Determining the DNA features that contribute to X-specific recruitment of the dosage compensation complex in C. *elegans*

HAMD BILAL TAHIR

Location: The Ercan Lab of Developmental Genetics at NYU

Based on:

Cooperation between a hierarchical set of recruitment sites targets the X chromosome for dosage compensation.

Albritton SE, Kranz AL, Winterkorn LH, Street LA, Ercan S[1]

BACKGROUND

-Dosage Compensation is a natural process that equalizes the X chromosome gene expression between the two sexes in C. *elegans* (roundworm): males (one X chromosome) and hermaphrodites (two X chromosomes)

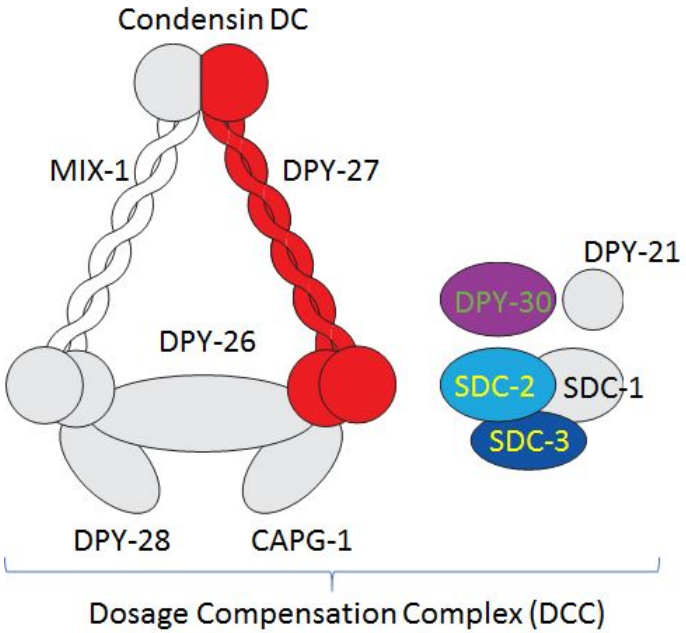
-In C. *elegans* hermaphrodites (that have both male and female characteristics), this process reduces both X chromosomes' transcription (expression) to about half of the initial so that the overall X-linked gene expression becomes equal in the somatic cells of both sexes

-This is critical because aberrations in chromosome count (having an extra chromosome or missing a chromosome) give rise to developmental problems or lethality in species (e.g. trisomy of chromosome 21,
aka Down's syndrome)

BACKGROUND

-Dosage compensation is achieved by a protein complex: the Dosage compensation complex (DCC)

-The DCC binds DNA in two step process. It first binds DNA at specific recruitment sites (Rex sites) on the X chromosome and then allows for the sequential DCCs to bind at other sites (spreading sites).



BACKGROUND

-Albritton et al. 2017 (Elife) found 12bp DNA motifs in C. *elegans* that should show strong binding to the Dosage Compensation Complex (DCC)

DNA Motif: A small DNA pattern that is widespread and has, or is conjectured to have, a biological significance.

-However, several supposedly strong motifs were not bound.
This meant that several of the recruitment sites with strong motif matches were not recruiting DCC at all.

-The Ercan lab made two hypotheses:

1. **A structural feature of the DNA sequences surrounding the motif is important for the ability of the motif to be recognized by the DCC recruiter protein (and hence recruit DCC)**

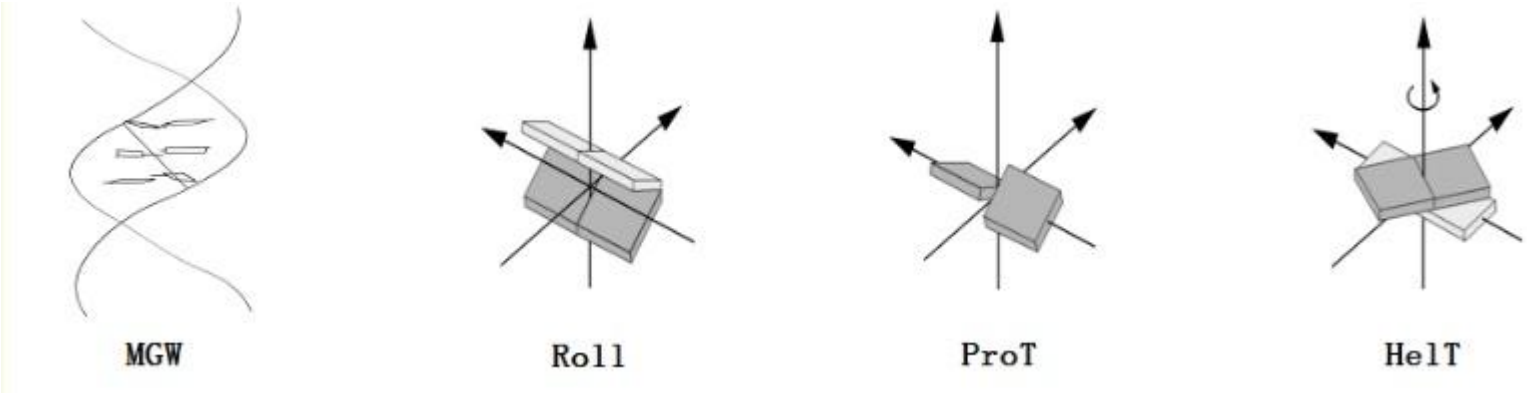
2. **The nucleosome occupancy of the DNA motifs is important the ability of the motif to be recognized by the DCC recruiter protein (and hence recruit DCC)**

BACKGROUND

-Nucleosome Occupancy: Wrapping of DNA at specific locations, around histone proteins to make “ball-like” nucleosomes, inhibiting the DNA from being exposed for potential binding to molecules.

BACKGROUND

-DNA features: Specific DNA configurations at base-pair level



BIOLOGICAL QUESTION

Why is the recruitment of the dosage compensation complex not seen at several of the the DNA motifs that were categorized as strong matches for the DCC recruitment?

GOAL

Choosing the first hypothesis: Analyze DCC bound and unbound DNA motifs to determine whether there are any DNA features that distinguish the bound motifs from the unbound ones.

Methods

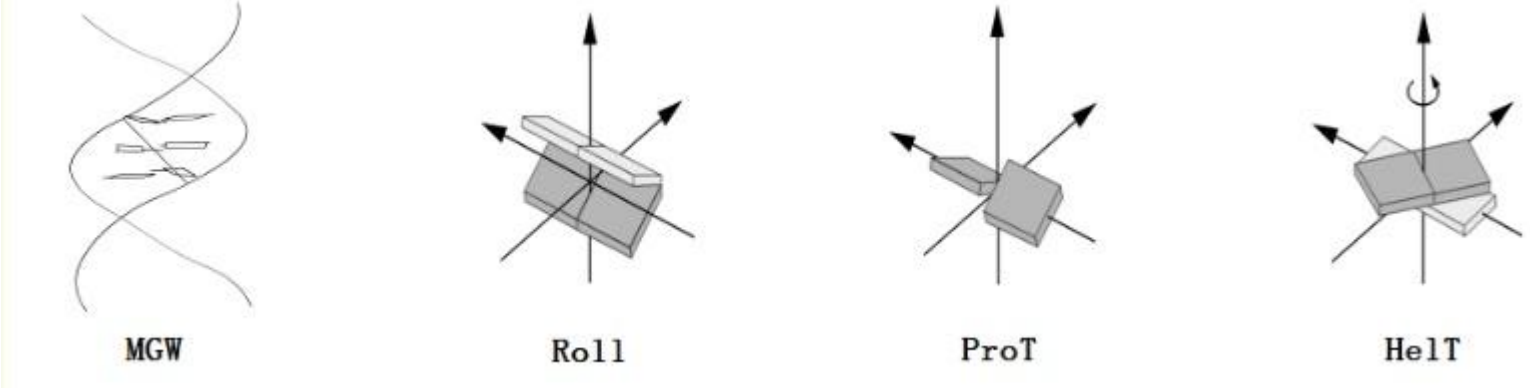
-Four distinct DNA shape features tested (courtesy of DNAshapeR package in R)

-Minor Groove Width (MGW): Width of the strand backbones that are closer together

-Propeller Twist (ProT): Bent structure induced by the rotation of one base with respect to the other in a base pair

-Roll (Roll): Rotation around the slide axis, making successive base pairs less parallel

-Helix Twist (HelT): Twisting of the DNA helix



Methods

-Data of Motifs, Rex Sites and Spreading Sites taken from Supplemental Albritton et al. 2017

-Intersection of the motifs with the Rex and Spreading data separately to find overlapping hits (Bound motifs)

-The rest of the motifs were (Unbound motifs)

Methods

-The input motifs have been ranked with scores from 1-10 by Albritton et al. 2017

-A cutoff of 7 was chosen for motifs

-200bp DNA windows were produced for the resulting Rex, Spreading and Unbound motifs that centered at the 12bp motifs

Methods

-Granges of those 200bp Rex, Spreading and Unbound windows were produced

-DNA sequences were extracted by overlapping the windows' coordinates with BSgenome C. *elegans*

-DNAshapeR package was used to produce the scores regarding each of the 4 features (MGW, ProT, HelT and Roll) separately, for the 200bp DNA windows

Granges: Represent and manipulate genomic intervals and variables defined along a genome

BSgenome: Contains full genome sequence

Methods

-Each base in a 200bp sequence was given a score through a pentamer sliding window method used by the DNAshapeR package

