

Worked example: Querying for APP gene at NCBI

1. Go to <http://www.ncbi.nlm.nih.gov/>
2. Type in APP into the query box and press the Go button
3. Query results across all NCBI Entrez databases are returned
4. Look over databases and describe most popular and notice that there are >1000 results for Gene
5. Click on result number next to Gene
6. Our record of interest (APP for human) is at the top because we're lucky
7. Point out gene name and organism in summary hit information
8. Demonstrate the use of the Preview/Index tab for field selection
9. Click on Preview/Index tab
10. At the bottom of the page, there is a dropdown menu with field names
11. Explain what field names in databases are
12. In this dropdown menu, select Gene Name, type in APP and click the Index button
13. Explain what index terms are and that the numbers in parenthesis are the numbers of record that contain that term
14. Select app in the list of indexed terms and click the AND button
15. Erase "APP AND" in the query box
16. Click Preview button, and notice that a new recent query appears, along with the number of results
17. Click on the linked number of results and you'll go to the page with results
18. Notice that results are now ordered differently and human APP is not on the top
19. Exercise: use Preview/Index to search the Organism field for Homo sapiens
20. Click on APP to view full record

Worked example: APP Entrez Gene record

1. Go to <http://www.ncbi.nlm.nih.gov/>
2. Select Gene from the Search dropdown menu on the upper left
3. Type in the following query: `app[Gene Name] AND human[Organism]`
4. Notice we used human instead of homo sapiens this time, and one result is retrieved
5. Click on Details tab
6. The Query Translation box contains the actual query that was performed.
7. Notice human has been converted by the query parser into homo sapiens, since they are synonyms, but the latter is the actual indexed term
8. Hit the browser's go back button
9. Click on the one result named APP and let's try to answer all the questions we can....
10. The APP gene is located on the negative strand at cytogenetic band 21q21.2-21.3, between genes GABPA and CYR1
11. Its genomic sequence is given in genomic contig reference sequence NC_000021.7, nucleotides 26,174,732-26,465,003
12. Explain RefSeq accession number conventions--ie, NC_=chromosome, NM_=mRNA, NP_=protein
13. Explain how versions are indicated--ie, with the # suffix--this record has seven versions
14. Explain the difference between accession and gi number
15. APP has three annotated alternately transcribed transcripts, and three protein isoforms
16. Any of these sequences may be retrieved from the linked RefSeq accession numbers in the "genomic regions, transcripts and products" section, via a dropdown format menu, via the "Reference sequence details" section near the bottom of the record
17. Let's try this for the genomic sequence records, and one each of the mRNA and protein records....
18. Click on the linked NC_ accession, and select GenBank format
19. Notice the record contains (luckily) just the genomic sequence associated with APP gene, and not that of the whole chromosome assembly, and that we are viewing the (correct) reverse complemented strand
20. By scrolling down in this record, we can see the annotated sequence features include STS (sequence tag sites), gene, mRNA and CDS (coding DNA sequence) features, and these always include nucleotide position references relative to the current record
21. Hit the go back button on the browser to go back to the APP gene record
22. Click on one of the linked mRNA accessions, and select GenBank format--eg, NM_000484.2 (what version of this accession will we be looking at?)
23. Notice all of the citations to journal articles on APP sequence-based and functional studies, as well as GeneRIFs (gene references into function).
24. Notice the comment just before the features section that gives some curatorial information about this RefSeq record
25. Scroll through the features and notice, once again, that nucleotide positions are given for all and are registered to the current record
26. Notice that mRNA-specific features are given, such as canonical polyA signal and site
27. Hit the go back button on the browser to go back to the APP gene record, and inspect

