Materials and Methods

Study sample

A total of 3,852 breast cancer cases and controls from the pooled data from the European Prospective Investigation into Cancer and Nutrition (EPIC, n=2,772) and the Nurses' Health study (NHS, n=1,080) were analyzed in the present study. The EPIC sample consisted of 937 cases and 1,835 controls. 1,037 women were premenopausal at the time of blood donation, 1,735 women were postmenopausal (1). Women were considered premenopausal at the time of blood donation if they reported having had at least nine menstrual periods over the previous 12 months. Women who had missing or incomplete questionnaire data on menstrual periods or who had had a hysterectomy were considered premenopausal if they were younger than 42 years because among the female EPIC participants who had complete questionnaire data, 99.5% of those younger than age 42 years were premenopausal. Women were considered postmenopausal when they reported not having had any menses over the past 12 months or when they reported bilateral ovariectomy. When questionnaire data were missing or incomplete or when women reported previous hysterectomy, women were considered postmenopausal when they were older than 55 years (1;2).

The NHS sample consisted of 359 cases and 721 controls. All women from NHS were postmenopausal. A postmenopausal participant was defined as a woman who reported having a natural menopause or a bilateral oophorectomy or as a woman who reported having a hysterectomy with either one or both ovaries remaining when she was older than 56 years, if she was a nonsmoker, or older than 54 years if she was a current smoker. At these ages the natural menopause had occurred in 99% of the women in these groups (3).

The MEC subsample used for the analysis here consisted only of postmenopausal controls, for which hormone measurements were available (119 African-Americans, 84 Asians, 79 Caucasians, 70 Hawaiians, and 102 Latinos). The study participants of the MEC were assigned to the ethnicities according to self-reported affiliation (4). The MEC sample included women who were older than 56 years at the time of blood draw, who did not report a history of breast, endometrial or ovarian cancer on the baseline

questionnaire, who had body weight and body mass index (BMI) information available (5). In all three cohort studies, women who did use postmenopausal hormones at baseline or at the time of blood draw were excluded. Informed consent was obtained from patients prior to sample collection for all cohort studies. Ethics approvals were obtained for all cohort studies involved.

Breast cancer cases were identified in each cohort by self report, with subsequent confirmation of the diagnosis, including tumor details, from medical records, and/or linkage with population-based tumor registries. Controls were matched to the breast cancer cases by ethnicity and age, and in some cohorts, additional matching criteria were employed (e.g., EPIC matched on country of residence). In EPIC and NHS, controls and cases were also matched for time at blood donation to account for circadian rhythm of hormone levels. Informed consent was obtained from each woman. Further details about the study participants and available covariates such as tumor stage, age, menopausal status, and BMI can be found in Table 1.

Hormone Measurement

Hormone concentrations of dehydro-epiandrosterone sulphate (DHEAS), delta-4 androstenedione (Δ4), testosterone (TESTO), estrone (E1), estradiol (E2) and sex-hormone binding globulin (SHBG) were measured in blood samples of women who subsequently developed breast cancer as well as in matched control subjects, as part of previous prospective cohort studies EPIC, NHS and MEC. All hormones were measured on non-users of exogenous hormones (oral contraceptive or hormone replacement therapy). E1 (n=2,433 control subjects from EPIC/NHS, n=422 cross-sectional sample of MEC) and E2 (n=2,721 EPIC/NHS, n=424 MEC) were measured in postmenopausal women only, because of the large effect of menopause on their blood concentrations, and because of the large variations of these hormones in premenopausal women during the menstrual cycle. Δ4 (n=3,530 EPIC/NHS), DHEAS (n=3,547 EPIC/NHS), SHBG (3,796 EPIC/NHS, n=451 MEC) and TESTO (n=3,752 EPIC/NHS) were measured in both pre- and postmenopausal women for EPIC/NHS, and for postmenopausal women in MEC. DHEAS measures were not available for MEC. TESTO and Δ4 were not analyzed in MEC, because no significant

associations were identified in EPIC/NHS. The measurements were made by direct radioimmunoassays for EPIC (1;2), and by a mixture of direct and indirect radioimmunoassays for NHS (3;6) and MEC (5); details about these methods are given in Table 1.

Genetic data

General outline for the Identification of polymorphisms and tagging SNPs

The methods of identifying tagging SNPs, genotyping and quality control in BPC3 are described elsewhere (7). The final set of SNPs was derived in a three stage procedure, explained in detail below. First, an extensive catalog of SNPs in the candidate genes was generated by systematic resequencing of these genes in breast and prostate cancer cases with different ethnic backgrounds. Secondly, all SNPs with minor allele frequency MAF >1% were then genotyped in larger sets, so as to provide more detailed insight in the linkage disequilibrium (LD) pattern. This data set combined with information from the HapMap project, was used to select the subset of tagging SNPs. The aim of the marker selection procedure was to identify tagging SNPs that capture most of the common variants within the candidate genes. Thirdly, these SNPs were genotyped in all individuals in BPC3 and finally undergone extensive quality control.

Gene sequencing

All candidate genes were sequenced at three BPC3 collaborating genome centers (USC/Broad Institute, CEPH, and NCI) in order to establish a comprehensive catalogue of their common genetic variants. Exons, intron/exon junctions and evolutionarily conserved (at least 80% homology with mouse sequence over at least 200bp) sequences in introns and sequences up to 30kb 5' of transcription start and 10kb 3' of translation end of each gene were sequenced in a panel of 95 advanced breast cancer cases from the MEC and EPIC (19 of each ethnic group represented in the study: African American, Latino, Japanese, Native Hawaiian, and Caucasian). SNPs with minor allele frequency greater than 5% in any of the five ethnic groups or greater than 1% overall were selected for further work.

Genotyping

Genotyping of tagging SNPs in the breast cancer cases and controls was performed in five laboratories (USC; NCI; Harvard School of Public Health; Imperial College (London); and the International Agency for Research in Cancer, IARC). During the first part of the study, 172 SNPs were typed using a TaqMan platform with a fluorescent 5′ endonuclease assay and the ABI-PRISM 7900 for sequence detection. We then switched to the GoldenGate assay and Illumina BeadArray™ technology and created an OPA of 1,536 SNPs, which are located in the 36 genes of the sex-steroid pathway that are part of this analysis as well as 23 genes that are part of the insulin-like growth factor signaling pathway. Thirty trios of European descent from Utah (CEU) were genotyped in all laboratories to evaluate inter-laboratory reproducibility. These trios were collected by the Centre d'Étude du Polymorphisme Humain (CEPH), Paris, France, and used in HapMap to represent common variation in the Caucasian population. For the Illumina OPA, the overall concordance was 99.5% (before excluding failed SNPs or samples), and the concordance rate of blinded duplicate samples (~5%) ranged from 97.2-99.9% across the various cohort studies contributing to BPC3. The very high reliability of both TaqMan and Illumina was also constantly monitored by stringent inter- and intra-lab quality control all along the project. The method of (8) was used to select SNPs based on pairwise LD, and the approach of (9) to identify a set of tagging SNPs that optimize the predictability of common haplotypes.

Quality control

Any individual in which more than 25% of the OPA SNPs failed was dropped from the OPA analysis. All SNPs that failed on 25% or more samples in the OPA or 10% or more with TaqMan were excluded from further analysis, as were all SNPs that showed statistically significant ($p<10^{-5}$) deviations from Hardy-Weinberg equilibrium among European-ancestry controls, and all SNPs with MAF<1%. Any SNP that was missing in more than three of the genotyped five to six study cohorts, or that exhibited large differences in European-ancestry allele frequencies across cohorts was also excluded from further analysis (Fixation index $F_{st}>0.02$).

Imputation of missing genotypes and non-genotyped SNPs

We imputed 2,671 SNPs that were polymorphic in any of the HapMap reference panels using observed genotypes from the BPC3 subjects (OPA and TaqMan) and phased haplotypes from HapMap samples (release #21) for all study samples within the BPC3 consortium using the software MACH 1.0, a Markov Chain based haplotype algorithm that infers missing genotypes for both genotyped and non-genotyped SNPs in samples of unrelated individuals. Genotypes for European-ancestry subjects were imputed using the CEPH European (CEU) reference panel; those for Japanese Americans were imputed using the combined Han Chinese and Japanese panels (CHB+JPT). The remaining subjects (African Americans, Latinos, Native Hawaiians) were imputed using a "cosmopolitan" panel of all HapMap samples (CEU+CHB+JPT+YRI) (10). Imputation was performed stratified by study and ethnicity. SNPs with an average estimated correlation between the imputed and true genotypes of less than 30% were excluded from analysis. Details for the selection and the genotyping of SNPs can be found in the supplement (Supplementary Table S1).

Statistical methods

Descriptive analysis

The hormone concentrations were measured in different laboratories by either direct or indirect RIA. To adjust for differences in the absolute levels of sex steroid hormones observed between the cohorts, we took the natural logarithm of the hormone levels regressed on breast cancer case-control status, BMI, menopausal status (for pre- and postmenopausal women pooled) and age at blood donation. We adjusted further for assay batch to account for possible laboratory-induced differences in hormone measurement, and thereby also adjusted for cohort membership. We used the residuals of the regression models as outcome variables for the subsequent association analysis. For the analysis of the MEC sample, we further adjusted additionally for ethnicity.

To analyze potential population stratification between EPIC and NHS and to detect genetic outliers within the cohorts, we performed principal components analysis using the software EIGENSTRAT (11;12). No population stratification in the pooled sample of EPIC and NHS could be detected and no outliers were identified.

Association tests

Association of genes with variation in hormone levels

We tested for associations of genes as well as for epistasis with variation in hormone levels in the pooled sample of EPIC and NHS. The genes, which were found to be significantly associated after adjustment for multiple testing and correlation, were reanalyzed in the five ethnic groups of MEC and in all groups combined in the MEC.

To identify globally significant genes, we combined the "MAX" test proposed by Freidlin et al. (2002) with the step-down-min-p algorithm proposed by Westfall and Young (1993). This combination controls for the family wise error rate, while adjusting for multiple testing and correlation between the test statistics (13-15). We performed linear regression of the residuals on each SNP independently for four models: recessive ("Rec"), dominant ("Dom"), the additive model with 1 degree of freedom (df) that corresponds to the usual trend test ("Trend"), and the additive model with 2 dfs ("Add"). For each SNP, we defined $p_{min}=\min(p_{Rec}, p_{Dom}, p_{Trend}, p_{Add})$ as the minimum p-value of the p-values derived by the respective tests. For all SNPs within a gene, we applied the step-down-min-p-algorithm based on 1,000 permutations to adjust for multiple testing, i.e. the number of SNPs and of the four association tests at each SNP ("Rec", "Dom", "Trend" and "Add"), taking into account the correlation between the SNPs due to LD (14;16;17). Finally, the adjusted p-values of the SNPs were further Bonferroni-corrected by the number of genes studied. The minimum adjusted p-value among the SNPs for any given gene was considered as the global p-value of this gene.

We considered a gene (and the corresponding SNP) to be globally significant if the global p-value, adjusted for multiple testing within any gene, was lower than the significance level $\alpha=0.05$. If the step-

down-min-p algorithm yielded a p-value of p= 0.0 based on 1,000 permutations, it indicates $p < 0.001$. Thus the corresponding global p-value is $p_{global} < 0.036$, as calculated by Bonferroni correction for 36 genes. For the confirmatory analysis in the MEC samples, where we studied whether significant associations observed within the combined sample of EPIC and NHS were also detectable in women of other ethnic backgrounds, a more lenient p-value of 0.05 was used to denote a SNP significant.

The effect of the SNPs are presented as differences in geometric means of the effect estimates of the linear regression, with the homozygous major genotype as reference group. The sign of the difference indicates the direction, i.e. a positive difference is associated with higher levels, while a negative difference is associated with lower levels. However, the focus on common variants selected for their ability to capture the genetic variation within genes may hamper the comparison with results from other studies.

We tested for gene - gene interactions (GxG) using a two step procedure. First, we tested for each hormone all 226,312 models that included the main effects of two SNPs from two different genes plus their interaction term, using the software Plink (18). The SNPs were coded as the number of minor alleles, corresponding to the trend model. If this test against the null model showed a p-value $< 10^{-3}$, we performed a likelihood ratio test with one degree of freedom of the model with interaction term against the reduced model, which included only the main effects of the two SNPs.

Association of genes with breast cancer risk

The genes identified to be significantly associated with variation in hormone levels were tested for association with breast cancer risk in a logistic model, adjusted for age at blood donation, BMI, batch, and for menopausal status, if required. The same strategy as outlined above was used to correct for multiple testing and correlation between the test statistics.

Software

We used the following software programs for the statistical analysis:

R: Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Plink: http://pngu.mgh.harvard.edu/purcell/plink/ a toolset for whole-genome association and population-based linkage analysis (18)

Mach 1.0: www.sph.umich.edu/csg/abecasis/MACH/

Reference List

1. **Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PH, Biessy C, Dossus L, Lukanova A, Bingham S, Khaw KT, Allen NE, Bueno-de-Mesquita HB, van Gils CH, Grobbee D, Boeing H, Lahmann PH, Nagel G, Chang-Claude J, Clavel-Chapelon F, Fournier A, Thiebaut A, Gonzalez CA, Quiros JR, Tormo MJ, Ardanaz E, Amiano P, Krogh V, Palli D, Panico S, Tumino R, Vineis P, Trichopoulou A, Kalapothaki V, Trichopoulos D, Ferrari P, Norat T, Saracci R, Riboli E** 2005 Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. Endocr Relat Cancer 12(4):1071-1082.

2. **Kaaks R, Berrino F, Key T, Rinaldi S, Dossus L, Biessy C, Secreto G, Amiano P, Bingham S, Boeing H, Bueno de Mesquita HB, Chang-Claude J, Clavel-Chapelon F, Fournier A, van Gils CH, Gonzalez CA, Gurrea AB, Critselis E, Khaw KT, Krogh V, Lahmann PH, Nagel G, Olsen A, Onland-Moret NC, Overvad K, Palli D, Panico S, Peeters P, Quiros JR, Roddam A, Thiebaut A, Tjonneland A, Chirlaque MD, Trichopoulou A, Trichopoulos D, Tumino R, Vineis P, Norat T, Ferrari P, Slimani N, Riboli E** 2005 Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). J Natl Cancer Inst 97(10):755-765.

3. **Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE** 2004 Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. J Natl Cancer Inst 96(24):1856-1865.

4. **Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS** 2000 A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 151(4):346-357.

5. **Setiawan VW, Haiman CA, Stanczyk FZ, Le ML, Henderson BE** 2006 Racial/ethnic differences in postmenopausal endogenous hormones: the multiethnic cohort study. Cancer Epidemiol Biomarkers Prev 15(10):1849-1855.

6. **Hankinson SE, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, Barbieri RL, Speizer FE** 1998 Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 90(17):1292-1299.

7. **Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, Haynes RB, Henderson BE, Kaaks R, Stram DO, Thomas G, Thun MJ, Blanche H, Buring JE, Burtt NP, Calle EE, Cann H, Canzian F, Chen YC, Colditz GA, Cox DG, Dunning AM, Feigelson HS, Freedman ML, Gaziano JM, Giovannucci E, Hankinson SE, Hirschhorn JN, Hoover RN, Key T, Kolonel LN, Kraft P, Le ML, Liu S, Ma J, Melnick S, Pharaoh P, Pike MC, Rodriguez C, Setiawan VW, Stampfer MJ, Trapido E, Travis R, Virtamo J, Wacholder S, Willett WC** 2005

A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Genet 5(12):977-985.

8. **Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D** 2002 The structure of haplotype blocks in the human genome. Science 296(5576):2225-2229.

9. **Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC** 2003 Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum Hered 55(1):27-36.

10. **de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D** 2005 Efficiency and power in genetic association studies. Nat Genet 37(11):1217-1223.

11. **Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D** 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904-909.

12. **Patterson N, Price AL, Reich D** 2006 Population structure and eigenanalysis. PLoS Genet 2(12):e190.

13. **Freidlin B, Zheng G, Li Z, Gastwirth JL** 2002 Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered 53(3):146-152.

14. Westfall PH, Young SS. Resampling-based multiple testing : examples and methods for P-value adjustment. New York: Wiley, 1993.

15. **Ziegler A, Ewhida A, Brendel M, Kleensang A** 2008 More powerful haplotype sharing by accounting for the mode of inheritance. Genet Epidemiol.

16. **Ge YC, Dudoit S, Speed TP** 2003 Resampling-based multiple testing for microarray data analysis. Test 12(1):1-77.

17. **Obreiter M, Fischer C, Chang-Claude J, Beckmann L** 2005 SDMinP: a program to control the family wise error rate using step-down minP adjusted P-values. Bioinformatics 21(14):3183-3184.

18. **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC** 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559-575.

Table 1: Summary table of the genes analyzed. Presented are details about sequencing, genotyping and tagging SNP selection.

					Sequencing / genotyping					Genotyping of tagging SNPs [g]			
Gene symbol	**Gene name**	**Chr**	**Location**	**Target region [a]**	**Sample [b]**	**N identified polymorphisms**	**Sample[c]**	**Method[d]**	**Lab**	**N SNPs genotyped [e]**	**Technology[d]**	**N SNPs**	**N imputed SNPs**	**Total**
ACTHR	adrenocorticotropic hormone receptor	18	18p11.2	43,088	S190	44	SM1	Ill, SNP, Taq	CEPH	49	Ill, Taq	18	69	87
ACVR1	activin A receptor, type I	2	2q23-q24	131,830	S95	23	SM2	Ill, Seq	USC/Broad	55	Ill	18	49	67
ACVR2	activin A receptor, type II	2	2q22.3	112,939	S95	7	SM2	Ill, Seq	USC/Broad	55	Ill	8	53	61
AKR1C3	aldo-keto reductase family 1, member C3	10	10p15-p14	57,665	S190	69	SM1	Ill, SNP, Taq	CEPH	74	Ill	34	135	169
CGA	glycoprotein hormones, alpha polypeptide	6	6q12-q21	39,603	S95	2	SM2	Ill, Seq	USC/Broad	62	Ill	15	35	50
COMT	catechol-O-methyltransferase	22	22q11.21	59,098	S102	96 (entire gene)	SM3	Ill, Taq	NCI-CGF	24	Taq	8	48	56
CYP11A1	cytochrome P450, family 11, subfamily A, polypeptide 1	15	15q23-q24	28,487	S95	13	SM2	Ill, Seq	USC/Broad	37	Taq	2	26	28
CYP17A1	cytochrome P450, family 17, subfamily A, polypeptide 1	10	10q24.3	35,927	S190	20	SM2	Ill, Seq	USC/Broad	32	Taq	7	32	39
CYP19A1	cytochrome P450, family 19, subfamily A, polypeptide 1	15	15q21.1	38,005	S95	24	SM2	Ill, Seq	USC/Broad	105	Taq	16	168	184
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	15	15q24.1	59,864	S102	56 (entire gene)	SM3	Ill, Taq	NCI-CGF	22	Taq	5	3	8
CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2	15	15q24.1	36,885	S102	72 (entire gene)	SM3	Ill, Taq	NCI-CGF	36	Taq	2	9	11
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	2	2p21	159,126	S102	140 (entire gene)	SM3	Ill, Taq	NCI-CGF	39	Taq	5	39	44
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	7	7q21.1	48,890	S102	132 (entire gene)	SM3	Ill, Taq	NCI-CGF	111	Ill	5	7	12
ESR1	estrogen receptor 1	6	6q25.1	325,721	S95	36	SM2	Ill, Seq	USC/Broad	351	Ill	101	415	516
ESR2	estrogen receptor 2	14	14q23.2	91,224	S190	8	SM2	Ill, Seq	USC/Broad	40	Taq	4	73	77
FSHB	follicle stimulating hormone, beta polypeptide	11	11p13	31,898	S95	6	SM2	Ill, Seq	USC/Broad	12	Taq	2	21	23
FSHR	follicle stimulating hormone receptor	2	2p21-p16	221,979	S95	12	SM2	Ill, Seq	USC/Broad	362	Ill	102	266	368
GNRH1	gonadotropin-releasing hormone 1	8	8p21-p11.2	35,144	S95	6	SM2	Ill, Seq	USC/Broad	17	Taq	2	15	17
GNRHR	gonadotropin-releasing hormone receptor	4	4q21.1	45,607	S95	5	SM2	Ill, Seq	USC/Broad	36	Taq	6	41	47
HSD17B1	hydroxysteroid (17-beta) dehydrogenase 1	17	17q11-q21	33,248	S190	11	SM2	Ill, Seq	USC/Broad	26	Taq	2	8	10
HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	16	16q24.1-q24.2	93,275	S95	10	SM2	Ill, Seq	USC/Broad	209	Taq	6	75	81
HSD17B3	hydroxysteroid (17-beta) dehydrogenase 3	9	9q22	108,398	S190	76	SM1	Ill, SNP, Taq	CEPH	97	Ill	27	86	113

Gene symbol	Gene name	Chr	Location	Target region [a]	Sample [b]	N identified polymorphisms	Sample[c]	Method[d]	Lab	N SNPs genotyped [e]	Technology[d]	N SNPs	N imputed SNPs	Total
HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	5	5q21	130,500	S190	96	SM1	Ill, SNP, Taq	CEPH	81	Ill	22	83	105
HSD3B1+HSD3B2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	1	1p13.1	42,772	S190	38	SM1	Ill, SNP, Taq	CEPH	28	Ill	14	135	149
INHA	inhibin, alpha	2	2q33-q36	33,474	S95	3	SM2	Ill, Seq	USC/Broad	24	Ill	5	1	6
INHBA	inhibin, beta A	7	7p15-p13	41,264	S95	1	SM2	Ill, Seq	USC/Broad	58	Ill	15	32	47
INHBB	inhibin, beta B	2	2cen-q13	35,976	S95	4	SM2	Ill, Seq	USC/Broad	31	Ill	7	1	8
LHCGR	luteinizing hormone/choriogonadotropin receptor	2	2p21	98,849	S95	21	SM2	Ill, Seq	USC/Broad	176	Ill	73	127	200
PGR	progesterone receptor	11	11q22-q23	122,153	S95	31	SM2	Ill, Seq	USC/Broad	122	Ill, Taq	18	135	153
POMC	proopiomelanocortin	2	2p23.3	50,849	S190	28	SM1	Ill, SNP, Taq	CEPH	38	Ill, Taq	8	23	31
PRL	prolactin	6	6p22.2-p21.3	39,729	S95	15	SM2	Ill, Seq	USC/Broad	62	Ill	29	70	99
PRLR	prolactin receptor	5	5p13.2	195,929	S95	15	SM2	Ill, Seq	USC/Broad	202	Ill	73	182	255
SHBG	sex hormone-binding globulin	17	17p13-p12	33,180	S95	10	SM2	Ill, Seq	USC/Broad	28	Ill	8	11	19
SRD5A1	steroid-5-alpha-reductase, alpha polypeptide 1	5	5p15	63,860			SM2	Ill, Seq	USC/Broad	93	Ill	19	78	97
SRD5A2	steroid-5-alpha-reductase, alpha polypeptide 2	2	2p23	86,385	S95	12	SM2	Ill, Seq	USC/Broad	54	Ill	9	45	54
UGT2B7/17	UDP glucuronosyltransferase 2 family, polypeptide B7	4	4q13	43,511	f)	f)	f)	f)	f)	f)	Ill	5	75	80
Sum				**2,856,332**		**646**				**2848**		**700**	**2671**	**3371**

a: Target region: 1st exon to last exon + 20kb up, 10kb down
b: Samples used for sequencing:
S95: 95 samples: 19 advanced breast cancer (African-Americans, Asians, Hawaiians, Caucasians, Latinas)
S190: 190 samples = (19 advanced breast cancer + 19 advanced prostate cancer) x (African-Americans, Asians, Hawaiians, Caucasians, Latinos)
S102 102 samples of SNP500cancer:31 Caucasians, 24 African/African American heritage, 23 Hispanic heritgae, 24 Pacific Rim heritage
c: Samples used for identification of tagging SNPs
SM1: 90 CEU (trios) + multiethnic panel from HGDP (119 W Africa, 166 E Asia, 75 Native Americans) + 58 Hawaii from MEC + 190 samples
SM2: multiethnic panel from MEC (70 African-Americans, 68 Latinos, 72 Japanese, 70 Caucasians, 69 Hawaiians)
SM3: 102 samples of SNP500cancer: 31 Caucasians, 24 African/African American heritage, 23 Hispanic heritgae, 24 Pacific Rim
d: Abbreviations for the methods used: Ill: Illumina, Taq: Taqman, SNP: SNPlex, Seq: Sequenom
e: SNPs found with resequencing were complemented with SNPs from dbSNP and HapMap.
f: Tag SNPs selection based exclusively on CEU HapMap data (phase II).
g: Labs for genotyping: EPIC: Imperial College(London); NHS: Harvard School of Public Health; MEC: USC

Table 2: Correlation between measured hormone levels and both age at blood donation (AGE) and body mass index (BMI) in the pooled sample of EPIC and NHS. Presented are the Pearson correlation coefficient with 95% confidence interval (95% CI) as well as the corresponding p-value.

		AGE				BMI			
Phenotype	Menopausal status	Correlation coefficient	95% CI		p-value	Correlation coefficient	95% CI		p-value
Δ4	post/pre	-0.41	-0.44	-0.38	$<10^{-32}$	-0.04	-0.07	0.00	0.027
DHEAS	post/pre	-0.36	-0.39	-0.33	$<10^{-32}$	-0.01	-0.05	0.02	0.416
SHBG	post/pre	-0.05	-0.08	-0.02	0.003	-0.39	-0.42	-0.37	$<10^{-32}$
TESTO	post/pre	-0.30	-0.33	-0.27	$<10^{-32}$	0.03	0.00	0.06	0.069
E1	post	-0.09	-0.13	-0.05	9.37×10^{-6}	0.27	0.23	0.31	0.000
E2	post	-0.09	-0.13	-0.06	8.31×10^{-7}	0.25	0.21	0.28	0.000

Table 3: Characteristics of associations of blood sex steroid hormone levels with SNPs in sex hormone metabolizing genes in postmenopausal women from MEC. Presented are estimates and corresponding standard deviations (SD). P-values are presented for the model with the minimum p-value.

