



CEITEC

Central European Institute of Technology  
BRNO | CZECH REPUBLIC



**Bi7420: Moderní metody pro  
analýzu genomu**

# NGS data analysis introduction

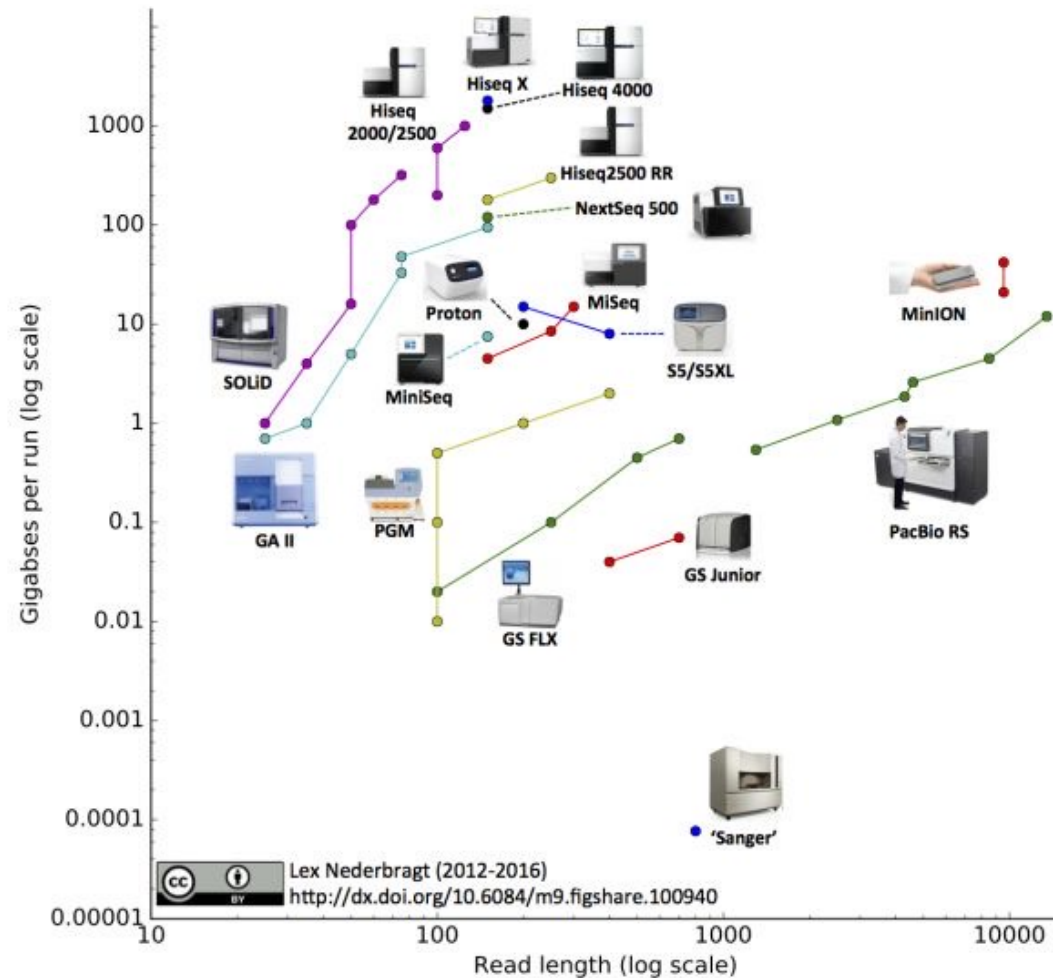
Vojta Bystry  
[vojtech.bystry@ceitec.muni.cz](mailto:vojtech.bystry@ceitec.muni.cz)

# Plan for Bi7420

- Next generation sequencing methods overview
  - Focus on experiment planning and result interpretation
- 1. Introduction to NGS technology
- 2. miRNA, lncRNA in cancer - Marek Mráz
- 3. Basic QC, DNA resequencing
- 4. DNA resequencing, Chip-seq (CLIP-seq)
- 5. Chip-seq (CLIP-seq)
- 6. RNA-seq
- 7. Single-cell RNA-seq, Spatial transcriptomics

# What is NGS?

- Next generation sequencing
  - New generation sequencing
  - HTP = High throughput
  - Massively parallel sequencing
- Contrast to Sanger sequencing

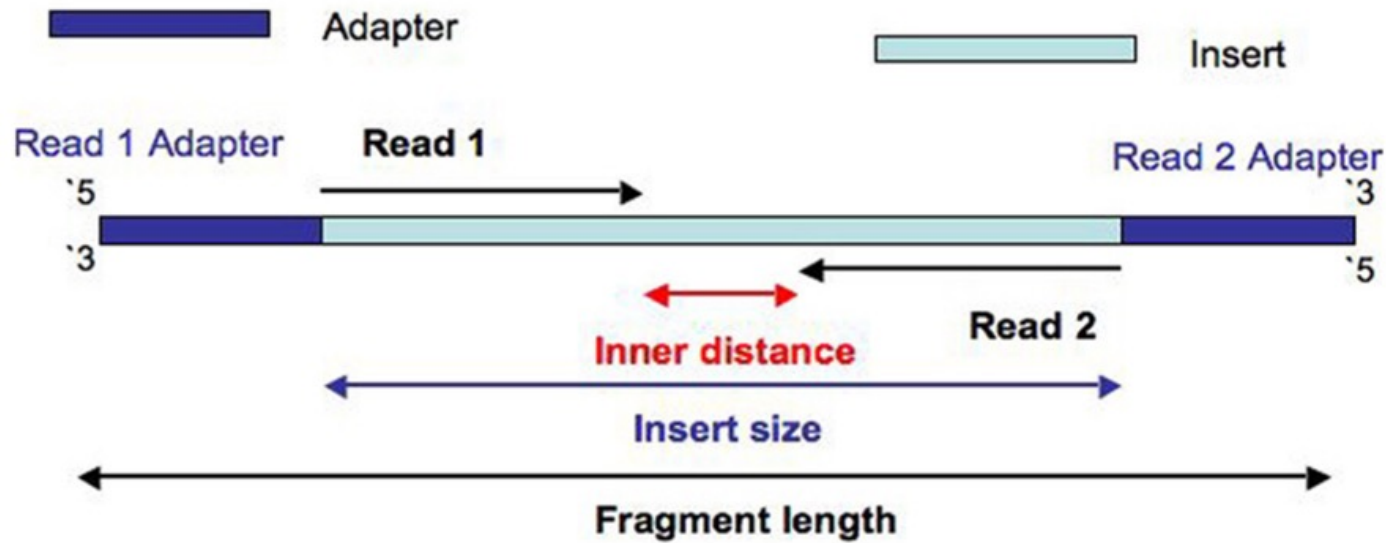


# What is NGS?

- Illumina – sequencing by synthesis – short-read sequencing
- Oxford Nanopore – Nanopore sequencing
- Pacific Bioscience - Single Molecule, Real-Time (SMRT)

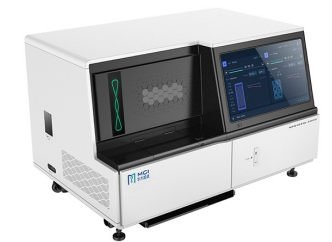
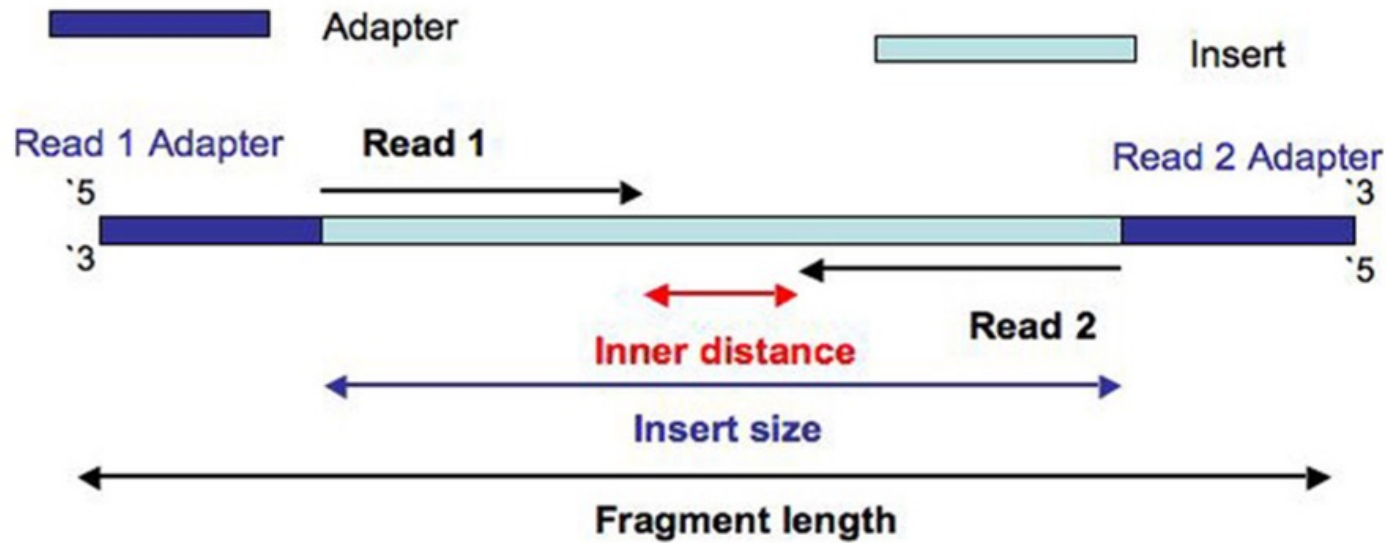
# What is NGS?

- Illumina – sequencing by synthesis – short-read sequencing
- Oxford Nanopore – Nanopore sequencing
- Pacific Bioscience - Single Molecule, Real-Time (SMRT)



# What is NGS?

- Illumina – sequencing by synthesis – short-read sequencing
- Oxford Nanopore – Nanopore sequencing
- Pacific Bioscience - Single Molecule, Real-Time (SMRT)



# Short-read sequencing result

```
>read_no_1
CGGCCTGGAGGCCCTGCAGAACCTGCTGGGCTACAGGTTCCGGCAGCAGGG

>read_no_2
GCAGCTGGAGGCCCATCATGGGCAACCCCCAGGTGAAGGCCACGGCAAGA

>read_no_3
GGGAGACACCCGCACGTGTGGCCCGCATGTATGCTGAGCTCTTCCGCGGAT

>read_no_4
TTTGCCCCGCATCGAGCGGGCTGTGCGGAAATCCTTCTGGCTGTAGGCGA

>read_no_5
CCTGTGGGGCAAGGTGAACCCCGTGGAGATCGGCGCCGAGAGCCTGGCCAG

>read_no_6
GAGGAGGCCAGGATCCACCAGAGGAAGGGCCTGCTGTGGTTCATCCCCGC

>read_no_7
CTGCACACGACTACAACCTGACCTGGTACAGGAACGGCAGCAACATGCCC

>read_no_8
GTGCTGGGCCTGGCCATCAGCCACTTCCCTGCTGGAGCAGTCCCCGACTAC

>read_no_9
AACCTGGGCGAGTACCTGCTGCTGGGCAAGGGCGAGGAGATGACCGGCGGC

>read_no_10
GTTCCCCGACTACAACGAGGGCGAGCTGAGCAGGCTGAGGAGCGCCATCGT

>read_no_11
CTTCAGCAAGTTCGGCGACCTGAGCAGCGTGAGCGCCATCATGGGCAACCC

>read_no_12
ACCAGAGGAAGGGCCTGCTGTGGTTCATCCCCGCCCCCTGGAGGACAGCG

>read_no_13
AAGGGCGAGGAGATGACCGGCGGCAGGAGGAAGGCCAGCCTGCTGGCCGAC
```

