

PROCESO SELECTIVO POR EL SISTEMA DE ACCESO LIBRE PARA INGRESO EN LA ESCALA DE TECNICOS SUPERIORES ESPECIALIZADOS DE LOS ORGANISMOS PÚBLICOS DE INVESTIGACIÓN, CONVOCADO POR RESOLUCION DE 16 DE DICIEMBRE DE 2020 (BOE Nº 341 DE 31 DE DICIEMBRE)

Cuestionario del primer ejercicio

Programa: Bioinformática aplicada a la biología y biomedicina

- No abra el **CUESTIONARIO** ni empiece el examen hasta que se le indique.
- Solo se calificarán las respuestas marcadas en la **HOJA DE RESPUESTAS**
- El cuestionario consta de **100 preguntas** (25 de ellas corresponderán a los temas recogidos en el grupo de materias comunes y las otras 75 pertenecerán a los temas previstos en el grupo de materias específicas del programa por el que se presenta), cada una de ellas con **cuatro respuesta alternativas**, de las cuales **sólo una de ellas es correcta**.
- Una vez abierto el cuestionario, compruebe que consta de todas las páginas y preguntas y que sea legible. En caso contrario solicite uno nuevo al personal del aula.
- Las **contestaciones erróneas se PENALIZARÁN** con un 25 % de su valoración.
- Lea atentamente las **instrucciones** para contestar la **HOJA DE RESPUESTAS**, que figuran al dorso de la misma.
- Cumplimente los datos personales y firme la **HOJA DE RESPUESTAS**.
- El tiempo para la realización de este ejercicio será de **noventa (90) minutos**.
- **NO SEPARE** ninguna de las copias de la **HOJA DE RESPUESTAS**. Una vez finalizado el ejercicio, el personal del aula le indicará los pasos a seguir.
- El **CUESTIONARIO** se podrá utilizar como borrador y se podrá llevar por el opositor al finalizar el tiempo marcado para el ejercicio.

1. Según el artículo 116 de la Constitución española:

A) El estado de alarma será declarado por el Gobierno mediante decreto acordado en Consejo de Ministros por un plazo máximo de quince días, dando cuenta al Congreso de los Diputados, reunido inmediatamente al efecto y sin cuya autorización no podrá ser prorrogado dicho plazo

B) El estado de alarma será declarado por el Gobierno mediante decreto acordado en Consejo de Ministros por un plazo máximo de quince días, dando cuenta al Senado, reunido inmediatamente al efecto y sin cuya autorización no podrá ser prorrogado dicho plazo

C) El estado de alarma será declarado por el Gobierno mediante decreto acordado en Consejo de Ministros por un plazo máximo de seis meses, dando cuenta al Congreso de los Diputados, reunido inmediatamente al efecto y sin cuya autorización no podrá ser prorrogado dicho plazo

D) El estado de alarma será declarado por el Gobierno mediante decreto acordado en Consejo de Ministros previa autorización del Congreso de los Diputados. La autorización y proclamación del estado de alarma deberá determinar expresamente los efectos del mismo, el ámbito territorial a que se extiende y su duración, que no podrá exceder de treinta días, prorrogables por otro plazo igual, con los mismos requisitos

2. ¿Cuál de las siguientes es una competencia exclusiva del Estado en virtud del artículo 149.1 de la Constitución española?

A) Sanidad e higiene dentro del territorio nacional

B) Fomento y coordinación general de la investigación científica y técnica

C) Asistencia social

D) Todas las respuestas son correctas

3. Según el artículo 103.1 de la Constitución española:

A) La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho

B) La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de transparencia, jerarquía, centralización, concentración y coordinación, con sometimiento pleno a la ley y al Derecho

C) La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, sometida exclusivamente a la Constitución

D) La Administración Pública sirve con objetividad sus intereses particulares y actúa de acuerdo con los principios de transparencia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho

4. Según la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, la transparencia de la actividad pública se aplica a los siguientes sujetos, EXCEPTO:

A) Las corporaciones de Derecho Público, en lo relativo a sus actividades sujetas a Derecho Administrativo

B) Las fundaciones del sector público previstas en la legislación en materia de fundaciones

C) Las corporaciones de Derecho Público, en lo relativo a sus actividades no sujetas a Derecho Administrativo

D) Los partidos políticos, organizaciones sindicales y organizaciones empresariales

5. Según el artículo 14 de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, ¿cuál de los siguientes supuestos NO se contempla como causa de limitación del derecho de acceso a la información?:

- A) La seguridad nacional
- B) Los intereses de los partidos políticos
- C) La defensa
- D) La seguridad pública

6. Según el artículo 35 de la Ley 39/2015, de 1 de octubre, de Procedimiento Administrativo Común de las Administraciones Públicas, señale la respuesta INCORRECTA sobre qué actos de derecho administrativo deben ser motivados con sucinta referencia de hechos y fundamentos de derecho:

- A) Los actos que limiten derechos subjetivos o intereses legítimos
- B) Los actos que se separen del criterio seguido en actuaciones precedentes o del dictamen de órganos consultivos
- C) Los actos que rechacen pruebas propuestas por los interesados
- D) Ningún acto de derecho administrativo debe ser motivado con sucinta referencia de hechos y fundamento de derecho

7. De acuerdo con el Texto Refundido de la Ley del Estatuto Básico del Empleado Público, aprobado por el Real Decreto Legislativo 5/2015, de 30 de octubre, los servicios extraordinarios prestados por los funcionarios fuera de la jornada normal de trabajo se retribuyen a través de:

- A) Las retribuciones complementarias
- B) Las gratificaciones
- C) Las horas extra
- D) Las retribuciones extraordinarias

8. La aprobación de los Presupuestos Generales del Estado corresponde a:

- A) El Consejo de Ministros
- B) El Gobierno
- C) Las Cortes Generales
- D) El Ministerio de Hacienda

9. A tenor del artículo 47.2 de la Ley 47/2003, de 26 de noviembre, General Presupuestaria, el número de ejercicios a que pueden aplicarse los gastos de carácter plurianual NO será superior a:

- A) Dos
- B) Cuatro
- C) Cinco
- D) Tres

10. Según la Ley 47/2003, de 26 de noviembre, General Presupuestaria, la agrupación de los créditos en capítulos, artículos, conceptos y subconceptos corresponde a la clasificación:

- A) Orgánica
- B) Contable
- C) Económica
- D) Por programas

11. De acuerdo con el artículo 32 de la Ley Orgánica 2/2012, de 27 de abril, de Estabilidad Presupuestaria y Sostenibilidad Financiera, en el supuesto de que la liquidación presupuestaria del Estado se sitúe en superávit, éste debe destinarse a:

- A) Reducir el gasto público
- B) Reducir el déficit presupuestario
- C) Reducir el endeudamiento neto
- D) Al fondo de contingencia

12. ¿Qué derechos tiene el personal técnico al servicio de los Organismos Públicos de Investigación de la Administración General del Estado según establece la Ley 14/2011 de la Ciencia, la Tecnología y la Innovación? Indicar la frase correcta.

- A) El derecho a ser autor o coautor de los trabajos de carácter técnico en los que no haya participado
- B) Tiene derecho a participar en los beneficios que obtengan las entidades para las que presta servicios, como consecuencia de la eventual explotación de los resultados de la actividad, aunque no haya participado el personal técnico
- C) Tiene derecho a la consideración y respeto de su actividad
- D) Tiene derecho a participar en la toma de decisiones de los servicios

13. ¿Qué agentes de financiación están adscritos al Ministerio de Ciencia e Innovación?

- A) La Agencia Estatal de Investigación (AEI)
- B) El Consejo de Desarrollo Tecnológico (CDT)
- C) La Agencia Científica de Programación Técnica (ACPT)
- D) El Instituto de Sanidad Carlos II (ISCII)

14. El Real Decreto 2/2020, de 12 de enero, por el que se reestructuran los departamentos ministeriales, en su artículo 19 define exactamente al Ministerio de Ciencia e Innovación como:

- A) El encargado de proponer los salarios del personal investigador
- B) El encargado de la propuesta y ejecución de la política del Gobierno en materia de ciencia, desarrollo tecnológico, investigación universitaria e innovación
- C) El encargado de la propuesta y ejecución de la política del Gobierno en materia de Ciencia, Desarrollo Tecnológico e Innovación
- D) El encargado de la propuesta y ejecución de la política del Gobierno en materia de Ciencia, Desarrollo Tecnológico, Transferencia de Tecnología e Innovación

15. Según la Ley 14/2011, de 1 de junio, de la Ciencia, Tecnología y la Innovación, NO es función de TODOS los Organismos Públicos de Investigación de la Administración General del Estado:

- A) La ejecución directa de actividades de investigación científica y técnica
- B) La ejecución directa de actividades de financiación de la investigación científica y técnica
- C) La ejecución directa de actividades de prestación de servicios tecnológicos
- D) La ejecución directa de otras actividades de carácter complementario, necesarias para el adecuado progreso científico y tecnológico de la sociedad

16. Dentro del contexto del Sistema Español de Ciencia, Tecnología e Innovación, las Comunidades Autónomas forman parte de forma directa de:

- A) El Consejo de Política Científica, Tecnológica y de Innovación
- B) El Consejo Asesor de Ciencia, Tecnología e Innovación
- C) La Comisión Delegada del Gobierno para Política Científica, Tecnológica y de Innovación
- D) El Comité Español de Ética de la Investigación

17. Dentro del contexto del Sistema Español de Ciencia, Tecnología e Innovación, las Comunidades Autónomas NO son:

- A) Agentes de financiación
- B) Agentes de coordinación
- C) Agentes de ejecución
- D) Agentes inactivos

18. Uno de los objetivos del Espacio Europeo de Investigación (ERA) es:

- A) Crear estructuras estatales en las que la investigación realizada en cada Estado Miembro sea transferida exclusivamente a empresas de terceros países
- B) Crear un área unificada europea en la que los investigadores puedan moverse libremente e interactuar sin dificultades, trabajando con infraestructuras de calidad mundial y con redes de trabajo excelentes
- C) Obstaculizar el libre movimiento de los investigadores europeos
- D) Crear un área unificada europea en la que los investigadores puedan establecer redes de trabajo con terceros países financiadas por estos últimos

19. Indicar cuál de las siguientes afirmaciones NO es un objetivo estratégico del Programa Horizonte 2020:

- A) Crear una ciencia de excelencia que permita reforzar la posición de la UE en el panorama científico mundial
- B) Potenciar el establecimiento de estructuras de gobernanza avanzada en las universidades europeas
- C) Desarrollar tecnologías y sus aplicaciones para mejorar la competitividad europea
- D) Investigar en las grandes cuestiones que afectan a los ciudadanos europeos

20. En virtud de los acuerdos alcanzados entre la Unión Europea (UE) y distintos Gobiernos, existe una serie de países asociados cuyas personas jurídicas pueden participar en el Programa Horizonte 2020:

- A) En igualdad de condiciones con los Estados miembros de la UE
- B) Supeditado a que los resultados obtenidos por los países asociados sean transferidos a empresas de países no asociados
- C) Horizonte 2020 fue un proyecto fallido que no llegó a desarrollarse
- D) Sólo pueden participar en Horizonte 2020 los Estados miembros de la Unión Europea

21. ¿Cuáles son las formas más comunes de transferencia de tecnología?

- A) Las cesiones de patrimonio.
- B) Los contratos de explotación, cesiones de derechos y/o licencias, así como los acuerdos de transferencia de material
- C) Las donaciones, préstamos y reinversiones
- D) Los préstamos de infraestructura

22. Con relación a las patentes y su ámbito de validez, éstas son:

- A) Derechos territoriales, generalmente exclusivos que solo tienen validez en el país o la región en los que se ha concedido
- B) Derechos nominales de ámbito ilimitado en todo el mundo, que son extrapolables y escalables
- C) Derechos exclusivos que se protegen de manera inespecífica
- D) Derechos generales, que se aplican a diferentes ámbitos territoriales independientemente del país o la región en los que se ha concedido

23. ¿Cuál es el organismo responsable de la recepción, estudio y concesión de la propiedad industrial en España?

