



Omics

Name: 苗奔奔

Student ID: 2111901034

Email: benben.miao@outlook.com

Date: 2019-12-04

Github: <https://github.com/benben-miao/>



Catalog

**1.
Genomics**

**2.
Transcripto
mics**

**3.
Microorga
nism**

**4.
Proteomics**

**5.
Metabolo
mics**

**6.
Epigenomi
cs**

**7.
Single Cell**

1. Genomics

Whole Genome Sequencing

RAD/GBS

GWAS (Genome-wide Association Study)

mtDNA (Mitochondrial DNA)

cpDNA (Chloroplast DNA,)

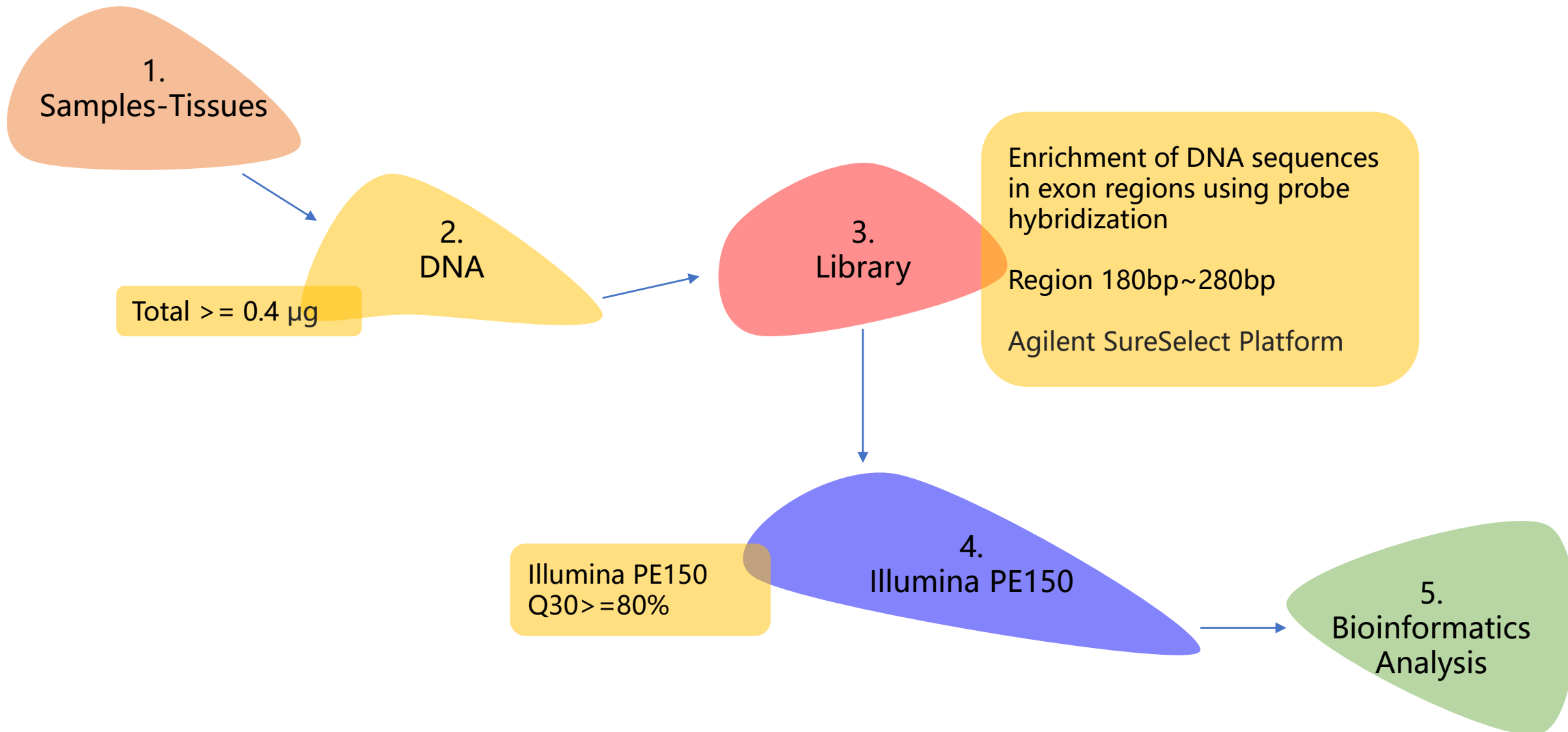
Genetic map

WES-Seq (Whole Exome Sequencing)

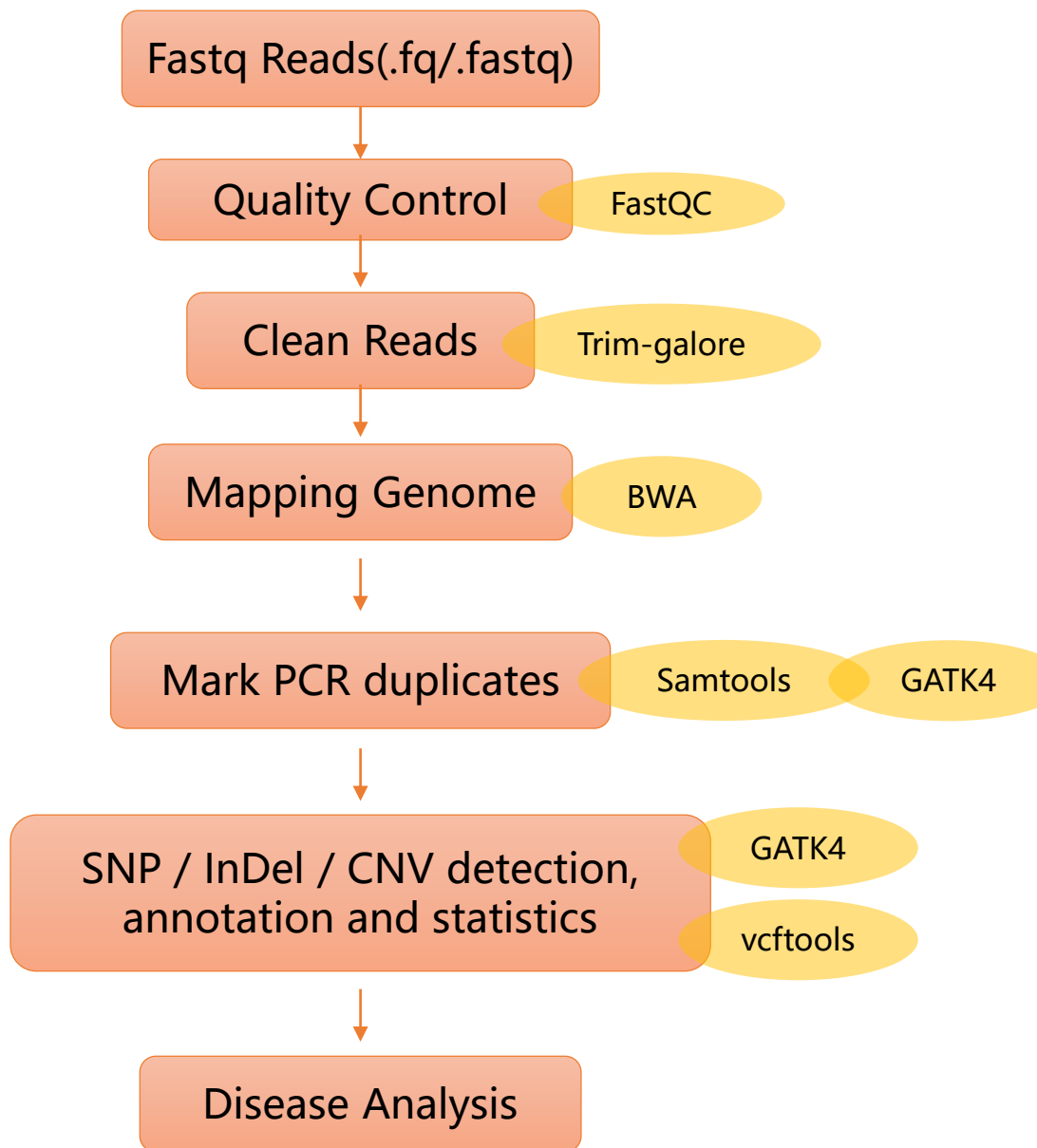
eQTL-Seq (expression quantitative trait Loci)

