

BIOINFORMÁTICA

La filosofía de los datos - ANÁLISIS DE DATOS

Hipótesis libre - ANÁLISIS COMPUTACIONAL

Evolución y enfermedad - EPIGENÉTICA

El difícil camino fácil - BIOLOGÍA DE SISTEMAS

Esto o nada - MODELIZACIÓN COMPUTACIONAL

JÓVENES INVESTIGADORES

El idioma de los datos



OBJETIVO: MEDICINA DEL FUTURO MÁS PRECISA

La actividad científica de vanguardia de los investigadores del CIC bioGUNE explora diversos aspectos de las ciencias de la vida, en la frontera entre la biomedicina, la biología y la química. Muchos de los proyectos en curso generan una gran cantidad de datos de distinta tipología, cuyo análisis es esencial para responder a preguntas relacionadas con salud y enfermedad. El objetivo último es desarrollar una medicina más precisa para el futuro.

La medicina de precisión puede definirse sucintamente como el enfoque más apropiado para maximizar la efectividad en la prevención y el tratamiento de enfermedades, considerando la variabilidad de cada individuo en sus genes, el medio ambiente en el que vive y su estilo de vida. El éxito en una iniciativa de medicina de precisión se consigue combinando avances científicos y tecnológicos en distintos ámbitos de las ciencias de la vida y de la ciencia de datos, con un cambio en la cultura de la práctica clínica y la investigación en salud. En este sentido, en nuestro Centro se generan mensualmente millones de datos, que provienen de distintas medidas experimentales, cuyo análisis riguroso representa un gran reto científico y tecnológico.

“El análisis de datos nos permite estar en condiciones de proporcionar un diagnóstico, pronóstico o una respuesta clínica más precisa y eficaz”

Así, durante los últimos tres años hemos apostado por ampliar nuestra especialización para dar respuesta a este reto a través de la Bioinformática. La bioinformática es un campo interdisciplinario que aprovecha las ciencias de la computación, las matemáticas, la física y la biología, esencial para la gestión de Big Data en el ámbito de la medicina de precisión. En concreto, los datos generados proceden de distintas herramientas científico-tecnológicas que miden la presencia de ácidos nucleicos, proteínas, lípidos, glicanos y metabolitos. El análisis de estos datos, unidos a los datos clínicos y los metadatos de los individuos, nos permite estar en condiciones de proporcionar un diagnóstico, un pronóstico o una respuesta clínica más precisa y eficaz.

Jesús Jiménez Barbero

Director Científico - CIC bioGUNE



Urko Martínez FILOSOFÍA de los datos

“Biólogo de números”, poco dotado para el trabajo en un laboratorio tradicional, reconvertido a la investigación genética de las enfermedades a través del análisis masivo de datos. “No sé si soy experto en ADN o en técnicas para analizar montañas de datos”. Urko Martínez llegó a CIC bioGUNE desde Estados Unidos, donde se especializó en el estudio de las diferencias genéticas para predecir más o menos riesgo de padecer una enfermedad.

“No podemos predecir qué enfermedad vamos a padecer, sí cuantificar riesgos de desarrollar una u otra”

Pregunta.- ¿Ahora en CIC bioGUNE?

Urko Martínez.- Estudiamos cómo aplicar a la prevención casi de cualquier enfermedad un buen predictor que desarrollamos en Estados Unidos. Ya no observamos sólo el ADN sino que lo hacemos con tendencias a la activación o desactivación de determinados genes, las diferencias entre personas con igual ADN. Aplicado a la enfermedad de Cronh, podríamos prever si los casos van a ser más o menos graves.

P.- Concretamos un poco más

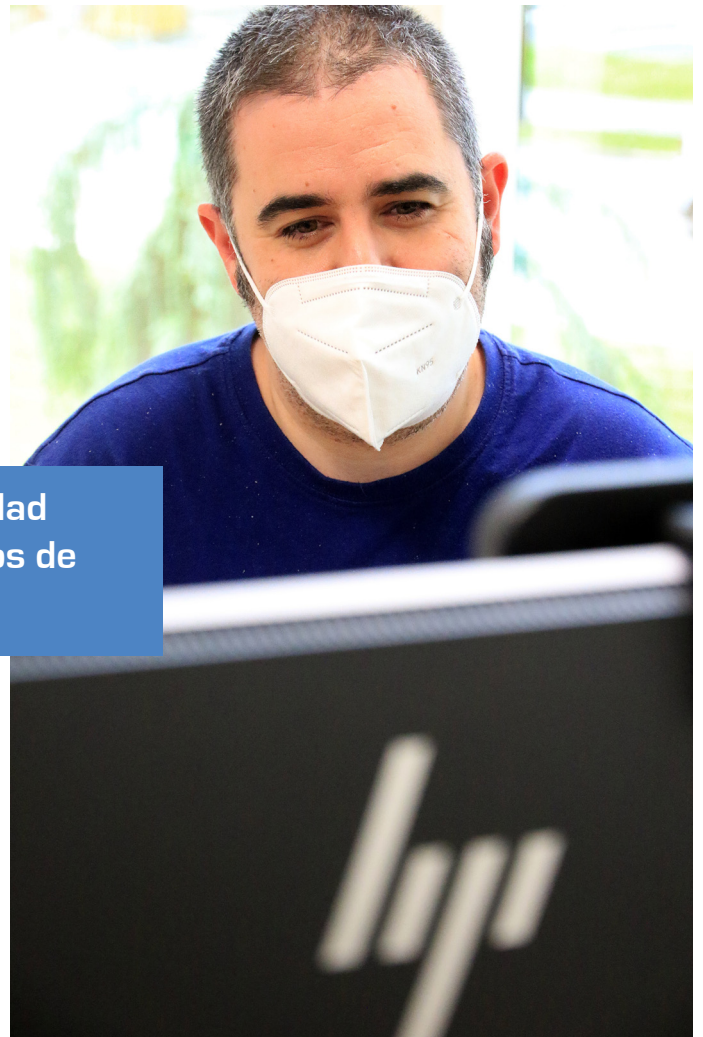
U.M.- El genoma hace ADN, este ARN, el ARN hace proteínas, éstas células, después tejidos, órganos y personas. Yo estudio el grado de activación de los genes, el transcriptoma, el que condiciona todo ese proceso. Soy especialista en datos ómicos, en el estudio de fenómenos biológicos, cómo se producen.

P.- ¿Cómo trabajas? ¿Qué datos manejas, qué buscas en ellos, cómo los manejas?

U.M.- Estoy centrado en la enfermedad Cronh. Trabajamos con muestras de sangre de pacientes con esta patología. Secuenciamos el ADN y observamos la distribución de moléculas y la relación entre genes. Trabajo con tablas muy aburridas con datos de miles de personas. Tengo millones de mutaciones de cada genoma. Busco aquellas que se encuentran más en personas enfermas que en personas sanas, mutaciones que aumentan el riesgo pero que no conducen necesariamente a la enfermedad. Busco combinaciones de mutaciones que me permitan predecir quién va a tener tal enfermedad y quién no.

