Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2018 September ; 27(9): 1101–1109. doi:
10.1158/1055-9965.EPI-17-0728.

Transcriptomic Characterization of Endometrioid, Clear Cell, and High Grade Serous Epithelial Ovarian Carcinoma

Brooke L. Fridley^{1,2,*}, Junqiang Dai¹, Rama Raghavan¹, Qian Li^{1,2}, Stacey J. Winham³, Xiaonan Hou⁴, S. John Werooha^{4,5}, Chen Wang³, Kimberly R. Kalli⁴, Julie M. Cunningham⁶, Kate Lawrenson^{7,8}, Simon A. Gayther^{8,9}, Ellen L. Goode³

¹Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA

²Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA

³Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

⁴Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA

⁵Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

⁶Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

⁷Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

⁸Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

⁹Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

Abstract

Background: Endometrioid carcinoma (EC) and clear cell carcinoma (CC) histotypes of epithelial ovarian cancer are understudied compared to the more common high grade serous carcinomas (HGSCs). We therefore sought to characterize EC and CC transcriptomes in relation to HGSC.

Methods: Following bioinformatics processing and gene abundance normalization, differential expression analysis of RNA sequence data collected on fresh frozen tumors was completed with nonparametric statistical analysis methods (55 ECs, 19 CCs, 112 HGSCs). Association of gene expression with progression free survival (PFS) was completed with Cox proportional hazards models. Eight additional multi-histotype expression array datasets (N = 852 patients) were utilized for replication.

Results: In the discovery set, tumors generally clustered together by histotype. Thirty-two protein-coding genes were differentially expressed across histotype ($p < 1 \times 10^{-10}$) and showed similar associations in replication datasets, including *MAP2K6*, *KIAA1324*, *CDH1*, *ENTPD5*,

*Correspondence: Brooke L. Fridley, Moffitt Cancer Center, 12902 Magnolia Dr, Tampa, FL, USA 33612, Phone: (813)745-1461, Fax: (813) 745-6107, brooke.fridley@moffitt.org.

Conflict of Interest: None declared

LAMB1 and *DRAM1*. Nine genes associated with PFS ($p < 0.0001$) showed similar associations in replication datasets. In particular, we observed shorter PFS time for CC and EC patients with high gene expression for *CCNB2*, *CORO2A*, *CSNK1G1*, *FRMD8*, *LIN54*, *LINC00664*, *PDK1*, and *PEX6*, whereas, the converse was observed for HGSC patients.

Conclusions: The results suggest important histotype differences which may aid in the development of treatment options, particularly those for patients with EC or CC.

Impact: We present replicated findings on transcriptomic differences and how they relate to clinical outcome for two of the rarer ovarian cancer histotypes of EC and CC, along with comparison to the common histotype of HGSC.

Keywords

ovarian cancer; progression free survival; differential expression; high-throughput sequencing

Introduction

Epithelial ovarian cancer (EOC) makes up approximately 90% of all ovarian tumors and is the fifth leading cause of cancer death among women in the United States (1); five-year overall survival remains around 45% for all stages (27% for distant disease). One reason for the lack of success with the treatment of EOC can be attributed to disease heterogeneity from both the morphological and molecular perspectives (2, 3). The most common histological subtype (histotype) of EOC is high-grade serous carcinoma (HGSC), representing 70–74% of EOC cases followed by endometrioid carcinoma (EC) (7–24%) and clear-cell carcinoma (CC) (10–26%) (4). HGSC has been extensively studied in terms of its molecular profile and genomic landscape (3, 5–9), in part due to its higher frequency. Some features of HGSC tumors are the following: believed to originate from fallopian tube secretory epithelium and/or ovarian surface epithelium cells; majority have somatic *TP53* mutations; and several germline risk alleles associate almost exclusively with serous EOC (10, 11). In contrast, features of ECs and CCs include: appear to arise from endometrial epithelium (12); often harbor *ARID1A* mutations (13); and appear to have some unique risk alleles (10, 14, 15). Due to the rareness of these two EOC histotypes, limited research has been conducted on EC and CCs tumors.

We sought to characterize tumor transcriptomes of EC and CC in relation to HGSC, and to evaluate clinical associations using RNA-Seq in a clinically annotated EOC patient population, followed by replication of findings in publicly available expression array datasets. The determination of the molecular features that differ between these histotypes and how these transcriptomic differences relate to clinical response might lead to new insights in histotype-specific treatments for EOC patients.

Methods and Materials

Study Participants, RNA Sequencing, and Bioinformatics

Eligible patients were women aged 20 years or above ascertained between 2000 and 2014 at the Mayo Clinic within one year of diagnosis with pathologically confirmed primary EOC.

Clinical diagnoses and histotypes were confirmed by immunohistochemistry-guided re-review by a gynecologic pathologist. The pathologist also verified tumor grade and reviewed each fresh frozen tissue to ensure 70% tumor content prior to RNA extraction. All cases provided informed written consent to protocols approved by the Mayo Clinic Institutional Review Board. Supplemental Table 1 summarizes the characteristics of the 186 EOC patients included in this study. Transcriptomic sequencing of RNA was performed in four batches using Illumina TruSeq library preparation (Stranded Total RNA Library Prep Kit or RNA Library Preparation Kit v2) and sequenced on the Illumina HiSeq 2000 sequencer with paired end reads (50- or 100-nucleotide, Supplementary Table 2). No technical replicates were included in the study. Primary analysis and de-multiplexing was performed using Illumina's *CASAVA* software, followed by alignment to the GRCh37 genome assembly using *TopHat2* (16) and abundance estimation at the gene level using *RSEM* (17).

