nf-core/bytesize: nascent

Edmund Miller / @e_miller88 / @emiller88

THE UNIVERSITY OF TEXAS AT DALLAS

Overview

- Background on Nascent Transcript Identification
- History of Development
- Pipeline Overview

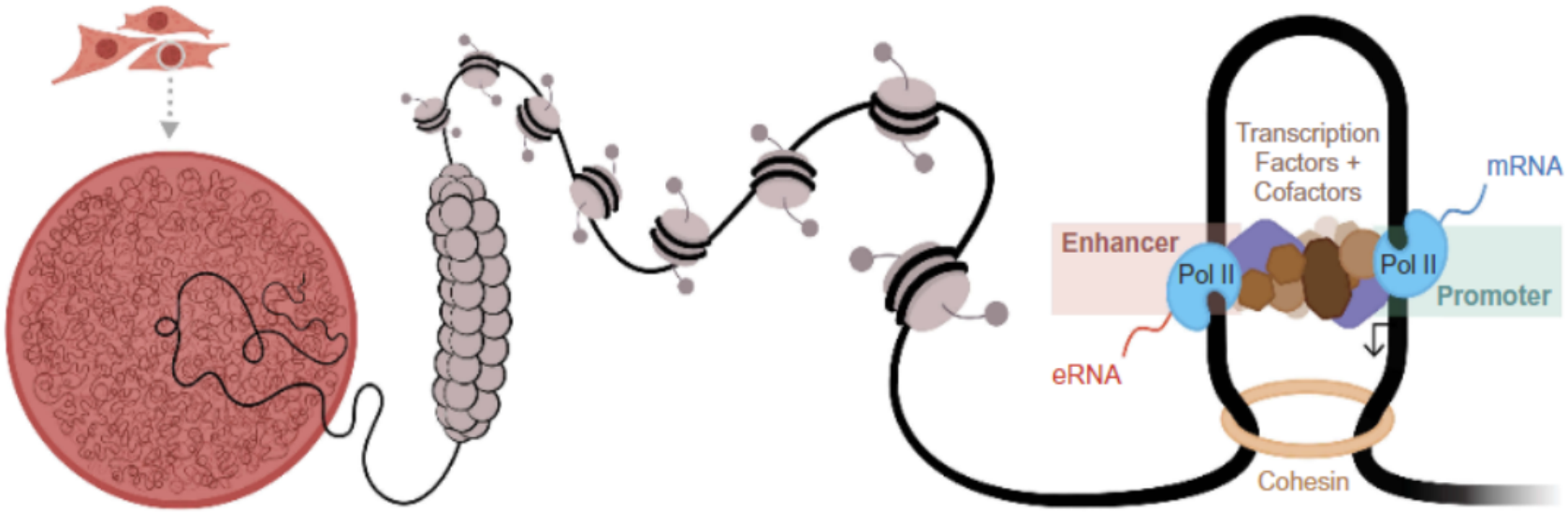
Background on Nascent Transcript Identification

- Goal is to identify changes in transcription
  - Rather than RNAseq which isolates all of the RNA in cell
  - Pulling out the transcriptionally activate sites **through metabolical labelling**
- Covering a lot of different assays
- Slight variation in computational pipeline can result in **25% change in results**

