Gynecologic Oncology 152 (2019) 194–201

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Review Article

RNA sequencing (RNA-Seq) and its application in ovarian cancer

Jinglu Wang [a,b], Dylan C. Dean [b], Francis J. Hornicek [b], Huirong Shi [a,*], Zhenfeng Duan [a,b,**]

[a] Department of Obstetrics and Gynecology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China
[b] Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

HIGHLIGHTS

- Absence of biomarkers in early detection and drug resistance are principal causes of treatment failure in ovarian cancer.
- RNA-Seq has expanded the knowledge of molecular tumorigenesis and novel transcriptional changes in ovarian cancer.
- RNA-Seq has revealed novel biomarkers and advanced the areas of ovarian cancer diagnostics and classification.

ARTICLE INFO

Article history:
Received 31 July 2018
Received in revised form 29 September 2018
Accepted 1 October 2018
Available online 5 October 2018

Keywords:
RNA-Seq
Ovarian cancer
Early detection
Drug resistance
Non-coding RNA

ABSTRACT

Despite the surgical and chemotherapeutic advances over the past few decades, ovarian cancer remains the leading cause of gynecological cancer-related mortality. The absence of biomarkers in early detection and the development of drug resistance are principal causes of treatment failure in ovarian cancer. Recent progress in RNA sequencing (RNA-Seq) with Next Generation Sequencing technology has expanded the understanding of the molecular pathogenesis of ovarian cancer. As compared to previous hybridization-based microarray and Sanger sequence-based methods, RNA-Seq provides multiple layers of resolutions and transcriptome complexity, with less background noise and a broader dynamic range of RNA expression. Beyond quantifying gene expression, the data generated by RNA-Seq accelerates the identification of alternatively spliced genes, fusion genes, mutations/SNPs, allele-specific expression, novel transcripts and non-coding RNAs. RNA-Seq has been successfully applied in ovarian cancer research for earlier detection, ascertaining pathological origin, and defining the aberrant genes and dysregulated molecular pathways across patient groups. This review outlines the distinct advantages of RNA-Seq compared to other transcriptomics methods and its recent applications in ovarian cancer.

Contents

* Correspondence to: H. Shi, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Zhengzhou University, 1 Jianshe East Road, Zhengzhou, Henan 450052, China.
** Correspondence to: Z. Duan, Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, 615 Charles E. Young. Dr. South, Los Angeles, CA 90095, USA.
E-mail addresses: hrshi56@126.com (H. Shi), zduan@mednet.ucla.edu (Z. Duan).

https://doi.org/10.1016/j.ygyno.2018.10.002
0090-8258/

1. Introduction

Ovarian cancer is the deadliest gynecologic cancer and ranks as the fifth leading cause of cancer-related death among women in the United States. The American Cancer Society estimates that approximately 22,240 new ovarian cancer cases and 14,070 deaths will occur in the United States in 2018 alone [1]. Current standard treatment for ovarian cancer patients consists of primary maximal debulking surgery followed by adjuvant combination platinum-taxane based chemotherapy [2]. Due to the absence of specific symptoms at early stages and lack of approved population screening tools, >60% of ovarian cancer patients already have advanced stage III and IV upon diagnosis [1]. When epithelial ovarian cancer is detected in the early stages, the 5-year survival rate is >70%. However, the 5-year survival rate for advanced ovarian cancer patients drops to 29% [1]. Therefore, early detection is critical for achieving long-term disease control and mortality reduction. The high mortality rate associated with ovarian cancer is not only due to challenges in early stage diagnosis, however, but also a result of the development of multidrug resistance (MDR). More than half of advanced ovarian cancer patients relapse within the first 5 years and acquire resistance to both standard chemotherapy as well as cross-resistance to other functionally and structurally unrelated chemotherapeutic drugs [3]. Intrinsic or acquired MDR to chemotherapy severely limits the long-term success of chemotherapy in ovarian cancer treatment. At present, several altered gene expression profiles and mutations contributing to MDR have been identified, such as TP53 or BRCA1/2 mutations and MDR1 or CCNE1 amplifications [4]. In addition to these well-known aberrations, there are many genetic and epigenetic alterations responsible for MDR which can now be detected by Next Generation Sequencing (NGS) [5]. On the basis of this technology, identifying new biomarkers may shorten the detection period and elucidate the molecular mechanisms of MDR, therefore enabling the generation of novel therapies while improving outcomes for ovarian cancer patients.

The transcriptome contains coding messenger RNAs (mRNAs) as well as noncoding RNAs (ncRNAs) such as microRNA (miRNA, miR), long noncoding RNA (lncRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). Although several methodologies have been developed to characterize these RNAs in human disease, NGS has emerged as the central platform for the thorough and unbiased profiling of cancer genomes and transcriptomes [6]. RNA sequencing (RNA-Seq) is a high-throughput technology used to provide a comprehensive view of the entire transcriptome, including isoform and gene fusion detection, gene expression profiling, targeted sequencing and single-cell analysis [7]. The term high-throughput refers to the massive parallel sequencing of hundreds of millions of DNA/RNA strands simultaneously, producing a rich dataset in less time [8]. RNA-Seq facilitates the ability to identify new genes, allele-specific expression, fusion genes, disease-associated single nucleotide polymorphisms (SNP), post-transcriptional modifications, ncRNAs and differential gene expression across different groups or treatments [9]. Based on the RNA-Seq technology, the Cancer Genome Atlas (TCGA) project generates a comprehensive transcriptome profile of 33 cancer types (including ovarian cancer) to identify potential therapeutic targets for personalized therapy [10–12]. Similarly, the Gene Expression Omnibus (GEO) is a public database that archives and distributes microarray, NGS and other forms of high-throughput functional genomic data sets [13]. The goal of these projects is to provide open access datasets for clinicians and scholars to advance diagnostic methods, treatment standards, and ultimately prevent cancer [14]. RNA-Seq technology has emerged as a powerful tool for identifying functional genes and pathways in anti-tumor research. In this review, we discuss the rationale, methods, advantages and translational applications of RNA-Seq in ovarian cancer.

