Supplementary materials for Tamura et al.

(one supplemental movie, four supplemental tables, six supplemental figures, one supplemental methods section)

Movie S1: HGSC cells exhibit elevated microtubule assembly rates. Shown is an EB3-GFP microtubule tip tracking movie from a COV318 HGSC cell treated with monastrol and filmed every 2 s (see Materials and methods for details).

Table S2: All mutations identified from whole genome sequencing of seven HGSC cell lines (including *TP53*, *BRCA1* and *BRCA2)*.

Table S3: RNAseq data of all transcriptional changes in HGSC cell lines, relative to FNE1 or FNE2, changes unique to each cell line, and CIN phenotype-specific changes (for **Figure S5c**).

Table S4: Lists of significant gene transcription changes for specific CIN phenotypes, plus GO-pathways associated with specific CIN phenotypes (for **Figure S6a,b**).

Table S1: Details of HGSC cell lines used in this study.

Cell Line	Source	Tissue	Notes
FNE1[1]	University of Miami	Fallopian tube epithelium	45n, -15, -X, t9:15, iso9p
FNE2[1]	University of Miami	Fallopian tube epithelium	
COV318[2]	F. Balkwill	Ovary (peritoneal ascites)	
G33[3]	F. Balkwill		Omental metastases of a patient with HGSC after chemotherapy
G164	F. Balkwill		
Kuramochi[4]	JCRB (Purchased)	Ovary (ascites)	Ovarian cancer, undifferentiated carcinoma
OVKATE[5]	JCRB (Purchased)	Ovary	Established from a patient with ovarian carcinoma, stage IIIc, CAP(3) EP(6) treatment done previously
OVSAHO[5]	F. Balkwill		
SNU119[6]	KCLB (Purchased)	Abdominal metastatic focus	Established from a patient with ovarian carcinoma, stage IIIc, FAMT(15) CFF(6) treatment done previously

Figure S1:

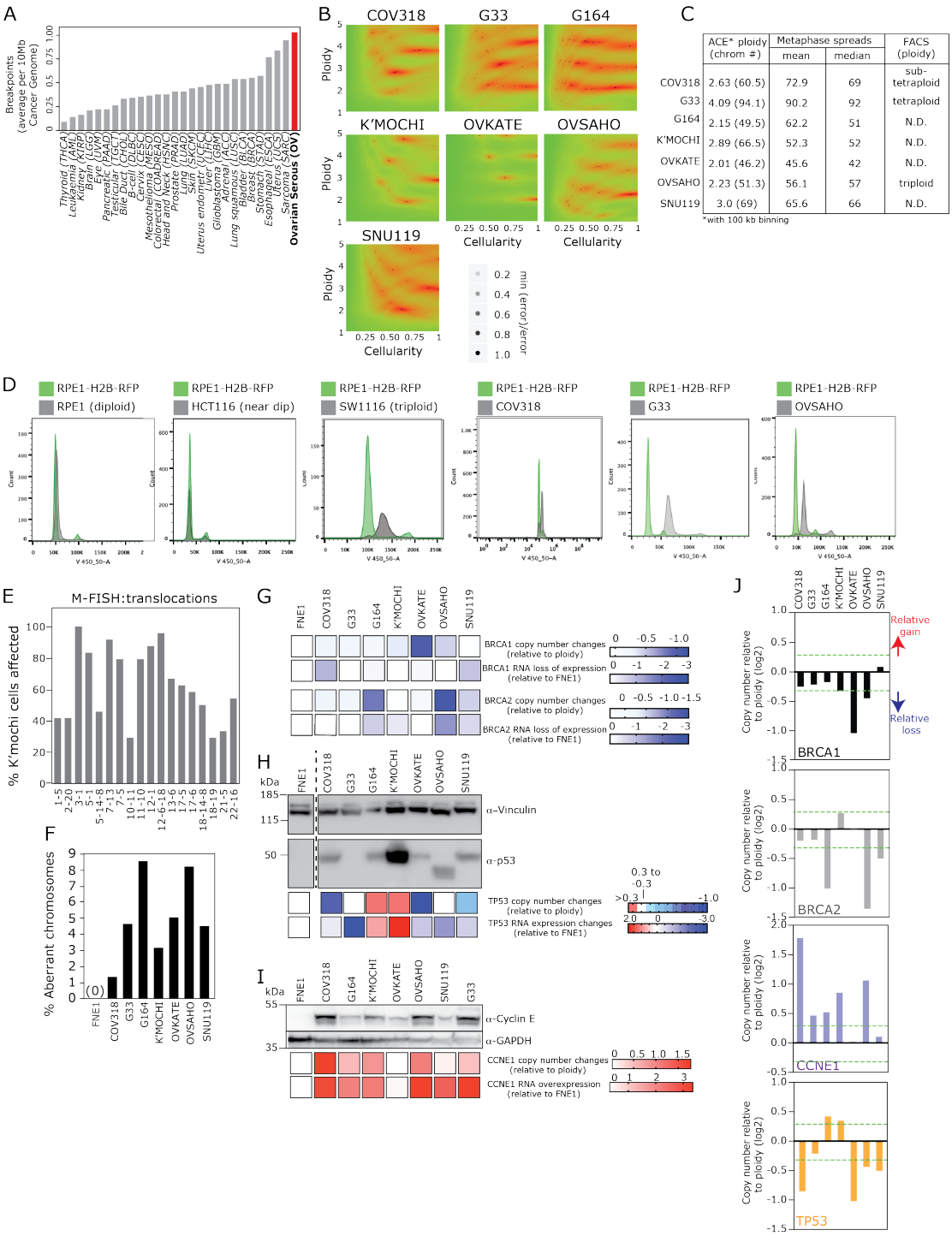

Figure S1 (relating to Figure 1): HGSC cell lines demonstrate a range of ploidies and exhibit genomic changes in key ovarian cancer marker genes. A). Frequency of breakpoints occurring in tumour genomes across a range of different cancer types. The number of breakpoints per 10 Mb genome segment is shown, as calculated from the segmentation copy number data of the TCGA datasets. **B)** Matrix of errors plots derived from the loopsquaremodel function of ACE to indicate ploidy and cellularity probabilities. Intensity of red zones represent most likely ploidy and cellularity

(% sample that is tumour cells), the latter of which should be close to 1 for cell lines (see Materials and methods for details). Most models found best fits with high cellularity, but notably cell line COV318 most likely has a ploidy between 3N and 4N based on our available data for the model. **C)** Table showing the best ploidy values obtained from ACE analysis, compared to mean and median chromosome counts from metaphase chromosome spreads. **D)** Flow cytometry analysis of three HGSC cell lines, compared to diploid cell lines. Each 'test' cell line sample was spiked with diploid RPE1-H2B-RFP cells to provide an internal control. CRC cell lines HCT116 (near diploid) and SW1116 (near triploid) were also analysed as positive controls for diploidy and triploidy respectively. **E)** Analysis of M-FISH images of 24 Kuramochi metaphase spreads (see **Figure 1d**). Percentage of cells showing each indicated translocation. **F)** Percentage of all chromosomes analysed from metaphase spreads which demonstrated a structural defect (i.e. dicentric or acentric) (893-1511 chromosomes analysed across two experiments for each cell line). **G)** Heatmap indicates changes in DNA copy number and RNA expression (compared to FNE1) for BRCA1 and BRCA2. **H)** Analysis for p53 expression by Western blot (with vinculin as loading control) with heatmaps for DNA copy number and RNA expression (compared to FNE1). **I)** Similarly, an analysis of Cyclin E expression (with GAPDH as loading control). **J)** Graph showing DNA copy number alterations of key genes on log2 scale, derived from calculating gene reads coverage as a function of mean genome read depth (see Methods for details). Green dotted lines indicate range of "normal" copy number relative to ploidy (between +0.3 and -0.3).

