Gene networks and expression quantitative trait loci associated with platinum-based chemotherapy response in high-grade serous ovarian cancer

Jihoon Choi[1], Anastasiya Tarnouskaya[2], Sean Nesdoly[2], Danai G. Topouza[1], Madhuri Koti[1], Qing Ling Duan[1,2]

[1]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada; [2]School of Computing, Queen's University, Kingston, Ontario, Canada;

Correspondence: QL Duan (qingling.duan@queensu.ca)

Abstract

Background: A major impediment in the treatment of ovarian cancer is the relapse of platinum-resistant tumors, which occurs in approximately 25% of patients. A better understanding of the biological mechanisms underlying platinum-based response will improve treatment efficacy through genetic testing and novel therapies.

Methods: Using data from high-grade serous ovarian carcinoma (HGSOC) patients in the Cancer Genome Atlas (TCGA), we classified those who remained progression-free for 12 months following platinum-based chemotherapy as "chemo-sensitive" (N=160) and those who had recurrence within six months as "chemo-resistant" (N=110). Univariate and multivariate analysis of expression microarrays identified differentially expressed genes and co-expression gene networks associated with chemotherapy response. Moreover, we integrated genomics data to determine expression quantitative trait loci (eQTL).

Results: Differential expression of the Valosin-containing protein (VCP) gene and five co-expression gene networks were associated with chemotherapy response in HGSOC. VCP and the gene networks contribute to protein processing in the endoplasmic reticulum, which has been implicated in chemotherapy response. These findings were successfully replicated using independent replication cohort. Furthermore, 192 QTLs were associated with these gene networks and *BRCA2* expression.

Conclusion: This study implicates both known and novel genes as well as biological networks underlying response to platinum-based chemotherapy among HGSOC patients.

Introduction

Ovarian cancer is the most lethal gynecological malignancy and the 8th leading cause of cancer death in women around the world.[1] According to the Global Cancer Observatory report in 2012, ovarian cancer accounts for 3.6% of all cancer cases and 4.3% of all cancer related deaths worldwide.[2] High-grade serous ovarian carcinoma (HGSOC) is the most malignant form of ovarian cancer that accounts for up to 70% of all cases.[3] Routine diagnosis is often difficult due to the lack of mass screening methods and heterogenous manifestations of the cancer symptoms, which result in approximately 75% of patients diagnosed with advanced stages.[4] The average 5-year survival rates are 39% for Stage 3 and 17% for Stage 4 cancers.[1]

The current standard of care for ovarian cancer is aggressive cytoreductive surgery followed by platinum-based chemotherapy.[4] However, this standard of care is not effective for all patients, with approximately 25% experiencing relapse within six months following platinum-based therapy, likely due to the development of antineoplastic resistance.[5] The median survival time for recurrent ovarian cancers range from 12-24 months.[6,7] Treatment options for patients with recurrent ovarian cancer include non-platinum-based chemotherapy regimens, immunotherapy, and molecular targeted therapy.[7,8]

Ovarian cancer has a multifactorial etiology that includes genetic and non-genetic risk factors. An estimated 23% of cases are hereditary, but the majority are sporadic with multiple reported risk factors such as history of gravidity, infertility, and late age menopause.[9,10] A better understanding of the etiology of ovarian cancer, as well as the genetic mechanisms underlying variable response to platinum-based chemotherapy is needed for improved diagnosis and treatment. For example, previous studies reported that *BRCA1* and

BRCA2 genes, associated with increased risk of ovarian cancer, harbor mutations associated with platinum drug sensitivity and survival time.[11] Similarly, tumor suppressor genes such as *RB1*, *NF1*, *RAD51B*, *PTEN* have been associated with acquired chemotherapy resistance.[12] Earlier studies have also highlighted the importance of the immune system in the treatment of ovarian cancer. For example, loss of chemokines and disruptions to the IFN-γ pathway have been associated with poor treatment outcomes in HGSOC paients[13] whereas the NFκB signaling pathway and elevated expression of *STAT1* were associated with increased response to platinum therapy.[14,15,16] However, these known genetic variations do not account for all of the variability in chemotherapy response among HGSOC patients and there is currently no screening method to accurately predict prognosis prior to start of chemotherapy. Thus, further studies are necessary to determine additional modulators of chemotherapy response, which can be used as biomarkers for genetic testing.

The majority of earlier studies of chemotherapy response in ovarian cancer patients used univariate analysis of gene expression data known as differential gene expression (DGE) analysis.[17,18] This analysis assumes that each gene functions in isolation within the genome and fails to capture the effects of complex gene-gene interactions. This study applies a multivariate approach to identify groups of interconnected genes, which may contribute to common biological pathways. These genes may each have modest effects that are not detected by conventional univariate analysis. Specifically, we applied Weighted Gene Co-expression Network Analysis[19] (WGCNA), which uses an unsupervised machine-learning algorithm to identify clusters of highly correlated or co-expressed genes. Moreover, we correlated sequence variations with co-expressed gene networks to identify expression Quantitative Trait Loci (eQTLs), which are potentially regulatory variants associated with co-expression gene networks. A better understanding of the biological mechanisms regulating

chemotherapy response will enable more effective treatments by improving the accuracy of genetic testing and through the identification of novel therapies for cancer patients.

Methods

Patient Classification:

From the Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) portal[20], we retrieved 587 high-grade serous ovarian carcinoma (HGSOC) patients with available clinical data using the TCGAbiolinks R/Bioconductor package.[21] The interval between patient's last primary platinum treatment and the onset of a recurrent tumor or progression of an existing tumor was used as a metric for determining chemotherapy sensitivity. Patients who developed a new tumor in less than six months following their last platinum treatment were defined as resistant (N=110). In contrast, those who did not have a recurrent tumor event for over a year after their last primary platinum treatment were defined as sensitive (N=160). Individuals who had a recurrent tumor event between six months to one year following chemotherapy were excluded from the study. This strategy for dichotomizing resistant from sensitive patients was used to enrich for the genetic differences.

