Published in final edited form as:

Methods Mol Biol. 2013 ; 1049: 53–64. doi:10.1007/978-1-62703-547-7_5.

Survival prediction based on inherited gene variation analysis

Mine S. Cicek, Ph.D.,

Department of Health Sciences Research; Mayo Clinic College of Medicine; 200 First Street SW; Rochester, MN 55905, USA; 507/266-5196; cicek.mine@mayo.edu

Matthew J. Maurer, and

Department of Health Sciences Research; Mayo Clinic College of Medicine; 200 First Street SW; Rochester, MN 55905, USA; 507/266-2064; maurer.matthew@mayo.edu

Ellen L. Goode, Ph.D., M.P.H.

Department of Health Sciences Research; Mayo Clinic College of Medicine; 200 First Street SW; Rochester, MN 55905, USA; 507/266-7997; egoode@mayo.edu

SUMMARY

There is significant variation of outcome among ovarian cases. Clinical features such as age, stage, co-morbidities or degree of de-bulking are known prognostics factors for the disease. However, additional variation remains unexplained, some of which may be due to inherited factors. Here, we describe identification of survival-associated inherited variants in ovarian cancer that can enhance our current prognostic capabilities.

Keywords

Survival prediction; inherited variation; tagSNP; candidate genes; pathway analysis; ovarian cancer related pathways

1. INTRODUCTION

Ovarian cancer is the most lethal of gynecologic malignancies in the United States (1), yet much remains unknown about reasons for the observed differences in outcome across patients. Paclitaxel (Taxol) and platinum-based chemotherapy (either cisplatin or carboplatin) represent standard therapy after surgical debulking (2). Unfortunately, even with modern chemotherapy, most patients with advanced disease relapse and die of ovarian cancer; ten-year overall survival remains around 30% (3–5). Clinical features such as age, stage, co-morbidities, or degree of debulking are known prognostic factors (6). However, additional variation in outcome among patients remains unexplained, some of which may be due to inherited factors (germline, as opposed to tumor or somatic changes) that vary among patients. Inherited variation may influence outcome through differential metabolism of chemotherapeutic agents (e.g., variants in *ERCC1* and *ABCBI* modifying chemo-response), variable immune response, or through a variety of other host-related factors. For example, inherited *BRCA1* and *BRCA2* mutations are known to correlate with ovarian cancer survival (7).

To identify survival-associated variants, one can examine genome-wide or candidate genes. As an example of a candidate pathway approach, we hypothesized that inherited variants in mechanistic ovarian cancer-related pathways may contribute to genetically determined differences in ovarian cancer outcome. In particular, variants related to tumor angiogenesis may influence outcome as a major contributor to pathogenesis; anti-angiogenic therapy with bevacizumab represents the most successful application of a biologically-targeted ovarian cancer therapy (8). Aberrant inflammatory processes can impact multiple tumorigenic pathways (9). Detoxifying molecules are key in response to oxidative stress (10) and the processing of chemotherapeutic agents. Alterations of cell cycle regulators cooperate in ovarian tumor development, and tumor molecular analyses show frequent mutation of these regulators (11). Host defense mechanisms that eliminate mutated cells through permanent growth arrest (cellular senescence) are considered crucial tumor suppressive mechanisms (12). Apoptotic pathways have been shown to be frequently altered in both tumor progression and drug resistance (13). Glycosylation (14–16), one carbon transfer (17), and other processes may also be susceptible to inherited changes which lead to differential ovarian cancer outcome.

A unique aspect of our approach was the clinical, demographic, and treatment homogeneity of the patients studied which, when combined with high data completeness, allowed us to assess the role of inherited variants above and beyond known clinical factors. Use of a clinic-based population provided the ability to account for detailed clinical data, which are not routinely available in population-based studies, while restriction to patients living within a defined surrounding geographic region minimized the potential for the referral bias sometimes seen in clinic-based studies. Centralized pathology review, high-quality genotyping, covariate adjustment, and conservative interpretation of results are also strengths of this study. The downside of an observational study, is that compared to a randomized controlled trial, follow-up (e.g., measurement of toxicities) is less standardized. Our comprehensive efforts in selecting ovarian cancer related polymorphisms, genes and pathways were critical in this work. We used bioinformatics tools to help us choose genes included in this study in a systematic way.

