Journal of Cell Identity

ExSeq: Expansion Sequencing for Single-Cell Spatial Transcriptomics

Andra Fortner[1,2], Antoanela Tanca[2,3], Octavian Bucur[2,4,5] *

1 Medical School, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany
2 Victor Babes National Institute of Pathology, Bucharest, Romania
3 Carol Davila University of Medicine and Pharmacy, Bucharest, Romania
4 Viron Molecular Medicine Institute, Boston, MA 02108, USA
5 Genomics Research and Development Institute, Bucharest, Romania

Abstract

The innovation of spatial transcriptomics allows for the study of gene expression while preserving spatial information on the detected transcripts. Expansion Sequencing (ExSeq) is an advanced spatial transcriptomic method combining *in situ* sequencing with expansion microscopy, thus enablingthe sequencing of RNAs in a physically enlarged tissue. The experiment involves the steps of tissue expansion, RNA library preparation and multiplexed sequencing by rounds of fluorescent dye addition, imaging and fluorescent dye removal. ExSeq significantly enhances resolution capacity of RNA transcripts up to the nanoscale level, thus identifying subcellular location of RNAs. The method is applicable to exploring cellular RNA in an untargeted approach, as well as to specifically detecting certain RNA types in a targeted approach, thus making the method suitable for a variety of research questions. However, limitations to the method include referring to the readout length of transcripts obtainable during sequencing and the availability of the method.

ExSeq has the potential to be used in multiple tissues to allow for new insights into cellular identity, heterogeneity, interactions and implications in diseases and could possibly be adapted to inquire other cellular molecules, such as DNA or proteins on a nanoscale level.

Keywords:
spatial transcriptomics, *in situ* sequencing, expansion microscopy, ExSeq.

DOI:
https://www.doi.org/xxxxxxxxxxjocixxxxxxxxx

*** Corresponding author:**
Octavian Bucur, MD, PhD, Next Generation Pathology Group,
Genomics Research and Development Institute, Bucharest, Romania"
Victor Babes National Institute of Pathology, Bucharest, Romania and
Viron Molecular Medicine Institute, Boston, MA 02108, USA;
octavian.bucur@ivb.ro; octavian.bucur@gmail.com

1. Introduction

Transcriptomics as one of the subdisciplines of omics has greatly evolved in recent years. The innovation of new transcriptomic methods continuously impacts the study of gene expression in cells and tissues, allowing for discoveries to be made regarding cell identity, heterogeneity and their implications in diseases. Spatial transcriptomics additionally preserves locational information about the RNAs inside the cell examined during transcriptomic experiments. One approach to spatial transcriptomics is in situ sequencing (ISS). ISS methods are carried out directly on the permeabilized cell and can be expanded to analyze tissue sections. Examples of ISS methods are ISS using padlock probes (Ke et al., 2013), FISSEQ (fluorescent in situ sequencing) (Lee et al., 2015), STARmap (spatially resolved transcript amplicon readout mapping) (Wang et al., 2018) and ExSeq (expansion sequencing) (Alon et al., 2021). ExSeq is an enhanced version of ISS which can be obtained by applying Expansion Microscopy (ExM) on ISS technology. Chen et al. developed ExM in 2015 as a method of introducing a physically expandable gel into cells to enlarge the tissue, so that the resolution limit of tissue structures as seen under a light microscope can exceed the normal maximum of about 250 nm to to reach 70 nm (Chen et al., 2015; Bucur et al., 2020). ExSeq was introduced by Alon et al. and has significantly enhanced resolution capacity of RNA transcripts, allowing localisation in the subcellular compartments (Alon et al., 2021). ExSeq can be performed in a targeted or untargeted mode. While in a targeted ExSeq approach cells are specifically screened for targeted RNAs, untargeted ExSeq detects RNAs in an unbiased manner (Alon et al., 2021).

2. The ExSeq method

2.1 Untargeted ExSeq

ExSeq starts with **cell fixation** and **permeabilization** (Alon et al., 2021). Firstly, expansion has to be performed before ISS. Expansion consists of the three steps: gelling, digestion and expansion (Chen et al., 2015). Before starting the expansion protocol, cell(s) are treated with LabelX diluted in MOPS (3-(N-morpholino)propanesulfonic acid) buffer which alkylates RNA bases to **anchor RNAs to the cellular matrix** (Alon et al., 2021).

