

Question 8

How can one find all the members of a human gene family?

doi:10.1038/ng973

The HUGO Gene Nomenclature Committee (<http://www.gene.ucl.ac.uk/nomenclature/>) has been working to develop a unique symbol, as well as a longer and more descriptive name, for each human gene. Thus, members of many gene families, previously cloned in different laboratories and known by a variety of terms, now share a common gene symbol. A text search in any of the genome browsers will often return links to all named members of a gene family that have been mapped to the genome. Whereas Ensembl and UCSC currently return lists of the genes, the NCBI presents both a list and a graphical overview.

Go to the NCBI home page at <http://www.ncbi.nlm.nih.gov/> and click on the *Human map viewer* link on the right side to access the Map Viewer search page. Enter the term 'ADAM*[sym]' in the text query box. The asterisk, or wild card, will match any character, whereas the term [sym] limits the search to items with ADAM as their gene symbol. Other advanced search options are available by clicking the *Advanced Search* box or by reading the online documentation. The search returns 41 hits, which include members of the ADAM family as well as other related families whose names start with the term 'ADAM', such as ADAMTS and ADAMDEC. To limit the search to ADAM genes only, eliminate the undesired gene symbols with the Boolean NOT term, using the query ADAM*[sym] NOT ADAMTS*[sym] NOT ADAMDEC1*[sym]. The graphic at the top of the returned page shows the location of each gene with a red tick mark (Fig. 8.1). It is immediately clear that the 19 mapped ADAM genes are distributed among 11 chromosomes, and that some, such as those at the tips of the q arms of chromosomes 10 and 14, are close together. The list at the bottom of the page presents links to the 19 genes.

Another way to search for homologous genes in the genome is through a basic local alignment search tool (BLAST) search at the NCBI or Ensembl. BLAT searches at UCSC are not as sensitive as BLAST searches and may not find as many homologous genes. In this example, all genomic sequences homologous to the ADAM2 protein will be found using the Ensembl BLAST interface. From the Ensembl Human home page at http://www.ensembl.org/Homo_sapiens/, click on the link to

BLAST. Paste the sequence of the ADAM2 protein (GenBank accession NP_001455.2) into the query box (having obtained the protein sequence from the NCBI's Entrez database by following the steps in Question 5). Set the database to *Homo sapiens, genomic sequence* to search the Ensembl genome assembly, and choose TBLASTN as the executable (Fig. 8.2). Use the default parameters for the remaining settings. When done, click *Search*. The returned page will contain a retrieval ID (Fig. 8.3), which, when the search is finished, will link to the search results page (Fig. 8.4).

The top of the results page shows a graphical overview of the locations of hits. These hits may be to the entire protein or just to a single domain. The hits are colored by BLAST score, red being most similar, blue least similar and green intermediate. Some of the hits, like the pairs on the q arms of chromosomes 10 and 14, lie in positions similar to those of ADAMs mapped by the NCBI (Fig. 8.1), but others, such as those on chromosomes 12 and Y, are unique to the BLAST search. These unique hits may represent real members of the ADAM family that have not yet been named and would therefore not show up in a text-based search. Alternatively, they may be unnamed pseudogenes or nonsignificant BLAST hits. One gene on chromosome 1 is found in the text-based search at the NCBI but not in the BLAST search at Ensembl. The similarity between this gene and ADAM2 is not high enough for it to appear in the BLAST search using the default Ensembl parameters.

Clicking on an arrow next to one of the hits shown in Figure 8.4 activates a pop-up menu that gives the details of the BLAST report and provides links to the BLAST alignment and the ContigView (Figs 8.5 and 8.6, respectively, for the hit on chromosome 12). The hit on chromosome 12 contains a stop codon and is probably an intronless pseudogene. The bottom of the results page (Fig. 8.4) shows a summary of the BLAST hits. Clicking on a hit links to the BLAST alignment (Fig. 8.5). A link in the middle of the results page (Fig. 8.4) provides the entire BLAST report in standard format. Clicking on a hit in the BLAST report retrieves the ContigView for the region around the hit (similar to what is shown in Fig. 8.6).

