

## Question 6

### How would one retrieve the sequence of a gene, along with all annotated exons and introns, as well as a certain number of flanking bases for use in primer design?

doi:10.1038/ng1194

This type of search can be initiated at the UCSC Genome Browser home page, located at <http://genome.ucsc.edu>. Select *Human* from the pull-down menu labeled *Organism*, and then click on *Browser*. This brings the user to the Human Genome Browser Gateway, from which a number of text- or position-based searches can be performed on current or older versions of the genome assembly. In this case, select the *Nov. 2002* assembly, type the name of the gene of interest (PTPN1) into the *position* box, and then click *Submit*. The Browser returns all genes starting with the characters 'PTPN1' (Fig. 6.1). The gene of interest here is the one called *PTPN1*; click on the hyperlinked *PTPN1* (arrow, Fig. 6.1) to view the genomic context of this gene (Fig. 6.2).

The text box at the top of Fig. 6.2 gives the absolute base pair position of this gene (chromosome 20, positions 48815311-48889509) and indicates that the gene spans 74 kb. The track labeled *Chromosome Band* shows that *PTPN1* is located at 20q13.13. Finally, the track marked *Known Genes* shows that the gene is on the forward strand, as the arrows on that track are pointing to the right. The exons within this gene are indicated by the vertical lines in the *Known Genes* track.

One way to obtain sequence upstream of a gene is described in Question 7. Here, we explain how to retrieve flanking sequence on both sides of a gene. To retrieve an adequate amount of sequence with which to design primers, one can increase the size of the region displayed by changing the position numbers within the *position* box at the top of the figure. To add an additional 1,000 nt at the 5' end and an additional 200 nt at the 3' end, for example, change the text in the *position* box to 'chr20:48814311-48889709' and click *Jump*. This now redraws the graphic with the new boundaries.

To obtain the actual sequence within the region, click on the *DNA* link in the blue bar at the top of the page. This produces a new page, entitled *Get DNA in Window* (Fig. 6.3). Change the

Sequence Formatting Options to *All lower case*, and click on the button called *Extended case/color options*. By selecting this option, the user can highlight features in the sequence by changing the format (case, underline, bold, italic) and/or color (red, green, blue) of the text. Colors can be varied in darkness and mixed together by changing the values in the boxes under *Red*, *Green* and *Blue* to any number between 0 and 255; examples of how to specify in RGB (red-green-blue) format color are given below the table. At this point, check the *Toggle Case* box in the *Known Genes* row, change the red saturation to 255 and leave the other color values set at zero (Fig. 6.4). Once the user clicks *Submit*, a new page is presented with the entire length of the sequence specified above (chr20:48814311-48889709) and the exons within this range are shown in red in capital letters (Fig. 6.5). This genomic sequence can now be saved and imported into a primer design or sequence assembly package for further analysis.

The Extended DNA Case/Color Options page can be used to combine and differentiate between genomic tracks. For example, return to the Options page, leave the *Known Genes* row as before but now also check the *Underline* square in the *Mouse Cons* row of the table. Clicking *Submit* produces a page on which the human exons still appear in red capital letters, but regions of alignment with the mouse are now shown as underlined text (Fig. 6.6). In this section of the gene, the conserved mouse sequence overlaps with the exons.

One way to retrieve sequence for a defined chromosomal region at the NCBI is with the *seq* link on the MapViewer, visible when the Gene\_Seq map is the master (Fig. 1.2). At Ensembl, export genomic nucleotide sequence with the *Export→FASTA* link in any ContigView window (Fig. 1.12, center yellow bar).

