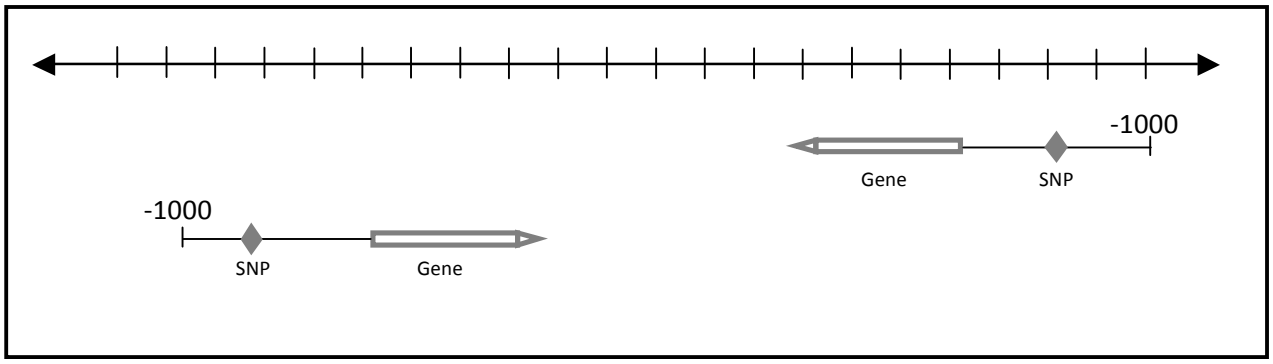
EXERCISE IV: Find all protein coding genes in *P. falciparum* that have SNPs in the region 1000 bp upstream of the gene's start site. The SNPs can be located on either strand. Only include SNPs with allele frequency between 0.2 and 0.5 in at least 10 strains. Go to http://plasmodb.org to begin your search.

Hint: This search requires a two-step strategy: a gene type search and a SNP search.



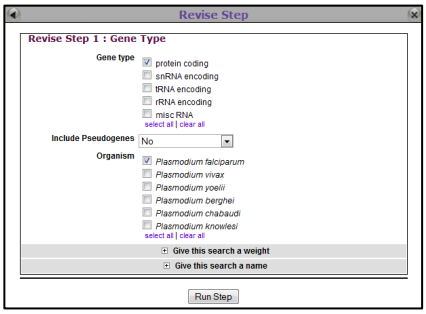
Run your first search.

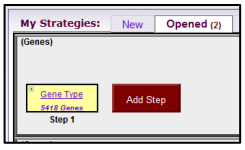
1. What search did you run?

2. How many IDs did you get?

3. What do you think you should do next?

SEARCH ONE HINTS
Identify Genes by:
Gene Attributes:
Gene Type:
Protein coding, only *P. falciparum*
Get answer = 5418 genes

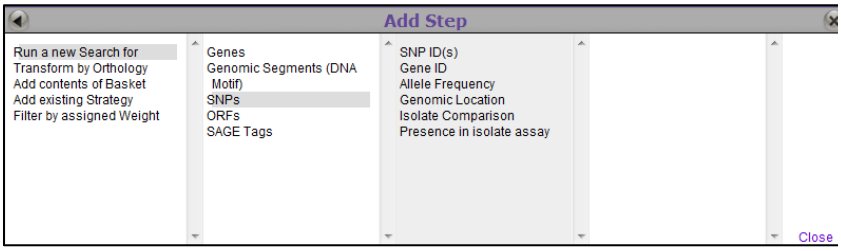


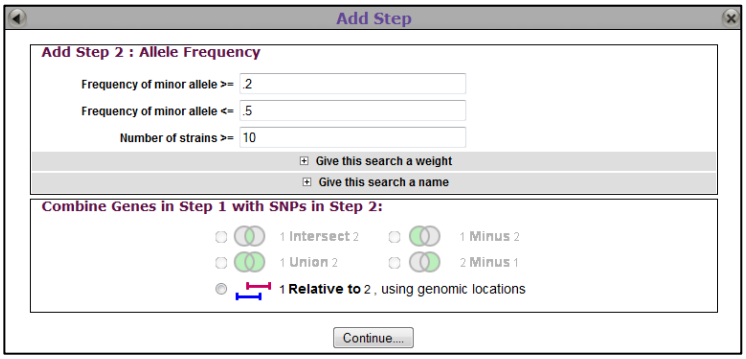


Click on "Add Step" to run a second search.

