

– Entrez

מנוע חיפוש במאגרי מידע באתר NCBI (National Center for Biotechnology Information) "המרכז הלאומי למידע ביוטכנולוגי". בעזרת שאילתה המורכבת ממילות חיפוש או מילות מפתח, הכלי מאתר במאגרי מידע רשומות העונות לתנאי החיפוש.



Entrez - מנוע חיפוש המאפשר לאתר פרטי מידע (רשומות) במאגרי מידע מתחומי הבריאות, הביולוגיה והביוטכנולוגיה באמצעות מילות חיפוש, הנקראות גם מילות מפתח. חיפוש מושכל כולל - בנוסף לצירוף של מילות החיפוש - גם הגבלות מתאימות, כגון הגבלה למאגר מידע שמתאים לסוג המולקולה הרצויה, הגבלה לשדה החיפוש ועוד. מקור השם Entrez בצרפתית, ומשמעותו - "הזן". המשתמשים מזינים את מילות המפתח למנוע החיפוש, ומנוע החיפוש מחפש רשומות אשר מכילות את מילות המפתח בכל מאגרי המידע המשויכים אליו. כדאי להקפיד בבחירת מילות מפתח מתאימות ולהפעיל הגבלות מושכלות על החיפוש כדי לקבל תוצאות נכונות וממוקדות.

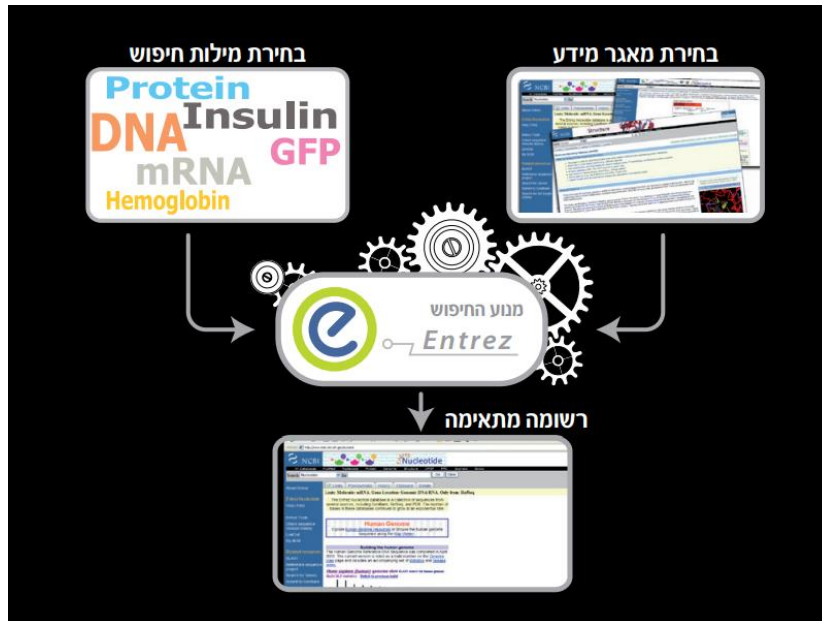
בין מאגרי המידע השונים שבהם מתבצע החיפוש ניתן למצוא מאגרים של רצפי נוקלאוטידים, מאגרים של רצפים או מבנים של חלבונים, מפות גנומיות, מאגרי ספרות ועוד. מאגרי המידע מכילים מספר רב של רשומות, לכן השימוש במילות מפתח ובהגבלות מתאימות הנו שלב חיוני לאיתור המידע המבוקש. השלב של חיפוש ואיתור רשומות רלוונטיות ממאגרי המידע באמצעות מנוע החיפוש הוא לרוב שלב ראשוני וחיוני במחקרים רבים.



Entrez - גישה לכלי בכתובת: <https://www.ncbi.nlm.nih.gov/search/>



ברוכים הבאים לסיור המודרך של מנוע החיפוש Entrez באתר NCBI. Entrez הוא מנוע חיפוש המאפשר לחפש פרטי מידע, הנקראים גם רשומות, במאגרי מידע מתחומי הבריאות, הביולוגיה והביוטכנולוגיה. חיפוש מושכל מתבצע באמצעות מילות מפתח, בדומה למנוע החיפוש Google, והגבלה למאגרי מידע מתאימים.