Figure 8.3

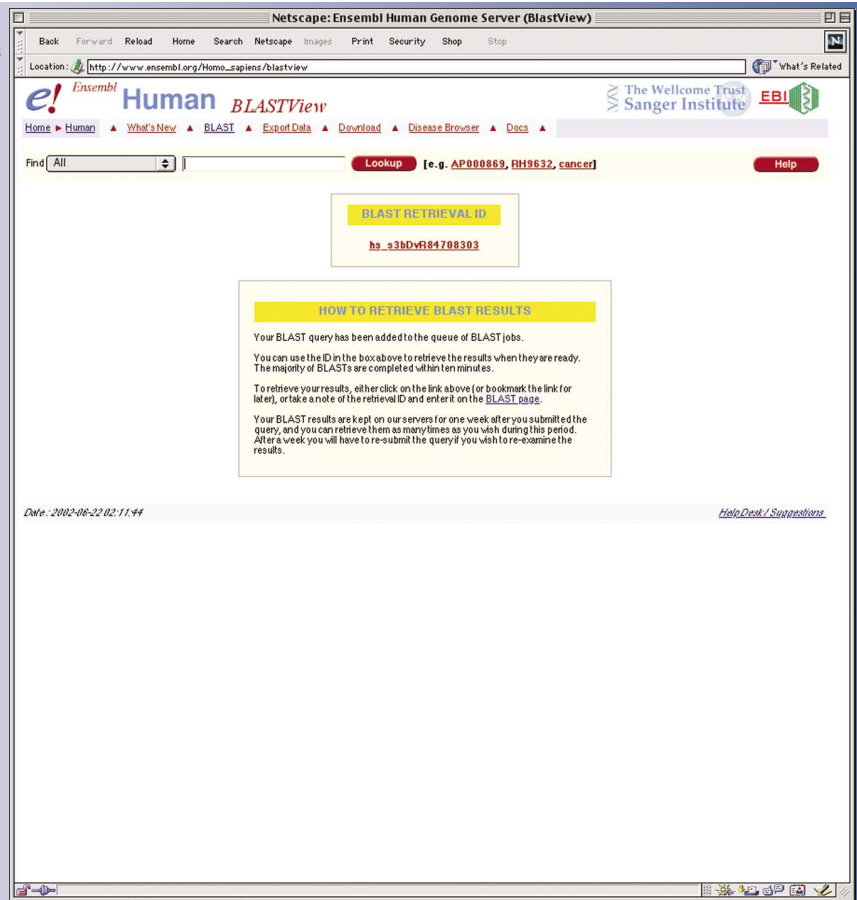
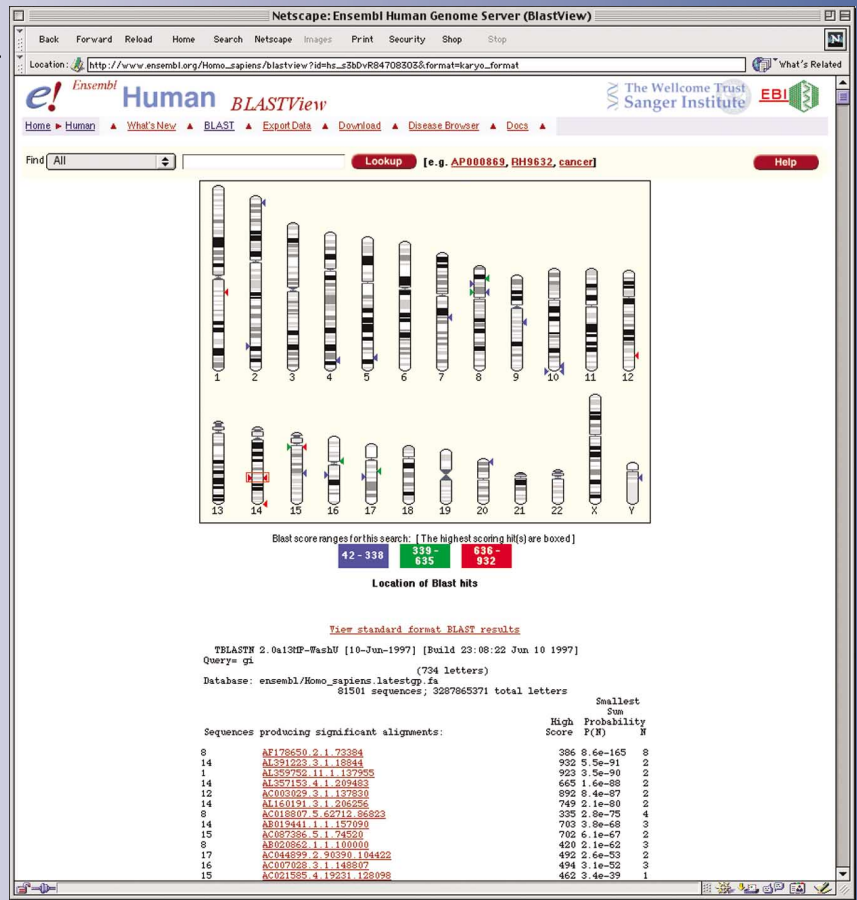


