BOISE STATE UNIVERSITY

IGEM # 19-002: Nucleic Acid Memory

July 1, 2018 – June 30, 2019 Annual Report

Will Hughes
Chad Watson
Elton Graugnard
Tim Andersen
Reza Zadegan
Eric Hayden
Wan Kuang
George Dickinson
Will Clay

I. Project Summary

The Idaho Global Entrepreneurial Mission (IGEM) and State Board of Education Higher Education Research Council (HERC) have provided three years of funding to help meet emerging state economic development, research, and workforce needs in the area of Nucleic Acid Memory (NAM). This report summarizes the activities during the first year of the project.

II. Project Overview

In 2016, the digital universe produced 16 ZB (1 ZB = 1 trillion GB) of data. In 2025 it will create 163 ZB. These data, once generated, cascade through the information lifecycle — from primary storage media in the form of hard disks and solid-state drives to archival media such as tape. While the semiconductor industry maximizes the density, stability, and energy efficiency of electronic and magnetic memory, both are fast approaching their physical and economic finish lines. As envisioned by the new Semiconductor Synthetic Biology Roadmap, DNA-based massive information storage is a fresh start for memory manufacturing in the United States. According to our study with Micron, Harvard, and the Semiconductor Research Corporation (SRC), DNA has a retention time that ranges from thousands to millions of years, 1 kg of DNA can store the projected digital universe in 2040, and DNA's energy of operation is 100 million times less than current electronic memory. As a result, nucleic acid memory has become a global conversation, a national investment, an industrial opportunity, and a local strength in Idaho.

Our vision is to pioneer a digital data storage paradigm in Idaho by designing, building, and testing accessible, editable, and non-volatile nucleic acid memory (NAM) technologies that are inspired by DNA circuits and made possible by our innovations in DNA nanotechnology. With support from IGEM-HERC, we are creating a Nucleic Acid Memory Institute to meet critical innovation, economic, and workforce development needs in Idaho. To expedite our vision of Idaho becoming a global leader in NAM, five tasks will be met over the life of the IGEM-HERC: **Task 1** – Create efficient algorithms for coding information into data strands. Error correction strategies will account for DNA insertions, deletions, and substitutions, as well as screen for biological sequences to ensure that the data has no genetic function. **Task 2** – Create a high-throughput, integrated analytical engine to design and select data strands using quantitative metrics based on an in-house, algorithm. **Task 3** – Create synthetic biological factories for manufacturing DNA scaffolds using rapid design-build-test cycles of genomes. Genome size and structure will be engineered. **Task 4** – Design and fabricate NAM storage platforms using the DNA scaffolds, and validate the functionality of genome scaffolds using atomic force microscopy. **Task 5** – Read arbitrary data files into NAM storage nodes using super-resolution microscopy. Realize sub-nanometer imaging resolution to enable high areal density data storage.

This annual report spans July 1, 2018 to June 30, 2019 and hence is inclusive of our prior progress report from July 1, 2018 to January 1, 2019. Listed below is a summary of our accomplishments to date.

III. Summary of project accomplishments

Task 1 – Create improved algorithms for coding information into data strands.

T1.1 Kelsey Suyehira successfully defended her Master of Science in Computer Science in September 2018. Her thesis topic was entitled, *Using DNA for Data Storage: Encoding and Decoding Algorithm Development*. Briefly described, when encoding binary data into sequences of DNA, algorithms should account for biological constraints representing the idiosyncrasies of working with nucleic acids. In response, Kelsey created the REDNAM software package (a.k.a. Robust Encoding and Decoding of Nucleic Acid Memory). REDNAM includes a novel-mapping scheme that converts digital information into codons while accounting for important constraints when working with DNA. For example, it removes biologically active codes—such as start codons and some known promoter regions—avoids multiple repeats of unique nucleotides, and excludes repeating sequence strings. In doing so, Kelsey developed a schema mimicking how information has evolved to be efficiently encoded into natural DNA while also accounting for the errors that often arise when working with synthetic DNA. She also integrated her mapping scheme into a fountain code in an implementation that balanced information density with error correction. The result is that REDNAM recovers 100% of its data in spite of introducing random errors into the DNA. It also achieved a speed up of 2x for encoding and 435x for decoding digital information when compared to state-of-the-art fountain codes found in the literature. As shown below, Kelsey's thesis resulted in one publication and two conference presentations that established a foundation for this award, with two more publications that are in preparation.

- K. Suyehira, S. Llewellyn, R.M. Zadegan, W.L. Hughes, T. Anderson, "A Coding Scheme for Nucleic Acid Memory (NAM)," IEEE Workshop on Microelectronic and Electron Devices, pp 1-3, 2017.
- R.M. Zadegan, K. Suyehira, S. Llewellyn, T. Andersen, W.L. Hughes, A Coding Scheme for Digital Data Storage in Nucleic Acid Memory (NAM), DNA 23, (September 2017), Austin, TX, USA.
- R.M. Zadegan, K. Suyehira, S. Llewellyn, T. Andersen, W.L. Hughes, A Biologically Inspired Coding Scheme for Nucleic Acid Memory, FNANO, (April 2017), Snowbird, UT, USA.

T1.2 Drs. Reza Zadegan and George Dickinson created the first proofs-of-concept that digital Nucleic Acid Memory (dNAM) is possible by encoding and decoding "NAM" and "IGEM" onto blocks of DNA that are 98 nm x 71 nm. **Figure 1** shows how an ASCII file—that was converted into a binary file—can be encoded onto a DNA block. The DNA block was optically decoded via a super-resolution microscope and error-correction was performed using hard encoded indexing and structure averaging of the resolved images.

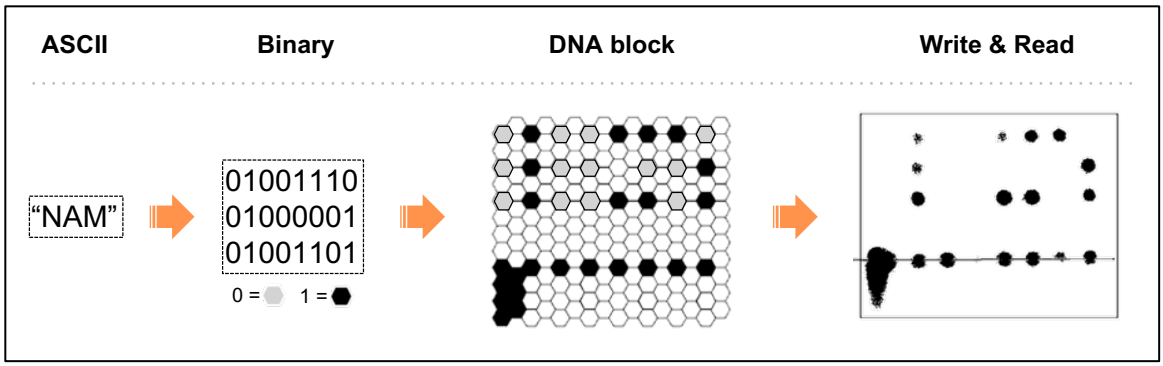

Figure 1. Implementation scheme and data for ASCII encoding of digital information into dNAM with hard encoded indexing and structure averaging for error correction. “NAM” is written in ASCII on a DNA block made from DNA origami. The rightmost image is a multi-structure average of super-resolution images of DNA blocks with the encoded ASCII data and index. Structure averaging constitutes error correction through redundant structures. The DNA block and super-resolved images are ~98 nm x 71 nm.

