

La secuenciación del genoma humano completada al inicio de este siglo está impulsando decisivamente los avances en la biología y la genética molecular y ha revolucionado la investigación biomédica, de modo particular en enfermedades de origen molecular y génico como el cáncer. Además, una nueva generación de **nanotecnologías biomoleculares** y *biochips* de altas prestaciones están proporcionando **cantidades masivas de datos** sobre **miles de entidades moleculares** como genes, proteínas, etc, medidas en muestras biológicas. Dichos datos esconden muchos de los secretos que permitirán avanzar en la comprensión y curación de enfermedades complejas como el cáncer. Sin embargo, debido a la cantidad, complejidad y diversidad de la información disponible, es necesario la utilización de herramientas computacionales que permitan extraer, organizar y almacenar el conocimiento biológico subyacente. Ello ha dado lugar a una nueva rama de la ciencia llamada **bioinformática o biología computacional**.

Desde una perspectiva histórica, la primera mitad del siglo XX fue dominada por los avances en la física, la química y las matemáticas. La aparición de nuevos instrumentos científicos como los aceleradores de partículas y el desarrollo de herramientas matemáticas posibilitaron grandes avances en la física. La segunda mitad del siglo XX ha sido dominada por la biología y las ciencias de la computación. La **biología computacional** se sitúa en la intersección de ambas áreas. De la misma manera que los avances de la física fueron propiciados por nuevos instrumentos de medida acompañados por el desarrollo de las matemáticas, los avances de la biología y la medicina requerirán el desarrollo de la biología computacional y la bioinformática para traducir la información proporcionada por los nuevos instrumentos de observación molecular en conocimiento biológico.

De modo particular, en los últimos años la **bioinformática** se ha convertido en una **herramienta indispensable** en la **investigación** mas avanzada en el ámbito biomédico del **cáncer**. Si se realiza una búsqueda simultánea de las palabras “*bioinformatics*” y “*cancer*” en Google se obtienen más de 10 millones de referencias. Las más relevantes, corresponden a nuevos micro y nano-dispositivos (“*biochips*”) desarrollados para análisis masivo de datos de genes, del genoma y del DNA. Por tanto, es razonable pensar que la utilización de la bioinformática en el estudio del cáncer ha sido motivada principalmente por la necesidad de analizar grandes cantidades de datos complejos, proporcionados por diferentes tecnologías y chips diseñados para estudios “genómicos”.

La biología molecular ha sufrido una revolución en los últimos años impulsando decisivamente la investigación contra el cáncer. La **exploración de los genes y las moléculas** productos génicos (que son las proteínas) se realiza a través de una nueva generación de **tecnologías ómicas** que permiten analizar el estado de miles de entidades moleculares simultáneamente. Ello ha dado lugar a los campos de la biología conocidos como **genómica, proteómica, transcriptómica,**

etc, que estudian no sólo las entidades moleculares individuales sino los complejos mecanismos de comunicación e interacción entre ellas. El almacenamiento, gestión y extracción de conocimientos biológicos a partir de los datos proporcionados por las tecnologías *ómicas* debe ser realizado por un conjunto de disciplinas heterogéneas que constituyen la **bioinformática**.

Así, la *minería de datos* ha permitido obtener conocimiento de cantidades masivas de datos, la *estadística* o la *inteligencia artificial* han descubierto patrones moleculares asociados a enfermedades, dianas moleculares o han generado modelos predictivos para mejorar el diagnóstico y tratamiento. Las *tecnologías de la información* y *teoría de comunicaciones* han ayudado a entender los mecanismos de interacción entre diferentes tipos de moléculas y han facilitado el almacenamiento y consulta de información biológica.

La colaboración entre bioinformáticos y biólogos ha generado importantes avances en el **conocimiento** de las **bases moleculares del cáncer**. Además, la **bioinformática** ha sido aplicada con éxito en el diseño de nuevas estrategias de diagnóstico precoz y tratamiento personalizado del cáncer dando lugar a lo que se denomina **medicina molecular**. Así, Una nueva generación de biochips de DNA ha permitido avanzar en la evaluación de los riesgos de sufrir cáncer y su prevención (Bromberg, 2013), el diagnóstico e identificación de nuevos subtipos de cáncer (Martín-Merino, 2009; 2009a; 2012) que pueden requerir un tratamiento diferenciado, la predicción del pronóstico y respuesta a la quimioterapia (Bueno, 2013; Hatzis, 2011) permitiendo tratamientos mejor dirigidos o el diseño de drogas más eficaces y seguras (Yildirim, 2007; Motter, 2010).

