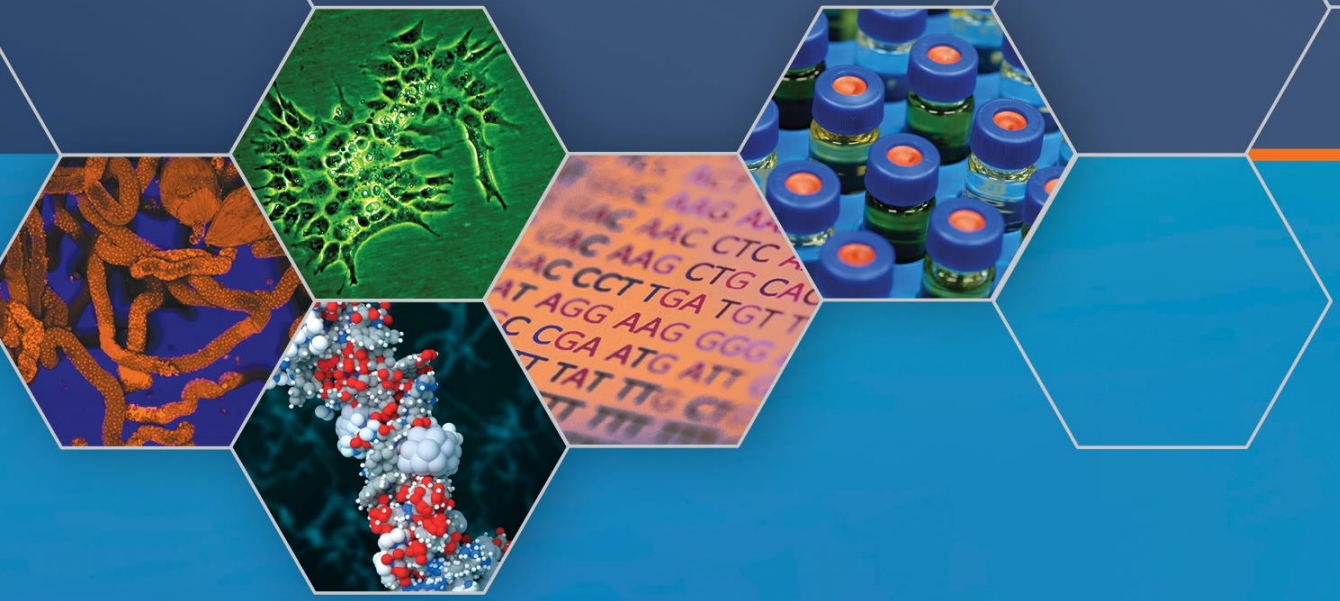
ICBR

Interdisciplinary Center
for Biotechnology Research



UF | UNIVERSITY of FLORIDA

Bioinformatics 101 – Lecture 6

Other NGS applications: SNP calling , ChIP-Seq, methylation analysis

Alberto Riva (ariva@ufl.edu), J. Lucas Boatwright (jlboat@ufl.edu)
ICBR Bioinformatics Core

UF | UNIVERSITY of FLORIDA

Applications: SNP Discovery

- Goal: to identify different forms of variation between individuals at the genomic level.
- NGS allows detecting SNPs, short insertions and deletions, large-scale rearrangements.
- Example application: cancer profiling (very high mutation rates even within populations of cells from same tissue).

Applications: SNP Discovery

Position		1	2	3	4	5	6	7	8	9	10	11	12	
Reference	…	**C**	G	G	**A**	C	G	**A**	C	A	T	C	G	…
	…	C	G	G	**C**	C	G	**A**	C	A	T	C	G	…
	…	C	G	G	**A**	C	G	**G**	C	A	T	C	G	…
Aligned reads	…	C	G	G	**C**	C	G	**A**	C	A	T	C	G	…
	…	C	G	G	**A**	C	G	**A**	C	A	T	C	G	…
	…	C	G	G	**A**	C	G	**A**	C	A	T	C	G	…

Allelic frequencies: Position 4, A=60%, C=40%
Position 7, A=80%, G=20%

- Accurate frequency estimation requires high coverage (at least 20X) and enough observations of each allele.

UF ICBR BIOINFORMATICS

UFHealth CANCER CENTER

Applications: SNP Discovery

- Tools: GATK, freebayes, samtools, bcftools.

- Simulation:
  1. Generate synthetic reference sequence with SNPs;
  2. Simulate reads from reference sequence – at SNP positions, bases are chosen at random from alleles;
  3. Align reads to reference, call SNPs from alignment;
  4. Compare detected SNPs with “true” list.

