Oncogene
https://doi.org/10.1038/s41388-018-0644-y

ARTICLE

Hybrid sequencing-based personal full-length transcriptomic analysis implicates proteostatic stress in metastatic ovarian cancer

Ying Jing[1,2] · Yi Zhang[1,2] · Hui Zhu[3] · Ke Zhang[4] · Mei-Chun Cai[5] · Pengfei Ma[1] · Peiye Shen[1] · Zhenfeng Zhang[5] · Minghui Shao[6] · Jing Wang[6] · Minhua Yu[1,2] · Xia Yin[1,2] · Meiying Zhang[1,2] · Yuan Hu[1,2] · Danni Chen[7] · Wen Di[1,2] · Xiaojie Wang[7] · Guanglei Zhuang[1,2]

Received: 16 June 2018 / Revised: 16 October 2018 / Accepted: 4 December 2018

Abstract

Comprehensive molecular characterization of myriad somatic alterations and aberrant gene expressions at personal level is key to precision cancer therapy, yet limited by current short-read sequencing technology, individualized catalog of complete genomic and transcriptomic features is thus far elusive. Here, we integrated second- and third-generation sequencing platforms to generate a multidimensional dataset on a patient affected by metastatic epithelial ovarian cancer. Whole-genome and hybrid transcriptome dissection captured global genetic and transcriptional variants at previously unparalleled resolution. Particularly, single-molecule mRNA sequencing identified a vast array of unannotated transcripts, novel long noncoding RNAs and gene chimeras, permitting accurate determination of transcription start, splice, polyadenylation and fusion sites. Phylogenetic and enrichment inference of isoform-level measurements implicated early functional divergence and cytosolic proteostatic stress in shaping ovarian tumorigenesis. A complementary imaging-based high-throughput drug screen was performed and subsequently validated, which consistently pinpointed proteasome inhibitors as an effective therapeutic regime by inducing protein aggregates in ovarian cancer cells. Therefore, our study suggests that clinical application of the emerging long-read full-length analysis for improving molecular diagnostics is feasible and informative. An in-depth understanding of the tumor transcriptome complexity allowed by leveraging the hybrid sequencing approach lays the basis to reveal novel and valid therapeutic vulnerabilities in advanced ovarian malignancies.

Introduction

Epithelial ovarian cancer (EOC) represents the most lethal gynecologic malignancy, and is hallmarked by early metastatic dissemination and pervasive genomic variations [1, 2]. Recent advances in high-throughput second-generation sequencing techniques have aided a systematic and comprehensive characterization of the genetic basis

These authors contributed equally: Ying Jing, Yi Zhang, Hui Zhu

Supplementary information The online version of this article (https://doi.org/10.1038/s41388-018-0644-y) contains supplementary material, which is available to authorized users.

✉ Xiaojie Wang
wangxiaojie2008@hotmail.com

✉ Guanglei Zhuang
zhuangguanglei@gmail.com

[1] State Key Laboratory of Oncogenes and Related Genes, Department of Obstetrics and Gynecology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2] Shanghai Key Laboratory of Gynecologic Oncology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[3] Department of Anesthesiology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[4] Department of Oncology, Rizhao People's Hospital, Shandong, China

[5] State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[6] Novogene Bioinformatics Technology Co., Ltd, Beijing, China

[7] Department of Obstetrics and Gynecology, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Published online: 07 January 2019

underlying ovarian pathogenesis [3–5]. In particular, large-scale copy number alterations are predominant in EOC [6], leading to aberrant repertoire of transcribed elements and subsequent transcriptional addiction [7, 8]. Therefore, a more thorough delineation of tumor genomic and transcriptomic traits promises to provide an enormous capability for nominating new therapeutic strategies against this deadly disease [9].

However, there are multiple critical limitations intrinsically associated with previous cancer profiling data generated from second-generation sequencing platforms. Specifically, the amplification and fragmentation steps are required during library preparation, both of which may potentially introduce substantial biases and errors in the obtained short reads and downstream sequence assembly [10–12]. These key issues pose formidable challenges especially for the analysis of extremely complex cancer transcriptome [13]. Owing to the prodigious diversification of possible alternative transcripts in neoplasms, current workflows often fail to resolve the transcription start sites, splice junctions, polyadenylation positions, and gene fusion events with sufficient accuracy [14].

Third-generation sequencing technologies, such as the single-molecule real-time (SMRT) system from Pacific Biosciences (PacBio), offer a unique opportunity for full-length transcriptome construction [15–17]. PacBio SMRT technology produces exceptionally long-read lengths (>10 kb), which usually span the entire transcript molecule and obviate the need for isoform assembly [18]. Although the sequencing error rates are typically high (~15%), various correction approaches have been designed to address the erroneous long reads, including traversing circularized templates multiple times to derive consensus sequences [19], and a hybrid method that takes advantage of second-generation, high-fidelity short fragments [20]. Recently, PacBio isoform sequencing (Iso-Seq) has been widely applied to analyze the transcriptional landscapes in fission yeast, plants, human organs and cancer cell lines [21–27]. Nonetheless, an encyclopedic survey of the human cancer transcriptome composition in primary tumor specimens is by far lacking.

In this work, we present hybrid sequencing-based personal transcriptomic analysis on a patient suffering from disseminated epithelial ovarian carcinoma. Our data provide the first detailed portrait on the transcription start, splice, polyadenylation and fusion sites of paired primary and metastatic lesions at high resolution. The information assists to infer the phylogenetic structures of malignant EOC, and further enables the implication of proteostatic stress in the process of ovarian carcinogenesis. An imaging-based anticancer small-molecule screen, followed by pharmacological and mechanistic validation studies, establishes protein homeostasis as a bona fide therapeutic target and justifies the clinical translation of proteasome inhibitors in invasive ovarian cancer.

Results

Integrative genetic and functional analyses of metastatic ovarian cancer

Early intraperitoneal spread is a distinct feature of ovarian cancer compared with most other solid malignancies, largely due to its innate aggressive nature and lack of physical barriers in the peritoneal cavity [28]. In order to gain new insights into EOC disease progression and treatment at personal level, we exploited integrative analyses that combined deep sequencing and experimental approaches to systematically depict the genomic, transcriptomic, and functional properties of a clinical case (Fig. 1). A patient diagnosed with pathologically confirmed metastatic serous carcinoma (Supplementary Fig. 1) underwent laparoscopic debulking surgery in our institution, during which blood cells, normal ovarian tissue, primary tumor and distal lesion were collected with written informed consent. Paired-end libraries of genomic DNA from blood cell control and cancer samples, and rRNA-depleted total RNA from normal and matched ovarian tumor tissues were sequenced on an Illumina HiSeq X Ten platform. Meanwhile, we subjected Iso-Seq libraries of unfragmented polyA + full-length transcripts from ovarian epithelial and neoplastic specimens to third-generation RNA sequencing by using the PacBio RS II system. Benefiting from the multidimensional information, we were able to comprehensively infer the personal landscape of somatic mutations, copy number alterations, transcriptome variations, and gene aberrations at an unprecedented scope (Fig. 1a). In parallel, an unbiased imaging-based high-throughput drug screen was performed to pinpoint effective therapeutic modalities against ovarian tumor cells (Fig. 1b). These accompanying complementary sequencing and screening profiles were consolidated for identifying emergent cancer-specific vulnerabilities that might arise as a consequence of genetic or transcriptomic abnormalities.