- A) La Oficina de Transferencia de Resultados de Investigación (OTRI)
- B) La Oficina Española de Patentes y Marcas (OEPM)
- C) La “European Patent Office” (EPO)
- D) La “World Intellectual Property Organization” (WIPO)

24. El personal laboral al servicio de las Administraciones Públicas se rige por:

- A) Exclusivamente por la legislación laboral
- B) La legislación laboral y por las demás normas convencionalmente aplicables, y por los preceptos del Estatuto Básico del Empleado Público que así lo dispongan
- C) Exclusivamente por el articulado específico a esta modalidad de contratación incluido en el Estatuto Básico del Empleado Público
- D) Una Administración Pública a la que están vinculados a través de un nombramiento legal

25. Según la Ley 7/2007, de 12 de abril, del Estatuto Básico del Empleado Público, se considera una retribución básica de los funcionarios:

- A) El complemento de productividad
- B) Las pagas extraordinarias
- C) El complemento de destino
- D) El complemento específico

26. Según un criterio estrictamente matemático ¿cuántos codones diferentes podrían formarse combinando los nucleótidos A, C, G y T?

- A) 64
- B) 20
- C) 21
- D) 32

27. ¿Cuál de los siguientes términos NO hace referencia a un método de secuenciación de ADN?

- A) Secuenciación de Maxam-Gilbert
- B) Método de Sanger
- C) Pirosecuenciación
- D) Degradación de Edman

28. Cuando se mide la expresión génica con un microarray de ADN ¿Qué proceso biológico se suele cuantificar?

- A) El plegamiento de las cadenas de aminoácidos
- B) La traducción de las proteínas
- C) La transcripción de los genes
- D) La replicación del ADN genómico

29. ¿Cuál de las siguientes es una definición válida del término "metagenómica"?

- A) La ciencia que estudia los genes implicados en el catabolismo del metano
- B) La aplicación de técnicas genómicas sin la necesidad de aislar y cultivar las especies individuales presentes en una muestra
- C) El conjunto de métodos aplicados al estudio de genomas que constan de más de un millón de nucleótidos entre todos sus cromosomas
- D) La ciencia que estudia las características de secuencia del origen de replicación de los genomas circulares

30. En bioinformática ¿qué son las bases de datos secundarias?

- A) Son las que contienen información obtenida directamente de datos experimentales
- B) Son las que contienen información de secuencias biológicas
- C) Son las que contienen información de estructuras de moléculas biológicas
- D) Son las que contienen información derivada de las bases de datos primarias

31. ¿Cuáles son subgrupos de la base de datos EnsemblGenomes?

- A) EnsemblPlants, EnsemblMetazoa, EnsemblProtists, EnsemblFungi, EnsemblBacteria
- B) EnsemblHuman, EnsemblMouse, EnsemblRat, EnsemblArabidopsis, EnsemblDrosophila
- C) EnsemblPlants, EnsemblMammals, EnsemblBacteria
- D) EnsemblBacteria, EnsemblEukarya, EnsemblArchaea

32. Una de las herramientas más utilizadas para comparar secuencias biológicas es BLAST ¿Qué significan las letras que componen dicho nombre?

- A) Basic Limited Alignment Search Tool
- B) Basic Local Alignment Search Tool
- C) Best Linked Aminoacid Search Tool
- D) Binary Local Aminoacid Search Tool

33. ¿Cuál de las siguientes características de los genes NO se considera una anotación funcional?

- A) Lista sistemática de términos de Gene Ontology asociados al gen
- B) Longitud de la secuencia de nucleótidos del gen
- C) Conjunto de rutas metabólicas o procesos biológicos en general, en las que interviene el gen
- D) Descripción de la función del gen, escrita en texto libre

34. ¿Qué información contienen los archivos en formato GFF?

- A) Coordinadas, identificadores, sinónimos y otros detalles relativos a genes y otros elementos genómicos
- B) Secuencias de ácidos nucleicos en formato 2bit
- C) Secuencias de ácidos nucleicos o de aminoácidos en formato FASTA
- D) Coordinadas y medidas de calidad de alineamientos de fragmentos de secuencia obtenidos por secuenciación masiva

35. En un análisis de datos basado en comparar grupos de réplicas biológicas, ¿cuál de las siguientes afirmaciones es correcta?

- A) La potencia estadística disminuye al aumentar el número de réplicas
- B) La potencia estadística es mayor cuanto menor es la variabilidad de las réplicas
- C) La potencia estadística es mayor cuanto mayor es la variabilidad de las réplicas
- D) La potencia estadística es independiente del número de réplicas

36. En un experimento de RNA-seq, ¿cuántas réplicas biológicas se necesitan para detectar cambios de expresión génica mayores de 2, con una tasa de falsos descubrimientos (FDR) menor de 0.05?

- A) Al menos 4 réplicas biológicas
- B) Depende, entre otras cosas, de la variabilidad de las muestras
- C) Al menos 3 réplicas biológicas
- D) Depende, entre otras cosas, de la calidad de las anotaciones funcionales de los genes

37. ¿Cuál de los siguientes valores relativos a un conjunto de números NO es una medida de dispersión?

- A) Promedio
- B) Varianza
- C) Rango
- D) Desviación estándar

38. En estadística descriptiva ¿qué es una población con una distribución normal?

- A) Aquella cuya función de densidad es simétrica respecto al valor promedio
- B) Aquella cuya función de densidad tiene forma de campana de Gauss
- C) Aquella cuya función de densidad es asimétrica respecto al valor promedio
- D) Aquella cuya función de densidad es continua y derivable

39. En un análisis estadístico ¿qué se entiende por "error tipo I"?

- A) Es el resultado de dividir la desviación estándar de un conjunto de valores entre el número de valores
- B) Es la diferencia entre el máximo y el mínimo de un conjunto de valores numéricos
- C) Es cualquier error cometido al transcribir valores numéricos, obtenidos con un aparato de medida, en una hoja de cálculo
- D) Es el error que se comete cuando se rechaza la hipótesis nula, siendo ésta verdadera en la población investigada

40. ¿Qué relación hay entre la varianza y la desviación estándar de una población?