- $10^5 - 10^{10}$  reads
- 75 – 300Bp
- Could be pair-end

# NGS experiment workflow



Experimental design



Library preparation



Sequencing



Data analysis

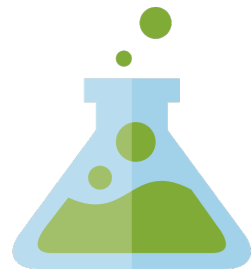


# NGS experiment workflow



Experimental design

Why we sequence



Library preparation

What we sequence



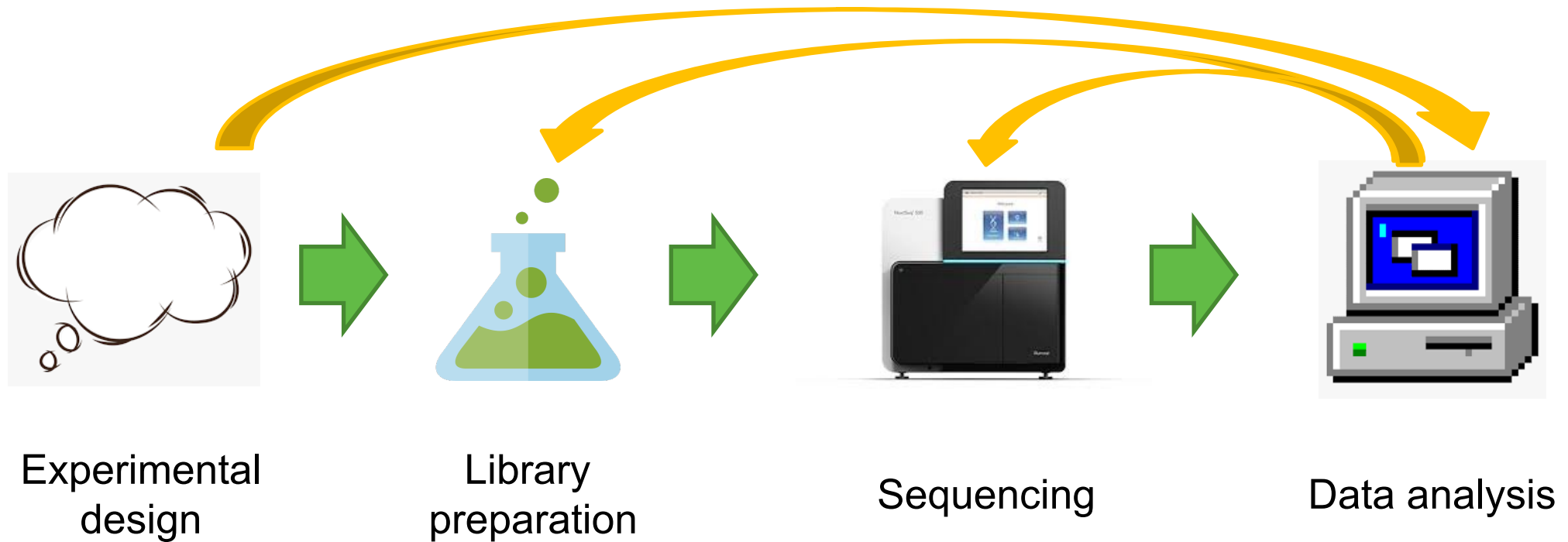
Sequencing

How we sequence



Data analysis

# NGS experiment workflow



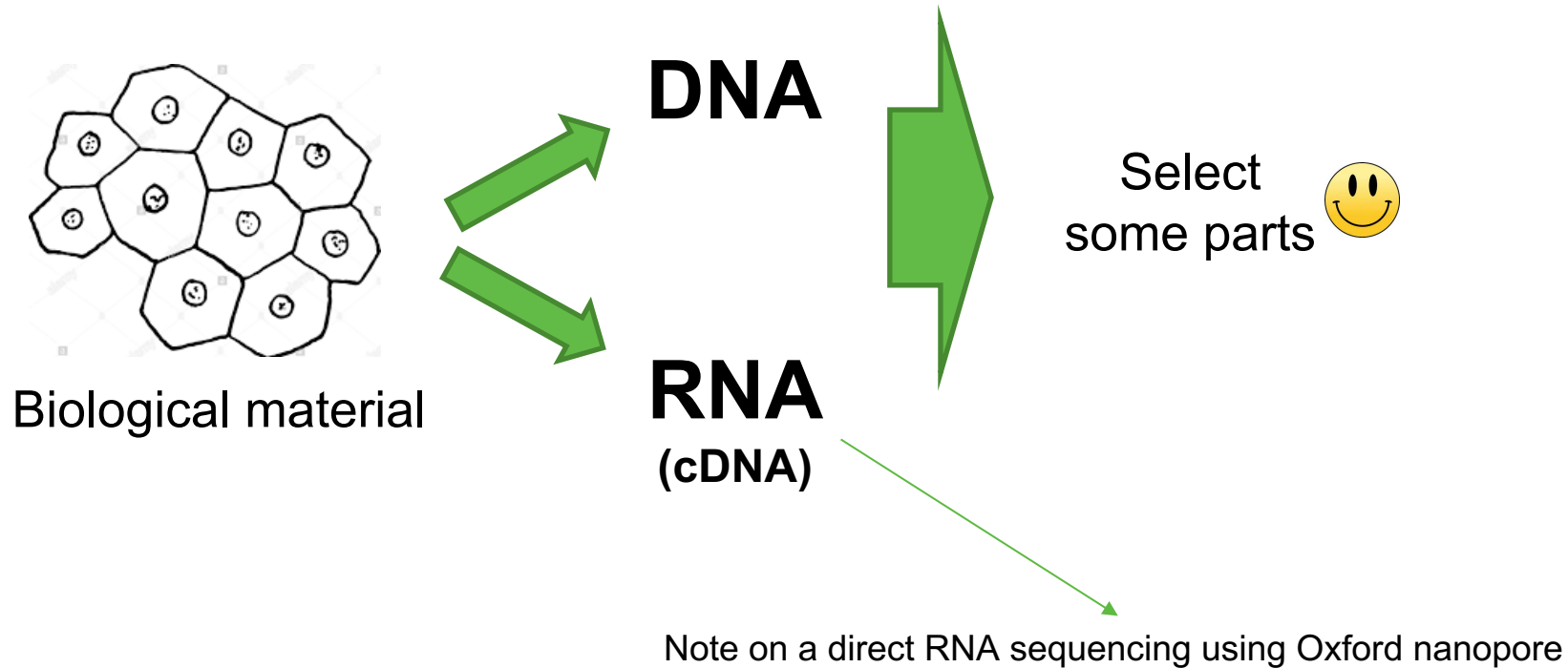
Why we sequence

What we sequence

How we sequence

**Consultation regarding data analysis is highly advisable.**

# NGS library preparation - What we sequence



# For all you seq...

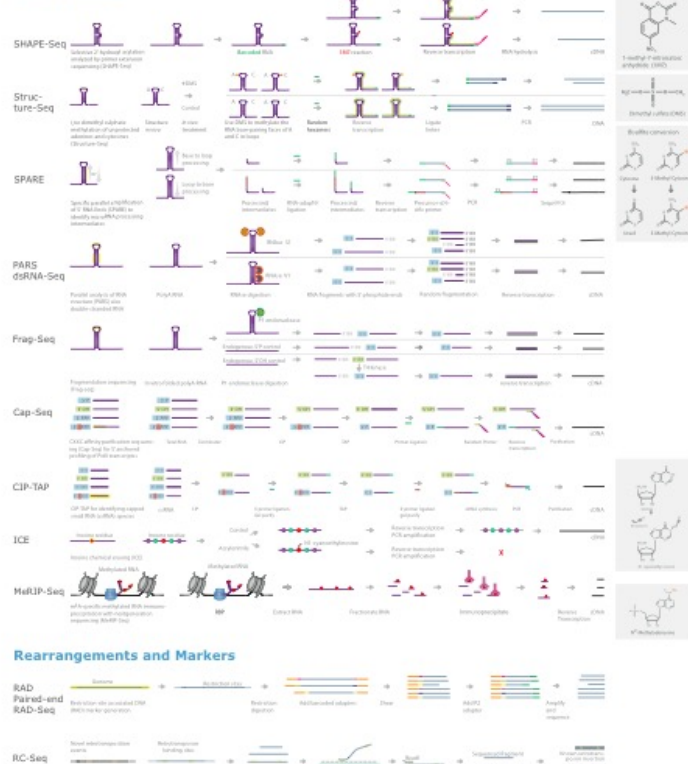
[www.illumina.com/LibraryPrepMethods](http://www.illumina.com/LibraryPrepMethods)



