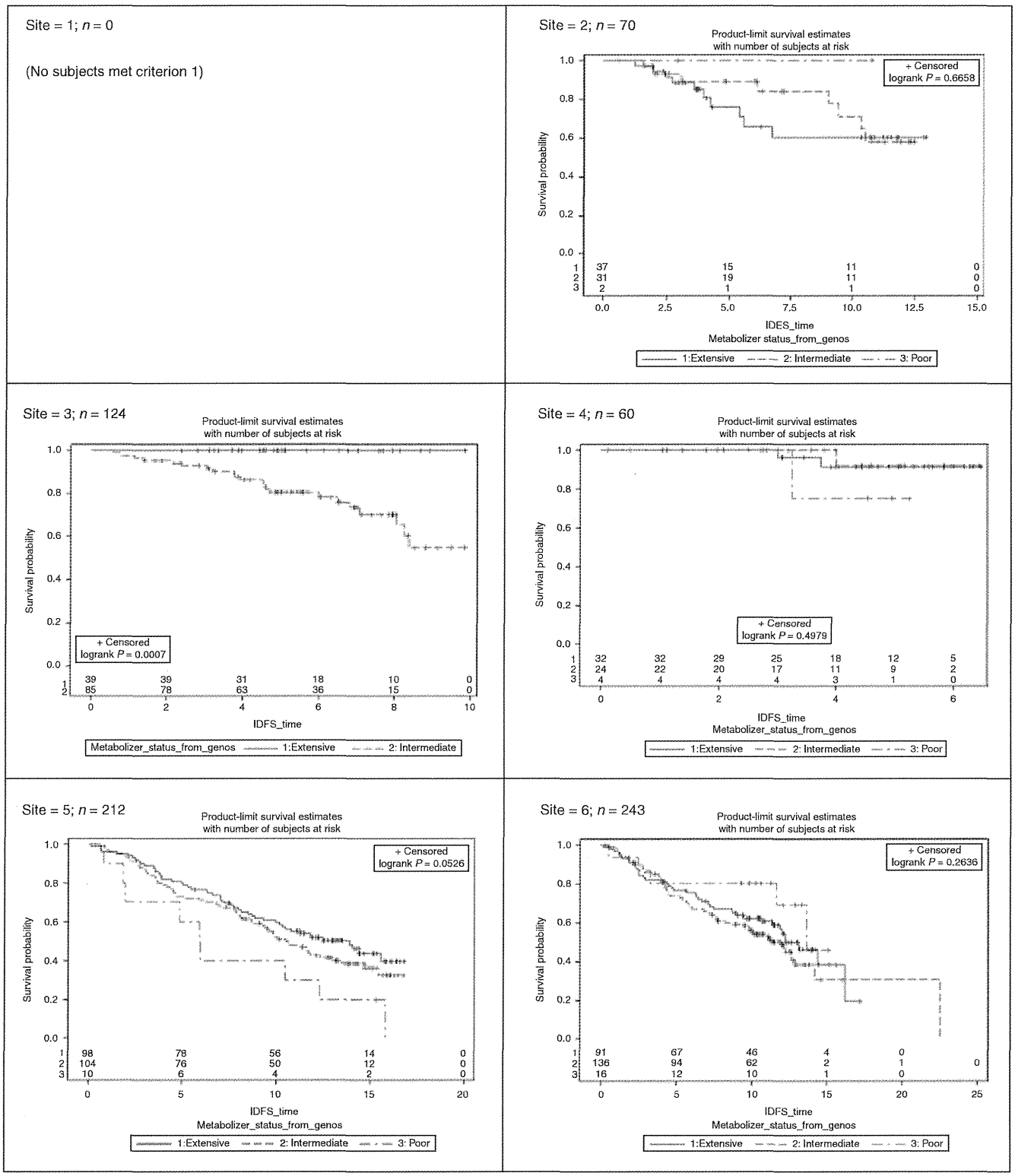

Figure 2 Site-specific effects of CYP2D6 metabolizer status on clinical outcomes for subjects meeting inclusion criterion 1 (outcome = invasive disease-free survival (IDFS)).

formalin-fixed–paraffin-embedded (FFPE) tissue. However, several sites used multiple platforms to validate their genotyping data, reducing potential genotyping errors across the entire data set. Importantly, *CYP2D6* genotypes obtained from blood-derived DNA reflect the patients' germ-line genotypes, known to influence endoxifen plasma concentrations. By contrast, *CYP2D6* genotypes from tumor-derived DNA may be subject to error due to somatic mutation by loss of heterozygosity, known to affect the *CYP2D6* locus at 22q13 in up to 30% of breast tumors.[20–22] Thus, when *CYP2D6* genotype is derived from tumor samples, an excess number of homozygotes may result as a consequence of loss of heterozygosity. This form of genotyping error is revealed by Hardy–Weinberg Equilibrium (HWE) testing, as was observed in the Breast

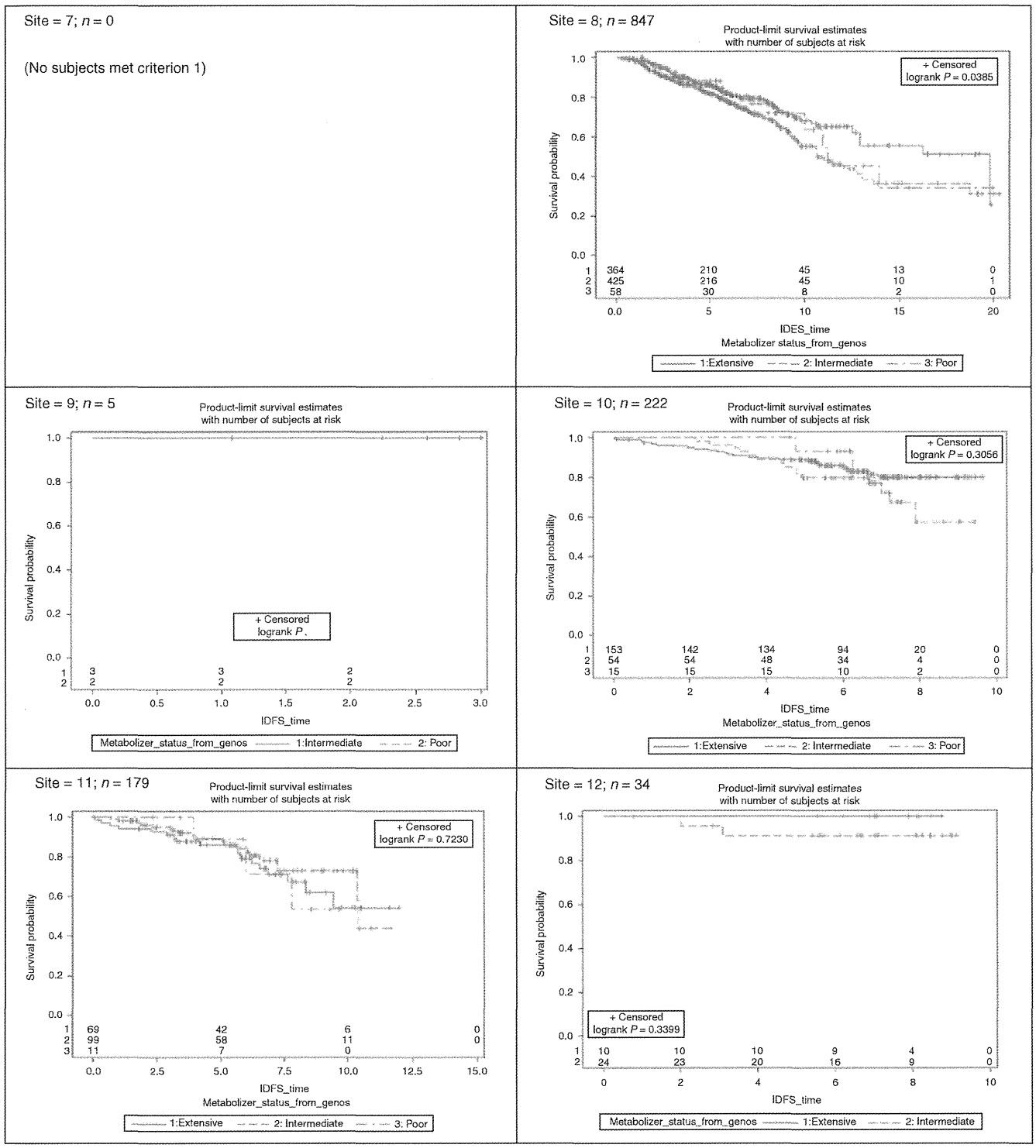

Figure 2 Continued

International Group 1-98 study, in which strong departures from HWE (to a magnitude of 10^{-92}) were observed, leading to a call for retraction of this article.[3,12,16]

For criterion 1, 49.9% of our patient DNA samples originated from blood, 21.6% from fresh-frozen tissues, and 28.5% from FFPE tissues. For criterion 2, 55.0% samples originated from blood, 20.5% were fresh-frozen tissues, and 24.5% from FFPE tissues. For criterion 3, 50.9% of DNA samples originated from blood, 31.9% from fresh-frozen tumor, 13.4% from FFPE tumor tissues, and 3.5% from FFPE normal tissue. Although we cannot exclude the presence of somatic events leading to misclassification of *CYP2D6* genotype, as evident from HWE deviation identified in data from