Data Normalization and Tumor Purity

The normalization across RNA-Seq batch effects consisted of two steps: estimation of surrogate variables and normalization. The first step involved surrogate variable analysis (SVA) conducted on the residual matrix generated from a linear model fit of the gene expression abundance against histotype. We adopted the methodology proposed by Buja and Eyuboglu to identify surrogate variables that represent known and unknown confounders other than the primary factor of histotype (18). The algorithm to carry out the analysis was provided in the Bioconductor package *SVA* via function *svaseq*, with two approaches to determine the number of significant confounders or surrogate variables (18, 19). For the second step involving across sample normalization, we employed the *ComBat* function in Bioconductor to adjust for the surrogate variables and/or confounders other than histotype, using an empirical Bayesian approach proposed by Johnson et al (20). Lastly, we adjusted for the known factor (i.e. batch) and visually inspected the data with principal component analysis (PCA) to determine if the batch effects were adequately removed (see Supplemental Methods).

Following normalization of the data to account for technical artifacts, we determined if any differences in tumor purity remained between the samples and associated with histotype. Estimates of tumor purity were calculated using the program *ESTIMATE* (21), where we observe the association of tumor purity estimates with histotype to be non-significant ($p > 0.10$). *SVA* and *ComBat* normalization effectively removed the differences in library sizes between samples and difference in tumor purity between histotypes, thus eliminating the need for additional normalization of these factors (see Supplemental Methods).

Statistical Analyses

Statistical analyses utilized the normalized count data, treating it as a continuous measurement, where both positive and negative values are possible. To assess similarity of global tumor transcriptomes, PCA was completed, along with unsupervised clustering using a parametric model-based approach implemented in *mclust* (22) using the top 75% most variable genes. Determination of number of optimal clusters was completed using the Bayesian Information Criteria (BIC). After assessing model fit for a random sample of ten genes using the normalized gene expression data and three modeling methods (i.e., negative

binomial model using edgeR(23), ANOVA, and nonparametric Kruskal-Wallis (KW) tests), we choose to use the conservative nonparametric KW tests to assess per-gene differential expression between all histotypes (2 degree of freedom tests)(see Supplemental Methods). In addition to testing whether mean gene expression abundances differed between the three histotypes, we also completed three sets of pairwise differential expression analyses (1 degree of freedom tests) using non-parametric two-sample t-tests (Wilcoxon tests).

Association of gene expression with outcome was completed with progression free survival (PFS) defined as time from the date of diagnosis to the date that second-line therapy was initiated for a clinically-actionable tumor recurrence or death, if progression were unknown. Cox proportional hazards models were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for association of each gene expression with PFS. Analyses adjusted for covariates associated with PFS ($p < 0.05$ based on score test), including stage (coded as I/II and III/IV) and surgical debulking status (coded as optimal, sub-optimal). Censoring at 10 years was performed to minimize competing causes of mortality. We excluded four patients with missing or unknown debulking status from the PFS analysis. We applied a less stringent statistical significance threshold for PFS analysis compared to the differential gene expression analysis due to the smaller effect sizes one expects to see for gene expression association with clinical outcome. To determine if a different relationship exists between gene expression and PFS by histotype, a gene expression by histotype interaction term was also included in the Cox proportional hazards model. Q-values were computed to control for the false-discovery rate (24). Due to limited sample size, molecular subgroup analyses were not completed.

Replication in Public Expression Array Studies

Data from eight EOC studies where more than one histotype group was included from the curatedOvarianData (9, 25) were used to assess replication of key findings. The TCGA ovarian cancer study, which included only serous EOC tumors, was not included as one of the replication studies. We did not restrict serous tumors to high-grade due to the limited sample size and lack of information related to grade in some of the replication studies. A summary of these studies are presented in Supplemental Table 3. These eight studies consisted of the following studies in GEO (Gene Expression Omnibus): GSE51088(26), GSE44104(27), GSE26193(28), GSE14764(29), GSE2109 (IGC's Expression Project for Oncology), GSE30161(30), GSE6008(31–34), GSE9891(5). Five of these eight studies also had clinical outcome(s): GSE51088 (overall survival (OS)), GSE26193 (OS and PFS), GSE14764 (OS), GSE30161 (OS and PFS), GSE9891 (OS and PFS). The gene expression data in the GEO studies were all generated from microarray technologies, whereby all the datasets have been previously normalized for analysis, as described in Waldron et al (9) and Ganzfried et al (25).

Statistical analyses were completed in a similar manner as completed for the Mayo Clinic transcriptome study. We used a rigorous threshold for determining replication of histotype-specific differentially expressed genes, whereby a gene would be initially considered as “replicated” if more than one replication study had a KW $p < 0.001$. As the KW test does not look at direction of the effect, we next looked to verify similarity of gene expression pattern

for the three histotypes between the discovery set (Mayo Clinic) and the replication set (GEO studies). Replication for PFS was considered if at least one of the 5 GEO studies with outcome data had an association $p < 0.05$.

