

Supplementary Iib, part1 (chr1-11): Context dependent prediction in DNA sequence using neural networks. Fourier, human genome, LSTM200, high frequencies.

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

April 23, 2022

This file and its companion, part2, contains the second 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); this second file/set covers the frequencies from 40000 to 140000 and used a window length of 5000 (and step size of 100). Note that, since the window length is here five times larger than for the lower frequency range (first set), the scale of the amplitudes (y-axis) is five times higher than there.

Here the chromosomes 1-11 are covered; part2 covers chromosomes 12-22.

Fourier plots, LSTM200, frequency range 40000 to 140000

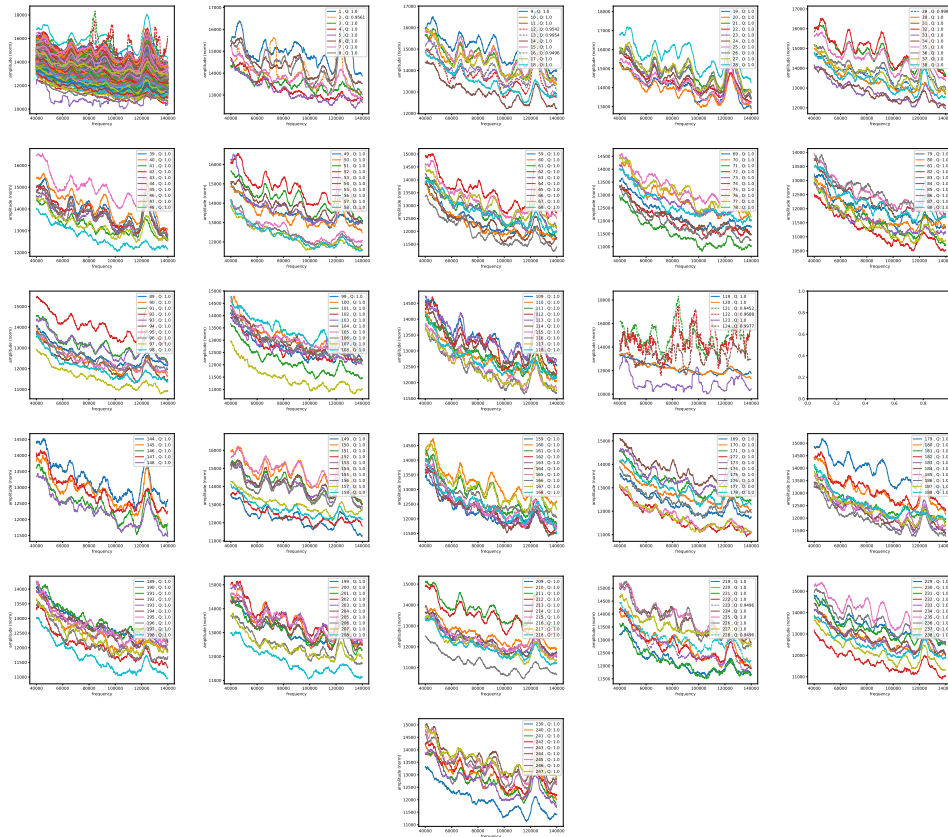


Figure 1: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr1. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