2. RNA-Seq technology and its advantages

RNA-Seq provides a more detailed and unbiased view of the complex and dynamic nature of the transcriptome. RNA-Seq workflow is divided into several steps: RNA extraction, library construction, sequencing, and data analysis [15]. The first step in transcriptome sequencing is isolation of high-quality RNA from solid tissue or cancer cells. The total RNA content is composed of rRNA, mRNA and various categories of ncRNA [15]. Removal of rRNA is crucial for precise library construction because it accounts for >95% of the total cellular RNA and is irrelevant for prognostics [15,16]. This has led to the development of several separation techniques. The 3′ polyadenylated (poly-A) tail of mRNA may be extracted using oligo-dT primer beads by selecting for poly-A RNAs (termed a poly-A library). A portion of the lncRNAs is excluded from the poly-A library due to the absence of a poly-A tail [17]. Alternatively, rRNA can be removed using commercially available kits, such as Ribo-Zero and RiboMinus [16]. Specific extraction kits have been developed to selectively isolate small RNAs, including miRNA and piwi-interacting RNA (piRNA), which are short (15–30 nt), sparse, and lack a poly-A tail [15]. After the targeted/fractionated RNA is isolated from the total RNA content, it is converted to a library of complementary DNA (cDNA) fragments with adaptors ligated to one or both ends [18]. These adaptors can directly ligate to the small RNAs and subsequently undergo reverse transcription. However, long RNA or cDNA molecules must first be fragmented into smaller pieces (~200–250 bp) in order to be compatible with Next Generation Sequencing (NGS) technologies (Fig. 1).

Each cDNA fragment, with or without PCR amplification, is then sequenced in a high-throughput manner to obtain millions of short sequences derived from single-end sequencing or pair-end sequencing. There are variations in this sequencing workflow, however, as illustrated by the new Illumina platforms whereby cDNA libraries are amplified via cluster generation chemistry. In this process, fragments from the cDNA library are loaded onto a so-called flow cell, which consists of a glass slide with physically separated lanes coated with a lawn of surface-bound oligos complementary to the library adapters. The single-stranded adapter ligation fragment is then hybridized to the flow cell surface, with repeated denaturation and extension cycles resulting in clonal clusters by bridge amplification. When cluster generation is complete, each cluster on the flow cell produces a single sequencing read [8]. Sequencing by synthesis (SBS) is a widely adopted NGS technology which uses four fluorescently labeled nucleotides to sequence massive clusters in parallel [19,20]. During each sequencing cycle, a single labeled deoxynucleotide triphosphate (dNTP) is added to the nucleic acid chain. This nucleotide label serves as a "reversible terminator" for polymerization. After dNTP incorporation, the fluorescent dye is identified through laser excitation, revealing a specific color based on the incorporated nucleotide. After imaging, the fluorophore is cleaved and the terminator is reversed, allowing for the incorporation of the next base within the sequence. This process is repeated as necessary until the predetermined read length has been met. The reads range from 30 to 400 bp depending on which NGS technology was applied; options include Illumina HiSeq 2500 (125 bp), Applied Biosystems SOLiD (120 bp) and Roche 454 Life Science (400 bp) [9]. Following sequencing, the produced reads are aligned to a reference genome and then assembled into transcripts according to annotated reference transcripts. Next, the expression level of each gene is measured by counting the number of produced reads that align to each exon or full-length transcript. Alternatively, these reads can create a genome-scale transcriptome map without the genomic sequence using de novo assembly approaches. In addition, downstream analyses with RNA-Seq can detect differential gene expression between samples, allele-specific expression, alternatively spliced genes, and novel transcripts (Fig. 2) [15,18]. Generally, the Tuxedo Suite consisting of Bowtie, Tophat, and Cufflinks are appropriate open-source software programs for detecting differential gene expression between samples [21]. In addition, there are several other unique platforms, including DESeq, DEGseq, baySeq, and edgeR [15]. MISO and DEXseq are programs capable of mapping reads according to the individual isoform or exon, thereby identifying alternative splicing within a sample [22]. Given the variations in data analysis

