DEPARTMENT OF COMPUTATIONAL
DCM&B
MEDICINE & BIOINFORMATICS

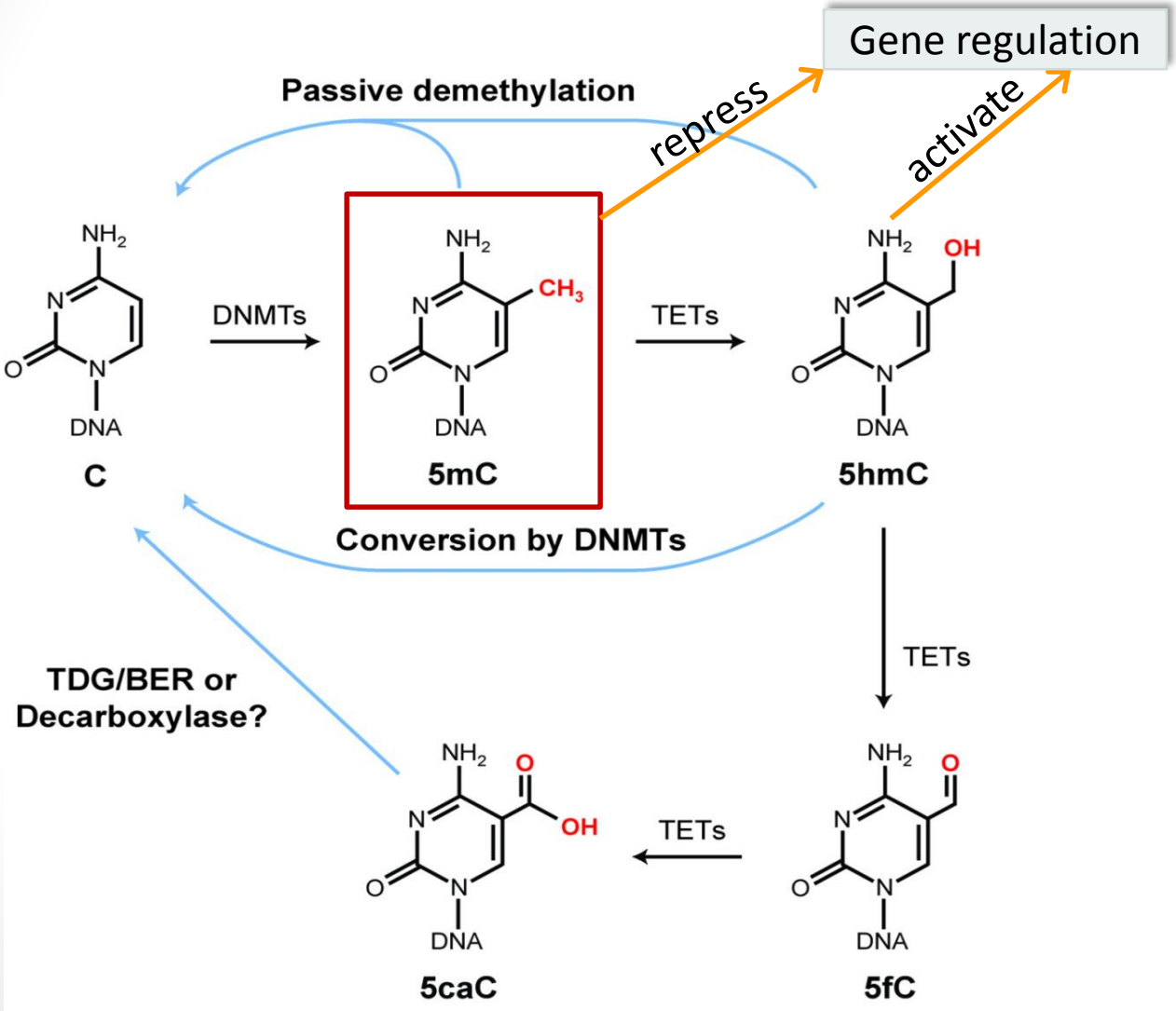
`mint` (and `annotatr`): Assessing DNA methylation and hydroxymethylation signatures in cancers

Maureen Sartor
Associate Professor
Department of Computational Medicine and Bioinformatics
University of Michigan

Overview of project

- R01 Supplement – 2 yrs
- **mint** – methylation integration
  - Command line pipeline
  - Galaxy implementation
- **annotatr** – R package for fast & flexible annotation and visualization (on bioRxiv)
- Builds off of tools developed in my lab
  - **PePr** – differential binding
  - **methylSig** – differential methylation for bisulfite sequencing data

