

BASIC DETAILS:

Subject:	ANÁLISIS ÓMICO COMPUTACIONAL		
Id.:	33708		
Programme:	DOBLE GRADO EN FARMACIA Y BIOINFORMÁTICA. PLAN 2019		
Module:	BIOINFORMÁTICA		
Subject type:	OBLIGATORIA		
Year:	3	Teaching period:	Primer Cuatrimestre
Credits:	6	Total hours:	150
Classroom activities:	59	Individual study:	91
Main teaching language:	Inglés	Secondary teaching language:	Castellano
Lecturer:	RANERA BELTRAN, BEATRIZ (T) ROIG MOLINA, FRANCISCO JOSE	Email:	branera@usj.es fjroig@usj.es

PRESENTATION:

Next generation sequencing techniques, NGS, have allowed great improvements in different biological research fields. The vast amount, the quality and the easiness of data generation lead to researchers to perform new investigations with economical feasible methodologies respect to the previous techniques. As a result, new transversal and complementary research lines have emerged to support the classical ones. These new fields, calle -omics are one the main focus of Bioinformatics.

This subject will provide a general idea of the omics techniques, from the experimental design, in order to guide the students to identify the type of analysis to be performed, the accurate tools and the expected results.

PROFESSIONAL COMPETENCES ACQUIRED IN THE SUBJECT:

General programme competences	G01	Use learning strategies autonomously for their application in the continuous improvement of professional practice.
	G02	Perform the analysis and synthesis of problems of their professional activity and apply them in similar environments.
	G03	Cooperate to achieve common results through teamwork in a context of integration, collaboration and empowerment of critical discussion.
	G04	Reason critically based on information, data and lines of action and their application on relevant issues of a social, scientific or ethical nature.
	G05	Communicate professional topics in Spanish and / or English both orally and in writing.
	G06	Solve complex or unforeseen problems that arise during the professional activity within any type of organisation and adapt to the needs and demands of their professional environment.
	G07	Choose between different complex models of knowledge to solve problems.
	G09	Apply information and communication technologies in the professional field.
	G10	Apply creativity, independence of thought, self-criticism and autonomy in the professional practice.
	Specific programme competences	E02
E03		Apply the fundamental concepts of mathematics, logic, algorithmics and computational complexity to solve problems specific to bioinformatics.
E04		Program applications in a robust, correct, and efficient way, choosing the paradigm and the most appropriate programming languages, applying knowledge about basic algorithmic procedures and using the most appropriate types and data structures.
E05		Implement well-founded applications, previously designed and analysed, in the characteristics of the databases.
E06		Apply the fundamental principles and basic techniques of intelligent systems and their practical application in the field of bioinformatics.
E07		Apply the principles, methodologies and life cycles of software engineering to the development of a project in the field of bioinformatics.
E12		Apply the principles and techniques of protein computational modelling to predict their biological function, their activity or new therapeutic targets (Structural Bioinformatics, Computational Toxicology).

E13	Apply omics technologies for the extraction of statistically significant information and for the creation of relational databases of biodata that can be updated and publicly accessible to the scientific community.
E14	Use programming languages, most commonly used in the field of Life Sciences, to develop and evaluate techniques and/ or computational tools.
E15	Infer the evolutionary history of genes and proteins through the creation and interpretation of phylogenetic trees.
E16	Plan linkage and association studies for medical and environmental purposes.
E17	Induce complex relationships between samples by applying statistical and classification techniques.
E18	Apply statistical and computational methods to solve problems in the fields of molecular biology, genomics, medical research and population genetics.
E21	Apply computational and data processing techniques for the integration of physical, chemical and biological concepts and data for the description and/ or prediction of the activity of a substance in a given context.

PRE-REQUISITES:

Not previous requirements are necessary.

