# Nascent Transcript Identification

Edmund Miller

2022-03-02 Wed

# Background

# Previous Work

## Global transcriptional activity dynamics reveal functional enhancer RNAs

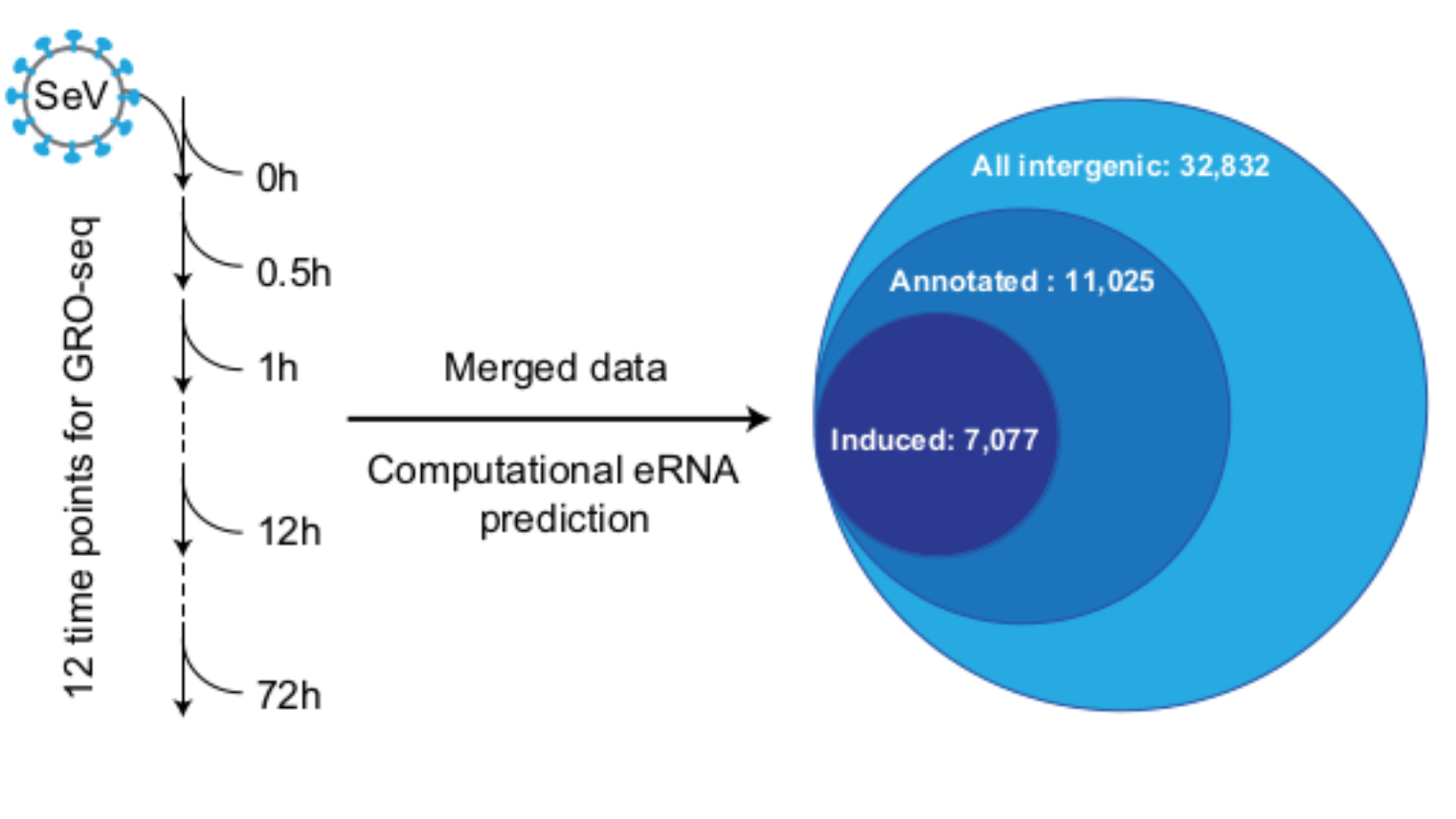
Yoon Jung Kim,[1,2] Peng Xie,[1,2] Lian Cao,[1] Michael Q. Zhang,[1] and Tae Hoon Kim[1]