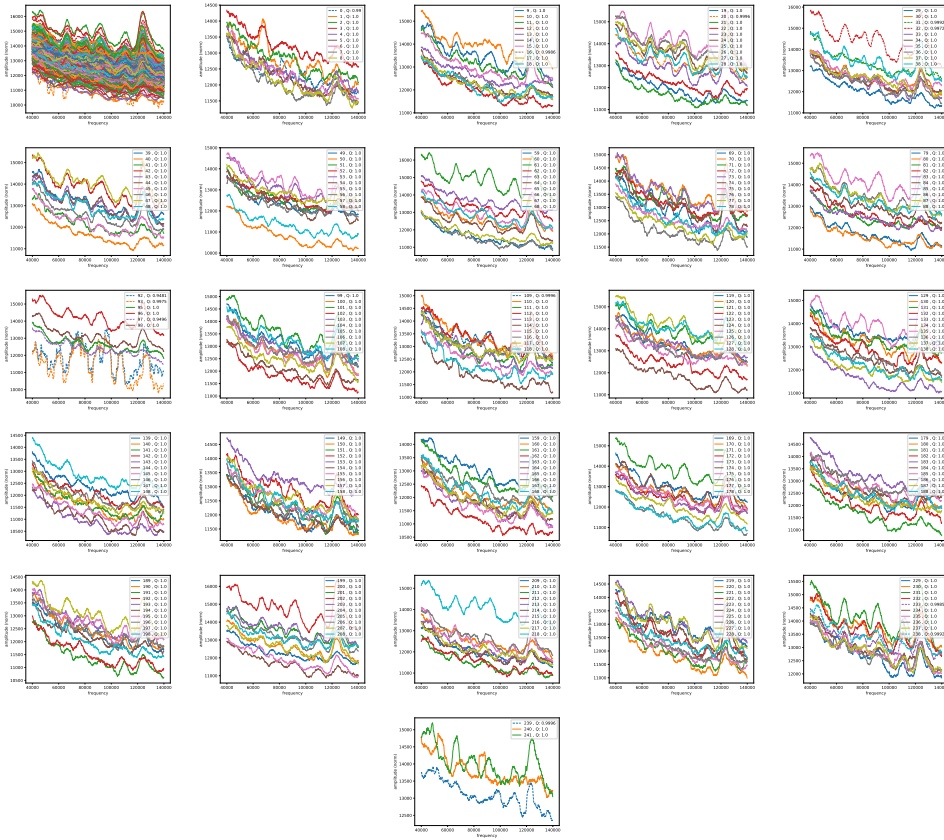


Figure 2: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr2. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

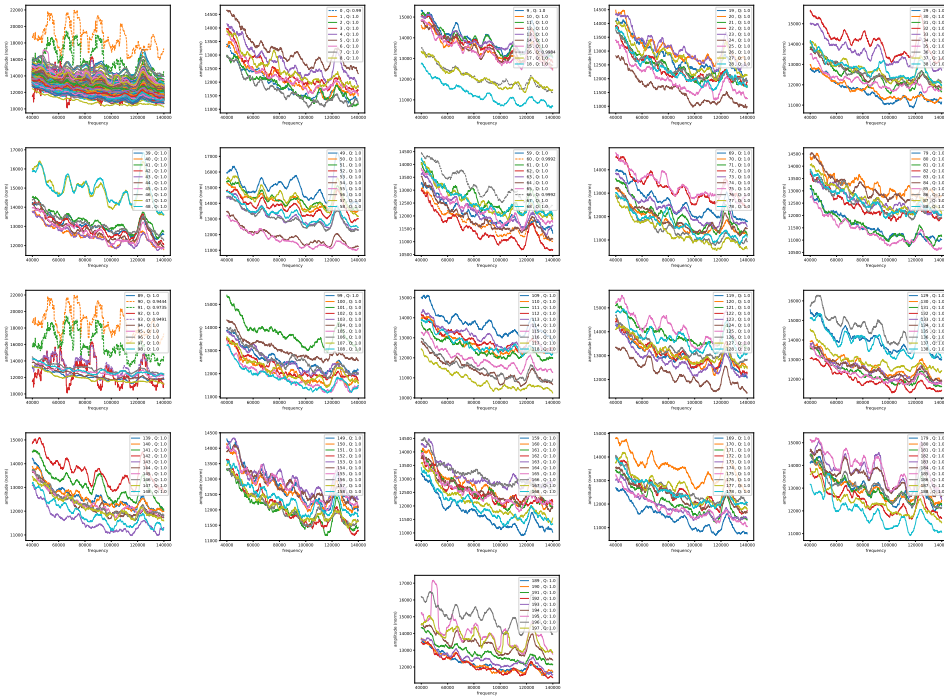


Figure 3: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr3. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

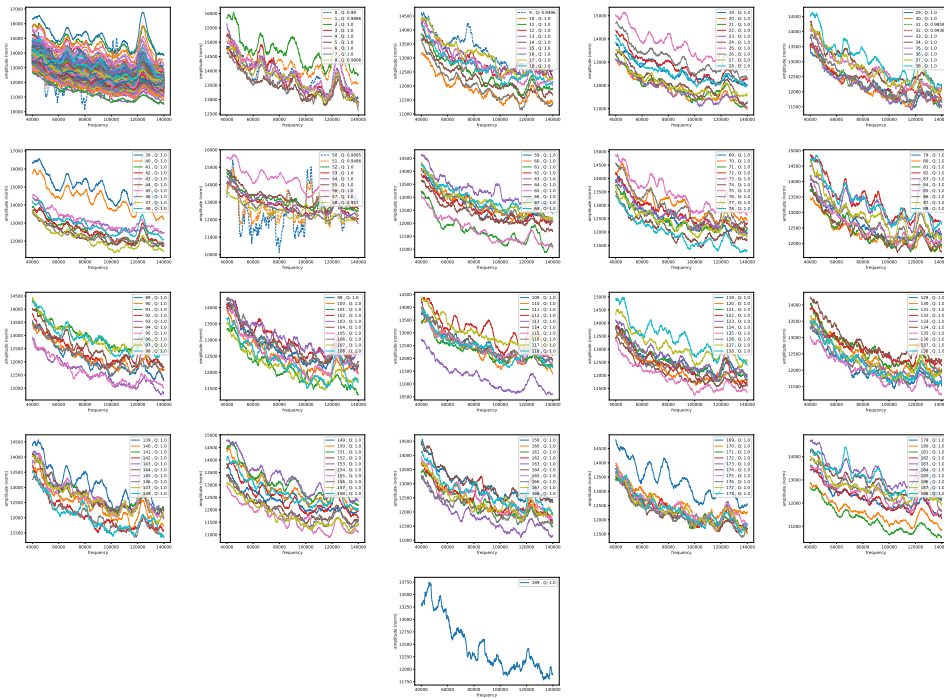


Figure 4: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr4. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

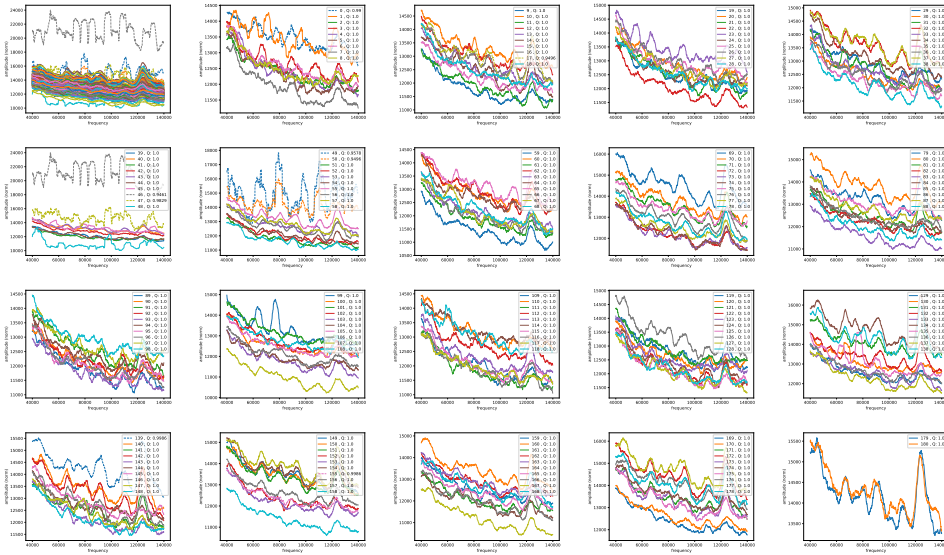


Figure 5: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr5. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

