



Godei
Editorial



Academia de Ciencias Matemáticas,
Físico-Químicas y Naturales de Granada

EL PUZLE GENÉTICO Y LA TECNOLOGÍA

DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICO CORRESPONDIENTE POR EL

ILMO. SR. D. IGNACIO ROJAS RUIZ

GRANADA, 2021





**Academia de Ciencias Matemáticas,
Físico-Químicas y Naturales de Granada**

EL PUZLE GENÉTICO Y LA TECNOLOGÍA

**DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICO CORRESPONDIENTE POR EL**

ILMO. SR. D. IGNACIO ROJAS RUIZ

Granada, 2021

EL PUZLE GENÉTICO Y LA TECNOLOGÍA

© IGNACIO ROJAS RUIZ

I.S.B.N.: 978-84-17970-63-5

Depósito Legal: GR 915-2021

Edita e imprime: Godel Impresiones Digitales SL

“No hay problema que resista el ejercicio
continuo del pensamiento”

François-Marie Arouet, Voltaire

EL PUZLE GENÉTICO Y LA TECNOLOGÍA

IGNACIO ROJAS RUIZ

**Excmo. Sr. Presidente de la Academia de Ciencias Matemáticas, Físico-Químicas y Naturales de Granada,
Excelentísimos e Ilustrísimos Señoras y Señores Académicos,
Señoras y Señores,**

En primer lugar, quisiera expresar mi más sincero agradecimiento a la Academia de Ciencias Matemáticas, Físico-Químicas y Naturales de Granada, y en especial a su Sección de Ciencias Físico-Químicas por haber propuesto mi candidatura para ingresar como académico correspondiente en esta ilustre Institución. Gracias a todos los académicos por su unánime acogida. Para mí es un honor pertenecer a esta

Academia, y manifiesto mi compromiso con sus objetivos fundacionales.

Hoy estoy ante personas de Ciencia, algunas de ellas, profesores que me adentraron en el fascinante campo de la Física y las Matemáticas, profesores a los que respeto y admiro, que forjaron mis características de investigador y docente.

Muchas gracias a su Presidente, Prof. Francisco González Lodeiro, por su apoyo. Le aseguro que pondré toda mi ilusión, saber y empeño, por estar a la altura de mis ilustres compañeros y cumplir con los propósitos y obligaciones de esta Institución.

Mi agradecimiento especial al Prof. Alberto Prieto Espinosa, anterior Presidente de la Sección de Físico-Químicas, porque gracias a su perseverancia, interés, afecto y cariño, hoy estoy aquí pronunciando, emocionado e ilusionado, este discurso. El Prof. Prieto fue mi profesor en el último año de mis estudios de Ciencias Físicas, pero Alberto ha sido siempre la persona que me ha ayudado a crecer como profesor y científico. Gracias, muchas gracias por ofrecerme la oportunidad de

comenzar a tu lado el camino universitario, y hoy por ser mi padrino y pronunciar el discurso de contestación.

Tecnología y biología computacional

La expansión, propagación y difusión de las Tecnologías de la Información y las Comunicaciones (TIC) es un proceso irrefrenable, que impregna prácticamente a todos nuestros hábitos de vida y estamentos sociales.

El desarrollo de nuevos dispositivos tecnológicos, software y aplicaciones, involucran tanto a los usuarios que las utilizan, como a las empresas e instituciones a hacer uso de ellos. Rápidamente se produce una evolución tecnológica, que implica descartar lo hasta la fecha utilizado (por decremento en la falta de prestaciones, potencia de cálculo, capacidad de memoria o requisitos tecnológicos que los hacen caer en desuso, siendo inapropiados para satisfacer los exigentes requisitos actuales), abrazándose velozmente hacia nuevas tecnologías que irrumpen en el mercado.

Varios son los avances tecnológicos que han irrumpido en los últimos años en el campo de las TIC, como son la llegada de 5G, internet de las cosas, la nube y cloud computing,

inteligencia artificial y ciberseguridad. Son tecnologías que evolucionan rápidamente y que se integran en nuestras vidas sin apenas notarlo [1]. Como ejemplo destacar la evolución de dispositivos avanzados, que con un hardware y software potente e interfaces amigables, nos permiten el acceso a multitud de aplicaciones e información.

En nuestro nuevo horizonte temporal, la generación de datos (de todo tipo) cobra una relevancia vital. Los datos son el nuevo "petróleo" o como también se ha definido, el oro del Siglo XXI. Según un estudio IBM, más del 90% de los datos generados a nivel mundial, se han producido en los dos últimos años. Durante el próximo minuto de este discurso se habrán realizado cerca de 4 millones de búsquedas en Google. Se estima que la generación de información generada desde el inicio de la humanidad, hasta el año 2003, podría ser del orden de cinco exabytes. Un exabyte son aproximadamente mil millones de gigabytes. Sin embargo, tan solo en el año 2007 se produjeron 281 exabytes y en 2011 se alcanzó los 1800 exabytes [2]. Analizando la evolución en la generación de información, es probable que podamos escuchar afirmaciones como: “cada día de 2026 generamos tanta información y datos como en todo el año 2021”, y esto será posible gracias al desarrollo y evolución en las infraestructuras TIC.

Mencionar como ejemplos simples y comunes de la generación de datos, el navegar por internet, escribir en redes sociales, envío de mensajes, descargar una app o comprar online. Esta información, aunque no seamos conscientes, es de un gran valor para las empresas y grandes compañías tecnológicas [3]. Estamos ante el conocido “Big Data”, con todos los retos tecnológicos que conlleva, tales como su gestión, manipulación, aplicación y niveles de seguridad [4], [5]. El “tráfico de datos” también presenta cifras en continuo crecimiento, estimando que tan solo con nuestros móviles, para el año 2020, el tráfico de datos supere el zettabyte (un billón de gigabytes). Para el año 2023 se estima que los datos a nivel mundial generados lleguen a los 72 zettabytes. Según cálculos que realiza Alberto Prieto en uno de sus vídeos de YouTube [6]; si utilizásemos unidades de disco USB actuales de 4 Terabytes; es decir de muy alta capacidad, para almacenar dicha información, poniéndolos uno encima de otro tendrían una altura de más de 388.000 Km; es decir, superior a la distancia de la Tierra a la Luna.

Pero, ¿toda esta información que producimos se procesa? No, por ahora generamos mucho y procesamos poco. Somos ricos en datos, pero pobres en conocimiento.

Los cambios tecnológicos afectan a la mayor parte de los ámbitos sociales, junto a las estructuras e infraestructuras de las instituciones, siendo siempre el principal, y más en los tiempos que vivimos, la salud.

Hay veces que el cine juega el papel de predictor del futuro. Voy a referirme a una película realizada en 1997, cuyo título es “Gattaca” (correspondería con una secuencia de ADN: Guanina, Adenina, Timina, Timina, Adenina, Citosina, Adenina) y que me ayudará a ir centrándonos en la evolución de la tecnología y su repercusión en la biología computacional, la genómica y principalmente, en la secuenciación del ADN.

El argumento de la película podría clasificarse de distopía, al presentar una sociedad ficticia indeseable en sí misma, siendo el objetivo principal la transformación de la población gracias al desarrollo de nuevas tecnologías, que perfeccionen las capacidades humanas, mejorando las competencias intelectuales y físicas de los individuos. El argumento se basa en una sociedad, donde los padres recurren a la ayuda de los avances tecnológicos en ingeniería genética, para la concepción de sus hijos, creando bebés que estén libre de enfermedades genéticas o problemas cardíacos, alta resistencia

física, buena visión e inteligencia superior, lo que les permitirá ser seleccionados como astronautas y colonos espaciales.

La competitividad se vuelve muy agresiva, casi feroz, dividiendo la sociedad entre los concebidos en laboratorios con un código genético en principio “superior”, con excelentes condiciones salariales, expectativas de éxito y con la posibilidad de conquistar otros planetas, y los nacidos de manera natural, consignados según la propia sociedad a tareas poco enriquecedoras o gratificantes. La película se centra en un niño nacido de forma natural, sin ayuda de ingeniería genética, que debe rivalizar contra aquellos niños que si recibieron dicho apoyo genético para su concepción, incluyendo a su propio hermano. Debido a su naturaleza “no-perfecta”, se le destina a tareas de escasa relevancia, pero de forma clandestina, se registra en la academia encargada de la formación e instrucción de astronautas que viajaran al espacio, realizando todo tipo de artimañas y argucias para que su genética “no-perfecta” no sea detectada. El desenlace de la película termina con el descubrimiento por parte del protagonista de que la ayuda genética no es la responsable de que exista disimilitud entre capacidad y no capacidad, cumpliendo finalmente su objetivo de viajar al espacio.

Cuando se realizó esta película, la secuencia del ADN humano era desconocida. En la actualidad, como describiré en el presente discurso, uno de los proyectos más sobresalientes en la historia de la Ciencia, ha permitido descifrar este enigma, el Proyecto del Genoma Humano.

Gracias a la colaboración conjunta de múltiples disciplinas, entre las que las Tecnologías de la Información y las Comunicaciones han jugado también un papel fundamental, se ha logrado un desarrollo e impulso muy fructífero en la investigación genómica y proteómica.

Millones de Gigabytes (es decir de Petabytes) de datos se generan y almacenan anualmente en la mayor parte de los Hospitales y Centros de Asistencia Sanitaria. Sin embargo, la mayor parte de esta información, tan solo queda almacenada y no procesada para poder ser utilizada en la mejora de la práctica clínica o descubrir novedosos avances científico-técnicos.

Estamos en los inicios del “Big Data Clinic”, que podría definirse como la aplicación de la metodología y tecnología de Big Data en el ámbito de la salud, utilizando todos los datos disponibles de la práctica clínica, aun siendo heterogéneos y

de diferentes bases de datos, para desarrollar una medicina de precisión y personalizada, que haga viable integrar, procesar y analizar la información global del paciente: historia clínica, imágenes médicas, pruebas diagnósticas, dietética, genómica, proteómica, etc. Estas nuevas tecnologías permitirán una transformación en el paradigma de la asistencia sanitaria a los pacientes, junto con relevantes avances en la investigación de nuevos fármacos. En la actualidad, los ensayos clínicos en nuevas terapias avanzadas (génicas, celulares y tisulares) son cada vez más ajustados al perfil de los pacientes gracias a las tecnologías que involucran el Big Data Clinic.

Los avances en medicina personalizada requieren de una nueva estrategia transversal, que además de las ventajas que hemos enumerado y que brindan las nuevas tecnologías, concilie la protección de los datos de carácter personal, las cuestiones legales y éticas. En efecto, la privacidad y confidencialidad de los datos es fundamental, y las nuevas plataformas deben de contar siempre con todas las garantías de protección.

Pero, ¿cómo ha sido la influencia de las Tecnologías de la Información y las Comunicaciones en el campo de la biología computacional? ¿Cuáles han sido las máquinas que han

permitido descifrar el enigma de la codificación de nuestro ADN?

Hubo un tiempo en el que el DNA era algo desconocido.

En la actualidad, son ya bien conocidas por los estudiantes en sus libros de texto respuestas a preguntas que se realizaban los científicos a principios del siglo XX, tales como: ¿De qué manera se transmite la información entre generaciones? ¿Qué molécula podía contener dicha información? ¿Cómo es su estructura? ¿Cómo interacciona?

