Gene expression and sequencing techniques of transcriptome to elucidate novel function for phospholipase (s) enzyme

M G Tyagi, S Saha, D Mukherjee, A Ramaswamy, PA Vora

Department of Biotechnology, Vellore Institute of Technology, Vellore, Tamilnadu and Department of Paramedical Technology, NIMS University, Jaipur, Rajasthan 303121

Received 01 September 2018; Accepted 26 September. 2018

ABSTRACT:
Sequencing of RNA, or RNA-Seq, is now a common method to analyze gene expression and to uncover novel RNA species. Aspects of RNA biogenesis and metabolism can be evaluated with specialized methods for cDNA library preparation. In this study, the review examines current RNA-Seq methods for general analysis of gene expression and several specific applications, including isoform and gene fusion detection, targeted sequencing and single-cell analysis. Elucidating the transcriptome may help in better understanding of phospholipase function and novel application in diagnostics and clinical therapeutics. Therefore this review article focuses on the transcriptome analysis of the phospholipase enzymes i.e phospholipase A2, D and C isozymes and characterizing their functional roles in specialized disorders or for therapy.

Key words: RNA, sequencing, transcriptome, phospholipase D, gene, expression

INTRODUCTION
Lipids play a critical role in signal transduction as they often act as secondary messengers to activate downstream pathways [1]. Phospholipid mediated signal transduction and its role in pathophysiological states is an area of intensive research. It has been found for e.g that the overexpression of myostatin propeptide activated the phosphatidic acid (PA) signaling pathway, which generates signaling lipids and participates in the cell cycle and cell growth [2,3]. In that study, they found that Plcd4 and Dgkz, two key enzymes of the PA signaling pathway, were up-regulated in myostatin propeptide transgenic mice. These results suggested a connection between the myostatin signaling and PA signaling pathways. Plcd4 is a member of the delta class of phospholipase C enzymes. Phospholipase C enzymes can catalyze hydrolyzing phosphatidylinositol 4,5-bisphosphate (PIP2) to two intracellular secondary messengers, inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG) [4]. Previous studies reported that DAG could be phosphorylated by Dgkz, and the phosphorylated DAG could be digested to phosphatidic acid (PA). PA can reportedly bind to mTOR to induce muscle fiber hypertrophy [5]. Other studies reported that myostatin mediates myoblast differentiation and myotube hypertrophy by inhibiting Akt/mTOR/p70S6 protein signaling [6-7]. Overall, it can be inferred that myostatin can negatively regulate the PA signaling pathway to inhibit the mTOR signaling pathway. In
addition to activating the mTOR signaling pathway, PA also activates p21-activated kinase 1 (PAK1) to initiate the release of RhoGDI from Rac1, and RhoGD1 subsequently changes the actin dynamics [8]. These results suggest novel function of phospholipases in skeletal muscle system. On the other hand, phospholipase A2 expression is affected by growth factors like the EGF and PDGF and the promoter region of the phospholipase A2 gene contains the consensus sequence for binding to transcription factors. Molecular techniques involving the sequencing of RNA and single cell sequencing are unravelling some undescribed functions of the phospholipase enzymes. This study focuses on recent sequencing and gene expression studies elucidating novel roles of the three phospholipase A2, C and D enzymes in biology and therapeutics.

**General principles of RNA-Seq:**

Directing sequencing of RNA molecules is possible and most RNA-Seq experiments are carried out on instruments that sequence DNA molecules due to the technical maturity of commercial instruments designed for DNA-based sequencing [9]. Therefore, cDNA library preparation from RNA is a required step for RNA-Seq. Each cDNA in an RNA-Seq library is composed of a cDNA insert of certain size flanked by adapter sequences, as required for amplification and sequencing on a specific platform. The cDNA library preparation method varies depending on the RNA species under investigation, which can differ in size, sequence, structural features and abundance. Major considerations are as following i.e (1) how to capture RNA molecules of interest; (2) how to convert RNA to double-stranded cDNAs with defined size ranges; and (3) how to place adapter sequences on the cDNA ends for amplification and sequencing.