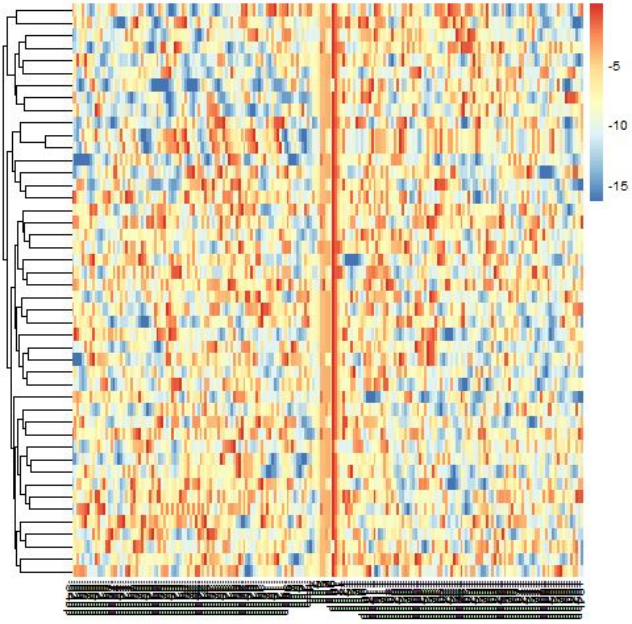
-DNA feature heatmaps of the Rex, Spreading and Unbound motifs were produced

-In each heatmap, the rows depict motifs and the columns depict the base-pairs from 1-200

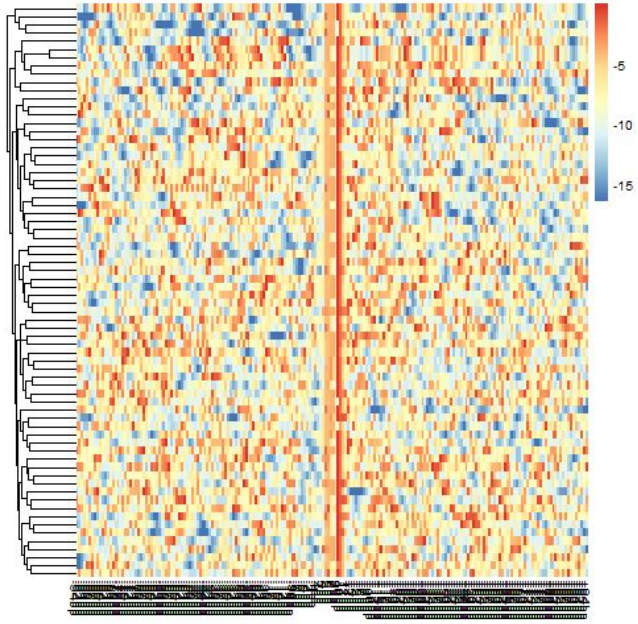
Results 200bp window (ProT)

-Rex vs Spreading vs Unbound motifs

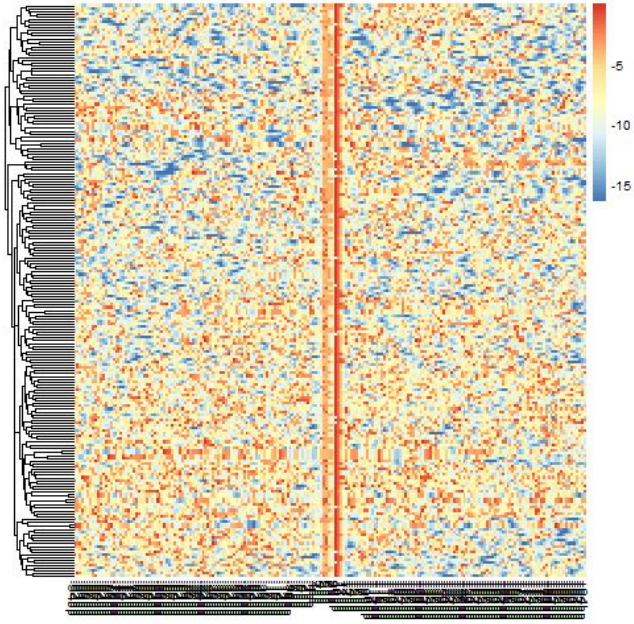
Rex



Spreading



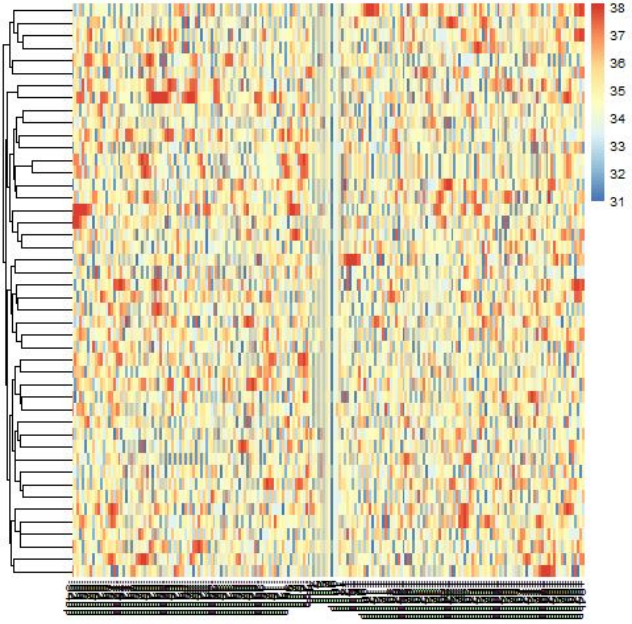
Unbound



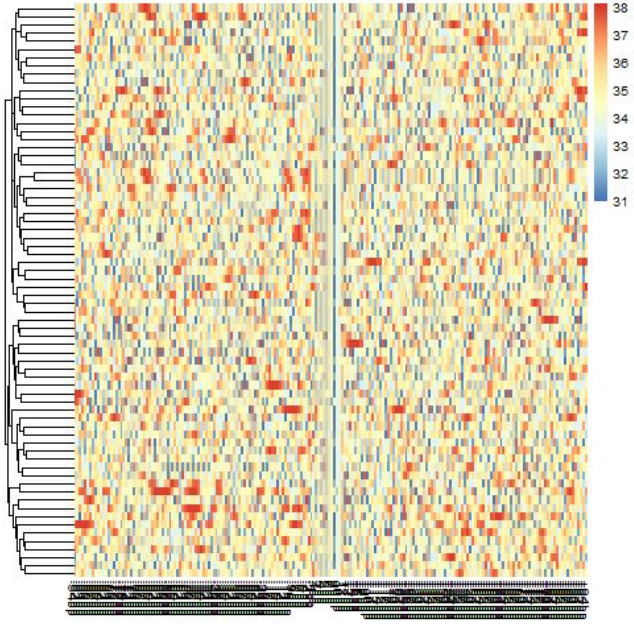
Results 200bp window (HelT)

-Rex vs Spreading vs Unbound motifs

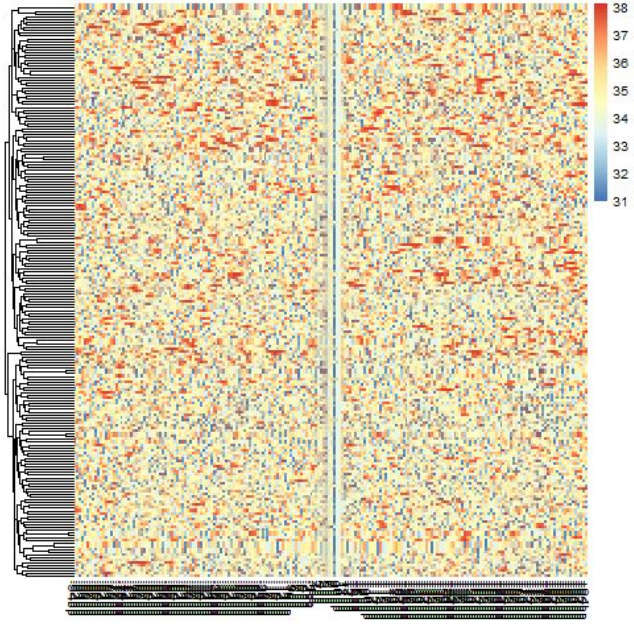
Rex



Spreading



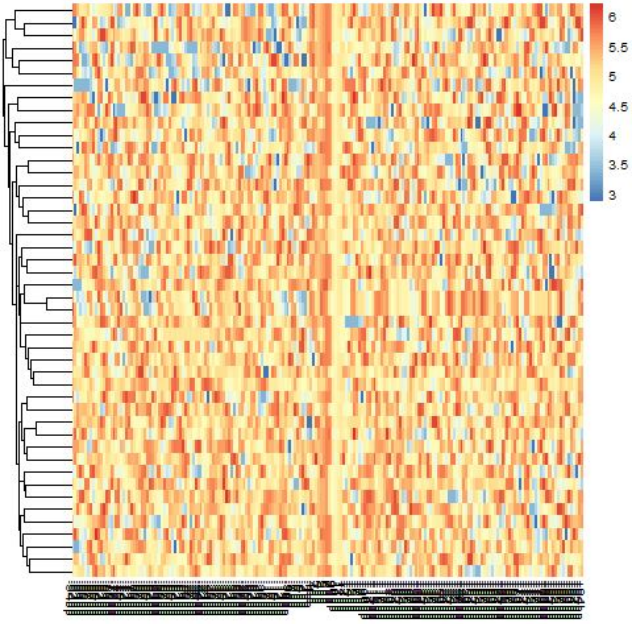
Unbound



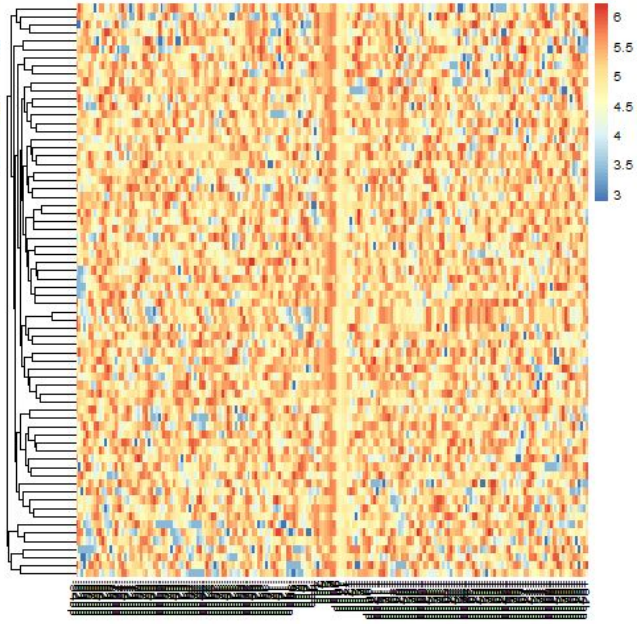
Results 200bp window (MGW)

