High-Spatial-Resolution Transcriptomic Map of the Mouse Lymph Node Microenvironment Using Deterministic Barcoding

Archibald Enninful¹, Yang Liu¹, Rong Fan¹

¹Department of Biomedical Engineering, Yale University

Abstract
Spatial transcriptomics is an emerging approach which characterizes gene expression profiles for a more nuanced understanding of biological processes at a tissue level. This offers significant advantages over traditional omics which require the digestion of tissues and subsequent isolation of cells, during which the spatial information is completely lost. Lymph nodes are an integral part of the immune system and an in-depth analysis of its spatial organization will provide useful insights which can be applicable in the development of novel immunotherapies. In this study, the mouse lymph node is characterized using the newly developed microfluidic-based approach, Deterministic Barcoding in Tissue for spatial omics sequencing (DBiT-seq). Unsupervised spatial clustering analysis via the Seurat R package yielded 8 unique clusters, 3 of which were identified as lymphatic muscle cells, macrophages and T and NK cells, respectively. SpatialDE and GO analysis further elucidated the biological processes associated with the patterns of gene expression.

INTRODUCTION
The mouse lymph node is a complex anatomical structure which plays an important role in the adaptive immune response. It can be divided into three main anatomical regions: the outer capsule, the medulla, and the cortex. Lymph nodes serve multiple functions such as serving as filtering proteins and molecules from the afferent lymph node (Clement et al., 2018). The lymphatics have been also implicated in the spread of cancers by serving as a conduit system for lymph fluids and immune cells (Takeda et al., 2019). Lymph nodes are also known to contain a variety of cell types including lymphatic muscle cells, endothelial cells and reticular cells. Given the important role the lymph node plays in both the circulatory and immune systems, it is very necessary to probe and understand its microenvironment. The knowledge obtained could shape immunotherapies and inform how certain cancers are treated. The work done by Stahl et. al, 2016 opened the door to this new field. Prior approaches involved bulk sequencing when the RNA from a bulk tissue sample was analyzed (Burgess et al., 2019; Rodrigues et al., 2019; Butler et al., 2018). A difficulty with this approach is the masking of cellular heterogeneity. Single-cell RNA sequencing resolves this problem by revealing the heterogeneity and sub-population expression since the sequencing is done on a cell by cell basis. Here, using the Deterministic Barcoding in-tissue (DBiT-seq) approached developed by Liu et. al 2019 (bioRxiv, preprint), ST pipeline, Seurat R package, SpatialDE and Web Gestalt, the mouse lymph node tissue sample was probed to unravel its microenvironment and spatial organization⁵,⁶,⁷.

MATERIALS AND METHODS
All experiments were performed in accordance with the Yale Environmental Health and Safety Regulations.

Fabrication of PDMS chips
A series of microfluidic chips were fabricated with 50 parallel microfluidic channels in the center that are 25μm in width (Fig. 1a and 1b). The PDMS chip containing 50 parallel channels was placed directly on a tissue slide and, if needed, the center region was clamped using two acrylic plates and screws. All 50 inlets were open holes ~2mm in diameter and capable of holding ~13μL of solution. Different barcode solutions were pipetted to these inlets and drawn into micro-chan-
nels via vacuum applied to outlets situated on opposite side of PDMS chip. This device is a universal approach to realize spatially defined delivery of DNA barcodes to tissue surface at a resolution of down to 10μm or even better.

**Tissue Preparation (Fig. 1c-d)**

Fresh frozen tissue sections were cut into sections with thickness < 10 μm and placed at the center of a glass slide and stored at -80 °C until use. Sections were taken from refrigerator and allowed to thaw to room-temperature and cleaned using PBS-RI, 1x PBS + 0.05U/μL RNase Inhibitor (Enzymatic, 40 U/μL). To fix, 500μL freshly prepared 4% formaldehyde solution (in 1x PBS) was added onto sections. Tissues were fixed for 20 min at room-temperature, 0.5 % Triton X-100 in PBS was added onto tissue, and tissues were permeabilized for 20 min at room temperature. To clean up, 0.5 X PBS-RI was flashed through each channel. To stop permeabilization, 0.5 X PBS-RI was used to clean up the tissue sections.

