Molecular landscape and functional characterization of centrosome amplification in ovarian cancer

Carolin M. Sauer[1,2,$], James A. Hall[1,2], Dominique-Laurent Couturier[1,3], Thomas Bradley[1,2], Anna M. Piskorz[1,2], Jacob Griffiths[1,2], Ashley Sawle[1,2], Matthew D. Eldridge[1,2], Philip Smith[1,2], Karen Hosking[2], Marika A.V. Reinius[1,2,4], Lena Morrill Gavarró[1,2], Anne-Marie Mes-Masson[5], Darren Ennis[6], David Millan[7], Aoisha Hoyle[8], Iain A. McNeish[6], Mercedes Jimenez-Linan[4], Filipe Correia Martins[1,4], Julia Tischer[1,2], Maria Vias[1,2], James D. Brenton[1,2,4,$]

1. Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.

2. Cancer Research UK Major Centre–Cambridge, University of Cambridge, Cambridge CB2 0RE, UK.

3. Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, UK.

4. Cambridge University Hospital NHS Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK.

5. Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM) and Institut du Cancer de Montréal (ICM), CRCHUM, 900 Saint-Denis, Montreal, QC, Canada.

6. Department of Surgery and Cancer, Ovarian Cancer Action Research Centre, Imperial College London, London, UK.

7. Department of Pathology, Queen Elizabeth University Hospital, Glasgow, UK.

8. Deparment of Pathology, University Hospital Monklands. NHS Lanarkshire, UK.

[$]Correspondence: carolin.sauer@cruk.cam.ac.uk; james.brenton@cruk.cam.ac.uk

Abstract

High grade serous ovarian carcinoma (HGSOC) is the major cause of gynaecological cancer-related deaths in the Western world, and is characterised by extreme chromosomal instability (CIN). CIN is a key mediator of clonal diversity and treatment resistance and is associated with poor disease outcome. An increased understanding of underlying mechanisms driving CIN in HGSOC is therefore critical to predict outcomes and identify novel therapeutic approaches. The centrosome is the main microtubule organising centre of a cell and plays a crucial role during cell division and chromosome segregation. Missegregation of chromosomes at high rates results in CIN; and supernumerary centrosomes have previously been associated with aneuploidy and poor disease outcome in several cancers. However, relatively little is currently known about centrosome abnormalities and associated molecular features in HGSOC. Here, we developed high-throughput microscopy approaches to provide the first large-scale characterisation of centrosome amplification (CA) in >300 clinical tissue samples and >70 ovarian cancer cell lines. We report that HGSOC frequently shows supernumerary centrosomes with marked intra-tissue heterogeneity. Imaging and shallow whole genome sequencing experiments further revealed that CA in HGSOC is associated with the presence of micronuclei, increased genomic abnormality and, interestingly, genomic subclonality, highlighting CA as a potential driver of tumour evolution. Further, using RNA sequencing, and drug assays on selected cell lines we found that cells with high CA show upregulated NFκB signalling and increased resistance to standard-of-care paclitaxel as well as other therapeutic agents. In addition, our detailed phenotypic, genomic, and transcriptomic characterisation of >70 ovarian cancer cell lines will provide an important platform for future studies on CIN and CA in ovarian cancer.

Keywords

Chromosomal instability; high grade serous ovarian cancer; centrosomes; micronuclei; copy number aberrations; cell line characterisation

Introduction

High grade serous ovarian carcinoma (HGSOC) accounts for the majority of ovarian cancer cases and ovarian cancer related deaths with a five-year survival rate of less than 30%. Overall survival of HGSOC patients has not changed over the last 2 decades mainly due to the severe ongoing chromosomal instability (CIN) characterising and driving this disease. The lack of a clear understanding of the drivers of this genomic complexity has significantly impeded the development of targeted therapies, with the major exception being the successful targeting of BRCA1/2 dysfunction with PARP inhibitors. Importantly, CIN, which is present in virtually all HGSOCs[1], has been shown to drive resistance to anti-cancer therapies[2–6], and is the main cause of regenerating subclonal diversity in response to treatment-induced selective pressures[7]. As such, an increased understanding of the mechanisms and roles of CIN in HGSOC is now crucial to facilitate the development of novel therapeutic approaches.

The single most prominent cause of CIN is chromosome missegregation, which can be caused by weakened mitotic checkpoint, improper chromosome attachment to the mitotic spindle and centrosome amplification (CA; typically defined as an abnormal number of centrosomes >1 in non-dividing cells)[8–10]. Centrosomes are the main microtubule organising centre (MTOC) of eukaryotic cells, forming the mitotic spindle poles during cell division. The centrosome consists of two centrioles contained within the pericentriolar material (PCM), a proteinaceous scaffold with microtubule-nucleating activity[11,12]. During the cell cycle, each centrosome duplicates once in coordination with chromosome duplication in a tightly controlled manner. This ensures that exactly two centrosomes are present during mitosis to guide spindle formation and chromosome segregation[13]. As such, centrosome abnormalities, both structural and numerical, can lead to the missegregation of chromosomes.

CA has recently emerged as a potential therapeutic target, and several drugs targeting centrosome duplication[14] or associated survival mechanisms[15,16] are now in development or clinical trials[17]. An improved understanding of the prevalence, mechanisms and origins of CA is therefore critical for the clinical exploitation of this phenotype. Several studies have previously reported CA-induced formation of chromosome bridges[16], and the occurrence of lagging chromosomes, which can result in micronuclei (MN) formation[8,18,19], leading to aneuploidy and CIN. In addition, CA has also been associated with invasive cell phenotypes, paracrine tumour signalling, and spontaneous tumourigenesis[18,20–23]. However, the vast majority of studies investigating mechanisms and consequences of CA have relied on experimentally induced or inhibited CA in cell lines and other model systems. Importantly, the high degree of variability in centrosome numbers observed across cancer cell populations suggests the existence of a CA “set point” or equilibrium[24]. This was first proposed following observations that cell lines recovered to their initial levels of CA after inhibition of centriole biogenesis[25]. Other studies have also demonstrated the loss of extra centrosomes over time following the inhibition of cytokinesis to induce CA[26]. Together, these observations strongly suggest that cell lines have different tolerance thresholds for CA and maintain centrosome numbers through an equilibrium of CA mechanisms and negative selection. Consequently, the experimental induction or inhibition of CA may trigger a range of different cellular responses that would not otherwise be

observed without these artificial perturbations. These data highlight the need for detailed systematic analyses of baseline CA profiles along with their molecular and genomic associations in large panels of cancer cell lines. In addition, because of significant challenges associated with detecting centrosomes in clinical human tissue samples, only a handful of studies have investigated CA in clinical tumour specimens[27,28], and the prevalence of CA in HGSOC (both cell lines and tumour tissues) remains largely unclear.

Here, we develop a high-throughput microscopy-based assay to reliably detect and quantify centrosomes in formalin-fixed paraffin-embedded (FFPE) tissues and analyse centrosome profiles in 287 HGSOC tumours. Further, we extend these observations to in-depth profiling of >70 ovarian cancer cell line models and provide the first comprehensive phenotypic, transcriptomic, and genomic characterisation of cell lines with supernumerary centrosomes.

Results

Supernumerary centrosomes are present in the majority of HGSOC tissues and show marked inter- and intra-tumour heterogeneity

To investigate the presence and extend of supernumerary centrosomes in HGSOC, we selected full-face FFPE tumour tissue sections (n=93) from the OV04 study, and tissue microarrays (TMAs; n=194) from the British Translational Research Ovarian Cancer Collaborative (BriTROC) study (total of n=287 tumour tissues; see methods). In addition, we included 24 normal fallopian tube (FT) samples, a spleen, and a kidney samples as negative controls. Normal liver samples were included as positive control tissues for CA, as the liver is the only organ in the human body that displays centrosome amplification in a healthy state[29–31]. An overview of all analysed tissue samples (n=318) is shown in **Supplementary Table 1**. To comprehensively survey and obtain an accurate estimate of centrosome numbers in HGSOC tissues, we developed a highly automated, reliable and high-throughput immunofluorescent microscopy-based imaging approach using the confocal Operetta CLS™ imaging system. This allowed us to unambiguously detect centrosomes in thick (25 µm) FFPE tissues sections across up to 50 randomly placed imaging fields per tissue sample, providing good tissue representation. Combined, our centrosome profiling analyses comprised a total of 3,632,389 nuclei across 11,811 imaging fields. To allow cross-comparisons across both cohorts and to account for potential staining (batch effect) differences, CA scores were normalised to the median CA score of normal control tissues within each cohort. In both study cohorts, CA scores were significantly higher in tumour tissues than normal control tissues (FT, spleen and kidney; Kruskal-Wallis, $p = 1.9 \times 10^{-10}$). As expected, control liver tissues showed the highest CA score (**Fig. 1d**). To investigate the prevalence of CA in HGSOC, we determined a CA score threshold using the 95% confidence interval of the FT sample with the highest CA score (**Fig. 2**), resulting in a CA cutoff value of ~1.83. Using this cutoff, we determined that ~73% of OV04 and ~59% of BriTROC tumour samples displayed significant CA (**Fig. 1d**). In contrast, only one tissue sample showed significantly lower CA scores than normal tissues (**Fig. 2**). In addition, we found that centrosomes in HGSOC tissues had significantly larger PCM areas (measured as width × length determined from 2D maximum intensity projection images) than centrosomes detected in normal control tissues (Kruskal-Wallis, $p = 2.6 \times 10^{-8}$; **Fig. 1e**).

Strikingly, as shown by the boxplots in **Fig. 2**, individual tissue sections not only showed large inter-tissue variability in median CA scores, but also substantial differences in the intra-tumoural spread of CA scores across different imaging fields, even for tissue samples with similar CA score medians. To investigate this heterogeneity further, we developed a hierarchical statistical model able to handle the potential within-tissue dependence in mean and variance CA scores by means of random effects and allowing for heavy-tailed error distributions (see Methods for details). Model estimates confirmed the significant differences in average CA levels between FT and tumour tissues (p □ 0.0001), as well as the marked presence of within-tissue-heterogeneity ($p < 0.0001$). However, the CA mean to variance relationship was not significantly different between FT and tumour tissues ($p = 0.725$). To quantify the observed heterogeneity in CA phenotypes, we estimated the standard deviation of log-transformed CA scores across all imaging fields for each tissue sample. Examples of CA high and low tissues with varying tissue heterogeneity scores are shown as spatial heatmaps in

Fig. 3a-b and **Supplementary Fig. 1**. Heterogeneity scores were significantly higher in OV04 samples (full face sections) than in BriTROC samples (TMA cores; Wilcoxon, $p = 6.6 \times 10^{-11}$) consistent with a wider tissue representation of imaging fields placed across full face sections compared to adjacent imaging fields covering 0.6mm TMA cores. In addition to the intra-tumour CA heterogeneity, we also observed significant differences in CA levels across tumour tissues collected from the same patient (n = 36; **Fig. 3c**). Using the CA cutoff of ~1.83, for the majority of patients (~64%), analysed tissues fell into the same CA category (i.e. either significant CA or no CA), whereas for ~36% of cases, tissues within either category were found. Importantly, no differences in CA scores were observed between different anatomical tissue sites (Kruskal-Wallis, p = 0.53; **Fig. 3d**), while tissue samples collected from the omentum showed highest CA tissue heterogeneity (Kruskal-Wallis, p = 0.011; **Fig. 3e**). However, this might be caused by the sampling differences between the two study cohorts.

Overall, these analyses show a very high prevalence of CA in HGSOC cases and highlight marked within-tissue and between-tissue heterogeneity of CA phenotypes. Both the high degree of CA heterogeneity and the confounding effect of stromal cells underscore the importance of acquiring multiple imaging fields per tumour sample and highlight a key strength of our imaging approach (see **Supplementary Methods**).

Centrosome amplification is not associated with disease outcome in HGSOC

CA has been frequently associated with increased disease aggressiveness and poor patient prognosis in various cancer types[27,32]. We therefore investigated whether CA was associated with clinical or molecular features and overall survival in ovarian cancer.

For patients for which multiple tissue samples were analysed, the median CA score across these tissues was used. The BriTROC cohort, contained four patients with low grade endometrioid ovarian cancer, which have improved prognosis compared to HGSOC patients. These patients showed significantly lower CA compared to the remaining HGSOC patients from both cohorts (Wilcoxon, p = 0.02; **Fig. 4a**). No significant differences in CA scores were observed across different tumour stages (FIGO stages 1–4; Kruskal-Wallis, p = 0.18), *BRCA1* and *BRCA2* germline mutation status (where available; Kruskal-Wallis, p = 0.26), or whether analysed samples were treated (delayed primary surgery [DPS] post chemotherapy initiation) or treatment-naïve (immediate primary surgery [IPS] prior to chemotherapy; Wilcoxon, P = 0.26) (**Fig. 4b–d**). We did, however, find an increase in CA tissue heterogeneity scores in treated (DPS) samples compared to treatment-naïve (IPS) samples (Wilcoxon, P = 0.015; **Fig. 4e**) as well as an increase in centrosome size (Wilcoxon, p = 0.00036; data not shown), suggesting that chemotherapy might have an effect on tissue-wide CA characteristics.