- A) Son dos maneras diferentes de llamar a una misma medida
- B) La desviación estándar es el logaritmo natural de la varianza
- C) La varianza es el logaritmo natural de la desviación estándar
- D) La desviación estándar es la raíz cuadrada positiva de la varianza

41. ¿Cuál de las siguientes afirmaciones sobre el coeficiente de correlación de Pearson NO es correcta?

- A) Es una medida de correlación lineal entre dos conjuntos de datos
- B) Es la covarianza de dos variables, dividida entre el producto de sus desviaciones estándar
- C) Tiene siempre un valor comprendido entre 0 y 1
- D) Es una prueba paramétrica

42. En lo relativo a la seguridad de los sistemas de información, ¿cuál de las siguientes afirmaciones es correcta?

- A) Se entiende por seguridad física de un sistema informático al conjunto de hábitos recomendados para proteger la salud de los técnicos que trabajan con pantallas de visualización de datos
- B) La seguridad lógica de un sistema informático consiste en la aplicación de barreras y procedimientos que protejan el acceso a los datos y a la información contenida en él
- C) Todos los sistemas de información diseñados después del año 2000 están totalmente libres de vulnerabilidades
- D) La principal ventaja de los discos duros mecánicos frente a los discos de estado sólido es que, en los primeros, la información siempre se mantiene libre de errores por tiempo ilimitado

43. El proyecto Bioconductor consta de un conjunto de paquetes de funciones escritos en lenguaje:

- A) R
- B) Visual Basic para Excel
- C) SPSS
- D) SAS

44. ¿Cuál es la causa del problema de las comparaciones múltiples?

- A) Se debe al uso de demasiadas réplicas biológicas y el consecuente sobreentrenamiento del algoritmo de normalización
- B) Se debe a la eliminación excesiva de datos extremos en la normalización de valores de expresión génica
- C) Se debe a la presencia de varias sondas para un mismo gen en ciertos microarrays de ADN y dificulta la estimación de su valor de expresión real
- D) Se debe al incremento en el error de tipo I como consecuencia del uso repetido de test estadísticos

45. ¿Cuál de estos tipos de gráficos es el más adecuado para representar una serie de datos con sus cuartiles y valores atípicos?

- A) Diagrama de cajas (en inglés, "box plot")
- B) Diagrama de dispersión (en inglés, "scatter plot")
- C) Gráfico circular (en inglés, "pie chart")
- D) Mapa de calor (en inglés, "heat map")

46. ¿Qué tipo de herramienta bioinformática es un navegador genómico?

- A) Es una base de datos de genomas secuenciados y sus elementos genómicos
- B) Es una herramienta estadística para calcular parámetros cuantitativos de los elementos genómicos
- C) Es una herramienta gráfica para agrupar elementos genómicos, normalmente genes codificantes de proteínas, según sus patrones de expresión
- D) Es una herramienta gráfica para mostrar, de forma interactiva, información de datos genómicos

47. En procesamiento digital de imágenes, ¿qué se conoce por segmentación?

- A) La detección de todas las líneas rectas que aparecen en la imagen, clasificándolas en horizontales, verticales y otras
- B) La compresión del archivo digital sin pérdida de calidad
- C) La división de la imagen en sus partes constituyentes hasta un nivel de subdivisión en el que se aíslen las regiones u objetos de interés
- D) La compresión del archivo digital con algo de pérdida de calidad

48. En diseño gráfico, ¿qué es un gráfico vectorial?

- A) Es aquel que representa vectores de longitud igual a 1
- B) Es un gráfico que muestra las relaciones de los cambios en los datos en un período de tiempo, en forma de líneas, marcando cada tiempo con un símbolo
- C) Es aquel formado por objetos geométricos (segmentos, polígonos, arcos, muros, etc.), cada uno de ellos definido por atributos matemáticos
- D) Es cualquier archivo de imagen guardado en formato JPEG

49. ¿Qué es la Interactómica?

- A) El conjunto de herramientas, disponibles en internet, para el estudio y la visualización de los genomas
- B) Es el estudio del conjunto de interacciones entre distintas biomoléculas en un contexto determinado
- C) Es el conjunto de protómeros que constituyen la estructura cuaternaria de una proteína oligomérica
- D) Es el estudio de la magnitud de la fuerza de la interacción enzima-sustrato en una reacción enzimática

50. ¿Cuál de los siguientes nombres corresponde a un motor de búsqueda que utiliza datos de espectrometría de masas para identificar proteínas en bases de datos de secuencias de aminoácidos?

- A) BLAST
- B) MACS2
- C) MySQL
- D) Mascot

51. ¿Cómo se suelen representar los datos de un experimento de metabolómica cuyo objetivo sea comparar los niveles de una serie de metabolitos en varias muestras?

- A) Como una imagen microscópica de alta resolución, por cada muestra, en la que se pueden distinguir, separados, los metabolitos a cuantificar
- B) Como una matriz de dos dimensiones en la que una dimensión corresponde a las muestras y la otra corresponde a los niveles de metabolito
- C) Con una ecuación diferencial de tercer grado para cada metabolito
- D) Con un grafo acíclico dirigido por cada grupo de muestras

52. ¿Cuál es la información principal que contiene la base de datos KEGG PATHWAY?

- A) Rutas metabólicas y los genes que las componen
- B) Compendio de estructuras tridimensionales y sitios activos de proteínas humanas con función estructural
- C) Listado permanentemente actualizado de genomas secuenciados
- D) Niveles de expresión de los genes humanos más abundantes en cada órgano

53. ¿Para qué sirve el comando "paste" del intérprete bash?