Gelling is performed by embedding the sample into a gelling solution containing PBS (phosphate-buffered saline), NaCl, the monomer sodium acrylate, the co-monomer acrylamide, the cross-linker N,N'methylenebisacrylamide, the catalyst 4-hydroxy-TEMPO, the initiator ammonium persulfate (APS), and the accelerator tetramethylethylenediamine (TEMED) (Alon et al., 2021). Then, radical polymerization of the gel is carried out in a gel chamber which takes 1,5-2 hours (Alon et al., 2021). For the second step, **digestion**, the gelled cells are incubated overnight at 37°C with Proteinase K which is responsible for protein cleavage and release of nucleic acids (Alon et al., 2021). This is important because it allows isotropic expansion of the protein scaffold in the cell during the expansion step. The digestive solution also contains Tris, EDTA (thylenediaminetetraacetic acid), Triton X-100 and 2-Amino-5-methoxybenzoic acid which removes formaldehyde adducts (Alon et al., 2021). Lastly, **expansion** is performed by incubation of the cells in double-distilled water (ddH2O) for three times 45 minutes, which results in an expansion factor of about 3.3 (Alon et al., 2021).

After the expansion protocol has been completed, **re-embedding** is performed (Alon et al., 2021). Re-embedding is the placement of the cell-containing expanded gel into a polyacrylamide gel which is non-expandable. This fixes the cell-containing expanded gel to prevent any conformational changes that the cell-containing expanded gel could possibly undergo during its use in the following procedure. Moreover, **fluorescent beads** can be added which help in the later image processing steps (Alon et al., 2021).

The next step is **passivation**. During passivation carboxylic acid groups of sodium acrylate in the expansion gel have to be removed which is done by binding ethanolamine to them (Alon et al., 2021). The reason for removing this groups is that they work as inhibitors for enzymes needed in the later sequencing steps (Alon et al., 2021).

At this point, the cells are prepared for the following **RNA library preparation**. Altogether, FISSEQ-based ISS contains the steps of DNA removal, reverse transcription, complementary DNA (cDNA) anchoring and circulization and rolling circle amplification (RCA) (Alon et al., 2021).

DNA removal

The cell's genome is digested by DNAse I to avoid mix-up with cDNA that will be obtained in the experiment after RNA preparation (Alon et al., 2021).

Reverse transcription of RNA into cDNA

RNAs present in the cells are reverse transcribed into cDNA by firstly primer annealing, secondly complementary strand synthesis by a reverse transcriptase and thirdly denaturation to separate the cDNA from the RNA strand (Alon et al., 2021). For the untargeted approach random RNA primers are used. During complementary strand synthesis, aminoallyl dUTP molecules are used additionally to the normal deoxynucleotide triphosphates (dNTPs) (Alon et al., 2021). Furthermore, adaptors are added to the ends of the cDNA because those are needed later to perform SOLiD sequencing (Alon et al., 2021). More precisely, the P1 adaptor is added to the 5' end, whereas the P2 adaptor is attached to the 3' end of the cDNA (Alon et al., 2021).

cDNA anchoring and circulization

To anchor the cDNA to the scaffold of the cell, BS(PEG)9 (PEGylated bis(sulfosuccinimidyl)suberate)) is used (Alon et al., 2021). This substance triggers a chemical reaction between the aminoallyl dUTPs that have been incorporated into the cDNA, and amine groups of cellular biomolecules (Alon et al., 2021). Moreover, the RNA is degraded after this step to prevent competitive inhibition of the enzyme CircLigase which is needed for cDNA circulization (Alon et al., 2021). CircLigase is a thermostable enzyme catalyzing the ligation of single stranded DNA into a circle (Li et al., 2021).

Rolling circle amplification (RCA)

Circularized cDNA serves as a template during rolling circle amplification. RCA is a technique for nucleic acid amplification which in contrast to the widely used polymerase chain reaction (PCR) operates isothermally and uses a circular DNA as a template (Garafutdinov et al., 2021). It involves 3 steps:

(1) Annealing of a starting primer to the single-stranded circular DNA template,
(2) synthesis of the complementary strand along the circle by a special polymerase with strand displacement activity and
(3) displacement of the copy from the circular DNA template as soon as the 5' end of the primer is reached, while at the same time continuing the synthesis along the circle (Garafutdinov et al., 2021).