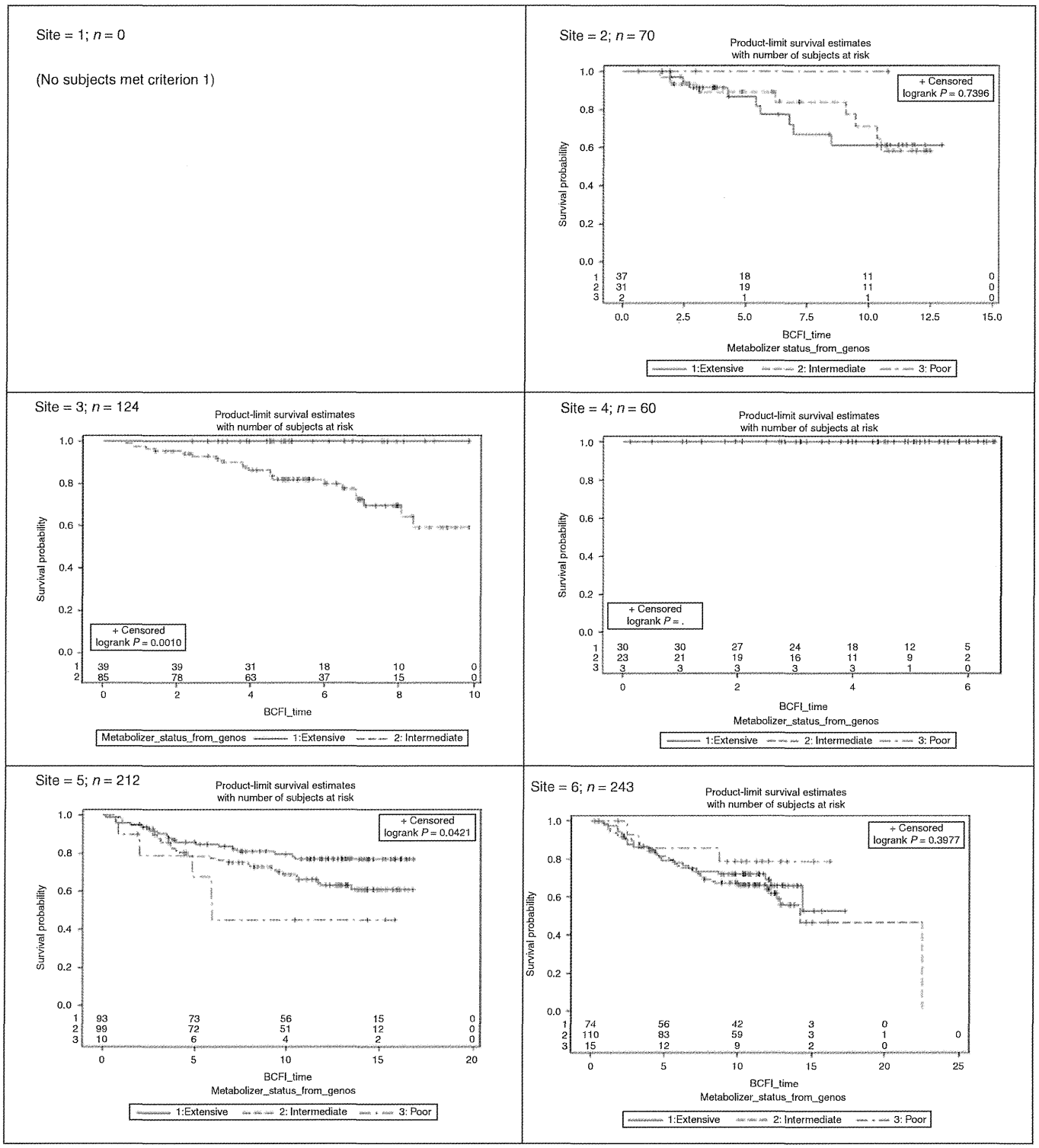

Figure 3 Site-specific effects of CYP2D6 metabolizer status on clinical outcomes for subjects meeting inclusion criterion 1 (outcome = breast cancer–free interval (BCFI)).

some sites, comprehensive testing for HWE did not reveal significant violations across most sites. Moreover, the extent of deviation from HWE in the **4* allele was not associated with sites that evinced less clinical benefit from tamoxifen in patients who were assessed to be PMs in terms of their *CYP2D6* status. This suggests that genotyping errors are unlikely to be a major issue in our analyses.

Our findings are subject to the shortcomings commonly encountered when performing retrospective "biomarker" studies. In our study, most sites were unable to collect or control for the factors known to alter endoxifen exposure, including dose and duration of tamoxifen administration and patients' adherence to the regimen. Although tamoxifen adherence is increasingly recognized as a critical factor for drug efficacy,[23] most studies evaluating tamoxifen biomarkers have not controlled for adherence. Other confounders include limited *CYP2D6* allele coverage and lack of information

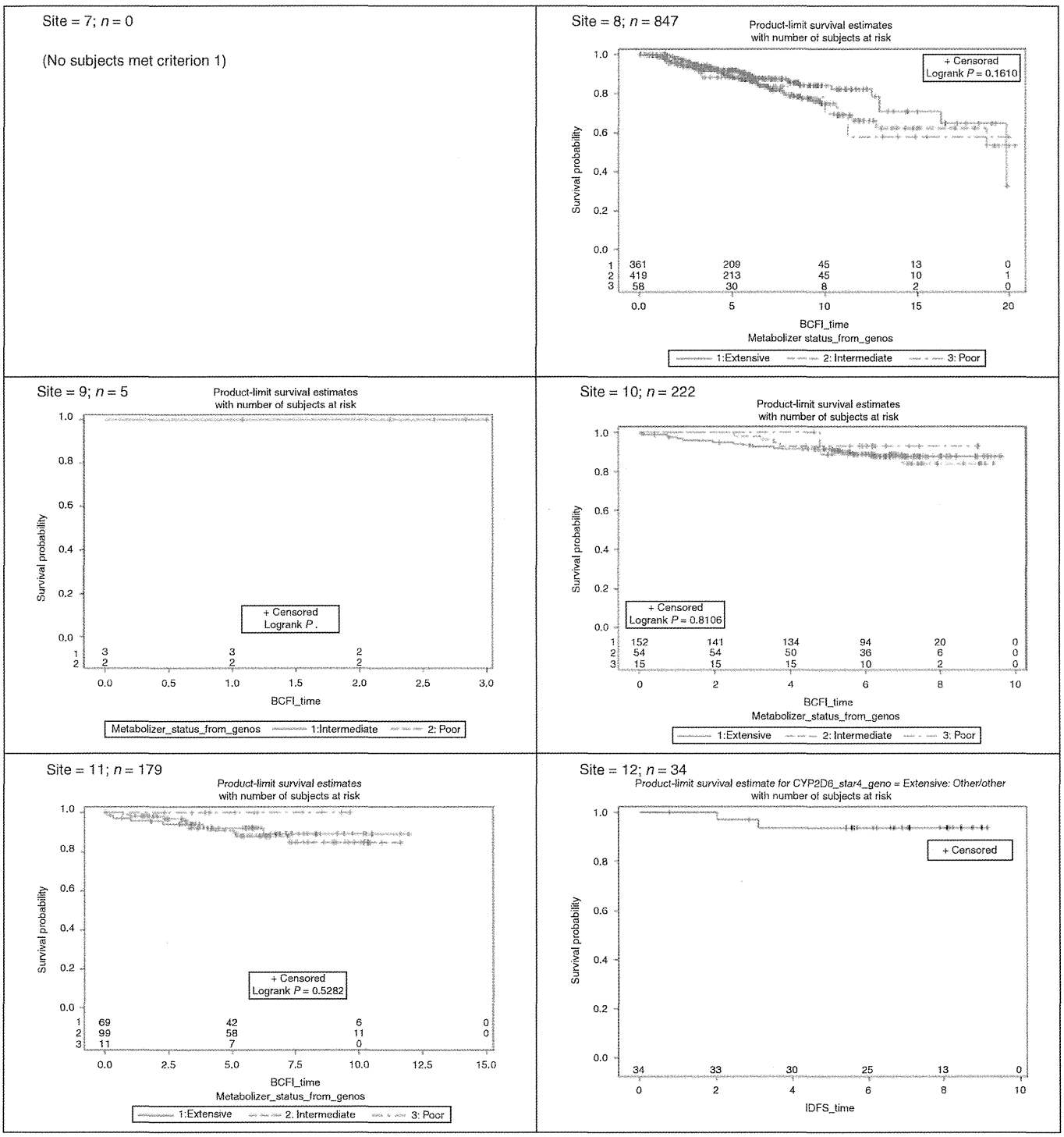

Figure 3 Continued

regarding the coadministration of CYP2D6 inhibitors, leading to potential misclassification of the CYP2D6 drug metabolism phenotype. Therefore, our meta-analysis results depend heavily on which subgroup of patients we include. If we accept that utmost precautions must be applied to avoid the distortion of results from influences derived from the aforementioned shortcomings, it follows that merely increasing the numbers of subjects without controlling the quality of input data, as done in our preliminary overview analysis,[5] may result in heterogeneity that masks the effect of a pharmacokinetic biomarker such as *CYP2D6*. From this, we conclude that until results from prospective adjuvant studies are available, women who meet criterion 1 as established in this and other independent cohorts (ABCSG 8) should be counseled regarding the potential impact of *CYP2D6* on the effectiveness of adjuvant tamoxifen, and potent CYP2D6 inhibitors should be avoided in these patients. Prospective adjuvant studies are needed to determine whether genotype-guided selection of hormonal therapy will improve the outcomes of women with early-stage ER-positive breast cancer, and results from ongoing prospective studies in the metastatic setting are eagerly awaited. A similarly motivated study on warfarin is currently being conducted in the Clarification of Optimal Anticoagulation through Genetics trial.[24]

By strict clinical and genotype criteria, reduced CYP2D6 metabolism is associated with a higher risk of recurrence (as measured by IDFS) in tamoxifen-treated women. However, the heterogeneity observed across sites contributing data to the ITPC points to the likely influence of critical confounding factors unlikely to be controllable in global retrospective studies. This study demonstrates the complexity of performing a retrospective biomarker study that focuses on the genetic factors that affect exposure to an active metabolite, endoxifen, for a drug, tamoxifen, administered for 5 years. Our observation that <50% of the patients in this study met the basic eligibility criteria—in terms of similar disease, treatment, and control for critical pharmacological factors such as dose and duration of tamoxifen—provides insight into possible reasons for the discrepancies in the literature on *CYP2D6* and tamoxifen. Although *CYP2D6* is a predictor of IDFS in a subset of patients treated with tamoxifen, the lack of an effect in the entire heterogeneous study population suggests that prospective studies are necessary to finally establish whether genotype-guided selection of hormonal therapy improves clinical outcomes of women with ER-positive breast cancer.

