Estrogen exposure and breast cancer risk: interaction between estrogen related genetic and non-genetic factors

By

Jingjing Zhu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

August 31st, 2019

Nashville, Tennessee

Approved:

Xiao-Ou Shu, M.D., Ph.D.
Jirong Long, Ph.D.
Loren Lipworth-Elliot, Sc.D.
Ran Tao, Ph.D.

Copyright Page

Acknowledgements

For the past years I have received a lot of guidance, help, and support from many people for my training and completion of projects including my thesis work.

First, I would like to express my most sincere gratitude to Dr Xiao-Ou Shu, my thesis advisor, mentor as well as friend, for providing the best level of mentorship, support, advice, and help. During my research training, she had spent a lot of time and effort on guiding me through the process, from mastering relevant scientific knowledge to accumulating research experience in various projects, and improving and polishing my skills in scientific writing. She is always extremely enthusiastic and passionate for science, which keeps influencing me all the time. She keeps providing the best advice for my projects and training, and provides the strongest support for my scientific development. I am incredibly grateful to have had the opportunity to work with her for the past years.

I also want to extend my gratitude to other thesis committee members, Dr. Jirong Long, Dr. Loren Lipworth-Elliot, and Dr. Ran Tao. I received a lot of guidance, support and help from my thesis committee for better developing my thesis project and my scientific training. Dr. Long provided guidance and help to develop my expertise in genetic epidemiology. Dr. Ran Tao provided a lot of help for the power calculation aspects of my thesis project. Dr. Loren Lipworth-Elliot provided valuable suggestions for the development of my thesis project.

I would like to thank many other faculty members and colleagues in the Vanderbilt Epidemiology Center, who have been extremely supportive and helpful for my training and scientific development. I have learned a lot from their discussion and interaction during the scientific meetings involving Drs. Wei Zheng, Hui Cai, Danxia Yu, Xingyi Guo, Qiuyin Cai, Lang Wu, Wanqing Wen. I received lots of help from colleagues Drs. Xiang Shu, Yaohua Yang, and Zhishan Chen for my thesis project and other projects.

I also want to thank a lot of collaborators, without whom it would be impossible for my thesis project to be completed. I am thankful to investigators in the Breast Cancer Association Consortium (BCAC) and the Asia Breast Cancer Consortium (ABCC) for making data available for my thesis project.

The Vanderbilt Epidemiology PhD program provided lots of assistance throughout my PhD studies. I am grateful to the program leaders, Drs Todd L. Edwards, Peter F Rebeiro, and Katherine E. Hartmann, as well as program manager Molly Knowlton for their help. I also would like to thank my academic advisor Dr. Digna Velez Edwards who has been continuously providing guidance for my scientific and career development.

Finally, I would like to thank my family for their forever love, support and encouragement during my PhD studies.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1. Breast cancer is the most prevalent cancer among women

Approximately 1 in 8 women in the United States will develop invasive breast cancer in their lifetime, which accounts for about one third of all cancers diagnosed among women. It is also the second most common cause of cancer death in US women. Although the incidence rate of breast cancer has begun to decrease in the US, possibly attributed to the reduced use of hormone HRT in menopause, the incidence rate of breast cancer has continued to increase worldwide, especially in the low- and middle-income countries. This trend can be possible explained by the adoption of western lifestyle in those developing areas, including high fat and high protein diet, sedentary life, and menstrual and reproductive patterns. As the risk of having breast cancer substantially increases as age goes up, lengthened lifespan of humans has resulted in a steady grow of the aged population as well as corresponding global burden of breast cancer. A better understanding of the factors associated with breast cancer is of utmost public health importance for developing strategies to prevent this disease and improve women's health.(1–3)

Breast cancer is a group of biologically diverse cancers. Based on the intrinsic molecular features, four main subtypes of breast cancer have been identified: Luminal A, Luminal B, HER2-enriched and Basal-like. Intrinsic subtypes can be distinguished either by the expression level of RNA/protein or by DNA mutations. Compared to Luminal A tumors, Luminal B tumors show higher expression level of proliferation/cell cycle-related genes or proteins (e.g. FOXA, progesterone receptor). Also, Luminal B tumors have a higher number of mutations across the genome, higher number of chromosomal copy-number changes, more TP53 mutations (29% vs. 12%), and less PIK3CA (29% vs. 45%) and MAP3K1 (5% vs. 13%) mutations. The HER2-enriched subtype is

distinguished by its high expression level of HER2- and proliferation-related genes and proteins (e.g. ERBB2 and HER2), moderate expression level of luminal-related genes and proteins (e.g. ESR1 and PGR), and low expression level of basal-related genes and proteins (e.g. keratin5 and FOXC1). Additionally, HER2-enriched tumors reveal the highest number of mutations across the genome (e.g. 72% are TP53 mutated and 39% are PIK3CA mutated). The Basal-like subtype is differentiated by the high expression level of proliferation-related genes (e.g. MKI67) and keratins, moderate expression level of HER2-related gens, and very low expression level of luminal-related genes. Furthermore, the Basal-like tumors display the second highest number of mutations across the genome, with the vast majority of them being hypomethylated(4).

Traditionally, breast cancer is classified based on the presence of three receptors on its cells: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2–neu (HER2). A cancer is called hormone-receptor positive if one or more of the receptors are present. Otherwise, the cancer is called hormone-receptor negative. Approximately two-thirds of breast cancers are ER and/or PR positive.(5)

2. Level of estrogen is among the most important risk factors for breast cancer development

Estrogen, as the primary steroid female sex hormone, plays important roles in many physiological functions, such as regulating the menstrual cycle and reproduction, maintaining bone density and brain function, or controlling energy balance and glucose homeostasis.(1,2) It also plays and important role in the breast development not only during puberty, but also during each menstrual cycle in preparation for pregnancy. There are three types of main estrogens in human body, which are estrone (E1), estradiol (E2), and estriol (E3). E1and E2 are biochemically interconvertible by the enzyme 17β-estradiol dehydrogenase. All estrogens are derived from cholesterol, but can be metabolized through different pathways.

Accumulated evidence from epidemiological studies and randomized clinical trials has consistently suggested that high circulating estrogen levels are associated with an increased risk of breast cancer. A cohort study conducted in early 1900s involving over 6,000 women found that bilateral oophorectomy before age 40 reduced the risk of developing breast cancer by nearly 75%.(6) In 2002, results from a pooling analysis of nine prospective studies, which included 663 breast cancer cases and 1765 controls, indicated that higher level of circulating sex hormone was strongly associated with an elevated breast cancer risk among postmenopausal women.(7) The respective relative risk (RR) was 2.00 [95% CI: 1.47, 2.71] and 2.19 [95% CI: 1.48, 3.22] by comparing the highest to the lowest quintile of total estradiol and estrone levels, respectively. Several similar observational studies were conducted afterwards which revealed similar results.(8–12) The Women's Health Initiative study tested the relationship of HRT, which is to use medications containing female hormones to replace its reduced amount after menopause, with breast cancer risk in a clinical trial.(13,14) This study was terminated after an average follow-up length of five years because of increased risks of several diseases in the group receiving HRT. Specifically, they observed an inverse association for estrogen alone use, but a positive association for estrogen plus progestin use with breast cancer risk.

There are two possible explanations for the underlying mechanisms of the association between estrogen level and breast cancer risk (Figure 1). The first potential explanation is through estrogen receptor-dependent pathway. Estrogens can stimulate cell proliferation, which can increase the possibility of errors in DNA replication and mutation accumulations in breast tissue. Cell proliferation has been detected to be at its greatest extent when estrogen and progesterone are at their peak levels.(15) The other pathway that is not mediated by estrogen receptor is induced through the estrogen metabolism. The metabolism of estrone/estradiol results in the formation of catechol estrogens [including 4-OHE_1 (E_2) and 2-OHE_1 (E_2)] and 16α-hydroxylation. Catechol estrogens can be further oxidized to electrophilic catechol estrogen quinones, which may react with DNA and form

unstable depurinating DNA adducts. These adducts then generate apurinic sites in the DNA. Error-prone DNA repair of these apurinic sites can result in point mutations(16–18) and cells transformation, which eventually could contribute to the initiation of breast cancer.(19–21) 4-OH metabolites have been shown to be more carcinogenic than the 2-OH metabolites due to a much higher level of the formation of depurinating DNA adducts.(22) The other major metabolites, 16α-hydroxylation, may also play an important role in initiating breast cancer development by increasing unscheduled DNA synthesis in mammary cells.(23)

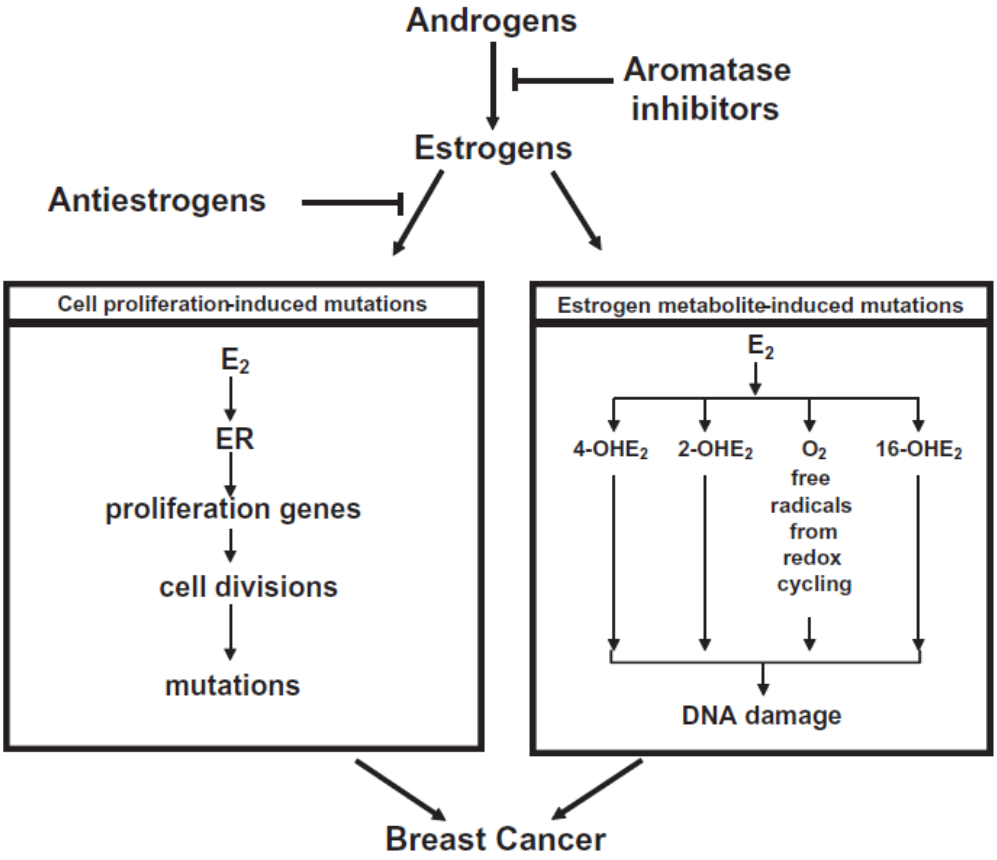


Figure 1. Pathways in the effects on estrogen on breast cancer development.(24)

Reference: Santen RJ, Yue W, Wang J-P. Estrogen metabolites and breast cancer. *Steroids*. 2015;99(Pt A):61-66. doi:10.1016/j.steroids.2014.08.003.

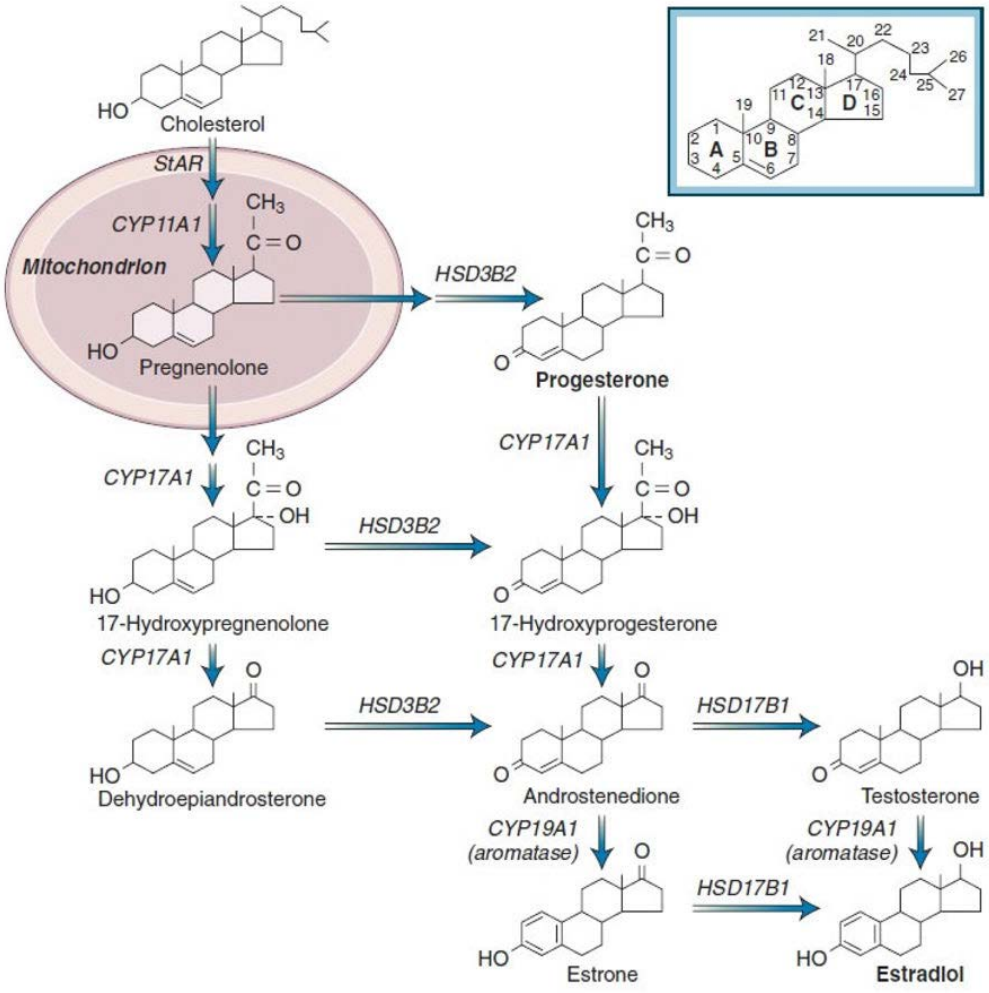


Figure 2. Pathways of steroid hormone synthesis in humans.(25)

Reference: Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. *Cancer Lett*. 2015;356(2):231-243. doi:10.1016/j.canlet.2014.04.018.

3. Endogenous estrogen level is affected by the expression level of estrogen related genes

3.1. Estrogen synthesis pathway (include current literature on SNP associations)

There are 10 related genes involved in the estrogen synthesis pathway, which included: *STAR, CYP11A1, CYP17A1, HSD3B2, HSD17B1, CYP19A1, HSD17B11, AKR1B15, HSD17B14*, and

HSD17B2. The first step in this pathway is regulated by the steroidogenic acute regulatory protein (StAR), which transported cholesterol into mitochondrion. (26) Then cholesterol is converted to pregnenolone, which is the precursor for all steroid hormones. This process is catalyzed by the mitochondrial side-chain cleavage enzyme complex, controlled by the *CYP11A1* gene. Next, pregnenolone can be transferred into progesterone or androstendione through different pathways. Ultimately, androstenedione can be metabolized to estrogens directly or indirectly through testosterone.(27) As for premenopausal women, the most important estrogen circulating in the body is E2, which is synthesized in the ovaries. While for postmenopausal women, E1 is the main estrogen, which is synthesized in peripheral, especially the adipose tissues.

3.2. Estrogen metabolism pathway

Many genes are involved in the estrogen metabolism pathway which include *STs (ARSA, ARSB, STS, ARSC2, ARSD, ARSE, ARSF, ARSG, ARSH, ARSI, ARSJ, ARSK, GALNS, GNS, IDS, SGSH, SUPF1, SULF2), SULTs (SULT1A1, SULT1A2, SULT1A3, SULT1A4, SULT1B1, SULT1C2, SULT1C3, SULT1C4, SULT1D1P, SULT1E1, SULT2A1, SULT2B1, SULT4A1, SULT6B1), CYP1A1, CYP1A2, CYP1B1, CYP3A (CYP3A4, CYP3A5), COMT, UGTs (B3GAT1, B3GAT2, B3GAT3, UGT1A1, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT1A10, UGT2A1, UGT2A2, UGT2A3, UGT2B4, UGT2B7, UGT2B10, UGT2B11, UGT2B15, UGT2B17, UGT2B28*). E1 and E2 are metabolized in three pathways. At the first step, E1 and E2 are hydroxylated by the CYP enzymes (CYP1A1, CYP1B1, CYP1A2, CYP3A) and converted to catechol estrogens (2-hydroxyestrone, 4-hydroxyestrone, 2-hydroxyestradiol, 4-hydroxyestradiol) and 16α-hydroxyestrone. Next, catechol estrogens are further methylated to methoxyestrogens by the catechol-Omethyltreansferace (COMT) enzyme. Besides hydroxylation and methylation, E1 and E2 and their metabolites can also be conjugated with sulfate by sulftransferases (SULTs), however, their sulfates can be deconjugated by sulfatases (STs). In addition to sulfate, estrogen metabolites may also be

conjugated with glucuronic acid by glucoronyltrasferases (UGTs). Both the conjugation with sulfate or glucuronic acid are considered as estrogen elimination process, by which hormones are either converted to water soluble substance and removed out of the body, or become more lipophilic with increased half-lives.(25)

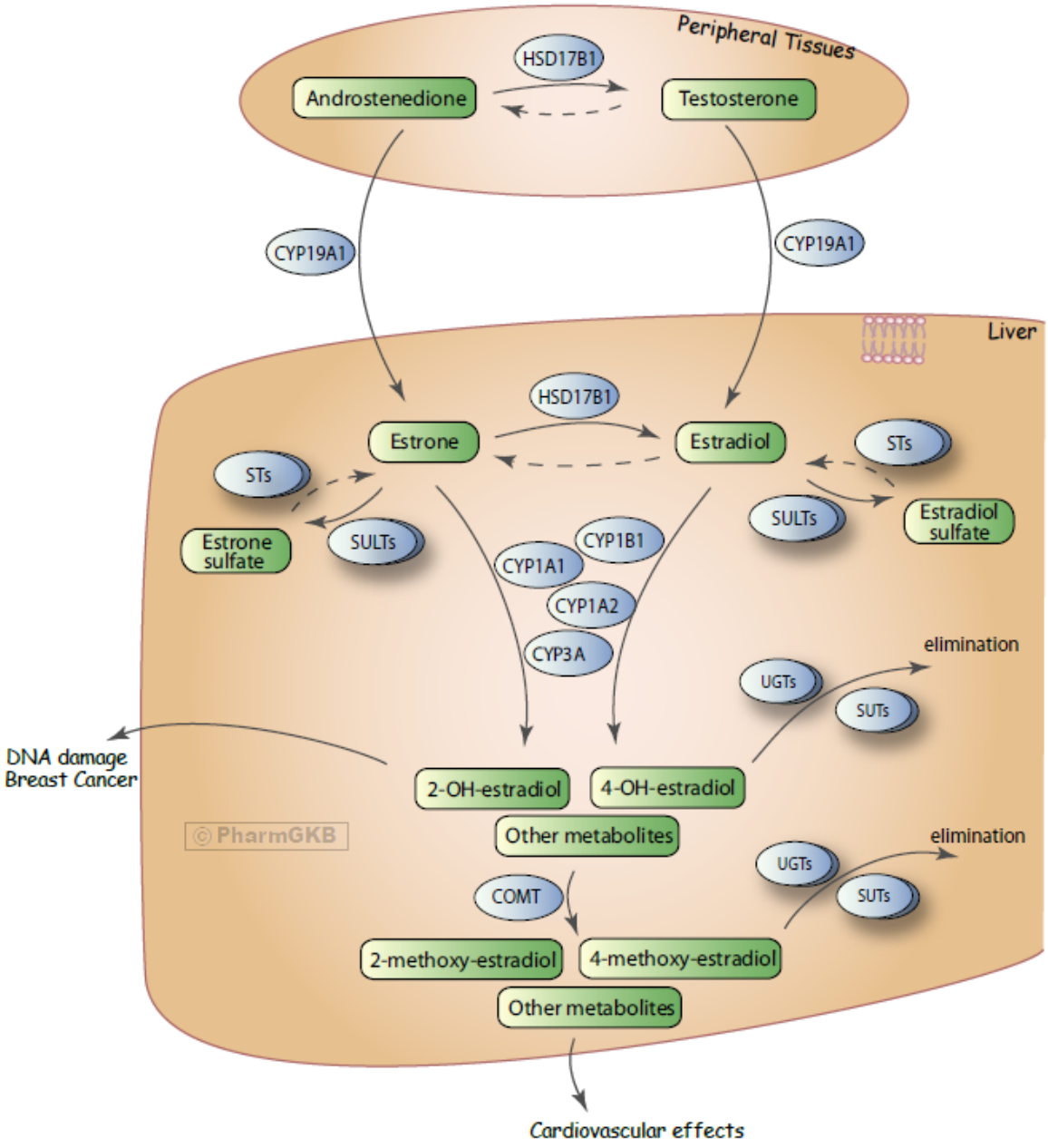


Figure 3. Estrogen metabolism pathway.

3.3. Estrogen receptors (ERs) and Sex hormone-binding globulin (SHBG)

Estrogen receptors are a group of ligand-activated enhancer protein which stays inside cells. The biological functions of estrogen are mediated by binding to ERs in targeted organs. There are two types of estrogen receptor, ERα and ERβ, which are encoded by two distinct genes, *ESR1* and *ESR2*, respectively. ERα is mainly expressed in the uterus, ovaries, breast and liver.(28) ERβ is highly expressed in ovaries, bone marrow, breast, and brain.(29) Accumulating evidence has suggested that ERα and ERβ may play different roles in the development of breast cancer. It is found that estrogen stimulated cell proliferation and tumor formation are only present given the existence of ERα.(30) Whereas, ERβ is considered as a potential carcinogenesis-inhibitor, given the evidence that its expression level is decreased in breast cancer tissue compared to in normal breast tissue.[30] In healthy breast tissue, ERα expression is much lower than ERβ. However, in neoplastic breast tissue, ERα is expressed at a higher level than ERβ.(32)

Sex Hormone Binding Globulin (SHBG), encoded by gene *SHBG*, is a steroid binding protein that is produced in liver. It can modulate the bioavailability of estrogens in the target tissue by delivering estrogens to target tissues and binding to the circulating estrogens. A defect in SHBG synthesis could result in excessive bioavailable sex hormones. SHBG also functions as part of the steroid-signaling system, which can regulates estrogen signaling at the cell membrane.(33)

3.4. Existing studies on examining the association between genetic variants in estrogen-related genes and sex hormone levels / breast cancer risk.

Several studies have shown that genetic variants in estrogen-related genes are related to sex hormone levels. In 2004, Dunning et al. have reported that CYP19 SNPs (rs10046 and [TCT]+/-) were associated with differences in circulating levels of estradiol and estradiol.(34) In another study of the Multiethnic Cohort Study in 2005, genetic variation at the *SHBG* locus is shown to be potentially associated with serum level of SHBG.(35) A pooled analysis, using data from the National Cancer

Institute Breast and Prostate Cancer Cohort Consortium, further confirmed an association between SHBG level and *SHBG* SNPs. They also suggested an association of circulating levels of E1 and E2 with polymorphisms in *CYP19* and *ESR1*.(36) In 2012, Johnson et al. found a single *CYP3A* SNP (rs10273424) to be associated with E1 level among premenopausal women.(37)

The associations between genetic variants in estrogen-related genes and breast cancer risk are, however, mostly inconsistent. A review article has summarized the associations in estrogen biosynthesis (*CYP17*, *CYP19*) and metabolism (*CYP1A1*, *CYP1B1*, *COMT*) genes.(38) For *CYP17*, 3 studies identified significant associations between its polymorphisms and breast cancer risk. However, null results were suggested in 10 other studies. Among the published studies on *CYP19* polymorphism, 6 found a significant association with breast cancer risk, while 9 other studies did not find sufficient evidence for supporting the association. With regards to *CPY1A1*, only 6 out of 17 published articles found significant associations. Similarly, only a small fraction of existing studies observed a significant relationship between SNPs variations in *CYP1B1*/*COMT* and breast cancer risk. The inconsistency of these results could be partly attributed to the underpowered study design, differences across populations, or small effect sizes. It is known that the conventional analytical methods only evaluate single variant effects in a limited number of genes, resulting in insufficient power for detecting the associations.

4. Lifestyle and non-genetic factors can affect the estrogen level as well

4.1. Adiposity

Overall and central adiposity are important predictors of estrogen concentrations for postmenopausal women.(39) After menopause, ovaries cannot produce measurable plasma level of estrogens any more. Instead, the primary source of estrogens is synthesized in adipose tissue.(40) Several studies have found that postmenopausal women with overall or central obesity have increased total and free plasma estrogen concentrations compared to those with healthy weight.(41–

44) Several possible underlying mechanisms have been proposed to explain this phenomenon. First, adipose tissues also produce enzymes that promote the aromatization of testosterone and androstenedione into E2 and E1, respectively.(45) Moreover, obesity-induced inflammation has been shown to cause the overexpression of aromatase and eventually increase the estrogen production in adipose stromal cells of the breasts.(46) Second, increased weight and BMI have been reported to be associated with reduced levels of SHBG, which results in an elevated concentrations of the bioavailable estrogen.(47)

Numerous evidences from epidemiological studies have investigated the association of obesity with breast cancer risk and found that the direction of the association between obesity and breast cancer risk depends on menopausal status. It has been consistently reported that measured adult BMI and breast cancer are positively related among post-menopausal women, however, inversely associated among pre-menopausal women.(48–56) However, a Mendelian randomization study conducted among European descendants found that BMI predicted by genome-wide association studies-identified variants is inversely associated with breast cancer risk among both pre- and postmenopausal women. The explanation for this inverse association includes that the genetic influence on BMI may better reflect the early-life BMI, which is reported to be inversely associated with breast cancer risk. (57)

4.2. Menstrual and reproductive factors, exogenous hormone use

Menstrual and reproductive factors including age at menarche, age at menopause, parity (the number of livebirths in a woman’s lifetime), as well as oral contraceptive (OC) use, and hormone replacement therapy (HRT) usage, are important determinants for a women’s lifetime exposure to estrogens. Pregnancy can increase serum estrogen levels as much as 100-fold(58). However, after pregnancy, studies have shown that estrogen levels are lower among parous women compared to nulliparous women. Consequently, women who have given birth to her first child at a younger age or

women who have higher parity are more likely to have overall lower lifetime exposure to estrogen.(59) Similarly, women with early menarche and late menopause are also more likely to have longer periods of exposure to high levels of estrogen, since estrogen levels are elevated during reproductive ages.(60,61) Commonly used OC contain two types of synthetic hormone, estrogen and progesterone which is referred to combined hormonal contraceptives. OC aims at disturbing the natural alteration of estrogens levels during each menstrual cycle and ultimately prohibiting ovulation. Women who take OC regularly were reported to have a lower level of circulating estrogen compared to women who have natural menstrual cycles.(62,63) Studies have shown that this difference persists even after the discontinuation of the OC use.(64,65) Furthermore, postmenopausal women who have a history of OC use showed reduced level of circulatory estrogen compared to those who have never used it.(66,67) HRT, used to be a commonly used treatment to relieve menopausal symptoms, works by elevating hormone levels that are declined after menopause. The Women's Health Initiative randomized, controlled clinical trial found that after one year's use of HRT among healthy post-menopausal women, their serum estrogen levels increased by 3.6-, 2.7-, and 1.8-fold for E1, E2, and E3, respectively, compared to placebo group at baseline.(68)

At the same time, menstrual and reproductive factors also have significant impact on the breast cancer risk. Studies have shown that for every one year increase in age at menarche, the risk of breast cancer can be reduced from 10% to 20%.(69,70) Age at the birth of first child is positively associated breast cancer risk.(71–73) Higher parity is inversely associated breast cancer risk.(74,75) Older age at menopause is associated with increased risk in breast cancer.(76,77) It has been estimated that for every 5 year increase in age at menopause, the risk of breast cancer elevates by approximately 17%. The association between OC use and breast cancer risk is mixed. A meta-analysis summarizing 54 observational studies reported that current OC users or women who have used it in the past 10 years have about 1.24-fold risk of breast cancer compared to never users.(78) However, there is no significant association observed between OC use and the risk of breast cancer

diagnosed 10 or more years after cessation of use. However, a more recent case-control study found that current or former oral-contraceptive use was not associated with breast cancer risk(79). HRT is also found to be associated with increased breast cancer risk.(80)

4.3. Soybeans or soy products consumption

Soybeans and soy products are one of the major sources of protein intake among Asian populations. Soy foods are rich in isoflavones, which are a class of phytoestrogens that has a very similar chemical structure to E2. As a result, isoflavones can compete with estrogens to bind to estrogen receptors. Although isoflavones can bind to both ERα and ERβ, they are more prone to binding to ERβ(81). Isoflavones have also been shown to have many other protective effects against breast carcinogenesis, including inhibiting the activity of aromatase enzyme and epidermal growth factor receptor tyrosine kinase, modulating eicosanoid metabolism, promoting apoptosis and antioxidant activities, arresting cell cycle progression, suppressing angiogenesis, and regulating gene expression by epigenetic means.(82–84)

Findings from epidemiological studies have showed that soyfood consumption was associated with a reduced risk of breast cancer among pre-menopausal women or Asian population, while the results from studies in western populations have been mixed.(85,86) The heterogeneity of these findings could be due to the fact that the effect of soy foods on breast cancer prevention may depend on some other facts, including low intake level, poor measurements or estrogen levels.

5. The interaction between genetic variants and other estrogen-related factors are not well investigated for breast cancer risk

Limited studies have been conducted to evaluate estrogen-related gene-environment interactions on breast cancer risk. Among those existing studies, most of them were focused on Caucasians. Two early studies observed potential interaction between polymorphisms in gene *CYP17*

and age at menarche on breast cancer risk. Both of these studies show similar results that the association between a later age at menarche and reduced breast cancer risk tends to be only observed in women with specific genetic background of gene *CYP17*.(87,88) In 2005, a nested case-control study from the Multiethnic Cohort Study using over 1339 cases and 1370 controls reported no significant interaction between estrogen metabolism genes (including *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP3A4*, *COMT*, *SULT1A1*) and environmental factors (including age at menarche, BMI, HRT) on breast cancer risk.(89) Similarly, Involving 677 breast cancer cases and 905 controls, a study published in 2007 found no evidence of effect modification of HRT on the association between genetic variants in estrogen metabolism genes (including *COMT*, *CYP1A1*, *CYP1A2*, *CYP1B1*, *SULT1A1*, *SULT1E1*) and breast cancer risk.(90) One year later, BMI and HRT were shown to modulate the association between estrogen biosynthesis gene *CYP17* variant C allele and breast cancer risk from a study with over 1000 cases and 1000 controls.(91) Another study published in the same year found that BMI might modulate the relationship of genetic variants in gene *COMT* and *HSD17B1* with breast cancer risk. However, no formal statistical test had been used to test the significance level of the interaction.(92) At the same time, a study with a relatively small sample size (324 cases and 651 controls) found statistically significant interactions between genetic variants in gene *CYP1A1*and *CYP1B1* and HRT use on influencing breast cancer risk.(93) In 2010, a study with larger sample size (including over 3,000 cases and over 5,000 controls) observed a significant interaction of *CYP17A1* and HRT use on breast cancer risk. With a total sample size of 1,644 cases and 1,451 controls, a study published in 2012 examined the interactions of variations in estrogen metabolism related genes with HRT use on breast cancer. Their analysis suggested that HRT use might modify the associations between SNPs within *CYP1B1* and *CYP17A1* and breast cancer risk.(94) By using a total of 658 cases and 715 controls of European ancestry, a recent study published in 2014 has found that the effect of *ESR1* SNPs on breast cancer risk was modified by life-time exposure to endogenous estrogen and HRT among women of European ancestry. The

association between *ESR1* gene and breast cancer risk was found to be stronger among women with longer life-time estrogen exposure among all European ancestry women. In addition, rs1801132 and rs3020314 in *ESR1* were associated with breast cancer risk only among postmenopausal women, and rs2046210 in *ESR1* only associated with breast cancer among postmenopausal women without HRT use.(95) To our knowledge, only one study has investigated the interactions between isoflavone intake and genetic variants in the risk of breast cancer.(96) The study examined the interaction of five SNPs in gene *ESR1* and *ESR2* with isoflavone intake in breast cancer risk. Their results showed that the protective effect of higher intake of isoflavone-rich foods only exist among postmenopausal women with GG genotype of the rs4986938 in the estrogen receptor beta gene in the Japanese population. In conclusion, the existing studies have only found limited evidence for estrogen-related gene-environment interactions, which include HRT (with *CYP17A1*, *CYP1B1*, and *ESR1*) and soy food intake (wit *ESR2*).

6. Summary

At the current stage, we still lack sufficient and conclusive evidence to establish the associations between genetic variants in estrogen-related genes and breast cancer risk. Furthermore, very limited studies are available to illuminate whether the genetic variants in those genes could modify the association between estrogen-related non-genetic factors and breast cancer risk.
To fill in these important gaps in understanding, I conducted a study evaluating genetically predicted expression levels of estrogen synthesis, metabolism and bioavailability pathway genes with breast cancer risk, including assessing potential interaction effects with known estrogen-related risk factors.

The proposed study has the following specific aims:

Aim 1: To build expression prediction models for estrogen synthesis, metabolism and bioavailability pathway genes using SNP and gene expression data [data from Genotype-Tissue Expression (GTEx)] focusing on breast, liver, adipose, ovaries, and cross tissue models;

Aim 2: To investigate associations of predicted expression levels of estrogen synthesis, metabolism and bioavailability pathway genes with breast cancer risk in both Europeans and Asians;

Aim 3: To conduct a gene-environment interaction study assessing potential interaction of predicted expression levels of estrogen related genes with adiposity [body mass index (BMI), waist circumference, and waist-hip ratio (WHR)], menstrual and reproductive factors [i.e. age at menarche, age at menopause and parity, as well as with exogenous hormone exposure, i.e. oral contraceptive (OC) use, hormone replacement therapy (HRT)], and soy food intake on breast cancer risk in both Europeans and Asians.

We hypothesized that genetic determined expression levels of estrogen-related pathway genes were associated with breast cancer risk, in both Europeans and Asians. Furthermore, the genetically predicted expression levels of these genes may modify the associations of hormonal related factors (menstrual and reproductive history, exogenous hormone exposure), body fat and soy foods intake, with breast cancer risk in these populations.

The proposed study would be the first study to systematically examine the association between genetically predicted expression levels of estrogen-related pathway genes and breast cancer risk. This study would also be the first one to comprehensively investigate the potential effect modification of estrogen-related environmental factors on the association between genetically predicted gene expression levels and breast cancer risk. Our innovative design of aggregating potential regulatory effect of genetic variants into one testing unit – the gene expression and very large sample size hold the promise of overcoming the limitation of low statistical power from previous individual SNP-based studies to great extent. Overall, our results are expected to add new information to the current state of knowledge on this topic and may provide a foundation for directing future studies to better understand the complicated interplay between genetic background and estrogen-related exposures in influencing breast cancer development.

CHAPTER 2

SPECIFIC AIM 1

The goal of this aim is to build expression prediction models for estrogen synthesis, metabolism and bioavailability pathway genes using SNP and gene expression data focusing on breast, liver, adipose, ovaries, and cross tissue models.

1. Methods

1.1. Statistical Analyses

We used the most updated transcriptome and SNP data from the whole-genome sequencing (WGS) (version 7) from the GTEx(97) to build expression prediction models for estrogen related genes. Models for all genes were built in breast, liver, ovaries, subcutaneous adipose, visceral adipose and cross tissues.

In GTEx, whole genome sequencing (WGS) was performed by the Broad Institute's Genomics Platform on DNA samples from 652 GTEx donors at an average coverage of 30X, including 68 sequenced on Illumina HiSeq 2000 using 101-bp paired-end reads, and 584 samples on Illumina HiSeq X using 151-bp paired-end reads, all using blood sample DNA. Genotype data were processed according to the GTEx protocol (https://www.gtexportal.org/home/documentationPage). In brief, BAMs were processed using Picard pipeline. With BWA-MEM, the reads were aligned to the human reference genome build hg19/GRCh37. GATK's HaplotypeCaller v3.4 was further used for joint variant calling across all subjects. Genotype posterior probabilities were then calculated for all calls based on allele frequency in 1000 Genomes Project Phase 3 version 1, and genotype quality (GQ) scores were updated using GATK's CalculateGenotypePosteriors. Regarding subject QC, replicates, Klinefelter subjects, a chromosome 17 trisomy subject, and an individual from a related pair were

excluded, resulting 635 subjects remained, including 406 males and 229 females. Variant QC included Variant Quality Score Recalibration (VQSR) filtering; removal of variants in low-complexity regions or with inbreeding coefficient ≤ -0.3; assigning calls with allelic imbalance (AB) < 0.2 or > 0.8 and/or GQ < 20 to missing; excluding variants with genotyping call rate < 85%, with HWE P-value < 1×10^{-6} (using European samples only and females for chrX), associated with library construction batch or sequencing technology ($P < 1\times10^{-8}$), or with heterozygous haploid genotypes in non-pseudoautosomal regions of sex chromosomes in males. These yielded overall 10,526,813 variants at MAF ≥ 1%.

GTEx genotype data were downloaded from dbGap in the format of a vcf file, which has gone through the standard quality control (QC) as described above. Only SNPs with minor allele frequency ≥ 5%, biallelic, and those that are not strand ambiguous were retained for use. Genotypes were encoded as a continuous variable ranging from 0 to 2, with the value representing the estimated count of the effect allele.

In GTEx, RNA-seq was performed using the Illumina TruSeq library construction protocol (non-stranded, polyA+ selection), generating 76-bp paired-end reads. Gene-level expression was quantified as Transcripts Per Kilobase Million (TPM) using RNA-SeQC v1.1.8(98). RNA-seq expression outliers were identified and excluded using a multidimensional extension of the statistic(99). Genes with a median expression level of 0 TPM across samples were removed. We normalized the TPM values to the average empirical distribution observed across samples. Then we quantile normalized the TPM values to a standard normal distribution for each gene across samples. The gene expression data were adjusted for covariates including sex, top 3 genotyping principal components (PCs), sequencing platform and probabilistic estimation of expression residuals (PEER) factors.

The GTEx v7 data contains 620 participants with both genotype and gene expression data available from at least one of 48 different tissues. Among them, 34.2% are females. Considering the

recognized sex difference for estrogen, in our study we focused on data generated from females. The sample size for each specific tissue that were used in our study is presented in Table 1. Among them, most were with European ancestry.

All expression prediction models were built by using the elastic net method as implemented in the glmnet R package, with alpha=0.5. The genetically regulated expression for each gene was estimated by including variants within the 2 MB flanking region of each gene. Ten-fold cross-validation were used to validate the models internally. Prediction R^2 values (the square of the correlation between predicted and observed expression) were generated to estimate the prediction performance of each of the gene prediction models established.

Table 1. Number of sample sizes for gene expression model building from different tissues.

Tissue	**Sample Size**	
	European women	Asian women
Breast	85	3
Liver	47	2
Ovaries	99	2
Visceral adipose	94	1
Subcutaneous adipose	121	2
Cross-tissue (include the above 5 tissues)	179	3

1.2. Model building

Our research group at Vanderbilt has generated whole genome RNA-seq data of adjacent normal breast tissue obtained from 200 subjects diagnosed with benign breast diseases or breast cancer recruited to the Shanghai Breast Cancer Study (all women of East Asian ancestry) using Illumina HiSeq. Of them, 151 subjects also had genetic data available (genotyped using MEGA or

GWAS arrays). The detailed information for genetic data QC and imputation is described in the next Chapter. The raw RNA-seq reads were processed using a previously described pipeline[19]. The sequencing reads were mapped to human genome (hg19) using the Bowtie2 tool[20]. The mapped reads were then used to determine expression for all coding genes and noncoding RNAs by Cufflinks (version 2.2.1)[21]. The Gencode (release 27) was used to annotate coding genes and noncoding RNAs[22]. The fragments per kilobase of transcript per million mapped reads (FPKM) for expression of each gene was then determined. We further performed a series of QC and normalization. Specifically, genes with a median expression level of 0 FPKM across samples were removed, and the FPKM values of each gene were log2 transformed. The quantile normalization was performed to bring the expression profile of each subject to the same scale, and inverse quantile normalization was performed for each gene to map each set of expression values to standard normal. We adjusted for the top ten principal components derived from genotype data, the top 30 probabilistic estimation of expression residual factors to correct for batch effects and experimental confounders in model building, and status of subject (benign disease or breast cancer patients) for model building using data from all 151 subjects. Furthermore, we built another set of models using data focusing only on the 65 subjects with benign breast diseases, for which we adjust for the top 15 probabilistic estimation of expression residual factors to correct for batch effects and experimental confounders in model building.

1.3. Utility

We also evaluate performance of the models built in the European descents in Asians by comparing the predicted gene expression level for each gene with the expression level measured by RNAseq in the 151 Asian breast tissue sample from 151 subjects using the Spearman's correlation. Models showed a performance at $R^2 \geq 0.01$ in GTEx dataset and correlation $R > 0.1$ in the in Asian

were deemed reliable models for predicting expression in Asians and were used in association analyses with breast cancer risk in Asians.

1.4. Power calculation

We performed a simulation analysis estimating the model prediction performance that can be achieved according to the varying sample size of the reference dataset for model building and the cis-heritability (h2) which we aim to capture using gene expression prediction models (R2) (Figure 4). Based on a simulation analysis, it seems the performance of prediction models does not change very much when the reference dataset achieves a sample size of 100 or higher, in each scenario of the cis-heritability (h2). Our study thus should have reasonable power for gene expression prediction model building for both Caucasian and Asians considering that the available reference datasets for the relevant tissues (except for liver) involve a sample size close in scale to the 100 threshold, if not higher (see Table 1 above).

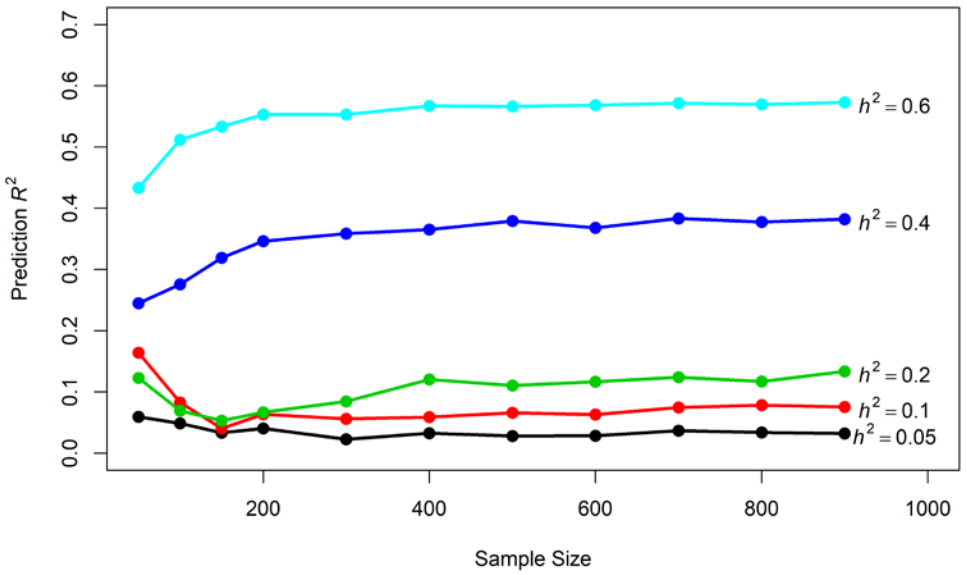


Figure 4. The performance of gene expression prediction models based on different sample size in each scenario of cis-heritability

2. Results

The distributions of the gene expression levels were shown in supplementary figure 1-5 for ovary, liver, breast, subcutaneous adipose and visceral adipose tissues, respectively. Only genes with a median expression greater than 0 were included. Since quite a number of the distribution were skewed, normalization for the gene expression levels were conducted before the prediction model building.

Genes that showed a prediction performance (R^2) of at least 0.01 (0.1 correlation between predicted and observed gene expression levels) were presented for models that were built for each tissue among only European descendants or both European and Asian descendants.

Of the ovary tissue prediction models built for 74 genes, 21 showed satisfactory prediction ($R^2 \geq 0.01$) performance. Genes showed the best prediction performance were SULT1C2 (R2 = 0.38), SULT1A2 ($R^2 = 0.25$), and HSD17B14 ($R^2 = 0.22$). The first two SULT genes are in the estrogen metabolism pathway, and HSD17B14 is in the estrogen synthesis pathway. European-only model predicted 5 more genes with an R2 greater than 0.01 compared to model including both Europeans and Asians. (Table 2)

For models that were built for liver tissue, the number of genes that showed good prediction performance were 26 among Europeans. Gene *ARSB* ($R^2 = 0.38$) in the estrogen metabolism pathway showed the best performance. (Table 3)

Regarding models built in breast tissue, estrogen metabolism pathway genes *SULT1C2* ($R^2 = 0.51$) and *UGT2B7* ($R^2 = 0.33$) showed the best prediction performance for either exclusive Europeans or Europeans and Asians. 29 genes showed satisfactory prediction performance. Two additional genes were built for European-only model compared to the mixed-race model. (Table 4)

Genes showed very distinct prediction performance between the two sets of models for subcutaneous adipose. Thirteen extra genes were successfully built in the Europeans. Genes with best prediction performance were also very different between the two models. For European only

model, genes showed the best performance were *HSD17B11* and *SULT1A4*, which were in estrogen synthesis- and metabolism- pathways, respectively. For models included Asians, two estrogen-metabolism pathway genes, *SULT1C2* and *GSTM1* showed the best prediction performance. (Table 5)

Similar results were observed for the two models build in visceral adipose tissue, with 24 and 23 genes showing a prediction performance greater than 0.01 for the European only and Asian included models, respectively. The top 4 genes showing the best prediction performance included *SULT1C2*, *SULT1A2*, *UGT2B4* and *ARSF*, which were all in the estrogen-metabolism pathway. (Table 6)

For the cross-tissue models, information of the gene expression levels across all 38 tissues collected in GTEx was used. The prediction performances were similarly good between the two sets of models. The top two genes showed the best prediction performance were *SULT1C2* and *SULT1A2*, which were the same as the visceral adipose tissue model. (Table 7)

Table 2. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in ovary tissue.

Ovary (only Europeans) N=99	R^2	Ovary (Europeans and Asians) N=101	R^2
SULT1C2	0.377	*SULT1C2*	0.358
SULT1A2	0.253	*HSD17B14*	0.214
HSD17B14	0.221	*SULT1A2*	0.190
ARSA	0.174	*CYP11A1*	0.189
CYP11A1	0.129	*ARSA*	0.177
CYP1A1	0.124	*SULT1A1*	0.154
SULT1A1	0.111	*ARSJ*	0.120
ARSJ	0.102	*GSTM1*	0.092
IDS	0.077	*ARSD*	0.063
CYP1B1	0.077	*UGT2B7*	0.054
HSD17B1	0.071	*IDS*	0.052
B3GAT2	0.061	*ARSB*	0.049
ARSD	0.045	*HSD17B1*	0.045
CYP3A4	0.045	*STS*	0.039

UGT2B7	0.045	*ESR1*	0.032
CYP19A1	0.045	*CYP1A1*	0.028
GNS	0.045	*CYP3A4*	0.025
GSTM1	0.045	*GNS*	0.019
ARSB	0.045	*ARSE*	0.014
ARSG	0.045	*B3GAT2*	0.012
ARSE	0.045	*SHBG*	0.011

Table 3. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in liver tissue.

Liver (only Europeans) N=47	R^2	Liver (Europeans and Asians) N=49	R^2
ARSB	0.383	*ARSB*	0.559
HSD3B2	0.251	*UGT2B17*	0.413
UGT2B17	0.242	*UGT1A6*	0.203
CYP11A1	0.217	*UGT2B4*	0.168
UGT1A6	0.215	*HSD3B2*	0.154
ARSE	0.153	*UGT2B28*	0.131
SULF2	0.140	*UGT1A4*	0.112
UGT1A4	0.140	*SULF2*	0.092
AKR1B15	0.134	*ARSG*	0.079
UGT2B4	0.115	*ARSA*	0.078
CYP1B1	0.113	*ARSJ*	0.077
UGT2B28	0.094	*ESR2*	0.073
B3GAT1	0.086	*ARSE*	0.069
B3GAT2	0.086	*IDS*	0.066
SULT4A1	0.085	*STAR*	0.065
ESR2	0.081	*HSD17B1*	0.064
STAR	0.077	*UGT1A7*	0.063
CYP1A2	0.069	*CYP1B1*	0.059
ARSA	0.064	*B3GAT2*	0.050
SULT1A1	0.049	*SULT4A1*	0.042
SULT2B1	0.040	*B3GAT3*	0.038
GSTA1	0.040	*SULT1A1*	0.017
SGSH	0.040	*CYP1A2*	0.016
GSTM1	0.029	*SULT1C2*	0.011
COMT	0.028		
HSD17B11	0.025		
B3GAT3	0.019		
HSD17B14	0.015		
ARSJ	0.011		

Table 4. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in breast tissue.

Breast (only Europeans) N=85	R^2	Breast (Europeans and Asians) N=88	R^2
SULT1C2	0.509	*SULT1C2*	0.449
UGT2B7	0.326	*UGT2B7*	0.298
UGT2B17	0.249	*GSTA1*	0.255
SULT1A2	0.226	*SULT1A2*	0.253
GSTM1	0.212	*GSTM1*	0.251
SULT2B1	0.208	*SGSH*	0.244
CYP1A2	0.178	*UGT2B17*	0.197
IDS	0.164	*SULT2B1*	0.166
ARSE	0.154	*ARSE*	0.165
SULT4A1	0.137	*SULT4A1*	0.159
GSTA1	0.120	*COMT*	0.146
ARSA	0.112	*B3GAT2*	0.119
B3GAT2	0.106	*HSD3B2*	0.111
CYP1A1	0.102	*IDS*	0.111
SGSH	0.080	*ARSA*	0.093
GALNS	0.067	*CYP1A1*	0.070
B3GAT3	0.062	*GALNS*	0.062
HSD3B2	0.062	*ARSD*	0.044
UGT2B11	0.051	*B3GAT3*	0.042
HSD17B1	0.046	*ARSB*	0.040
CYP3A4	0.045	*CYP3A4*	0.034
SULT1B1	0.039	*HSD17B1*	0.032
ARSK	0.030	*SULT1A1*	0.030
CYP11A1	0.026	*HSD17B2*	0.022
COMT	0.023	*SULT1C4*	0.021
ARSF	0.020	*CYP19A1*	0.019
SHBG	0.015	*UGT2B11*	0.010
ARSB	0.014		
CYP3A5	0.012		

Table 5. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in subcutaneous adipose tissue.

Subcutaneous adipose (only Europeans) N=121	R^2	Subcutaneous adipose (Europeans and Asians) N=123	R^2
HSD17B11	0.344	*SULT1C2*	0.510
SULT1A4	0.204	*GSTM1*	0.267

SULF2	0.192	*STS*	0.205
B3GAT2	0.181	*SULT1C4*	0.158
AKR1B15	0.178	*HSD17B14*	0.155
STAR	0.135	*SULT1A2*	0.133
HSD3B2	0.118	*HSD17B1*	0.124
SULT1A1	0.113	*B3GAT1*	0.121
NQO1	0.111	*ARSA*	0.117
CYP1B1	0.099	*SULT2B1*	0.089
GNS	0.093	*GALNS*	0.075
GALNS	0.060	*SULT1A4*	0.069
SULT4A1	0.027	*SULT1B1*	0.057
ARSG	0.018	*GNS*	0.056
		ARSE	0.054
		UGT2B4	0.053
		SULT1A3	0.051
		HSD3B2	0.040
		SULT1A1	0.035
		ARSB	0.034
		SULF1	0.034
		CYP1A2	0.032
		SULF2	0.030
		HSD17B11	0.019
		ARSG	0.015
		SULT1E1	0.014
		ESR2	0.012

Table 6. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in visceral adipose tissue.

Visceral adipose (only Europeans) N=94	R^2	Visceral adipose (Europeans and Asians) N=95	R^2
SULT1C2	0.598	*SULT1C2*	0.556
SULT1A2	0.346	*SULT1A2*	0.311
UGT2B4	0.302	*ARSF*	0.268
ARSF	0.284	*UGT2B4*	0.261
GSTM1	0.246	*GSTM1*	0.206
ARSA	0.193	*CYP3A4*	0.151
CYP1A1	0.146	*HSD17B14*	0.096
SULT2B1	0.145	*HSD17B1*	0.095
HSD17B14	0.124	*SULT1B1*	0.089
HSD17B1	0.096	*COMT*	0.085
CYP1B1	0.077	*SULT2B1*	0.073
SULT1B1	0.071	*ARSA*	0.064

CYP3A4	0.070	*CYP1B1*	0.057
ESR1	0.066	*SULF2*	0.056
COMT	0.065	*ESR2*	0.044
SULT1E1	0.034	*ESR1*	0.038
ESR2	0.024	*SULT1A3*	0.035
ARSD	0.019	*SULF1*	0.032
IDS	0.016	*ARSK*	0.030
ARSJ	0.011	*ARSE*	0.030
SULF1	0.011	*HSD17B2*	0.024
SULF1	0.011	*IDS*	0.021
SULF2	0.010	*GNS*	0.018
SULT4A1	0.010		

Table 7. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in cross tissue model

Cross tissue (only Europeans) N=195	R^2	Cross tissue (Europeans and Asians) N=198	R^2
SULT1C2	0.600	*SULT1C2*	0.559
SULT1A2	0.420	*SULT1A2*	0.407
SULT2B1	0.380	*SULT2B1*	0.357
GSTM1	0.292	*GSTM1*	0.299
ARSA	0.236	*ARSA*	0.212
UGT2B17	0.205	*UGT2B17*	0.191
HSD17B14	0.183	*HSD17B14*	0.170
CYP17A1	0.147	*CYP11A1*	0.126
SULT1A1	0.096	*SULT1A1*	0.103
CYP11A1	0.080	*GALNS*	0.066
AKR1B15	0.080	*CYP1B1*	0.062
CYP1B1	0.073	*SULT1E1*	0.060
UGT2B7	0.061	*UGT2B7*	0.055
B3GAT1	0.052	*AKR1B15*	0.053
SULT1E1	0.045	*CYP3A4*	0.044
HSD3B2	0.040	*ARSI*	0.038
CYP3A4	0.039	*B3GAT1*	0.038
ARSF	0.036	*SULT1A4*	0.036
GALNS	0.035	*SULF2*	0.032
SULF2	0.029	*HSD17B2*	0.031
ARSI	0.027	*ARSK*	0.029
SULT1A3	0.024	*CYP3A5*	0.028
SULT4A1	0.024	*HSD3B2*	0.025
STS	0.022	*STS*	0.024
GSTA1	0.021	*SULT4A1*	0.024

ARSD	0.019	*CYP19A1*	0.017
SULF1	0.019	*SULF1*	0.017
HSD17B2	0.016	*GNS*	0.016
STAR	0.016	*ARSD*	0.014
GNS	0.012		
SULT1A4	0.010		

For Asian breast tissue models built, the set of models using data of the 65 begin breast disease subjects showed better performance overall, compared with the set of models using data of the 151 subjects with tissue from either begin breast disease or breast cancer patients. We thus focused on using the models built using the 65 tissues from the begin breast disease subjects. For the other genes with no satisfactory model built using this dataset, we checked performance of the prediction models built using GTEx subjects of both European and Asian breast tissue expression data to predict expression in Asians. Models for four of the genes showed an $R^2 \geq 0.1$ in the Asian dataset (*CYP19A1*, *GALNS*, *ARSA*, *B3GAT2*). These four models plus other 15 genetic models built using Asian data were analyzed for their association with breast cancer risk in Asians. (Table 8)

Table 8. Prediction models built with an $R^2 \geq 0.01$ for estrogen-related pathway genes in breast tissue model using samples from Shanghai, China.

Breast tissue (adjacent normal and benign disease) N=151	R^2	Breast tissue (benign disease only) N=65	R^2
GSTM1	0.196	*GSTM1*	0.389
SULT1C2	0.099	*AKR1B15*	0.236
SULT4A1	0.079	*SULT4A1*	0.145
UGT2B28	0.077	*SULT1C2*	0.141
COMT	0.062	*SULT1A2*	0.119
CYP3A5	0.054	*CYP3A4*	0.064
ESR2	0.037	*UGT2B28*	0.031
SGSH	0.029	*GSTA1*	0.025
SULT2B1	0.027	*UGT2B15*	0.024
SULT1A4	0.022	*CYP17A1*	0.019
B3GAT2	0.021	*SULT1A4*	0.018
ARSI	0.014	*HSD17B1*	0.018

		ARSK	0.016
		SULF1	0.012
		ESR2	0.011

To check the performance of the prediction models we built among European descendants. We have also tried to use the models we built in Europeans to predict breast cancer patients' expression level using TCGA data from breast tissues. Table 9 shows the results for the correlation between predicted and measured expression level in TCGA. Among a total of 24 GTEx built breast tissue gene prediction models using data of Europeans and Asians with a good prediction performance, 5 showed an external performance correlation greater than 0.1 by using the TCGA data. 16 of the genes showed a correlation greater than 0, and 8 of these genes showed a negative correlation in the external validation analysis. (Table 9)

Table 9. Correlation between the predicted and measured expression level using TCGA data in breast tissue.

