Storing digital data in DNA-based tape

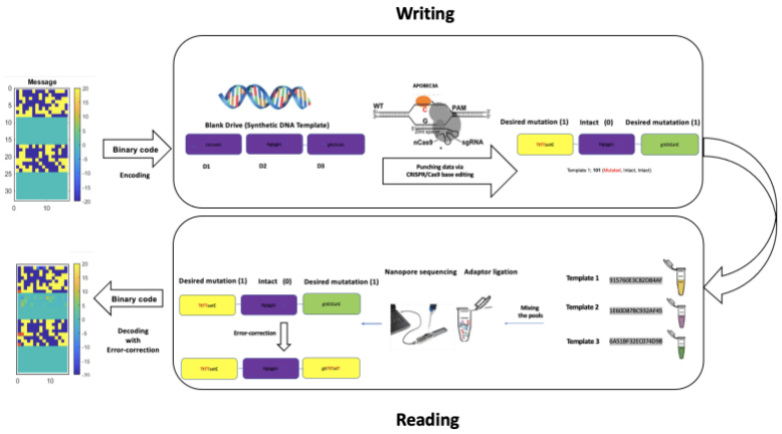
Afsaneh Sadremomtaz[1,3], Robert Glass[2,3], Jorge Guerrero[1,3], Micheal Brandon Reed [1,3], Eric Josephs[2,3]*, Reza Zadegan[1,3]*

[1] Department of Nanoengineering, NC A&T State University, Greensboro, NC, USA

[2] Department of Nanoscience, UNC Greensboro, Greensboro, NC, USA

[3] Joint School of Nanoscience and Nanoengineering, Greensboro, NC, USA

Global digital data will surpass 1.4 Yottabits by 20251. While demands for data storage increases, the conventional digital storage systems are reaching their physical limits2. Therefore, alternative information storage materials including DNA storage systems that could provide viable energetics, spatial capacity, data retention, and economics solutions are becoming a trend[2]. DNA as an alternative for data storage potentially enables great physical density and scalability[1,2]. Additionally, rapid development of DNA synthesis and sequencing has helped the community to consider DNA as an alternative information storage medium[2]. Considerable challenges that need to be addressed for DNA to become a main-stream DNA storage include sustainability, latency, and scalability. While majority of the community are attempting to address scalability, we believe de novo DNA synthesis –which in most cases results in large amounts of toxic waste– is a major bottleneck[1,2]. In this report, we used DNA Mutational Overwriting Storage (DMOS) to overwrite the sequence content of greenly synthesized template DNA domains (bits) to write the digital data. Our DMOS DNA templates contain 16-bit domains and addressing strings. We wrote the digital information by allowing or prohibiting localized mutations the DNA templates via guided synthetic chimeric enzymes (hereby called punchers). Our puncher enzyme (APOBEC3A) recognizes the desired domain by aid of our predefined library of ribonucleic protein complexes (CRISPR-dead Cas9 and guide RNAs)[3]. The developed cell-free punchers perform cytidine deamination of single-stranded DNA strands, resulting in base substitution mutations of base-C (intact state) to base-T (mutated state) in the targeted domains[3]. Also, we developed error-correction and DNA sequencing codes that enable high data retention in our DMOS system. As schematically demonstrated in Fig.1, we performed mutational edits to write the data, added addressing strings to the encoded DNA pools and combined them to store the data, and performed nanopore sequencing to read the data[4].



"This work was funded in part by the National Science Foundation (MCB 2027738)"

1- Zhirnov, V., Zadegan, RM., Sandhu, GS., Church, GM., Hughes, WL. Nucleic acid memory. *Nat Mater*. 2016;15(4):366-370. doi:10.1038/nmat4594.

2- Zhrinov, V., Rasic, D. 2018 Semiconductor Synthetic Biology Roadmap. (Retrieved on 03/14/2020 from https://www.src.org/library/publication/p095387/p095387.pdf).

3- Marshall, R., Maxwell, C. S., Collins, S. P., Jacobsen, T., Luo, M. L., Begemann, M. B., Gray, B. N., January, E., Singer, A., He, Y., Beisel, C. L. & Noireaux, V. Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free TranscriptionTranslation System. Molecular cell 69, 146-157.e143 (2018), doi:10.1016/j.molcel.2017.12.007.

4- Timp, W., Comer, J. & Aksimentiev, A. DNA base-calling from a nanopore using a Viterbi algorithm. Biophysical Journal 102, L37-L39 (2012), doi:10.1016/j.bpj.2012.04.009.