**Daniel Johnson**

**A53072130**

**BGGN213 2019 Midterm “Find a Gene Project”**

**[Q1]Protein Name**

Protein Name: **CBFA2T1** (Product of *RUNX1T1* Gene)

Species: **Human**

Accession: **NX\_Q06455**

**[Q2] Blast Search Results**

**Blast Input:**

>sp|Q06455|MTG8\_HUMAN Protein **CBFA2T1** OS=**Homo sapiens** OX=9606 GN=**RUNX1T1** PE=1 SV=2

MISVKRNTWRALSLVIGDCRKKGNFEYCQDRTEKHSTMPDSPVDVKTQSRLTPPTMPPPP

TTQGAPRTSSFTPTTLTNGTSHSPTALNGAPSPPNGFSNGPSSSSSSSLANQQLPPACGA

RQLSKLKRFLTTLQQFGNDISPEIGERVRTLVLGLVNSTLTIEEFHSKLQEATNFPLRPF

VIPFLKANLPLLQRELLHCARLAKQNPAQYLAQHEQLLLDASTTSPVDSSELLLDVNENG

KRRTPDRTKENGFDREPLHSEHPSKRPCTISPGQRYSPNNGLSYQPNGLPHPTPPPPQHY

RLDDMAIAHHYRDSYRHPSHRDLRDRNRPMGLHGTRQEEMIDHRLTDREWAEEWKHLDHL

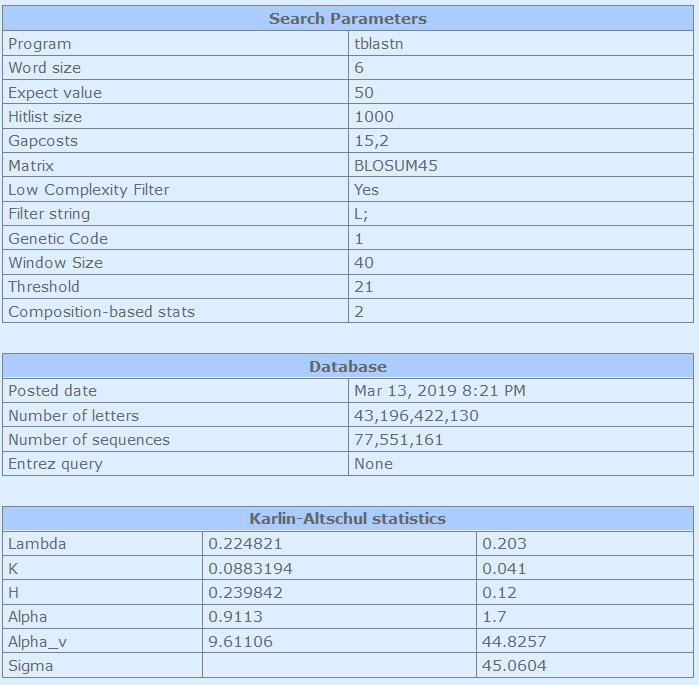
LNCIMDMVEKTRRSLTVLRRCQEADREELNYWIRRYSDAEDLKKGGGSSSSHSRQQSPVN