- A) Cruza la información de dos archivos y muestra las partes que se repiten
- B) Toma la primera línea de cada archivo y las combina para formar una línea de salida
- C) Muestra las diferencias entre dos archivos
- D) Cambia una cadena de caracteres por otra

54. ¿Para qué se usa el carácter "&" al final de un comando en bash?

- A) Para ejecutar el comando con la máxima prioridad
- B) Para cerrar el terminal de forma automática al acabar la ejecución del comando
- C) Para ejecutar el comando en segundo plano
- D) Para impedir la sobrescritura de cualquier archivo de salida que ya existiese antes de ejecutar el comando

55. ¿Qué significan las letras del protocolo SFTP?

- A) System File Translation Protocol
- B) System Fast Transfer Protocol
- C) Secure File Transfer Protocol
- D) Secure Fast Transfer Protocol

56. ¿Para qué sirve principalmente MySQL?

- A) Para gestionar información en archivos de texto plano con columnas separadas por comas
- B) Para gestionar información en archivos de texto plano con columnas separadas por tabuladores
- C) Para gestionar información en bases de datos relacionales
- D) Para gestionar información en bases de datos no relacionales

57. ¿Para qué sirve el comando pip?

- A) pip es un editor de textos muy utilizado por programadores en Python
- B) pip es un calculador del punto isoeléctrico de un péptido
- C) pip es un buscador de proteínas que contengan uno o varios péptidos de una lista
- D) pip es un sistema de gestión de paquetes utilizado para instalar y administrar paquetes escritos en Python

58. ¿Cómo se delimita el bloque de instrucciones dentro de un bucle "for" en Python?

- A) Con los caracteres "abrir paréntesis" y "cerrar paréntesis"
- B) Con los caracteres "abrir llave" y "cerrar llave"
- C) Sangrando las líneas del bloque con un espacio o un tabulador respecto al comando "for"
- D) Poniendo todas las instrucciones del bloque en la misma línea, separadas por "punto y coma"

59. En el contexto de la secuenciación masiva de fragmentos de ARN ¿qué se entiende por "lecturas de extremos emparejados" (en inglés, "paired-end reads")?

- A) Son secuencias palindrómicas
- B) Son secuencias resultantes de secuenciar los dos extremos de un mismo fragmento de ARN
- C) Son secuencias de ARN con la misma cantidad de pares "GC" que de pares "AU"
- D) Son secuencias de ARN en las que se han sustituido las "U" por "C" para evitar emparejamientos internos tipo "GU"

60. ¿En qué consiste el ensamblaje de genomas?

- A) En reconstruir secuencias de ADN genómico mediante el solapamiento de fragmentos más pequeños de secuencias de ADN
- B) En ordenar los cromosomas de un organismo, en orden alfabético, en un archivo FASTA
- C) En reemplazar los polimorfismos de un solo nucleótido (SNP) por el nucleótido de original del genoma de referencia
- D) En ordenar los cromosomas de un organismo, de menor a mayor, en un archivo FASTA

61. ¿Qué es el Proyecto Microbioma Humano?

- A) Es una extensión del Proyecto Genoma Humano para la caracterización de los microbios presentes en distintas localizaciones del cuerpo humano
- B) Es un conjunto de funciones útiles para analizar datos obtenidos en el Proyecto Genoma Humano, escritas en lenguaje R
- C) Es un proyecto para encontrar características comunes en moléculas con capacidad para inhibir el crecimiento de microorganismos patógenos en el intestino
- D) Es una extensión del Proyecto Genoma Humano para el estudio de los elementos genómicos del ADN mitocondrial

62. ¿Qué información contienen los archivos en formato FASTQ resultantes de una secuenciación por Illumina?

- A) Entre otros detalles, cromosoma, posición inicial y posición final de los distintos elementos genómicos de un genoma
- B) Secuencias de nucleótidos y sus calidades
- C) Coordenadas de alineamientos genómicos de secuencias obtenidas por técnicas de secuenciación masiva
- D) Información sobre la posición, en bytes, del nucleótido inicial de cada cromosoma en un archivo FASTA accesorio

63. ¿Qué se entiende por metatranscriptoma?

- A) Es el conjunto de transcritos comunes de todos los organismos secuenciados del grupo de los metazoos
- B) Es el conjunto de moléculas resultantes de la traducción de un transcriptoma completo
- C) Es el conjunto de transcritos presentes, en un momento dado, en las células de una mezcla de organismos
- D) Es el conjunto mínimo de transcritos, suficiente para que una célula tenga capacidad de reproducirse y, sus descendientes, también la tengan

64. ¿En qué tres aspectos describen los términos de Gene Ontology nuestro conocimiento del dominio biológico?

- A) Eucariota, bacteriano y arqueobacteriano
- B) Animal, vegetal y procariota
- C) Funciones moleculares, procesos biológicos y componentes celulares
- D) Tamaño de los genes, sintenia y abundancia de intrones

65. ¿Qué significan las letras del método GSEA?

- A) Gene Sequence Extended Analysis
- B) General Systematic Expression Analysis
- C) Gene Set Enrichment Analysis
- D) Gene Synteny Extended Analysis

66. En RNA-seq ¿para qué sirve la normalización de datos?

- A) Para corregir los factores que impiden comparar directamente los valores crudos de expresión de varias muestras
- B) Para igualar el número de transcritos a considerar por cada gen cuantificado
- C) Para asignar un valor positivo pequeño a los genes cuyo valor crudo de expresión es cero
- D) Para representar todos los genes con una longitud en una escala de 0 a 100 unidades arbitrarias

67. ¿Qué se entiende por secuenciación de exomas completos?

- A) La determinación de la secuencia del ADN mitocondrial
- B) La determinación de la secuencia de moléculas de ADN presentes en el citoplasma, fuera del núcleo celular
- C) La determinación de la secuencia de los exones de los elementos genómicos
- D) La determinación de la secuencia de las moléculas de ADN presentes en regiones extracelulares

68. ¿Cuál de los siguientes métodos de secuenciación se utiliza, particularmente, para detectar, mediante análisis bioinformático posterior, citosinas metiladas a nivel genómico?