Mutational landscape and dysregulated expression profiles

Whole-genome sequencing (WGS) achieved an average of 50× coverage with 98.62% of targets being sequenced to a depth of at least 10× (Supplementary Table S1). We detected 7053 and 6852 somatic alterations, including a total of 50 and 47 nonsynonymous exonic substitutions, from primary and metastatic tumors, respectively (Supplementary Table S2). $KRAS^{G12R}$ and $FGFR2^{T14I}$ were

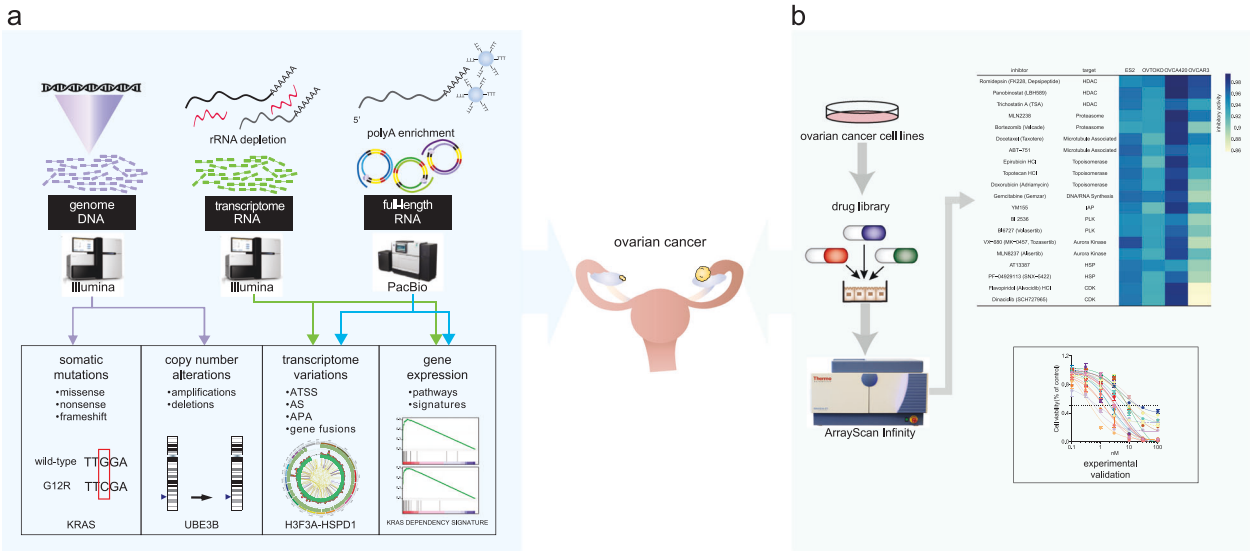


Fig. 1 Integrative genetic and functional analyses of metastatic ovarian cancer. **a** Genetic analysis was performed by running whole-genome sequencing (purple lines) and transcriptome sequencing (green lines) on the Illumina HiSeq X Ten system, or by running full-length RNA sequencing (blue lines) on the Pacific Biosciences RS II platform. Results of each sequencing strategy were integrated to identify tumor-specific aberrations including nucleotide acid mutations, copy number alterations, transcriptome variations, and differential gene expression. **b** Functional analysis was conducted using high-content imaging-based drug screen and cell proliferation assays. Effective small-molecule candidates were subjected to experimental validation in a larger ovarian cancer cell line panel. ATSS alternative transcription start sites, AS alternative splicing, APA alternative polyadenylation sites

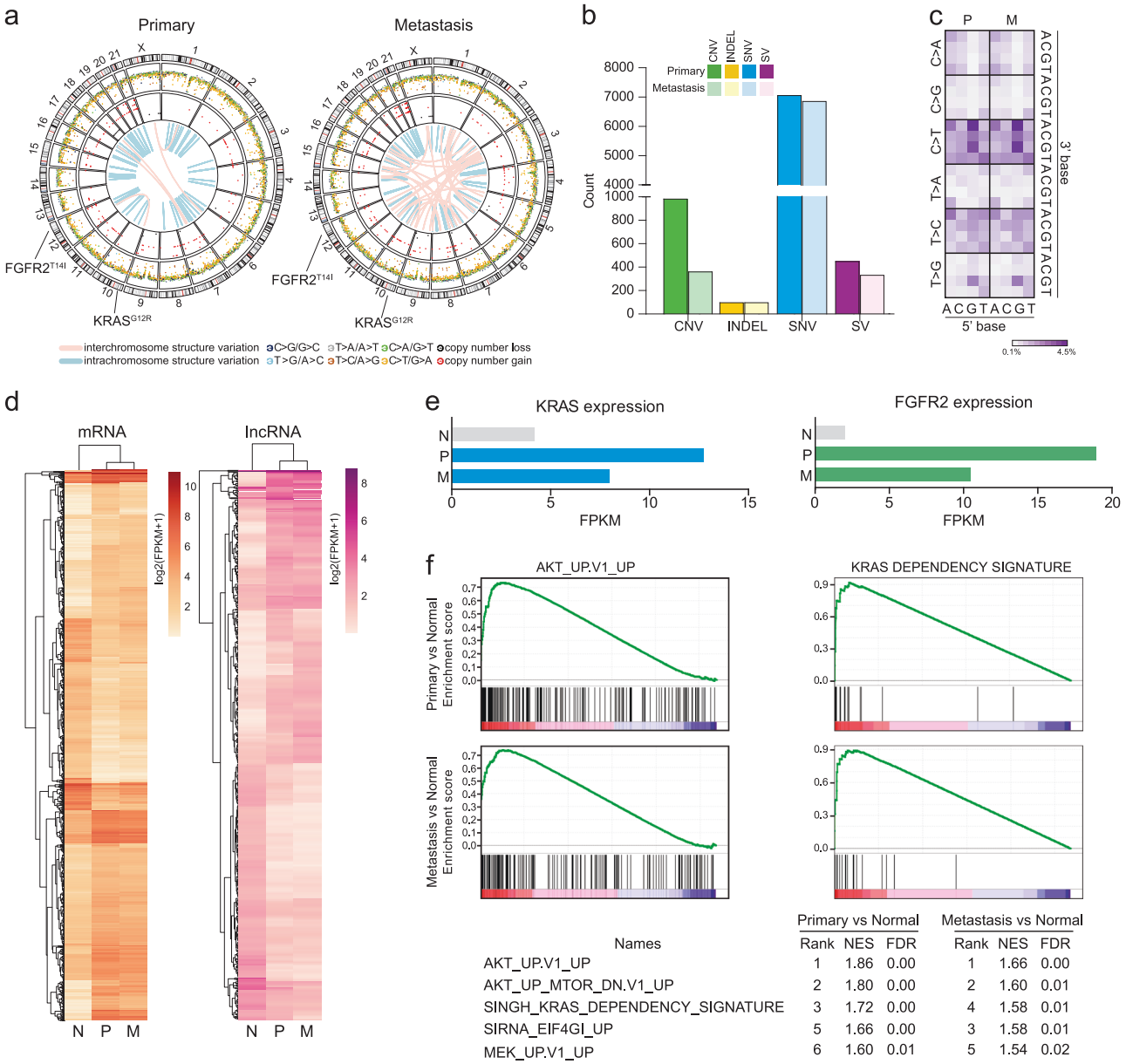
consistently identified as putative driver mutations regardless of neoplastic lesions (Fig. 2a), and considered to be clonal events on the basis of cancer cell fraction (CCF) analysis (the fraction of tumor cells harboring the variants) (Supplementary Fig. 2a). Additionally, the repertoire of genomic aberrations (Fig. 2b; Supplementary Tables S2-4), mutation spectra (Supplementary Fig. 2b), and mutational signatures (Fig. 2c; Supplementary Fig. 2c) of single-nucleotide variations (SNV) were fairly similar between the two samples. Both primary and metastatic cancers exhibited notable intratumor heterogeneity, as evidenced by inferred oligoclonal structure using PyClone (Supplementary Fig. 3a) and SciClone (Supplementary Fig. 3b). Two-dimensional cluster correlation of subclonal composition indicated that the original neoplasm and disseminated counterpart were constituted of broadly uniform genetic deposits (Supplementary Fig. 3c).

Short reads RNA sequencing-based gene expression data (Supplementary Table S5) and long noncoding RNA (lncRNA) profiles (Supplementary Table S6) unambiguously distinguished normal and tumor specimens following unsupervised hierarchical clustering (Fig. 2d). Interestingly, the transcriptional levels of two mutated driver oncogenes, *KRAS* and *FGFR2*, were both elevated in malignant tissues relative to normal ovarian epithelium (Fig. 2e). Accordingly, gene set enrichment analysis of differentially expressed transcripts pinpointed cancer-specific upregulation of *KRAS* and *AKT* pathways (Fig. 2f). These results suggested that in line with identical somatic driver mutations, the paired primary and metastatic ovarian tumors phenotypically resembled each other to a large extent solely based on the gene expression measurements derived from standard second-generation RNA sequencing, underscoring the potential value of more sophisticated transcriptome assessment using emerging technologies.

Full-length transcriptome and molecular evolution

A fundamental paradigm shift is underway from "gene-centric" to "isoform-centric" understanding of human transcriptome, owing to the frequent generation of diverse transcripts from a single genomic locus. Alternative transcriptional initiation or termination (transcriptional regulation) can give rise to different pre-mRNAs, which may further undergo alternative splicing (post-transcriptional regulation), resulting in multiple mRNA variants from the same gene [25, 29–35]. In addition, chimeric gene fusions, which have been widely discovered across various malignancies, further increase the complexity of cancer transcriptome [36, 37]. Although numerous bioinformatics tools have been developed to decode the complicated messages, current pipeline for processing second generation RNA sequencing data still presents major challenges in transcript assembly, arising from the ambiguous assignment of short reads to gene products [12, 38].

