**Version 07/04/2022**

1. **Work with the following sequence obtained after sequencing (also in Moodle):**

TACTGTTTTCGTACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTATTCATACCGTCCCACCAT

CGGGCGCGGATCTTTTTATCTAGCATAGCCAAAAAGAAAGAGCTTGCACATATGGAGAGATCAAACAGCA

CAGCTTCTATGGCCGTGCAAGAACTTCACCATGGAGCTATGGAGATTATGATAATTGCCAACAGGATCAT

GATTATCTTCTAGGGTTTTCATGGCCACCAAGATCCTACACTTGCAGCTTCTGCAAAAGGGAATTCAGAT

CGGCTCAAGCACTTGGTGGCCACATGAATGTTCACAGAAGAGACAGAGCAAGACTCAGATTACAACAGTC

TCCATCATCATCTTCAACACCTTCTCCTCCTTACCCTAACCCTAATTACTCTTACTCAACCATGGCAAAC

TCTCCTCCTCCTCATCATTCTCCTCTAACCCTATTTCCAACCCTTTCTCCTCCATCCTCACCAAGATATA

GGGCAGGTTTGATCCGTTCCTTGAGCCCCAAGTCAAAACATACACCAGAAAACGCTTGTAAGACTAAGAA

ATCATCTCTTTTAGTGGAGGCTGGAGAGGCTACAAGGTTCACCAGTAAAGATGCTTGCAAGATCCTGAGG

AATGATGAAATCATCAGCTTGGAGCTTGAGATTGGTTTGATTAACGAATCAGAGCAAGATCTGGATCTAG

AACTCCGTTTGGGTTTC

* Determine if this sequence is contaminated with vector, rewrite the purified sequence in FASTA format. 



>purified sequence

TTTTATCTAGCATAGCCAAAAAGAAAGAGCTTGCACATATGGAGAGATCAAACAGCACAG

CTTCTATGGCCGTGCAAGAACTTCACCATGGAGCTATGGAGATTATGATAATTGCCAACA

GGATCATGATTATCTTCTAGGGTTTTCATGGCCACCAAGATCCTACACTTGCAGCTTCTG

CAAAAGGGAATTCAGATCGGCTCAAGCACTTGGTGGCCACATGAATGTTCACAGAAGAGA

CAGAGCAAGACTCAGATTACAACAGTCTCCATCATCATCTTCAACACCTTCTCCTCCTTA

CCCTAACCCTAATTACTCTTACTCAACCATGGCAAACTCTCCTCCTCCTCATCATTCTCC

TCTAACCCTATTTCCAACCCTTTCTCCTCCATCCTCACCAAGATATAGGGCAGGTTTGAT

CCGTTCCTTGAGCCCCAAGTCAAAACATACACCAGAAAACGCTTGTAAGACTAAGAAATC

ATCTCTTTTAGTGGAGGCTGGAGAGGCTACAAGGTTCACCAGTAAAGATGCTTGCAAGAT

CCTGAGGAATGATGAAATCATCAGCTTGGAGCTTGAGATTGGTTTGATTAACGAATCAGA

GCAAGATCTGGATCTAGAACTCCGTTTGGGTTTC

• Does the purified sequence encode a protein? How long is the longest open reading frame (ORF)? the longest ORF: 414nt=137aa





• What organism does this sequence likely come from? *A. thaliana*



• Will the purified sequence be cleaved by the following enzymes: EcoRI, KpnI or MseI?

EcoRI 1x

KpnI 0x

MseI 1x







• What fragments are formed after cleavage by all these enzymes at once?

Three fragments: 400bp,188bp,46bp



1. **Find a human protein sequence called FOX1**
* What is the accession number and function of this protein?

**Q9NWB1 / NP\_665898**

RNA-binding protein that regulates alternative splicing events by binding to 5'-UGCAUGU-3' elements. Regulates alternative splicing of tissue-specific exons and of differentially spliced exons during erythropoiesis.

• Does this protein have any transmembrane regions? None



* How many cysteines does the sequence have? 1 cysteine



• Compare how similar the protein is to the respective mouse homologue?

Laling: 

Needle



* Design primers to amplify the CDS of respective gene.

>NM\_145891.3:158-1414 Homo sapiens RNA binding fox-1 homolog 1 (RBFOX1), transcript variant 1, mRNA

ATGCTGGCGTCTCAAGGAGTTCTCCTGCATCCTTATGGCGTGCCTATGATTGTACCGGCAGCTCCTTACC

TTCCTGGACTGATTCAGGGTAATCAGGAAGCAGCCGCTGCCCCTGACACAATGGCTCAGCCTTACGCTTC

GGCCCAGTTTGCTCCCCCGCAGAACGGTATCCCCGCGGAATACACGGCCCCTCATCCCCACCCCGCGCCA

GAGTACACAGGCCAGACCACGGTTCCCGAGCACACATTAAACCTGTACCCTCCCGCCCAGACGCACTCCG

AGCAGAGCCCGGCGGACACGAGCGCTCAGACCGTCTCTGGCACCGCCACACAGACAGATGACGCAGCACC

GACGGATGGCCAGCCCCAGACACAACCTTCTGAAAACACGGAAAACAAGTCTCAGCCCAAGCGGCTGCAT

GTCTCCAATATCCCCTTCAGGTTCCGGGATCCGGACCTCAGACAAATGTTTGGTCAATTTGGTAAAATCT

TAGATGTTGAAATTATTTTTAATGAGCGAGGCTCAAAGGGATTTGGTTTCGTAACTTTCGAAAATAGTGC

CGATGCGGACAGGGCGAGGGAGAAATTACACGGCACCGTGGTAGAGGGCCGTAAAATCGAGGTAAATAAT

GCCACAGCACGTGTAATGACAAATAAAAAGACCGTCAACCCTTATACAAATGGCTGGAAATTGAATCCAG

TTGTGGGTGCAGTCTACAGTCCCGAATTCTATGCAGGCACGGTCCTGTTGTGCCAGGCCAACCAGGAGGG

ATCTTCCATGTACAGTGCCCCCAGTTCACTTGTATATACTTCTGCAATGCCAGGCTTCCCGTATCCAGCA

GCCACCGCCGCGGCCGCCTACCGAGGGGCGCACCTGCGAGGCCGCGGTCGCACCGTGTACAACACCTTCA

GGGCCGCGGCGCCCCCGCCCCCGATCCCGGCCTACGGCGGAGTAGTGTATCAAGAGCCTGTGTATGGCAA

TAAATTGCTGCAGGGTGGTTATGCTGCATACCGCTACGCCCAGCCTACCCCTGCCACTGCCGCTGCCTAC

AGTGACAGTTACGGACGAGTTTATGCTGCCGACCCCTACCACCACGCACTTGCTCCAGCCCCCACCTACG

GCGTTGGTGCCATGAATGCTTTTGCACCTTTGACTGATGCCAAGACTAGGAGCCATGCTGATGATGTGGG

TCTCGTTCTTTCTTCATTGCAGGCTAGTATATACCGAGGGGGATACAACCGTTTTGCTCCATACTAA

F: ATGCTGGCGTCTCAAGGAG

R: TTAGTATGGAGCAAAACGGTTG

F:



R:

