

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

Transcripción de una presentación de Marc Riedel (Universidad de Minnesota), 16 de julio de 2021



Título: [EAGER: Predicción computacional y caracterización de la respuesta inmune a las infecciones virales](#)

[Perfil de Marc Riedel en la base de datos del CIC](#)

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés)#: [2036064](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de Julio 2021](#)

Editora de la Transcripción: Macy Moujabber

Editora de la Traducción: Isabella Graham Martínez

Transcripción

Diapositiva 1

Así que, hola a todos. Voy a hablar hoy sobre el trabajo computacional sobre el problema de la unión de péptidos y esta es una colaboración con George Vasmatzis de la Clínica Mayo.

Diapositiva 2

Así que, con COVID-19, todos hemos oído que la enfermedad se comporta como cualquier enfermedad. Hay un rango en la gravedad de los síntomas que la gente experimenta, pero es quizás más pronunciado con COVID-19 que otras enfermedades. Una buena fracción de las personas que contraen el virus no muestran ningún síntoma. La mayoría no presenta síntomas o son leves, pero una fracción significativa muestra síntomas graves. Y por supuesto, una pequeña fracción muestra síntomas críticos o incluso experimenta la muerte como resultado. Por lo tanto, hay muchos factores que influyen en la gravedad de la enfermedad y los expertos pueden hablar mejor de esto que yo. La edad, el sexo y, en particular, las comorbilidades son muy significativas y la exposición previa a virus similares juega un papel, pero una parte de la ecuación aquí en términos de gravedad de la enfermedad es también las diferencias innatas en nuestros sistemas inmunológicos. Esencialmente, los genes que hemos heredado de nuestros padres juegan un papel en lo bien que nuestros cuerpos combaten esta enfermedad o la mayoría de las enfermedades. Así que nuestro tema de estudio aquí es predecir computacionalmente un aspecto de la diferencia innata en la llamada respuesta celular del sistema inmune de un individuo.

Diapositiva 3

Por lo tanto, la respuesta celular es la primera línea de defensa que nuestros cuerpos tienen en respuesta a cualquier infección viral. Es el mecanismo por el cual los péptidos extraños introducidos en las células son troceados en pequeños fragmentos llamados péptidos. Estos luego pueden ser transportados a la superficie celular donde pueden unirse a receptores de superficie celular llamados MHC [Major Histocompatibility Complex] moléculas de clase 1. Si se unen de esta manera, las células infectadas se convierten en objetivos para las células T asesinas que pueden salir y matar a las células infectadas, ya sea muertas por una infección viral o como resultado, esta es la defensa más eficaz que tenemos contra el cáncer. Las células cancerosas también se matan de esta manera. Esta es la primera línea de defensa porque si se activa una vez que estamos infectados, el sistema inmunológico, las células T asesinas pueden matar a todas las células infectadas antes de que tengan la oportunidad de ponerse en marcha. Pero si la respuesta celular falla, entonces las células infectadas se convierten en fábricas. Comienzan a producir muchas copias del virus, y otros aspectos del sistema inmunológico tienen que hacerse cargo.

