

Guia N° 1

Alumnos: Micaela Vargas, Francina Zuliani, Santiago Vietto

Materia: Introducción a la bioinformática

Docente: Danilo Ceschin

Año: 2020

1)_

1.1) El núcleo de la célula eucariota es el elemento más destacado, es el depósito de la información genética de la célula, el mismo está rodeado por una membrana nuclear. La membrana nuclear es una doble membrana que delimita el núcleo de la célula. La función de la misma es mantener separados el contenido del núcleo, llamado nucleoplasma y, el citoplasma. ✓

1.2)

- La **replicación** es el mecanismo que permite al ADN duplicarse (es decir, sintetizar una copia idéntica). Cuando una célula se divide, en primer lugar, debe duplicar su genoma para que cada célula hija contenga un juego completo de cromosomas. ✓
- La **expresión** génica es el proceso que permite obtener proteínas a partir de genes. Los genes son secuencias de nucleótidos de ADN que codifican la información necesaria para la síntesis de proteínas. ✓
- Se le llama **mutación genética**, a los cambios que alteran la secuencia de nucleótidos del ADN. Estas **mutaciones** en la secuencia del ADN pueden llevar a la sustitución de aminoácidos en las proteínas resultantes. ✓
- Es el proceso por medio del cual todos los organismos, tanto procariotas como eucariotas transforman la información codificada por los ácidos nucleicos en las proteínas necesarias para su desarrollo, funcionamiento y reproducción con otros organismos. ✓
- El material genético se emplea para **guardar la información** genética de una forma de vida orgánica y, en eucariotas, está almacenado en el núcleo de la célula (**almacenaje**). ✓

1.3) En el **modelo** de **Watson y Crick**, las dos cadenas de la **doble hélice** del ADN se mantienen unidas por puentes de hidrógeno entre las bases nitrogenadas en cadenas opuestas.

La doble hélice es la descripción de la estructura de una molécula de ADN. Una molécula de ADN consiste en dos cadenas que serpentean una alrededor de la otra como una escalera de caracol. Cada cadena tiene una espina dorsal en la cual se alternan un azúcar (desoxirribosa) y un grupo fosfato. A cada azúcar se une una de las cuatro bases: adenina (A), citosina (C), guanina (G) o timina (T). Las dos cadenas se mantienen unidas por enlaces entre las bases nitrogenadas, adenina formando enlaces con la timina, y citosina con la guanina. ✓

1.4) Se conoce como **gen** a la cadena de ácido desoxirribonucleico (ADN), una estructura que se constituye como una unidad funcional a cargo del traspaso de rasgos hereditarios. Un gen, es una serie de nucleótidos que almacena la información que se requiere para sintetizar a una macromolécula que posee un rol celular específico. ✓

En los eucariotas (organismo un poco más complejos con el ADN dentro de un núcleo) los genes tienen dos tipos de regiones: los exones o regiones codificantes y los intrones o regiones no codificantes. Un **intrón** es una región del ADN que forma parte de la transcripción primaria de ARN, pero a diferencia de los exones, son eliminados del transcrito

maduro, previamente a su traducción. Un **exón** es la porción de gen que codifica aminoácidos.

2)_

2.1) Sabiendo que A: 27%; G: 35%; C: 25% y T: 13%, y la suma total es 100%:

G + C = 56% entonces **G= 28%** y **C=28%**

por lo tanto A + T = 44% entonces **A=22%** y **T = 22%**

2.2) La proposición es falsa, ya que una molécula de ADN no puede estar compuesta únicamente por guanina y por timina, debe existir adenina y citosina también. Además, según la Ley de Chargaff, el número total de bases púricas es igual al número total de bases pirimidinas y en este caso, el número de bases pirimidinas sobrepasa al de bases púricas

2.3) Se ha comprobado que la temperatura aumenta con el contenido de G+C. Como el par de bases G-C está unido por tres puentes de hidrógeno (a diferencia del par A-T que sólo presenta 2) se requiere una temperatura más alta para desnaturalizarlo.

3)_

3.1)

i) Un **codón** es una secuencia de tres nucleótidos de ADN o ARN que corresponde a un aminoácido específico y almacena la información genética para el ensamblaje del mismo. El código genético describe la relación entre la secuencia de bases del ADN (A, C, G y T) en un gen y la secuencia correspondiente de la proteína que codifica. La célula lee la secuencia del gen en grupos de tres bases. Existen 64 codones diferentes: 61 son específicos de aminoácidos, mientras que los tres restantes se utilizan como señales de parada.

ii) Un codón es un triplete de nucleótidos. En el código genético, cada aminoácido está codificado por uno o varios codones. El codón es la unidad de información básica en el proceso de traducción del ARNm. Cada uno de los codones codifica un aminoácido y esta correlación es la base del código genético que permite la traducción de la secuencia de ARNm a la secuencia de aminoácidos que componen la proteína.

