BACKGROUND

Introduction to DNA Computing. In 1994, Adleman first demonstrated that DNA can be used for computing operations,[1] initiating the field of DNA computation. Though current DNA computation devices lack the scalability and computational speed of silicon-based devices,[2] biocomputing systems have distinct advantages.[3] DNA can store 5.5 x 10^{15} petabits per mm^3, which is several orders of magnitude better than comparable electronic devices.[4] Further, DNA computation offers massive parallel processing capabilities; this capability allowed Adleman's initial 1994 experiment with 1,018 DNA strands[1] to compute at 100 teraflops,[2] which outperformed today's best consumer computers by orders of magnitude. Importantly, DNA computation devices can interact with inputs and outputs inaccessible to conventional devices. **Specifically, they can sense and respond to biological and chemical environments, which is an important step towards developing *in vivo* cellular computation, and can interface with genetic regulatory molecules for direct analysis and biological response.**

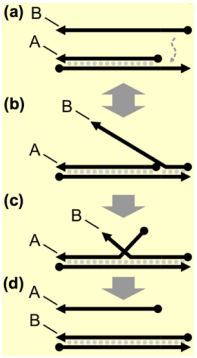

Figure 1. Schematic of toehold mediated strand displacement, showing binding (**a**) and hybridization (**b**) leading to branch migration (**c**) and strand displacement (**d**). Adapted from Chen, X. & Ellington, A.D. *Curr. Opin. Biotechnol.* (2010).

DNA-based devices have been engineered to recognize nucleic acid strands as inputs via sequence-specific interactions, then to respond with outputs that feed into downstream synthetic circuit components.[5] The DNA logic gates that make up these circuits function by toehold-mediated strand displacement. A toehold comprises a short single-stranded DNA sequence (approximately 6 bases) extended from a DNA duplex. The toehold is designed to bind complementary sequences on other strands (**Fig. 1A**), resulting in hybridization (**Fig. 1B**), branch migration (**Fig. 1C**), and strand displacement (**Fig. 1D**). The first step – binding of the input oligo to the toehold sequence – generally determines the kinetics of the process. Although it typically occurs rapidly using 6 DNA bases (k_1 = 10^6 M^{-1} s^{-1}),[6] the rate varies widely with concentration, sequence, and toehold length.[7] Total computation time increases substantially with increasing circuit size.[8] We expect to improve the overall kinetics of these processes in **Aim 2** of this proposal.

Introduction to microRNAs. Our group recently discovered that DNA logic gates are functional in human cells[9] and that they can differentiate between cell types based on patterns of microRNA (miRNA) expression (see **Preliminary Results**). MiRNAs are single-stranded RNA molecules of approximately 22 nucleotides in length.[10] MiRNAs control an estimated 60% of human genes[11] and they govern fundamental cellular processes, including development, differentiation, proliferation, survival, and death.[12] Dysregulation of miRNAs' switch-like behavior can lead to uncontrolled cellular proliferation (i.e., cancer) and is best understood by precisely analyzing miRNA networks and patterns.[13] Existing methods for analyzing miRNA patterns – such as microarrays, quantitative real-time PCR (qRT-PCR), and RNA sequencing[14,15] – require complicated sample workup, instrumentation, and data processing. Importantly, these methods' outputs cannot interface with biological systems, e.g., optical markers or small molecule inhibitors as responses to miRNA signatures. DNA computation in live cells is uniquely suited to analyze miRNA patterns and to respond with defined outputs such as fluorescence or small molecule activation. **This proposal will advance cellular DNA computation of miRNA patterns and interfacing with the biological environment.** We **hypothesize** that detection, computation, and biological response functions can be combined in DNA-based computation circuity ***in live cells***.

