Question 11

An investigator has identified and cloned a human gene, but no corresponding mouse ortholog has yet been identified. How can a mouse genomic sequence with similarity to the human gene sequence be retrieved?

doi:10.1038/ng976

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

For purposes of this example, assume that the user does not already have the human sequence of interest to hand. The first step will be to locate the human gene of interest using the UCSC Genome Browser. Begin by pointing to the UCSC Genome Browser home page, at http://genome.ucsc.edu. Select *Human* from the *Organism* pull-down menu and then click on *Browser*; both are located on the blue navigation bar at the left side of the page. This will take the user to the Human Genome Browser Gateway. Select the *Dec. 2001* version of the UCSC genome assembly, type the gene symbol 'AGPS' into the *position* box, and then click *Submit*. On the resulting page, follow the link for *AGPS* in the *Known Genes* section.

The result of the search on AGPS is shown in Fig. 11.1. In the main figure are a series of 'tracks', which are labeled along the left-hand side. The *Known Gene* track is for *AGPS*, corresponding to the query. Clicking on *AGPS* returns a summary of information on that gene, including the full name of the protein product (alkylglycerone phosphate synthase precursor), a link to the GeneCards database at the Weizmann Institute[20] and links to the translated protein, mRNA and genomic sequences. Focus now on the track labeled *Mouse Translated Blat Alignments*. What is shown in this track are the results of aligning the November 2001 version of the mouse genome assembly with the human genome using the program BLAT[8] in its translated protein mode. More details about the BLAT algorithm and about how the mouse BLAT track is automatically generated can be found by clicking on the *Mouse Blat* hyperlink found below the main graphical display.

Click anywhere within the *Mouse Blat* track to expand the single BLAT track so that it now shows each individual mouse sequence that aligns with human sequence in the region of interest (Fig. 11.2). Especially in a translated mode, mouse and human gene sequences are usually more similar in exons than in introns. Look carefully at the two alignments that derive from a mouse sequence called *chr3 81178k* (Fig. 11.2, arrow). On the *Mouse Blat* track, the brown vertical lines represent alignments and the horizontal lines are gaps. These alignments correspond to the blue vertical lines indicating the exons of *AGPS* on the *Known Genes* track.

To see the kind of information available for a translated BLAT alignment, click on the mouse genomic sequence labeled *chr3 81178k*. The resulting page (Fig. 11.3) provides the details of the alignment of the trace with the human genome assembly. This mouse genomic sequence is 607 nt in length and aligns with the human sequence in eight blocks. Within the blocks, the mouse and human sequences are 78% identical. To view the alignment itself, click on the *View details of parts of alignment. . .* link. On the resulting page (Fig. 11.4), the mouse sequence is shown on top, with the region of alignment in blue. The human genomic sequence is shown next, and a side-by-side alignment of the human and mouse sequences is at the bottom of the web page (not shown).

The NCBI's UniGene_Mouse map shows alignments of mouse mRNA and EST sequences with the human genome. Add this map using *Maps & Options* (Fig. 3.9). The easiest way to find the mouse ortholog of a human gene is probably to use Ensembl's precomputed *Homology Matches*. These matches, where available, link directly from a human gene to a putative mouse homolog (Fig. 1.9).