Figure 2 is a super-resolved DNA origami block that includes a hard-encoded index. The image physically reveals, and experimentally validates, that the binary data “IGEM” can be written and read without error. Error correction was performed via averaging multiple super-resolution microscopy images together—a process known as structure averaging. For scale, the DNA origami block (the image) is ~98 nm x 71 nm in dimensions.

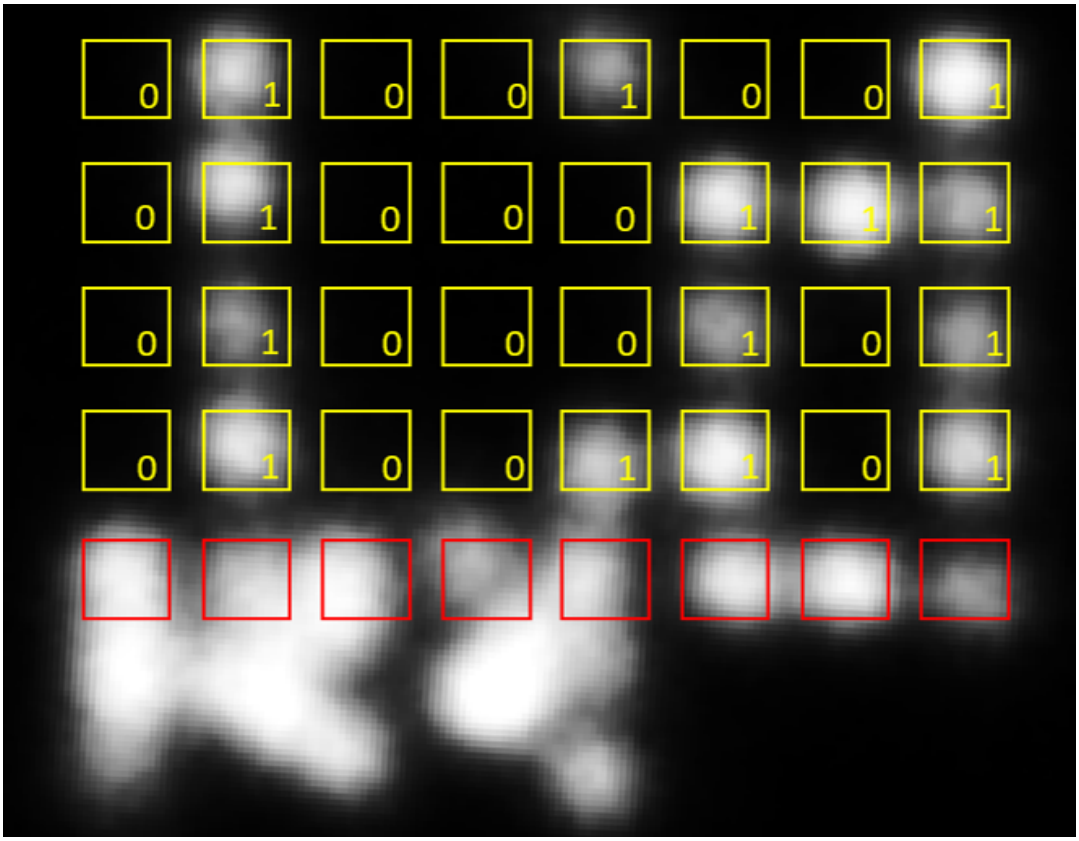

Figure 2. Super-resolved DNA origami block with a hard encoded index. The image includes structure averaging and reveals the binary data “IGEM”. The DNA block is ~98 nm x 71 nm.

Task 2 – Create a high-throughput, integrated analytical engine to design select data strands using quantitative metrics based on an in-house, evolutionary algorithm.

T2.1 Michael Tobiason is completing his PhD in the Micron School of Materials Science & Engineering. With an expected graduation date of December 2019, his dissertation topic is entitled, *Engineering Kinetically Uniform DNA Devices*. Briefly described, the relationship between DNA sequence and the rate of DNA reactions is not well understood in the literature. As highlighted in the rightmost image in **Figure 3**, sequence and hence kinetic variability among DNA strands *may* contribute to site-to-site variability when performing super-resolution microscopy on DNA blocks.

In response, Mike has hypothesized that observed kinetic variations in the literature arise due to unintentional base pairing in DNA. He has found that ranking model DNA devices based first on the size (in base-pairs) of the largest unintentional structure and then the count (number of structures of this size) reliably identifies sequences with improved kinetic reproducibility (**Figure 3**). Mike did this by creating and then validating a software package called Device Profiler (*DevPro*) that selects kinetically reproducible sequences in the literature via his Total-Fitness Score.