P.- La capacidad de predecir enfermedades... ¿posibilidad o realidad?

U.M.- Nunca va a ser así. Mirando el genoma nunca vamos a saber qué enfermedad vamos a tener, en qué vamos a ser buenos o no. Tiene que ver con todo lo que le sucede a una persona y con nada. Lo que sí podemos



hacer es cuantificar riesgos, predecir quién tiene más posibilidades de desarrollar una enfermedad u otra.

P.- Algún ejemplo concreto...

U.M.- Sí, el caso de la Diabetes Melitus tipo 1. Es una enfermedad de las que mejor se pueden predecir... al menos el nivel de riesgo. Costaría poco estudiar el riesgo de los niños de padecer esta grave enfermedad, analizando un gen concreto, determinante en el desarrollo de una diabetes. Conocer el riesgo te podría ayudar a evitar comportamientos y hábitos peligrosos con consecuencias fatales...

P.- Volvemos a los datos... trabajas siempre con un ordenador.

U.M.- Uso la informática para aprender sobre el genoma. La capacidad de analizar datos te permite reciclar aquellos de los que dispones para utilizarlos en tus investigaciones. Al final todos nos beneficiamos de todo. De hecho, más del 90% de los datos con los que trabajo los han generado otros equipos, en la inmensa mayoría de los casos para estudios muy diferentes al mío.

P.- ¿Cómo manejas tantos datos?

U.M.- El 90% de nuestro trabajo es identificar fuentes de información y homogeneizar datos para

combinarlos y hacer un estudio estadístico. Buscamos puntos en común de diferentes estudios que puedan servirnos para desarrollar un proyecto concreto.

P.- Son los datos los que hablan o hay una decisión, una selección previa.

U.M.- En el hospital de Galdakao estudiamos células concretas, las que después generan anticuerpos. Pero la tecnología de la que disponemos nos permite también analizar datos sin una pregunta previa, sin hipótesis de

partida. También me permite responder a preguntas concretas, dirigir la investigación por un camino y con un fin concreto. Digamos que aquí hay dos modelos diferentes, no necesariamente enfrentados: por un lado el científico puro, el que plantea una hipótesis que después refrenda o descarta, por otro dejar que sean los datos los que definan el camino, los que de alguna generen diversas hipótesis.

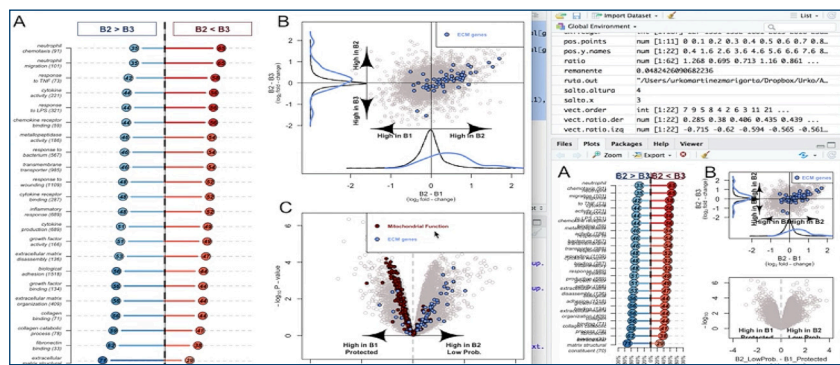
P.- Pero es difícil evitar un sesgo, tanto al inicio como durante el camino.

U.M.- Ciertamente la hipótesis libre no lo es del todo porque ponemos la cámara donde sospechamos que vamos a obtener unos datos. Y esa es la parte del riesgo del análisis masivo de datos, que puedes incorporar demasiados sesgos... Un ejemplo, la policía Estados Unidos analiza quién comete más delitos. Para ello coloca cámaras y observa que es mayoritaria la población negra delincuente. Por ello en los controles para más a negros que a blancos... pero ¿dónde ha colocado esas cámaras? Este es un sesgo que condiciona el resultado de toda la investigación.

“El análisis masivo de datos ha propiciado un cambio filosófico”.

P.- Esto obliga a estar siempre alerta para que el sesgo no contamine el resultado...

U.M.- El análisis masivo de datos ha propiciado un cambio filosófico, ha confirmado que hace falta validar todo lo que ves. Y en muchos casos no se hace. A veces caminamos, aunque sea de forma inconsciente, hacia donde nos gusta ir, no hacia donde deberíamos ir. Sólo validando confirmas el impacto del sesgo, más presente en las investigaciones de lo que nos gusta reconocer. Ahora, al aplicar el AMD, podemos replicar todo lo que se invente. Pero los sesgos ahora, los inconscientes o conscientes, se producen muchas más veces. A veces



Una pantalla de trabajo

nos dirigimos, no siempre de forma consciente, hacia donde nos gusta ir.

P.- ¿Vale todo con tal de publicar?

U.M.- Todos los científicos recibimos importantes cantidades de dinero público. Y, de alguna manera, tenemos que dejar registrados los resultados de nuestras investigaciones... Publicar es lo que te permite decir que tienes una pieza de información única, que antes nadie ha descubierto. Ese artículo, dependiendo del tipo de publicación que lo recoja, te da más o menos prestigio. Te da credenciales como experto en temas concretos y, por lo tanto, te facilita el camino para obtener financiación a proyectos que puedas plantear en el futuro.

P.- ¿Qué es más importante, el análisis de datos o el dinero?

U.M.- A veces cuesta convencer al partner de la importancia de invertir dinero en la recogida de datos. Un ejemplo: aún hoy se registra la altura y peso de los niños en cartillas de papel... Si todos esos datos se recogieran de una determinada forma, en tablas ordenadas en bases de datos, dispondríamos de un volumen notable de ellos que nos permitiría llegar a conclusiones interesantes.

P.- Un alegato a favor del análisis masivo de datos...

U.M.- Tener datos o no nos permite hacer preguntas y obtener respuestas de manera diferente a como lo hacen otros laboratorios. Lleva implícito un cambio filosófico... aún con todos los sesgos, disponemos de tantos datos que puedo dejarles que me lleven a un determinado lugar. Hay decenas de genes interesantes pero al combinar tablas y multitud de datos me ha salido que este gen es especialmente interesante. En los laboratorios tradicionales el investigador ha de enamorarse previamente de un gen, elegir bien y después aprenderlo todo de él. El riesgo de equivocación es tremendo y la frustración grande. El análisis masivo de datos nos permite apuntar mejor, acertar más.

“Ahora dirijo a gente, sin dejar de hacer colaboraciones. Tomo decisiones acerca de qué tablas de qué estudios coger, de cómo combinarlas y qué metodología estadística usar. Valoramos resultados y los comparamos para validar hipótesis”



HIPÓTESIS LIBRE... dejar hablar a los datos



Desde Italia a Euskadi, pasando por Oxford, Estocolmo y Melbourne. Biólogo de formación, investigador de las enfermedades gastrointestinales, **Mauro d'Amato** deja hablar a los datos, cruza y compara montañas de ellos para identificar polimorfismos, evaluar riesgos y, quizás, diseñar test para focalizarlos y minimizar su posible impacto.

