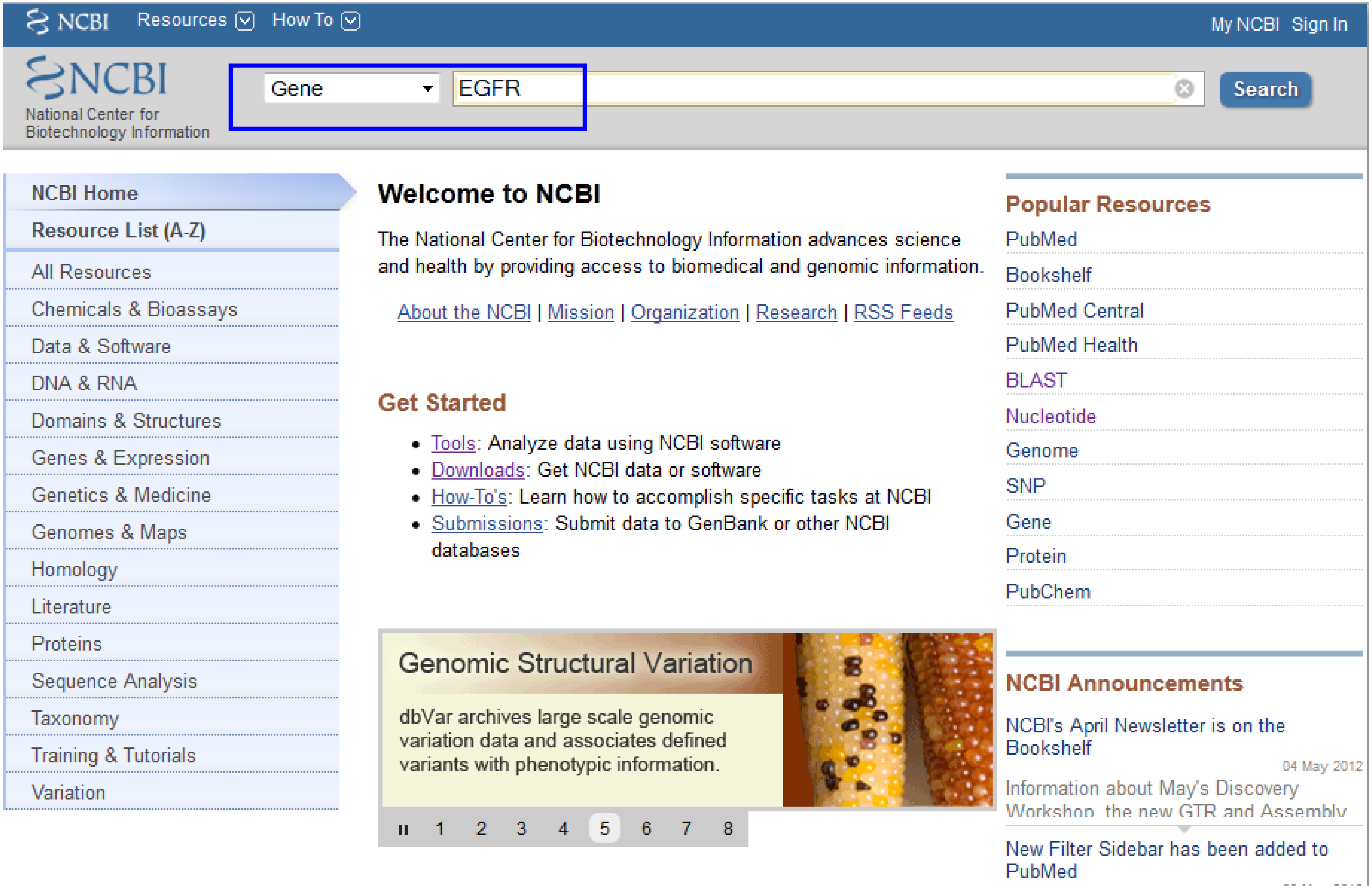
Download GenBank from NCBI

- Download NG or NC Accession
- Download NT Accession
- Save GenBank



Gene ▾ EGFR

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

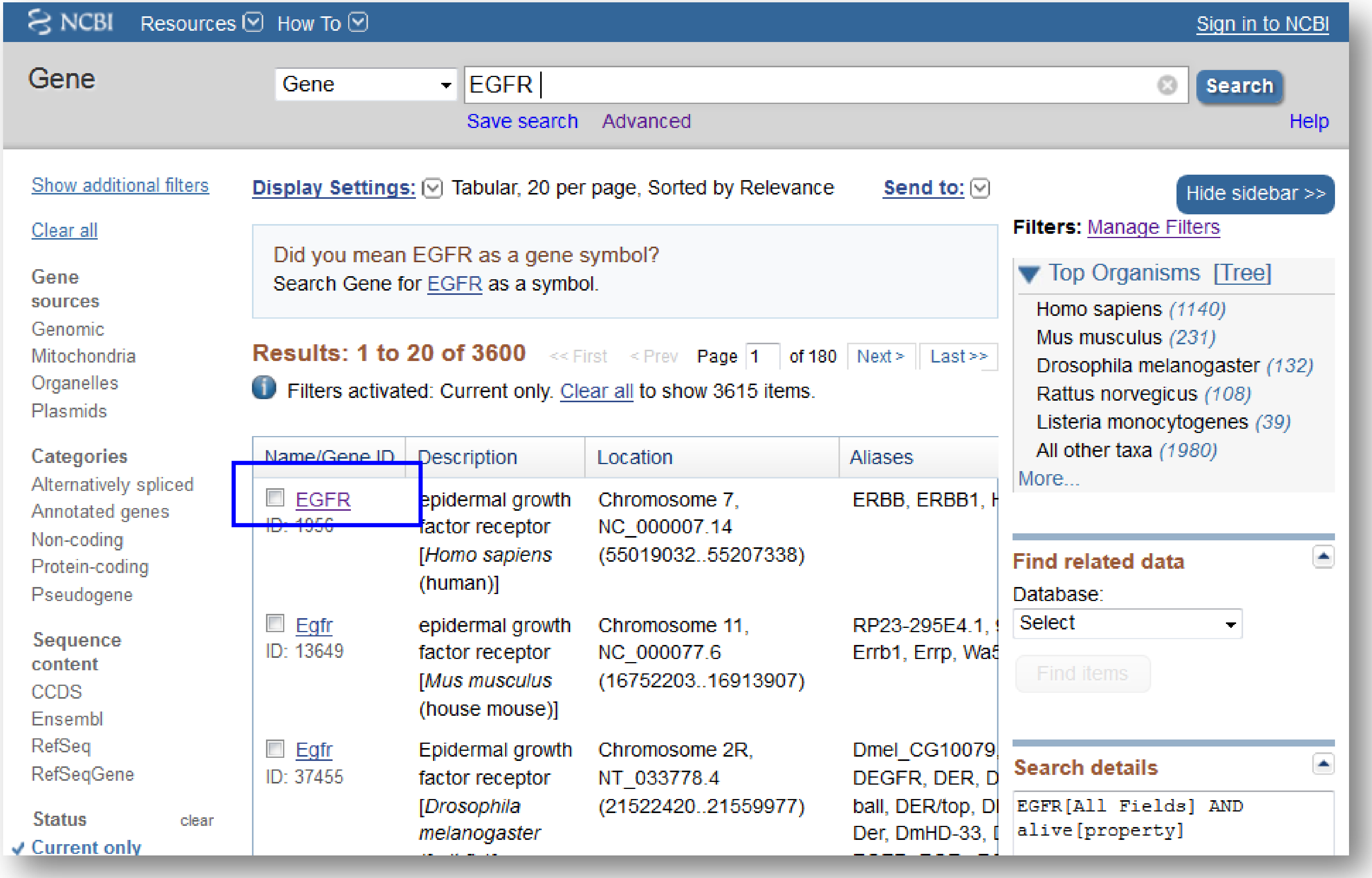
NCBI Announcements

NCBI's April Newsletter is on the Bookshelf

04 May 2012

Information about May's Discovery Workshop, the new GTR and Assembly

New Filter Sidebar has been added to PubMed



Search: EGFR

Display Settings: Tabular, 20 per page, Sorted by Relevance

Did you mean EGFR as a gene symbol?
Search Gene for EGFR as a symbol.

Results: 1 to 20 of 3600

Page 1 of 180

Filters activated: Current only. Clear all to show 3615 items.

Name/Gene ID	Description	Location	Aliases
☐ EGFR ID: 1956	epidermal growth factor receptor [*Homo sapiens* (human)]	Chromosome 7, NC_000007.14 (55019032..55207338)	ERBB, ERBB1, H
☐ Egfr ID: 13649	epidermal growth factor receptor [*Mus musculus* (house mouse)]	Chromosome 11, NC_000077.6 (16752203..16913907)	RP23-295E4.1, Errb1, Errp, Wa5