Diapositiva 4

Así, la respuesta celular es- como con todos los aspectos del complejo inmunológico. Entenderlo es muy importante. Es fundamental para comprender y predecir la gravedad de nuevos virus como el SARS-CoV-2, para el desarrollo de vacunas dirigidas a virus, y también para comprender los impactos de las mutaciones virales. Cómo diferentes virus afectarán a diferentes personas a través de la- su respuesta inmune celular innata, y como sigo volviendo al tema de la inmunoterapia contra el cáncer, que es el área de experiencia de mi colaborador- como muchas personas dirigimos nuestra atención a COVID-19, y reutilizado los conjuntos de habilidades. Así que, en este caso, las predicciones computacionales están dirigidas originalmente a la inmunoterapia contra el cáncer. Por lo tanto, y por supuesto toda esta misma teoría se aplica a las enfermedades autoinmunes. Así, la respuesta inmune celular, en su núcleo, es un problema computacional. Si tenemos los péptidos azules en la figura de la derecha este es el fragmento de proteína asociado con el virus. La pregunta es: ¿ese péptido azul se unirá dentro de un surco o una hendidura dentro de la molécula de la superficie de la célula amarilla, la molécula MHC 1? Y la respuesta inmune dependerá entonces de si este fragmento de proteína se une, y si se une lo suficientemente bien como para que las células T asesinas lo reconozcan. Los fragmentos azules que provienen del virus son novedosos. Por lo tanto, dado un nuevo virus, tendremos un conjunto completamente nuevo de péptidos. Las moléculas de la superficie de las células amarillas son determinadas por nuestros genes y cada individuo tiene hasta seis moléculas diferentes de MHC 1 determinadas por nuestra herencia, tres de cada padre. Y las moléculas de la superficie celular, las moléculas de MHC, están entre las más diversas en nuestro genoma, hay aproximadamente 21.000 variantes en la población humana. Esto no es un accidente. La evolución ha asegurado que este aspecto de nuestro sistema inmunológico sea muy diverso para que podamos sobrevivir a las infecciones virales del pasado.

Diapositiva 5

Pero esto también plantea un desafío computacional muy significativo, como voy a describir. Por lo tanto, el problema central que estamos abordando es cómo aplicar la informática para predecir si el péptido azul se unirá dentro de la molécula de la superficie de la célula amarilla, y para hacer esto para una gama muy amplia de péptidos, todos los asociados con el virus, y para la muy amplia gama de receptores de superficie celular las moléculas MHC 1. Ha habido mucho trabajo previo computacional.

Ha habido trabajo experimental que por supuesto tiene. Así, hay un paquete llamado NetMHC que se ha entrenado extensamente en los datos experimentales con la fuerza de unión para los pares conocidos de las moléculas amarillas azules- los pares del péptido MHC 1. Y basado en una estructura de red neuronal, este programa puede predecir, dado un nuevo péptido, qué tan fuerte predecirá. Y esta inferencia se basa simplemente en la secuencia de péptidos, por lo que la secuencia de letras, estos son los aminoácidos en el péptido se combinan con una etiqueta que corresponde a la molécula de superficie celular, la molécula MHC 1. Hay un punto fuerte. Esos son todos los datos experimentales. Y así, dado muchos miles de tales pares, la red neuronal puede ser entrenada. Y una vez que se entrena, se le da una nueva secuencia de péptidos, una nueva secuencia de aminoácidos, puede predecir qué tan bien ese péptido se unirá a una molécula MHC 1 dada. Así aquí, la secuencia del péptido sería la molécula del péptido azul en el dibujo, y la etiqueta correspondería al receptor superficial amarillo de la célula, la molécula del MHC 1. Esto es genial y las redes neuronales son poderosas en el sentido de que pueden entrenarse fácilmente y pueden hacer predicciones muy efectivas basadas en los datos que se les dan. Pero la limitación aquí es que no hay absolutamente ningún dato molecular, simplemente estamos entrenando etiquetas y letras, y también los datos de entrenamiento se toman para un conjunto muy diverso de datos experimentales. Muchos de ellos, por ejemplo, proviene del VIH, un virus diferente, y los péptidos asociados con el VIH, y el problema se da un nuevo virus como el SARS-CoV-2, la mayoría de los péptidos nunca se han visto antes, y una red neuronal hará predicciones que son falsas porque está haciendo inferencias de una región muy diferente del espacio del péptido. proporcionado información molecular sobre la estructura de estas moléculas, y el trabajo computacional que ha sido muy exitoso ha sido aplicar un aprendizaje automático, no es de extrañar para aquellos que tienen una formación en ciencias de la computación y redes neuronales.