El descubrimiento de la estructura del ácido desoxirribonucleico (ADN), fue un paso fundamental en el campo de la biología, siendo James Watson y Francis Crick, los que en 1953 consiguieron solventar este enigma, descubriendo la distribución espacial del ADN y resolviendo así este “puzle de cuatro piezas” (adenina, timina, citosina y guanina).

James Watson (licenciado con 21 años en los estudios de Física por la University College London) estaba fascinado por la molécula de ADN, pensando que era ella la garante de almacenar y transmitir la información genética. No obstante, la

estructura química tridimensional del ADN era desconocida y un gran reto a dilucidar. Las claves para dicho descubrimiento se basó tanto en los modelos moleculares tridimensionales de Robert Corey, Linus Pauling y Walter Koltun, como a los avances en cristalografía de rayos X, especialmente impulsada por Rosalind Franklin. James Watson realizó en 1951 una estancia en el laboratorio Cavendish en Cambridge, trabajando conjuntamente con Francis Crick (bioquímico inglés, que ejerció como físico militar en la Segunda Guerra Mundial, y que en aquellos momentos realizaba su doctorado en biología).

James Watson y Francis Crick, utilizando los estudios y experimentos en análisis cristalográficos por rayos X realizados por Rosalind Franklin (que de forma resumida es la técnica experimental dedicada al estudio de los patrones de difracción que crean los rayos X al rebotar en los átomos que se disponen de forma regular en un cristal), propusieron en un artículo la estructura del ADN. Este artículo publicado en la prestigiosa revista Nature [7], muy breve para los estándares actuales (con tan solo dos páginas) describe como es la estructura en doble hélice de la molécula de ADN.

Diez años más tarde, en 1962, el premio Nobel de Fisiología y Medicina, fue otorgado a James Watson, Francis Crick y

Maurice Wilkins por sus descubrimientos de la estructura molecular tridimensional del ADN y su relevancia y aportación en la transferencia y almacenamiento de información de los seres vivos.

Rosalind Franklin, que había sido la que realmente había tomado las imágenes de alta calidad de difracción de rayos X en una molécula de ADN, había fallecido cuatro años antes de cáncer y a pesar de su importante contribución, ella no obtuvo este reconocimiento, ya que el Nobel no se concede póstumamente. Como reconocieron los tres premiados, la determinación de la estructura del ADN “habría sido bastante improbable, si no imposible”, de no ser por la fotografía de rayos X que Rosalind Franklin obtuvo.

Francis Crick, unos años más tarde, en 1958, realizó otro importante descubrimiento, conocido como el dogma central de la biología molecular, al analizar y describir los mecanismos e instrucciones que permiten tanto la transmisión, como la expresión genética. En su trabajo inicial, formula que hay una unidireccionalidad en el mecanismo por el cual se expresa la información existente en los genes de nuestras células (que después replanteó en un artículo de Nature publicado en 1970 [8]).

El comienzo de la secuenciación del ADN

En la historia de la secuenciación del ADN existe un investigador de extraordinaria relevancia, Frederick Sanger [9], bioquímico inglés dos veces galardonado con el Premio Nobel de Química.

Su primer Premio Nobel, obtenido en 1958, fue por la obtención de la secuencia de los aminoácidos de la insulina, comprobando la estructura específica de las proteínas. Su segundo Premio Nobel, concedido en 1980, fue debido al método de secuenciación de ADN, conocido como el método enzimático de terminación de cadena, realizado a mediados de los años 70 [10]. Con su técnica se pudo secuenciar el genoma del bacteriófago Phi-X174 (tan solo tiene 11 genes en 5.386 bases) el primer organismo del que se secuenció totalmente el genoma [11].

En esta primera generación de la secuenciación conviene mencionar también el método utilizado en 1977 por Maxam y Gilbert [12], con la inicial ventaja sobre la secuenciación de Frederick Sanger de utilizar el ADN purificado directamente, aunque su complejidad técnica y el elevado coste en los kits

estándar de biología molecular, hizo que el método enzimático fuese el más popular en todos los laboratorios del mundo en los años 80. Este revolucionario método enzimático de terminación de cadena supuso la primera herramienta para el inicio de la secuenciación del ADN, utilizándose en 1984 por científicos del Medical Research Council para obtener los 172.281 pares de bases del virus de Epstein-Barr. ¿Nos parecen muchos pares de bases teniendo en cuenta que unos años antes la secuenciación de unas cinco mil pares de bases había supuesto la obtención de un Premio Nobel?

La velocidad a la que avanza la tecnología en este campo de la secuenciación crece de forma impresionante con los avances científico-técnicos, muchos de ellos gracias al desarrollo de las TIC. La carrera para la secuenciación del Genoma Humano, y por tanto llegar a secuenciar 3.000 millones de pares de bases, comenzó a divisarse en el horizonte.

Segunda generación de secuenciadores (Next Generation Sequencing)

De forma análoga a la evolución en el mundo de la informática, existen también varias generaciones en el campo de la secuenciación del ADN. El desarrollo y avance

tecnológico en el área de los computadores comienza con la primera generación (años 40-56) donde la tecnología predominante eran los tubos de vacío. Continúa la segunda generación (años 56-64) con la utilización de transistores, una tercera generación con los circuitos integrados (1965-71), una cuarta generación con VLSI (años 1971-83) y una quinta generación (1983-1999-presente) siendo protagonista el computador portátil, la computación de altas prestaciones y el paralelismo, la inteligencia artificial y las redes de comunicaciones.

La segunda generación en el campo de la secuenciación de ADN (denominada como secuenciación de próxima generación, NGS) se basa en el principio de la secuenciación masiva y el procesamiento paralelo [13]. Esta tecnología surgió en 1994 y ha estado disponible comercialmente desde 2005 [14], [15].

Una fase inicial y de gran relevancia en las diferentes tecnologías de secuenciación es la preparación de una biblioteca genética de alta calidad (lo que podemos denominar como genoteca). Este concepto involucra la construcción de insertos de DNA genómico o RNA (cDNA específico de tejido), donde se guarda la información de los genes que se

requieran analizar (regiones codificantes y fragmentos que contendrán intrones). Gracias a esta genoteca, nuestra búsqueda de una “aguja (gen) en un pajar (genoma/mRNA)” será más eficiente. Las plataformas tecnológicas más relevantes de esta segunda generación han sido:

- Secuenciador Roche 454. Esta tecnología se considera el primer método comercial disponible de NGS [16], realizada por la empresa americana 454 Life Sciences.
- Illumina (Solexa). En el año 2008 la empresa americana Illumina [17], fundada en 1998, implementó comercialmente su tecnología de secuenciación masiva, conocida como secuenciación de terminador reversible o secuenciación por síntesis (SBS) [18]. En 2014 fue nombrada por el MIT Technology Review como "la compañía más inteligente del mundo". Los sistemas Illumina, con sus plataformas HiSeq y MiSeq, pueden realizar una secuenciación paralela masiva, reduciendo considerablemente el tiempo y el costo de la secuencia. Concretamente, el secuenciador HiSeq es el secuenciador más barato de la segunda generación.
- Ion Torrent: Esta metodología de secuenciación fue lanzada al mercado en febrero de 2010 por el sistema Ion Torrent [19] [20].
- ABI/Solid: SOLiD Sequencing es una tecnología de NGS realizada por la empresa Life Technologies y que estuvo

comercialmente disponible en el año 2006 [21]. Es muy potente ya que tiene la capacidad de generar cientos de millones de lecturas de secuencia al mismo tiempo. Es un método de secuenciación de gran precisión (99.94%).

Tercera generación de secuenciadores

Las tecnologías de secuenciación de segunda generación, siendo Illumina la predominante en el mercado, necesitan un proceso de amplificación a través de PCR. Este paso de amplificación es generalmente costoso y requiere un tiempo considerable. Además, estas tecnologías crean lecturas relativamente cortas que necesitan una fase de ensamblaje compleja posterior. Con el objetivo de resolver estas particularidades, los científicos desarrollaron la secuenciación de tercera generación, que no necesita amplificación por PCR y permiten la lectura de largas cadenas de ADN (varias kilobases) lo que facilita el problema de ensamblaje y la secuenciación de nuevas especies. La característica más relevante es por tanto la secuenciación de molécula única o completa (single molecule sequencing -SMS). La primera tecnología de SMS se desarrolló en el laboratorio de Stephen Quake [22] [23], luego comercializado por Helicos

BioSciences. Las plataformas tecnológicas más relevantes son las desarrolladas por las siguientes empresas:

- Pacific Biosciences [24], que fue la primera compañía en desarrollar el método SMRT (Single Molecule Real Time) [25] para secuenciación de una molécula de ADN completa en tiempo real [26]. Esta metodología tiene como principal ventaja la facilidad en la preparación de la muestra que se requiere analizar, junto con la obtención de longitudes de lectura largas (actualmente de unos 10 kb (kilo pares de bases) y llegando incluso a los 60 kbp). La desventaja de este sistema de secuenciación es su alta tasa de error de alrededor del 13%.
- Oxford Nanopore, fundada en 2005 para desarrollar un sistema de detección disruptivo, electrónico y de una única molécula de ADN (SMRT) basado en nanoporos. El primer producto, MinION, se presentó en 2014 [27] y se puso a disposición comercial en 2015. Es un dispositivo pequeño, que puede conectarse a un simple portátil para realizar la secuenciación en tiempo real (en un dispositivo de coste inferior a 1000 euros, de menos de medio kilo de peso, se pueden realizar secuenciación en tiempo real de hasta 30 Gb). Esta tecnología se ha utilizado para generar secuencias de referencia de genomas bacterianos [28] o trabajar en colaboración con la secuenciación de Illumina [29]. Las ventajas de este método, basado en la secuenciación SMRT,

son bajo costo y secuenciación en tiempo real y su portabilidad al ser un dispositivo de pequeño tamaño y requiriendo como fuente de energía tan solo el puerto USB de un computador. Esta tecnología permite “descentralizar” la secuenciación y realizarla en cualquier lugar, como lo demostraron Joshua Quick y Nicholas Loman, cuando secuenciaron el virus del Ébola en Guinea dos días después de la recolección de la muestra [30]. Esto crea un punto de inflexión en comparación con las tecnologías predecesoras. Su desventaja, es de nuevo su alta tasa de error, que puede llegar al 12 %.

Proyecto del Genoma Humano

Si fue un gran avance en la historia de la ciencia aeroespacial el alunizaje de un hombre en la Luna, su equivalente en biología es el Proyecto del Genoma Humano (PGH [35]), que podría ser uno de los proyectos científicos más relevantes realizado hasta la fecha.

El 26 de junio de 2000, el Primer Ministro del Reino Unido, junto con el presidente de los Estados Unidos, Bill Clinton, anunciaron en una conferencia de prensa la obtención del primer prototipo del Genoma Humano. Se obtuvo por primera

vez la secuencia de los 3000 millones de pares de bases de nuestro ADN, suponiendo un gran éxito para el proyecto lanzado en 1990 por el Departamento de Energía y el Instituto Nacional de la Salud de los Estados Unidos. El investigador responsable y líder del consorcio público, formado por múltiples científicos de diferentes países (EEUU, Reino Unido, Alemania, Japón, Francia y China como principales países), fue Francis Collins. Inicialmente se proyectó que se necesitarían del orden de 15 años para la realización de este desafiante proyecto, teniendo un presupuesto de partida que de unos 3 mil millones de dólares.