Gene name	Spearman correlation coefficient
ARSK	0.249
CYP1A1	0.172
HSD3B2	0.153
SHBG	0.110
GALNS	0.107
B3GAT3	0.090
GSTA1	0.075
ARSA	0.074
CYP1A2	0.057
HSD17B1	0.053
CYP3A5	0.041
ARSB	0.036
GSTM1	0.035
UGT2B7	0.029
SULT1C2	0.017
SULT2B1	0.016
CYP11A1	-0.015
B3GAT2	-0.051
CYP3A4	-0.063
SGSH	-0.065

SULT4A1	-0.081
COMT	-0.186
UGT2B11	-0.190
SULT1A2	-0.241

To check the performance of our prediction model. We also using the breast tissue model built among European and Asian in GTEx to predict the expression level in breast tissue in Shanghai Breast Cancer Study. Table 10 shows the results for the correlation between the predicted and measured expression in patients with benign breast disease (BBD) in Shanghai. Among a total of 28 genes, 8 of the genes have a correlation greater than 0.1. 15 of the genes have a correlation greater than 0, and 13 of the genes have a negative association. (Table 10)

Table 10. Correlation between the predicted and measured expression level using Shanghai BBD data in breast tissue.

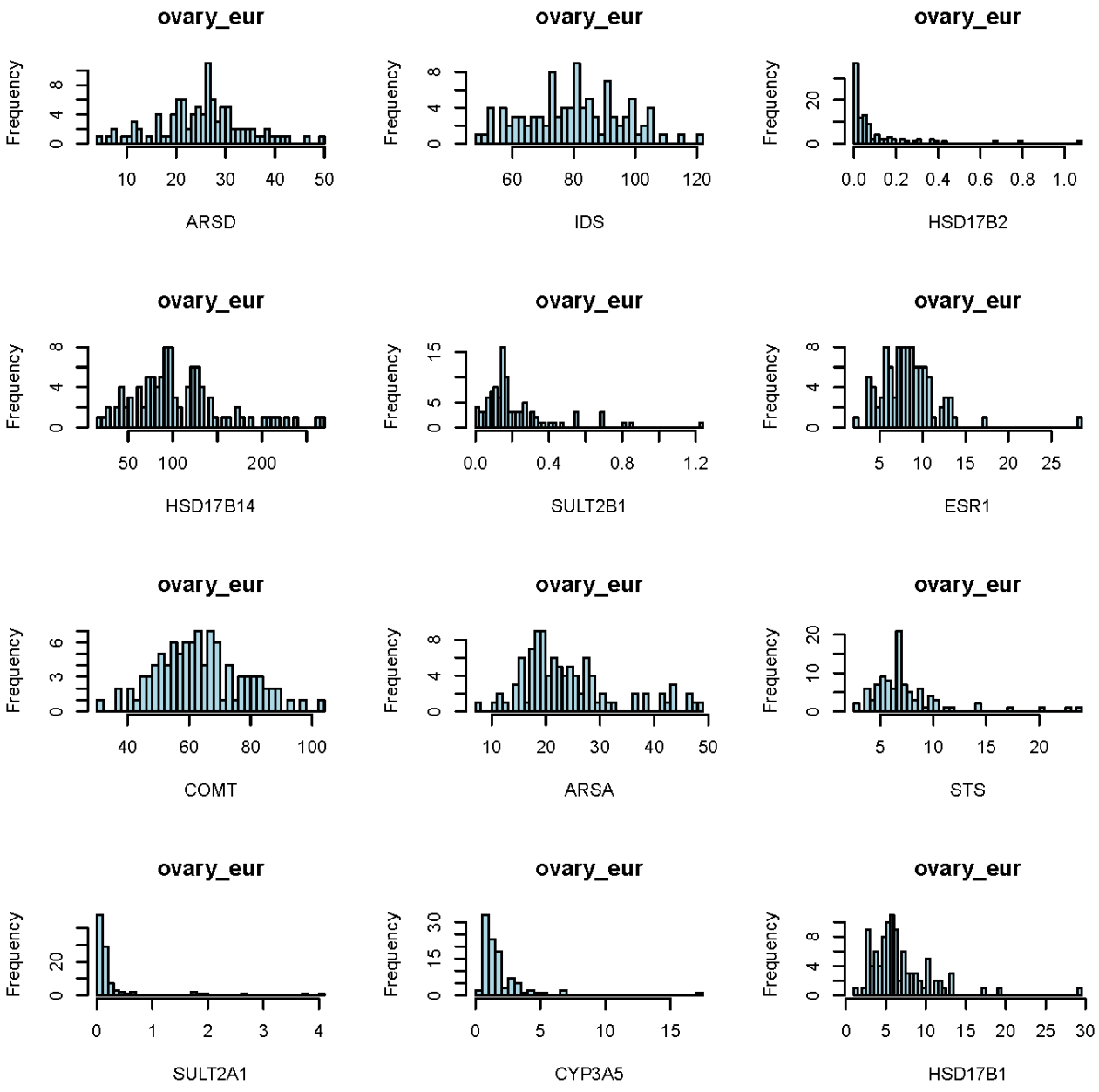
Gene name	Spearman correlation coefficient
SULT1C2	0.429
GSTM1	0.399
ARSA	0.198
SULT1A2	0.156
GALNS	0.141
B3GAT2	0.135
CYP19A1	0.104
GSTA1	0.102
CYP3A5	0.089
SGSH	0.073
ARSB	0.057
SULT1B1	0.052
B3GAT1	0.025
SULT1A1	0.015
CYP1B1	0.005
UGT2B28	-0.010
UGT2B17	-0.026
SULT2B1	-0.034
HSD17B1	-0.044
SHBG	-0.091
STAR	-0.104
UGT2B11	-0.122

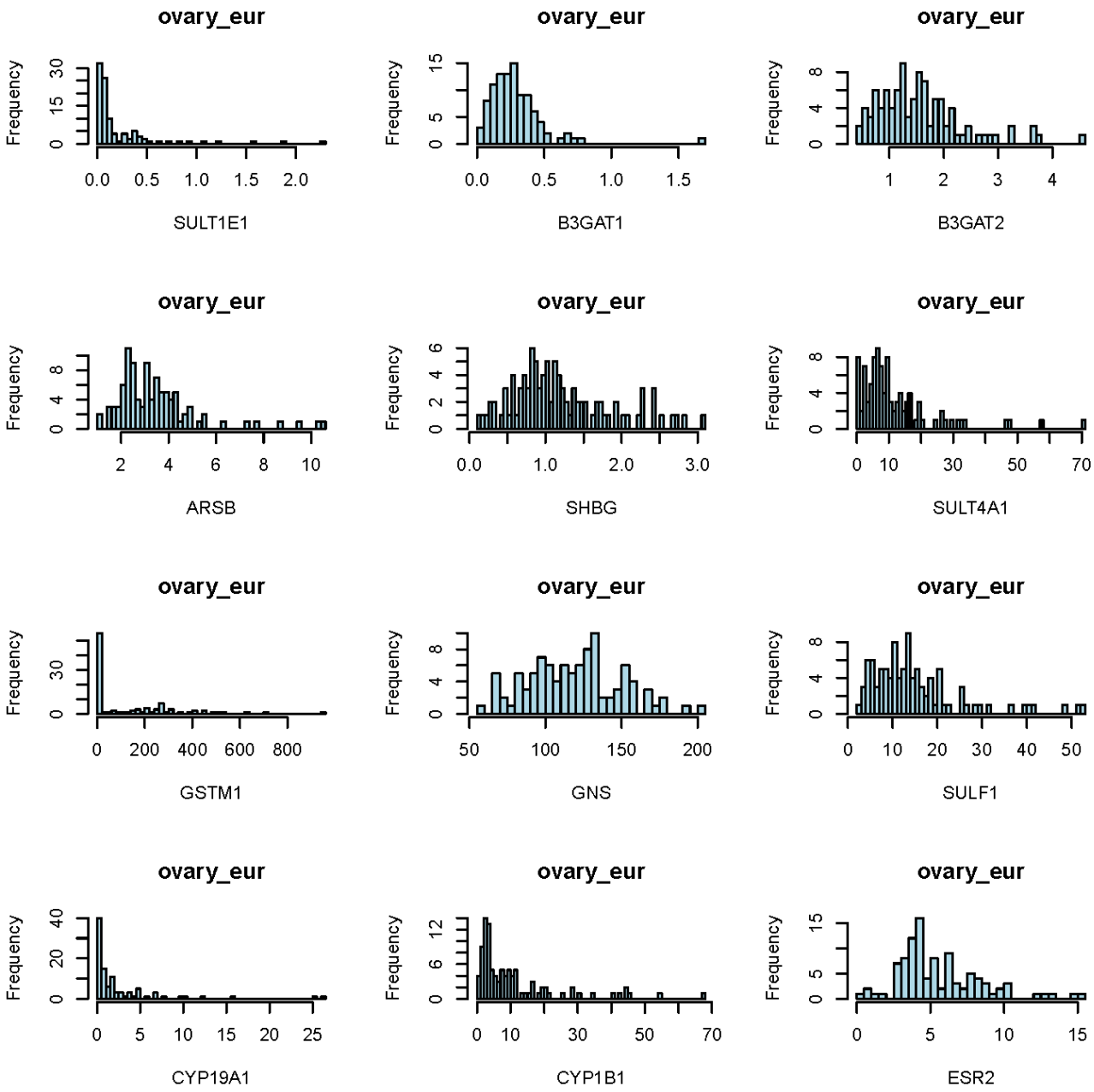
SULT1C4	-0.150
SULT4A1	-0.156
CYP3A4	-0.195
HSD17B2	-0.201
COMT	-0.246
B3GAT3	-0.255

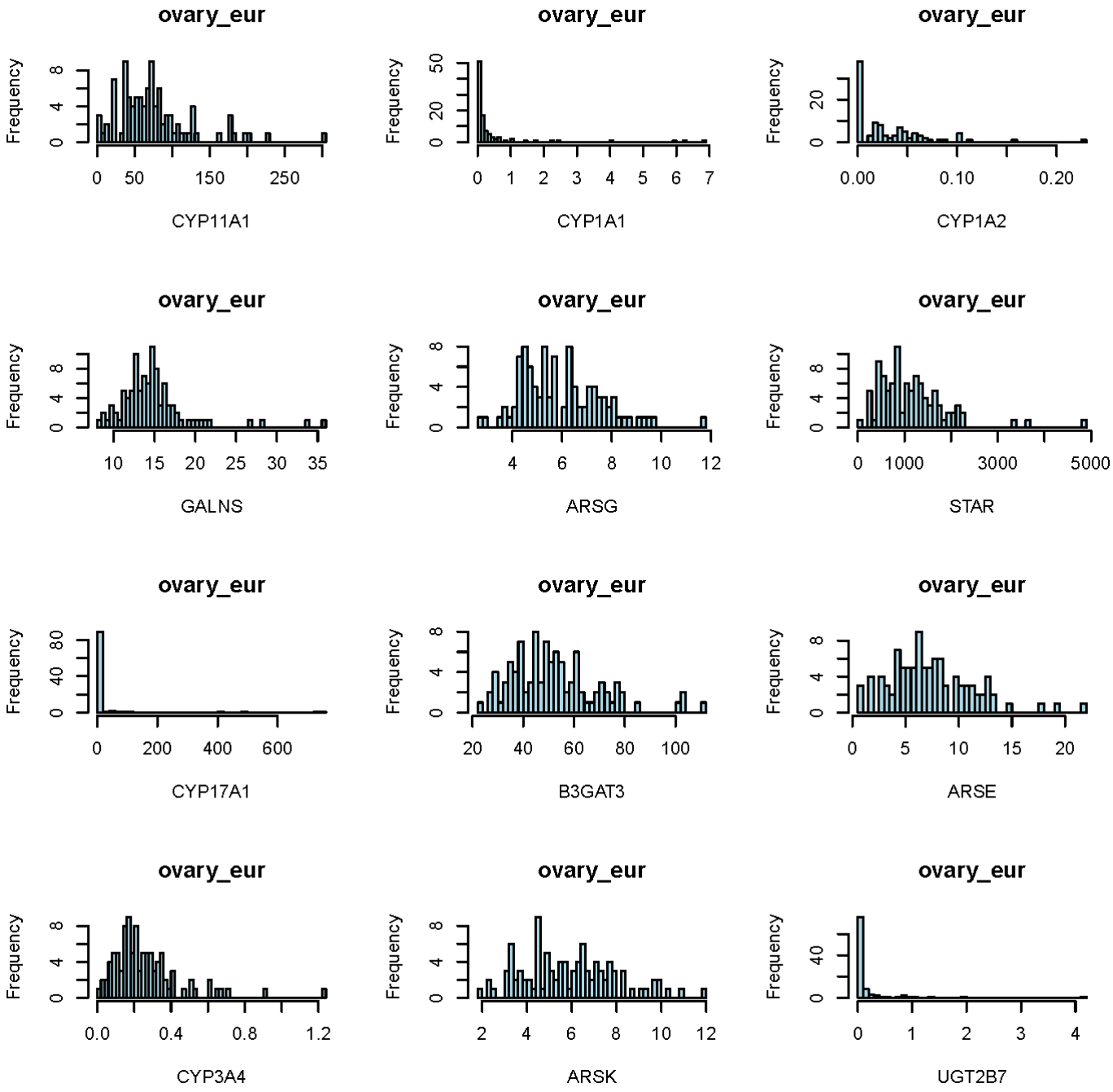
For breast tissue gene expression prediction models, by using GTEx data, there were models for 29 genes showing a prediction performance R^2 of at least 0.01 by using data of 85 Europeans. Using data of 85 Europeans and 3 Asians, there were models for 27 genes showing a prediction performance R^2 of at least 0.01. In comparison, by using data focusing on 65 East Asian begin breast disease subjects, there were only 15 genes showing a prediction performance R^2 of at least 0.01. When using data of 65 East Asian begin breast disease subjects and 86 breast cancer patients, the number of models built with R^2 of at least 0.01 was 12.

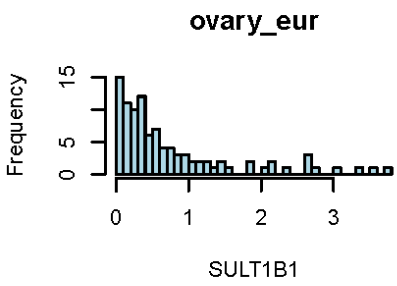
We evaluated performance of GTEx built breast tissue models in TCGA dataset focusing on tumor-adjacent normal tissue, as well as external performance of GTEx built breast tissue models in Shanghai BBD dataset. As expected, a large proportion but not all (67%) of tested models built using GTEx data showed positive correlation in external validation using TCGA data. It is known that some of the gene expression might have changed in the tumor-adjacent normal tissues such as those in TCGA, and thus it is anticipated that some genes may not show high prediction performance in TCGA data due to potential influence of tumor growth (100,101). In Asian women with BBD, only 8 genes showed correlation of above 0.1.

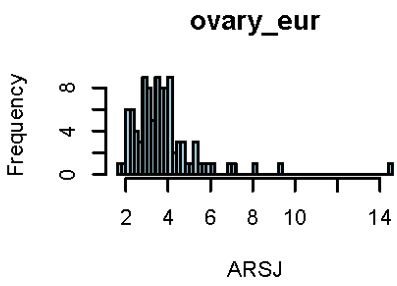
Figure 5. Distribution of gene expression levels for all estrogen-related genes in ovary tissue among European descendants.

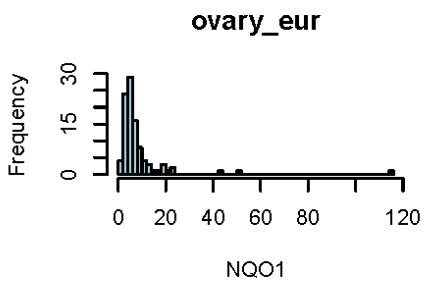


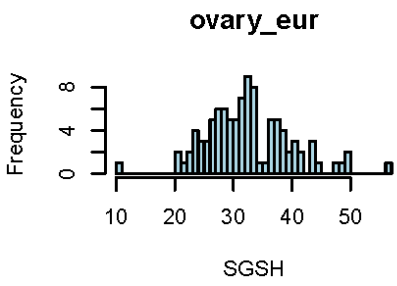

ovary_eur
Frequency
SULT1E1
ovary_eur
Frequency
B3GAT1
ovary_eur
Frequency
B3GAT2
ovary_eur
Frequency
ARSB
ovary_eur
Frequency
SHBG
ovary_eur
Frequency
SULT4A1
ovary_eur
Frequency
GSTM1
ovary_eur
Frequency
GNS
ovary_eur
Frequency
SULF1
ovary_eur
Frequency
CYP19A1
ovary_eur
Frequency
CYP1B1
ovary_eur
Frequency
ESR2

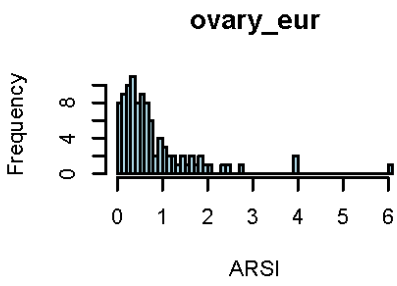

ovary_eur
Frequency
0 4 8
0 50 150 250
CYP11A1
ovary_eur
Frequency
0 20 50
0 1 2 3 4 5 6 7
CYP1A1
ovary_eur
Frequency
0 20
0.00 0.10 0.20
CYP1A2
ovary_eur
Frequency
0 4 8
10 15 20 25 30 35
GALNS
ovary_eur
Frequency
0 4 8
4 6 8 10 12
ARSG
ovary_eur
Frequency
0 4 8
0 1000 3000 5000
STAR
ovary_eur
Frequency
0 40 80
0 200 400 600
CYP17A1
ovary_eur
Frequency
0 4 8
20 40 60 80 100
B3GAT3
ovary_eur
Frequency
0 4 8
0 5 10 15 20
ARSE
ovary_eur
Frequency
0 4 8
0.0 0.4 0.8 1.2
CYP3A4
ovary_eur
Frequency
0 4 8
2 4 6 8 10 12
ARSK
ovary_eur
Frequency
0 40
0 1 2 3 4
UGT2B7

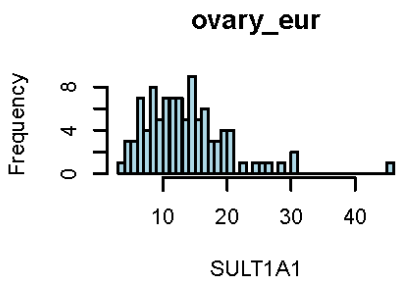

ovary_eur
Frequency
0
5
15
0
1
2
3
SULT1B1

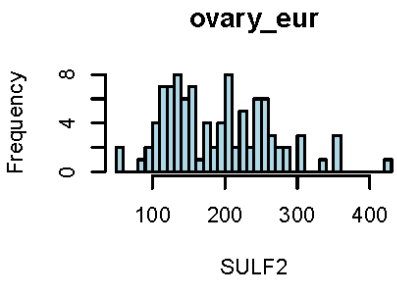

ovary_eur
Frequency
0
4
8
2
4
6
8
10
14
ARSJ

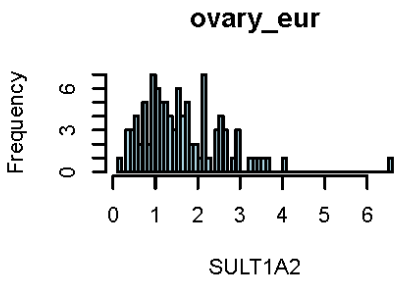

ovary_eur
Frequency
0
15
30
0
20
40
60
80
120
NQO1

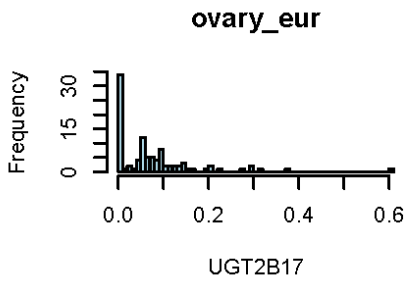

ovary_eur
Frequency
0
4
8
10
20
30
40
50
SGSH

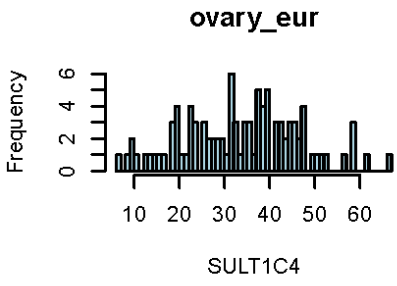

ovary_eur
Frequency
0
4
8
0
1
2
3
4
5
6
ARSI

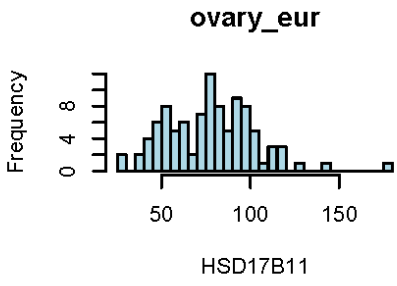

ovary_eur
Frequency
0
4
8
10
20
30
40
SULT1A1

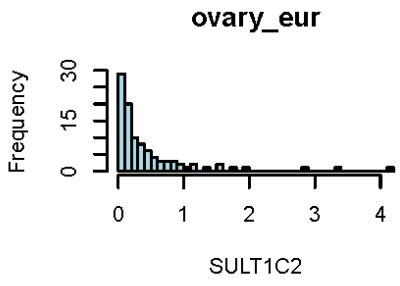

ovary_eur
Frequency
0
4
8
100
200
300
400
SULF2

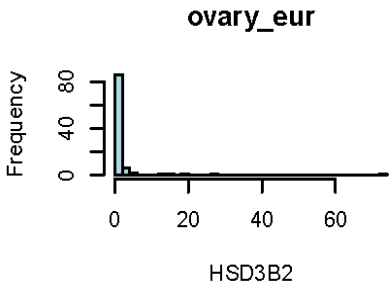

ovary_eur
Frequency
0
3
6
0
1
2
3
4
5
6
SULT1A2

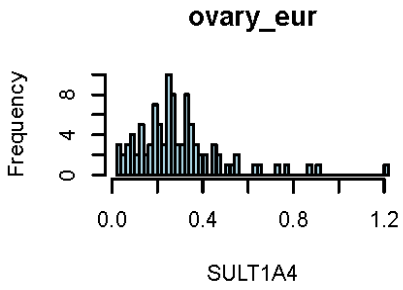

ovary_eur
Frequency
0
15
30
0.0
0.2
0.4
0.6
UGT2B17

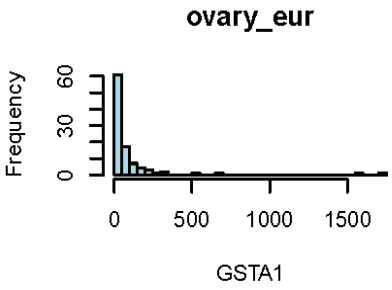

ovary_eur
Frequency
0
2
4
6
10
20
30
40
50
60
SULT1C4


ovary_eur
Frequency
0
4
8
50
100
150
HSD17B11


ovary_eur
Frequency
0
15
30
0
1
2
3
4
SULT1C2


ovary_eur
Frequency
0
40
80
0
20
40
60
HSD3B2


ovary_eur
Frequency
0
4
8
0.0
0.4
0.8
1.2
SULT1A4


ovary_eur
Frequency
0
30
60
0
500
1000
1500
GSTA1

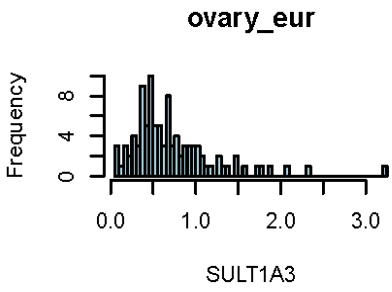
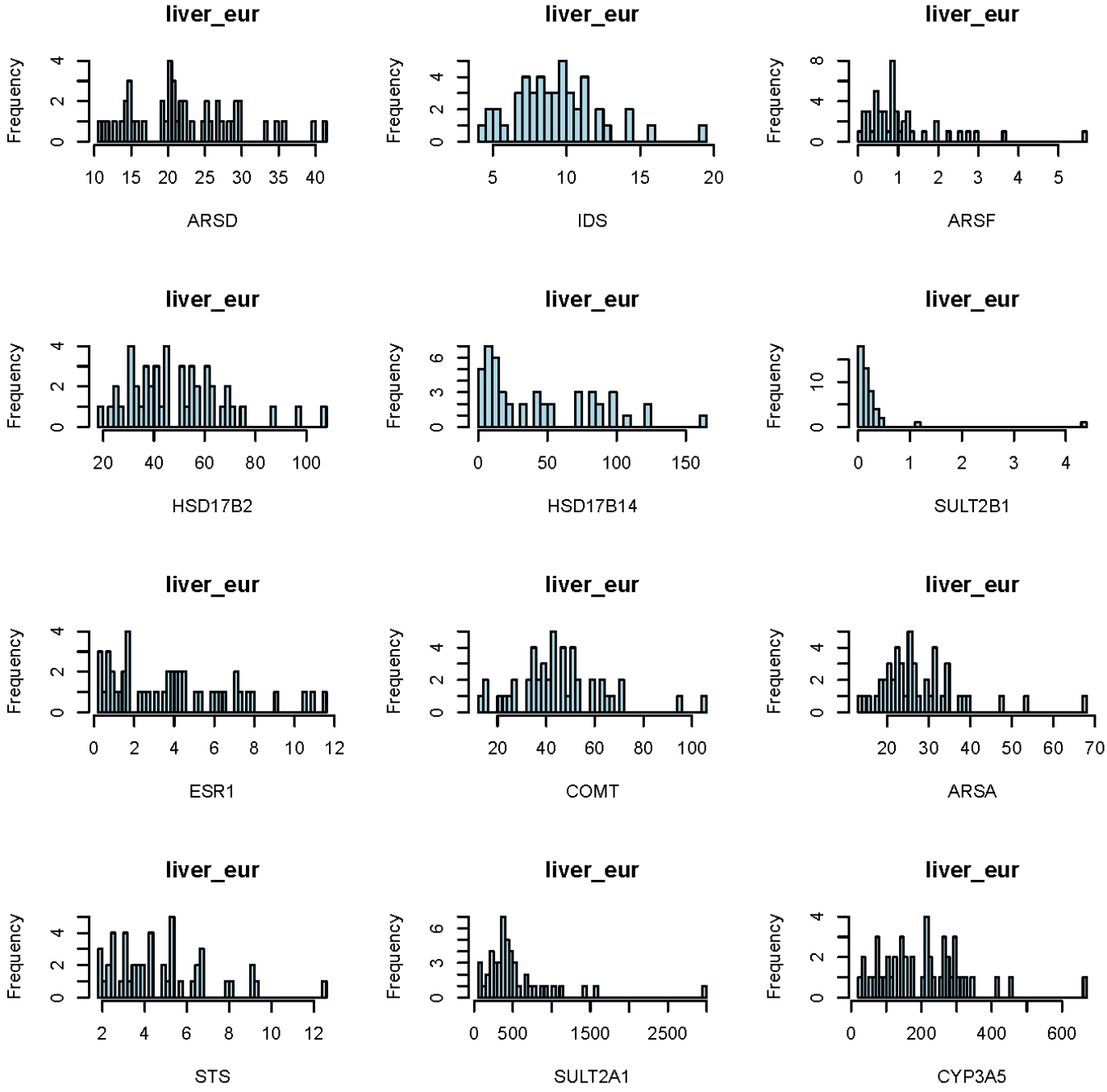
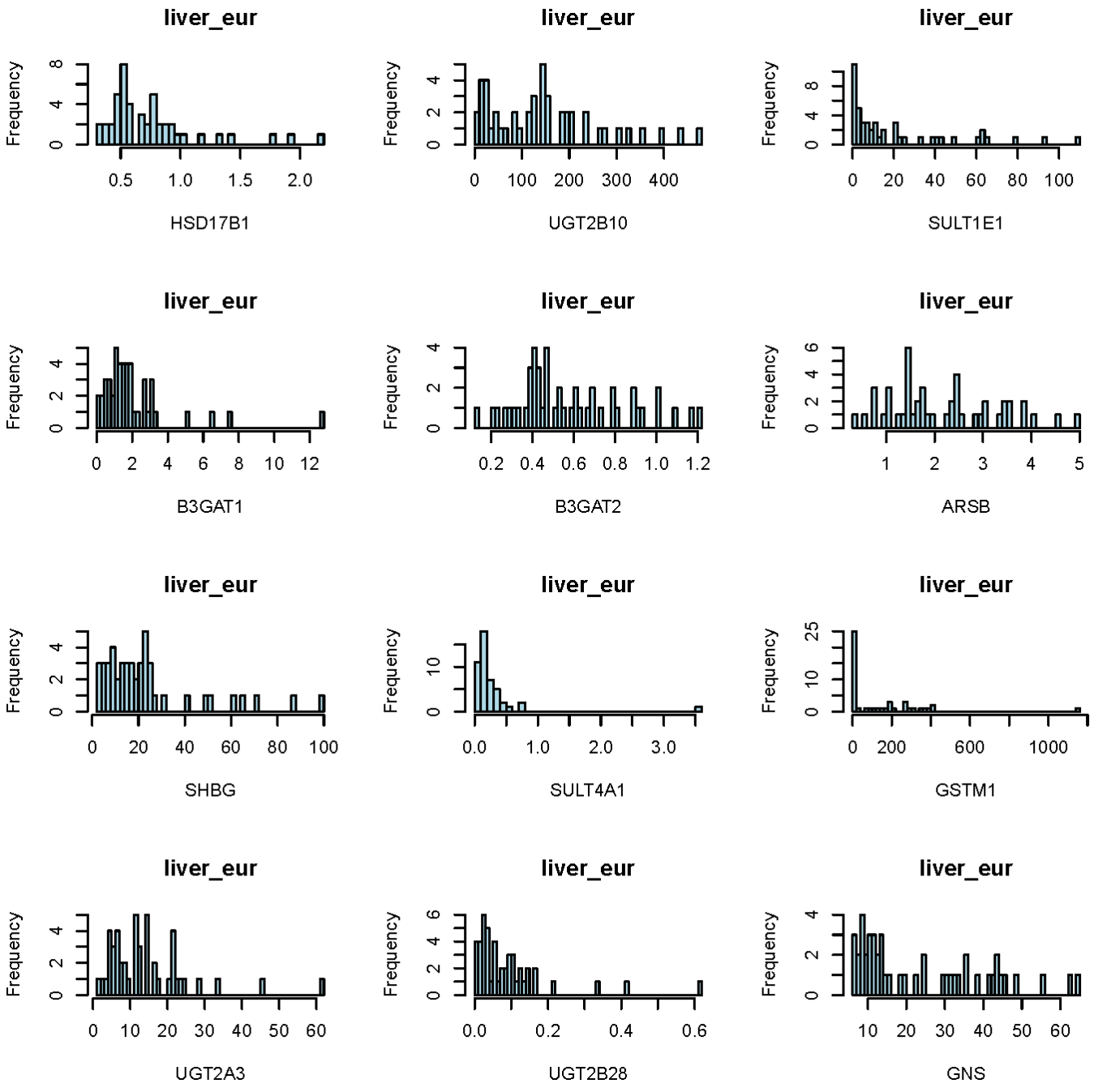
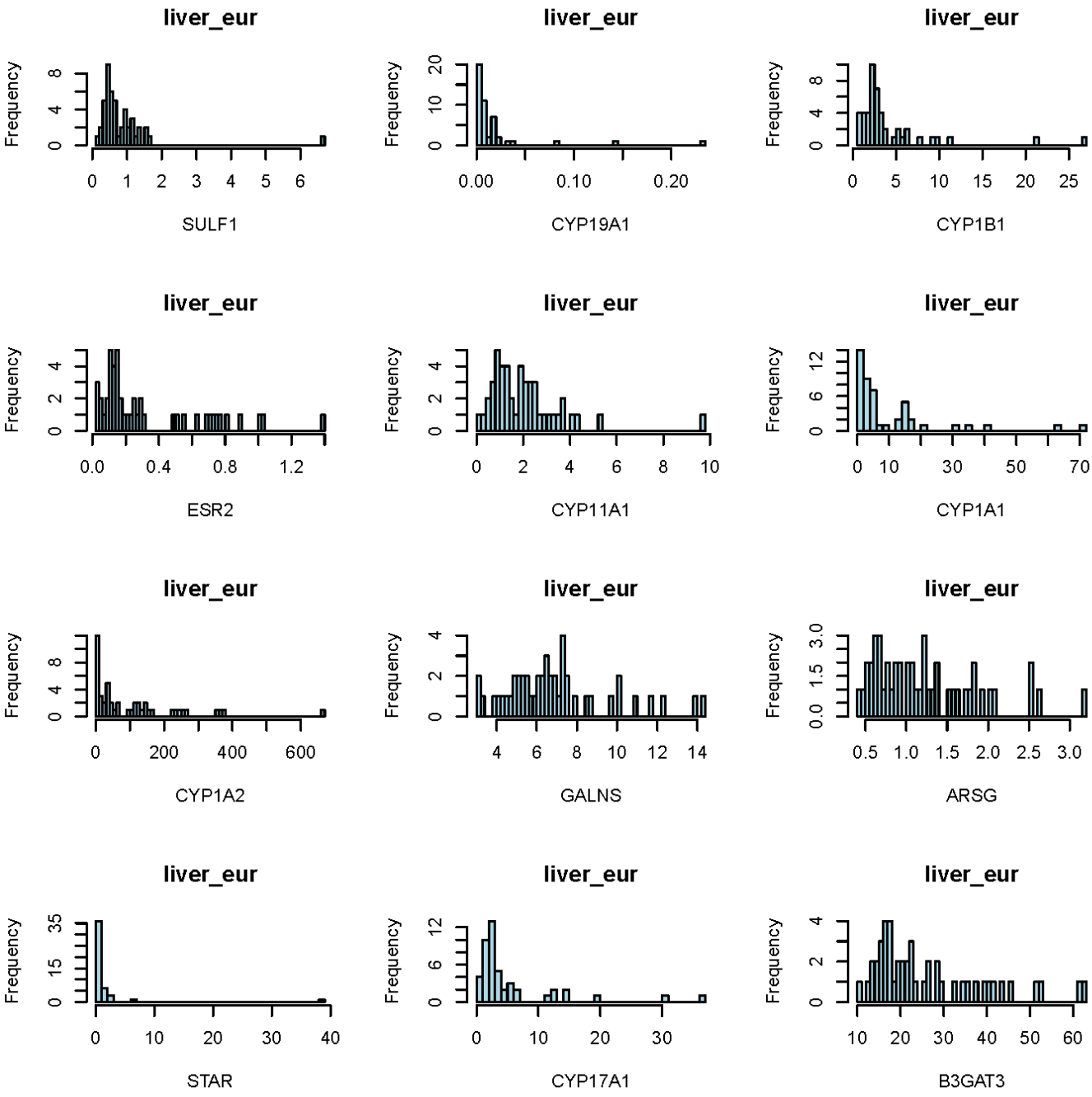
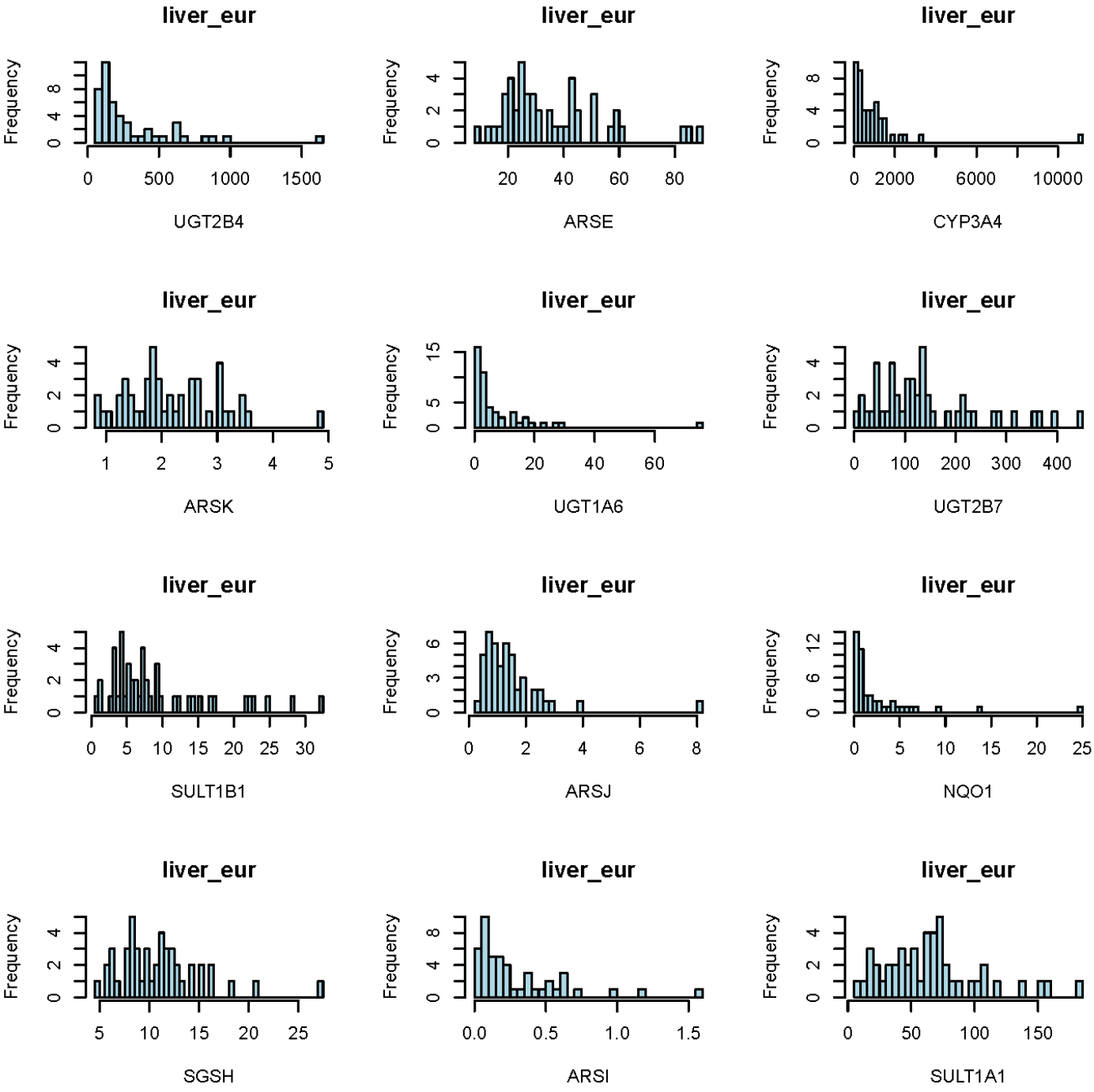

ovary_eur
Frequency
0
4
8
0.0
1.0
2.0
3.0
SULT1A3

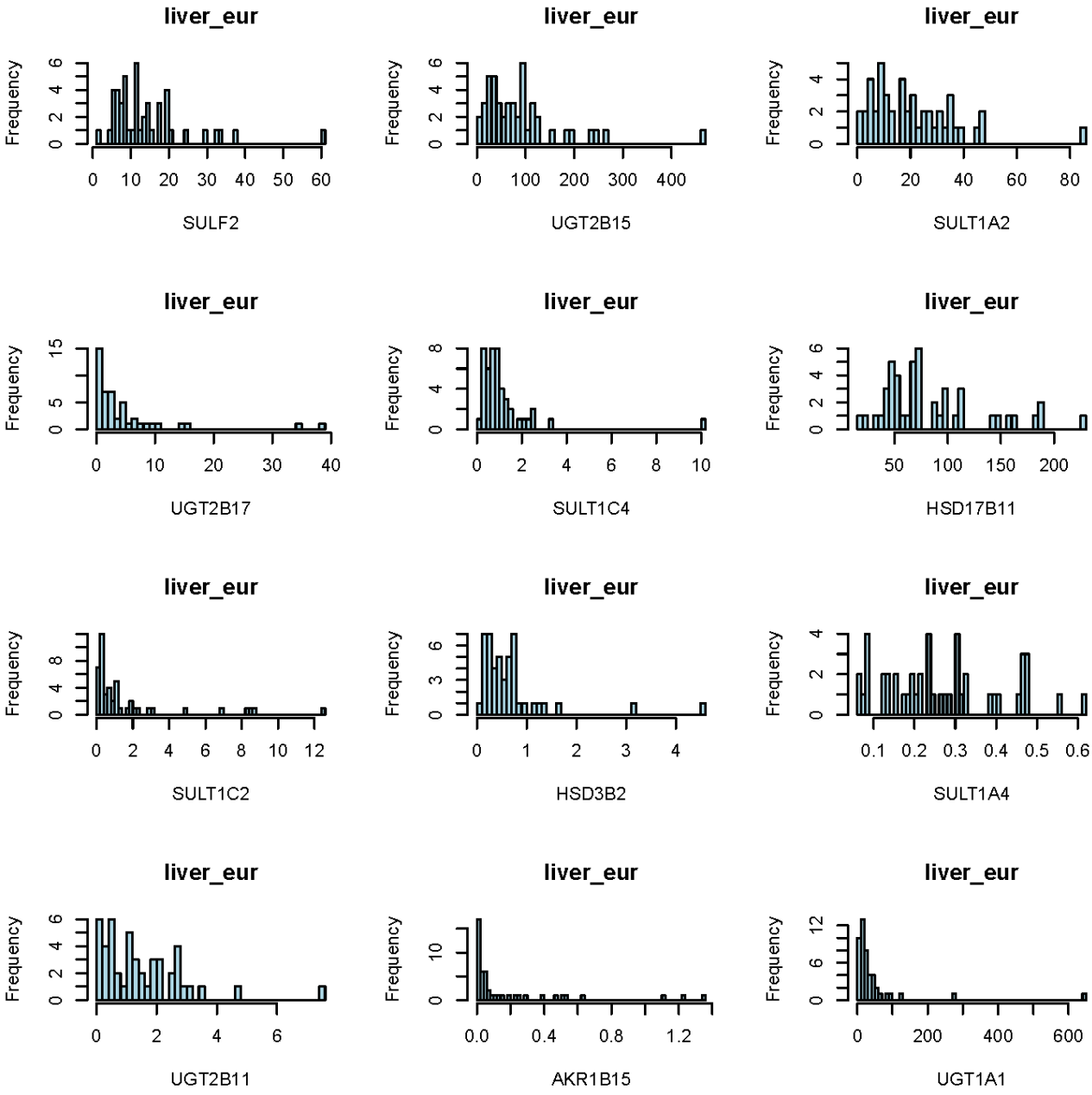
Figure 6. Distribution of gene expression levels for all estrogen-related genes in liver tissue among European descendants

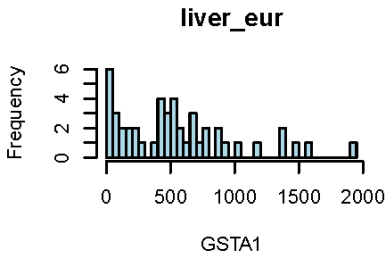


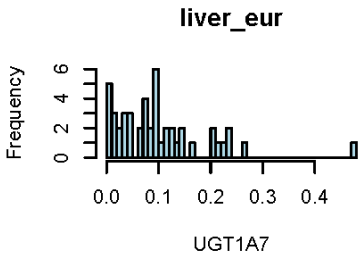

liver_eur
Frequency
HSD17B1
liver_eur
Frequency
UGT2B10
liver_eur
Frequency
SULT1E1
liver_eur
Frequency
B3GAT1
liver_eur
Frequency
B3GAT2
liver_eur
Frequency
ARSB
liver_eur
Frequency
SHBG
liver_eur
Frequency
SULT4A1
liver_eur
Frequency
GSTM1
liver_eur
Frequency
UGT2A3
liver_eur
Frequency
UGT2B28
liver_eur
Frequency
GNS

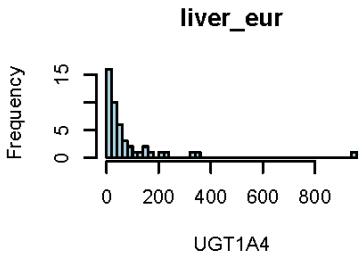

liver_eur
Frequency
SULF1
liver_eur
Frequency
CYP19A1
liver_eur
Frequency
CYP1B1
liver_eur
Frequency
ESR2
liver_eur
Frequency
CYP11A1
liver_eur
Frequency
CYP1A1
liver_eur
Frequency
CYP1A2
liver_eur
Frequency
GALNS
liver_eur
Frequency
ARSG
liver_eur
Frequency
STAR
liver_eur
Frequency
CYP17A1
liver_eur
Frequency
B3GAT3


liver_eur
Frequency
UGT2B4
liver_eur
Frequency
ARSE
liver_eur
Frequency
CYP3A4
liver_eur
Frequency
ARSK
liver_eur
Frequency
UGT1A6
liver_eur
Frequency
UGT2B7
liver_eur
Frequency
SULT1B1
liver_eur
Frequency
ARSJ
liver_eur
Frequency
NQO1
liver_eur
Frequency
SGSH
liver_eur
Frequency
ARSI
liver_eur
Frequency
SULT1A1


liver_eur
Frequency
SULF2
liver_eur
Frequency
UGT2B15
liver_eur
Frequency
SULT1A2
liver_eur
Frequency
UGT2B17
liver_eur
Frequency
SULT1C4
liver_eur
Frequency
HSD17B11
liver_eur
Frequency
SULT1C2
liver_eur
Frequency
HSD3B2
liver_eur
Frequency
SULT1A4
liver_eur
Frequency
UGT2B11
liver_eur
Frequency
AKR1B15
liver_eur
Frequency
UGT1A1


liver_eur
Frequency
0 2 4 6
0 500 1000 1500 2000
GSTA1


liver_eur
Frequency
0 2 4 6
0.0 0.1 0.2 0.3 0.4
UGT1A7


liver_eur
Frequency
0 5 15
0 200 400 600 800
UGT1A4

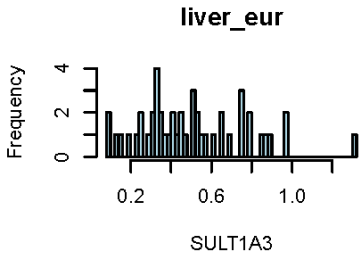
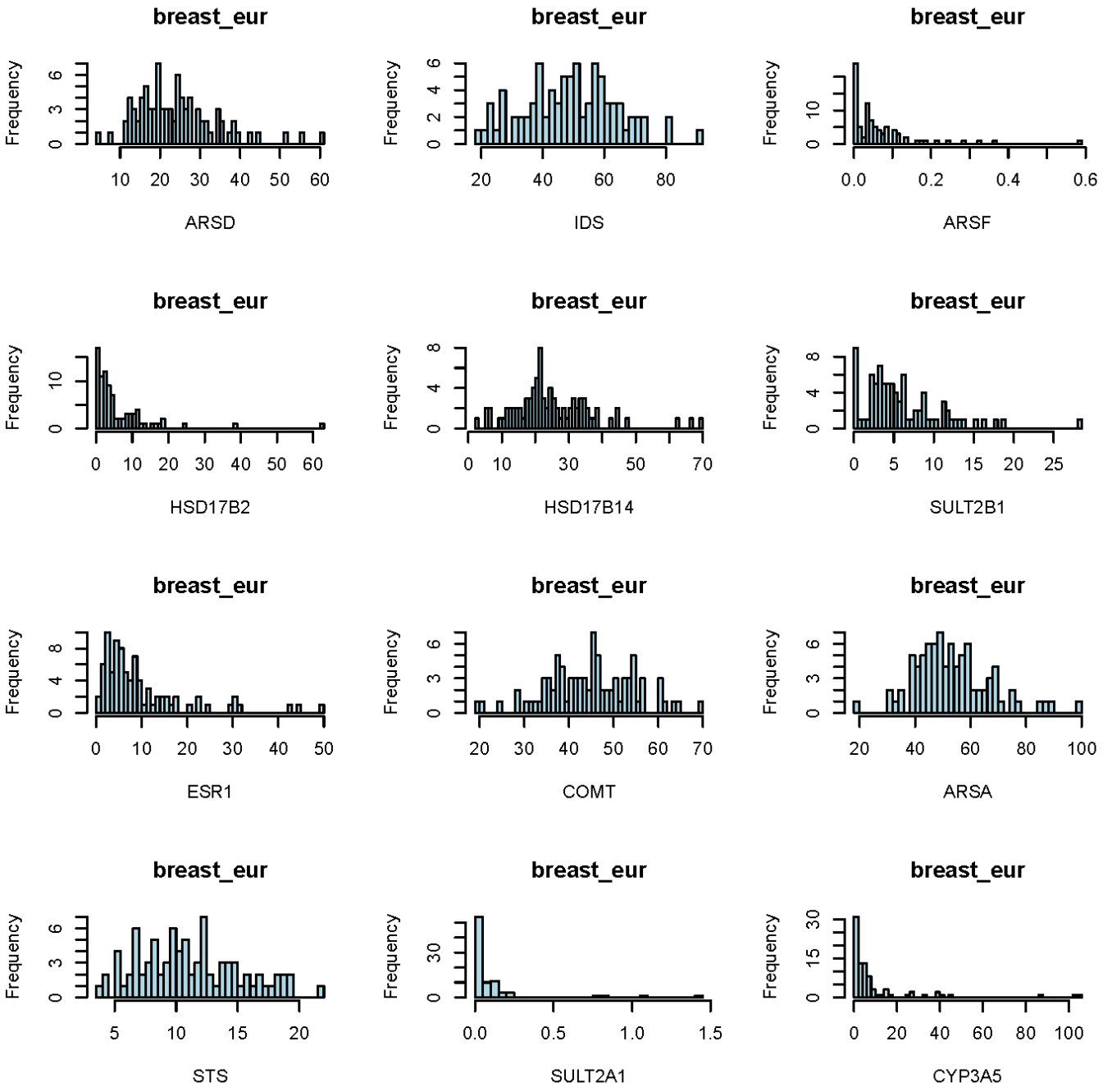
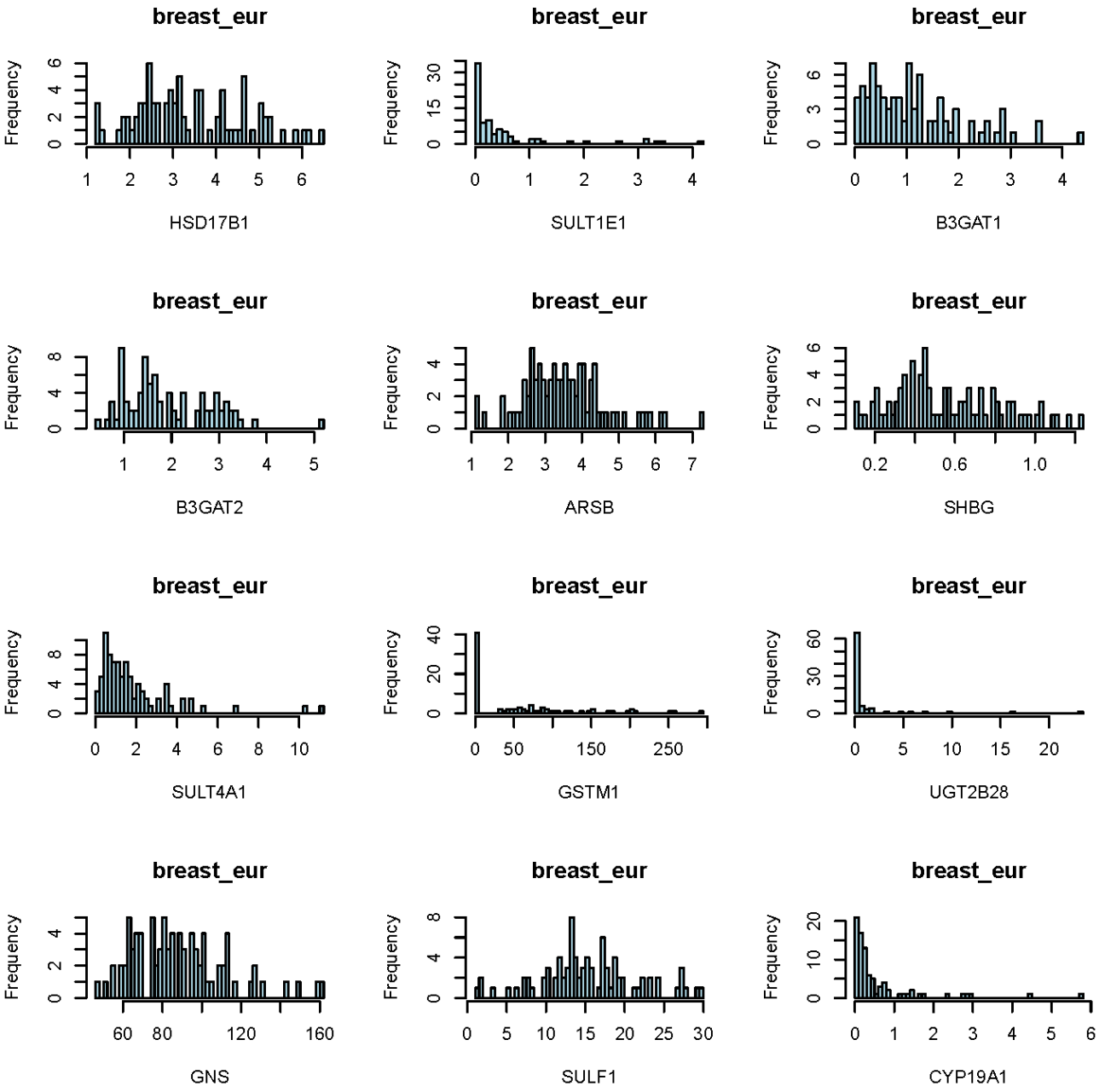
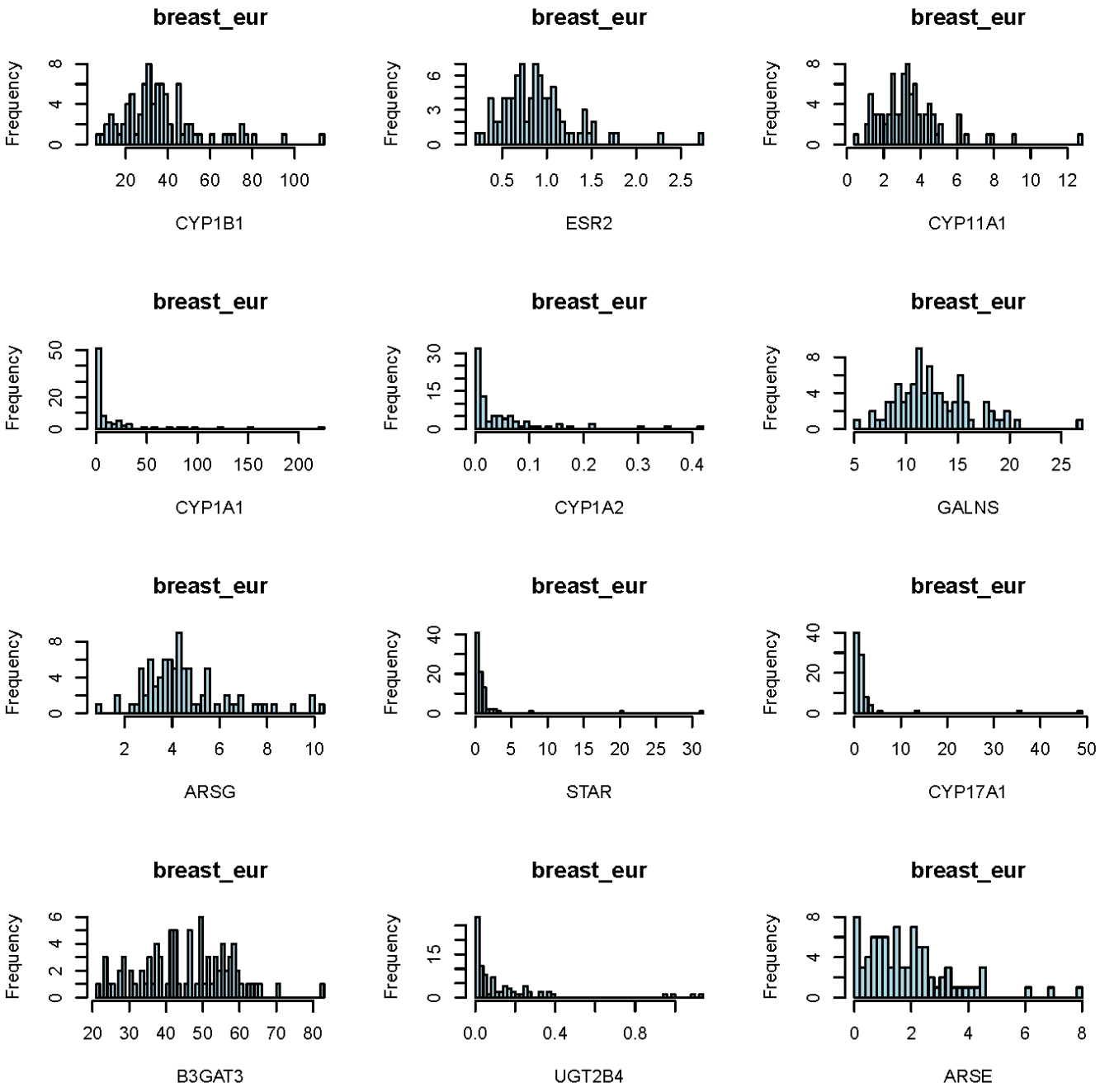
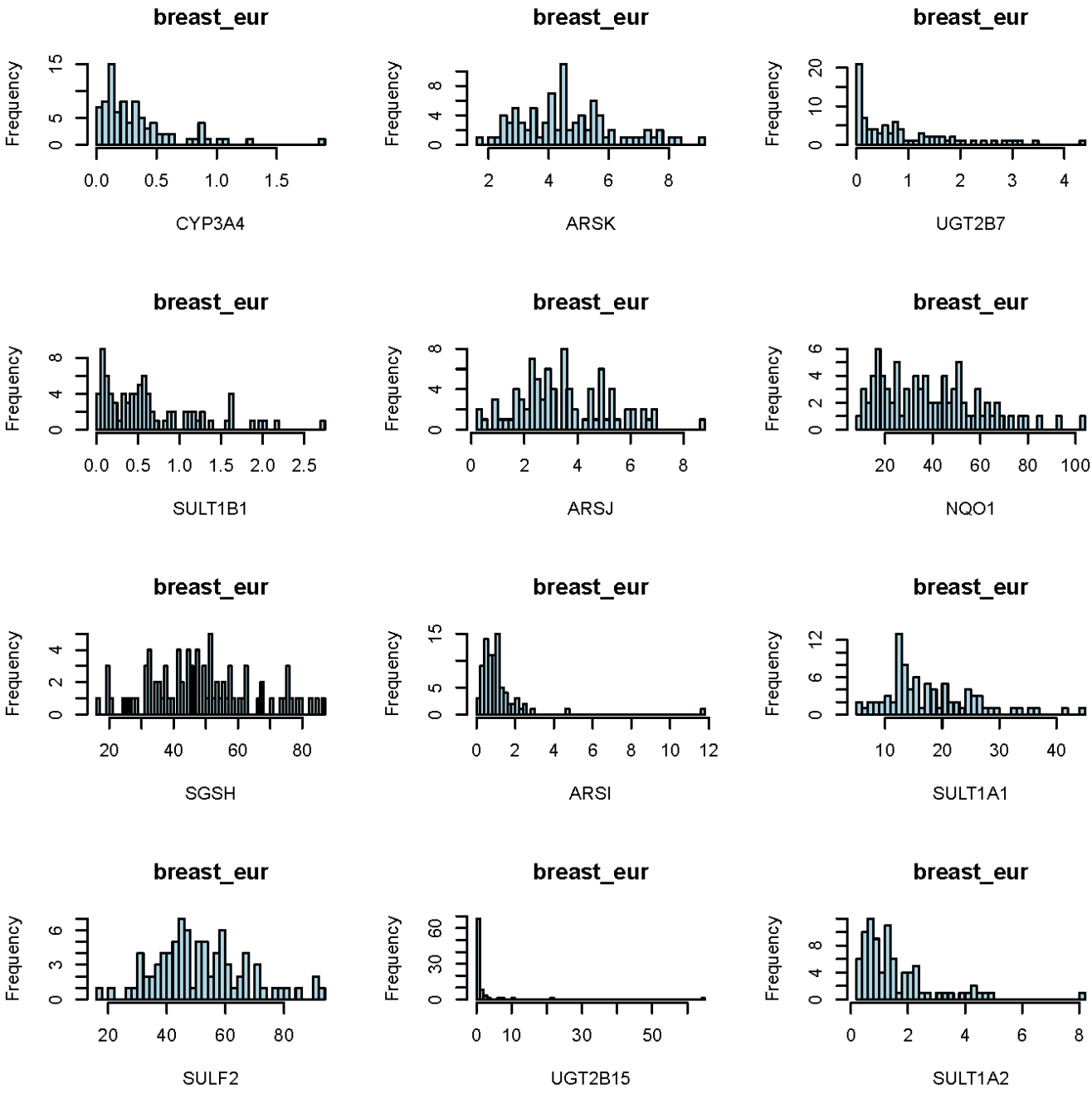

liver_eur
Frequency
0 2 4
0.2 0.6 1.0
SULT1A3

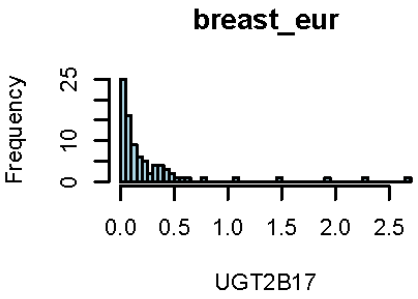
Figure 7. Distribution of gene expression levels for all estrogen-related genes in breast tissue among European descendants.

