CYP3A7*1C allele is associated with reduced levels of 2-hydroxylation pathway estrogen metabolites.

Deepti Sood[1], Nichola Johnson[2], Pooja Jain[1], Alexandros P Siskos[1], Mark Bennett[3], Clare Gilham[4], Marta Cecilia Busana[4], Julian Peto[4], Isabel dos-Santos-Silva[4], Hector C. Keun[1] and Olivia Fletcher[2*]

[1]Department of Surgery and Cancer, Faculty of Medicine, Imperial College, London, SW7 2AZ, UK

[2]Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, SW7 3RP, UK

[3]Department of Life Sciences, Imperial College, London, SW7 2AZ, UK

[4]Department of Non-communicable Disease Epidemiology, The London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.

Corresponding author:

Olivia Fletcher

Breast Cancer Now Toby Robins Research Centre

The Institute of Cancer Research,

London SW7 3RP, UK

Tel: +44 (0) 207 153 5177

Fax: +44 (0) 207 153 5340

E-mail: Olivia.Fletcher@icr.ac.uk

Running title: *CYP3A7*1C* allele and estrogen metabolites

Abstract

Background: Endogenous sex hormones are well established risk factors for breast cancer; the contribution of specific estrogen metabolites (EMs) and/or ratios of specific EMs is less clear. We have previously identified a *CYP3A7*1C* allele that is associated with lower urinary estrone (E_1) levels in premenopausal women. The purpose of this analysis was to determine whether this allele was associated with specific pathway EMs.

Methods: We measured successfully 12 EMs in mid-follicular phase urine samples from 30 *CYP3A7*1C* carriers and 30 non-carriers using HPLC-MS/MS.

Results: In addition to having lower urinary E_1 levels, *CYP3A7*1C* carriers had significantly lower levels of four of the 2-hydroxylation pathway EMs that we measured (2-hydroxyestrone, P=1.1x10^{-12}; 2- hydroxyestradiol, P=2.7x10^{-7}; 2-methoxyestrone, P=1.9x10^{-12}, 2-methoxyestradiol, P=0.0009). By contrast, 16α-hydroxylation pathway EMs were slightly higher in carriers and significantly so for 17-epiestriol (P=0.002).

Conclusions: The *CYP3A7*1C* allele is associated with a lower urinary E_1 levels, a more pronounced reduction in 2-hydroxylation pathway EMs and a lower ratio of 2-hydroxylation:16α-hydroxylation EMs in premenopausal women. To further characterise the association between parent estrogens, EMs and subsequent risk of breast cancer, characterisation of additional genetic variants that influence estrogen metabolism, and large prospective studies of a broad spectrum of EMs will be required.

(199 words)

Key words: *CYP3A7*1C* allele, estrogen metabolites, breast cancer

Introduction

Endogenous sex hormones are well established risk factors for breast cancer. Pooled analyses of data from prospective studies have estimated that a doubling of circulating estradiol (E_2), free E_2 or estrone (E_1) is associated with a 20-30% or 30-50% increase in breast cancer risk in pre- and postmenopausal women, respectively (Key *et al*, 2002; Key *et al*, 2013). The contribution of specific estrogen metabolites (EMs) to breast cancer risk is less clear. Briefly, interconversion between the parent estrogens, E_2 and E_1 occurs by reversible oxidation at the 17α-position of the steroid ring; conversion of parent estrogens to EMs is by irreversible hydroxylation at the 2-, 4- or 16-positions (Figure 1 (Badawi *et al*, 2001; Samavat & Kurzer, 2015; Tsuchiya *et al*, 2005)). A recent review, summarising evidence from four prospective studies of estrogen metabolism and breast cancer risk, concluded that there was consistent evidence that enhanced 2-hydroxylation was associated with a reduction in risk of breast cancer that was independent of the strong positive associations of unconjugated parent estrogens (E_2 and E_1) with breast cancer risk (Ziegler *et al*, 2015).

In an analysis of single nucleotide polymorphisms (SNPs) tagging genes that are involved in estrogen synthesis and metabolism we identified one SNP, rs10273424, that was associated with a 22% reduction in levels of urinary estrone glucuronide (E_1G) in premenopausal women (Johnson *et al*, 2012). rs10273424 maps to the cytochrome P450 3A (*CYP3A*) gene cluster at 7q22.1; the *CYP3A* genes (*CYP3A5, CYP3A7* and *CYP3A4*) encode enzymes that catalyze the oxidative metabolism of a wide range of exogenous and endogenous substrates including parent estrogens (E_1 and E_2; Figure 1). The metabolic capacity of the CYP3A enzymes differ, depending on the substrate (Williams *et al*, 2002); with respect to estrogen metabolism specifically, 2-hydroxylation of E_1 to 2-OHE_1 is catalysed by CYP3A4, 4-hydroxylation to 4-OHE_1 is catalysed by CYP3A4 and CYP3A5 and 16α-hydroxylation to 16α-OHE_1 is catalysed by CYP3A4, CYP3A5 and CYP3A7 (Figure 1; (Badawi *et al*, 2001; Lee *et al*, 2003; Tsuchiya *et al*, 2005)). Fine-mapping of the 7q22.1 association signal for urinary E_1G levels implicated

the *CYP3A7*1C* allele as the causal allele (Johnson *et al*, 2016). This allele, which comprises seven highly correlated single base changes mapping to the *CYP3A7* promoter, results in expression of the fetal *CYP3A7* gene in adult carriers of the *CYP3A7*1C* allele (Burk *et al*, 2002; Gonzalez, 1988; Kuehl *et al*, 2001).

The purpose of this current analysis was to determine whether, in addition to the association between *CYP3A7*1C* carrier status and lower urinary E_1G levels, the *CYP3A7*1C* allele was associated with a reduction in levels of specific pathway EMs.

