

**UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA** 

# Manual de prácticas de laboratorio Introducción a la Bioinformática



Ma. Dolores García Suárez

Héctor Serrano

#### UNIVERSIDAD AUTÓNOMA METROPOLITANA

Dr. Salvador Vega y León *Rector General* 

Mtro. Norberto Manjarrez Álvarez Secretario General

#### UNIDAD IZTAPALAPA

Dr. Javier Velázquez Moctezuma Rector de Unidad

Dr. Miguel Ángel Gómez Fonseca Secretario de Unidad

Dr. Rubén Román Ramos Director de la División de Ciencias Biológicas y de la Salud

Dra. Milagros Huerta Coria Coordinadora de Extensión Universitaria

Lic. Adrián Felipe Valencia Llamas Jefe de la Sección de Producción Editorial

Primera Impresión 2013

UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA Av. San Rafael Atlixco No. 186, Col. Vicentina, Del. Iztapalapa, C.P 09340, México D.F. Tel.: 5804 4600

Impreso y hecho en México/Printed in Mexico

## Índice

Prólogo		5
Capítulo 1	Búsqueda de literatura especializada	7
Capítulo 2	Actualización automatizada de literatura científica	15
Capítulo 3	Obtención y análisis de secuencias de proteínas utilizando un servidor en línea	21
Capítulo 4	Simulación de la estructura de genes a partir de la secuencia de aminoácidos	33
Capítulo 5	Búsqueda de información general de enfermedades genéticas	43
Capítulo 6	Diseño de oligonucleótidos para amplificación	49
Referencias ger	ierales	62



## Prólogo

El material que está a punto de analizar y seguir, fue pensado para permitir que los alumnos del nivel Licenciatura y las personas que se están incorporando a la investigación científica en general, puedan tener una guía simple para poder integrarse a la corriente actual de la actividad de la investigación.

En primer lugar, este manual está enfocado a las necesidades de la Genética Molecular Humana. Sin embargo, dada la generalidad y estructura del texto, puede ser aplicable a otros campos. Las herramientas que se sugieren no son lo más sofisticadas que podrían encontrarse disponibles. De hecho, esta simplicidad permite su amplia aplicación y la facilidad de interpretación que normalmente requiere un curso introductorio. En cada una de las páginas electrónicas que se utilizan se encuentra material de apoyo que permite un uso amigable.

La mayoría de las plataformas utilizadas permite que sean modificadas por lo que si se tienen los conocimientos avanzados, permiten llenar los requerimientos de lo que sería el desarrollo de procesos de búsqueda, comparación, análisis y evaluación de resultados específicos que requiera un laboratorio de investigación en alguna de las disciplinas "ómicas" que se estén desarrollando.

Los autores de este manual esperamos que les sea útil y les permita adentrarse en el uso de herramientas que hagan aún más atractiva la actividad de investigación que estén desarrollando.



## Capítulo 1 Búsqueda de literatura especializada

La labor del investigador es una acción continua y que requiere en primera instancia de una documentación adecuada, confiable y que permita el avance del campo del conocimiento en el que se desempeña. En la mayoría de los casos, trata de ser vanguardista por lo que es necesario mantener un acervo no sólo de las evidencias experimentales que apoyan los resultados de uno u otro grupo de trabajo, sino que abarque la importancia de ambos, evalúe en la medida correcta los aciertos y remarque los campos que son necesarios apoyar para dar congruencia a un fenómeno específico.

Podemos dividir en dos grandes etapas la labor de documentación del investigador a partir del desarrollo de una herramienta brindada por la red de redes informáticas. Era común que un investigador pasara varias horas en la biblioteca de la institución a la que pertenecía leyendo o fotocopiando los artículos de las revistas donde comúnmente se publicaban los resultados de su campo. Esta costumbre tenía la limitante de la cantidad de revistas a las que estaba suscrita la institución y era más o menos solventada por los intercambios bibliotecarios, con la consiguiente ampliación del tiempo que requería para leer un artículo que muchas veces cuando llegaba, ya no tenía el impacto de cuando fue requerido.

Un servicio de alerta se ofreció a través de una compañía denominada Institute for Scientific Information (ISI) que tenía sus oficinas generales en Philadelphia, Estados Unidos. El servicio de documentación Current Contents© primeramente por correo ordinario y posteriormente apoyados por computadora permitían una actualización relativamente aceptable. De manera general, con una sola suscripción se podían atender varios laboratorios de una sola Institución y cada uno de los miembros tenía un tiempo y prioridad de revisión especificado para hacer eficiente dentro de lo posible el sistema.

Con el desarrollo de una función alterna a la que se le denominó "Medline", se entró en el uso de las comunicaciones vía computadora. Este servicio, que originalmente tenía un costo de suscripción permitía mantener búsquedas retrógradas de hasta 5 años de la literatura publicada en un tema particular o tener un listado semanal de los artículos publicados sobre temas específicos. Al igual que con Current Contents, el acceso era dependiente de la capacidad económica del laboratorio y la disponibilidad de al menos un aparato de fax o acceso a la red institucional que en México a finales de la década de 1980, era restringido solamente a la UNAM y CONACyT.

La fundación del Centro Nacional de Información Biotecnológica (NCBI) como una división específica de la Biblioteca Nacional de Medicina (NLM) dependiente de los Institutos Nacionales de Salud (NIH) utilizando la plataforma y base de datos de Medline consolidará la herramienta de búsqueda poderosa que conocemos como PubMed y su central de artículos sin costo PubMed Central. Esta base de datos se actualiza cada 24 horas y es alimentada fundamental aunque no exclusivamente por las oficinas editoriales de las revistas.

#### Objetivo

Mostrar el uso de la base de datos PubMed en el acopio de literatura especializada.

#### Desarrollo:

Acceda a la página principal del NCBI a través de la dirección http://ncbi.nlm.nih.gov

Puede llegar a la página de PubMed ya sea desplegando el menú en la casilla de Búsqueda (SEARCH) o en la liga de Recursos comunes (POPULAR RESOURCES) (Figura 1.1). Si lo hace en la ventana del dialogo, puede incluir inmediatamente la palabra clave que identifique su búsqueda. Si lo hace por la liga de Popular resources, una vez que se despliega la página de PubMed escriba la palabra clave.



En ambos casos, puede utilizar diversos tipos de entradas (AUTOR, KEY WORD, JOURNAL) ya que el identificador incorporado tiene esta facilidad. Al presionar el botón SEARCH se despliega la información requerida. La duración de la búsqueda depende de la complejidad del campo, la actividad que tenga en ese momento y el volumen de información que se tenga.

S NCBI Resources (	] How To 🔍			Sign in to NCBI
SNICH	Recent			
SINCBI	✓ All Databases			Search
Biotechnology Information	PubMed			
	PMC			
NCBI Home	All	CBI		
Recourse Liet (A.7)	All Databases	00.		Popular Resources
Resource List (M-L)	Assembly	for Biotechnology Information advances science and health	h by providing access to	PubMed
All Resources	BioProject	mic information.		Bookshelf
Chemicals & Bioassays	BioSample	fission   Organization   Research   NCBI News		PubMed Central
Data & Software	BioSystems			PubMed Health
DNA & RNA	Books			BLAST
Domains & Structures	ClinVar			Nucleotide
Genes & Expression	Clone	e data using NCBI software		Genome
Genetics & Medicine	Conserved Domains	set NCBI data or software		SNP
Genomes & Maps	dbGaP	Submit data to GenBank or other NCBI databases		Gene
Homology	DbVar			Protein
Literature	Enigenomics			BubCham
Destalas	EST			- uconum
Proteins	Gene	pok page		
Sequence Analysis	Genome	news about NCBI		NCBI Announcements
Taxonomy	GEO Data Sate	ticipate in community		Tenth Anniversary of RefSeg FTP
Training & Tutorials	GEO Brofiles	GO		Releases
Variation	0201101103			The July 2013 RefSen FTP release
	HemoleCone	5678		
	MadCas			RefSeq Release 60 is Available for FTP
	Medden			The complete RefSec release 60
	Mean			contains 40,913,699 records, 32,504,738
	NCBI Web Site			
	NLM Catalog			Pandoravirus Sequences are Accessible
	Nucleotide			in GenBank*
	OMIA			Jul 24, 2013
	OMIM			· ·
	PMC			More
	PopSet			
	Probe			
You are here: NCBI > Nation	Protein			Write to the Help Desk
GETTING STARTED	Protein Clusters	POPULAR	FEATURED	NON INFORMATION
NC8I Education	PubChem BioAssay	PubMed	Genetic Testing Registry	About NCBI
NCBI Help Manual	PubChem Compound	Bookshelf	PubMed Health	Research at NCBI
NC8I Handbook	PubChem Substance	PubMed Central	GenBank	NCBI News
Training & Tutorials	PubMed	PubMed Health	Reference Sequences	NGBI FTP Site
	PubMed Health	BLAST	Gene Expression Omnibus	NCBI on Facebook
	SNP	Receide	Map Viewer	NCBI on Twitter
	SRA	SNP	Mouse Genome	
	Structure	Gene	Influenza Virus	
	Taxonomy	Protein	Primer-BLAST	
	ToolKit	PubChem	Sequence Read Archive	
	ToolKitAll			
	ToolKitBook			
	UniGene			

Figura 1.1a. Pantalla de inicio del NCBI.

S NCBI Resources (	🛛 How To 🖸				Sign in to NCBI
Pub	(PubMed 1)				Search
US National Library of Medicine National Institutes of	Advanced				Help
Health					
	PubMed PubMed compris MEDLINE, Iffe so to full-text conten	es more than 23 million clations for biom ence journals, and online books. Clation from PubMed Central and publisher wet	edical literature from s may include links o sites.	PubReader A whole new wey to read scientific literature at PubMed Centrel	
Using PubMed		PubMed Tools		More Resource	ces
PubMed Quick Start Gui	de	PubMed Mobile		MeSH Database	
Full Text Articles		Single Citation Matcher		Journals in NCBI	Databases
PubMed FAQs		Batch Citation Matcher		Clinical Trials	
PubMed Tutorials		Clinical Queries		E-Utilities	
New and Noteworthy		Topic-Specific Queries		LinkOut	
You are here: NCBI > Literate	ure > PubMed				Write to the Help Desk
GETTING STARTED	RESOURCES	POPULAR	FEATURED		NCBI INFORMATION
NCBI Education	Chemicals & Bioassay	s PubMed	Genetic Tes	ing Registry	About NCBI
NCBI Help Manual	Data & Software	Bookshelf	PubMed Hea	lth	Research at NCBI
NCBI Handbook	DNA & RNA	PubMed Central	GenBank		NCBI News
Training & Tutorials	Domains & Structures	PubMed Health	Reference S	equences	NCBI FTP Site
	Genetics & Medicine	Nucleotide	Map Viewer	Ision Omnious	NCBI on Twitter
	Genomes & Maps	Genome	Human Gen	me	NCBI on YouTube
	Homology	SNP	Mouse Geno	me	
	Literature	Gene	Influenza Vir	us	
	Proteins	Protein	Primer-BLAS	т	
	Sequence Analysis	PubChem	Sequence R	sad Archive	
	Taxonomy				
	Variation				
Copyright   Disclaimer   Pri-	vacy   Browsers   Accessibility   Contac	1			
National Center for Biotechno	ology Information, U.S. National Library o	- f Medicine			M Ch Term
8600 Rockville Pike, Betheso	da MD, 20894 USA				HARDEN SALES CLEAR USA.gov

Figura 1.1b. Pantalla PubMed.

La búsqueda en PubMed Central brinda un menor número de artículos referidos pero casi siempre disponibles para ser adquiridos inmediatamente sin costo o requerimientos (Figura 1.2). A esta iniciativa se le conoce como OPEN ACCESS y varias revistas están haciendo esfuerzos por poner a disposición de la comunidad científica su contenido. Algunas otras permiten el Open Access a revistas de sus archivos que fueron publicadas un lapso variable de tiempo atrás; finalmente, algunos autores han pagado el costo que representa a los editores para que la información sea catalogada como Open Access, básicamente mediante apoyos otorgados por el NIH o el gobierno de Estados Unidos.



Figura 1.2a. Resultados simples utilizando la palabra clave "Phytoestrogen\*" en las bases PubMed.



Figura 1.2b. PubMed Central.

En el desarrollo del curso, estaremos interesados en tres tópicos fundamentales: la regulación de la esteroidogénesis en mamíferos, el uso de marcadores moleculares de diabetes y enfermedades asociadas al aparato circulatorio y las pruebas de paternidad.

Para cada uno de estos tópicos será necesaria la documentación de los antecedentes acumulados durante los 10 años previos. Para llenar este requisito, abra la liga LIMITS y en la ventana de dialogo donde aparece "PUBLICATION DATE" indique 10 YEARS. Asegúrese de haber introducido la palabra clave en la ventana de diálogo principal (SEARCH). Si es necesario, indique los parámetros adicionales. Nuestro interés es con estudios en animales de sexo indistinto, artículos publicados únicamente en Inglés, Francés, Alemán, Italiano, Portugués y Español. Indique la búsqueda oprimiendo el botón SEARCH del final de la página (Figura 1.3).



S NCBI Resources ⊙ Ho	w To 🕑			<u>Sign in to N</u>
PMC Home Help				
MC Advanced Search	Builder			
History deleted.				
Use the bu	ilder below to create your search			
Edit				Clear
Builder				
0	All Fields \$		0	Show index list
(AND \$)	All Fields \$		0	Show index list
Search	or Add to history			
There is no	recent history			
ou are here: NCBI > Literature > P	ubMed Central (PMC)			Write to the Help
ETTING STARTED	RESOURCES	POPULAR	FEATURED	NCBI INFORMATION
281 Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry	About NCBI
31 Help Manual	Data & Software	Bookshelf	PubMed Health	Research at NCBI
BI Handbook	DNA & RNA	PubMed Central	GenBank Perference Sequences	NCBI News
ining & Totonais	Genes & Excession	Publied Health BLAST	Gene Expression Omniture	NODI PTP Stel
	Genetics & Medicine	Nucleotide	Map Viewer	NCBI on Twitter
	Genomes & Maps	Genome	Human Genome	NCBI on YouTube
	Homology	SNP	Mouse Genome	
	Literature	Gene	Influenza Virus	
	Proteins	Protein	Primer-BLAST	
	Sequence Analysis	PubChem	Sequence Read Archive	
	Taxonomy			
	Training & Tutorials			
	Variation			
pyright   Disclaimer   Privacy   I	srowsers   Accessionity   Contact			
anona conta to biotechnology in	normation, e.e. reasonal borary of medicine			

Figura 1.3a. Ventana de diálogo de búsqueda de avanzada de PubMed.

	10 🕑			Sign in to NCBI
US National Library of Medicine National Institutes of Health	C () Journal List Advanced			Search Help
Limits				
Embargoed Articles				
Limit by Author			Limit by Journal	
Add Author		Clear	Add Journal	Clear
Dates				
Published in the Last Added to PMC in the Last	: Any date ¢			
Article Type				
Research & Review Articles     Articles Cited by Other PMC     Author Manuscripts     Open Access Articles     Corrected or Retracted Artik	: Articles		<ul> <li>All of these article types (AND)</li> <li>Any of these article types (OR)</li> <li>None of these article types (NOT)</li> </ul>	
Search Field Tags				
Field: All Fields	•			
		Reset	Search	
You are here: NCBI > Literature > Pub	Med Central (PMC)			Write to the Help Desk
You are here: NCBI > Literature > Pub GETTING STARTED	Med Central (PMC) RESOURCES	POPULAR	FEATURED	Write to the Help Desk
You are here: NCBI > Literature > Pub GETTING STARTED NCBI Education	Med Central (PMC) RESOURCES Chemicals & Bioassays	POPULAR PubMed	FEATURED Genetic Testing Registry	Write to the Help Desk NCBI INFORMATION About NCBI
You are here: NCBI > Literature > Pub GETTING STARTED NCBI Education NCBI Help Manual	Med Central (PMC) RESOURCES Cherricala & Biosssays Data & Software	POPULAR PubMed Bookshelf	FEATURED Genetic Testing Registry PubMed Health	Write to the Help Desk NCBI INFORMATION About NCBI Research at NCBI
You are here: NCBI > Literature > Pub GETTING STARTED NCBI Education NCBI Hejn Manual NCBI Handbook	Med Central (PMC) RESOURCES Oherricals & Biossays Data & Scheare DNA & RNA	POPULAR PubMed Bookshelf PubMed Central	FEATURED Genetic Testing Registry PubMed Health GenBank	Write to the Help Desk NCBI INFORMATION About NCBI Research at NCBI NCBI News
You are here: NCBI > Literature > Pub CETTING STARTED NCBI Education NCBI Help Manual NCBI Help Monual NCBI Help Noticotic Training & Tutorials	Med Central (PMC) RESOURCES Chemicals & Bioassays Data & Schware DNA & RNA Domains & Structures	POPULAR PubMed Bookshelf PubMed Central PubMed Health	FEATURED Genetic Testing Registry PubMed Health GenBank Reference Sequences	Write to the Help Desk NCBI INCORTON About NCBI Research at NCBI NCBI Hens NCBI FTP. Site
You are here: NCBI > Literature > Pub CETTING STATED NCBI Education NCBI Heip Manual NCBI Handbook Training & Tutorials	Med Central (PMC) RESOURCES Chemicals & Biosssays Data & Schware DAVA & FINA Domine & Structures Genes & Expression	POPULAR PubMed Bookshelf PubMed Central PubMed Health BLAST	FEATURED Great: Training Registry PoMMed Health Greatering Regulatories Gene Expression Ommibus	Write to the Help Desk About ACBI Research AICBI NOBI Here NOBI FTP Site NOBI or Research
You are here: NOBI > Literature > Pub <b>GETTING STARTED</b> NOBI Education NOBI Herbook NOBI Herbook Training & Tutorials	Med Central (PMC) RESOURCES Chericies & Sossays Data & Software DNA & RNA Domaina & Structures Genes & Expression Genesia & Medicine	POPULAR PubMed Bookshelf PubMed Central PubMed Health BLAST Nucleotide	FEATURED Gentis Testing Registry PubMed Health Gentlank Reference Sequences Gene Expression Ombus May Viewer	Write to the Help Desk NCBI INFORMATION About NCBI Research at NCBI NCBI Heves NCBI FIP Site NCBI on Twitter NCBI on Twitter
You are here: NCBI > Literature > Put <b>GETTING STARTED</b> NCBI Recursion NCBI Hendbook NCBI Handbook Training & Tutorials	Med Central (PMC) RESOURCES Oranicais & Blosssays Data & Schware DNA RPW Domains & Structures Gene & Expression Genetics & Medicine Genetics & Mas	POPULAR PelMed Booksheff PelMed Health BLAST Nucleotide Genome	FEATURED Genetic Testing Registry Publick Health Ordhank Reference Securices Gene Espession Ormhou May Newer Huma Ginnom Huma Ginnom	Write to the Help Desk NOBI INFORMATION Adda NOBI Research a NOBI NOBI The Sta NOBI on Facebook NOBI on Facebook NOBI on Facebook NOBI on Facebook
You are here: NOBI > Ukerature > Pub GETTING STARTED NOBI Education NOBI Heip Manual NOBI Handbook Training & Tutohials	Med Central (FMC) RESOURCES Omnical & Bossays Data & Software Dirk & Software Dirk & Software Dirk & Software Dirk & Software Centrals & Modicine Centrals & Modicine C	POPULAR PubMed Booksheff PubMed Central PubMed Health BLAST Nucleotide Genome SNP	FEATURED Genetic Testing Registry Publick Health Genbark Reference Sequences Gene Expression Ombos May Univer Human Occome Manan Second May Second	White to the Help Desk Accel INFORMATION Accel NOI Reserved In NOI NOI IFT Sea NOI INT Sea NOI INT Sea NOI IN There NOI IN There NOI IN YouTube
You are here: NCBI > Literature > Pub GETTING STARTED NCBI Receiption NCBI Herotook NCBI Herotook Training & Tutorias	Med Central (PMC) RESOURCES Drancas & Botessay Das & Software Drak & Software Drak & Software Drak & Software Drak & Software Genomes & Maccine Genomes & Maccine Hendringy Unstance Products	POPULAR PubMed Bookshelf PubMed Central PubMed Health BLAST Nucleotide Genome SNP Gene Posten	FEATURED Grants: Terring Registry Grants: Reference desparation Reference desparation Reference Manual Genome Name Genome Name Genome Name Genome Name Bell MI	With to the Help Desk NCEL INFORMATION Account NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI NCBI
You are hore: NCBI > Literature > Put CETTING STARTED NCBI Hord Monual NCBI Hord Monual NCBI Hord Monual NCBI Hord Monual Tracing & Tutorials	Med Central (PAC) RESOURCES Omnicals & Biostays Data & Software Data & Software Data & Software Data & Software Omnia & Software Genome & Massi Homology Literature Social Annalis Social Annalis	POPULAR PubMed Bookshelf PubMed Central PubMed Heath BLAST Nucleotide Genome SNP Gene Protein PubCoven	FEATURED Genetic Testing Registry Publick eleath Genbark Reference Sequencies Gene Expression Ombos May Univer Haman Genome Manara et Vinn Primer BLAT Bosume Ref Arthyse	White to the Help Deak Accu NOSI Reserved at NOSI NOSI News NOSI PTP Site NOSI on Technology NOSI on Technology NOSI on Technol NOSI on YouTube
You are here: NCBI > Liberature > Pub GETTING STARTED DOEI Education NCBI Hery Manual NCBI Hery Manual NCBI Hery Annual NCBI Hery Annual Training & Tudontis	Med Central (PMC) RESOURCES Drancas & Bossasy Dra & Software Dra & RNA Drann & Sonastin General & Macines General & Macines General & Macines Dranses Hendroy Unstance Prodes Sequero Anglis Tatorom	POPULAR PubMed Bookshelf PubMed Gentral PubMed Heatin BLAST Nucleotide Genome SNP Gene Protein PubChem	FEATURED Genetic Testing Registry Contention Interesce Securices Gene Expression Commons May Visear Human Genome Indianas Visa Prome-Build Degrame Testa Pachrie	With to the Help Desk NCBI INFORMATION Accol KCBI Mod Kenet KCBI Kenet KCBI Kenet KCBI Kenet KCBI on Telson KCBI on Telson KCBI on Telson KCBI on Telson
You are here: NCBI > Literature > Puto <b>GETTHOS STARTED</b> NCBI Horizon NCBI Horizon	Med Central (PAC) PRESURCES Omnicals & Bossays Data & Software Data & Software Data & Software Data & Software Data & Software Connes & Software Generals & Monites Homogra Liberature Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pactanian Pac	POPULAR PubMed Bookshelf PubMed Central PubMed Central BLAST Nucleotide Genome SNP Gene Protein PubChem	FEATURED Genetic Testing Registry Publick eleath Genbark Reference Sequences Gene Expression Ormbia May Univer Hama Genome Manage Genetic Indexed Unit Beguence Read Active	White to the Help Deak Accu NOSI Reserved at NOSI NOSI PTP Site NOSI PTP Site NOSI on Resetook NOSI on Resetook NOSI on NouTube
You are here: NCBI + Literature + Put <b>GETTING STARTED</b> NCBI Horization NCBI Horization KCBI Horization Training & Tutorian	Med Central (PMC) RESOLUCES Data & Schware Data & Schware Data & Schware Data & Schware Data & Schware Data & Schware Controls & Schware Genome & Massi Honology Literature Proteins Sequere Analysis Sequere Analysis Taming & Litoxies Versito	POPULAR Pubble Bookhard Pubble Charal Pubble Charal BLAST Nucleickie Genome SAP Genom Pubblem	FEATURED Grants: Terring Rigistry Grants: Methodiek Reference Separation Amerika Mark Genome Marka Genome Mar	White to the Help Deck NCEL INFORMATION Allow NCEL NCEL No. NCEL NO. NC
You are hare: NCB > Literature > Put GETTING STARTED NCB1 Honitoux NCB1 Honitoux NCB1 Honitoux Training & Tutorate Copyright   Disclaimer   Privacy   BT	Med Central (PAC) RESOURCES Commonità & Bossays Data & Schware Data & Schware Data & Schware Data & Schware Central & Schware Genera & Schware Genera & Schware Central & Macine Central & Macine Central & Macine Central & Macine Central & Macine Central & Macine Central & Schware Central (Schware Central & Schware Central & Schware Ce	POPULAR Pubble Bootshaf Pubble Gernal Pubble Health BLAST Buble Buble Same Protein Pubblem	FEATURED Carenic Tenning Registry PubMed Feath Carelline Seatons Carelline Seatons Carelline Seatons Mark Vener Henna Genome Influenza Visa Prime-BLAST Sequence Read Archive	With to the Help Desk Accel INFORMATION Accel And Reserver at IAGE IAGE INFORMATION ACCEL INFORMATION IAGE IN Frances IAGE INFORMATION IAGE INFORMATION IA

Figura 1.3b. PubMed Central. PMC ofrece la opción de delimitar la búsqueda utilizando los criterios que aparecen en esta ventana. PubMed lo hace directamente en la ventana de resultados (vea el campo izquierdo de la figura de arriba).

Después que el buscador le entregue los resultados, seleccione sólo aquellos Open Access pulsando la liga FREE FULL TEXT que aparece dentro del recuadro REFINE YOUR SEARCH localizado en la porción de herramientas (lado derecho de la pantalla).

Una vez que se tiene acceso al listado de artículos gratuitos, aumente el número de referencias que puede ver por página abriendo el cuadro de dialogo localizado al inicio del encabezado del listado bajo el renglón DISPLAY SETTINGS. Obtendrá un cuadro. Al terminar de hacer la selección, pulse el botón APPLY que está al fondo del cuadro de diálogo

#### Actividad para reportar

Utilizando las palabras clave STEROID\*, MOLECULAR MARKERS, TYPE 2 DIABETES MELLITUS, ADRENERG\* RECEPTOR\* y PATERNITY TESTING haga una búsqueda de la información que se hubiese publicado sobre el tema en los últimos 10 y 5 años.

Anote la cantidad de artículos originales, de revisión y editoriales que se tenga en cada lapso. Anote la cantidad de artículos disponibles bajo el esquema de OPEN ACCESS.

Anote la cantidad relativa de artículos que se escribieron en cada uno de los idiomas principales que estamos utilizando en el curso.

Anote el número de estudios que se han realizado en animales y humanos así como la proporción de estudios que se hicieron por sexo de los individuos.

En el caso de estudios en humanos, indique el número de artículos en los cuales están contenidos datos de estudios con aplicación clínica.

Obtenga un artículo de investigación relacionado con los temas de estudio del curso y uno de revisión o meta-análisis, si es aplicable.



Figura 1.4a. Ventana de acceso a las bases de datos de la biblioteca UAM-I.

WEB O		C THOMSON REUTERS
Go to mobile site	Sign In   Marked List (0)   EndNote   ResearcherID   Citation Alerts   Journ	al Alerts   Saved Searches   Log Out   Help
All Databases Search Searc	Select a Database Web of Science Additional Resources	Universidad Autonoma Metropolitana Information For new usens
Search	La Ses	Get EndNote X7 now! Store your references and PDFs, find full text, and sync your library across desktop, online and iPad.Try it now!
AND 🗘	Example: O'Brian C* O'R OBrian C* in Author : Select from Index	Training and Support  Download Quick Recorded Training Find us on
AND 🛟	Example: Cancer* OR Journal of Cancer in Publication Name : Select from Index Research and Clinical Oncology	• Facebook
	Add Another Field >> Search Clear	Data Citation Index <sup>SM</sup> ; Discover, use and cite research data. More information.     More of What's New
Limits: (To sa	ve these permanently, sign in or register.)	Customize Your Experience Sign In   Register
⊞- Tir €	nespan (All years 0)	Save and run manage your references     online with EndNote – freely available     and fully integrated.     Save and run searches

Figura 1.4b. Acceso a ISI Web of knowledge si se está conectado a una computadora de la Unidad o mediante el acceso remoto permitido.



Utilizando la base de datos electrónica ISI WEB of Knowledge (bases de datos disponible en línea dentro del portal de la Biblioteca UAM-I ya sea mientras está utilizando computadoras conectadas a la red de la Universidad o mediante el ingreso de su clave de identificación (DIRECCION IP) para poder tener acceso remoto), realice una búsqueda similar (Figura 1. 4). Utilice al menos otras dos bases de datos electrónicas y realice búsquedas similares. Como punto de partida, utilice la opción de Google Académico (http://scholar.google.com.mx ) y Microsoft Academic Search ( http://academic.research.microsoft.com , Figura 1.5), las páginas electrónicas más conocidas y de mayor consulta. En cada caso obtenga un artículo de investigación de cada tema del curso y una revisión o meta-análisis, si es aplicable.



Figura 1.5a. Ventana de acceso a Google Académico.



Figura 1.5b. Microsoft Academic Search popular buscadores.

Indique las ventajas y limitantes que presenta el uso de ISI Web of Knowledge y las otras dos bases de datos que consultó cuando se comparan con PubMed y PubMed Central.

## Referencias

- PubMed Help [Internet]. National Center for Biotechnology Information. 2005. Disponible en el URL <u>http://www.ncbi.nlm.nih.</u> gov/books/NBK3830
- Sawyers, E., Barret, T., Benson, D.A., Bolton, E., Bryant, SH y otros 38 co-autores. 2012. Database resources of the National Center for Biotechnology Information. Nucleic. Acids. Res. 40:D13-D25. Disponible en el URL http://nar.oxfordjournals.org/content/40/D1/D13.full.pdf+html



## Capítulo 2

## Actualización automatizada de literatura científica

La generación de conocimiento que implica la investigación requiere de una actualización constante. La búsqueda en las bases de datos puede ser repetitiva y para poder ser eficiente, debería ser al mismo ritmo que es actualizada la base de datos en línea que se consulta. En el caso de PubMed y PubMed Central, esta actividad diaria podría implicar que no se encuentren nuevos artículos y la consulta sea improductiva.

Una posible alternativa es recibir el índice de las revistas en las cuales se publica normalmente o preferencialmente el tipo de estudios en los cuales estamos interesados. Esto tiene como ventaja el estar no sólo actualizado respecto a la literatura impresa sino que incluso se puede tener acceso a los temas que la revista publicará en breve. La desventaja es que perderíamos aquellos artículos en los cuales por no tener el enfoque particular y reportar sólo avances metodológicos, no sean aceptables para la revista en cuestión.

Esta limitante podría tener una salida de búsqueda compulsiva equivalente a ir a la biblioteca al menos una vez a la semana y ver cuales son las novedades o, en la actual era tecnológica, recurrir a la búsqueda rutinaria de las fuentes electrónicas confiables (PubMed, PubMed Central, Scopus, etc) teniendo que iniciar una sesión cada vez que se haga la consulta haciéndolo tedioso y en la mayoría de los casos, frustrante. Una posibilidad es utilizar un servicio de actualización que nos haga la búsqueda rutinaria. Varias compañías han desarrollado diferentes buscadores de literatura con capacidades y costos diferentes. Sin embargo, la comunidad científica ha tratado de mantener una actitud más abierta y sin fines de lucro. Con este objeto, varios servidores ofrecen la posibilidad de ocupar un espacio en el cual se pueda recopilar la literatura específica de un investigador para después enviarle las características mínimas para que el investigador pueda tener acceso a ella, básicamente a través del correo electrónico.

#### Objetivo

Mostrar un sistema automatizado de actualización bibliográfica mediante un buscador remoto.

#### Desarrollo

En el inicio de su buscador, escriba la dirección http://pubcrawler.gen.tcd.ie

Después de que se visualice la ventana de acceso, registre la apertura de una cuenta ingresando los datos necesarios para ello (Figura 2.1).

Una vez creada la cuenta y dado de alta el buzón electrónico, introduzca los términos de la búsqueda. Asegúrese que en el despliegue del menú se seleccione "ALL FIELDS". Indique que el lapso de actualización como "WEEKLY"

Al terminar de introducir los términos de la misma manera que la sesión anterior, envíe la consulta. En algunos casos, la respuesta es inmediata por lo que puede abrir su dirección de correo para analizar las referencias que sean depositadas en su buzón.



	nttp://pubcrawier.gen.tcd.ie/
60 []] 🏭 Apple 4	shared Yahoo! Google Maps Wikipedia Noticias Populares Flow
	PubCrawler - an Update Alerting Service for PubMed and GenBank PubCrawler is a free "alerting" service that scans daily updates to the NCBI Medline (PubMed) and GenBank databases. PubCrawler helps keeping scientists informed of the current contents of Medline and GenBank, by listing new database entries that match their research interests.
Crawlor	Now users - join BubCrawley's Web services
	It's free Chose a username (race-sensitive) and click the Vale button
- you go to the pub(TM)	usemame:
About	
Download	Existing users - log in:
Sample Results	Otherwise, enter your username and password and click the 'log in' button.
News	usemame: <u>(Send password)</u> password: <u>(Log in)</u>
Contact	(Reset)
	Disclaimer and Copyright
	PubCrawler was developed and is hosted by <b>Ken Wolfe's lab</b> in the Genetics Department, Trinky College Dublin, Initiad, This service is free and comes with ABSOLUTELY NO WARRANTY. It is provided to the public in the hope that it is useful. PubCrawler queries databases provided by the National Library of Medicine, but we have no affailation with NCBI. Please take a look at <b>NCBI'S Disclaimter</b> for more information about their disclaimers and copyrights.
	Last modified at \$Date: 2008/10/23 22:08:58 \$ For question or remarka, please e-mail <b>pubcrawlerREMOVECAPShelp@gmail.com</b> .

Figuras 2.1a. Ventana de inicio - registro.

	PubCrawler - an Update Alerting Service for PubMed and Ge	nBank
	Overview for user 'hectorserrano' (log in as different user)	
	Results: The latest results contain 183 new hits. They were generated on Thu Aug 15 19:20:04 2013 (Eire).	alts) eries!
It goes to the library - you go to the pub(TM)	Queries:	
	3 queries are defined. Last modified on Change gu Wed Oct 17 04:21:37 2012 (Eire).	eriest
About	Schedule:	
Sample Results	Queries are carried out on Sundays. (Change set	edule
News	Mail and results options: Results are stored in Summary format.	tionel
FAQ	when new hits were found.	
Contact	Parameters: Database entries from the last 180 days are searched.	
	The 800 most recent hits are retrieved, of which 100 entries are shown in full. Any excess number of hits can Change para be retrieved through a link combining up to 400 items. Previous hits will be listed for 10 days.	meters
	Cookies:	ies
	Beenwordt	
	Click the button to change your password. Change pas	sword
	(Refresh!) (Log out)	

Figuras 2.1b. Desarrollo de perfil de búsqueda de PubCrawler.

Cada una de las entradas (QUERY o ALIAS) puede tener varias palabras clave. Si la investigación en la cual está inmiscuido requiere de la separación o agrupación de palabras clave, cada una de ellas puede obtenerse por separado.

Una vez acreditada la cuenta, puede ingresar a ella utilizando las características de usuario y clave de acceso para ampliar, modificar o suprimir las consultas y recopilaciones solicitadas (Figura 2.2)

A lo largo del curso, determine el impacto que tiene el campo de acuerdo al número de citas y la relación con el curso que tenga.

	<b>PubCrawler -</b> an Update Alerting Service for PubMed and GenBank Query configuration for user 'hectorserrano
Crawler Leges to the Brary you go to the Publith	Existing queries:     0 1 pubmed 'phyotestrogens' Phytoestrogen* (All Fields) OR     phytoestrogen* (All Fields) OR counestrol (All Fields) OR Estrogen     receptor (All Fields) OR Genistein (All Fields) OR Zearalonone     0 2 pubmed* Steroid synthesis' Steroid hormone (All Fields) OR     Aromatase (All Fields) OR Androgen* (All Fields) OR Progest* (All     Fields)
About Download	0 3 pubmed 'Placenta' Placenta (Ali Fields') OR Steroldogenesis (Ali Fields') OR aromatase (Ali Fields') OR StAR (Ali Fields')           Action for selected query:         ••••••••••••••••••••••••••••••••••••
Sample Results	New PubMed query: Alias: Deery Aliac: (replace with descriptive term)
Contact	search term search field connector AND = AND = AND = AND = AND = AND = AND = Same Search
	Type: © PubMed O PubMed Neighbour O Nucleotide O Nucleotide Neighbour Terms: 4 Charget

Figura 2.2a. Ventana para especificar las revisiones y cambios.



Figura 2.2b. Resultados acumulados por campo requisitado.



El sitio le permite analizar un máximo de 800 citas que pueden ser modificadas en cantidad, forma (puede almacenar los resúmenes de las diferentes citas recibidas, el resumen en formato gráfico, si está disponible, etc). Cada 10 días, se remueven las 100 citas más antiguas de cada uno de los campos de búsqueda requeridos. Sin embargo y con algunas limitantes, puede tener acceso a ellos utilizando la herramienta "PREVIOUS SEARCHES" que aparece una vez que ha accedido a los listados de resultados.

#### Actividades para reportar

A partir del portal de la Biblioteca o de PubMed, permita que al menos 10 revistas le envíen su tabla de contenidos. Dentro de ellas asegúrese de incluir Nature (varias de ellas ahora están especializadas), Science y Nucleic Acids Research y para una lectura ligera, Scientific American y La Recherche.

Haga una tabla en la cual muestre el número de artículos sobre el tema que seleccionó y la frecuencia con la cual obtuvo actualizaciones a lo largo del trimestre. Esta actividad será evaluada y discutida en el grupo.

## Referencias

- Hokamp, K., Wolfe, K.H. 2004. PubCrawler: keeping up comfortably with PubMed and GenBank. Nucl. Acids Res. 32:W16-W19. Disponible en la URL <a href="http://nar.oxfordjournals.org/content/32/suppl\_2/W16.full.pdf+html">http://nar.oxfordjournals.org/content/32/suppl\_2/W16.full.pdf+html</a>
- Ma'ayan, A. 2008. PubMedAlertMe- standalone Windows-based PubMed SDI software application. Comput. Biol. Med. 38:620-622. Disponible en la URL <u>http://www.ncbi.nlm.nih.gov/pmc/srticles/PMC2431148/pdf/nihms51645.pdf</u>



## Búsqueda y análisis de bases de datos

## Capítulo 3

# Obtención y análisis de secuencias de proteínas utilizando un servidor en línea

En los laboratorios de investigación en Genética Molecular, es necesario conocer los productos génicos con los que se está tratando, sobre todo si están asociados a patologías o características morfológicas que pueden determinarse de manera simple. Como se había planteado desde el inicio del curso, se utilizarán marcadores moleculares asociados a diferentes patologías. En una de ellas, el equilibrio hemodinámico proporcionado por la actividad de los receptores adrenérgicos se encuentra alterada.

Como producto del esfuerzo de la comunidad científica en conocer la secuencia de cada uno de los nucleótidos que conforman nuestro genoma, una gran cantidad de datos se han generado. Durante las últimas dos décadas, la información referente a las secuencias de nucleótidos han dominado el escenario científico. Sin embargo, durante una buena parte del siglo pasado, la función, regulación y constitución de las proteínas mantuvo ocupados a bioquímicos y fisiólogos de diversos campos.

Algunos de los mejores premios en Ciencia se otorgaron a grandes representantes del campo de estudio de las proteínas. Pocas personas han recibido no uno sino dos premios Nobel por el impacto e importancia de sus investigaciones. El primer premio Nobel de Linus Pauling sobre la estructura de una proteína fue seguido por el de Frederick Sanger sobre el desarrollo de una metodología para conocer la secuencia de aminoácidos de una proteína. Mientras Sanger obtuvo su segundo Nobel por su metodología de secuenciación de ADN, Pauling lo obtuvo por su activismo a favor de la paz.

Las publicaciones de las secuencias de aminoácidos crecieron mucho antes de que se tuviese un acervo de secuencias nucleotídicas importante. En ambos casos, la conformación de bases de datos conteniendo información estructural, de las secuencias de aminoácidos, de las funciones y variaciones de las proteínas de diferentes sistemas celulares ha permitido el avance de la investigación con frentes, metodologías y estrategias basadas en la esencia de la investigación básica.

#### Objetivo

En este ejercicio, haremos el análisis de los diferentes receptores adrenérgicos que se presentan en el humano y se encuentran depositadas en las bases de datos. El objetivo de este análisis es determinar los participantes proteicos de este tipo de proteínas, su alineamiento en cada aminoácido, las características topográficas locales y la posible relación filogenética existente entre ellas.

#### Desarrollo

En su buscador, ingrese la dirección <u>http://web.expasy.org/docs/swiss-prot\_guideline.html</u> Esto desplegará la página inicial de SwissProt and TrEMBL. En el segundo párrafo, active la liga UniProt Website lo que abrirá la página de Uni-Prot en la pestaña SEARCH (Figura 3.1).

Introduzca en la ventana de diálogo la palabra "Human adrenergic receptor" y pulse el botón SEARCH. Esto abrirá una página en la cual se encuentran diferentes proteínas asociadas. Marque los registros de las proteínas correspondientes a los diferentes tipos de receptores adrenérgicos de humano. Al final de la revisión de las páginas consta la búsqueda, pulse el botón "RETRIEVE". En este momento, abrirá la pestaña "RETRIEVE" del servidor en la cual dentro de una ventana aparecen los identificadores de las proteínas seleccionadas (Figura 3.2).





Figura 3.1a. Ventana del portal de inicio de Swissprot.



Figura 3.1b. Arranque del sitio UniProt.

				Trainan automigic receptor in control to		
> • O +	http://www.uniprot.org/uniprot/h	query=Human+adrenergic+re	<pre>ceptor&amp;sort=score</pre>		C Q+ Coople	
CD III Apple 4sh	nared Yahoo! Coosie Maps Wikip	edia Noticias * Populares */	Tiew			
						Developerty Context, Decument
100						
Search Diast	X Algn Rotrieve	ID Mapping *				
arch in	Query					
totelle Knowledgebase (DistProce)	.d (Human adveningic receptor	(freet) (Advanced Search +) (P				
5 of 164 results for Human II	AND advenergic II AND receptor II in Uni*	.ProtKB sorted by score descending	48			
drowse by taxonomy, keywor	and, gene ontology, enzyme class or pathway I	Reduce sequence redundancy f	to 100%, 90% or 50%			<u>07</u>
						Page 1 of 7
and the party of the local division of the						
KINS (CELICARD)						
row only reviewed (85) = (U	JNPhotK2kSwbs-Photj or unreviewed (ny) = a	(UniProRS/TrCMSL) entries				
vote terms: "advenenge recep	ptor"					
estrict term "nemail" to prove	A name (1), organism (78), taxonowy (78)	and the web research (1)				
Janet form "receptor" to per	while family (as), gene ontology (127), keywor	4 AMD (SS), was reason a (1),	aana 19			
And the second s	an an of the growth and the second	the second se	ue (a)			
Entry	Entry name	Status	Protein names	Luizi Cene names	Organiam	Leign
P35348	ADA1A_HUMAN		Alpha-1A adrenergic receptor	ADRA1A ADRA1C	Homo sapiens (Human)	
P08913	ADASA_HUMAN		Alpha-2A adrenergic receptor	ADRA2A ADRA/R ADRA/R	Hono sapiens (Human)	
P25100	ADA1D_HUMAN		Alpha-1D adrenergic receptor	ADRA1D ADRA1A	Hono sapiens (Human)	
P07550	ADRIS2_HUMAN		Beta-2 adrenergic receptor	ADRB2 ADR52R 52AR	Homo sapiena (Human)	
P18825	ADA2C_HUMAN		Alpha-2C adrenergic receptor	ADRAZO ADRAZIZ ADRAZRILZ	Hono sapiers (Human)	
P13045	ADRIBO_HUMAN	*	Beta-3 advenergic receptor	ADRES ADRESS BAAR	Homo sapiens (Human)	
P18089	ADA28_HUMAN	*	Alpha-38 adrenergic receptor	ADRA28 ADRA31 ADRA311	Homo sapiens (Human)	
P25368	ADA10_HUMAN		Alpha-18 adrenergic receptor	ADRA18	Homo sapiena (Human)	
POBS66	ADR51_HUMAN		Beta-1 adrenergic receptor	ADRD1 ADRD1R D1AR	Hono sapiens (Human)	
P25068	ARBRI_HUMAN		Beta-adrenergic receptor kinose 1	ADRBK1 BARK BARK1 CRK2	Hono sapiens (Human)	
P15823	ADA1B_BAT	*	Alpha-18 adrenergic receptor	Advatb	Battus norvogicus (Bat)	
P25062	ADR83_MOUSE	*	Bota-3 adrenergio receptor	Adribă Adribăr Böbar	Mus musculus (Mouse)	
P35626	ARBK2_HUMAN	*	Beta-adrenergic receptor kinase 2	ADRBK2 BARK2 GRK3	Homo sapiens (Human)	
P22044	ADA1D_RAT		Alpha-1D adrenergic receptor	Adrata Adrata	Ratus rorvegicus (Rat)	
08/670	AA2DA, DANNE	*	Alpha-2Da adrenergic receptor	edra2de	Danio rerio (Zebrafish) (Brachydanio rerio)	
05/069	AA208_DANRE		Alpha-2Db advenergic receptor	ebfarbe	Danio rerio (Zebrafish) (Brachydanio rerio)	
Q90WY6	ADA2C_DANRE	*	Alpha-2C adrenergic receptor	adražo	Danio renia (Zebrafish) (Brachydanio renio)	
GROWINS	ADASE_DANRE	*	Alpha-3B adrenergic receptor	dimbe	Danio renio (Zebrafish) (Brachydanio renio)	
P30545	ADA2D_MOUSE	*	Alpha-3B adrenergic receptor	Adra2b	Mue musculus (Mouse)	
Q20524	ADRIBS_MADAU		Beta-3 adrenergic receptor	ADRES DOAR	Macaca mulata (Rhesus macaque)	
Q60483	ADR83_CAVPO	*	Beta-3 adrenergic receptor	ADR83	Cavia porcellus (Guinea pig)	
Q90WY4	ADA2A_DANRE	*	Alpha-2A adrenengic receptor	edra2e	Danio renio (Zebrafish) (Brachydanio renio)	
Q801U1	Georul_DANRE	*	Novel protein similar to human adrenergic rec	adribi2 (2221D18.1-001	Danio renio (Zebrafish) (Brachydanio renio)	
E7EW16	E7EW16_HUMAN	*	Alpha-1A adrenergic receptor	ADRASA	Hono sapiens (Human)	
GROBIO	OSOBLO_MOUSE	*	Adrenergic receptor, alpha 1b	Adratb mCG_21004	Mus musculus (Mouse)	
	/					Page 1 of 2
				© 2002-2013 UniProt Consortium I License & Discibilitier I Contact		
				CMDI CDI MIN		

Figura 3.2a. Listado resultante de la búsqueda de receptores adrenérgicos.



Figura 3.2b. Depuración para humanos de UniProt.

Pulse el botón "ALIGN" localizado en el extremo inferior derecho de la pantalla. Esto abrirá la pestaña correspondiente en el servidor. Si el número de selecciones es pequeño, se abrirá un cuadro que contiene el código de una letra de los aminoácidos que forman las proteínas (Figura 3.3).

Al término de la revisión, la pestaña correspondiente se activa y muestra el alineamiento que tienen las diferentes proteínas seleccionadas. El programa de alineamiento presenta un algoritmo que optimiza la "llamada" del aminoácido y, cuando no existe un alineamiento adecuado, inserta un guión. En este punto es importante hacer notar las herramientas de análisis disponibles en la zona derecha del inicio de la secuencia alineada. En esta herramienta es posible localizar los diferentes tipos de aminoácidos o la presencia de modificaciones y dominios. En ejemplo mostrado en la figura 3.3, se ha pedido la localización de los dominios transmembranales que se ha resaltado en amarillo.

Si activa las ventanas de las características de los aminoácidos (polares, pequeños, hidrofóbicos, etc) se irán marcando en diferentes colores los aminoácidos que cumplan con los requisitos. Todo esto permitirá hacer el análisis de las diferentes regiones que conforman las proteínas y que podrían estar asociadas con la función.



Augment (competed)	
od 💭 🎹 Apple Ashared Yahool Google Maps Wildpedia Noticiasy Popularesy Flow	
upperf) - Allen	Downloads - Contact - Documentation/Help
Overen Della Alla Parte Contra	
appointing of Polar A contract of the Contract Contract of Contract Contract Of Contract On Contract O	Help
Value and A state of the state	To align several protein sequences, enter:
WHICH AND ADDRESS ADDRE	<ul> <li>two or more sequences in PAGLA format</li> <li>two or more UniProt identifiers, e.g.:</li> </ul>
Summi softwork water and a software and a softwar	17A_3019AS
	More_
	text tree fasts jaiview
12 <sup>1</sup> Alignment : Tree : Annotation : Job Information (Dualization and an an	
Alignment	
	head at the
There have to print this alignment in color	Chidities verifies
0 P2204 AddA_328A3	Dista straid
1 117111111717101970110405010400010404471004700171970400000114044710171041001719710400001140441371044	Decleotide Minding
1	() Two
1	Clicitize site
E 27255 ANAL TRANS	Cheiz
1	Citate Citate had
1	Chitageoseis
6 (0777) Walk	Chain
1 TETALLY DEVELOPMONDANCE AND THE INFORMATION FOR A CONTRACT OF A CONTRACT	() Hetoral variant
1 19 PROF. POST.	C Dunaia
1 TPL: MALECER TO PAPER DALLOF THOUGH IN TO T 77534 ANALA JEED	Cho-terminal residee
8	Chipidation
14	OBALS
29	Compositional bias
9 INTRAMULTINGTOWPTORT-UPUTTATION UPUTTATION (1) / 2 P2000 AND 2000 AND	Clegins
The second	Awino acid properties
2	Cliniarity
(1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (1)     (	Depative
1) PUIDPEDARTO	Cleashave
20	() tay
40 ILTELEBE-LEPTETTERATALLITETTETF-ADDILITETTETEF-ADDILITETTETF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTETEF-ADDILITETTEF-ADDILITETTEF-ADDILITETTEF-ADDILITETTEF-ADDILITETTEF-ADDILITETTEF-ADDILI	Caronatán
55 IIATTBAUKNEUUT-ANELIATUDF-LAINENTINGS 15 19993 ARAA READ 18 IIATEBAUNIVERTINGANELIANELI-FARTANINAN 16 2210 ARAA READ	[]Smll
25 TALATED	[]Polar
27 DAILNTDUCDTTTT:SAAACU	Cierciae Threeniae
67 III/III/III/III/III/III/III/III/III/II	
7 TATTERA-DARDETARACEN-ATDEFACEORDATION 12 MARCHINAL AND ADDRET 12 MARCHINAL ADDRET 12	
9 1.1116/00/000171811/0-0401121011-0401121001040017- 9 10.111111 - 901121001040017- 9 10.111111 - 90112111111-0401121001040017- 9 10.111111 - 90112111111-0401121001040017- 9 10.1111111 - 90112111111-0401121001040017- 9 10.11111111-0401121001040017- 9 10.1111111-0401121001040017- 9 10.111111-0401121001040017- 9 10.111111-0401121001040017- 9 10.111111-0401121001040017- 9 10.111111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-0401121001040017- 9 10.11111-04011210010400017- 9 10.11111-04011210010400017- 9 10.11111-0401121000400000- 9 10.1111-04011200000000000000000000000000000	
11 2016/00/	
27 JAARTING AND THE AND A	
99	
16	
16	
110 CONTRACTOR CONTRAC	
11 CONTRACTORISTICATED 14 PONA ALLEPENA CONTRACTORISTICATED 14 PONA ALLEPENA 10 CONTRACTORISTICATED 18 PONA ALLEPENA	
24 INFORMATION AND A CONTRACT AND A	
10 CETTOTACTALITICALS IN NO.	
10 CHEVEN CONTRACTANTICALARS 10 HOUR MEAN	
15 8120171	
10 1995TO	
[3] BTATETEOLLINGSFTLUTETSET 14 TOTAL ABBL/DBAT BETWEEN	
105 FLATTEFLA-FRAITENCATIONTITT	

Figura 3.3a. Ventana de alineamiento.

R 🖸 🖯 🕹		
A b P O t mtp://www.uniprec.org/align/2013081540078UG4Y2	Exercise C Q+ Coople	
100 🌐 🖽 Apple Ashared Yaheot Google Mapa Millipedia Neticiaa+ Popularea+ Flow		
unnu j Algn		Downloads Contact Documentation/Help
Services In EATA Assess (Tables) Review ID Mapping		
Sould'STAREADAA KAMAN Agaba-La Admengiji: exceptor Os-Henro sagares 👘 Align		Help
Di-Addia R-1 W-2		<ul> <li>two or more sequences in FASTA format</li> </ul>
SIGNED CONTROL OF AN ALL AND A		TPA_MEAN
		TPA_FID Non
PC- Alexenser - Trans- Anseteries - Adv Information (Definition and P		Soot tree fasts jakiev
Alighment		
When her to print this slipsest in color	Indiziei veide	
1	Divis straid	
1 HTTPELL IS STREAMED SEALON CONCUMPANT OF A STATE OF A	OAlternative segmence	
1	Chotise site	
1 1001100- 001100-001000000000000000000	C Boost	
1	Discullide bood	
1	C thit operation	
t HTTPELL IS STREPTISED SECONDAL CONFINCTION OF THE STREPTISED SECOND STREPTISED S	f Trapmenbrum	
1	Densin Officers/ation	
1	Son-terminal residue	
61 SECRETSENDED	Disployical domain	
2	Compositional Man	
9 INTERANTIFICATION CONTINUES CONTINUES IN CONTINUES (CONTINUES) (CONTINUES) (CONTINUES (CONTINUES)	Degian	
1	industry	
Introduction of the second sec	Directive	
1 I HEORIS GALERES - ANALOCINOLATION TRADUCTIVATIAN MAILL   1   1221   1221   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1222   1	Doublive Tableback	
	Office	
40 LITEGORGA-LEVIELTETTTTTTTTALAULU	D Charged	
2 INLATE-LOTTETTTTTLANDU	Delar	
99 TALGETER-LAUGE LEVELAALTE	Stig Source Threatine	
07 ILTRACTOR-LICTOTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT		
40 LUPACON-LUPITUTUTALALLI-TITUTUT-LUPACUTUN997- 91 12014 LUPACUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTUTUT		
41         1.17#COBLETERTITIATALILIFITTUTI-ATATU/WAV977         9         10221         BEIED, REAL           5         2.0722         GBIT2, SERAL         5         2.0722         GBIT2, SERAL		
19         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         19.4         1		
16 CILLATIVITSTICALED 19 1991 4442 PRA		
16 CETEVICIAL DI PIESO ANO DI PIESO ANDE DE		
10 million of the second secon		
P		
110 CHARTER CONTRACT CONTRACT OF A DESCRIPTION OF A DESCR		
154 STEMTT		
121 NTCHT		
110 STANTE		
		(1000) (100) (100) (100)

Figura 3.3b. Dominios transmembranales del extremo amino terminal de los receptores adrenérgicos humanos.

#### Actividades para reportar

Utilizando el alineamiento del receptor adrenérgico humano, identifique los residuos hidrofóbicos, dominios polares, variantes y sitios de glicosilación. Identifique los residuos de aminoácidos polares pequeños, de lipidización y los dominios funcionales que pueda presentar el receptor.

Haga un análisis similar utilizando por separado los diferentes tipos de receptores adrenérgicos resaltando las posiciones donde tenga homología.

Repita el análisis utilizando las secuencias de Calpaína 10 humana y los receptores estrogénicos humanos.

## Referencias

- O'Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R. 2002. High-quality protein knowledge resource: Swiss-Prot and TrEMBL. Brief. Bioinf. 3:275-284. Disponible en la URL <a href="http://bib.oxfordjournals.org/content/3/3/275.full.pdf+html">http://bib.oxfordjournals.org/content/3/3/275.full.pdf+html</a>
- Uniprot Consortium. 2012. Reorganizing the protein space at Universal Protein Resource (UniProt). Nucleic Acid Res. 40:D71-D75. Disponible en la URL <u>http://nar.oxfordjournals.org/content/40/D1/D71.full.pdf+html</u>



### Actividad 2

#### Analisis de homología y evolucion de grupos proteicos

Además de comparar e ingresar nuevas secuencias, la labor de los investigadores requiere el conocimiento de las alteraciones que puedan presentarse en una proteína o enzima que afecta una función particular de la célula o tejido que se está estudiando. El estudio detallado de este tipo de alteraciones y que se realizó durante una buena parte del siglo XX, se ha facilitado por el empleo de las metodologías de ADN recombinante y el posterior desarrollo de algoritmos que permiten su alineamiento, como lo vimos en la actividad anterior.

Es posible seguir el camino evolutivo o divergente de una proteína o familia proteica utilizando la información contenida en su propia constitución. El marco teórico en el cual se encuadra la presunción básica es que si cada una de las proteínas realiza una función específica en un organismo determinado, en el origen de ella, el proceso de mutación que afecta al gen que la codifica podría haber afectado una región no importante para la función de la proteína. Si al mismo tiempo el gen codificante se duplicó, el efecto de estas mutaciones y la selección natural en cada uno de los genes sería diferente: mientras uno fuese fuertemente seleccionado para sufrir cambios en lugares diferentes al sitio activo o de importancia para la función proteica, el otro tendría una presión de selección menor. La acumulación de los cambios en cada uno de los genes a lo largo de la historia evolutiva permitiría que si bien se mantienen características similares, existan diferencias que permitan no solo distinguir los genes que codifican a estas dos proteínas, sino de las mismas proteínas que permitirían identificarlas como distintas.

Uno de estos escenarios podría implicar, por ejemplo, que una mutación interna en la región codificante de un exón ensamblara un codón de término. La proteína producto de este gen sería más corta que la proteína homóloga. Si el residuo de la proteína alterada implica, por ejemplo una región de regulación alostérica, esta versión de la proteína no respondería a este tipo de regulación, por lo que la función sería diferente o se llevaría a cabo en un lugar diferente.

Mediante el uso de las herramientas desarrolladas a partir de las bases de datos, es posible hacer una estimación y simulación de este tipo de eventos.

#### Objetivo

El objetivo de esta actividad es mostrar la forma de analizar las relaciones entre diferentes miembros de las familias proteicas cuyas secuencias se encuentran depositadas en los bancos de secuencias como UniProt.

#### Desarrollo

En la página inicial de SwissProt (http://www.uniprot.org) direccione hacia la página de búsqueda (SEARCH del punto 3 de la actividad anterior) introduzca "Human Adrenergic Receptor" en el cuadro de diálogo; despliegue el cuadro de menú de la base de datos localizada a la izquierda del cuadro de diálogo y seleccione "SEQUENCE CLUSTERS (UniRef)" como base de datos para la búsqueda (Figura 3.4a). Pulse el botón SEARCH para desplegar la hoja de resultados en la pestaña SEARCH.





Figura 3.4a. Ventana de inicio de búsqueda con el menú desplegado.

🖲 🔿 C	,				Human a	drenergic receptor in Unitef				
4 F	• 0 + ) http://www.unic	arct.org/uniref/?	query=Human+adrenergic+receptor&sort=score			C Cr Coople				_
70 QQ	III Apple Ashared Yahool Got	ogle Mapa - Wiki	oodia Noticiaa+ Popularea+ Row							
entra)	+ UniRef						Downloads -	Contact - Dec	our en aronn	-
50	arch Disat. Algo	Retieve	ID Mapping							
Search	in Query									
Sequen	or Charters Khilde() [2] Haman adnor	ergic receptor	(horn) (Monand Joseph Class)							
- 25 of 4	Prosults for Human 1: AND advancegic 1: AN	ID receptors in Un	Perf sorted by sears descending to							_
in Brons	e by toxonomy I up Map to Unificated or Unifi	ake .							Contra La	2
-								age -	ALC: NO	7
Results	Custovice									
- Show o	nly clusters with an identity of \$275, \$275, 1027	s								
Ouste 1	erme: "achenergic receptor"									
Restric	Harm "human" to cluster name (1)									
- mepear	Sector in University (194									
	- Cluster ID	Statua	Cluster name	Size	Cluster member(s)	Coper lane	Longth	Ident	thy	
					ASTUSO	Rome speers (Network) Ganila polita (Schend polita)				
					H2P540	Penge aboli (Sumatan caraputar) (Pongo pygmanux aboli) Macaaa Sacularia (Cara) - Andreading macaputar (Comonigae noniney)				
0	Unifiel50,(P18088	*	Cluster: Alpha 25 adrenongic receptor		205 G7PMD1	Cultifui jacohus (Wilde-sufad-sizr mannoset)		450		66
					UPRO02855E5A	Basis servita (Physical International) Managementary (Physical International)				
					UPI00027431CF	Fells catus (Cat') (Fells silvestris catus)				
					+09 P05505	Hore splere (rieman)				
					99038 05030	Bythesis construct Maccac methods (Phase macagos)				
					P47999 UP900000000000	7 Normanicus Inaccentrus (Northern white sitested object) 5-https://www.inaccentrus/				
0	Creeking, young		Custor: bitts-1 solarwige Neeptor		G150R9	Past toglodyna (Chinpana e)		***		٦I
					UPI0001D37PPB	Massie gebruik Ann (European domestie ferret) (Musicia fant)				
					-49	eas sanak (rig) 430 -				
					P25008 H20482	Hone support (human) Pan toodorkin (Chimaeted				
					10MDH3	Epermanetas Extenseminantes (Thistee-Intel spond suivite) (Effetiniys Milocentineatus) Encon telli (Il instatus consent Alfonco naturates a tellio)				
0	United Pasos	*	Cluster: Beta-adranenzia receptor kinase 1		IN HIPCIS	Macaca mulatis (Phense macagos)		680		~
					UPRODUBILIZED 7	Promotogradus guore (Autori mole na)				
					P21146 J909902	Des teuras (Devine) Tankips huncelas (Allantic bottle resed dobhit) (Delphinus truncatus)				
					+126	x87 Home successful Humani				
					064.006	9 Des tradición (Olimonana)				
					P38348-2	Makea mulate (Phenios makeaun) Genetia exercita (Phenios makeaun)				
0	UnRel50_P05348	*	Cluster: Apha-SA adrenengic receptor		DOG47N	Califini (pote (pha (construction)) Califini (pote (pha (construction))		466	5	8
					8084P2	Madada Isabalia ( ) - Bo-bang Madaga ( ) ( - yinomgas Intervery ) Nomacia se locaciony ( ) ( Northern white-Island gibban ( ) ( - yinotatis locaciony) (				
					-91	Mustel patrian fure (European domesic forest) (Mustels fure)				
					P07550	Home stapiers (Nerman) Analise and the control control of the cont				
					HORUYE	Tupin changing (Chinese year show) Tupin changing (Chinese year show)				
0	UniPutto Pozsia	*	Cluster: Beta-2 advancesis recentor		48 LERINDS	Crysteliger centralius (Rebb)		413	5	
					G1TAV7	Autopola relational planta (Philae-lined ground aquited) (Ictionrys tridocentinastus)				
					C2HBA/7 KIMUE3	7 Macaon multite (Phreus macague)				
					48	420 And				
					H20462	Post togeting (Partogeting (Par				
					P21146	Bes tarris of occurs in the second seco				
0	United0_P25038	*	Cluster: Bets-advancegic receptor kinase 1		53 F1N723	Turbop functions (where both-could objivit) (Deprivating functions) Mesocrites survives (Society homesor)		680		6
					COMME	Mas musices (Movie) Bates non-exists (Mat)				
					COULAR COULAR	Conit familiaria (Dog) (Canit lupus familiaria)				
					P07550	Home sloping (Homan)				
					ROBUMB	Tapala china (Lonas Lonas				
	100000 00000		Cluster Date 2 winnersis seconds		LINDICO LINDICO	Une sees (sones) - Control (Passo) - Control (Pa		412		
~		- C -	and the second se		BIGG28 GITAV7	(Alaropoda relandeuca (diam parda) Seemendika tidoceminatasi (Thitsee-Ined second soult of (cidomya tidoceminatus)				
					COHEN/ CONUES	Const limiters (Dog) (Carls Live Events) Macalan mixed (Parks Indexa)				
					dt					
					01060	Police Electric (Internal) Mariana Francisca (Columbia) Advancementaria (Columbia)		_		

Figura 3.4b. Resultados agrupados bajo el criterio UniRef.

			-
കമ	Apple 4s	Abrocoma bennettii (Bennett's chinchilla rat) (2)	1
		Cuniculus (2)	
UniProt	) → UniRef	Cuniculus taczanowskii (Mountain paca) (Agouti (1)	
••		Cuniculus paca (Lowland paca) (Agouti paca) (1)	
S	earch Bla		
J Č		Trichys fasciculata (long-tailed porcupine) (1)	-
Search	in in	Atherurus macrourus (Asian brush-tailed porcupine) (1)	- 11
Conus	nco Clusters (UniPof)	Ctenomyidae (tuco-tucos) (3)	-
Caeque	nce clusters (officer)	Octodontomys aliroides (Mountain deau) (2)	24
01	Alacha OD ada	Ctenomus boliviensis (Bolivian tuco-tuco) (1)	
Cluste	r: Alpha-2B adre	Echimuidae (American spiny rats) (2)	- 11
Publishee	d July 24, 2013	Honoremus aumausus (Armored rat) (Echimus aumausus) (1)	
		Echimur chargeners (White faced spinu tree rat) (1)	
Built on	seed sequence UPI00	Determine control of the set of t	n!
Dumon			
(10)- EII	ter · Members · Sequ	Cryptomys damarensis (Damaraland mole rat) (Fuk (1)	
		Bathyergus suillus (Cape dune mole rat) (1)	
		Sciurognathi (37)	
Filter			
Filter		Ctenodactylus gundi (Northern gundi) (1)	
1.250	f 236 members from 1	Dipodidae (jerboas) (11)	
D-40		Euchoreutes naso (Long-eared jerboa) (1)	
Datas	<b>UL</b>	Dipodinae (8)	
UniPr	ot (236) 🔹		(
		Dipus sagitta (Northern three-toed jerboa) (2)	
Memb	ers Customize	laculus (4)	
ment	one one of the office	laculus jaculus (Lesser Fountian Jerhoa) (2)	
		laculus blanfordi (1)	
	Cluster member(s)	Cardiographica (2)	5
		Californiae (2)	
	P18089	Salpingotus koziovi (1)	
			-
0	A2RUS0	Muroidea (24)	
-		Calomyscus mystax (Afghan mouse-like hamster) (1)	
	CODMM	Nesomyidae (6)	
	Ganivivo		
	H2P540		
		Macrotarsomys ingens (greater big-footed mouse) (1)	_
	G7PMQ1	Hypogeomys antimena (Malagasy giant rat) (1)	
<u> </u>	and man	Eliurus sp. C24 (1)	
-		Brachytarsomys albicauda (White-tailed rat) (1)	
	F/IPN4	Muridae (10)	
		Acomys (spiny mice) (2)	
	UPI0002655E5A	Acomys (spiny mice) (2)	
		Acomys rabisinus (Colter spiny mouse) (40	_
	LIPI00027EB93B	Musicas (C)	
	011000211 0000		
-		Mus musculus (Mouse) (4)	
8	UPI00027431CF		
		Otomys angoniensis (Angoni vlei rat) (1)	-
	UPI00027F4E8B		
	G35902	Dipodillus dasyurus (Wagner's dipodil) (1)	
	000002	Cricetidae (7)	
		Neotominae (2)	
	UPI0000047181	Peromyscus polionotus (Oldfield mouse) (1)	
		Neotoma fuscipes (Dusky-footed woodrat) (1)	
	Q38PT0	Arvicolinae (2)	
	0961.06	Microtus socialis (Social vole) (1)	
	GBGL00	Cricetinae (hamsters) (3)	
-		Mesocricetus auratus (Golden hamster) (2)	
8	G7NAM7	Cricetulus griseus (Chinese hamster) (Cricetulu (1)	
		Primates (17)	
8	Q0W9D8	Otalemur asreattii (Small, aared galage) (Careat (7)	-
-		Similarment (15)	
-	MOVAET.	Similtormes (15)	
	MOATE?		
		Cercopithecinae (4)	-
	Q9GL17	Cercopithecus solatus (Sun-tailed monkey) (1)	
		Macaca (macaques) (3)	
Ð	F7IPB3	Macaca mulatta (Rhesus macaque) (2)	
-		Macaca fascicularis (Crab-eating macaque) (Cyno (1)	
	0.501/00	Hominoidea (7)	
8	G5GVH6		
8	UPI0002C58146		
		Homininae (5)	
	LIPI0003346C07	Gorilla gorilla gorilla (Lowland gorilla) (2)	
	0110000040007	Homo saniens (Human) (3)	
-	0.001	Platurchini (New World monkow) (4)	
	G5CV17	Callithriv incohur (White tufted and marmoret) (3)	
		ATPIES 1/1	

Figura 3.4c. Receptores adrenérgicos y la selección del filtro para primates.

En esta pestaña, se encontrarán los grupos de proteínas agrupadas en relación a la identidad de los residuos de aminoácidos y longitud de la homología (Figura 3.4 b). Active la liga que aparece como "UniRef50\_ P18089" que abrirá la página de resultados parciales. En el listado del filtro, seleccione "PRIMATES" y active el botón "FILTER" para mostrar solamente los resultados de los diferentes primates con que cuenta la base de datos. (Figura 3.4 c)

Cuando aparezca la página de resultados, active el botón "ALIGN" para obtener los resultados con el algoritmo CLUSTALW para análisis de grupos (cluster analysis). Este algoritmo evalúa el grado de identidad, homología y regiones de cambio. Para ver los detalles de las distancias euclideanas entre los diferentes nodos, utilice la herramienta "TREE" que se encuentra al inicio de la zona de alineamiento (Figura 3.5). Coloree los diferentes tipos de aminoácidos y obtenga las relaciones filogenéticas y distancias entre las diferentes proteínas.







```
Image: Image:
 & 💭 🎹 Apple 4shared Yahoo! Google Maps Wikipedia Noticias 🔻 Populares 🔻 Flow
sp | P25098 | ARBK1_HUMAN: 0.0263889
tr|Q59F73|Q59F73_HUMAN:0.0263889
):0.417971
sp | P18089 | ADA2B_HUMAN: 0.271585
sp | P08913 | ADA2A_HUMAN: 0.262917
sp|P18825|ADA2C_HUMAN:0.0427807
tr | D6RGL0 | D6RGL0_HUMAN: 0.0427807
):0.220137
):0.00866759
):0.111746
sp | P25100 | ADA1D_HUMAN:0
tr BOZBEO BOZBEO HUMAN:0
):0.270575
sp | P35368 | ADA1B_HUMAN: 0.258446
tr | B02BD3 | B02BD3_HUMAN: 0.0232647
sp | P35348 | ADA1A_HUMAN: 0.00468384
tr|E7EW16|E7EW16_HUMAN:0.00468384
):0.0185808
):0.235181
):0.0121288
):0.0917689
tr | B8X6N1 | B8X6N1_HUMAN: 0.00183824
sp | P13945 | ADRB3_HUMAN: 0.00122549
tr | B8X6N0 | B8X6N0_HUMAN:0.00122549
):0.000612745
):0.28125
sp | P07550 | ADRB2_HUMAN: 0.256659
sp | P08588 | ADRB1_HUMAN:0.256659
):0.0264297
):0.0792553
):0.0209878
):0.0610283
```

Figura 3.5b. Resultado numérico para la distancia entre los nodos para el mapa.

#### Actividades para reportar

Haga un análisis similar al que se mostró en la actividad 1 pero utilizando los receptores alfa y beta adrenérgicos por separado. Utilizando los resultados del árbol filogenético de CLUSTALW, indique el posible origen de cada tipo de receptores y las posibles razones que pudieron haber originado los diferentes tipos y variedades.

Haga un análisis similar al ejercicio 2 para los receptores adrenérgicos alfa y beta por separado incluyendo el grupo de los mamíferos. Coloree e imprima los resultados que obtenga al activar las regiones conservadas. Con los datos del árbol filogenético, indique la distancia que existe entre el humano y los roedores, en particular con la rata de laboratorio.

Utilizando esta herramienta, analice los siguientes marcadores: PPAR alfa, PPAR gamma y CAPN10.



## Referencias

- Pearson WR. 1997. Identifying distantly related protein sequences. Comput. Appl. Biosci. 13:325-332. Dispoible en la URL <a href="http://bioinformatics.oxfordjournals.org/contents/13/4/325.full.pdf+html">http://bioinformatics.oxfordjournals.org/contents/13/4/325.full.pdf+html</a>
- ThompsonJD, Higgins DG, Gibson TJ. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignement through sequence weighting, position-specific gap penalties and weigh matrix choice. Nucleic Acid Res. 22:4673-4680. Disponible en la URL <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/pdf/nar00046-0131.pdf">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/pdf/nar00046-0131.pdf</a>

## Capítulo 4

# Simulación de la estructura de genes a partir de la secuencia de aminoácidos

En la mayoría de los casos, la obtención de un gene proviene de la secuencia de aminoácidos de un fragmento de la proteína que nos interesa. Actualmente, el uso de las herramientas bioinformáticas nos permiten no sólo hacer la búsqueda de la proteína que nos interesa, sino también determinar el grado de conservación que presenta en diferentes organismos, como se hizo en el ejercicio de la sesión 3.

Partiendo del hecho que se conoce la proteína que se pretende estudiar, el análisis experimental puede hacerse utilizando una estrategia definida. Si quisiéramos hacer un mapeo fino de la proteína para determinar los parámetros básicos de su función, por ejemplo debida a la presencia de un dominio de aminoácidos polares que le den un dominio hidrofílico con una estructura tridimensional específica para que se exponga un residuo lateral de serina del cual depende la función, una estrategia adecuada sería clonar específicamente o de manera restringida la región en cuestión y expresarla en cantidades abundantes o de manera constante para realizar los experimentos en cantidad y número necesario para obtener resultados confiables.

Usualmente, la clonación de genes o fragmentos de genes requiere el uso de sitios de restricción específicos tanto en el vehículo molecular (plásmido, bacteriófago, cósmido, etc.) por lo que es necesario conocer el sitio potencial de clonación así como aquellas enzimas de restricción que pueden ser utilizadas en el análisis o mapeo fino del gene y así diseñar alelos modificados (mutantes), que carezcan de una región (con deleciones) o que contengan un codón de paro al inicio que evite ser traducido.

#### Objetivo

Que el alumno pueda desarrollar una estrategia de investigación de un gen a partir del conocimiento de su secuencia de aminoácidos

#### Desarrollo

Obtenga la secuencia de la proteína aromatasa de la oveja (Cytochrome p450 19A1) utilizando el código de acceso NP\_001116472 en la base de datos PROTEIN del NCBI (<u>http://www.ncbi.nlm.nih.gov/protein</u> Fig 4.1) o Q9XS28 de UniProt (<u>http://www.uniprot.</u> org Fig 4.2).





Figura 4.1a. Ventana de acceso a Protein del NCBI.

Satari A	Archivo Edición Visualización Historial Favoritos Ventana Ayuda	
	cytochrome P450 19A1 (UVIs aries) - Protein - NUBI	t) On Granda
	Annie Ashared Vahoel Google Mans Wikinedia Noticias z Populares z Elow	C Coogle
S NCBI Re	sources 🖂 How To 🖂	Sign in to NCBI
Destain		
Protein	Protein	Search
	Advanced	Help
Display Settin	gs: ☑ GenPept Send to:	Change region shown
cytochro	ome P450 19A1 [Ovis aries]	Customize view
NCBI Referen	ce Sequence: NP_001116472.1	
FASTA Grap	<u>bhics</u>	
<u>Go to:</u> 🕑		Run BLAST
LOCUS	NP_001116472 503 aa linear MAM 18-APR-2013	Identify Conserved Domains
ACCESSION	cytochrome P450 19A1 [Ovis aries]. NP 001116472	Highlight Sequence Features
VERSION	NP_001116472.1 GI:172088165	Find in this Sequence
KEYWORDS	RefSeq.	
SOURCE	Ovis aries (sheep)	Articles about the CYP19 gene
ORGANISH	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia: Eutheria: Laurasiatheria: Cetartiodactyla: Ruminantia:	Aromatase immunoreactivity in fetal
	Pecora; Bovidae; Caprinae; Ovis.	The ovine sexually dimorphic nucleus
AUTHORS	1 (residues 1 to 503) Roselli,C.E. and Stormshak,F.	aroma [J Steroid Biochem Mol Biol. 2010]
TITLE	The ovine sexually dimorphic nucleus, aromatase, and sexual partner preferences in sheep	DNA methylation and chromatin accessibility of th [Mol Reprod Dev. 2008]
JOURNAL	J. Steroid Biochem. Mol. Biol. 118 (4-5), 252-256 (2010)	See all
REMARK	GeneRIF: The study used the domestic ram to study the role of	
	aromatase for the sexual differentiation of the brain in this long gestation species and, in particular for the development of	Identical proteins for
	male-typical sexual partner preferences.	NP_001116472.1
REFERENCE	2 (residues 1 to 503)	RecName: Full=Cytochrome P4t[Q6YI21]
AUTHORS	Mura,A., Gadau,S., Lepore,G., Balzano,F., Zedda,M., Mura,E. and Farina V	aromatase cytochrome P450 [AAN23836]
TITLE	Expression and distribution of P450-aromatase in the ovine	RecName: Full=Cytochrome P4[Q9XS28]
JOURNAL	hypothalamus at different stages of fetal development Neuro Endocrinol. Lett. 31 (5), 690-699 (2010)	See all
PUBMED	21173749	
REMARK	GeneRIF: In conclusion, the sheep can be included among the species where aromatase is highly expressed in the hypothalamus during the	Reference sequence
	whole period of sexual differentiation.	information
AUTHORS	Lepore,G., Gadau,S., Mura,A., Zedda,M. and Farina,V.	RefSeq mRNA See reference mRNA sequence for the
TITLE	Aromatase immunoreactivity in fetal ovine neuronal cell cultures	CYP19 gene (NM_001123000.1).
JOURNAL	Eur J Histochem 53 (4), E28 (2009)	
PUBMED	22073360	More about the gene CYP19
REPARK	cultures exposed to oxidative injury.	CYP19 gene
REFERENCE	4 (residues 1 to 503) Vancelay I Solimum P, and Furbace P	Also Known As: CYP19A1
TITLE	DNA methylation of placenta-specific Cyp19 promoters of cattle and	
1	sheep	<b></b>

Safari Ard	hivo Edición Visualización Historial Favoritos	Ventana	Ayuda
		UniProt	
	1 + _) http://www.uniprot.org/		C Q Google
⇔ []] IIII Ap	ple 4shared Yahoo! Google Maps Wikipedia Noticias 🔻	Populares *	Flow
UniProt			Downloads · Contact · Documentation/He
Search	Blast Align Retrieve	ID Map	oping
Search in	Query		
Protein Knowle	dgebase (UniProtKB)	Search	h) (Advanced Search ») (Clear)
WELCOME			NEWS
The mission of Ur	iProt is to provide the scientific community with a		
comprehensive, h	igh-quality and freely accessible resource of protein sequence		UniProt release 2013 08 - Jul 24, 2013
and functional info	ermation.		Girls just want to have IFNE I Cross-references to GeneWiki I
What we provid	de		Changes to cross-references to GlycoSuiteDB to UniCarbKB I Removal of cross-references to GermOnline
UniProtKB	Protein knowledgebase, consists of two sections:		Statistics for UniProtKB: Swiss-Prot · TrEMBL
	Swiss-Prot, which is manually annotated and		Forthcoming changes
	TENDI shish is showing the		INEWS ACCIVES     ME Follow Runippot 648 followere
	TrEMBL, which is automatically annotated and is not reviewed.		Pollow euniproc 040 tolowers
	Includes complete and reference proteome sets.		SITE TOUR
UniRef	Sequence clusters, used to speed up sequence similarity searches.		
UniParc	Sequence archive, used to keep track of sequences and their identifiers.		Professional and the second se
Supporting data	Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.		Market and a set of the set
Getting started			A - Sea Mark     Here     The sea Mark     Here
<ul> <li>Text search</li> </ul>	h		
<ul> <li>Sequence</li> </ul>	similarity searches (BLAST)		Learn how to make best use of the tools and data on this site.
Sequence	alignments		
<ul> <li>Batch retri</li> <li>Database i</li> </ul>	eval identifier mapping (ID Mapping)		PROTEIN SPOTLIGHT
			the root of the problem
			July 2013
			And we all go about it in the most ingenious ways. Humans go
			hunting in supermarkets. Dogs wait for food to appear in their
			ooma, moaquitoes suck bioou, mains seep in igni
	© 2002–2013 UniProt Consor	rtium I Licens	e & Disclaimer I Contact
	EMBL-EBI	🔹 😰	SIB
	CMDL-EDI	• 🛩	SIB

Figura 4.1b. Desplegado de la aromatasa de oveja.

Figura 4.2a. Venatana de acceso a UniProt.

🕷 Safari Archivo Edici	ión Visualización Historial Favoritos Ventana Ayuda
00	Cytochrome P450 19A1 - Ovis aries (Sheep)
🔹 🕨 🕐 🔴 🕂 🗅 h	ttp://www.uniprot.org/uniprot/Q9XS28
😔 💭 🎹 Apple 4shared	Yahoo! Google Maps Wikipedia Noticias Populares Flow
UniProt > UniProtKB	Downloads - Contact - Documentation/Help
Search Bla	ast * Align Retrieve ID Mapping *
Search in	Query
Protein Knowledgebase (UniPro	otK8) 🗘 (Search) (Advanced Search ») (Clear)
Q9XS28 (CP19A_SHEEF Last modified April 3, 2013. Ver	)      Reviewed, UniProtKB/Swiss-Prot     Send feedback     Gend feedback     Reviewed, On add your own     Send feedback     Gend feedback     Send fee
Clusters with 100%, 90%, 8	50% identity I 🕒 Documents (1) I 💷 Third-party data 🛛 🔤 text 🛛 xml 🔤 rdif/xml 🔤 ff fasta
Names · Attributes · Gen Customize order	eral annotation · Ontologies · Sequence annotation · Sequences · References · Cross-refs · Entry info · Documents
Numes and origin	
Cone comen	Cytochrome P450 1941           EC=1.14.4.4.1           Alternative name(s):           Arromatase           CYPXIX           Cytochrome P450AROM           Estrogen synthase
Gene names	Name: CYP19A1 Synonyms:CYP19
Organism	Ovis aries (Sheep)
Taxonomic identifier	9940 [NCBI]
Taxonomic lineage Protein attributes	Eukaryota · Metazoa · Okordata · Caraniata · Vertebrata · Euteleostomi · Mammalia · Eutheria · Laurasiatheria · Cetarilodactyfa · Ruminantia · Pecora · Bovidae · Caprinae · Ovidenti
Sequence length	503 AA.
Sequence status	Complete.
Protein existence	Evidence at transcript level
General annotation (Co	mments)
Function	Catalyzes the formation of aromatic C18 estrogens from C19 androgens.
Catalytic activity	RH + reduced flavoprotein + O <sub>2</sub> = ROH + oxidized flavoprotein + H <sub>2</sub> O.
Cofactor	Heme group (By similarity).
Subcellular location	Membrane; Peripheral membrane protein.

Figura 4.2b. Desplegado de la aromatasa de oveja.



En cualesquiera de los casos que hubiese seleccionado, active la liga para obtener la secuencia en formato FASTA (Figura 4.3) Seleccione la secuencia de aminoácidos de la proteína y cópiela en un archivo de Word o wordpad. El archivo que obtenga de UniProt requiere que elimine los números de los aminoácidos y los espacios que separan los grupos.

🗯 Safari /	Archivo	Edició	n Visu	alización	Historial	Favoritos	Ventana	Ayuda		
00					cytochron	ne P450 194	1 [Ovis arie	s] – Proteir	1 – NCBI	
	0	+ Shtt	p://www.i	ncbi.nlm.nil	h.gov/protein	/172088165	?report=fasta	L		
ee 🖽 🎹	Apple	4shared	Yahoo!	Google Map	os Wikipedia	Noticias 🔻	Populares <b>*</b>	Flow		
S NCBI Re	esources	✓ How	To 🖂							
Protein		Prot	ein	•						
				Adva	anced					
Display Settin	Display Settings: ♥ FASTA Send to: ♥ cvtochrome P450 19A1 [Ovis aries]									
NCBI Referen	nce Sequ	uence: N	P 00111	- 6472.1	-					
GenPept Gr	aphics		_							
GenPept Graphics >gi 172088165 ref NP_001116472.1  cytochrome P450 19A1 [Ovis aries] MLLEVLNPRHINNTSMVSEVVPIASIAILLLTGFLLLVWNYEDTSSIPGPSYLGIGPLISHCRFLWMGI GSACNYYNKMYGEFMRVWVCGEETLIISKSSSMFHVMKHSHYISRFGSKLGLQFIGHHEKGIIFNNPAL WKAVRPFFTKALSGPGLVRAVTICADSITKHLDRLEEVCNDLGYVDVLTLMRRIMLDTSNILFIGIPLDE SAIVVKIQGYFDAWQALLLKPDIFFKISWLCRKYEKSVKDLKDAMEILIEEKRHRISTAEKLEDCIDFAT ELIFAEKRGELTKENVNQCILEMLIAAPDTMSVSVFFMLFLIAKHPQVEEAMMREIQTVVGERDIRIDDM QKLKVVENFINESMRYQPVVDLVMRKALEDDVIDGYPVKKGTNIILNLGRMHRLEFFFKPNEFTLENFAK NVPYRYF0PFGFGPRACAGKYIAMVMMKVILVTLLRRFHVQTLQGRCVEKMQKKNDLSLHPDETSDRLEM IFIPRNSDKCLEC										

Figura 4.3a. Ventana que muestra la secuencia en lenguaje FASTA de la base PROTEIN NCBI.

📕 Fi	nder Are	chivo	e Edición	Visualizad	tión Ir V	entana	Ayuda							
00	0				C	ytochrome	P450	19A1 -	Ovis aries (Shee	ep)				
			+ 📿 http://	www.uniprot	.org/uniprot/	Q9XS28					RSS (		gle	
<del>со</del> Д	🗋 🏭 Ар	ple	4shared Yal	hoo! Google	Maps Wikip	edia Noti	cias 🔻	Populare	s≖ Flow					
Cus Seq	lames · Attr tomize orde luences	ribute er	es · General :	annotation ·	Ontologies	Sequence	e annot	ation	Sequences · Re	oferences	· Cross-refs	Entry info	Documents	
	Sequenc	е			L	ength Ma	ss (Da)	Tools						
	Q9XS28 Last modi Checksun	(UniP fied No h: FA9	arc]. ovember 1, 19 FC7510691B3	99. Version 1. 3D8	FASTA	503	57,970	Blast	•	go				
	MLLEVL	1 <u>0</u> NPRH	2 <u>0</u> YNVTSMVSEV	3 <u>0</u> VPIASIAILL	4 <u>0</u> LTGFLLLVWN	5 YEDTSSIPG	0 P SYFL	6 <u>0</u> GIGPLI						
	SHCRFL	WMGI	GSACNYYNKM	YGEFMRVWVC	GEETLIISKS	SSMFHVMKH	S HYIS	RFGSKL						
	GLQFIG	13 <u>0</u> MHEK	14 <u>0</u> GIIFNNNPAL	15 <u>0</u> WKAVRPFFTK	16 <u>0</u> ALSGPGLVRM	17 VTICADSIT	0 K HLDR	18 <u>0</u> LEEVCN						
	DLGYVD	19 <u>0</u> VLTL	20 <u>0</u> MRRIMLDTSN	21 <u>0</u> ILFLGIPLDE	22 <u>0</u> SAIVVKIQGY	23 FDAWQALLL	0 K PDIF	24 <u>0</u> FKISWL						
	CRKYEK	25 <u>0</u> SVKD	26 <u>0</u> LKDAMEILIE	27 <u>0</u> EKRHRISTAE	28 <u>0</u> KLEDCIDFAT	29 ELIFAEKRG	0 E LTKE	30 <u>0</u> NVNQCI						
	LEMLIA	31 <u>0</u> APDT	32 <u>0</u> MSVSVFFMLF	33 <u>0</u> LIAKHPQVEE	34 <u>0</u> AMMREIQTVV	35 GERDIRIDD	0 M QKLK	36 <u>0</u> VVENFI						
	NESMRY	37 <u>0</u> QPVV	38 <u>0</u> DLVMRKALED	39 <u>0</u> DVIDGYPVKK	40 <u>0</u> GTNIILNLGR	41 MHRLEFFPK	0 P NEFT	42 <u>0</u> LENFAK						
	NVPYRY	43 <u>0</u> FQPF	44 <u>0</u> GFGPRACAGK	45 <u>0</u> YIAMVMMKVI	46 <u>0</u> LVTLLRRFHV	47 QTLQGRCVE	<u>0</u> к мокк	48 <u>0</u> NDLSLH						
	PDETSD	49 <u>0</u> RLEM	50 <u>0</u> IFIPRNSDKC	LEC										
	« Hide													

Figura 4.3b. UniProt.

Abra el servidor del sitio GeneDesign en la URL <u>http://54.235.254.95/gd/</u> que lo llevará al menú de operaciones del sitio. Seleccione la opción "REVERSE TRANSLATION" que abre la ventana correspondiente (Figura 4.4).

🟓 💛 😁	GeneDesign											
		+ 😚 htt	p://54.2	85.254.95/gd/					Ċ	Q- G00	gle	$\supset$
⇔ m III	Apple 4	4shared	Yahoo!	Google Maps	Wikipedia	Noticias 🔻	Populares <b>*</b>	Flow				
		design	n a gene						read	I the mar	nual	
							ne Sigi	N				
re tran now FA	everse Islation STA friend	lly		restriction si addition	te	sho	ort sequence addition	e	Building Block Desig (restriction site overla	n p)	sequence analysis	
c jug new algorithm	odon ggling a, new visu	alization	,	restriction si subtraction mproved algorit	te hm	sha SL impr	ort sequence ubtraction roved algorithm	e m	Building Block Desig (USER overlap)	n a	ambiguous sequence translation	
cod gra all new and	on bias aphing d FASTA fri	iendly	re more	striction enzy filter enzymes, more	yme criteria				Building Block Desig (constant length overla	n ap) s	random nucleotide sequence generation	
gener V all new and	rate RSCU alues d FASTA fri	iendly		vector filter						n	eed the old GeneDesign?	
email: notadoctor@jhmi.edu						rep download ti	oort an erron he source fror	r m github	3.0	Rev 12/0	7/09	
						ge	et the paper					

Figura 4.4a. Ventana de ingreso a Gene Design.



Figura 4.4b. Herramienta REVERSE TRANSLATION con la secuencia FASTA de la aromatasa ovina.



En esta nueva pestaña, se encuentra un encabezado con inscripciones simples y una ventana de diálogo en el cual debe pegar el archivo FASTA que hizo en Word o wordpad. En este sitio, puede seleccionar las características de los codones preferenciales que son utilizados en algunos de los modelos celulares más utilizados sin embargo, tiene la capacidad de utilizar los de la especie con la que se está trabajando. Para una referencia sobre los codones más utilizados en diferentes especies, puede consultar el artículo Sharp. PM, Cowe E, Higgins DG, Shields DC, Wolfe KH. Wright F. Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens;* a review of the considerable within-species diversity. Nucl. Acids Res. 16:8207-8211. 1988. Disponible en la URL <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC338553/pdf/nar00159-0016.pdf">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC338553/pdf/nar00159-0016.pdf</a>

Seleccione "H. sapiens" para el análisis. Pulse el botón "REVERSE TRANSLATE" lo que traerá una ventana en la cual se tiene la secuencia de nucleótidos que tendría el mRNA que codificaría para la proteína. Guarde esta secuencia en un archivo con formato FASTA.

Regrese a la página de resultados y pulse el botón "SEQUENCE ANALYSIS" lo que traerá una ventana en la cual está en una ventana simple la secuencia del mRNA que obtuvo, dos mapas verticales en el cual se indican los sitios de restricción que están presentes en el gene que codifica a este mRNA y uno más en el cual se encuentran las posibles pautas de lectura que pueden codificar a la proteína (ORF del inglés Open Reading Frame). En el espacio lateral se encuentra un listado de los sitios de restricción que no se encuentran en las ORF del gen que codificaría para la proteína (Fig 4.5). Guarde toda esta información en un documento que pueda visualizar posteriormente.

00					(	GeneDesigr	n: Reverse Tr	anslatic	on					
	0	+ 🕙 htt	tp://54.2	35.254.95/cgi	-bin/gd/gdF	RevTrans.cgi	i			Le	ctor 🖒	Q- Google	2	
	Apple	4shared	Yahoo!	Google Maps	Wikipedia	Noticias 🔻	Populares <b>v</b>	Flow						
GEN DES	ie Ign								Revers	e Trar	nslati	on		
				Your amino	acid sequenc S	es have bee ee the man	en successfull ual for more	y revers informa	se translated tion.	i to nucleo	tides.			
Your re ATGCT ATGCA TACTA AACCA CACTA CACAC CACTA CACAC AAGTA AAGCA AGTAA GAAAT	verse tra ACTAGAA ACTACGA CAACAAA GATTCGG AGCGGAC CTACTAA AGCGGAC ATGCTAA AGGAGAA AGGAGAA AGGAGAA CTACTAC GATATTC GATATTC	nslated nur GTACTAAAC AGACACAAC AGGACAAAC AGGACAAAC AAGCAAAC AAGCAATAAT AACCAGAC ACAGCACAA AAGAACATTAT AACAACACAT TTCCCAAAA AAGAACATTA TACCAAGA	cleotide s iCCAAGAC iCAGCAT iCAGATTA GGAATCA TAGGACTA TAGACG iCAGACA AAAACTAA AAAACTAA AAAACTAA AAAACTAA ITGACCG ICACTAA ICACAGCC	Sequences: CACTACAACGTA ACCAGGACAA ATGACAGTATOC ACAATTCATACC TGGTAACAATA CAAAATAACGT CAAAAATAACGT CAAAAATAACCTACA GACGACCATCAA GACGACCTACA GACGACCTACA GACAACATCCACTACA GACAACATCCACTACA GACAACATCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACATCCCACTACA GACAACTACACACTACA GACAACTACACACTACA GACAACTACACACTACA GACAACTACACACACACACACACACACACACACACACACA	ACAAGCATGG SCTACTTCCT. GTATGCGGA AATCGACGA ACTATTCCTG ACTATTCCTG AGCTTTGCA GGCTATTCCT AGCATCGCA AGGAACATGG GAATGC GAATGC	ΤΤΑΑΟርGΑΑΙ ΑGGAATAGG, JAAGAAACAI AAAAGCGAAT GGATAACAA GGAATACCA AAAATACCA AAAATACCA CAAACATAA TCATGCTATT GTAGTATAGAA CAAACATAA TGGGATTCG CGGATCG CGGATCG	СТАСТАССААТ АССАСТААТАА ААТАТТСААС АААТАТСААС ААТАТССААС СТАСАСТАА АААААСССТАСА ССТААТАССАА АСТСАТААС СААССААС	AGCAAG GCCACT CAAAAG ACAACC AGACTA GCGCAA AAAGAC CGAAAG CGGAAAC CGAAAG ATGCGC AAAACC	ICATAGCAAT. IGCAGATTCC ICACGAGCAT ICACAGCAT ICACAGCAAGTA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCA ICACAAGCAAGCA ICACAAGCAAGCA ICACAAGCAAGCA ICACAAGCAAGCA ICACAAGCAAGCA ICACAAGCAAGCAAGCA ICACAAGCAAGCAAGCA ICACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGC	АСТАСТАС ТАТБСДАТСА СБАААССА ТБССААСССА ССААТССААС ССААТССАА ССААССААС ССААССАА	TAACAGG GGAATAC TAATGAA GTAAGAC ICTAGAC ICTAGAC ICTACCA TAGTAGA GATCATC GACGAAA	ΑΤΤΟ ΕΓΑ ΕΓΑ GAAGEGEAT ACACAGCCA CATTECTTCA TAGGAGCAT TAGGAAGAAT ACACATE CCTACTAAT AAACCAATE CCTACTAAT AAACCAATE CAAGEGAAAC CAAGEGACA	СТАСТ СТАСАТ СТАСАТ ЗААААС GTACT GCAAG WAACGA GAATC GACTA GACTA	
_Com	position :	39% GC	Take th	nis sequence	You can	take this se	quence to an RE Site Subtra	other m	nodule now.	ing Sea	uence Ana	Ivsis		
				BB Design	(RE Overlap)	BB Design	n (USER Overlag	b) BB	Design (Sequ	ience Overla	ap)	114315		
					You	can save th	nis output as	a (FAST	TA file					

Figura 4.5a. Secuencia de nucleótidos.

\varTheta 🔿 🔿 GeneDesign: Sequence Analysis	
	Q- Google
Go III Apple 4shared Yahoo! Google Maps Wikipedia Noticias Populares Flow	
GNE Sequence Analys For a description of the formulas used, see the manual.	is
Please Note: •You do not have to go back to put in a new sequence. Simply type or paste the new sequence into the box and hi	t 'Analyze Again'.
Your sequence:	
ATGCTACTAGAAGTACTAAACCCCAACACACTACAACGTAACAGCATGGGAAGGCGAAGTAGTACCAATAGCAAGCA	
_Base Count : 1509 bp (677 A, 250 T, 305 C, 277 G) _Composition : 39% GC, 61% AT	
(Analyze Again)	
Absent Sites     ORFs       Aari Aaril Acii Acci IA dal Awa Awa Apal     Amili Acii Acci IA dal Awa	
Take this sequence to:         RE Site Addition         RE Site Subtraction         Codon Juggling	
(BB Design (RE Overlap)) (BB Design (USER Overlap)) (BB Design (Sequence Overlap))	

Figura 4.5b. Mapa de restricción obtenidos después de ingresar la secuencia de aminoácidos de la aromatasa de ovinos.

Con la secuencia de nucleótidos en FASTA que guardó, compárela con la que se encuentra depositada en el GenBank. Abra la página de inicio del NCBI y espere a que aparezca la página de inicio. En el campo de las herramientas disponibles, active la liga que lo lleve a la herramienta BLAST (Basic Local Alignment Search Tool). Esta activación lo llevará al menú de inicio de la herramienta. Identifique la porción correspondiente a "nucleotide blast" dentro de la sección "Basic BLAST" y active la liga.

La nueva ventana tiene una ventanilla donde puede pegar la secuencia en FASTA que obtuvo con GeneDesign. Asegúrese de que no se tienen espacios en la secuencia pues de otra forma el programa lo reconoce como una proteína independiente que debe alinear y provocar que se utilice un bucle del algoritmo que reporta un mensaje de error.





Figura 4.6a. Página de inicio de la herramienta BLAST.



Figura 4.6b. Consulta en donde se incorpora la secuencia FASTA obtenida en GeneDesign.



Figura 4.6c. Alineamiento de las secuencias y grado de homología.

Una vez terminada la depuración, indique que se trata de una secuencia para oveja (Sheep) en el cuadro de diálogo correspondiente. Marque el cuadro que indica que quiere los resultados en una nueva ventana y pulse el botón BLAST (Fig 4.5). En la nueva ventana aparecerá un mapa a color de los resultados del alineamiento y los genes o porciones de genes con las que se tiene alineamiento, junto con parámetros importantes para hacer la evaluación. Copie a un archivo los resultados para tener un registro permanente.

#### Actividades para reportar

Obtenga las secuencias de las proteínas de los receptores adrenérgicos humanos. Clasifíquelos y haga un análisis similar al realizado en las actividades descritas. Identifique las diferencias primordiales que se obtendría si se hiciera una estrategia de clonación en un bacteriófago del tipo  $\lambda$ gt11 en un sitio diferente al reconocido por *EcoRI*.

En un análisis redundante, obtenga la secuencia de nucleótidos del mRNA de la calpaína 10 y replique la actividad realizada. Discuta las diferencias que encontró en el proceso.



## Referencias

- Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., Schäffer, A.A. 2008. Database Indexing for Production MegaBLAST Searches, Bioinformatics 24:1757-1764. Disponible en la URL <a href="http://bioinformatics.oxfordjournals.org/content/24/16/1757.full.pdf">http://bioinformatics.oxfordjournals.org/content/24/16/1757.full.pdf</a>+http://bioinformatics.oxfordjournals.org/content/24/16/1757.full.pdf</a>+http://bioinformatics.oxfordjournals.org/content/24/16/1757.full.pdf</a>+http://bioinformatics.oxfordjournals.org/content/24/16/1757.full.pdf</a>+
- Richardson SM, Wheelan SJ, Yarrington RM, Boeke, JD. 2006. GeneDesign: Rapid, automated design of multikilobase synthetic genes. Genome Res. 16:550-556. Disponible en la URL <u>http://genome.cshlp.org/content/16/4/550.full.pdf+html</u>
- Zhang, Z., Schwartz, S., Wagner L., Miller W. 2000. A greedy algorithm for aligning DNA sequences", J Comput Biol 7:203-14. Disponible en la URL <u>http://online.liebertpub.com/doi/pdf/10.1089/10665270050081478</u>

# Aplicación de datos para el diseño de experimentos de genotipificación

## Capítulo 5

## Búsqueda de información general de enfermedades genéticas

Una de las metas de la investigación en Genética Humana es encontrar la razón fundamental de las enfermedades o la predisposición a desarrollarlas. Los estudios y esfuerzos continuos de Victor McKusick utilizado las metodologías básicas de la citogenética, inició el acopio de las relaciones entre las alteraciones cromosómicas y el desarrollo de enfermedades (o síndromes) y que se asociaban a una herencia mendeliana clásica. McKusick plasmó estas asociaciones en su libro "Mendelian Inheritance of Man" y a quien se refería en forma coloquial como "EL MAPA MORBIDO DEL HUMANO" que concretó primeramente en un texto de 12 volúmenes.

El inicio de la aplicación de las metodologías de manipulación y modificación de los ácidos nucleicos que actualmente confundimos con "Ingeniería Genética" o aún más con "Biología Molecular" permitió que se encontrara el sustrato fino de algunas enfermedades cuya atención médica era imperante. Así pues, la clonación del gene normal que se encuentra alterado en la Fibrosis Quística y su función como transportador (CFTR) permitió visualizar el potencial que tenía esta aplicación metodológica y que superaba pero no reemplazaba a su predecesora.

Con la implementación del proyecto HUGO, el desarrollo de las bases de datos, los adelantos en la informática, los algoritmos de correlaciones entre genes expresados (EST), porciones transcritas (ORF), mapeo de genes por cercanía (CONTIGS) y variaciones de un solo nucleótido (SNP) han permitido que el asentamiento físico de los genes cuya expresión anómala pueda ser asociado a una región y secuencia específica de genes.

La unidad encargada de mantener actualizada estas relaciones se encuentra en la Universidad John Hopkins, última institución a la que perteneció el Dr. McKusick. El proyecto forma parte del acervo en línea que maneja el NCBI. Al estar disponible en línea, la base de datos permite no sólo hacer una revisión de las regiones que han sido asociadas al desarrollo de una enfermedad, sino que también tiene una referencia cruzada con la base de datos en la cual se reportan los diferentes marcadores moleculares asociados, principalmente del tipo SNP y loci microsatelitales.

Además de las anotaciones propias, OMIM (Online Mendelian Inheritance of Man), se puede tener acceso a otras bases de datos, algunas de acceso abierto como las asociadas a las mutaciones del genoma mitocondrial (MITOMAP) mientras que otras requieren de un registro, aunque son libres de pago (Human Gene Mutation Database, HGMD).

#### Objetivo

Manejar una base de datos especializada en información médica. Utilizar la información relevante para diseñar una estrategia experimental que permita aportar conocimiento de punta al Área de la Salud.

#### Desarrollo

Adquiera el acceso a la página principal del NCBI (<u>http://www.ncbi.nlm.nih.gov</u>). Despliegue el menú de la ventana de selección de la base de datos. Seleccione OMIM. Opcionalmente, puede acceder ingresando directamente la dirección (<u>http://www.ncbi.nlm.nih.gov/omim</u>). En la ventana de diálogo, escriba "Non-insulin dependent Diabetes mellitus" y oprima el botón "SEARCH". Esto abrirá la página de OMIM y le brindará en una ventana los primeros 20 registros de la enfermedad. Debajo de cada una de los diferentes registros, se encuentra la localización de los genes asociados, incluyendo los de susceptibilidad (Fig 5.1). Modifique la página para poder ver los diferentes genes que recibe la página. Abra aquellos genes que no sea explícita su participación en la enfermedad.





Figura 5.1a. Ventana de acceso a la base de datos OMIM.



Figura 5.1b. Registros iniciales de Diabetes no dependiente de insulina.

Seleccione los registros asociados a la enfermedad. Elabore una lista de aquellos directamente asociados. En la pestaña de OMIM UniSTS confirme que su listado sea el asociado de acuerdo a la expresión de genes únicos. Haga un listado de la localización de estos genes. Abra la pestaña de OMIM dbSNP para ver la posición de los marcadores (Figura 5.2).

Safari Archivo Edición Visualización Historial Favoritos Ventana Ayuda	
O O Non-insulin dependent Diabetes Mellitus - OMIM - NCBI	
P     O	C Qr Google
🕀 🛄 🎹 Apple 4shared Yahoo! Google Maps Wikipedia Noticias* Populares* Flow	
S NCBI Resources V How To V	Sign in to NCBI
OMIM	
OMIM I Non-Insulin dependent Diabetes Meliitus	Search Hole
Jave search Linna Advarced	Link
Display Settings: Summary, 200 per page Send to:	Filter your results:
	All (33)
Results: 7	OMIM UniSTS (7)
#125853 - DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM	OMIM dbSNP (25)
<ol> <li>INSULIN RESISTANCE, SUSCEPTIBILITY TO, INCLUDED</li> </ol>	Manage Filters
Cytogenetic locations: 17q12, 17p13, 15q21-q23, 13q34, 13q12.1, 12q24.2, 11p12-p11.2, 11p15.1, 11p15.1	
2024.1.20013.1-013.2.217012.17013.15021-023.13034.13012.1.12024.2.11012-011.2.11015.1.	Find related data
11p15.1, 10q25.3, 8q24.11, 7q32, 7p15-p13, 7p21, 6q22-q23, 6p22.3, 5q34-q35.2, 4p16.1, 3q28, 3p25, 2q36	Database: Select
2q32, 2q24.1, 20q13.1-q13.2,	Find items
Gene summaries Genetic tests Medical literature	
<u>%603694 - DIABETES MELLITUS, NONINSULIN-DEPENDENT, 3</u>	Search details
<ul> <li>Cytogenetic locations. 20(12-(13.1 OMIM: 603694</li> </ul>	Non-insulin[All Fields] AND dependent[All Fields] AND
Gene summaries Genetic tests Medical literature	Diabetes[All Fields] AND
#601283 - DIABETES MELLITUS NONINSULIN-DEPENDENT 1: NIDDM1	Refficus(Aff Fields)
3. Cytogenetic locations: 2q37.3	Paarah
OMIM: 601283	See more
Gene summaries, Generic lesis, medical inerature	
#125850 - MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 1; MODY1	Recent activity
<ol> <li>Cytogenetic locations: 20q12-q13.1</li> </ol>	Turn Off Clear
Gene summaries Genetic tests Medical literature	Mellitus (33) OMIM
	See more
ANGIOTENSIN I-CONVERTING ENZYME, PLASMA LEVEL OF INCLUDED	
Cytogenetic locations: 17q23	
OMIM: 106180	
Gene summares, Generic resis, medical merature	
<u>*138570 - GLYCOGEN SYNTHASE 1: GYS1</u>	
<ul> <li>Cytogenetic locations: 19q13.3 ONUM: 19870</li> </ul>	
Gene summaries Genetic tests Medical literature	
Cytogenetic locations: 9o22.2-o22.3	
OMIM: 611570	
Gene summaries. Genetic tests. Medical literature	
Display Settings: V Summary, 200 per page Send to: V	

Figura 5.2a. Ventana de resultados de las bases de datos de secuencias estandarizadas únicas, UniSTS.



Figura 5.2b. Marcadores moleculares de variaciones de una sola base, SNP de Diabetes no dependiente de insulina.



Active la liga de los genes asociados para desplegar el mapa que muestra la localización del gene. Active la liga del número MIM para desplegar las características del gen (estructura, función, evidencias de asociación con la enfermedad, etc). Identifique el papel que tienen en la posible causa de la enfermedad (Figura 5.3). Utilizando la liga MAPVIEWER que aparece, indique la búsqueda de los genes asociados a DIABETES MELLITUS. Identifique por cromosoma la cantidad de genes que se han asociado.

<b>Safari</b> Archiv	vo Edición Visualización Historial Favori	tos Venta	na Ayuda			
	OMIM Entry - # 125853 - DIABE	TES MELLITU	S, NONINSU	LIN-DEPENDENT; NIDDM		_
	+ s http://omim.org/entry/125853		-1	<del>ن</del>		2
မော် ကျူးဆား Apple	4shared Yahoo! Google Maps Wikipedia Notici	as▼ Populare	ST Flow			
Google tradu	JCTOT Ver esta página en: español Ver esta página en:	r de Google	Desactivar	para: inglés	Opciones ▼ X	
Home   About   Sta	tistics -   Downloads/API -   Help -   External L	Links   Term	is of Use 👻	Contact Us	🛃 Seleccionar idioma   🔻	
			Court In		deted	l
Search OMIM		Search	) Sort by:	🖲 Relevance 🕒 Date	updated	l
Advanced Search: OM Search History: View,	IIM, Clinical Synopses, OMIM Gene Map <b>Toggle:   </b> changes highli Clear	ghted				l
					Table of Contents - #125853	L
#125853				ICD+	External Links:	L
. 120000					Clinical Resources	L
DIABET	ES MELLITUS, NONINSULIN-DI	EPENDE	ENT; NI	DDM	Animal Models	L
					► Cellular Pathways	L
Alternative +	itles: sumbols					[
Piner nulloe li					Centers for Mendelian Genomics	
DIABETES	MELLITUS, TYPE II; T2D					
NONINSUI	LIN-DEPENDENT DIABETES MELLITUS					
MATURITY	-ONSET DIABETES					L
						L
Other entiti	es represented in this entry:					L
IN ICT II IN						L
INSULI	N RESISTANCE, SUSCEPTIBILIT	Y 10, IN	ICLUDI	ΞD		L
						L
Phenotype	Gene Relationships					L
Location	Phenotype	Phenotype	Gene/Locus	Gene/Locus		L
		MIM number	CINDA	MIM number		L
2q24.1 2q31.3	{Diabetes, type 2, susceptibility to}	125853	GPD2 NEUROD1	138430		L
2036.3	{Diabetes mellitus, noninsulin-dependent}	125853	IRS1	147545		
3p25.2	{Diabetes, type 2}	125853	PPARG	601487		L
3q27.2	{Diabetes mellitus, noninsulin-dependent, susceptibility to}	125853	IGF2BP2	608289		
4p16.1	{Diabetes mellitus, noninsulin-dependent, association with}	125853	WFS1	606201		
5q34-q35.2	{Diabetes mellitus, noninsulin-dependent}	125853	NIDDM4	608036		ſ
6p22.3	{Diabetes mellitus, noninsulin-dependent, susceptibility to}	125853	CDKAL1	611259		Ľ
6p21.31	{Diabetes mellitus, noninsulin-dependent, susceptibility to}	125853	HMGA1	600701		ſ
6q23.2	{Diabetes mellitus, non-insulin-dependent, susceptibility to}	125853	ENPP1	1/3335		ſ
7p13 7c22.1	Diabetes mellitus, noninsulin-dependent, late onset	125853	GCK PAYA	1380/9		ſ
7q32.1 8a24.11	Diabetes mellitus, type 2 (Diabetes mellitus, noninsulin-dependent, susceptibility to)	125853	SLC30A8	611145		ſ
10g25.2-g25.3	{Diabetes mellitus, type 2, susceptibility to}	125853	TCF7L2	602228		
11p15.1	{Diabetes mellitus, type 2, susceptibility to}	125853	KCNJ11	600937		ſ
11p15.1	Diabetes mellitus, noninsulin-dependent	125853	ABCC8	600509		
11p11.2	{Diabetes mellitus, noninsulin-dependent}	125853	MAPK8IP1	604641		ſ
11q14.3	{Diabetes mellitus, type 2, susceptiblity to}	125853	MTNR1B	600804		ſ
12q24.31	{Diabetes mellitus, noninsulin-dependent, 2}	125853	HNF1A	142410		
13q12.2	{Diabates mellitus, type II, susceptibility to}	125853	IPF1 IPC2	600707		ſ
13q34	(Diabetes mellitus, noninsulin-dependent)	120803	LIPC	151670		
17p13.1	{Diabetes mellitus, noninsulin-dependent}	125853	SLC2A4	138190		ſ
17g12	Diabetes mellitus, noninsulin-dependent	125853	HNF1B	189907		
17q25.3	{Diabetes mellitus, noninsulin-dependent}	125853	GCGR	138033		Ľ
19p13.2	{Hypertension, insulin resistance-related, susceptibility to}	125853	RETN	605565		Ľ
19p13.2	{Diabetes mellitus, noninsulin-dependent, susceptibility to}	125853	RETN	605565		ſ
19q13.2	Diabetes mellitus, type II	125853	AKT2	164731		Ľ
20q12-q13.1	{Diabetes mellitus, noninsulin-dependent}	125853	NIDDM3	603694		Ľ
20q13.12	{Diabetes mellitus, noninsulin-dependent}	125853	HNF4A	600281		Ľ
20q13.13	{Insulin resistance, susceptibility to}	125853	PTPN1	176885		Ľ

Figura 5.3a. Ventana informativa sobre el registro 125853 de diabetes mellitus no dependiente de insulina. Genes asociados a la enfermedad.

Safari Archivo Edición Visualización Historial Favoritos Ventana Ayuda		
OMIM Entry – # 125853 – DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDD	м	
Image: Second	Ċ	Q- Google
🕀 🛄 🗰 Apple 4shared Yahoo! Google Maps Wikipedia Noticias 🔻 Populares 🔻 Flow		
Ciffical oynopolo		(
TEXT		
A number sign (#) is used with this entry because of evidence that more than one gene is		u di seconda
involved in the causation of noninsulin-dependent diabetes mellitus (NIDDM).		
See 601283 for description of a form of NIDDM linked to 2q, which may be caused by		. Table of Contents #125.952
mutation in the gene encoding calpain-10 (CAPN10; 605286). See 601407 for description of a		Frable of Contents - #125855
chromosome 12q locus, NIDDM2, found in a Finnish population. See 603694 for description		External Links:
of a locus on chromosome 20, NIDDM3.		Clinical Resources
		Animal Models
A mutation has been observed in hepatocyte nuclear factor-4-alpha (HNF4A; 600281.0004)		Cellular Pathways
in a French family with NIDDM of late onset. Mutations in the NEUROD1 gene (601724) on		
chromosome 2q32 were found to cause type II diabetes mellitus in 2 families. Mutation in		Centers for Mendelian Genomics
the GLUT4 glucose transporter was associated with NIDDM in 1 patient (138190.0001) and		
in the GLUT2 glucose transporter in another (138160.0001). Mutation in the MAPK8IP1		
gene, which encodes the islet-brain-1 protein, was found in a family with type II diabetes in		
individuals in 4 successive generations (604641.0001). Polymorphism in the KCNJ11 gene		
(600937.0014) confers susceptibility. In French white families, Vionnet et al. (2000) found		
evidence for a susceptibility locus for type II diabetes on 3q27-qter. They confirmed the		
diabetes susceptibility locus on 1q21-q24 reported by Elbein et al. (1999) in whites and by		
Hanson et al. (1998) in Pima Indians. A mutation in the GPD2 gene (138430.0001) on		
chromosome 2q24.1, encoding mitochondrial glycerophosphate dehydrogenase, was found		
in a patient with type II diabetes mellitus and in his glucose-intolerant half sister.		
Mutations in the PAX4 gene (167413) have been identified in patients with type II diabetes.		
Triggs-Raine et al. (2002) stated that in the Oji-Cree, a gly319-to-ser change in HNF1-alpha		
(142410.0008) behaves as a susceptibility allele for type II diabetes. Mutation in the HNF1B		
gene (189907.0007) was found in 2 Japanese patients with typical late-onset type II diabetes.		
Mutations in the IRS1 gene (147545) have been found in patients with type II diabetes.		
Reynisdottir et al. (2003) mapped a susceptibility locus for type II diabetes to chromosome		
5q34-q35.2 (NIDDM4; 608036). A missense mutation in the AKT2 gene (164731.0001) caused		
autosomal dominant type II diabetes in 1 family. A (single-nucleotide polymorphism) SNP		
in the 3-prime untranslated region of the resistin gene (605565.0001) was associated with		
susceptibility to diabetes and to insulin resistance-related hypertension in Chinese		
subjects. Susceptibility to insulin resistance has been associated with polymorphism in the		
TCF1 (142410.0011), PPP1R3A (600917.0001), PTPN1 (176885.0001), ENPP1 (173335.0006),		
IRS1 (147545.0002), and EPHX2 (132811.0001) genes. The K121Q polymorphism of ENPP1		
(173335.0006) is associated with susceptibility to type II diabetes; a haplotype defined by 3		
SNPs of this gene, including K121Q, is associated with obesity, glucose intolerance, and		
type II diabetes. A SNP in the promoter region of the hepatic lipase gene (151670.0004)		
predicts conversion from impaired glucose tolerance to type II diabetes. Variants of		
transcription factor 7-like-2 (TCF7L2; 602228.0001), located on 10q, have also been found to		
confer risk of type II diabetes. A common sequence variant, $rs10811661$ , on chromosome		
9p21 near the CDKN2A (600160) and CDKN2B (600431) genes has been associated with risk		
of type II diabetes. Variation in the PPARG gene (601487) has been associated with risk of		
type 2 diabetes. A promoter polymorphism in the IL6 gene (147620) is associated with		
susceptibility to NIDDM. Variation in the KCNJ15 gene (602106) has been associated with		
T2DM in lean Asians. Variation in the HMGA1 gene (600701.0001) is associated with an		
increased risk of type II diabetes. Mutation in the MTNR1B gene (600804) is associated with		

Figura 5.3b. Descripción de síntomas clínicos, localización cromosómica y asociación génetica para Diabetes no dependiente de insulina.

#### Actividades para reportar

Utilizando las herramientas desarrolladas en esta sesión, haga el análisis de las patologías asociadas a los receptores adrenérgicos Alfa.

De la misma manera, analice los genes asociados al síndrome vasovagal.



## Referencias

- Goh KI., Cusick, ME., Valle, D., Childs, B., Vidal, M., Barabási, A-L.2007. The human disease network. Proc. Nat. Acad. Sci. 104:8685-8690. Disponible en la URL http://www.pnas.org/content/104/21/8685.full.pdf+html
- Schriml, LM., Arze, C., Nadendla, S., Chang, Y-W W., Mazaitis, M., Felix, V., Feng, G., Kibbe WA.2012. Disease Ontology: a back bone for disease semantic integration. Nucleic Acid Res. 40:D940-D946. Disponible en la URL <a href="http://nar.oxfordjournals.org/content/40/D1/D940.full.pdf+html">http://nar.oxfordjournals.org/content/40/D1/D940.full.pdf+html</a>

## Capítulo 6 Diseño de oligonucleótidos para amplificación

Una de las metodologías que ha permitido el avance de la Biología Molecular y la Genética Molecular es la reacción en cadena de la polimerasa, mejor conocida por sus siglas en inglés, PCR. Esta reacción permite la obtención de cantidades de ADN en cantidades adecuadas para ser utilizadas o analizadas por otras técnicas a partir de pequeñas cantidades de material inicial. Si simplificamos el procedimiento, se trata de una reacción de replicación del material genético bajo condiciones in vitro. Como en el caso del proceso in vivo, se requiere de un sustrato que puede ser ADN o ARN, reactivos necesarios que servirán de unidades estructurales a partir de las cuales se construirán las nuevas cadenas de ADN, componentes que inician las reacción y aquellos que la mantienen, incluyendo las sales que permiten tener una fuerza iónica adecuada para que el funcionamiento de la enzima pueda realizarse bajo las mejores condiciones. Podemos encuadrar en cada uno de estos rubros a los desoxinucleótidos trifosfatados (dNTP), iones Mg<sup>2+</sup>, glicerol, amortiguadores de Tris-hidroximetil-amino-metano (Tris), los cebadores de la polimerasa y la propia enzima, comúnmente se utiliza la ADN polimerasa termoestable proveniente de la arqueobacteria *Thermus aqcuaticus*, mejor conocida como Taq polimerasa.

Bajo condiciones naturales, las polimerasas requieren de un oligonucleótido producido por la enzima DNA primasa por lo que se denomina "primer" que puede ser traducido como cebador o iniciador. Este es un oligonucleótido de sólo 30 a 50 bases que proporcionan el sitio de reconocimiento de la ADN polimerasa: el extremo 3'hidroxilo de un nucleótido pareado a la cadena que se copiará. Al unirse a este extremo, la ADN polimerasa continúa con la elongación de la cadena incorporando el desoxinucleótido trifosfatado que complemente la doble cadena, de acuerdo a las reglas de apareamiento de Watson-Crick: si el molde tiene Adenina, en la cadena complementaria se incorpora Timidina; si el molde presenta Citosina, la complementaria incorporará una Guanina.

En la caso de la reacción in vitro, se copia únicamente una región específica por lo que se utiliza un par de iniciadores cuya secuencia es complementaria a las zonas que se van a amplificar. La especificidad, eficiencia y posibles variaciones dependen entonces de un diseño adecuado de estos oligonucleótidos. A manera de características generales, los aspectos que deben cuidarse en el diseño de estos oligonucleótidos son su secuencia, temperatura de alineamiento, que no sean autocomplementarios, que no formen estructuras tridimensionales complejas de más de 3 bases, que no se unan a secuencias similares internas (sitios secundarios de hibridización), y que la unión en el extremo 3'del propio oligonucleótido esté estabilizado por una base capaz de formar 3 puentes de hidrógeno con su base complementaria.

#### Objetivo

Diseñar un par de cebadores capaces de ser utilizados en una reacción de amplificación, determinar las características que tendría la reacción de amplificación y la posible longitud que tendría el fragmento amplificado (amplicón) utilizando un recurso en línea.

#### Desarrollo

#### Obtención de iniciadores mediante el uso de un portal de herramientas bioinformáticas

El inicio del análisis implica tener una secuencia blanco a la cual estará dirigida la metodología. Obtenga la secuencia en FASTA del acceso NM\_023085 correspondiente a la calpaina 10 humana depositada en el GenBank (http://www.ncbi.nlm.nih.gov/gene). Obtenga la secuencia en FASTA activando la liga que se encuentra en la parte superior derecha del mapa cromosómico (figura 6.1). Seleccione todo el contenido, copie y guarde ya sea en la memoria temporal de su computadora o en un archivo de un procesador de textos.



Safari Archivo Edicion Vis	sualizacion Historial	Favoritos Ventana	Ayuda	
00		Home - Gene - N	ICBI	
	w.ncbi.nlm.nih.gov/gene/			C Google
↔ 💭 🇰 Apple 4shared Yahoo!	Google Maps Wikipedia	n Noticias ▼ Populares ▼	Flow	
🗧 NCBI Resources 🖸 How To 🗹				Sign in to NCBI
Gene				Search
	Limits Advancer	1		Help
and all the state of the state	Gen	e		
	Gene ir	ntegrates information fro	om a wide range of	species. A record may include nomenclature,
	Referen	hce Sequences (RefSec vne- and locus-specific	s), maps, pathway	rs, variations, phenotypes, and links to genome-,
				<u> </u>
Using Gene	Gene T	ools		Other Resources
Gene Quick Start	Submit G	eneRIFs		HomoloGene
FAQ	Submit Co	orrection		OMIM
Download/FTP	Statistics			RefSeq
RefSeq Mailing List	BLAST			RefSeqGene
Gene News 🔊	Genome	Norkbench		UniGene
Factsheet	Splign			Protein Clusters
Representative queries				
Find genes by	Search text			
free text	human muscular dystrop	<u>hy</u>		
chromosome and symbol	(II[chr] OR 2[chr]) AND a	dh*[sym]		
partial name and multiple species	alive[prop] AND transport	ter[title] AND ("Drosophila	melanogaster"[orgn]	OR "Mus musculus"[orgn])
associated sequence accession	M11313[accn]			
gene name (symbol)	BRCA1[sym]			
publication (PubMed ID)	11331580[PMID]			
	"coll adhesion"[GO]			
Gene Ontology (GO) terms or identifiers	10030[GO]			
	10000[00]			
genes with short variants of medical	gene_snp_clin[filter]			
chromosome and species		<u>JNI</u>		
Enzyme Commission (EC) numbers	<u>1.9.3.1[EC]</u>			

Figura 6.1a. Ventana de acceso al GenBank.



Figura 6.1b. Región del cromosoma 2 humano donde se localiza el gen de la calpaina 10.



Satari Archivo Edición Visualización Historial Favoritos Ventana Ayuda	and the second se
🔴 🔿 🧑	leotide –
Image: A state of the s	
Annie Ashared Vahool Google Mans Wikinedia Noticiasy Popularesy Flow	
Nucleotide	
Nucleotide	
Limits Advanced	
<u>Display Settings:</u> () FASTA	<u>Send:</u> [∨]
Homo sapiens calpain 10 (CAPN10), transcript variant 1, mRNA	
NCBI Reference Sequence: NM 023083.3	
GenBank Graphics	
>di 305855091 ref NM 023083.3  Homo sapiens calpain 10 (CAPN10), transcript	
variant 1, mRNA	
CGCCTGGTTACCAATGGGAGACTAGCGGGCCGGCGTACTGGCCTGGTCCAGCACCTGCGGGGCCCTCGGG	
CTTGGAGGGCTGGGCCGGGGGGAACGGGCGGGGGGGGGG	
TCCGGGGCTGCGACCCCGAGGCAACCGGCTGCAGATGGGAGCCCGGGGCCGGAGCCGGGGCGGGGCGGGC	
CACCCCGGCTGTTTCCAGATGACCCACGGGAGGGCAGGCA	
CCTGTGTGCCTGCGCCGCGCCTGCAGAAGAGCAGCACCTCCTGGACCAGGTCATTCCTCCGGGACAGCCG	
AGCTGGGCCCACCAGGAGTACCGGGGCTCCTTCACCTGTCGCATTTGGCAGTTGGACGCTGGGTGGAGG	
TGACCACAGATGACCGCCTGCCGTGCCTTGCAGGGAGACTCTGTTTCTCCCCGCTGCCAGAGGAGGATGT	
GTTCTGGCTCCCCTTACTGGAAAAGGTCTACGCCAAGGTCCTACGAGCACCTGTGGGCCGGG	
GACCAGTGTCTGATCAGCTGCTGCGCGCGCCCAGCCCCCAGCGGGGGGGG	
GCCTTCATTGTCTCGGACCTGCGGGAGCTCCAGGGCGGGC	
AGAACCCCTGGGGCCGGCGGTGCTGGCAGGGGGCTCTGGAGGGGGGGG	
TGCAGCGGTAGCATCTGAGCTCCTGTCCCAGCTCCAGGAAGGGGAGTTCTGGGTGGAGGAGGAGGAGTTC	
CTCAGGGAGTTTGACGAGCTCACCGGTGGCTACCCGGTCACGGCGGCCGCCACCTGCAGAGCCTCTACA	
CAGABAGGCTGCTGCGCGTGCGGGGGGGGGGGGGGGGGGG	
ATTECCGTCCTGCAGAGATCCAGGCTGCACGCGCGCGCCCGGGCCCCGGCCCTGGTGGGTG	
ACAGTCATACTTCGTGGAGCCCAGCGAGCATCCCGGGCAAGCACTACCAGGCTGTGGGTCTGCACCTCTG	
GAAGGTAGAGAAGCGGCGGGTCAATCTGCCTAGGGTCCTGTCCATGCCCCCGTGGCTGGC	
CATGCATACGACCGGGAGGTCCACCTGCGTTGTGGGCTCTCCACCGGGCTACTACCTGGCTGTCCCCAGCA	
CCTTCCTGAAGGACGCGCCAGGGGAGTTCCTGCTCCGAGTCTTCTCTCACCGGGCGAGTCTCCCTTAGCGC	
CALCAGEGETETTEGACAETCECCAECCECEGEGEGEGECETECEGEGEGETETEGEGEGEGECEGEC	
CCTGCTTCCCCTTCTCGGTCCCCGAGGGCCCTGGCCCCCGCTGCGTCCGCATCACTCTGCATCAGCACTG	
CCGGCCCAGTGACACCGAGTTCCACCCATCGGCTTCCATATCTTCCAGGGTCCCAGAGGGTGGAAGGAGC	
CAGGACGCACCCCACTGCTGCTGCAGGAGCCGCTGCTGAGCTGCGTGCCACATCGCTACGCCCAGGAGG	
TGAGCCGGCTCTGCCTCCTGCCGGCGCCCCTACAAGGTTGTGCCCCTCCACCTACCT	
CAGTTCCTACAGTGACCATCGCAACCAGGAGATGACAGGCCATCCAT	
GGTGACTGGAGCCCGAGGGCCTGACAGGTTCCCAGCAGCTGGGCCGGCC	
GTCCTGAGTCTTGGCCTGCCTCCCAGCCCTGCCAGGAGGCTGCGGCCTAGGGGTCCACGGGAAGCCTCCG	
TCAGGAGAGACGCAGCCCTGGGGGCCAGCTGGTGCTGCAAGGAAGG	
GCGCCACTGAGACGGCAGAGACCCCAGGATCCCAGAGCTTCCCAGGATCCCTCCC	
TUUATATUGAGGUUTUAUAGAUUUUAGAGGUTAGGUUAGATUTTTTTTTATAGUTATTTATT	
AA	

Figura 6.1 c Secuencia de nucleótidos del gen capn10.

Ingrese al portal de herramientas de ciencia y tecnología de Israel en la URL <u>http://www.scienc.co.il/Biomedical/Primer-Tools.asp</u> lo que le abrirá la página que se muestra en la figura 6.1 Dentro del menú de sitios, seleccione la opción PRIMER3 que lo llevará a la URL <u>http://frodo.wi.mit.edu/primer3</u> y le desplegará la ventana que se muestra en la figura 6.2 En la ventana de diálogo, pegue la secuencia FASTA que obtuvo del GenBank, identifique su secuencia como "Calpaina10" y la región que queremos amplificar que en este caso corresponde a la región 4 bases alrededor del nucleótido 1469. Haga esta indicación ingresando "1469,4" en el cuadro de dialogo marcado como "TARGET". Delimite a un amplicon de 300 a 400 nucleótidos de longitud seleccionando este valor en el cuadro marcado "Product size ranges"

00		Primer design tools	
	http://www.science.co.il,	Biomedical/Primer-Tools.asp	C Q- Google
↔ □ III Apple 4shar	ed Yahoo! Google Map	Wikipedia Noticias Populares Flow	
*	Israel Scienc	e and Technology Homepage	
Home Search About Contact	Biomedical :	Online Tools for Primer Design and Analysis	
Biomedical directory	Online Tool	Description	
Biochemical	AutoPrime	Primer design for real-time PCR measurement of eukaryotic gene exp	pression.
nomenclature	CODEHOP	COnsensus-DEgenerate Hybrid Oligonucleotide Primers designed from	n protein multiple sequence alignments.
Biomedical Databases	ExonPrimer	Design intronic primers for PCR amplification of exons. Input needed: sequence.	: a cDNA and the corresponding genomic
Carbohydrate Databases	IDT AntiSense Design	Antisense oligo design and selection tool.	
Gene Expression	IDT Oligo Analyzer	Online calculation of oligonucleotide parameters such as melting temp performs Blast.	perature. Shows self-dimers, hairpin, and
Databases	IDT PrimerQuest	Primer and probe design and selection.	
Genome Databases	NetPrimer	Java applet for primer design.	
Genome, Plant	PathoGene	A CDS Finding and Primer Design Tool for Microorganisms.	
Genome, Prokaryotic	Primer3	Utility for locating oligonucleotide primers for PCR amplification of DN	A sequences.
Glossaries	PrimerX	Automated design of mutagenic primers for site-directed mutagenesis	s.
Lipid Databases	Primo Pro	PCR Primer Design.	
Metabolic Pathways	QuantPrime	Automatic high-throughput primer pair design and specificity testing for	for realtime qPCR on any organism.
Mutation Databases	RF-Cloning	Online tool for the design of primers for restriction-free cloning project	cts.
Nomenclature - Gene	RNAi Design	Design duplexed RNA oligos for RNA interference.	
Organelle Databases	SiteFind	Design of oligonucleotide primers for site-directed-mutageneis that in marker of successful mutation.	clude a novel restriction site for use as a
Plasmid Databases	UCSC In-Silico PCR	In-Silico PCR searches a genome sequence database with a given pair	ir of PCR primers.
Protein Databases	Web Primer	Primer design and sets for amplifying yeast ORFs.	
Protein Nomenclature			
Protein Interaction Databases	Primer Desig	n Considerations	
RNA Databases	Desired character	stics of automated DNA sequencing primer design.	
Sequence Databases	<ul> <li>Based on</li> </ul>	accurate sequence	
Species Databases	<ul> <li>Melting ter</li> </ul>	nperature (Tm): 52°C to 65°C	
Structure Databases	<ul> <li>Absence of</li> </ul>	f self-hybridization	
Mars Deteksors	<ul> <li>Absence of</li> </ul>	f significant hairpin formation (>3 bp)	
Virus Databases	<ul> <li>Lack of se</li> </ul>	condary priming sites	
Primer Tools	<ul> <li>Low specir</li> </ul>	ic binding at the 3' end (ie. lower GC content to avoid mispriming)	
Protein Tools	<ul> <li>Working v</li> </ul>	ith PCR	
Frequence Teols	<ul> <li>PrimerBar</li> </ul>	k: PCR primers for gene expression detection or quantification	
Sequence Tools	<ul> <li>How Real-</li> </ul>	Time PCR Works	
	Notes on I	rimer Design in PCR	
	<ul> <li>Primers for</li> </ul>	r Automated Sequencing	
	Copy	This wash also included a second second and the second sec	s Reserveu. Terms or use.
		This web site is validated for HTMLS.	

Figura 6.2a. Ventana de aceso al portal de apoyo bioinformático en el área de la Biomedicina de Israel.



Figura 6.2b. Consulta del algoritmo Primer3 del MIT.



Las características que queremos de los cebadores son tamaño comprendido entre 20 y 30 nucleótidos con un óptimo de 25, la temperatura de linearización entre 54 y 64 °C con una óptima de 60, contenido de CG entre 40 y 60% con óptimo de 58. Coloque estos parámetros en los cuadros correspondientes en la sección "GENERAL PRIMER PICKING CONDITIONS". De los 3 métodos para calcular las características de corrección en el cálculo de la temperatura de linearización (Tm), utilice el modelo de SantaLucia (1998) de corrección de interacciones secuenciales múltiples dependiente de las condiciones de la reacción. Al final de estas indicaciones, active el botón de acción localizado en el final de la sección como "PICK PRIMERS". El programa le indicará que se encuentra trabajando en el diseño de los cebadores. Cuando encuentre algunos, se desplegará la página de resultados (figura 6.3). La tabla de resultados le indica las características de los cebadores con mejor eficiencia. Además de la secuencia, le indica la temperatura de hibridización y en la parte inferior de la tabla, el posible tamaño del amplicon seguido de la secuencia, el programa regresa pares de iniciadores adicionales. Esto permite hacer una selección dependiendo del interés del investigador.

Primer3 Output (primer3_results.cgi release 4.	.0.0)
Image: State of the state o	
оо дд ни Арріс Чанаса тапоо. Сооденара инкреша поссазо горианса той	
Primer3 Output	
No mispriming library specified Uaing 1-based sequence positions OLIGO <u>start len tm got any th 3' th hairpin seq</u> LEFT PRIMER 1399 20 55.93 45.00 4.15 0.00 0.00 TGACAGTCATAC' RIGHT PRIMER 1770 20 58.44 55.00 0.00 0.00 0.00 ACTCTCCAAGAA SEQUENCE SIZE: 2662 INCLUDED REGION SIZE: 2662	TTCGTGGA CCCCGTAG
PRODUCT SIZE: 372, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00 TARGETS ( <u>start lon</u>	
1 CGCCTGGTTACCAATGGGAGACTAGCGGGCCGGCGTACTGGCCTGGTCCAGCACCTGCGG	
61 GGCCTTCGGGCTTGGAGGGCTGGGCCGGGGGAACGGGCGGG	
121 GECGGCTGACTCECCTTCTCCCGGGGCTGCGACCCCGAGGCAACCGGCTGCAGATGGGA	
181 GCCCGCGGAGCCGAGGATGCGGGCGGGGCGGGGCGCGGGGGGGG	
241 GGACGCCGCCTTCCCCGCCGGGACTCCTCGCCTCTTCTGCGACTTGTCTACGCCGCTGGC	
301 CCASTTCCGCGAGGACATCACGTGGAGGCGGCCCCAGGAGATTTGTGCCACACCCCGGCT	
361 GTTTCCAGATGACCCACGGGAAGGGCAGGTGAAGCAGGGGCTGCTGGGGGATTGCTGGTT	
421 CCTGTGTGCCCGCCGCCGCGCGCAGAAGAGCAGGCACCTCCTGGACCAGGTCATTCCTCC	
481 GGGACAGCCGAGCTGGGCCGACCAGGAGTACCGGGGGCTCCTTCACCTGTCGCATTTGGCA	
541 GTTTGGACGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGG	
601 CTGTTTCTCCCGCTGCCAGAGGGAGGATGTGTTCTGGCTCCCCTTACTGGAAAAGGTCTA	
661 CGCCAAGGTCCATGGGTCCTACGAGCACCTGTGGGCCGGGCAGGTGGCGGATGCCCTGGT	
721 GGACCTGACCGGCGGCCTGGCAGAAAGATGGAACCTGAAGGGCGTAGCAGGAAGCGGAGG	
781 CCAGCAGGACAGGCCAGGCCGCTGGGAGCACAGGACTTGTCGGCAGCTGCTCCACCTGAA	
841 GGACCAGTGTTGATCAGCTGCTGCGGGGCTCAGCCCCAGAGCAGGTGCCCGGGAGCTGGG	
901 GEAGTTCCATGCCTTCATTGTCTCGGACCTGCGGGAGCTCCAGGGTCAGGCGGGCCAGTG	
961 CATCCTGCTGCGGATCCAGAACCCCTGGGGCCGGCGGTGCTGGCAGGGGCTCTGGAG	
1021 AGAGGGGGGTGAAGGGTGGAGCCAGGTAGATGCAGCGGTAGCATCTGAGCTCCTGTCCCA	
1081 GCTCCAGGAAGGGGAGTTCTGGGTGGAGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACGAGCT	
1141 CACCETTEGCTACCEGCTCACGEAGECCEGCCACCTECAGAGECCTCTACACAGAGAGECT	
1201 GCTCTGCCATACGCGGGCGCTGCCTGGGGCCTGGGTCAAGGGCCAGTCAGCAGGAGGCTG	
1261 CCGGAACAACAGCGGCTTTCCCAGCAACCCCAAATTCTGGCTGCGGGTCTCAGAACCGAG	

Figura 6.3a. Resultados iniciales del diseño Prime3 con la secuencia y características de los oligonucleótidos iniciadores así como la posible longitud del amplicón.





## Diseño de iniciadores utilizando la herramienta "Primer-BLAST" del Centro Nacional de Información Biotecnológica, NCBI

Acceda a la página principal del NCBI (http://ncbi.nlm.nih.gov) y en el campo de las herramientas populares, active la liga "BLAST" que lo llevará a la HERRAMIENTA LOCAL BASICA DE ALINEAMIENTO DE SECUENCIAS (figura 6.4a). En la sección de BLAST Especiales, active la liga de "Primer-BLAST" que despliega la ventana de la herramienta que utilizaremos (figura 6.4b). En el cuadro de dialogo, ingrese el número "3046857" correspondiente al mRNA de la aromatasa de ratón. Modifique el organismo sobre el cual se hará el diseño. De manera habitual, el programa hace el diseño en humanos, en este caso, cambie a "Mus musculus" y conforme va ingresando esta información, el programa le da algunas opciones de taxa, dependiendo de la subespecie (no importante para este ejercicio).

Active la liga de diseño avanzado (figura 6.4c) para indicar además que se requiere un iniciador interno que nos permitiría diseñar un experimento de PCr cuantitativa, también denominada de "tiempo real" y que en la literatura podemos encontrar como qPCR o RTPCR. Introduzca en las ventanas de dialogo de los iniciadores internos el tamaño (20 nucleótidos de longitud como optimo), la temperatura de alineamiento (60°C) y contenido de CG (50%). Active la liga "GET PRIMERS" localizada en el margen inferior izquierdo.



Ú.	Safari	Archivo	- Edició	n Visi	ualización	Historial	Favoritos	Venta	ina Avuda	
0	) ()		Luicio		anzación	BLAS	T: Basic Loo	cal Aligni	ment Search Tool	
			+ Shtt	n://blast	nchi nlm nih			J.		¢
		Annle	Ashared	Vahool	Google Mans	Wikinedia	Noticias	Popular	es T Flow	
-		a	43marcu	Tanoo.	doogie maps	mapeula	Noticitas	ropular		
3	BLAST					Basic	Local Ali	gnment	Search Tool	
	Home	Recen	nt Results	Save	d Strategies	Help				
► N	CBI/ BLAST	Home								
	BLAST fi	inds reg	ions of sir	nilarity l	etween biolo	gical seque	nces. more			
			New	DELT	A-BLAST,	a more ser	nsitive pro	otein-pro	otein search	
	BLAST	Assen	nbled Re	efSeq 0	Senomes					
	Choose a	species	genome to	search.	or list all gen	omic BLAS	databases	5.		
			0	,				-		
	Huma	an			Oryza	sativa			Gallus gallus	
	D Mous	e			□ <u>Bos ta</u>	aurus			Pan troglodytes	
	n Arabi	donsis	thaliana		Danio	nhila melan	ogaster		Anis mellifera	
							- Juoron	_	<u>- pro monto a</u>	
	Basic B	LAST								
	0	DIAOT								
	Choose a	BLAST	program to	run.						
			Soarch	a nucleo	tida databasa		leetide que	<b>D</b> /		
	nucleoti	de blast	Alg	orithms:	blastn, megab	last, disconti	guous mega	ablast		
			Soarch	protoin (	latabasa using	a a protoin a	100			
	prote	ein blast	Alg	orithms:	blastp, psi-bla	st, phi-blast,	delta-blast			
		blach	Soarah	protoin (	latabasa using	a a translato	t nucleotid			
		DIASIA	Search	protein	เลเลมสรช บริการุ	a liansiale	i nucleoliu	e query		
		tblastn	Search	translate	d nucleotide	database us	ng a <b>protei</b>	n query		
		tblastx	Search	translate	d nucleotide	database us	no a transla	ated nucl	leotide querv	
							0			
	Speciali	ized D	ACT							
	Special		LASI							
	Choose a	type of s	specialized	search (	or database n	ame in paren	theses.)			
		Makes	pecific prir	nore with	Primer-BLAS	ат				
	-	Search	trace arch	nives						
		Find co	onserved d	lomains	in your seque	nce (cds)				
		Find se	equences w	ith simila	ar conserved	domain arch	itecture (co	dart)		
		Search	sequence	s that ha	ve gene expre	ession profil	<u>es</u> (GEO)			
		Search	Immunog	lobulins Janke	and I cell re	ceptor sequ	ences (IgBI	LAST)		
		Screen	sequence	for vecto	or contaminat	ion (vecscre	en)			
		Align t	wo (or mor	e) seque	nces using BL	AST (bl2seq	)			
		Search	protein or	nucleo	ide targets in	PubChem Bi	oAssay			
		Search	SRA by e	xperime	nt					
		Needlo	man-Wure	Protein I	A Sequence	Iignment T	ool			
		Search	RefSeqGe	ene ene	a ocquence /	angiment I				
						E	LAST is a reg	istered trad	demark of the National Library	of Medicine.
Сору	right   Disclai	imer   Priv	acy   Access	ibility   Cor	tact   Send feedl	back				

Figura 6.4a. Ventana de acceso a la herramienta de diseño de iniciadores.

<b>Safari</b> Archivo Edi	ción Visualización Historial Favoritos Ventana Ayuda
	Primer designing tool
	http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome C Qr Google 3
C Drimor-BLAST	a tanoo: Google Maps Wikipedia Noticias+ Populares+ Plow
	A tool for many specific primers
NCBI/ Primer-BLAST: Findin	primers specific to your PCR template (using Primer3 and BLAST). More Tips for finding specific primers
PCR Template	Reset page Save search parameters Retrieve recent results
Enter accession, gi	, or FASTA sequence (A refseq record is preferred) 🛞 <u>Clear</u> Range
3046857	From To
	Reverse primer
Or, upload FASTA	ile Seleccionar archivo ningúnionado
Primer Parame	ters
Use my own forwar	d primer
(5'->3' on plus strar Use my own revers	d) e primer Q Clear
(5'->3' on minus str	and) Min Max
PCR product size	70 1000
# of primers to retu	rn 5
Primer melting tem	Min Opt Max Max Tm difference
(Tm)	
Exon/intron sel	ection A refseq mRNA sequence as PCR template input is required for options in the section (a)
Exon junction spar	No preference
Exon junction mate	h Exon at 5' side Exon at 3' side
Intron inclusion	
Intron length range	Min Max
	1000 0
	Note: Parameter values that differ from the default are highlighted in yellow
Primer Pair Sp	acificity Checking Parameters
Specificity check	M Enable search for primer pairs specific to the intended PCR template 🈡
Database	Refseq mRNA
Organism	Mus musculus (taxid:10090)
	Add more organisms
Exclusion (oplonal)	Exclude predicted Refseq transcripts (accession with XM, XR prefix) Exclude uncultured/environmental sample sequences 😣
Entrez query (option	al)
Primer specificity s	tringency Primer must have at least 2 total mismatches to unintended targets, including
	at least 2 🗘 mismatches within the last 5 🗘 bps at the 3' end. 😡
	Ignore targets that have 6 ; or more mismatches to the primer.
Misprimed product deviation	size 4000 @
Splice variant hand	ling Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input) @
Get Primers	Show results in a new window 🕅 Use new graphic view 📦
Advanced paramet	ers Note: Parameter values that differ from the default are highlighted in vellow
Copyright   Disclaimer   Privacy   Ac	cessibility   Contact   Send feedback on new interface NCRI   NI M   NIH   DHHS
Copyright   Disclaimer   Privacy   Ac	2essibility   Contact   Send feedback on new interface NCBI   NLM   NIH   DHHS

Figura 6.4b. Características de la ventana de dialogo inicial.



Safari Archivo Edición	Visualización Historial Favoritos Ventana Ayuda	
00	Primer designing tool	
< ▶ 🔋 🛈 + Shttp:	//www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome C Q Google	0
🖙 🛄 🋄 Apple 4shared Y	′ahoo! Google Maps Wikipedia Noticias▼ Populares▼ Flow	
▼ Advanced parameters	Note: Parameter values that differ from the default are highlighted in yellow	ń
Primer Pair Specificity Ch	ecking Parameters	- 11
Max number of Blast target	50000 0	- 11
Blast expect (E) value	30000 🛊 😡	- 11
Blast word size	7 ; .	- 11
Max primer pairs to screen	500 0	
Max targets to show (for designing new primers)	20 💿	
Max targets to show (for pre- designed primers)	1000	
Primer Parameters		
PCR Product Tm	Min Opt Max	
	Min Opt Max	
Primer Size	15 20 25 Min Max	
Primer GC content (%)	20.0 80.0	
GC clamp	0 😡	
Max Poly-X	5	
Max 3' Stability	9 😥	
Max GC in primer 3' end	5	
Secondary Structure Alignment Methods	Use Thermodynamic Oligo Alignment Use Thermodynamic Template Alignment (warning: search may be very slow with this option on) 🤬	
Alignment metroda	Primer Pair	
TH: Max Template Mispriming	40.00 [70.00 [For thermodynamic alignment model only]	
maprining	Any 3' 45.0 (For thermodynamic alignment model only)	
TH: Max Self Complementarity	Any 3'	
oomplementanty	45.0 35.0 (For thermodynamic alignment model only)	
TH: Max Pair Complementarity	24.0 (For thermodynamic alignment model only)	
TH: Max Primer Hairpin	Primer Pair	
Max Template Misoriming	12.00 24.00 (For old secondary structure alignment model only)	
Max Self Complementarity	Any 3' 8.00 3.00 (For old secondary structure alignment model only)	
max den complementanty	Any 3'	
Max Pair Complementarity	8.00 3.00 (For old secondary structure alignment model only)	
Excluded regions		
Overlap junctions		
	5'side overlaps 3'side overlaps	
	Minimal number of nucleotides that the left or the right primer must have at the 5' or 3' side of the junctions	
Concentration of	50.0	
Concentration of divalent	1.5	
Concentration of dNTPs	0.6	
Salt correction formula	SantaLucia 1998 🔹 😥	
Table of thermodynamic parameters	SantaLucia 1998 🛊 😡	
Annealing Oligo	50.0 😡	
SNP handling	Primer binding site may not contain known SNP 😡	
Repeat filter	Automatic + 0	
Low complexity filter	✓ Avoid low complexity region for primer selection	
Laboration to the first state of the		
Internal hybridization oligo	o parameters	
riyondization oligo	Hine Ont Nav	
Hyb Oligo Size	18 20 27	
	Min Opt Max	
Hyb Oligo tm	57.0 60.0 63.0	
Hyb Oligo GC%	Min Opt Max	
110 0190 00 //	1000 VV	
Get Primers	Show results in a new window 🗹 Use new graphic view 😡	
onvright   Disclaimer   Privacy   Accossibility	Contart   Send feedbark on new interfere NOT I M M I MUT PULL	A
statistis i statistici i rimary i ciccessionity		- L

Figura 6.4c. Ventana de característica Avanzadas.