☐ Egfr ID: 37455	Epidermal growth factor receptor [*Drosophila melanogaster*	Chromosome 2R, NT_033778.4 (21522420..21559977)	Dmel_CG10079, DEGFR, DER, D ball, DER/top, DI Der, DmHD-33, [

Gene sources
Genomic
Mitochondria
Organelles
Plasmids

Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Pseudogene

Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene

Status clear
✓ Current only

Top Organisms [Tree]

Homo sapiens	*(1140)*
Mus musculus	*(231)*
Drosophila melanogaster	*(132)*
Rattus norvegicus	*(108)*
Listeria monocytogenes	*(39)*
All other taxa	*(1980)*

More...

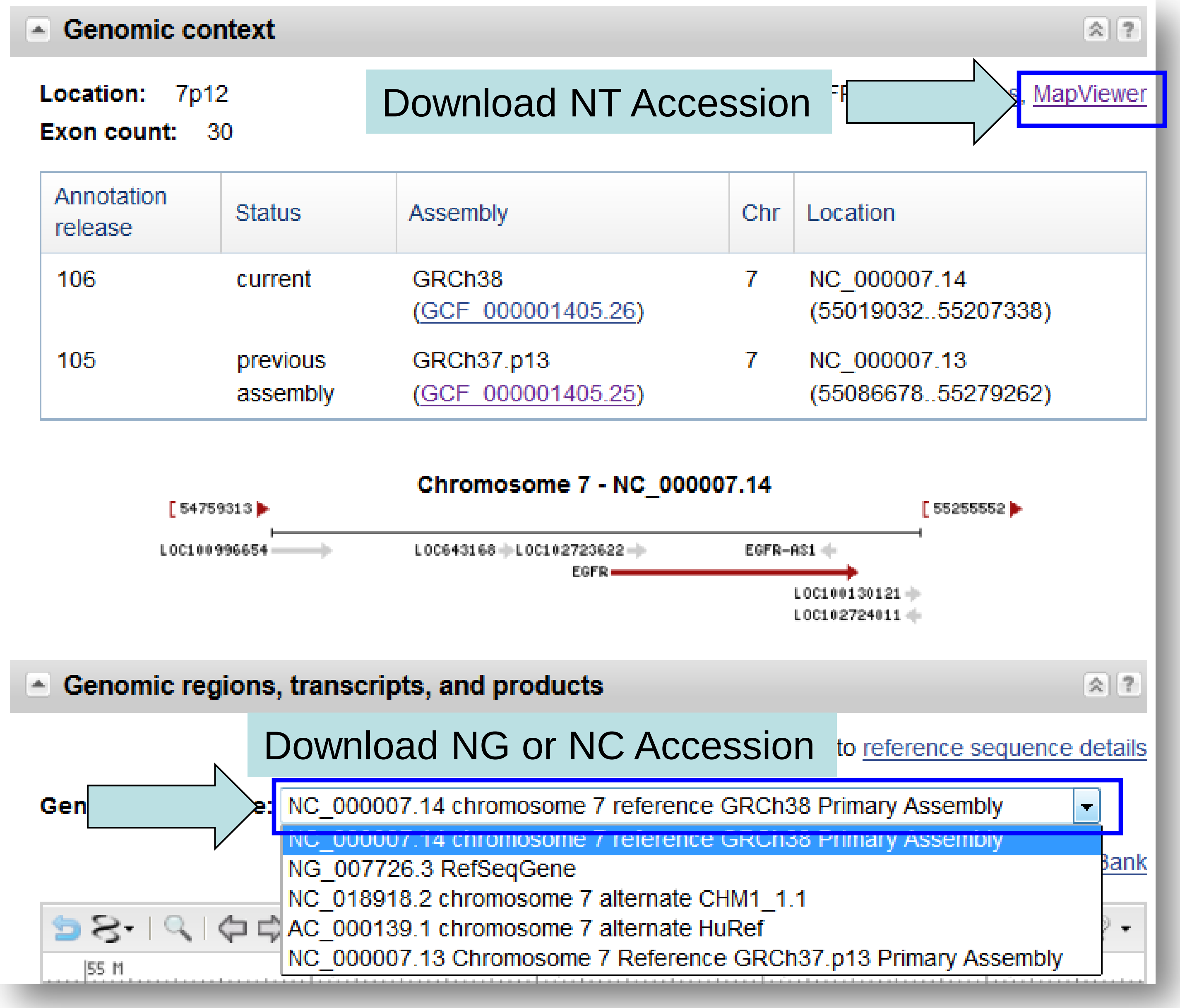
Find related data

Database: Select

Search details

```
EGFR[All Fields] AND
alive[property]
```

Scroll down to Genomic regions, select the appropriate assembly



Download NT Accession

Download NG or NC Accession

Genomic context

Location: 7p12

Exon count: 30

MapViewer

Annotation release	Status	Assembly	Chr	Location
106	current	GRCh38 (GCF_000001405.26)	7	NC_000007.14 (55019032..55207338)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	7	NC_000007.13 (55086678..55279262)