הגישה למנוע החיפוש נעשית מדף החיפוש של אתר NCBI (קישור מצורף בדף קודם). כאשר אנו רוצים למצוא מידע רלוונטי למחקר, עלינו לבחור את מילות המפתח המתאימות לחיפוש המבוקש וניתן גם לקבוע את מאגר המידע המתאים לסוג המידע שאנו מחפשים. בדף החיפוש של הכלי מוצגת רשימה של מאגרי המידע המתוחזקים על ידי NCBI ושניתן לבצע בהם חיפוש. לדוגמה, מאגר ה-Protein מכיל רצפי חלבונים, מאגר ה-Nucleotide מכיל רצפי נוקלאוטידים כגון מולקולות RNA של (mRNA), גנים ועוד.



Search NCBI

Search NCBI

Search

NCBI Databases

Literature

The World's largest repository of medical and scientific abstracts, full-text articles, books and reports

Bookshelf

Books and reports

MeSH

Ontology used for PubMed indexing

NLM Catalog

Books, journals and more in the NLM Collections

PubMed

Scientific and medical abstracts/citations

PubMed Central

Full-text journal articles

Genes

Gene sequences and annotations used as references for the study of orthologs structure, expression, and evolution

Gene

Collected information about gene loci

GEO DataSets

Functional genomics studies

GEO Profiles

Gene expression and molecular abundance profiles

HomoloGene

Homologous genes sets for selected organisms

PopSet

Sequence sets from phylogenetic and population studies

Proteins

Protein sequences, 3-D structures, and tools for the study of functional protein domains and active sites

Conserved Protein Domains Database

Conserved protein domains

Identical Protein Groups

Protein sequences grouped by identity

Protein

Protein sequences

Protein Clusters

Sequence similarity-based protein clusters

Sparcle

Functional categorization of proteins by domain architecture

Structure

Experimentally-determined biomolecular structures

Genomes

Genome sequence assemblies, large-scale functional genomics data, and source biological samples

Assembly

Genome assembly information

BioCollections

Museum, herbaria, and other biorepository collections

BioProject

Biological projects providing data to NCBI

BioSample

Description of biological samples

Genome

Genome sequencing projects by organism

Nucleotide

DNA and RNA sequences

Probe

Sequence-based probes and primers

SRA

High-throughput sequence reads

Genetics

Heritable DNA variations, associations with human pathologies, and clinical diagnostics and treatments

ClinVar

Human variations of clinical significance

dbGaP

Genotype/phenotype interaction studies

dbSNP

Short genetic variations

gnomAD

Genetic variation in diverse populations

Genetic Testing Registry

Genetic testing registry

MedGen

Medical genetics literature and links

OMIM

Online mendelian inheritance in man

Chemicals

Repository of chemical information, molecular pathways, and tools for bioactivity screening

BioSystems

Molecular pathways with links to genes, proteins and chemicals

PubChem BioAssay

Bioactivity screening studies

PubChem Compound

Chemical information with structures, information and links

PubChem Substance

Deposited substance and chemical information

חיפוש במאגרי מידע

בדוגמה שלפנינו נרצה לחפש את רצף הגן המקודד לנשא החמצן מיוגלובין באדם. כעת עלינו להזין את שאילתת החיפוש המכילה מילות מפתח בחלון החיפוש. בחירת מילות מפתח נכונות הנה חיונית לקבלת תוצאות מדויקות. בשלב הראשון נזין את מילת החיפוש Myoglobin בחלון החיפוש ונלחץ על Search.