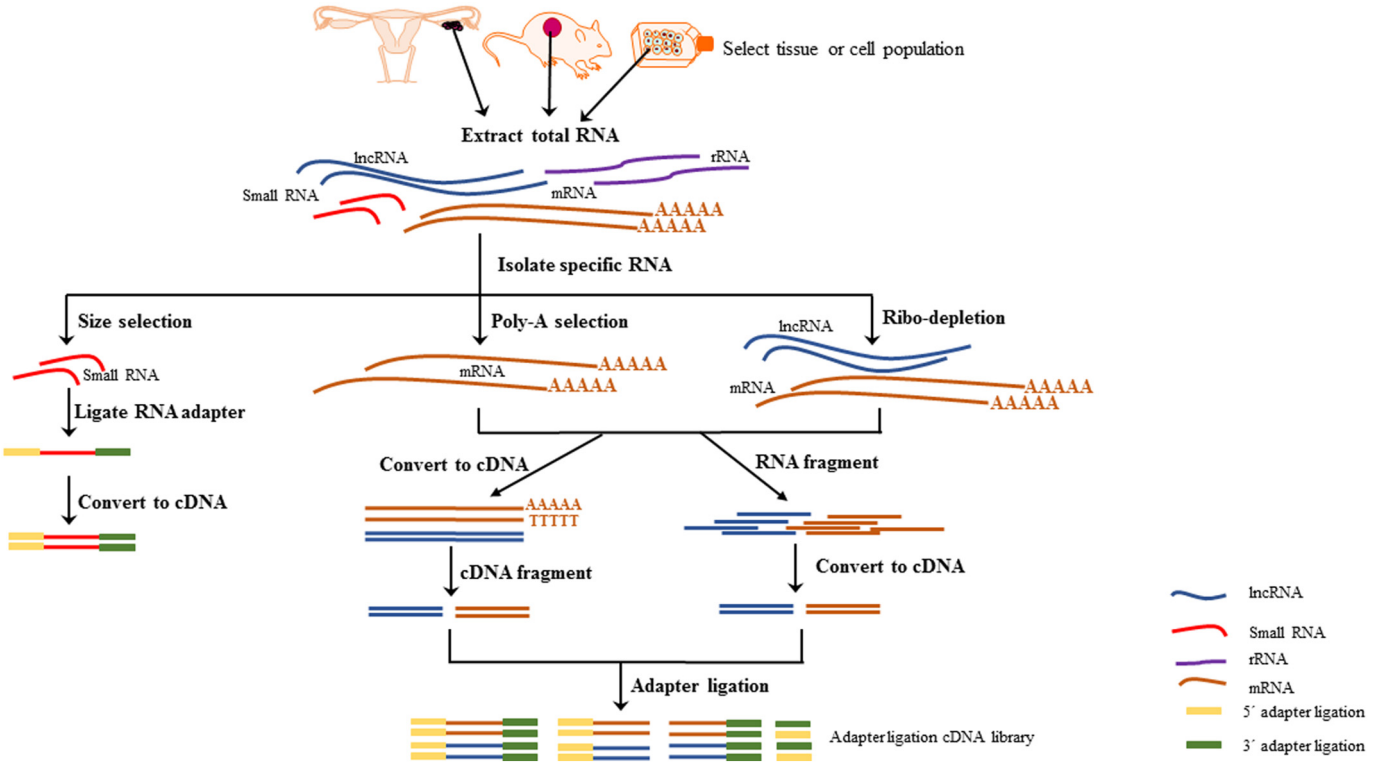

Fig. 1. Workflow of cDNA library construction. First, total RNA is extracted from ovarian tumor tissues or cancer cells. Second, subsets of RNA molecules are isolated using a specific protocol. mRNA can be extracted using poly-A selection, however, a portion of long noncoding RNAs (lncRNAs) is excluded from the poly-A library due to the absence of a poly-A tail. rRNA can be removed using the ribo-depletion method to obtain mRNA and lncRNA. Size selection has been developed to selectively isolate small RNAs, which are short (15–30 nt), sparse, and lack a poly-A tail. Next, adaptors can be directly ligated to the small RNAs, followed by reverse transcription. However, long RNA or cDNA molecules must be fragmented into smaller pieces to be compatible with NCG technologies. Finally, the RNA is converted to a library of cDNA fragments with adaptors ligated to either or both ends.

among software platforms, it is clear that choosing the most appropriate RNA-seq software, based on experimental design, is a vital component for capturing robust data.

RNA-Seq offers several distinct advantages over previous hybridization-based microarray technology (Table 1). In general, microarray-based gene expression profiling is limited to hybridizing gene probes to known genomic sequences within target genes of tumor specimens [23]. Preferably, RNA-Seq enables the detection and discovery of novel gene transcripts and ncRNAs [18]. RNA-Seq can also detect transcription boundaries with a single nucleotide resolution, reveal alternative splicing, allele-specific expression, and SNPs in the transcribed regions [15,24]. RNA-Seq has remarkably low background signaling owing to its precisely mapped sequenced reads of unique genomic regions. Microarrays, in contrast, have much higher background noise (due to cross-hybridization) and saturation of the hybridization signals, and consequently, have difficulty detecting weakly or very highly expressed genes and have a shorter range for detecting expression (one hundred-fold to a few hundred-fold) [25,26]. As there is no upper limit for sequence quantification, RNA-Seq permits a more comprehensive range of expression level detection than microarray alone (>9000-fold) [18,26]. Furthermore, comparing gene expression across experiments is often complicated for hybridization-based microarray technology, and may require dataset normalization because the signal intensity of the probe does not necessarily correlate with gene expression [25]. Sanger sequencing technology, in comparison, uses cDNA without adaptor ligation and is frequently low throughput, relatively expensive, and generally not quantitative [20,27]. Depending on the detected sequence, RNA-Seq provides a more precise and quantitative profile of gene expression in diverse groups or treatments.

There are two novel platforms for targeted genotyping, mutation analysis, and RNA-Seq validation. The Agena MassARRAY® System is a non-fluorescent high-throughput platform utilizing mass spectrometry to accurately measure PCR-derived amplicons [28]. The MassARRAY platform provides a highly cost-effective workflow for determining the expression of targeted genes and SNP genotypes without the need for complicated and expensive data pipelines. Another platform, the NanoString gene expression panel, can digitally detect the expression of up to 800 genes within a single reaction [29]. Briefly, extracted RNA is hybridized to both target-specific capture probes and color-coded reporter probes without converting the mRNA to cDNA. The reporter probes contain precisely ordered fluorescent barcodes for each target gene of interest. An electric current then aligns the probe-target complexes, and the expression of each gene is quantitated according to the number of times its corresponding fluorescent barcode is detected.

3. Applications of RNA-Seq in ovarian cancer

Recently, RNA-Seq based studies have expanded the knowledge of the pathological origin and molecular tumorigenesis of ovarian cancer. Analyzing gene expression and transcriptome changes also aids in the understanding of ovarian cancer classification and progression (Table 2).

3.1. Identification of the pathological origin of ovarian cancer

Ovarian tumors are divided into three clinicopathological subtypes with distinct histopathological features: epithelial, sex cord-stromal and germ cell. Of these, epithelial ovarian cancer (EOC) is the most common, representing 85–90% of ovarian cancers [30]. According to histological and morphological differences, EOC is classified into five major categories: high-grade serous, low-grade serous, mucinous, endometrioid and clear cell. High-grade serous ovarian cancer (HGSOC) is the most frequently observed and lethal histotype, causing nearly 75% of all EOC-related mortalities [30]. Historically, HGSOC was thought to originate