Figure S2:

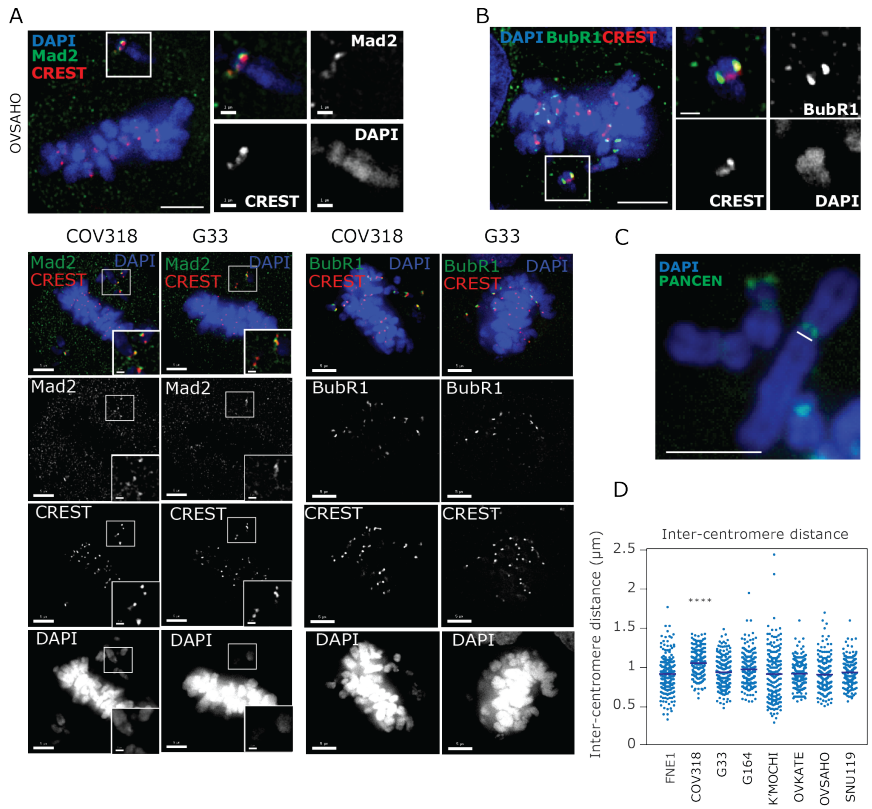

Figure S2 (relating to Figure 2): The mitotic checkpoint is functional and sister chromatid cohesion is normal in HGSC. A,B) Prometaphase cells from Ovsaho were probed with antibodies against CREST (red, marks centromeres) and either Mad2 (in A) or BubR1 (in B), components of the spindle assembly checkpoint that accumulate on unattached kinetochores and delay anaphase onset. Unaligned centric chromosomes demonstrated robust loading of Mad2 and BubR1. **C,D)** Measurements (white line) of inter-sister chromatid centromere distances (n>200 per cell line) from metaphase spread (scale bar indicates 5 µm). Statistical test is one-way ANOVA with Dunnett's correction for multiple testing (significant difference to FNE1 control shown, **** =p<0.0001).

Figure S3:

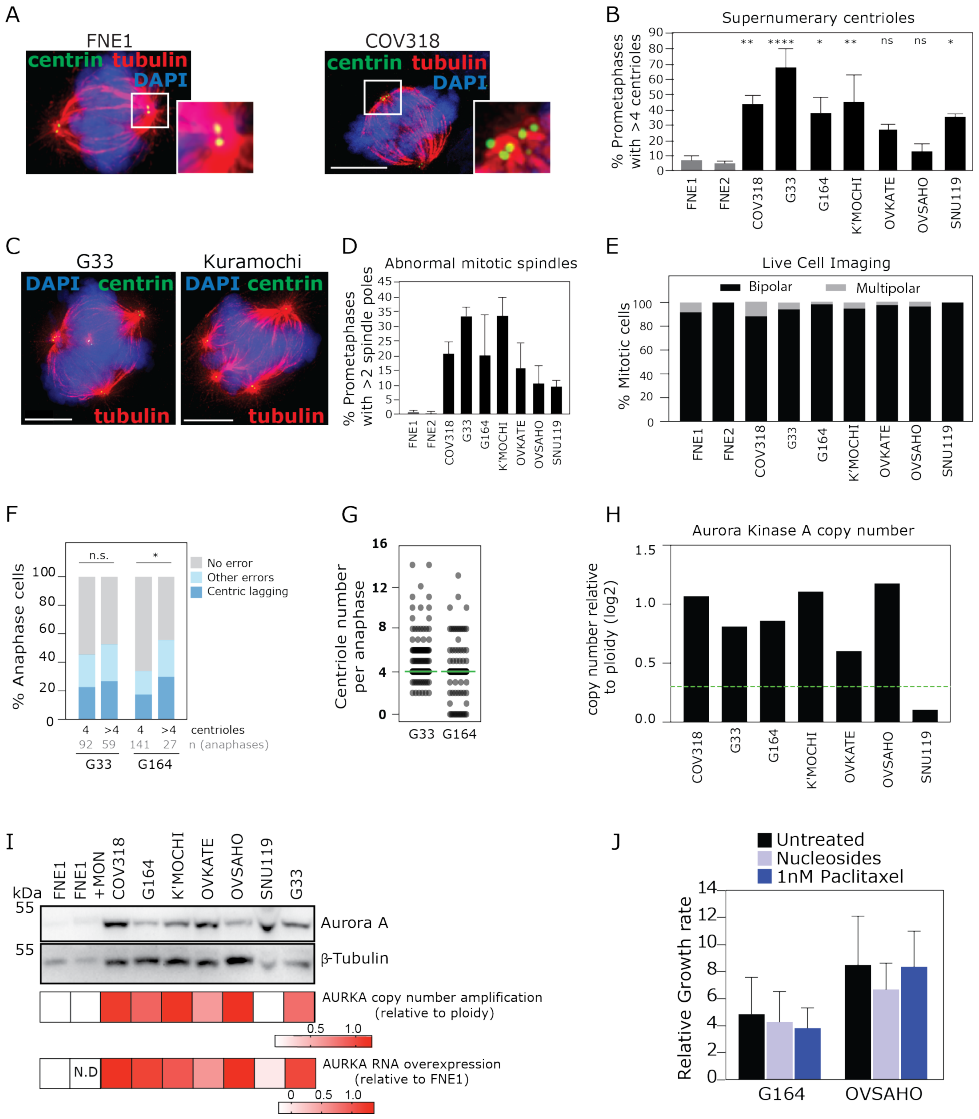

Figure S3 (relating to Figure 2): Supernumerary centrioles, multipolar divisions and Aurora Kinase A overexpression in HGSC. A) Prometaphase cells probed with antibodies against centrin (green) to mark centrioles and tubulin (red) to mark microtubules. **B)** Percentage of prometaphase cells with abnormal (>4) centrioles. Statistical test is one-way ANOVA with Dunnet's correction for multiple testing. Differences between HGSC cancer cell lines compared to FNE1 are shown. **C)** Examples of prometaphase cells with abnormal mitotic spindles. **D)** Percentage of cells with abnormal mitotic spindles. **E)** Quantification of multipolar and bipolar cell divisions from live cell imaging. Summary of two experiments, n=28-101 cells. **F)** Chromosome segregation error rates in anaphase cells with normal (4) and abnormal (>4) centrioles (n = total anaphases from two independent experiments). Significance from Fishers exact tests for all errors versus no errors between 4 and >4 centriole populations is shown. **G)** Quantification of centriole number in mitotic cells in indicated cell lines. **H)** Graph showing DNA copy number alterations of Aurora Kinase A (*AURKA*) derived from

whole genome sequencing, calculating gene reads coverage as a function of mean genome read depth (see Methods for details). Dotted line at 0.3 indicates level of relative normal ploidy, beyond this is considered gene amplification. **I)** Western blot of untreated cell lines (or FNE1 treated with monastrol: +MON), with membrane probed for antibodies against Aurora A, or β-tubulin (loading control). Heatmaps of *AURKA* gene DNA copy number (relative to median cell line ploidy) and of AURKA RNA overexpression (relative to FNE1 expression). N.D indicates not determined. **J)** Proliferation rates of cell lines treated with nucleosides or low dose (1 nM) paclitaxel.