METHODS

Data collection and study cohorts. The ITPC invited any research group from across the world that had published or unpublished *CYP2D6* data to participate in this meta-analysis. The ITPC comprises 12 research projects for a total of 4,973 breast cancer patients treated with tamoxifen. This retrospective study does not include a control group not treated with tamoxifen. These data were curated at the PharmGKB (Pharmacogenomics Knowledge Base, http://www.pharmgkb.org). Consent for participation in the ITPC and DNA collection, *CYP2D6* genetic testing, and submission of data was obtained under local ethical review board permissions.

We collected information on clinical factors previously shown to be associated with breast cancer therapy and prognosis that were available from the information received from the sites. These data included demographic characteristics, cancer history, cancer recurrence, use of other therapies, use of concomitant medications known to affect *CYP2D6* phenotype, ER status, and classic prognostic factors such as tumor size and number of affected lymph nodes. Information was also collected regarding the presence of *CYP2D6* genetic variants (**2, *3, *4, *5, *6, *10, *17,* and **41*, categorized by their DNA sources), for which coverage of these alleles varied by site. For 1,635 subjects, *CYP2D6* variants assessable from blood DNA using the AmpliChip CYP450 test (Roche) were collected. A complete list of the information collected is detailed in **S1–S3** online, including the project-specific *CYP2D6* genotype assays used and the DNA source. Independent confirmation of *CYP2D6* genotypes was not performed owing to lack of access to subjects' samples. The clinical outcome variable was either breast cancer–free interval or IDFS, as previously defined.[25] The complete data set of genotypes and clinical variables is available at http://www.pharmgkb.org.

Statistical analysis. Because the ITPC was not a prospectively defined multicenter study with a common protocol, there is potential for considerable study-to-study heterogeneity. Therefore, we did not analyze the combined data as a single series even though we had access to individual-level data from all studies. Rather, we applied a random-effects meta-analysis strategy. This provided estimates of the effect of *CYP2D6* in each study's data separately, allowing us to examine the consistency of the results across sites. The meta-analysis is a two-stage procedure. In the first stage, we fit proportional-hazards models to the data from each of the ITPC sites separately, predicting clinical outcome after surgery from *CYP2D6* genotype and other relevant covariates. These analyses produced a set of 12 parameter estimates of the HRs of *CYP2D6* genotypes on outcome, along with their corresponding SEs (one for each site). In the second stage, we used a random-effects meta-analysis procedure[26] to test for study heterogeneity (i.e., whether the 12 studies met the assumptions of the meta-analysis sufficiently so as to be combinable using that method). When the heterogeneity was not significant, we combined the log-HRs into a single, meta-analysis estimate of the effect of *CYP2D6* on tamoxifen-treated recurrence and/or survival outcomes. The DerSimonian and Laird method also provides a penalty in its test of overall association for moderate levels of study-to-study heterogeneity (i.e., for heterogeneity that is not so severe as to be statistically significant). This method is therefore conservative in its conclusions when heterogeneity is a potential issue.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/cpt

ACKNOWLEDGMENTS

The authors acknowledge useful conversations with Donald A. Berry (The University of Texas MD Anderson Cancer Center). The complete data set of genotypes and clinical variables, analysis codes, and full analyses is available to registered PharmGKB users at http://www.pharmgkb.org. We are grateful to all breast cancer patients for their participation. We thank the physicians and other hospital staff, scientists, research assistants, and study staff who contributed to the patient recruitment, data collection, and sample preparation.

This work was supported by grants from the National Institutes of Health (National Institute of General Medical Sciences, the National Cancer Institute, and the National Institute of Child Health and Human Development R24 GM61374, CA-25224, CA-37404, CA-15083, CA-35113, CA-35269, CA-63849, CA35103, CA-35195, CA-35272, CA-35101, CA-37417, CA-35415, CA-52352, CA-35448, CA-60274, 1R01CA133049-01, U19 GM61388, CA-116201, CA-58207, CA-51008, T32 GM007546,

and CA051008; Breast Cancer Research (Scotland); Tayside Tissue Bank; California Breast Cancer Research Program grant 14OB-0166; Cancer Research UK; Deutsches Krebsforschungszentrum, Heidelberg, Germany; Robert Bosch Foundation, Stuttgart, Germany; 7FP EU Marie Curie Initial Training Network "FightingDrugFailure" (GA 238132); GrantStichting Emmanuel van der Schueren (scientific partner of the Vlaamse Liga tegen Kanker); and the National Project for Personalized Genomic Medicine, Ministry for Health & Welfare, Republic of Korea (A111218-PG02). W.S. is supported by grant Deutsche Forschungsgemeinschaft SCHR 1323/2-1, Germany.

The institutions in the International Tamoxifen Pharmacogenomics Consortium are listed below.:

Writing committee:

Michael A. Province, Matthew P. Goetz, David A. Flockhart, Hiltrud Brauch, and Teri E. Klein

Data contributors:

Deutsches Krebsforschungszentrum, Heidelberg, Germany: Ute Hamann, Julia Boländer, and Hans-Ulrich Ulmer, Frauenklinik Städtisches Klinikum Karlsruhe, Germany

Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology Stuttgart and University of Tuebingen, Tuebingen, Germany: Hiltrud Brauch, Werner Schroth, Matthias Schwab, Stefan Winter, Michel Eichelbaum, Peter Fritz, and Wolfgang Simon

Division of Research, Federal Institute for Drugs and Medical Devices, and University of Bonn Medical Faculty, Bonn, Germany: Julia C. Stingl

Indiana University and University of Michigan, USA: David A. Flockhart and Anne T. Nguyen

Inje University College of Medicine and Department of Clinical Pharmacology, Busan Paik Hospital, Busan, Korea: Jae-Gook Shin and Ji-Yeob Choi

Mayo Clinic in Rochester, Minnesota, USA: Matthew P. Goetz, James N. Ingle, Vera J. Suman, and Richard M. Weinshilboum

Ninewells Hospital and Medical School, Dundee, UK: Colin A. Purdie and Lee B. Jordan

Örebro University, Sweden: Pia Wegman

RIKEN Center for Genomic Medicine and University of Tokyo, Japan: Hitoshi Zembutsu, Taisei Mushiroda, and Kazuma Kiyotani

Roswell Park Cancer Institute, USA: Christine B. Ambrosone

University Breast Center Franconia, University Hospital Erlangen, Erlangen, Germany: Peter A Fasching, Reiner Strick, and Matthias W Beckmann

University Hospitals Leuven, Belgium: Patrick Neven and Anne-Sophie Dieudonné

University Institute of Medical and Chemical Laboratory Diagnostics, Paracelsus Private Medical University, Salzburg, Austria: Elisabeth Haschke-Becher

University of California, San Francisco, San Francisco, USA: Alan H.B. Wu, Wendy Lorizio, and Elad Ziv

University of Dundee, UK: Philip Quinlan, Lee B. Jordan, and Colin A. Purdie

Department of Gynecology and Obstetrics, University of Mainz, Mainz, Germany: Marcus Schmidt and Heinz Koelbl

University of Manchester, Manchester, UK: William G. Newman, Roberta Ferraldeschi, Anthony Howell, and Ayse Latif

Vesalius Research Center, Catholic University of Leuven, Leuven, Belgium: Diether Lambrechts

Yonsei University Health System, Korea: Byeong-Woo Park

Statistical Analysis:

Mayo Clinic in Rochester, MN, USA: Vera J. Suman

Stanford University, Stanford, CA, USA: Teri E. Klein and Ryan Whaley

Washington University in St. Louis, MO, USA: Michael A. Province

Data Curation:

Stanford University, Stanford, CA, USA: Teri E. Klein, Ryan Whaley, Joan M. Hebert, and Li Gong

Site Principal Investigators:

Russ B. Altman and Teri E. Klein (Stanford University), Christine A. Ambrosone (Roswell Park Cancer Institute), Hiltrud Brauch (IKP Stuttgart), Jae-Gook Shin (Inje University), Matthew P. Goetz (Mayo Clinic), William G. Newman (University of Manchester), Patrick Neven (Belgium), Alastair M. Thompson (University of Dundee), Sten Wingren (Örebro University), Hitoshi Zembutsu (University of Tokyo), and Elad Ziv (University of California, San Francisco)

Steering Committee:

Russ B. Altman, Hiltrud Brauch, Michel Eichelbaum, David A. Flockhart, Matthew P. Goetz, James N. Ingle, V. Craig Jordan, Teri E. Klein, Kent Osborne, Yusuke Nakamura, and Richard M. Weinshilboum

Participating sites:

The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA

Department of Clinical Medicine, Örebro University, Örebro, Sweden

Department of Genetics, Stanford University, Stanford, CA, USA

Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

Department of Gynecology and Obstetrics, University Hospitals Leuven, Leuven, Belgium

Department of Oncology, Catholic University Leuven, Leuven, Belgium

Department of Medicine/GIM and Clinical Pharmacology and Experimental Therapeutics, and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA

Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

Department of Oncology, Georgetown University, Washington, DC, USA

Department of Oncology, Mayo Clinic, Rochester, MN, USA

Department of Oncology and Pharmacology, Mayo Clinic, Rochester, MN, USA

Department of Pharmacology; Mayo Clinic, Rochester, MN, USA

Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine and Department of Clinical Pharmacology, Inje University, Busan Paik Hospital, Busan, Korea

Department of Surgery, Yonsei University Health System, Seoul, Korea

Deutsches Krebsforschungszentrum, Heidelberg, Germany

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

Division of Clinical Pharmacology, School of Medicine, Indiana University, IN, USA

Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology Stuttgart, and University of Tuebingen, Germany

Dundee Cancer Centre, Dundee, UK

Genetic Medicine, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine, Yokohama, Japan

Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum, Heidelberg, Germany

Ninewells Hospital and Medical School, Dundee, UK

CONFLICT OF INTEREST

A.M.T. and W.G.N. report Roche funding for genotyping H.B. and M. Schwab report that they have initiated scientific collaborations in 2009 with Roche Molecular Diagnostics and Siemens Healthcare Diagnostics Products, respectively P.A.F. and M.W.B. report Novartis research funding. M.-T.M.L. is a paid consultant of YongLin Health Foundation. R.B.A. is a founder, equity holder, and consultant for Personalis. The other authors declared no conflict of interest.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

✓ There has been extensive controversy with regard to the association between *CYP2D6* genetic variants and the clinical outcomes of tamoxifen use.

WHAT QUESTION DID THIS STUDY ADDRESS?

✓ The ITPC was established to address this controversy and to determine the association of *CYP2D6* status with IDFS in tamoxifen-treated early-stage, ER-positive breast cancer.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

✓ We found that *CYP2D6* genotype was associated with a higher risk of recurrence in patients meeting the strict criterion. However, the observation of substantial heterogeneity in cohorts 2 and 3 suggests that study design factors that cannot be controlled retrospectively may obscure the predictive utility of *CYP2D6* genotype. This study demonstrates the complexity of performing a retrospective biomarker study.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

✓ Although *CYP2D6* is a predictor of IDFS in a subset of patients treated with tamoxifen monotherapy, the lack of an effect in the entire heterogeneous study population suggests that prospective studies are necessary to fully establish the value of *CYP2D6* genotyping in tamoxifen therapy.

1. Jordan, V.C. Tamoxifen: a most unlikely pioneering medicine. *Nat. Rev. Drug Discov.* **2**, 205–213 (2003).
2. Davies, C. *et al.* Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* **378**, 771–784 (2011).
3. Johnson, M.D. *et al.* Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res. Treat.* **85**, 151–159 (2004).
4. Stearns, V. *et al.* Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J. Natl. Cancer Inst.* **95**, 1758–1764 (2003).
5. Goetz, M.P., Berry, D.A. & Klein, T.E. 31136 Findings from the International Tamoxifen Pharmacogenomics Consortium. Proceedings of the San Antonio Breast Cancer Symposium, San Antonio, Texas, 9–13 December 2009. Abstract 33.
6. Mürdter, T.E. *et al.*; German Tamoxifen and AI Clinicians Group. 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.* **89**, 708–717 (2011).
7. Irvin, W.J. Jr *et al.* Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: a multicenter study. *J. Clin. Oncol.* **29**, 3232–3239 (2011).
8. Rae, J.M. *et al.*; COBRA investigators. Cytochrome P450 2D6 activity predicts discontinuation of tamoxifen therapy in breast cancer patients. *Pharmacogenomics J.* **9**, 258–264 (2009).
9. Schroth, W. *et al.* Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* **302**, 1429–1436 (2009).
10. Goetz, M.P. *et al.* Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J. Clin. Oncol.* **23**, 9312–9318 (2005).
11. Rae, J.M. *et al.*; ATAC trialists. CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J. Natl. Cancer Inst.* **104**, 452–460 (2012).
12. Regan, M.M. *et al.*; Breast International Group (BIG) 1-98 Collaborative Group. CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the breast international group 1-98 trial. *J. Natl. Cancer Inst.* **104**, 441–451 (2012).
13. Brauch, H. *et al.* Tamoxifen use in postmenopausal breast cancer: CYP2D6 matters. *J. Clin. Oncol.* **31**, 176–180 (2013).
14. Nakamura, Y., Ratain, M.J., Cox, N.J., McLeod, H.L., Kroetz, D.L. & Flockhart, D.A. Re: CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1-98 trial. *J. Natl. Cancer Inst.* **104**, 1264; author reply 1266–1268 (2012).
15. Stanton, V. Jr. Re: CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1-98 trial. *J. Natl. Cancer Inst.* **104**, 1265–6; author reply 1266–8 (2012).
16. Goldberg, P. Experts claim errors in breast cancer study, demand retraction of practice-changing paper. *The Cancer Letter* **38**, 1–11 (2012).
17. Goetz, M.P. *et al.* CYP2D6 metabolism and patient outcome in the Austrian Breast and Colorectal Cancer Study Group trial (ABCSG) 8. *Clin. Cancer Res.* **19**, 500–507 (2013).
18. Schroth, W. *et al.* CYP2D6 polymorphisms as predictors of outcome in breast cancer patients treated with tamoxifen: expanded polymorphism coverage improves risk stratification. *Clin. Cancer Res.* **16**, 4468–4477 (2010).
19. Zanger, U.M., Turpeinen, M., Klein, K. & Schwab, M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal. Bioanal. Chem.* **392**, 1093–1108 (2008).
20. Castells, A., Gusella, J.F., Ramesh, V. & Rustgi, A.K. A region of deletion on chromosome 22q13 is common to human breast and colorectal cancers. *Cancer Res.* **60**, 2836–2839 (2000).
21. Hirano, A. *et al.* Allelic losses of loci at 3p25.1, 8p22, 13q12, 17p13.3, and 22q13 correlate with postoperative recurrence in breast cancer. *Clin. Cancer Res.* **7**, 876–882 (2001).
22. Loo, L.W. *et al.* Differential patterns of allelic loss in estrogen receptor-positive infiltrating lobular and ductal breast cancer. *Genes. Chromosomes Cancer* **47**, 1049–1066 (2008).
23. Thompson, A.M. *et al.* Comprehensive CYP2D6 genotype and adherence affect outcome in breast cancer patients treated with tamoxifen monotherapy. *Breast Cancer Res. Treat.* **125**, 279–287 (2011).
24. French, B. *et al.*; COAG (Clarification of Optimal Anticoagulation through Genetics) Investigators. Statistical design of personalized medicine interventions: the Clarification of Optimal Anticoagulation through Genetics (COAG) trial. *Trials* **11**, 108 (2010).
25. Hudis, C.A. *et al.* Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J. Clin. Oncol.* **25**, 2127–2132 (2007).
26. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).

Breast Cancer RESEARCH

RESEARCH ARTICLE Open Access

A genome-wide association study of chemotherapy-induced alopecia in breast cancer patients

Suyoun Chung[1,6†], Siew Kee Low[3,6†], Hitoshi Zembutsu[6], Atsushi Takahashi[3], Michiaki Kubo[4], Mitsunori Sasa[5] and Yusuke Nakamura[1,2,6*]

Abstract

Introduction: Chemotherapy induced alopecia is one of the most common adverse events caused by conventional cytotoxic chemotherapy, yet there has been very little progress in the prevention or treatment of this side effect. Although this is not a life threatening event, alopecia is very psychologically difficult for many women to manage. In order to improve the quality of life for these women, it is important to elucidate the molecular mechanisms of chemotherapy induced alopecia and develop ways to effectively prevent and/or treat it. To identify the genetic risk factors associated with chemotherapy induced alopecia, we conducted a genome wide association study (GWAS) using DNA samples from breast cancer patients who were treated with chemotherapy.

Methods: We performed a case control association study of 303 individuals who developed grade 2 alopecia, and compared them with 880 breast cancer patients who did not show hair loss after being treated with conventional chemotherapy. In addition, we separately analyzed a subset of patients who received specific combination therapies by GWASs and applied the weighted genetic risk scoring (wGRS) system to investigate the cumulative effects of the associated SNPs.

Results: We identified an SNP significantly associated with drug induced grade 2 alopecia (rs3820706 in *CACNB4* (calcium channel voltage dependent subunit beta 4) on 2q23, $P = 8.13 \times 10^{-9}$, OR = 3.71) and detected several SNPs that showed some suggestive associations by subgroup analyses. We also classified patients into four groups on the basis of wGRS analysis and found that patients who classified in the highest risk group showed 443 times higher risk of antimicrotubule agents induced alopecia than the lowest risk group.

Conclusions: Our study suggests several associated genes and should shed some light on the molecular mechanism of alopecia in chemotherapy treated breast cancer patients and hopefully will contribute to development of interventions that will improve the quality of life (QOL) of cancer patients.

Introduction

Breast cancer is the most common malignancy among women worldwide [1]. Although treatment of breast cancer has been significantly improved by the development of molecular-targeted drugs in the past few decades, a subset of patients do not receive benefit from these modalities [2,3]. Such patients and the majority of relapsed patients are treated with conventional cytotoxic chemotherapy that can often cause various adverse events including hair loss.