Materials and methods

Study population

The study population from which the women we included in this analysis were drawn has been described previously (Johnson *et al*, 2012). Briefly, they comprised 729 premenopausal women who were first-degree relatives and friends of breast cancer cases participating in the British Breast Cancer study (Johnson *et al*, 2005) or participants in the intervention arm of a trial of annual mammographic screening in young women conducted in Britain (Mammography Oestrogens and Growth factors study (Walker *et al*, 2009)). To be eligible women had to be having regular menstrual cycles, not using hormone replacement therapy or oral contraceptives and not to have been diagnosed with breast cancer at recruitment to the study. All women were of self-reported White British ancestry. To be included in the original analysis and this subsequent analysis, women had to have provided serial urine samples (six follicular phase and one luteal phase), at pre-specified days of their menstrual cycle for measurement of creatinine adjusted urinary E_1G using an in-house enzyme-linked immunosorbent assay (Johnson *et al*, 2012). To maximise the statistical efficiency of this analysis of the *CYP3A7*1C* allele, which has a minor allele frequency (MAF) of just 4% in Northern European populations (Johnson *et al*, 2016), we selected 60 women on the basis of genotype; a random sample of 30 *CYP3A7*1C* carriers and 30 *CYP3A7*1C* non-carriers. We further

selected the two periovulatory samples (samples three and four of the six sequential follicular phase samples) on the basis that estrogen levels would be at their peak in these samples and this would maximise any differences in levels between *CYP3A7*1C* carriers and non-carriers. To minimise random variation, we used the average of these two sequential samples as our outcome variable.

Ethics

The study was conducted in accordance with the tenets of the Declaration of Helsinki and all participants provided written informed consent.

Genotyping

Genotyping of the tag SNP rs45467892, which is perfectly correlated with the *CYP3A7*1C* allele has been described previously (Johnson *et al*, 2016). Briefly, genotyping was by Taqman (Life Technologies, Paisley, UK). The call rate was 96.9% and concordance between duplicate samples was 100%.

HPLC-MS/MS analysis

LC-MS/MS analysis was carried out for 14 EMs using the method of Xu and colleagues (Xu *et al*, 2007; Xu *et al*, 2005). We were unable to measure one of the 15 EMs measured by Xu and colleagues (16-Epiestriol) as there was no commercially available standard for this EM.

Briefly, two aliquots of frozen urine per woman were sent to the Mass Spectroscopy Facility for Quantitative Analysis, Faculty of Natural Science, Imperial College, London for analysis. Hydrolysis of glucuronide and sulfate conjugated EMs to form free EMs was carried out by mixing 500ul of freshly thawed urine with 20ul of internal standards (comprising 2ng of each of five deuterium-labelled estrogen metabolites (d-EMs); 17β-estradiol-d4, estriol-d3, 2-hydroxy-17β-estradiol-d5, 2-methoxy-17bestradiol-d5 and 16-epiestriol-d3; Qmx Laboratories Ltd, UK) and 500ul of freshly prepared enzymatic hydrolysis buffer (100 µL of β-glucuronidase from *Helix pomatia* (Type H-2; Sigma-Aldrich)

in 10mL 0.15M sodium acetate buffer, pH4.6 containing 2 mg of ascorbic acid). Samples were incubated at 37°C overnight before extraction with dichloromethane and dansyl chloride dervitazation. The final derivatized samples (200 µL) were transferred to HPLC vials. Urine samples were randomly allocated to one of six analytical batches. Each batch contained one matrix blank, one matrix blank spiked with internal standards, eight-point calibration standards, three quality control (QC) samples and twenty urine samples. QC samples were prepared using charcoal stripped human urine (Golden West Biological Inc., CA, USA) with no detectable levels of estrogen metabolites, spiked with all 14 EMs at a concentration of 2ng/ml.

Samples (10 µl) were then analysed by HPLC-electrospray ionisation/MS-MS using an Agilent 1100 HPLC coupled to an SCIEX QTRAP 6500 mass spectrometer (SCIEX, California, USA) running in multiple reaction monitoring (MRM) mode. Chromatographic separation was carried out on a Phenomenex Synergi Hydro-RP 4 µm x 150 mm × 2.0 mm column, at 40°C. The solvent gradient used was 35%A (99.8% H_2O: 0.2% CHOOH) to 85%B (99.8% MeOH: 0.2% CHOOH) over 60 min. Solvent B was held at 85% for 4 min then the solvent returned to 35% A for 10 min equilibration prior to the next injection. The solvent flow rate was 250 µl/min. The ESI source (type: Ion Drive Turbo V) parameters were set to - TEM 500°C, Curtain Gas 45psi, GS1 50psi, GS2 60psi, MS parameters were CAD gas Medium, DP 80, EP 10, CE 45, CXP 5. A scheduled detection method was used, the MRM detection window was 120 sec with a target scan time of 1 sec. Transitions and retention times are listed in Supplementary Table 1.

Analyst® 1.6.2 software was used for quantitation of the EMs. Peak quantitations were carried out using deuterium labelled-EM internal standards and constructing matrix matched (charcoal stripped human urine) 8-point calibration curves for each of the six analytical batches. The calibration curves were evaluated by plotting the peak area ratios of dansyl-EM/d-dansyl-EM against concentration (ng/mL) of EMs in the standard and using linear regression with 1/X weighting to fit the data. Using this linear function, the amounts of EMs in the urine sample were interpolated.

The intra- and inter-batch coefficients of variation (CVs), evaluated from three QC samples per analytical batch, in six independent consecutive batches (N=18 QC samples) ranged from 6% to 10% (intra-batch CV) and 6% to 14% (inter-batch CV; Supplementary Table 2). The two highest inter-batch CVs were for 4-methoxyestrone (14%) and 4-methoxyestradiol (13%), the two estrogen metabolites that had the lowest concentrations and which were subsequently excluded from the analysis. The LLOQ was estimated as 80 pg/mL, where the intra- and inter-batch precision of all the EMs were consistently less than 10% and the intra- and inter batch accuracies were between 97 and 105%. The LOD for the entire assay (all 14 EMs) was estimated to be 8 pg/mL.

