Title

Integrated Genomic, Epigenomic, and Expression Analyses of Ovarian Cancer Cell Lines.

Journal

Cell Reports, 25(9)

Authors

Papp, Eniko
Hallberg, Dorothy
Konecny, Gottfried
et al.

Publication Date

2018-11-27

DOI

10.1016/j.celrep.2018.10.096

Peer reviewed

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2019 April 24.

Published in final edited form as:

Cell Rep. 2018 November 27; 25(9): 2617–2633. doi:10.1016/j.celrep.2018.10.096.

Integrated Genomic, Epigenomic, and Expression Analyses of Ovarian Cancer Cell Lines

Eniko Papp[1,5], **Dorothy Hallberg**[1,5], **Gottfried E. Konecny**[2], **Daniel C. Bruhm**[1], **Vilmos Adleff**[1], **Michaël Noë**[1], **Ioannis Kagiampakis**[1], **Doreen Palsgrove**[1], **Dylan Conklin**[2], **Yasuto Kinose**[3], **James R. White**[1], **Michael F. Press**[4], **Ronny Drapkin**[3], **Hariharan Easwaran**[1], **Stephen B. Baylin**[1], **Dennis Slamon**[2], **Victor E. Velculescu**[1,*], and **Robert B. Scharpf**[1,6,*]

[1]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[2]Division of Hematology and Oncology, Jonsson Comprehensive Cancer Center, David Geffen School of Medicine at UCLA, Los Angeles, CA 90024, USA

[3]Department of Obstetrics and Gynecology Penn Ovarian Cancer Research Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Department of Pathology, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA

[5]These authors contributed equally

[6]Lead Contact

SUMMARY

To improve our understanding of ovarian cancer, we performed genome-wide analyses of 45 ovarian cancer cell lines. Given the challenges of genomic analyses of tumors without matched normal samples, we developed approaches for detection of somatic sequence and structural changes and integrated these with epigenetic and expression alterations. Alterations not previously implicated in ovarian cancer included amplification or overexpression of *ASXL1* and *H3F3B,* deletion or underexpression of *CDC73* and TGF-beta receptor pathway members, and rearrangements of *YAP1-MAML2* and *IKZF2-ERBB4.* Dose-response analyses to targeted therapies revealed unique molecular dependencies, including increased sensitivity of tumors with *PIK3CA* and *PPP2R1A* alterations to PI3K inhibitor GNE-493, *MYC* amplifications to PARP inhibitor BMN673, and *SMAD3/4* alterations to MEK inhibitor MEK162. Genome-wide

*Correspondence: velculescu@jhmi.edu (V.E.V.), rscharpf@jhu.edu (R.B.S.).

AUTHOR CONTRIBUTIONS

Conceptualization, G.E.K., S.B.B., D.S., V.E.V., and R.B.S.; Methodology, E.P., D.H., G.E.K., V.A., I.K., H.E., J.R.W., S.B.B., V.E.V., and R.B.S.; Data Cu- ration, E.P., D.H., G.E.K., M.N., and D.P.; Software, D.H., D.C.B., and R.B.S.; Validation, D.H., V.A., and M.N.; Formal Analysis, E.P., D.H., D.C.B., I.K., H.E., J.R.W., and R.B.S.; Investigation, E.P., D.H., G.E.K., V.A., D.C., I.K., R.D., H.E., V.E.V., and R.B.S.; Writing, E.P., D.H., G.E.K., M.N., D.P., J.R.W., H.E., S.B.B., R.D., V.E.V., and R.B.S.; Visualization, E.P, D.H., M.N., V.E.V., and R.B.S.; Supervision, H.E., S.B.B., D.S., V.E.V., and R.B.S.; Project Administration, D.S. and V.E.V.; Resources, G.E.K., D.C., Y.K., M.F.P., R.D., D.S., and V.E.V.; Funding Acquisition, G.E.K., S.B.B., D.S., and V.E.V.

SUPPLEMENTAL INFORMATION

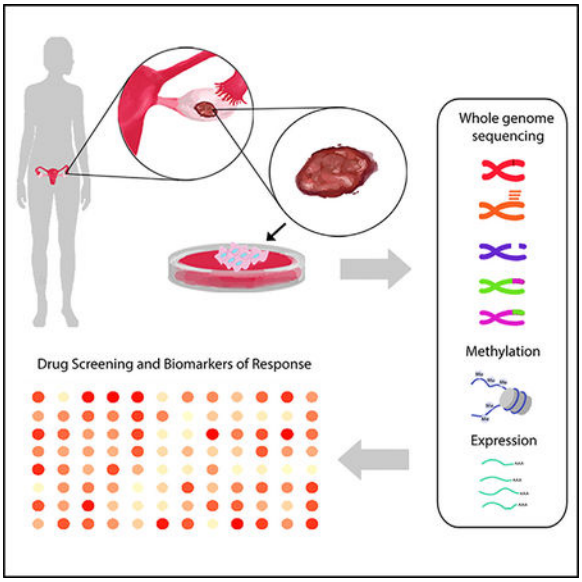
Supplemental Information includes 11 figures and 1 table can be found with this article online at https://doi.org/10.1016/j.celrep.2018.10.096.

rearrangements provided an improved measure of sensitivity to PARP inhibition. This study provides a comprehensive and broadly accessible resource of molecular information for the development of therapeutic avenues in ovarian cancer.

In Brief

The overall survival of patients with late-stage ovarian cancer is dismal. To identify therapeutic opportunities, Papp et al. integrate genomic, epigenomic, and expression analyses to provide a resource of molecular abnormalities in ovarian cancer cell lines and use these to identify tumors sensitive to PARP, MEK, and PI3K inhibitors.

Graphical Abstract



INTRODUCTION

Despite significant advances in therapies for other solid tumor malignancies, the overall survival of patients with late-stage ovarian cancer has remained dismal with few new options for treatment. The standard therapy involves debulking surgery followed by chemotherapy. Part of the reason for the lack of novel therapies for ovarian cancer has been an inadequate understanding of the underlying molecular characteristics of this disease, especially in the context of cancer cell models that can facilitate the development of various cancer treatments.

Recent studies have highlighted the genomic complexity and heterogeneity of ovarian cancer. These have included a catalog of sequence mutations, focal changes in DNA copy number, gene expression, and methylation alterations in high-grade serous ovarian cancer (Cancer Genome Atlas Research Network, 2011), as well as whole-exome analyses of ovarian clear-cell carcinoma and low-grade serous carcinoma (Jones et al., 2012, 2015). Genome-wide sequence analyses of high-grade serous ovarian cancer identified drivers associated with primary and acquired resistance to chemotherapy (Patch et al., 2015; Labidi-

Galy et al., 2017). More recently, a catalog of proteomic alterations in high-grade serous The Cancer Genome Atlas (TCGA) samples has been integrated with structural alterations and correlated with clinical outcomes (Zhang et al., 2016). Hypothesis-generating pharmacogenomic studies involving cancer cell lines, some of which were ovarian, have revealed genetic- and expression-based alterations associated with resistance or sensitivity to a panel of drugs (Garnett et al., 2012; Barretina et al., 2012).Cell line studies have evaluated high-grade serous, clear-cell, and other cancers using targeted genomic and other molecular analyses (Domcke et al., 2013; Anglesio et al., 2013; Ince et al., 2015). These initial efforts were extended to demonstrate the similarity of molecular alterations in cell lines to those in corresponding tissues, to develop approaches for incorporating multiple data types to model sensitivity, and to apply these models to larger drug panels (Iorio et al., 2016). Despite these advances, a comprehensive analysis of genome-wide structural alterations, including intra- and inter-chromosomal translocations and gene fusions, and integration of these data with whole-exome sequence, epigenetic, and expression information are not available for many histological subtypes of ovarian cancer. Furthermore, the therapeutic response of these ovarian cancer subtypes to common targeted therapies is not well understood.

Here, we performed complementary molecular analyses of 45 ovarian cell lines of different histologies, including serous, clear-cell, endometrioid, and mucinous cancers. As these cell lines do not have matched normal tissues, we developed approaches to characterize tumor-specific alterations at sequence and structural levels, and integrated these with methylation and transcript changes to identify the compendium of alterations within genes and pathways. Using the same cell lines, we evaluated the effect of a few targeted agents of common pathways using *in vitro* cell survival assays. Our analyses identified molecular alterations not previously reported in ovarian cancer, delineated genes modulated by genetic and epigenetic changes, and highlighted specific sequence, structural, and epigenetic alterations associated with sensitivity and resistance to common pathway inhibitors.

RESULTS

Overall Approach

We aimed to assemble a collection of ovarian cancer cell lines that would be representative of the different histological subtypes. These encompassed both publicly available as well as newly generated cell lines, ultimately comprising 19 serous, 9 clear-cell, 3 mucinous, 2 undifferentiated, 2 endometrioid, 1 mixed, and 9 unclassified subtypes (Table S1a). Information related to the original description and classification of these cell lines is indicated in Table S1a, and the origin of the lines was confirmed using unique short tandem repeat (STR) analyses (Table S1b). To identify sequence and structural changes in these ovarian cancer cell lines, we performed next-generation whole-genome analyses at an average coverage of 32x and 116.6 Gb per sample (Table S1c). As matched normal DNA was not available for these samples, we also sequenced a set of 18 unmatched DNA samples from normal blood or lymphoblastoid cell lines from individuals of various ethnicities. We developed approaches to focus on likely tumor-specific sequence and genomewide structural changes, including amplifications, deletions, and rearrangements. In parallel, genome-wide

methylation analyses were performed and integrated with genomic and expression data in order to obtain a comprehensive molecular profile of these samples (Figure 1).

Sequence Analyses

A high-sensitivity analysis of sequence alterations, including single-base substitutions and small insertions and deletions, was performed for the exomes of these samples. Given the challenges of characterizing tumor-specific (somatic) changes in tumor samples without matched normal tissue (Jones et al., 2015), we developed stringent bioinformatic approaches to determine likely somatic mutations. Removal of common germline variants resulted in an average of 928 alterations per cell line exome, comprising 41,768 changes that included rare germline and somatic alterations. Six cell lines (two clear cell and one each of endometrioid, serous, unclassified, and mixed lineage) were hypermutated, having alterations in mismatch repair (MMR) genes *MLH1, MSH2, MSH6,* or *PMS2,* and six times as many sequence changes compared to those tumors that were MMR proficient (Table S1d). To focus on likely somatic alterations involved in tumorigenesis, we analyzed the sequence alterations in each cell line and identified changes that have been previously detected in the coding genomes of other cancer patients (Forbes et al., 2010). We also identified nonsense orframeshift inactivating mutations in a panel of tumor suppressor genes (Table S1e). Through these analyses, we discovered 659 putative driver somatic mutations across 45 ovarian cell lines (Table S1f).

The most frequently mutated gene was the *TP53* tumor suppressor gene (altered in 25 non-hypermutated and 3 hypermu- tated tumors). Excluding hypermutated samples, other genes frequently mutated included *ARID1A* (14 cancer cell lines), *PIK3CA* (6), *SMAD4* (4), *KRAS* (4), *APC* (3), *CREBBP* (3), and *PPP2R1A* (3). Mutations were predominantly CpG transitions C→T or G→A (48%) followed by non-CpG transitions A↔G or C→T (25%) (Figure S1A). These observations are consistent with mutations in *TP53* in a high fraction of serous ovarian cancers (Cancer Genome Atlas Research Network, 2011), and *PIK3CA, ARID1A,* and *PPP2R1A* in clear-cell tumors (Jones et al., 2012). Analysis of mutation signatures aggregated by ovarian cancer subtypes revealed that serous, mucinous, and undifferentiated tumor cell lines had an age-related signature previously reported in ovarian adenocarcinomas (Alexandrov et al., 2013). Clear-cell and serous ovarian cancers also had a profile consistent with a MMR-associated mutation signature, likely due to the subset of tumors with MMR defects in the cell lines (Patch et al., 2015) (Figure S1B). Overall, both the compendium of mutated genes as well as mutation-associated signatures were representative of previous ovarian cancer genome analyses (Table S1f).

Structural Variant Analyses

Given the importance of structural alterations in the development of ovarian cancer (Cancer Genome Atlas Research Network, 2011; Patch et al., 2015; Zhang et al., 2016; Labidi-Galy et al., 2017), we used whole-genome sequence data to characterize copy number changes as well as rearrangements that may affect key driver genes. We first considered existing approaches for whole-genome analyses, including DELLY and LUMPY, but these typically use matched normal sequences to accurately identify tumor-specific rearrangements (Rausch et al., 2012; Layer et al., 2014). Given the multitude of tumor cell lines and other cancer

specimens where matched normal DNA is not available, we developed a framework for structural variant detection called Trellis that could be used with tumor genome sequence data directly. Additionally, because many structural changes are linked genomically (i.e., an amplified gene has both copy number changes and rearrangements that can be located in multiple locations of the genome), we aimed to connect the multiple changes that were related to individual genetic targets. The features of this approach include (1) detection of tumor-only structural changes through removal of germline and artifactual changes, (2) distinction of focal homozygous deletions and amplifications from larger structural changes, (3) connection of apparently disparate copy number regions using paired sequences in the same amplicons, (4) detection of homozygous and hemizygous deletions through copy number and rearrangement data, (5) confirmation of rearrangements using a stringent local realignment to detect and remove spurious paired read and split alignments, and (6) identification of in-frame rearrangements that would likely lead to gene fusions.

To implement the Trellis approach, we excluded low complexity sequences by mappability, as well as regions of germline copy number variants (CNVs) and rearrangements detected in the genomes of 18 samples derived from normal blood cells. We divided the remaining 2.7 Gb of the genome into 1-kb bins and examined areas of increased read density (>2.75-fold) to identify copy number gains, and regions of decreased read density (<0.6-fold) to detect hemizygous or homozygous deletions greater than 2 kb using approaches similar to digital karyotyping (Wang et al., 2002; Leary et al., 2008). We identified rearrangements from atypical orientation or spacing of paired reads as well as split read alignments (see Experimental Procedures:Implementation of Trellis).