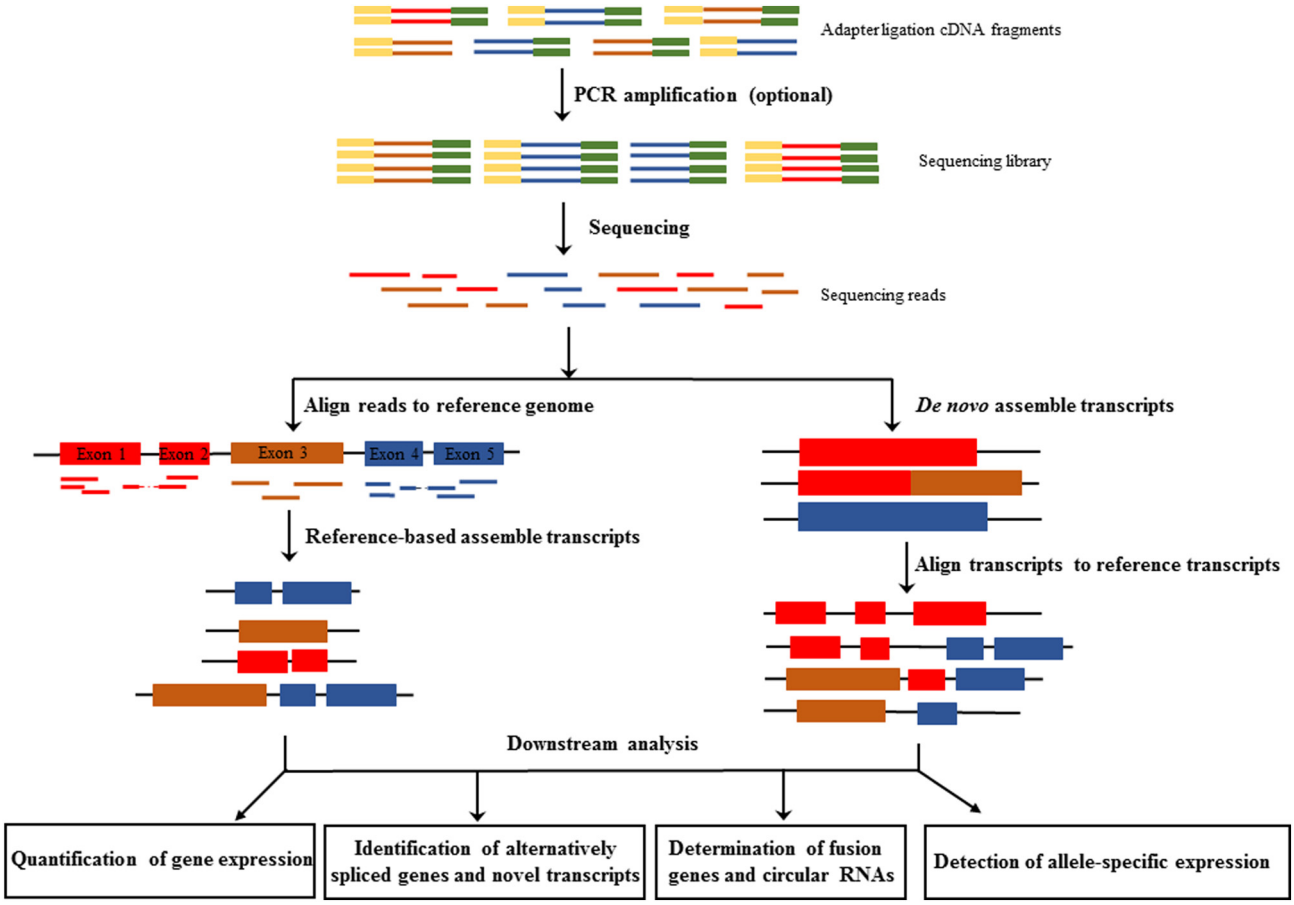

Fig. 2. Schematic of RNA-Seq technology and data analysis. After construction of adapter ligation cDNA library, each cDNA fragment, with or without PCR amplification, is then sequenced in a high-throughput manner to obtain millions of short sequences derived from single-end sequencing or pair-end sequencing. Following sequencing, the produced reads are aligned to a reference genome and then assembled into transcripts based on annotated reference transcripts. Additionally, the reads can create a genome-scale transcriptome map without the genomic sequence using de novo assembly approaches. Downstream analyses with RNA-Seq include detecting for differential gene expression between different samples, testing allele-specific expression, and identifying alternatively spliced genes and novel transcripts.

from the ovarian surface epithelium; however, recent work indicates the majority of advanced HGSOC may arise from the fallopian tube, especially in the fimbriae [31,32]. The establishment of an HGSOC model derived from transformed fallopian tube secretory epithelium has further supported the hypothesis that HGSOC principally originates from fallopian tube epithelium [32,33]. Because there have been relatively fewer cases of low-grade serous ovarian cancer (LGSOC) available for study, its origin is less clear than HGSOC. However, molecular evidence from RNA-Seq has suggested that the majority of LGSOC also originate from the fallopian tube [34]. A total of 31 tissue samples were collected, including 6 LGSC, 6 HGSC, 6 serous borderline tumors, 5 fallopian tube epithelia, 4 ovarian surface epithelia, and 4 peritoneal mesothelia. HTSeq v0.6.1 was applied to count the read numbers mapped from each gene. Unsupervised hierarchical cluster analysis was performed to display differentially expressed genes among each sample. After testing the correlation across the 6 groups by Rank-sum analysis combined with Pearson correlation analysis, LGSOC showed a significantly closer relationship with fallopian tube epithelium than with ovarian surface epithelia. In addition, PAX8 was highly expressed in both serous tumor and fallopian tube epithelium samples but not in the ovarian surface epithelium using Kruskal-Wallis contrasts. The Cancer Genome Atlas (TCGA) project, which catalogs genetic alterations in cancer, labels PAX8 as a survival gene essential for ovarian cancer cell proliferation [35]. Accumulating research has confirmed that HGSOC is derived from PAX8-expressing cells within the

Table 1
Advantages of RNA-Seq compared with other transcriptomics methods.

Technology	RNA-Seq	Expression microarray	cDNA or EST sequencing
Detection methods	High-throughput sequencing	DNA hybridization	Sanger sequencing
Reliance on reference genome	In some cases	Yes	No
Resolution	Single nucleotide	Probe length (from several to 100 bp)	Single nucleotide
Throughput	High	High	Low
Background noise	Low	High	Low
Cost for mapping transcriptomes of large genomes	Relatively low	High	High
Comparison of gene expression	Counts of reads	Relative intensities	Limited for gene expression
Identification of novel gene and isoform	Yes	No	Yes
SNP detection in the transcribed regions	Yes	No	Yes
Ability to distinguish allelic expression	Yes	No	Yes
Dynamic range of expression levels	>8000-fold	One hundred to a few hundred-fold	Not practical

Note: EST, expressed sequence tag; SNP, single nucleotide polymorphism.

Table 2
Applications of RNA-Seq in ovarian cancer.

Application	Origin	Data analysis	Result	Reference
Identification of the pathological origin	Patients	HTSeq, Unsupervised hierarchical clustering	LGSOC originated from fallopian tube	34
Identification of the fusion gens	Patients	Fusion map	Two fusion genes were discovered	40
	Patients	Illumina Genome Analyzer II (GAII)	One fusion gene has been reported	42
Identification of biomarkers in early detection	Patients	ELAND32, ERANGE	IGFBP-4 was highly expressed in tumors across all stages	48
Identification of genes in the development of MDR	Cells	Q2 Solutions-EA Genomics	DUOXA1 was up-regulated in resistant cells	50
	Patients	Cuffdiff, VARSCAN	Four new SNPs were correlated with platinum resistance	51
	Patients	TopHat, Cufflinks, Cuffdiff	IRF1 was related with platinum sensitivity	52
	PDX	Aligner STAR, RPKM	ABCG1 was up-regulated in chemotherapy treated tumors	55
	cells	Cufflinks, DESeq	Overexpression of miR-31 increased cisplatin resistance	58
	patients	Unsupervised hierarchical clustering, R/Limma package	Seven lncRNAs were identified as predictors of platinum resistance	59
Single-cell sequencing	66 single cells from one patient	Unsupervised hierarchical clustering, K-means clustering	Intratumor heterogeneity were characterized	62

Note: LGSOC, low-grade serous ovarian cancer; MDR, multidrug resistance; SNP, single nucleotide polymorphisms; IRF1, interferon regulatory factor 1; PDX, patient-derived xenograft; RPKM, reads per kilobase million; lncRNA, long noncoding RNA.

distal fallopian tube [36,37]. The characterization of ovarian cancer origin and associated molecular mechanisms made possible by RNA-Seq may lead to novel prevention strategies and shorter detection periods for ovarian cancer patients.