**Technique for sequencing the ribose nucleic acid:** RNA-seq libraries are prepared from either total RNA or a sub-fraction of RNA i.e. poly-adenylated, ribosomal RNA depleted, 5’ cap containing, small RNA by random shearing of RNA into 100-400 base pair fragments. Barcodes and sequencing primer binding sites are then added, and the RNA is briefly amplified, followed by size selection of 100-400 bp fragments also called inserts. This size selection captures most protein coding RNAs and long non-coding RNAs. For small RNA sequencing, which captures mature microRNAs, piwi-interacting RNAs, and other small RNAs, fragmentation is skipped and inserts < 40 bp are selected. Following insert selection, around 30 bp up to >150 bp of one end single end read or both ends paired-end reads of each fragment are sequenced. Sequences reads are then computationally mapped by finding unique usually matches between the annotated genome/transcriptome and the sequenced end of the RNA fragment. If paired-ends are sequenced, then each end is matched with the expected distance constrained to the approximate expected fragment size. If libraries are stranded, then the read can be mapped directionally to either the positive or negative strand of DNA [10].

**Novel uses of sequencing the RNA:**

Almost all multi-exon genes display alternative splicing (AS). AS plays an important role in regulation of cellular processes, and aberrations of the process are associated with many human diseases [11]. Some RNA-Seq reads cover exon-exon junctions, providing substantial evidence of AS. In addition, reads mapped to internal exonic regions can be used to predict the AS pattern using statistical inference methods. A more direct approach to examine AS is to sequence the exon-exon junction region directly. Similar to splicing, gene fusion events can place two noncontinuous genomic regions together in a single transcript. Created by chromosomal rearrangements, gene fusions are present in approximately 20% of cancer [12]. Fusion
events can be detected using RNA-Seq data along with specific bioinformatic methods [13]. Detection of a fusion event is actually revealed by reads containing fusion junctions or by differences in expression between the 5′ and 3′ ends of genes that are fused. Regular RNA-Seq methods are typically not sufficiently sensitive to detect fusion junctions. Several methods have been developed, including (1) enrichment of RNA-Seq reads for genes of interest, (2) exon capture, and (3) amplicon sequencing. In a recent study, exon capture of 467 cancer-related genes was successfully employed for the detection of their fusion events.

**Single cell RNA sequence:**

A single mammalian cell contains approximately 5–15 pg of RNA [14]. However, the RNA-Seq methods described above are generally not suitable for sequencing RNA with less than 1ng. Thus special RNA/DNA amplification or enhanced efficiency for sample processing is needed for single-cell RNA-seq. Some methods, such as CEL-Seq and MARS-Seq, introduce a T7 promoter sequence with oligo(dT) during RT, which enables linear amplification of input RNA by in vitro transcription [15]. Second strand cDNA synthesis has been enhanced by template-switching during RT (SMART-Seq) and poly(A) tailing of cDNA [16-17]. These methods, however, do not provide strand-specific information, because the cDNA is subsequently fragmented and ligated to a second set of adapters. Recently, the multiple annealing and looping-based cycles (MALBAC) method, originally designed for single-cell genomic DNA amplification, was applied to RNA-Seq [18]. State-of-the-art technologies are in development to characterize transcriptomes inside the cells, providing spatial information of RNA expression. One novel method, TIVA, is based on introduction of biotinylated tags to cells in tissue, followed by targeted activation of these tags by a laser in selected cells. The laser activates poly(U) tracts on the biotinylated tags, enabling them to bind mRNAs in the targeted cell. The bound mRNAs are then selected by streptavidin, cloned, and sequenced. In another technique, fluorescent in situ RNA sequencing (FISSEQ), combines in situ amplification with sequencing by oligonucleotide ligation [19].

**A novel phospholipase A2 inhibitor based on snake venom PLA2 transcriptome?**

PLA2 inhibitors (PLIs) isolated from snake blood may present differences among different snake species PLIs [20]. These molecules are oligomeric, globular, acidic glycoproteins with molecular mass ranging between 75 and 180 kDa. The mechanism of action of secreted PLA2 inhibitors is through the formation of a soluble complex between the inhibitor and the target enzyme [21]. These inhibitors are classified into three groups, α, β and γ, based on their structural characteristics and may be present in a single snake, whether poisonous or non-poisonous in nature. Inhibitors from the three groups have been isolated from diverse snake species and families [22]. Furthermore, the study of PLA2 inhibitors is a critical area of research because of the pharmacological potential of these components in the treatment of inflammation and as a tool to investigate the role of PLA2s in physiological functions of cell death and injury. In recent years, synthetic derived peptides from PLIs were showed to have therapeutic potential in PLA2-related diseases and as anti-venom-like bioactive molecules for snake venom neutralization [23,24]. At.PLI1 and At.PLI2 amino acid sequences show homology to the PLIs ‘three-finger fold, which are also present in urokinase-type plasminogen activator receptors (u-PAR) and in cell surface antigens from the superfamily Ly-6 and CD59. The similarity in the structure of PLIs and uPAR suggest that structure activity relationship studies can result in PLA2 inhibitors which could be used for anti-inflammatory or anti-venom therapy. PLI transcriptome could lead
to even developments of SiRNAs for inhibiting the snake venom PLA2 [25].

