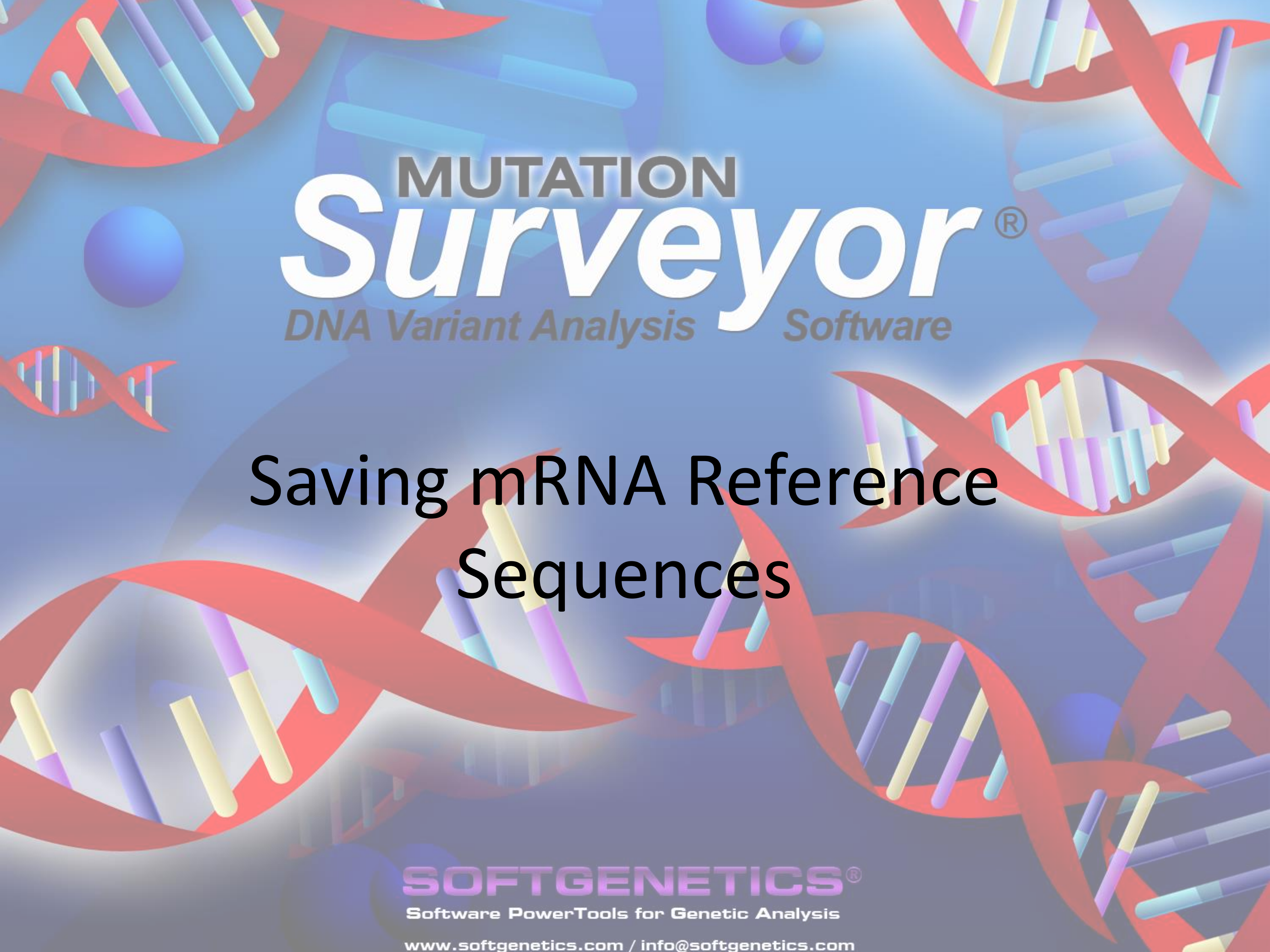


MUTATION Surveyor®

DNA Variant Analysis Software

Saving mRNA Reference Sequences

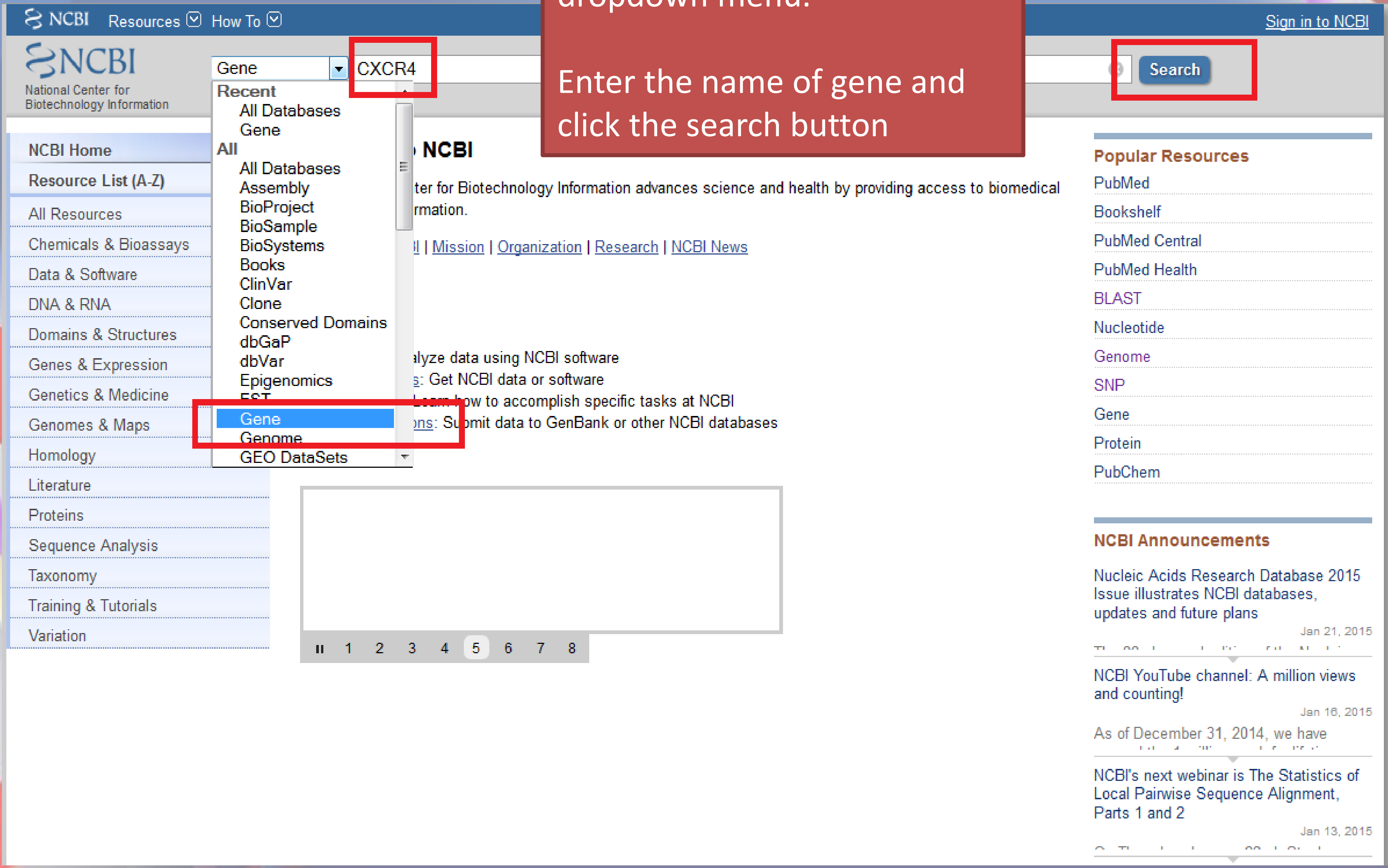
SOFTGENETICS®

Software PowerTools for Genetic Analysis

www.softgenetics.com / info@softgenetics.com

Select “Gene” from the dropdown menu.

