

# Supplementary IIa: Context dependent prediction in DNA sequence using neural networks. Fourier, human genome, LSTM200, low frequencies.

Christian Grønbaek, Yuhu Liang, Desmond Elliott, and Anders Krogh

February 6, 2021

This file contains the first of two sets of plots based on the Fourier analysis showing the L2-norm of the Fourier coefficients in a running window. This first set covers the frequency range from 200 to 45000 using a window length of 1000 (and a step size of 100); the second set covers the frequencies from 40000 to 140000 and used a window length of 5000 (and step size of 100) and is placed in a separate file.

# Fourier plots, LSTM200, frequency range 200 to 45000

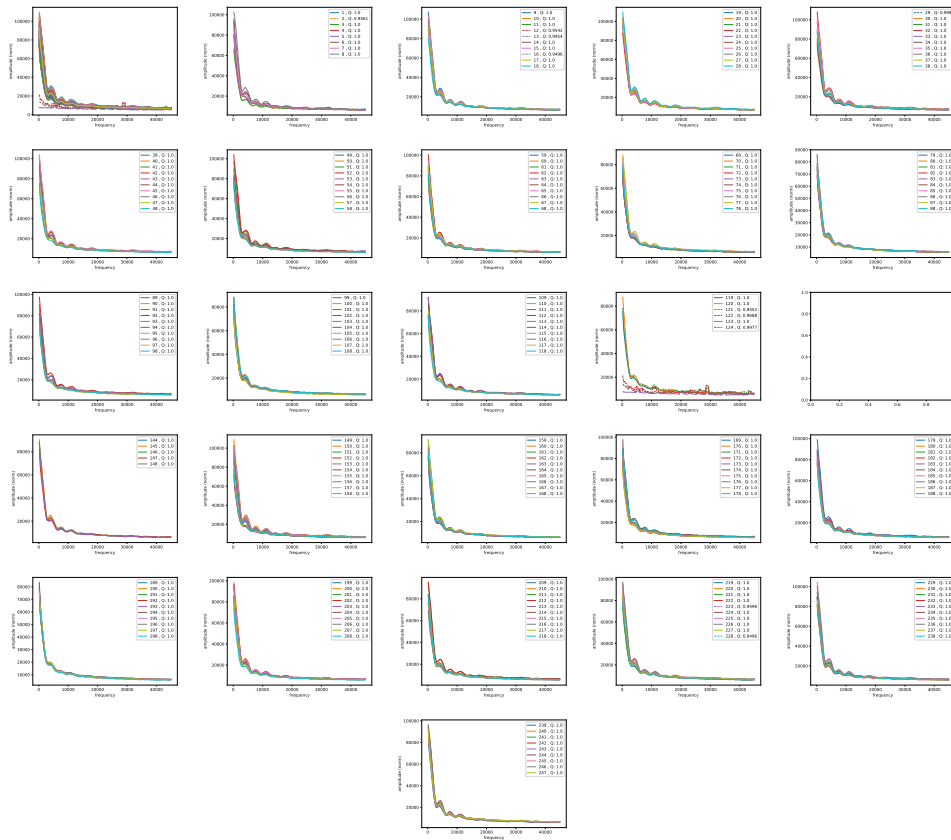


Figure 1: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr1. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

