

Universidad del Sureste.

Campus Tuxtla Gutiérrez.

Iris Rubí Vázquez Ramírez.

Lic. En medicina humana.

Cuarto semestre.

Actividad 3: resumen.

Materia: biología molecular.

Dr. José Miguel Culebro Ricaldi.

Viernes 20 de mayo del 2022.

Lost in translation: análisis bioinformático de las variaciones que afectan al codón de iniciación de la traducción en el genoma humano.

El inicio de la traducción es un paso crucial que está sujeto a una regulación sustancial para garantizar la correcta formación de una proteína por parte de un ARNm (Lorsch y Dever, 2010). Se construye el complejo de preiniciación formado por las subunidades pequeñas del ribosoma, por el ARNt que transporta la metionina y por varios factores de iniciación. Luego este complejo de preiniciación se lleva a un ARNm en su capuchón terminal 5' para iniciar un proceso de exploración de 5' a 3' para encontrar un sitio de inicio de la traducción (TIS), cuyo componente principal es el codón AUG de iniciación. La secuencia que flaquea el codón de iniciación es importante para que el complejo de preiniciación reconozca AUG como un codón de iniciación. La secuencia consenso GCC(A/G)CCAUGG, también conocida como la secuencia de Kozak (Kozak, 1986), cuyos nucleótidos más importantes están en negrita, describe un contexto óptimo para el reconocimiento del AUG subrayado como codón de iniciación. Así, el ribosoma subunidad grande se agrega al complejo, que procede a la próxima etapa de traducción: elongación (Preiss y Hentze, 2003).

Por tanto, la mayoría de las variaciones genéticas que afectan al codón de iniciación debería tener consecuencia patológica debido a la importancia biológica del codón.

Hasta donde sabemos, no existe un estudio global sobre las mutaciones que afectan el codón de iniciación en el genoma humano. Por lo tanto, nuestros principales objetivos son (i) realizar un análisis descriptivo extrayendo todos los datos sobre estas variaciones genéticas de la base de datos Ensembl; (ii) realizar un análisis comparativo entre variaciones frecuentes y raras, considerando las características de los marcos abiertos de lectura (ORF) provocados por el uso de codones de iniciación alternativos, para identificar factores diferenciadores; y (iii) medir el impacto de estas variaciones en un individuo concreto a través del análisis de exomas completos de 5 sujetos.

MATERIAL Y METODO.

Se recuperó un conjunto de objetos de alelo de variaciones transcripción (TVA) de Ensembl. Estos objetos son tuplas de la forma <transcripción, variación, alelo>, de modo que cada TVA identifica un alelo de una variación que afecta a un transcrito concreto. Cada fila del archivo TSV almacena información en un TVA. La misma variación puede aparecer más de una vez porque puede afectar varios transcritos y puede tener varios alelos.

Se buscaron codones de iniciación alternativos para cada TVA en las secuencias mutadas de acuerdo con tres enfoques:

- i. El primer codón UAG encontrado en la región codificante.
- ii. El primer codón UAG en la región de codificación cuyo contexto tiene una eficiencia mayor o igual a 87 según estudios. Este estudio proporciona una lista con todos los TIS posibles que abarcan la posición -6 a 5 junto con una medida de eficiencia. Esta eficiencia se calculó teniendo en cuenta las interacciones de los dinucleótidos.
- iii. El primer UAG encontrado en un fuerte contexto Kozak en la región de codificación. Para identificar el contexto de Kozak en las transcripciones, se compiló una matriz de peso de posición por medio de 10 nucleótidos por lado alrededor del codón de iniciación de 14, 160 muestras de transcripción recuperadas de Ensembl. Después de eso, la matriz formada se utilizó para escanear los transcritos mutados por medio de la biblioteca R PWMEnrich.

Las consecuencias de la mutación también se calcularon comparando la longitud, el marco de lectura y la conservación del péptido señal del ORF de tipo salvaje con los causados por el uso del codón de iniciación alternativo encontrado por los enfoques 1, 2 y 3. El software Phobius (Kallet al.,2004) se empleó para predecir las posiciones relativas al péptido señal en el transcrito de tipo salvaje. Esta estrategia nos permitió calcular la conservación del péptido señal según el codón de iniciación alternativo utilizado. Además, se analizaron las metioninas en la región no traducida (UTR) 5' de la transcripción de tipo salvaje. Se obtuvieron sus posiciones y marcos de lectura con respecto al codón de iniciación de tipo salvaje. También identificamos la existencia de un codón de terminación antes de la región codificante.