**Phospholipase C and its role in leukemia:**

A general mechanism for PLC activation, first stated clearly by Hicks et al. [26] involves movement of the X-Y linker in the TIM barrel domain away from the active site, which it apparently occludes. The X-Y linker is evidently inhibitory because both proteolysis and genetic deletion [27] of sections of this linker activate PLC-β, -γ, -δ, and -ε. Furthermore, when the N- and C-terminal portions of PLC-β and PLC-γ are expressed as separate proteins, they can bind to form an active enzyme that has elevated basal activity [28]. PLC-ζ, which has a positively charged X-Y linker rather than the more common negative charge cluster, is constitutively active, and removal of this linker causes inhibition [29] (Fig.1). Hicks et al also based their proposal on the observation that PLC-δ and a fragment of PLC-β that lacks the C-terminal coiled-coil domain show very little if any change in overall tertiary structure upon activation, as indicated by X-ray crystallography [30]. This suggested the need for a mechanism other than a conventional conformational isomerization of this core enzyme. This idea was supported by the fact that the X-Y linker regions of PLC-β and PLC-δ appear to be either highly mobile or disordered or even both because neither is observed by X-ray diffraction. PLC-γ1 and PLC-γ2 are similar in structure, and their regulation appears to be similar in most cases. PLC-γ1 is widely expressed, whereas PLC-γ2 is expressed primarily in the immune cell lineage, but many cells express both isoforms. To generalize, PLC-γ1 appears to be important primarily for control of cell growth and differentiation in response to both receptor tyrosine kinases and soluble tyrosine kinases that are recruited to the plasma membrane by assorted inputs. In immune cells, PLC-γ2 acts downstream of soluble tyrosine kinases recruited by T and B cell receptors and modulate more acute responses [31]. In cells that express both enzymes, each can exert non overlapping functions, and one enzyme generally cannot compensate for depletion of the other [32]. In a recent study, PLC-γ1 expression was significantly higher in t(8;21) AML compared to other karyotypes. The PLC-γ1 protein expression was suppressed in AML1-ETO knock down cells indicating that it might induce kasumi-1 cell death. ShRNA-mediated PLC-γ1 knockdown in kasumi-1 cells significantly blocked cell growth, induced apoptosis and cell cycle arrest which was explained by the increased activation of apoptotic related and cell cycle regulatory protein expressions. Gene expression array analysis showed the up-regulation of apoptotic and DNA damage response genes together with the downregulation of cell growth, proliferation and differentiation genes in the PLC-γ1 suppressed kasumi-1 cells, consistent with the observed phenotypic effects. Importantly, PLC-γ1 suppressed kasumi-1 cells showed higher chemosensitivity to the chemotherapeutic drug treatments and lower cell proliferation upon hypoxic stress. Taken together, these *in vitro* findings strongly suggested an important role for PLC-γ1 in the survival of t(8;21) AML mimicking kasumi-1 cells and identify PLC-γ1 as a probable therapeutic target for t(8;21) AML treatment.

**Phospholipase D enzyme**

The RNA sequencing of transcriptome and related signal transduction studys has revealed newer roles for phospholipase D isoenzymes i.e PLD1 to PLD6 have found roles in various disorders ranging from autoimmune to cancer.

