



Curso 2018-2019

Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

1. ASIGNATURA / COURSE TITLE

ANÁLISIS DE DATOS MASIVOS Y BIOLOGÍA DE SISTEMAS (BDC7) / SYSTEMS BIOLOGY AND HIGH-THROUGHPUT DATA ANALYSIS (BDC7)

1.1. Código / Course number

32854

1.2. Materia / Content area

ANÁLISIS DE DATOS MASIVOS Y BIOLOGÍA DE SISTEMAS (BDC7) / SYSTEMS BIOLOGY AND HIGH-THROUGHPUT DATA ANALYSIS (BDC7)

1.3. Tipo / Course type

Optativa / Optional

1.4. Nivel / Course level

Máster / Master

1.5. Curso/ Year

Primero / First

1.6. Semestre / Semester

Segundo / Second

1.7. Idioma / Language

Inglés / English

1.8. Requisitos previos / Prerequisites

Conocimientos de Genética Molecular (nivel grado).
Técnicas de Biología Molecular (nivel grado).
Matemáticas y estadística (nivel de un grado de ciencias).
Inglés escrito y hablado (nivel B2).

Molecular Genetics (University Degree Level)



Curso 2018-2019

Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

Molecular Biology techniques (University Degree Level)
Mathematics and Statistics (Science Degree Level)
Spoken and written English (Level B2)

1.9. Requisitos mínimos de asistencia a las sesiones presenciales / **Minimum attendance requirement**

Esta asignatura forma parte de un Master presencial. Para obtener la calificación de aprobado en la asignatura será requisito que el alumno haya asistido al menos al 80% de todas las actividades presenciales. / **This course is a part of a Master in which attendance is mandatory. As a requisite to obtain the "pass" mark in this course, the student will have to attend to at least 80 % of the classroom activities.**

Asistencia no obligatoria pero recomendada / Attendance not required but strongly recommended.

1.10. Datos del equipo docente / **Faculty data**

Coordinador / **Coordinator:** Gonzalo Gómez López
Departamento de / **Department of:** Bioquímica / Biochemistry
Facultad / **Faculty:** Medicina / Medicine
Despacho - Módulo / **Office - Module:** Unidad Bioinformática, Centro Nacional de Investigaciones Oncológicas (CNIO) / Bioinformatics Unit, National Cancer Research Centre (CNIO)
Teléfono / **Phone:** +34 91 732 80 00 ext.3062
Correo electrónico/**Email:** ggomez@cni.es, gonzalo.gomez@uam.es
Página web/**Website:** www.cni.es, http://bioinfo.cni.es/
Horario de atención al alumnado/**Office hours:** Pedir cita / Under appointment

Coordinador / **Coordinator:** Alberto López Bueno
Departamento de / **Department of:** Biología Molecular / Molecular Biology
Facultad / **Faculty:** Ciencias / Science
Despacho - Módulo / **Office - Module:** Lab. 224 CBMSO
Teléfono / **Phone:** +34 91 196 45 89
Correo electrónico/**Email:** alopezbueno@cbm.uam.es
Página web/**Website:** www.cbms0.es/alopezbueno
Horario de atención al alumnado/**Office hours:** Pedir cita / Under appointment

Profesor/**Teacher:** Ramón Díaz-Uriarte (UAM).



Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

email: ramon.diaz@iib.uam.es

Facultad / Faculty: Medicina / Medicine, Lab B-25

Teléfono / Phone: +34 91 497 24 12

Página web/Website: <http://ligarto.org/rdiaz>

Horario de atención al alumnado/Office hours: Pedir cita / Under appointment

1.11. Objetivos del curso / Course objectives

Esta asignatura facilitará la adquisición por los estudiantes de las siguientes competencias generales y específicas del Master en Biomedicina Molecular:

Competencias generales:

CG1.- Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con la Biomedicina y Biología Molecular.

CG2.- Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

CG3.- Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CG4.- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida auto-dirigido o autónomo.

Competencias específicas:

CE1. - Comprender la necesidad del análisis cuantitativo en biología molecular y biomedicina y las ventajas y oportunidades que ofrece el uso de modelos matemáticos.

CE2 - Adquirir nociones básicas para analizar sistemas complejos y grandes volúmenes de datos. Demostrar una buena capacidad de acceder por



Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

búsquedas electrónicas en bases de datos a la literatura científica y técnica. Demostrar unos conocimientos mínimos y una destreza en el manejo de las herramientas bioinformáticas básicas de mayor relevancia en la Biomedicina y Biología Molecular.

Objetivos y relación con las competencias:

Esta asignatura tiene cinco objetivos generales fundamentales:

1. Permitir a los estudiantes la elección, con criterio fundamentado, entre métodos bioinformáticos.
2. Convertir a los estudiantes en lectores críticos de la literatura científica primaria que desarrolla y usa métodos bioinformáticos.
3. Facilitar el acceso a la investigación en bioinformática.