DNA (de)methylation overview

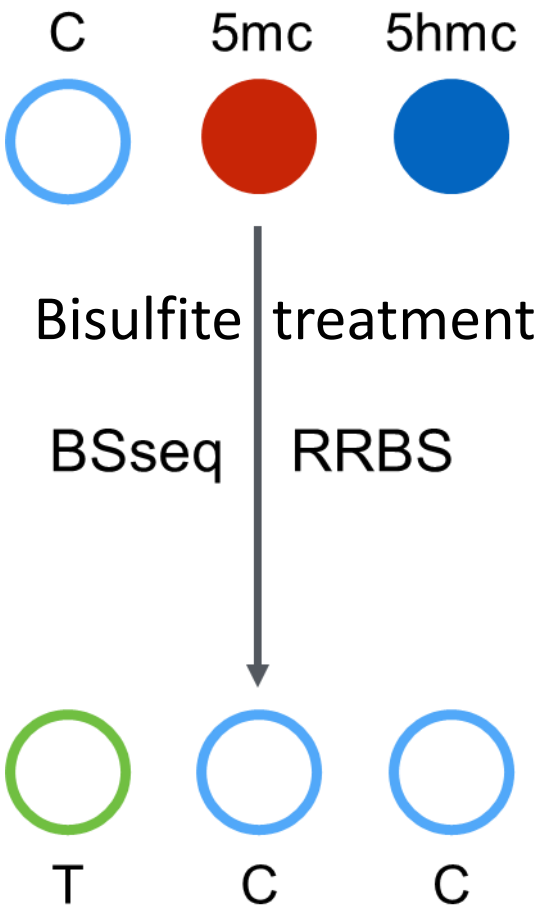


Why study 5mC and 5hmC?

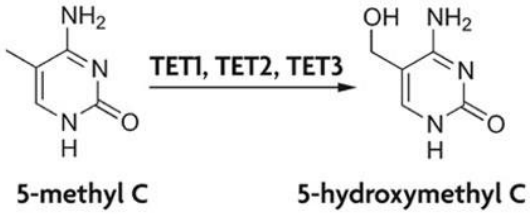
- DNA methylation (5mC) is long known to play an important role in cancers
  - Tumor suppressors are often hypermethylated, while intergenic regions are hypomethylated → DNA instability
  - Changes due to viral infections or environmental exposures and can lead to future cancer development
  - In some cancers, many of the top mutated genes are epigenetic drivers
- Hydroxymethylation (5hmC) has been observed to:
  - correlate better with gene expression levels than 5mC
  - Serve as a good biomarker for recently activated genes
  - Often occur near exon/intron boundaries and correspond to differential splicing

bisulfite sequencing

- Bisulfite treatment plus sequencing allows quantitative distinction between methylated and unmethylated CpG's
- But it does NOT distinguish between 5mC and 5hmC



Genome-wide approaches to assess DNA methylation/hydroxymethylation



Methods to assess 5mC + 5hmC:

- HumanMethylation450K BeadChip
- **WGBS (whole genome bisulfite sequencing)**
- **RRBS (reduced representation bisulfite sequencing)**

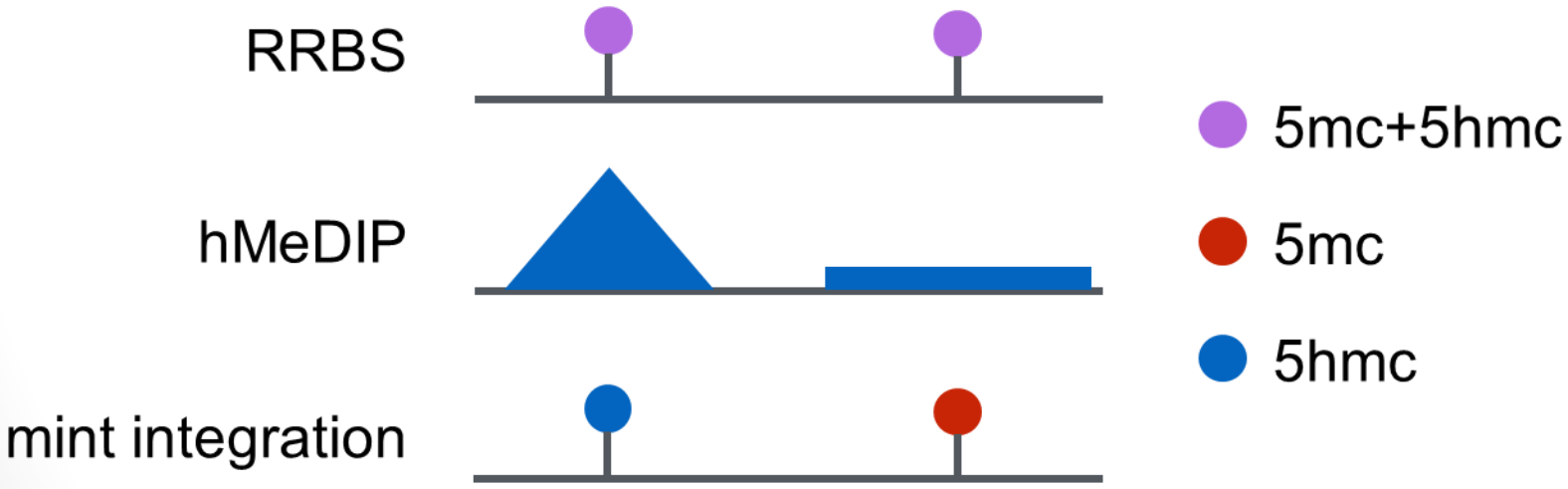
Methods to assess 5mc only:

- **MeDIP-seq or MethylCap-seq (affinity-based)**
- oxBS-seq – in parallel with bis-seq, measure 5mC only

Methods to assess 5hmC:

- TAB-seq – in parallel with bis-seq, measure 5hmC only
- **hmeDIP-seq or hMe-Seal (affinity-based)**

- Bisulfite sequencing methods – WGBS, RRBS
- Affinity-based methods – meDIPseq, hmeDIPseq



Main goals

- Complex workflows and analysis designs are required for analysis/ integration/ interpretation of 5mC/5hmC experiments

→ quick reproducibility

→ flexibility

→ transparency

Supported workflow designs

Hybrid 

Bisulfite-based assay and pulldown assay.

or

Pulldown 

Pulldown assays for mc and hmc.

and

Sample-wise

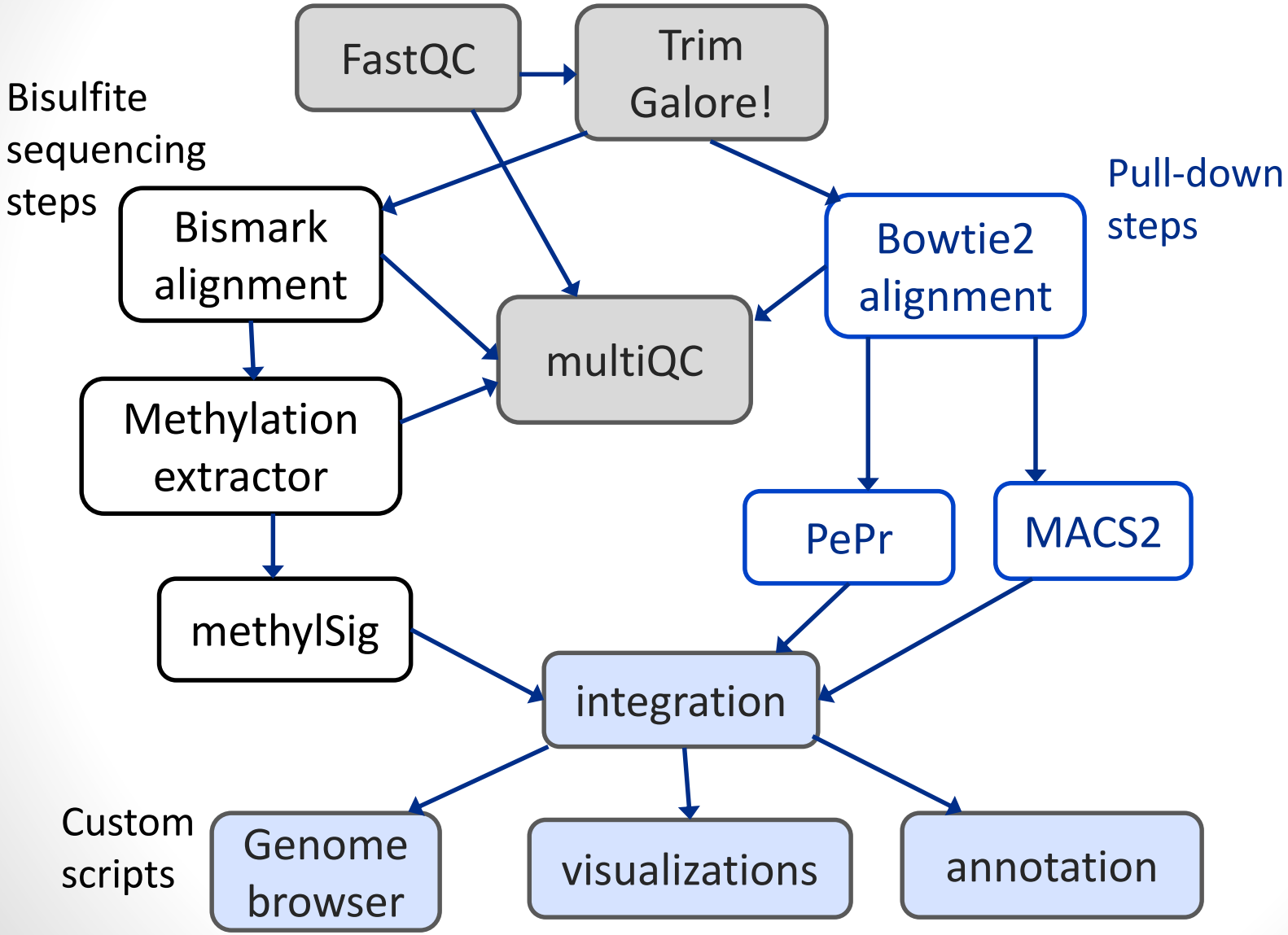
Integration on a per-sample basis.

and or

Comparison-wise

Integration on a per-group basis.