Enter the name of gene and click the search button



NCBI

ter for Biotechnology Information advances science and health by providing access to biomedical rmation.

| Mission | Organization | Research | NCBI News

lyze data using NCBI software

: Get NCBI data or software

Learn how to accomplish specific tasks at NCBI

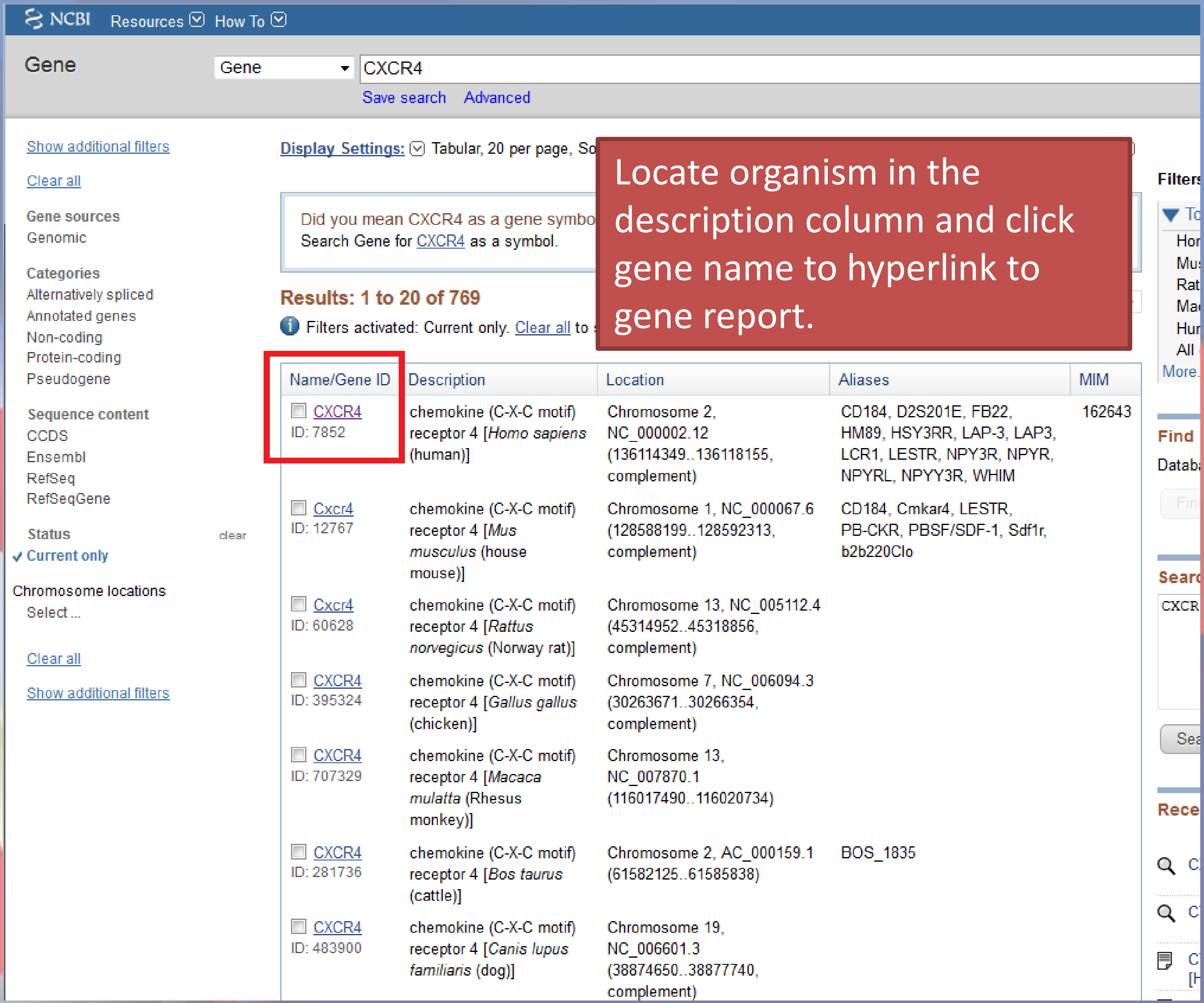
ns: Submit data to GenBank or other NCBI databases

Nucleic Acids Research Database 2015 Issue illustrates NCBI databases, updates and future plans
Jan 21, 2015

NCBI YouTube channel: A million views and counting!
Jan 16, 2015

As of December 31, 2014, we have

NCBI's next webinar is The Statistics of Local Pairwise Sequence Alignment, Parts 1 and 2
Jan 13, 2015



Display Settings: Tabular, 20 per page, So

Did you mean CXCR4 as a gene symbo
Search Gene for CXCR4 as a symbol.