Statistical analysis

EMs were converted from nanograms per ml of urine to picomoles per mg creatinine (pmol/mg) using the molecular weight of the unconjugated form of each of the EMs and the amount of creatinine (measured in mg/ml) for each of the samples. This allowed us to create pathway variables as described by Faupel-Badger and colleagues (Faupel-Badger *et al*, 2010). Where both of the two samples per woman were measured successfully we used the mean of the two measurements, where one of two measurements from a woman was missing, we used the one available measurement, where both measurements were missing we excluded the woman from the analysis of this EM. For the majority of EMs there were no missing values. For four EMs (2-hydroxyestrone (2-OHE1), 2-methoxyestradiol (2-MeOE_2), 2-hydroxyestrone-3-methyl ether (3-MeOE_1) and 16-ketoestradiol (16-ketoE_2)) there were <10% missing values (Supplementary Table 3). For two EMs (4-methoxyestrone (4-MeOE_1) and 4-methoxyestradiol (4-MeOE_2)) levels were undetectable in the majority of samples (92 (76.7%) and 107 (89.2%) for 4-MeOE_1 and 4-MeOE_2, respectively). These two EMs were excluded from further analysis. For the 12 EMs that we were able to measure, we estimated the within woman variation based on the two sequential samples per woman; intra-class correlation coefficients (ICCs) for individual EMs and grouped EMs are shown in Supplementary Table 4.

For individual and pathway EMs we calculated geometric means and 95% confidence intervals (CIs) on the natural logarithm scale and exponentiated the values back to the original scale. Linear regression models of the $\log_e$ transformed EMs were used to estimate percent differences between *CYP3A7*1C* carriers and non-carriers for individual and pathway EMs. We carried out unadjusted analyses and analyses that were adjusted for measurement batch (1 -6), body mass index (BMI; quartiles), age at first full term pregnancy (quartiles) and parity (0, 1-2, >2). Adjustment for these covariates did not alter effect sizes substantially and, therefore, unadjusted results are presented. We used *t* tests of the linear regression coefficient to estimate *P* values. We applied a Bonferroni correction to establish a statistical significance level of $P < 0.003$, based on our measuring of 14 individual EMs and five grouped EMs. Statistical analyses were carried out using R (version 3.2.3; http://cran.r-project.org). All reported *P* values are two-sided.

Results

Characteristics of the 60 women who were included in this analysis are presented in Table 1; there were no differences between *CYP3A7*1C* carriers and non-carriers. Levels of EMs in *CYP3A7*1C* carriers and non-carriers are shown in Figure 2 (full details of individual EMs, grouped EMs and ratios of EMs are in Supplementary Table 3). The predominant estrogen/EMs were estriol (E_3), E_1 and 2-hydroxyestrone (2-OHE_1) with geometric mean concentrations in non-carriers of 63.6, 28.6 and 21.8 pmol/mg creatinine respectively. The least abundant EMs were the methylated 2-catechol EMs, 2-methoxyestradiol (2-$MeOE_2$) and 2-Hydroxyestrone-3-methyl ether (3-$MeOE_1$) and 17-Epiestriol (17-$epiE_3$) with geometric mean concentrations in non-carriers of 0.67, 0.74 and 0.31 pmol/mg creatinine, respectively. Levels of the methylated 4-catechol EMs, 4-methoxyestrone (4-$MeOE_1$) and 4-methoxyestradiol (4-$MeOE_2$) were below the limits of detection in 92 (76.7%) and 107 (89.2%) of the 120 samples that we assayed.

Comparing *CYP3A7*1C* carriers to non-carriers, levels of the parent estrogen E_1 were 45.3% lower in *CYP3A7*1C* carriers (P = 0.0005, Table 2). For the 2 catechol EMs (2-OHE_1 and 2-OHE_2) and the methylated 2 catechol EMs (2-$MeOE_1$ and 2-$MeOE_2$) the reduction in urinary levels was more extreme; compared with non-carriers levels in *CYP3A7*1C* carriers were -78.3% ($P = 1.1 \times 10^{-12}$), -67.9% ($P = 2.7 \times 10^{-7}$), -81.2% ($P = 1.9 \times 10^{-12}$), and -62.8% (P = 0.0009) for 2-OHE_1, 2-OHE_2, 2-$MeOE_1$ and 2-$MeOE_2$, respectively (Table 2). By contrast, levels of the 16-pathway EMs were slightly higher in *CYP3A7*1C* carriers, i.e. +25.5% (P = 0.11), +91.6% (P = 0.007) and +160.1% (P = 0.002) higher for E_3, 16α-hydroxyestrone (16α-OHE_1) and 17-epiE_3, respectively (Table 2). Adjusting for measurement batch, body mass index, age at first full term pregnancy and parity did not alter these results substantially (Supplementary Table 5).

EMs from the three different estrogen metabolism pathways (2-hydroxylation, 4-hydroxylation and 16α-hydroxylation) have been associated with different estrogenic and genotoxic properties (Faupel-Badger *et al*, 2010; Ziegler *et al*, 2015) with a high 2-hydroxylation:16α-hydroxylation pathway ratio, generally being considered to be associated with a reduction in breast cancer risk. Comparing *CYP3A7*1C* carriers to non-carriers, the 2-OHE_1:16α-OHE_1 ratio in carriers (0.39, 95% CI 0.26 to 0.57) was significantly lower than the ratio in non-carriers (3.86, 95% CI 2.53 to 5.89; $P = 1.0 \times 10^{-11}$; Supplementary Table 3). Similarly, combining all 2-hydroxylation pathway and 16α-hydroxylation pathway metabolites the ratio in *CYP3A7*1C* carriers (0.10, 95% CI 0.07 to 0.13) was much lower than the ratio in non-carriers (0.51, 95% CI 0.38 to 0.67; $P = 1.7 \times 10^{-9}$; Supplementary Table 3).

