



**APLICACIONES DE LAS TÉCNICAS
DE BIOLOGIA MOLECULAR EN
MEDICINA FORENSE**

GENÉTICA FORENSE

La **genética forense** se puede definir como la parte de la genética y de la medicina legal y forense que analiza la variabilidad genética humana para resolver problemas de identificación.

Los principales problemas que pretende resolver la genética forense son:

- **Estudios de paternidad**
- **Identificación de restos cadavéricos**
- **Investigaciones criminalísticas a partir de muestras biológicas humanas (sangre, pelo, piel, semen, saliva...)**

- El avance de las técnicas de biología molecular, hace posible analizar directamente el ADN
- La **genética forense** evoluciona rápidamente hacia el análisis de **regiones polimórficas del ADN** (regiones que presentan variabilidad entre los distintos individuos → **variación genética interindividual**)

ORGANIZACIÓN DEL ADN HUMANO

Para entender las aplicaciones de la biología molecular en la medicina legal y forense, es necesario saber primero cómo se organiza el ADN humano.

En el genoma humano:

- Solo una proporción mínima (**< 4%**) es **ADN codificante**, es decir, codifica para proteínas.
- El resto es **ADN no codificante**.
 - El 25% del ADN total lo constituyen secuencias relacionadas con genes, como **intrones** y **secuencias reguladoras**
 - La gran mayoría del ADN humano (alrededor del 70%) es ADN **extragénico**, sin función conocida.



ORGANIZACIÓN DEL ADN HUMANO

Para analizar la organización del ADN, utilizaremos el criterio de repetitividad, que va a permitir definir varios tipos de secuencias:

- **ADN de secuencia única**, tanto **codificante** como **no codificante**. Constituye aproximadamente el 50% del ADN humano.
- **ADN repetitivo**. Constituye el otro 50% del ADN humano. Lo integran secuencias que se repiten en número variable, desde unas pocas hasta miles de repeticiones.

ADN repetitivo codificante.

Supone una parte pequeña del ADN repetitivo. Engloba unos pocos genes que se repiten en tándem (como los genes que codifican el ARN ribosómico) y algunas familias multigénicas.

ADN repetitivo no codificante.

Supone la mayor parte del ADN repetitivo. Hay dos tipos

ADN repetido en tándem: secuencias que se repiten en una copia a continuación de otra, con una localización concreta (locus específico).

ADN repetido disperso por todo el genoma: formado por secuencias de tamaño variable que se repiten miles o cientos de miles de veces, pero no en tándem, sino dispersas a lo largo de todo el genoma.

POLIMORFISMOS

Se definen como variaciones en la secuencia de un locus específico, con una incidencia en la población superior al 1%.

Cada una de las secuencias alternativas del mismo locus se denomina **alelo**. El conjunto de todos los alelos que porta un individuo constituye su genotipo.

- Los polimorfismos en loci codificantes se traducen en diferentes fenotipos.
- Los polimorfismos en loci no codificantes no tienen reflejo en el fenotipo del individuo, pero **contribuyen a la variabilidad genética interindividual, por lo que tienen una gran utilidad en genética forense.**
- Los polimorfismos pueden ser de dos tipos:
 - De longitud
 - De secuencia.

TIPOS DE POLIMORFISMOS

Los **polimorfismos de longitud** son variaciones en la longitud de la secuencia de un locus debido a deleciones o inserciones de uno o más nucleótidos.

- En genética forense tienen gran importancia los polimorfismos de longitud de ciertos minisatélites y microsatélites.
- Estas regiones de ADN son hipervariables y muy polimórficas
- **Su polimorfismo consiste en el número de veces que se repite la secuencia básica.**
- Este polimorfismo de minisatélites y microsatélites **permite diferenciar un individuo de otro, por lo que se habla de marcadores genéticos.**
- Como los distintos alelos de un marcador se diferencian por el número de copias de la secuencia básica que contienen, esta información se utiliza para nombrarlos.
- Cada alelo se denomina con el nombre del marcador seguido del número de repeticiones de la secuencia básica.
 - Por ejemplo, TH01-7 corresponde al alelo del marcador TH01 que contiene 7 copias de la secuencia básica.

TIPOS DE POLIMORFISMOS

Polimorfismos de longitud

Repeticiones en tándem de número variable, VNTR

(variable number tandem repeats) → Son minisatélites polimórficos hipervariables con un tamaño relativamente grande, entre 500 y 10 000 nucleótidos.

Repeticiones en tándem cortas, STR

(short tandem repeats) → Son microsatélites polimórficos (varios alelos), con una secuencia básica de 4-6 nucleótidos que se repite un número de veces diferente en cada alelo, normalmente no superior a 50. El tamaño final del alelo más largo es inferior a 500 nucleótidos.

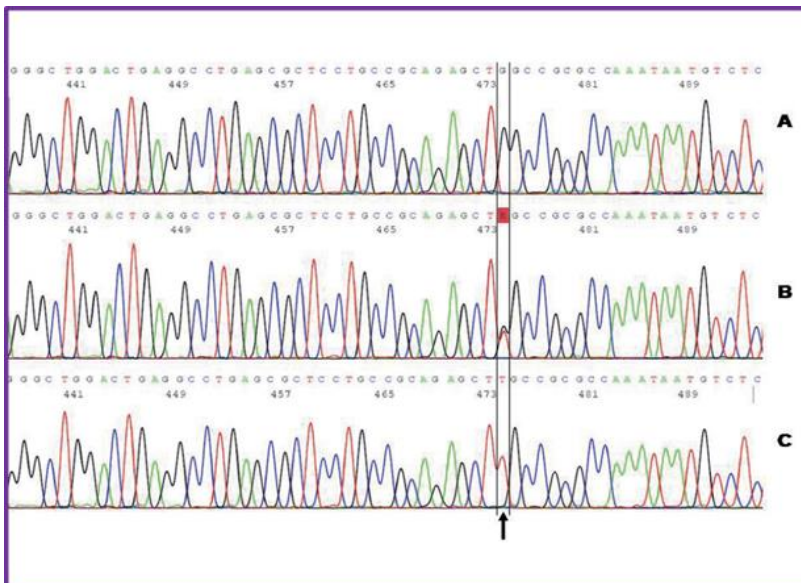
- Los individuos **nos diferenciamos por el número de repeticiones de esas secuencias**, así dos seres humanos no relacionados es poco probable que tengan en común los mismos marcadores genéticos.
- Cada individuo presenta dos valores (iguales o distintos) para cada uno de los microsatélites o STRs. un valor procede de la madre y otro del padre.
- Al conjunto de polimorfismos específico de cada persona se le conoce como **perfil genético**.

TIPOS DE POLIMORFISMOS

Los **polimorfismos de secuencia** consisten en variaciones de uno o más nucleótidos en la secuencia de un locus.

Los más interesantes para la genética forense son los **polimorfismos de nucleótido simple o único, SNP**, (single nucleotide polymorphism), consistentes en variaciones puntuales de un único nucleótido.

Habitualmente son polimorfismos dialélicos y su frecuencia en el genoma humano es aproximadamente de 1 por cada 1000 nucleótidos.



- **individuo A** es homocigoto para el alelo G
- **individuo B** es heterocigoto, portando ambos alelos G y T;
- **individuo C** es homocigoto para el alelo T

HUELLA GENÉTICA

Se puede definir la **huella genética** como la combinación de alelos de varios marcadores genéticos que posee un individuo.

Sus aplicaciones son diversas:

- Medicina forense,
- Pruebas de paternidad,
- Para identificar a personas sin documentación
- Estudios en la compatibilidad en la donación de órganos
- Incluso generar hipótesis sobre las migraciones de los seres humanos en la prehistoria.
- Investigación para detectar y combatir enfermedades tales como dolencias congénitas, tumores, distrofia.

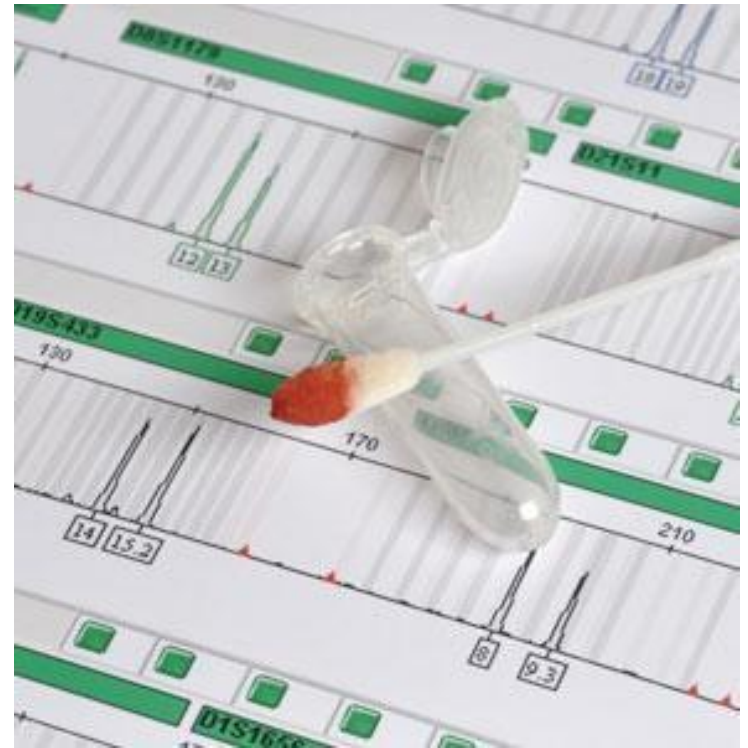


HUELLA GENÉTICA

- El perfil genético individual hace posible diferenciar a cualquier persona

Excepción → los **gemelos monocigóticos**, ya que en este caso **comparten la misma secuencia de ADN**.

- El perfil genético caracteriza a cualquier individuo igual o mejor que sus huellas dactilares, por lo que también recibe el nombre de huella genética.
- Ésta aporta la ventaja de que es mucho más precisa que otros métodos de identificación además, el ADN se halla en todas y cada una de las células del cuerpo humano, por lo que puede obtenerse de cualquier muestra biológica.
- La huella genética es **única e invariable a lo largo de la vida**.



HUELLA GENÉTICA

PRUEBAS DE LABORATORIO

La tecnología usada en el laboratorio para las prueba de la huella genética son:

- **RFLP (*Restriction Fragment Length Polymorphism*)**. Basada en la digestión del ADN por un enzima de restricción y posterior amplificación. Los fragmentos obtenidos son específicos para cada individuo.
- **STRs (*Short tandem repeat*)**. *Es hoy la técnica de elección, se basa en evaluar zonas específicas altamente variables del DNA.*
La variabilidad de éstas zonas nos permite distinguir un ADN de otro.
- El **análisis de SNP**. es el futuro de la huella genética y representa el paso de las metodologías que analizan polimorfismos de longitud a las metodologías que analizan polimorfismos de secuencia.
- **Análisis del ADN mitocondrial (*mtDNA*)**. El *DNA mitocondrial* es de origen exclusivamente materno (el padre sólo aporta DNA nuclear), a partir de las *mitocondrias del ovocito*.

ANÁLISIS DEL ADN MITOCONDRIAL

Es un recurso muy interesante en **estudios forenses** de



Muestras muy **antiguas**



Muy **degradadas**



Con **ausencia de ADN nuclear**

- El **ADN mitocondrial** es de origen **exclusivamente materno** (el padre sólo aporta DNA nuclear), a partir de las mitocondrias del ovocito.
- Es una molécula **circular**, mucho **más estable** y **más resistente** a la acción de las exonucleasas que el ADN nuclear lineal.
- Cada mitocondria tiene **múltiples copias de ADNmt**, por lo que el número de moléculas de ese tipo de ADN por célula es muy superior al del ADN nuclear.
- **No hay diferencias entre los individuos de un mismo linaje materno**
- Se basa en **polimorfismos de secuencia**; Tiene dos regiones hipervariables, denominadas **HVI** (342 pb) y **HVII** (268 pb), que son las utilizadas para este tipo de estudios.

HUELLA GENÉTICA

MUESTRAS REQUERIDAS

Básicamente se utiliza:

- **Sangre periférica** extraída con EDTA como anticoagulante
- **Frotis bucal** con torunda desprovista de medio de transporte.

Otras muestras que se pueden utilizar son:

- Manchas sobre soporte sólido (tejidos, ropas..)
- Pelo arrancado (para ADN nuclear),
- Pelo cortado (ADN mitocondrial),
- Piezas tumorales (en fresco o incluida en parafina)
- Semen.



ANÁLISIS DE LOS POLIMORFISMOS DEL CROMOSOMA Y

Es un cromosoma acrocéntrico pequeño que prácticamente no sufre recombinación durante la meiosis.

Esto lo hace especialmente interesante para estudios forenses, puesto que prácticamente todo el cromosoma constituye un grupo de ligamiento que se hereda junto.

Su análisis con fines forenses se basa en polimorfismos de longitud en STR (**Y-STR**).

Todos los **Y-STR** analizados se encuentran en la zona no recombinante del cromosoma Y, por lo que se heredan en bloque de padres a hijos varones,

Constituye lo que se denomina un haplotipo, que se mantiene generación tras generación sin más cambios que los producidos por mutaciones.

Esta propiedad es muy útil en:

- Estudios antropológicos para estudiar la evolución de linajes paternos.
- Estudios de paternidad de individuos varones cuando no existe muestra del presunto padre, pero se puede tener acceso a muestras de otros miembros del linaje paterno, como hermanos, tíos, abuelo paterno, primos, etc.
- En criminalística es en casos de delitos sexuales.

ANÁLISIS ESTADÍSTICO DE LOS DATOS

ANÁLISIS EN LA IDENTIFICACIÓN DE RESTOS Y CRIMINALÍSTICA

Se compara el perfil genético de una muestra problema con el perfil de una muestra de referencia para ver si coinciden

El análisis estadístico debe aportar una evidencia científica razonable, a ser posible inequívoca, de que la coincidencia de ambos perfiles es debido a que ambas muestras proceden del mismo individuo y no es producto del azar.

Para ello, se calcula la probabilidad que tiene el perfil obtenido de encontrarse en una población determinada.

Su cálculo varía según los marcadores analizados, según sean **Y-STR** o **STR autosómicos**.

Análisis de STR autosómicos → Se analizan 15 STR, la probabilidad de una combinación determinada de alelos (una huella genética) es del orden de 10^{-18}

Análisis del cromosoma Y → Se estima la frecuencia de un haplotipo concreto a partir de bases de datos que recogen miles de haplotipos

ANÁLISIS ESTADÍSTICO DE LOS DATOS

ANÁLISIS EN LOS ESTUDIOS DE PATERNIDAD

Se compara el perfil genético de STR autosómicos de un individuo (hijo o hija) con el de un supuesto padre o madre.

Hay tres casos posibles:

- Al menos en dos marcadores de todos los estudiados, no existe ninguna coincidencia entre los alelos de la persona y los de su supuesto progenitor. Esto constituye un criterio de exclusión, por lo que se descarta la paternidad biológica, independientemente de que en todos los demás STR haya coincidencia de un alelo.
- Existe coincidencia de un alelo del hijo o hija y uno del supuesto padre o madre para todos los STR estudiados excepto uno. En este caso no se aplica criterio de exclusión, puesto que ese único caso de no concordancia puede ser debido a una mutación. En estos casos hay que ampliar el estudio mediante STR autosómicos adicionales, Y-STR o RFLP.
- Existe concordancia en todos los STR estudiados entre la persona descendiente y la supuesta persona progenitora, es decir, uno de los alelos de cada STR del hijo o hija coincide con uno de los alelos del mismo STR del supuesto padre o madre.

ANÁLISIS ESTADÍSTICO DE LOS DATOS

ANÁLISIS EN LOS ESTUDIOS DE PATERNIDAD

En estos casos hay que validar estadísticamente el resultado, calculándose el **índice de paternidad (IP)** o la **probabilidad de paternidad (PP)**.

El **índice de paternidad** es el cociente entre la probabilidad de que el hijo o la hija haya recibido los alelos que integran su perfil genético del presunto padre o madre y la probabilidad de que los haya recibido de otro individuo distinto de la población.

- El valor del IP global es el producto de los IP individuales de cada STR analizado.
- El IP de un STR se calcula en función de la frecuencia del alelo coincidente en la población a la que pertenece el individuo en estudio.

La **probabilidad de paternidad** se calcula mediante la expresión matemática:

$$PP = IP \cdot 100 / (1 + IP)$$

Con los kits que analizan 15 STR, se consiguen probabilidades de paternidad superiores a 99,9%.

BIOINFORMÁTICA

Es una parte de la informática que utiliza herramientas computacionales para gestionar y analizar datos biológicos y médicos.

La gestión de datos incluye su **adquisición, organización, almacenamiento, análisis, comparación y visualización.**

En biología molecular la bioinformática es una herramienta cuyas aplicaciones resultan imprescindibles para:

- Localizar y comparar secuencias.
- Diseñar cebadores y sondas.
- Obtener información sobre el inserto que se quiere clonar, etc.

BIOINFORMÁTICA

Se han ido desarrollando:

- Bases de datos genómicas para almacenar la información generada por la secuenciación completa de múltiples organismos, incluido el genoma humano.
- Bases de datos de proteínas.
- Algoritmos y aplicaciones para análisis de secuencias, mediante alineamiento y comparación con secuencias conocidas.
- Software específico para la creación de mapas cromosómicos y genómicos.
- Algoritmos para diseño de cebadores y sondas.
- Herramientas para análisis de modelos evolutivos y creación de árboles filogenéticos.
- Aplicaciones para la interpretación de micro-arrays y análisis de expresión génica.
- Bases de datos especializadas de mutaciones somáticas y de la línea germinal, SNP, STR, etc.
- Modelos de predicción de la estructura 3D de proteínas (estructura secundaria, terciaria y cuaternaria).

BIOINFORMÁTICA

El último logro de la biinformática ha sido el diseño de herramientas para la llamada investigación en silico o investigación simulada por ordenador.

Se basa en tres conceptos:

- **MODELADO**: consiste en la generación de modelos biológicos virtuales a partir de los datos extraídos de sistemas biológicos reales.
- **SIMULACIÓN**: consiste en predecir de forma realista la evolución del modelo biológico virtual en el tiempo suponiendo determinados estímulos.
- **VISUALIZACIÓN**: hace referencia a la presentación gráfica de los resultados obtenidos en la simulación.

BIOINFORMÁTICA

En la actualidad se han diseñado innumerables bases de datos que recogen secuencias, tanto de ADN genómico, como de ADN clonado, ARNm, etc.

Entre las bases de datos especializadas se pueden mencionar:

- **GeneCards**. Es una base de datos de genes humanos
- **Cosmic**. Es un catálogo de mutaciones somáticas en el cáncer
- **HGVS**. Reúne varias bases de datos sobre variaciones y mutaciones del genoma humano

Respecto a las **bases de datos genéricas**, existe un proyecto de colaboración internacional, **International Nucleotide Sequence Database Collaboration**, que agrupa a las tres mayores bases de datos de secuencias nucleotídicas:

- Base de datos de secuencias nucleotídicas de los Institutos Nacionales de Salud de Estados Unidos (NIH, National Institutes of Health), denominada **GenBank**.
- Base de datos de secuencias nucleotídicas del Laboratorio Europeo de Biología Molecular (**EMBL**, European Molecular Biology Laboratory).
- Base de datos de secuencias nucleotídicas de Japón, denominada **DNA DataBank** (DDBJ).

BIOINFORMÁTICA

BASE DE DATOS DE GenBank

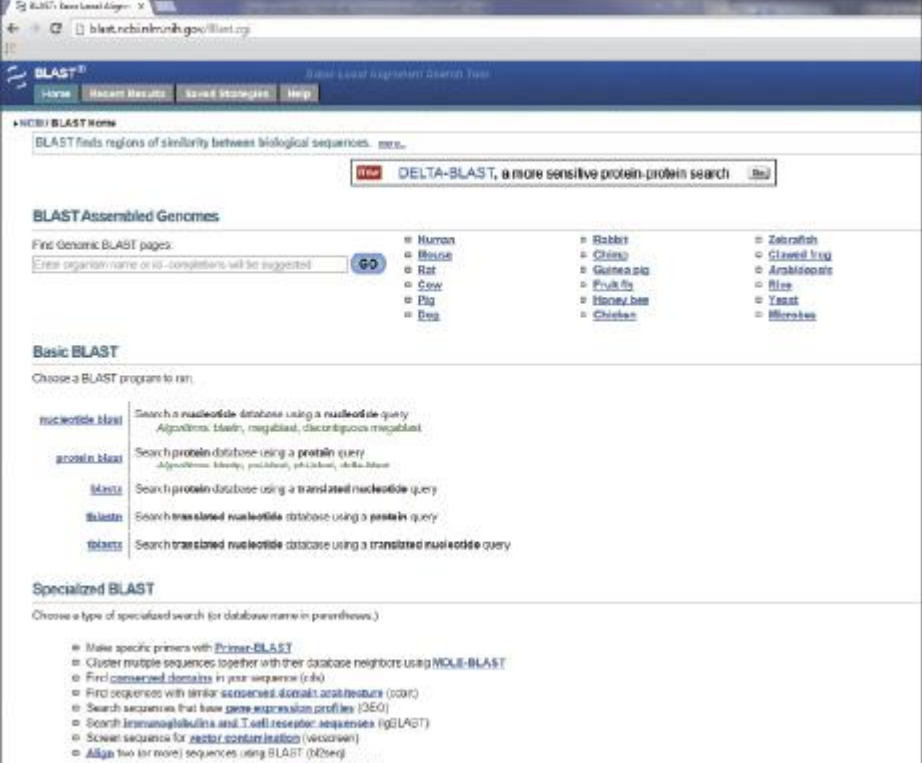
- Es la más utilizada, está disponible en línea y es de acceso público gratuito.
- Se accede a través de la página web del Centro Nacional para la Información Biotecnológica (NCBI, National Center for Biotechnology Information).
- Mantiene una amplia colección de bases de datos, tanto de secuencias génicas, como de proteínas, publicaciones científicas, etc
- La entrada se realiza desde el link «**Nucleotide**» en el apartado «**Genomes**», que abre un nuevo motor de búsqueda.
- Para acceder a una secuencia se puede buscar por el nombre del gen, del locus o su número de acceso.
- Una vez que se accede a la secuencia, se obtiene información sobre:
 - Locus, tamaño de la secuencia y tipo de ADN.
 - Número de acceso.
 - Organismo.
 - Referencias de las publicaciones en las que ha aparecido.
 - Estructura de la secuencia (exones, intrones, variaciones, etc.).
 - Secuencia completa.

BIOINFORMÁTICA

ANÁLISIS DE SECUENCIAS

La más común consiste en la comparación de una secuencia problema con todas las secuencias de una base de datos nucleotídica para identificar la secuencia y detectar posibles variaciones y/o mutaciones.

- La aplicación más utilizada se llama «**BLAST**», perteneciente al NCBI, con acceso gratuito en línea.
- Para comparar secuencias nucleotídicas se selecciona la opción «nucleotide blast» que lleva a una nueva pantalla en la que se introduce la secuencia que se quiere analizar, bien tecleándola directamente, bien cargándola desde un archivo.



The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with 'Home', 'Recent Results', 'Saved Searches', and 'Help'. Below that, a search bar is visible with the text 'BLAST finds regions of similarity between biological sequences. more...'. A prominent button for 'DELTA-BLAST, a more sensitive protein-protein search' is shown. The main content area is titled 'BLAST Assembled Genomes' and includes a search box for 'Enter organism name or id - completions will be suggested' with a 'GO' button. To the right of this box is a list of organisms: Human, Mouse, Rat, Cow, Pig, Dog, Rabbit, Guinea pig, Poultry, Honey bee, Chicken, Zebrafish, Citrus fruit, Arabidopsis, Rice, Yeast, and Microbes. Below this is the 'Basic BLAST' section, which offers several options: 'nucleotide blast' (Search a nucleotide database using a nucleotide query), 'protein blast' (Search protein database using a protein query), 'tblastn' (Search protein database using a translated nucleotide query), 'tblastx' (Search translated nucleotide database using a protein query), and 'tblastx' (Search translated nucleotide database using a translated nucleotide query). At the bottom, there's a 'Specialized BLAST' section with a list of specialized search types, including 'Make specific primers with Primer-BLAST', 'Cluster multiple sequences together with their database neighbors using MCL-BLAST', 'Find conserved domains in your sequence (rdb)', 'Find sequences with similar conserved domain architecture (cdart)', 'Search sequences that have gene expression profiles (GEO)', 'Search immunoglobulin and T cell receptor sequences (igBLAST)', 'Screen sequences for zeta/zeta interaction (vicoview)', 'Align two (or more) sequences using BLAST (Mseq)', and 'Search protein or nucleotide targets in Pfam/BioAssay'.

BIOINFORMÁTICA

PORTALES BIOINFORMÁTICOS

En internet se encuentran también portales bioinformáticos que ofrecen acceso a múltiples aplicaciones.

La página web del NCBI constituye quizá uno de los portales bioinformáticos más importantes.

