Estrogen Exposure, Metabolism, and Enzyme Variants in a Model for Breast Cancer Risk Prediction

Fritz F. Parl¹, Kathleen M. Egan², Chun Li³ and Philip S. Crooke⁴

¹Departments of Pathology, ³Biostatistics, ⁴Mathematics and Integrative Cancer Biology Center, Vanderbilt University, Nashville, TN 37232. ²Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa FL 33612.

Abstract: Estrogen is a well-known risk factor for breast cancer. Current models of breast cancer risk prediction are based on cumulative estrogen exposure but do not directly reflect mammary estrogen metabolism or address genetic variability between women in exposure to carcinogenic estrogen metabolites. We are proposing a mathematical model that forecasts breast cancer risk for a woman based on three factors: (1) estimated estrogen exposure, (2) kinetic analysis of the oxidative estrogen metabolism pathway in the breast, and (3) enzyme genotypes responsible for inherited differences in the production of carcinogenic metabolites. The model incorporates the main components of mammary estrogen metabolism, i.e. the conversion of 17β-estradiol (E₂) by the phase I and II enzymes cytochrome P450 (CYP) 1A1 and 1B1, catechol-O-methyltransferase (COMT), and glutathione S-transferase P1 (GSTP1) into reactive metabolites, including catechol estrogens and estrogen quinones, such as E₂-3,4-Q which can damage DNA. Each of the four genes is genotyped and the SNP data used to derive the haplotype configuration for each subject. The model then utilizes the kinetic and genotypic data to calculate the amount of E₂-3,4-Q carcinogen as ultimate risk factor for each woman. The proposed model extends existing models by combining the traditional “phenotypic” measures of estrogen exposure with genotypic data associated with the metabolic fate of E₂ as determined by critical phase I and II enzymes. Instead of providing a general risk estimate our model would predict the risk for each individual woman based on her age, reproductive experiences as well as her genotypic profile.

Keywords: breast cancer, risk assessment, hormonal carcinogenesis, computational biology, model

Introduction

Estrogens have long been recognized as the primary risk factor for the development of breast cancer.¹,² Epidemiologic studies have indicated that breast cancer risk is higher in women with early menarche and late menopause, who have longer exposure to estrogens.³ A pooled analysis of nine prospective studies found that circulating estrogen levels were directly related to risk of breast cancer in postmenopausal women.⁴ Based on these data, current models of breast cancer risk prediction are mainly based on cumulative estrogen exposure and include such factors as age, age at menarche, and age at first live birth;⁵,⁶ (www.cancer.gov/bcrisktool). While all these studies implicate estrogens as risk factor for the development of breast cancer, they leave open two important questions that need to be answered to advance from an empirical, global risk assessment to a truly etiological, individualized assessment. The questions are: (1) How do estrogens cause breast cancer? and (2) Since all women are exposed to estrogens, how do we better delineate risk? To close these gaps in our knowledge we need to explain mechanisms of estrogen carcinogenesis and inter-individual risk variation and our approach is to examine the dynamics of a pathway for estrogen metabolism and use its prediction of the level of DNA corrupting compounds as a predictor of breast cancer risk.

Carcinogenesis is usually viewed as a stepwise process beginning with genotoxic effects (initiation) followed by enhanced cell proliferation (promotion). The main estrogen, 17β-estradiol (E₂), is a substrate for the phase I enzymes, cytochrome P450 (CYP) 1A1 and 1B1 and a ligand for the estrogen receptor. In its dual role of substrate and ligand, E₂ has been implicated in the development of breast cancer by simultaneously causing DNA damage via its oxidation products, the 2-OH and 4-OH catechol estrogens, and by stimulating cell proliferation and gene expression via the estrogen receptor. Thus, E₂ and its oxidative metabolites are unique carcinogens that affect both tumor initiation and promotion.⁷,⁸

As shown in Figure 1, E₂ is oxidized to catechol estrogens by CYP1A1 and CYP1B1. These enzymes further oxidize the catechol estrogens to semiquinones and quinones. The highly reactive estrogen

Correspondence: Philip Crooke, Department of Mathematics, Vanderbilt University, Nashville, TN 37240.
E mail: philip.s.crooke@vanderbilt.edu

Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: http://creativecommons.org/licenses/by/3.0/.
quinones form Michael addition products with deoxyribonucleosides. Thus, estrogen quinones share a common feature of many chemical carcinogens, i.e. the ability to covalently modify DNA. Furthermore, estrogen semiquinones and quinones undergo redox-cycling, which results in the production of reactive oxygen species that can cause oxidative DNA damage.