- A) Secuenciación con bisulfito
- B) Secuenciación por el método del dideoxi-terminal
- C) Secuenciación por tecnología capilar
- D) Secuenciación de Sanger

69. En el ensamblaje de genomas ¿qué es el valor N50?

- A) Es la longitud más frecuente, con un margen de 50 bases, de los fragmentos obtenidos en el proceso de secuenciación
- B) Es la media truncada (sin contar los 50 valores más extremos) de las longitudes de los fragmentos obtenidos en el proceso de secuenciación
- C) Es la mediana de las longitudes de los fragmentos obtenidos en el proceso de secuenciación
- D) Es la longitud de fragmento tal que, usando fragmentos de igual o mayor tamaño, se cubre la mitad de las bases del genoma

70. ¿En qué se basan los métodos bioinformáticos para la detección de regiones genómicas diferenciales en experimentos de ChIP-seq?

- A) En los niveles de fluorescencia, convertidos en valores numéricos, producidos al hibridar ARN mensajero con un microarray (o chip) compuesto por sondas complementarias a los genes
- B) En las diferencias de cobertura regional entre una muestra genómica enriquecida y un control no enriquecido
- C) En la diferencia de la abundancia de pares de nucleótidos tipo purina/purina y pirimidina/pirimidina entre diferentes muestras
- D) En la diferencia de señal generada al pasar diferentes nucleótidos de una molécula de ADN a través de un poro asociado a un gradiente eléctrico

71. ¿Qué información contienen los archivos en formato SAM?

- A) Parejas de genes o proteínas que interactúan, física o funcionalmente, y detalles sobre la calidad de la predicción de dicha interacción
- B) Información sobre tamaños de oligopéptidos teóricos generados tras la simulación de la digestión enzimática de una proteína
- C) Valores de similitud entre todas las posibles sustituciones de aminoácidos, calculados en función de sus características físico-químicas
- D) Secuencias de nucleótidos mapeadas, normalmente en cromosomas, con medidas de calidad y otras características

72. ¿Cuál es el objetivo principal del análisis bioinformático de datos de ATAC-Seq?

- A) Identificar características relacionadas con la mayor o menor accesibilidad de la cromatina
- B) Estimar cambios en la transcripción génica, a nivel transcriptómico, entre grupos de muestras
- C) Detectar diferencias en eventos de procesamiento de ARN mensajeros ("splicing") entre grupos de muestras
- D) Determinar diferencias significativas en el contenido en nucleótidos A y T de múltiples regiones genómicas

73. ¿Cuál de estos enunciados describe uno de los pasos del procesamiento bioinformático de datos fisiológicos obtenidos por la técnica de "neurofeedback"?

- A) Extracción de características de interés de señales cerebrales obtenidas por técnicas de electroencefalografía
- B) Determinación de la presencia de regiones genómicas enriquecidas, es decir, que tienen mayor cobertura de lo esperado en una secuenciación de ADN genómico
- C) Determinación de polimorfismos en secuencias genómicas
- D) Detección de eventos de splicing alternativo entre dos grupos de muestras de ARN

74. ¿Qué se entiende por "minería de datos" en bioinformática?

- A) Es el conjunto de procesos para el descubrimiento de modelos y patrones interpretables a partir de grandes cantidades de datos biológicos
- B) La conversión de grandes cantidades de datos biológicos no procesados en resultados patentables
- C) Es el conjunto de tareas encaminadas a corregir errores sistemáticos en grandes cantidades de datos biológicos, generalmente numéricos y presentados en forma tabulada
- D) El conjunto de técnicas para la obtención de datos genómicos de los seres vivos que viven bajo tierra

75. En el contexto de los algoritmos de clasificación, ¿qué es y para qué sirve la curva ROC?

- A) Es la línea suavizada imaginaria (en n dimensiones) que mejor clasifica dos conjuntos de datos pertenecientes a grupos diferentes
- B) Es una representación gráfica del rendimiento del algoritmo de clasificación que sirve para mostrar la distribución de las fracciones de verdaderos positivos y de falsos positivos
- C) Es la forma de la gráfica, en dos dimensiones, que se obtiene al representar valores numéricos en escala lineal (en un eje) y en escala logarítmica (en el otro eje)
- D) Es la representación gráfica que relaciona la variabilidad entre los elementos de un grupo de datos con la variabilidad entre grupos diferentes, en los distintos pasos de un algoritmo de clasificación

76. En el contexto del "cloud computing", ¿qué se entiende por "multicloud"?

- A) El análisis de datos, en un servidor remoto, de los datos de dos o más genomas de diferentes organismos
- B) La división de un análisis de datos complejo en tareas más sencillas, que pueden ejecutarse, en paralelo, en un servidor remoto
- C) Es un tipo de organización de archivos que separa, en redes de discos diferentes, los archivos intermedios generados en cualquier análisis bioinformático sin completar, de los resultados finales
- D) Un enfoque de nube compuesto por al menos dos servicios de nube que proporcionan, por lo menos, dos proveedores de nube pública o privada

77. ¿Cuáles de los siguientes conceptos se refieren a los tipos fundamentales de modelos de servicios de "cloud computing"?

- A) Cloud computing basada en GPUs (GBCC) y cloud computing basada en CPU (CBCC)
- B) Servicio mono-core (MoCS), servicio dual-core (DuCS) y servicio multi-core (MuCS)
- C) La infraestructura como servicio (IaaS), la plataforma como servicio (PaaS) y el software como servicio (SaaS)
- D) Servicio basado en Linux (LbS), servicio basado en Windows (WbS) y servicio basado en OS X (ObS)

78. ¿A qué cuatro magnitudes alude la expresión "las cuatro V del Big Data"?

- A) Viscosidad, volumen, velocidad y volatilidad
- B) Volumen, variedad, visibilidad y velocidad
- C) Volumen, variedad, velocidad y veracidad
- D) Valor, variedad, velocidad y volumen

79. ¿Cuál de las siguientes es una definición adecuada del concepto de armonización de datos?

