

1. **1.25 pts.** There are two attached files: znf214_mrna.txt and znf214_genomic.txt. Use Splign to find the mRNA and CDS coordinates in the genomic DNA. Load the genomic DNA into Sequin. Annotate the mRNA and CDS locations in the genomic DNA using Sequin. Export the results in GenBank format. **Hint:** When annotating CDS locations in Sequin, you can enter all the coding exons at once.

- 0.5 pts: Correct GenBank format
- 0.5 pts: Correct mRNA locations
- 0.25 pts: Correct CDS locations

Below are screenshots of my results. I also attached the GeneBank File to my submission.

NCBI Splign

HOME SEARCH SITE MAP Overview Online Download Documentation Contacts

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	Homo(+)	chromosome:GRCh38:11:6998718:7020968:-1(+)	614-21651	100.00	100.00	100.00	0.00	0.00

Graphics|Text

#	Query	Subject	Idty	Len	Q.Start	Q.Fin	S.Start	S.Fin	Type	Details
+1	Homo	chromosome:GRCh38:11	1	283	1	283	614	896	<exon>GC	M283
+1	Homo	chromosome:GRCh38:11	1	147	284	430	18114	18260	AG<exon>GT	M147
+1	Homo	chromosome:GRCh38:11	1	2238	431	2668	19414	21651	AG<exon>	M2238

Help: for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

Model 1

Coverage	100.00%	CDS	0.00%	Mismatches and indels	0
Overall	100.00%	In-frame	0.00%	Exons (min/max/ave), bp	147 / 2238 / 889
Exon	100.00%	Primary transcript	2668 bp	Introns (min/max/ave), bp	1154 / 17218 / 9186

Homo (+) sapiens zinc finger protein 214 (ZNF214), mRNA

chromosome:GRCh38:11:6998718:7020968:-1 (+)

614 21651

Segments Alignment

1 2 3

```

1  AAGACCTGGGACTTCGGTTTCGGTCCAGCCGGGCTGCGGCCATTGTTTGTGTGGACTGAGTGTTTGGCC
614 AAGACCTGGGACTTCGGTTTCGGTCCAGCCGGGCTGCGGCCATTGTTTGTGTGGACTGAGTGTTTGGCC

71  ATCCCGGTCCACTCTCACAGGCTCCGTTAAGTGACATGAACTCTCAGGAGGACTGAGCCAGAAGCGGACA
684 ATCCCGGTCCACTCTCACAGGCTCCGTTAAGTGACATGAACTCTCAGGAGGACTGAGCCAGAAGCGGACA

141 GGGCAAGACGAGCTGTGCTTGAAGGAAGAGGGGCAGAAAATCGAGGGTCAGGGACTGAGAAGCTACTCCG
754 GGGCAAGACGAGCTGTGCTTGAAGGAAGAGGGGCAGAAAATCGAGGGTCAGGGACTGAGAAGCTACTCCG

211 GTTTAGAAACCCAGAGACACCCGTGTAGATGGGTACATCACGGCTTCTCTCCACGTTAAGACTTAATA
824 GTTTAGAAACCCAGAGACACCCGTGTAGATGGGTACATCACGGCTTCTCTCCACGTTAAGACTTAATA
  