Results: 1 to 20 of 769

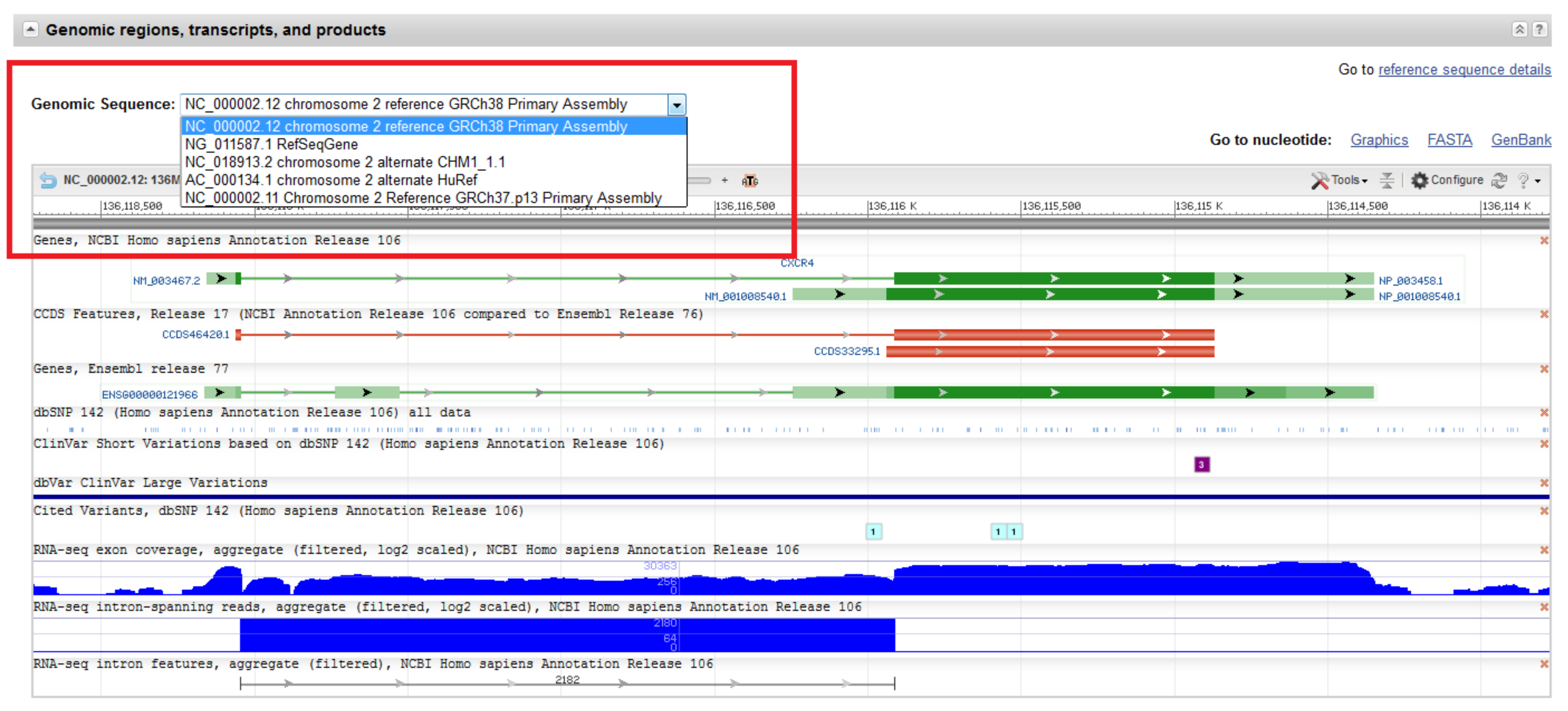
Filters activated: Current only. Clear all to

Name/Gene ID	Description	Location	Aliases	MIM
☐ CXCR4 ID: 7852	chemokine (C-X-C motif) receptor 4 [*Homo sapiens* (human)]	Chromosome 2, NC_000002.12 (136114349..136118155, complement)	CD184, D2S201E, FB22, HM89, HSY3RR, LAP-3, LAP3, LCR1, LESTR, NPY3R, NPYR, NPYRL, NPYY3R, WHIM	162643
☐ Cxcr4 ID: 12767	chemokine (C-X-C motif) receptor 4 [*Mus musculus* (house mouse)]	Chromosome 1, NC_000067.6 (128588199..128592313, complement)	CD184, Cmkar4, LESTR, PB-CKR, PBSF/SDF-1, Sdf1r, b2b220Clo	
☐ Cxcr4 ID: 60628	chemokine (C-X-C motif) receptor 4 [*Rattus norvegicus* (Norway rat)]	Chromosome 13, NC_005112.4 (45314952..45318856, complement)		
☐ CXCR4 ID: 395324	chemokine (C-X-C motif) receptor 4 [*Gallus gallus* (chicken)]	Chromosome 7, NC_006094.3 (30263671..30266354, complement)		
☐ CXCR4 ID: 707329	chemokine (C-X-C motif) receptor 4 [*Macaca mulatta* (Rhesus monkey)]	Chromosome 13, NC_007870.1 (116017490..116020734)		
☐ CXCR4 ID: 281736	chemokine (C-X-C motif) receptor 4 [*Bos taurus* (cattle)]	Chromosome 2, AC_000159.1 (61582125..61585838)	BOS_1835	
☐ CXCR4 ID: 483900	chemokine (C-X-C motif) receptor 4 [*Canis lupus familiaris* (dog)]	Chromosome 19, NC_006601.3 (38874650..38877740, complement)		

Locate organism in the description column and click gene name to hyperlink to gene report.

Select transcript from “Genomic regions, transcripts, and products” section.

NC accession files contain multiple transcripts/isoforms.