This procedure thus creates one long single-stranded DNA which can contain between 10 to a few thousand copies of the template (Garafutdinov et al., 2021; Lee et al., 2015). The amount of copies in the ExSeq performed by Alon et al. amounts to 100 (Alon et al., 2021). In this experiment aminoallyl dUPTs are also used during RCA and cross-linking via BS(PEG)9 is again being performed in order to form the long amplified DNA copy into a tuft called the RCA amplicon (Alon et al., 2021).

SOLiD sequencing

After these steps have been completed, the cells are ready for ISS. Herein, sequencing by oligonucleotide ligation and detection (**SOLiD**) is used which is carried out at room temperature under a fluorescence microscope in a computer-controlled flow chamber (Alon et al., 2021). To perform SOLiD sequencing, the cDNA fragments have to be comprised by the P1 and P2 adaptors on their opposite ends respectively, which has already been dealt with during reverse transcription of the RNA.

In the subsequent step, probes consisting of eight nucleotides each are introduced. These probes are linked to a fluorescent dye. They comprise four distinct segments arranged from 3' to 5'.

The first part consists of a pair of bases such as AT, CT, or GG, resulting in a total of 16 possible combinations due to the four bases.

The second part includes three degenerate bases capable of binding to any of the four bases: A, T, C, or G.

The third segment comprises three inosine bases, and finally, the fourth part is the fluorescent dye itself. (Alon et al., 2021; Voelkerding et al., 2009).

These probes are made up of four distinct parts which from 3' to 5' are:

(1) a combination of two bases, e.g. AT, CT, GG. As there are four bases, a permutation of 16 bases is possible.
(2) three degenerate bases able to bind to any of the four bases A, T, C and G.
(3) three inosine bases and(
(4) a fluorescent dye (Alon et al., 2021; Voelkerding et al., 2009). In the experiment four different types of probes are used,

i.e. four colors of fluorescent dye with each color corresponding to four distinct combinations of bases in the first two nucleotides of the probe (Voelkerding et al., 2009; Alon et al., 2021).

(5) the SOLiD ligase is added and sequencing can start.

After primer binding to P1, a complementary probe anneals to the DNA fragment adjacent to the primer and SOLiD ligase joins the 5' phosphate group of the primer with the 3' hydroxyl group of the probe together (Voelkerding et al., 2009). After a washing step the fluorescent signal of the probe can be measured and noted, which is followed by cleavage of the fluorescent dye including the adjacent three inosine bases of the probe and a wash (Voelkerding et al., 2009; Alon et al., 2021). Thus, the remaining probe on the DNA fragment is only five bases long. Seven cycles of probe annealing, ligation, washing, fluorescence measurement and washing are repeated (Alon et al., 2021; Voelkerding et al., 2009). To this point, the data collected only comprises the fluorescent measurement for each 5th base on the DNA fragment. To identify the bases that lie in between, denaturation is performed to remove the newly synthesized strand and to reuse the DNA fragment (Voelkerding et al., 2009; Alon et al., 2021). Next, a new primer off-set by 1 base (n-1) is added and the procedure of probe annealing, ligation and fluorescence measurements is repeated as explained above (Voelkerding et al., 2009; Alon et al., 2021). In total, 5 rounds are performed (n, n-1, n-2, n-3, n-4) (Alon et al., 2021; Voelkerding et al., 2009). This ensures that the data about every base on the DNA fragment is obtained, even more so, it means that each base of the DNA fragment has been read out twice. The data is then interpreted.

Data processing in ExSeq is carried out using MATLAB software and can be divided into image processing and data analysis (Alon et al., 2021). On the one hand, for image processing, images are first deconvoluted, color correction is performed with the help of the fluorescent beads, the images are used to create a 3-dimensional field, background signal was removed and the single fluorescent spots are detected and registered (Alon et al., 2021). On the other hand, during data analysis the obtained data is interpreted to obtain base sequences for each RNA detected during the experiment (Alon et al., 2021).

