Research Update

Edmund Miller

2021-01-27 Wed

Overview of GRO-Seq pipeline

Global Transcription

Research

Global transcriptional activity dynamics reveal functional enhancer RNAs

Yoon Jung Kim,[1,2] Peng Xie,[1,2] Lian Cao,[1] Michael Q. Zhang,[1] and Tae Hoon Kim[1]

[1]*Department of Biological Sciences and Center for Systems Biology, University of Texas at Dallas, Richardson, Texas 75080, USA*

Active enhancers of the human genome generate long noncoding transcripts known as enhancer RNAs (eRNAs). How dynamic transcriptional changes of eRNAs are physically and functionally linked with target gene transcription remains unclear. To investigate the dynamic functional relationships among eRNAs and target promoters, we obtained a dense time series of GRO-seq and ChIP-seq data to generate a time-resolved enhancer activity map of a cell undergoing an innate antiviral immune response. Dynamic changes in eRNA and pre-mRNA transcription activities suggest distinct regulatory roles of enhancers. Using a criterion based on proximity and transcriptional inducibility, we identified 123 highly confident pairs of virus-inducible enhancers and their target genes. These enhancers interact with their target promoters transiently and concurrently at the peak of gene activation. Accordingly, their physical disassociation from the promoters is likely involved in post-induction repression. Functional assessments further establish that these eRNAs are necessary for full induction of the target genes and that a complement of inducible eRNAs functions together to achieve full activation. Lastly, we demonstrate the potential for eRNA-targeted transcriptional reprogramming through targeted reduction of eRNAs for a clinically relevant gene, *TNFSF10*, resulting in a selective control of interferon-induced apoptosis.

[Supplemental material is available for this article.]

Recap

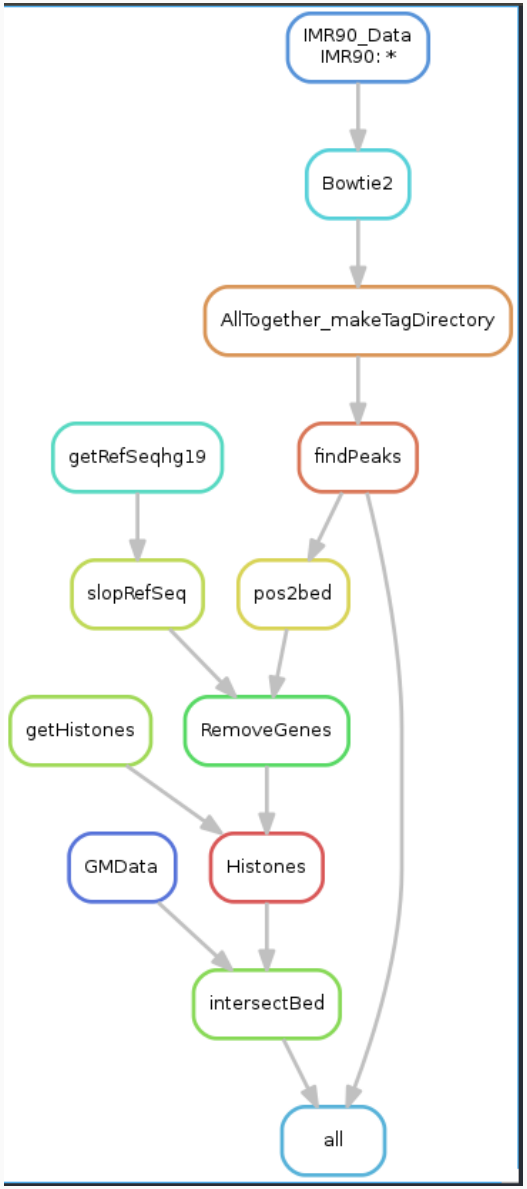
Overview

- Reproducing GM18
- Predicted IMR90 eRNAs
- Compared IMR90 Predicted Enhancers to GM
- Used Homer scripts to find DE of eRNAs and Genes
- Gene Centric vs. Enhancer Centric

Reproducing GM18

- hg18 vs hg19
- Overpredicting eRNA transcripts
- Past Issue
    - What I thought Peng sent me
    - hg18 -> eRNAs -> Me
    - What actually happened
    - hg18 -> eRNAs -> LiftOver -> hg19 -> Me
- Main issue is homer uniqmap