Search NCBI

myoglobin



Search

לחצן להפעלת החיפוש

שאלתת החיפוש

Results by database

Results found in 27 databases for: **myoglobin**

Literature

Bookshelf	344
MeSH	14
NLM Catalog	40
PubMed	15,338
PubMed Central	21,642

Genes

Gene	1,331
GEO DataSets	125
GEO Profiles	5,900
HomoloGene	2
PopSet	230

Proteins

Conserved Domains	6
Identical Protein Groups	793
Protein	27,139
Protein Clusters	2
Sparcle	23
Structure	1,331

Genomes

Assembly	0
BioCollections	0
BioProject	0
BioSample	0
Genome	0
Nucleotide	29,794
Probe	122
SRA	0
Taxonomy	0

Genetics

ClinVar	94
dbGaP	0
dbVar	124
GTR	1
MedGen	20
OMIM	18

Chemicals

BioSystems	4,481
PubChem BioAssay	22
PubChem Compound	0
PubChem Substance	1,200

קישור למאגר רצפי נוקלאוטידים

Nucleotide

תוצאות החיפוש

המידע במאגרים מסודר ברשומות. התוצאה המתקבלת בעקבות החיפוש כוללת את מספר הרשומות הכוללות את שאלתת החיפוש, הנמצאות במאגרי המידע השונים וקישור אליהן. לדוגמה, במאגר המידע Nucleotide נמצא רשומות של רצפי נוקלאוטידים, גנים, מולקולות RNA ועוד. כל רצף ברשומה נפרדת.

בדוגמה שלנו, ניתן לראות בצילום המסך את מספר הרשומות שמכילות במקום כלשהו בתוכן, את המילה Myoglobin. לדוגמה, במאגר המידע Nucleotide נמצאו 29,794 רשומות. לחיצה על המאגר המבוקש יוביל לדף תוצאה הכולל הפניה לרשומות עצמן. מכיוון שנחפש רצף mRNA, נלחץ על מאגר הנוקלאוטידים.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide myoglobin Search

מאגר המידע הנבחר

רשומות על פי ממלכות עולם החי

רשומות על פי סוג המולקולה

רשומות על פי מאגרי המידע

שאלתת החיפוש

כלל הרשומות בהן מופיעה המילה Myoglobin

תוצאות חיפוש על פי אורגניזם

קישור לרשומה

סוג המולקולה ואורכה

רשומות שנמצאו בחיפוש

Results by taxon

Top Organisms Tree

Caenorhabditis elegans (53)

Crassostrea virginica (49)

Crassostrea gigas (42)

Strongylocentrotus purpuratus (40)

Biomphalaria glabrata (37)

All other taxa (1469)

Find related data

Database: Select

Find items

Search details

myoglobin[All Fields] AND (biomol_mrna[PROP] AND refseq[filter])

Search

See more...

Recent activity

Turn Off Clear

myoglobin AND (biomol_mrna[PROP] AND refseq[filter]) (1690) Nucleotide

Myoglobin and human AND (biomol_mrna[PROP] AND refseq[filter]) Nucleotide

דף התוצאה המתקבל כולל את רשימת רשומות. לכל רשומה מופיע תיאור קצר וקוד זיהוי ייחודי לה. בקישורים משמאל התוצאות מחולקות לקבוצות שונות. לדוגמה, אם נלחץ על קישור mRNA, נקבל את כל הרשומות שמתארות מולקולות RNA שליה ושמציינת בהן המילה מיוגלובין. אפשר גם לעיין ברשומות שמקורן בממלכות שונות בעולם החי לדוגמה Bacteria, במאגרי מידע שונים ועוד. מצד ימין של דף התוצאה יופיע קישור לרשומות של אורגניזמים שונים להם קיימות רשומות רבות עם מילת החיפוש המבוקשת.

חיפוש מושכל – שילוב מילות חיפוש

מספר התוצאות שהתקבל גדול מאוד, ולא כל הרשומות אכן מתארות את הגן מיוגלובין מאדם. כעת נמקד את החיפוש לאורגניזם הנכון, בדוגמה זו אדם. לשם כך נוסיף לשאלתה את המילה Human. בין שתי מילות המפתח עלינו להוסיף את המילה AND.