3.2. Identification of fusion genes

Fusion genes are a subtype of hybrid genes consisting of parts from two or more previously independent genes, formed through chromosomal deletion, translocation or inversion [38,39]. More than 25,000 fusion genes across 33 different tumor types have been identified in the TCGA Fusion Gene Data Portal [10–12]. Interestingly, integrated data on gene expression, copy number, and fusion annotation show oncogene related fusion genes tend to exhibit increased expression, whereas tumor suppressor associated genes tend to be poorly expressed [10]. However, fusion gene data within ovarian cancer is remains limited. Stemming from the combined analysis of karyotype information, RNA-Seq, and exon-level gene expression microarray data, the two fusion genes DPP9-PPP6R3 and DPP9-PLIN3 were discovered in high-grade ovarian cancer [40]. RNA-Seq data on 19 karyotyped ovarian cancer tissues was analyzed with FusionMap software, and RT-PCR and Sanger sequencing validated the presence of the two fusion genes. The DPP9 protein participates in the regulation of survival and proliferation pathways and has several tumor-suppressing abilities, such as inducing apoptosis and suppressing epidermal growth factor (EGF)-mediated AKT (protein kinase B) activation [41]. Despite having different fusion breakpoint positions, both rearrangements were associated with a shortened transcript of the 3′ end of DPP9, leading to a loss of the functional domains at the DPP9 protein level. The fusion gene MUC1-TRIM46-KRTCAP2 has also been reported in high-grade ovarian cancer, as revealed by high-throughput transcriptome sequencing data of 130 HGSOC patient samples from the TCGA [42]. This fusion gene has six splicing isoforms not present in normal ovary, yet is detected in three out of four ovarian cancer tumors. Previous studies have shown MUC1 overexpression to be indicative of metastatic and advanced ovarian cancers, as well as a valuable diagnostic marker of metastasis and progression [43,44]. Despite this, the considerable variability in MUC1 expression among tumor and normal samples challenge its reliability within the clinic as a biomarker [44]. Overall, these fusion genes are readily detectable in tumor samples, not normal ovary, and are therefore promising diagnostic biomarkers and therapeutic targets within ovarian cancer.

3.3. Identification of biomarkers in early detection

The identification of specific and sensitive serum-based biomarkers in ovarian cancer has long been sought after, yet only circulating cancer antigen 125 (CA125) is currently approved for clinical use for early detection and disease monitoring [45,46]. However, because CA125 levels are regularly elevated in many other malignant tumors, and may only increase in 50–60% of patients with early-stage disease, its clinical reliability cannot be overestimated in ovarian cancer diagnostics [47]. This limitation has driven the discovery of additional biomarkers. For example, insulin-like growth factor binding protein (IGFBP-4) has shown significant serum elevation in primary and recurrent ovarian cancer patients even in cases where CA125 levels are within the normal range [48]. Mechanistically, IGFBP-4 binds with IGF-I and IGF-II to prolong the half-life of the IGFs while altering their interaction with cell surface receptors (IGF-IRs) [49]. Based on RNA-Seq transcriptome analysis of 22 ovarian cancer patient-derived tumors across all stages, IGFBP-4 was highly expressed in early, late, and recurrent tumor samples [48]. Sequence alignments were processed with the software program ELAND32, and expression levels were quantified by running ERANGE. Subsequently, 21 ovarian cancer patient samples were selected to quantify IGFBP-4 serum protein expression levels via ELISA assay and were found to have marked elevation compared to controls. Interestingly, three of the six early-stage cases with normal CA125 levels had up-regulated IGFBP-4 levels. A larger case study was consistent with these findings, as ovarian cancer (all stages) had an average IGFBP-4 of 1344.09 ng/ml compared to 400.9 ng/ml for healthy controls and 394.6 ng/ml for benign controls. Thus, IGFBP-4 was not only up-regulated in early stage disease in the absence of elevated CA125, but was also highly expressed in malignant versus benign disease [48].

3.4. Identification of novel genes and molecular pathways in the development of MDR

The development of MDR is a major obstacle for ovarian cancer patients undergoing standard chemotherapy. Alleviation of MDR is hindered by the ability of drug-resistant tumors to manifest multiple resistance mechanisms simultaneously. Understanding the details of MDR events may lead to the development of novel therapeutic strategies to overcome drug resistance. Genetic analysis using RNA-Seq has advantages in throughput, sensitivity, and accuracy over microarray and qPCR methods. RNA-Seq has recently been applied to compare

cisplatin-sensitive and cisplatin-resistant ovarian cancer cells by Q2 Solutions-EA Genomics. Dual oxidase maturation factor 1 (DUOXA1) was notably up-regulated in cisplatin-resistant ovarian cancer cells, resulting in overproduction of reactive oxygen species (ROS) [50]. DUOXA1-mediated-ROS activation consequently sustains the DNA damage response in resistant ovarian cancer cells by stimulating the ATR-Chk1 pathway. Specifically, the combined use of RNA-Seq and PathwayNet analysis (http://pathwaynet.princeton.edu) revealed a total of 108 genes to be overexpressed in cisplatin-resistant cells and linked to Chk1 function. Additionally, heightened expression of modified genes was closely associated with poor patient prognosis [50]. In another study, ovarian cancer RNA-Seq data was gathered from the GEO database, including specimens from six platinum-sensitive and six platinum-resistant ovarian cancer samples [51]. After Btrim software processing, the data were analyzed with Cuffdiff for differentially expressed genes, then evaluated for specific mutations using VARSCAN software in the platinum-resistant samples. The data identified 38 new SNPs, and that ESRP1, LDHA, DDX5, and HEXA correlated with platinum-resistance in ovarian cancer. Similarly, RNA-Seq was performed on tumors from seven ovarian cancer patients and revealed four were chemosensitive and three were chemoresistant. All samples were aligned using TopHat (v2.0.4) to the hg19 RefSeq transcriptome, and the transcript assembly and quantification of each sample was analyzed using Cufflinks and Cuffdiff. Transcriptome analysis confirmed interferon regulatory factor 1 (IRF1), a transcription factor with functions in immune regulation and tumor suppression, to be related with platinum sensitivity and also a strong prognostic marker for both progression-free and overall survival in HGSOC [52]. For the past three decades, cell lines have been integral in investigating the mechanisms and pathways of drug resistance. However, the clonal nature of cancer cell lines limits their ability to mimic the intratumoral and interpatient heterogeneity seen in many tumors [53]. The patient-derived xenograft (PDX) model addresses this shortcoming, as tumors are collected from patients and immediately implanted into mice all while retaining their heterogeneity. The PDX model is, therefore, an attractive platform to test the efficacy of pre-clinical therapies and their related clinical response [53,54]. An ovarian PDX model has been established to study the development of chemoresistance [55]. Mice were either given an intraperitoneal injection of carboplatin and paclitaxel or a vehicle every week. The tumors were harvested six days after they received four weeks of treatment so that the chemotherapy derived acute tumor effects were minimalized. To investigate the genetic and pathway-specific alterations arising from chemotherapy treatment, RNA-Seq was conducted on six pairs of treated and untreated PDX tumors. Significant changes were found in 299 genes from the treated PDX samples compared to untreated. The reads per kilobase of exon model per million mapped reads (RPKM)-normalized reads were calculated and the expression levels of genes were estimated. One of most prominently up-regulated genes was ABCG1 (BCRP1), a member of the White family of ATP-Binding cassette (ABC) transporters.

Previous works characterizing the molecular biology of MDR in ovarian cancer has traditionally focused on genomic or epigenomic deregulation of protein-coding genes to identify the genetic alterations responsible for developing drug resistance. More recent studies, however, have demonstrated ncRNAs such as miRs and lncRNAs to be robust contributors to MDR in ovarian cancer [56–59]. Overexpression of miR-31 significantly decreased KCNMA1 and increased cisplatin resistance in ovarian cancer cells [58]. When RNA-Seq was used to identify pathways and genes involved in cisplatin resistance, KCNMA1 RNA was reduced more than a thousand-fold in the resistant cell line [58]. Transcript abundance was estimated using Cufflinks, and the relative level of transcripts between samples were compared using DESeq. In another study, a panel of seven prognosis-related lncRNAs was found to improve predictability of platinum-based chemoresistance within HGSOC [59]. Additional analysis of GEO and TCGA databases revealed four up-regulated lncRNAs and three downregulated lncRNAs to be highly sensitive markers for predicting chemoresistance and poor prognosis. These lncRNAs are useful prognostic markers and predictors of platinum resistance in HGSOC.

3.5. Single-cell sequencing in ovarian cancer

Intratumor heterogeneity is a confirmed major cause of treatment failure and drug resistance in ovarian cancer and other cancer types [60]. Challenges arise in the bulk sequencing of tumor samples, as small subpopulations of tumor cells are masked when analyzed alongside other low and high RNA expressing cells within a bulk tumor sample. Single-cell sequencing of tumor cells addresses this issue by identifying subpopulations of cancer cells within a single patient, hence facilitating the ability to characterize intratumor heterogeneity [61,62]. RNA-Seq was performed on 66 individual cells isolated from a tumor specimen obtained from a patient with HGSOC during primary debulking surgery [62]. Based on gene expression patterns as revealed by unsupervised hierarchical clustering, K-means clustering, and principle components analysis, cells were divided into two groups: epithelial cells and stromal cells. The majority of the epithelial cells belonged to the proliferative group, which expressed PAX8 and/or CA125 and genes associated with proliferation. The majority of stromal cells fell into the mesenchymal group and was characterized by increased expression of extracellular matrix (ECM) and epithelial-to-mesenchymal transition (EMT) genes. Both groups had gene expression patterns similar to cancer stem cells. Neither of these cell groups expressed known drug resistant genes, a finding which was consistent with the clinical data, as the patient in this study had no recurrence 19 months post-surgery. The RNA-Seq results revealed the presence of multiple molecular subclassifications within the tumor [62]. Single-cell sequencing may prove to be instrumental in understanding the pathogenesis, treatment response, progression and drug resistance in ovarian cancer.

4. Conclusion and future perspectives

Recent developments in RNA-Seq technology have enabled a more accurate representation of the genome-wide changes in ovarian cancer, including dysregulated gene expression, alternative splicing, SNP and ncRNA. RNA-Seq has been successfully used in understanding the pathological origin, identifying fusion genes and biomarkers, and uncovering novel mechanisms responsible for drug resistance in ovarian cancer. However, RNA-Seq presents several barriers which limit its utility. The high cost of RNA-Seq is a major deterrent for more widespread use, and large studies may therefore be impractical for smaller sized laboratories. Furthermore, precise transcript quantification is a prerequisite for downstream differential analysis of drug treatment response, differences between healthy and diseased conditions, and other research variables. Biases have also been observed in RNA-Seq datasets formed by library preparation, sequencing and read mapping. A large portion of sampling biases has been produced during cDNA library preparation depending on the applied protocols [63]. The fragments are not uniformly sampled and sequenced, as there is variability in sequencing depth across the transcriptome due to variable primers, preferred fragmentation sites, and tagged nucleotide composition effects [25]. PCR amplification is also known to introduce bias in library preparation, as GC-neutral fragments are amplified more efficiently than GC-rich or AT-rich fragments [63]. In addition, sequence reads from RNA-Seq are typically short and may originate from more than one transcript. Paralogous gene families and high sequence similarity between alternatively spliced isoforms of the same gene are the leading causes of ambiguous mapping [25,64]. Long reads support accurate and efficient mapping, which results in easily identifiable patterns of transcript variants. Although RNA-Seq brings novel algorithmic and logistical challenges for large-scale data analysis and storage, as sequencing technology advances, computational tools will likely evolve to solve these technical challenges while ensuring its novel applications.

In summary, RNA-Seq is immensely useful for large high-throughput genetic studies and has dramatically expanded our ability to recognize previously unknown transcriptional changes in ovarian cancer diagnoses and treatment. As the costs continue to decrease, read lengths increase, and computational tools become more precise, RNA-Seq technology will likely see an increased application in determining the transcriptome structure and dynamics of ovarian cancer. As a result, greater exploration of the complex and heterogeneous nature of ovarian cancer will likely promote the discovery of novel biomarkers and therapeutic strategies.

Conflict of interest statement

The authors declare that no conflict of interest exists.

Acknowledgments

This work was supported, in part, by a Joint Research Fund (228282) devoted to clinical pharmacy and precision medicine (Z.D). Support has also been provided by the Gattegno and Wechsler funds. Dr. Duan is supported, in part, a Grant from National Cancer Institute (NCI)/National Institutes of Health (NIH), UO1, CA151452-01. Dr. Wang is supported by scholarship from the China Scholarship Council (201607040013).

Authors' contributions

Conception and design: Zhenfeng Duan.

Acquisition of data: Jinglu Wang.

Writing, review, and/or revision of the manuscript: Jinglu Wang, Dylan C. Dean, Francis J. Hornicek, Huirong Shi, Zhenfeng Duan.

