

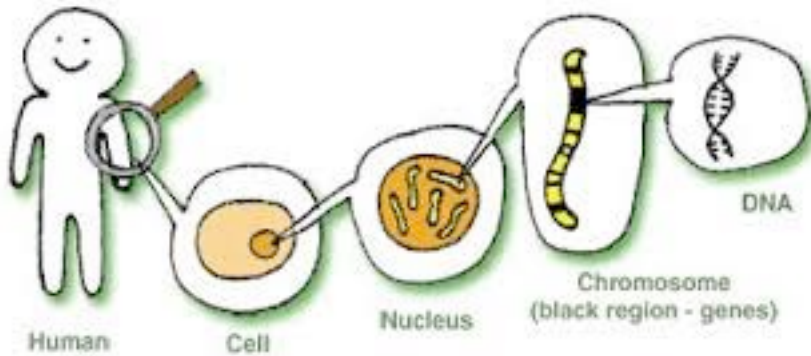
A key to unlocking the function of non-coding variations in disease: data sharing nascent transcription reanalysis

What makes one person
different from another?
Can the cell's read-out (RNA) tell us?

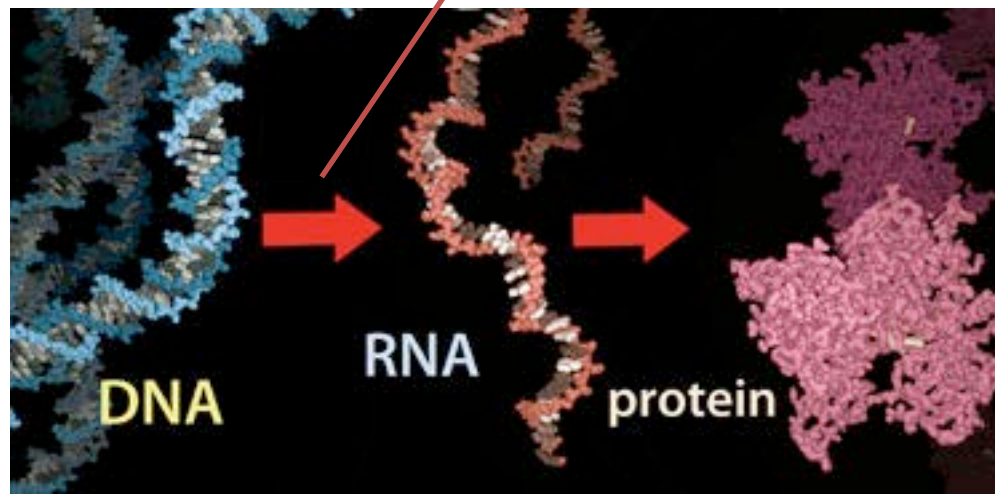
Mary Ann Allen
The Dowell Lab



The stuff of life

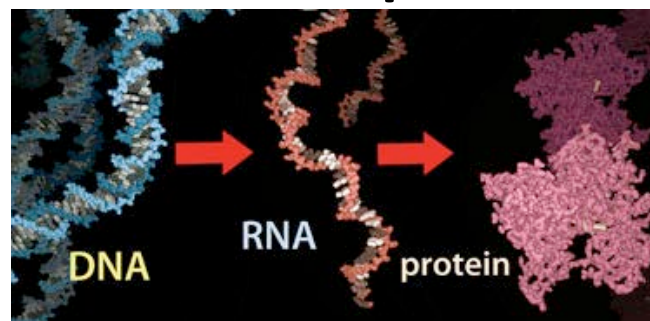


The Dowell Lab:
How genetic variation (differences in
DNA) affect transcription (DNA→RNA)





Studying the planet: Where are the important places?



DNA



Annotation =
labels on a map

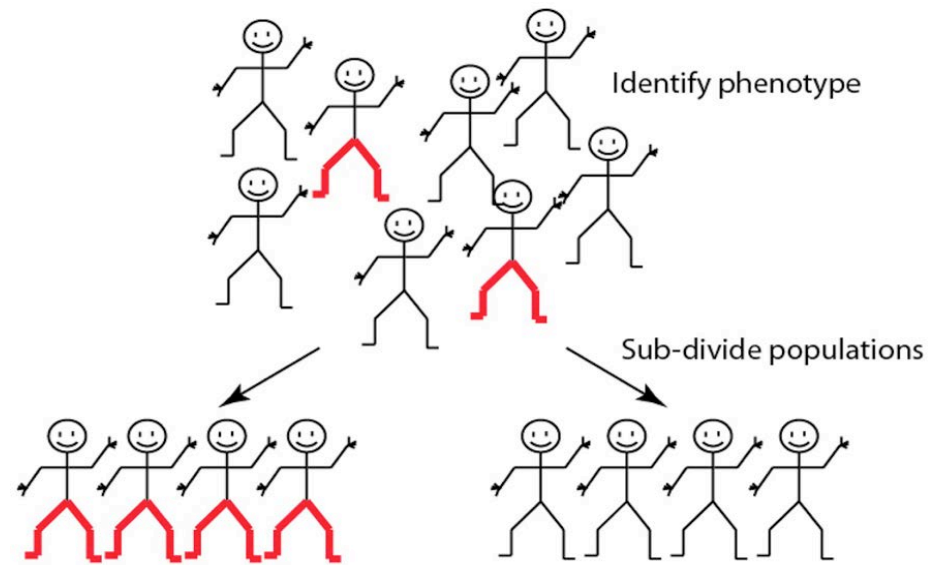
Protein coding gene



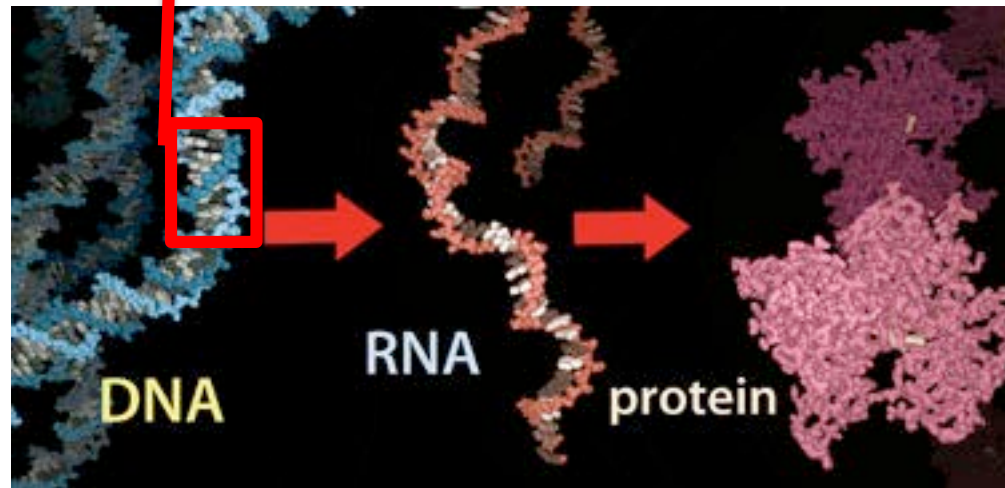
2%



Discovering the cause of a disease: genome-wide association studies



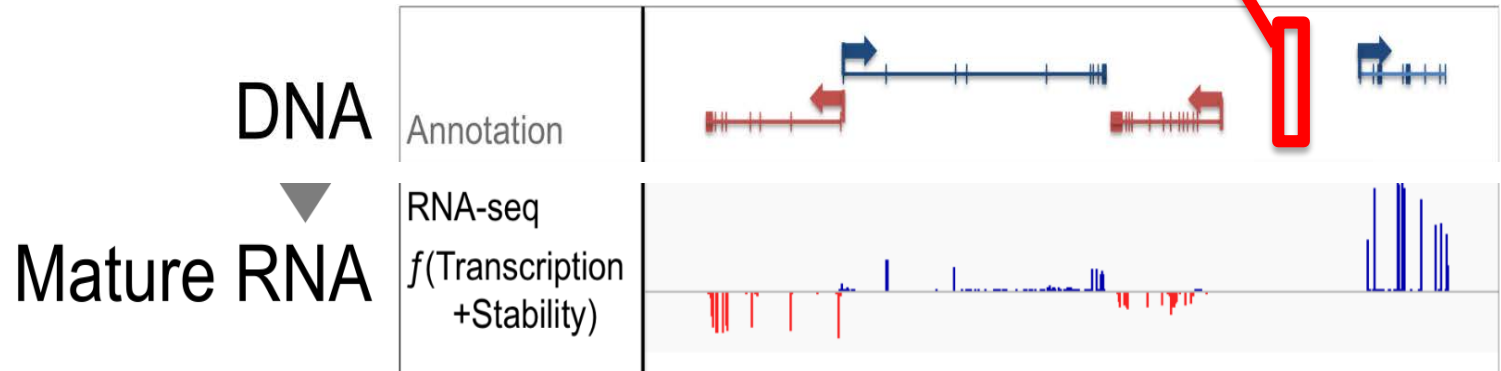
SNP



[Schierding](#) W. et al. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. Front. Genet., 18 February 2014