Transcriptomics Data Processing and Analysis:

Expression Microarrays: Of the 270 HGSOC subjects classified as sensitive or resistant to chemotherapy, 238 (138 sensitive, 100 resistant) had microarray expression data available (Affymetrix ht_hg_u133a chip) in the GDC portal. The robust multi-array average (RMA) method[22] in the *affy* package from Bioconductor[23] was used for background correction, log-transformation, and quantile normalization of the probe intensities. Two potential outliers and two duplicated samples were removed from the study during the quality control step using the *arrayQualityMetrics*[24] package (see **Supplemental Data 1** for steps of pre-processing), resulting in 135 sensitive and 99 resistant HGSOC subjects in the expression set. Next, probes were filtered using the median absolute deviation (MAD) whereby the top 50% with

highest variation (n=11,107) were selected for analysis. Thus, this non-specific filtering step removed probes with no or low variability, which are not likely to be differentially expressed between sensitive and resistant patients. Moreover, this step reduces the number of multiple testing corrections and therefore, the likelihood of false positives.

Differential Gene Expression Analysis: The *Limma*[25] package in Bioconductor[26] was used to identify differentially expressed genes between chemo-sensitive and resistant groups using linear models. The false discovery rate (FDR) method was employed as a measure for multiple testing correction to control for type I error. Age at diagnosis and stage of tumor were included as covariates in this analysis (**Supplemental Data 2**).

Weighed Gene Co-expression Network Analysis (WGCNA): We performed hierarchical clustering of genes using the R package *WGCNA*[19], which groups genes based on their similarity in expression. This was achieved by first creating a similarity matrix using Pearson correlations of expression among all genes. The resulting matrix was raised to a power of 9, as suggested by the soft-thresholding power estimation plot (**Supplemental Figure 1**). This assumes that raising the correlation matrix to a power will enrich differences between weak and strong signals, allowing for better quantification of gene-gene interactions. In addition, because information between two genes are often too weak and noisy to interpret, similarity matrix is transformed to Topological Overlap Matrix (TOM), where strength of association between a pair of genes is reinforced by the common neighbors shared by them. To avoid excessive splitting of genes into smaller modules, minimum module size was set to 30, split sensitivity (deep split) to 4, and modules with similar expression profiles were merged at a height of 0.5 (**Supplemental Figure 2**). Using principal component analysis, we calculated the module eigengene for each co-expression cluster to summarize module expression into a single measure. Each module eigengene was tested for association with platinum

chemotherapy response using generalized linear models, while adjusting for age at diagnosis and stage of cancer as covariates. Finally, we used *Cytoscape*[27], an open source bioinformatics platform, to visualize gene co-expression networks.

Gene function and pathway annotations: The Database for Annotation, Visualization and Integrated Discovery (DAVID) [28] was applied to identify biological pathways and functions that were enriched in each significant gene co-expression module. We also screened significant genes in the GeneMANIA[29] database to identify for functional connections reported in published literature. Next, we searched the UCSC TFBS conservation sites track with DAVID to identify enriched motifs of transcription factors that may co-regulate genes within each cluster. Finally, we used Drug–Gene Interaction database (DGIdb)[30], a public database with curation of data describing relationships between genes, chemicals, drugs, and pathological phenotype, to identify genes with prior reported associations with chemotherapeutic agents.

Validation of differentially expressed gene: The Kaplan–Meier plotter was used to cross-validate the effect of differentially expressed genes in an independent ovarian cancer cohort (Geo accession identifier: GSE9891).[31] This replication cohort included gene expression profiling (Affymetrix Human Genome U133 Plus 2.0 Array) of 285 ovarian tumor samples. Patients were filtered to include those with cancer histology of serous carcinoma and who received chemotherapy containing a platin compound to allow close comparison with the TCGA ovarian cancer cohort. This step omitted a total of 60 subjects from analysis, which included 21 with endometrioid carcinoma cases and 43 who did not receive platinum therapy (4 overlapping subjects). Thus, 225 patients remained for replication analysis. Patient survival was evaluated using Cox proportional hazards model and progression-free survival (PFS) was the primary outcome used in the replication analysis.[32]

Validation of co-expression network:

For validation of co-expression networks, SurvExpress database was used, which allow users to validate the combined effect of multiple gene expression measures with a target trait.[33] Same cohort and filtering steps were used for validation (GSE9891, N = 225). Survival curve was evaluated using Cox proportional hazards model and progression-free survival (PFS) was the primary outcome used in the replication analysis.

Genomics Data Processing and Analysis:

Genomics Data: Single nucleotide polymorphisms (SNPs) data from the germline tissues (DNA extracted from blood or solid non-tumor ovarian tissue) were obtained from the TCGA legacy database. The Affymetrix Genome-Wide Human SNP Array 6.0 was used to capture genetic variations, which detected 906,600 SNPs. Of the 270 subjects from TCGA classified as resistant or sensitive to platinum-based chemotherapy, 266 (157 sensitive and 109 resistant) had genotype data available.

Imputations: The imputation of autosomal chromosomes was performed using the Michigan imputation server pipeline[34]. We used the 1000 Genome Project phase 3 sequencing data (version 5)[35] reference panel for the imputation of missing genotypes. We then used Eagle v.2.3[36] for phasing of the genotypes to their respective chromosomes. For the imputation of variants on the X chromosome, SHAPEIT[37] was used for phasing in combination with the 1000 genomes project phase 3 (version 5) reference panel (**Supplemental Data 3**).

Quality control:

Subject level: Two pairs of individuals had a relatedness coefficient (pi-hat) > 0.9, which are likely duplicated samples. One subject from each pair was randomly removed from the

dataset. Next, inbreeding coefficients (F) were computed for each subject using PLINK[38]. A total of 18 subjects with high homozygosity (F>0.05) or heterozygosity (F<-0.05) rates were excluded. Moreover, genetic sex was estimated based on heterozygosity rates (F) of the X chromosome, and four subjects who had undefined (F>0.2) genetic sex were removed from the study.

SNP level: SNPs with minor allele frequencies (MAF) less than 1% or with genotyping call rate less than 90% were removed. This step removed 38,430,595 SNPs with MAF < 0.01, resulting in 9,528,963 SNPs to be used for further analysis.

Genome-wide Association Study: After imputations and quality control, 240 subjects (N= 142 sensitive, 98 resistant) and a total of 9,528,963 SNPs (MAF > 0.1) remained available for analysis (**Supplemental Data 4**). Plink (v.1.90) was used to compute genome wide and BRCA1/2 targeted association analysis using a logistic regression model. Age at diagnosis and tumor stages were included as covariates.