Study supervision: Huirong Shi, Zhenfeng Duan.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J. Clin. 68 (2018) 7–30.
[2] C. Della Pepa, G. Tonini, C. Pisano, M. Di Napoli, S.C. Cecere, R. Tambaro, et al., Ovarian cancer standard of care: are there real alternatives? Chin. J. Cancer 34 (2015) 17–27.
[3] S. Sato, H. Itamochi, Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy, Therapeutic Advances in Medical Oncology 6 (2014) 293–304.
[4] A.M. Patch, E.L. Christie, D. Etemadmoghadam, D.W. Garsed, J. George, S. Fereday, et al., Whole-genome characterization of chemoresistant ovarian cancer, Nature 521 (2015) 489–494.
[5] J. Sun, X. Cai, M.M. Yung, W. Zhou, J. Li, Y. Zhang, et al., miR-137 mediates the functional link between c-Myc and EZH2 that regulates cisplatin resistance in ovarian cancer, Oncogene (2018).
[6] V.G. Leblanc, M.A. Marra, Next-generation sequencing approaches in cancer: where have they brought us and where will they take us? Cancer 7 (2015) 1925–1958.
[7] R. Hrdlickova, M. Toloue, B. Tian, RNA-Seq methods for transcriptome analysis, Wiley Interdisciplinary Reviews RNA (2017) 8.
[8] V. Chaitankar, G. Karakulah, R. Ratnapriya, F.O. Giuste, M.J. Brooks, A. Swaroop, Next generation sequencing technology and genomewide data analysis: perspectives for retinal research, Prog. Retin. Eye Res. 55 (2016) 1–31.
[9] S.A. Byron, K.R. Van Keuren-Jensen, D.M. Engelthaler, J.D. Carpten, D.W. Craig, Translating RNA sequencing into clinical diagnostics: opportunities and challenges, Nat. Rev. Genet. 17 (2016) 257–271.
[10] Q. Gao, W.W. Liang, S.M. Foltz, G. Mutharasu, R.G. Jayasinghe, S. Cao, et al., Driver fusions and their implications in the development and treatment of human cancers, Cell Rep. 23 (2018) 227–238 (e3).
[11] F. Sanchez-Vega, M. Mina, J. Armenia, W.K. Chatila, A. Luna, K.C. La, et al., Oncogenic signaling pathways in the cancer genome atlas, Cell 173 (2018) 321–337 (e10).
[12] C. Hutter, J.C. Zenklusen, The cancer genome atlas: creating lasting value beyond its data, Cell 173 (2018) 283–285.
[13] E. Clough, T. Barrett, The gene expression omnibus database, Methods Mol. Biol. 1418 (2016) 93–110.
[14] K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, Contemp. Oncol. 19 (2015) A68–A77.
[15] K.R. Kukurba, S.B. Montgomery, RNA sequencing and analysis, Cold Spring Harb Protoc 2015 (2015) 951–969.
[16] C. Peano, A. Pietrelli, C. Consolandi, E. Rossi, L. Petiti, L. Tagliabue, et al., An efficient rRNA removal method for RNA sequencing in GC-rich bacteria, Microbial Informatics and Experimentation 3 (1) (2013).
[17] W. Zhao, X. He, K.A. Hoadley, J.S. Parker, D.N. Hayes, C.M. Perou, Comparison of RNA-Seq by poly(A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling, BMC Genomics 15 (2014) 419.
[18] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[19] Y. Yoshinaga, C. Daum, G. He, R. O'Malley, Genome sequencing, Methods Mol. Biol. 2018 (1775) 37–52.
[20] J.M. Heather, B. Chain, The sequence of sequencers: the history of sequencing DNA, Genomics 107 (2016) 1–8.
[21] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, et al., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nat. Protoc. 7 (2012) 562–578.
[22] J.E. Hooper, A survey of software for genome-wide discovery of differential splicing in RNA-Seq data, Human Genomics 8 (2014) 3.
[23] A. Sud, B. Kinnersley, R.S. Houlston, Genome-wide association studies of cancer: current insights and future perspectives, Nat. Rev. Cancer 17 (2017) 692–704.
[24] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, et al., A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (2016) 13.
[25] S. Zhao, W.P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells, PLoS One 9 (2014), e78644. .
[26] D.G. Robinson, J.Y. Wang, J.D. Storey, A nested parallel experiment demonstrates differences in intensity-dependence between RNA-seq and microarrays, Nucleic Acids Res. 43 (2015), e131. .
[27] M.S. Boguski, C.M. Tolstoshev, D.E. Bassett Jr., Gene discovery in dbEST, Science 265 (1994) 1993–1994.
[28] C. Khoo, T.M. Rogers, A. Fellowes, A. Bell, S. Fox, Molecular methods for somatic mutation testing in lung adenocarcinoma: EGFR and beyond, Translational Lung Cancer Research 4 (2015) 126–141.
[29] H.F. Tsang, V.W. Xue, S.P. Koh, Y.M. Chiu, L.P. Ng, S.C. Wong, NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics, Expert. Rev. Mol. Diagn. 17 (2017) 95–103.
[30] G.C. Jayson, E.C. Kohn, H.C. Kitchener, J.A. Ledermann, Ovarian cancer, Lancet 384 (2014) 1376–1388.
[31] S.I. Labidi-Galy, E. Papp, D. Hallberg, N. Niknafs, V. Adleff, M. Noe, et al., High grade serous ovarian carcinomas originate in the fallopian tube, Nat. Commun. 8 (2017) 1093.
[32] R. Perets, G.A. Wyant, K.W. Muto, J.G. Bijron, B.B. Poole, K.T. Chin, et al., Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca;Tp53;Pten models, Cancer Cell 24 (2013) 751–765.
[33] K. Nakamura, K. Nakayama, N. Ishikawa, M. Ishikawa, R. Sultana, T. Kiyono, et al., Reconstitution of high-grade serous ovarian carcinoma from primary fallopian tube secretory epithelial cells, Oncotarget 9 (2018) 12609–12619.
[34] C. Qiu, N. Lu, X. Wang, Q. Zhang, C. Yuan, S. Yan, et al., Gene expression profiles of ovarian low-grade serous carcinoma resemble those of fallopian tube epithelium, Gynecol. Oncol. 147 (2017) 634–641.
[35] H.W. Cheung, G.S. Cowley, B.A. Weir, J.S. Boehm, S. Rusin, J.A. Scott, et al., Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 12372–12377.
[36] A.A. Tone, S. Salvador, S.J. Finlayson, A.V. Tinker, J.S. Kwon, C.H. Lee, et al., The role of the fallopian tube in ovarian cancer, Clin. Adv. Hematol. Oncol.: H&O 10 (2012) 296–306.
[37] M.H. Roh, D. Kindelberger, C.P. Crum, Serous tubal intraepithelial carcinoma and the dominant ovarian mass: clues to serous tumor origin? Am. J. Surg. Pathol. 33 (2009) 376–383.
[38] F. Mertens, B. Johansson, T. Fioretos, F. Mitelman, The emerging complexity of gene fusions in cancer, Nat. Rev. Cancer 15 (2015) 371–381.
[39] I. Panagopoulos, J. Thorsen, L. Gorunova, F. Micci, S. Heim, Sequential combination of karyotyping and RNA-sequencing in the search for cancer-specific fusion genes, Int. J. Biochem. Cell Biol. 53 (2014) 462–465.
[40] M.L. Smebye, A. Agostini, B. Johannessen, J. Thorsen, B. Davidson, C.G. Trope, et al., Involvement of DPP9 in gene fusions in serous ovarian carcinoma, BMC Cancer 17 (2017) 642.
[41] T.W. Yao, W.S. Kim, D.M. Yu, G. Sharbeen, G.W. McCaughan, K.Y. Choi, et al., A novel role of dipeptidyl peptidase 9 in epidermal growth factor signaling, Mol. Cancer Res.: MCR 9 (2011) 948–959.
[42] K. Kannan, G.K. Kordestani, A. Galagoda, C. Coarfa, L. Yen, Aberrant MUC1-TRIM46-KRTCAP2 chimeric RNAs in high-grade serous ovarian carcinoma, Cancer 7 (2015) 2083–2093.
[43] T.M. Horm, J.A. Schroeder, MUC1 and metastatic cancer: expression, function and therapeutic targeting, Cell Adhes. Migr. 7 (2013) 187–198.
[44] J. Deng, L. Wang, H. Chen, L. Li, Y. Ma, J. Ni, et al., The role of tumour-associated MUC1 in epithelial ovarian cancer metastasis and progression, Cancer Metastasis Rev. 32 (2013) 535–551.
[45] N. Scholler, N. Urban, CA125 in ovarian cancer, Biomark. Med 1 (2007) 513–523.
[46] M. Felder, A. Kapur, J. Gonzalez-Bosquet, S. Horibata, J. Heintz, R. Albrecht, et al., MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress, Mol. Cancer 13 (2014) 129.
[47] D. Sasaroli, G. Coukos, N. Scholler, Beyond CA125: the coming of age of ovarian cancer biomarkers. Are we there yet? Biomark. Med 3 (2009) 275–288.
[48] R.A. Mosig, M. Lobl, E. Senturk, H. Shah, S. Cohen, E. Chudin, et al., IGFBP-4 tumor and serum levels are increased across all stages of epithelial ovarian cancer, Journal of Ovarian Research 5 (2012) 3.
[49] R. Durai, M. Davies, W. Yang, S.Y. Yang, A. Seifalian, G. Goldspink, et al., Biology of insulin-like growth factor binding protein-4 and its role in cancer (review), Int. J. Oncol. 28 (2006) 1317–1325.
[50] Y. Meng, C.W. Chen, M.M.H. Yung, W. Sun, J. Sun, Z. Li, et al., DUOXA1-mediated ROS production promotes cisplatin resistance by activating ATR-Chk1 pathway in ovarian cancer, Cancer Lett. 428 (2018) 104–116.

