All the ”-omics”

C. Ryan Campbell

Duke University

c.ryan.campbell@duke.edu

26 Sept 2017

Overview

Today's Goals

- What other questions can genomics answer?
- What technologies underlie these methods?

- Group dataset selection

Methods in Context

My goal is to present the methods as:

- Biological Question

  (Broad enough to apply to whatever system)
- Underlying NGS
- A paper that utilizes the method

RNAseq

- Biological Question:

RNAseq

- Biological Question:
  - What genes are being expressed in a cell?

RNAseq

- Biological Question:
  - What genes are being expressed in a cell?
- Underlying NGS:

RNAseq

- Biological Question:
  - What genes are being expressed in a cell?
- Underlying NGS:
  - Illumina NGS with special library prep

RNAseq

- Biological Question:
  - What genes are being expressed in a cell?
- Underlying NGS:
  - Illumina NGS with special library prep
- Faherty. 2016. Gene Expression Profiling in the Hibernating Primate, Cheirogaleus Medius.

ASE

- Allele-Specific Expression

ASE

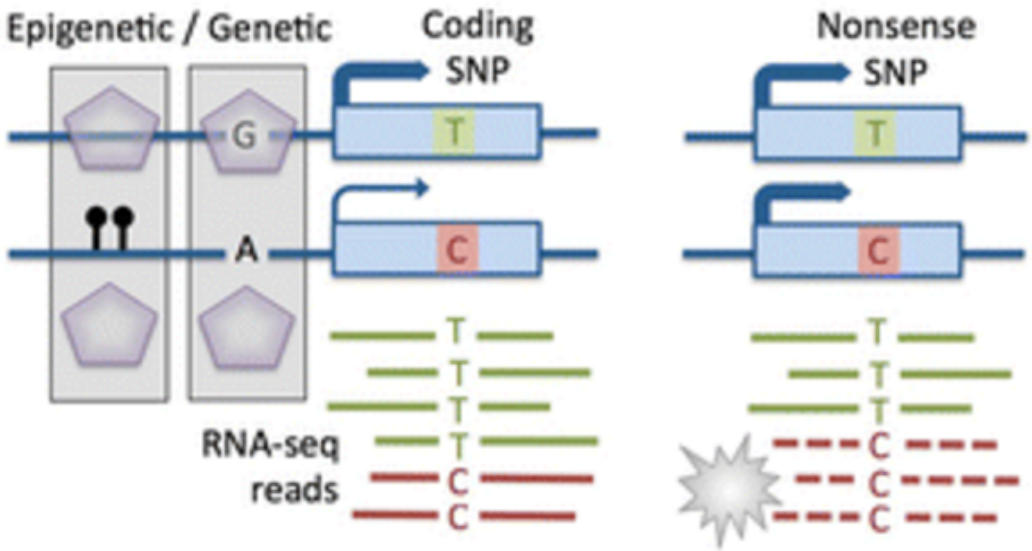
- Allele-Specific Expression
- Quantification of gene expression by diploid copy

ASE

- Allele-Specific Expression
- Quantification of gene expression by diploid copy
- Can determine effect of small-scale changes on expression

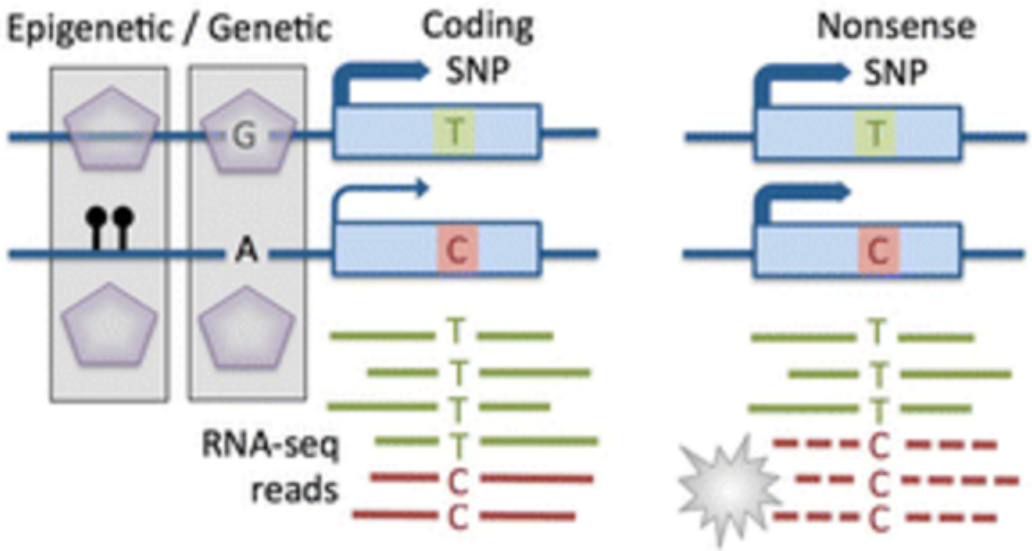
ASE

- Comparing haplotypes allows for detection of differences:



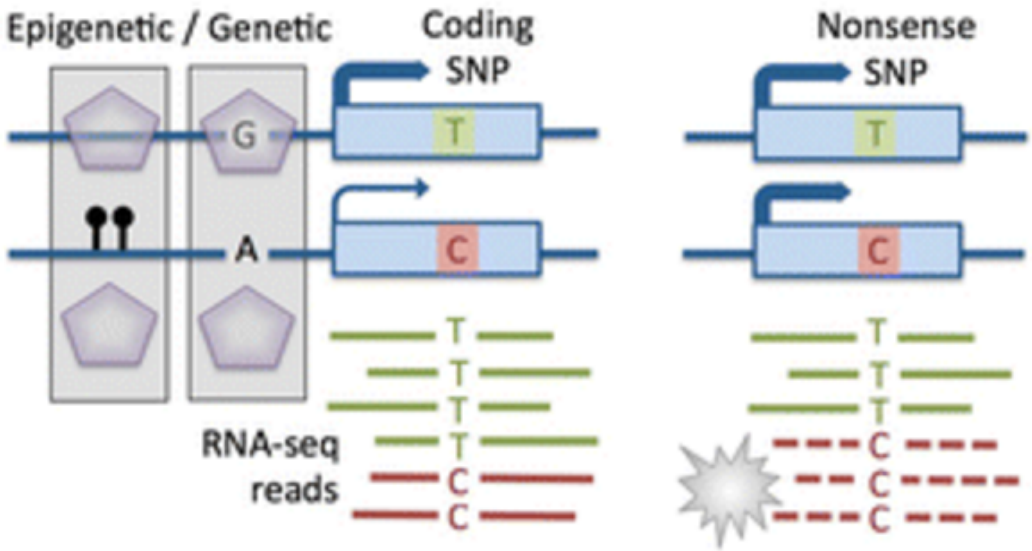
ASE

- Comparing haplotypes allows for detection of differences:
- Stop Codons



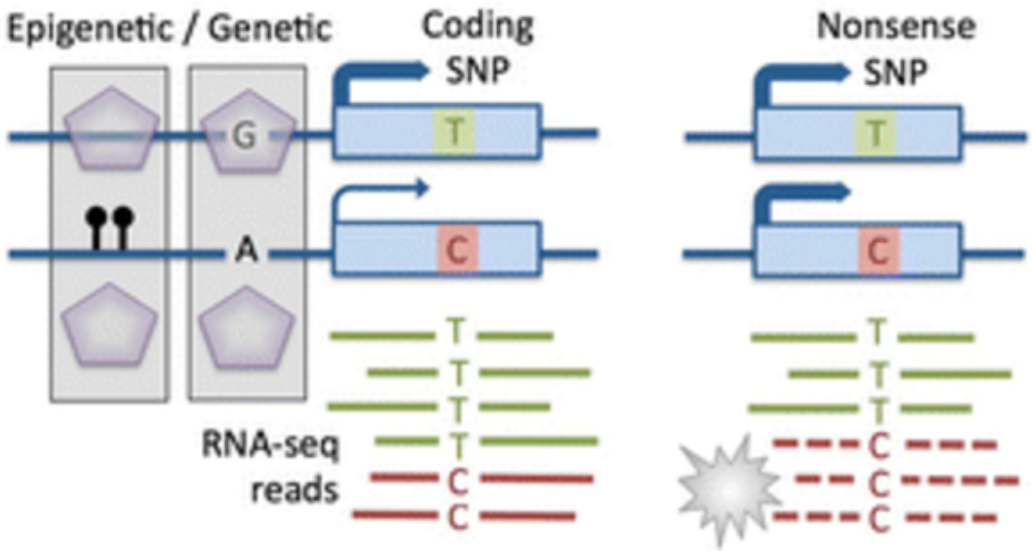
ASE

- Comparing haplotypes allows for detection of differences:
- Stop Codons
- SNPs



ASE

- Comparing haplotypes allows for detection of differences:
- Stop Codons
- SNPs
- TF binding sites or methylation



ASE

- Biological Question:

ASE

- Biological Question:
  - What of two dipliod gene copies are being expressed?

