

QuantSeq 3' mRNA-Seq Service – 최적의 DEG 분석 방법

Differentially Expressed Genes (DEG) 분석은 두 샘플 이상의 시료에서 transcriptome sequencing 이후에 샘플간, 그룹간 유전자 발현 증감 및 조절 양상의 차이를 비교하는 방법입니다. 1990 년대에 등장한 Microarray chip 은 쉽고 빠르게 DEG 분석을 할 수 있게 해주었으나, 각 제조사마다 디자인한 probe 정보가 다르고, chip 위에 설계되어 있는 probe list 에 한해서만 분석할 수 있는 닫힌 시스템으로써 한계를 가지고 있으며, 또한 hybridization 을 통한 형광을 detection 해야 하기 때문에 높은 퀄리티의 시료가 필요했습니다. 우리는 NGS(Next Generation Sequencing)를 통해 이러한 한계점 없이 분석가능한 방법을 찾아냈습니다.



그것이 바로 QuantSeq 3'mRNA-Seq 서비스이며, 이바이오젠에서는 poly(a) tail 이 존재하는 생물 종(species)을 대상으로 total RNA 를 분리하고 mRNA 3' UTR 부분을 짧게 library 로 제작, sequencing 하여 유전자 발현 분석을 진행하는 실험 기법이며, 오랜 기간 서비스하고 있습니다. QuantSeq 3' mRNA-Seq 서비스는 퀄리티가 높은 RNA 샘플 뿐만 아니라 FFPE(Formalin-Fixed, Paraffin-Embedded) 등과 같이 상대적으로 퀄리티가 낮은 RNA 샘플에서도 수행 가능한 프로토콜을 제공하며, transcript 마다 하나의 read 조각만을 생성하여 gene expression 분석 목적으로 빠르고 저렴하게 데이터를 확인할 수 있는 최적의 방법입니다.

Whole Transcriptome Library for Expression Profiling



Assembly, Align and RPKM



Many reads for 1 transcript are needed

Expression Profiling Library for Expression Profiling



Only 1 read for 1 transcript is needed if we can sequence the same positions of all different transcripts.

QuantSeq 3' mRNA-Seq Service Process

Preparation

최소 10ng 이상의 total RNA 시료로 빠르고 간단하게 library 제작이 가능합니다. QuantSeq library prep kit 는 poly(a) capture 나 rRNA depletion 과정 없이 진행되며, oligo dT priming 과 random priming 을 통한 3' UTR 부분의 sequence 에 초점을 두기 때문에 FFPE 와 같은 퀄리티가 좋지 않은 시료에도 적용 가능한 장점이 있습니다. 3' UTR 부분의 영역은 유전자마다의 특이적인 서열로 유전자를 식별하는 데 사용될 수 있는 영역입니다.

3' Untranslated region

(noun) /θri: prəm ,ʌntɹæns 'leitɪd 'ri: dʒən/

3' UTR 은 mRNA 영역 중 종결 코돈 다음에 위치하는 영역이다. 3' UTR 의 서열과 길이는 유전자마다 매우 다양하지만, poly(A) tail 영역 근처에서 몇 가지 공통적인 motif 를 가지고 있다. 3' UTR 은 각 유전자를 식별하는 데 사용할 수 있을 뿐만 아니라 다른 연구에서 poly(A) tail 영역을 재설정 할 때도 사용할 수 있다. 결론적으로, 3' UTR 은 각 유전자마다의 특이적인 서열이라고 할 수 있다.

Sequencing

QuantSeq library prep 과정은 transcript 마다 단 하나의 read 를 생성하고, 그 짧은 read 만을 sequencing 합니다. Illumina 의 NextSeq 550 장비를 이용하여 single end 75bp 를 읽어내고, 약 10M read 의 데이터를 생산합니다.

Data Processing

Sequencing raw data 를 Q20 의 기준으로 read 를 trimming 하고 Bowtie2 를 이용하여 reference genome 에 mapping, raw count 를 도출하고 Edge R 을 이용하여 TMM + CPM 으로 normalization, 이후 정리된 expression data 를 제공합니다.

