

LOST IN TRASLATION: BIOINFORMATIC ANALYSIS OF VARIATIONS AFFECTING THE TRANSLATION INITIATION CODON IN THE HUMAN GENOME

La traducción es un proceso biológico clave controlado en eucariotas por el codón AUG de iniciación. Las variaciones que afectan a este codón pueden tener consecuencias patológicas al perturbar el correcto inicio de la traducción. Desafortunadamente, no existe un estudio sistemático que describa estas variaciones en el genoma humano. Además, nuestro objetivo era desarrollar nuevas herramientas para en silicopredicción de la patogenicidad de las variaciones genéticas que afectan a los codones AUG, ya que hasta la fecha estos defectos genéticos se han clasificado erróneamente como sin sentido.

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 ARN. Este es un proceso complejo que involucra diferentes elementos y pasos. Brevemente, se construye el complejo de preiniciación formado por la subunidad pequeña 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 flanquea el codón de iniciación es importante para que el complejo de preiniciación reconozca AUG como un codón de iniciación. Por tanto, la mayoría de las variaciones genéticas que afectan al codón de iniciación deberían tener consecuencias patológicas debido a la importancia biológica de este codón. De hecho, existen numerosos ejemplos de mutaciones de este codón asociadas a diversas enfermedades.

En resumen, se recuperó un conjunto de objetos de alelo de variación de transcripción (TVA) de Ensembl. Estos objetos son tuplas de la forma <transcrito, 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 a 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 AUG encontrado en la región codificante.

- ii. El primer codón AUG en la región de codificación cuyo contexto tiene una eficiencia mayor o igual a 87 según un estudio. Ese estudio proporciona una lista con todos los TIS posibles que abarcan las posiciones -6 a +5 junto con una medida de eficiencia. Esta eficiencia se calculó teniendo en cuenta las interacciones de los dinucleótidos.

- iii. El primer AUG 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.

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.

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.

El archivo TSV con todas las variaciones que afectan a los codones de iniciación contiene 15 312 TVA, incluidas 11 261 variaciones diferentes, que afectan a 9591 transcripciones de 7205 genes. Cada gen tenía una media de 2,12 TVA relacionados que afectaban a los codones de iniciación. Los codones AUG alternativos encontrados por los Métodos 2 y 3 también estaban más cerca del codón de iniciación de tipo salvaje para las variaciones con MAF alto que para aquellas con MAF bajo.

La búsqueda de variaciones que afecten a los codones de iniciación en exomas completos de 5 personas reveló la media de 12 variaciones con estas características de la media de 45 696 variaciones identificadas en cada paciente. Se identificaron veintiocho variaciones diferentes que afectan a los codones de iniciación en estos 5 sujetos, y 14 de las variaciones se identificaron en más de 1 paciente. En total, 7 de estas variaciones no tenían MAF informado, mientras que las 21 mutaciones restantes tenían un MAF medio de 0,232.

La traducción es un proceso biológico clave que probablemente no ha recibido la debida atención, particularmente en el contexto de las patologías. Por lo tanto, es bastante normal encontrar informes de mutaciones identificadas en pacientes con diferentes enfermedades que afectan el codón de iniciación de muchos genes diferentes que aún mantienen el cambio de sentido erróneo

previsto. Este fenómeno se debe al simple análisis de la secuencia primaria, cuando las consecuencias de este tipo de variaciones se consideran como un simple cambio de nucleótidos en lugar de considerar posibles truncamientos o interrupciones del marco de lectura traduccional.

Una comparación de predicciones basadas en un residuo de aminoácido y nuestro método puede producir dos resultados: acuerdo y desacuerdo. Acuerdo significa que se obtuvo la misma predicción aplicando métodos completamente diferentes. La distorsión del contexto de iniciación, particularmente si afecta el codón de iniciación AUG, puede dar como resultado un inicio diferente de la traducción que podría generar una proteína bastante diferente. En este estudio, exploramos todas las variaciones reportadas en Ensembl que afectan el codón de iniciación AUG. Encontramos factores diferenciadores que pueden facilitar el desarrollo de nuevas herramientas predictivas para calcular el efecto deletéreo de este tipo de variación. Estos factores se pueden resumir en dos situaciones bien definidas.

En conclusión, los datos mostrados en este artículo sugieren que estamos perdidos en la traducción. Se necesitan nuevas herramientas que tengan en cuenta las características que se muestran en este estudio para predecir el efecto de las variaciones genéticas que afectan a los codones de iniciación AUG. Este efecto puede ser absolutamente inesperado y no estar relacionado con las funciones asignadas originalmente a la proteína de tipo salvaje.

DIEGO FABRICIO GONZÁLEZ MELLANES

BIBLIOGRAFÍA

- Adzhubei, IA et al. (2010) Un método y servidor para predecir errores dañinos mutaciones de sentido. *Nat. Métodos*, 7, 248.
- Asano, K. (2014) ¿Por qué la selección del codón de inicio es tan precisa en eucariotas? *Traducción*, 2, e28387.
- Böhm, S. et al. (2008) Expresión génica de la familia de distrobrevina y distrofina en pez cebra *Expr. gen. Patrones*, 8, 71–78.
- Cazola, M. y Skoda, RC (2000) Fisiopatología traslacional: una novela mecanismo molecular de la enfermedad humana. *Sangre*, 95, 3280–3288 .
- Cenciarelli, C. et al. (1999) Identificación de una familia de proteínas F-box humanas. *actual Biol.*, 9, 1177–1179.
- Cervantes, J. et al. (2012) TLR8: el pariente olvidado reivindicado. *Celula mol. inmunol.*, 9, 434–438.
- Chorev, D. et al. (2015). Exponer la diversidad de subunidades y la modularidad de complejos de proteínas mediante enfoques de espectrometría de masas estructural.
- Cobaleda, C. et al. (2007) Pax5: el guardián de la identidad y función de las células B.