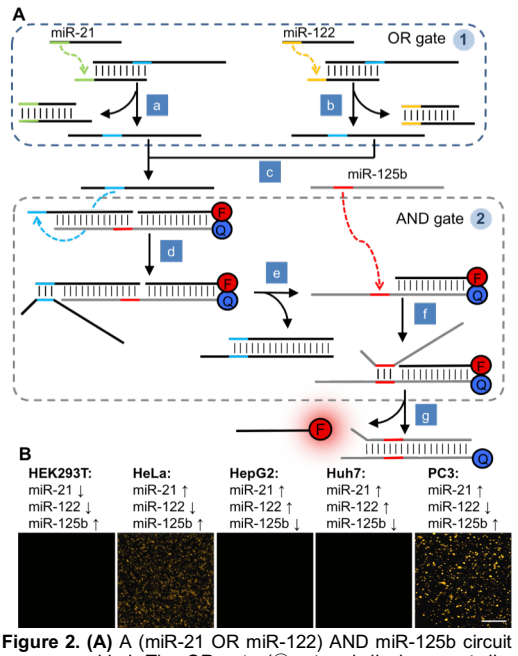

Figure 2. (A) A (miR-21 OR miR-122) AND miR-125b circuit was assembled. The OR gate (①; *strand displacement diagrams for different gates are labeled with circled numbers throughout this proposal*) delivers an identical output (steps a & b) in response to miR-21 or -122. The output serves as an input to AND gate ② by unmasking a toehold (steps d & e) to which miR-125b can bind (step f). The final step (g) releases a single-stranded oligo that becomes fluorescent (λ_{ex} = 532 nm). **(B)** Successful DNA circuit use in live cells. Only cells expressing (miR-21 OR miR-122) AND miR-125b were fluorescently labeled, as in the case of HeLa and PC3 cells. The logic gates were assembled by combining purchased oligos (IDT) in a test tube and transfecting into the five cell lines, then imaging after 4 h (the signal is detectable after 1.5 h as well). F = TAMRA, Q = BHQ-2.

PRELIMINARY RESULTS

This proposal builds on our group's strong foundation designing approaches for controlling DNA logic gates[16] and miRNAs[17,18] as well as our work on DNA computation with small molecule outputs[19] and in live human cells.[9] Recently, we developed an OR/AND logic gate circuit for detecting patterns of miRNA expression in mammalian cells. We used an internal fluorophore:quencher pair to generate the readout (**Fig. 2**). Here, an upstream OR gate feeds into an AND gate that displaces the fluorophore from the quencher to produce a fluorescent readout. This circuit detected miRNA signatures in five different cell lines, successfully reporting the cell lines that match (miR-21 OR miR-122) AND miR-125 logic. The concentration of miR-21 and miR-122 is estimated to be only 5-10 nM (1-2 $\times$ 10^4 copies/cell),[20–22] demonstrating high sensitivity. Gate duplexes were stable in human cells and plasma, and no degradation was observed within 24 h. As the gate designs are simple and modular, they can integrate into large circuits of increasing complexity.

RESEARCH PLAN

Aim 1. Detect complex cellular miRNA patterns using Boolean logic operations encoded in DNA and respond with fluorescent and small molecule outputs. We have identified miRNA patterns reported in the literature, which – based on our **preliminary results** – will allow us to test increasingly complex DNA circuit designs in live cells. A distinct pattern of four up- and down-regulated miRNAs[23] occurs in the cell lines EKVX and HOP92[24] (available from the NCI Biological Testing Branch): (miR-21↑ AND -201↑) NOT (miR-126↓ OR -486↓). Both of these lung cancer cell lines (adenocarcinoma and large cell carcinoma, respectively) come from the NCI-60 cancer cell line panel and are thus key targets for improved diagnostics and treatments.[24] Our DNA circuit design – an AND/NOT/OR circuit – leverages a first layer of an AND gate and an OR gate providing the inputs for a second layer INHIBIT gate (**Fig. 3A** and **3B**). We will tune toehold length[6,7] to ensure OR gate computation (2-3 nucleotides longer) before AND gate computation in order to prevent premature reporter activation. EKVX and HOP92 cells will be differentiated from HEK293T, Huh7, and HeLa cells, which do not express this particular pattern of miRNAs. We will image the fluorescence output, as shown in **Fig. 2**, and will quantify via cell counting (ImageJ), flow cytometry (University of Pittsburgh Cancer Institute, of which my advisor, Dr. Alexander Deiters, is a member), and plate reader measurements. If the miRNA concentrations are too low to generate a measureable signal, the output will be amplified using a hybridization chain reaction (HCR). HCR and related designs can detect nucleic acids down to 0.1 nM and can function in complex biological environments;[25,26] we recently applied HCR to signal amplification DNA computation,[27] and will extend it here using the polymerization of pyrene fluorophore-labeled hairpin monomers.