PDPVALDAHREFLHRPASGYVPEEIWKKAEEAVNEVKRQAMTELQKAVSEAERKAHDMIT

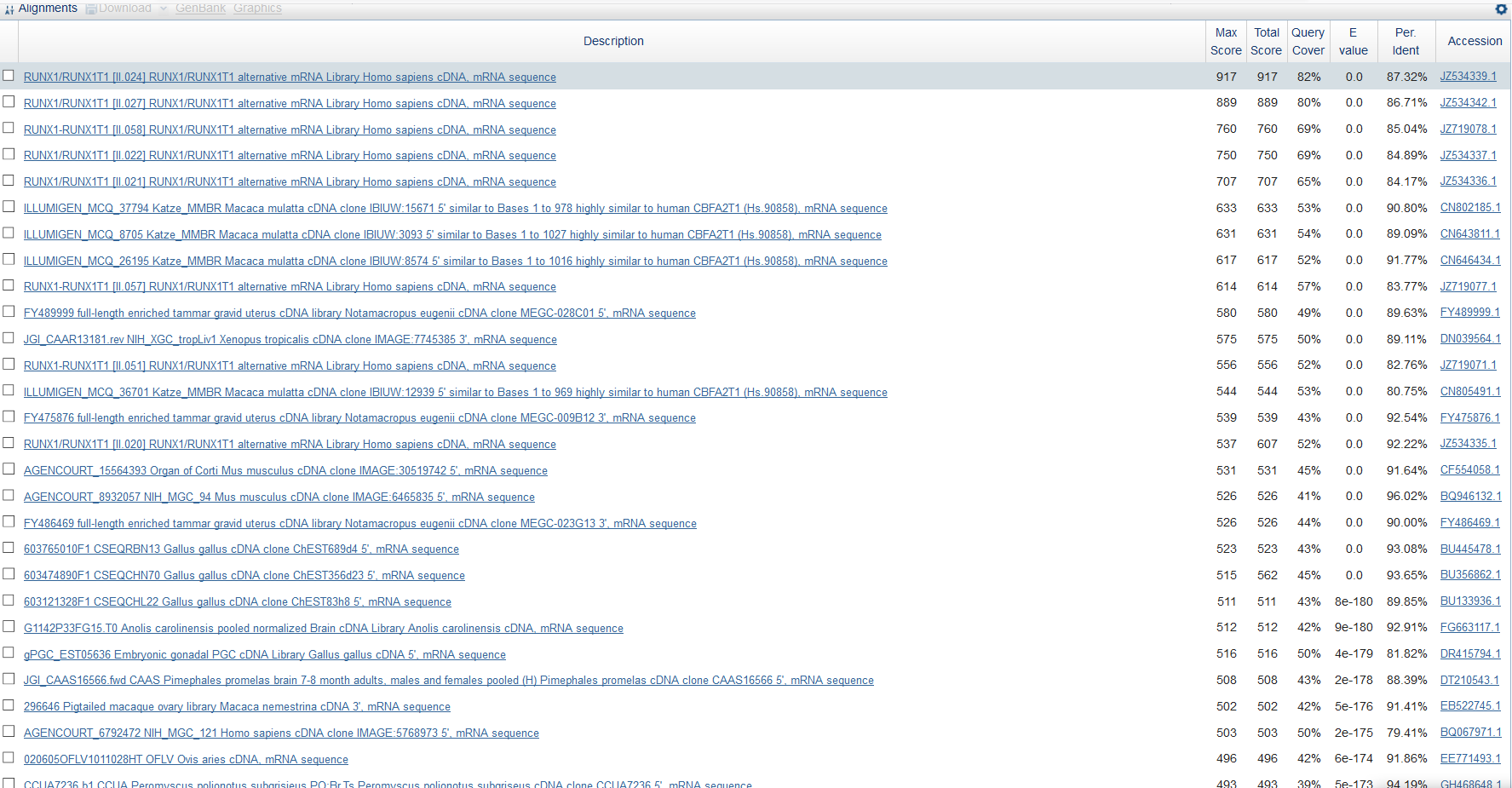
TERAKMERTVAEAKRQAAEDALAVINQQEDSSESCWNCGRKASETCSGCNTARYCGSFCQ

HKDWEKHHHICGQTLQAQQQGDTPAVSSSVTPNSGAGSPMDTPPAATPRSTTPGTPSTIE

TTPR

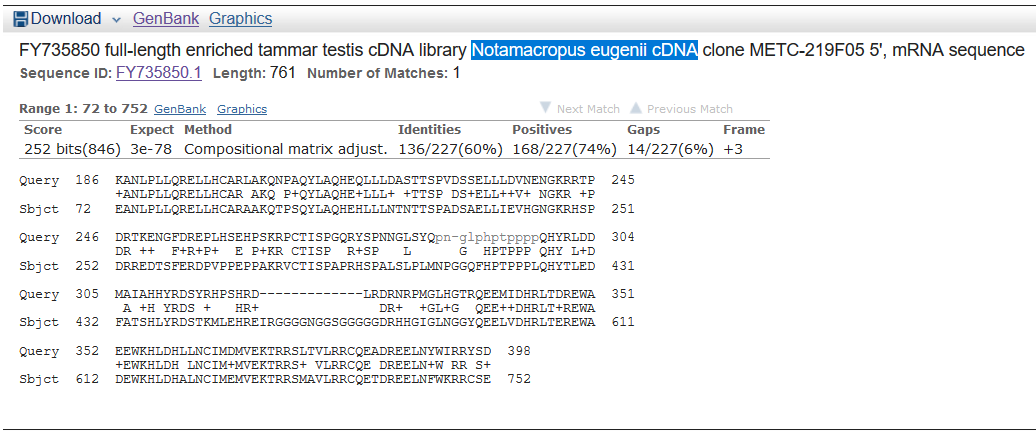


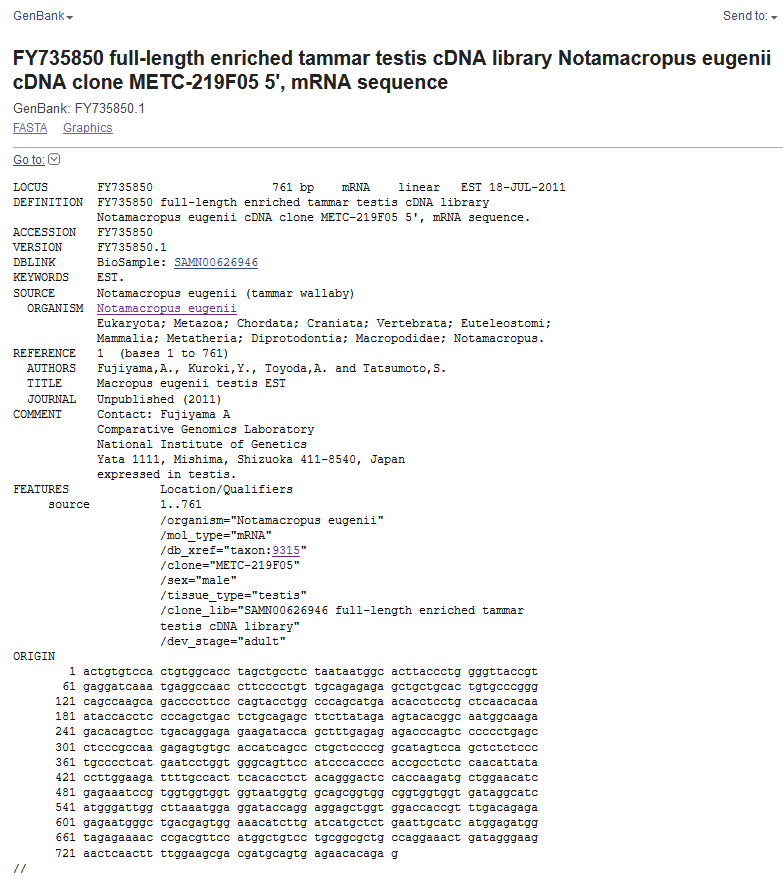
**Blast Output:**





**Selected homolog:**





**[Q3] Novel Protein Information**

Species: Notamacropus Eugenii (Tammar Wallaby)