Identifying survival-associated inherited variants can help to elucidate the most important drivers of ovarian cancer, representing ideal targets for tailored therapeutics, and may also enhance our current prognostic capabilities. Causality is not necessarily established by a correlation between a set of genes and clinical endpoints. In addition, a variety of mechanisms regulate gene expression including DNA methylation, histone deacetylation, copy-number changes and targeting by micro-RNAs. Therefore, an integration procedure that incorporates this biologically useful knowledge is highly desirable.

2. MATERIALS

2.1. Participants

Recruitment of patients from Mayo Clinic's gynecologic surgery and medical oncology departments, including administration of questionnaires and venipuncture, used established protocols (18). The protocol was approved by the Institutional Review Board, and all participants gave written informed consent. Eligible patients were women aged 20 years or

older living in MN, IA, WI, IL, ND, or SD (the upper Midwest, USA) and ascertained within one year of a diagnosis of pathologically-confirmed primary invasive epithelial ovarian cancer. Three hundred and forty-four invasive patients recruited between December 14, 1999 and March 19, 2006 and diagnosed within the preceding year (median time from recruitment to diagnosis = 4 days) were studied. Prior to analysis, additional pathologic review excluded nine patients found to be of non-epithelial or non-ovarian origin, resulting in 325 analyzed patients. Data on clinical features of disease including histology, surgical outcomes, and chemotherapy were abstracted by an experienced research nurse with review by gynecologic and medical oncology clinicians. Epidemiologic risk factor data including family history, reproductive factors, and exposure to known and suspected ovarian cancer risk factors were obtained through administration of a 16-page written questionnaire. Data on vital status was obtained from several sources including the Mayo Clinic computerized medical record, the Mayo Clinic Tumor Registry, which follows patients annually for overall survival who were diagnosed or received initial treatment at Mayo Clinic, and the National Death Index. Follow-up data was obtained through July 31, 2008.

2.2. Gene selection

With the goal of comprehensiveness, key genes within the pathways of angiogenesis (N=18 genes), apoptosis (N=18), cell cycle (N=39), cellular senescence (N=33), detoxification (N=15), glycosylation (N=26), inflammation (N=26), and one-carbon transfer (N=21) were identified in October 2005 using a number of sources including peer-reviewed published literature, the Cancer Genome Anatomy Project, Biokarta, and KEGG pathway databases. In addition, we included 25 genes previously shown to have a 50% or more reduction in expression in ovarian tumor versus normal tissue using the Affymetrix U133A GeneChip array (19). Finally, based on unpublished data and recent reports, a small number of genes related to NF- κ B (*NFKBIA*, *NFKBIB*) and hormone metabolism (*CYP19A1*, *SULT1E1*) as well as the granulatin-epithelin precursor (*GRN*) (20), and transcription elongation factor A (SII)-like 7 (*TCEAL7*) (21) were included. The complete list of 227 included genes is provided in Table 1.

2.3. Single nucleotide polymorphism (SNP) selection

SNP selection strategies varied slightly across pathways due to logistical issues related to panel development. Generally, for all pathways except glycosylation and inflammation, tagSNPs were identified via analysis of unrelated Caucasian genotypes from the Hapmap Consortium's release 21a (genome build 34) (22), Perlegen Sciences (genome build 34) (23), SeattleSNPs (<http://pga.mbt.washington.edu>, genome build 35) and Panel 2 of the National Institute for Environmental Health Science SNPs (www.egp.gs.washington.edu, genome build 35). SNPs for each gene were chosen from each source from the region within 5 kb of the most 5' and 3' exons of each gene. tagSNPs were identified by applying the ldSelect algorithm (24) to group (i.e., bin) SNPs with minor allele frequency (MAF) ≥ 0.05 and pair-wise linkage disequilibrium (LD) threshold of $r^2 \geq 0.8$. Following binning, we selected tagSNPs for analysis from the source with the greatest number of SNPs with MAF ≥ 0.05 and the greatest number of LD bins that also met criteria for predicted likelihood of successful genotyping using the Illumina Golden Gate Assay^{TM} quality score metrics. In addition to the tagSNPs, we included all additional putative-functional SNPs (within 1 kb