Expression Quantitative Trait Loci (eQTL) Analysis:

Common SNPs (MAF > 0.01) were tested for association with differentially expressed genes and gene networks using the *matrixeQTL* R package[39]. Loci correlated with nearby gene expression indicates potential regulatory function of the SNP on the corresponding gene, known as cis-expression Quantitative Trait Loci (cis-eQTL). Cis-eQTLs are defined as correlated SNPs found within 1 Mb from the gene transcriptional start site (TSS).

Results

Differential gene expression (DGE) analysis identified that low expression of a probe

(208648_at) mapping to the Valosin Containing Protein (*VCP*) gene was significantly associated with resistance to chemotherapy (FDR adjusted p-value < 0.05; **Figure 1**). In addition, 606 probes mapping to 521 unique genes were nominally correlated with chemotherapy response (unadjusted p-value < 0.05; **Supplemental Table 1**). The Kaplan–Meier survival plotter demonstrated that low expression of VCP is associated with poor progression-free survival (p = 0.015) and shorter median survival time (**Figure 2**) in an independent ovarian serous cancer cohort following treatment with platinum antineoplastic agents. Functional annotation analysis showed that the 521 genes were enriched for numerous oncogenic processes, such as cellular response to DNA damage stimulus (GO:0006974), DNA repair (GO:0006281), positive regulation of cell growth (GO:0030307), and positive regulation of apoptotic process (GO:0043065). Pathway analysis identified many immune related pathways including the mitogen-activated protein kinase (MAPK) and B-Cell antigen receptor (BCR) signaling pathways. A detailed list of functional annotations and pathways are shown in **Supplemental Data 5**.

The hierarchical clustering of genes using *WGCNA* resulted in 86 unique modules of co-expressed genes (**Supplemental Table 2**). Each module was assessed for association with chemotherapy response (results shown in **Supplemental Figure 3**). Five gene clusters (*honeydew1, lightcyan1, lightpink3, orangered4, and skyblue3*) were co-downregulated in platinum-resistant patients (p < 0.05); one gene cluster remained significant after Bonferroni correction (*honeydew1*, p-value=7e-05) (**Figure 3**). This signal was validated using SurvExpress in an independent cohort (Tothill et al.), which demonstrates that downregulation of the genes in five modules result in significantly reduced survival of patients (**Supplemental Figure 4**). The five significant modules were annotated using DAVID, which identified gene enrichment for biological pathways including apoptosis,

negative regulation of the Wnt signaling pathway, transcription, immune responses, and DNA double-strand break processing involved in repair via single-strand annealing. GeneMANIA analysis showed that genes in these modules were previously reported in 49 publications, some of which documented associations with oncogenic pathways and chemotherapeutic outcomes. Our search of these genes in the gene-drug interaction database (DGIdb) found that 35 were associated with chemotherapeutic agents such as carboplatin and paclitaxel. Furthermore, we identified over-represented conserved transcription factor binding sites located in genes from each module. A detailed list of functional annotations, transcription factors and pathways related to gene modules can be found in **Supplemental Data 6.**

Our GWAS of SNPs did not identify any variants correlated for platinum chemotherapy response after multiple testing correction. The Manhattan plot (**Supplemental Figure 5**), demonstrates that none of the SNPs meet the genome-wide significance threshold (p<5x10-8), as indicated in the by the red horizontal line. This is likely due to insufficient statistical power resulting from low number of subjects in TCGA-OV cohort. Next, we performed a targeted association analysis of two well-known genes associated with ovarian cancer and chemotherapeutic outcomes: *BRCA1* and *BRCA2*. Of the 238 SNPs in *BRCA1* and 256 in *BRCA2*, we identified 34 SNPs in *BRCA2* and 1 SNP in *BRCA1* that were significantly associated with chemotherapy response (**Supplemental Table 3**).

Next, SNPs were tested for correlation with expression of the 606 differentially expressed probes and 5 co-expression modules. This identified 5,242 cis-eQTL associated with the expression of differentially expressed probes, and 192 cis-eQTL associated with co-expression networks (**Supplemental Data 7**).

Discussion

In this manuscript, we identified novel genes and gene networks correlated with variable response to platinum-based chemotherapy in HGSOC patients. Using a univariate analysis approach, we identified a differentially expressed locus encoding the valosin-containing protein (*VCP*) associated with sensitivity to platinum-based chemotherapy. In addition, we applied a multivariate, co-expression analysis method for identifying groups of interconnected genes that could contribute to common biological pathways. This method identified 5 clusters of co-expressed genes nominally correlated with chemo-response, one of which remained significant after correction of multiple testing. Genes in these modules have been annotated to be associated with biological pathways such as apoptosis, transcription, immune response, negative regulation of the Wnt signaling pathway and DNA double-strand break processing involved in repair via single-strand annealing.

The most significantly associated probe identified in the DGE analysis was for a gene encoding Valosin-containing protein (*VCP*, p = 3.91E-06). We have confirmed that this signal is replicated in an independent ovarian cancer cohort with statistical significance (p=0.015). *VCP* plays a critical role in disintegrating large polypeptide cellular structures for further degradation by proteolytic enzymes. It functions to regulate important pathways of DNA repair, replication and cell cycle progression by removing faulty polypeptide structures from chromatin material, ribosomes, endoplasmic reticulum and mitochondria. This gene has been previously identified as a potential biomarker for predicting the success of platinum-based chemotherapy in lung cancer patients.[40]

Many of the nominally correlated genes from our DGE analysis are enriched for pathways known to be critical for oncogenesis and chemotherapeutic drug resistance. For

example, probe 213532_at mapping to the A Disintegrin and Metalloproteinase-17 (*ADAM17*) gene was upregulated in the chemo-resistant group (p = 0.017). This is consistent with prior studies, which reported that overexpression of *ADAM17* results in reduced cisplatin-induced apoptosis in hepatocellular carcinoma cells and may contribute to cisplatin resistance via the EGFR pathway[41]. Moreover, we identified another up-regulated probe 205239_at (p = 0.003) for the gene encoding Amphiregulin (AREG), a protein found in the EGFR signaling pathway, which has been reported to promote ovarian cancer stemness and drug resistance to anti-cancer therapy.[42] Taken together, our findings support a potential role of the AREG-EGFR signaling pathway in the development of platinum-resistance in ovarian carcinoma.