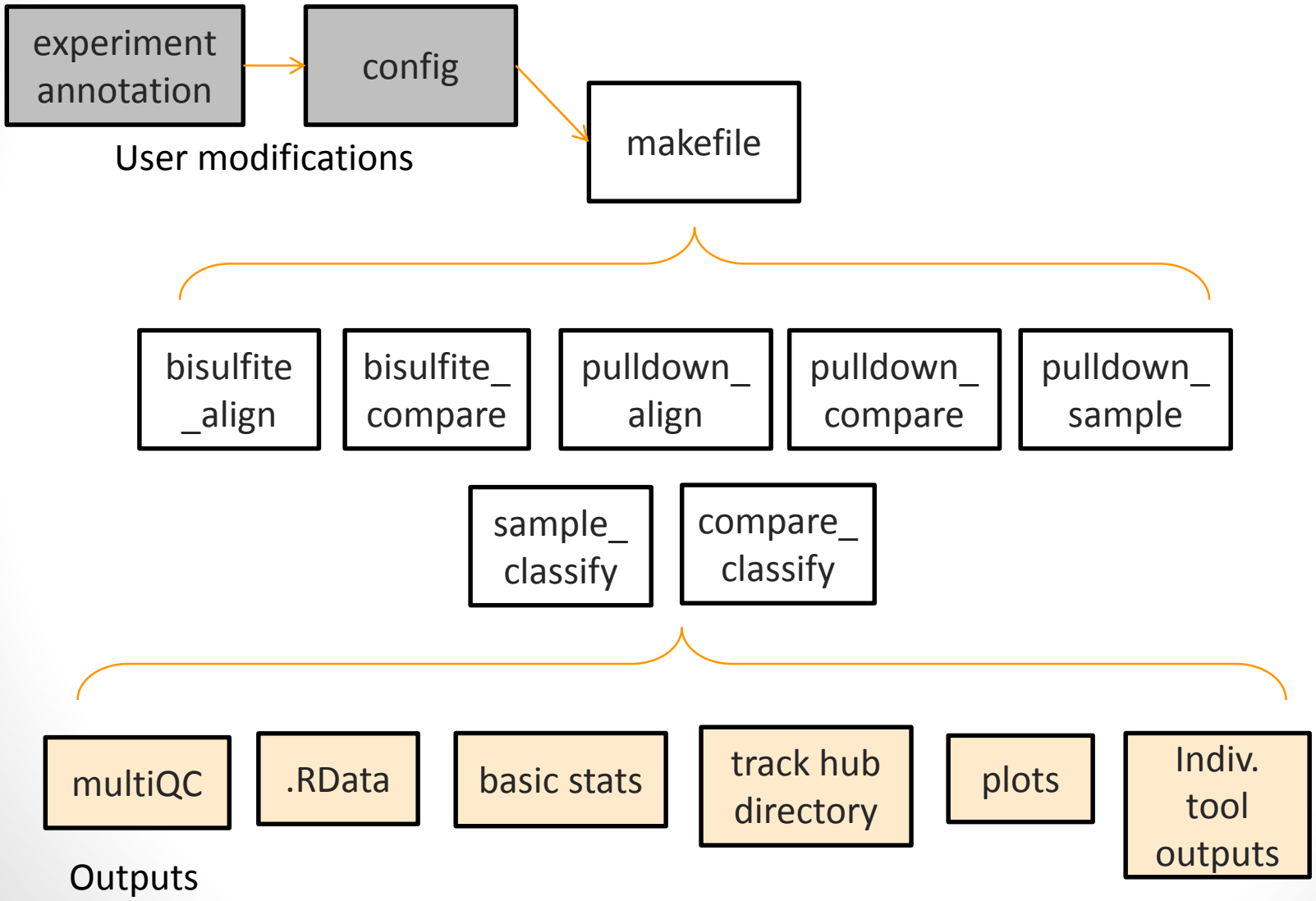
Typical hybrid workflow (without mint)