Applications: SNP Discovery

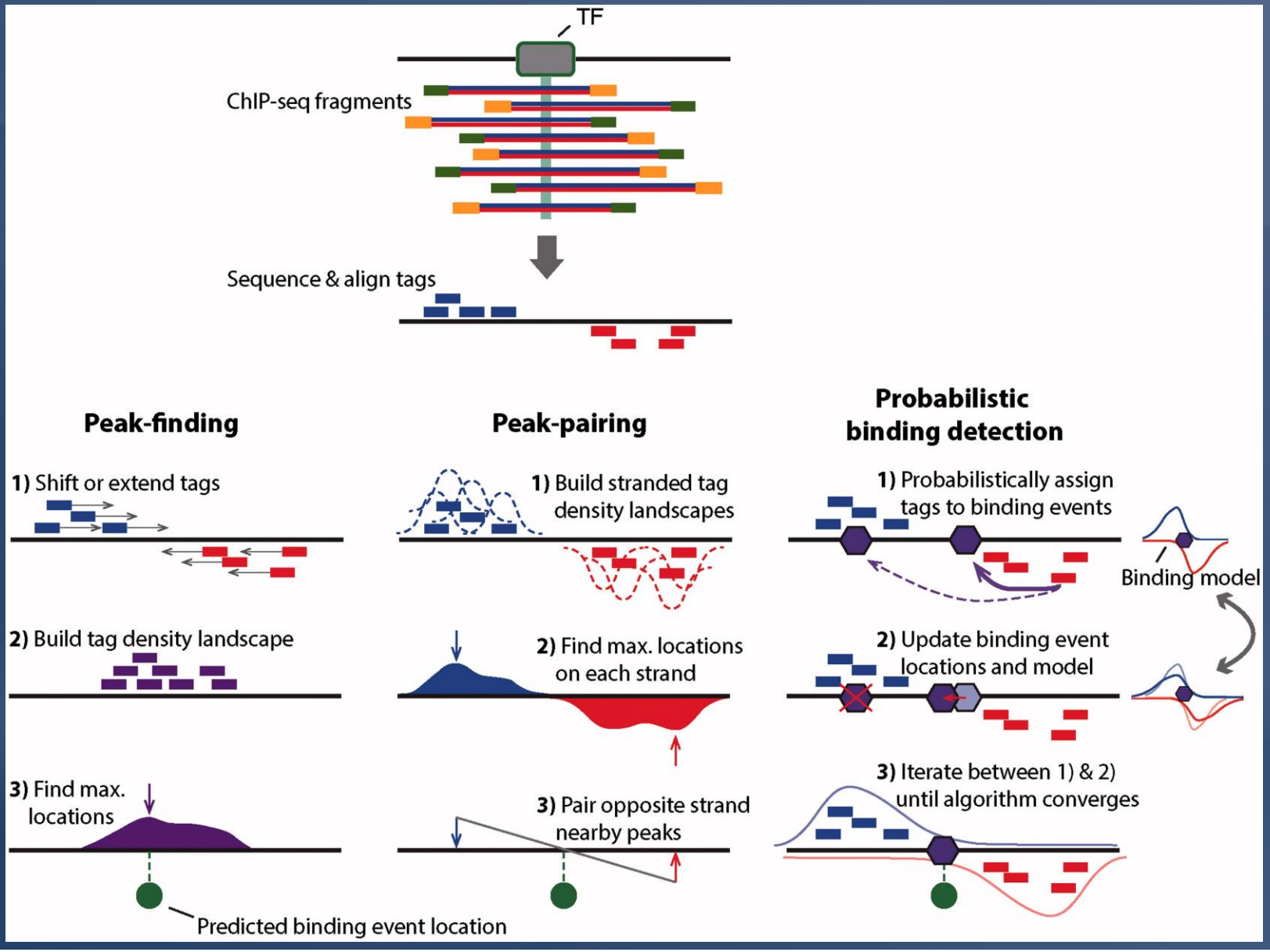
Output is typically a **VCF** file:

- Each row represents a variant;
- The first nine columns provide information about the variant: genomic position, alleles, quality, annotations;
- One or more columns containing genotypes in different samples.

Downstream analysis: **functional annotation** of variants, to identify potentially deleterious ones.
Annotation tools: annovar, snpEff.

Applications: chromatin analysis

- ChIP-Seq analyzes protein interactions with DNA. Bound DNA fragments are precipitated and sequenced.
- Peaks in the short-read alignment indicate the location of binding sites.
- ATAC-Seq is a similar technique that identifies open chromatin regions.
- Analysis tools: MACS, HOMER.



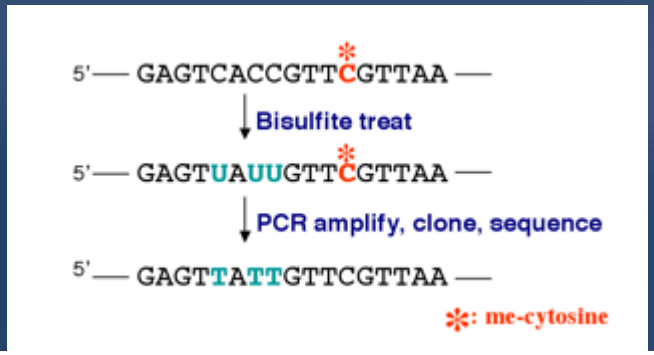
UF | ICBR BIOINFORMATICS

UFHealth CANCER CENTER

Applications: methylation analysis

- DNA methylation can be studied at the genome-wide level using NGS methods.

- Bisulfite Sequencing: sodium bisulfite treatment converts Cs to Ts, except when they are methylated. Treated DNA is sequenced and compared with the reference sequence to detect conversion.



- Methylation rate = fraction of non-converted Cs (in contexts where C can be methylated, e.g. CG).

UF | ICBR BIOINFORMATICS

UFHealth CANCER CENTER

Applications: methylation analysis

Methylation is not totally present or absent because of:

- Cell heterogeneity;
- Incomplete bisulfite reaction.

UF | ICBR BIOINFORMATICS

UFHealth CANCER CENTER

Applications: methylation analysis

Downstream analysis:

- Methylation in promoters and other gene regions;
- Differential methylation analysis;
- Identification of differentially methylated regions;
- Classification of DMRs in genomic contexts (e.g. promoters, enhancers, intergenic).

Analysis tools: Bismark, MOABS, CSCALL.

Conclusion

- All tools and pipelines we described are available on HiPerGator (in their own modules, e.g. dibig_tools);
- We will make the Nextflow pipelines freely available;
- For more information or help with projects please contact:

✓ICBR Bioinformatics core – ICBR-Bioinformatics@ad.ufl.edu
✓UFHCC BQS Shared resource - jihyun.lee@ufl.edu

UF | ICBR BIOINFORMATICS

UFHealth CANCER CENTER