Results

We completed robust normalization of RNA-Seq gene expression data measured on 186 Mayo Clinic EOC patient tumors, resulting in 22,234 Ensembl genes suitable for differential expression analyses. SVA and normalization accounted for both known (i.e., tumor purity, library size, library prep) and unknown differences between samples. To determine the extent to which EC, CC, and HGSC histotypes clustered separately based on gene expression, unsupervised model-based clustering and principal component analysis (PCA) were performed, with results presented in Figure 1. The three histotypes clearly separated from one-another (Figure 1A), and the model-based cluster assignments agreed well with the histotypes. The HGSC tumors primarily grouped into two clusters (Cluster 2 and Cluster 5), while Cluster 1 primarily comprised EC tumors and Cluster 3 and Cluster 4 represented CC tumors (Figure 1B, Supplemental Table 4). In comparison of the model based cluster groups with previously reported subtypes related to outcome (i.e., CLOVAR signature by Verhaak et al(6)), we observed a moderate association between the two different cluster assignments (all EOC tumors p -value = 0.01; HGSC tumors only $p = 0.35$) (Supplemental Table 4). In particular, Cluster 4 is enriched for “differentiated” (DIF) tumors (7 out of 9 or 78% of EOC tumors in Cluster 4 were denoted as DIF tumors), with the other cluster groups having moderate representation of EOC tumors across the four CLOVAR subtypes. However, it should be emphasized that the motivating questions and the approaches for generating the two cluster assignments were different. Measures of global expression, defined as the median gene expression level, differed between the three histotypes ($p = 2.7 \times 10^{-9}$) (Figures 2A & 2B), but were not related to PFS. However, median global gene expression was found to be associated with debulking status ($p = 0.004$) with higher expression observed in tumors that were optimally surgically debulked (Figure 2C), noting that debulking status was not associated with histotype ($p = 0.1$, Supplemental Table 1).

We identified 1,254 individual genes which showed differential expression across the EOC histotypes (KW $p < 1 \times 10^{-10}$; Supplemental Table 5); 606 showed differential expression between HGSC and EC (474 up-regulated in EC) and 244 genes showed differential expression between CC and HGSC (238 upregulated in CC) based on pairwise histotype comparisons (Wilcoxon $p < 1 \times 10^{-10}$). Only four genes were up-regulated in HGSC compared to both EC and CC tumors: three protein-coding genes (*GFRA3*, *ASCL5*, and *NPTX1*) and a lncRNA (ERVH45-1). It should be noted that these three pairwise histotype comparisons have varied power to detect true differences due to the differences in the sample size in each histotype. Analysis involving the eight GEO publically available EOC studies, 73 of the 1,254 differentially expressed genes showed replication, defined as KW $p < 0.001$ in more than one study (Supplemental Table 6). Plots of the mean expression levels for the histotypes for these 73 genes can be found in Supplemental Figure 1, with the top plot displaying the data from the Mayo Clinic study (discovery study) and the bottom plot showing the data from the various GEO study (replication studies). Based on this visual assessment, we determined that 32 protein-coding genes showed consistent direction of

effect in the discovery and replication studies. The gene expression data from the Mayo Clinic study for these 32 genes are presented in Table 1, Figure 3 and Supplemental Figure 2, which include the following biologically relevant genes: *CDH1*, *DRAM1*, *ENTPD5*, *FERMT2*, *GLRX*, *GPX3*, *KIAA1324*, *LAMB1*, *MAP2K6*, *PRKAB1*, *RHOB*, *SLC22A5*, and *TSPAN1*. Additionally, 6 of the genes denoted in Table 1 (*ANXA4*, *FLOT1*, *GLRX*, *LAMB1*, *MITF*, and *SGPP1*) were also found to be differentially expressed between histotypes by Zorn et al (35).

Analysis of PFS with gene expression levels in the Mayo Clinic discovery study revealed that expression of *BNIP3P17* was prognostic in all three histotypes and in the same direction (i.e., additive effect) and indicated that 22 additional genes were associated with PFS in a different manner by histotypes (i.e., interaction effect between histotype and gene expression) ($p < 0.0001$; Supplemental Table 7). Therefore, we assessed the association of clinical outcome with these 23 genes in the five publically available GEO studies in which clinical outcome were reported; results are shown in Supplemental Table 8. It should be noted that many of the replication studies had small sample sizes and the model was only adjusted for stage (when available), as none of the studies had debulking status recorded. Association between PFS and *BNIP3P17* did not replicate in any of the studies. However, the following nine genes showed evidence of a histotype-by-gene expression interaction effect ($p < 0.05$) replication in at least one replication study (Table 2 and Figure 3): *CCNB2*, *CORO2A*, *CSNK1G1*, *FRMD8*, *LIN54*, *LINC00664*, *PDK1*, *PEX6*, and *USP31*. Figure 4 presents the predicted PFS curves for these nine genes for the Mayo Clinic study for each histotype by high or low gene expression level. In particular, we observed that for 8 of these 9 genes (i.e., all but *USP31*), lower expression levels were associated with better outcome (longer PFS times) for EC and CC EOC patients, while the converse was observed for HGSC EOC patients (i.e., higher expression levels associated with longer PFS time), after adjusting for age of diagnosis, stage and debulking status. However, one of these eight genes, *PEX6*, was also found to have the expression level associated with disease stage ($p = 0.001$, Spearman correlation = -0.24).

Discussion

Little progress has been made in the treatment of EOC, due in part to incomplete knowledge about disease heterogeneity with regards to histotype. Using the largest known next-generation sequencing study of the three most common histotypes, we sought to determine transcriptomic differences between EC, CC, and HGSC. We found 1,254 genes with $p < 1 \times 10^{-10}$ (FDR q-value $< 3 \times 10^{-10}$), of which 32 replicated with consistent direction of effect in publicly available data from GEO (Table 1). Additionally, we replicated the association of nine genes with differing relationships with PFS between the three histologies (Table 2).