“Las enfermedades son diferentes, no los métodos. Los de la genética son los mismos. Lo que nosotros hacemos es utilizar datos que obtenemos de biobancos, de todo tipo de cohortes internacionales, porque necesitamos muchos datos. Investigamos enfermedades complejas, en las que la contribución de cada gen es muy pequeña. Hay cientos de ellos que tienen su propia contribución a la enfermedad y para identificarla, necesitas tener muchísimas muestras y muchos pacientes”.

Mauro d'Amato valora como una ventaja no necesitar una hipótesis de partida. Observa todo el genoma, variaciones de frecuencia, polimorfismos, “pequeños cambios en el ADN de las personas” y si son más comunes en pacientes con una determinada enfermedad o en personas sanas. “Al final podemos describir un riesgo más alto en pacientes en relación con estos polimorfismos simplemente porque son más comunes en los pacientes”.

El objeto de estudio preferente son enfermedades complejas en las que hay decenas de genes que de una u otra forma contribuyen al desarrollo de la patología. Pretende graduar el nivel de riesgo a partir del número de polimorfismos y variaciones genéticas.

“Ahora podemos hacer estudios de un tamaño enorme con millones de muestras. Podemos hacer estudios del genoma de toda la población. Vamos buscando enfermedades concretas, comparando un polimorfismo con los de personas que no padecen esa patología. Todo es trabajo de ordenador. Todo análisis computacional. Si obtenemos resultados muy concretos, cuando has

identificado el gen y el polimorfismo, el cambio del ADN responsable del riesgo, pasamos al siguiente escenario, el análisis funcional en laboratorio. Y para cerrar el círculo CIC bioGUNE es perfecto porque trabajamos con investigadores experimentales altamente especializados”.

Los últimos 4 o 5 años ha coordinado un proyecto en el que han participado más de 100 autores de decenas de instituciones de todo el mundo, una investigación con premio porque en breve será publicada en Nature Genetics, “un estudio de todo el genoma relacionado con el colon irritable, del que aún no puedo avanzar muchos detalles”. D'Amato explica que es una de las enfermedades que más le interesan, una enfermedad funcional en la que no está implicado órgano alguno, que se desconoce por qué se desarrolla en algunos individuos. “En estos casos es difícil encontrar factores genéticos de riesgo porque las razones son múltiples. Es muy complicado encontrar los genes que pueden contribuir al desarrollo de la patología. En este campo cualquier hallazgo es muy importante”.

Enfermedad funcional y predisposición genética...

Pero se sabe que en el colon irritable juega un papel fundamental la predisposición genética, “si no nuestra investigación no tendría sentido”. Para realizar el estudio es preciso dividir el colon en subtipos. “Y hemos identificado a pacientes que tienen un problema genético que les complica la digestión de hidratos de

carbono. Tienen un problema con una encima concreta, un defecto genético del que antes no se sabía nada". Saberlo abrió la puerta al tratamiento, bien modificando hábitos nutricionales y de vida, bien suministrando al paciente una encima recombinante que les permite corregir el defecto genético concreto.

El reto ahora es desarrollar un test concreto que permita identificar con cierta facilidad aquellos pacientes con colon irritable y problemas para digerir carbohidratos. "Nosotros haríamos el test y el análisis del ADN. Si localizamos la mutación podemos caracterizarla in vitro en laboratorio, verificar que la encima no funciona y, a través de un nutricionista del equipo, ver con el paciente qué se puede hacer. Pero necesitamos algunos socios inversores para desarrollar este proyecto".

"Estos son los estudios genéticos de enfermedades gastrointestinales a gran escala a lo que dedico todo mi tiempo. Trabajamos con muchas enfermedades: funcionales como el colon irritable y genéticas como las hemorroides. Ahora estamos trabajando sobre la colitis microscópica, una enfermedad que afecta a mujeres mayores, una patología muy importante de la que nada se sabe y de la que hemos obtenido resultados muy interesantes".

Pero el logro más importante sería contribuir desde el análisis masivo de datos a la prevención de enfermedades. Identificada la actividad biológica se puede intervenir y corregir. El perfil genético de cada persona se puede caracterizar desde el nacimiento. Se

pueden observar polimorfismos y calcular riesgos, con valores específicos para cada enfermedad. "Y te puedo decir qué enfermedades son las que más riesgo tienes de padecer a lo largo de tu vida. Ahora estamos en una fase avanzada de la investigación y nuestro objetivo es desarrollar una utilidad clínica a partir de la información

"El futuro es la medicina personalizada. Antes es preciso invertir, tener caracterizada a la población"

que hemos obtenido".

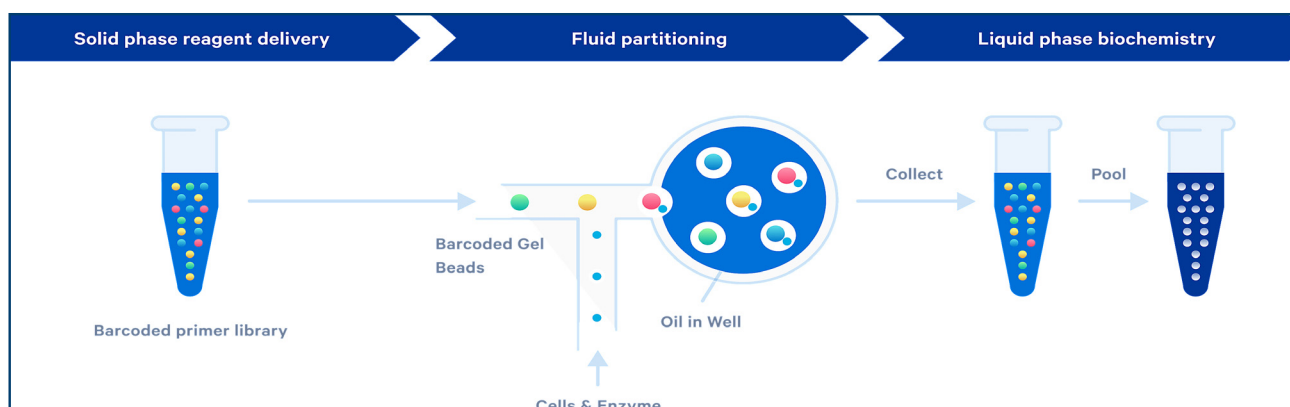
El futuro es la medicina personalizada, y para ello es preciso tener caracterizada a cada población, evaluar el riesgo de cada persona de padecer determinadas enfermedades. "Si pretendes desarrollar en el País Vasco la medicina personalizada basa en la genética, has de disponer de esa información, y un buen punto de partida es el Biobanco vasco. Hace falta inversión y confío en que el Gobierno Vasco finalmente lo haga".

"En el Biobanco hay muestras biológicas y datos clínicos. Pedimos las pruebas, hacemos el genotipado y después analizamos los datos genéticos en relación con las variables clínicas. En el País Vasco hay un montón de datos monitorizados pero falta la información genética. Y para generalizar formas de actuar, de tratar, necesitas trabajar con datos de las personas que viven en un entorno concreto, en este caso en Euskadi".

