



Guía docente

Guía docente				
Datos Identificativos				2022/23
Asignatura (*)	Gestión de Datos Ómicos y Modelización		Código	614G02042
Titulación	Grao en Ciencia e Enxeñaría de Datos			
Descriptor				
Ciclo	Periodo	Curso	Tipo	Créditos
Grado	2º cuatrimestre	Cuarto	Optativa	6
Idioma	Castellano			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da Información			
Coordinador/a	Fernández Lozano, Carlos	Correo electrónico	carlos.fernandez@udc.es	
Profesorado	Fernández Lozano, Carlos	Correo electrónico	carlos.fernandez@udc.es	
Web	cafernandezlo.github.io/es_github_cafernandezlo/teaching/			
Descripción general	A lo largo de los últimos años, la cantidad de datos biológicos disponibles está creciendo de manera exponencial. Esto hace posible que diferentes disciplinas científicas puedan disponer de ellos para estudiar a nivel molecular diferentes organismos con el objetivo de generar nuevo conocimiento y/o verificar el existente. El perfil de un investigador que utilice datos ómicos debe ser computacional pero, a su vez, debe llevar asociado un perfil de interés biológico para que las técnicas de análisis a aplicar y los resultados obtenidos cobren sentido. Dichos conocimientos biológicos se integrarán durante las clases de forma sencilla para que el alumnado adquiriera las competencias necesarias en el desarrollo de los análisis. La reducción de costes y el incremento de la capacidad de cómputo de los últimos tiempos ha logrado acercar las plataformas de secuenciación masiva al día a día de la investigación y, en algunos casos, a la asistencia clínica diaria. Como ejemplo, a día de hoy, existen paneles ómicos que se utilizan para determinar si una determinada paciente sufre un cáncer de mama, el estadio en el que se encuentra, el subtipo concreto que padece y se generan modelos de predicción de recidivas y de respuesta a fármaco. Todo ello es posible gracias a los datos ómicos y a análisis bioinformáticos. Gracias a las aproximaciones de ciencia de datos es posible comprender el funcionamiento de fenómenos biológicos complejos a diferentes niveles (genoma, proteoma, microbioma, etc) y es necesario aplicar nuevos paradigmas capaces de gestionar y analizar el enorme volumen de datos del que se dispone.			

Competencias del título

Código	Competencias del título
A2	CE2 - Capacidad para resolver problemas matemáticos, planificando su resolución en función de las herramientas disponibles y de las restricciones de tiempo y recursos.
B2	CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
B3	CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética
B4	CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado
B7	CG2 - Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables.
B8	CG3 - Ser capaz de mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo.
B9	CG4 - Capacidad para abordar con éxito todas las etapas de un proyecto de análisis de datos: exploración previa de los datos, preprocesado, análisis, visualización y comunicación de resultados.
B10	CG5 - Ser capaz de trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones.



C1	CT1 - Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida.
C4	CT4 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad.

Resultados de aprendizaje			
Resultados de aprendizaje		Competencias del título	
Conocimiento y aplicación de técnicas de análisis de datos ómicos desde la generación de datos científicos de secuencias de ácidos nucleicos o secuencias de proteínas hasta el estudio de diferencias entre subtipos de una determinada patología		A2	B2 B3 B4 B7 B8 B9 B10
Manejar las herramientas y entornos más actuales en el ámbito de análisis de datos biológicos		A2	B9 B10
			C1 C4