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

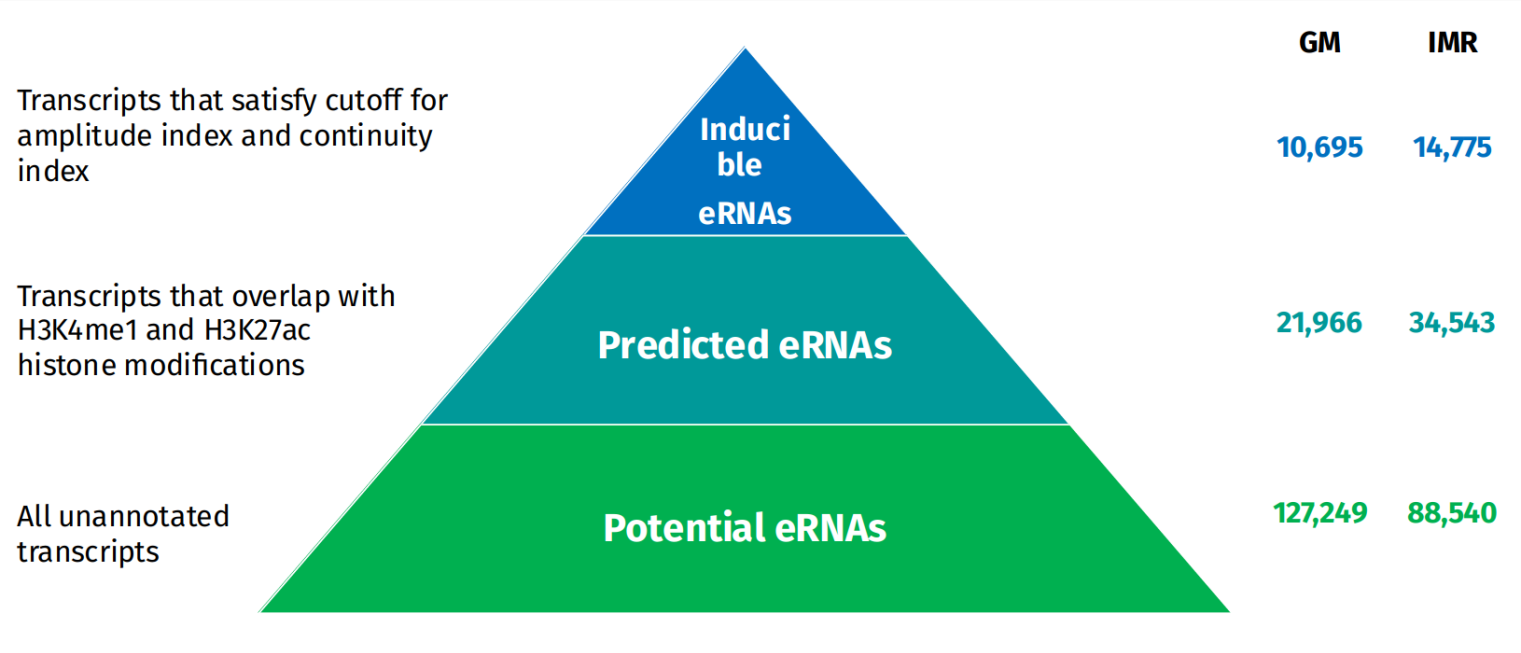
Active enhancers of the human genome generate long noncoding transcripts known as enhancer RNAs (eRNAs). How dynamic transcriptional changes of eRNAs are physically and functionally linked with target gene transcription remains unclear. To investigate the dynamic functional relationships among eRNAs 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.]

