

# Supplementary IVb: Context dependent prediction in DNA sequence using neural networks. Fouriers, zebrafish genome, LSTM50, high frequencies.

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

April 16, 2022

This file 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 amplitudes seen here are about five times higher than there.

## Fourier plots, LSTM50 predictions, frequency range 40000 to 140000

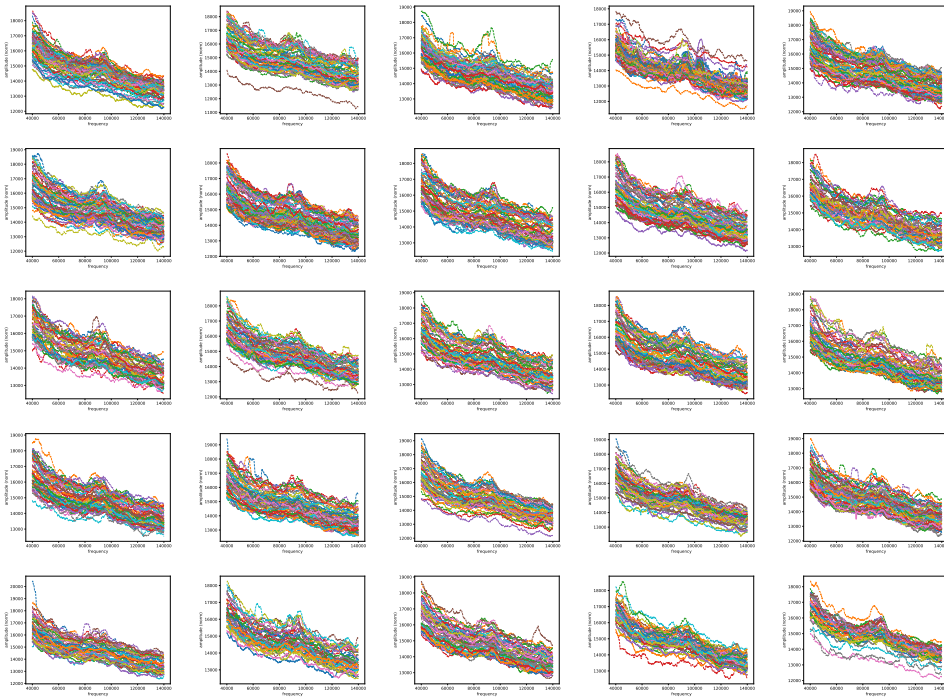


Figure 1: Zebrafish LSTM50. Fouriers on reference-base probability, high frequency range. Each plot covers one chromosome, listed in increasing order (chr1 to chr25). The genome string is divided in adjacent segments of 1Mb (per chromosome); each plot shows the results for all segments in the chromosome (with ratio of qualified positions  $> 0.9$ ).