**In-tissue Reverse Transcription**

For a 50x50 device, we first made RT mixture with 50μL of 5X RT Buffer, 32.8μL of RNase-free water, 1.6μL RNase Inhibitor (Enzymatic), 3.2μL Superscript RNase Inhibitor (Ambion), 12.5μL of 10 mM dNTPs each (ThermoFisher), 25μL of Maxima H Minus Reverse Transcriptase (ThermoFisher), and 100μL 0.5X PBS-RI. In a 96-well plate, for each well, we added 4μL of above RT mixture. For each of the well, 1μL of 25uM Barcode A was added and mixed with the RT mixture. The final mixture (5μL in total) was injected into each of the 50 inlets of the 1st PDMS and vacuumed through the channels, making sure all channels were filled and incubated at 37 °C for 30 mins.

**Lysis and Sub-library Generation**

Channels were washed using a wash buffer (4 mL of 1X PBS, 40μL of 10% Triton X-100 and 10μL of SUPERase In RNase Inhibitor) for 5 minutes. The 2nd PDMS was peeled off. A small PDMS solution reservoir was used to cover the whole barcoded tissue. (10-20μL should be enough for 25 um resolution, and 50μL for 50 um resolution) lysis solution made of 50μL 1x PBS, 50μL of 2X lysis buffer (20 mM Tris (pH 8.0), 400 mM NaCl, 100 mM EDTA (pH 8.0), and 4.4% SDS) and 10μL of proteinase K solution (20mg/mL), was added into the reservoir. The reservoir was placed in a humidified chamber and incubated at 55˚C for 2 hours to reverse formaldehyde crosslinks. Afterwards, lysate was collected into 1.5 mL tube and kept at -80°C until use.

**Purification of cDNA**

We prepared 40μL Dynabeads MyOne Streptavidin C1 beads (ThermoFisher) per sublibrary by washing them 3x with 800μL of 1X B&W buffer with 0.05% Tween-20 (refer to manufacturer’s protocol for B&W buffer), before resuspending beads in 100μL 2X B&W buffer (with 2μL of SUPERase In RNase Inhibitor) per sample. To inhibit residual proteinase K activity, we added 5μL of 100μM PMSF in ethanol to each thawed lysate and incubated at room temperature for 10 minutes. We added 100μL of resuspended Dynabeads MyOne Streptavidin C1 (ThermoFisher) magnetic beads to each lysate. Binding allowed to occur for 60 min at room-temperature (with agitation on a microtube foam insert). Beads were washed twice with 1X B&W buffer and once with 10mM Tris containing 0.1% Tween-20 (with each wash including 5 min
of agitation after resuspension of beads).

**Template Switch**

Streptavidin beads with bound cDNA molecules were resuspended in a solution containing 44μL of 5X Maxima RT buffer (ThermoFisher), 44μL of 20% Ficoll PM-400 solution, 22μL of 10 mM dNTPs each (ThermoFisher), 5.5μL of RNase Inhibitor (Enzymatics), 11μL of Maxima H Minus Reverse Transcriptase (ThermoFisher), and 5.5μL of 100uM of a template switch primer (BC_0127). Beads were incubated at room temperature for 30 minutes and at 42°C for 90 minutes with gentle shaking.