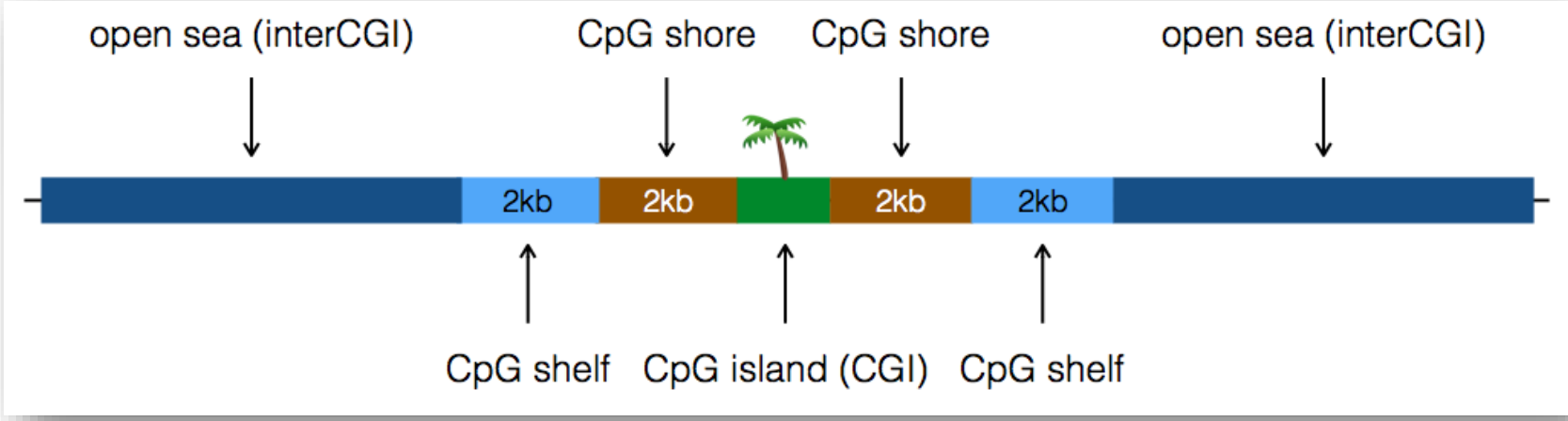
New additions in mint

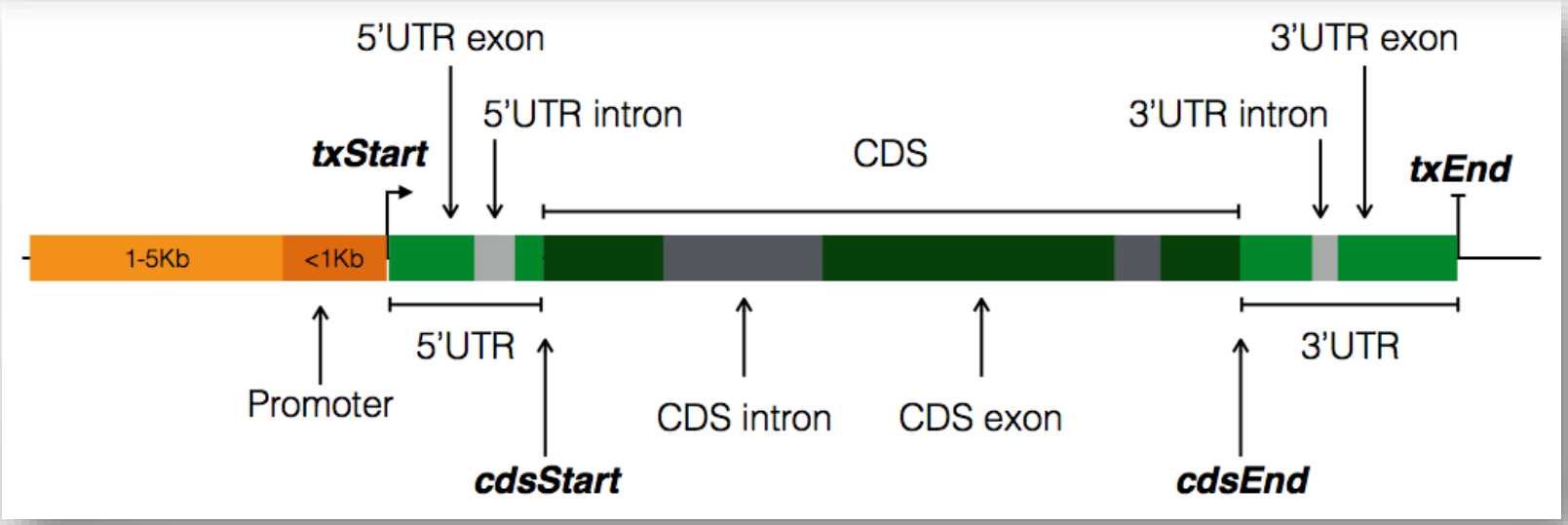
- Made use of many current tools
- New additions include:
  - Common SNP filter (C→T, G→A), human only
  - Significantly reduced run-time for methylSig and PePr
  - Added alternative experimental designs for methylSig
  - Classification of CpG sites or regions by 5mC/5hmC status
  - Automatic genome browser track hub creation
  - annotatr – to annotate and visualize results

