Supplementary data

S 1. Complementary bibliography

The bibliography included in this section was compiled aiming to provide additional examples on general resources and methodological articles to complement those cited in the main version of this tutorial review. It only represents a partial selection from the large amount of available literature, and it is expected to serve as a guide for young researchers that want to further explore the omics tools covered in this tutorial. For practical reasons, the bibliography is sort by date of publication within different topics.

S_1 Transcriptomics

General resources

  This review article provides an overview of the functioning of RNA-seq technology, and discusses its challenges.


  These articles review the developments in DNA sequencing technologies over the past 10 years. They also discuss the benefits and drawbacks of different approaches used in next-generation sequencing (NGS), and emerging applications.

Library preparation
The authors review the occurrence of biases reported in literature in NGS library preparation protocols for DNA-seq and RNA-seq, and provide some suggestions on how to improve library quality.

**Experimental design**


These articles examine in simulated (Robles and co-workers) and real (Schurch and co-workers) datasets the number of biological replicates needed in RNA-seq experiments for the detection of differential expression. Both articles compare the performance of different analytical tools to detect these differences.


This review discusses the trade-off between sequencing effort (sequencing depth) and replication (biological replicates) in RNA-seq experimental designs.

**Pre-processing**


Description and performance evaluation of some of the tools used in pre-processing NGS reads.

**Transcriptome assembly**


These articles describe the use of Trinity platform for *de novo* transcriptome assembly of RNA-Seq data. Transcriptome reconstruction is necessary in the absence of reference genome, which is the case for most non-model organisms.

**Annotation**


This review explains the most relevant computational challenges for transcriptome annotation, which the authors divide in three categories: (i) read mapping, (ii) transcriptome reconstruction and (iii) expression quantification.

**Differential expression analysis**

These articles present different open-source software tools to perform expression analysis.


The authors evaluate the performance in terms of specificity and sensitivity of RNA-seq measurements from seven different pipelines.

**Single cell transcriptomics**


The article describes single-cell digital gene expression profiling assay, and discuss the importance of this type of analysis to examine transcriptome complexity in individual cells.


The authors review the computational methods, their benefits and drawbacks, for the design and analysis of single cell RNA-seq experiments.

**General methodological considerations**

This book serves as a general resource for a broad audience seeking to understand the bases of NGS as well as available sequencing platforms. It covers several aspects of the analysis of NGS data including genomics, transcriptomics, and metagenomics. It provides the tools to know how, to interpret, visualize and present this type of data.


This review provides an overall picture of each of the major steps in RNA-seq data analysis. It may be very useful especially for beginners because it gives the overall map of the sequence processing with comments and suggestions that may help to improve the experimental design and analysis.

**S_2 Proteomics**

*General resources in MS-based Proteomics*


It represents the first attempt to generate proteomics data and gene annotation from proteomics-derived information.


This article describes the standard protocol used in MS-based proteomics, including experimental design, data collection, data analysis and visualization. The authors discuss the advantages of MS-based protein identification over traditional biochemical approaches.

*Protein enrichment and isolation steps*
The authors review some of the workflows used in classical proteomics analyses, either from digestion or LC separation steps. They also provide a clear description of MS and MS/MS concepts.

**Identification of peptides with tandem mass spectrometry**


A comprehensive review of database searching algorithms. It presents the basic concepts used by most algorithms such as Mascot, SEQUEST, PeptideSearch or X!Tandem, in order to understand how the rationale behind them.

**De novo sequencing and homology-based database search for organisms with unsequenced genomes**


The authors review computer algorithms and programs for automated sequencing such as PEAKS, Lutefisk and PepNovo. They also provide guidelines for homology-based database search for organisms with unsequenced genomes using tools such as MS-BLAST and FASTS.

**Quantification**


These two articles are seminal papers at the beginning of labelling quantification proteomics.


This article describes MaxQuant, an integrated suite of algorithms specifically developed for high-resolution, quantitative MS data, which allow the identification of peptides in large amount of raw data.


A step-by-step instruction for a quantitative derivatization method.