Discussion

To our knowledge, comprehensive data on urinary EMs in premenopausal women, measured using LC-MS have been previously reported in three studies (Eliassen *et al*, 2012; Eliassen *et al*, 2009; Faupel-Badger *et al*, 2010; Maskarinec *et al*, 2012). The first of these was the Nurses' Health Study II (NHS II), a prospective study of North American registered nurses aged 25 – 42 years at recruitment

(Eliassen *et al*, 2012; Eliassen *et al*, 2009). The second was a population based study of incident breast cancer among women of Asian ancestry living in California and Hawaii (Faupel-Badger *et al*, 2010) and the third was a randomised trial of the effect of consuming soy foods on hormonal outcomes, conducted in women of Caucasian, Native Hawaiian and Asian ancestry and living in Hawaii (Maskarinec *et al*, 2012).

Comparing absolute levels of EMs in our study with those reported by these other studies is not straightforward. Levels of individual EMs and all EMs combined may differ between ethnicities (Maskarinec *et al*, 2012) and women from several different ethnicities have been included in the reports to date (Asian, (Faupel-Badger *et al*, 2010); African-American, Asian, Hispanic and Caucasian, (Eliassen *et al*, 2012); Caucasian, Native Hawaiian and Asian, (Maskarinec *et al*, 2012)). In addition, the 30 non-carriers that we analysed may not be representative of the general British population as they were selected on genotype (with a MAF of 4% we would expect 2.2 *CYP3A7*1C* carriers in a population-based sample of 30 women) and six (20%) were first-degree relatives of breast cancer cases suggesting that they may be at higher risk than the general population. Our analysis is based on the average of two periovulatory samples, Faupel-Badger and colleagues analysed both non-luteal phase and luteal phase samples, whereas the studies reported by Eliassen and colleagues and Maskarinec and colleagues focussed on luteal phase samples. There is, however, no evidence that the 2-OHE_1:16α-OHE_1 ratio differs according to the phase of the menstrual cycle (Faupel-Badger *et al*, 2010). Comparing this ratio across studies, the 2-OHE_1:16α-OHE_1 ratio in our non-carriers (3.9) was similar to the 2-OHE_1:16α-OHE_1 ratios reported by Faupel-Badger and colleagues (2.2 and 2.1 in luteal phase and non-luteal phase samples, respectively) and by Eliassen and colleagues (4.35) but somewhat lower than the ratios reported by Maskarinec and colleagues (8.4 and 13.0 in Asian and non-Asians, respectively).

We have previously demonstrated that the *CYP3A7*1C* allele is associated with lower levels of urinary E_1G (Johnson *et al*, 2016). Here we have demonstrated that *CYP3A7*1C* carriers have a more

pronounced reduction in 2-hydroxylation pathway EMs, increased 16α-hydroxylation pathway EMs and markedly reduced two-hydroxylation pathway:16α-hydroxylation pathway ratios as measured by the 2-OHE_1:16α-OHE_1 ratio (0.39 in *CYP3A7*1C* carriers compared to population estimates of 2.1 to 13.0) or all 2- and 16α-pathway metabolites combined (0.10 in *CYP3A7*1C* carriers compared to population estimates of 0.90 or 0.96 (Eliassen *et al*, 2012; Eliassen *et al*, 2009)). Our data are consistent with expression of the fetal *CYP3A7* gene in adult carriers of the *CYP3A7*1C* allele resulting in (i) a modest reduction in levels of the parent estrogen E_1 and (ii) a specific bias towards 16α-hydroxylation (which can be catalysed by CYP3A7) over 2- or 4- hydroxylation (which are not catalysed by CYP3A7 (Lee *et al*, 2003)).

In their recent review of estrogen metabolism and breast cancer, Ziegler and colleagues concluded that the combined evidence from prospective studies using LC-MS to measure EMs in pre-diagnostic serum, plasma and urine was consistent with the hypothesis that enhanced 2-hydroxylation is associated with reduced risk of breast cancer (Ziegler *et al*, 2015). They further concluded that the inverse association with enhanced 2-hydroxylation (specifically 2-hydroxylation:16α-hydroxylation and 2-hydroxylation:parent estrogens ratios) was independent of the strong positive associations between unconjugated estradiol and estrone with postmenopausal breast cancer risk (Ziegler *et al*, 2015). Our analysis supports these conclusions if we assume that breast cancer risk in carriers of the *CYP3A7*1C* allele is influenced by two components with opposite effects. Based on the combined evidence of prospective studies of premenopausal hormone levels and breast cancer risk, lifetime lower levels of E_1 and E_2 of approximately 45% and 27% in carriers of the *CYP3A7*1C* allele would be predicted to result in a substantial reduction in risk for these individuals. By contrast, an unfavourable (reduced) 2-hydroxylation:16α-hydroxylation ratio would be expected to increase risk. Our previous analyses demonstrate that the reduction in breast cancer risk for carriers of the rare allele of rs10235235 (which is correlated with the *CYP3A7*1C* allele $r^2 = 0.25$, D' = 1.0) is very modest (heterozygote OR = 0.98, $P = 0.2$; homozygote OR = 0.80, $P = 0.004$) (Johnson *et al*, 2014) consistent

with an unfavourable (reduced) 2-hydroxylation:16α-hydroxylation counteracting a potentially more substantial beneficial effect of lower levels of parent estrogens.

Strengths of our study include our comprehensive analysis of urinary EMs using HPLC-MS/MS and, since genotypes are effectively randomised at birth (Ebrahim & Davey Smith, 2008), our focus on a genetic variant (the *CYP3A7*1C* allele) which minimises the potential for confounding by unmeasured environmental factors. There are several limitations to our study; due to the frequency of the *CYP3A7*1C* allele the numbers are small and, given our selection procedure, the non-carriers may not be representative of the White British population. Our choice of using two consecutive periovulatory samples makes it difficult to compare EM levels in our study with other published reports, which have mainly analysed a single luteal phase sample (Eliassen *et al*, 2012; Eliassen *et al*, 2009; Faupel-Badger *et al*, 2010; Maskarinec *et al,* 2012). In addition, we do not have prospective data on *CYP3A7*1C* carrier status, hormone levels and breast cancer risk in large numbers of women and we cannot compare breast cancer risk in women with low levels of the parent estrogen E_1 according to their *CYP3A7*1C* status (and hence 2-hydroxylation:16α-hydroxylation ratio) directly.

