



Universidad
Internacional
de Valencia

Guía didáctica

Secuenciación genómica y análisis de variantes

Título: *Máster Universitario en Bioinformática*

Materia: *Bioinformática genómica*

Créditos: 6 ECTS

Código: 04MBIF

Índice

1. Organización general.....	3
1.1. Datos de la asignatura.....	3
1.2. Equipo docente	3
1.3. Introducción a la asignatura.....	3
1.4. Competencias.....	4
2. Contenidos/temario	5
3. Metodología	8
4. Actividades formativas	8
5. Evaluación.....	9
5.1. Sistema de evaluación.....	9
5.2. Sistema de calificación	10
6. Bibliografía.....	10
6.1. Bibliografía de referencia.....	10

1. Organización general

1.1. Datos de la asignatura

MATERIA	Bioinformática genómica
ASIGNATURA	Secuenciación genómica y análisis de variantes 6 ECTS
Carácter	Obligatorio
Cuatrimestre	Primero
Idioma en que se imparte	Castellano
Requisitos previos	No existen
Dedicación al estudio por ECTS	25 horas

1.2. Equipo docente

Profesor/a	Dra. María de Toro Hernando <i>maria.detoro@campusviu.es</i>
-------------------	--

1.3. Introducción a la asignatura

Esta asignatura se centra en profundizar en una de las ramas de la Bioinformática Genómica más aplicadas, el análisis DNaseq, o análisis de genomas. Veremos todo el procesamiento de las muestras, desde el diseño experimental, preparación de las muestras para secuenciar, secuenciación de las muestras, obtención de datos, análisis y resultados. Nos centraremos en la secuenciación genómica y análisis de variantes, tanto por técnicas de resecuenciación y de mapeado, como análisis “de novo”. Veremos este análisis aplicado a genomas eucariotas (humano), así como a patógenos de interés, como bacterias y virus (SARS-CoV-2). Para ello, utilizaremos las herramientas más actuales y de mayor distribución, basándonos en software libre. Esta asignatura forma parte de la materia del Máster Bioinformática Genómica (18 ECTS).

Durante esta materia se conocerán con detalle tres flujos de trabajo muy utilizados en bioinformática (DNaseq, RNAseq y metagenómica). El objetivo de cada asignatura en esta materia será el de mostrar desde el diseño experimental hasta el análisis de resultados pasando por el pipeline bioinformática para cada uno de los casos.

1.4. Competencias

COMPETENCIAS BÁSICAS

CB6: Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7: Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB8: 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.

CB9: 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.

CB10: Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS ESPECÍFICAS DE LA ASIGNATURA

C.E.3.- Saber utilizar herramientas de Python en el entorno de la bioinformática.

C.E.4.- Ser capaz de analizar ficheros de datos biológicos mediante el lenguaje de programación Python.

C.E.5.- Saber interpretar los resultados de los análisis bioinformáticos en el lenguaje de programación Python.

C.E.6.- Saber utilizar herramientas de R en el entorno de la bioinformática.

C.E.7.- Saber analizar ficheros de datos biológicos mediante el lenguaje de programación R.

C.E.8.- Saber interpretar los resultados de los análisis bioinformáticos en el lenguaje de programación R.

C.E.9.- Saber utilizar herramientas de conexión remota a centros de procesamiento de datos (CPD) en la resolución de problemas específicos de bioinformática.

C.E.10.- Ser capaz de seleccionar las técnicas bioestadísticas adecuadas para el análisis en bioinformática.

C.E.11.- Saber analizar los principales formatos de secuencias en la aplicación de datos ómicos.

C.E.12.- Ser capaz de extraer la información necesaria de las principales bases de datos de depósito de información biológica, mediante herramientas de automatización o scripting, en la resolución de problemas bioinformáticos.

C.E.14.- Saber establecer los distintos parámetros que definen la calidad de las secuencias que se obtienen de los secuenciadores.

C.E.15.- Ser capaz de aplicar los principales métodos de selección y mejora de calidad de secuencias en la bioinformática.

