**Modeling *in vivo* dynamics of RNA Polymerase II meeting Nucleosomes**

Roozbeh Abedini-Nassab\* and Xu Zhang

Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27707, US

**Supplementary Materials**

1. **Simulations**

The simulation is performed using a Matlab code. The algorithms are as follows:





In Algorithm 2, X and T are the inputs for training and desired output of the RBF network, respectively. X*test* and Y are testing data set and the RBF network outputs during the test, respectively.

1. **Extracting Nucleosomes position from MNase data**



**Figure S1. Mapping 3′NT data for CG9246 gene on MNase-seq data.** Using MNase-seq data to define nucleosome position results in RNAP II reads peaks downstream nucleosome dyad, which is not what we expect. Data is taken from [3].

1. **Size optimization for RBF networks**

To find optimal number of kernels in hidden layer of RBF networks (N) and avoid over-fitting we calculated training and prediction errors for different nucleosomes. Results, demonstrated in Figure S2, show that there is a trade-off between training error and prediction error. To avoid over-fitting, prediction error needs to be at the range of training error. Thus, we used N=15 for our simulations in this paper, resulting no over-fitting in simulations for +3 nucleosomes, and reducing it in +1 and +2 nucleosomes. We could use smaller N to decrease over-fitting, even more, in cost of increasing training error. Here, we also provide results of simulations with N=500 for comparison (See Figures S3, S4, and S5). It can be seen that increasing N results in over-fitting.



**Figure S2. RBF Network over-fitting analysis.** Error (%) of training (red) and prediction (green) for (a) +1, (b) +2, and (c) +3 Nucleosomes, for different network size (N).



**Figure S3. Modeling RNA Polymerase II reads when it approaches +1 Nucleosome, based on RBF Network (N=500).** Results of model training (red) and prediction (black) are compared with experimental data (blue) from [3]. (a) Acon, (b) CG31627, (c) CG9246, (d) Mcm10, (e) bur, (f) CG9243, and (g) CG9247.



**Figure S4. Modeling RNA Polymerase II reads when it approaches +2 Nucleosome, based on RBF Network (N=500).** Results of model training (red) and prediction (black) are compared with experimental data (blue) from [3]. (a) Acon, (b) CG9246, (c) Mcm10, (d) bur, and (e) CG9243.



**Figure S5. Modeling RNA Polymerase II reads when it approaches +3 Nucleosome, based on RBF Network (N=500).** Results of model training (red) and prediction (black) are compared with experimental data (blue) from [3]. (a) CG9246, (b) Mcm10, (c) bur, and (d) CG9243.