

SeattleSNPs Variation Discovery Workshop
Sept 12-13, 2005
Mark Rieder, PhD

Interactive Tutorial: SNP Database Resources

The tutorial is designed to take you through the steps necessary to access SNP data from the primary database resources:

1. Entrez SNP/dbSNP
2. UCSC Genome Browser
3. Perlegen Genotype Browser
4. HapMap Genome Browser
5. SeattleSNPs Variation Discovery Resource
6. SeattleSNPs Genome Variation Server (GVS)

Note: Answers to questions from this tutorial are included at the end of this document

As a launching point, we will begin our searching at the Entrez cross-database browser. This can be accessed on the NCBI home page (<http://www.ncbi.nlm.nih.gov/>). For these exercises we will be accessing data for the gene: chemokine-like factor (HUGO name: CKLF).

For a cross-database search:

1. Enter the gene symbol (CKLF) into the empty box next to the 'Search All Databases', type CKLF into the empty box and click on the GO button, or simply hit the return key on your keyboard.
Which NCBI database gives the most number of results?
What is the database? Hint: Mouse over on the '?' next to the icon and click for a popup explanation of this database.
2. On the left column note the results returned for the 'SNP' and 'Gene' databases.
3. How many results were returned for the 'SNP' and 'Gene' databases?
4. Why did the 'Gene' database return more than one result?

Entrez Gene

5. From the cross database search, click on the 'Gene' database icon.
6. Click on the result that corresponds to the 'homo sapiens' CKLF gene.
7. CKLF maps to which chromosome?
8. What are the genes 5' and 3' of CKLF? (Hint: look at the genomic context).
9. On the far right of the page next to the CKLF gene name and description, note the word 'Links'(see Figure 1 below).
10. Scroll down this list and select 'SNP: Geneview.'

NCBI Entrez Gene

My NCBI
Welcome mrieder. [Sign Out]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

Search Gene for [] Go Clear [x] current records only

Limits Preview/Index History Clipboard Details

Display Full Report Show 5 Send to

All: 1 Gene Genotype: 1 Genes Genomes: 1

1: **CKLF chemokine-like factor** [*Homo sapiens*]
GeneID: 51192 Locus tag: [HGNC:13253](#) updated 02-Aug-2005

Summary

Official Symbol: CKLF and **Name:** chemokine-like factor **provided by** [HUGO Gene Nomenclature Committee](#)
Gene type: protein coding
Gene name: CKLF
Gene description: chemokine-like factor
RefSeq status: Reviewed
Organism: [Homo sapiens](#)
Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo*
Gene aliases: C32; CKLF1; CKLF2; CKLF3; CKLF4; UCK-1; HSPC224
Summary: The product of this gene is a cytokine. Cytokines are small proteins that have an essential role in the immune and inflammatory responses. This gene is one of several chemokine-like factor genes located in a cluster on chromosome 16. The protein encoded by this gene is a potent chemoattractant for neutrophils, monocytes and lymphocytes. It also can stimulate the proliferation of skeletal muscle cells. This protein may play important roles in inflammation and in the regeneration of skeletal muscle. Alternatively spliced transcript variants encoding different isoforms have been identified.

Transcripts and products

[RefSeq below](#)

NC_000016

[65143973] [65157472]

5' 3'

NM_016951
 NM_181641
 NM_191849
 NM_016326
 NP_058647 isoform a
 NP_057592 isoform d
 NP_057591 isoform b
 NP_057411 isoform c

Entrez Gene Home
 Table Of Contents
 Summary
 Transcripts and products
 Genomic context
 Bibliography
 General gene information
 General protein information
 Reference Sequences
 Related Sequences
 Additional Links
 Links
 Conserved Domains
 GEO Profiles
 HomoloGene
 Map Viewer
 Nucleotide
 Full text in PMC
 Protein
 PubMed
 PubMed (GeneRIF)
 SNP
 SNP: Genotype
 SNP: GeneView
 Taxonomy
 UniSTS
 AceView
 Ensembl
 Evidence Viewer
 HGNC
 KEGG
 MGC
 ModelMaker
 UCSC
 UniGene
 LinkOut

dbSNP

1. The initial dbSNP Geneview only shows SNPs that are located in the coding region of the gene (cSNPs).
2. How many cSNPs are found in dbSNP for CKLF? How many are validated? Under the 'Gene Model' heading, use the button selectors to view all SNPs in the 'gene region' (select that button) and then select the 'view rs' button.
3. After selecting this, the page will update and show all SNPs in this gene.
4. How many SNPs are found in dbSNP for CKLF?
5. Note: this number will appear just above the SNP map picture of the gene.
6. How many SNPs shown have been validated by the HapMap project (i.e., SNPs with an 'H' symbol in the validation column)?
7. How many SNPs have frequency data (i.e., a heterozygosity value) associated with them? Hint: count the number without this data and subtract from total.
8. Click on the rs# SNP link that is validated by the HapMap (rs3785087).
9. How many submitters have recorded a discovery of this SNP?
10. Click on the ss# (ss28446109) next to the 'PGA-UW-FHCRC|CKLF-005513' SNP submission.
11. On this page, scroll down and find the frequency data for this SNP in each of the two populations studied by this submitter (PGA-EUROPEAN-PANEL, PGA-AFRICAN-PANEL). What are the allele frequencies of the C and G alleles in each of these populations?