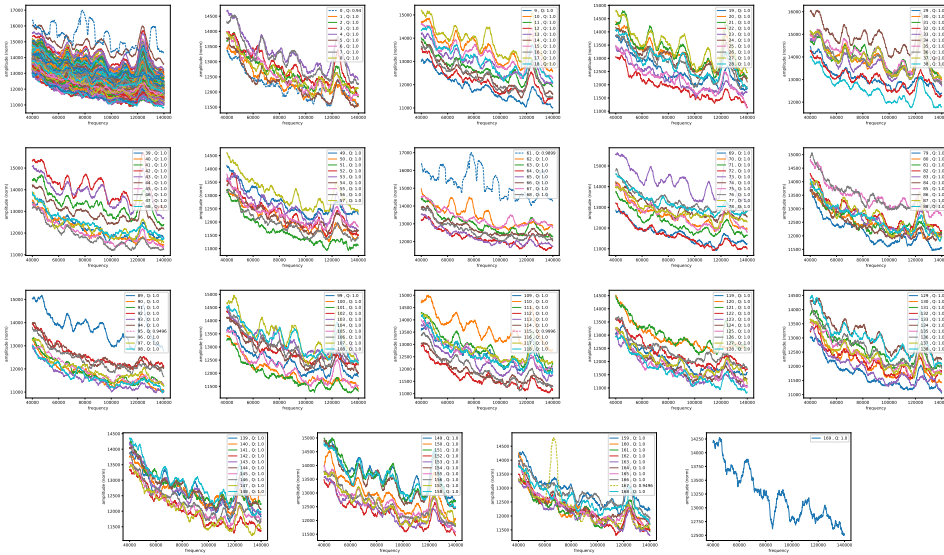


Figure 6: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr6. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

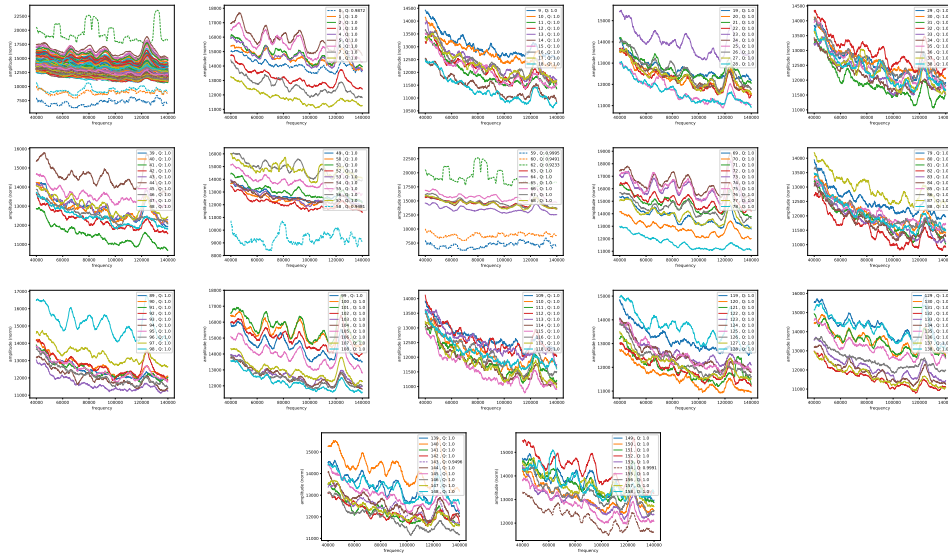


Figure 7: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr7. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

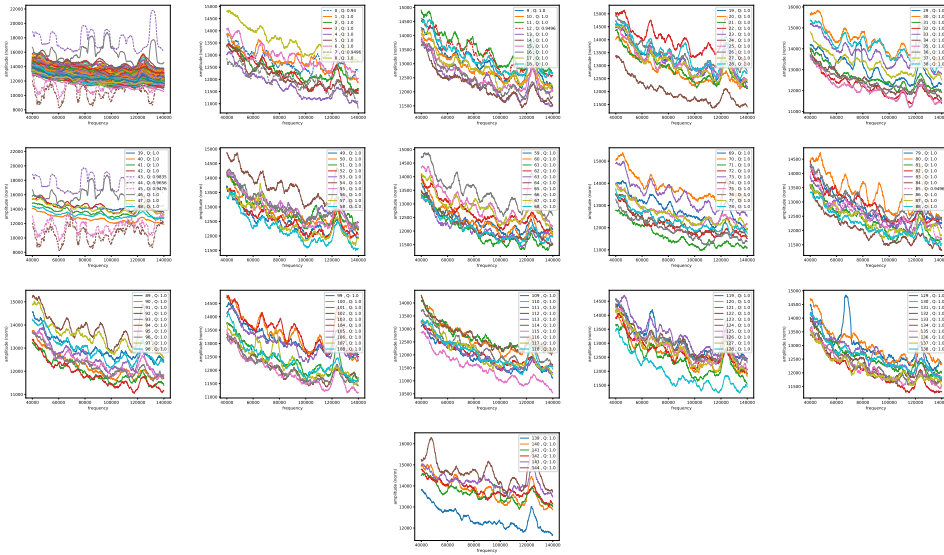


Figure 8: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr8. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