**PCR and cDNA purification**

After washing beads once with 10 mM Tris and 0.1% Tween-20 solution and once with water, beads were resuspended into a solution containing 110μL of 2X Kapa HiFi Hot-Start Master Mix (Kapa Biosystems), 8.8μL of 10μM stocks of primers BC_0062 and BC_0108, and 92.4μL of water. PCR thermocycling was performed as follows: 95°C for 3 mins, then five cycles at 98°C for 20 seconds, 65°C for 45 seconds, 72°C for 3 minutes. After these five cycles, Dynabeads beads were removed from PCR solution and EvaGreen (Biotium) was added at a 1X concentration. Samples were placed in a qPCR machine with the following thermocycling conditions: 95°C for 3 minutes, cycling at 98°C for 20 seconds, 65°C for 20 seconds, and 72°C for 3 minutes, followed by a single 5 minutes at 72°C after cycling. Once the qPCR signal began to plateau, reactions were removed. PCR reactions were purified using a 0.8X ratio of SPRI beads (Kapa Pure Beads, Kapa Biosystems). 80μL of KAPA Pure Beads was added to the 100 μL fragmented DNA sample and mixed thoroughly by vortexing and pipetting up and down multiple times. The plate/tube(s) was incubated at room temperature for 10 min to bind the DNA to the beads and placed on a magnet to capture the beads. When the liquid became clear, the supernatant was carefully removed and discarded. Keeping the plate/tube(s) on the magnet 200 μL of 80% ethanol was added and the plate/tube(s) incubated on the magnet at room temperature for ≥30 sec. The ethanol was carefully removed and discarded. Beads were dried at room temperature for 3 – 5 min, or until all the ethanol evaporated. Beads were resuspended in 15 uL of elution buffer or PCR-grade water, depending on the downstream application and incubated at room temperature for 5 min to elute the DNA off the beads. Elution time may be extended up to 10 min if necessary, to improve DNA recovery. The plate/tube(s) was placed on a magnet to capture the beads and incubated until the liquid is clear. The clear supernatant was transferred to a new plate/tube(s). Do Qubit. Proceed with your downstream application such as Bioanalyzer, or store DNA at 4°C for 1 – 2 weeks, or at -20°C.

**Library Preparation**

For library preparation, a Nextera XT Library Prep Kit (FC-131-1024) was used (Illumina). 750pg or 1 ng of purified cDNA was diluted in water to a total volume of 5μL. 10μL of Tagment DNA buffer and 5μL of Amplicon Tagment mix was added to bring the total volume to 20μL. After mixing, the solution was incubated at 55°C for 5 minutes. 5μL of NT buffer was added and the solution was mixed before incubation at room temperature for another 5 minutes. 15μL volume of PCR master mix, 8μL of water, and 1μL of each primer (P5 primer and indexed P7 primer) at a stock concentration of 10μM was added to the mix, making a total volume of 50μL. PCR was performed as follows: 95°C for 30 seconds, followed by 12 cycles of 95°C for 10 seconds, 55°C for 20 seconds, and 72°C for 3 minutes, followed by a single 5 minutes at 72°C after cycling. Once the qPCR signal began to plateau, reactions were removed. PCR reactions were purified a 0.7X ratio of Kapa Pure Beads to generate an Illumina-compatible sequencing library.

**DBiT-seq Raw Data Processing**

To obtain transcriptomics data, the Read 2 was processed by extracting the UMI, Barcode A and Barcode B. The processed read was trimmed, mapped against the mouse genome(GRCh38), demultiplexed and annotated(Gencode release M11) using the ST pipeline v1.7.2 (Navarro et al., 2017), which generated the gene expression matrix for downstream analysis. Spatially variable genes were identified by SpatialDE (Svensson et al., 2018b).

**DBiT-seq data analysis using Seurat Pipeline**

Spatially variable genes generated by SpatialDE was used to conduct the clustering analysis as well as the Seurat
package in R. Clustering was done using UMAP. GO analysis was performed using Web Gestalt.

Note: Figures 1(B)-(D) used with permission from Liu et.al, 2019 (bioRxiv preprint) doi: https://doi.org/10.1101/788992

RESULTS

Optical microscope images of mouse lymph node tissue

For a fresh frozen mouse lymph node tissue sample (Figure 2a-d), the entire DBiT-seq workflow was conducted, followed subsequently by H&E staining. The hematoxylin stains the nuclei blue whereas the eosin stains the extracellular matrix pink. From the staining, the dense blue population in the cortical and medullary portions of the lymph node may correspond to immune cell populations like macrophages and T-cells which are known to be present in the lymph node.