It is also important to note that one of the most enriched pathways from our DGE analysis was the protein processing in endoplasmic reticulum (hsa04141) pathway. Using DAVID, we have identified 18 differentially expressed probes mapping to 15 unique genes within this pathway. Prior studies have highlighted that endoplasmic reticulum (ER) stress may cause cisplatin resistance in ovarian carcinoma by inducing autophagy in cancer cells, allowing them to escape apoptosis.[43,44] Findings from this pathway analysis suggest that future studies of the contribution of ER stress in cisplatin sensitivity is needed to improve our understanding of platinum-resistance in ovarian cancer.

In our co-expression network analysis, the gene module "*honeydew1*" showed the most significant correlation with chemotherapy response (p = 6.53e-05). This association signal was validated with statistical significance (p = 5.88e-07) in an independent replication cohort. This module includes two probes that map to *VCP*, a gene that was associated with platinum-based chemotherapy response in our DGE analysis. Additional genes in this module were associated with positive regulation of mitochondrial membrane potential, protein ubiquitination, mitosis, alternative splicing, and apoptotic processes. Pathway analysis

showed that this module is involved in protein processing in the endoplasmic reticulum. A prior study has found that *VCP* plays a crucial role in cancer cell survival through extraction and degradation of unfolded proteins in endoplasmic reticulum, and noted that lower expression of *VCP* was associated with poor response to platinum based chemotherapy.[45] In alignment with this finding, genes co-expressed in the *honeydew1* module were co-downregulated in chemo-resistant patients.

The *honeydew1* module is composed of 76 probes mapping to 54 unique genes, and of these, 44 genes are located in chromosome 9, demonstrating the importance of chromosome 9 in the regulation of chemo-resistance in ovarian cancer. These findings support previous studies, where genetic imbalance and alterations in chromosome 9 have been associated with progression of ovarian cancer and increased cisplatin resistance.[46] Analysis of overrepresented transcription factor binding sites demonstrated that genes in this module may be co-regulated by a common transcription factor known as organic cation transporter 1 (OCT1). We found that over 96% of genes in this module (49/54 genes) contain a nucleotide motif bound by OCT1. Prior studies have reported that silencing OCT1 impaired cisplatin-induced apoptosis in esophageal cancer cells, and that cisplatin-resistant cells were already expressing significantly reduced levels of OCT1.[47] Taken together, these findings characterize a network of co-expressed genes that is associated with platinum-resistance in ovarian cancer. Genes within this module may be co-regulated by the OCT1 transcription factor, which may be used as a novel potential target for ovarian cancer therapies

The other four co-expression modules, which we were also able to replicate in an independent cohort, include genes known to be involved in oncogenic process and drug response outcomes. For example, the *orangered4* module is composed of genes that are critical for regulation of immune response, which have been shown to play a role in

chemotherapy response in HGSOC.[14,15] Genes in this module are associated with functional annotation terms including immunoglobulin receptor binding, antigen binding, B cell receptor signaling pathway, and phagocytosis. In addition, 10 of the 58 genes in this module are enriched for a common transcription factor binding site: acute myeloid leukemia 1 protein (AML1). This transcription factor is involved in haematopoiesis process and immune functions such as thymic T-cell development. Studies have reported that the AML1 transcription factor is overexpressed in ovarian cancer patients, and may contribute to cancer cell proliferation, migration and invasion.[48] In addition, we found that the *lightpink3* module is strongly associated with transcription regulation process, which plays a pivotal role in cancer progression. Finally, genes in modules *lightcyan1* and *skyblue3* are target regions of chemotherapeutic agents such as tyrosine kinase inhibitors (**Supplemental data 6**).

Targeted analysis of *BRCA1* and *BRCA2* SNPs demonstrated that 28 out of 35 variants associated with chemotherapy response were also cis-acting eQTLs, correlated with the expression of *BRCA2* as well as neighboring genes *N4BP2L1*, *N4BP2L2*, *FRY*, and *STARD13* (nominal p-value <0.05). Both *BRCA2* and *STARD13* are well known tumor-suppressors, and upregulation of *N4BP2L1* and *N4BP2L2* have been associated with positive prognosis in ovarian cancer cases.[49] This finding shows the potential regulatory effect of the variants in *BRCA2*.

In addition, our annotation results show that 47% of platin-response associated variants in *BRCA2* are linked with LDL/HDL cholesterol levels (**Supplemental Table 3**). Prior studies of lung and ovarian cancers consistently reported that cholesterol levels may affect the efficacy of platinum-based chemotherapeutic agents.[50,51] Our findings indicate a new link between genetic variants in *BRCA2* and platinum-based chemotherapy response through cholesterol level regulation.

Conclusion

In this study, we identified genes and gene networks correlated with platinum-based chemotherapy response in high-grade serous ovarian cancer patients, which implicate known and novel biological mechanisms. Specifically, we identified that reduced expression of *VCP* is associated with platinum-resistance. This gene is critical for removing unfolded proteins from the endoplasmic reticulum and have been correlated with cancer cell survival and platinum-based chemotherapy response in prior studies. In addition, we report potentially regulatory variants in the *BRCA2* gene correlated with chemotherapy response and the expression of genes that determine cholesterol levels. Moreover, we identified a novel group of potentially co-regulated genes on chromosome 9 that are correlated with platinum resistance using a machine-learning algorithm. In addition, genes from this module are also involved in protein processing in the endoplasmic reticulum. This manuscript supports earlier studies which implicated *VCP* and *BRCA2* genes in chemotherapy response. We also identified regulatory variants in *BRCA2* and additional genes co-regulated with *VCP* on chromosome 9 that also contribute to protein removal in the ER. Findings from our study could facilitate genetic testing through the identification of gene signatures that may predict chemotherapy response as well as lead to novel drug targets, given a better understanding of the biological mechanisms underlying chemotherapy response.

Figure Legends

Figure 1. Differential expression analysis of platinum-based chemotherapy response in HGSOC patients. Volcano plot showing univariate association analysis results. One probe,

208648_at, which maps to the Valosin-Containing Protein (*VCP*) gene is significantly differentially expressed and correlated with chemotherapy outcome after multiple testing correction. Red horizontal line demonstrates FDR-corrected p value (< 0.05) threshold. A total of 606 probes mapping to 521 unique genes are also differentially expressed with nominal significance, as indicated by the green line (p = 0.05).