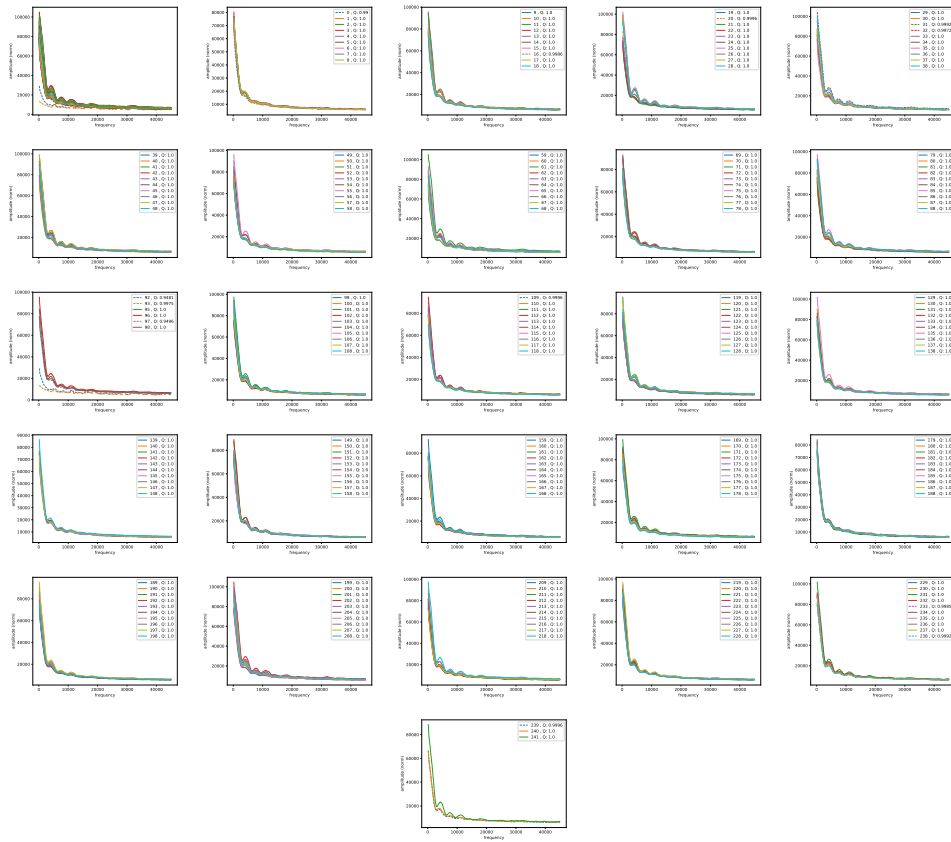


Figure 2: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr2. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

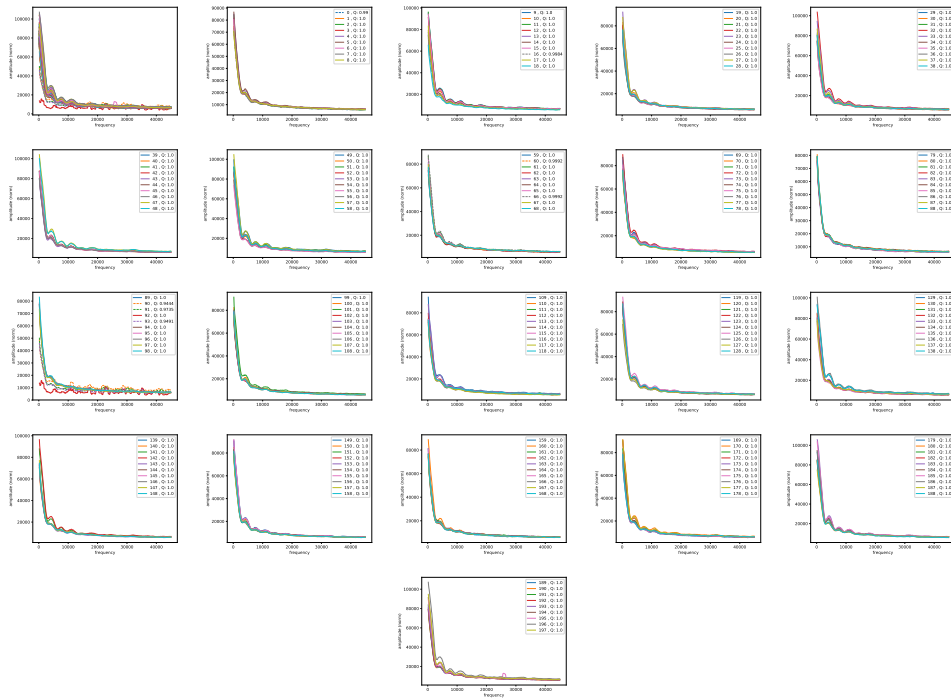


Figure 3: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr3. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.