The authors review the state-of-the-art of ionization interfaces in MS. They explain and discuss some of the organic-free-matrix MS that would be able to achieve yoctomoles scales, such as nanostructure-initiator mass spectrometry (NIMS) and laser-induced silicon microcolumn arrays (LISMA).


The authors present a protocol for hypothesis-driven studies in MS-Based Proteomics.

The author review most of the recent technical updates from the last 10 years, and provide some descriptions and advises on the workflows used in targeted proteomics.

*Top-Down Proteomics*


The authors describe the functioning of top down proteomics, and describe how this approach can overcome some of the ambiguous or incomplete protein characterization derived from bottom up process.

*A general guide for proteomics workflows*


These review articles discuss benefits and drawbacks of some of the protocols used in proteomics

*Pipelines of data collection and analysis*


Summarizes recent developments in the PRIDE database and related tools, which represents a great effort in public repositories and integrative data analysis.
S 1.3 Metabolomics

General resources


These articles present and discuss how to treat the large-scale data obtained from metabolomic analysis, from MS and NMR, including statistical tools that can improve data visualization.

Separation techniques


These articles present detailed chromatographic concepts, including different separation techniques and when and how to apply them.

**MS-based metabolomics**


These articles show how to develop metabolomics studies using mass spectrometry approaches, including direct-infusion, GC-MS, LC-MS, LC-MS/MS among others.

**MS-based imaging**


These articles present and discuss the use of different strategies, such as MALDI-IMS and LC-MS, to investigate the distribution of molecules within biological systems through direct analysis from tissue sections and through 3D molecular mapping.

**NMR-based metabolomics**


These articles discuss the advantages and disadvantages of using NMR technique for metabolomics analyses.

**General approaches for studying chemical interactions using metabolomics**


These articles present some strategies commonly used to identify/investigate chemical molecules that mediate ecological interactions between different organisms. They also report the main analytical tools employed in such investigations considering metabolomics approaches.

### S 1.4 Hologenomics

**General resources**

The bibliography suggested for transcriptomics may also be useful in hologenomic analysis. For instance see chapter 14 in the book “Next-generation DNA Sequencing Informatics”

**Identification of uncultured microorganisms**


This article describe for the first time how rRNA sequence can be used to identify uncultured bacteria from natural samples.


These review describe the origin and progresses made in the analysis of genetic content from uncultured microorganism after 10 years of the seminal paper of Amman and co-workers.

**DNA barcode: analytical considerations**


This article may be particularly useful for those researchers that are not so familiar with taxonomic classification of bacteria using DNA barcode data.


The authors describe conservative fragments in 16S rRNA genes that have average coverage rate of 96%, and discuss potential application for metagenomic studies.

**Analysis of metagenomic DNA sequences**


The authors demonstrate that single-end Illumina reads, between 75 and 100 bases, enable to have an excellent consistency in taxonomic recovery and recapture the diversity patterns observed with longer
reads. These results have opened the possibility of conducting large-scale studies of microbial communities analysing thousands of samples simultaneously.


Kraken is classification programme based on exact alignment of $k$-mers that allows assigning taxonomic labels to metagenomic DNA sequences.


The authors introduce MEGAN Community Edition (CE), which is an open source program that provides a straightforward and complete pipeline for the analysis of metagenome shotgun sequences.


Deblur is an open source platform that allows identification of real ecological differences between taxa whose amplicons differ by a single base pair. It is based on a novel sub-operational-taxonomic-unit (sOTU) approach that takes into account putative sequences errors from Illumina MiSeq and HiSeq sequencing platforms.

**Prokaryotic genome annotation**


The author introduces Prokka, a command line software tool that allows full bacterial genome annotation on a typical desktop computer. Prokka is freely available.

**Databases**

The author provides a survey of key databases on microorganisms. It includes databases with different scopes, such as SILVA and GREENGENS for characterization of microbial diversity, and SEED and IMG for annotation and analysis of microbial genomes.

**S 1.6 Omics integration**


A very inspiring book section that discuss how the developments from different omics could be integrated by chemical ecologist to investigate the role of infochemicals in populations and communities.


In this review the authors present an overview of the recent emergence of trans-omic studies from a biochemical perspective using a reconstruction approach of a biochemical trans-omic network through prior knowledge of interactions.