El código genético es compartido por todos los organismos conocidos, incluyendo virus y orgánulos, aunque pueden aparecer pequeñas diferencias. Este hecho indica que el código genético ha tenido un origen único en todos los seres vivos conocidos. La palabra "universal" en este contexto aplica solamente a la vida en la Tierra, ya que no se ha establecido la existencia de vida en otro planeta

3.2)

i, ii, iii, iv) El ARN y el ADN son polímeros formados por largas cadenas de nucleótidos. Un nucleótido está formado por una molécula de azúcar (ribosa en el ARN o desoxirribosa en el ADN) unido a un grupo fosfato y una base nitrogenada. Las bases utilizadas en el ADN son la adenina (A), citosina (C), guanina (G) y timina (T). En el ARN, la base uracilo (U) ocupa el lugar de la timina. El ácido ribonucleico (ARN) es una molécula similar a la de ADN. A diferencia del ADN, el ARN es de cadena sencilla. Una hebra de ARN tiene un eje constituido por un azúcar (ribosa) y grupos de fosfato de forma alterna. Unidos a cada azúcar se encuentra una de las cuatro bases adenina (A), uracilo (U), citosina (C) o guanina (G). Hay diferentes tipos de ARN en la célula: ARN mensajero (ARNm), ARN ribosomal (ARNr) y ARN de transferencia (ARNt).

La **síntesis de proteínas** se realiza en los ribosomas situados en el citoplasma celular. En el **proceso de síntesis**, los aminoácidos son transportados por ARN de transferencia correspondiente para cada aminoácido hasta el ARN mensajero donde se unen en la posición adecuada para formar las nuevas proteínas.

Un **ARN no codificante** es una molécula de ARN funcional, que a diferencia del RNA mensajero no se traduce en una proteína.

Los **ARNs no codificantes** pertenecen a diversos grupos y participan en muchos procesos celulares. Estos van desde los ncARNs de vital importancia que se conservan en la mayoría de la vida celular a través de ncARNs más transitorios, que son específicos para una o unas pocas especies relacionadas.

4)_

Sabiendo que purinas(A-G) y pirimidinas(C-T), complemento (A-T y G-C):

Para la hebra de DNA 5'-TACGATCATAT-3' la hebra de DNA complementario correcta es:

3'-ATGCTAGTATA-5'

5)_

a) Para el nucleótido A T C G U T A G C T U A G la secuencia complementaria es:

U A G C A A U C G A A U C

b) Esta secuencia no podría existir en la naturaleza porque contiene Uracilo y Timina, estas dos bases nitrogenadas no pueden coexistir ya que una pertenece al ADN (timina) y la otra al ARN (uracilo), además el uracilo sería el reemplazo de la timina en el ARN. El uracilo se degrada más rápido que la timina (más estable).

6)_

- a. 335
- b. 112**
- c. 168
- d. 111
- e. cualquier valor ≤ 111
- f. ninguna de estas opciones



Justificación: Hay 336 nucleótidos y se codifica cada 3 nucleótidos donde este es un codón, e incluye los codones de iniciación y terminación. Esto sería $336/3$.



7)_

_ Toda síntesis de proteína comienzan con el codón de iniciación que es AUG (metionina).

Tras la síntesis, se podría esperar obtener una proteína, con una secuencia de dos aminoácidos alternantes diferentes

Dado que se ha supuesto que la iniciación puede proceder sin el codón AUG para la metionina, la traducción puede comenzar en cualquier lugar pero debe proceder en unidades de tres nucleótidos a partir del punto de inicio. Así, son codones posibles CAC, ACA, CAC, ACA independientemente del punto de iniciación. Dos codones pueden dar una proteína compuesta de dos aminoácidos alternantes.



8)_

a) Debemos encontrar el marco de lectura del ARNm (ORF - Open reading frame). El AUG (metionina) es el codón de iniciación siempre, entonces identificamos en nuestra cadena donde está este, y la proteína va a ser hasta un codón de finalización que pueden ser UAA, UAG o UGA.