SUBJECT PROGRAMME:

Subject contents:

1 - Omics Science: General Principles
1.1 - What the omics are
1.2 - Omics and Bioinformatics
1.3 - Applications
2 - Linux Command Line
2.1 - Introduction to UNIX environments. Basic commands.
2.2 - Commands for access to file contents
2.3 - Permissions
2.4 - Scripting
2.5 - File Processing Language: GAWK
2.6 - Software installation
3 - Omics Revolution: from Sanger to NGS
3.1 - Sequencing
3.2 - De novo sequencing and resequencing
3.3 - Genomics: description and applications
3.4 - Transcriptomics: description and applications
3.5 - Metagenomics and metatranscriptomics: description and applications
3.6 - MethySeq and ChipSeq: description and applications
4 - Present and future of omics science
4.1 - Role of omics science in health, science and industry
5 - Current software for omics analysis
5.1 - Quality analysis and preprocessing
5.2 - Genomics and transcriptomics
5.3 - Metagenomics and metatranscriptomics
5.4 - Annotation
5.5 - Phylogenomics
6 - Proteins and omics science
6.1 - Proteomics: description and applications
6.2 - Metabolomics: description and applications
6.3 - Interactomics: description and applications
6.4 - Relationship between proteomics and NGS: from computer to real life
7 - Basic principles of omics analysis
7.1 - Quality control
7.2 - Preprocessing
7.3 - Assembly
7.4 - Annotation

7.5 - Mapping
7.6 - Variant Calling
7.7 - Quantification
7.8 - Normalization
7.9 - Differential expression analysis
8 - Experimental Design
8.1 - Basic principles of experimental design
8.2 - From the idea to the experimental process

Subject planning could be modified due unforeseen circumstances (group performance, availability of resources, changes to academic calendar etc.) and should not, therefore, be considered to be definitive.

TEACHING AND LEARNING METHODOLOGIES AND ACTIVITIES:

Teaching and learning methodologies and activities applied:

Master classes:

Teacher will explain the theory using TIC and physical resources on presential classes. Material will be available on the PDU in advance for previous reading. Students are highly recommended to perform the reading task.

Theoretical-practical classes:

Omics needed from practise. In this type of sessions the students will perform a reproduction of command execution done in theory and the completion of proposed exercises, in order to assure correct comprehension.

Self-learning based on critical-thinking:

Students will explore scientific publication and discuss between each other about oral presentation of the rest of students. This way, basic knowledge of the field will be reinforce bias for scientific process.

Learning based on projects:

Three projects will be performed time course of the subject in order to apply the acquire knowledge in the experimental design area,

Student work load:

Teaching mode	Teaching methods	Estimated hours
Classroom activities	Master classes	25
	Practical work, exercises, problem-solving etc.	2
	Other practical activities	20
	Test in class	2
	Classroom tutorials	5
	Evaluation tests (questionnaires and other instruments)	5
Individual study	Individual study	15
	Individual coursework preparation	30
	Compulsory reading	14
	Application of investigation techniques and information search	24

Video lessons/Webinars/ podcast	8
Total hours:	150

ASSESSMENT SCHEME:

Calculation of final mark:

Written tests:	5 %
Individual coursework:	30 %
Group coursework:	10 %
Final exam:	40 %
Evaluation of presentations:	15 %
TOTAL	100 %

*Las observaciones específicas sobre el sistema de evaluación serán comunicadas por escrito a los alumnos al inicio de la materia.

BIBLIOGRAPHY AND DOCUMENTATION:

Basic bibliography:

Arivaradarajan, P., Misra, G. Omics Approaches, Technologies And Applications Integrative Approaches For Understanding OMICS Data. Springer. 2019.
Newham, C. Learning the bash Shell, 3rd Edition. O'Reilly Media, Inc. 2005.

