

¿Qué podemos aprender del genoma del cáncer? | Claudia Arnedo

Claudia Arnedo, estudiante de doctorado en Barcelona Biomedical Genomics Lab, llevó a cabo una conferencia donde nos explica respuestas que pueden hacerse la pregunta qué podemos aprender de la genética del cáncer, la mayor parte de la investigación que realiza es bioinformática, también habla de aplicaciones bioinformáticas que sirven para analizar el genoma del cáncer.

El cáncer es un grupo heterogéneo de enfermedades que se caracteriza por el crecimiento descontrolado de células que son anormales, además según el tejido celular afectado pueden invadir el tejido en el que están y a otras localizaciones del organismo desencadenando la metástasis.

La causa, en el más del 90% de los tumores, es genética, hay alteraciones en el ADN de las células que hace que esta se desregule y empiece a proliferar sin control. Las alteraciones genéticas en este caso se pueden dividir en dos grandes grupos: germinales y somáticas. A lo largo de la vida vamos acumulando mutaciones en cada célula, puede ser por procesos endógenos o por daños exógenos, de esta manera, la mayoría de los cánceres son de tipo somático.

Las células tumorales presentan características particulares, como, por ejemplo, mantienen el ciclo celular siempre activo, inactivan genes que evaden la proliferación, invaden el sistema inmune, destruyen células del SI. Estas características interesan porque si sabemos de qué depende el cáncer, se pueden desarrollar terapias específicas.

El cáncer se entiende como una evolución de células que van adquiriendo mutaciones *driver*, este proceso involucra a Darwin por sus estudios sobre el proceso evolutivo. Podemos aplicar este sistema al cáncer, en vez de población de especies, será población celular. La variabilidad genética del cáncer viene dada por cómo se van acumulando las mutaciones somáticas. En un linaje celular donde se van acumulando mutaciones, donde la siguiente tiene más mutaciones que la anterior, por lo que se van sumando mutaciones drivers que acaban por desarrollar el cáncer.

Un gen *driver* tiene mutaciones *drivers* en algunos residuos específicos que confieren una ventaja selectiva, nos interesa conocer estos drivers para conocer como funciona el cáncer, para mejorar la vida de los pacientes oncológicos y adelantarnos a dar un mejor pronóstico.

Para conocer los drivers tenemos que secuenciar los tumores, después, se compara con un genoma de referencia (consenso). Además, cada individuo tiene sus propias variantes por lo que se debe de secuenciar un tejido sano, de esta manera, se conocen qué mutaciones son germinales y cuáles somáticas.

La mayoría de los tumores tienen más de 100 mutaciones somáticas, hay mucha variabilidad, depende de varios factores. Los tumores que acumulan un número mayor de mutaciones son el melanoma, pulmón, esófago y colon. Se puede observar que son tejidos expuestos a un mayor número de mutágenos. Otro tipo de tumores relacionados con un alto número de mutaciones son los pediátricos, o tejidos que no tienen una tasa de proliferación celular elevada, como el sistema nervioso.

La mayoría de los tumores acumula entre 1 y 5 mutaciones *drivers*, el número de mutaciones *drivers*, son muy similares entre diferentes tipos de cáncer. El problema que se plantea está en

qué se analizan un elevado número de mutaciones, pero sólo una pequeña fracción son mutaciones *drivers*. ¿Cómo se identifican mutaciones o gen *drivers*?

Hay varios métodos que identifiquen gen *drivers*, entre ellos se encuentra detectar genes *drivers* que muestren señales de selección positiva. En mucho de los proyectos de investigación, el protocolo que se lleva a cabo es secuenciar gran cantidad de biopsias tumorales de pacientes enfermos, a continuación, se seleccionan las mutaciones somáticas de cada uno de ellos, y, se buscan unos patrones especiales en la distribución de estas mutaciones. Son patrones que aparecen en función del proceso evolutivo de la tumoración, entre estas señales, podemos destacar la **recurrencia**, gran cantidad de pacientes tienen la misma mutación en un gen, además no se espera que dicho gen este tan mutado por azar, esto quiere decir que estás células pueden haber adquirido una ventaja selectiva para proliferar y, por tanto, haber desarrollado el tumor. Otras de las señales, **impacto funcional**, mutaciones a lo largo de la proteína, y, **clustering**, mutaciones distribuidas en una región específica que no esperaríamos encontrar. Mutaciones que tienen un impacto funcional grande, como truncar la proteína, confieren una fuerte señal positiva, de esta manera se puede conocer con mayor facilidad cuál es el gen *driver*. Estas señales descritas, son complementarias, es decir, no todos los genes *drivers* tienen que desarrollar las tres señales. Por ejemplo, los oncogenes acostumbran a tener mutaciones de tipo *clustering*, mientras que, los genes supresores de tumores suelen presentar mutaciones de impacto funcional.

La identificación de señales positivas se lleva a cabo comparando la distribución de mutaciones observadas con las que cabe esperar, siempre y cuando, estas mutaciones no tengan relación con la proliferación celular. Además, las mutaciones no se acumulan de forma uniforme a lo largo del genoma, son bastante variables, depende de covariables a nivel de genoma completo y a nivel de la secuencia específica de nucleótidos, como, por ejemplo, en la cromatina más compacta se acumulan un mayor número de mutaciones debido a que no puede entrar la maquinaria de reparación de ADN, además, la compactación de la cromatina depende del tipo celular, por lo que la distribución es específica de cada tipo de tumor. Una covariable interesante, es la distribución de mutaciones y cómo cambian con la secuencia del genoma, según el tipo de daño o de exposición de mutágenos en cada tejido, las mutaciones son diferentes, por ejemplo, la luz ultravioleta refleja un tipo de mutación en melanomas.

A continuación, Claudia Arnedo, nos explica el método OncodriveCLUSTL, su utilidad principal, es encontrar un gen *driver*, se distribuye la secuencia de tal manera que podemos saber cómo de probable puede ser un *cluster*, lo comparemos con las mutaciones que observamos en los pacientes, si encuentras un *cluster* demasiado grande a lo que cabría esperar, es posible que estemos delante de un gen con señales de selección positiva por lo que, podría ser un gen *driver*. Este análisis cuenta con complementos 3D de señal de *cluster*, donde se puede comparar los *cluster* lineales con los que están en la proteína en disposición tridimensional, esto tiene la utilidad de comparar los métodos de detección de genes *drivers* bioinformáticos, se parte de genes *drivers* conocidos, por lo que se puede inferir en la evaluación del método.

Por último, Claudia nos explica uno de los métodos que integran diferentes submétodos y algoritmos que permiten identificar todos los genes *drivers* obtenidos de manera individual. InTOGen, es un método para identificar todos los *drivers*. Se analizan los pacientes *cohort* a *cohort*, es decir, se analizan aquellos que están secuenciados en un lugar en concreto. Aquí se miran las señales de selección positiva, además de que InTOGen incorpora más de cien métodos de análisis, los principales son los tres descritos anteriormente de manera que algunos métodos

dependen de estos primordiales. Por último, la parte clave es como se combinan los datos obtenidos de los diferentes métodos.

IntOGen, es una tecnología prometedora, ya que aún faltan encontrar genes drivers de menor frecuencia o genes drivers que no codifiquen a proteína, Claudia, hace referencia de qué esto último, es un campo dónde requiere investigación.