Support for the carcinogenic activity of estrogens and their oxidative products, the catechol estrogens, comes from experiments in animal models. Treatment with either E2 or the 2-OH or 4-OH catechol estrogens caused kidney cancer in male Syrian hamsters and endometrial cancer in female CD1 mice, the latter compounds being the most carcinogenic agents. However, there is no animal model for estrogen-induced breast cancer and even in the hamster and mouse models the precise mechanism of DNA damage is uncertain. Thus, there is a need to understand estrogen metabolism in the human breast in order to elucidate the role of endogenous and exogenous estrogens in mammary carcinogenesis. To advance this understanding requires not only characterization of the various estrogen metabolites but equally important, a precise definition of the responsible enzymes. Several investigators have proposed a qualitative model of mammary estrogen metabolism regulated by oxidizing phase I and conjugating phase II enzymes. The oxidative estrogen metabolism pathway starts with E2 and E1, which are oxidized to the 2-OH and 4-OH catechol estrogens by the phase I enzymes CYP1A1 and CYP1B1. As described above, the P450-mediated estrogen metabolism is expected to lead to the formation of both estrogen and oxidative DNA adducts, all of which have been shown to possess mutagenic potential. It is postulated that the genotoxicity of the oxidative estrogen metabolism pathway is mitigated by alternate reactions of the metabolites with phase II enzymes. Specifically, catechol-O-methyl transferase (COMT) catalyzes the methylation of catechol estrogens to methoxy estrogens, which lowers the catechol estrogens available for conversion to estrogen quinones. In turn, the estrogen quinones undergo conjugation with glutathione (GSH) via the catalytic action of glutathione S-transferase GSTP1. The formation of GSH-estrogen conjugates would reduce the level of estrogen quinones and thereby lower the potential for DNA damage.

The current models of mammary estrogen metabolism have limitations. Firstly, only single enzymes, e.g. CYP1B1 and COMT, have been analyzed to date with simple substrate-product kinetics, which clearly generates an incomplete picture of the metabolic pathway. Secondly, while the model incorporates the functional roles of the phase I and II enzymes, it does so only qualitatively and it remains uncertain how the enzymes interact quantitatively. Third, each of the phase I and II enzymes contains genetic polymorphisms. Studies from several laboratories have examined the functional implications of the polymorphisms on estrogen metabolism, again focusing on single enzymes. However, the multitude of potential kinetic reactions resulting from the complex genetic variations of the phase I and II enzymes is completely outside the scope of the current model of estrogen metabolism. In contrast to the relatively small number of functional studies of estrogen metabolism, multiple epidemiological studies have investigated breast cancer risk in relation to genetic variation in the critical enzymes involved in estrogen metabolism with inconsistent findings. A drawback of any purely genetic assessment is the lack of information about functional interactions inherent in complex metabolic pathways such as the estrogen metabolism pathway. Thus, such studies cannot assess the underlying metabolic interactions in the pathway. A pathway-based functional and quantitative approach is necessary to overcome the current limitation in genotype assessment.

We have developed an experimental in vitro model of mammary estrogen metabolism, in which we combined purified, recombinant phase I enzymes CYP1A1 and CYP1B1 with the phase II enzymes COMT and GSTP1 to determine how E2 is metabolized. We employed both gas and liquid chromatography with mass spectrometry (GC/MS and LC/MS) to measure the parent hormone E2 as well as eight metabolites, i.e. the catechol estrogens, methoxyestrogens, and estrogen-GSH conjugates. With this important experimental data, an in silico model of the metabolic pathway has been developed.

Methods
A mathematical model for the estrogen metabolism pathway that is shown in Figure 1 can be constructed using some basic assumptions.
Figure 1. Oxidative estrogen metabolism causes DNA adduct formation. The estrogen metabolism pathway is regulated by oxidizing phase I and conjugating phase II enzymes. CYP1A1 and CYP1B1 catalyze the oxidation of E2 to catechol estrogens 2-OHE2 and 4-OHE2. The catechol estrogens are either methylated by COMT to methoxyestrogens (2-MeOE2, 2-OH-3-MeOE2, 4-MeOE2) or further oxidized by CYPs to semiquinones (E2-2,3-SQ, E2-3,4-SQ) and quinones (E2-2,3-Q, E2-3,4-Q). The estrogen quinones are conjugated by GSTP1 to GSH-conjugates (2-OHE2-1-SG, 2-OHE2-4-SG, 4-OHE2-2-SG). Alternatively, the quinones can form quinone-DNA adducts (e.g. 4-OHE2-N7-guanine, 2-OH-E2-N2-deoxyguanosine) or cause oxidative adducts (e.g. 8-OH-deoxyguanosine) via reactive oxygen species resulting from redox-cycling between semiquinones and quinones. The three adducts and their estrene (E1) and adenine counterparts have been identified in human breast tissues.\textsuperscript{56,57} Recently, we demonstrated experimentally that CYP1B1-mediated oxidation of E2 in the presence of deoxyguanosine caused the formation of the 4-OHE2-N7-guanine adduct.\textsuperscript{48} Our results provide direct evidence that metabolism of the parent hormone can initiate DNA damage.
about the kinetics of the reactions in this figure. We assume that each reaction in the pathway \((A \rightarrow B, \text{a generic step in the pathway})\) is an enzyme-catalyzed reaction of the form: 