```

Sequin File Edit Search Options Misc Annotate

chromosome:GRCh38:11:6

Target Sequence

Format Mode Style

mRNA: mRNA

```

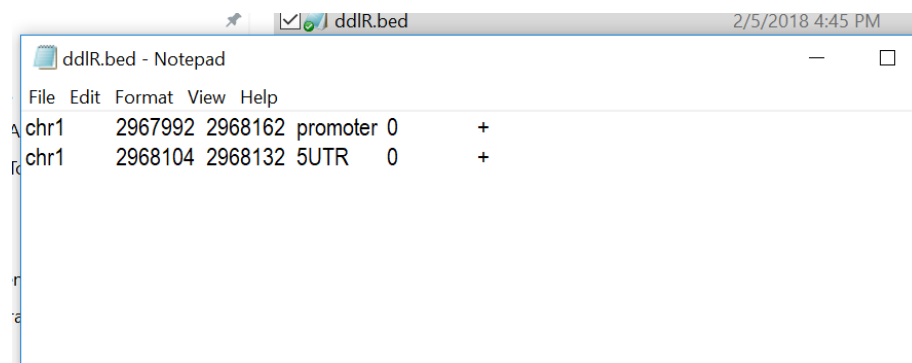
LOCUS       chromosome:GRCh38:11:6998718:7020968:>22251 bp    DNA Linear
PRI05-FEB-2018
DEFINITION   Homo sapiens.
ACCESSION
VERSION
KEYWORDS
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
              cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa;
              Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata;
              Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii;
              Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria;
              Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini;
              Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo;
              Homo sapiens.
REFERENCE    1 (bases 1 to 22251)
AUTHORS      Garcia,J.P.
TITLE        znf214 annotation
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 22251)
AUTHORS      Garcia,J.P.
TITLE        Direct Submission
JOURNAL      Submitted (05-FEB-2018) Biotechnology, Johns Hopkins University,
              409 Stoneridge Ct., Grand Junction, CO 81507, USA
FEATURES             Location/Qualifiers
     source           1..22251
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
     mRNA             join(614..896,18114..18260,19414..21651)
     CDS              join(18034..18249,19413..21107)
                     /codon_start=1
                     /product="CDS"
                     /translation="-
                     IRLATVLC*MSLSLSIFSLRFLFPRKPDLPDGSNI*RCOVYFVYMGVEIPGFFSKK
                     TLQGGHVGLHKKCNWNSYKSQEEKFRYLEVENFSVWQGWNNAGAQMYENQNVGETV
                     QGTDSKDLTQDRSQCEWLILSTQVPGVGNVELTFESKSLANLKYKNFMPWQSLETK
                     TTQDYGREIYMSGSHGFQGGAYRLGISAKNLSMEKEQKLIQHSYIPVEERALPQYVGV
                     ICQEDLLRDSMEEKYCGCNKCKGIYYWNSACVFKRNQPGENLCQCSI CKACFSQASD
                     LYRHPNNHIGKKLYGCDVGNFHQSSGVHFHQRVHIGVPPYSCNACGKSFQISSLH
                     NQRVHTEEFYKIECDKLSRNSLLHIHQRLHIGKPFKCNQCGKSFNASSVLHVHQ
                     RVHTGEKPYKDCGCGKFSQSSNLRIHQLVHTGEKSKCEDCGKGTQRSNLQIHQRV
                     HTGEKPYKCDGCGKDFSHSSDLRIHQRVHTGEKPYTCPECGKGFSSKSLHTRQVHT
                     GEKPYKCECGKGFSSQSHLLIHQRVHTGEKPYKCHDCGKGFSSSNLHIHQRVHTGE
                     KPVQCAKCGKGFSSSALRIHQRVHAGEKPYKCREVYKGFQHNHSHLHNNHRRGNL"
BASE COUNT      6399 a   3868 c   4118 g   7866 t
ORIGIN
      1  ttctgtgtgtggt cccaggacac tgaggagggg ctccttctca tggtcttatt taagctccac
     61  tatagatttc cctttctctc ctaggatctc cagagagagc tagagattta cagagagagc
  
```

2. **1.25 pts.** Create a BED6 file with 2 lines based on the attached paper (Takenaka_et_al-2015-FEBS_Journal.pdf). Figure 3 shows the location of transcription factor DdIR binding to the promoter region of the ddIR-ddl operon in *Brevibacillus brevis*. The chromosomal location of the ddIR CDS is 2968133..2969623. The zero-based BED6 file should contain the location information of two genomic regions:

a. 0.625 pts: The region bound by the DdIR transcription factor, which we will call the *promoter*. It is 170 bp in length, begins 140 nucleotides upstream from the start codon, and ends 29 nucleotides downstream from the start codon.