Figure 6.1

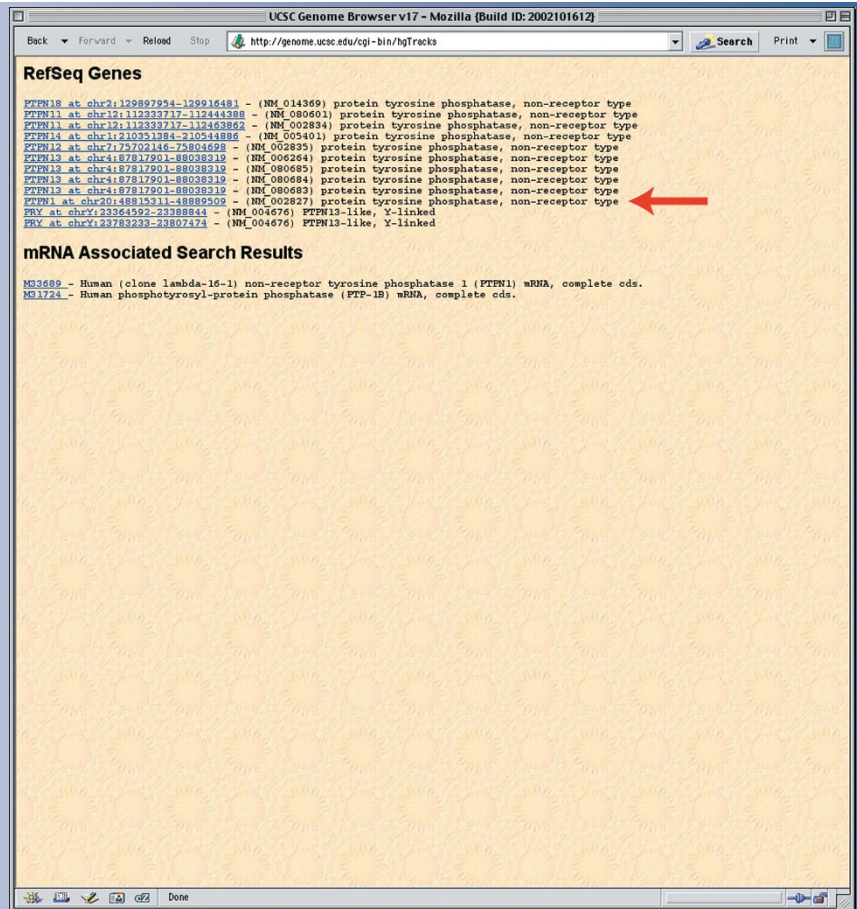
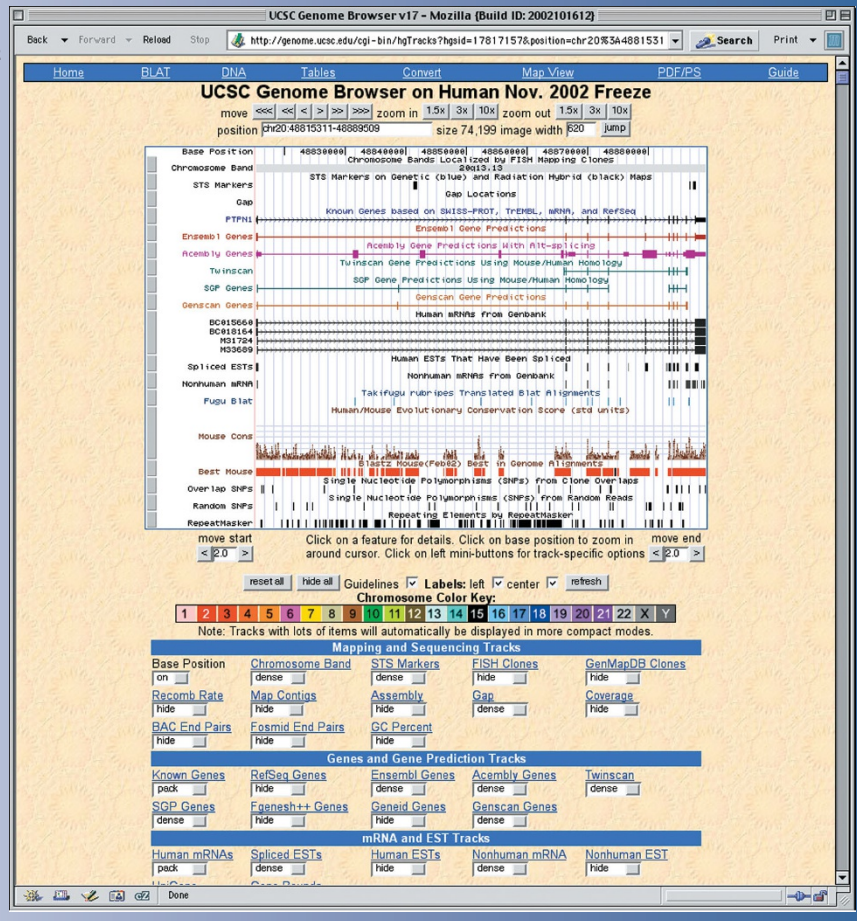