Experimental design

WES-Seq (Whole Exome Sequencing)



Analysis Process



```
1 source activate wes
2 GATK=/home/jmzeng/biosoft/gatk4/gatk-4.0.6.0/gatk
3 ref=/public/biosoft/GATK/resources/bundle/hg38/Homo_sapiens_assembly38.fasta
4 snp=/public/biosoft/GATK/resources/bundle/hg38/dbsnp_146.hg38.vcf.gz
5 indel=/public/biosoft/GATK/resources/bundle/hg38/Mills_and_1000G_gold_standard.indels.hg38.vcf
6
7 for sample in {7E5239.L1,7E5240,7E5241.L1}
8 do
9 echo $sample
10 # Elapsed time: 7.91 minutes
11 $GATK --java-options "-Xmx20G -Djava.io.tmpdir=/" MarkDuplicates \
12 -I $sample.bam \
13 -O ${sample}_marked.bam \
14 -M $sample.metrics \
15 1>${sample}_log.mark 2>&1
16
17 ## Elapsed time: 13.61 minutes
18 $GATK --java-options "-Xmx20G -Djava.io.tmpdir=/" FixMateInformation \
19 -I ${sample}_marked.bam \
20 -O ${sample}_marked_fixed.bam \
21 -SO coordinate \
22 1>${sample}_log.fix 2>&1
23
24 samtools index ${sample}_marked_fixed.bam
25
26 ## 17.2 minutes
27 $GATK --java-options "-Xmx20G -Djava.io.tmpdir=/" BaseRecalibrator \
28 -R $ref \
29 -I ${sample}_marked_fixed.bam \
30 --known-sites $snp \
31 --known-sites $indel \
32 -O ${sample}_recal.table \
33 1>${sample}_log.recal 2>&1
34
35 $GATK --java-options "-Xmx20G -Djava.io.tmpdir=/" ApplyBQSR \
36 -R $ref \
37 -I ${sample}_marked_fixed.bam \
38 -bqsr ${sample}_recal.table \
39 -O ${sample}_bqsr.bam \
40 1>${sample}_log.ApplyBQSR 2>&1
41
42 ## 使用GATK的HaplotypeCaller命令
43 $GATK --java-options "-Xmx20G -Djava.io.tmpdir=/" HaplotypeCaller \
44 -R $ref \
45 -I ${sample}_bqsr.bam \
46 --dbsnp $snp \
47 -O ${sample}_raw.vcf \
48 1>${sample}_log.HC 2>&1
49
50 done
51
```

Databases:

NCBI (*The National Center for Biotechnology Information*) : <https://www.ncbi.nlm.nih.gov/>

EMBL (The European Bioinformatics Institute) : <https://www.ebi.ac.uk/>

DDBJ () : <https://www.ddbj.nig.ac.jp/index-e.html>

Plant Genome DB: http://ensembl.gramene.org/genome_browser/index.html

Human/Mouse/Zebrafish: <http://asia.ensembl.org/index.html>

Fungi(真菌) Genome Database: <http://asia.ensembl.org/index.html>

Bacteria Genome Database : <http://bacteria.ensembl.org/index.html>

Biodb Collection (benben-miao) : <https://biodb.neocities.org/>

Softwares:

GATK4: <https://software.broadinstitute.org/gatk/>

Samtools: <http://samtools.sourceforge.net/>

Vcftools : <http://vcftools.sourceforge.net/>

BWA : <http://bio-bwa.sourceforge.net/>

2. Transcriptomics

Eukaryotic reference transcriptome

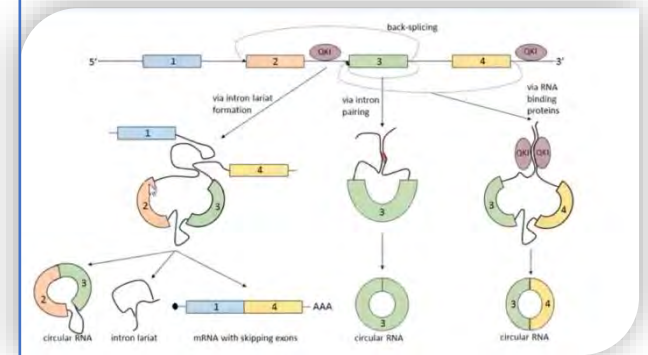
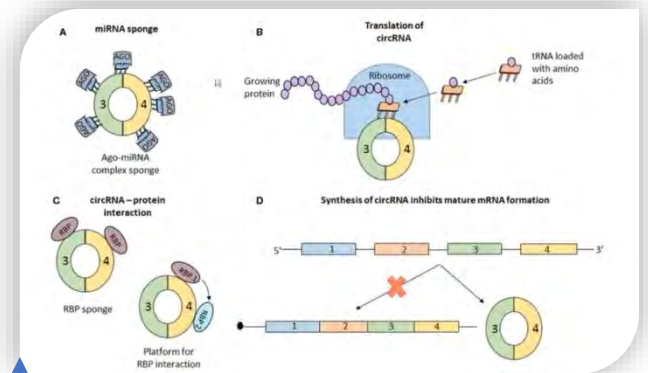
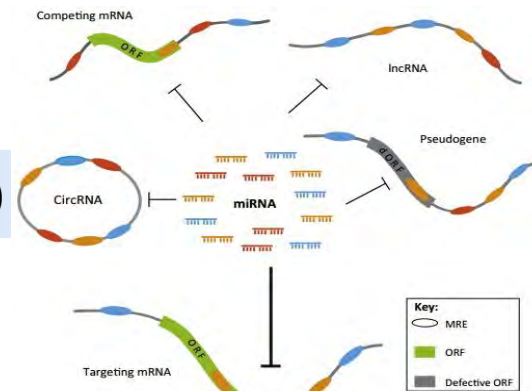
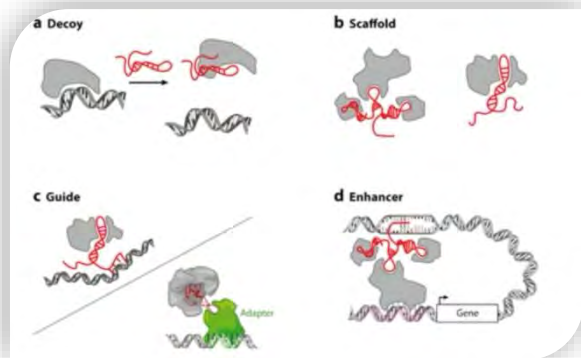
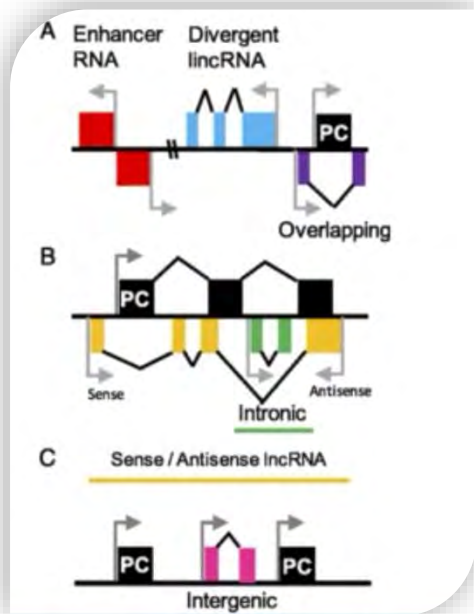
Eukaryotic referenceless transcriptome

LncRNA-Seq (Long No Code RNA)

miRNA-Seq (Micro RNA)