Figure S4:

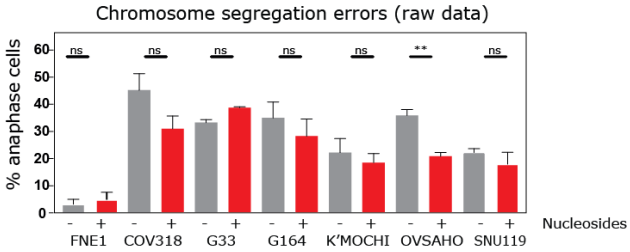

Figure S4 (relating to Figure 3): Reduction of segregation errors upon nucleoside supplementation. Segregation error rates for untreated cells vs cells treated with nucleosides, used to generate **Figure 3F**. T tests between pairs (treated vs untreated) are indicated.

Figure S5:

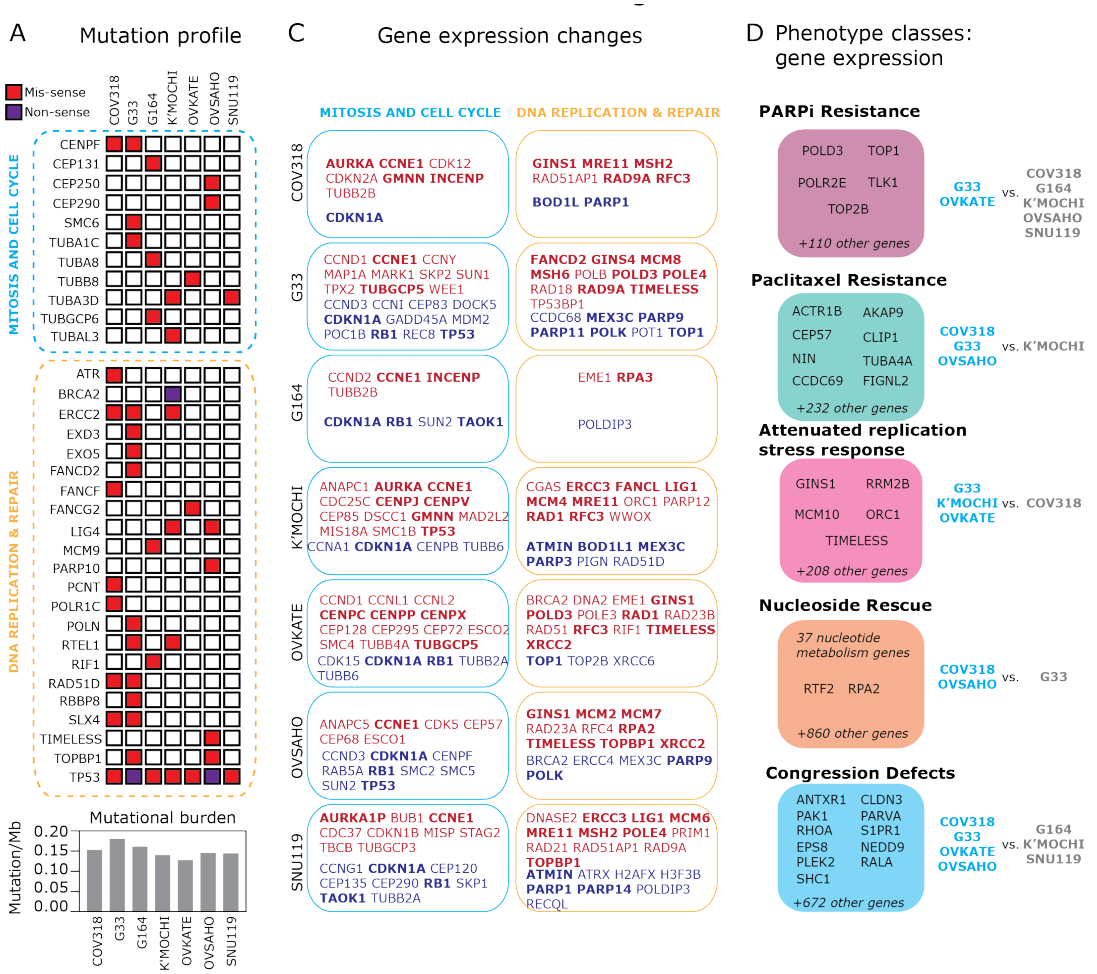

Figure S5: Genetic and transcriptomic analyses reveals potential causative CIN genes.