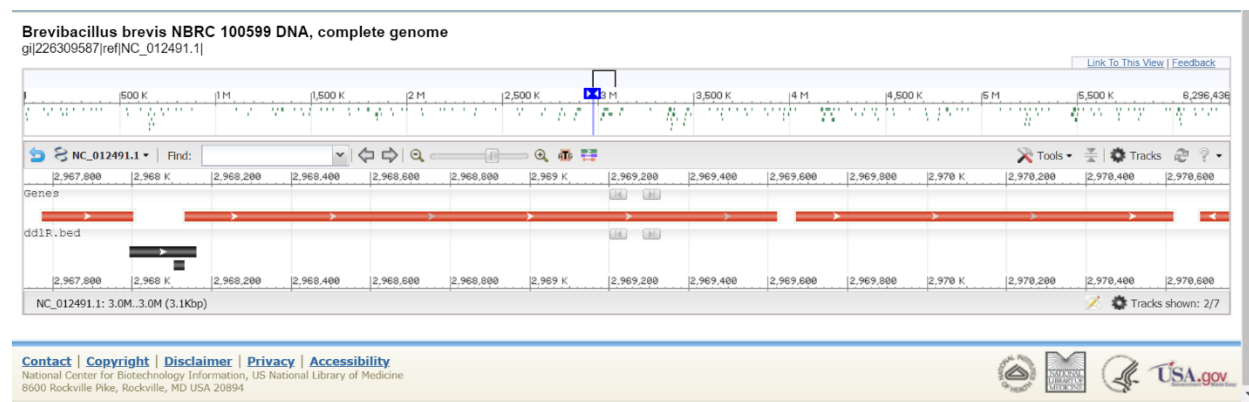
b. 0.625 pts: The 5' UTR, noting that the transcription start site, as predicted by BPROM, begins 28 nucleotides upstream from the start codon. The 5' UTR is defined as the region from the transcription start site through the nucleotide that immediately precedes the start codon.

Here is a screenshot of my BED6 file and I also attached it.



3. 1.25 pts. Submit a screenshot of the BED6 from Problem 2. Using the NCBI Genome Browser for *Brevibacillus brevis* [NBRC 100599](#), load your BED6 file. Take a screenshot showing the entire promoter, 5' UTR region, and CDS of ddlR. Be sure to zoom in so that these regions take up a majority of the shot.

See above for the screenshot of the BED6 file. Below is a screenshot of NCBI Genome Browser with BED6 file uploaded. The promoter and 5'UTR regions are in black and overlap the CDS region above them in red.



4. 1.25 pts. Use the web-based Biomart in Ensembl to create a dataset and save it as a TSV, CSV, or XLS file. Use the following parameters to make the dataset:

Dataset:

Ensembl Genes 91

Mouse genes (GRCm38.p5)

Filters:

Chromosome 11

Band E2 only

Transcript count ≥ 7

Limit to genes with RefSeq protein IDs only

Attributes:

Default attributes

Add "RefSeq Protein ID"

Get all the results, export the results to a file, and submit the file.

The file is attached to my submission and below is a screenshot of my results.

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

New | Count | Results

URL | XML | Perl | Help

Dataset 64 / 53946 Genes

Mouse genes (GRCm38.p5)

Filters

Chromosome/scaffold: 11

Band Start: E2

Band End: E2

Transcript count ≥ 7

With RefSeq peptide ID(s): Only

Attributes

Gene stable ID

Transcript stable ID

RefSeq peptide ID

Dataset

None Selected]

Export all results to

File

TSV

Unique

Email notification to

View

10

rows as

HTML

Unique results only

Gene stable ID	Transcript stable ID	RefSeq peptide ID
ENSMUSG00000020758	ENSMUST00000068981	NP_598424
ENSMUSG00000020758	ENSMUST00000169928	NP_001005608
ENSMUSG00000020770	ENSMUST00000021116	NP_766157
ENSMUSG00000020770	ENSMUST00000106452	NP_001272935
ENSMUSG00000057948	ENSMUST00000075036	NP_001009573
ENSMUSG00000025137	ENSMUST00000026129	NP_077191
ENSMUSG00000025137	ENSMUST00000106188	NP_001334544
ENSMUSG00000025138	ENSMUST00000080202	NP_694696
ENSMUSG00000033857	ENSMUST00000135383	NP_766161
ENSMUSG00000020773	ENSMUST00000106441	NP_766158