2.2 Targeted ExSeq

To perform ExSeq in a targeted approach a few supplementary steps have to be carried out at the beginning of the experiment and a second sequencing platform is used.

Selection of targeted RNAs and probe design

First, it must be determined which RNAs are to be analyzed in the ExSeq experiment (Alon et al., 2021). Next, padlock probes are designed specifically for each of the targeted RNAs (Alon et al., 2021). Per RNA type a barcode is assigned that can be found on the corresponding padlock probe.

Another main feature of these padlock probes is a complementary region that is specific to one of the selected RNA types. Thus, when generating the complementary sequences, the sequence has to be screened for homology with other RNAs in order to prevent off-target binding of the padlock probe (Alon et al., 2021).

Referring to the structure of the padlock probes, a padlock probe is made up of the following elements: (1) a region (16 nucleotides long) complementary to one end of the RNA target region, (2) a barcode at the 5' end for later SOLiD sequencing, (3) an initiator region for the primers of RCA and ISS, (4) a barcode at the 3' end for later Illumina sequencing and (5) a region (16 nucleotides long) complementary to one the other end of the RNA target region (Alon et al., 2021). When binding to their assigned RNA type the padlock probes' ends lie directly next to each other on the RNA target sequence. In situ, the padlock probes can be ligated to form a circularized DNA that can be used for RCA in order to generate the amplicon. Thus, in comparison to the untargeted ExSeq, reverse transcription is not required for library preparation in targeted ExSeq.

Sequencing platform

Another difference between the targeted and untargeted approach is the sequencing platform used. Alon et al. used SOLiD sequencing as well as Illumina sequencing-by-synthesis (Alon et al., 2021). For example, they sequenced RNA in metastatic breast tumor cells using SOLiD (Alon et al., 2021). However, in their experiments on the visual cortex and hippocampus, they failed to apply SOLiD sequencing and

used Illumina sequencing-by-synthesis as an alternative (Alon et al., 2021). After re-embedding has been performed, the steps are (Alon et al., 2021):

Blocking 3' ends of cellular DNA
Terminal deoxynucleotidyl transferase (TdT) is added to the probe to cap free 3' ends to prevent the binding of bases and thus false fluorescence signals during sequencing.

Hybridization of the Illumina primer
To prepare for Illumina sequencing, a primer has to hybridize to the amplicons.

Illumina sequencing-by-synthesis
Illumina sequencing-by-synthesis reads out the barcodes to determine the RNA type. It consists of the steps: elongation, imaging and fluorescence removal. During elongation the fluorescent-labeled complementary base is added to the primer. Each of the four bases A, T, C and G corresponds to a different color. ***Next***, the probe is imaged for the four color channels to detect the fluorescent signals. ***Lastly***, the fluorescent dye gets cleaved from the strand in order to allow for the next base to be added. This procedure is repeated four times.

After sequencing is completed, image processing and data analysis takes place as de-