ASE

- Biological Question:
  - What of two dipliod gene copies are being expressed?
- Underlying NGS:

ASE

- Biological Question:
  - What of two dipliod gene copies are being expressed?
- Underlying NGS:
  - Illumina NGS with special library prep and post-processing

ASE

- Biological Question:
  - What of two dipliod gene copies are being expressed?
- Underlying NGS:
  - Illumina NGS with special library prep and post-processing
- Lappalainen et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans.

Iso-Seq

- Isoform Sequencing

Iso-Seq

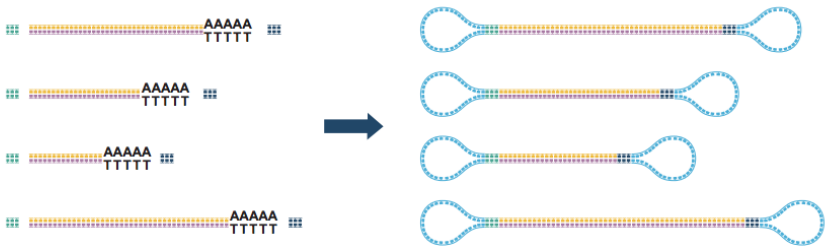
- Isoform Sequencing
- Long read RNA sequencing

Iso-Seq

- Isoform Sequencing
- Long read RNA sequencing
- Allows for the full transcript to be sequenced

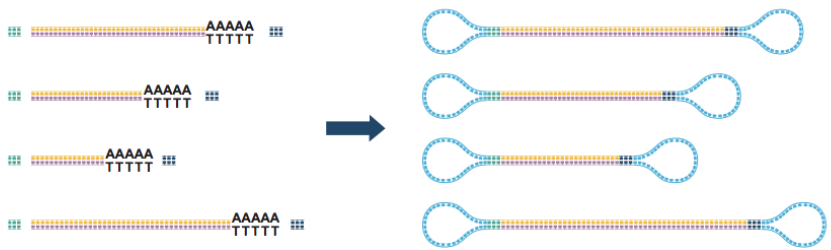
Iso-Seq

- RNAseq library preparation



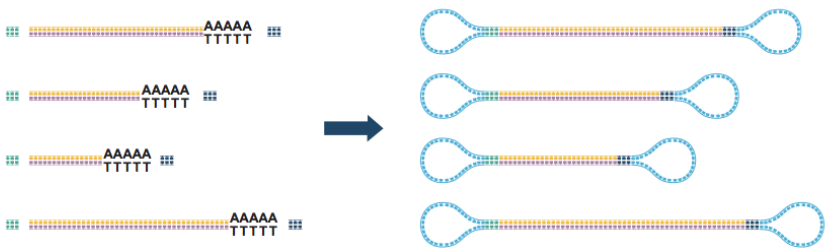
Iso-Seq

- RNAseq library preparation
- No fragmentation and keep full transcripts



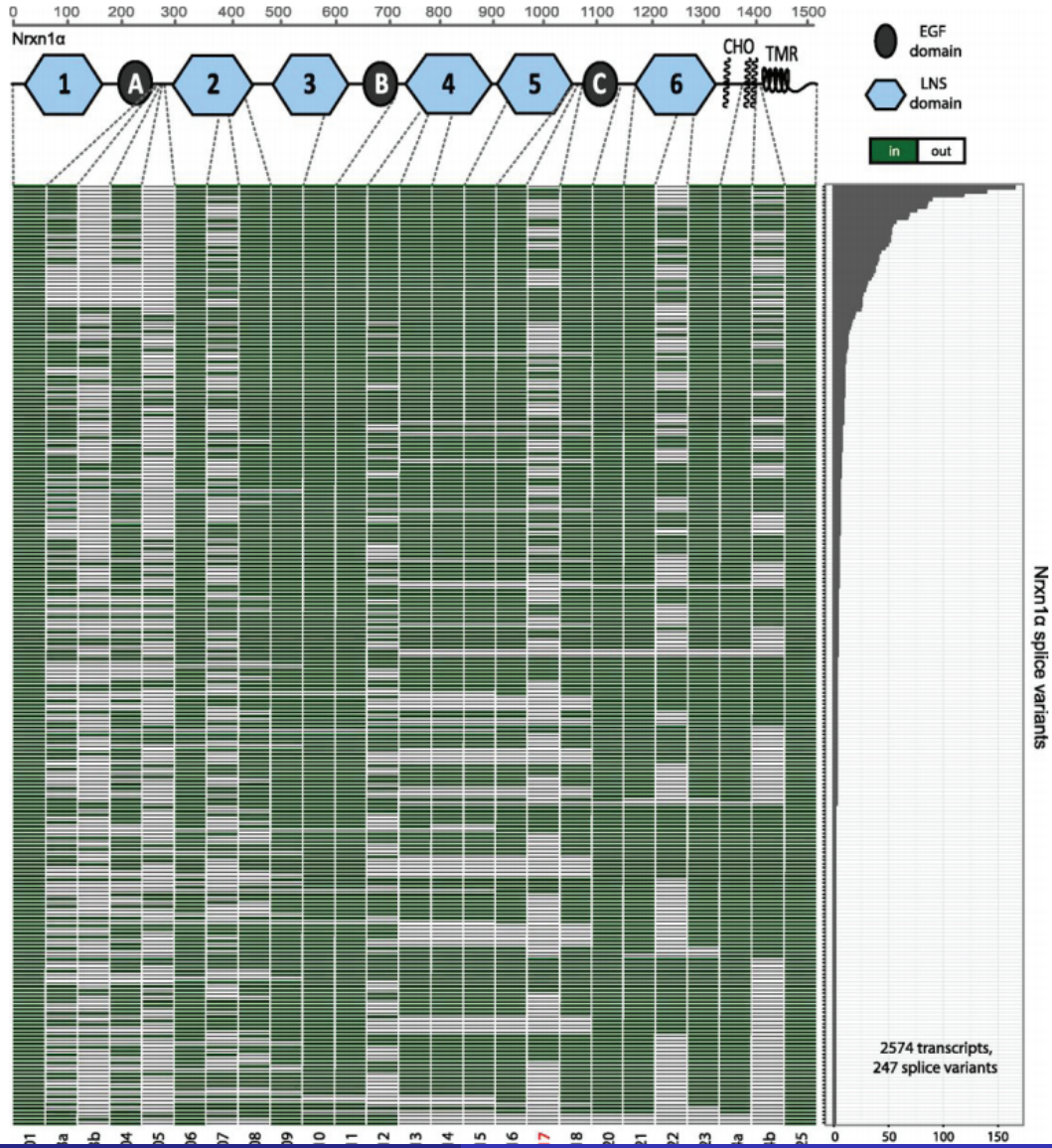
Iso-Seq

- RNAseq library preparation
- No fragmentation and keep full transcripts
- Prepare and sequence on the PacBio platform



