Transcriptome Profiling of Korean colon cancer by cDNA PCR Nanopore sequencing

Jungwoo Lee[1], Hye-ran Shim[2], Jin-Young Lee[2], Yutae Kim[1], Ji-Yeon Lee[3], Min-Hyuk Jung[2], Won-Young Choi[1], Ji-Hyun Hwang[1], Lark Kyun Kim[3], Young-Joon Kim[1,2]

[1]Department of Integrated Omics for Biomedical Science, Graduate School, Yonsei University, Seoul, Korea,
[2]Department of Biochemistry, College of Life Science & Technology, Yonsei University, Seoul, Korea,
[3]Severance Biomedical Science Institute and BK21 PLUS Project for Medical Sciences, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

Jungwoo Lee
Ljw90607@hotmail.com

Abstract

- We have sequenced 20 pairs of Korean colon cancer patients using **cDNA PCR** Nanopore technology to identify **colon cancer specific novel biomarkers** that can be used for **clinical usage**.
- We attempted to identify **novel isoform, non-coding RNAs and fusion genes**. Some candidates that were found were shown to have potential to be new colon cancer diagnostic biomarkers.
- By using more precise and accurate data to detect such biomarkers, we want not only to be able to use them for clinical usage, but also to further propose a **new molecular mechanism** to better understand colon cancer tumorigenesis using Nanopore data.

Aim: Finding Colon cancer specific biomarker

Sample preparation

- Colon cancer patient tissue
- Collaboration with hospitals

SNUH
Asan Medical Center
SAMSUNG MEDICAL CENTER

Transcriptome analysis

- GridION cDNA-PCR
- Biomarker candidate selection (Isoform, lncRNA, Fusion)

Clinical usage

- Validation of newly found colon cancer specific predictive or diagnosis marker

Result

Fig 1. Quality Control for cDNA PCR Nanopore Data

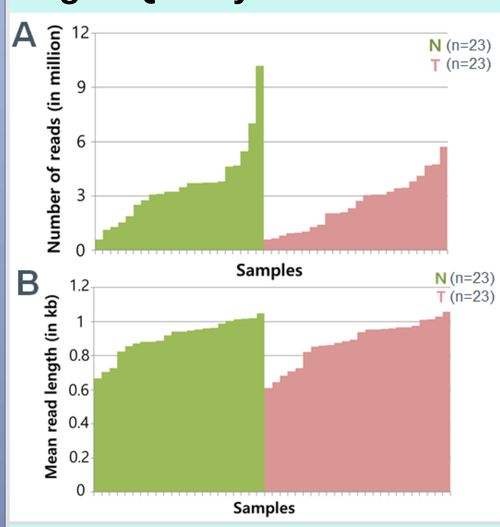


Figure 1.
(A,B) number of reads, mean read length for normal and tumor samples (average 3 million reads, 800bp)
(c) Filtering out the quality reads from pooled data set. (read Q >5, & read bp >200bp)

Fig 2. Profiling of transcripts using SQANTI for isoform detection

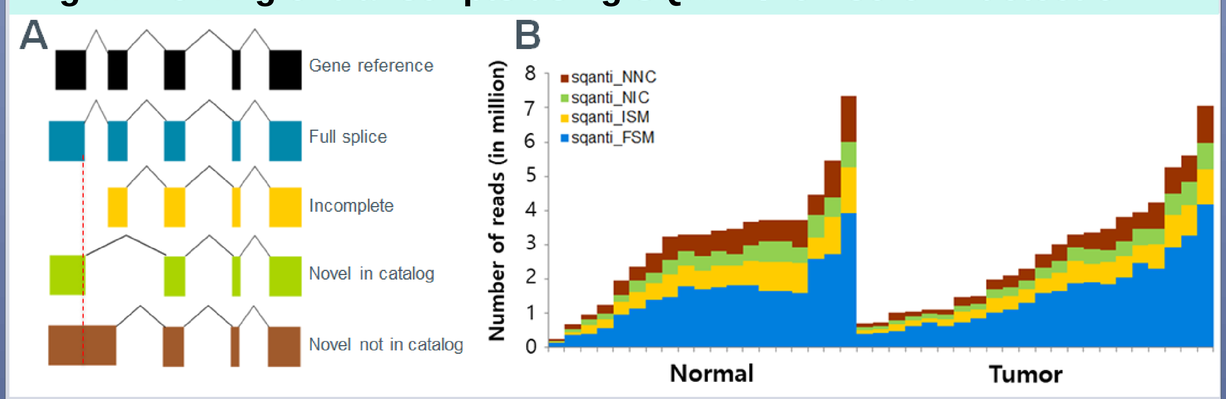


Figure 2. Transcriptomic structure annotation using SQANTI
(A) Transcripts can be divided into 4 structural categories comparing to the gene reference; Full splice, Incomplete splice, Novel in catalog, Novel not in catalog (FSM, ISM, NIC, NNC).
(B) Number of reads categorized to each structural annotation from SQANTI2. Samples in both groups (Normal and Tumor) are sorted in ascending order of the number of total reads within the group.

Reference

1. Tardaguila M. (2018) SQANTI:. Genome Res. 2018. 28: 396-411
2. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**:3094-3100. \
3. Göke J. (2018) Beyond gene expression: long-read RNA-sequencing of the cancer transcriptome. Short talk. https://nanoporetech.com/jp/node/82536#11&
4. Tseng E. (2019) Cupcake ToFU. Tool. https://github.com/Magdoll/cDNA_Cupcake/wiki/Cupcake-ToFU%3A-supporting-scripts-for-Iso-Seq-after-clustering-step

Figure 3. Colon cancer specific isoform identification

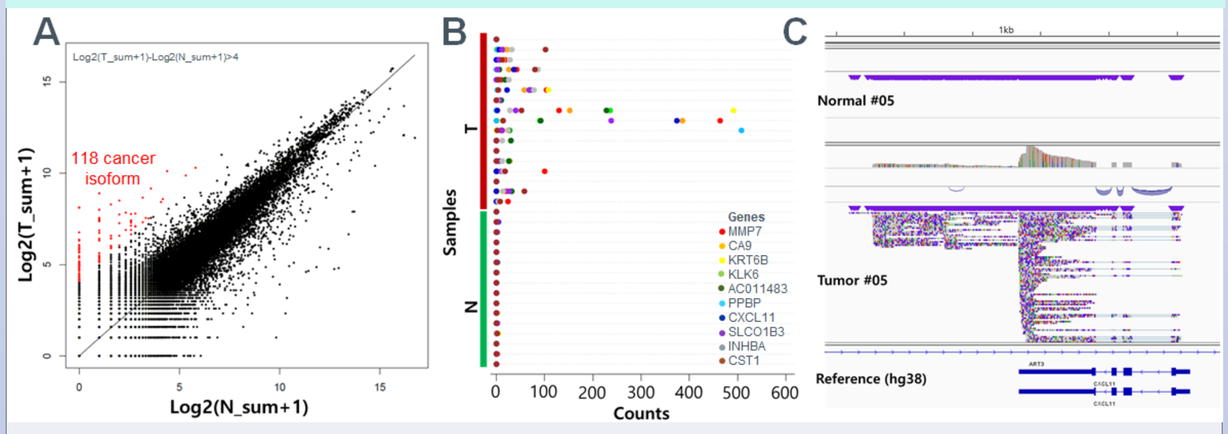
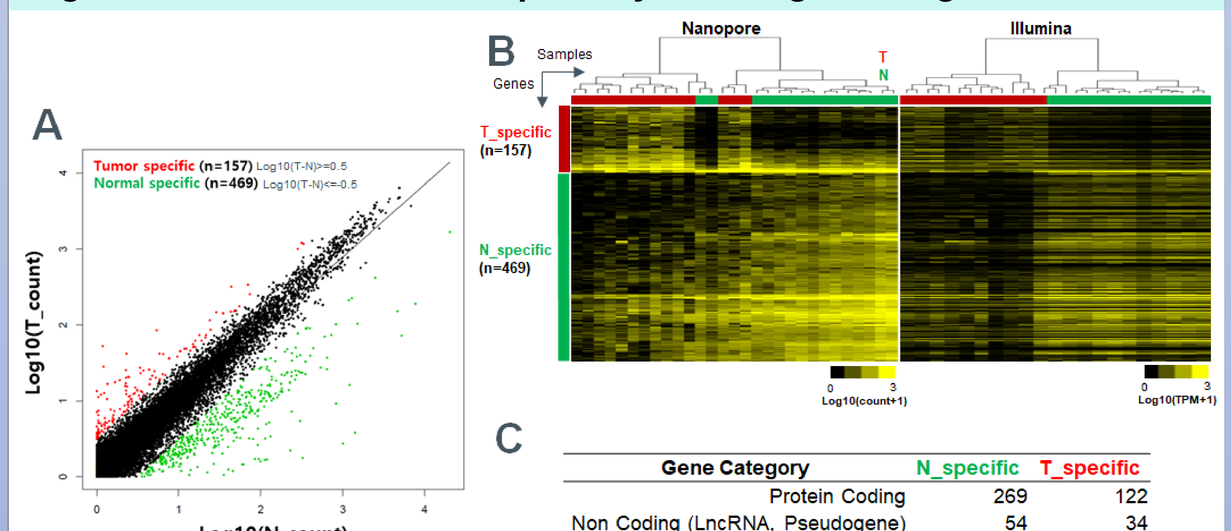


Figure 3. (A) Most differential colon cancer specific isoform (NIC, NNC only) were selected. (log2(T_sum+1)-log2(N_sum+1)>0.4 ~ 16 fold difference)
(B) List of top 10 cancer specific isoform selected (Tumor in red, normal in green). Each dot color represents genes. X axis indicates how many read counts there are in the sample.
(C) CXCL11 as an example of colon cancer specific isoform (in one of the paired samples)

Figure 4. Differential transcript analysis using full length reads



C

Gene Category	N_specific	T_specific
Protein Coding	269	122
Non Coding (LncRNA, Pseudogene)	54	34
Other (IG, undetermined)	146	1
Total	469	157

Figure 4. (A) Tumor and normal specific differentially expressed full length genes (Total averaged by each group then, (log10(T-N)>=0.5, log10(T-N)<=-0.5))
(B) Clustering of N/T group specific genes. Matched gene expression for Illumina is on the right side. **(C)** N/T specific gene categorization by their genomic location.