-Rex vs Spreading vs Unbound motifs

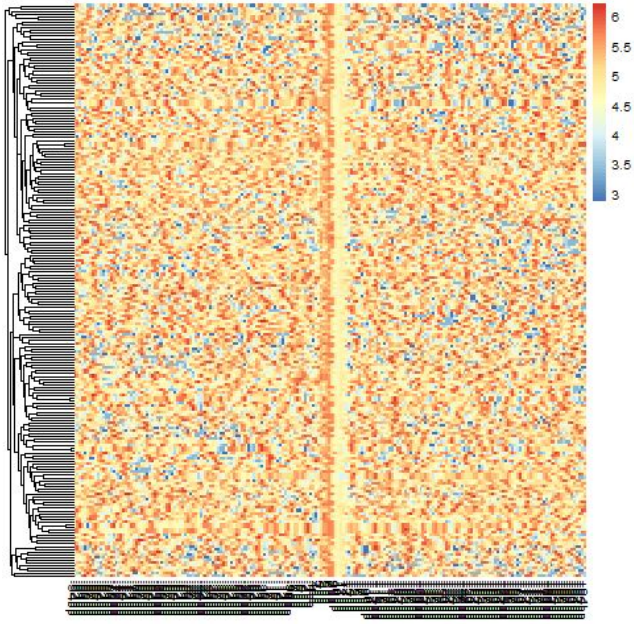
Rex



Spreading



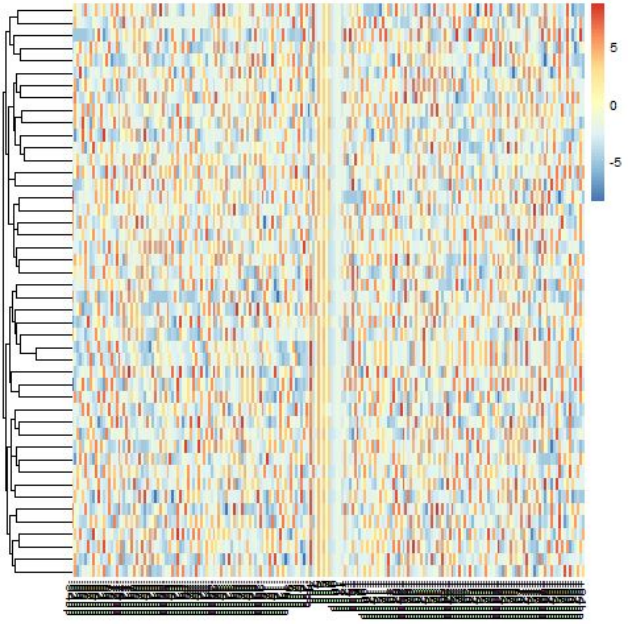
Unbound



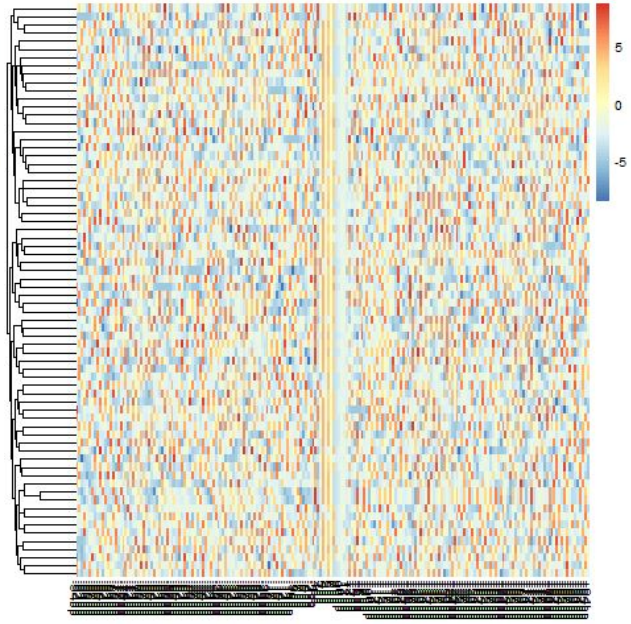
Results 200bp window (Roll)

-Rex vs Spreading vs Unbound motifs

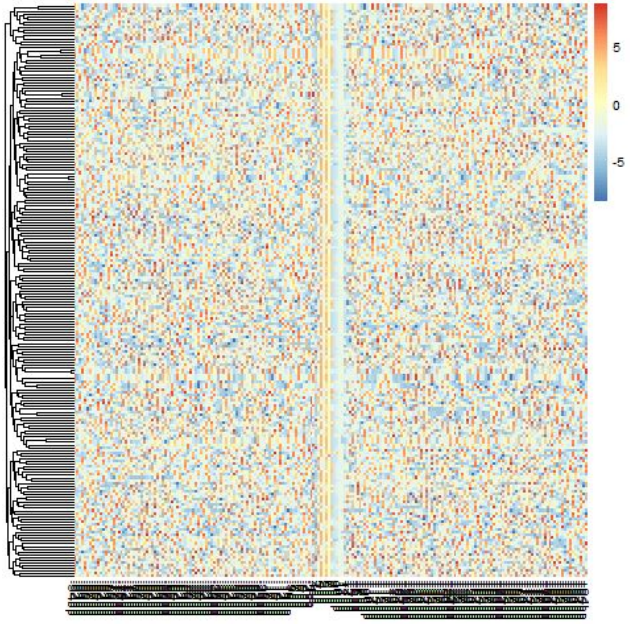
Rex



Spreading



Unbound



Results

-The heatmaps of the 200bp windows showed no significant differences between the DNA features of Bound (Rex and Spreading) and Unbound motifs

Next Steps

-20bp DNA windows were produced for the resulting Rex, Spreading and Unbound motifs that centered at the motifs (for better clarity)

-Granges of those 20bp Rex, Spreading and Unbound windows were produced

-DNA sequences were extracted by overlapping the windows' coordinates with BSgenome C. *elegans*

-DNAshapeR package was used to produce the scores regarding each of the 4 features (MGW, ProT, HelT and Roll) separately, for the 20bp DNA windows

Granges: Represent and manipulate genomic intervals and variables defined along a genome

BSgenome: Contains full genome sequence

Methods

-Each base in a 20bp sequence was given a score through a pentamer sliding window method used by the DNAshapeR package

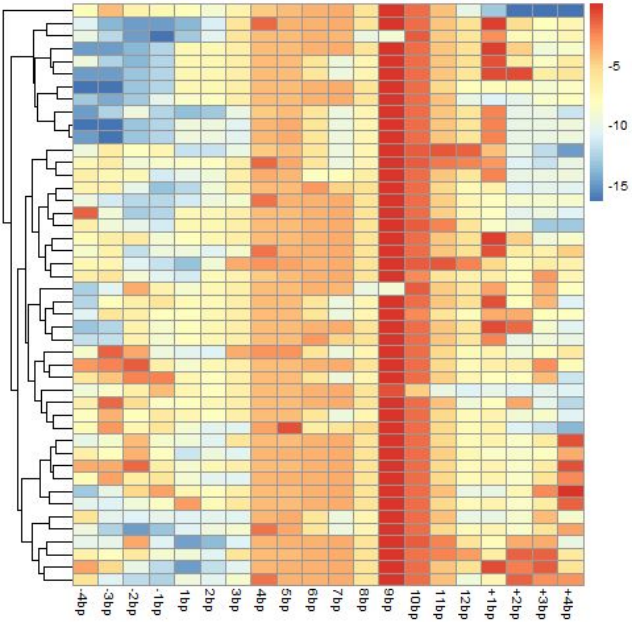
-DNA feature heatmaps of the Rex, Spreading and Unbound motifs were produced

-In each heatmap, the rows depict motifs and the columns depict the base-pairs from 1-20

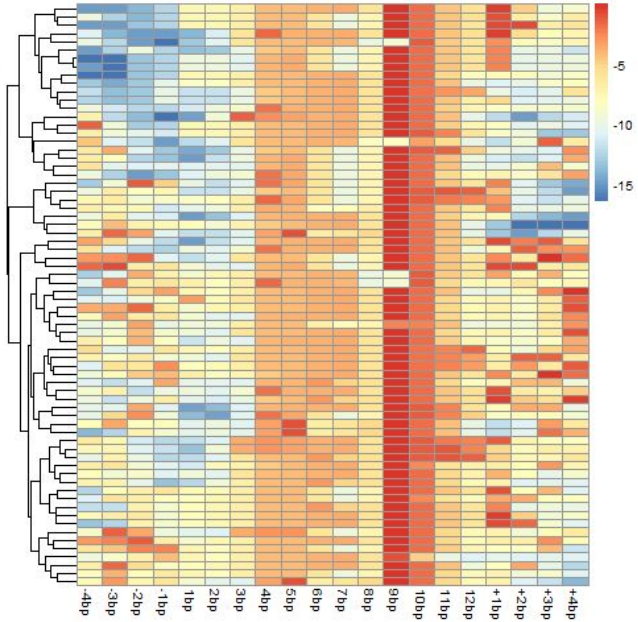
Results 20bp window (ProT)