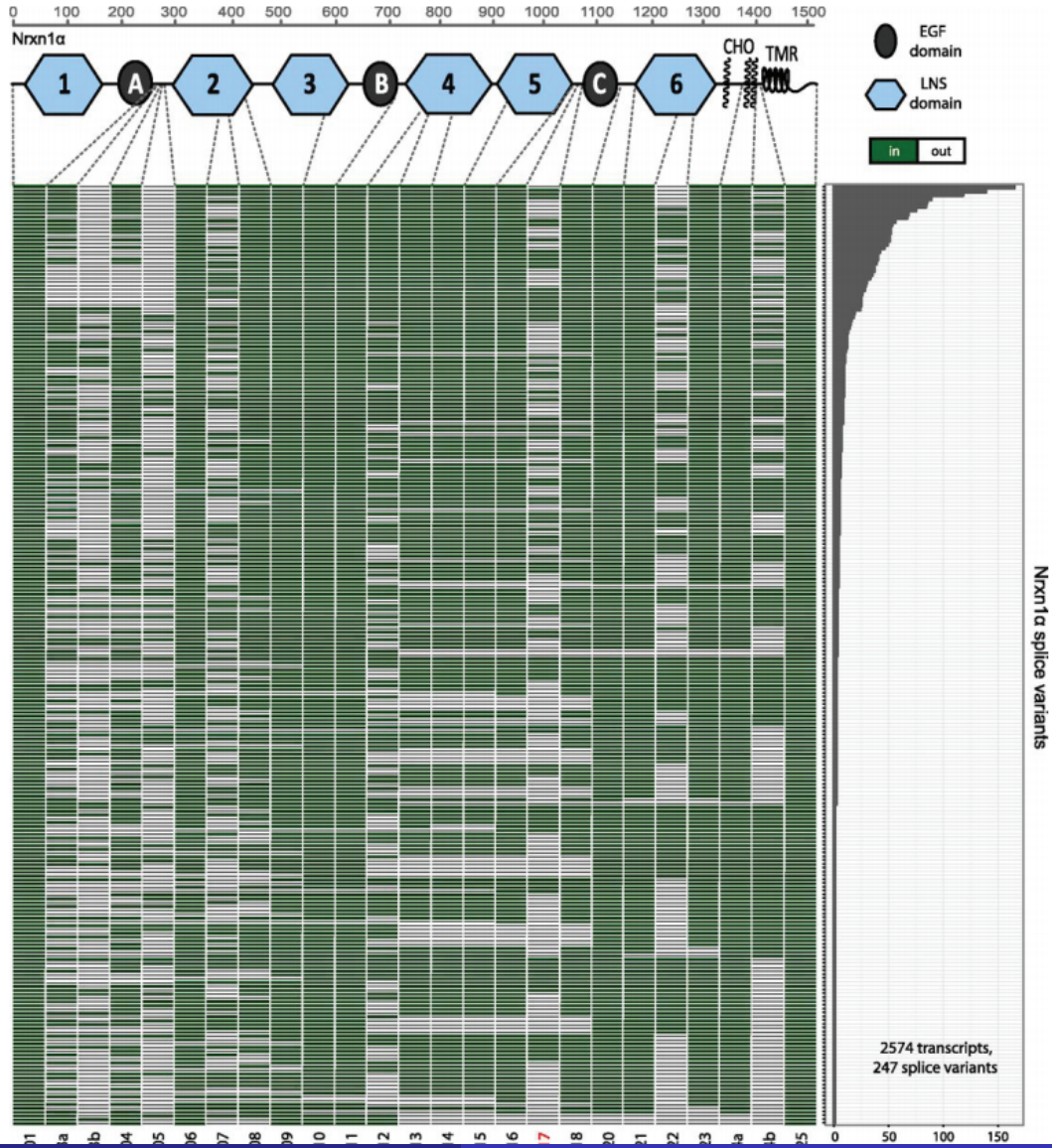
Iso-Seq

- Get longer data



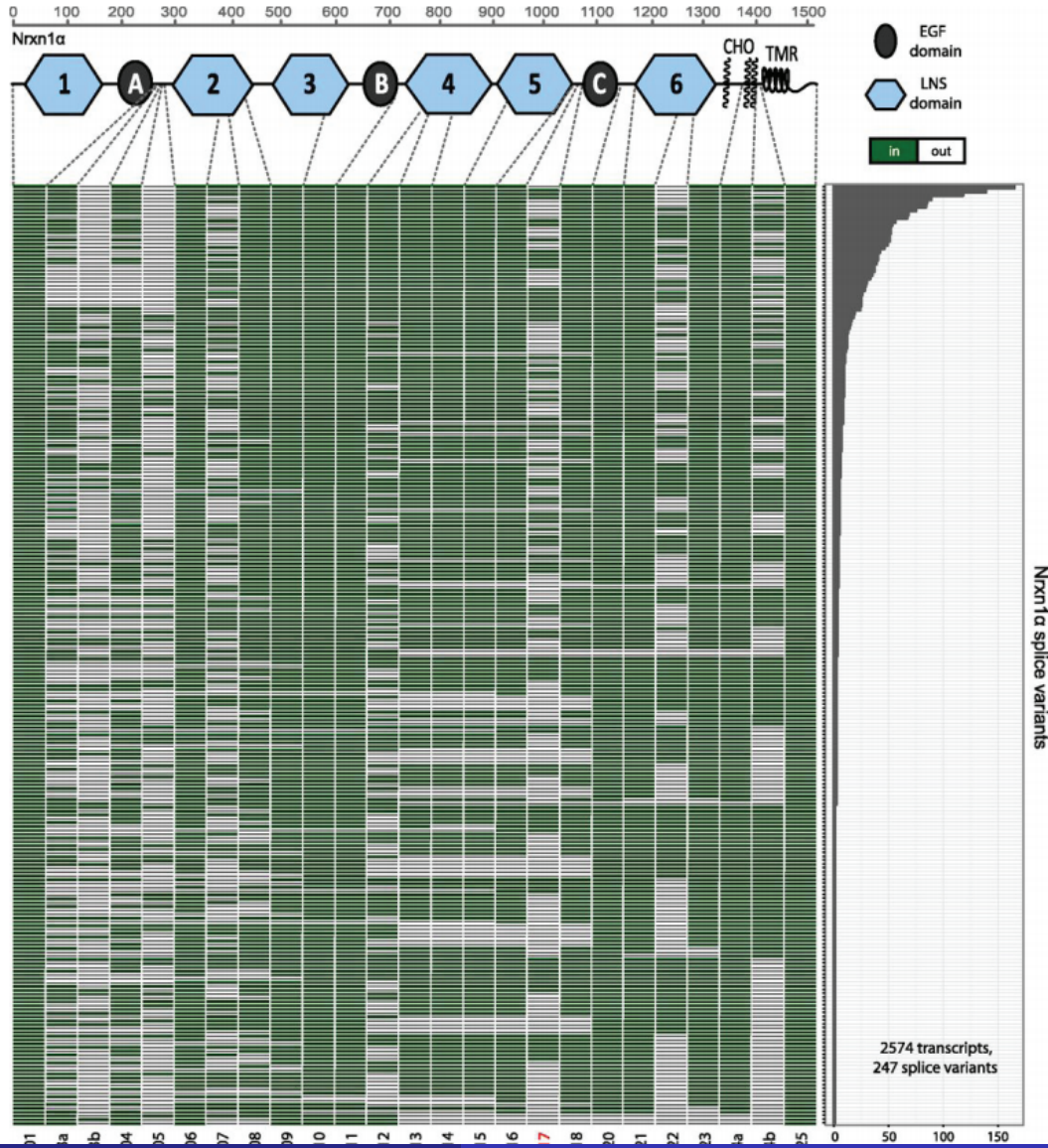
Iso-Seq

- Get longer data
- Don't have to assemble full transcripts



Iso-Seq

- Get longer data
- Don’t have to assemble full transcripts
- Actually “see” all the isoforms



Iso-Seq

- Biological Question:

Iso-Seq

- Biological Question:
  - What isoforms of a gene are being expressed?

Iso-Seq

- Biological Question:
  - What isoforms of a gene are being expressed?
- Underlying NGS:

Iso-Seq

- Biological Question:
  - What isoforms of a gene are being expressed?
- Underlying NGS:
  - RNAseq using PacBio to sequencing whole transcripts

Iso-Seq

- Biological Question:
  - What isoforms of a gene are being expressed?
- Underlying NGS:
  - RNAseq using PacBio to sequencing whole transcripts
- Cheng et al. 2017. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts.

Microbiome

- What is the community of microbes that live within vertebrates?

