Supplementary Materials and Methods

Specimen collection. This study was conducted with Institutional Review Board approval and is registered on ClinicalTrials.gov (NCT00858585). Individuals over the age of 21 with a histologically confirmed diagnosis of breast cancer were eligible to donate specimens of whole blood, formalin fixed paraffin embedded (FFPE) lymph node tissue that is unaffected by tumor and FFPE tumor tissue. Study exclusion criteria included those with insufficient specimen availability, Stage IV disease unless there was documentation of less than 5 circulating tumor cells in 7 ml of whole blood within 4 weeks of study phlebotomy and there was no evidence of clinical progression at the time of study participation. Peripheral blood was collected from 141 patients after providing written informed consent. Of these, 122 had available FFPE unaffected lymph node and tumor specimens. Limited clinical data were extracted from the medical record and included age at time of diagnosis, histologic features of the tumor including grade, hormone receptor status and HER2 status and systemic and local antineoplastic therapies administered prior to the donation of specimens. The specimens and clinical data were assigned corresponding codes and the patient identifiers were removed prior to research investigations. Genetic analysis was performed on de-identified, coded specimens.

DNA Extraction. DNA was extracted from whole blood cells (WBC) using the Qiagen DNeasy® Blood and Tissue Kit (kit 69506, Valencia, California, USA) and from formalin-fixed paraffin-embedded (FFPE) tumor and unaffected lymph node as described previously. Areas of malignancy within the FFPE tumor specimens were identified by H&E staining of an adjacent section and three 0.6 mm diameter cores were obtained. Unaffected FFPE lymph node specimens were identified for each patient with available WBC and FFPE tumor specimens and three 0.6 mm diameter cores were also obtained from these tissues.

CYP2D6 Genotype Determinations. We used Taqman Allelic Discrimination assays (Applied Biosystems, Inc.) to genotype DNA extracted from the 3 separate tissue types. Six single nucleotide polymorphisms (SNPs) in the *CYP2D6* gene were assayed including: rs3892097 (assay id: C__27102431_DO), rs1135840 (assay id: C__27102414__10), rs1065852 (C__11484460_40), rs28371725 (C__34816116__20), rs35742686 (C__32407232_50) and rs5030655 (C__32407243_20). All assays were run in single 25ul reactions on a CFX96 Real-time detection system (Bio-Rad) using optimized cycling conditions. Germline DNA was genotyped using the following PCR conditions: 95ºC for 10 minutes followed by 40 cycles at 92ºC for 15s and 60ºC for 1 minute. The FFPE DNA was genotyped using the following PCR conditions: 95ºC for 10 minutes followed by 60 cycles at 92ºC for 15s and 60ºC for 1 minute (for rs3892097 we used the optimal extension temperature of 57ºC). For genotype quality control, 10% of samples were randomly selected and retested for *CYP2D6* genotype by an independent blinded observer with all retested providing 100% concordant genotypes. All genotyping reactions were prepared in a designated template-free zone in a vertical laminar flow hood (AirClean 600, AirClean Systems, Raleigh, NC), with HEPA filtration.

Copy Number Assays. To determine germline *CYP2D6* gene copy number variants (CNV) we used the assay from Applied Biosystems Inc., (Foster City, California, USA, assay id: Hs04502391_cn (Int6)). The assay was multiplexed with an internal control- RNase P Taqman Copy number Reference (part: 4403326) and performed using the CFX96 Real-time detection system (Bio-Rad) following the manufacturer's provided protocol as described previously.(16) Briefly, 20 ng of DNA was assayed in quadruplicate 20 ul reactions using the following PCR cycling conditions: 95ºC for 10 minutes, followed by 40 cycles of 95ºC for 15 s and 60ºC for 1 minute. Relative *CYP2D6* CNV was calculated using the $\Delta\Delta C_T$ method relative to a known sample with 2 copies of *CYP2D6* and a known heterozygous CYP2D6*5 was used as a control

for gene deletions. For CYP2D6 CNV quality control, 10% of samples were randomly selected and retested with all but 1 sample demonstrating similar CNV status.

Statistical considerations. Proportions with exact 95% confidence intervals, and simple κ coefficients with asymptotic 95% CI are reported for concordance and agreement. Exact tests for deviation from Hardy-Weinberg equilibrium are reported. The study protocol aimed to enroll up to 150 patients, and under the assumption of variant SNP prevalence of 69% (for a SNP of interest unrelated to CYP2D6), then 90 to 126 patients having paired samples would provide 80% power to detect a true κ of 0.97 versus null κ of 0.80 to 0.84, for a one-sided hypothesis test with α=0.05, For this study of CYP2D6, a target sample size of 120 paired samples provided 90% power to detect a true κ of 0.94 versus κ of 0.80 for a two-sided hypothesis test with α=0.05 and expected phenotype frequencies of 0.08, 0.42, and 0.50 for PMs, IMs and EMs, respectively; and provided a 95% CI within ±0.039 for a concordance proportion of 0.95.