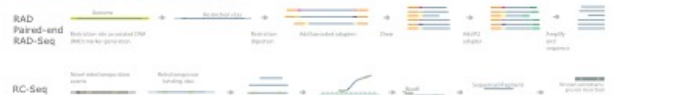
## RNA Transcription



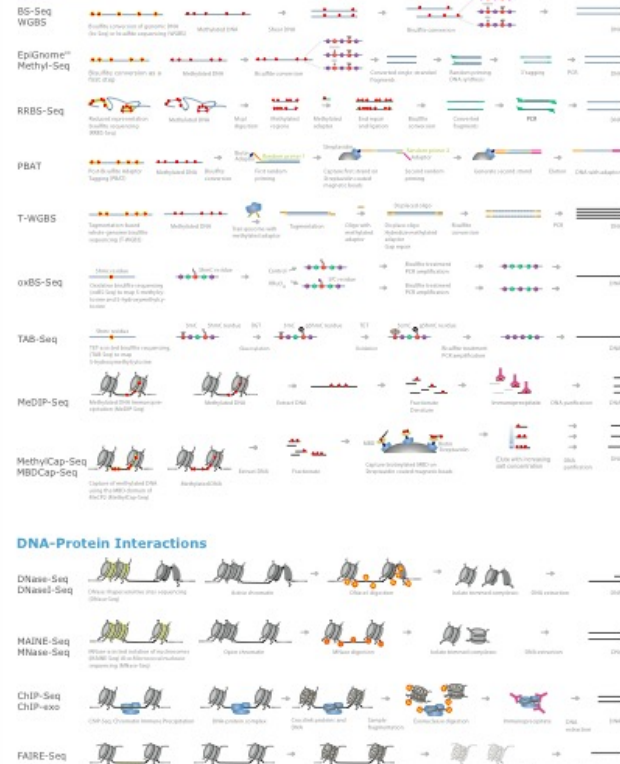
## RNA Structure



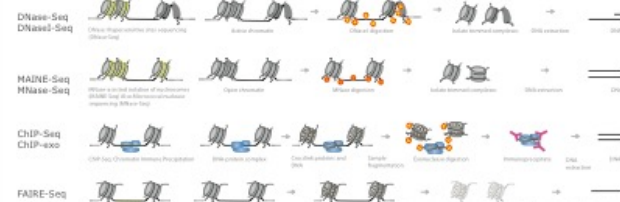
## Rearrangements and Markers



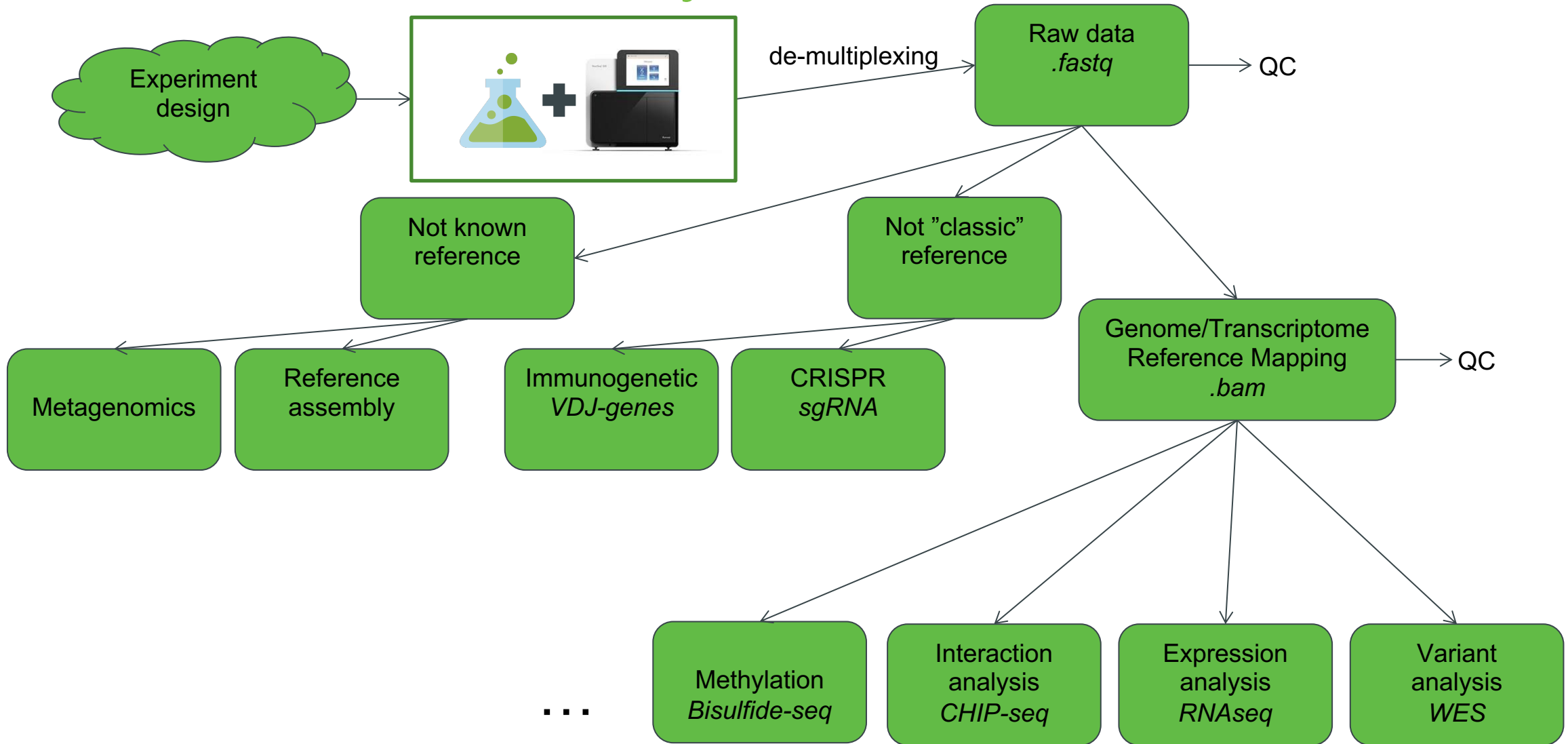
## Methylation



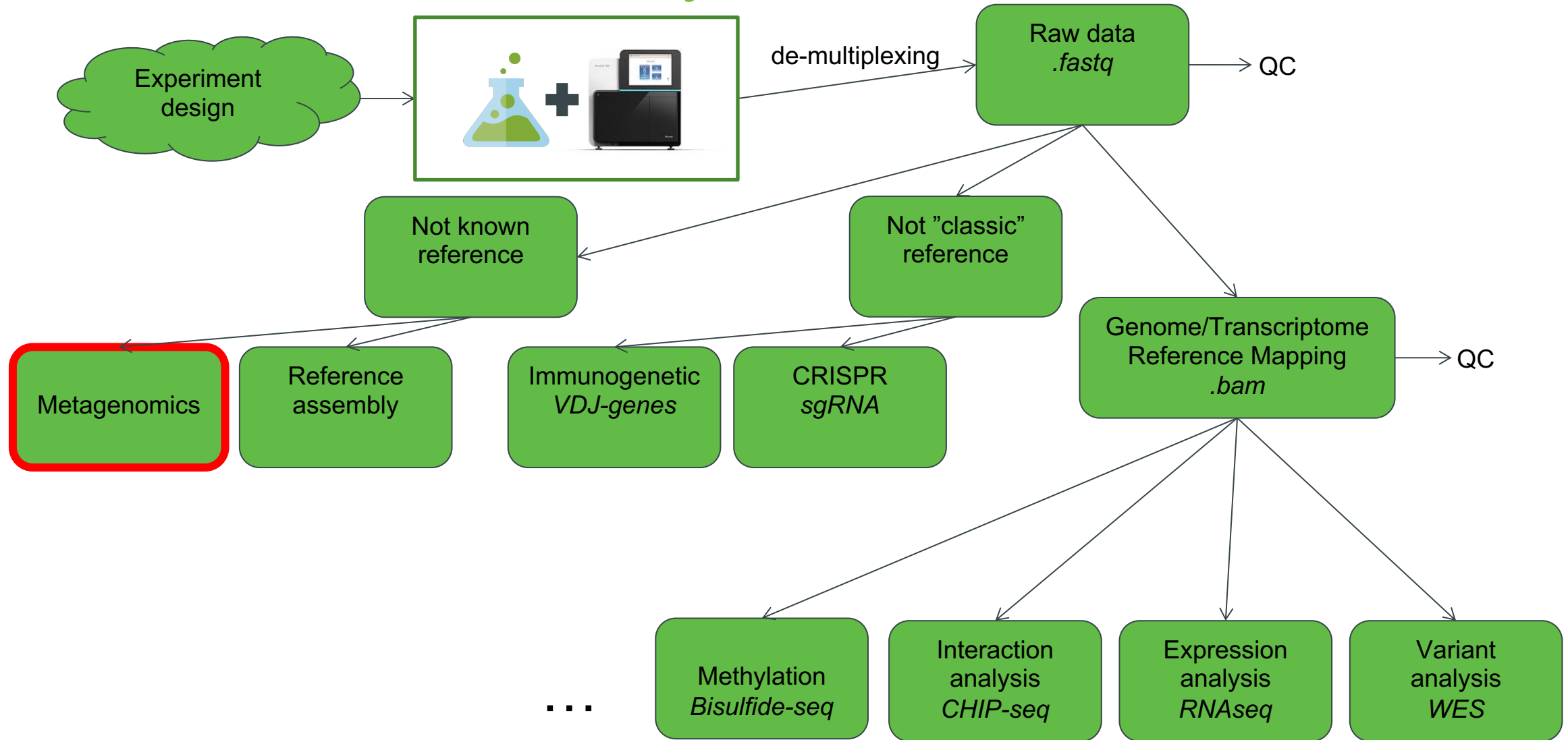
## DNA-Protein Interactions



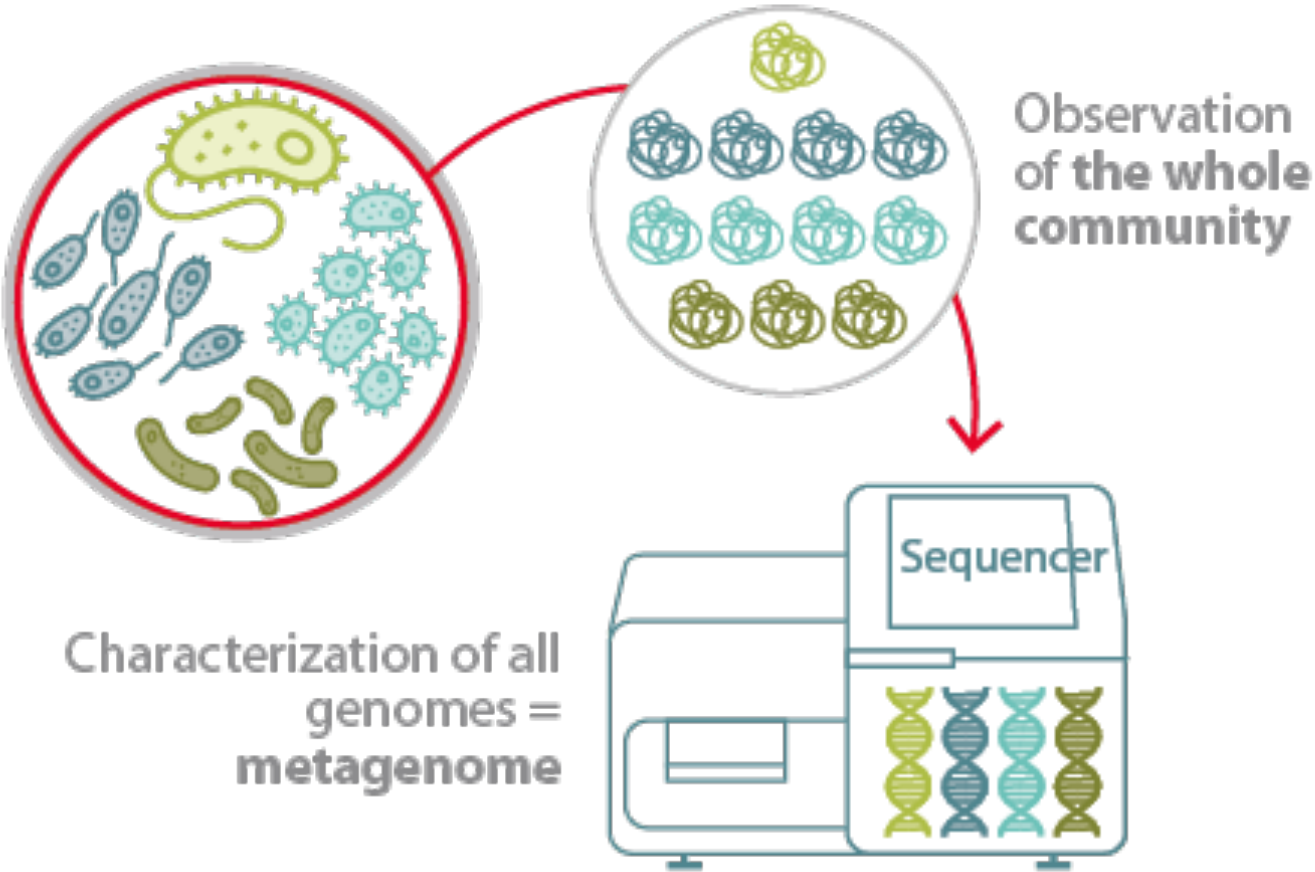
# NGS data analysis



# NGS data analysis

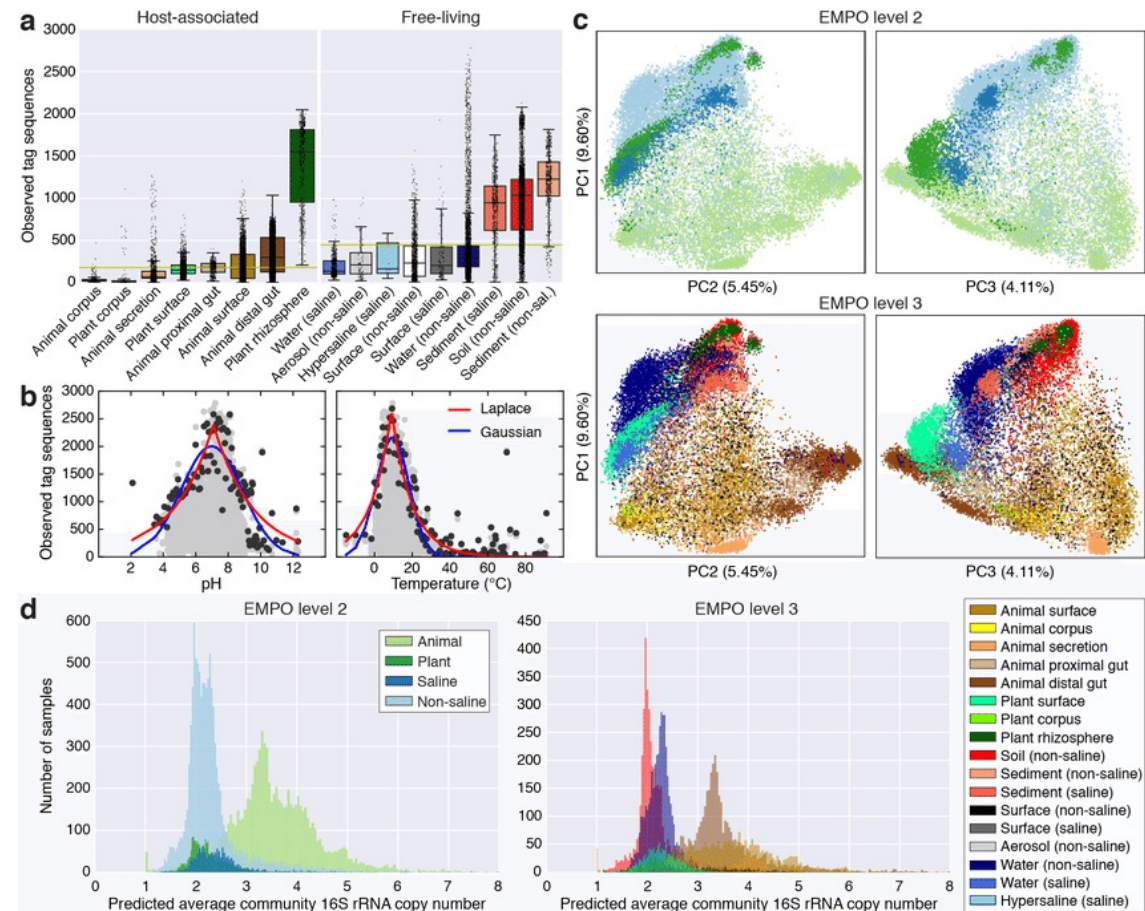


# Metagenomics



# Metagenomics results

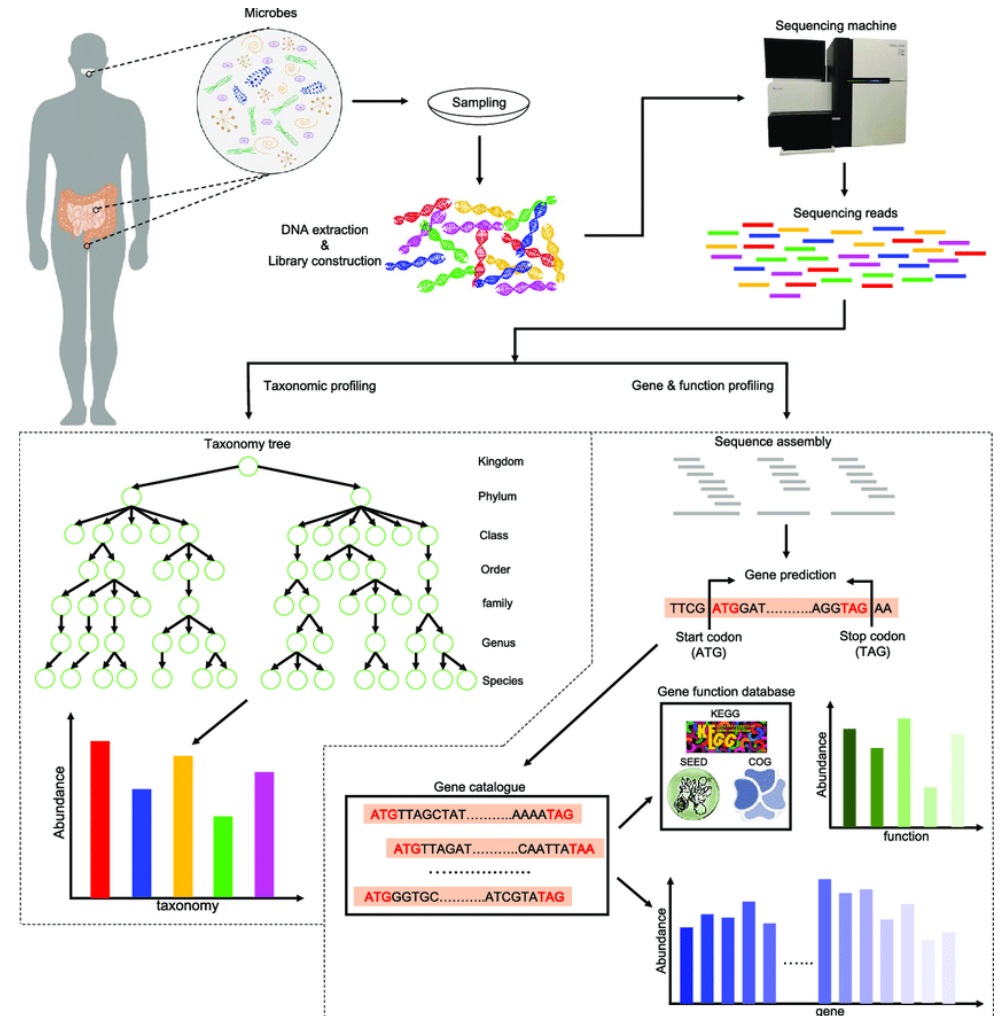
- Environmental statistics about populations
  - alpha, beta, gamma diversity





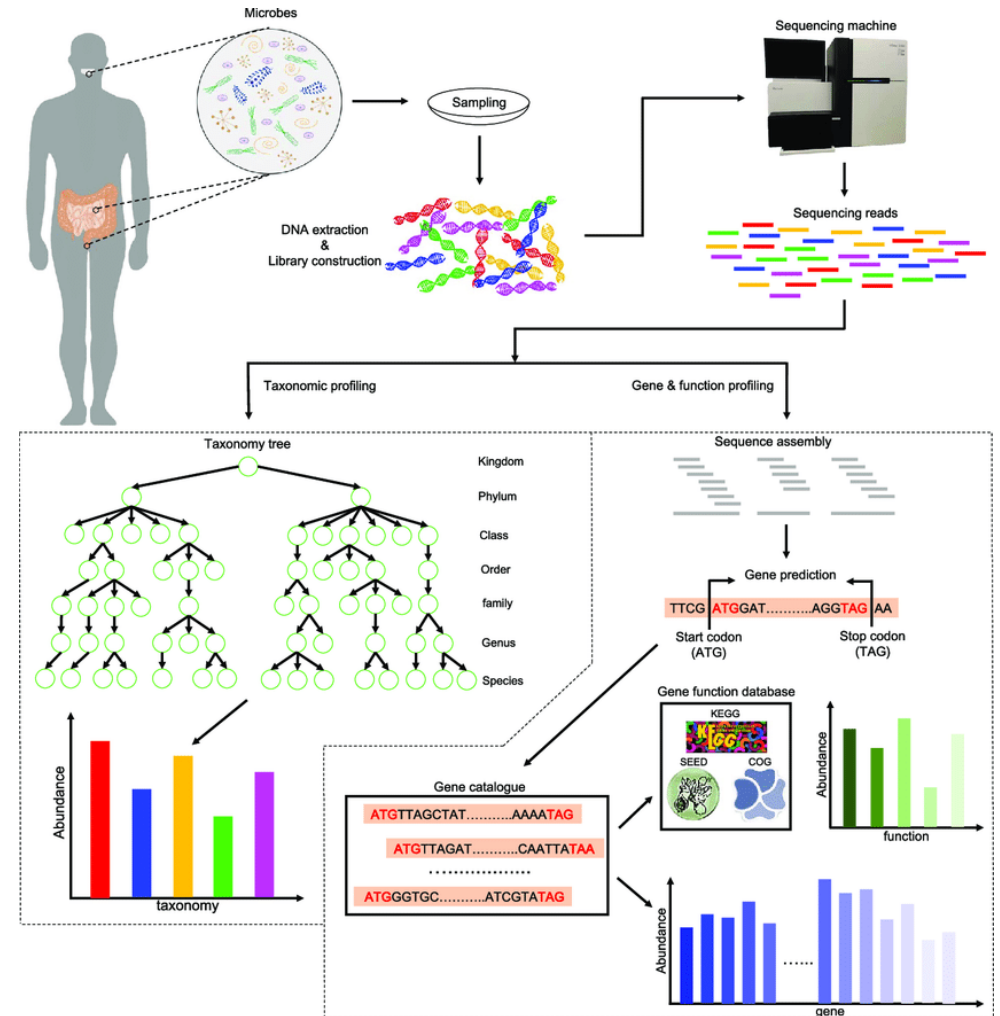
# Metagenomics results

- Environmental statistics about populations
  - identify known bacterial species
    - taxonomy profiling
  - eventually functional profiling
    - E.g. antimicrobial resistance genes