To evaluate the specificity of our approach in a set of nontumor samples where we expected very few somatic structural changes, we used a leave-one-out cross-validation analysis among the 10 unmatched normal blood samples. Using Trellis, these analyses identified no focal high copy gains. On average, we identified 5 hemizygous deletions (interquartile range, 2–15) and 1 homozygous deletion (interquartile range, 0–8) in the normal samples (Figure 2). Likewise, the average number of rearrangements observed per sample was 3 (interquartile range, 0–6). These observations are consistent with previous descriptions of germline structural changes in normal DNA, in particular in lymphoblastoid DNA (Shirley et al., 2012), and suggest a high specificity of our approach for detection of bona fide somatic alterations (mean specificity, 0.97).

By contrast, analysis of normal samples with DELLY or LUMPY detected hundreds to thousands of structural changes in each normal DNA sample (Figure 2). With DELLY and LUMPY, the average numbers of focal high-quality copy number alterations were 13 and 21, respectively. The average numbers of intra- and inter-chromosomal rearrangements identified by DELLY were 297 and 433, respectively, and for LUMPY these were higher, at 511 and 2203, respectively. The number of alterations observed by DELLY using low-stringency settings was higher yet (Figure 2). False positives for copy number changes appeared to largely be due to inclusion of single-copy gains and losses, with neither DELLY nor LUMPY distinguishing hemizygous from homozygous losses or single-copy gains from high-copy amplifications. The source of the rearrangement false positives appeared to be

largely the result of mapping artifacts due to low sequence complexity in putative rearrangements (Figure S2).

To assess the sensitivity of this approach, we sequenced 16 cell lines using high-coverage next-generation sequencing of 111 genes comprising 585,216 bp. Computing the fold change of read depth at these targeted regions, we found four high-copy amplifications with fold change ≥6, nine low- copy amplifications with fold change ≥3 and <6, and nine homozygous deletions. Trellis detected all four high-copy amplifications, including amplifications of *AKT2, CCNE1,* and *KRAS.* All nine regions identified as low-copy amplifications by targeted sequencing were also determined to be low-copy amplifications by Trellis, corroborating quantitative and qualitative characteristics of the amplifications. Similarly, all nine deletions discovered by targeted sequencing, comprising *CDKN2A* (8) and *NF1* (1), were also characterized as homozygous deletions by Trellis. Overall, these analyses established that the Trellis approach had both high specificity and sensitivity for detection of structural alterations that are currently not possible with tumor-only samples using existing approaches.

Linked Amplicons—We focused our analysis of amplifications to regions smaller than 3 Mb that were present at >2.75-fold compared to the modal genome copy number, as such alterations have historically been linked to amplified driver genes (Leary et al., 2008). An analysis of the 45 ovarian cancer samples identified 538 focal amplicons, or an average of 12 amplicons per tumor (Table S1g). As multiple amplicons within the same tumor may be derived from an amplification of a single target gene localized to different chromosomal regions (Leary et al., 2008; Campbell et al., 2008; Greenman et al., 2016), we examined the possibility that amplicons may be linked. Using our paired read whole-genome analyses, we found that reads at the edges of many amplicons were linked with aberrant spacing and/or orientation with respect to the reference genome. In order to identify links between apparently distant amplicons, we visualized these as undirected graphs where the nodes were amplicons and edges between amplicons were defined by multiple paired reads aligned to both genomic locations (e.g., Figures 3A and 3B). Our analyses discovered 57 amplicon groups from the 538 amplicons across the ovarian tumor cell lines. Among tumors with at least one am- plicon, the median number of amplicon groups was 2 and the median number of amplicons within an amplicon group was 4 (interquartile range, 2–9). The majority of cell lines (15/28) with an amplicon group contained known driver genes. As an example, cell line ES-2 had 41 apparent amplicons, but through this approach we determined that 38 of the amplicons were linked to a single group that contained the *CCND1* driver gene (Figure S4). Both the copy number ($t = 3.3$, $p = 0.003$) and number of connections between amplicons ($t = 5.3$, $p < 0.001$) was significantly higher for amplicon groups containing known drivers compared to amplicon groups without known drivers (Figure 3C).

Driver genes that were amplified in two or more cell lines as part of amplicon groups that have previously been observed in ovarian cancer included well-known oncogenes such as *MYC* (4), *ERBB2* (2), *CCND1* (2), *CCNE1* (2), *FGFR4* (2), and *KRAS* (2). Interestingly, we identified amplifications of cancer driver genes that have not been previously appreciated in ovarian cancer, including epigenetic regulator *ASXL1* (2), H3 histone family member *H3F3B* (2), NOTCH family receptor *NOTCH4* (1), repair and recombination paralog

RAD51C (1), and ubiquitin ligase *RNF43* (1). *ASXL1* amplification and overexpression have been previously identified in cervical cancer (Katoh, 2015), and over-expression of *H3F3B* has been reported in several tumor types but not ovarian cancer (Ayoubi et al., 2017). Several of these genes have been observed as being part of larger structural alterations in recent TCGA high-grade serous ovarian carcinoma analyses (Cancer Genome Atlas Research Network, 2011) but have not been identified as target genes in those cases.

Overall, these analyses greatly simplified the observed amplification events and revealed that many focal amplicons would not have been associated with driver genes had they not been linked in specific amplicon groups. The observed amplicons were consistent with previously detected genes in ovarian cancer, but we also identified genes not previously implicated in this disease.

Deletions—We used a combination of stringent analyses of segmented read depth and aberrant read pair spacing to identify homozygous and hemizygous deletions. As deletions may occur in the germline, we removed deletions that were in or near structural alterations observed in the normal controls in order to identify those deletions that were most likely to be somatic. These analyses revealed 674 hemizygous^{+}, 41 overlapping hemizygous^{+}, 286 homozygous, and 263 homozygous^{+} deletions, where "+" denotes evidence for deletion supported by rearranged read pairs in addition to read depth (Figures 3D; Table S1h). Deletion breakpoints with rearranged read pairs were more precise (typically within 100 bp), while deletions without rearranged read pairs had a resolution of 1–5 kb. We included homozygous deletions from segmentation analyses even if these were without rearranged read pairs as these could have been missed in read pair analyses due to the limited mappability at one or both deletion breakpoints. The median number of homozygous and hemizygous deletions per tumor was 10.5 (interquartile range, 8–16) and 11.0 (interquartile range, 6–18), respectively. Genes that were recurrently deleted included cell cycle regulators *CDKN2A* (9) and *CDKN2B* (8), tyrosine kinase receptor *ERBB4* (5), neuro- fibromin genes *NF1* (3) and *NF2* (3), transcriptional regulator *CDC73* (2), polycomb-group repressor *EZH2* (2), and serine/threonine kinase *STK11* (2) (Table S1h), of which *CDKN2A, NF1, NF2,* and *STK11* have been previously reported to be altered in high-grade serous ovarian carcinomas (Cancer Genome Atlas Research Network, 2011; Huang et al., 2012). Genes that have been implicated through somatic deletion in other tumors but that had not been previously implicated in ovarian cancer include *CDC73, ERBB4, EZH2, MLH1*, as well as TGF-beta pathway members *TGFBR2, SMAD3,* and *SMAD4,* estrogen receptor *ESR1,* cell cycle kinase *CDK6,* notch receptor *NOTCH1*, cohesin member *STAG2*, and epigenetic regulator *ATRX* (Table S1h). In a fashion similar to amplifications, several of these genes have been observed as being part of larger structural alterations in recent TCGA high-grade serous ovarian carcinoma analyses (Cancer Genome Atlas Research Network, 2011) but have not been identified as target genes in those cases or other histologic subtypes. The absence or low frequency of such alterations in previous studies may in part reflect the challenges of identifying bona fide deletions through existing approaches in primary tumors.

Other recurrent deletions occurred in genes encompassing large genomic regions (>1 Mb) that were more likely to be affected by structural alterations, including a member of the low-density lipoprotein receptor family *LRP1B* (7), fragile histidine triad involved in purine

metabolism *FHIT* (11), a member of the short-chain dehydrogenases/reductases protein family *WWOX* (15), and the deacetylase *MACROD2* (7). *FHIT* and *WWOX* occur in fragile sites and are often deleted in cancers, and some evidence suggests they encode putative tumor suppressors (Ohta et al., 1996; Zochbauer-Muller et al., 2000; Roy et al., 2011; Aldazet al., 2014). *LRP1B* deletion has been associated with chemotherapy resistance in high-grade serous ovarian cancers and is a putative tumor suppressor (Cowin et al., 2012). Because of their proximity to *CDKN2A,* the methylthioadenosine phosphorylase *MTAP* and the transcription factor *DMRT1* are commonly co-deleted with *CDKN2A* (Zhang et al., 1996), and use of compounds exploiting the loss of *MTAP* has been proposed as a potential therapeutic avenue (Marjon et al., 2016) for tumors with *CDKN2A* deletions.

Rearrangements and Fusions—We next examined structural rearrangements that were not associated with segmental copy number changes. We detected 834 inter-chromosomal and 2,277 intra-chromosomal rearrangements (Table S1i). The median per sample of inter- and intra-chromosomal rearrangements was 16 (interquartile range, 5–31) and 37 (interquartile range, 17–62), respectively, with many of these rearrangements involving inversions (median of 8 and 6, respectively).

Among rearrangements for which the sequence junction was within the intron or exon of a gene, we detected 128 in-frame fusions of two genes (Table S1j). Several of these in-frame fusions have not been observed in ovarian cancer but have been previously reported in other cancers. For example, *YAP1-MAML2* has been reported in nasopharyngeal carcinoma and salivary cancers (Tonon et al., 2003; Coxon et al., 2005; Valouev et al., 2014), *IKZF2-ERBB4* has been reported in T cell lymphomas (Boddicker et al., 2016), and fusions involving CCND1 were identified in a patient with leukemic mantle cell lymphoma (Gruszka-Westwood et al., 2002). Our study discovered the *YAP1-MAML2* fusion in cell line ES-2 after exon 6 of *YAP1* and before exon 2 of *MAML2*, preserving the transactivation domain of *MAML2* and its likely role in Notch signaling (Figure S6). The predicted amino acid sequence of the affected *MAML2* gene is the same as reported in nasopharyngeal carcinoma and salivary gland cancers (amino acid 172) (Tonon et al., 2003; Coxon et al., 2005; Valouev etal., 2014).

The *IKZF2-ERBB4* fusion we identified in ovarian tumor KK involves the first three exons of *IKZF2* and exons 2–27 of *ERBB4,* a member of the epidermal growth factor receptor (EGFR) family. This *IKZF2-ERBB4* junction is nearly identical to that reported by Boddicker et al. (2016) in T cell lymphoma and mucinous lung adenocarcinoma, involving the same exons of *ERBB4* and leaving the *ERBB4* kinase domain intact. Gene expression analyses indicated that the *ERBB4* transcript, including the fusion transcript, was overexpressed (Figure S7A). *ERBB4* overexpression has been associated with resistance to platinum-based therapy in ovarian serous carcinomas (Saglam et al., 2017), suggesting a potentially important role for this translocation event for therapeutic selection. In ovarian tumor ES-2, *CCND1* was amplified and also participated in a fusion where the promoter of *SHANK2* was linked to the coding region of *CCND1* (Figure S7B). An amplification and fusion involving *CCND1* has been previously identified in a patient with leukemic mantle cell lymphoma (Gruszka-Westwood et al., 2002). Additional gene fusions not previously observed in ovarian cancer involved the negative regulator of the RAS pathway *NF1,* the

tumor suppressor regulating mTORC1 signaling *TSC2,* and the member of the F-box protein family *FBXW7.* The fusion of *NF1 (NF1-MYO1D)* occurred after the first exon of this gene and would be expected to disrupt its function, consistent with its tumor-suppressive role Cancer Genome Atlas Research Network (2011). Similarly, the fusion of *MLST8-TSC2* would be expected to result in a TSC2 protein lacking the first 373 aa, disrupting the key region of interaction with TSC1 (Guertin and Sabatini, 2005). As detailed below, the fusion of full-length *FBXW7* to the promoter of *FAM160A1* was also likely deleterious due to decreased expression under the new promoter (Figure S7C). For each of the nine predicted fusions involving at least one gene previously identified in other cancer fusions, we independently validated the novel sequence junction using PCR and Sanger sequencing and a recently developed droplet digital PCR approach (Cumbo et al., 2018) (Figures S8 and S9).

Epigenomic and Expression Analyses

We next examined genome-wide methylation profiles in order to evaluate the role of epigenetic alterations in these ovarian cancer cell lines. Analyses of over 850,000 methylation sites were performed using Infinium MethylationEPIC arrays. Methylation levels were evaluated at individual CpG sites within gene promoter regions (1,500 bp upstream of the transcription start site) or within individual genes. We compared methylation levels in the ovarian cell lines to methylation levels in the normal lymphoblastoid cells, as well as to 8 TCGA normal fallopian tissue and 533 TCGA ovarian cancers. Among the 18,619 CpG probes shared by the Infinium HumanMethylation27 BeadChip array (27,578 probes) and the MethylationEPIC array, we estimated the proportion of methylated CpG sites as the fraction of CpG probes with $\beta > 0.3$. Consistent with previous studies (Smiraglia et al., 2001; Paz et al., 2003; Varley et al., 2013), we found that the overall proportion of methylated CpG sites in the lymphoblastoid (median, 0.35) and ovarian cell lines (median, 0.41) was higher than the proportion in fallopian tissues (median, 0.30) and ovarian cancers (median, 0.29) (Figure 4A). Previous studies have suggested that differences in methylation between cell cultures and primary tumors often occur at genes involved in cell cycle regulation (Varley etal., 2013), and may also result from contamination of normal cells in the primary tissues. To examine methylation profiles of the cell lines at individual CpG sites in the broader context of ovarian cancer methylation profiles, we identified 96 genes that were differentially methylated between normal fallopian tissue and 100 randomly sampled TCGA ovarian tumors (Figure 4B). While we excluded both the lymphoblastoid cell lines and the ovarian cancer cell lines from the probe selection procedure, the normal lymphoblastoid cell lines were more highly correlated to the normal fallopian tissues while the ovarian cancer cell lines were more correlated to the TCGA ovarian cancers. Taken together, these analyses indicate that the ovarian cell lines retain epigenetic profiles of genes commonly methylated in ovarian cancer and that the methylation of these genes is unlikely to be related to growth in culture.