Chromosome 7 - NC_000007.14

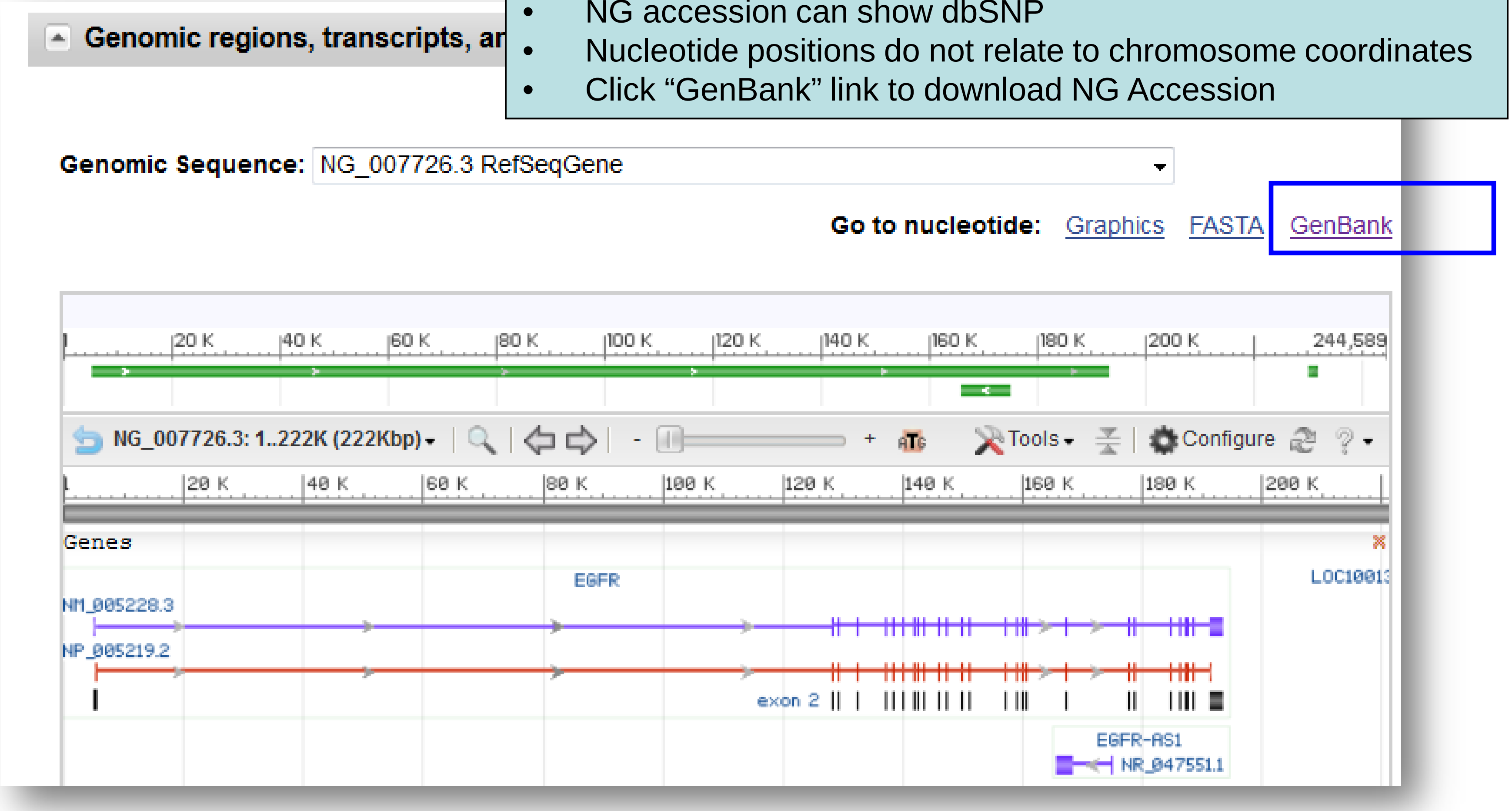
Genomic regions, transcripts, and products

to reference sequence details

Gen **e:** NC_000007.14 chromosome 7 reference GRCh38 Primary Assembly

- NC_000007.14 chromosome 7 reference GRCh38 Primary Assembly
- NG_007726.3 RefSeqGene
- NC_018918.2 chromosome 7 alternate CHM1_1.1
- AC_000139.1 chromosome 7 alternate HuRef
- NC_000007.13 Chromosome 7 Reference GRCh37.p13 Primary Assembly

- NG accession is the RefSeq
- Most RefSeq GenBanks contain only a single transcript
- NG accession can show dbSNP
- Nucleotide positions do not relate to chromosome coordinates
- Click “GenBank” link to download NG Accession

