Native Elongating Transcript Sequencing Provides Dynamic Picture of Transcription

Briana Mittleman, Kristen Patterson, Sebastian Pott, Yang Li, Yoav Gilad

Section of Genetic Medicine; Department of Human Genetics; Committee on Genetics, Genomics and Systems Biology

Introduction

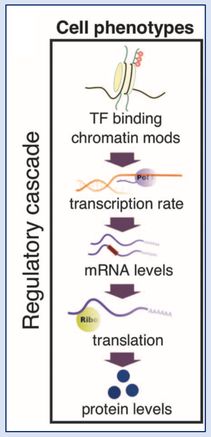


Figure adapted from Li et al. 2016

A major goal in human genetics research is to understand the mechanisms by which the genome is regulated. To answer regulatory genomics questions, we utilize genome wide assays of phenotypes along the regulatory cascade with next generation sequencing technologies. By connecting natural genetic variation to these phenotypes at a genome wide scale, we can start to understand how genetic variation directs gene regulation. Many studies in the Gilad lab have focused on assaying gene expression using RNA-seq, chromatin accessibility through ATAC sequencing, and histone modifications with CHIP-seq (1). While these assays have helped us learn about how genetic variation impacts steady state RNA, we have not been able to reliably model transcription along genes as a dynamic model. By using Native Elongating Transcript Sequencing (NET-seq) to sequence RNA transcripts currently being transcribed in the cell, I will gain a new perspective on the transcription process and therefore gene regulation.

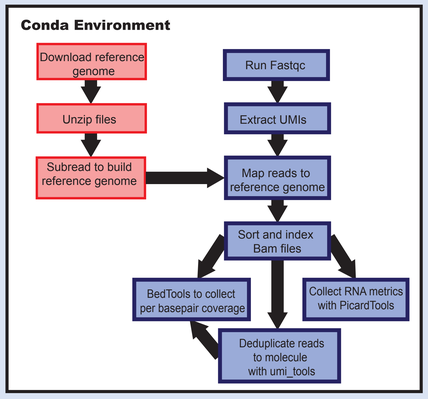
Data processing with Snakemake on Midway2

```
#Snakefile

#path to data:
dir_proj= config["dir_proj"]

#rules
rule example:
  input: "file.fastq"
  output: "file.bam"
  shell: "command_to_run {input} > {output}"

rule all:
  input:
    dir_proj + "file.bam"
```

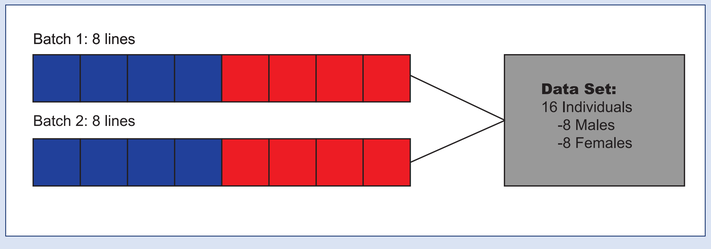


Left: Snakemake (3) is a rule based workflow management tool. **Right:** My Snakemake workflow to process raw NET-seq reads. This allows for parallel job running and prevents file recreation if a file is used multiple times in the workflow.

Initial Data Visualization

Mapped reads can be visualized using the IGV browser. Most reads are transcribed on the coding strand (red reads). We also see antisense transcription at the TSS of ACTB (blue reads).

NET-seq Protocol and study design



Right: The Mayer et al. 2016 protocol (2) halts transcription with a compound called α-Amanatin, then isolates the transcripts that were in the process of being transcribed. **Above:** I performed NET-seq on 16 human Lymphoblastoid cell lines from the HapMap project in 2 batches. The advantage of using these cell lines is they have been previously used for many assays in the lab. I will be able to perform combined analysis between the NET-seq data and other data types

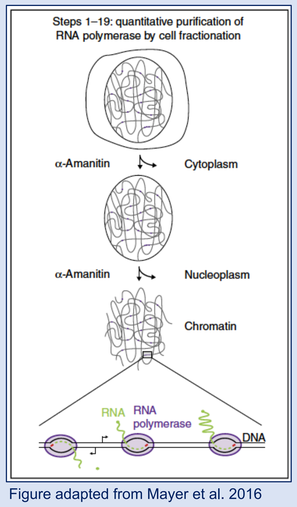
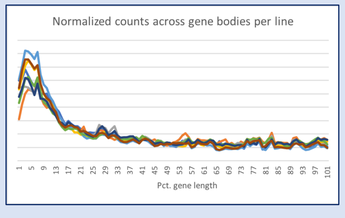
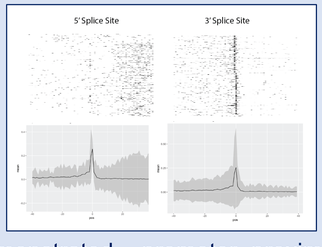


Figure adapted from Mayer et al. 2016

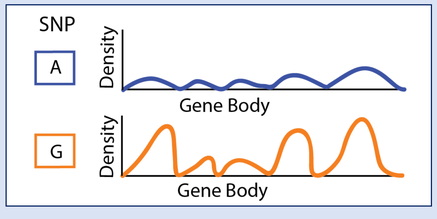
QC and preliminary results





Left: Mayer et al. 2015 (4) demonstrated promoter proximal pausing of PolII. As expected under this model, we see a 5' bias in NET-seq reads at a gene level. In addition, under the kinetic coupling model, PolII pausing also occurs at splice sites upon exons being splice out of the mRNA (4,5). **Right**: Our NET-seq data shows enrichment at the 3' and 5' splice sites.

Future Directions



The preliminary data comes from half of the data (8 individuals). With the second 8 individuals we will be able to detect genetic variants that correlate with differences in NET-seq read counts.

References: 1. Li et al. "RNA splicing is the primary link between genetic variation and disease." Science 2016. 2. Mayer, Andreas and Churchman L Sterling. "Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing." Nature Protocols 2016. 3. Köster, Johannes and Rahmann, Sven. "Snakemake - A scalable bioinformatics workflow engine". Bioinformatics 2012. 4. Mayer et al. "Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution." Cell 2015. 5. Kornblihtt, "Alberto R. Chromatin, transcript elongation and alternative splicing." Nature Structural & Molecular Biology 2006.

Supported By: NIH Genetics and Regulation Training Grant (T32 GM07197)