We integrated our genomic and epigenetic analyses with expression data previously obtained for these cell lines through the Agilent 44K array (Konecny et al., 2011). We assessed whether specific genes affected by deletions or other structural changes in some tumors may be silenced through methylation and low expression in others. Among genes that were

methylated or deleted, expression analyses revealed lower expression for many of these genes. Cell lines RMG-I and IGROV-1 both had hemizygous deletion and loss of expression of *CDC73.* Of the 13 drivers homozygously deleted in at least one tumor, five genes, including *CDKN2A* and *ESR1,* displayed loss of expression and concomitant promoter methylation in additional ovarian cancers (Figures 4C and S10). When promoter methylation and underexpression were considered, the fraction of tumors with alterations in *CDKN2A* more than doubled from 23% to 55%, highlighting the multiple mechanisms by which *CDKN2A* function can be compromised. Similarly, *MLH1* was mutated in a single case, but was mutated and/or underexpressed in an additional seven cancers. For *ESR1,* the inactivating methylation is thought to be associated with age and has been previously observed in both ovarian cancers and ovarian cancer cell lines (Imura et al., 2006; Wiley et al., 2006). Lower expression also resulted from abnormal fusion of non-adjacent promoters to the full coding sequence of target genes. In OVCAR-8, the fusion of the promoter of *FAM160A1* with the full-length *FBXW7* gene resulted in dramatically decreased expression of *FBXW7* (Figure S7C), consistent with the tumor-suppressive function of this protein and inactivating mutations observed in other tumor types (Akhoondi et al., 2007).

We also examined the possibility of increased expression for genes with structural changes. We identified 17 genes with focal amplification in one or more cancer cell lines and evidence of bimodal expression across the samples analyzed. For these genes, 20 of the 22 tumors (91%) with focal amplification also had increased expression (Figure S11). Genes associated with amplification and fusion had particularly high expression, suggesting that the combination of genetic alterations led to increased overall transcription of these genes. The amplification of *CCND1* and its fusion to *SHANK2* in sample ES-2 increased the expression of *CCND1* relative to other ovarian cell lines without the amplification and fusion (Figure S7B). The *YAP1-MAML2* fusion, which was also duplicated in the same sample, resulted in expression of *MAML2* that was higher than 85% of the other ovarian cancer cell lines (Figure S6). For driver genes that were amplified, we examined whether additional tumors may be identified with increased expression of these genes. We found increased expression of *CCNE1, ERBB2, KRAS,* and *AKT2* in eight additional cases without genomic alterations in these genes (Figures 5 and S11). These analyses indicate the importance of integrated genomic, epigenetic, and expression analyses and have resulted in an expansion of the number of tumors with alterations in key driver genes. These observations also highlight the functional consequences of genomic and epigenomic alterations in human cancer at the RNA level.

Combining sequence and structural variants with methylation and differential gene expression, we found that nearly all ovarian cancer subtypes had alterations in cell cycle, chromatin remodeling, DNA repair, RAS, Notch, PI3K, or TGFB signaling pathways (Figure 5). Alterations in the cell cycle pathway genes, including *CDKN2A,* were the most common with one or more alterations in 60%−70% of the three most represented subtypes (serous, adenocarcinoma, and clear cell). Chromatin modifications occur in 5/7 (71%) of the clear-cell subtypes but in only 2/21 (11%) of the serous samples. We see evidence of mutual exclusivity between *CDKN2A, CCNE1,* and *RB1* within the cell cycle pathway, but not mutual exclusivity between cell cycle and *KRAS* pathways, underscoring that clonal selection often involves multiple drivers regulating distinct molecular processes.

Sensitivity and Resistance to Pathway Inhibitors

To begin to understand the relationship between genomic, epigenetic, and expression alterations and response to pathway inhibitors, we developed a screening platform for evaluating cellular proliferation in the presence of candidate therapeutic agents. As an example of the analyses that can be performed and the genotype-phenotype connections that can be obtained, we measured IC_{20}, IC_{50}, and IC_{80} after 7 days of incubation for three inhibitors, GNE-493, BMN673, and MEK162, targeting PI3K, PARP, and MEK proteins, respectively (Table S1k). Aggregating the molecular information from multiple platforms to the gene level, we limited our analyses to genes that were altered in 3 or more of the 45 cell lines. Alterations that tend to be mutually exclusive between cell lines were combined, including genes in the PI3K pathway *(PIK3CA* and *PPP2R1A)* and the genes in the TGFBR pathway *(SMAD3* and *SMAD4).* As tumors with homologous recombination deficiencies (HRDs) have been known to be sensitive to PARP inhibitors, we additionally added covariates summarizing the extent of genome-wide structural alterations for the PARP inhibitor BMN673. *A priori,* we hypothesized that most alterations would not modulate response to the targeted inhibitors. Implementing a Bayesian model averaging approach to variable selection as has been previously considered for other biomarkers (Viallefont et al., 2001; Neto et al., 2014; Meisner et al., 2018), we specified a positive prior probability that the coefficient for each gene is exactly zero. Given the genes or combination of genes and structural variant summaries, we explored the space of possible single and multi-variate models for $\log IC_{50}$ by Markov chain Monte Carlo. Relevant posterior summaries available for each inhibitor include the probability that the regression coefficient is non-zero and the posterior distribution of the regression coefficients. We used this approach to focus on those features that were present in at least half of the models as these had a higher probability of being predictive for drug response (Figure 6A).

For PARP inhibition by BMN673, our analyses revealed that the number of genome-wide rearrangements and amplification of *MYC* were important predictors of drug sensitivity (Figure 6). Importantly, the two cell lines with inactivating BRCA1/2 mutations, as well as the HRD score as defined by Abkevich et al. or Swisher et al. and applied through our whole-genome analyses, and PARP1 expression showed a trend toward increased sensitivity to PARP inhibition but were not statistically significant (Figure 6). We found that amplification of *MYC* or an increase in the number of genome-wide rearrangements, including inversions and intra-chromosomal rearrangements, were significantly associated with sensitivity to this therapy, appearing in 94% of the single-variate and multi-variate models. We estimated the difference of the mean log IC_{50} between the group of tumors with alterations in these features and the group of tumors without such changes, revealing a 93% (90% confidence interval [CI], 99%−64%) and 86% (90% CI, 96%−43%) increased sensitivity to PARP inhibition for cell lines with MYC amplification and increased rearrangements, respectively (Figure 6). Although other genomic signatures and *PARP1* expression have been suggested as biomarkers for PARP sensitivity (Nik-Zainal et al., 2016), *MYC* amplification and rearrangements have not been previously identified as markers of PARP sensitivity in serous and endometrioid ovarian cancers. Taken together, these observations suggest that alterations of common drivers along with large-scale structural alterations in ovarian cancer may identify tumors with high sensitivity to this therapy.

For inhibition of the PI3K pathway by GNE-493, mutations of *PPP2R1A* or *PIK3CA* appeared in more than 75% of the models evaluated. Cancer cell lines with mutations in *PARP1* or *PPP2R1A* had a 66% increased sensitivity to GNE-493 (Figure 6). *PPP2R1A* is a subunit of protein phosphatase 2A (PP2A), a known tumor suppressor and regulator of PI3K signaling through AKT inhibition (Basu, 2011). Somatic mutations in *PIK3CA* and *PPP2R1* have been previously reported in ovarian and uterine cancers (Shih et al., 2011; McConechy et al., 2011; Jones et al., 2015; Labidi-Galy et al., 2017). Our observations suggest that PI3K inhibitors counter the loss of PI3K pathway regulation from inactivating mutations of *PPP2R1A* and activating mutations of *PIK3CA.*

For the MEK pathway, mutations or deletions in *SMAD3* or *SMAD4* were predictive of IC_{50} levels in response to the inhibitor MEK-162. These were selected in more than 85% of the models and resulted in an increased sensitivity of 89% to this therapy. SMAD3 and SMAD4 form a complex that activates transcription of TGF-beta-regulated genes important in cellular growth control (Zhang et al., 2004). Our results are consistent with previous observations showing that loss of *SMAD4* can lead to activation of Smad-independent MEK/ERK pathway signaling and that inhibition of this pathway with MEK inhibitors can reverse tumorigenic effects (Ai et al., 2013).

DISCUSSION

These analyses provide the most comprehensive molecular analyses of ovarian cancer cell lines to date. Through integration of sequence, copy number, rearrangement, methylation, and expression analyses, we were able to detect frequently altered driver genes and pathways. These efforts revealed that ovarian cancer cell lines were largely representative of primary ovarian cancers and identified a variety of alterations not previously appreciated in these tumors.

One of the challenges of genomic analyses of tumor cell lines has been the lack of matched normal samples for determination of whether alterations are truly tumor-specific. Through the development of Trellis, we were able to identify high-confidence copy number and rearrangement alterations in the absence of matched normal DNA. Additionally, the grouping of structural changes allowed linking of apparently disparate events that would normally have been considered separate copy number alterations. These efforts provide a method for detecting somatic structural changes in cancer cell lines and other specimens in the absence of a matched normal, without the hundreds to thousands of false-positive changes typically detected through other approaches. The method also permits identification of likely driver copy number changes by detecting the most connected regions of amplicons among the many such passenger alterations typically observed in human cancers. The identification of likely somatic structural and sequence alterations in a tumor-only setting is conservative. Cancer signature analyses, for example, currently require a matched normal sample to comprehensively identify somatic signatures across the genome, and our analyses based on a subset of somatic alterations may not reflect overall somatic mutation signatures.

These efforts revealed alterations in ovarian cancer that may provide insights into the biology of this disease. The integration of methylation and expression changes together with

sequence and structural changes identified a larger fraction of tumors that were altered in specific genes and pathways than whole-genome sequencing alone. For example, *CDKN2A* was altered in over half of the serous, clear-cell, and mucinous cancer cell lines analyzed, revealing a higher fraction than previously thought to have alterations in this gene. Additionally, changes of a variety of genes, including *ASXL1*, *H3F3B,* and *CDC73,* as well as fusions of *YAP1-MAML2, IKFZ2-ERBB4,* and those disrupting *FBXW7, NF1*, and *CCND1* were detected in these tumors, providing pathways of dysregulation within ovarian cancer.

The integrative molecular analyses of ovarian cancer cell lines provide opportunities for characterizing differences in susceptibility to targeted therapies. As an example, ovarian tumors with sequence alterations in *PIK3CA* and *PPP2R1A* had enhanced sensitivity to the GNE-493 inhibitor to achieve the same therapeutic response as tumors without these alterations. Mutations in *PIK3CA* and *PPP2R1A* are common in ovarian clear-cell cancers, suggesting that PI3K inhibitors may be an effective therapeutic strategy for this subtype.

Cancers with HRD are more prone to genomic errors resulting in loss or duplication of chromosomal regions and chromosomal instability (Wang et al., 2006). Drugs that inhibit PARP1 cause multiple double strand breaks, and in tumors with HRD such DNA damage cannot be efficiently repaired, leading to cell death. We found that serous and endometrioid tumors with *MYC* amplification were highly sensitive to PARP inhibitors. PARP inhibitors have been previously shown to lead to mitotic catastrophe in neuroblastoma with amplification of the related gene *MYCN* on chromosome 2p (Colicchia et al., 2017), and *MYC* amplification on chromosome 8q has been shown to suppress BIN1, thereby increasing the activity of PARP1 (Pyndiah et al., 2011). Additionally, *MYCN* amplified neuroblastoma cell lines were more sensitive to the PARP inhibitor BYK204165 compared to cell lines without such alterations (Hallett et al., 2016). In our study, *MYC* was the central driver in multiple ovarian cancers cell lines with highly linked amplicon groups (Table S1g; Figures S3, S4, and S5), suggesting that this gene plays a major role in a subset of ovarian cancers that may now be targeted therapeutically. Interestingly, we found that genome-wide rearrangements were a more useful biomarker of sensitivity to this therapy compared to the commonly used HRD scores (Abkevich et al., 2012; Frey and Pothuri, 2017; Swisher et al., 2017), perhaps because these latter measurements only identify a subset of patients that respond to PARP inhibitors, especially as shown in clinical trials that enriched for tumors with previous sensitivity to DNA-damaging agents (Mirza et al., 2016; Coleman et al., 2017). Although there is a correlation between the HRD score and the number of inversions and intra-chromosomal rearrangements (Spearman correlation coefficient = 0.5), the genome-wide rearrangements we identified may provide a more informative signature of the underlying recombination deficiency in these tumors and together with *MYC* amplification offer potentially improved opportunities for identifying responsive patients.

Although our analyses provide a comprehensive integration of molecular alterations for ovarian cancer, certain types of changes have not been evaluated. In the future, proteomic, metabolomic, and carbohydrate changes can be added to the compendium of genomic, epigenomic, and transcriptomic information for these cell lines. Additionally, further efforts will be needed to demonstrate that these observations can be translated broadly to ovarian

cancer patients. Nevertheless, these data provide a foundation for using ovarian cell line models in screening novel therapeutic strategies. Evaluation of additional compounds using these well-characterized tumor cell lines will provide rational translational opportunities for development of new therapies in ovarian cancer.

