

NOTA PERSONAL

Siendo sincero, el estudio del cáncer no es una de mis pasiones, pero disfruté muchísimo esta presentación (además la dinámica de las tarjetas para interactuar con nosotros estuvo muy guay). Me pareció súper interesante, sobre todo, por dos razones principales: que ella se formase en bioinformática relativamente “tarde” en su recorrido educativo, y que este proyecto surgiese de la colaboración entre personas de distintas disciplinas.

Quizá sea por mi inexperiencia, pero me parece que su iniciativa es muy ambiciosa a nivel de programación, y es súper motivador ver cómo alguien que apenas había comenzado desde cero al empezar su proyecto haya sido completamente capaz de ir más allá a un nivel complejo. Puede que para ella lo que está haciendo es sencillo, pero desde mi perspectiva es una hazaña, y creo que para todos nosotros es un chute de motivación y un toque de atención a intentar no infravalorar lo que aprenderemos en un futuro.

Respecto al *background* tan diverso de sus compañeros, personalmente me encanta que nos haya mostrado cómo la transversalidad en ciencia sí funciona. En las últimas décadas se ha tendido a una superespecialización en (que bueno, también considero que tiene su razón de ser y por qué puede ser necesaria), pero eso no debe empañar la importancia de unir los conocimientos de disciplinas a priori poco solapadas para resolver problemas tan relevantes como éste. De hecho, para áreas tan puntillosas como la medicina personalizada que promete este proyecto, seguramente deba ser fundamental unir fuerzas. Además, también es un ejemplo perfecto de lo versátil que es la bioinformática, y como puede ser un recurso perfecto en manos tanto de un biólogo como de una ingeniera.



¿Cómo podemos analizar el genoma del cáncer?

Conferencia por Claudia Arnedo

El cáncer, a pesar de su concepción coloquial, se refiere a un grupo de enfermedades caracterizadas por un crecimiento anormal e invasión de tejidos por partes de células del cuerpo. En más del 90% de los tumores, la causa es genética ya sea por alteraciones germinales o somáticas; en la mayoría de casos, son mutaciones acumuladas a lo largo de la vida. Estos tumores comparten una serie de propiedades o *hallmarks*: mantener las señales de proliferación activas, suprimir genes que inhiben el ciclo celular, evadir el sistema inmune, ... Por lo que es importante conocerlas para el desarrollo de terapias con el objetivo de contrarrestarlas.

Existe un componente evolutivo de selección en el desarrollo del cáncer, puesto que en los tejidos existe variabilidad entre las células debido a la acumulación aleatoria de mutaciones. Si alguna de ellas recoge mutaciones *drivers* que incrementen la proliferación, se dividirá más que el resto y acumulará a su vez nuevas mutaciones que aporten nuevas *hallmarks*. La identificación de genes y mutaciones *driver* es un objetivo de la investigación oncológica ya que permite prevenir, diagnosticar, tratar y pronosticar la enfermedad.

Para descubrir qué mutaciones responsables son somáticas, se secuencian el tumor y se compara, por un lado, con el genoma de referencia, y por otro, con la secuencia de otro tejido que esté sano (sangre, por ejemplo). La mayoría de tumores tienen más de 100 mutaciones, pero algunos pueden acumular hasta 10k-100k, sobre todo en los tejidos más proliferativos o expuestos a mutágenos, como la luz UV o el tabaco. De hecho, la tasa de mutación entre tipos de cáncer es muy variable, pero no así el número de *drivers*. La gran parte únicamente tienen 4 o 6, lo que supone un problema a la hora de identificarlas entre todas las demás. Por otro lado, también depende del tipo celular en cuanto a que la conformación de la cromatina cambia entre ellos.

La mayoría de mutaciones son *passengers*, no afectan a la evolución del tumor.

Los genes *driver*, susceptibles a sufrir alguna mutación de este tipo, pueden detectarse a partir de señales de selección positiva. Tras secuenciar muchos pacientes, se determina qué patrones asociados a tumorigénesis son compartidos entre ellos. Estos patrones de selección positiva son: recurrencia, impacto funcional grande (truncar la proteína, por ejemplo), y clustering (oncogenes); son complementarios, y no siempre aparecen los 3 a la vez. Así, es posible calcular la probabilidad de una mutación usando la frecuencia de la misma en un grupo de pacientes con cáncer.

Un método de clustering basado en secuencias es OncodriveCLUSTC, que acepta como input mutaciones somáticas y regiones genómicas, analizando las mutaciones observadas y simulando las esperadas, repitiendo el proceso miles de veces hasta obtener un clustering observado y otro simulado. La comparación de ambos devuelve un p-value que determina cómo de diferentes son. Si se combina con algoritmos de identificación de genes oncogénicos, se obtendrán mejores resultados.

El método que agrupa todos los algoritmos complementarios para aumentar la fiabilidad de los resultados sería **IntoGEN**. Para ello no se secuenciarían pacientes, sino que se utilizarían datos de bases de datos (~28.000 pacientes), tres métodos de clustering con características ligeramente diferentes (localización de clusters, etc.), métodos de impacto funcional, y otros de recurrencia (ver si el contexto de la mutación es el esperado o no). En total, se combinan resultados de siete métodos diferentes. Por cada cohorte, se da un peso diferente a cada método según su fiabilidad en ese contexto concreto para encontrar así genes *driver* (de momento se han identificado 500-600, ya que sólo se han comprobado los genes codificantes).

Un gen *driver* puede serlo únicamente para tipos de cáncer en concreto, algunos aparecen en muchos, mientras que otros son más específicos.

Por ejemplo, boostDM es un algoritmo de machine learning que predice mutaciones *driver*, creando un modelo que compare las mutaciones con las ya conocidas como *driver* o

passenger, y explicando incluso por qué lo son (si ocurren en un dominio kinasa, por ejemplo). El cancer genome interpreter funciona como un algoritmo que puede identificar las mutaciones potencialmente oncogénicas tras introducir como input las de un paciente, así como una prescripción *in silico* con posibles medicamentos que pudiesen tratar al paciente de manera personalizada.

Como perspectivas de futuro en la genómica del cáncer, se han de mirar nuevos y más tipos de cáncer, intentando disminuir el umbral de frecuencia bajo el cual los tumores más raros son indetectables. Del mismo modo, se debe recoger más información respecto a las mutaciones metastásicas.