Contenidos	
Tema	Subtema
1. Introducción a los datos ómicos	1.1. El ADN 1.2. El dogma central de la biología molecular 1.3. Las ómicas
2. Trabajo con secuencias moleculares	2.1. El formato FASTA 2.2. Bases de datos abiertas 2.3. Herramientas para el análisis de secuencias moleculares: BLAST, Clustal, Galaxy...
3. Tecnologías de secuenciación masiva (NGS)	3.1. El origen con la secuenciación Sanger 3.2. Nuevas tecnologías NGS 3.3. Illumina, PacBio, MinION, Solexa... 3.4. Diferencias entre plataforma de secuencia corta y larga, aplicaciones más frecuentes
4. Análisis de calidad y filtrado de secuencias	4.1. Formato FASTAQ 4.2. Control y evaluación de calidad de las secuencias 4.3. Filtrado de las secuencias
5. Ensamblaje de genomas y metagenomas	5.1. Ensamblaje de novo 5.2. Ensamblaje contra genoma de referencia 5.3. Herramientas software de ensamblaje 5.4. Anotación de las secuencias
6. Análisis de expresión génica mediante RNA-Seq	6.1. Preprocesado 6.2. Análisis de expresión diferencial con R/Bioconductor: edgeR, DESeq2, ... 6.3. Aplicación de machine learning a datos transcriptómicos: Caret, mlr3, ... 6.4. The Cancer Genome Atlas (TCGA)
7. Análisis del metagenoma	7.1. Secuenciación del genoma completo (shotgun) 7.2. Secuenciación del genoma bacteriano 16S rRNA 7.3. Anotación basada en asignación de OTUs 7.4. Análisis de diferencias en equilibrio microbiano 7.5. American Gut Project (AGP), Human Microbiome Project (HMP)

Planificación



Metodologías / pruebas	Competencias	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales
Prácticas de laboratorio	A2 B2 B9 B10 C1 C4	16	16	32
Prueba de respuesta múltiple	B7 B8	3	21	24
Trabajos tutelados	A2 B2 B3 B4 B7 B9 B10 C1 C4	6	30	36
Sesión magistral	B7 B8 C1 C4	21	36	57
Atención personalizada		1	0	1
(*) Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	Descripción
Prácticas de laboratorio	Resolución de problemas y diferentes partes individuales del pipeline de análisis de datos ómicos
Prueba de respuesta múltiple	Prueba de evaluación escrita en la que el/la estudiante deberá demostrar los conocimientos adquiridos
Trabajos tutelados	Elaboración supervisada de análisis de datos ómicos reales mediante técnicas aprendidas a lo largo de las sesiones magistrales y de prácticas de laboratorio de la asignatura
Sesión magistral	Impartición teórica de la materia de la asignatura. Se alternarán la exposición de nuevos conceptos teóricos con la revisión de ejemplos y la resolución de pequeños ejercicios

Atención personalizada	
Metodologías	Descripción
Prácticas de laboratorio Trabajos tutelados	El profesor guiará el proceso de resolución de los problemas y solventará las dudas que surjan en el desarrollo

Evaluación			
Metodologías	Competencias	Descripción	Calificación
Prácticas de laboratorio	A2 B2 B9 B10 C1 C4	Las/os estudiantes deberán realizar pequeñas partes de los pipelines de análisis de datos ómicos para asentar el conocimiento teórico y prepararse para el trabajo tutelado	10
Trabajos tutelados	A2 B2 B3 B4 B7 B9 B10 C1 C4	Las/os estudiantes deberán analizar datos ómicos reales aplicando las técnicas/aproximaciones aprendidas en la asignatura. Deberá acompañarse de un informe replicable de resultados, código y conclusiones realizado en entorno R justificando el pipeline de análisis utilizado y las conclusiones obtenidas	60
Prueba de respuesta múltiple	B7 B8	En el periodo de evaluación se realizará una prueba de conocimientos teórico-prácticos para evaluar la adquisición de las competencias por parte del alumno	30

