

Algorithms in Bioinformatics II, SS2004

Assignment sheet # 4

Daniel Huson

May 10, 2004

1 Run Genscan on annotated sequence (5 points)

Please view the NCBI entry of the “nucleotide sequence of the human ornithine decarboxylase gene” with accession number X16277. Refer to the gene described there as the *actual gene*. Run the complete DNA sequence through Genscan using <http://genes.mit.edu/GENSCAN.html>. Refer to Genscan’s result as the *predicted gene*.

Please compare the exons of the actual and predicted gene, reporting the number of *true positive*, *true negative*, *false positive* and *false negative* classifications of individual bases, and also of whole exons.

Do the same for the NCBI accession J03733, which contains the mouse ornithine decarboxylase gene.

2 Homology-assisted gene finding (2 points)

To obtain a better prediction of the mouse ornithine decarboxylase gene present in the nucleotide sequence of accession J03733, use the homology-assisted gene finder *Genomscan* running at <http://genes.mit.edu/genomscan.html>.

To do so, paste in the DNA sequence of J03733 as target sequence and the *protein* sequence reported in X16277 as the protein homolog sequence.

Compare the output of this program with the actual gene structure reported in J03733. Can you explain the improvement?

3 Suffix tree (3 points)

Please draw the suffix tree for the text

```
tomtomtomato$.
```

Assignments due: **Monday, May 17, 10am**