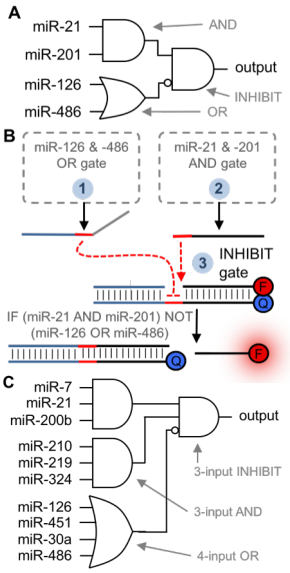

Figure 3. (A) Circuit diagram for recognizing a four-miRNA pattern. **(B)** Circuit design in which an OR gate (①) provides an output that inhibits the reporter gate ③ if miR-126 or miR-486 is present. Otherwise, the presence of miR-21 and miR-201 will trigger AND gate ② to release an output that activates fluorescence by displacing the quencher. The AND gate for miR-21 and miR-201 will follow the strand-exchange scheme in **Fig. 2** with redesigned sequences. **(C)** Circuit diagram for recognizing a ten-miRNA pattern. The strand exchange reaction simply increases the number of inputs for each gate.

To further advance the complexity of the miRNA signatures and corresponding circuit designs, we will detect an alternative pattern of 10 miRNAs in EKVX and HOP92 cells (**Fig. 3C**): (miR-7-2↑ AND miR-21↑ AND miR-200b↑ AND miR-210↑ AND miR-219↑ AND miR-324↑) NOT (miR-126↓ OR miR-451↓ OR miR-30a↓ OR miR-486↓).[28] Analysis using miRBase revealed no significant sequence homology between the miRNA targets in this aim and other human miRNAs, minimizing the potential for false positives. Sequences will be designed manually (as we and others have done before), although if problems occur (e.g., oligo cross-reactivity in larger circuits), we will apply *in silico* modeling tools (e.g., NuPack or Visual DSD).[29,30]

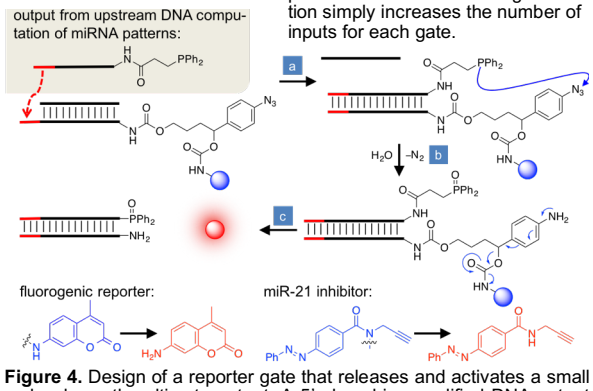

Figure 4. Design of a reporter gate that releases and activates a small molecule as the ultimate output. A 5' phosphine-modified DNA output from the miRNA computation circuit undergoes a strand-displacement reaction with a 3' azide-linker modified gate (step a). The close proximity and thus high local concentration of the phosphine and the azide induces a Staudinger reduction (step b). The resulting well-documented fragmentation reaction[32,43] releases and activates the small molecule (step c; blue → red sphere), yielding a fluorophore or miR-21 inhibitor.

Our group recently published the first DNA computation circuits that release small molecules as outputs.[19] We will build on this discovery to develop – for the first time – DNA computation circuitry that releases small molecule outputs ***in live cells***. Our approach applies a biorthogonal Staudinger reduction of an azide by a phosphine. The reaction partners are kept separate (i.e., at low concentration) until DNA hybridization brings them into close proximity and dramatically increases the local concentration. Related reactions have been used extensively in bioconjugations in cells and animals[31] – including in our group[32] – and are non-toxic and orthogonal to cellular chemistries.[33] We will validate the approach by releasing our published latent fluorophore in cells;[19] next, we will activate small molecules that modify cellular function by releasing one of our discovered miR-21 inhibitors (**Fig. 4**).[18] This will generate a negative feedback loop, which we will verify by qRT-PCR measurements of miR-21. Thus, the development of this methodology may provide a fundamentally new approach to precisely releasing bioactive molecules in response to cell state recognition through biocomputing.

