Concorso pubblico, per titoli ed esami, per la copertura di n. 1 posto di Tecnologo di I livello – posizione economica EP1 a tempo determinato, con regime di impegno a tempo pieno presso il Dipartimento di Bioscienze, Biotecnologie e Ambiente dell'Università degli Studi di Bari Aldo Moro, nell'ambito del Piano Nazionale di Ripresa e Resilienza, Missione 4 "Istruzione e ricerca" Componente 2 Investimento 3.1 "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" finanziato dall'Unione europea – NextGenerationEU, per il Progetto "ELIXIR x NextGenerationIT" (codice progetto R0000010 _ CUP B53C22001800006), indetto con DDG n. 663 del giorno 22/06/2023

Domande elaborate dalla commissione esaminatrice per l'espletamento della prova orale.

Prova n. 1

Il/La candidato/a descriva:

- la procedura bioinformatica per il pre-processamento dei dati trascrittomici
- un foglio di calcolo Excel ed uno dei programmi per l'allineamento di dati RNAseq al genoma di riferimento
- il ruolo e la composizione del Senato Accademico

Prova n. 2

Il/La candidato/a illustri:

- un workflow bioinformatico per l'analisi dei dati prodotti con tecnologia PacBio
- la modalità per inserire una formula in un foglio di calcolo Excel ed uno dei programmi per il *trimming* delle sequenze prodotte dai più moderni sequenziatori
- il ruolo e la composizione del Consiglio di Amministrazione

Prova n. 3

Il/La candidato/a illustri:

- un workflow bioinformatico per l'analisi dei dati prodotti con tecnologia PacBio
- la modalità per inserire una formula in un foglio di calcolo Excel ed uno dei programmi per il *trimming* delle sequenze prodotte dai più moderni sequenziatori
- il ruolo e la composizione del Consiglio di Amministrazione

La verifica della conoscenza della lingua inglese è stata condotta sul testo allegato.

Il Segretario della Commissione
Sig. Fabio CORPOSANTO

ACC. VERBACE 3

Conesa *et al. Genome Biology* (2016) 17:13
DOI 10.1186/s13059-016-0881-8

Genome Biology

REVIEW

Open Access

A survey of best practices for RNA-seq data analysis

Ana Conesa[1,2*], Pedro Madrigal[3,4*], Sonia Tarazona[2,5], David Gomez-Cabrero[6,7,8,9], Alejandra Cervera[10], Andrew McPherson[11], Michał Wojciech Szcześniak[12], Daniel J. Gaffney[3], Laura L. Elo[13], Xuegong Zhang[14,15] and Ali Mortazavi[16,17*]

Abstract

RNA-sequencing (RNA-seq) has a wide variety of applications, but no single analysis pipeline can be used in all cases. We review all of the major steps in RNA-seq data analysis, including experimental design, quality control, read alignment, quantification of gene and transcript levels, visualization, differential gene expression, alternative splicing, functional analysis, gene fusion detection and eQTL mapping. We highlight the challenges associated with each step. We discuss the analysis of small RNAs and the integration of RNA-seq with other functional genomics techniques. Finally, we discuss the outlook for novel technologies that are changing the state of the art in transcriptomics.

Background

Transcript identification and the quantification of gene expression have been distinct core activities in molecular biology ever since the discovery of RNA's role as the key intermediate between the genome and the proteome. The power of sequencing RNA lies in the fact that the twin aspects of discovery and quantification can be combined in a single high-throughput sequencing assay called RNA-sequencing (RNA-seq). The pervasive adoption of RNA-seq has spread well beyond the genomics community and has become a standard part of the toolkit used by the life sciences research community. Many variations of RNA-seq protocols and analyses have been published, making it challenging for new users to appreciate all of the steps necessary to conduct an RNA-seq study properly.

There is no optimal pipeline for the variety of different applications and analysis scenarios in which RNA-seq can be used. Scientists plan experiments and adopt different analysis strategies depending on the organism being studied and their research goals. For example, if a genome sequence is available for the studied organism, it should be possible to identify transcripts by mapping RNA-seq reads onto the genome. By contrast, for organisms without sequenced genomes, quantification would be achieved by first assembling reads de novo into contigs and then mapping these contigs onto the transcriptome. For well-annotated genomes such as the human genome, researchers may choose to base their RNA-seq analysis on the existing annotated reference transcriptome alone, or might try to identify new transcripts and their differential regulation. Furthermore, investigators might be interested only in messenger RNA isoform expression or microRNA (miRNA) levels or allele variant identification. Both the experimental design and the analysis procedures will vary greatly in each of these cases. RNA-seq can be used solo for transcriptome profiling or in combination with other functional genomics methods to enhance the analysis of gene expression. Finally, RNA-seq can be coupled with different types of biochemical assay to analyze many other aspects of RNA biology, such as RNA–protein binding, RNA structure, or RNA–RNA interactions. These applications are, however, beyond the scope of this review as we focus on 'typical' RNA-seq.

Every RNA-seq experimental scenario could potentially have different optimal methods for transcript quantification, normalization, and ultimately differential expression analysis. Moreover, quality control checks should be applied pertinently at different stages of the analysis to ensure both reproducibility and reliability of the results. Our focus is to outline current standards

* Correspondence: aconesa@ufl.edu; pm12@sanger.ac.uk; ali.mortazavi@uci.edu
[1] Institute for Food and Agricultural Sciences, Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32603, USA
[3] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
[16] Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697-2300, USA
Full list of author information is available at the end of the article