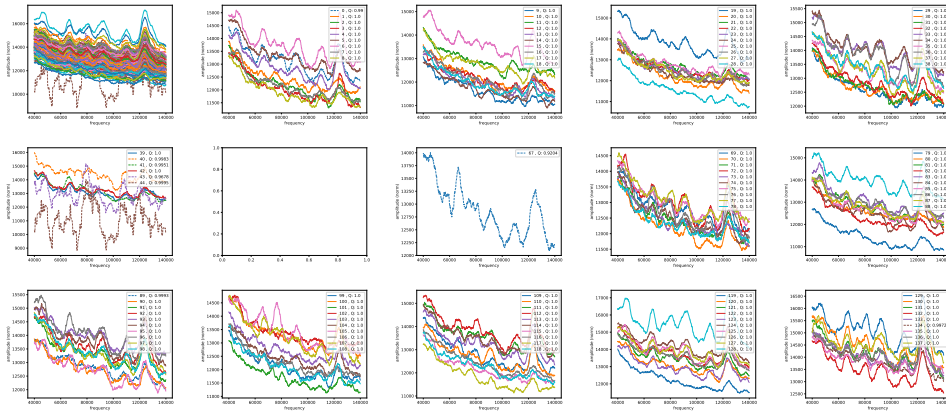


Figure 9: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr9. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

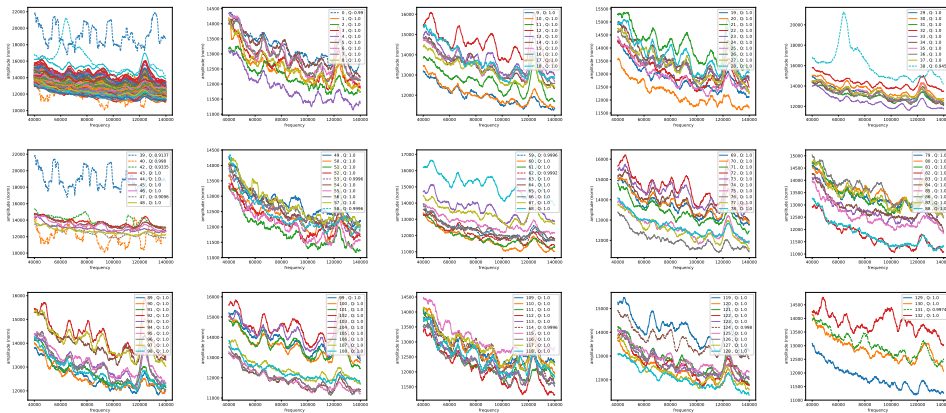


Figure 10: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr10. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.

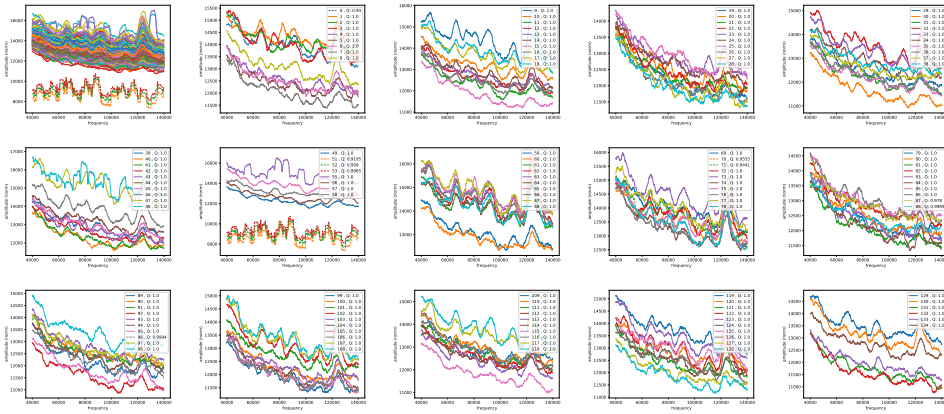


Figure 11: LSTM200. Fouriers on reference-base probability for chromosome: hg38_chr11. The genome string is divided in adjacent segments of 1Mb; several adjacent segments are covered in each plot, segment numbers are indicated in the legend. First plot (upper left) shows the results for all segments in this chromosome.