mint pipeline

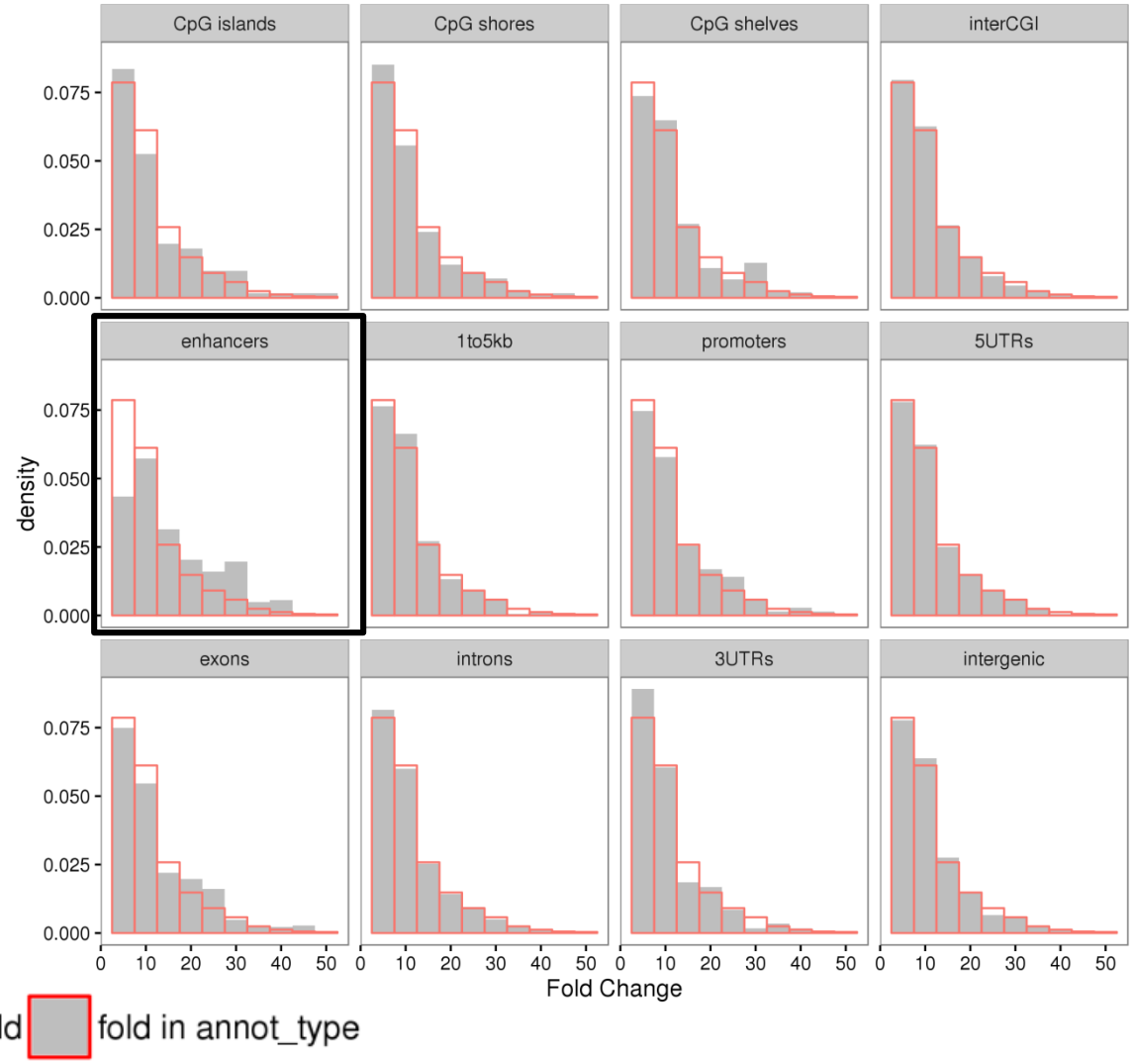


annotatr: simple, fast & flexible annotation of genomic regions





Example: high fold change DhMRs enriched in enhancers

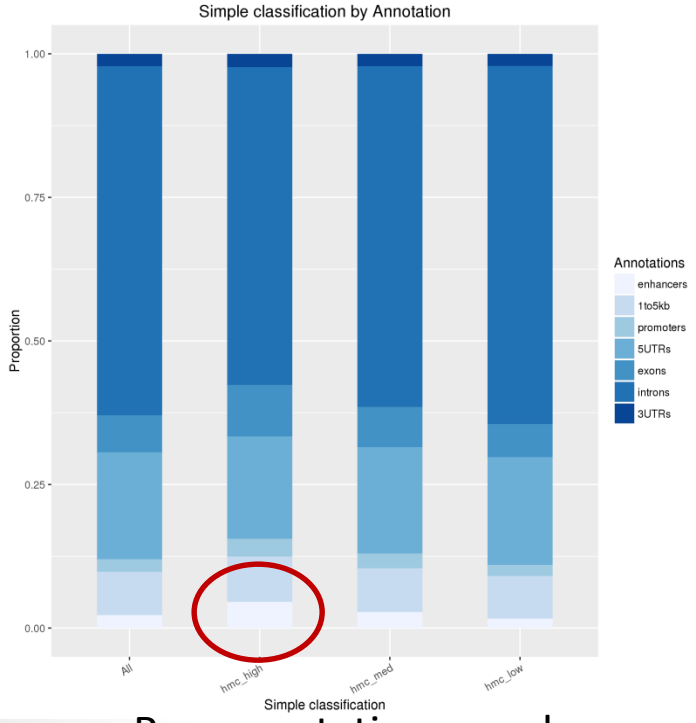


applications

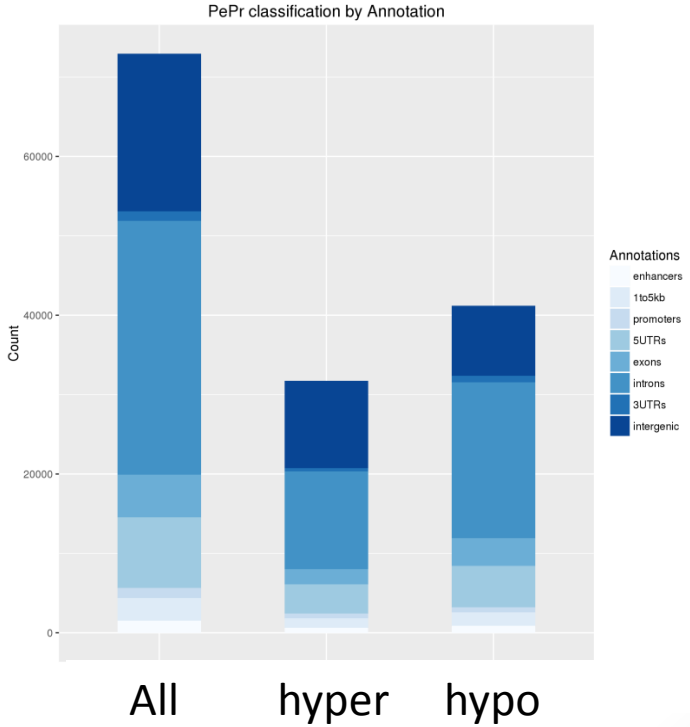
1. Head and neck squamous cell carcinomas (Sartor lab)
   - WGBS + hmeDIP-seq
2. Mouse model age/sex, expsoures (Pb, pthalate) (Dolinoy lab)
   - ERRBS + hmeDIP-seq
3. Leukemia (AML and MDS) (Figueroa lab)
   - ERRBS + hmeDIP-seq

Head and neck cancers

- 18 HPV-positive; 18 HPV-negative
- WGBS & hmeDIP-seq
- Used mint to analyze hmeDIP data
- HPV(+) samples tend to be hypo-hydroxymethylated; high 5hmC was enriched in enhancers