\[
A + E \stackrel{k_1}{\longrightarrow} C \stackrel{k_2}{\longrightarrow} B + E
\]

where \(E\) denotes the enzyme, \(C\) is the enzyme-substrate complex, and \(k_i, i = 1, 2, 3\), are the rate constants of the reaction. For these types of reaction we approximate the kinetics using the quasi-steady state assumption: 

\[
CE \stackrel{K_m}{\Longleftrightarrow} A \stackrel{k_1}{\rightarrow} k_2 2 3
\]

where \(E^*\) denotes the initial enzyme concentration. We can “prove” this approximation by looking at the differential equations for \(A + E \stackrel{k_1}{\longrightarrow} C \stackrel{k_2}{\longrightarrow} B + E\):

\[
dA\frac{dt}{dt} = -k_1AE + k_2C, A(0) = A_0
\]
\[
 dB\frac{dt}{dt} = k_1C, B(0) = 0
\]
\[
 dC\frac{dt}{dt} = k_1AE - (k_2 + k_3)C, C(0) = 0
\]
\[
 dE\frac{dt}{dt} = -k_1AE + (k_2 + k_3)C, E(0) = E_0
\]

This system of differential equations yields two conservation laws:

\[
A(t) + C(t) + B(t) = A_0
\]
\[
C(t) + E(t) = E_0
\]

With the conservation laws, we can reduce the four differential equations to two differential equations:

\[
 dA\frac{dt}{dt} = -k_1E_0S + (k_1A + k_2)C, A(0) = A_0
\]

\[
\frac{dC}{dt} = k_1E_0A - (k_1A + k_2 + k_3)C, C(0) = 0
\]

\[
\Rightarrow \frac{dC}{dt} = k_1E_0A - (k_1A + k_2 + k_3)C
\]

\[
\Rightarrow \frac{dA}{dt} = -k_1E_0S + (k_1A + k_2)C
\]

We now assume that the reaction has progressed to the state that

\[
\frac{dC}{dt} \approx 0 \Rightarrow \frac{dC}{dt} = k_1E_0A - (k_1A + k_2 + k_3)C
\]

\[
\Rightarrow C(A) = \frac{k_1E_0A}{k_1A + k_2 + k_3} = \frac{E_0A}{k_1} = \frac{E_0A}{K_m + A}, K_m = \frac{k_2 + k_3}{k_1}
\]

Hence, the formation of \(B(t)\) is approximately given by \(\frac{dB}{dt} = k_3(\frac{E_0A}{K_m + A}) = k_{cat}E^*A/(K_m + A)\) where \(k_{cat} = k_3\) and \(E^* = E_0\) is the initial enzyme level.

More information about the quasi-steady state approximation can be found in Parl et al.\(^43\) Using this approach for the individual reactions in Figure 1, we can write down to a system of nonlinear, ordinary differential equations for the concentrations of the compounds in the pathway:

\[
\frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]

\[
\Rightarrow \frac{d(E_2)}{dt} = \frac{k_{cat}E_{CYP1B1}E_2}{K_m + E_2} - \frac{k_{cat}E_{CYP1A1}E_2}{K_m + E_2}
\]
Here $k_{catj}$ and $K_{mj}$ are constants and $E_{enzyme}$ are the enzyme levels in the reactions. There are parts of the pathway for which kinetic data is not available. In particular, rate constants are not known for the reactions: $2-OHE_2 \rightarrow E_2-2,3-SQ \rightarrow E_2-2,3-Q$ and $4-OHE_2 \rightarrow E_2-3,4-SQ \rightarrow E_2-3,4-Q$, respectively. Our first simplification is to collapse these reaction to single reactions, $2-OHE_2 \rightarrow E_2-2,3-Q$ and $4-OHE_2 \rightarrow E_2-3,4-Q$, respectively. The next simplification is to assume that each of these quinone production reactions $(OHE_k \rightarrow EQ_l^j)$ satisfy dynamics of the form: $\frac{dEQ_l^j}{dt} = V_{maxQ} (OHE_k^{\alpha})^\beta / K_{Q_j} + (OHE_k^{\sigma})^\sigma$ where $V_{maxQ}$, $K_{Q_j}$ and $\sigma$ are constants. For the mathematical
model to be a tractable computational model of the metabolism pathway, it is necessary to have estimates of these unknown constants. We next look at a technique for estimating $\alpha, \beta, V_{\text{max},Q1}$, $V_{\text{max},Q2}$, $K_{mQ1}$, and $K_{mQ2}$.