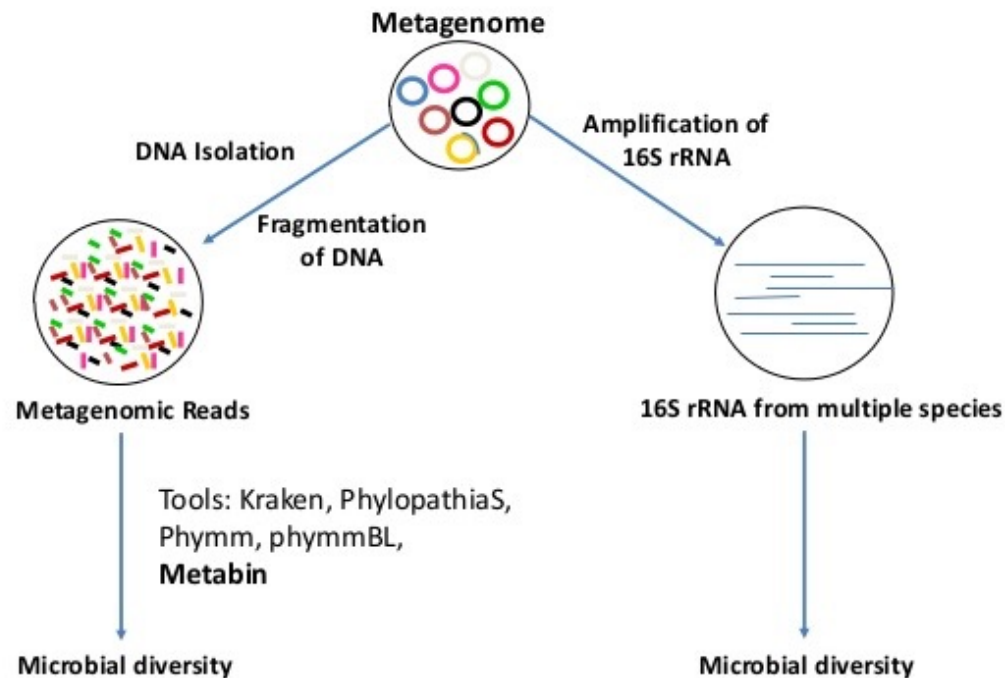
# Metagenomics results

- Environmental statistics about populations
  - identify known bacterial species
    - taxonomy profiling
  - eventually functional profiling
    - E.g. antimicrobial resistance genes
- Sequencing techniques
  - 16S rRNA sequencing
  - Shotgun metagenomic sequencing



# Metagenomics – 16S rRNA vs. Shotgun

## Metagenomic reads vs 16S rRNA for microbial diversity identification



Factors	16S rRNA sequencing	Shotgun Metagenomic Sequencing
Cost	~\$50 USD	Starting at ~\$150 but price will depend on sequencing depth required
Sample preparation	Similar complexity to shotgun sequencing	Similar complexity to 16S rRNA sequencing
Functional profiling (profile microbial genes)	No (but 'predicted' functional profiling is possible)	Yes (but it only reveals information on functional potential)
Taxonomic resolution: Genus, species, strain?	Bacterial genus (sometimes species); dependent on region(s) targeted	Bacterial species (sometimes strains and single nucleotide variants, if sequencing is deep enough)
Taxonomic coverage	Bacteria and archaea	All taxa, including viruses
Bioinformatics requirements	Beginner to intermediate expertise	Intermediate to advanced expertise
Databases	Established, well-curated	Relatively new, still growing
Sensitivity to host DNA contamination	Low (but PCR success depends on the absence of inhibitors and the presence of a detectable microbiome)	High, varies with sample type (but this can be mitigated by calibrating the sequencing depth)
Bias	Medium to high (retrieved taxonomic composition is dependent on selected primers and targeted variable region)	Lower (while metagenomics is "untargeted", experimental and analytical biases can be introduced at various stages)

# Metagenomics – 16S rRNA vs. Shotgun

- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**

# Metagenomics – 16S rRNA vs. Shotgun

- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**
    - 16S rRNA sequencing may provide more taxonomic resolution

# Metagenomics – 16S rRNA vs. Shotgun

- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**
    - 16S rRNA sequencing may provide more taxonomic resolution
  - **Changes in microbiome composition and antimicrobial gene carriage following fecal transplant**

# Metagenomics – 16S rRNA vs. Shotgun

- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**
    - 16S rRNA sequencing may provide more taxonomic resolution
  - **Changes in microbiome composition and antimicrobial gene carriage following fecal transplant**
    - shotgun sequencing to assess both compositional and functional differences

# Metagenomics – 16S rRNA vs. Shotgun

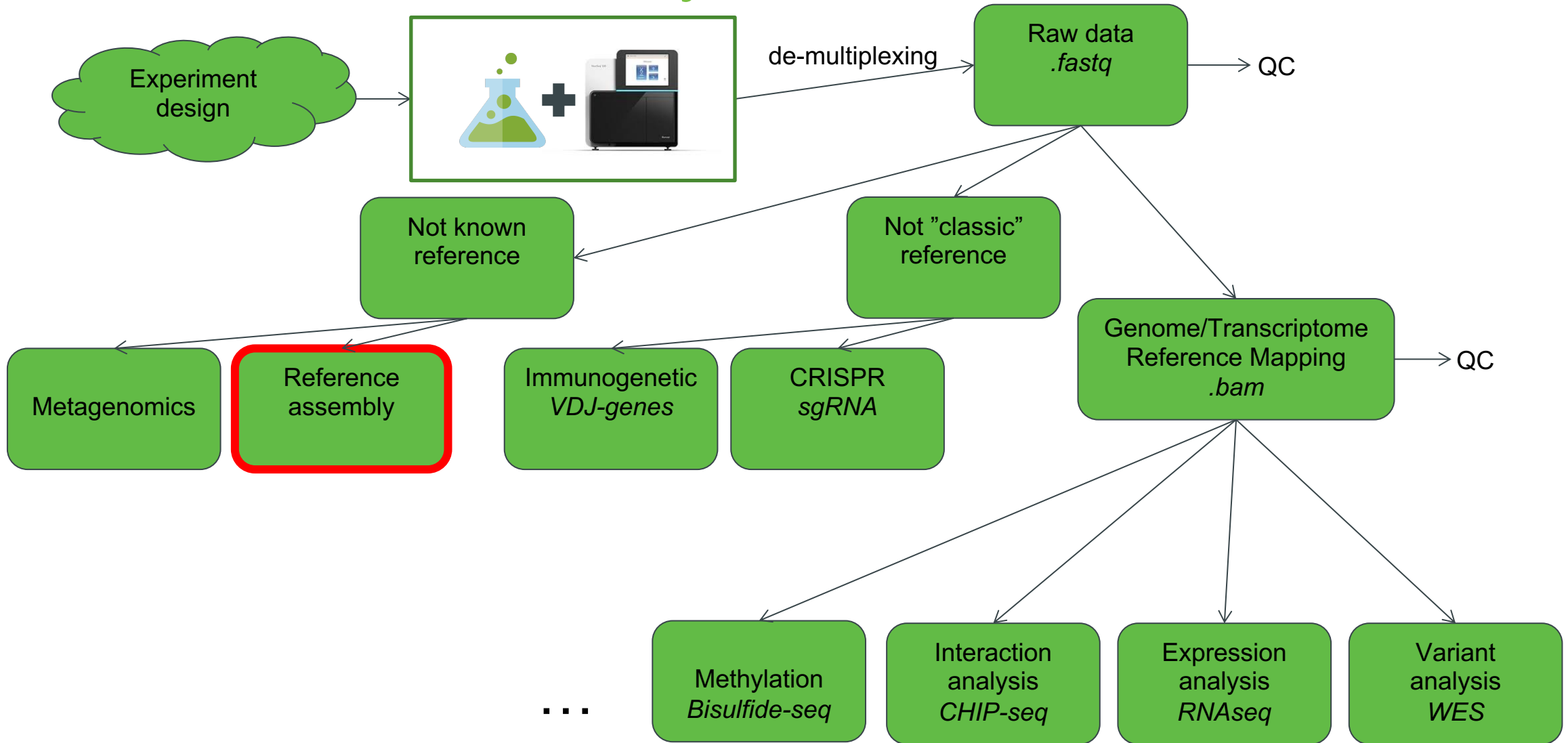
- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**
    - 16S rRNA sequencing may provide more taxonomic resolution
  - **Changes in microbiome composition and antimicrobial gene carriage following fecal transplant**
    - shotgun sequencing to assess both compositional and functional differences
  - **Daily fluctuations in gut microbiome following 2 week dietary fiber intervention**



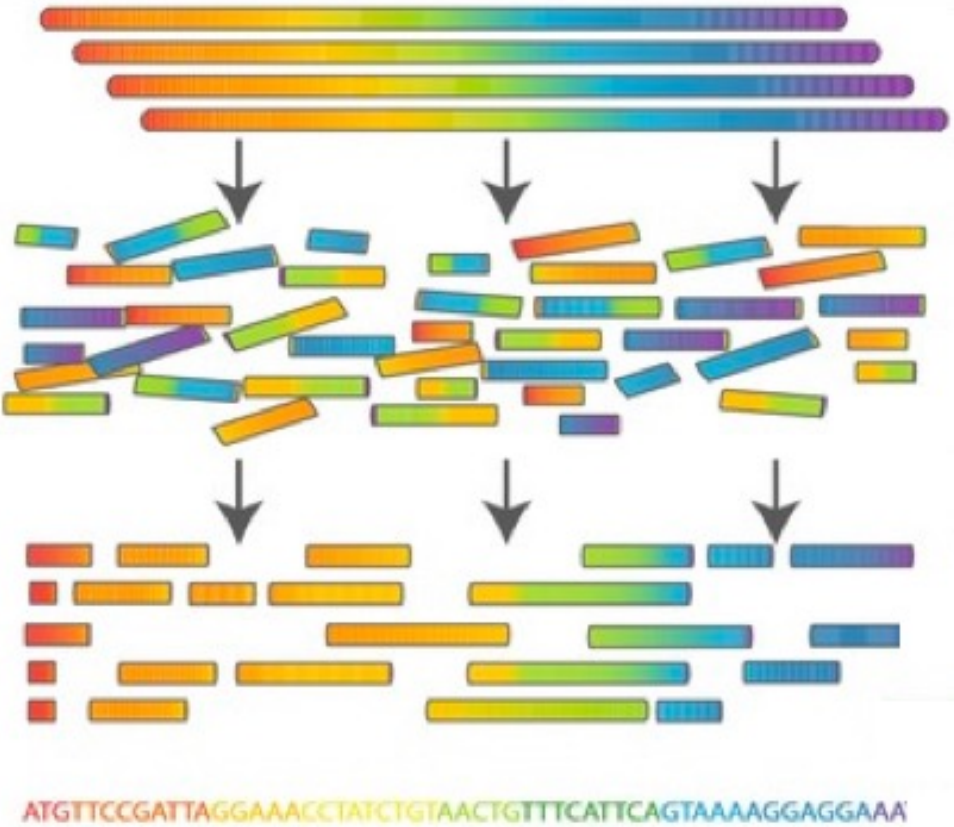
# Metagenomics – 16S rRNA vs. Shotgun

- Study Examples
  - **Assessment of the bacterial microbiome of Amazonian soil**
    - 16S rRNA sequencing may provide more taxonomic resolution
  - **Changes in microbiome composition and antimicrobial gene carriage following fecal transplant**
    - shotgun sequencing to assess both compositional and functional differences
  - **Daily fluctuations in gut microbiome following 2 week dietary fiber intervention**
    - shotgun sequencing or 16S rRNA
      - assess both compositional and functional differences
      - cheaper and in this case can use 'predicted' functional profiling