Overall, we found that the EOC tumors clustered together based on histotype. In particular, we observed most EC tumors (78%, 43 out of 55 EOC tumors) cluster together in one cluster (C1), while CC tumors cluster evenly into two clusters (C3 and C4) with 95% of CC tumors falling into C3 and C4 cluster groups. Lastly, we observed that clusters 2 and 5 (C2, C5) are predominately HGSC clusters, with 84% (44 out of 52) and 93% (66 out of 71) of tumors in C2 and C5, respectively, being HGSC (Supplemental Table 4, Figure 1). We also

observed globally, as measured by median expression level, that CC tumors had higher global expression levels, followed by EC and HGSC ($p = 2.7 \times 10^{-9}$); optimally debulked patients had higher transcriptomic expression levels compared to tumors from non-optimally debulked patients ($p = 0.0044$) (Figure 2). This observed difference in global gene expression could be due to copy number and genomic instability, known to be a hallmark of serous EOC and triple negative breast cancers (36). That is, the observed lower levels of median global gene expression in HGSC tumors could be due to the higher genomic instability compared to the other tumor histotypes of EOC (i.e., higher genomic instability might lead to fewer “functional” genes and thus a lower level of global gene expression levels). Much less is known about how this compares in the rarer histotypes of ECs and CCs, with some evidence that ECs and CCs have lower levels of genomic stability (12). We did not observe any association of median global gene expression level with PFS ($p = 0.32$), after adjusting for debulking status, stage and histotype, and no observed histotype-specific effects for association of median global expression level with PFS (interaction effect, median global gene expression by histotype p -value = 0.73).

As the primary focus of this research, we identified 32 genes that were differentially expressed across histotypes, along with another nine genes associated with PFS differently between the three histotypes (i.e., interaction effect between gene expression level and histotype with respect to relationship with PFS). Out of the 32 replicated differentially expressed genes, 29 genes were significantly up-regulated (Wilcoxon $p < 0.0001$) in CC compared to HGSC and EC (Table 1, Supplemental Figure 2). *KIAA1324*, important in cellular response to stress, and *MAP2K6*, involved in cell cycle arrest and apoptosis, were found to be up-regulated in EC compared to both HGSC and CC tumors, while *FLOT1*, *SLC3A1* and *C12orf75* were found to be down-regulated in EC tumors (Wilcoxon $p < 0.0001$). Mitogen-activated protein kinases (MAPK) have been found to be metastasis suppressor proteins (37); therefore, high expression of *MAP2K6* observed in EC tumors might be related to reduced metastases and subsequent longer survival. Recently, *miR-625-3p* has been identified as a biomarker for oxaliplatin resistance in colon cancer, where *MAP2K6* is the direct target for this microRNA (miRNA). The subsequent down-regulation of *MAP2K6* impairs p38-MAPK stress signaling and thus leads to oxaliplatin resistance by preventing apoptosis (38). As most EOC patients are treated with combination platinum-taxane chemotherapies, a similar mechanism might underlie differences in clinical outcome, whereby EC EOC patients generally have better clinical outcome as compared to HGSC patients. Lastly, *TALDO1* and *UPK3B* were down-regulated in HGSC compared to EC and CC tumors; little is known about how these two genes relate to carcinogenesis and EOC.

One limitation of replication study was the inclusion of both low and high-grade serous tumors due to the limited sample size or lack of grade information in many of the studies. As the majority of serous EOC tumors are high-grade, with only 5–10% of serous tumors being low grade (39), we believe the impact of including possible low-grade serous EOC tumors in the replication study to be small. A second limitation of the replication study is that the majority of studies involved patients enrolled prior to the start of the enrollment for the Mayo Clinic study (Supplemental Table 3). A final limitation of the study is the focus on only transcriptomic differences between the three histologies, without orthogonal validation of downstream proteins known to be discriminant of the different histologies. None of our

32 validate genes included genes known to be related to histotype based on immunohistochemistry (IHC) studies (*VIM*, *PGR*, *ARID1A*, *NAPSA*, *TP53*, *CDKN2A*, *TFF3*, *WT1*)(40); however, all but *TP53* had a KW $p < 0.05$ for association with histotype (Supplemental Table 9).

Of particular interest among the nine PFS related genes, two genes (*LIN54* and *CCNB2*) are related to cell-cycle and two genes (*CSNK1G1* and *PDPK1*) are kinase-related genes. For all nine genes, with the exception of *USP31*, we observed an increased risk of progression in EC patients with higher gene expression levels (Table 2, Figure 4). To a lesser extent, we saw a similar association between high gene expression and increased risk of progression in CC for these genes. In contrast to EC and CC, high gene expression for these genes, with the exception of *USP31*, was associated with better clinical outcome in HGSC. For example, EC and CC patients with higher levels of *CCNB2* had worse outcome than patients with lower expression. This finding is not unique to EOC, as it has also been observed in breast cancer (41). In contrast, HGSC patients with high expression of *CCNB2* had a reduce risk of progression. (41) While elevated *PDK1* has been found to promote tumor growth and metastasis in breast cancer (42), HGSC patients with high expression of *PDK1* have been found to have improved survival compared to patients negative for protein expression (43), supporting our findings that HGSC patients with higher *PDK1* expression have better outcome compared to patients with low *PDK1* expression (Figure 4). Interesting, we observed the opposite effect of *PDK1* for patients with CC and EC tumors, with low expression levels resulting in better clinical outcome. Lastly, *CSNK1G1* (i.e., casein kinase I, gamma 1 or CK1 γ 1) is involved in many cancer related pathways, including hedgehog signaling, mTOR signaling and p53 interaction pathway, and as such, many CK1 inhibitors are currently under investigation (44, 45).

In conclusion, we present replicated findings on transcriptomic differences and how they appear to relate to PFS for two of the rarer EOC histotypes of EC and CC, along with comparison to the common histotype of HGSC. One of the strengths of this report is the incorporation of eight additional public expression datasets for replication. Future research is needed to assess these results in a larger cohort of EC and CC patients, along with the incorporation of epigenetic, copy number and protein expression information to determine molecular differences between the histotypes of EOC and their impact on clinical outcome. These findings suggest important biological insights into the differences between the three most common histotypes of EOC and may aid in the development of targeted treatment options, particularly those related to the treatment for patients with EC and CC EOC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by the University of Kansas Cancer Center (P30 CA168524) through a summer student research project award (Dai), The Mayo Foundation (Fridley; sequencing), R01 CA122443 (Goode), P50 CA136393 (Kalli, Goode, Fridley), P20 GM103418 (Fridley, Raghavan), R21 CA182715 (Fridley) and R00CA184415 (Lawrenson).