Observaciones evaluación



La/el estudiante deberá obtener una nota mínima de 3 sobre 10 puntos en la prueba objetiva y una nota mínima de 4 sobre 10 puntos en los trabajos tutelados. Las prácticas de laboratorio no tienen nota mínima. De no alcanzar esta nota mínima en alguno de dichos apartados, la nota de la materia será la menor de las dos. Las entregas correspondientes a los trabajos tutelados en las fechas indicadas y la asistencia a las clases prácticas son obligatorias para aprobar la asignatura. Un/una estudiante se considerará presentado/a en una convocatoria si hace entrega de los trabajos tutelados o si se presenta a la prueba objetiva. El trabajo entregado deberá ser original del/de la estudiante. De acuerdo al artículo 14, apartado 4, de la normativa*, la entrega de trabajos no originales o con partes duplicadas (sea por copias entre compañeros o por obtención de otras fuentes...) llevará una nota global de SUSPENSO en la convocatoria correspondiente, tanto para el/la estudiante que presente material copiado como a quien lo haya facilitado, invalidando cualquier otra calificación obtenida en las actividades evaluables.* Normativa de evaluación, revisión y reclamación de las calificaciones de los estudios de grado y máster universitario, aprobada por el Consejo de Gobierno de la Universidade da Coruña el 19 de diciembre de 2013. Sobre la responsabilidad compartida de los trabajos en grupo. En las actividades que se llevan a cabo en grupos, tales como los trabajos tutelados, todos los miembros del grupo serán responsables solidarios del trabajo realizado y entregado, así como de las consecuencias que se deriven del incumplimiento de las normas de autoría del mismo. Segunda oportunidad y convocatorias posteriores En la segunda oportunidad, se mantiene la nota obtenida en los trabajos tutelados. Solo en el caso de no haber presentado los trabajos tutelados en la primera oportunidad o que estas hubiesen obtenido la calificación de SUSPENSO (calificación menor que 5), se podrán presentar nuevamente los trabajos tutelados en la segunda oportunidad. En caso de suspender la asignatura, los trabajos tutelados con nota igual o superior a 5 se guardarán para cursos posteriores con calificación de aprobado (5). En cada curso, el/la estudiante tendrá la opción de entregar un nuevo trabajo tutelado que sustituirá la nota de la anterior. Los trabajos no se guardarán más de un curso. Matrícula con dispensa académica Para estudiantes con matrícula con dispensa académica se elimina la obligatoriedad de asistencia a las clases de prácticas. Deberán entregar los trabajos tutelados en las fechas establecidas. Es responsabilidad de dichos/as estudiantes poner en conocimiento del profesor su circunstancia. Tutorías A mayores, las tutorías se consideran una parte importante dentro del desarrollo de la asignatura. Están orientadas de tal manera que las/los estudiantes tengan y/o puedan consultar distintas cuestiones como: 1. Problemas en el desarrollo de las prácticas 2. Maneras de enfocar/organizar los trabajos tutelados 3. Resolución de dudas sobre las cuestiones teóricas. Se pedirá a las/los estudiantes que soliciten cita a los/las profesores/as responsables para realizar reuniones presenciales dentro de los horarios de tutorías del profesorado establecido en espazos.udc.es. El uso de Teams deberá limitarse a cuestiones puntuales organizativas.

Fuentes de información

Básica	- Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. <i>Genome Biology</i> - Chen Y, Lun AAT, Smyth GK (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. <i>F1000Research</i> - Enis Afgan, Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, Martin ?ech, John Chilt (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. <i>Nucleic Acids Research</i> - TCGA Consortium (2022). The Cancer Genome Atlas. https://portal.gdc.cancer.gov/ - NCBI Gene Expression Omnibus (2022). NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/ - Michael Love, Wolfgang Huber y Simon Anders. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. <i>Genome Biology</i> - Malachi Griffith y col. (2015). Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. <i>Plos Computational Biology</i>
Complementaria	- Liñares-Blanco J., Fernandez-Lozano C., Seoane JA y López-Campos G. (2022). Machine Learning Based Microbiome Signature to Predict Inflammatory Bowel Disease Subtypes. <i>Frontiers in Microbiology</i> - Fernández-Edreira D., Liñares-Blanco J. y Fernandez-Lozano C. (2021). Machine Learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. <i>Expert Systems with Applications</i> - Liñares-Blanco, J., Gestal, M., Dorado, J., y Fernandez-Lozano, C. (2019). Differential gene expression analysis of RNA-seq data using machine learning for cancer research. <i>Machine Learning Paradigms. Learning and Analytics in Intelligent Systems. Vol 1. Springer, Cham.</i>

Recomendaciones

Asignaturas que se recomienda haber cursado previamente



Análisis Estadístico de Datos Complejos/614G02031

Aprendizaje Automático I/614G02019

Cálculo Multivariable/614G02006

Asignaturas que se recomienda cursar simultáneamente

Asignaturas que continúan el temario

Análisis Estadístico de Datos Complejos/614G02031

Aprendizaje Automático a Gran Escala/614G02032

Otros comentarios

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías