



Universidad
Internacional
de Valencia

Guía didáctica

Análisis transcriptómicos de la expresión génica

Título: *Máster Universitario en Bioinformática*

Materia: *Bioinformática genómica*

Créditos: 6 ECTS

Código: 05MBIF

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.....	3
2. Contenidos/temario	5
3. Metodología	6
4. Actividades formativas	6
5. Evaluación.....	7
5.1. Sistema de evaluación.....	7
5.2. Sistema de calificación	8
6. Bibliografía.....	9
6.1. Bibliografía de referencia	9
6.2. Bibliografía complementaria.....	9

1. Organización general

1.1. Datos de la asignatura

MATERIA	Bioinformática genómica
ASIGNATURA	Análisis transcriptómicos de la expresión génica 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. Nuria Mauri Panadero <i>nuria.mauri@campusviu.es</i>
-------------------	---

1.3. Introducción a la asignatura

Esta asignatura ofrece una formación fundamental dentro de las Ciencias que estudian la actividad de los genes tales como la Biología Molecular, la Biomedicina o el Laboratorio Clínico. Recoge una visión actualizada de la Transcriptómica, cuyo desarrollo ha estado marcado recientemente por la aparición de la secuenciación de nueva generación (NGS). Hoy, los datos de secuencia son el nuevo escenario de los estudios de expresión génica y su complejidad, por el volumen de información que supone, puede apabullar al investigador o técnico. Este curso pretende dar al estudiante los conceptos y la capacidad técnica necesarios para analizar computacionalmente el conjunto de los transcritos, de manera adecuada, y extraer toda la información beneficiosa para el interés del estudio.

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.

CE20: Saber aplicar herramientas bioinformáticas avanzadas en el análisis de expresión génica, de poblaciones y de expresión diferencial de proteínas en datos ómicos.

2. Contenidos/temario

CAPÍTULO 1. Introducción a las técnicas transcriptómicas actuales y emergentes.

- 1.1. Pre-transcriptómica.
- 1.2. Microarrays.
- 1.3. Secuenciación de ARN de alto rendimiento (RNA-seq)
 - 1.3.1 Secuenciación de segunda generación (“lectura corta”).
 - 1.3.2 Secuenciación de tercera generación (“lectura larga”).
- 1.4. Transcriptómica espacial.

CAPÍTULO 2. Estudios de expresión génica con datos de NGS.

- 2.1. ¿Por qué elegir la NGS y qué tipo?
- 2.2. Protocolo básico de NGS.
 - 2.2.1. Preparación de la biblioteca.
 - 2.2.2. Generación de los haces de copias.
 - 2.2.3. Secuenciación (SGS) Preparación de la biblioteca.
- 2.3. Experimentos de NGS en las bases de datos.

CAPÍTULO 3. Análisis de datos de NGS.

- 3.1. Introducción al flujo de trabajo con datos NGS y sus archivos resultantes.
 - 3.1.1. Archivo de datos de lecturas FASTQ.
 - 3.1.2. Archivo de datos genómicos FASTA.
 - 3.1.3. Archivo de alineamientos SAM/BAM.
 - 3.1.4. Archivo de anotaciones GFF/GFT.
- 3.2. Preprocesado de las lecturas.
 - 3.2.1. Control de calidad.
 - 3.2.2. Eliminación de adaptadores.
- 3.3. Mapeo de las lecturas.
 - 3.3.1. Genoma de referencia.
 - 3.3.2. Mapeadores de empalme.
- 3.4. Recuento de los alineamientos.

CAPÍTULO 4. Análisis estadístico de la diferencia de expresión.

- 4.1. Normalización de los recuentos.
- 4.2. Eliminación de genes de baja expresión.
- 4.3. Estudio de la variabilidad entre muestras.
 - 4.3.1. Variabilidad técnica.
 - 4.3.2. Variabilidad biológica.
- 4.4. Comparativas entre grupos tipo ANOVA.
 - 4.4.1. Pruebas de significancia.

4.4.2. Nivel de cambio.

4.5. Simulación en R.

CAPÍTULO 5. Exploración y visualización de resultados.

5.1. Visualización por significancia y nivel de cambio: gráfico Volcano.

5.2. Comparación de conjuntos de genes: diagrama de Venn.

5.3. Agrupamiento de perfiles de expresión: gráfico heatmap.

5.4. Enriquecimiento funcional: términos GO.

5.6. Navegadores genómicos.

5.7. Simulación en R.

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 %

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).

***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

- Gentleman, R.C., Carey, V.J., Bates, D.M. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5, R80 (2004). <https://doi.org/10.1186/gb-2004-5-10-r80>
- Wang, Z., Gerstein, M. & Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63 . <https://doi.org/10.1038/nrg2484>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3), R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, 11(9), 1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS computational biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Arita, M., Karsch-Mizrachi, I., & Cochrane, G. (2021). The international nucleotide sequence database collaboration. *Nucleic acids research*, 49(D1), D121–D124. <https://doi.org/10.1093/nar/gkaa967>.

6.2. Bibliografía complementaria

- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of molecular biology*, 147(1), 195–197. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5)
- Boguski, M. S., Lowe, T. M., & Tolstoshev, C. M. (1993). dbEST--database for "expressed sequence tags". *Nature genetics*, 4(4), 332–333. <https://doi.org/10.1038/ng0893-332>

- Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (New York, N.Y.)*, 270(5235), 467–470. <https://doi.org/10.1126/science.270.5235.467>
- Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, 30(1), 207–210. <https://doi.org/10.1093/nar/30.1.207>
- Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), 3960–3964. <https://doi.org/10.1073/pnas.0230489100>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9), 1105–1111. <https://doi.org/10.1093/bioinformatics/btp120>
- Ozsolak, F., Platt, A. R., Jones, D. R., Reifengerger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M., & Milos, P. M. (2009). Direct RNA sequencing. *Nature*, 461(7265), 814–818. <https://doi.org/10.1038/nature08390>
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5), 377–382. <https://doi.org/10.1038/nmeth.1315>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., ... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*, 323(5910), 133–138. <https://doi.org/10.1126/science.1162986>
- Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), 9546–9551. <https://doi.org/10.1073/pnas.0914005107>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome research*, 21(12), 2213–2223. <https://doi.org/10.1101/gr.124321.111>
- Clark, M. J., Chen, R., & Snyder, M. (2013). Exome sequencing by targeted enrichment. *Current protocols in molecular biology*, Chapter 7, Unit7.12. <https://doi.org/10.1002/0471142727.mb0712s102>
- Li, J., & Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical methods in medical research*, 22(5), 519–536. <https://doi.org/10.1177/0962280211428386>
- Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: more sequence or more replication?. *Bioinformatics (Oxford, England)*, 30(3), 301–304. <https://doi.org/10.1093/bioinformatics/btt688>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>

- Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., & Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC bioinformatics*, 16(1), 169. <https://doi.org/10.1186/s12859-015-0611-3>
- Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., ... Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)*, 353(6294), 78–82. <https://doi.org/10.1126/science.aaf2403>
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., & Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific reports*, 6, 25533. <https://doi.org/10.1038/srep25533>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome biology*, 21(1), 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Maynard, K.R., Collado-Torres, L., Weber, L.M. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 24, 425–436 (2021). <https://doi.org/10.1038/s41593-020-00787-0>