Experimental values for rate constants of the CYP1A1, CYP1B1, COMT, and GSTP1 catalyzed reactions are available.$^{26,29,32,41,42}$ Furthermore, the concentrations over time for each non-quinone compound in the complete pathway have been measured for a particular starting concentration of $E_2$ i.e. $E_2(0) = E_{2r}$. Using this data, a searching algorithm was written in Mathematica (Wolfram Research, Inc.) to find values for $V_{\text{max},Q}$, $K_{mQ}$ and $\sigma$ in each of the two quinones reactions that fit the experimental data in a certain metric using numerical solutions of the differential equation system. The constants, $\alpha, \beta, V_{\text{max},Q1}$, $V_{\text{max},Q2}$, $K_{mQ1}$, and $K_{mQ2}$, were obtained in this manner.

**Results**

Figure 2 shows comparisons between the model (solid curves) and the data$^{41}$ over a simulation covering 30 minutes for some of the components in the pathway. These simulations used the estimates for $\alpha, \beta, V_{\text{max},Q1}, V_{\text{max},Q2}, K_{mQ1}$, and $K_{mQ2}$ calculated above as well as the published values of the other kinetic parameters. In the simulations of the pathway it was assumed that initially all quantities are zero, except for $E_2(0) = E_{2r}$. Enzyme concentrations used in the simulation are consistent with those reported previously.$^{41}$

Having all of the parameters of the system, one can view the model as giving functional relations between $E_2(t)$ and the estrogen quinone concentrations: $EQ_{23}^k(t)$ and $EQ_{24}^k(t)$. Figure 3 shows the time-wise buildup and decay of the estrogen quinones. In an attempt to give a simple measure of the quinone concentrations over the course of time, we introduce the Area Under the Curve (AUC) metric: $AUC_k = \int_0^T EQ_k^2(t)dt$ where $k = 23, 24$ and $T = 30$ min. It is possible to introduce other measures e.g. $EQ_{2max}^k = \max_{0\leq t\leq T} EQ_k^2(t)$ which is the highest concentration achieved during the time interval $[0, T]$. We have chosen the AUC metric because it incorporates both concentration level and time.

The mathematical model for the estrogen metabolism pathway provides a relationship between an input $E_2$ and two outputs $AUC_{23}$ and $AUC_{24}$. It can also be view as connecting the area under the curve outcomes to the kinetic parameters, $k_{\text{cat}}$ and $K_m$, embedded in the model. The model permits one to analyze the behavior of the area under curve variables as functions of the kinetic parameters, either for a single step in the pathway or a combination of steps. This analysis allows one to view how variations in the kinetic parameters, which are the result of polymorphism of the enzymes, affect the area under the curve outcomes.

Each of the phase I and II enzymes involved in estrogen metabolism possesses genetic variants that (a) are associated with altered enzyme function and (b) occur in a sizable portion of the population.$^{38,44}$ We and others have determined the enzymatic rate constants ($k_{\text{cat}}$ and $K_m$) of the common CYP1A1,
CYP1B1, and COMT variants and compared their activity to the respective wild-type enzymes. These studies were limited to individual enzyme reactions and did not take the entire estrogen metabolism pathway into account. To obtain a more realistic and inclusive view of estrogen metabolism in the female population, we utilized the model to simulate how variations in the kinetic parameters resulting from polymorphisms of the enzymes impact the metabolite concentrations. We examined 4 CYP1A1, 16 CYP1B1, and 2 COMT alleles. Thus, our simulations are based on the examination of 4\(^1\)\(6\)\(^2\)\(16\) = 128 genetic combinations to demonstrate the utility of the model. Although each of the metabolites can be modeled, we concentrated our analysis on the catechols and quinones because of their documented carcinogenic activity.

Since women may differ in their combination of enzyme variants, they will have different rate constants, resulting in differences of 4-OHE\(_2\) and E\(_2\)-3,4-Q production. As shown in Figure 3 modeling of the 128 haplotype combinations produced a spectrum of catechol and quinone concentrations over time, as expressed by a range of AUC values. The simulations identified the haplotype combinations producing the highest and lowest AUCs. For example, the maximum AUCs for 4-OHE\(_2\) and E\(_2\)-3,4-Q were produced by the haplootypes CYP1A1\(^{461}\)Asn-462Ile, CYP1B1\(^{484}\)Arg-119Ser-432Val-453Asn, COMT\(^{108}\)Met, which were 2.6- and 4.6-fold higher, respectively, than the minimum AUCs produced by haplotype CYP1A1\(^{461}\)Thr-462Val, CYP1B1\(^{148}\)Gly-119Ala-432Val-453Ser, COMT\(^{108}\)Val. While 2.6 to 4.6-fold differences may not appear large, it is important to consider that they impact on lifetime exposure, which is consistent with the hormonal risk model presented by Pike.

