Quantitative RNA-FISH Probe Design

Jason J. Hughes[1], Ben Kesler[1], and Gregor Neuert[1]

Single-molecule fluorescent in-situ hybridization (smFISH) of RNA molecules is a common method to detect RNA molecules in fixed cells and tissues. This method requires the design of specific short oligonucleotide probes that can hybridize to the RNA of interest. Various research groups and companies have created smFISH probe design software across multiple platforms. However, there has been no comprehensive computational or experimental comparison of smFISH probe design software. We propose a new RNA-FISH probe design software approach using genome-wide gene expression data and BLAST to design RNA-FISH probes and a thermodynamic kinetic model to predict quantitative experimental metrics for comparing FISH probe sets. Using our software, we will compare the efficacy of various probe design software both computationally and experimentally.

***Keywords* — RNA-FISH Probe Design, DNA/RNA Hybridization Thermodynamics, Spatial-Transcriptomics**

I. PURPOSE

Fixed cell single-molecule fluorescent in-situ hybridization (smFISH) experimental approaches such as RNA-FISH and DNA-FISH enable high-resolution interrogation of gene expression and gene regulatory dynamics in single cells [1-2]. These approaches allow the characterization of many aspects of gene regulation. An essential step in using these approaches to study the function of specific DNA and RNA elements is the design of smFISH probe sets. FISH probes are oligonucleotide sequences with high homology to target DNA or RNA sequences conjugated to a fluorophore or tagged with a homologous sequence to a fluorophore-conjugated secondary probe [1,3]. The fluorescent signals generated from FISH probe sets must be specific to the target genes and transcripts to quantify RNA expression distributions correctly while minimizing off-target fluorescent signals that can alter the background signal or cause false-positive FISH spots [4]. However, RNA-FISH probe design is often an overlooked step in building and interpreting smFISH gene regulatory models and experiments.

Several open-source and commercial probe design software packages aim to address this challenge. However, there has been no comprehensive computational and experimental comparison of their efficacy. Various probe design software developers evaluate their software using specific platforms or end-user cases, which might not generalize their software to address the variety of hypotheses about gene regulation and RNA biology. Several factors play a role in determining the efficacy of RNA-FISH probe sets, including genome sequence, thermodynamic and experimental parameters. These include off-target matches to highly expressed transcripts, homologous genes, and increased genome sizes in higher eukaryotes. Probes also need optimal thermodynamic properties such as optimal GC content, melting temperature, and minimal secondary structures. Experimental conditions such as hybridization temperature and salt concentration, as well as differences in expression in different cell-type and cell lines can also affect the efficacy of RNA-FISH probes.

Our approach uses previously established publicly accessible genome-wide cell/tissue-type specific RNA expression datasets and reference genomes to design RNA-FISH probe sets. Our approach tiles the entire target DNA or RNA molecule to evaluate every possible probe's potential to have off-target binding and unfavorable thermodynamic properties. We then use thermodynamic kinetic models to determine how strongly probes will bind these mismatches compared to on-target matches and how likely probes are to form hairpin loops, self-dimers, and cross-dimerize with other probes. After this, we use a mathematical model to predict useful experimental metrics for any given probe set, determining the probability of observing on-target and off-target spots. We then use an algorithm that ranks and selects FISH probes to use experimentally based on their likelihood of off-target binding and thermodynamic properties. We seek to compare our software computationally and experimentally to published RNA-FISH probe design methods for several organisms.

II. CONCLUSION

Our software allows end-users to design smFISH probe sets to study various problems, including mature RNA, 5' vs. 3' end nascent RNA, Intronic RNA, sense-antisense transcription, alternative splicing, allelic variation, and tissue/cell-type specificity.

Acknowledgements: This work was funded by NSF GRFP Grant No. 1937963 and NIH grants DP2GM114849, R01GM115892, and R01GM140240.

[1]Department of Molecular Physiology and Biophysics, Vanderbilt University. Corresponding author E-mail: gregor.neuert@vanderbilt.edu

REFERENCES

[1] Raj, A., et. al. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods*

[2] Chen, K. H., et. al. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*

[3] Tsanov, N., et. al. (2016). smiFISH and FISH-quant–a flexible single RNA detection approach with super-resolution capability. *NAR*

[4] Liu, G., & Zhang, T. (2021). Single Copy Oligonucleotide Fluorescence In Situ Hybridization Probe Design Platforms: Development, Application and Evaluation. *IJMS*