BMEG 3105 Data Analytics for Personalized Genomics and Precision Medicine

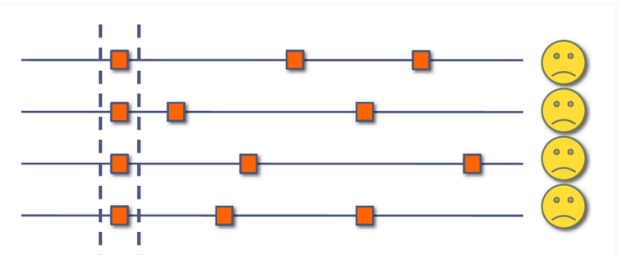
Scribing assignment

31/10/2025 Lec-16

Jettapat Vanasrisawasd 1155199844

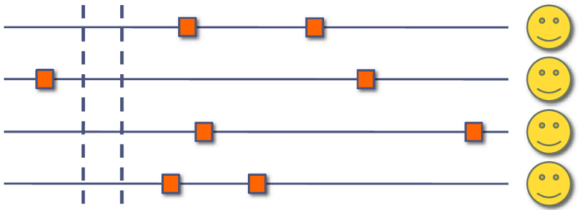
Genome Wide Association Studies (GWAS)

> Determine whether specific variants in many individuals = disease/trait



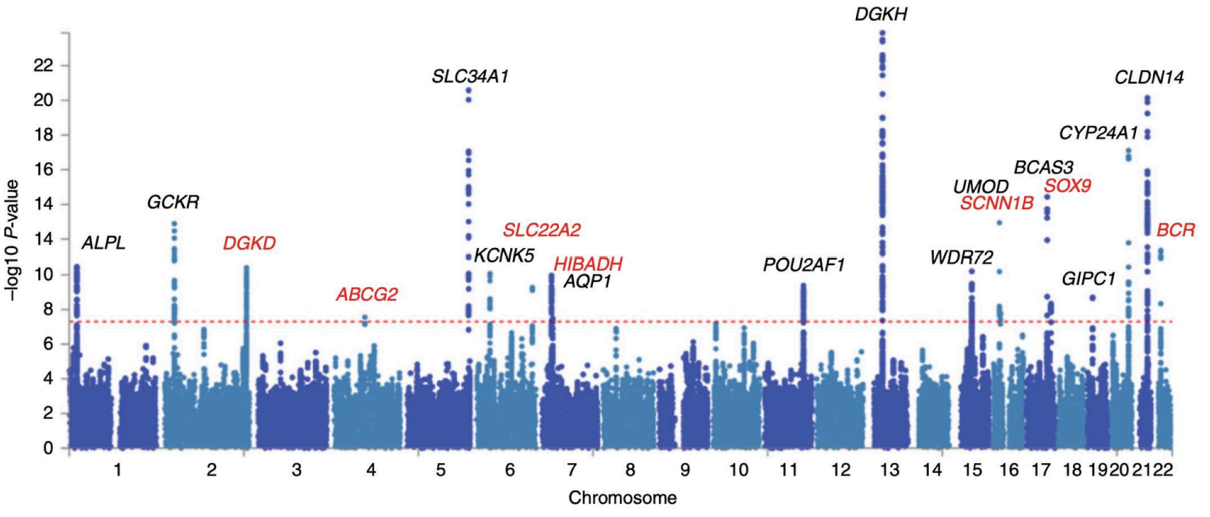
> in the above image, the individuals **are affected** and they have a common variant.

Compared to the **unaffected** below where the variant is absent.



That is the **ideal case.**

In reality,



> 1.5 mil Singlenucleotide Polymorphism (SNPs)

Bonferroni Correction = p-value / # of tests

Example : 1 mil SNPs

$$\text{Adjusted p-value} = \frac{0.05}{1{,}000{,}000}$$

$$\text{Adjusted p-value} = 5 * 10^{-8}$$

RNA data seq

Step process:

1. Exploration : Finding a genome, taking the whole seq (RNA)
2. Choose informative genes : Conduct tests to get expressions (Bonferroni, FDR)
3. Choose algorithm : Algorithm training and overfitting (Cross validation)
4. Validation : Training and validation sets
5. Implementation : Nice, FDA, RT-PCR, NanoStrinq

