

BioProject: Differentially regulated genes in TOR knockdown Arabidopsis lines

Rapamycin-sensitive transgenic Arabidopsis lines (BP12) expressing yeast FK506 Binding Protein12 (FKBP12) were developed. Inhibition of TOR in BP12 plants by rapamycin resulted in slower overall root, leaf and shoot growth and development leading to poor nutrient uptake and light energy utilization. Genetic and physiological studies together with RNA-Seq and metabolite analysis of TOR-suppressed lines revealed that TOR regulates development and lifespan in Arabidopsis by restructuring cell growth, carbon and nitrogen metabolism, gene expression, ribosomal RNA and protein synthesis. Overall design: Arabidopsis WT (Col) and BP12-2 (TOR knockdown line) seedlings at 15 DAG were treated with rapamycin for 3 days by transferring from 0.5 MS medium to 0.5 MS+10 µg/ml rapamycin. Triplicate samples of rapamycin treated WT and BP12-2 seedlings were used for RNA-Seq analysis (Illumina Hiseq 2000)

Dados

Os arquivos de *reads* estão no diretório:

```
/data/userXX/libs_trimmed
```

São seis arquivos no formato fastq, sendo as bibliotecas SRR634969, SRR634970, SRR634971 referentes à “BP12-2”, e as bibliotecas SRR634972, SRR634973, SRR634974 são referente à “WT”.

O genoma de referência de Arabidopsis está no diretório:

```
/data/homedir/genoma/Athaliana_167_TAIR9.fa
```

A anotação dos transcritos conhecidos de Arabidopsis está no diretório:

```
/data/homedir/genoma/Athaliana_167_TAIR10.gene.gff3
```

1 – Montagem usando o genoma de referência

Objetivos:

1. Utilizar o genoma de referência de *Arabidopsis thaliana* e os seus transcritos conhecidos para fazer a montagem dos dados de RNA-Seq. A vantagem dessa metodologia é que ela permite a descoberta de novas variantes de splicing e novos genes (codificantes e não codificantes).
2. Calcular a expressão dos transcritos da montagem.
3. Identificar os transcritos diferencialmente expressos.

Pipeline

1. Criar diretório com o nome “montagem_referencia” no seu home.
2. Entrar dentro do diretório “montagem_referencia” e criar um diretório com o nome “tophat”
3. Entrar dentro do diretório “tophat” e para cada uma das seis bibliotecas realizar o alinhamento dos *reads* de RNA-Seq no genoma de Arabidopsis.

Dicas e Sugestões

- Não se esqueça de mudar o nome do diretório de saída para cada alinhamento.
- Rode os alinhamentos utilizando os comandos Linux “nohup” e “&”. Ex “nohup comando &”. Dessa forma o terminal é liberado e o processo não fica preso a sua sessão, assim é possível fechar o terminal sem que o processo seja perdido.
- Todos os alinhamentos vão demorar umas 12 horas para terminar, então tentem deixar todos rodando antes de ir embora.
- Por questão de tempo os seguintes parâmetros são de uso obrigatório:

--no-coverage-search

--b2-very-fast

4. Sair do diretório tophat, criar o diretório cufflinks e entrar dentro dele. Realizar uma montagem com o software cufflinks para cada biblioteca utilizando os alinhamentos da etapa anterior.

Dicas e Sugestões

- Não se esqueça de mudar o nome do diretório de saída para cada alinhamento.
- Todas as montagens vão levar em torno de duas horas para rodar, então tentem deixar todas as montagens rodando.
- Rode os alinhamentos utilizando os comandos Linux “nohup” e “&”. Ex “nohup comando &”. Dessa forma o terminal é liberado e o processo não fica preso a sua sessão, assim é possível fechar o terminal sem que o processo seja perdido.

5. Juntar as montagens utilizando o software cuffmerge
6. Comparar as montagens utilizando o software cuffcompare
7. Criar o arquivo fasta da montagem utilizando o software gtf_to_fasta
8. Calcular a expressão dos transcritos da montagem utilizando o software kallisto

Dicas e Sugestões

- Primeiramente crie o índice da referência (arquivo fasta com a montagem dos transcritos).
- Não se esqueça de mudar o nome do diretório de saída para cada biblioteca.
- Rode o kallisto utilizando os comandos Linux “nohup” e “&”. Ex “nohup comando &”. Dessa forma o terminal é liberado e o processo não fica preso a sua sessão, assim é possível fechar o terminal sem que o processo seja perdido.

9. Criar a tabela com os read counts utilizando o script “roda_cat.sh”.

Dicas e Sugestões

- Você deve passar para esse script os arquivos de expressão de cada biblioteca. Ex:

```
rodaCat.sh SRR634969/abundance.tsv SRR634970/abundance.tsv  
SRR634971/abundance.tsv SRR634972/abundance.tsv SRR634973/abundance.tsv  
SRR634974/abundance.tsv > read_counts.txt
```

- Ele irá criar no seu diretório atual um arquivo chamado “read_counts.txt” que tem o formato tabular, sendo a primeira coluna o nome do transcrito e as demais colunas os reads counts para cada uma das bibliotecas.
- A ordem das bibliotecas no arquivo de saída é a mesma dos argumentos.

10. Identificar os genes diferencialmente expressos utilizando o pacote em R “DESeq2”.

Dicas e Sugestões

- Fazer a análise de expressão no computador local.
- Para copiar o arquivo readcounts.txt para a máquina local, baixe e utilize o programa WinSCP:

(<https://winscp.net/download/winscp574.zip>)

- Primeiramente fazer um PCA e/ou HEATMAP das bibliotecas
- Seguir o manual do DESeq2 para fazer a análise e criar um arquivo no formato csv com os resultados.