

[Centro de Información de COVID \(CIC\): Charlas científicas de relámpago](#)



Transcripción de una presentación de Ali Rahnavard (George Washington University), octubre de 2022

Título: [Una nueva plataforma para la integración de datos y el aprendizaje profundo en COVID-19](#)

[Perfil de Ali Rahnavard en la base de datos del CIC](#)

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés): [2028280](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de Octubre 2020](#)

Editora de la Transcripción: Saanya Subasinghe

Traductor: Isabella Graham Martínez

Transcripción

Diapositiva 1

Muchas gracias por la oportunidad. Este es un trabajo de colaboración en el Instituto de Biología Computacional de la Universidad George Washington.

Diapositiva 2

En resumen, el enfoque de lo que hacemos en mi laboratorio es - con las tecnologías de alto rendimiento, podemos medir muchos ómicos diferentes. Y nuestro objetivo es ver si podemos obtener una mejor imagen utilizando la integración de estos datos ómicos diferentes, incluyendo genómica, metabolómica, proteómica, y genomas virales también.

Diapositiva 3

Así que, de hecho, el objetivo o propuesta de Subvención de la NSF era desarrollar una plataforma computacional que incluye dos conjuntos diferentes de enfoques analíticos para invertir - enfoques analíticos y software para investigar datos ómicos relacionados con COVID-19.

Diapositiva 4

Así que como has visto antes, en la otra presentación, la fuente principal de esta pandemia es el virus, y nos estamos centrando en el genoma del virus como un recurso para investigar cómo se comporta el virus, y si hay proteínas o regiones específicas en el genoma del virus a las que debemos dirigirnos para el desarrollo de vacunas.

Diapositiva 5

La manera en que analizamos los datos es con los datos de secuenciación, obtenemos el genoma del virus a través de todos los infectados, los individuos en la pandemia. Y nos enfocamos en la variación del genoma también, en la región específica del genoma, incluyendo las proteínas y otras regiones como las proteínas no estructurales. La forma en que hacemos eso, la forma en que obtenemos estas secuencias del genoma para todos los individuos de nuestra población, intentamos calcular la variación que se explica en cada una de estas muestras en comparación con las otras muestras, e intentamos evaluar y ver qué significan esas variaciones.

Diapositiva 6

Aquí, estoy mostrando una hoja de ruta de cómo podemos calcular - cómo podemos ver las variaciones del genoma en regiones específicas. En la primera fila, lo que se ve es la variación del genoma, la correlación con otras regiones del genoma. Y vemos cuatro regiones, cinco aquí, que están altamente correlacionadas con la variación que tienen el genoma, que explica a través de las poblaciones. Y dos de esas regiones son realmente interesantes. Una de ellos es la región de la proteína de la espiga, que, ese número usted ve, 32.5, dice que la variación que vemos en el genoma del virus a través de la población se correlaciona entre el S y el genoma entero.

Y también vemos para las tres regiones NSP. En las otras tres regiones que se ven en el amarillo, hay grandes regiones y tienen subconjuntos. Es por eso que no nos enfocamos demasiado en ellos, porque esperábamos que cuando se toma una gran región del genoma, y se trata de ver cómo eso lleva a la información que todo el genoma toma, debe ser altamente correlacionado. Entonces, vemos aquí que hay lugares específicos en el genoma del virus, como esta proteína espiga, ya sabemos que facilita la entrada del virus a las células humanas. Esto está altamente correlacionado a la proteína no estructural 3, y esos dos juntos, se correlacionan con la variación entera del genoma que vemos en la población.

Como parte de nuestra subvención de la NSF, desarrollamos una estrategia llamada 'OmeClust'. Lo que hace, le das un conjunto de puntos aquí, los puntos son los individuos o las cepas del virus, y los tres - cuando lo ejecutamos utilizando la información de las tres regiones diferentes, en primer lugar, todo el genoma del virus. Segundo, la proteína espiga y tercero, la proteína no estructural. Vemos que esos tres nos dan las mismas comunidades. Y eso es consistente en todos ellos. Esto sugiere que estas dos distribuciones son importantes para ser focalizadas e investigadas más a fondo.

Diapositiva 7

También, observamos esta variación que se calcula a través de la población para ver cómo se correlacionan con los datos epidemiológicos. Por lo tanto, no teníamos mucha información,

excepto que el sexo y la edad son un buen ejemplo aquí, que el genoma del virus no se correlaciona bien con el sexo y la edad como se esperaba. El resultado podría no estar relacionado, pero no el genoma de la variación del genoma.

Diapositiva 8

También en nuestro trabajo, lo que estamos haciendo, en parte, es integrar los datos ómicos que medimos de individuos infectados, como proteínas o metabólicos. Queremos ver cómo esa información, cuando unimos todo junto, nos lleva a la hipótesis de los datos que podemos apuntar. Y, por ejemplo, lo que están viendo aquí, es un enfoque que estamos desarrollando llamado btest. Lo que hace, nos dan la información de los metabolitos de los pacientes y las proteínas del paciente y del cuerpo, nos dicen —nos da un bloque de relaciones, cómo, cuáles son los metabolitos que se relacionan con las proteínas en o los individuos infectados. Además, estamos desarrollando un enfoque de aprendizaje profundo, que le da los datos de secuenciación y le dice - individualmente, en la persona que se infecta, ¿cómo se verá la gravedad en función del genoma del virus que obtenemos de la persona infectada. Así que, como ven, estamos desarrollando diferentes, ya saben, enfoques, métodos, enfoques de aprendizaje profundo, para investigar los datos ómicos, el primer enfoque fue en el genoma, el genoma viral, y ahora nos estamos moviendo a las proteínas metabolizadas y tratar de integrar toda esta información utilizando el enfoque de aprendizaje profundo.

Diapositiva 9

Así que aquí hay varios estudiantes en mi laboratorio y también co-PIs de la biología computacional, del Instituto de Biología Computacional de Washington, colaborando en este trabajo. Y es un esfuerzo de equipo, para que podamos aprender más sobre nuestros enfoques y los resultados de COVID-19. Los publicamos en una página web de mi laboratorio. Y también los softwares que estamos desarrollando que podrán utilizar para investigar sus datos. Con eso, quiero agradecerles su atención.