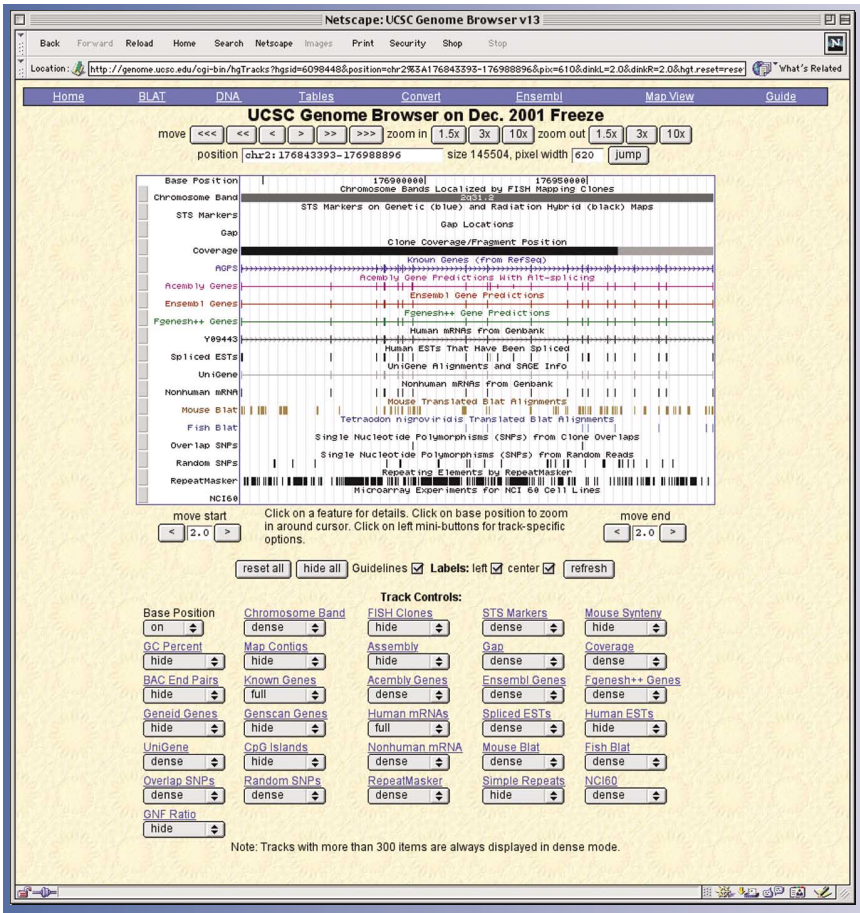

Figure 11.1

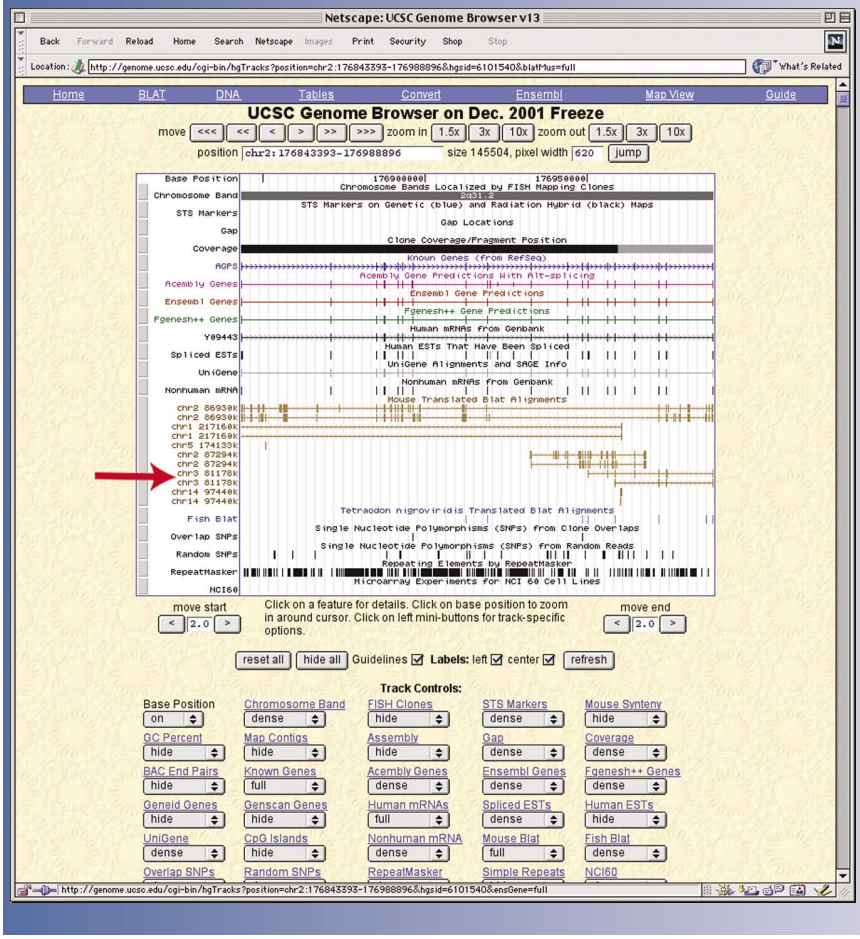

Figure 11.2