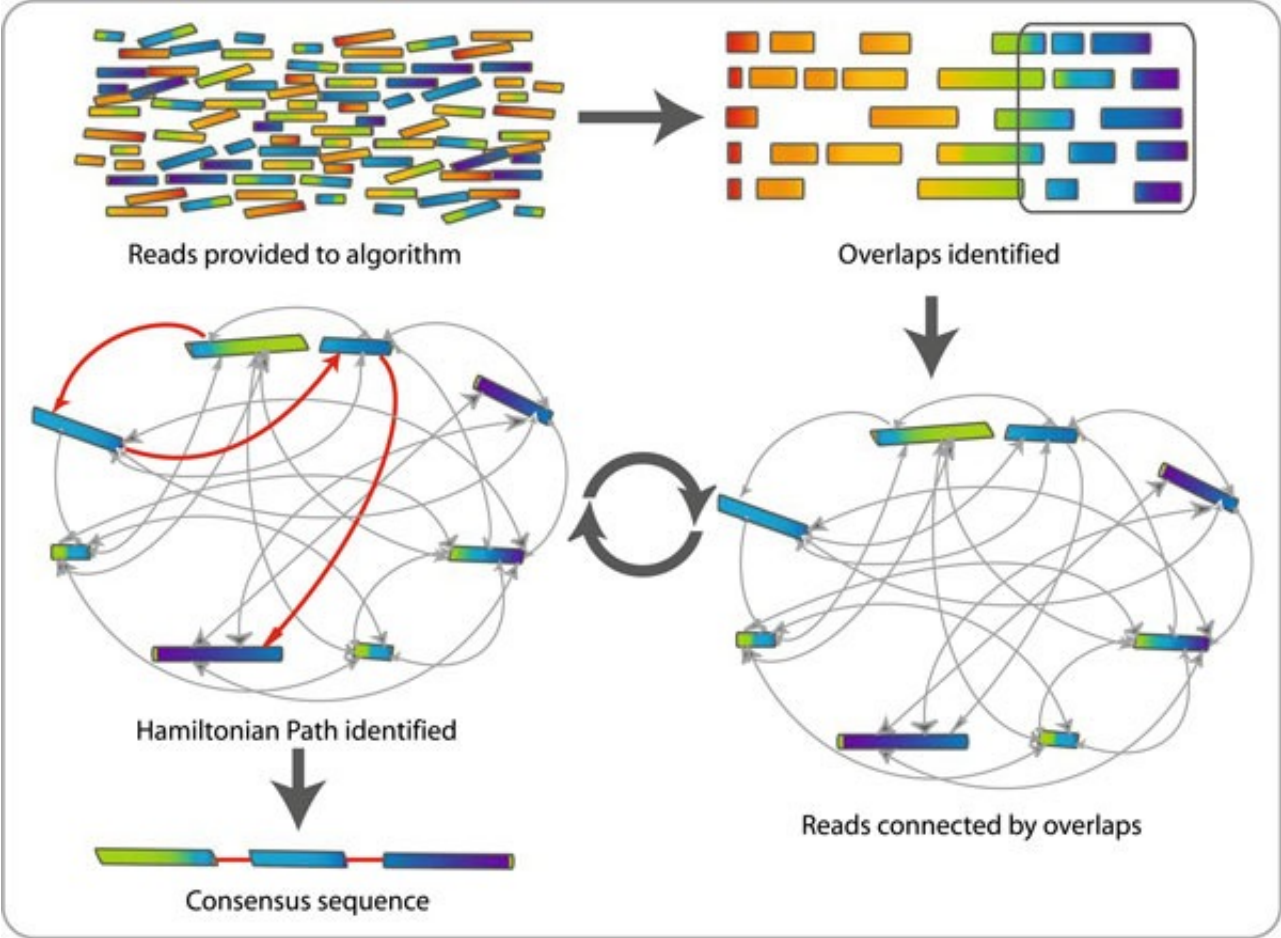
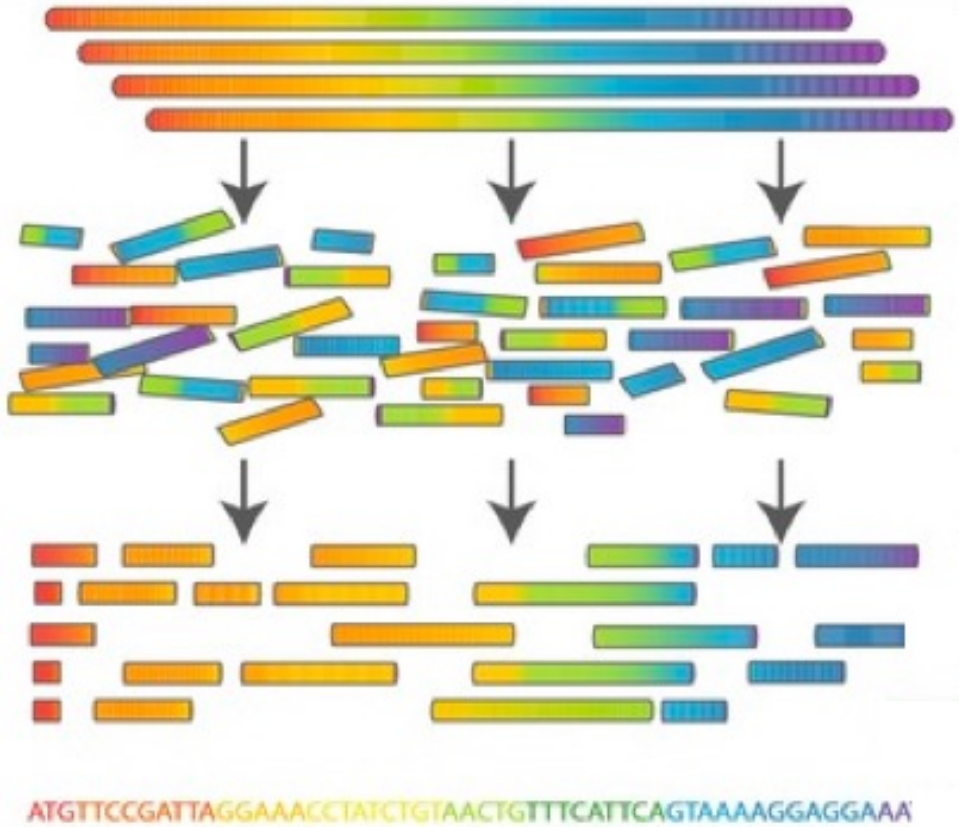
# NGS data analysis



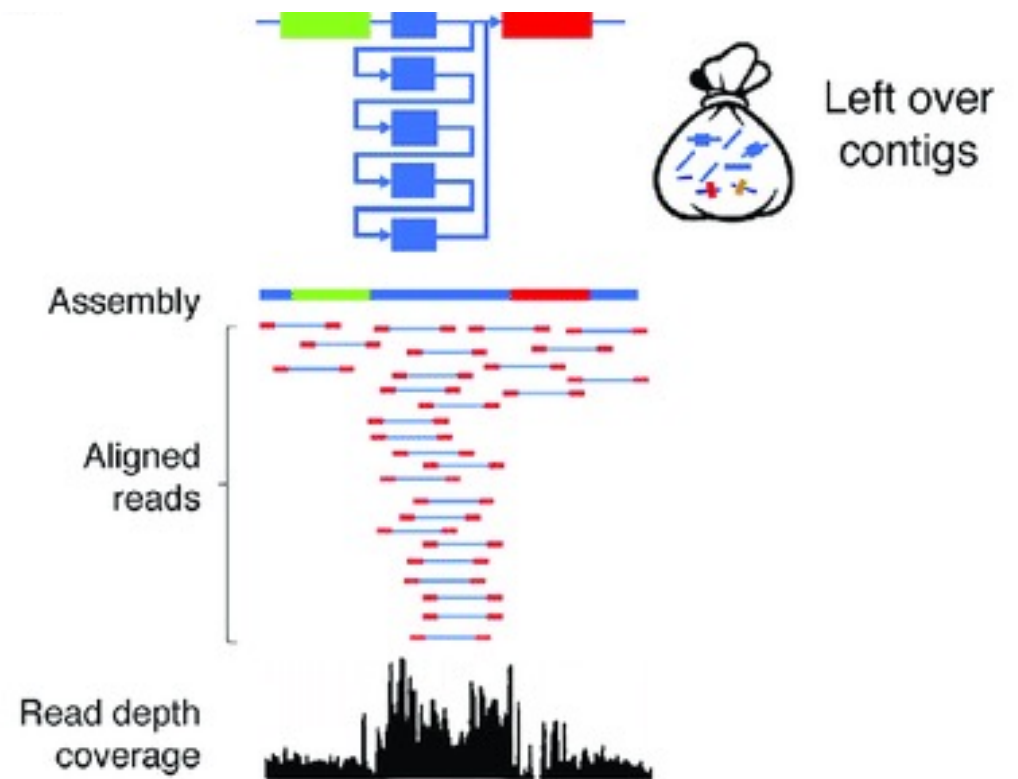
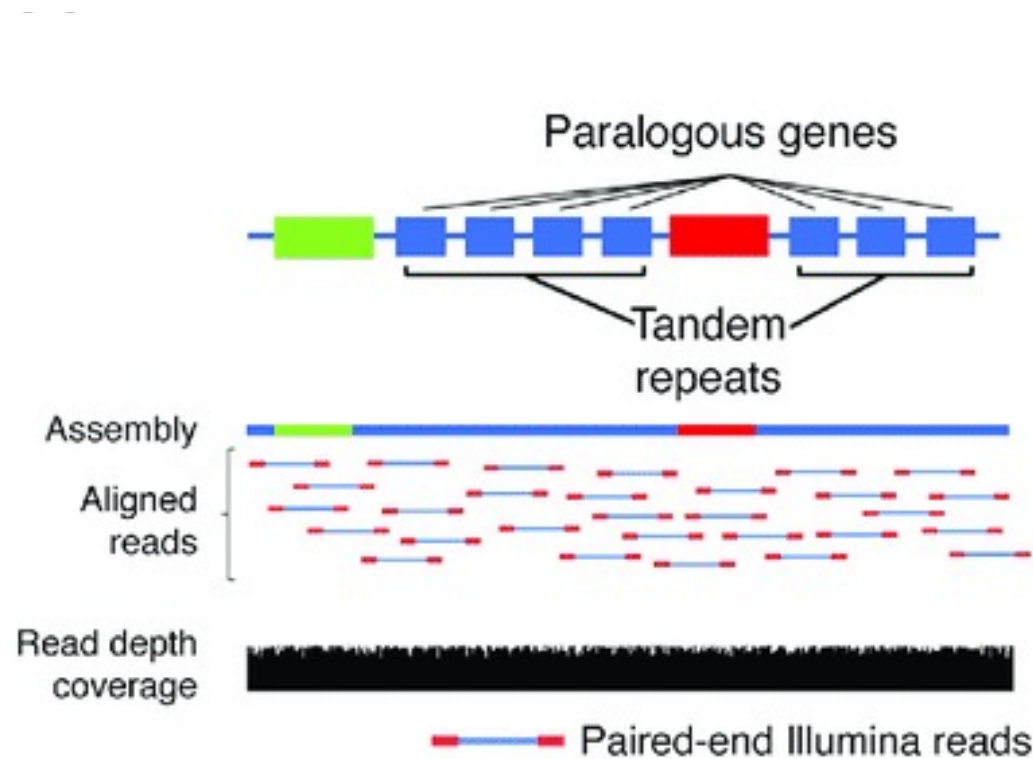
# Reference Assembly



# Reference Assembly

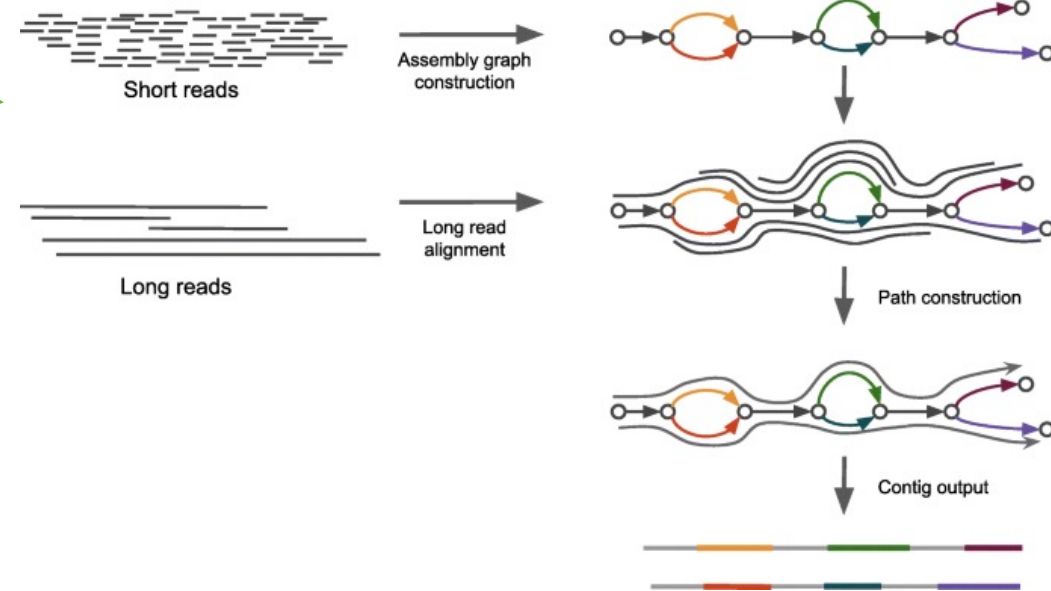
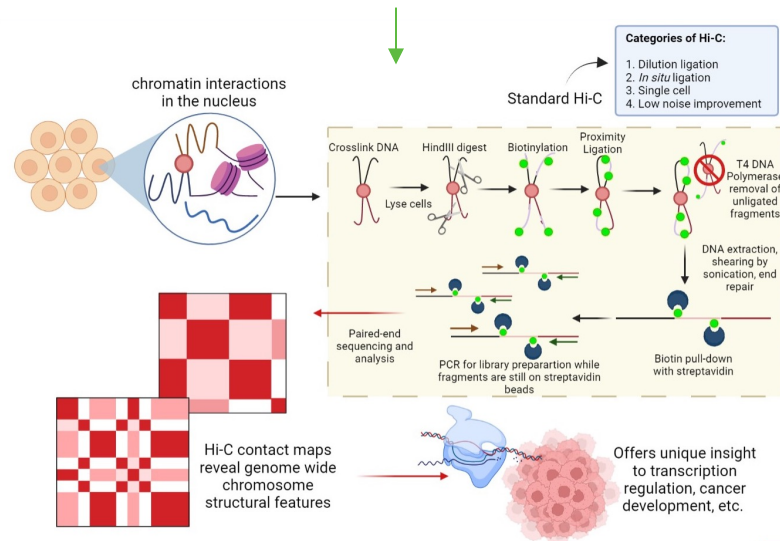