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin* 2015;65:5–29. [PubMed: 25559415]
2. Kobel M, Kalloger SE, Boyd N, McKinney S, Mehl E, Palmer C, et al. Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. *PLoS Med* 2008;5:e232. [PubMed: 19053170]
3. TCGA. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609–15. [PubMed: 21720365]
4. Alvarez RD, Karlan BY, Strauss JF. “Ovarian cancers: Evolving paradigms in research and care”: Report from the Institute of Medicine. *Gynecol Oncol* 2016;141:413–5. [PubMed: 27138456]
5. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* 2008;14:5198–208. [PubMed: 18698038]
6. Verhaak RG, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *The Journal of clinical investigation* 2013;123:517–25. [PubMed: 23257362]
7. Konecny GE, Wang C, Hamidi H, Winterhoff B, Kalli KR, Dering J, et al. Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *Journal of the National Cancer Institute* 2014;106.
8. Rieger M, Wei W, Waldron L, Culhane AC, Trippa L, Oliva E, et al. Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples. *Journal of the National Cancer Institute* 2014;106.
9. Waldron L, Haibe-Kains B, Culhane AC, Rieger M, Ding J, Wang XV, et al. Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. *Journal of the National Cancer Institute* 2014;106.
10. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet* 2010;42:874–9. [PubMed: 20852632]
11. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet* 2010;42:880–4. [PubMed: 20852633]
12. Kurman RJ, Shih Ie M. The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded. *Am J Pathol* 2016;186:733–47. [PubMed: 27012190]
13. McCluggage WG. Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis. *Pathology* 2011;43:420–32. [PubMed: 21716157]
14. Shen H, Fridley BL, Song H, Lawrenson K, Cunningham JM, Ramus SJ, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. *Nat Commun* 2013;4:1628. [PubMed: 23535649]
15. Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet* 2015;47:164–71. [PubMed: 25581431]
16. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11. [PubMed: 19289445]
17. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323. [PubMed: 21816040]
18. Buja A, Eyuboglu N. Remarks on Parallel Analysis. *Multivariate Behav Res* 1992;27:509–40. [PubMed: 26811132]
19. Leek JT. Asymptotic conditional singular value decomposition for high-dimensional genomic data. *Biometrics* 2011;67:344–52. [PubMed: 20560929]
20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27. [PubMed: 16632515]

21. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612. [PubMed: 24113773]
22. Yeung KY, Fraley C, Murua A, Raftery AE, Ruzzo WL. Model-based clustering and data transformations for gene expression data. *Bioinformatics* 2001;17:977–87. [PubMed: 11673243]
23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40. [PubMed: 19910308]
24. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440–5. [PubMed: 12883005]
25. Ganzfried BF, Riester M, Haibe-Kains B, Risch T, Tyekucheva S, Jazic I, et al. curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database : the journal of biological databases and curation* 2013;2013:bat013.
26. Karlan BY, Dering J, Walsh C, Orsulic S, Lester J, Anderson LA, et al. POSTN/TGFBI-associated stromal signature predicts poor prognosis in serous epithelial ovarian cancer. *Gynecol Oncol* 2014;132:334–42. [PubMed: 24368280]
27. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* 2014;33:3432–40. [PubMed: 23934190]
28. Mateescu B, Batista L, Cardon M, Gruosso T, de Feraudy Y, Mariani O, et al. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med* 2011;17:1627–35. [PubMed: 22101765]
29. Denkert C, Budczies J, Darb-Esfahani S, Gyorffy B, Sehouli J, Konsgen D, et al. A prognostic gene expression index in ovarian cancer - validation across different independent data sets. *J Pathol* 2009;218:273–80. [PubMed: 19294737]
30. Ferriss JS, Kim Y, Duska L, Birrer M, Levine DA, Moskaluk C, et al. Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance. *PLoS One* 2012;7:e30550. [PubMed: 22348014]
31. Wu R, Hendrix-Lucas N, Kuick R, Zhai Y, Schwartz DR, Akyol A, et al. Mouse model of human ovarian endometrioid adenocarcinoma based on somatic defects in the Wnt/beta-catenin and PI3K/Pten signaling pathways. *Cancer Cell* 2007;11:321–33. [PubMed: 17418409]
32. Hendrix ND, Wu R, Kuick R, Schwartz DR, Fearon ER, Cho KR. Fibroblast growth factor 9 has oncogenic activity and is a downstream target of Wnt signaling in ovarian endometrioid adenocarcinomas. *Cancer Res* 2006;66:1354–62. [PubMed: 16452189]
33. Bommer GT, Feng Y, Iura A, Giordano TJ, Kuick R, Kadikoy H, et al. IRS1 regulation by Wnt/beta-catenin signaling and varied contribution of IRS1 to the neoplastic phenotype. *J Biol Chem* 2010;285:1928–38. [PubMed: 19843521]
34. Schwartz DR, Kardia SL, Shedden KA, Kuick R, Michailidis G, Taylor JM, et al. Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Res* 2002;62:4722–9. [PubMed: 12183431]
35. Zorn KK, Bonome T, Gangi L, Chandramouli GV, Awtrey CS, Gardner GJ, et al. Gene expression profiles of serous, endometrioid, and clear cell subtypes of ovarian and endometrial cancer. *Clin Cancer Res* 2005;11:6422–30. [PubMed: 16166416]
36. Fehrmann RS, Karjalainen JM, Krajewska M, Westra HJ, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* 2015;47:115–25. [PubMed: 25581432]
37. Hickson JA, Huo D, Vander Griend DJ, Lin A, Rinker-Schaeffer CW, Yamada SD. The p38 kinases MKK4 and MKK6 suppress metastatic colonization in human ovarian carcinoma. *Cancer Res* 2006;66:2264–70. [PubMed: 16489030]
38. Rasmussen MH, Lyskjaer I, Jersie-Christensen RR, Tarpgaard LS, Primdal-Bengtson B, Nielsen MM, et al. miR-625–3p regulates oxaliplatin resistance by targeting MAP2K6-p38 signalling in human colorectal adenocarcinoma cells. *Nat Commun* 2016;7:12436. [PubMed: 27526785]
39. Plaxe SC. Epidemiology of low-grade serous ovarian cancer. *Am J Obstet Gynecol* 2008;198:459 e1–8; discussion e8–9. [PubMed: 18395040]