"Si pretendes trabajar con una población concreta has de disponer de su careacterización genética. Necesitamos trabajar con personas de un entorno concreto. La genética vasca es muy interesante, muy antigua".

Células: fuente importante de información

La tecnología Chromium Single Cell 3 'Solution de 10x Genomics' single-cell RNA-seq posibilita analizar transcriptomas célula por célula mediante el uso de particiones microfluídicas para capturar células individuales y preparar bibliotecas de ADN de secuenciación de próxima generación con código de barras.





EVOLUCIÓN VS ENFERMEDAD

De la biología a la epigenética, pasando por la estadística, la programación y la informática. Pocas palabras que hacen escasa justicia al desarrollo profesional de Isabel Mendizabal, desde hace apenas dos años en CIC bioGUNE. “Diferentes tipos de células desempeñan funciones completamente distintas aunque todas comparten idéntico ADN.

La epigenética es la ciencia que hace que cada célula apague o encienda interruptores para que desempeñe una u otra función”.

“La epigenética es la ciencia que hace que cada célula apague o encienda interruptores para que desempeñe una u otra función”.

Fue en Estados Unidos, en Atlanta, donde se involucró en un proyecto para entender la evolución del cerebro humano y la forma en la que afecta al desarrollo de enfermedades neuropsiquiátricas. En concreto, se trataba de comprender la relación entre la evolución reciente del cerebro humano y la esquizofrenia. “Es una pregunta que los científicos nos hacemos desde hace mucho tiempo. Se da la circunstancia de que la esquizofrenia es una enfermedad únicamente humana. Quizás sea porque es el cerebro humano el que más se ha desarrollado en muy poco espacio de tiempo”.

Posteriormente intentamos saber si existía alguna relación entre las metilaciones y el desarrollo de enfermedades mentales. Comparamos los datos de los que disponíamos de personas esquizofrénicas con otras sanas y comprobamos que las regiones del ADN que están implicadas en el desarrollo de esas enfermedades son las que han evolucionado de manera única en nuestra especie. Al final sabemos qué nos pasa a nosotros que no les sucede a los primates, animales que pueden sufrir ansiedad o estrés, pero que no desarrollan patologías psicóticas”.

“Sabía cómo, tenía las herramientas y la necesidad de ayudar. Incluso ahora la siento”

El punto de partida es relativamente obvio: comparar áreas equivalentes de cerebros humanos, de chimpancé y de macaco e identificar las variaciones que sólo se producen en los primeros. Una vez identificadas se trataba de ir más allá de la foto fija del ADN y su secuencia. Lo que se pretendía era estudiar la actividad en dicho ADN, qué genes estaban activos y cuáles no. “Nos centramos en las metilaciones, esas modificaciones químicas que de alguna forma determinan la función de cada célula en el organismo.

Se trató de una investigación básica. aún sin aplicación clínica. Pero ha servido para dotar a los científicos de herramientas para entender la esquizofrenia. Y sirve para enfocar mejor, por ejemplo, la investigación con modelos animales porque las estructuras cerebrales relevantes para la esquizofrenia han evolucionado de manera muy diferente en nuestra especie. “Hemos sentado las bases para que otras personas, otros científicos, desarrollen investigaciones más prácticas”. Hacia la aplicación clínica...Ahora, en CIC bioGUNE

intenta dotar a la epigenética de mayor aplicabilidad, trabajando con clínicos, conectar la investigación básica con la aplicación clínica. “La investigación desarrollada en Estados Unidos podría ser útil en el futuro, quién sabe. Ahora está de moda la generación de organoides para trabajar con ellos y es posible que quien desarrolle proyectos en ese campo algún día utilice nuestros resultados”.

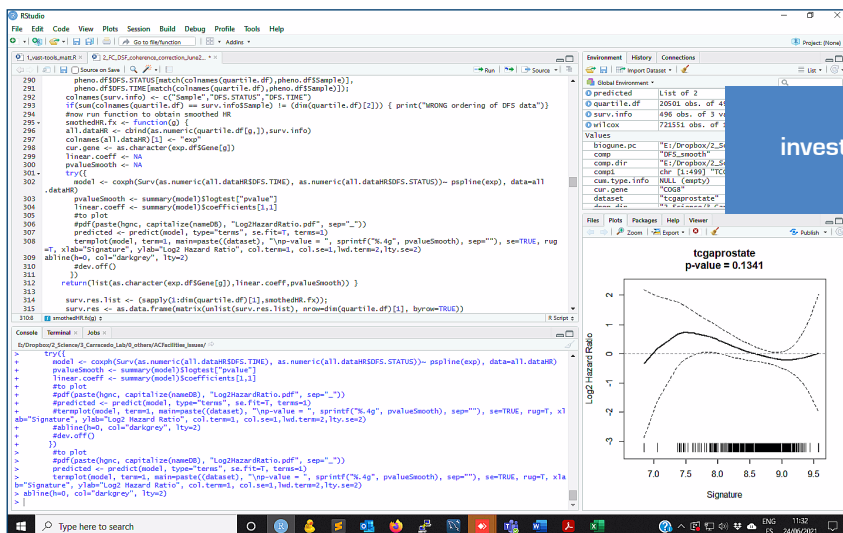
El día a día en el País Vasco es muy diferente. Isabel Mendizabal trabaja con un equipo de experimentalistas de CIC bioGUNE liderado por Arkaitz Carracedo, centrado en la investigación del cáncer a partir de modelos animales y sus colaboradores clínicos. “A él le interesaba focalizar el análisis masivo de datos y a mí entrar en contacto con un laboratorio con gran bagaje como el suyo, con publicaciones importantes en el campo del cáncer de próstata”.

Disponen de un importante volumen de datos de pacientes con esta enfermedad a los que se ha hecho un seguimiento desde hace años. “Intentamos entender

son esos indicadores de una mayor o menor actividad. “Estamos dando los primeros pasos. Llevamos alrededor de un año comparando datos recogidos en los últimos 14 o 15 años. De momento soy sólo yo quien bucea en montañas de archivos, pero la idea es reforzar pronto el equipo para desarrollar mejor esta tarea concreta”.

Son los primeros pasos de un proyecto que ha sido posible al unir análisis de datos y experimentación. “Llegué al CIC bioGUNE con conocimientos en biología evolutiva y epigenética, también en el manejo de montañas de datos. Pero nada sería posible sin la implicación de otros investigadores, de clínicos y de pacientes. Yo identifico genes, en el laboratorio se trabaja con ellos. El punto de partida: tejidos recogidos por clínicos a pacientes con cáncer de próstata. De hecho, el cáncer es una enfermedad muy evolutiva, de ahí que el salto que he dado en el campo de la investigación tenga mucho sentido”

La propia investigación está sujeta a una evolución



“Nada sería posible sin la implicación de otros investigadores, de clínicos y de pacientes. Yo identifico genes, en el laboratorio se trabaja con ellos.”

las razones de la diferente agresividad de este tipo de tumores, lo que redundaría en un mejor tratamiento, más ajustado a cada tipo de paciente y tumor. De alguna forma, aplicando nuestros conocimientos de epigenética, podemos predecir su evolución, si va a ser más o menos agresivo, más o menos grave”.