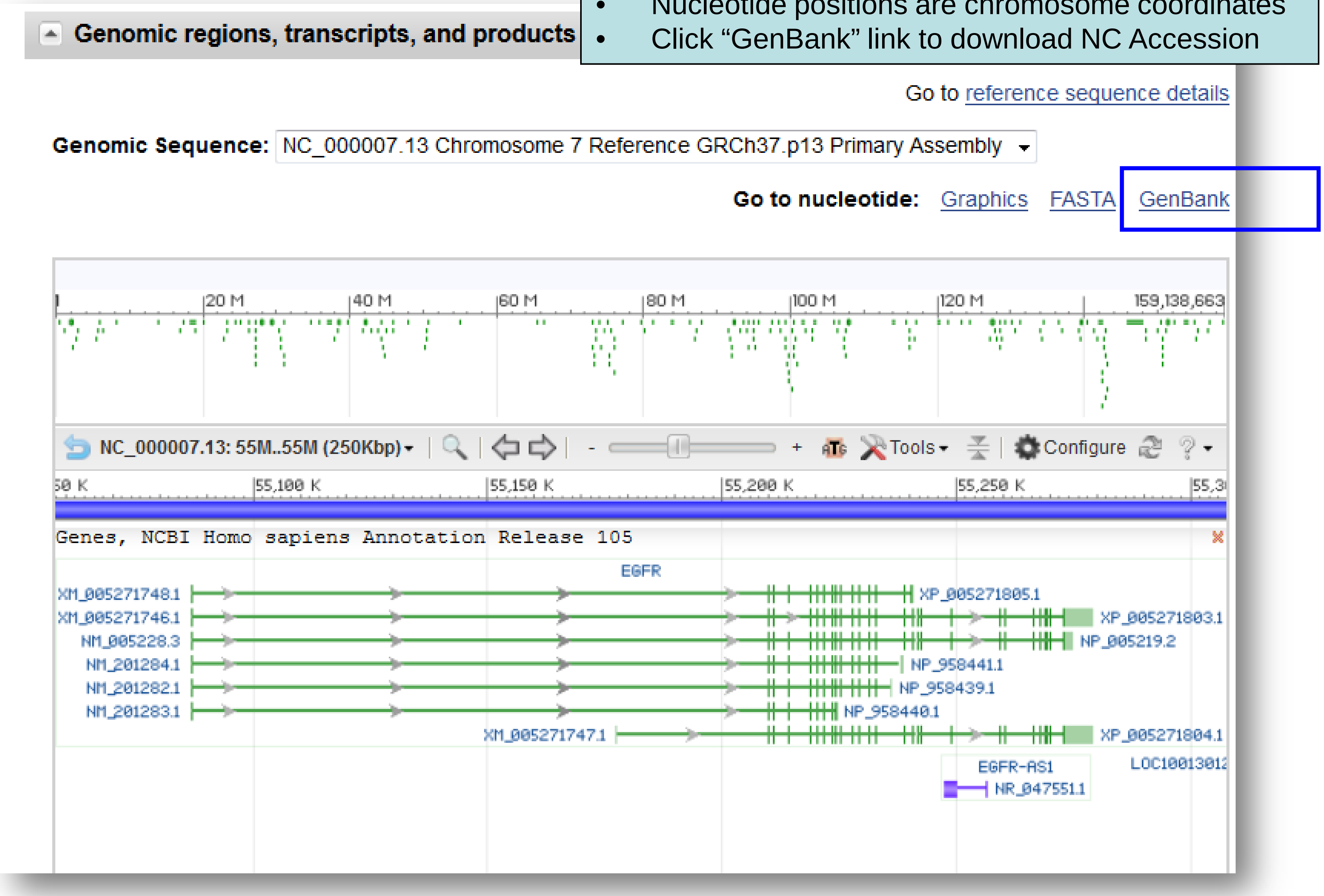


Genomic regions, transcripts, ar

Genomic Sequence: NG_007726.3 RefSeqGene

Go to nucleotide: Graphics FASTA GenBank

- NC Accessions often contains Several Transcripts
- NC Accessions do not have dbSNP
- Nucleotide positions are chromosome coordinates
- Click “GenBank” link to download NC Accession

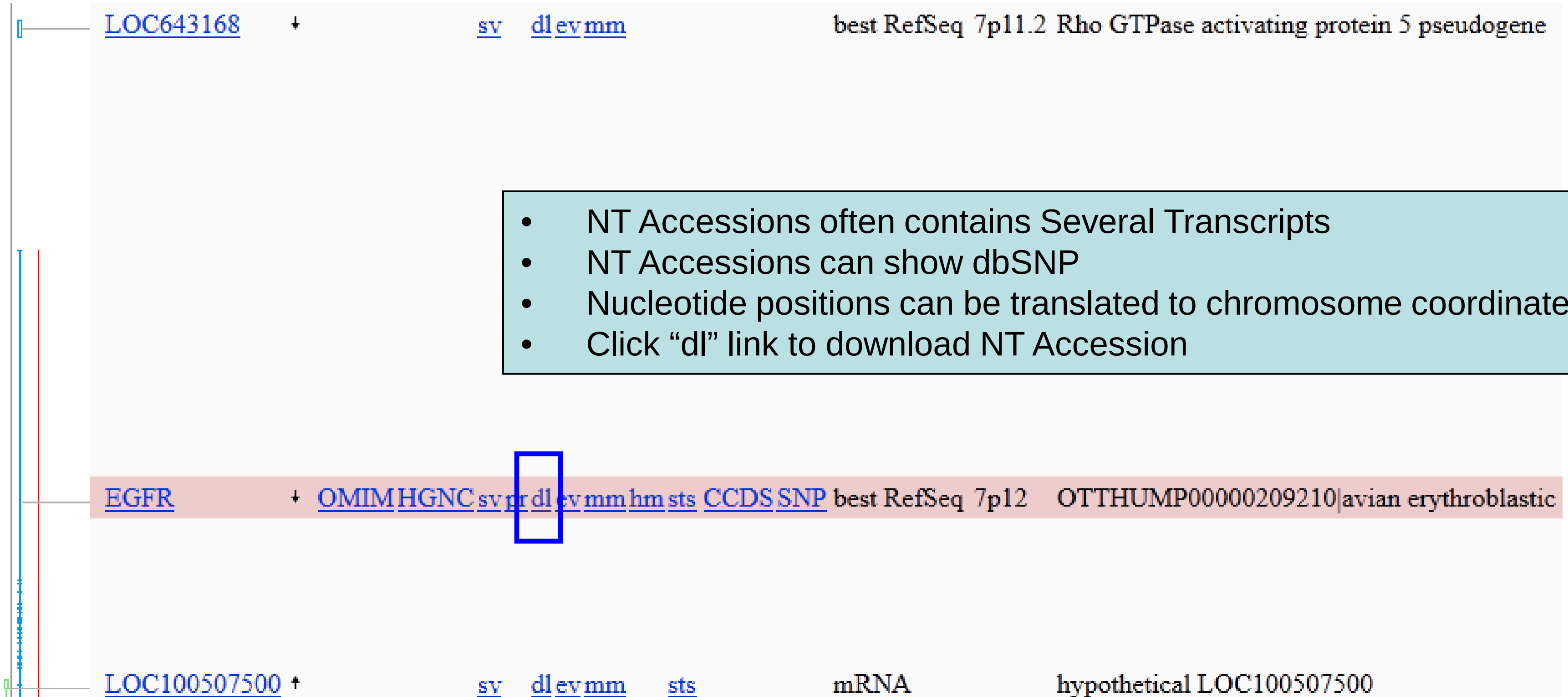


Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_000007.13 Chromosome 7 Reference GRCh37.p13 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

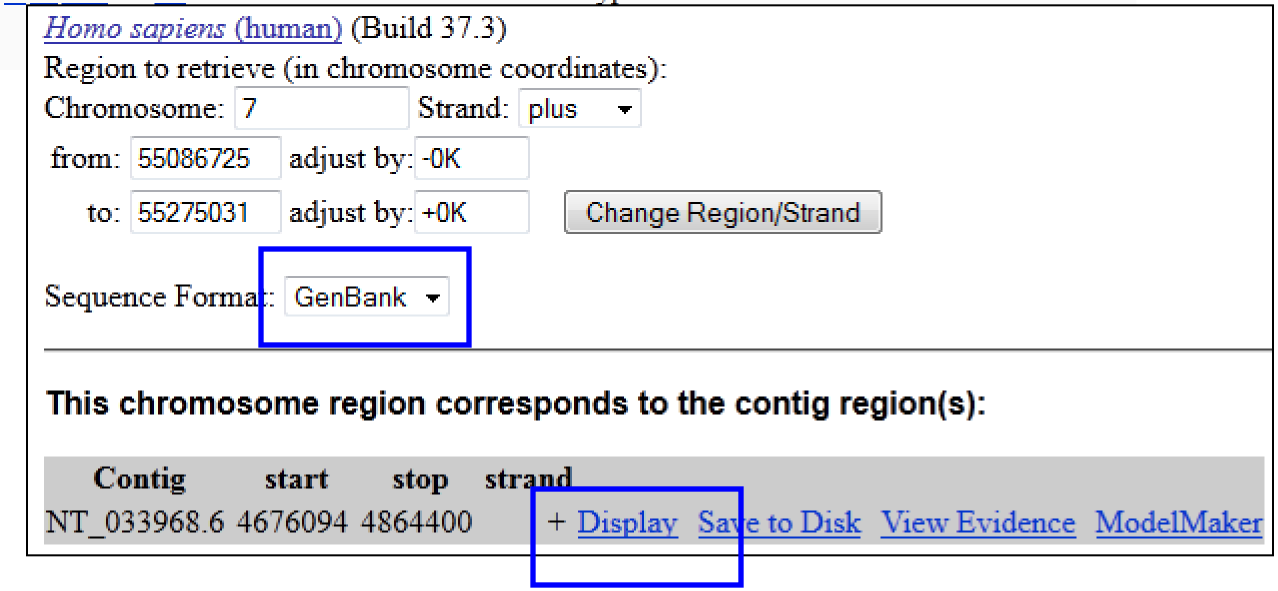


LOC643168 sv dl ev mm best RefSeq 7p11.2 Rho GTPase activating protein 5 pseudogene

EGFR OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq 7p12 OTTHUMP00000209210|avian erythroblastic

LOC100507500 sv dl ev mm sts mRNA hypothetical LOC100507500

- NT Accessions often contains Several Transcripts
- NT Accessions can show dbSNP
- Nucleotide positions can be translated to chromosome coordinates
- Click “dl” link to download NT Accession



Homo sapiens (human) (Build 37.3)

Region to retrieve (in chromosome coordinates):

Chromosome: 7 Strand: plus

from: 55086725 adjust by: -0K

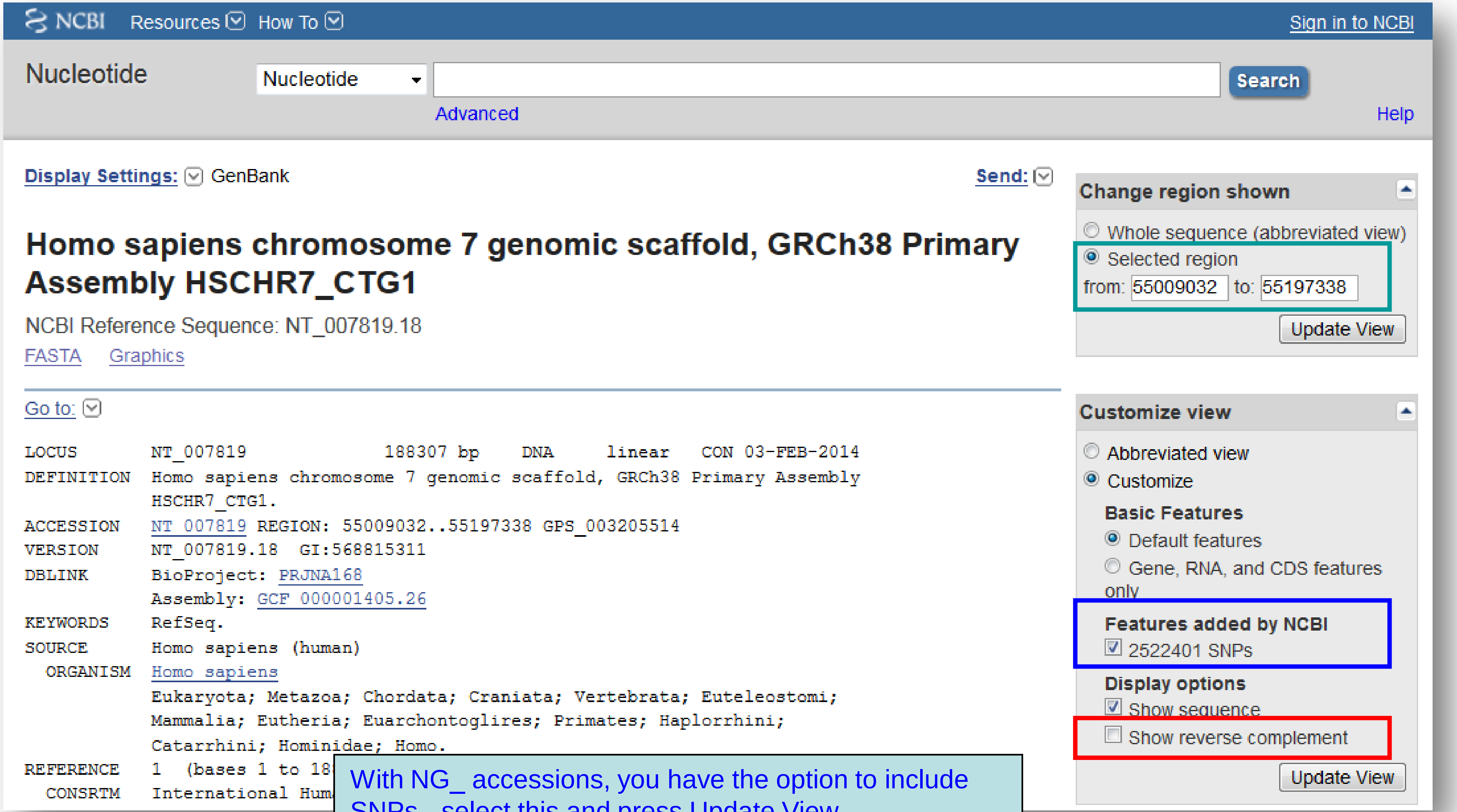
to: 55275031 adjust by: +0K

Sequence Format: GenBank

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand
NT_033968.6	4676094	4864400	