Entre las competencias y destrezas específicas a adquirir se incluyen:

- Desarrollar la capacidad de analizar genomas y sus anotaciones.
- Entender algunos de los algoritmos clásicos en bioinformática.
- Entender las etapas implicadas en el análisis estadístico de datos de alto rendimiento (high-throughput) en particular el análisis de expresión diferencial y el desarrollo de modelos de clasificación a partir de datos genómicos. Comprender la metodologías básicas para el estudio de metagenomas.
- Familiarizarse con algunas de las áreas de investigación activa en la bioinformática actual.
- Promover la interacción entre estudiantes con diferente formación académica

This subject will facilitate the acquisition of the next general and specific abilities related to Biomedicine and Molecular Biology:

Generic Skills:

CG1.- To ensure students acquire the knowledge and ability they will need to identify problems, and find practical and creative solutions, as well as to apply them in a research or professional context in the field of Biomedicine and Molecular Biology.



Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

CG2.- To enable students to integrate knowledge and elaborate critical opinion from data and information received. This includes social and ethical responsibilities linked to the applications of their own expertise.

CG3.- To develop their ability to effectively communicate scientific advances in the field of Molecular Biomedicine to a non-specialist audience, as well as their ethical and social implications.

CG4.- To enable students to acquire the self-study skills they will need to continue learning autonomously throughout their lives.

Specific skills:

CE1.- To understand the need and utility of quantitative analysis demonstrating the advantages and opportunities that mathematical models offers to Molecular Biology and Biomedicine.

CE2.- To acquire basics in data analysis of complex systems and genome-wide experiments. To develop the ability to readily access scientific and technical literature databases using electronic searches. Demonstration of a good knowledge and the ability to use the most fundamental Molecular Biology and Biomedicine bioinformatics tools.

Objectives and relationships with the skills:

This course has three main objectives:

1. Allow the students to make choices, with rigorous criteria, among bioinformatics methods.
2. Turn the students into critical reader of the primary scientific literature that uses and develops bioinformatics methods.
3. Ensure that students are familiar with crucial elements in a variety of bioinformatics areas.

In this sense, some of the specific skills acquired include:

- Developing of the capacity to analyse genomes and their annotations.
- Understanding some classic bioinformatics algorithms.
- Understanding the key steps involved in the statistical analysis of high throughput experiments. In particular, differential expression and



- classification models for genomic data. To understand fundamentals in metagenomics.
- Becoming familiar with active areas of research in bioinformatics.
 - Encouraging the interaction between students with diverse academic backgrounds.

1.12. Contenidos del programa / Course contents

Tema 1. Introducción a las técnicas high-throughput (arrays y NGS). Transcriptómica. (1 sesión)

Tema 2. Expresión diferencial y testeo multiple; modelos lineales. (1 sesión)

Tema 3. Análisis funcional: métodos de sobrerrepresentación y métodos de bloques. (2 sesiones)

Tema 4. DNA-seq para la identificación de mutaciones y SNPs. ChIP-seq (1 sesión)

Tema 5. Clasificación, clustering y análisis de supervivencia. Validación cruzada y bootstrapping para estimar el error del clasificador. (4 sesiones)

Tema 6. Seminario Genómica y Transcriptómica. (1 sesión)

Tema 7. Introducción a la metagenómica. Repositorios y bases de datos. Secuenciación al azar de genomas fragmentados y secuenciación de genes marcadores. Unidades taxonómicas operativas (OTUs). Ensamblaje *de novo*: solapamiento y grafos De Bruijn. Taxonomía (¿Quién hay ahí?) y función (¿Qué funciones hacen?). Curvas de rarefacción y estimación de alfa y beta diversidad microbiana. Comparación de metagenomas basado en perfiles taxonómicos, composición nucleotídica y alineamientos. (3 sesiones)

Tema 8. Seminario de metagenómica. (1 sesión)

Topic 1. Introduction to high-throughput methodologies (arrays and next-generation sequencing). Transcriptomics. (1 session)

Topic 2. Differential expression and multiple testing; lineal models. (1 session)



Topic 3. Functional analysis: overrepresentation approaches and gene set-based methods. (2 sessions)

Topic 4. DNA-seq to determine mutations and SNPs. ChIPseq. (1 session)

Topic 5. Classification, clustering and survival analysis. Cross-validation and bootstrapping to estimate error in classification. (4 sessions)

Topic 6. Seminar on NGS applications. (1 session)