Hair loss (alopecia) is one of the most common side effects caused by chemotherapy in cancer patients, particularly in women with breast cancer. Although molecular-targeted drugs such as trastuzumab do not cause alopecia, these drugs are given together with other chemotherapeutic agents. Most of the cytotoxic agents cause alopecia, but the severity in individual patients and the incidence by the types of drugs are significantly different: more than 80% of patients treated with antimicrotubule agents, more than 60% of those with

* Correspondence: ynakamura@bsd.uchicago.edu
†Equal contributors
[1]Department of Medicine, The University of Chicago, 5801 South Ellis Avenue, Chicago, IL 60637, USA
[2]Department of Surgery, The University of Chicago, A27 S Maryland Avenue, Chicago, IL 60637, USA
Full list of author information is available at the end of the article

alkylating agents, 60 to 100% of those with topoisomerase inhibitors, and 10 to 50% of those with antimetabolite-based drugs experience severe alopecia [4]. It is also well known that the incidence and the severity are increased when patients are treated with a combination of multiple drugs rather than a single agent [4,5]. Usually, hair loss begins one to two weeks after the start of chemotherapy and a patient's hair can be completely lost in a one- to two-month period. Hair starts to regrow after chemotherapy is completed or discontinued [6,7]. This drug-induced hair loss is not a life-threatening side effect, however, it can strongly influence cosmetic appearance and psychological stresses, and often affects the quality of life (QOL) of the patients [7]. Several studies have demonstrated that the majority of women patients are distressed due to treatment-related alopecia and that 8% of the women avoid chemotherapy because they are unwilling to deal with hair loss [7-10]. Moreover, one study reported that the hair loss was harder to manage than the loss of a breast in some patients [11].

It is known that there are three cycles during hair growth: anagen is the growth phase; catagen is the involuting or regressing phase; and telogen is the resting or quiescent phase [12,13]. It is thought that chemotherapeutic agents target highly proliferative hair matrix cells in the anagen phase, called the anagen effluvium [4,14], but the molecular mechanism is still largely unknown. Scalp cooling with cold air or liquid is the most widely used method since the 1970s to prevent or minimize drug-induced alopecia. However, it is not always effective and it is not easy to standardize the system of scalp cooling [4,15]. Since medications such as minoxidil or AS101, which are widely used for aging-related hair loss, failed to show any protective effect in the case of chemotherapy-induced alopecia [16-19], there is currently no good option to prevent or treat drug-induced alopecia.

In this study, we conducted a genome-wide association study (GWAS) using mono- or combination-chemotherapy-treated breast cancer cases to identify common genetic factors that are associated with drug-induced alopecia. We have identified some loci that are likely to be associated with increased risk of chemotherapy-induced alopecia. These results can provide new insight into the molecular mechanisms of hair loss induced by anticancer drugs and may contribute to development of drugs that can prevent or treat this emotionally devastating side effect.

Methods

Participants

All samples used in this study were obtained from the BioBank Japan located at the Institute of Medical Science at the University of Tokyo. The BioBank Japan project [20], which began in 2003, is a collaborative network of 66 hospitals in Japan [21]. The project achieved a collection of genomic DNA, serum, and clinical information from a total of 330,000 cases (200,000 patients) that had at least 1 of 47 defined diseases. Adverse drug reaction (ADR) information was collected from the patients' medical records by medical coordinators. From the BioBank Japan, we selected 1,367 individuals who had been diagnosed with breast cancer and had received conventional chemotherapy. Of them, 303 patients had experienced grade 2 alopecia (ADR), 184 revealed grade 1 alopecia, and the remaining 880 patients were reported to have had no alopecia (non-ADR). Grade 2 alopecia is defined as complete hair loss, which is the most severe grade in this adverse reaction (National Cancer Institute Common Toxicity Criteria (NCI-CTC) version 3.0). In addition, samples from 23 breast cancer patients with grade 2 alopecia were collected at the Tokushima Breast Care Clinic to further verify the findings of the initial GWAS study; all of the 23 patients were treated with a combination therapy of docetaxel and cyclophosphamide. The detailed clinical information is summarized in Additional file 1. All participants provided written informed consent. This project was approved by the Institutional Review Board of the Institute of Medical Science, the University of Tokyo, and RIKEN Center for Genomic Medicine.

Genotyping and quality control

For GWAS, all DNA samples were genotyped using Illumina Human OmniExpress BeadChip kits (Illumina, San Diego, CA, USA). Sample quality control was performed by identity-by-state clustering across all samples to evaluate cryptic relatedness for each sample and by use of principal component analysis to exclude genetically heterogeneous samples from further analysis. We applied SNP quality control by excluding SNPs with a call rate of <0.99, a P value of the Hardy-Weinberg equilibrium test of $\leq 1.0 \times 10^{-6}$, and non-polymorphic SNPs in the dataset. Quantile-quantile (Q-Q) plots and lambda values, which were used for further evaluation of population substructure, were calculated between observed P value from Fisher's exact test allelic model against expected P value. For genotyping of additional samples, we used the multiplex PCR-based Invader assay (Third Wave Technologies, Madison, WI, USA) as described previously [22].

Statistical analysis

In the GWAS, Fisher's exact test was applied to three genetic models: an allele frequency model, a dominant inheritance model, and a recessive inheritance model. SNPs were rank-ordered according to the lowest P value among the three models. Odds ratio (OR) and confidence intervals (CIs) were calculated for the allelic model using a non-risk allele or a non-risk genotype as a reference. A Manhattan plot was generated by using the minimum

P value among three genetic models. For the combined analysis, the genotype count of the additional samples was added to that of the GWAS. All statistical analyses and plots were carried out using R statistical environment version 2.13.2 [23], and PLINK version 1.07 [24,25]. Haploview software was used for haplotype analysis, to draw the Manhattan plot and linkage disequilibrium (LD) map.

Scoring system using weighted genetic risk score (wGRS)

The scoring analysis was performed by utilizing SNPs with *P* min of $<1.0 \times 10^{-5}$ after exclusion of SNPs that show strong LD (r^2 >0.8) of each GWAS. wGRSs were calculated according to a method reported by De Jager *et al.* [26]. Briefly, we first determined the effect size of each SNP, calculated the cumulative genetic risk scores by multiplying the number of risk alleles for each SNP by its corresponding weight, and subsequently took the sum across the total number of SNPs that were taken into consideration of each GWAS set. We classified the genetic risk score into four different groups, which were created from the mean and standard deviation (SD) as follows: <mean −1 SD for group 1; mean −1 SD to average for group 2; average to mean +1 SD for group 3; >mean +1 SD for group 4. Odds ratio (OR), 95% confidence interval (CI), *P* value, sensitivity, and specificity were calculated using group 1 as reference.

Results

Genome-wide association for chemotherapy-induced alopecia in breast cancer

We performed a GWAS of 303 individuals who developed grade 2 alopecia, and compared them with 880 breast cancer patients who did not show any hair loss after being treated with conventional chemotherapy. The Q-Q plot and lambda (λ) value (λ <1.000) indicated no evidence of population stratification between the cases and controls we analyzed (Additional file 2). After the data was quality controlled, association analysis was carried out for 555,600 autosomal SNPs by Fisher's exact test on the basis of three genetic models: allelic-effect, dominant-inheritance, and recessive-inheritance models. Among the SNPs analyzed in the GWAS, we identified a locus that reached genome-wide significance (rs3820706 near *CACNB4*, minimum $P = 8.13 \times 10^{-9}$, ORrec = 3.71, 95% CI: 2.24 to 6.15) and five additional loci that revealed suggestive association with chemotherapy-induced alopecia with a *P* value of $<10^{-6}$ (Additional file 3 and Table 1). We further validated the top nine SNPs that revealed the smallest *P* value on the three loci in the GWAS result, using 23 additionally obtained alopecia cases. The combined analysis slightly improved the association with the rs3820706 locus (combined minimum $P = 1.85 \times 10^{-9}$, ORrec = 2.38, 95% CI: 1.44 to 3.93) and a nearby SNP rs16830728 (combined minimum $P = 2.60 \times 10^{-8}$, ORrec = 3.61, 95% CI: 2.17 to 5.98; Table 2). As these two SNPs are in strong LD with r^2 of >0.8, we performed haplotype analysis, but the association was not as strong as those of single SNPs (Additional file 4 and Additional file 5).

Association studies for drug subgroups and specific drugs

We also performed subgroup analyses for different types of chemotherapy, namely the CEF (cyclophosphamide + epirubicin +/− 5-FU)-treated and CAF (cyclophosphamide + doxorubicin +/− 5-FU)-treated groups. Detailed sample demographics are described in Additional file 1. In the GWAS of the CEF-treated group, genetic variants in the *ALOX5AP* gene on chromosome 13 were most significantly associated with chemotherapy-induced alopecia (rs3885907, minimum $P = 1.38 \times 10^{-6}$, OR = 2.66, 95% CI: 1.71 to 4.13). The GWAS analysis for the CAF-treated group identified SNP rs594206 located in an intronic region of *BCL9* on chromosome 1 to be most strongly associated (minimum $P = 5.91 \times 10^{-7}$, OR = 36.3, 95% CI: 4.58 to 287; Additional file 3 and Additional file 6). Although the *P* values for these variants did not exceed the genome-wide significance, it is notable that OR for the identified SNP for the CAF analysis is very large. In addition, we analyzed the association with antimicrotubule agents, paclitaxel monotherapy and docetaxel monotherapy because of their high incidence of alopecia, and found that rs1858231 (minimum $P = 1.95 \times 10^{-6}$, OR = 2.71, 95% CI: 1.79 to 4.12), rs11059635 (minimum $P = 2.05 \times 10^{-7}$, OR = 6.63, 95% CI: 2.95 to 14.9) and rs4262906 (minimum $P = 6.62 \times 10^{-7}$, OR = 4.36, 95% CI: 2.41 to 7.89) were most significantly associated, respectively (Additional file 6).