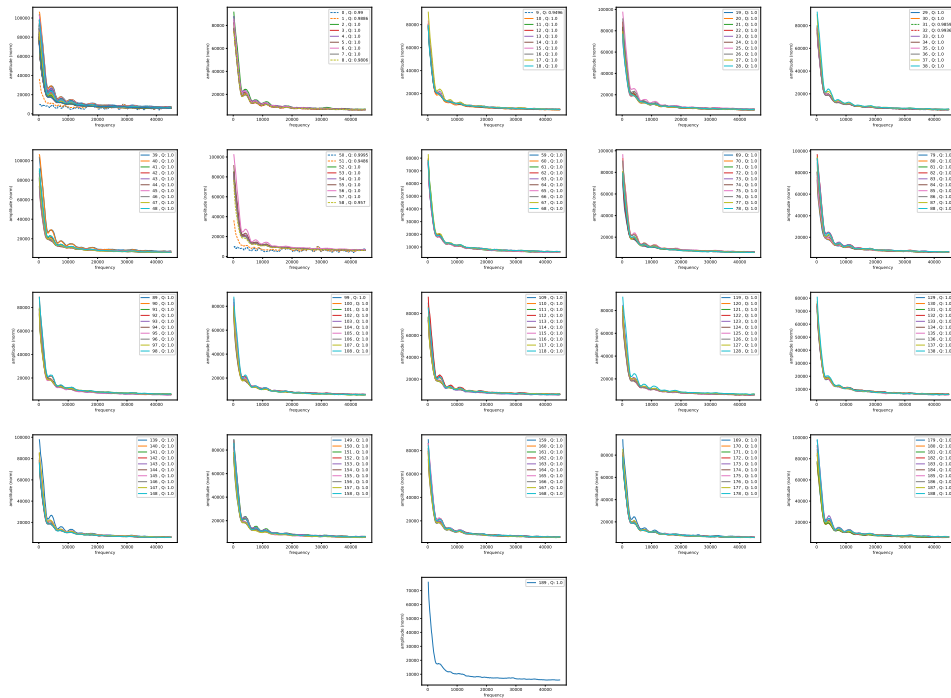


Figure 4: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr4. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

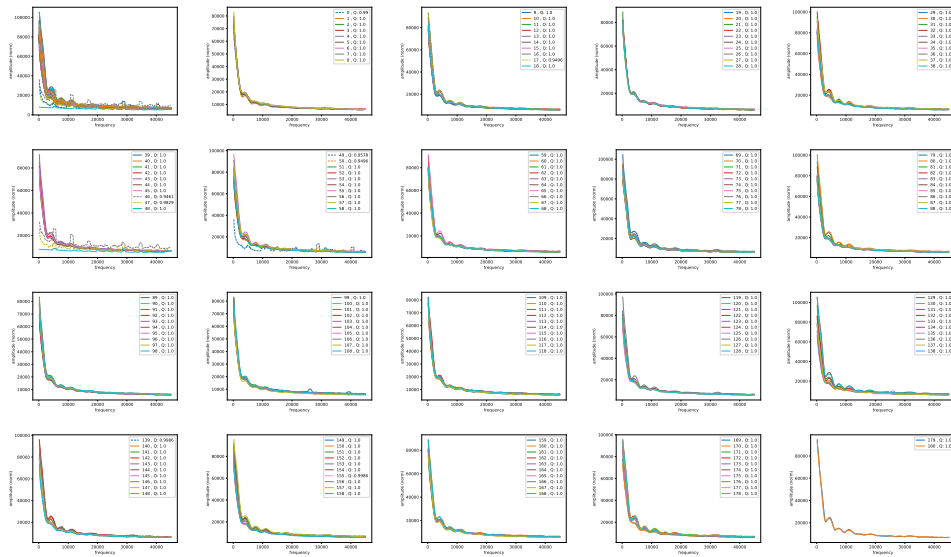


Figure 5: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr5. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