Figure 2. Kaplan- Meier plot of two ovarian cancer cohorts for cross-validation of differentially expressed gene VCP. Kaplan-Meier (KM) plot shows the Progression Free Survival (PFS) of the replication ovarian cancer cohort post platinum-based chemotherapy treatment. Red line indicates the PFS of patients with high VCP expression, while black line indicates the PFS of patients with low VCP expression. Patients with high expression of VCP is associated with better PFS with statistical significance.

Figure 3. Gene co-expression modules correlated with platinum-based chemotherapy response. Network plot displaying the five significant gene co-expression modules from *WGCNA*: honeydew1 - centre, lightcyan1 - left, lightpink3 - top, orangered4 - bottom, and skyblue3 - right. Nodes represent probes and edges are connections among the probes. Co-expressed probes (i.e. belonging to a single module) are indicated by the same color.

Acknowledgements

Computations in this manuscript were performed on resources and with support provided by the Centre for Advanced Computing (CAC) at Queen's University in Kingston, Ontario. The CAC is funded by: the Canada Foundation for Innovation, the Government of Ontario, and Queen's University. JC is funded by Queen's University, Faculty of Health Sciences Dean's Doctoral Award. QLD receives funding from the Canadian Institutes of Health Research and

Queen's University.

Conflict of interest

The authors declare no conflicts of interest.

Author contribution

J.C. performed the data analyses and drafted the manuscript. D.G.T., S.N. and A.T. assisted in the data analyses. Q.L.D. designed the research project, supervised data analyses and assisted in the writing of the manuscript. M.K. assisted in the study design and editing of the manuscript.

Data availability

Transcriptomics, genomics, and clinical data used for the analysis of TCGA-OV cohort can be accessed/downloaded from the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/). Gene expression and clinical data of the replication cohort can be accessed/downloaded from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9899).

References

1. Reid, F. World Ovarian Cancer Coalition 2018. *World Ovarian Cancer Coalit.* (2018).
2. International Agency for Research on Cancer Website. Globocan 2012 - Home. *Webpage* (2012). doi:NO:11
3. Brett M., R. *et al.* Epidemiology of ovarian cancer: a review. *Cancer Biol. Med.* (2017). doi:10.20892/j.issn.2095-3941.2016.0084
4. Cannistra, S. A. Cancer of the Ovary. *N. Engl. J. Med.* **351**, 2519–2529 (2004).

5. Miller, D. S. *et al.* Phase II Evaluation of Pemetrexed in the Treatment of Recurrent or Persistent Platinum-Resistant Ovarian or Primary Peritoneal Carcinoma: A Study of the Gynecologic Oncology Group. *J. Clin. Oncol.* **27**, 2686–2691 (2009).

6. Armstrong, D. K. Relapsed ovarian cancer: challenges and management strategies for a chronic disease. *Oncologist* **7 Suppl 5**, 20–28 (2002).

7. Foley, O. W., Rauh-Hain, J. A. & del Carmen, M. G. Recurrent epithelial ovarian cancer: an update on treatment. *Oncology (Williston Park).* **27**, 288–94, 298 (2013).

8. Ozols, R. F. Recurrent ovarian cancer: Evidence-based treatment. *Journal of Clinical Oncology* **20**, 1161–1163 (2002).

9. Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl. Acad. Sci.* (2011). doi:10.1073/pnas.1115052108

10. Booth, M., Beral, V. & Smith, P. Risk factors for ovarian cancer: A case-control study. *Br. J. Cancer* (1989). doi:10.1038/bjc.1989.320

11. Vencken, P. M. L. H. *et al.* Chemosensitivity and outcome of BRCA1- and BRCA2-associated ovarian cancer patients after first-line chemotherapy compared with sporadic ovarian cancer patients. *Ann. Oncol.* (2011). doi:10.1093/annonc/mdq628

12. Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* (2015). doi:10.1038/nature14410

13. Hao, D. *et al.* Immunogenomic analyses of advanced serous ovarian cancer reveal immune score is a strong prognostic factor and an indicator of chemosensitivity. *Clin. Cancer Res*. (2018). doi:10.1158/1078-0432.CCR-17-3862

14. Koti, M. *et al.* A distinct pre-existing inflammatory tumour microenvironment is associated with chemotherapy resistance in high-grade serous epithelial ovarian cancer. *Br. J. Cancer* (2015). doi:10.1038/bjc.2015.81

15. Au, K. K. *et al.* STAT1-associated intratumoural TH1 immunity predicts chemotherapy resistance in high-grade serous ovarian cancer. *J. Pathol. Clin. Res*. (2016). doi:10.1002/cjp2.55

16. Koti, M. *et al.* Identification of the IGF1/PI3K/NF κB/ERK gene signalling networks associated with chemotherapy resistance and treatment response in high-grade serous epithelial ovarian cancer. *BMC Cancer* **13**, 549 (2013).

17. Cancer, T. & Atlas, G. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–15 (2011).

18. Li, J., Wood, W. H., Becker, K. G., Weeraratna, A. T. & Morin, P. J. Gene expression response to cisplatin treatment in drug-sensitive and drug-resistant ovarian cancer cells. *Oncogene* **26**, 2860–2872 (2007).

19. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

20. Grossman, R. L. *et al.* Toward a shared vision for cancer genomic data. *New England*

Journal of Medicine (2016). doi:10.1056/NEJMp1607591

21. Colaprico, A. *et al.* TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkv1507

22. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

23. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).

24. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009).

25. Smyth, G. limma: Linear Models for Microarray Data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (2005). doi:citeulike-article-id:5722720

26. Gentleman, R. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

27. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

28. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

29. Warde-Farley, D. *et al.* The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, (2010).

30. Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1143

31. Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* (2008). doi:10.1158/1078-0432.CCR-08-0196

32. Gyorffy, B., Lánczky, A. & Szállási, Z. Implementing an online tool for genomewide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr. Relat. Cancer* (2012). doi:10.1530/ERC-11-0329

33. Aguirre-Gamboa, R. *et al.* SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS One* (2013). doi:10.1371/journal.pone.0074250

34. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* (2016). doi:10.1038/ng.3656

35. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

36. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium

panel. *Nat. Genet.* (2016). doi:10.1038/ng.3679

37. Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* (2014). doi:10.1038/ncomms4934

38. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

39. Shabalin, A. A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).

40. Peng, J. *et al.* VCP gene variation predicts outcome of advanced non-small-cell lung cancer platinum-based chemotherapy. *Tumor Biol.* **34**, 953–961 (2013).