# Reference Assembly problematic with short read



# Genome Assembly

- Very hard and costly (in eukaryota)
- Multiple sequencing types needed
  - Pair-end short reads
  - Long reads
  - Mate-pairs (e.g. Hi-C)

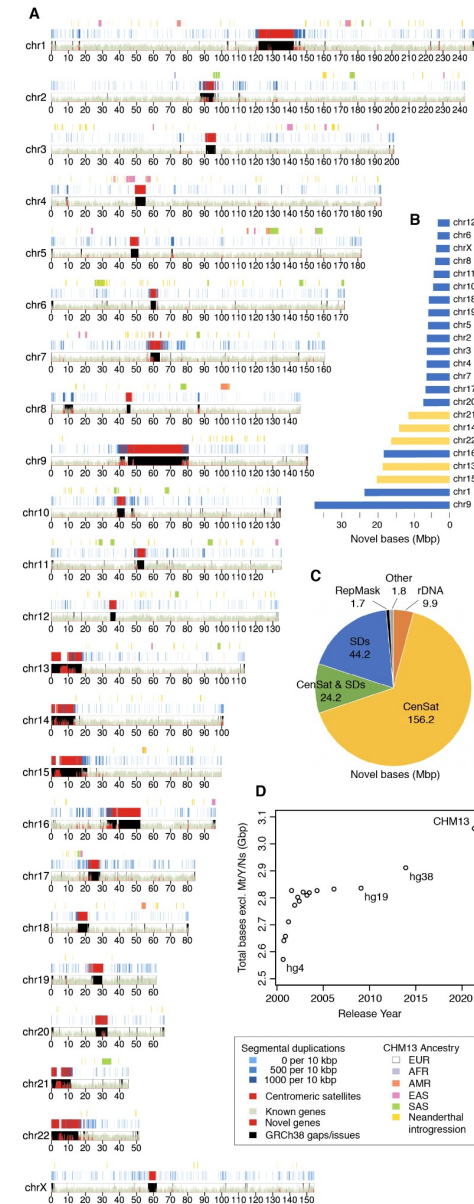


Created in BioRender.com bio

# Genome Assembly

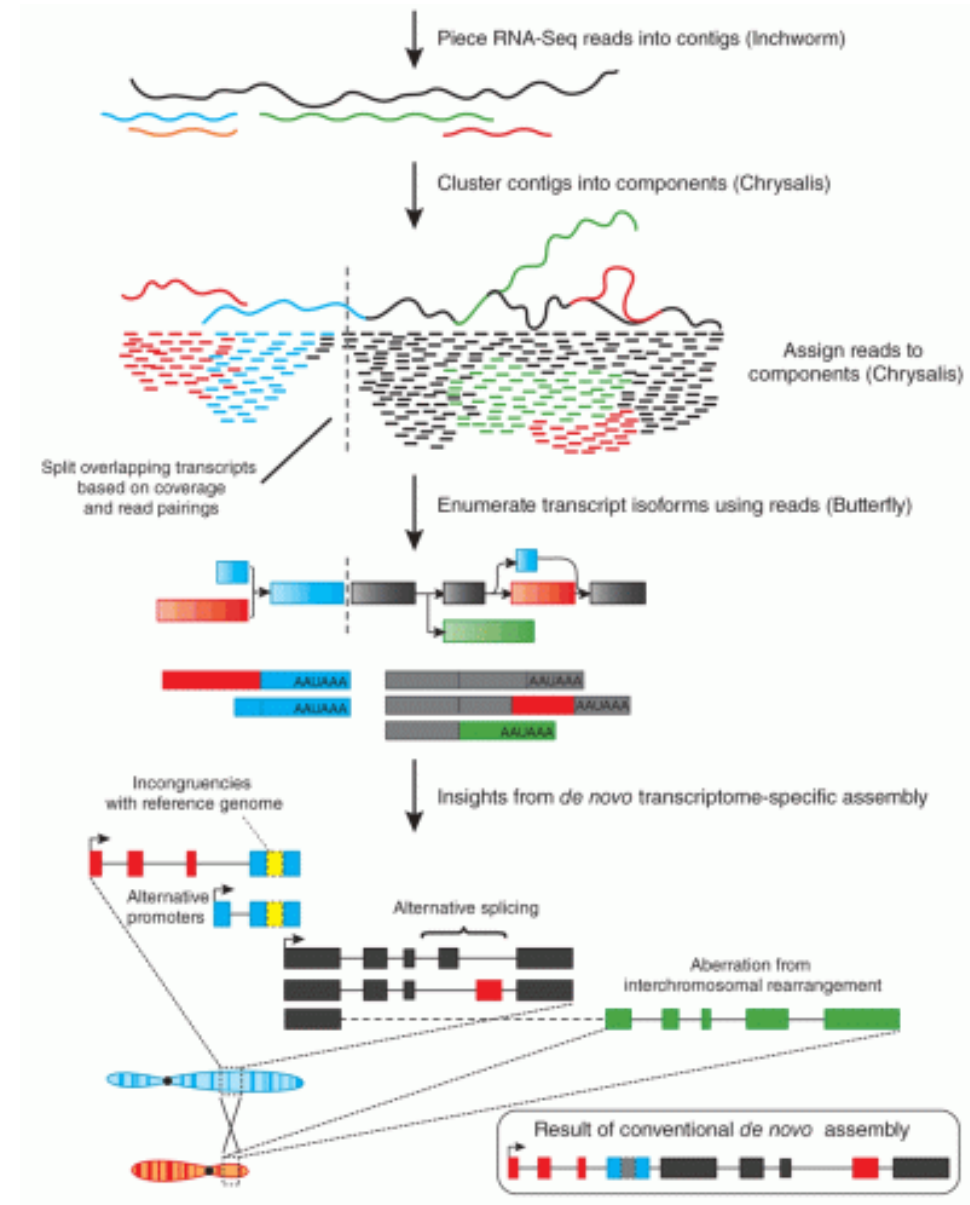
- Very hard and costly (in eukaryota)
- Multiple sequencing types needed
  - Pair-end short reads
  - Long reads
  - Mate-pairs (e.g. Hi-C)

T2T-CHM13



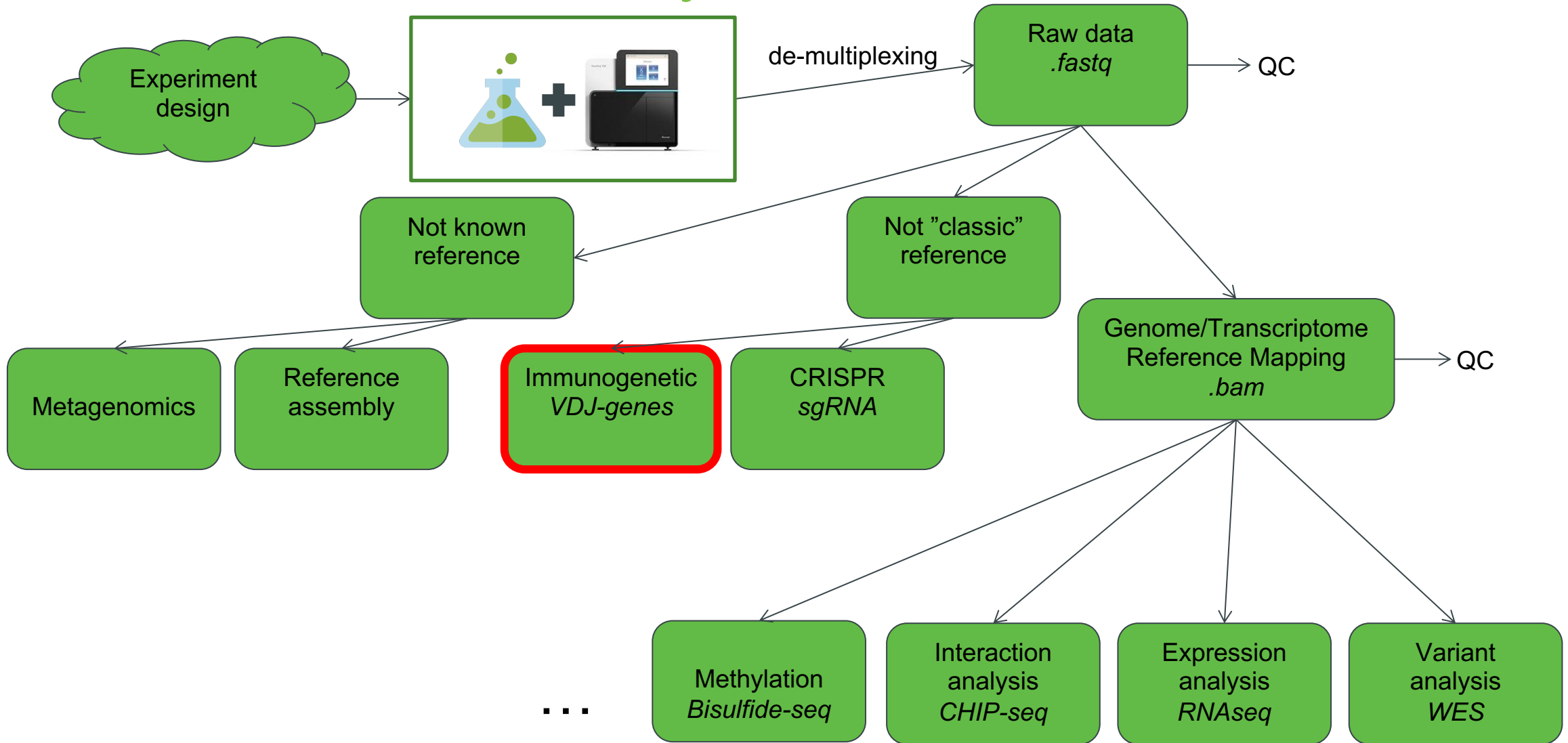
# Transcriptome Assembly

- Assemble RNA fragments
  - Similar reference helpful
- Genome guided assembly
  - Good for poorly annotated organisms with known genomic reference