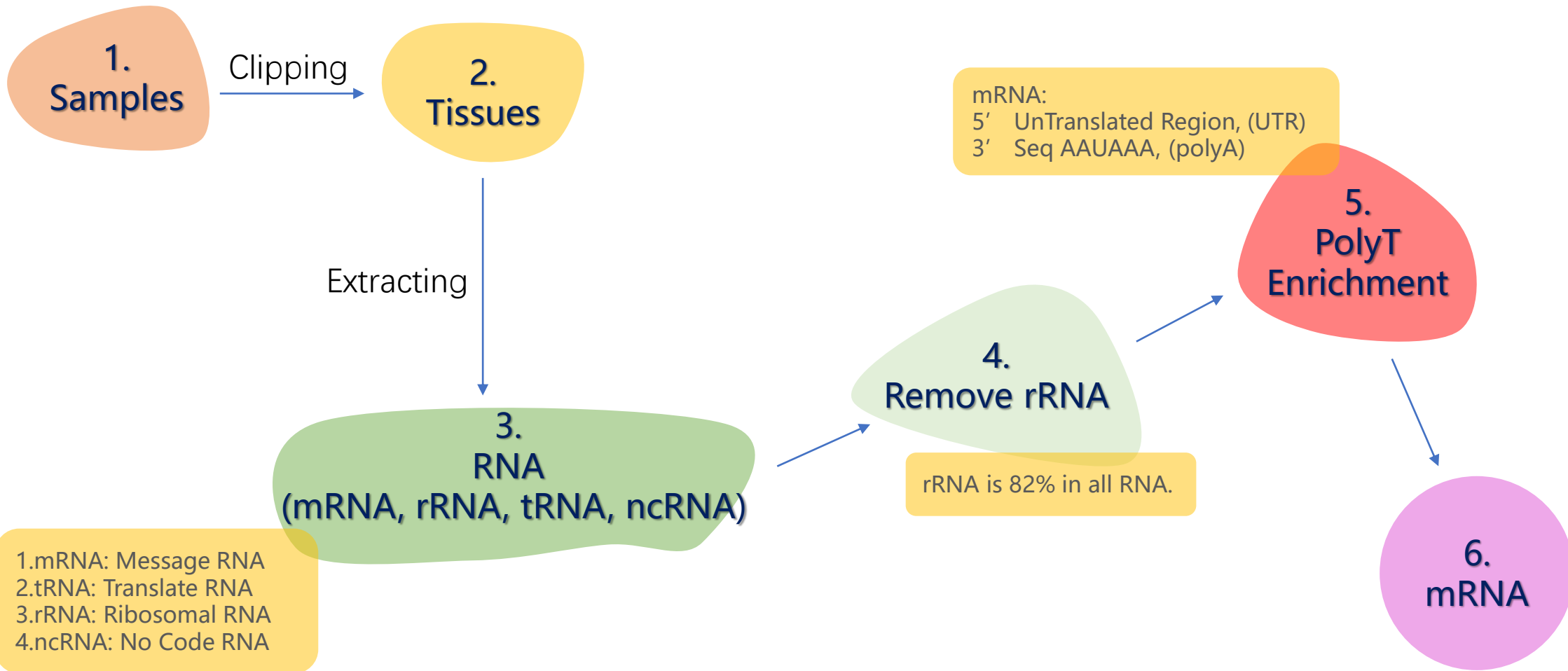
ceRNA-Seq(competing endogenous RNAs)

circRNA-Seq(Circular RNA)

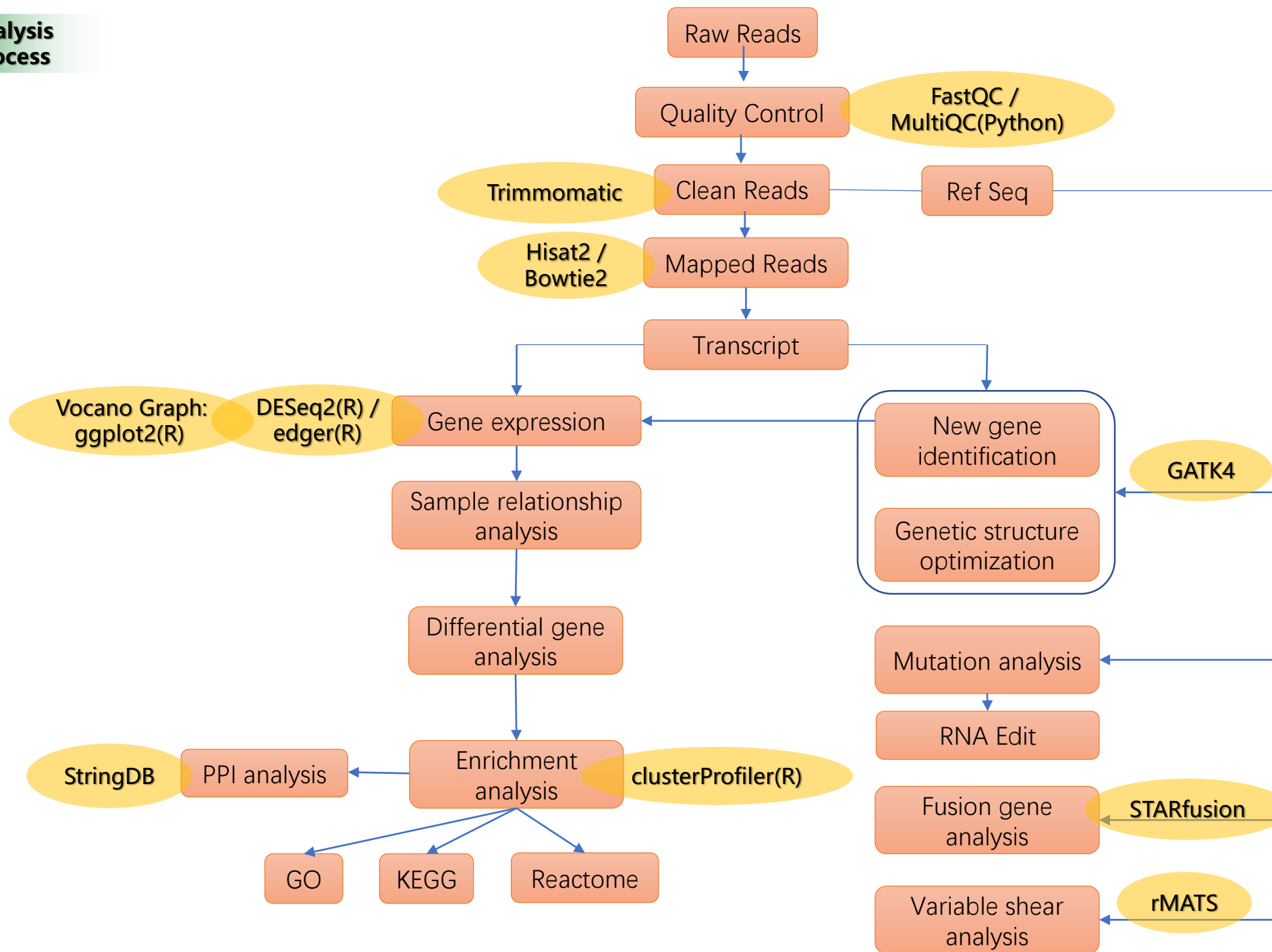


Experimental design

Eukaryotic reference transcriptome



Analysis Process



FastQC (Quality Control) : <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Hisat2(Map to Complete Genome): <http://www.psc.edu/user-resources/software/hisat2>

Bowtie2(Map to Complete Genome): <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>

STRING(Interaction of genes or proteins): <https://string-db.org/>

R(A beautiful data computer and visual language): <https://www.r-project.org/>

Rstudio(IDE for R): <https://rstudio.com/>

DESeq2(Differential Expression Sequences): <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>

clusterProfiler(GO and KEGG): <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>

3. Microorganism

16S, 18S, ITS Sequences

"S" (settling speed) represents a measure of the settling speed, not the mass. The sedimentation speed of each subunit is affected by both its shape and its mass.

Prokaryotic ribosomes have a total of 70S containing 23S, 16S and 5S.

Eukaryotic ribosomes have a total of 80S, including 28S, 18S, 5.8S and 5S.

Metagenomic Sequencing

Metagenomic sequencing is a method that uses high-throughput sequencing technology to complete the detection and functional analysis of the genomes of all species in the microbial community.

Microbial transcriptome sequencing

Microbial transcriptome sequencing technology is a method for sequencing single colony microbial transcripts.

