

Aula prática: RNAseqII. Explorando dados de NGS.

06/07/2017

Professor: Jorge Estefano Santana de Souza, jorge@imd.ufrn.br;

Monitores: Danilo Lopes Martins, danilolmartins@gmail.com;
Luan Pereira, luanpereira00@outlook.com.

Objetivos:

Utilizar as ferramentas básicas de RNAseq para obter o padrão de expressão dos mirRNAs de uma amostras.

Ferramentas:

- 1- Linux.
- 2- WebServer.
- 3- cutadapt
- 4- mapper
- 5- miRDeep2

Comandos Básicos:

Durante a execução dos tutoriais necessitaremos saber alguns comandos básicos do Linux. Podem procurar mais informação no site:

<http://wiki.ubuntu-br.org/ComandosBasicos>

Login Servidor:

Inicialmente vamos fazer o login no servidor, abra um terminal no linux e digite:

```
ssh -p 4422 bif@10.7.5.38
```

Irá pedir uma senha, digite:

```
bif0003
```

*ps. não aparece a digitação, o teclado não quebrou não!

Regras para login no servidor:

Interno à UFRN:

```
ssh -p 4422 bif@10.7.5.38
```

Senha: bif0003

Externo à UFRN:

```
ssh -p 4422 bif@177.20.147.141
```

Senha: bif0003

Dados brutos (raw data):

Durante a execução dos tutoriais necessitaremos de alguns dados iniciais, em via de regra estarão disponíveis no diretório:

```
/home/treinamento/NGS/RNAseq/
```

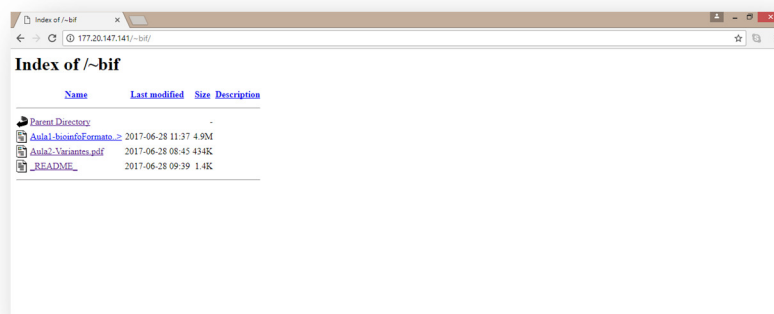
Servido WEB:

Como os trabalhos realizados no servidor são de difícil visualização, iremos necessitar de uma área web para facilitar nossa tarefa, todos os arquivos copiados para o diretório:

```
/home/bif/public_html/
```

Estarão disponíveis via navegador web em:

```
http://177.20.147.141/~bif/
```



Iniciando o Workflow:

1) Vamos começar pelo básico, certifique-se de que a pasta atual é:

```
/home/bif
```

Para isso digite o comando:

```
pwd
```

2) Crie um diretório contendo o seu nome, digite o comando:

```
mkdir SeuNome
```

3) Entre no diretório criado:

```
cd SeuNome
```

4) Crie um diretório chamado rna2:

```
mkdir rna2
```

5) Entre no diretório criado:

```
cd rna2
```

6) certifique-se de que a pasta atual é a correta:

```
pwd
```

O diretório atual deve ser: /home/bif/SeuNome/rna2

7) Crie links simbólicos para os arquivos:

```
ln -s /home/treinamento/NGS/RNAseq/sample_data/SRR326279_R1.fastq .  
ln -s /home/treinamento/NGS/RNAseq/sample_data/SRR326280_R1.fastq .  
ln -s /home/treinamento/NGS/RNAseq/sample_data/SRR326281_R1.fastq .  
ln -s /home/treinamento/NGS/RNAseq/sample_data/SRR326282_R1.fastq .
```

*ps. esses são os sequenciamentos de nossas amostras.

Arquivos de Referência:

8) Agora vamos necessitar de:

- Arquivo fasta com o genoma de referência;
- Arquivos de index do genoma de referência;
- Arquivo fasta com os miRNAs referência para a espécie (utilizaremos o miRBase);
- Arquivo fasta com os miRNAs maduros para a espécie (utilizaremos o miRBase);
- Arquivo fasta de predição dos loops das sequências dos miRNAs para a espécie, os hairpins (utilizaremos o miRBase);

Esses arquivos já foram baixados faça um link para eles:

```
ln -s /home/treinamento/NGS/RNAseq/small_ref/ .
```

Preparando o dado inicial:

9) Antes de executar o miRDeep2, os dados devem ser pré-processados para remover adaptadores. Isso pode ser feito usando o cutadapt: .

```
cutadapt -b AATCTCGTATGCCGTCTTCTGCTTGC -O 3 -m 17  
-f fastq SRR326279_R1.fastq > SRR326279_R1.ct.fastq
```

*ps. o comando deve ser digitado em apenas uma linha

```
cutadapt -b AATCTCGTATGCCGTCTTCTGCTTGC -O 3 -m 17  
-f fastq SRR326280_R1.fastq > SRR326280_R1.ct.fastq
```

```
cutadapt -b AATCTCGTATGCCGTCTTCTGCTTGC -O 3 -m 17  
-f fastq SRR326281_R1.fastq > SRR326281_R1.ct.fastq
```

```
cutadapt -b AATCTCGTATGCCGTCTTCTGCTTGC -O 3 -m 17  
-f fastq SRR326282_R1.fastq > SRR326282_R1.ct.fastq
```

Mapeamento:

10) Usaremos o script mapper.pl para processar as leituras e mapear-las contra o genoma de referência:

```
mapper.pl SRR326279_R1.ct.fastq
-e
-p small_ref/hg19_chr1
-s SRR326279.pr.fa
-t SRR326279.mr.arf
-h -m -i -j
```

*ps. o comando deve ser digitado em apenas uma linha

```
mapper.pl SRR326280_R1.ct.fastq
-e
-p small_ref/hg19_chr1
-s SRR326280.pr.fa
-t SRR326280.mr.arf
-h -m -i -j
```

*ps. o comando deve ser digitado em apenas uma linha

```
mapper.pl SRR326281_R1.ct.fastq
-e
-p small_ref/hg19_chr1
-s SRR326281.pr.fa
-t SRR326281.mr.arf
-h -m -i -j
```

*ps. o comando deve ser digitado em apenas uma linha

```
mapper.pl SRR326282_R1.ct.fastq
-e
-p small_ref/hg19_chr1
-s SRR326282.pr.fa
-t SRR326282.mr.arf
-h -m -i -j
```

*ps. o comando deve ser digitado em apenas uma linha

Rodando o miRDeep2:

Identificação de miRNAs conhecidos e novos nos dados de sequenciamento:

```
miRDeep2.pl SRR326279.pr.fa small_ref/hg19_chr1.fa
SRR326279.mr.arf small_ref/mature.hsa.dna.fa
none small_ref/hairpin.hsa.dna.fa
-t Human 2> report.log
```

*ps. o comando deve ser digitado em apenas uma linha

```
miRDeep2.pl SRR326280.pr.fa small_ref/hg19_chr1.fa
            SRR326280.mr.arf small_ref/mature.hsa.dna.fa
            none             small_ref/hairpin.hsa.dna.fa
            -t Human         2> report.log
```

*ps. o comando deve ser digitado em apenas uma linha

```
miRDeep2.pl SRR326281.pr.fa small_ref/hg19_chr1.fa
            SRR326281.mr.arf small_ref/mature.hsa.dna.fa
            none             small_ref/hairpin.hsa.dna.fa
            -t Human         2> report.log
```

*ps. o comando deve ser digitado em apenas uma linha

```
miRDeep2.pl SRR326282.pr.fa small_ref/hg19_chr1.fa
            SRR326282.mr.arf small_ref/mature.hsa.dna.fa
            none             small_ref/hairpin.hsa.dna.fa
            -t Human         2> report.log
```

*ps. o comando deve ser digitado em apenas uma linha

Agora vamos olhar os resultados.

Referências:

- 1- Simple Combinations of LineageDetermining Transcription Factors Prime cisRegulatory Elements Required for Macrophage and B Cell Identities. Heinz S, Benner C, Spann N, Bertolino E et al.
- 2- Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Cole Trapnell, Brian Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Jeltje van Baren, Steven Salzberg, Barbara Wold, Lior Pachter. *Nature Biotechnology*, 2010
- 3- Cutadapt removes adapter sequences from high-throughput sequencing reads. MARTIN, Marcel. **EMBnet.journal**, [S.l.], v. 17, n. 1, p. pp. 10-12, may. 2011. ISSN 2226-6089. Available at: <<http://journal.embnet.org/index.php/embnetjournal/article/view/200>>. Date accessed: 08 Jul. 2017. doi:<http://dx.doi.org/10.14806/ej.17.1.200>.
- 4- Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N. 'Discovering microRNAs from deep sequencing data using miRDeep', *Nature Biotechnology*, 26, 407-415 (2008)