A) List of mutations in CIN pathways identified in cell lines after screening out common SNPs. **B)** Mutational burden in HGSC cell lines. **C)** List of CIN-related genes which are upregulated or downregulated in HGSC (statistically significant changes in RNA expression in HGSC cell line when compared to both FNE1 and to FNE2 controls). Genes (or similar genes in same pathway) found in more than one cell line are listed in bold. Blue font indicates downregulation, red indicates upregulation. **D)** CIN-related genes with significantly altered expression (compared to both FNE1 and FNE2) in panels of cell lines sharing the phenotypes indicated. Listed are: For PARP inhibitor (PARPi) Resistance, five genes in the replication stress response pathway. For Paclitaxel Resistance, nine genes in positive regulation of microtubule polymerisation. For Attenuated Replication Stress response, five genes in DNA replication. For Nucleoside Rescue, thirty-seven genes in nucleotide metabolism. For Congression Defects, a significantly enriched pathway (FDR=3.39E-03) of eleven genes in Actin Reorganisation pathway was identified.

Figure S6:

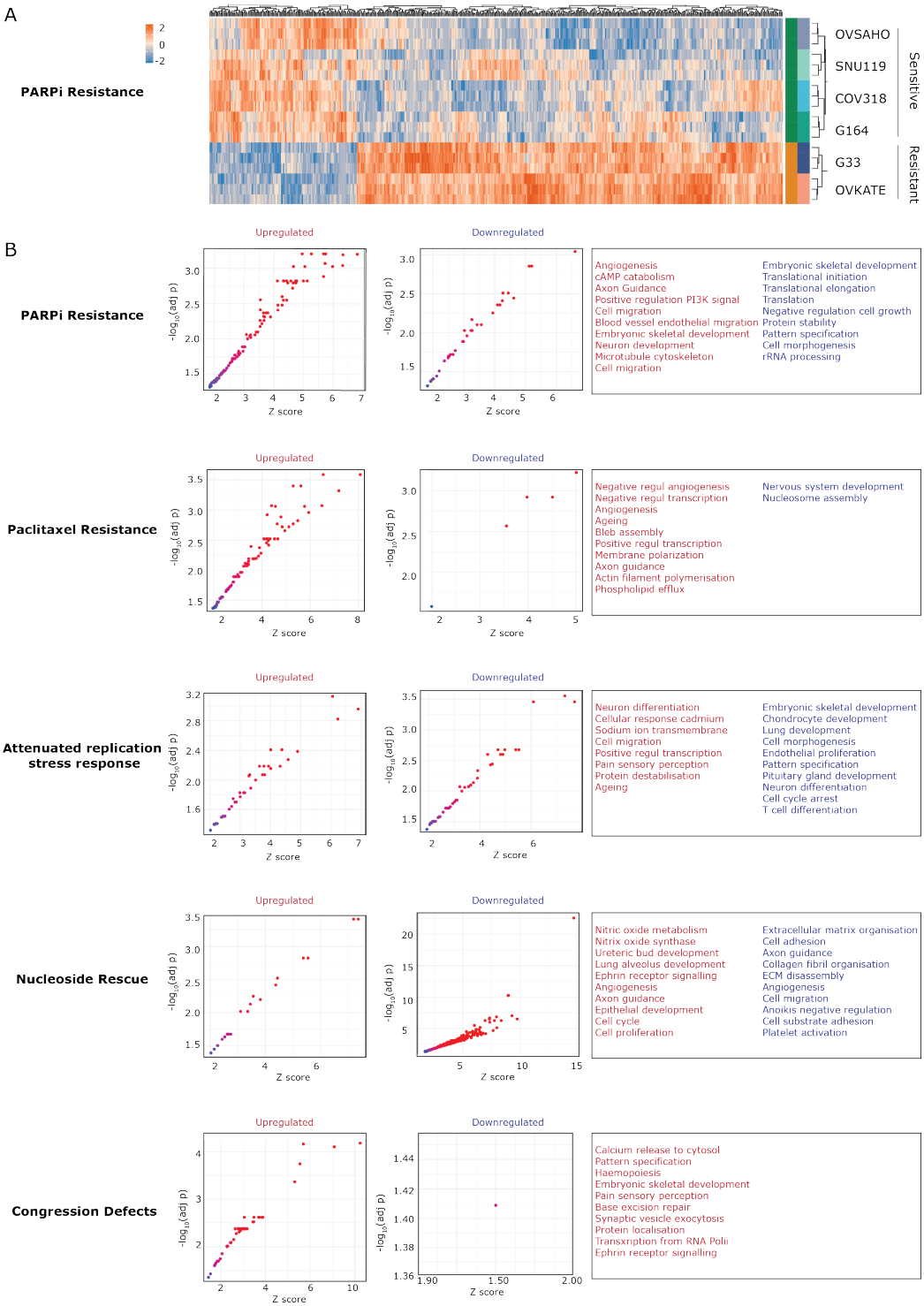

Figure S6: Analysis of transcriptional changes and pathway enrichment. A) Representative heatmap showing gene expression changes in two HGSC cell lines (Ovkate, G33) that demonstrated resistance to PARPi, vs five HGSC cell lines (COV318, G164, Kuramochi, Ovsaho, Snu119) that do not. Full list of genes for this and four other CIN phenotypes found in Supplementary Table S4. **B)** Plots of ontology pathways that are enriched for upregulated or downregulated genes for cell

lines that share particular CIN phenotypes, each with a listing of the top ten pathways with a Z-score >4. Full lists in Supplementary Table S4.

Supplementary Methods

Whole Genome Sequencing:

For data preprocessing, FastQC was used to perform quality control (QC) of the raw sequencing files[7]. Each file was then aligned to the Hg38 genome build using HISAT2[8]. Resulting .sam files were compressed and converted to .bam files using samtools 1.7[9]. Picard version 2.6.0-SNAPSHOT was used to sort the .bam files and to mark duplicate reads (http://broadinstitute.github.io/picard/). Duplicate reads were removed and .bam files were indexed using samtools and picard respectively. QualiMap Bamqc v.2.2.1. was used to conduct QC of the processed .bam file[10].