DAG of workflow

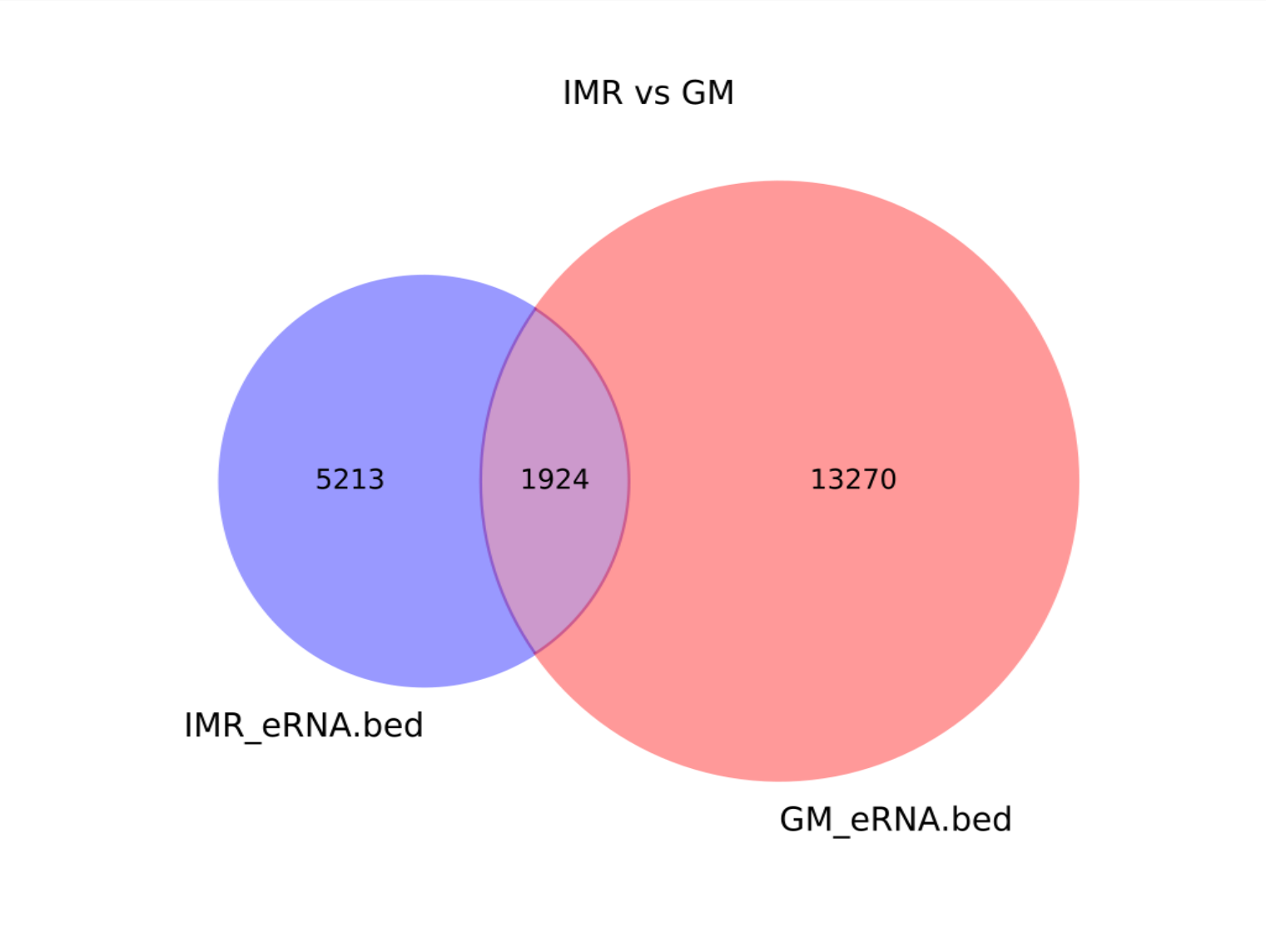


Enhancer Transcript Identification

Changes from GM18

- hg19
- No liftover

IMR/GM enhancer transcript overlaps

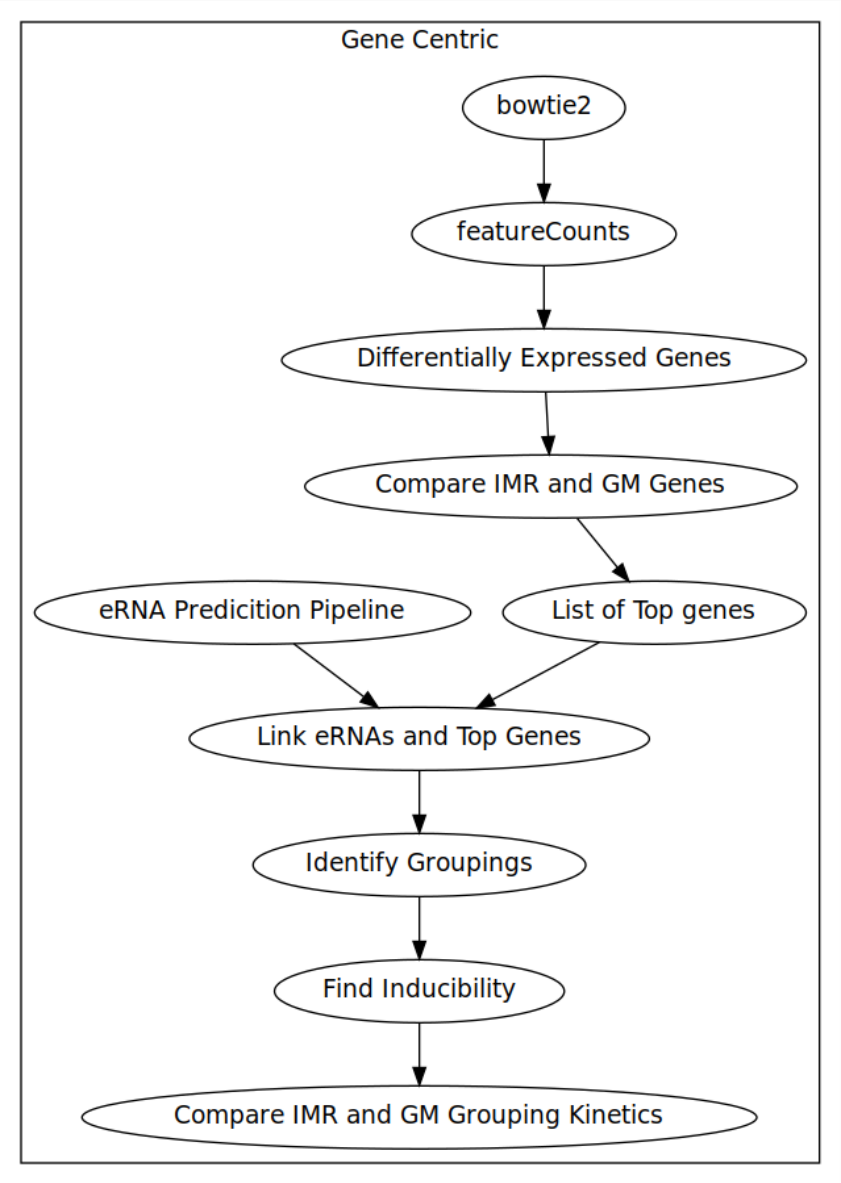


IMR/GM enhancer transcripts linked to DGEs

Gene Centric vs. Enhancer Centric

- Peng's approach
  - Took enhancers that were expressed deferentially
  - Linked them to Genes within 200Kb
- New approach
  - Find genes that are deferentially expressed
  - Link the Enhancers to those genes