The BriTROC study recruited women with relapsed HGSOC that underwent either image-guided biopsy or intra-operative biopsy[33] during secondary debulking surgery. As a result, the BriTROC cohort is enriched for better-outcome patients with a median survival of 4.9 years (patients had to live long enough to relapse and be well enough to undergo biopsy sampling). In comparison, patients of the OV04 cohort had a median survival of 2.4 years (see **Supplementary Table 1**). We

therefore performed survival analyses on both cohorts independently. Using a multivariable Cox proportional hazard model, we found that the two cohorts showed opposing trends, but that the presence of CA was not significantly associated with overall survival in either of the two cohorts (OV04: p = 0.461, HR = 0.8, CI = 0.43–1.5; BriTROC: p = 0.088, HR = 1.58, CI = 0.93–2.7; **Fig. 4f–g**). To investigate this further, we next focused on data available from The Cancer Genome Atlas (TCGA) database, using the previously described CA20 gene expression signature[32,34] as a proxy marker for CA. Estimating CA20 gene expression scores across the pan-cancer TCGA dataset confirmed the high prevalence of CA in ovarian cancer, which was amongst the highest CA20 expressing cancer types[32] (**Supplementary Fig. 2a**). Moreover, we found a strong correlation between CA20 scores and the CIN gene expression signature, CIN25[35] (Spearman's R = 0.96, $p = 2.2 \times 10^{-16}$; **Supplementary Fig. 2b**). This is consistent with previous studies reporting that CA induces aneuploidy and CIN[8,23]. Importantly, we noticed that cancer types in which CA has previously been implicated as a poor prognostic variable[27,32] (highlighted by blue arrowheads in **Supplementary Fig. 2a**) generally had lower than pan-cancer average CA20 and CIN25 signature expression (see **Supplementary Fig. 2a**). We hypothesised that the survival differences observed in these cancers is caused by CA identifying cases with higher CIN. Further, we reasoned that in cancer subtypes that are primarily driven by severe CIN such as HGSOC[36], the CA phenotype does not have a large effect on poor disease outcome. We therefore separated the TCGA cancer subtypes into CIN-high and CIN-low cancers (median split). When performing multivariable Cox proportional hazard analyses on CIN-low cancers alone confirmed previous findings that high CA20 was associated with poor survival (P < 0.001, HR = 1.5, CI = 1.2–1.9; **Supplementary Fig. 2c**). However, this association was not observed in CIN-high cancers (P < 0.971, HR = 1.0, CI = 0.89–1.1; **Supplementary Fig. 2d**). In addition, high CA20 was also not associated with disease outcome when looking at TCGA ovarian cancer cases alone (patients divided by CA20 median; log-rank test, p = 0.63), confirming our observations in stained tissue samples.

In-depth characterisation of ovarian cancer cell lines shows frequent centrosome amplification through centriole overduplication

Our data showed that CA is a highly prevalent and heterogenous feature in HGSOC tumour tissues. However, several unanswered questions remain. Our tissue staining approach was limited to PCM-targeting antibodies (PCNT and CDK5RAP2) owing to FFPE-associated technical staining challenges. PCM proteins are highly dynamic and can form assemblies without centrioles[37]. As such, it remains to be elucidated whether supernumerary centrosomes detected in HGSOC tissues are truly functional and intact centrosomes or possibly acentriolar PCM foci caused by PCM fragmentation. To investigate this possibility, we next focussed our analysis on the in-depth characterisation of ovarian cancer cell lines. To this end, we developed a semi-automated imaging approach (summarised in **Fig. 5a**) using four independent immunofluorescent stains. Across these four stains, antibody panels were designed to allow quantification of centrosomes (PCNT, CETN3, and CEP164) and micronuclei (Hoechst) at the single cell level (cytokeratin stains were used to estimate cell boundaries), as well as the estimation of mitotic indices (pHH3), and the detection of DNA damage (γH2AX). Using this

approach, we screened 73 ovarian cancer cell lines, comprising a total of 388,748 cells (**Fig. 5b**). Our analyses showed that ovarian cancer cell lines recapitulate the high prevalence of CA previously observed in HGSOC tissue samples. CA frequencies (non-mitotic cells with two or more centrosomes) ranged from 0.5–50.4% with a mean of 26.2%. Examples of cell lines with high and low CA frequencies are shown in **Fig. 6a,b**.

To examine whether supernumerary centrosomes were intact centrosomes or aggregates of PCM, we analysed PCNT co-staining with centriole markers Centrin 3 (CETN3) and CEP164. CEP164 is a distal appendage marker which can only be found on the mother centriole of mature centrosomes. Accordingly, acentriolar PCM assemblies would be expected to be negative for both markers. Extra centrosomes caused by the incorrect or premature segregation of centrioles would be expected to result in CETN3-positive PCM foci of which approximately only 50% would show positive CEP164 staining. Additionally, CA arising from true centriole overduplication would be expected to result in fully functional centrosomes containing two centrioles, a daughter and a CEP164-positive mother centriole, surrounded by PCM. We found that both markers, CETN3 and CEP164, were present in the large majority (>95%) of detected PCM foci (**Fig. 5b**). This suggests that CA in ovarian cancer cells is predominantly caused by true centriole overduplication resulting in intact and therefore likely fully functional centrosomes.

Given the high degree of CIN observed in HGSOC in comparison to other ovarian cancer subtypes, we next asked whether HGSOC cell lines showed higher levels of CA, and whether CA correlated with other molecular features of interest, including DNA content and the presence of micronuclei. We were able to detect micronuclei in the majority of analysed cell lines with micronuclei frequencies ranging from 2.7–48.9% (mean 15.6%), defined as single non-mitotic cells which contain at least one or more micronuclei. High level CA was associated with increased MN frequencies (Spearman's $R = 0.42$, $p \ll 0.001$), providing further evidence that CA might induce MN formation through chromosome missegregation. In addition, we found moderate to strong correlations between CA frequencies and nuclei intensities (surrogate marker for DNA content; Spearman's $R = 0.32$, $p = 0.006$) as well as increased mitotic indices (Spearman's $R = 0.38$, $p < 0.001$). No correlation was observed between CA frequencies and DNA damage (measured as percentage of cells with γH2AX-positive nuclei Spearman's $R = 0.14$, $p = 0.3$; **Fig. 6c**). Importantly, we did also not observe significant differences in CA frequencies across different histological subtypes, including the three non-cancerous (immortalised) cell lines, FT194, FT246 and IOSE4 (Wilcoxon, $p = 0.17$; **Fig. 6d**). These observations further suggest that the transformation and/or immortalisation of normal cells may induce oncogenic features including CA and that these lines may not represent suitable normal controls for CA/CIN studies. In contrast, MN frequencies were significantly higher in HGSOC cell lines compared to all other histological cell line subtypes (Wilcoxon, $p \ll 0.0001$; **Fig. 6e**).

A recent study showed that hypoxia can promote CA via HIF1α-dependent induction of PLK4[38]. Out of the 73 cell lines screened, 25 were derived and maintained in low oxygen growth conditions (5% O2; termed here as low-O2 cell lines). When comparing these lines to those grown in normoxia (21% O2), we found that low-O2 cell lines had significantly higher CA and MN frequencies (Wilcoxon, $p = 0.00044$ and $p \ll 0.0001$, respectively; **Fig. 5f-g**; these results might be confounded

by the fact that the subgroup of low-O2 cell lines only contained HGSOC lines, which showed higher MN frequencies than lines of other ovarian cancer subtypes). To investigate whether increased CA and MN frequencies were caused by lower oxygen growth conditions in our study setting, eight HGSOC cell lines that are normally grown in normoxia were transferred to 5% oxygen for 48 hours and subsequently screened. This showed that reduction in oxygen levels significantly increased CA (Wilcoxon, $p = 2\times10^{-6}$) but had no effect on MN frequencies (Wilcoxon, $p = 1$; **Fig. 5h-i**). Longer exposure to 5% oxygen conditions might be required to observe MN formation. Nevertheless, this suggests that short-term reduction in oxygen levels induces CA but not MN, and that, at least in this setting, CA precedes MN formation.

Centrosome amplification does not correlate with distinct patterns of copy number aberrations but is associated with increased genome subclonality

Multiple studies have previously reported putative associations between centrosome abnormalities and aneuploidy/CIN[8,23,32]. However, possible underlying mechanisms remain poorly understood. We hypothesised that supernumerary centrosomes might present a previously undescribed mutational process in HGSOC and induce distinct patterns of genomic aberrations. We therefore performed low-coverage/shallow whole genome sequencing (sWGS) and absolute copy number (ACN) fitting of ovarian cancer tissue samples and cell lines to identify copy number signatures[39] that might be associated with the CA phenotype.

For the HGSOC tumour tissues, we only included samples with high confidence ACN fits in this analysis (n=54 OV04 and n=84 BriTROC samples). No significant correlations were observed between copy number signatures and CA scores following p-value adjustment for multiple comparisons (**Supplementary Fig. 3a,b**). Consistently, no associations were found between CA and the number of breakpoints, copy number change points, oscillating copy number or copy number segment size (**Supplementary Fig. 3c-d**). Importantly, we did also not observe a correlation between CA scores and tumour ploidy (**Supplementary Fig. 3e**), suggesting that the high prevalence of CA in HGSOC is not linked to whole genome duplication or cytokinesis failure.

ACN fitting from sWGS data from FFPE tumour tissue samples remains highly challenging[40,41] and can be confounded by poor sequencing quality as a result of FFPE processing, tumour heterogeneity and low tumour purity. Given the high purity of cell lines, ACN and derived copy number signature data are more robust in this sample set and might therefore allow the identification of potential CA-associated patterns of genomic aberration that we were unable to detect in FFPE tumour tissues. We therefore next performed tagged amplicon sequencing (TAm-Seq) and sWGS on our ovarian cancer cell line collection. An overview of collated imaging and genomic data across all cell lines included in this study is provided in **Supplementary Fig. 4**. Importantly, genomic analyses confirmed that the sequenced cell lines were very representative of genomic features observed in the overall HGSOC case population[39,42,43]. To further investigate the relationship between CA and CIN, we first quantified genome-wide copy number aberrations by estimating trimmed median absolute deviation from copy number neutrality (tMAD) scores[44] across all cell lines (**Supplementary Fig. 4 and Fig. 7b-d**). Both CA and MN frequencies significantly correlated with tMAD scores (Spearman's

R = 0.36, p = 0.0063; Spearman's R = 0.46, p = 0.00043, respectively); and cell lines which had both, high CA and high MN frequencies showed most extreme genomic aberration (Spearman's R = 0.48, p = 0.00019; **Fig. 7b-d**). Consistent with our HGSOC tissue data, there were no significant associations between CA and copy number signatures (**Fig. 7e**). However, MN frequencies showed a significant negative correlation with signature 1 (Spearman's R = −0.52, p << 0.0001, p.adj = 0.0003), and were positively correlated with signature 4 (Spearman's R = 0.42, p = 0.0011, p.adj = 0.007), and signature 5 (Spearman's R = 0.4, p = 0.0015, p.adj = 0.007; **Fig. 7f**). Given the compositional nature of copy number signatures, we used a fixed effect model on the additive log ratio (ALR)-transformed copy number data to confirm our results and test for global differential abundance of copy number signature exposures between both, CA high vs. low, and MN high vs. low cell lines (**Fig. 7g,h**). These analyses confirmed no statistically significant shift in copy number signature activities between CA high and low cell lines (Wald test, p = 0.913), but showed a significant change in copy number signature abundances observed in MN high vs. low cell lines (Wald test, p = 0.0083). This shift is most likely caused by the decrease in signature 1 (corresponding to ALR1; the signature with the most extreme coefficient) in MN high cell lines. Copy number signature 1 is characterised by a low number of breakpoints and large copy number segment size[39]. Consistent with these results, we found that MN frequencies were strongly correlated with a higher number of breakpoints per chromosome arm (Spearman's R = 0.52, p << 0.001) and negatively correlated with mean segment size (Spearman's R = -0.51, p = 0.000039). We did not observe any correlations between CA frequencies and genomic features, including segment size, number of breakpoints, copy number change points, and oscillating copy number (**Supplementary Fig. 5**). Importantly, in alignment with our tissue-based observations we did also not observe significant correlations between ploidy and CA (Spearman's R = 0.13, p = 0.34, respectively; **Fig. 7a**). This further confirms that cytokinesis failure and other processes impacting tumour ploidy are not involved in driving CA in ovarian cancer.

Finally, ongoing CIN is considered a principal mediator of intra-tumour heterogeneity[9]. Given the strong link between CA and CIN, we next estimated fractions of subclonal segments (genome subclonality) from sWGS data as previously described[40,45] for all sequenced cell lines. Strikingly, CA frequencies were significantly correlated with genome subclonality (Spearman's R = 0.36, p = 0.0069, p = 0.054), while a similar trend was observed for MN (Spearman's R = 0.26, p = 0.054; **Fig. 7i-j**). This suggests that while CA is not associated with specific patterns of genome aberrations in our experimental setting (as read out by sWGS), it might play an important role in driving intra-tumour heterogeneity and tumour evolution.

Cell lines with high centrosome amplification show upregulated survival signalling pathways, and have decreased drug sensitivity

Having shown that our large collection of ovarian cancer cell lines phenocopied patient samples, we next investigated potential drivers and consequences of CA using RNA sequencing of ovarian cancer cell lines. To understand other potential confounding factors that might influence gene expression independent of CA and MN status, we first performed differential gene expression (DGE) analyses comparing different histological subtypes and cell culture growth conditions (see

Supplementary Fig. 6). As expected, in comparison to "normal" cell lines, HGSOC cells showed upregulation of E2F and MYC targets (GSEA; **Supplementary Fig. 6b**). Importantly, however, we did not observe significant upregulation of hypoxia hallmark/response genes in low-O_2 cell lines compared to cell lines cultured at 21% oxygen, and no additive effect was detected between CA and oxygen growth conditions when performing DGE model checks. This suggests that it is the reduction of oxygen levels (compared to baseline) that induces CA (see **Fig. 6h-i**), and that the higher frequencies of CA and MN observed in low-O_2 cell lines are cell line intrinsic rather than caused by differing oxygen growth conditions. Based on these results, we used an additive model accounting for differences in histological subtypes to investigate differentially expressed genes in cell lines with high vs. low CA, and high vs. low MN frequencies (**Fig. 8a-b**, showing volcano plots of differentially expressed genes). Given the correlation between CA and MN (Spearman's $R = 0.38$, $p = 0.0015$; see **Fig. 6c**), cell lines with either high CA or high MN frequencies in part showed similar gene expression signatures, e.g. downregulation of E2F target genes, and upregulation of genes involved in interleukin and interferon signalling (**Fig. 8c-d**). However, the most upregulated pathway in MN high cell lines was interferon alpha response signalling. In contrast, the most significantly upregulated pathway in CA high cell lines was TNFα via NFκB signalling (**Fig. 8c-f**).