Experimental design

16S Seq Metagenomics

1. Microbial community

2. DNA

3. DNA Library

4. Sequences

5. Bioinformatics Analysis

Concentration $\geq 5\text{ng}/\mu\text{L}$
Total $\geq 150\text{ng}$

PCR amplify 16S
Add adapter

Ion S5 XL Platform
NovaSeq 6000 Platform
PE250/SE400

1. Microbial community

2. DNA

3. DNA Library

4. Sequences

5. Bioinformatics Analysis

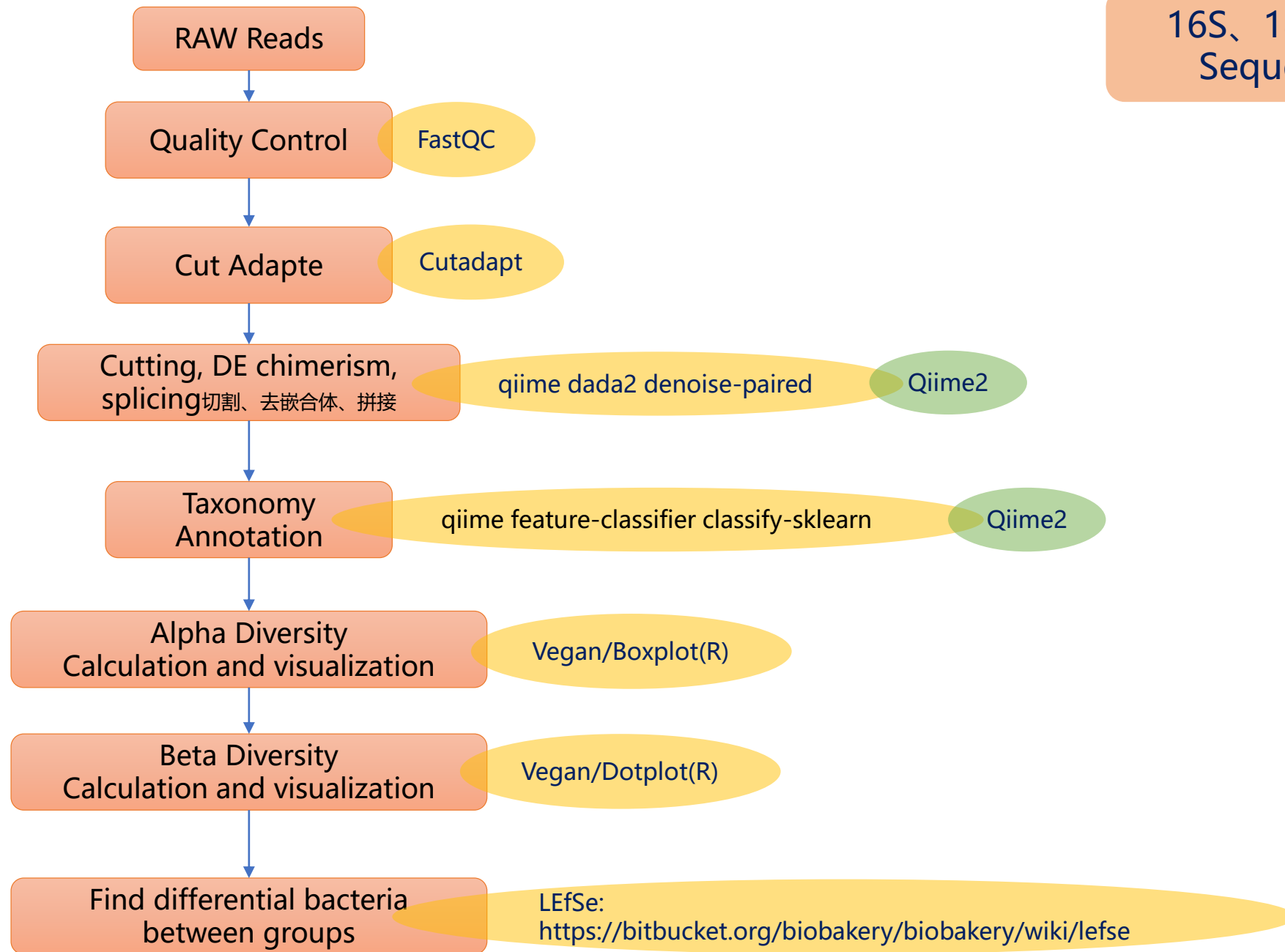
Total $\geq 200\text{ng}$

1. DNA Fragments
2. Add adapter to two ends
3. 350bp library

Illumina Sequences platform
PE150

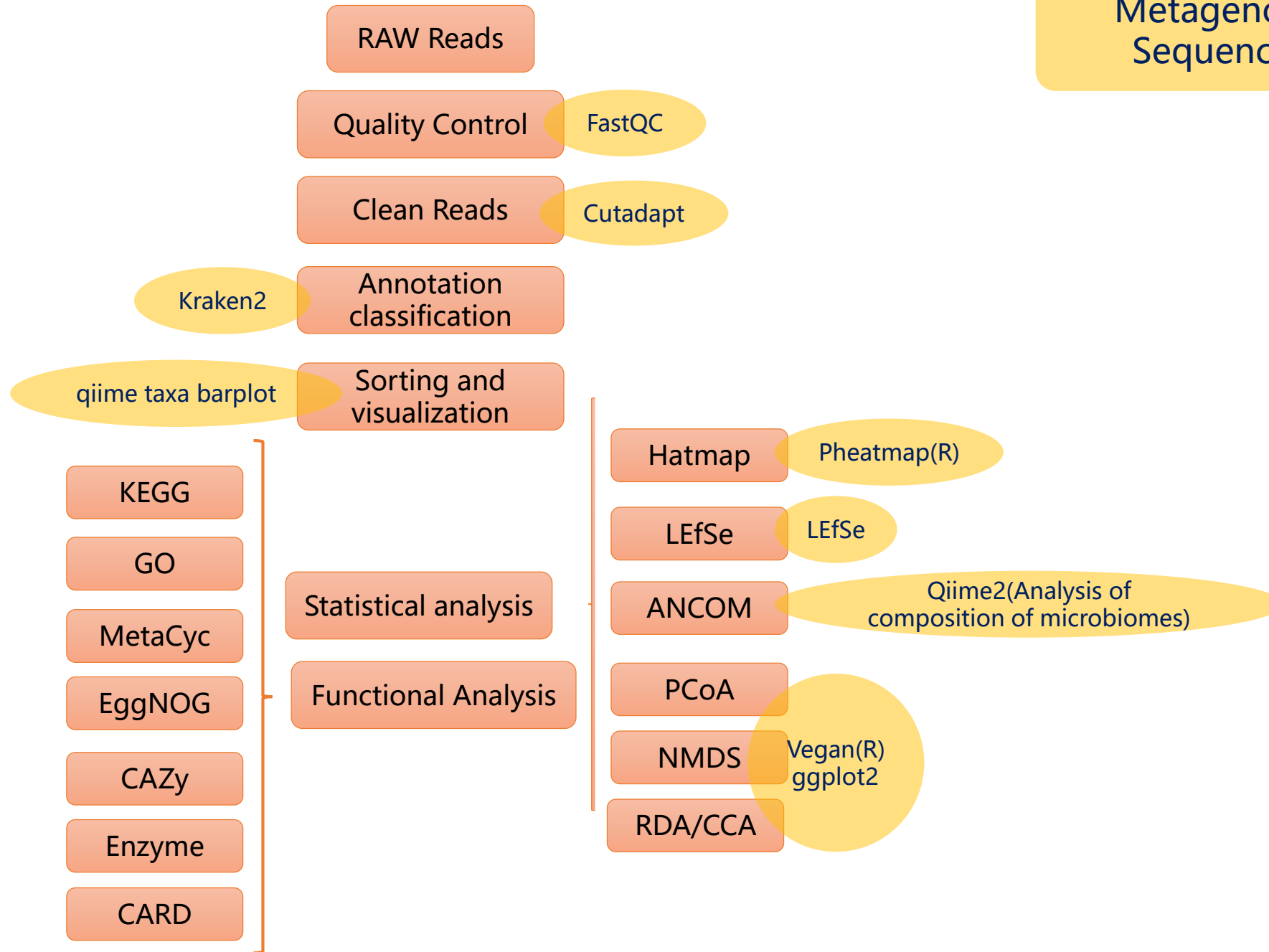
**Analysis
Process**

16S、18S、ITS
Sequences



Analysis Process

Metagenomic Sequencing



FastQC (Quality Control) : <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Cutadapt (Remove adapt at reads) : <https://cutadapt.readthedocs.io/en/stable/>