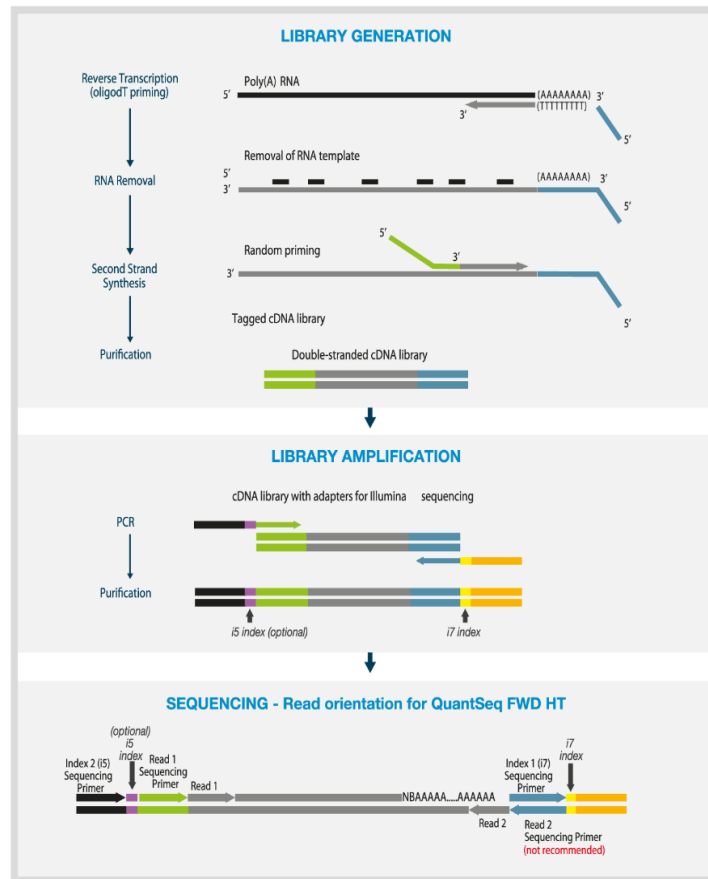
Analysis

정리된 expression data 는 이바이오젠 분석툴인 ExDEGA (Excel based Differentially Expressed Gene Analysis)를 통해 제공되며, DEG 분석 및 annotation 정보, align summary 등을 확인할 수 있습니다.