SNP rs3820706 on *CACNB4*, which showed the strongest association with chemotherapy-induced alopecia with the genome-wide significance in the analysis of all-combined samples, showed modest associations in all of the subgroup analyses (Additional file 7). Although the numbers of samples in these subgroup analyses were relatively limited, these data may provide fundamental information that will contribute to a better understanding of chemotherapy-induced alopecia.

Scoring system for prediction of chemotherapy-induced alopecia

We then evaluated the cumulative effects of the candidate loci (SNPs showing $P < 10^{-5}$ in Table 1 and Additional file 6) using a weighted genetic risk scoring (wGRS) method [26]. We first selected eight SNPs from the GWAS of the combination of all samples and calculated wGRS. As shown in Additional file 8, only 17 of 190 patients belonging to group 1 showed severe hair loss (grade 2) while 54 of 82 patients in group 4 revealed it. Cumulative risk scores for the risk of drug-induced alopecia were calculated to be

Table 1 Summary of association results of the genome-wide association study

				ADR[b]			Non-ADR[c]			RAF		*P* value				
CHR	SNP	Gene	Allele 1/2 (risk)	11	12	22	11	12	22	ADR	Non-ADR	Allelic	Dominant	Recessive	OR[a]	95% CI
2	rs3820706	*CACNB4*	A/G (G)	18	169	116	167	421	291	0.66	0.57	8.26E-05	1.07E-01	8.13E-09	3.71	(2.24-6.15)
2	rs6725180	*CACNB4*	A/C (C)	17	152	134	135	429	316	0.69	0.60	7.90E-05	1.11E-02	3.84E-06	3.05	(1.81-5.14)
8	rs16908658	*FAM135B*	G/A (G)	30	93	180	23	286	571	0.25	0.19	1.07E-03	9.68E-02	9.93E-07	4.09	(2.34-7.17)
10	rs7476422	*PCDH15*	T/G (G)	4	47	252	34	245	601	0.91	0.82	1.20E-07	3.77E-07	3.58E-02	2.17	(1.60-2.93)
10	rs857373	*PCDH15*	G/A (A)	5	55	243	43	255	581	0.89	0.81	5.16E-07	3.15E-06	1.11E-02	2.00	(1.51-2.66)
10	rs857392	*PCDH15*	G/A (A)	5	55	243	42	252	584	0.89	0.81	9.08E-07	5.95E-06	1.60E-02	1.97	(1.48-2.62)
10	rs1319836	*PCDH15*	C/T (T)	5	55	243	42	254	583	0.89	0.81	9.10E-07	4.34E-06	1.60E-02	1.98	(1.49-2.63)
10	rs7919725	*PCDH15*	A/G (G)	5	56	242	42	256	580	0.89	0.81	9.94E-07	4.68E-06	1.60E-02	1.97	(1.48-2.60)
10	rs857369	*PCDH15*	T/C (C)	1	32	270	18	178	684	0.94	0.88	2.29E-06	7.25E-06	5.87E-02	2.33	(1.60-3.39)
10	rs9416306	*PCDH15*	G/T (T)	1	32	270	18	178	682	0.94	0.88	2.29E-06	7.13E-06	5.88E-02	2.34	(1.61-3.39)
10	rs1219862	*PCDH15*	C/T (T)	2	31	270	17	182	681	0.94	0.88	2.73E-06	5.08E-06	1.85E-01	2.28	(1.58-3.30)
13	rs7318267	*FARP1*	C/T (T)	11	149	143	108	387	385	0.72	0.66	6.69E-03	3.15E-01	4.09E-06	3.71	(1.97-7.01)
13	rs2282048	*FARP1*	T/C (C)	11	148	144	107	387	386	0.72	0.66	5.72E-03	2.84E-01	6.24E-06	3.68	(1.95-6.93)
17	rs1530357	*LOC100506974*	A/G (A)	57	170	76	114	417	349	0.47	0.37	1.11E-05	4.29E-06	1.39E-02	1.96	(1.45-2.63)
17	rs1530361	*LOC100506974*	A/G (A)	53	165	85	99	408	372	0.45	0.35	8.83E-06	1.12E-05	7.04E-03	1.54	(1.27-1.86)
19	rs11666971	*LASS4*	G/A (G)	46	119	138	56	379	445	0.35	0.28	1.64E-03	1.43E-01	8.13E-06	2.63	(1.74-3.96)

[a]ORs and C s are ca cu ated according to the associated genetic mode ; [b]individua s who deve oped grade 2 a opecia; [c]individua s who did not deve oped any ADRs after chemotherapy. CHR, chromosome; SNP, sing e nuc eotide po ymorphism; ADR, adverse drug reaction; RAF, risk a e e frequency; OR, odds ratio; C , confidence interva .

Table 2 Summary of combined results of the genome-wide association study and additional genotyped data

						ADR[c]				Non-ADR[d]				*P* value				
SNP	CHR	Chromosome position[a]	Gene	Allele 1/2 (risk)		11	12	22	RAF	11	12	22	RAF	Allelic	Dominant	Recessive	*P* min	OR[b] (95% CI)
rs3820706	2	152957411	*CACNB4*	A/G	GWAS	18	169	116	0.66	167	421	291	0.57	8.26E-05	1.07E-01	8.13E-09	8.13E-09	3.71 (2.24-6.15)
				(G)	2nd	1	12	10	0.70	167	421	291	0.57	9.80E-02	3.70E-01	1.00E-01	9.80E-02	1.72 (0.91-3.25)
					Comb ne	19	181	126	0.66	167	421	291	0.57	3.16E-05	7.65E-02	1.85E-09	1.85E-09	2.38 (1.44-3.93)
rs16830728	2	152981335	*STAM2*	G/T	GWAS	17	163	123	0.68	153	422	304	0.59	1.11E-04	6.16E-02	7.24E-08	7.24E-08	3.54 (2.11-5.96)
				(T)	2nd	1	11	11	0.72	153	422	304	0.59	9.40E-02	1.91E-01	1.55E-01	9.40E-02	1.79 (0.94-3.43)
					Comb ne	18	174	134	0.68	153	422	304	0.59	3.49E-05	4.30E-02	2.60E-08	2.60E-08	3.61 (2.17-5.98)
rs7476422	10	56204291	*PCDH15*	T/G	GWAS	4	47	252	0.91	34	245	601	0.82	1.20E-07	3.77E-07	3.58E-02	1.20E-07	2.17 (1.60-2.93)
				(G)	2nd	0	7	16	0.85	34	245	601	0.82	8.45E-01	1.00E+00	1.00E+00	8.45E-01	1.21 (0.53-2.72)
					Comb ne	4	54	268	0.91	34	245	601	0.82	2.63E-07	1.15E-06	2.41E-02	2.63E-07	2.06 (1.54-2.75)

[a]On the basis of NCB 36 genome assemb y; [b]ORs and C s are ca cu ated according to the associated genetic mode ; [c]individua s who deve oped grade 2 a opecia; [d]individua s who did not deve oped any ADRs after chemotherapy. The same contro s were used in the GWAS and second stages ana ysis. SNP, sing e nuc eotide po ymorphism; CHR, chromosome; ADR, adverse drug reaction; RAF, risk a e e frequency; *P* min, minimum *P* va ue; OR, odds ratio; C , confidence interva .

4.44 in group 3 and 19.6 in group 4 ($P = 3.44 \times 10^{-9}$, 95% CI: 2.62 to 7.53; $P = 1.44 \times 10^{-21}$, 95% CI: 9.99 to 38.6, respectively), compared with patients in group 1.

Similarly, in the subgroup analysis, an individual belonging to group 4 with the highest risk score in each of the CEF, CAF, antimicrotubules, paclitaxel, and docetaxel analyses was estimated to have 86.2 times, 891 times, 858 times, 1,680 times, and 441 times higher risk for the drug-related alopecia than those in group 1, respectively (Additional file 8). Due to the clinical importance of antimicrotubule agents (paclitaxel and docetaxel), which cause chemotherapy-induced alopecia at nearly 80% frequency, we further investigated the wGRS scoring method using cases with grade 1 alopecia. Interestingly, the association levels and odds ratios of patients with grade 1 alopecia induced by the antimicrotubule agents were intermediate, compared with those of grade 2 alopecia (Table 3). Not only antimicrotubule agents, but other subgroups (all, CEA or CEF) also showed similar results, and the association level of grade 1 was intermediate compared with grade 2. These results further support a possible association of these variants in alopecia development (Additional file 9). As shown in Figure 1, the proportion of grade 2 alopecia increased according to the increase of the wGRS score; for example, in the case of docetaxel, only one (3.4%) of the 29 patients in group 1 revealed grade 2 alopecia, while 52 (83%) of 63 patients belonging to groups 3 and 4 developed grade 2 alopecia. These results indicate that our scoring system may be applied to predict severe chemotherapy-induced alopecia and might provide useful information for better understanding of the hair-loss mechanism, even though further verification using an additional independent set(s) of samples is warranted.