Save GenBank



Display Settings: GenBank

Send:

Homo sapiens chromosome 7 genomic scaffold, GRCh38 Primary Assembly HSCHR7_CTG1

NCBI Reference Sequence: NT_007819.18

FASTA Graphics

Go to:

```
LOCUS       NT_007819            188307 bp    DNA     linear   CON 03-FEB-2014
DEFINITION  Homo sapiens chromosome 7 genomic scaffold, GRCh38 Primary Assembly
            HSCHR7_CTG1.
ACCESSION   NT_007819 REGION: 55009032..55197338 GPS_003205514
VERSION     NT_007819.18  GI:568815311
DBLINK      BioProject: PRJNA168
            Assembly: GCF_000001405.26
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 18
  CONSRTM   International Hum
```

Change region shown

○ Whole sequence (abbreviated view)
● Selected region
from: 55009032 to: 55197338

Update View

Customize view

○ Abbreviated view
● Customize

Basic Features

● Default features
○ Gene, RNA, and CDS features only

Features added by NCBI

☑ 2522401 SNPs

Display options

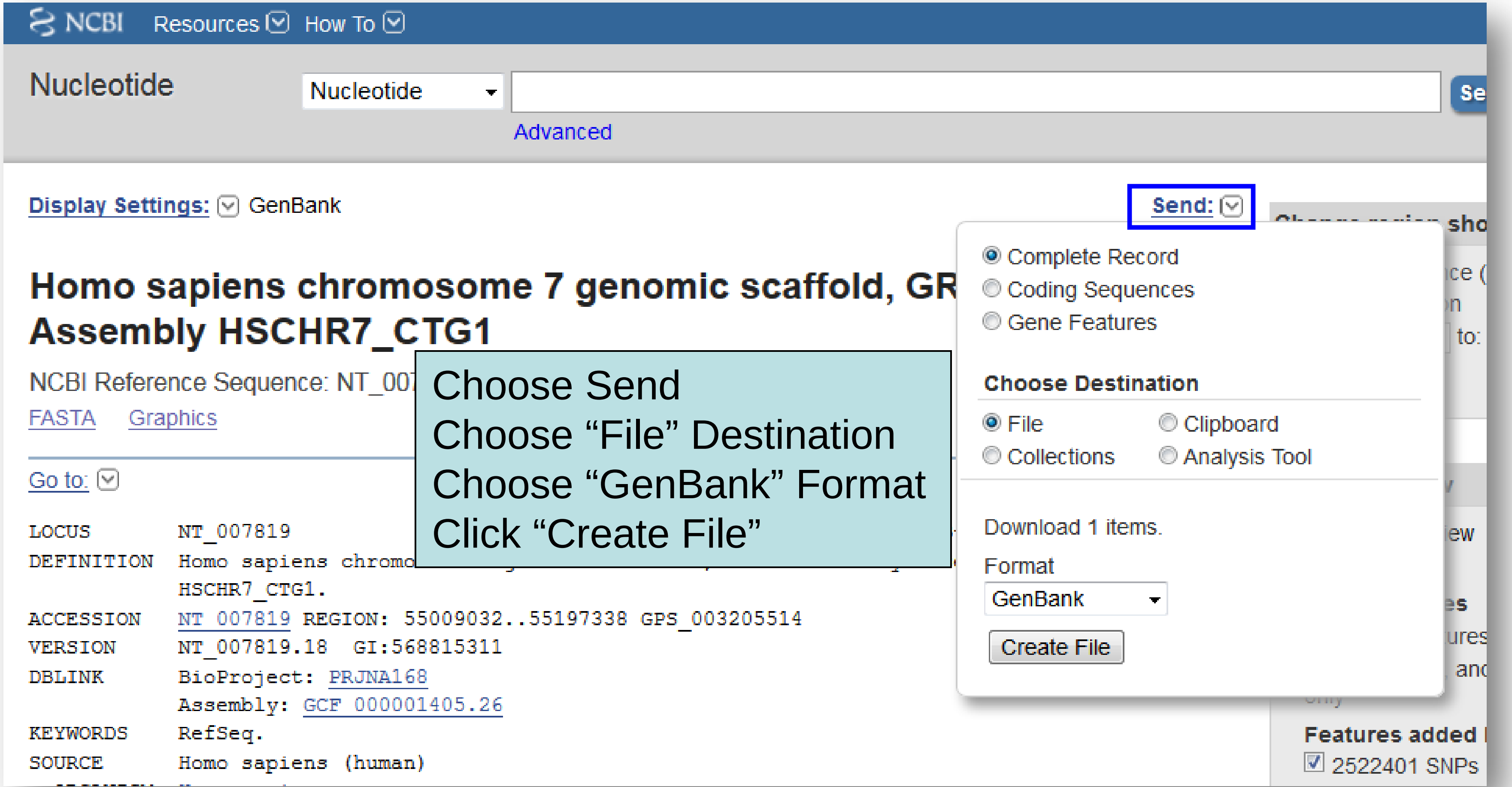
☑ Show sequence
☐ Show reverse complement

Update View

With NG_ accessions, you have the option to include SNPs - select this and press Update View.

For genes located on minus strand, deselect Reverse complemented strand and press Update View.

Expand the range if you are interested in the gene’s promoter region or other information outside of the gene, then press Update View.