En esta lección inaugural deseo transmitirles 3 ideas fundamentales:

- Las **tecnologías ómicas** han revolucionado la biología molecular del cáncer permitiendo analizar complejos procesos biológicos en los que intervienen cientos de moléculas que interaccionan entre si.
- La colaboración de la **bioinformática** y **biología computacional** es totalmente **indispensable** para transformar la ingente y compleja información proporcionada por las tecnologías ómicas en conocimiento biológico que pueda ser utilizado y consultado por médicos y biólogos.
- Los avances conjuntos de la **biología molecular** y la **bioinformática** se están traduciendo ya en una mejora del diagnóstico y tratamiento del cáncer.

BIOINFORMÁTICA EN EL ESTUDIO DE LA BASE GENÓMICA DEL CÁNCER

El cáncer es conocido como la enfermedad de los genes. Desde que en 1981 Eugenio Santos (director del CIC de Salamanca) descubrió el primer oncogen, se ha comprendido que el comportamiento aberrante de las células

cancerígenas es causado, ante todo, por mutaciones en determinados genes.

La mayor parte de los **casos de cancer** son producidos por **mutaciones somáticas de varios genes** (Vazquez, 2007). Estas mutaciones se van acumulando durante la vida del organismo como respuesta por ej. a la exposición a toxinas o fallos en la reparación y copia del DNA antes de la división celular. Las **tecnologías ómicas** han desarrollado biochips de altas prestaciones para identificar mutaciones o medir el grado de **activación** (nivel de expresión dirían los genéticos) de **cientos de genes simultáneamente** en muestras de cáncer. Con la ayuda de estos biochips es posible descubrir patrones génicos característicos de cada tipo de cáncer y predecir su pronóstico o respuesta al tratamiento.

La utilización de los perfiles de expresión de los genes ha permitido avanzar significativamente en la investigación y tratamiento del cáncer. Ello ha requerido el desarrollo simultáneo de diferentes algoritmos y técnicas bioinformáticas. A continuación destaco varios avances importantes.

- **Diagnóstico más preciso** de diferentes tipos y **subtipos de cáncer** atendiendo a los perfiles moleculares. Ello ha permitido reducir los errores y mejorar el tratamiento. Para ello es necesario la aplicación de *modelos predictivos* desarrollados por la inteligencia artificial.
- Identificación de **biomarcadores de pronóstico y respuesta** a la quimioterapia. Ello ha permitido realizar un diagnóstico más precoz, estratificar los pacientes y ofrecer un tratamiento más personalizado. Para avanzar en este campo ha sido necesario el desarrollo y aplicación de test estadísticos, algoritmos de minería de datos o teoría de la información.
- Reconstrucción de **redes de interacción de genes**. Como hemos comentado anteriormente los genes no actúan solos, sino que interaccionan con otros para realizar una determinada función a través de una serie de productos moleculares y proteínas. Además, los genes se organizan jerárquicamente de tal forma que, unos pocos genes regulan la activación de un gran número de genes. Conocer las redes de interacción de genes y, lo que es más importante, su estructura regulatoria tiene gran interés biológico para descubrir dianas moleculares y módulos o grupos de genes que realizan una determinada función. El análisis de las redes de interacción implica la utilización de numerosas técnicas de análisis de datos desarrolladas por la comunidad bioinformática.

Para ayudar en la obtención de nuevos conocimientos biológicos las **tecnologías ómicas** que analizan la expresión de los genes deben **afrentar numerosos desafíos**. En primer lugar, la mejora de los biochips para capturar nueva información sobre la biología del cáncer es continua. En segundo lugar,

nuevos avances en la bioinformática son necesarios para superar las limitaciones de muchas de las tecnologías ómicas existentes. Por último, cada vez que se desarrolla una nueva tecnología de observación biológica, es necesario el desarrollo de nuevos métodos computacionales específicos que transformen los datos en conocimiento biológico.