To overcome these limitations, we leveraged full-length transcriptomic analysis provided by the PacBio third-generation Iso-Seq platform, and obtained a total of



Names	Primary vs Normal			Metastasis vs Normal		
	Rank	NES	FDR	Rank	NES	FDR
AKT_UP.V1_UP	1	1.86	0.00	1	1.66	0.00
AKT_UP_MTOR_DN.V1_UP	2	1.80	0.00	2	1.60	0.01
SINGH_KRAS_DEPENDENCY_SIGNATURE	3	1.72	0.00	4	1.58	0.01
SIRNA_EIF4GI_UP	5	1.66	0.00	3	1.58	0.01
MEK_UP.V1_UP	6	1.60	0.01	5	1.54	0.02

Fig. 2 Mutational landscape and dysregulated expression profiles. **a** Circos plots of genetic mutations depicting DNA sequence aberrations, copy number changes, and structure variations in the primary and metastatic ovarian tumors from whole-genome sequencing. Chromosomes were arranged circularly end-to-end with rainfall plot of single nucleotide variations in the second ring (skyblue: T > G/A > C; darkorange3: T > C/A > G; darkgray: T > A/A > T; darkgoldenrod1: C > T/G > A; darkblue: C > G/G > C; chartreuse3: C > A/G > T). The third ring displayed copy number alterations (red dots, amplification; black dots, deletion). Within the inner ring each blue line denoted an intrachromosomal rearrangement and each pink line represented an interchromosomal rearrangement. **b** Numbers of single nucleotide variations (SNVs), small insertions and deletions (INDELs), copy number variations (CNVs), and structural variations (SVs) as detected by whole-genome sequencing in primary tumor and metastasis samples. **c** Heatmap of frequencies of the 96 trinucleotide mutation types in sequenced ovarian tumors. P Primary tumor, M Metastatic tumor. **d** Hierarchical clustering of mRNA and long noncoding RNA (lncRNA) expression values in ovarian cancers and normal tissues. N Normal tissue, P Primary tumor, M Metastatic tumor. **e** Upregulated gene expression of *KRAS* and *FGFR2* in ovarian tumors compared with normal tissues. N Normal tissue, P Primary tumor, M Metastatic tumor, FPKM fragments per kilo-million. **f** Top, Gene Set Enrichment Analysis (GSEA) plots indicated upregulation of *AKT* and *KRAS* dependency signature gene sets in primary and metastatic ovarian cancer relative to normal tissues. Bottom, Top-five-ranked upregulated oncogenic signature gene sets in ovarian tumor specimens

approximately 6 million sub-reads from three size-selected libraries per sample with an average read-length of 2489 bp (Supplementary Table S7). We custom designed a stringent analytic workflow (Supplementary Fig. 4a), incorporating circular-consensus reads clustering [19], hybrid error correction [20], and TAPIS algorithm [21], to build high-

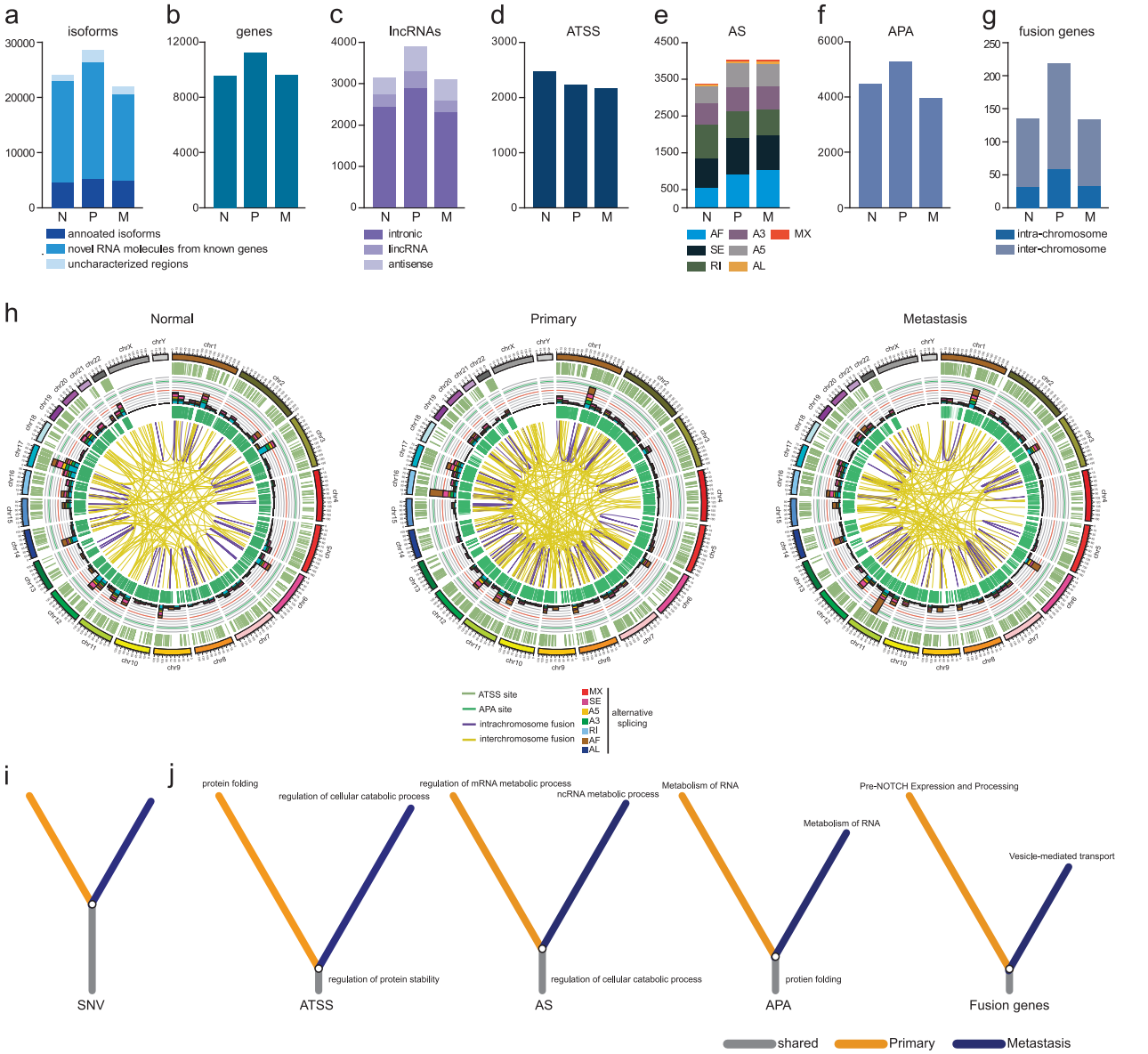


Fig. 3 Full-length transcriptome and molecular evolution. **a**–**g** Numbers of isoforms (**a**), genes (**b**), lncRNAs (**c**), alternative transcription start sites (ATSS) (**d**), alternative splicing (AS) (**e**), alternative polyadenylation sites (APA) (**f**), and fusion genes (**g**) detected by full-length transcriptome sequencing in each sample. N Normal tissue, P Primary tumor, M Metastatic tumor, MX mutually exclusive exons, SE skipping exon, A5 alternative 5′ splice-site, A3 alternative 3′ splice-site, RI retained intron, AF alternative first exon, AL alternative last exon. **h** Circos visualization for the full-length transcriptome characteristics of three sequenced samples. From outer rings to inner rings: karyotype of human genome, alternative transcription start sites, number of alternative splicing events per 30 Mb of seven categories (MX, SE, A5, A3, RI, AF and AL), alternative polyadenylation sites, and fusion genes. **i**, **j** The evolutionary course of ovarian cancer was depicted on the basis of all somatic SNV (**i**), ATSS, AS, APA and fusion genes (**j**)

quality nonredundant personal transcriptome of normal, primary tumor and metastasis tissues (Supplementary Table S8). Of note, only full-length sequences with PacBio library adaptors and identifiable polyA tails were retained, and the distribution of aligned reads across human chromosomes validated the superb performance of our procedure (Supplementary Fig. 4b). A combined 58,410 unique transcripts (normal: 24,150; primary tumor: 28,581; metastasis: 21,962) were identified (Fig. 3a), among which 13.4, 78.8 and 7.8% were assigned to annotated isoforms, novel RNA molecules of known genes and uncharacterized regions of reference genome, respectively. The length density of third-generation transcriptome data suggested a genuine representation of existing reference records (Supplementary Fig. 5a). Intriguingly, fully annotated isoforms exhibited relatively higher transcriptional levels (Supplementary Fig. 5b). Overall, 12,180 cataloged genes (Fig. 3b) in the Ensembl annotation and 8885 inferred lncRNAs (Fig. 3c)

were covered by PacBio sequencing (Supplementary Tables S8-9), and as expected, their expression values hierarchically classified normal and tumor samples (Supplementary Fig. 5c; Supplementary Tables S10-11). For comparison, 11,543 isoforms, 10,607 genes, and 409 lncRNAs (fragments per kilobase of transcript per million mapped reads > 10) were detected using Illumina data (Supplementary Fig. 6a–c), demonstrating a relatively equal discovery rate of expressed transcripts.