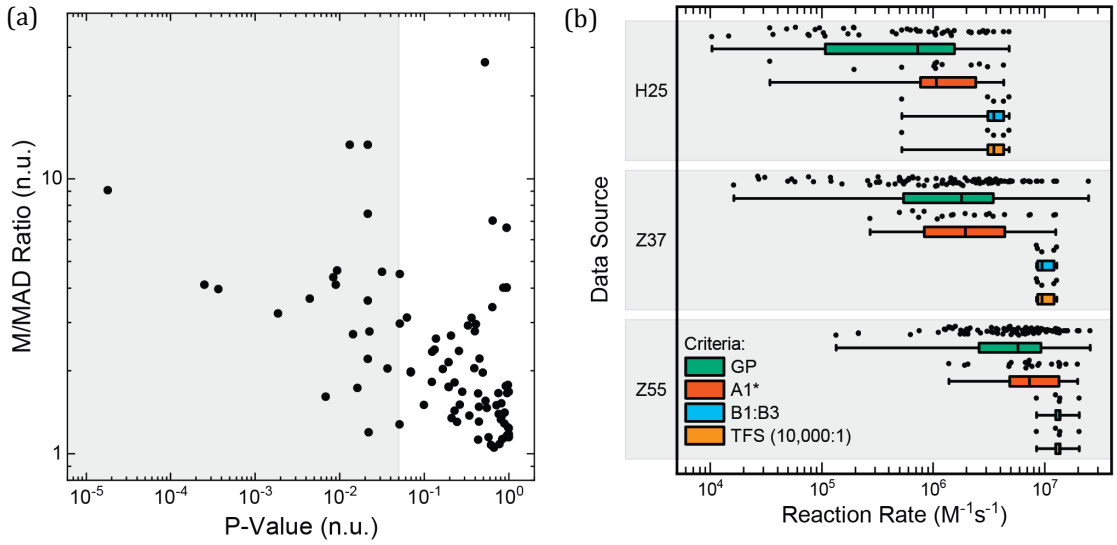

Figure 3. The sequences reported in 3 self-consistent sets of experimental data were analyzed using DevPro. The reaction rates of each top-quality population were statistically analyzed and summarized using two values: The P-value resulting from a comparison of the reaction rates of top-quality devices to the rates of the entire dataset (referred to as the general population) via a two-sample Kolmogorov-Smirnov test, and the ratio of the Median to the Median-Absolute-Deviation (M/MAD). **(a)** The P-values and M/MAD ratio calculated for each data point were used to identify criteria exhibiting both a statistically-significant (P-value < 0.05, shaded area) impact on device kinetics and kinetic reproducibility when compared to the general population. **(b)** Reaction-rates of top-quality devices resulting from the application of three selected criteria to one of the three datasets. Reaction rates from three sources are analyzed: Hata *et al.* at 25°C (1, labeled H25), Zhang et al. at 37°C (2, labeled Z37), and Zhang *et al.* at 55°C (2, labeled Z55). The unfiltered general

population (GP) of each dataset is presented for reference. The first criteria, A1, selects devices based on the length of the longest interference event present and is equivalent to the state-of-the-art sequence-symmetry tools in the literature (*). The second criterion, B1:B3, selects devices based first on the length of the longest intramolecular interference event, and then based on the number of intramolecular interference events of this length. The third criterion, TFS, is the Total-Fitness-Score metric and is a slight variation from B1:B3.

To engineer DNA devices based on his Total-Fitness Score, Mike created and validated an evolutionary algorithm called Sequence Evolver (*SeqEvo*) (**Figure 4**). Through large and small random mutations, sequences evolve in *SeqEvo* through multiple cycles and only sequences with a better (lower) fitness score are retained for the next cycle. Stated more clearly, *SeqEvo* optimizes DNA sequences to minimize their inter- and intra-molecular interactions and in doing so increases their engineering reliability.

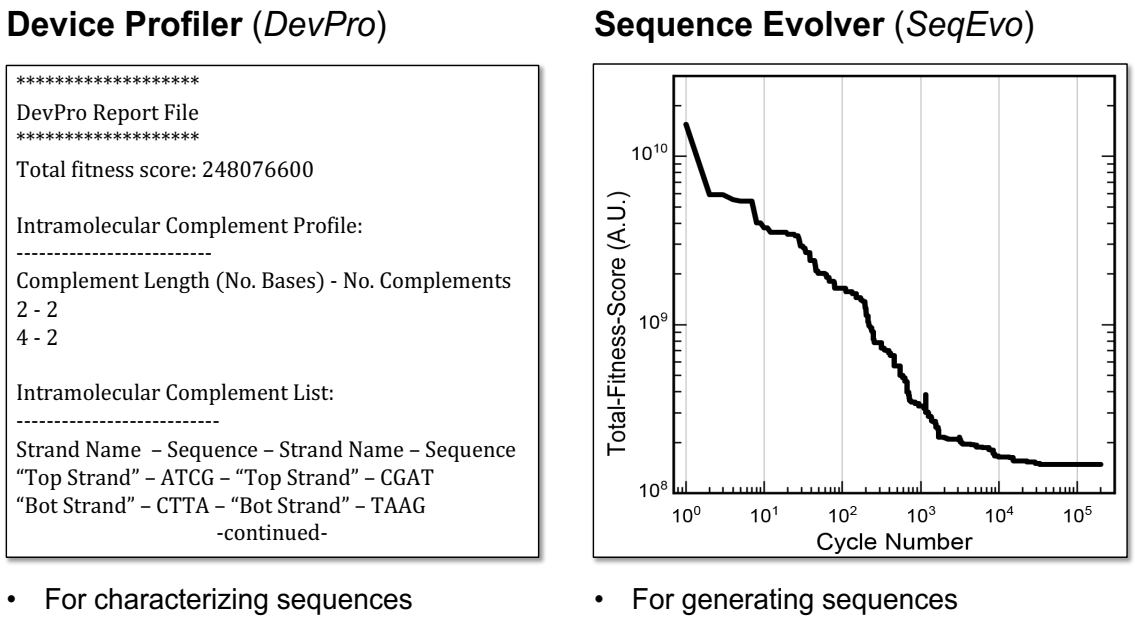

Figure 4. Examples of developed computational tools *DevPro* sequence analyzer and *SeqEvo* sequence optimizer. Through random mutations, sequences evolve through multiple cycles and only sequences with a better (lower) fitness score are retained for the next cycle.

SeqEvo theoretically out-performs all known sequence generation tools at minimizing the number and length of interference events. In support of this claim, Mike tasked *SeqEvo* and its competitors with generating single 35 base-pair duplexes; which is the smallest duplex that requires the introduction of a three base-pair long interference event, and is thus expected to highlight performance differences between the tools. For each method, default parameters were applied in three independent trials. The interference profiles of the resulting eighteen devices were characterized using the *DevPro* software. The median interference profile (as judged by the Total Fitness Score) was selected for each design method and is reported in **Table 1**.