Recommended bibliography:

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. Molecular biology of the cell. New York: Garland Science. 2002.
Altschul, S. F., et al. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." <i>Nucleic Acids Res.</i> 25(17): 3389-3402. 1997.
Bar-Even A. et al. Noise in protein expression scales with natural protein abundance. <i>Nat. Genet.</i> 38: 636-643. 2006.
Barh, D., Blum K., and Madigan M. OMICS: Biomedical Perspectives and Applications. CRC Press.. 2011.
Brent, M. R. Genome annotation past, present, and future: How to define an ORF at each locus. <i>Genome Res.</i> , 15:1777-1786. 2006.
Cameron Newham, Bill Rosenblatt, and Gigi Estabrook. Learning the Bash Shell. O'Reilly and associates, Inc., USA. 1998.
Chee-Seng, K., L. En Yun, P. Yudi, and C. Kee-Seng. Next Generation Sequencing Technologies and Their Applications. John Wiley and Sons. 2010.
Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. Ribosomal Database Project: data and tools for high throughput rRNA analysis <i>Nucl. Acids Res.</i> 42(Database issue):D633-D642. 2014.
D'Alessandro, Angelo. High-Throughput Metabolomics. Nature Springer. 2019.
Eidhammer, I., Flikka, K., Martens, L., and Mikalsen, S. O. Computational Methods for Mass Spectrometry Proteomics Wiley-Interscience. 2008.
Futami R., Munoz-Pomer A, Viu JM, Dominguez-Escriba L Covelli L, Bernet GP, Sempere JM, Moya A and Llorens C. GPRO The professional tool for annotation, management and functional analysis of omic databases. <i>Biotechnava Bioinformatics: SOFT3.</i> 2011.
Goff L, Trapnell C and Kelley D. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.14.0. 2013.
Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. <i>Microbiol. Mol. Biol. Rev.</i> , 68: 669-685. 2004.
Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ and O'Donovan C. The GOA database: gene Ontology annotation updates for 2015. <i>Nucleic Acids Res</i> 43: D1057-1063. 2015.
Kotera M, Hirakawa M, Tokimatsu T, Goto S and Kanehisa M. The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. <i>Methods Mol Biol</i> 802: 19-39. 2012.
Kulski, J. K. Next-generation sequencing—an overview of the history, tools, and “omic applications. <i>Handbook of next generation sequencing—advances, applications and challenges.</i> Intech, 2016.
Martin, M. "Cutadapt removes adapter sequences from high-throughput sequencing reads." <i>EMBnet journal Vol 17(1).</i> 2011.
Metzker, M. L. Sequencing technologies the next generation. <i>Nat. Rev. Genet.</i> , 11: 31-46. 2010.

Meyer, F., D. Paarmann, Souza, M.D, Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. and Edwards, R.A. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. 2008.

Myers, C. L., Discovery of biological networks from diverse functional genomic data. *Genome Biology*, 6: R114. 2005

Pérez-Ortín, J.E.; Alepuz, P. y Moreno; J. Genomics and gene transcription kinetics in yeast. *Trends Genet.* 23, 250-257. 2007.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Opens external link in new window Nucl. Acids Res.* 41 (D1): D590-D596. 2013

Schmieder R and Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 27 (6): 863-864. 2011.

Schulz, M. H., et al. "Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels." *Bioinformatics* 28(8): 1086-1092. 2012.

Teh, CK., Ong, A. and Kwong, QB. *Applications of NGS Data: A Practical Handbook of Next Generation Sequencing and Its Applications.* World Scientific. 2017.

The Gene Ontology Consortium: Gene ontology: tool for the unification of biology. *Nature Genetics*, 25:25-29. 2000

Trapnell C, Pachter L and Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105-1111. 2009

Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *NatProtoc* 7: 562-578. 2012.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16), 5261–5267. 2007.

Xu, Y., and Gogarten, J. P. *Computational Methods for Understanding Bacterial and Archaeal Genomes.* Series on Advances in Bioinformatics and Computational Biology, vol. 7. Imperial College Press, London. 2008.

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Opens external link in new window Nucl. Acids Res.* 42:D643-D648. 2014

Young MD, Wakefield MJ, Smyth GK and Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11, pp. R14. 2010.

Zerbino, D. R. and E. Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." *Genome Res.* 18(5): 821-829. 2008

Recommended websites:

Fastqc

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>