NFκB/TNFα signalling play pleiotropic roles in various aspects of cell homeostasis and stress response[46,47], and might therefore play an important role in regulating cell survival in response to CA. Moreover, previous studies have shown that experimentally expanding the degree of aneuploidy in cells can increase their ability to adapt and develop resistance to chemotherapeutic agents[2,5,6]. In this study, CA showed strong correlations with both, genome subclonality and CIN (see **Fig. 7b,d,i**). We therefore hypothesised that CA might be associated with increased resistance to chemotherapy in ovarian cancer cell lines. We therefore selected seven CA-high and seven CA-low lines to test whether the extent of CA would impact drug sensitivities, using a panel of eleven drugs containing both standard-of-care chemotherapeutics and novel anti-mitotic and anti-centrosome agents (**Fig. 9a**). These included oxaliplatin and paclitaxel, as well as drugs targeting centrosome clustering, the spindle assembly checkpoint (SAC) and different components of the centrosome duplication and cell cycle (see Methods for detailed drug screen set-up). We assessed drug sensitivities using the growth rate inhibition (GR) metric[48,49], to adequately control for variations in cell proliferation/division rates and assay duration, and estimated both the potency (GR50, i.e. the concentration of a drug at which cell proliferation is reduced by 50%) and efficacy (GRmax, i.e. the maximum effect of a drug at the highest concentration tested) for each therapeutic agent (Note that an ideal drug would have high potency and high efficacy). We observed that drugs targeting SAC components or components of the centrosome duplication/cell cycle were highly effective in all tested cell lines, inducing cytotoxic responses (**Fig. 9b**). CW069, which impedes centrosome clustering by inhibiting the microtubule motor protein HSET, was less effective and only induced partial cytotoxicity in four out of the 14 lines. Paclitaxel was the most potent drug, followed by the PLK1 inhibitors, BI2536 and Volasertib. Interestingly, cell lines with higher CA frequencies showed overall decreased sensitivity to drug perturbations compared to cell lines with low CA frequencies. This observation was most striking for Paclitaxel treatment. Paclitaxel only induced cytotoxicity in CA low cell lines, whereas in CA high cell

lines it resulted in partial cell growth inhibition only (Spearman's R = 0.65, p = 0.014; **Fig. 9c**). Similarly, BI2536, Volasertib (both PLK1 inhibitors) and AZ3146 (Mps1 inhibitor) showed significantly lower potencies, while Oxaliplatin, and Barasertib (Aurora B inhibitor) showed significantly lower efficacies in CA-high compared to CA-low cell lines (**Fig. 9c**). This differential response to drug agents, and in particular to Paclitaxel, in relation to centrosome numbers highlights the clear implications of centrosome biology on the clinical management of HGSOC.

Discussion

Centrosome amplification is now considered a key feature in many cancers. However, the in-depth characterisation of centrosome abnormalities and associated molecular features in HGSOC is limited. We developed a high-throughput microscopy-based approach to analyse >300 clinical tissue samples and found CA to be highly prevalent feature of HGSOC. Interestingly, a recent study in prostate cancer reported centrosome loss as another centrosomal abnormality that was associated with aneuploidy and disease progression[50], and studies on frozen ovarian cancer tissues suggest that regions of centrosome loss may be common[28]. In our study, only one tissue sample showed significantly lower CA scores than those detected in normal control tissues. Given the high-throughput data-driven approach deployed in this study, and the large amount of imaging data collected (>10,000 images across >300 FFPE tissues), we did not further inspect individual tissue regions for loss of centrosomes. However, consistent with the previous study [28], we observed large variability in the spread of CA scores across all imaging fields within the same tissue, confirming that the CA phenotype displays marked inter- and intra-tissue heterogeneity in HGSOC tumours[24,28]. However, it is important to note that, although selected and imaged tissue samples were enriched for tumour regions, the presence of normal/stromal cells may influence the observed CA score and heterogeneity. In contrast to previous observations in other cancer types[27,32], we did not observe a correlation between the CA phenotype and disease outcome. This might be explained by the high degree of severe CIN observed in virtually all HGSOC cases[1] which is itself associated with poor outcome.

In addition to the large-scale characterisation of tumour tissues, our in-depth characterisation and profiling of >70 ovarian cancer cell lines further confirmed that CA through centriole overduplication is a widespread phenomenon in ovarian cancer cells, recapitulating the high prevalence of CA observed in HGSOC tissues. In addition, our cell line imaging data showed that cells with higher CA were associated with the presence of micronuclei, increased nuclei intensity (often used as a surrogate for DNA content) and mitotic indices, suggesting that cells with CA are delayed in mitosis. This is consistent with previous studies which have reported that cells with multiple centrosomes can take up to twice as long to complete mitosis, as additional time is required during the SAC to cluster centrosomes and align chromosomes in the metaphase plate[15,51–53].

HGSOC is characterised by frequent whole genome duplication and severe CIN; and we recently reported the development of copy number signatures summarising distinct patterns of genomic features associated with CIN[39]. Importantly, we did not observe correlations between CA and ploidy in either our HGSOC tissue sections or our ovarian cancer cell lines. This strongly suggests that while whole genome duplication through cell fusion, mitotic slippage and/or cytokinesis failure may induce CA in some cells, it is not the main driver of CA in ovarian cancer cell lines and tissues. In this study, CA did also not correlate with known copy number signatures, in neither HGSOC tissue samples nor ovarian cancer cell lines. Importantly, early and ubiquitous *TP53* mutations in HGSOC provide a permissive environment for a compendium of multiple mutational processes, including CA, to evolve and simultaneously shape the HGSOC genome[10,39]. This makes it very challenging to decipher out distinct mutational patterns linked to CA. Single-cell sequencing or third generation long-

read sequencing technologies might be required instead to identify genomic patterns associated with CA. Nevertheless, our findings further support the notion that CA correlates with increased genomic aberration, can result in micronuclei formation, and thus likely presents a key driver of CIN[8,18,54] in HGSOC. Importantly, we also observed that high level CA is associated with increased genome subclonality. CA might therefore play an important role in driving intra-tumour heterogeneity and tumour evolution.

Ongoing CIN and tumour heterogeneity have been associated with increased treatment resistance by providing tumours with the ability to more readily adapt to environmental stresses[2,4–6,9]. In line with these findings, we observed decreased sensitivities to several therapeutic agents in cell lines with higher levels of CA. This observation was most severe in cells treated with paclitaxel. While paclitaxel induced cytotoxic responses in CA low cell lines, CA high cell lines only showed partial growth inhibition. Importantly, supernumerary centrosomes have increased capacities for microtubule nucleation[55]. Consequently, higher concentrations of paclitaxel might be required in cells with high CA to inhibit spindle assembly and thus cell division. Centrosome-driven microtubule nucleation is regulated by the PCM, the size of which changes throughout the cell cycle and in preparation for spindle formation. We reported that HGSOC tumours frequently contain centrosomes with enlarged PCM foci in comparison to normal control tissues. However, due to FFPE staining limitations, we were not able to determine whether these size variations were true structural centrosome abnormalities or caused by staining artefacts and/or overlapping centrosomes that were not correctly segmented during spot counting. Nevertheless, further analyses on PCM size variations should be performed to investigate the role of enlarged PCM in CIN and paclitaxel sensitivities in future HGSOC studies.

Lastly, RNA sequencing experiments revealed significant upregulation of TNFα/NFκB signalling in cell lines with high CA. Importantly, TNFα and NFκB signalling have been implicated in various aspects of oncogenesis[46,47,56,57]. For example, activation of NFκB results in the transcriptional upregulation of anti-apoptotic and antioxidant genes, promoting cell proliferation and survival. We therefore propose that NFκB signalling might be a key mediator of the CA “set point” allowing cells to tolerate higher levels of CA. Interestingly, a recent study illustrated CA-associated induction of the oxidative stress response via increased reactive oxygen species (ROS)[20]. High levels of ROS are toxic to cells and can induce cell senescence and apoptosis. Constitutive NFκB activity might therefore provide an important antioxidant mechanism resulting in cell survival and CA tolerance. NFκB-mediated upregulation of survival factors, together with increased levels of CIN and subclonality, might also contribute toward the CA-associated drug resistance observed in this study. In addition, NFκB signalling induces the transcription of many cytokines and chemokines, including IL6 and IL8, which regulate inflammatory responses, as well as cell invasion and migration[57]. Our RNA sequencing experiments showed that IL6 signalling was among the most upregulated signalling pathway in CA high cell lines. These findings further support the recent discovery that CA induces non-cell autonomous invasion through the secretion of pro-invasive factors including both, IL6 and IL8[20]. Together, these data highlight NFκB as a potential mediator of CA-associated oncogenic features, such as increased resistance and invasion. However, future studies will be needed to further investigate the exact mechanisms and relationship between NFκB signalling and CA, and to evaluate

NFκB signalling as a potential vulnerability for therapeutic exploitation in cancers with high-level CIN and CA.

In summary, our large-scale characterisation of centrosome profiles in cell lines and tissue samples provides important evidence for the high prevalence of CA in HGSOC; and for the first time, links supernumerary centrosomes to genomic subclonality and thus tumour evolution. In addition, we highlight the centrosome and CA-associated survival mechanisms as potential and promising targets for novel therapeutic approaches in HGSOC granting further investigation. Finally, our work provides an important resource detailing the phenotypic, genomic, and transcriptomic characterisation of >70 ovarian cancer cell lines. This will facilitate the informed selection of suitable cell line models for future pre-clinical research investigating CIN, CA and associated vulnerabilities in ovarian cancer.

Methods

Clinical samples and primary tissue processing and selection

Tumour samples were obtained from patients enrolled in the Cambridge Translational Cancer Research Ovarian Study 04 (CTCROV04) study approved by the Institutional Ethics Committee (REC08/H0306/61). Tumour samples were processed following standardised operating protocols as outlined in the OV04 study design. For the generation of FFPE tissue blocks, tumour tissues were suspended in 10% neutral buffered formalin (NBF) for 24 hours and subsequently transferred into 70% ethanol for paraffin embedding and sectioning. Tissues for the OV04 sample set was chosen to be representative of the overall HGSOC case population, while taking into account tissue availability. The sample set originally comprised a total of 120 FFPE tissues from 105 OV04 patients. However, 27 samples were excluded prior to centrosome analyses owing to poor tissue quality and purity. BriTROC TMAs were previously generated by the McNeish Laboratory (Imperial College London) using a total of 247 samples from 172 BriTROC patients. Three 1 mm cores were collected from each tissue sample and placed into three identically laid-out sister-TMAs. Prior to centrosome analyses, H&E stains of resulting TMAs were inspected, and tissue cores/samples that mostly contained stromal or necrotic regions were excluded (n = 53). An overview of all remaining tumour samples (n = 287) and patients (n = 226) from both cohorts included in this study is shown in **Supplementary Table 1**. In addition to the 287 tumour samples, we also included 24 normal fallopian tube (FT) samples, one normal spleen, one normal kidney and five normal liver samples. Importantly, the liver is the only organ in the human body that displays centrosome amplification in a healthy state, likely caused by multi-nucleation and polyploidisation during tissue development[29–31]. Liver tissues were therefore used as positive controls in this study.

Cell Lines

Cell lines used in this study are listed in **Supplementary Table 2**, together with their associated growth conditions and culture media. In general, cell lines were grown at 37℃ and 5% CO_2 according to ATCC/ECACC recommendations. OSE medium is composed of 50:50 medium 199 (Sigma-5017) and medium 105 (Sigma-6395). Cells were tested for mycoplasma contaminations on a regular basis using the qPCR PhoenixDx Mycoplasma kit (Procomcure Biotech), and cell line identities were confirmed prior to DNA extractions using our in-house human short tandem repeat (STR) profiling cell authentication service.

Cell line growth curves

Cell line growth curves were generated to estimate optimal cell seeding densities for subsequent drug and centrosome characterisation assays. Cell lines were seeded into 96-well plates at different densities ranging from 5,000–20,000 cells/well. Plates were placed inside a real-time live-cell system (IncuCyte®, Sartorius) and monitored for four days. Phase contrast images were acquired at 3 hour intervals and confluence measures were collected. Cell seeding densities that resulted in 80–90% confluency at the end of the assay (e.g. four days) were chosen for each cell line.

Immunofluorescent centrosome staining

1. FFPE tissue samples (also see **Supplementary Methods**)

Tissue samples were fixed in 10% NBF for 24 hours, embedded in paraffin and cut into 25 µm thick sections. Subsequently, tissue slides were deparaffinised and rehydrated by resuspending them into 1xPBS three times for 5 min each. Antigen retrieval was performed via heat-induced epitope retrieval (HIER) in universal antigen retrieval buffer (ab208572, Abcam). Deparaffinised tissue slides were immersed in antigen retrieval buffer which was brought to boil using a conventional microwave. Once boiling temperature was reached, tissue slides were boiled for 10 mins at medium-high power and subsequently incubated at room temperature for 25–30 mins to allow buffer and slides to cool down. Slides were washed three times in 1xPBS for 5 min each. Subsequently, tissue sections were permeabilised in permeabilisation buffer (5% BSA, 0.3% TritonX100) for 25–30 mins, and blocked in 5% BSA for 2 hours at room temperature.

