Planning project Advanced Bioinformatics

Group Mango

1. Data format: platform, paired end, size (dimensions)
2. Quality control: statistics of input data (fastqc)
3. Quality improvement
4. Mapping – Bowtie(2), BWA
5. PICARD: remove PCR duplicates (overrepresented)
6. Mapping quality: IGV view
7. Count reads – Tophat cufflinks
8. Data normalization – RPKM/FPKM/Trimmed mean
9. Missing values – KNN imputation
10. Differential Expression – 2 conditions –Genes DeSeq or Numpy
11. Enrichment
12. Biological explanation
13. Coexpression
14. Gene interactions (MEME) – transcription factor binding sites

MoSCow

* MUST
  + Data format
  + Quality control FASTQC  **GEERT**
  + Quality improvement FASTX **GEERT**
  + Mapping TOPHAT **NIKOS**
  + Count reads CUFFLINKS **ADITHI**
  + Data normalization CUFFLINKS **ADITHI**
  + Differential expression CUFFDIFF **MARCEL**
  + Enrichment **NICK**
* SHOULD
  + Chose appropriate software for:
    - Mapping
    - Count reads
    - Normalization
  + Mapping quality
  + Enrichment
* COULD (future)
  + Aligner: Stampy as comparison
  + Differential expression: EdgeR as comparison
  + De novo assembly
  + Missing values
  + Coexpression (?)
  + Networks
  + Motifs

|  |  |
| --- | --- |
| Monday | * Divide tasks * Find example file per subpart |
| Tuesday | * Check together FASTQC output and define thresholds * **Planning presentation** * Scripting |
| Wednesday | * Scripting |
| Thursday | * Scripting |
| Friday | * **Everybody finished their part** * **Progress presentation** |
| Monday | * Clubbing * Bug fixing |
| Tuesday | * Biological inference/writing |
| Wednesday | * Biological inference/writing |
| Thursday | * Add ‘coulds’ * Make powerpoint/writing * Have a final report ready |
| Friday | * **Powerpoint presentations** |

Every working day will be from 9:00 to 17:00

From 9:00 to 9:30. Meetings are making a round in the group. Everybody tells what they have done, and tell their day planning.