Qiime2 (microbiome bioinformatics platform) : <https://qiime2.org/>

LEfSe (Difference analysis) : <https://bitbucket.org/biobakery/biobakery/wiki/lefse>

Vegan (Statistics) : <https://www.rdocumentation.org/packages/vegan/versions/2.4-2>

Pheatmap (Heatmap) : <https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap>

kEGG (Kyoto Encyclopedia of Genes and Genomes): <https://www.kegg.jp/>

GO (Gene Ontology) : <http://geneontology.org/>

MetaCyc (Database of metabolic pathways): <https://metacyc.org/>

EggNOG (Clusters of Orthologous Groups of proteins) : <http://eggnogdb.embl.de/>

CAZy (Enzyme) : <http://www.cazy.org/>

ENZYME (Enzyme) : <https://enzyme.expasy.org/>

4. Proteomics

Qualitative proteome

1. Identification of Protein Glue Strips
2. Identification of protein interactions
3. Proteome full spectrum identification

Quantitative Proteome

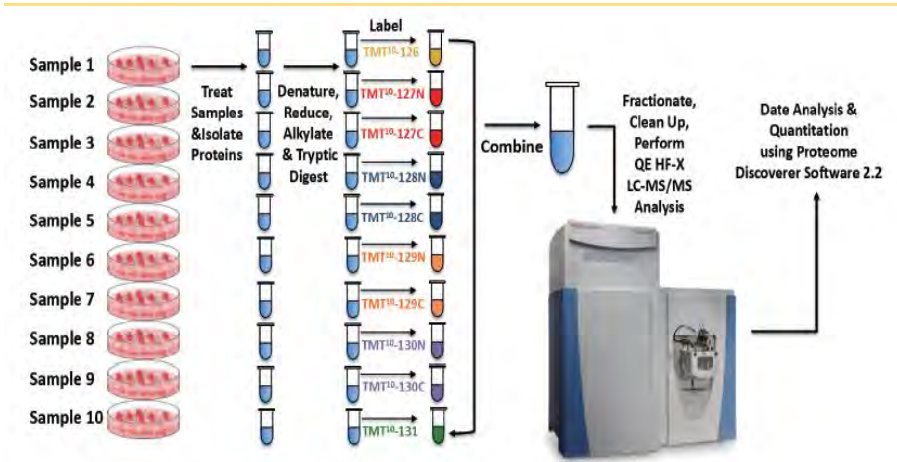
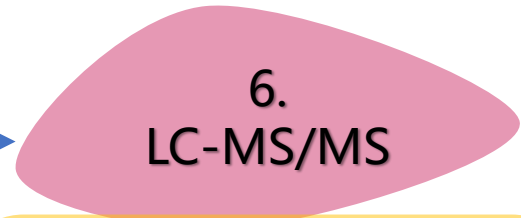
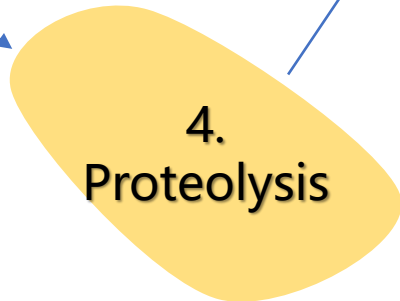
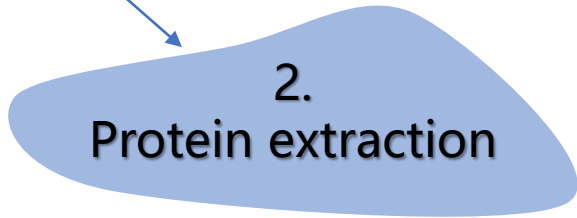
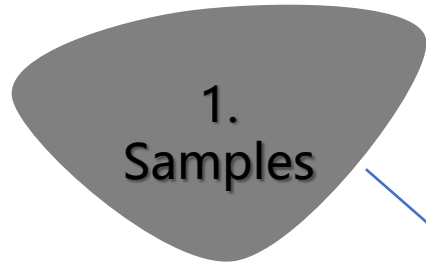
1. DDA (Data Dependent Acquisition): Label-free, **iTRAQ/TMT**
2. DIA (Data Independent Acquisition): DIA, SWATH

Modified proteome

1. Phosphorylated (磷酸化) proteome
2. Acetylated (乙酰化) protein
3. Ubiquitinated (泛素化) proteome
4. N-glycosylated (糖基化) protein
5. Succinylated (琥珀酰化) protein
6. Propionylated (丙酰化) proteome

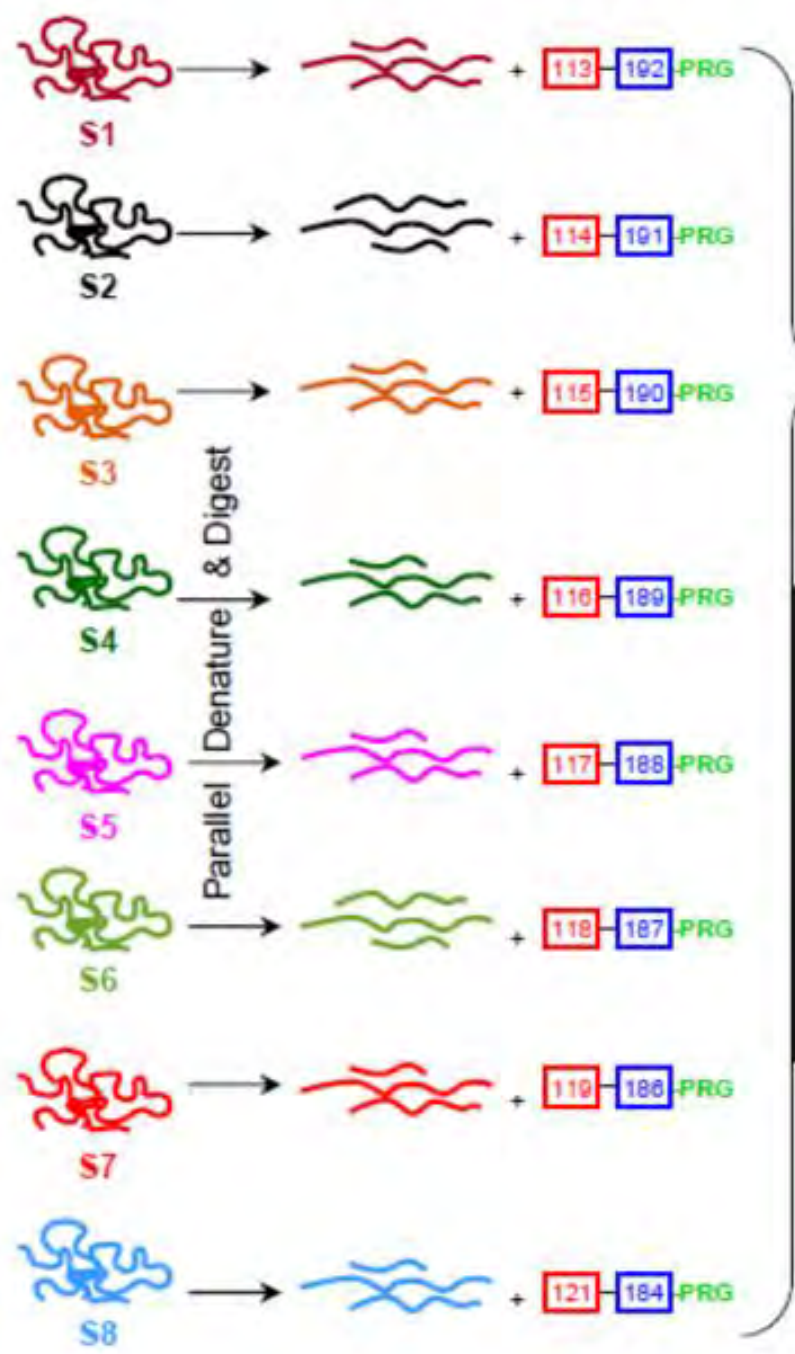
Experimental design

Quantitative Proteome

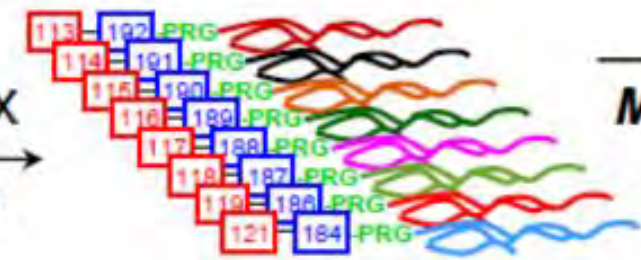


iTRAQ: Isobaric Tags for Relative and Absolute Quantitation
TMT: Tandem Mass Tags