12. Using the 'BACK' button in your browser, return to the Entrez Gene page for CKLF.

Entrez SNP

1. Starting from the Entrez Gene page again, use the 'Links' menu on the right side to view the linkout choices and select the 'SNP' option.
2. This will automatically query the Entrez SNP database for all SNPs in dbSNP for the CKLF gene for species you are viewing (i.e., 'homo sapiens').
3. How many SNPs are returned?
4. Below the search box and tabbed menu choices (i.e., 'Limits', 'Preview/Index', etc.), set the 'Display' feature menu to show this list as a 'FASTA'. The page should automatically update when you make your selection.
5. In the 'Send To' drop down menu, select the 'Text' option. The page should update the results in plain text format. This selection can be directly copied to a file on your computer.
6. Use the 'BACK' button on your browser. Alternatively this data can be "Sent To" a 'File' directly, that is saved on your computer.
7. Select the 'Limits' tab below the main search box. In the main search box type the gene name 'CKLF'.
8. Select the following search limits from the selections on this page:
 - a. Organism: Homo sapiens
 - b. Validation: 2hit-2allele
9. After making these selections, use the 'Go' button next to the main search box to get the result.
10. How many results are returned for validated 2hit-2allele SNPs in this gene?
11. Experiment with saving these in different formats using both the 'Send To' → 'Text' option and 'Send To' → 'File' option.
12. Go to the 'Limits' option again and select the following search items:
 - a. Organism: Homo sapiens
 - b. Has Genotype: True
13. How many results are returned for SNPs with genotype data?
14. In the 'Display' drop down menu select the 'Genotype'. Genotype data for each rs# SNP will be displayed.
15. Make sure that the checkbox in the 'Limits' tab is unchecked. Finally, to demonstrate the ability of using search term fields directly in the main search box, type the following:

CKLF[gene] AND "PGA-UW-FHCRC"[handle]

How many total entries are in dbSNP for this gene and submitted by the SeattleSNPs PGA project (our handle is PGA-UW-FHCRC)?

Perlegen Genotype Browser:

The Perlegen Genotype Browser can be accessed at:

<http://genome.perlegen.com/browser/>

1. On the main page for the Perlegen Browser enter 'CKLF' in the 'Enter Gene Name' search box. Click on the 'Go' button to access the browser.
2. How many SNPs are shown in this gene view?
3. Click on the first SNP on the left-hand side (afd1639142).
4. What is the reference allele frequency for this SNP in each of the populations studied?
5. Go to the 'LD Analysis' in the Chinese population.
6. Click on the LD bin.
7. What is the length of that LD bin?
8. How many SNPs are present in this bin?
9. Return to previous page and select the 'Haplotype Analysis' for each population in the 'Tracks' lower control panel.
10. What is the length of the longest haplotype for the European American population? And for the African American haplotype?

HapMap Browser:

The HapMap Genome Browser is linked directly from the main page at hapmap.org or can be accessed directly at: <http://hapmap.org/cgi-perl/gbrowse/gbrowse/hapmap>.

1. In the main search box enter the gene name 'CKLF'.
2. A list of result from many genes will be displayed – selected the one specific to CKLF.
3. The browser page with tracks will be presented.
4. How many HapMap SNPs in this gene?
5. Note the display of frequency data for each population using the pie graphs for each SNP. Click on the first HapMap SNP in this gene (rs3785087).
6. Note the allele frequencies for each population. Select the 'retrieve genotypes' link on the far right column. Genotype data for this SNP and population will be displayed.
7. Use the 'BACK' button to get to the main gene view.
8. In the lower section (highlighted in yellow), in the 'Dumps, Searches, and other Operations' select 'Dump SNP Genotype Data' from the drop down menu. Click on the 'Go' button.
9. Genotype data for all HapMap SNPs in this view will appear.

UCSC Genome Browser:

The UCSC browser can be accessed directly at: <http://www.genome.ucsc.edu>/or from the 'Links' menu at Entrez Gene (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene).