5'-CCUCAU **AUG** CGCCAUUA **UAA** **UGA** CACACA-3'

5.1

Como tengo dos codones de finalización separo en cadenas de tres, y determino la que corresponde:

5'-CCUCAU **AUG** CGC CAU UAU AAG **UGA** CACACA-3'

La proteína va a ser:

AUGCGCCAUUAUAAGUGA

Met(metionina), Arg(Arginina), His(Histidina), Tyr(Tirosina), Lys(Lisina), stop



b) La secuencia de la hebra de ADN que codifica para este ARNm AUGCGCCAUUAUAAGUGA ES:

3'-GGAGTATACGCGGTAATATTCAGTGTGTGT-5'

c) El ADN molde es el complementario de 5'-CCUCAUAUGCGCCAUUAUAAGUGACACACA-3' :

3'-GGAGTATACGCGGTAATATTCAGTGTGTGT-5'



5'-TGTGTGTCACTTATAATGGCGCATATGAGG-3'

9)_

La epigenética es un campo emergente de la ciencia que estudia los cambios hereditarios causados por la activación y desactivación de los genes sin ningún cambio en la secuencia de ADN subyacente del organismo.



10)_

La diferenciación celular es un proceso muy importante pero también complejo; es necesario tener en cuenta que mediante la diferenciación se logra que células con idéntica información contenida en sus genomas presenten fenotipos (morfología, funciones, etcétera) tan diferentes como una célula ósea, una neurona o una fibra muscular. Los procesos de diferenciación son parte esencial de los procesos de desarrollo en seres multicelulares. Por todo ello, la diferenciación es un proceso al cual se está dedicando mucha investigación en la actualidad. Se cree que las células madre pueden un día, cuando se conozcan a fondo los mecanismos de diferenciación, ayudar a la sanación de graves lesiones (por ejemplo, medulares) o a curar importantes enfermedades, tales como las neurodegenerativas.



Guia N° 2

Alumnos: Micaela Vargas, Francina Zuliani, Santiago Vietto

Materia: Introducción a la bioinformática

Docente: Danilo Ceschin

Año: 2020

1)_

Algunos hitos históricos:

- **2003:** se funda en España el Instituto Nacional de Bioinformática, soportado por la Fundación Genoma España (fundada, a su vez, un año antes y que pretende constituirse en instrumento del estado para potenciar la investigación en este campo).
- **2004:** la estadounidense FDA (*Food and Drug Administration*, agencia para la administración de alimentos y fármacos) autoriza el uso de un chip de ADN por primera vez.
- **2005:** se completa el proyecto HapMap (catalogación de variaciones genéticas en el ser humano).
- **2008:** la revista *Science* ya señalaba la reprogramación celular como el hallazgo estrella de la investigación de 2008. Desarrollada en 2006 en ratones, la técnica posibilita, entre otras cosas, que una célula de la piel o de un cabello se convierta en una neurona o en cualquier otro tipo celular de los 220 que componen nuestro organismo. Eso significa que gracias a esto, se puede borrar la memoria del desarrollo de una célula, convirtiéndola en un tipo totalmente diferente después de haberla devuelto a su estado embrionario. El padre de esta técnica, el japonés Shinya Yamanaka, recibió en 2012, junto con el científico británico John B. Gurdon, el Nobel de Medicina por sus investigaciones pioneras en clonación y células madre.
- **2009:** los servicios informáticos básicos son clasificados por el EBI en tres categorías
 - similitudes entre secuencias (SSS)
 - Alineamientos múltiples de secuencias (MSA)
 - análisis de secuencia bioinformático (BSA)
- **2012:** Gracias a las investigaciones del proyecto ENCODE (la investigación de mayor envergadura en el campo de la genómica en la que participan varios biomédicos españoles) se descubrió que el llamado ADN basura es mucho más útil de lo que se pensaba. Y es que, en realidad, este es esencial para que los genes humanos funcionen ya que regula su actividad.

2)_

Definiciones:

- “La **bioinformática** es una disciplina que permite aplicar **herramientas de la informática al estudio y la gestión de datos** de la **biología**. En su campo interactúan diversas áreas del conocimiento como las ciencias de la computación, la estadística y la química.” - Julián Pérez Porto y María Merino
- “La bioinformática es un área emergente interdisciplinaria que se ocupa de la aplicación de la informática a la recopilación, almacenamiento, organización, análisis, manipulación, presentación y distribución de información relativa a los datos biológicos o médicos, tales como macromoléculas (por ejemplo DNA o proteínas).” - Marta Coronado Zamora.
- “Bioinformática es un campo de la ciencia en el cual confluyen varias disciplinas tales como: biología, computación y tecnología de la información. El fin último de este campo es facilitar el descubrimiento de nuevas ideas biológicas así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología” - Centro Nacional para la Información Biotecnológica "National Center for Biotechnology Information"

En nuestra opinión, la definición que nos parece más válida sería la C, porque habla de la interrelación de la computación, biología y la tecnología para formar lo que conocemos como Bioinformática, ahí queda claro que esta ciencia surge de la ayuda mutua entre los campos nombrados, en cambio, otras definiciones expresan que la biología utiliza a las tecnologías como una mera herramienta. Además expresa el objetivo de la bioinformática y su importancia en el campo de la biología.