Primary antibodies (anti-Pericentrin, Abcam ab28144; anti-CDK5RAP2, Gergely Lab) were diluted in 5% BSA, added to the tissue slides and incubated at 4 °C overnight. The anti-CDK5RAP2 antibody was used at a concentration of 2 µg/ml and the ab28144 anti- Pericentrin antibody was used at a concentration of 5 µg/ml. Following incubation with primary antibodies, tissue slides were thoroughly washed three times in 0.1% Tween 20 in 1×PBS for 8 mins each. Secondary antibodies (Alexa Fluor 488 and Alexa Fluor 555, cat No.: A-11001 and A-21429, Invitrogen) were diluted in 5% BSA to a concentration of 4 µg/ml (1:500) and incubated for 1 hour at 37 °C. Subsequently, slides were washed again three times in 0.1% Tween 20 in 1×PBS for 8 mins each, rinsed with distilled water and counterstained with 1 µg/ml Hoechst (33342, B2261 Sigma-Aldrich) for 15 mins at room temperature. Tissue slides were mounted in glycerol / n-propyl gallate medium (4% n-propyl gallate (02370, Sigma-Aldrich), 8 ml 100% glycerol, and 2 ml 1× PBS) and incubated for 24–48 hours protected from light to allow tissue equilibration and diffusion of mounting media prior to imaging.

2. Cell lines

Optical 96–well plates (CellCarrier-96 Ultra, Perkin Elmer) were coated with sterile Poly-L-Lysine (P4832, Sigma). Cells were seeded using previously estimated cell seeding densities and cultured for four days until confluent (see **Fig. 5a** for an overview of the experimental set-up). Cells were fixed in 100% ice-cold methanol for 5 mins at −20 °C, subsequently washed three times in 1×PBS and stored in 200 µl PBS at 4 °C until further use. Cells were rinsed with 0.1% Tween 20 in 1×PBS and permeabilised with 1% TritonX100 + 0.5% NP40 in 1×PBS for 5 mins at room temperature. Subsequently, cells were blocked for 1 hour at room temperature using 5% BSA in 1×PBS. Primary non-conjugated antibodies were diluted in blocking buffer, added to the cells and incubated overnight at 4 °C. An overview of antibodies, provider information, and concentrations for each of the four stains is shown in **Supplementary Table 3**. Following primary antibody incubation, cells were washed three times with 0.1% Tween 20 in 1×PBS. Secondary antibodies (Alexa Fluor 555 and Alexa Fluor 647, cat No.: A-21429 and A-21236, Invitrogen) were diluted to 1 µg/ml (1:2000) in blocking buffer, added to the cells and incubated for 1 hour at 37 °C. Subsequently, cells were washed three times with 0.1% Tween 20 in 1×PBS, and re-blocked with 5% BSA in 1×PBS for 30 mins at room temperature. Subsequently, conjugated primary antibodies (Table 7.3) were diluted, added to the cells and

incubated for 1 hour at room temperature, after which cells were washed again three times with 0.1% Tween 20 in 1×PBS. After the last washing step, cells were rinsed in Milli-Q water. Hoechst 33342 was diluted to 1 µg/ml in MilliQ water, added to the cells and incubated for 15-30 mins at room temperature. Cells were washed and stored in 1×PBS until further use.

Microscopy

1. SP8

For method optimisation purposes (see **Supplementary Methods**), sequential confocal images were acquired using a Leica TCS SP8 microscope. Z-stacks were collected at a step size of 0.35 µm. The pinhole diameter for channels containing centrosome staining signals was set to 0.5 AU. White light lasers with wavelengths of 488 and 555 nm were used to excite AlexaFluor 488 and AlexaFluor 555, respectively. Hoechst signal was excited using a laser diode 405.

2. Operetta

For large scale centrosome profiling experiments, the Operetta CLS™ high-content analysis system was used. For FFPE tissue sections, the Operetta PreciScan™ feature was used to allow automatisation of imaging processes – whole microscopy slides were initially scanned a 5× (PreScan), global images were filtered using Gaussian blur and tissue regions were identified based on Hoechst 33342 signal. The geometric centre of the tissue was then calculated and expanded to create a region of interest covering most of the tissue. Following morphological properties calculations, 50 independent non-overlapping imaging fields were placed into the tissue region of interest and subsequently imaged (ReScan) at high resolution using a confocal 63× 1.15NA water objective (see **Supplementary Methods**). For each imaging field, Z-stacks were generated at a step size of 0.5µm (100 planes in total) across all imaging channels (i.e. 405 nm, 488 nm and 555 nm wavelengths).

Cell lines were imaged in filtered 1 × PBS as mounting medium (200 µl/well) using a confocal 40× 1.1NA water objective. Images from five independent non-overlapping imaging fields were acquired from each well (see **Fig. 5a**). For each imaging field, Z-stack images were collected at a step size of 0.5 µm (28 planes in total) across all imaging channels (i.e. 405 nm, 488 nm, 555 nm and 647 nm wavelengths).

Image Analysis

Image analyses were performed using the Harmony software. Image analysis scripts were optimised for different sample types and immunofluorescent stains. In brief, images were reconstructed as maximum projections (collapsed Z-stacks) using basic brightfield correction. Image analyses included segmentation of nuclei, removal of border objects to only include whole cells, identification of cytoplasm and regions of interest, and subsequent centrosome detection and spot counting (see **Supplementary Methods**). Each image analysis step included a series of quality controls based on morphological and intensity properties of identified objects. The centrosome amplification score (CA score) was estimated as the total number of centrosomes divided by the total number of nuclei detected in each imaging field.

Statistical modelling of centrosome amplification score (CA score) variabilities

To model the distribution of CA scores across imaging fields within tissue samples and within tissue types (displayed in **Fig. 2**), we developed a linear mixed model[58] allowing

1. to detect differences in average CA scores between tissue types by means of fixed effects,
2. to take the CA within-tissue dependence into account by means of normally distributed random intercepts,
3. the spread (variance) of within-tissue CA scores to vary between tissues by means of gamma distributed standard deviations and to vary between tissue types by means of fixed effects,
4. the mean and variance of within-tissue CA scores to be dependent by using a Gaussian copula to model the association between the random intercepts of point 2 and random standard deviations of point 3,
5. the model residuals to be heavy-tailed by considering the generalised T (taking the Gaussian distribution as limit case) as conditional error distribution (i.e., distribution of the error terms given the fixed and random effects described in points 1 to 4).

Parameter estimates and corresponding inference were respectfully obtained by means of the iterative bootstrap[59] and the indirect inference method[60]. Monte Carlo simulations showed that the estimator was consistent and that the confidence interval coverages were close to the 0.95 nominal values.

H&E purity estimation

H&E sections from FFPE tissues were sent to our pathologist for tumour marking and purity estimation. In addition, H&E sections were scanned and subjected to HALO, an image analysis platform for quantitative tissue analysis in digital pathology. HALO's random forest classifier was used to separate the H&E image into tumour, stroma and microscope glass slide, allowing tumour purity estimation.

DNA and RNA extraction

1. FFPE tissue samples

For each FFPE sample, multiple sections at 10 µm thickness were cut depending on tissue size and tumour cellularity assessed by a pathologist, who marked tumour areas on separate Haematoxylin and Eosin (H&E) stained sections to guide microdissection for DNA extraction. Marked tumour areas from unstained FFPE sections were scraped off manually using a scalpel, dewaxed in xylene and subsequently washed with 100% ethanol. Following complete removal and evaporation of residual ethanol (10 mins at 30℃) DNA was extracted using the AllPrep DNA/RNA FFPE Kit (Qiagen). DNA was eluted in 40 µl Elution buffer.

2. Cell Lines

Cell pellets of approximately $1x10^6$ cells were generated from cultured cells for each cell line outlined above and stored at -80℃ until further use. DNA and RNA were extracted from cell pellets using the

Maxwell® RSC Cultured Cells DNA and RNA Kit (Promega, AS1620 and AS1390, respectively) with the Maxwell® RSC 48 Instrument (Promega, AS8500).

DNA sequencing

1. Tagged-Amplicon Sequencing (TAmSeq)

Extracted DNA samples were diluted to a final concentration of 10 ng/ml using PCR certified water. Tagged-Amplicon deep sequencing was performed as previously described[61]. In short, libraries were prepared in 48.48 Juno Access Array Integrated Fluidic Circuits chips (Fluidigm, PN 101-1926) on the IFC Controller AX instrument (Fluidigm) using primers designed to assess small indels and single nucleotide variants across the coding region of the *TP53* gene. Following target-specific amplification and unique sample indexing, libraries were pooled and subsequently purified using AMPure XP beads. Quality and quantity of the pooled library were assessed using a D1000 genomic DNA ScreenTape (Agilent 5067-5582) on the Agilent 4200 TapeStation System (G2991AA), before submitting the library for sequencing to the CRUK CI Genomics Core Facility using 150bp paired-end mode on either the NovaSeq 6000 (SP flowcell) or HiSeq 4000 system. Sequencing reads were aligned to the 1000 Genomes Project GRCh37-derived reference genome (i.e. hs37d5) using the 'BWA-MEM' aligner. Data analysis and variant calling was performed as previously described[62].

2. Shallow Whole Genome Sequencing (sWGS)

DNA extractions were performed as described above, and quantified using Qubit quantification (ThermoFisher, Q328851). DNA samples were diluted to 75 ng in 15 µl of PCR certified water, and sheared by sonication with a target of 200-250bp using the LE220-plus Focused-Ultrasonicator (Covaris) with the following settings: 120 sec at room temperature; 30% duty factor; 180W peak incident power; 50 cycles per burst. Sequencing libraries were prepared using the SMARTer Thruplex DNA-Seq kit (Takara), with each FFPE sample undergoing 7 PCR cycles, and all other samples undergoing 5 PCR cycles for unique sample indexing and library amplification. AMPure XP beads were used following manufacturer's recommendations to clean prepared libraries, which were subsequently eluted in 20 µl TE buffer. sWGS libraries were quantified and quality-checked using D5000 genomic DNA ScreenTapes (Agilent 5067-5588) on the Agilent 4200 TapeStation System (G2991AA) before pooling the uniquely indexed samples in equimolar ratios. Pooled libraries were sequenced at low coverage (~ 0.4 × coverage) on either NovaSeq 6000 S1 flowcells with paired-end 50 bp reads for clinical tissue samples, or the HiSeq 4000 with single 50 bp reads, at the CRUK CI Genomic Core Facility. Resulting sequencing reads were aligned to the 1000 Genomes Project GRCh37-derived reference genome (i.e. hs37d5) using the 'BWA' aligner (v.0.07.17) with default parameters.

Copy number analyses

Relative copy number data was obtained using the QDNAseq[63] R package (v1.24.0) to count reads within 30 kb bins, followed by read count correction for sequence mappability and GC content, and copy number segmentation. QDNAseq data were then subjected to downstream analyses using the Rascal R package[40] for ploidy and cellularity estimation and absolute copy number fitting. Putative

driver amplifications were detected and identified using the Catalogue Of Somatic Mutations In Cancer (COSMIC; https://cancer.sanger.ac.uk/cosmic/help/cnv/overview) definitions and thresholds for high level amplifications and homozygous deletions: *Amplification* – average genome ploidy ≤2.7 and total copy number ≥5; or average genome ploidy >2.7 and total copy number ≥9. *Deletion* – average genome ploidy ≤2.7 and total copy number = 0; or average genome ploidy >2.7 and total copy number <(average genome ploidy − 2.7). In contrast, copy number gains and losses were defined as copy number > ploidy + 1, or copy number < ploidy − 1, respectively, and are not defined as distinct copy number alterations in the same manner as amplifications and deletions.

Copy number signatures (CNS)

Copy number signatures were estimated as previously described[39] from absolute copy number data, following computation of the distributions of six genomic features: I. the breakpoint count per 10 Mb, II. the copy number of each segment, III. the copy number change points (i.e. the difference in copy number between adjacent segments), IV. the breakpoint count per chromosome arm, V. the lengths of chains of oscillating copy number segments, and VI. the size of segments.

Fixed effect model (for CNS analyses)

To account for the compositional nature of copy number signatures (the sum of copy number signature exposures within each samples = 1), we performed additive log ratio (ALR) transformation[64] in a sample-wise manner, choosing signature 7 as the denominator. Subsequently a fixed effect regression model was applied to investigate differences in signature abundances between CA and MN high vs. low samples. For determining global statistically significant shifts in copy number signature abundances we used a Wald test.

tMAD scores

Trimmed median absolute deviation from CN neutrality (tMAD) scores were estimated as a genome-wide summary measure of genomic abnormality as described before[44]. The R code was adjusted from https://github.com/sdchandra/tMAD. Samples were downsampled to 3×10^6 reads and allocated into 500 kb non-overlapping bins as recommended by the authors. The diploid normal ovarian surface epithelium cell line, IOSE4, which was the cell line with the least (no) copy number aberrations, was chosen as normal/control sample for normalisation.

TCGA dataset analysis

Publicly available clinical data and and RNAseq V2 data from The Cancer Genome Atlas (TCGA; https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) were downloaded using Firebrowse (http://firebrowse.org/). Centrosome amplification gene expression (CA20) signatures were calculated as previously described[32]. In brief, gene-level read counts were quantile normalised using Voom[65] and the CA20 score was estimated for each sample as the sum of the log2 median-centred gene expression levels of the CA20 signature genes: *AURKA*, *CCNA2*, *CCND1*, *CCNE2*, *CDK1*, *CEP63*, *CEP152*, *E2F1*, *E2F2*, *LMO4*, *MDM2*, *MYCN*, *NDRG1*, *NEK2*, *PIN1*,

PLK1, *PLK4*, *SASS6*, *STIL*, and *TUBG1*. The CIN25 gene expression signature, as previously defined[35], was estimated in the same manner using the following genes: *C20orf24*/*TGIF2*, *CCNB2*, *CCT5*, *CDC45*, *CDC2*, *ESPL1*, *FEN1*, *FOXM1*, *H2AFZ*, *KIF20A*, *MAD2L1*, *MCM2*, *MCM7*, *MELK*, *NCAPD2*, *PCNA*, *PRC1*, *RAD51AP1*, *RFC4*, *RNASEH2A*, *TOP2A*, *TPX2*, *TRIP13*, *TTK*, *UBE2C*.