In conclusion, we have demonstrated that the *CYP3A7*1C* allele has a profound effect on levels of the parent estrogen E_1 and the ratio of 2-hydroxylation:16α-hydroxylation EMs in premenopausal women. To characterise the association between parent estrogens, EMs and subsequent risk of breast cancer fully, identification of additional genetic variants that influence parent estrogens and particular pathway EMs, and further prospective studies that analyse a broad spectrum of EMs will be required.

Funding: NJ and OF are funded by Breast Cancer Now. The Mammography Oestrogens and Growth factors study was funded by Cancer Research UK grants C405/A2144 and C405/A14565; the

latter also provided support for MCB. JP and CG are funded by Cancer Research UK (C1178/A10292). HK, DS & AS are supported by the EC FP7 project HELIX (grant agreement number 603864). PJ is supported by the EC FP7 project EXPOSOMICS (grant agreement number 308610). The authors declare no conflict of interest.

References

Badawi AF, Cavalieri EL, Rogan EG (2001) Role of human cytochrome P450 1A1, 1A2, 1B1, and 3A4 in the 2-, 4-, and 16alpha-hydroxylation of 17beta-estradiol. *Metabolism: clinical and experimental* **50**(9)**:** 1001-3

Burk O, Tegude H, Koch I, Hustert E, Wolbold R, Glaeser H, Klein K, Fromm MF, Nuessler AK, Neuhaus P, Zanger UM, Eichelbaum M, Wojnowski L (2002) Molecular mechanisms of polymorphic CYP3A7 expression in adult human liver and intestine. *The Journal of biological chemistry* **277**(27)**:** 24280-8

Ebrahim S, Davey Smith G (2008) Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human genetics* **123**(1)**:** 15-33

Eliassen AH, Spiegelman D, Xu X, Keefer LK, Veenstra TD, Barbieri RL, Willett WC, Hankinson SE, Ziegler RG (2012) Urinary estrogens and estrogen metabolites and subsequent risk of breast cancer among premenopausal women. *Cancer research* **72**(3)**:** 696-706

Eliassen AH, Ziegler RG, Rosner B, Veenstra TD, Roman JM, Xu X, Hankinson SE (2009) Reproducibility of fifteen urinary estrogens and estrogen metabolites over a 2- to 3-year period in premenopausal women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **18**(11)**:** 2860-8

Faupel-Badger JM, Fuhrman BJ, Xu X, Falk RT, Keefer LK, Veenstra TD, Hoover RN, Ziegler RG (2010) Comparison of liquid chromatography-tandem mass spectrometry, RIA, and ELISA methods for measurement of urinary estrogens. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **19**(1)**:** 292-300

Gonzalez FJ (1988) The molecular biology of cytochrome P450s. *Pharmacological reviews* **40**(4)**:** 243-88

Johnson N, De Ieso P, Migliorini G, Orr N, Broderick P, Catovsky D, Matakidou A, Eisen T, Goldsmith C, Dudbridge F, Peto J, Dos-Santos-Silva I, Ashworth A, Ross G, Houlston RS, Fletcher O (2016) Cytochrome P450 Allele CYP3A7*1C Associates with Adverse Outcomes in Chronic Lymphocytic Leukemia, Breast, and Lung Cancer. *Cancer research*

Johnson N, Dudbridge F, Orr N, Gibson L, Jones ME, Schoemaker MJ, Folkerd EJ, Haynes BP, Hopper JL, Southey MC, Dite GS, Apicella C, Schmidt MK, Broeks A, Van't Veer LJ, Atsma F, Muir K, Lophatananon A, Fasching PA, Beckmann MW, Ekici AB, Renner SP, Sawyer E, Tomlinson I, Kerin M, Miller N, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Guenel P, Truong T, Cordina E, Menegaux F, Bojesen SE, Nordestgaard BG, Flyger H, Milne R, Zamora MP, Arias Perez JI, Benitez J, Bernstein L, Anton-Culver H, Ziogas A, Clarke Dur C, Brenner H, Muller H, Arndt V, Dieffenbach AK, Meindl A, Heil J, Bartram CR, Schmutzler RK, Brauch H, Justenhoven C, Ko YD, Network G, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Matsuo K, Dork T, Bogdanova NV, Antonenkova NN, Lindblom A, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Chenevix-Trench G, Beesley J, kConFab I, Australian Ovarian Cancer Study G, Wu AH, Van den Berg D, Tseng CC, Lambrechts D, Smeets D, Neven P, Wildiers H, Chang-Claude J, Rudolph A, Nickels S, Flesch-Janys D, Radice P, Peterlongo P, Bonanni B, Pensotti V, Couch FJ, Olson JE, Wang X, Fredericksen Z, Pankratz VS, Giles GG, Severi G, Baglietto L, Haiman C, Simard J, Goldberg MS, Labreche F, Dumont M, Soucy P, Teo S, Yip CH, Phuah SY, Cornes BK, Kristensen VN, Grenaker Alnaes G, Borresen-Dale AL, Zheng W, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Devillee P, Figueroa J, Chanock SJ, Lissowska J, Sherman ME, Hall P, Schoof N, Hooning M, Hollestelle A, Oldenburg RA, Tilanus-Linthorst M, Liu J, Cox A, Brock IW, Reed MW, Cross SS, Blot W, Signorello LB, Pharoah PD, Dunning AM, Shah M, Kang D, Noh DY, Park SK, Choi JY, Hartman M, Miao H, Lim WY, Tang A, Hamann U, Forsti A, Rudiger T, Ulmer HU, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Slager S, Toland AE, Vachon C, Yannoukakos D, Shen CY, Yu JC, Huang CS, Hou MF, Gonzalez-Neira A, Tessier DC, Vincent D, Bacot F, Luccarini C, Dennis J, Michailidou K, Bolla MK, Wang J, Easton DF, Garcia-Closas M, Dowsett M, Ashworth A, Swerdlow AJ, Peto J, dos Santos Silva I, Fletcher O (2014) Genetic variation at CYP3A is associated with age at menarche and breast cancer risk: a case-control study. *Breast cancer research : BCR* **16**(3)**:** R51