Display Settings: GenBank

Homo sapiens chromosome 7 genomic scaffold, GR
Assembly HSCHR7_CTG1

NCBI Reference Sequence: NT_007

```
LOCUS       NT_007819
DEFINITION  Homo sapiens chromo
            HSCHR7_CTG1.
ACCESSION   NT_007819 REGION: 55009032..55197338 GPS_003205514
VERSION     NT_007819.18  GI:568815311
DBLINK      BioProject: PRJNA168
            Assembly: GCF_000001405.26
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
```

● Complete Record
○ Coding Sequences
○ Gene Features

Choose Destination

● File ○ Clipboard
○ Collections ○ Analysis Tool

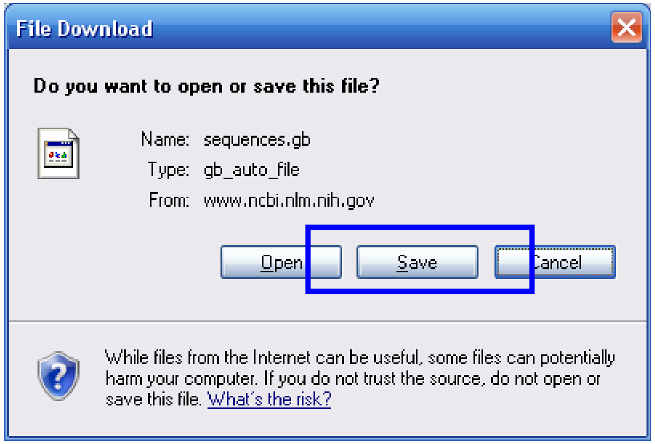
Download 1 items.

Format: GenBank

Features added

☑ 2522401 SNPs

Choose Send
Choose “File” Destination
Choose “GenBank” Format
Click “Create File”

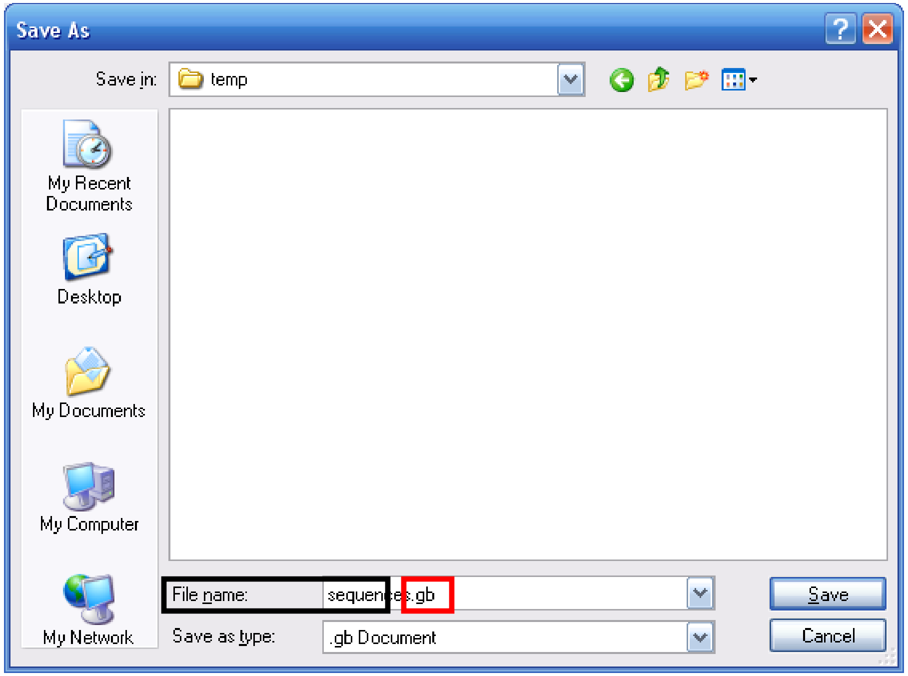


Do you want to open or save this file?

Name: sequences.gb
Type: gb_auto_file
From: www.ncbi.nlm.nih.gov

While files from the Internet can be useful, some files can potentially harm your computer. If you do not trust the source, do not open or save this file. What's the risk?

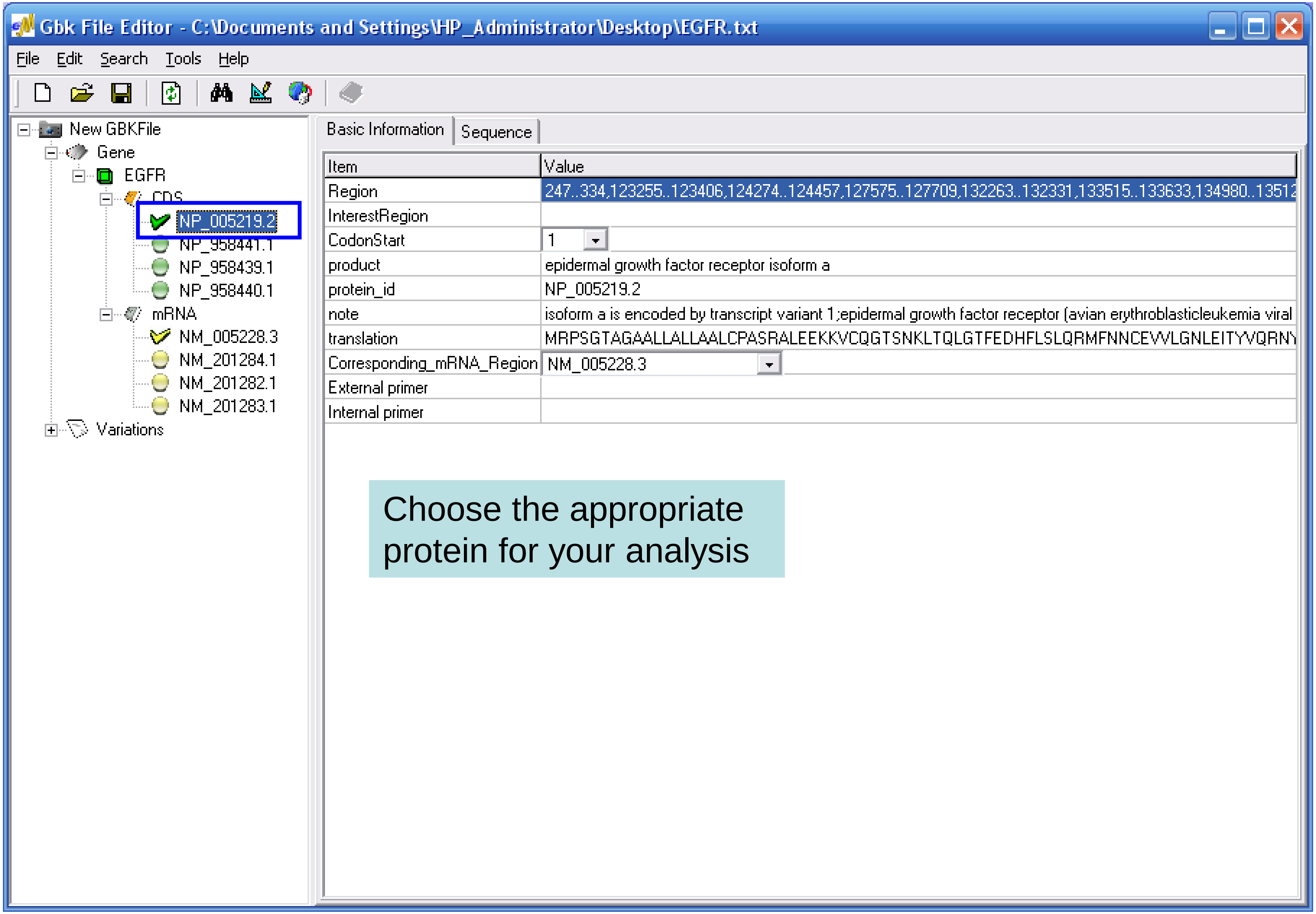
Choose to Save File.