						Heterozygote		Homozygote	
Hormone	**Gene**	**SNP**	**Sample**	**P-value**	**Model**	**Estimate**	**SD**	**Estimate**	**SD**
SHBG	**SHBG**	rs1619016	African-american	0.261	recessive	0.017	0.108	0.125	0.122
			Asian	0.109	dominant	-0.14	0.1	-0.182	0.14
			Latino	0.501	dominant	-0.091	0.125	-0.037	0.192
			Hawaiian	0.001	dominant	-0.355	0.106	-0.266	0.246
			Whites	0.021	dominant	-0.259	0.113	-0.186	0.216
			MEC all	0.002	dominant	-0.17	0.048	-0.054	0.069
		rs2955617	African-american	0.055	dominant	-0.376	0.194	-0.349	0.192
			Asian	0.034	recessive	0.034	0.139	-0.187	0.151
			Latino	0.214	recessive	0.056	0.132	0.219	0.168
			Hawaiian	0.093	dominant	-0.166	0.106	-0.182	0.188
			Whites	0.012	dominant	-0.255	0.106	-0.252	0.133
			MEC all	0.009	dominant	-0.125	0.054	-0.139	0.061
		rs9898876	African-american	0.263	dominant	0.126	0.098	-0.02	0.206
			Asian	0.51	recessive	0.051	0.096	-0.142	0.253
			Latino	0.136	trend	0.127	0.124	0.32	0.251
			Hawaiian	0.062	dominant	0.311	0.159	0.094	0.354
			Whites	0.079	recessive	-0.204	0.099	0.363	0.245
			MEC all	0.242	trend	0.04	0.05	0.114	0.116
		rs9913778	African-american	0.52	dominant	0.093	0.107	-0.122	0.249
			Asian	0.715	dom/tr/add	0.042	0.115	NA	NA
			Latino	0.126	recessive	-0.002	0.135	0.435	0.285
			Hawaiian	0.371	dom/tr/add	0.177	0.196	NA	NA
			Whites	0.164	recessive	-0.186	0.128	0.567	0.424
			MEC all	0.356	recessive	0.011	0.058	0.16	0.171
E1	**CYP19**	rs10046	African-american	0.009	recessive	-0.138	0.094	0.437	0.186
			Asian	0.355	trend	0.055	0.098	0.1	0.108
			Latino	0.02	dominant	0.3	0.133	0.312	0.156
			Hawaiian	0.011	recessive	0.089	0.106	0.437	0.161
			Whites	0.306	dominant	0.112	0.114	0.117	0.155
			MEC all	0.006	recessive	0.048	0.047	0.192	0.065
		rs4646	African-american	0.14	recessive	-0.025	0.096	-0.26	0.174
			Asian	0.008	recessive	-0.025	0.087	-0.397	0.146
			Latino	0.016	recessive	-0.111	0.113	-0.44	0.165
			Hawaiian	0.001	dominant	-0.378	0.112	-0.316	0.128
			Whites	0.279	recessive	0.039	0.1	-0.137	0.148
			MEC all	$p<10^{-3}$	trend	-0.084	0.046	-0.251	0.067
		rs6493494	African-american	0.29	dominant	0.111	0.095	0.007	0.194
			Asian	0.065	dominant	0.169	0.089	0.119	0.111
			Latino	0.5	dominant	0.076	0.12	0.083	0.167
			Hawaiian	0.215	trend	0.1	0.105	0.236	0.236
			Whites	0.637	dominant	-0.048	0.1	-0.028	0.196
			MEC all	0.051	dominant	0.09	0.046	0.064	0.077
		rs727479	African-american	0.005	trend	-0.211	0.094	-0.5	0.242
			Asian	0.002	recessive	0.002	0.085	-0.419	0.135
			Latino	0.001	trend	-0.151	0.114	-0.49	0.142
			Hawaiian	0.047	recessive	-0.052	0.111	-0.297	0.145
			Whites	0.101	trend	-0.121	0.105	-0.23	0.146
			MEC all	$p<10^{-3}$	trend	-0.101	0.045	-0.355	0.069

	rs749292	African-american	0.03	dominant	0.2	0.108	0.25	0.121
		Asian	0.365	dominant	0.105	0.09	0.015	0.11
		Latino	0.013	trend	0.236	0.128	0.401	0.157
		Hawaiian	0.016	dominant	0.265	0.108	0.193	0.156
		Whites	0.61	recessive	-0.005	0.103	0.084	0.184
		MEC all	$p<10^{-3}$	dominant	0.163	0.048	0.172	0.063
ESR1	rs1884053	African-american	0.169	dominant	-0.216	0.19	-0.258	0.18
		Asian	0.252	recessive	-0.003	0.097	-0.197	0.178
		Latino	0.383	recessive	-0.014	0.114	0.072	0.132
		Hawaiian	0.66	recessive	-0.021	0.146	0.034	0.141
		Whites	0.479	trend	0.052	0.105	0.167	0.262
		MEC all	0.695	trend	-0.013	0.053	-0.023	0.058
	rs2347871	African-american	0.341	dominant	-0.079	0.126	-0.083	0.127
		Asian	0.575	recessive	0.027	0.096	-0.045	0.134
		Latino	0.566	dominant	-0.097	0.137	-0.056	0.139
		Hawaiian	0.582	recessive	0.02	0.106	0.151	0.262
		Whites	0.252	recessive	-0.003	0.097	-0.197	0.178
		MEC all	0.491	dominant	-0.017	0.049	-0.047	0.061
	rs9341016	African-american	0.178	dom/trend/add	-0.381	0.282	NA	NA
		Asian	0.713	dom/trend/add	0.041	0.111	NA	NA
		Latino	0.902	dom/trend/add	-0.015	0.12	NA	NA
		Hawaiian	0.416	dom/trend/add	0.161	0.197	NA	NA
		Whites	0.546	dom/trend/add	-0.103	0.17	NA	NA
		MEC all	0.723	dom/trend/add	-0.025	0.071	NA	NA
FSHR	rs10495968	African-american	0.665	dominant	0.041	0.099	0.037	0.166
		Asian	0.568	trend	0.02	0.098	0.062	0.109
		Latino	0.193	recessive	-0.028	0.137	0.131	0.149
		Hawaiian	0.784	recessive	0.036	0.118	-0.013	0.145
		Whites	0.115	dominant	-0.162	0.097	-0.073	0.164
		MEC all	0.629	dominant	-0.032	0.048	0.003	0.062
	rs11125215	African-american	0.17	dominant	0.132	0.097	0.099	0.181
		Asian	0.568	trend	0.02	0.098	0.062	0.109
		Latino	0.195	recessive	0.017	0.133	0.164	0.148
		Hawaiian	0.711	recessive	0.053	0.115	-0.02	0.146
		Whites	0.159	dominant	-0.147	0.097	-0.059	0.163
		MEC all	0.652	trend	0.01	0.048	0.029	0.063
	rs1157876	African-american	0.776	dominant	0.025	0.099	0.031	0.166
		Asian	0.568	trend	0.02	0.098	0.062	0.109
		Latino	0.192	recessive	0.019	0.132	0.169	0.15
		Hawaiian	0.784	trend	0.013	0.117	0.042	0.147
		Whites	0.115	dominant	-0.147	0.096	-0.144	0.175
		MEC all	0.584	recessive	-0.034	0.048	0.014	0.063
	rs1394205	African-american	0.667	dominant	0.04	0.098	0.037	0.166
		Asian	0.568	trend	0.02	0.098	0.062	0.109
		Latino	0.193	recessive	0.014	0.134	0.16	0.146
		Hawaiian	0.784	recessive	0.017	0.119	-0.025	0.146
		Whites	0.146	dominant	-0.15	0.097	-0.065	0.164
		MEC all	0.67	dominant	-0.029	0.048	0.004	0.062
	rs4331540	African-american	0.665	dominant	0.041	0.099	0.037	0.166
		Asian	0.568	trend	0.02	0.098	0.062	0.109
		Latino	0.193	recessive	0.014	0.134	0.16	0.146
		Hawaiian	0.733	dominant	0.055	0.117	-0.003	0.144
		Whites	0.115	dominant	-0.162	0.097	-0.073	0.164
		MEC all	0.73	recessive	-0.022	0.048	0.008	0.062

		rs4971637	African-american	0.318	dominant	0.138	0.115	-0.425	0.485
			Asian	0.54	dominant	-0.051	0.091	-0.051	0.153
			Latino	0.41	recessive	-0.059	0.115	0.094	0.159
			Hawaiian	0.235	recessive	0.019	0.105	0.435	0.361
			Whites	0.689	recessive	0.01	0.101	-0.065	0.175
			MEC all	0.712	trend	0.015	0.048	0.024	0.095
		rs4971665	African-american	0.665	dominant	0.037	0.1	0.051	0.154
			Asian	0.693	dominant	0.041	0.099	0.029	0.108
			Latino	0.362	recessive	-0.01	0.138	0.097	0.148
			Hawaiian	0.611	recessive	-0.001	0.12	0.06	0.137
			Whites	0.115	dominant	-0.176	0.097	-0.025	0.155
			MEC all	0.536	recessive	-0.039	0.049	0.013	0.06
		rs4971884	African-american	0.099	recessive	0.135	0.107	-0.537	0.342
			Asian	0.725	recessive	0.009	0.1	0.037	0.11
			Latino	0.12	recessive	0.001	0.133	0.18	0.145
			Hawaiian	0.599	dominant	0.077	0.116	0.009	0.145
			Whites	0.322	trend	-0.068	0.099	-0.143	0.164
			MEC all	0.714	dominant	0.024	0.048	-0.002	0.063
		rs7606570	African-american	0.302	recessive	0.087	0.118	-0.047	0.119
			Asian	0.617	recessive	0.002	0.1	0.045	0.108
			Latino	0.257	trend	0.059	0.139	0.16	0.146
			Hawaiian	0.484	dominant	0.101	0.118	0.026	0.141
			Whites	0.232	trend	-0.095	0.098	-0.156	0.164
			MEC all	0.501	recessive	0.018	0.051	-0.023	0.057
E2	**CYP19**	rs10046	African-american	0.061	recessive	-0.097	0.09	0.295	0.179
			Asian	0.245	recessive	0.033	0.107	0.131	0.115
			Latino	0.105	dominant	0.238	0.15	0.224	0.176
			Hawaiian	0.015	trend	0.14	0.107	0.407	0.162
			Whites	0.423	trend	-0.075	0.125	-0.132	0.17
			MEC all	0.038	recessive	0.035	0.049	0.146	0.067
		rs6493494	African-american	0.474	dominant	0.077	0.09	-0.04	0.183
			Asian	0.292	dominant	0.104	0.098	0.078	0.123
			Latino	0.294	dominant	0.145	0.136	0.107	0.189
			Hawaiian	0.117	trend	0.112	0.104	0.333	0.234
			Whites	0.191	dominant	-0.149	0.109	-0.07	0.214
			MEC all	0.127	dominant	0.07	0.047	0.064	0.079
		rs727479	African-american	0.03	trend	-0.133	0.089	-0.449	0.231
			Asian	0.054	trend	-0.142	0.096	-0.241	0.154
			Latino	0.01	recessive	-0.055	0.133	-0.413	0.166
			Hawaiian	0.051	recessive	0.044	0.11	-0.251	0.148
			Whites	0.174	trend	0.098	0.117	0.22	0.163
			MEC all	0.002	recessive	-0.044	0.047	-0.233	0.072
		rs749292	African-american	0.03	dominant	0.188	0.102	0.234	0.114
			Asian	0.629	recessive	0.005	0.1	0.055	0.119
			Latino	0.027	trend	0.23	0.146	0.396	0.176
			Hawaiian	0.033	dominant	0.228	0.108	0.186	0.156
			Whites	0.565	dominant	-0.076	0.114	0.023	0.203
			MEC all	0.004	trend	0.12	0.05	0.171	0.064
	FSHR	rs10454135	African-american	0.277	dominant	-0.093	0.09	-0.099	0.172
			Asian	0.544	dominant	0.107	0.131	0.044	0.13
			Latino	0.142	recessive	-0.191	0.224	0.016	0.221
			Hawaiian	0.137	dominant	-0.217	0.128	-0.122	0.137
			Whites	0.337	dominant	-0.121	0.114	-0.05	0.154
			MEC all	0.101	dominant	-0.103	0.053	-0.043	0.059

Table 4: Results from logistic regression of breast cancer status on SNPs for genes, which were found to be associated with variation in hormone levels, adjusted for age at blood donation, BMI, batch, and, if required, for menopausal status.

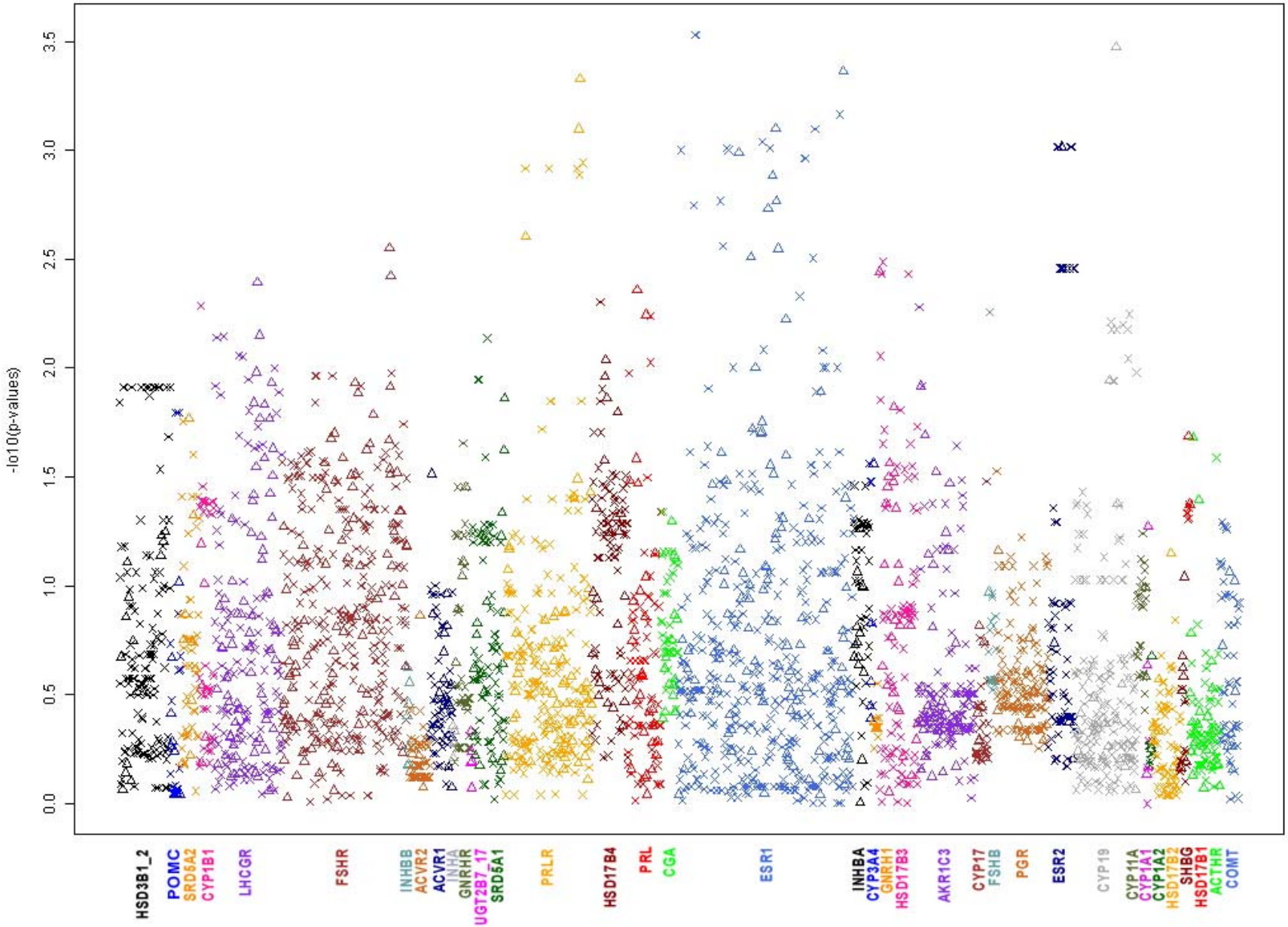
Sample	Gene	SNP	Trend	Add	Dominant	Recessive	Global[(a)]	Model for minimum p-value	OR (95% CI) for allele dosage
Pre- and postmenopausal									
	SHBG	rs1619016	0.95	0.94	0.98	0.74	1.00	Recessive	1.00 (0.86, 1.15)
		rs2955617	0.44	0.67	0.61	0.39	1.00	Recessive	0.96 (0.87, 1.06)
		rs9898876	0.42	0.56	0.59	0.30	1.00	Recessive	1.05 (0.93, 1.19)
		rs9913778	0.29	0.18	0.19	0.29	1.00	Additive	0.90 (0.75, 1.09)
	FSHR	rs12713034	0.62	0.86	0.58	0.82	1.00	Dominant	1.02 (0.93, 1.13)
		rs1290100	0.88	0.96	0.81	0.96	1.00	Dominant	0.99 (0.9, 1.1)
	AKR1C3	rs10752001	0.98	0.66	0.78	0.44	1.00	Recessive	1.00 (0.88, 1.14)
Postmenopausal									
	Cyp19	rs10046	0.83	0.91	0.96	0.70	1.00	Recessive	1.01 (0.90, 1.14)
		rs4646	0.28	0.34	0.53	0.15	1.00	Recessive	1.07 (0.94, 1.22)
		rs6493494	0.21	0.40	0.44	0.19	1.00	Recessive	0.93 (0.82, 1.04)
		rs727479	0.31	0.16	0.88	0.06	1.00	Recessive	0.94 (0.84, 1.00)
		rs749292	0.24	0.49	0.40	0.28	1.00	Trend	0.93 (0.83, 1.05)
	ESR1	rs1884053	0.18	0.40	0.25	0.30	1.00	Trend	1.08 (0.96, 1.22)
		rs2347871	0.23	0.48	0.27	0.40	1.00	Trend	1.08 (0.96, 1.21)
		rs9341016	0.73	0.93	0.75	0.79	1.00	Trend	1.04 (0.82, 1.32)
	FSHR	rs10495968	0.43	0.25	0.19	0.56	1.00	Dominant	1.05 (0.93, 1.19)
		rs11125215	0.11	0.05	0.03	0.62	1.00	Dominant	1.12 (0.98, 1.28)
		rs1157876	0.45	0.70	0.40	0.83	1.00	Dominant	1.05 (0.93, 1.19)
		rs1394205	0.49	0.40	0.27	0.66	1.00	Dominant	1.04 (0.92, 1.19)
		rs4331540	0.50	0.43	0.29	0.70	1.00	Dominant	1.04 (0.92, 1.18)
		rs4971637	0.09	0.20	0.08	0.56	1.00	Dominant	1.13 (0.98, 1.30)
		rs4971665	0.21	0.25	0.11	0.94	1.00	Dominant	1.08 (0.96, 1.22)
		rs4971884	0.09	0.11	0.04	0.94	1.00	Dominant	1.13 (0.98, 1.29)
		rs7606570	0.09	0.10	0.04	0.96	1.00	Dominant	1.12 (0.98, 1.28)
		rs10454135	0.14	0.02	0.01	0.70	1.00	Dominant	1.09 (0.97, 1.22)

a: adjusted for multiple testing and correlation as well as for the number of genes tested, here 6.

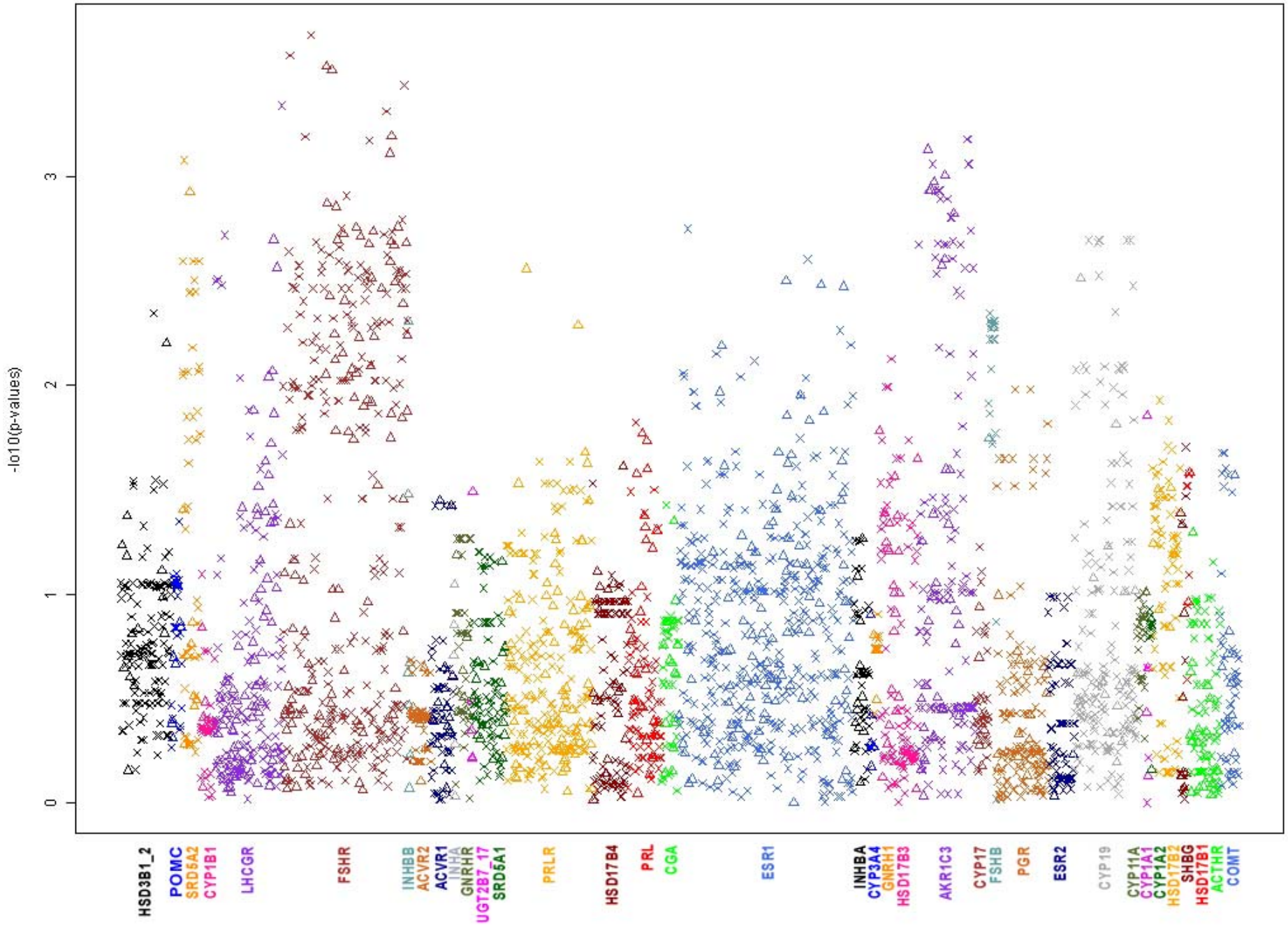
Figure 1: Plot of the $-\log_{10}$ of the unadjusted p-values from linear regression of the natural logarithm of the sex steroid hormone levels on 700 measured SNPs and 2,671 imputed SNPs. Genes are ordered by chromosomal position. Different colors denote different genes.