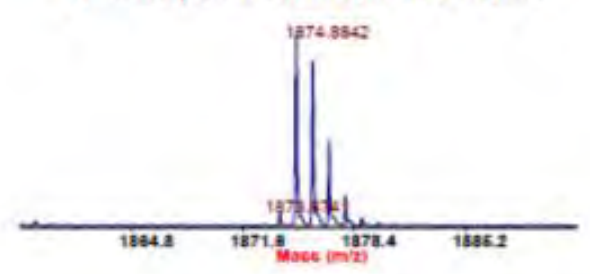
1.LC-MS/MS:
Liquid Chromatography - Mass Spectrometry
2.TOF: Time of Flight
3. ESI: Electron Spray Ionization
4.MALDI:
Matrix-assisted Laser Desorption/ionization



Mix
MS

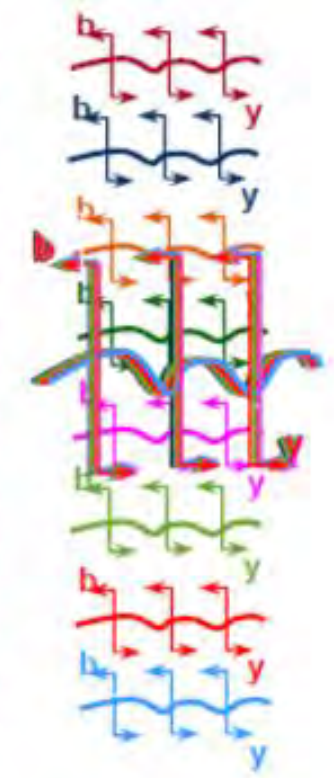


-Reporter-Balance-Peptide INTACT
- 8 samples *identical m/z*

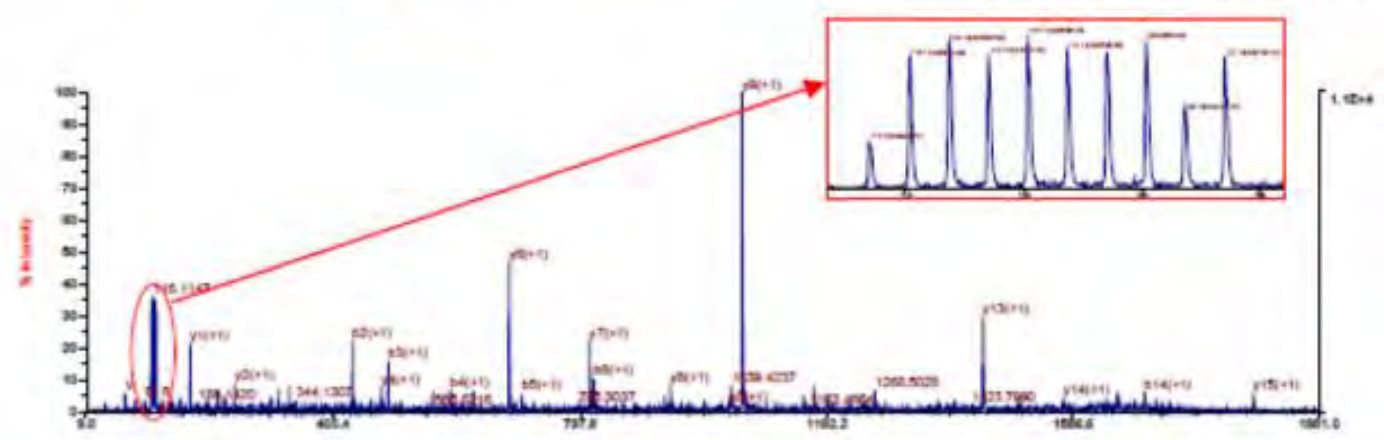


MS/MS

- [113]
- [114]
- [115]
- [116]
- [117]
- [118]
- [119]
- [121]



- Reporter ions DIFFERENT
- Peptide fragments EQUAL



Analysis Process

Raw File

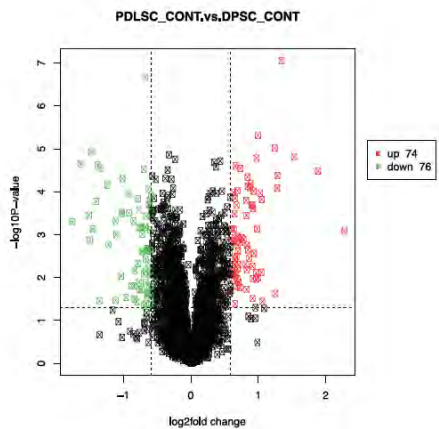
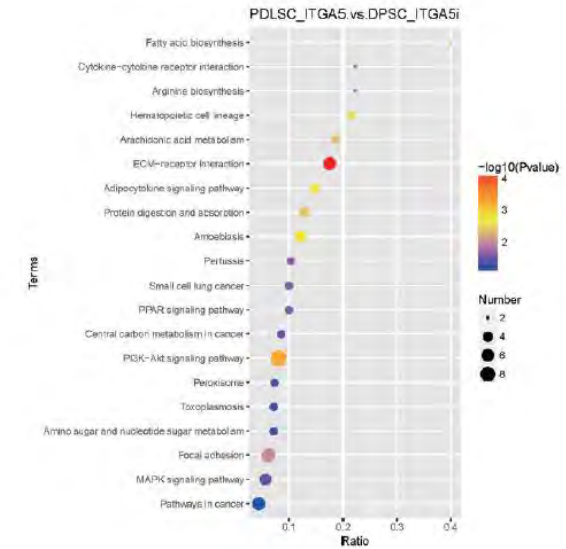
Proteome Xchange:
<http://www.proteomexchange.org/>

Protein DB Search

MaxQuant(offline)/
SEQUEST /
MASCOT(offline/online)

Vocano Graph:
ggplot2(R)

Different expression
Proteins



Protein
Function Annotation

GO Anotation

GO Graph:
barplot/dotplot/ggplot2(R)

KEGG Anotaion

KEGG Graph:
barplot/dotplot/ggplot2(R)

COG Anotation

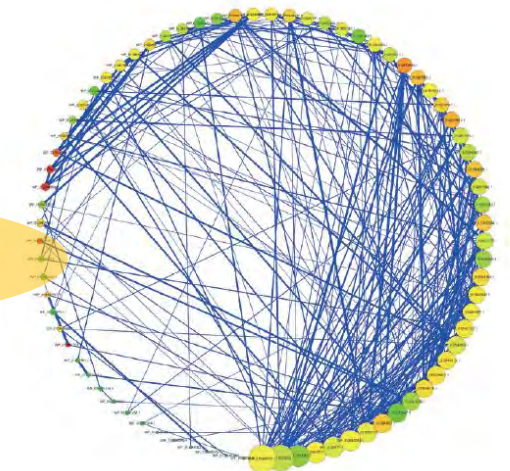
Protein structure
domains predictor

SMART:
<https://smart.embl-heidelberg.de/>

Interaction Network

STRING:
<https://string-db.org/>

Proteom analysis
With transcriptom



Mascot (Protide search engineer): <http://www.matrixscience.com/>

Uniprot (Most Protein Database): <https://www.uniprot.org/>

Maxquant (Protide search engineer): <https://www.maxquant.org/>

SEQUEST: <http://www.thermo.com>

pFind (Search protein): <http://www.pfind.net/>

ExPASy(Expert Protein Analysis): <https://www.expasy.org/>

STRING(Interaction of genes or proteins): <https://string-db.org/>

R(A beautiful data computer and visual language): <https://www.r-project.org/>

Rstudio(IDE for R): <https://rstudio.com/>

5. Metabolomics

Untargeted metabolomics

Qualitative and quantitative analysis of all small molecule compounds (molecular weight less than 1000) in biological samples.