En buena parte gracias al trabajo de quienes durante décadas han ido recogiendo muestras, hoy se pueden comparar los datos de pacientes muy diversos. “Podemos comparar a personas que han convivido sin excesivos problemas con un cáncer de próstata, con otras que han sufrido las consecuencias de tumores de gran agresividad”. De hecho, el objetivo de la búsqueda

constante, a cambios muy notables, especialmente en las dos últimas décadas. El ritmo se aceleró gracias a la secuenciación del genoma humano, para ir cada vez más rápido gracias al imparable desarrollo tecnológico. De las montañas de papel a los potentes ordenadores en los que se recopilan, cruzan y analizan datos casi infinitos.

“Yo trabajo con datos, con enormes ficheros, haciendo comparaciones. Un billón de datos por cada individuo, multiplicado por 100. Trabajamos con ordenadores muy potentes, utilizando técnicas estadísticas y programas muy precisos. Es preciso saber programar, pero aún es más importante saber dónde y qué buscar”.

Colaboradores clínicos: Miguel Unda -Hospital Universitario de Basurto- y Enrique González Billalabeitia -Hospital 12 de Mayo y Hospital Morales Meseguer-.
Investigación apoyada por Ikerbasque y CRIS contra el cáncer.



El complicado reto de **HALLAR EL CAMINO MÁS FÁCIL**



Antonio Del Sol

En biología hay muchos problemas sin resolver, muchas preguntas a las que personas con formación en física teórica, como es el caso de Antonio Del Sol, pero también en matemáticas o informática, podrían ayudar a responder. Esta posibilidad, además de “la cantidad de temas interesantes” que ofrece aquella rama de la ciencia, fueron lo que le animaron a iniciar un camino que le ha llevado a centrar buena parte de su trabajo como investigador en la Biología Informática o Computacional.

“En biología hay ahora una gran aportación de datos. Es necesario desarrollar herramientas computacionales y modelos con los que trabajar”.

Palabras como “fácil” o “sencillo” no parecen asociables a la biología de sistemas. Y sin embargo lo son... “buscamos entre todas las combinaciones posibles la mínima para llegar allí donde queremos llegar, la forma más fácil y sencilla”, aunque se esté refiriendo a la tarea modificar la relación entre genes, entre células, entre tejidos... también para tratar enfermedades.

Antonio del Sol, desde Luxemburgo, explica que “trato de entender los procesos biológicos con un enfoque holístico, intento comprender escalas de organización en una célula, sus redes metabólicas, su señalización, las redes de regulación génica, cómo los genes se relacionan como un conjunto para que las células desempeñen una función determinada. Es como estudiar el planeta en su conjunto, las relaciones entre países, no sólo un país”.

Biología de sistemas

La biología de sistemas comenzó a nivel intracelular para, por ejemplo, comprender las redes de señalización entre células involucradas en un cáncer. Pero, cada vez más, los objetivos crecen en ambición y la comunidad científica trata ya de conocer cómo se relacionan las

células en un tejido, cómo la desregulación entre ellas puede conducir a una enfermedad, al envejecimiento también.

“El estudio de la enfermedad es cada vez menos reduccionista. En general son muy complejas, involucran a multitud de genes, requieren una visión más global... y esa visión nos la ofrece la biología computacional”. De hecho, Su búsqueda se centra en nuevos conceptos, en formas diferentes de abordar diferentes fenómenos biológicos... la enfermedad, la regeneración celular, la modificación de células.

Reconoce tener una cierta ventaja con respecto a otros investigadores ya que dispone de información detallada del punto de partida, también del objetivo final. “Construimos redes basadas en datos experimentales. El problema es que los investigadores estudian sólo aquellos genes que le interesan. Nosotros lo ponemos todo junto y generamos una importante base de datos que se irá completando con el paso del tiempo. Disponemos de una base de datos experimentales que desvela quién regula a quién. El objetivo final es la generación de modelos”.

Modelos para, por ejemplo, aplicar a la modificación de células, convertir unas en otras con un enorme

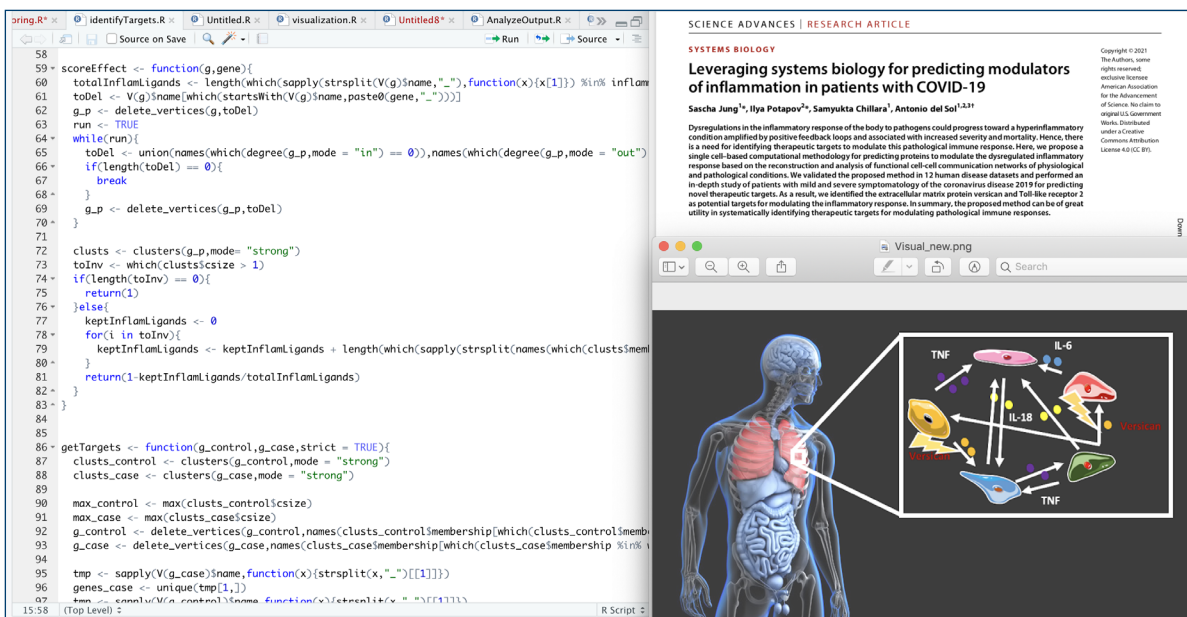
potencial terapéutico, especialmente en enfermedades caracterizadas por la degeneración celular, también por la no regeneración de células necesarias para sustituir aquellas que se van perdiendo y que dejan de desempeñar su función. “Nosotros desarrollamos modelos computacionales que construyan la asociación de los diferentes genes... Haciendo estos modelos computacionales a nivel de sistema podemos identificar el grupo reducido de genes cuya perturbación puede cambiar el tipo celular”. Es una información fundamental.