Gene fusion

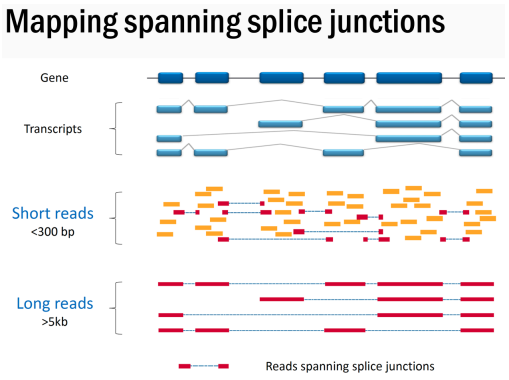
Gene formed by two distinct wild-type genes

In cancer: produced by somatic genome rearrangements

For chromosomal translocation, exon cans transfer between genes which causes fusion in the chromosome.

For interstitial deletion, introns might get deleted so two exons can cause gene fusion.

- Breakpoints are at **INTRONS**
- **Need** whole genome sequencing, exome sequencing not sufficient

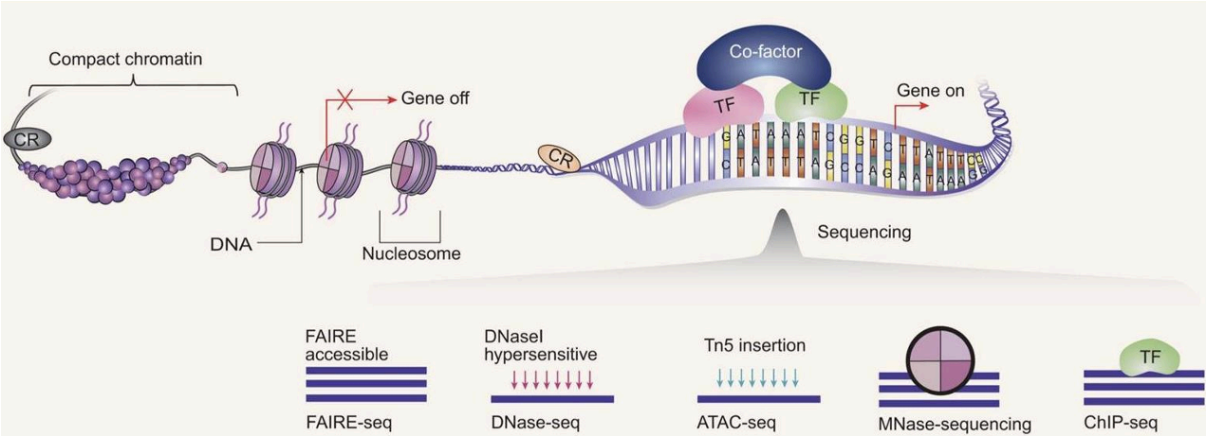


Detected via RNA-seq where it requires less sequencing than WGS with long reads.

Epigenome

Epigenomic variant sequencing: isolate DNA methylation, chromatin modding, DNase I hypersensitive sites, and transcription factors to monitor sequencing protocols.