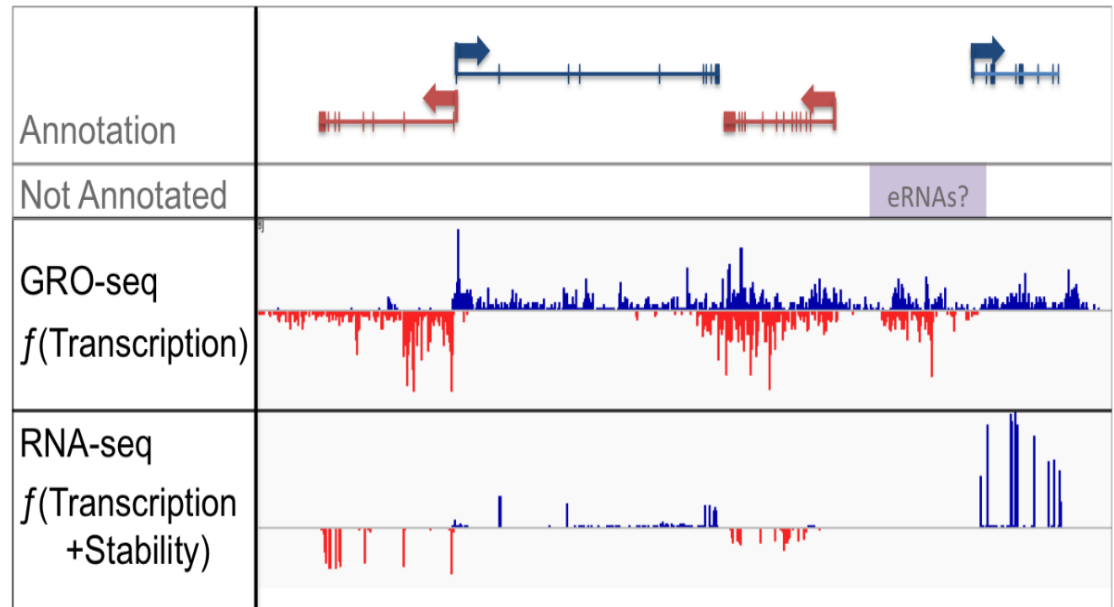
Know what you are measuring: “Transcriptome”

95% of disease causing SNPS are not in protein coding gene



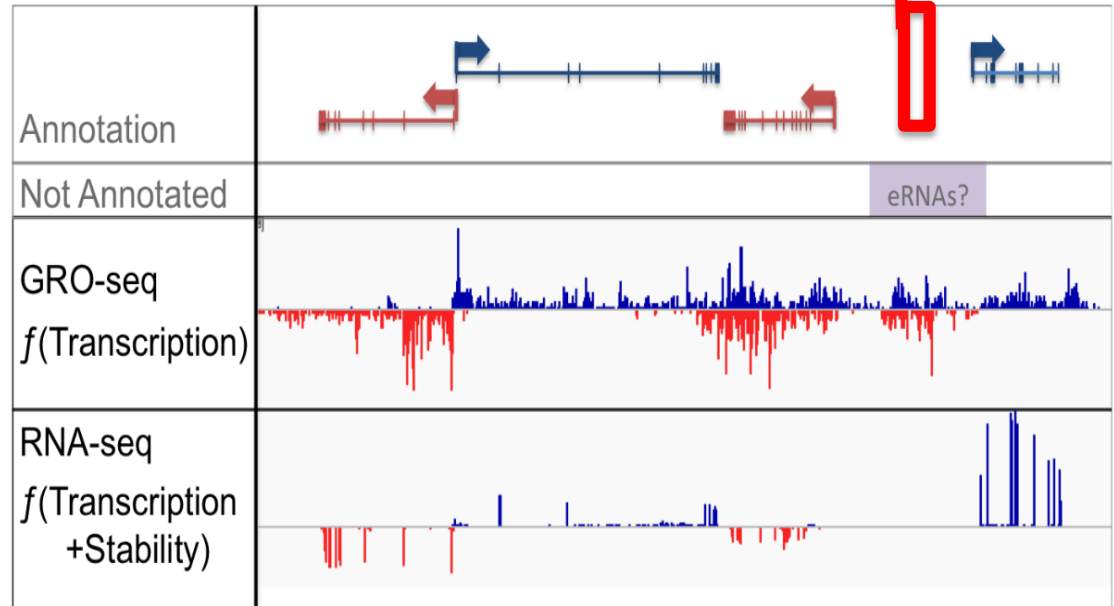
70% of the genome is transcribed,
10% of the genome encodes stable RNA

70% DNA
↓
Nascent RNA
↓
10% Mature RNA



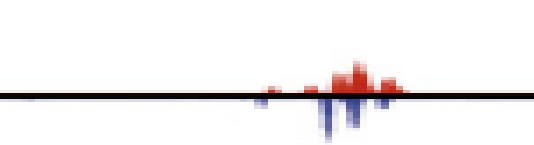
Know what you are measuring

70% DNA
↓
Nascent RNA
↓
10% Mature RNA

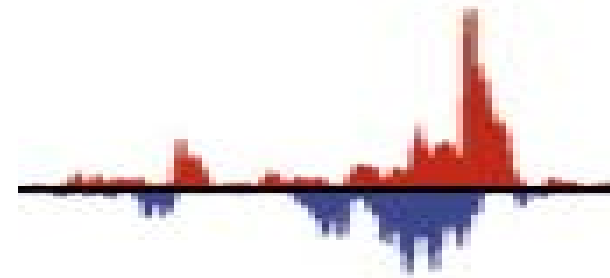
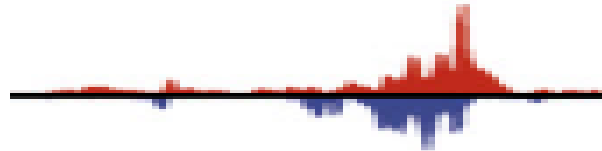