Topic 7. Introduction to metagenomics. Repositories and databases. Whole Genome Shotgun sequencing and deep sequencing of maker genes (16S). Operational Taxonomic Units (OTUs). *De novo* assembly (Overlapping Layout Consensus and De Bruijn Graphs). Taxonomy (Who is there?) and functions (What do they do?). Rarefaction curves and estimation of alpha and beta microbial diversity. Comparative metagenomics based on alignments, signatures and taxonomic profiles. (3 sessions)

Topic 8. Seminar on metagenomics. (1 session)

1.13. Referencias de consulta / [Course bibliography](#)

La mayor parte de la bibliografía serán artículos de la literatura científica actual que serán distribuidos o indicados a los alumnos antes o durante las clases. La mayoría de los temas expuestos se pueden estudiar en:

Most of the bibliography will be papers from the current literature that will be distributed/pointed to during or before the class. Most of the topics for teaching are covered in:

- Dziuda DM. 2010. Data mining for genomics and proteomics. Wiley.
- Simon et al. 2004. Design and analysis of DNA microarray investigations. Springer.
- Thomas J Sharpton. 2014. An introduction to the analysis of shotgun metagenomic data.
- Jason R. Miller et al. 2010. Assembly Algorithms for Next-Generation Sequencing Data. Genomics.
- Logares R et al. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. J Microbiol Methods.



Curso 2018-2019

Asignatura: Análisis de datos masivos y biología de sistemas.
Código: 32854
Centro: UAM
Titulación: Master en Biomoléculas y Biodinámica Celular
Nivel: Master
Tipo: Optativa
Nº de créditos: 3 ECTS

2. Métodos docentes / **Teaching methodology**

Clases Teóricas apoyadas con material multimedia
Resolución de problemas o casos prácticos en el aula
Sesiones de discusión en grupo guiadas por el Profesor
Seminarios y/o conferencias a cargo de expertos
Aprendizaje basado en problemas
Análisis crítico de la literatura científica
Metodologías e-learning

Theoretical lectures supported by multimedia contents.
Problem solving and practical cases in the class.
Guided discussion sessions
Seminars by experts in the area
Problem-based learning
Critical analysis of scientific literature
E-learning



Tiempo de trabajo del estudiante / **Student workload**

		Horas/Hours	%
Presencial/ Presential	Clases teóricas / Theoretical lectures	22,5	
	Seminarios / Seminars	3	
	Exámen final / Final exam	1.5	
	Tutorías / Tutorials	1	
	TOTAL	28	
No presencial/ Non-presential	Estudio semanal / Study per week	42	
	Preparación seminario / Seminar set up	0	
	Preparación del examen / Exam preparation	5	
	TOTAL	47	
Carga total de horas de trabajo / Total working hours		75	100

3. **Métodos de evaluación y porcentaje en la calificación final / Evaluation procedures and weight of components in the final grade**

Esta asignatura forma parte de un Master presencial. Para obtener la calificación de aprobado en la asignatura será requisito que el alumno haya asistido al menos al 80% de todas las actividades presenciales. / **This course is a part of a Master in which attendance is mandatory. As a requisite to obtain the "pass" mark in this course, the student will have to attend to at least 80 % of the classroom activities.**

Examen escrito (50%). Ejercicios en casa (45%). Participación en clase (5%). El peso en la evaluación de cada bloque será proporcional al número de horas asistenciales. Los ejercicios no presentados tendrán una calificación de cero puntos.



Written examination (50%). Homework (45%). Class participation (5%). The weight of the evaluation of each teaching block will be proportional to the number of presential hours. Missed exercises will be rated as zero.

Convocatoria extraordinaria:

Para los alumnos suspendidos, en la convocatoria extraordinaria la evaluación consistirá en un examen final escrito (100% de la nota).

Extraordinary examination:

Students suspended in the final exam will be re-evaluated through a written exam (100% of the final mark).

4. Cronograma* / Course calendar*

Semana Week	Contenido Contents	Horas presenciales Contact hours	Horas no presenciales Independent study time
1	Intro. métodos masivos. Transcriptómica. <i>Intro. High-throughput methods. Transcriptomics.</i>	2	3
1	Expresión diferencial y multiple testing. <i>Differential expression and multiple testing.</i>	2	4
1	DNaseq. ChIPseq.	2	4
1	Análisis Funcional Functional Analysis	4	8
2	Clasificación, clustering, análisis de supervivencia. <i>Classification, clustering, survival analysis.</i>	8	15
2	Seminario de NGS <i>Seminar on NGS</i>	1.5	0
3	Metagenómica <i>Metagenomics</i>	4,5	8
3	Seminario de metagenómica <i>Seminar on metagenomics</i>	1.5	0

*Este cronograma tiene carácter orientativo.

*This is a tentative course calendar.