C.E.16.- Saber diseñar el flujo de trabajo aplicando los principios generales del diseño de experimentos ómicos.

C.E.17.- Ser capaz de aplicar los principales algoritmos de alineamiento de secuencias de datos ómicos.

C.E.18.- Saber identificar las principales herramientas de análisis de datos para la medicina personalizada.

C.E.19.- Ser capaz de analizar grandes volúmenes de datos mediante aplicaciones bioinformáticas en medicina personalizada.

2. Contenidos/temario

Tema 1. La estructura del genoma humano y patrones de transmisión de enfermedades genéticas.

- 1.1. Estructura del genoma humano
 - 1.1.1. Genoma mitocondrial
 - 1.1.2. Genoma nuclear
 - 1.1.3. ¿Qué tipos de secuencias encontramos en el genoma humano?
 - 1.1.4. Secuencias repetidas y secuenciación masiva
- 1.2. Patrones de transmisión de las enfermedades genéticas.
 - 1.2.1. Variabilidad del genoma
 - 1.2.2. Tipos de enfermedades genéticas

Tema 2. Introducción a la Medicina Preventiva Personalizada. Exposoma. Tipos de datos.

- 2.1. Medicina Preventiva Personalizada
 - 2.1.1. Definición y objetivos
 - 2.1.2. ¿Qué aporta la genómica en este área?
 - 2.1.3. Aplicaciones de la Medicina Preventiva Personalizada.
 - 2.1.4. Retos de la Medicina Preventiva Personalizada.
- 2.2. Exposoma.
 - 2.2.1. Definición.
 - 2.2.2. Factores no genéticos.
 - 2.2.3. ¿Cómo se estudia el exposoma?
 - 2.2.3. Patologías más comunes y su relación con exposoma.
 - 2.2.4. Retos asociados.
- 2.3. Datos de Medicina Personalizada y de Precisión.
 - 2.3.1. Datos integrados.
 - 2.3.2. Tipos, fuentes y características de los datos.
 - 2.3.5. Retos asociados.

Tema 3. ¿Cómo analizamos un genoma eucariota? Paneles de captura de genes vs genoma completo.

- 3.1. Introducción al análisis de genomas eucariotas

- 3. 2. Estrategias generales de análisis
 - 3.2.1. Paneles de genes o regiones de interés.
 - 3.2.2. Secuenciación de exoma (WES)
 - 3.2.3. Secuenciación de genoma completo (WGS)
- 3.3. Protocolo de análisis
 - 3.3.1. Extracción del ADN
 - 3.3.2. Preparación de la genoteca/librería
 - 3.3.3. Secuenciación de la genoteca/librería.

Tema 4. ¿Cómo analizamos un genoma eucariota? Análisis bioinformático de paneles de captura de genoma humano

- 4.1. Análisis primario. Calidad y filtrado de secuencias
 - 4.1.1. ¿Por qué es importante evaluar la calidad de las lecturas y filtrarlas?
 - 4.1.2. Herramientas de control de calidad
 - 4.1.3. Análisis de calidad con FastQC
 - 4.1.4. Herramientas de filtrado por calidad de las secuencias y eliminación de adaptadores
- 4.2. Mapeo de secuencias. Herramientas de mapeo, visualización y análisis de calidad,
 - 4.2.1. El proceso de mapeo
 - 4.2.2. Herramientas para el mapeo.
 - 4.2.3. Archivos de mapeo: el formato SAM.
 - 4.2.4. Herramientas para el manejo de los archivos SAM. Generación de archivos BAM.
 - 4.2.5. Visualización del mapeo.
 - 4.2.6. Análisis de calidad del mapeo.
- 4.3. Identificación de variantes.
 - 4.3.1. Preprocesamiento: identificación de secuencias duplicadas
 - 4.3.2. Identificación de variantes
 - 4.3.3. Post-procesado: filtrado de variantes y etiquetado
- 4.4. Anotación de variantes.
 - 4.4.1. Variant Effect Predictor (VEP)
 - 4.4.2. Annovar

Tema 5. Análisis de genomas bacterianos

- 5.1. Genoma procarionta
 - 5.1.1. Características generales de los genomas procariontas
 - 5.1.2. Genomas multipartitos
 - 5.1.3. Estructura detallada del genoma procarionta
 - 5.1.4. Número de genes y función general.
- 5.2. Epidemiología Genómica. Retos y oportunidades en el análisis de genomas bacterianos.