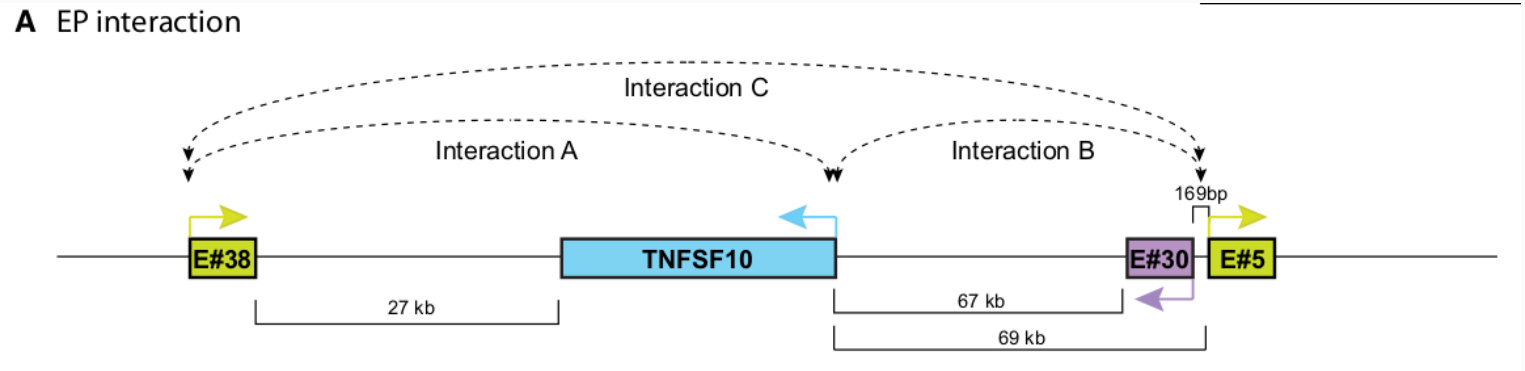
# Experimental Design



# eRNA Identification



# Viral Inducible eRNA-gene pairs




## Inducibility criteria

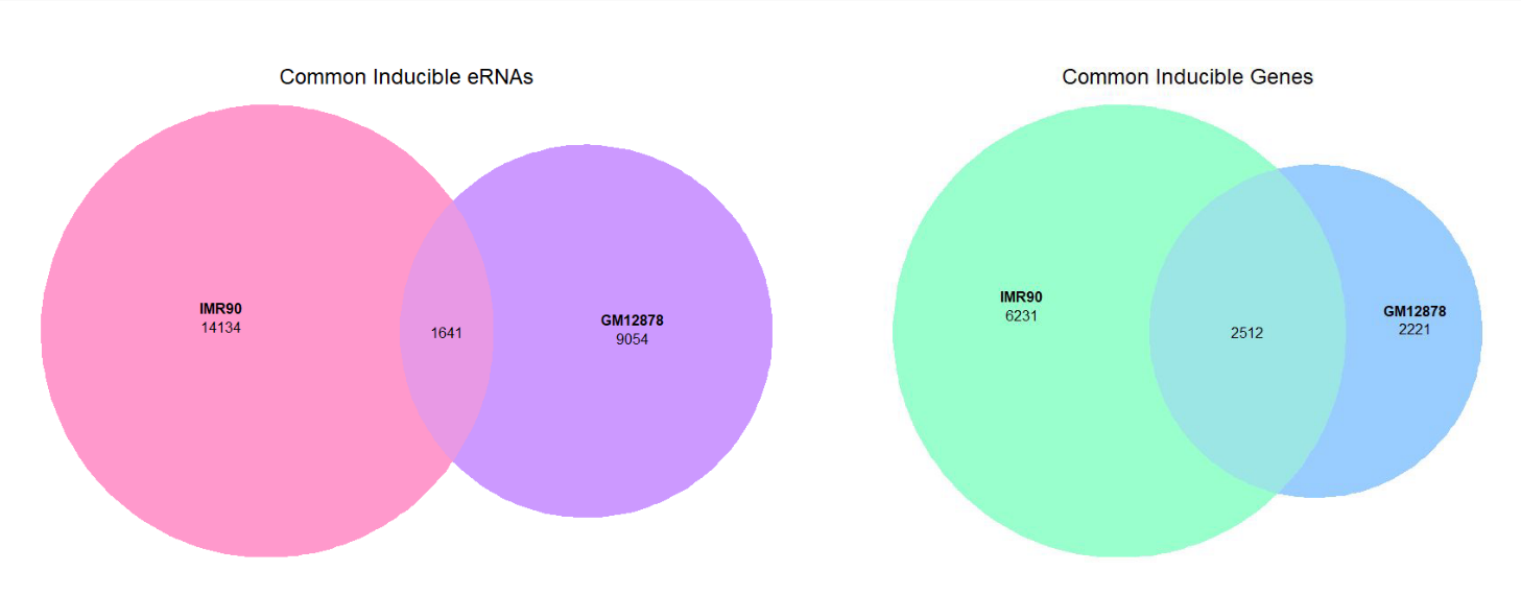
$$AI = \log_2 \left[ \frac{\max_{t \leq 24\,\mathrm{h}} e(t)}{e(0)} \right].$$

$$CI = \mathrm{correlation}\{[e(t_1), \ldots, e(t_{n-1})], [e(t_2), \ldots, e(t_n)]\}.$$

# Reproduction with IMR Dataset

- Wrote workflow using snakemake
- Goal was to reproduce GM results
- Achieved 80% of predicted eRNAs