Microbiome

- What is the community of microbes that live within vertebrates?
- Gut is most commonly studied, but there are plenty of other targets

Microbiome

- Often longitudinal:

Microbiome

- Often longitudinal:
  - How does this community change over a lifespan?

Microbiome

- Often longitudinal:
  - How does this community change over a lifespan?
  - How do antibiotics alter this community?

Microbiome

- Often longitudinal:
  - How does this community change over a lifespan?
  - How do antibiotics alter this community?
- Also can compare across populations (diet effect)

Microbiome

- Often longitudinal:
  - How does this community change over a lifespan?
  - How do antibiotics alter this community?
- Also can compare across populations (diet effect)
- Can often tell a differnce based on lifestyle

16S Region

- 16S is a “barcode” region of bacterial genomes

16S Region

- 16S is a “barcode” region of bacterial genomes
  - Often stable (encodes ribosomal RNA molecule)

16S Region

- 16S is a “barcode” region of bacterial genomes
  - Often stable (encodes ribosomal RNA molecule)
  - Used to distinguish “species”

16S Region

- 16S is a “barcode” region of bacterial genomes
  - Often stable (encodes ribosomal RNA molecule)
  - Used to distinguish “species”
- The most common form of microbiome research amplifies this region

16S Region

- 16S is a “barcode” region of bacterial genomes
  - Often stable (encodes ribosomal RNA molecule)
  - Used to distinguish “species”
- The most common form of microbiome research amplifies this region
- Then sequencing is performed to assess the community

Metagenomics

- Broad, non-targeted method of censusing the microbiome

Metagenomics

- Broad, non-targeted method of censusing the microbiome
- Gather a sample and prepare all the DNA therein for sequencing

Metagenomics

- Broad, non-targeted method of censusing the microbiome
- Gather a sample and prepare all the DNA therein for sequencing
- Catches everything, unbiased

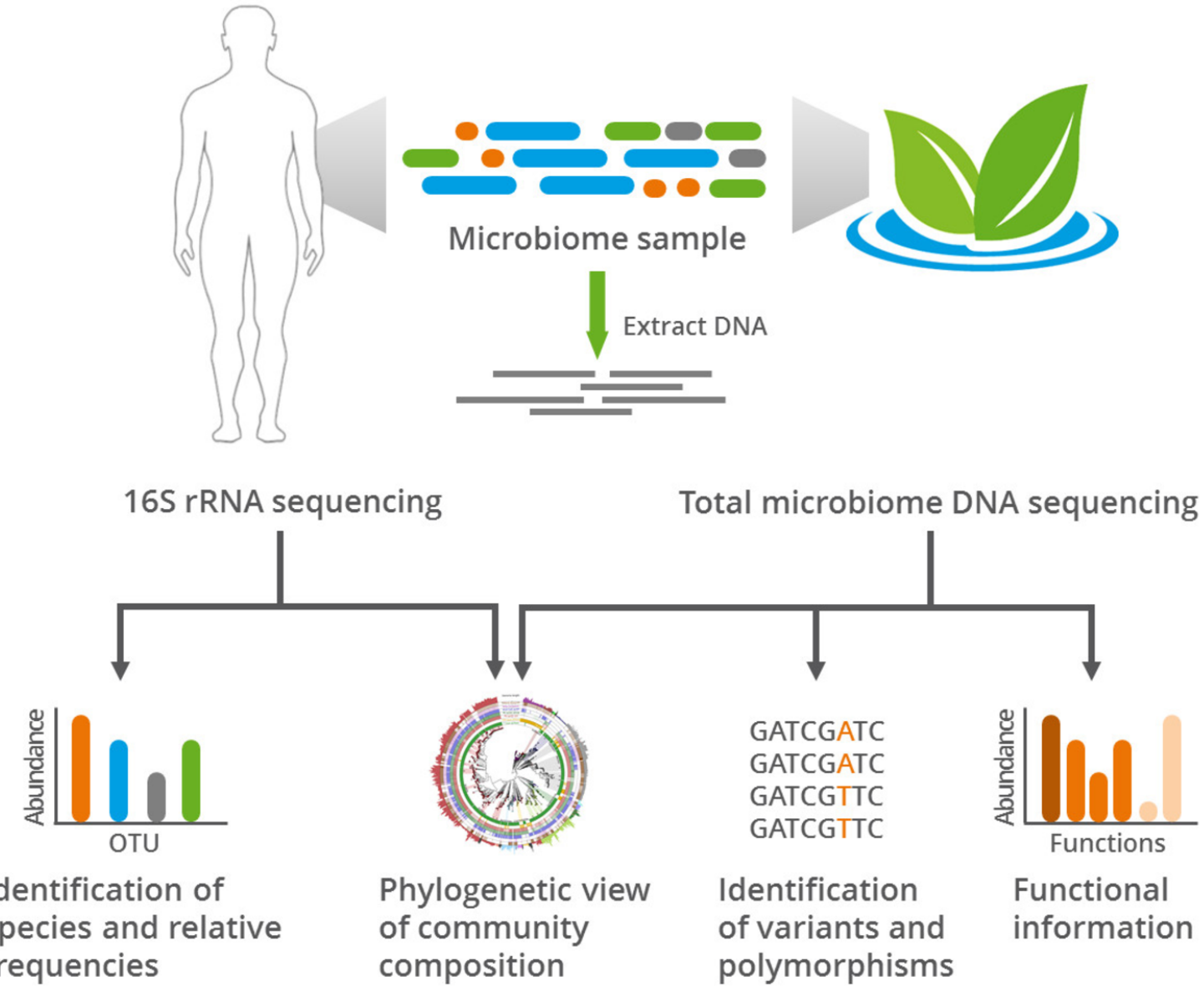
Metagenomics

- Broad, non-targeted method of censusing the microbiome
- Gather a sample and prepare all the DNA therein for sequencing
- Catches everything, unbiased
- Categorization is often difficult

Metagenomics

- Broad, non-targeted method of censusing the microbiome
- Gather a sample and prepare all the DNA therein for sequencing
- Catches everything, unbiased
- Categorization is often difficult
- Captures functional capabilities of microbiota

16S v. Metagenomics



Microbiome

- Biological Question:

Microbiome

- Biological Question:
  - What bacteria live in the gut?

Microbiome

- Biological Question:
  - What bacteria live in the gut?
- Underlying technology:

Microbiome

- Biological Question:
  - What bacteria live in the gut?
- Underlying technology:
  - Illumina NGS with a narrow input (16S region) and special processing

Microbiome

- Biological Question:
  - What bacteria live in the gut?
- Underlying technology:
  - Illumina NGS with a narrow input (16S region) and special processing
  - Illumina NGS with a broad input (any DNA) and special processing

Microbiome

- Biological Question:
  - What bacteria live in the gut?
- Underlying technology:
  - Illumina NGS with a narrow input (16S region) and special processing
  - Illumina NGS with a broad input (any DNA) and special processing
- Davie et al. 2014. Diet rapidly and reproducibly alters the human gut microbiome.

Epi-

- Epi- means?

Epi-

- Epi- means?
- On, upon

Epi-

- Epi- means?
- On, upon
- Above

Epigenomics

- Changes that occur “on” the genome

Epigenomics

- Changes that occur “on” the genome
- More broadly - changes that aren’t encoded in the DNA

Epigenomics

- Changes that occur “on” the genome
- More broadly - changes that aren’t encoded in the DNA
  - Methylation of Cytosines

Epigenomics

- Changes that occur “on” the genome
- More broadly - changes that aren’t encoded in the DNA
  - Methylation of Cytosines
  - Chromatin status

Epigenomics

- Changes that occur “on” the genome
- More broadly - changes that aren’t encoded in the DNA
  - Methylation of Cytosines
  - Chromatin status
  - microRNAs

