Recap: **Gene enrichment analysis**

A **biological pathway** is a series of interactions among molecules in a cell that leads to a certain product or a change in a cell. Such a pathway can trigger the assembly of new molecules, such as a fat or protein. Pathways can also turn genes on and off or spur a cell to move.

->KEGG pathway database

->Each pathway contains a set of genes

Recap: **Fisher's exact test**

->a statistical significance test used in the analysis of contingency tables.

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{a+b+c+d}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!(a+b+c+d)!}$$

Recap: **Study of cancer** (a genomic disease)

->use genomics/multi-omics methods to study it e.g. Genome/ Epigenome/ Transcriptome/Proteome/Metabolome

✧   CIGAR=Concise Idiosyncratic Gapped Alignment report
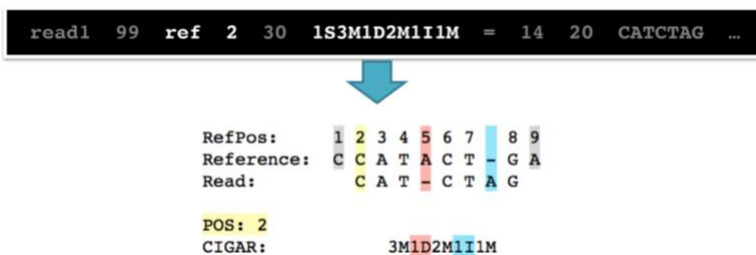
We need to know:

1.The reason that we need to do the steps

2.The ability to read the records in those files

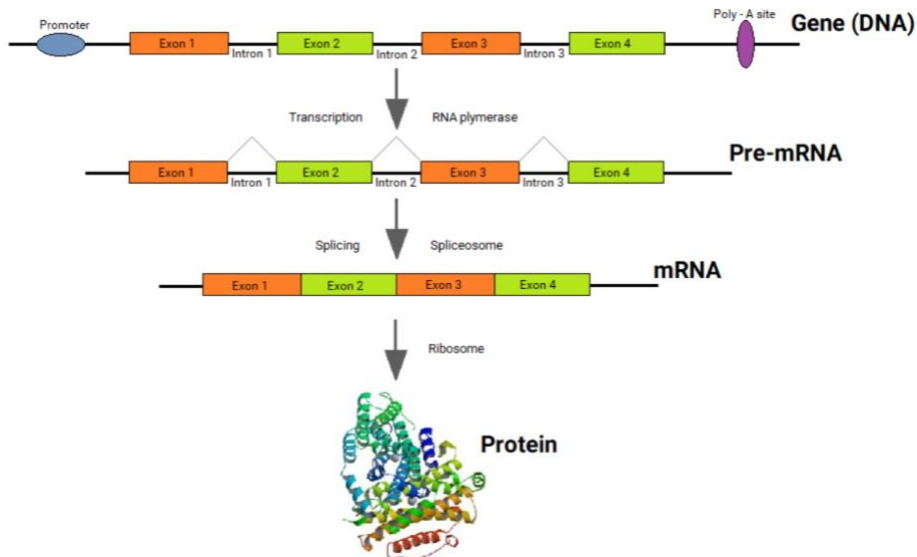3.How different factors affect the quality of the mapping and the variant calling