upstream, 5' UTR, 3' UTR or non-synonymous) with MAF ≥ 0.05 identified in Ensembl version 34. For *CDC25B*, *CUL1*, *E2F3*, and *SMARCA2*, SNPs tagging at least one other SNP were selected using pairwise binning in Tagger for SNPs in HapMap with Caucasian MAF ≥ 0.05 , $r^2 \geq 0.8$, within 10 kb of genes (25). For glycosylation genes, a tagSNP approach was not used; rather all non-intronic SNPs within 10 kb of each gene with a MAF > 0.05 were attempted. For inflammation genes, SNPs within 10 kb of each gene with MAF > 0.05 among the CEU samples of Hapmap Consortium's release 21a (genome build 34) (22) were binned if $r^2 \geq 0.8$, and all non-synonymous SNPs were included. For *CYP11A1* variants were selected using resequencing data among 60 Caucasian-American subjects (26) via both the tagSNP method (MAF ≥ 0.05 , $r^2 \geq 0.8$) (24) and a haplotype-tagging method (MAF ≥ 0.02 , haplotype frequency ≥ 0.01 , $r^2 \geq 0.9$). Finally, each SNP was evaluated for compatibility with Illumina Golden Gate Assay™ ensuring all designability scores ≥ 0.6 and no SNP pairs within 60 bp. A total of 1,536 SNPs were selected.

3. METHODS

3.1. Genotyping

Genotyping was performed using the Illumina GoldenGate™ BeadArray assay. DNA was extracted from 10 to 15 mL fresh peripheral blood using the Gentra AutoPure LS Purgene salting out methodology (Gentra, Minneapolis, MN) and stored at 4°C. Samples were bar-coded to ensure accurate and reliable processing and plated with duplicates and lab standards. Following DNA activation, incubation, amplification, and automated genotype calling, quality control exclusions were performed based on poor clustering, low call rates ($<95\%$) and duplicate samples with unusually high number of discrepancies. No samples were excluded from cases eligible for the current analysis. Of the 1,536 attempted SNPs, we excluded SNPs which failed completely (or nearly completely), had low call rates ($<95\%$), and were monomorphic among all samples. Of the remaining 1,479 SNPs, prior to the analysis, additional review excluded tagSNPs found to be incorrect due to a chromosome 7 misalignment (*ABCB1*, *CDK6*, *PPP1R3A*, *NOS3*), SNPs found to be > 10 kb from the gene of interest in current genome build 36.3 (*GSTM1* rs366631; *TIMP1* rs2283736; *MGAT5* rs1257189, rs1257187; *AKT1* rs4983559; *TFDP2* rs733996, rs7644383, rs11720300), and SNPs with observed MAF < 0.01 in the current sample subset (*ADH4* rs13112924, *RBL1* rs1780703, *TCEAL7* rs17340307, *ERCC3* rs1803541, *CUL1* rs6958001, *IL4R* rs1801275, *SHMT1* rs2168781, *BCL2* rs4987792, *FUT6* rs364637 and *TIMP3* rs1962223). Thus, 1,416 SNPs are included in the analyses.

3.2. Statistical analysis

We used Cox proportional hazards regression (27) to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the associations with overall survival, adjusted for known prognostic factors. In order to adjust for known factors, a clinical risk score was created and used as a covariate in genetic models. The risk score was created from the linear combination of the following variables multiplied by their Cox regression coefficients (i.e., $x\beta$): race (white, other), age (<50 , 50–59, 60–69, 70–79s, 80+), location of debulking surgery (Mayo, elsewhere), stage (III/IV, I/II, missing), debulking (optimal, suboptimal, missing), ascites (yes, no, missing), grade (1/2, 3/4, missing), histology (serous,

endometrioid, clear cell, mucinous, mixed, other or unknown), pre-surgery CA125 ($\log_{10}$), and chemotherapy (taxane, platinum). Using a single variable for the risk score allowed for reduced degrees of freedom over inclusion of all individual prognostic factors in Cox models (28–30). The clinical risk score was highly predictive of survival time (HR 2.7, 95% CI 2.2 – 3.3). Genotype-specific survival curves from a Cox model were created using the median clinical risk score as a baseline covariate.

Association testing proceeded for SNPs, for genes, and for pathways. First, for each SNP, a primary association using an ordinal coding (i.e., 0, 1, 2 copies of minor allele) was tested (analogous to the Armitage test for trend for binary endpoints), and HRs with 95% CIs were estimated per-allele as well as for heterozygotes and minor allele homozygotes compared to major allele homozygotes separately. Gene level tests were completed using principal component analysis (PCA) in a manner similar to that described previously (31). Briefly, PCs were created using linear combinations of ordinally-scaled SNPs (based on number of minor alleles), and the smallest set of resulting PCs that explained at least 90% of the SNP variance within the gene was then included in Cox models. The rationale for this approach over inclusion of the entire collection of SNPs in a multivariate model is the resulting reduction in degrees of freedom, avoidance of model fitting issues due to multi-collinearity from LD, and potential improvement in statistical power. Finally, to assess overall significance of a pathway, we computed the tail strength (32) of the set of gene level tests from each pathway and performed permutation testing based on 1,000 simulations of our entire association-testing process. Tail strength is closely related to the false discovery rate and assesses the relative strength of the collection of an observed set of empirical p values. Positive tail strength indicates more associations than expected by chance, and negative tail strength indicates fewer associations than expected by chance. Analyses were repeated on the subset of serous patients, due to possible genetic heterogeneity of associations by histological subtype.