Genomic regions, transcripts, and products

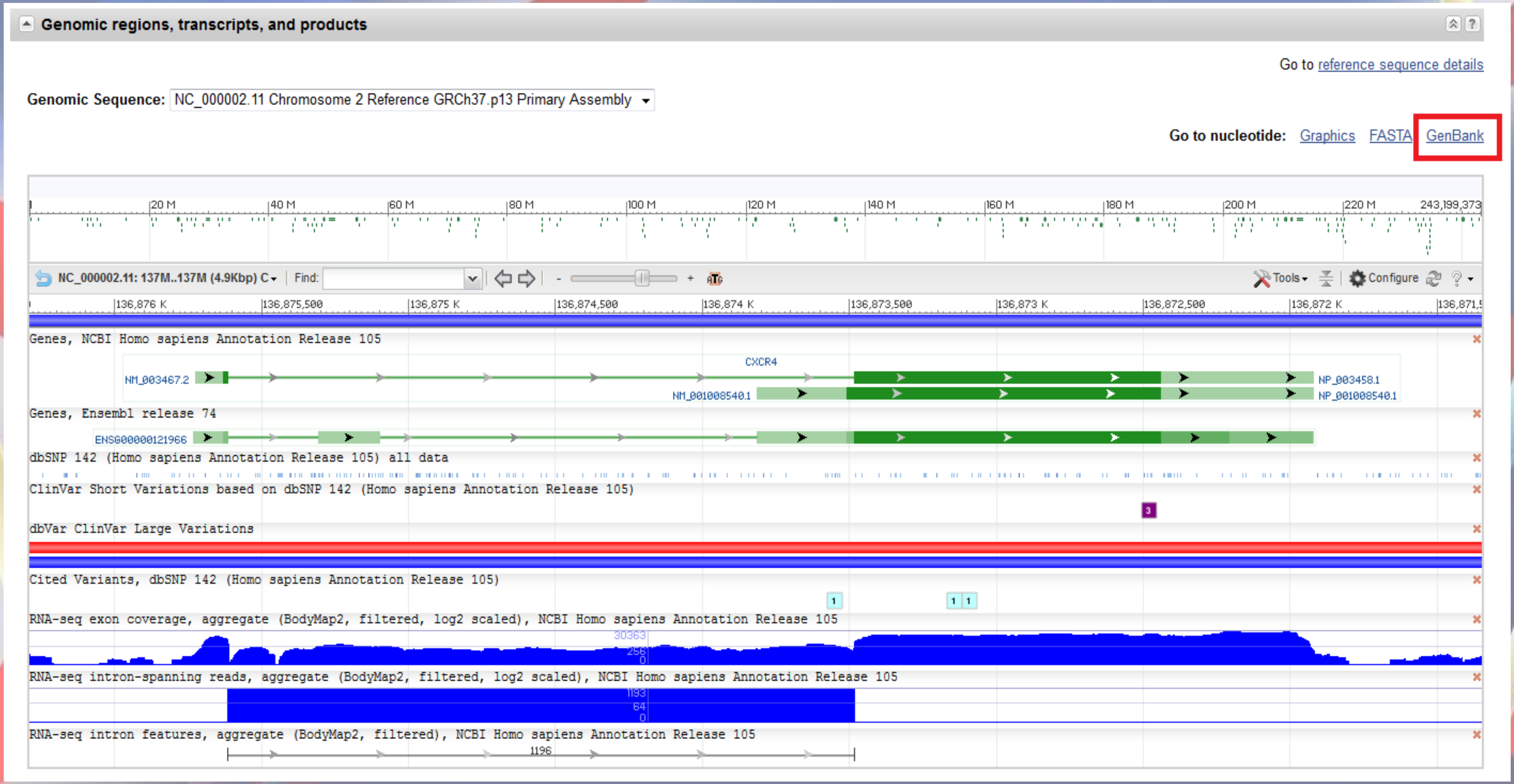
Go to reference sequence details

Genomic Sequence: NC_000002.12 chromosome 2 reference GRCh38 Primary Assembly

- NC_000002.12 chromosome 2 reference GRCh38 Primary Assembly
- NG_011587.1 RefSeqGene
- NC_018913.2 chromosome 2 alternate CHM1_1.1
- AC_000134.1 chromosome 2 alternate HuRef
- NC_000002.11 Chromosome 2 Reference GRCh37.p13 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

After transcript is selected, click the GenBank option to the right of the page.



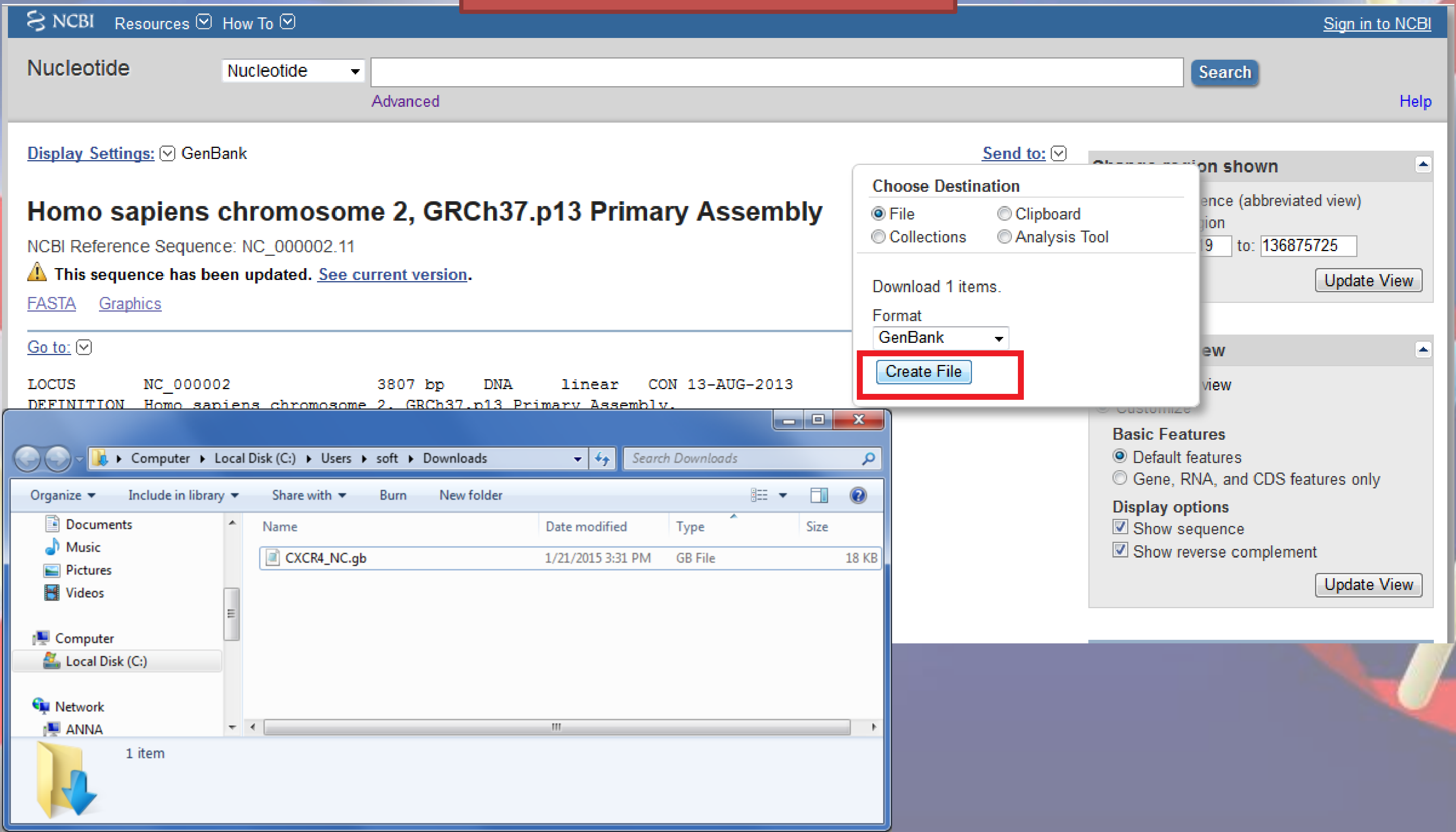
Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_000002.11 Chromosome 2 Reference GRCh37.p13 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

1. Navigate to “Send to”
2. Select File
3. Select GenBank format
4. Select Create file
5. Save file to computer



Nucleotide: Nucleotide

Display Settings: GenBank

Send to:

Choose Destination
● File ○ Clipboard
○ Collections ○ Analysis Tool

Download 1 items.

Format: GenBank

Create File

Homo sapiens chromosome 2, GRCh37.p13 Primary Assembly

NCBI Reference Sequence: NC_000002.11

⚠ **This sequence has been updated. See current version.**

FASTA Graphics

Go to:

```
LOCUS       NC_000002               3807 bp    DNA     linear   CON 13-AUG-2013
DEFINITION  Homo sapiens chromosome 2, GRCh37.p13 Primary Assembly.
```