Δ: measured SNPs, x: imputed SNPs

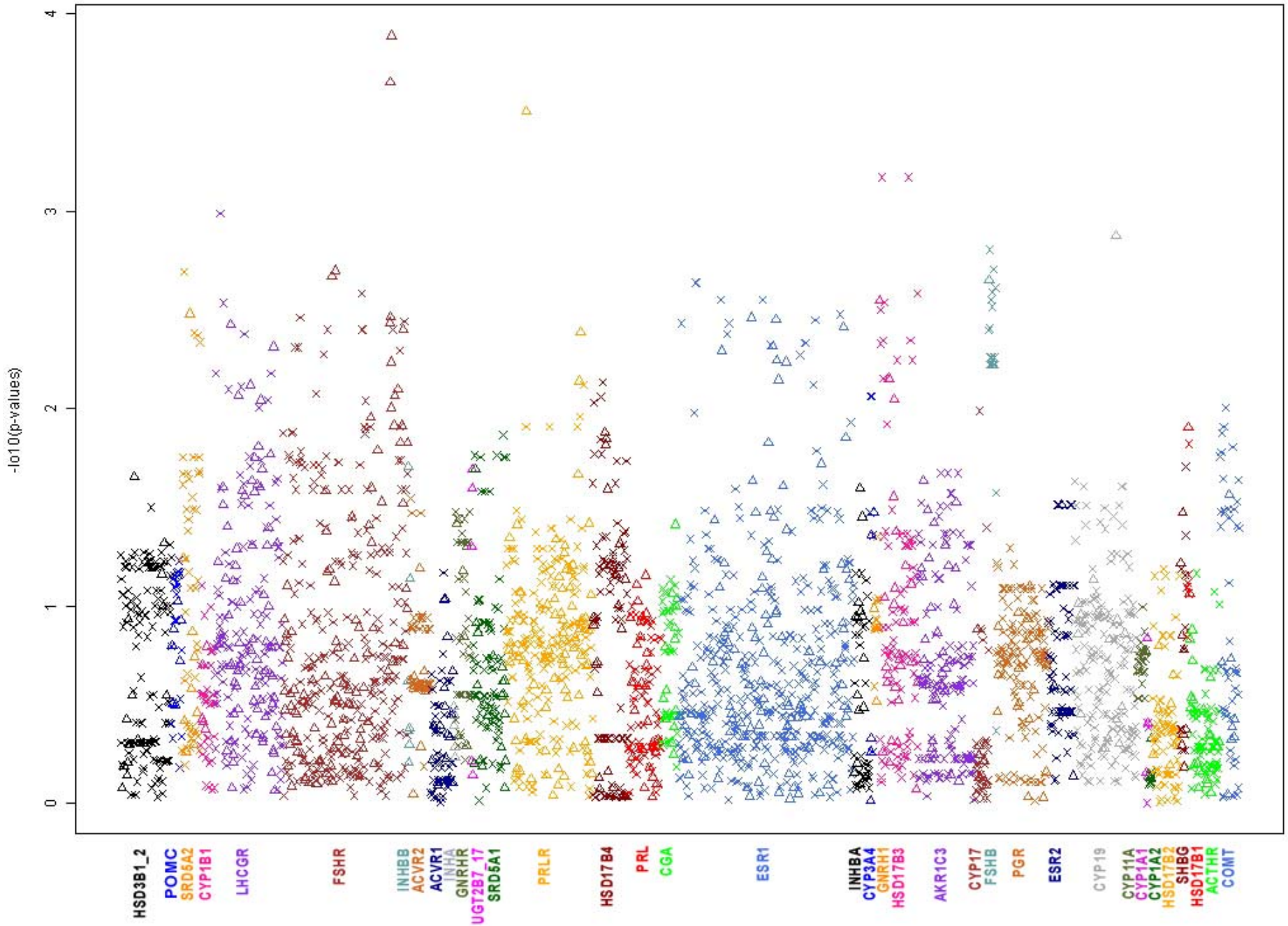
a) Δ4 in pre- and postmenopausal women



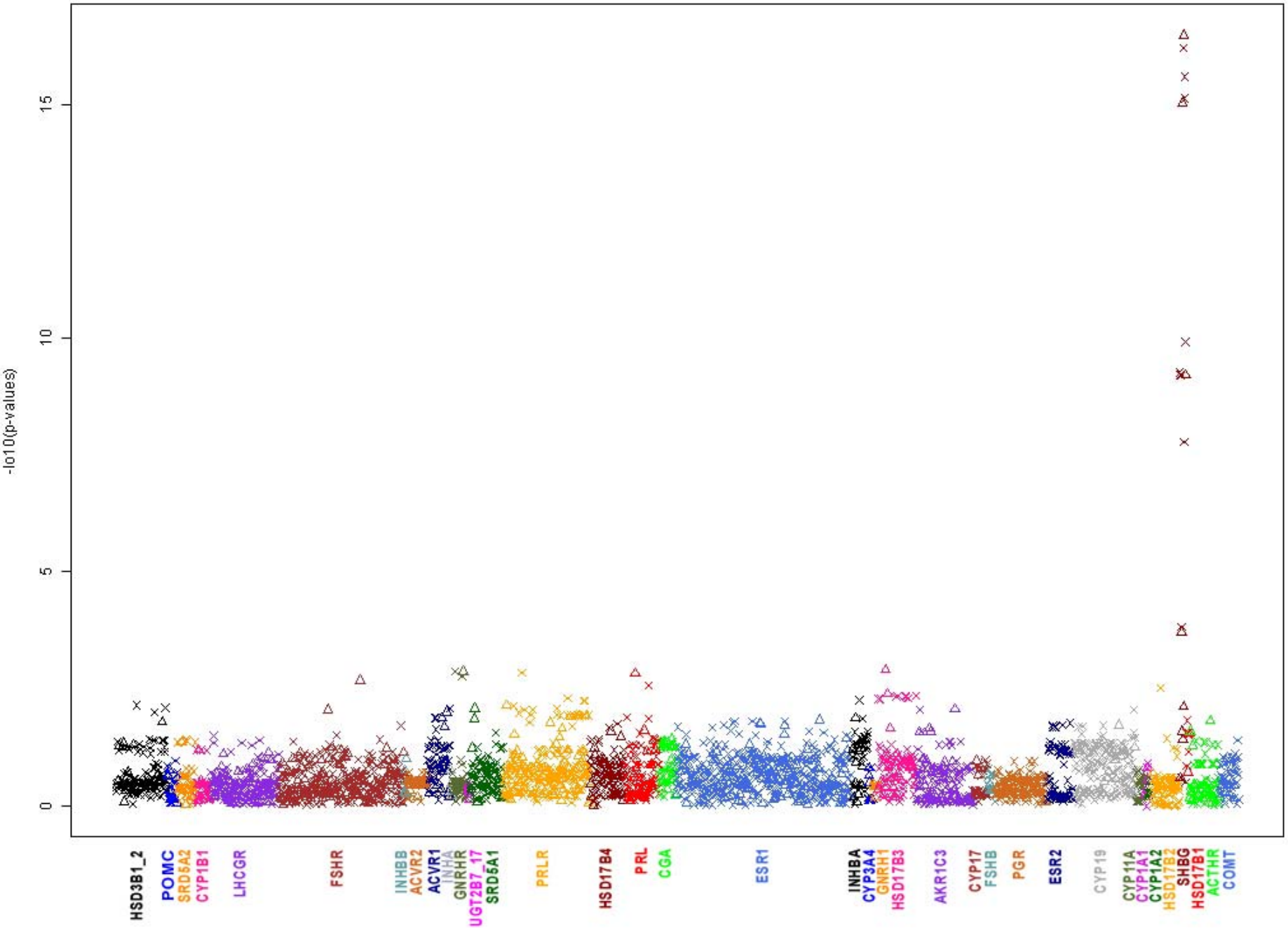
b) DHEAS in pre- and postmenopausal women



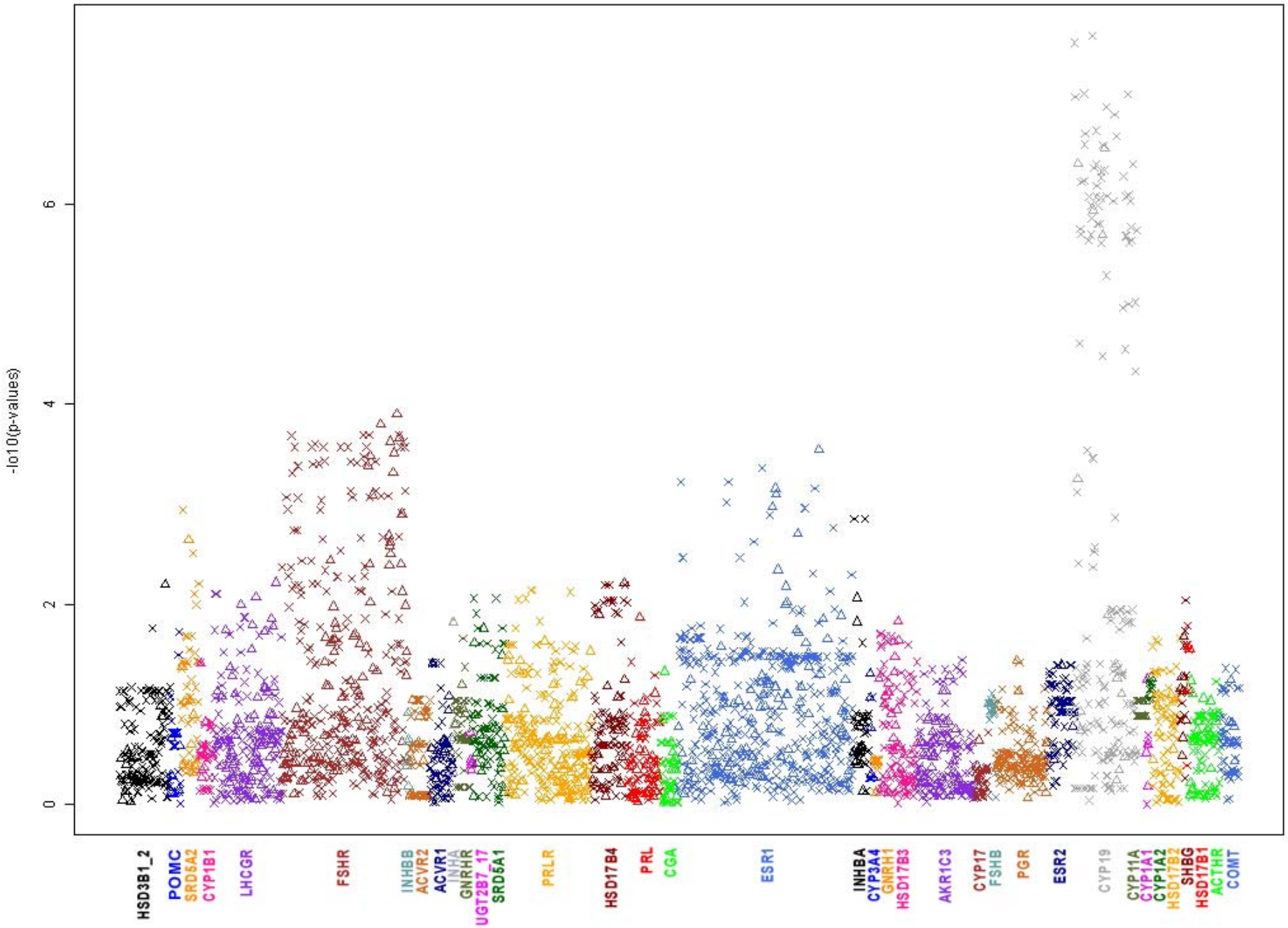
c) TESTO in pre- and postmenopausal women