EXPERIMENTAL PROCEDURES

Cell Lines and Growth Analyses

Cell lines were obtained from multiple sources (Table S1a) (Selby et al., 1980; Motoyama, 1981; Bast et al., 1981; Eva et al., 1982; Simon et al., 1983; Hamilton et al., 1983, 1984; Wilson, 1984; Uehara et al., 1984; Buick et al., 1985; Kidera et al., 1985; Bénard et al., 1985; Fogh, 1986; Yoshiya, 1986; Langdon et al., 1988; Nozawa et al., 1988; Hills et al., 1989; Lau et al., 1989; Sakayori et al., 1990; Schilder et al., 1990; Berchuck et al., 1992; van den Berg-Bakker et al., 1993; Hattori et al., 1994; Gorai et al., 1995; Wilson et al., 1996; Yanagi-bashi et al., 1997; Conover et al., 1998; Lounis et al., 1998; Emoto et al., 1999; Yamada et al., 1999; Provencher et al., 2000; Steinmeyer et al., 2003; Barretina et al., 2012). All samples were obtained under Institutional Review Board-approved protocols with informed consent for research use or were publicly available. Cells were plated into 24-well tissue culture plates at a density of 2×10^5 to 5×10^5 cells per well and grown in cell line-specific medium without or with increasing concentrations of their respective drugs (ranging between 0.001 and 10 μm/L).

Cells were counted on day 7 using an automated cell viability assay (Vi-CELL XR Cell Viability Analyzer; Beckman Coulter, Fullerton, CA), a video imaging system that uses an automated trypan blue exclusion protocol. Both adherent and floating viable cells were counted for treatment and control wells. Growth inhibition (GI) was calculated as a percentage of untreated controls. The log of the fractional GI was then plotted against the log of the drug concentration and the IC_{50} values were interpolated from the resulting linear regression curve fit (CalcuSyn; Biosoft, Ferguson, MO). Experiments were performed thrice in duplicate for each cell line.

STR Analyses

Genomic DNA from all cell lines was PCR amplified using a Geneprint 10 System (Promega, Madison, WI) that contains eight STR loci plus Amelogenin, a gender-determining marker. The PCR amplification was carried out in a GeneAmp PCR System 9700 following the manufacturer's protocol. The PCR products were electrophoresed on a ABI Prism 3730xl Genetic Analyzer using Internal Lane Standard 600 (Promega) for sizing. Data were analyzed using GeneMapper v. 4.0 software (Applied Biosystems, Foster City, CA). We compared our STR profiles (Johns Hopkins University [JHU]) for these cell lines to external STR profiles, including Korch et al. (2012), COSMIC (v83; https://cancer.sanger.ac.uk/cosmic), the RIKEN BioResource Center (https://www.jove.com/institutions/AS-asia/JP-japan/20278-riken-bioresource-center), or Yu et al. 2015 (Table S1b). The average percent similarity between JHU STRs and external STRs was 98%. An external STR was not available for five cell lines.

Whole-Genome Next-Generation Sequencing

DNA was extracted from cell lines using a QIAamp DNA Blood Mini QIAcube Kit (QIAGEN, Valencia, CA). In brief, the samples were incubated in proteinase K for 16 hr before DNA extraction. DNA purification was performed using the QIAamp DNA Blood Mini QIAcube kit following the manufacturer's instructions (QIAGEN, Valencia, CA). Genomic DNA from tumor samples were used for Illumina TruSeq library construction (Illumina, San Diego, CA) according to the manufacturer's instructions. Paired-end sequencing resulting in 100 bases from each end of the fragments was performed using Illumina HiSeq2000 instrumentation.

PCR and Sanger Sequencing

PCR and Sanger sequencing confirmed the presence of fusion candidates generated by Trellis. Primers were designed 200 bp on either side of the junction (Table S1m). Primerswere purchased from IDT (Coralville, IA) and purified by desalting. Primers and probes were resuspended to 100 μM in IDTE (10 mM Tris, pH 8.0; 0.1 mM EDTA) buffer and stored at − 20° C. Using the primers specific for each fusion, PCR amplification was performed in a 50-μL reaction volume in quadruplicate, consisting of 10 μL of 5X Phusion buffer, 1 μL of 10 mM dNTP, 2.5 mL of each primer at 10 mM, 0.5 μL of HotStart Phusion, and 10 ng of cell line DNA. PCR was performed using a Bio-Rad S1000 Thermal Cycler. The thermal cycle was programmed for 30 s at 98° C for initial denaturation, followed by 34 cycles of 10 s at 98°C for denaturation, 30 s at 59°C for annealing, 30 s at 72°C for extension, and 5 min at 72°C for final extension. Human mixed genomic DNA (Promega, Madison, WI) and no template were used as negative controls. PCR products were purified using Nucleospin Gel and PCR cleanup as per the manufacturer's instructions (Macherey-Nagel, Duren, Germany). PCR products were then subjected to Sanger sequencing using the Applied Biosystems 3730xl DNA Analyzer as per manufacturer's instructions (Thermo Fisher, Waltham, MA). Output was compared to original candidate fusion sequence and confirmed.

Droplet Digital PCR

The translocation primers were designed on both sides of the translocation. One of these primers was used as a common primer for both the translocation and the control. A third primer was designed to be used in combination with the common primer to amplify the wild-type sequence of one of the two translocation partners. The hydrolysis probes labeled with the FAM-fluorochrome at the 5′ end were designed to bind specifically to the translocation PCR product, while the probes labeled with the HEX-fluorochrome were designed to bind specifically to the control PCR product. As quenchers, a ZEN quencher was used as an internal quencher, while the Iowa Black FQ-quencher was added to the 3′ end of the probes. Probes were designed to have a higher melting temperature than the primers. The primers and hydrolysis probes were purchased from IDT (Coralville, IA). The primers were purified by desalting, while the hydrolysis probes were purified using high-performance liquid chromatography. Upon arrival, primers and probes were resuspended to 100 μM in IDTE (10 mM Tris, pH 8.0; 0.1 mM EDTA) buffer and stored at − 20° C. 20-μL droplet digital PCR (ddPCR) reactions were prepared, using 10 mL of 2× ddPCR SuperMix

for Probes (No dUTP) (Bio-Rad, Hercules, CA), 5–30 ng of gDNA, as quantified by the Qubit dsDNA high-sensitivity assay kit (Thermo Fisher Scientific, Waltham, MA), primers (each at a final concentration of 900 nM), probes (each at a final concentration of 250 nM), and nuclease-free water. Human mixed genomic DNA (Promega) was used as negative control. Droplets were generated using the QX200 droplet generator (Bio-Rad) by loading the DG8 cartridge (Bio-Rad)with 20 mL of the reaction mixture and 70 μL of droplet generation oil for probes (Bio-Rad). 40 mLof droplet/oil mixture was transferred to a ddPCR 96-well plate (Bio-Rad). The plate was heat-sealed with a pierceable foil heat seal (Bio-Rad). A S1000 Thermal Cycler (Bio-Rad) was used with the following amplification protocol: enzyme activation at 95°C for 10 min, followed by six cycles: denaturation at 54°C for 30 s; annealing/extension at 60°C for 1 min, followed by 34 cycles: denaturation at 58°C for 30 s; and annealing/extension at 60°C for 1 min. Following cycling, the samples were held at 98°C for 10 min. Upon completion of the PCR protocol, the plate was read using theQX200 droplet reader(Bio-Rad). Droplet counts and amplitudes were analyzed with QuantaSoft software (v1.7) (Bio-Rad).

Alignment and Identification of Sequence Alterations

Priorto mutation calling, primary processing of sequence data for samples was performed using Illumina CASAVA software (v1.8.2), including masking of adaptor sequences. Sequence reads were aligned against the hg19 human reference genome using ELAND. Candidate somatic mutations in the exome, consisting of point mutations, insertions, and deletions were identified using VariantDx (Jones et al., 2015). To detect mutations that were more likely to be somatic, we excluded mutations that appeared in >10% of the distinct reads and mutations tagged as COMMON or MULT in dbSNP VCF files. Additionally, we excluded mutations without a record in COSMIC as well as in-frame deletions (COSMIC v72). Exceptions to the COSMIC requirement were mutations that predicted truncations in relevant pathways or tumor suppressor genes (Table S1e). SNPs flagged as clinically associated or reported in COSMIC were not excluded regardless of heterozygosity or percentage of distinct reads. All candidate somatic mutations were confirmed by visual inspection. Samples with more than 2,000 alterations after dbSNP filtering were considered hypermutators. Mutational signatures were based on the fraction of mutations in each of the 96 trinucleotide contexts (Alexandrov et al., 2013). The contribution ofeach signature to each tumorsample was estimated using the deconstructSigs R package (Table S1n for R package versions).

Implementation of DELLY and LUMPY

Identifying probable somatic structural variants in tumor-only experimental designs is a major challenge. False positives arise from germline variants incorrectly reported as somatic and spurious alignments misinterpreted as biological variation. We considered two established tools, DELLY and LUMPY, for detection of structural variants (Rausch et al., 2012; Layer et al., 2014). Reads were aligned to the hg19 reference genome using BWA-MEM (version 0.7.10) (Li and Durbin, 2009) as recommended by these methods. DELLY (version 0.7.7) and LUMPY (version 0.2.13) were implemented using default parameters.

A simple leave-one-outcross-validation experiment was implemented using 10 lymphoblastoid controls to evaluate the specificity of these methods for identifying somatic structural variants in a tumor-only experimental design. Specifically, we treated the held out sample as a *tumor* and identified germline structural alterations in the training set. Excluding structural variants identified in the training set, we considered any alteration identified in the held-out sample as a false positive.

Implementation of Trellis

Germline Filters—Using 10 lymphoblastoid cell lines and 8 normal ovarian samples, we developed sequence and germline filters for the hg19 reference genome to flag regions prone to alignment artifacts and/or germline structural variation. Sequence filters for the hg19 reference genome that were masked prior to copy number analyses comprised 326.4 Mb of the genome and included non-overlapping 1-kb genomic intervals (bins) with average mappability less than 0.75 or GC percentage less than 10%, as well as the gaps track from the University of California, Santa Cruz (UCSC) genome browser that includes heterochromatin, centromeric, and subtelomeric regions (Fujita et al., 2011). After removing these sequence filters as well as chrY (all cell lines were derived from women), we normalized the read depth for the remaining 2,680,222 bins. For each bin, we computed the GC-adjusted, $\log_2$-transformed count of aligned reads. GC normalization was implemented using a loess smoother with span 1/3 fitted to a scatterplot of the bin-level GC and log2 count. We denote the GC-adjusted $\log_2$ ratios (the residuals from the loess correction) by $\bar{R}$, the mean R for a genomic region by R, and the median absolute deviation of the autosomal Rs by S. Because some bins had an unusually high or low number of aligned reads in multiple controls, we defined bin i in normal control j as an outlier if $| R_i | > (3 \times S_j)$. Bins identified as an outlier in two or more normal controls were flagged. These analyses flagged 55,764 genomic regions totaling 75.9 Mb of sequence. To identify somatic copy number alterations, we segmented the Rs using circular binary segmentation implemented in the R package DNAcopy with settings alpha = 0.001, undo.splits = ‘sdundo’, and undo.SD = 2 (Olshen et al., 2004; Venkatraman and Olshen, 2007). To exclude regions that were either copy number altered in the lympho-blastoid cell lines as well as segments that span difficult regions to genotype, we flagged segments having $|\bar{R}| > 1$. We flagged a total of 919 segments (46.8 Mb) across the 18 normal controls.

To characterize copy neutral rearrangements including inversions and translocations in the normal controls, we extracted all read pairs from the BAM file that were *improperly* paired and for which the intra-mate distance between paired reads was at least 10 kb. We defined a cluster of improper read pairs as a genomic region where at least one base is spanned by five or more improper reads and for which the union of the aligned regions is at least 115 bp. Next, we linked these clusters by the mates of the constituent reads. Clusters that could not be linked byat least five read pairs were excluded from further analysis. For all linked clusters, we required at least 90% of the linking read pairs to support the same structural variant group (Table S11). Linked clusters for which the type of rearrangement was not consistent among the linking read pairs were excluded from further analysis. For the remaining linked clusters, we realigned all the reads supporting the link using the local aligner BLAT (Kent, 2002). A command-line version of BLAT was utilized for this step

(Standalone BLAT v. 35). Confirmation by BLAT required that the reads only align to one location with a BLAT score R 90% in the hg19 reference genome. These germline rearrangements were used to screen candidate somatic rearrangements as described in greater detail below.

Somatic Deletions—Putative focal homozygous and hemizygous deletions greater than 2 kb and less than 3 Mb in the ovarian cell lines were identified by $\bar{R} < -3$ and $\bar{R} \in (-3, -0.75]$, respectively. We excluded focal deletions in tumor samples if ≥ 75% of the region was identified as a deletion in a control sample. For each deletion, we investigated whether any improperly paired reads were aligned within 5 kb of the segmentation boundaries. When five or more rearranged read pairs were aligned near the segmentation boundaries, the distribution of the improper read pair alignments was used to further resolve the genomic coordinates ofthe deletion boundaries. Resolution ofthe deletion breakpoints using this approach depends on the intra-mate distance of the improperly paired reads. On average, the intra-mate distance in the ovarian tumors was 262 bp (5th and 95th percentiles, 183 and 353). With multiple rearranged read pairs, we expect that the resolution of the deletion breakpoints was generally less than 100 bp. As previously described, realignment by BLAT was used to confirm that the rearranged read pairs supporting the deletion mapped uniquely and with high fidelity to this region of the genome. Hemizygous and homozygous deletions supported by rearranged read pairs were indicated by hemizygous + or homozygous +, respectively. Any deletion for which the outlier bins or germline CNVs occupied 75% or more of the width were excluded. Hemizygous deletions not supported by rearranged read pairs were also excluded. All deletions were confirmed by visual inspection.

Somatic Amplifications—To identify focal amplicons and establish how these amplicons were linked in the tumor genome, we seeded a graph with high-copy focal amplicons. Specifically, putative amplifications were identified as segments with $\bar{R} > 1.46$, or a 2.75-fold increase from the mean ploidy of the cell line, and between 2 kb and 3 Mb in length. Properly paired reads were used to link seed amplicons to adjacent low-copy duplications (segments with $\bar{R} > 0.81$ or fold change of 1.75). When five or more links were established, the low-copy segments were added as nodes to the graph with an edge indicating the connection between the high- and low-copy amplicons. Similarly, we established links between the low- and high-copy amplicons that were non-adjacent with respect to the reference genome by analysis of improperly paired reads as previously described.