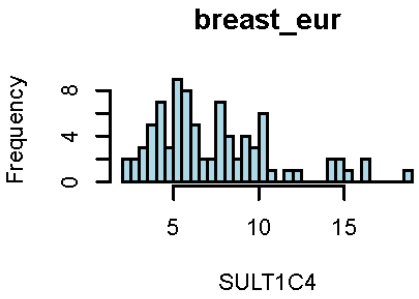


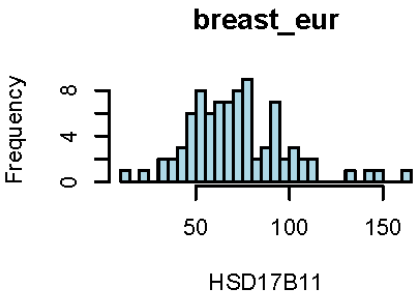

breast_eur
Frequency
HSD17B1
breast_eur
Frequency
SULT1E1
breast_eur
Frequency
B3GAT1
breast_eur
Frequency
B3GAT2
breast_eur
Frequency
ARSB
breast_eur
Frequency
SHBG
breast_eur
Frequency
SULT4A1
breast_eur
Frequency
GSTM1
breast_eur
Frequency
UGT2B28
breast_eur
Frequency
GNS
breast_eur
Frequency
SULF1
breast_eur
Frequency
CYP19A1


breast_eur
Frequency
CYP1B1
breast_eur
Frequency
ESR2
breast_eur
Frequency
CYP11A1
breast_eur
Frequency
CYP1A1
breast_eur
Frequency
CYP1A2
breast_eur
Frequency
GALNS
breast_eur
Frequency
ARSG
breast_eur
Frequency
STAR
breast_eur
Frequency
CYP17A1
breast_eur
Frequency
B3GAT3
breast_eur
Frequency
UGT2B4
breast_eur
Frequency
ARSE

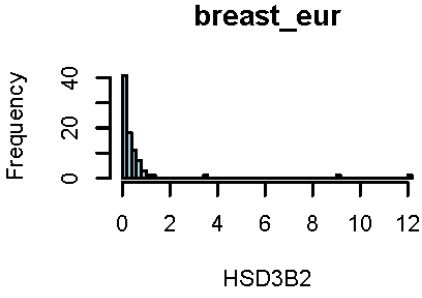

breast_eur
Frequency
CYP3A4
breast_eur
Frequency
ARSK
breast_eur
Frequency
UGT2B7
breast_eur
Frequency
SULT1B1
breast_eur
Frequency
ARSJ
breast_eur
Frequency
NQO1
breast_eur
Frequency
SGSH
breast_eur
Frequency
ARSI
breast_eur
Frequency
SULT1A1
breast_eur
Frequency
SULF2
breast_eur
Frequency
UGT2B15
breast_eur
Frequency
SULT1A2

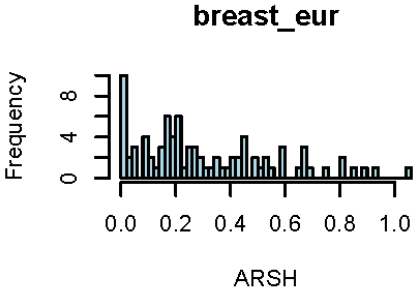

breast_eur
Frequency
0 10 25
0.0 0.5 1.0 1.5 2.0 2.5
UGT2B17

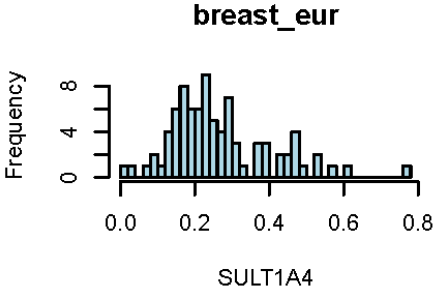

breast_eur
Frequency
0 4 8
5 10 15
SULT1C4

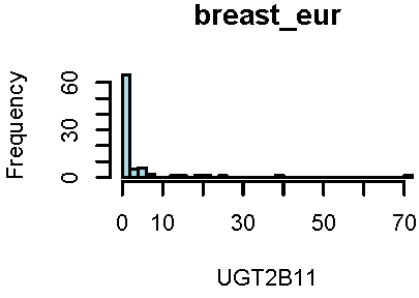

breast_eur
Frequency
0 4 8
50 100 150
HSD17B11

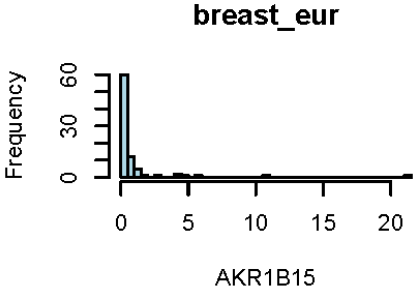

breast_eur
Frequency
0 10
0 1 2 3 4 5
SULT1C2

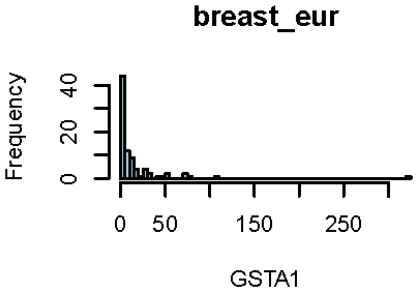

breast_eur
Frequency
0 20 40
0 2 4 6 8 10 12
HSD3B2


breast_eur
Frequency
0 4 8
0.0 0.2 0.4 0.6 0.8 1.0
ARSH


breast_eur
Frequency
0 4 8
0.0 0.2 0.4 0.6 0.8
SULT1A4


breast_eur
Frequency
0 30 60
0 10 30 50 70
UGT2B11


breast_eur
Frequency
0 30 60
0 5 10 15 20
AKR1B15


breast_eur
Frequency
0 20 40
0 50 150 250
GSTA1

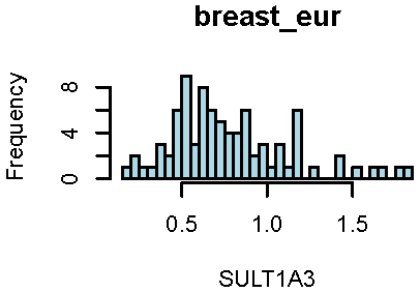
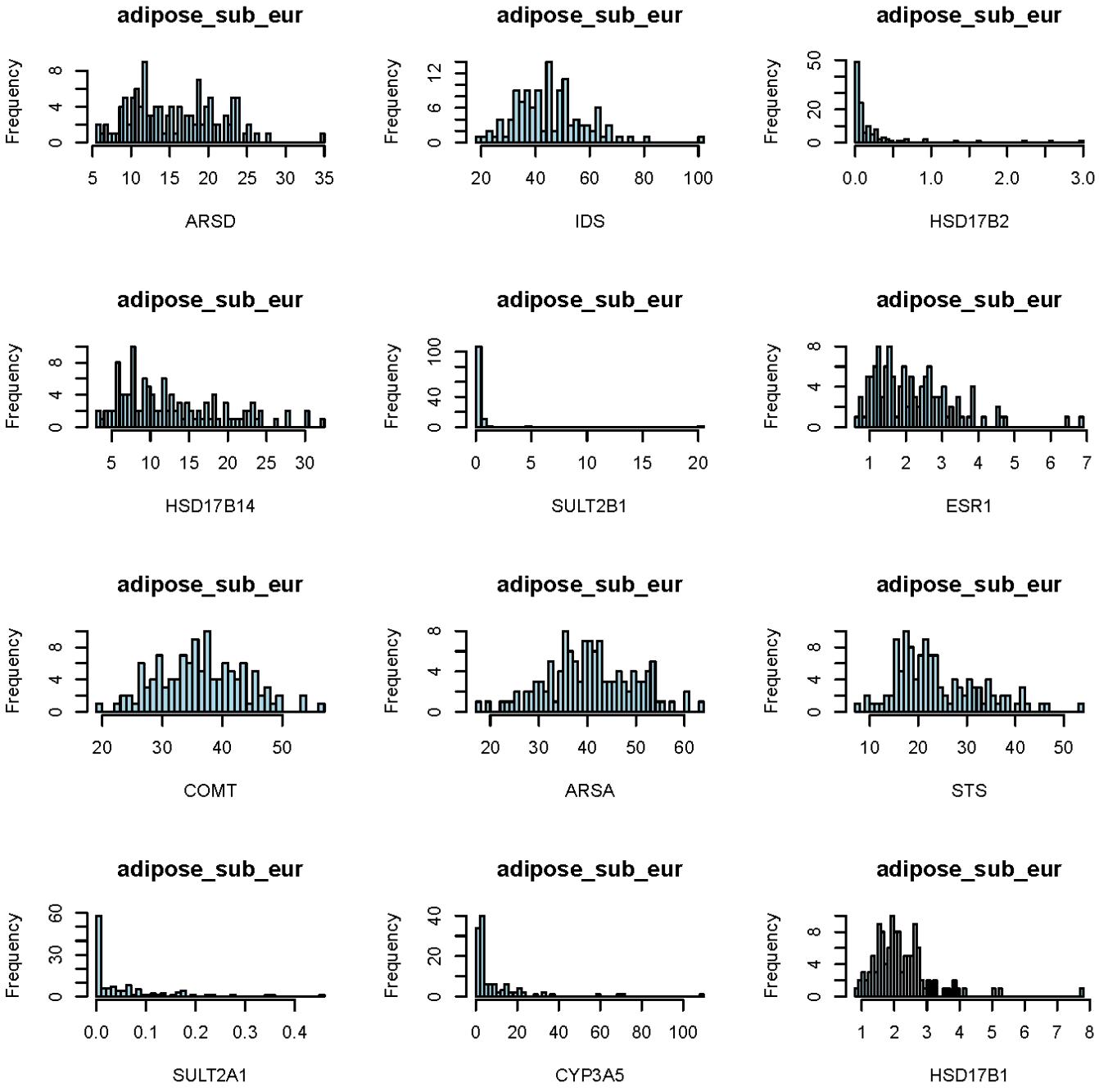
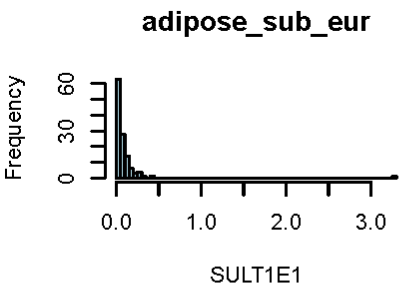
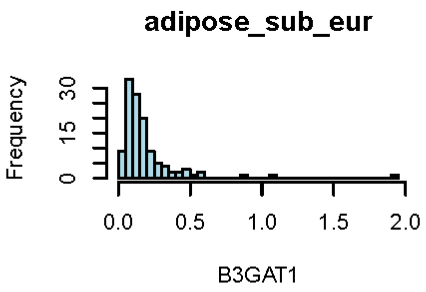
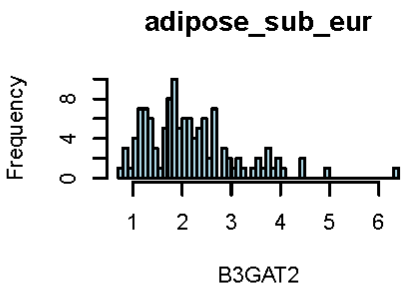

breast_eur
Frequency
0 4 8
0.5 1.0 1.5
SULT1A3

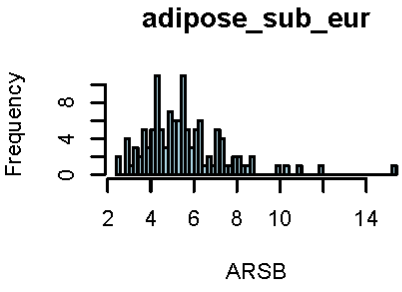
Figure 8. Distribution of gene expression for all estrogen-related genes in subcutaneous adipose tissue among Europeans.

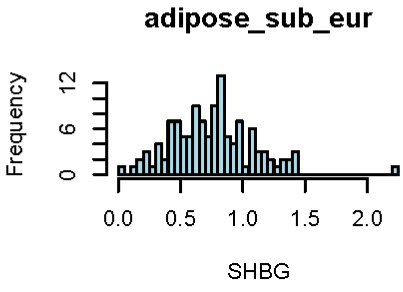


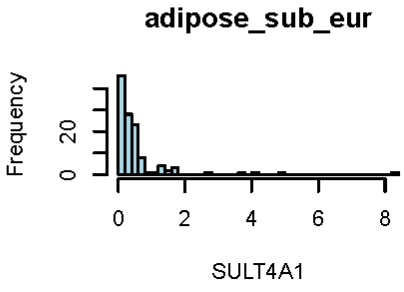

adipose_sub_eur
Frequency
0
30
60
0.0
1.0
2.0
3.0
SULT1E1

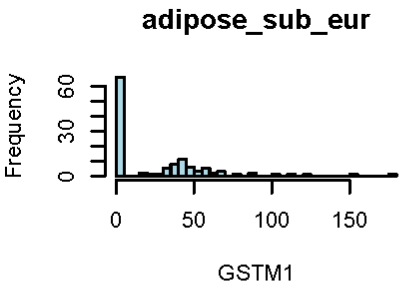

adipose_sub_eur
Frequency
0
15
30
0.0
0.5
1.0
1.5
2.0
B3GAT1

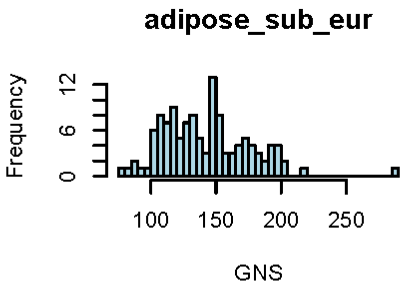

adipose_sub_eur
Frequency
0
4
8
1
2
3
4
5
6
B3GAT2

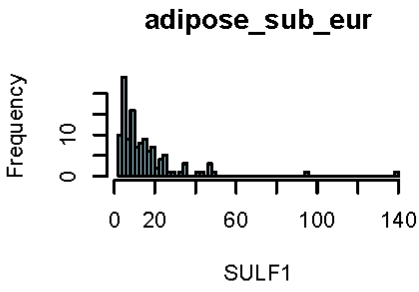

adipose_sub_eur
Frequency
0
4
8
2
4
6
8
10
14
ARSB

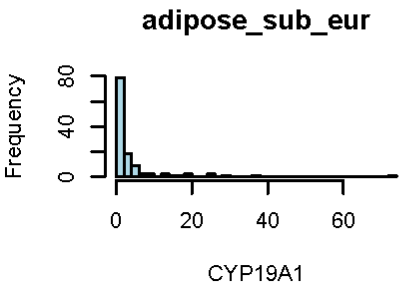

adipose_sub_eur
Frequency
0
6
12
0.0
0.5
1.0
1.5
2.0
SHBG

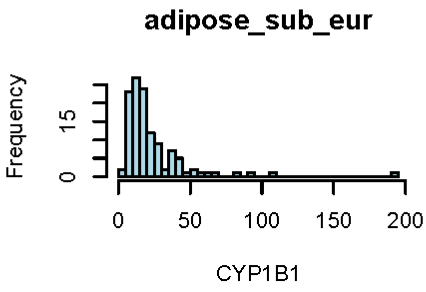

adipose_sub_eur
Frequency
0
20
0
2
4
6
8
SULT4A1

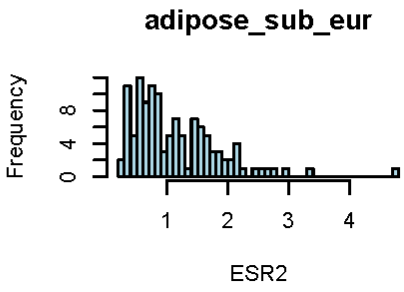

adipose_sub_eur
Frequency
0
30
60
0
50
100
150
GSTM1

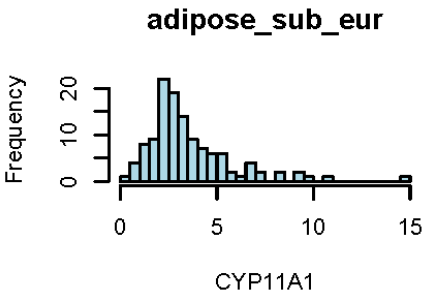

adipose_sub_eur
Frequency
0
6
12
100
150
200
250
GNS

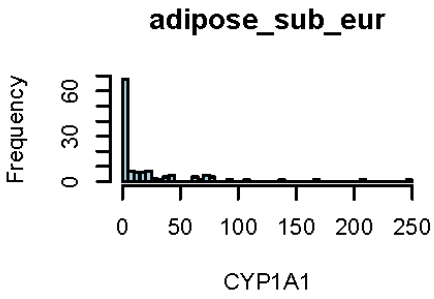

adipose_sub_eur
Frequency
0
10
0
20
60
100
140
SULF1

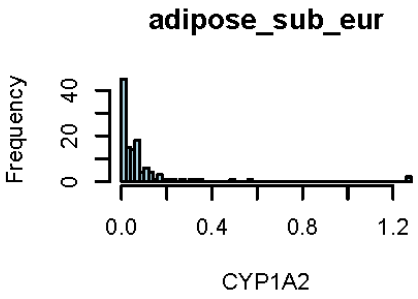

adipose_sub_eur
Frequency
0
40
80
0
20
40
60
CYP19A1

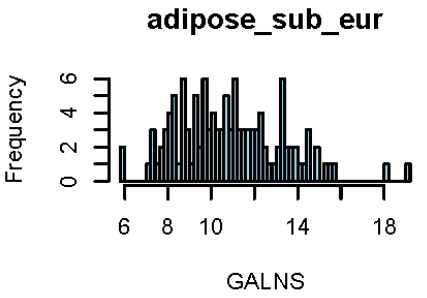

adipose_sub_eur
Frequency
0
15
0
50
100
150
200
CYP1B1

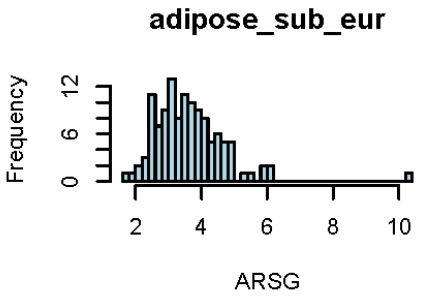

adipose_sub_eur
Frequency
0
4
8
1
2
3
4
ESR2

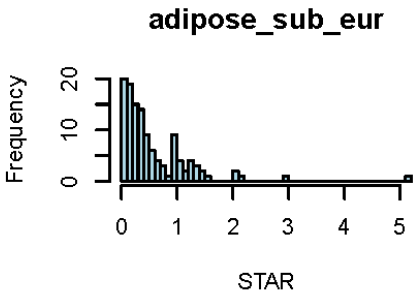

adipose_sub_eur
Frequency
0
10
20
0
5
10
15
CYP11A1

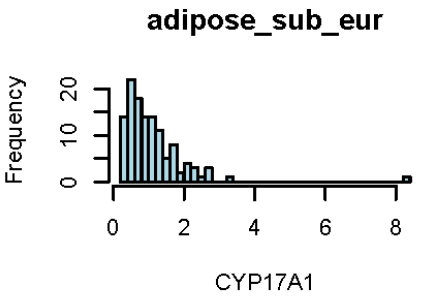

adipose_sub_eur
Frequency
0
30
60
0
50
100
150
200
250
CYP1A1

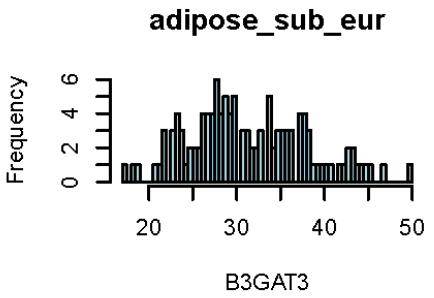

adipose_sub_eur
Frequency
0
20
40
0.0
0.4
0.8
1.2
CYP1A2

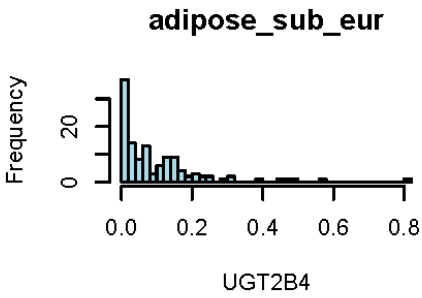

adipose_sub_eur
Frequency
0
2
4
6
6
8
10
14
18
GALNS

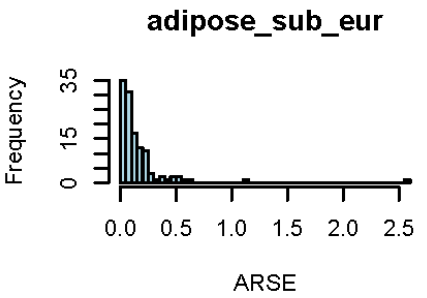

adipose_sub_eur
Frequency
0
6
12
2
4
6
8
10
ARSG

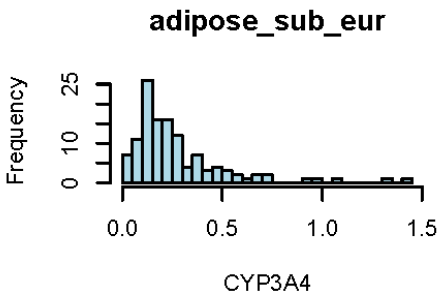

adipose_sub_eur
Frequency
0
10
20
0
1
2
3
4
5
STAR

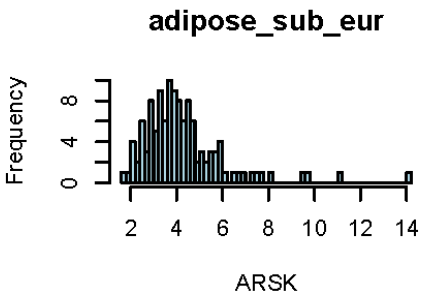

adipose_sub_eur
Frequency
0
10
20
0
2
4
6
8
CYP17A1

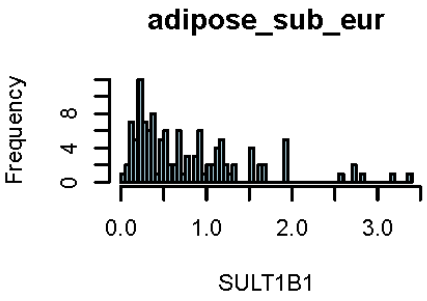

adipose_sub_eur
Frequency
0
2
4
6
20
30
40
50
B3GAT3

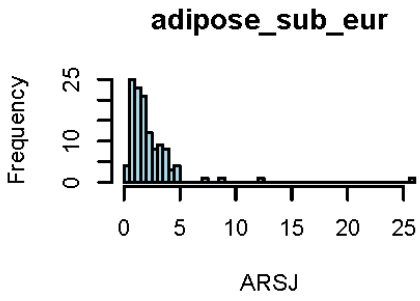

adipose_sub_eur
Frequency
0
20
0.0
0.2
0.4
0.6
0.8
UGT2B4

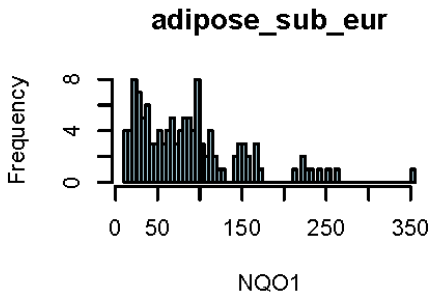

adipose_sub_eur
Frequency
0
15
35
0.0
0.5
1.0
1.5
2.0
2.5
ARSE

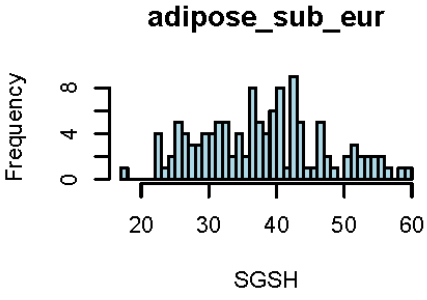

adipose_sub_eur
Frequency
0
10
25
0.0
0.5
1.0
1.5
CYP3A4

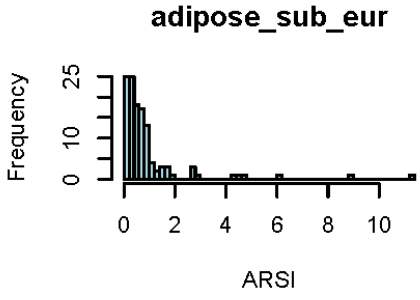

adipose_sub_eur
Frequency
0
4
8
2
4
6
8
10
12
14
ARSK

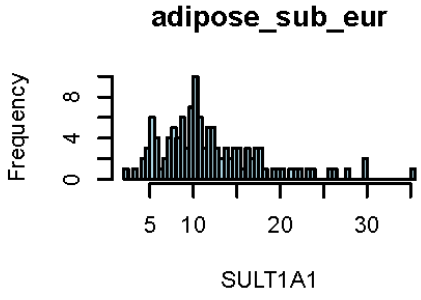

adipose_sub_eur
Frequency
0 4 8
0.0 1.0 2.0 3.0
SULT1B1

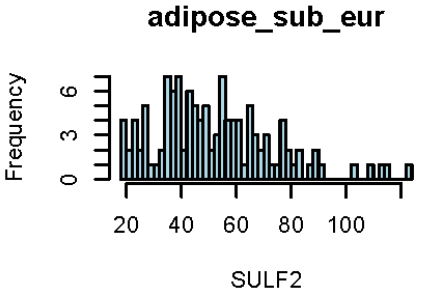

adipose_sub_eur
Frequency
0 10 25
0 5 10 15 20 25
ARSJ

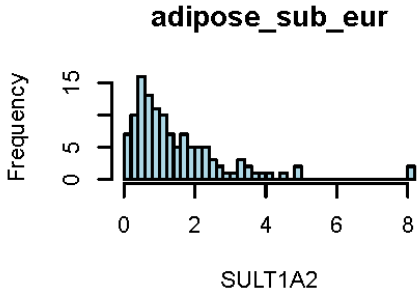

adipose_sub_eur
Frequency
0 4 8
0 50 150 250 350
NQO1

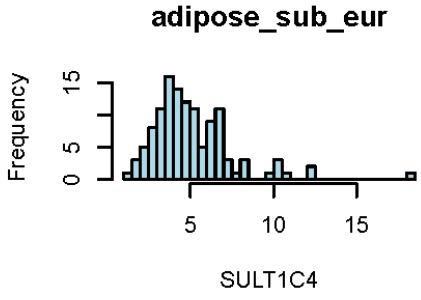

adipose_sub_eur
Frequency
0 4 8
20 30 40 50 60
SGSH

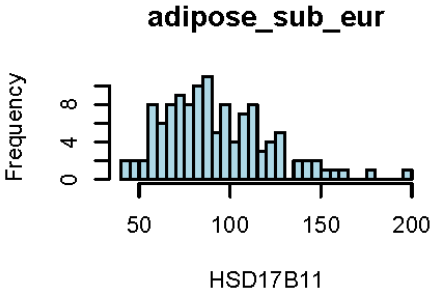

adipose_sub_eur
Frequency
0 10 25
0 2 4 6 8 10
ARSI

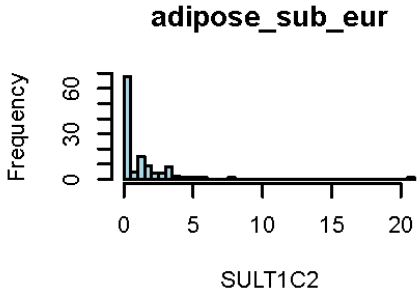

adipose_sub_eur
Frequency
0 4 8
5 10 20 30
SULT1A1

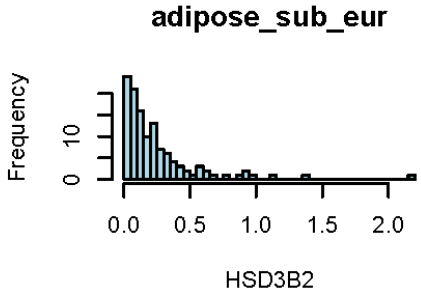

adipose_sub_eur
Frequency
0 3 6
20 40 60 80 100
SULF2

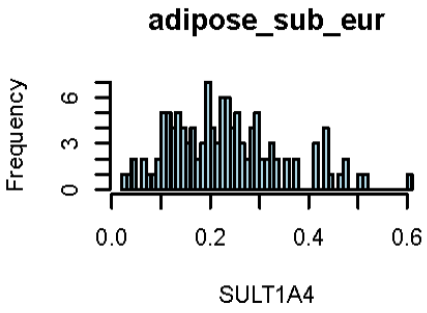

adipose_sub_eur
Frequency
0 5 15
0 2 4 6 8
SULT1A2

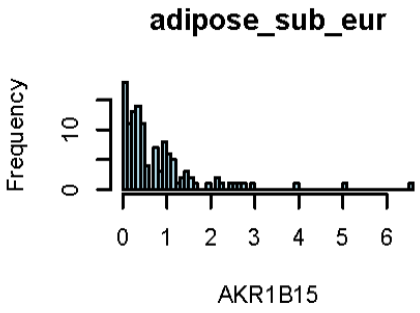

adipose_sub_eur
Frequency
0 5 15
5 10 15
SULT1C4

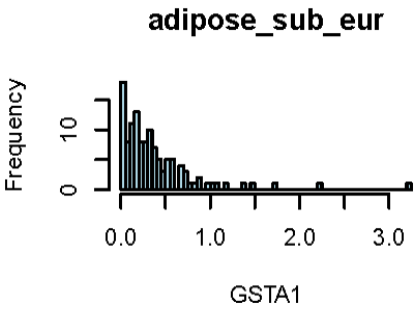

adipose_sub_eur
Frequency
0 4 8
50 100 150 200
HSD17B11


adipose_sub_eur
Frequency
0 30 60
0 5 10 15 20
SULT1C2


adipose_sub_eur
Frequency
0 10
0.0 0.5 1.0 1.5 2.0
HSD3B2


adipose_sub_eur
Frequency
0
3
6
0.0
0.2
0.4
0.6
SULT1A4


adipose_sub_eur
Frequency
0
10
0
1
2
3
4
5
6
AKR1B15


adipose_sub_eur
Frequency
0
10
0.0
1.0
2.0
3.0
GSTA1

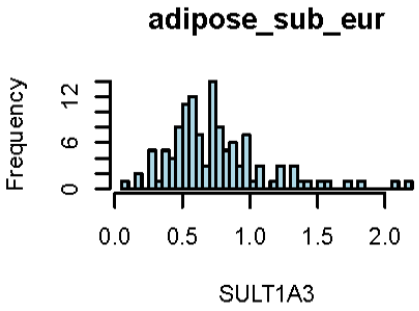
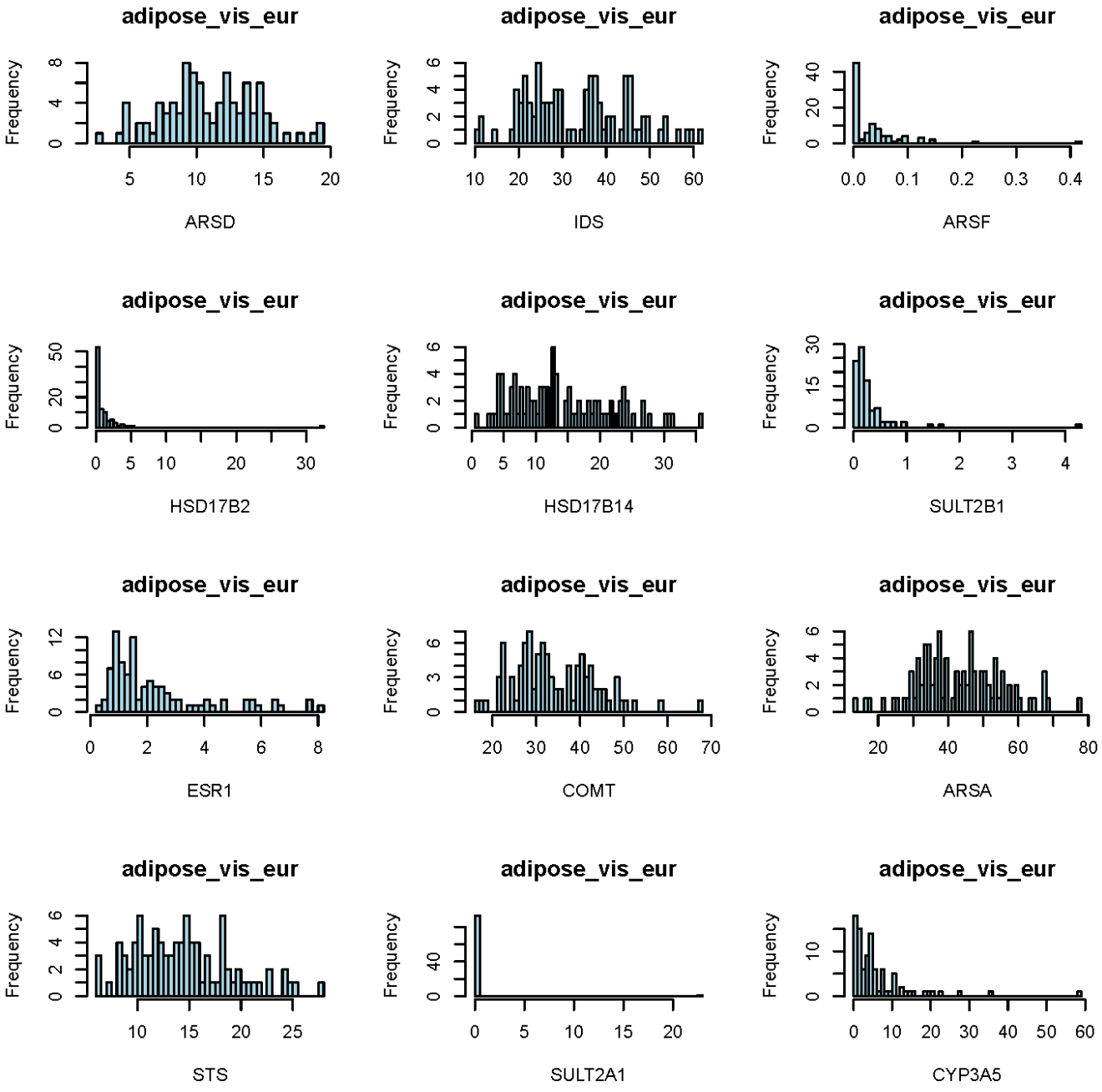
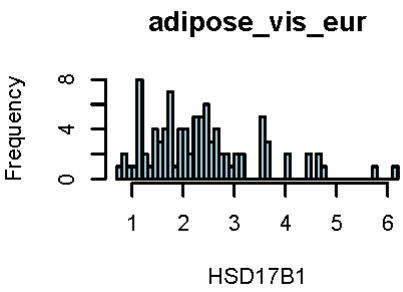
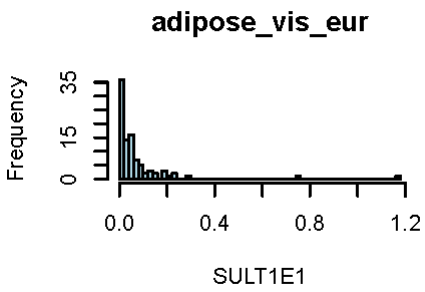
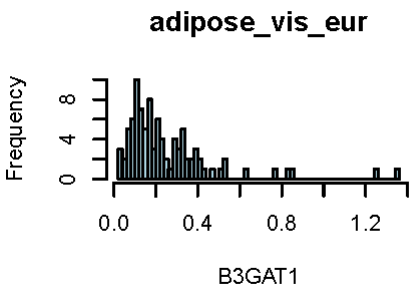

adipose_sub_eur
Frequency
0
6
12
0.0
0.5
1.0
1.5
2.0
SULT1A3

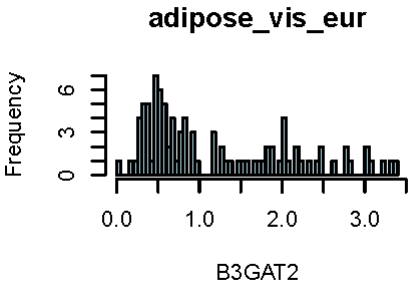
Figure 9. Distribution of gene expression levels for all estrogen-related genes in visceral adipose tissue among Europeans.

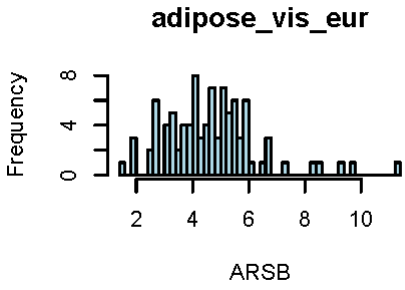


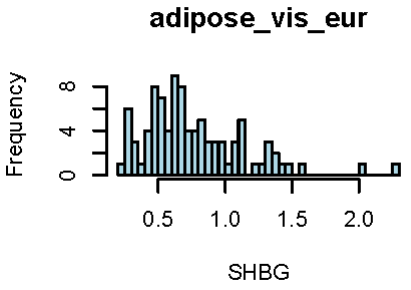

adipose_vis_eur
Frequency
0
4
8
1
2
3
4
5
6
HSD17B1

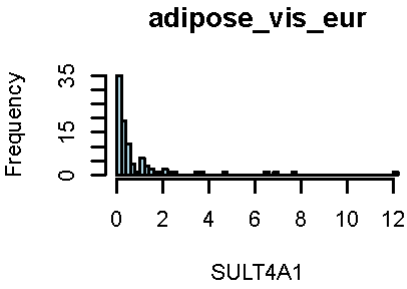

adipose_vis_eur
Frequency
0
15
35
0.0
0.4
0.8
1.2
SULT1E1

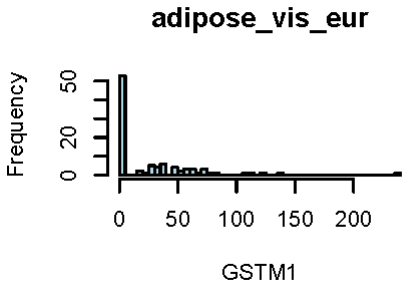

adipose_vis_eur
Frequency
0
4
8
0.0
0.4
0.8
1.2
B3GAT1

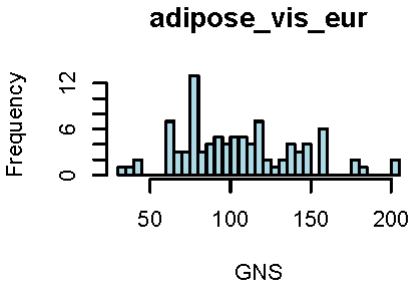

adipose_vis_eur
Frequency
0
3
6
0.0
1.0
2.0
3.0
B3GAT2

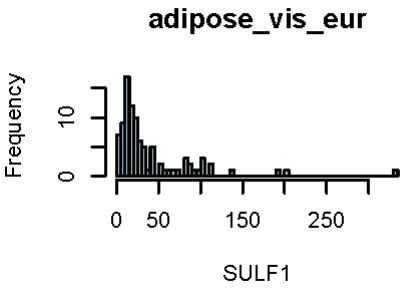

adipose_vis_eur
Frequency
0
4
8
2
4
6
8
10
ARSB

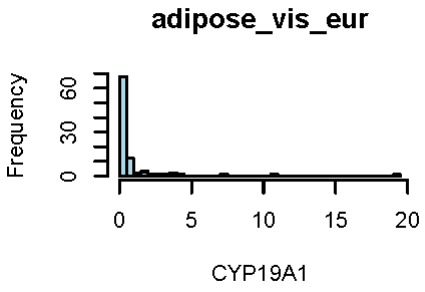

adipose_vis_eur
Frequency
0
4
8
0.5
1.0
1.5
2.0
SHBG

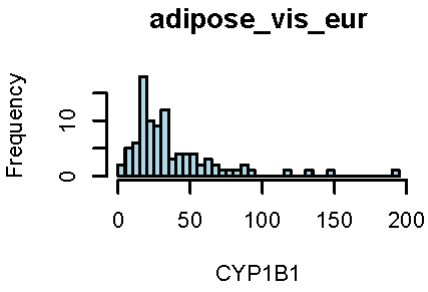

adipose_vis_eur
Frequency
0
15
35
0
2
4
6
8
10
12
SULT4A1

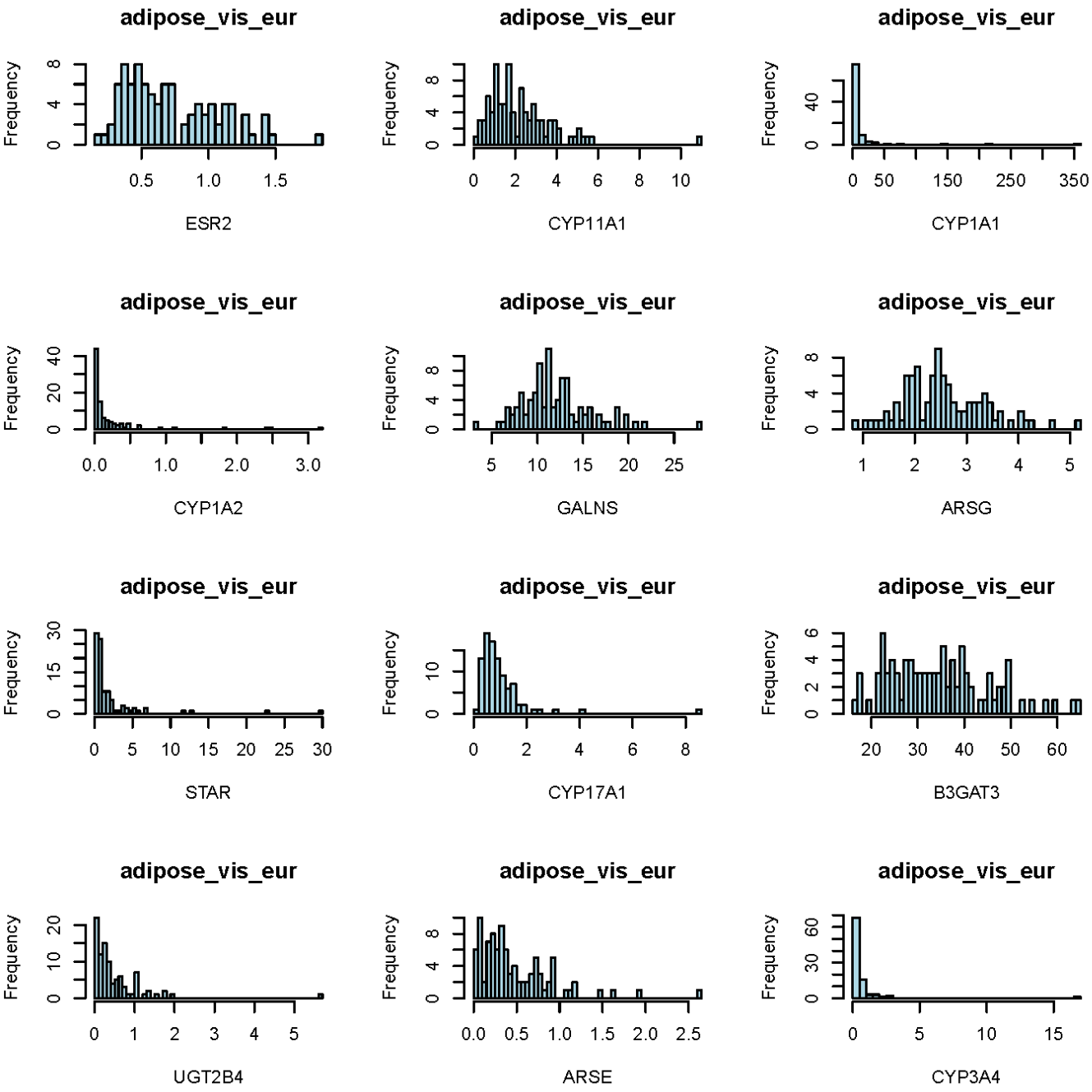

adipose_vis_eur
Frequency
0
20
50
0
50
100
150
200
GSTM1

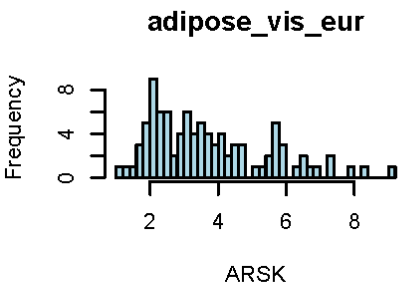

adipose_vis_eur
Frequency
0
6
12
50
100
150
200
GNS

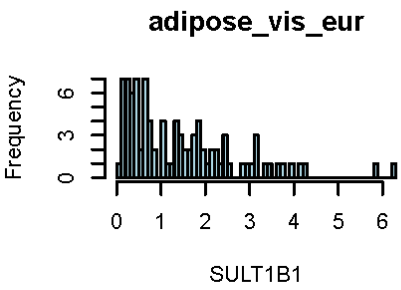

adipose_vis_eur
Frequency
0
10
0
50
150
250
SULF1

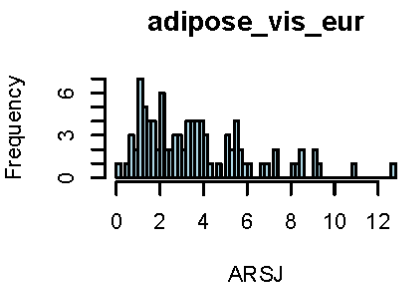

adipose_vis_eur
Frequency
0
30
60
0
5
10
15
20
CYP19A1

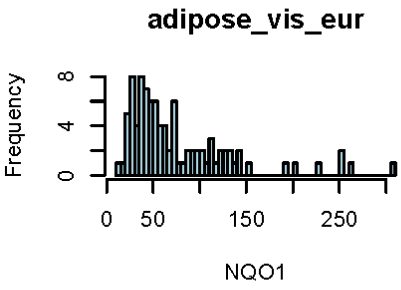

adipose_vis_eur
Frequency
0
10
0
50
100
150
200
CYP1B1

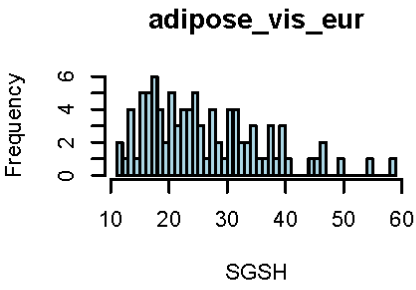

adipose_vis_eur
Frequency
ESR2
adipose_vis_eur
Frequency
CYP11A1
adipose_vis_eur
Frequency
CYP1A1
adipose_vis_eur
Frequency
CYP1A2
adipose_vis_eur
Frequency
GALNS
adipose_vis_eur
Frequency
ARSG
adipose_vis_eur
Frequency
STAR
adipose_vis_eur
Frequency
CYP17A1
adipose_vis_eur
Frequency
B3GAT3
adipose_vis_eur
Frequency
UGT2B4
adipose_vis_eur
Frequency
ARSE
adipose_vis_eur
Frequency
CYP3A4

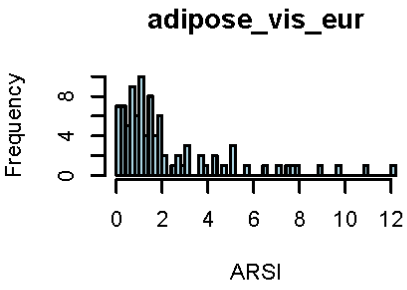

adipose_vis_eur
Frequency
0 4 8
2 4 6 8
ARSK

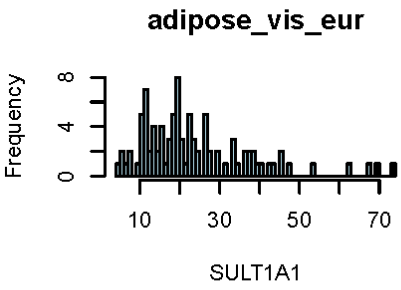

adipose_vis_eur
Frequency
0 3 6
0 1 2 3 4 5 6
SULT1B1

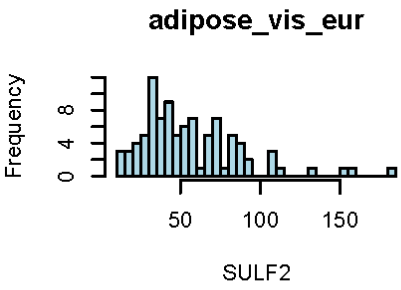

adipose_vis_eur
Frequency
0 3 6
0 2 4 6 8 10 12
ARSJ

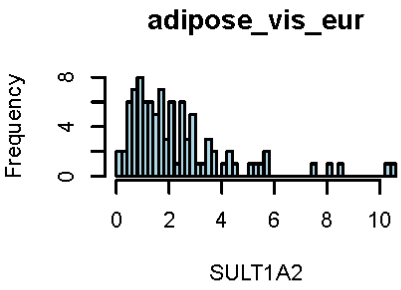

adipose_vis_eur
Frequency
0 4 8
0 50 150 250
NQO1

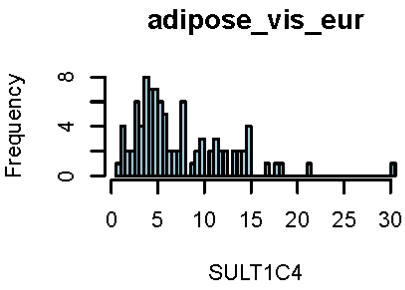

adipose_vis_eur
Frequency
0 2 4 6
10 20 30 40 50 60
SGSH

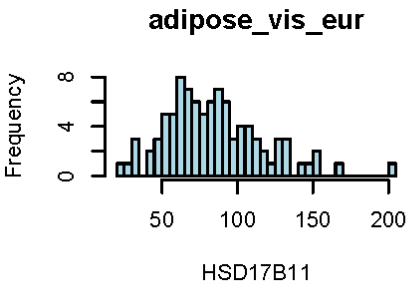

adipose_vis_eur
Frequency
0 4 8
0 2 4 6 8 10 12
ARSI

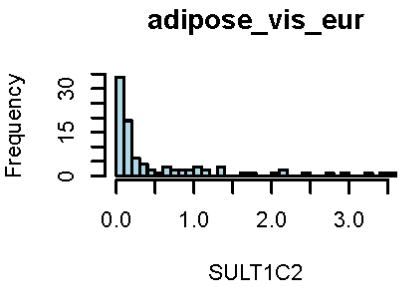

adipose_vis_eur
Frequency
0 4 8
10 30 50 70
SULT1A1

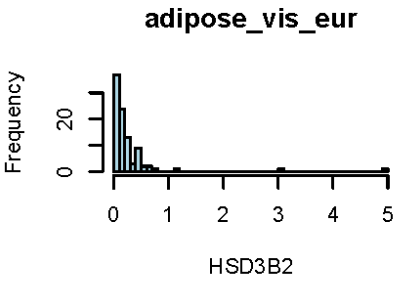

adipose_vis_eur
Frequency
0 4 8
50 100 150
SULF2

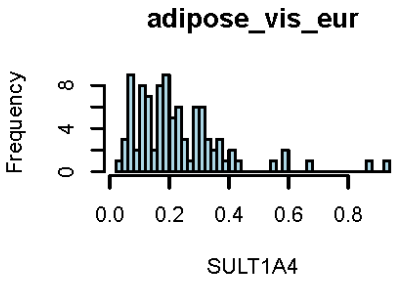

adipose_vis_eur
Frequency
0 4 8
0 2 4 6 8 10
SULT1A2

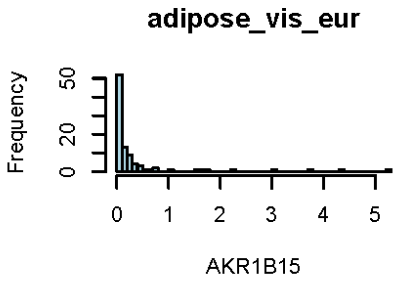

adipose_vis_eur
Frequency
0 4 8
0 5 10 15 20 25 30
SULT1C4

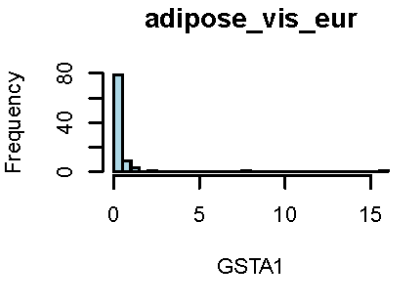

adipose_vis_eur
Frequency
0 4 8
50 100 150 200
HSD17B11

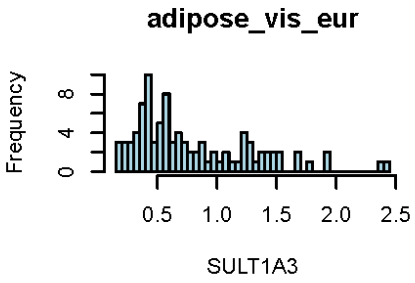

adipose_vis_eur
Frequency
0 15 30
0.0 1.0 2.0 3.0
SULT1C2


adipose_vis_eur
Frequency
0
20
0
1
2
3
4
5
HSD3B2


adipose_vis_eur
Frequency
0
4
8
0.0
0.2
0.4
0.6
0.8
SULT1A4


adipose_vis_eur
Frequency
0
20
50
0
1
2
3
4
5
AKR1B15


adipose_vis_eur
Frequency
0
40
80
0
5
10
15
GSTA1


adipose_vis_eur
Frequency
0
4
8
0.5
1.0
1.5
2.0
2.5
SULT1A3

CHAPTER 3

SPECIFIC AIM 2

The objective of this aim is to investigate associations of predicted expression levels of estrogen synthesis, metabolism and bioavailability pathway genes with breast cancer risk in both Europeans and Asians;

1. Methods

1.1. Study participants

We used individual level data from two sources, the Breast Cancer Association Consortium (BCAC), and the Asia Breast Cancer Consortium (ABCC), for Europeans and Asians, respectively. For analyses conducted among European ancestries, individual level data of 105,421 cases and 91,920 controls from BCAC were used. For association analyses among Asian populations, we included 12,213 cases and 12,840 controls from BCAC, and 11,993 cases and 11,935 controls from ABCC. Duplicates for Asian participants in BCAC were excluded. Consequently, a total of 24,206, cases and 24,775 controls were included for analyses among Asians. Most of the included studies was population-bases case-control studies, or nested case-control studies from population-bases cohort studies. All participating studies have been approved by their local institutional review board and all participants have provided informed consent.