# GM12878/IMR90 Comparison



# nf-core/nascent

# 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

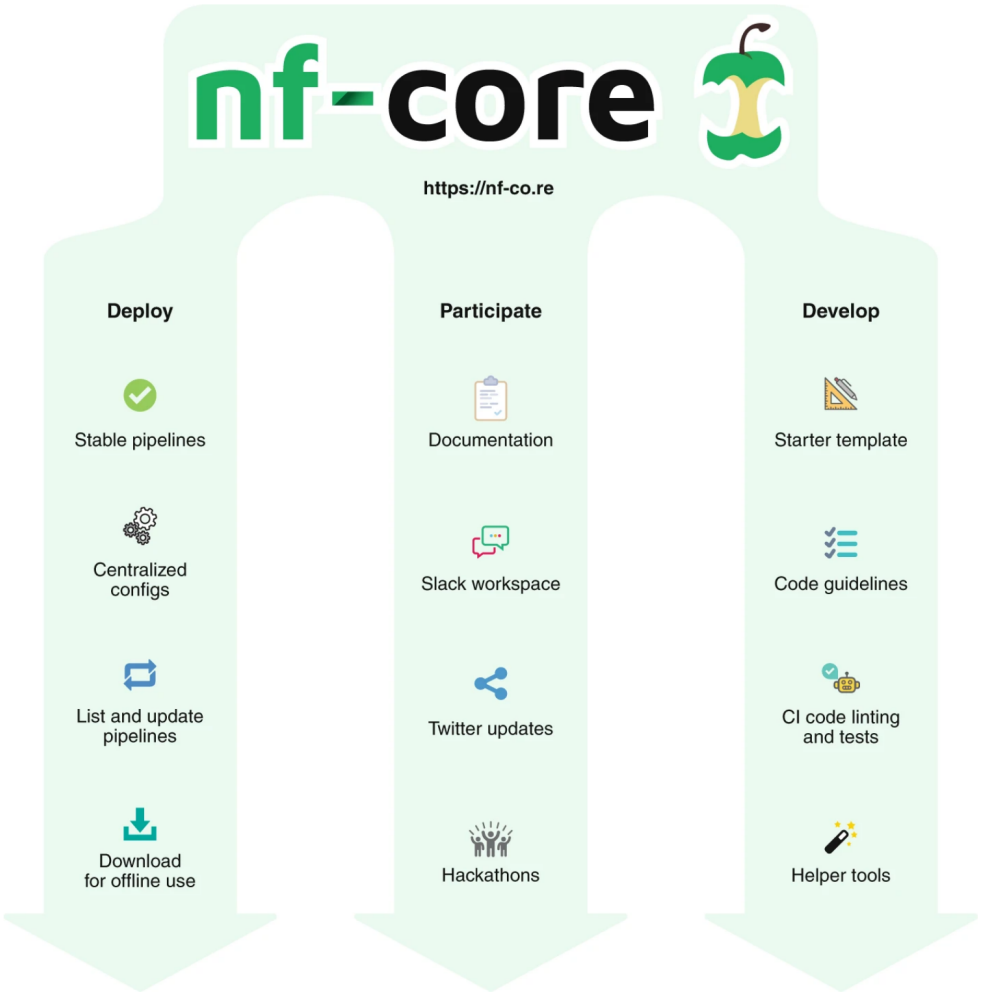
Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso & Sven Nahnsen