	Interference Type, Length, and Count									
	Intramolecular Length (bp)			Intermolecular Length (bp)						
Sequence Source	**4**	**3**	**2**	**8**	**7**	**6**	**5**	**4**	**3**	**2**
SeqEvo (10,000:1)	0	0	0	0	0	0	6	36	90	176
SeqEvo (1:1)	0	0	34	0	0	0	0	0	2	154
Domain Design	0	8	54	0	0	0	2	14	36	218
EGNAS	0	8	58	0	0	0	0	0	30	204
DSG	0	10	66	0	0	0	0	0	36	206
Random Sequences	4	16	64	0	0	0	10	50	202	250
Uniquimer3D	4	26	86	2	4	8	12	22	72	250
NUPACK	6	22	70	0	0	2	6	30	76	232

Table 1. Typical interference profiles of the 35 bp duplexes generated using various sequence-generation tools. Interference structures within each device were characterized using the *DevPro* software and represent a typical design based on three independent design trials. Interference structures are binned based on their length (bold column headings), with the number of interferences of each given length reported in the table. Intramolecular and intermolecular interferences are reported separately in the left and right groupings of columns. To demonstrate the effects of parameter tuning, SeqEvo was characterized using two sets of scoring weights (10,000:1 and 1:1).

Building upon the theoretical performance of *SeqEvo*, Mike experimentally validated the kinetic reproducibility of his tool by directly comparing randomly generated sequences to optimized sequences (**Figure 5**).

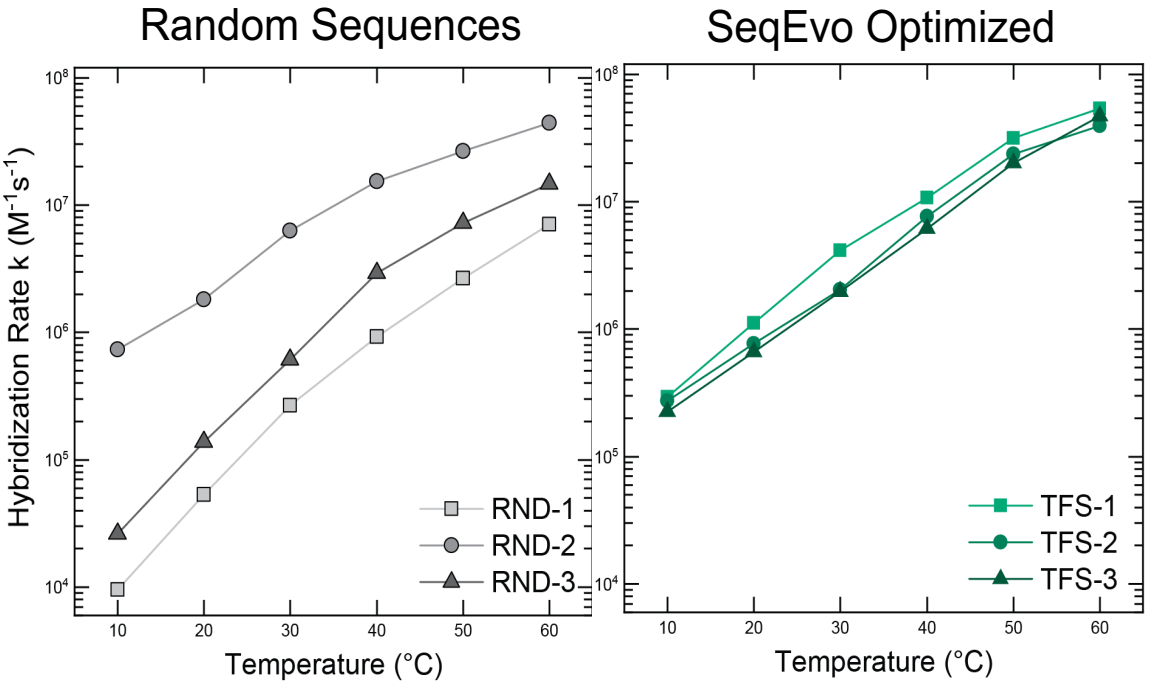

Figure 5. Example optimization of sequences using *SeqEvo*. In this implementation, total sequence fitness was optimized to reduce the variability of DNA hybridization rates and greater predictability of DNA interactions, which will be critical for throughput in dNAM and seqNAM.

By engineering DNA devices with favorable interference profiles using his software packages, Mike experimentally demonstrated that DNA kinetics vary by a factor of two or less when his sequences satisfy four conditions: (1) no intramolecular interferences longer than 2 base-pairs, (2) no intermolecular interferences longer than 4 base-pairs, (3) no stretches of consecutive cytosines or guanines longer than 3 base-pairs, and (4) no stretches of consecutive adenines or thymines longer than 6 base-pairs. Taken together, his findings support the hypothesis that kinetic variation arise due to interfering events and that kinetic reproducibility is possible through sequence optimization. These insights will help inform the design, build, and test cycle of Nucleic Acid Memory at Boise State. To date, they have contributed to two poster presentations and one oral presentation. A manuscript is in preparation.

- M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Reproducible DNA Devices, *6th Annual Conference on Foundations of Nanoscience: Self-Assembled Architectures and Devices (FNANO19)*, Snowbird, UT. **Poster Presentation**.
- M. Tobiason, B. Yurke, and W.L. Hughes (2019). Engineering Kinetically Reproducible DNA Devices, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*, College Park, MD. **Poster Presentation**.
- C.M. Green, M. Tobiason, R.M. Zadegan, W.L. Hughes (2019). Nucleic Acid Memory, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*, College Park, MD. **Oral Presentation**.

Task 3 – Create a synthetic biological factory for manufacturing DNA scaffolds using a rapid design, build, and test cycle of genomes.

T3.1 Steven Burden is completing his PhD in Biomolecular Sciences and is expected to graduate December 2019. His dissertation topic is the development of nucleic acid biosensors with allosteric fluorescence signals. With mentorship from Co-PI Hayden, Steven has established the experimental and instructional infrastructure in support of over 10 undergraduate students participating in a NAM-centric Vertically Integrated Project (VIP). The undergraduate students have developed eight training modules on synthetic biology concepts needed to understand, design, and produce customizable single-stranded DNA from phagemids in E. coli. The training module format is based on a watch-one, do-one, teach-one approach, fostering both peer-to-mentor and peer-to-peer learning.

T3.2 Building on this workforce development effort, four of the VIP students were hired as undergraduate researchers this summer. One of these summer students was awarded an INBRE fellowship based off their VIP work and academic performance. All of these students are working on designing and building DNA plasmids needed to produce custom scaffolds for DNA origami. To date, they have produced 3 kilobase and 6 kilobase single-stranded DNA scaffolds. Working in parallel, and in preparation for student training, Dr. Natalya Hallstrom adopted/adapted asymmetric polymerase chain reactions to produce single-stranded DNA up to 50 kilobases in length.

Task 4 – Design and fabricate NAM storage nodes using the DNA scaffolds.