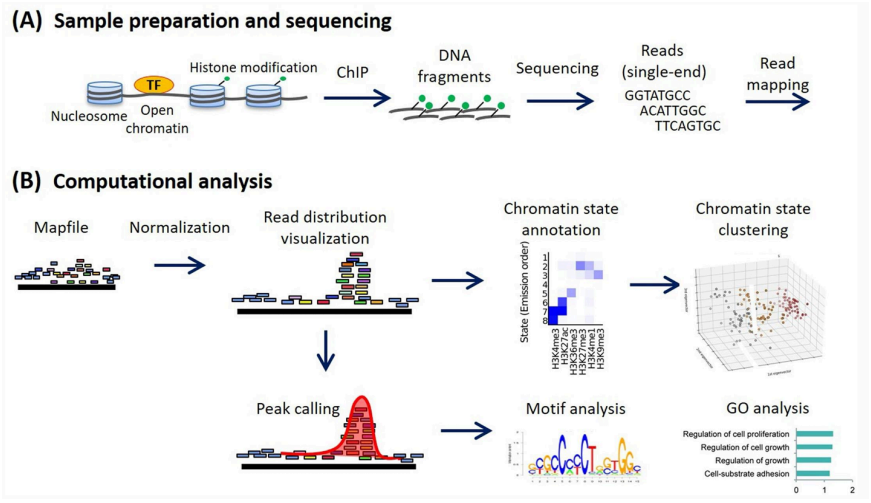
Sequencing protocols



Data analytics pipeline

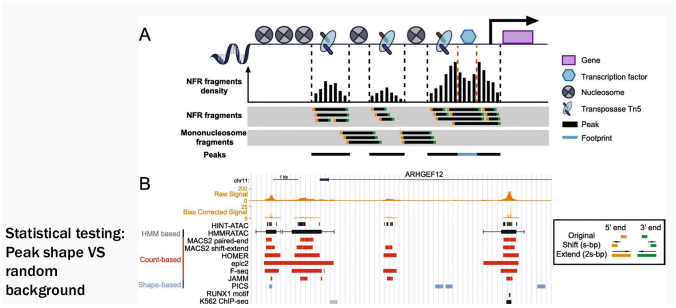
a) Sample prep and seq

b) Computational

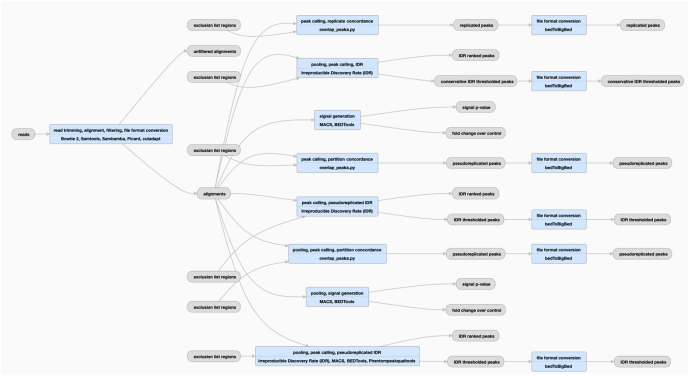


Peak calling

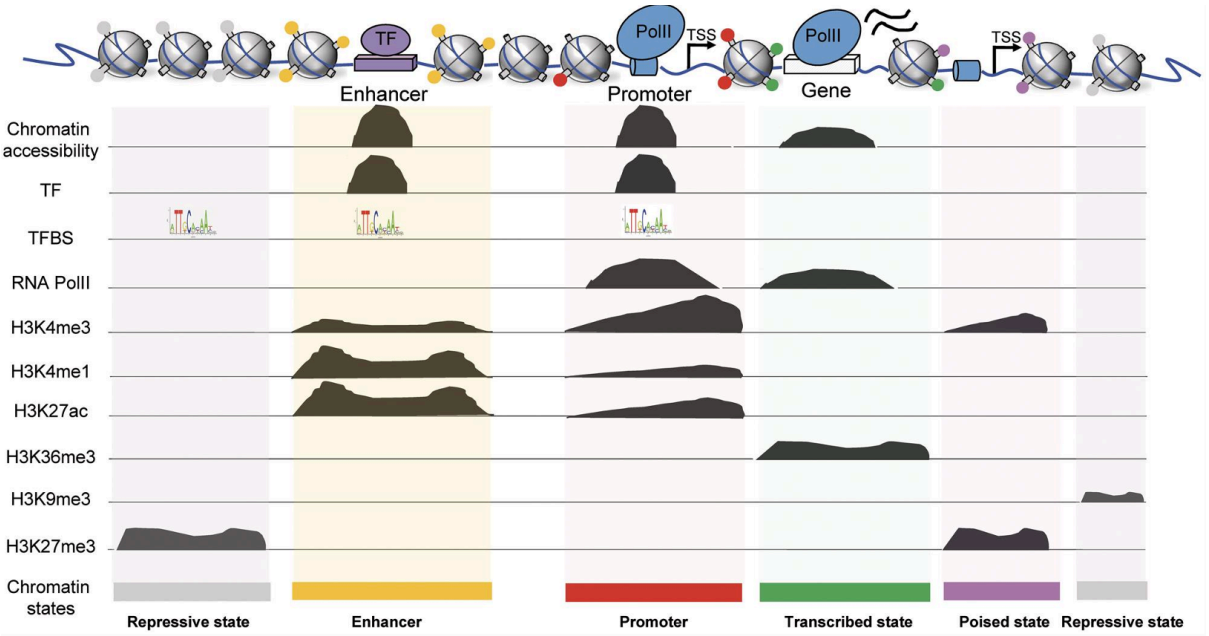
Compare peak shape with random background to show correlation.



The entire detailed pipeline is very tedious and complex.



Histone marks and chromatin accessibility:



With different protocols, different histone and chromatin accessibilities can be **detected.**