d) SHBG in pre- and postmenopausal women



e) E1 in postmenopausal women



f) E2 in postmenopausal women

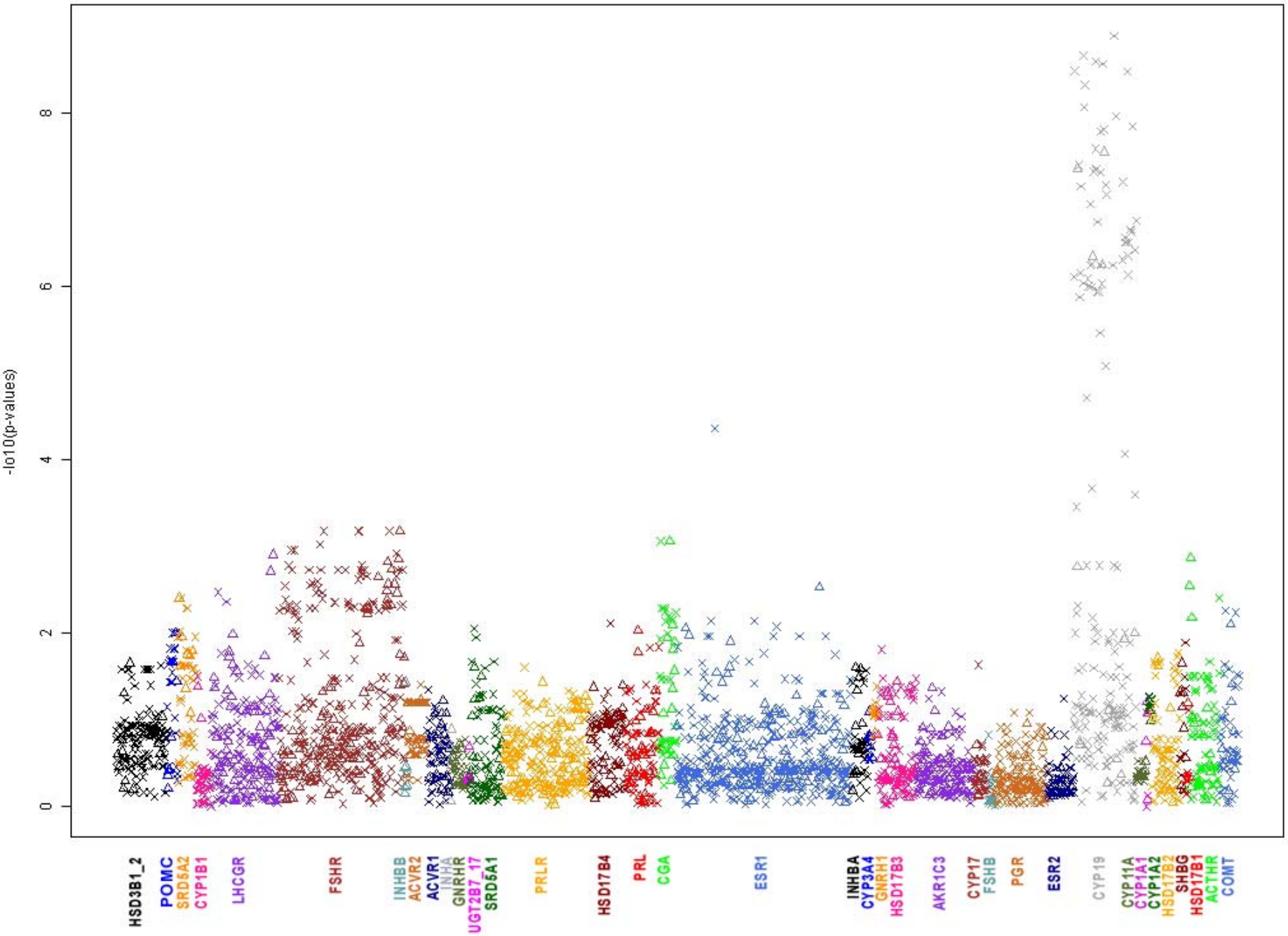
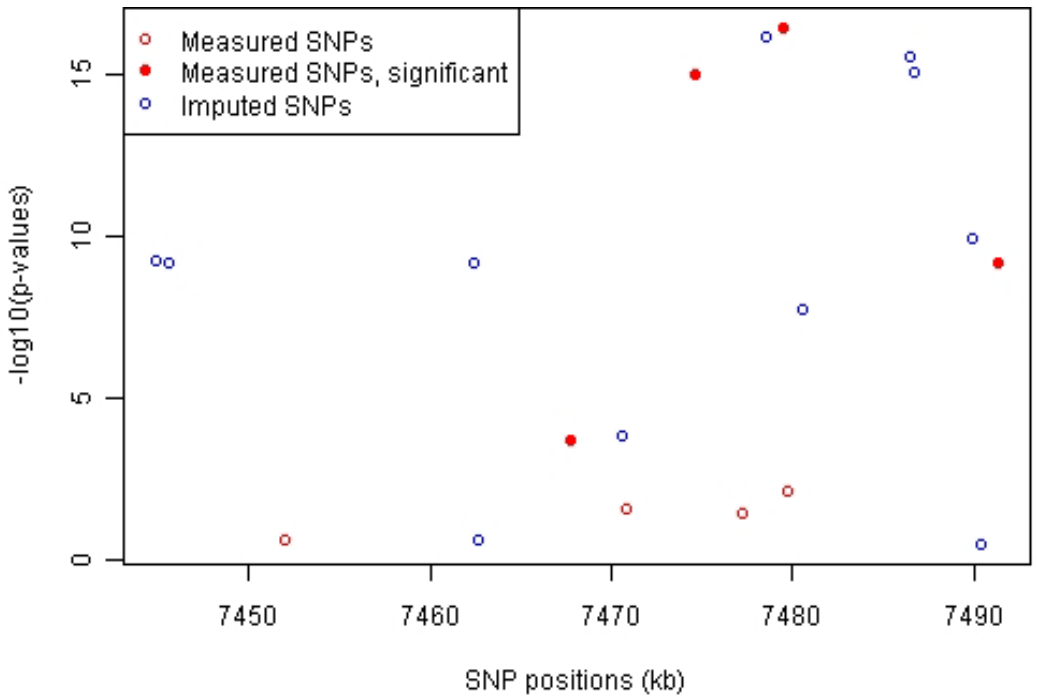
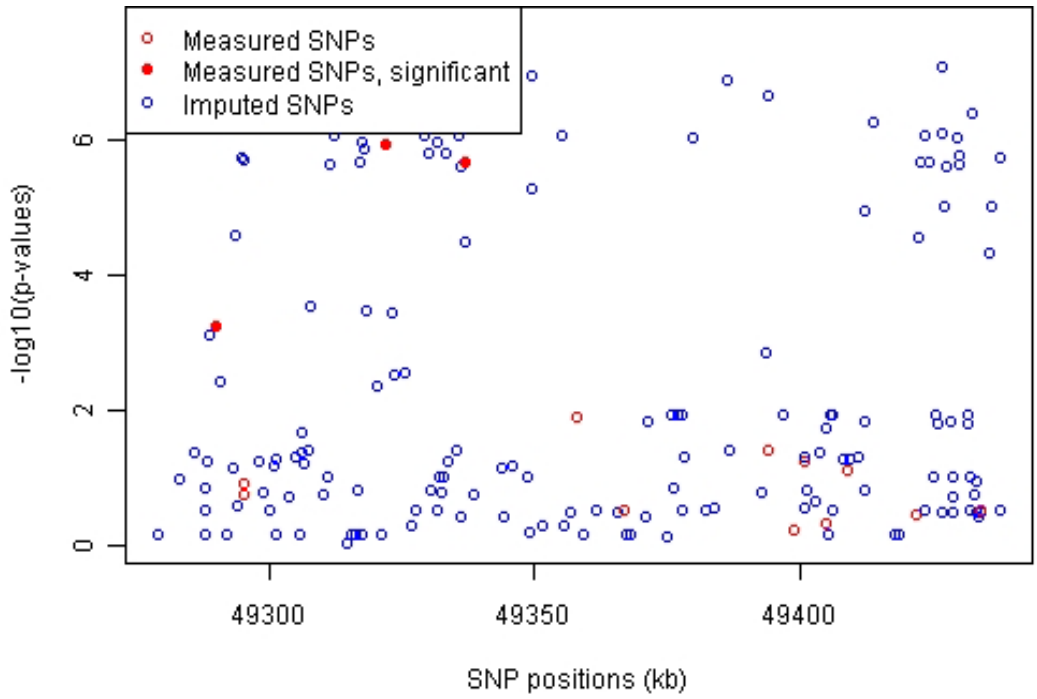


Figure 2: : Plot of the $-\log_{10}$ of the unadjusted minimum p-values from linear regression of the natural logarithm of the sex steroid hormone levels on measured and imputed SNPs for selected genes.