הנה כללים עיקריים לחיפוש נכון באמצעות מילות חיבור אשר נרשמות באותיות גדולות:

תנאי החיפוש	הפקודה
הרשומה תכיל בהכרח את שתי מילות החיפוש: מילה א' AND מילה ב'	מילה א' AND מילה ב'
הרשומה תכיל לפחות אחת משתי מילות החיפוש, אך לא בהכרח את שתיהן: מילה א' OR מילה ב'	מילה א' OR מילה ב'

כפי שניתן לראות בתמונה הבאה, ההגבלות שביצענו צמצמו מאוד את מספר הרשומות, וכל הרשומות שקיבלנו יכולו את שתי מילות המפתח. כדי לבחור את הרשומות המכילות רצפי אדם *Homo sapiens*, נבחר בצד ימין ברשומות אלו.

הגבלת החיפוש לסוג מולקולה ומאגר המידע המבוקשים

קישור לרשומות שכוללות רצפים שמקורם באדם

מספר הרשומות בהן מופיעות המילים Human-ו Myoglobin

מספר הרשומות בהן מופיעות המילים Human-ו Myoglobin

NCBI Nucleotide search results for "myoglobin AND human". The search filters show "mRNA (46)" and "RefSeq (46)". The results list shows "Homo sapiens myoglobin (MB), transcript variant 4, mRNA" and "Homo sapiens myoglobin (MB), transcript variant 1, mRNA". The "RefSeq Sequences" section shows the sequence details for "Homo sapiens myoglobin (MB), transcript variant 4, mRNA".

חיפוש מושכל - הגבלות חיפוש נוספות

אנו יכולים למקד יותר את החיפוש על ידי שימוש בהגבלות שונות המופיעות במסך התוצאה מצד שמאל. במסך זה ניתן להגביל את החיפוש לסוג המולקולה, לסוג מאגר המידע ועוד. בדוגמה זו נחפש רק מולקולות שהן RNA שליח. בשדה Molecule types ניתן לראות כי כלי החיפוש בחר באופן אוטומטי באפשרות mRNA, אך ניתן לשנות אפשרות זאת במידת הצורך. בנוסף, נרצה גם להגביל את סוג המאגר שבו מתבצע החיפוש. המאגרים השונים נבדלים זה מזה בתכונות שונות. מאגר ה-RefSeq הוא מאגר אמין לחיפוש רשומות נוקלאוטידים, לכן בשדה ה-Source Database כלי החיפוש יציע לבחור את הקישור לרשומות הרלוונטיות במאגר Refseq. במידה וכלי החיפוש לא בחר בצורה אוטומטית נכונה את הגבלות החיפוש המתאימות, ניתן לבחור או לשנות אותן.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (myoglobin AND human) AND "Homo sapiens"[porgn:__bid9606] Search

Create alert Advanced Help

Species Animals(14) Customize ...

Molecule types clear mRNA(14) Customize ...

Source databases clear

RefSeq(14) Customize ...

Sequence Type clear

Nucleotide(14)

Sequence length Custom range...

Release date Custom range...

Revision date Custom range...

Clear all Show additional filters

Summary 20 per page Sort by Default order

Filters activated: mRNA

Items: 14

1. [Homo sapiens myoglobin \(MB\), transcript variant 4, mRNA](#)
1,058 bp linear mRNA
Accession: NM_001362846.2 GI: 1674987968
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Homo sapiens myoglobin \(MB\), transcript variant 1, mRNA](#)
1,070 bp linear mRNA
Accession: NM_005368.3 GI: 1619243913
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

3. [Homo sapiens myoglobin \(MB\), transcript variant 3, mRNA](#)
1,153 bp linear mRNA
Accession: NM_203378.1 GI: 44955887
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

4. [Homo sapiens myoglobin \(MB\), transcript variant 2, mRNA](#)
1,170 bp linear mRNA
Accession: NM_203377.1 GI: 44955884
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

5. [Homo sapiens cytochrome b5 type A \(CYB5A\), transcript variant 3, mRNA](#)
3,201 bp linear mRNA

Send to: Filters: Manage Filters

Analyze these sequences
Run BLAST
Find in these sequences

Find related data
Database: Select
Find items

Search details
(myoglobin[All Fields] AND ("Homo sapiens"[Organism] OR human[All Fields])) AND "Homo sapiens"[porgn] AND (biomol_mrna[PROP] AND refseq[filter] AND is_nucleotide[filter])
Search See more...