The transcription level of the RNA approximates the function level

Hardly working



Working hard

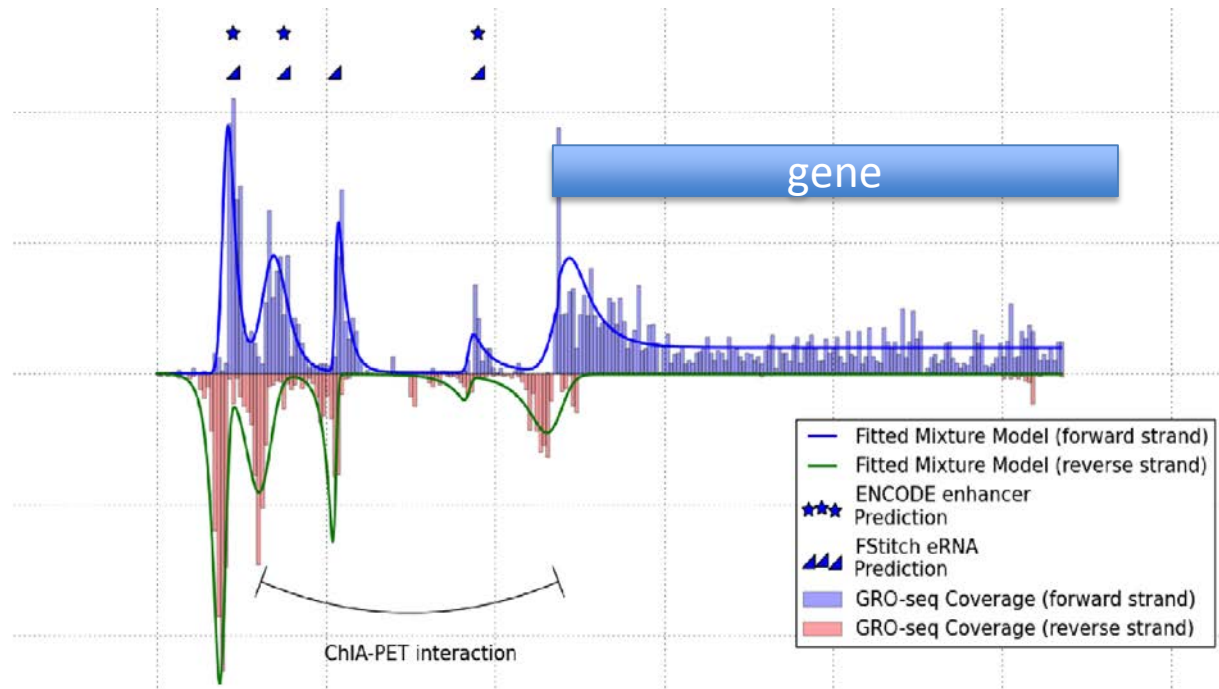


- X Different cell types
- X Different conditions
- ? Different people

We are building new algorithms for nascent data and are analyzing public data



Joey Azofeifa



Making the public data useful: a heroic effort from a talented undergrad



Josephina Hendrix

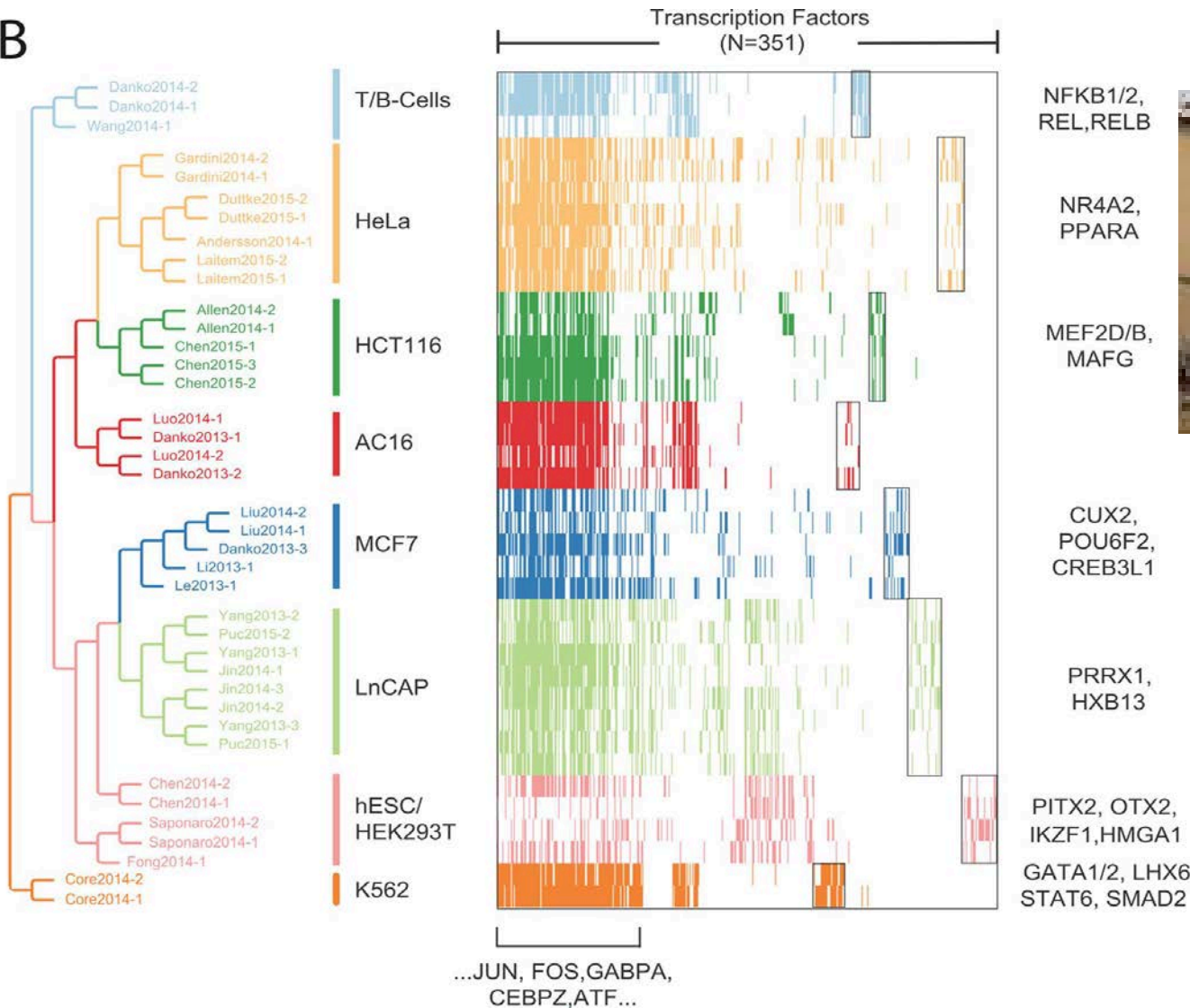
Data Reanalysis of all published GRO-seq data (summer 2015)

- Over 60 papers
 - 50 cell types
 - 20 Terabytes of data
 - 60,000 CPU hours on pando
 - A computing cluster with 2000 CPU cores

**SRA -> FASTQ -> SAM FILES (BOWTIE2) ->
BAM FILES (SAMTOOLS) -> BEDGRAPH (BEDTOOLS)
-> BIGWIG (BEDGRAPH2BIGWIG)**

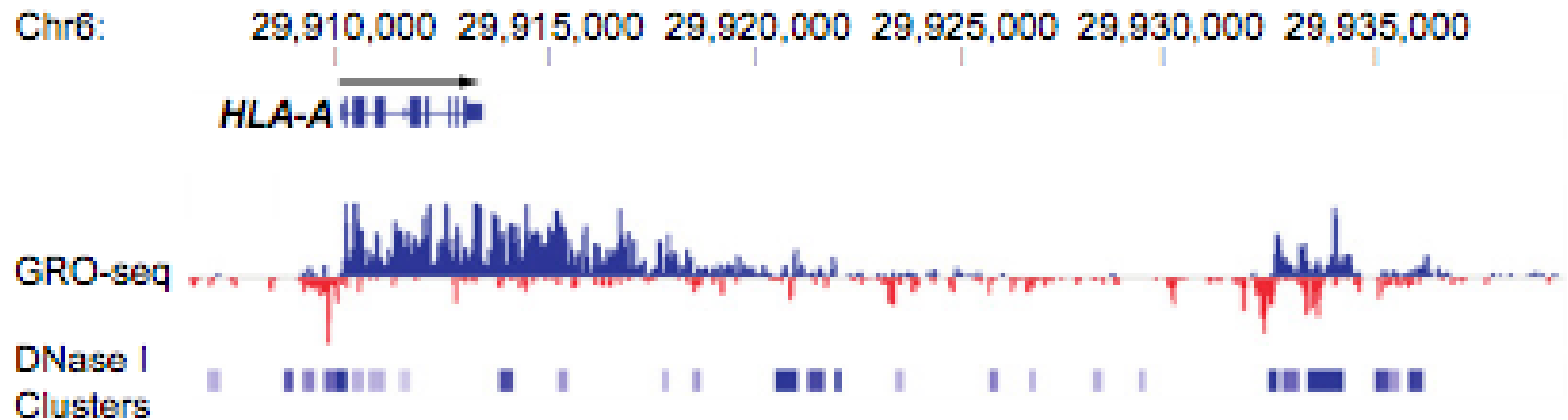
But these data took a ton of time and computing to make them useful...