-Rex vs Spreading vs Unbound motifs

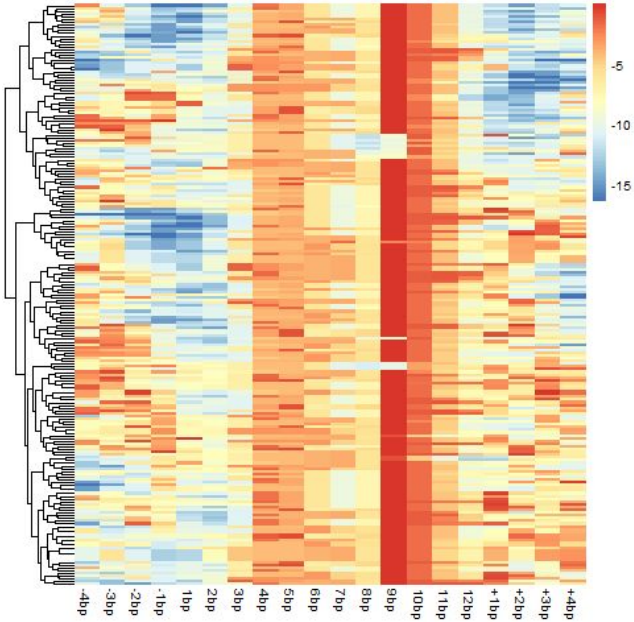
Rex



Spreading



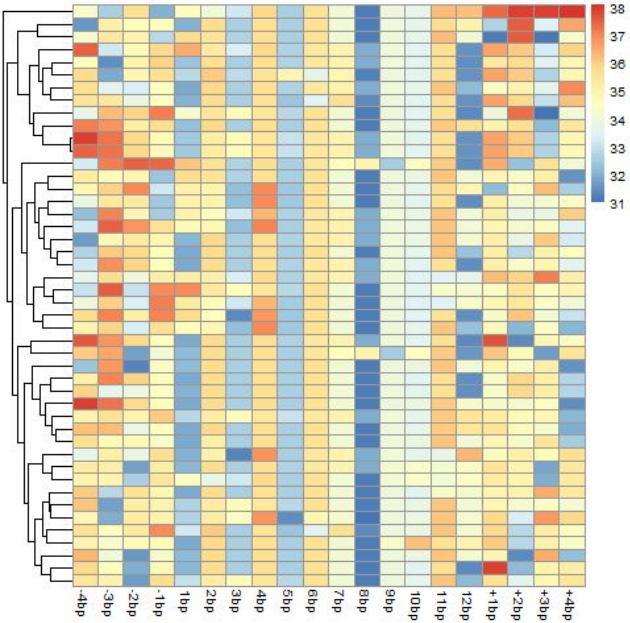
Unbound



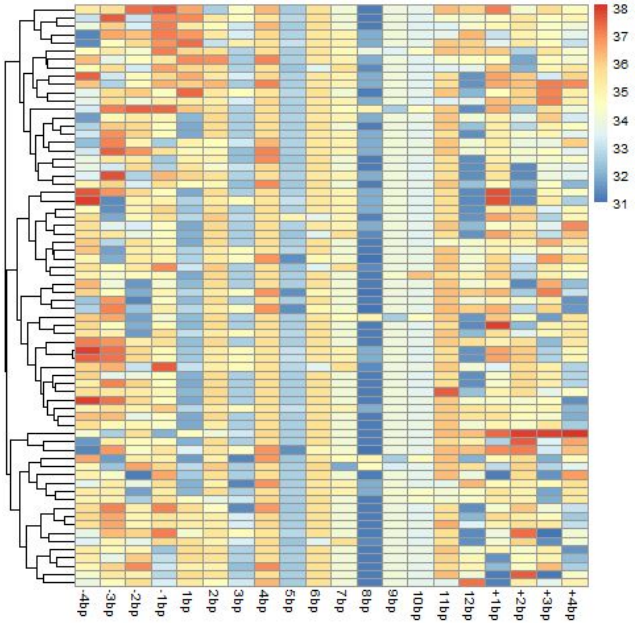
Results 20bp window (HelT)

-Rex vs Spreading vs Unbound motifs

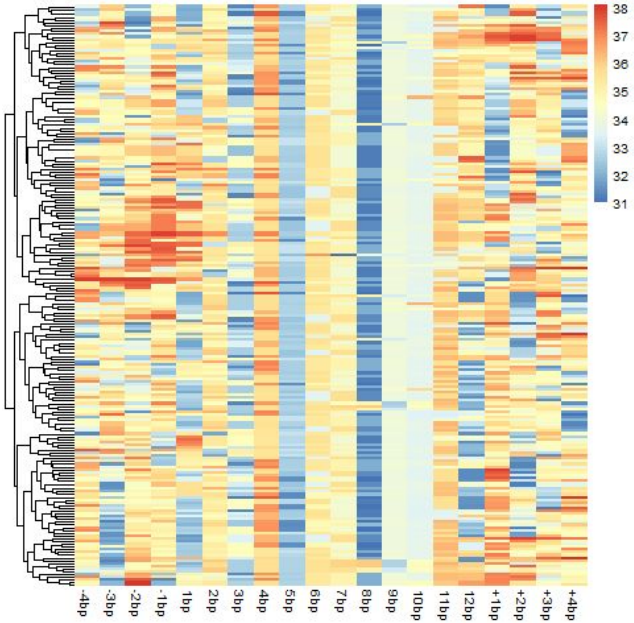
Rex



Spreading



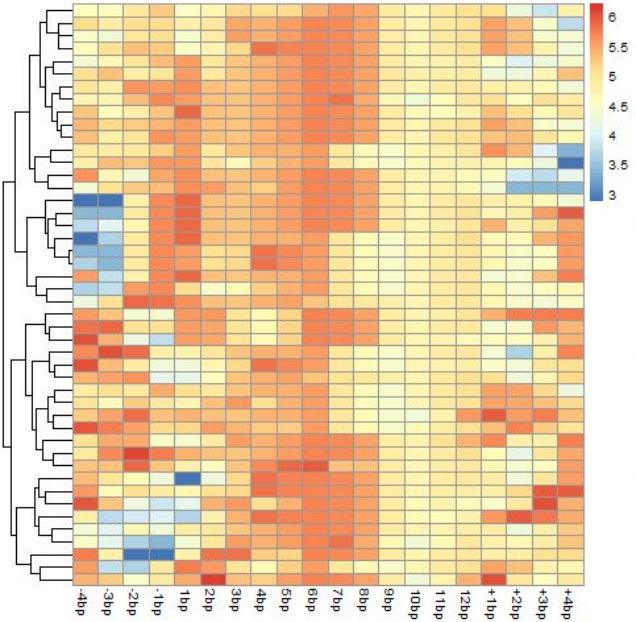
Unbound



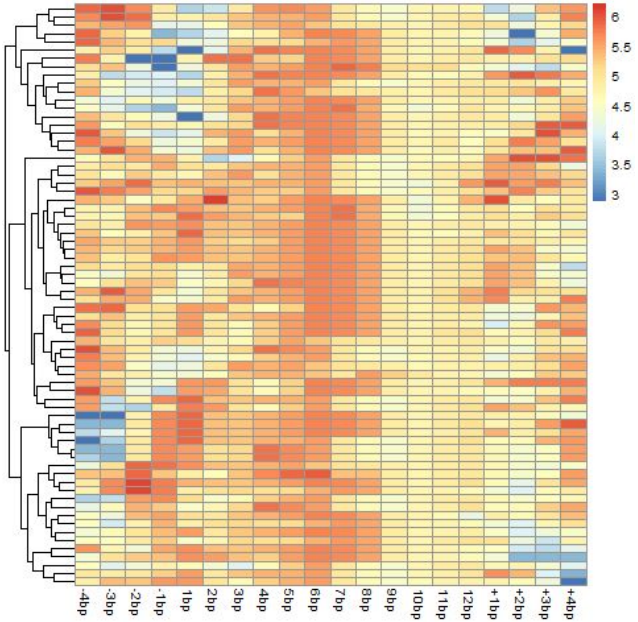
Results 20bp window (MGW)

