Transcriptional Regulatory Elements and Transcription Factors that Control Kaposi's Sarcoma Human Virus Genes.

Manjula Kasoji, Georgia Institute of Technology
Dr. Bino John, University of Pittsburgh

Kaposi's sarcoma-associated herpesvirus (**KSHV**)

- A virus that causes Kaposi's sarcoma, a cancer
- It was originally described by Moritz Kaposi, an Austro-Hungarian dermatologist
- Commonly occurs in AIDS patients
- Fourth most common cancer caused by an infectious agent worldwide following gastric, cervical, and hepatic cancers.

KSHV

- causes abnormal tissue growth under the skin and leads to lesions.
- usually not life threatening unless it reaches to the internal organs
- Functional studies of KSHV indicate that
  - its genes interfere with established tumor suppressor pathways
  - modify the host cellular environment

Goal

- The goal of this experiment is to identify the transcriptional regulatory sites and factors that are involved in the transcriptional regulation of the viral genome.

Methods

1. Find a good web program that predicts transcription factors based on a database of known transcription factor motifs.
   - TRANSFAC*
   - JASPER
   - Chosen web program: CLOVER
   - http://biowulf.bu.edu/MotifViz/

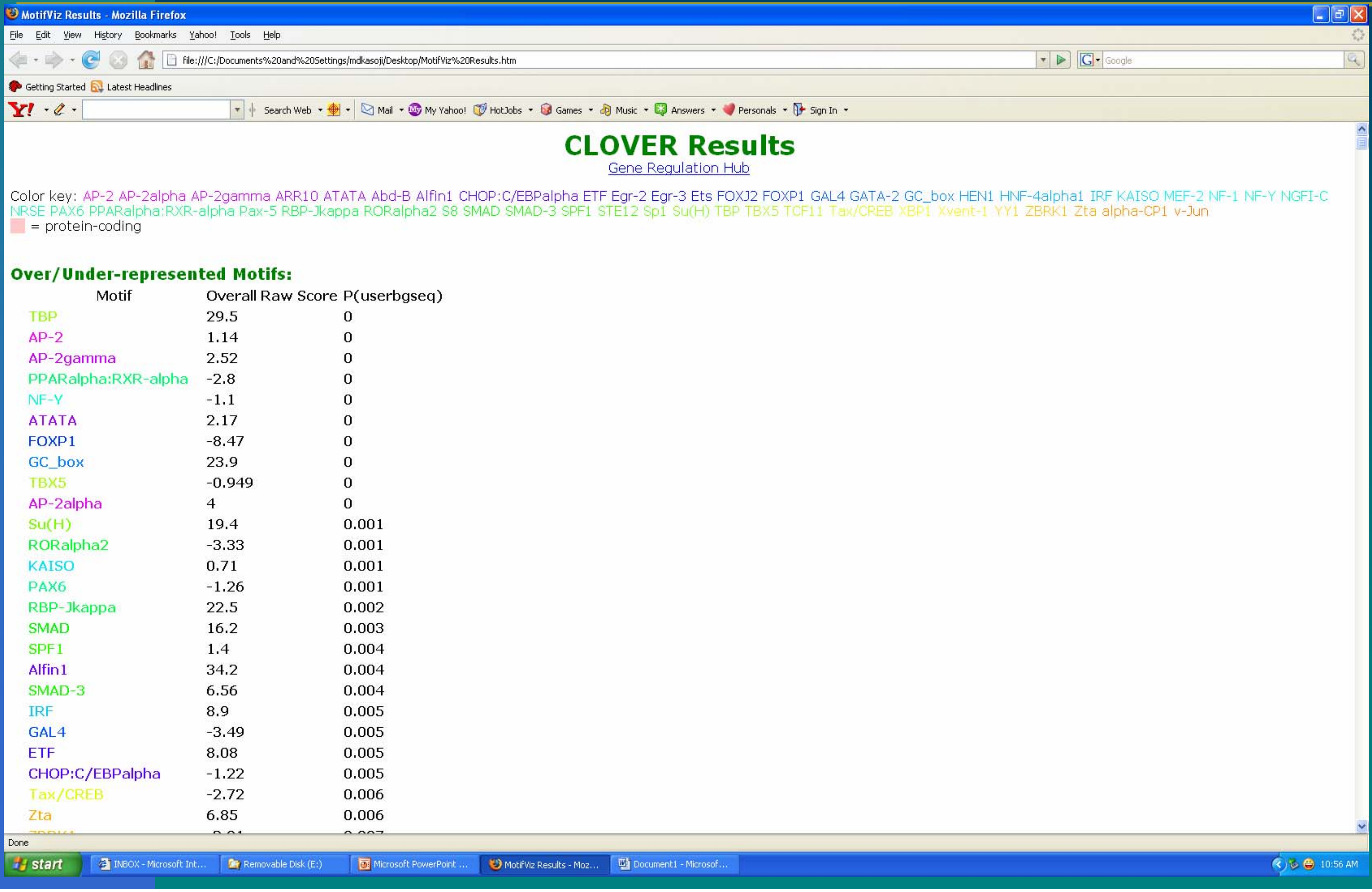
CLOVER

- CLOVER predicts transcription factors based on a database of known transcription factors and motifs.
- The database used was TRANSFAC.
- Takes in 3 files:
  - Target sequences (upstream/promoter regions of all protein coding genes)
  - Background sequences
  - TRANSFAC (.dat) file

Methods cont…

2. Obtain genome and protein coding genes
    - http://www.ncbi.nlm.nih.gov/
3. Obtain 1000 nucleotides (nt) upstream of all protein coding genes
4. generate random sequences of 2000 nt from same genome. These are used as background sequences.
5. Obtain TRANSFAC data file (.dat)