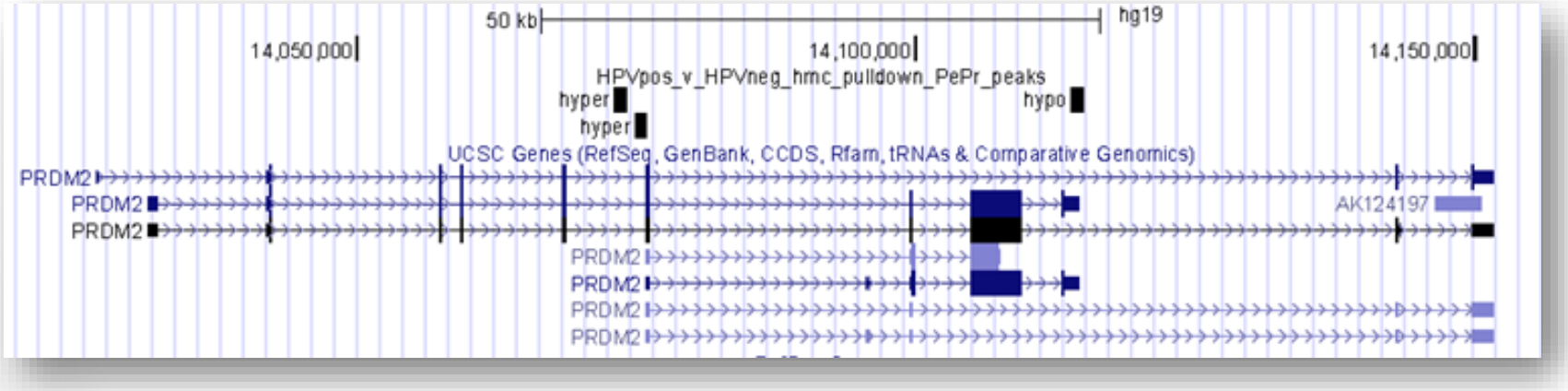


Representative sample



Example: PRDM2 DhMR

- PRDM2 is a tumor suppressor that binds to Rb; Rb is disrupted by the HPV E7 oncogene
- Multiple DhMRs near exon-intron boundaries or alternative TSS or TES.



Galaxy tools for mint

MINT TOOLS

Bedtools

Mint-Pulldown Sample

Mint-Preprocessing

Mint-Classification

Mint-Pulldown Align

Mint-Pulldown Compare

Mint-Utilities

Samtools

Mint-Bisulfite Align

Mint-Bisulfite Compare

Mint-Classification

- Classify Simple R script Convert Pepr outpur for for Ucsc genome browser
- Annotatr classification combined all tools for mint
- Classify to annotatr Awk coomand to convirt claasify to annotatr