40. Kobel M, Rahimi K, Rambau PF, Naugler C, Le Page C, Meunier L, et al. An Immunohistochemical Algorithm for Ovarian Carcinoma Typing. *Int J Gynecol Pathol* 2016;35:430–41. [PubMed: 26974996]
41. Shubbar E, Kovacs A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdottir K, et al. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC cancer* 2013;13:1. [PubMed: 23282137]
42. Du J, Yang M, Chen S, Li D, Chang Z, Dong Z. PDK1 promotes tumor growth and metastasis in a spontaneous breast cancer model. *Oncogene* 2016;35:3314–23. [PubMed: 26455327]
43. Lohneis P, Darb-Esfahani S, Dietel M, Braicu I, Sehouli J, Arsenic R. PDK1 is Expressed in Ovarian Serous Carcinoma and Correlates with Improved Survival in High-grade Tumors. *Anticancer Res* 2015;35:6329–34. [PubMed: 26504072]
44. Schitteck B, Sinnberg T. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. *Mol Cancer* 2014;13:231. [PubMed: 25306547]
45. Delehouze C, Godl K, Loaec N, Bruyere C, Desban N, Oumata N, et al. CDK/CK1 inhibitors roscovitine and CR8 downregulate amplified MYCN in neuroblastoma cells. *Oncogene* 2014;33:5675–87. [PubMed: 24317512]

Figure 1:

3D plots of the first principal component (PC1) vs the second principal component (PC2) vs the third principal component (PC3) based on RNA sequences of 186 EOC tumors. Each point represents a tumor which is colored based on (A) histotype or (B) model-based cluster assignment. Figure 1A shows the tumors of the same histologies clustering together, while Figure 1B shows how tumors cluster into 5 clusters based on model based clustering. Cluster 1 is primarily comprised of EC tumors, cluster 2 and 5 are primarily HGSC tumors and CC tumors split between cluster 3 and 4. A cross-tabs of the cluster assignment vs histologies can be found in Supplemental Table 4.

Figure 2:

Plot of distribution of gene expression levels for the tumor samples. (A) Plot of middle 60% of gene expression measurements for the 186 tumors with tumors sorted by median level of gene expression. Blue represents HGSC tumors, green represents EC tumors and red represents CC tumors. This plot illustrated that in general the CC tumors had higher gene expression levels (globally) as compared to EC and HGSC tumors (B) Boxplots of the median gene expression level and variance in gene expression level for each tumor by histotype group; (C) Boxplots of the median gene expression level and variance of gene expression level for each tumor by debulking status. These plots show that optimally debulked tumors tend to have higher median global gene expression.

Figure 3: Heatmap of the 32 replicated differentially expressed genes and 9 replicated progression related genes for the Mayo Clinic Study. The heatmap shows that the majority of genes that were differentially expressed between the three histologies had higher levels of expression in the CC tumors.

Figure 4:

Predicted PFS curves from Cox proportional hazards model which included stage (low/high), debulking status (optimal / sub-optimal), histotype, gene expression level (high/low) and interaction between histotype and gene expression level. Using the results from the fitted model on the Mayo Clinic Study (HGSC = 109, EC N = 54, CC N = 19) we predicted the average PFS curves for EOC patients with high stage (III or IV) and optimally-debulked tumors for the various histotypes and level of gene expression (high/low). Two curves represent PFS for each histotype (HGSC, EC or CC) with either high (dashed line) or low (solid line) gene expression levels based on a median threshold within each histotype. For example, CC histotype with low expression is represented with the solid black line while CC histotype with high expression level is represented with the dashed black line. HGSC = red lines; EC = blue lines; CC = black lines. These plots show that for 8 of these 9 genes (i.e., all but gene *USP31*) lower expression levels were associated with better outcome (longer PFS times) for EC and CC EOC patients, while the converse was observed for HGSC EOC patients (i.e., higher expression levels associated with longer PFS time), after adjusting for

age of diagnosis, stage and debulking status. Note: gene expression levels for PEX6 were also found to be associated with stage ($p = 0.001$, Spearman correlation = -0.235).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

32 Replicated Differentially Expressed Genes (Mayo Clinic study KW $p < 10^{-10}$ and KW $p < 0.001$ in two GEO studies and pattern of association between three histologies consistent between GEO studies and Mayo Clinic Study). All results presented are based on the Mayo Clinic Study subjects. The full set of results for 1,254 genes with KW $p < 10^{-10}$ in Mayo Clinic study can be found in Supplemental Table 5, while results from the GEO replication studies can be found in Supplemental Table 6.