B



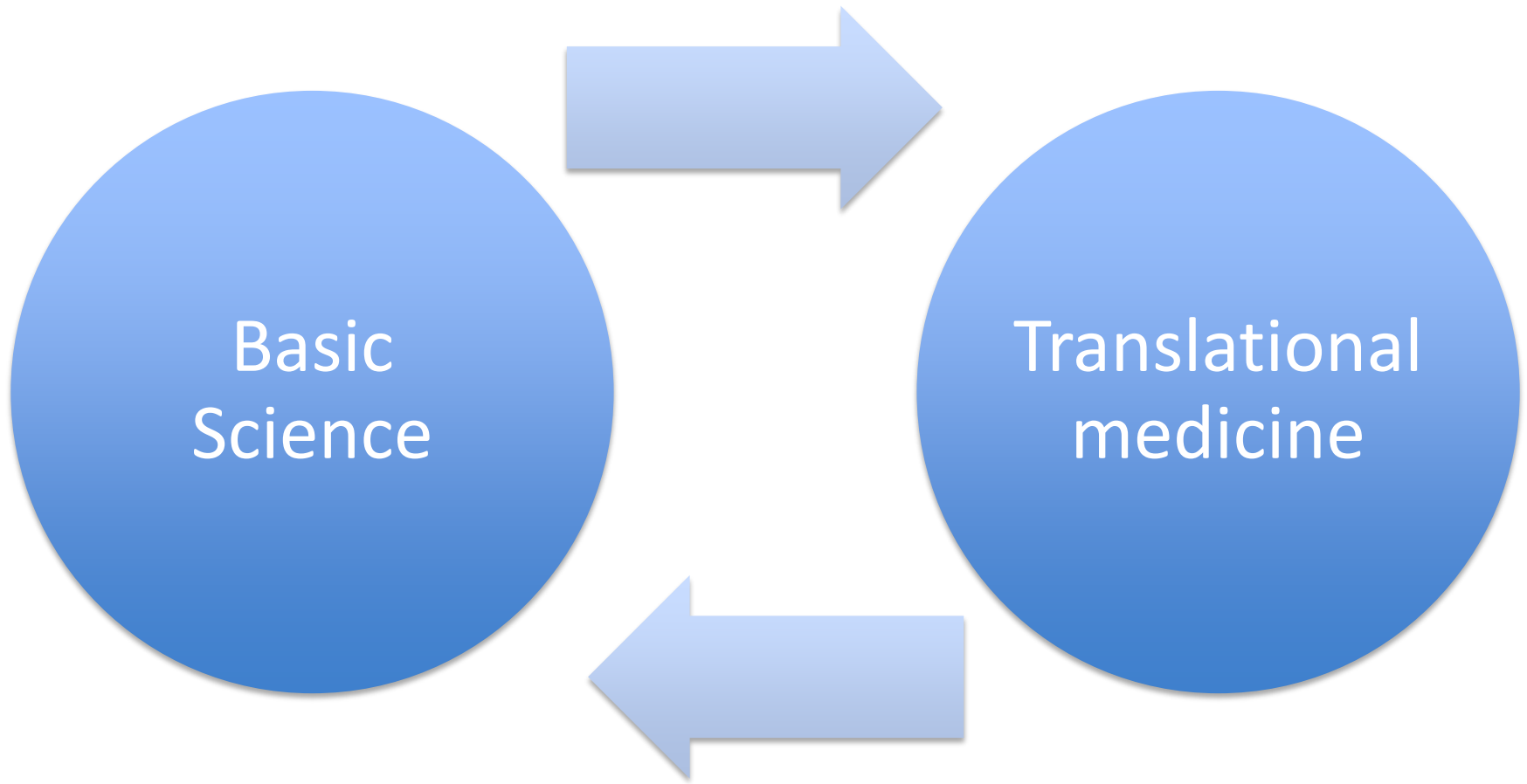
Why should
someone
repeat that?

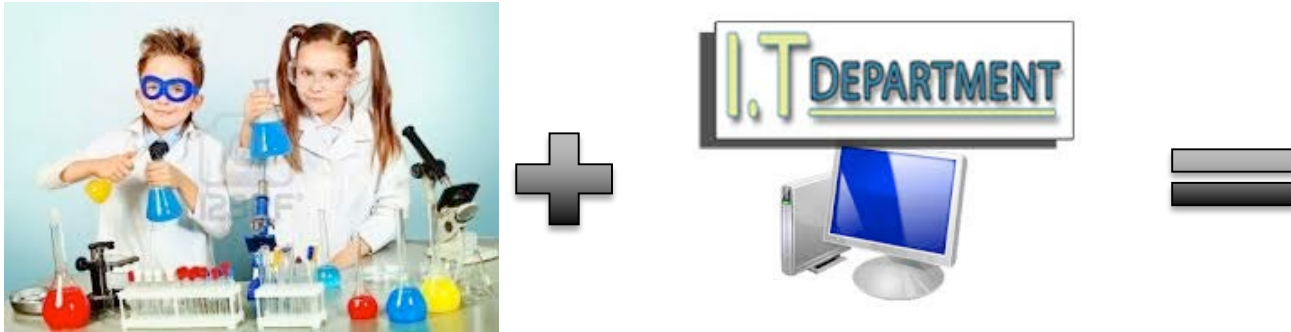
Sharing eRNA information one investigator at a time is productive but slow



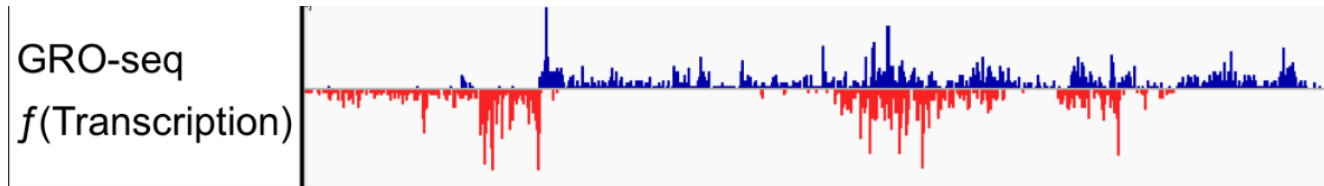
Hayashi et. Al. Autoimmune vitiligo is associated with gain-of-function by a transcriptional regulator that elevates expression of HLA-A*02:01 in vivo

We want to share information about nascent RNA





A nascent transcription repository



Building a GRO-seq genome browser: Another talented undergrad



Jonathan DeMasi
(starting a job at Apple)

Independent study:

GRO-seq Visualization Exploration
Project

- Tried several difference interfaces
- Settled on biodalliance

Then we needed a website:

[Cassidy Thompson](#)

[Mary Allen](#)

[Josephina Hendrix](#)

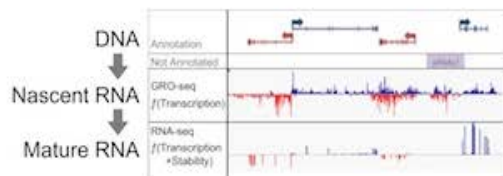


Have a favorite non-coding SNP?

<http://gro-seq.colorado.edu/>



Right now the repository is just human (we need to add more data)
Planning on adding our annotations from our novel algorithms
Wish to add better querying



[Log in](#)

Search Site

Search

☐ only in current section

[Home](#)

[Gro-Seq Genome Browser](#)

[Dowell Lab](#)

[Data Reanalysis](#)

[About Us](#)

[Home](#)

Nascent Transcription Repository

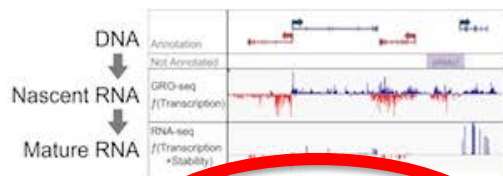
Here we present Nascent Transcription data (GRO-seq) from many publications. This data can be used to find RNAs in non-coding regions which are not annotated.

The misnomer "transcriptome"

For years people have performed RNA-seq and RNA microarray experiments calling them the transcriptome. However, these assays actually measure steady state RNA.

Imagine I told you to measure how much coffee was made at your work in one day. Would you do that by asking how much coffee is in the coffee pot?





Log in

Search Site

Search

☐ only in current section

Home

Gro-Seq Genome Browser

Dowell Lab

Data Reanalysis

About Us

Home

Nascent Transcription Repository

Here we present Nascent Transcription data (GRO-seq) from many publications. This data can be used to find RNAs in non-coding regions which are not annotated.

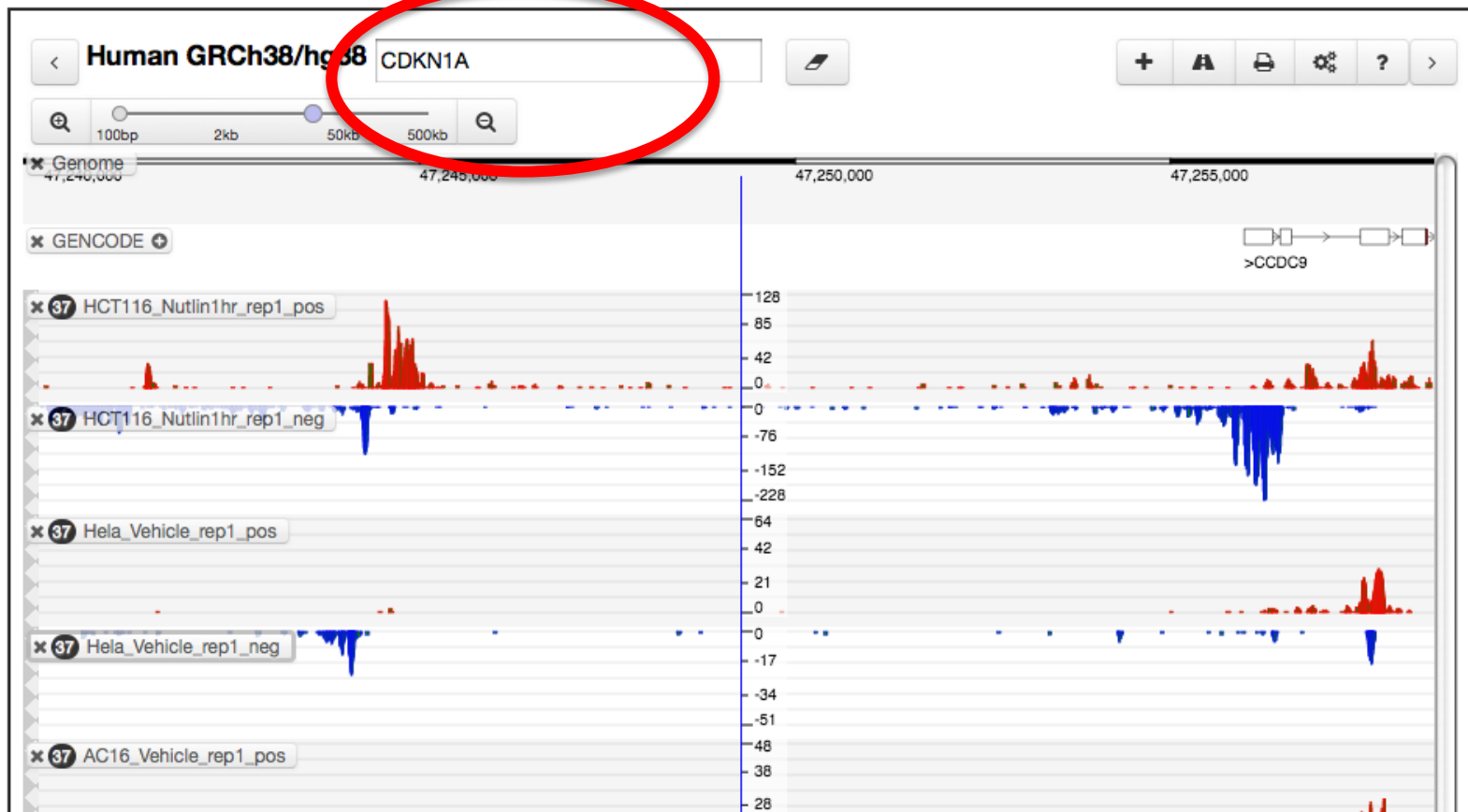
The misnomer "transcriptome"

For years people have performed RNA-seq and RNA microarray experiments calling them the transcriptome. However, these assays actually measure steady state RNA.

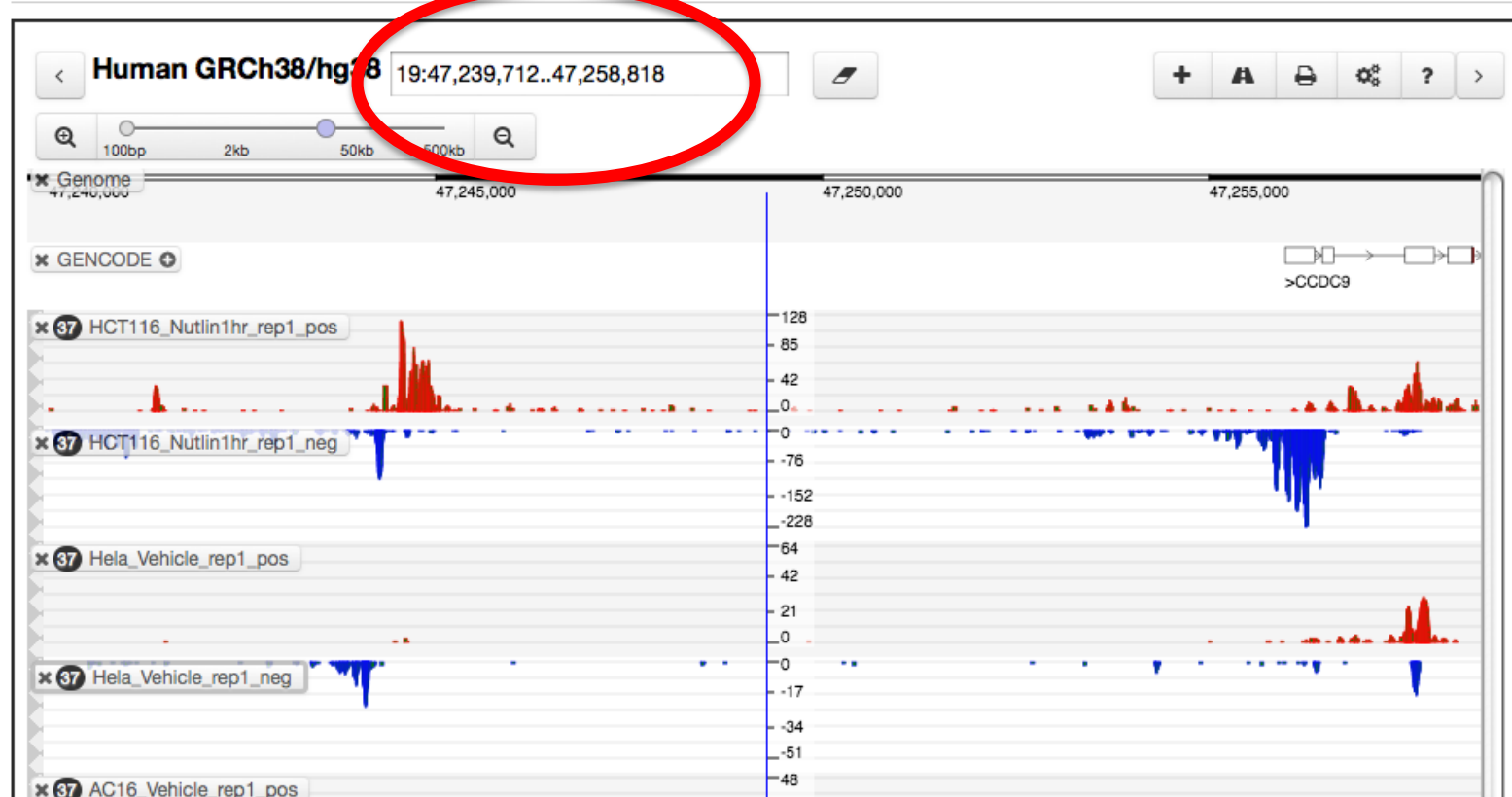
Imagine I told you to measure how much coffee was made at your work in one day. Would you do that by asking how much coffee is in the coffee pot?



