SUPPLEMENTARY METHODS:

All figures and methods generated using R can be reproduced using the sweave[1] format document provided as an additional supplementary document. Note that figures in the sweave-generated automatic output are numbered in order of presentation and may not necessarily correspond to the figure number in the main text. For clarity, the figure number corresponding to that in the manuscript and in the supplementary data is annotated in the sweave figure legend.

DESCRIPTION OF PATIENT SAMPLES AND SOURCES OF GENOMIC DATA

Genomic data of ovarian cancer patients from three different cohorts, referred to as AOCS, TCGA, and Japan are used in this study and are described below. The samples used for initial genomic and immunohistochemical (IHC) analyses were derived from Peter MacCallum Cancer Centre and are outlined in Table 1. Samples used for validation IHC and genomics were from Vancouver, Japan and TCGA. Gene expression data from human ovarian surface epithelium (HOSE) was used for comparison with OCCA and HGSC Affymetrix data.

Peter MacCallum Cancer Centre samples

Most of the samples used for genomic analysis were collected as part of the Australian Ovarian Cancer Study (AOCS; (1)). Details of these samples, and the analyses performed with them, can be found in Table 1 and Supplementary Table 2. Of these, ten OCCA samples used in SNP6.0 analysis were obtained thought the Peter MacCallum tissue bank. Further detail on these samples may be obtained from Ramakrishna et al. (2).

TCGA data

The Cancer Genome Atlas (TCGA) is a comprehensive and multi-institutional effort to identify the genetics of cancer[2]. A subset of data (157 samples) from this cohort is used to compute the copy number changes found in HGSC (See(3) for details).

Japanese samples

250K Affymetrix SNP array data was available for 18 OCC samples from Jikei University (Tokyo, Japan) to provide a comparison group for copy number analysis. Summary of clinical annotations for these samples is provided in Supplementary Table 1.

---

[1] http://www.stat.uni-muenchen.de/~leisch/Sweave/

[2] http://cancergenome.nih.gov/

Vancouver samples

Vancouver tumor bank cohort used for TMA consisted of a single hospital-based set of cases from the Gynaecologic Tissue Bank at Vancouver General Hospital from patients diagnosed with ovarian carcinoma between 2001 - 2008. These samples represent high quality tissue with short devitalization times and standardized fixation and tissue processing (4).

Human ovarian surface epithelium

Data from 10 samples of human ovarian surface epithelium was obtained from Professor Michael Birrer as part of the data set described in Bonome et al (5).

GENE EXPRESSION DATA ANALYSIS

Raw microarray data of HOSE (5), HGSC (1, 6), and OCCA tumours were profiled on microarray (Affymetrix HGU133Plus_2) platform and were normalized using the GCRMA(7) method available in R package affy (8).

All probe sets with mean expression level greater than 7 and standard deviation greater than 1 were used to identify the structure in gene expression space in an unsupervised manner. Samples were hierarchically clustered using Euclidean distance as the distance measure and average linkage agglomeration method(9). To identify co-expressed genes, Pearson correlation coefficient was used as the similarity measure and average linkage agglomeration method.

Relative expression levels of selected genes is shown for HGSC and OCCA samples are shown as boxplots in Supplementary Figure 1A-1F. In our previous studies, HGSC are associated with the C1, C2, C4, and C5 molecular subtypes (1). C3 subtype is associated with *RAS* pathway activated borderline serous cancers and C6 with low-grade endometrioid tumours. Therefore C1, C2, C4, and C5 data were used for comparison with OCCA in which a Wilcoxon Rank Sum Test was used to compute the statistical significance of differences in expression.

The correlation between expression levels of *IL6, HIF2A* and *PTHLH* are also shown in Supplementary Figure 1. The statistical significance of the best fitting lines is shown, computed using the function available in R.

Comparison of the differential gene expression between OCCA tumours with or without ARID1A mutation was done using Limma(10).