Methylation

- Methyl group is added to a cytosine

Methylation

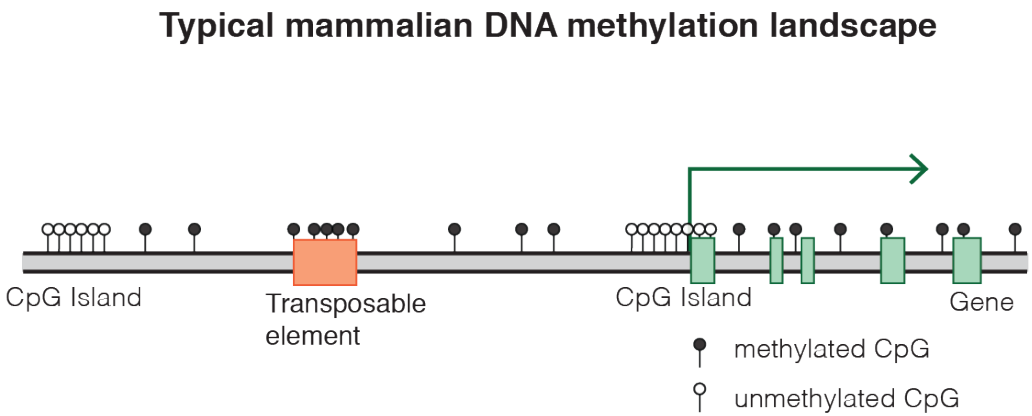
- Methyl group is added to a cytosine
- (Adenine can too, just less common)

Methylation

- Methyl group is added to a cytosine
- (Adenine can too, just less common)
- Supresses transcription if it occurs in a gene promoter region

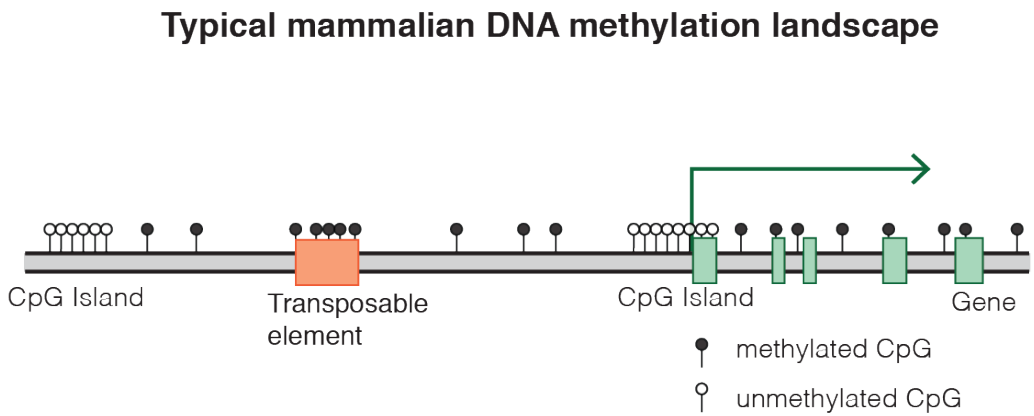
Methylation

- Often present near TE's and Genes



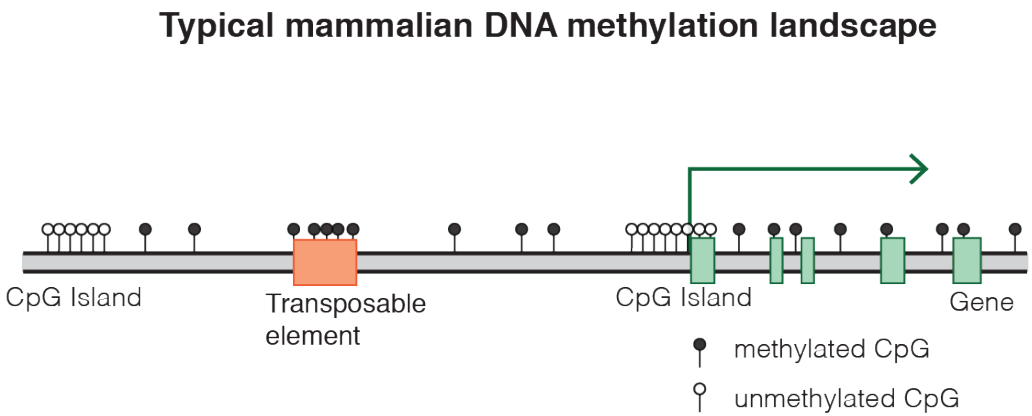
Methylation

- Often present near TE's and Genes
- Presence mutes TE's



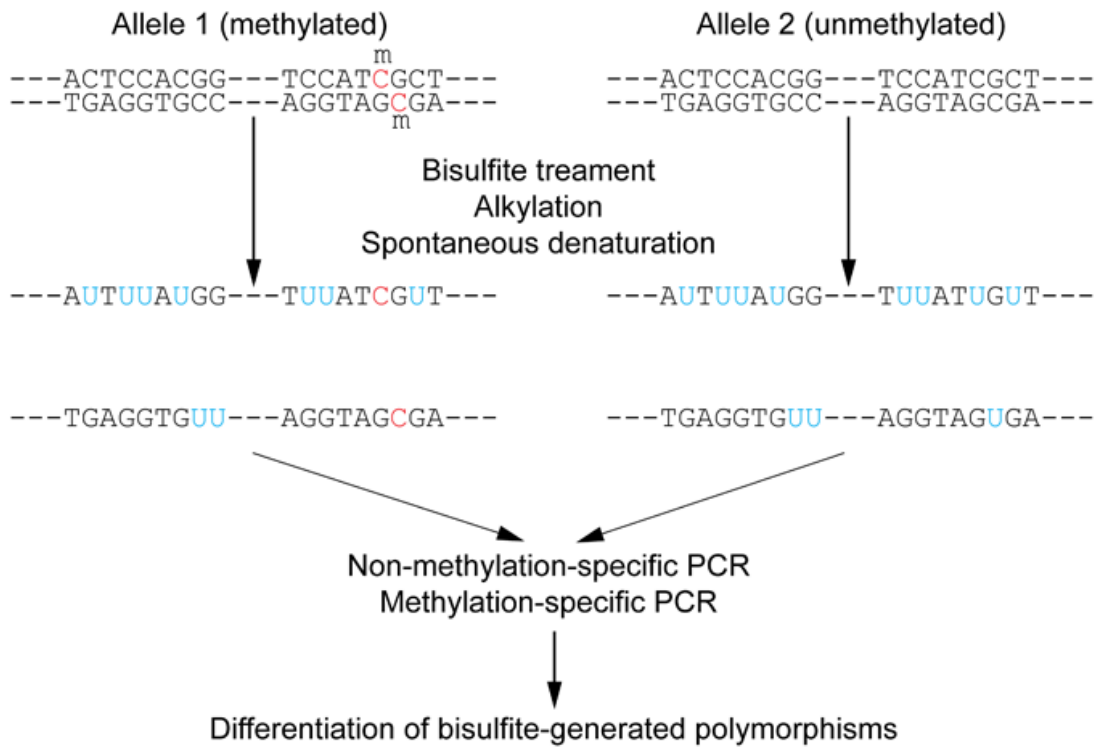
Methylation

- Often present near TE's and Genes
- Presence mutes TE's
- Absence allows gene transcription