41. Wang, X. J., Feng, C. W. & Li, M. ADAM17 mediates hypoxia-induced drug resistance in hepatocellular carcinoma cells through activation of EGFR/PI3K/Akt pathway. *Mol. Cell. Biochem.* (2013). doi:10.1007/s11010-013-1657-z

42. Tung, S. L. *et al.* miRNA-34c-5p inhibits amphiregulin-induced ovarian cancer stemness and drug resistance via downregulation of the AREG-EGFR-ERK pathway. *Oncogenesis* (2017). doi:10.1038/oncsis.2017.25

43. Tian, J., Liu, R. & Qu, Q. Role of endoplasmic reticulum stress on cisplatin resistance in ovarian carcinoma. *Oncol. Lett.* (2017). doi:10.3892/ol.2017.5580

44. Xu, Y. *et al.* Tolerance to endoplasmic reticulum stress mediates cisplatin resistance in human ovarian cancer cells by maintaining endoplasmic reticulum and mitochondrial homeostasis. *Oncol. Rep.* (2015). doi:10.3892/or.2015.4283

45. Bastola, P., Neums, L., Schoenen, F. J. & Chien, J. VCP inhibitors induce endoplasmic reticulum stress, cause cell cycle arrest, trigger caspase-mediated cell death and synergistically kill ovarian cancer cells in combination with Salubrinal. *Mol. Oncol.* (2016). doi:10.1016/j.molonc.2016.09.005

46. Devlin, J., Elder, P. A., Gabra, H., Steel, C. M. & Knowles, M. A. High frequency of chromosome 9 deletion in ovarian cancer: Evidence for three tumour-suppressor loci. *Br. J. Cancer* (1996). doi:10.1038/bjc.1996.75

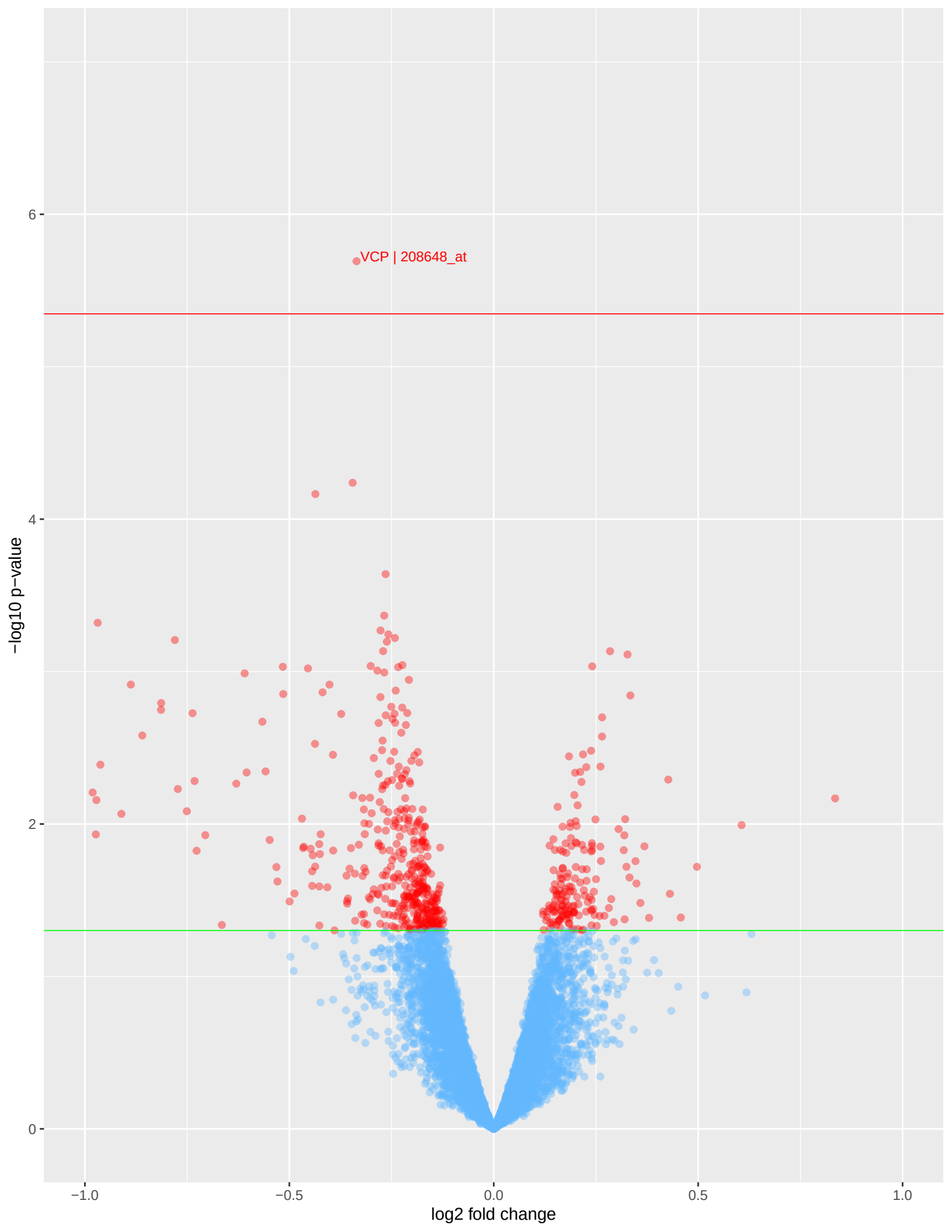
47. Lin, R. *et al.* Long-term cisplatin exposure promotes methylation of the OCT1 gene in human esophageal cancer cells. *Dig. Dis. Sci.* (2013). doi:10.1007/s10620-012-2424-9

