

Claudia Arnedo (IRB): HOW CAN WE ANALYZE CANCER GENOME?

Introducción

Explicación sobre la *fitness* i el proceso de selección que se da también en las células cancerígenas. Los drivers del cáncer son alteraciones génicas o mutaciones puntuales que confieren cualidades favorecedoras a la aparición de cáncer. La acumulación de drivers deriva al tumor y la metástasis.

Identificar los drivers es esencial para caracterizar los mecanismos moleculares del cáncer y desarrollar la medicina de precisión para estos casos. Para lograrlo se secuencian el genoma en muestras tumorales, el genoma de muestras de sangre del paciente y se comparan con el genoma de referencia (discernir mutaciones propias del genoma del paciente de las somáticas que se han dado en el cáncer).

IntOGen

Pipeline para el estudio. Secuenciamos una cohorte de pacientes y seleccionamos las mutaciones somáticas. Después buscamos patrones de selección positiva en el genoma (recurrencia → más mutaciones de lo esperado por azar en un gen, impacto funcional → efectos sobre el producto de un gen, *clustering* → acumulación de mutaciones en una región concreta de la secuencia, hecho asociado a los dos casos anteriores ya que se localizan especialmente en regiones de impacto funcional) para identificar los *drivers*. En el caso genes supresores hay menos *clustering* pero si acumulan mutaciones de impacto funcional, mientras que en los oncogenes sí se detectará bastante *clustering*, por tanto, la pipeline debe incluir diversos métodos de detección).

También debemos tener en cuenta el contexto genómico para conocer la tasa de mutación por nucleótido (el contexto varía en función de la secuencia *upstream/downstream*). Muchas veces sin embargo no tenemos todos los datos para todos los individuos.

La pipeline combina 6 métodos de detección: dos de recurrencia, tres de clustering y uno de impacto funcional. Cada uno aporta unos *drivers* con un p-value y el resultado combinatorio final se hace dando un peso distinto a cada método en base a su capacidad de detección (analizado con un control +).

A través de su uso se han podido detectar drivers nuevos e incluso drivers ya asociados a un cáncer que también estén asociados a otros tipos.

OncodriveCLUSTL

Uno de los algoritmos de detección. El input coge las secuencias nucleotídicas de las mutaciones somáticas, detecta los clusters y les asigna un valor (basado en la distancia entre mutaciones dentro del *cluster* y el contexto). Para cada mutación abrimos una ventana para analizar en 1000 pruebas como esperaríamos la distribución de mutaciones alrededor (contexto) y, al comparar esta prueba con el resultado observado conseguimos evaluar si hay diferencias.

Se evaluó si el método es más útil que el que se utilizaba previamente (OncodriveCLUST vs OncodriveCLUSTL). Se evalúa el porcentaje de drivers que son capaces de detectar de forma progresiva (Top1 si ambos lo detectan 100% en ambos, Top2 para dos loci si solo uno detecta tenemos 100% para uno y 66,7% para el otro). Así pues a mayor área bajo la curva mejor detección.

También comparamos con otro algoritmo (HotMAPS) que se basa en detectar clusters en la estructura 3D de los productos (algo que no puede ver el algoritmo explicado). Sin embargo, HotMAPS no puede detectar mutaciones *truncating* o de las que no se tengan datos 3D.

Cancer Genome Interpreter

Buscar los *drivers* en un paciente y analizar que mutaciones son susceptibles de recibir terapia génica (eventos accionables → análisis *in silico*).