The single-molecule transcriptomic profiling allowed for accurate multifaceted exploration of alternative transcription start, splice, polyadenylation and fusion events undergoing in primary ovarian and neoplastic tissues. In total, 6086 alternative transcription start sites (ATSS) (Fig. 3d; Supplementary Table S12), 9112 alternative splicing events (AS) (Fig. 3e; Supplementary Table S13), 12,061 alternative polyadenylation positions (APA) (Fig. 3f; Supplementary Table S14) and 489 fusion transcripts (Fig. 3g; Supplementary Table 15) were observed in full-length transcriptomes of normal ovarian, primary tumor and metastasis tissues, all of which were notably different from the corresponding estimates with Illumina data (Supplementary Fig. 6d–g). As short reads-based analysis was potentially prone to artifacts, the results of our hybrid approach suggested that widespread transcriptional regulation (ATSS and APA) possibly contributed to isoform diversity more significantly than genomic (fusions) or post-transcriptional regulation (AS).

Taking advantage of the comprehensive genomic (Fig. 2a) and transcriptomic information (Fig. 3h), we performed phylogenetic reconstruction to estimate the evolutionary trajectories from primary to metastatic ovarian tumors. Reminiscent of previous observations [39–43], the repertoire of SNV displayed pronounced temporal difference, indicating genotypic divergence (Fig. 3i). More strikingly, a larger proportion of differential ATSS, AS, APA or fusion occurrence relative to SNV discrepancy was evident between the two samples (Fig. 3j), in agreement with an earlier functional alteration conceivably associated with cancer cell dissemination, as illuminated by biological pathway analysis.

Aberrant gene products and proteostatic stress

Alternative transcription start, splice, polyadenylation, and fusion events produce diverse forms of transcripts and consequently enhance the transcriptome complexity. Indeed, the number of gene isoforms positively correlated with that of alternative transcriptional incidents in each sample (Fig. 4a). Compared with Ensembl annotations, the PacBio full-length transcriptome contained a higher proportion of multi-isoform genes (Supplementary Fig. 7). Notably, multi-isoform genes displayed significantly higher expression levels in comparison to single-isoform genes (Fig. 4b). We sought to determine whether aberrantly transcribed gene products associated with transcriptome diversification might contribute to ovarian cancer development. To this end, a list of genes with the most frequent transcripts variety (ATSS, AS and APA sites) were subjected to pathway analysis, which showed neoplasm-specific activation of many processes related to oncogenic signaling and metabolism (Supplementary Table S16). Of particular interest, molecules involved in protein folding and response to topologically incorrect protein were enriched in primary tumor and metastasis, respectively (Fig. 4c), implying the functional impact of proteomic instability on ovarian pathogenesis. Furthermore, a subset of proteotoxic stress responders contained differentially expressed transcripts (Fig. 4d; Supplementary Table S17). Considering the isoform diversity and impact of proteostasis, we selected the proteostatic regulators for detailed analysis. Meanwhile, we confirmed that at the isoform level, proteostatic regulators were often altered in multifarious ways (Fig. 4e). For example, *HSPD1* encodes a mitochondrial protein of the chaperonin family that is essential for the folding and assembly of newly imported polypeptides [44]. ATSS, splice junctions, polyadenylation positions and increased expression of *HSPD1* gene were all observed in both primary and metastatic tumor transcriptomes (Fig. 4f). Corroborating the important role of proteome homeostasis, members within the chaperone [45], ubiquitin-proteasome [46, 47], and unfolded protein response systems [48], responsible for counteracting proteostatic stress, commonly underwent genetic and transcriptomic regulation in the sequenced case of EOC (Fig. 4g).

Therapeutically targeting proteostasis inhibits ovarian cancer

In addition to proteostatic regulators, an in silico query of the Cancer Drivers Actionability Database [49] pinpointed a number of neoplasm-specific targetable genomic or transcriptomic alterations informed by integrative sequencing analyses (Fig. 5a). To complement the molecular interrogation of primary EOC samples from clinic, we experimentally explored possible intervention opportunities by conducting an unbiased high-throughput drug screen in ovarian cancer cell lines. Multiple models, including ES-2, OVTOKO, OVCA420 and OVCAR3, were assayed against a library of clinically relevant inhibitors and the effect of each compound was quantified by a high-content imaging-based viability test (Supplementary Table S18). The putative targets of top-ranked inhibitors, with IC50 < 0.5 μM across all four lines, participated in multiple crucial cell-regulatory processes, spanning epigenetic modification (HDAC), cell cycle (PLK, CDK, Aurora kinase) and most