**Structure, regulation and implicated roles in diseases:** PLD1 and PLD2, which are ~50% identical in protein sequence and have almost the same protein domain organization, are widely expressed in different tissues and cell types and are activated by a variety of signaling molecules including protein kinase C and the small GTPases RhoA and ARF [33]
The PLD catalytic site is defined by the presence of two highly-conserved His-x-Lys-x-x-x-Asp sequences (x is any amino acid) termed the HKD motif, or more broadly, the PLDc domain, each of which creates half of the split-catalytic site. The HKD motifs are important for PLD enzymatic activity. PLD1 has been proposed to play important roles in the invasive migration of glioma cells, and in glioma cell proliferation, cell adhesion, and viability [34-36]. The tumor signaling pathways and mechanisms relevant to PLD1 function are complex and have been proposed to include activation of AKT, upregulation of HIF1-α, and increased VEGF and MMP-2 secretion [37]. Overlapping roles have been suggested for PLD2 [38]. Small molecule inhibitors of PLD1 and PLD2 such as FIP1 or isoform-selective analogs have been shown to have dramatic effects on human glioma cell lines in tissue culture studies in the context of the PLD-driven roles above. Roles for PLD2 in thrombotic disease [39-40], cancer, Alzheimer’s disease and immune function based on animal model studies have recently been summarized [41]. PLD2 and Cancer PLD2 polymorphisms, as well as up-regulated protein activity levels have been observed in several types of cancer including gastric, colorectal, kidney and breast [41-42]. In a particularly interesting recent report, it was observed that expression of microRNA (miR)-203 in high WHO grade glioma tissues was significantly lower than in low WHO grade gliomas and normal brain tissue. Transfection of a miR-203 mimic into human glioma cells strongly and directly downregulated PLD2 expression and in parallel suppressed proliferation and invasion of the glioma cells, whereas PLD2 overexpression rescued the effects induced by the miR-203 mimic. PLD3 and Alzheimer’s disease. PLD1, PLD2 and PLD3 have all been implicated in Alzheimer’s disease (AD). PLD3 is highly expressed in the brain, including in but not limited to mature neurons of the forebrain, the hippocampus and cortex [43]. Rare coding variants in PLD3 have been associated with up to 9% of late-onset AD in 14 families of European ancestry. PLD4 and autoimmune diseases as is the case for PLD3, it is not known whether PLD4 has a bona fide enzymatic function. Nonetheless, PLD4 clearly has important functional roles. Initial reports described PLD4 expression in microglia, the macrophage-like innate immune cells of the CNS, as well as in splenic cells, presumably macrophages. PLD4 expression increases with microglial activation, which is also characterized by increased phagocytic capacity. siRNA knockdown of PLD4 suppressed phagocytosis, suggesting a role for PLD4 in the setting of CNS injury and infection [44]. Despite having no catalytic activity, PLD5 has been linked to a number of diseases, including a profibrotic uterine phenotype that occurs during childbearing years, and PLD5 polymorphisms may be associated with an increased risk of tumor progression in multiple cutaneous and uterine leiomyomatosis syndrome [45-46]. PLD6-deficient mice, which cannot generate piRNAs to suppress transposon mobilization during spermatogenesis, are completely sterile, but are otherwise grossly normal to inspection. PLD6 mutations do not appear to be a major cause of human infertility; sequencing of PLD6 in 400 azoospermic European men did not uncover any PLD6 polymorphisms. Nonetheless, PLD6 may have other, less obvious roles per se.
Fig.1: Phospholipase C zeta enzyme mRNA expression in smooth muscle

Fig.2: Phospholipase A2 motif linked to vasopressin hormone

Conclusion:
Since its inception about a decade years ago, RNA-Seq has become a widely used and indispensable approach for studying gene expression and interrogating aspects of RNA biogenesis and metabolism particularly in tissues where expression levels are low. As sequencing technologies continue to advance, many methods have been developed and new RNA-Seq methods are expected to emerge in the future. It is believed that several areas are particularly relevant to RNA analysis. First, current sequencing platforms have size limitations. Secondly, current sequencing chemistry cannot handle homopolymers efficiently. Instruments that can handle long reads and offer high read output would be particularly beneficial for quantitative analysis of transcript isoforms, including those generated by alternative initiation and alternative polyadenylation, and gene fusions. In addition to expression, RNA-seq analysis provides a tool for examining transcripts, and their plausible roles in cell function and disease, at several other layers within the cell terrain including the 5′ cap and poly-A tail, the 5′ and 3′ UTRs, intron retention, alternative splicing, strand, and single nucleotide sequences. RNA sequencing can also reveal the R loop hybrid formations implicated in many neurodegenerative disorders. This is particularly relevant when sequencing the poly(A) tail region, which plays important roles in transcript metabolism.

References:


17. Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, et al. Full-length mRNA-Seq from single-cell levels of RNA and


32. Hasan Mahmud, Frank J.G. Scherpen, Tiny Meeuwsen de Boer, Harm-Jan Lourens, Caroline Schoenherr, Matthias Eder, Michaela Scherr, Victor Guryev and