Finally, we simulated the sample number that is required to verify our scoring system. In BioBank Japan, a total of 279 patients received antimicrotubule agents (paclitaxel and/or docetaxel). Among them, 119 (43%) patients developed grade 2 alopecia, 55 (20%) developed grade 1 alopecia and 105 (37%) did not show any adverse events. Among 156 patients who received paclitaxel monotherapy, 57 (37%) developed grade 2 alopecia, 36 (23%) developed grade 1 alopecia and 63 (40%) did not develop any adverse reactions. When we assume that 100 patients who receive antimicrotubule agents (or paclitaxel monotherapy) are registered, the incidences of alopecia are estimated as shown in Table 4. If we categorize the patients by wGRS according to the data in Table 3, 100 additional patients should provide the sufficient statistical power to verify our results with P value of <0.01. Even if two individuals in each of groups 1 and 4 are not correctly predicted, the calculated P value is still 0.001 by Fisher's exact test.

Table 3 wGRS results of antimicrotubule agents, docetaxel, and paclitaxel-induced alopecia

Cat	Score	G2[a]	G1[b]	G0[c]	% G2	% G1	% G0	G2 vs. G0			G1 vs. G0		
								OR[d*]	95%CI	*P* value	OR[d*]	95%CI	*P* value
Antimicrotubule (6 SNPs)													
1	<5.56	2	7	34	0.05	0.16	0.79		Ref			Ref	
2	5.56 7.60	25	20	50	0.26	0.21	0.53	8.50	1.89 38.3	1.66E 03	1.94	0.74 1.42	2.52E 01
3	7.60 9.63	65	17	19	0.64	0.17	0.19	58.2	12.8 265	4.93E 14	4.35	1.53 12.4	6.42E 03
4	>9.63	26	6	1	0.79	0.18	0.03	442	38.0 5140	2.71E 14	29.1	3.02 282	8.39E 04
	Total	**118**	**50**	**104**									
Docetaxel (4 SNPs)													
1	<2.26	1	5	23	0.03	0.17	0.79		Ref			Ref	
2	2.26 4.70	9	6	14	0.31	0.21	0.48	14.8	1.69 130	4.39E 03	1.97	0.51 7.68	4.88E 01
3	4.70 7.15	33	5	4	0.79	0.12	0.10	190	19.9 1810	1.01E 11	5.75	1.12 29.4	4.08E 02
4	>7.15	19	2	0	0.90	0.10	0.00	611	23.5 15900	2.50E 11	21.4	0.89 511	4.83E 02
	Total	**62**	**18**	**41**									
Paclitaxel (7 SNPs)													
1	<3.24	2	4	28	0.06	0.12	0.82		Ref			Ref	
2	3.24 7.48	4	14	22	0.10	0.35	0.55	2.55	0.43 15.2	4.01E 01	4.46	1.28 1.92	2.60E 02
3	7.48 11.7	35	12	11	0.60	0.21	0.19	44.6	9.12 218	9.55E 10	7.64	2.02 28.9	2.30E 03
4	>11.7	16	5	0	0.76	0.24	0.00	376	17.0 8320	1.54E 10	69.7	3.26 1490	2.89E 04
	Total	**57**	**35**	**61**									

[a]Individuals who developed grade 2 alopecia; [b]individuals who developed grade 1 alopecia; [c]individuals who did not developed any ADRs after chemotherapy; [d]ORs and CIs are calculated using category (group) 1 as reference. *OR calculated after Haldane's correction: adding 0.5 to all the cells of a contingency table if any of the cell expectations would cause a division by zero error. Cat, category; OR, odds ratio; CI, confidence interval; Ref, reference.

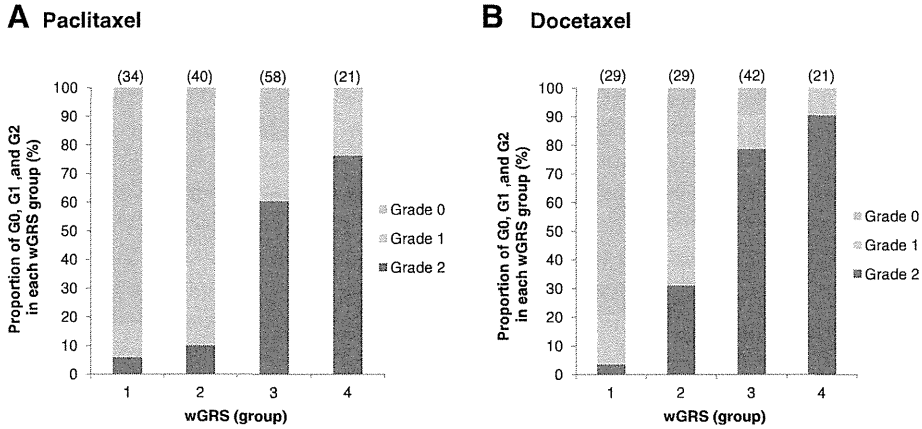

Figure 1 The proportions of patients by alopecia grade in each weighted genomic risk score. The proportions of patients who developed no adverse reaction (G0), grade 1 alopecia, or grade 2 alopecia in each of the weighted genomic risk score (wGRS) groups. The number in parentheses indicates the number of samples in each group. **(A)** paclitaxel monotherapy, **(B)** docetaxel monotherapy.

Discussion

Recent pharmacogenomics studies focus on prediction of drug response as well as the risk assessment of toxic events due to administration of drugs. Whole-genome association studies have been proven to be a powerful strategy to identify genetic factor(s) associated with various adverse reactions caused by certain drugs. In this study, we conducted the first GWAS for chemotherapy-induced alopecia in Japanese breast cancer patients, and identified one locus including two SNPs, rs3820706 on chromosome 2q23 and its nearby SNP rs16830728, which showed a strong association with genome-wide significance, and found several SNPs showing suggestive associations.

SNP rs3820706 is located near a gene encoding calcium channel voltage-dependent subunit beta 4 (*CACNB4*), a member of a beta subunit family of the voltage-dependent calcium channel (VDCC) complex. Calcium (Ca^{2+}) functions as a second messenger in many cellular signal transduction pathways such as cell proliferation and apoptosis. When VDCC is activated it depolarizes membrane potentials, it allows Ca^{2+} to enter into cells [27]. We are not aware of any previous reports indicating that there is a relationship between the Ca^{2+} channel and alopecia. However, a potassium channel opener, minoxidil, was approved for the treatment of alopecia by the US FDA in 1988 [28] and has proven to be effective in a subset of alopecia patients. Although the mode of action of minoxidil is still not well known, the clinical outcome implies the involvement of ion channels for K^+ and probably Ca^{2+} in the pathogenesis of alopecia. Intriguingly, the second most significantly associated locus that we found in our study is a region containing the *PCDH15* gene on chromosome

Table 4 Estimation of required sample number for verification

	Cat	G2	G0	OR	95% CI	*P* value*
Antimicrotubule (paclitaxel and docetaxel) (N = 100)						
	1	1	12		Ref	
	2	9	18	6.00	0.67 53.7	1.24E 01
	3	24	7	41.1	4.53 374	2.48E 05
	4	9	1	108	5.92 1970	1.15E 04
	Total	43	38			
Paclitaxel (N = 100)						
	1	1	18		Ref	
	2	3	15	3.60	0.34 38.3	3.40E 01
	3	23	7	59.1	6.66 525	8.02E 07
	4	10	0	259	9.66 6950	5.49E 07
	Total	37	40			

**P* values are calculated by Fisher's exact test. Cat, category; OR, odds ratio; CI, confidence interval; Ref, reference.

10. *PCDH15* encodes a protocadherin-related protein, which is involved in calcium-dependent cell-cell adhesion. Additionally, among the 70 loci in the top 100 SNPs found in our GWAS study, five loci are implicated to be ion channels or proteins related to ion channels (data not shown). Ion channels have shown to have important roles not only in cell maintenance but also in stem/progenitor cells [29]. Because cytotoxic agents damage the proliferating progenitor cells in the hair matrix [13], we suspect that several ion channels might be involved in chemotherapy-induced alopecia and be promising targets for development of novel treatments.

However, since rs3820706 is strongly linked to rs16830728, which is located within a gene encoding a signal transducing adaptor molecule 2 (*STAM2*), we cannot exclude the possibility that *STAM2* is a candidate gene for chemotherapy-induced alopecia. STAM2 is a member of the STAM family, which is an adaptor protein involved in the downstream signaling of cytokine receptors that contain an SH3 domain and the immunoreceptor tyrosine-based activation motif (ITAM). STAM2 is involved in the signaling through GM-SCF and IL-2 stimulation, and has a crucial role in T cell development [30,31]. As most studies of STAM2 focused on immune cells, its functions in other cell types like hair follicle cells are not fully understood.

In addition, we performed subgroup analyses in which we identified multiple loci that might be associated with drug-induced alopecia. rs3885907, which was most significantly associated in CEF-treated patients, was located in an intron of *ALOX5AP*. *ALOX5AP*, arachidonate 5-lipoxygenase-activating protein, is related to the inflammatory responses and possibly to vascular diseases [32,33]. Detailed biological mechanisms in hair growth cycle are not well characterized, but one paper reported involvement of the *ALOX5AP* upregulation in scarring alopecia [34]. According to GWAS, for alopecia areata [35] that identified genes related in both innate and adaptive immunity, inflammatory or immune responses seem to be important in alopecia development. The mechanisms of hair loss in alopecia areata and in drug-induced alopecia may not be same, but our result suggests a possible relationship of the immune response with chemotherapy-induced alopecia.

A SNP in the *BCL9* gene was most significantly associated with hair loss in the CAF-treated group with very high OR of 36.3. The *BCL9* gene encodes B-cell lymphoma 9 which was reported to interact with β-catenin. The β-catenin signaling pathway is involved in hair follicle morphogenesis during embryogenesis and, interestingly, hair is completely lost when β-catenin is depleted even after hair follicles have been formed [36,37]. Similarly, *CDH7*, one of the cadherin family members, showed an association with severe hair loss in the CAF-treated group with high OR of 32.5. This cadherin has been reported to be expressed in hair follicles and regulate hair growth [38,39]. These results, in combination with our GWAS results, imply possible roles for BCL9 and CDH7 in chemotherapy-induced alopecia. If so, these two molecules as well as CACNB4 and other ion channel proteins could be promising targets for the development of new treatments. However, further validation is still needed.