*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
```

# Inheriting nf-core Nascent

- Breaking our analysis up into smaller pieces
- nf-core portion includes Quality Checks, alignment, graph generation, transcript identification, and transcript quantification
- Downstream analysis is then a seperate nextflow workflow
- Data engineering/Data Science split

# Future

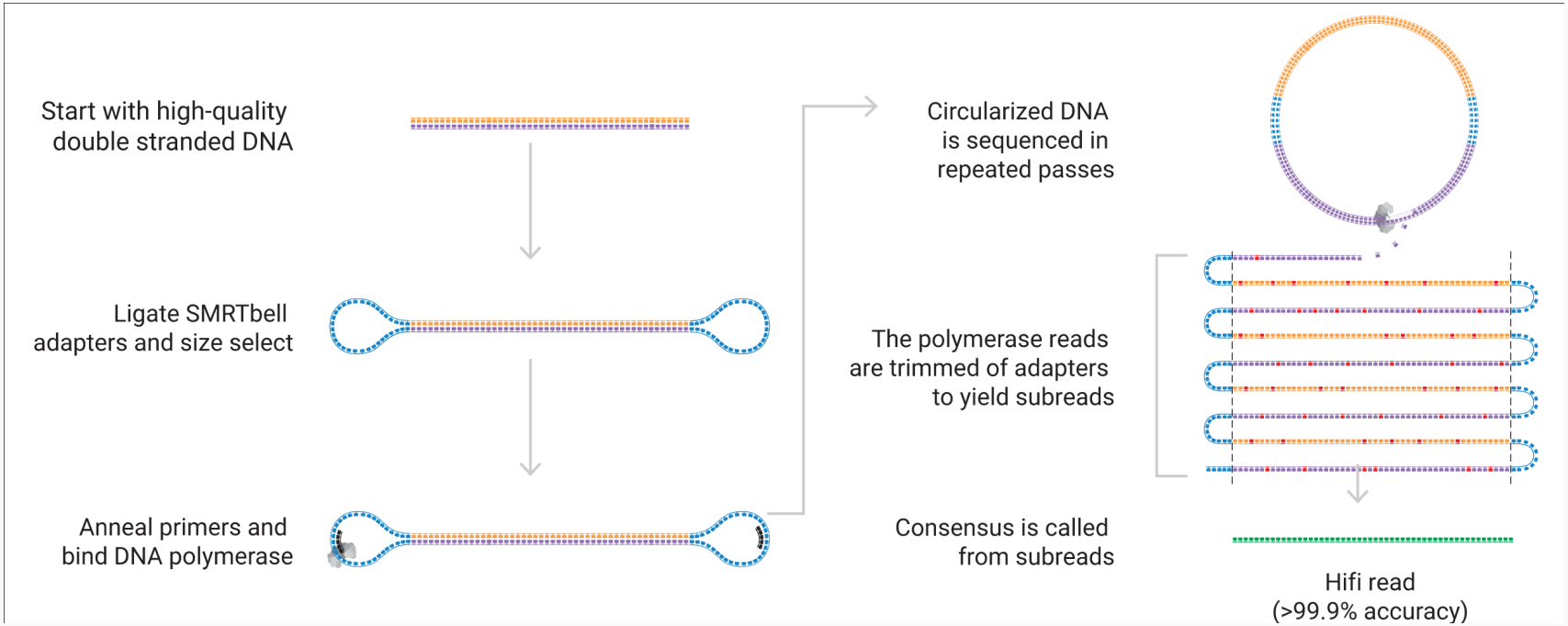
# Future

- CHM13 centromeres
- Utilizing PINTs transcriptional regulatory elements (TREs)
- WGS viral sequences

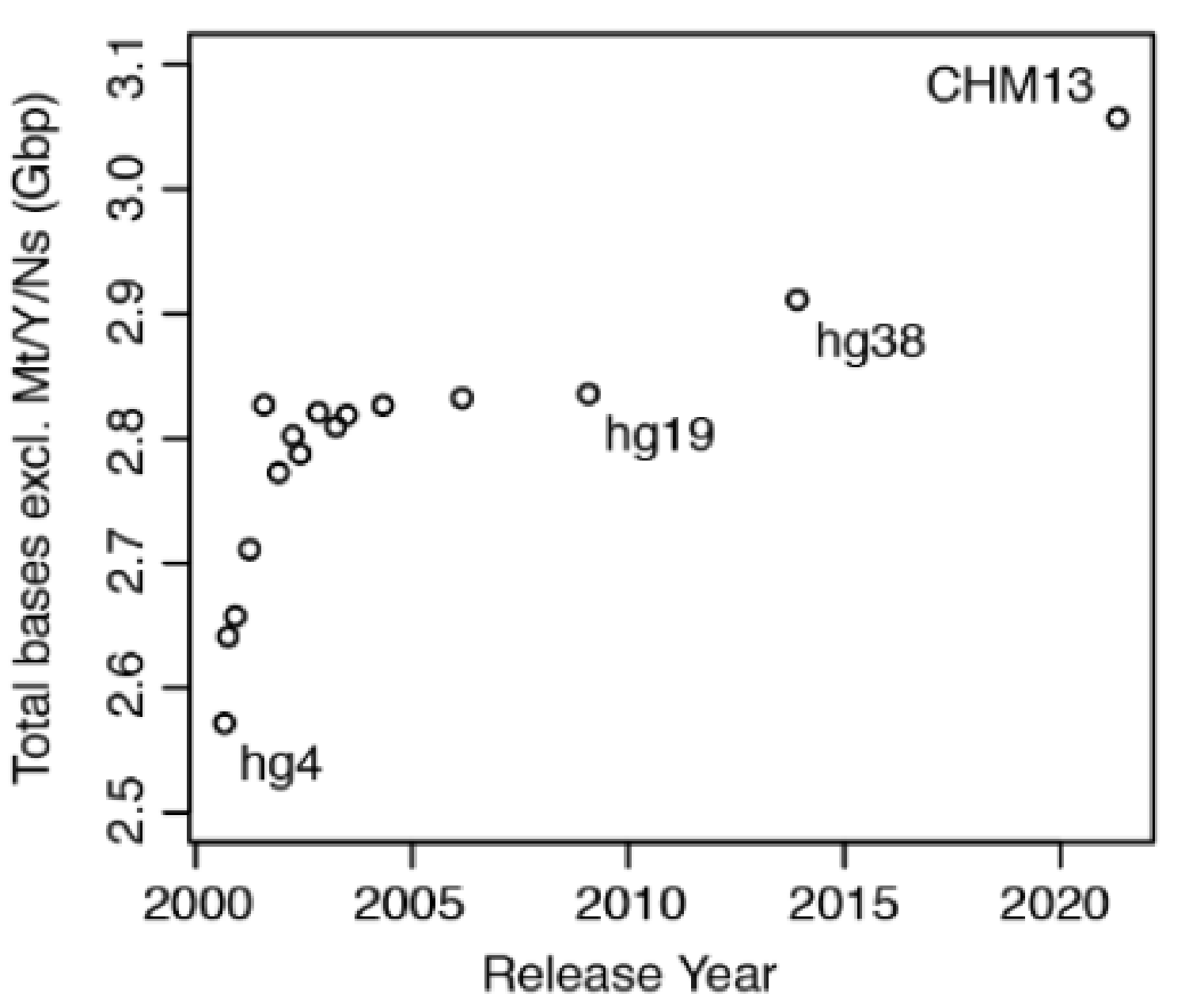
# CHM13

- Used PacBio's HiFi, and Nanopore's "ultra-long" reads to resolve complex forms of structural variation and gaps in GRCh38