1.2. Genotyping array, quality control and imputation

a. OncoArray

61,282 breast cancer cases and 45,494 controls of European ancestry and 7,454 cases and 6,883 controls of Asian population were genotyped with the Illumina OncoArray BeadChip, which is a

570,000-SNP custom array with genome-wide coverage. About half of the SNPs on this platform were common variants, which are the GWAS backbone (Illumina HumanCore). And the other half of the SNPs was selected based on their associations with diseases from previous studies. Genotypes were called using the Gentrain2 algorithm. Samples that (1) are possible duplicates or relatives within or across studies; (2) had a call rate < 95%; (3) had extreme heterozygosity; (4) Ancestry outliers as determined with reference to the HapMap (v.2) populations based on the first two principal components were also excluded. Additionally, SNPs that (1) had a call rate < 95%; (2) were not in Hardy-Weinberg equilibrium ($P < 10^{-12}$ in cases and $P < 10^{-7}$ in controls); (3) were with concordance <98% among duplicate sample pairs; (4) could not be linked to the 1000 Genomes Project reference; (5) had a significant deviation in the frequency from the 1000 Genomes Project dataset were also eliminated. Genotypes for around 21 million SNPs were imputed with the 1000 Genomes Project (Phase 3) as the reference panel using a two-stage imputation approach: SHAPEIT2 for phasing and IMPUTE version 2 for imputation. Among them, about 11.5 million of the imputed SNPs that have an imputation accuracy $r^2 > 0.3$ and MAF > 0.005 were included in the current study.

b. iCOGs

46,785 cases and 42,892 controls of European ancestry and 4,760 cases and 5,957 controls of Asian population were genotyped with iCOGs, which is a custom Illumina iSelect genotyping array that is designed to test genetic variation related to three hormone related diseases including breast, ovary and prostate cancer. This array includes SNPs associated with these three cancers in GWAS studies, SNPs associated with breast or ovarian cancer in BRCA1 or BRCA2 carriers, SNPs associated with other traits (e.g. ER-negative breast cancer, serous ovarian cancer and aggressive prostate cancer, cancer survival, ag at menarche, breast density), and functional candidate variants. A total of 211,155 SNPs were included in the array.

Genotypes were called using Illumina's proprietaty GenCall algorithm. Samples that (1) were genotypically not female (XX); (2) had an overall call rate < 95%; (3) had an outlier value for heterozygosity ($P < 1\times10^{-6}$); (4) were discordant with previously determined genotypes; (5) were duplicate were excluded. For first-degree relative pairs, the control or otherwise the individual with a lower call rate was excluded. Ancestry outliers, which were determined based on the first two principal components (>15% minority ancestry), were also excluded. Furthermore, SNPs that (1) had a call rate < 95%; (2) deviated from Hardy-Weinberg equilibrium in controls; (3) for which the genotypes were discordant in more than 2% of duplicate samples across all COGS data were also excluded. For SNPs in each new region that a genome-wide significant association was observed, genotype intensity cluster plots were checked manually. SNPs showed poor clustering were deleted. Genotypes for 11.6 million SNPs were imputed using the 1000 Genomes Project March 2012 release as the reference dataset. For common SNPs (MAF > 0.5), the overall imputation accuracy r^2 is greater than 0.5, and for 37% of them, the r^2 is greater than 0.9. The same imputation exclusion criteria were used as OncoArray data.

c. Asian GWAS data

For a majority of the 5,285 Chinese females of our study (4,985), they were genotyped using Affymetrix Genome-Wide Human SNP Array 6.0 platform(102) except for 300 samples which was genotyped using the Affymetrix GeneChip Mapping 500K Array Set. (103)

In GWAS studies, one negative control and at least three positive QC samples from Coriell Cell Repositories (http://ccr.coriell.org/) were included in each of the 96-well plates for Affymetrix SNP Array 6.0 genotyping. A total of 273 positive QC samples were successfully genotyped with an average concordance of 99.9% with a median value of 100%. Samples that (1) had a genotype call rate per sample of < 95%; (2) were genetically identical (i.e., PI_HAT>0.9) or duplicates; (3) had an assay-determined sex that did not concur with reported sex; (4) were first-or second-degree relatives

(i.e., PI_HAT >0.25); (5) had an outlier value for ethnicity; or (6) had an outlier value for heterozygosity were excluded. Furthermore, SNPs that had (1) a genotype call rate of < 95%; (2) a MAF of < 0.01; (3) concordance with the QC sample were excluded as well. Imputation was performed for autosomal SNPs using MInimac2 with the 1000 Genomes Project Phase 3 as the reference.

d. Asian Exomechip data

We also used the data generated from the Asian Exomechip in our study. The Exomechip is an expanded IIIumina HumanExome-12v1_A Beadchip including 247,870 markers of the original exome array combined with ~60K customer content variants(104). The 247,870 markers included in the original exome array mainly focused on protein-coding regions selected from >12,000 samples with exome and genome sequencing data. Most of those sequenced samples were from European ancestry populations, and around 600 of them were from Asians. Detailed information on SNP contents and characteristics have been described elsewhere(105). Briefly, nonsynonymous variants observed ≥3 times in at least two studies, splicing and stop-altering variants observed ≥2 times in at least two studies, variants observed to be associated with complex traits in previous GWAS, human leukocyte antigen tags, ancestry-informative markers, markers for identity-by-descent estimation, along with random synonymous SNPs were included. For our focus on Asian populations, to improve the coverage for less common variants in our population of interest, ~60K variants based on additional sequencing data of Asians and top variants selected from GWAS were added to the Exomechip as well. Specifically, by using exome sequencing in 581 Shanghai Chinese women from Shanghai breast cancer study (SBCS), exome sequencing in 496 Singapore Chinese, and Asian data in the 1000 Genomes Project, nonsynonymous, splicing, and stop-altering variants observed ≥2 times in any one of these three datasets or once in any two of the three datasets, were added (*N* = 33,342).

Additional 28,637 common variants for various GWAS follow-up and GWAS loci fine-mapping projects were included as well. Overall, this chip covers well common SNPs of the whole genome. All samples were genotyped at the Genome Quebec Innovation Centre (Montreal, Quebec, Canada) following Illumina's protocol. On each 96-well plate, two blinded duplicate samples and two HapMap samples were included for QC. Genotype calling was performed using Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries were determined using study samples. After clustering, approximately 80,000 variants were further manually reviewed and clusters were edited for 27,506 variants. The concordance between HapMap samples genotyped and data from the 1000 Genomes Project (http://www.1000genomes.org/) were evaluated. Principal components analyses (PCA) were conducted based on 3,200 ancestry informative markers (AIMs) on the HumanExome Beadchipusing EIGENSTRAT (http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm) to identify population outliers with 1000 Genomes Project data as the reference. The pair-wise proportion of IBD was also assessed to identify potential genetically identical or unexpected duplicate samples or close relatives. Samples that had (1) a genotype call rate of < 98%, (2) concordance with the QC sample genotyping of < 99%, (3) an outlier value for heterozygosity, or (4) an outlier value for ethnicity, as well as (5) samples that appeared to be closely related, (6) duplicate samples with a consistency of < 99%, and (7) samples whose assay-determined sex did not concur with reported sex were excluded. Furthermore, SNPs that had (1) a MAF = 0, (2) a genotype call rate of < 98%, (3) concordance with the QC sample genotyping of < 98%, or (4) a HWE test P of $<1 \times 10^{-5}$, as well as (5) SNPs that were redundant and (6) SNPs designated by the Illumina Infinium assay design group as “treat with caution” were also excluded from analyses. The imputation methods for exome-chip samples were similar to the GWAS studies in ABCC mentioned above.

e. Asian Mega^{EX} Array data

By leveraging content from carrier screening panels and multiple resources to generate a multipurpose, multiethnic array, the MEGA^{EX} array has substantially improved coverage of relatively rarer variants [minor allele frequency (MAF) between 0.5 - 10%]. The datasets used for selecting variants in the MEGA^{EX} array include Phase 3 of the 1000 Genomes Project, Consortium on Asthma among African-ancestry Populations in the Americas, Population Architecture using Genomics and Epidemiology, T2D-Genes Consortium, OMIM, ACGM, and ClinVar. Overall, it includes about 2,036,060 variants in the array, including over 28K for indels, 466K for exonic variants, 22K missense variants, 21K nonsense variants, 26K synonymous variants, and 31K untranslated region (UTR) variants. Overall, this chip covers well common SNPs of the whole genome.

Genotype calling has been performed using Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries will be determined using study samples. After clustering, variants will be further manually reviewed, and clusters will be edited when needed. Principal components analyses (PCA) will be conducted based on 3,200 ancestry informative markers (AIMs) using EIGENSTRAT (http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm) to identify population outliers with 1000 Genomes Project data as the reference. The pair-wise proportion of IBD will also be assessed to identify potential genetically identical or unexpected duplicate samples or close relatives. Samples that have (1) a genotype call rate of < 98%, (2) an outlier value for heterozygosity, or (3) an outlier value for ethnicity, as well as (4) samples that appear to be closely related, (5) duplicate samples with a consistency of < 99%, and (6) samples whose assay-determined sex do not concur with reported sex will be excluded. Furthermore, SNPs that have (1) a MAF = 0, (2) a genotype call rate of < 98%, or (3) a HWE test P of $<1 \times 10^{-5}$, as well as (4) SNPs that are redundant will also be excluded from analyses. SNPs were imputed using similar methods as the other data from ABCC.

1.3. Non-genetic variables assessment

a. Anthropometric measurements

Anthropometric measurements were taken either after cancer diagnosis for case-control studies or at study recruitment for cohort studies. Measured and self-reported heights and weights were used to calculate body mass index (BMI, kg/m2) at each age level.

b. Menstrual and reproductive factors

Information on menstrual and reproductive factors, including age at menarche, age at menopause, menopausal status, parity (the number of livebirths in a woman's lifetime), parous (yes or no), age at the end of first full time pregnancy, oral contraceptive (OC) use, and hormone replacement therapy (HRT) usage, were collected through structured questionnaires. Menopausal status (premenopausal or postmenopausal) was defined by whether participants reported ceased menstrual periods in the 12 months prior to completing the questionnaire; bleeding due to use of female hormones was not included as menstrual periods.

c. Soy and soy products intake

Soy and soy products intake were measured by self-reported food frequency questionnaires. Food items involved in this category include soy milk, tofu, fried/dried tofu, dried soy beans, soy bean sprouts, and fresh soy beans. Total soy and soy products intake was calculated by grams/day of consumption of soy milk + tofu + fried/dried tofu + dried soy beans+ soy bean sprouts + fresh soy beans* 63%.

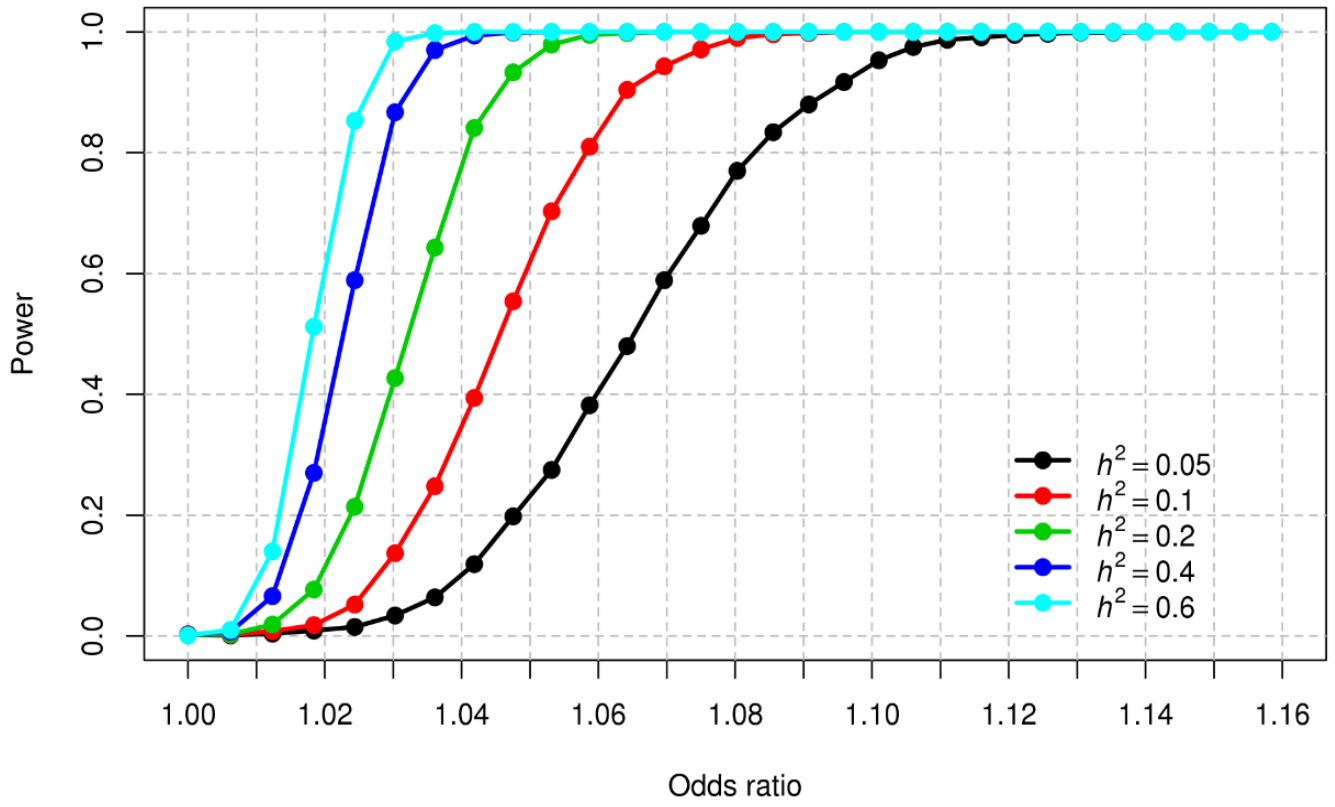


Figure 10. Expected power per one standard deviation change in the gene expression among European descendants.

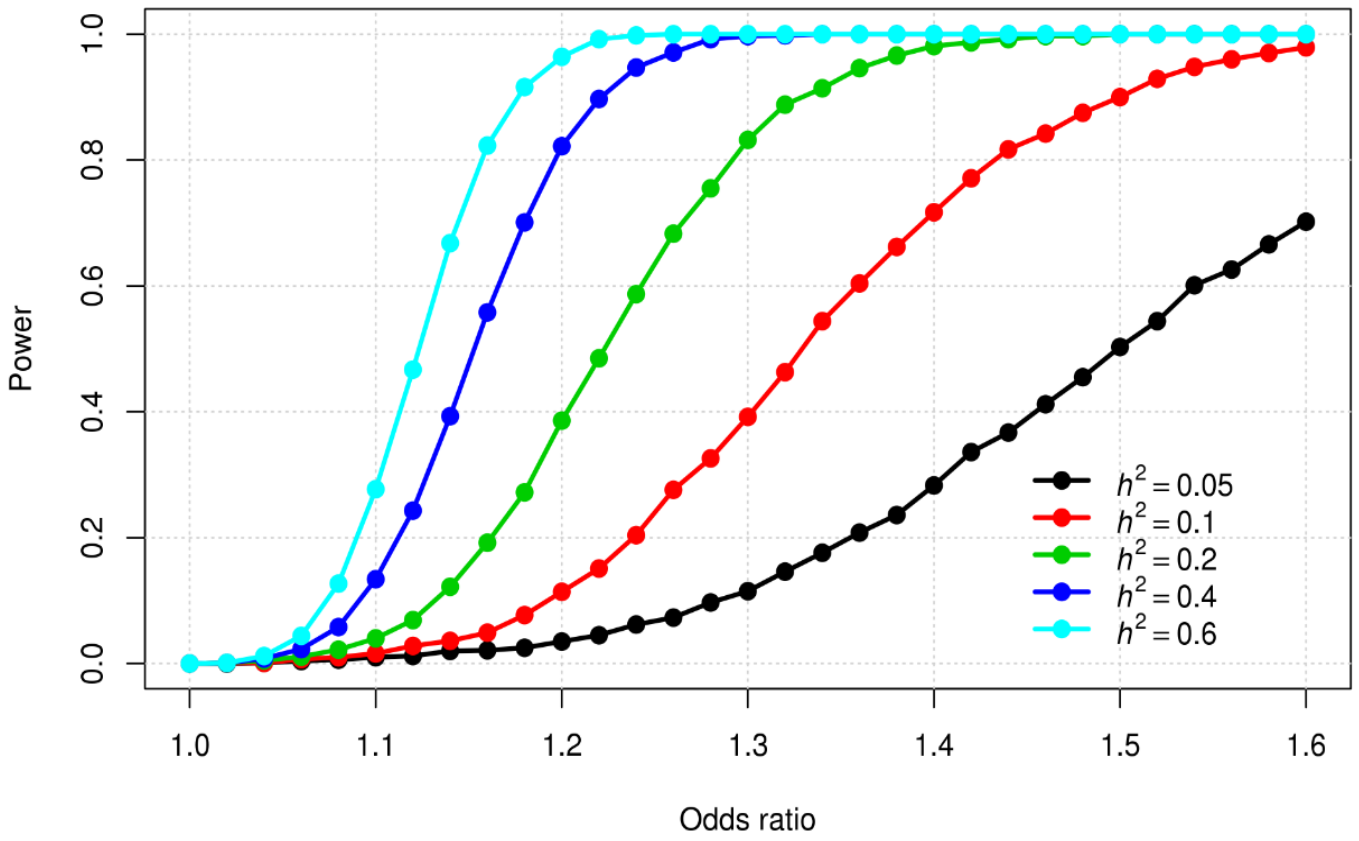


Figure 11. Expected power per one standard deviation change in the gene expression among Asian population.

1.4. Power calculation

We performed two simulation analyses based on 122,977 cases and 105,974 controls among European descendants (Figure 5), and 28,310 cases and 27,933 controls among Asian population (Figure 6). The gene expression was generated from the empirical distribution of predicted gene expression levels in the *Breast Cancer* Association Consortium (BCAC). Statistical power was calculated at $P<7.25\times10^{-4}$ (0.05/69) according to cis-heritability (h^2) which we aim to capture using gene expression prediction models (R^2). The figure shows results per one standard deviation increase (or decrease) in the gene expression based on 1000 replicates. Based on the power calculation, our association analysis is expected to have 80% power to detect a minimum odds ratio (OR) for the association between gene expression level and breast cancer risk of 1.03 or 1.06 for breast cancer risk per one standard deviation (SD) change in the expression level of a gene whose cis-heritability is 40% among Europeans and Asians, respectively.

1.5. Statistical analyses

PrediXcan method were used to generate the predicted expression levels of estrogen synthesis, metabolism and bioavailability pathway genes in breast cancer cases and controls based on the built prediction models and genotyped or imputed SNP information of these subjects. Only SNPs with an imputation quality $R^2 \geq 0.3$ were used. After establishing the genetically predicted gene expression, associations between the predicted gene expression and breast cancer risk were estimated using logistic regression separately for European populations, adjusting for covariates of age, PCs,

and study site. For Asian population, since for the studies from Japan and Korea, only summarized statistics instead of individual-level SNP were available, MetaXcan methods is used to perform the association analyses(106). The formula for MetaXcan is as the follows:

$$Zg \approx \sum_{l \epsilon Modelg} w_{tg} \frac{\hat{\sigma}_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{se(\beta_l)}$$

FDR-adjusted P-values were calculated and compared to the results from the traditional threshold of 0.05.

2. Results for European data

2.1. Main results

The associations of predicted expression of each gene with breast cancer risk across the tested prediction models were shown in Tables 11-16. Overall, there are 16 genes demonstrated a significant association at $P < 0.05$. For example, *ARSA* showed an inverse association based on both breast tissue model (beta=-0.04; $p=2.33 \times 10^{-5}$) and visceral adipose model (beta=-0.09; $p=5.95 \times 10^{-5}$); *ARSJ* showed a positive association based on visceral adipose model (beta=0.06; $p=4.18 \times 10^{-2}$); *B3GAT2* and *CYP3A5* showed positive associations based on breast tissue model (beta=0.04; $p=8.82 \times 10^{-3}$); *CYP19A1* showed a positive association based on ovary tissue model (beta=0.09; $p=9.93 \times 10^{-3}$); *CYP1B1* showed a negative association based on ovary tissue model (beta=-0.04; $p=5.70 \times 10^{-3}$); *GSTM1* showed an inverse association based on both visceral adipose model (beta=-0.04; $p=2.16 \times 10^{-2}$) and ovary model (beta=-0.09; $p=1.77 \times 10^{-3}$); *HSD17B11* showed a negative association based on subcutaneous adipose model (beta=-0.03; $p=6.33 \times 10^{-3}$); *SULT1A1* showed a positive association

based on cross tissue model (beta=0.67; p=1.53 × 10^{-2}); *SULT1A4* showed a negative association based on cross tissue model (beta=-0.18; p=2.43 × 10^{-2}); *SULT2B1* and *UGT2B4* showed a negative association based on liver model (beta=-0.02; p=1.58 × 10^{-2}; beta=-0.03; p=1.00 × 10^{-2} respectively); *UGT1A6* showed a positive association based on liver model (beta=0.04; p=2.06 × 10^{-2}); and *UGT2B7* showed a negative association based on ovary model (beta=-0.06; p=3.12 × 10^{-2}). Some of the associations tended to be restricted to post-menopausal women or pre-menopausal women only.

Table 11. Associations between predicted expression levels of estrogen-related genes in breast tissue and breast cancer risk

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.03	0.01	1.17E-02	-0.06	0.02	2.26E-04	-0.04	0.01	2.33E-05	1.61E-03	--
SULT4A1	0.06	0.02	1.14E-02	0.02	0.02	1.60E-01	0.04	0.01	8.82E-03	1.73E-01	++
CYP3A5	0.46	0.13	2.59E-04	0.06	0.07	4.27E-01	0.15	0.06	1.39E-02	1.82E-01	++
B3GAT2	-0.07	0.03	2.33E-02	-0.01	0.03	6.70E-01	-0.05	0.02	4.34E-02	2.93E-01	--
HSD3B2	-0.20	0.09	2.11E-02	-0.03	0.06	5.72E-01	-0.08	0.05	8.27E-02	4.67E-01	--
ARSK	-0.18	0.37	6.21E-01	0.60	0.25	1.59E-02	0.36	0.21	8.47E-02	4.67E-01	-+
ARSF	-0.46	0.21	2.57E-02	-0.08	0.08	3.44E-01	-0.12	0.07	9.28E-02	4.68E-01	--
UGT2B7	-0.26	0.09	3.69E-03	0.02	0.07	7.60E-01	-0.09	0.06	1.11E-01	5.30E-01	-+
B3GAT3	0.03	0.03	3.18E-01	0.02	0.02	3.91E-01	0.02	0.02	2.03E-01	6.26E-01	++
GALNS	0.10	0.06	8.90E-02	0.01	0.03	6.74E-01	0.03	0.03	2.41E-01	6.65E-01	++
SHBG	0.05	0.08	5.46E-01	0.06	0.06	3.29E-01	0.06	0.05	2.55E-01	6.85E-01	++
IDS	0.00	0.03	8.74E-01	0.09	0.04	3.14E-02	0.03	0.02	2.75E-01	6.90E-01	-+
SULT1B1	-0.08	0.34	8.26E-01	-0.27	0.23	2.48E-01	-0.21	0.19	2.81E-01	6.90E-01	--
GSTM1	0.04	0.05	3.37E-01	-0.07	0.03	5.05E-02	-0.03	0.03	3.22E-01	7.08E-01	+-
CYP1A2	0.01	0.01	4.79E-01	0.01	0.01	5.25E-01	0.01	0.01	3.43E-01	7.08E-01	++
UGT2B11	-0.08	0.05	1.14E-01	0.00	0.04	9.43E-01	-0.03	0.03	3.53E-01	7.08E-01	-+
CYP1A1	-0.02	0.08	7.93E-01	-0.04	0.04	3.64E-01	-0.03	0.04	3.58E-01	7.08E-01	--
COMT	-0.05	0.09	5.46E-01	-0.04	0.06	4.97E-01	-0.04	0.05	3.69E-01	7.08E-01	--

SGSH	0.04	0.04	3.05E-01	0.01	0.03	7.79E-01	0.02	0.02	3.85E-01	7.22E-01	++
GSTA1	0.03	0.04	4.01E-01	0.01	0.03	6.68E-01	0.02	0.03	3.87E-01	7.22E-01	++
UGT2B17	0.01	0.02	6.43E-01	-0.02	0.02	1.78E-01	-0.01	0.01	4.50E-01	7.54E-01	+-
HSD17B1	0.00	0.04	9.56E-01	0.03	0.03	3.29E-01	0.02	0.02	4.68E-01	7.60E-01	-+
ARSB	-0.05	0.10	6.19E-01	-0.03	0.07	6.24E-01	-0.04	0.06	4.96E-01	7.87E-01	--
SULT1A2	-0.05	0.04	2.05E-01	0.04	0.02	1.33E-01	0.01	0.02	5.77E-01	8.01E-01	-+
ARSE	-0.27	0.11	1.71E-02	0.00	0.02	9.21E-01	-0.01	0.02	6.77E-01	8.67E-01	-+
SULT2B1	0.02	0.03	5.44E-01	0.00	0.02	9.55E-01	0.01	0.02	6.91E-01	8.67E-01	++
CYP3A4	0.09	0.09	3.07E-01	-0.04	0.04	3.86E-01	-0.01	0.04	7.30E-01	8.88E-01	+-
SULT1C2	0.00	0.02	9.82E-01	0.00	0.01	9.45E-01	0.00	0.01	9.44E-01	9.72E-01	++

Table 12. Associations between predicted expression levels of estrogen-related genes in liver tissue and breast cancer risk

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
UGT2B4	-0.02	0.02	1.77E-01	-0.04	0.02	2.36E-02	-0.03	0.01	1.00E-02	1.73E-01	--
SULT2B1	-0.03	0.01	3.22E-02	-0.02	0.01	1.82E-01	-0.02	0.01	1.58E-02	1.82E-01	--
UGT1A6	0.07	0.03	5.51E-03	0.01	0.03	6.78E-01	0.04	0.02	2.06E-02	1.99E-01	++
SULT4A1	-0.04	0.03	1.74E-01	-0.03	0.02	9.05E-02	-0.03	0.01	3.07E-02	2.39E-01	--
UGT2B28	0.10	0.07	1.37E-01	0.07	0.05	1.84E-01	0.08	0.04	5.01E-02	3.15E-01	++
CYP1B1	-0.03	0.04	4.72E-01	-0.03	0.02	2.06E-01	-0.03	0.02	1.46E-01	6.26E-01	--
SULT1A1	-0.03	0.42	9.35E-01	-0.39	0.24	1.07E-01	-0.30	0.21	1.50E-01	6.26E-01	--
STAR	-0.05	0.03	1.41E-01	-0.02	0.03	5.69E-01	-0.03	0.02	1.51E-01	6.26E-01	--
B3GAT1	0.01	0.03	7.79E-01	-0.05	0.03	6.98E-02	-0.03	0.02	2.40E-01	6.65E-01	+-
B3GAT2	0.03	0.02	6.79E-02	0.00	0.01	9.60E-01	0.01	0.01	2.64E-01	6.86E-01	++
GSTM1	-0.09	0.07	1.96E-01	-0.02	0.04	6.20E-01	-0.04	0.03	2.89E-01	6.90E-01	--
HSD17B11	0.14	0.25	5.59E-01	0.11	0.12	3.76E-01	0.12	0.11	2.93E-01	6.90E-01	++
UGT2B17	0.01	0.02	8.23E-01	-0.03	0.02	1.41E-01	-0.01	0.01	2.99E-01	6.90E-01	+-
ARSA	-0.04	0.04	2.50E-01	0.06	0.03	4.13E-02	0.02	0.02	3.48E-01	7.08E-01	-+
AKR1B15	-0.11	0.51	8.35E-01	0.50	0.39	1.95E-01	0.28	0.31	3.66E-01	7.08E-01	-+
COMT	-0.04	0.04	2.76E-01	-0.01	0.02	6.81E-01	-0.02	0.02	3.68E-01	7.08E-01	--
SGSH	0.04	0.18	8.01E-01	0.15	0.16	3.56E-01	0.10	0.12	3.93E-01	7.23E-01	++

HSD3B2	0.00	0.02	7.93E-01	-0.02	0.01	2.20E-01	-0.01	0.01	3.99E-01	7.25E-01	+-
ARSB	-0.01	0.01	3.85E-01	0.00	0.01	7.88E-01	-0.01	0.01	4.59E-01	7.54E-01	--
ARSJ	-0.15	0.11	1.55E-01	0.00	0.05	9.80E-01	-0.03	0.04	5.36E-01	8.01E-01	--
GSTA1	0.03	0.03	3.44E-01	0.00	0.02	9.90E-01	0.01	0.02	5.86E-01	8.01E-01	++
B3GAT3	0.03	0.03	3.14E-01	-0.01	0.02	8.19E-01	0.01	0.02	6.54E-01	8.67E-01	+-
SULF2	0.00	0.05	9.37E-01	0.01	0.03	8.22E-01	0.01	0.02	8.14E-01	9.35E-01	++
UGT1A4	0.00	0.01	7.96E-01	0.00	0.01	9.26E-01	0.00	0.01	8.24E-01	9.35E-01	--
CYP11A1	0.00	0.21	9.97E-01	-0.01	0.12	9.23E-01	-0.01	0.11	9.35E-01	9.72E-01	+-
ESR2	-0.02	0.03	4.90E-01	0.03	0.03	3.59E-01	0.00	0.02	9.39E-01	9.72E-01	-+
HSD17B14	0.03	0.04	4.15E-01	-0.03	0.04	4.50E-01	0.00	0.03	9.53E-01	9.74E-01	+-
ARSE	0.09	0.13	4.83E-01	0.00	0.03	8.62E-01	0.00	0.02	9.66E-01	9.76E-01	+-
CYP1A2	0.01	0.02	7.44E-01	0.00	0.01	8.39E-01	0.00	0.01	9.83E-01	9.83E-01	+-

Table 13. Associations between predicted expression levels of estrogen-related genes in ovary tissue and breast cancer risk

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
GSTM1	-0.05	0.05	2.88E-01	-0.11	0.04	1.82E-03	-0.09	0.03	1.77E-03	8.16E-02	--
CYP1B1	-0.01	0.02	5.42E-01	-0.05	0.02	2.29E-03	-0.04	0.01	5.70E-03	1.73E-01	--
CYP19A1	0.14	0.06	1.75E-02	0.06	0.04	1.48E-01	0.09	0.03	9.93E-03	1.73E-01	++
UGT2B7	-0.08	0.05	7.45E-02	-0.05	0.04	1.87E-01	-0.06	0.03	3.12E-02	2.39E-01	--
HSD17B14	-0.08	0.04	7.31E-02	-0.02	0.03	5.62E-01	-0.04	0.03	1.25E-01	5.77E-01	--
SULT1C2	-0.04	0.05	3.91E-01	-0.05	0.05	2.47E-01	-0.05	0.03	1.55E-01	6.26E-01	--
ARSB	0.13	0.10	2.05E-01	0.07	0.10	4.86E-01	0.10	0.07	1.65E-01	6.26E-01	++
ARSA	-0.02	0.01	4.54E-02	0.00	0.01	8.15E-01	-0.01	0.01	1.91E-01	6.26E-01	--
ARSG	0.01	0.05	8.88E-01	-0.05	0.03	1.15E-01	-0.03	0.03	2.10E-01	6.29E-01	+-
CYP3A4	0.05	0.14	7.42E-01	0.11	0.11	3.34E-01	0.08	0.09	3.38E-01	7.08E-01	++
B3GAT2	0.01	0.03	7.06E-01	0.02	0.02	4.46E-01	0.01	0.02	4.10E-01	7.29E-01	++
ARSJ	0.01	0.03	8.40E-01	0.02	0.02	3.86E-01	0.01	0.02	4.13E-01	7.29E-01	++
CYP11A1	0.03	0.02	2.03E-01	-0.01	0.02	7.44E-01	0.01	0.01	5.55E-01	8.01E-01	+-
ARSD	-0.04	0.10	6.80E-01	-0.03	0.08	6.85E-01	-0.04	0.06	5.65E-01	8.01E-01	--
HSD17B1	-0.01	0.04	7.93E-01	-0.02	0.03	6.15E-01	-0.01	0.03	5.79E-01	8.01E-01	--

SULT1A2	0.01	0.04	8.02E-01	-0.02	0.03	5.02E-01	-0.01	0.02	6.65E-01	8.67E-01	+-
SULT1A1	0.05	0.04	2.53E-01	-0.01	0.03	7.88E-01	0.01	0.02	6.72E-01	8.67E-01	+-
ARSE	0.01	0.07	8.92E-01	-0.03	0.06	6.44E-01	-0.01	0.04	7.70E-01	9.17E-01	+-
CYP1A1	0.03	0.05	4.96E-01	-0.01	0.03	8.59E-01	0.01	0.03	8.16E-01	9.35E-01	+-
IDS	0.00	0.05	9.79E-01	0.01	0.05	8.93E-01	0.00	0.03	9.41E-01	9.72E-01	-+

Table 14. Associations between predicted expression levels of estrogen-related genes in subcutaneous adipose tissue and breast cancer risk

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
HSD17B11	0.01	0.02	6.24E-01	-0.05	0.01	2.90E-04	-0.03	0.01	6.33E-03	1.73E-01	+-
SULT4A1	0.00	0.04	9.87E-01	0.06	0.03	3.41E-02	0.04	0.02	9.50E-02	4.68E-01	-+
GNS	0.09	0.61	8.86E-01	0.87	0.55	1.11E-01	0.52	0.41	2.00E-01	6.26E-01	++
HSD3B2	0.16	0.11	1.42E-01	0.02	0.10	8.29E-01	0.09	0.07	2.40E-01	6.65E-01	++
B3GAT2	-0.04	0.02	4.04E-02	0.00	0.01	7.86E-01	-0.01	0.01	3.00E-01	6.90E-01	-+
NQO1	-0.03	0.02	8.66E-02	0.00	0.01	9.85E-01	-0.01	0.01	3.55E-01	7.08E-01	--
CYP1B1	0.00	0.11	9.92E-01	-0.09	0.09	3.02E-01	-0.06	0.07	4.17E-01	7.29E-01	+-
ARSG	0.25	0.20	2.09E-01	-0.30	0.16	5.20E-02	-0.09	0.12	4.58E-01	7.54E-01	+-
GALNS	-0.02	0.03	5.58E-01	-0.01	0.02	7.04E-01	-0.01	0.02	5.08E-01	7.88E-01	--
SULF2	-0.05	0.07	5.23E-01	0.05	0.04	2.79E-01	0.02	0.04	5.43E-01	8.01E-01	-+
SULT1A4	-0.01	0.02	5.64E-01	0.00	0.02	9.85E-01	0.00	0.01	7.25E-01	8.88E-01	-+
STAR	0.02	0.02	3.41E-01	0.00	0.01	6.68E-01	0.00	0.01	8.93E-01	9.72E-01	+-
SULT1A1	-0.01	0.01	2.39E-01	0.01	0.01	3.56E-01	0.00	0.01	9.32E-01	9.72E-01	-+

Table 15. Associations between predicted expression levels of estrogen-related genes in visceral adipose tissue and breast cancer risk

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.05	0.03	9.49E-02	-0.14	0.03	1.74E-06	-0.09	0.02	5.95E-06	8.21E-04	--
SULT4A1	-0.40	0.17	2.02E-02	-0.25	0.22	2.71E-01	-0.34	0.14	1.19E-02	1.82E-01	--

GSTM1	-0.03	0.03	2.61E-01	-0.04	0.02	4.31E-02	-0.04	0.02	2.16E-02	1.99E-01	--
ARSJ	0.02	0.06	7.39E-01	0.07	0.03	3.24E-02	0.06	0.03	4.18E-02	2.93E-01	++
SULT1C2	-0.06	0.03	3.07E-02	-0.02	0.02	4.18E-01	-0.04	0.02	4.46E-02	2.93E-01	--
UGT2B4	0.11	0.07	1.35E-01	0.07	0.05	2.20E-01	0.08	0.04	5.98E-02	3.59E-01	++
ESR1	-0.02	0.05	7.09E-01	-0.06	0.03	7.80E-02	-0.05	0.03	9.25E-02	4.68E-01	--
ARSF	0.05	0.03	1.34E-01	0.01	0.02	4.90E-01	0.02	0.02	1.65E-01	6.26E-01	++
COMT	-0.04	0.03	1.39E-01	-0.01	0.03	6.70E-01	-0.03	0.02	1.73E-01	6.26E-01	--
SULF1	0.05	0.26	8.58E-01	-0.29	0.17	9.07E-02	-0.19	0.14	1.89E-01	6.26E-01	+-
SULT1A2	0.11	0.05	2.48E-02	0.00	0.04	9.05E-01	0.04	0.03	2.04E-01	6.26E-01	+-
CYP1B1	0.12	0.04	2.26E-03	-0.06	0.04	1.02E-01	0.03	0.03	2.99E-01	6.90E-01	+-
CYP3A4	-0.01	0.02	6.35E-01	-0.01	0.02	4.24E-01	-0.01	0.01	3.53E-01	7.08E-01	--
SULT1B1	-0.13	0.13	3.16E-01	-0.03	0.07	6.96E-01	-0.05	0.06	4.24E-01	7.31E-01	--
SULT1E1	-0.02	0.03	4.62E-01	0.00	0.02	8.87E-01	-0.01	0.02	5.69E-01	8.01E-01	--
HSD17B1	-0.03	0.03	2.86E-01	0.01	0.02	7.34E-01	-0.01	0.02	6.50E-01	8.67E-01	-+
ESR2	-0.17	0.22	4.42E-01	0.11	0.12	3.63E-01	0.04	0.11	6.80E-01	8.67E-01	-+
IDS	-0.10	0.05	7.24E-02	0.04	0.03	1.69E-01	0.01	0.03	7.32E-01	8.88E-01	-+
SULF2	-0.03	0.04	4.24E-01	0.01	0.03	8.28E-01	-0.01	0.02	7.34E-01	8.88E-01	-+
SULT2B1	-0.02	0.05	6.64E-01	0.00	0.04	9.01E-01	0.00	0.03	8.87E-01	9.72E-01	-+
HSD17B14	0.01	0.04	8.81E-01	-0.01	0.04	7.87E-01	0.00	0.03	9.08E-01	9.72E-01	+-
CYP1A1	0.99	0.86	2.52E-01	-0.64	0.74	3.88E-01	0.05	0.56	9.23E-01	9.72E-01	+-
ARSD	0.27	0.13	3.75E-02	-0.02	0.04	5.78E-01	0.00	0.04	9.26E-01	9.72E-01	+-

Table 16. Associations between predicted expression levels of estrogen-related genes and breast cancer risk using cross tissue prediction models

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
SULT1A1	0.95	0.33	3.53E-03	-0.05	0.52	9.26E-01	0.67	0.28	1.53E-02	1.82E-01	++
SULT1C2	0.04	0.02	4.65E-02	0.02	0.02	1.57E-01	0.03	0.01	2.14E-02	1.99E-01	--
SULT1A4	-0.20	0.13	1.26E-01	-0.16	0.10	9.53E-02	-0.18	0.08	2.43E-02	2.09E-01	--
STAR	0.17	0.13	1.78E-01	0.07	0.30	8.20E-01	0.15	0.12	1.84E-01	6.26E-01	++
SULT2B1	-0.05	0.02	3.49E-02	0.00	0.01	9.17E-01	-0.02	0.01	1.99E-01	6.26E-01	++
ARSA	-0.04	0.02	1.06E-01	0.03	0.05	6.15E-01	-0.03	0.02	2.04E-01	6.26E-01	--

UGT2B7	0.24	0.18	1.84E-01	0.06	0.08	4.59E-01	0.09	0.07	2.26E-01	6.63E-01	++
SULF1	-2.56	1.83	1.62E-01	0.25	0.19	1.99E-01	0.22	0.19	2.58E-01	6.85E-01	--
CYP17A1	0.13	0.08	1.23E-01	0.01	0.06	8.47E-01	0.05	0.05	3.10E-01	7.02E-01	++
GALNS	0.01	0.05	9.08E-01	-0.04	0.03	2.91E-01	-0.02	0.03	4.39E-01	7.47E-01	-+
SULT4A1	0.05	0.85	9.54E-01	-0.63	0.69	3.59E-01	-0.36	0.54	4.99E-01	7.87E-01	-+
SULF2	-0.06	0.21	7.74E-01	0.15	0.15	3.03E-01	0.08	0.12	5.02E-01	7.87E-01	--
GSTM1	0.00	0.03	9.06E-01	-0.01	0.02	4.97E-01	-0.01	0.01	5.19E-01	7.96E-01	+-
CYP3A4	-0.06	0.08	4.31E-01	0.07	0.05	2.21E-01	0.03	0.04	5.55E-01	8.01E-01	+-
SULT1A2	-0.07	0.04	5.79E-02	0.02	0.03	4.68E-01	-0.01	0.02	5.69E-01	8.01E-01	-+
CYP1B1	0.04	0.15	8.08E-01	0.05	0.10	6.23E-01	0.04	0.08	5.85E-01	8.01E-01	--
CYP11A1	-0.11	0.06	9.59E-02	0.02	0.04	5.91E-01	-0.02	0.03	6.56E-01	8.67E-01	-+
HSD17B14	-0.03	0.09	7.01E-01	-0.01	0.07	8.39E-01	-0.02	0.06	6.90E-01	8.67E-01	+-
STS	-0.08	0.18	6.84E-01	-0.01	0.15	9.63E-01	-0.03	0.12	7.71E-01	9.17E-01	+-
AKR1B15	-0.07	0.08	4.27E-01	0.04	0.05	4.53E-01	0.01	0.04	8.23E-01	9.35E-01	-+
ARSI	-0.62	0.34	6.81E-02	0.24	0.24	3.11E-01	-0.04	0.20	8.24E-01	9.35E-01	--
UGT2B17	0.04	0.06	5.72E-01	-0.02	0.07	7.83E-01	0.01	0.05	8.27E-01	9.35E-01	+-
SULT1E1	0.00	0.04	9.03E-01	0.00	0.07	9.82E-01	0.00	0.03	9.25E-01	9.72E-01	-+
B3GAT1	0.04	0.06	5.34E-01	-0.04	0.06	4.48E-01	0.00	0.04	9.34E-01	9.72E-01	-+
ARSD	-0.55	0.39	1.64E-01	0.13	0.20	5.12E-01	-0.01	0.18	9.69E-01	9.76E-01	-+

Table 17. Summary for the significant genes (P value less than 0.05) for the association with breast cancer risk from prediction models built for different tissues

Breast			Liver			Subcutaneous adipose			Visceral adipose		
Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value
ARSA	-0.04	2.33E-05	*UGT2B4*	-0.03	1.00E-02	*HSD17B11*	-0.03	6.33E-03	*ARSA*	-0.09	5.95E-06
SULT4A1	0.04	8.82E-03	*SULT2B1*	-0.02	1.58E-02				*SULT4A1*	-0.34	1.19E-02
CYP3A5	0.15	1.39E-02	*UGT1A6*	0.04	2.06E-02				*GSTM1*	-0.04	2.16E-02
B3GAT2	-0.05	4.34E-02	*SULT4A1*	-0.03	3.07E-02				*ARSJ*	0.06	4.18E-02
									SULT1C2	-0.04	4.46E-02

Ovary			Cross		
Gene	Effect	P-value	Gene	Effect	P-value
GSTM1	-0.09	1.77E-03	*SULT1A1*	0.67	1.53E-02
CYP1B1	-0.04	5.70E-03	*SULT1C2*	0.03	2.14E-02
CYP19A1	0.09	9.93E-03	*SULT1A4*	-0.18	2.43E-02
UGT2B7	-0.06	3.12E-02			

Gene	Breast		Liver		Subcutaneous adipose		Visceral adipose		Ovary		Cross	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.04	2.33E-05					-0.09	5.95E-06				
ARSJ							0.06	4.18E-02				
B3GAT2	0.04	8.82E-03										
CYP19A1									0.09	9.93E-03		
CYP1B1									-0.04	5.70E-03		
CYP3A5	0.04	8.82E-03										
GSTM1							-0.04	2.16E-02	-0.09	1.77E-03		
HSD17B11					-0.03	6.33E-03						
SULT1A1											0.67	1.53E-02
SULT1A4											-0.18	2.43E-02
SULT1C2							-0.04	4.46E-02			0.03	2.14E-02
SULT2B1			-0.02	1.58E-02								
SULT4A1	0.04	8.82E-03	-0.03	3.07E-02			-0.34	1.19E-02				
UGT1A6			0.04	2.06E-02								
UGT2B4			-0.03	1.00E-02								
UGT2B7									-0.06	3.12E-02		

2.2. Results for associations according to ER/PR status.

Table 18. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in breast tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.03	0.01	2.20E-02	-0.06	0.02	1.24E-03	-0.04	0.01	1.49E-04	1.03E-02	--
SGSH	0.06	0.04	1.57E-01	0.08	0.04	2.03E-02	0.08	0.03	6.98E-03	1.34E-01	++
UGT2B7	-0.31	0.10	2.70E-03	-0.05	0.08	5.65E-01	-0.15	0.06	2.15E-02	2.27E-01	--
CYP3A5	0.53	0.15	2.60E-04	0.04	0.08	6.40E-01	0.15	0.07	3.34E-02	2.77E-01	++
ARSF	-0.28	0.24	2.37E-01	-0.15	0.09	9.31E-02	-0.16	0.08	4.71E-02	3.03E-01	--
SULT4A1	0.06	0.03	2.00E-02	0.01	0.02	4.48E-01	0.03	0.02	5.11E-02	3.03E-01	++
ARSK	-0.23	0.43	5.95E-01	0.60	0.28	3.19E-02	0.35	0.23	1.30E-01	5.70E-01	-+
B3GAT2	-0.03	0.03	3.16E-01	-0.03	0.04	4.29E-01	-0.03	0.03	2.02E-01	6.64E-01	--
UGT2B17	-0.02	0.03	5.50E-01	-0.02	0.02	3.01E-01	-0.02	0.02	2.37E-01	6.99E-01	--
UGT2B11	-0.14	0.06	2.19E-02	0.01	0.05	7.83E-01	-0.04	0.04	2.38E-01	6.99E-01	-+
GSTA1	0.04	0.05	3.80E-01	0.03	0.04	4.36E-01	0.03	0.03	2.48E-01	7.03E-01	++
HSD17B1	0.01	0.04	8.47E-01	0.04	0.03	2.34E-01	0.03	0.03	2.91E-01	7.03E-01	++
ARSE	-0.24	0.13	6.01E-02	-0.02	0.03	5.34E-01	-0.03	0.03	3.18E-01	7.07E-01	--
GALNS	0.10	0.07	1.47E-01	0.01	0.04	7.75E-01	0.03	0.03	3.59E-01	7.26E-01	++
SULT1A2	-0.01	0.04	8.32E-01	0.03	0.03	2.27E-01	0.02	0.02	3.65E-01	7.26E-01	-+
B3GAT3	0.02	0.03	5.88E-01	0.01	0.02	5.38E-01	0.02	0.02	4.16E-01	7.66E-01	++
SULT1B1	0.07	0.40	8.68E-01	-0.26	0.26	3.05E-01	-0.17	0.22	4.41E-01	7.91E-01	+-
IDS	-0.01	0.03	7.37E-01	0.08	0.05	8.69E-02	0.02	0.03	4.69E-01	8.02E-01	-+
ARSB	-0.08	0.12	5.27E-01	-0.03	0.07	6.68E-01	-0.04	0.06	4.85E-01	8.02E-01	--
CYP1A1	-0.03	0.09	7.23E-01	-0.03	0.05	5.52E-01	-0.03	0.04	4.89E-01	8.02E-01	--
HSD3B2	-0.20	0.10	5.07E-02	0.03	0.06	6.60E-01	-0.04	0.05	5.04E-01	8.02E-01	-+
SULT2B1	0.01	0.04	8.55E-01	0.02	0.03	4.96E-01	0.01	0.02	5.06E-01	8.02E-01	++
COMT	-0.04	0.10	7.09E-01	-0.03	0.06	6.17E-01	-0.03	0.05	5.34E-01	8.10E-01	--
SHBG	-0.02	0.09	7.92E-01	0.06	0.07	3.85E-01	0.03	0.06	6.01E-01	8.42E-01	-+
GSTM1	0.08	0.05	1.30E-01	-0.07	0.04	8.57E-02	-0.02	0.03	6.09E-01	8.42E-01	+-
SULT1C2	-0.01	0.02	6.99E-01	0.00	0.01	7.54E-01	-0.01	0.01	6.40E-01	8.42E-01	--
CYP1A2	0.00	0.01	8.88E-01	0.01	0.01	5.44E-01	0.00	0.01	7.27E-01	9.17E-01	-+

CYP3A4	0.12	0.10	2.09E-01	-0.04	0.05	3.67E-01	-0.01	0.04	7.67E-01	9.36E-01	+-

Table 19. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in breast tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.03	0.02	6.94E-02	-0.08	0.02	6.72E-05	-0.05	0.01	9.01E-05	6.73E-03	--
CYP3A5	0.56	0.17	8.27E-04	0.11	0.09	1.93E-01	0.20	0.08	7.27E-03	1.45E-01	++
SGSH	0.10	0.05	5.98E-02	0.08	0.04	5.32E-02	0.08	0.03	7.37E-03	1.45E-01	++
UGT2B7	-0.34	0.12	4.20E-03	-0.07	0.09	4.45E-01	-0.17	0.07	1.97E-02	2.52E-01	--
CYP1A1	-0.04	0.10	6.62E-01	-0.10	0.05	5.92E-02	-0.09	0.05	6.11E-02	4.44E-01	--
SULT4A1	0.05	0.03	1.02E-01	0.02	0.02	3.40E-01	0.03	0.02	8.86E-02	5.04E-01	++
ARSF	-0.28	0.27	3.00E-01	-0.14	0.10	1.60E-01	-0.15	0.09	9.49E-02	5.04E-01	--
ARSB	-0.08	0.14	5.33E-01	-0.11	0.08	1.78E-01	-0.10	0.07	1.40E-01	6.22E-01	--
SULT1A2	-0.01	0.05	8.39E-01	0.05	0.03	7.88E-02	0.04	0.03	1.64E-01	6.48E-01	-+
GSTA1	0.05	0.05	3.86E-01	0.04	0.04	3.17E-01	0.04	0.03	1.86E-01	6.60E-01	++
B3GAT3	0.04	0.04	3.24E-01	0.02	0.02	4.20E-01	0.03	0.02	2.25E-01	7.48E-01	++
GALNS	0.14	0.08	6.34E-02	0.01	0.04	7.10E-01	0.04	0.03	2.47E-01	7.92E-01	++
HSD17B1	0.02	0.05	6.60E-01	0.04	0.04	2.93E-01	0.03	0.03	2.69E-01	8.14E-01	++
ARSK	-0.20	0.49	6.79E-01	0.42	0.31	1.69E-01	0.25	0.26	3.43E-01	8.23E-01	-+
COMT	-0.04	0.11	7.19E-01	-0.04	0.07	5.39E-01	-0.04	0.06	4.77E-01	8.89E-01	--
HSD3B2	-0.16	0.11	1.73E-01	0.11	0.07	9.96E-02	0.04	0.06	4.83E-01	8.89E-01	-+
B3GAT2	-0.02	0.04	5.18E-01	-0.01	0.04	8.23E-01	-0.02	0.03	5.27E-01	9.10E-01	--
CYP3A4	0.11	0.11	3.11E-01	-0.05	0.05	2.73E-01	-0.03	0.05	5.53E-01	9.13E-01	+-
UGT2B17	-0.01	0.03	6.35E-01	-0.01	0.02	7.83E-01	-0.01	0.02	6.14E-01	9.13E-01	--
SULT1B1	0.04	0.45	9.21E-01	-0.17	0.28	5.52E-01	-0.11	0.24	6.52E-01	9.13E-01	+-
IDS	-0.04	0.04	2.80E-01	0.11	0.05	3.38E-02	0.01	0.03	6.91E-01	9.13E-01	-+
ARSE	-0.18	0.15	2.24E-01	0.00	0.03	8.76E-01	-0.01	0.03	6.92E-01	9.13E-01	--
GSTM1	0.10	0.06	1.05E-01	-0.06	0.04	1.31E-01	-0.01	0.03	7.52E-01	9.13E-01	+-
UGT2B11	-0.11	0.07	1.24E-01	0.04	0.05	4.41E-01	-0.01	0.04	7.66E-01	9.19E-01	-+
SHBG	-0.04	0.11	7.01E-01	0.05	0.08	5.66E-01	0.01	0.06	8.20E-01	9.43E-01	-+

SULT2B1	0.02	0.05	5.87E-01	-0.01	0.03	8.39E-01	0.00	0.02	8.98E-01	9.77E-01	+-
CYP1A2	0.00	0.01	6.65E-01	0.00	0.01	6.70E-01	0.00	0.01	9.75E-01	9.82E-01	-+
SULT1C2	0.00	0.03	9.39E-01	0.00	0.02	9.66E-01	0.00	0.01	9.99E-01	9.99E-01	+-

Table 20. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in liver tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
SULT4A1	-0.05	0.03	8.43E-02	-0.04	0.02	3.82E-02	-0.04	0.02	7.47E-03	1.34E-01	--
UGT2B4	-0.03	0.02	1.34E-01	-0.04	0.02	3.98E-02	-0.04	0.01	1.14E-02	1.58E-01	--
UGT2B28	0.11	0.08	1.29E-01	0.10	0.06	7.81E-02	0.10	0.04	2.05E-02	2.27E-01	++
STAR	-0.11	0.04	6.83E-03	-0.01	0.04	7.34E-01	-0.06	0.03	3.61E-02	2.77E-01	--
UGT1A6	0.07	0.03	1.77E-02	0.01	0.03	6.31E-01	0.04	0.02	4.13E-02	3.00E-01	++
B3GAT2	0.04	0.02	6.99E-02	0.01	0.02	4.78E-01	0.02	0.01	1.04E-01	4.80E-01	++
SULT2B1	-0.02	0.02	2.56E-01	-0.01	0.01	3.87E-01	-0.01	0.01	1.64E-01	6.22E-01	--
GSTM1	-0.13	0.08	9.78E-02	-0.03	0.04	5.14E-01	-0.05	0.04	1.75E-01	6.22E-01	--
AKR1B15	0.37	0.58	5.26E-01	0.49	0.43	2.53E-01	0.45	0.34	1.95E-01	6.64E-01	++
UGT1A4	-0.01	0.02	5.41E-01	-0.01	0.01	3.59E-01	-0.01	0.01	2.70E-01	7.03E-01	--
HSD17B11	0.24	0.29	4.11E-01	0.11	0.14	4.09E-01	0.14	0.12	2.72E-01	7.03E-01	++
ARSA	-0.02	0.04	5.69E-01	0.05	0.03	8.11E-02	0.03	0.03	2.81E-01	7.03E-01	-+
ARSE	0.03	0.16	8.52E-01	-0.03	0.03	2.64E-01	-0.03	0.03	2.86E-01	7.03E-01	+-
SULF2	0.04	0.06	4.69E-01	0.02	0.03	4.48E-01	0.03	0.03	3.15E-01	7.07E-01	++
UGT2B17	-0.01	0.03	8.37E-01	-0.02	0.02	3.08E-01	-0.01	0.02	3.44E-01	7.26E-01	--
CYP1B1	-0.03	0.04	4.94E-01	-0.01	0.03	7.79E-01	-0.01	0.02	5.48E-01	8.14E-01	--
SULT1A1	0.40	0.49	4.17E-01	-0.30	0.27	2.75E-01	-0.13	0.24	5.75E-01	8.42E-01	+-
B3GAT1	0.03	0.04	4.56E-01	-0.04	0.03	1.82E-01	-0.01	0.03	5.92E-01	8.42E-01	+-
HSD17B14	0.02	0.04	6.37E-01	-0.05	0.04	2.62E-01	-0.01	0.03	6.44E-01	8.42E-01	+-
ESR2	-0.02	0.03	5.57E-01	0.04	0.03	1.92E-01	0.01	0.02	6.65E-01	8.50E-01	-+
COMT	-0.04	0.04	3.16E-01	0.00	0.02	8.45E-01	-0.01	0.02	7.51E-01	9.34E-01	-+
HSD3B2	0.01	0.02	5.56E-01	0.00	0.01	9.18E-01	0.00	0.01	8.02E-01	9.44E-01	+-
CYP1A2	0.01	0.02	7.05E-01	-0.01	0.01	6.22E-01	0.00	0.01	8.44E-01	9.44E-01	+-
B3GAT3	0.03	0.03	4.55E-01	-0.01	0.03	7.40E-01	0.00	0.02	8.49E-01	9.44E-01	+-
ARSJ	-0.13	0.12	2.95E-01	0.01	0.05	7.90E-01	-0.01	0.05	8.56E-01	9.44E-01	-+
CYP11A1	0.17	0.25	4.93E-01	-0.03	0.14	8.10E-01	0.02	0.12	8.99E-01	9.46E-01	+-
SGSH	0.04	0.21	8.33E-01	-0.06	0.18	7.44E-01	-0.01	0.13	9.14E-01	9.47E-01	+-
GSTA1	0.04	0.04	3.51E-01	-0.02	0.03	5.03E-01	0.00	0.02	9.60E-01	9.66E-01	+-
ARSB	0.00	0.01	8.32E-01	0.00	0.01	9.20E-01	0.00	0.01	9.65E-01	9.66E-01	-+