If a subject’s haplotypes can be resolved for all genes (i.e. she has at most one heterozygous SNP for each gene), then the in silico model can be used directly to derive the E\(_2\)-3,4-Q production, as depicted in Figure 4. When a subject’s haplotype configurations are uncertain for some genes because of the presence of two or more heterozygous SNPs (e.g. CYP1B1), we first calculate the distribution of all haplotype configurations using PHASE\(^{47}\) (stephenslab.uchicago.edu/software.html). Then we derive the E\(_2\)-3,4-Q production value for each haplotype configuration, and calculate the weighted average of all E\(_2\)-3,4-Q production values, using the probabilities of haplotype configurations as weights. It can be shown that this weighted average is the expected E\(_2\)-3,4-Q production given the genotypes. This way, we incorporate information from all genotyped SNPs and each haplotype configuration is apportioned appropriately. Application of the model to a breast cancer case-control population (438 pre- and postmenopausal women with 221 invasive breast cancer cases and 217 controls) defined the estrogen quinone E\(_2\)-3,4-Q as a potential breast cancer risk factor. This exploratory analysis identified a subset of women at increased breast cancer risk based on their enzyme haplotype and consequent E\(_2\)-3,4-Q production.\(^{42}\) Based on the E\(_2\)-3,4-Q AUC values, cases predominated in the top tier of the population. For example, among the 10 women with the highest E\(_2\)-3,4-Q values in the entire study population, there were nine cases and one control (p-value = 0.01). These results suggest for the first time the possibility that breast cancer risk prediction may be enhanced by incorporation of inherited differences in estrogen metabolism.

Obviously, the model requires testing and as a first step we have examined the contribution of the

![Figure 3. Kinetic-genomic modeling of catechol estrogen. (A) 4-OHE\(_2\) and estrogen quinone (B) E\(_2\)-3,4-Q using rate constants for wild-type and variant CYP1A1, CYP1B1, and COMT. The Area Under the Curve = AUC represents the metabolite production over time. Only the highest, lowest, and wild-type (dotted line) AUCs are shown.](image-url)
estrogen concentration on \(E_2-3,4-Q\) production and the associated breast cancer risk. Numerous epidemiological studies have implicated estrogens in the development of breast cancer.\(^3\) For example, a pooled analysis of nine prospective studies of serum estrogen levels and breast cancer in 2428 postmenopausal women revealed a strong association of serum \(E_2\) concentrations with breast cancer risk.\(^4\) The relative risk of breast cancer for women whose free \(E_2\) levels were in the top quintile was 2.58 compared with 1.00 for those women whose levels were in the bottom quintile. Since the nine studies employed different methods to measure \(E_2\), there were considerable differences in the median \(E_2\) values reported. In spite of this variability, the median serum \(E_2\) concentrations in seven of the nine studies were higher in the case patients than in the control subjects. To incorporate the different levels into our simulations, we introduced a ratio that is defined as \(EE_{\text{case}}^{E_2}/EE_{\text{control}}^{E_2}\). Ratios varied, the fraction of cases in the top \(E_2-3,4-Q\) AUC values of the women also changed. With an \(E_2\) Ratio of 0.908 (Rancho Bernardo), only 31 women in the top AUC cases and controls had breast cancer compared to 65 women (Washington Country) at 1.06 (\(p = 0.037\)) and 102 (SOF) at 1.34 (\(p < 0.00001\)).

The results of these simulations demonstrate that relatively small changes in the concentration of the parent hormone \(E_2\) result in markedly increased production of the carcinogenic estrogen quinone metabolite, \(E_2-3,4-Q\), which, in turn, is reflected in a higher fraction of women with breast cancer in the top tier of our study population. Thus, testing of our model with estrogen concentrations reported in the literature confirms the striking influence of serum \(E_2\) concentrations on breast cancer risk. Importantly, the model offers a risk assessment of individual women by combining the hormone level with the genotype.