En 1998 surgió con fuerza una empresa privada, Celera Genomics, siendo el responsable de la misma Craig J. Venter, biólogo y empresario estadounidense, que vio una gran oportunidad en el ámbito comercial en la secuenciación, impulsando la tecnología denominada "shotgun sequencing". La empresa Celera Genomics utilizó el ADN de cinco personas para obtener la secuencia patrón o genoma de referencia humano, existiendo sospechas de que entre uno de los voluntarios anónimos del proyecto, fue el mismo Craig Venter. Se inició una investigación intensa y competencia frenética en todos los laboratorios relacionados con la secuenciación del genoma humano, puesto que quien primero

obtuviese la secuenciación de un cromosoma, rápidamente lo patentaría.

Además de obtener el valor de los nucleótidos del genoma, también se quería disponer de información adicional, como la localización de la posición de inicio/final de los genes para cada uno de los 23 pares de cromosomas que tiene nuestro ADN. En el primer borrador del Proyecto del Genoma Humano, el número de genes del ser humano oscilaba entre 26.000 y 38.000, lejana a la predicción que hacían los genetistas que suponían que nuestro genoma tendría unos 100.000 genes. A medida que se fue corrigiendo y puliendo la secuencia de ADN del genoma humano, nuestro número de genes codificante se modificó, siendo en la actualidad del orden de 20.000 (de los 3100 millones de pares de bases del ADN, aproximadamente el 2% pertenecería a genoma codificante). Tenemos aproximadamente el mismo número de genes que un gusano y la mitad que el arroz. Afortunadamente para nosotros, el número de genes que tiene una especie no tiene una relación directa con la complejidad de la misma.

Durante el 15 y el 16 de febrero de 2001, Nature (con los datos aportados por el consorcio público liderado por Francis Collins [36]) y Science (con la información de Celera

Genomics [37]), publicaron simultáneamente la secuenciación del Genoma Humano, obteniendo una elevada fiabilidad del 99.9% y adelantando en más de un año el tiempo estimado para la finalización del proyecto.

Es importante mencionar el riguroso mantenimiento y confidencialidad de la información obtenida durante este proyecto, y en general, en todo el ámbito de la bioinformática.

Desde mi punto de vista, las bases de datos que contienen la información surgida del Proyecto Genoma Humano y de otros múltiples proyectos relacionados en el campo de la bioinformática y biomedicina, son un ejemplo de rigurosidad y generosidad. Rigurosidad por la excelente accesibilidad a grandes volúmenes de datos de forma eficiente y ordenada. Generosidad por permitir, que datos y resultados de proyectos que han involucrado miles de millones de dólares puedan ser utilizados y estén disponibles, de forma gratuita, a la comunidad científica.

Con el proyecto del genoma humano termina una primera fase para el conocimiento del ADN, pero todavía queda mucho por conocer y descubrir. Por mi afinidad con la docencia en estructura de computadores, se podría realizar la siguiente

pregunta: ya hemos descubierto una herramienta para desensamblar un programa, pero ¿conocemos bien qué hace y cuál es la función de cada una de las partes del programa (funciones, datos, pila, etc.)? Algo similar podríamos afirmar con el final del proyecto del genoma humano, tenemos información sobre los nucleótidos que componen el ADN, pero la total funcionalidad de nuestros genes (y zonas intergénicas) es un campo apasionante, en especial en el ámbito médico, todavía desconocido y que supone un gran reto científico.

Podemos resumir que con el Proyecto Genoma Humano, se inicia una nueva era en la investigación biomédica y la biología computacional en particular, que afecta transversalmente a múltiples disciplinas, como biología, farmacología, bioquímica y medicina. Es el primer pilar hacia un nuevo reto en salud: la medicina personalizada o de precisión.

Muchos de estos avances anteriormente indicados no se podrían haber realizado sin una plataforma de computación apropiada. El gran volumen de datos (decenas/centenares de Gigabytes por experimento) y la necesidad de procesamiento paralelo son características intrínsecas de la actual biología

computacional. Un pilar básico para el avance de esta tecnología de secuenciación es la sinergia con los métodos, herramientas y procedimientos desarrollados en el ámbito de la Arquitectura y Tecnología de Computadores.

La bioinformática la podríamos definir como un campo multidisciplinar, enfocado a la investigación, aplicación y desarrollo de metodologías y técnicas computacionales para el procesamiento de datos biológicos y médicos, incluyendo la adquisición, almacenamiento, tratamiento, organización, visualización, análisis e interpretación de dichos datos [13]. No solo es necesario una correcta organización y almacenamiento de los datos (como he expresado anteriormente, las bases de datos en bioinformática son excepcionales), es también necesario el procesamiento relacional y deductivo de dichos datos, para la obtención de información, tarea de gran relevancia en el ámbito de los sistemas inteligentes.

Llegados a este punto quisiera enfatizar de nuevo la vertiginosa velocidad a la que evoluciona el mundo de la bioinformática y en particular, la evolución de la secuenciación del ADN.

Por hacer una comparación gráfica, podemos utilizar un índice que mida de forma conjunta la eficiencia de la secuenciación, teniendo en cuenta el número de pares de bases secuenciadas, el número de personas involucradas y el tiempo requerido. Le asignamos a este índice de eficiencia el valor 1 sobre el año 1962, cuando Watson, Crick y Wilkins obtuvieron el premio Nobel de Fisiología y Medicina. Para tener una referencia intuitiva de la evolución relacionemos este valor 1 de eficiencia a una velocidad muy pequeña, como, por ejemplo, un metro por segundo. En 1965 Robert Holley [38] fue capaz de producir la primera secuencia completa de ácido nucleico, la de tRNA de *Saccharomyces cerevisiae* y en 1968 Ray Wu [39] analiza los extremos de las moléculas del ADN del bacteriófago λ , concluyendo que las cadenas terminadas en 5' son 20 nucleótidos más largas que las cadenas terminadas en 3'. Aquí la eficiencia de la secuenciación sería de valor 15 (en tan solo 6 años ya vamos corriendo, pero no a cualquier velocidad, corremos más rápido que el conocido atleta, con ocho medallas olímpicas como velocista, Usain Bolt).

Tomemos la fecha de 1977, cuando Frederick Sanger desarrollaba el método enzimático de terminación de cadena que le condujo a obtener su segundo Premio Nobel. La eficiencia sería de valor 1500. Ahora nuestro símil en

velocidad nos permitiría ir de Granada a Moscú en menos de 45 minutos.

En 1986 fueron decisivos los nuevos secuenciadores didesoxi, como la gama ABI PRISM desarrollado a partir de la investigación de Leroy Hood [40], producido por Applied Biosystems, que permitió la secuenciación simultánea de cientos de muestras, el cual fue utilizado en los inicios del Proyecto Genoma Humano. La eficiencia se dispara a valores de 25.000. En términos comparativos de velocidad, podríamos dar una vuelta al perímetro de la Tierra en unos 26 minutos.

Lynx Therapeutics publicó y comercializó la secuencia de firma paralela masiva (massively parallel signature sequencing MPSS), en 2000. Este método incorporó una tecnología de secuenciación paralela, que sirvió como el primer método de secuenciación de "próxima generación" disponible comercialmente [41]. Cuando en febrero de 2001, las revistas Nature y Science, hicieron públicas la secuenciación del Genoma Humano, la eficiencia tendría un valor de uno 5×10^7 . Comparativamente, tardaríamos unos 15 segundos en ir y volver desde la Tierra a la Luna.

Si hasta ahora el incremento de eficiencia en la secuenciación del ADN (por comparación el de velocidad) es sorprendente, entre el 2007 y el 2009 la famosa Ley de Moore establecida para la tecnología electrónica deja de cumplirse (en especial en el coste económico de la secuenciación) gracias al desarrollo tecnológico de Illumina. El precio por secuenciar un genoma humano pasa de los 100 millones de dólares del año 2001 a los menos de 10.000 dólares en 2011 (en la actualidad estamos por debajo de la barrera de los 1000 dólares). La eficiencia en 2007 se dispara a valores de 1×10^{11} ; lo que equivale a realizar mil viajes de ida y vuelta a la Luna en menos de 8 segundos.

Medicina de precisión: un nuevo enfoque en el tratamiento del cáncer

La medicina de precisión engloba varios campos de la salud, destacando su aplicación contra el cáncer y enfermedades genéticas. Este enfoque precisa de nuevos hallazgos biológicos, la aplicación de tecnologías innovadoras, la revisión exhaustiva de biomarcadores clave de cáncer, la aplicación de metodologías analíticas integradoras y plataformas computacionales avanzadas [42], [43]. Un aspecto clave para acelerar el desarrollo de la medicina de precisión es una caracterización genómica de las muestras de pacientes

más profunda, más rápida y más amplia. El conocimiento exhaustivo del genoma y la secuenciación del transcriptoma facilitan el descubrimiento de nuevas moléculas más fiables para el diagnóstico y evolución de la enfermedad [44]. Este código genético es idéntico al 99,9% en todos los humanos, pero en ese resto de 0,1% radican diferencias esenciales, que repercutirán en el aspecto del individuo, su predisposición a determinadas enfermedades de componente genético, la eficiencia y respuesta a medicamentos, etc. Por tanto, podemos afirmar que ese 0,1% de diferencias en nuestro ADN interviene en que no existan, a nivel médico, dos pacientes iguales.

La medicina genómica, es una potente herramienta que permite implementar en la práctica clínica el modelo sanitario de medicina de precisión y personalizada. Para ello es necesario analizar y conocer las características genéticas y potenciales biomarcadores de cada paciente, con el objetivo de realizar un diagnóstico preciso y un tratamiento eficaz [45]. Esto contrasta con el modelo de medicina poblacional, fundamentada en analizar a un grupo de pacientes y extrapolar los resultados y conclusiones obtenidas al resto de población o posibles afectados, perdiendo las posibles variaciones individuales de cada paciente.

Actualmente, las especialidades de oncología, neurología, cardiología, pediatría y farmacogenética [46], entre otras, tienen en la medicina genómica una potente y eficaz tecnología.

Permítanme contarles el resumen de un artículo, en la prestigiosa revista *The New England Journal of Medicine* del año 2014 [47] que les comento a mis alumnos del master en Ciencia de Datos e Ingeniería de Computadores. El artículo se refiere a cómo las tecnologías de secuenciación ayudan en el diagnóstico de enfermedades, ante posibles contagios de virus o bacterias en nuestro organismo.

En concreto se refiere al caso de un niño de 14 años con inmunodeficiencia combinada severa, el cual se presentó tres veces a un centro médico durante un período de 4 meses con fiebre y dolor de cabeza que progresó a hidrocefalia y estado epiléptico, al que finalmente tuvieron que producirle un coma médicamente inducido. La evaluación diagnóstica, incluida la biopsia cerebral, no fue reveladora. Se procedió a solicitar ayuda al equipo de bioinformática. La secuenciación (mediante NGS) del líquido cefalorraquídeo identificó 475 de las 3.063.784 lecturas de secuencia (0.016%) correspondientes

a la infección por leptospira. Sin embargo, los ensayos clínicos para la leptospirosis realizados fueron negativos. Al final se confió en el procesamiento bioinformático, administrando fármacos antimicrobianos dirigidos y el paciente fue dado de alta 32 días después. La reacción en cadena de la polimerasa (PCR) y las pruebas serológicas en los centros para el control y la prevención de enfermedades de EEUU confirmaron posteriormente la evidencia de infección por *Leptospira santarosai*.