Johnson N, Fletcher O, Naceur-Lombardelli C, dos Santos Silva I, Ashworth A, Peto J (2005) Interaction between CHEK2*1100delC and other low-penetrance breast-cancer susceptibility genes: a familial study. *Lancet* **366**(9496)**:** 1554-7

Johnson N, Walker K, Gibson LJ, Orr N, Folkerd E, Haynes B, Palles C, Coupland B, Schoemaker M, Jones M, Broderick P, Sawyer E, Kerin M, Tomlinson IP, Zvelebil M, Chilcott-Burns S, Tomczyk K, Simpson G, Williamson J, Hillier SG, Ross G, Houlston RS, Swerdlow A, Ashworth A, Dowsett M, Peto J, Dos Santos Silva I, Fletcher O (2012) CYP3A variation, premenopausal estrone levels, and breast cancer risk. *Journal of the National Cancer Institute* **104**(9)**:** 657-69

Key T, Appleby P, Barnes I, Reeves G, Endogenous H, Breast Cancer Collaborative G (2002) Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *Journal of the National Cancer Institute* **94**(8)**:** 606-16

Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, Berrino F, Krogh V, Sieri S, Brinton LA, Dorgan JF, Dossus L, Dowsett M, Eliassen AH, Fortner RT, Hankinson SE, Helzlsouer KJ, Hoff man-Bolton J, Comstock GW, Kaaks R, Kahle LL, Muti P, Overvad K, Peeters PH, Riboli E, Rinaldi S, Rollison DE, Stanczyk FZ, Trichopoulos D, Tworoger SS, Vineis P (2013) Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *The Lancet Oncology* **14**(10)**:** 1009-19

Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski MS, Schuetz E (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nature genetics* **27**(4)**:** 383-91

Lee AJ, Conney AH, Zhu BT (2003) Human cytochrome P450 3A7 has a distinct high catalytic activity for the 16alpha-hydroxylation of estrone but not 17beta-estradiol. *Cancer research* **63**(19)**:** 6532-6

Maskarinec G, Heak S, Morimoto Y, Custer L, Franke AA (2012) The relation of urinary estrogen metabolites with mammographic densities in premenopausal women. *Cancer epidemiology* **36**(5)**:** e310-6

Samavat H, Kurzer MS (2015) Estrogen metabolism and breast cancer. *Cancer letters* **356**(2 Pt A)**:** 231-43

Tsuchiya Y, Nakajima M, Yokoi T (2005) Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer letters* **227**(2)**:** 115-24

Walker K, Fletcher O, Johnson N, Coupland B, McCormack VA, Folkerd E, Gibson L, Hillier SG, Holly JM, Moss S, Dowsett M, Peto J, dos Santos Silva I (2009) Premenopausal mammographic density in relation to cyclic variations in endogenous sex hormone levels, prolactin, and insulin-like growth factors. *Cancer research* **69**(16)**:** 6490-9

Williams JA, Ring BJ, Cantrell VE, Jones DR, Eckstein J, Ruterbories K, Hamman MA, Hall SD, Wrighton SA (2002) Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug metabolism and disposition: the biological fate of chemicals* **30**(8)**:** 883-91

Xu X, Keefer LK, Ziegler RG, Veenstra TD (2007) A liquid chromatography-mass spectrometry method for the quantitative analysis of urinary endogenous estrogen metabolites. *Nature protocols* **2**(6)**:** 1350-5

Xu X, Veenstra TD, Fox SD, Roman JM, Issaq HJ, Falk R, Saavedra JE, Keefer LK, Ziegler RG (2005) Measuring fifteen endogenous estrogens simultaneously in human urine by high-performance liquid chromatography-mass spectrometry. *Analytical chemistry* **77**(20)**:** 6646-54

Ziegler RG, Fuhrman BJ, Moore SC, Matthews CE (2015) Epidemiologic studies of estrogen metabolism and breast cancer. *Steroids* **99**(Pt A)**:** 67-75

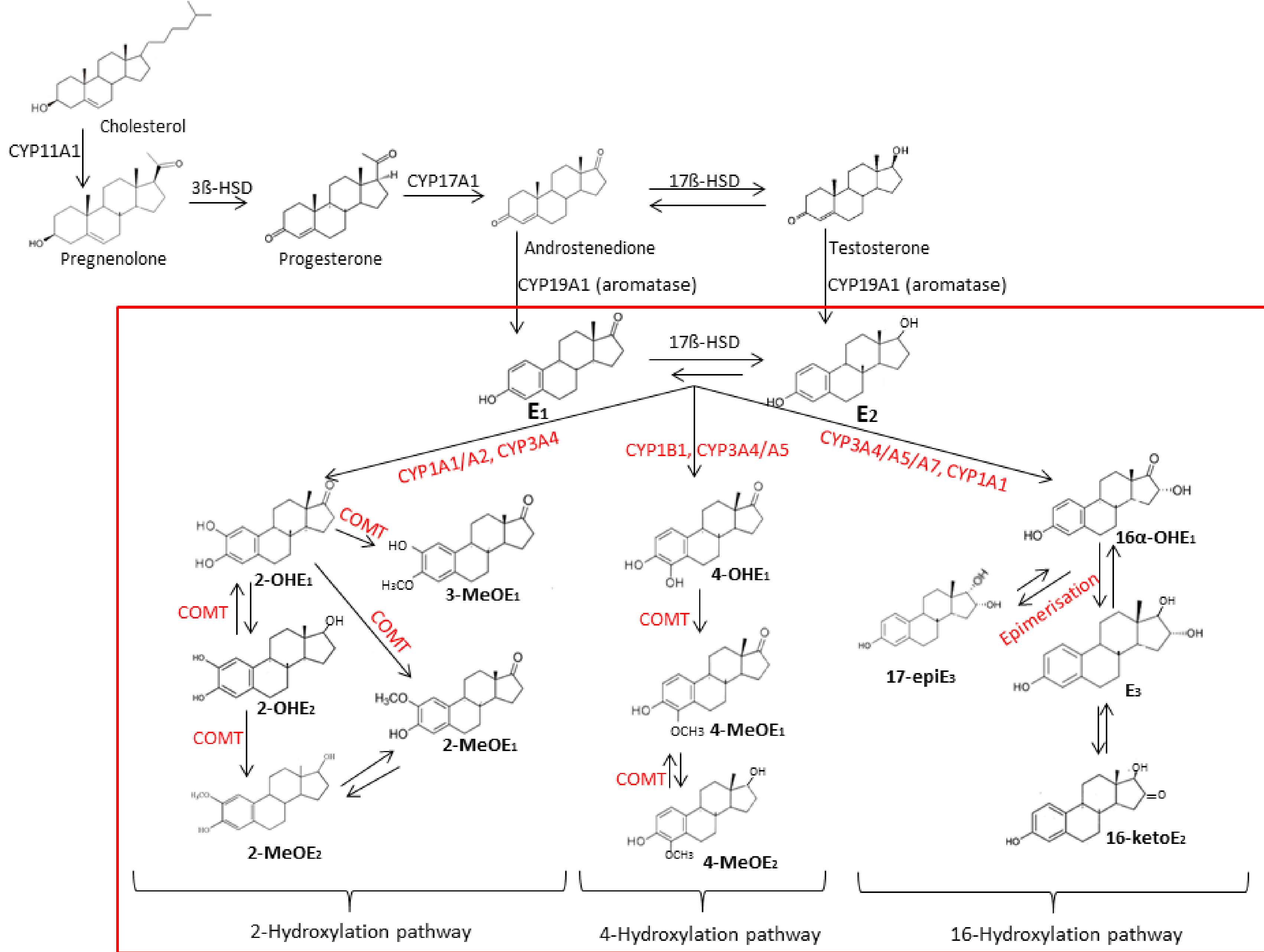
Figure Legends:

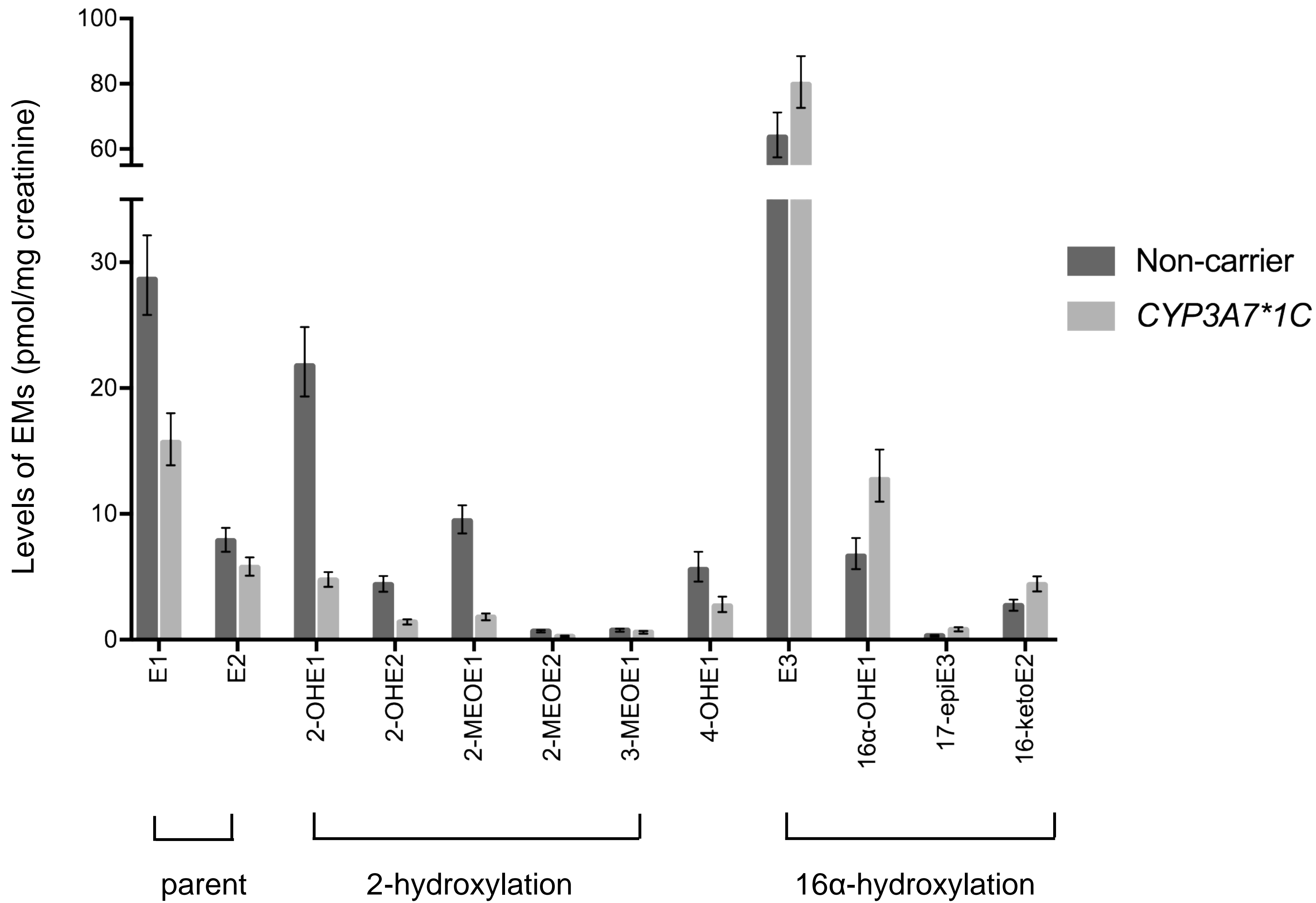
Figure 1: Steroid hormone synthesis and endogenous estrogen metabolism in humans.