**Discussion**

A strength of the *in silico* model is that it can incorporate each woman’s actual lifetime endogenous and exogenous estrogen exposures, in
addition to her genotype, when predicting cumulative $E_2$-3,4-$Q$ exposure. This is schematically shown in Figure 5, which displays the interaction of estrogens, enzyme genotypes, and resulting $E_2$-3,4-$Q$ production as a three-dimensional graph. The graph is built on the two-dimensional Figure 3, in which we used a fixed $E_2$ level to model the $E_2$-3,4-$Q$ AUC for wild-type and variant enzyme genotypes and displayed only the lowest, highest, and wild-type $E_2$-3,4-$Q$ AUCs. In the three-dimensional graph, we plot the available genotypes, from lowest to highest, separated into quintiles based on their respective $E_2$-3,4-$Q$ production. A new component in the three-dimensional graph is the variation in $E_2$ concentration. As illustrated in the overall pathway in Figure 1 and in experimental studies, the input concentration of the parent hormone $E_2$ determines the output concentration of the oxidative metabolites, such as 4-OHE$_2$ and $E_2$-3,4-$Q$. Thus, in the graph we display estrogen exposure in quintiles. Estrogen exposure can be represented by actual $E_2$ values, measured in pmol/L, in combination with semiquantitative estimates each woman’s overall exposure to estrogen. The latter is derived by taking into account her total years of ovulation as a function of current age, age at menarche, age at menopause, numbers of full-term pregnancies and lactation experience for each, and the dosage and duration of the use of exogenous estrogens. With regard to exogenous estrogens, all estrogens including equine estrogens used in hormone replacement therapy are metabolized via the same CYP-mediated oxidative pathway to generate catechols and quinones, which, in turn, cause DNA damage. For example, cell culture experiments showed that 4-OH-equilenin via its quinone induced DNA damage in breast cancer cell lines and cellular transformation in vitro. Thus, as far as the model is concerned, exogenous and endogenous estrogens can be combined although their precise contribution to estrogen exposure and the production of carcinogenic metabolites is presently unknown.

In designing Figure 5, we assumed that the difference in estrogen exposure between individual women is no more than twofold, with the quintiles 1.0, 1.25, 1.5, 1.75, 2.0. We chose this two-fold difference based on the range of median serum $E_2$ values seen in post-menopausal women and the variation in mammary tissue $E_2$ concentrations. This range is conservative since up to fivefold differences have been reported. Regardless of the scale used for the estrogen exposure axis, the production of carcinogenic $E_2$-3,4-$Q$ would be expected to be greater in women with more

### Table 1. Correlation of postmenopausal serum estradiol concentration by study and case-control status with breast cancer risk

| Study, country                | Estradiol pmol/L | Cases | Controls | Ratio | m   | n   | n/(m + n) | % Cases | p-value |
|------------------------------|------------------|-------|----------|-------|-----|-----|-----------|---------|---------|
| Columbia, MO United States   | 55.1             | 51.4  | 1.07     | 63    | 42  | 60.0| 0.034     |
| Guernsey, United Kingdom     | 45.5             | 35.0  | 1.30     | 100   | 4   | 96.2| <0.001    |
| Nurses’ Health Study, United States | 29.4      | 25.7  | 1.14     | 77    | 31  | 71.3| <0.001    |
| NYU WHS, United States      | 134              | 101   | 1.33     | 102   | 4   | 96.2| <0.001    |
| ORDET, Italy                | 21.9             | 21.7  | 1.01     | 59    | 55  | 51.7| >0.1      |
| Rancho Bernardo, United States | 36.7      | 40.4  | 0.908    | 31    | 73  | 29.8| >0.1      |
| RERF, Japan                 | 63.1             | 64.5  | 0.978    | 39    | 69  | 36.1| >0.1      |
| SOF, United States          | 29.4             | 22.0  | 1.34     | 102   | 4   | 96.2| <0.001    |
| Washington Country, United States | 62.4      | 58.7  | 1.06     | 65    | 44  | 59.6| 0.037     |
| Average of all centers      |                  |       |          | 1.126 | 78  | 35  | 69.0      | <0.001  |
endogenous (more ovulatory cycles) or exogenous (hormone replacement therapy, oral contraceptives) estrogen exposure. It is evident from Figure 5 that our in silico model extends existing models by combining the traditional “phenotypic” measures of estrogen exposure with genotypic data. It is also evident from the 3D graph that the combined phenotypic and genotypic data appear to have not just an additive, but also a multiplicative effect on \( E_2-3,4-Q \) production.

Current models of breast cancer risk prediction are mainly based on cumulative estrogen exposure but do not reflect mammary estrogen metabolism;\(^\text{16}\) (www.cancer.gov/bcrisktool). Moreover, they do not address genetic variability between women in exposure to estrogen metabolites. Our model addresses the unique genetic trait of each woman and combines the genetic information with the metabolomic information in order to predict individual-level mammary estrogen metabolism. Some genetic traits are currently available in the patient care setting, such as BRCA and CYP2D6 testing. The availability of rapid genetic testing for BRCA1 and BRCA2 mutations has made it possible to follow unaffected carriers in greater numbers and to search for inherited mutations in women with a severe family history of breast cancer. The potential effect of CYP2D6 genetic variants on clinical response to tamoxifen treatment in breast cancer patients has gained much interest.\(^\text{53}\) The Food and Drug Administration recommended an update in the tamoxifen package insert in 2006 to reflect the increased risk of breast cancer recurrence in postmenopausal estrogen receptor-positive patients, who are CYP2D6 poor metabolizers. Thus, the CYP2D6 genotype has the potential to become a useful predictive marker for tamoxifen response. Certain characteristics are beneficial for a marker to become successful clinically.\(^\text{54}\) Testing of this marker should be cost-effective as well as easy to apply in daily practice, both of which are increasingly realized for DNA analysis. Thus, analysis of multiple genes encoding the enzymes in the estrogen metabolism pathway can readily be achieved. Estrogen is a universal breast cancer risk factor; by helping to define high-risk subgroups, the proposed model should advance the overall goal of reducing breast cancer mortality through improved screening and the early detection and treatment of disease. Rates of obesity, an important source of estrogen after menopause are on the rise in most of the world,