Aim 2. Develop PNA-based biological computation that performs faster and is fully insulated from cellular systems. With increasing circuit complexity comes an increased risk of leakiness leading to background signal in the absence of miRNA inputs. Leakiness results from: (1) premature gate activation from stochastic DNA "breathing" of sequestered toeholds, (2) non-specific interactions with endogenous DNA/RNA, (3) oligonucleotide defects introduced during synthesis, and (4) degradation during extended computations.[34–36] **To preemptively address all of these potential causes of leakiness, while increasing computing kinetics and decreasing the size of the logic gates, we will use peptide nucleic acids (PNA) as the material for encoding and performing biological computing.** PNAs have unnatural polyamide linkages that prevent degradation by proteases and nucleases,[37] allowing for extended intracellular computations without risk of degradation. Specifically, we will use γPNAs, which

are conformationally preorganized into a helix because a chiral center is installed in the γ-backbone (**Fig. 5A**).[38] Depending on the absolute configuration, either left-handed (LH) or right-handed (RH) γPNAs can be generated; native PNAs are achiral and non-helical (NH).[37] Whereas **NH-PNA strongly binds both LH-γPNA and natural RNA/DNA[37] and can therefore serve as a bridge between these nucleic acids, LH-γPNA is conformationally mismatched and cannot recognize natural RNA/DNA,[39] insulating it from biological oligonucleotides and the cellular environment.** NH-PNA can participate in in toehold-mediated strand displacement reactions with DNA in a test tube;[40] extending this process to living cells and using it to translate signals to LH-γPNA is a novel and promising leap for DNA computation.

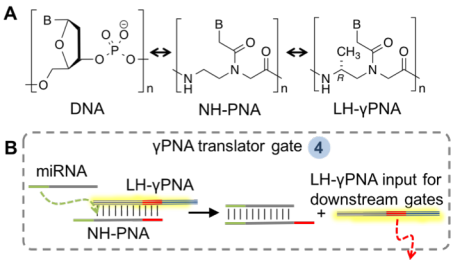

Figure 5. (A) Structures of DNA, PNA, and γPNA monomers. Altering the stereochemistry at the γ carbon to the *S* configuration gives RH-γPNA. B = nucleobase (A, T, G, C); n = oligomer length. **(B)** Strand-exchange diagram for an miRNA–to–LH- γPNA translator gate. PNAs will be transfected using Endo-Porter[44] (Gene-Tools, Inc) or by appending short transduction peptides.[45]

We will determine the strand displacement kinetics and specificity for NH:LH- and LH:LH-γPNA duplexes to determine design rules for oligos (specifically, ideal lengths of toehold and branch migration domains). We expect to be able to halve overall oligo lengths based on expected gate stability and sequence specificity gains,[37] minimizing the potential for errors in synthesis. Enhanced hybridization efficiency per base-pair and the resulting increased free-energy release is expected to improve strand-displacement kinetics and overall circuit processing, particularly for larger multi-gate circuits.[6,41] We will convert the miRNA pattern computation circuits tested in **Aim 1** into γPNA circuits by simply converting miRNA inputs to LH-γPNA oligos using translator gates (**Fig. 5B**) before carrying out any computations. We will validate circuit designs in test tube experiments, and then transition them into cell-based experiments as we have done before.[9] PNAs and γPNAs are commercially available (Panagene, PNA Bio Inc., and PNA Innovations). Should any γPNA-specific issues arise, we will consult with experts (Drs. Ly and Armitage) at Carnegie Mellon University's Center for Nucleic Acids Science and Technology (CNAST), of which Dr. Deiters is a member.

Relevance to DoD Goals. The present proposal will develop fundamental, broadly-applicable design rules for translating increasingly complex DNA computation circuits into live mammalian cells. Therefore, it will advance the Office of Naval Research's goal to "extend the natural capabilities of living organisms… using synthetic biology" by "improving tools needed to advance the field such as… novel information processing circuits… *in vitro* and in cells; and an expanded toolkit of orthogonal, validated genetic regulatory elements."[42] Additionally, analysis of miRNA patterns associated with lung cancer (**Aim 1**) enables fundamental leaps in our understanding of the disease and its diagnostic markers in line with the DoD Congressionally Directed Medical Research Program (http://cdmrp.army.mil/lcrp/default). This work lays the foundation for rapid, sensitive detection of combinations of pathogens and vectors in the field; intraoperative labeling of solid tumor margins for complete resections followed by post-operative surveillance; and precision medicine in which combinations of biomarkers trigger drug release.