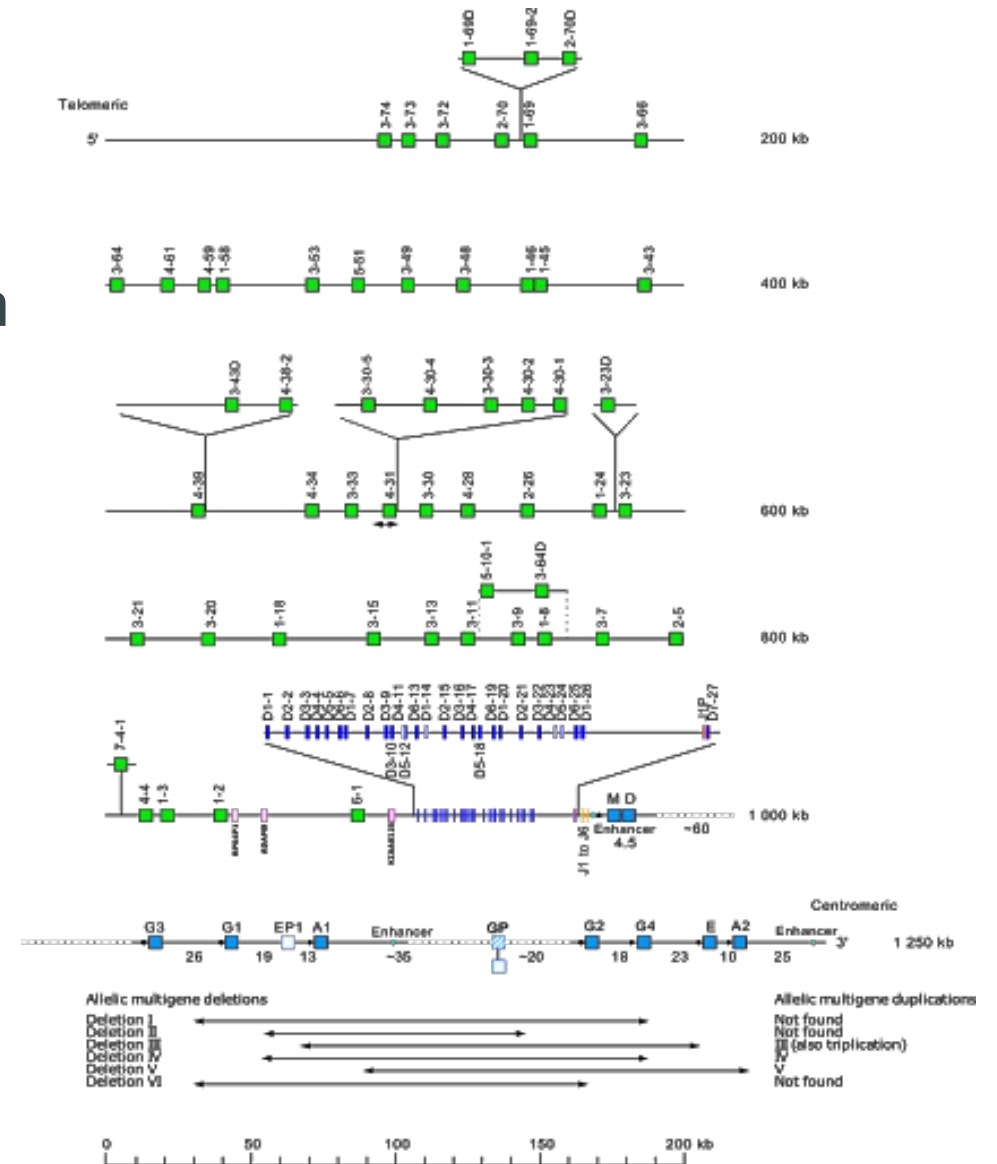


# NGS data analysis



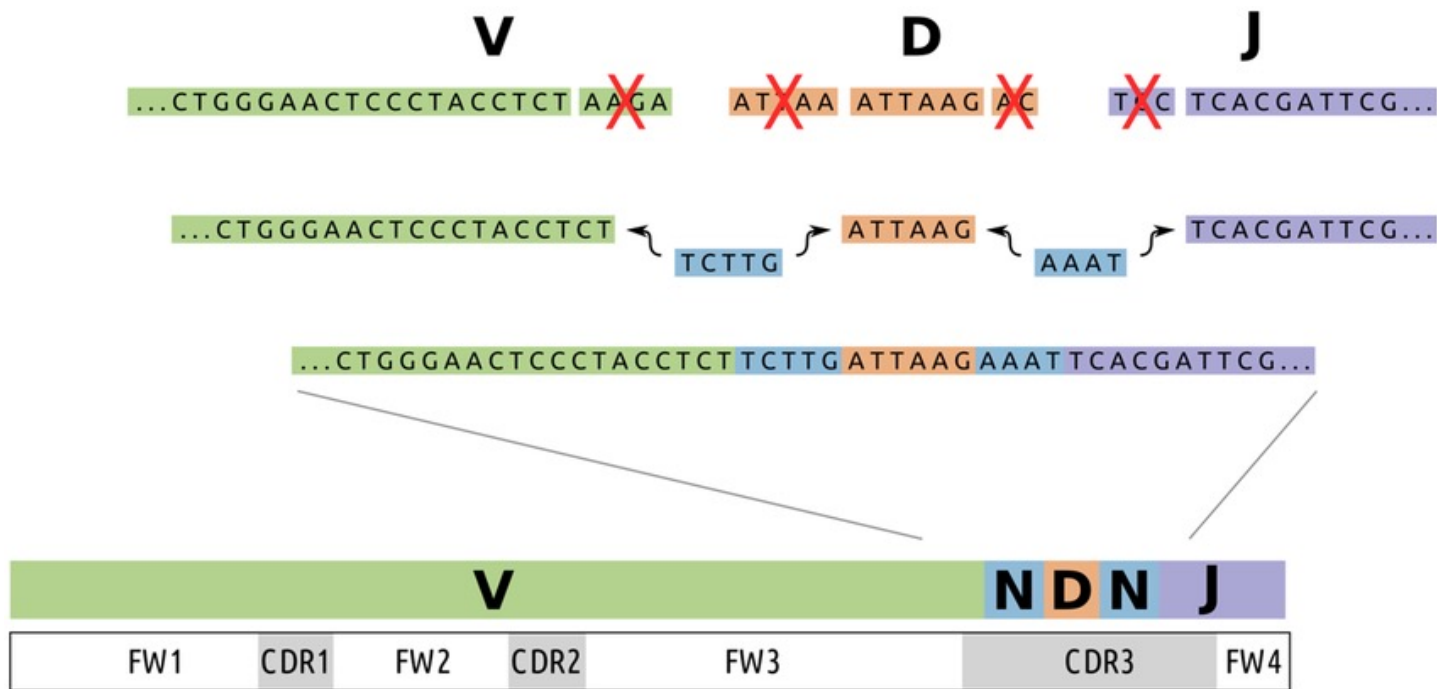
# Immunogenetic

- T-cell receptor , Immunoglobulin – (B-cell)
- Gene rearrangement during cell maturation
  - VDJ recombination



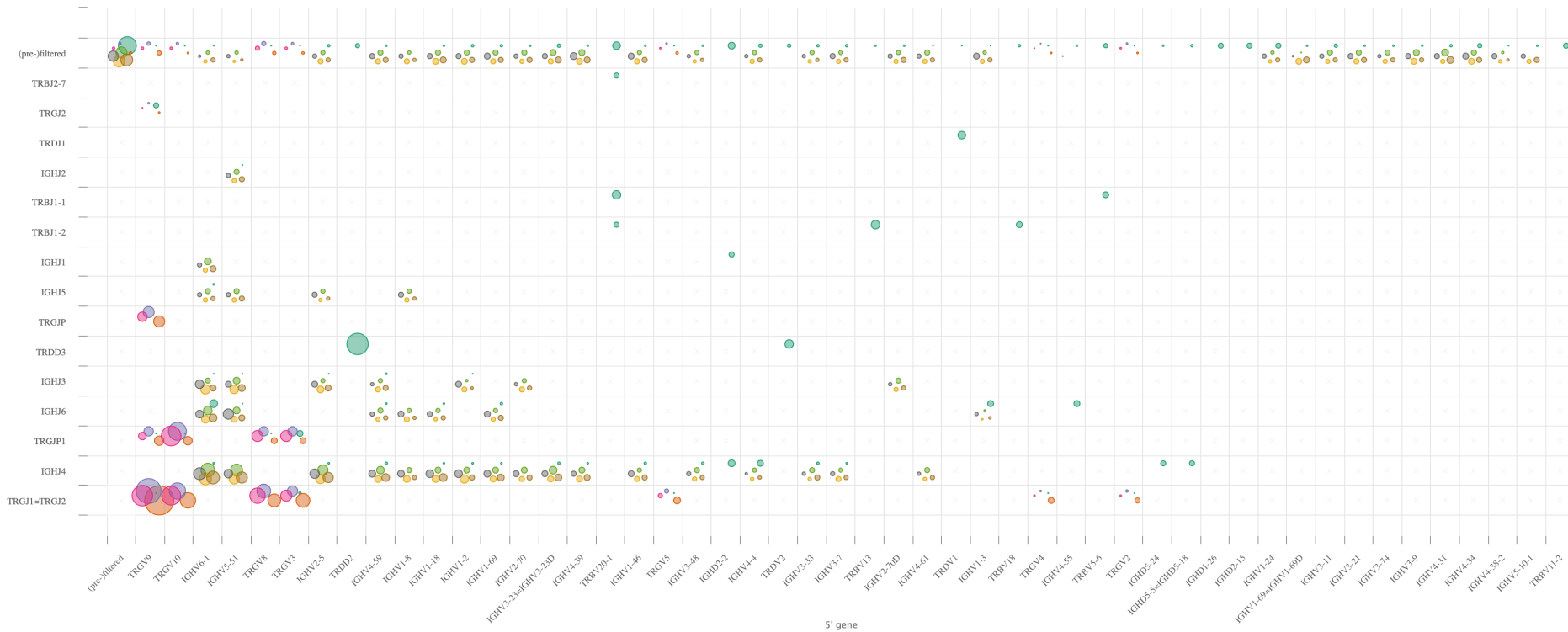
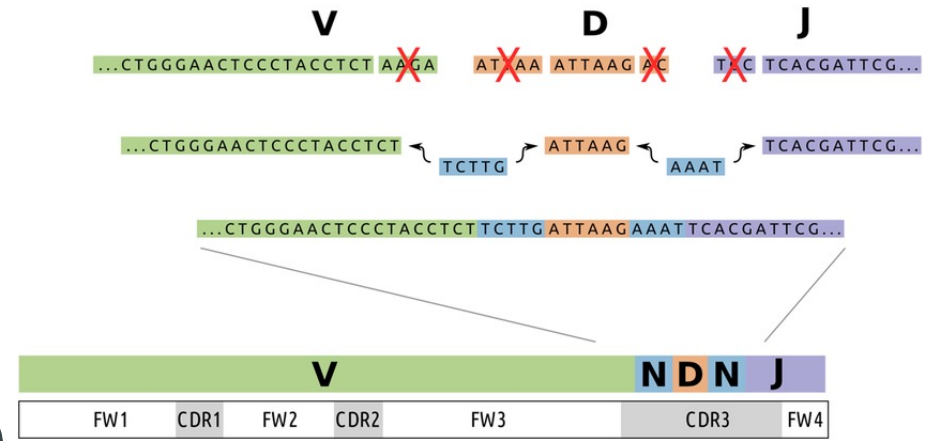
# Immunogenetic

- T-cell receptor , Immunoglobulin – (B-cell)
- Gene rearrangement during cell maturation
  - VDJ recombination

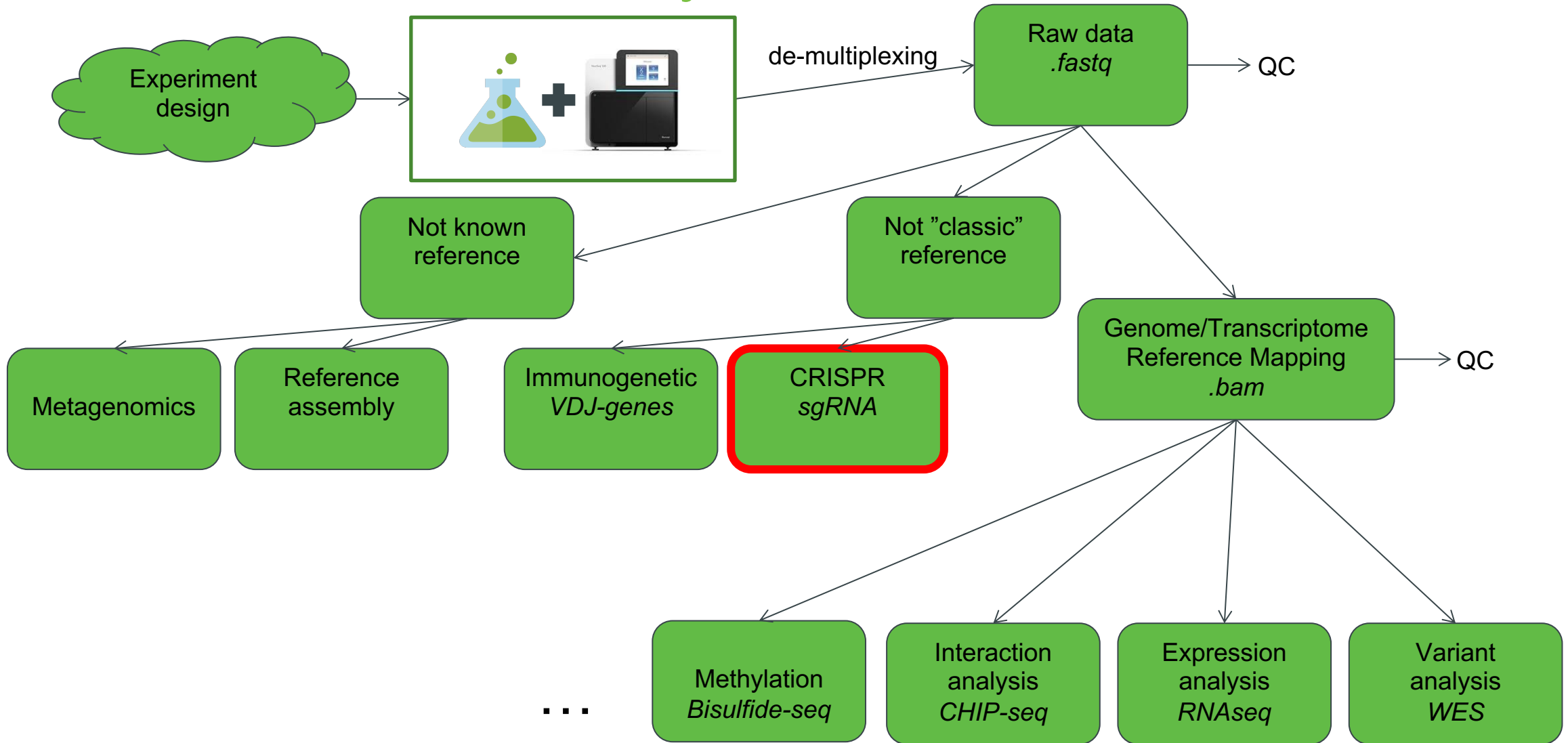


# Immunogenetic

- Different cell populations
  - Clonal studies
  - Repertoire usage
- Main usage – blood malignancies (leukemias)

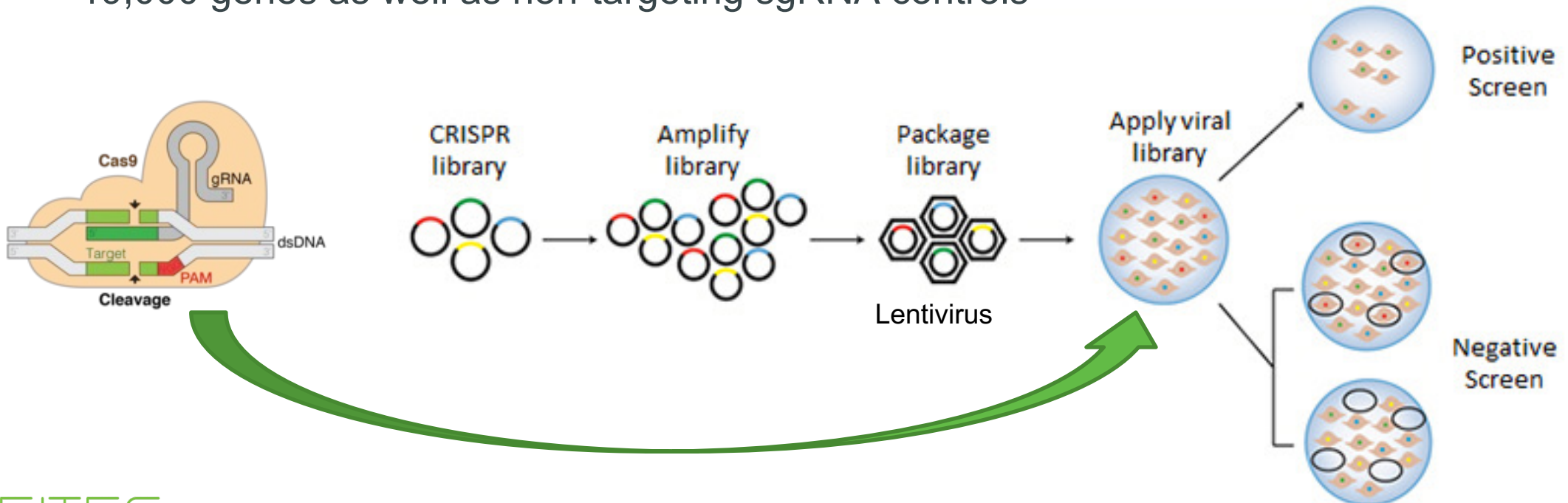


# NGS data analysis



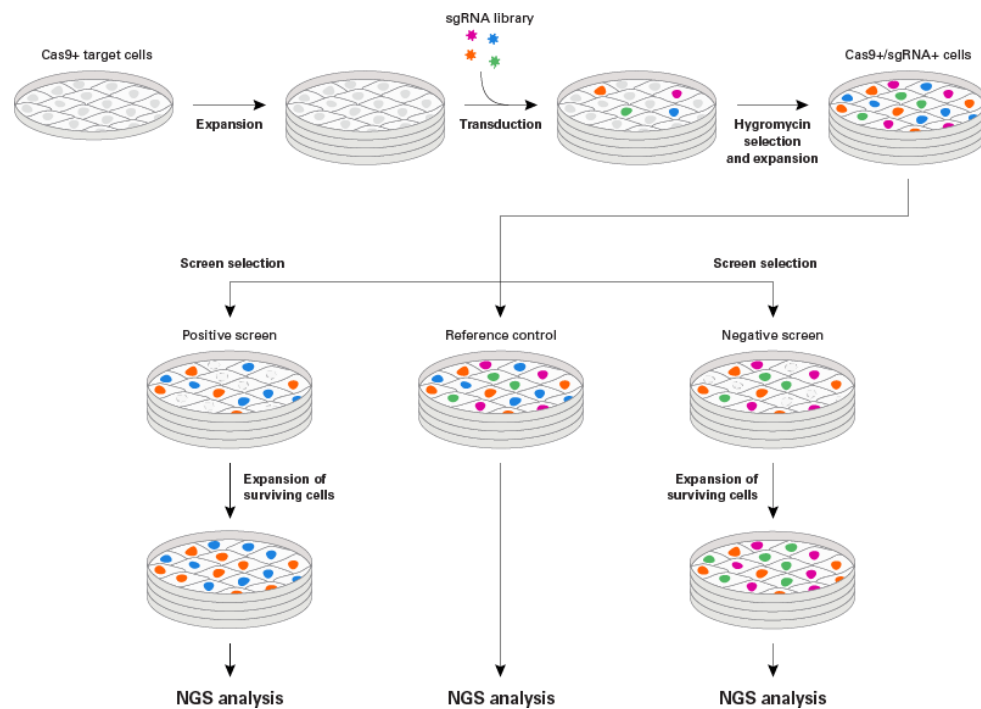
# Genome-wide CRISPR-Cas9 knockout screens