Figure 5. Three-dimensional graph displaying estrogen metabolomic-genomic model of breast cancer risk. The risk is represented by the amount of carcinogenic estrogen quinone, \( E_2-3,4-Q \) AUC, which is produced by the metabolism of estrogen catalyzed by the enzymes CYP1A1, CYP1B1, and COMT. In theory, all enzyme genotype combinations could be plotted. However for clarity, we have plotted representative combinations from lowest to highest, separated into quintiles based on their respective \( E_2-3,4-Q \) production: (1) CYP1A1\(^{*1777/-}\)COMT\(^{*}\), (2) CYP1A1\(^{*}1777-462\)\text{AUC} , (3) CYP1B1\(^{*}48\)\text{AUC} , (4) CYP1B1\(^{*}48\)\text{AUC} , (5) CYP1B1\(^{*}48\)\text{AUC} . Following the characterization of GSTP1 variants, we will include GSTP1 genotype data in the model. Cumulative estrogen exposure is displayed in quintiles. Actual \( E_2 \) values, measured in pmol/L, could be plotted, in combination with semi-quantitative estimates of each woman’s overall exposure to estrogen. The latter is derived by taking into account her total years of ovulation as a function of current age, age at menarche, age at menopause, numbers of full-term pregnancies and lactation experience for each, and the dosage and duration of the use of exogenous estrogens. (The authors acknowledge work of Eric Parl in the design and preparation of this figure).
underscoring the importance of establishing the impact of estrogen metabolites on breast cancer risk before and after menopause. Women who carry a germline mutation of BRCA1 frequently develop breast cancer at an early age. However, in any given kindred the age of onset can vary substantially and an important unresolved question is the extent to which other risk factors modify the cancer risk in carriers. Estrogen exposure appears to play an important role since prophylactic oophorectomy is associated with a significant reduction in the risk of breast cancer. A practical clinical application of the model in the premenopausal age group would be the differentiation of BRCA1 carriers into low- and high-risk based on their genetic profile of estrogen metabolism. Another clinical application of the model would be in the postmenopausal age group with the distinction of low- and high-risk women. The former could benefit from hormone replacement therapy whereas the latter should avoid such treatment.

In summary, our in silico model integrates pathway-specific genetic testing with diverse types of data and for the first time offers the opportunity to combine exposure, metabolic, and genetic data in assessing estrogens in relation to breast cancer risk. In order to achieve such comprehensive risk assessment, the model will require extensive validation.

Acknowledgments
Supported in part by NIH 1R01CA/ES83752, 5U54CA113007.

Disclosure
The authors report no conflicts of interest.