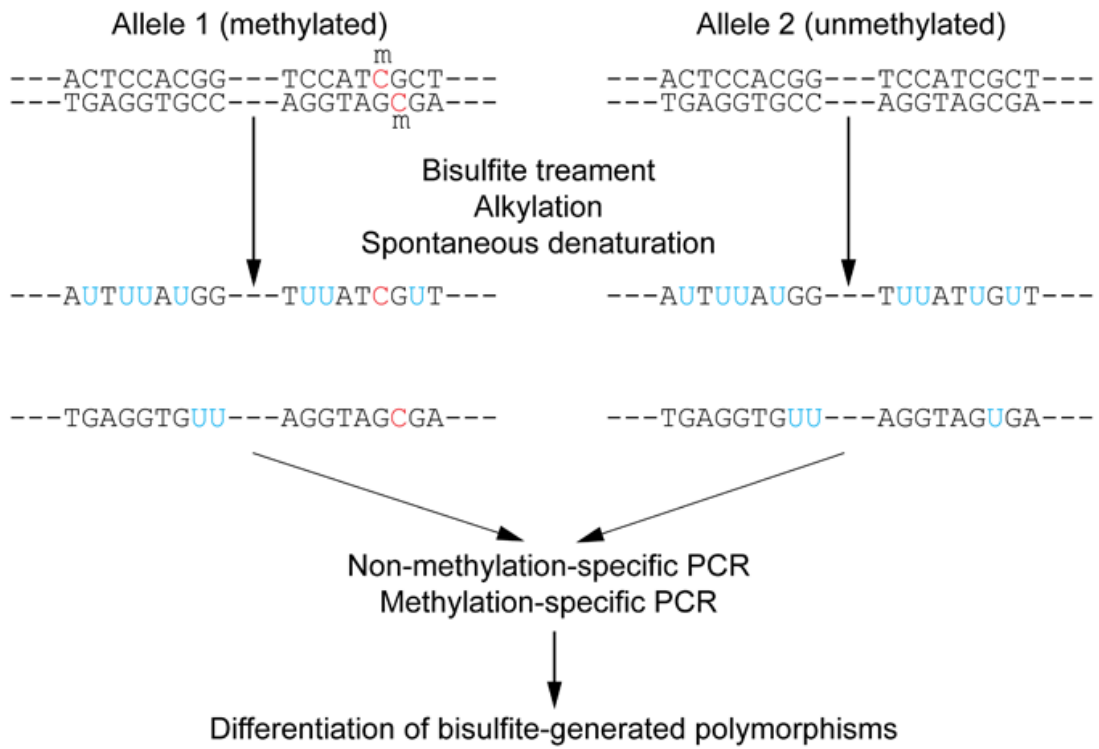
Bisulfite Sequencing

- Library prep converts unmethylated C's to U's



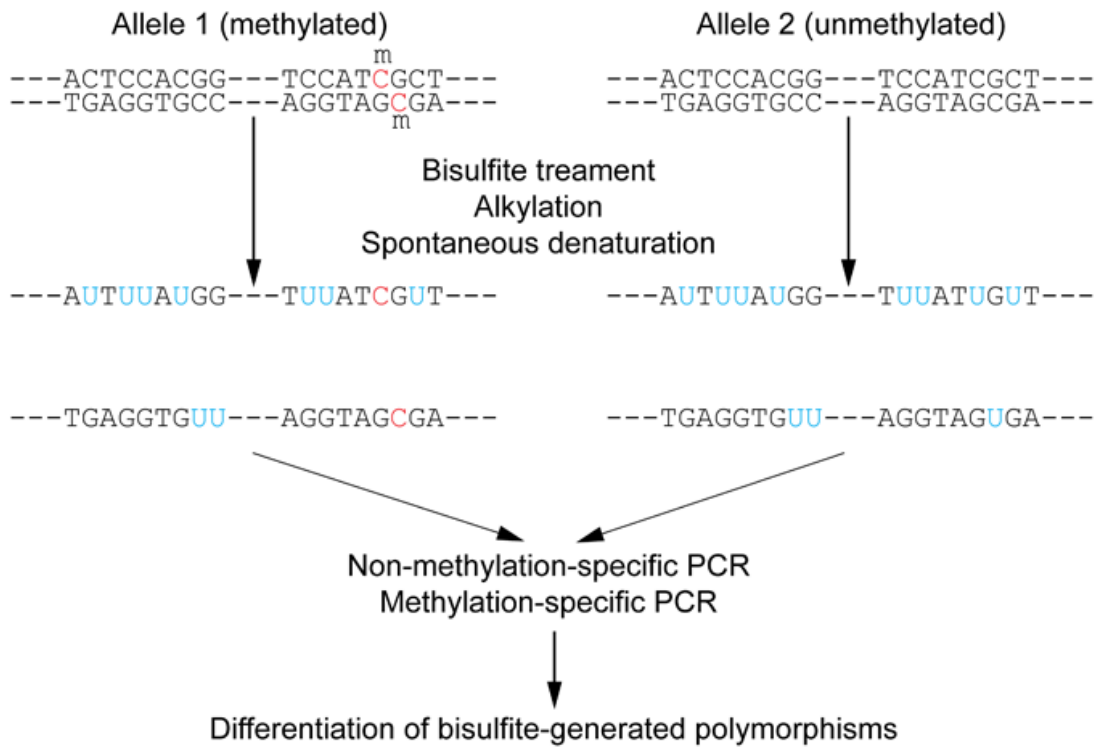
Bisulfite Sequencing

- Library prep converts unmethylated C's to U's
- Leaves methylated C's as C's



Bisulfite Sequencing

- Library prep converts unmethylated C's to U's
- Leaves methylated C's as C's
- Sequence to determine where methylated sites are



Bisulfite Sequencing

- Biological Question:

Bisulfite Sequencing

- Biological Question:
  - Which bases in the genome are methylated?

Bisulfite Sequencing

- Biological Question:
  - Which bases in the genome are methylated?
- Underlying NGS:

Bisulfite Sequencing

- Biological Question:
  - Which bases in the genome are methylated?
- Underlying NGS:
  - Illumina NGS with bisulfite conversion

Bisulfite Sequencing

- Biological Question:
  - Which bases in the genome are methylated?
- Underlying NGS:
  - Illumina NGS with bisulfite conversion
- Lea et al. 2016. Resource base influences genomewide DNA methylation levels in wild baboons (Papio cynocephalus).

ATAC-seq

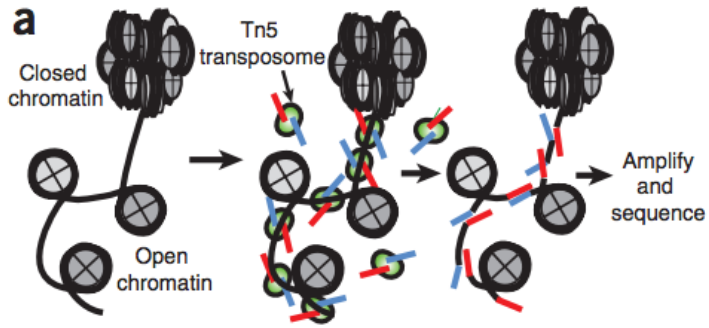
- Assay for Transposase-Accessible Chromatin using sequencing

ATAC-seq

- Assay for Transposase-Accessible Chromatin using sequencing
-

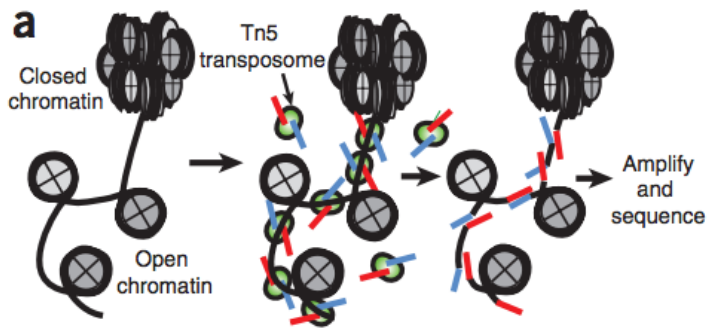
ATAC-seq

- Transposase (green)



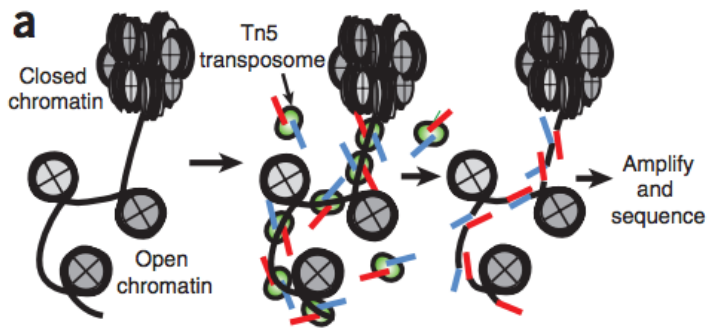
ATAC-seq

- Transposase (green)
- Sequencing adaptors (red and blue)