Table 1. Benefits and Limitations of ExSeq

Benefits	Limitations
• **Higher resolution of RNA detection.** The application of ExM increases the distance between RNA transcripts in situ, thus reducing optical crowding. • **Detailed mapping of RNA in subcellular compartments.** Combining physical isotropic expansion of cellular tissue with ISS greatly enhances the resolution up to the subcellular compartments. For instance, certain RNA transcripts can be detected in the somata, axons, along the dendrites or spines of neurons by ExSeq. • **Accessibility to more RNA transcripts for ISS.** The application of ExM enables better binding accessibility for molecules used during sequencing and thus detection of more RNAs per cell. E.g. ExSeq was able to read out approximately 326 transcripts per hippocampal neuron (Alon et al., 2021; Zhang et al., 2023). Although the total number of RNA transcripts detected might be lower than in other spatial transcriptomic methods, ExSeq has higher RNA detection rates per single cell (Zhang et al., 2023). • **Can be applied in both untargeted and targeted approaches.** The targeted approach allows for the detection of specific RNAs to be studied, whereas in an untargeted approach RNAs are randomly sequenced and no prior selection of target RNAs has to be performed (Alon et al., 2021). The latter enables the unbiased discovery of gene expression in cells. • **Facilitates the discovery of novel RNA splicing isoforms.** In the untargeted approach ExSeq can identify alternative splicing isoforms and previously unknown isoforms can be discovered as reported by Alon et al. • **Detailed Mapping of Cellular Gene Expression in Health and Disease.** Cell types can be identified according to their spatial gene expression patterns and visualized in cell clusters (Zhuang, 2021). This enables a more detailed study of physiological and pathological tissues including cancers. On a cellular level, changes in gene expression pointing towards pathways in development of various diseases can be analyzed (Zhuang, 2021; Vu et al., 2022; Lewis et al., 2021).	• **Lower RNA detection rate than in situ hybridization (ISH).** While the ISH method ExFISH can achieve an mRNA detection of approximately 70%, targeted ExFISH yielded only 62% (Alon et al., 2021). • **Time-consuming.** The completion of an ExSeq experiment takes a few days, with SOLiD sequencing especially time-consuming (Alon et al., 2021). For instance, the sequencing of 20 bases takes approximately 100 hours. • **Requires many reagents and specialized equipment.** ExSeq experiments necessitate a wide range of reagents and highly specialized equipment, which are costly and can limit accessibility. This requirement also increases the complexity of the protocol, demanding specialized training and expertise for successful implementation. • **Ability to read only short transcript lengths inherent to ISS methods.** Sequencing capacity currently still counts below 30 bases per RNA (Lee et al., 2014; Alon et al., 2021). This is due to damage to the RNA transcripts during multiplexing that is induced by laser during the imaging rounds and by the chemical reactions taking place. To improve RNA identification, Alon et al. performed ex situ sequencing immediately after ISS to match the data and achieve more precise RNA identification.

scribed above. RNA signals can be traced to their subcellular compartment (Alon et al., 2021).

3. Applications, Benefits and Limitations

ExSeq has been applied and validated on a number of fixed tissues and cell types, such as the mouse hippocampus, the nematode C. elegans, embryos of Drosophila and HeLa cells (Alon et al., 2021). The benefits and limitations of the ExSeq method are presented in **Table 1.**

4. Conclusions

As spatial transcriptomic studies enable new discoveries about cellular identity and function, many innovative technologies have emerged in recent years. ExSeq as one of the latest developments in ISS techniques bears the potential to significantly contribute to advancements made in transcriptomic science.

The ExSeq method combines ISS with ExM protocols to enlarge cellular structures and to increase accessibility to RNAs for ISS. Thus, the combination of ExM with ISS allows for more RNAs per cell to be sequenced and for better resolution of single RNAs under the microscope, so that RNAs can be traced back to their nanoscale subcellular compartments. Furthermore, ExSeq is applicable to explore cellular RNA in an untargeted approach, as well as to specifically detect certain RNA types in a targeted approach, thus making the method suitable for a variety of research questions. For instance, the untargeted approach could help to uncover previously unknown RNA splicing isoforms and to trace RNAs of unknown function to determine their role in the cell.

Although the development of ISS and ExSeq brings many advancements, there are limitations that are difficult to overcome. For example, the possibility to read out only short transcript lengths is inherent to ISS methods and RNA detection rate is usually lower when compared to ISH methods. Furthermore, another problem is the availability of performing the ExSeq experiment. Many reagents and special equipment are needed to perform ExSeq and the experiment is very time-consuming as it involves many steps to perform expansion, RNA library preparation and finally, multiplexed sequencing. Thus, it remains to be seen if improvements can be made referring to the efficiency and feasibility of ExSeq. In conclusion, ExSeq can enable detailed RNA detection to their subcellular location and has the potential to be used in multiple tissues to allow for new insights into cellular identity, heterogeneity and interactions and could possibly be adapted to inquire other cellular molecules such as DNA or proteins on a nanoscale level.