RNA sequencing

Following RNA extractions, RNA integrity number (RIN) values were estimated using the Bioanalyzer system (Agilent). RIN values ranged from 8.3–10 with a mean of 9.7. Library preparation and RNA sequencing (RNAseq) were performed by the CRUK CI Genomics Core Facility: In brief libraries were prepared using the Illumina TruSeq standard mRNA kit and sequenced using 50 bp paired-end mode on the NovaSeq 6000 sequencer (S1 flowcell).

RNAseq analyses

Sequencing reads were mapped to the GRCh38 (hg38) reference genome using STAR aligner v2.7.6a (https://github.com/alexdobin/STAR/releases). Following standard quality control (MultiQC[66]), raw sequencing reads were quantified against the same reference genome (GRCh38) using Salmon v1.4.0[67], https://combine-lab.github.io/salmon/). Quantified transcripts (gene-level) were imported into R using the tximport[68] in preparation for downstream analyses. Low-abundance genes were filtered out, keeping only genes with at least 25 reads in at least two samples. Gene count normalisation and differential gene expression analyses were performed using DESeq2 v1.26.0[69]. Genes with p-value < 0.05, FDR < 0.1 and a fold-change of at least 1.5 were considered to be differentially expressed. Likelihood ratio tests were used to assess which model (simple, additive or interactive) was most appropriate for each comparison. Data was visualised using the lfcShrink() function (DESeq2 package) and the ComplexHeatmap R package v2.2.0[70]. Gene set enrichment analyses (GSEA) were performed using the fgsea R package v1.12.0 (https://github.com/ctlab/fgsea). Genes were ranked by statistical significance and fold change; and 50 hallmark sets from the human C5 MSigDB collection were used for running GSEA. 1,000 permutations were used to calculate p-values. In addition, the clusterProfiler R package v3.14.3[71] was used for gene ontology-based overrepresentation analyses.

Drug sensitivity assays – GR metrics

Cell lines were seeded into opaque-walled 384 well plates compatible with the CLARIOstar® Plus (BMG Labtech) luminometer using a CombiDispenser (Thermo ScientificTM). For each cell line, two plates were generated – a control plate to correct for the number of divisions taking place over the course of the assay, and a test plate. 24 hours after seeding, 10 μl of CellTiter-Glo® (Promega) were added to the control plate using the CombiDispenser (Thermo ScientificTM), and after an incubation time of 30 mins the plate was read using the CLARIOstar® Plus luminometer (BMG Labtech). At the same time (24 hours post seeding), drugs were added to the test plate using a D300e Digital Dispenser (Tecan). Plate layouts were randomised, normalised for DMSO volumes across wells, and drugs were dispensed as 10-point half-log dilution series with a DMSO concentration of 0.01%. The following compounds were used: oxaliplatin (Selleckchem S1224), paclitaxel (Selleckchem S1150),

764 tozasertib (Selleckchem S1048), alisertib (Selleckchem S1133), barasertib (Selleckchem S1147),
765 cw069 (Selleckchem S7336), bi2536 (Selleckchem S1109), volasertib (Selleckchem S2235),
766 cfi400945 (Selleckchem S7552), az3146 (Selleckchem S2731), and gsk923295 (Selleckchem
767 S7090). The highest drug concentration used for paclitaxel was 0.1 μM. For all other compounds, the
768 highest tested concentration was 100 μM. Drug screens were performed in triplicates. Following the
769 addition of drugs, treatment plates were incubated for 72 hours, after which cell viabilities were
770 assessed as described above. To estimate drug sensitivity, as well as drug potency and efficacy for
771 each cell line and compound, the growth rate inhibition metrics were estimated as previously
772 described[48,49] using the GRmetrics R package v1.12.2
773 (https://bioconductor.org/packages/release/bioc/html/GRmetrics.html).

Acknowledgements

We thank all patients who participated in and donated tissue samples to this study. The Addenbrookes Human Research Tissue Bank is supported by the NIHR Cambridge Biomedical Research Centre. We also thank the OV04 study team for their help with clinical tissue samples. We would like to thank the Cancer Research UK Cambridge Institute Microscopy, Compliance & Biobanking, Genomics, Bioinformatics, and IT & Scientific Computing core facilities for their support with various aspects of this study. We would like to thank D. Provencher and A.-M Mes-Masson for kindly donating a subset of 25 ovarian cancer cell lines to us that were included in this study. These cell lines were derived with the support of the Banque de tissus et de données of the Réseau de recherche sur le cancer of the Fonds de recherche du Québec - Santé (FRQS) affiliated with the Canadian Tumor Repository Network (CTRNet). We also thank Fanni Gergely for her kind donation of the CDK5RAP2 antibody, and Gayathri Chandrasekaran for her help and advice on the centrosome staining protocol.

We acknowledge funding and support from Cancer Research UK, and the Cancer Research UK Cambridge Centre (A22905, A25177, A25117). L.M.G. was supported by the Wellcome Trust PhD programme in Mathematical Genomics and Medicine (grant number RG92770). This research was also supported by the Ovarian Cancer Action (grant number 006), and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). Work in the Cancer Molecular Diagnostics Laboratory/Blood Processing Laboratory was supported by the NIHR Cambridge Biomedical Research Centre, Cancer Research UK Cambridge Centre and the Mark Foundation Institute for Integrated Cancer Medicine. The views expressed are those of the authors and not necessarily those of Cancer Research UK, the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization, C.M.S., J.T., M.V. and J.D.B.; Methodology, C.M.S. and D.L.C; Validation, C.M.S.; Formal Analysis, C.M.S. and D.L.C.; Investigation, C.M.S., J.A.H., A.M.P., J.G., J.T., M.V.; Resources, C.M.S., D.L.C., T.B., A.P., A.S., M.D.E., P.S., K.H., M.A.V.R., L.M.G, A-M.M-M., D.E., D.M., A.H., I.A.M., M.J.L., F.C.M., and J.T.; Data Curation, C.M.S., D.L.C., T.B., P.S., and K.H.; Writing - Original Draft, C.M.S.; Writing - Review & Editing, C.M.S., J.T., M.V., and J.D.B.; Visualisation, C.M.S.; Supervision, C.M.S., J.T., M.V., and J.D.B.; Project Administration, C.M.S. and M.V. Funding Acquisition, J.D.B.

Conflicts of Interest

J.D.B. is a co-founder and shareholder of Tailor Bio Ltd. There are no further conflicts of interests to declare.

References

1. Morden, C. R. *et al.* Chromosome instability is prevalent and dynamic in high-grade serous ovarian cancer patient samples. *Gynecol Oncol* (2021) doi:10.1016/j.ygyno.2021.02.038.
2. Ippolito, M. R. *et al.* Gene copy-number changes and chromosomal instability induced by aneuploidy confer resistance to chemotherapy. *Dev Cell* **56**, 2440-2454.e6 (2021).
3. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893.e13 (2018).
4. Lee, A. J. X. *et al.* Chromosomal instability confers intrinsic multidrug resistance. *Cancer Res* **71**, 1858–1870 (2011).
5. Lukow, D. A. *et al.* Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev Cell* **56**, 2427-2439.e4 (2021).
6. Shaikh, N. & McClelland, S. E. Diversity in chromosome numbers promotes resistance to chemotherapeutics. *Dev Cell* **56**, 2399–2400 (2021).
7. Minussi, D. C. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302–308 (2021).
8. Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278–282 (2009).
9. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: Therapeutic and diagnostic challenges. *EMBO Rep* **13**, 528–538 (2012).
10. Tamura, N. *et al.* Specific Mechanisms of Chromosomal Instability Indicate Therapeutic Sensitivities in High-Grade Serous Ovarian Carcinoma. *Cancer Res* **80**, 4946–4959 (2020).
11. Conduit, P. T. *et al.* The centrosome-specific phosphorylation of Cnn by Polo/Plk1 drives Cnn scaffold assembly and centrosome maturation. *Dev Cell* **28**, 659–669 (2014).
12. Conduit, P. T., Wainman, A. & Raff, J. W. Centrosome function and assembly in animal cells. *Nat Rev Mol Cell Biol* **16**, 611–624 (2015).
13. Nigg, E. A. & Holland, A. J. Once and only once: mechanisms of centriole duplication and their deregulation in disease. *Nat Rev Mol Cell Biol* **19**, 297–312 (2018).
14. Mason, J. M. *et al.* Functional characterization of CFI-400945, a polo-like kinase 4 inhibitor, as a potential anticancer agent. *Cancer Cell* **26**, 163–176 (2014).
15. Kwon, M. *et al.* Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes Dev* **22**, 2189–2203 (2008).
16. Marteil, G. *et al.* Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. *Nat Commun* **9**, (2018).
17. Tischer, J. & Gergely, F. Anti-mitotic therapies in cancer. *Journal of Cell Biology* **218**, 10–11 (2019).
18. Godinho, S. A. & Pellman, D. Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130467–20130467 (2014).

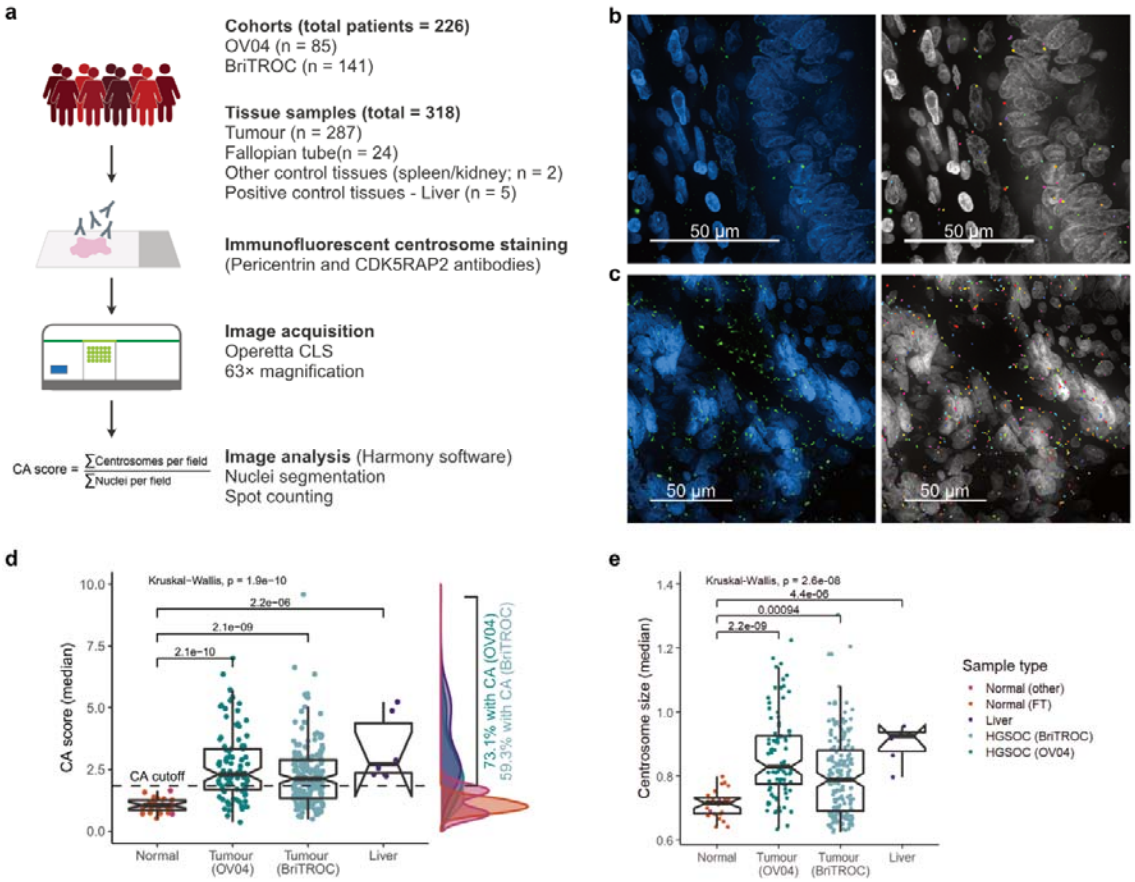
19. Silkworth, W. T., Nardi, I. K., Scholl, L. M. & Cimini, D. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS One* **4**, (2009).
20. Arnandis, T. *et al.* Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. *Dev Cell* **47**, 409-424.e9 (2018).
21. Levine, M. S. *et al.* Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Dev Cell* **40**, 313-322.e5 (2017).
22. Pihan, G. A. Centrosome Dysfunction Contributes to Chromosome Instability, Chromoanagenesis, and Genome Reprograming in Cancer. *Front Oncol* **3**, 1–26 (2013).
23. Raff, J. W. & Basto, R. Centrosome Amplification and Cancer: A Question of Sufficiency. *Dev Cell* **40**, 217–218 (2017).
24. Dias Louro, M. A., Bettencourt-Dias, M. & Bank, C. Patterns of selection against centrosome amplification in human cell lines. *PLoS Comput Biol* **17**, 1–22 (2021).
25. Wong, Y. L. *et al.* Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science (1979)* **348**, 1155–1160 (2015).
26. Galofré, C. *et al.* Centrosome reduction in newly-generated tetraploid cancer cells obtained by separase depletion. *Sci Rep* **10**, 1–12 (2020).
27. Chan, J. A clinical overview of centrosome amplification in human cancers. *Int J Biol Sci* **7**, 1122–1144 (2011).
28. Morretton, J.-P. *et al.* A catalog of numerical centrosome defects in epithelial ovarian cancers. *EMBO Mol Med* **14**, e15670 (2022).
29. Donne, R., Sangouard, F., Celton-Morizur, S. & Desdouets, C. Hepatocyte polyploidy: Driver or gatekeeper of chronic liver diseases. *Cancers (Basel)* **13**, 1–19 (2021).
30. Guidotti, J. E. *et al.* Liver cell polyploidization: A pivotal role for binuclear hepatocytes. *Journal of Biological Chemistry* **278**, 19095–19101 (2003).
31. Margall-Ducos, G., Celton-Morizur, S., Couton, D., Brégerie, O. & Desdouets, C. Liver tetraploidization is controlled by a new process of incomplete cytokinesis. *J Cell Sci* **120**, 3633–3639 (2007).
32. de Almeida, B. P., Vieira, A. F., Paredes, J., Bettencourt-Dias, M. & Barbosa-Morais, N. L. Pan-cancer association of a centrosome amplification gene expression signature with genomic alterations and clinical outcome. *PLoS Comput Biol* **15**, 1–31 (2019).
33. Goranova, T. *et al.* Safety and utility of image-guided research biopsies in relapsed high-grade serous ovarian carcinoma - Experience of the BriTROC consortium. *Br J Cancer* **116**, 1294–1301 (2017).
34. Ogden, A., Rida, P. C. G. & Aneja, R. Prognostic value of CA20, a score based on centrosome amplificationassociated genes, in breast tumors. *Sci Rep* **7**, 1–11 (2017).
35. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043–1048 (2006).

36. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45**, 1127–1133 (2013).
37. Woodruff, J. B., Wueseke, O. & Hyman, A. A. Pericentriolar material structure and dynamics. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130459- (2014).
38. Mittal, K. *et al.* Hypoxia drives centrosome amplification in cancer cells via HIF-1α-dependent induction of polo-like kinase 4. *Molecular Cancer Research* **4**, molcanres.MCR-20-0798-E.2020 (2021).
39. Macintyre, G. *et al.* Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet* **50**, 1262–1270 (2018).
40. Sauer, C. M. *et al.* Absolute copy number fitting from shallow whole genome sequencing data. *bioRxiv* 2021.07.19.452658 (2021).
41. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413–421 (2012).
42. Bell, D. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
43. Bowtell, D. D. *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* **15**, 668–679 (2015).
44. Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* **10**, 1–14 (2018).
45. van Dijk, E. *et al.* Chromosomal copy number heterogeneity predicts survival rates across cancers. *Nat Commun* **12**, 1–12 (2021).
46. Luo, J. L., Kamata, H. & Karin, M. IKK/NF-κB signaling: Balancing life and death - A new approach to cancer therapy. *Journal of Clinical Investigation* **115**, 2625–2632 (2005).
47. Tornatore, L. *et al.* Cancer-Selective Targeting of the NF-κB Survival Pathway With Gadd45β/Mkk7 Inhibitors. *Cancer Cell* **26**, 495–508 (2014).
48. Hafner, M., Niepel, M. & Sorger, P. K. Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics. *Nat Biotechnol* **35**, 500–502 (2017).
49. Hafner, M., Niepel, M., Chung, M. & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods* **13**, 521–527 (2016).
50. Wang, M., Nagle, R. B., Knudsen, B. S., Cress, A. E. & Rogers, G. C. Centrosome loss results in an unstable genome and malignant prostate tumors. *Oncogene* **39**, 399–413 (2020).
51. Acilan, C. & Saunders, W. S. A Tale of Too Many Centrosomes. *Cell* **134**, 572–575 (2008).
52. Basto, R. *et al.* Flies without Centrioles. *Cell* **125**, 1375–1386 (2006).
53. Yang, Z., Lončarek, J., Khodjakov, A. & Rieder, C. L. Extra centrosomes and/or chromosomes prolong mitosis in human cells. *Nat Cell Biol* **10**, 748–751 (2008).
54. Vitre, B. D. & Cleveland, D. W. Centrosomes, chromosome instability (CIN) and aneuploidy. *Curr Opin Cell Biol* **24**, 809–815 (2012).
55. Godinho, S. a *et al.* Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* **510**, 167–171 (2014).
56. Godwin, P. *et al.* Targeting nuclear factor-kappa B to overcome resistance to chemotherapy. *Front Oncol* **3 MAY**, 1–10 (2013).

57. Taniguchi, K. & Karin, M. NF-κB, inflammation, immunity and cancer: Coming of age. *Nat Rev Immunol* **18**, 309–324 (2018).
58. McCulloch, C. E., Searle, S. R. & Neuhaus, J. M. Generalized, linear, and mixed models, 2nd edn. 424 (2008).
59. Kuk, A. Y. C. Asymptotically Unbiased Estimation in Generalized Linear Models with Random Effects. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 395–407 (1995).
60. Guerrier, S., Dupuis-Lozeron, E., Ma, Y. & Victoria-Feser, M. P. Simulation-Based Bias Correction Methods for Complex Models. *https://doi.org/10.1080/01621459.2017.1380031* **114**, 146–157 (2018).
61. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, (2012).
62. Piskorz, A. M. *et al.* Methanol-based fixation is superior to buffered formalin for next-generation sequencing of DNA from clinical cancer samples. *Annals of Oncology* **27**, 532–539 (2016).
63. Scheinin, I. *et al.* DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* **24**, 2022–2032 (2014).
64. Aitchison, J. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)* **44**, 139–160 (1982).
65. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, 1–17 (2014).
66. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
67. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
68. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1–19 (2016).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).
70. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
71. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

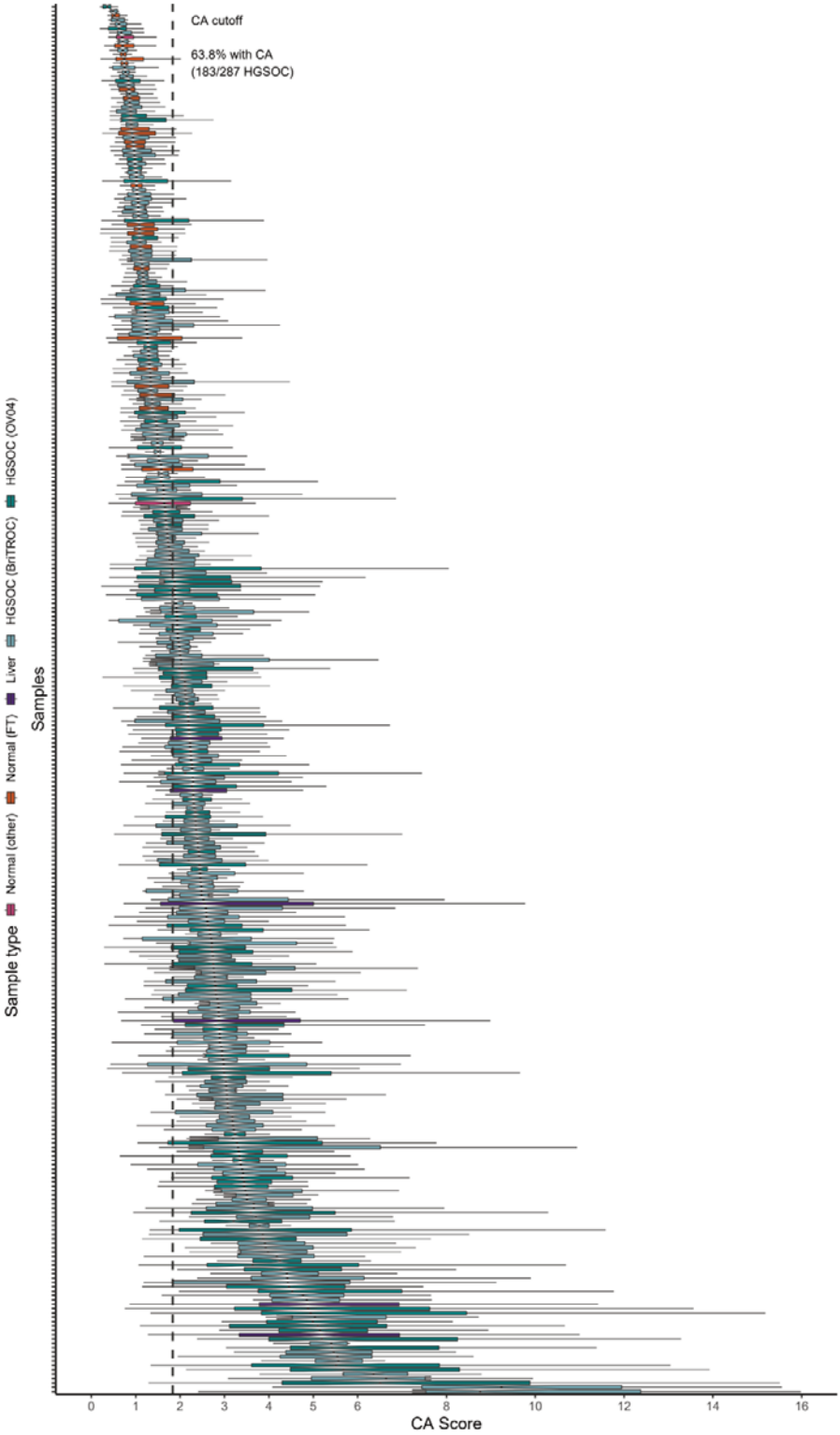
Main Figures

Figure 1 – Centrosome characterisation in HGSOC tissue samples.



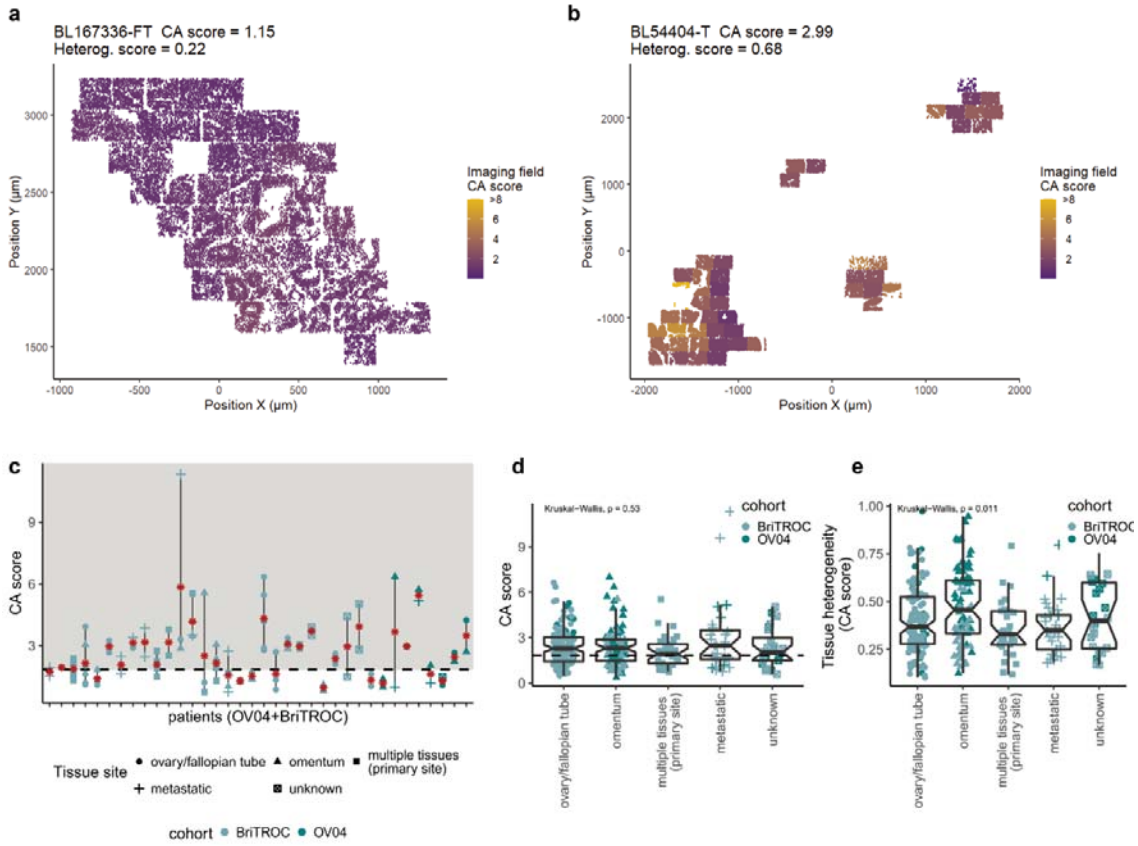
(**a**) Basic workflow of centrosome characterisation in clinical tissue samples (also see **Supplementary Methods**). Example confocal immunofluorescent images (max. intensity projections) of a fallopian tube and HGSOC tissue are shown in (**b**) and (**c**), respectively (left panel). Nuclei are shown in blue, centrosomes are shown in green. Individually detected centrosomes following automated image analysis are highlighted in rainbow colours (right panel; nuclei shown in grey). Scale bars = 50 µm. (**d**) and (**e**) show sample type comparison of centrosome amplification (CA) scores and centrosome size, respectively. Sample types are indicated by different colours. Dashed horizontal line indicates CA cutoff threshold of ~1.83.

Figure 2 – Centrosome amplification profiles across 318 tissue samples.