y desarrollar modelos de interacción celular que permitan entender los procesos de regeneración, muy potentes en este órgano. “El objetivo final es desarrollar modelos computacionales que nos permitan identificar las moléculas que puedan aumentar el potencial regenerador de los hígados más viejos”. Los investigadores disponen ya de la secuenciación de los tres tipos de ratón y se están analizando datos y generando predicciones. Básico para focalizar, para avanzar, para explorar, es la discusión. Discusión con colegas, con estudiantes,

“Nuestro trabajo es no repetir lo que otros hacen, en busca de nuevos conceptos, de nuevos enfoques, diferentes”

Antonio del Sol ha participado en proyectos que han logrado crear modelos para entender cómo un nicho de células madre modula el tipo celular y cómo contrarrestarlo en casos patológicos. “Por ejemplo, con la edad el cerebro de los ratones pierde células madres neuronales, además de tener un número creciente de células inactivas. Si comparamos un cerebro joven con uno viejo se observan diferencias claras... y se trata de saber cómo afecta el nicho en este proceso. Nosotros desarrollamos un modelo que trata de identificar las rutas de señalización y después comprobamos en laboratorio que si bloqueamos algunas de ellas hay un desplazamiento de la población de células pasivas a activas... Batalla contra el envejecimiento neuronal que todavía no se ha validado en humanos”. En bioGUNE investiga la regeneración del hígado. De nuevo con ratones, trabajan en el perfil genético de hígados jóvenes, viejos y geriátricos. Secuenciar célula por célula, obtener el mapa de expresión génica,

con experimentalistas, con informáticos, también con clínicos. “Nuestro trabajo es no repetir lo que otros hacen, en busca de nuevos conceptos, de nuevos enfoques, diferentes”. El objetivo es ambicioso: identificar mapas de nodos y links, redes de regulación génica, redes de comunicaciones intercelulares, redes de células en tejidos, de tejidos en órganos, de órganos en el cuerpo humano. “Ya hablamos más de redes que de nodos... es la biología de redes”. El objetivo futuro es el de desarrollar la biología computacional de tal forma que nos permita resolver problemas de solución imposible sin ella. “Seguir avanzando en el área de la regeneración como medicina del futuro. Contribuir desde el punto de vista computacional a resolver problemas médicos concretos. Elevar nuestros modelos y predicciones a la aplicación clínica, ver el resultado hasta el final. Y, cómo no, hacer predicciones y descubrir moléculas para el rejuvenecimiento celular”.



Probetas, pipetas, cultivos... datos y más datos



“Las movidas químicas de Gonzalo que él aun las hace más complicadas... tiene la molécula de la que conoce la estructura y usa modelos por los que deja funcionar al ordenador libremente, que se muevan a su antojo para después recoger los resultados. Es un estudio sin hipótesis de partida, una forma de crear realidad falsa”. “Lo de Gonzalo... a mí me explota la cabeza cuando intento comprender qué hace”. Son palabras de compañeros de Gonzalo Jiménez Osés en CIC bioGUNE, personas que trabajan en el análisis masivo de datos, reflejo de la complejidad la tarea de este químico de laboratorio que aterrizó en la modelización computacional durante una estancia en California. Y en ella se quedó.

ESTO O NADA



“Ahora somos seis, pero quiero llegar a 10, con una potente parte teórica y otra experimental”

El equipo que dirige Jiménez-Osés trabaja en dos direcciones... para explicar y experimentar, también para predecir. El punto de partida es una realidad en movimiento a la que se incorpora el factor tiempo, para saltar la foto fija. “Los glicanos son flexibles, se mueven, y nosotros, basándonos en su estructura y utilizando los principios de la física, la proyectamos en el tiempo. ¿Qué sucede si variamos la temperatura, si aportamos agua o iones, si los ponemos en condiciones realistas? Sucede que todo se empieza a animar, a mover en el ordenador. Añadimos la dimensión temporal, lo que nos da mucha información de procesos que sabes que

tienen que estar sucediendo, pero que no ves”.

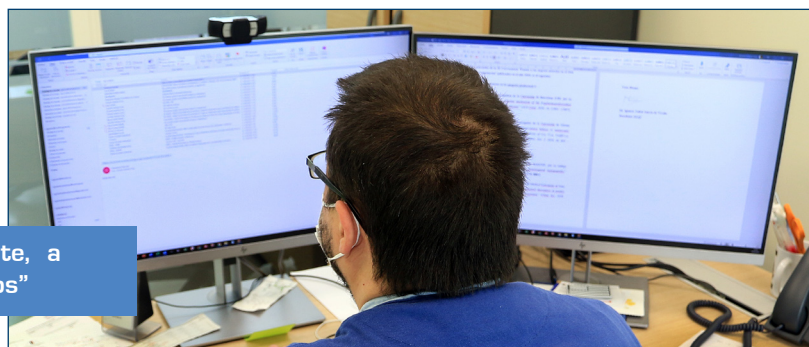
Pero todo se complica cuando ni siquiera se conoce cómo son las cosas, cuando desconoces la estructura. Entran en juego las hipótesis, “una especie de fake, pero es eso o nada”. En este caso la simulación se utiliza como banco de pruebas que después quizás sean validadas en laboratorio. “No sabes hasta el final del proceso si estabas en lo cierto o no”.

La colaboración con los laboratorios se antoja como una pieza fundamental del puzle “porque me quedo más tranquilo cuando valido el modelo”. En el diseño de enzimas, laboratorio suele ser sinónimo de frustración

porque en un elevadísimo número de casos es allí donde se demuestra que los modelos desarrollados en el ordenador no son viables, no son ciertos.

“Estamos obligados a ir siempre por delante, a imaginar escenarios nuevos y avanzar hacia ellos. Es uno de los males y de los beneficios de nuestra investigación. Nunca estamos quietos. No somos un grupo que haya logrado una estabilidad notable expresando cosas. Cada mes trabajamos con una metodología nueva. Siempre aprendiendo, siempre en terreno inestable... cambios de técnica, campos de fuerza nuevos, estamos probando siempre el primer modelo de vehículo. Esto o nada”.

“Estamos obligados a ir siempre por delante, a imaginar escenarios nuevos y avanzar hacia ellos”



Líneas de trabajo

El equipo de Jiménez-Osés trabaja en la modelización de la modificación de proteínas y anticuerpos, imaginando anticuerpos que puedan actuar en tejidos tumorales concretos. “Ayudamos a diseñar este tipo de uniones, nada trivial”.