Recent activity
Turn Off Clear
Q (myoglobin AND human) AND "Homo sapiens"[porgn] AND (biomol_mrna[PROP] AND refseq[filter] AND is_nucleotide[filter])
Q myoglobin AND human AND (biomol_mrna[PROP] AND refseq[filter] AND is_nucleotide[filter])
Q myoglobin AND human AND (biomol_mrna[PROP] AND refseq[filter] AND is_nucleotide[filter])

לאחר הגבלת חיפוש הרשומות על פי האורגניזם המבוקש, סוג המולקולה ומאגר הרשומות, מספר הרשומות העונות לתנאי החיפוש הצטמצם מאוד, וכעת יהיה לנו קל לבדוק בדקדקנות כל רשומה בנפרד. שימו לב, אם ידוע לנו קוד הזיהוי של הרשומה, ניתן להכניס אותו ישירות לחלון החיפוש. בחלק זה למדנו כיצד לבצע חיפוש יעיל במאגר מידע מסוים על ידי שימוש במילות חיפוש מתאימות. ראינו שלרוב חיפוש פשוט מניב הרבה תוצאות, ולכן יש למקד את החיפוש באמצעות הגבלות שונות.

ניתוח רשומה

נבחר רשומה מסוימת ונלמד להפיק ממנה מידע רלוונטי. לחיצה על הקישור לרשומה תפתח את הרשומה. הרשומות הן בעלות מבנה קבוע הכולל את שמות השדות בצדו השמאלי של הדף ופירוט המידע בצדו הימני של הדף. במאגר מסוים השדות קבועים בכל הרשומות, ורק המידע הנלווה אליהם משתנה מרשומה לרשומה.

שדות ברשומה

פירוט המידע בכל שדה

Homo sapiens myoglobin (MB), transcript variant 2, mRNA

NCBI Reference Sequence: NM_203377.1

LOCUS	NM_203377	1170 bp	mRNA	linear	PRI 24-DEC-2010
DEFINITION	Homo sapiens myoglobin (MB), transcript variant 2, mRNA.				
ACCESSION	NM_203377				
VERSION	NM_203377.1 GI:44955884				
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
COMMENT	REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from CB140824.1, BQ956082.1 and BC014547.1. This sequence is a reference standard in the RefSeqGene project. Summary: This gene encodes a member of the globin superfamily and is expressed in skeletal and cardiac muscles. The encoded protein is a haemoprotein contributing to intracellular oxygen storage and transcellular facilitated diffusion of oxygen. At least three alternatively spliced transcript variants encoding the same protein have been reported. [provided by RefSeq]. Transcript Variant: This variant (2) differs in the 5' UTR compared to variant 1. All three variants encode the same protein.				

מאפיינים: קוד זיהוי, אורך הרצף וסוג המולקולה

הגדרה: תיאור קצר של הרצף ומקורו

מקור: האורגניזם שממנו לקוח הרצף

מאפיינים חשובים ברצף

גן: גבולות הגן ושמו

אקסון: גבולות האקסון

CDS (coding sequence): גבולות הרצף המקודד

הרצף: פירוט רצף ה-mRNA

FEATURES

source

gene

exon

CDS

ORIGIN

Location/Qualifiers

1..1170

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="22"

/map="22q13.1"

1..1170

/gene="MB"

/gene_synonym="MGC13548; PVALB"

/note="myoglobin"

/db_xref="GeneID:4151"

/db_xref="HGNC:6915"

/db_xref="HPRD:01170"

/db_xref="MIM:160000"

1..164

/gene="MB"

/gene_synonym="MGC13548; PVALB"

/inference="alignment:Splign"

/number=1

173..637

/gene="MB"

/gene_synonym="MGC13548; PVALB"

/codon_start=1

/product="myoglobin"

/protein_id="NP_976311.1"

/db_xref="GI:44955885"

/db_xref="CCDS:CCDS13917.1"

/db_xref="GeneID:4151"

/db_xref="HGNC:6915"

/db_xref="HPRD:01170"