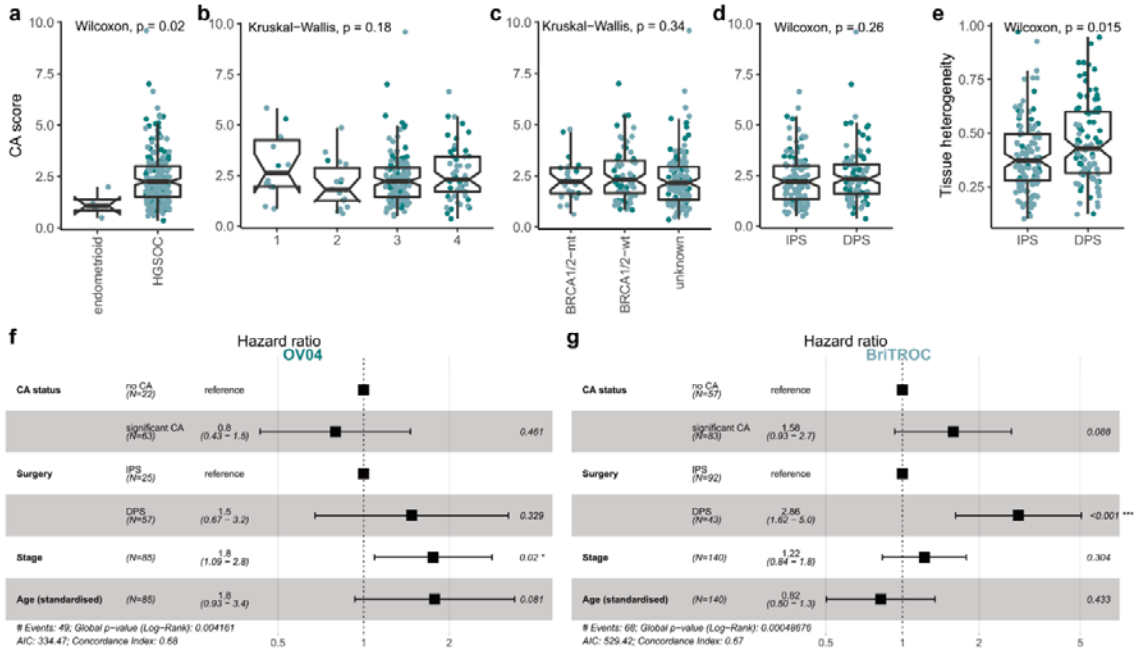
(**a**) Centrosome amplification scores across 287 HGSOC tissue samples, 24 fallopian tube, a spleen, a kidney, and 5 liver samples. Each boxplot represents CA scores across up to 50 imaging fields from individual tissue samples. Different tissue types are indicated by different colours. Dashed vertical line indicates the CA cutoff threshold of ~1.83. Samples with median CA scores > CA cutoff were considered to have significant CA.

Figure 3 – Centrosome amplification shows inter- and intra-tissue heterogeneity.



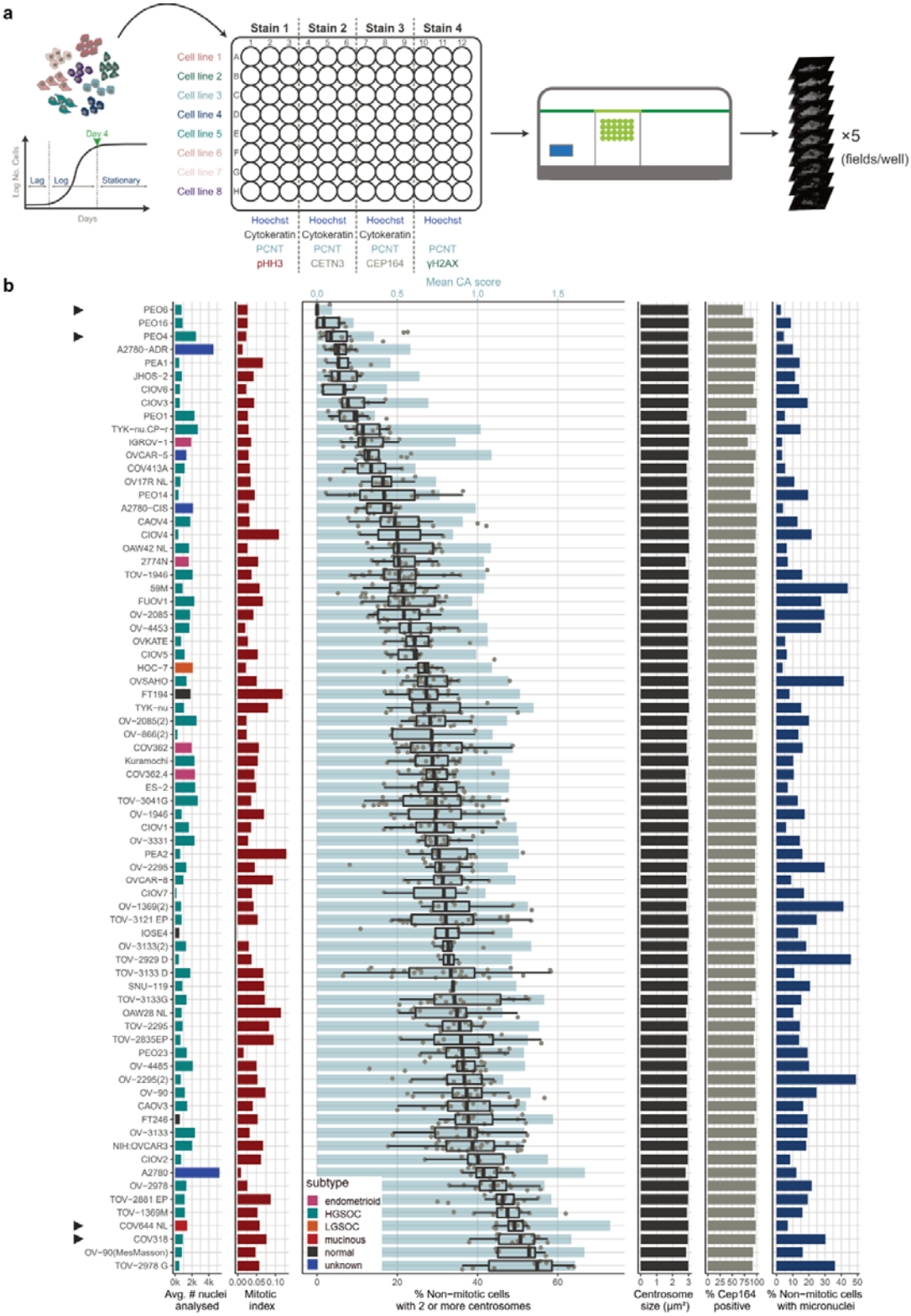
Spatial heatmap of centrosome amplification in a normal fallopian tube (**a**) and HGSOC tumour tissue (**b**). Additional examples are shown in **Supplementary Fig. 1**. CA scores are indicated by a colour gradient (purple = low, yellow = high) for each imaging field. Points represent individual nuclei detected during image analyses and were plotted in relation to their physical position (µm) on the microscope slide (x and y axis). Note that CMYK printing may obscure differences on the CA score colour scale. (**c**) Comparison of CA scores across tissues collected from individual patients in a total of 36 cases. Red stars indicate CA score medians for patient with multiple tissue samples. (**d**) Comparison of CA scores across different tissue sites. (**e**) Comparison of CA tissue heterogeneity across different tissue sites. Tissue sites are indicated by different shapes. Cohorts are indicated by teal (OV04) and light blue (BriTROC) colours. Dashed horizontal lines indicate the CA score cutoff of ~1.83.

Figure 4 – Centrosome amplification is not associated with clinical features and disease outcome in HGSOC.



Comparison of CA scores across different (**a**) histotypes, (**b**) disease stages at diagnosis, (**c**) germline *BRCA1*/*BRCA2* mutation status, and (**d**) surgery types during sample collection (IPS = immediate primary surgery; DPS = delayed primary surgery). (**e**) Comparison of CA tissue heterogeneity in IPS vs. DPS cases. Forest plots of multivariable Cox proportional hazard modelling on overall survival for (**f**) OV04 and (**g**) BriTROC patients with and without CA. Adjusted covariates included surgery type, stage and age. OV04 samples are shown in teal, BriTROC samples are shown in light blue.

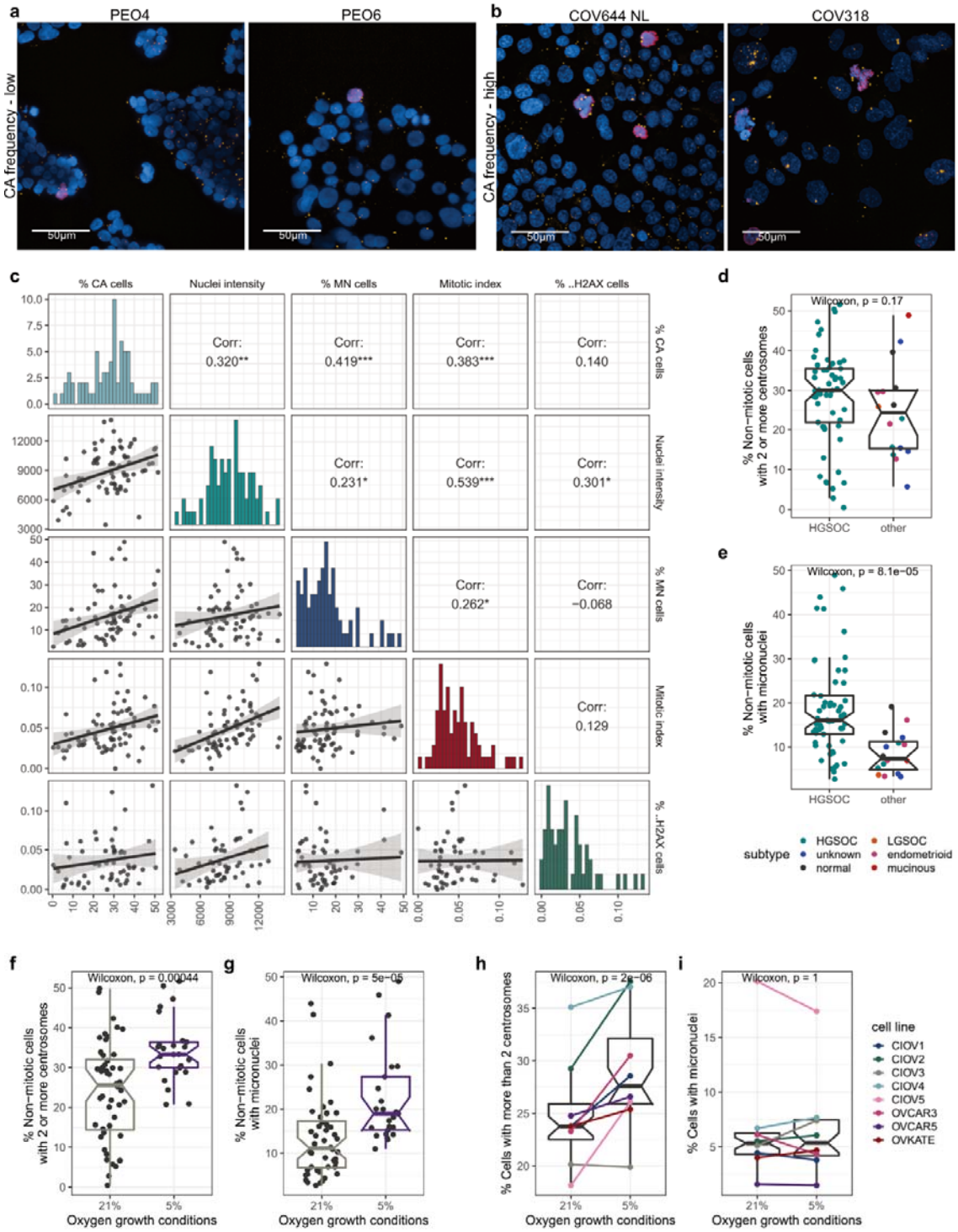
Figure 5 – In-depth imaging-based characterisation of ovarian cancer cell lines.



(**a**) Cell line characterisation workflow. Growth curves were estimated for each cell line to determine optimal seeding densities. Cells were then seeded according to estimated seeding densities into an optical, poly-L-lysine coated 96 well plate and incubated for 4 days until they reached confluence (late Log/early Stationary phase). Cells were fixed with 100% methanol and stained against indicated proteins. For each well, five independent

fields were imaged using the indicated Operetta CLS™ system (40× confocal water objective). (**b**) Immunofluorescent imaging results for 68 screened cell lines ordered by the median percentage of non-mitotic cells with 2 or more centrosomes (boxplots). Cell lines highlighted by grey arrowheads are depicted in **Fig. 6a-b**. Panel 1 shows the average number of nuclei included in each of the four staining screens. Bar plots are coloured by cell line subtype (from publicly available information (Cellosaurus) and Domcke et al. (2013)). Panel 2 indicates the mitotic index for each cell line. The fraction of non-mitotic cells with two or more centrosomes is shown in panel 3 (boxplots). Each point represents an individual imaging field. The mean (population wide) CA score (= $\frac{\sum \text{Centrosomes}}{\sum \text{Nuclei}}$) is shown by light blue bars. Panel 4 shows the centrosome size (estimated from max. intensity projection images), and panel 4 indicates the percentage of CEP164-positive centrosomes. Micronuclei (MN) frequencies, estimated as percentage of non-mitotic cells with at least one or more MN, are shown in panel 6.