También persigue un objetivo “tremendamente complicado”, el diseño de enzimas. “Conferir actividad catalítica a una proteína que no la tiene para que haga algo para lo que no ha sido diseñada. Infinitas combinaciones, una forma estructural determinada en el espacio... técnicamente mucho más complicado que encontrar una aguja en el pajar”. Pero se han obtenido resultados... en concreto cuando se ha publicado hace poco el resultado de una investigación que arrancó en 2013. “Diseñamos la primera metaloenzima diseñada con índices de selectividad mejores que los naturales para provocar una reacción que no existe y que tiene aplicación industrial”. Pero la publicación llegó tarde, en el peor momento de la pandemia “y sólo conseguimos publicar en Nature Chemistry cuando lo natural habría sido aparecer en publicaciones de primer orden”. Sabor agridulce tras 8 años de investigación.

Trabajan también en el descubrimiento de nuevos fármacos, “una línea más débil pero que debe estar en bioGUNE. Quizás lo sea porque los métodos de cribado masivo no son buenos... hay variables que no se consideran porque el objetivo es cribar mucho”. Y desarrollan modelos de reconocimiento de los carbohidratos con las lectinas, “campo, el de la glicobiología en el que somos muy valiosos porque estructuralmente hay muy poca información”.

El futuro...

“Hay que trabajar en proyectos asequibles, difíciles e imposibles”. Las líneas más productivas están consolidadas, “pero tenemos que replantearnos el diseño de enzimas porque las posibilidades de éxito son muy escasas”. En este contexto, Jiménez-Osés considera que hay que elegir entre las colaboraciones o las producciones propias, el primero un camino más sencillo con publicaciones seguras, el segundo mucho más complejo pero necesario para captar fondos para el desarrollo de proyectos.

El debate interno no cesa, en ocasiones con un componente filosófico notable. “En el mundo de la investigación hay mucho sesgo hacia lo experimental, lo que en ocasiones te hace sentir un poco comparsa. Pero en CIC bioGUNE el ambiente es diferente, nos sentimos útiles. Otros investigadores necesitan nuestros modelos y nos los pide”.

De alguna forma, él y su equipo son como un corcho en el océano, al capricho de las corrientes y las mareas, también de las modas. “Si todo es Covid-19, estamos obligados a sumergirnos en ese mar, aunque tendamos a escapar de aquello que todos hacen. Pero hay que tener en cuenta que es donde se reparte buena parte de el dinero”.

Gonzalo Jiménez-Osés quiere “llegar al punto de la bata, del investigador de toda la vida”. Su objetivo a medio plazo es conformar un grupo con una parte teórica y otra experimental... un equipo de 10 personas, 4 más de las que ahora lo conforman. Ahora son suministradores de posibilidades, de caminos a seguir, a explorar. Es una parte de la investigación, la suya, que rara vez se ve, que carece de visibilidad. No hay investigador experimental que en el laboratorio se ponga a mezclar a lo loco. Previamente hay un potente trabajo teórico, “duro trabajo que aquí, en CIC bioGUNE si goza de reconocimiento. Pero fuera no tanto, es más bien escaso”.

“El público rara vez se para a pensar en los cálculos previos a la construcción de un puente. Se queda en el puente físico. Y rara es la persona que sabe que hay modelización previa a una operación de corazón. Pero hay desarrollos teóricos previos, se llegan a simular intervenciones con modelos 3D”.



Suplementar magnesio mejora la evolución del hígado graso Jorge Simón...

La suplementación de magnesio ha demostrado ser una forma eficaz de reducir la mortalidad derivada de enfermedades hepáticas. De hecho, la hipomagnesemia es una condición asociada a patologías como la cirrosis o el cáncer. En esta línea, en enero de este mismo año el equipo publicó en la revista *Biomedicine* una recopilación de perturbaciones del magnesio asociadas a fallecimientos por hígado graso no alcohólico, tales como la obesidad, la hipertensión, las enfermedades cardiovasculares o la resistencia a la insulina.

La sobre-expresión de la proteína M4 (CNNM4), transportadora de magnesio, es la causante de la desregulación del catión en el desarrollo del hígado graso no alcohólico, tal y como se ha probado en modelos animales de ratón y en muestras de pacientes con distintos grados de la enfermedad (publicado en

Journal of Hepatology).

Estudios preclínicos *in vitro* han probado que silenciando dicha proteína reduce la acumulación de grasa por los hepatocitos, consecuencia confirmada *in vivo* en ratones a los que se sometía a diversos tipos de dieta. Silenciar la CNNM4 frenaba el desarrollo de la enfermedad hepática en la acumulación de grasas, la inflamación y la fibrosis. Y es más, ese silenciamiento promueve también la eliminación de grasas por parte del hígado.

Como aplicación clínica, en colaboración con Silence Therapeutics, el equipo ha desarrollado una molécula con RNA silenciador que permite evitar su degradación en el organismo y liberar la partícula de manera específica en el hígado. Un primer estudio pre-clínico ha obtenido ya resultados prometedores.

FAH, diana para el tratamiento de Tirosinemia tipo 1 Jon Gil...

La HT1 es una enfermedad rara, congénita, con una incidencia de 1 caso por cada 100.000 nacimientos. Daña fundamentalmente el hígado y el riñón y de no ser tratada a tiempo puede provocar la muerte del paciente antes de los 2 años. Un error en el catabolismo de la tirosina con consecuencias que pueden ser fatales.

Existe un tratamiento con nitisinona que actúa en la segunda de las cinco reacciones del metabolismo de este aminoácido, pero es en la quinta aquella en la que interviene la enzima FAH para catalizar esta última reacción. Con nitisinona y con una severa dieta mejora la evolución de la enfermedad, si bien es un tratamiento que no evita la acumulación de la tirosina y el riesgo de desarrollar cáncer hepático.

¿Qué provoca la disfunción de la FAH? Mutaciones genéticas que desestabilizan la estructura dimérica de la proteína. De ahí la hipótesis de que determinadas moléculas podrían servir para interactuar con la FAH,

estabilizar su estructura y formar el dímero. Fue el punto de partida de un muestreo con 2.500 compuestos, uno de los cuales se demostró que favorecía el estado dimérico de la FAH mutante.

A partir del compuesto líder se crearon cuatro nuevas moléculas, todas ellas con interacciones positivas con la FAH en la región de interés. Posteriormente se generó un modelo celular de HT1 al que se había incorporado una mutación de la enzima y, en presencia de algunos de los nuevos compuestos, se vio que la cantidad de los intermediarios tóxicos que favorecen la enfermedad se reducía de forma notable, con un efecto sanador de aquellos.

Un trabajo que demuestra que el tratamiento de pacientes con HT1 con las nuevas moléculas podría mejorar el ya indicado con nitisinona. Propone una base estructural para refinar moléculas administrables como chaperonas farmacológicas.

Mejorando la serología para la detección de anticuerpos Leire Egia...

La capacidad multiplex de nanopartículas detectables por citometría de flujo, ha permitido desarrollar una técnica capaz de detectar simultáneamente anticuerpos frente a tres antígenos del SARS-Cov-2, facilitando la interrogación de la presencia de anticuerpos en pocos microlitros de suero.

Se ha demostrado un método más sensible que los ELISA, también que la combinación de tres antígenos virales permite la detección de la seroconversión

completa. Esta técnica ha permitido detectar anticuerpos reactivos en individuos Covid-19 negativos, así como observar diferencias entre los niveles de anticuerpos de pacientes, dependiendo de la severidad de su reacción al SARS-CoV-2.