- 5.3. Tipificación de genomas bacterianos.
 - 5.3.1. Análisis basado en polimorfismos de un único nucleótido (SNP)
 - 5.3.2. Análisis gen-por-gen
 - 5.3.3. Identidad Nucleotídica Promedio (Average Nucleotide Identity, ANI).
- 5.4. Reconstrucción del genoma: ensamblaje de novo
 - 5.4.1. Genotecas para el ensamblado de novo
 - 5.4.2. Conceptos generales del ensamblaje
 - 5.4.3. Tipos de algoritmos de ensamblaje
 - 5.4.4. Evaluación previa al ensamblaje. Análisis de k-meros.
 - 5.4.5. Ensamblaje de secuencias cortas con SPAdes.
 - 5.4.6. Evaluación de la calidad de ensamblaje: continuidad y contenido.
- 5.5. Reconstrucción del genoma: anotación del genoma ensamblado.
 - 5.5.1. Anotación utilizando Prokka.
 - 5.5.2. Visualización de las anotaciones.
- 5.6. Reconstrucción del genoma: Elementos Genéticos Móviles (EGMs)
 - 5.6.1. Herramientas de ensamblaje y extracción de plásmidos.
 - 5.6.2. Herramientas de identificación mediante marcadores genéticos.
 - 5.6.3. Herramientas basadas en los grafos de ensamblaje de novo.
- 5.7. Detección y anotación de regiones de interés: genes de resistencia a antibióticos y virulencia.

Tema 6. Análisis de genomas de virus. El caso de SARS-Cov-2.

- 6.1. Estructura de los virus. Aspectos generales
- 6.2. El genoma de SARS-CoV-2
- 6.3. Clasificación de SARS-CoV-2. Variantes de interés y linajes. Bases de datos.
 - 6.3.1. Clasificación de la OMS: Variantes de preocupación, variantes de interés y variantes bajo vigilancia.
 - 6.3.2. Monitorización de la evolución de SARS-CoV-2 a tiempo real. Nextstrain.
 - 6.3.3. GISAID. El repositorio de todos los genomas SARS-CoV-2 secuenciados.
 - 6.3.4. Clasificación de linajes y constelaciones de mutaciones: PANGO.
 - 6.3.5. El análisis de mutaciones de interés. Bases de datos de monitorización.
- 6.4. Análisis bioinformático de linajes de SARS-CoV-2.
 - 6.4.1. Limpieza de las lecturas crudas. Eliminación del genoma del hospedador.
 - 6.4.2. Mapeo al genoma de referencia y determinación del linaje.
 - 6.4.3. Ensamblaje de novo y anotación.

3. Metodología

La metodología de la Universidad Internacional de Valencia (VIU) se caracteriza por una apuesta decidida en un modelo de carácter e-presencial. Así, siguiendo lo estipulado en el calendario de actividades docentes del Título, se impartirán en directo un conjunto de sesiones, que, además, quedarán grabadas para su posterior visionado por parte de aquellos estudiantes que lo necesiten. En todo caso, se recomienda acudir, en la medida de lo posible, a dichas sesiones, facilitando así el intercambio de experiencias y dudas con el docente.

En lo que se refiere a las metodologías específicas de enseñanza-aprendizaje, serán aplicadas por el docente en función de los contenidos de la asignatura y de las necesidades pedagógicas de los estudiantes. De manera general, se impartirán contenidos teóricos y, en el ámbito de las clases prácticas se podrá realizar la resolución de problemas, el estudio de casos y/o la simulación.