Somatic Copy-Neutral Intra- and Inter-chromosomal Translocations and Inversions—Candidate somatic copy-neutral rearrangements were identified as previously described in the control samples. However, rearrangements in the ovarian tumor cell lines that overlapped any rearrangement identified in the controls samples were excluded. In addition to improperly paired reads, we required at least one split read supporting the rearrangement. To identify split read alignments, we extracted all read pairs for which only one read in the pair was aligned within 5 kb of the candidate rearrangement. For all such read pairs, we re-aligned the unmapped mate using BLAT (Kent, 2002). A read aligned by BLAT to both ends of the candidate sequence junction with a combined score ≥90% constituted a split read (e.g., Figure S6).

In-Frame Gene Fusions—To report candidate gene fusions, we identified all candidate somatic rearrangements for which both ends of the novel adjacency in the tumor genome was in a coding region of the genome or a promoter of a gene defined as within 5 kb of the transcription start site. Rearrangements in which both ends resided in the same gene were excluded as these may represent alternative isoforms. For each candidate fusion, we evaluated two possible orientations of the regions joined in the tumor genome, and for each orientation we extracted the full amino acid sequence of both the 5′ and 3′ transcripts as well as the candidate amino acid sequence that would be created by the fusion. We considered the fusion to be in-frame if the amino acid sequence of the 3′ partner was a subsequence of the reference amino acid sequence.

Genome-wide Methylation Analyses

We pre-processed and normalized raw IDAT files from the Infinimum Methyl-ationEPIC array using the preprocess Funnorm function in the R package minfi Aryee et al. (2014). Probes on chromosomes X or Y, probes with detection p value greater than 0.5, or probes overlapping a SNP with dbSNP minor allele frequency greater than 10% were excluded. In order to understand the similarity of ovarian cells lines with human ovarian cancer, we compared the ovarian cells lines with human ovarian cancer samples available from Genomic Data Commons (https://gdc.cancer.gov/). The Genomic Data Commons contained 533 human methylation profiles of ovarian cancer and eight normal fallopian tissue samples. Methylation ofTCGA ovarian cancers was assessed using In-finium HumanMethylation27 BeadChip array (27,578 probes). The number of probes in common between the HumanMethylation27 platforms and the Meth-ylationEPIC platform was 18,016. On the common set of 18,016 probes, we quantified overall methylation in the TCGA samples and the ovarian cell lines as the fraction of CpG sites with $\beta > 0.3$. To identify differentially methylated CpG sites comparing normal fallopian tissue to TCGA ovarian cancers, we selected probes from the common set of 18,016 that were hyper-methylated in TCGA ovarian cancer (average $\beta > 0.4$) and unmethylated in normal fallopian tissue (average $\beta < 0.2$). In addition, we also selected probes that were hypo- methylated in TCGA ovarian cancer (average $\beta<0.1$) and hyper-methylated in normal fallopian (average $\beta>0.3$).

Gene Expression Analyses

Pre-processing and normalization of the 44k Agilent microarray for the ovarian cell lines has been previously described, and normalized expression data were available for 44 of the 45 tumors (Konecny et al., 2011). For copy number altered genes with known clinical relevance to cancer, we assessed whether amplified genes were overexpressed and whether deleted genes were under-expressed. The probability that a gene was overexpressed or underexpressed was estimated by a two-component pooled variance mixture model implemented in the R package CNPBayes (https://bioconductor.org/packages/release/bioc/html/CNPBayes.html). The location of the non-differentially expressed mixture component was fixed at the median log2 expression of unmethylated cell lines without copy number alterations. A gene was considered differentially expressed if the posterior probability of membership in the overexpressed or underexpressed mixture component exceeded 0.5.

Dose Response Models

Bayesian Model Averaging—We considered models of the form

$$\log C_i = \gamma_1 x_{i,1} + \ldots + \gamma_p x_{i,p}, + \varepsilon_i, \text{ where}$$

C_i denotes the logIC5$_0$ and $x_{i,j}$ is an indicator for the alteration status (0, not altered; 1,altered) of feature j in cell line i. The regression coefficient for feature j is the product of abinary indicator z_j and a real number h_j. We used a modified g-prior for γ such that γ_j was zero whenever z_j was zero (Hoff, 2009). For the vector of γ's with non-zero z's, we used a multivariate normal prior. We explored the space of the possible 2^p models using a Gibbs sampler. The binary features comprising the $\boldsymbol{x}$'s included somatic mutations, somatic structural variants (deletions, amplifications, in-frame fusions), methylation, and underexpression or overexpression. For the PARP inhibitor, we additionally considered the number of intra-chromosomal rearrangements and the HRD score as potential markers for HRD. For rearrangements, we computed the mean of the square-root-transformed frequency across all cell lines and defined a binary covariate for whether the square-root-transformed statistic was greater than the mean. We used the HRD score without transformation for Bayesian model averaging. For the univariate analyses described in the next section, we defined a binary covariate for HRD according to whether the score was larger than the mean. We obtained qualitatively similar inferences using the continuous HRD score (data not shown). For the inhibitor of the MEK pathway, one of the logIC$_{50}$ concentrations was missing. For this cell line, we used the posterior mean from the imputation described in greater detail below.

Univariate Analysis of Selected Features—For a given feature, our sampling model for the length-3 vector of inhibitor concentrations inducing 20%, 50%, and 80% cell death is

$$\log \boldsymbol{C}_{i,\text{altered}} = \boldsymbol{\mu} + \boldsymbol{\delta} + \boldsymbol{\epsilon}_{i,\text{altered}}$$

for a cell line with an alteration in this feature and

$$\log \boldsymbol{C}_{i,\text{WT}} = \boldsymbol{\mu} - \boldsymbol{\delta} + \boldsymbol{\epsilon}_{i,\text{WT}}$$

for a cell line without an alteration. With inhibitor concentrations on the log scale, the residuals are approximately multivariate-normal:

$$\boldsymbol{\epsilon}_{i,j} \sim \text{i.i.d. MVN}(0, \Sigma).$$

Computationally convenient conjugate priors for the unknown parameters in this model are

$$p(\boldsymbol{\mu}, \boldsymbol{\delta}, \Sigma) = p(\boldsymbol{\mu}) p(\boldsymbol{\delta}) p(\Sigma),$$

$$\boldsymbol{\mu} \sim \mathrm{MVN}(\boldsymbol{\mu}_0, \Sigma_0),$$

$$\boldsymbol{\delta} \sim \mathrm{MVN}(\boldsymbol{\delta}_0, \Psi_0), \text{ and}$$

$$\Sigma^{-1} \sim W\left(\nu_0, S_0^{-1}\right).$$

For some cell lines, inhibitor concentrations were incomplete. As the log$\boldsymbol{C}$ were highly correlated across cell lines, we imputed missing observations from the observed data using a Gibbs sampler. Inference regarding differences in mean log$\boldsymbol{C}$, given by the posterior distribution of 2S, was based on the marginal probability of the observed data integrating over the missing data. We reported 90% highest posterior density (HPD) intervals for the difference in the mean logIC_{50}.

DATA AND SOFTWARE AVAILABILITY

The accession number for the sequencing data reported in this paper is European Genome-Phenome Archive at ENSEMBL-EBI: EGAS00001002998. The R package Trellis for identifying somatic structural variants in tumor-only analyses are available from github (https://github.com/cancer-genomics/trellis).Code and data for reproducing figures and tables are available from Google Drive (http://bit.ly/cell_reports).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to thank Carolyn Hruban for help with artwork and members of our laboratories for critical review of this manuscript. This work was supported in part by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation; the Commonwealth Foundation; The Cigarette Restitution Fund; NIH grants CA121113, CA006973, and CA180950; and Department of Defense CDMRP Team Innovator Award W81XWH-14–1-0385. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

DECLARATION OF INTERESTS

V.E.V. is a founder of the Personal Genome Diagnostics, is a member of its Scientific Advisory Board and Board of Directors, and owns a Personal Genome Diagnostics stock, which is subject to certain restrictions under university policy. V.E.V. is also on the Scientific Advisory Board for Ignyta. Under separate licensing agreements between Agios Pharmaceuticals, Exact Sciences Corporation, Myriad Genetics, Personal Genome Diagnostics, Qiagen, Sysmex-Inostics, and the Johns Hopkins University, V.E.V. is entitled to a share of royalty and milestone payments received by the University on sales of products related to research described in this manuscript. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies. D.S. has a leadership role with Bio- marin, is a consultant for and has recieved research funding from Novartis, owns stock of and has recieved research funding from Pfizer, and is a consultant for Eli Lilly.

REFERENCES

Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, et al. (2012). Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br. J. Cancer 107, 1776–1782. [PubMed: 23047548]

Ai X, Wu Y, Zhang W, Zhang Z, Jin G, Zhao J, Yu J, Lin Y, Zhang W, Liang H, et al. (2013). Targeting the ERK pathway reduces liver metastasis of Smad4-inactivated colorectal cancer. Cancer Biol. Ther 14, 1059–1067. [PubMed: 24025354]

Akhoondi S, Sun D, von der Lehr N, Apostolidou S, Klotz K, Maljukova A, Cepeda D, Fiegl H, Dafou D, Marth C, et al. (2007). FBXW7/hCDC4 is a general tumor suppressor in human cancer. Cancer Res. 67, 9006–9012. [PubMed: 17909001]

Aldaz CM, Ferguson BW, and Abba MC (2014). WWOX at the crossroads of cancer, metabolic syndrome related traits and CNS pathologies. Biochim. Biophys. Acta 1846, 188–200. [PubMed: 24932569]

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, B0rresen-Dale A-L, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature 500, 415–421. [PubMed: 23945592]

Anglesio MS, Wiegand KC, Melnyk N, Chow C, Salamanca C, Prentice LM, Senz J, Yang W, Spillman MA, Cochrane DR, et al. (2013). Type- specific cell line models for type-specific ovarian cancer research. PLoS One 8, e72162. [PubMed: 24023729]

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, and Irizarry RA (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369. [PubMed: 24478339]

Ayoubi HA, Mahjoubi F, and Mirzaei R (2017). Investigation of the human H3.3B (H3F3B) gene expression as a novel marker in patients with colorectal cancer. J. Gastrointest. Oncol 8, 64–69. [PubMed: 28280610]

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607. [PubMed: 22460905]

Bast RC Jr., Feeney M, Lazarus H, Nadler LM, Colvin RB, and Knapp RC (1981). Reactivity of a monoclonal antibody with human ovarian carcinoma. J. Clin. Invest 68, 1331–1337. [PubMed: 7028788]

Basu S (2011). PP2A in the regulation of cell motility and invasion. Curr. Protein Pept. Sci 12, 3–11. [PubMed: 21190527]

Bénard J, Da Silva J, De Blois MC, Boyer P, Duvillard P, Chiric E, and Riou G (1985). Characterization of a human ovarian adenocarcinoma line, IGROV1, in tissue culture and in nude mice. Cancer Res. 45, 4970–4979. [PubMed: 3861241]

Berchuck A, Rodriguez G, Olt G, Whitaker R, Boente MP, Arrick BA, Clarke-Pearson DL, and Bast RC Jr. (1992). Regulation of growth of normal ovarian epithelial cells and ovarian cancer cell lines by transforming growth factor-beta. Am. J. Obstet. Gynecol 166, 676–684. [PubMed: 1536252]

Boddicker RL, Razidlo GL, Dasari S, Zeng Y, Hu G, Knudson RA, Greipp PT, Davila JI, Johnson SH, Porcher JC, et al. (2016). Integrated mate-pair and RNA sequencing identifies novel, targetable gene fusions in peripheral T-cell lymphoma. Blood 128, 1234–1245. [PubMed: 27297792]

Buick RN, Pullano R, and Trent JM (1985). Comparative properties of five human ovarian adenocarcinoma cell lines. Cancer Res. 45, 3668–3676. [PubMed: 4016745]

Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet 40, 722–729. [PubMed: 18438408]

Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615. [PubMed: 21720365]

Coleman RL, Oza AM, Lorusso D, Aghajanian C, Oaknin A, Dean A, Colombo N, Weberpals JI, Clamp A, Scambia G, et al.; ARIEL3 investigators (2017). Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 300, 1949–1961.

Colicchia V, Petroni M, Guarguaglini G, Sardina F, Sahún-Roncero M, Carbonari M, Ricci B, Heil C, Capalbo C, Belardinilli F, et al. (2017). PARP inhibitors enhance replication stress and cause mitotic catastrophe in MYCN-dependent neuroblastoma. Oncogene 36, 4682–4691. [PubMed: 28394338]

Conover CA, Hartmann LC, Bradley S, Stalboerger P, Klee GG, Kalli KR, and Jenkins RB (1998). Biological characterization of human epithelial ovarian carcinoma cells in primary culture: the insulin-like growth factor system. Exp. Cell Res 238, 439–449. [PubMed: 9473353]

Cowin PA, George J, Fereday S, Loehrer E, Van Loo P, Cullinane C, Etemadmoghadam D, Ftouni S, Galletta L, Anglesio MS, et al.; Australian Ovarian Cancer Study (2012). LRP1B deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin. Cancer Res. 72, 4060–4073. [PubMed: 22896685]

Coxon A, Rozenblum E, Park Y-S, Joshi N, Tsurutani J, Dennis PA, Kirsch IR, and Kaye FJ (2005). Mect1-Maml2 fusion oncogene linked to the aberrant activation of cyclic AMP/CREB regulated genes. Cancer Res. 65, 7137–7144. [PubMed: 16103063]

Cumbo C, Impera L, Minervini CF, Orsini P, Anelli L, Zagaria A, Coccaro N, Tota G, Minervini A, Casieri P, et al. (2018). Genomic BCR- ABL1 breakpoint characterization by a multi-strategy approach for “personalized monitoring” of residual disease in chronic myeloid leukemia patients. On-cotarget 9, 10978–10986.