Gro-Seq Genome Browser



Gro-Seq Genome Browser



Right now the repository is just human (we need to add more data)
Planning on adding our annotations from our novel algorithms
Wish to add better querying

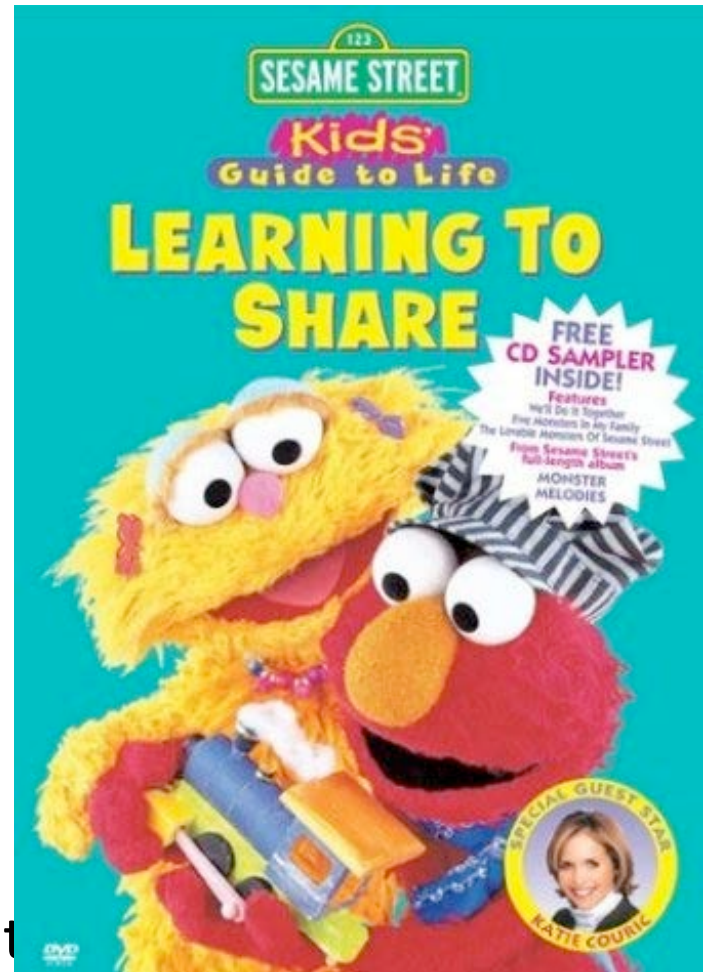
Learning to Share Data

What we needed

- Data (reanalysis)
- Organization
- Big computer(s)
- Awesome team
 - Scientists and IT cooperated
 - UNDERGRADS ARE AWESOME!

What still need

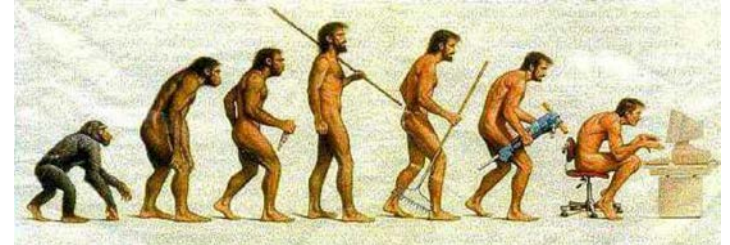
- Funding for sharing projects?
 - A maintenance plan
- Growing compute resources
- Informing (GWAS) investigators about this resource



Acknowledgements

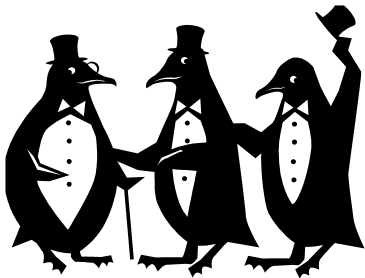
Dowell Laboratory,
University of Colorado Boulder

Robin Dowell, Mary Allen, Tim Read, Amber
Sorenson, [Joey Azofeifa](#), [Josephina
Hendrix](#), Jonathon Rubin, Andrew
Blumenfeld



BioFrontiers IT (BIT),
University of Colorado Boulder

Dan Timmons (BioFrontiers IT Director), Scott
Davis, Matt Hynes-Grace, [Cassidy
Thompson](#), [Jonathan DeMasi](#)



Funding Sources:

- Butcher Seed Grant
- National Science Foundation
- Boettcher Foundation
- Linda Crnic Institute for Down syndrome
- Sloan Foundation Fellowship