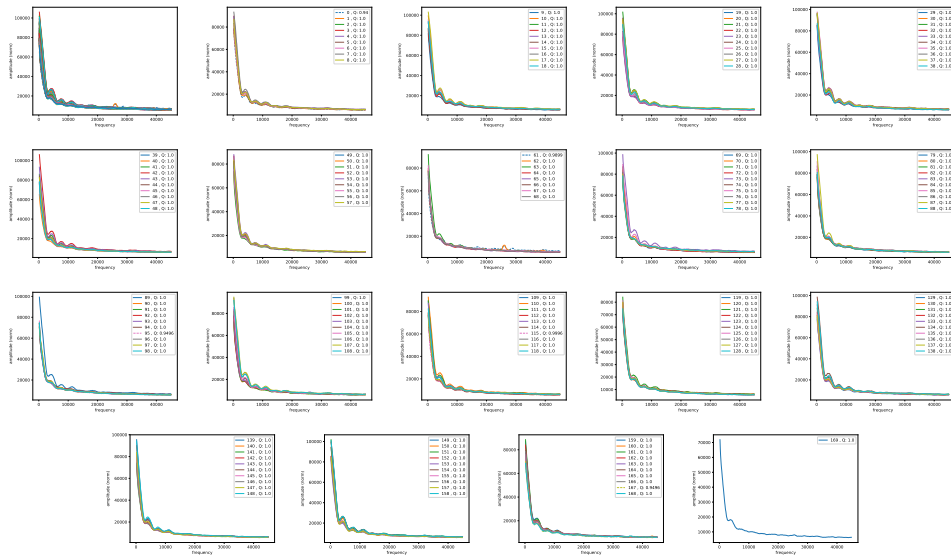


Figure 6: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr6. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

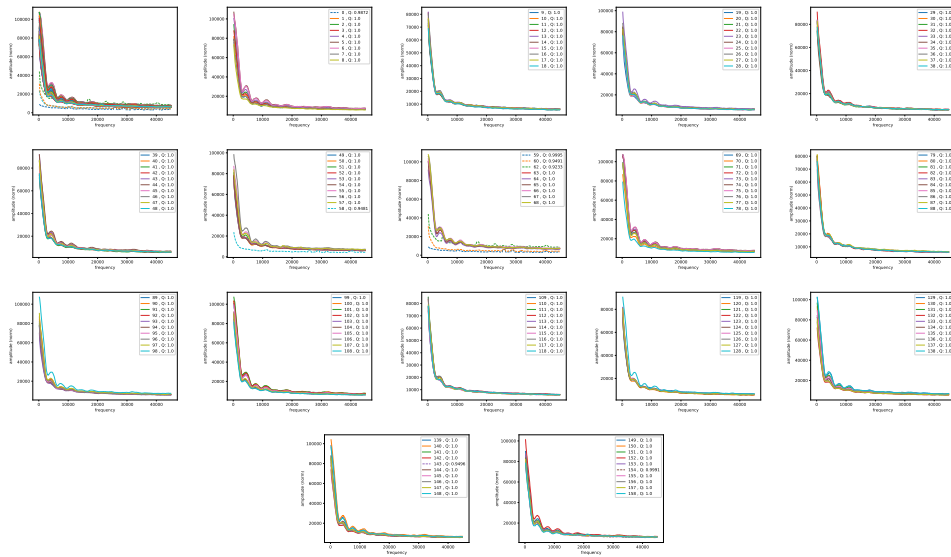


Figure 7: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr7. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