References
1. McMahon B, Feinleib M. Breast cancer in relation to nursing and menopausal history. J Natl Cancer Inst. 1960;24:733–53.
2. Pike MC, Krailo MD, Henderson BE, et al. ‘Hormonal’ risk factors, ‘breast tissue age’ and the age-incidence of breast cancer. Nature. 1983;303:767–70.
3. Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis. 2000;21:427–33.
4. The Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. J Natl Cancer Inst. 2002;94:606–16.
5. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81:1879–86.
6. Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst. 2001;93:358–66.
7. Parl FF. Estrogens, Estrogen Receptor and Breast Cancer. Amsterdam, IOS Press. 2000.
8. Yue W, Santen RJ, Wang JP, et al. Genotoxic metabolites of estradiol in breast: potential mechanism of estradiol induced carcinogenesis. J Steroid Biochem Mol Biol. 2003;86:477–86.
9. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. New Engl J Med. 2006;354:270–82.
10. Stack DE, Byun J, Gross ML, et al. Molecular characteristics of catechol estrogen quinones in reactions with deoxyribonucleosides. Chem Res Toxicol. 1996;9:851–9.
11. Akanni A, Abul-Hajj YJ. Estrogen-nucleic acid adducts: reaction of 3,4-estriene o-quinone radical anion with deoxyribonucleosides. Chem Res Toxicol. 1997;10:760–6.
12. Bolton JL, Pisha E, Zhang F, et al. Role of quinoids in estrogen carcinogenesis. Chem Res Toxicol. 1998;11:1113–27.
13. Bolton JL, Trush MA, Penning TM, et al. Role of quinones in toxicity. Chem Res Toxicol. 2000;13:135–60.
14. Cavalieri EL, Li KM, Buhu N, et al. Catechol ortho-quinones: the electrophilic compounds that form depurinating DNA adducts and could initiate cancer and other diseases. Carcinogenesis. 2002;23:1071–7.
15. Li KM, Todorovic R, Devanesan P, et al. Metabolism and DNA binding studies of 4-hydroxyestradiol and estriadiol-3,4-quinone in vitro and in female ACI rat mammary gland in vivo. Carcinogenesis, 2004;25:289–97.
16. Gaikwad NW, Yang L, Muti P, et al. The moldeular etiology of breast cancer: Evidence from biomarkers of riks. Int J Cancer. 2008;122:1949–57.
17. Liehr JG, Roy D. Free radical generation by redox cycling of estrogens. Free Radical Biol Med. 1990;8:415–23.
18. Nutter LM, Wu YY, Ngo EO, et al. An o-quinone form of estrogen produces free radicals in human breast cancer cells: correlation with DNA damage. Chem Res Toxicol. 1994;7:23–8.
19. Han X, Liehr JG. Microsomal-mediated 8-hydroxylation of guanine bases of DNA by steroid estrogens: correlation of DNA damage by free radicals with metabolic activation to quinones. Carcinogenesis. 1995;16:2571–4.
20. Liehr JG, Fang WF, Sirbasku DA, et al. Carcinogenicity of catechol estrogens in Syrian hamsters. J Steroid Biochem. 1986;24:353–6.
21. Li JJ, Li SA. Estrogen carcinogenesis in Syrian hamster tissues: role of metabolism. Fed Proc. 1987;46:1858–63.
22. Newbold RR, Liehr JG. Induction of uterine adenocarcinoma in CD-1 mice by catechol estrogens. Cancer Res. 2000;60:235–7.
23. Yager JD, Liehr JG. Molecular mechanisms of estrogen carcinogenesis. Annu Rev Pharmacol Toxicol. 1996;36:203–32.
24. Cavalieri EL, Stack DE, Devanesan PD, et al. Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. Proc Natl Acad Sci U S A. 1997;94:10937–42.
25. Hayes CL, Spink DC, Spink BC, et al. 17β-estradiol hydroxylation catalyzed by human cytochrome P450 1B1. Proc Natl Acad Sci U S A. 1996;93:9776–81.
26. Hanna IH, Dawling S, Roodi N, et al. Cytochrome P450 1B1 (CYP1B1) pharmacogenetics: association of polymorphisms with functional differences in estrogen hydroxylation activity. Cancer Res. 2000;60:3440–4.
27. Shibutani S Takeshita M, Grollman A. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431–4.
28. Terasima I, Suzuki N, Shibutani S. Mutagenic properties of estrogen-quinone derived DNA adducts in Simian kidney cells. Carcinogen. 2000;21:2035–46.
29. Dawling S, Roodi N, Mernaugh RL, et al. Catechol-O-methyltransferase catalyzed catechol estrogen methylation. Pharmacogenetics. 2002;12:517–28.
## Nomenclature

| Symbol | Quantity | Symbol | Quantity |
|--------|----------|--------|----------|
| $E_2$  | 17β-estradiol | $E_{CYP1A1}$, $E_{CYP1B1}$, $E_{COMT}$, $E_{GSTP1}$ |
| $OHE_2^2$ | 2-OHE$_2$-catechol estrogen | $k_{cat}, i = 1, ..., 14$ |
| $OHE_2^4$ | 4-OHE$_2$-catechol estrogen | $K_{mi}, i = 1, 2, ..., 14$ |
| $MeOHE_2^2$ | 2-MeOE$_2$-methoxyestrogen | $k_1$ and $k_2$ |
| $MeOHE_2^{23}$ | 2-OH-3-MeOE$_2$-methoxyestrogen | $\alpha$ and $\beta$ |
| $MeOHE_2^4$ | 4-MeOE$_2$-methoxyestrogen | $V_{max Q1}$, $V_{max Q2}$, $K_{mQ1}$, $K_{mQ2}$ |
| $OHE_2^{21}SG$ | 2-OHE$_2$-1-SG-GSH-conjugate | Rate Constants for decay of Quinones |
| $OHE_2^{24}SG$ | 2-OHE$_2$-4-SG-GSH-conjugate | Hill exponents |
| $OHE_2^{42}SG$ | 4-OHE$_2$-2-SG-GSH-conjugate | Rate Constants for Quinone Formation |
| $EQ_2^{23}$ | $E_2$-2,3-Q-quinone | |
| $EQ_2^{34}$ | $E_2$-3,4-Q-quinone | |
| $t$ | time | |
| $\int_{0}^{T} EQ_2^{34} dt$ | Area under the Curve (AUC) | |