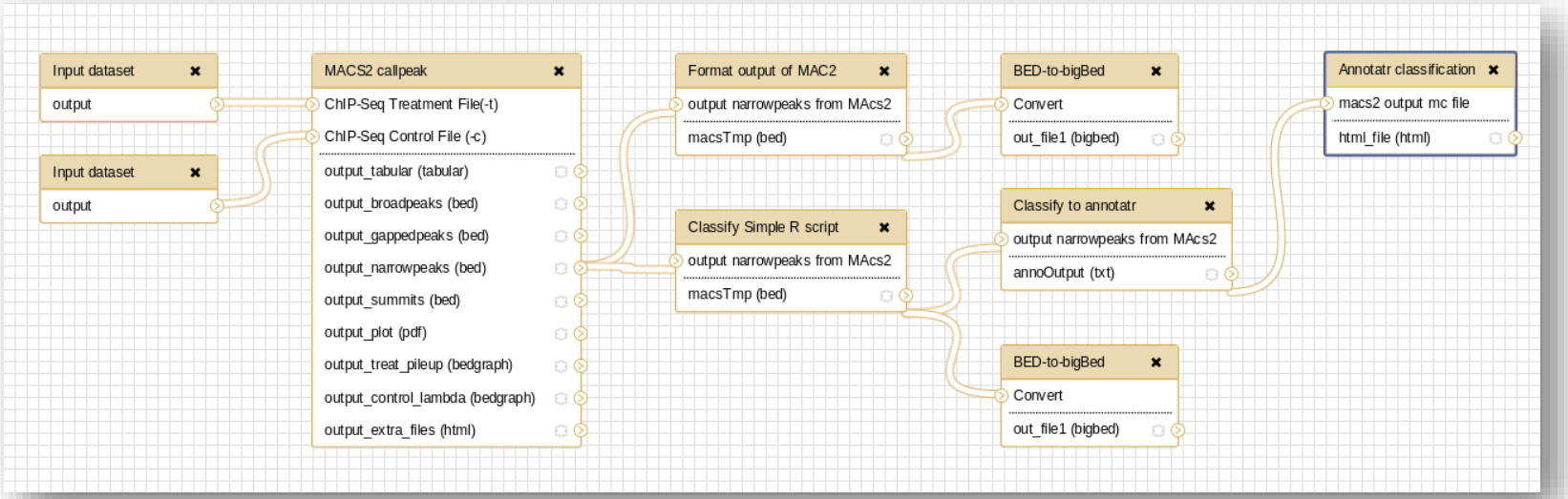
Mint-Pulldown Compare

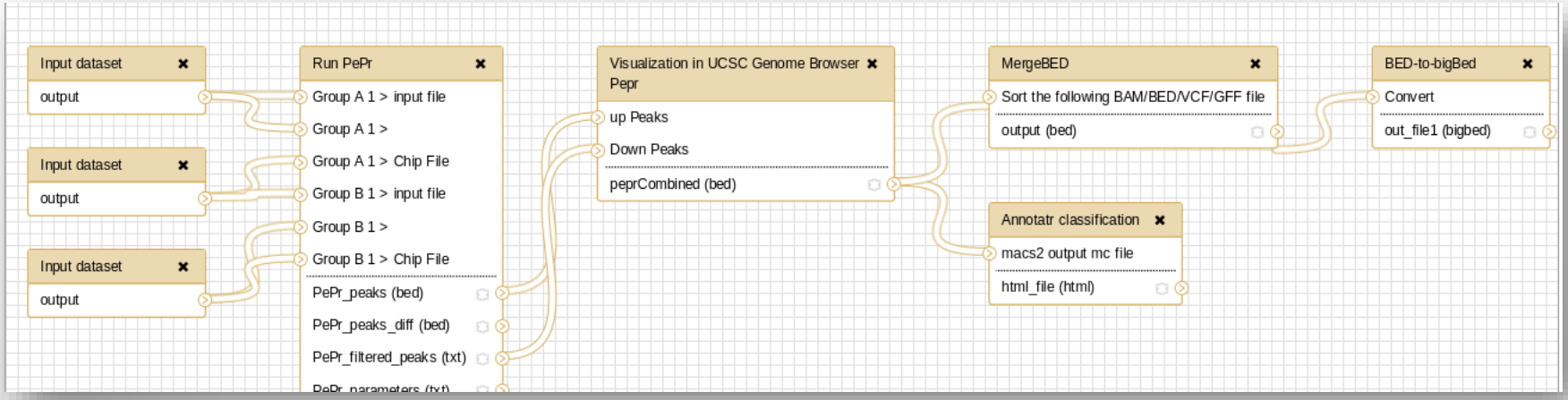
- Run PePr ChIP-Seq Peak-calling and Prioritization pipeline (PePr)
- Visualization in UCSC Genome Browser Pepr Convert Pepr outpur for for Ucsc genome browser

Mint-Bisulfite Compare

- MethylSig for each sequence in a file
- MethylSig To Annotatr combined all tools for mint
- Annotatr Bisulfite Align Annotatr bis align
- From MethylSig To Bedgraph-to-bigWig combined all tools for mint

Galaxy workflows





To do: Integrate with RNAseq data

- Incorporate statistical tests that integrate 5mC and/or 5hmC data with RNAseq or microarray data.

1. Link epigenetic changes to target genes and report genes mostly likely differentially regulated by 5mC &/or 5hmC. Consider both gene level and isoform level expression
2. Test for correlation across samples between 5mC or 5hmC and expression levels. Options for summarizing scores for a genomic feature (eg, CpG islands, enhancers) and assigning the genomic features to genes.
3. Prioritize DMCs/DMRs in terms of causing a functional change in gene expression, based on the significance levels and distance between the site and the gene.

- Incorporate association tests with regions of CNVs, CpG islands, TF binding sites, cancer related sets, etc.

Thank you!

Yanxiao Zhang **Raymond Cavalcante** **Snehal Patil**

Laura Rozek, Co-PI

Maria Figueroa, Co-I

Terry Weymouth

Yongseok Park, Post-doc (now at U. Pitt)

genome.gov
National Human Genome Research Institute
National Institutes of Health

NATIONAL CANCER INSTITUTE