Bonferroni correction ->Adjusted p-value = p-value/number of tests

e.g. we have 1 million SNPs to test -> Adjusted p-value = $5 * 10$^-8
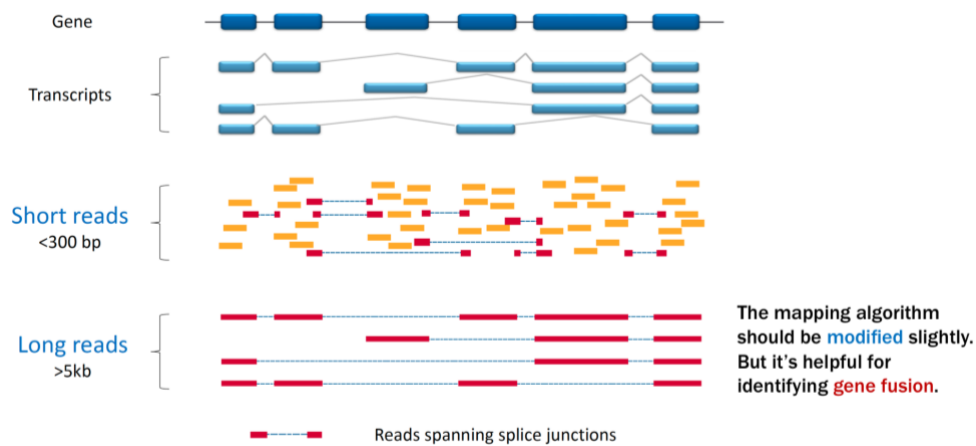
1. RNA-seq data analysis
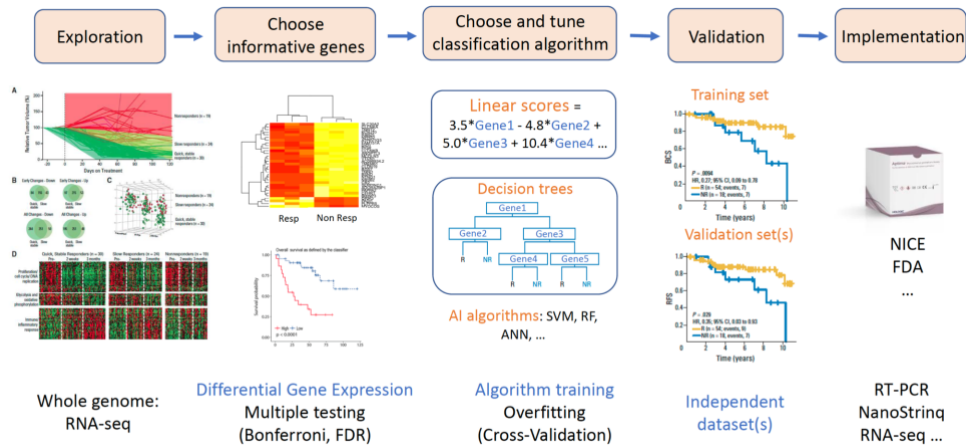
Transcription, splicing and translation of a eukaryotic gene:
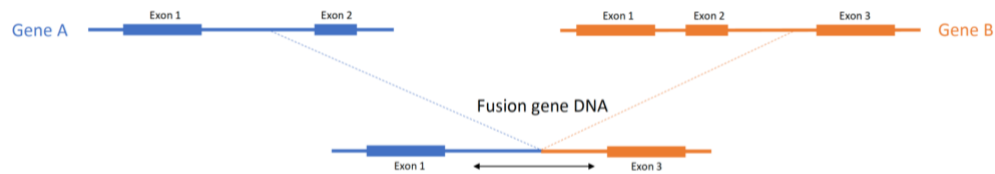

Mapping spanning splice junctions:

1.2 Gene fusion (a specific kind of structural variant related to cancer)

-The first fusion gene was described in cancer cells in the early 1980s

-Novel gene formed by fusion of two distinct wild type genes

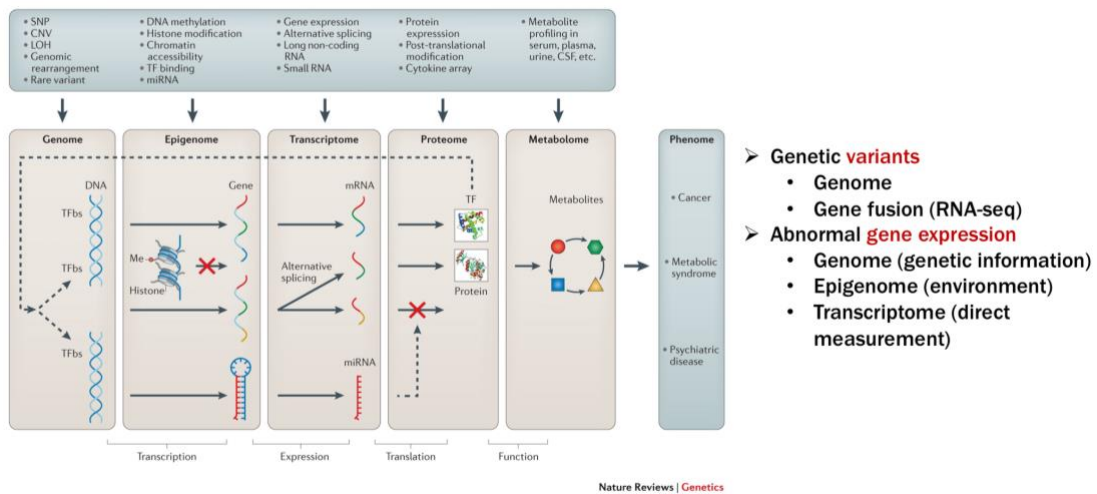-In cancer: produced by somatic genome rearrangements