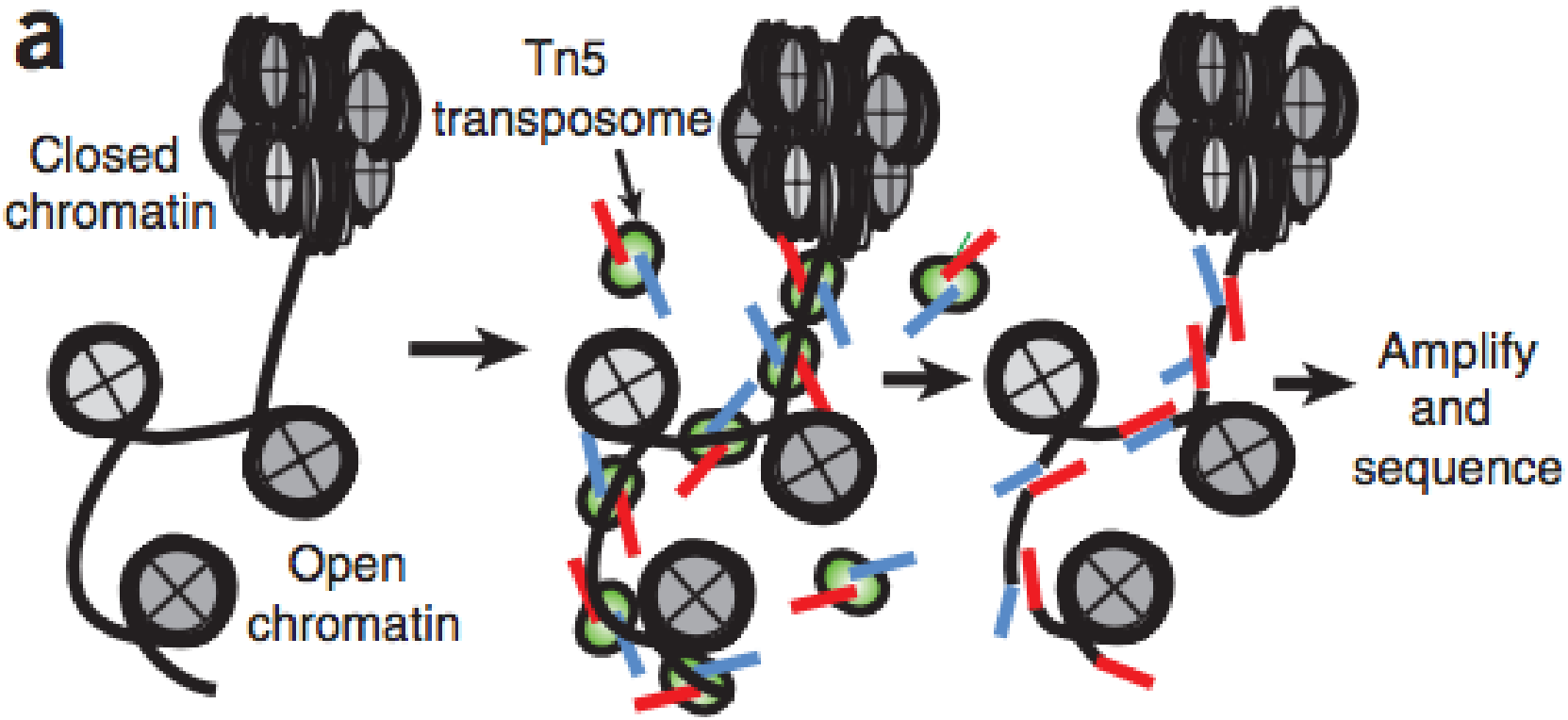


ATAC-seq

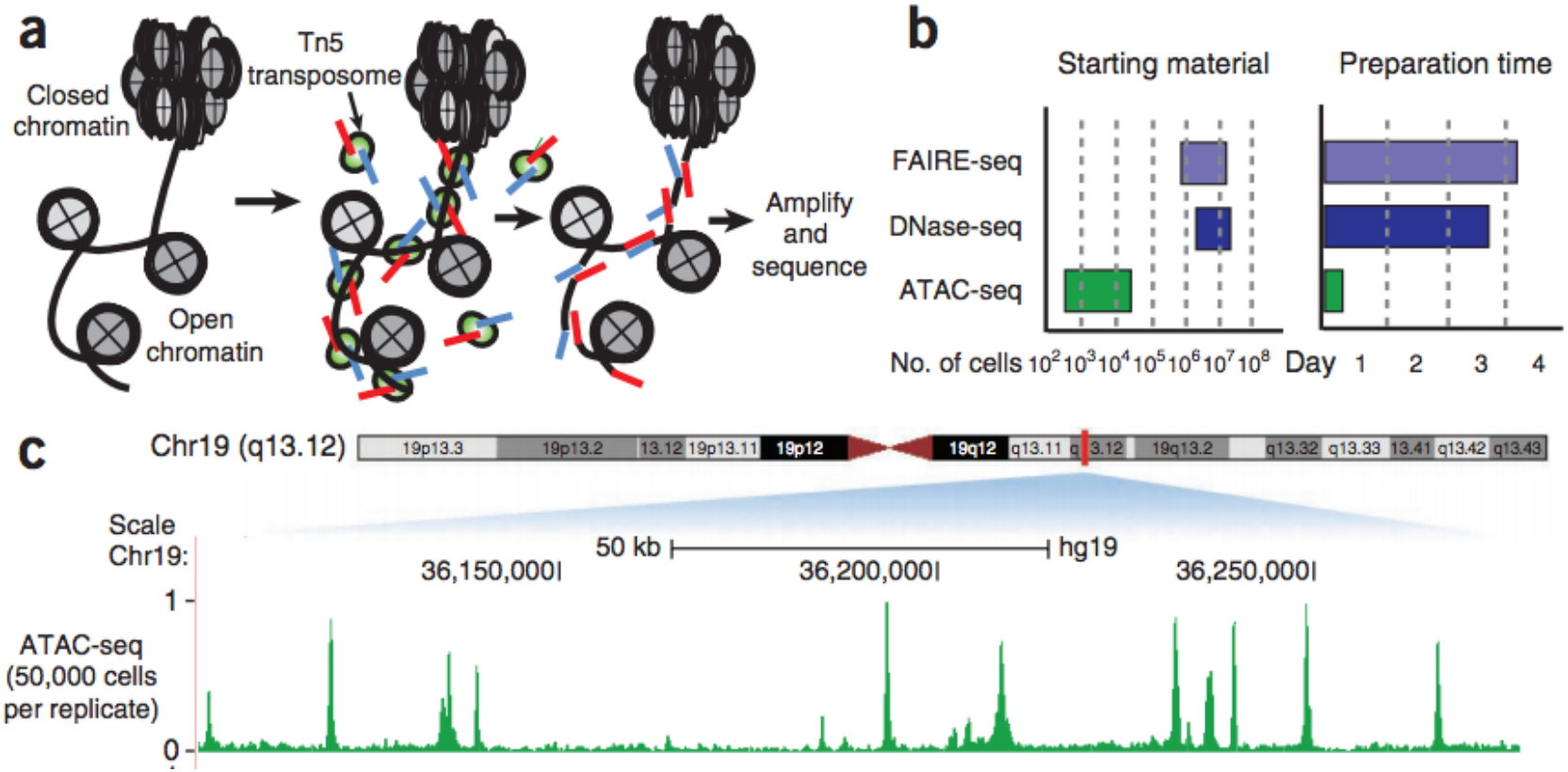
- Transposase (green)
- Sequencing adaptors (red and blue)
- Inserts only in regions of open chromatin, between nucleosomes



ATAC-seq


a
Closed chromatin
Tn5 transposome
Open chromatin
Amplify and sequence

ATAC-seq Results



ATAC-seq

- Biological Question:

ATAC-seq

- Biological Question:
  - Which regions of the genome are accessible?

ATAC-seq

- Biological Question:
  - Which regions of the genome are accessible?
- Underlying NGS:

ATAC-seq

- Biological Question:
  - Which regions of the genome are accessible?
- Underlying NGS:
  - Cutting open regions out and prep for Illumina NGS

ATAC-seq

- Biological Question:
  - Which regions of the genome are accessible?
- Underlying NGS:
  - Cutting open regions out and prep for Illumina NGS
- Ackerman et al. 2016. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes.