Figure 1. DBiT-seq technology and workflow

(A) Microfluidic device used in DBiT-seq, fabricated with 50 parallel 25μm microfluidic channels in the center. Barcodes are pipetted into the 50 inlets and drawn into the microchannels using a vacuum

(B) AutoCAD design of PDMS chip with 25μm width channels

(C) DBiT-seq workflow. The steps involved in a sequential order are tissue preparation, in-tissue reverse transcription, in-cell ligation, lysis and sub-library generation, purification of cDNA, template switching and PCR, cDNA purification, library preparation and sequencing and data analysis

(D) DBiT-seq barcodes. Each unique barcode A1-A50 contains a ligation linker, and a poly-T sequence(16mer), which detects mRNAs and proteins through binding to poly-A tails. Second barcodes B1-B50 include a ligation linker complementary to that of barcodes A producing to a 2D array of spatially distinct barcodes

Figure 2. Optical microscope images of mouse lymph node tissue on a slide.

(A) Mouse lymph node sample

(B) An H&E stained mouse lymph node sample

(C) Mouse lymph node sample with the second direction PDMS chip (25um channel width) covering it

(D) Image of mouse lymph node sample with an overlay of both first and second direction PDMS chip
Assessing data quality

The size, distribution and quality of the cDNA produced during the in-tissue reverse transcription step were analyzed. It was observed that the average size of the cDNA was approximately 1500 bp long with a size distribution of close to 86% (Figure 3a) These results were optimal and what would be expected since it shows the reverse transcription enzyme was not hindered in its processes. The number of genes and unique molecular identifiers (UMI) for each spot (pixel) is shown in Figure 3b and 3c. The average number of genes and UMI was in the range of 200-300. Typically, this method generates >5000 unique transcripts per pixel, >1500 genes per pixel. One possible reason for the lower number of detected genes may be because the tissue sample was of sub-optimal quality.

Spatial Clustering of Genes

After dimensional reduction using Principal Component Analysis (PCA) and clustering using the UMAP algorithm in the Seurat R package, a spatial map of the mouse lymph node was obtained (Figures 4a and 4b). Each tiny square represents a pixel and is the intersection of the first and second direction of the channels. The long vertical and horizontal lines across the tissue may represents artefacts from the flow in the channels. Higher number of genes are observed in the medullary portion of the lymph node sample. From the UMAP clustering, 8 unique clusters were observed. The top genes differentially expressed by each cluster was plotted in a heatmap (Figure 4d). Some of the differentially expressed genes for the clusters included Myh6, Zfp775, Gpc1 and Hint2. Known marker genes for various anatomical regions of the lymph node such as ACKR4, LYVE1, MARCO and MRC1 which represent the ceiling of the subcapsular sinus lymph endothelial cells (cLECs), floor of the subcapsular sinus lymph endothelial cells (fLECs), the medullary LECs and interfollicular sinuses respectively were however not identified in this analysis. However, GO analysis of the biological processes associated with each cluster identified clusters 0, 2 and 5 to be most likely lymphatic muscle cells, macrophages and T and NK cells respectively.

Differentially Expressed Genes and Gene Enrichment Analysis of Clusters

After clustering, each cluster of cells were probed further in the analysis to identify which genes are differentially expressed. Gene ontology analysis was used to identify the specific biological processes each cluster is associated with using the differentially expressed genes. WebGestalt, a web-based platform was used for all such analysis. Identification of the clusters was made based on this analysis. Cluster 5 showed the expression of genes that are associated with positive regulation of the interferon gamma signaling pathway and interferon response. Thus, cluster 5 was most likely...
T-cells and NK cells. The genes associated with Cluster 2 are involved with the positive regulation of macrophage differentiation. Cluster 0 also expressed genes associated with lymphatic muscle contraction. Other biological processes associated with the other clusters included regulation of necroptotic processes, posttranscriptional regulation of gene expression and peptide biosynthetic processes. Further experimentation and identification of gene markers will shed more light on the specific identities of these cell clusters.

Figure 4. Spatial transcriptomic analysis of lymph node tissue sample

(A) Spatial clustering showing 8 unique clusters. Dimensional reduction and clustering done using PCA and UMAP respectively

(B) UMAP clustering showing 8 unique clusters. Clusters 0, 2 and 5 correspond to lymphatic muscle cells, macrophages and T and NK cells, respectively

(C) Gene enrichment analysis showing biological processes associated with each cluster

(D) Gene expression heatmap of the 8 clusters

SpatialDE to Detect Tissue Features

Spatial DE, a recently developed algorithm which identifies spatially variable genes and measures and assigns a score of expression levels as a spatial coordinate of samples was used to identify spatial differential expressed genes. 10 distinct spatial gene expression patterns were identified. GO analysis of the biological process associated with the 10 distinct patterns confirmed the unique biological processes associated with four of the patterns (Figure 5a-d). Pattern 1 genes are associated with muscle contraction which may be correlated with cluster 0 from the UMAP cluster. Pattern 2 is associated with cytokine biosynthetic processes. Patterns 4 and 7 are associated with mRNA splicing and snoRNA localization, respectively.