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

Figure 11.3

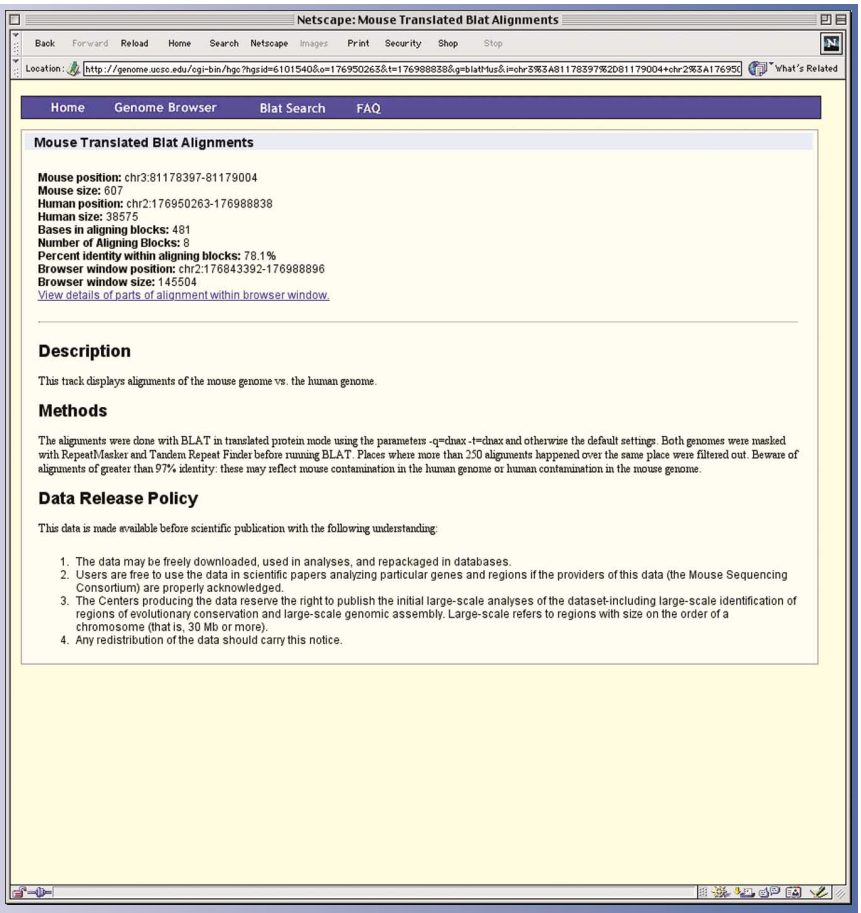
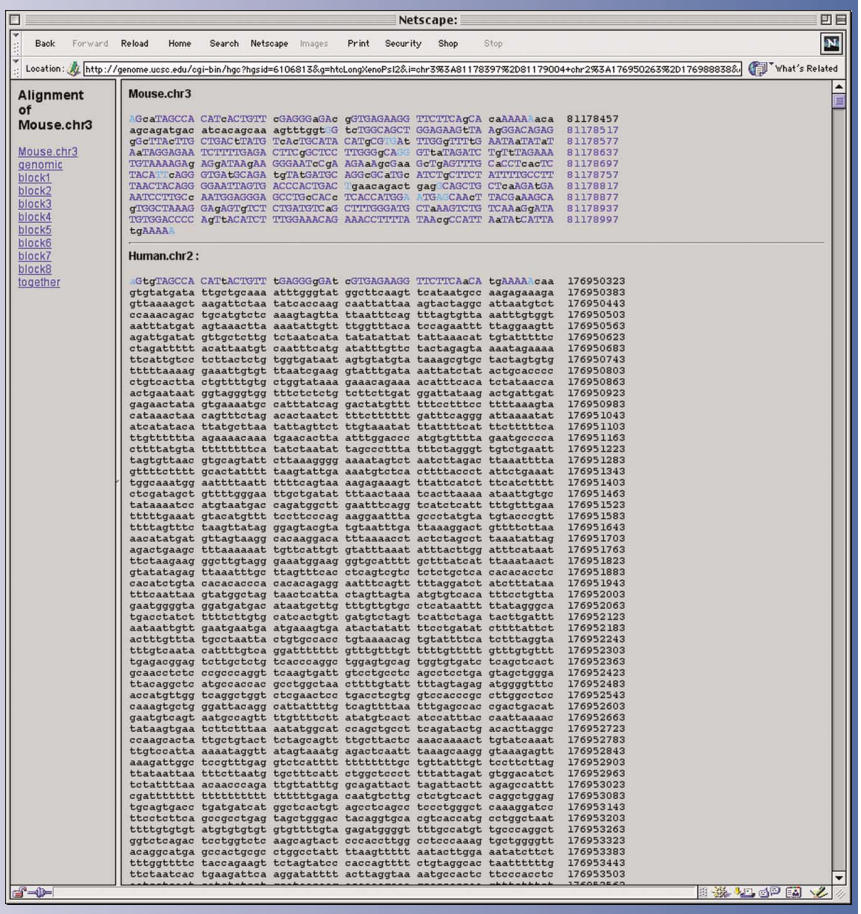
Figure 11.4