- CHM13 Cell line a complete hydatidiform mole (CHM) cell line, "essentially haploid nature"

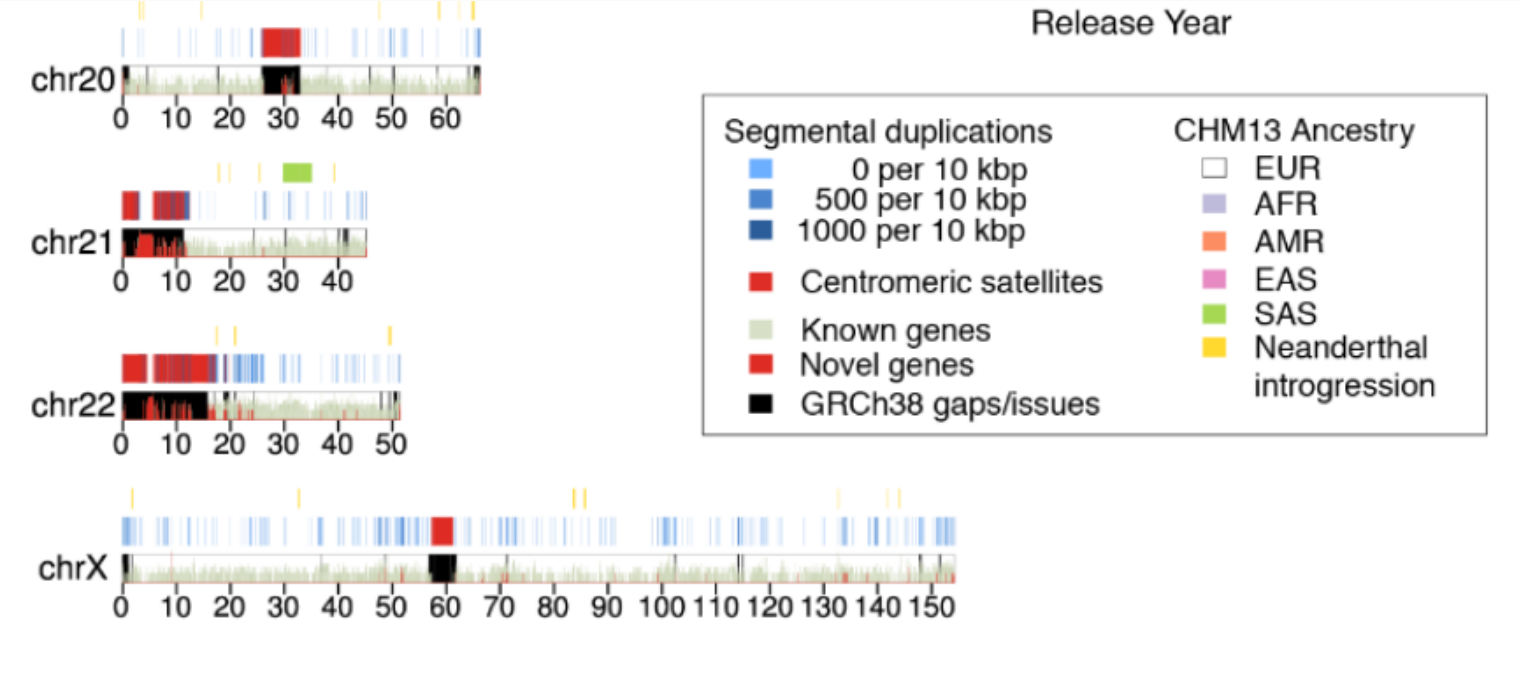
# Pacbio HiFi Reads



# CHM13


CHM13
hg38
hg19
hg4
Total bases excl. Mt/Y/Ns (Gbp)
2.5
2.6
2.7
2.8
2.9
3.0
3.1
2000
2005
2010
2015
2020
Release Year

# CHM13


Release Year
chr20
0 10 20 30 40 50 60
chr21
0 10 20 30 40
chr22
0 10 20 30 40 50
chrX
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
Segmental duplications
0 per 10 kbp
500 per 10 kbp
1000 per 10 kbp
Centromeric satellites
Known genes
Novel genes
GRCh38 gaps/issues
CHM13 Ancestry
EUR
AFR
AMR
EAS
SAS
Neanderthal introgression

# The intersection of Homer identified peaks and centromeres

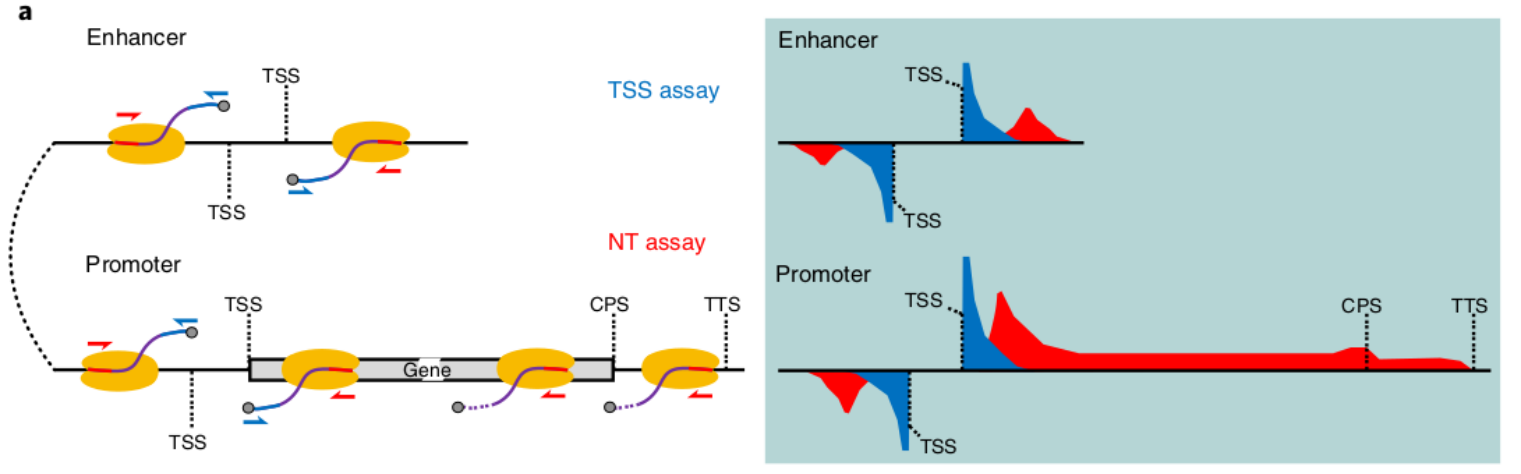