to: 136875725

Update View

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

Display options
☑ Show sequence
☑ Show reverse complement

Update View

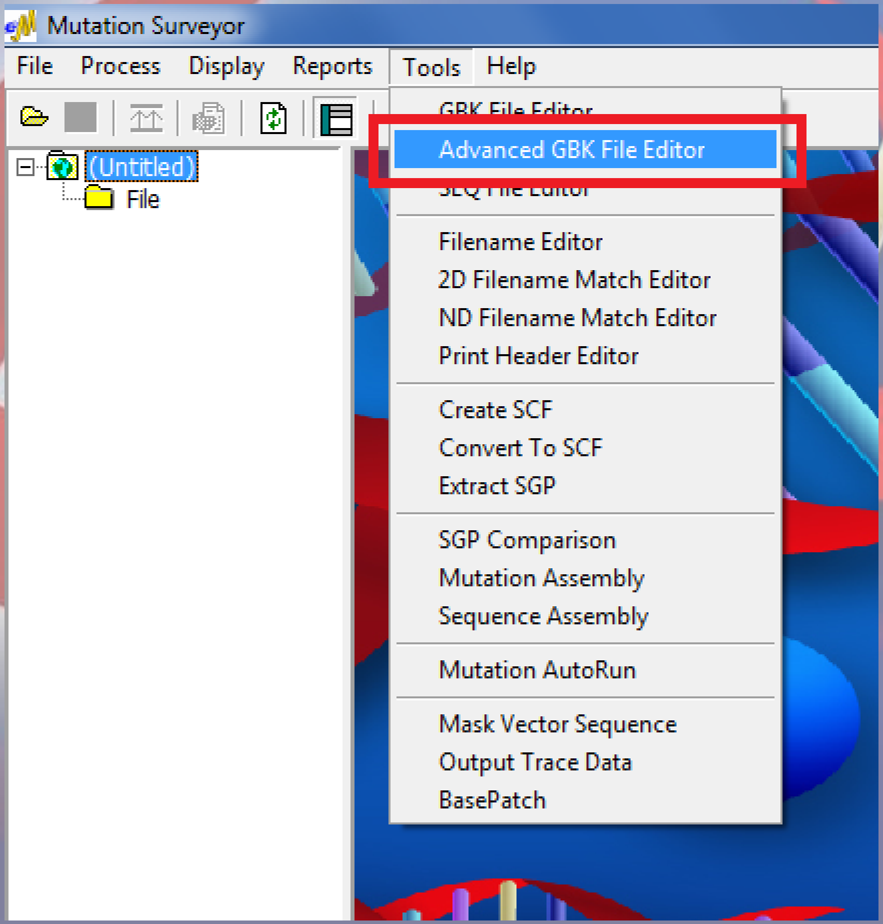
Computer ▸ Local Disk (C:) ▸ Users ▸ soft ▸ Downloads

Name	Date modified	Type	Size
CXCR4_NC.gb	1/21/2015 3:31 PM	GB File	18 KB

1 item

Launch Mutation Surveyor

Navigate to Tools > Advanced GBK File Editor

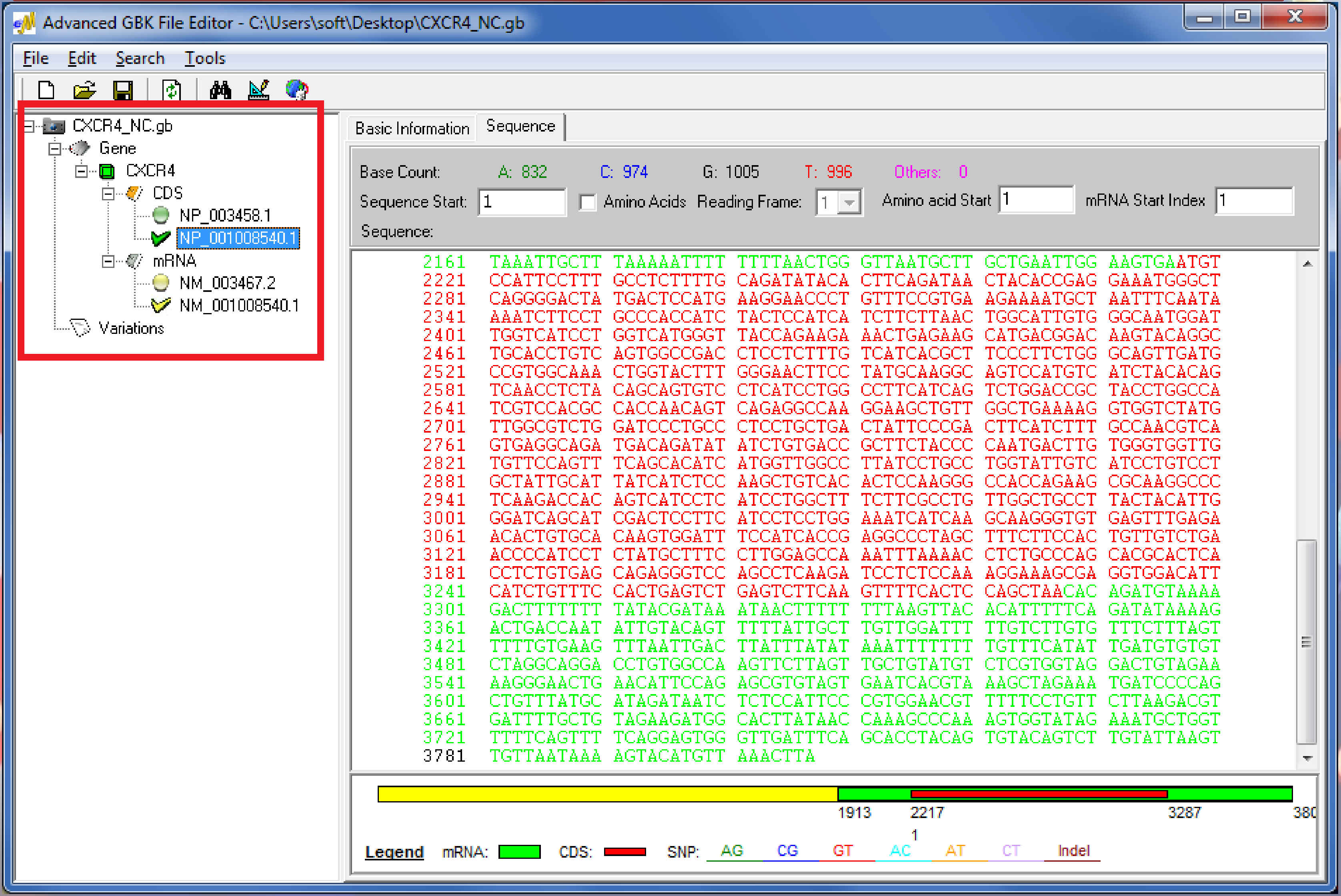


Open GenBank file

Isoform information will populate in left hand column

Click NP number to select isoform- a check mark will indicate that it is selected

Click the “Sequence” tab to review transcript



Base Count: A: 832 C: 974 G: 1005 T: 996 Others: 0

Sequence Start: 1 ☐ Amino Acids Reading Frame: 1 Amino acid Start 1 mRNA Start Index 1