Domcke S, Sinha R, Levine DA, Sander C, and Schultz N (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nat. Commun 4, 2126. [PubMed: 23839242]

Emoto M, Oshima K, Ishiguro M, Iwasaki H, Kawarabayashi T, and Kikuchi M (1999). Establishment and characterization of a serous papillary adenocarcinoma cell line of the human ovary in a serum-free culture. Pathol. Res. Pract 195, 237–242. [PubMed: 10337661]

Eva A, Robbins KC, Andersen PR, Srinivasan A, Tronick SR, Reddy EP, Ellmore NW, Galen AT, Lautenberger JA, Papas TS, et al. (1982). Cellular genes analogous to retroviral onc genes are transcribed in human tumour cells. Nature 295, 116–119. [PubMed: 6173755]

Fogh J (1986). Human tumor lines for cancer research. Cancer Invest. 4, 157–184. [PubMed: 3518877]

Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, et al. (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. 38, D652–D657. [PubMed: 19906727]

Frey MK, and Pothuri B (2017). Homologous recombination deficiency (HRD) testing in ovarian cancer clinical practice: a review ofthe literature. Gynecol. Oncol. Res. Pract. 4, 4. [PubMed: 28250960]

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 39, D876–D882. [PubMed: 20959295]

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, et al. (2012). Systematic identification ofgenomic markers ofdrug sensitivity in cancer cells. Nature 483, 570–575. [PubMed: 22460902]

Gorai I, Nakazawa T, Miyagi E, Hirahara F, Nagashima Y, and Minaguchi H (1995). Establishment and characterization of two human ovarian clear cell adenocarcinoma lines from metastatic lesions with different properties. Gynecol. Oncol 57, 33–46. [PubMed: 7535723]

Greenman CD, Cooke SL, Marshall J, Stratton MR, and Campbell PJ (2016). Modeling the evolution space of breakage fusion bridge cycles with a stochastic folding process. J. Math. Biol 72, 47–86. [PubMed: 25833184]

Gruszka-Westwood AM, Atkinson S, Summersgill BM, Shipley J, Elne-naei MO, Jain P, Hamoudi RA, Kaeda JS, Wotherspoon AC, Matutes E, and Catovsky D (2002). Unusual case of leukemic mantle cell lymphoma with amplified CCND1/IGH fusion gene. Genes Chromosomes Cancer 33, 206–212. [PubMed: 11793447]

Guertin DA, and Sabatini DM (2005). An expanding role for mTOR in cancer. Trends Mol. Med 11, 353–361. [PubMed: 16002336]

Hallett RM, Seong ABK, Kaplan DR, and Irwin MS (2016). Transcript signatures that predict outcome and identify targetable pathways in MYCN-amplified neuroblastoma. Mol. Oncol 10, 1461–1472. [PubMed: 27599694]

Hamilton TC, Young RC, McKoy WM, Grotzinger KR, Green JA, Chu EW, Whang-Peng J, Rogan AM, Green WR, and Ozols RF (1983). Characterization of a human ovarian carcinoma cell line (NIH:OVCAR-3) with androgen and estrogen receptors. Cancer Res. 43, 5379–5389. [PubMed: 6604576]

Hamilton TC, Young RC, and Ozols RF (1984). Experimental model systems of ovarian cancer: applications to the design and evaluation of new treatment approaches. Semin. Oncol 11, 285–298. [PubMed: 6385258]

Hattori R, Sase K, Eizawa H, Kosuga K, Aoyama T, Inoue R, Sasayama S, Kawai C, Yui Y, Miyahara K, et al. (1994). Structure and function of nitric oxide synthases. Int. J. Cardiol 47 (1, Suppl), S71–S75. [PubMed: 7537723]

Hills CA, Kelland LR, Abel G, Siracky J, Wilson AP, and Harrap KR (1989). Biological properties often human ovarian carcinoma cell lines: calibration in vitro against four platinum complexes. Br. J. Cancer 59, 527–534. [PubMed: 2653399]

Hoff P (2009). A First Course in Bayesian Statistical Methods (Springer).

Huang RY, Chen GB, Matsumura N, Lai H-C, Mori S, Li J, Wong MK, Konishi I, Thiery J-P, and Goh L (2012). Histotype-specific copy- number alterations in ovarian cancer. BMC Med. Genomics 5, 47.

Imura M, Yamashita S, Cai L-Y, Furuta J, Wakabayashi M, Yasugi T, and Ushijima T (2006). Methylation and expression analysis of 15 genes and three normally-methylated genes in 13 Ovarian cancer cell lines. Cancer Lett. 241,213–220. [PubMed: 16303245]

Ince TA, Sousa AD, Jones MA, Harrell JC, Agoston ES, Krohn M, Selfors LM, Liu W, Chen K, Yong M, et al. (2015). Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. Nat. Commun 6, 7419. [PubMed: 26080861]

Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Goncalves E, Barthorpe S, Lightfoot H, et al. (2016). A land-scape of pharmacogenomic interactions in cancer. Cell 166, 740–754. [PubMed: 27397505]

Jones S, Wang TL, Kurman RJ, Nakayama K, Velculescu VE, Vogel- stein B, Kinzler KW, Papadopoulos N, and Shih leM. (2012). Low-grade serous carcinomas of the ovary contain very few point mutations. J. Pathol 226, 413–420. [PubMed: 22102435]

Jones S, Anagnostou V, Lytle K, Parpart-Li S, Nesselbush M, Riley DR, Shukla M, Chesnick B, Kadan M, Papp E, et al. (2015). Personalized genomic analyses for cancer mutation discovery and interpretation. Sci. Transl. Med 7, 283ra53.

Katoh M (2015). Functional proteomics of the epigenetic regulators ASXL1, ASXL2 and ASXL3: a convergence of proteomics and epigenetics for translational medicine. Expert Rev. Proteomics 12, 317–328. [PubMed: 25835095]

Kent WJ (2002). BLAT—the BLAST-like alignment tool. Genome Res. 12, 656–664. [PubMed: 11932250]

Kidera Y, Yoshimura T, Ohkuma Y, Iwasaka T, and Sugimori H (1985). [Establishment and characterization of a cell line derived from mucinous cys- tadenocarcinoma of human ovary]. Nippon Sanka Fujinka Gakkai Zasshi i, 1820–1824.

Konecny GE, Winterhoff B, Kolarova T, Qi J, Manivong K, Dering J, Yang G, Chalukya M, Wang H-J, Anderson L, et al. (2011). Expression of p16 and retinoblastoma determines response to CDK4/6 inhibition in ovarian cancer. Clin. Cancer Res 17, 1591–1602. [PubMed: 21278246]

Korch C, Spillman MA, Jackson TA, Jacobsen BM, Murphy SK, Les- sey BA, Jordan VC, and Bradford AP (2012). DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. Gynecol. Oncol 127, 241–248. [PubMed: 22710073]

Labidi-Galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M, Bhattacharya R, Novak M, Jones S, Phallen J, et al. (2017). High grade serous ovarian carcinomas originate in the fallopian tube. Nat. Commun 8, 1093. [PubMed: 29061967]

Langdon SP, Lawrie SS, Hay FG, Hawkes MM, McDonald A, Hayward IP, Schol DJ, Hilgers J, Leonard RC, and Smyth JF (1988). Characterization and properties of nine human ovarian adenocarcinoma cell lines. Cancer Res. 48, 6166–6172. [PubMed: 3167863]

Lau DH, Lewis AD, and Sikic BI (1989). Association of DNA cross-linking with potentiation of the morpholino derivative of doxorubicin by human liver micro somes. J. Natl. Cancer Inst 81, 1034–1038. [PubMed: 2733045]

Layer RM, Chiang C, Quinlan AR, and Hall IM (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 15, R84. [PubMed: 24970577]

Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, Jones S, Sjoblom T, Park B-H, Parsons R, et al. (2008). Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc. Natl. Acad. Sci. USA 105, 16224–16229. [PubMed: 18852474]

Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]

Lounis H, Mes-Masson AM, Dion F, Bradley WE, Seymour RJ, Pro-vencher D, and Tonin PN (1998). Mapping of chromosome 3p deletions in human epithelial ovarian tumors. Oncogene 17, 2359–2365. [PubMed: 9811467]

Marjon K, Cameron MJ, Quang P, Clasquin MF, Mandley E, Kunii K, McVay M, Choe S, Kernytsky A, Gross S, et al. (2016). MTAP deletions in cancer create vulnerability to targeting ofthe MAT2A/PRMT5/RIOK1 axis. Cell Rep. 15, 574–587. [PubMed: 27068473]

McConechy MK, Anglesio MS, Kalloger SE, Yang W, Senz J, Chow C, Heravi-Moussavi A, Morin GB, Mes-Masson A-M, Carey MS, et al.; Australian Ovarian Cancer Study Group (2011). Subtype-specific mutation of PPP2R1A in endometrial and ovarian carcinomas. J. Pathol 223, 567–573. [PubMed: 21381030]

Meisner A, Kerr KF, Thiessen-Philbrook H, Wilson FP, Garg AX, Shlipak MG, Kavsak P, Whitlock RP, Coca SG, and Parikh CR (2018). Development of biomarker combinations for postoperative acute kidney injury via Bayesian model selection in a multicenter cohort study. Biomark. Res 6,3. [PubMed: 29344362]

Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, Fabbro M, Ledermann JA, Lorusso D, Vergote I, et al.; ENGOT-OV16/ NOVA Investigators (2016). Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. N. Engl. J. Med 375, 2154–2164. [PubMed: 27717299]

Motoyama T (1981). [Biological characterization including sensitivity to mito-mycin C of cultured human ovarian cancers (author's transl)]. Nippon Sanka Fujinka Gakkai Zasshi 33, 1197–1204. [PubMed: 6268719]

Neto EC, Jang IS, Friend SH, and Margolin AA (2014). The Stream algorithm: computationally efficient ridge-regression via Bayesian model averaging, and applications to pharmacogenomic prediction of cancer cell line sensitivity. Pac. Symp. Biocomput 2014, 27–38.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54. [PubMed: 27135926]

Nozawa S, Tsukazaki K, Sakayori M, Jeng CH, and Iizuka R (1988). Establishment of a human ovarian clear cell carcinoma cell line (RMG-I) and its single cell cloning-with special reference to the stem cell of the tumor. Hum. Cell 1, 426–435. [PubMed: 3154025]

Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palazzo J, Siprashvili Z, Mori M, McCue P, Druck T, et al. (1996). The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell 84, 587–597. [PubMed: 8598045]

Olshen AB, Venkatraman ES, Lucito R, and Wigler M (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557–572. [PubMed: 15475419]

Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, Nones K, Cowin P, Alsop K, Bailey PJ, et al.; Australian Ovarian Cancer Study Group (2015). Whole-genome characterization of chemoresistant ovarian cancer. Nature 521, 489–494. [PubMed: 26017449]

Paz MF, Fraga MF, Avila S, Guo M, Pollan M, Herman JG, and Estel- ler M (2003). A systematic profile of DNA methylation in human cancer cell lines. Cancer Res. 63, 1114–1121. [PubMed: 12615730]

Provencher DM, Lounis H, Champoux L, Tetrault M, Manderson EN, Wang JC, Eydoux P, Savoie R, Tonin PN, and Mes-Masson AM (2000). Characterization of four novel epithelial ovarian cancer cell lines. In Vitro Cell. Dev. Biol. Anim 36, 357–361. [PubMed: 10949993]

Pyndiah S, Tanida S, Ahmed KM, Cassimere EK, Choe C, and Sakamuro D (2011). c-MYC suppresses BIN1 to release poly(ADP-ribose) polymerase 1: a mechanism by which cancer cells acquire cisplatin resistance. Sci. Signal 4, ra19. [PubMed: 21447800]

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, and Korbel JO (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28, i333–i339. [PubMed: 22962449]

Roy D, Sin S-H, Damania B, and Dittmer DP (2011). Tumor suppressor genes FHIT and WWOX are deleted in primary effusion lymphoma (PEL) cell lines. Blood 118, e32–e39. [PubMed: 21685375]

Saglam O, Xiong Y, Marchion DC, Strosberg C, Wenham RM, Johnson JJ, Saeed-Vafa D, Cubitt C, Hakam A, and Magliocco AM (2017). ERBB4 expression in ovarian serous carcinoma resistant to platinum-based therapy. Cancer Control 24, 89–95. [PubMed: 28178720]

Sakayori M, Nozawa S, Udagawa Y, Chin K, Lee SG, Sakuma T, Iizuka R, Wada Y, Yoshida S, and Takeda Y (1990). [Biological properties of two newly established cell lines (RMUG-S, RMUG-L) from a human ovarian mucinous cystadenocarcinoma]. Hum. Cell 3, 52–56. [PubMed: 2083224]

Schilder RJ, Hall L, Monks A, Handel LM, Fornace AJ Jr., Ozols RF, Fojo AT, and Hamilton TC (1990). Metallothionein gene expression and resistance to cisplatin in human ovarian cancer. Int. J. Cancer 45, 416–422. [PubMed: 2307530]

Selby PJ, Thomas JM, Monaghan P, Sloane J, and Peckham MJ (1980). Human tumour xenografts established and serially transplanted in mice immunologically deprived by thymectomy, cytosine arabinoside and whole-body irradiation. Br. J. Cancer 41, 52–61. [PubMed: 7362779]

Shih IeM., Panuganti PK, Kuo K-T, Mao T-L, Kuhn E, Jones S Velcu- lescu VE, Kurman RJ, and Wang T-L (2011). Somatic mutations of PPP2R1A in ovarian and uterine carcinomas. Am. J. Pathol 178, 1442–1447. [PubMed: 21435433]

Shirley MD, Baugher JD, Stevens EL, Tang Z, Gerry N, Beiswanger CM, Berlin DS, and Pevsner J (2012). Chromosomal variation in lympho-blastoid cell lines. Hum. Mutat 33, 1075–1086. [PubMed: 22374857]

Simon WE, Albrecht M, Hansel M, Dietel M, and Holzel F (1983). Cell lines derived from human ovarian carcinomas: growth stimulation by gonadotropic and steroid hormones. J. Natl. Cancer Inst 70, 839–845. [PubMed: 6405069]

Smiraglia DJ, Rush LJ, Frühwald MC, Dai Z, Held WA, Costello JF, Lang JC, Eng C, Li B, Wright FA, et al. (2001). Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum. Mol. Genet 10, 1413–1419. [PubMed: 11440994]