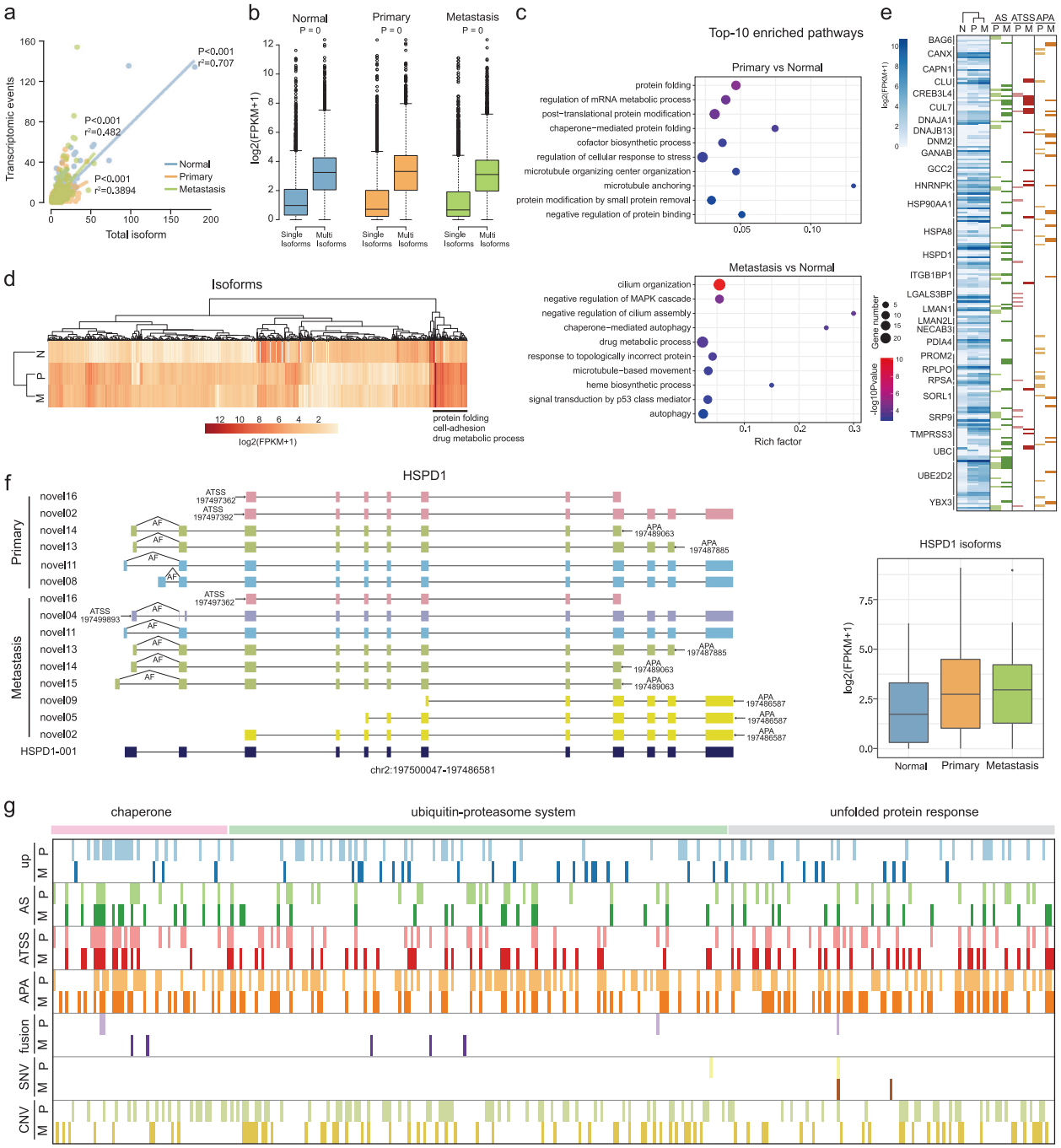


Fig. 4 Aberrant gene products and proteostatic stress. **a** Correlation analysis (two-tailed Pearson correlation test) on the numbers of transcriptomic events (ATSS, AS and APA) and detected numbers of isoforms for normal ovarian tissue, primary tumor, and distal metastasis. **b** Expression of genes with multiple isoforms was compared with those with a single isoform in normal tissue, primary and metastatic tumors (unpaired Student's *t* test, two-sided). **c** Pathway analysis of top genes with more transcriptomic events (ATSS, AS and APA) in ovarian tumors as compared to normal tissues. Top-ten-ranked pathways were presented. **d** Unsupervised hierarchical clustering of isoform expression in normal tissue and ovarian tumors. The highlighted cluster of isoforms showing increased expression participated in protein folding, cell adhesion, and drug metabolic process. N Normal tissue, P Primary tumor, M Metastatic tumor. **e** Isoform-level alterations of selected proteostatic regulators in the pathway analysis. N Normal tissue, P Primary tumor, M Metastatic tumor. **f** Left, representative isoform structures of *HSPD1* in primary and metastatic tumors. Boxes: exon; lines: intron. AF alternative first exon, ATSS alternative transcription start sites, APA alternative polyadenylation sites. Right, comparison of expression levels of *HSPD1* isoforms in normal ovarian tissue, primary and metastatic tumors. **g** Identified somatic genetic and transcriptomic aberrations in genes involved in proteostatic stress regulation. P Primary tumor, M Metastatic tumor

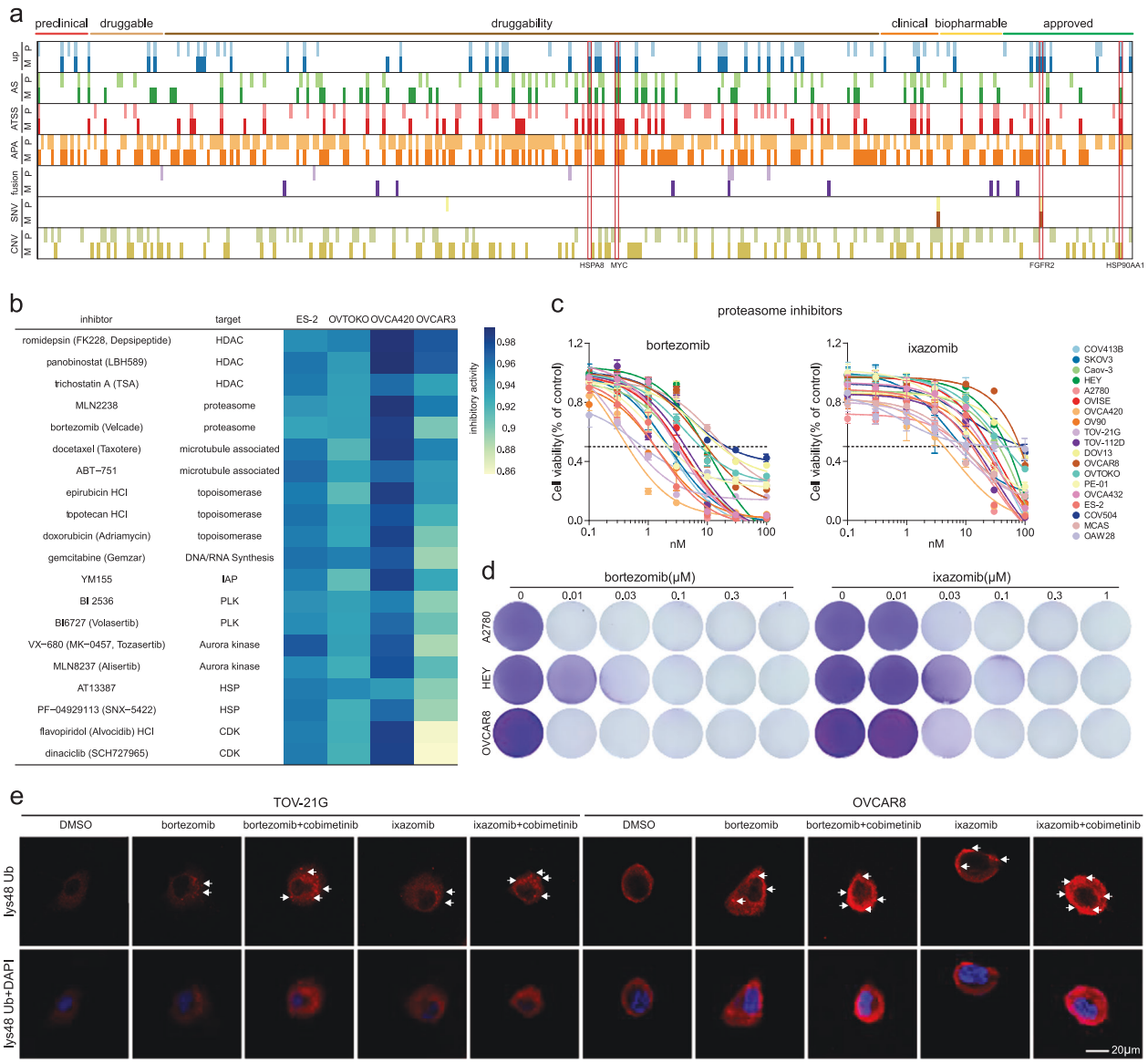


Fig. 5 Therapeutically targeting proteostasis inhibits ovarian cancer. **a** Classification of genetic and transcriptomic aberrations with potentially druggable implications according to the Cancer Drivers Actionability Database. P Primary tumor, M Metastatic tumor. **b** Heatmap of top-20-ranked inhibitors suppressing ovarian cancer cell growth and their corresponding targets. **c** Cell viability of 19 distinct EOC cell lines treated with various concentrations of bortezomib or ixazomib. **d** Crystal violet staining of A2780, HEY and OVCAR8 treated with bortezomib or ixazomib at indicated concentrations for 3 days. **e** TOV-21G and OVCAR8 were treated with 10 nM bortezomib or 100 nM ixazomib, in the presence or absence of 1 μM cobimetinib, and stained with Lys48-specific ubiquitin antibodies. White arrows designated ubiquitin-positive aggregates. Scale bar: 20 μm

prominently, proteostasis (proteasome, HSP) (Fig. 5b). Of note, proteasome inhibition was known to promote apoptosis in both hematologic and solid tumor malignancies, and had proved to be an effective anticancer therapeutic approach [50]. To further validate targeting protein homeostasis as a treatment option for ovarian cancer, we extended our findings in drug screen to a sizable collection of 19 distinct EOC cell lines and evaluated the cytotoxic activity of two FDA-approved proteasome inhibitors, bortezomib and ixazomib. Both compounds showed consistent evidence of potent efficacy toward ovarian cancer cells, as demonstrated by dose−response curves (Fig. 5c) and crystal violet staining (Fig. 5d). We probed the mechanism of action underlying proteasome inhibitors and found that ubiquitin-positive cytoplasmic aggregates were markedly induced upon bortezomib or ixazomib exposure, reinforcing the importance of protein quality-control machinery in ovarian tumor cells. In accordance with previous reports implicating MAPK signaling in guarding proteomic integrity [51], cobimetinib cotreatment largely augmented punctate foci

formation (Fig. 5e). Collectively, we concluded that pharmaceutically interfering cellular proteostasis might serve as a valid strategy to control ovarian cancer.

