Khan Inan
10.4 project

1. For sequence 1, I found that the sequence was a direct 100% match with a section of chromosome 4. However when the sequence is aligned and compared to chr8 or chr5 then there are gaps that appear in the sequence indicating splice sites/fusions

CHR8:

cDNA YourSeq

```
aagcggcggg gctcactcct gagcgcCCTG cCCGCAGGTA CATGATcATG  50
CGGGAgTGCT GGCATGCcGc GCCCTCcCAG AGgCCCACCT TCAAGCAG
```

we can see several instances of lower case nucleotides interrupting the sequence

CHR5:

cDNA YourSeq

```
aagcggcggg gctcactcct gagcgccctg ccCGCAGGTA CatGaTcATG  50
CGgGAGTGCT GGCAtGCcGC GCCCTCCCAG AGGCCcACCT TCAAGCAG
```

The same case is seen here, where the several lower case gaps in the sequence indicated the fusions of incomplete strands of cDNA

2. For sequence 2, it was a 100% match for chr4,chr8,chr10,chr2 and chr5 but in every instance it was incomplete, or there were additional nucleotides present that didn't belong

CHR4:

cDNA YourSeq

```
cagaGTGCTT GTGTGGGTAC CTGTGCTCCT GGCCTGGTCG TGTAGAAGAT  50
GAGGTGTGGG GCGGGCCTTC TGGGGCACAG CCTGGGCACA GAGGTGGC
```

we can see the fusion of "caga" at the start

CHR8:

cDNA YourSeq

```
cagagtgctt gtgtgggtac ctgtgctcct ggcctggtcg tgtagaagat  50
gaggtgtggg gcgggccttc tggGGCACAG CCTGGGCACA GAGGTGGC
```

Here, only the end of the sequence is a match, and the rest is fused in addition
CHR10:

cDNA YourSeq

```
cagagtgctt gtgtgggtac ctgtgctcct ggcctggtcg tgtagaagat  50
gaggtgtggg gcGGGCCTTC TGGGGCACAG CCTGggcaca gaggtggc
```

We can see that the match occurs in the middle of the sequence. This suggests that the fusion sites are at the G and G, where the start has an added sequence and the end has an added sequence that doesn't belong

CHR2:

cDNA YourSeq

```
cagagtgctt gtgtgggtac ctgtgctcct ggcctggtcg tgtagaagat  50
gaggtgtggg gcgggccttc tggggcACAG CCTGGGCACA GAGGTGgc
```

This is the same fusion case as with CHR10, where the sequence is located in the middle

CHR5:

cDNA YourSeq

```
cagagtgctt gtgtgggtaC CTGTGCTCCT GGCCTGGTCg tgtagaagat  50
gaggtgtggg gcgggccttc tggggcacag cctgggcaca gaggtggc
```

3. For the third sequence I found that the sequence was a direct match with chr4, but with chr17,chr15,chr11,chr13,and chr 15 there is a lot of evidence of fusions and splices

CHR4:

cDNA YourSeq

```
GCCCTTTCCT CCAGGAGAGA GCCTTGAACT CTGCCAGCAC CTCGCTTCCC  50
ACAAGCTGTC CAGGCAGTGA GCCAGTGCCC ACCCATCAGC AGGGGCAG
```

Here we can see a direct match

CHR17:

cDNA YourSeq

```
gccctttcct ccaggAGAGA GCCTTGAACT CTGCCAGcAC CTCgcttccc  50
acaagctgtc caggcagtga gccagtgccc acccatcagc aggggcag
```

For chr17 we can see a lower case c interrupting the sequence as well as many light blue letters. These are all splice sites and locations were multiple smaller sections of cDNA joined together

CHR15:

cDNA YourSeq

```
gccctttcct ccaggagaga gccttgaact ctgccagcac ctcgcttccc  50
acaagctgtc caggcagtgA GCCAGTGCCC ACCCATCaGC AGGGGcag
```

Here we see a lower case “a” interrupt the sequence

CHR11:

cDNA YourSeq

```
gccctttcct ccagGAGAGA GCCTTGAACT CTGCCAGCac ctcgcttccc  50
acaagctgtc caggcagtga gccagtgccc acccatcagc aggggcag
```

here we see that there are light blue letters in the midst of the sequence. This tells us that there are two sequences from different parts of the chromosome that were joined together at this point.

CHR13:

cDNA YourSeq

```
gccctttccT CCAGGAGAGA GCCTTGAACT ctgccagcac ctcgcttccc  50
acaagctgtc caggcagtga gccagtgccc acccatcagc aggggcag
```

The sequence is in the middle of the chromosome, so essentially there if “fluff” on either end of our sequence that doesn’t belong

CHR15:

cDNA YourSeq

```
gccctttcct ccaggagaga gccttgaact ctgccagcac ctcgcttccc  50
acaagctgtc caggcagtga gccagtgcCC ACCCATCAGC AGGGGCAG
```

The sequence is in the middle of the chromosome

The cancer associated with these types of proteins is most likely to be some kind of Lymphoma because I saw that there are several instances of in-frame mutations that are synonymous with lymphoma related tumors

An example of a gene that should not be expressed in the cancer cells but instead are present is chromosome 17. This chromosome is responsible for making instructions for the retinoic acid receptor, that essentially controls the activity of specific genes. Whereas the other chromosomal deficiencies make sense since they all provide instructions for making proteins, the alteration of chromosome 17 in particular is probably likely to result in more benign cancers or tumors