-Rex vs Spreading vs Unbound motifs

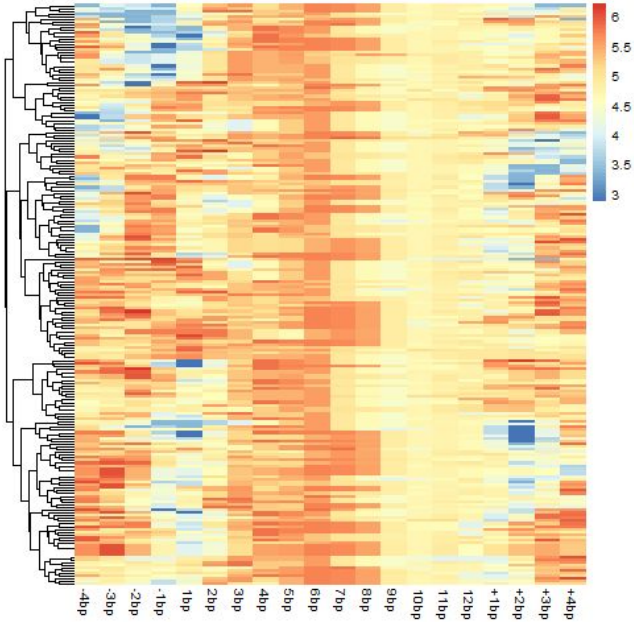
Rex



Spreading



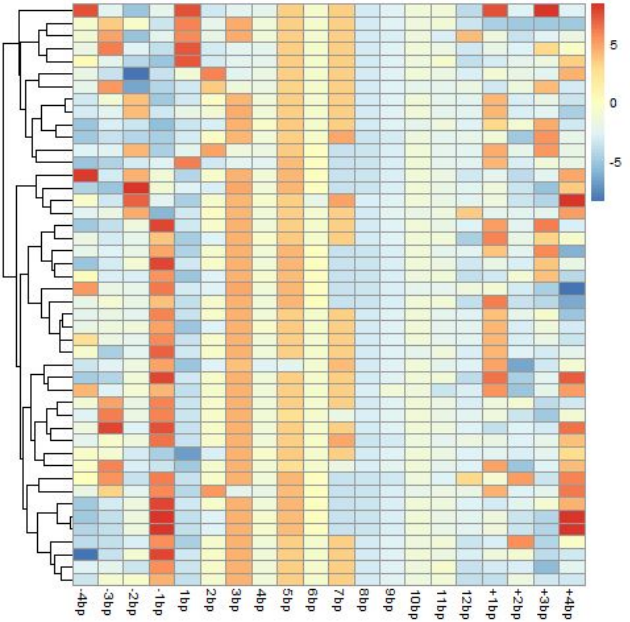
Unbound



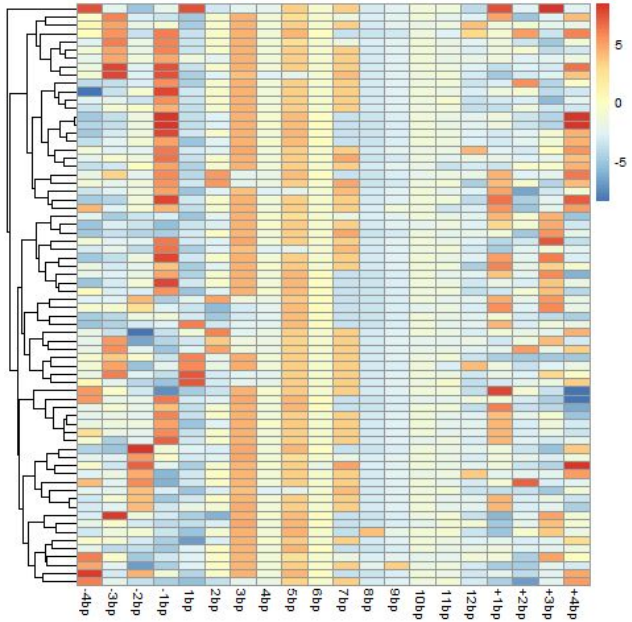
Results 20bp window (Roll)

-Rex vs Spreading vs Unbound motifs

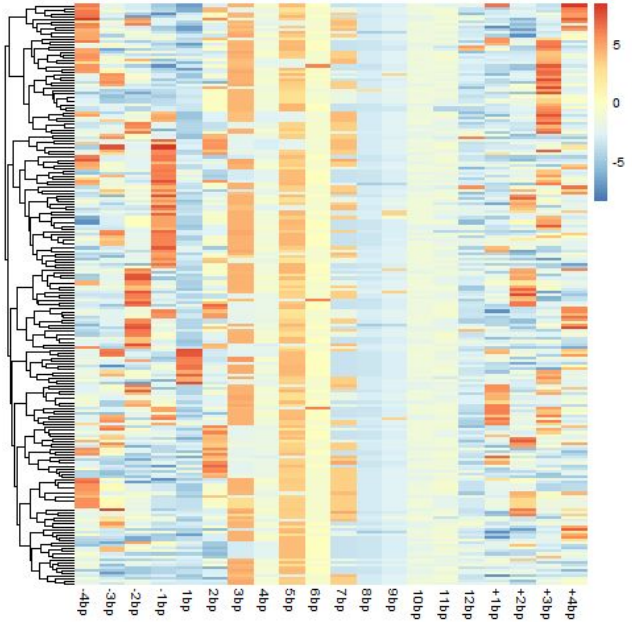
Rex



Spreading



Unbound



Results

-The heatmaps of the 20bp windows still showed no significant differences between the DNA features of Bound (Rex and Spreading) and Unbound motifs

What next?

-The Rex and Spreading sites data (from Albritton et. al 2017) were also previously ranked (just like the motif data)

-Contrary to earlier, only “Strong” ranked Rex Sites were chosen

-Only the the upper quartile scoring Spreading sites were chosen

-Like before, motifs scoring 7 and above were chosen to intersect with the Rex and Spreading Sites

What next?

-Intersection of the chosen motifs with the chosen Rex and Spreading data separately to find overlapping hits (Bound motifs)

-The rest of the motifs were (Unbound motifs)

-The most frequent motif in the three datasets was chosen: ATCGCGAAGGGA

-200bp DNA windows were produced for the resulting Rex, Spreading and Unbound motifs that centered at that motif

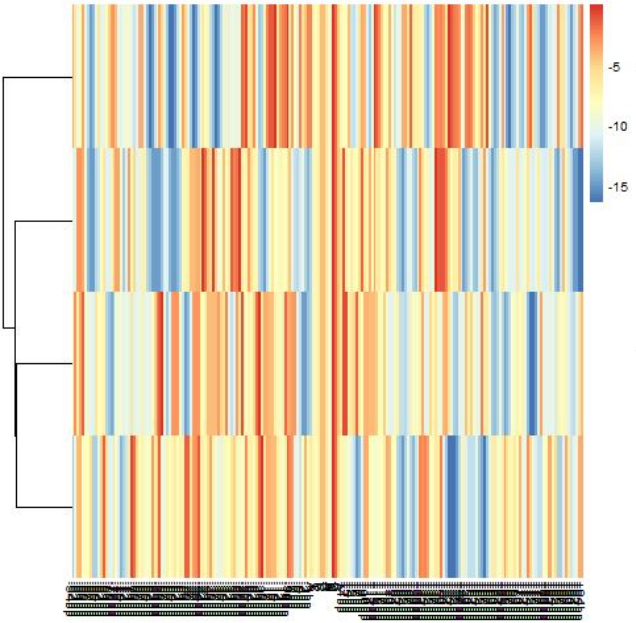
-DNAshapeR package was used to produce the scores regarding each of the 4 features (MGW, ProT, HelT and Roll) separately for the 200bp windows

Results 200bp window (ProT)

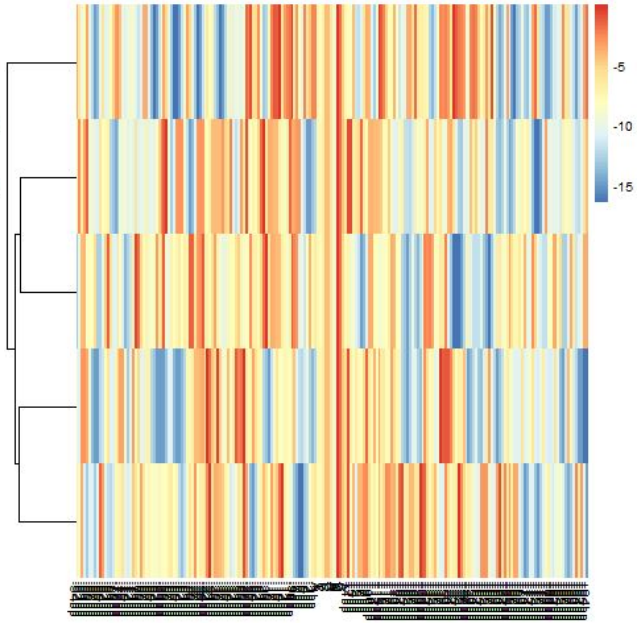
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

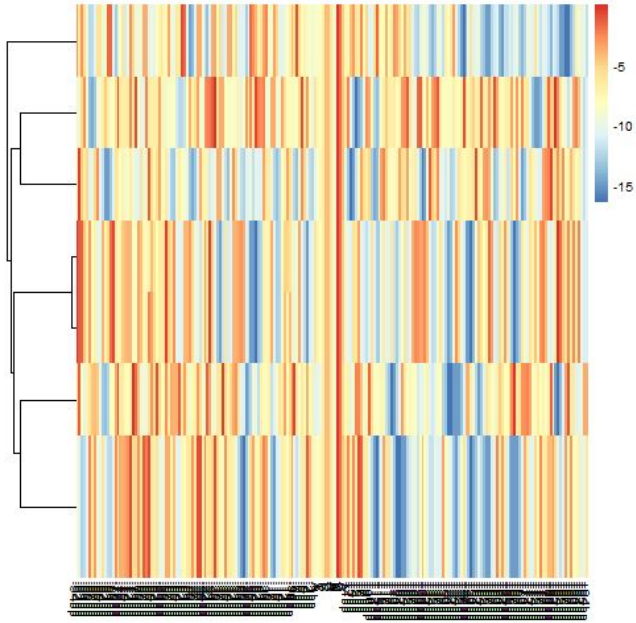
Rex



Spreading



Unbound

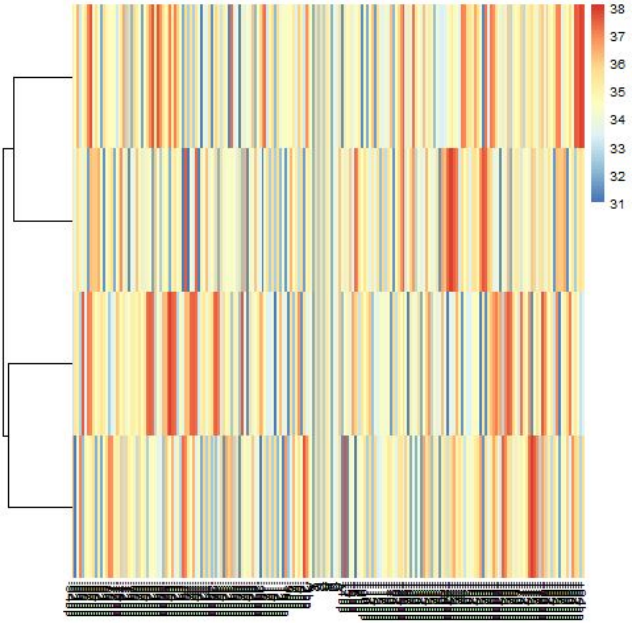


Results 200bp window (HelT)