4. What search did you run?

5. How will you combine these results with your previous results?





SEARCH TWO HINTS

Add Step:

Run a new Search for:

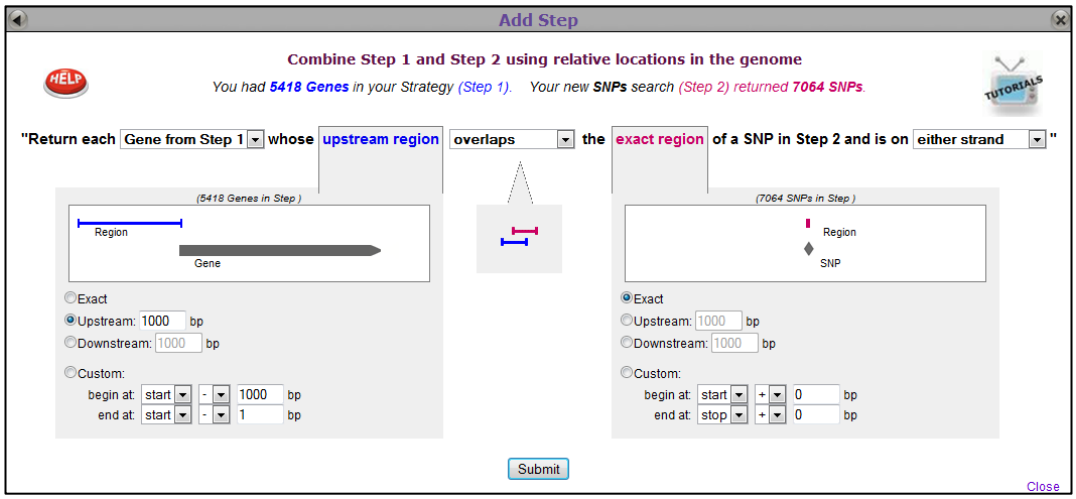
SNPs:

Allele frequency:

1 relative to 2:

Continue:

Now a new "Add Step" form appears. Use the form to combine the results of the two searches you just ran.
(Combine Step 1 and Step 2 using relative locations in the genome.)

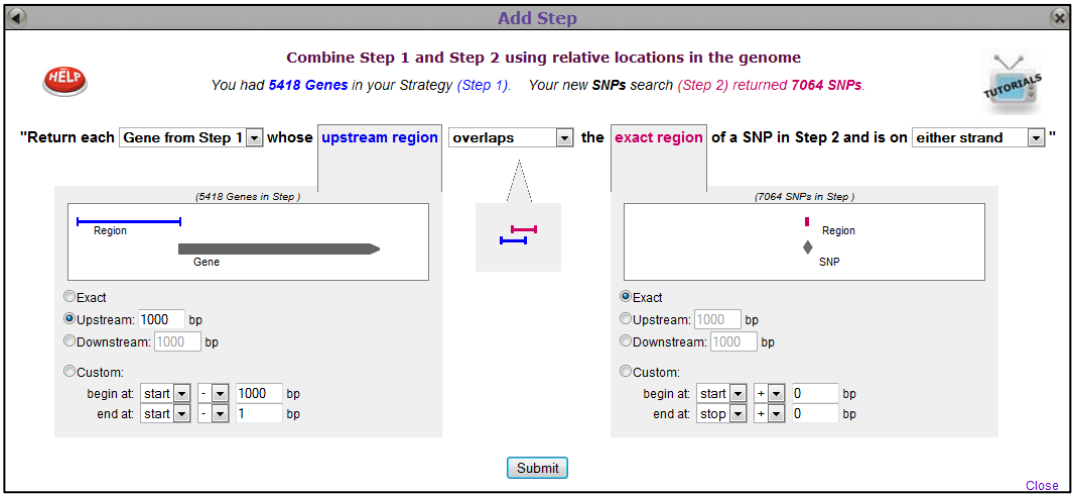


6. What options did you select in this form?

7. How many genes did you find?

HINTS FOR COMBINING STEPS:

There are **409 genes** in *P. falciparum* whose upstream 1000 bp contain SNPs with allele frequencies between 0.2 and 0.5 in at least 10 strains.



The following questions relate to the pop-up window used to combine search results (shown above).

8. Which term best describes the principle underlying the way search results are combined in this example?
   a. Feature Proximity
   b. Co-location
   c. Relative genomic location
   d. Span Logic
   e. Other _____

9. Did you notice the logic statement (the sentence: "Return each … whose … the ... and is on … strand") without the hints in page 3?

10. Did you notice the 5 selections (parameter choices for what to return, region, type of overlap, region, strand) that you need to make?

11. What do the blue and red color codings represent?

12. Did the color coding help you understand how to arrange the statement?

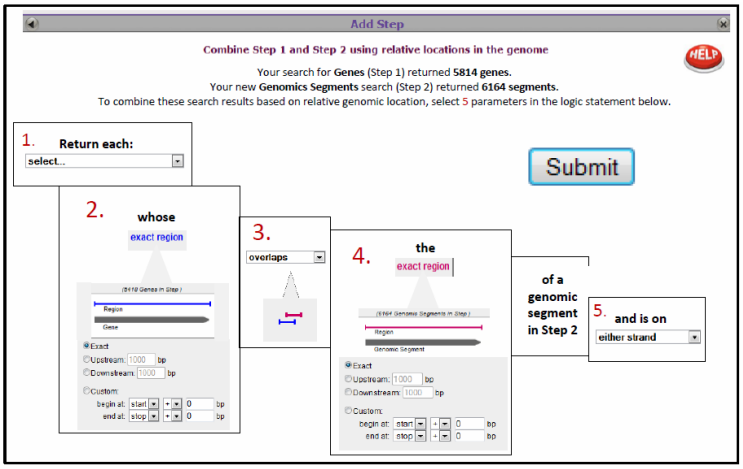
13. Did you notice the help button?