CLOVER Output



CLOVER Results

Gene Regulation Hub

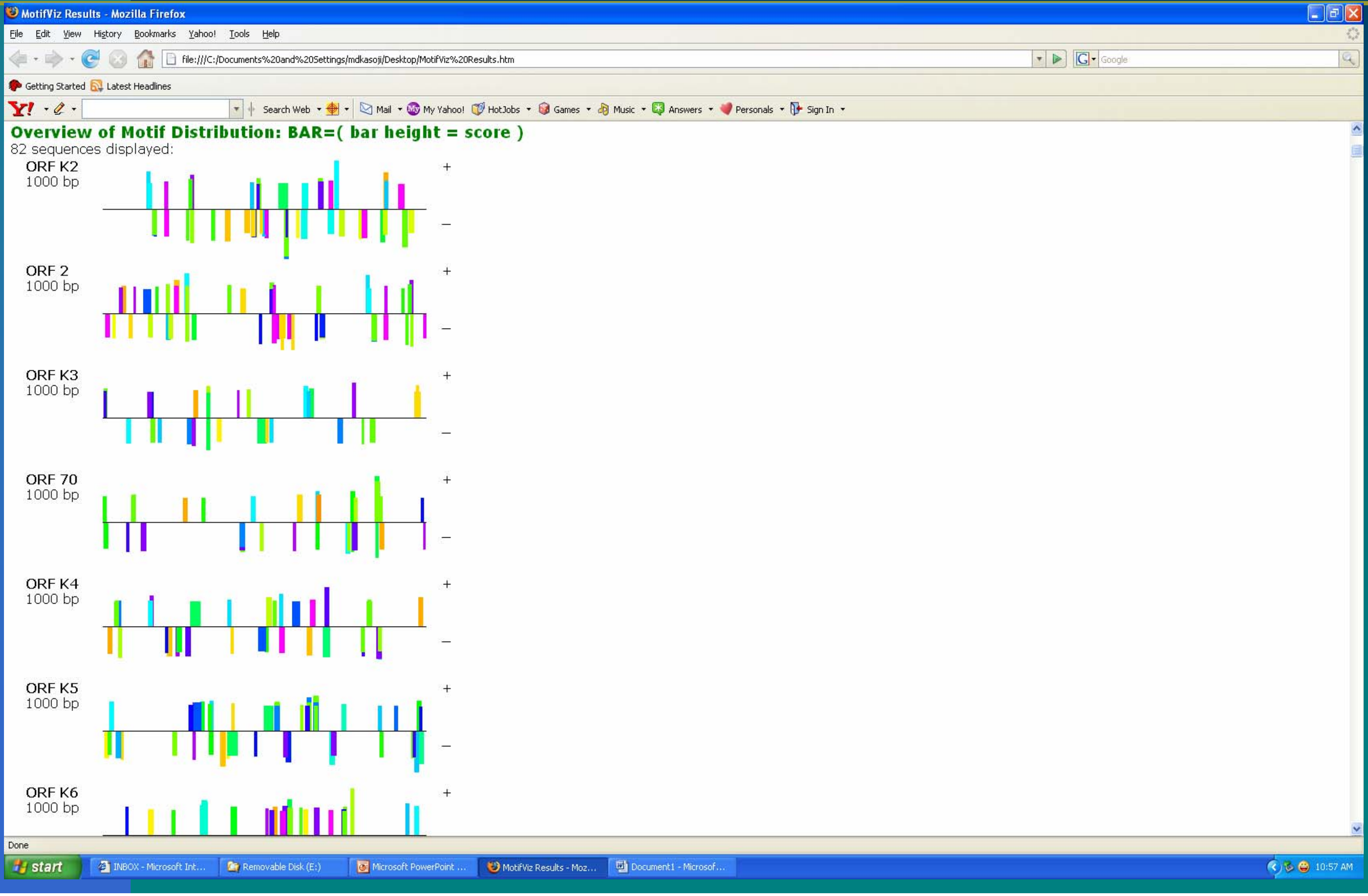
Color key: AP-2 AP-2alpha AP-2gamma ARR10 ATATA Abd-B Alfin1 CHOP:C/EBPalpha ETF Egr-2 Egr-3 Ets FOXJ2 FOXP1 GAL4 GATA-2 GC_box HEN1 HNF-4alpha1 IRF KAISO MEF-2 NF-1 NF-Y NGFI-C NRSE PAX6 PPARalpha:RXR-alpha Pax-5 RBP-Jkappa RORalpha2 S8 SMAD SMAD-3 SPF1 STE12 Sp1 Su(H) TBP TBX5 TCF11 Tax/CREB XBP1 Xvent-1 YY1 ZBRK1 Zta alpha-CP1 v-Jun

🟥 = protein-coding

Over/Under-represented Motifs:

Motif	Overall Raw Score	P(userbgseq)
TBP	29.5	0
AP-2	1.14	0
AP-2gamma	2.52	0
PPARalpha:RXR-alpha	-2.8	0
NF-Y	-1.1	0
ATATA	2.17	0
FOXP1	-8.47	0
GC_box	23.9	0
TBX5	-0.949	0
AP-2alpha	4	0
Su(H)	19.4	0.001
RORalpha2	-3.33	0.001
KAISO	0.71	0.001
PAX6	-1.26	0.001
RBP-Jkappa	22.5	0.002
SMAD	16.2	0.003
SPF1	1.4	0.004
Alfin1	34.2	0.004
SMAD-3	6.56	0.004
IRF	8.9	0.005
GAL4	-3.49	0.005
ETF	8.08	0.005
CHOP:C/EBPalpha	-1.22	0.005
Tax/CREB	-2.72	0.006
Zta	6.85	0.006