Figure 8.4



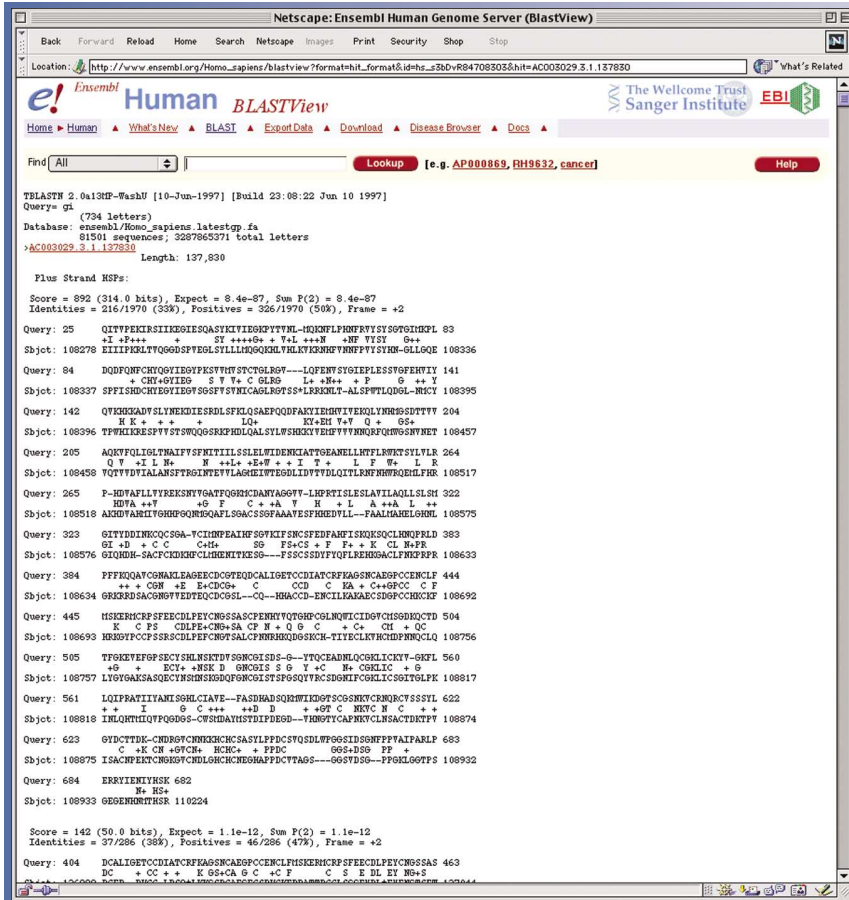


Figure 8.5

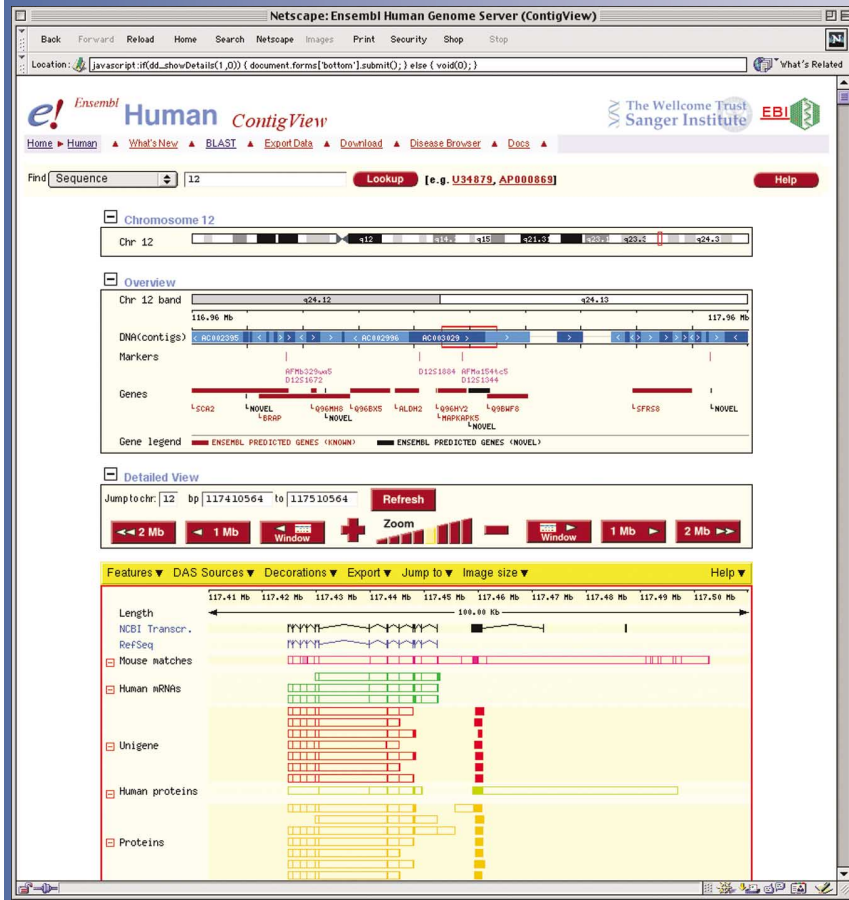


Figure 8.6