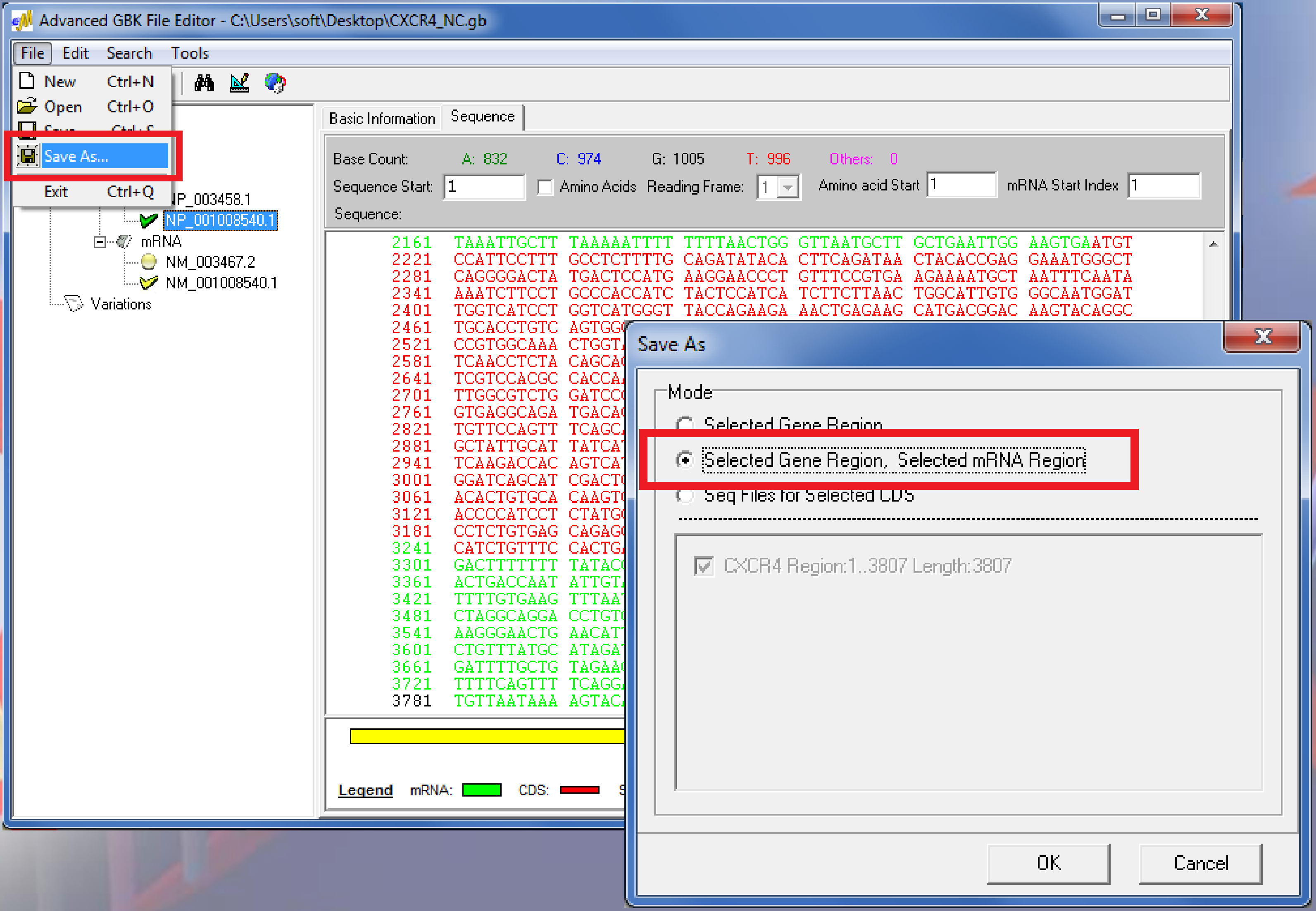
Sequence:

```
2161  TAAATTGCTT TAAAAATTTT TTTTAACTGG GTTAATGCTT GCTGAATTGG AAGTGAATGT
2221  CCATTCCTTT GCCTCTTTTG CAGATATACA CTTCAGATAA CTACACCGAG GAAATGGGCT
2281  CAGGGGACTA TGACTCCATG AAGGAACCCT GTTTCCGTGA AGAAAATGCT AATTTCAATA
2341  AAATCTTCCT GCCCACCATC TACTCCATCA TCTTCTTAAC TGGCATTGTG GGCAATGGAT
2401  TGGTCATCCT GGTCATGGGT TACCAGAAGA AACTGAGAAG CATGACGGAC AAGTACAGGC
2461  TGCACCTGTC AGTGGCCGAC CTCCTCTTTG TCATCACGCT TCCCTTCTGG GCAGTTGATG
2521  CCGTGGCAAA CTGGTACTTT GGGAACTTCC TATGCAAGGC AGTCCATGTC ATCTACACAG
2581  TCAACCTCTA CAGCAGTGTC CTCATCCTGG CCTTCATCAG TCTGGACCGC TACCTGGCCA
2641  TCGTCCACGC CACCAACAGT CAGAGGCCAA GGAAGCTGTT GGCTGAAAAG GTGGTCTATG
2701  TTGGCGTCTG GATCCCTGCC CTCCTGCTGA CTATTCCCGA CTTCATCTTT GCCAACGTCA
2761  GTGAGGCAGA TGACAGATAT ATCTGTGACC GCTTCTACCC CAATGACTTG TGGGTGGTTG
2821  TGTTCCAGTT TCAGCACATC ATGGTTGGCC TTATCCTGCC TGGTATTGTC ATCCTGTCCT
2881  GCTATTGCAT TATCATCTCC AAGCTGTCAC ACTCCAAGGG CCACCAGAAG CGCAAGGCCC
2941  TCAAGACCAC AGTCATCCTC ATCCTGGCTT TCTTCGCCTG TTGGCTGCCT TACTACATTG
3001  GGATCAGCAT CGACTCCTTC ATCCTCCTGG AAATCATCAA GCAAGGGTGT GAGTTTGAGA
3061  ACACTGTGCA CAAGTGGATT TCCATCACCG AGGCCCTAGC TTTCTTCCAC TGTTGTCTGA
3121  ACCCCATCCT CTATGCTTTC CTTGGAGCCA AATTTAAAAC CTCTGCCCAG CACGCACTCA
3181  CCTCTGTGAG CAGAGGGTCC AGCCTCAAGA TCCTCTCCAA AGGAAAGCGA GGTGGACATT
3241  CATCTGTTTC CACTGAGTCT GAGTCTTCAA GTTTTCACTC CAGCTAACAC AGATGTAAAA
3301  GACTTTTTTT TATACGATAA ATAACTTTTT TTTAAGTTAC ACATTTTTCA GATATAAAAG
3361  ACTGACCAAT ATTGTACAGT TTTTATTGCT TGTTGGATTT TTGTCTTGTG TTTCTTTAGT
3421  TTTTGTGAAG TTTAATTGAC TTATTTATAT AAATTTTTTT TGTTTCATAT TGATGTGTGT
3481  CTAGGCAGGA CCTGTGGCCA AGTTCTTAGT TGCTGTATGT CTCGTGGTAG GACTGTAGAA
3541  AAGGGAACTG AACATTCCAG AGCGTGTAGT GAATCACGTA AAGCTAGAAA TGATCCCCAG
3601  CTGTTTATGC ATAGATAATC TCTCCATTCC CGTGGAACGT TTTTCCTGTT CTTAAGACGT
3661  GATTTTGCTG TAGAAGATGG CACTTATAAC CAAAGCCCAA AGTGGTATAG AAATGCTGGT
3721  TTTTCAGTTT TCAGGAGTGG GTTGATTTCA GCACCTACAG TGTACAGTCT TGTATTAAGT
3781  TGTTAATAAA AGTACATGTT AAACTTA
```