- A) Consiste en convertir los datos en un conjunto de valores numéricos de valor promedio 0 y desviación estándar 1
- B) Es el proceso de recopilar datos desde diferentes fuentes y formatos para interconectarlos racionalmente, convirtiéndolos en una masa cohesionada
- C) Es un procedimiento que consiste en completar los datos no numéricos, de una base de datos, en cadenas de caracteres de la misma longitud
- D) Consiste en la eliminación de los valores numéricos que sobrepasen el rango Inter-cuartil tanto por el rango superior como por el inferior

80. En el ámbito de la visualización gráfica de datos masivos, el análisis de componentes principales (PCA) es útil para:

- A) Determinar el promedio, la mediana, la moda y los cuatro cuartiles de un conjunto grande de valores numéricos
- B) Dividir consecutivamente un grupo de valores, representados por vectores, en dos grupos cada vez, hasta formar un árbol binario con grupos de un solo elemento en las hojas
- C) Representar dos o tres listados de elementos como círculos solapados en los que, en cada intersección, se muestran los elementos comunes a las listas implicadas en la misma
- D) Reducir la dimensionalidad de los datos para poder representarlos, perdiendo algo de información, en dos o tres dimensiones

81. Una de las herramientas más utilizadas en bioinformática aplicada a la secuenciación es IGV. ¿Cuál es su principal utilidad?

- A) Sirve para calcular parámetros estadísticos de varios conjuntos de datos
- B) Sirve para representar gráficamente las intersecciones entre conjuntos de datos
- C) Sirve para descartar secuencias de baja calidad en archivos FASTQ
- D) Sirve para visualizar alineamientos de secuencias de nucleótidos

82. En la comparación de secuencias de nucleótidos, ¿para qué sirven las matrices de puntos ("dot plots")?

- A) Para calcular el porcentaje de G+C de cada secuencia de nucleótidos
- B) Para determinar la longitud relativa de dos secuencias de nucleótidos
- C) Para calcular el tamaño final de un archivo FASTA que contenga todas las secuencias a comparar
- D) Para determinar, de forma gráfica, qué regiones de dos secuencias tienen cierta homología

83. Una de las familias de métodos bioinformáticos para la predicción de sitios de corte de proteasas incluye los métodos basados en funciones de puntuación (scoring function-based methods). ¿Cuál es la asunción principal en la que se basan dichos métodos?

- A) Secuencias similares comparten funciones biológicas similares
- B) Secuencias pertenecientes a un mismo operón comparten funciones biológicas similares
- C) Secuencias pertenecientes a un mismo cromosoma comparten funciones biológicas similares
- D) Secuencias con un contenido de G+C similar comparten funciones biológicas similares

84. ¿En qué consiste el "algoritmo de la fuerza bruta" para calcular la homología entre dos secuencias biológicas?

- A) En apilar nucleótidos (o aminoácidos) correspondientes a las mismas coordenadas de cada secuencia, sin introducir ningún gap
- B) En generar todos los alineamientos posibles, puntuarlos según un esquema de valoración y elegir el que obtenga mayor puntuación
- C) Se aplica a secuencias de proteínas y consiste en puntuar con "1" cada aminoácido coincidente y con "0" el resto, sin tener en cuenta las similitudes entre grupos de aminoácidos
- D) En alinear dos secuencias, comenzando por sus extremos 5' y 3', y extendiendo el alineamiento hacia las coordenadas interiores

85. Al comparar secuencias de nucleótidos, ¿cuál es la principal diferencia entre alineamientos globales y alineamientos locales?

- A) Se consideran alineamientos globales los realizados entre secuencias de distintos orígenes geográficos. Alineamientos locales son los realizados entre secuencias obtenidas en zonas geográficas cercanas
- B) Los alineamientos entre secuencias con un 70% o más de homología se consideran alineamientos locales. Todos los demás se consideran alineamientos globales
- C) Los alineamientos locales son los realizados entre secuencias del mismo organismo. Los alineamientos globales sirven para comparar secuencias de distintos organismos
- D) En un alineamiento global se intentan alinear las secuencias completas de extremo a extremo y en un alineamiento local se buscan regiones locales con alta homología

86. ¿Para qué sirve la matriz BLOSUM 62?

- A) Sirve para poder comparar nucleótidos del tipo purina (A, G) con los del tipo pirimidina (T, C)
- B) Sirve para determinar un conjunto de genes patrón, cuya transcripción no varía apenas en ninguna condición fisiológica de la célula
- C) Sirve para estimar la homología entre secuencias de aminoácidos
- D) Sirve para calcular la distancia entre grupos de patrones de expresión génica

87.Cuál de las siguientes variantes de BLAST sirve para comparar una secuencia problema de nucleótidos, que será traducida por el algoritmo, frente una base de datos de secuencias de proteínas.

- A) blastx
- B) blastp
- C) blastn
- D) tblastn

88. ¿Cuál es el objetivo principal del proyecto GENCODE?

- A) Mejorar los métodos de detección "Ab initio", en genomas, de secuencias codificantes de genes
- B) Detectar empíricamente nuevas equivalencias codón/aminoácido del código genético
- C) Secuenciar millones de genomas mitocondriales de individuos humanos para detectar patrones de secuencia asociados a enfermedades
- D) Identificar y clasificar todos los elementos génicos en el genoma humano y en el de ratón, con una alta precisión basada en evidencias biológicas

89. ¿Cuál de las siguientes afirmaciones, acerca de los alineamientos múltiples de secuencias (AMS) de nucleótidos o aminoácidos, NO es correcta?