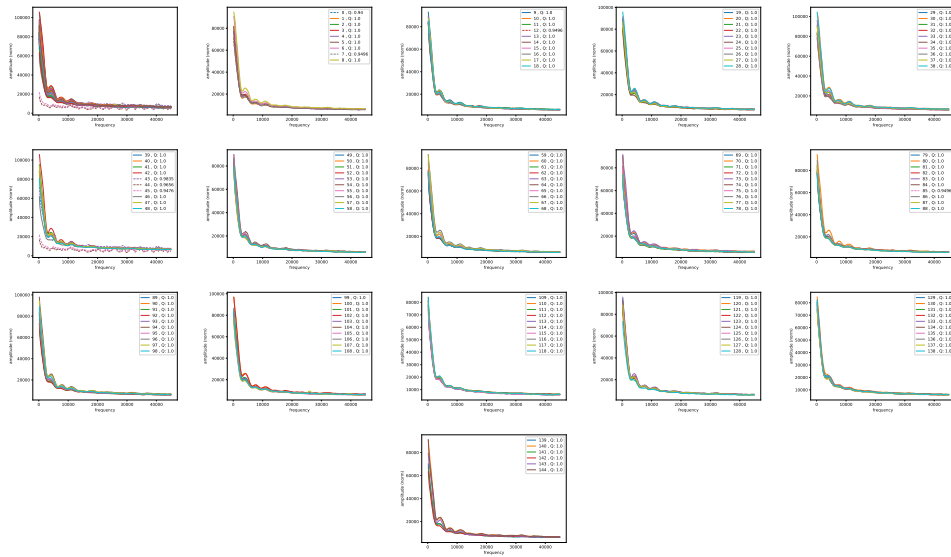


Figure 8: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr8. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

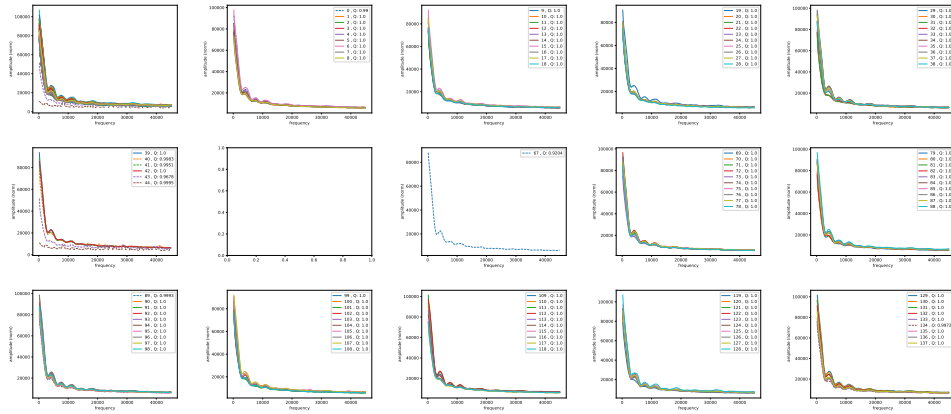


Figure 9: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr9. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

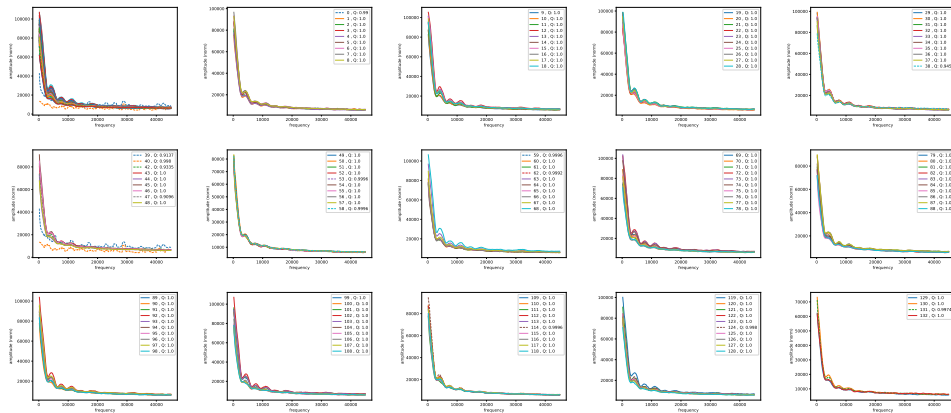


Figure 10: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr10. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