Discussion

We have presented a proof-of-concept study illustrating the unique power of hybrid second- and third-generation sequencing technologies in the clinical oncology setting, as well as the complementary utility of exhaustive transcriptomic and functional data in uncovering rational therapeutics for EOC. Through this innovative approach, we portrayed the first personal full-length transcriptome of metastatic ovarian carcinoma, unveiled proteostatic stress as a susceptible feature, and verified the preclinical effectiveness of proteasome inhibitors against EOC. Future application of the method described here to a larger patient cohort and various other tumor types will most likely facilitate neoplastic disease understanding and personalized cancer medicine.

The PacBio SMRT platform, by directly sequencing full-length transcripts, offers novel insights into human transcriptome complexity. An increasing number of novel isoforms in human tissues and cell lines have been recently described on the basis of Iso-Seq analysis [21–24, 26, 27, 52]. However, to the best of our knowledge, this work provides for the first time a global long-read inventory of different RNA species and alternative transcriptional events within matched normal, primary and metastatic ovarian lesions, which can be used to define aberrant gene products associated with tumorigenicity and guide therapeutic development targeting cancer-addicted pathways. In summary, integrating whole-genome and hybrid transcriptome sequencing data, we identified 7053 and 6852 genomic aberrations, 1966 and 1841 ATSS, 3239 and 3104 AS, 4608 and 3284 APA, 177 and 112 fusion transcripts, 7189 and 6328 differentially expressed genes, 2083 and 1522 differentially expressed isoforms, in primary tumor and distal metastasis as compared to normal ovarian tissue, respectively. Whether these multifaceted processes are coordinated to shape ovarian oncogenesis remains to be elucidated. Nevertheless, the massive genome-wide information extracted by our analytic pipeline clearly shows the potential of comprehensive molecular characterization to systematically survey somatic and transcriptional alterations in a given individual, to expand the diversity of candidate cancer drivers, and to refine the evolutionary model underpinning malignant progression. In line with the latter notion, phylogenetic reconstruction of metastatic ovarian cancer assisted by interrelated genomic and transcriptomic profiles reproduced detailed spatiotemporal patterns of tumor spreading and revealed early transcriptional divergence preceding genetic branching. It is reasonable to speculate that phenotypic plasticity, rather than genotypic heterogeneity, might act as a driving force of ovarian cancer dissemination, in contrast with well-studied scenario of hematogenous metastasis [53].

The delicate cancer transcriptome anatomized by our hybrid sequencing effort emphasizes the necessity of investigating gene regulation and function at the isoform level, since the vast majority of detected transcripts contain multiple variants as a result of transcriptional or post-transcriptional regulatory mechanisms. The consequent expansive protein multiplicity derived from multi-isoform genes conceivably elicits proteotoxic insults in neoplastic cells. More importantly, we discovered that myriad regulators of proteome homeostasis, such as heat-shock factors, unfolded protein responders and proteasome subunits, were rigorously modulated at the genetic, transcriptomic and expression levels in epithelial ovarian carcinoma. Therefore, these findings add further rationale that excessive proteostatic stress may represent an intrinsic property of EOC and can be exploited as a therapeutic vulnerability for cancer management. Indeed, our high-throughput drug screen and experimental validation established that ovarian tumor cells were generally sensitive to proteasome perturbation, substantiating major conclusions of previous reports [54]. It is noteworthy that a whole array of proteasome inhibitors, including tested bortezomib and orally available ixazomib, have been approved in multiple myeloma [55], and hence could be considered for repurposing opportunities to treat EOC patients. Encouragingly, initial clinical trials assessing intraperitoneal bortezomib in conjunction with carboplatin have yielded favorable tolerability and efficacy outcome in recurrent ovarian cancer population [56, 57].

In conclusion, hybrid personal transcriptomic analysis that combines short-read and full-length sequencing is feasible in clinical practice and found to be a powerful tool for depicting the entire breadth of alternative transcription start, splice, polyadenylation and fusion incidents. The resultant multidimensional high-resolution dataset forms a valuable resource to identify molecular abnormalities and determine disease courses in EOC and presumably other human malignancies. Isoform-specific variations implicate proteostatic stress as a consequence of transcriptome diversification during tumorigenesis and highlight that proteasome inhibition can point to a promising therapeutic avenue to improve survival rates. Last but not least, the bioinformatics and experimental approaches presented here will help inform the retrospective and prospective studies of larger series, which may ultimately lead to more tailored care for cancer patients.

Materials and methods

Patient cohort

The study was conducted in accordance with ethical guidelines of U.S. Common Rule and approved by the Ren Ji Hospital Ethics Committee. Written informed consent was obtained from the patient in this study. Primary tumor and distal metastatic samples (the proportion of tumor cell nuclei (>50%) and necrosis (<30%)) for sequencing were collected from a 63-year-old woman diagnosed with pathologically confirmed metastatic serous carcinoma during laparoscopic debulking surgery at Ren Ji Hospital. Fresh-frozen primary and metastatic tumors with matched blood were subjected to WGS. Fresh-frozen primary and metastatic tumors with paired normal tissue were subjected to RNA sequencing and full-length transcriptome sequencing.

Genomic DNA and total RNA preparation

DNA of primary and metastatic tumor was extracted from tissue shavings of fresh-frozen specimens using QIAamp DNA Tissue Kit. Paired blood cell DNA was extracted following instructions using QIAamp DNA Blood Mini Kit. DNA was quantified by Qubit (Life Technologies) and DNA integrity was examined by agarose gel electrophoresis. Total RNA of primary tumor, metastatic tumor and paired normal tissue was extracted from shavings of fresh-frozen specimens using RNeasy Plus Kit (Qiagen) according to the manufacturer's protocol. RNA purity and integrity were assessed by the NanoPhotometer spectrophotometer (IMPLEN) and RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies), respectively. Total RNAs with RNA Integrity Number (RIN) of >8 were subjected to full-length transcriptome sequencing.

Whole-genome sequencing

Extracted genomic DNA (1 μg) was used as input for the DNA library preparation. Sequencing library was prepared using Truseq Nano DNA HT Sample Prep Kit (Illumina) following the manufacturer's instructions. Briefly, genomic DNA sample was fragmented, end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing, followed by further PCR amplification. The clustering of samples was performed on a cBot Cluster Generation System using Hiseq PE Cluster Kit (Illumina) according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced on Illumina X Ten sequencing platform and 150 bp paired-end reads were generated (Novogene, Beijing). The targeted sequencing depth was 50×.

Sequence alignment and variant calling

Clean reads in FastQ format generated by the Illumina platform were aligned to UCSC hg19 human genome by Burrows−Wheeler Aligner (BWA) [58]. SAMtools, Picard (http://broadinstitute.github.io/picard/), and GATK were used for filtering, local realignment, and base quality recalibration to generate final BAM files for computation of the sequence coverage and depth [59, 60]. Somatic SNVs were detected by MuTect [61]. Somatic INDELs were detected by GATK Somatic Indel Detector. ANNOVAR was performed to do annotation for VCF (Variant Call Format) [62]. Variants obtained from previous steps were compared against SNPs present in the dbSNP and 1000 Genomes databases (1000 Genomes Project Consortium) to discard known SNPs. The retained nonsynonymous SNVs were submitted to PolyPhen and SIFT for functional prediction [63, 64]. Control-FREEC was utilized to detect somatic copy number variations [65]. Breakdancer was implemented to identify potential structural variants [66].

RNA sequencing and analysis

Total amount of 3 μg RNA for each sample was used as input materials for the RNA library preparation. Sequencing libraries were generated using the rRNA-depleted RNA with NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). The index-coded libraries were clustered on a cBot Cluster Generation System using TreSeq PE Cluster Kit v3-cBot-HS (Illumina) and sequenced on an Illumina Hiseq X Ten platform to generate 125 bp paired-end reads (Novogene, Beijing). Clean data were obtained from FastQ raw data by removing adapter, ploy-N sequences and low-quality reads. All the downstream analyses were based on the clean data with high quality. Index of the reference genome was built using Bowtie v2.0.6 and paired-end clean reads were aligned to the reference genome (Ensembl hg38 human genome) using TopHat v2.0.9 [67]. The mapped reads of each sample were assembled by Cufflinks (v2.1.1) [68] in a reference-based approach. Differential expression analysis and FPKM (fragments per kilo-base of exon per million fragments) of both lncRNAs and transcripts of genes in each sample were calculated using Cuffdiff (v2.1.1) [68]. Transcripts with P values of <0.05, $q < 0.1$ and absolute value of log2 fold change of >1 were assigned as differentially expressed transcripts. The counts of mapped reads of genes were calculated by featureCounts (v1.6.1) [69]. Differential expression analysis and FPKM (fragments per kilo-base of exon per million fragments) calculation of genes in each sample were performed by edgeR [70]. Genes with P values of <0.05, FDR <0.1 and absolute value of log2 fold change of >1 were assigned as differentially expressed genes. SOAPfuse was

implemented to identify fusion genes [71]. AS was analyzed by SUPPA (v2.2.1) [72]. APA and ATSS were analyzed by DarPars (v0.9.1) [73].