Steinmeyer C, Berkholz A, Gebauer G, and Jager W (2003). The expression of hCG receptor mRNA in four human ovarian cancer cell lines varies considerably under different experimental conditions. Tumour Biol. 24,13–22. [PubMed: 12743422]

Swisher EM, Lin KK, Oza AM, Scott CL, Giordano H, Sun J, Konecny GE, Coleman RL, Tinker AV, O'Malley DM, et al. (2017). Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. Lancet Oncol. 18, 75–87. [PubMed: 27908594]

Tonon G, Modi S, Wu L, Kubo A, Coxon AB, Komiya T, O'Neil K, Stover K, El-Naggar A, Griffin JD, et al. (2003). t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nat. Genet 33, 208–213. [PubMed: 12539049]

Uehara S, Abe H, Hoshiai H, Yajima A, and Suzuki M (1984). Establishment and characterization of ovarian endometrioid carcinoma cell line. Gynecol. Oncol 17, 314–325. [PubMed: 6584390]

Valouev A, Weng Z, Sweeney RT, Varma S, Le Q-T, Kong C, Sidow A, and West RB (2014). Discovery of recurrent structural variants in nasopharyngeal carcinoma. Genome Res. 24, 300–309. [PubMed: 24214394]

van den Berg-Bakker CA, Hagemeijer A, Franken-Postma EM, Smit VT, Kuppen PJ, van Ravenswaay Claasen HH, Cornelisse CJ, and Schrier PI (1993). Establishment and characterization of 7 ovarian carcinoma cell lines and one granulosa tumor cell line: growth features and cytogenetics. Int. J. Cancer 53, 613–620. [PubMed: 8436435]

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–567. [PubMed: 23325432]

Venkatraman ES, and Olshen AB (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23, 657–663. [PubMed: 17234643]

Viallefont V, Raftery AE, and Richardson S (2001). Variable selection and Bayesian model averaging in case-control studies. Stat. Med 20, 3215–3230. [PubMed: 11746314]

Wang T-L, Maierhofer C, Speicher MR, Lengauer C, Vogelstein B, Kinzler KW, and Velculescu VE (2002). Digital karyotyping. Proc. Natl. Acad. Sci. USA 99, 16156–16161. [PubMed: 12461184]

Wang M, Wu W, Wu W, Rosidi B, Zhang L, Wang H, and Iliakis G (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res. 34, 6170–6182. [PubMed: 17088286]

Wiley A, Katsaros D, Chen H, Rigault de la Longrais IA, Beeghly A, Puopolo M, Singal R, Zhang Y, Amoako A, Zelterman D, and Yu H (2006). Aberrant promoter methylation of multiple genes in malignant ovarian tumors and in ovarian tumors with low malignant potential. Cancer 107, 299–308. [PubMed: 16773633]

Wilson AP (1984). Characterization of a cell line derived from the ascites ofa patient with papillary serous cystadenocarcinoma of the ovary. J. Natl. Cancer Inst 72, 513–521. [PubMed: 6583437]

Wilson AP, Dent M, Pejovic T, Hubbold L, and Radford H (1996). Characterisation of seven human ovarian tumour cell lines. Br. J. Cancer 74, 722–727. [PubMed: 8795574]

Yamada K, Tachibana T, Hashimoto H, Suzuki K, Yanagida S, Endoh H, Kimura E, Yasuda C and characterization of cell lines derived from serous adenocarcinoma (JHOS-2) and clear cell adenocarcinoma (JHOC-5, JHOC-6) of human ovary. Hum. Cell 12, 131–138. [PubMed: 10695020]

Yanagibashi T, Gorai I, Nakazawa T, Miyagi E, Hirahara F, Kitamura H, and Minaguchi H (1997). Complexity of expression of the intermediate filaments of six new human ovarian carcinoma cell lines: new expression of cytokeratin 20. Br. J. Cancer 76, 829–835. [PubMed: 9328139]

Yoshiya N (1986). [Establishment of a cell line from human ovarian cancer (undifferentiated carcinoma of FIGO classification) and analysis of its cell-biological characteristics and sensitivity to anticancer drugs]. Nippon Sanka Fujinka Gakkai Zasshi 38, 1747–1753. [PubMed: 3782955]

Yu M, Selvaraj SK, Liang-Chu MM, Aghajani S, Busse M, Yuan J, Lee G, Peale F, Klijn C, Bourgon R, et al. (2015).Are source for cell line authentication, annotation and quality control. Nature 520, 307–311. [PubMed: 25877200]

Zhang H, Chen ZH, and Savarese TM (1996). Codeletion of the genes for p16INK4, methylthioadenosine phosphorylase, interferon-alpha1, interferon-beta1, and other 9p21 markers in human malignant cell lines. Cancer Genet. Cytogenet 86, 22–28. [PubMed: 8616780]

Zhang L, Duan CJ, Binkley C, Li G, Uhler MD, Logsdon CD, and Simeone DM (2004). A transforming growth factor beta-induced Smad3/ Smad4 complex directly activates protein kinase A. Mol. Cell. Biol 24, 2169–2180. [PubMed: 14966294]

Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou J-Y, Petyuk VA, Chen L, Ray D, et al.; CPTAC Investigators (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell 166, 755–765. [PubMed: 27372738]

Zöchbauer-Müller S, Wistuba II, Minna JD, and Gazdar AF (2000). Fragile histidine triad (FHIT) gene abnormalities in lung cancer. Clin. Lung Cancer 2, 141–145. [PubMed: 14731325]

Highlights

- Compendium of somatic sequence and structural variants in ovarian cancer cell lines
- Development of Trellis algorithm for tumor-only structural variant detection
- Integration of DNA, mRNA, and methylation alterations to predict drug sensitivity

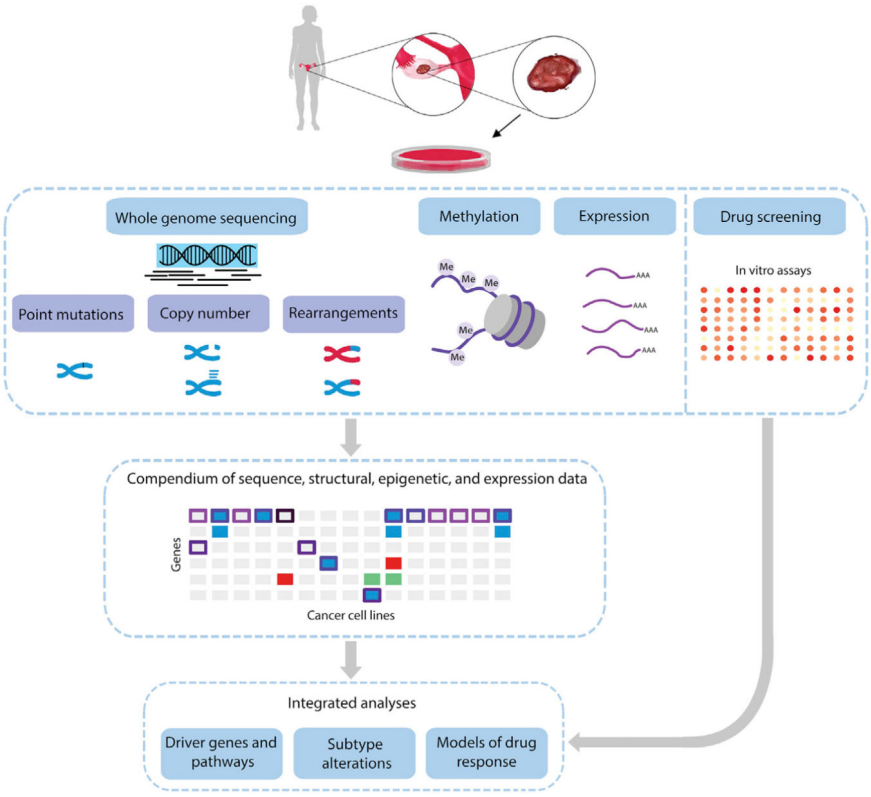


Figure 1. Overview of Genomic, Epigenomic, Expression, and Therapeutic Analyses of Ovarian Cancer Cell Lines

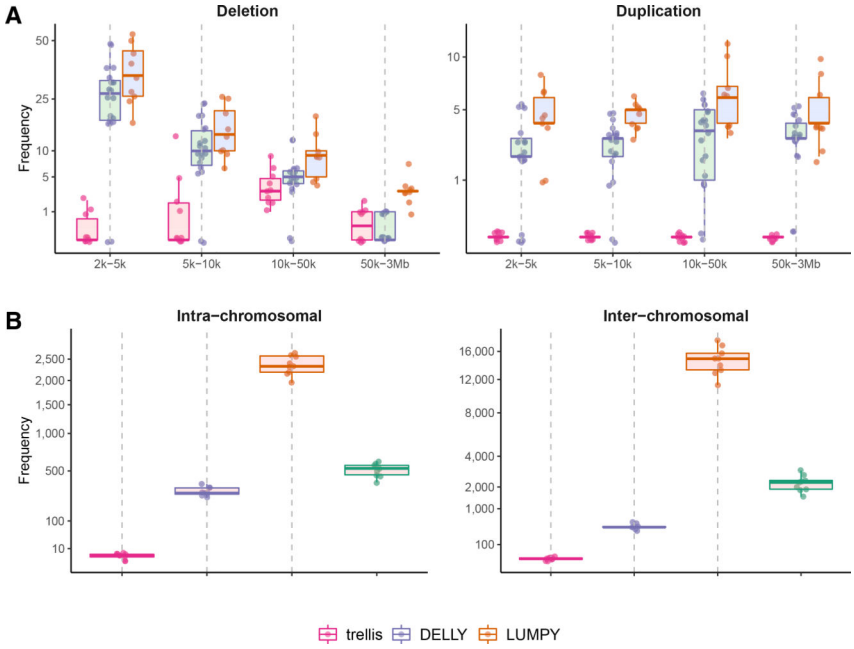


Figure 2. Number of False-Positive Somatic Structural Variant Identifications in Lympho-blastoid Cell Lines

(A and B) Estimated number of false-positive somatic deletions and duplications (A) and somatic intrachromosomal and inter-chromosomal rearrangements (B).

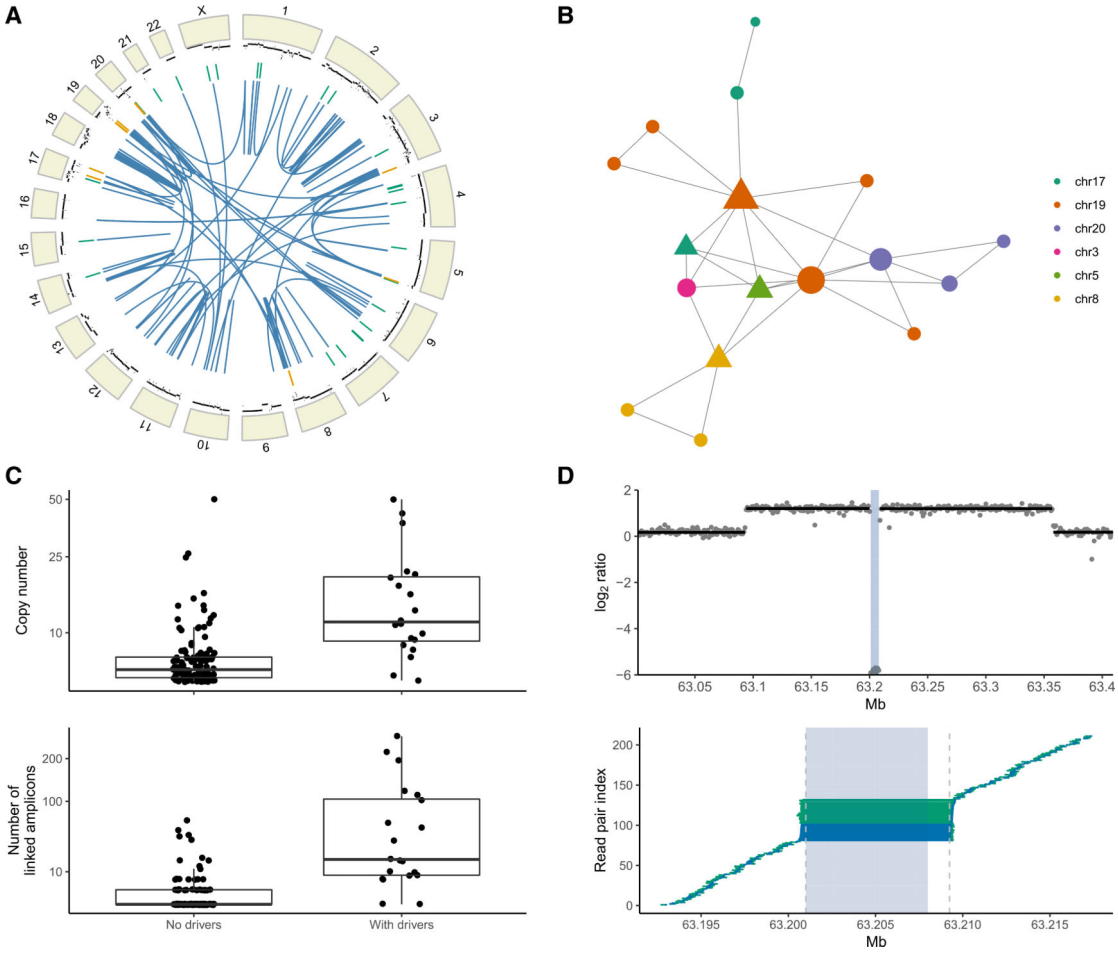


Figure 3. Trellis Approach for Characterization of Genomic Structural Alterations

(A) Circos plot displaying focal deletions (green), amplifications (orange), and intra- and inter-chromosomal rearrangements (blue) for cell line FU-OV-1.

(B) Improperly paired reads established connections (edges) between distant amplicons (nodes) visualized as a graph. The size of the plotting symbols is proportional to the number of sites in which the amplicon was inserted, and the triangle shape indicates an amplicon involving a known driver.