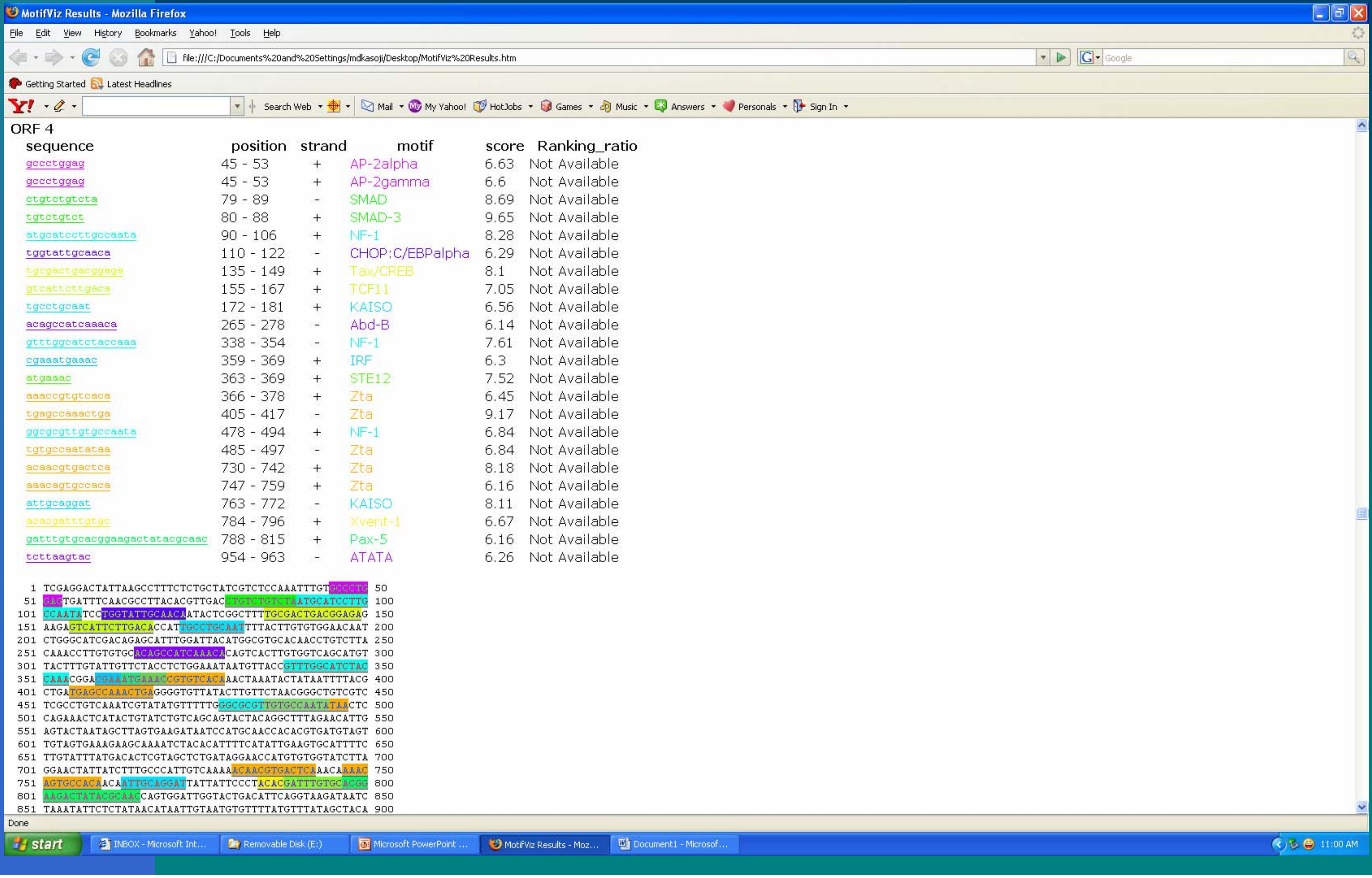
CLOVER Output



Overview of Motif Distribution: BAR=(bar height = score)

82 sequences displayed:

CLOVER Output



ORF 4

sequence	position	strand	motif	score	Ranking_ratio
gccctggag	45 - 53	+	AP-2alpha	6.63	Not Available
gccctggag	45 - 53	+	AP-2gamma	6.6	Not Available
ctgtctgtcta	79 - 89	-	SMAD	8.69	Not Available
tgtctgtct	80 - 88	+	SMAD-3	9.65	Not Available
atgcatccttgccaata	90 - 106	+	NF-1	8.28	Not Available
tggtattgcaaca	110 - 122	-	CHOP:C/EBPalpha	6.29	Not Available
tgcgactgacggaga	135 - 149	+	Tax/CREB	8.1	Not Available
gtcattcttgaca	155 - 167	+	TCF11	7.05	Not Available
tgcctgcaat	172 - 181	+	KAISO	6.56	Not Available
acagccatcaaaca	265 - 278	-	Abd-B	6.14	Not Available
gtttggcatctaccaaa	338 - 354	-	NF-1	7.61	Not Available
cgaaatgaaac	359 - 369	+	IRF	6.3	Not Available
atgaaac	363 - 369	+	STE12	7.52	Not Available
aaaccgtgtcaca	366 - 378	+	Zta	6.45	Not Available
tgagccaaactga	405 - 417	-	Zta	9.17	Not Available
ggcgcgttgtgccaata	478 - 494	+	NF-1	6.84	Not Available
tgtgccaatataa	485 - 497	-	Zta	6.84	Not Available
acaacgtgactca	730 - 742	+	Zta	8.18	Not Available
aaacagtgccaca	747 - 759	+	Zta	6.16	Not Available
attgcaggat	763 - 772	-	KAISO	8.11	Not Available
acacgatttgtgc	784 - 796	+	Xvent-1	6.67	Not Available
gatttgtgcacggaagactatacgcaac	788 - 815	+	Pax-5	6.16	Not Available
tcttaagtac	954 - 963	-	ATATA	6.26	Not Available

```
  1 TCGAGGACTATTAAGCCTTTCTCTGCTATCGTCTCCAAATTTGTGCCCTG  50
 51 GAGTGATTTCAACGCCTTACACGTTGACCTGTCTGTCTAATGCATCCTTG 100
101 CCAATATCCTGGTATTGCAACAATACTCGGCTTTTGCGACTGACGGAGAG 150
151 AAGAGTCATTCTTGACACCATTGCCTGCAATTTTACTTGTGTGGAACAAT 200
201 CTGGGCATCGACAGAGCATTTGGATTACATGGCGTGCACAACCTGTCTTA 250
251 CAAACCTTGTGTGCACAGCCATCAAACACAGTCACTTGTGGTCAGCATGT 300
301 TACTTTGTATTGTTCTACCTCTGGAAATAATGTTACCGTTTGGCATCTAC 350
351 CAAACGGACGAAATGAAACCGTGTCACAAACTAAATACTATAATTTTACG 400
401 CTGATGAGCCAAACTGAGGGGTGTTATACTTGTTCTAACGGGCTGTCGTC 450
451 TCGCCTGTCAAATCGTATATGTTTTTGGGCGCGTTGTGCCAATATAACTC 500
501 CAGAAACTCATACTGTATCTGTCAGCAGTACTACAGGCTTTAGAACATTG 550
551 AGTACTAATAGCTTAGTGAAGATAATCCATGCAACCACACGTGATGTAGT 600
601 TGTAGTGAAAGAAGCAAAATCTACACATTTTCATATTGAAGTGCATTTTC 650
651 TTGTATTTATGACACTCGTAGCTCTGATAGGAACCATGTGTGGTATCTTA 700
701 GGAACTATTATCTTTGCCCATTGTCAAAAACAACGTGACTCAAACAAAAC 750
751 AGTGCCACAACAATTGCAGGATTATTATTCCCTACACGATTTGTGCACGG 800
801 AAGACTATACGCAACCAGTGGATTGGTACTGACATTCAGGTAAGATAATC 850
851 TAAATATTCTCTATAACATAATTGTAATGTGTTTTATGTTTATAGCTACA 900
```

Conclusion

- With the identification of these transcription factors we hope to better understand the regulation and infection of the Kaposi virus.

Future Work

- Find out which transcription factors regulate miRNAs
  - miRNAs are a class of endogenous, small RNAs that are thought to negatively regulate protein production.
  - aberrant expression of miRNAs is linked to cancer and other diseases

Acknowledgements

- Dr. Bino John
- Dr. Shaun Mahoney
- Ankur Agarwal
- Andrej Savol