Targeted metabolomics

Targeted metabolomics targeted analysis of a single or a small amount of target metabolites in biological samples, and the absolute abundance information of target metabolites obtained by comparison analysis with standards.

Broadly targeted metabolome

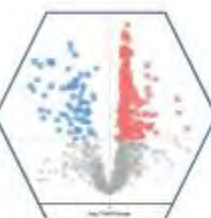
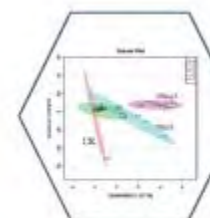
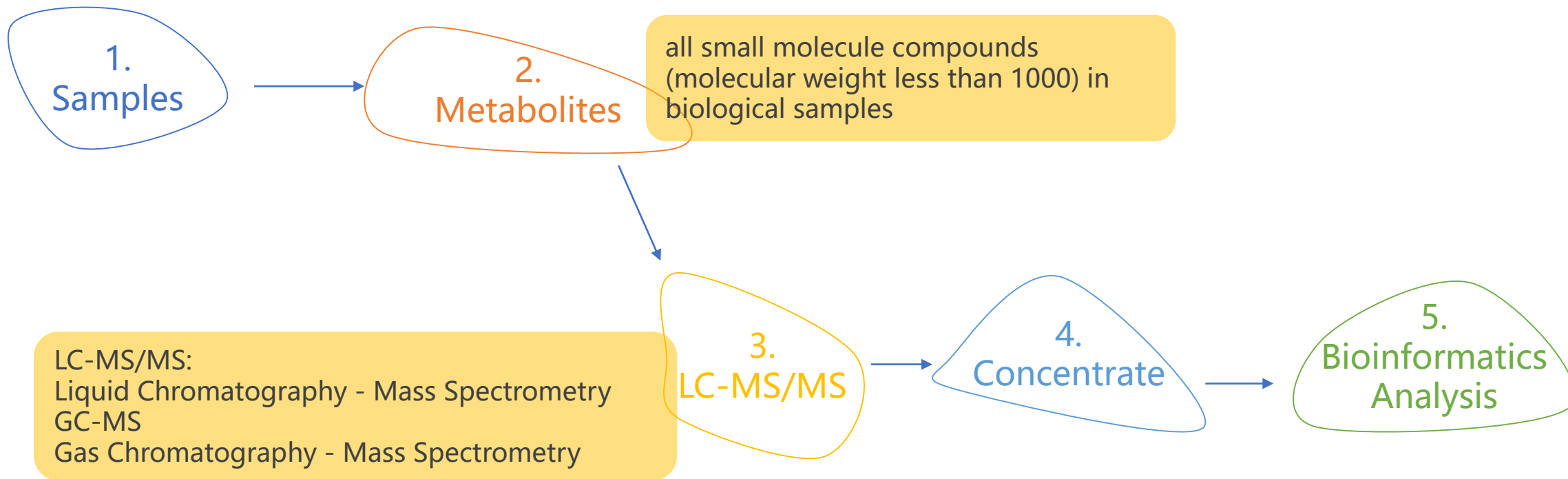
Combines advantages of non-targeted and targeted metabolite detection technology.

AB Sciex QTRAP® 4500 LC-MS/MS

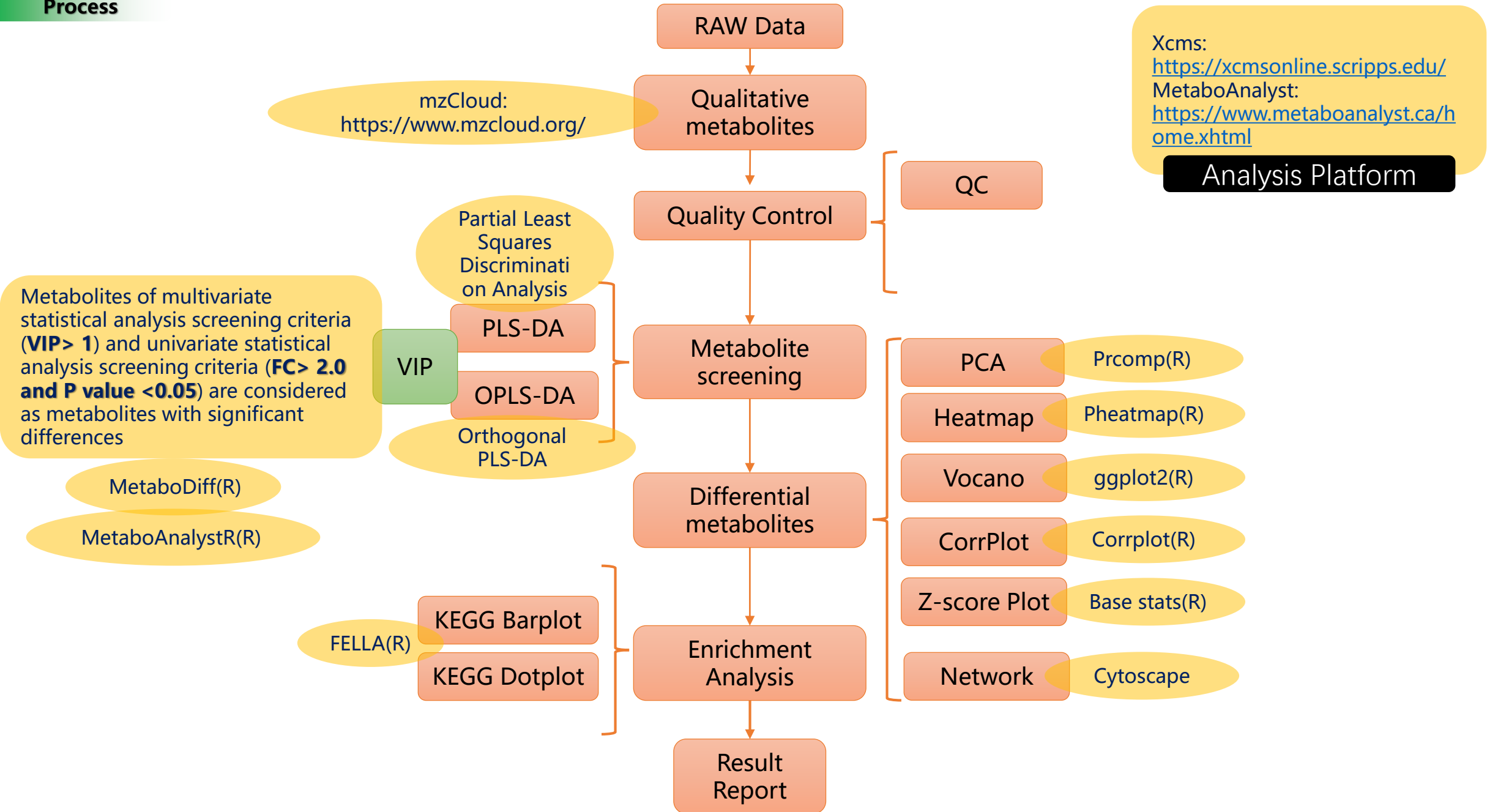
Database of primary and secondary metabolites (more than 2,200 substances)
Medical metabolite database (more than 1500 substances)

Experimental design

Untargeted metabolomics



Analysis Process



MetaboAnalyst (statistical, functional and integrative analysis of metabolomics data)

: <https://www.metaboanalyst.ca/home.xhtml>

XCMX: <https://xcmsonline.scripps.edu/>

R : <https://www.r-project.org/>

MetaboDiff : <https://github.com/andreamock/MetaboDiff/>

MetaboAnalystR : <https://github.com/xia-lab/MetaboAnalystR>

Cytoscape : <https://cytoscape.org/>

FELLA (Interpretation and enrichment for metabolomics data)

: <http://www.bioconductor.org/packages/release/bioc/html/FELLA.html>

Prcomp (Principle component analysis)

: <https://www.rdocumentation.org/packages/twidlr/versions/0.0.0.9000/topics/prcomp>