To identify genes whose expression was associated with clinical outcome, samples are stratified into two groups based on the median expression level of each gene. When multiple probesets were mapped to a gene, the probeset with the largest mean expression is taken as representative probeset for the gene. Association with progression-free or overall survival was computed using Cox proportional hazard model and the log rank test p-value is reported.

Ontology & Pathway Analysis

Expression data was analysed using Bioconductor 2.6 running on R. Normalised expression value for 32 OCCA samples and 235 HGSC samples(1) were calculated by using Affymetrix package Robust Multichip Average (GCRMA) default method(8). Differential gene expression was assessed between OCCA and HGSC samples using an empirical Bayes t-test (Limma, (10)); p values were adjusted for multiple testing by using the Benjamini-Hochberg method(11). Differentially expressed probes sets were those that exhibited an adjusted p value of 0.05 or less and were divided into two sets based on their fold change relative to HGSC samples. Up and down regulated probe sets were individually assessed for pathway and biological processes enrichment using Genego's Metacore [GeneGo Inc, St. Joseph, MI] pre-annotated gene sets. The analysis employs a hyper-geometric distribution to determine the most enriched gene-set relative to the background chip (HgU133plus2). Significantly enriched pathways and processes for the up-regulated and down-regulated gene lists can be found in the supplementary tables.

Up- and down-regulated genes present in the OCCA cell line derived signature described by Yamaguchi et al(12) were analysed against OCCA tumour data using GSEA(13).

DNA COPY NUMBER DATA ANALYSIS

Data were generated using Affymetrix SNP6.0 genechip arrays. Raw data from all the batches of array experiments were separately normalized using procedures available in R-package aroma.affymetrix(14). Methods to remove systematic biases including GC biases and fragment length biases were applied. Samples were copy normalized using pooled normal samples from the same cohort.

Copy normalized samples were segmented using circular binary segmentation (CBS) algorithm available in R package DNAcopy(15). Segmentation values were used to make a call of gain and loss across the genome for every sample. Segments with mean value above 0.3 were considered to be gained and those with less than -0.3 were considered to be lost.

The frequency of aberrations in OCCA were compared against HGSC samples using data obtained from TCGA, as well as from an independent cohort of Japanese OCCA samples (Supplementary Table 2). For each cytoband the number of samples affected by gain ($\log_2$ ratio of copy number > 0.3) or loss ($\log_2$ ratio of copy number > -0.3) in each cohort was tested using Fisher exact test. A plot of the p-values (2-tail) is shown.

Minimal regions of gain and loss were defined using a semi-automated method described previously(16), a list of regions is contained in Supplemental Table 4. A genomic locus is considered to be gained if the segmented value is greater than a predefined threshold (0.3) and lost if the segmented value is less than a threshold (-0.3).

Length of the amplicon containing the *MET* gene varies across the samples. Distribution of the length quantized to steps of 1MB lengths is plotted for both OCCA and HGSC samples as bar graphs in Figure 2D.

Significant regions of DNA copy number change were identified using GISTIC(17) and these correlated with clinical outcome using Cox proportional hazard model and the log rank test p-value is reported.

REFERENCES

1. 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.

2. RAMAKRISHNA M, WILLIAMS LH, BOYLE SE, BEARFOOT JL, SRIDHAR A, SPEED TP, ET AL. IDENTIFICATION OF CANDIDATE GROWTH PROMOTING GENES IN OVARIAN CANCER THROUGH INTEGRATED COPY NUMBER AND EXPRESSION ANALYSIS. PLOS ONE. 2010;5:E9983.

3. GORRINGE KL, GEORGE J, ANGLESIO M, RAMAKRISHNA M, ETEMADMOGHADAM D, COWIN P, ET AL. GENOME-WIDE SURVEY OF COPY NUMBER ASSOCIATIONS IN OVARIAN CANCER. PLOSONE. 2010;IN PRESS.

4. KALLOGER SE, KOBEL M, LEUNG S, MEHL E, GAO D, MARCON KM, ET AL. CALCULATOR FOR OVARIAN CARCINOMA SUBTYPE PREDICTION. MOD PATHOL. 2010.