3)_

a. Organización de la Información: es una estrategia que nos permite mejorar la capacidad de comprensión y elevar la eficacia al estudiar organizando la información, para así, poder entender el tema de forma efectiva.



b. Acceso a la Información: el **acceso a la información** se refiere al conjunto de técnicas para buscar, categorizar, modificar y acceder a la información que se encuentra en un sistema: bases de datos, bibliotecas, archivos, Internet. Se aplica a información que ya ha sido procesada por el entendimiento humano o por algún tipo de sistema de procesamiento automático, por lo que el objetivo no es tanto encontrar la manera más eficiente de clasificarla y archivarla; como sí lo es encontrar la mejor manera de obtener de manera inequívoca la información deseada utilizando para ello el menor número de recursos.



c. Algoritmos: es un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar

10.1
datos y llevar a cabo otras tareas o actividades.² Dados un estado inicial y una entrada, siguiendo los pasos sucesivos se llega a un estado final y se obtiene una solución. ✓

10.2 10.3
d. Búsqueda de secuencias: una **secuencia** es una sucesión de letras representando la estructura primaria de una molécula real o hipotética de ADN o banda, con la capacidad de transportar información.

10.4
Cuando los proyectos de secuenciación genómica descubren una nueva secuencia biológica es muy poco lo que se sabe de ella. Por ese motivo, una de las actividades más frecuentes en Bioinformática consiste en consultar las bases de datos en busca de secuencias idénticas o parecidas a la secuencia recién descubierta.

e. Genómica: es un campo de la ciencia dentro del campo de la biología molecular. Un genoma es un conjunto completo de ADN dentro de una sola célula de un organismo, y como tal, la genómica se enfoca en la estructura, función, evolución y mapeo de los genomas. La genómica tiene como objetivo la caracterización colectiva y la cuantificación de los genes, que dirigen la producción de proteínas con la ayuda de enzimas y moléculas mensajeras. La genómica también implica la secuenciación y el análisis de genomas. ✓

f. Transcriptómica: estudio de todas las moléculas de ARN en una célula. El ARN se copia de piezas de ADN y contiene información para elaborar proteínas y realizar otras funciones importantes en la célula. La transcriptómica se usa para aprender más acerca de la manera en que los genes se transforman en diferentes tipos de células y cómo esto puede ayudar a la presentación de ciertas enfermedades como el cáncer. Anteriormente a la transcriptómica, lo que se hacía era analizar gen a gen si podían participar en un proceso biológico o no, pero muchas veces el análisis gen a gen para saber si estos participan en funciones biológicas puede ser muy tedioso. La transcriptómica permite detectar cambios en niveles de expresión en situaciones fisiológicas concretas, agilizando el proceso. ✓

g. Proteómica: la proteómica representa la línea de conocimiento que quiere establecer las identidades, las cantidades, las estructuras, las funciones bioquímicas y celulares de todas las proteínas de un organismo, órgano y las formas como las anteriores propiedades varían en el espacio y el tiempo y con el estado fisiológico y anímico. ✓

El objetivo fundamental de la proteómica es comprender tanto la composición (incluyendo modificaciones postranscripcionales) como la interacción de proteínas en una muestra de técnicas empleadas para explorar, simultáneamente, la presencia y abundancia de miles de proteínas en una muestra biológica determinada. ✓

Otro objetivo más o menos inmediato de la proteómica es la aplicación de las tecnologías y datos a la investigación en clínica humana.

h. Metabolómica: la metabolómica detecta y analiza los metabolitos presentes en el organismo, lo que permite diagnosticar o determinar el riesgo de padecer una enfermedad a largo plazo, como obesidad o diabetes, y diseñar dietas personalizadas.