La posibilidad cada vez más cercana de seleccionar tratamientos individualizados para cada paciente en base a la presencia de determinados biomarcadores abre unas expectativas enormes en la mejora del pronóstico de dicho paciente [48]. El eje central de su aplicación gira en torno a la posibilidad de detectar biomarcadores genómicos/transcriptómicos que puedan ser vinculados con respuestas a fármacos [49]. Por tanto, la determinación de nuevos perfiles de biomarcadores de enfermedad que permitan el diagnóstico más preciso de la patología y la posibilidad de que estos grupos de biomarcadores indiquen la quimiosensibilidad o quimirresistencia de un paciente a un tratamiento son un eje esencial de la investigación en la oncología actual [50].

Aparte del enfoque genómico/bioinformático anteriormente expuesto, la medicina personalizada de precisión es un modelo médico que incluye también la caracterización de fenotipos y genotipos de los pacientes (por ejemplo los perfiles moleculares, la información de su expresión genética, las imágenes médicas, las señales médicas de biosensores, los datos sobre estilo de vida, la historia clínica, etc) para diseñar la estrategia terapéutica óptima en un determinado momento, y/o determinar la predisposición hacia la enfermedad y/o desarrollar una prevención adecuada en el momento preciso. Por tanto, la medicina personalizada de precisión no puede ser un sinónimo de medicina genética, ya que, aunque la información genética desempeña un papel decisivo en esta medicina, no es plausible que se puedan adoptar decisiones sobre el tratamiento más adecuado para un paciente sin utilizar toda la información disponible sobre él.

La medicina personalizada aborda diferentes campos de actuación, tales como la determinación de la predisposición a padecer una determinada enfermedad, la detección precoz (asintomática) de enfermedades, el diagnóstico diferencial de enfermedades y su clasificación (en qué fase se encuentra un paciente), el pronóstico de la enfermedad (predicción de la evolución del paciente), la selección de la mejor terapia

(determinación de que tratamientos y en qué cantidad/tiempo funcionaran de forma óptima).

Nuestro grupo de investigación del Departamento de Arquitectura y Tecnología de Computadores, en colaboración y con la ayuda de otros Departamentos, Universidades, Centros de Investigación y Hospitales, se ha centrado en los últimos años en el desarrollo de nuevas metodologías bioinformáticas para la selección de biomarcadores y clasificadores inteligentes, en diferentes tipos de cáncer ([51]-[69]).

La detección precoz de la mayoría de los cánceres es una tarea difícil, pero de gran prioridad en la comunidad científica, ya que una detección temprana precisa incrementa las posibilidades de un paciente de tener un tratamiento exitoso, prolonga su supervivencia y mejora significativamente su calidad de vida.

De la implementación electrónica al ámbito de la medicina personalizada, ¿Cómo ha sido el camino?

Con el ánimo de intentar concretar cómo ha sido el camino, desde mis inicios como investigador, hasta el estado actual,

abordo la evolución de los temas y proyectos de investigación en los que he participado, junto con el grupo de compañeros que continuamente me han estado enseñando, ayudando y colaborando (y no solamente en las facetas puramente científico-técnicas).

Mis comienzos en el campo de la investigación se iniciaron como estudiante de último curso de Físicas en el ámbito del Electromagnetismo. Por aquel entonces, trabajando bajo la supervisión del Prof. Rafael Gómez Martín, estuve involucrado en el diseño y simulación de la propagación electromagnética en antenas, utilizando como lenguaje de programación el FORTRAN 77. Al terminar mis estudios de Física en junio de 1992, el Prof. Alberto Prieto, en aquel momento en el seno del Departamento de Electrónica, me ofreció la posibilidad de unirme a su Grupo de investigación y solicitar una Beca. El Prof. Prieto, mi mentor, compañero y amigo Alberto, me ha guiado desde aquel momento en esta apasionante aventura de la investigación, docencia y gestión en la Universidad de Granada. Muchas gracias, muchísimas gracias por haberte fijado en ese estudiante para que formarse parte de tu grupo de investigación.

El camino universitario de la mano del Prof. Prieto y los compañeros del actual Departamento de Arquitectura y Tecnología de Computadores ha sido siempre ilusionante y lleno de retos en el ámbito científico. Pero esto no se puede lograr sin lo más importante, el apoyo personal, generosidad, cariño y afecto que he recibido durante mi formación hasta la actualidad, por mis compañeros.

El punto de partida de la senda investigadora es el desarrollo de la tesis doctoral. Gracias a mis tutores el Prof. Julio Ortega y el Prof. Francisco Pelayo, comencé a caminar en el ámbito del desarrollo e implementación electrónica de sistemas difusos (título de la tesis). En aquel tiempo comenzaba a fraguarse una estrecha cooperación entre el Grupo de investigación del Prof. Prieto de la Universidad de Granada, con, el del Prof. Karl Goser del Departamento de Electrónica de la Universidad de Dortmund (Alemania). Eran las conocidas Acciones Integradas, y cuando se solicitaron en nuestro Grupo de Investigación voluntarios para cooperar en dichas acciones, no dudé en levantar la mano. Parte de mis estudios pre-doctorales los desarrollé en las tres estancias que realicé en la Universidad de Dortmund.

Allí pude realizar un aprendizaje intenso en el campo de la implementación electrónica y realización de dispositivos analógicos, ya que en la Universidad de Dortmund tenían su propia sala para la fabricación de dispositivos y todo el instrumental necesario para realizar el test posterior.

Agradecer al Prof. Karl Goser, además de su apoyo como científico en las tareas de investigación, su afectividad y amabilidad en todas las estancias que realice en su grupo de investigación. Su brillante trayectoria científica y su relación con la Universidad de Granada le condujeron a ser nombrado Doctor Honoris Causa de la Universidad de Granada en 2006. Destacar que el Prof. Karl Goser fue director de departamento en la división de circuitos integrados de Siemens AG, de Munich. Durante su dirección, fue el responsable de la fabricación de diferentes circuitos integrados comerciales, obteniendo un gran número de patentes. Gracias a esta colaboración, tuve el privilegio de co-dirigir junto al Prof. Karl Goser dos tesis doctorales, en el seno de la empresa Siemens AG, de Munich. Además de los conocimientos que me aportó esta cooperación, la experiencia más destacable fue conocer de forma activa la sinergia y reciprocidad entre el mundo universitario y su aplicación práctica en el campo empresarial,

algo que en las universidades alemanas tiene una gran implicación y dilatada experiencia.

Mis estancias en Dortmund, acompañado por mis compañeros el Prof. Julio Ortega y el Prof. Eduardo Ros me permitieron avanzar en mis conocimientos de electrónica, pero sobre todo y más relevante, me permitieron establecer amistad con miembros de la Universidad de Dortmund, y en cierto sentido, comenzar a “germanizarme”. En especial destacar al Prof. Peter Gloesekoetter, actualmente en la Universidad de Münster, compañero y gran amigo que tanto me ha ayudado y tantas experiencias positivas hemos vivido juntos en Münster.

El profesor universitario es un gran privilegiado, al poder involucrarse en temas de investigación, que más que suponer un trabajo, suponen una fascinación y a veces una seducción, disfrutando de aquello que se va aprendiendo. Además, se goza de la excepcional característica de evolucionar y modificar el campo de investigación que se analiza, lo que me permitió, en mi caso, desarrollar mi tema de investigación hacia el mundo de sistemas inteligentes, cambiando ya a mitad de la realización de mi tesis doctoral del hardware al software.

Mis comienzos fueron en el campo de la lógica difusa y tuve la fortuna de formarme en la Universidad de Granada, una de las universidades a nivel internacional más relevantes en este ámbito. En efecto, la Universidad de Granada, y en particular E.T.S. Ingenierías Informática y de Telecomunicación, tenía (y por supuesto tiene en la actualidad) una destacada presencia internacional en el ámbito de la lógica difusa y en sistemas inteligentes en general. Aunque hay muchos índices que avalan el prestigio de nuestra institución en ranking internacionales, destacar la edición de 2020 del Academic Ranking of World Universities (ARWU) que sitúa a la Universidad de Granada en la primera posición de España, en el área de Ciencias e Ingeniería de la Computación.

En cierta medida, esta posición de excelencia internacional posibilitó el fructífero avance en el campo de sistemas inteligentes y otra de las estancias internacionales decisivas en mi formación en 1998. En ese curso académico realicé una estancia de tres meses en el Departamento de Ingeniería Electrónica y Ciencias de la Computación, en la Universidad de Berkeley, bajo la dirección del Prof. L. Zadeh, Doctor Honoris Causa de la Universidad de Granada (1996).

Para los que hemos trabajado en el campo de la lógica difusa, el Prof. L.Zadeh es bien conocido y de una excepcional relevancia internacional en este campo. Fue pionero al introducir en el año 1965 la teoría de conjuntos difusos y lógica difusa [70]. Nació en la antigua República Soviética de Azerbaiyán, estudió en la Universidad de Teherán, y después continuó sus estudios en el MIT, en la Universidad de Columbia y finalmente en la Universidad de Berkeley. Por su excepcional y brillante carrera como investigador en el campo de la lógica difusa y la teoría de la posibilidad, el Prof. L.Zadeh ha recibido un gran número de premios, distinciones y doctorados honoris causa de varias instituciones del mundo. Mencionar la Medalla Richard W. Hamming IEEE en 1992 por sus logros sobresalientes en ciencias de la información, sistemas de información y tecnología de la información. Destacar también su apego a diferentes universidades españolas, y a la de Granada en particular, recibiendo doctorados honoris causa, además de por nuestra Universidad, por las de Oviedo (1995) y la Politécnica de Madrid (2007). De nuevo tuve la fortuna de colaborar en el seno de un grupo de investigación liderado por un gran científico y una persona de gran talla.

La investigación realizada en el campo de la lógica difusa fue aplicada tanto al campo de la aproximación funcional, como al área del control. Quisiera mencionar mi primera tesis doctoral dirigida, realizada por el Prof. Héctor Pomares. Ser tutor en una tesis es un honor, pero también una excelente oportunidad de aprender de la labor y trabajo que realiza el doctorando. Dicha tesis se centró en la construcción de una nueva metodología para la caracterización y diseño automático de la estructura y componentes de un sistema difuso, utilizando la información proporcionada por un conjunto de datos de entrada/salida. Adicionalmente, se desarrollaron algoritmos para el diseño automático de controladores difusos.

Como mencioné anteriormente, tuve la fortuna de aprender mucho (y lo continúo haciendo) de mi compañero y amigo Héctor. A esta tesis siguieron otras que involucraban el campo de las redes neuronales, sistemas expertos, sistemas inteligentes en plataformas de cómputo paralelas y sistemas de reconocimiento auto-configurables, entre otros. De nuevo, tuve el privilegio de instruirme y entender nuevas metodologías bajo la realización de las tesis doctorales de mis compañeros del Departamento, Prof. Jesús González, Prof. Luis Javier Herrera, Prof. Alberto Guillén y Prof. Oresti Baños.

El acercamiento al campo de la biomedicina y la bioinformática comenzó en el 2006, cuando el Prof. Alberto Prieto realizó una reunión conjunta con el grupo de investigación liderado por la Prof. Antonia Aránega, del Departamento de Anatomía y Embriología Humana, de la Universidad de Granada. La primera colaboración fue en la utilización de sistemas inteligentes para el desarrollo de un modelo bioinformático como plataforma predictiva del comportamiento y trazabilidad de células madre, que pudiese ser utilizado con éxito en la diferenciación miocárdica en base a factores de inducción exógenos.

Recuerdo que la Prof. Antonia Aránega con su forma directa de comunicarse, siempre amable, pero sin rodeos, nos comentó su dificultad para poder analizar un fichero Excel, donde tenían unas cuarenta columnas (correspondientes a pacientes enfermos y controles relacionados con el cáncer de páncreas) y unas 30.000 mil filas (correspondientes a los valores de expresión de los genes de dichos pacientes, además de genes de control). Sus palabras motivadoras: “me estoy volviendo loca, tengo más de un millón de datos. Esto lo tenéis que resolver los informáticos” dio pie a que nuestro grupo de

investigación comenzase con el fascinante problema de la expresión diferencial de genes.

Por aquellos tiempos (año 2009) la tecnología que empezamos a utilizar fue la de los microarrays [71], y el problema que nos planteamos era el cáncer de páncreas [56]. De forma resumida, los datos de microarrays se utilizan para medir los niveles de expresión de múltiples genes simultáneamente. La base fundamental de los microarrays de ADN es el proceso de hibridación. Dos cadenas de ADN se hibridan si son complementarias entre sí. La complementariedad se refleja según regla de Watson-Crick de que la adenina (A) se une a la timina (T) y la citosina (C) se une a la guanina (G).

El material hibridado es excitado por un láser y es detectado por un escáner de luz en la superficie del microarrays (que también es denominado chip de ADN). Debido a que se conoce la ubicación de cada sonda de oligonucleótidos, se puede cuantificar la cantidad de muestra hibridada a partir de la imagen generada por la exploración y determinar el valor de expresión de los genes analizados. Estos fueron nuestros inicios, utilizando microarrays de Affimetrix para la patología del cáncer de páncreas.

La investigación en el adenocarcinoma ductal pancreático (PDAC) [54] contó con la colaboración de Inst. Biopatología y Medicina Regenerativa (IBIMER), el departamento Bioquímica y Biología Molecular de la Universidad de Granada, así como la sección de oncología del Hospital Universitario Virgen de la Victoria de Málaga, del Hospital Virgen de las Nieves de Granada, del Hospital Universitario Ramon y Cajal de Madrid y del Hospital Universitario Reina Sofía de Córdoba.

Indicar que debido a la agresividad de esta enfermedad, la detección rápida de PDAC parece ser la estrategia más relevante para mejorar la supervivencia a largo plazo de los pacientes. Nuestro objetivo inicial fue identificar nuevos marcadores en sangre periférica que diferencien entre pacientes enfermos y controles sanos.

Los estudios de microarrays identificaron 87 genes expresados diferencialmente en muestras de sangre periférica de pacientes con PDAC. Cuatro de estos genes fueron seleccionados para su análisis por RT-qPCR (PCR cuantitativa en tiempo real), que confirmó los cambios observados previamente. En nuestro estudio de validación, la combinación de CLEC4D e IRAK3

predijo el diagnóstico de PDAC con una precisión del 93%, con una sensibilidad del 86% y una especificidad del 100%.

Estos primeros resultados en el campo de la bioinformática dieron paso a la consecución de proyectos de investigación conjunta entre miembros de nuestro departamento y de la Facultad de Medicina, la realización de tesis doctorales en este apasionante mundo de la bioinformática y la obtención conjunta de patentes. Agradecer en este punto la colaboración del Prof. José Carlos Prados Salazar y Prof. Octavio Caba Pérez.

Nuestro grupo de investigación de bioinformática se ha centrado en los últimos años en la hibridación de fuentes de información para la obtención de una huella génica robusta, desarrollando herramientas que han sido puestas a disposición de la comunidad científica en el conocido repositorio de Bioconductor [59]. Dado que el cáncer se considera fundamentalmente una enfermedad de los genes, todos los esfuerzos han sido destinados a extraer conocimiento desde principalmente dos fuentes de información a nivel transcriptómico y genómico: la expresión de gen y el número de copias de gen.

Adicionalmente a la investigación en el adenocarcinoma ductal pancreático, las patologías fructíferamente abordadas por nuestro grupo, que han permitido la defensa de varias tesis doctorales y publicación de relevantes artículos en revistas internacionales, han sido la leucemia [51], cáncer de piel [60], cáncer de mama [55] y cáncer de pulmón [59]. Por mencionar un ejemplo, en leucemia, que es un cáncer con génesis en las células primitivas productoras de sangre, se obtuvieron sistemas de clasificación inteligente con una precisión del 96.40% para el problema multi-clase (tanto leucemia aguda y crónica, en sus variantes mieloide y linfoblástica) con 10 genes expresados diferencialmente, utilizando siempre muestras provenientes de células de médula ósea. Se integraron conjuntos de datos de diferentes tecnologías y plataformas (Microarray y RNA-seq) con el propósito de realizar la identificación de una huella génica precisa. En contraposición a lo ocurrido con el trabajo inicialmente mencionado del cáncer de páncreas, los datos utilizados proceden de bases de datos públicas, en especial de GDC [72].

GDC (Genomic Data Commons) es una plataforma que en la actualidad aglutina más de 65 proyectos, de 29 tipos de cáncer diferentes, con más de 84.000 pacientes y con un total de unos 570.000 archivos. Es un portal de referencia en el ámbito de la

bioinformática, financiado por el NIH (National Institutes of Health and National Cancer Institute, EEUU), mencionando con orgullo, la colaboración en esta plataforma de uno de nuestros becarios de investigación, el Dr. Francisco Ortuño, que actualmente pertenece al Cancer Genomics Researcher en el Clinical Bioinformatics Area de la Fundación Progreso y Salud, ubicada en Sevilla en el Hospital Universitario Virgen del Rocío. En dicha unidad de bioinformática existen como miembros varios doctorandos que se han formado en nuestro grupo.

Durante la redacción del presente Discurso, hemos recibido la noticia de la aceptación de un nuevo proyecto titulado: “Desarrollo de una plataforma inteligente que permita la integración de fuentes de información heterogénea (imágenes, genética y proteómica) para la caracterización y predicción de la virulencia y patogenicidad de pacientes con COVID-19”, dentro de la Convocatoria de subvenciones a Proyectos de investigación sobre Sars-Cov-2 y la enfermedad Covid-19, de la Junta de Andalucía. Debido al desfase que ha existido entre la redacción inicial de este documento y su actual lectura, recientemente ha sido concedido el proyecto titulado: “Integración de fuentes heterogéneas de información biomédica utilizando computación de altas prestaciones.

Aplicación en problemas complejos en medicina personalizada y de precisión”, afín a la temática del presente discurso.

En la situación actual, resulta de una especial relevancia, ante la gravedad y rápida evolución de la COVID-19, ser capaces de abordar un diagnóstico, tratamiento y pronóstico rápidos ante un nuevo enfermo de COVID-19. La metodología utilizada debe ser adaptable y flexible según la información obtenida sobre el paciente. La finalidad es utilizar, analizar e integrar todas las fuentes de información disponibles para un paciente, empleando como soporte sistemas de inteligencia artificial. El mencionar este proyecto me obliga a agradecer a todos los integrantes del mismo el esfuerzo realizado, ya que las herramientas y metodologías desarrolladas en patologías del cáncer (en especial tratamiento de imágenes médicas, genómica y proteómica), se pretenden trasladar y aplicar al problema del COVID-19.

Estamos ante el inicio de una gran revolución en el ámbito clínico, debido a que en la actualidad existen cada vez más datos biomédicos/bioinformáticos, máquinas capaces de generar enormes volúmenes de dichos datos, incluyendo datos genómicos, a un coste realmente asequible. Cada vez se genera más cantidad de datos, pero ¿Cuál es el porcentaje de los datos

que realmente se procesan y analizan? Es decir, ¿se extrae todo el jugo, toda la información posible a los datos que se generan?

El uso de nuevos algoritmos o metodologías que permitan la transformación de mediciones genómicas en mediciones de funcionalidad celular, que analicen los mecanismos de la enfermedad y los mecanismos de la acción farmacológica, junto con la integración de múltiples fuentes de información biomédica, permitirá en última instancia la transición real de la medicina empírica actual a la medicina de precisión y proporcionará una medicina cada vez más personalizada. Citando a Voltaire, no hay problema complejo cuya solución resista el ejercicio continuo del pensamiento.

La extracción de la información de los diferentes y heterogéneos tipos de datos que disponemos es un reto en la actualidad. Pero este Big Data Clinic nos abre una ventana hacia el futuro del descubrimiento de nuevos tratamientos más precisos para una determinada enfermedad y para un paciente en particular, además de nuevas formas de diagnóstico más exactas.

He dejado sin mencionar las implicaciones legales, éticas y morales que conllevan parte de esta nueva disciplina de la medicina personalizada y de precisión, y no lo hago, porque citando de nuevo a Voltaire, el secreto de no aburrir es no tratar de decirlo todo.

Agradecimientos

Quisiera terminar este discurso con mi reconocimiento a todas las personas que me ha apoyado a lo largo de este camino. Aunque el agradecimiento es un acto personal e íntimo, y con el riesgo de dejar sin mencionar muchos de los apoyos recibidos (algunos nombres los he mencionado en las secciones anteriores), quisiera empezar por mi profesor, mi director, mi compañero y amigo, el Prof. Alberto Prieto. Sus méritos como profesor y científico son indiscutibles y bien conocidos. Pero el profesor y científico antes que nada es persona, y la generosidad personal de Alberto se ha caracterizado por guiar, proteger y ayudar a las personas que le rodean y especial a los miembros de su Departamento. Me dio la oportunidad de empezar a dar mis primeros pasos en la Universidad de Granada, aconsejándome en cada una de las etapas de este largo y a la vez, ilusionante camino.

A los profesores Julio Ortega y Francisco Pelayo, que fueron mis directores de tesis y me ofrecieron el primer tema de investigación en el ámbito de la implementación electrónica de sistemas difusos. Agradecerles su generosidad, su tiempo y los conocimientos transferidos. Mencionar la libertad y confianza que depositaron en ese joven doctorando, para que pudiese descubrir y explorar la senda que estimase más conveniente en cada momento (aun a riesgo de equivocarme), pero siempre sintiendo sus apoyos.

Gracias a todos los investigadores que me dieron la oportunidad, distinción y confianza de ser su director de tesis, siendo algunos de ellos actualmente profesores en mi departamento (Héctor, Jesús, Luis Javier, Alberto y Oresti). Hemos disfrutando, como compañeros y amigos, los momentos ilusionantes en los que el éxito nos sonreía y también en los que las dificultades se vislumbraban como insalvables. Gracias a mis veinticinco doctorandos por compartir y enriquecerme con vuestros conocimientos. Es muy difícil trabajar en el ámbito de la investigación de forma individual, y como dice un proverbio africano: “Si quieres ir rápido ve solo. Si quieres llegar lejos ve acompañado”. He tenido la fortuna de ir muy bien acompañado.

Al resto de compañeros del Departamento de Arquitectura y Tecnología de Computadores, agradecerles su apoyo imprescindible en muchas tareas, tanto docentes (muchas gracias Antonio, Javier y Gustavo por tener siempre actualizada la asignatura que compartimos) como investigadoras.

Gracias a todos mis alumnos por enseñarme a enseñar, y a todos los que, en estos veinte y siete años de camino, he compartido un tiempo de docencia, investigación y gestión universitaria. De todos ellos he aprendido mucho, deseando seguir haciéndolo por algún tiempo más. Respecto a la gestión en esta institución, agradecer al Prof. Buenaventura Clares por otorgarme la confianza para caminar a su lado durante ocho años como subdirector de mi querida E.T.S. Ingenierías Informática y de Telecomunicación. A mis subdirectores del Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC-UGR), el Prof. Alberto Palma y el Prof. Andrés Cano, a la Secretaria del Centro, Prof. Nuria Medina y personal de administración, Bea, Paco y Carolina, gracias por vuestra generosidad, tiempo y empeño en la gestión de este Centro.

Gracias a la institución que ha albergado mi camino académico, la Universidad de Granada, que ha sido, es y será para mí mucho más que mi lugar de trabajo.

Y dejando para el final los agradecimientos más importantes, quisiera enfatizar que todo mi recorrido personal y profesional no hubiese sido posible sin la confianza, apoyo y afecto de mi familia y amigos. Mi mujer, Olga, fuente constante de ánimo, cariño y energía, junto a mis hijos Ignacio, Laura y Olga, por hacerme disfrutar con vuestra sonrisa de una de las mayores satisfacciones de la vida: la paternidad. A mis padres, Carmen y Torcuato, y mis hermanos, Javier, Pablo y Fernando, por vuestro esfuerzo, afecto y entusiasmo. No había en mi familia ninguna tradición universitaria, aunque sí docente, y hoy ya somos varios los que compartimos la satisfacción de pertenecer a la Universidad de Granada.

Mi más profundo agradecimiento a todos por acompañarme en este acto.

Bibliografía

[1] S.K. Singh, S. Rathore y J.H. Park, “BlockIoTIntelligence: A Blockchain-enabled Intelligent IoT

Architecture with Artificial Intelligence”, Future Generation Computer Systems, vol. 110, pp. 721-743, 2020.

[2] C.A. Hardy y S.P. Williams, “Managing Information Risks and Protecting Information Assets in a Web 2.0 era”, in 23rd Bled eConference eTrust: Implications for the Individual, Enterprises and Society, Bled, Slovenia, 2010.

[3] I.A. Hashem, I. Yaqoob, N.B. Anuar et ál., “The rise of "big data" on cloud computing: Review and open research issues”, Information Systems, vol. 47, pp. 98-115, 2015.

[4] D. Boyd y K. Crawford, “Critical questions for big data”, Information Communication & Society, vol. 15, no. 5, pp. 662-679, 2012.

[5] X. Zhu, S. Shang y M. Kim, “Privacy-preserving representation learning for big data”, Neurocomputing, vol. 406, pp. 293-294, 2020.

[6] A. Prieto, “¿Qué es la Ciencia de Datos, Minería de Datos y Big Data?”, YouTube, 2018.

[7] J.D. Watson y F. Crick, “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”, Nature, vol. 171, pp. 737-738, 1953.

[8] F. Crick, “Central Dogma of Molecular Biology”, Nature, vol. 227, pp. 561-563, 1970.

[9] F. Sanger, “F. Sanger, Biographical”, [Online]. Available: <http://www.nobelprize.org/>.

- [10] F. Sanger, S. Nicklen y A.R. Coulson, “DNA sequencing with chain-terminating inhibitors”, *Proc Natl Acad Sci USA*, vol. 74, no. 12, pp. 5463-7, 1977.
- [11] F. Sanger, G.M. Air, B.G. Barrell et ál., “Nucleotide sequence of bacteriophage phi X174 DNA”, *Nature*, vol. 265, p. 687–695, 1977.
- [12] A.M. Maxam y W. Gilbert, “A new method for sequencing DNA”, *Proc Natl Acad Sci USA*, vol. 74, no. 2, pp. 560-4, 1977.
- [13] D.A. Wheeler, M. Srinivasan, M. Egholm et ál., “The complete genome of an individual by massively parallel DNA sequencing”, *Nature*, vol. 452, p. 872–876, 2008.
- [14] H. Kambara, T. Nishikawa, Y. Katayama y T. Yamaguchi, “Optimization of parameters in a DNA sequenator using fluorescence detection”, *Bio-technology*, vol. 6, p. 816–820, 1988.
- [15] T. Hunkapiller, R. Kaiser, B. Koop y L. Hood, “Large-scale and automated DNA sequence determination”, *Science*, vol. 254, p. 59–67, 1991.
- [16] T. Holtegaard, D. Oglesbee, S. Middha et al., “Evaluation of Two Next Generation Sequencing Platforms, Illumina Genome Analyzer and Roche 454, for Sequencing the Entire Mitochondrial Genome”, *Journal of Molecular Diagnostics*, vol. 11, no. 6, p. 617, 2009.

- [17] Illumina.[Online]. Available: <https://www.illumina.com/>
- [18] H. Li y R. Durbin, “Fast and accurate short read alignment with Burrows-Wheeler transform”, *Bioinformatics*, vol. 25, no. 14, pp. 1754-1760, 2009.
- [19] M.A. Quail, M. Smith, P. Coupland et ál., “A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers”, *BMC Genomics*, vol. 13, no. 341, 2012.
- [20] N. Loman, R.V. Misra, T.J. Dallman et ál, “Performance comparison of benchtop high-throughput sequencing platforms”, *Nature Biotechnology*, vol. 30, no. 5, p. 434, 2012.
- [21] B. Lin, J. Wang y Y. Cheng, “Recent Patents and Advances in the Next-Generation Sequencing Technologies”, *Recent patents on biomedical engineering*, vol. 1, pp. 60-67, 2008.
- [22] I. Braslavsky, B. Hebert, E. Kartalov y S.R. Quake, “Sequence information can be obtained from single DNA molecules”, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, p. 3960–3964, 2003.
- [23] T.D. Harris, P.R. Buzby, H. Babcock et ál., “Single-molecule DNA sequencing of a viral genome”, *Science*, vol. 320, p. 106–109, 2008.
- [24] Pacific Biosciences. [Online]. Available: <https://www.pacb.com/>.

- [25] A. Rhoads y K.F. Au, “PacBio Sequencing and Its Applications”, *Genomics Proteomics & Bioinformatics*, vol. 13, no. 5, pp. 278-289, 2015.
- [26] M. Metzker, “Sequencing in real time”, *Nature Biotechnology*, pp. 150-151, 2009.
- [27] C. Hayden, “Data from pocket-sized genome sequencer unveiled”, *Nature*, 2014.
- [28] N.J. Loman, J. Quick y J.T. Simpson, “A complete bacterial genome assembled de novo”, *Nature Methods*, p. 11–14, 2015.
- [29] M.A. Madoui, S. Engelen, C. Cruaud et ál., “Genome assembly using nanopore-guided long and error-free DNA reads”, *BMC Genomics*, pp. 1-11, 2015.
- [30] E.C. Hayden, “Pint-sized DNA sequencer impresses first users,” *Nature*, vol. 521, pp. 15-16, 2015.
- [31] T. Miyake, T. Tani, H. Sonobe et ál., “Real-time imaging of single-molecule fluorescence with a zero-mode waveguide for the analysis of protein-protein interaction”, *Analytical Chemistry*, vol. 80, no. 15, pp. 6018-6022, 2008.
- [32] R. M. Leggett y M.D. Clark, “A world of opportunities with nanopore sequencing”, *Journal of Experimental Botany*, vol. 68, no. 20, pp. 5419-5429, 2017.

- [33] C. Belser, B. Istace, E. Denis et ál., “Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps”, *Nature Plants*, vol. 4, no. 11, p. 879, 2018.
- [34] OxfordNanopore, “<https://nanoporetech.com>,” [Online].
- [35] “The Human Genome Project,” [Online]. Available: <https://www.genome.gov/human-genome-project>.
- [36] International Human Genome Sequencing Consortium, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, pp. 860-921, 2001.
- [37] J.C. Venter, M.D. Adams, E.W. Myers et al., “The Sequence of the Human Genome”, *Science*, vol. 291, no. 5507, pp. 1304-1351, 2001.
- [38] R.W. Holley, J. Apgar, G. Everett et ál., “Structure of a ribonucleic acid”, *Science*, vol. 147, p. 1462–1465, 1965.
- [39] R. Wu y A.D. Kaiser, “Structure and base sequence in the cohesive ends of bacteriophage lambda DNA”, *Journal of Molecular Biology*, vol. 35, no. 3, pp. 523-537, 1968.
- [40] L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes et ál., “Fluorescence detection in automated DNA sequence analysis”, *Nature*, vol. 321, p. 674–679, 1986.
- [41] S. Brenner, M. Johnson, J. Bridgham et ál., “Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays”, *Nature Biotechnology*, vol. 18, no. 6, p. 630–634, 2000.

- [42] A. Re, C. Nardella, A. Quattrone y A. Lunardi, “Precision Medicine in Oncology”, *Frontiers in Oncology* , vol. 8, 2018.
- [43] R.J. Arnall, R. Petro, J.N. Patel y L. Kennedy, “A clinical pharmacy pilot within a Precision Medicine Program for cancer patients and review of related pharmacist clinical practice”, *Journal of Oncology Pharmacy Practice*, vol. 25, pp. 179-186, 2019.
- [44] G. Curigliano, “Cancer Evolution as the New Frontier of Precision Medicine”, *Handbook of Experimental Pharmacology*, vol. 249, pp. 289-297, 2018.
- [45] M.A. Hamburg y F.S. Collins, “The Path to Personalized Medicine,” *New England Journal of Medicine*, vol. 363, no. 4, pp. 301-304, 2010.
- [46] O. Strianese, F. Rizzo, M. Ciccarelli et ál., “Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease”, *Genes*, vol. 11, no. 7, 2020.
- [47] M.R. Wilson, S.N. Naccache, E. Samayoa et ál., “Actionable diagnosis of neuroleptospirosis by next-generation sequencing,” *The New England Journal of Medicine*, vol. 370, no. 25, pp. 2408-2417, 2014.

- [48] A. Heider y R. Alt, “virtualArray: a R/bioconductor package to merge raw data from different microarray platforms”, *BMC Bioinformatics.*, pp. 14-75, 2013.
- [49] M. Shibata y M.O. Hoque, “Development of biomarkers for real precision medicine”, *Translational Lung Cancer Research*, vol. 7, no. 3, pp. S228-S231, 2018.
- [50] S. Mondal, D. Roy, S.S. Bhattacharya, L. Jin et al., “Therapeutic targeting of PFKFB3 with a novel glycolytic inhibitor PFK158 promotes lipophagy and chemosensitivity in gynecologic cancers”, *International Journal of Cancer*, vol. 144, no. 1, pp. 178-189, 2019.
- [51] D. Castillo, J.M. Galvez, L.J. Herrera, F. Rojas, O. Valenzuela, O. Caba, J. Prados e I. Rojas, “Leukemia multiclass assessment and classification from Microarray and RNA-seq technologies integration at gene expression level,” *Plos One*, vol. 14, no. 2, 2019.
- [52] O. Valenzuela, F. Rojas, I. Rojas y P. Glosekotter, “Main findings and advances in bioinformatics and biomedical engineering- IWBBIO 2018”, *BMC Bioinformatics*, vol. 21, no. 7, 2020.
- [53] J. M. Galvez, D. Castillo, L.J. Herrera, B.S. Román, O. Valenzuela, F. M. Ortuño e I. Rojas, “Multiclass classification for skin cancer profiling based on the integration of

heterogeneous gene expression series”, Plos One, vol. 13, no. 5, 2018.