Break-points are in introns
We need whole genome sequencing
Whole exome sequencing is not enough

Fusion gene RNA

Detecting fusion in RNA-seq requires much less sequencing
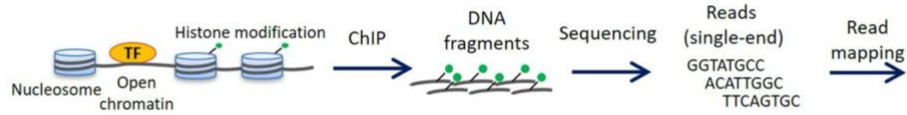than WGS, especially with long reads

> Genetic variants
> • Genome
> • Gene fusion (RNA-seq)
> Abnormal gene expression
> • Genome (genetic information)
> • Epigenome (environment)
> • Transcriptome (direct measurement)

Nature Reviews | Genetics

2.Epigenomics (Gene expression regulation: structure and environment)

The overall data analytics pipeline for epigenomics
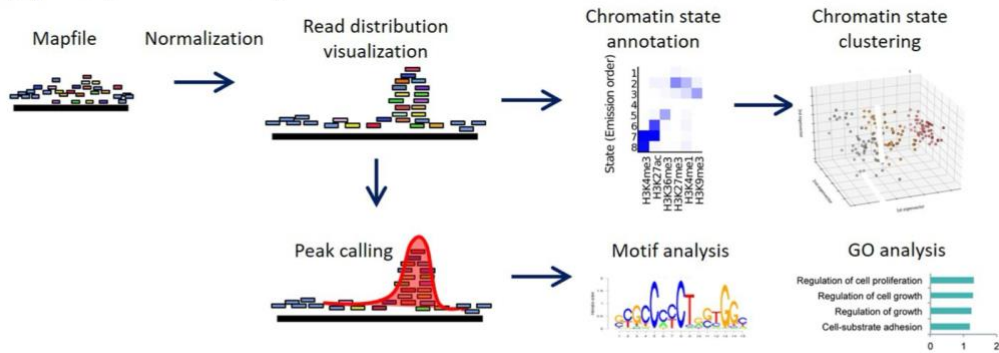
2.1 Peak calling
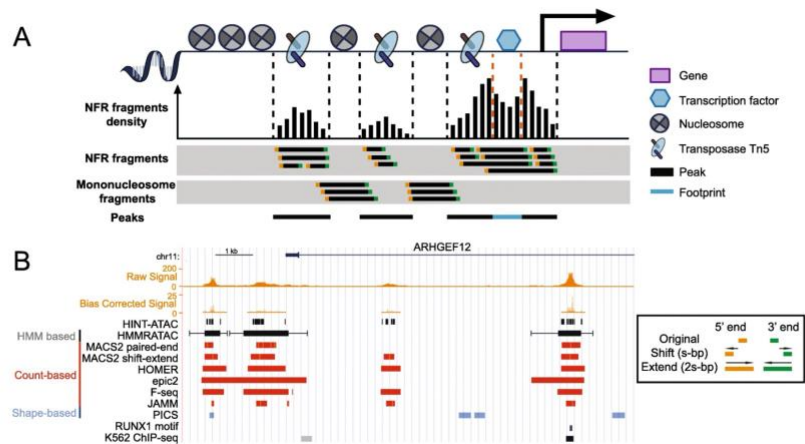

Statistical testing:
Peak shape VS
random
background

-Output-BED (Browser Extensible Data) file