BOETTCHER
FOUNDATION

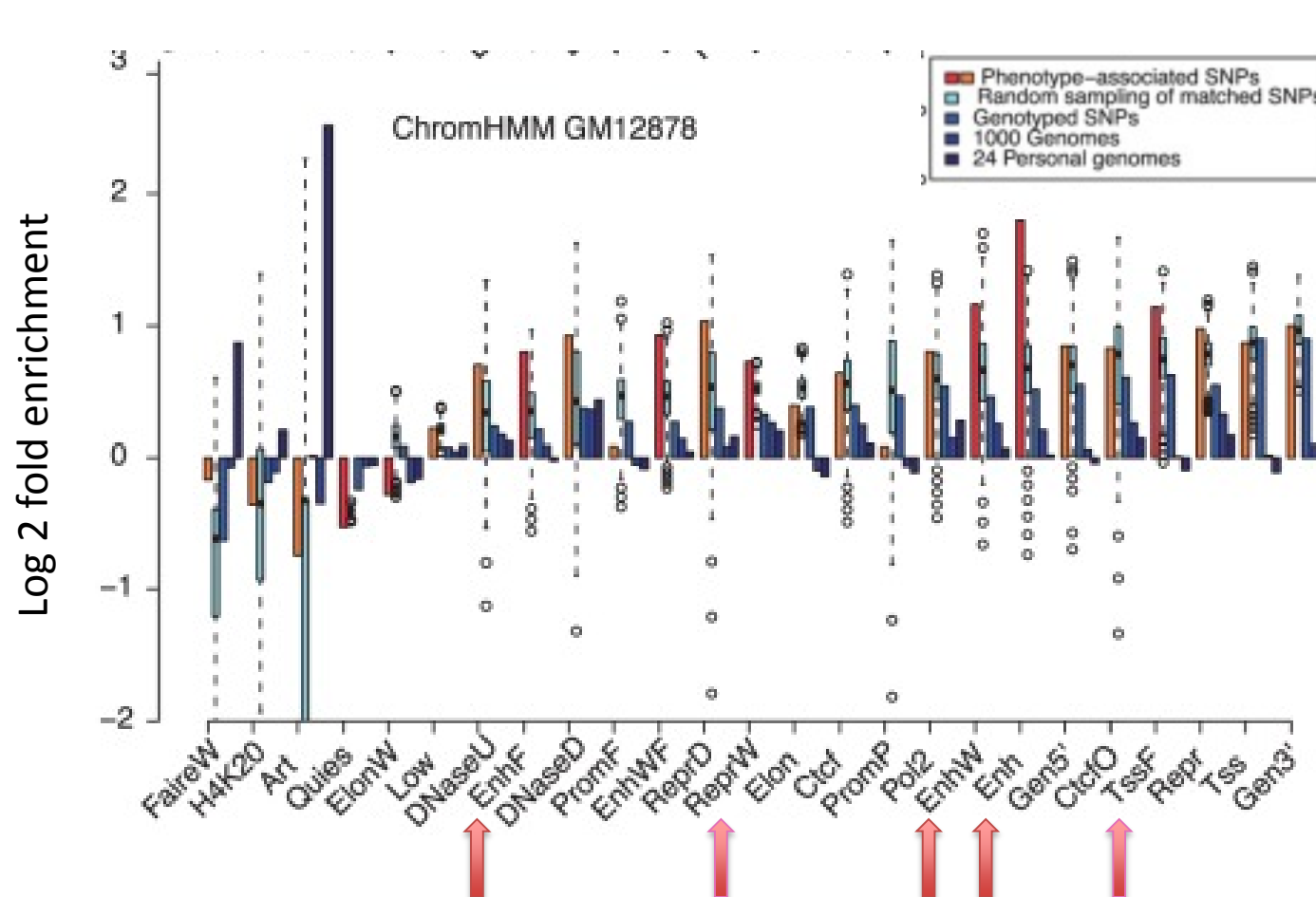


Sequencing Analysis: Just do it

- June 6th-17th
- tinyurl.com/j3zqobu

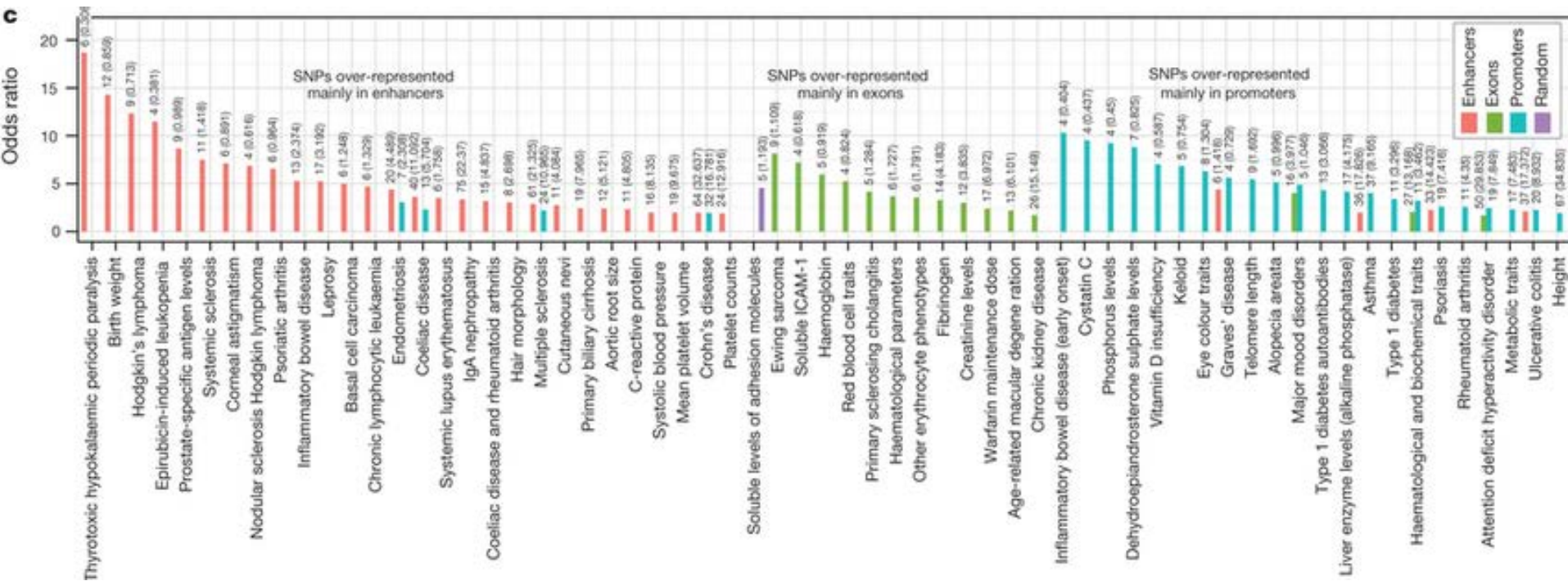


Many of the disease-associated SNPs are in enhancers



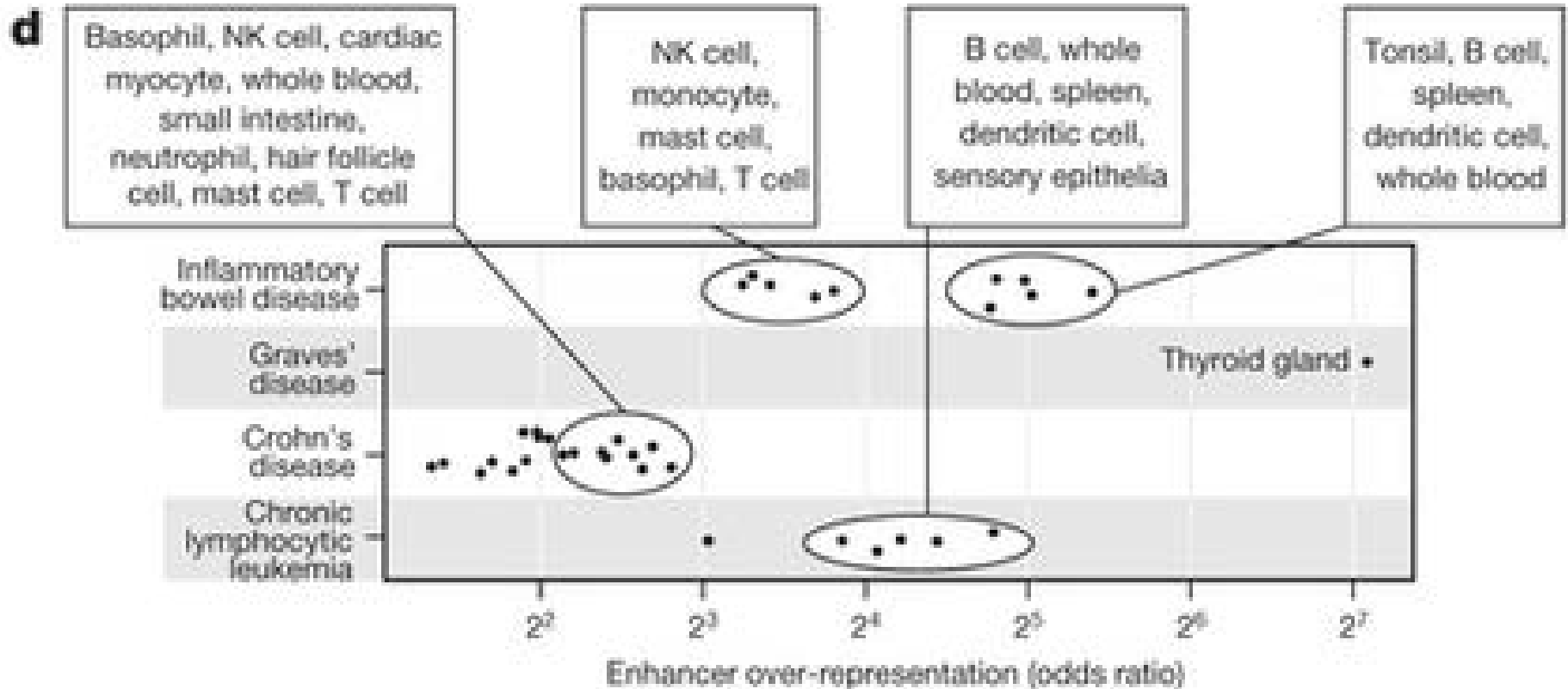
[Hoffman](#) M et al. Integrative annotation of chromatin elements from ENCODE data.
Nucleic Acids Res. 2013

Many of the disease-associated SNPs are in enhancers



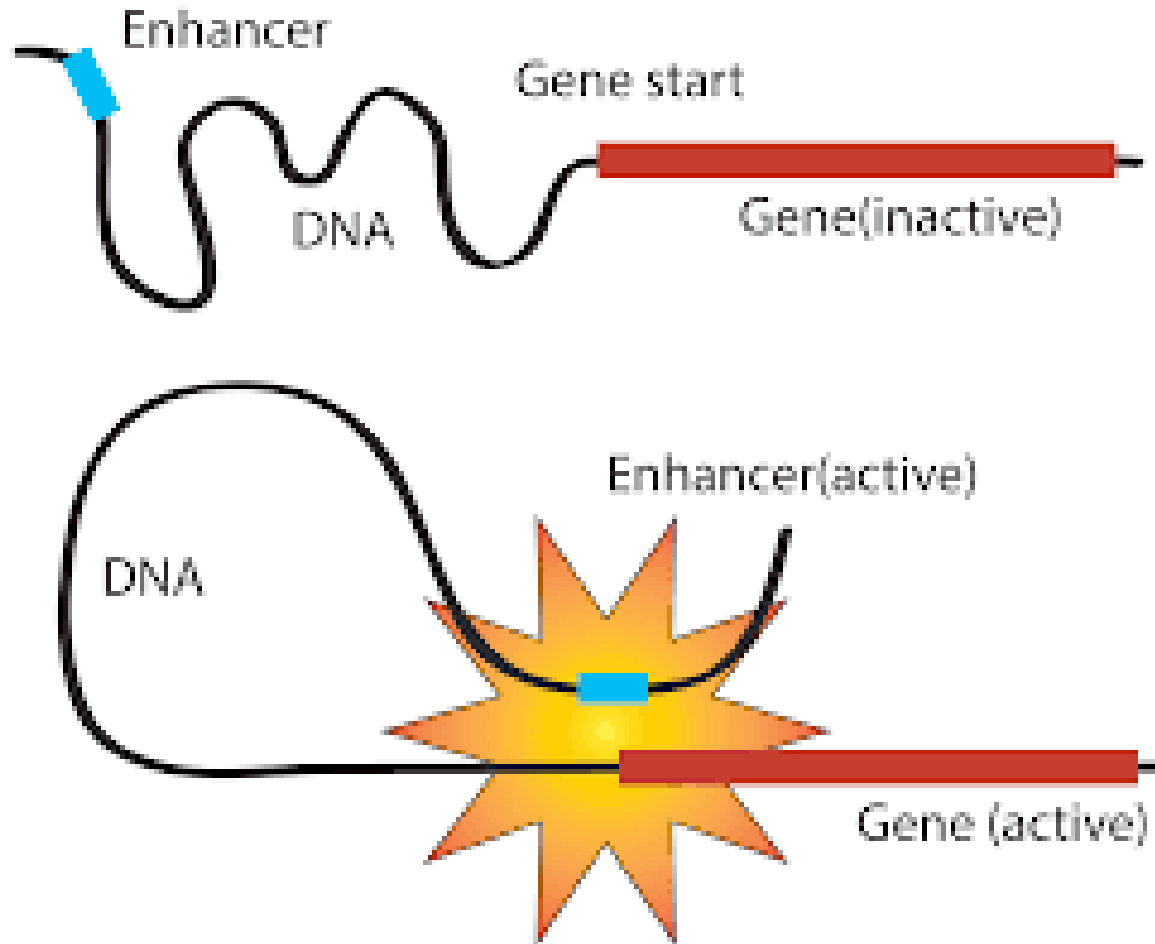
[Andersson](#) et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461 (27 March 2014)

Disease SNPs are often seen in cell type specific enhancers



[Andersson](#) et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461 (27 March 2014)

What is an enhancer?



