1. Bioinformatics Analysis

1.1 Data Quality Control

|  |
| --- |
| **Data Quality Control** |
| Distribution of Sequencing Error Rate |
| Distribution of A/T/G/C Base Content |
| Data Filtering: Removal of reads with adaptor contamination, uncertain nucleotides(N)with > 10% content and nucleotides of low quality |
| Statistic Summary of Data Quality: Amount of data output, Error rate, Q20%, Q30% and GC content, etc |

1.2 Mapping Only Analysis

|  |
| --- |
| **Only Mapping Analysis** |
| Distribution of Sequencing Error Rate |
| Distribution of A/T/G/C Base Content |
| Data Filtering: Removal of reads with adaptor contamination, uncertain nucleotides(N)with > 10% content and nucleotides of low quality |
| Statistic Summary of Data Quality: Amount of data output, Error rate, Q20%, Q30% and GC content, etc |
| Mapping Clean Reads to Reference Genome |

1.3 Advanced PDX Analysis (For Medical Species Only)

|  |
| --- |
| **Advanced PDX Filter Analysis (Workflow A)** |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality |
| Mapping PDX Reads and Control Reads to **Human** Reference Genome Separately |
| Human **Standard** Analysis/Human **Quantification** Analysis |
| **Filter A**: Removal of Differential Expression and/or Variation of Control |

|  |
| --- |
| **Advanced PDX Filter Analysis (Workflow B)** |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality |
| Mapping PDX Reads to **Mouse** Reference Genome |
| **Filter B:** Removal of Mouse Reads |
| Mapping Reads to **Human** Reference Genome |
| Human **Standard** Analysis/Human **Quantification** Analysis |

|  |
| --- |
| **Advanced PDX Filter Analysis (Workflow C)** |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality |
| Mapping Reads to **Combine**d Reference Genome |
| **Filter C**: Picking up **Human** Mapped Reads |
| Human **Standard** Analysis/Human **Quantification** Analysis |

1.4 Quantification Analysis

1.4.1 Quantification Analysis for Medical Species

|  |  |
| --- | --- |
| **Quantification Analysis (human/mouse[Medical Species])** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| Mapping Clean Reads to Reference Genome | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Disease Ontology (**DO**) Enrichment Analysis (**Only for Human samples**) |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| Reactome Pathway Enrichment Analysis |
| DisGeNET Enrichment Analysis of Differentially Expressed Genes (DEGs) (two or more groups of samples and only for human samples) |
| GSEA Enrichment Analysis of Expressed Genes (two or more groups of samples) |
| Oncogene Functional Annotation analysis |
| Protein Protein Interaction Analysis |

1.4.2 Quantification Analysis for Agricultural Species

|  |  |
| --- | --- |
| **Quantification Analysis (Species with Reference[Agricultural Species])** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| Mapping Clean Reads to Reference Genome | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| GSEA Enrichment Analysis of Expressed Genes (two or more groups of samples) |
| Protein Protein Interaction Analysis |

1.4.3 Quantification Analysis for Species without Reference

|  |  |
| --- | --- |
| ***De novo* Quantification Analysis (Species without Reference)** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| *De novo* Transcriptome Assembly | |
| Gene Functional Annotation | |
| GO, COG, KEGG Classification | |
| Coding Sequence (**CDS**) Prediction | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| Protein Protein Interaction Analysis |

1.5 Standard Analysis

1.5.1 Standard Analysis for Medical Species

|  |  |
| --- | --- |
| **Standard Analysis (human/mouse[Medical Species])** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| Mapping Clean Reads to Reference Genome | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Disease Ontology (**DO**) Enrichment Analysis (**Only for Human samples**) |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| Reactome Pathway Enrichment Analysis |
| DisGeNET Enrichment Analysis of Differentially Expressed Genes (DEGs) (two or more groups of samples and only for human samples) |
| GSEA Enrichment Analysis of Expressed Genes (two or more groups of samples) |
| Oncogene Functional Annotation analysis |
| Protein Protein Interaction Analysis |
| Structural Analysis | Alternative Splicing (**AS**) Quantification and Differential Expression Analysis (**Only for Grouping with control samples**) |
| SNP/InDel Detection and Annotation Analysis |
| Fusion Gene Analysis (**Only for Grouping with TUMOR or CANCER tissue/cell line**) |

1.5.2 Standard Analysis for Agricultural Species

|  |  |
| --- | --- |
| **Standard Analysis (Species with Reference[Agricultural Species])** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| Mapping Clean Reads to Reference Genome | |
| Novel Gene Prediction | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| GSEA Enrichment Analysis of Expressed Genes (two or more groups of samples) |
| Protein Protein Interaction Analysis |
| Structural Analysis | Alternative Splicing (**AS**) Quantification and Differential Expression Analysis (**Only for Grouping with control samples**) |
| SNP/InDel Detection and Annotation Analysis |

1.5.3 Standard Analysis for Species without Reference

|  |  |
| --- | --- |
| ***De novo* Standard Analysis (Species without Reference)** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| *De novo* Transcriptome Assembly | |
| Gene Functional Annotation | |
| GO, COG, KEGG Classification | |
| Coding Sequence (**CDS**) Prediction | |
| Structural Analysis | SNP/InDel Detection and Annotation Analysis |
| Simple Sequence Repeats (SSR) Analysis |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| Protein Protein Interaction Analysis |

Note: there might be multiple Data Quality Control presented in the tables which does not mean repetition of quality control; regularly data quality control is considered to be one-time required.