Figure 6.2



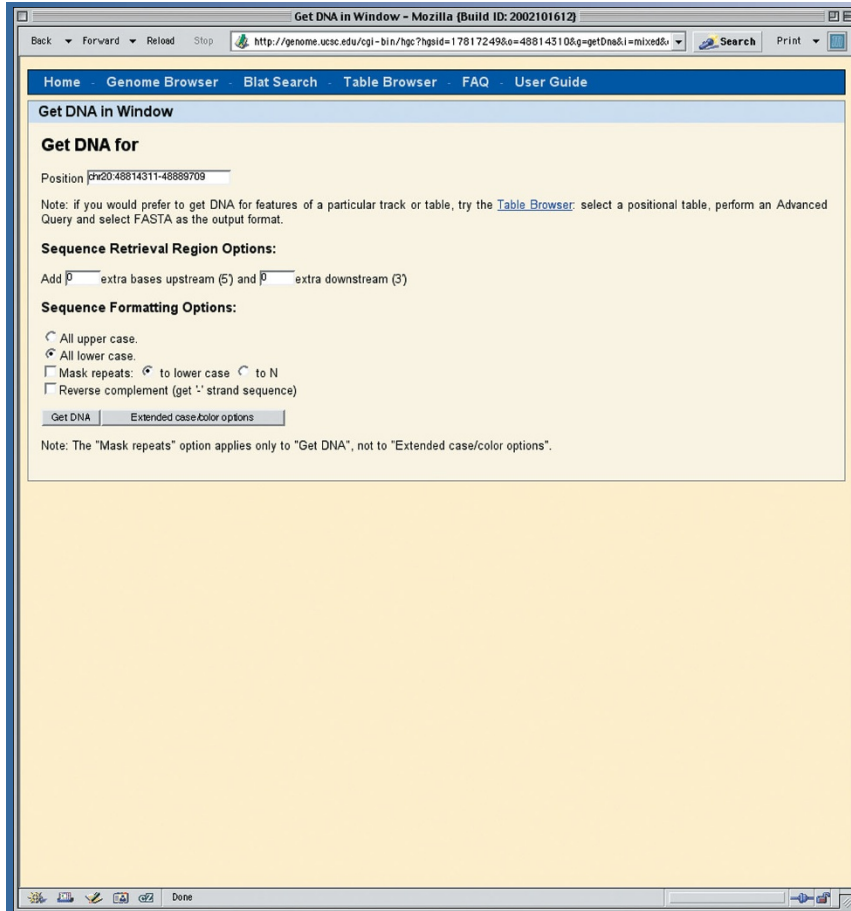


Figure 6.3

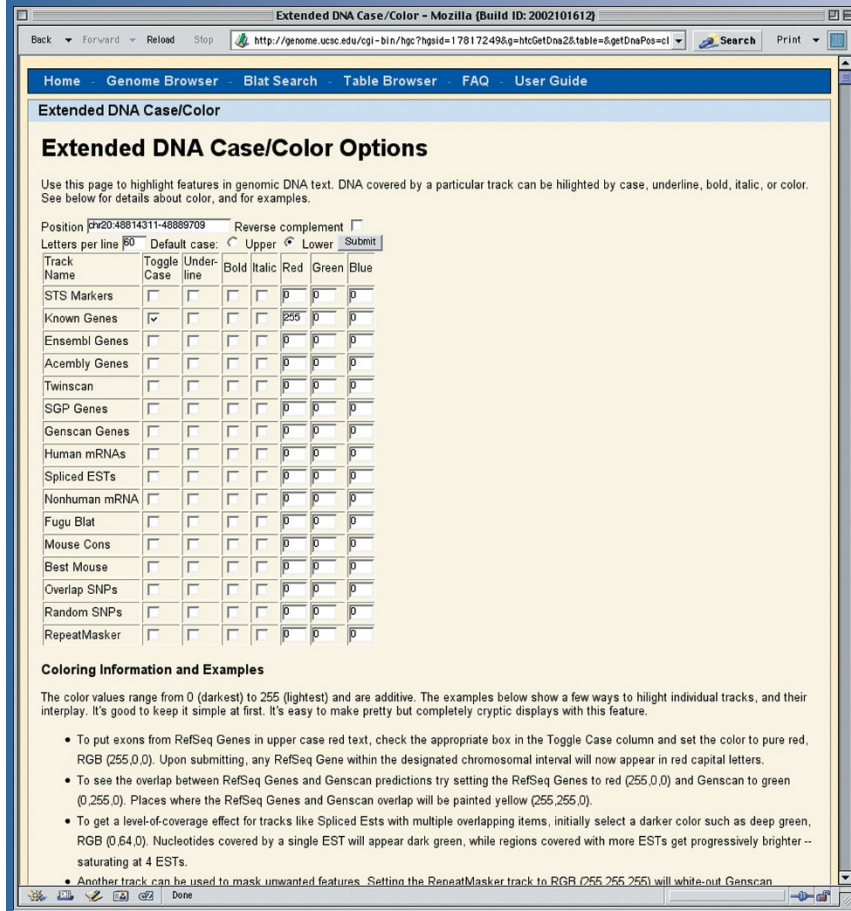