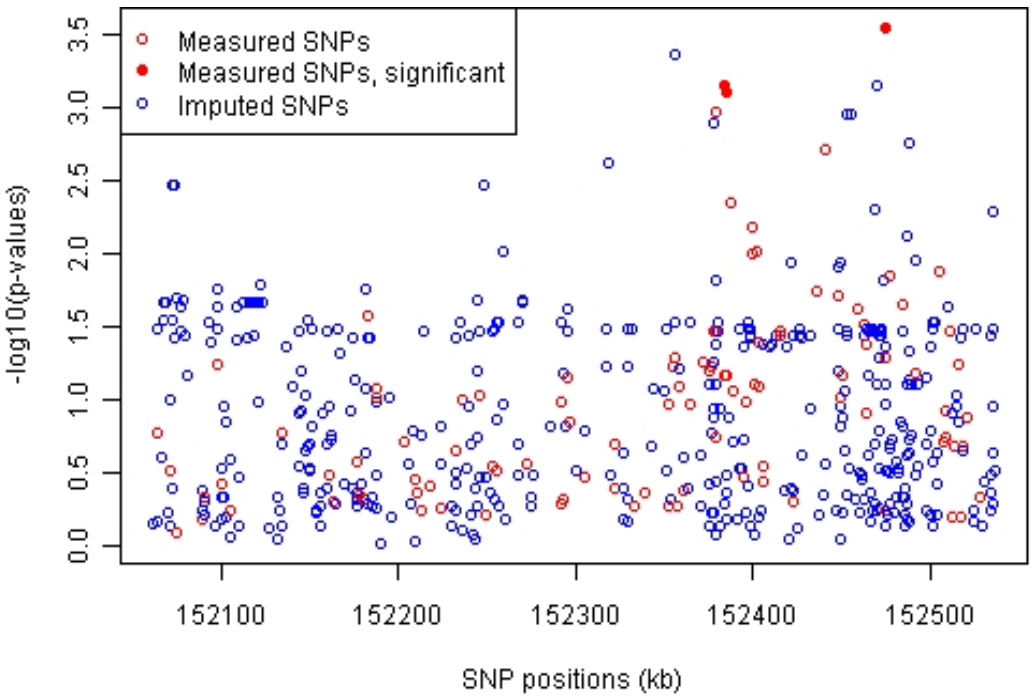
a) SHBG and *SHBG*



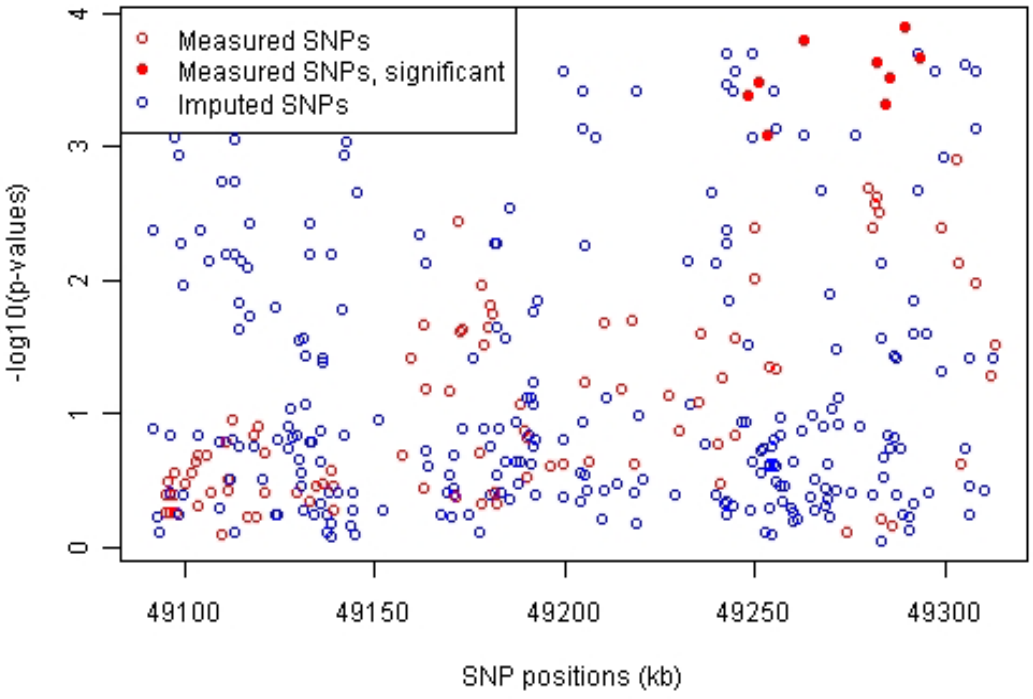
b) E1 and *CYP19*



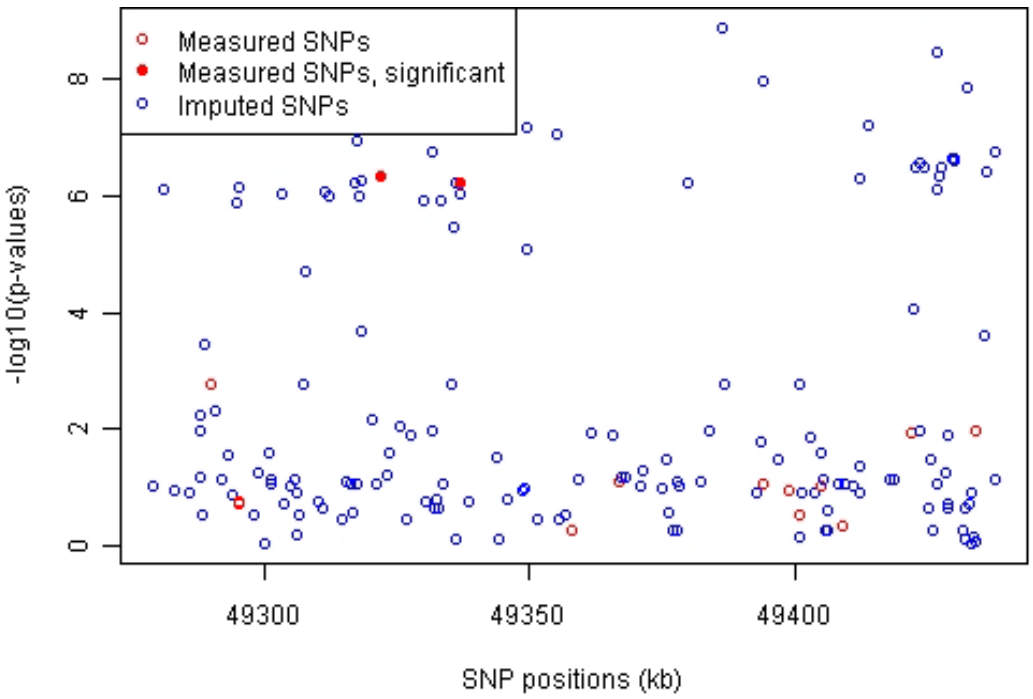
c) E1 and *ESR1*



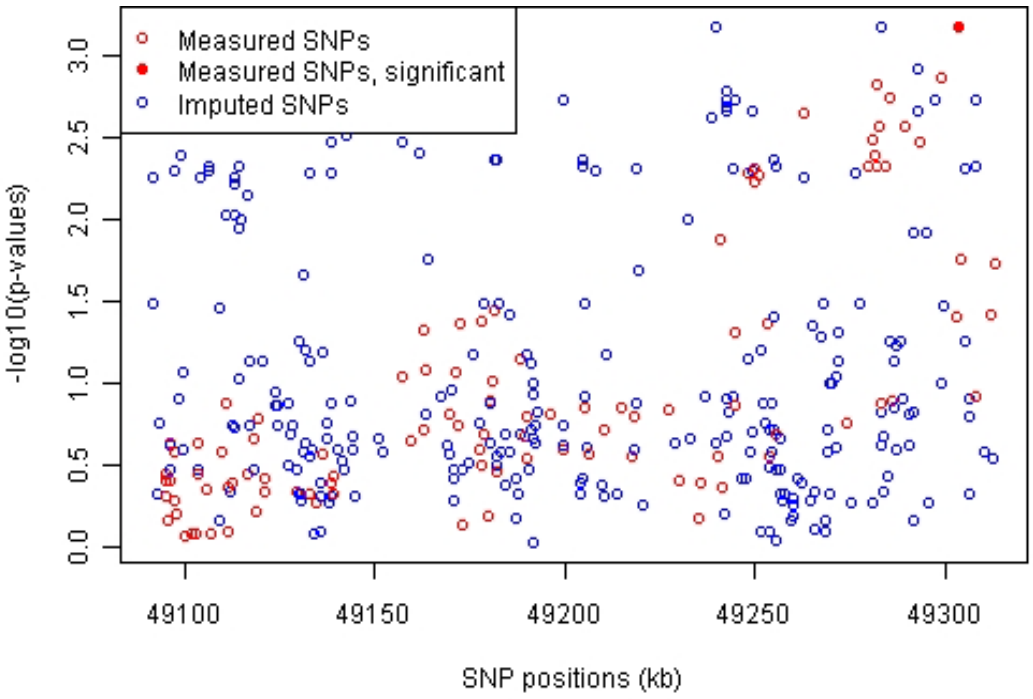
d) E1 and *FSHR*



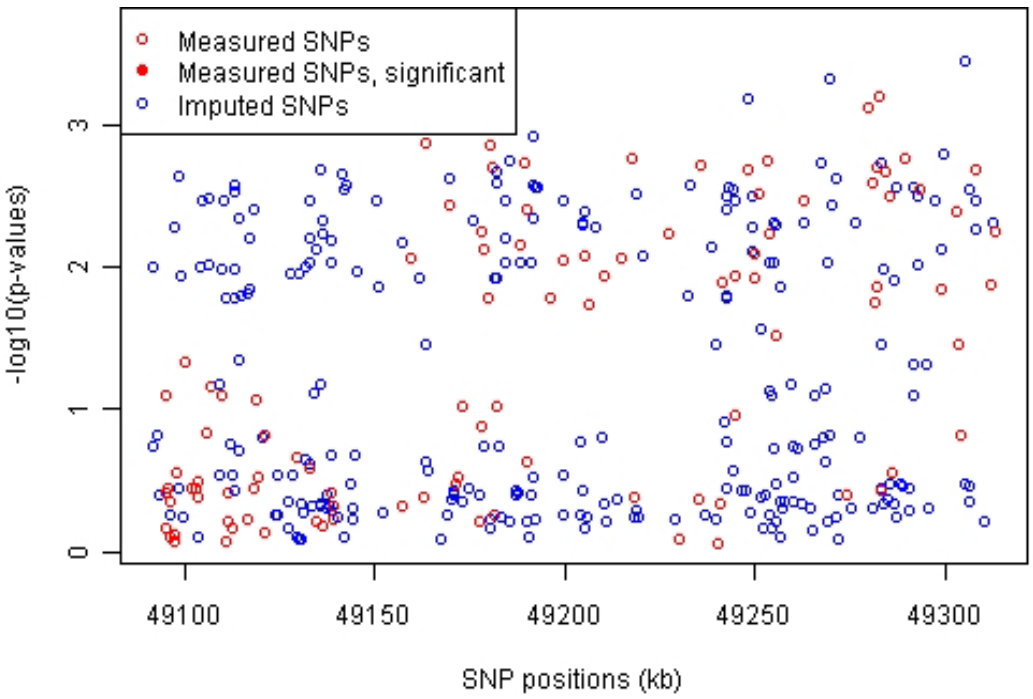
e) E2 and *CYP19*



f) E2 and *FSHR*



g) DHEAS and *FSHR*



h) DHEAS and *AKR1C3*

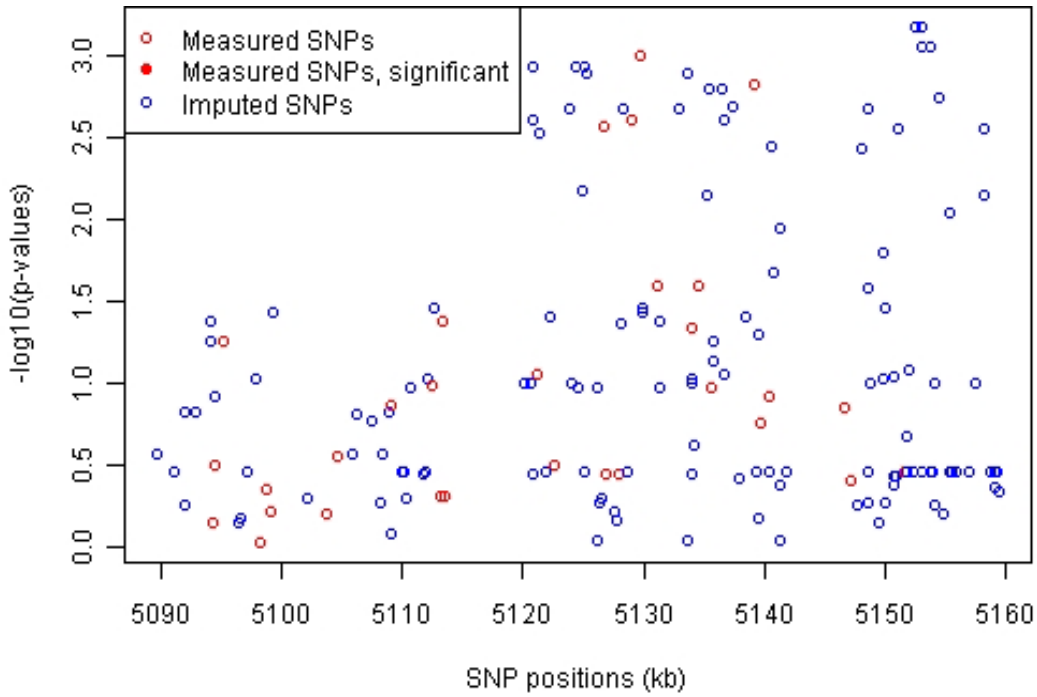
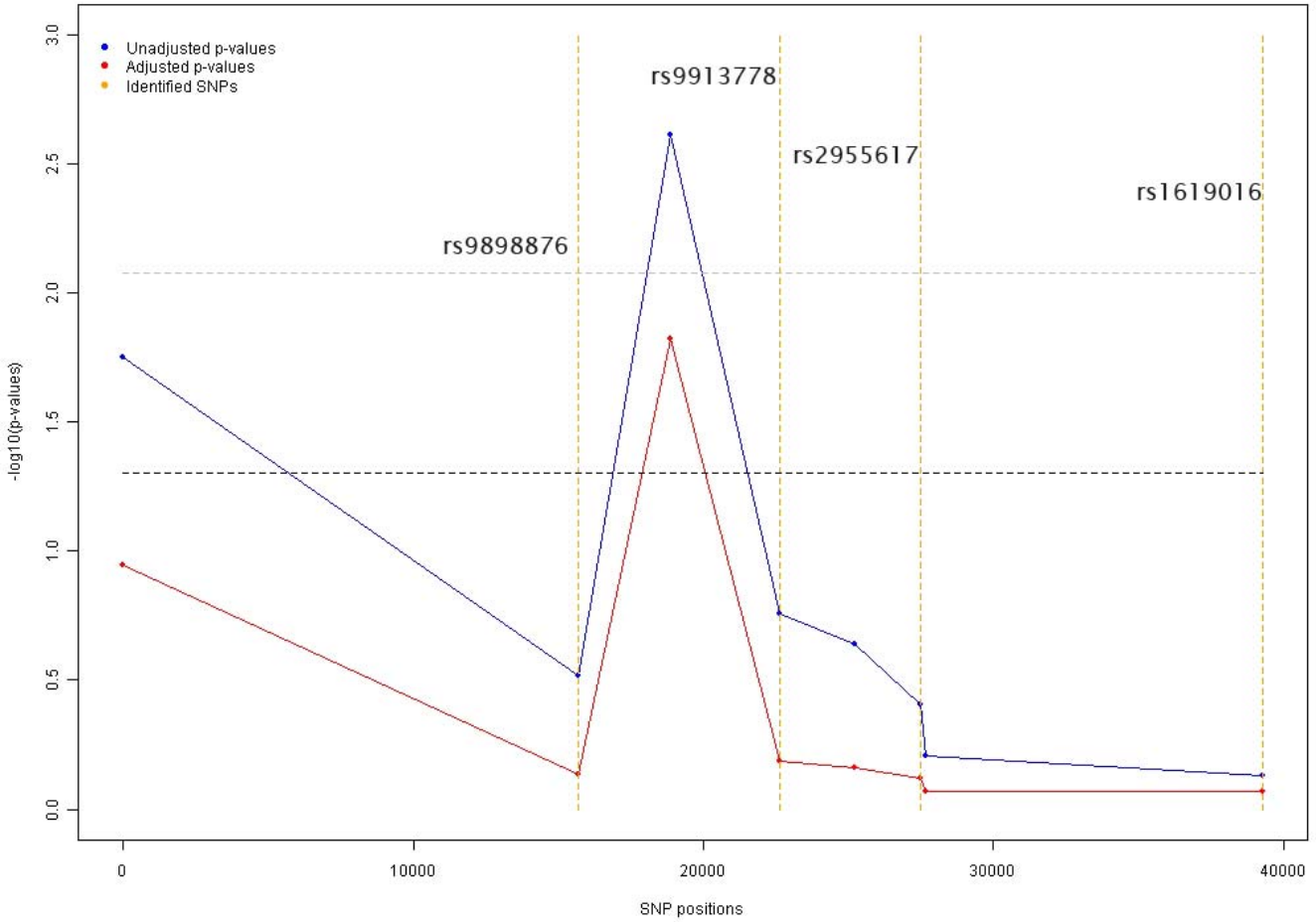
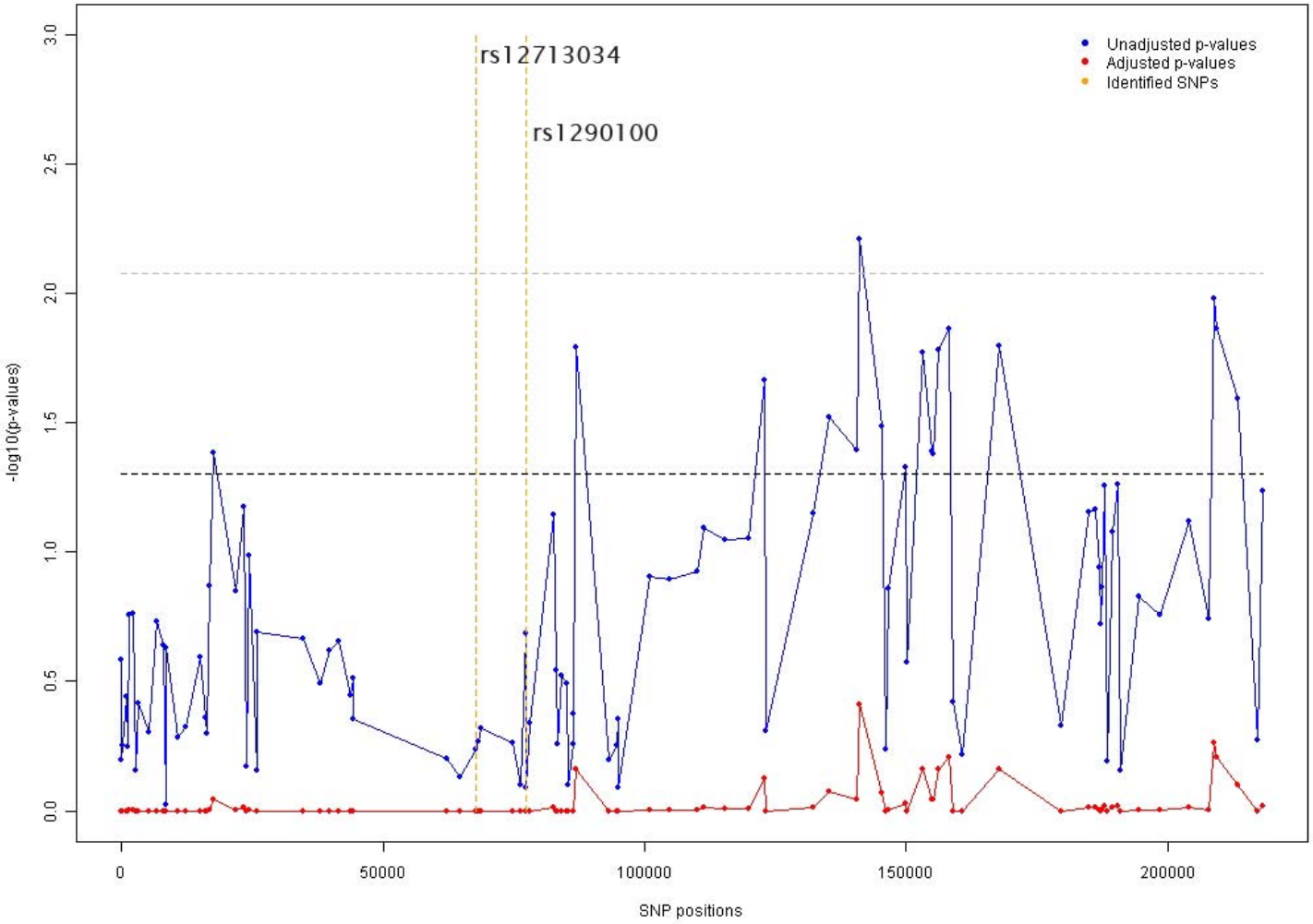


Figure 3: Plot of the $-\log_{10}$ of the unadjusted and adjusted minimum p-values from logistic regression of breast cancer status on SNPs for genes, which were found to be associated with variation in hormone levels.