Downsampling and absolute copy number estimation and visualization: To generate absolute copy numbers using unmatched tumour data, we first align raw .fastq files to the Hg37 Human Genome Reference using HISAT2 and pre-processed as above. Samtools was used to obtain 0.1X downsampled .bams. Using these downsampled .bam files as input, we performed absolute copy number estimation using the R package ACE[34]. The loopsquaremodel function of ACE and resulting matrixplots were used to generate ploidy and cellularity models with the lowest error rates. The most likely biologically relevant model was then manually picked based on the following criteria: 1) Cellularity estimates should be ~1.0 due to the nature of our cell line samples and 2) Ploidy estimates should be concordant with chromosome numbers counted from metaphase spreads. The ggplot2 package in R was used to plot the segmented copy number profiles.

Somatic mutation and copy number variation (CNV) analysis

In absence of matched normal samples, GATK version 4.1.4.1. was applied on processed .bam files to perform somatic SNV and CNV calling, according to the GATK best practices workflow[11]. First, a panel of normals (PoN) was assembled using twelve 30X PCR-free high coverage samples from the 1000 Genomes Project[12]. Mutations were then called using Mutect2 and the assembled PoN, after which raw variant calls were further processed, filtered and annotated. For CNV calling, read coverage counts were collected across pre-processed intervals (bin length 1000bp), and denoised against a CNV PoN. Next, allelic counts for both the reference and samples were collected, and copy number data were segmented. Finally, amplification, copy-neutral and deletion events were called. Low and high copy-neutral segment ratio bounds were defined as 0.7 and 1.3 respectively. Gene

specific copy number log-ratios were obtained by matching the gene coordinates with the segmented copy number data.