- the GenPept record for the isoform a of APP
28. Notice the extensive post-translational modifications (including cleavages, phosphorylations and acetylations) that are possible
 29. As an exercise, have the participants identify the annotated Beta-amyloid peptide fragment, and identify the smallest size of this fragment
 30. Explain what >705 means in this context
 31. Hit the go back button on the browser to go back to the APP gene record
 32. Scroll down through the rest of the APP gene record and explain each of the remaining sections.
 33. Briefly review the GeneRIFs, interactions (could be more useful in other contexts) and markers sections (again, perhaps more useful under other circumstances)
 34. Depending on the level of gene annotation, it is possible to assess disease and normal function in four ways: GeneRIFs, GeneOntology, OMIM and Conserved Domains.
 35. APP has extensive annotations in all four of these and they can be reviewed by direct viewing within the APP gene record (GeneRIFs and GO), or by following links to OMIM and Conserved Domain records.
 36. Transcript expression profiles can be examined by clicking on the GEO Profiles link on the right sidebar
 37. By clicking that link, you will usually retrieve many thousands of profiles over many gene expression (usually microarrays, but also SAGE) datasets, and APP is no exception
 38. Show participants how to restrict results to conditions of interest....
 39. Click on Preview/Index, and note the search number that corresponds to the link from the APP gene record
 40. In the query box, type that search number plus “AND Alzheimer’s disease” and press the Go button
 41. A handful of records are retrieved.
 42. Click on one of the profile bar charts to understand the details of the conditions under which the profile was generated and if there is any interesting behavior
 43. In GDS810 record, it doesn’t appear there is any association between APP expression and AD stage
 44. Click the go back button in the browser to get back to the GEO Profiles query results page
 45. If an interesting profile was found, we could find similarly-behaving genes by clicking the Profile Neighbors link
 46. Click on one of those links to see what happens and confirm that the profiles look similar
 47. Click on the browser’s go back button a few times until you get back to the APP Gene record
 48. Click on the KEGG link on the right sidebar
 49. Explain that KEGG is a pathway database that initially focused on metabolic pathways, but which has now started to describe disease pathways, such as AD
 50. Click on the Alzheimer’s disease record with id hsa05010
 51. Spend a little time tracing out what we have discovered up to now about APP

Worked example: APP in the UCSC Genome Browser

1. Go to <http://genome.ucsc.edu>
2. Click on the Genome Browser link on the left sidebar
3. Verify that genome: Human and the latest assembly (currently, Mar. 2006) are selected
4. Type in app into the query box and click the Submit button
5. In the list of results, click on the link for APP, isoform a (but isoforms b and c are just as good since all the positions are the same)
6. Orient the participants to the display...
7. at the top are some browsing and zooming controls
8. on the next line are the position jumping controls
9. next is the chromosome overview
10. then we have the browser window with data tracks displayed as position-registered, layered horizontal lines
11. and at the bottom are track display controls.
12. Let's make sure we're all looking at the same tracks...
13. Scroll down to the track display controls and make sure everything is set to "hide" by clicking on the hide all button
14. Set the following tracks to "dense": Base Position, Known Genes, RefSeq Genes, Ensembl Genes, Human mRNAs, Spliced ESTs, Human ESTs, UniGene, TFBS Conserved, Conservation, Most Conserved, Repeat Masker, Simple Repeats
15. Click on the refresh button
16. Point out the different marks on the gene tracks: exons: large ticks, UTR: small ticks, line: intron, arrow heads: transcription direction
17. Click on the Known Genes, RefSeq Genes and Ensembl Genes to expand them to packed display
18. Known genes are UCSC pick of "best" gene models
19. Ensembl gene predictions are analogous to RefSeq, but they are made by EMBL (more about that later)
20. Human mRNAs, Spliced ESTs and Human ESTs are derived from sequencing records.
21. Notice that the UniGene clusters for this sequence are fractured and don't combine exons. Possible explanations include too few mRNA and EST sequences to form robust clusters, bad EST coverage of exons won't allow good sequence overlaps, or misalignment of cluster sequences with the genome.
22. To check out the number of EST sequences, click on Human ESTs. It appears that there is very good EST coverage of this gene, but that they are very fragmented with numerous 3' ends. This could confuse the UniGene clustering algorithm since it weights 3' ends more highly than 5' ends.
23. To check out how big the exons are, click on one of the RefSeq sequences, and then click on the linked Entrez Gene id.
24. In the new browser window, change the display to Gene Table and scan through the exon lengths. No problem here, since the exons seem to be about average size (about 100nt long).