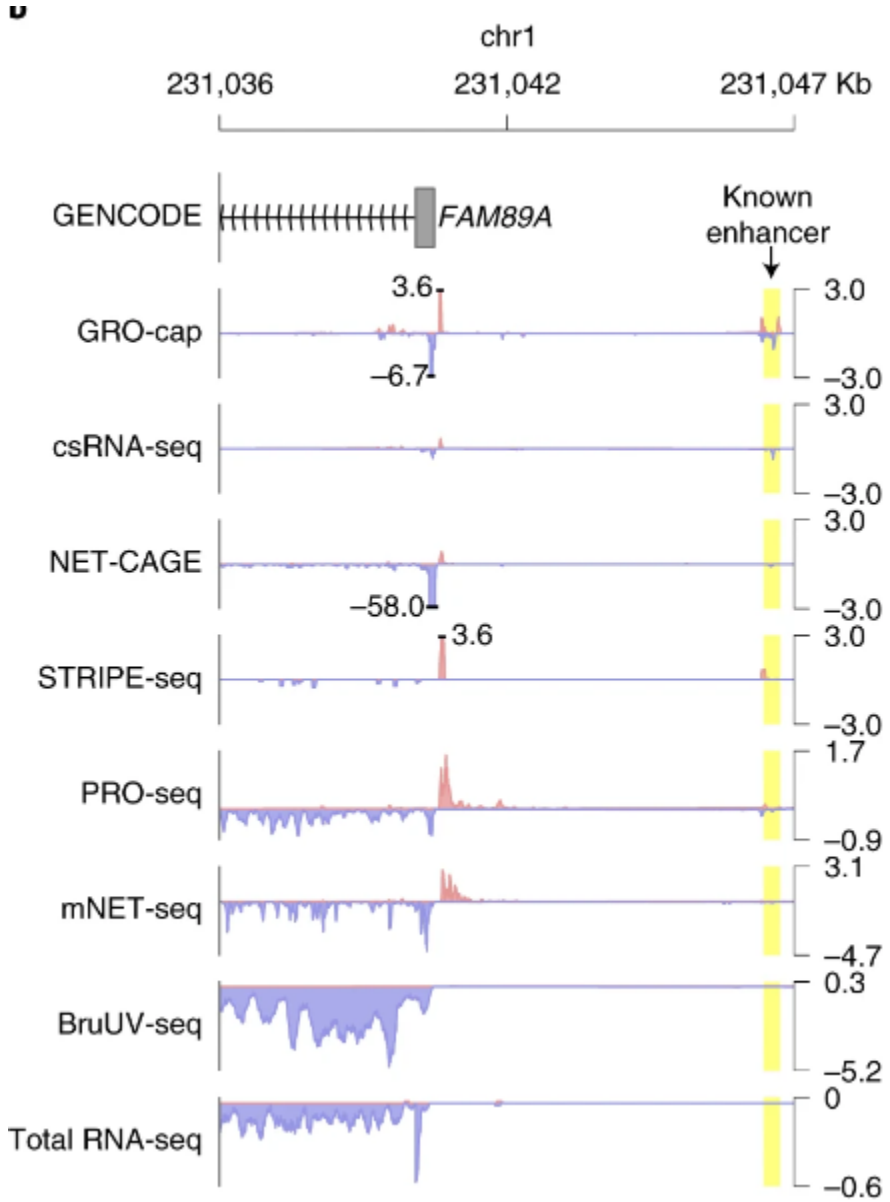
Enhancers

- Cis-acting DNA sequences that can increase the transcription of genes
- The human genome contains hundreds of thousands of enhancers
- Evidence of Enhancer-Promoter interaction from cross-linking assays (3c)
- enhancer RNAs (**eRNA**) have a short half-life so in low abundance