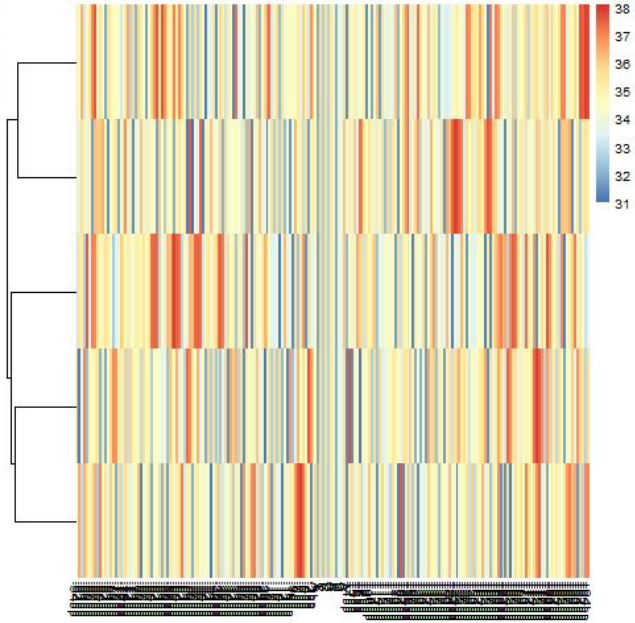
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

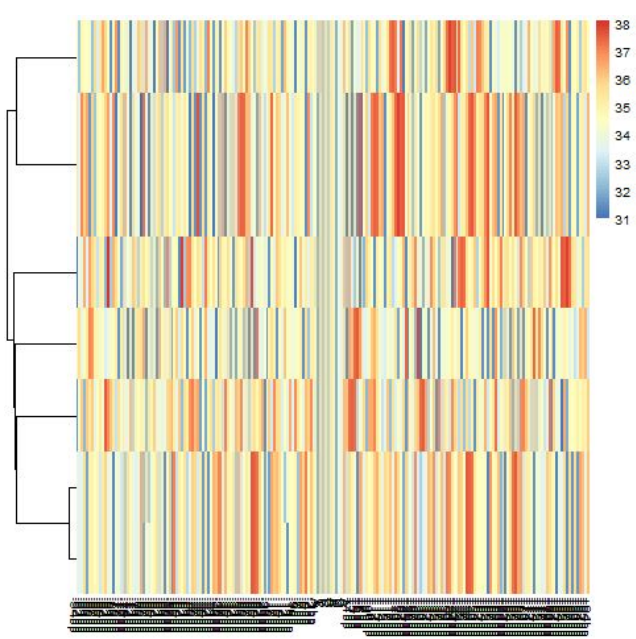
Rex



Spreading



Unbound

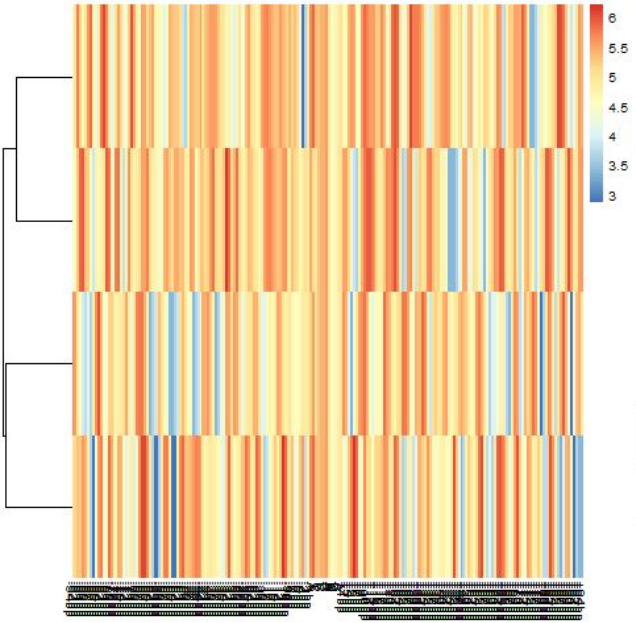


Results 200bp window (MGW)

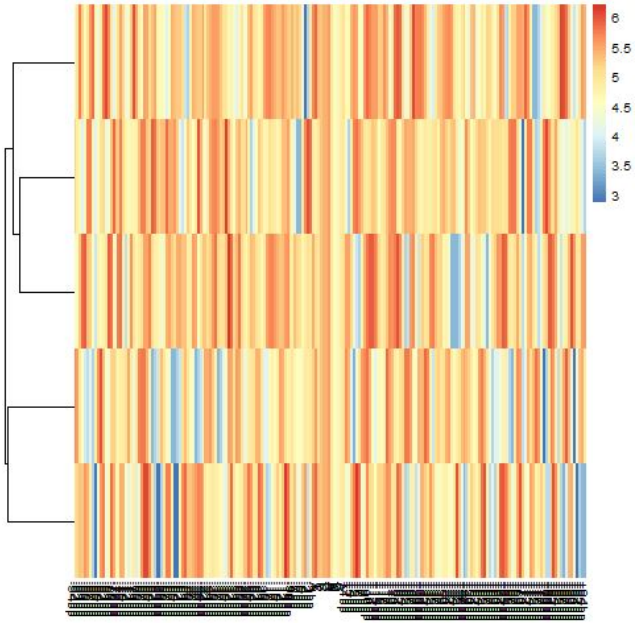
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

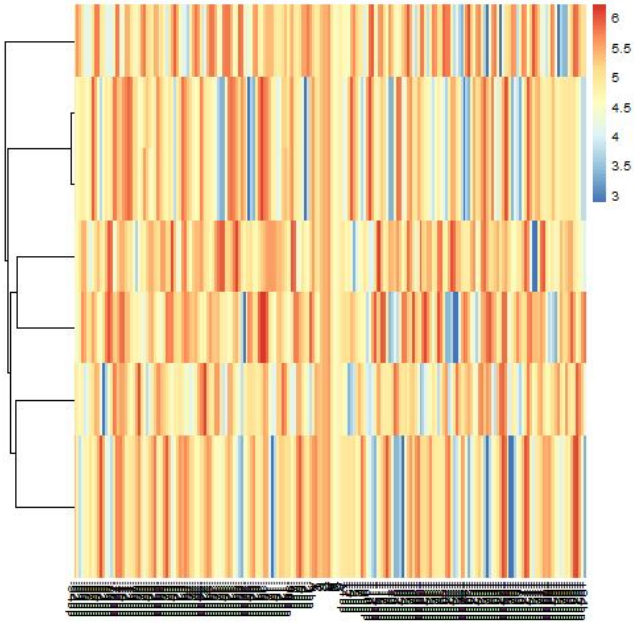
Rex



Spreading



Unbound

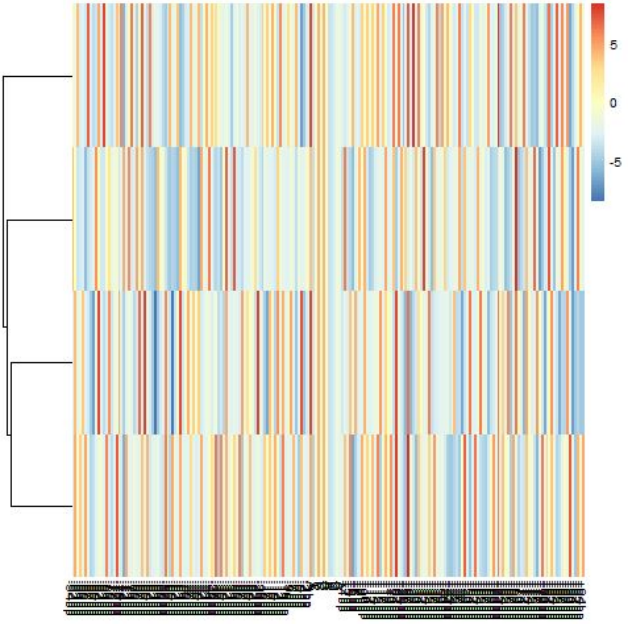


Results 200bp window (Roll)

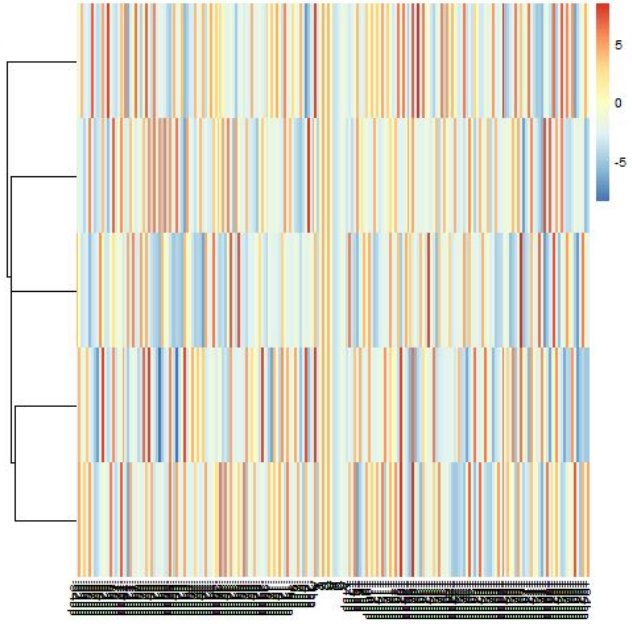
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