Non-Epigenomics

- These are other measures “above” the DNA-level

Non-Epigenomics

- These are other measures “above” the DNA-level
- However they’re not heritable

Non-Epigenomics

- These are other measures “above” the DNA-level
- However they’re not heritable
- So not technically epigenetic

Non-Epigenomics

- These are other measures “above” the DNA-level
- However they’re not heritable
- So not technically epigenetic
- (But the methods are very similar)

CHIPseq

- Chromatin Immunoprecipitation sequencing

CHIPseq

- Chromatin Immunoprecipitation sequencing
- Capture DNA that is currently bound to a protein

CHIPseq

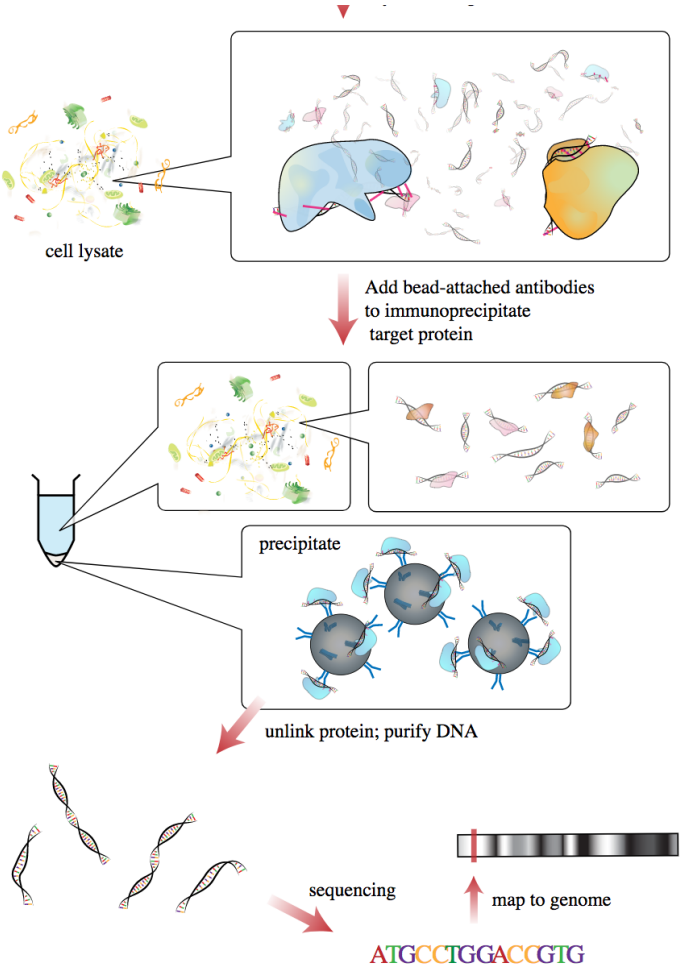
- Chromatin Immunoprecipitation sequencing
- Capture DNA that is currently bound to a protein
- Works for any protein of interest

CHIPseq

- Chromatin Immunoprecipitation sequencing
- Capture DNA that is currently bound to a protein
- Works for any protein of interest
- Often used in Transcription Factor work

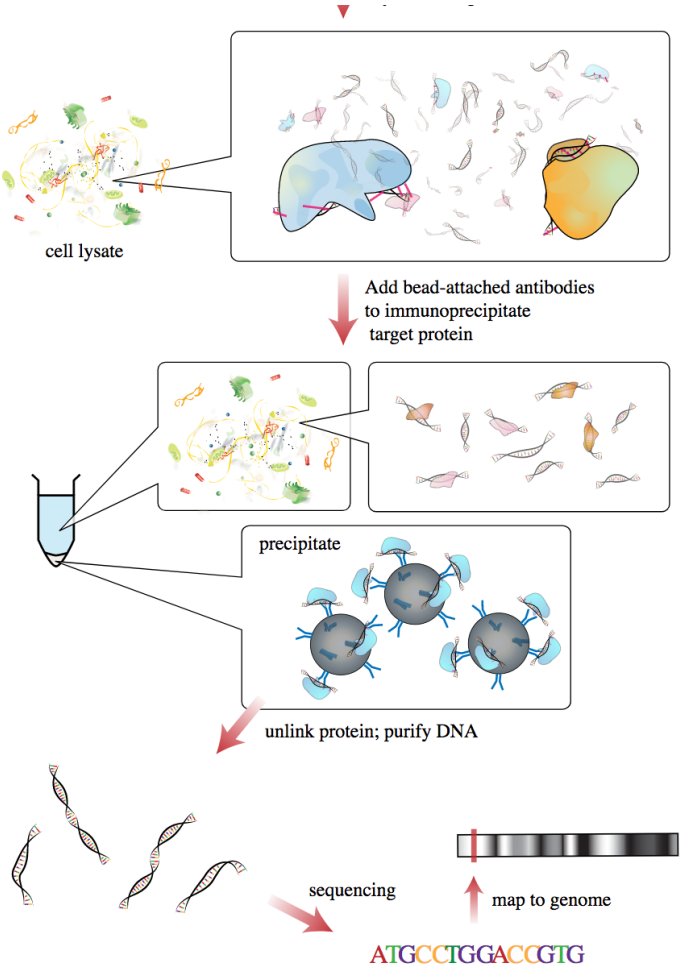
ATAC-seq

- Fix protein to DNA



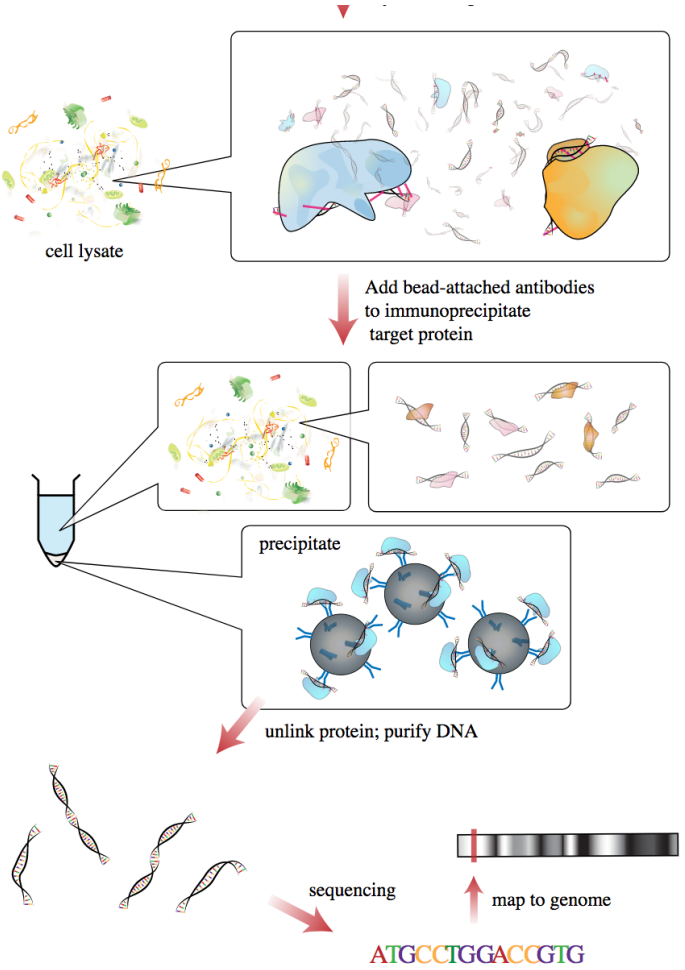
ATAC-seq

- Fix protein to DNA
- Break up genome



ATAC-seq

- Fix protein to DNA
- Break up genome
- Capture and sequence only DNA that is attached to protein



CHIPseq

- Biological Question:

CHIPseq

- Biological Question:
  - Where is a given protein bound to DNA?

CHIPseq

- Biological Question:
  - Where is a given protein bound to DNA?
- Underlying NGS:

CHIPseq

- Biological Question:
  - Where is a given protein bound to DNA?
- Underlying NGS:
  - Protein pulldown and library prep for Illumina NGS

CHIPseq

- Biological Question:
  - Where is a given protein bound to DNA?
- Underlying NGS:
  - Protein pulldown and library prep for Illumina NGS
- Schmidt. 2010. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding.

Get with your Groups:

- Jennifer, Nayib, Chris
- Austin, Alan, Raymond - flies and bacteria
- Othmane, Kevin, Alvin
- Helena, Jake, Rahul - flies and devo
- Hank, Sisi, Joy

- Today's Tasks:
- Meet with your group
- Discuss possible datasets
- Settle on a preference and clear it with me
- Find sra link for Thursday

The End