QuantSeq 3'mRNA-Seq Service Process

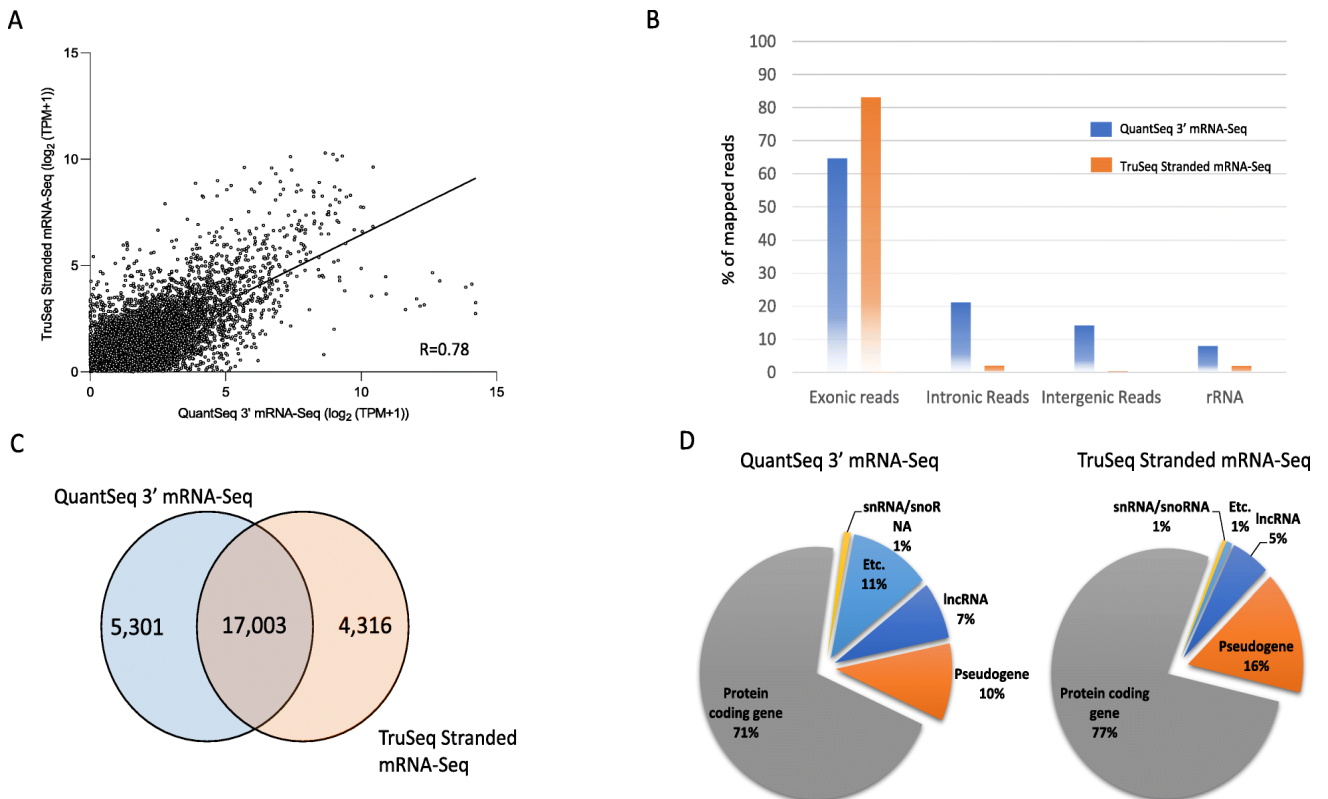


Library Construction Workflow



QuantSeq 3' mRNA-Seq Service Paper

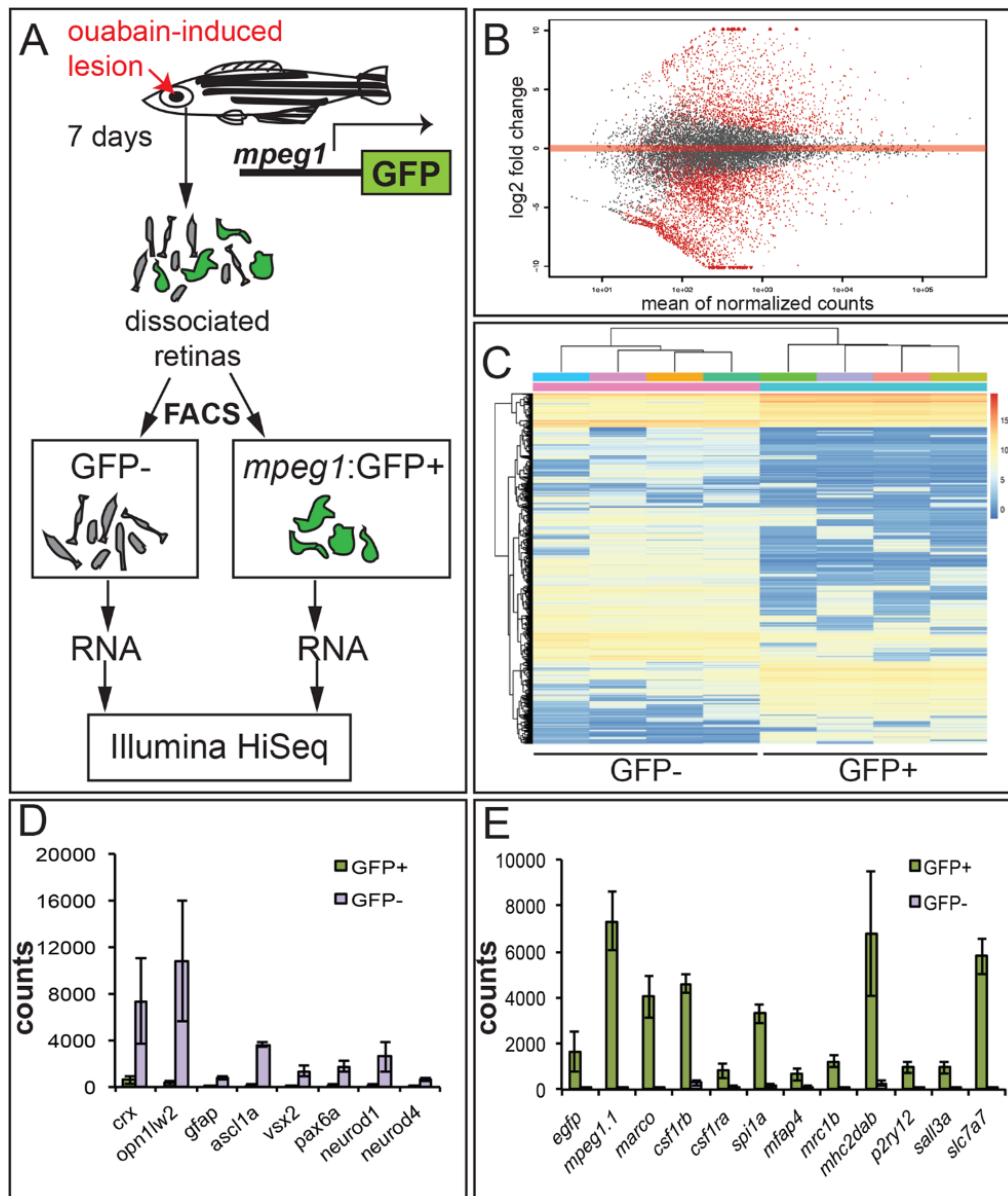
Application of QuantSeq / Correlation with mRNA-seq