Novel Protein Name: **CBFA2T1**

**Novel Protein Translated.**

>EMBOSS\_001\_Frame 1

TVSTVAPSCL\*\*WHLPWGYREDQMRPTFPCCRESCCTVPGQPSRPLPSTWPSMNTSCSTQ

IPPPQLTLQSFL\*KYTAMARDTVLTGEKIPALRETQSPLSLPPRECAPSALLPGIVQLSP

CPS\*ILVGSSIPPHRLSNIIPWKILPLHTSTGTPPRCWNIEKSVVVVVMVAAVAVVVIGI

MGLA\*MEDTRRSWWTTV\*QRENGLTSGNILIML\*IASWRW\*RKPDVPWLSCGAARKLIGK

NSTFGSDDAVRTQX

>EMBOSS\_001\_Frame 2

LCPLWHLAASNNGTYPGVTVRIK\*GQPSPVAERAAALCPGSQADPFPVPGPA\*TPPAQHK

YHLPS\*LCRASYRSTRQWQETQS\*QERRYQL\*ERPSPP\*ASRQESVHHQPCSPA\*SSSLP

APHESWWAVPSHPTASPTLYLGRFCHFTPLQGLHQDAGTSRNPWWWW\*WWQRWRWW\*\*AS

WDWLKWRIPGGAGGPPFDRERMG\*RVETS\*SCSELHHGDGRENPTFHGCPAALPGN\*\*GR

TQLLEATMQ\*EHRX

**>EMBOSS\_001\_Frame 3 CBFA2T1 Predicted Sequence (from EST)**

CVHCGT\*LPLIMALTLGLP\*GSNEANLPLLQRELLHCARAAKQTPSQYLAQHEHLLLNTN

TTSPADSAELLIEVHGNGKRHSPDRREDTSFERDPVPPEPPAKRVCTISPAPRHSPALSL

PLMNPGGQFHPTPPPLQHYTLEDFATSHLYRDSTKMLEHREIRGGGGNGGSGGGGGDRHH

GIGLNGGYQEELVDHRLTEREWADEWKHLDHALNCIMEMVEKTRRSMAVLRRCQETDREE

LNFWKRRCSENTE

>EMBOSS\_001\_Frame 4

LCSHCIVASKS\*VLPYQFPGSAAGQPWNVGFSLPSP\*CNSEHDQDVSTRQPILSLSNGGP

PAPPGILHLSQSHDAYHHHRHRCHHYHHHHGFLDVPASWWSPCRGVKWQNLPRYNVGEAV

GWDGTAHQDS\*GAGRELDYAGEQG\*WCTLSWREAQGGLGLSQSWYLLSCQDCVSCHCRVL

L\*EALQSQLGRWYLC\*AGGVHAGPGTGKGSAWLPGHSAAALSATGEGWPHLILTVTPG\*V

PLLEAARCHSGHS

>EMBOSS\_001\_Frame 5

SVFSLHRRFQKLSSSLSVSWQRRRTAMERRVFSTISMMQFRA\*SRCFHSSAHSLSVKRWS

TSSSWYPPFKPIP\*CLSPPPPPLPPLPPPPRISRCSSILVESL\*RCEVAKSSKV\*CWRGG

GVGWNCPPGFMRGRERAGLCRGAGLMVHTLLAGGSGGTGSLSKLVSSLLSGLCLLPLPCT

SIRSSAESAGEVVFVLSRRCSCWARYWEGVCLAARAQCSSSLCNRGRLASFDPHGNPRVS

AIIRGS\*VPQWTQX

>EMBOSS\_001\_Frame 6

LCVLTASSLPKVEFFPISFLAAPQDSHGTSGFLYHLHDAIQSMIKMFPLVSPFSLCQTVV

HQLLLVSSI\*ANPMMPITTTATAATITTTTTDFSMFQHLGGVPVEV\*SGKIFQGIMLERR

WGGMELPTRIHEGQGESWTMPGSRADGAHSLGGRLRGDWVSLKAGIFSPVRTVSLAIAVY

FYKKLCRVSWGGGICVEQEVFMLGQVLGRGLLGCPGTVQQLSLQQGKVGLI\*SSR\*PQGK

CHY\*RQLGATVDTV

**[Q4]Proof of Novelty**

**>EMBOSS\_001\_Frame 3 CBFA2T1 Predicted Sequence (from EST)**

GSNEANLPLLQRELLHCARAAKQTPSQYLAQHEHLLLNTNTTSPADSAELLIEVHGNGKR

HSPDRREDTSFERDPVPPEPPAKRVCTISPAPRHSPALSLPLMNPGGQFHPTPPPLQHYT

LEDFATSHLYRDSTKMLEHREIRGGGGNGGSGGGGGDRHHGIGLNGGYQEELVDHRLTER

EWADEWKHLDHALNCIMEMVEKTRRSMAVLRRCQETDREELNFWKRRCSENTE