Se trata de un inmunoensayo de sencilla implementación, pero con una notable capacidad para detectar cantidades bajas de anticuerpos en comparación con los actuales métodos serológicos.

Jon Imano Quintana...

Galectina 4, proteína reconocedora de patógenos

Las galectinas conforman una de las familias de proteínas más relevantes para el reconocimiento de hidratos de carbono. Son proteínas que intervienen en numerosas actividades biológicas, que están involucradas en varias enfermedades.

El objetivo de esta investigación fue la galectina 4, con dos dominios capaces de reconocer carbohidratos, uno de ellos capaz de reconocer determinantes antigénicos de los grupos sanguíneos. Un estudio reciente demostró que las células tumorales son capaces de producir esta proteína en la superficie de sus células para interactuar con células sanguíneas y así extenderse. Pero las galectinas 4 también tienen la capacidad de reconocer patógenos que imitan a los grupos sanguíneos para provocar la muerte celular, de ahí

la necesidad de entender esta interacción para el desarrollo de fármacos con los que modular dichos acontecimientos.

La Resonancia Magnética Nuclear ha demostrado ser la mejor técnica para explicar cómo ocurre el reconocimiento entre moléculas. Ha proporcionado un modelo 3D que hace posible identificar qué parte del carbohidrato está involucrada en la interacción con la proteína, así como la parte que resulta afectada por la presencia de azúcar. Todo ello ha permitido determinar qué azúcares y de qué dominio son los preferidos por el primer dominio de la galectina 4, también qué tipo de patógenos reconoce mejor.

Este estudio sienta las bases para el diseño de moléculas reguladores de la galectina 4.

Marco Gómez Redondo...

Galectina 8 y estructuras supramoleculares

Las galectinas también juegan un papel importante en la respuesta inmune del organismo, en los procesos inflamatorios y en la regulación de la muerte celular. En este estudio concreto, el equipo investigador se ha centrado en la galectina 8, del subtipo "de repetición en tándem", muy presente en órganos como los riñones, el hígado, el corazón y los pulmones. Son proteínas capaces de formar estructuras grandes y complejas al poder relacionarse con receptores celulares que contengan carbohidratos diferentes. Están involucradas en la infección bacteriana o la autofagia. El estudio de la forma completa de la galectina 8 ha

permitido observar que sus dos dominios están unidos por una cadena peptídica. Los investigadores se han fijado como primer objetivo entender las propiedades de este nexo, lo que ha desvelado que en cada proteína sus dominios de reconocimiento de carbohidratos (CRD), cada uno de ellos, se comporta de forma independiente, uniéndose de manera específica a sus azúcares característicos.

La galectina 8 es, en consecuencia, una proteína de nuestro organismo que puede unirse a distintas entidades de manera simultánea y, de esta manera, formar estructuras supramoleculares.

Diego Barriales...

Inflamación del corazón por la bacteria Borrelia

La enfermedad de Lyme es la infección causada por la bacteria Borrelia que llega a los humanos a través de la picadura de una garrapata. Sin tratamiento antibiótico, dicha bacteria puede colonizar articulaciones, el sistema nervioso y el corazón, causando patologías severas.

En la infección cardíaca, carditis de Lyme, la presencia de aquella se contrarresta con la infiltración de macrófagos. Pero la bacteria es capaz de persistir en el corazón durante mucho tiempo mientras que la actividad inflamatoria de las células del sistema inmune altera la actividad cardíaca y genera diversas patologías. De ahí la importancia de estudiar cómo Borrelia es capaz de modular la respuesta de los fármacos para, en su caso, identificar nuevas dianas terapéuticas.

Partiendo de modelos de ratón: 1. Observar los cambios producidos en el corazón en el curso de la infección

(represión de genes y proteínas que aportan energía a las células cardíacas); 2. Con macrófagos, comparar la respuesta de estas células en un primer contacto con Borrelia con la de otras que ya habían estado en contacto con la bacteria (mejor respuesta en este segundo caso); 3. Analizar la expresión génica de este último grupo de células cardíacas (expresión diferencial de genes productivos de inflamación, una alteración notable del metabolismo en las células entrenadas, destacando el aumento de su actividad glucolítica).

El control de esta ruta metabólica en ratones ha derivado en un mejor control de la enfermedad, reduciéndose el número de bacterias, la infiltración de macrófagos y la inflamación, además de recuperar la expresión de genes reprimidos durante esta.

El paisaje metabólico del hígado

Jon Bilbao...

La complejidad del hígado dificulta de forma muy notable la caracterización precisa de un órgano que juega un papel central en todos los procesos metabólicos del organismo. En CIC bioGUNE se desarrolló un método basado en la RMN de ^{31}P para la identificación y cuantificación simultánea de multitud de metabolitos fosforilados, tanto hidrófilos como hidrófobos.

Con esta técnica se definió el panorama metabólico en hígados de ratón con porfiria eritropoyética congénita (CEP) e hígado graso no alcohólico, tanto genético como

con un suministro extra de grasas. Fueron reportadas alteraciones en las concentraciones de metabolitos en cuestión, alteraciones que en ratones con CEP mejoraron de forma notable gracias al tratamiento con ciclopirox.

El equipo concluye que estos hallazgos indican que la fosforómica basada en Resonancia Magnética Nuclear puede usarse para desentrañar los fenotipos metabólicos de la lesión hepática y para identificar el mecanismo de acción del fármaco.

El papel de una proteína en el desarrollo del cáncer

Laura Bozal...

Las ciliopatías son enfermedades consecuencia del mal funcionamiento de los cilios, las antenas de las células de los mamíferos, pequeñas protuberancias flexibles que emergen de la superficie celular para detectar y transmitir señales externas. Estos se ensamblan con la ayuda del citoesqueleto de actina, una red dinámica de filamentos en forma de malla que se extiende por el interior de la célula.

En CIC bioGUNE se ha comprobado que las células de los pacientes con el síndrome de Townes-Brocks (TBS) -malformaciones de los dedos, discapacidad auditiva, problemas renales y cardíacos...- tienen cilios defectuosos. Y también ha sido reconocida la existencia de una proteína, la LUZP1, que regula la formación de citoesqueleto y cilios primarios. Y dado que las

alteraciones en el citoesqueleto y el cilio primario dan lugar a cambios en la progresión del cáncer y en el desarrollo de metástasis, el equipo investigó si esta proteína también está implicada en esta patología.

Mutaciones de LUZP1 están presentes en 4 de cada 10 pacientes de muchos tipos de cáncer, mientras que la expresión de esta proteína es significativamente menor en el caso del cáncer de próstata. Y se observó que la eliminación total de LUZP1 promueve la migración e invasión celular, al tiempo que reduce la viabilidad de las células y aumenta su muerte por apoptosis.

Todo ello prueba el importante papel de esta proteína en la regulación de las características del cáncer a través del control del citoesqueleto de actina.