The 14 estrogen metabolites (EMs) formed by hydroxylation of parent estrogens (E_1 and E_2), at the 2, 4-, and 16α-positions of the carbon ring that were measured are shown within the red box. Enzymes (cytochrome P450 (CYPs) and catechol O-methyltransferase (COMT)) involved in estrogen metabolism are in red. Estrone glucuronide (E_1G), measured in our previous analysis (Johnson *et al*, 2012) is an estrone conjugate, present in urine. The E_1 measured in this analysis (after hydrolysis of glucuronide and sulphate conjugates in the first step of the LC-MS/MS protocol) is highly correlated with E_1G (Spearman's correlation, ρ=0.70, P<0.0001). Abbreviations: E_1= estrone, E_2= estradiol, 2-OHE_1=2-hydroxyestrone, 2-OHE_2= 2-hydroxyestradiol, 2-$MeOE_1$= 2-methoxyestrone, 2-$MeOE_2$= 2-methoxyestradiol, 3-$MeoE_1$= 2-hydroxyestrone-3-methyl ether, 4-OHE_1= 4-hydroxyestrone, 4-$MeOE_1$= 4-methoxyestrone, 4-$MeOE_2$= 4-methoxyestradiol, 16α-OHE_1= 16α-hydroxyestrone, E_3= estriol, 16-$KetoE_2$= 16-ketoestradiol, 17-$epiE_3$= 17-epiestriol and 17ß-HSD = 17ß-hydroxysteroid.

Figure 2: Geometric mean levels (picomoles/mg creatinine) of 12 EMs that we measured in urine samples from 60 premenopausal women; 30 carriers of the *CYP3A7*1C* allele (light grey) and 30 non-carriers (dark grey). Estimates are the average of two samples per woman, taken on sequential days, calculated to be at or around ovulation based on the woman's usual menstrual cycle length. Error bars represent standard errors. Levels of two of the 4-hydroxylation pathway EMs (4-$MeOE_1$ and 4-$MeOE_2$) were below detection in 92 (4-$MeOE_1$) and 107 (4-$MeOE_2$) of the samples analysed. These two EMs were, therefore, excluded.

Figure 1





Levels of EMs (pmol/mg creatinine)
100
80
60
30
20
10
0
Non-carrier
CYP3A7*1C
E1
E2
2-OHE1
2-OHE2
2-MEOE1
2-MEOE2
3-MEOE1
4-OHE1
E3
16α-OHE1
17-epiE3
16-ketoE2
parent
2-hydroxylation
16α-hydroxylation

Table 1. Characteristics of the 30 *CYP3A7*1C* carriers and 30 *CYP3A7*1C* non-carriers included in this analysis.

	*CYP3A7*1C* non-carriers	*CYP3A7*1C* carriers	*P* value
Age at urine collection	44.1 (16 – 53)	45.5 (27– 50)	0.38
BMI at urine collection	24.0 (19.2 – 39.9)	25.5 (18.3 – 39.0)	0.64
Parity			
Nulliparous	10 (33.3)	6 (20.0)	
1-2 children	17 (56.7)	15 (50.0)	
>2 children	3 (10.0)	9 (30.0)	0.13
Age at first full term pregnancy[#]	26.3 (20 – 31)	27.2 (17 – 32)	0.39

[#]in parous women. For quantitative traits (age at urine collection, BMI at urine collection and age at first full term pregnancy) means and ranges are presented. For parity, the number and percentage of women in each category is shown. For quantitative traits *P* values were from *t* tests; for parity the *P* value was from a Fisher's exact test.

Table 2: Percentage difference (95% CI) in urinary EM levels comparing *CYP3A7*1C* carriers with non-carriers

Grouped EM	Individual EM	% Difference in EM[#$] (95% CI)	*P*
Total EMs		-15.1 (-32.7, 7.0)	0.16
Parent EMs		-41.0 (-57.2, -18.5)	0.002
	Estrone	-45.3 (-60.6, -24.1)	0.0005
	Estradiol	-26.7 (-47.5, 2.2)	0.07
Catechol EMs		-71.0 (-80.1, -57.9)	1.2×10^{-8}
2-Catechol EMs		-76.1 (-83.0, -66.6)	8.4×10^{-12}
	2-Hydroxyestrone*	-78.3 (-84.5, -69.6)	1.1×10^{-12}
	2- Hydroxyestradiol	-67.9 (-78.3, -52.6)	2.7×10^{-7}
4-Catechol EMs			
	4-Hydroxyestrone	-51.8 (-72.9, -14.2)	0.01
Methylated Catechol EM			
Methylated 2-Catechol EM		-73.5 (-81.2, -62.7)	1.9×10^{-10}
	2-Methoxyestrone	-81.2 (-87.1, -72.6)	1.9×10^{-12}
	2-Methoxyestradiol*	-62.8 (-78.9, -34.7)	0.0009
	2-Hydroxyestrone-3-methyl ether*	-21.9 (-50.3, 22.6)	0.28
Methylated 4-Catechol EMs			
	4-Methoxyestrone	n/a	n/a
	4-Methoxyestradiol	n/a	n/a
2-Hydroxylation pathway EMs		-74.9 (-81.6, -65.8)	1.6×10^{-12}
4-Hydroxylation pathway EMs		-51.8 (-72.9, -14.2)	0.01
16-Hydroxylation pathway EMs		34.7 (1.6, 78.5)	0.04
	16α-Hydroxyestrone	91.6 (20.2, 205.1)	0.007
	17-Epiestriol	160.1 (46.0, 363.4)	0.002
	Estriol	25.5 (-5.1, 66.0)	0.11
	16-Ketoestradiol*	62.4 (8.7, 142.7)	0.02

[#] unadjusted analysis; [$]a negative difference may be interpreted as lower levels in *CYP3A7*1C* carriers compared to non-carriers; *missing values: 2-OHE$_2$, 1 sample (non-carrier); 2-MeOE$_2$, 11 samples (11 women, 2 non-carriers, 9 carriers); 3-MeOE$_1$, 11 samples (9 women, 2 non-carriers, 7 carriers), 16-ketoE$_2$, 5 samples (3 women, 2 non-carriers, 1 carrier). n/a = not available