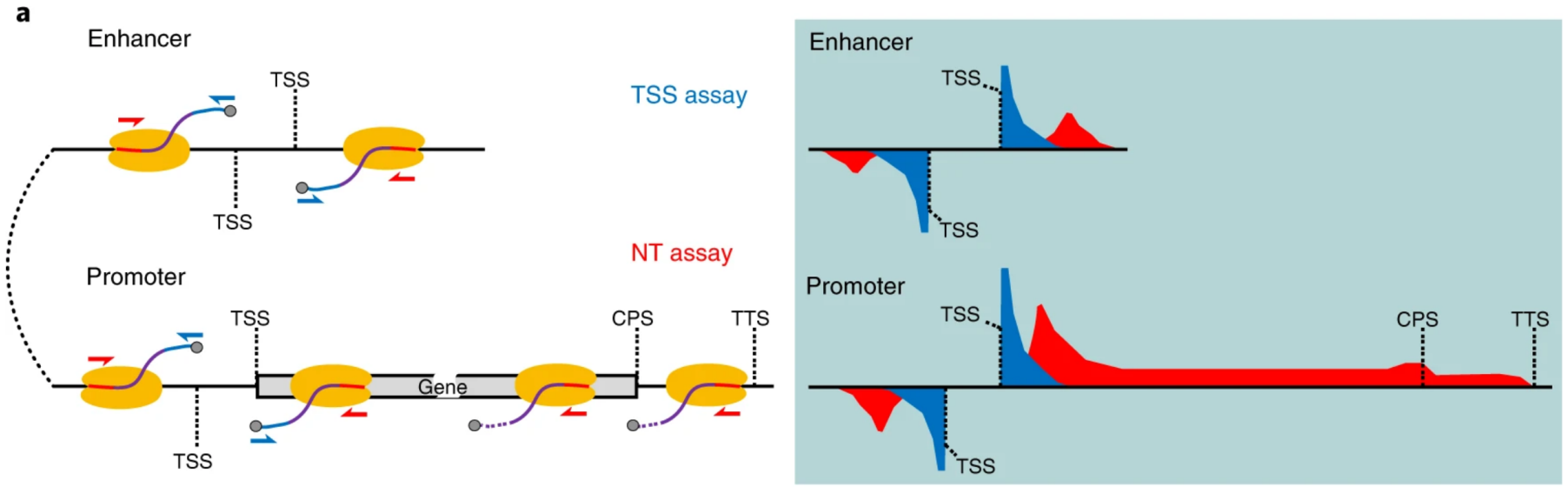
Enhancer-promoter Looping



What the Reads Look Like



Nascent Transcript(NT) and Transcription Start Sites(TSS)

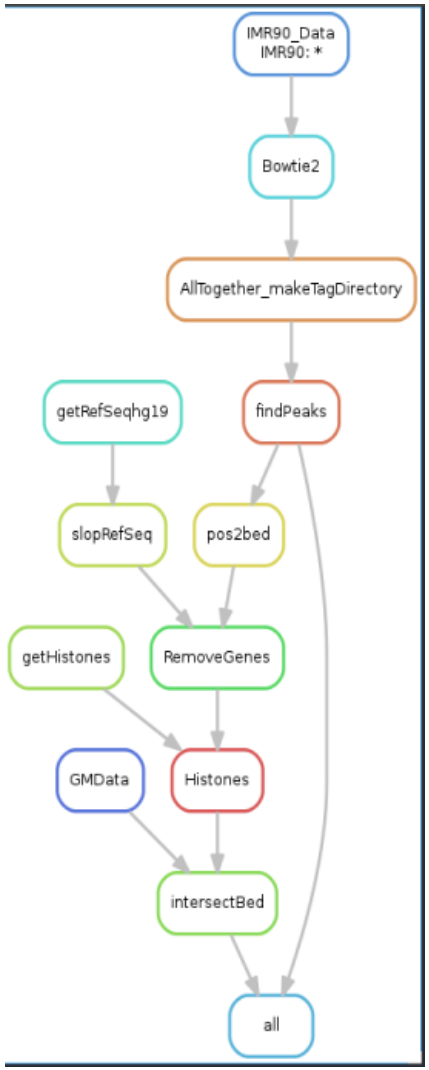


- Currently 13+ assays for nascent transcript identification
- “minor changes in sample processing could lead to changes of up to >20% in the final

