



CENTER of EXCELLENCE
for ENGINEERING BIOLOGY



Pilot Project Proposal
(Not to exceed two pages)

Name of Project: DNA Fountain

Proposer and Contact Information:

Yaniv Erlich, PhD

Assistant Professor of Computer Science, Columbia University

Core Member, New York Genome Center

E-mail: yaniv@cs.columbia.edu

Background:

DNA is an attractive medium to store digital information.

Technical Idea:

Here we report a storage strategy called DNA Fountain that is highly robust and approaches the information capacity per nucleotide. Using our approach, we stored a full computer operating system, movie, and other files with a total of 2.14×10^6 bytes in DNA oligonucleotides and perfectly retrieved the information from a sequencing coverage equivalent to a single tile of Illumina sequencing. We also tested a process that can allow 2.18×10^{15} retrievals using the original DNA sample and were able to perfectly decode the data. Finally, we explored the limit of our architecture in terms of bytes per molecule and obtained a perfect retrieval from a density of 215 petabytes per gram of DNA, orders of magnitude higher than previous reports.