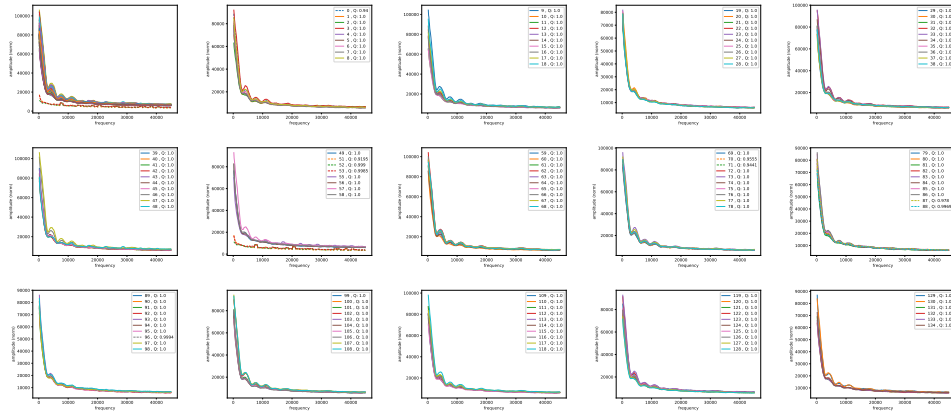


Figure 11: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr11. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

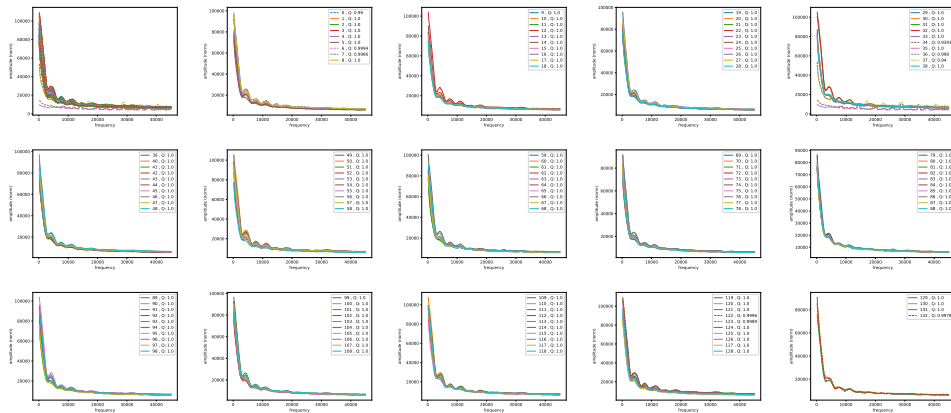


Figure 12: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr12. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

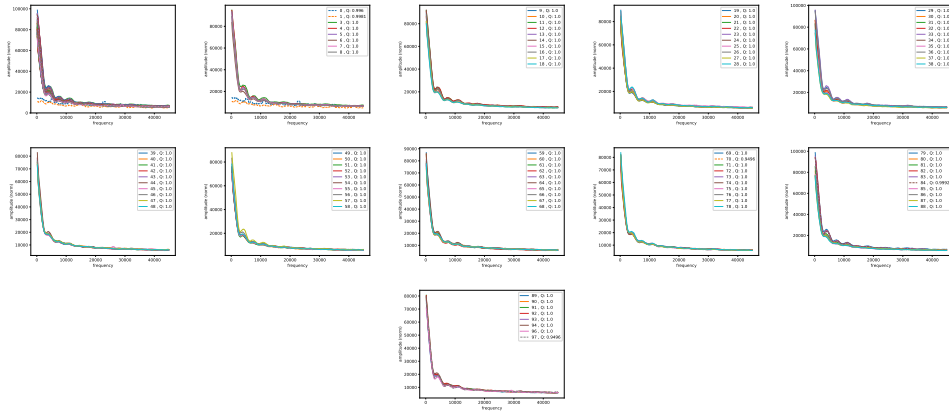


Figure 13: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr13. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

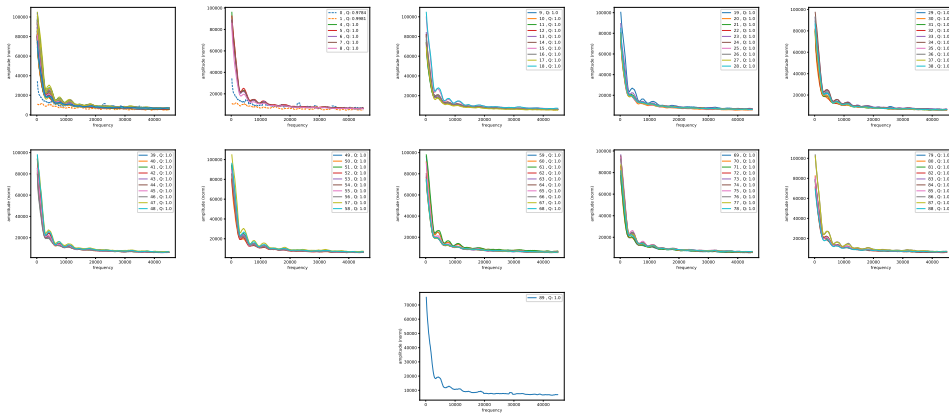


Figure 14: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr14. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.