T4.1 Sadao Takabayashi is attempting to complete his PhD in the Micron School of Materials Science and Engineering while working full time at Micron. His dissertation topic is *Patterning and Fabricating with DNA*. Foundational to this IGEM-HERC award, Sadao demonstrated high density and selective adsorption of DNA origami onto boron implanted silicon substrates made by Micron Technology, which resulted in the below listed publication.

- S. Takabayashi, S. Kotani, J. Flores-Estrada, E. Spears, J.E. Padilla, L. Godwin, E. Graugnard, W. Kuang, S. Sills, W.L. Hughes, "Boron-Implanted Silicon Substrates for Physical Adsorption of DNA Origami," International Journal of Molecular Sciences, vol 19, issue 9, number 2513, pp 1-12, 2018.

Leveraging this publication, DNA origami array precursors were dispersed onto silicon dioxide wafers composed of 5µm × 5µm boron-implanted silicon patterns (**Figure 6**). Upon self-assembly, Sadao observed that the boron-implanted silicon was consistently covered with arrays, even though DNA origami were not detected on the silicon dioxide. This is valuable when reflecting on how one might site-specifically deposit dNAM onto semiconductor-grade substrates in preparation for super-resolution microscopy.

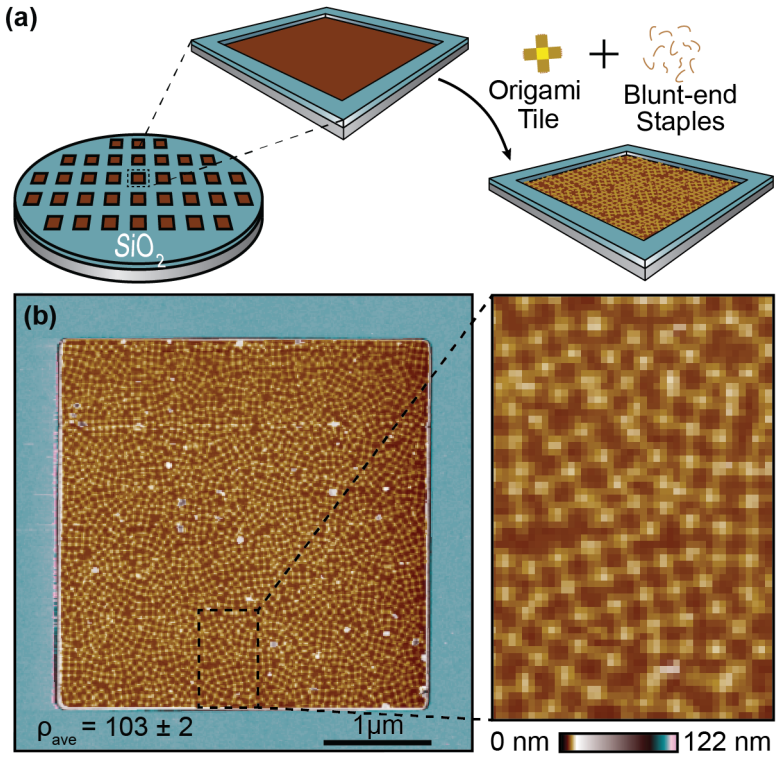

Figure 6. Selective DNA origami array assembly on 5 x 5µm boron-implanted silicon (brown) surrounded by raised silicon dioxide (blue). **(a)** Schematic of the pre-patterned substrate and DNA origami arrays. **(b)** AFM image of the origami arrays assembled onto the boron-implanted substrates. The DNA arrays with visible grain boundaries are fully covered by the boron-implanted silicon surface.

Sadao has also observed that surface adsorption is inversely proportional to the pattern feature size, and the smaller the pattern, the more pronounced the effect (**Figure 7**). This is important when reflecting on how the surface density of dNAM may vary based on its local environment.

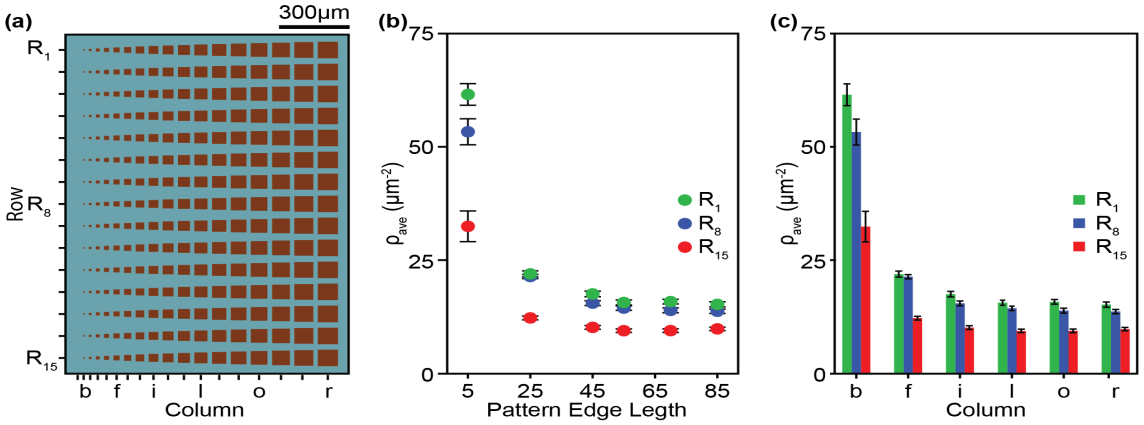

Figure 7. **(a)** A schematic of boron implanted patterns (brown) that are carved in silicon dioxide (blue). Each of the 15 rows consist of 18 square patterns with their edge length range from 5μm to 85μm. **(b)** Origami surface density as a function of the boron-implanted silicon pattern size. The green, blue, and red dots are surface density of boron-implanted silicon features in R_1, R_8, and R_{15}, respectively. The surface density increases exponentially as pattern size decreased. **(c)** Origami surface density plotted versus columns. The green, blue, and red bars are the origami surface density of the features in R_1, R_8, and R_{15}, respectively. The DNA origami surface density was consistently greater for the patterns R_1, followed by R_8 and then R_{15}. The error bars indicate 95% confidence limit.

T4.2 As outlined in section T1.2, the research team has prototyped digital NAM (dNAM) structures. Contributing to this effort, Chris Green is completing his PhD in the Micron School of Materials Science & Engineering and is expected to graduate December 2019. Among other things, his dissertation topic correlates atomic force microscopy with super-resolution microscopy to study defects. As shown in **Figure 8**, correlated images allow Chris to explore structure-property relationships in dNAM. For example, incorporated yet inaccessible or defective DNA data strands are responsible for errors in super-resolution microscopy, and thus digital Nucleic Acid Memory. Equally as important, Chris created a new protocol to immobilize DNA nanostructures onto glass without needing protein binding (BSA-biotin). Similar to the insights gained from Mike Tobiason, when optimizing DNA sequences via *SeqEvo*, the insights gained from Chris will help inform the design, build, and test cycle of Nucleic Acid Memory. To date, they have contributed one oral presentation already listed and one additional poster presentation captured here. A manuscript on Chris' work is in preparation.