History of Development

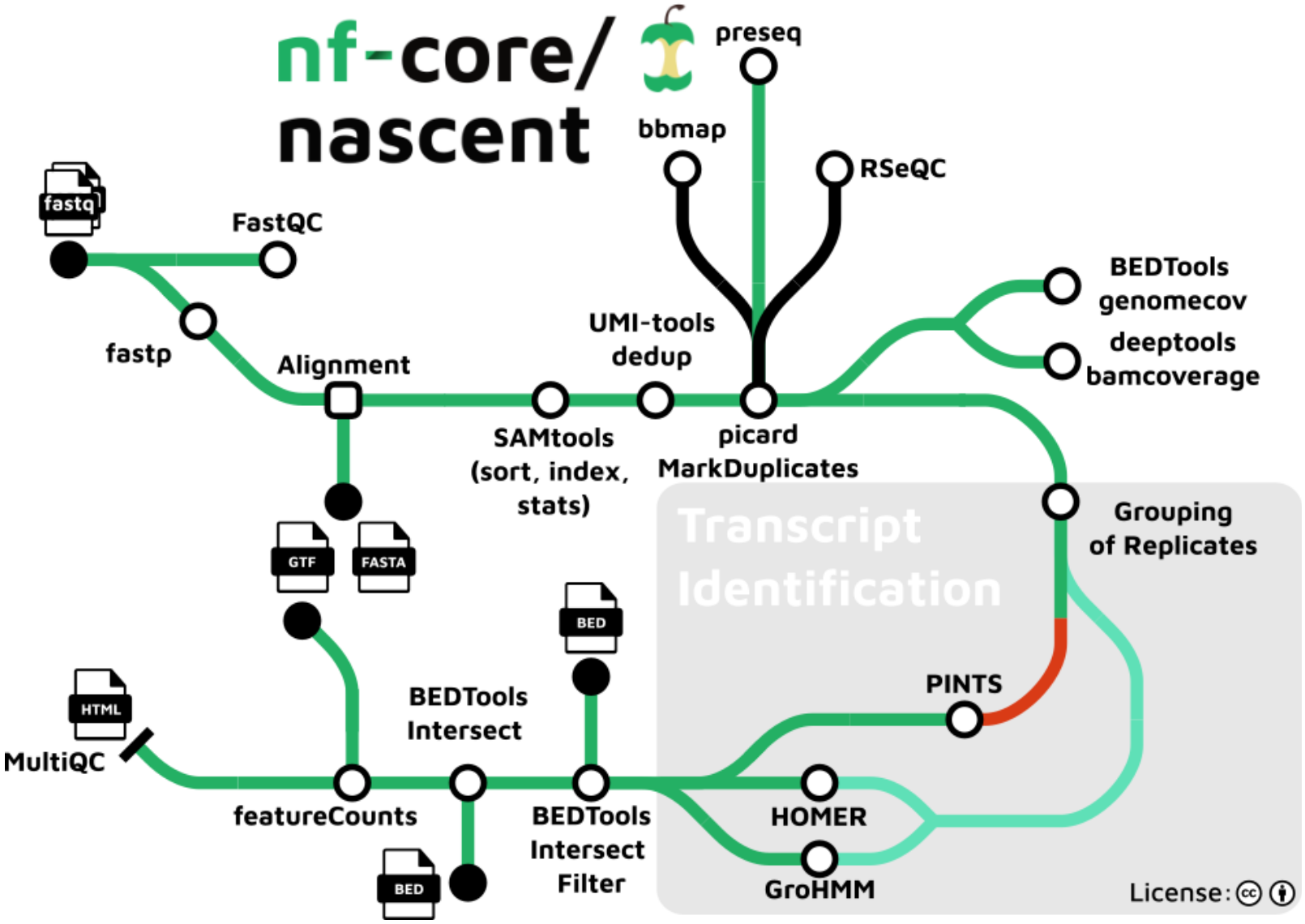
- Version 1.0 was developed by Ignacio Tripodi and Margaret Gruca and released Apr 16, 2019
- In 2017, THK lab started working to reproduce Kim et al. 2018 in a second dataset
- Struggled with building a reproducible pipeline in snakemake and creating CI/CD workflows and templates in January 2020
- Found nf-core in March 2020