Table 21. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in liver tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
STAR	-0.14	0.05	2.69E-03	-0.04	0.04	3.63E-01	-0.08	0.03	7.27E-03	1.45E-01	--
UGT1A6	0.08	0.03	1.56E-02	0.03	0.03	4.10E-01	0.05	0.02	2.19E-02	2.52E-01	++
UGT2B28	0.13	0.09	1.31E-01	0.10	0.06	9.31E-02	0.11	0.05	2.49E-02	2.64E-01	++
SULT4A1	-0.04	0.03	2.81E-01	-0.03	0.02	1.93E-01	-0.03	0.02	9.45E-02	5.04E-01	--
UGT2B4	-0.01	0.02	5.69E-01	-0.03	0.02	1.57E-01	-0.02	0.02	1.50E-01	6.45E-01	--
SULF2	0.07	0.07	3.13E-01	0.02	0.03	4.62E-01	0.03	0.03	2.69E-01	8.14E-01	++
AKR1B15	0.64	0.66	3.32E-01	0.31	0.47	5.12E-01	0.42	0.39	2.72E-01	8.14E-01	++
B3GAT2	0.04	0.03	1.43E-01	0.00	0.02	7.81E-01	0.01	0.01	2.93E-01	8.14E-01	++
CYP1B1	-0.06	0.05	2.04E-01	-0.01	0.03	6.60E-01	-0.03	0.03	3.02E-01	8.14E-01	--
CYP1A2	-0.02	0.02	4.00E-01	-0.01	0.02	5.08E-01	-0.01	0.01	3.11E-01	8.14E-01	--
UGT1A4	-0.01	0.02	7.23E-01	-0.01	0.01	3.45E-01	-0.01	0.01	3.23E-01	8.14E-01	--
HSD17B11	0.48	0.33	1.44E-01	0.06	0.15	6.80E-01	0.13	0.14	3.24E-01	8.14E-01	++
ESR2	-0.02	0.03	5.75E-01	0.07	0.04	5.67E-02	0.02	0.03	3.76E-01	8.23E-01	-+
B3GAT3	0.03	0.04	3.84E-01	0.01	0.03	8.48E-01	0.02	0.02	5.02E-01	8.89E-01	++
CYP11A1	0.17	0.28	5.35E-01	0.04	0.15	7.67E-01	0.07	0.13	5.78E-01	9.13E-01	++
ARSE	-0.11	0.18	5.41E-01	-0.01	0.03	6.74E-01	-0.02	0.03	6.04E-01	9.13E-01	--
HSD3B2	-0.02	0.02	4.11E-01	0.00	0.02	9.94E-01	-0.01	0.01	6.52E-01	9.13E-01	-+
UGT2B17	0.00	0.03	9.97E-01	-0.01	0.02	6.18E-01	-0.01	0.02	6.84E-01	9.13E-01	+-
HSD17B14	0.04	0.05	4.91E-01	-0.06	0.05	2.54E-01	-0.01	0.04	7.28E-01	9.13E-01	+-
GSTM1	-0.03	0.09	7.66E-01	-0.01	0.05	8.37E-01	-0.01	0.04	7.49E-01	9.13E-01	--
SULT2B1	-0.01	0.02	6.53E-01	0.00	0.01	9.68E-01	0.00	0.01	7.54E-01	9.13E-01	--
COMT	-0.05	0.05	2.79E-01	0.01	0.02	7.74E-01	-0.01	0.02	7.95E-01	9.42E-01	-+
ARSA	-0.04	0.05	4.68E-01	0.03	0.03	4.22E-01	0.01	0.03	8.09E-01	9.42E-01	-+
B3GAT1	0.01	0.04	7.84E-01	-0.02	0.04	5.96E-01	-0.01	0.03	8.12E-01	9.42E-01	+-
SULT1A1	0.16	0.56	7.78E-01	0.02	0.30	9.55E-01	0.05	0.26	8.55E-01	9.64E-01	++
ARSB	0.00	0.02	9.99E-01	0.00	0.01	8.61E-01	0.00	0.01	8.86E-01	9.77E-01	--
SGSH	0.02	0.23	9.23E-01	-0.04	0.20	8.45E-01	-0.01	0.15	9.31E-01	9.77E-01	+-
ARSJ	-0.14	0.14	3.27E-01	0.03	0.06	6.16E-01	0.00	0.06	9.42E-01	9.77E-01	-+
GSTA1	0.02	0.05	5.99E-01	-0.01	0.03	6.84E-01	0.00	0.02	9.49E-01	9.77E-01	+-

Table 22. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in ovary tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
GSTM1	-0.04	0.06	4.24E-01	-0.15	0.04	2.02E-04	-0.11	0.03	5.16E-04	2.37E-02	--
CYP19A1	0.18	0.07	8.91E-03	0.10	0.05	3.82E-02	0.13	0.04	1.34E-03	4.63E-02	++
CYP1B1	-0.04	0.02	1.34E-01	-0.04	0.02	2.73E-02	-0.04	0.01	7.76E-03	1.34E-01	--
UGT2B7	-0.09	0.05	1.16E-01	-0.06	0.04	1.38E-01	-0.07	0.03	3.24E-02	2.77E-01	--
ARSJ	0.03	0.03	3.22E-01	0.03	0.02	2.63E-01	0.03	0.02	1.38E-01	5.75E-01	++
HSD17B14	-0.09	0.05	5.73E-02	-0.01	0.04	7.21E-01	-0.04	0.03	1.64E-01	6.22E-01	--
SULT1C2	-0.01	0.05	8.66E-01	-0.09	0.05	1.00E-01	-0.05	0.04	1.97E-01	6.64E-01	--
ARSB	0.12	0.12	3.19E-01	0.09	0.12	4.44E-01	0.10	0.08	2.14E-01	6.87E-01	++
ARSG	0.05	0.06	4.08E-01	-0.06	0.03	7.09E-02	-0.03	0.03	2.69E-01	7.03E-01	+-
IDS	-0.05	0.05	3.62E-01	-0.03	0.05	5.44E-01	-0.04	0.04	2.83E-01	7.03E-01	--
CYP11A1	0.03	0.02	2.50E-01	0.00	0.02	8.55E-01	0.01	0.02	3.79E-01	7.26E-01	++
ARSA	-0.03	0.01	3.56E-02	0.00	0.01	7.51E-01	-0.01	0.01	3.84E-01	7.26E-01	-+
CYP3A4	0.09	0.16	5.90E-01	0.07	0.12	5.61E-01	0.08	0.10	4.30E-01	7.80E-01	++
ARSD	0.00	0.12	9.69E-01	-0.08	0.09	3.67E-01	-0.05	0.07	4.94E-01	8.02E-01	+-
SULT1A1	0.01	0.05	9.00E-01	-0.03	0.03	3.90E-01	-0.02	0.03	5.16E-01	8.05E-01	+-
B3GAT2	0.01	0.03	7.74E-01	0.01	0.02	6.80E-01	0.01	0.02	6.17E-01	8.42E-01	++
ARSE	0.06	0.09	4.91E-01	-0.07	0.06	2.69E-01	-0.02	0.05	6.18E-01	8.42E-01	+-
HSD17B1	0.01	0.05	9.07E-01	-0.02	0.04	6.62E-01	-0.01	0.03	7.86E-01	9.44E-01	+-
SULT1A2	0.00	0.05	9.30E-01	0.00	0.03	8.68E-01	0.00	0.03	8.51E-01	9.44E-01	--
CYP1A1	0.04	0.05	4.14E-01	-0.02	0.03	5.63E-01	0.00	0.03	9.66E-01	9.66E-01	+-

Table 23. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in ovary tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
CYP1B1	-0.07	0.03	8.60E-03	-0.05	0.02	1.42E-02	-0.06	0.02	4.09E-04	1.88E-02	--
GSTM1	-0.05	0.06	4.75E-01	-0.14	0.04	1.79E-03	-0.11	0.04	3.03E-03	1.05E-01	--
UGT2B7	-0.07	0.06	2.62E-01	-0.11	0.05	2.26E-02	-0.09	0.04	1.27E-02	2.19E-01	--
ARSJ	0.03	0.04	4.63E-01	0.04	0.03	8.68E-02	0.04	0.02	6.75E-02	4.44E-01	++
CYP19A1	0.15	0.08	5.37E-02	0.04	0.05	4.22E-01	0.08	0.04	7.92E-02	4.97E-01	++
ARSG	0.05	0.06	4.01E-01	-0.10	0.04	1.19E-02	-0.06	0.03	8.31E-02	4.98E-01	+-
B3GAT2	0.01	0.03	7.94E-01	0.03	0.03	1.83E-01	0.02	0.02	2.27E-01	7.48E-01	++
ARSB	0.14	0.14	3.07E-01	0.09	0.13	4.87E-01	0.11	0.09	2.28E-01	7.48E-01	++
CYP11A1	0.03	0.03	2.29E-01	0.01	0.02	7.03E-01	0.02	0.02	2.91E-01	8.14E-01	++
IDS	-0.06	0.06	3.35E-01	-0.02	0.06	7.74E-01	-0.04	0.04	3.80E-01	8.23E-01	--
CYP3A4	0.13	0.18	4.76E-01	0.08	0.14	5.85E-01	0.10	0.11	3.88E-01	8.23E-01	++
HSD17B14	-0.08	0.06	1.72E-01	0.00	0.04	9.54E-01	-0.02	0.03	4.67E-01	8.89E-01	-+
ARSA	-0.04	0.02	4.92E-03	0.01	0.01	3.47E-01	-0.01	0.01	4.94E-01	8.89E-01	-+
ARSE	0.00	0.10	9.80E-01	-0.05	0.07	4.53E-01	-0.04	0.06	5.28E-01	9.10E-01	--
SULT1A2	-0.02	0.06	7.23E-01	-0.01	0.03	6.56E-01	-0.02	0.03	5.73E-01	9.13E-01	--
ARSD	0.04	0.13	7.66E-01	-0.08	0.10	4.44E-01	-0.03	0.08	6.68E-01	9.13E-01	+-
SULT1A1	0.02	0.05	7.41E-01	0.01	0.03	8.08E-01	0.01	0.03	7.03E-01	9.13E-01	++
SULT1C2	0.03	0.06	6.75E-01	-0.05	0.06	3.91E-01	-0.01	0.04	7.40E-01	9.13E-01	+-
CYP1A1	0.06	0.06	3.67E-01	-0.04	0.04	3.51E-01	-0.01	0.03	7.53E-01	9.13E-01	+-
HSD17B1	-0.01	0.05	8.41E-01	-0.01	0.04	8.76E-01	-0.01	0.03	8.06E-01	9.42E-01	--

Table 24. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in subcutaneous adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
HSD17B11	0.02	0.02	3.92E-01	-0.05	0.01	2.81E-04	-0.03	0.01	9.81E-03	1.50E-01	+-
SULT4A1	-0.04	0.04	3.91E-01	0.07	0.03	1.43E-02	0.04	0.02	1.44E-01	5.84E-01	-+
GNS	-0.18	0.70	8.01E-01	1.02	0.61	9.31E-02	0.51	0.46	2.69E-01	7.03E-01	-+
NQO1	-0.04	0.02	6.41E-02	0.00	0.01	9.62E-01	-0.01	0.01	3.60E-01	7.26E-01	-+
HSD3B2	0.11	0.12	3.76E-01	0.04	0.11	6.92E-01	0.07	0.08	3.73E-01	7.26E-01	++
GALNS	-0.01	0.03	7.11E-01	-0.02	0.02	5.02E-01	-0.01	0.02	4.48E-01	7.93E-01	--
ARSG	0.25	0.23	2.77E-01	-0.30	0.17	8.41E-02	-0.10	0.14	4.79E-01	8.02E-01	+-
SULT1A4	-0.01	0.02	5.33E-01	-0.01	0.02	7.46E-01	-0.01	0.01	5.22E-01	8.05E-01	--
B3GAT2	-0.04	0.02	5.41E-02	0.01	0.02	5.23E-01	-0.01	0.01	5.25E-01	8.05E-01	-+
SULF2	-0.10	0.08	2.55E-01	0.00	0.05	9.24E-01	-0.02	0.04	6.31E-01	8.42E-01	-+
STAR	0.01	0.02	5.53E-01	0.00	0.01	8.52E-01	0.00	0.01	6.44E-01	8.42E-01	++
CYP1B1	0.08	0.13	5.34E-01	-0.07	0.10	4.65E-01	-0.02	0.08	8.29E-01	9.44E-01	+-
SULT1A1	-0.01	0.01	5.32E-01	0.01	0.01	5.09E-01	0.00	0.01	8.40E-01	9.44E-01	-+

Table 25. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in subcutaneous adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
HSD17B11	0.038	0.025	1.30E-01	-0.042	0.016	6.68E-03	-0.020	0.013	1.34E-01	6.22E-01	+-
SULT4A1	-0.047	0.047	3.11E-01	0.061	0.033	6.84E-02	0.024	0.027	3.71E-01	8.23E-01	-+
GNS	-0.238	0.801	7.66E-01	0.927	0.667	1.65E-01	0.450	0.513	3.81E-01	8.23E-01	-+
ARSG	0.373	0.258	1.49E-01	-0.383	0.192	4.63E-02	-0.114	0.154	4.59E-01	8.89E-01	+-
STAR	-0.003	0.024	9.10E-01	0.012	0.014	3.86E-01	0.008	0.012	4.86E-01	8.89E-01	-+
HSD3B2	0.033	0.140	8.15E-01	0.074	0.125	5.52E-01	0.056	0.093	5.49E-01	9.13E-01	++
CYP1B1	0.227	0.150	1.30E-01	-0.040	0.107	7.10E-01	0.050	0.087	5.64E-01	9.13E-01	+-
SULT1A4	0.002	0.027	9.55E-01	-0.016	0.021	4.50E-01	-0.009	0.017	5.71E-01	9.13E-01	+-
NQO1	-0.021	0.022	3.41E-01	0.001	0.013	9.22E-01	-0.004	0.011	6.93E-01	9.13E-01	-+
GALNS	0.000	0.036	9.96E-01	-0.010	0.026	7.01E-01	-0.007	0.021	7.54E-01	9.13E-01	--
B3GAT2	-0.018	0.023	4.31E-01	0.006	0.017	7.36E-01	-0.003	0.014	8.53E-01	9.64E-01	-+
SULT1A1	0.000	0.013	9.88E-01	0.002	0.009	8.26E-01	0.001	0.007	8.60E-01	9.64E-01	-+
SULF2	-0.012	0.095	8.98E-01	-0.006	0.052	9.03E-01	-0.008	0.046	8.66E-01	9.64E-01	--

Table 26. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in visceral adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.03	0.03	3.44E-01	-0.15	0.03	6.57E-06	-0.09	0.02	1.11E-04	1.03E-02	--
GSTM1	-0.03	0.03	4.50E-01	-0.07	0.02	3.37E-03	-0.06	0.02	4.52E-03	1.25E-01	--
ARSF	0.07	0.04	4.93E-02	0.03	0.02	1.25E-01	0.05	0.02	1.95E-02	2.27E-01	++
SULT1C2	-0.05	0.03	1.63E-01	-0.03	0.03	1.94E-01	-0.04	0.02	5.91E-02	3.26E-01	--
ARSJ	-0.03	0.07	6.58E-01	0.08	0.04	2.52E-02	0.06	0.03	7.61E-02	4.04E-01	-+
UGT2B4	0.10	0.08	2.06E-01	0.07	0.06	2.30E-01	0.08	0.05	8.58E-02	4.39E-01	++
SULT4A1	-0.31	0.20	1.14E-01	-0.17	0.25	5.05E-01	-0.26	0.15	9.80E-02	4.75E-01	--
COMT	-0.02	0.03	4.42E-01	-0.04	0.03	2.45E-01	-0.03	0.02	1.73E-01	6.22E-01	--
CYP3A4	-0.01	0.03	5.66E-01	-0.02	0.02	2.80E-01	-0.02	0.01	2.24E-01	6.99E-01	--
ESR1	0.03	0.06	5.49E-01	-0.06	0.04	7.61E-02	-0.04	0.03	2.35E-01	6.99E-01	+-
SULF1	0.08	0.30	7.94E-01	-0.26	0.19	1.71E-01	-0.16	0.16	3.10E-01	7.07E-01	+-
ESR2	-0.21	0.25	4.10E-01	0.22	0.14	1.11E-01	0.12	0.12	3.15E-01	7.07E-01	-+
SULT1E1	-0.04	0.03	2.67E-01	-0.01	0.03	6.81E-01	-0.02	0.02	3.16E-01	7.07E-01	--
CYP1B1	0.12	0.04	7.65E-03	-0.06	0.04	1.94E-01	0.03	0.03	3.45E-01	7.26E-01	+-
SULT1A2	0.11	0.06	4.48E-02	-0.01	0.04	7.36E-01	0.03	0.03	3.57E-01	7.26E-01	+-
IDS	-0.01	0.06	8.14E-01	0.04	0.03	2.48E-01	0.03	0.03	3.62E-01	7.26E-01	-+
ARSD	0.29	0.15	5.06E-02	-0.06	0.05	1.70E-01	-0.03	0.04	4.55E-01	7.95E-01	+-
SULT1B1	-0.14	0.15	3.49E-01	-0.02	0.07	7.39E-01	-0.05	0.07	4.79E-01	8.02E-01	--
HSD17B1	-0.04	0.03	2.08E-01	0.01	0.03	7.85E-01	-0.01	0.02	5.43E-01	8.14E-01	-+
HSD17B14	0.00	0.05	9.38E-01	0.01	0.04	8.49E-01	0.01	0.03	8.43E-01	9.44E-01	++
SULF2	-0.06	0.04	1.60E-01	0.03	0.04	3.59E-01	0.00	0.03	8.79E-01	9.46E-01	-+
CYP1A1	1.27	0.99	1.99E-01	-0.73	0.83	3.76E-01	0.09	0.63	8.88E-01	9.46E-01	+-
SULT2B1	-0.02	0.06	6.85E-01	0.02	0.04	6.82E-01	0.00	0.03	9.05E-01	9.46E-01	-+

Table 27. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in visceral adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
ARSA	-0.04	0.04	3.09E-01	-0.16	0.04	9.63E-06	-0.10	0.03	9.76E-05	6.73E-03	--
ARSF	0.10	0.04	1.30E-02	0.03	0.03	1.98E-01	0.05	0.02	1.75E-02	2.52E-01	++
GSTM1	-0.02	0.04	5.55E-01	-0.06	0.03	1.82E-02	-0.05	0.02	2.20E-02	2.52E-01	--
CYP3A4	-0.06	0.03	4.36E-02	-0.02	0.02	2.01E-01	-0.03	0.02	3.06E-02	3.02E-01	--
UGT2B4	0.15	0.09	1.07E-01	0.09	0.07	1.69E-01	0.11	0.05	3.99E-02	3.44E-01	++
COMT	-0.04	0.04	2.88E-01	-0.05	0.03	1.29E-01	-0.05	0.03	6.75E-02	4.44E-01	--
ARSJ	-0.08	0.08	2.98E-01	0.09	0.04	2.08E-02	0.06	0.04	1.12E-01	5.54E-01	-+
SULF1	-0.04	0.34	9.18E-01	-0.35	0.21	9.50E-02	-0.27	0.18	1.39E-01	6.22E-01	--
SULT2B1	0.02	0.07	7.90E-01	0.06	0.04	1.38E-01	0.05	0.04	1.63E-01	6.48E-01	++
SULT4A1	-0.30	0.23	1.77E-01	-0.14	0.28	6.04E-01	-0.24	0.17	1.70E-01	6.48E-01	--
CYP1B1	0.14	0.05	6.90E-03	-0.05	0.05	2.78E-01	0.04	0.04	2.82E-01	8.14E-01	+-
ESR1	0.00	0.07	9.53E-01	-0.05	0.04	2.09E-01	-0.04	0.03	2.96E-01	8.14E-01	+-
SULT1C2	-0.01	0.04	7.74E-01	-0.03	0.03	2.92E-01	-0.02	0.02	3.13E-01	8.14E-01	--
HSD17B1	-0.02	0.04	5.30E-01	-0.02	0.03	5.35E-01	-0.02	0.02	3.80E-01	8.23E-01	--
SULT1A2	0.13	0.07	3.98E-02	-0.02	0.05	6.88E-01	0.03	0.04	3.86E-01	8.23E-01	+-
SULT1E1	-0.02	0.04	6.16E-01	-0.02	0.03	5.83E-01	-0.02	0.02	4.59E-01	8.89E-01	--
SULT1B1	-0.03	0.18	8.69E-01	-0.06	0.08	4.74E-01	-0.05	0.07	4.72E-01	8.89E-01	--
ARSD	0.15	0.17	3.73E-01	-0.05	0.05	3.37E-01	-0.03	0.05	5.03E-01	8.89E-01	+-
IDS	-0.06	0.07	4.29E-01	0.04	0.04	3.16E-01	0.02	0.03	5.92E-01	9.13E-01	-+
SULF2	-0.09	0.05	8.56E-02	0.03	0.04	4.44E-01	-0.01	0.03	6.66E-01	9.13E-01	-+
ESR2	-0.44	0.29	1.29E-01	0.06	0.15	6.86E-01	-0.05	0.13	7.34E-01	9.13E-01	-+
CYP1A1	1.03	1.13	3.63E-01	-0.57	0.91	5.29E-01	0.06	0.71	9.38E-01	9.77E-01	+-
HSD17B14	-0.02	0.06	7.80E-01	0.01	0.04	9.03E-01	0.00	0.03	9.43E-01	9.77E-01	-+

Table 28. Association between the predicted expression level of estrogen-related genes and ER-positive breast cancer risk using prediction model built in cross tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
SULT1A4	-0.21	0.15	1.64E-01	-0.19	0.11	7.23E-02	-0.20	0.09	2.31E-02	2.27E-01	--
SULT1A1	0.80	0.38	3.29E-02	0.34	0.58	5.56E-01	0.67	0.32	3.49E-02	2.77E-01	++
SULT2B1	-0.06	0.03	1.88E-02	-0.01	0.02	4.05E-01	-0.03	0.01	4.63E-02	3.03E-01	--
ARSA	-0.07	0.03	1.64E-02	0.03	0.06	5.90E-01	-0.05	0.03	5.13E-02	3.03E-01	-+
GALNS	-0.03	0.05	5.99E-01	-0.08	0.04	4.44E-02	-0.06	0.03	5.26E-02	3.03E-01	--
UGT2B7	0.39	0.21	6.19E-02	0.09	0.09	3.17E-01	0.13	0.08	9.98E-02	4.75E-01	++
STAR	0.21	0.15	1.42E-01	0.13	0.33	6.87E-01	0.20	0.13	1.32E-01	5.70E-01	++
CYP17A1	0.22	0.10	1.94E-02	0.01	0.06	9.16E-01	0.07	0.05	1.76E-01	6.22E-01	++
SULT1C2	0.01	0.03	6.51E-01	0.02	0.02	3.08E-01	0.02	0.01	2.72E-01	7.03E-01	++
CYP1B1	0.19	0.18	2.75E-01	0.04	0.11	6.99E-01	0.08	0.09	3.73E-01	7.26E-01	++
SULT4A1	-0.14	0.98	8.88E-01	-0.71	0.77	3.57E-01	-0.49	0.61	4.16E-01	7.66E-01	--
B3GAT1	0.03	0.07	6.04E-01	0.02	0.07	7.99E-01	0.03	0.05	5.85E-01	8.42E-01	++
CYP11A1	-0.08	0.07	3.04E-01	0.06	0.05	2.29E-01	0.02	0.04	6.28E-01	8.42E-01	-+
STS	-0.14	0.21	5.23E-01	-0.01	0.16	9.32E-01	-0.06	0.13	6.47E-01	8.42E-01	--
SULF1	-3.92	2.07	5.89E-02	0.14	0.21	5.20E-01	0.10	0.21	6.56E-01	8.46E-01	-+
AKR1B15	-0.12	0.10	2.16E-01	0.02	0.06	7.41E-01	-0.02	0.05	7.31E-01	9.17E-01	-+
SULF2	-0.07	0.24	7.70E-01	0.09	0.16	5.66E-01	0.04	0.14	7.59E-01	9.36E-01	-+
GSTM1	0.02	0.03	4.27E-01	-0.01	0.02	4.58E-01	0.00	0.01	8.01E-01	9.44E-01	+-
UGT2B17	0.07	0.07	3.49E-01	-0.05	0.07	5.05E-01	0.01	0.05	8.44E-01	9.44E-01	+-
ARSI	-0.49	0.39	2.08E-01	0.29	0.27	2.77E-01	0.04	0.22	8.49E-01	9.44E-01	-+
SULT1E1	0.01	0.04	8.35E-01	-0.01	0.07	9.33E-01	0.01	0.04	8.92E-01	9.46E-01	+-
ARSD	-0.38	0.45	3.98E-01	0.12	0.22	5.76E-01	0.03	0.20	8.92E-01	9.46E-01	-+
HSD17B14	-0.03	0.10	7.58E-01	0.01	0.08	9.46E-01	-0.01	0.06	8.93E-01	9.46E-01	-+
CYP3A4	-0.10	0.09	2.79E-01	0.03	0.06	5.72E-01	-0.01	0.05	9.19E-01	9.47E-01	-+
SULT1A2	-0.06	0.04	1.71E-01	0.03	0.03	2.87E-01	0.00	0.03	9.58E-01	9.66E-01	-+

Table 29. Association between the predicted expression level of estrogen-related genes and PR-positive breast cancer risk using prediction model built in cross tissue.

Gene Name	Icogs			Onco_array			Meta-analysis				Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	FDR-adjusted P	icogs/onco-array
SULT1A1	0.69	0.43	1.09E-01	0.87	0.64	1.72E-01	0.75	0.36	3.65E-02	3.36E-01	++
GALNS	-0.04	0.06	5.57E-01	-0.08	0.04	5.06E-02	-0.07	0.03	5.22E-02	4.24E-01	--
UGT2B7	0.41	0.24	8.59E-02	0.13	0.10	1.99E-01	0.17	0.09	6.62E-02	4.44E-01	--
SULT1A4	-0.22	0.17	1.87E-01	-0.12	0.12	2.95E-01	-0.16	0.10	1.07E-01	5.45E-01	--
SULT1C2	0.03	0.03	3.03E-01	0.02	0.02	3.42E-01	0.02	0.02	1.74E-01	6.48E-01	++
ARSA	-0.06	0.03	5.74E-02	0.06	0.07	4.08E-01	-0.04	0.03	1.77E-01	6.48E-01	--
SULT2B1	-0.06	0.03	4.93E-02	-0.01	0.02	7.35E-01	-0.02	0.02	1.79E-01	6.48E-01	--
STS	-0.10	0.24	6.75E-01	-0.14	0.18	4.23E-01	-0.13	0.15	3.72E-01	8.23E-01	--
SULF1	-3.86	2.31	9.47E-02	0.25	0.24	2.86E-01	0.21	0.24	3.74E-01	8.23E-01	--
CYP17A1	0.15	0.11	1.57E-01	0.00	0.07	9.81E-01	0.04	0.06	4.42E-01	8.89E-01	++
CYP1B1	0.34	0.20	9.19E-02	-0.02	0.12	8.77E-01	0.07	0.10	4.83E-01	8.89E-01	-+
AKR1B15	-0.14	0.11	2.15E-01	-0.01	0.06	9.32E-01	-0.04	0.06	4.91E-01	8.89E-01	++
HSD17B14	0.00	0.12	9.84E-01	0.06	0.09	5.32E-01	0.04	0.07	6.09E-01	9.13E-01	--
STAR	0.12	0.17	4.72E-01	-0.13	0.36	7.29E-01	0.08	0.15	6.11E-01	9.13E-01	--
SULT4A1	0.20	1.13	8.59E-01	-0.59	0.84	4.85E-01	-0.31	0.68	6.52E-01	9.13E-01	+-
B3GAT1	0.01	0.08	8.46E-01	-0.06	0.07	4.22E-01	-0.02	0.05	6.55E-01	9.13E-01	--
GSTM1	0.03	0.03	4.20E-01	-0.02	0.02	3.43E-01	-0.01	0.02	6.59E-01	9.13E-01	-+
SULT1A2	-0.07	0.05	1.84E-01	0.01	0.03	6.79E-01	-0.01	0.03	6.70E-01	9.13E-01	-+
SULT1E1	-0.01	0.05	9.11E-01	-0.05	0.08	5.82E-01	-0.02	0.04	7.02E-01	9.13E-01	-+
CYP3A4	-0.11	0.11	3.22E-01	0.07	0.06	2.98E-01	0.02	0.06	7.09E-01	9.13E-01	+-
ARSD	-0.26	0.52	6.18E-01	0.09	0.24	7.17E-01	0.03	0.22	9.07E-01	9.77E-01	-+
UGT2B17	0.01	0.08	8.75E-01	-0.02	0.08	7.74E-01	-0.01	0.06	9.24E-01	9.77E-01	+-
ARSI	-0.53	0.45	2.39E-01	0.20	0.29	4.92E-01	-0.02	0.25	9.43E-01	9.77E-01	-+
SULF2	-0.25	0.27	3.54E-01	0.10	0.18	5.72E-01	-0.01	0.15	9.67E-01	9.82E-01	+-
CYP11A1	-0.15	0.09	8.73E-02	0.05	0.05	2.89E-01	0.00	0.04	9.68E-01	9.82E-01	+-

Table 30. Summary results for associations between predicted gene expression level and ER-positive breast cancer.

Breast			Liver			Subcutaneous adipose			Visceral adipose		
Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value
ARSA	-0.04	1.49E-04	SULT4A1	-0.04	7.47E-03	*HSD17B11*	-0.03	9.81E-03	*ARSA*	-0.09	1.11E-04
SGSH	0.08	6.98E-03	UGT2B4	-0.04	1.14E-02				*GSTM1*	-0.06	4.52E-03
UGT2B7	-0.15	2.15E-02	UGT2B28	0.10	2.05E-02				*ARSF*	0.05	1.95E-02
CYP3A5	0.15	3.34E-02	STAR	-0.06	3.61E-02						
ARSF	-0.16	4.71E-02	UGT1A6	0.04	4.13E-02						

Ovary			Cross		
Gene	Effect	P-value	Gene	Effect	P-value
GSTM1	-0.11	5.16E-04	*SULT1A4*	-0.20	2.31E-02
CYP19A1	0.13	1.34E-03	*SULT1A1*	0.67	3.49E-02
CYP1B1	-0.04	7.76E-03	*SULT2B1*	-0.03	4.63E-02
UGT2B7	-0.07	3.24E-02			

Gene	Breast		Liver		Subcutaneous adipose		Visceral adipose		Ovary		Cross	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.04	1.49E-04					-0.09	1.11E-04				
ARSF	-0.16	4.71E-02					0.05	1.95E-02				
CYP19A1									0.13	1.34E-03		
CYP1B1									-0.04	7.76E-03		
CYP3A5	0.15	3.34E-02										
GSTM1							-0.06	4.52E-03	-0.11	5.16E-04		
HSD17B11					-0.03	9.81E-03						
SGSH	0.08	6.98E-03										
STAR			-0.06	3.61E-02								
SULT1A1											0.67	3.49E-02
SULT1A4												
SULT2B1											-0.03	4.63E-02
SULT4A1			-0.04	7.47E-03							-0.20	2.31E-02
UGT1A6			0.04	4.13E-02								
UGT2B28			0.10	2.05E-02								
UGT2B4			-0.04	1.14E-02								
UGT2B7	-0.15	2.15E-02							-0.07	3.24E-02		

Table 31. Summary results for associations between predicted gene expression level and PR-positive breast cancer.

Breast			Liver			Subcutaneous adipose			Visceral adipose		
Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value
ARSA	-0.05	9.01E-05	*STAR*	-0.08	7.27E-03				*ARSA*	-0.10	9.76E-05
CYP3A5	0.20	7.27E-03	*UGT1A6*	0.05	2.19E-02				*ARSF*	0.05	1.75E-02
SGSH	0.08	7.37E-03	*UGT2B28*	0.11	2.49E-02				*GSTM1*	-0.05	2.20E-02
UGT2B7	-0.17	1.97E-02							*CYP3A4*	-0.03	3.06E-02
									UGT2B4	0.11	3.99E-02

Ovary			Cross		
Gene	Effect	P-value	Gene	Effect	P-value
CYP1B1	-0.06	4.09E-04	*SULT1A1*	0.75	3.65E-02
GSTM1	-0.11	3.03E-03			
UGT2B7	-0.09	1.27E-02			

Gene	Breast		Liver		Subcutaneous adipose		Visceral adipose		Ovary		Cross	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.05	9.01E-05					-0.10	9.76E-05				
ARSF							0.05	1.75E-02				
CYP1B1									-0.06	4.09E-04		
CYP3A4							-0.03	3.06E-02				
CYP3A5	0.20	7.27E-03										
GSTM1							-0.05	2.20E-02	-0.11	3.03E-03		
SGSH	0.08	7.37E-03										
STAR			-0.08	7.27E-03								
SULT1A1											SULT1A1	0.75
UGT1A6			0.05	2.19E-02								
UGT2B28			0.11	2.49E-02								
UGT2B4							0.11	3.99E-02				
UGT2B7	-0.17	1.97E-02							-0.09	1.27E-02		

2.3. Results for associations according to menopausal status.

Table 32. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in breast tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
ARSA	-0.04	0.02	2.43E-02	-0.06	0.02	1.45E-02	-0.05	0.02	1.02E-03	--
CYP3A5	0.62	0.19	9.92E-04	0.19	0.10	6.48E-02	0.28	0.09	1.51E-03	++
SGSH	0.16	0.06	4.08E-03	0.02	0.05	7.01E-01	0.07	0.04	3.67E-02	++
ARSK	0.00	0.55	9.97E-01	0.81	0.36	2.35E-02	0.57	0.30	5.76E-02	-+
SULT4A1	0.07	0.04	5.72E-02	0.02	0.02	4.16E-01	0.03	0.02	8.03E-02	++
B3GAT2	-0.06	0.04	1.33E-01	-0.02	0.05	7.05E-01	-0.05	0.03	1.66E-01	--
SULT1B1	-0.18	0.51	7.21E-01	-0.46	0.33	1.64E-01	-0.38	0.28	1.74E-01	--
HSD17B1	0.05	0.06	3.39E-01	0.04	0.04	3.55E-01	0.05	0.03	1.88E-01	++
IDS	0.01	0.04	7.68E-01	0.11	0.06	6.50E-02	0.05	0.03	1.88E-01	++
CYP1A1	-0.07	0.11	5.50E-01	-0.07	0.06	2.45E-01	-0.07	0.05	1.91E-01	--
SULT2B1	0.07	0.05	1.69E-01	0.02	0.03	6.06E-01	0.03	0.03	2.30E-01	++
SHBG	0.06	0.12	6.02E-01	0.10	0.09	2.80E-01	0.09	0.07	2.41E-01	++
ARSB	-0.10	0.15	5.32E-01	-0.08	0.10	4.10E-01	-0.08	0.08	3.03E-01	--
COMT	0.02	0.13	8.79E-01	0.09	0.08	2.84E-01	0.07	0.07	3.23E-01	++
GALNS	0.06	0.09	5.17E-01	0.03	0.05	4.77E-01	0.04	0.04	3.53E-01	++
ARSF	-0.44	0.31	1.53E-01	-0.05	0.11	6.55E-01	-0.10	0.11	3.60E-01	--
HSD3B2	-0.19	0.13	1.39E-01	0.00	0.08	9.85E-01	-0.05	0.07	4.40E-01	-+
B3GAT3	0.01	0.04	8.36E-01	0.02	0.03	5.37E-01	0.02	0.02	5.30E-01	++
CYP1A2	0.00	0.01	8.64E-01	-0.01	0.01	5.27E-01	-0.01	0.01	5.62E-01	--
GSTM1	0.08	0.07	2.10E-01	-0.02	0.05	7.05E-01	0.02	0.04	6.68E-01	+-
UGT2B11	-0.04	0.08	6.43E-01	-0.01	0.06	8.64E-01	-0.02	0.05	6.76E-01	--
CYP3A4	0.03	0.13	7.88E-01	0.01	0.06	8.25E-01	0.02	0.05	7.55E-01	++
ARSE	-0.28	0.17	9.06E-02	0.02	0.04	5.68E-01	0.01	0.03	8.36E-01	-+
UGT2B7	-0.08	0.13	5.32E-01	0.07	0.10	4.81E-01	0.01	0.08	8.67E-01	-+
SULT1A2	-0.05	0.06	3.70E-01	0.03	0.03	4.44E-01	0.01	0.03	8.67E-01	-+
SULT1C2	0.01	0.03	8.44E-01	-0.01	0.02	7.60E-01	0.00	0.02	8.72E-01	+-
GSTA1	0.00	0.06	9.67E-01	0.01	0.05	9.16E-01	0.00	0.04	9.14E-01	++
UGT2B17	0.02	0.03	5.12E-01	-0.01	0.03	6.83E-01	0.00	0.02	9.34E-01	+-

Table 33. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in breast tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
UGT2B11	-0.21	0.12	7.07E-02	-0.24	0.09	9.78E-03	-0.23	0.07	1.64E-03	--
HSD3B2	-0.29	0.20	1.34E-01	-0.27	0.12	2.78E-02	-0.28	0.11	7.80E-03	--
SULT1A2	0.01	0.08	8.62E-01	0.10	0.05	5.26E-02	0.08	0.05	8.51E-02	++
IDS	0.05	0.06	4.19E-01	0.15	0.09	9.45E-02	0.08	0.05	1.05E-01	++
ARSA	-0.02	0.03	5.88E-01	-0.07	0.04	5.31E-02	-0.04	0.02	1.09E-01	--
B3GAT2	-0.10	0.07	1.30E-01	-0.02	0.08	7.52E-01	-0.07	0.05	1.76E-01	--
SULT4A1	0.07	0.05	2.12E-01	0.03	0.04	4.98E-01	0.04	0.03	2.04E-01	++
HSD17B1	0.01	0.08	8.81E-01	0.09	0.07	1.53E-01	0.06	0.05	2.22E-01	++
CYP3A4	0.34	0.19	7.90E-02	0.03	0.09	7.40E-01	0.09	0.08	2.97E-01	++
ARSE	-0.15	0.25	5.42E-01	-0.05	0.05	3.83E-01	-0.05	0.05	3.27E-01	--
GSTA1	0.14	0.09	1.25E-01	-0.01	0.07	9.26E-01	0.05	0.06	3.72E-01	+-
GSTM1	-0.04	0.10	6.92E-01	-0.06	0.08	4.57E-01	-0.05	0.06	4.04E-01	--
SGSH	-0.16	0.09	7.57E-02	0.03	0.07	6.55E-01	-0.04	0.05	4.45E-01	-+
COMT	-0.14	0.20	4.74E-01	-0.05	0.12	6.83E-01	-0.08	0.11	4.67E-01	--
GALNS	0.17	0.13	1.99E-01	0.01	0.07	8.80E-01	0.04	0.06	4.67E-01	++
SHBG	-0.23	0.18	2.05E-01	0.04	0.14	7.93E-01	-0.07	0.11	5.59E-01	-+
B3GAT3	-0.03	0.06	6.55E-01	0.04	0.04	3.34E-01	0.02	0.04	5.85E-01	-+
SULT1C2	0.00	0.04	9.37E-01	0.02	0.03	5.16E-01	0.01	0.02	6.06E-01	-+
CYP1A1	-0.01	0.17	9.43E-01	-0.05	0.09	5.94E-01	-0.04	0.08	6.16E-01	--
CYP3A5	0.00	0.29	9.97E-01	0.08	0.15	5.89E-01	0.06	0.13	6.35E-01	-+
ARSF	-0.32	0.47	5.00E-01	0.11	0.17	5.14E-01	0.06	0.16	7.07E-01	-+
CYP1A2	-0.02	0.02	2.22E-01	0.01	0.02	5.02E-01	0.00	0.01	7.27E-01	-+
UGT2B7	-0.12	0.20	5.37E-01	0.02	0.16	9.11E-01	-0.04	0.13	7.67E-01	-+
SULT1B1	-0.30	0.78	7.00E-01	0.05	0.51	9.25E-01	-0.06	0.43	8.95E-01	-+
SULT2B1	-0.07	0.08	3.35E-01	0.03	0.05	6.04E-01	-0.01	0.04	9.01E-01	-+
ARSB	0.03	0.23	9.06E-01	-0.03	0.15	8.51E-01	-0.01	0.12	9.23E-01	+-
ARSK	-0.05	0.83	9.54E-01	0.06	0.54	9.10E-01	0.03	0.46	9.50E-01	-+
UGT2B17	0.00	0.05	9.28E-01	-0.01	0.04	8.87E-01	0.00	0.03	9.56E-01	+-

Table 34. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in liver tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	icogs/onco-array
UGT2B28	0.20	0.10	3.65E-02	0.14	0.07	5.85E-02	0.16	0.06	5.71E-03	++
UGT2B4	-0.05	0.03	6.14E-02	-0.04	0.02	6.45E-02	-0.05	0.02	8.64E-03	--
SULT4A1	-0.07	0.04	8.27E-02	-0.03	0.02	2.77E-01	-0.04	0.02	6.50E-02	--
UGT1A6	0.10	0.04	8.70E-03	-0.01	0.04	8.69E-01	0.05	0.03	7.55E-02	+-
ARSB	-0.01	0.02	4.82E-01	-0.02	0.01	1.04E-01	-0.02	0.01	8.48E-02	--
GSTA1	0.04	0.05	3.91E-01	0.05	0.03	1.77E-01	0.05	0.03	1.10E-01	++
STAR	-0.11	0.05	2.42E-02	0.00	0.05	9.53E-01	-0.06	0.03	1.14E-01	--
B3GAT1	0.00	0.05	9.91E-01	-0.08	0.04	5.48E-02	-0.05	0.03	1.43E-01	+-
B3GAT3	0.07	0.04	1.09E-01	0.02	0.03	5.85E-01	0.04	0.03	1.58E-01	++
SULT1A1	0.15	0.64	8.16E-01	-0.59	0.35	9.27E-02	-0.42	0.31	1.73E-01	+-
HSD17B11	0.22	0.37	5.43E-01	0.19	0.18	2.70E-01	0.20	0.16	2.09E-01	++
GSTM1	-0.10	0.10	3.26E-01	-0.05	0.06	3.77E-01	-0.06	0.05	2.14E-01	--
COMT	-0.10	0.05	4.69E-02	-0.01	0.03	7.66E-01	-0.03	0.03	2.22E-01	--
CYP1B1	0.02	0.06	7.58E-01	-0.06	0.04	1.16E-01	-0.04	0.03	2.43E-01	+-
CYP11A1	-0.44	0.32	1.70E-01	-0.07	0.18	6.84E-01	-0.16	0.16	3.05E-01	--
ESR2	-0.01	0.04	7.68E-01	0.08	0.04	6.40E-02	0.03	0.03	3.11E-01	-+
SULT2B1	-0.02	0.02	2.62E-01	-0.01	0.02	7.07E-01	-0.01	0.01	3.13E-01	--
ARSE	0.36	0.20	7.35E-02	0.02	0.04	4.99E-01	0.03	0.04	3.27E-01	++
SGSH	-0.02	0.26	9.37E-01	-0.28	0.23	2.30E-01	-0.17	0.17	3.38E-01	--
UGT1A4	-0.02	0.02	3.04E-01	-0.01	0.01	6.96E-01	-0.01	0.01	3.71E-01	--
UGT2B17	0.02	0.03	6.56E-01	-0.03	0.02	1.69E-01	-0.02	0.02	3.95E-01	+-
B3GAT2	0.05	0.03	7.04E-02	-0.01	0.02	6.22E-01	0.01	0.02	5.25E-01	+-
HSD3B2	0.01	0.03	7.46E-01	-0.02	0.02	4.03E-01	-0.01	0.02	6.13E-01	+-
ARSA	-0.06	0.06	2.56E-01	0.06	0.04	1.70E-01	0.01	0.03	6.60E-01	-+
ARSJ	-0.21	0.16	1.74E-01	0.07	0.07	3.34E-01	0.02	0.06	7.44E-01	-+
AKR1B15	-0.51	0.75	4.95E-01	0.08	0.55	8.89E-01	-0.13	0.44	7.70E-01	-+
HSD17B14	0.07	0.06	2.06E-01	-0.05	0.06	3.62E-01	0.01	0.04	8.04E-01	+-
CYP1A2	-0.01	0.03	7.12E-01	0.01	0.02	6.37E-01	0.00	0.02	8.53E-01	-+
SULF2	-0.04	0.08	6.39E-01	0.01	0.04	7.30E-01	0.00	0.03	9.26E-01	-+

Table 35. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in liver tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	Effect	SE	P	Effect	SE	P	icogs/onco-array
B3GAT3	-0.13	0.07	6.51E-02	-0.11	0.05	3.67E-02	-0.12	0.04	5.43E-03	--
SGSH	0.45	0.41	2.71E-01	0.70	0.35	4.74E-02	0.59	0.27	2.65E-02	++
UGT1A4	0.05	0.03	1.72E-01	0.03	0.02	1.30E-01	0.04	0.02	4.33E-02	++
HSD3B2	-0.05	0.04	2.02E-01	-0.05	0.03	1.16E-01	-0.05	0.02	4.35E-02	--
ARSE	-0.06	0.30	8.34E-01	-0.08	0.06	1.38E-01	-0.08	0.05	1.34E-01	--
CYP1B1	0.04	0.09	6.36E-01	0.08	0.05	1.59E-01	0.07	0.05	1.49E-01	++
ARSA	0.00	0.08	9.57E-01	0.11	0.06	7.62E-02	0.07	0.05	1.64E-01	-+
UGT2B17	-0.04	0.05	4.95E-01	-0.04	0.04	2.99E-01	-0.04	0.03	2.15E-01	--
SULT4A1	-0.06	0.06	3.43E-01	-0.03	0.04	4.20E-01	-0.04	0.03	2.35E-01	--
SULT2B1	-0.04	0.03	2.23E-01	-0.01	0.03	6.68E-01	-0.02	0.02	2.63E-01	--
UGT2B4	0.00	0.04	9.21E-01	-0.06	0.04	1.15E-01	-0.03	0.03	2.73E-01	+-
ARSJ	-0.29	0.24	2.36E-01	-0.04	0.11	6.93E-01	-0.08	0.10	3.98E-01	--
GSTA1	0.00	0.08	9.69E-01	-0.05	0.05	3.04E-01	-0.04	0.04	4.03E-01	+-
AKR1B15	0.22	1.13	8.47E-01	0.75	0.85	3.78E-01	0.56	0.68	4.12E-01	++
ARSB	-0.04	0.03	9.58E-02	0.01	0.02	7.00E-01	-0.01	0.02	4.92E-01	-+
ESR2	0.03	0.06	6.25E-01	0.03	0.07	6.90E-01	0.03	0.04	5.29E-01	++
COMT	0.06	0.08	4.21E-01	-0.04	0.04	3.44E-01	-0.02	0.04	6.65E-01	+-
SULF2	0.07	0.11	5.20E-01	0.01	0.06	9.17E-01	0.02	0.05	6.96E-01	++
UGT2B28	-0.03	0.15	8.58E-01	0.05	0.11	6.33E-01	0.02	0.09	7.84E-01	-+
STAR	0.02	0.08	8.37E-01	-0.04	0.07	5.73E-01	-0.01	0.05	7.88E-01	+-
UGT1A6	0.00	0.06	9.39E-01	0.02	0.06	7.63E-01	0.01	0.04	7.93E-01	++
SULT1A1	-0.31	0.95	7.42E-01	0.25	0.54	6.35E-01	0.12	0.47	8.01E-01	-+
HSD17B11	0.63	0.56	2.58E-01	-0.21	0.27	4.46E-01	-0.05	0.24	8.46E-01	+-
HSD17B14	0.00	0.09	9.95E-01	-0.02	0.09	7.92E-01	-0.01	0.06	8.56E-01	+-
B3GAT2	0.04	0.04	4.03E-01	-0.02	0.03	4.36E-01	0.00	0.03	8.81E-01	+-
GSTM1	0.01	0.16	9.36E-01	0.01	0.08	9.05E-01	0.01	0.07	8.86E-01	++
CYP1A2	-0.01	0.04	7.88E-01	0.00	0.03	9.55E-01	0.00	0.02	9.16E-01	-+
CYP11A1	0.41	0.48	3.92E-01	-0.14	0.27	6.11E-01	-0.01	0.24	9.83E-01	+-
B3GAT1	-0.03	0.08	6.70E-01	0.02	0.06	7.28E-01	0.00	0.05	9.92E-01	-+

Table 36. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in subcutaneous adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
STAR	0.02	0.03	4.30E-01	0.04	0.02	1.23E-02	0.04	0.01	1.06E-02	++
HSD17B11	0.04	0.03	1.85E-01	-0.06	0.02	8.20E-04	-0.03	0.02	3.73E-02	+-
SULT4A1	0.01	0.05	8.79E-01	0.09	0.04	2.52E-02	0.06	0.03	5.90E-02	++
ARSG	0.10	0.29	7.37E-01	-0.45	0.22	4.71E-02	-0.24	0.18	1.71E-01	+-
CYP1B1	0.00	0.17	9.87E-01	-0.16	0.12	1.93E-01	-0.10	0.10	3.01E-01	+-
SULT1A4	-0.02	0.03	6.26E-01	-0.02	0.02	4.44E-01	-0.02	0.02	3.67E-01	--
NQO1	-0.02	0.02	4.65E-01	0.02	0.01	1.73E-01	0.01	0.01	4.29E-01	-+
B3GAT2	-0.01	0.03	6.02E-01	0.02	0.02	2.15E-01	0.01	0.02	4.93E-01	-+
SULT1A1	-0.01	0.02	6.48E-01	0.01	0.01	3.20E-01	0.00	0.01	5.67E-01	-+
GALNS	-0.04	0.04	2.82E-01	0.01	0.03	8.09E-01	-0.01	0.02	6.45E-01	-+
GNS	-0.02	0.91	9.83E-01	0.16	0.78	8.39E-01	0.08	0.59	8.89E-01	-+
SULF2	-0.06	0.11	5.76E-01	0.02	0.06	7.12E-01	0.00	0.05	9.63E-01	-+
HSD3B2	0.05	0.16	7.48E-01	-0.05	0.15	7.37E-01	0.00	0.11	9.78E-01	+-

Table 37. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in subcutaneous adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
STAR	0.00	0.04	9.60E-01	-0.06	0.02	1.74E-02	-0.04	0.02	4.43E-02	+-
B3GAT2	-0.05	0.04	1.80E-01	-0.02	0.03	5.76E-01	-0.03	0.02	2.10E-01	--
ARSG	0.03	0.45	9.48E-01	-0.53	0.34	1.22E-01	-0.32	0.27	2.34E-01	+-
NQO1	-0.04	0.04	3.36E-01	-0.02	0.02	4.24E-01	-0.02	0.02	2.37E-01	--
HSD3B2	0.10	0.24	6.62E-01	-0.44	0.22	4.82E-02	-0.19	0.16	2.52E-01	+-
CYP1B1	-0.15	0.25	5.60E-01	-0.11	0.19	5.65E-01	-0.12	0.15	4.18E-01	--
SULF2	-0.09	0.16	5.73E-01	-0.04	0.09	6.40E-01	-0.06	0.08	4.93E-01	--
SULT1A1	-0.03	0.02	1.76E-01	0.00	0.02	8.84E-01	-0.01	0.01	5.23E-01	-+
SULT4A1	0.13	0.08	1.00E-01	-0.04	0.06	5.39E-01	0.02	0.05	6.18E-01	+-
SULT1A4	0.01	0.05	8.95E-01	0.01	0.04	7.14E-01	0.01	0.03	7.12E-01	++
HSD17B11	-0.01	0.04	8.69E-01	-0.01	0.03	8.18E-01	-0.01	0.02	7.77E-01	--
GNS	-0.96	1.36	4.82E-01	1.15	1.20	3.37E-01	0.23	0.90	7.98E-01	-+
GALNS	0.01	0.06	9.18E-01	0.00	0.05	9.37E-01	0.00	0.04	9.00E-01	++

Table 38. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in visceral adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
GSTM1	-0.09	0.04	3.15E-02	-0.08	0.03	7.40E-03	-0.09	0.03	6.11E-04	--
SULT1E1	-0.04	0.04	3.18E-01	-0.08	0.03	1.89E-02	-0.06	0.03	1.35E-02	--
ARSA	-0.04	0.04	3.23E-01	-0.09	0.04	2.80E-02	-0.07	0.03	2.40E-02	--
CYP3A4	-0.03	0.03	3.32E-01	-0.03	0.02	1.48E-01	-0.03	0.02	8.14E-02	--
ESR1	-0.07	0.07	3.21E-01	-0.06	0.05	2.19E-01	-0.06	0.04	1.17E-01	--
SULT1C2	-0.06	0.04	1.53E-01	-0.02	0.03	5.67E-01	-0.04	0.03	1.79E-01	--
SULT4A1	-0.34	0.26	1.83E-01	-0.14	0.32	6.74E-01	-0.26	0.20	1.93E-01	--
ARSF	0.02	0.05	7.46E-01	0.03	0.03	2.38E-01	0.03	0.03	2.40E-01	++
CYP1A1	0.07	1.28	9.54E-01	-1.36	1.06	2.00E-01	-0.77	0.82	3.43E-01	+-
COMT	-0.08	0.04	5.06E-02	0.03	0.04	5.25E-01	-0.03	0.03	3.49E-01	-+
ARSJ	-0.08	0.09	3.85E-01	0.07	0.05	1.51E-01	0.04	0.04	3.86E-01	-+
CYP1B1	0.15	0.06	8.46E-03	-0.10	0.06	8.80E-02	0.03	0.04	5.30E-01	+-
UGT2B4	0.08	0.11	4.36E-01	0.01	0.08	8.93E-01	0.04	0.06	5.68E-01	++
HSD17B1	-0.04	0.04	2.88E-01	0.01	0.03	8.17E-01	-0.01	0.03	6.06E-01	-+
SULT1B1	0.02	0.20	9.26E-01	-0.05	0.10	5.79E-01	-0.04	0.09	6.46E-01	+-
SULT1A2	0.13	0.07	8.59E-02	-0.04	0.05	4.72E-01	0.02	0.04	6.55E-01	+-
SULF2	0.01	0.06	9.05E-01	0.02	0.05	7.08E-01	0.01	0.04	7.13E-01	++
ARSD	0.35	0.19	7.17E-02	-0.01	0.06	8.34E-01	0.02	0.06	7.50E-01	+-
IDS	-0.11	0.08	1.71E-01	0.04	0.04	3.00E-01	0.01	0.04	7.82E-01	-+
HSD17B14	0.03	0.07	6.47E-01	-0.03	0.05	5.00E-01	-0.01	0.04	8.02E-01	+-
SULF1	0.52	0.39	1.82E-01	-0.27	0.25	2.65E-01	-0.05	0.21	8.14E-01	+-
ESR2	-0.25	0.32	4.36E-01	0.08	0.18	6.36E-01	0.01	0.15	9.70E-01	-+
SULT2B1	-0.08	0.08	3.33E-01	0.03	0.05	5.11E-01	0.00	0.04	9.85E-01	-+

Table 39. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in visceral adipose tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
ARSA	-0.05	0.06	4.49E-01	-0.22	0.06	4.21E-04	-0.14	0.05	2.47E-03	--
CYP3A4	-0.01	0.05	8.95E-01	-0.07	0.03	3.20E-02	-0.05	0.03	6.43E-02	--
CYP1B1	0.15	0.09	8.01E-02	0.05	0.09	5.38E-01	0.10	0.06	9.43E-02	++
UGT2B4	0.05	0.16	7.42E-01	0.22	0.12	7.21E-02	0.16	0.10	1.00E-01	++
SULT4A1	-0.32	0.38	4.11E-01	-0.53	0.50	2.85E-01	-0.40	0.30	1.92E-01	--
SULT1E1	-0.07	0.06	2.98E-01	0.12	0.05	1.52E-02	0.05	0.04	2.04E-01	-+
SULT1C2	-0.06	0.06	3.11E-01	-0.03	0.05	5.31E-01	-0.05	0.04	2.60E-01	--
ESR1	0.07	0.11	5.43E-01	-0.11	0.07	1.38E-01	-0.06	0.06	3.56E-01	+-
SULF1	-0.56	0.59	3.49E-01	-0.17	0.38	6.59E-01	-0.28	0.32	3.80E-01	--
HSD17B1	-0.10	0.06	9.30E-02	0.02	0.05	7.49E-01	-0.03	0.04	3.93E-01	-+
ARSD	0.29	0.29	3.28E-01	0.05	0.09	5.67E-01	0.07	0.09	4.04E-01	++
SULT2B1	0.21	0.12	6.99E-02	-0.02	0.08	7.97E-01	0.05	0.06	4.23E-01	+-
ARSJ	0.13	0.13	3.17E-01	0.02	0.07	8.01E-01	0.04	0.06	4.88E-01	++
ARSF	0.06	0.07	3.89E-01	0.01	0.05	8.55E-01	0.02	0.04	5.41E-01	++
CYP1A1	1.58	1.92	4.11E-01	-0.02	1.63	9.92E-01	0.65	1.24	5.99E-01	+-
IDS	-0.15	0.12	2.14E-01	0.07	0.06	2.41E-01	0.03	0.06	6.48E-01	-+
SULT1B1	-0.14	0.30	6.38E-01	-0.02	0.15	8.66E-01	-0.05	0.13	7.20E-01	--
SULT1A2	0.06	0.11	5.88E-01	-0.07	0.08	4.02E-01	-0.02	0.07	7.29E-01	+-
GSTM1	-0.01	0.07	8.33E-01	-0.01	0.05	7.88E-01	-0.01	0.04	7.33E-01	--
HSD17B14	0.09	0.10	3.53E-01	-0.03	0.08	6.68E-01	0.01	0.06	8.08E-01	+-
SULF2	-0.04	0.09	6.65E-01	0.02	0.07	8.16E-01	0.00	0.05	9.30E-01	-+
ESR2	-0.49	0.49	3.16E-01	0.13	0.27	6.35E-01	-0.01	0.23	9.52E-01	-+
COMT	0.00	0.06	9.37E-01	-0.01	0.06	8.99E-01	0.00	0.04	9.74E-01	+-

