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

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 Institute20 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 BLAT8 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

UCSC Genome Browser on Dec. 26th Freeze

Move start Click on a feature for details. Click on base position to zoom around current. Click on scrollbar buttons for zoom-specific options.

Base: Position

Track Controls:

Figure 11.2

UCSC Genome Browser on Dec. 26th Freeze

Move start Click on a feature for details. Click on base position to zoom around current. Click on scrollbar buttons for zoom-specific options.

Base: Position

Track Controls:
Figure 11.3

Mouse Translated Blat Alignments

Mouse position: cyt93179378·1799829
Mouse Locus: 937
Human position: cyt1799829·17998310
Human Locus: 30978
Database scope: 4511
Number of aligned blocks: 1
Percent identity within aligned blocks: 70.1%
Browser window position: cyt1799829·17998996
Browser window size: 1400x4

Description

This tool displays alignments of the mouse genome vs. the human genome.

Methods

The alignments were done with BLAT in translated protein mode using the parameters: -lt250, -m=1, -score=200, and -score=1000. The results were visualized with RepeatMasker and Toor's Repeat Pileup before running BLAT. Please refer to the README file for details. Because of the high percentage of identity, these may not reflect true conserved regions in the human genome.

Data Release Policy

This data is made available before scientific publication with the following restrictions:

1. The data may be freely distributed, used in analyses, and repurposed in databases.
2. Users are free to use the data in scientific papers analyzing particular genes and regions if the providers of the data (the Mouse Sequencing Consortium) are properly acknowledged.
3. The Centers producing the data reserve the right to publish the initial large-scale analyses of the dataset including large-scale identification of regions of evolutionary conservation and large-scale genomic information. Large-scale refers to regions with size on the order of 50 kb or more.
4. Any redistribution of the data should carry this notice.