1. Using the 'Genome Browser' link at the top of the left column go to the Genome Browser Gateway. From the 'Genome' drop down menu select 'Human' and the 'assembly' from 'May 2004' (aka hg17 or Build 35 of the human genome).
2. Enter the gene name CKLF in the browser coordinate search box.
3. From the 'RefSeq Genes' results select the first entry:
 - a. CKLF at chr16:65,143,972-65,157,471
4. Below the gene view window, select the 'default tracks' button.
5. Near the bottom of the gene view window a track for 'Simple Nucleotide Polymorphisms (SNPs)' will be present.
6. Note that on this track, SNPs are color coded by location with synonymous SNPs appearing green, nonsynonymous appearing red, and other intronic SNPs gene regions appearing blue. A single SNP near the 3' end of that is color coded red (rs14835). Click on that SNP.
7. Flanking sequence surrounding this SNP can be found by clicking on the 'View DNA for this feature' link.
8. From the 'Get DNA' page, use the sequence retrieval options to set the upstream (5') and downstream sequence (3') to 100 bp. Click the 'get DNA' button at the bottom of the page to view this sequence.
9. Go 'BACK' to the 'Get DNA' page and follow the 'Table Browser' link.
10. The Table Browser allows you to access annotation data in text or tabular form. The 'Group' selection should be set to 'Variation and Repeats' and the 'Track' selection should be set to 'SNPs'. In the 'region' selection the position button should be selected with the chr16 coordinates of CKLF. Click on the 'Summary/Stats' button at the bottom of the page. How many SNPs are mapped to this region?
11. Go 'BACK' – now click the 'Get Output' button at the bottom of the page. A listing of summary data for all SNPs in this region.

SeattleSNPs.

<http://pga.gs.washington.edu>

The variation data for the CKLF can be accessed through the search box on the top right part of the home page, or via the 'Genes sequenced for SNPs' link in the left hand column under 'Gene Resources'.

Using this last link, find CKLF in the alphabetical listing of genes.

1. Under the 'Mapping Data' section, click on the 'cSNPs' link.
2. How many cSNPs were discovered in this gene? What is the synonymous SNP location in our reference sequence?

3. What was the cDNA position of this SNP?
4. In which population was it discovered?
5. Go 'BACK'. In the 'Genotyping' category, click on the 'Visual Genotype' link. An image of all the genotyping data for this gene is displayed. Using the SNP location of the synonymous SNP, determine which individual carries this polymorphism?
6. Explore other links in the 'Mapping Data', 'Genotyping Data' and 'Predictive Data' sections. The 'Linkage Data' and 'Haplotype Data' will be covered in a subsequent talk and tutorial. Compare the data at SeattleSNPs to that found in the other database resources.

Genome Variation Server:

<http://gvs.gs.washington.edu/GVS/>

This can also be accessed off the SeattleSNPs home page.

Note: Check that your browser is set to NOT block pop-up windows. GVS uses pop-up windows to display the images and tables.

1. Click the Gene Name search button
2. Enter 'CKLF'
3. A table of 'Populations' and 'Submitters' will appear along with the number of genotyped SNPs for each entry.
4. Select the button for 'PGA-European-Panel'. Do not change any of the display parameters.
5. Select the 'Display Genotypes' button at the bottom of the page.
6. Two windows should pop up, displaying a visual genotype image and a table with the genotypes for every sample at every SNP site.
7. Experiment with the other buttons for displaying results (e.g. display tagSNPs, etc.).
8. Parameters can be used to customize these results.

Answer Key:

Cross-database Search

1. Geo Profiles
3. SNP = 117 and Gene = 6
4. Sequences from many organisms each has a RefSeq

Entrez Gene

7. CKLF is on chromosome 16
8. 5' – TK2 and 3' CKLFSF1

dbSNP

2. How many cSNPs are found in dbSNP for CKLF? 3 (2-synonymous, 1-nonsynonymous) How many are validated? None.
4. 57
6. 5 genes HapMap confirmed
7. Hint: count the number without this data and subtract from total – 46
9. 4
11. (EUROPEAN = G = 0.763, A = 0.237, AFRICAN = G = 0.957, A = 0.043)

Entrez SNP

3. 72 snp
10. 28
13. 18 (if you don't release previous search) if you clear Validation: 2hit-2allele then 73
15. How many total entries are in dbSNP for this gene and submitted by the SeattleSNPs PGA project (our handle is PGA-UW-FHCRC)? 54

Perlegen Genotype Browser

2. How many SNPs are shown in this gene view? 5
4. Reference_Allele_Frequency_African_American: 0.909
Reference_Allele_Frequency_Chinese: 0.917
Reference_Allele_Frequency_European_American: 1.0
7. What is the length of that LD bin? 290105
8. How many SNPs are present in this bin? 12
10. What is the length of the longest haplotype for the European American population? 89096 And for the African American haplotype? 43717

HapMap Browser:

4. How many HapMap SNPs in this gene? 5

UCSC:

10. How many SNPs are mapped to this region? 57

SeattleSNPs Variation Data:

1. How many cSNPs were discovered in this gene? What is the synonymous SNP location in our reference sequence? 15467
2. What was the cDNA position of this SNP? 483
3. In which population was it discovered? European
4. Go 'BACK'. In the 'Genotyping' category, click on the 'Visual Genotype' link. An image of all the genotyping data for this gene is displayed. Using the SNP location of the synonymous SNP, determine which individual carries this polymorphism? E003

Genome Variation Server:

No answers to report.