- Near identical number of reads mapped(There's only 5% more to align to)
- Had to use hg19 centromeres for the intersection(Found a T2T liftOver chainfile)
- There were no hits
- Confirmed with call to aligned reads for centromere regions

# The intersection of Homer identified peaks and centromeres

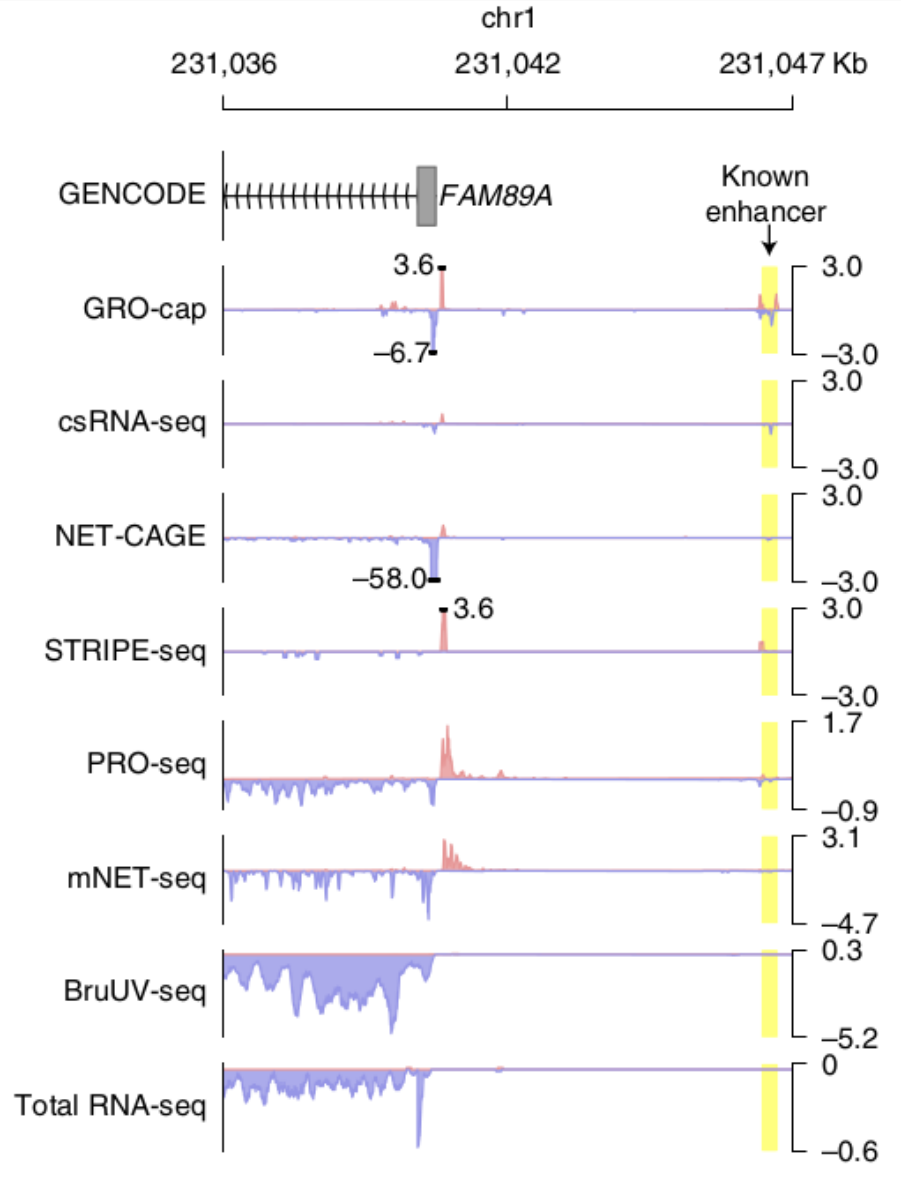
- Reads are probably being thrown out for two reasons
  - Centromeres are high in repeats so tossed
  - They then are high in repeats so the probably have multiple alignments and get tossed

# PINTS - transcriptional regulatory elements (TREs)

- PINTS is modeled after MACS2
- Utilized TSS assays (GRO/PRO-cap, CoPRO, Start-seq, CAGE)
- GRO/PRO-seq are NT assays



# Picking up Nascent Reads



# PINTS Highlights

- Reasonable computational requirements (Feasible to run on a typical personal computer)
- Noticed minor changes in sample processing could lead to changes of up to >20% in the final results
- Well packaged and usable by others

# WGS viral sequences

- Take 1000 genomes data from Illumina
- Take unaligned reads from bams and convert back to raw reads
- Realign to viral genomes and quantify through nf-core/sarek WGS pipline
- Looking for possible correlation between viral infection and phenotypes

# Call to Actions

- Take Applied Genomics this summer!
- Apply for the nf-core mentorship program!
- Element Biosciences launch event on monday!