



IQB-723 – BIOINFORMÁTICA: GENOMAS E TRANSCRIPTOMAS

Professor

Rafael Dias Mesquita (IQ-UFRJ) – rdmesquita@iq.ufrj.br

Carga horária: 30 horas

Disciplina teórico-prática

Créditos: 2

Vagas: 10

Objetivo

A disciplina tem por objetivo principal ensinar conceitos de Bioinformática relacionados à genômica e transcriptômica e treinar o aluno no uso de programas diversos para análises genômicas e transcriptômicas, especialmente dentro do ambiente web galaxy.

Ementa

Os principais tópicos do curso são: (i) Sequenciamento (tecnologia Illumina); (ii) Formatos de arquivos; (iii) Controle de qualidade de sequenciamento; (iv) Montagem de genomas e transcriptomas; (v) Controle de qualidade de montagens; (vi) Predição gênica em genomas; (vii) Análise de expressão gênica baseado em sequenciamento de transcriptoma.

Programa Analítico

1. Sequenciamento Illumina: introdução, preparação de bibliotecas (barcode, single x paired, comprimento e adaptadores) e tecnologia do sequenciamento Illumina.
PRÁTICA: Login no ambiente galaxy, operações básicas do galaxy, *download* de um genoma (procarioto) e de um transcriptoma (procarioto, com no mínimo 2 condições, com 2 replicatas cada) – Sequenciados por Illumina, transformação de formato SRA para fastq.
2. Controle de qualidade: formato de arquivos, métricas de controle de qualidade de sequenciamento, limpeza da sequência e reanálise de qualidade.
PRÁTICA: Controle de qualidade com o fastQC, limpeza com o cutadapt e reanálise de qualidade, todos em ambiente galaxy.
3. Montagem de transcriptomas: introdução, grafos De Bruijn, montagem de novo e baseada em genoma, software Trinity, polimorfismos (SNPs), paralogia e *splicing* alternativo.
PRÁTICA: Montagem de transcriptoma usando o Trinity e o Cuflinks, todos em ambiente galaxy.
4. Montagem de genomas de procariotos: introdução, heterozigotidade, correção de erros, grafos De Bruijn normais e duplos, software Spades.
PRÁTICA: Montagem de genoma usando o Spades, em ambiente galaxy.
5. Controle de qualidade de transcriptomas e genomas: métricas para genomas (cobertura,



N50, maior contig, histograma de tamanhos) e transcriptomas (cobertura, N50, maior transcrito, histograma de tamanhos, sequencias completas).

PRÁTICA: Controle de qualidade de montagens de genomas (Spades quality, assemblyStats e QAST web). Controle de qualidade de transcriptomas (assemblyStats e BUSCO).

6. Análise de expressão gênica: alinhamento baseado em genoma e transcriptoma, TopHat, Bowtie2, Regiões e alinhamentos repetitivos, RPKM e FPKM.

PRÁTICA: Alinhamento usando o software Bowtie2 dos reads do transcriptoma contra o transcriptoma montado previamente. Quantificação de reads por gene usando o eXpress.

7. Estatística relacionada e expressão gênica: determinação dos genes diferencialmente expressos usando o EdgeR (cobertura, variância, ruídos, distribuição e significância) e clusterização de perfis de expressão usando o Expander.

PRÁTICA: Determinação dos genes diferencialmente expressos usando o EdgeR. Clusterização de padrões de expressão gênica usando o Expander.

8. Predição gênica em genomas: regiões gênicas – início, fim, exons e introns, regiões promotoras, fim de transcrição, operons x genes, frames de tradução, estatística na predição gênica, HMMs.

PRÁTICA: Predição gênica com os softwares Prodigal e Glimmer usando o genoma montado previamente.

Literatura recomendada

Introdução e Galaxy

- Bioinformatics – A practical guide to the analysis of genes and proteins. Baxevanis, AD and Ouellette, BFF. Wiley. 3rd edition. 2005.
- Galaxy Project: <https://wiki.galaxyproject.org/Support> e <https://wiki.galaxyproject.org/Learn>
- SRA knowledgebase: <http://www.ncbi.nlm.nih.gov/books/NBK56551/>
- Developing bioinformatics computer skills, Gibas e Jambeck. O'Reilly. 2001.

Controle de qualidade e limpeza de dados de sequenciamento

- FASTQC: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/>
- Cutadapt: <http://cutadapt.readthedocs.org/en/stable/guide.html>

Montagem de Transcriptomas

- Trinity: <https://github.com/trinityrnaseq/trinityrnaseq/wiki>
- Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29(7):644-652. (<http://dx.doi.org/10.1038/nbt.1883>)
- Cufflinks: <http://cole-trapnell-lab.github.io/cufflinks/manual/>
- Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7(3):562-578. (<http://www.nature.com/nprot/journal/v7/n3/full/nprot.2012.016.html>)

Comparações e Revisões

- Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet



12(10): 671-682. (<http://www.nature.com/nrg/journal/v12/n10/full/nrg3068.html>)

- Clarke K, Yang Y, Marsh R, Xie L, Zhang KK (2013) Comparative analysis of de novo transcriptome assembly. Science China Life sciences 56(2):156-162. (<http://link.springer.com/article/10.1007%2Fs11427-013-4444-x>)
- Zhao QY, et al. (2011) Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12 Suppl 14:S2. (<http://www.biomedcentral.com/1471-2105/12/S14/S2>)

Montagem de genomas de procariotos

- Spades: <http://spades.bioinf.spbau.ru/release3.6.1/manual.html>
- Velvet: <https://www.ebi.ac.uk/~zerbino/velvet/>

Controle de qualidade de genomas

- QUAST: <http://bioinf.spbau.ru/quast>

Controle de qualidade de Transcriptomas

- BUSCO: <http://buscos.ezlab.org/files/BUSCO-Simao-Waterhouse-Bioinformatics-2015.pdf>

Expressão gênica - Alinhamento

- TopHat: <https://ccb.jhu.edu/software/tophat/index.shtml>
- Kim D, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4):R36.
- Bowtie2: <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357-359.

Expressão gênica

- eXpress: <http://bio.math.berkeley.edu/eXpress/overview.html>
- EdgeR: <https://bioconductor.org/packages/release/bioc/html/edgeR.html>
- Expander: <http://acgt.cs.tau.ac.il/expander/index.html>

Predição Gênica

- Prodigal: <http://prodigal.ornl.gov>
- Glimmer: <https://ccb.jhu.edu/software/glimmer/>