Gene Centric vs. Enhancer Centric



nf-core

Standardizing Snakemake

- January 2020
- Template
- Universal Commands
- Testing
- CI/CD
- Wrappers

nf-core Paper

Correspondence | Published: 13 February 2020

The nf-core framework for community-curated bioinformatics pipelines

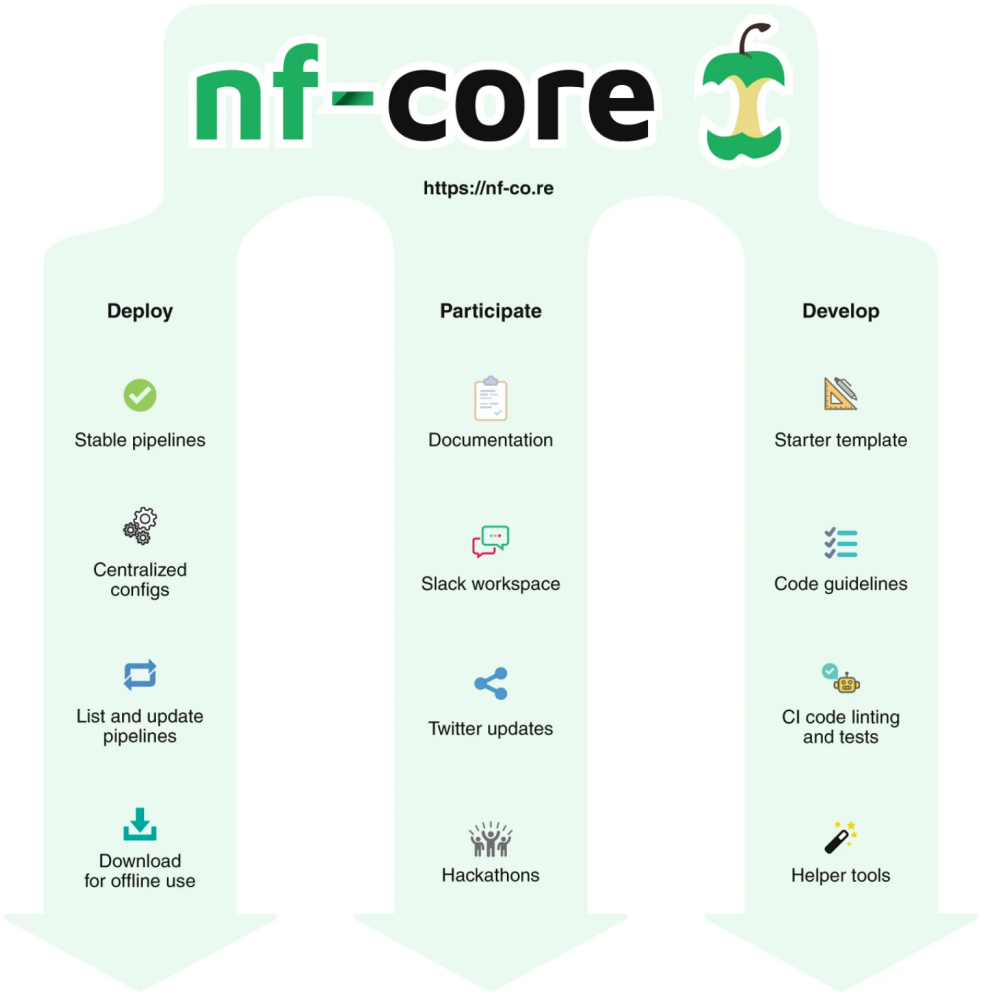
Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso & Sven Nahnsen

Nature Biotechnology **38**, 276–278(2020) | Cite this article

nf-core Features

Fig. 1: Main concepts of nf-core.

From: The nf-core framework for community-curated bioinformatics pipelines



Best-practice pipelines are available to be deployed on virtually any computational infrastructure. Community-built tools help pipeline developers to create new pipelines and adhere to nf-core guidelines. Slack, Twitter and events such as hackathons allow both users and developers to actively participate in the nf-core community. CI, continuous integration.

nf-core Getting started

```
# Install nextflow
curl -s https://get.nextflow.io | bash
mv nextflow ~/bin/

# Launch the RNAseq pipeline
nextflow run nf-core/rnaseq \
    --input samplesheet.csv \
    --genome GRCh37 \
    -profile docker
```

nf-core Contributions

- Rewriting modules that weren't dogfood
- Added indepth testing for modules using pytest-workflow
  - Testing in Scientific Research and Academia - Martin Héroux - Test & Code : P...
- Coming Soon
  - Pytest-workflow tests for pipelines

Example pytest-workflow

```
- name: Run fastqc paired-end test workflow
  command: nextflow run ./tests/software/fastqc/ -profi
  tags:
    - fastqc
  files:
    - path: output/test_paired_end/test_1_fastqc.html
    - path: output/test_paired_end/test_2_fastqc.html
    - path: output/test_paired_end/test_1_fastqc.zip
    - path: output/test_paired_end/test_2_fastqc.zip
```

nf-core Improvements to GRO-Seq pipeline

- Handling of multiple genomes
- Cluster configs
- SRA-download
- Plugging in down stream analysis

Future

- Adding Total Functional Score of Enhancer Elements (TFSEE) Model
  - ChroHMM alternative
- Integrate genomic data indicating open regions of chromatin (ATAC-seq, DNase-seq, or MNase-seq)
- Applying the pipeline to our datasets
- Applying the pipeline to outside datasets