How far we've come

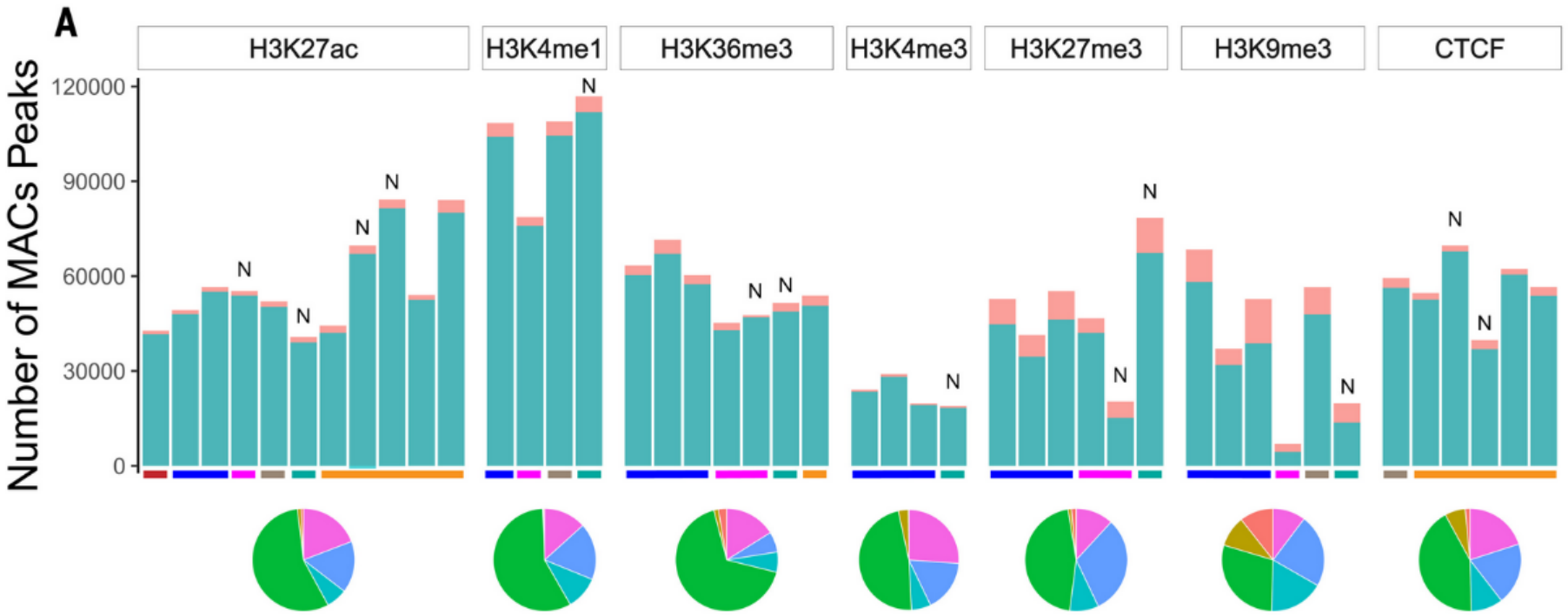


Original 2018 presentation

Obligatory Metro Map



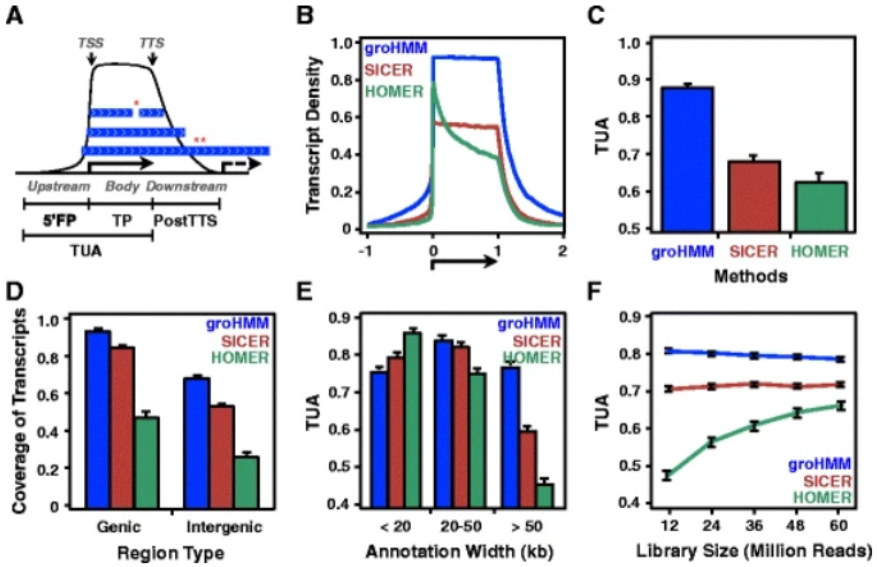
CHM13 Support



Transcript Identification

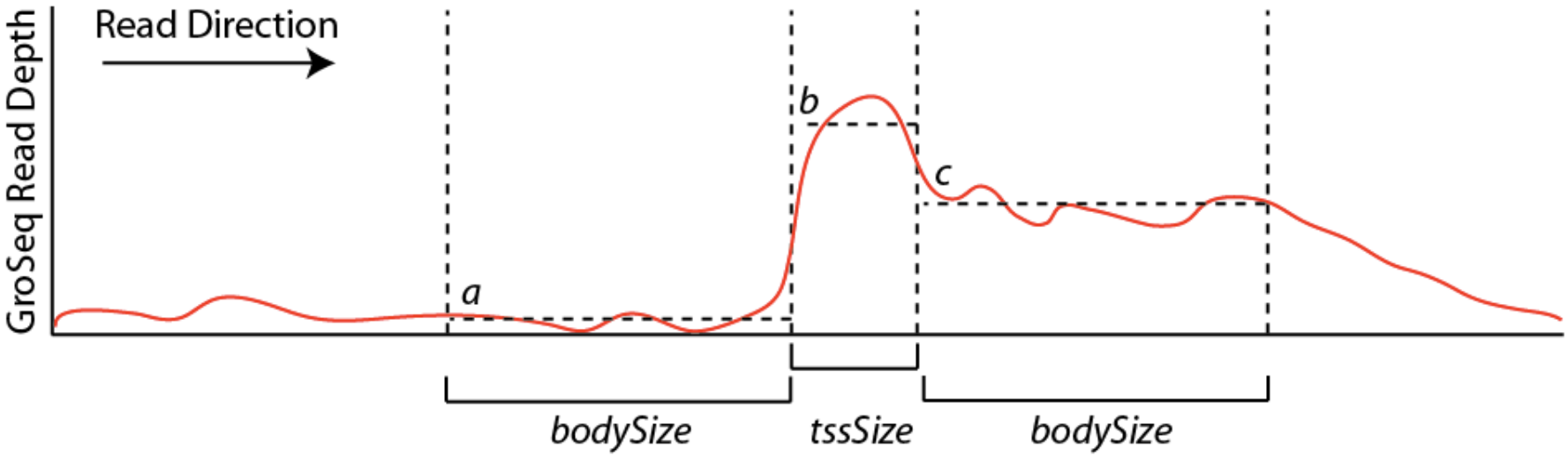
- GroHMM - Predicts transcripts from aligned GROSeq data in the form of bed files.
- HOMER - Transcript identification from GROSeq data
- PINTS - Identifies transcriptional regulatory elements (TREs) identified from nascent-transcript sequencing.

GroHMM



- R package released in 2015 by Minho Chae, Charles Danko, & Lee Kraus
- Draw back is it requires tuning and is quite memory hungry
- Recommended we use tunits instead