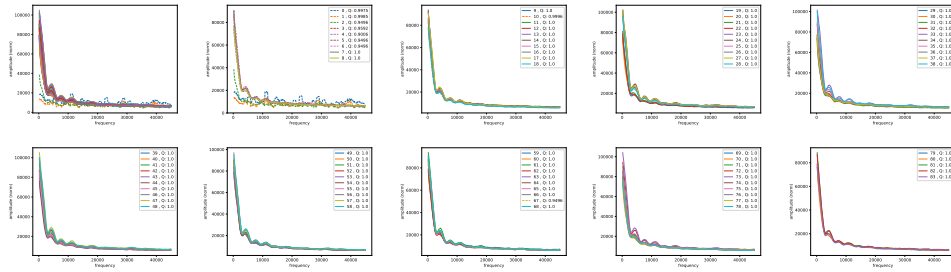


Figure 15: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr15. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

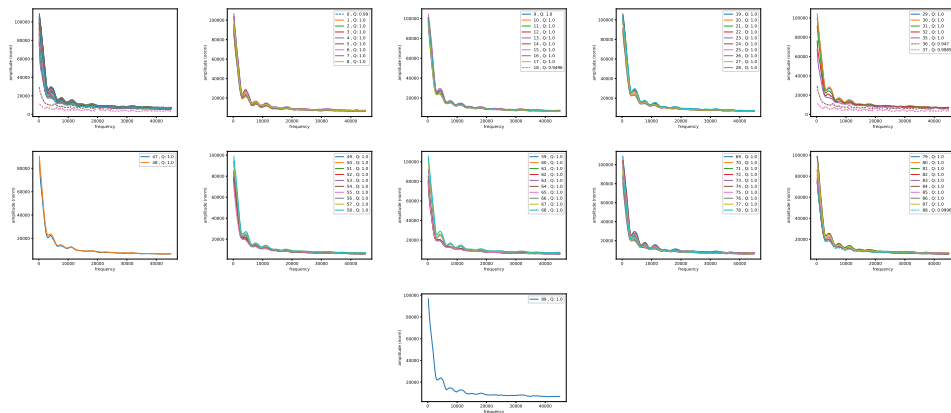


Figure 16: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr16. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

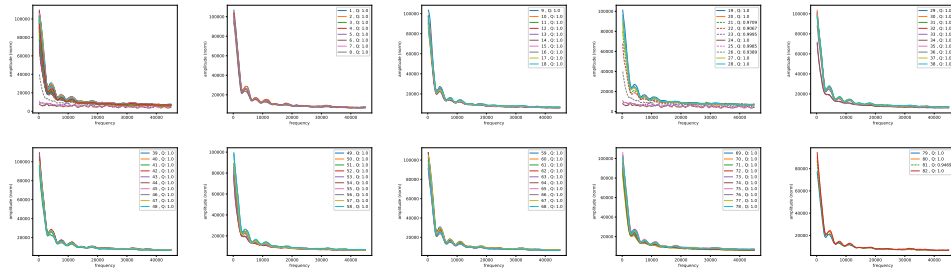


Figure 17: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr17. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

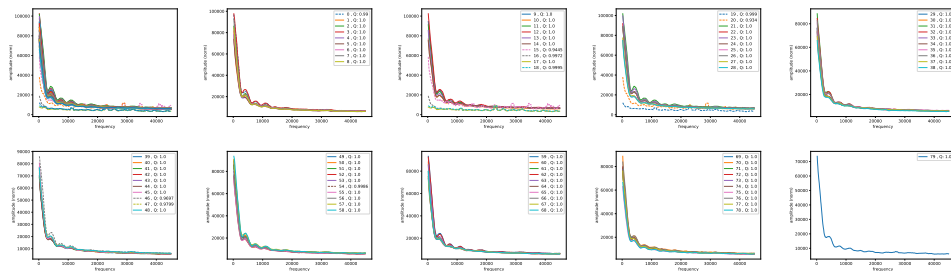


Figure 18: LSTM200. Fouriers on reference-base probability for chromosome: hg38\_chr18. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