Por otro lado, la Universidad y sus docentes ofrecen un acompañamiento continuo al estudiante, poniendo a su disposición foros de dudas y tutorías para resolver las consultas de carácter académico que el estudiante pueda tener. Es importante señalar que resulta fundamental el trabajo autónomo del estudiante para lograr una adecuada consecución de los objetivos formativos previstos para la asignatura.

4. Actividades formativas

Durante el desarrollo de cada una de las asignaturas se programan una serie de actividades de aprendizaje que ayudan a los estudiantes a consolidar los conocimientos trabajados.

A continuación, se relacionan las actividades que forman parte de la asignatura:

1. Actividades de carácter teórico

Se trata de un conjunto de actividades guiadas por el profesor de la asignatura destinadas a la adquisición por parte de los estudiantes de los contenidos teóricos de la misma. Estas actividades, diseñadas de manera integral, se complementan entre sí y están directamente relacionadas con los materiales teóricos que se ponen a disposición del estudiante (manual y material complementario).

2. Actividades de carácter práctico

Se trata de un conjunto de actividades guiadas y supervisadas por el profesor de la asignatura vinculadas con la adquisición por parte de los estudiantes de las competencias asociadas. Estas actividades, diseñadas con visión de conjunto, están relacionadas entre sí para ofrecer al estudiante una formación completa e integral.

3. Tutorías

Se trata de sesiones, tanto de carácter síncrono como asíncrono (e-mail), individuales o colectivas, en las que el profesor comparte información sobre el progreso académico del estudiante y en las que se resuelven dudas y se dan orientaciones específicas ante dificultades concretas en el desarrollo de la asignatura.

4. Trabajo autónomo

Se trata de un conjunto de actividades que el estudiante desarrolla autónomamente y que están enfocadas a lograr un aprendizaje significativo y a superar la evaluación de la asignatura. La realización de estas actividades es indispensable para adquirir las competencias y se encuentran entroncadas en el aprendizaje autónomo que consagra la actual ordenación de enseñanzas universitarias. Esta actividad, por su definición, tiene carácter asíncrono.

5. Prueba objetiva final

Como parte de la evaluación de cada una de las asignaturas (a excepción del Trabajo fin de Máster), se realiza una prueba objetiva (examen). Esta prueba se realiza en tiempo real (con los medios de control antifraude especificados) y tiene como objetivo evidenciar el nivel de adquisición de conocimientos y desarrollo de competencias por parte de los estudiantes. Esta actividad, por su definición, tiene carácter síncrono.

5. Evaluación

5.1. Sistema de evaluación

El Modelo de Evaluación de estudiantes en la Universidad se sustenta en los principios del Espacio Europeo de Educación Superior (EEES), y está adaptado a la estructura de formación virtual propia de esta Universidad. De este modo, se dirige a la evaluación de competencias.

Sistema de Evaluación	Ponderación
Portafolio*	70 %
<i>Se desarrolla a lo largo de todo el curso. Los elementos que componen esta evaluación son los trabajos que realizan los estudiantes en el marco de las clases prácticas (estudio de casos, resolución de problemas, revisión bibliográfica, simulación, trabajo cooperativo, diseño de proyectos, etc.).</i>	
Sistema de Evaluación	Ponderación
Prueba final*	30 %
<i>Valoración del nivel de adquisición por parte del estudiante de las competencias asociadas a la asignatura, empleando diversas tipologías de pregunta (preguntas de tipo test, preguntas de desarrollo, preguntas de respuesta breve o cualquier combinación de estas).</i>	

***Es requisito indispensable para superar la asignatura aprobar cada apartado (portafolio y prueba final) con un mínimo de 5.0 para ponderar las calificaciones.**

Los enunciados y especificaciones propias de las distintas actividades serán aportados por el docente, a través del Campus Virtual, a lo largo de la impartición de la asignatura.

Atendiendo a la Normativa de Evaluación de la Universidad, se tendrá en cuenta que la utilización de **contenido de autoría ajena** al propio estudiante debe ser citada adecuadamente

en los trabajos entregados. Los casos de plagio serán sancionados con suspenso (0) de la actividad en la que se detecte. Asimismo, el uso de **medios fraudulentos durante las pruebas de evaluación** implicará un suspenso (0) y podrá implicar la apertura de un expediente disciplinario.