위의 자료는 Illumina의 Truseq stranded mRNA-Seq 데이터와 Lexogen의 QuantSeq 데이터 비교 자료입니다. 동량의 RNA input을 했을 때 mapping rate 정보와 mapping된 RNA read의 종류들을 확인할 수 있습니다. Truseq stranded mRNA-Seq에 비해서 단백질 코딩 유전자 mapping rate가 상대적으로 낮지만, 17,003개의 유전자를 공유하는 것을 확인할 수 있으며, R=0.78의 상관관계를 보이고 있습니다. 이외에도 FFPE와 fresh frozen tissue의 상관관계, FFPE 샘플에서의 input 양에 따른 상관관계, RNA exome capture 데이터와의 상관관계를 통해서 QuantSeq 데이터의 gene expression 분석에 대한 재현 가능성을 보여주었습니다.

Discovery of Transcriptional Signature / Low input RNA

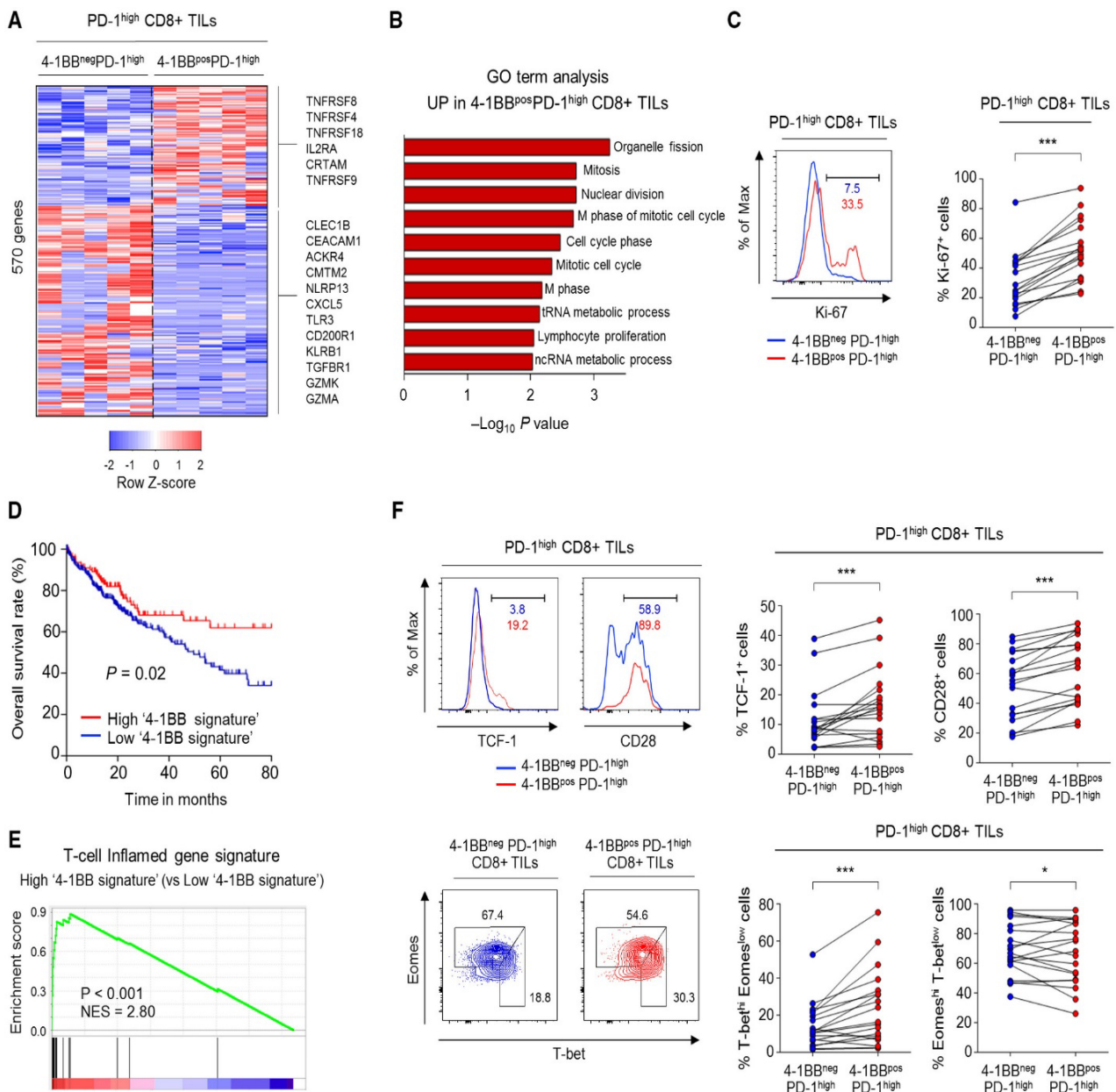
아래의 자료는 zebrafish 망막 조직에서 dissociation 한 세포들을 FACS sorting 하여 GFP negative 세포와 GFP positive 세포를 비교한 자료입니다. GFP positive 세포는 mpeg1 positive 로, microglia / macrophage 에서 유전자 발현 패턴이 어떻게 달라지는지에 대한 연구를 진행하였습니다. 이를 통해 zebrafish의 여러가지 초기 망막 조직 손상, 세포 사멸 기간 후에 조직 재생에 대한 새로운 데이터를 발견하였습니다. 이 실험은 zebrafish의 GFP positive, negative 세포를 FACS sorting 을 통해 분리하여 굉장히 적은 양의 RNA(14~29ng)로 진행하였는데, 이것은 QuantSeq의 아주 큰 장점이라고 할 수 있습니다.