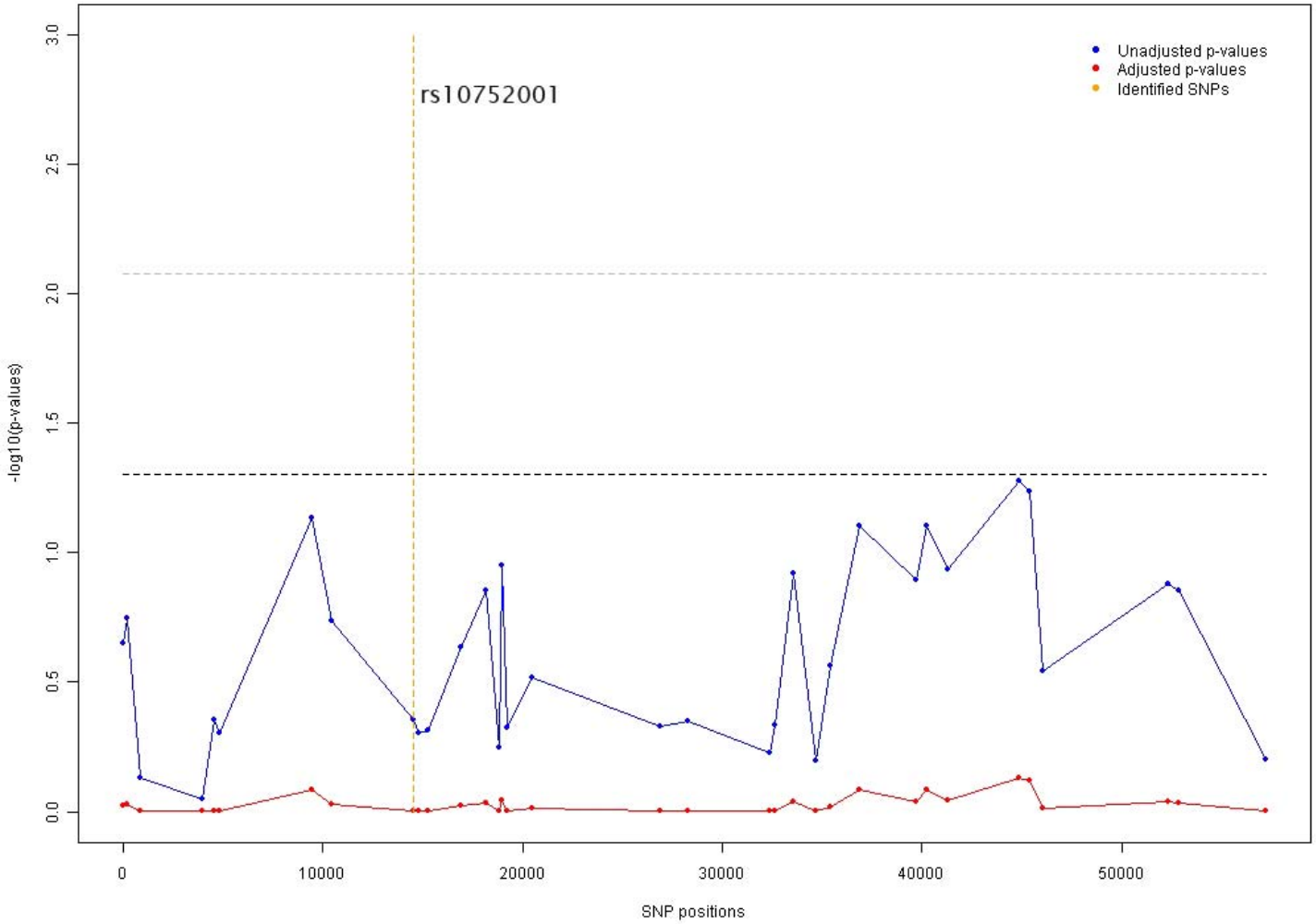
a) Pre- and postmenopausal women and the *SHBG* gene.



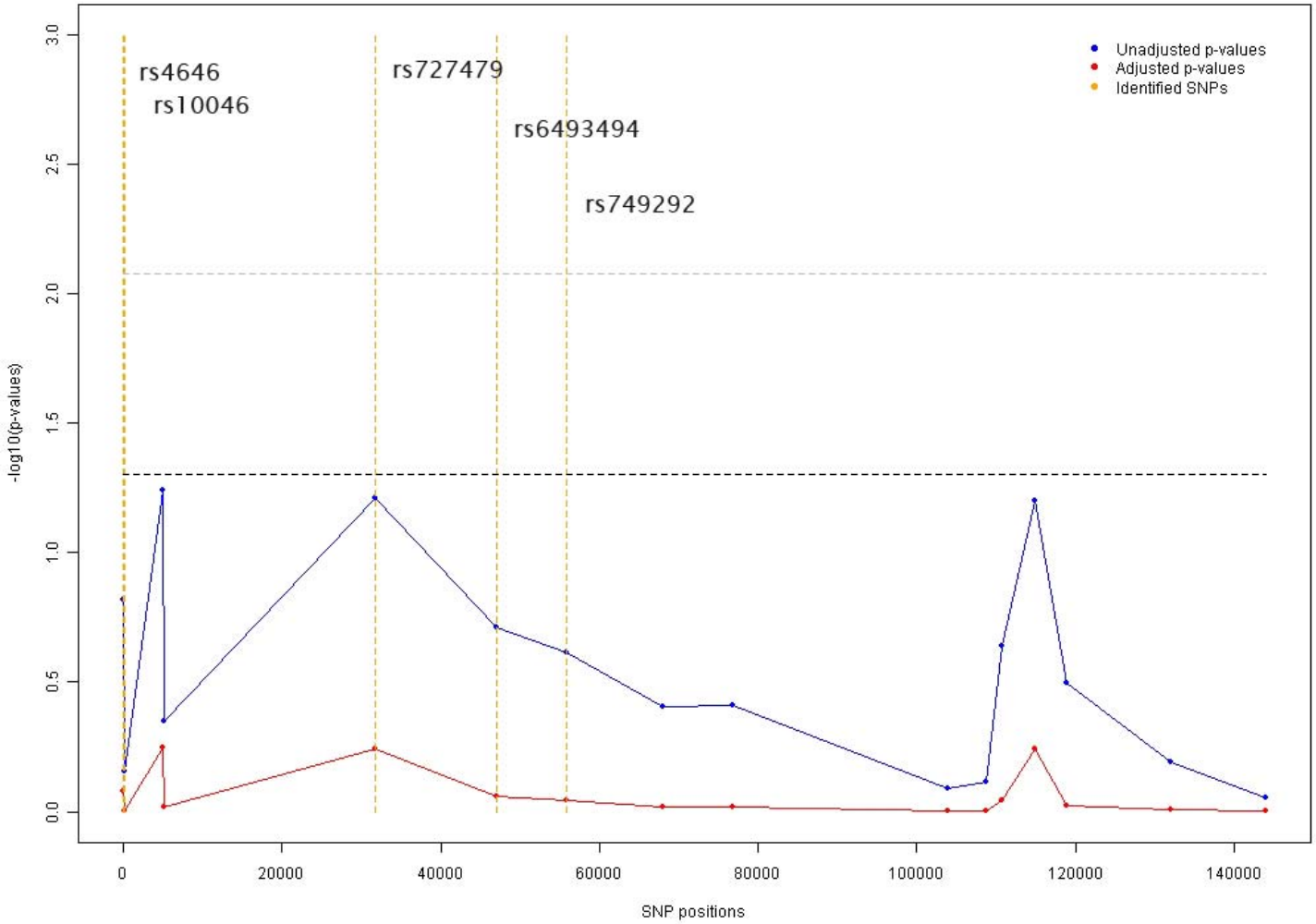
b) Pre- and postmenopausal women and FSHR



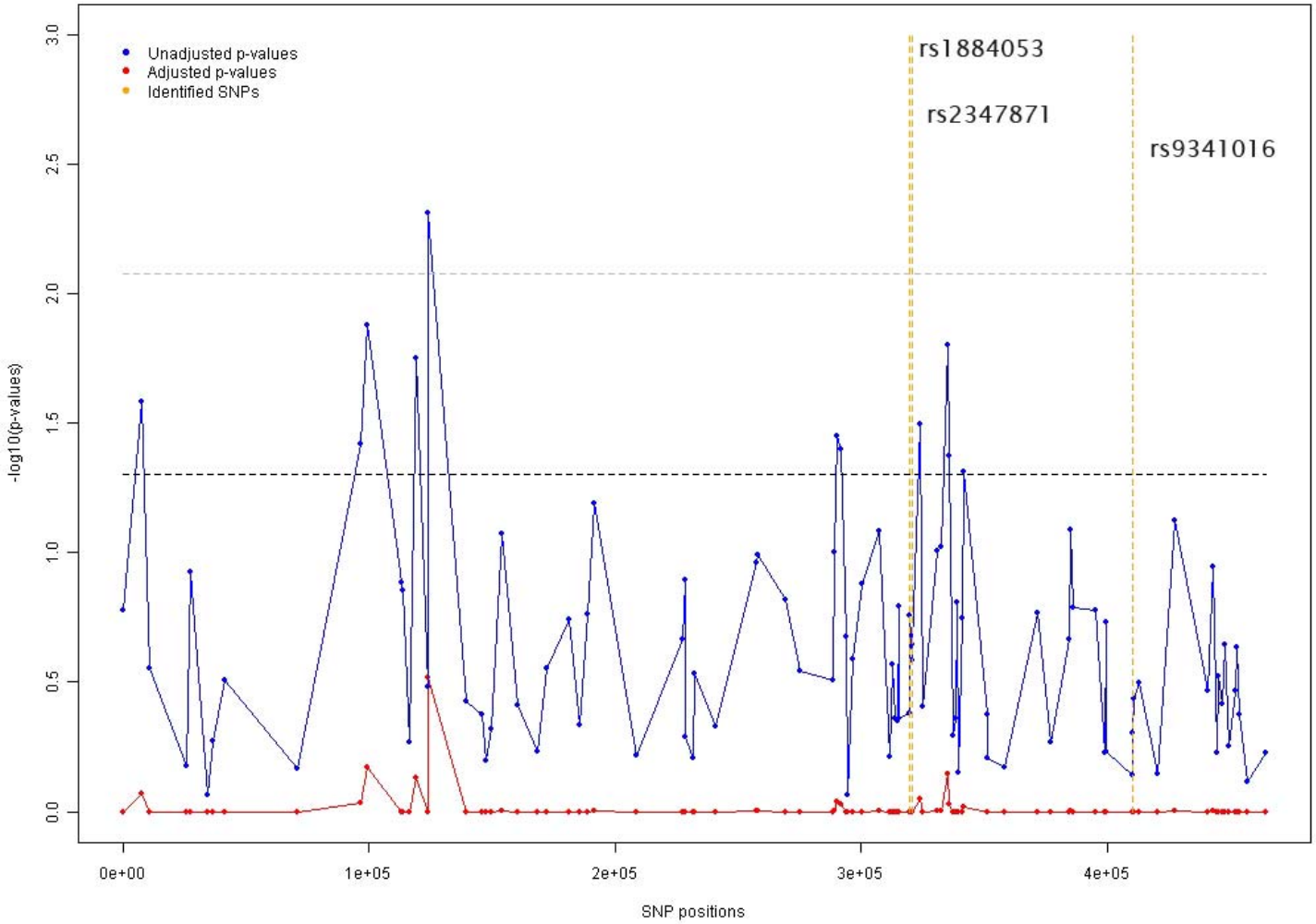
c) Pre- an d postmenopausal women and *AKR1C3*.



d) Postmenopausal women and *CYP19*.



e) Postmenopausal women and *ESR1*.



f) Postmenopausal women and *FSHR*.