Figure 5. Differential transcript analysis using full length reads

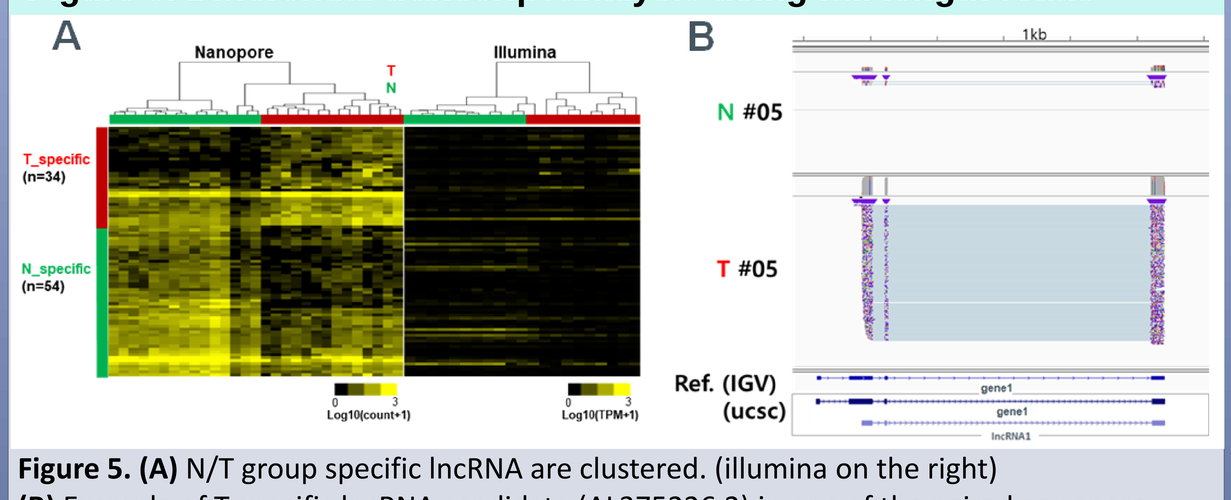


Figure 5. (A) N/T group specific lncRNA are clustered. (illumina on the right)
(B) Example of T specific lncRNA candidate (AL375226.2) in one of the paired nanopore sample.

Figure 6. Detection of fusion gene candidates

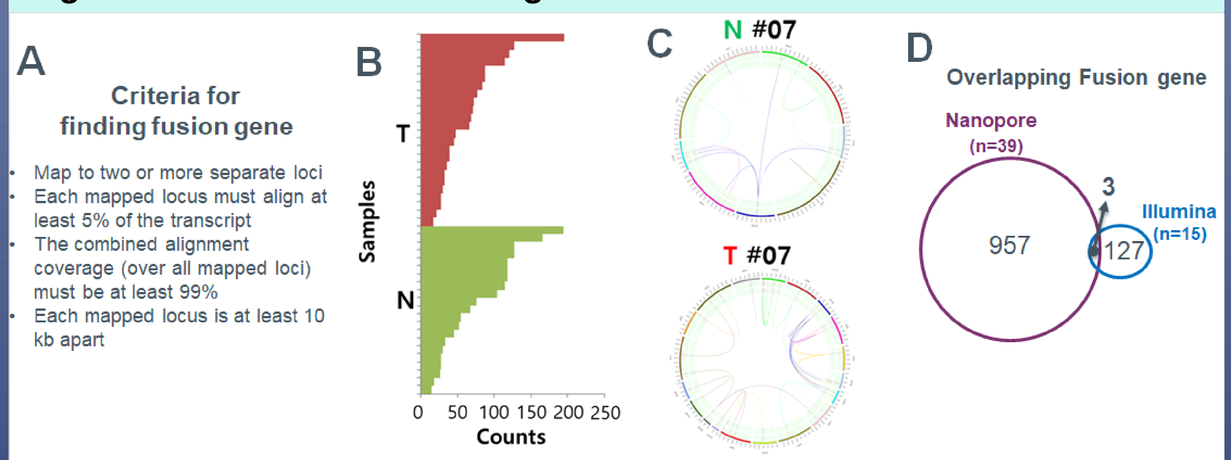


Figure 6. (A) Fusion finding Filtering criteria (applied from Pacbio Fusionfinder algorithm)
(B) Number of fusion counts called from cDNA-PCR nanopore samples (in N/T samples)
(C) Circos plot to show candidate fusion gene event in one of the paired samples
(D) Overlapping fusion gene list between Nanopore & Illumina pooled fusion gene query.

Supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (grant number : NRF-2017M3A9A7050614).