Figure 6.4

Figure 6.5

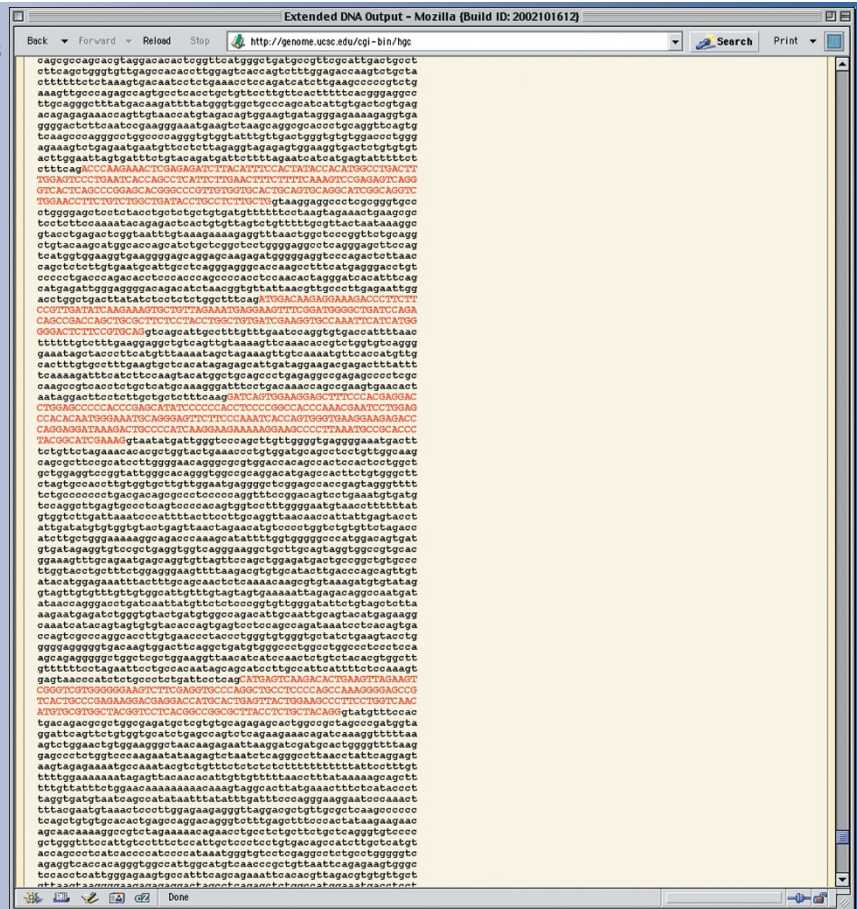


Figure 6.6