5.2. Sistema de calificación

La calificación de la asignatura se establecerá en los siguientes cálculos y términos:

Nivel de aprendizaje	Calificación numérica	Calificación cualitativa
Muy competente	9,0 - 10	Sobresaliente
Competente	7,0 - 8,9	Notable
Aceptable	5,0 -6,9	Aprobado
Aún no competente	0,0 -4,9	Suspenso

Sin detrimento de lo anterior, el estudiante dispondrá de una **rúbrica simplificada** en el aula que mostrará los aspectos que valorará el docente, como así también los **niveles de desempeño que tendrá en cuenta para calificar las actividades vinculadas a cada resultado de aprendizaje**.

La mención de «**Matrícula de Honor**» podrá ser otorgada a estudiantes que hayan obtenido una calificación igual o superior a 9.0. Su número no podrá exceder del cinco por ciento de los estudiantes matriculados en una materia en el correspondiente curso académico, salvo que el número de estudiantes matriculados sea inferior a 20, en cuyo caso se podrá conceder una sola «Matrícula de Honor».

6. Bibliografía

6.1. Bibliografía de referencia

- Álvarez-Molina, A., de Toro, M., Alexa, E. A., & Álvarez-Ordóñez, A. (2021). Applying Genomics to Track Antimicrobial Resistance in the Food Chain. *Comprehensive Foodomics*, 188–211. <https://doi.org/10.1016/b978-0-08-100596-5.22700-5>
- Baker, M. (2012). De novo genome assembly: What every biologist should know. *Nature Methods*, 9(4), 333–337. <https://doi.org/10.1038/nmeth.1935>
- Brown, T. A. (2017a). Genomes 3. In *The Yale Journal of Biology and Medicine* (Vol. 90, Issue 4). Yale Journal of Biology and Medicine. [/pmc/articles/PMC5733849/](https://doi.org/10.1093/yjbm/bax004)
- Cordero, P., & Ashley, E. (2012). Whole-genome sequencing in personalized therapeutics. *Clinical Pharmacology and Therapeutics*, 91(6), 1001–1009.

- <https://doi.org/10.1038/CLPT.2012.51>
- Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., Chaffron, S., Van Der Vossen, J., Tang, S., Katase, M., McClure, P., Kimura, B., Ching Chai, L., Chapman, J., & Grant, K. (2019). The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiology*, 79, 96–115. <https://doi.org/10.1016/J.FM.2018.11.005>
- Olea, N., Casas, M., Castaño, A., Mendiola, J., Vrijheid, M., Arenas, J., Carracedo, Á., Lapunzina, P., & Martín-Sánchez, F. (2020). *Informes Anticipando. Exposoma*. <https://www.instituto-roche.es/observatorio/exposoma>
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M. R., Zschocke, J., & Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, 15(2), 256. <https://doi.org/10.1093/BIB/BBS086>
- Petersen, B. S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genetics*, 18(1), 1–13. <https://doi.org/10.1186/s12863-017-0479-5>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W., Hedge, M., Lyon, E., Spector, E., Voelkerding, K., & HL, R. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 17(5), 405–424. <https://doi.org/10.1038/GIM.2015.30>
- Seaby, E. G., Pengelly, R. J., & Ennis, S. (2016). Exome sequencing explained: A practical guide to its clinical application. *Briefings in Functional Genomics*, 15(5), 374–384. <https://doi.org/10.1093/bfgp/elv054>
- Sun, Y., Ruivenkamp, C. AL, Hoffer, M. J., Vrijenhoek, T., Kriek, M., van Asperen, C. J., den Dunnen, J. T., & Santen, G. W. (2015). Next Generation Diagnostics: gene panel, exome or whole genome? *Hum Mutat*, 36(6), 648–655. <https://doi.org/10.1002/humu.22783>.This
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>