	)			cron		Isuali	izaci	ion	Histo	orial I	Favori P	tos \ rimer-l	/entar Blast re	na A esults	Ayuda									
	•		+ 8	http	://ww	ww.nct	bi.nln	n.nih.	gov/to	ols/prir	ner-bl	ast/prir	nertool	.cgi?ct	tg_time	e=1376	67060	36&job	_ key	¢ (	Q- G(	ogle		
m		Apple	4share	d 1	Yahoo	o! Go	oogle	Maps	Wiki	ipedia	Notici	as▼ Po	opulare	s▼ Fl	low									
Pri	imer-E	BLAST									Prime	r-Blast	result	5										initi
	Primer-	BLAST :	results	s: Jo	b id=J	JSID_(	01_36	63478	_130.1	4.22.21	9004_	primert	ool m	ore										
s	Input	: PCR te	emplat Rang prime	te je rs	DI 1 pr	00659 - 239	<u>9.1</u> M 94 5 may	1us m y <b>not</b>	usculu be spe	us CYP r	nRNA f	or aron	natase templa	P450, ate as	comp	lete cd	s e founc	l in sele	ected	databa	ase:Re	efseq n	nRNA	
		Other	repor	ts	(C	Organi Searc	ism li ch Su	imitec Imma	i to Mu ry	is musc	ulus)	. <u>help or</u>	n specif	fic prin	ners									
	Grap	hical v	iew o	fpr	imer	r pair	rs							N 1 (				_	a a		1.184		~	0
	D Temp	00659.1: ate  2	12.4K	(2.4K	l400	Find	1 on Se	equenc	e: 0  800	0 1900	1 K	1.100	1.200	≥  -   1.300	1.400	1.500	1.600	+ AIG 1.700  1	× 1 008.1	1 1.900	2 K	Configu	ire 🥰	2.3
													1.1-0.0		17.00	1,000		1.1.1						
	Genes							CYF	,															
	BAA005	51.1		>				>					>			_								
	Prime	r pairs	for	D006	59.1																			
						Primer																		
						r men	- 8								Prin	mer 3 🗦								
						Primor	5 <b>b</b>								Prim	er 4		-4						
	<u>P</u>	rimer	pair	r 1																				
				Seq	uenc	e (5'->	>3')		1	Templa	te	Leng	th Star	t Stop	Tm	GC% <sup>S</sup>	Self			Self	3'			-
	F	orward		Seq	uenc	e (5'->	>3')		1	Templa strand	te	Leng	th Star	t Stop	Tm	GC% c	Self	menta	rity	Self com	3' Iplem	entarit	у	-
	F	orward rimer		Seq CGT	uenc rggA	e (5'-> TGTC	> <b>3')</b> Эттб	ACC	T S CTCA F	Templa strand Plus	te	Leng 20	th Star 598	<b>t Stop</b> 617	Tm 59.97	GC% c	Self comple	ementa	rity	Self com 3.00	3' Iplem	entarit	y	-
	F P R	orward rimer everse rimer		Seq CGT GGC	I <mark>uenc</mark> TGGA CCAC	e <b>(5'-&gt;</b> TGTC	>3') Эттс тстт	BACCO CACT	T S CTCA F	<b>Templa</b> strand Plus Minus	te	<b>Leng</b> 20 20	<b>th Star</b> 598 1570	t Stop 617 ) 1551	Tm 59.97 60.18	GC% 55.003	Self comple 3.00	ementa	rity	Self com 3.00 1.00	3' Iplem	entarit	ey.	-
	F P R P	orward rimer everse rimer roduct l	length	<b>Seq</b> CGT GG0 973	IUENC FGGA CCAC	e (5'-> .TGTC :ACCT	<b>&gt;3')</b> ЭТТС ТСТТ	BACCO CACT	T S CTCA F IGTT I	<b>Templa</b> strand Plus Vlinus	te	<b>Leng</b> 20 20	<b>th Star</b> 598 1570	t Stop 617 ) 1551	Tm 59.97 60.18	<b>GC% <sup>S</sup>c</b> 55.00 3 55.00 4	Self comple 3.00	ementa	rity	Self com 3.00 1.00	3' Iplem	entarit	у	-
	F P R P	orward rimer everse rimer roduct l	length	Seq CGT GGC 973	uenc rgga ccac	e (5'-> TGTC ACCT	>3') GTTG TCTT	BACCO CACT	T CTCAF IGTT I	Templa strand Plus Minus	te	Leng 20 20	th Star 598 1570	t Stop 617 ) 1551	Tm 59.97 60.18	GC% c 55.00 3 55.00 4	Self comple 3.00	ementa	rity	Self com 3.00 1.00	3' Iplem	entaril	y	-
	F P R P P	orward rimer everse rimer roduct l roducts	length s on pc 810.3 M	Seq CGT GGC 973 otent Mus	uenc rGGA CCAC tially musc	e (5'-> TGTC ACCT unint	>3') GTTG TCTT tende	GACCO CACT ed ten	TGTT I nplates	Templat strand Plus Minus s	te 19, su	Leng 20 20 bfamily	th Star 598 1570 a, poly	t Stop 617 0 1551 /peptid	Tm 59.97 60.18 de 1 (C	GC% c 55.00 3 55.00 4 55.00 4	<b>Self</b> comple 3.00 3.00	ementa	rity	Self com 3.00 1.00	3' plem	entarif	y	-
	F P P P >j F	orward rimer deverse rimer roduct l Products NM_007 roduct product product	length s on pc 810.3 M leng prim e	Seq CGT GGC 973 otent Mus th er	tially musc 1 593	e (5'-> ATGTC ACCT unint culus c	>3') GTTG TCTT tende cytoc	GACCO CACT ad ten throme	TGTT I nplates P450	Templa strand Plus Minus s , family ccctcA	19, su 19, su 10, su	Leng 20 20 bfamily	th Star 598 1570 a, poly	t Stop 617 0 1551 /peptid	Tm 59.97 60.18 de 1 (C	GC% 55.00 3 55.00 4 Syp19a	Self comple 3.00 1), mRI	ementa	rity	Self com 3.00 1.00	3' plem	entarit	y	_
	F P R P P S S T R T	orward rimer everse rimer roduct l roducts NM_007 roduct product product product product product product	length s on pc 810.3 M leng prim e prim e	Seq CGT GGC 973 otent Mus th er	uenc rGGA CCAC tially musc = 97 1 593 1 156	unint cG cG cG cG cG cG cG cG cG	>3') GTTG TCTT tende cytoc TGGA	GACCO CACT ed ten hrome	TGTT I nplates P450	Templat strand Plus Minus s , family CCCTCA CACTGT	te 19, su 612 TT 20 . 15	Leng 20 20 bfamily	th Star 598 1570 a, poly	617 0 1551 vpeptid	Tm 59.97	<b>GC% <sup>S</sup>c</b> 55.00 3 55.00 4	Self comple .00 .00	ementa	rity	Self com 3.00 1.00	3' plem	entarit	у	_
	F P P F F T R R T	orward rimer leverse rimer roduct l roducts NM 007 roducts everse emplat	e length leng prim e 128151	Seq CGT GGC 973 otent Mus th er her	rGGA CCAC tially musc = 97 1 593 1 156 Mus m	e (5'-> TGTC CACCT unint culus c 3 CG CG CG CG S 5	>3') GTTG ICTT tende cytoc GCCA	CACI	TGTT I nplatee e P450 STTGAG	Templa strand Plus Vinus s , family CCCTCA CACTGT	te 19, su 612 TT 20 • 15	Leng 20 20 bbfamily 2 5 46 5 46 5 46 5 46	th Star 598 1570 a, poly	617 0 1551 ypeptid	Tm 59.97 60.18 de 1 (C	<mark>вс% s</mark> 55.003 55.004 Хур19а <sup>-1</sup>	Self .00 .00 1), mRI	menta	rity	Self com 3.00 1.00	3' plem	entarit	у	_
	F P P P P P P P P P P P P P P P P P P P	orward rimer keverse rimer roducts NM_007 roducts orward emplat everse emplat NM_001 roduct everse emplat	length s on pc 810.3 k leng prim e prim e 128151 leng prim e	Seq CGT GGC 973 otent Mus th er 1.1 M th er	uenc (GGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797	e (5'-> TGTC CACCI unint culus c 3 CG 5	>3') GTTTG TCTT tende cytoc GCCP GCCP GCCP	ACCCC CACT ad terr ATGTCC ACACCC	TGTT / rplates b P450 GTTGA CTCTTY C Syndr	Templai strand Plus Minus s t, family ccctcA cactGI	19, su 20 612 T 2( . 15 T 2( G 79	Leng 20 20 bfamily 3 3 3 46 yme reg	598 1570 a, poly	617 0 1551 ypeptid	Tm 59.977 60.18 de 1 (C e 2 (Ce	GC% S c 555.003 555.004 Xyp19a	Gelf comple 00 00	menta	rity	Self com 3.00 1.00	3' plem	entarit	у	_
	F P P P P P P P P P P P P P P P P P P P	orward rimer rimer roduct I roducts NM 007/ roduct exerse emplat.	length son pc 810.3 M prim e 128151 leng prim e prim e	Seq CGT GGC 973 otent Mus th her 1.1 M th ier	uenc rGGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797 1 454	e (5'-> TGTC CACCT unint culus c 3 CG 5 · 6 5 · 6 5 · 7 24 22 C 22 C 29 T	>3') 3TTG TCTT tende cytoc TGGA GGCCA GGCCA C GCCA C C	CACI CACI cd ten hrome ATGTC ACACC ACACC C	rGTT / nplate: e P450 STTGA/ STTGA/ STTGA/ STTGA/ STTGA/ STTGA/ STTGA/ STTCTTY	Templai strand Plus Winus s t, family ccctcA cactGI cactGI	te 19, su 20 612 T 20 72 20 612 T 20 G 75 T 20 G 45 T 20 G 45	Leng 20 20 bfamily 22 ) 546 ome reg	th Star 598 1570 a, poly	617 0 1551 ypeptid	Tm 59.97 60.18 de 1 (C	GC% <sup>S</sup> c c 55.003 55.004 Xyp19a	Self .00 .00	ementa	rity	Self com 3.00	3' Iplem	entarif	у	_
	F P P P P F F T T R R T T R R T T R R	orward rimer weverse rimer roducts NM_007 roducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts coducts codu	length s on pc 810.3 k leng prim e leng prim e prim e 801.2 k	Seq CGT GGC 973 otent Mus tth her 1.1 N tth er Mus	uenc rGGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797 1 454 musc	e (5'-> TGTC CACCT unint culus c 3 CG 5	>3') GTTG TCTT tende cytoc TTGGA GCCA GCCA GCCA GCCA CTC mucir	ed terr cACT hrome ATGTC ACACC ACACC C ACACC C	TGTT / s CTCA F IGTT / nplates a P450 STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTGA STTG	Templai strand Plus Viinus s t, family ccctcta cactgt cactgt cactgt cactgt b, tracl	te 19, su 20 612 T 20 T 20 G 75 T 20 G 45 T 20 G 45 T 20	Leng 20 20 bbfamily 3 3 3 446 9 553 3 568 anchial (I	598 1570 a, poly	t Stop 617 ) 1551 //peptid ndidate	Tm 59.97 60.18 de 1 (C e 2 (Ce	GC% <sup>S</sup> c 55.003 55.004 355.004	self .oople .00 1), mRI	menta	rity	Self com 3.00 1.00	3' plem	entarit	y	_
	F P P P P P P P P P P P P P P P P P P P	orward rimer roduct I roducts NM 007/ roducts emplat: everse emplat. VM 001 roduct everse emplat.	length son pc 810.3 M prim e 128151 leng prim e 801.2 M leng prim e	Seq CGT GGC 973 otent Mus th her 1.1 N th her Wus th her	uenc rGGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797 1 454 musc = 97 1 127	uninte caccor uninte culus of caccor auscul 24 caccor auscul 22 caccor auscul 22 caccor auscul 22 caccor auscul 22 caccor caccor auscul 22 caccor auscul 22 caccor caccor auscul 22 caccor caccor caccor auscul 22 caccor caccor auscul 22 caccor caccor auscul 22 caccor caccor auscul 22 caccor caccor auscul 22 caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caccor caco caccor caccor cacco caccor caccor caco cacco caccor caco caco	>3') GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP GECCP	CACCO	IGTT I s cTCAF IGTT I IGTT	Templa strand Plus Vlinus s h, family ccctcA cactGT CactGT B, track GACCCT	19, su 20 612 T 2( T 2( G 75 T 2( G 45 T 2( G 45 T 2( CA 2 1	Leng 20 20 bbfamily 346 y53 355 668 nnchial (1 20 20 22740	th Star 598 1570 a, poly ion, car	t Stop 617 ) 1551 //peptid ndidate	Tm 59.97 60.18 de 1 (C e 2 (Ce	GC% S 555.003 555.004 Xyp19a*	self .oople .00	nmenta	rity	Self com 3.00 1.00	3' plem	entarii	У	
	F P P P P P P P P F T T T T T T T T T T	orward rimer roducts NM 007/ roducts emplat everse emplat roduct everse emplat roduct everse emplat roduct everse emplat	length son pc 810.3 N leng prim e 128151 leng prim e 801.2 N leng prim e e	Seq CGT GGC 973 Jotent Mus thh ier 1.1 M th ier Mus th ier Mus th	uenc rGGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797 1 454 musc = 97 1 127 1 127 1 127	a (5'-> ACC1 unint sulus c G CG CG CG CG CG CG CG CG CG	>3') GTTG TCTT tende cytoc GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA GCCA	CACCO CACT CACT CACT CACTO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACACO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CACTO CAC	TGTT I s rgTT I rgTT I s rcTTA s yndn c c c c t c t t t y c c c t t y e s yndn c c c t y c t g t g t y e s y t g t y e s y t s y s y t s y s y s y s y s y s y	Templaistrand Plus Vinus s t, family ccctCA cactGT cactGT B, tracl gaccct a	19, su 19, su 20 612 TT 20 G 75 TT 20 G 45 TT 20 G 45 TT 20 CA 2 A. 1 TT 2 A. 1	Leng 20 20 bbfamily 546 0 553 0 568 0 568 0 568 0 22740 20 22740 20 22740 20 22740 20 220 220 20 20 20 20 20 20 20 20 20	th Star 598 1570 a, poly ion, car	t Stop 617 ) 1551 vpeptid	Tm 59.97 60.18 de 1 (C e 2 (Ce	GC% S 55.003 55.004 Xyp19a'	self comple 00 00	NA	rity	Self com 3.00 1.00	3' plem	entarit	у	
	FPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP	orward rimer roduct I roducts NM 007/ roduct emplat verse emplat verse emplat verse emplat verse emplat verse emplat verse emplat verse emplat	length son pc 810.3 N leng prim e 1281511 leng prim e 801.2 N leng prim e 1177454	Seq CGT GGC 973 otent Mus er 1.1 M th er Mus th er Mus	uenc rGGA CCAC tially musc = 97 1 593 1 156 Mus m = 34 1 797 1 454 musc = 97 1 157 1 157 1 156 Mus m = 34 1 157 1 157 1 156 157 1 156 157 1 156 157 1 156 157 1 156 157 1 156 157 1 156 157 1 156 157 1 156 157 1 157 1 156 157 1 156 157 1 156 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 157 1 1 157 1 1 1 1 1 1 1 1 1 1 1 1 1	a (5'-3) ACCTI unint culus c 3 CG 5 CG 5 CG 6 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG 7 CG CG CG CG CG CG CG CG CG CG	>3') GTTG TCTT tende cytoc TGGZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ GCCZ	ed terr cACT hrome ACACC ACACC ACACC C ACACC C ACACC C ACACC C ACACC C ACACC C ACACC C ACACC C ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACT ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACAC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC ACACC AC	TGTT I splates b P450 STTGAA STTGAA STTGAA STTGAA STTGAA STGTTA STGTTA STGTTA STGTTA STGTTA STGTTA	Templa strand Plus Minus s t, family CCCTCA CACTGT CACTGT CACTGT CACTGT GACCCT CACTGT CACTGT CACTGT CACTGT CACTGT CACTGT CACTGT CACTGT	19, su 20 612 TT 20 G 75 TT 20 G 45 TT 20 G 45 TT 20 G 45 TT 20 CA 2 X. 1 S (Zfp5	Leng 20 20 bbfamily 2 546 546 568 nnchial (1 20 22740 20 11807 16), tran	th Star 598 1570 a, poly ion, car Muc5b)	t Stop 617 ) 1551 ndidate	Tm 59.97 60.18 de 1 (C e 2 (Ce IA	GC% S c55.003 555.004 Xyp19a <sup>+</sup>	self comple .00 .00	nna	rity	Self com 3.00 1.00	3' Iplem	entarit	у	_

Figura 6.4d. Ventana de resultados Primer-BLAST para el gene de la aromatasa de ratón.

Después de que se haga el trabajo, se abrirá una ventana mostrando los resultados (figura 6.4d). De manera gráfica, puede verse la posición de los mejores iniciadores. Recuerde que en este caso se pidió el diseño de un tercer iniciador que actuaría como sonda para la región. En cada caso, el apartado que marca cada par de iniciadores brinda la información adicional sobre la especificidad con la especie que se desea. En esta misma sección de resultados el programa muestra las características del amplicon que se obtendría así como las especificaciones necesarias para diseñar el perfil térmico para realizar la amplificación.



#### Actividad para reportar

Utilizando el programa Primer3 del MIT, obtenga 5 pares de iniciadores que podría utilizar para hacer la amplificación del receptor estrogénico beta del ratón a partir de ADN. Localice dentro del mRNA producto de la transcripción del gen cada uno de los pares de iniciadores sugeridos. Realice un análisis similar utilizando la herramienta Primer-BLAST.

Construya el perfil térmico que debería programar en el Termociclador para cada par de iniciadores obtenido en cada uno de los programas suponiendo que utiliza Taq polimerasa como enzima de amplificación. De acuerdo a las características anotadas en cada uno de los programas relacionadas a la concentración de iones monovalentes y bivalentes, indique la concentración de MgCl, y dNTP necesarias para la reacción de amplificación.

## Referencias

- SantaLucía, J. 2007. Physical principles and visual OMP software for optimal PCR design. En: Methods in Molecular Biology vol 40, PCR primer design (A. Yurev, ed). Humana Press (Totowa, NJ): 3-33.
- Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. 2012. Primer3- new capabilities and interfaces. Nucleic Acids res. 40:e115. Disponible en la URL <a href="http://nar.oxforjournals.org/content/40/15/e15.full.pdf">http://nar.oxforjournals.org/content/40/15/e15.full.pdf</a>+http://nar.oxforjournals.org/content/40/15/e15.full.pdf</a>+http://
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache I., Rozen, S., Madden, T.L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134 Disponible en el URL <a href="http://www.biomedcentral.com/content/pdf/1471-2105-13-134.pdf">http://www.biomedcentral.com/content/pdf/1471-2105-13-134.pdf</a>



## **Referencias generales**

- Baxevanos AD, Ouellette BFF. 2002. Bioinformatics: A practical guide to the analysis of genes and proteins. Jonh Wiley and Sons (New York).
- McEntyre A.L. 1998. Coffe break: tutorials for NCBI tools. Nat. Academic Press (Washington).
- Olson S, Beachy SH, Giammaria CF, Berger AC. 2002. Integrating large-scale genomic information into clinical practice. Nat. Academic Press (Washington).
- Pool R, Esnayra J. 2000. Bioinformatics: Converting data into knowledge. Nat. Academic Press (Washington).



#### Manual de prácticas de laboratorio. Introducción a la Bioinformática

Se terminó de imprimir en septiembre de 2013, con un tiraje de 200 ejemplares, más sobrantes para reposición.



UNIDAD IZTAPALAPA División de Ciencias Biológicas y de la Salud

Av. San Rafael Atlixco No.186, Col. Vicentina C.P. 09340, Del. Iztapalapa, México D.F. Tel.: (01) 58044600