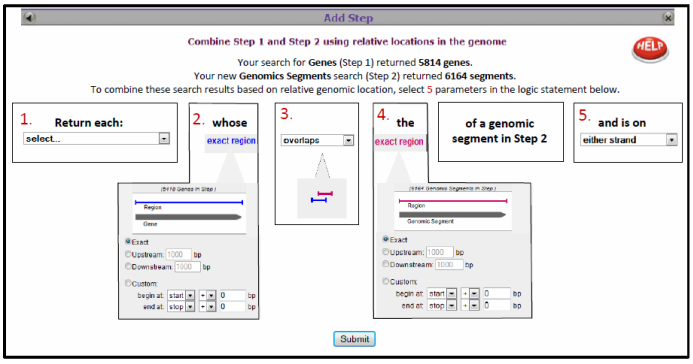
14. Did you use the help button?

15. Would you have used the help button from your desk at your home institution?

16. Which font style carries the greatest emphasis / most attracts your attention to notice the sentence?

    Return each **Return each** Return each Return each **Return each**





Here are two alternatives to the form used to combine search results using relative genomic location. Please answer the following questions.

17. In which of the forms are the five choices for parameters most apparent? (1= most apparent, 3 = least)

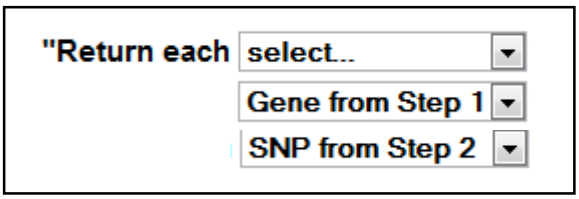
    ___ Top of page 4 ? ____ Above left? ___Above right?

18. Rank the forms windows for clarity (1 = most clear, 3 = least clear).

    ___ Top of Page 4 ___ Above left ___ Above right

19. Would this type of query be easier to use if the parameter choices were presented as a series of five pop-up windows with each window prompting the user to define a new parameter?

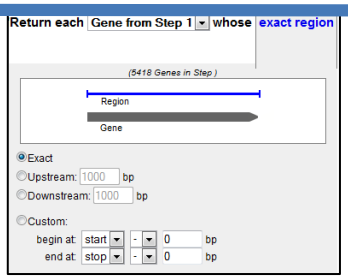
20. What is the purpose of this pull-down menu?

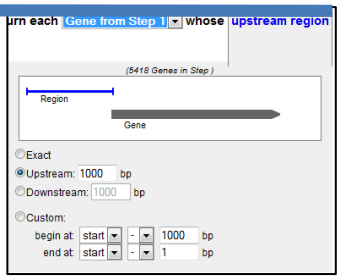


21. Did you understand this right away?

22. If you did not, then how did you figure this out?

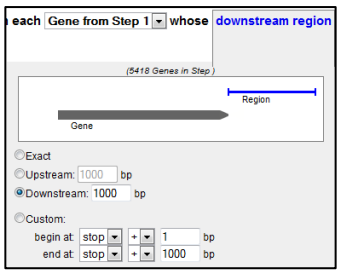
23. What is the purpose of these pull-down menus?

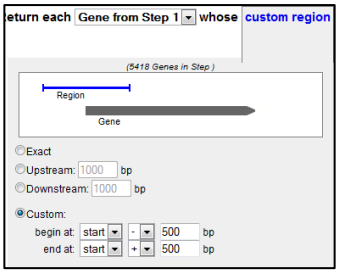




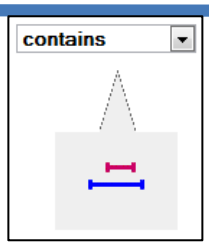
24. Did you understand this right away?

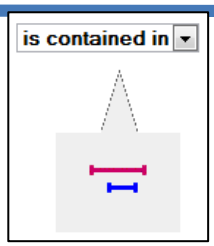
25. If not, then how did you figure this out?



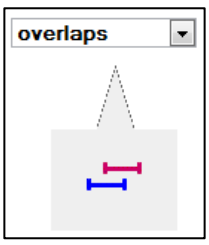


26. What is the purpose of these pull-down menus?





27. Did you understand this right away?



28. If not, then how did you figure it out?