Figure 5. Spatial differential gene expression of lymph node tissue sample
Using SpatialDE to detect tissue features and number of spatial differentially expressed genes. The associated biological processes unique to each pattern were further explored using WebGestalt. Patterns 1, 2, 4 and 7 genes are associated with muscle contraction, cytokine biosynthetic processes, mRNA splicing and snoRNA localization, respectively.

**DISCUSSION AND CONCLUSION**

The architecture of the mouse lymph node has been shown by various other research groups to have some similarities to that of humans. This thus opens the door to applying the knowledge acquired from research on mouse lymph nodes to research on human lymph nodes which could have profound impacts on immunotherapies. Here, the mouse lymph node tissue sample was analyzed using Deterministic Barcoding in-tissue, developed in the Fan Lab and has been previously applied in studying mouse embryo samples. The number of genes and UMI obtained was approximately 300 per pixel. This number is lower down the >1500 genes per pixel typically observed using the DBiT-seq technology. This could be due to the sub-optimal quality of the lymph node tissue sample used. Downstream analysis using the Seurat R package revealed 8 unique clusters of cells, 3 of which could be positively identified as muscle cells, macrophages and T and NK cells respectively from gene ontology analysis which revealed the biological processes associated with these unique clusters of cells. However, it is worth noting that the typical marker genes, associated with structures in the lymph node such as the fLECs, cLECs and medullary sinuses like ACKR4, LYVE1 and MARCO respectively, were not detected in the preliminary experiments for this publication. However, positive identification of lymphatic muscle cells, macrophages and T and NK cells shows that this project is on the right track. Analysis of the spatial differential gene expression using the SpatialDE pipeline yielded 10 unique patterns. The biological processes associated with four of these patterns were identified using GO analysis. Pattern 1 is associated with muscle contraction which may be correlated with cluster 0 from the UMAP which was identified to be lymphatic muscle cells. These findings are very promising giving the crucial role lymphatics play in normal physiology and even in diseased states such as cancer. Taken together, this work demonstrates an application of the DBiT-seq technology, which has been previously applied to the mouse embryo (Liu et al., 2019) to a new tissue sample (lymph node) with promising results and highlights the immense potential of the emerging field of spatial transcriptomics.

**FUTURE DIRECTIONS**

Based on these early promising results, these experiments can be redone using channels of reduced size (20-10um) in diameter which will greatly improve the resolution. Multiple mouse lymph node samples (inguinal, axillary, cervical) could be used to determine if gene expression patterns and microarchitecture are preserved in samples from different body parts. The data from the mouse lymph node can be compared with human samples to determine if cross-species conservation of gene expression exists. Work done by Takeda et.al, 2019 shows that this might be the case, and this will be explored in future experiments. Fluorescent-labelled antibodies will also shed light on the anatomical locations of the clusters identified and can help in the positive identification of clusters. Comparisons of experimental data from cancerous and non-cancerous lymph nodes from future experiments will shed light on the metastatic changes that occur when cancers spread to the lymph node. Further analysis of the SpatialDE data could identify unique patterns and associated biological processes that can broaden our knowledge of the architecture and biological processes of the mouse lymph node.

**ACKNOWLEDGEMENTS**

This research conducted as part of the Senior Thesis Project (BENG 473 &474) and was made possible with support of my Principal Investigator, Professor Rong Fan, PhD and my mentor, Yang Liu, PhD at the Yale School of Engineering and Applied Sciences. Special acknowledgement to all other members of the Fan Lab and the instructors for the Senior Thesis Project and the Director of Undergraduate Studies for Biomedical Engineering, Professor James Duncan.
REFERENCES