Gene	Normalized Mean Abundances			HGSC vs EC vs CC (KW test)		HGSC vs EC		HGSC vs CC		EC vs CC	
	HGSC	EC	CC	P-value	Q-value	P-value	Q-value	P-value	Q-value	P-value	Q-value
<i>ANXA4*</i>	4,733	3,059	32,385	3.3×10^{-12}	1.7×10^{-11}	5.4×10^{-2}	5.4×10^{-2}	7.7×10^{-12}	9.1×10^{-10}	1.0×10^{-10}	2.2×10^{-8}
<i>AP2A2</i>	1,923	1,906	6,274	8.3×10^{-11}	2.5×10^{-10}	7.3×10^{-1}	3.6×10^{-1}	3.1×10^{-11}	1.2×10^{-9}	7.5×10^{-10}	2.9×10^{-8}
<i>ATP6V0A1</i>	936	1,053	2,397	2.8×10^{-12}	1.5×10^{-11}	2.2×10^{-2}	2.7×10^{-2}	8.9×10^{-12}	9.1×10^{-10}	2.3×10^{-10}	2.2×10^{-8}
<i>CI2orf75</i>	1,964	661	9,382	1.5×10^{-14}	1.7×10^{-13}	3.0×10^{-6}	1.3×10^{-5}	4.7×10^{-11}	1.4×10^{-9}	5.9×10^{-10}	2.6×10^{-8}
<i>CDH1</i>	7,245	6,142	20,517	8.2×10^{-12}	3.6×10^{-11}	3.4×10^{-2}	3.8×10^{-2}	1.8×10^{-11}	9.3×10^{-10}	4.7×10^{-10}	2.3×10^{-8}
<i>CYB5R3</i>	2,512	2,715	8,536	6.0×10^{-12}	2.8×10^{-11}	2.1×10^{-1}	1.5×10^{-1}	4.9×10^{-12}	9.1×10^{-10}	1.0×10^{-10}	2.2×10^{-8}
<i>DRAM1</i>	657	566	2,031	4.4×10^{-11}	1.4×10^{-10}	2.5×10^{-1}	1.7×10^{-1}	4.3×10^{-11}	1.4×10^{-9}	3.1×10^{-10}	2.2×10^{-8}
<i>EIF1</i>	7,635	9,523	14,593	2.5×10^{-11}	8.9×10^{-11}	1.5×10^{-6}	7.3×10^{-6}	3.2×10^{-8}	1.4×10^{-7}	1.4×10^{-5}	5.1×10^{-5}
<i>ENTPD5</i>	670	638	2,034	1.2×10^{-11}	4.9×10^{-11}	7.9×10^{-1}	3.8×10^{-1}	5.6×10^{-12}	9.1×10^{-10}	1.0×10^{-10}	2.2×10^{-8}
<i>FERMT2</i>	1,627	1,351	3,934	2.9×10^{-11}	1.0×10^{-10}	5.1×10^{-2}	5.3×10^{-2}	2.2×10^{-10}	3.7×10^{-9}	1.0×10^{-10}	2.2×10^{-8}
<i>FLOT1*</i>	56	-5	75	1.7×10^{-11}	6.6×10^{-11}	3.4×10^{-11}	6.3×10^{-10}	1.2×10^{-1}	5.1×10^{-2}	3.4×10^{-6}	1.6×10^{-5}
<i>FOSL2</i>	6,195	5,696	14,950	6.0×10^{-11}	1.9×10^{-10}	2.8×10^{-1}	1.9×10^{-1}	7.9×10^{-11}	1.9×10^{-9}	2.1×10^{-10}	2.2×10^{-8}
<i>GABARAPL1</i>	1,643	985	9,612	4.9×10^{-13}	3.3×10^{-12}	1.5×10^{-3}	3.0×10^{-3}	1.8×10^{-11}	9.3×10^{-10}	1.3×10^{-10}	2.2×10^{-8}
<i>GLRX*</i>	559	350	4,809	2.6×10^{-12}	1.4×10^{-11}	6.9×10^{-2}	6.6×10^{-2}	4.3×10^{-12}	9.1×10^{-10}	1.0×10^{-10}	2.2×10^{-8}
<i>GNS</i>	6,235	7,616	16,582	3.8×10^{-11}	1.3×10^{-10}	4.1×10^{-3}	6.7×10^{-3}	2.3×10^{-10}	3.8×10^{-9}	1.6×10^{-8}	2.2×10^{-7}
<i>GPX3</i>	11,658	8,328	98,645	1.2×10^{-11}	5.1×10^{-11}	5.2×10^{-1}	2.9×10^{-1}	7.4×10^{-12}	9.1×10^{-10}	1.0×10^{-10}	2.2×10^{-8}
<i>IGSF3</i>	1,000	1,015	4,917	3.0×10^{-11}	1.0×10^{-10}	4.8×10^{-1}	2.7×10^{-1}	8.5×10^{-12}	9.1×10^{-10}	7.5×10^{-10}	2.9×10^{-8}
<i>KIAA1324</i>	1,117	3,612	-997	1.7×10^{-18}	5.8×10^{-17}	1.0×10^{-11}	2.0×10^{-10}	2.5×10^{-10}	3.9×10^{-9}	4.7×10^{-10}	2.3×10^{-8}
<i>LAMB1*</i>	5,670	3,168	29,521	2.4×10^{-12}	1.3×10^{-11}	1.3×10^{-2}	1.7×10^{-2}	1.6×10^{-11}	9.1×10^{-10}	1.7×10^{-10}	2.2×10^{-8}
<i>MAP2K6</i>	367	977	239	1.4×10^{-13}	1.2×10^{-12}	6.1×10^{-13}	1.8×10^{-11}	5.9×10^{-2}	2.8×10^{-2}	3.6×10^{-7}	2.5×10^{-6}
<i>MITF*</i>	422	409	1,964	3.1×10^{-11}	1.1×10^{-10}	7.1×10^{-1}	3.5×10^{-1}	1.5×10^{-11}	9.1×10^{-10}	2.3×10^{-10}	2.2×10^{-8}

