

Prática 7 - Mapeamento de Transcritos no Genoma

Ricardo de Marco Alessandro S. Nascimento

19 de Outubro de 2016

1 Uso do programa *Splign* para mapeamento de transcritos

1. Selecione a sequência **NM_001135146.1** na base de nucleotídeos do NCBI.
2. 1) Vá a site do programa [BLAST](#) no NCBI e selecione o programa *BLASTN*.
3. Na opção “*database*” escolha a opção “*reference genomic sequences*”;
4. No campo “*organism*” escreva “*Homo sapiens*”;
5. Abra a porção denominada “*algorithm parameters*” e na opção “*Filter*” deselegione a opção “*low complexity regions*”. Os parâmetros ajustados nos passos 3, 4 e 5 permitirão que você faça uma busca bom o BLAST somente em versões depositadas do genoma humano e sem filtrar os trechos com baixa complexidade, permitindo assim um alinhamento completo do transcrito.
6. Aperte o botão BLAST.
7. **Serão gerados quatro alinhamentos com versões diferentes do genoma humano. Concentre-se apenas em um dos alinhamentos. Você consegue descrever quantos *exons* este gene possui? Em qual base do cromossomo o gene começa e em qual ele termina? Anote estes números. Existem bases do transcrito que de acordo com o BLAST poderiam estar localizadas em dois exons diferentes? (Q1)**
8. Selecione a sequência genômica do alinhamento que você examinou clicando no hiperlink contendo o numero de acesso;
9. Na página de registro da sequência genômica (representando todo o cromossomo 4), adicione nos campos “*change region shown*” os valores que você anotou para o início e o fim do gene.
10. No campo “*Display*” selecione a opção FASTA;
11. Na sua tela deverá aparecer a sequência no formato FASTA representando a região do genoma contendo seu gene.

12. Abra a página do [Splign](#). Clique em "*Online*" na barra superior para utilizar a versão online do programa.
13. No campo cDNA, digite os seguintes códigos de acesso:
 - NM_022154
 - NM_001135146
 - NM_001135147
 - NM_001135148
14. No campo "Genomic", digite o código de acesso da sequência genômica que você obteve através da busca no BLAST.
15. Clique em "*Align*";
16. **Examine os alinhamentos produzidos. É possível ter uma melhor definição dos *exons* que um simples alinhamento no BLAST? No que diferem as diferentes formas de *splicing*? A sequência NM_001135147 é mapeada completamente? O que deve ter acontecido e como isso pode ser corrigido? (Q2)**

2 Predição de Transcritos Utilizando o Genscan

Nesta atividade iremos utilizar o programa genscan para realizar a predição de genes utilizando uma sequência de DNA genômico de camundongo como ponto de partida. Iremos comparar as predições realizadas com a sequência bruta com aquela com mascaramento para elementos repetitivos e de baixa complexidade. Finalmente iremos comparar os dados destas predições com aqueles de transcritos reais depositados nos bancos públicos.

1. Faça o download do arquivo [sequencia-atividade-8.txt](#)
2. Vá até a webpage do [Genscan](#)
3. Copie toda a sequência no espaço reservado e aperte o botão "*run gene Scan*". Isso fará com que o programa realize a predição de genes na sequência que você submeteu.
4. Copie as coordenadas previstas e a sequência de proteína para um arquivo de texto. Este transcrito previsto será denominado Transcrito-1.
5. Vá até a webpage do [RepeatMasker](#).
6. Copie a sequência do arquivo "*sequencia-atividade-8.txt*" no espaço reservado.
7. Vá ao campo "*DNA source*" e selecione a opção *mouse*.
8. Aperte o botão "*Submit Sequence*".
9. Na página resultante inspecione a quantidade de repetições previstas e vá até o fim da página e aperte o hiperlink do lado do campo "masked file";

10. Uma página contendo a sequência mascarada será aberta. Copie a sequência contida na página;
11. Repita a predição do *Genscan* utilizando esta nova sequência.
12. **Copie as coordenadas previstas e a sequência de proteína para um arquivo de texto. Este transcrito previsto será denominado Transcrito-2. O transcrito predito possui a mesma estrutura do transcrito 1? (Q3)**
13. **Utilize as proteínas codificadas pelo transcrito 1 e 2 em um BLASTP contra a base de dados não-redundante (nr). Que resultado é obtido? Algum dos dois modelos é igual a alguma proteína do banco? O que deve ter ocorrido? (Q4)**
14. Selecione a proteína do banco *nr* que seja mais parecida com os seus transcritos e obtenha a sequência de cDNA (somente a região do CDS) que codifique esta proteína.
15. **Utilizando o programa *Splign* mapeie este cDNA no DNA do arquivo sequencia-atividade-8.txt. Anote as coordenadas do gene e a sequência proteica em um arquivo de texto. Este transcrito previsto será denominado Transcrito-3. Qual a diferença de estrutura do transcrito-3 em relação aos outros? Tendo em vista o resultado obtido, discuta a utilidade de programas de predição de genes. (Q5)**