Select File-Save As

Select the option to save “Selected Gene Region, Selected mRNA Region”

Rename file and select the directory to save the file



Base Count: A: 832 C: 974 G: 1005 T: 996 Others: 0

Sequence Start: 1 ☐ Amino Acids Reading Frame: 1 Amino acid Start 1 mRNA Start Index 1

Sequence:

```
2161 TAAATTGCTT TAAAAATTTT TTTTAACTGG GTTAATGCTT GCTGAATTGG AAGTGAATGT
2221 CCATTCCTTT GCCTCTTTTG CAGATATACA CTTCAGATAA CTACACCGAG GAAATGGGCT
2281 CAGGGGACTA TGACTCCATG AAGGAACCCT GTTTCCGTGA AGAAAATGCT AATTTCAATA
2341 AAATCTTCCT GCCCACCATC TACTCCATCA TCTTCTTAAC TGGCATTGTG GGCAATGGAT
2401 TGGTCATCCT GGTCATGGGT TACCAGAAGA AACTGAGAAG CATGACGGAC AAGTACAGGC
2461 TGCACCTGTC AGTGG
2521 CCGTGGCAAA CTGGT
2581 TCAACCTCTA CAGCA
2641 TCGTCCACGC CACCA
2701 TTGGCGTCTG GATCC
2761 GTGAGGCAGA TGACA
2821 TGTTCCAGTT TCAGC
2881 GCTATTGCAT TATCA
2941 TCAAGACCAC AGTCA
3001 GGATCAGCAT CGACT
3061 ACACTGTGCA CAAGT
3121 ACCCCATCCT CTATG
3181 CCTCTGTGAG CAGAG
3241 CATCTGTTTC CACTG
3301 GACTTTTTTT TATAC
3361 ACTGACCAAT ATTGT
3421 TTTTGTGAAG TTTAA
3481 CTAGGCAGGA CCTGT
3541 AAGGGAACTG AACAT
3601 CTGTTTATGC ATAGA
3661 GATTTTGCTG TAGAA
3721 TTTTCAGTTT TCAGG
3781 TGTTAATAAA AGTAC
```