Full-length transcriptome sequencing and analysis

Five micrograms total RNA was used for polyA+ selection using oligo(dT)25-coated Dynabeads (ThermoFisher Scientific). The Isoform Sequencing (Iso-Seq) library was prepared using polyA+ RNA. According to the Iso-Seq library protocol, first-strand cDNAs were generated and amplified by Clontech SMARTer PCR cDNA Synthesis Kit. Large-scale PCR products were purified with AMPure PB beads (Beckman Coulter). cDNA products were then size-selected using the BluePippin system (Sage Science) into three separate bins: 1–2 kb, 2–3 kb, and 3–10 kb. Three libraries were constructed with the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). Every library was sequenced on the PacBio RS II small-molecule real-time (SMRT) sequencing platform by two SMRT cells. Raw sequence reads from PacBio RS II was converted to circular consensus (CCS) reads by Permissive CCS 2. All three size fractions from one sample were pooled and analyzed with the SMRT link 4.0 to obtain polished consensus reads, using --minLength = 200, --minReadScore = 0.75, --minPasses = 1, --minPredictedAccuracy = 0.8. The CCS reads were classified into full-length nonchimeric and non-full-length nonchimeric reads. Only CCS reads contained 5′ primer, 3′ primer and polyA tail but not primer sequences were classified as full-length non-chimeric reads. Polished consensus reads were generated by clustering and polishing full-length nonchimeric reads. Next, polished consensus reads were corrected by Illumina clean reads (aligned with hisat2 (v2.1.0) [74]) using proovread (v2.13.12) [75] to generate consensus reads. The analysis of transcriptome aberrations was based on consensus reads. Consensus reads were mapped to the Ensembl human hg38 genome using GMAP [76] with the following option: --no-chimeras -K 50000 -n 1 --expand-offsets 1. Isoform generation and annotation, novel gene and novel isoform determination, ATSS and APA sites analysis were completed using default parameters of TAPIS (Transcriptome Analysis Pipeline for Isoform Sequencing) [21]. We used CNCI [77], CPC [78], Pfam-scan [79] and PLEK [80] to predict the coding potential of isoforms, and those predicted to have no coding potential by these four tools were considered as lncRNAs. AS was analyzed by SUPPA (v2.2.1) [72]. Fusion transcripts were identified following three criteria: (1) at least 10% of consensus read sequences were aligned to two or more locations of >100 kb distance in human genome (GMAP options: --no-chimeras -K 50000 -n 0 --expand-offsets 1); (2) the coverage of alignment was at least 99%; (3) at least two Illumina junction reads supported the fusion events. Differential expression analysis and FPKM calculation of identified isoforms in each sample were performed using Cuffdiff (v2.1.1) [68]. Isoforms with P values of <0.05, $q < 0.1$ and absolute values of log2 fold change of >1 were assigned as differentially expressed isoforms. The counts of mapped reads of identified genes were calculated by featureCounts (v1.6.1) [69]. Differential expression analysis and FPKM calculation of genes in each sample were performed using edgeR [70]. Genes with P values of <0.05, FDR < 0.1 and absolute values of log2 fold change of >1 were assigned as differentially expressed genes.

Bioinformatics and statistical analysis

The sequencing data have been deposited in NCBI Sequence Read Archive (SRA) under the accession number SRP137947. All somatic SNVs were included to calculate relative weights of mutational signatures in a given sample. The R package "deconstructSigs", based on the Wellcome Trust Sanger Institute Mutational Signature Framework, was used to statistically quantify the contribution of each signature for each tumor [81]. SciClone R package and PyClone were employed to detect subclonality of a given sample [82, 83]. For each variant, the CCF was defined as $\mathrm{VAF} = \frac{\alpha \times \mathrm{CCF}}{\mathrm{CT} \times \alpha + \mathrm{CN}(1-\alpha)}$, where CT is the copy number of the tumor, CN is the copy number of the matched normal sample and α is the tumor purity. Tumor purity and copy number were determined by ABSOLUTE [84]. Variant Allele Frequency (VAF) was defined with respect to the number of reads supporting the variant allele (x^{var}) and the number of reads supporting the reference allele (x^{ref}): $\mathrm{VAF} = \frac{x^{\mathrm{var}}}{x^{\mathrm{var}} + x^{\mathrm{ref}}}$. Pathway enrichment was analyzed in Metascape website (http://metascape.org/gp/index.html#/main/step1). Statistical analysis was performed with GraphPad Prism 5.0 software. In all experiments, comparisons between two groups were based on unpaired two-sided Student's t test. P values of <0.05 were considered statistically significant.

Phylogenetic tree reconstruction

The input data of phylogenetic tree reconstruction were somatic and exonic SNVs, somatic ATSS, somatic AS, somatic APA and somatic fusion transcripts. Somatic ATSS, AS, APA and fusion transcripts were calculated by subtraction of events that occurred in the tumor sample with those in the normal sample. Phylogenetic tree construction was performed using PHYLIP version 3.695.

Cell culture and reagents

Tumor cell lines were originally obtained from ATCC or JCRB in 2014, where cell characterization was performed

using polymorphic short tandem repeat profiling. The cells were regularly examined to ensure that they were free of mycoplasma contamination. Cells were cultured in RPIM1640 (Life Technologies) supplemented with 10% fetal bovine serum (Millipore). Bortezomib, ixazomib, cobimetinib and inhibitors library were purchased from Selleck Chemicals. All inhibitors were reconstituted in Dimethyl Sulfoxide (DMSO) (Sigma-Aldrich) at a stock concentration of 10 mM. Cell viability assays were performed after 30,000 cells were seeded in six-well plates and treated with a serial dilution of inhibitors for 3 days. Cells were fixed with formalin and stained with 0.5% crystal violet.

High-throughput drug screen and validation

Drug screening was performed in a 96-well format. Cells were seeded at optimal density and treated with the indicated inhibitors at the same concentration (500 nM). Fresh medium and drugs were changed every 3 days. After 6 days of drug exposure, cells were imaged and viability was calculated using ArrayScan Infinity (Thermo Scientific). To quantify the IC50 of bortezomib and ixazomib, seven concentrations of compounds were applied at a stepwise threefold dilution series. Cell viability was evaluated using CellTiter-Glo reagent according to the manufacturer's instructions (Promega). Estimates of IC50 were derived from the seven dose−response curves plotted by GraphPad Prism 5.

Immunofluorescence staining

After treated with indicated inhibitors for 24 h, cells were fixed with 4% formaldehyde, blocked with 5% goat serum in Phosphate Buffer Saline (PBS) containing 0.3% Triton X-100 at room temperature for 1 h, and incubated with Lys48-specific ubiquitin antibodies (Millipore, 05-1307) overnight. Cells were washed three times with Tris Buffered Saline Tween (TBST) and incubated with Alexa Fluor 594 goat anti-rabbit IgG Abs (Life Technologies), followed by 4,6-Diamino-2-Phenyl Indole (DAPI) (Sigma) nuclear staining. Fluorescent images were captured using a Zeiss confocal microscope.