As a conservative approach, overall interpretation of associations began with examination at the pathway level, followed next by assessment of gene- and SNP-level results. This tiered testing approach was carried out in the spirit of Fisher's protected least significance difference test; individual associations on a lower hierarchy level or tier were not considered statistically significant in the absence of significance at higher tier(s). At the pathway and gene levels, we focused on results significant at $p=0.10$, and, at the SNP level, we focused on results significant at $p=0.05$. Because of this tiered interpretation of results and the exploratory nature of this analysis, results were not corrected for multiple testing. Plots of LD displaying r^2 for each gene were created using Haploview v 4.1 (33).

4. NOTES

4.1

Survival (or more generally time-to-event) genotyping studies present different sets of challenges than etiology studies. Power is limited by the number of events in the dataset, whether the event is death or a measure of disease progression. Covariates are also different in a time to event model, as you hope to account for heterogeneity in case outcome, rather than potential confounders of case vs. control status. Researchers need to decide whether to

examine the association of SNPs alone or after adjustment for known clinical factors that influence outcome heterogeneity in their population.

4.2

Factors such as stage, morphology, tumor grade, residual disease from primary debulking surgery and others are associated with overall survival in patients with ovarian cancer, and we chose to adjust for these factors in our study. The final adjustment list would require approximately 20 degrees of freedom to include in the survival model, which was not desirable given the sample size of the study. These clinical factors have been established in other studies and precise estimation of their effect size and direction was not as important as examining the SNP effect in the models. Thus, we developed a single degree of freedom risk score with the clinical factors. This single degree of freedom risk score was then used as the lone adjustment variable in our SNP Cox models. We have used this approach with success on other studies, and simulation studies have shown this approach yields similar results as including all the clinical variables in the model, with the 1 degree of freedom risk score tending to results in tighter estimates of the SNP effects.

4.3

Gene level tests provide some benefits over single SNP approaches. Our simulations have shown that principal components analysis (PCA) based gene tests have good power to detect effects under a wide range of genetic models while greatly reducing the number of discovery tests performed, thus reducing the multiple testing concerns. PCA tests do not provide information about the nature of the association (protective or deleterious) but further investigation of SNPs within a significant gene can help further identify the nature of associations with outcome.

4.4

We also employed a simulation-based approach to examine whether the genes among varied biological based pathway showed greater significance than expected due to chance. These gene set approaches can help a researcher further identify meaningful associations within biological or other pre-defined gene groupings. See Fridley and Biernacka for a expanded discussion of these analysis approaches (34).

Acknowledgement

This work was supported by the National Cancer Institute (R01-CA86888, R01-CA122443, P50-CA136393-01A1), the Fred C. and Katherine B. Andersen Foundation, and the Mayo Foundation

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2012; 62:10–29. [PubMed: 22237781]
2. Cannistra SA. Cancer of the ovary. N Engl J Med. 2004; 351:2519–2529. [PubMed: 15590954]
3. Hoskins WJ, et al. The influence of cytoreductive surgery on recurrence-free interval and survival in small-volume stage III epithelial ovarian cancer: a Gynecologic Oncology Group study. Gynecol Oncol. 1992; 47:159–166. [PubMed: 1468693]