MetaCyc : <https://metacyc.org/>

BioCyc : <https://biocyc.org/>

Corrplot : <https://cran.r-project.org/web/packages/corrplot/index.html>

mzCloud: <https://www.mzcloud.org/>

ChIP-Seq
Chromatin immunoprecipitation

BS-Seq
Bisulfite Sequencing

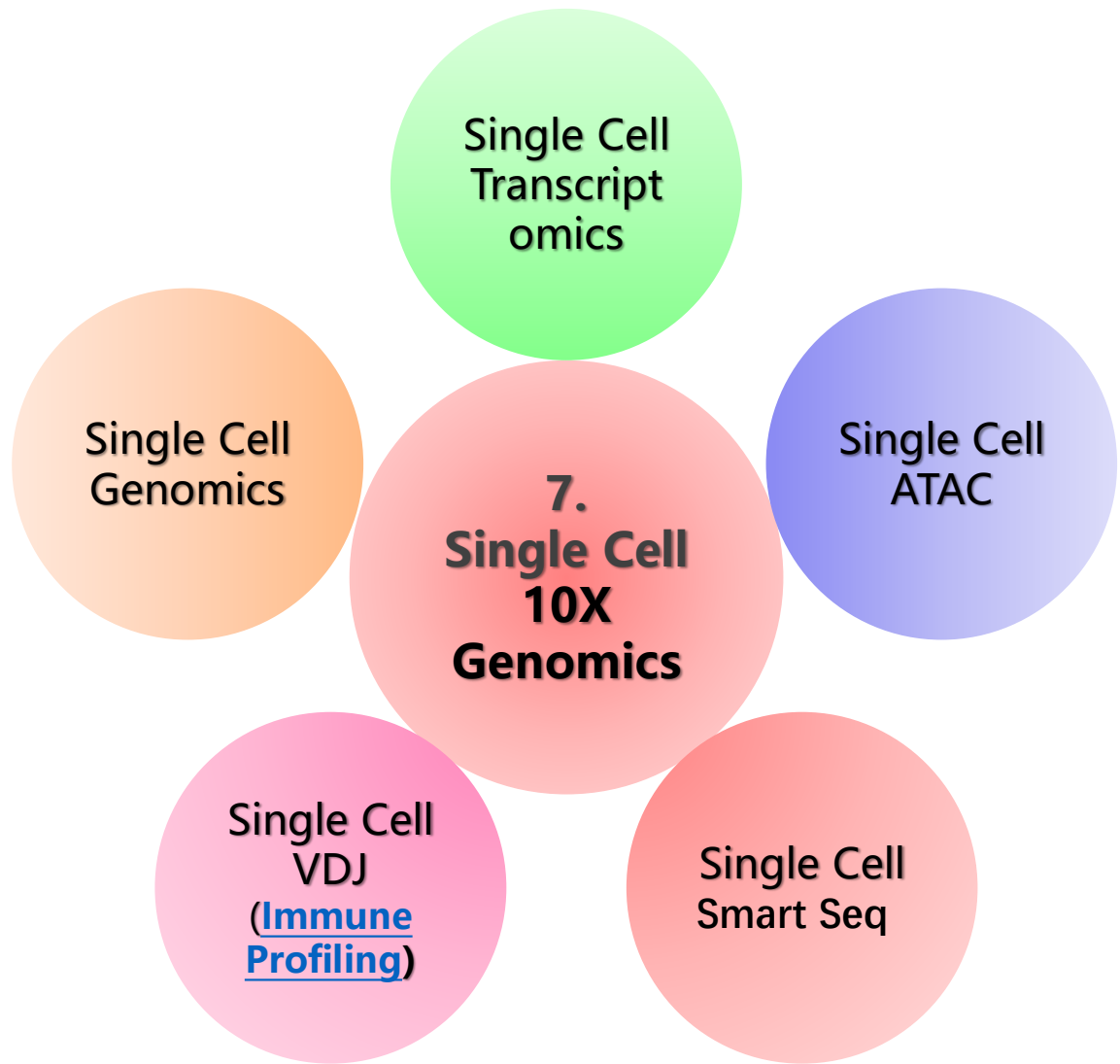
ATAC-seq
Assay for Transposase-Accessible Chromatin

RRBS-Seq
Reduced Representation Bisulfite Sequencing

**6.
Epigenomics**

RIP-seq
RNA Binding Protein Immunoprecipitation Assay

MeRIP m⁶A-Seq
Methylated RNA Immunoprecipitation



Single Cell
Transcript
omics

Single Cell
Genomics

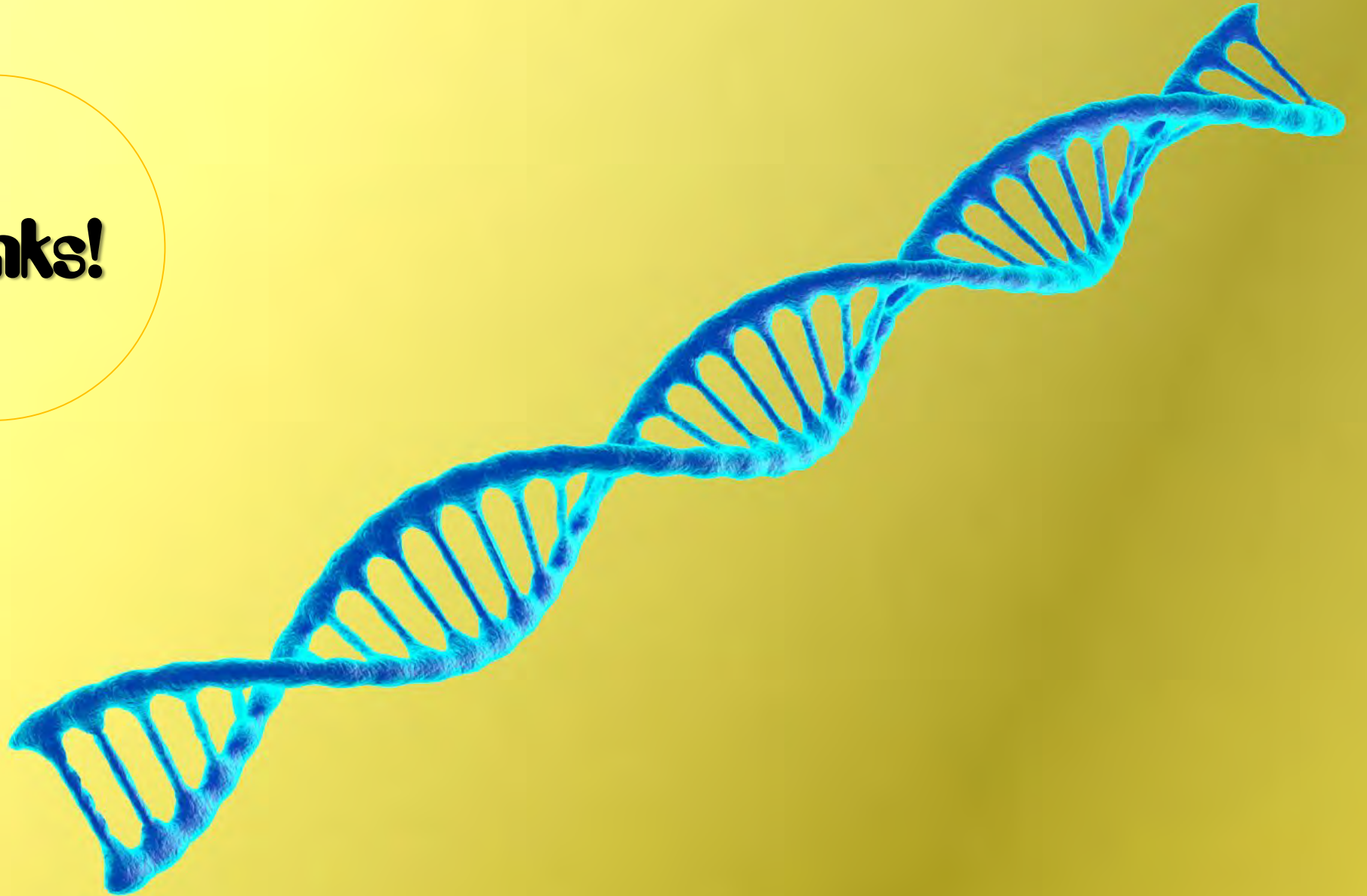
Single Cell
ATAC

7.
Single Cell
10X
Genomics

Single Cell
VDJ
(Immune
Profiling)

Single Cell
Smart Seq

Thanks!



Designer: Benben Miao
Copyright @benben-miao