ANALISIS DE EXOMA

Se analizaron exomas completos de cinco pacientes para estimar la presencia de variaciones que afectan los codones de iniciación en un individuo concreto. Se empleó Ion Proton Platform (Ion Torrent) para la secuenciación. El software de esta plataforma genera archivos CSV con información sobre las variaciones genómicas encontradas en el paciente. Esta información incluye la posición de una variación en una región codificante, lo que permite a los investigadores aplicar un filtro para obtener las variaciones que afectan a los codones de iniciación. Además, se eliminaron las variaciones con cobertura < 20 para evitar posibles errores de lectura.

RESULTADO.

PAX5 fue el gen más afectado, con 72 TVA (nueve variaciones diferentes), todas presentes en COSMIC, afectando a ocho transcritos del gen. Este gen es un factor de transcripción cuya función es esencial para el compromiso de los progenitores linfoides con el linaje de linfocitos B. (Cobaleda et al., 2007). El segundo gen más afectado es dtna, con 38 TVA (ocho variaciones diferentes), una presente en dbSNP y siete en COSMIC, que afectan a ocho transcripciones diferentes. Este gen codifica una proteína de la familia de las distrobrevinas que desempeña funciones estructurales y de señalización en la membrana plasmática de muchos tipos de células. (Böhmet et al., 2008).

Nuestro estudio identificó 87 TVA (62 variaciones que afectan a 75 transcripciones) con $MAF > 0.01$, siendo rs11107 (dbSNP) la variación más frecuente identificada ($MAF = 0.48742$). Este SNP (ATG/ATA) afecta a una transcripción del FBXO7 gen, que codifica una proteína F-box cuya función está implicada en los procesos de ubiquitinación (Cenciarelli et al., 1999). La segunda variación más frecuente ($MAF = 0.463841$) es rs3764880 (dbSNP). Este es otro SNP (ATG/GTG) que afecta solo a una transcripción de TLR8, un gen que codifica el receptor tipo Toll 8 involucrado en el reconocimiento de patógenos extraños.

1446 TVA (1008 variaciones diferentes que afectan a 1190 transcripciones) tenían un MAF bajo (< 0.01). Una de las variaciones más raras ($MAF = 0.0002$) es rs139267617 (dbSNP), una variante de un solo nucleótido [SNV] (ATG/GTG) que afecta a una transcripción de ITLN1, un gen que codifica la intelectina-1, que participa en el reconocimiento de los glucanos microbianos (Tsujiet al., 2001; Wesener et al., 2015). Otra mutación rara es rs587701888

(dbSNP) (MAF¹/40.0002), un SNV (ATG/GTG) que afecta a tres transcripciones diferentes de SEMA6C, un gen que codifica una proteína de la familia de las semaforinas, que está involucrada en la regeneración neural.

A continuación, probamos mediante Phobius si el uso del codón AUG alternativo podría afectar al péptido señal. Se detectó un péptido señal en la transcripción de tipo salvaje de solo 288 TVA del grupo de variaciones con MAF bajo y 13 TVA en el grupo de MAF alto. El uso del codón AUG alternativo encontrado por el Método 1 en esos TVA dio como resultado una conservación media del péptido señal para variaciones con MAF bajo de 32,2 y 51,0 % para variaciones con MAF alto (la mediana de conservación es 0 % para MAF bajo y 45,7 %). % para MAF alto). Los valores fueron similares cuando consideramos los AUG alternativos encontrados por los Métodos 2 y 3, aunque solo los encontrados por el Método 2 alcanzaron significación estadística.

Los tres enfoques arrojan un resultado similar sobre el estado del marco de lectura: la proporción de codones de iniciación alternativos que mantienen el marco de lectura es mayor para las variaciones de MAF alto (60,92 frente a 48,48 % para el Método 1, 62,07 frente a 41,01 % para el Método 2 y 67,82 frente a 50,97 % para el Método 3). Buscamos codones AUG corriente arriba en la UTR 5' de la transcripción de tipo salvaje. Es de destacar que las variaciones con MAF alto tenían más de 5; codones UTR AUG que variaciones con bajo MAF.