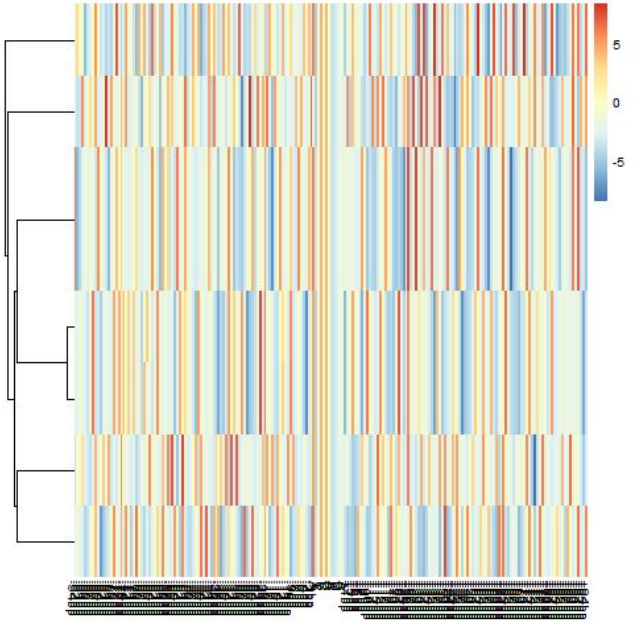
Rex



Spreading



Unbound



Results

-The sequence specific heatmaps of the 200bp windows were not clear and did not show any significant differences between the DNA features of Bound (Rex and Spreading) and Unbound motifs

Finally

-For better clarity, 20bp DNA windows were produced for the same data that centered at that motif

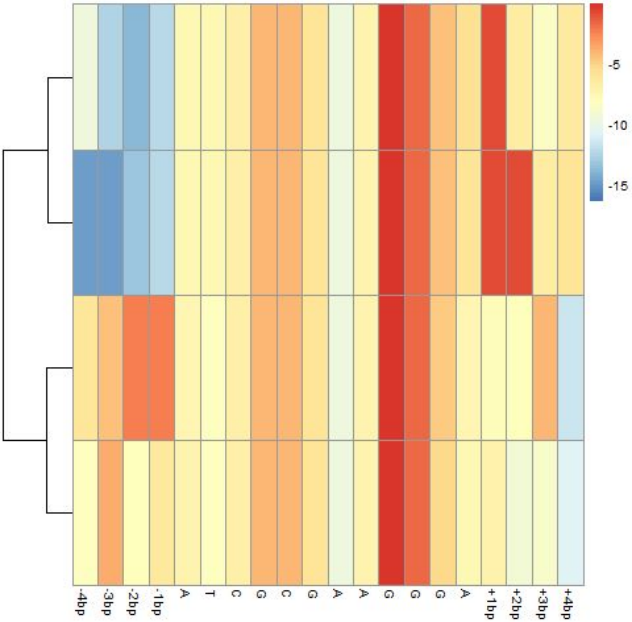
-DNAshapeR package was used to produce the scores regarding each of the 4 features (MGW, ProT, HelT and Roll) separately for the 20bp windows

Results 20bp window (ProT)

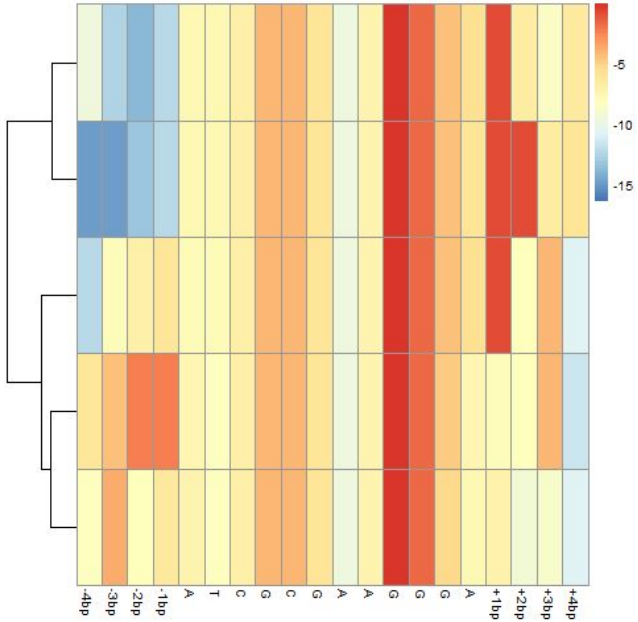
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

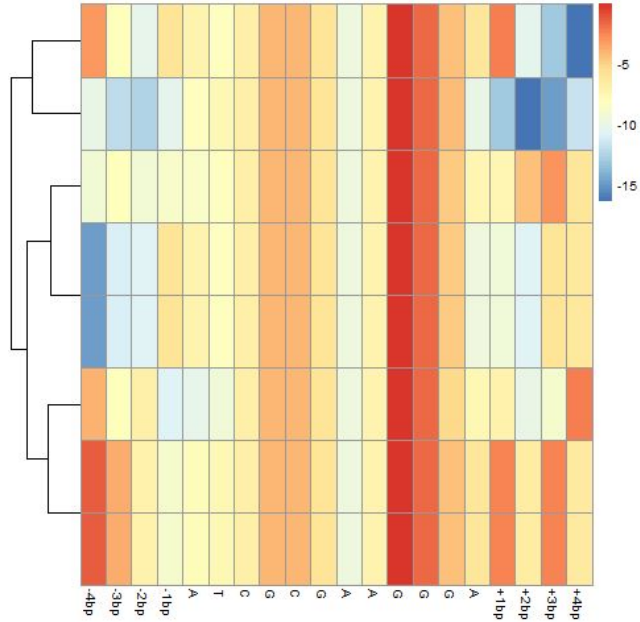
Rex



Spreading



Unbound

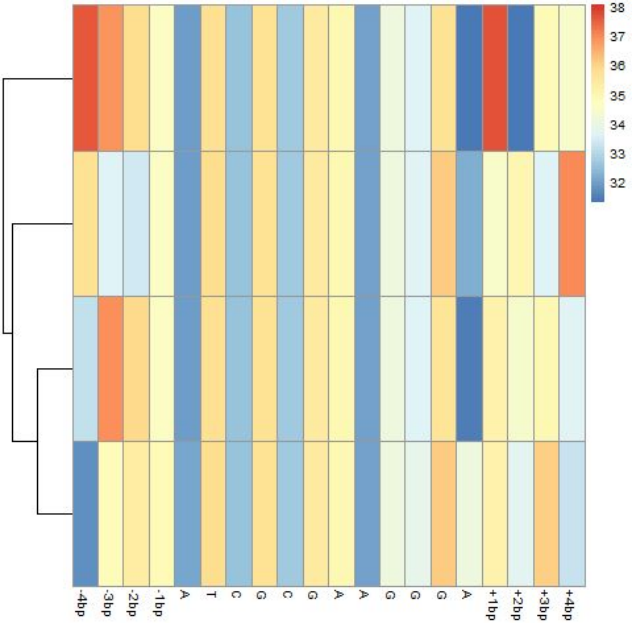


Results 20bp window (HelT)

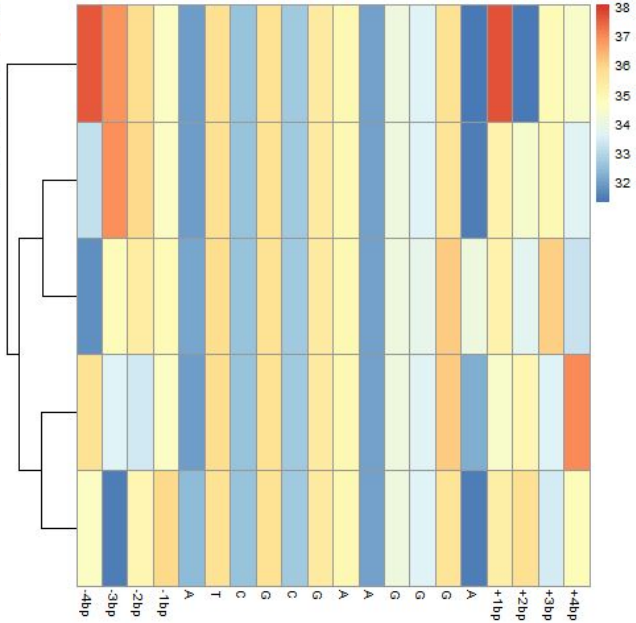
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