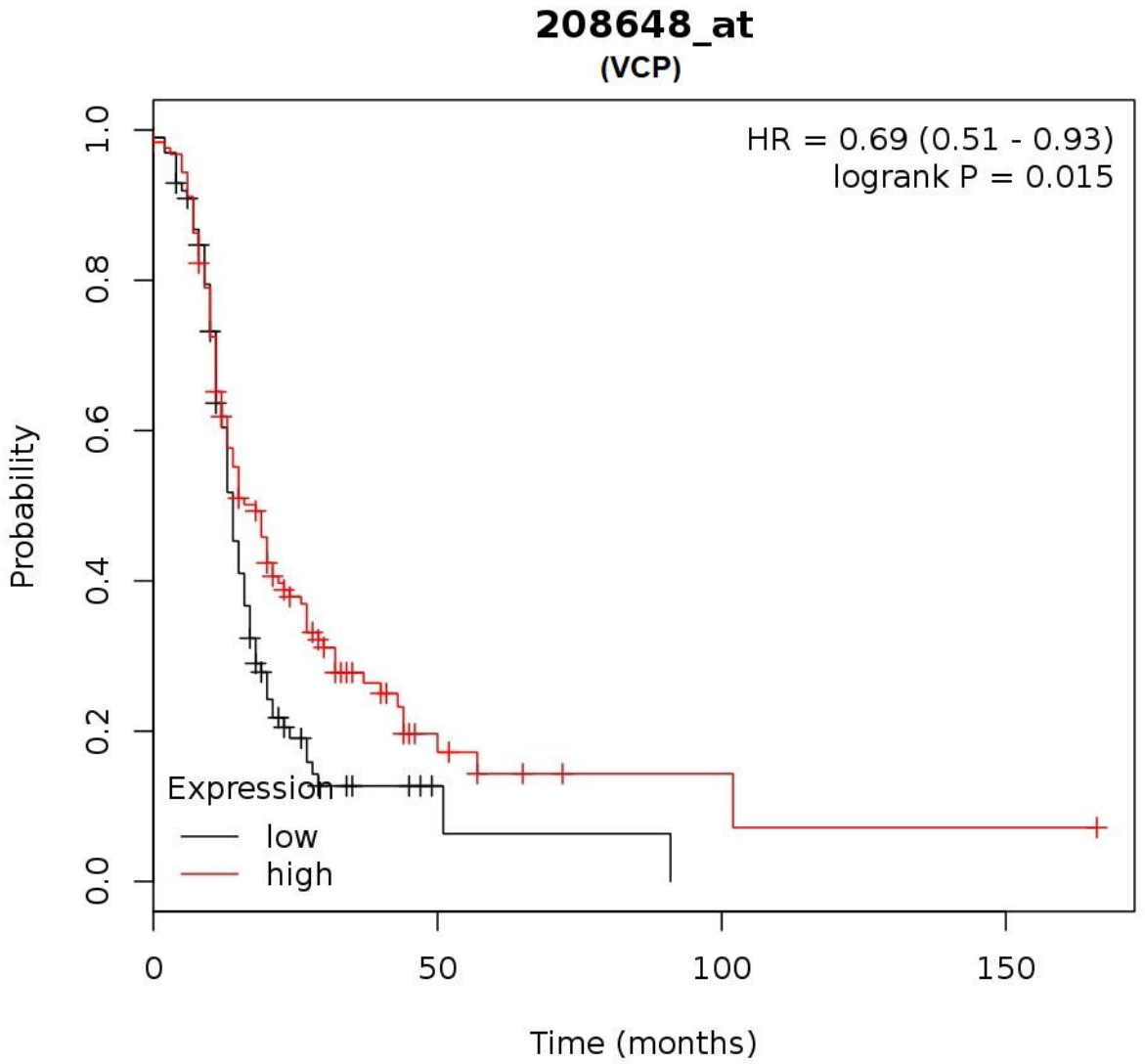
48. Keita, M. *et al.* The RUNX1 transcription factor is expressed in serous epithelial ovarian carcinoma and contributes to cell proliferation, migration and invasion. *Cell Cycle* (2013). doi:10.4161/cc.23963

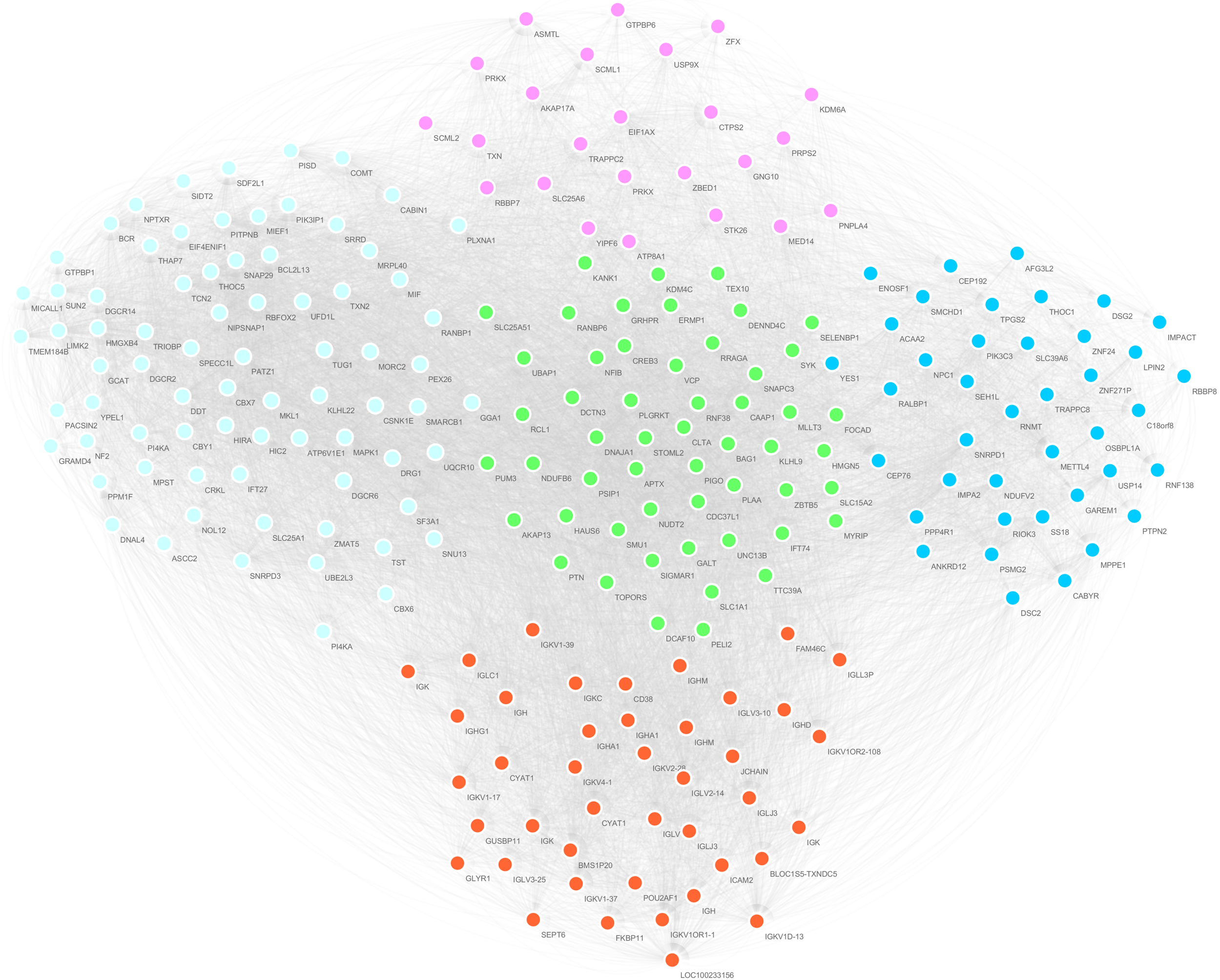
49. Koussounadis, A., Langdon, S. P., Harrison, D. J. & Smith, V. A. Chemotherapy-induced dynamic gene expression changes in vivo are prognostic in ovarian cancer. *Br. J. Cancer* **110**, 2975–2984 (2014).

