



Entrez, The Life Sciences Search Engine

Centro Nacional para la Información Biotecnológica de los Estados Unidos: un palacio de la información para la medicina molecular

Lic. Rubén Cañedo Andalia ^I

Lic. Roberto Rodríguez Labrada ^{II}

MsC. Yaimeé Vázquez Mojena ^{III}

^I Licenciado en Información Científico-Técnica y Bibliotecología. Departamento de Fuentes y Servicios de Información. Centro Nacional de Información de Ciencias Médicas-Infomed. Ciudad de La Habana, Cuba.

^{II} Licenciado en Microbiología. Departamento de Neurofisiología Clínica. Centro para la Investigación y la Rehabilitación de las Ataxias Hereditarias. Holguín, Cuba.

^{III} Máster en Virología. Departamento de Biología Molecular. Centro para la Investigación y la Rehabilitación de las Ataxias Hereditarias. Holguín, Cuba.



Fecha de actualización

6 de mayo de 2011. Actualizado a partir de Cañedo Andalia R, Rodríguez Labrada R, Vázquez Mojena Y. Centro Nacional para la Información Biotecnológica de los Estados Unidos: un palacio de la información para la medicina molecular. *Acimed* 2009; 19(4). Disponible en:

http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352009000400003&lng=es&nrm=iso&tlng=es

Correspondencia

Lic. *Rubén Cañedo Andalia*. Grupo de Alfabetización Informacional. Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas Holguín. Ave. V. I. Lenin No. 4 e/ Aguilera y Agramonte. CP. 80100. Holguín. Cuba. Correo electrónico: ruben@infomed.sld.cu

Derechos de copia



Copyright: © Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas Holguín. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 3.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original. La imagen del cabezal de la portada se tomó de: <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>.

Contenido

Marchamos aceleradamente hacia la era de la medicina molecular donde cada individuo gozará de una atención clínica personalizada sobre la base de sus particularidades génicas en cada momento de su vida. Las escuelas de medicina introducen con rapidez cambios en sus currículos y los sistemas de información se alistan para la era que nace. Se realiza una panorámica de los múltiples recursos de información que ofrece el Centro Nacional para la Información Biotecnológica de los Estados Unidos. Con ella se pretende estimular el uso de sus excepcionales recursos de información, porque estamos firmemente convencidos de su utilidad para la formación de nuestros estudiantes y jóvenes profesionales, así como para la realización de investigaciones que puedan situarse a la "altura" de las exigencias actuales a escala internacional.

Introducción

El sistema de información del National Center for Biotechnology Information, de los Estados Unidos, es para nosotros la experiencia real más cercana a la soñada Web 3.0 en las esferas de las ciencias de la vida y de la salud.

LOS AUTORES

El influjo de las ciencias biológicas y de increíbles tecnologías está cambiando dramáticamente el panorama de las ciencias médicas, la clínica y los servicios de salud. Los últimos avances en el área de la biotecnología, que han marcado científicamente al 2008, nos acercan más a la era de la genómica personal, una nueva etapa en la que será posible leer completamente lo que "dice" nuestro genoma (anexo). En los últimos 5 años se han desarrollado plataformas de secuenciación masiva en paralelo, que permiten distinguir hasta cuatro millones de variaciones individuales simples, que afectan a una sola base, conocidos como polimorfismos de nucleótido simple (SNPs por sus siglas en inglés) y centenares de miles de variaciones genéticas de mayor tamaño,¹ en forma mucho más rápida, precisa y, sobre todo, más barata. Actualmente, un equipo de esta clase cuesta alrededor de 400 000 de euros y puede leer hasta 10 millones de pequeños fragmentos de ADN en varios días, un proceso que se prolonga durante algunas semanas si lo que se desea es secuenciar el genoma completo, formado por 3 000 millones de pares de bases o nucleótidos y unos 20-25 000 genes.²

Los estudios del genoma humano completo aportaron información sobre varias enfermedades, como el trastorno afectivo bipolar, el cáncer de mama y colorrectal, la diabetes mellitus tipos I y II, ciertas afecciones cardíacas, la hipertensión arterial, la esclerosis múltiple y la artritis reumatoide.³ Pero no existe un genoma único como referencia, sino muchos. Más del 10 % del genoma es distinto entre las personas. Y es precisamente en las regiones que cambian y son más susceptibles a la variabilidad donde más se debe trabajar. Son estas las que pueden ofrecernos información muy importante con respecto a cómo funcionan ciertos aspectos de nuestra biología que son fundamentales para la adaptación al entorno y que pueden contener las claves sobre la susceptibilidad a determinadas enfermedades.⁴

A partir de la secuencia del genoma humano, los científicos comenzaron a seguir variaciones diminutas llamadas polimorfismos de nucleótido simple. Esas variaciones o polimorfismos son claves para comparar el ADN de miles de individuos con o sin una enfermedad y determinar cuáles variantes genéticas pueden suponer un riesgo de

padecerla. Al determinar la variación genética humana, se halló que entre miles de millones de bases del ADN, muchas pueden perderse, añadirse o copiarse de forma que alteren su estructura genética en unas cuantas generaciones.³ No obstante, la certeza de padecer cierta enfermedad a partir del examen del mapa genético de un individuo es sólo posible en algunas de ellas. En muchas, depende también de la interacción de éste con el medio.

Pero la información genética codificada en las moléculas de ADN no es la única portadora de la herencia biológica: la epigenética influye también en la transmisión de rasgos y enfermedades. Se denomina epigenética al conjunto de modificaciones químicas que experimentan el ADN y las proteínas que lo envuelven y que regulan su expresión, a partir de dos procesos básicos: la metilación y la acetilación.⁵

Cada vez está más claro que en muchas enfermedades los genes no son los únicos que tienen la palabra. Estos "hablan" o "callan" según las circunstancias y el entorno en el que se encuentren. Procesos como la metilación, entonces, adquieren una gran importancia, porque pueden, sin cambiar la secuencia de los genes, modificar su expresión como resultado de la exposición a algún factor medioambiental. En general, la interacción entre factores genéticos y medioambientales puede conducir a modificaciones en el desarrollo de la vida de un individuo que pueden tener un carácter duradero.⁶ Conocer ciertas predisposiciones genéticas puede ayudar a desarrollar un comportamiento personal que coadyuve a evitar el padecimiento, demorarlo o reducirlo.⁷

La fármaco-genómica, una disciplina que pretende conquistar el campo de la medicina personalizada, no es ya un sueño lejano. Las nuevas metodologías de secuenciación, que se encuentran actualmente en fase de desarrollo, prometen en 5-10 años poder leer un genoma humano completo en 15 minutos.⁸ Esto, sin dudas, constituirá un gran paso de avance para el ejercicio de la clínica y el desarrollo de las ciencias médicas en su conjunto porque, entre otros beneficios, posibilitará la identificación de genes o de sus interacciones, que conducen con frecuencia a la aparición de ciertas enfermedades cuyo tratamiento podrá enfrentarse con opciones terapéuticas adecuadas a las características de cada individuo.

Marchamos entonces, con pasos de gigantes, hacia una medicina molecular donde cada individuo gozará de una atención clínica personalizada sobre la base de sus particularidades génicas en cada momento de su vida, determinadas, como dijimos, por la interacción con ciertos factores ambientales internos y externos nutricionales,

metabólicos, microbiológicos y otros que intervienen en el cambio, silenciamiento o expresión de algunas de estas unidades de la herencia.

Durante los últimos años, se ha producido un espectacular crecimiento de la investigación, la literatura y los recursos de información en el área biológica con claras implicaciones para el ejercicio de la clínica.

Observemos un ejemplo sencillo. Una búsqueda como: *Breast Neoplasms/genetics[Mesh]*, con límites temporales y sin estos:

Publicados en los últimos:	Número de referencias	Acumulado (%)
3 años	4 585	23,68
5 años	7 513	38,80
10 años	13 224	68,30
Sin límite temporal (total)	19 359	100,00

Fuente: PubMed-Medline. 4 de febrero de 2009.

El primer año en el cual se registró un artículo que trató sobre la genética del cáncer de mama en el sexo femenino bajo el descriptor utilizado fue 1957, hace 52 años. Sin embargo, en los últimos 10 años se ha registrado casi el 70 % de las referencias a esta clase de artículos. Los últimos cinco años acaparan aproximadamente el 40 % de ellas y los últimos tres años comprenden casi una cuarta parte del total de las publicaciones procesadas durante los más de 50 años en los que se puede seguir el comportamiento de la producción científica sobre este tema de investigación de interés crucial para la medicina.

Hoy, la biología se transforma aceleradamente como consecuencia del crecimiento explosivo de los datos procedentes de los laboratorios de investigación distribuidos por todo el mundo. El reto consiste en transformar esos datos en conocimiento, un conocimiento que lleve a una mejor comprensión de los procesos biológicos que subyacen, tanto en la salud, como en la enfermedad. La clave para obtener nuevos conocimientos, a partir de los datos acumulados, es desarrollar programas para supercomputadoras capaces de explorar integralmente conjuntos de datos masivos y complejos imposibles de analizar de otra manera.⁹

Comprender el lenguaje elegante, pero silencioso, de la naturaleza de las células vivas es el propósito de la biología molecular. De un alfabeto de tan solo cuatro letras, que representan las unidades elementales del ADN, emerge la sintaxis de los procesos de la

vida, que adquieren su mayor complejidad en el hombre. El uso de este alfabeto para formar nuevas "palabras y frases" es el propósito central de la biología molecular.¹⁰

La biología molecular, como indica su nombre, estudia la vida a un nivel molecular. Ella busca comprender las interacciones de los diferentes sistemas de la célula, que comprende múltiples relaciones, entre ellas, las del ADN con el ARN, la síntesis de proteínas, el metabolismo energético y cómo estas interacciones se regulan para conseguir un refinado funcionamiento de la célula.¹¹

Evidentemente los paradigmas están cambiando y los datos por sí solos son abrumadores. Las escuelas de medicina introducen aceleradamente cambios en sus currículos y hasta los sistemas de información se alistan para la nueva era que está naciendo. La práctica de la medicina está siendo y será "arrastrada" cada vez de forma más rápida por las nuevas corrientes de conocimientos y tecnologías procedentes del desarrollo de la biología molecular, la genética y otras disciplinas afines.

Hoy día, es prácticamente imposible escribir un informe de investigación original sobre una entidad clínica cualquiera sin referirse en una de sus partes, al menos, a la información genética y molecular pertinente.

El Centro Nacional para la Información Biotecnológica de los Estados Unidos

Fundado el 4 de noviembre de 1988, y ubicado en Bethesda, Maryland, el *National Center for Biotechnology Information* (NCBI- <http://www.ncbi.nlm.nih.gov/>) es una división de la *National Library of Medicine*, uno de los Institutos Nacionales de Salud de los Estados Unidos. Como recurso nacional de información sobre biología molecular y uno de los más poderosos en las llamadas ciencias de la vida en general, el centro desarrolla constantemente nuevas tecnologías de información para ayudar a comprender, tanto los procesos genéticos, como moleculares que controlan la salud y la enfermedad.^{10,12}

El sitio del NCBI, conformado por un amplio y diverso banco de bases de datos, herramientas y otros medios, posibilita la exploración integral de sus recursos mediante un sistema de recuperación (un gran motor de búsqueda) de interfaz única denominado *Entrez, The Life Sciences Search Engine* (<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>), distribuido por primera vez en 1991, que permite la búsqueda simultánea de información generalmente de carácter público sobre los más diversos aspectos de las ciencias de la vida y la salud en decenas de bases de datos de una calidad excepcional (fig. 1).¹³

Fig. 1. *Entrez, The Life Sciences Search Engine.*

The screenshot displays the NCBI Entrez search engine interface. At the top, the NCBI logo and the text "Entrez, The Life Sciences Search Engine" are visible. Below the header, there are navigation tabs for "HOME", "SEARCH", "SITE MAP", "PubMed", "All Databases", "Human Genome", "GenBank", "Map Viewer", and "BLAST". The search bar contains the text "ATXN2" and buttons for "GO", "Clear", and "Help". Below the search bar, a message states: "- Result counts displayed in gray indicate one or more terms not found". The search results are organized into two columns of database-specific results, each with a result count in a gray box and a small icon representing the database. The results are as follows:

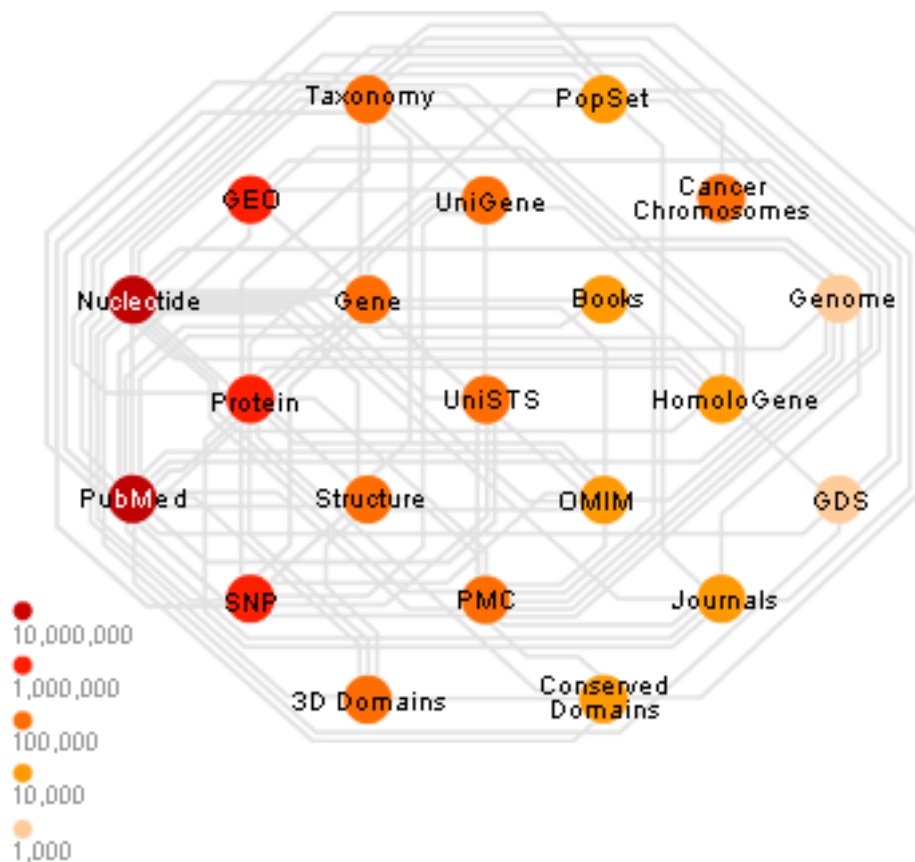
Database	Result Count	Description
PubMed	25	biomedical literature citations and abstracts
PubMed Central	51	free, full text journal articles
Site Search	none	NCBI web and FTP sites
Books	14	online books
Images	21	images from full text resources at NCBI
OMIM	12	online Mendelian Inheritance in Man
Nucleotide	70	Core subset of nucleotide sequence records
EST	9	Expressed Sequence Tag records
GSS	5	Genome Survey Sequence records
Protein	31	sequence database
Genome	14	whole genome sequences
Structure	none	three-dimensional macromolecular structures
Taxonomy	none	organisms in GenBank
dbGaP	4	genotype and phenotype
UniGene	8	gene-oriented clusters of transcript sequences
EDD	none	conserved protein domain database
UniSTS	27	markers and mapping data
PopSet	none	population study data sets
GEO Profiles	4278	expression and molecular abundance profiles
GEO DataSets	none	experimental sets of GEO data

Por la vía de *Entrez*, es posible acceder, tanto a las principales bases de datos de secuencias de proteínas y ADN disponibles como *Nucleotide*, *dbSNP*, *Proteins*, *PubChem*, *Gene*, entre otras como a una gran parte de la mejor literatura biomédica mundial procesada por bases como *PubMed-Medline* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>), la más popular y una de las más prestigiosas base de datos en el área de las ciencias médicas a escala mundial, producida por la *National Library of Medicine* y OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>), un compendio autorizado y exhaustivo de los genes humanos y los fenotipos genéticos, desarrollada por la *Johns Hopking University*. El acceso libre al texto completo de la información disponible es también cada vez más frecuente.

El NCBI agrupa sus bases de datos esenciales en tres grandes sectores: *Literature Databases*, *Molecular Databases* y *Genomes*.¹⁴ Estas dos últimas clases comprenden un grupo amplio y diverso de bases de datos biológicas cuya información procede básicamente de los resultados de experimentos científicos, suministrados directamente por los laboratorios o instituciones que los realizan o publicados en la literatura científica especializada, donde con frecuencia se aplican tecnologías de experimentación de muy alto rendimiento y el análisis computacional. La información contenida en estas bases de datos comprende: funciones, estructura y localización (tanto celular como cromosómica) de los genes, los efectos clínicos de las mutaciones, así como las similitudes entre secuencias y estructuras biológicas.¹⁵

El nivel de entrecruzamiento de la información registrada en las diferentes bases de datos es tan alto que prácticamente por cualquier punto que se "enganche" el "palangre" (arte de pesca de arrastre) es posible recuperarlo completo con su preciada carga de "peces", en este caso, con una gran cosecha de información procedente de múltiples fuentes de gran calidad (fig. 2).

Fig. 2. Sistema integrado de búsqueda.



Literature databases

Comprende cuatro grandes bases de datos:

- *PubMed* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>)

PubMed es el recurso bibliográfico más utilizado en el área de la salud en Internet. Su acceso es gratuito. *Medline*, su base de datos principal, es insigne y la más popular en el campo de la salud desde hace varias décadas. *PubMed-Medline* procesa más de 5 200 revistas de unos 80 países y comprende los campos de la medicina, la enfermería, la estomatología, la veterinaria, la gestión de salud, las ciencias preclínicas y algunas áreas de las ciencias de la vida.¹⁶ Los archivos de *PubMed* contienen más de 20 millones y medio de referencias desde el año 1865 hasta la fecha. *PubMed* recibe docenas de millones de visitas anuales de usuarios de todo el mundo por la calidad de su información y por su acceso gratuito, a diferencia, por ejemplo, del *Web of Science*, cuya exploración requiere de una suscripción previa. Con *PubMed* es posible, además, acceder a múltiples bases de datos en el área de las ciencias de la vida, en particular en los campos de la

genética y la farmacología, a partir de las facilidades que ofrece la plataforma del NCBI.

- *PubMed Central* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Pmc>)
PubMed Central (PMC) es el archivo digital de revistas biomédicas y en ciencias de la vida de los Institutos Nacionales de Salud de los Estados Unidos de Norteamérica, desarrollado y administrado por el NCBI y la Biblioteca Nacional de Medicina. No es una casa editora, sino una gran hemeroteca digital que ofrece acceso libre y gratuito a más de 1 200 revistas y 2 100 000 artículos en la esfera de las ciencias de la vida.¹⁷
- *Images Database*. (<http://www.ncbi.nlm.nih.gov/images/>) Provee acceso a imágenes compiladas a partir de los recursos a texto completo que posee el NCBI.
- *Books* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>)
El primer libro disponible en el sitio del NCBI fue la tercera edición del *Molecular Biology of the Cell*, de los autores *Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts y James D. Watson*, editado por *Garland Publishing*. Este es uno de los libros de texto más utilizados por los estudiantes en el área de la biología celular y molecular. *Books* es una colección de libros biomédicos que crece progresivamente, que puede explorarse fácilmente con solo escribir algunas palabras clave en su ventana de búsqueda.¹⁹

Molecular databases

Se divide en seis clases principales:¹⁴

- *Nucleotide Sequences*
(<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>)
- *Protein Sequences* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>)
- *Structures* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Structure>)
- *Genes* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>)
- *Gene Expression* (<http://www.ncbi.nlm.nih.gov/geo/>)
- *Taxonomy* (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>)

Estas clases principales están, a su vez, compuestas por bases de datos y subconjuntos de ellas.

Nucleotide sequences

Comprende bases de datos como:

- *Nucleotides* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>)

Descifrar totalmente el genoma humano, tanto su estructura como funciones de cada compuesto hasta el nivel de los nucleótidos, requiere aún de años de trabajo. Téngase en cuenta que estamos hablando de unos 3 000 millones de piezas. *Nucleotides* valida y actualiza diariamente información procedente de múltiples organismos e instituciones empeñados en todo el mundo en descifrar nuestro genoma y coloca a disposición de la comunidad científica internacional cada nuevo descubrimiento, cada nueva porción de datos hallada, repetimos, cada día. Esta base permite la exploración simultánea de tres bases de datos: *dbEST*, *dbGSS*, cuyas descripciones aparecen más adelante, y *CoreNucleotide*, que comprende secuencias de nucleótidos no incluidas en las antes referidas. Esto permite la rápida exploración y la recuperación de resultados con mayor especificidad de los registros de secuencias de nucleótidos. Los componentes de esta base de datos, juntos, contienen el total de los datos de secuencias registrados en *GenBank*, el *European Molecular Biology Laboratory* (EMBL) y el *DNA DataBank* (DDBJ) de Japón, miembros de la *International Collaboration of Sequence Databases*. *Nucleotides* posibilita buscar por campos específicos como: *Gene Name*, *Organism*, *Protein Name* y otros.²⁰

- *dbSNP (Single Nucleotide Polymorphisms)*

[<http://www.ncbi.nlm.nih.gov/SNP/index.html>]

Un polimorfismo de nucleótido simple, como referíamos, es una variación comúnmente de una sola base que ocurre en el ADN humano con una frecuencia aproximada de uno por cada 1 000 bases. Estas variaciones se pueden emplear para rastrear patrones de herencia familiar. *Single Nucleotide Polymorphisms* (dbSNP) es un archivo digital central que registra, tanto sustituciones de nucleótidos simples, como deleciones —pérdida de un fragmento de ADN de un cromosoma; la deleción de un gen o de parte de un gen puede ocasionar una enfermedad o una anomalía— cortas y polimorfismos de inserción (una inserción es un tipo de anomalía cromosómica en el cual un segmento de ADN se inserta en un lugar diferente a la que comúnmente se encuentra; en algunos casos, altera la estructura y función normal de un gen). Las variaciones de secuencias existen en posiciones definidas del genoma y son responsables de ciertas características fenotípicas individuales, incluida la predisposición de algunas personas hacia enfermedades como las del corazón y el cáncer. Como herramientas para entender estas variaciones y la genética molecular de los humanos, pueden

estudiarse las variaciones de secuencias para el mapeo de genes, la definición de la estructura de las poblaciones y la ejecución de estudios funcionales. *Single Nucleotide Polymorphisms* es un archivo de dominio público con una amplia colección de polimorfismos genéticos simples. Esta colección incluye sustituciones de nucleótidos simples (conocidos además como polimorfismos de nucleótido simple), deleciones o inserciones de bases múltiples a pequeña escala (también llamado polimorfismos de inserción y deleción) e inserciones de elementos transponibles y variaciones de repeticiones de microsatélites (además llamado repeticiones cortas en *tandem*). Una búsqueda para el gen de la ataxia espinocerebelosa tipo 2 (ATXN2) limitada a *Homo sapiens*, permitió identificar, el 9 de febrero, 760 variaciones de este tipo en el gen indicado (fig. 3). Obviamente, que muchas de ellas no guardan relación alguna con esta enfermedad.^{21,22}

- *Probe* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=probe>)

Probe es un registro público de reactivos químicos utilizados para la detección de secuencias específicas de ácidos nucleicos de uso en una amplia variedad de investigaciones biomédicas, como el desarrollo de los mapas génicos, la determinación de la expresión y el silenciamiento genético y de variaciones genéticas o polimorfismos. Contiene información detallada sobre las técnicas y procedimientos reacción en cadena de la polimerasa (PCR por sus siglas en inglés), resecuenciación, entre otras utilizadas para esta clase de estudios, así como sobre sus compañías distribuidoras. *Probe* (sonda), como término, se refiere a una secuencia específica de bases que se utiliza para detectar una secuencia complementaria de bases. La determinación, por ejemplo, de una variación genética puede servir para el diagnóstico temprano, la prevención y el tratamiento de diversas enfermedades. Las explicaciones que ofrece la base de datos sobre las tecnologías empleadas para esta clase de investigaciones son muy útiles para quienes se inician en esta esfera del conocimiento.²³

- *UniSTS (Sequence Tagged Sites)*

[<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unists>]

Los sitios de secuencias marcadas: *sequence tagged sites* (STSs, por sus siglas en inglés) son segmentos cortos de ADN que aparecen sólo una vez en el genoma humano, en una posición exacta y el orden de sus bases es conocido. *UniSTS* ofrece una vista unificada, no redundante, de los llamados STSs e integra, tanto datos de los mapas (localización y otros), como información sobre los marcadores genéticos en una colección única a partir de diversos recursos de información públicos. Sus fuentes de información comprenden bases como *dbSTS* (se describe más adelante), *RHdb (Radiation Hybrid data base* [<http://www.ebi.ac.uk/RHdb>]) y *GDB (Human Genome Database* [<http://www.gdb.org>]); varios mapas genéticos,

como el *Whitehead RH Map*, *Whitehead YAC*, *Stanford RH map*, *NHGRI chr 7 physical map* y el *WashU chrX physical map*, así como otros del ratón como: *Mouse RH map*, *Whitehead Mouse YAC* y el *Jackson Laboratory's MGD*.^{24,25}

- *PopSet (Evolutionary Relatedness)*

[<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PopSet>]

Es un conjunto de secuencias de ADN reunidas para analizar el parentesco en la evolución de una población. Contiene secuencias alineadas, obtenidas como resultado de estudios de mutaciones, filogenéticos o de poblaciones. Estas alineaciones describen eventos evolutivos y las variaciones ocurridas en la población que se estudia. Las secuencias pueden ser, tanto de nucleótidos, como de proteínas.^{26,27}

Fig. 3. dbSNP.

The screenshot displays the NCBI dbSNP interface. At the top, there's a search bar with 'SNP' selected and 'ATXN2' entered. Below the search bar, there are tabs for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The main content area shows search results for 'homo sapiens'. It includes a 'Display' dropdown set to 'Graphic Summary', a 'Show' dropdown set to '20', and a 'Sort by' dropdown. Below this, it shows 'All: 760', 'Cited in PubMed: 5', 'Clinical/LSDb Submissions: 0', and 'Human: 760'. The results list items 1-20 of 760. Three items are visible:

- Item 1: rs63740561 [Homo sapiens]. Sequence: AAGTAAATGTATGCCCTACATAAAAAA [~/AAAAAAA] AAAAGCACACAATACGGAAAGACAT. HGVS Names: [NM_002973.2:c.1177-1262T>TTTTTTTT] [MT_009775.16:g.2529548A>AAAAAAA].
- Item 2: rs63653160 [Homo sapiens]. Sequence: CCCAAACACAGTGAAGCTAGAGTTTT [~/TTTTG] TTTTGTTTTGTTTTGTTTTGTTTT. HGVS Names: [NM_002973.2:c.732-1307C>CAAAA] [MT_009775.16:g.2576309G>TTTTG].
- Item 3: rs63638660 [Homo sapiens]. Sequence: GCAGAAATAGGTTTAAAGGAAAAAAA [~/AA] AAAAAAAAAAGTCTGCTGAACTCTT.

También existen subconjuntos de secuencias, entre los que se destacan:¹⁴

- *GenBank* (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>)

GenBank es una base de datos exhaustiva, una colección anotada de las secuencias de nucleótidos disponibles al público de decenas de miles de organismos. La anotación comprende la secuencia codificadora (CDS en la base de datos); es decir, la región del nucleótido que corresponde con la secuencia de

aminoácidos en una proteína (la localización contiene el codón de inicio y fin). Incluye también una herramienta para la revisión de los cambios que han ocurrido en los datos de la secuencia registrada a lo largo de la historia (<http://www.ncbi.nlm.nih.gov/entrez/sutils/girevhist.cgi>). La mayoría de los registros procede de miles de laboratorios de investigación distribuidos por todo el mundo. Forma parte de un proyecto de colaboración internacional: la *International Nucleotide Sequence Database Collaboration*, que comprende el EMBL y el DDBJ, de Japón. El intercambio diario de datos entre estos laboratorios asegura a *GenBank* un cubrimiento mundial de la información publicada. La edición bimestral de esta base de datos y su actualización diaria puede descargarse desde el FTP del sitio. Una explicación detallada de las posibilidades de su sistema de búsqueda y el contenido de los registros de la base puede obtenerse en:

<http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html#CDSB>.^{28,29}

- *RefSeq (Reference Sequences)* [<http://www.ncbi.nlm.nih.gov/RefSeq/>]
RefSeq provee un conjunto exhaustivo, integrado, no redundante y correctamente anotado de secuencias de nucleótidos del ADN y ARN, así como de proteínas. Contiene referencias estables para la anotación del genoma, la identificación y la caracterización de genes, el análisis de polimorfismos y mutaciones (especialmente los registros curados por *RefSeq*), el hallazgo de estudios de expresión y la ejecución de análisis comparativos. Comprende diversos organismos biológicos. Cada registro contiene información sobre una molécula natural simple de un organismo. Su propósito es desarrollar un conjunto de datos exhaustivo y normalizado que represente una secuencia de información propia de una especie. Es un subproducto de *GenBank*, pero difiere de este en que cada uno de sus registros contiene una síntesis de información, no una unidad de archivo de datos primarios de investigación. Esto los asemeja a los conocidos artículos de revisión tan comunes en la literatura científica. Cada registro representa una consolidación de la información para un grupo particular de organismos en un momento específico.^{30,31}
- *dbEST (Expressed Sequence Tags)* [<http://www.ncbi.nlm.nih.gov/dbEST/>]
dbEST es una división de *GenBank* que contiene datos de secuencias e información sobre secuencias de ADN complementario "de simple pase" o marcas de secuencias expresadas provenientes de diversos organismos.³²
- *dbGSS (Sequence Tagged Sites)* [<http://www.ncbi.nlm.nih.gov/dbSTS/>]
GSS es una división similar a *EST*, con la excepción de que la mayoría de las secuencias son génicas por su origen, en lugar de proceder de ADN complementario (ARN mensajero).³³

- *dbSTS (Sequence Tagged Sites)* [<http://www.ncbi.nlm.nih.gov/dbSTS/>]
Contiene datos de mapas y secuencias de *STSs*. Es una alternativa a la colocación de estas secuencias en *GenBank*. Fue especialmente diseñada para la publicación de secuencias de *STSs* largas.³⁴

Proteins sequences

Abarca bases de datos como:

- *Proteins* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>)
Las proteínas son moléculas compuestas por aminoácidos. Sus funciones en el organismo son diversas y muy importantes. *Proteins* contiene secuencias de proteínas de diversos organismos provenientes de bases de datos como: *Protein Information Resource* (PIR- <http://pir.georgetown.edu/>), *SWISS-PROT* (<http://www.ebi.ac.uk/swissprot/>), *Protein Research Foundation* (PRF- <http://www.proteinresearch.net/>) y *Protein Data Bank* (PDB- <http://www.rcsb.org/>), así como de la traducción de las regiones codificables de las secuencias de ADN anotadas, disponibles en *GenBank*, *RefSeq*, *EMBL* y *DDBJ*. Registra, además, secuencias de estructuras descubiertas. Entre otras opciones, permite la búsqueda por el nombre del gen y de la proteína de interés.³⁵
- *Protein Clusters* (<http://www.ncbi.nlm.nih.gov/proteinclusters>). El rápido incremento de las capacidades de secuenciación de ADN ha producido un incremento notable de los datos generados por las investigaciones del genoma procariótico. Ello es una ventaja importante para los científicos que estudian la evolución de los microorganismos y aquellos que desean comprender los fundamentos de los sistemas microbiológicos. La base de datos *NCBI Protein Clusters* fue creada para conservar eficientemente y mantener el creciente número de datos obtenidos hasta hoy. *ProtClustDB* contiene grupos de proteínas curadas y no curadas, agrupadas por similitud de secuencias.³⁶
- *CDD (Conserved Domain Database)*
[<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>]
Es una base de datos de dominios de proteínas.³⁷ Las proteínas con frecuencia están compuestas por uno o varios dominios. *CDD* comprende una colección de alineamientos de secuencias múltiples de dominios y proteínas completas ancestrales. La exploración de esta base permite identificar los dominios conservados hasta el presente en una secuencia buscada de alguna proteína en particular.³⁸

- *HIV-1, Human Protein Interaction Database* (<http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/>). Documentar la interacción de las proteínas del virus de inmunodeficiencia humana tipo 1 (VIH-1) con las proteínas de la célula hospedera es crucial para comprender el proceso de replicación y patogénesis del VIH-1, y proveer una base esencial para el desarrollo de terapias seguras y efectivas como estrategias de prevención contra el SIDA. Aunque numerosas interacciones se han reportado en la literatura científica, no existe actualmente una fuente pública disponible para acceder eficientemente a esta información. En consecuencia, la Division of Acquired Immunodeficiency Syndrome (DAIDS) del National Institute of Allergy and Infectious Diseases (NIAID) ha iniciado un proyecto, en colaboración con el Southern Research Institute y el National Center for Biotechnology Information, para compilar una base de datos detallada que ofrezca a los científicos en el campo de las investigaciones de VIH/SIDA, un resumen concreto y detallado de todas las interacciones conocidas de las proteínas del VIH-1 con proteínas de las células hospederas, con otras proteínas del VIH-1 o con proteínas de organismos enfermos, asociados con el VIH/SIDA.³⁹

Structures

Incluye bases de datos como:

- *3D. Macromolecular Structures. MMDB (Molecular Modeling Database)* [<http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>] Comprende la base de datos *Molecular Modeling Database* (MMDB) (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure>) y un conjunto de herramientas bioinformáticas. MMDB ofrece abundante información sobre estructuras biomoleculares. Experimentalmente determinadas, estas estructuras tridimensionales proveen información amplia sobre su función biológica, los mecanismos de las funciones, así como sobre la evolución histórica y las relaciones entre macromoléculas. La mayoría de los datos sobre estructura 3D se obtienen por medio de cristalografía de rayos X y espectroscopia de resonancia magnética nuclear.⁴⁰
- *PubChem Substance* (<http://pubchem.ncbi.nlm.nih.gov/>) Presenta una caracterización detallada, que incluye las estructuras químicas y los resultados del seguimiento de la actividad biológica de millones de sustancias y moléculas químicas. Ofrece abundante información farmacológica, así como

múltiples enlaces a otros recursos con información sobre cada una de las sustancias y moléculas registradas en la base de datos. Comprende: ⁴¹⁻⁴³

- 1) *PubChem Compound* (<http://pubchem.ncbi.nlm.nih.gov/>), que contiene información validada de la representación química que provee *PubChem Substance* para describir las sustancias. Las estructuras almacenadas en esta base de datos se agrupan, y sus referencias se entrecruzan, según la identidad y similitud de los grupos.^{44,45}
 - 2) *PubChem BioAssay* (<http://pubchem.ncbi.nlm.nih.gov/>), contiene información sobre los estudios de bioactividad de las sustancias químicas descritas en *PubChem Substance*. Provee una descripción explorable de cada bioensayo, que incluye una síntesis de las condiciones y los resultados de la aplicación del protocolo de estudio.^{46,47}
- *BioSystems* (<http://www.ncbi.nlm.nih.gov/biosystems/>). Un biosistema o sistema biológico, es un grupo de moléculas que interactúan entre sí en un sistema biológico. Un tipo de biosistema es una vía biológica, que puede consistir en la interacción de genes, proteínas y pequeñas moléculas. Otro tipo de biosistema es una enfermedad, que puede involucrar componentes como genes, biomarcadores y drogas. La base de datos *BioSystems* colecciona información sobre conjuntos de interacciones de biomoléculas involucradas en rutas metabólicas y de señalización, etapas de enfermedades y otros procesos biológicos.⁴⁸
 - *BioSample* (<http://www.ncbi.nlm.nih.gov/biosample>). Contiene descripciones de materiales de origen biológico utilizados en ensayos experimentales. *BioSystems* colecciona información sobre conjuntos de interacciones de biomoléculas involucradas en rutas metabólicas y de señalización, etapas de enfermedades y otros procesos biológicos. Los registros de *BioSystems* se conectan con la literatura relacionada, genes, secuencias de proteínas, estructuras, datos químicos y se enlazan a *BioSample*.⁴⁹
 - *Database of Genomic Structural Variation* (dbVar - <http://www.ncbi.nlm.nih.gov/dbvar>). La variación estructural (SV, por sus siglas en inglés) es generalmente definida como una región de ADN de aproximadamente 1kb o más que puede incluir inversiones y translocaciones balanceadas o desequilibrios genómicos (inserciones o deleciones), comúnmente referidas como variantes del número de copias (CNVs, por siglas en inglés). Estas CNVs con frecuencia se superponen con duplicaciones de segmentos, regiones de ADN >1 kb presentes más de una vez en el genoma, copias de las cuales son >90% idénticas. Si se presenta >1% en una población, una CNV puede ser referida como un polimorfismo del número de copias (CNP, por sus siglas en inglés).⁵⁰

Genes (comprende genética general y clínica)

Contiene bases de datos como:

- *Gene* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>)

Es una base de datos para la información específica sobre genes. No comprende todos los genes conocidos o previstos, sino que se centra en los genomas completamente secuenciados de diversos organismos, los que presentan una comunidad de investigadores activos que estudian y aportan sistemáticamente información específica sobre uno o varios de ellos, o que se prevé un análisis intenso de su secuencia. Sus registros son el producto de la revisión, edición e integración automática de los datos procedentes de *RefSeq*, ciertas bases de datos creadas bajo modelos de cooperación y otras muchas disponibles en el NCBI. Cada registro recibe un número entero único, permanente y rastreable. Sus datos: nomenclatura, localización en el mapa, productos génicos y sus características, marcadores, fenotipos; así como enlaces a citas, secuencias, detalles de variaciones, mapas, expresiones, homólogos, dominios de proteínas e información existente en otras bases de datos, se actualizan en la medida en que aparece información nueva relacionada con el gen registrado.⁵¹

- *OMIM* (*Online Mendelian Inheritance in Man*)

[<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>]

OMIM, producida por *Johns Hopkins University School of Medicine*, es un compendio exhaustivo, autorizado y actualizado, de información sobre los genes humanos y los fenotipos que estos producen en su interacción con el medio. Los textos completos de sus reseñas contienen información sobre todos los trastornos hereditarios conocidos y sobre más de 12 000 genes. OMIM se centra en la relación fenotipo-genotipo. Se actualiza diariamente y posee abundantes enlaces, tanto a la información registrada por PubMed, como por otros recursos para la información genética.¹⁸

Una simple búsqueda por el nombre del gen causante de la ataxia espinocerebelosa tipo 2 (ATXN2), la más frecuente en Cuba, nos presenta una extensa información sobre su ubicación, estructura y función, así como sobre la genética molecular y otros muchos datos sobre esta enfermedad, no sólo directamente en los registros recuperados, sino también a partir de una extensa red de vínculos a la literatura citada. OMIM además permite la exploración de:

- *OMIM Gene Map* (<http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi>), donde con sólo dar un clic sobre sobre la *Location* del gen que aparece en la columna siguiente de la tabla que nos presenta el sistema (ATXN2), nos muestra mediante el *Map Wiewer* (<http://www.ncbi.nlm.nih.gov/mapview/static/MVstart.html>) del NCBI, que provee una visión integral de los mapas cromosómicos de más de 40 organismos diferentes, un mapa con el gen deseado sombreado (601517) y que con un nuevo clic sobre él nos informa el *Número MIM*, el *símbolo*, el *identificador para Gene* (otra base de datos), así como el *nombre* y el *código de la enfermedad* (en este caso, 183090) que produce (fig. 4).

- *OMIM Morbid Map* (<http://www.ncbi.nlm.nih.gov/Omim/getmorbid.cgi>). Al introducir el código de la enfermedad en su ventana de búsqueda y dar clic sobre el nombre de la enfermedad en la lista que nos devuelve, *Spinocerebellar ataxia-2*, obtenemos una amplia información sobre ella.

- *Online Mendelian Inheritance in Animals* (OMIA - <http://www.ncbi.nlm.nih.gov/omia>). Base de datos de genes, desórdenes hereditarios y rasgos en especies animales (otras especies además de ratones y humanos), con información textual y referencias, así como enlaces relevantes a registros de otras bases de datos del NCBI, como *PubMed* y *Gene*.⁵²
 - *UniGene* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>)
Cada registro contiene un conjunto de secuencias de transcritos que provienen del mismo *locus* de transcripción (expresión de genes o pseudogenes), e información sobre similitud de proteínas, expresión de genes, reactivos para la clonación de ADN complementario y localización genómica.⁵³
 - *HomoloGene* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene>)
Es un sistema para la detección automática de genes homólogos entre secuencias completas de genomas de organismos eucariotas.⁵⁴
 - *GeneTests* (<http://www.ncbi.nlm.nih.gov/sites/GeneTests/?db=GeneTests>). Es una base de datos reciente, ofrece información autorizada sobre pruebas genéticas y su uso en el diagnóstico, manejo y asesoramiento genético. *GeneTests* promueve el uso apropiado del servicio genético en el cuidado de pacientes y la toma de decisiones personales. Esta es una fuente de información médico-genética pública desarrollada por médicos, otros profesionales de la salud e investigadores. Está disponible sin costo alguno para las personas interesadas.⁵⁵
- Database of Major Histocompatibility Complex* (dbMHC-<http://www.ncbi.nlm.nih.gov/gv/mhc/main.cgi?cmd=init>). Una de las regiones del genoma humano más profundamente estudiada es la correspondiente al *Complejo mayor de histocompatibilidad* (MHC, por sus siglas en inglés). Este es, un grupo

de genes que abarcan aproximadamente de 4-6 megabases en el brazo corto del cromosoma 6. Los genes MHC, conocidos en humanos como genes de *Antígeno humano leucocitario* (HLA, siglas en inglés), son altamente polimórficos y codifican moléculas involucradas en la respuesta inmune. Provee una plataforma pública accesible para datos de ADN y datos clínicos relacionados con el MHC humano. La base de datos proporciona una plataforma donde la comunidad científica que estudia los HLAs puede presentar, editar, observar e intercambiar datos relacionados con el MHC humano. Consiste en un visor de configuración interactiva para HLA y genes relacionados, una base de datos de microsatélites de MHC, un sitio de interpretación de secuencias por tipaje basado en secuencias y una base de datos cebador/sonda.⁵⁶

- *Genes and Disease* (<http://www.ncbi.nlm.nih.gov/books/NBK22183/>). Ofrece información resumida sobre más de 80 desórdenes genéticos con discusiones sobre las mutaciones causantes y características clínicas, así como conexiones a otras bases de datos y organizaciones relacionadas. Se accede a la base de datos a través de *NCBI's Bookshelf*. *Genes and Disease* es una colección de artículos donde se analizan genes y las enfermedades que estos pueden causar. Los desórdenes genéticos se organizan según la parte del cuerpo que afectan. Como algunas enfermedades afectan varios sistemas del cuerpo, éstas aparecen en más de un capítulo. Con cada desorden genético, se discute la mutación(es) subyacente(es) junto con las características clínicas y se proporcionan enlaces a sitios Web relacionados.⁵⁷

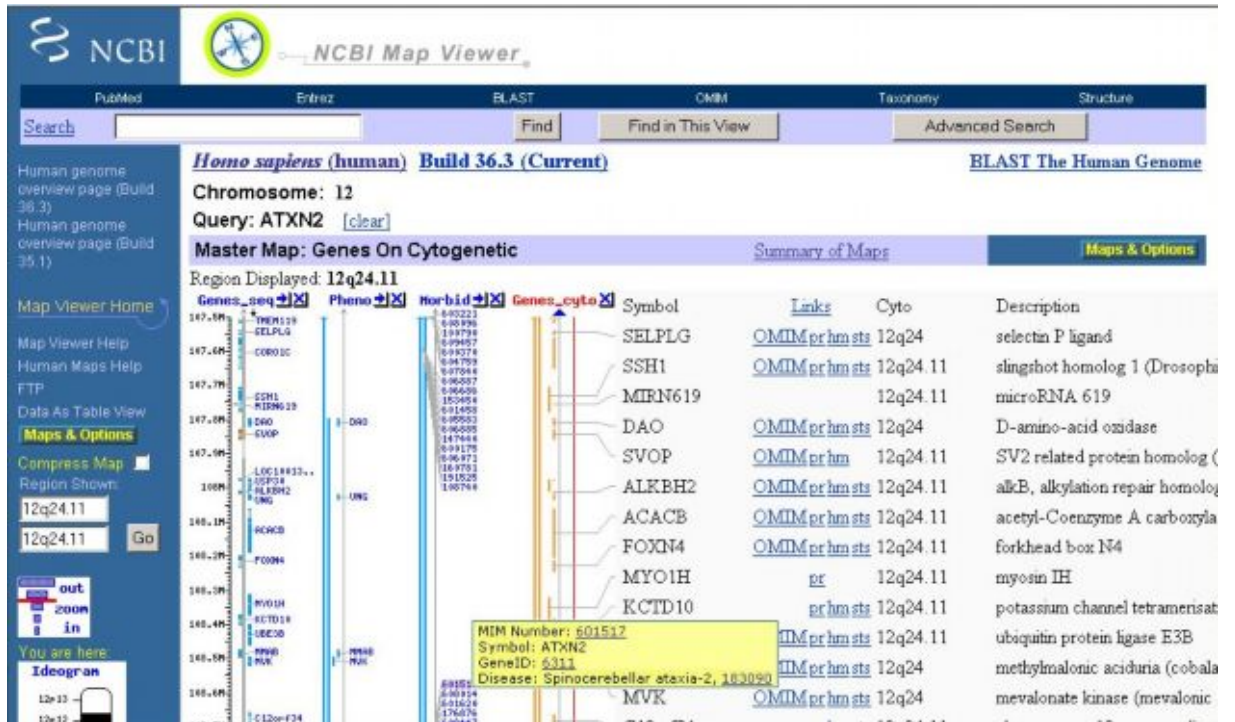
- *Cancer Chromosomes*

(<http://www.ncbi.nlm.nih.gov/sites/entrez?db=cancerchromosomes>)

Para catalogar los datos sobre las aberraciones cromosómicas en el cáncer, provenientes de la aplicación de novedosas técnicas citogenéticas, e integrar dichos datos con los mapas genómicos, se utilizó información de la base de datos del NCBI y el *National Cancer Institute*, denominada SKY/M-FISH & CGH (<http://www.ncbi.nlm.nih.gov/sky/>) y de *Cancer Chromosomes*. Se trata entonces de un esfuerzo por facilitar a los investigadores, la colocación y el análisis de datos citogenéticos de uso clínico o útiles a la investigación. Contiene una herramienta para el análisis de cariotipos que permite convertir automáticamente su forma corta, según la denominación en la *International System for Human Cytogenetics Nomenclature* (ISCN), en una representación detallada que incluye un ideograma. Posee, a su vez, una herramienta para comparar los perfiles obtenidos mediante CGH (*Comparative Genomic Hybridization*) de múltiples casos. *Cancer Chromosomes* integra actualmente información de las bases SKY/M-FISH & CGH, *Mitelman Database of Chromosome Aberrations in Cancer*

(<http://cgap.nci.nih.gov/Chromosomes/Mitelman>) y *Recurrent Chromosome Aberrations in Cancer* (<http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations>).⁵⁸⁻⁶⁰ Esta base de datos se retirará próximamente, sus contenidos estarán disponibles vía *dbVar*.

Fig. 4. Mapa genético, ATXN2.



Gene expression

Comprende bases de datos como:

- *Entrez GEO Profiles* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=geo>)

Es un repositorio de resultados de estudios realizados mediante las conocidas micromatrices, una tecnología que se emplea fundamentalmente para el análisis de la expresión y la variación genética. Ella se emplea no sólo en la investigación biomédica, sino también para el diagnóstico de múltiples enfermedades. Registra perfiles individuales de expresión génica y abundancia molecular, ensamblados a partir de los datos depositados en *Gene Expression Omnibus* (GEO-<http://www.ncbi.nlm.nih.gov/geo/>), un repositorio para el registro de los hallazgos de esta clase de estudios. Los perfiles de abundancia molecular se

basan en el perfil de frecuencia de generación de transcritos (moléculas de ARN) en una muestra determinada. Forma parte de las micromatrices de genes, con las que se estudia la expresión diferencial de los genes en varias condiciones sobre la base del análisis de la abundancia de sus ARN mensajeros, y sus resultados se colocan a disposición del público interesado en repositorios como GEO. El NCBI comenzó el proyecto GEO en 1999 en respuesta a la demanda creciente de un repositorio público para registrar los datos generados en experimentos realizados con micromatrices de alta densidad, una tecnología que permite, por ejemplo, el estudio simultáneo de la expresión de miles de genes. GEO presenta un diseño abierto y flexible que posibilita la colocación, almacenamiento y recuperación de diversos conjuntos de datos, como aquellos procedentes de experimentos de expresión de alcance amplio genes, hibridación génica y la determinación de anticuerpos.⁶¹

Database of Genotypes and Phenotypes (dbGaP - <http://www.ncbi.nlm.nih.gov/gap>) *Genotypes and Phenotypes* fue desarrollada para archivar y distribuir los resultados de los estudios sobre la interacción genotipo-fenotipo. Estos incluyen estudios de asociación a nivel de genoma completo, secuenciación médica, ensayos de diagnóstico molecular, así como asociación entre genotipos y rasgos no clínicos. La introducción de métodos comerciales de alta resolución para el genotipaje y la secuenciación ha provisto poderosas herramientas que permiten la generación de gran número de datos genotípicos necesarios para estos análisis.⁶²

- *Entrez Gene DataSets* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gds>)
Contiene conjuntos de datos de expresión génica y abundancia molecular reunidas a partir de la información almacenada en GEO.⁶³
- *GENSAT(Gene Expresión Nervous System Atlas)* [<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gensat>]
Es un proyecto dirigido a la creación de un mapa de la expresión de los genes en el sistema nervioso central de los ratones, mediante el uso de la hibridación *in situ* y de técnicas transgénicas.⁶⁴

Taxonomy

Incluye:

- *Entrez Taxonomy* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy>)
Contiene los nombres de todos los organismos representados con al menos una secuencia de nucleótidos o de proteína en las bases de datos de información genética del NCBI.⁶⁵

Genomes

El tercer sector ofrece, entre otros recursos:

- *Genome* (<http://www.ncbi.nlm.nih.gov/sites/genome>). *Genome* contiene secuencias y datos de mapas del genoma completo de alrededor de 100 especies o cepas. Contiene genomas totalmente secuenciados y aquellos que la secuenciación está en progreso. Los tres principales dominios de la vida están representados (bacteria, archaea y eukaryota), así como varios virus, fagos, viroides, plásmidos y organelos.⁶⁶
- *dbGaP (Genotypes and Phenotypes)* [<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>] Permite archivar y distribuir los resultados de los estudios que investigan la interacción genotipo-fenotipo. Comprende estudios amplios de asociación genotipo-fenotipo, de secuenciación, de diagnóstico molecular, así como de la relación entre el genotipo y algunos marcadores no clínicos.⁶⁷
- *Clone DB* (<http://www.ncbi.nlm.nih.gov/clone/>). *NCBI Clone DB* reemplaza y se expande sobre la antigua *NCBI Clone Registry*, establecida durante la era de los proyectos del genoma humano y del ratón. *Clone DB* representa un registro general de organismos eucarióticos y tipos de clones. Integra información sobre los clones y librerías, e incluye datos de secuenciación, posiciones en mapas y distribuidores de información. Esta consolidación provee una vista más detallada de datos asociados a clones y ayudará a los usuarios a tomar ventaja en las fuentes de clones. Los clones y librerías de clones toman varias formas y desempeñan diversos papeles en la investigación biológica. En términos generales, un clon puede considerarse un sistema de auto-replicación que contiene un fragmento de ADN de interés. Una librería de clones es una colección de clones que comparten propiedades definidas por el usuario, como el origen de la fuente de ADN o clase de vector.⁶⁸
- *Influenza Virus Resource* (<http://www.ncbi.nlm.nih.gov/genomes/FLU/>) Presenta datos obtenidos a partir del *NIAID Influenza Genome Sequencing Project* y del *GenBank*, combinados con herramientas para el análisis de la secuencia de la gripe, anotaciones y propuestas a *Genbank*. Además, suministra vínculos a otras fuentes que contienen secuencias de influenza, publicaciones e información general sobre los virus de la gripe.⁶⁹
- *Viral Genomes*

(<http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?>). Es un registro amplio de fuentes, que incluyen un breve resumen de la biología de los virus, conexiones a secuencias de genomas virales en *Entrez Genome*, e información sobre *Reference Sequences* (virales), es una colección de secuencias de referencia para miles de genomas virales.⁷⁰

- *Virus Variation*

(<http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/>). Es una extensión del *Influenza Virus Resource* a otros organismos. Posee una interfaz para descargar colecciones de secuencias de virus seleccionados; herramientas de análisis, incluye páginas de BLAST (*Basic Local Alignment Search Tool*); para virus-específicos y vías de anotación genómica (en desarrollo).⁷¹

El NCBI posee además un vasto sistema de herramientas para la captura de datos en sus bases de datos, la presentación de la información disponible en las bases de datos, la minería de datos..., BLAST, por ejemplo, posibilita el alineamiento de secuencias de ADN o proteínas, la comparación de secuencias y el hallazgo de las más similares y diversos *software*, agrupadas en el acápite denominado *Tools* (<http://www.ncbi.nlm.nih.gov/Tools/>) y otras secciones dirigidas a divulgar las investigaciones que realiza el centro, los *software* y bases de datos en desarrollo, los recursos para la educación disponibles para sus usuarios, así como un sitio FTP desde el que es posible descargar, tanto las actualizaciones de las bases de datos, como diversos programas para el manejo de información en las áreas temáticas que cubre el NCBI e información para contactar con la institución.⁷²

Consideraciones finales

El recorrido realizado por los principales recursos de información que ofrece el Web del NCBI durante la presente exposición no deja lugar a dudas sobre la abundancia y calidad de las bases de datos y otros medios para la consulta y análisis de la información biológica, biomédica, clínica y de salud que atesora el sitio referido. El Web del NCBI de los Estados Unidos es, por tanto, un espacio de consulta obligatoria y un punto de referencia extremadamente importante cuando se trata de la realización de investigaciones biomédicas y clínicas. La investigación molecular tiene en este sitio un verdadero palacio de la información.

La aceleración que experimenta la investigación molecular y su profundo impacto en el desarrollo y el ejercicio de la medicina a corto y mediano plazos nos llevará rápidamente a la era de la medicina personalizada. Desconocer sus avances es como «cerrar los ojos ante un alud que arrastrará tu casa en cuestión de minutos». La medicina y su ejercicio profesional ha cambiado dramáticamente en los últimos años y los cambios que se avecinan serán cada vez más increíbles y espectaculares.

Disponemos de excepcionales recursos de información totalmente gratuitos.

Lamentablemente, ellos con frecuencia también se encuentran sin uso alguno o con un uso totalmente deficiente. Es nuestra responsabilidad como profesionales de la información y la salud, informar y aconsejar a profesores, investigadores, gerentes, asistentes, tutores, estudiantes y otros muchos el uso de estos recursos. Pero corresponde a las autoridades, sobre todo en los campos de la docencia y la investigación, desarrollar los mecanismos que compulsen dicha consulta, porque estamos firmemente convencidos de que esa información a las que nos referimos es esencial para la formación de nuestros estudiantes y jóvenes profesionales, así como para la realización de investigaciones que puedan situarse a la «altura» de las exigencias actuales a escala internacional.

Finalmente es oportuno advertir que la presente contribución no se propuso en ningún momento describir todos o con todo detalle la abundante cantidad de recursos de información y otros medios disponibles en el Web del NCBI, sino tan sólo atraer la atención de los profesionales del sector de la salud sobre la disponibilidad de recursos de excelencia en el campo de la biología molecular y para el desarrollo de una medicina molecular personalizada.

Agradecimientos

A *Ernesto Galbán Rodríguez* y *Raúl Alejandro Valdés Pavón*, del Centro de Ingeniería Genética y Biotecnología, así como a *Luís Mederos Almaguer*, *Tania Cruz Mariño* y *Luís Velázquez Pérez*, del Centro para la Investigación y Rehabilitación de la Ataxias Hereditarias, de Holguín.

Referencias bibliográficas

1. Presentan tres nuevas secuencias del genoma humano. AL Día. 5 de noviembre de 2008. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=23383> [Consultado: 7 de febrero de 2011].
2. Última generación de secuenciadores, avvicina la era de la genómica personal. 24 de diciembre del 2008. Al Día. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=23799> [Consultado: 7 de febrero de 2011].
3. Variación genética humana, acontecimiento científico del 2007. Al Día. 20 de diciembre de 2007. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=20712> [Consultado: 7 de febrero de 2011].
4. Más del 10% del genoma es distinto entre las personas. Al Día. 22 de noviembre de 2006. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=17207> [Consultado: 8 de febrero de 2009].
5. Herencia biológica no está solo en el ácido desoxirribonucleico. Al Día. 18 de enero de 2009. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=23978> [Consultado: 8 de febrero de 2011].
6. Estudio revela cambios en el ADN del cerebro de suicidas. Al Día. 24 de octubre de 2008. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=23295> [Consultado: 8 de febrero de 2011].
7. Mapa genético de un hombre muestra que el ADN aún es un misterio. Al Día. 4 de septiembre de 2007. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=19743> [Consultado: 8 de febrero de 2011].
8. Última generación de secuenciadores, avvicina la era de la genómica personal. 24 de diciembre del 2008. Al Día. Disponible en: <http://www.sld.cu/servicios/aldia/view.php?idn=23799> [Consultado: 8 de febrero de 2011].
9. NCBI. NCBI at a glance. A history of discovery. From sequence to survival: the search to understand disease. Disponible en: http://www.ncbi.nlm.nih.gov/About/glance/story_discovery.html [Consultado: 27 de enero de 2011].
10. NCBI. NCBI at a glance. Our misión. Disponible en: <http://www.ncbi.nlm.nih.gov/About/glance/ourmission.html> [Consultado: 27 de enero de 2011].
11. Wikipedia. Biología molecular. Disponible en: http://es.wikipedia.org/wiki/BiologC3ADa_molecular [Consultado: 27 de enero de 2011].

12. Wikipedia. Centro Nacional para la Información Biotecnológica. Disponible en: <http://es.wikipedia.org/wiki/NCBI> [Consultado: 27 de enero de 2011].
13. Wikipedia. Entrez. Disponible en: <http://en.wikipedia.org/wiki/Entrez> [Consultado: 27 de enero de 2011].
14. NCBI. Site map. Disponible en: <http://www.ncbi.nlm.nih.gov/Sitemap/index.html> [Consultado: 4 de febrero de 2011].
15. Wikipedia. Base de datos biológica. Disponible en: http://es.wikipedia.org/wiki/Base_de_datos_biologica [Consultado: 7 de febrero de 2011].
16. NCBI. PubMed. PubMed Overview. Disponible en: <http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html> [Consultado: 6 de febrero de 2011].
17. NCBI. PMC Overview. Disponible en: <http://www.pubmedcentral.nih.gov/about/intro.html> [Consultado: 16 de febrero de 2011].
18. NCBI. OMIM - Online Mendelian Inheritance in Man. Disponible en: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM> [Consultado: 6 de febrero de 2011].
19. NCBI. Bookshelf. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=books> [Consultado: 6 de febrero de 2011].
20. NCBI. Entrez help. Nucleotides. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 4 de febrero de 2011].
21. National Human Genome Research Institute. Glosario. Disponible en: <http://www.um.es/bbmbi/AyudasDocentes/Glosario/Glosario/> [Consultado: 5 de febrero de 2011].
22. Kitts A, Sherry S. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. Disponible en: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.ch5> [Consultado: 5 de febrero de 2011].
23. NCBI. Probe. Database overview. Disponible en: <http://www.ncbi.nlm.nih.gov/genome/probe/doc/Overview.shtml> [Consultado: 6 de febrero de 2011].
24. Wikipedia Sequence-tagged site. Disponible en: http://en.wikipedia.org/wiki/Sequence_tagged_site [Consultado: 6 de febrero de 2011].
25. NCBI. The databases. UniSTS. Disponible en: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helpentrez.section.EntrezHelp.The_Data_bases [Consultado: 6 de febrero de 2011].

26. NCBI. Romiti M, Cooper P. Entrez help. PopSet. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 7 de febrero de 2011].
27. NCBI. PopSet. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=popset> [Consultado: 7 de febrero de 2011].
28. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2008 Jan; 36: D25-30.
29. NCBI. NCBI Resource guide. GenBank. General information. Disponible en: <http://www.ncbi.nlm.nih.gov/Sitemap/ResourceGuide.html#Overview>
30. Pruitt K, Tatusova T, Maglott D. The Reference Sequence (RefSeq) Project. Disponible en: <http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/handbook/ch18.pdf> [Consultado: 8 de febrero de 2011].
31. NCBI. RefSeq. Disponible en: <http://www.ncbi.nlm.nih.gov/RefSeq/> [Consultado: 8 de febrero de 2011].
32. NCBI. dbEST. Disponible en: <http://www.ncbi.nlm.nih.gov/dbEST/> [Consultado: 8 de febrero de 2011].
33. NCBI. dbGSS. Disponible en: <http://www.ncbi.nlm.nih.gov/dbSTS/> [Consultado: 8 de febrero de 2011].
34. NCBI. dbSTS: database of «Sequence Tagged Sites». Disponible en: <http://www.ncbi.nlm.nih.gov/dbSTS/> [Consultado: 8 de febrero de 2009].
35. NCBI. Protein. Disponible en: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein> [Consultado: 9 de febrero de 2011].
36. Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, et al. The National Center for Biotechnology Information's Protein Clusters Database. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686591/pdf/gkn734.pdf> [Consultado: 4 de mayo de 2011].
37. NCBI. Conserved Domains. Disponible en: <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> [Consultado: 9 de febrero de 2011].
38. NCBI. Protein Clusters. Disponible en: <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> [Consultado: 9 de febrero de 2011].
39. National Institute of Allergy and Infectious Diseases. HIV-1, Human Protein Interaction Database. Disponible en: <http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/#overview>. [Consultado: 3 de mayo de 2011].

40. NCBI. Molecular Modeling Database (MMDB). Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure> [Consultado: 28 de abril de 2011].
41. NCBI. Romiti M, Cooper P. Entrez help. PubChem. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 10 de febrero de 2011].
42. NCBI. PubChem. Disponible en: <http://pubchem.ncbi.nlm.nih.gov/> [Consultado: 10 de febrero de 2009].
43. Wikipedia. PubChem. Disponible en: <http://en.wikipedia.org/wiki/Pubchem> [Consultado: 10 de febrero de 2011].
44. NCBI. Romiti M, Cooper P. Entrez help. PubChem Compound. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 10 de febrero de 2011].
45. NCBI. PubChem. Disponible en: <http://pubchem.ncbi.nlm.nih.gov/> [Consultado: 10 de febrero de 2011].
46. NCBI. Romiti M, Cooper P. Entrez help. Pubchem BioAssay. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 10 de febrero de 2011].
47. NCBI. PubChem. Disponible en: <http://pubchem.ncbi.nlm.nih.gov/> [Consultado: 10 de febrero de 2011].
48. NCBI. BioSystems. Disponible en: <http://www.ncbi.nlm.nih.gov/biosystems/> [Consultado: 4 de mayo de 2011].
49. Romite M, Cooper P. Entrez help. BioSample. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 5 de mayo de 2011].
50. NCBI. Overview of structural variation. Disponible en: <http://www.ncbi.nlm.nih.gov/dbvar/content/overview/> [Consultado: 5 de mayo de 2011].
51. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2005 January 1; 33(Database Issue): D54D58. Disponible en: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=15608257> [Consultado: 10 de febrero de 2011].
52. NCBI. Online Mendelian Inheritance in Animals (OMIA). Disponible en: <http://www.ncbi.nlm.nih.gov/omia> [Consultado: 5 de mayo de 2011].
53. NCBI. Unigene. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene> [Consultado: 11 de febrero de 2011].
54. NCBI. HomoloGene. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene> [Consultado: 11 de febrero de 2011].
55. NCBI. About GeneTests. Disponible en: <http://www.ncbi.nlm.nih.gov/projects/GeneTests/static/about/aboutindex.shtml> [Consultado: 5 de mayo de 2011].

56. Kitts A, Feolo M. The Major Histocompatibility Complex Database, dbMHC. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK21080/> [Consultado: 5 de mayo de 2011].
57. NCBI. Genes and Disease. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK22183/> [Consultado: 5 de mayo de 2011].
58. NCBI. Cancer Chromosomes. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=cancerchromosomes> [Consultado: 12 de febrero de 2011].
59. NCBI. Romiti M, Cooper P. Entrez help. Cancer Chromosomes. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 12 de febrero de 2011].
60. Knutsen T, Gobu V, Knaus R, Padilla-Nash H, Augustus M, Strausberg RL, et.al. The interactive online SKY/M-FISH & CGH database and the Entrez cancer chromosomes search database: linkage of chromosomal aberrations with the genome sequence. *Genes Chromosomes Cancer*. 2005 Sep;44(1):52-64. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/15934046?dopt=Abstract> [Consultado: 12 de febrero de 2011].
61. Edgar R, Lash A. The Gene Expression Omnibus (GEO): A Gene Expression and Hybridization Repository. Disponible en: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.ch6> [Consultado: 11 de febrero de 2011].
- 62 NCBI. dbGaP Overview. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK22183/> [Consultado: 4 de mayo de 2011].
63. NCBI. Entrez Gene DataSets. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gds> [Consultado: 12 de febrero de 2011].
64. NCBI. GENSAT Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gensat> [Consultado: 12 de febrero de 2011].
65. NCBI. Taxonomy. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy> [Consultado: 12 de febrero de 2011].
66. Romiti M, Cooper P. Entrez help. Genome. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 6 de mayo de 2011].
67. Romiti M, Cooper P. Entrez help. dbGap. Genome. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK3837/> [Consultado: 6 de mayo de 2011].
68. NCBI. Clone DB. Disponible en: <http://www.ncbi.nlm.nih.gov/clone/content/overview/> [Consultado: 5 de mayo de 2011].
69. NCBI. Influenza Virus Resource. <http://www.ncbi.nlm.nih.gov/genomes/FLU/> [Consultado: 5 de mayo de 2011].

70. NCBI. All resources. Viral Genomes. Disponible en: <http://www.ncbi.nlm.nih.gov/guide/all/> [Consultado: 12 de febrero de 2011].
71. NCBI. All resources. Virus Variation. Disponible en: <http://www.ncbi.nlm.nih.gov/guide/all/> [Consultado: 12 de febrero de 2011].
72. Wikipedia. BLAST. Disponible en: <http://es.wikipedia.org/wiki/BLAST> [Consultado: 12 de febrero de 2009].

Anexos.

Anexo 1. Glosario.

Ácidos nucleicos: Biomoléculas formadas por polímeros de nucleótidos, o polinucleótidos. Están presentes en todas las células y constituye la base material de la herencia. Existen dos tipos: el ácido desoxirribonucleico (ADN) y el ácido ribonucleico (ARN).¹

ADN (ácido desoxirribonucleico): Un ácido nucleico compuesto de dos cadenas de polinucleótidos, antiparalelas y complementarias entre ellas, que se disponen alrededor de un eje central formando una doble hélice, capaz de autorreplicarse, y que posee la información necesaria para, entre otras funciones, realizar la síntesis de las proteínas en la célula. Su unidad básica, el nucleótido.¹

ADN recombinante: Molécula de ADN formado por recombinación de fragmentos de ADN de orígenes diferentes.¹

Adaptación (del latín *adaptare* = acomodar): Tendencia de un organismo a «adecuarse» a su medio ambiente. Uno de los principales puntos de la teoría de la evolución por selección natural de Charles Darwin: los organismos se adaptan a su medio ambiente. Aquellos organismos mejor adaptados tendrán mayor probabilidad de sobrevivir y transmitir sus genes a la siguiente generación.¹

Alelo: Una de las variantes o alternativas de un gen en un locus. Diferentes alelos de un gen producen variaciones en las características hereditarias como el color del cabello o ciertas enfermedades.² Es uno o más de los estados alternativos de un gen. Se heredan en forma separada de cada padre por ejemplo, en el locus para el color de ojos puede haber un alelo para ojos azules o uno para ojos negros.³

Aminoácido: Es la unidad estructural básica de la proteína. Conforman un grupo de 20 tipos diferentes de moléculas pequeñas que se unen entre ellas por medio de enlaces peptídicos.²

Anticodón: Secuencia de 3 nucleótidos en el ARN de transferencia que se emparejan de forma complementaria con un codón específico del ARN mensajero durante la síntesis proteica para colocar un aminoácido concreto en la cadena polipeptídica.³

ARN (ácido ribonucleico): El componente químico que resulta de la transcripción del ADN. Su nucleótido, consiste de una molécula del azúcar ribosa, un grupo fosfato, y una de estas cuatro bases nitrogenadas: adenina, uracilo, citosina y guanina.²

ARN de transferencia (del latín *transferre* = transportar, der. de *ferre*= llevar): Ácido ribonucleico, pequeño, de una sola cadena, que se enlaza al aminoácido correspondiente y lo lleva al codón apropiado en el ARNm.²

ARN mensajero (ARNm): Molécula de ARN que representa una copia en negativo de las secuencias de aminoácidos de un gen. Previo a su conformación, se eliminan las secuencias no codificantes (intrones) del gen.²

ARN ribosómico: Es un componente estructural de los ribosomas.²

Autosoma (del griego *autos* = uno mismo; *soma* = cuerpo): Cromosoma que no interviene en la determinación del sexo. El genoma humano diploide consiste en 46 cromosomas compuestos de 22 pares de autosomas y un par de cromosomas sexuales (X e Y).¹

Autosoma dominante: Un gen en uno de los autosomas, que si está presente, producirá casi siempre una enfermedad o rasgo específico. La probabilidad de pasar el gen (y por lo tanto la enfermedad) a los hijos, es de 50:50 en cada embarazo.²

Biomoléculas: Elementos arquitectónicos básicos de los seres vivos. Las biomoléculas inorgánicas son sobretodo agua, sales minerales y gases como oxígeno y dióxido de carbono. Los grupos de compuestos orgánicos exclusivos de los seres vivos son cuatro: carbohidratos, lípidos, proteínas y ácidos nucleicos.¹

Cariotipo: Se refiere al ordenamiento de los cromosomas en una célula. Comprende número, tamaño y forma de cada cromosoma.⁴ En el caso de los humanos es 46XX en el sexo femenino y 46XY en el sexo masculino.²

Código del triplete: Sucesión de tres bases de tres nucleótidos en la molécula de ADN que cifra un aminoácido.

Código genético: Conjunto de 64 codones que codifican para los diferentes aminoácidos durante la síntesis de proteínas.⁵

Codón: Secuencia de tres nucleótidos consecutivos en un gen o molécula de ARNm, que especifica la posición de un aminoácido en una proteína.¹

Cromosoma (del griego *chroma* = color; *soma* = cuerpo): Son corpúsculos intracelulares alargados que constan de ADN, asociado con proteínas, y constituido por una serie lineal de unidades funcionales conocidas como genes.¹ Los cromosomas se encuentran en el núcleo de las células y diferentes especies tienen diferente número y morfología de cromosomas. Los humanos tenemos 23 pares de cromosomas, 46 en total: 44 autosomas y 2 cromosomas sexuales. Cada uno de los progenitores aporta un cromosoma a cada par, de manera que los hijos reciben la mitad de los cromosomas de la madre y la mitad del padre.²

Deleción: Un tipo especial de mutación que consiste en la pérdida de un fragmento de ADN de un cromosoma. La deleción de un gen o de parte de un gen puede ocasionar una enfermedad o una anomalía.²

Diploide (del griego *di*= doble; *ploion*= nave): Organismo o fase nuclear que tiene los dos juegos de cromosomas. Número cigótico de cromosomas (2n), por oposición al número gamético (n) o haploide.¹

Desórdenes poligénicos: Resultan de la acción combinada de los alelos de más de un gen por ejemplo, enfermedades cardíacas, diabetes y ciertos tipos de cáncer. Si bien estos desórdenes son heredables, dependen de la presencia simultánea de muchos alelos.¹

Dominante: Es la condición por la cual un miembro de un par alélico se manifiesta en el fenotipo de un individuo excluyendo la expresión del otro alelo.²

Dominio de proteínas: Porción de una proteína con una función propia. La combinación de las funciones de los dominios de una proteína determina su función global.¹ Son unidades estructurales compactas y estables en una proteína, formadas por segmentos de una cadena peptídica que se pliegan de forma independiente, con una estructura y función distinguible de otras regiones de la proteína. Los dominios pueden considerarse como unidades elementales de la estructura y evolución de las proteínas, que son capaces de plegarse y de funcionar de una manera relativamente autónoma.⁶

Duplicación: Un tipo particular de mutación que consiste en la producción de una o más copias de un fragmento de ADN, que puede ser un gen o un cromosoma completo.²

Enzima (del griego *en* = en; *zyme* = levadura): Catalizador biológico de naturaleza proteica, que media y promueve un proceso químico sin que ella se altere o destruya.³

Expresión: En genética, proceso por el cual la información codificada en los genes se convierte en estructuras operacionales presentes en la célula. Desde una perspectiva epigenética, pudiera decirse que es un proceso altamente específico en el cual un gen se «enciende» en un momento determinado y comienza la producción de su proteína.²

Expresión génica: El proceso por el cual la información codificada en los genes se convierte en estructuras operacionales presentes en las células. Los genes expresados incluyen a aquellos que fueron transcritos al ARNm y luego se tradujeron en proteínas y aquellos que fueron transcritos al ARN pero no se tradujeron en proteínas.¹

Eucariotas: Se denomina eucariotas a todas las células (y organismos) que tienen su material hereditario fundamental (su información genética) encerrado en una doble membrana, la envoltura nuclear, que delimita el núcleo celular.⁷

Frecuencia fenotípica: Porcentaje de individuos que expresan una cualidad del fenotipo que se estudia en relación con el total de individuos que conforman la población objeto de análisis.⁸

Frecuencia genotípica: Porcentaje o proporción en que aparecen cada uno de los genotipos generados, en relación con el total de genotipos (que será igual al total de individuos) estudiados.⁸

Frecuencia génica: Proporción en que un alelo se encuentra presente en relación con el número total de alelos de la población en estudio para un locus determinado.⁸

Fenotipo: Rasgos o características visibles de un organismo, por ejemplo, el color del cabello, el peso o la presencia o ausencia de una enfermedad. Los rasgos fenotípicos no son necesariamente genéticos.²

Gen: La unidad física y funcional de la herencia, que se pasa de padres a hijos. Los genes están compuestos por ADN y la mayoría de ellos contiene la información para elaborar una proteína específica.²

Genoma: Todo el ADN contenido en un organismo o célula, que incluye tanto los cromosomas dentro del núcleo como el ADN en las mitocondrias,¹ o los cloroplastos en las plantas.

Genotipo: La identidad genética de un individuo que no se muestra como características externas.²

Hibridación: Apareamiento de bases de hebras únicas de ADN o ARN,² según complementariedad de bases.

Hibridación «*in situ*» (del latín «en el lugar»): El uso de sondas de ADN o ARN para detectar la presencia de secuencias de ADN complementarias a ellas, en bacterias clonadas o células eucariotas cultivadas.¹ Se emplea frecuentemente en el ejercicio de la práctica clínica para detectar ciertas enfermedades hereditarias o mutaciones que producen enfermedades de especial interés.

Homólogo: Dos genes de dos especies distintas son homólogos si se derivan de un mismo gen ancestral. La secuencia de nucleótidos de un gen se transmite de padres a hijos. Cuando examinamos el genoma de dos especies esperamos encontrar los genes equivalentes en ambas, con una secuencia algo diferente, más cuanto más remoto en el tiempo se ubica el antepasado común. La expresión *homología de secuencia* se refiere a la correspondencia entre las cadenas nucleotídicas de esos dos genes, que es precisamente la que permite reconocer que son homólogos.⁹

Inserción: Un tipo de anomalía cromosómica en el cual un segmento de ADN se inserta en un lugar diferente, y en algunos casos alterando la estructura y función normal de un gen.²

Locus: El lugar del cromosoma donde está localizado un gen específico, es la dirección física del gen. El plural es «loci».²

Macromoléculas: Son moléculas que tienen una masa molecular elevada, formadas por un gran número de átomos. Generalmente se pueden describir como la repetición de una o unas pocas unidades mínimas o monómeros, formando los polímeros.

Mapa citogenético: Configuración de las bandas coloreadas de los cromosomas observada en el microscopio óptico después de su tinción.¹

Mapa cromosómico: Diagrama que muestra las posiciones relativas en un cromosoma de los *loci* de los genes, basado en la frecuencia en que se heredan conjuntamente. Las distancias se miden en centimorgans.¹

Mapa físico: Mapa de la localización de las marcas identificables en el ADN -por ejemplo, genes, sitios de corte de enzimas de restricción, bandas- que se confecciona prescindiendo de los fenómenos hereditarios. Las distancias se miden en pares de bases. En el genoma humano, el mapa físico de menor resolución lo constituye los patrones de bandeo de los 24 cromosomas diferentes (22 autosómicos, el X y el Y). El de mayor resolución lo constituirá la secuencia completa de nucleótidos de los cromosomas.¹

Mapa genético: Diagrama descriptivo de los genes en cada cromosoma.¹

Mapeo genético: Determinación de la posición relativa de los genes en la molécula de ADN (plásmido o cromosoma) y las distancias (unidades de «unión») entre ellos.¹

Marcador (del inglés *marker*; pob. del italiano *marcare* = «señalar una cosa para que se distinga de otra»): Una posición física identificable en un cromosoma cuya herencia puede seguirse (por ejemplo, un gen o un sitio que corta una enzima de restricción). Los marcadores pueden ser una región del ADN que se expresa (gen) o un segmento de ADN que no se conoce que codifica pero que se puede seguir su manera de heredarse.¹

Mapa génico: Mapa basado en las posiciones relativas de los genes en un cromosoma y de la distancia entre ellos.²

Marcador genético: Un segmento de ADN con una ubicación física identificable en un cromosoma y cuya herencia se puede rastrear. Un marcador puede ser un gen, o puede ser alguna sección del ADN sin función conocida. Debido a que los segmentos del ADN que se encuentran contiguos en un cromosoma tienden a heredarse juntos, los marcadores se utilizan a menudo como formas indirectas de rastrear el patrón hereditario de un gen que todavía no ha sido identificado, pero cuya ubicación aproximada se conoce.²

Microsatélite: Son pequeñas regiones de ADN que contienen múltiples copias de secuencias repetitivas cortas -1-5 pares de bases, en grupos de 10-50 copias- y que se emplean como marcadores genéticos para rastrear la herencia familiar o mapear enfermedades en el genoma.² Prácticamente, la mitad del ADN eucariota corresponde a secuencias nucleotídicas repetidas. Estas secuencias se clasifican en: satélites, minisatélites y microsatélites, según el grado de repetición que presentan.¹⁰

Monómero (del griego *mono*, uno y *meros*, parte): Se refiere a una molécula de pequeña masa molecular que unida a otros monómeros, a veces cientos o miles, por medio de enlaces químicos, generalmente covalentes, forman macromoléculas llamadas polímeros. Además son unidades básicas o moléculas orgánicas relativamente simples. Los aminoácidos son los monómeros de las proteínas; los nucleótidos de los ácidos nucleicos; los monosacáridos de los glúcidos y los ácidos grasos y el glicerol de los lípidos.¹¹

Mutación (del latín *mutare* = cambiar): Alteración del material genético que puede involucrar uno o pocos nucleótidos o alterar el cariotipo.⁵

Núcleo: La estructura celular central que contiene los cromosomas.²

Nucleótido: Uno de los componentes estructurales o unidades constituyentes del ADN o del ARN. Un nucleótido consta de una base (adenina, timina, guanina, uracilo o citosina), más una molécula de azúcar y una de ácido fosfórico.²

Par de bases: Dos bases que forman un «peldaño de la escalera del ADN». Un nucleótido del ADN está compuesto por una molécula de azúcar, una molécula de ácido fosfórico y una base nitrogenada. Las bases son las «letras» del código genético. En el ADN, las letras del código son A, T, G y C, que corresponden a adenina, timina, guanina y citosina respectivamente. Al formar los pares, la adenina siempre se une a timina y la guanina siempre se une a citosina.²

Péptido: Un péptido está formado por 2 o más aminoácidos ensamblados por un enlace peptídico.²

Polímero (del griego *polys* = muchos, *meros* = parte): Son macromoléculas (generalmente orgánicas) formadas por la unión de moléculas más pequeñas llamadas monómeros. Una molécula compuesta por muchas subunidades idénticas o similares.³

Polimorfismos de nucleótido simple: Son variaciones comunes de una sola base que ocurren en el ADN humano con una frecuencia aproximada de 1 en cada 1000 bases. Estas variaciones se pueden emplear para rastrear patrones de herencia familiar.²

Proteínas: Polímeros de aminoácidos ligados por enlaces peptídicos que desempeñan funciones vitales en las células como la catálisis de reacciones bioquímicas, transporte de sustancias, de señalización, etcétera.⁵

Recesivo: Es el término aplicado al miembro de un par alélico imposibilitado de manifestarse cuando el alelo dominante está presente. Para que este alelo se observe en el fenotipo debe estar presente en doble dosis, proveniente uno de la madre y otro del padre.²

Replicación (del ¿latín? *replere* = rellenar): Proceso anterior a la división celular por medio del cual se duplica el ADN. Es un proceso por el que una molécula de ADN origina otra idéntica a la preexistente.¹

Ribosomas: Pequeñas organelas, donde se produce la síntesis proteica.²

Secuencia complementaria: Secuencia de bases en una molécula de ADN que puede formar una doble hebra al aparear las bases de otra. Por ejemplo, la secuencia complementaria de G-T-A-C es C-A-T-G.¹

Secuencia de ADN: Orden de encadenamiento de las bases nitrogenadas de los nucleótidos que constituyen el ADN y que cifra toda la información genética. Cuando es codificante (exón), define el orden de los aminoácidos que forman la proteína correspondiente.¹

Secuenciar: Determinación del orden de los nucleótidos en el ADN o ARN o el orden de los aminoácidos en las proteínas.¹

Sonda (del francés *sonde*, abreviación del anglosajón *sundline*, en su acepción de elemento exploratorio): Molécula de ADN o ARN monocatenario con una secuencia de bases específica, marcada radiactivamente o inmunológicamente, que se utiliza para detectar secuencias de bases complementarias por hibridación.¹

Sustitución: Reemplazo de un nucleótido en una secuencia del ADN por otro nucleótido, o reemplazo de un aminoácido en una proteína, por otro aminoácido.²

Terapia génica: Conjunto de los procesos destinados a la introducción *in situ* o *in vivo* de un gen normal en células, germinales o somáticas, en las que el mismo gen, anormal, provoca una deficiencia funcional, origen de una enfermedad, o la de un gen codificador de una proteína, por ejemplo, con una acción antitumoral en las células cancerosas, o antivírica en células infectadas por un virus patógeno.¹

Transcripción: Síntesis de moléculas de ARN a partir de moldes de moléculas de ADN según complementación de bases.⁵⁻¹²

Traducción de la información genética: Paso de la información genética contenida en el ARNm a las proteínas, mediante la síntesis de estas últimas, según el código genético.⁵⁻¹²

Referencias bibliográficas

1. Zamudio T. Regulación jurídica de las biotecnologías. Glosario. Disponible en: <http://www.biotech.bioetica.org/glosario.htm> [Consultado: 22 de febrero de 2009].
2. National Human Genome Research Institute. Glosario. Disponible en: <http://www.um.es/bbmbi/AyudasDocentes/Glosario/Glosario/> [Consultado: 22 de febrero de 2009].
3. Universidad Nacional del Nordeste. Argentina. Hipertextos del área de la biología. Síntesis proteica. Disponible en: <http://www.biologia.edu.ar/adn/adntema2.htm#Glosario> [Consultado: 22 de febrero de 2009].
4. Wikipedia. Cariotipo. Disponible en: <http://es.wikipedia.org/wiki/Cariotipo> [Consultado: 22 de febrero de 2009].
5. De Robertis EMF, Hib J, Ponzio R. Biología Celular y Molecular. Buenos Aires: El Ateneo; 1998.
6. Temas de bioquímica. Disponible en: <http://temasdebioquimica.wordpress.com/2008/10/12/estructuras-supersecundarias-motivos-y-dominios/> [Consultado: 22 de febrero de 2009].

7. Wikipedia. Célula eucariota. Disponible en:
http://es.wikipedia.org/wiki/C%C3%A9lula_eucariota [Consultado: 22 de febrero de 2009].
8. Lantigua Cruz A. Introducción a la genética médica. La Habana: Editorial Ciencias Médicas. 2004.
9. Wikipedia. Homólogo. Disponible en:
[http://es.wikipedia.org/wiki/Homolog%C3%ADa_\(biolog%C3%ADa\)](http://es.wikipedia.org/wiki/Homolog%C3%ADa_(biolog%C3%ADa)) [Consultado: 22 de febrero de 2009].
10. Universidad Nacional del Nordeste. Argentina. Hipertextos del área de la biología. Control de la expresión de los genes eucariotas. Disponible en:
<http://www.biologia.edu.ar/adn/adntema4.htm> [Consultado: 22 de febrero de 2009].
11. Wikipedia. Monómero. Disponible en: <http://es.wikipedia.org/wiki/Mon%C3%B3mero> [Consultado: 22 de febrero de 2009].
12. Zamudio T. La transcripción y la traducción. Disponible en:
<http://www.biotech.bioetica.org/clase1-14.htm> [Consultado: 22 de febrero de 2009].

¿Cómo citar este documento?

Cita (Vancouver): Cañedo Andalia R, Rodríguez Labrada R, Vázquez Mojena Y. Centro Nacional para la Información Biotecnológica de los Estados Unidos: un palacio de la información para la medicina molecular. En: Cañedo Andalia R, Rodríguez Labrada R, Fernández Valdés MM, Zayas Mujica R, Nodarse Rodríguez M, Sánchez Tarragó N, *et al.* Lecturas avanzadas para la alfabetización informacional en salud. Holguín: Centro Provincial de Información de Ciencias Médicas. Universidad de Ciencias Médicas de Holguín; 2011. [citado día mes año]. Disponible en: http://www.hlg.sld.cu/sitios/CPICM/index.php?option=com_jdownloads&Itemid=87&view=viewcategory&catid=5