(C) The average maximum copy number (top) and mean number of amplicon links (bottom) for amplicon groups with and without drivers.

(D) Top: Segmented normalized coverage identified a homozygous deletion (shaded). Bottom: Rearranged read pairs improved the precision of the deletion breakpoints. Lines connecting the read pairs indicate whether the positive or negative strand was sequenced (blue, positive; green, negative).

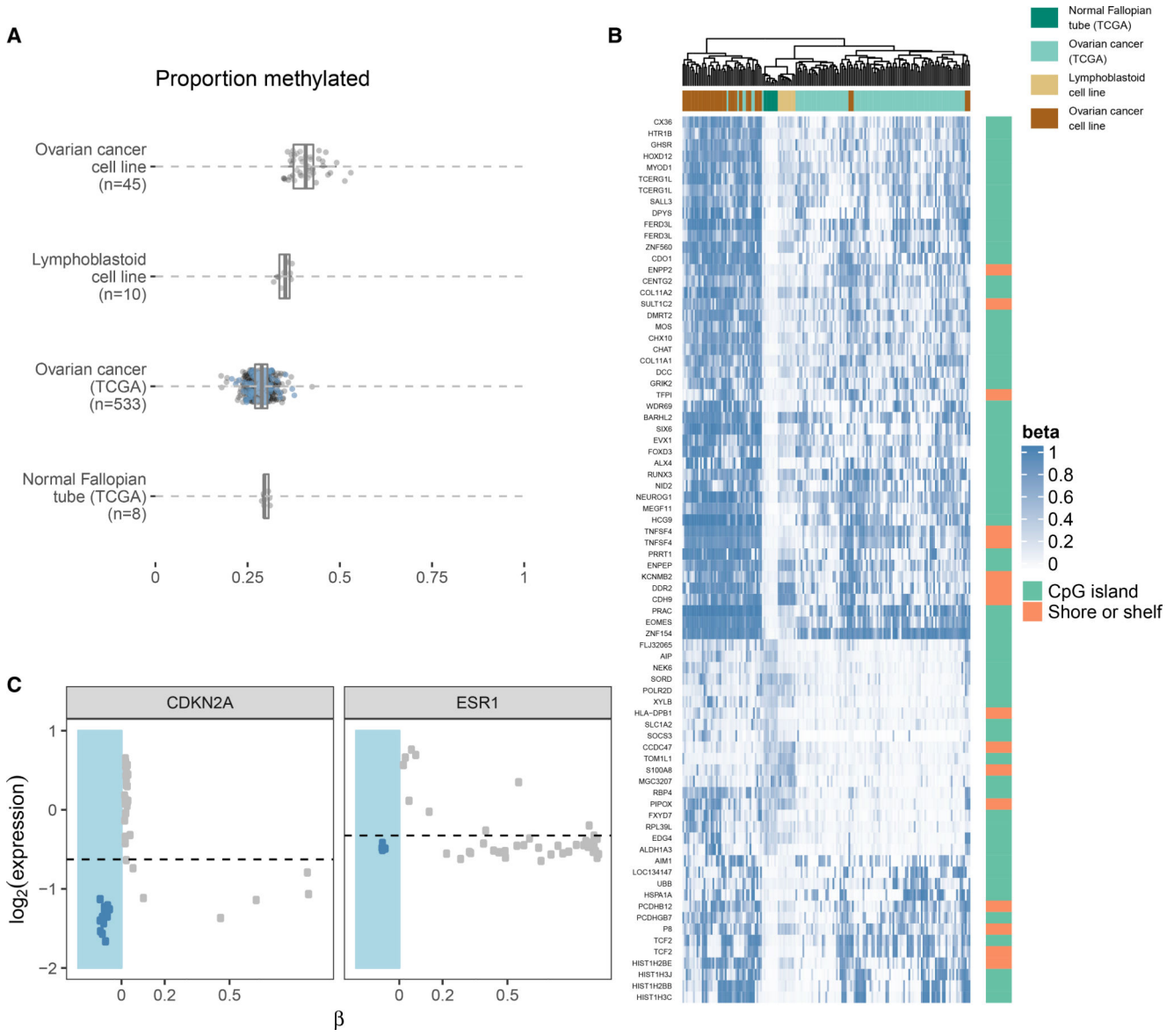


Figure 4. Methylation of CpG Sites in Ovarian Cancers and Normal Fallopian Tissue

(A) The proportion of methylated CpG sites (mean $\beta > 0.3$) in the lymphoblastoid cell lines, ovarian cell lines, TCGA ovarian cancers, and TCGA normal fallopian tissues.

(B) Plotted are 96 probes that were differentially methylated between normal TCGA fallopian tissue and 100 randomly selected TCGA ovarian tumors (blue points, A).Among these probes, the lymphoblastoid cell lines were most correlated with normal fallopian tissue and the ovarian cell lines were most correlated with TCGA ovarian tumors, suggesting that the cell line effect does not dominate among probes that were differentially methylated in these tissues. Among probes that were methylated in TCGA ovarian and unmethylated in TCGA fallopian, the ovarian cell lines were predominantly methylated and have quantitatively higher β values. While copy number analyses suggested that the purity in the ovarian cell lines was ≈00%, the median tumor purity of TCGA ovarian tumors was 85% (interquartile range, 78%−88%).

(C) Genes *CDKN2A* and *ESR1* exhibit bimodal gene expression explained by homozygous copy number deletions (blue points in x-axis margin) or methylation levels above 0.2.

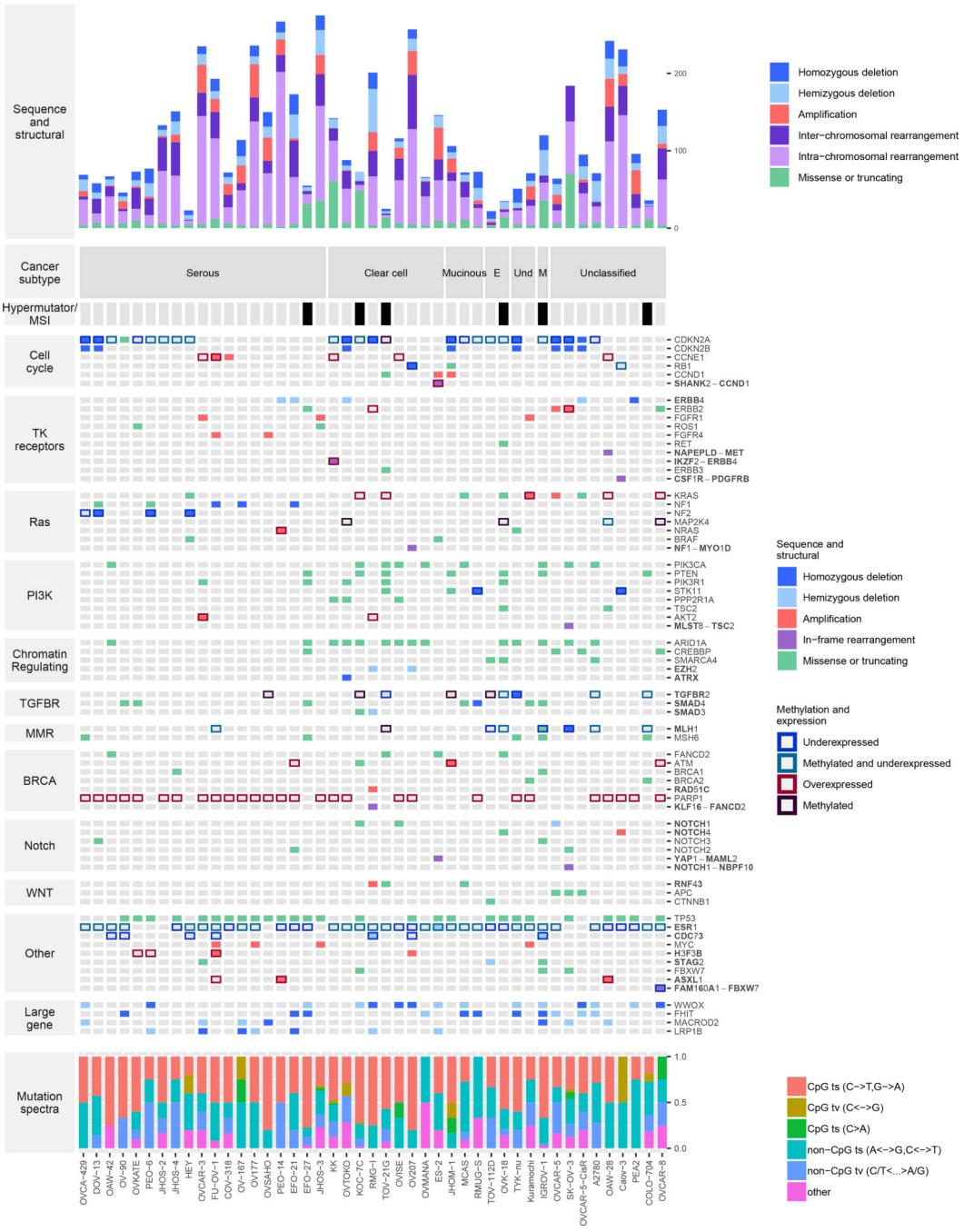


Figure 5. Sequence, Structural, Epigenomic, and Expression Alterations in Ovarian Cancer Cell Lines

Cell lines were grouped by tumor subtype (E, endometrioid; Und, undifferentiated; M, mixed). For many of the pathways, mutual exclusivity of genomic alterations within the pathway is evident (e.g., cell cycle, TK receptors, TGFBR, BRCA, and WNT). The group indicated as Other contains genes that are clinically relevant for ovarian cancer but cannot be easily categorized by a single molecular process. Methylation and expression were not evaluated for the Large Gene group.

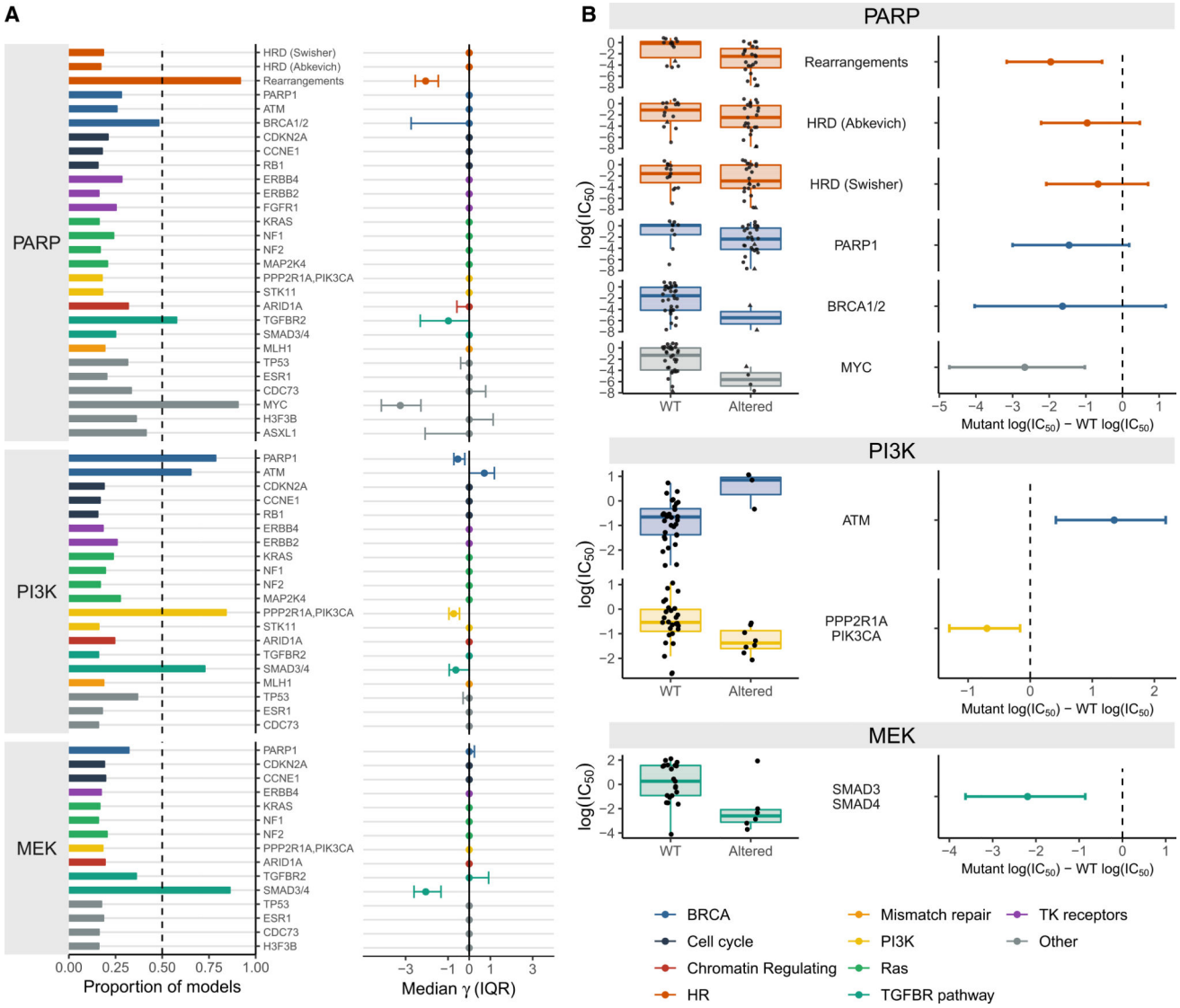


Figure 6. Sensitivity and Resistance to Pathway Inhibitors

(A) Bayesian model averaging was used to identify features associated with response to drug. Features selected in fewer than half of the multi-variate models have a posterior probability of being non-zero ≤ 0.5 (vertical dashed line, left) and a posterior median of zero (right).

(B) Boxplots of inhibitor concentrations for features selected by this approach, as well as HRD, *PARP1,* and *BRCA1/2* (left). The two cell lines with *BRCA1/2* mutations are indicated by triangles inthe PARP pathway. Right: The difference in mean logIC50 concentrations by alteration status and the90% highest posterior density (HPD) interval for the difference.