Otros portales importantes son:

- **EMBOSS explorer**. Es un portal con más de doscientas aplicaciones para trabajar con ácidos nucleicos y proteínas.
- **Instituto Europeo de Bioinformática (EBI)**, que forma parte del EMBL. En su portal se encuentran algunas de las mejores aplicaciones y herramientas para alineamiento pareado y múltiple de secuencias (aplicaciones «Clustal»), imprescindibles para la realización de árboles filogenéticos.
- **Ensembl**. Contiene múltiples bases de datos de genomas de vertebrados (mamíferos, aves, peces, reptiles), así como herramientas para comparar genes entre especies, expresión génica en tejidos, etc.

GeneCards®: The Human Gene Database

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. It automatically integrates gene-centric data from ~125 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information.



Explore a Gene



GO

Jump to section for this gene:

[Aliases](#)
[Disorders](#)
[Domains](#)
[Drugs](#)
[Expression](#)
[Function](#)
[Genomics](#)
[Localization](#)
[Orthologs](#)
[Paralogs](#)
[Pathways](#)
[Products](#)
[Proteins](#)
[Publications](#)
[Sources](#)
[Summaries](#)
[Transcripts](#)
[Variants](#)

GeneCardsSuite

NGS Analysis Tools



Affiliated Databases



Analysis Tools



GENECARDS: base de datos de genes humanos (www.genecards.org)

COSMIC v76

eg: *Braf*, *COLO-829*, *Carcinoma*, *V600E*, *BRCA-UK*, *Campbell*

SEARCH

R Resources

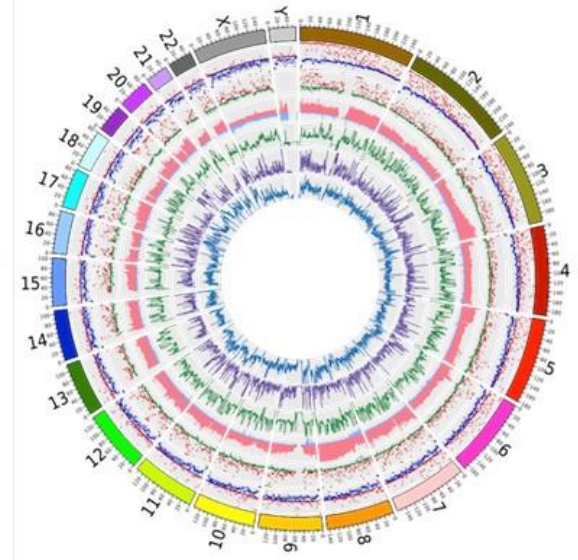
Key COSMIC resources

- [Cell Lines Project](#)
- [COSMIC Whole Genomes](#)
- [Cancer Gene Census](#)
- [Drug Sensitivity](#) 
- [Mutational Signatures](#)
- [GRCh37 Cancer Archive](#) 

T Tools

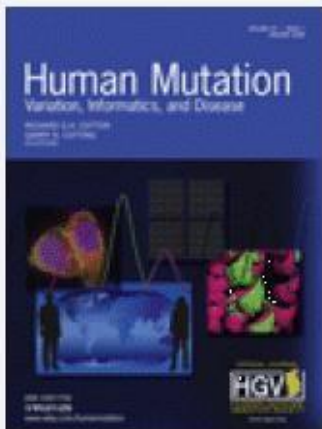
Additional tools to explore COSMIC

- [Cancer Browser](#)
- [Genome Browser](#)
- [GA4GH Beacon](#) ^{New}
- [COSMIC Mart](#)
- [CONAN](#)



Genomic Landscape of Cancer

COSMIC: Catálogo de mutaciones somáticas en el cáncer



HUMAN MUTATION

Members who choose to subscribe to *Human*

ABOUT THE SOCIETY

Prof. Richard (Dick) G.H. Cotton AM, BAgSc., Ph.D, D.Sc. was the Founding President of HGVS.

In Memoriam - Richard G.H Cotton (1940-2015)

The Society aims to foster discovery and characterization of genomic variations including population distribution and phenotypic associations. Promote collection, documentation and free distribution of genomic variation information and associated clinical variations. Endeavor to foster the development of the

FORTHCOMING EVENTS

The Next HGVS Meeting

Clinical Interpretation of Variants from Next-Generation Sequencing
20 May 2016
Barcelona Spain
(A Satellite of the ESHG Meeting)

HGVS: Reúne varias bases de datos sobre variaciones y mutaciones del genoma humano