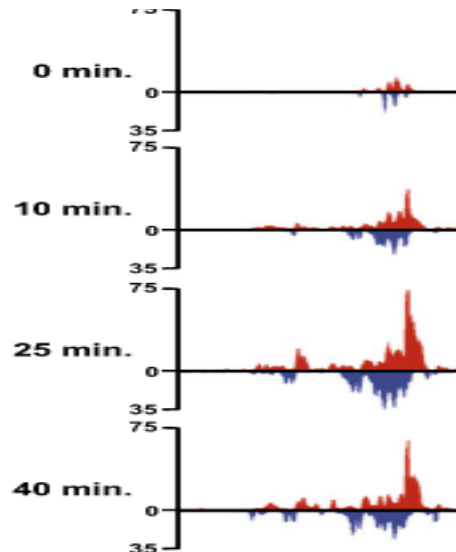
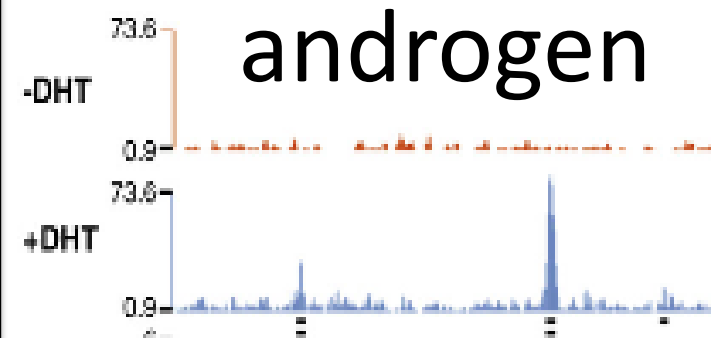
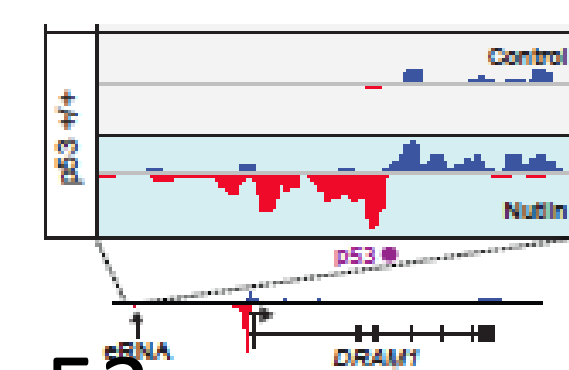
eRNAs' transcription levels are controlled by transcription factors (TF)

eRNAs

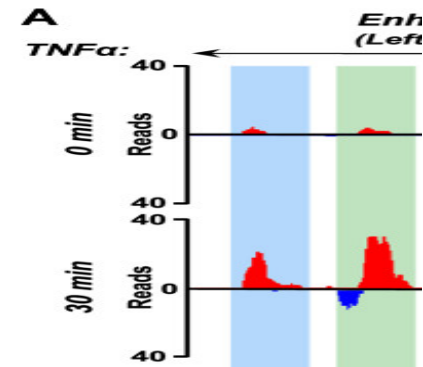


p53

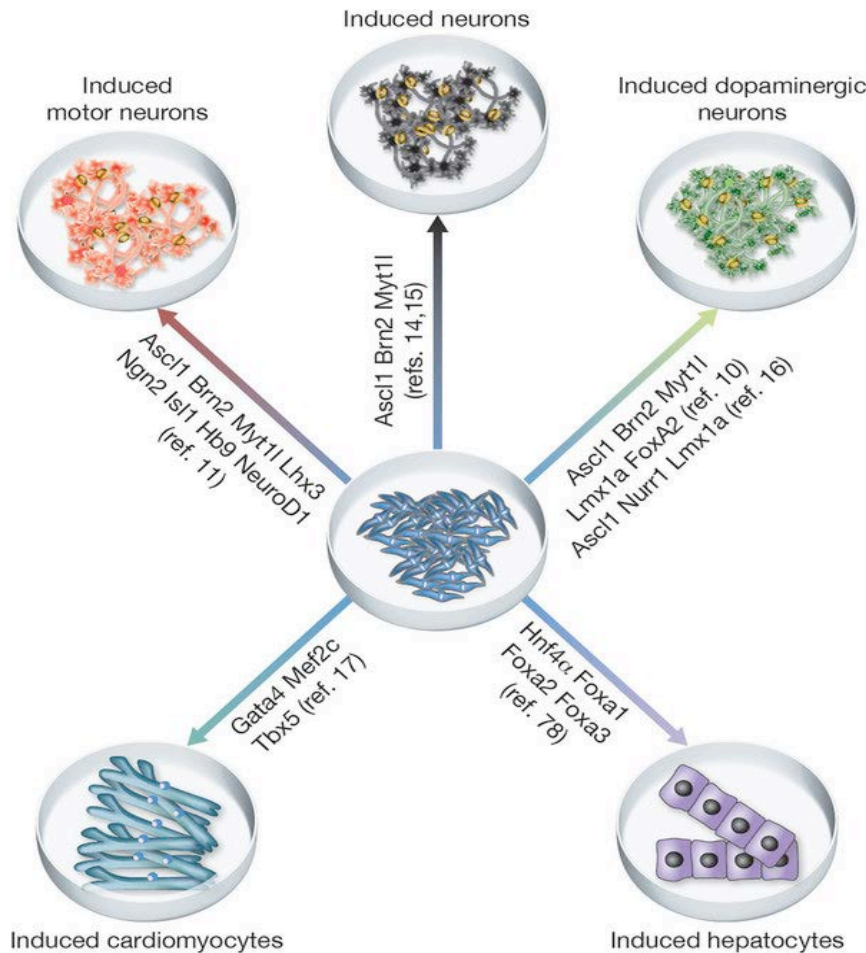
estrogen



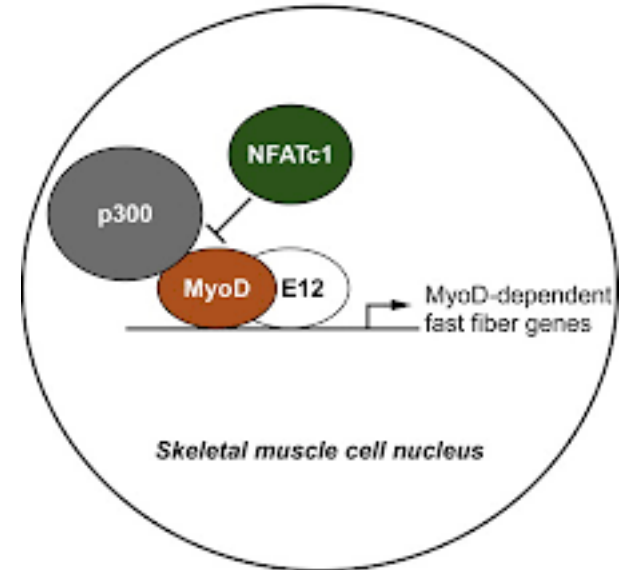
NF- κ B



Transcription factors make a muscle cell a muscle cell and a stem cell a stem cell



Transcription Factor



Medical Application: Mogrify – a new computational tool for direct reprogramming of cell type

Degradation of an RNA can occur so quickly it is not detectable by RNA-seq