4. McGuire V, Jessor CA, Whittemore AS. Survival among U.S. women with invasive epithelial ovarian cancer. *Gynecol Oncol.* 2002; 84:399–403. [PubMed: 11855877]
5. Barnholtz-Sloan JS, et al. Ovarian cancer: changes in patterns at diagnosis and relative survival over the last three decades. *Am J Obstet Gynecol.* 2003; 189:1120–1127. [PubMed: 14586365]
6. Aletti GD, et al. Current management strategies for ovarian cancer. *Mayo Clin Proc.* 2007; 82:751–770. [PubMed: 17550756]
7. Bolton KL, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA: The Journal of the American Medical Association.* 2012; 307:382–389.
8. Nimeiri HS, et al. Efficacy and safety of bevacizumab plus erlotinib for patients with recurrent ovarian, primary peritoneal, and fallopian tube cancer: a trial of the Chicago, PMH, and California Phase II Consortia. *Gynecol Oncol.* 2008; 110:49–55. [PubMed: 18423560]
9. Inoue J, et al. NF-kappaB activation in development and progression of cancer. *Cancer Sci.* 2007; 98:268–274. [PubMed: 17270016]
10. Storz P. Reactive oxygen species in tumor progression. *Front Biosci.* 2005; 10:1881–1896. [PubMed: 15769673]
11. D'Andrilli G, Giordano A, Bovicelli A. Epithelial ovarian cancer: the role of cell cycle genes in the different histotypes. *Open clinical cancer journal.* 2008; 2:7–12. [PubMed: 18665245]
12. Chiantore MV, et al. Senescence and cell death pathways and their role in cancer therapeutic outcome. *Current medicinal chemistry.* 2009; 16:287–300. [PubMed: 19149578]
13. Hajra KM, Tan L, Liu JR. Defective apoptosis underlies chemoresistance in ovarian cancer. *Adv Exp Med Biol.* 2008; 622:197–208. [PubMed: 18546629]
14. Ho SB, et al. Heterogeneity of mucin gene expression in normal and neoplastic tissues. *Cancer Res.* 1993; 53:641–651. [PubMed: 7678777]
15. Baldus SE, Engelmann K, Hanisch FG. MUC1 and the MUCs: a family of human mucins with impact in cancer biology. *Critical reviews in clinical laboratory sciences.* 2004; 41:189–231. [PubMed: 15270554]
16. Saldova R, et al. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Disease markers.* 2008; 25:219–232. [PubMed: 19126966]
17. Yap TA, Carden CP, Kaye SB. Beyond chemotherapy: targeted therapies in ovarian cancer. *Nat Rev Cancer.* 2009; 9:167–181. [PubMed: 19238149]
18. Sellers TA, et al. Estrogen bioactivation, genetic polymorphisms, and ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2005; 14:2536–2543. [PubMed: 16284375]
19. Lancaster JM, et al. Identification of genes associated with ovarian cancer metastasis using microarray expression analysis. *Int J Gynecol Cancer.* 2006; 16:1733–1745. [PubMed: 17009964]
20. Jones MB, et al. The granulin-epithelin precursor/PC-cell-derived growth factor is a growth factor for epithelial ovarian cancer. *Clin Cancer Res.* 2003; 9:44–51. [PubMed: 12538450]
21. Chien J, et al. Epigenetic silencing of TCEAL7 (Bex4) in ovarian cancer. *Oncogene.* 2005; 24:5089–5100. [PubMed: 15870691]
22. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–861. [PubMed: 17943122]
23. Hinds DA, et al. Whole-genome patterns of common DNA variation in three human populations. *Science.* 2005; 307:1072–1079. [PubMed: 15718463]
24. Carlson CS, et al. Mapping complex disease loci in whole-genome association studies. *Nature.* 2004; 429:446–452. [PubMed: 15164069]
25. de Bakker PI, et al. Efficiency and power in genetic association studies. *Nat Genet.* 2005; 37:1217–1223. [PubMed: 16244653]
26. Olson JE, et al. A comprehensive examination of CYP19 variation and risk of breast cancer using two haplotype-tagging approaches. *Breast Cancer Res Treat.* 2007; 102:237–247. [PubMed: 17004113]
27. Cox DR. Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Series B.* 1972; 34:187–220.

28. Cerhan JR, et al. Prognostic significance of host immune gene polymorphisms in follicular lymphoma survival. *Blood*. 2007; 109:5439–5446. [PubMed: 17327408]
29. Habermann TM, et al. Host immune gene polymorphisms in combination with clinical and demographic factors predict late survival in diffuse large B-cell lymphoma patients in the pre-rituximab era. *Blood*. 2008; 112:2694–2702. [PubMed: 18633131]
30. Wang SS, et al. Polymorphisms in DNA repair and one-carbon metabolism genes and overall survival in diffuse large B-cell lymphoma and follicular lymphoma. *Leukemia*. 2008
31. Gauderman WJ, et al. Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol*. 2007; 31:383–395. [PubMed: 17410554]
32. Taylor J, Tibshirani R. A tail strength measure for assessing the overall univariate significance in a dataset. *Biostatistics*. 2006; 7:167–181. [PubMed: 16332926]
33. Barrett JC, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–265. [PubMed: 15297300]
34. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. *European journal of human genetics : EJHG*. 2011; 19:837–843. [PubMed: 21487444]
35. Dressman HK, et al. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol*. 2007; 25:517–525. [PubMed: 17290060]