Gene	Normalized Mean Abundances			HGSC vs EC vs CC (KW test)		HGSC vs EC		HGSC vs CC		EC vs CC	
	HGSC	EC	CC	P-value	Q-value	P-value	Q-value	P-value	Q-value	P-value	Q-value
<i>PNP</i>	1,411	1,268	4,591	4.5×10^{-11}	1.5×10^{-10}	1.1×10^{-1}	9.5×10^{-2}	3.3×10^{-11}	1.2×10^{-9}	1.6×10^{-9}	4.5×10^{-8}
<i>PPP1R3B</i>	1,462	799	6,511	1.4×10^{-13}	1.2×10^{-12}	2.2×10^{-4}	5.6×10^{-4}	2.2×10^{-11}	1.0×10^{-9}	1.8×10^{-10}	2.2×10^{-8}
<i>PRKAB1</i>	1,020	1,062	2,885	6.6×10^{-11}	2.0×10^{-10}	8.9×10^{-1}	4.0×10^{-1}	1.3×10^{-11}	9.1×10^{-10}	1.6×10^{-9}	4.5×10^{-8}
<i>RHOB</i>	4,443	5,020	21,197	1.7×10^{-11}	6.6×10^{-11}	2.0×10^{-1}	1.5×10^{-1}	8.9×10^{-12}	9.1×10^{-10}	5.9×10^{-10}	2.6×10^{-8}
<i>SESTD1</i>	956	1,084	3,087	4.1×10^{-11}	1.4×10^{-10}	3.0×10^{-1}	2.0×10^{-1}	1.3×10^{-11}	9.1×10^{-10}	1.8×10^{-9}	4.7×10^{-8}
<i>SGPP1</i> *	474	430	1,351	8.0×10^{-11}	2.4×10^{-10}	3.7×10^{-1}	2.3×10^{-1}	5.5×10^{-11}	1.6×10^{-9}	5.0×10^{-10}	2.4×10^{-8}
<i>SLC22A5</i>	379	346	1,001	3.7×10^{-11}	1.3×10^{-10}	3.3×10^{-1}	2.1×10^{-1}	3.4×10^{-11}	1.2×10^{-9}	2.1×10^{-10}	2.2×10^{-8}
<i>SLC3A1</i>	380	126	2,364	1.4×10^{-13}	1.1×10^{-12}	2.2×10^{-5}	7.7×10^{-5}	8.2×10^{-11}	2.0×10^{-9}	1.1×10^{-10}	3.6×10^{-8}
<i>TALDO1</i>	2,893	3,553	6,046	3.9×10^{-11}	1.3×10^{-10}	9.7×10^{-5}	2.8×10^{-4}	2.5×10^{-9}	2.0×10^{-8}	5.6×10^{-10}	3.5×10^{-8}
<i>TSPAN1</i>	3,065	1,629	11,969	3.1×10^{-12}	1.6×10^{-11}	1.5×10^{-4}	4.2×10^{-4}	8.7×10^{-10}	9.5×10^{-9}	1.6×10^{-10}	4.5×10^{-8}
<i>UPK3B</i>	297	1,416	1,381	4.4×10^{-12}	2.2×10^{-11}	1.4×10^{-11}	2.7×10^{-10}	8.6×10^{-5}	1.0×10^{-4}	3.3×10^{-1}	2.0×10^{-1}

* genes also detected by Zom et al (2005).

9 replicated progression-related genes (Mayo Clinic study $p < 0.0001$ and $p < 0.05$ in one GEO studies). All results presented are based on the Mayo Clinic Study subjects. The full set of results for 23 genes with $p < 0.0001$ in Mayo Clinic study can be found in Supplemental Table 7, while results from the GEO replication studies can be found in Supplemental Table 8.

Table 2:

Gene	Test for gene main effect*		Test for gene-histotype interaction effect*		HGSC [^]		EC [^]		CC [^]	
	p-value	q-value	p-value	q-value	HR	p-value	HR	p-value	HR	p-value
<i>CSNK1G1</i>	0.254	0.923	6.0×10^{-6}	0.018	HR < 1	0.044	HR > 1	3.1×10^{-5}		0.848
<i>LIN54</i>	0.377	0.939	1.0×10^{-5}	0.018	HR < 1	0.05	HR > 1	0.002		0.224
<i>FRMD8</i>	0.121	0.877	1.6×10^{-5}	0.021		0.057	HR > 1	0.001		0.326
<i>CORO2A</i>	0.356	0.935	3.5×10^{-5}	0.035		0.092	HR > 1	0.004		0.193
<i>LINC00664</i>	0.032	0.779	4.0×10^{-5}	0.037		0.276		0.127		0.848
<i>PDPK1</i>	0.660	0.961	6.3×10^{-5}	0.045	HR < 1	0.007	HR > 1	0.023		0.088
<i>CCNB2</i>	0.019	0.753	7.9×10^{-5}	0.045		0.14		0.274		0.16
<i>LSP31</i>	0.387	0.941	8.6×10^{-5}	0.045	HR < 1	3.8×10^{-4}	HR > 1	0.016		0.57
<i>PEX6</i>	0.133	0.888	8.7×10^{-5}	0.045		0.858	HR < 1	0.001	HR < 1	0.014

* adjusted for debulking status, stage (low/high) and histotype

[^] Additive gene effect from analysis by histotype with no adjustment for covariates. Due to limited sample size, direction of HR presented only for genes with $p < 0.05$