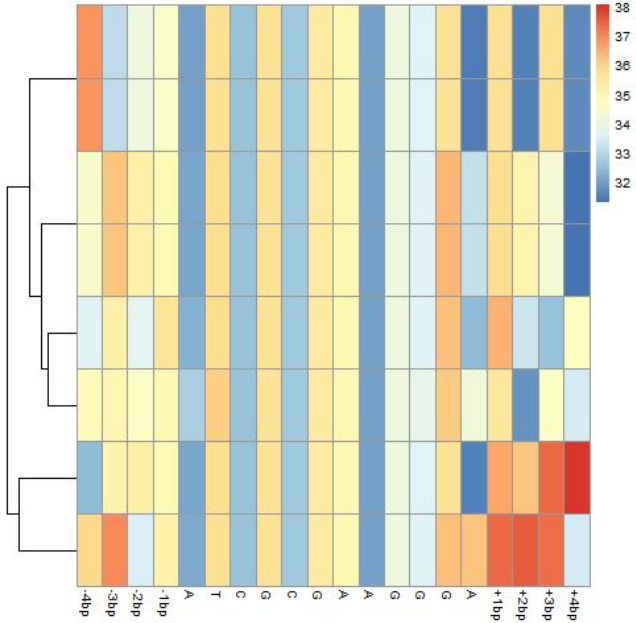
Rex



Spreading



Unbound

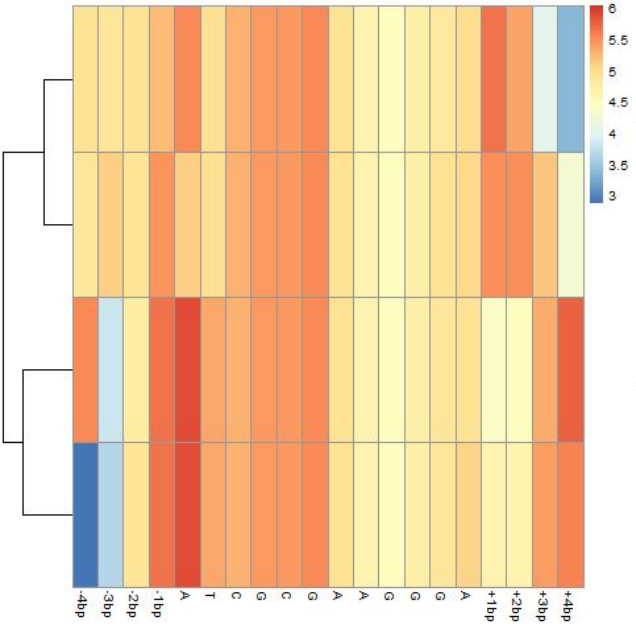


Results 20bp window (MGW)

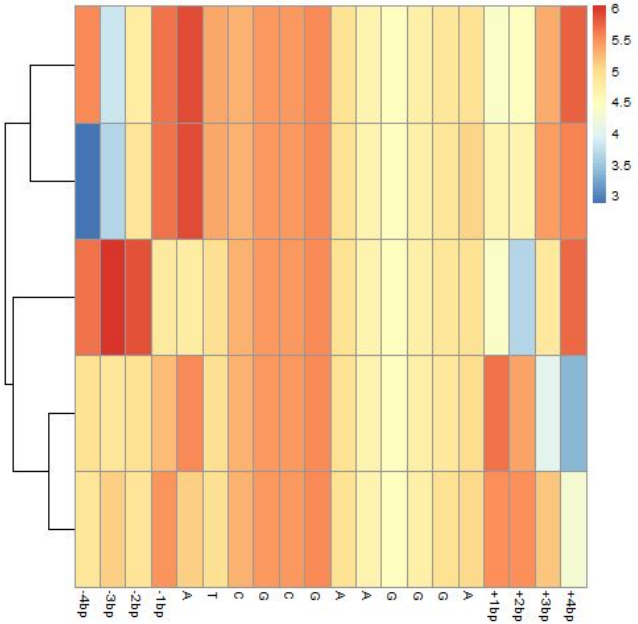
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

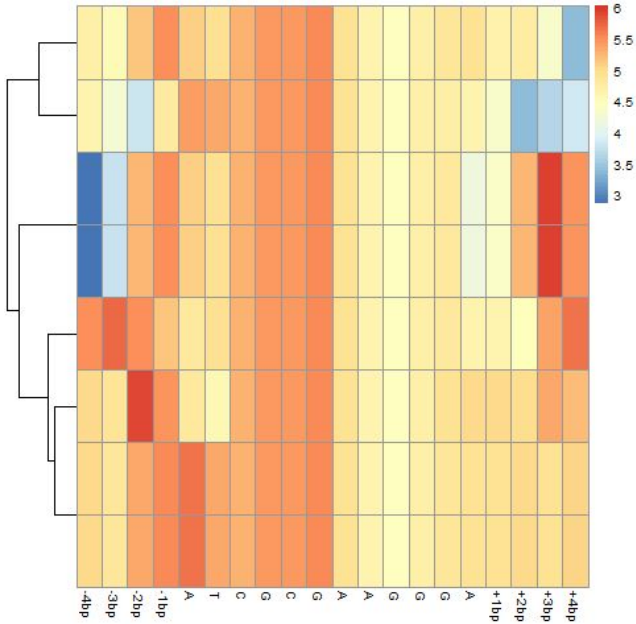
Rex



Spreading



Unbound

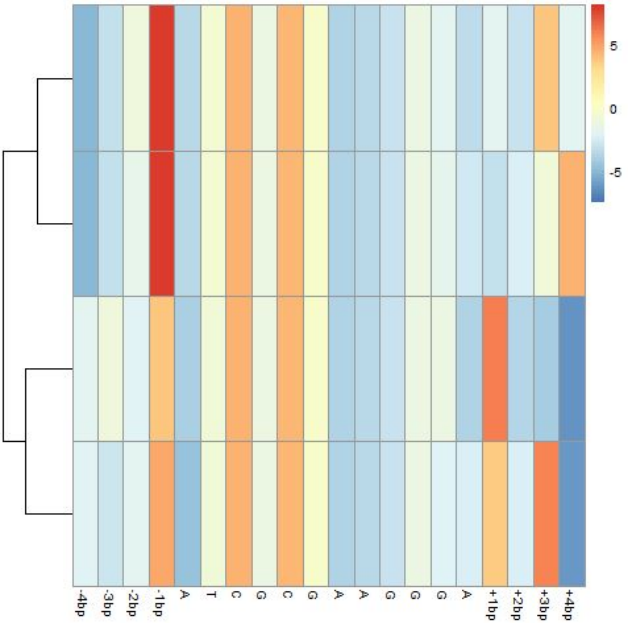


Results 20bp window (Roll)

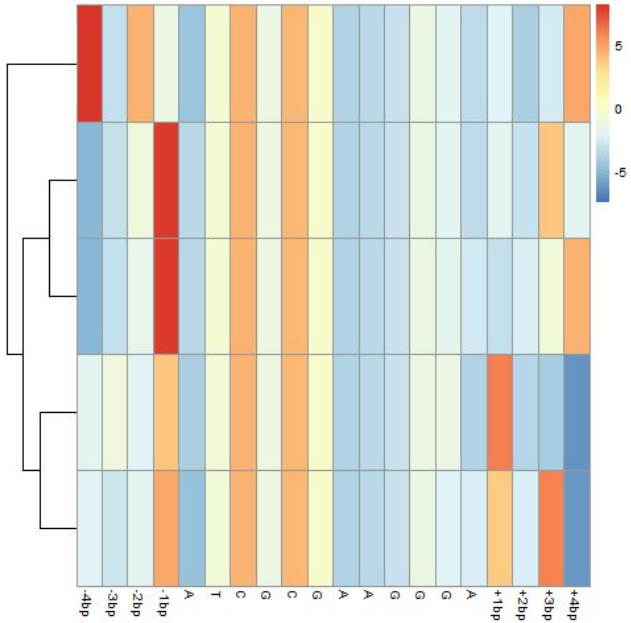
-Rex vs Spreading vs Unbound motifs

-The motif ATCGCGAAGGGA was chosen for this sequence specific analysis

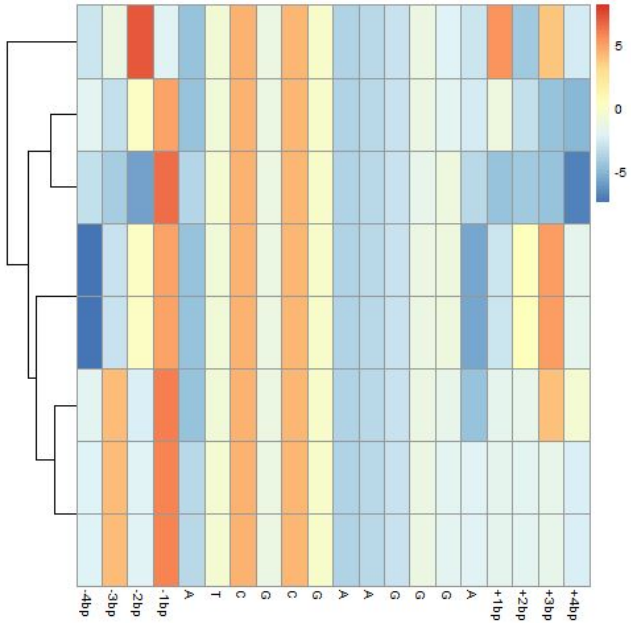
Rex



Spreading



Unbound



Results

-The sequence specific heatmaps of the 20bp windows showed no significant differences between the DNA features of Bound (Rex and Spreading) and Unbound motifs

CONCLUSIONS & FURTHER APPLICABLE WORK

-Our analysis suggests no clear pattern that can distinguish the Bound motifs from the Unbound motifs

-Our analysis shows that the DNA shape is sequence specific

-DNA shape is not able to explain why some motifs recruit and others do not

-Although our hypothesis was disproved, the results from this research are very important in proving that we have to focus on the other explanation for DCC binding to motifs at Rex and Spreading sites, specifically nucleosome occupancy as suggested by Albritton et al. 2017

Bibliography

-Albritton, S. E., A. L. Kranz, L. H. Winterkorn, L. A. Street, and S. Ercan (2017, May 30). Cooperation between a Hierarchical Set of Recruitment Sites Targets the X Chromosome for Dosage Compensation. Retrieved from https://www.ncbi.nlm.nih.gov/m/pubmed/28562241/