25. To check for misalignment of cluster sequences with the genome, take note of one of the RefSeq sequence ids (say NM_000484), click on one of the UniGene ids, and then the Human UniGene id.
26. In the UniGene record browser in the new browser window, search for the RefSeq sequence. We find that Hs.651215 is large cluster that defines APP.
27. Close that browser window and search for Hs.651215 in the UCSC genome browser. It is not found and therefore the problem must be misalignment.
28. Press the browser's back button until you're showing the APP gene again
29. Click the hide all button
30. Set the following tracks to "dense": Base Position, Known Genes, RefSeq Genes, Ensembl Genes, TFBS Conserved, Conservation, Most Conserved, SNPs, Repeat Masker, Simple Repeats
31. Most tracks have the following display types: hide, dense, squish, pack, full
32. The TFBS Conserved track is a good one to try out the different track display types-- have participants try each type on their own--remember, nothing will happen unless you click the refresh button
33. Make all the tracks dense again, we're going to look upstream for possible regulatory regions.
34. Click on the TFBS Conserved and Conservation tracks to flip them into pack display
35. The Conservation track shows a histogram with degree of conservation after comparison to other vertebrate genomes.
36. The TFBS Conserved track shows high scoring transcription factor binding sites in regions of high vertebrate conservation. These are computationally derived.
37. Click on the double arrow right to pan the browser view to the right.
38. Identify the TFBS Conserved site upstream of the putative transcription start site for APP.
39. This is V\$MEF2_01. Click on it. This is the record for myogenic enhancing factor 2 family of transcription factors.
40. To see if this is known to be related to APP, open a new browser window, go to NCBI and search all databases for MEF2 AND APP.
41. You get two hits in PubMed and two hits in Gene. Click on the PubMed hits.
42. Click on the Burton et al paper to read the abstract and see that APP seems to mediate phosphorylation of MEF2, which regulates neuronal survival. There was differential activity of mutated APP in familial AD. There is also a GeneRIF record in the APP gene record which references this paper.
43. Hypothesis: perhaps APP activates MEF2 and MEF2 negatively regulates APP transcription?

Worked example: APP in EBI/Ensembl resources

1. Go to <http://www.ebi.ac.uk/>
2. Enter app in the EB-eye search box with All Databases selected in the drop down menu and click the Go button
3. Expand the Genomes databases, and select the Ensembl link
4. Scan down the retrieved records and select the gene record for human APP
5. Click on the [Peptide info] link for the first transcript
6. Note extensive graphical display of protein features
7. In the Protein Sequence section select Exons and SNPs in the feature dropdown, and Yes in the Number residues dropdown, and click the Refresh button
8. How many coding sequence variations are listed?
9. Mouse over each to show alternatives.
10. Scroll down to the bottom of the page to see the tabulated details for features and variations
11. Click on the linked SNP ID for the 1nt deletion, frameshift variation
12. Note that there are links to all transcripts containing this SNP
13. Scroll down to the SNP Context and note all the SNPs in this area (our SNP is centered, pink and boxed)
14. Note that all other SNPs but one are intronic SNPs (blue), but there is one splice site SNP in orange.
15. Scroll back up to the top and click on the SNP ID link, which will take us back to the source record, dbSNP at NCBI
16. Scroll down to the GeneView section and note that this SNP could not be mapped to the genome for some reason, but we don't really know why
17. Click the browser's back button a few times to get back to the APP gene record
18. Scroll down to the Orthologue Prediction section and note the list of gene records in other species that have been determined to be potential functional orthologs
19. Scroll down to the details for the first transcript.
20. Note the rich list of identifiers that Ensembl has mapped to these transcripts--these include microarray platform ids (such as for Affymetrix) and PDB structures.
21. Scroll back up to the top of the gene record, and click on the [Exon info] link for the first transcript.
22. Scroll down to the tabulated Exon Information section to view details, included exact chromosome start and stops for all exons, introns and UTR, segment length, as well as full nt sequence and codon phases for the former
23. Note that there are some function buttons to display the table in different ways