Cancer genome breakpoint analysis: TCGA segmentation copy number data was downloaded from cBioportal[13,14] and the number of breakpoints were calculated using an R-script adapted from available code as written by Macintyre et al[15].

RNAseq Analysis: Raw reads were mapped to the human genome (hg38, Genome Reference Consortium GRCh38) using HISAT2[16]. Number of reads aligned to the exonic region of each gene were counted using HTSeq[17] based on the Ensembl annotation. Only genes that achieved at least one read count per million reads (cpm) in at least twenty-five percent of the samples were kept and a log2cpm expression matrix was subsequently generated. Differential expression analysis of each of the cell lines versus the control cell lines FNE1 and FNE2 was performed using the 'limma' R package[18]. Gene-set enrichment analysis (GSEA) of these genes was performed using the GSEA software[19] for Gene Ontology Biological Processes. To identify genes differentially expressed between multiple cell-line classes (multi-class comparisons) we used the function sam from R package siggenes. Enrichment analysis on these genes was performed by hypergeometric test for Gene Ontology Biological Processes using the dEnrichr function of the R package dnet. Heatmaps illustrating the expression pattern of the genes were generated using R package ComplexHeatmap. Row clustering was performed on euclidean distance and "complete" clustering method. Dotplots of significantly enriched pathways (BH-adjusted $p < 0.05$) were generated using R package ggplot2. RNA-Seq data have been deposited in Gene Expression Omnibus (GEO) under the accession number GSE155310. For pathway identification of genes enriched relative to FNE1/2, pathways were identified using PantherDB.org.

Supplementary References:

1 Merritt, M. A. *et al*. Gene expression signature of normal cell-of-origin predicts ovarian tumor outcomes. *PloS one* **8**, e80314, doi:10.1371/journal.pone.0080314 (2013).

2 van den Berg-Bakker, C. A. *et al*. Establishment and characterization of 7 ovarian carcinoma cell lines and one granulosa tumor cell line: growth features and cytogenetics. *International journal of cancer. Journal international du cancer* **53**, 613-620 (1993).

3 Milagre, C. S. *et al*. Adaptive Upregulation of EGFR Limits Attenuation of Tumor Growth by Neutralizing IL6 Antibodies, with Implications for Combined Therapy in Ovarian Cancer. *Cancer research* **75**, 1255-1264, doi:10.1158/0008-5472.CAN-14-1801 (2015).

4 Motoyama, T. [Quantitative analysis on in vitro drug sensitivity of cultured human ovarian cancer cell lines (author's transl)]. *Nihon Sanka Fujinka Gakkai Zasshi* **34**, 308-314 (1982).

5 Yanagibashi, T. *et al*. Complexity of expression of the intermediate filaments of six new human ovarian carcinoma cell lines: new expression of cytokeratin 20. *British journal of cancer* **76**, 829-835 (1997).

6 Yuan, Y. *et al*. Establishment and characterization of human ovarian carcinoma cell lines. *Gynecologic oncology* **66**, 378-387, doi:10.1006/gyno.1997.4785 (1997).

7 Andrews, S. FastQC: A Quality Control tool for High Throughput Sequence Data [Online]. . http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

8 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907-915, doi:10.1038/s41587-019-0201-4 (2019).

9 Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

10 Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294, doi:10.1093/bioinformatics/btv566 (2016).

11 McKenna, A. *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

12 Genomes Project, C. *et al*. A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

13 Gao, J. *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).

14 Cerami, E. *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).

15 Macintyre, G. *et al*. Copy number signatures and mutational processes in ovarian carcinoma. *Nature genetics* **50**, 1262-1270, doi:10.1038/s41588-018-0179-8 (2018).

16 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).

17 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

18 Ritchie, M. E. *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).

19 Subramanian, A. *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* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).