- Cas9 (CRISPR associated protein 9) is a protein which plays a vital role in the immunological defense of certain bacteria against DNA viruses
- sgRNA libraries
  - Each sgRNA knockout specific gene
  - 76,000 guide RNAs (sgRNAs) with four highly active guides per gene, targeting about 19,000 genes as well as non-targeting sgRNA controls



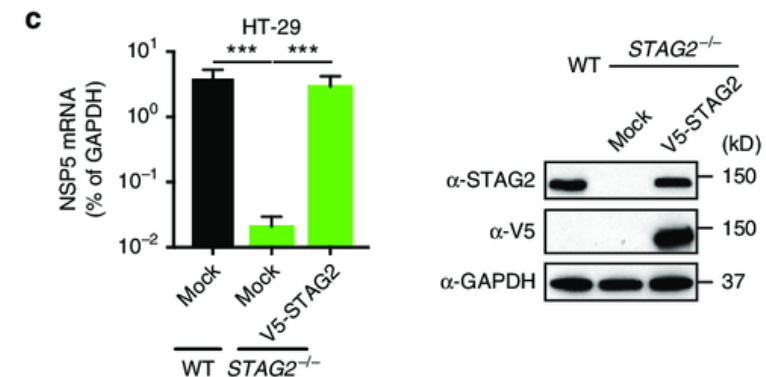
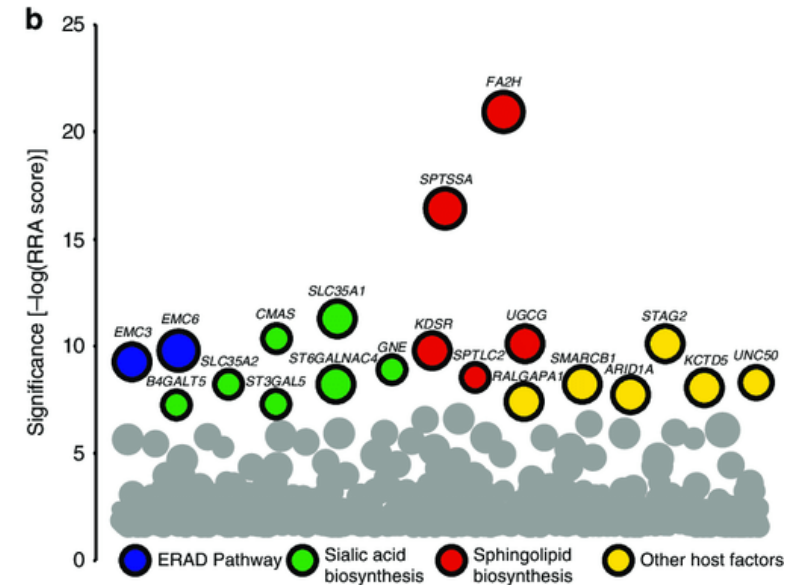
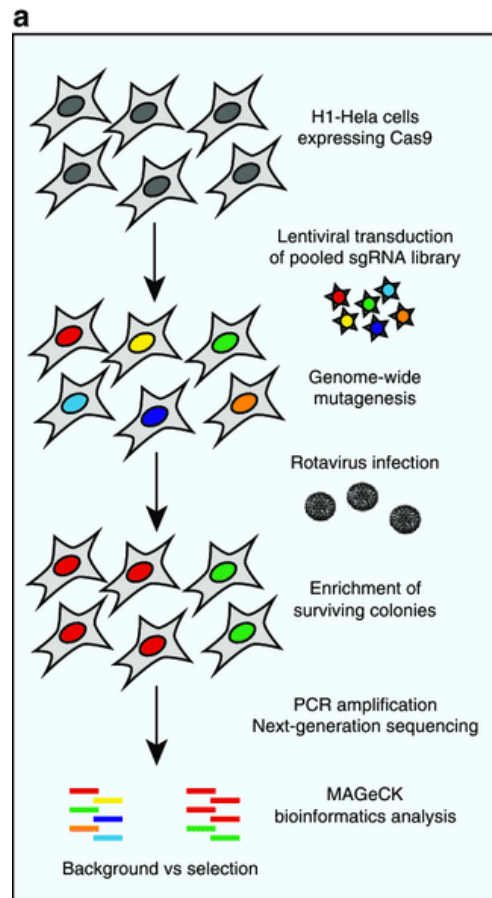
# Genome-wide CRISPR-Cas9 knockout screens

- Screen selection + expansion/enrichment of surviving cells
- NGS sequencing



# Genome-wide CRISPR-Cas9 knockout screens

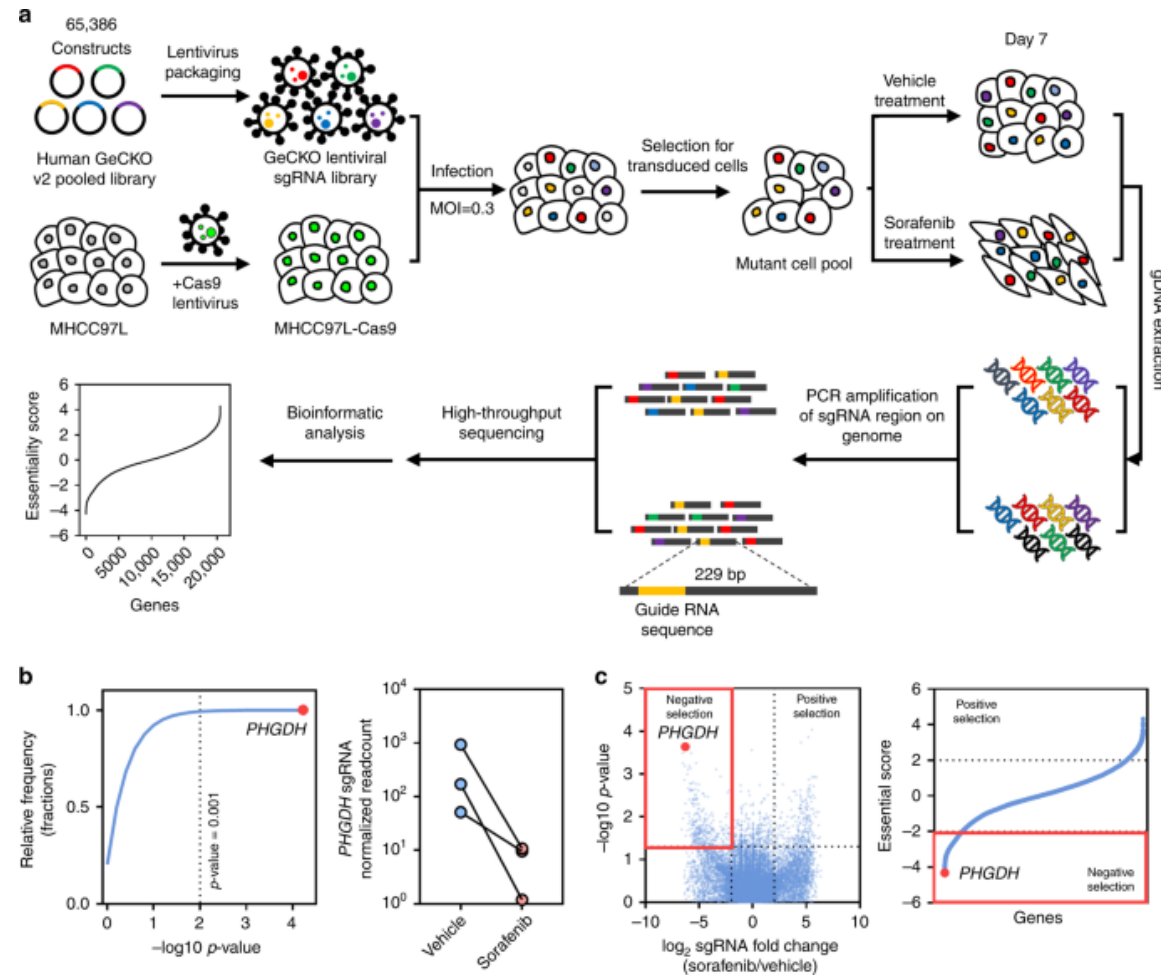
- NGS data analysis
  - Counting cells with different genes KD
  - Counting sgRNA fragments
  - Compare conditions





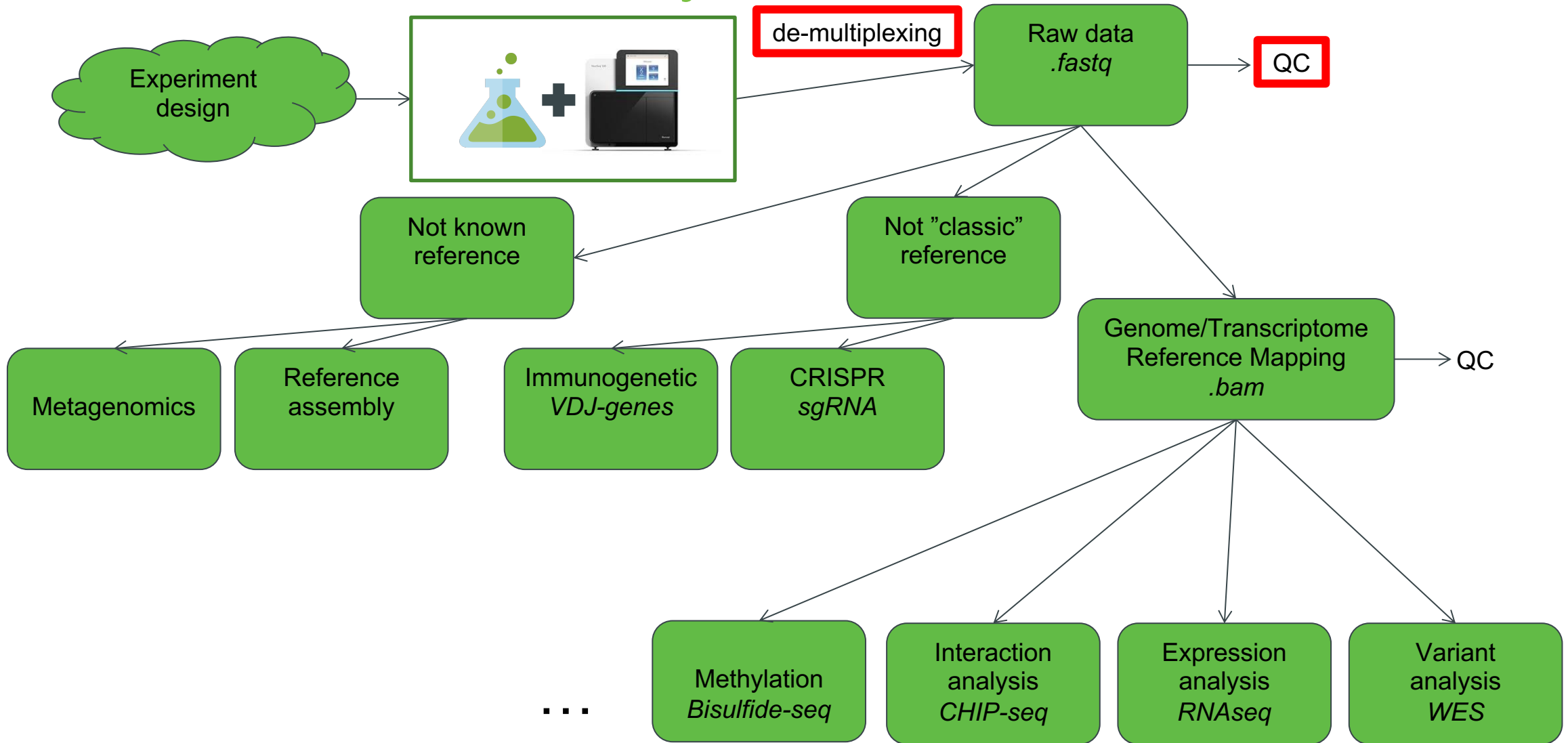
# Genome-wide CRISPR-Cas9 knockout screens

- Example study



Wei, L., Lee, D., Law, CT. *et al.* Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC. *Nat Commun* 10, 4681 (2019). <https://doi.org/10.1038/s41467-019-12606-7>

# NGS data analysis





CEITEC



@CEITEC\_Brno

Thank you for your attention!