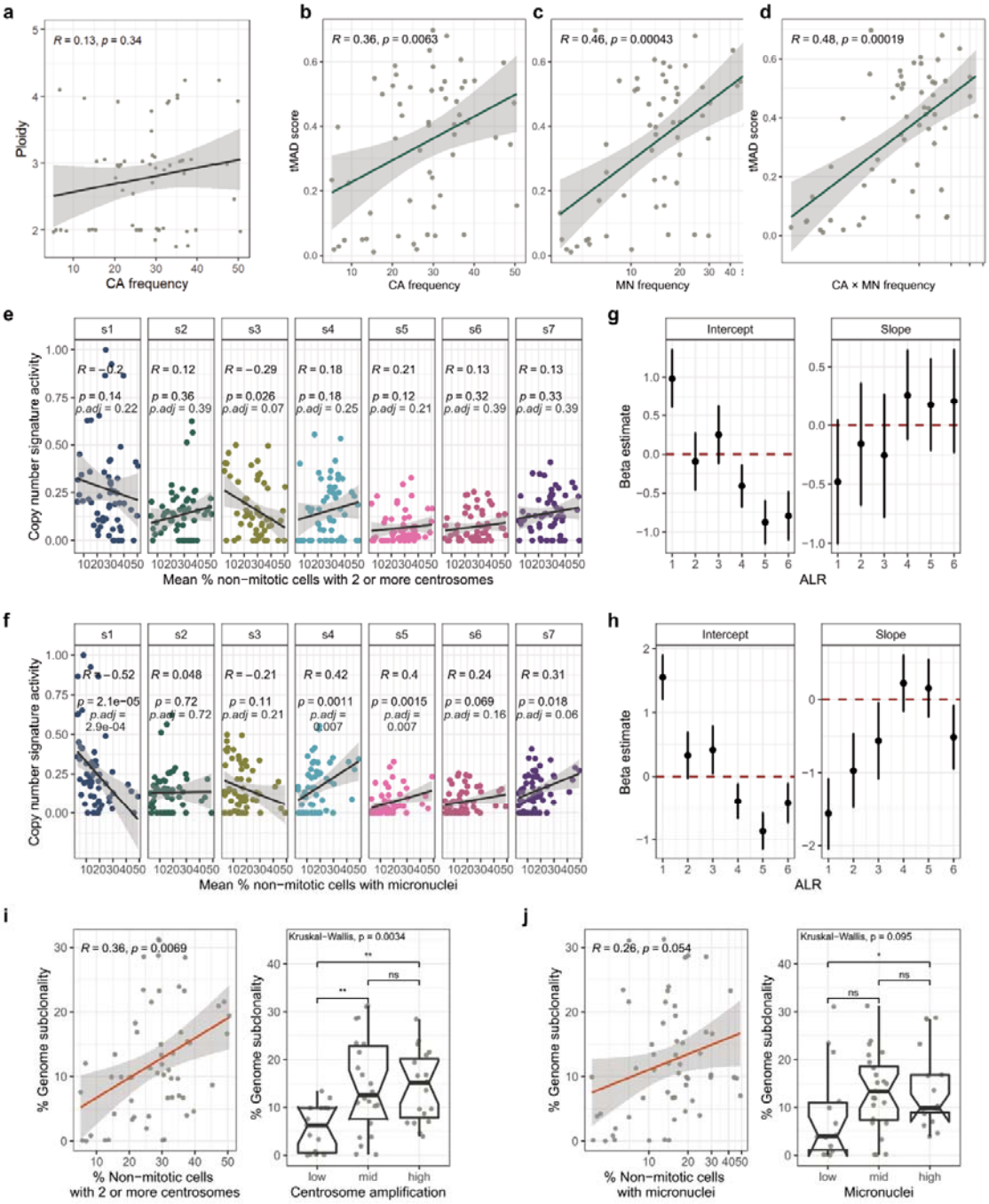
Figure 6 – Centrosome amplification is associated with increased micronuclei frequencies and can be induced by low oxygen growth conditions.



Confocal immunofluorescent staining images (zoomed in from 40× max. intensity projection images) for (**a**) CA-low and (**b**) CA-high cell lines as highlighted in **Fig. 5b**. Cells were stained for Pericentrin (yellow), pHH3 (mitotic cells; red), DNA (Hoechst; blue) and cytokeratin (not shown for easier centrosome visualisation). Scale bar = 50 µm. (**c**) Cell line characterisation correlation plot whosing Spearman's rank correlations of analysed immunofluorescent staining results. P-values: *** P < 0.001; ** P < 0.01; * P < 0.05; . P < 0.1. (**d**-**e**) Cell line subtype comparison of centrosome amplification and micronuclei frequencies, respectively. Histological subtypes

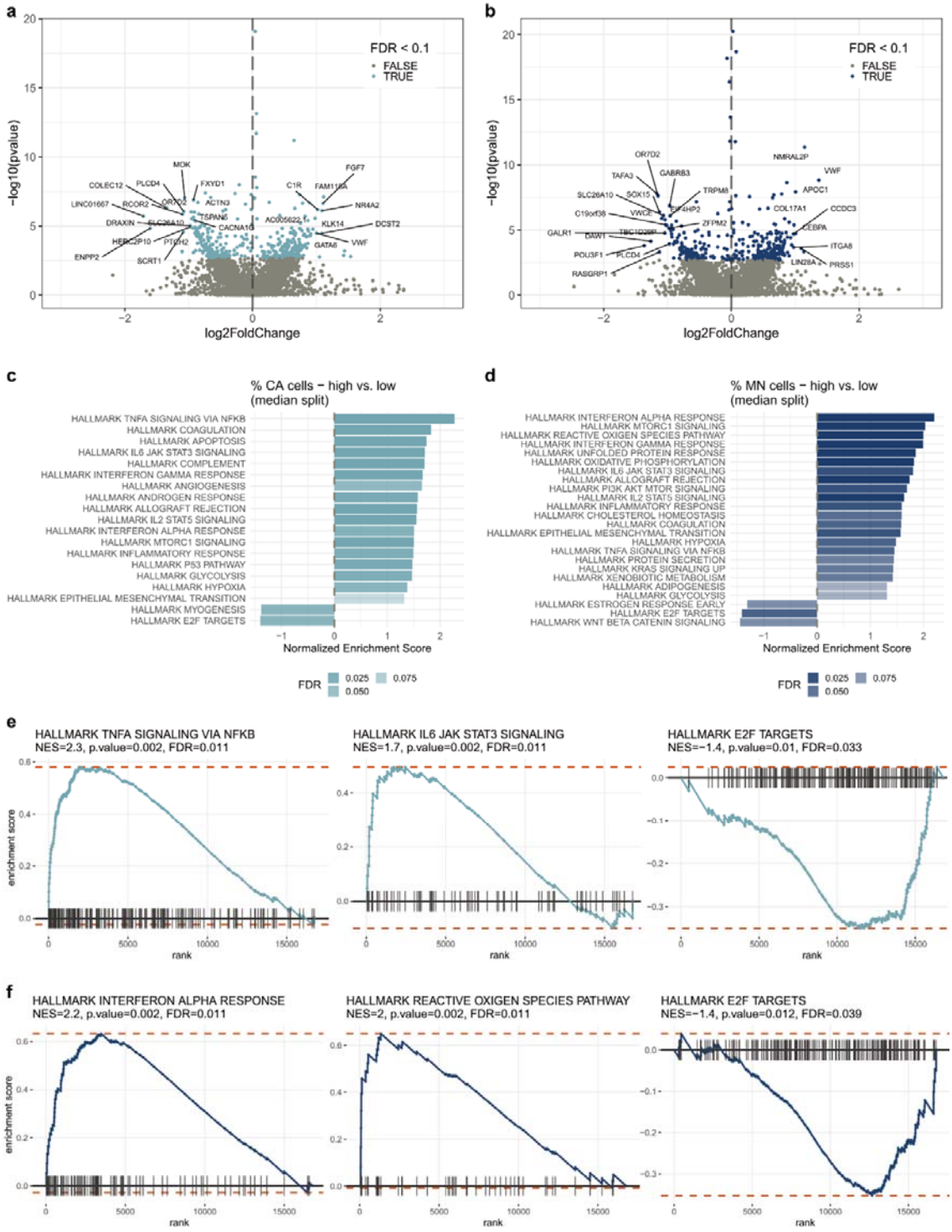
are indicated by different colours. Note that the three “normal” cell lines were grouped together with other subtypes, as these cell lines are transformed and show some cancer characteristics. (**f**-**g**) Centrosome amplification and micronuclei frequencies, respectively, in cell lines grown at 21% vs. 5% oxygen. (**h**-**i**) Cell lines, which are normally grown in normoxic conditions (21% oxygen), were transferred into 5% oxygen and centrosome amplification (**h**) and micronuclei frequencies (**i**) were estimated.

Figure 7 – Genomic characterisation of centrosome amplification and micronuclei.



(**a**) Correlation of centrosome amplification (CA) with cell line ploidy. Coefficients were estimated using Spearman's rank correlations. (**b**-**d**) Correlations of cell line tMAD scores with CA frequencies, micronuclei (MN) frequencies and CA × MN (combined) frequencies, respectively. tMAD = trimmed median absolute deviation from copy number neutrality. Note that normal cell lines were excluded from this analysis. (**e**-**f**) Correlations of copy number signature exposures with CA and MN frequencies, respectively. Different signatures are indicated by colours. Intercepts and coefficients (slopes) and their standard errors of the fixed effect model of ALR-transformed copy number signature data for CA and MN frequencies are shown in (**g**) and (**h**). (**i**-**j**) Correlations of genome subclonality estimated from ACN fits with CA and MN frequencies. All correlation coefficients were estimated using Spearman's rank correlations.

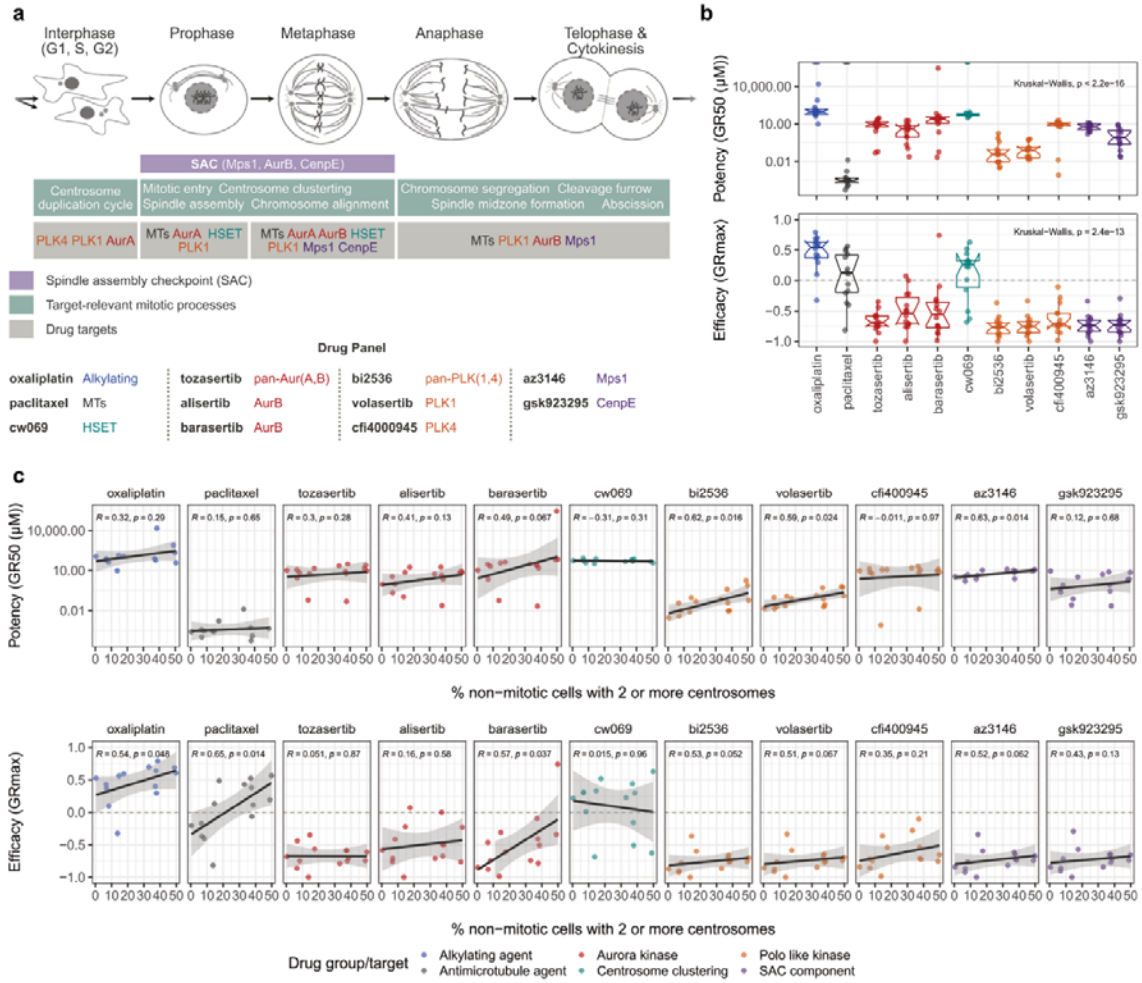
Figure 8 – Differential gene expression in cells with high centrosome amplification and micronuclei frequencies.



Differential gene expression volcano plots of all genes significantly regulated by (**a**) CA and (**b**) MN frequencies (high vs. low; median split). FDR < 0.1 highlighted in lightblue/blue. The top 25 most significantly up- or downregulated genes are labelled. (**c**-**d**) Gene set enrichment analysis (GSEA) of centrosome and micronuclei high vs. low cell lines, showing hallmark pathways for which p < 0.05 and FDR < 0.1. Examples of enrichment score plots for centrosome and micronuclei GSEA pathway results are shown in (**e**) and (**f**), respectively. Light

blue/blue lines illustrate the running sum for each gene set shown. Maximum and minimum enrichment score (ES) are indicated by orange dotted lines. Leading-edge gene subset is indicated by vertical black lines.

Figure 9 – Cell lines with high centrosome amplification show decreased drug sensitivities.



(**a**) Ovarian cancer cell line drug-screen overview showing drugs used in our drug sensitivity assay and their respective targets within the cell cycle. (**b**) Variation in potency and efficacy of selected drugs a cross all tested ovarian cancer cell lines (n=14). (**c**) Correlation of centrosome amplification frequencies and drug potency (GR50; top panel) and drug efficacy (GRmax; bottom panel). Spearman's rank correlations are shown for each drug. Drug agents are colour-coded according to their respective drug target groups. Note that higher GR values correspond to lower potency or efficacy. The signs of GRmax values indicated drug response: GRmax > 0 indicates partial growth inhibition; GRmax = 0 indicates complete cytostasis; GRmax < 0 indicates cell death (cytotoxicity).