Diapositiva 6

Y nuestro enfoque es hacer simulaciones a nivel molecular, y por supuesto ha habido mucho trabajo previo sobre este tema. Técnicas de simulación molecular muy sofisticadas son conocidas y ampliamente utilizadas. Uno se llama Dinámica Molecular. También hay simulaciones basadas en Monte Carlo. Usan una técnica llamada Recocido Simulado, y Acoplamiento Molecular es otro enfoque. El software disponible para tales simulaciones de nivel molecular es ampliamente utilizado, pero son especiales. Son de propósito general. Han sido desarrollados para clases amplias de moléculas y son extremadamente intensivos computacionalmente. Para tomar un péptido, un par de MHC 1, y para usar el software existente para simularlo, toma días, a veces semanas para simular un evento de unión único. Así que, semanas de tiempo de supercomputación para hacer una sola predicción. Y el alcance del problema que enfrentamos es que tenemos 21.000 variantes de las moléculas de la superficie de la célula amarilla, las moléculas MHC1, y para SARS-CoV-2, si nos centramos solo en la proteína spike y la cortamos en pedacitos para los péptidos, tenemos alrededor de 38.000 de esos. Entonces, estamos hablando de mil millones de combinaciones que queremos simular en términos de la cepa, y si toma una semana de tiempo de supercomputación, obviamente no tenemos mil millones de semanas para estudiar esto.

Diapositiva 7

Por lo tanto, nuestro enfoque es doble. Por un lado, estamos creando un software altamente personalizado para la simulación molecular y estamos utilizando los detalles del dominio en el que estamos trabajando. Empezamos con el péptido. Los péptidos no varían tanto en términos de su longitud o de su forma. Estamos empezando con los péptidos correctamente alineados dentro de la hendidura de

la molécula de superficie celular, la molécula MHC 1, por lo que no lo hacemos , y realizamos toda la búsqueda en el espacio torsional. Así que en lugar de mover la molécula alrededor del espacio tridimensional simplemente giramos y giramos sus enlaces para tratar de encontrar la configuración óptima. La otra contribución es que estamos desplegando esto a escala. Por lo tanto, estamos utilizando GPU, y luego, finalmente, la infraestructura de computación en la nube, para tirar la energía informática en el problema. Y nuestro objetivo, como dije en el título, es convertir mil millones de días en un millón de minutos o quizás un mes de tiempo de computación en la nube. pasar mucho tiempo simplemente rotando todo el péptido en el espacio. Lo colocamos exactamente donde debería estar

Diapositiva 8

Así que, el desafío está aquí y voy a pasar por alto esto- Estoy casi fuera de tiempo. Una cosa: los datos experimentales no están completos. No tenemos modelos moleculares completos, no solo de las 21.000 variantes, sino que ni siquiera tenemos una buena distribución geográfica. Esto está relacionado con algunas de las conversaciones anteriores. La mayoría de los datos experimentales son para la varianza de las moléculas MHC 1 de la demografía del Cáucaso occidental. Por lo tanto, el otro enfoque es que debemos inferir la estructura de las moléculas y luego simularlas.

Diapositiva 9

Y así, diapositiva final: el impacto de este trabajo. Bueno, es relativamente sencillo determinar para un individuo qué genes tienen ese código para las moléculas relevantes, las moléculas MHC 1. Se puede hacer a través de la tipificación de HLA, que se realiza para la prueba de paternidad. Dada esa información y dada nuestra infraestructura computacional, seremos capaces de predecir para un individuo, cómo ese individuo responderá a un nuevo patógeno. a un virus, a variantes del virus para diferentes individuos, y por el efecto de diferentes vacunas en diferentes virus para diferentes individuos. Y por supuesto, como dije al principio, este trabajo se aplicará no solo a los virus, sino también potencialmente a la inmunoterapia contra el cáncer y a las enfermedades autoinmunes. Así que me detendré ahí. Muchas gracias a todos.