GENÓMICA EN EL DIAGNÓSTICO Y TRATAMIENTO DEL CÁNCER

Las tecnologías de las que he hablado anteriormente han proporcionado nuevos conocimientos biológicos que han permitido avanzar en la comprensión de los mecanismos genéticos del cáncer. Además, dichos conocimientos están siendo trasladados a la clínica (Gómez-López, 2008) generando mejoras en su diagnóstico y tratamiento. Aunque son frecuentes los estudios genómicos con aplicación médica comentaré 2 a modo de ejemplo.

En (Hatzis, 2011) se proporcionan los resultados de una investigación liderada por el “*Anderson Cancer Center*” en el que se ha evaluado la capacidad predictiva de un **subconjunto de genes** (seleccionados mediante algoritmos bioinformáticos) para determinar el **pronóstico y respuesta** a la quimioterapia de pacientes diagnosticados de cáncer de mama con HER2 negativo. Los pacientes fueron sometidos a quimioterapia estándar (Taxane-Anthracycline) previamente a ser operados y posteriormente sometidos a tratamiento endocrino si eran positivos a receptores de estrógenos (ER+). Utilizando **biochips genómicos** y modelos estadístico/bioinformáticos se diseñaron **predictores genómicos** para 3 eventos biológicos relevantes en cáncer de mama: sensibilidad al tratamiento endocrino, resistencia a la quimioterapia y sensibilidad. Utilizando dichos índices fue posible identificar los pacientes con alta probabilidad de sobrevivir con el tratamiento clásico de quimioterapia separándolos de aquellos que tenían alto riesgo de recaída. Dicho predictor genómico tiene gran relevancia médica ya que, aquellos pacientes para los que el predictor genómico sugiere mala respuesta al tratamiento pueden ser separados y sometidos a otro tratamiento potencialmente más eficaz.

Continuando con el estudio anterior, en (Bueno, 2013) se han presentado los resultados preliminares de una investigación que desarrolla un nuevo algoritmo para obtener **firmas génicas estables** relacionadas con el **pronóstico** y con otras variables clínicas como **sensibilidad a la quimioterapia**. Algunos de dichos genes han sido comprobados por estudios biológicos mientras que otros deben ser estudiados. Conocer los genes causantes de la respuesta a la quimioterapia y pronóstico es muy importante a nivel de obtener **nuevas dianas moleculares y diseñar nuevos tratamientos**.

MEDICINA DE REDES: UN NUEVO ENFOQUE DE LA BIOLOGÍA COMPUTACIONAL Y UN DESAFÍO PARA LA BIOINFORMÁTICA

A pesar de los 3 billones de dólares invertidos en descifrar el genoma humano y de los 138 millones de dólares en identificar genes asociados a diferentes tipos de cáncer estamos lejos de entender su etiología. Además, no se

han obtenido los conocimientos esperados para afrontar la curación ni para personalizar su tratamiento (González, 2012). Ello es debido a varios factores que dificultan el estudio de enfermedades complejas como el cáncer.

- El cáncer es una **enfermedad multigénica**, producto del efecto combinatorio de pequeñas alteraciones en diferentes genes que interaccionan entre si y, con frecuencia, difíciles de detectar. Además, genes, productos de genes como proteínas, pequeñas moléculas y factores ambientales interaccionan entre si formando una compleja red de interacción (Cho, 2012). Un **mismo perfil genético** puede dar lugar a **enfermedades diferentes**, dependiendo de factores ambientales o alteraciones en otros productos de la red.
- **Perfiles genéticos diferentes** pueden dar lugar a enfermedades similares (Cho, 2012; Barabási, 2011). Ello **dificulta** significativamente el estudio genético de la enfermedad y la **obtención de dianas moleculares**. Esta observación se puede explicar porque los genes forman grupos que interaccionan entre si para realizar funciones biológicas. Alteraciones diferentes en cualquiera de los componentes del grupo dan lugar a disfunciones similares.
- Para entender la aparición y desarrollo del cáncer no basta con entender su base genética, hay que estudiar los **fundamentos moleculares**, en los que las **proteínas** juegan un papel esencial (González, 2012). Las proteínas actúan de manera cooperativa para llevar a cabo rutas moleculares formando complejas redes de interacción. Las interacciones entre proteínas controlan los estados de salud y enfermedad y, profundizar en su conocimiento es esencial para comprender la evolución y aparición de la enfermedad y, desarrollar tratamientos eficaces. Sin embargo, descifrar la base molecular del cáncer es aún más complejo que su base genética.