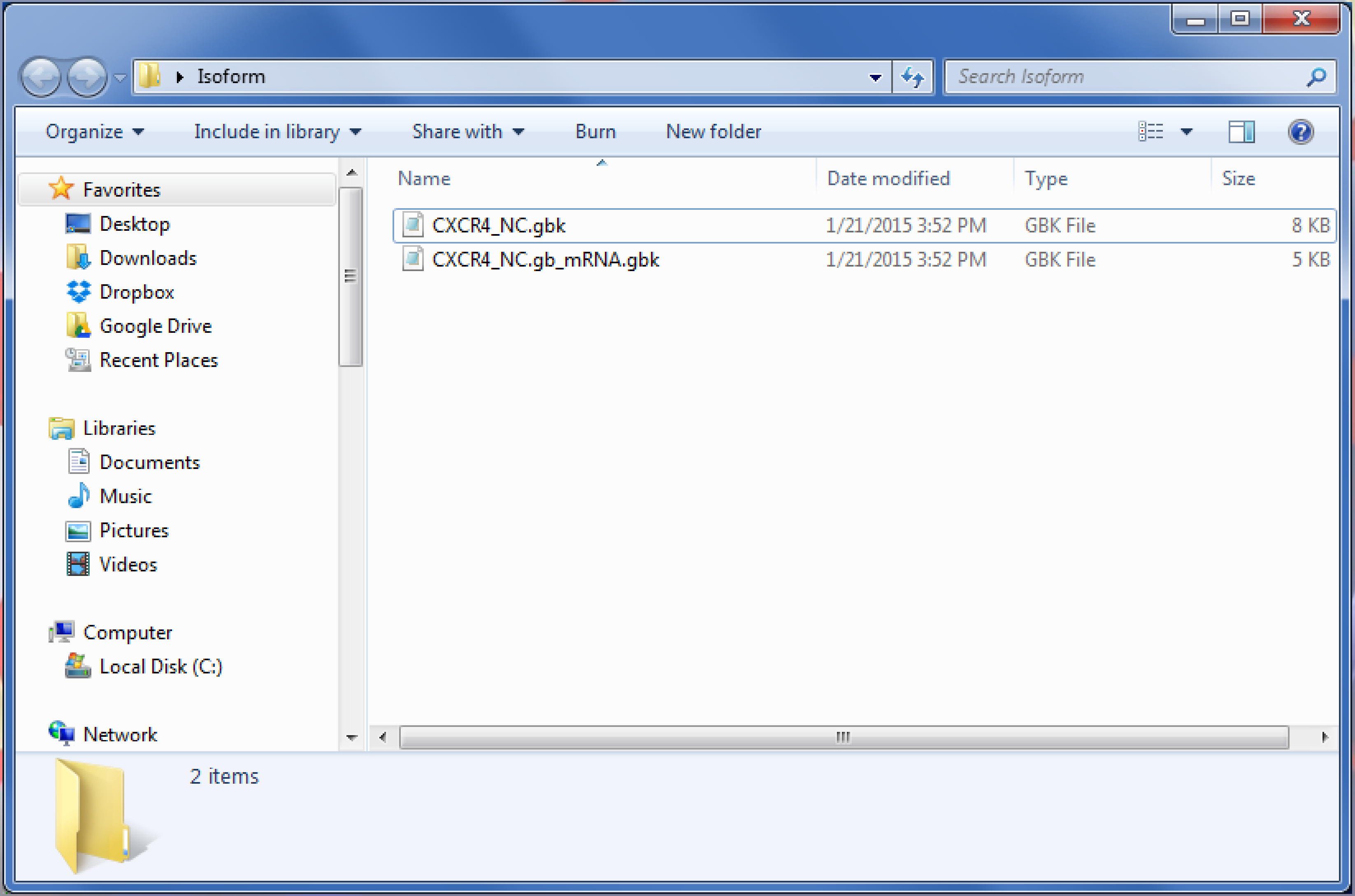
○ Selected Gene Region

● Selected Gene Region, Selected mRNA Region

○ Seq Files for Selected CDS

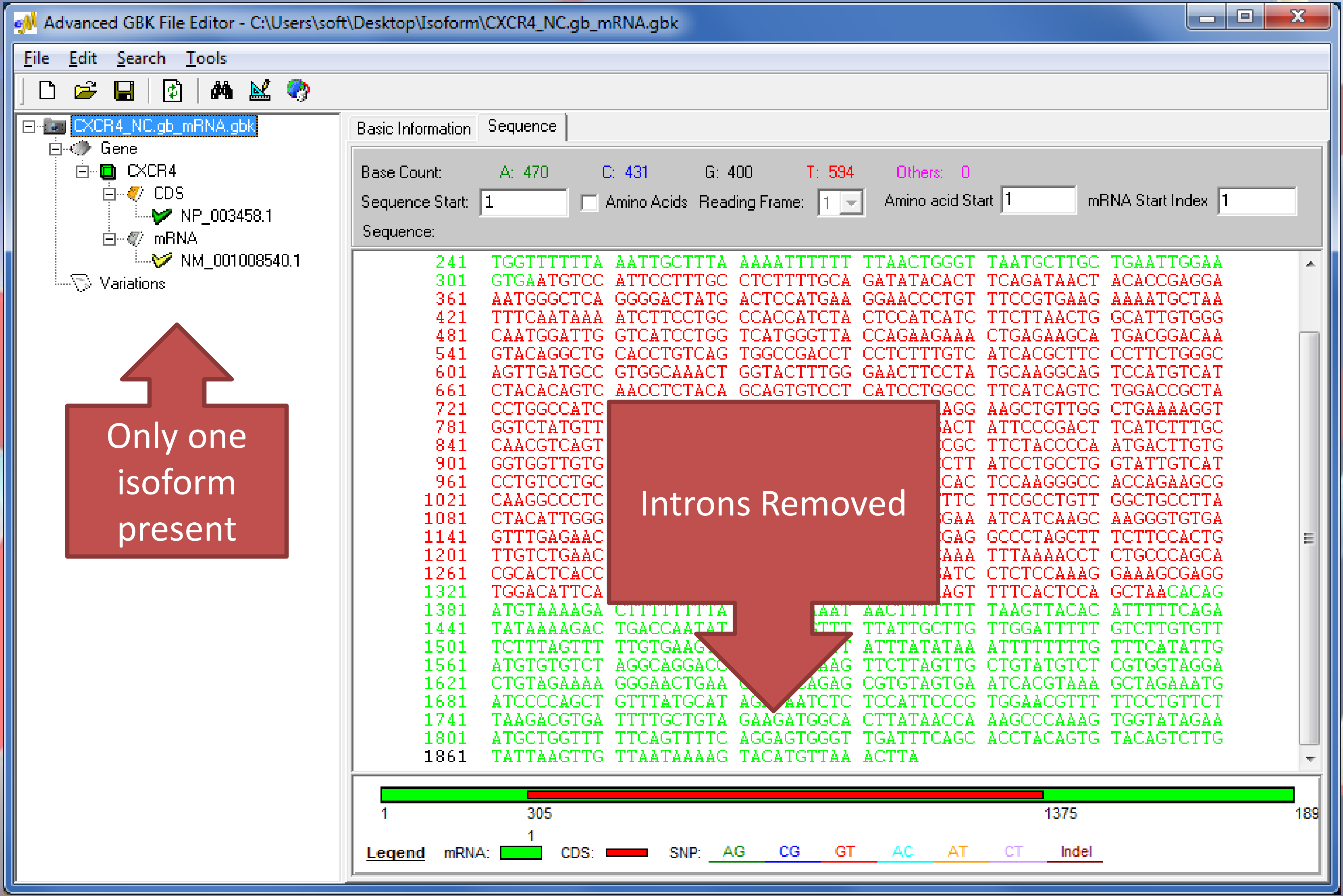
☑ CXCR4 Region:1..3807 Length:3807

Two files will save:
filename.gbk = Full GenBank file for selected isoform
*filename*_mRNA.gbk = mRNA file for selected isoform, introns removed



Name	Date modified	Type	Size
CXCR4_NC.gbk	1/21/2015 3:52 PM	GBK File	8 KB
CXCR4_NC.gb_mRNA.gbk	1/21/2015 3:52 PM	GBK File	5 KB

Confirm by loading mRNA file into Advanced GBK File Editor



Base Count: A: 470 C: 431 G: 400 T: 594 Others: 0

Sequence Start: 1 ☐ Amino Acids Reading Frame: 1 Amino acid Start 1 mRNA Start Index 1

Sequence:

```
 241  TGGTTTTTTA AATTGCTTTA AAAATTTTTT TTAACTGGGT TAATGCTTGC TGAATTGGAA
 301  GTGAATGTCC ATTCCTTTGC CTCTTTTGCA GATATACACT TCAGATAACT ACACCGAGGA
 361  AATGGGCTCA GGGGACTATG ACTCCATGAA GGAACCCTGT TTCCGTGAAG AAAATGCTAA
 421  TTTCAATAAA ATCTTCCTGC CCACCATCTA CTCCATCATC TTCTTAACTG GCATTGTGGG
 481  CAATGGATTG GTCATCCTGG TCATGGGTTA CCAGAAGAAA CTGAGAAGCA TGACGGACAA
 541  GTACAGGCTG CACCTGTCAG TGGCCGACCT CCTCTTTGTC ATCACGCTTC CCTTCTGGGC
 601  AGTTGATGCC GTGGCAAACT GGTACTTTGG GAACTTCCTA TGCAAGGCAG TCCATGTCAT
 661  CTACACAGTC AACCTCTACA GCAGTGTCCT CATCCTGGCC TTCATCAGTC TGGACCGCTA
 721  CCTGGCCATC                              AGG AAGCTGTTGG CTGAAAAGGT
 781  GGTCTATGTT                              ACT ATTCCCGACT TCATCTTTGC
 841  CAACGTCAGT                              CGC TTCTACCCCA ATGACTTGTG
 901  GGTGGTTGTG                              CTT ATCCTGCCTG GTATTGTCAT
 961  CCTGTCCTGC                              CAC TCCAAGGGCC ACCAGAAGCG
1021  CAAGGCCCTC                              TTC TTCGCCTGTT GGCTGCCTTA
1081  CTACATTGGG                              GAA ATCATCAAGC AAGGGTGTGA
1141  GTTTGAGAAC                              GAG GCCCTAGCTT TCTTCCACTG
1201  TTGTCTGAAC                              AAA TTTAAAACCT CTGCCCAGCA
1261  CGCACTCACC                              ATC CTCTCCAAAG GAAAGCGAGG
1321  TGGACATTCA                              AGT TTTCACTCCA GCTAACACAG
1381  ATGTAAAAGA CTTTTTTTTA              AAT AACTTTTTTT TAAGTTACAC ATTTTTCAGA
1441  TATAAAAGAC TGACCAATAT              TTT TTATTGCTTG TTGGATTTTT GTCTTGTGTT
1501  TCTTTAGTTT TTGTGAAG                  T ATTTATATAA ATTTTTTTTG TTTCATATTG
1561  ATGTGTGTCT AGGCAGGAC               AAG TTCTTAGTTG CTGTATGTCT CGTGGTAGGA
1621  CTGTAGAAAA GGGAACTGAA C           AGAG CGTGTAGTGA ATCACGTAAA GCTAGAAATG
1681  ATCCCCAGCT GTTTATGCAT AG        AATCTC TCCATTCCCG TGGAACGTTT TTCCTGTTCT
1741  TAAGACGTGA TTTTGCTGTA GAAGATGGCA CTTATAACCA AAGCCCAAAG TGGTATAGAA
1801  ATGCTGGTTT TTCAGTTTTC AGGAGTGGGT TGATTTCAGC ACCTACAGTG TACAGTCTTG
1861  TATTAAGTTG TTAATAAAAG TACATGTTAA ACTTA
```

Only one isoform present

Introns Removed