Table 40. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in ovary tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
CYP1B1	-0.05	0.03	1.29E-01	-0.08	0.02	5.97E-04	-0.07	0.02	2.66E-04	--
GSTM1	-0.08	0.07	2.49E-01	-0.16	0.05	2.54E-03	-0.13	0.04	1.84E-03	--
CYP19A1	0.20	0.09	1.83E-02	0.12	0.06	6.27E-02	0.15	0.05	3.89E-03	++
ARSB	0.27	0.16	8.14E-02	0.13	0.15	3.74E-01	0.20	0.11	6.48E-02	++
ARSJ	0.07	0.04	1.19E-01	0.03	0.03	2.86E-01	0.04	0.02	7.71E-02	++
CYP3A4	0.15	0.21	4.60E-01	0.19	0.16	2.49E-01	0.17	0.13	1.73E-01	++
CYP1A1	0.12	0.07	7.95E-02	0.01	0.05	7.64E-01	0.05	0.04	2.28E-01	++
B3GAT2	0.03	0.04	4.44E-01	0.03	0.03	4.00E-01	0.03	0.02	2.56E-01	++
HSD17B14	-0.02	0.06	7.88E-01	-0.05	0.05	2.50E-01	-0.04	0.04	2.75E-01	--
SULT1C2	-0.02	0.07	7.85E-01	-0.08	0.07	2.32E-01	-0.05	0.05	2.97E-01	--
ARSG	0.00	0.07	9.91E-01	-0.05	0.04	2.27E-01	-0.04	0.04	3.07E-01	+-
ARSE	0.08	0.11	4.61E-01	0.05	0.08	4.88E-01	0.06	0.06	3.21E-01	++
UGT2B7	0.01	0.07	8.98E-01	0.05	0.06	3.94E-01	0.03	0.04	4.56E-01	++
ARSA	-0.01	0.02	6.31E-01	0.00	0.01	7.80E-01	0.00	0.01	6.21E-01	--
SULT1A2	0.03	0.06	6.28E-01	-0.03	0.04	4.67E-01	-0.01	0.03	7.12E-01	+-
ARSD	-0.08	0.15	6.12E-01	0.03	0.12	8.08E-01	-0.01	0.09	9.00E-01	-+
CYP11A1	0.02	0.03	5.57E-01	-0.01	0.03	5.64E-01	0.00	0.02	9.47E-01	+-
IDS	0.02	0.07	7.14E-01	-0.03	0.07	6.83E-01	0.00	0.05	9.79E-01	+-
HSD17B1	-0.06	0.06	3.21E-01	0.04	0.05	4.55E-01	0.00	0.04	9.82E-01	-+
SULT1A1	0.04	0.06	4.85E-01	-0.02	0.04	6.58E-01	0.00	0.03	9.91E-01	+-

Table 41. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in ovary tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
UGT2B7	-0.17	0.10	1.15E-01	-0.12	0.09	1.53E-01	-0.14	0.07	3.54E-02	--
CYP19A1	0.20	0.13	1.26E-01	0.06	0.10	5.38E-01	0.11	0.08	1.61E-01	++
CYP11A1	-0.01	0.05	7.77E-01	-0.04	0.04	2.68E-01	-0.03	0.03	3.04E-01	--
SULT1C2	-0.11	0.10	3.00E-01	-0.04	0.10	6.90E-01	-0.07	0.07	3.11E-01	--
B3GAT2	0.00	0.06	9.48E-01	0.06	0.05	2.29E-01	0.03	0.04	3.37E-01	++
IDS	-0.13	0.10	2.23E-01	0.00	0.10	9.72E-01	-0.06	0.07	3.98E-01	-+
HSD17B14	0.01	0.10	9.17E-01	0.07	0.07	3.40E-01	0.05	0.06	4.04E-01	++
ARSG	0.06	0.11	5.96E-01	-0.09	0.07	2.05E-01	-0.05	0.06	4.25E-01	+-
GSTM1	-0.06	0.11	5.63E-01	-0.04	0.08	6.58E-01	-0.05	0.06	4.85E-01	--
SULT1A1	0.10	0.09	2.95E-01	0.01	0.06	8.90E-01	0.04	0.05	4.86E-01	++
SULT1A2	0.12	0.09	1.86E-01	0.00	0.06	9.85E-01	0.03	0.05	4.94E-01	+-
CYP3A4	-0.08	0.31	7.96E-01	-0.13	0.25	5.96E-01	-0.11	0.19	5.64E-01	--
CYP1B1	-0.02	0.05	7.29E-01	-0.01	0.04	7.73E-01	-0.01	0.03	6.60E-01	--
ARSE	0.17	0.17	3.20E-01	-0.03	0.12	7.92E-01	0.04	0.10	7.19E-01	+-
ARSD	0.18	0.23	4.39E-01	-0.03	0.18	8.58E-01	0.05	0.14	7.33E-01	+-
ARSB	-0.06	0.24	7.97E-01	-0.05	0.23	8.28E-01	-0.06	0.17	7.38E-01	--
CYP1A1	-0.11	0.10	3.11E-01	0.02	0.07	7.90E-01	-0.02	0.06	7.39E-01	-+
ARSJ	-0.06	0.07	3.26E-01	0.02	0.05	6.58E-01	-0.01	0.04	8.37E-01	-+
ARSA	-0.06	0.03	3.43E-02	0.03	0.02	1.11E-01	0.00	0.01	8.72E-01	-+
HSD17B1	0.06	0.09	4.92E-01	-0.04	0.07	5.95E-01	0.00	0.06	9.85E-01	+-

Table 42. Association between the predicted expression level of estrogen-related genes and breast cancer risk among post-menopausal women using prediction model built in cross tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
SULT1A4	-0.29	0.19	1.27E-01	-0.28	0.14	4.68E-02	-0.28	0.11	1.23E-02	--
SULT1C2	0.05	0.03	1.14E-01	0.04	0.02	7.50E-02	0.04	0.02	1.83E-02	++
SULF1	-0.16	2.78	9.55E-01	0.47	0.28	9.56E-02	0.46	0.28	9.85E-02	-+
ARSI	0.01	0.51	9.82E-01	0.66	0.34	5.43E-02	0.46	0.28	1.09E-01	++
SULT1A1	0.77	0.48	1.11E-01	0.20	0.75	7.94E-01	0.60	0.41	1.39E-01	++
UGT2B7	0.07	0.27	7.89E-01	0.15	0.11	1.89E-01	0.14	0.11	1.89E-01	++
ARSD	-0.60	0.58	3.08E-01	-0.24	0.28	3.88E-01	-0.31	0.25	2.22E-01	--
UGT2B17	0.15	0.09	1.10E-01	-0.01	0.09	8.85E-01	0.07	0.07	3.03E-01	+-
GALNS	0.00	0.07	9.93E-01	-0.06	0.05	2.64E-01	-0.04	0.04	3.64E-01	--
B3GAT1	-0.08	0.09	3.70E-01	-0.03	0.08	7.41E-01	-0.05	0.06	3.87E-01	--
SULT1E1	-0.05	0.06	3.59E-01	0.00	0.10	9.92E-01	-0.04	0.05	4.27E-01	--
CYP11A1	-0.01	0.10	9.40E-01	0.05	0.06	3.66E-01	0.04	0.05	4.65E-01	-+
SULT1A2	-0.06	0.06	2.48E-01	0.00	0.04	1.00E+00	-0.02	0.03	4.96E-01	--
SULT4A1	0.58	1.28	6.50E-01	0.48	0.99	6.30E-01	0.52	0.78	5.10E-01	++
STAR	-0.10	0.19	6.04E-01	-0.13	0.42	7.53E-01	-0.10	0.17	5.47E-01	--
SULT2B1	-0.03	0.03	2.87E-01	0.03	0.02	1.98E-01	0.01	0.02	6.29E-01	-+
CYP17A1	0.20	0.12	1.12E-01	-0.04	0.08	6.56E-01	0.03	0.07	6.32E-01	+-
CYP1B1	0.02	0.23	9.28E-01	-0.06	0.14	6.80E-01	-0.04	0.12	7.59E-01	+-
STS	-0.14	0.27	6.10E-01	0.01	0.21	9.52E-01	-0.04	0.17	7.91E-01	-+
HSD17B14	0.12	0.14	3.62E-01	-0.11	0.10	3.08E-01	-0.02	0.08	7.97E-01	+-
CYP3A4	-0.10	0.12	4.24E-01	0.06	0.08	4.59E-01	0.01	0.06	8.43E-01	-+
AKR1B15	-0.08	0.13	5.51E-01	0.04	0.07	6.09E-01	0.01	0.06	8.89E-01	-+
SULF2	0.15	0.31	6.16E-01	-0.10	0.21	6.53E-01	-0.01	0.17	9.34E-01	+-
GSTM1	-0.02	0.04	5.61E-01	0.01	0.02	8.03E-01	0.00	0.02	9.41E-01	-+
ARSA	0.00	0.04	9.08E-01	-0.01	0.08	8.85E-01	0.00	0.03	9.66E-01	+-

Table 43. Association between the predicted expression level of estrogen-related genes and breast cancer risk among pre-menopausal women using prediction model built in cross tissue.

Gene Name	Icogs			Onco_array			Meta-analysis			Direction
	Effect	SE	P	icogs/onco-array	SE	P	Effect	SE	P	icogs/onco-array
B3GAT1	0.22	0.13	9.42E-02	0.07	0.13	6.12E-01	0.14	0.09	1.23E-01	++
SULT1A4	-0.37	0.29	2.06E-01	-0.15	0.22	4.87E-01	-0.23	0.17	1.89E-01	--
ARSI	-1.94	0.77	1.23E-02	0.09	0.52	8.70E-01	-0.55	0.43	2.06E-01	-+
SULT4A1	0.13	1.91	9.45E-01	-2.50	1.52	1.01E-01	-1.47	1.19	2.15E-01	+-
HSD17B14	-0.42	0.20	3.88E-02	0.06	0.16	7.22E-01	-0.12	0.13	3.20E-01	-+
UGT2B17	-0.03	0.14	8.04E-01	-0.15	0.14	3.02E-01	-0.09	0.10	3.70E-01	--
SULT2B1	-0.01	0.05	8.72E-01	-0.03	0.03	3.52E-01	-0.02	0.03	3.89E-01	--
STS	-0.04	0.41	9.27E-01	-0.31	0.32	3.41E-01	-0.21	0.26	4.20E-01	--
AKR1B15	-0.10	0.19	6.00E-01	-0.05	0.11	6.35E-01	-0.07	0.10	4.99E-01	--
CYP3A4	-0.14	0.18	4.56E-01	-0.03	0.12	8.21E-01	-0.06	0.10	5.54E-01	--
SULF2	-0.03	0.47	9.42E-01	0.24	0.32	4.62E-01	0.15	0.27	5.73E-01	-+
SULT1C2	0.00	0.05	9.38E-01	0.02	0.03	5.33E-01	0.02	0.03	5.77E-01	++
CYP17A1	0.10	0.19	6.06E-01	0.03	0.12	8.01E-01	0.05	0.10	6.22E-01	++
SULT1A2	-0.08	0.08	3.20E-01	0.01	0.06	8.24E-01	-0.02	0.05	6.77E-01	-+
CYP1B1	0.29	0.34	4.06E-01	-0.20	0.21	3.27E-01	-0.07	0.18	6.86E-01	+-
GALNS	-0.06	0.10	5.38E-01	0.07	0.08	3.61E-01	0.02	0.06	7.10E-01	-+
UGT2B7	0.55	0.42	1.83E-01	-0.04	0.18	8.37E-01	0.05	0.16	7.41E-01	+-
ARSD	-0.56	0.88	5.28E-01	0.29	0.44	5.09E-01	0.12	0.39	7.55E-01	-+
SULT1A1	0.70	0.73	3.38E-01	-1.13	1.13	3.20E-01	0.16	0.61	7.91E-01	+-
STAR	0.04	0.28	8.90E-01	0.08	0.65	9.01E-01	0.05	0.26	8.60E-01	++
SULF1	-3.94	4.26	3.55E-01	0.11	0.43	7.97E-01	0.07	0.43	8.71E-01	-+
SULT1E1	0.02	0.09	7.94E-01	-0.10	0.15	5.10E-01	-0.01	0.07	9.13E-01	+-
GSTM1	-0.04	0.06	5.40E-01	0.01	0.03	6.72E-01	0.00	0.03	9.48E-01	-+
ARSA	-0.03	0.05	5.70E-01	0.14	0.12	2.36E-01	0.00	0.05	9.81E-01	-+
CYP11A1	-0.17	0.15	2.46E-01	0.06	0.09	4.82E-01	0.00	0.08	9.87E-01	-+

Table 44. Summary results for associations between predicted gene expression level and breast cancer among post-menopausal women.

Breast			Liver			Subcutaneous adipose			Visceral adipose		
Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value
ARSA	-0.05	1.02E-03	*UGT2B28*	0.16	5.71E-03	*STAR*	0.04	1.06E-02	*GSTM1*	-0.09	6.11E-04
CYP3A5	0.28	1.51E-03	*UGT2B4*	-0.05	8.64E-03	*HSD17B11*	-0.03	3.73E-02	*SULT1E1*	-0.06	1.35E-02
SGSH	0.07	3.67E-02							*ARSA*	-0.07	2.40E-02

Ovary			Cross		
Gene	Effect	P-value	Gene	Effect	P-value
CYP1B1	-0.07	2.66E-04	*SULT1A4*	-0.28	1.23E-02
GSTM1	-0.13	1.84E-03	*SULT1C2*	0.04	1.83E-02
CYP19A1	0.15	3.89E-03			

Gene	Breast		Liver		Subcutaneous adipose		Visceral adipose		Ovary		Cross	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.05	1.02E-03					-0.07	2.40E-02				
CYP19A1									0.15	3.89E-03		
CYP1B1									-0.07	2.66E-04		
CYP3A5	0.28	1.51E-03										
GSTM1							-0.09	6.11E-04	-0.13	1.84E-03		
HSD17B11					-0.03	3.73E-02						
SGSH	0.07	3.67E-02										
STAR					0.04	1.06E-02						
SULT1A4											-0.28	1.23E-02
SULT1C2											0.04	1.83E-02
SULT1E1							-0.06	1.35E-02				
UGT2B28			0.16	5.71E-03								
UGT2B4			-0.05	8.64E-03								

Table 45. Summary results for associations between predicted gene expression level and breast cancer among pre-menopausal women.

Breast			Liver			Subcutaneous adipose			Visceral adipose		
Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value	Gene	Effect	P-value
UGT2B11	-0.23	1.64E-03	*B3GAT3*	-0.12	5.43E-03	*STAR*	-0.04	4.43E-02	*ARSA*	-0.14	2.47E-03
HSD3B2	-0.28	7.80E-03	*SGSH*	0.59	2.65E-02						
			UGT1A4	0.04	4.33E-02						
			HSD3B2	-0.05	4.35E-02						

Ovary			Cross		
Gene	Effect	P-value	Gene	Effect	P-value
UGT2B7	-0.14	3.54E-02			

Gene	Breast		Liver		Subcutaneous adipose		Visceral adipose		Ovary		Cross	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA							-0.14	2.47E-03				
B3GAT3			-0.12	5.43E-03								
HSD3B2	-0.28	7.80E-03	-0.05	4.35E-02								
SGSH			0.59	2.65E-02								
STAR					-0.04	4.43E-02						
UGT1A4			0.04	4.33E-02								
UGT2B11	-0.23	1.64E-03										
UGT2B7									UGT2B7	-0.14		

2.4. The associations between predicted gene expression level and breast cancer after adjustment for potential breast cancer risk factors.

Table 46. Association after adjustment for covariates in breast tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.04	2.33E-05	-0.05	6.23E-05	-0.05	6.05E-05	-0.06	2.70E-04
SULT4A1	0.04	8.82E-03	0.04	2.45E-02	0.04	2.68E-02	0.02	4.12E-00
CYP3A5	0.15	1.39E-02	0.16	2.25E-02	0.16	2.72E-02	0.22	6.65E-03
B3GAT2	-0.05	4.34E-02	-0.07	6.14E-03	-0.07	6.55E-03	-0.06	6.34E-02

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.04	2.33E-05	-0.06	1.07E-05	-0.06	1.89E-05	-0.05	1.31E-04	-0.06	1.05E-03	-0.06	1.71E-03
SULT4A1	0.04	8.82E-03	0.03	1.10E-01	0.03	1.05E-01	0.03	5.96E-02	0.03	2.62E-01	0.02	3.68E-01
CYP3A5	0.15	1.39E-02	0.14	6.86E-02	0.22	7.76E-03	0.21	5.43E-03	0.29	5.08E-03	0.29	4.76E-03
B3GAT2	-0.05	4.34E-02	0.02	3.41E-01	-0.09	5.65E-03	-0.07	1.66E-02	-0.05	2.11E-01	-0.07	1.24E-01

Table 47. Association after adjustment for covariates in liver tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
UGT2B4	-0.03	1.00E-02	-0.04	4.19E-03	-0.04	4.37E-03	-0.03	4.90E-02
SULT2B1	-0.02	1.58E-02	-0.01	2.93E-01	-0.01	3.03E-01	-0.01	4.72E-01
UGT1A6	0.04	2.06E-02	0.04	9.21E-02	0.04	1.11E-01	0.04	1.92E-01
SULT4A1	-0.03	3.07E-02	-0.03	5.26E-02	-0.03	5.19E-02	-0.03	1.02E-01

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
UGT2B4	-0.05	1.08E-03	-0.03	4.51E-02	-0.05	1.89E-03	-0.03	3.25E-02	-0.04	4.31E-02	-0.04	7.77E-02
SULT2B1	-0.02	1.16E-01	-0.01	5.99E-01	-0.002	8.55E-01	-0.01	3.78E-01	-0.04	6.06E-01	-0.01	6.19E-01
UGT1A6	0.03	1.37E-01	0.05	8.41E-02	0.03	3.09E-01	0.03	1.47E-01	0.03	3.50E-01	0.03	3.80E-01
SULT4A1	-0.03	5.24E-02	-0.03	1.68E-01	-0.03	1.33E-01	-0.03	8.11E-02	-0.03	1.72E-01	-0.03	1.86E-01

Table 48. Association after adjustment for covariates in subcutaneous adipose tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
HSD17B11	-0.03	6.33E-03	-0.03	1.75E-02	-0.03	2.00E-02	-0.02	9.71E-02

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
HSD17B11	-0.02	7.05E-02	-0.03	3.87E-02	-0.03	6.50E-02	-0.03	2.67E-02	-0.03	6.19E-02	-0.04	4.92E-02

Table 49. Association after adjustment for covariates in visceral adipose tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.09	5.95E-06	-0.11	5.21E-06	-0.11	4.69E-06	-0.12	6.60E-05
SULT4A1	-0.34	1.19E-02	-0.27	9.52E-02	-0.29	7.31E-02	-0.25	2.16E-01
GSTM1	-0.04	2.16E-02	-0.08	1.21E-04	-0.08	9.87E-05	-0.06	7.38E-03
ARSJ	0.06	4.18E-02	0.06	5.51E-02	0.06	6.73E-02	0.04	2.55E-01
SULT1C2	-0.04	4.46E-02	-0.04	4.60E-02	-0.04	5.30E-02	-0.04	9.29E-02

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
ARSA	-0.11	1.74E-05	-0.12	2.11E-05	-0.02	5.00E-01	-0.12	3.37E-06	-0.13	3.89E-02	-0.10	5.83E-03
SULT4A1	-0.26	1.24E-01	-0.21	2.84E-01	-0.38	4.45E-02	-0.27	1.20E-01	-0.53	1.65E-01	-0.35	1.83E-01
GSTM1	-0.07	1.78E-03	-0.07	1.58E-03	-0.08	3.41E-04	-0.06	9.22E-03	-0.15	2.37E-02	-0.08	1.33E-02
ARSJ	0.05	1.14E-01	0.04	2.74E-01	0.09	1.66E-02	0.06	7.95E-02	0.02	8.83E-01	0.03	5.06E-01
SULT1C2	-0.03	1.18E-01	-0.04	1.25E-01	-0.03	2.82E-01	-0.05	2.25E-02	-0.06	3.16E-01	-0.02	6.50E-01

Table 50. Association after adjustment for covariates in ovary tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
GSTM1	-0.09	1.77E-03	-0.14	5.99E-05	-0.14	3.70E-05	-0.14	5.55E-04
CYP1B1	-0.04	5.70E-03	-0.04	5.04E-03	-0.04	3.94E-03	-0.05	2.08E-03
CYP19A1	0.09	9.93E-03	0.13	1.50E-03	0.13	1.81E-03	0.14	3.17E-03
UGT2B7	-0.06	3.12E-02	-0.04	2.62E-01	-0.04	2.55E-01	-0.01	8.30E-01

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
GSTM1	-0.12	4.13E-04	-0.15	7.93E-05	-0.15	1.65E-04	-0.11	1.78E-03	-0.18	7.80E-02	-0.14	7.55E-03
CYP1B1	-0.04	8.76E-03	-0.05	2.57E-03	-0.04	1.01E-02	-0.05	1.80E-03	-0.08	7.23E-02	-0.08	2.82E-04
CYP19A1	0.12	4.00E-03	0.12	1.08E-02	0.12	9.12E-03	0.13	2.09E-03	0.18	1.56E-01	0.09	1.23E-01
UGT2B7	-0.04	3.20E-01	-0.04	3.80E-01	-0.03	4.25E-01	-0.07	8.08E-02	-0.07	4.87E-01	0.04	4.67E-01

Table 51. Association after adjustment for covariates in cross tissue model.

	Adjustment							
Gene	None		parous		parity		HRT	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
SULT1A1	0.67	1.53E-02	0.44	1.91E-01	0.41	2.19E-01	0.32	4.53E-01
SULT1C2	0.03	2.14E-02	0.04	1.53E-02	0.03	2.22E-02	0.05	4.97E-03
SULT1A4	-0.18	2.43E-02	-0.27	2.83E-03	-0.28	2.54E-03	-0.27	1.44E-02

	Adjustment											
Gene	BMI		OC use		age at first full time pregnancy		age at menarche		age at menopause		reproductive years	
	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
SULT1A1	0.30	3.85E-01	0.30	4.70E-01	0.32	4.17E-01	0.63	8.01E-02	-0.09	9.03E-01	0.08	9.12E-01
SULT1C2	0.03	4.21E-02	0.04	9.75E-03	0.04	9.47E-03	0.03	7.57E-02	0.04	3.70E-01	0.03	4.99E-01
SULT1A4	-0.28	3.08E-03	-0.21	4.34E-02	-0.30	4.39E-03	-0.23	1.71E-02	0.23	4.19E-01	0.18	5.33E-01

2.5. Results for Asian population

Table 52. Association between the predicted expression level of estrogen-related genes and breast cancer risk using prediction model built in breast tissue of samples from either GTEx or Shanghai (SH), China.

Gene name	Prediction model	Effect	SE	P
CYP19A1	GTEx	0.78	0.42	6.28E-02
SULT4A1	SH	-0.02	0.01	8.11E-02
HSD17B1	SH	-0.07	0.05	1.54E-01
ARSK	SH	0.08	0.06	1.92E-01
SULT1A2	SH	-0.04	0.04	2.42E-01
CYP3A4	SH	-0.76	0.67	2.62E-01
UGT2B28	SH	0.05	0.04	2.98E-01
SULT1C2	SH	-0.03	0.02	3.04E-01
GALNS	GTEx	0.04	0.05	3.72E-01
ARSA	GTEx	0.10	0.12	3.89E-01
AKR1B15	SH	-0.61	0.79	4.40E-01
GSTA1	SH	-0.03	0.05	5.57E-01
UGT2B15	SH	-0.01	0.04	7.40E-01
GSTM1	SH	-0.01	0.02	7.46E-01
SULF1	SH	-0.02	0.08	7.55E-01
B3GAT2	GTEx	0.01	0.03	8.61E-01
ESR2	SH	0.00	0.07	9.72E-01
CYP17A1	SH	0.00	0.13	9.72E-01
SULT1A4	SH	0.00	0.19	9.80E-01

Some of the tested genes showing an association with breast cancer risk at $P<0.05$ have been previously reported to potentially play important roles in the development of breast cancer. For example, for *CYP19A1*, research suggests that endocrine-disrupting compounds could up-regulate aromatase *CYP19A1* mRNA levels, the aromatase-induced biosynthesis of the breast carcinogen 17β-estradiol, as well as increased ERα-positive breast cell proliferation (107). This is consistent with the detected positive association between predicted expression and breast cancer risk in our analyses (Table 17). Furthermore, such an association tends to be restricted within

post-menopause women. Very interesting, previous research suggested that *CYP19A1* mRNA was significantly elevated in postmenopausal patients. Regarding *CYP1B1*, research suggested that its expression was downregulated in breast cancerous tissue compared with normal adjacent tissue (108). This is consistent with our finding of an inverse association between the genetically predicted expression and breast cancer risk. Furthermore, in our study the association for *CYP1B1* tends to be restricted to post-menopausal women. Interestingly, previous research conducted by our group suggested that *CYP1B1* genetic polymorphisms may be associated with the natural onset of menopause (109). For *GSTM1*, research suggested that there tended to have higher copy number deletion in breast cancer cases versus controls and most of the measurements of the expression level of this gene were zero (110). For *UGT2B4*, research showed that its transfection resulted in significantly fewer tumor cells compared with untreated condition (111). Regarding *UGT2B7,* previous findings suggested decreased expression in breast tissue of breast cancer cases versus healthy controls (112). These findings tend to be consistent with the observed inverse association between predicted expression of *UGT2B4* and *UGT2B7* with breast cancer risk.

CHAPTER 4

SPECIFIC AIM 3

The objective of this aim is to conduct a gene-environment interaction study assessing potential interaction of predicted expression levels of estrogen related genes with BMI, menstrual and reproductive factors (age at menarche, age at menopause, age at the end of first full time pregnancy, parity, parous, OC use, HRT use) and soy food intake (only Asians) on breast cancer risk.

1. Methods

We first evaluated association of estrogen related non-genetic covariates including BMI, HRT use, OC use, parity, parous, age at FFTP, age at menarche and age at menopause with breast cancer risk. All statistically analyses were conducted among European and Asian population separately. The distributions of those variables were compared between breast cancer cases and controls by treating them as both a continuous variable and a categorical variable if applicable. BMI was classified into groups of underweight (<18.5 kg/m^2), normal (18.5-24.9 kg/m^2), overweight (25-29.9) kg/m^2), and obese (≥3 0 kg/m^2) according to the BMI cut-off points recommended by the World Health Organization(113). Age at menarche was categorized as (< 12, 12, 13, 14, 15, 16, ≥ 17); age at menopause as (<40, 40-44, 45-48, 49-54, ≥ 55); parity as (0, 1, 2, 3, 4, ≥ 4); age at FFTP as (< 20, 20-24, 25-29, 30-34, ≥ 35); soy food intake was grouped by quartile distribution.

The associations between those non-genetic factors and breast cancer risk were evaluated using logistic regressions to derive the mutually adjusted odds ratios (ORs) and 95% confidence intervals. The related associations were also examined by menopausal status in both European and Asian populations, and by ER or PR status of the breast cancer cases among Caucasians. Presence of an interaction was assessed with the Wald test. The interaction terms were created by multiplying all predicted gene expression levels as a continuous variable and non-genetics factors treated as either a continuous variable (BMI, parity, age at FFTP, age at menarche, age at menopause and soy food intake) or as a binary variable (HRT use, OC use, and parous). ORs were adjusted for study site and top principal components.

Given the objective of exploratory nature of this study, we did not adjust for multiple comparison in our main analysis. Instead, we presented the interactions by P-interaction less than 0.05, 0.01 and 0.005, respectively. For those with a P value for the interaction term less than 0.005, we carried out further stratified analyses to display the association between the non-genetic factors and breast cancer risk by quartiles distribution of the predicted gene expression. In addition, subgroup analyses were conducted to evaluate the association of BMI and HRT use with all estrogen related genes only among post-menopausal women. Given the substantial alteration of the function of ovaries after menopause, interactions were assessed restricted among pre-menopausal women for gene expression levels predicted by models built in ovaries.

To reduce the possibility of false positive findings due to multiple comparisons, we created a new index to combine the genes in estrogen synthesis, metabolism and bio-availability, respectively, based on the proposed role of these genes. The predicted

gene expression level for each gene within each tissue was first categorized by quartiles from the controls and assigned a value of one of the following: 0; 1; 2; 3; with "0" representing the group with lowest and "3" with the highest predicted expression level. For gene *STS*, whose role is to deconjugate the conjugated estrogen and its metabolites, we multiplied the assigned value by -1 to convert the number to be negative considering of their function in the estrogen metabolism pathway. Next, a summed value was calculated within each pathway by adding the assigned value representing the expression level together. Thus, for each participant, there were three summed set of scores developed indicating her overall expression level in estrogen synthesis, metabolism and bioavailability pathway based on different tissue models built. The interaction terms were created by multiplying the summed score for predicted gene expression levels by the non-genetics factors as individual gene analysis, and Wald test was used to test the significance of the interactions with $P<0.05$ as the cut-off. All statistical analyses were conducted using R version 3.4.1.

2. Power calculation

The power calculation was conducted using Power V3.0 software. Assuming that for each quintile increase in the non-genetic factors, the OR is 1.1, the ORs for the group of high genetically predicted expression is 0.9 and 1.5, respectively, for those in the lowest and highest quintile of non-genetic variable, to achieve 80% power would require a sample of 4,670 cases and 4,670 controls. The proposed study thus should have sufficient power, considering that the analysis will be based on 97,694 cases and 87,560 controls for European descendants and 11,074 cases and 7,776 for Asians.

3. Results for European descendants

3.1. Non-genetic variables.

Table 53 showed the characteristics of our study population of European descendants. In general, compared to controls, breast cancer cases were more likely to be older, nulliparous, have higher BMI, and menopause at an older age. They were also less likely to use oral contraceptive, have given birth to more than one child or give birth to their first child at an earlier age.

Table 53. Distribution of the known breast cancer risk factors by breast cancer status among European descendants, *BCAC*.

	Total No. of cases	Total No. of controls	Case	Control	P-value
Age at interview	97694	87560	57.3 (49.0-65.3)	57 (48.0-64.0)	<.0001
BMI (kg/m2)	74020	66593	25.3 (22.7-28.9)	25.1 (22.6-28.4)	<.0001
BMI (categorized)	74020	66593			<.0001
18.5-24.9			45.7	47.4	
< 18.5			1.5	1.4	
25-29.9			32.9	33	
≥ 30			19.9	18.2	
Hormone replacement therapy use (% of yes)	55137	52183	42.9	43.5	0.049
Oral conceptive use (% of yes)	56121	57723	61.7	65.6	<.0001
Post-menopausal (% of yes)	73577	67138	68.1	70.7	<.0001
Age at Menarche (years)	71432	63247	13 (12-14)	13 (12-14)	0.147
Age at Menarche (categorized)	71432	63247			0.364
< 12			15.7	16	
12			23.4	23.6	
13			25.9	25.6	
14			19.9	20	
15			9.1	9	
16			4	3.9	
≥ 17			2	1.9	
Age at Menopause (years)	38167	34168	50 (46-52)	50 (45-52)	<.0001
Age at Menopause (categorized)	38167	34168			<.0001
< 40			7.8	8.8	
40-44			11.3	11.9	
45-48			21.1	21.2	
49-54			48.1	47.3	
>=55			11.7	10.9	
Parous (% of no)	75244	72169	15.4	13.3	<.0001
Parity	74640	71196	2 (1-3)	2 (1-3)	<.0001
Parity (categorized)	74640	71196			<.0001
0			15.5	13.4	
1			17.4	14.9	
2			38.7	39.1	
3			18.7	20.7	
4			6.6	7.6	

>4			3.2	4.3	
Age at end of first full-term pregnancy (years)	58058	53012	25 (22-28)	24 (22-28)	<.0001
Age at end of first full-term pregnancy (categorized)	58058	53012			<.0001
< 20			10.8	10.1	
20-24			39	41.4	
25-29			32.5	32.9	
30-34			13	11.7	
≥ 35			4.8	3.9	

a. Numbers were presented in median (interquartile) or percentage for continuous and categorical variables, respectively.
b. P-values were calculated from ANOVA test or Chi-square test for continuous and categorical variables, respectively.

Table 54 displayed the mutually adjusted odds ratios (ORs) between the evaluated breast cancer risk factors and breast cancer risk in our European study population. BMI was observed to be associated with an increased risk of breast cancer among post-menopausal women, and with a decreased risk of breast cancer among pre-menopausal women. HRT use was associated with an elevated risk of breast cancer (OR=1.08, 95% confidence interval (CI):1.04, 1.13). Higher age at menarche and parity were both associated with a lower likelihood of having breast cancer. Higher age at menopause and first full-term pregnancy were both associated with an increased risk of breast cancer.

Table 54. Odds ratios (95% confidence intervals) for the associations of body mass index, menstrual and reproductive factors with breast cancer risk among European descendants.

	All	Post-menopausal	Pre-menopausal	ER-positive breast cancer	ER-negative breast cancer
No. of control/case	33388/28959	23625/20196	9763/8763	33388/19406	33388/15401
BMI					
18.5-24.9	1.0	1.0	1.0	1.0	1.0
<18.5	1.03 (0.88, 1.20)	0.86 (0.71, 1.04)	1.50 (1.14, 1.98)	0.99 (0.83, 1.17)	1.34 (1.01, 1.78)
25-29.9	1.01 (0.98, 1.05)	1.04 (1.00, 1.09)	0.92 (0.85, 0.99)	1.01 (0.97, 1.06)	1.00 (0.92, 1.08)
≥ 30	1.04 (0.99, 1.09)	1.11 (1.05, 1.17)	0.80 (0.73, 0.88)	1.04 (0.99, 1.10)	0.97 (0.88, 1.07)
Hormone replacement therapy use	1.07 (1.03, 1.11)	1.06 (1.02, 1.11)	0.94 (0.84, 1.05)	1.13 (1.08, 1.19)	0.90 (0.82, 0.98)
Oral conceptive use	1.08 (1.04, 1.13)	1.08 (1.03, 1.13)	0.99 (0.90, 1.08)	1.06 (1.01, 1.11)	1.14 (1.04, 1.25)
Age at Menarche	0.97 (0.96, 0.98)	0.97 (0.96, 0.99)	0.95 (0.93, 0.97)	0.97 (0.96, 0.98)	0.97 (0.94, 0.99)

Age at Menopause					
<40		1.0		1.0	1.0
40-44		0.95 (0.87, 1.04)		0.98 (0.89, 1.09)	
45-48		1.00 (0.92, 1.09)		1.05 (0.96, 1.16)	
49-54		1.04 (0.97, 1.13)		1.09 (1.00, 1.18)	
>=55		1.15 (1.05, 1.26)		1.19 (1.07, 1.32)	
Parity	0.93 (0.91, 0.94)	0.93 (0.91, 0.95)	0.90 (0.86, 0.93)	0.93 (0.91, 0.95)	0.92 (0.89, 0.95)
Age at end of first full-term pregnancy					
<20	1.0	1.0	1.0	1.0	1.0
20-24	1.00 (0.94, 1.06)	1.01 (0.94, 1.08)	0.98 (0.86, 1.12)	1.02 (0.95, 1.09)	0.97 (0.86, 1.09)
25-29	1.13 (1.06, 1.21)	1.18 (1.10, 1.27)	1.04 (0.91, 1.18)	1.18 (1.10, 1.27)	0.97 (0.86, 1.10)
30-34	1.29 (1.19, 1.39)	1.35 (1.23, 1.48)	1.14 (0.99, 1.32)	1.40 (1.29, 1.53)	0.98 (0.85, 1.14)
>=35	1.42 (1.28, 1.58)	1.54 (1.34, 1.76)	1.16 (0.96, 1.38)	1.50 (1.33, 1.68)	1.31 (1.07, 1.60)

3.2. Interactions between estrogen-related genetically predicted gene expression level and non-genetic variables on breast cancer risk.

The results of the assessed interaction between estrogen-related pathway genes assessed by different tissue based model and estrogen-related non-genetic factors were presented from Table 55 to Table 102. The summarized results for those interactions were included from Table 103 to Table 105 based on different cutoffs of p-values. If P<0.01 is used to determine the significance of the examined interactions, significant interactions were observed for two genes (ARSA and AKR1B15) and HRT on breast cancer risk across three tissues. Two genes (HSD3B2 and ARSD) had a significant interaction with OC use on breast cancer risk. Three genes (SULT1A4, B3GAT1 and IDS) interacted with parous, and two genes (HSD17B11 and SULF1) interacted with age at FFTP. If a stricter threshold was used (P < 0.005), the AKR1B15 and HRT use interaction remained significant (subcutaneous adipose tissue model, P = 0.003, cross tissue model, P = 0.004). Additionally, B3GAT1 (P for interaction = 0.0004) and IDS (P for interaction = 0.001) had a significant interaction with parous in

the cross and ovary tissue mode; HSD17B11 (P for interaction = 0.001) had a significant interaction with age at FFTP on breast cancer risk.

Table 55. Interaction between Body mass index and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
COMT	-0.027	0.011	0.011
SULT4A1	0.008	0.003	0.015
ARSA	-0.006	0.003	0.016
UGT2B11	0.015	0.008	0.056
HSD17B1	0.010	0.005	0.058
UGT2B17	-0.006	0.003	0.066
SULT1A2	-0.007	0.004	0.106
GSTM1	-0.008	0.006	0.195
B3GAT2	0.007	0.005	0.217
ARSK	0.056	0.048	0.237
ARSF	0.018	0.016	0.258
ARSB	-0.014	0.013	0.269
SULT1B1	-0.046	0.044	0.302
GSTA1	-0.006	0.006	0.325
CYP1A1	0.008	0.008	0.368
SULT1C2	-0.002	0.002	0.371
CYP3A5	0.011	0.014	0.422
UGT2B7	0.008	0.013	0.521
SGSH	-0.003	0.006	0.633
GALNS	-0.003	0.006	0.668
SHBG	-0.005	0.012	0.677
SULT2B1	-0.002	0.005	0.724
B3GAT3	0.001	0.004	0.731
CYP1A2	0.000	0.001	0.761
IDS	-0.001	0.006	0.822
ARSE	0.001	0.005	0.881
CYP3A4	0.001	0.008	0.902
HSD3B2	0.000	0.011	0.971

Table 56. Interaction between Body mass index and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
CYP1A2	0.004	0.003	0.085
SULT2B1	0.003	0.002	0.151

ARSB	-0.002	0.002	0.163
SULT4A1	0.004	0.003	0.255
AKR1B15	-0.074	0.072	0.303
UGT1A6	-0.004	0.004	0.321
SGSH	-0.025	0.028	0.375
UGT2B17	-0.003	0.003	0.391
STAR	-0.005	0.006	0.392
CYP1B1	-0.004	0.005	0.405
SULF2	-0.004	0.005	0.471
ESR2	0.003	0.005	0.489
B3GAT1	-0.003	0.005	0.536
UGT2B4	-0.002	0.003	0.569
UGT2B28	-0.005	0.009	0.585
GSTA1	-0.002	0.004	0.644
HSD17B11	-0.010	0.025	0.690
ARSA	-0.002	0.005	0.696
HSD17B14	0.002	0.007	0.751
COMT	0.001	0.004	0.761
ARSJ	0.003	0.010	0.768
UGT1A4	-0.001	0.002	0.813
ARSE	-0.001	0.005	0.816
CYP11A1	-0.006	0.024	0.819
GSTM1	-0.002	0.008	0.846
HSD3B2	0.000	0.003	0.888
B3GAT2	0.000	0.003	0.924
SULT1A1	0.003	0.047	0.948
B3GAT3	0.000	0.004	0.975

Table 57. Interaction between Body mass index and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
HSD17B14	0.009	0.006	0.115
IDS	-0.011	0.008	0.158
SULT1A1	0.007	0.005	0.181
UGT2B7	0.008	0.007	0.238
GSTM1	-0.008	0.007	0.242
ARSA	0.001	0.002	0.361
CYP3A4	-0.017	0.020	0.404
SULT1C2	-0.007	0.008	0.409
SULT1A2	-0.004	0.005	0.498
ARSD	-0.007	0.015	0.643
CYP11A1	0.001	0.003	0.654
HSD17B1	-0.003	0.006	0.664
CYP1A1	0.002	0.006	0.684

B3GAT2	-0.001	0.004	0.717
CYP1B1	0.001	0.003	0.797
ARSG	0.001	0.006	0.866
ARSE	-0.001	0.010	0.902
CYP19A1	0.001	0.008	0.930
ARSJ	0.000	0.004	0.967
ARSB	-0.001	0.018	0.968

Table 58. Interaction between Body mass index and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
NQO1	0.004	0.002	0.042
SULT1A4	-0.005	0.003	0.113
CYP1B1	-0.022	0.016	0.162
B3GAT2	-0.003	0.003	0.229
STAR	0.003	0.002	0.245
HSD17B11	-0.003	0.002	0.274
GALNS	0.003	0.004	0.481
GNS	0.049	0.095	0.603
HSD3B2	-0.007	0.017	0.668
ARSG	-0.010	0.029	0.740
AKR1B15	0.001	0.003	0.763
SULT1A1	0.000	0.001	0.786
SULF2	0.001	0.008	0.949
SULT4A1	0.000	0.005	0.997

Table 59. Interaction between Body mass index and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
CYP1B1	-0.015	0.007	0.024
SULT4A1	-0.045	0.033	0.172
GSTM1	-0.005	0.004	0.176
SULT1A2	0.008	0.007	0.204
HSD17B1	0.005	0.004	0.255
CYP1A1	-0.136	0.131	0.300
ESR2	0.025	0.024	0.305
SULT2B1	0.007	0.007	0.323
ARSJ	0.006	0.006	0.339

ARSD	0.008	0.008	0.361
IDS	0.003	0.006	0.564
ARSA	-0.003	0.005	0.593
COMT	-0.002	0.005	0.603
SULT1C2	-0.002	0.004	0.614
ARSF	-0.002	0.004	0.645
SULT1B1	0.006	0.013	0.655
HSD17B14	0.003	0.006	0.658
UGT2B4	-0.003	0.010	0.768
SULT1E1	0.001	0.004	0.785
ESR1	0.002	0.006	0.787
SULF2	-0.001	0.006	0.803
SULF1	-0.007	0.034	0.834
CYP3A4	0.001	0.003	0.855

Table 60. Interaction between Body mass index and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
B3GAT1	0.018	0.010	0.062
HSD17B14	0.022	0.013	0.086
ARSI	0.063	0.045	0.157
ARSD	0.052	0.039	0.182
SULF2	0.030	0.028	0.278
CYP17A1	0.011	0.011	0.288
UGT2B17	0.010	0.011	0.377
STAR	-0.024	0.030	0.419
GALNS	0.005	0.006	0.420
CYP3A4	0.008	0.010	0.434
UGT2B7	-0.012	0.016	0.457
SULF1	0.030	0.042	0.466
SULT1A1	0.043	0.069	0.535
SULT4A1	-0.075	0.126	0.553
SULT1E1	-0.005	0.008	0.557
SULT2B1	-0.002	0.003	0.564
GSTM1	0.002	0.003	0.574
CYP11A1	-0.004	0.008	0.604
SULT1A2	-0.002	0.005	0.730
SULT1A4	0.006	0.018	0.733
SULT1C2	-0.001	0.003	0.774
STS	-0.005	0.027	0.843
ARSA	0.001	0.006	0.856
CYP1B1	-0.002	0.018	0.924
GSTA1	-0.001	0.010	0.937
AKR1B15	0.001	0.010	0.952

Table 61. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
ARSA	-0.085	0.031	0.006
B3GAT2	0.155	0.066	0.019
GSTM1	-0.148	0.077	0.054
SULT1B1	-0.996	0.542	0.066
GALNS	-0.138	0.076	0.068
SULT2B1	0.072	0.055	0.191
IDS	-0.085	0.072	0.237
UGT2B7	-0.177	0.160	0.270
ARSE	-0.068	0.063	0.274
CYP3A4	-0.107	0.099	0.278
SULT1A2	0.057	0.053	0.282
SHBG	0.147	0.142	0.301
COMT	-0.129	0.132	0.327
CYP3A5	-0.160	0.167	0.337
UGT2B17	-0.033	0.039	0.405
GSTA1	0.061	0.074	0.408
SULT4A1	0.030	0.039	0.432
HSD17B1	0.047	0.067	0.481
CYP1A1	0.066	0.101	0.514
CYP1A2	-0.011	0.017	0.516
B3GAT3	0.028	0.046	0.552
HSD3B2	-0.069	0.133	0.605
ARSK	-0.298	0.575	0.605
ARSB	0.047	0.155	0.764
SULT1C2	-0.006	0.029	0.835
UGT2B11	0.019	0.093	0.836
SGSH	-0.013	0.070	0.852
ARSF	-0.021	0.198	0.916

Table 62. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
ESR2	0.123	0.059	0.036
HSD17B11	0.602	0.298	0.043
SULT4A1	-0.078	0.040	0.051
ARSA	-0.104	0.063	0.097
SGSH	-0.520	0.342	0.128

UGT2B28	-0.161	0.112	0.151
UGT2B4	0.049	0.036	0.164
AKR1B15	1.128	0.856	0.187
SULT1A1	-0.710	0.580	0.221
UGT1A4	0.027	0.024	0.258
COMT	0.051	0.048	0.287
B3GAT3	0.055	0.053	0.299
ARSE	-0.059	0.064	0.357
SULT2B1	-0.023	0.026	0.376
B3GAT1	-0.044	0.064	0.487
UGT2B17	0.023	0.039	0.550
STAR	0.040	0.069	0.558
SULF2	-0.038	0.065	0.563
GSTM1	0.052	0.092	0.572
CYP1B1	-0.032	0.057	0.581
CYP1A2	0.016	0.030	0.600
B3GAT2	-0.016	0.032	0.618
GSTA1	0.027	0.055	0.626
HSD3B2	-0.013	0.030	0.679
ARSB	0.005	0.020	0.825
ARSJ	0.024	0.120	0.840
UGT1A6	-0.010	0.054	0.852
CYP11A1	-0.039	0.294	0.894
HSD17B14	0.001	0.082	0.989

Table 63. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
ARSG	-0.158	0.072	0.028
ARSB	0.456	0.214	0.033
HSD17B1	-0.115	0.074	0.119
ARSA	0.022	0.018	0.229
ARSE	-0.133	0.124	0.285
IDS	-0.103	0.097	0.290
ARSD	-0.174	0.181	0.336
SULT1A1	-0.052	0.065	0.426
CYP1A1	0.051	0.072	0.480
B3GAT2	0.033	0.047	0.480
CYP1B1	-0.025	0.036	0.488
HSD17B14	-0.046	0.071	0.517
SULT1C2	0.051	0.097	0.598
UGT2B7	0.038	0.085	0.653
CYP11A1	-0.014	0.038	0.722
CYP3A4	0.067	0.248	0.787

CYP19A1	0.019	0.098	0.845
ARSJ	-0.008	0.048	0.861
SULT1A2	0.009	0.062	0.881
GSTM1	0.012	0.081	0.883

Table 64. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
AKR1B15	0.095	0.032	0.003
STAR	0.054	0.026	0.038
B3GAT2	-0.060	0.030	0.048
SULT1A1	-0.022	0.016	0.169
SULF2	0.113	0.101	0.260
SULT1A4	-0.040	0.037	0.287
NQO1	-0.024	0.024	0.320
GALNS	-0.032	0.048	0.497
ARSG	-0.198	0.347	0.569
HSD3B2	0.113	0.212	0.594
SULT4A1	0.029	0.061	0.631
HSD17B11	0.012	0.030	0.687
GNS	0.438	1.173	0.709
CYP1B1	-0.004	0.194	0.982

Table 65. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
SULT1B1	-0.397	0.164	0.015
SULF2	-0.131	0.070	0.060
IDS	0.120	0.070	0.084
UGT2B4	-0.201	0.123	0.100
ARSA	-0.096	0.059	0.104
COMT	-0.065	0.058	0.259
SULT1C2	0.058	0.052	0.268
CYP1A1	-1.733	1.611	0.282
SULT1A2	-0.076	0.080	0.343
HSD17B14	0.072	0.078	0.354
SULT4A1	-0.345	0.411	0.402
ESR1	-0.051	0.075	0.498

CYP1B1	-0.053	0.080	0.506
ARSF	-0.031	0.048	0.517
CYP3A4	-0.020	0.036	0.565
GSTM1	0.017	0.049	0.724
ARSJ	0.027	0.077	0.729
SULT2B1	-0.021	0.082	0.794
ESR2	0.067	0.292	0.819
ARSD	0.022	0.102	0.828
SULT1E1	-0.011	0.051	0.833
HSD17B1	-0.010	0.051	0.845
SULF1	0.005	0.400	0.991

Table 66. Interaction between Hormone replacement therapy use and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
AKR1B15	0.347	0.122	0.004
STAR	-1.012	0.378	0.007
SULT4A1	-3.453	1.520	0.023
SULT1C2	0.070	0.035	0.046
SULT1A4	-0.374	0.218	0.086
SULT2B1	0.058	0.034	0.088
SULT1A2	0.091	0.064	0.158
SULF1	-0.673	0.494	0.174
GSTA1	0.135	0.123	0.273
CYP17A1	-0.133	0.127	0.294
SULT1A1	-0.764	0.857	0.373
STS	0.244	0.326	0.455
B3GAT1	-0.090	0.120	0.457
HSD17B14	-0.101	0.160	0.529
UGT2B7	0.095	0.195	0.628
CYP3A4	0.058	0.123	0.637
GALNS	-0.033	0.077	0.668
GSTM1	-0.014	0.036	0.695
SULF2	-0.105	0.335	0.753
ARSA	0.021	0.071	0.768
ARSD	0.068	0.471	0.886
CYP11A1	-0.010	0.097	0.920
ARSI	-0.048	0.547	0.930
CYP1B1	0.008	0.222	0.972
SULT1E1	0.004	0.105	0.973
UGT2B17	-0.001	0.133	0.993

Table 67. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
HSD3B2	-0.252	0.129	0.051
CYP1A2	0.033	0.017	0.052
SULT4A1	-0.070	0.038	0.062
SULT2B1	0.091	0.054	0.094
CYP3A5	-0.267	0.163	0.101
SHBG	-0.220	0.141	0.118
ARSF	0.268	0.192	0.163
SULT1A2	0.065	0.051	0.204
ARSB	0.162	0.151	0.284
UGT2B7	-0.167	0.158	0.292
UGT2B11	-0.095	0.092	0.302
ARSE	-0.062	0.060	0.304
B3GAT3	-0.046	0.045	0.310
UGT2B17	0.028	0.038	0.459
GALNS	0.045	0.074	0.539
ARSA	0.016	0.031	0.606
CYP1A1	0.050	0.100	0.613
B3GAT2	0.031	0.066	0.636
SULT1B1	0.231	0.526	0.661
HSD17B1	0.024	0.066	0.712
ARSK	-0.157	0.562	0.779
IDS	0.019	0.072	0.788
GSTA1	0.018	0.073	0.804
SGSH	-0.015	0.068	0.826
CYP3A4	0.019	0.096	0.842
COMT	-0.011	0.129	0.931
GSTM1	0.002	0.076	0.979
SULT1C2	-0.001	0.028	0.985

Table 68. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
AKR1B15	1.597	0.848	0.060
B3GAT1	0.115	0.062	0.064
B3GAT2	0.055	0.031	0.076
UGT2B17	0.062	0.038	0.104
SULF2	0.100	0.064	0.116
SULT1A1	-0.881	0.564	0.118
UGT1A6	0.083	0.054	0.123

CYP1B1	0.084	0.057	0.137
COMT	-0.062	0.047	0.183
UGT1A4	0.027	0.023	0.250
ARSE	-0.069	0.062	0.263
CYP1A2	-0.030	0.029	0.301
HSD3B2	-0.029	0.030	0.335
ARSB	0.016	0.020	0.432
HSD17B14	0.055	0.081	0.500
GSTA1	-0.036	0.053	0.502
SGSH	-0.177	0.338	0.601
UGT2B28	-0.049	0.110	0.654
B3GAT3	0.022	0.052	0.668
STAR	-0.027	0.069	0.693
ESR2	-0.019	0.059	0.741
HSD17B11	0.095	0.290	0.743
UGT2B4	-0.010	0.035	0.782
ARSJ	0.021	0.118	0.856
SULT2B1	-0.004	0.026	0.867
SULT4A1	-0.006	0.039	0.871
ARSA	-0.006	0.062	0.929
GSTM1	-0.006	0.090	0.946
CYP11A1	0.009	0.285	0.974

Table 69. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
ARSB	-0.511	0.212	0.016
ARSJ	-0.106	0.047	0.023
CYP1A1	-0.106	0.071	0.134
CYP1B1	0.052	0.035	0.139
HSD17B1	-0.098	0.073	0.183
IDS	-0.125	0.096	0.192
ARSD	0.212	0.178	0.235
CYP11A1	-0.036	0.038	0.338
HSD17B14	0.059	0.070	0.394
CYP3A4	-0.191	0.244	0.435
UGT2B7	-0.065	0.084	0.438
GSTM1	-0.059	0.080	0.466
SULT1A2	-0.044	0.061	0.468
SULT1C2	-0.046	0.096	0.630
SULT1A1	-0.027	0.064	0.671
ARSA	0.007	0.018	0.677
ARSG	-0.010	0.071	0.892
B3GAT2	-0.004	0.046	0.931

CYP19A1	0.002	0.096	0.982
ARSE	0.001	0.122	0.992

Table 70. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
HSD3B2	-0.577	0.209	0.006
GNS	-2.084	1.157	0.072
CYP1B1	0.304	0.192	0.113
SULT1A1	-0.020	0.016	0.194
SULT4A1	0.066	0.060	0.267
NQO1	0.022	0.024	0.360
STAR	-0.020	0.025	0.434
SULT1A4	0.027	0.037	0.469
ARSG	0.232	0.341	0.497
SULF2	-0.052	0.098	0.598
AKR1B15	-0.011	0.031	0.710
HSD17B11	-0.005	0.029	0.859
B3GAT2	-0.003	0.030	0.912
GALNS	0.004	0.047	0.931

Table 71. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
ARSD	0.262	0.099	0.008
CYP1B1	0.166	0.080	0.038
SULF1	0.655	0.390	0.093
UGT2B4	-0.197	0.121	0.104
CYP3A4	-0.052	0.035	0.138
SULT1A2	-0.115	0.079	0.143
ESR2	-0.326	0.286	0.254
IDS	-0.075	0.067	0.264
SULT1C2	-0.056	0.051	0.280
SULT2B1	-0.082	0.080	0.306
ARSF	-0.039	0.047	0.404
COMT	-0.047	0.057	0.410
ESR1	-0.056	0.074	0.449
ARSJ	-0.055	0.075	0.458

SULT1E1	0.036	0.050	0.475
SULT1B1	0.110	0.158	0.488
ARSA	0.037	0.058	0.528
SULF2	-0.033	0.069	0.632
CYP1A1	0.664	1.574	0.673
SULT4A1	0.164	0.412	0.691
HSD17B1	-0.020	0.050	0.697
GSTM1	0.016	0.048	0.736
HSD17B14	-0.016	0.077	0.836

Table 72. Interaction between oral contraceptive use and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
SULT1E1	-0.267	0.106	0.012
SULT1C2	0.087	0.035	0.012
SULT1A4	-0.403	0.214	0.060
CYP17A1	0.192	0.125	0.125
ARSI	-0.775	0.536	0.148
GSTM1	-0.047	0.035	0.181
STS	0.413	0.323	0.201
SULF1	0.561	0.487	0.250
ARSD	0.528	0.465	0.256
CYP3A4	-0.104	0.121	0.388
STAR	0.333	0.388	0.390
HSD17B14	-0.102	0.158	0.516
CYP11A1	0.060	0.094	0.524
AKR1B15	0.068	0.119	0.567
GALNS	-0.035	0.076	0.649
CYP1B1	0.098	0.218	0.654
UGT2B7	-0.082	0.191	0.669
SULT1A1	-0.324	0.862	0.707
SULT4A1	0.519	1.505	0.730
UGT2B17	0.041	0.132	0.759
ARSA	0.021	0.073	0.772
SULT2B1	0.009	0.034	0.788
GSTA1	-0.020	0.120	0.871
B3GAT1	0.006	0.119	0.961
SULF2	0.008	0.328	0.980
SULT1A2	0.000	0.063	0.998

Table 73. Interaction between age at menarche and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
UGT2B17	0.021	0.011	0.056
HSD3B2	-0.071	0.038	0.058
HSD17B1	0.031	0.019	0.103
GALNS	0.034	0.022	0.115
CYP3A4	-0.040	0.029	0.163
SULT1A2	0.021	0.015	0.165
B3GAT3	-0.017	0.013	0.195
SULT1C2	-0.010	0.008	0.228
CYP1A1	0.031	0.029	0.293
ARSA	-0.009	0.008	0.294
B3GAT2	-0.019	0.018	0.305
UGT2B11	0.027	0.026	0.309
CYP1A2	-0.005	0.005	0.318
ARSK	0.156	0.162	0.337
ARSB	0.040	0.044	0.359
SULT4A1	-0.010	0.011	0.361
SHBG	0.035	0.040	0.383
SULT1B1	0.126	0.151	0.406
COMT	-0.022	0.037	0.561
CYP3A5	-0.027	0.048	0.567
ARSE	0.010	0.018	0.587
GSTM1	0.011	0.022	0.604
SGSH	0.009	0.020	0.650
GSTA1	0.009	0.021	0.676
IDS	0.006	0.019	0.766
UGT2B7	-0.011	0.045	0.804
SULT2B1	-0.002	0.016	0.901
ARSF	0.005	0.058	0.934