- A) La mayoría de los programas para obtener AMS óptimos entre un gran número de secuencias biológicas usan métodos de optimización global
- B) De un AMS se puede inferir el grado de homología entre las secuencias biológicas que lo forman
- C) En general, los métodos que se requieren para generar AMS de tres o más secuencias biológicas son más complejos que los que sirven para comparar solo dos secuencias biológicas
- D) En un AMS, todas las secuencias resultantes, compuestas por nucleótidos o aminoácidos y "gaps", tienen la misma longitud

90. Cuando se usa un algoritmo progresivo para construir un alineamiento múltiple de secuencias, ¿para qué sirve el "árbol guía"?

- A) Para encontrar las secuencias con mayor contenido en G+C y comenzar a construir el alineamiento con el mayor número de enlaces fuertes
- B) Para determinar el organismo al que pertenece cada secuencia
- C) Para descartar secuencias biológicas demasiado cortas
- D) Para encontrar las parejas de secuencias más similares y comenzar a construir el alineamiento final a partir de ellas

91. ¿Cuál de las siguientes herramientas bioinformáticas NO sirve para representar alineamientos múltiples de secuencias biológicas?

- A) Boxshade
- B) MultiExperiment Viewer
- C) Jalview
- D) MView

92. En el contexto del análisis de secuencias biológicas, ¿qué se entiende por huella dactilar (en inglés "fingerprint")?

- A) El porcentaje de nucleótidos G+C en una región definida de la secuencia
- B) El patrón de estructuras secundarias (hélices alfa y cadenas beta) en una región definida de la secuencia
- C) Una secuencia de ADN, generalmente palindrómica, a la que se une un factor de transcripción
- D) Un grupo de motivos extraído de regiones conservadas de un alineamiento de secuencias

93. En los resultados de BLAST ¿qué es el E-Value?

- A) Es la proporción de nucleótidos G+C que pertenecen a la región de secuencia alineada
- B) Es la fracción de nucleótidos o aminoácidos alineados respecto a la longitud de la secuencia de menor tamaño
- C) La suma de puntuaciones de cada nucleótido o aminoácido alineado menos la suma de penalizaciones por los gaps
- D) Indica en qué medida, la similitud de secuencia encontrada, se podría deber al azar

94. ¿Cuál de las siguientes afirmaciones sobre análisis filogenéticos NO es correcta?

- A) Los métodos de reconstrucción filogenética más habituales asumen que todas las secuencias proceden de un ancestro común
- B) Un análisis filogenético permite estimar relaciones jerárquicas entre secuencias, pero no da información sobre la distancia relativa entre ellas
- C) Es posible hacer análisis filogenéticos con secuencias de nucleótidos o de aminoácidos
- D) A partir del resultado de un alineamiento múltiple de secuencias es posible hacer un análisis filogenético para estimar relaciones evolutivas entre ellas

95. ¿Cómo es la estructura de un archivo PDB típico que describe la estructura de una proteína?

- A) Un archivo de texto que contiene una cabecera con detalles sobre la proteína, seguido de la secuencia y una lista de átomos y sus coordenadas
- B) Un archivo binario y comprimido (solo compatible con visualizadores especiales) que contiene la secuencia y las coordenadas de todos los átomos
- C) Una imagen vectorial, en formato SVG, que representa la estructura terciaria de la proteína
- D) Una serie de ecuaciones paramétricas cuya representación en el espacio de 3 dimensiones (x, y, z) da como resultado una imagen esquemática de la estructura terciaria de la proteína

96. Las proteínas humanas de membrana son las dianas para muchos fármacos actuales. ¿Cuál de los siguientes términos NO es el nombre de una base de datos especializada en estructuras de proteínas de membrana?

- A) TrEMBL
- B) mpstruct
- C) MemProtMD
- D) PDBTM

97. Según la mayoría de métodos que predicen la estructura secundaria de una proteína, ¿a qué tres principales estados puede pertenecer un aminoácido dado?

- A) Pertenece a los aminoácidos polares, a los aminoácidos apolares o no pertenece a ninguno de los dos tipos
- B) Pertenece a los aminoácidos ácidos, a los aminoácidos básicos o no pertenece a ninguno de los dos tipos
- C) Pertenece a una hélice alfa, a una cadena beta o no pertenece a ninguna de las dos conformaciones
- D) Pertenece a las proteínas globulares, a las proteínas fibrosas o no pertenece a ninguno de los dos grupos

98. En el contexto de la predicción de la estructura secundaria del ARN ¿cuál de las siguientes frases es correcta?

- A) El ARN mensajero es más o menos lineal y no estructurado, mientras que el ARN ribosómico y el ARN de transferencia sólo son funcionales cuando forman estructuras secundarias y terciarias particulares
- B) Gracias a la abundancia de estructuras resueltas en las bases de datos, es más sencillo predecir la estructura terciaria del ARN que su estructura secundaria
- C) Los tipos de estructuras secundarias del ARN más abundantes son las hélices alfa y las cadenas beta
- D) Además de las interacciones Watson-Crick entre nucleótidos (A:U y G:C), en el ARN, tiene relevancia la interacción A:G

99. ¿A qué nos referimos cuando decimos que un método de predicción de estructura de una proteína es "Ab initio"?

A) A que se basa en la comparación de la secuencia de aminoácidos de la proteína con una base de datos de secuencias con estructuras conocidas

B) A que se basa en analizar fotografías microscópicas de la proteína a estudiar, cristalizada en distintas conformaciones

C) A que se basa en encontrar en el genoma, los codones STAR y STOP, y los sitios de splicing de cada exón del gen que da lugar a la proteína cuya estructura se quiere predecir

D) A que se basa, exclusivamente, en las propiedades de la secuencia de aminoácidos de la proteína cuya estructura se quiere predecir

100. Las interacciones entre macromoléculas biológicas y otras moléculas de menor peso molecular es muy importante para el desarrollo de fármacos. En este contexto, ¿cuáles de las siguientes son dos enfoques populares para la predicción bioinformática de este tipo de interacciones?

A) Agrupamiento jerárquico y agrupamiento no jerárquico

B) Complementariedad geométrica y simulación del proceso de anclaje

C) Análisis de componentes principales (PCA) y análisis de la curva ROC

D) Ajuste por Bonferroni y ajuste por Benjamini-Hockberg