HOMER

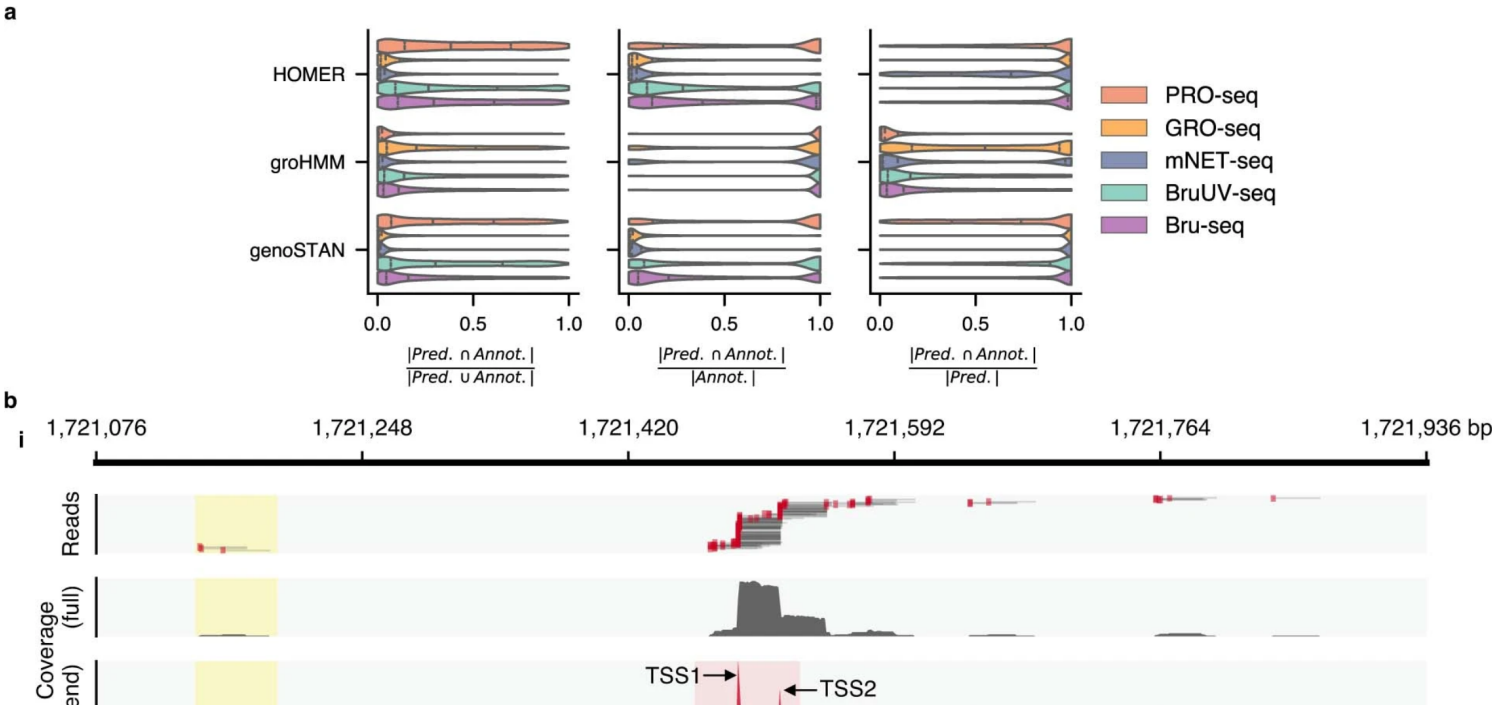


$$tssFold = \frac{b}{a} \quad bodyFold = \frac{c}{a}$$

- Released in 2010 from the Glass lab and continued by Chris Brenner
- Created in a land before docker

  `configureHomer.pl`

PINTS Identification



PINTS

- Released in 2022 by Yu Lab and Lis Lab
- Determines the TSS site(opposed to entire transcription units)
- PINTS achieves optimal balance among resolution, robustness, sensitivity, specificity and computational resources required.
- Supports 10 assays out of the box

Cunningham’s Law

The best way to get the right answer on the internet is not to ask a question; it’s to post the wrong answer.

- Please open a issue or drop into slack!

Need help?

Repository: `nf-core/nascent`

Tutorial: `https://nf-co.re/nascent/2.0.0/usage`

Chat: `https://nf-co.re/join` #nascent

Follow nf-core on

https://nf-co.re/

Chan Zuckerberg Initiative

Icons:
openmoji.org

References

Yao, L., Liang, J., Ozer, A. et al. A comparison of experimental assays and analytical methods for genome-wide identification of active enhancers.

Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. Nat Rev Genet. 2013 Apr;14(4):288-95. doi: 10.1038/nrg3458. PMID: 23503198; PMCID: PMC4445073.

Chae M, Danko CG, Kraus WL (2015). "groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data." BMC Bioinformatics, 16(222).

Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-589. doi:10.1016/j.molcel.2010.05.004