- C.M. Green, G. Dickinson, R.M. Zadegan, W.L. Hughes, E. Graugnard, W. Kuang (2019). Correlative metrology and defect analysis of DNA origami, *Semiconductor Research Corporation: Semiconductor Synthetic Biology (SemiSynBio) Annual Review*, College Park, MD. **Poster Presentation**.

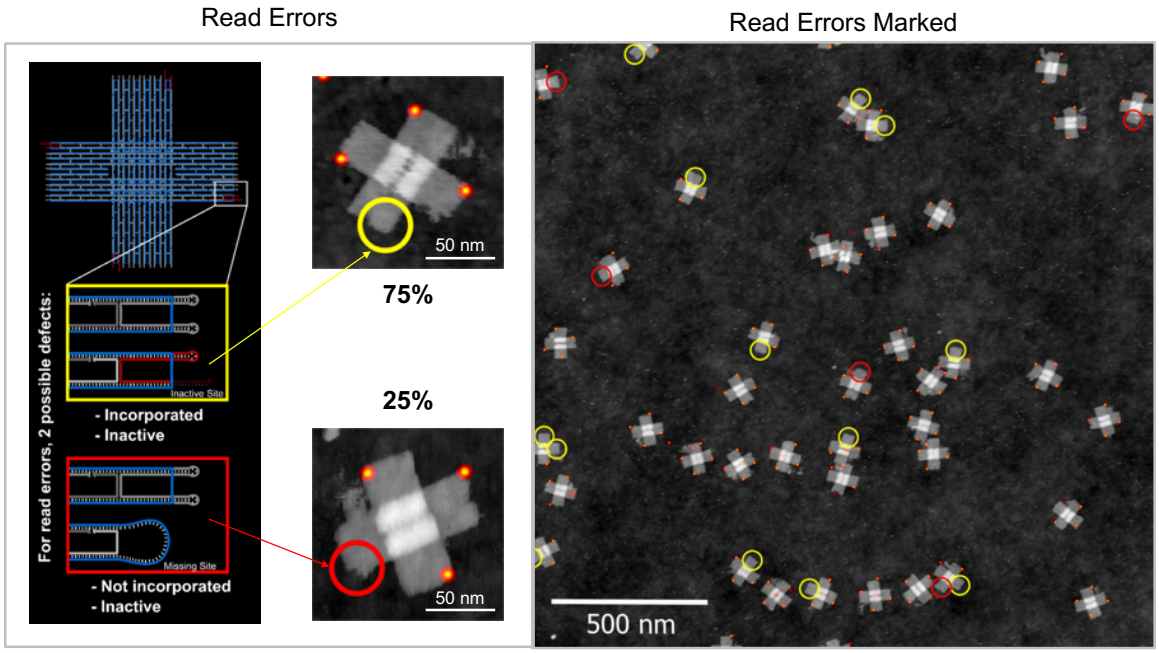

Figure 8. Correlated atomic force microscopy (AFM) and super-resolution microscopy (SRM). The correlated images indicate that unincorporated data strands would also be discernable through AFM. Yet many SRM errors appear structurally intact in AFM. Thus, incorporated yet inaccessible or defective data strands are responsible for errors in SRM, and thus dNAM.

Task 5 – Read arbitrary files into NAM storage nodes using super-resolution microscopy.

T5.1 With expertise in super-resolution microscopy, coding, biology, and imaging, Drs. George Dickinson and William Clay have been recruited as key members of the Nucleic Acid Memory Institute. Dr. Dickinson is currently pushing the resolution of our existing super-resolution microscope, while Dr. Clay is designing a future generation microscope. In collaboration with Chris Green, Dr. Dickinson has implemented methods to reduce thermal drift and photo-oxidation when performing super-resolution microscopy. Alongside Dr. Zadegan, they have designed and imaged calibration standards made from DNA origami. As a team, they have achieved ~5 nm point-to-point resolution using super resolution microscopy (**Figure 9**).

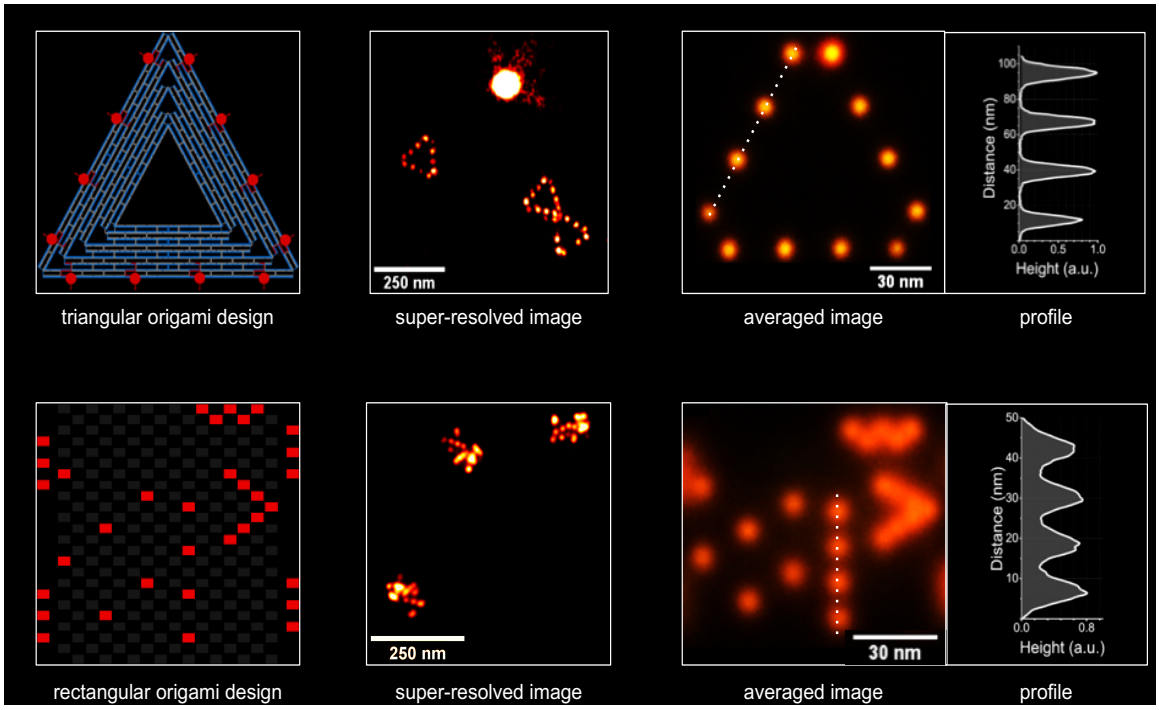

Figure 9. With improved image drift correction and error reduction, sub-5 nm super-resolution imaging was achieved and demonstrated with triangular and rectangular DNA origami test structures, both for individual and averaged structures.