Abbreviations

APS –	ammonium persulfate
BS(PEG)9 –	PEGylated bis(sulfosuccinimidyl)suberate)
cDNA –	complementary DNA
ddH2O –	double–distilled water
dNTP –	deoxynucleotide triphosphate
EDTA –	thylenediaminetetraacetic acid
ExM –	Expansion Microscopy
ExSeq –	Expansion Sequencing
FISSEQ –	fluorescent in situ sequencing
ISH –	In situ hybridization
ISS –	In situ sequencing
MOPS –	3–(N–morpholino)propanesulfonic acid
PBS –	phosphate–buffered saline
PCR –	polymerase chain reaction
RCA –	rolling circle amplification
SOLiD – tection	sequencing by oligonucleotide ligation and de-
STARmap – mapping	spatially resolved transcript amplicon readout
TdT –	terminal deoxynucleotidyl transferase
TEMED –	tetramethylethylenediamine

Author contributions

AF designed, organized the review, analyzed and summarized the data and wrote the manuscript. OB supervised the work and contributed to the design, organization and writing of the manuscript. AT contributed to the writing of the manuscript.

Funding

OB is funded by a grant of the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-III-P4-ID-PCE-2020-2027, within PNCDI III. The authors would like to acknowledge the funding from the Ministry of Research, Innovation, and Digitization in Romania, under Program 1—The Improvement of the National System of Research and Development, Subprogram 1.2—Institutional Excellence—Projects of Excellence Funding in RDI, Contract No. 31PFE/30.12.2021.

Acknowledgements

The authors would like to acknowledge the excellent environment and support from our host institutions.

Conflict of Interest

The authors declare no conflicts of interest.

References

Alon, S. et al. (2021) 'Expansion Sequencing: Spatially Precise In Situ Transcriptomics in Intact Biological Systems', Science. 371. doi:10.1126/science.aax2656.

Bucur, O. et al. (2020) 'Nanoscale Imaging of Clinical Specimens Using Conventional and Rapid Expansion Pathology', Nat Protoc. 15:1649. doi:10.1038/S41596-020-0300-1.

Chen, F., Tillberg, P.W., Boyden, E.S. (2015) 'Optical imaging. Expansion microscopy', Science. 347:543–548. doi:10.1126/science.1260088.

Garafutdinov, R.R. et al. (2021) 'Rolling Circle Amplification as a Universal Method for the Analysis of a Wide Range of Biological Targets', Russ J Bioorg Chem. 47:1172. doi:10.1134/S1068162021060078.

Ke, R. et al. (2013) 'In situ sequencing for RNA analysis in preserved tissue and cells', Nature Methods 2013 10:9. 10:857–860. doi:10.1038/nmeth.2563.

Lee, J.H. et al. (2015) 'Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues', Nat Protoc. 10:442. doi:10.1038/nprot.2014.191.

Lee, J.H. et al. (2014) 'Highly multiplexed subcellular RNA sequencing in situ', Science. 343:1360. doi:10.1126/science.1250212.

Lewis, S.M. et al. (2021) 'Spatial omics and multiplexed imaging to explore cancer biology', Nature Methods 2021 18:9. 18:997–1012. doi:10.1038/s41592-021-01203-6.

Li, Q. et al. (2021) 'Programming CircLigase catalysis for DNA rings and topologies', Anal Chem. 93:1801–1810. doi: 10.1021/acs.analchem.0c04668.

Shendure, J. et al. (2005) 'Accurate multiplex polony sequencing of an evolved bacterial genome', Science. 309:1728–1732. doi:10.1126/science.1117389.

Voelkerding, K.V., Dames, S.A., Durtschi, J.D. (2009) 'Next-Generation Sequencing: From Basic Research to Diagnostics', Clin Chem. 55:641–658. doi:10.1373/clinchem.2008.112789.

Vu, T. et al. (2022) 'Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis', Nat Commun. 13. doi:10.1038/S41467-021-27798-0.

Wang, X. et al. (2018) 'Three-dimensional intact-tissue sequencing of single-cell transcriptional states', Science. 361. doi:10.1126/science.aat5691.

Zhang, Y. et al. (2023) 'Reference-based cell type matching of in situ image-based spatial transcriptomics data on primary visual cortex of mouse brain', Sci Rep. 13. doi:10.1038/S41598-023-36638-8.

Zhuang, X. (2021) 'Spatially resolved single-cell genomics and transcriptomics by imaging', Nat Methods. 18:18. doi:10.1038/S41592-020-01037-8.

Targeted ExSeq — In Situ Sequencing (Illumina Chemistry). [Accessed October 22, 2023] Available at: https://www.protocols.io/view/targeted-exseq-in-situ-sequencing-illumina-chemist-yxmvmxbd6l3p/v1?step=9.