Table 74. Interaction between age at menarche and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
HSD17B11	-0.149	0.086	0.082
CYP1B1	-0.027	0.016	0.106
UGT2B4	0.016	0.010	0.107
UGT1A6	0.023	0.015	0.131
STAR	0.025	0.019	0.186
ARSA	-0.023	0.018	0.188

B3GAT3	-0.020	0.015	0.191
ESR2	-0.019	0.016	0.246
B3GAT2	0.008	0.009	0.366
AKR1B15	-0.218	0.243	0.369
SULT2B1	-0.006	0.007	0.382
HSD17B14	0.020	0.023	0.383
ARSJ	-0.027	0.035	0.430
SULT4A1	0.009	0.011	0.438
CYP11A1	-0.065	0.084	0.439
SULF2	0.014	0.019	0.452
SGSH	0.071	0.096	0.456
ARSB	0.003	0.006	0.561
SULT1A1	0.088	0.164	0.592
COMT	-0.007	0.014	0.600
HSD3B2	0.003	0.009	0.686
CYP1A2	-0.003	0.008	0.719
GSTA1	0.005	0.015	0.742
GSTM1	0.006	0.026	0.809
B3GAT1	0.003	0.018	0.856
UGT2B17	0.001	0.011	0.896
ARSE	-0.002	0.019	0.898
UGT1A4	-0.001	0.007	0.935
UGT2B28	0.003	0.032	0.936

Table 75. Interaction between age at menarche and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
ARSD	-0.095	0.051	0.062
CYP11A1	-0.018	0.011	0.101
ARSG	-0.031	0.021	0.131
ARSA	0.008	0.005	0.145
ARSE	-0.049	0.035	0.166
B3GAT2	0.016	0.013	0.214
SULT1C2	-0.031	0.027	0.244
CYP1A1	0.023	0.021	0.262
UGT2B7	0.021	0.024	0.382
HSD17B1	0.018	0.021	0.396
CYP3A4	-0.058	0.070	0.402
SULT1A2	-0.014	0.018	0.412
GSTM1	0.018	0.023	0.426
CYP1B1	0.007	0.010	0.521
CYP19A1	0.016	0.027	0.550
ARSJ	-0.007	0.014	0.623
IDS	-0.010	0.027	0.723

HSD17B14	-0.006	0.020	0.768
ARSB	-0.013	0.060	0.825
SULT1A1	0.003	0.018	0.887

Table 76. Interaction between age at menarche and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
AKR1B15	0.021	0.009	0.022
NQO1	-0.010	0.007	0.139
HSD3B2	0.082	0.058	0.158
SULT1A4	-0.015	0.010	0.164
CYP1B1	-0.060	0.055	0.274
SULT4A1	-0.019	0.017	0.278
STAR	-0.007	0.008	0.379
GNS	-0.234	0.326	0.474
ARSG	0.052	0.097	0.593
GALNS	0.006	0.013	0.630
SULF2	-0.014	0.029	0.635
SULT1A1	0.002	0.005	0.685
HSD17B11	-0.003	0.008	0.726
B3GAT2	0.000	0.009	0.966

Table 77. Interaction between age at menarche and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
CYP1A1	-1.019	0.452	0.024
HSD17B14	-0.038	0.022	0.086
ARSF	-0.022	0.014	0.101
CYP1B1	0.029	0.022	0.188
SULT1B1	-0.060	0.047	0.198
UGT2B4	-0.044	0.035	0.208
SULT1E1	0.016	0.014	0.267
SULF2	-0.019	0.020	0.336
SULT2B1	-0.022	0.023	0.341
ESR1	0.020	0.021	0.357
IDS	0.018	0.020	0.370

ARSA	0.013	0.016	0.421
ARSJ	-0.015	0.022	0.483
HSD17B1	0.009	0.014	0.527
SULF1	-0.064	0.114	0.573
SULT1A2	-0.012	0.023	0.593
GSTM1	0.007	0.014	0.599
ESR2	0.038	0.084	0.654
CYP3A4	-0.003	0.010	0.796
COMT	0.003	0.016	0.856
ARSD	-0.005	0.030	0.859
SULT4A1	-0.019	0.113	0.865
SULT1C2	-0.002	0.015	0.903

Table 78. Interaction between age at menarche and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
SULT1E1	-0.049	0.028	0.082
AKR1B15	0.056	0.035	0.108
SULF1	-0.223	0.148	0.134
STAR	-0.142	0.096	0.141
GALNS	-0.031	0.022	0.155
GSTM1	-0.013	0.010	0.204
UGT2B17	-0.042	0.037	0.251
SULT1C2	-0.011	0.010	0.290
HSD17B14	0.048	0.045	0.293
UGT2B7	0.059	0.056	0.298
GSTA1	-0.033	0.035	0.355
ARSI	0.133	0.154	0.390
SULF2	-0.075	0.096	0.431
CYP3A4	0.027	0.035	0.441
SULT1A4	0.046	0.061	0.451
B3GAT1	-0.017	0.033	0.619
CYP1B1	0.030	0.064	0.639
SULT1A1	-0.088	0.227	0.697
CYP11A1	0.008	0.027	0.766
SULT2B1	-0.003	0.010	0.789
CYP17A1	-0.007	0.036	0.847
STS	-0.014	0.092	0.878
ARSA	-0.003	0.019	0.893
SULT4A1	-0.033	0.430	0.940
SULT1A2	-0.001	0.018	0.968
ARSD	0.001	0.136	0.997

Table 79. Interaction between age at menopause and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
SULT1B1	0.146	0.058	0.012
GSTA1	0.020	0.008	0.014
ARSK	-0.153	0.063	0.015
ARSE	-0.015	0.007	0.027
CYP1A2	0.003	0.002	0.082
SULT4A1	0.007	0.004	0.085
B3GAT3	0.008	0.005	0.127
SGSH	0.012	0.008	0.132
SULT1C2	-0.005	0.003	0.133
ARSF	-0.032	0.022	0.147
HSD3B2	0.019	0.015	0.189
SULT2B1	-0.007	0.006	0.230
SHBG	0.013	0.016	0.415
HSD17B1	-0.005	0.007	0.469
IDS	0.006	0.008	0.483
GALNS	0.006	0.008	0.489
SULT1A2	0.004	0.006	0.513
CYP1A1	0.006	0.011	0.577
UGT2B11	-0.004	0.010	0.670
CYP3A5	-0.008	0.018	0.674
GSTM1	-0.003	0.008	0.687
CYP3A4	-0.004	0.011	0.721
UGT2B17	-0.001	0.004	0.752
ARSB	0.005	0.017	0.785
ARSA	0.001	0.004	0.852
COMT	-0.002	0.015	0.876
B3GAT2	-0.001	0.007	0.909
UGT2B7	0.002	0.018	0.915

Table 80. Interaction between age at menopause and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
HSD3B2	0.008	0.003	0.025
ARSE	-0.016	0.007	0.027
HSD17B14	0.019	0.009	0.033
AKR1B15	-0.180	0.095	0.058
CYP1B1	-0.011	0.006	0.083

CYP1A2	0.006	0.003	0.097
SGSH	-0.060	0.038	0.113
ARSJ	-0.019	0.013	0.155
ESR2	0.007	0.007	0.268
B3GAT3	-0.006	0.006	0.296
COMT	0.005	0.005	0.329
B3GAT2	0.003	0.004	0.381
B3GAT1	-0.005	0.007	0.444
SULF2	0.004	0.007	0.546
UGT2B4	-0.002	0.004	0.653
UGT1A6	-0.003	0.006	0.656
SULT1A1	0.027	0.063	0.667
GSTM1	0.004	0.010	0.680
STAR	0.003	0.008	0.717
SULT2B1	-0.001	0.003	0.717
GSTA1	0.002	0.006	0.763
UGT1A4	-0.001	0.003	0.780
UGT2B28	-0.003	0.012	0.820
ARSA	-0.002	0.007	0.829
HSD17B11	0.006	0.033	0.845
SULT4A1	0.000	0.004	0.921
CYP11A1	-0.003	0.033	0.922
ARSB	0.000	0.002	0.969
UGT2B17	0.000	0.004	0.999

Table 81. Interaction between age at menopause and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
ARSD	-0.039	0.020	0.056
ARSB	0.041	0.025	0.090
SULT1A2	0.009	0.007	0.196
ARSA	-0.002	0.002	0.224
CYP1A1	0.009	0.008	0.281
HSD17B1	-0.006	0.008	0.444
B3GAT2	0.004	0.005	0.471
ARSJ	0.004	0.005	0.473
SULT1A1	-0.005	0.007	0.474
CYP1B1	0.003	0.004	0.477
ARSG	0.006	0.008	0.483
UGT2B7	0.006	0.009	0.514
CYP11A1	0.002	0.004	0.580
CYP19A1	0.003	0.011	0.793
CYP3A4	-0.006	0.028	0.831
HSD17B14	0.001	0.008	0.882

SULT1C2	0.002	0.011	0.884
IDS	0.001	0.011	0.940
GSTM1	-0.001	0.009	0.947
ARSE	-0.001	0.014	0.970

Table 82. Interaction between age at menopause and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
SULT1A1	-0.002	0.002	0.180
GALNS	-0.007	0.005	0.191
HSD17B11	0.004	0.003	0.212
SULT1A4	-0.005	0.004	0.222
NQO1	-0.002	0.003	0.389
CYP1B1	-0.016	0.021	0.469
GNS	-0.091	0.131	0.488
SULT4A1	0.004	0.007	0.598
HSD3B2	0.006	0.023	0.795
ARSG	-0.009	0.038	0.814
STAR	0.001	0.003	0.817
B3GAT2	-0.001	0.003	0.836
SULF2	-0.001	0.011	0.916
AKR1B15	0.000	0.003	0.950

Table 83. Interaction between age at menopause and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
ESR2	0.082	0.032	0.010
HSD17B1	0.010	0.006	0.065
CYP1A1	-0.291	0.175	0.096
UGT2B4	-0.022	0.014	0.099
ARSD	-0.014	0.011	0.206
SULF2	0.010	0.008	0.221
ARSJ	0.010	0.008	0.222
ARSF	0.006	0.005	0.227
SULT1B1	0.020	0.018	0.268
SULT4A1	-0.047	0.047	0.315
ARSA	-0.006	0.007	0.367
IDS	0.006	0.008	0.407

SULT1E1	-0.003	0.006	0.553
COMT	-0.004	0.006	0.561
CYP3A4	0.002	0.004	0.652
SULT1A2	0.004	0.009	0.671
SULT2B1	-0.004	0.009	0.692
CYP1B1	0.002	0.009	0.789
HSD17B14	0.002	0.009	0.807
ESR1	-0.002	0.008	0.848
GSTM1	0.000	0.005	0.960
SULF1	-0.001	0.044	0.992
SULT1C2	0.000	0.006	0.992

Table 84. Interaction between age at menopause and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
GSTM1	-0.009	0.004	0.018
SULT1A4	-0.041	0.024	0.085
AKR1B15	-0.020	0.013	0.143
HSD17B14	-0.024	0.018	0.172
SULT2B1	0.005	0.004	0.207
SULF1	-0.068	0.055	0.213
CYP3A4	0.014	0.014	0.307
UGT2B7	0.021	0.022	0.321
ARSA	-0.007	0.008	0.369
ARSD	0.036	0.052	0.488
SULT1A1	0.064	0.096	0.506
SULT1A2	0.005	0.007	0.512
B3GAT1	0.008	0.013	0.528
CYP11A1	0.006	0.011	0.573
GSTA1	-0.007	0.013	0.585
SULF2	-0.018	0.037	0.639
UGT2B17	0.006	0.015	0.702
SULT1E1	-0.004	0.012	0.757
GALNS	0.002	0.009	0.781
SULT4A1	0.037	0.168	0.827
SULT1C2	-0.001	0.004	0.848
CYP17A1	-0.003	0.014	0.849
CYP1B1	0.004	0.024	0.861
STS	-0.002	0.036	0.963
ARSI	-0.001	0.060	0.988
STAR	0.000	0.044	0.995

Table 85. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
UGT2B7	-0.033	0.016	0.042
CYP1A1	-0.021	0.011	0.047
COMT	-0.022	0.014	0.108
SULT2B1	-0.009	0.006	0.122
CYP1A2	0.002	0.002	0.313
GSTA1	0.006	0.007	0.417
UGT2B17	0.003	0.004	0.476
CYP3A4	-0.007	0.010	0.482
GALNS	0.005	0.008	0.495
HSD17B1	0.004	0.007	0.513
ARSA	0.002	0.003	0.515
B3GAT2	0.004	0.006	0.531
IDS	0.004	0.007	0.537
ARSB	-0.010	0.016	0.549
SULT1B1	-0.033	0.055	0.552
ARSK	-0.034	0.059	0.567
CYP3A5	-0.007	0.017	0.700
SHBG	-0.004	0.014	0.770
SULT1C2	-0.001	0.003	0.806
ARSF	0.004	0.021	0.853
ARSE	-0.001	0.007	0.865
GSTM1	0.001	0.008	0.872
B3GAT3	-0.001	0.005	0.876
UGT2B11	0.002	0.010	0.878
HSD3B2	0.002	0.014	0.893
SULT1A2	-0.001	0.005	0.901
SULT4A1	0.000	0.004	0.942
SGSH	0.000	0.007	0.969

Table 86. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
CYP1B1	0.011	0.006	0.053
UGT2B17	0.007	0.004	0.063
ARSA	0.011	0.006	0.086
GSTA1	0.009	0.006	0.093
UGT1A4	-0.004	0.002	0.110
UGT1A6	-0.009	0.005	0.110
B3GAT2	0.005	0.003	0.128
SULT2B1	0.004	0.003	0.176

B3GAT3	-0.006	0.005	0.253
COMT	-0.006	0.005	0.263
HSD17B14	0.009	0.008	0.272
ARSJ	-0.012	0.013	0.325
HSD3B2	0.003	0.003	0.350
ARSE	-0.006	0.007	0.363
HSD17B11	-0.018	0.031	0.573
ARSB	-0.001	0.002	0.632
ESR2	0.003	0.006	0.652
SULF2	0.002	0.007	0.742
B3GAT1	0.002	0.006	0.745
SGSH	0.011	0.034	0.756
SULT1A1	0.018	0.060	0.762
UGT2B4	0.001	0.004	0.775
UGT2B28	-0.003	0.011	0.785
CYP1A2	-0.001	0.003	0.806
STAR	0.002	0.007	0.809
GSTM1	0.001	0.010	0.903
CYP11A1	-0.003	0.030	0.916
AKR1B15	-0.008	0.088	0.931
SULT4A1	0.000	0.004	0.991

Table 87. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
HSD17B14	0.014	0.007	0.058
CYP1A1	0.013	0.007	0.071
SULT1C2	-0.017	0.010	0.077
CYP11A1	0.006	0.004	0.121
UGT2B7	0.013	0.009	0.125
ARSA	-0.003	0.002	0.128
CYP1B1	0.005	0.004	0.171
HSD17B1	-0.010	0.008	0.196
IDS	-0.010	0.010	0.289
CYP19A1	0.009	0.010	0.354
ARSG	0.006	0.007	0.444
ARSE	-0.006	0.013	0.658
GSTM1	0.004	0.008	0.676
SULT1A1	-0.002	0.007	0.717
CYP3A4	0.008	0.025	0.749
ARSB	-0.007	0.022	0.756
ARSJ	-0.001	0.005	0.813

B3GAT2	0.001	0.005	0.865
SULT1A2	-0.001	0.006	0.914
ARSD	0.001	0.018	0.945

Table 88. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
HSD17B11	-0.010	0.003	0.001
STAR	-0.006	0.003	0.031
CYP1B1	-0.038	0.020	0.051
HSD3B2	-0.036	0.021	0.090
B3GAT2	0.005	0.003	0.137
GALNS	-0.005	0.005	0.291
SULT1A1	-0.001	0.002	0.419
NQO1	-0.002	0.003	0.486
SULT1A4	-0.003	0.004	0.488
ARSG	0.023	0.035	0.521
AKR1B15	-0.002	0.003	0.580
GNS	0.038	0.117	0.747
SULT4A1	0.001	0.006	0.843
SULF2	-0.002	0.010	0.844

Table 89. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
SULT1C2	-0.013	0.005	0.014
SULF2	0.013	0.007	0.068
ARSJ	-0.013	0.008	0.106
COMT	0.009	0.006	0.126
ARSF	0.007	0.005	0.154
SULT1A2	-0.011	0.008	0.192
SULF1	0.048	0.041	0.243
HSD17B1	-0.004	0.005	0.382
ESR1	-0.006	0.008	0.441
CYP1B1	-0.006	0.008	0.476
CYP1A1	0.113	0.162	0.488
SULT1E1	0.004	0.005	0.502
HSD17B14	-0.005	0.008	0.510
SULT1B1	-0.011	0.017	0.528

SULT2B1	0.004	0.008	0.610
ESR2	0.012	0.030	0.687
UGT2B4	-0.005	0.012	0.706
SULT4A1	-0.012	0.040	0.763
ARSA	0.002	0.006	0.773
IDS	0.002	0.007	0.814
ARSD	0.002	0.011	0.834
GSTM1	0.001	0.005	0.898
CYP3A4	0.000	0.004	0.968

Table 90. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
SULF1	0.140	0.053	0.008
ARSD	-0.079	0.049	0.104
UGT2B17	0.021	0.013	0.117
GSTM1	-0.005	0.004	0.153
STS	0.040	0.033	0.222
ARSA	-0.007	0.007	0.261
SULT2B1	-0.003	0.004	0.367
CYP17A1	-0.011	0.013	0.381
CYP1B1	-0.017	0.023	0.459
SULF2	-0.024	0.034	0.484
SULT1E1	0.007	0.010	0.506
SULT4A1	0.090	0.153	0.554
UGT2B7	0.012	0.021	0.572
SULT1C2	0.001	0.004	0.738
SULT1A4	-0.007	0.022	0.748
CYP3A4	0.004	0.013	0.757
HSD17B14	0.005	0.016	0.769
CYP11A1	0.003	0.010	0.778
SULT1A2	0.002	0.007	0.780
AKR1B15	-0.003	0.013	0.820
B3GAT1	0.003	0.012	0.834
STAR	-0.007	0.035	0.841
GSTA1	0.002	0.013	0.879
GALNS	0.001	0.008	0.906
ARSI	0.006	0.056	0.912
SULT1A1	0.003	0.081	0.972

Table 91. Interaction between parity and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
SULT2B1	0.036	0.018	0.038
SULT1B1	0.343	0.171	0.045
GALNS	0.037	0.024	0.130
UGT2B11	-0.044	0.030	0.140
GSTA1	-0.033	0.023	0.152
SGSH	0.029	0.022	0.180
HSD3B2	-0.054	0.042	0.198
CYP3A5	0.066	0.053	0.217
SHBG	-0.055	0.045	0.217
CYP3A4	0.036	0.032	0.256
CYP1A2	0.005	0.005	0.315
SULT1A2	0.017	0.017	0.326
COMT	0.038	0.042	0.368
ARSF	-0.047	0.064	0.462
HSD17B1	-0.015	0.021	0.467
IDS	0.015	0.022	0.483
ARSE	-0.014	0.021	0.507
UGT2B17	-0.007	0.012	0.552
B3GAT3	0.008	0.015	0.567
B3GAT2	0.009	0.021	0.676
GSTM1	-0.008	0.024	0.733
ARSA	0.002	0.009	0.830
ARSK	-0.039	0.184	0.832
CYP1A1	0.003	0.033	0.917
SULT4A1	0.001	0.012	0.929
ARSB	0.002	0.049	0.969
SULT1C2	0.000	0.009	0.977
UGT2B7	0.000	0.051	1.000

Table 92. Interaction between parity and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
HSD17B11	0.216	0.096	0.025
UGT2B4	-0.025	0.011	0.028
STAR	-0.037	0.022	0.091
SULF2	-0.035	0.021	0.097
B3GAT1	-0.027	0.020	0.176
ARSA	0.026	0.020	0.189
HSD17B14	0.032	0.026	0.208

SULT1A1	-0.227	0.186	0.222
B3GAT2	0.011	0.010	0.253
B3GAT3	0.019	0.017	0.262
GSTA1	-0.019	0.017	0.267
SGSH	-0.117	0.108	0.278
AKR1B15	0.291	0.275	0.290
SULT4A1	0.013	0.013	0.304
ESR2	0.014	0.018	0.457
SULT2B1	0.006	0.008	0.497
ARSJ	0.019	0.039	0.620
GSTM1	0.013	0.030	0.671
ARSE	-0.009	0.021	0.674
UGT1A6	-0.006	0.017	0.706
CYP1A2	0.003	0.010	0.778
CYP1B1	-0.004	0.018	0.838
UGT2B28	-0.007	0.035	0.840
COMT	0.003	0.015	0.862
CYP11A1	-0.014	0.094	0.881
ARSB	0.001	0.006	0.886
HSD3B2	-0.001	0.010	0.926
UGT2B17	0.001	0.012	0.949
UGT1A4	0.000	0.008	0.966

Table 93. Interaction between parity and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
GSTM1	-0.057	0.026	0.028
CYP1B1	0.015	0.011	0.185
HSD17B14	0.025	0.022	0.260
CYP1A1	-0.024	0.023	0.291
SULT1A2	-0.018	0.020	0.358
CYP3A4	0.067	0.078	0.392
ARSG	-0.020	0.023	0.392
ARSJ	0.013	0.015	0.395
UGT2B7	-0.021	0.027	0.444
ARSE	0.029	0.039	0.459
ARSB	0.046	0.067	0.493
HSD17B1	-0.016	0.024	0.510
CYP19A1	-0.020	0.031	0.525
SULT1C2	0.017	0.030	0.571
B3GAT2	-0.006	0.015	0.671
SULT1A1	-0.006	0.021	0.756
ARSA	-0.001	0.006	0.892
ARSD	-0.007	0.057	0.910

CYP11A1	0.001	0.012	0.928
IDS	0.001	0.030	0.964

Table 94. Interaction between parity and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
SULT1A4	0.030	0.012	0.012
STAR	0.014	0.008	0.094
AKR1B15	-0.015	0.010	0.153
ARSG	0.144	0.110	0.190
HSD3B2	-0.077	0.066	0.243
SULT4A1	-0.017	0.019	0.376
HSD17B11	0.006	0.009	0.509
GNS	-0.166	0.364	0.649
SULT1A1	0.002	0.005	0.669
CYP1B1	-0.026	0.061	0.669
GALNS	0.005	0.015	0.724
B3GAT2	-0.003	0.010	0.748
SULF2	0.009	0.032	0.784
NQO1	0.002	0.008	0.840

Table 95. Interaction between parity and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
UGT2B4	-0.088	0.039	0.023
ARSJ	0.044	0.025	0.073
SULT1B1	0.080	0.053	0.134
GSTM1	-0.023	0.015	0.142
SULF2	-0.029	0.022	0.188
SULT4A1	-0.113	0.127	0.376
SULT1E1	0.013	0.016	0.433
SULT2B1	-0.019	0.026	0.464
ESR1	0.017	0.024	0.470
CYP1B1	0.018	0.025	0.472
ARSA	-0.011	0.019	0.546
SULT1A2	0.011	0.026	0.664
SULF1	0.056	0.128	0.664
ESR2	0.039	0.094	0.681
SULT1C2	0.006	0.016	0.732

COMT	-0.006	0.018	0.735
HSD17B1	0.005	0.016	0.747
IDS	-0.006	0.022	0.800
CYP3A4	0.002	0.011	0.834
ARSF	-0.003	0.015	0.852
HSD17B14	0.004	0.025	0.858
ARSD	0.006	0.033	0.867
CYP1A1	0.049	0.507	0.923

Table 96. Interaction between parity and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
UGT2B17	-0.089	0.041	0.030
SULT1A4	-0.125	0.068	0.067
CYP3A4	-0.063	0.039	0.109
SULT1E1	-0.047	0.032	0.140
B3GAT1	0.055	0.038	0.145
UGT2B7	0.089	0.064	0.163
HSD17B14	0.068	0.051	0.181
AKR1B15	0.047	0.039	0.233
SULF1	0.187	0.162	0.250
SULT2B1	0.011	0.011	0.298
GSTA1	0.040	0.039	0.307
SULF2	-0.102	0.108	0.341
CYP11A1	0.028	0.031	0.360
STS	-0.085	0.103	0.408
CYP1B1	-0.056	0.071	0.434
CYP17A1	-0.032	0.041	0.435
ARSI	0.128	0.175	0.465
SULT1C2	0.008	0.011	0.492
SULT4A1	0.272	0.478	0.570
SULT1A2	0.011	0.020	0.581
STAR	-0.049	0.113	0.667
GALNS	-0.006	0.025	0.812
ARSD	0.036	0.155	0.818
GSTM1	0.001	0.012	0.918
ARSA	-0.001	0.022	0.963
SULT1A1	-0.005	0.261	0.984

Table 97. Interaction between parous and predicted gene expression level on breast cancer risk in breast tissue model.

Gene name	Effect	SE	P-value
SULT1B1	1.298	0.651	0.046
GSTA1	-0.147	0.088	0.095
SULT2B1	0.111	0.067	0.097
SULT1A2	0.095	0.064	0.140
HSD17B1	-0.113	0.079	0.153
GALNS	0.115	0.094	0.222
HSD3B2	-0.190	0.161	0.239
CYP3A4	0.142	0.124	0.255
UGT2B7	-0.196	0.192	0.305
ARSK	0.696	0.699	0.320
UGT2B11	-0.110	0.112	0.323
ARSE	-0.073	0.080	0.361
CYP1A2	0.018	0.020	0.372
ARSB	-0.148	0.187	0.429
UGT2B17	0.033	0.046	0.477
B3GAT3	-0.033	0.056	0.551
ARSA	-0.020	0.035	0.562
IDS	-0.039	0.081	0.630
CYP3A5	0.079	0.207	0.704
GSTM1	0.034	0.093	0.714
SULT4A1	-0.016	0.046	0.728
COMT	0.047	0.161	0.770
SGSH	0.021	0.083	0.795
ARSF	0.063	0.249	0.799
CYP1A1	0.029	0.127	0.818
SULT1C2	0.003	0.035	0.937
SHBG	0.011	0.171	0.951
B3GAT2	0.003	0.076	0.968

Table 98. Interaction between parous and predicted gene expression level on breast cancer risk in liver tissue.

Gene name	Effect	SE	P-value
SULT2B1	0.079	0.031	0.011
AKR1B15	1.969	1.036	0.057
UGT2B4	-0.079	0.042	0.058
ARSE	-0.146	0.083	0.076
ESR2	0.115	0.068	0.091
B3GAT1	-0.125	0.075	0.095
UGT1A4	-0.047	0.029	0.100

HSD17B14	0.155	0.095	0.102
ARSA	0.120	0.075	0.112
UGT1A6	-0.089	0.063	0.160
SULF2	-0.112	0.080	0.164
B3GAT2	0.040	0.038	0.298
UGT2B28	-0.137	0.134	0.306
CYP11A1	-0.370	0.361	0.306
ARSB	-0.023	0.024	0.332
GSTA1	-0.059	0.066	0.376
CYP1B1	-0.060	0.070	0.391
GSTM1	0.094	0.114	0.409
HSD17B11	0.265	0.373	0.478
UGT2B17	0.033	0.047	0.478
HSD3B2	0.021	0.037	0.565
SGSH	0.231	0.405	0.568
ARSJ	0.053	0.151	0.727
B3GAT3	0.018	0.063	0.773
CYP1A2	0.006	0.037	0.866
SULT1A1	-0.117	0.712	0.870
COMT	0.008	0.059	0.891
STAR	-0.009	0.081	0.909
SULT4A1	-0.002	0.049	0.964

Table 99. Interaction between parous and predicted gene expression level on breast cancer risk in ovary tissue.

Gene name	Effect	SE	P-value
IDS	-0.385	0.113	0.001
ARSG	-0.187	0.088	0.034
HSD17B1	-0.184	0.088	0.037
HSD17B14	0.150	0.085	0.078
B3GAT2	0.082	0.055	0.140
ARSD	-0.308	0.214	0.151
CYP11A1	0.063	0.045	0.167
ARSB	0.279	0.252	0.269
GSTM1	-0.102	0.098	0.295
CYP19A1	-0.102	0.117	0.385
CYP1B1	0.036	0.043	0.399
ARSE	-0.118	0.149	0.430
SULT1C2	0.083	0.113	0.466
CYP1A1	-0.043	0.088	0.626
ARSJ	0.023	0.058	0.692
ARSA	0.008	0.022	0.724
SULT1A2	0.023	0.076	0.764
CYP3A4	-0.075	0.296	0.800

UGT2B7	0.016	0.102	0.876
SULT1A1	0.007	0.079	0.927

Table 100. Interaction between parous and predicted gene expression level on breast cancer risk in subcutaneous adipose tissue.

Gene name	Effect	SE	P-value
SULT1A4	0.125	0.045	0.005
CYP1B1	-0.36	0.234	0.123
ARSG	0.602	0.416	0.148
AKR1B15	-0.055	0.04	0.162
SULT1A1	0.017	0.019	0.392
B3GAT2	-0.028	0.037	0.453
HSD3B2	-0.135	0.245	0.583
SULT4A1	-0.039	0.073	0.591
SULF2	-0.051	0.124	0.683
GNS	-0.443	1.385	0.749
GALNS	-0.013	0.057	0.827
STAR	-0.002	0.032	0.95
NQO1	0.001	0.03	0.973
HSD17B11	0.001	0.036	0.975

Table 101. Interaction between parous and predicted gene expression level on breast cancer risk in visceral adipose tissue.

Gene name	Effect	SE	P-value
SULT1B1	0.431	0.204	0.035
SULT1C2	0.107	0.062	0.086
ARSD	-0.199	0.129	0.121
UGT2B4	-0.203	0.147	0.167
ARSJ	0.107	0.094	0.256
CYP1B1	0.090	0.094	0.336
IDS	0.071	0.086	0.405
GSTM1	-0.040	0.058	0.492
SULT1A2	-0.060	0.097	0.538
CYP3A4	0.026	0.043	0.549
HSD17B1	-0.031	0.060	0.607
HSD17B14	-0.047	0.094	0.618
SULT4A1	-0.219	0.464	0.637
SULF1	0.217	0.492	0.659
ESR1	0.034	0.092	0.712
ESR2	-0.114	0.358	0.749

SULT2B1	-0.025	0.099	0.798
SULT1E1	-0.014	0.061	0.825
CYP1A1	-0.411	1.897	0.829
ARSF	0.011	0.059	0.857
ARSA	-0.010	0.069	0.888
COMT	-0.005	0.067	0.942
SULF2	-0.005	0.083	0.949

Table 102. Interaction between parous and predicted gene expression level on breast cancer risk in cross tissue.

Gene name	Effect	SE	P-value
B3GAT1	0.498	0.140	0.0004
UGT2B17	-0.363	0.155	0.019
STS	-0.670	0.387	0.083
SULT1E1	-0.192	0.115	0.095
CYP17A1	-0.221	0.155	0.154
CYP1B1	-0.374	0.273	0.171
UGT2B7	0.323	0.245	0.187
SULT1A4	-0.335	0.265	0.205
GSTM1	0.055	0.044	0.210
GSTA1	0.184	0.152	0.228
ARSI	0.713	0.664	0.283
ARSA	0.078	0.077	0.315
SULT1C2	0.042	0.042	0.318
SULF2	-0.397	0.412	0.335
GALNS	-0.078	0.093	0.402
ARSD	0.469	0.592	0.427
AKR1B15	0.114	0.150	0.446
HSD17B14	0.145	0.191	0.449
SULT2B1	-0.018	0.042	0.675
CYP11A1	-0.044	0.118	0.707
STAR	-0.140	0.407	0.731
SULF1	0.178	0.640	0.781
CYP3A4	-0.029	0.151	0.846
SULT1A1	0.145	0.944	0.878
SULT4A1	-0.132	1.811	0.942
SULT1A2	-0.003	0.077	0.965

Table 103. Summary for the genes that have a significant (P < 0.05) interaction with breast cancer risk factors.

	breast	Subcutaneous adipose	Visceral adipose	cross	liver	ovary
BMI	*COMT* *SULT4A1* *ARSA*	*NQO1*	*CYP1B1*			
HRT	*ARSA* *B3GAT2*	*AKR1B15* *STAR* *B3GAT2*	*SULT1B1*	*AKR1B15* *STAR* *SULT4A1* *SULT1C2*	*ESR2* *HSD17B11*	*ARSG* *ARSB*
OC		*HSD3B2*	*ARSD* *CYP1B1*	*SULT1E1* *SULT1C2*		*ARSB* *ARSJ*
Parity	*SULT2B1* *SULT1B1*	*SULT1A4*	*UGT2B4*	*UGT2B17*	*HSD17B11* *UGT2B4*	*GSTM1*
Parous	*SULT1B1*	*SULT1A4*	*SULT1B1*	*B3GAT1* *UGT2B17*	*SULT2B1*	*IDS* *ARSG* *HSD17B1*
Age at the end of first full time pregnancy	*UGT2B7* *CYP1A1*	*HSD17B11* *STAR*	*SULT1C2*	*SULF1*		
Age at Menarche		*AKR1B15*	*CYP1A1*			
Age at menopause	*SULT1B1* *GSTA1* *ARSK* *ARSE*		*ESR2*	*GSTM1*	*HSD3B2* *ARSE* *HSD17B14*	

Table 104. Summary for the genes that have a significant (P < 0.01) interaction with breast cancer risk factors.

	breast	Subcutaneous adipose	Visceral adipose	cross	liver	ovary
BMI						
HRT	*ARSA* (0.006)	*AKR1B15* (0.003)		*AKR1B15* (0.004)		
OC		*HSD3B2* (0.006)	*ARSD* (0.008)			
Parity						
Parous		*SULT1A4* (0.005)		*B3GAT1* (0.0004)		*IDS* (0.001)
Age at the end of first full time pregnancy		*HSD17B11* (0.001)		*SULF1* (0.008)		
Age at Menarche						
Age at menopause						

Table 105. Summary for the genes that have a significant (P < 0.005) interaction with breast cancer risk factors.

	Breast	Subcutaneous adipose	Visceral adipose	Cross	Liver	Ovary
BMI						
HRT		*AKR1B15* (0.003)		*AKR1B15* (0.004)		
OC						
Parity						
Parous				*B3GAT1* (0.0004)		*IDS* (0.001)
Age at the end of first full time pregnancy		*HSD17B11* (0.001)				
Age at Menarche						
Age at menopause						

Table 107 to Table 109 showed the results for aggregated pathway analysis. Table 106 summarized number of genes available in the analysis for each pathway across different tissues. For the models of all 6 tissues, at least 2 genes were included in the estrogen synthesis pathway and at least 10 genes were included in the estrogen metabolism pathway. However, for estrogen bioavailability pathway, 2 genes were only available in the model of visceral adipose tissue.

Regarding the aggregated analysis in estrogen synthesis pathway, HRT was observed to have a significant interaction in both breast (P=0.026) and subcutaneous adipose (P=0.003) tissue models. OC use (P = 0.028) and age at FFTP (P = 6.32 * 10-5) also had a significant interaction in the model built within subcutaneous adipose tissue. (Table 107)

Table 106. Number of genes used for each aggregated gene expression of each pathway in each tissue.

Tissues	Estrogen synthesis pathway	Estrogen metabolism pathway	Estrogen bioavailability pathway
Breast	2	25	1
Liver	6	22	0
Ovary	4	16	0
Subcutaneous adipose	4	10	0
Visceral adipose	2	19	2
Cross	5	21	0

Table 107. P-values for the interactions between aggregated predicted gene expression level in estrogen **synthesis** pathway and breast cancer risk factors on breast cancer risk.

	Tissue					
	Breast	Subcutaneous adipose [a]	Visceral adipose	Liver	Ovary	Cross

BMI	0.204	0.949	0.356	0.756	0.357	0.839
HRT	0.855	0.003 (+)	0.583	0.326	0.809	0.622
OC	0.427	0.028 (-)	0.540	0.918	0.875	0.053
Parity	0.459	0.813	0.950	0.582	0.996	0.417
Parous	0.562	0.727	0.240	0.148	0.995	0.463
Age at the end of first full time pregnancy	0.630	$6.32*10^{-5}$ (-)	0.168	0.229	0.102	0.786
Age at Menarche	0.750	0.225	0.898	0.483	0.799	0.261
Age at menopause	0.781	0.171	0.127	0.091	0.808	0.427

a. Estrogen synthesis pathway genes involved in this tissue include: *STAR*, *HSD17B11*, *HSD3B2*, and *AKR1B15*.
b. Symbol in the parenthesis represent the direction of the interaction effect.

With regards to the estrogen metabolism pathway genes. The summed gene expression level of those genes was observed to have a significant interaction with HRT in breast tissue (P=0.030), and parous (P=0.020) in the subcutaneous adipose tissue. There were no significant interactions observed in the aggregated analysis for estrogen bioavailability pathway genes. (Table 108-109)

Table 108. P-values for the interactions between aggregated predicted gene expression level in estrogen **metabolism** pathway and breast cancer risk factors on breast cancer risk.

Non-genetic factors	Tissue					
	Breast [a]	Subcutaneous adipose [b]	Visceral adipose	Liver	Ovary	Cross
BMI	0.970	0.733	0.299	0.411	0.662	0.617
HRT	0.030 (+)	0.227	0.129	0.835	0.289	0.138
OC	0.413	0.241	0.536	0.384	0.960	0.620
Parity	0.494	0.380	0.406	0.542	0.412	0.927
Parous	0.810	0.020 (+)	0.369	> 0.99	0.070	0.882
Age at the end of first full time pregnancy	0.200	0.484	0.524	0.433	0.053	0.733
Age at Menarche	0.262	0.258	0.548	0.981	0.206	0.520

Age at menopause	0.246	0.438	0.170	0.991	0.460	0.700

a. Estrogen metabolism pathway genes involved in this tissue include: *IDS, ARSF, SULT2B1, COMT, ARSA, CYP3A5, B3GAT2, ARSB, SULT4A1, GSTM1, CYP1A1, CYP1A2, GALNS, B3GAT3, ARSE, CYP3A4, ARSK, UGT2B7, SULT1B1, SGSH, SULT1A2, UGT2B17, SULT1C2, UGT2B11, GSTA1.*

b. Estrogen metabolism pathway genes involved in this tissue include: *B3GAT2, SULT4A1, GNS, CYP1B1, GALNS, ARSG, NQO1, SULT1A1, SULF2, SULT1A4.*

c. Symbol in the parenthesis represent the direction of the interaction effect.

Table 109. P-values for the interactions between aggregated predicted gene expression level in estrogen **bioavailability** pathway and breast cancer risk factors on breast cancer risk.

	Visceral adipose tissue	Breast tissue
BMI	0.797	0.677
HRT	0.206	0.301
OC	0.415	0.118
Parity	0.766	0.217
Parous	0.731	0.951
Age at the end of first full time pregnancy	0.500	0.770
Age at Menarche	0.868	0.383
Age at menopause	0.202	0.415

3.3. Stratified analyses on the interactions of estrogen-related genetically predicted gene expression level with BMI and HRT on breast cancer risk.

Results for the interactions assessed in subgroups were shown from Table 110 to Table 115. 6 and 13 genes potentially interacted with BMI and HRT among post-menopausal women, respectively. The results for the examined interactions for all evaluated non-genetic factors according to menopausal status in ovary tissue model were shown from Table 112 to Table 114. There were no overlaps of the significant interactions observed between pre- and post-menopausal women. Significant interactions ($P < 0.05$) were observed for gene *HSD17B14* with parity, parous and age

at FFTP among menopausal women. Gene *ARSD* showed a significant interaction with both age at menarche and reproductive years among post-menopausal women.

Table 110. Stratified analyses on the interactions between estrogen-related genetically predicted gene expression level and **BMI** on breast cancer risk among post-menopausal women.

Tissue											
Breast		Subcutaneous adipose		Visceral adipose		Cross		Liver		Ovary	
Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value
COMT	0.011	*NQO1*	0.177	*CYP1B1*	0.033	*B3GAT1*	0.015	*CYP1A2*	0.064	*SULT1A1*	0.014
SULT4A1	0.024	*STAR*	0.187	*ARSD*	0.036	*SULF2*	0.124	*UGT1A6*	0.282	*CYP3A4*	0.172
HSD17B1	0.078	*B3GAT2*	0.247	*SULT2B1*	0.177	*CYP17A1*	0.151	*AKR1B15*	0.292	*CYP1B1*	0.287
ARSA	0.098	*GNS*	0.323	*CYP1A1*	0.285	*AKR1B15*	0.209	*HSD17B14*	0.300	*HSD17B14*	0.298
SULT1A2	0.151	*CYP1B1*	0.413	*SULT4A1*	0.288	*SULT1A4*	0.213	*ARSB*	0.325	*CYP19A1*	0.345
ARSB	0.183	*SULT1A4*	0.481	*ARSF*	0.323	*ARSD*	0.255	*UGT2B17*	0.366	*GSTM1*	0.358
ARSF	0.213	*HSD3B2*	0.525	*GSTM1*	0.337	*GALNS*	0.389	*GSTM1*	0.372	*IDS*	0.450
B3GAT2	0.236	*SULT1A1*	0.567	*SULF1*	0.391	*CYP3A4*	0.456	*SULT4A1*	0.395	*B3GAT2*	0.564
UGT2B17	0.237	*SULF2*	0.751	*SULF2*	0.419	*SULT1C2*	0.461	*SULT2B1*	0.440	*UGT2B7*	0.581
GSTM1	0.243	*HSD17B11*	0.771	*ESR2*	0.423	*CYP1B1*	0.495	*SGSH*	0.529	*SULT1C2*	0.672
UGT2B11	0.315	*GALNS*	0.795	*HSD17B1*	0.433	*SULT1A1*	0.501	*B3GAT3*	0.531	*ARSB*	0.706
SGSH	0.323	*ARSG*	0.801	*SULT1A2*	0.468	*SULT1E1*	0.528	*ESR2*	0.545	*ARSG*	0.715
SHBG	0.497	*AKR1B15*	0.975	*UGT2B4*	0.487	*GSTM1*	0.535	*ARSA*	0.547	*ARSA*	0.743
ARSK	0.532	*SULT4A1*	0.988	*COMT*	0.504	*UGT2B7*	0.536	*UGT2B28*	0.591	*ARSD*	0.773
CYP3A5	0.570			*SULT1E1*	0.530	*STAR*	0.537	*GSTA1*	0.699	*ARSE*	0.786
SULT1B1	0.577			*ARSJ*	0.646	*ARSI*	0.557	*SULF2*	0.712	*ARSJ*	0.858
ARSE	0.584			*CYP3A4*	0.699	*STS*	0.603	*STAR*	0.731	*HSD17B1*	0.889
B3GAT3	0.587			*ARSA*	0.717	*SULT1A2*	0.633	*SULT1A1*	0.783	*CYP1A1*	0.920
GSTA1	0.595			*IDS*	0.720	*GSTA1*	0.669	*ARSE*	0.794	*CYP11A1*	0.940
SULT2B1	0.595			*ESR1*	0.880	*HSD17B14*	0.678	*B3GAT1*	0.796	*SULT1A2*	0.987
CYP1A1	0.668			*SULT1B1*	0.881	*SULT4A1*	0.692	*HSD17B11*	0.826		
CYP1A2	0.696			*HSD17B14*	0.914	*UGT2B17*	0.734	*ARSJ*	0.836		
SULT1C2	0.732			*SULT1C2*	0.936	*ARSA*	0.756	*CYP11A1*	0.849		
CYP3A4	0.741					*SULT2B1*	0.814	*COMT*	0.873		
HSD3B2	0.767					*SULF1*	0.883	*UGT1A4*	0.874		
GALNS	0.786					*CYP11A1*	0.990	*HSD3B2*	0.916		
IDS	0.813							*UGT2B4*	0.924		
UGT2B7	0.836							*CYP1B1*	0.936		
								B3GAT2	0.967		

Table 111. Stratified analyses on the interactions between estrogen-related genetically predicted gene expression level and **HRT** on breast cancer risk among post-menopausal women.

Tissue											
Breast		Subcutaneous adipose		Visceral adipose		cross		liver		ovary	
Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value
ARSA	0.008	*STAR*	0.025	*SULF2*	0.029	*SULT1C2*	0.010	*SULT4A1*	0.032	*ARSG*	0.018
GSTM1	0.024	*AKR1B15*	0.071	*ARSA*	0.045	*CYP17A1*	0.027	*UGT2B28*	0.035	*ARSE*	0.026
SULT1A2	0.069	*HSD17B11*	0.081	*IDS*	0.054	*SULT4A1*	0.027	*UGT2B4*	0.039	*SULT1A1*	0.029
HSD3B2	0.082	*SULF2*	0.109	*SULT1A2*	0.066	*AKR1B15*	0.052	*COMT*	0.169	*ARSB*	0.040
GALNS	0.101	*B3GAT2*	0.112	*SULT1B1*	0.084	*STAR*	0.054	*SULT2B1*	0.185	*B3GAT2*	0.089
SGSH	0.107	*NQO1*	0.139	*CYP3A4*	0.130	*SULT2B1*	0.054	*B3GAT3*	0.187	*ARSA*	0.107
COMT	0.116	*SULT1A1*	0.284	*ARSF*	0.162	*SULT1A2*	0.167	*HSD17B11*	0.189	*HSD17B1*	0.151
CYP3A4	0.122	*GALNS*	0.338	*COMT*	0.245	*SULT1A4*	0.196	*UGT1A4*	0.197	*UGT2B7*	0.190
HSD17B1	0.167	*HSD3B2*	0.457	*HSD17B14*	0.263	*SULF1*	0.225	*B3GAT2*	0.202	*CYP11A1*	0.279
CYP3A5	0.173	*SULT4A1*	0.481	*UGT2B4*	0.303	*GSTA1*	0.260	*ARSE*	0.202	*ARSD*	0.322
B3GAT2	0.204	*ARSG*	0.683	*GSTM1*	0.381	*CYP1B1*	0.322	*STAR*	0.223	*CYP1A1*	0.426
SULT1B1	0.205	*SULT1A4*	0.771	*CYP1A1*	0.393	*SULF2*	0.453	*ESR2*	0.232	*SULT1C2*	0.490
ARSE	0.215	*CYP1B1*	0.828	*ESR2*	0.399	*SULT1A1*	0.465	*AKR1B15*	0.239	*ARSJ*	0.507
SHBG	0.222	*GNS*	0.939	*SULT4A1*	0.418	*CYP11A1*	0.502	*SULT1A1*	0.248	*SULT1A2*	0.674
CYP1A2	0.274			*SULT1C2*	0.427	*GALNS*	0.582	*CYP11A1*	0.386	*GSTM1*	0.678
CYP1A1	0.280			*SULF1*	0.569	*CYP3A4*	0.652	*SULF2*	0.425	*CYP1B1*	0.751
SULT4A1	0.310			*HSD17B1*	0.587	*UGT2B17*	0.682	*GSTM1*	0.482	*HSD17B14*	0.873
SULT2B1	0.322			*SULT1E1*	0.668	*GSTM1*	0.723	*ARSJ*	0.482	*IDS*	0.925
UGT2B7	0.322			*SULT2B1*	0.706	*STS*	0.818	*SGSH*	0.582	*CYP3A4*	0.983
GSTA1	0.382			*ARSD*	0.732	*ARSA*	0.842	*CYP1A2*	0.615	*CYP19A1*	0.986
IDS	0.386			*ARSJ*	0.770	*UGT2B7*	0.940	*ARSA*	0.623		
B3GAT3	0.431			*ESR1*	0.867	*SULT1E1*	0.943	*HSD17B14*	0.626		
UGT2B11	0.451			*CYP1B1*	0.924	*ARSI*	0.953	*GSTA1*	0.646		
UGT2B17	0.522					*B3GAT1*	0.966	*UGT1A6*	0.778		
ARSK	0.532					*ARSD*	0.966	*HSD3B2*	0.796		
SULT1C2	0.657					*HSD17B14*	0.997	*CYP1B1*	0.835		
ARSB	0.792							*ARSB*	0.853		
ARSF	0.893							*UGT2B17*	0.878		
								B3GAT1	0.989		

3.4. Interactions between estrogen-related genetically predicted gene expression level from prediction model of ovary tissue and non-genetic variables on breast cancer risk by menopausal status.

Table 112. Interactions of estrogen-related genetically predicted gene expression level from prediction model of **ovary** tissue with **BMI**, **HRT,** and **OC** use on breast cancer risk by menopausal status.

BMI				HRT				OC			
Pre-menopausal		Post-menopausal		Pre-menopausal		Post-menopausal		Pre-menopausal		Post-menopausal	
Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value
SULT1C2	0.088	*SULT1A1*	0.014	*IDS*	0.075	*ARSG*	0.018	*ARSJ*	0.023	*ARSB*	0.017
GSTM1	0.105	*CYP3A4*	0.172	*CYP1A1*	0.137	*ARSE*	0.026	*CYP1B1*	0.300	*IDS*	0.052
CYP11A1	0.230	*CYP1B1*	0.287	*ARSE*	0.246	*SULT1A1*	0.029	*GSTM1*	0.317	*HSD17B1*	0.070
IDS	0.253	*HSD17B14*	0.298	*ARSA*	0.377	*ARSB*	0.040	*CYP1A1*	0.342	*ARSJ*	0.121
HSD17B14	0.260	*CYP19A1*	0.345	*HSD17B1*	0.414	*B3GAT2*	0.089	*ARSG*	0.368	*CYP11A1*	0.209
SULT1A2	0.297	*GSTM1*	0.358	*SULT1A2*	0.443	*ARSA*	0.107	*CYP11A1*	0.402	*CYP1A1*	0.250
ARSA	0.424	*IDS*	0.450	*UGT2B7*	0.512	*HSD17B1*	0.151	*CYP3A4*	0.416	*SULT1C2*	0.271
SULT1A1	0.427	*B3GAT2*	0.564	*HSD17B14*	0.529	*UGT2B7*	0.190	*SULT1C2*	0.444	*CYP3A4*	0.348
ARSB	0.484	*UGT2B7*	0.581	*GSTM1*	0.574	*CYP11A1*	0.279	*CYP19A1*	0.493	*ARSD*	0.490
UGT2B7	0.594	*SULT1C2*	0.672	*B3GAT2*	0.648	*ARSD*	0.322	*IDS*	0.519	*ARSA*	0.532
HSD17B1	0.674	*ARSB*	0.706	*CYP1B1*	0.667	*CYP1A1*	0.426	*HSD17B1*	0.619	*SULT1A2*	0.608
B3GAT2	0.706	*ARSG*	0.715	*CYP19A1*	0.728	*SULT1C2*	0.490	*B3GAT2*	0.628	*CYP1B1*	0.674
ARSJ	0.728	*ARSA*	0.743	*CYP11A1*	0.739	*ARSJ*	0.507	*ARSD*	0.641	*HSD17B14*	0.715
ARSE	0.767	*ARSD*	0.773	*ARSJ*	0.754	*SULT1A2*	0.674	*HSD17B14*	0.654	*UGT2B7*	0.762
CYP19A1	0.788	*ARSE*	0.786	*ARSB*	0.784	*GSTM1*	0.678	*UGT2B7*	0.659	*B3GAT2*	0.802
CYP1A1	0.949	*ARSJ*	0.858	*ARSD*	0.800	*CYP1B1*	0.751	*ARSB*	0.805	*SULT1A1*	0.816
ARSD	0.978	*HSD17B1*	0.889	*CYP3A4*	0.932	*HSD17B14*	0.873	*ARSE*	0.867	*ARSG*	0.829
ARSG	0.979	*CYP1A1*	0.920	*SULT1A1*	0.961	*IDS*	0.925	*ARSA*	0.923	*GSTM1*	0.831
CYP3A4	0.984	*CYP11A1*	0.940	*ARSG*	0.991	*CYP3A4*	0.983	*SULT1A2*	0.996	*CYP19A1*	0.914
CYP1B1	0.996	*SULT1A2*	0.987	*SULT1C2*	0.991	*CYP19A1*	0.986	SULT1A1	0.997	*ARSE*	0.919

Table 113. Interactions of estrogen-related genetically predicted gene expression level from prediction model of **ovary** tissue with **parous, parity** and **age at first full time pregnancy** on breast cancer risk by menopausal status.

Parous				Parity				Age at first full time pregnancy			
Pre-menopausal		Post-menopausal		Pre-menopausal		Post-menopausal		Pre-menopausal		Post-menopausal	
Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value
ARSE	0.015	*HSD17B14*	0.034	*B3GAT2*	0.021	*HSD17B14*	0.040	*UGT2B7*	0.010	*HSD17B14*	0.025
ARSG	0.021	*IDS*	0.044	*GSTM1*	0.044	*CYP1B1*	0.054	*SULT1A2*	0.099	*CYP11A1*	0.064
HSD17B1	0.027	*B3GAT2*	0.064	*ARSG*	0.075	*GSTM1*	0.161	*CYP1B1*	0.130	*IDS*	0.072
IDS	0.065	*CYP1B1*	0.130	*ARSE*	0.086	*B3GAT2*	0.174	*SULT1C2*	0.134	*CYP19A1*	0.080
GSTM1	0.098	*ARSG*	0.191	*ARSJ*	0.149	*ARSE*	0.242	*ARSE*	0.171	*ARSA*	0.095
UGT2B7	0.206	*ARSD*	0.241	*IDS*	0.200	*IDS*	0.250	*B3GAT2*	0.192	*GSTM1*	0.187
ARSD	0.264	*SULT1A2*	0.258	*CYP3A4*	0.216	*CYP1A1*	0.317	*CYP1A1*	0.206	*SULT1A2*	0.229
B3GAT2	0.296	*CYP11A1*	0.286	*SULT1A2*	0.257	*CYP19A1*	0.366	*SULT1A1*	0.246	*ARSE*	0.249
ARSJ	0.351	*HSD17B1*	0.302	*ARSB*	0.313	*SULT1A2*	0.603	*CYP19A1*	0.290	*ARSG*	0.252
SULT1C2	0.371	*ARSB*	0.353	*CYP1B1*	0.381	*ARSG*	0.706	*ARSG*	0.426	*HSD17B1*	0.354
CYP1B1	0.374	*SULT1A1*	0.557	*SULT1C2*	0.388	*ARSB*	0.718	*ARSB*	0.455	*CYP3A4*	0.366
CYP1A1	0.460	*CYP3A4*	0.584	*CYP1A1*	0.410	*SULT1C2*	0.724	*ARSA*	0.520	*CYP1A1*	0.381
SULT1A1	0.505	*ARSJ*	0.616	*HSD17B14*	0.496	*ARSA*	0.768	*CYP3A4*	0.563	*UGT2B7*	0.466
CYP11A1	0.522	*UGT2B7*	0.617	*HSD17B1*	0.551	*CYP3A4*	0.797	*HSD17B14*	0.669	*CYP1B1*	0.529
CYP19A1	0.656	*CYP19A1*	0.663	*CYP19A1*	0.577	*HSD17B1*	0.842	*GSTM1*	0.703	*ARSB*	0.548
ARSB	0.697	*CYP1A1*	0.763	*CYP11A1*	0.632	*ARSD*	0.876	*CYP11A1*	0.819	*B3GAT2*	0.630
CYP3A4	0.723	*GSTM1*	0.788	*UGT2B7*	0.645	*CYP11A1*	0.909	*ARSJ*	0.865	*ARSD*	0.673
SULT1A2	0.749	*ARSE*	0.840	*SULT1A1*	0.733	*UGT2B7*	0.914	*ARSD*	0.882	*SULT1A1*	0.790
HSD17B14	0.759	*SULT1C2*	0.860	*ARSA*	0.751	*SULT1A1*	0.958	*IDS*	0.904	*SULT1C2*	0.943
ARSA	0.812	*ARSA*	0.980	*ARSD*	0.843	*ARSJ*	0.997	*HSD17B1*	0.904	*ARSJ*	0.945
ARSE	0.015	*HSD17B14*	0.034	*B3GAT2*	0.021	*HSD17B14*	0.040	*UGT2B7*	0.010	*HSD17B14*	0.025
ARSG	0.021	*IDS*	0.044	*GSTM1*	0.044	*CYP1B1*	0.054	*SULT1A2*	0.099	*CYP11A1*	0.064
HSD17B1	0.027	*B3GAT2*	0.064	*ARSG*	0.075	*GSTM1*	0.161	*CYP1B1*	0.130	*IDS*	0.072
IDS	0.065	*CYP1B1*	0.130	*ARSE*	0.086	*B3GAT2*	0.174	*SULT1C2*	0.134	*CYP19A1*	0.080
GSTM1	0.098	*ARSG*	0.191	*ARSJ*	0.149	*ARSE*	0.242	*ARSE*	0.171	*ARSA*	0.095
UGT2B7	0.206	*ARSD*	0.241	*IDS*	0.200	*IDS*	0.250	*B3GAT2*	0.192	*GSTM1*	0.187
ARSD	0.264	*SULT1A2*	0.258	*CYP3A4*	0.216	*CYP1A1*	0.317	*CYP1A1*	0.206	*SULT1A2*	0.229
B3GAT2	0.296	*CYP11A1*	0.286	*SULT1A2*	0.257	*CYP19A1*	0.366	*SULT1A1*	0.246	*ARSE*	0.249

Table 114. Interactions of estrogen-related genetically predicted gene expression level from prediction model of **ovary** tissue with **age at menarche, age at menopause, and reproductive years** on breast cancer risk by menopausal status.