T5.2 In working toward even high resolution, Drs. Dickinson and Clay are taking a two-pronged approach: (1) establishing a new resolution baseline on a recently purchased super-resolution microscope and (2) building an advanced super-resolution microscope designed specifically for imaging dNAM and seqNAM. When combined with our current microscope, we will have a faster design, build, and test cycle at resolutions that are currently not achievable.

IV. Demonstration of economic development and impact

Demonstration of Economic Development and Impact	Number
External Funding	$ 1,549,995
Number of External Grants	3
News Releases	3 articles
Private Sector Engagement	14 companies
University Engagement	11 universities
Federal Agency Engagement	5 agencies
Industry Involvement	2 companies
Patents	0
Copyrights	0
Plant Variety Protection Certificates	0
Technology Licenses Signed	0
Start-up Businesses Started	0
Jobs Created outside of Boise State University	0

After the IGEM-HERC award, the National Science Foundation (NSF) in collaboration with the Semiconductor Research Corporation (SRC) jointly awarded the research team $1,500,000 to address the scientific challenges facing NAM technologies. The funding mechanism was called *Semiconductor Synthetic Biology for Information Processing and Storage Technologies*. Boise State was one of the few universities in the country to receive the prestigious award in the first round of competition. Other awardees included: MIT, Stanford University, University of Washington, and UT Austin. Prior to the release of this award mechanism, Drs. Will Hughes and Reza Zadegan coauthored the Semiconductor Synthetic Biology Roadmap in collaboration with the SRC, which helped steer the federal investments.

In addition to the above listed award, the NSF awarded Hughes $49,995 to host the 2019 Germination Meeting at Boise State University on August 15-16, 2019. With the backing of the NSF Office of Emerging Frontiers and Multidisciplinary Activities, the meeting will discuss new approaches in cultivating risk-taking and impact-driven research culture. This investment follows 2 grants awarded to Hughes—prior to IGEM HERC funding—that helped seed the Nucleic Acid Memory Institute. The awards included: (1) EAGER Germination: Aligning Stakeholders and Structures to Enable Risk Taking ($99,991), and (2) EAGER Germination Renewal: Piloting a Center for Transformative Research at Boise State University ($299,930). When viewed together, the above listed awards have incentivized the NSF to visit Boise State this summer to discuss how to institutionalize a Center for Transformative Research that would be equivalent to our Center for Teaching

and Learning. Albeit indirect, this work is foundational to the future success of research center activity at Boise State University.

Because of the below listed Semiconductor Synthetic Biology (SemiSynBio) consortium, industry involvement on the research project includes Gurtej Sandhu (Micron Technology Vice President) and Victor Zhirnov (SRC Chief Scientist) who jointly serve as the co-chairs of the NAM Institute at Boise State University. According to Gurtej Sandhu, "*the leadership and innovation of this research team has brought them to the threshold of becoming a world class player in the research, development and education of nucleic acid memory*."

Industry Partners (14)	**University Partners (11)**	**Federal Partners (5)**
Autodesk GenoCAD Gingko Bioworks Globalfoundries IBM Intel International Data Corp Mentor Graphics Micron Microsoft Mubadala Technology Raytheon SynBioBeta Twist Biosciences	Boise State Boston University Brigham Young U. Columbia University Dartmouth Georgia Tech NC State University UCLA UIUC UNC Greensboro U. of Washington	- Army Research Office (ARO) - Department of Defense (DoD) - Office of Naval Research (ONR) - National Institute of Standards & Technology (NIST) - National Science Foundation (NSF) - Intelligence Advanced Research Projects Activity (IARPA)

For additional information, below is one forecast report, and three news releases related to our work.

- **Molecular Information STorage (MIST) Market and Forecast**: aka DNA Digital Storage, Nucleic Acid Memory, West Oxford Advisors, 2019.
  https://westoxadvisors.com/mist
- **New NSF awards support the creation of bio-based semiconductors**, Sarah Bates, National Science Foundation, July 16, 2018.
  www.nsf.gov/news/news_summ.jsp?cntn_id=295968&org=NSF
- **How Micron's business could change dramatically from this research at Boise State**, David Staats, Idaho Statesman, September 27, 2018.
  www.idahostatesman.com/news/business/article218442875.html
- **Boise State University awarded $3.5 Million to research storing data on DNA**, Sherry Squires, August 28, 2018.
  https://news.boisestate.edu/update/2018/08/28/boise-state-university-awarded-3-5-million-to-research-storing-data-on-dna

V. Numbers of student, staff, and faculty participation

Classification	Number
Tenured or Tenure Track Faculty	5 (*2 full professors, 3 associate professors*)
Research Faculty	1 (*starting a tenure-track faculty position*)
Project Manager	1 (*also focused on business development*)
Senior Lab Research Associate	1 (*manages the laboratory & supports team*)
Postdoctoral Fellows	2 (*+1 joining in August from Italy*)
Graduate Students	5 (*+1 working full time at Micron*)
Undergraduate Students	10 (*5 female and 5 male*)

Critical to the success of any research initiative are the people that make up the project team. As part of the IGEM-HERC, we have six faculty (Will Hughes, Tim Andersen, Wan Kuang, Elton Graugnard, Eric Haden, and Reza Zadegan), two postdoctoral research scientists (George Dickinson and Will Clay), and a project manager (Chad Watson). Reza Zadegan has been offered a tenure track faculty position at North Carolina A&T and will likely start in August. We have also transitioned three graduate students (Mike Tobiason – Task 2; Steven Burden – Task 3; Chris Green – Task 5) to this project and have recruited and hired two additional graduate students (Shoshi Llewellyn and Golam Md Mortuza – Task 1). In addition, we hired Kelsey Suyehira, who is a recent graduate student from Computer Science that completed her Master of Science on the project. Ms. Suyehira helped transition graduate students focused on Task 1, while also working toward two project-related publications in which she is the lead author. In support of Task 4, we have hired a new postdoctoral research scientist, Luca Piantanida. He is starting on August 5, 2019. His experience developing and characterizing DNA nanostructures brings us closer to meeting our goals. Through VIP, 10 undergraduate students; 5 of which are female are part of this multidisciplinary team as well (**Figure 10**). The VIP model integrates teaching and learning into one framework in support of work-force development of students that can work at the interface of semiconductor manufacturing and synthetic biology. These students are engaging in research activities aimed toward the production, purification, and quality control of new single-stranded DNA origami scaffolds. The students range from freshman to seniors and span four different majors: biology, pre-med, health sciences, and chemistry.