[54] A. Irigoyen, C. Jimenez-Luna, M. Benavides, O. Caba, J. Gallego, F.M. Ortuño, C. Guillen-Ponce, I.Rojas, R. Aranda C. Torres y J. Prados, “Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers”, Plos One, vol. 13, no. 4, 2018.

[55] D. Castillo, J.M. Galvez, L.J. Herrera, B. S. Román, F. Rojas e I. Rojas, “Integration of RNA-Seq data with heterogeneous microarray data for breast cancer profiling”, BMC Bioinformatics, vol. 18, 2017.

[56] O. Caba, A. Irigoyen, C. Jimenez-Luna, M. Benavides, F.M. Ortuño, J. Gallego, I.Rojas, C. Guillen-Ponce, C. Torres, E. Aranda y J. Prados, “Identification of gene expression profiling associated with erlotinib-related skin toxicity in pancreatic adenocarcinoma patients”, Toxicology and Applied Pharmacology, vol. 311, pp. 113-116, 2016.

[57] C. Torres, A. Linares, M.J. Alejandre, R.J. Palomino-Morales, O. Caba, J. Prados, A. Aránega, J. R. Delgado, A. Irigoyen, J. Martínez-Galán, F. M. Ortuño, I. Rojas y S. Perales, “Prognosis Relevance of Serum Cytokines in Pancreatic Cancer”, Biomed Research International, vol. 2015, 2015.

- [58] C. Torres, S. Perales, M.J. Alejandre, J. Iglesias, R.J. Palomino, M. Martin, O. Caba, J.C. Prados, A. Aránega, J.R. Delgado, A. Irigoyen, F. M. Ortuño, I.Rojas y A. Linares, “Serum Cytokine Profile in Patients With Pancreatic Cancer,” *Pancreas*, vol. 43, no. 7, pp. 1042-1049, 2014.
- [59] D. Castillo-Secilla, J.M. Galvez, F. Carrillo-Perez, M. Verona-Almeida, F.M. Ortuño, L.J. Herrera e I. Rojas, “KnowSeq R-Bioc package: The automatic smart gene expression tool for retrieving relevant biological knowledge”, *Computers in Biology and Medicine*, vol.133, 2021. Available:<https://bioconductor.org/packages/devel/bioc/html/KnowSeq.html>.
- [60] J. M. Gálvez, D. Castillo-Secilla, L. J. Herrera, O. Valenzuela, O. Caba, J. C. Prados, F.M. Ortuño e I. Rojas, “Towards Improving Skin Cancer Diagnosis by Integrating Microarray and RNA-Seq Datasets”, *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 2119-2130, 2020.
- [61] O. Caba, A. Irigoyen, C. Jimenez-Luna, M. Benavides, F.M. Ortuño, C. Guillen-Ponce, I. Rojas, E.A. Aguilar y J.C. Prados, “Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers”, *Annals of Oncology*, vol.28, 2017.

- [62] D. Castillo, J.M. Galvez, L.J. Herrera, I. Rojas, “Breast Cancer Microarray and RNASeq Data Integration Applied to Classification”, *Advances In Computational Intelligence, IWANN 2017*, vol.10305, 2017.
- [63] F.M. Ortuño, O. Valenzuela e I. Rojas, “Advances and New Perspectives in Medicinal Chemistry Engineering and Bioinformatics (from IWBBIO 2015)”, *Current Topics In Medicinal Chemistry*, vol.17, 2017.
- [64] I. Rojas y F. Ortuño, “Recent Advances in Bioinformatics and Biomedical Engineering (Selected Articles from IWBBIO 2014)”, *Current Bioinformatics*, vol.11, 2016.
- [65] F.M. Ortuño, O. Valenzuela, B. Prieto, M.J. Saez-Lara, C. Torres, H. Pomares e I. Rojas, “Comparing different machine learning and mathematical regression models to evaluate multiple sequence alignments”, *Neurocomputing*, vol.164, 2015.
- [66] F.M. Ortuño, C. Torres, P. Glosekotter e I. Rojas, “New trends in biomedical engineering and bioinformatics applied to biomedicine - special issue of IWBBIO 2014”, *Biomedical Engineering Online*, vol.14, 2015.
- [67] O. Caba, J. Prados, R. Ortiz, C. Jimenez-Luna, C. Melguizo, P.J. Alvarez, J.R. Delgado, A. Irigoyen, I. Rojas, J. Perez-Florido, C. Torres, S. Perales, A. Linares y A. Aránega, “Transcriptional Profiling of Peripheral Blood in Pancreatic

Adenocarcinoma Patients Identifies Diagnostic Biomarkers”, Digestive Diseases and Sciences, vol.59, 2014.

[68] F.M. Ortuño e I. Rojas, “Advances in bioinformatics and biomedical engineering - special issue of IWBBIO 2013”, Theoretical Biology and Medical Modelling, vol.11, 2014.

[69] F.M. Ortuño, O. Valenzuela, H. Pomares, F. Rojas, J.P. Florido, J.M. Urquiza e I. Rojas, “Predicting the accuracy of multiple sequence alignment algorithms by using computational intelligent techniques”, Nucleic Acids Research, vol.41, 2013.

[70] L.A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338-353, 1965.

[71] Y.H. Yang, S. Dudoit, P. Luu et ál., “Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation”, Nucleic Acids Research, vol. 30, no. 4, 2002.

[72] “Genomic Data Commons,” [Online]. Available: <https://portal.gdc.cancer.gov/>.

**Contestación al discurso de Ingreso en la Academia de
Ciencias Matemáticas, Físico-Químicas y Naturales de
Granada del**

Ilmo. Sr. D. Ignacio Rojas Ruiz

por el

Excmo. Sr. D. Alberto Prieto Espinosa

Excelentísimo Sr. Presidente,

Excelentísimos e Ilustrísimos miembros de la Academia

Señoras y Señores:

Deseo comenzar esta intervención expresando mi satisfacción y profundo agradecimiento hacia la Junta de Gobierno de esta Academia por el privilegio y honor que me ha concedido para contestar, en su nombre, en este acto de ingreso como nuevo académico, al discurso pronunciado por el Profesor Ignacio Rojas Ruíz, así como para destacar los aspectos más sobresalientes de su trayectoria y de su labor investigadora.

Como en las contestaciones que he dado a los ingresos de otros miembros en nuestra Academia, tengo que señalar que para mí este privilegio es tanto grato como fácil. Grato, por la estrecha relación que he mantenido con el nuevo académico, desde que fue alumno mío hace 29 años (curso 1991-1992), relación que tan generosamente ha calificado en su discurso; y fácil, por los excepcionales méritos académicos y científicos que concurren en él, que sin duda, como argumentaré más adelante, han contribuido notablemente a que nuestra Universidad ocupe un puesto excepcional dentro del ámbito de la Informática, tanto a nivel nacional como europeo.

Me voy a referir en primer lugar a su discurso y después a su trayectoria profesional.

Sobre el discurso

Considero que el Prof. Rojas nos ha descrito con extraordinaria claridad y precisión el desarrollo e hitos más relevantes de un ámbito científico-tecnológico de vital trascendencia como es el de la bioinformática. Creo que después de haber conocido el contenido del discurso del Prof. Rojas y de los recientes acontecimientos pandémicos nadie pondrá en duda lo adecuado de mi calificativo de “vital

transcendencia”. La descripción efectuada en el discurso tiene aún más mérito al haberla realizado en unas pocas páginas. No se puede efectuar un resumen tan interesante sin conocer en profundidad el tema.

Sería para mi pretencioso completar o añadir una coma al brillante discurso que acabamos de oír; únicamente deseo recordar una vez más lo que comenté en mi discurso de ingreso en esta Academia acerca de lo artificioso que resulta establecer distintas disciplinas dentro de las Ciencias Naturales. La Naturaleza, podríamos decir, que es un todo continuo y no distingue entre Física, Química, Biología o Geología. Realmente no existen fronteras entre las ciencias, sino demarcaciones prácticas o coyunturales. Así ocurre con la bioinformática que es un área de investigación que aplica la informática y las tecnologías de la información al procesamiento de datos del campo de la biología, y se encuentra en un punto de confluencia entre la informática, las ciencias biomédicas y la física, la química y las matemáticas. Este hecho lo corrobora la descripción de su desarrollo realizada por el Prof. Rojas en donde ha citado los nombres de biólogos (Francis Crick, Nicholas Loman, Craig Venter, Allan Maxam), físicos (James Watson, Maurice Wilkins y Stephen Quake), bioquímicos (Frederick Sanger, Robert William

Holley, y Ray Wu), químicos (Rosalind Franklin y Walter Gilbert), médicos (Leroy Hood) y matemáticos (Stephen Quake).

También deseo referirme brevemente a otro ámbito de investigación en el que el Prof. Rojas ha realizado contribuciones muy relevantes, y que él ha citado sólo de pasada, como es el de la predicción a partir de series temporales. En éste ámbito ha dirigido varias tesis doctorales, tiene muy relevantes publicaciones y ha sido además en varias oportunidades, la última en el pasado 2019, Presidente del Comité Organizador de la International Conference on Time Series and Forecasting.

Deseo citar el campo de la predicción científica o computacional porque está en pleno desarrollo y conectado con el ámbito de Big Data al que nuestro nuevo académico se ha referido en varias ocasiones. En este ámbito una vez más podemos decir que somos ricos en datos pero pobres en conocimiento ya que sólo nos aprovechamos mínimamente de las cantidades ingentes de datos disponibles para realizar proyecciones con ellos. Desde las medidas de fenómenos naturales como pueden ser la cantidad de precipitaciones o la temperatura de una determinada región, hasta el consumo de

electricidad de una ciudad o país, pasando por los precios de un determinado producto, todas ellas podemos considerarlas como una mera sucesión de valores obtenidos a intervalos normalmente regulares de tiempo, o **serie temporal**.

Conviene señalar que existen muchos tipos de series temporales originadas por muy diferentes fenómenos, y por tanto existen innumerables comportamientos diferentes. Si el fenómeno que origina la serie fuera conocido con exactitud podría construirse un modelo del mismo y la predicción podría ser directa y exacta. Sin embargo, la mayoría de los fenómenos fluyen con una dinámica compleja o desconocida o sin la suficiente precisión, por lo que resulta inviable obtener un modelo a partir del fenómeno. Esta imposibilidad la podemos soslayar si somos capaces de inferir el modelo a partir de los valores de la serie, que es lo que se pretende realizar el campo que estoy describiendo. La dificultad se incrementa notablemente al tener en cuenta que en la mayoría de los fenómenos, el conocimiento de valores pasados de la serie no contiene toda la información disponible para poder predecir valores futuros, sino que también deben considerarse variables externas o exógenas no pertenecientes a la serie. Estas variables pueden ser desde variables discretas (día de la semana) a variables continuas (temperatura exterior), e incluso

otras series temporales. Debido a lo anterior, desde el punto de vista práctico, cuando uno tiene una serie temporal de un proceso real y quiere extraer un modelo capaz de predecir qué va a suceder en la serie en instantes futuros, cualquier fuente de información, incluida o no en el modelo, debe ser tomada en cuenta.