References. 1. Adleman, L.M. *Science.* (2007) *266*, 1021. **2**. Parker, J. *EMBO Rep.* (2003) *4*, 7. **3**. Benenson, Y. *Nat. Rev. Genet.* (2012) *13*, 455. **4**. Church, G.M. et al. *Science.* (2012) *337*, 1628. **5**. Chen, X. & Ellington, A.D. *Curr. Opin. Biotechnol.* (2010) *21*, 392. **6**. Srinivas, N. et al. *Nucleic Acids Res.* (2013) *41*, 10641. **7**. Zhang, D.Y. & Winfree, E. *J. Am. Chem. Soc.* (2009) *131*, 17303. **8**. Seelig, G. et al. *Science.* (2006) *314*, 1585. **9**. Hemphill, J. & Deiters, A. *J. Am. Chem. Soc.* (2013) *135*, 10512. **10**. Wahlestedt, C. *Nat. Rev. Drug Discov.* (2013) *12*, 433. **11**. del C. Monroig, P. et al. *Adv. Drug Deliv. Rev.* (2015) *81*, 104. **12**. Garzon, R. et al. *Nat. Rev. Drug Discov.* (2010) *9*, 775. **13**. Bandyopadhyay, S. et al. *Silence* (2010) *1*, **14**. Mattie, M.D. et al. *Mol. Cancer* (2006) *5*, **15**. Chen, J. et al. *Nucleic Acids Res.* (2008) *36*, e87. **16**. Prokup, A. et al. *J. Am. Chem. Soc.* (2012) *134*, 3810. **17**. Connelly, C.M. et al. *Mol. Biosyst.* (2012) *8*, 2987. **18**. Gumireddy, K. et al. *Angew. Chemie Int. Ed.* (2008) *47*, 7482. **19**. Morihiro, K. et al. *J. Am. Chem. Soc.* (2017) *139*, 13909. **20**. Zhao, L. et al. *NMR Biomed.* (2008) *21*, 159. **21**. Chang, J. et al. *RNA Biol.* (2005) *2*, 17. **22**. Cermelli, S. et al. *PLoS One* (2011) *6*, e23937. **23**. Shen, J. et al. *Lab. Investig.* (2011) *91*, 579. **24**. Reinhold, W.C. et al. *Cancer Res.* (2012) *72*, 3499. **25**. Choi, H.M.T. et al. *Nat. Biotechnol.* (2010) *28*, 1208. **26**. Jiang, Y. (Sherry) et al. *J. Am. Chem. Soc.* (2013) *135*, 7430. **27**. Prokup, A. et al. *ACS Synth. Biol.* (2015) *4*, 1064. **28**. Boeri, M. et al. *Proc. Natl. Acad. Sci.* (2011) *108*, 3713. **29**. Zadeh, J.N. et al. *J. Comput. Chem.* (2011) *32*, 170. **30**. Lakin, M.R. et al. *Bioinformatics* (2011) *27*, 3211. **31**. Sletten, E.M. & Bertozzi, C.R. *Angew. Chemie Int. Ed.* (2009) *48*, 6974. **32**. Luo, J. et al. *Nat. Chem.* (2016) *8*, 1027. **33**. Sletten, E.M. & Bertozzi, C.R. *Acc. Chem. Res.* (2011) *44*, 666. **34**. Qian, L. et al. *Nature* (2011) *475*, 368. **35**. Chen, X. et al. *Proc. Natl. Acad. Sci.* (2013) *110*, 5386. **36**. Jiang, Y.S. et al. *Angew. Chemie Int. Ed.* (2014) *53*, 1845. **37**. Sacui, I. et al. *J. Am. Chem. Soc.* (2015) *137*, 8603. **38**. Dragulescu-Andrasi, A. et al. *J. Am. Chem. Soc.* (2006) *128*, 10258. **39**. Sahu, B. et al. *J. Org. Chem.* (2011) *76*, 5614. **40**. Kabza, A.M. et al. *J. Am. Chem. Soc.* (2017) *139*, 17715. **41**. Benenson, Y. et al. *Nature* (2004) *429*, 423. **42**. Office of Naval Research, *Code 34, Div. 342 Synthetic Biology for Naval Applications*. **43**. Alouane, A. et al. *Angew. Chemie Int. Ed.* (2015) *54*, 7492. **44**. Summerton, J.E. et al. *Ann. New York Acad. Sci.* (2005) *1058*, 62. **45**. Shiraishi, T. & Nielsen, P.E. *Methods Mol. Biol.* (2014) *1050*, 193.