Table 1

List of investigated genes (N=227) and number of SNPs analyzed by pathway

Pathway	Genes	N SNPs Analyzed (N Attempted)
Angiogenesis	<i>EIF2B5, ELAVL1, HIF1A, KDR, MMP1, MMP2, MMP26, MMP3, MMP9, NOS3, PDGFD, PDGFRA, TIMP1, TIMP2, TIMP3, VEGFA, VEGFC, VHL</i>	130 (147)
Apoptosis	<i>APAF1, ATM, BAD, BAX, BCL2L1, BID, CASP10, CASP3, CASP6, CASP7, CASP8, CASP9, CYC1, EIF2S1, PARP1, PTK2, STAT1, TLN1</i>	118 (122)
Cell Cycle	<i>ABL1, CCNA1, CCNA2, CCNB1, CCNB2, CCND1, CCND2, CCND3, CCNE1, CCNE2, CCNG1, CCNG2, CDC2, CDC25A, CDC25B, CDK2, CDK4, CDK6, CDK7, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C, CDKN2D, CUL1, E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, PLK1, RBL1, RBL2, SKP2, TFDP1, TFDP2</i>	222 (244)
Cellular Senescence	<i>ACTB, AKT1, ARID1A, BCL2, EGFR, ERCC3, FHIT, HRAS, IGF1, IGF1R, IGF2, KRAS, MYC, NF1, NR3C1, PIK3CB, POLR2A, PRKCA, SHC1, SMARCA2, SMARCA4, SMARCB1, SMARCC1, SMARCC2, SMARCE1, SOS1, SRC, TBP, TEP1, TERT, TNKS, TP53, XRCC5</i>	211 (223)
Detoxification	<i>ADH1A, ADH1B, ADH1C, ADH4, ADH5, ADH6, CYP1A1, CYP1A2, EPHX1, GSTM1, GSTP1, NAT1, NAT2, NQO1, NQO2</i>	108 (114)
Glycosylation	<i>FUT10, FUT3, FUT5, FUT6, FUT7, FUT8, FUT9, GALNT1, GALNT14, GALNT2, GALNT3, GALNT5, GALNT6, GALNT7, MGAT2, MGAT3, MGAT5, POFUT1, ST3GAL1, ST3GAL3, ST3GAL4, ST3GAL5, ST6GAL1, ST6GALNAC5, ST8SIA2, ST8SIA5</i>	95 (101)
Inflammation	<i>ALOX12, ALOX15, ALOX5, CCL11, CCL2, CCL3, CCR3, CRP, CXCL16, IL10, IL15RA, IL18, IL1A, IL1B, IL1RN, IL4R, IL6, IL6R, IL7R, IL8RA, IL8RB, IL9, PTGS1, PTGS2, TLR2, TNF</i>	154 (163)
One-carbon transfer	<i>AHCYL1, ALDH1L1, DHFR, DNMT1, DNMT3A, DNMT3B, DPYD, FOLR1, MAT2B, MBD4, MGMT, MTHFD1, MTHFD1L, MTHFD2, MTHFR, MTHFS, MTR, MTRR, SHMT1, SLC19A1, TYMS</i>	179 (187)
Under-expressed	<i>ABCB1, ACTG1, ARPC4, COL8A1, DKK3, DUSP1, EHD2, ELOVL4, F3, FGF2, FGF5, FOXD1, GALK1, HNT, KCNG1, LAMA4, MAPRE2, NRP2, P2RX5, PAPP, PPP1R3A, PPP3CC, TBX3, TPM1, ULBP2</i>	120 (154)
Miscellaneous	<i>CYP19A1, GRN, NFKBIA, NFKBIB, SULT1E1, TCEAL7</i>	79 (81)
		1,416 (1,536)

Under-expressed genes from Dressman *et al.*, J Clin Oncol, 25: 517-525, 2007 (35) and Lancaster *et al.*, Int J Gynecol Cancer, 16: 1733-1745, 2006 (19); N analyzed SNPs excludes 50 (3%) which failed genotyping and 70 (5%) with incorrect genomic coordinates or low MAF.