[51] X. Ye, Confluence analysis of multiple omics on platinum resistance of ovarian cancer, Eur. J. Gynaecol. Oncol. 36 (2015) 514–519.
[52] S. Cohen, R. Mosig, E. Moshier, E. Pereira, J. Rahaman, M. Prasad-Hayes, et al., Interferon regulatory factor 1 is an independent predictor of platinum resistance and survival in high-grade serous ovarian carcinoma, Gynecol. Oncol. 134 (2014) 591–598.
[53] M. Hidalgo, F. Amant, A.V. Biankin, E. Budinska, A.T. Byrne, C. Caldas, et al., Patient-derived xenograft models: an emerging platform for translational cancer research, Cancer Discovery 4 (2014) 998–1013.
[54] J.W. Cassidy, C. Caldas, A. Bruna, Maintaining tumor heterogeneity in patient-derived tumor xenografts, Cancer Res. 75 (2015) 2963–2968.
[55] Z.C. Dobbin, A.A. Katre, A.D. Steg, B.K. Erickson, M.M. Shah, R.D. Alvarez, et al., Using heterogeneity of the patient-derived xenograft model to identify the chemoresistant population in ovarian cancer, Oncotarget 5 (2014) 8750–8764.
[56] H. Xia, K.M. Hui, Mechanism of cancer drug resistance and the involvement of non-coding RNAs, Curr. Med. Chem. 21 (2014) 3029–3041.
[57] R.C. Pink, P. Samuel, D. Massa, D.P. Caley, S.A. Brooks, D.R. Carter, The passenger strand, miR-21-3p, plays a role in mediating cisplatin resistance in ovarian cancer cells, Gynecol. Oncol. 137 (2015) 143–151.
[58] P. Samuel, R.C. Pink, D.P. Caley, J.M. Currie, S.A. Brooks, D.R. Carter, Over-expression of miR-31 or loss of KCNMA1 leads to increased cisplatin resistance in ovarian cancer cells, Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine. 37 (2016) 2565–2573.
[59] J. Song, W. Zhang, S. Wang, K. Liu, F. Song, L. Ran, A panel of 7 prognosis-related long non-coding RNAs to improve platinum-based chemoresistance prediction in ovarian cancer, Int. J. Oncol. 53 (2018) 866–876.
[60] J.Y. Lee, J.K. Yoon, B. Kim, S. Kim, M.A. Kim, H. Lim, et al., Tumor evolution and intratumor heterogeneity of an epithelial ovarian cancer investigated using next-generation sequencing, BMC Cancer 15 (2015) 85.
[61] A.P. Patel, I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, et al., Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma, Science 344 (2014) 1396–1401.
[62] B.J. Winterhoff, M. Maile, A.K. Mitra, A. Sebe, M. Bazzaro, M.A. Geller, et al., Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells, Gynecol. Oncol. 144 (2017) 598–606.
[63] E.L. van Dijk, Y. Jaszczyszyn, C. Thermes, Library preparation methods for next-generation sequencing: tone down the bias, Exp. Cell Res. 322 (2014) 12–20.
[64] X. Liu, X. Shi, C. Chen, L. Zhang, Improving RNA-Seq expression estimation by modeling isoform- and exon-specific read sequencing rate, BMC Bioinformatics 16 (2015) 332.