50. Wu, Y. *et al.* Cholesterol reduces the sensitivity to platinum-based chemotherapy via upregulating ABCG2 in lung adenocarcinoma. *Biochem. Biophys. Res. Commun.* (2015). doi:10.1016/j.bbrc.2015.01.035

51. Kim, S., Lee, M., Dhanasekaran, D. N. & Song, Y. S. Activation of LXR□/$β$ by cholesterol in malignant ascites promotes chemoresistance in ovarian cancer. *BMC Cancer* **18**, 1232 (2018).


VCP | 208648_at
6
4
2
0
−log10 p−value
−1.0
−0.5
0.0
0.5
1.0
log2 fold change


208648_at
(VCP)
HR = 0.69 (0.51 - 0.93)
logrank P = 0.015
Probability
0.0
0.2
0.4
0.6
0.8
1.0
Expression
low
high
0
50
100
150
Time (months)


ASMTL
GTPBP6
ZFX
SCML1
USP9X
PRKX
AKAP17A
KDM6A
EIF1AX
CTPS2
SCML2
TXN
TRAPPC2
PRPS2
PISD
COMT
GNG10
SDF2L1
SIDT2
ZBED1
PRKX
RBBP7
SLC25A6
CABIN1
NPTXR
PIK3IP1
PNPLA4
STK26
MED14
BCR
PITPNB
MIEF1
SRRD
PLXNA1
YIPF6
EIF4ENIF1
ATP8A1
THAP7
MRPL40
AFG3L2
GTPBP1
SNAP29
BCL2L13
KANK1
CEP192
THOC5
KDM4C
TEX10
ENOSF1
TCN2
MIF
MICALL1
SUN2
DGCR14
TXN2
SMCHD1
THOC1
DSG2
RBFOX2
UFD1L
TPGS2
NIPSNAP1
GRHPR
ERMP1
DENND4C
RANBP1
SLC25A51
RANBP6
SELENBP1
ACAA2
IMPACT
TMEM184B
LIMK2
HMGXB4
TRIOBP
CREB3
RRAGA
PIK3C3
SLC39A6
ZNF24
SPECC1L
TUG1
MORC2
UBAP1
NFIB
VCP
SYK
YES1
NPC1
LPIN2
GCAT
DGCR2
PATZ1
PEX26
SNAPC3
SEH1L
ZNF271P
RBBP8
CBX7
RALBP1
DDT
KLHL22
DCTN3
PLGRKT
RNF38
CAAP1
TRAPPC8
YPEL1
MKL1
CSNK1E
SMARCB1
GGA1
RCL1
MLLT3
FOCAD
RNMT
C18orf8
PACSIN2
HIRA
CLTA
OSBPL1A
PI4KA
CBY1
HIC2
ATP6V1E1
MAPK1
DNAJA1
STOML2
BAG1
KLHL9
HMGN5
SNRPD1
GRAMD4
NF2
DRG1
UQCR10
PUM3
NDUFB6
PIGO
CEP76
METTL4
USP14
RNF138
MPST
APTX
PPM1F
CRKL
IFT27
DGCR6
PSIP1
PLAA
ZBTB5
SLC15A2
IMPA2
NDUFV2
GAREM1
CDC37L1
SF3A1
NUDT2
NOL12
AKAP13
HAUS6
MYRIP
PPP4R1
RIOK3
SS18
PTPN2
DNAL4
SLC25A1
ZMAT5
SMU1
IFT74
ASCC2
TST
SNU13
GALT
UNC13B
ANKRD12
PSMG2
MPPE1
SNRPD3
UBE2L3
PTN
SIGMAR1
TTC39A
TOPORS
CABYR
CBX6
SLC1A1
DSC2
DCAF10
PELI2
PI4KA
IGKV1-39
FAM46C
IGLC1
IGLL3P
IGK
IGHM
IGKC
CD38
IGH
IGLV3-10
IGHD
IGHG1
IGHA1
IGHM
IGKV1OR2-108
IGHA1
IGKV2-28
JCHAIN
CYAT1
IGKV4-1
IGLV2-14
IGKV1-17
IGLJ3
CYAT1
IGLV
GUSBP11
IGK
IGLJ3
IGK
BMS1P20
GLYR1
IGLV3-25
ICAM2
BLOC1S5-TXNDC5
IGKV1-37
POU2AF1
IGH
SEPT6
FKBP11
IGKV1OR1-1
IGKV1D-13
LOC100233156