Age at menarche				Age at menopause		Reproductive years	
Pre-menopausal		Post-menopausal		Post-menopausal		Post-menopausal	
Gene name	P-value	Gene name	P-value	Gene name	P-value	Gene name	P-value
CYP3A4	0.012	*ARSD*	0.030	*ARSD*	0.056	*ARSD*	0.047
ARSB	0.013	*SULT1A2*	0.087	*ARSB*	0.090	*SULT1A2*	0.157
CYP11A1	0.020	*ARSA*	0.136	*SULT1A2*	0.196	*ARSA*	0.175
HSD17B14	0.067	*SULT1C2*	0.156	*ARSA*	0.224	*ARSB*	0.282
CYP1A1	0.068	*B3GAT2*	0.177	*CYP1A1*	0.281	*CYP1A1*	0.379
GSTM1	0.118	*HSD17B14*	0.238	*HSD17B1*	0.444	*CYP11A1*	0.388
ARSA	0.140	*ARSG*	0.343	*B3GAT2*	0.471	*ARSJ*	0.390
ARSE	0.197	*ARSE*	0.457	*ARSJ*	0.473	*HSD17B1*	0.404
CYP1B1	0.239	*ARSB*	0.510	*SULT1A1*	0.474	*CYP3A4*	0.492
ARSG	0.250	*CYP1B1*	0.534	*CYP1B1*	0.477	*SULT1A1*	0.498
B3GAT2	0.257	*ARSJ*	0.615	*ARSG*	0.483	*UGT2B7*	0.516
HSD17B1	0.316	*SULT1A1*	0.617	*UGT2B7*	0.514	*HSD17B14*	0.528
CYP19A1	0.351	*CYP11A1*	0.652	*CYP11A1*	0.580	*CYP1B1*	0.570
SULT1A2	0.397	*CYP3A4*	0.668	*CYP19A1*	0.793	*ARSG*	0.575
SULT1A1	0.420	*UGT2B7*	0.698	*CYP3A4*	0.831	*B3GAT2*	0.632
IDS	0.460	*IDS*	0.704	*HSD17B14*	0.882	*SULT1C2*	0.757
ARSD	0.658	*GSTM1*	0.818	*SULT1C2*	0.884	*CYP19A1*	0.820
UGT2B7	0.808	*CYP19A1*	0.936	*IDS*	0.940	*ARSE*	0.828
SULT1C2	0.853	*HSD17B1*	0.949	*GSTM1*	0.947	*IDS*	0.846
ARSJ	0.954	*CYP1A1*	0.957	*ARSE*	0.970	*GSTM1*	0.873
CYP3A4	0.012	*ARSD*	0.030	*ARSD*	0.056	*ARSD*	0.047
ARSB	0.013	*SULT1A2*	0.087	*ARSB*	0.090	*SULT1A2*	0.157
CYP11A1	0.020	*ARSA*	0.136	*SULT1A2*	0.196	*ARSA*	0.175
HSD17B14	0.067	*SULT1C2*	0.156	*ARSA*	0.224	*ARSB*	0.282
CYP1A1	0.068	*B3GAT2*	0.177	*CYP1A1*	0.281	*CYP1A1*	0.379
GSTM1	0.118	*HSD17B14*	0.238	*HSD17B1*	0.444	*CYP11A1*	0.388
ARSA	0.140	*ARSG*	0.343	*B3GAT2*	0.471	*ARSJ*	0.390
ARSE	0.197	*ARSE*	0.457	*ARSJ*	0.473	*HSD17B1*	0.404

Table 115. Summary for the genes that have a significant (P < 0.05) interaction with BMI or HRT on breast cancer risk among post-menopausal women.

	Tissue					
	Breast	Subcutaneous adipose	Visceral adipose	Cross	Liver	Ovary
BMI	*COMT, SULT4A1*		*CYP1B1, ARSD*	*B3GAT1*		*SULT1A1*
HRT	*ARSA*, GSTM1*	*STAR*	*SULF2, ARSA*	*SULT1C2, CYP17A1, SULT4A1*	*SULT4A1, UGT2B28, UGT2B4*	*ARSG, ARSE, SULT1A1, ARSB*

*P<0.01

Table 116. Summary for the genes that have a significant (P < 0.05) interaction from **ovary** prediction model on breast cancer risk according to menopausal status.

	Pre-menopausal	Post-menopausal
BMI		*SULT1A1*
HRT		*ARSG*
OC	*ARSJ*	*ARSB*
Parity	*ARSE, ARSG, HSD17B1*	*HSD17B14, IDS*
Parous	*B3GAT2, GSTM1*	*HSD17B14*
Age at first full time pregnancy	*UGT2B7*	*HSD17B14*
Age at Menarche	*CYP3A4, ARSB, CYP11A1*	*ARSD*
Age at menopause		
Reproductive years		*ARSD*

3.5. Associations between predicted gene expression and breast cancer risk stratified by estrogen-related non-genetic factors.

Table 117. Associations between HRT and breast cancer risk according to quartiles of gene expression.

	Subcutaneous adipose		Cross	
Predicted gene expression	HRT use	P for interaction	HRT use	P for interaction
AKR1B15		0.003		0.004
Q1	1.07 (1.01, 1.14)		1.01 (0.95, 1.07)	
Q2	1.06 (1.00, 1.13)		1.13 (1.07, 1.20)	
Q3	1.11 (1.05, 1.18)		1.10 (1.04, 1.17)	
Q4	1.18 (1.11, 1.25)		1.17 (1.11, 1.25)	

The ORs for the association between estrogen related non-genetic factors and breast cancer risk according to predicted gene expression levels categorized by quartile were shown for the interactions terms that have a P value less than 0.005. HRT use showed a higher association with breast cancer risk among the group of people with a higher predicted expression level of the estrogen synthesis pathway gene, *AKR1B15*. In subcutaneous adipose tissue, the OR was 1.18 (95% CI: 1.11-1.25) in the highest quartile of *AKR1B15* expression level compared to 1.07 (95% CI: 1.01-1.14) in the lowest quartile. In the cross tissue model, HRT was also more significantly associated with breast cancer risk in the highest quartile of *AKR1B15* (OR=1.17, 95% CI: 1.11-1.25), compared to the lowest one (OR=1.01, 95% CI: 0.95-1.07). (Table 117)

The association between parous and breast cancer risk was observed to be more significant among subjects with a lower expression level predicted using the cross tissue model of the estrogen metabolism gene, *B3GAT1*. The OR was 0.82 (95% CI: 0.77-0.88) among subjects with the lowest expression level compared to 0.94 (95% CI: 0.88-1.00) among participants with the highest expression level. However, similar trend for this gene was not observed for other tissues (Table 118)

Table 118. Associations between parous and breast cancer risk according to quartiles of gene expression.

	Liver tissue		Cross	
Predicted gene expression	Parous	P for interaction	Parous	P for interaction
B3GAT1		0.095		0.0004
Q1	0.94 (0.88, 1.00)		0.82 (0.77, 0.88)	

Q2	0.86 (0.81, 0.92)		0.86 (0.81, 0.92)	
Q3	0.84 (0.78, 0.89)		0.91 (0.85, 0.97)	
Q4	0.89 (0.83, 0.95)		0.94 (0.88, 1.00)	

The association between parous and breast cancer risk also varied across the predicted gene expression level of the estrogen metabolism pathway gene, *IDS*, in ovary tissue model. Parous was observed to be more significantly associated with breast cancer risk among participants with a higher predicted expression level. The OR for the group of people with the highest expression level was 0.83 (95% CI: 0.78-0.89), compared to those with the lowest (OR=0.94, 95% CI: 0.88-1.01). However, similar trend for this gene was not observed for other tissues. (Table 119)

Table 119. Associations between parous and breast cancer risk according to quartiles of gene expression.

	Breast tissue		Ovary tissue		Visceral adipose	
Predicted gene expression	Parous	P for interaction	Parous	P for interaction	Parous	P for interaction
IDS		0.630		0.001		0.405
Q1	0.90 (0.84, 0.96)		0.94 (0.88, 1.01)		0.86 (0.80, 0.92)	
Q2	0.80 (0.75, 0.86)		0.86 (0.80, 0.92)		0.88 (0.83, 0.94)	
Q3	0.92 (0.86, 0.98)		0.90 (0.84, 0.96)		0.90 (0.84, 0.96)	
Q4	0.91 (0.85, 0.97)		0.83 (0.78, 0.89)		0.89 (0.83, 0.95)	

Table 120 showed the results for the stratified analysis for the association between age at FFTP and breast cancer risk by the predicted expression level the estrogen synthesis pathway gene, *HSD17B11*. Significant interaction was observed in the subcutaneous

adipose tissue model. Among participants with the lowest predicted expression level, age at FFTP was association with an increased risk of breast cancer, the OR was 1.32 (95% CI: 1.13-1.53) when comparing women with an age at FFTP of at least 35 years old to those with an age at FFTP of younger than 20 years old.

Table 120. Associations between age at the end of first full time pregnancy and breast cancer risk according to quartiles of gene expression.

Predicted gene expression	Liver tissue					
	age at the end of first full time pregnancy					
	< 20	20-24	25-29	30-34	≥ 35	P for interaction
HSD17B11						0.573
Q1	1.0	0.90 (0.82, 0.99)	0.97 (0.88, 1.06)	1.11 (1.00, 1.24)	1.15 (0.99, 1.33)	
Q2	1.0	0.95 (0.87, 1.04)	1.02 (0.93, 1.12)	1.14 (1.02, 1.27)	1.29 (1.11, 1.50)	
Q3	1.0	0.86 (0.79, 0.94)	0.93 (0.85, 1.02)	0.96 (0.87, 1.08)	1.08 (0.93, 1.26)	
Q4	1.0	0.91 (0.83, 0.99)	0.94 (0.86, 1.03)	1.07 (0.96, 1.20)	1.14 (0.98, 1.32)	
Predicted gene expression	Subcutaneous adipose tissue					
	age at the end of first full time pregnancy					
	< 20	20-24	25-29	30-34	≥ 35	P for interaction
HSD17B11						0.001
Q1	1.0	0.95 (0.87, 1.04)	1.04 (0.95, 1.14)	1.15 (1.03, 1.28)	1.32 (1.13, 1.53)	
Q2	1.0	0.98 (0.90, 1.07)	1.03 (0.94, 1.13)	1.24 (1.11, 1.39)	1.24 (1.07, 1.44)	
Q3	1.0	0.85 (0.77, 0.93)	0.91 (0.83, 1.00)	0.96 (0.86, 1.07)	1.06 (0.91, 1.23)	
Q4	1.0	0.85 (0.77, 0.93)	0.88 (0.80, 0.97)	0.96 (0.86, 1.08)	1.06 (0.92, 1.23)	

4. Results for Asian population

4.1. Non-genetic variables.

The results for the Asian population of Aim 3 were shown from Table 121 to Table 132. Table 121 showed the characteristics of our study population of Asians. Similar to the distribution of the European descendants, compared to controls, breast cancer cases were more likely to be older, nulliparous, have higher BMI, and

menopause at an older age. They were also less likely to have given birth to more than one child or give birth to their first child at an earlier age. No significant associations were observed between soy food intake and breast cancer risk.

Table 121. Distribution of the known breast cancer risk factors by breast cancer status among Asians, *ABCC*.

	Total No. of cases	Total No. of controls	Case	Control	P-value
Age at interview	11074	9776	47.5 (42.0-53.0)	47.0 (42.0-42.0)	<.0001
BMI (kg/m2)	10466	11954	23.1 (21.0-25.6)	23.0 (20.8-25.3)	<.0001
BMI (categorized)	10466	11954			<.0001
18.5-24.9			64.6	66.5	0.0002
< 18.5			5.3	5.5	
25-29.9			24.7	23.7	
≥ 30			5.5	4.4	
Hormone replacement therapy use (% of yes)	8296	9136	7.2	5.6	<.0001
Oral conceptive use (% of yes)	8586	10041	25.7	18.2	<.0001
Post-menopausal (% of yes)	10759	12145	51.0	47.2	<.0001
Age at Menarche (years)	11906	12978	14.0 (13.0-15.0)	14.0 (13.0-15.0)	<.0001
Age at Menarche (categorized)	11906	12978			<.0001
< 12			6.5	6.4	
12			16.3	16.1	
13			22.0	20.8	
14			20.4	19.5	
15			16.3	16.0	
≥ 16			18.5	21.2	
Age at Menopause (years)	4751	5326	50.0 (46.7-52.0)	50.0 (47.0-52.0)	0.132
Age at Menopause (categorized)	4751	5326			0.006
< 40			5.2	4.1	
40-44			10.4	9.5	
45-48			23.8	25.4	
49-54			52.1	53.2	
>=55			8.5	7.8	
Parous (% of no)	12269	12775	85.8	88.3	<.0001
Parity	12251	12708	2 (1-2)	2 (1-3)	<.0001

Parity (categorized)	12251	12708			<.0001
0			14.2	11.7	
1			31.7	27.1	
2			31.7	35.3	
3			14.4	16.8	
4			5.0	5.7	
>4			3.1	3.4	
Age at end of first full-term pregnancy (years)	9154	9456	26.9 (24.0-29.0)	26.0 (23.5-28.5)	<.0001
Age at end of first full-term pregnancy (categorized)	9154	9456			<.0001
<20			5.1	4.9	
20-24			25.5	30.1	
25-29			47.2	47.9	
30-34			17.5	13.8	
>=35			4.8	3.4	
Soy food intake (g/day)	4910	6680	104.1 (53.4-200.8)	104.7 (54.1-204.2)	0.454

a. Numbers were presented in median (interquartile) or percentage for continuous and categorical variables, respectively.
b. P-values were calculated from ANOVA test or Chi-square test for continuous and categorical variables, respectively.

Table 122 displayed the age-adjusted odds ratios (ORs) between the evaluated breast cancer risk factors and breast cancer risk in our Asian study population. Obesity was observed to be associated with an increased risk of breast cancer among post-menopausal women (OR=1.17, 95% CI: 1.03-1.33). Oral contraceptive use was associated with an elevated risk of breast cancer (OR=1.71, 95% CI:1.57-1.87). Parous (OR=0.85, 95% CI: 0.78-0.92) was associated with a lower likelihood of having breast cancer. Higher age at FFTP was associated with an increased risk of breast cancer. Higher soy food intake was associated with an increased risk of breast cancer among premenopausal women (OR=2.22, 95% CI:1.34-3.66, comparing the highest versus lowest intake).

Table 122. Age-adjusted odds ratios (95% confidence intervals) for the associations of body mass index, menstrual and reproductive factors and soy food intake with breast cancer risk.

	All	Post-menopausal	Pre-menopausal
No. of control/case	15091/14823	5279/5488	6416/5271
BMI			
18.5-24.9	1.0	1.0	1.0
<18.5	1.05 (0.91, 1.21)	0.98 (0.75, 1.28)	1.01 (0.85 (1.22)
25-29.9	1.08 (1.00, 1.17)	1.17 (1.03, 1.33)	1.02 (0.92, 1.14)
≥ 30	1.09 (0.94, 1.27)	1.26 (0.98, 1.61)	0.97 (0.79, 1.21)
Hormone replacement therapy use	0.96 (0.81, 1.14)	0.87 (0.71, 1.06)	1.14 (0.77, 1.69)
Oral conceptive use	1.71 (1.57, 1.87)	1.86 (1.61, 2.16)	1.56 (1.39, 1.75)
Age at Menarche	1.02 (1.00, 1.03)	1.02 (0.99, 1.05)	1.00 (0.97, 1.02)
Age at Menopause			
<40		1.0	
40-44		0.77 (0.57, 1.04)	
45-48		0.59 (0.45, 0.78)	
49-54		0.51 (0.39, 0.67)	
>=55		0.51 (0.35, 0.75)	
Parity			
Age at end of first full-term pregnancy			
<20	1.0	1.0	1.0
20-24	0.77 (0.64, 0.93)	0.66 (0.50, 0.86)	0.78 (0.60, 1.01)
25-29	1.05 (0.88, 1.25)	0.91 (0.70, 1.18)	1.05 (0.82, 1.35)
30-34	1.29 (1.06, 1.56)	1.21 (0.90, 1.63)	1.24 (0.95, 1.61)
>=35	1.26 (1.00, 1.61)	1.13 (0.76, 1.68)	1.24 (0.90, 1.72)
Soy food intake			
1st quartile	1.0	1.0	1.0
2nd quartile	1.16 (0.80, 1.69)	1.30 (0.56, 2.97)	1.18 (0.77, 1.81)
3rd quartile	2.16 (1.40, 3.32)	2.75 (1.00, 7.53)	2.07 (1.28, 3.35)
4th quartile	2.19 (1.40, 3.41)	2.02 (0.74, 5.48)	2.22 (1.34, 3.66)
Parous	0.85 (0.78, 0.92)	0.69 (0.57, 0.82)	1.01 (0.89, 1.13)

4.2. Interactions

The results of the assessed interaction terms between estrogen-related pathway genes and estrogen-related non-genetic factors across the models built in breast tissue were presented from Table 123 to Table 131. The summarized results for those interactions were included from Table 132. There were in total of 10 genes that have a P value for interaction less than 0.05 with the 6 non-genetic factors. Among them, the P

values for the interactions of OC use with the estrogen synthesis pathway gene *CYP19A1* and the estrogen metabolism pathway gene, *GSTM1,* were less than 0.01.

Table 123. Interaction between BMI and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
UGT2B28	-0.044	0.017	0.011
HSD17B1	-0.051	0.020	0.013
UGT2B15	-0.020	0.012	0.110
ARSK	0.032	0.026	0.209
SULT1A4	0.024	0.020	0.244
CYP3A4	-0.333	0.299	0.266
SULT1C2	-0.010	0.009	0.278
SULT4A1	-0.006	0.006	0.295
SULT1A2	0.016	0.017	0.328
ARSA	0.042	0.043	0.333
GSTM1	-0.006	0.008	0.439
B3GAT2	-0.009	0.015	0.535
GALNS	0.013	0.023	0.576
CYP17A1	-0.036	0.067	0.585
SULF1	-0.017	0.035	0.632
CYP19A1	-0.067	0.172	0.695
ESR2	-0.014	0.037	0.708
AKR1B15	0.108	0.298	0.716
GSTA1	-0.006	0.018	0.731

Table 124. Interaction between HRT use and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
CYP17A1	-2.045	1.145	0.074
SULF1	1.043	0.614	0.089
ARSK	-0.693	0.439	0.115
HSD17B1	-0.481	0.344	0.161
GSTM1	-0.175	0.130	0.179
SULT4A1	0.108	0.099	0.277
SULT1A2	0.290	0.294	0.325
UGT2B28	-0.283	0.288	0.326
SULT1A4	-0.325	0.356	0.362
CYP3A4	-3.947	4.827	0.413

ARSA	0.568	0.764	0.458
GALNS	0.291	0.394	0.461
SULT1C2	-0.089	0.161	0.578
CYP19A1	0.946	2.972	0.750
B3GAT2	0.073	0.249	0.769
GSTA1	-0.049	0.319	0.879
UGT2B15	-0.032	0.209	0.879
AKR1B15	-0.514	4.783	0.914
ESR2	-0.058	0.613	0.924

Table 125. Interaction between OC use and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
CYP19A1	-6.112	1.700	0.0003
GSTM1	-0.212	0.079	0.007
SULT1A2	-0.410	0.170	0.016
HSD17B1	0.493	0.209	0.018
GSTA1	0.370	0.180	0.040
B3GAT2	0.262	0.147	0.074
CYP3A4	3.784	3.191	0.236
UGT2B15	0.140	0.125	0.262
UGT2B28	0.200	0.179	0.264
ARSK	0.196	0.257	0.445
ARSA	0.326	0.476	0.494
CYP17A1	0.400	0.656	0.542
SULF1	-0.202	0.353	0.567
GALNS	0.116	0.236	0.625
SULT1A4	0.050	0.208	0.812
ESR2	-0.087	0.369	0.814
AKR1B15	-0.393	2.939	0.894
SULT4A1	0.004	0.060	0.952
SULT1C2	0.001	0.096	0.990

Table 126. Interaction between parity and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
SULT1C2	-0.064	0.025	0.010
AKR1B15	1.840	0.786	0.019
GSTA1	-0.111	0.048	0.020
HSD17B1	-0.116	0.054	0.030
CYP3A4	-1.391	0.811	0.086

ESR2	-0.147	0.100	0.141
UGT2B15	-0.037	0.033	0.264
B3GAT2	-0.042	0.041	0.305
UGT2B28	-0.030	0.046	0.515
SULF1	-0.063	0.097	0.516
SULT4A1	0.008	0.016	0.626
CYP19A1	0.188	0.443	0.671
GALNS	0.025	0.060	0.678
ARSK	0.020	0.070	0.771
SULT1A2	-0.012	0.046	0.794
GSTM1	-0.005	0.021	0.827
CYP17A1	-0.031	0.182	0.866
SULT1A4	-0.009	0.057	0.870
ARSA	0.003	0.120	0.982

Table 127. Interaction between parous and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
SULT1A2	0.283	0.169	0.094
CYP3A4	-4.476	2.932	0.127
SULF1	-0.538	0.361	0.136
UGT2B15	-0.133	0.127	0.295
GSTM1	0.071	0.079	0.364
SULT4A1	0.053	0.061	0.384
HSD17B1	-0.170	0.203	0.403
B3GAT2	0.098	0.152	0.518
SULT1C2	-0.060	0.096	0.531
GSTA1	-0.105	0.184	0.570
CYP17A1	-0.311	0.671	0.643
SULT1A4	0.061	0.205	0.767
AKR1B15	-0.704	3.008	0.815
ESR2	0.082	0.372	0.826
ARSK	0.046	0.261	0.860
ARSA	-0.059	0.450	0.895
UGT2B28	0.021	0.168	0.901
CYP19A1	0.192	1.735	0.912
GALNS	-0.001	0.219	0.996

Table 128. Interaction between age at first full time pregnancy and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value

AKR1B15	-0.684	0.276	0.013
UGT2B15	0.021	0.012	0.076
GSTA1	0.029	0.017	0.100
SULT4A1	-0.009	0.006	0.127
CYP3A4	0.418	0.289	0.149
HSD17B1	-0.023	0.019	0.240
SULT1A4	0.022	0.020	0.266
B3GAT2	-0.014	0.014	0.343
GALNS	0.020	0.022	0.359
ARSK	-0.020	0.024	0.407
ARSA	0.032	0.045	0.471
CYP19A1	0.114	0.165	0.489
SULF1	-0.023	0.034	0.503
SULT1A2	0.010	0.016	0.534
SULT1C2	0.005	0.009	0.558
GSTM1	-0.004	0.007	0.635
UGT2B28	-0.007	0.017	0.697
CYP17A1	0.022	0.063	0.724
ESR2	0.012	0.035	0.738

Table 129. Interaction between age at menarche and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
SULT1A2	0.080	0.031	0.010
UGT2B15	-0.049	0.022	0.026
SULF1	0.131	0.066	0.047
SULT1C2	-0.031	0.017	0.078
CYP19A1	0.496	0.309	0.109
HSD17B1	-0.058	0.038	0.121
UGT2B28	0.043	0.032	0.172
SULT1A4	-0.040	0.038	0.298
GSTM1	-0.013	0.014	0.353
CYP17A1	-0.099	0.121	0.413
B3GAT2	0.021	0.027	0.439
SULT4A1	0.008	0.011	0.475
ARSA	-0.057	0.082	0.488
ARSK	0.027	0.047	0.567
GSTA1	0.016	0.033	0.628
ESR2	0.021	0.067	0.757
CYP3A4	-0.112	0.562	0.842
GALNS	0.004	0.041	0.913
AKR1B15	0.025	0.546	0.964

Table 130. Interaction between age at menopause and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
ARSK	0.035	0.028	0.217
ESR2	-0.045	0.041	0.280
SULT1C2	0.011	0.011	0.297
UGT2B28	-0.019	0.019	0.315
SULT4A1	0.007	0.007	0.320
SULT1A4	0.016	0.022	0.456
CYP19A1	0.138	0.194	0.476
GSTA1	0.014	0.020	0.504
HSD17B1	-0.015	0.023	0.529
SULF1	0.018	0.039	0.648
GALNS	-0.011	0.025	0.651
ARSA	-0.021	0.051	0.683
CYP3A4	0.133	0.333	0.689
B3GAT2	-0.005	0.016	0.765
CYP17A1	-0.020	0.074	0.790
GSTM1	0.001	0.009	0.900
UGT2B15	0.001	0.014	0.946
SULT1A2	0.000	0.019	0.996
AKR1B15	-0.001	0.336	0.997

Table 131. Interaction between soy food intake and predicted gene expression level on breast cancer risk in breast tissue among Asian population.

Gene name	Effect	SE	P-value
SULF1	-0.004	0.002	0.043
SULT4A1	0.001	0.000	0.090
ESR2	-0.004	0.002	0.096
CYP3A4	0.113	0.068	0.098
CYP17A1	-0.006	0.004	0.111
GSTA1	-0.002	0.001	0.121
GSTM1	-0.001	0.001	0.264
AKR1B15	-0.020	0.018	0.265
SULT1C2	0.001	0.001	0.310
ARSK	-0.001	0.002	0.370
GALNS	0.002	0.002	0.385
HSD17B1	0.001	0.002	0.472
ARSA	0.004	0.006	0.486
SULT1A2	-0.001	0.001	0.492
B3GAT2	0.000	0.001	0.597
SULT1A4	0.001	0.002	0.638
CYP19A1	0.004	0.012	0.722
UGT2B28	-0.001	0.002	0.780

UGT2B15	0.000	0.001	0.829

Table 132. Summary for the genes that have a significant (P < 0.05) interaction on breast cancer risk among Asian population.

	BMI	OC use	parity	Age at first full term pregnancy	Age at menarche	Soy food intake
Genes (P-interaction<0.05)	*UGT2B28* *HSD17B1*	*CYP19A1** *GSTM1** *SULT1A2* *HSD17B1* *GSTA1*	*SULT1C2* *AKR1B15* *GSTA1* *HSD17B1*	*AKR1B15*	*SULT1A2* *UGT2B15* *SULF1*	*SULF1*

* P < 0.01

3. Discussion

In this large-scale gene-environment interaction study, we found that based on the subcutaneous adipose tissue, the predicted gene expression level in estrogen synthesis pathway genes may potentially modulate the association between HRT or age at the end of first full time pregnancy and breast cancer risk. To be more specific, the genetically predicted gene expression level of the estrogen synthesis pathway gene precited in either subcutaneous adipose tissue or in cross tissue, *AKR1B15* (p for interaction = 0.003 for subcutaneous adipose tissue, 0.004 for cross tissue), was observed to potentially modify the association between HRT and breast cancer risk. The positive association between HRT use and breast cancer risk is stronger among women with a higher genetically predicted expression level of *AKR1B15*. In addition, the genetically predicated gene expression level of the estrogen synthesis pathway gene, *HSD17B11* (P for interaction = 0.001),may also modulate the association between age at the end of first full time and breast cancer risk. The positive association between age

at the end of first full time pregnancy and breast cancer risk was stronger among women with a lower predicted expression of gene, *HSD17B11*. The estrogen metabolism gene, *B3GAT1* (P for interaction = 0.0004) and *IDS* (P for interaction = 0.001) predicted by the cross tissue and ovary tissue models, respectively, may modulate the association between parous and breast cancer risk. The inverse association between parous and breast cancer risk was stronger among women with lower gene expression level of the *UGT*s gene, *B3GAT1*. On the other hand, the inverse association between parous and breast cancer risk was stronger among women with higher gene expression level of the *ST*s gene, *IDS.* However, no similar findings were observed in the Asian population.

AKR1B15 plays a role in catalyzing the reduction of E2 to E1, which is a less active estrogen. Higher expression level of the gene, *AKR1B15,* in the adipose tissue, is considered to be associated with lower level of bioactive estrogen. Interestingly, gene, *HSD17B11*, plays the similar role as *AKR1B15,* which is to dehydrogenate E2 to biologically inactive form of estrogen, E1.

Older age at first full time pregnancy was associated with higher life time level of estrogen(114). Our study may indicate that women with the combination of lower biologically active estrogen and older age at first full time pregnancy may be more likely to develop breast cancer than their counterparts.

The estrogen metabolism gene, *B3GAT1,* is in the group of the *UGT* genes which play a role in the estrogen elimination process by conjugating estrogen metabolites with glucuronic acid. Lower expression level of this gene is considered to be

associated with higher bioavailable estrogen level in human body. Therefore, it might be theoretically plausible to assume that the lifetime estrogen reduction effect of pregnancy could be more helpful for women with a lower expression level of this gene. Significant interaction for parous was observed for the estrogen metabolism pathway in the subcutaneous adipose tissue (P for interaction = 0.020) in a consistent direction based on our hypothesis.

Some previous studies have also tried to test the interactions of our interest by using the measured plasma level of estrogen instead of genetically predicted gene expression. In 2004, a nested case-control study within the Nurses' Health Study has investigated whether the association between endogenous estrogen level and breast cancer risk in 322 cases and 643 controls; however, no significant interactions were observed (115). In 2013, a nested case-control study within the Nurses' Health Study II evaluated the potential effect modification of the association between endogenous steroid hormones and breast cancer risk among 634 cases and 1264 controls by: menopausal status at diagnosis, age at blood draw, BMI and duration of oral contraceptive use. Again, no significant findings were observed (116). Another nested case-control study from the Diet, Cancer. and Health' study using 254 cases and 442 controls has tested the interaction of serum estrogen level and HRT use on breast cancer risk. Also, no significant interaction was observed (117). However, a nested case-control study within the Women's Health Initiative randomized clinical trial of 348 cases and 348 controls found that the effect of HRT use on breast cancer risk was strongest in women with the lowest pretreatment level of total estradiol, bioavailable estradiol and estrone (P for interaction = 0.02), which is consistent with our findings

(118). One common limitation for the above-mentioned single case-control study is their small sample size, which conferred high possibility of insufficient power to detect the potential interactions.

To our knowledge, our study is by far the first and most comprehensive study to examine whether the genetically predicted gene expression level of estrogen related genes could modify the association of adiposity, menstrual and reproductive factors with breast cancer risk. Our innovative design and very large sample size minimize the limitation of low statistical power from previous individual SNP-based studies or single case-control studies to a great extent. However, this study also has several limitations. Firstly, the samples from GTEx used for model building is very limited, which may affect the reliability of the estimated weights for SNPs to some extent. However, at the current stage, no other resources with normal tissues are available for us to validate the models. TCGA data was used to check the validity of the model building of breast tissue in European population. However, samples from TCGA were tumor-adjacent normal tissue instead of tumor-free normal tissue as in the GTEx. Given the potential somatic alterations in tumor-adjacent normal tissues, the validation process may not accurately reflect the performance of our model building. Nevertheless, ten-fold cross-validation was used to validate the models internally. Based on a recent study conducted within our group, it seems such a strategy minimized the potential overfitting issue, although overfitting may not be completely controlled for. Similarly, the RNA-seq data from Shanghai, China was also not from completely healthy participants. We focused on models built using tissues from benign breast disease (BBD) patients in this study. However, it is likely that these samples don’t reflect normal tissue samples ant not

comparable to the those from GTEx, and the sample size is also very limited. Another limitation is the potential misclassification of some of the non-genetic factors. Since some of the included studies are case-control studies, the misclassification may not be non-differential. Lastly, false-positive results are plausible for the results of association and interaction in this study given the large number of tests included for each aim even if we have tied to controlled for multiple comparisons in this exploratory study.

In conclusion, our study found potential effect modification of the association of HRT use, parous and age at the end of first full time pregnancy with breast cancer risk by the genetically predicted expression level of estrogen synthesis or metabolism genes. Given the limitations embedded in this study, the interpretation of the results should be cautious, and additional studies are warranted to verify our findings. If verified in other studies, our findings may help identify subgroups of women with higher risk of breast cancer when using HRT or providing reference when make their reproductive life plan to reduce breast cancer risk.

Reference

1. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women: Breast Cancer Statistics, 2015. CA Cancer J Clin [Internet]. 2016 Jan [cited 2017 Sep 19];66(1):31–42. Available from: http://doi.wiley.com/10.3322/caac.21320

2. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev [Internet]. 2015 Oct 1 [cited 2017 Sep 19];24(10):1495–506. Available from: http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-15-0535

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016: Cancer Statistics, 2016. CA Cancer J Clin [Internet]. 2016 Jan [cited 2017 Sep 19];66(1):7–30. Available from: http://doi.wiley.com/10.3322/caac.21332

4. Prat A, Pineda E, Adamo B, Galván P, Fernández A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast Edinb Scotl. 2015 Nov;24 Suppl 2:S26-35.

5. Kang KS, Morita I, Cruz A, Jeon YJ, Trosko JE, Chang CC. Expression of estrogen receptors in a normal human breast epithelial cell type with luminal and stem cell characteristics and its neoplastically transformed cell lines. Carcinogenesis. 1997 Feb;18(2):251–7.

6. Feinleib M. Breast cancer and artificial menopause: a cohort study. J Natl Cancer Inst. 1968 Aug;41(2):315–29.

7. Key T, Appleby P, Barnes I, Reeves G, 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 Apr 17;94(8):606–16.

8. Manjer J, Johansson R, Berglund G, Janzon L, Kaaks R, Agren A, et al. Postmenopausal breast cancer risk in relation to sex steroid hormones, prolactin and SHBG (Sweden). Cancer Causes Control CCC. 2003 Sep;14(7):599–607.

9. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. J Natl Cancer Inst. 2004 Dec 15;96(24):1856–65.

10. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PHM, Biessy C, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. Endocr Relat Cancer. 2005 Dec;12(4):1071–82.

11. Baglietto L, Severi G, English DR, Krishnan K, Hopper JL, McLean C, et al. Circulating steroid hormone levels and risk of breast cancer for postmenopausal women. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2010 Feb;19(2):492–502.

12. Chlebowski RT, Manson JE, Anderson GL, Cauley JA, Aragaki AK, Stefanick ML, et al. Estrogen Plus Progestin and Breast Cancer Incidence and Mortality in the Women's Health Initiative Observational Study. JNCI J Natl Cancer Inst [Internet]. 2013 Apr 17 [cited 2017 Sep

22];105(8):526–35. Available from: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djt043

13. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SAA, Black H, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA. 2004 Apr 14;291(14):1701–12.

14. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA. 2002 Jul 17;288(3):321–33.

15. Hofseth LJ, Raafat AM, Osuch JR, Pathak DR, Slomski CA, Haslam SZ. Hormone replacement therapy with estrogen or estrogen plus medroxyprogesterone acetate is associated with increased epithelial proliferation in the normal postmenopausal breast. J Clin Endocrinol Metab. 1999 Dec;84(12):4559–65.

16. Chakravarti D, Mailander PC, Li KM, Higginbotham S, Zhang HL, Gross ML, et al. Evidence that a burst of DNA depurination in SENCAR mouse skin induces error-prone repair and forms mutations in the H-ras gene. Oncogene. 2001 Nov 29;20(55):7945–53.

17. Zhao Z, Kosinska W, Khmelnitsky M, Cavalieri EL, Rogan EG, Chakravarti D, et al. Mutagenic activity of 4-hydroxyestradiol, but not 2-hydroxyestradiol, in BB rat2 embryonic cells, and the mutational spectrum of 4-hydroxyestradiol. Chem Res Toxicol. 2006 Mar;19(3):475–9.

18. Mailander PC, Meza JL, Higginbotham S, Chakravarti D. Induction of A.T to G.C mutations by erroneous repair of depurinated DNA following estrogen treatment of the mammary gland of ACI rats. J Steroid Biochem Mol Biol. 2006 Nov;101(4–5):204–15.

19. Russo J, Lareef MH, Tahin Q, Hu YF, Slater C, Ao X, et al. 17Beta-estradiol is carcinogenic in human breast epithelial cells. J Steroid Biochem Mol Biol. 2002 Feb;80(2):149–62.

20. Russo J, Hasan Lareef M, Balogh G, Guo S, Russo IH. Estrogen and its metabolites are carcinogenic agents in human breast epithelial cells. J Steroid Biochem Mol Biol. 2003 Oct;87(1):1–25.

21. Fernandez SV, Russo IH, Russo J. Estradiol and its metabolites 4-hydroxyestradiol and 2-hydroxyestradiol induce mutations in human breast epithelial cells. Int J Cancer. 2006 Apr 15;118(8):1862–8.

22. Zahid M, Kohli E, Saeed M, Rogan E, Cavalieri E. The greater reactivity of estradiol-3,4-quinone vs estradiol-2,3-quinone with DNA in the formation of depurinating adducts: implications for tumor-initiating activity. Chem Res Toxicol. 2006 Jan;19(1):164–72.

23. Telang NT, Suto A, Wong GY, Osborne MP, Bradlow HL. Induction by estrogen metabolite 16 alpha-hydroxyestrone of genotoxic damage and aberrant proliferation in mouse mammary epithelial cells. J Natl Cancer Inst. 1992 Apr 15;84(8):634–8.

24. Santen RJ, Yue W, Wang J-P. Estrogen metabolites and breast cancer. Steroids. 2015 Jul;99(Pt A):61–6.

25. Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. Cancer Lett. 2015 Jan;356(2):231–43.

26. Burger HG. The endocrinology of the menopause. J Steroid Biochem Mol Biol. 1999 Jun;69(1–6):31–5.

27. Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. Cancer Lett [Internet]. 2015 Jan [cited 2017 Oct 4];356(2):231–43. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0304383514002365

28. Lane PH. Estrogen receptors in the kidney: lessons from genetically altered mice. Gend Med. 2008;5 Suppl A:S11-18.

29. Weiser MJ, Foradori CD, Handa RJ. Estrogen receptor beta in the brain: from form to function. Brain Res Rev. 2008 Mar;57(2):309–20.

30. Paruthiyil S, Parmar H, Kerekatte V, Cunha GR, Firestone GL, Leitman DC. Estrogen receptor beta inhibits human breast cancer cell proliferation and tumor formation by causing a G2 cell cycle arrest. Cancer Res. 2004 Jan 1;64(1):423–8.

31. Järvinen TA, Pelto-Huikko M, Holli K, Isola J. Estrogen receptor beta is coexpressed with ERalpha and PR and associated with nodal status, grade, and proliferation rate in breast cancer. Am J Pathol. 2000 Jan;156(1):29–35.

32. Speirs V, Walker R. New perspectives into the biological and clinical relevance of oestrogen receptors in the human breast. J Pathol [Internet]. 2007 Apr [cited 2018 Jan 3];211(5):499–506. Available from: http://doi.wiley.com/10.1002/path.2130

33. Selby C. Sex hormone binding globulin: origin, function and clinical significance. Ann Clin Biochem. 1990 Nov;27 (Pt 6):532–41.

34. Dunning AM, Dowsett M, Healey CS, Tee L, Luben RN, Folkerd E, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst. 2004 Jun 16;96(12):936–45.

35. Haiman CA, Riley SE, Freedman ML, Setiawan VW, Conti DV, Le Marchand L. Common genetic variation in the sex steroid hormone-binding globulin (SHBG) gene and circulating shbg levels among postmenopausal women: the Multiethnic Cohort. J Clin Endocrinol Metab. 2005 Apr;90(4):2198–204.

36. Beckmann L, Hüsing A, Setiawan VW, Amiano P, Clavel-Chapelon F, Chanock SJ, et al. Comprehensive analysis of hormone and genetic variation in 36 genes related to steroid hormone metabolism in pre- and postmenopausal women from the breast and prostate cancer cohort consortium (BPC3). J Clin Endocrinol Metab. 2011 Feb;96(2):E360-367.

37. Johnson N, Walker K, Gibson LJ, Orr N, Folkerd E, Haynes B, et al. CYP3A variation, premenopausal estrone levels, and breast cancer risk. J Natl Cancer Inst. 2012 May 2;104(9):657–69.

38. Miyoshi Y, Noguchi S. Polymorphisms of estrogen synthesizing and metabolizing genes and breast cancer risk in Japanese women. Biomed Pharmacother Biomedecine Pharmacother. 2003 Dec;57(10):471–81.

39. Vainio H, Kaaks R, Bianchini F. Weight control and physical activity in cancer prevention: international evaluation of the evidence. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP. 2002 Aug;11 Suppl 2:S94-100.

40. Simpson ER. Sources of estrogen and their importance. J Steroid Biochem Mol Biol. 2003 Sep;86(3–5):225–30.

41. Baglietto L, English DR, Hopper JL, MacInnis RJ, Morris HA, Tilley WD, et al. Circulating steroid hormone concentrations in postmenopausal women in relation to body size and composition. Breast Cancer Res Treat. 2009 May;115(1):171–9.

42. McTiernan A, Rajan KB, Tworoger SS, Irwin M, Bernstein L, Baumgartner R, et al. Adiposity and sex hormones in postmenopausal breast cancer survivors. J Clin Oncol Off J Am Soc Clin Oncol. 2003 May 15;21(10):1961–6.

43. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PHM, Biessy C, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. Endocr Relat Cancer. 2005 Dec;12(4):1071–82.

44. Rinaldi S, Key TJ, Peeters PHM, Lahmann PH, Lukanova A, Dossus L, et al. Anthropometric measures, endogenous sex steroids and breast cancer risk in postmenopausal women: a study within the EPIC cohort. Int J Cancer. 2006 Jun 1;118(11):2832–9.

45. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004 Aug;4(8):579–91.

46. Brown KA. Impact of obesity on mammary gland inflammation and local estrogen production. J Mammary Gland Biol Neoplasia. 2014 Jul;19(2):183–9.

47. Haffner SM, Katz MS, Dunn JF. Increased upper body and overall adiposity is associated with decreased sex hormone binding globulin in postmenopausal women. Int J Obes. 1991 Jul;15(7):471–8.

48. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. The Lancet [Internet]. 2008 Feb [cited 2017 Dec 21];371(9612):569–78. Available from: http://linkinghub.elsevier.com/retrieve/pii/S014067360860269X

49. Huang Z, Hankinson SE, Colditz GA, Stampfer MJ, Hunter DJ, Manson JE, et al. Dual effects of weight and weight gain on breast cancer risk. JAMA. 1997 Nov 5;278(17):1407–11.

50. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol. 2000 Sep 15;152(6):514–27.

51. Vatten LJ, Kvinnsland S. Prospective study of height, body mass index and risk of breast cancer. Acta Oncol Stockh Swed. 1992;31(2):195–200.

52. Ursin G, Longnecker MP, Haile RW, Greenland S. A meta-analysis of body mass index and risk of premenopausal breast cancer. Epidemiol Camb Mass. 1995 Mar;6(2):137–41.

53. Michels KB, Terry KL, Willett WC. Longitudinal study on the role of body size in premenopausal breast cancer. Arch Intern Med. 2006 Nov 27;166(21):2395–402.

54. Dowsett M, Folkerd E. Reduced progesterone levels explain the reduced risk of breast cancer in obese premenopausal women: a new hypothesis. Breast Cancer Res Treat. 2015 Jan;149(1):1–4.

55. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. The Lancet [Internet]. 2014 Aug [cited 2017 Dec 21];384(9945):755–65. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673614608928

56. Guo Y, Warren Andersen S, Shu X-O, Michailidou K, Bolla MK, Wang Q, et al. Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent. Beck AH, editor. PLOS Med [Internet]. 2016 Aug 23 [cited 2017 Dec 21];13(8):e1002105. Available from: http://dx.plos.org/10.1371/journal.pmed.1002105

57. Guo Y, Warren Andersen S, Shu X-O, Michailidou K, Bolla MK, Wang Q, et al. Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent. PLoS Med. 2016 Aug;13(8):e1002105.

58. Guan H-B, Wu L, Wu Q-J, Zhu J, Gong T. Parity and Pancreatic Cancer Risk: A Dose-Response Meta-Analysis of Epidemiologic Studies. Real FX, editor. PLoS ONE [Internet]. 2014 Mar 21 [cited 2017 Dec 21];9(3):e92738. Available from: http://dx.plos.org/10.1371/journal.pone.0092738

59. Bernstein L, Pike MC, Ross RK, Judd HL, Brown JB, Henderson BE. Estrogen and sex hormone-binding globulin levels in nulliparous and parous women. J Natl Cancer Inst. 1985 Apr;74(4):741–5.

60. Bandini L, Must A, Naumova E, Anderson S, Caprio S, Spadano-Gasbarro J, et al. Change in leptin, body composition and other hormones around menarche - a visual representation. Acta Paediatr [Internet]. 2008 Oct [cited 2017 Dec 21];97(10):1454–9. Available from: http://doi.wiley.com/10.1111/j.1651-2227.2008.00948.x

61. Zacur HA. Hormonal changes throughout life in women. Headache. 2006 Oct;46 Suppl 2:S49-54.

62. Arnold M, Tóth I, Faredin I. [Radioimmunological study of the effect of hormonal contraceptives upon the progesterone level of saliva (author's transl)]. Zahn Mund Kieferheilkd Zentralbl. 1980;68(7):713–8.

63. Basu J, Mikhail MS, Palan PR, Thysen B, Bloch E, Romney SL. Endogenous estradiol and progesterone concentrations in smokers on oral contraceptives. Gynecol Obstet Invest. 1992;33(4):224–7.

64. Balogh A, Ditrói F, Lampé LG. LH, FSH, estradiol and progesterone levels after discontinuation of hormonal contraception. Acta Univ Palacki Olomuc Fac Med. 1981;101:95–101.

65. Panzer C, Wise S, Fantini G, Kang D, Munarriz R, Guay A, et al. Impact of oral contraceptives on sex hormone-binding globulin and androgen levels: a retrospective study in women with sexual dysfunction. J Sex Med. 2006 Jan;3(1):104–13.

66. Fleischman DS, Navarrete CD, Fessler DMT. Oral Contraceptives Suppress Ovarian Hormone Production. Psychol Sci [Internet]. 2010 May [cited 2017 Dec 29];21(5):750–2. Available from: http://journals.sagepub.com/doi/10.1177/0956797610368062

67. Chan M-F, Dowsett M, Folkerd E, Wareham N, Luben R, Welch A, et al. Past oral contraceptive and hormone therapy use and endogenous hormone concentrations in postmenopausal women. Menopause N Y N. 2008 Apr;15(2):332–9.

68. Edlefsen KL, Jackson RD, Prentice RL, Janssen I, Rajkovic A, O'Sullivan MJ, et al. The effects of postmenopausal hormone therapy on serum estrogen, progesterone, and sex hormone-binding globulin levels in healthy postmenopausal women: Menopause [Internet]. 2010 Mar [cited 2017 Dec 29];1. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00042192-900000000-99264

69. MacMahon B, Trichopoulos D, Brown J, Andersen AP, Cole P, deWaard F, et al. Age at menarche, urine estrogens and breast cancer risk. Int J Cancer. 1982 Oct 15;30(4):427–31.

70. Apter D, Reinilä M, Vihko R. Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. Int J Cancer. 1989 Nov 15;44(5):783–7.

71. Wohlfahrt J, Andersen PK, Mouridsen HT, Melbye M. Risk of late-stage breast cancer after a childbirth. Am J Epidemiol. 2001 Jun 1;153(11):1079–84.

72. Neale RE, Darlington S, Murphy MFG, Silcocks PBS, Purdie DM, Talbäck M. The effects of twins, parity and age at first birth on cancer risk in Swedish women. Twin Res Hum Genet Off J Int Soc Twin Stud. 2005 Apr;8(2):156–62.

73. Lee SY, Kim MT, Kim SW, Song MS, Yoon SJ. Effect of lifetime lactation on breast cancer risk: a Korean women's cohort study. Int J Cancer. 2003 Jun 20;105(3):390–3.

74. Neale RE, Darlington S, Murphy MFG, Silcocks PBS, Purdie DM, Talbäck M. The effects of twins, parity and age at first birth on cancer risk in Swedish women. Twin Res Hum Genet Off J Int Soc Twin Stud. 2005 Apr;8(2):156–62.

75. Henderson BE, Ross RK, Judd HL, Krailo MD, Pike MC. Do regular ovulatory cycles increase breast cancer risk? Cancer. 1985 Sep 1;56(5):1206–8.

76. Sweeney C. Risk Factors for Breast Cancer in Elderly Women. Am J Epidemiol [Internet]. 2004 Nov 1 [cited 2017 Dec 29];160(9):868–75. Available from: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwh276

77. Kelsey JL, Fischer DB, Holford TR, LiVoisi VA, Mostow ED, Goldenberg IS, et al. Exogenous estrogens and other factors in the epidemiology of breast cancer. J Natl Cancer Inst. 1981 Aug;67(2):327–33.

78. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet Lond Engl. 1996 Jun 22;347(9017):1713–27.

79. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, et al. Oral Contraceptives and the Risk of Breast Cancer. N Engl J Med [Internet]. 2002 Jun 27 [cited 2018 May 22];346(26):2025–32. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa013202

80. Chlebowski RT, Anderson GL. Changing concepts: Menopausal hormone therapy and breast cancer. J Natl Cancer Inst. 2012 Apr 4;104(7):517–27.

81. An J, Tzagarakis-Foster C, Scharschmidt TC, Lomri N, Leitman DC. Estrogen Receptor β-Selective Transcriptional Activity and Recruitment of Coregulators by Phytoestrogens. J Biol Chem [Internet]. 2001 May 25 [cited 2018 Jan 3];276(21):17808–14. Available from: http://www.jbc.org/lookup/doi/10.1074/jbc.M100953200

82. Messina M, McCaskill-Stevens W, Lampe JW. Addressing the soy and breast cancer relationship: review, commentary, and workshop proceedings. J Natl Cancer Inst. 2006 Sep 20;98(18):1275–84.

83. Hilakivi-Clarke L, Andrade JE, Helferich W. Is Soy Consumption Good or Bad for the Breast? J Nutr [Internet]. 2010 Dec 1 [cited 2018 Jan 8];140(12):2326S-2334S. Available from: http://jn.nutrition.org/cgi/doi/10.3945/jn.110.124230

84. Lephart ED. Modulation of Aromatase by Phytoestrogens. Enzyme Res [Internet]. 2015 [cited 2018 Jan 8];2015:1–11. Available from: http://www.hindawi.com/journals/er/2015/594656/

85. Wu AH, Yu MC, Tseng C-C, Pike MC. Epidemiology of soy exposures and breast cancer risk. Br J Cancer [Internet]. 2008 Jan [cited 2018 Jan 8];98(1):9–14. Available from: http://www.nature.com/articles/6604145

86. Lee S-A, Shu X-O, Li H, Yang G, Cai H, Wen W, et al. Adolescent and adult soy food intake and breast cancer risk: results from the Shanghai Women's Health Study. Am J Clin Nutr [Internet]. 2009 Jun 1 [cited 2018 Jan 8];89(6):1920–6. Available from: http://www.ajcn.org/cgi/doi/10.3945/ajcn.2008.27361

87. Feigelson HS, Coetzee GA, Kolonel LN, Ross RK, Henderson BE. A polymorphism in the CYP17 gene increases the risk of breast cancer. Cancer Res. 1997 Mar 15;57(6):1063–5.

88. Haiman CA, Hankinson SE, Spiegelman D, Colditz GA, Willett WC, Speizer FE, et al. The relationship between a polymorphism in CYP17 with plasma hormone levels and breast cancer. Cancer Res. 1999 Mar 1;59(5):1015–20.

89. Le Marchand L, Donlon T, Kolonel LN, Henderson BE, Wilkens LR. Estrogen metabolism-related genes and breast cancer risk: the multiethnic cohort study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2005 Aug;14(8):1998–2003.

90. Rebbeck TR, Troxel AB, Shatalova EG, Blanchard R, Norman S, Bunin G, et al. Lack of effect modification between estrogen metabolism genotypes and combined hormone replacement therapy in postmenopausal breast cancer risk. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2007 Jun;16(6):1318–20.

91. Chen Y, Gammon MD, Teitelbaum SL, Britton JA, Terry MB, Shantakumar S, et al. Estrogen-biosynthesis gene CYP17 and its interactions with reproductive, hormonal and lifestyle factors in breast cancer risk: results from the Long Island Breast Cancer Study Project. Carcinogenesis. 2008 Apr;29(4):766–71.

92. Justenhoven C, Hamann U, Schubert F, Zapatka M, Pierl CB, Rabstein S, et al. Breast cancer: a candidate gene approach across the estrogen metabolic pathway. Breast Cancer Res Treat. 2008 Mar;108(1):137–49.

93. Diergaarde B, Potter JD, Jupe ER, Manjeshwar S, Shimasaki CD, Pugh TW, et al. Polymorphisms in genes involved in sex hormone metabolism, estrogen plus progestin hormone therapy use, and risk of postmenopausal breast cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2008 Jul;17(7):1751–9.

94. Reding KW, Chen C, Lowe K, Doody DR, Carlson CS, Chen CT, et al. Estrogen-related genes and their contribution to racial differences in breast cancer risk. Cancer Causes Control [Internet]. 2012 May [cited 2018 Jan 9];23(5):671–81. Available from: http://link.springer.com/10.1007/s10552-012-9925-x

95. Quan L, Hong C-C, Zirpoli G, Roberts MR, Khoury T, Sucheston-Campbell LE, et al. Variants of estrogen-related genes and breast cancer risk in European and African American women. Endocr Relat Cancer. 2014;21(6):853–64.

96. Iwasaki M, Hamada GS, Nishimoto IN, Netto MM, Motola Jr J, Laginha FM, et al. Isoflavone, polymorphisms in estrogen receptor genes and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. Cancer Sci [Internet]. 2009 May [cited 2018 Jan 9];100(5):927–33. Available from: http://doi.wiley.com/10.1111/j.1349-7006.2009.01118.x

97. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet [Internet]. 2013 May 29 [cited 2018 Jan 11];45(6):580–5. Available from: http://www.nature.com/doifinder/10.1038/ng.2653

98. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinforma Oxf Engl. 2012 Jun 1;28(11):1530–2.

99. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. Nat Genet. 2014 May;46(5):430–7.

100. Casbas-Hernandez P, Sun X, Roman-Perez E, D'Arcy M, Sandhu R, Hishida A, et al. Tumor Intrinsic Subtype Is Reflected in Cancer-Adjacent Tissue. Cancer Epidemiol Biomarkers Prev [Internet]. 2015 Feb 1 [cited 2019 Mar 11];24(2):406–14. Available from: http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-14-0934

101. Huang X, Stern DF, Zhao H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival – Evidence from TCGA Pan-Cancer Data. Sci Rep [Internet]. 2016 Aug [cited 2019 Mar 11];6(1). Available from: http://www.nature.com/articles/srep20567

102. Cai Q, Zhang B, Sung H, Low S-K, Kweon S-S, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nat Genet. 2014 Aug;46(8):886–90.

103. Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009 Mar;41(3):324–8.

104. Zhang Y, Long J, Lu W, Shu X-O, Cai Q, Zheng Y, et al. Rare coding variants and breast cancer risk: evaluation of susceptibility Loci identified in genome-wide association studies. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2014 Apr;23(4):622–8.

105. Michigan: Universtiy of MIchigan. Exomechip design [Internet] [cited 2018 Mar 26]. Available from: http://genome.sph.umich.edu/ wiki/Exome_Chip_Design. 2013 Mar.

106. Barbeira AN, Dickinson SP, Torres JM, Bonazzola R, Zheng J, Torstenson ES, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. bioRxiv [Internet]. 2017 Oct 3 [cited 2019 Mar 4]; Available from: http://biorxiv.org/lookup/doi/10.1101/045260

107. Williams GP, Darbre PD. Low-dose environmental endocrine disruptors, increase aromatase activity, estradiol biosynthesis and cell proliferation in human breast cells. Mol Cell Endocrinol. 2019 Feb 26;

108. Abdul Aziz AA, Md Salleh MS, Mohamad I, Krishna Bhavaraju VM, Mazuwin Yahya M, Zakaria AD, et al. Single-nucleotide polymorphisms and mRNA expression of CYP1B1 influence treatment response in triple negative breast cancer patients undergoing chemotherapy. J Genet. 2018 Dec;97(5):1185–94.

109. Long J-R, Shu X-O, Cai Q, Cai H, Gao Y-T, Jin F, et al. Polymorphisms of the CYP1B1 gene may be associated with the onset of natural menopause in Chinese women. Maturitas. 2006 Oct 20;55(3):238–46.

110. Kumaran M, Cass CE, Graham K, Mackey JR, Hubaux R, Lam W, et al. Germline copy number variations are associated with breast cancer risk and prognosis. Sci Rep. 2017 Nov 7;7(1):14621.

111. Dates CR, Fahmi T, Pyrek SJ, Yao-Borengasser A, Borowa-Mazgaj B, Bratton SM, et al. Human UDP-Glucuronosyltransferases: Effects of altered expression in breast and pancreatic cancer cell lines. Cancer Biol Ther. 2015;16(5):714–23.

112. Wang J, Scholtens D, Holko M, Ivancic D, Lee O, Hu H, et al. Lipid metabolism genes in contralateral unaffected breast and estrogen receptor status of breast cancer. Cancer Prev Res Phila Pa. 2013 Apr;6(4):321–30.

113. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. The Lancet [Internet]. 2004 Jan [cited 2018 Mar 20];363(9403):157–63. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673603152683

114. Bernstein L, Pike MC, Ross RK, Judd HL, Brown JB, Henderson BE. Estrogen and sex hormone-binding globulin levels in nulliparous and parous women. J Natl Cancer Inst. 1985 Apr;74(4):741–5.

115. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. J Natl Cancer Inst. 2004 Dec 15;96(24):1856–65.

116. Fortner RT, Eliassen AH, Spiegelman D, Willett WC, Barbieri RL, Hankinson SE. Premenopausal endogenous steroid hormones and breast cancer risk: results from the Nurses' Health Study II. Breast Cancer Res BCR. 2013 Mar 6;15(2):R19.

117. Würtz AML, Tjønneland A, Christensen J, Dragsted LO, Aarestrup J, Kyrø C, et al. Serum estrogen and SHBG levels and breast cancer incidence among users and never users of hormone replacement therapy. Cancer Causes Control CCC. 2012 Oct;23(10):1711–20.

118. Farhat GN, Parimi N, Chlebowski RT, Manson JE, Anderson G, Huang AJ, et al. Sex Hormone Levels and Risk of Breast Cancer With Estrogen Plus Progestin. JNCI J Natl Cancer Inst [Internet]. 2013 Oct 2 [cited 2019 Mar 11];105(19):1496–503. Available from: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djt243