Save in: temp

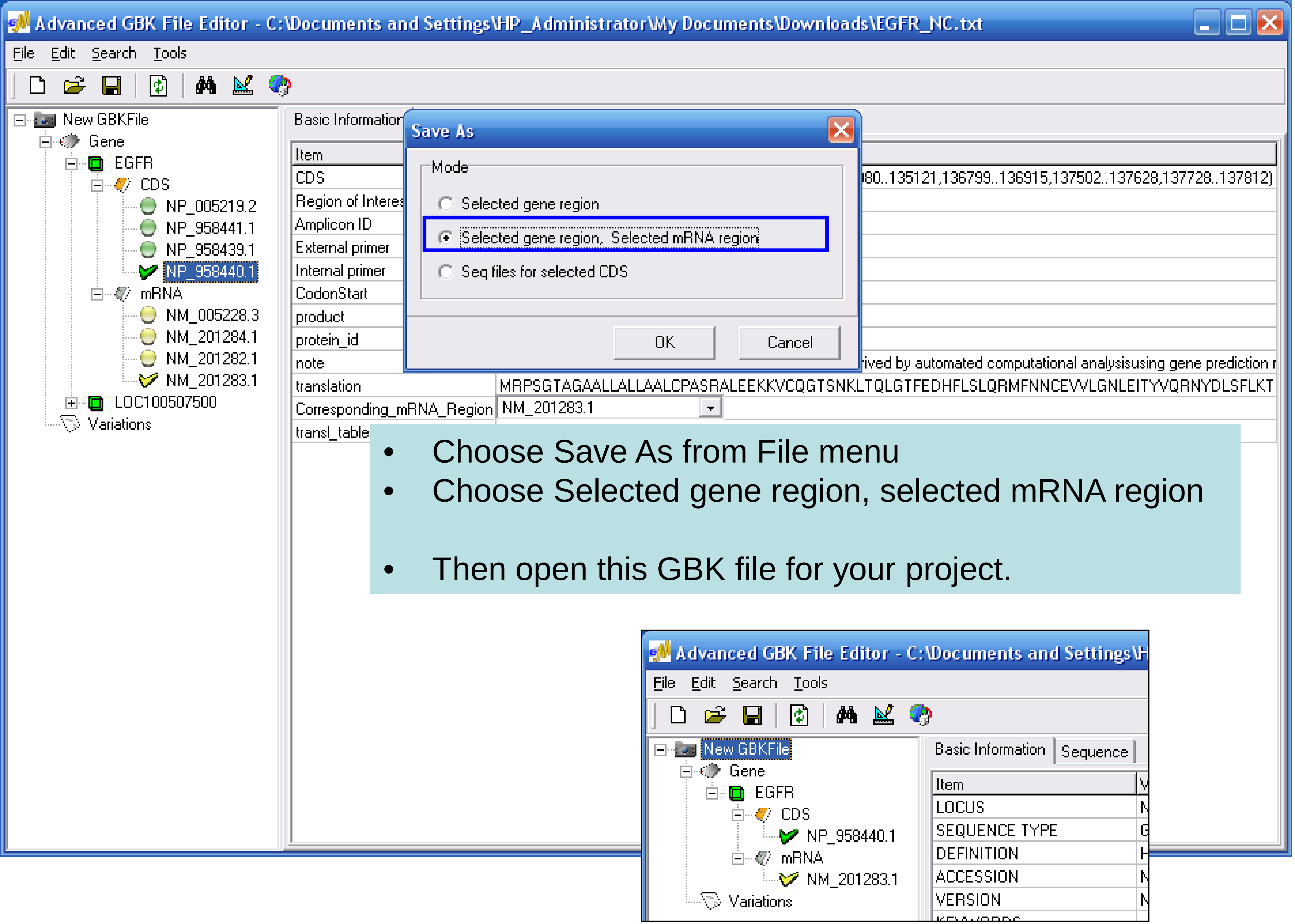
File name: sequences.gb

Save as type: .gb Document



Item	Value
Region	247..334,123255..123406,124274..124457,127575..127709,132263..132331,133515..133633,134980..1351
InterestRegion	
CodonStart	1
product	epidermal growth factor receptor isoform a
protein_id	NP_005219.2
note	isoform a is encoded by transcript variant 1;epidermal growth factor receptor (avian erythroblasticleukemia viral
translation	MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRN
Corresponding_mRNA_Region	NM_005228.3
External primer	
Internal primer	

Choose the appropriate protein for your analysis



Advanced GBK File Editor - C:\Documents and Settings\HP_Administrator\My Documents\Downloads\EGFR_NC.txt

Save As

Mode

○ Selected gene region

● Selected gene region, Selected mRNA region

○ Seq files for selected CDS

Item	
CDS	...80..135121,136799..136915,137502..137628,137728..137812)
Region of Interes...	
Amplicon ID	
External primer	
Internal primer	
CodonStart	
product	
protein_id	
note	...ived by automated computational analysisusing gene prediction r...
translation	MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEVVLGNLEITYVQRNYDLSFLKT
Corresponding_mRNA_Region	NM_201283.1
transl_table	

New GBKFile
- Gene
  - EGFR
    - CDS
      - NP_005219.2
      - NP_958441.1
      - NP_958439.1
      - ✓ NP_958440.1
    - mRNA
      - NM_005228.3
      - NM_201284.1
      - NM_201282.1
      - ✓ NM_201283.1
  - LOC100507500
- Variations

- Choose Save As from File menu
- Choose Selected gene region, selected mRNA region

- Then open this GBK file for your project.

Advanced GBK File Editor - C:\Documents and Settings\H...

New GBKFile
- Gene
  - EGFR
    - CDS
      - ✓ NP_958440.1
    - mRNA
      - ✓ NM_201283.1
- Variations

Item	V...
LOCUS	N...
SEQUENCE TYPE	G...
DEFINITION	H...
ACCESSION	N...
VERSION	N...