Acknowledgements This work was supported by the National Natural Science Foundation of China (81472537 and 81672714 to GZ; 81502597 to YJ; 81472426 to WD), the Grants from the State Key Laboratory of Oncogenes and Related Genes (SB17-06 to M-CC), the grants from Shanghai Jiao Tong University School of Medicine (DLY201505 to WD; YG2016MS51 to XY), Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20161313 to GZ), the Shanghai Institutions of Higher Learning (Eastern Scholar to GZ), Shanghai Rising-Star Program (16QA1403600 to GZ), Shanghai Municipal Commission of Health and Family Planning (2013ZYJB0202 and 15GWZK0701 to WD; 20174Y0189 to YJ; 20174Y0043 to M-CC), the grant from Shanghai Key Laboratory of Gynecologic Oncology (FKZL-2017-01 to YJ), and the grant from Science and Technology Commission of Shanghai Municipality (16140904401 to XY).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Mills K, Fuh K. Recent advances in understanding, diagnosing, and treating ovarian cancer. F1000Res. 2017;6:84.
2. Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. Nat Rev Dis Prim. 2016;2:16061.
3. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.
4. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015;521:489–94.
5. Wang YK, Bashashati A, Anglesio MS, Cochrane DR, Grewal DS, Ha G, et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. Nat Genet. 2017;49:856–65.
6. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45:1127–33.
7. Zhang Z, Ma P, Jing Y, Yan Y, Cai MC, Zhang M, et al. BET bromodomain inhibition as a therapeutic strategy in ovarian cancer by downregulating FoxM1. Theranostics. 2016;6:219–30.
8. Zhang Z, Peng H, Wang X, Yin X, Ma P, Jing Y, et al. Preclinical efficacy and molecular mechanism of targeting CDK7-dependent transcriptional addiction in ovarian cancer. Mol Cancer Ther. 2017;16:1739–50.
9. Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. Nat Biotechnol. 2018;36:46–60.
10. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Methods. 2011;8:61–65.
11. Niu B, Fu L, Sun S, Li W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinforma. 2010;11:187.
12. Steijger T, Abril JF, Engstrom PG, Kokocinski F, Consortium R, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013;10:1177–84.
13. Cieslik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2018;19:93–109.
14. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12:87–98.
15. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.
16. Rhoads A, Au KF. PacBio sequencing and its applications. Genom Proteom Bioinforma. 2015;13:278–89.
17. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. Nat Biotechnol. 2013;31:1009–14.
18. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol. 2013;14:405.

19. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010;38:e159.
20. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol. 2012;30:693–700.
21. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, et al. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. 2016;7:11706.
22. Au KF, Sebastiano V, Afshar PT, Durruthy JD, Lee L, Williams BA, et al. Characterization of the human ESC transcriptome by hybrid sequencing. Proc Natl Acad Sci USA. 2013;110:E4821–4830.
23. Kuang Z, Boeke JD, Canzar S. The dynamic landscape of fission yeast meiosis alternative-splice isoforms. Genome Res. 2017;27:145–56.
24. Li S, Yamada M, Han X, Ohler U, Benfey PN. High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. Dev Cell. 2016;39:508–22.
25. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 2017;18:18–30.
26. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun. 2016;7:11708.
27. Weirather JL, Afshar PT, Clark TA, Tseng E, Powers LS, Underwood JG, et al. Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. Nucleic Acids Res. 2015;43:e116.
28. Lengyel E. Ovarian cancer development and metastasis. Am J Pathol. 2010;177:1053–64.
29. Consortium F, the RP, CLST, Forrest AR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. Nature. 2014;507:462–70.
30. de Klerk E, t Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. 2015;31:128–39.
31. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol. 2013;14:153–65.
32. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–6.
33. Cui T, Zhang L, Huang Y, Yi Y, Tan P, Zhao Y, et al. MNDRv2.0: an updated resource of ncRNA-disease associations in mammals. Nucleic Acids Res. 2018;46:D371–D374.
34. Yi Y, Zhao Y, Li C, Zhang L, Huang H, Li Y, et al. RAID v2.0: an updated resource of RNA-associated interactions across organisms. Nucleic Acids Res. 2017;45:D115–D118.
35. Zhang T, Tan P, Wang L, Jin N, Li Y, Zhang L, et al. RNALocate: a resource for RNA subcellular localizations. Nucleic Acids Res. 2017;45:D135–D138.
36. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer. 2015;15:371–81.
37. Schram AM, Chang MT, Jonsson P, Drilon A. Fusions in solid tumours: diagnostic strategies, targeted therapy, and acquired resistance. Nat Rev Clin Oncol. 2017;14:735–48.
38. Engstrom PG, Steijger T, Sipos B, Grant GR, Kahles A, Ratsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods. 2013;10:1185–91.
39. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol. 2013;231:21–34.
40. Choi YJ, Rhee JK, Hur SY, Kim MS, Lee SH, Chung YJ, et al. Intraindividual genomic heterogeneity of high-grade serous carcinoma of the ovary and clinical utility of ascitic cancer cells for mutation profiling. J Pathol. 2017;241:57–66.
41. Hoogstraat M, de Pagter MS, Cirkel GA, van Roosmalen MJ, Harkins TT, Duran K, et al. Genomic and transcriptomic plasticity in treatment-naive ovarian cancer. Genome Res. 2014;24:200–11.
42. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet. 2016;48:758–67.
43. Yin X, Jing Y, Cai MC, Ma P, Zhang Y, Xu C, et al. Clonality, heterogeneity, and evolution of synchronous bilateral ovarian cancer. Cancer Res. 2017;77:6551–61.
44. Vilasi S, Bulone D, Caruso Bavisotto C, Campanella C, Marino Gammazza A, San Biagio PL, et al. Chaperonin of group I: oligomeric spectrum and biochemical and biological implications. Front Mol Biosci. 2017;4:99.
45. Chatterjee S, Burns TF. Targeting Heat Shock Proteins in Cancer: A Promising Therapeutic Approach. Int J Mol Sci. 2017;18:1978
46. Fulda S, Rajalingam K, Dikic I. Ubiquitylation in immune disorders and cancer: from molecular mechanisms to therapeutic implications. EMBO Mol Med. 2012;4:545–56.
47. Nandi D, Tahiliani P, Kumar A, Chandu D. The ubiquitin-proteasome system. J Biosci. 2006;31:137–55.
48. Faoro R, Bassu M, Mejia YX, Stephan T, Dudani N, Boeker C, et al. Aberration-corrected cryoimmersion light microscopy. Proc Natl Acad Sci USA. 2018;115:1204–9.
49. Rubio-Perez C, Tamborero D, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell. 2015;27:382–96.
50. Manasanch EE, Orlowski RZ. Proteasome inhibitors in cancer therapy. Nat Rev Clin Oncol. 2017;14:417–33.
51. Tang Z, Dai S, He Y, Doty RA, Shultz LD, Sampson SB, et al. MEK guards proteome stability and inhibits tumor-suppressive amyloidogenesis via HSF1. Cell. 2015;160:729–44.
52. Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single-molecule long-read transcriptome. Proc Natl Acad Sci USA. 2014;111:9869–74.
53. Turajlic S, Swanton C. Metastasis as an evolutionary process. Science. 2016;352:169–75.
54. Bazzaro M, Lee MK, Zoso A, Stirling WL, Santillan A, Shih Ie M, et al. Ubiquitin-proteasome system stress sensitizes ovarian cancer to proteasome inhibitor-induced apoptosis. Cancer Res. 2006;66:3754–63.
55. Kumar SK. New treatment options for the management of multiple myeloma. J Natl Compr Canc Netw. 2017;15:709–12.
56. Jandial DA, Brady WE, Howell SB, Lankes HA, Schilder RJ, Beumer JH, et al. A phase I pharmacokinetic study of intraperitoneal bortezomib and carboplatin in patients with persistent or recurrent ovarian cancer: an NRG Oncology/Gynecologic Oncology Group study. Gynecol Oncol. 2017;145:236–42.
57. Ramirez PT, Landen CN Jr., Coleman RL, Milam MR, Levenback C, Johnston TA, et al. Phase I trial of the proteasome inhibitor bortezomib in combination with carboplatin in patients with platinum- and taxane-resistant ovarian cancer. Gynecol Oncol. 2008;108:68–71.
58. Li H, Durbin R. Fast and accurate short read alignment with Burrows−Wheeler transform. Bioinformatics. 2009;25:1754–60.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce

framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
61. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.
62. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
63. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–4.
64. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Human Genet. 2013;Chapter 7:Unit7 20.
65. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28:423–5.
66. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009;6:677–81.
67. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
68. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
69. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
70. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
71. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 2013;14:R12.
72. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. 2018;19:40.
73. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nat Commun. 2014;5:5274.
74. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
75. Hackl T, Hedrich R, Schultz J, Forster F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014;30:3004–11.
76. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. Methods Mol Biol. 2016;1418:283–334.
77. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41:e166.
78. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35:W345–349.
79. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–285.
80. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinforma. 2014;15:311.
81. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
82. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol. 2014;10:e1003665.
83. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods. 2014;11:396–8.
84. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012;30:413–21.