DEG analysis Specialized / E-biogen ExDEGA analysis tool

이바이오젠에서는 QuantSeq 레포트시에 expression data 를 자체 제작한 ExDEGA analysis tool 을 함께 제공하고 있으며, reference genome 에 mapping 한 gene symbol 과 annotation 을 확인할 수 있습니다. 각 gene symbol 의 fold change, normalized data, p-value, raw count 정보를 한눈에 보기 쉽게 정리하였고 보고자 하는 수치의 조건을 연구자의 목적에 맞게 설정하여 필터된 리스트의 데이터를 비교, 분석할 수 있습니다.

또한 데이터를 단순히 정리하여 드리는 것만이 아닌 Gene Ontology editing 을 통한 GO 분석, Scatter / Volcano plot 과 Venn Diagram 을 쉽고 간편하게 제작할 수 있으며, third party support 를 제공함으로써 KEGG mapper, DAVID functional annotation, Clustering Heatmap, GSEA 분석 input data 를 클릭 한번으로 생성할 수 있습니다.



위와 같은 분석, 그림은 ExDEGA analysis tool 을 통해 만나 보실 수 있으며, 이바이오젠에서 서비스하고 있는 QuantSeq 은 DEG analysis 를 목적으로 최상의 데이터를 제공하고 있습니다.

다만, QuantSeq method 는 transcript 하나당 하나의 짧은 read 만을 sequencing 하여 분석하기 때문에 transcript level 이 아닌 gene level 에서의 DEG 분석만 가능합니다. 따라서 transcript level 까지의 분석이 필요한 분야, 예를 들어 isoform 분석이나 alternative splicing event 분석, novel transcript, gene fusion 과 같은 분석이 필요한 경우에는 QuantSeq 이 아닌 mRNA-seq 을 통해 제공이 가능합니다.

E-biogen's QuantSeq 3'mRNA-Seq Service

Service Information

Sample Requirement	>10ng total RNA
Library Method	QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lexogen, Cat#015)
NGS Sequencer	Illumina NextSeq 550, SE 75bp
Data Yield	~10M reads / sample
Turnaround Time	About 3 weeks after RNA QC
Possible Sample Type++	Cell, Tissue, Paxgene Whole blood, FFPE, Sorted Cell, etc.

++ 샘플의 양, 손상정도, 보관상태에 따라 RNA quality 를 check, QuantSeq 진행가능여부 확인이 가능합니다.

< 참고 문헌 >

1. Lexogen, *QuantSeq-Application-Note_2023-01-04, 2023*
2. Lexogen, *QuantSeq 3'mRNA-Seq Library Prep Kit FWD*, <https://www.lexogen.com/quantseq-3mrna-sequencing/>
3. E-biogen, *2023-2024 Service&Product Catalog*
4. JANG Jin Sung, et al. *Application of the 3'mRNA-Seq using unique molecular identifiers in highly degraded RNA derived from formalin-fixed, paraffin-embedded tissue. BMC genomics, 2021*
5. Mitchell, D.M., Sun, C., Hunter, S.S. et al. *Regeneration associated transcriptional signature of retinal microglia and macrophages. Scientific Reports, 2019*
6. Kim, Hyung-Don, et al. *4-1BB delineates distinct activation status of exhausted tumor-infiltrating CD8+ T cells in hepatocellular carcinoma. Hepatology, 2020*