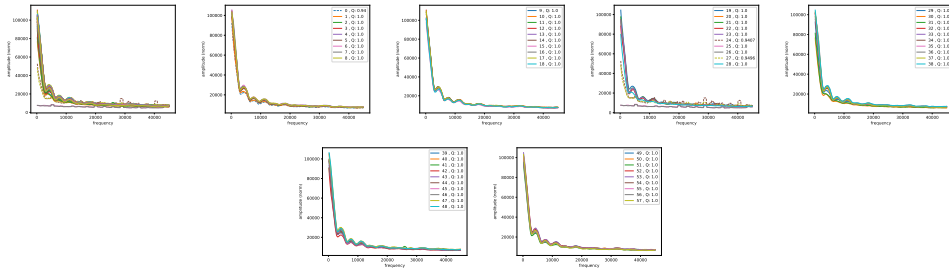


Figure 19: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr19. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

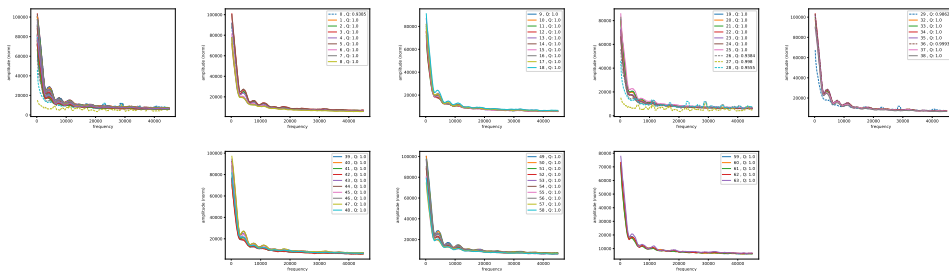


Figure 20: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr20. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

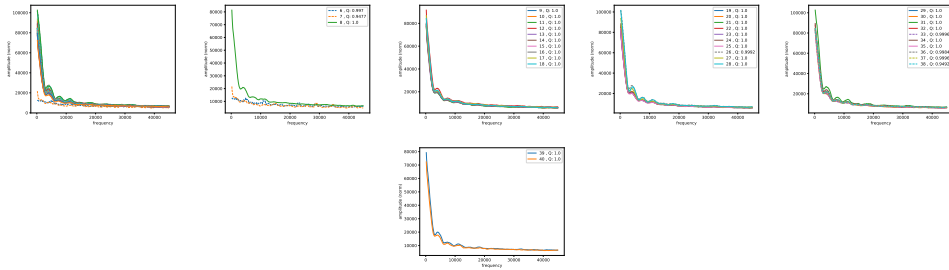


Figure 21: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr21. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

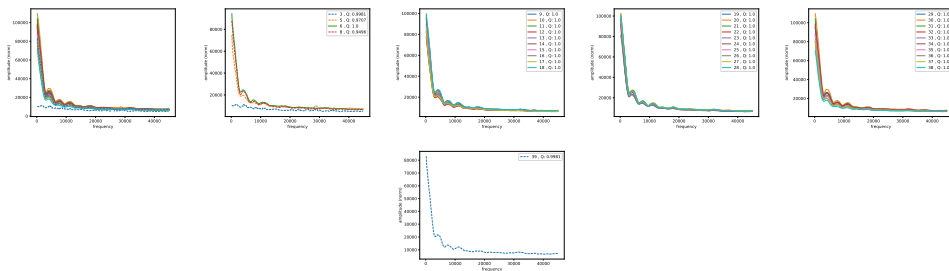


Figure 22: LSTM200. Fourier on reference-base probability for chromosome: hg38\_chr22. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.