En este sentido, recuerdo un trabajo realizado en el seno de nuestro grupo de investigación para EMASAGRA que trataba de predecir el consumo de agua en las distintas zonas de la ciudad al objeto de que los depósitos de agua estuviesen en todo momento suficientemente llenos para que no faltase el suministro en ningún punto de nuestra ciudad. Después de esbozado el modelo, al cotejarlo con la realidad observamos que había picos de consumo de agua en la ciudad que no podíamos pronosticar. Después de analizar el problema de forma distante, llegamos a la conclusión de que estos extraños picos de consumo se producían precisamente en los descansos de los partidos televisados de fútbol, en los que una gran cantidad de aficionados van al aseo. En definitiva, para ajustar el modelo en cierta medida hubo que introducir en él como variable exógena el calendario de la liga de fútbol. Esta anécdota se puede aplicar a cualquier predicción científica ya que pretenden tener un sentido de realidad, pero por lo general

se realizan sobre los incompletos supuestos que se consideran en el modelo. Lo habitual es que existan variables ocultas que no se han tenido en cuenta.

Aparte de anécdotas, el estudio de las series temporales es muy relevante para la sociedad en general, siendo sus aplicaciones amplísimas en campos de la ciencia y de la tecnología que tienen gran dificultad en la predicción exacta. En efecto, en algunos campos la complejidad de los procesos y de los datos hace muy difícil el pronóstico; entre otros se encuentran la predicción de pandemias, la evolución demográfica, la dinámica de la población, la meteorología, la predicción de los desastres naturales y, en general, el análisis de índices bursátiles y el modelado y predicción de cualquier proceso industrial, químico o natural. Las cantidades ingentes de datos digitales que se encuentran en la nube, y que se van acumulando año tras año, nos permiten, con ayuda de las técnicas computacionales de análisis de series temporales, predecir y ampliar nuestro conocimiento sobre fenómenos complejos.

Estamos en la Academia de Ciencias y deseo destacar que las investigaciones sobre bioinformática, biomedicina y la predicción a partir de series temporales entran dentro de lo que

en la actualidad se denomina **Ciencia de Datos**. Esta disciplina constituye una herramienta fundamental y eficaz para el avance de las distintas ciencias naturales y experimentales que acoge nuestra institución ya que se centra en métodos científicos, procesos y sistemas para extraer conocimiento implícito (oculto) o un mejor entendimiento de datos en sus diferentes formas. Así, está proyectada para trabajar con datos incompletos, procesar datos que suelen estar desordenados o no estructurados, y gestionar y procesar cantidades ingentes de datos heterogéneos de diferentes bases de datos, inabordables de administrar con técnicas convencionales.

Dentro de la Ciencia de Datos se incluyen conceptos tales como **Big Data** que se refiere a datos cuyo volumen, diversidad y complejidad requieren nuevas arquitecturas hardware, técnicas, algoritmos y análisis para gestionar y extraer valor y conocimiento implícitos (oculto) en ellos; y la **Minería de Datos** cuyo objetivo es a partir de una cantidad ingente de datos (la mina), extraer (minar) patrones de información ocultos, no triviales, desconocidos y potencialmente útiles (la mena) distinguiéndolo de la información irrelevante (la ganga).

En mi opinión el discurso de Ignacio Rojas es de gran interés ya que no es habitual escuchar el desarrollo de los últimos descubrimientos en biomedicina desde la perspectiva del procesamiento de datos y las expectativas que esto conlleva en el campo de la medicina personalizada o de precisión.

Trayectoria profesional

Una vez que he comentado el discurso del Prof. Ignacio Rojas que acabamos de oír, me voy a referir a su trayectoria profesional.

Nació hace 51 años (1969) y estudió EGB y bachillerato en la ciudad de Granada. Al terminar sus estudios de secundaria, ingresó en la Universidad de Granada como alumno de Ciencias Físicas en la Facultad de Ciencias, obteniendo el título de licenciado en 1992.

Una vez obtenido el título se vinculó al hoy denominado Dpto. Arquitectura y Tecnología de Computadores, donde desde entonces ejerce como profesor en la E.T.S. Ingenierías Informática y de Telecomunicación.

En 1996 se doctoró en Ciencias Físicas, y en el 2009 obtuvo el nombramiento de Catedrático de Universidad.

Su campo de investigación se centra en el estudio de sistemas multidimensionales complejos utilizando sistemas inteligentes, apoyado por plataformas de computación de altas prestaciones, enfocado a la resolución de problemas reales en diversos campos, tales como la bioinformática, la biomedicina y la predicción a partir de series temporales.

Como resultado de las investigaciones desarrolladas, ha realizado más de 270 publicaciones reflejadas en la base de datos ISI Web of Knowledge, obteniendo un total de 3.310 citas, e índice $h=28$. Cuarenta y nueve revistas en las que ha publicado son del primer cuartil (Q1) de la prelación de índice de impacto dentro de su ámbito. Según la base de datos Google Scholar, el número de citas es de 7.059, índice $h=39$, índice $i_{10}=138$. Sus publicaciones han obtenido a lo largo de los últimos cinco años un promedio de 265 citas/año. De forma complementaria a la publicación en revistas indexadas, ha participado con contribuciones científicas en más de 125 congresos internacionales relacionados con su ámbito de investigación, y ha dirigido 25 tesis doctorales.

Fue ganador de la competición internacional de predicción de series temporales convocada por el European Symposium on Time Series Prediction (2008).

A lo largo de su trayectoria investigadora, ha participado en 38 proyectos; siendo investigador principal en 8 del Plan Nacional de I+D+i, y 5 de Excelencia de la Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía. Destacar que también ha participado como investigador en 3 proyectos del 7º Programa Marco de la Comisión Europea (FP7) en el ámbito de las Tecnologías de la información y la comunicación (ICT).

Ha participado en un Proyecto CENIT y en 15 contratos de investigación de carácter innovador firmados a través de la Oficina de Transferencia de Resultados de la Investigación (OTRI).

Es coeditor de 15 volúmenes de la serie Lecture Notes on Computer Science de la editorial alemana Springer-Verlag, correspondientes a actas de congresos internacionales.

Entre sus 14 estancias en centros extranjero caben destacar la postdoctoral en la Universidad de Berkeley bajo la supervisión

del Prof. Lotfi A. Zadeh, las realizadas como profesor invitado en la Universidad de Dortmund bajo la supervisión del Prof. Karl Goser, y las más recientes en la Universidad de Münster, dentro del programa del Ministerio MECD-Salvador de Madariaga, bajo la supervisión de los profs. Dr. Peter Glösekötter y Xiaoyi Jiang.

Su actividad internacional no se limita sólo a sus estancias en el extranjero, sino que además ha ocupado puestos de gestión relevantes; así desde 2013 a 2018 fue Presidente de la Sección Española de la IEEE Computational Intelligence Society (CIS); es Editor-in-Chief de la revista internacional Forecasting (MDPI) y miembro del equipo editorial en cuatro otras revistas. En la actualidad es co-presidente de los comités organizadores de los siguientes congresos internacionales, que organiza periódicamente:

- International Work-Conference on Artificial Neural Networks (IWANN).
- International conference on Time Series and Forecasting (ITISE)
- International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO).

En cuanto a la labor de gestión universitaria, destacar que desde 2004 a 2012 fue Subdirector de infraestructura, Biblioteca y Gestión Económica de la Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada.

Por otra parte, desde 2013 es Director del Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC-UGR) de la Universidad de Granada. Este centro está formado por 175 investigadores senior, 111 investigadores en formación o contratados, y acoge a lo largo de cada año estancias de unos 70 investigadores procedentes de muy diversos países. Los investigadores senior del Centro son profesores principalmente de la ETSIIT y de la Facultad de Ciencias. Conviene destacar que en la última edición (2020) del Academic Ranking of World Universities (ARWU, Ranking de Shanghai) la Universidad de Granada, a través de la investigación desarrollada en el ámbito de Ciencias e Ingeniería de la Computación, ocupa la primera posición de España, encontrándose tan sólo 19 universidades europeas por delante de nosotros.

Perfil humano

Deseo en estos momentos complementar mis comentarios académicos y científicos haciendo referencia al perfil humano del protagonista de este acto.

La vocación del Prof. Rojas se puede resumir en las siguientes palabras de su discurso: “El profesor universitario es un gran privilegiado, al poder involucrarse en temas de investigación, que más que suponer un trabajo, suponen una fascinación y a veces una seducción, disfrutando de aquello que se va aprendiendo”. Estas ideas describen a la perfección su carácter personal que sirve de substrato de su talla como científico. Su ansia por la búsqueda en las fronteras del conocimiento es lo que ha marcado su trayectoria investigadora en un ámbito como es el de las tecnologías de la información que están en evolución permanente y vertiginosa.

Detrás del físico, informático y científico se encuentra un personaje honesto, generoso y noble. Su generosidad le ha llevado a mencionar los nombres de las principales personas que le hemos acompañado en su camino o que se han cruzado en él, mostrando su aprecio y sentimiento hacia sus compañeros.

Permanentemente ha estado comprometido con la Ciencia, la Universidad y la Sociedad, aportando su inteligencia, capacidad, perspicacia y perseverancia. Estos compromisos los comparte con su mujer Olga y sus hermanos Francisco Javier y Fernando, todos ellos profesores de la Universidad en los ámbitos de la Matemática Aplicada, la Educación Física y Deportiva, y la Arquitectura de Computadores, respectivamente. Sin duda sus tres hijos, Olga, Ignacio y Laura, están recibiendo una educación excelente a través del ejemplo de sus familiares más cercanos entre los que incluimos a sus abuelos.

Conclusión

Y concluyo satisfecho porque estoy convencido de que mis lazos de amistad no me han llevado a hacer un juicio imparcial sobre sus méritos, ya que me he limitado a exponer públicamente con datos y de forma objetiva los méritos del nuevo académico que han determinado su ingreso en nuestra Academia.

En definitiva, después de analizar la brillante trayectoria del Prof. Ignacio Rojas Ruíz, podemos afirmar categóricamente

que es una persona cabal, un excelente científico y un universitario integral, y por todo ello, excelentísimos e ilustrísimos miembros de la Academia, señoras y señores, me produce una gran satisfacción su ingreso en nuestra noble institución, como Académico Correspondiente, y contar con su participación en la misma. Además de sus excepcionales cualidades tanto científicas y académicas como humanas, y que he tratado de poner de manifiesto, el ingreso de esta relevante personalidad supone reforzar dentro de nuestra academia un campo, hoy día de importancia trascendental, como es el de la informática.

Ilustrísimo Profesor Rojas en nombre de la Academia y en el mío propio, le doy la bienvenida a esta noble institución con mis más cordiales felicitaciones, a la par que le agradecemos su disposición a colaborar en sus tareas propias, y le deseamos toda clase de venturas y felicidad tanto personales como familiares.

Muchas gracias.

Alberto Prieto Espinosa

