

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
(ФГАОУ ВПО КФУ)

Проректор по научной деятельности



МЕТОДИКА ТРАСКРИПЦИОННОГО ПРОФИЛИРОВАНИЯ ЭУКАРИОТИЧЕСКИХ
КЛЕТОК НА ОСНОВЕ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ

разработана в рамках Соглашения № 14.594.21.0003 от 15.08.2014 о предоставлении
субсидии по теме «Развитие протеогеномного направления Междисциплинарного ЦКП КФУ
для обеспечения клеточных, геномных и постгеномных исследований в Приволжском
регионе»

(уникальный идентификатор проекта RFMEFI59414X0003)

КАЗАНЬ 2015

1. Назначение и область применения

Настоящая лабораторная методика предназначена для проведения биоинформационического анализа транскриптомного профилирования эукариот.

2. Принцип методики

Методика основана на подсчете ридов, картированных на каждый ген в образце и дальнейшей оценке дифференциальной экспрессии генов между исследуемыми образцами.

3. Оборудование и материалы.

3.1. Оборудование для биоинформационического анализа транскриптомных данных

3.1.1. Вычислительный кластер Supermicro

3.1.2. Персональный компьютер

3.2. Материалы

3.2.1 Риды с высокопроизводительного секвенатора SOLiD xl 5500 Wildfire (Life Technologies, США) длиной 50 нуклеотидов или риды с высокопроизводительного секвенатора Ion Proton (Life Technologies, США) длиной до 200 нуклеотидов.

3.2.2 Референсный геном исследуемого организма (файл в формате fasta).

3.2.3 Аннотация генов исследуемого организма (файл в формате gtf или gff).

4. Методика проведения биоинформационического анализа транскриптомных данных эукариот

4.1 Оценка качества и фильтрация исходных ридов

Оценку качества исходных ридов проводили с помощью программы FastQC [1]. На рисунке 1 представлен отчет данной программы.

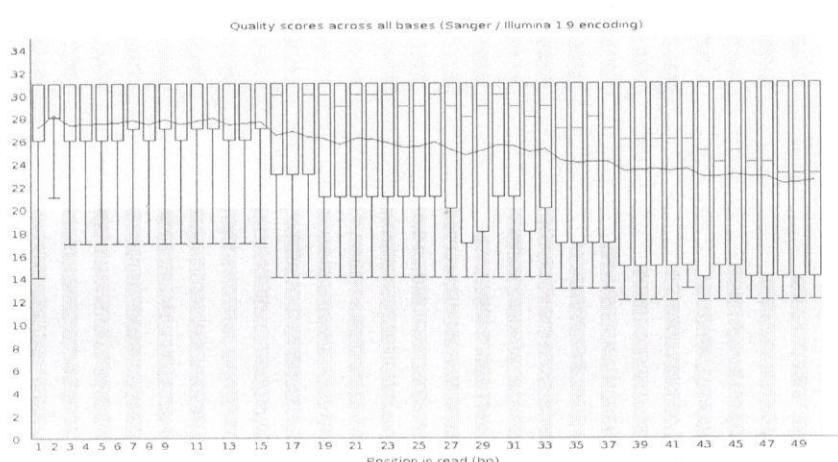


Рис. 1. Отчет о качестве ридов, программа FastQC.

Для ридов, полученных с секвенатора SOLiD 5500, качество каждого нуклеотида в каждом риде считается приемлемым, если оно больше 20. Для Ion Proton - больше 14. Качество нуклеотидов неизбежно снижается к концу рида, поэтому необходимо обрезать риды до такой длины, чтобы качество нуклеотидов было приемлемым. Для тримминга использовалась программа FastX [2].

4.2 Выравнивание на референсный геном, аннотация

В случае анализа данных с секвенатора SOLiD 5500 для выравнивания ридов на референсный геном исследуемого организма, аннотации найденных транскриптов, подсчета ридов, откарированных на каждый транскрипт, поиска новых транскриптов либо изоформ использовалась программа LifeScope (v2.5.1).

Для анализа данных с Ion Proton для выравнивания ридов на референсный геном использовалась программа BowTie [3] [4] и надстройка TopHat [5] [6]; для аннотации и подсчета ридов, откарированных на каждый транскрипт, и поиска новых транскриптов и изоформ использовался пакет программ Cufflinks [6].

4.3 Нормализация и оценка дифференциальной экспрессии

Нормализация – приведение значения уровня экспрессии каждого транскрипта к условной единице, которая учитывает не только число ридов, откарированных на транскрипт, но и длину транскрипта, и общее число ридов.

Для анализа данных с секвенатора SOLiD 5500 нормализацию и оценку дифференциальной экспрессии проводили в среде R с использованием пакета EdgeR [7]. Также с помощью этого пакета визуализировали полученные данные. В случае с Ion Proton для нормализации и оценки дифференциальной экспрессии использовали пакет программ Cufflinks [6], для визуализации результатов анализа использовали пакет CummeRbund [8] в среде R.

5. Список использованных источников и нормативных документов

1. Andrews S. et al. FastQC: A quality control tool for high throughput sequence data //Reference Source. – 2010.
2. Gordon A., Hannon G. J. Fastx-toolkit //FASTQ/A short-reads pre-processing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit. – 2010.
3. Langmead B., Salzberg S. L. Fast gapped-read alignment with Bowtie 2 //Nature methods. – 2012. – Т. 9. – №. 4. – С. 357-359.
4. Langmead B. Aligning short sequencing reads with Bowtie //Current protocols in

- bioinformatics. – 2010. – C. 11.7. 1-11.7. 14.
5. Trapnell C., Pachter L., Salzberg S. L. TopHat: discovering splice junctions with RNA-Seq //Bioinformatics. – 2009. – T. 25. – №. 9. – C. 1105-1111.
 6. Trapnell C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks //Nature protocols. – 2012. – T. 7. – №. 3. – C. 562-578.
 7. Robinson M. D., McCarthy D. J., Smyth G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data //Bioinformatics. – 2010. – T. 26. – №. 1. – C. 139-140.
 8. Goff L. A., Trapnell C., Kelley D. CummeRbund: visualization and exploration of Cufflinks high-throughput sequencing data //R package version. – 2012. – T. 2. – №. 0.