Our approach of using retrospective BioBank samples is not ideal for addressing this type of clinical problem and certainly a prospective analysis with well-defined clinical information would reduce the possibility of false-positive and false-negative results. However, considering the rapid progress of drug development or new combination therapies in recent years, it may not be wise to spend lots of effort, time and budget to do a prospective study, because the investigated regimen may not be used years later when the research results come out. One of the ways to effectively use the data and samples from the retrospective study is shown by the application of our wGRS system. The wGRS system indicated cumulative effect of multiple genetic variants for alopecia prediction. For example, the patients in group 4 who received paclitaxel showed 376 times increased risk of alopecia, compared with those belonging to group 1. Similarly, the patients in group 4 who received docetaxel showed 611 times higher risk of alopecia than those belonging to group 1. We understand the disadvantages and pitfalls of the retrospective design for the pharmacogenomics study such as the higher risk of false results. However, considering the very high OR obtained by the wGRS system, the advantage of this approach is that we are able to verify the results by using a relatively small number of additional prospective samples. We simulated the sample size needed to verify our results, as shown in Table 4, and suggest that the statistical power should be sufficient to validate with this small number of samples. We recognize that the clinical utility for this wGRS may not be as high as in other studies looking at life-threatening adverse events. However, identification of genetic factors associated with drug-induced hair loss should be the first step to understand the molecular mechanism and to contribute to the development of new drugs to prevent or treat alopecia.

For many years, breast cancer patients have had to accept the psychologically stressful side effect of alopecia caused by cytotoxic chemotherapies. It is known that a subset of patients will refuse to have chemotherapy because they do not want to lose their hair and therefore may lose the opportunity to receive the benefit of the chemotherapy and a chance to be cured of their disease. The QOL of these patients is extremely important and we believe it is urgent that we work to develop new treatment or prevention strategies to manage chemotherapy-induced alopecia. Although further validation of our findings is

required, our study identified some significant molecular alterations in genes such as ion channel-related genes and genes in the β-catenin signaling pathway. We welcome other groups to examine and validate our results and hope these findings will contribute to the development of interventions that will improve the quality of life (QOL) of breast cancer patients.

Conclusions

In summary, we identified strongly associated genetic variants near gene *CACNB4* and several suggestively associated SNPs with chemotherapy-induced alopecia in breast cancer patients. These results provide new information of the pathogenesis of chemotherapy-induced alopecia.

Additional files

Additional file 1: Table S1. Patients' characteristics.

Additional file 2: Quantile quantile plot of the genome wide association study.

Additional file 3: Manhattan plot of the genome wide association study for chemotherapy induced alopecia in breast cancer.

Additional file 4: Haplotype analysis.

Additional file 5: Table S2. Haplotype analysis of two SNPs.

Additional file 6: Table S3. Summary of genome wide association study for chemotherapy induced alopecia with each drug subgroup ($P < 10^{-6}$).

Additional file 7: Table S4. Association of rs3820706 in subgroups.

Additional file 8: Table S5. Weighted genomic risk score of each genome wide association study for chemotherapy induced alopecia.

Additional file 9: Table S6. Weighted genomic risk score results of all, CAF and CEF induced alopecia.

Abbreviations

ADR: Adverse drug reaction; CAF: Cyclophosphamide + doxorubicin +/− 5 FU; CEF: Cyclophosphamide + epirubicin +/− 5 ; CI: Confidence interval; GWAS: Genome wide association study; LD: Linkage disequilibrium; OR: Odds ratio; QOL: Quality of life; QQ plot: Quantile quantile plot; SD: Standard deviation; SNP: Single nucleotide polymorphism; wGRS: Weighted genomic risk score.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YN planned and supervised the study and obtained funding. SC and SKL designed the experiments and performed the GWAS and combined analysis. SKL performed the wGRS and statistical analysis. HZ and MS collected additional samples and medical information. MK genotyped all BioBank Japan samples. AT performed sample quality control. YN, SC and SKL wrote the manuscript. All authors revised and approved the manuscript for publication.

Acknowledgements

We express our heartfelt gratitude to all the participants. We thank Miss Kumi Matsuda for her outstanding technical assistance. We also thank all other members and staff for their contribution to the sample collection. This study was supported by a leading project of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author details

[1]Department of Medicine, The University of Chicago, 5801 South Ellis Avenue, Chicago, IL 60637, USA. [2]Department of Surgery, The University of Chicago, A27 S Maryland Avenue, Chicago, IL 60637, USA. [3]Laboratory for Statistical Analysis, Center for Genomic Medicine, 1 7 22 Suehiro cho, Tsurumi ku, Yokohama, Kanagawa 230 0045, Japan. [4]Laboratory for Genotyping Development, Center of Genomic Medicine, 1 7 22 Suehiro cho, Tsurumi ku, Yokohama, Kanagawa 230 0045, Japan. [5]Tokushima Breast Care Clinic, 4 7 7, Nakashimada cho, Tokushima 770 0052, Japan. [6]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4 6 1 Shirokanedai, Minato ku, Tokyo 108 8639, Japan.

Received: 1 May 2013 Accepted: 18 July 2013
Published: 11 September 2013

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ: **Cancer statistics, 2008.** *CA Cancer J Clin* 2008, **58**:71 96.
2. Sotiriou C, Pusztai L: **Gene expression signatures in breast cancer.** *N Engl J Med* 2009, **360**:790 800.
3. Bosch A, Eroles P, Zaragoza R, Vina JR, Lluch A: **Triple negative breast cancer: molecular features, pathogenesis, treatment and current lines of research.** *Cancer Treat Rev* 2010, **36**:206 215.
4. Trueb RM: **Chemotherapy induced alopecia.** *Semin Cutan Med Surg* 2009, **28**:11 14.
5. Carrick S, Parker S, Thornton CE, Ghersi D, Simes J, Wilcken N: **Single agent versus combination chemotherapy for metastatic breast cancer.** *Cochrane Database Syst Rev* 2009, **2**:CD003372.
6. Batchelor D: **Hair and cancer chemotherapy: consequences and nursing care a literature study.** *Eur J Cancer Care (Engl)* 2001, **10**:147 163.
7. McGarvey EL, Baum LD, Pinkerton RC, Rogers LM: **Psychological sequelae and alopecia among women with cancer.** *Cancer Pract* 2001, **9**:283 289.
8. Tierney AJ, Taylor J, Closs SJ: **Knowledge, expectations and experiences of patients receiving chemotherapy for breast cancer.** *Scand J Caring Sci* 1992, **6**:75 80.
9. Munstedt K, Manthey N, Sachsse S, Vahrson H: **Changes in self concept and body image during alopecia induced cancer chemotherapy.** *Support Care Cancer* 1997, **5**:139 143.
10. de Boer Dennert M, de Wit R, Schmitz PI, Djontono J, v Beurden V, Stoter G, Verweij J: **Patient perceptions of the side effects of chemotherapy: the influence of 5HT3 antagonists.** *Br J Cancer* 1997, **76**:1055 1061.
11. Freedman TG: **Social and cultural dimensions of hair loss in women treated for breast cancer.** *Cancer Nurs* 1994, **17**:334 341.
12. Stenn KS, Paus R: **Controls of hair follicle cycling.** *Physiol Rev* 2001, **81**:449 494.
13. Paus R, Haslam IS, Sharov AA, Botchkarev VA: **Pathobiology of chemotherapy induced hair loss.** *Lancet Oncol* 2013, **14**:e50 e59.
14. Chon SY, Champion RW, Geddes ER, Rashid RM: **Chemotherapy induced alopecia.** *J Am Acad Dermatol* 2012, **67**:e37 e47.
15. Grevelman EG, Breed WP: **Prevention of chemotherapy induced hair loss by scalp cooling.** *Ann Oncol* 2005, **16**:352 358.
16. Duvic M, Lemak NA, Valero V, Hymes SR, Farmer KL, Hortobagyi GN, Trancik RJ, Bandstra BA, Compton LD: **A randomized trial of minoxidil in chemotherapy induced alopecia.** *J Am Acad Dermatol* 1996, **35**:74 78.
17. Granai CO, Frederickson H, Gajewski W, Goodman A, Goldstein A, Baden H: **The use of minoxidil to attempt to prevent alopecia during chemotherapy for gynecologic malignancies.** *Eur J Gynaecol Oncol* 1991, **12**:129 132.
18. Rodriguez R, Machiavelli M, Leone B, Romero A, Cuevas MA, Langhi M, Romero Acuna L, Romero Acuna J, Amato S, Barbieri M, *et al*: **Minoxidil (Mx) as a prophylaxis of doxorubicin induced alopecia.** *Ann Oncol* 1994, **5**:769 770.
19. Wang J, Lu Z, Au JL: **Protection against chemotherapy induced alopecia.** *Pharm Res* 2006, **23**:2505 2514.
20. http://www.biobankjp.org.
21. Nakamura Y: **The BioBank Japan Project.** *Clin Adv Hematol Oncol* 2007, **5**:696 697.
22. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y: **A high throughput SNP typing system for genome wide association studies.** *J Hum Genet* 2001, **46**:471 477.
23. http://www.cran.r project.org/.
24. http://pngu.mgh.harvard.edu/purcell/plink/.