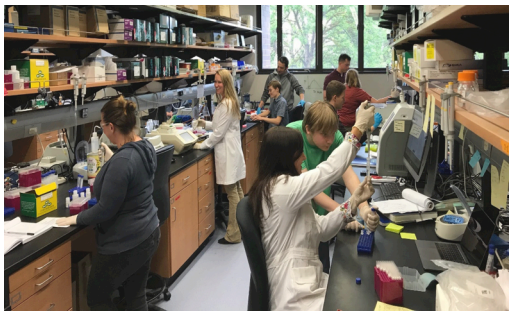

Figure 10. Together, and with mentorship from co-PI Hayden and PhD student Steven Burden, the VIP students are learning basic synthetic biology research skills including DNA primer design and validation in polymerase chain reaction, digital design and sharing of DNA sequences, bacterial transformation and cloning, gel electrophoresis, and DNA quantification and quality control using ultra-violet absorbance.

VI. Description of future plans

Team Management – Integration and graduation

- Support greater integration of project objectives via shared project personnel and increased meeting frequency.
- Support team transitions including the arrival of a third postdoctoral fellow from Italy and the graduation of two PhD students.
- Support the postdoctoral fellows to work as a team. Empower them to take greater intellectual ownership of the day-to-day aspects of the research project, as well as support the professional development of the students on the project.
- Transition the project manager from an internal to a more external-oriented role that includes grant writing, business development, and strategic planning.

Task 1 – Create improved algorithms for coding information into data strands.

- Implement fountain coding with error correction for dNAM, which will form the basis for a first publication of dNAM.
- Experimentally validate the encoding and decoding of digital information into synthetic DNA. Read the information using commercially available sequencing.
- Develop and publish a website for DNA-based encoding/decoding of data.

Task 2 – Create a high-throughput, integrated analytical engine to design select data strands using quantitative metrics based on an in-house, algorithm.

- Publish the initial manuscript describing *DevPro* and *SeqEvo*. Make both software tools publicly available via the GitHub repository and incorporate them into the design workflow for dNAM, starting with imager/docking strands, and then extending to staple and scaffold strands.

Task 3 – Create a synthetic biological factory for manufacturing DNA scaffolds using a rapid design, build, and test cycle of genomes.

- Building on the initial success of the VIP project, graduate and undergraduate students will: (1) select for novel terminator elements to improve scaffold purity, (2) produce unique scaffold sequences that have different base composition and/or are larger than M13mp18, (3) optimize scaffold isolation and purification protocols from bacterial cultures, and (4) develop a physical biobank of materials and data base for inventory.

Task 4 – Design and fabricate NAM storage nodes using the DNA scaffolds.

- DNA origami will be designed for dNAM using in-house scaffolds, and they will be tested using our super-resolution microscope.
- Synthesis yield and defect rates for origami structures will be quantified using super-resolution microscopy. Defect rates for commercial scaffolds (M13mp18) and in-house scaffolds will be compared.
- Initial origami structures will be designed and tested for sequence-based NAM (seqNAM).

Task 5 – Read arbitrary files into NAM storage nodes using super-resolution microscopy.

- Establish new resolution baseline for recently purchased super-resolution microscope.
- Incorporate upgrades to existing super-resolution microscope to improve both its resolution and extend its fluid handling.
- Toward higher resolution, the team will design and build a custom super resolution NAM reading platform capable of sub-5 nm resolution.

VII. Summary of Budget Expenditures

The below table summarizes expenditures associated with the project. O&E has helped support the postdoctoral research scientist searches and the purchase of modified and unmodified DNA oligos. The oligos are used to assemble NAM blocks and to perform super-resolution microscopy studies. Funds were also allocated to purchase polymerase enzymes and primer oligos necessary for asymmetrical Polymerase Chain Reactions (PCR). The research team also purchased a super-resolution microscope from Nikon for $219,957.45. All the capital from the IGEM was used to purchase this microscope; NSF and SRC funding was applied to the outstanding balance of $19,957.45. The remaining budget is a one-time occurrence and reflects an unanticipated delay to recruit and hire internationally competitive postdoctoral fellows at Boise State University; of which two of the fellows are actively contributing to the project and the third will start this summer. The remaining budget is vital to support the 3 fellows and should be rolled into FY20.

Category	Current Budget	Year-to-Date Expenditures	Encumbered	Remaining Budget
Salary	$277,297	$226,396.27	$16,652.16	$34,248.57
Fringe Benefits	$85,528	$67,983.70	$5,173.29	$12,371.01
Graduate Student Tuition and Fees	$33,099	$33,098.65	--	$0.35
Other Expenses	$70,576	$40,033.74	$30,000.00	$542.26
Capital	$200,000	--	$200,000.00	--
Total	$666,500	$367,512.36	$251,825.45	$47,162.19

VIII. Commercialization Revenue

Commercialization	Revenue
None.	$0

IX. Additional metrics established specific to individual project

Metrics	Number
External Funding	$ 1,549,995
Software Tools Created and Initially Validated	3
Master of Science Thesis Awarded	1
Peer-Reviewed Publications	1
Presentations	6 (1 oral and 5 poster)
Manuscripts in Preparation	4
VIP Program Enrollment (grad and undergrad)	10
National and International Postdoc Recruitment	42 (3 hired)

Note: Listed above are specific, objective, measurable, and realistic performance metrics to gauge project success and economic impact, many of which have been distributed throughout this report and are summarized here.

X. References

1. Hata, H., Kitajima, T. and Suyama, A. (2018) Influence of thermodynamically unfavorable secondary structures on DNA hybridization kinetics. *Nucleic Acids Res*, **46**, 782-791.

2. Zhang, J.X., Fang, J.Z., Duan, W., Wu, L.R., Zhang, A.W., Dalchau, N., Yordanov, B., Petersen, R., Phillips, A. and Zhang, D.Y. (2018) Predicting DNA hybridization kinetics from sequence. *Nat Chem*, **10**, 91-98.