# Predicting siRNA Sites and Sequences Based on DNA/RNA

## Group Members

Camrie Johnson   
Mica Haney

## Background

Small interfering RNA (siRNA) are small sequences of noncoding RNA around 20-30 nucleotides in length that can regulate gene expression(Dana et al., 2017). This is done by silencing complementary messenger RNA (mRNA). This technology is particularly interesting when designing therapeutics and has been applied to diseases such as COVID-19, cancers, and brain diseases (Chowdhury et al., 2021; Singh et al., 2018; Zheng et al., 2018). However, there are challenges in designing an effective siRNA therapy. Occasionally a design may induce off-target effects by interacting with near perfect mRNA complementary sequences (Ghosal et al., 2012). Computational methods can optimize and predict successful siRNA designs.

## Project Objective

In this study, we aim to create multiple machine learning algorithms to predict effective siRNA sequences. We will be analyzing datasets provided by MIT siRNA library (Sirna at MIT).

## Methods

Data will first be collected via either a manual download or a web scraper to produce a number of DNA or RNA sequences with accompanying siRNA sequences that will bind to an area of the DNA/RNA sequence. Both the DNA/RNA and siRNA sequences will then be processed using a mixture of bioinformatics algorithms as well as natural language processing (NLP) techniques (such as creating probability tables, one-hot encodings, etc.) to convert the sequences into a numerical format that allows for more advanced processing. Following this multiple machine learning (ML) algorithms and models such as Naive Bayes, LSTM, Transformers, and feedforward networks will be used to attempt to generate siRNA sequences based on a given DNA/RNA strand.

## Expected results

The output of the project will be multiple ML models of varying ability to generate siRNA sequences or predict siRNA binding sites, the effectiveness of which will need to be evaluated. If effective, this will create a tool that researchers can use to more precisely target siRNA binding or aid in creating siRNA sequences to reduce experimental overhead.

## References

Chowdhury, U. F., Sharif Shohan, M. U., Hoque, K. I., Beg, M. A., Sharif Siam, M. K., & Moni, M. A. (2021). A computational approach to design potential siRNA molecules as a prospective tool for silencing nucleocapsid phosphoprotein and surface glycoprotein gene of SARS-CoV-2. *Genomics*, *113*(1), 331–343. <https://doi.org/10.1016/j.ygeno.2020.12.021>

Dana, H., Mahmoodi Chalbatani, G., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., Mehmandoost, N., Moazzen, F., Mazraeh, A., Marmari, V., Ebrahimi, M., Menati Rashno, M., Jan Abadi, S., Gharagouzlo, E., & Mar-mari, V. (2017). Molecular Mechanisms and Biological Functions of siRNA. In *Int J Biomed Sci www.ijbs.org 48 INTERNATIONAL JOURNAL of BIOMEDICAL SCIENCE* (Vol. 13, Issue 2). [www.ijbs.org](http://www.ijbs.org)

Ghosal, S., Das, S., & Chakrabarti, J. (2012). Computational Approaches for Designing Efficient and Specific siRNAs. In *Bioinformatics*. InTech. <https://doi.org/10.5772/50125>

Sanketh, R. S. (2021, May 27). *Naive Bayes and LSTM based classifier models*. Medium. <https://towardsdatascience.com/naive-bayes-and-lstm-based-classifier-models-63d521a48c20>

Singh, A., Trivedi, P., & Jain, N. K. (2018). Advances in siRNA delivery in cancer therapy. In *Artificial Cells, Nanomedicine and Biotechnology* (Vol. 46, Issue 2, pp. 274–283). Taylor and Francis Ltd. <https://doi.org/10.1080/21691401.2017.1307210>

Sirna at MIT: Human sirnas. (n.d.-a). <http://web.mit.edu/sirna/sirnas-human.html#>

Wikimedia Foundation. (2023, September 7). *Small interfering RNA*. Wikipedia. <https://en.wikipedia.org/wiki/Small_interfering_RNA>

Zheng, M., Tao, W., Zou, Y., Farokhzad, O. C., & Shi, B. (2018). Nanotechnology-Based Strategies for siRNA Brain Delivery for Disease Therapy. In *Trends in Biotechnology* (Vol. 36, Issue 5, pp. 562–575). Elsevier Ltd. <https://doi.org/10.1016/j.tibtech.2018.01.006>