/db_xref="MIM:160000"

/translation="MGLSDGEWQLVLNVGKVEADIPGHGQEVILIRLFKQHPETLEKFKDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGMNKALELFRKDMASNYKELGFQC"

קישור לרשומת רצף החלבון

להלן מספר שדות מפתח במאגר רצפי נוקלאוטידים:

שם השדה	פירוט המידע
LOCUS	מציין את קוד הזיהוי, אורך הרצף וסוג המולקולה.
DEFINITION	מתאר בקצרה את הרצף ומקורו.
ORGANISM	מציין את האורגניזם שממנו לקוח הרצף.
FEATURES	מתאר מאפיינים חשובים ברצף.
תת-השדה GENE	מציין את גבולות הגן כולל אקסונים ואינטרונים.
תת-השדה EXON	מציין את הגבולות של כל אקסון.
תת-השדה CDS (CoDing Sequence)	מציין את גבולות הרצף המקודד בלבד.
ORIGIN	מפרט את רצף ה-RNA שליוח עצמו.

מבנה רשומה וקישור בין מאגרי מידע

מאגרי המידע השונים מקושרים מאוד זה לזה. לדוגמה, מתוך הרשומה המתארת את ה-RNA שליוח של מיוגלובין, ניתן לאתר רשומות רלוונטיות ממאגרי מידע אחרים כדוגמת רצף החלבון מיוגלובין ממאגר של רצפי החלבונים. גם כאן המידע מאורגן בשדות קבועים המכילים מידע רלוונטי וקישורים. לדוגמה, בשדה FEATURES קיים תת-שדה Protein המתאר את אורך החלבון ואת שמו; בתת השדה Region מצוינים אזורים בחלבון בעלי חשיבות תפקודית. המקטע המשתרע מחומצת אמינו 4 עד 143 הוא בעל חשיבות לתפקוד החלבון, ומקטע זה נקרא גלובין. תת-השדה Site מציין חומצות אמינו בעלות חשיבות תפקודית, כגון קישור קבוצת ההימ. בשדה ORIGIN ניתן לראות את רצף החלבון עצמו.

שדות ברשומה

פירוט המידע בכל שדה

מאפיינים:
קוד זיהוי, אורך הרצף וסוג המולקולה

הגדרה:
תיאור קצר של הרצף ומקורו

מקור:
האורגניזם שממנו לקוח הרצף

myoglobin [Homo sapiens]
NCBI Reference Sequence: NP_976311.1

LOCUS	NP_976311	154 aa	linear	PRI 24-DEC-2010
DEFINITION	myoglobin [Homo sapiens].			
ACCESSION	NP_976311			
VERSION	NP_976311.1 GI:44955885			
DBSOURCE	REFSEQ: accession NM_203377.1			
KEYWORDS	.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.			
COMMENT	REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from CB140824.1, BQ956082.1 and BC014547.1. This sequence is a reference standard in the RefSeqGene project. Summary: This gene encodes a member of the globin superfamily and is expressed in skeletal and cardiac muscles. The encoded protein is a haemoprotein contributing to intracellular oxygen storage and transcellular facilitated diffusion of oxygen. At least three alternatively spliced transcript variants encoding the same protein have been reported. [provided by RefSeq]. Transcript Variant: This variant (2) differs in the 5' UTR compared to variant 1. All three variants encode the same protein.			

מאפיינים חשובים ברצף

חלבון: [Protein](#)
אורך החלבון ושמו

איזור: [Region](#)
איזורים בחלבון בעלי חשיבות תפקודית

אתר: [Site](#)
חומצות אמינו בעלות חשיבות תפקודית

הרצף: [CDS](#)
פירוט רצף החלבון

סיכום

בסיום זה נחשפנו לחיפוש במאגרי המידע בעזרת מנוע החיפוש Entrez. למדנו על מבנה הרשומות וכיצד ניתן לנתח אותן ולהפיק מהן מידע ביולוגי שימושי.

אתם מוזמנים להתנסות ולתרגל את השימוש בכלי זה, ואנו מקווים שגם תיהנו. בהצלחה!