Su objeto de estudio es el metabolismo a gran escala, y para ello es necesario detectar y cuantificar el máximo número de metabolitos (pequeñas moléculas como la glucosa, el

colesterol, los aminoácidos o las vitaminas, que se producen al descomponer el organismo sustancias químicas, alimentos, o sus propios tejidos).

i. Análisis de Imágenes: el análisis de imagen es una técnica compleja que exige para su estudio un planteamiento eminentemente práctico y que se beneficia del avance de los medios informáticos para la obtención de datos cuantitativos en imágenes microscópicas y macroscópicas. El procesamiento y análisis de imágenes, puede definirse como una forma de extraer información cuantitativa a partir de imágenes y tiene un amplio rango de aplicaciones en Biología, que pueden ir desde la descripción y representación de variaciones de temperaturas y otras variables importantes del clima, la fisiología vegetal, hasta la utilización de marcadores para estudiar la distribución de células ricas en mitocondrias, entre muchas otras utilidades.

j. Visualización de Imágenes: creación de imágenes, diagramas o animaciones para comunicar un mensaje.

k. "Data Mining": es la extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de datos. Otra manera de definirlo podría ser: la exploración y el análisis -por medios automáticos o semiautomáticos- de grandes cantidades de datos con el fin de descubrir patrones con significado. Entonces, podemos decir que su objetivo es extraer conocimiento a partir de bases de datos.

El data mining nació con la idea de aprovechar dos cosas: la ingente cantidad de datos que se almacenaban en áreas como el comercio, la banca o la sanidad y la potencia de los nuevos ordenadores para realizar operaciones de análisis sobre esos datos.

l. Búsqueda de patrones: es una técnica de análisis que se puede aplicar en detección de objetos, reconocimiento, seguimiento y correspondencias. Acto de comprobación de una determinada secuencia de fichas para la presencia de los componentes de algún patrón.

m. Modelado: representación concreta o abstracta que responde a determinados propósitos. Esta constituidos por elementos y relaciones entre esos elementos, y se sustenta de una teoría científica.

n. Simulación: es una técnica que permite recrear situaciones o establecer la factibilidad de un experimento. A partir de la simulación, se logra visualizar a un sistema físico, haciendo una conexión entre lo abstracto y la realidad. Las simulaciones generan un ambiente de aprendizaje interactivo, lo que permite a los estudiantes explorar la dinámica de un proceso.

o. "System Biology": campo de investigación interdisciplinaria de los procesos biológicos en el que las interacciones de los elementos, internos y externos, que influyen en el desarrollo del proceso se representan con un sistema matemático. Permite comprender integralmente el funcionamiento de los sistemas (procesos) biológicos y profundizar en el entendimiento de cómo sus interacciones internas y con otros sistemas conllevan a la aparición (emergencia) de nuevas propiedades.

4)_

La bioinformática es un área multidisciplinar relativamente nueva, una interfase entre la biología y la informática que permite la aplicación de herramientas tecnológicas para responder a problemas biológicos y plantear soluciones. Para analizar sistemas biológicos complejos se requiere de herramientas informáticas que permitan plantear modelos y predecir comportamientos y funciones biológicas. La bioinformática está relacionada con la necesidad de procesar grandes cantidades de datos en el contexto de los marcos teóricos adecuados, con el objetivo de obtener resultados y predicciones más precisas y robustas en las diferentes áreas de aplicación, donde incorporar herramientas del campo de la informática resulta un proceso necesario.

Esta disciplina propone un acercamiento entre áreas, tanto para informáticos como para los profesionales de las ciencias biológicas, generando un espacio conceptual y un lenguaje en común, en el cual confluye el conocimiento científico y las herramientas tecnológicas. Es el cruce de dos mundos que se unen en la búsqueda de soluciones y en generar conocimiento científico, bien soportado estadísticamente, para poder confiar en las predicciones que uno puede llegar a hacer frente a determinados escenarios biológicos.



5)_

¿La bioinformática solamente conjuga disciplinas como la Biología y la Informática?.
Enuncie todas las disciplinas, explicando brevemente por qué y en que nutren a la bioinformática

La bioinformática tiene campos de estudios interdisciplinarios muy vinculados que requieren el uso o el desarrollo de diferentes técnicas estudiadas universitariamente en la Ingeniería Informática como ciencia aplicada de la disciplina informática. Entre estas pueden destacarse las siguientes: matemática aplicada, estadística, ciencias de la computación, inteligencia artificial, química y bioquímica con las que el Ingeniero Informático soluciona problemas al analizar datos, o simular sistemas o mecanismos, todos ellos de índole biológica, y usualmente a nivel molecular.



Índice de comentarios

- 5.1 No es que se separa la cadena en codones porque hay dos señales de "STOP". Se separa en codones por todas las características del código genético y en este caso, se concluye que realmente no es uno de los codones marcados como de finalización, en realidad no lo es.
- 8.1 Faltaron como 8 años :)
- 9.1 ¿Y en referencia con la bioinformática que significa?
- 10.1 ?
- 10.2 ?
- 10.3 No olvidar que puede ser ARN o secuencia de aminoácidos también
- 10.4 está bien pero no sólo significa eso