La solución a los problemas anteriores proviene de un **nuevo enfoque** en el estudio de los componentes moleculares basado en **redes de interacción** (Barabási, 2011). Los **genes se comunican** con otras moléculas y forman grupos para realizar funciones, de manera similar a como actúan los individuos de una **red social**. Algunos genes están fuertemente conectados y participan en múltiples funciones esenciales como los líderes de una red social. Diferentes rutas y grupos se entrelazan e interaccionan tal y como lo hacen diferentes grupos y tópicos en una red social.

El **enfoque de redes** presenta varias **ventajas** para avanzar en el estudio de enfermedades complejas como el cáncer:

- El enfoque de redes permite **identificar módulos** o subgrupos de genes y productos moleculares que cooperan para realizar una determinada **función**. Es lo que en biología se denomina **camino molecular**. De esta

manera es más fácil estudiar el *efecto cooperativo de múltiples genes* centrándose en la función alterada más que en genes individuales.

El enfoque de módulos explica por qué un **mismo gen participa en enfermedades diferentes**, ya que módulos asociados a funciones diferentes pueden estar solapados.

También explica por qué **diferentes perfiles genéticos** dan lugar a la misma enfermedad. Puesto que un módulo está compuesto por múltiples componentes, la alteración de cualquiera de ellos puede dar lugar a la interrupción de la misma función produciendo enfermedades similares.

- El enfoque de redes facilita la búsqueda de *dianas terapéuticas* que produzcan *efectos secundarios mínimos*. La identificación de módulos y funciones alteradas en el cáncer permite reducir la búsqueda de dianas terapéuticas. La mayor parte de las drogas utilizadas alteran la actividad de moléculas situadas en la vecindad del módulo afectado produciendo múltiples efectos secundarios. El enfoque de redes permite predecir dichos efectos en fases muy tempranas del desarrollo del medicamento.
- El enfoque de redes permite inferir **relaciones multivariantes** entre **grupos de genes** asociados a una función y la **enfermedad** desarrollada permitiendo **interpretar** fácilmente los **resultados** desde un punto de vista biológico.

¿ Por qué la **bioinformática** es totalmente **indispensable** para avanzar en el **conocimiento de las redes de interacción** ?

La red de interacción de genes, proteínas y otras moléculas esconde gran parte de los secretos que permitirán avanzar en la lucha contra el cáncer. Sin embargo dicha **red** es **extremadamente compleja**. Así, se han identificado unos 25.000 genes que codifican proteínas. Debido al *splicing* alternativo el número de proteínas es mucho más elevado y todavía desconocido, así como el número de moléculas funcionales de RNA. Ello quiere decir que la red de interacción a estudiar tiene más de 100.000 componentes (nodos) y un número mucho mayor de interacciones. La mayor parte de las interacciones son desconocidas. La **construcción** de dicha **red** y la **extracción de conocimientos biológicos** sólo es posible mediante la aplicación y desarrollo de técnicas computacionales avanzadas en el contexto de la bioinformática. La colaboración de científicos que trabajan en este área es necesaria en varias tareas clave:

- **Avanzar en la construcción** de la **red de interacción** de proteínas y otras entidades moleculares. Las tecnologías ómicas junto con herramientas bioinformáticas son una alternativa fiable y barata a las técnicas experimentales.

- **Identificación de módulos, caminos moleculares y de entidades esenciales.** Es un problema muy estudiado en el área de minería de grafos y teoría de la información. Son constantes las mejoras en los algoritmos propuestos por la comunidad bioinformática.
- Estudio de las **asociaciones entre módulos y enfermedades.** La biología moderna ha pasado de estudiar asociaciones entre genes individuales y cáncer a buscar módulos funcionales o caminos desregulados para cada tipo de cáncer. La bioinformática se enfrenta al desafío de incrementar la eficiencia computacional y reducir los errores.

MEDICINA MOLECULAR EN EL TRATAMIENTO Y DIAGNÓSTICO DEL CÁNCER

La **medicina molecular** promete avances muy importantes en el tratamiento y diagnóstico del cáncer pero se encuentra en su **infancia**.

Uno de los tumores en los que se ha comenzado a aplicar con buenos resultados ha sido el cáncer de mama. De manera experimental, el centro del cáncer de Salamanca con el que colaboro, analiza un número muy reducido genes en el cáncer de mama para **mejorar la predicción del pronóstico** y personalizar el tratamiento. Es necesario dar el paso de evaluar más genes y sus interacciones utilizando las tecnologías *ómicas*. De la misma manera, **nuevos fármacos** contra el cáncer de mama se dirigen a **dianas moleculares específicas** como el “factor de crecimiento epidérmico humano” (HER2). Las investigaciones actuales tratan de profundizar en la red de interacción involucrada en la proliferación celular para diseñar drogas que inhiban ciertas interacciones reduciendo los efectos secundarios.

A continuación comento 2 ejemplos en donde el formalismo de redes ha permitido avances en la comprensión y tratamiento del cáncer.

- **Redes de interacción de proteínas** para estudiar cómo **factores ambientales** interactúan con las proteínas para producir estados de cáncer. Los trabajos de (Charlesworth, 2010; Ding, 2008) investigan cómo afecta la exposición a un factor ambiental como el **humo del tabaco** en los caminos y redes de interacción de proteínas. Estos estudios encontraron que los genes cuya expresión está alterada en fumadores mapean principalmente en caminos involucrados en varios aspectos de muerte celular y funciones respiratorias. Además, las investigaciones muestran que la exposición al tabaco interfiere grandes subredes de proteínas relacionadas con la respuesta inmune inflamatoria. Ello puede ayudar a entender cómo el tabaco influye en la aparición del cáncer. La exposición al tabaco perturba varias interacciones entre proteínas deprimiendo el

sistema inmune y alterando la respuesta inflamatoria. Las investigaciones realizadas ayudan también a entender por qué dejar de fumar es beneficioso, ya que es posible revertir gran parte de las alteraciones transcriptómicas y restaurar la mayor parte de las interacciones alteradas entre proteínas.

- Identificación de **módulos de enfermedades** y caminos que se activan en estados de **enfermedad** para obtener **marcadores de pronóstico**. (Chuang, 2007) ha desarrollado un método para, utilizando datos de expresión de genes junto con una **red de interacción de proteínas**, identificar caminos relevantes que resultan activados en la progresión del tumor dando lugar a **metástasis** en **cáncer de mama**. Dichos marcadores basados en la red de interacción de proteínas tienen un poder predictivo de la posible aparición de metástasis mucho mayor que los basados en genes individuales o los marcadores clásicos.

Predecir el riesgo de metástasis es muy importante para prescribir una quimioterapia más agresiva, ya que la aparición de metástasis es la principal causa de muerte en cáncer de mama. Además, conocer los módulos funcionales involucrados en la aparición de metástasis está aportando conocimientos biológicos valiosos para avanzar en su tratamiento.

CONCLUSIONES Y REFLEXIONES FINALES

La aparición de las tecnologías “ómicas” ha proporcionado cantidades masivas de datos sobre entidades biológicas como genes o proteínas. Ello ha revolucionado el estudio de la base *genómica* del cáncer, mejorando el diagnóstico precoz, el pronóstico y la predicción de respuesta a la quimioterapia. Las tecnologías “ómicas” han proporcionado también valiosa información sobre el papel de las proteínas y sus redes de interacción con otras moléculas en el desarrollo y aparición del cáncer. Ello ha contribuido a identificar nuevos biomarcadores y dianas moleculares más eficaces y seguras para tratar el cáncer.

Debido a las cantidades masivas de datos generadas por las tecnologías “ómicas” y a su complejidad es necesario la aplicación y desarrollo de técnicas bioinformáticas que permitan almacenar, procesar y extraer nuevos conocimientos biológicos. La bioinformática se muestra como una herramienta esencial en el trabajo diario del biólogo sin la cual no podría interpretar los resultados de sus experimentos. Todas las tecnologías *ómicas* aparecidas para el estudio de los genes y proteínas han requerido de un desarrollo simultáneo de estrategias y métodos computacionales para el análisis de los datos generados. La colaboración multidisciplinar entre biólogos, médicos y bioinformáticos será cada vez más necesaria para avanzar en la lucha contra el cáncer.

El cáncer es una enfermedad muy compleja y, faltan bastantes pasos para