5. BONOME T, LEE JY, PARK DC, RADONOVICH M, PISE-MASISON C, BRADY J, ET AL. EXPRESSION PROFILING OF SEROUS LOW MALIGNANT POTENTIAL, LOW-GRADE, AND HIGH-GRADE TUMORS OF THE OVARY. CANCER RESEARCH. 2005;65:10602-12.

6. ANGLESIO MS, ARNOLD JM, GEORGE J, TINKER AV, TOTHILL R, WADDELL N, ET AL. MUTATION OF ERBB2 PROVIDES A NOVEL ALTERNATIVE MECHANISM FOR THE UBIQUITOUS ACTIVATION OF RAS-MAPK IN OVARIAN SEROUS LOW MALIGNANT POTENTIAL TUMORS. MOL CANCER RES. 2008;6:1678-90.

7. IRIZARRY RA, BOLSTAD BM, COLLIN F, COPE LM, HOBBS B, SPEED TP. SUMMARIES OF AFFYMETRIX GENECHIP PROBE LEVEL DATA. NUCLEIC ACIDS RES. 2003;31:E15.

8. GAUTIER L, COPE L, BOLSTAD BM, IRIZARRY RA. AFFY--ANALYSIS OF AFFYMETRIX GENECHIP DATA AT THE PROBE LEVEL. BIOINFORMATICS. 2004;20:307-15.

9. EISEN MB, SPELLMAN PT, BROWN PO, BOTSTEIN D. CLUSTER ANALYSIS AND DISPLAY OF GENOME-WIDE EXPRESSION PATTERNS. PROC NATL ACAD SCI U S A. 1998;95:14863-8.

10. SMYTH GK. LINEAR MODELS AND EMPIRICAL BAYES METHODS FOR ASSESSING DIFFERENTIAL EXPRESSION IN MICROARRAY EXPERIMENTS. STAT APPL GENET MOL BIOL. 2004;3:ARTICLE3.

11. BENJAMINI YHY. CONTROLLING THE FALSE DISCOVERY RATE: A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. J ROY STAT SOC. 1995;57:289-300.

12. YAMAGUCHI K, MANDAI M, OURA T, MATSUMURA N, HAMANISHI J, BABA T, ET AL. IDENTIFICATION OF AN OVARIAN CLEAR CELL CARCINOMA GENE SIGNATURE THAT REFLECTS INHERENT DISEASE BIOLOGY AND THE CARCINOGENIC PROCESSES. ONCOGENE. 2010.

13. SUBRAMANIAN A, TAMAYO P, MOOTHA VK, MUKHERJEE S, EBERT BL, GILLETTE MA, ET AL. GENE SET ENRICHMENT ANALYSIS: A KNOWLEDGE-BASED APPROACH FOR INTERPRETING GENOME-WIDE EXPRESSION PROFILES. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 2005;102:15545-50.

14. BENGTSSON H, IRIZARRY R, CARVALHO B, SPEED TP. ESTIMATION AND ASSESSMENT OF RAW COPY NUMBERS AT THE SINGLE LOCUS LEVEL. BIOINFORMATICS (OXFORD, ENGLAND). 2008;24:759-67.

15. VENKATRAMAN ES, OLSHEN AB. A FASTER CIRCULAR BINARY SEGMENTATION ALGORITHM FOR THE ANALYSIS OF ARRAY CGH DATA. BIOINFORMATICS. 2007;23:657-63.

16. GORRINGE KL, CAMPBELL IG. LARGE-SCALE GENOMIC ANALYSIS OF OVARIAN CARCINOMAS. MOL ONCOL. 2009;3:157-64.

17. BEROUKHIM R, GETZ G, NGHIEMPHU L, BARRETINA J, HSUEH T, LINHART D, ET AL. ASSESSING THE SIGNIFICANCE OF CHROMOSOMAL ABERRATIONS IN CANCER: METHODOLOGY AND APPLICATION TO GLIOMA. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 2007;104:20007-12.