

## Blog

### Blogger Profile



**Dan Kramer** is a graduate of McGill University, pursuing his PhD at UC Berkeley. He has a background in neuroscience and cancer research, and is ready to explore new frontiers in cell biology.

**email:** daniel.kramer@berkeley.edu  
**twitter:** @DJKramesauce

### Recent Posts

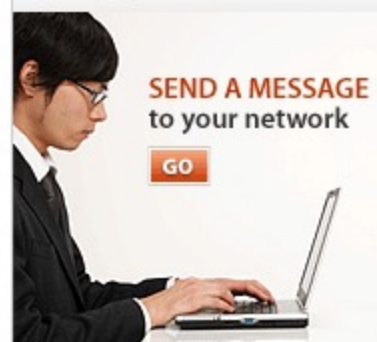
- ▶ **March 11, 2016**  
Rapid RNA analysis at a single cell level: a PL...
- ▶ **February 11, 2016**  
Free to roam: implantable optogenetic devices
- ▶ **January 23, 2016**  
The recipe for human serotonin neurons
- ▶ **January 08, 2016**  
Designing logical cells to be disease sensors

« Prev Next »

### Updates

- ▶ New post in **Green Science**: I leave and heave a sigh and say goodbye
- ▶ New post in **Saltwater Science**: Thanks for the good times!
- ▶ New topic in **Women in Science**: Where to find Laura Hoopes' Women in Science future thoughts?
- ▶ New post in **Saltwater Science**: Microplastic in the ocean
- ▶ New topic in **Women in Science**: LAB GIRL, exciting new woman-in-science biography

### Connect



# Bio 2.0

Dissecting the next revolution in biology



August 31, 2012 | By: **Eric Sawyer**

### Digital Data Storage in DNA

In today's world of high technology, we usually associate digital information with computers, satellites, and smart phones. But don't forget that digital information storage far predates the digital age. When Watson and Crick discovered the structure of DNA, they also discovered that at the heart of living things is a massive digital information storage molecule. The digital information in the human genome encodes the instructions for building the human body that would ultimately go on to build digital computers.

Bringing this full circle, a recent paper<sup>1</sup> explores the possibility of using synthetic DNA to store digital information of human design. We have a data problem. The future of science seems to be big data, and that's challenging and expensive to store. The CMS experiment at the Large Hadron Collider alone generates 1 terabyte of data every second and, more poetically, in its projected lifetime the LHC will give us a dataset comparable in size to a library containing every single word spoken by every human who has ever lived<sup>2</sup>. Big science aside, the internet and social media generate large quantities of digital information.

In their paper, George Church and colleagues combine recent advances in DNA synthesis and sequencing to create a data storage chip made of DNA. In the DNA chip they stored a HTML file of a **synthetic biology book**, containing 53,426 words of text, 11 black and white JPEG images, and a computer program written in Java. Since DNA contains four bases, two computer bits can be encoded per DNA base: e.g., A = 00, C = 01, G = 10, T = 11. The authors of this paper however decided to use a one base per bit code, A = 0, C = 0, G = 1, T = 1, which allows for many synonymous encodings to avoid sequences that are difficult to sequence such as long repeats.

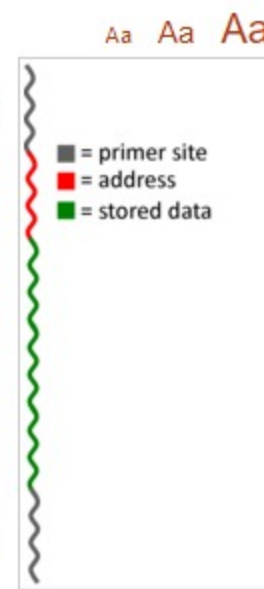
In all, they synthesized 54,898 short pieces of DNA. Each contained 2 primer binding sites for PCR amplification and sequencing, a unique 19-base address, and 96 bases (96 bits, or 12 bytes) of stored data (see the figure). Their DNA chip represents the densest information storage (nearly 10<sup>15</sup> bits/mm<sup>3</sup>) to date, beating the likes of *in vivo* experiments, flash memory, hard disks, and CDs. The error rate was an impressively low 10 bits out of the total 5.27 million.

The prospect of storing something like particle accelerator data on a synthetic DNA chip is intriguing and seductively poetic, but I'm a bit skeptical. The interface between computer chips and DNA chips will never be as seamless as the interface between computer chips and hard disks, flash memory storage devices, etc. And even though DNA sequencing is getting much cheaper, and even more portable, it seems unlikely that individuals would purchase sequencers. But maybe DNA will be used for long-term archiving as the authors suggest, where the need for high capacity, high longevity storage outweighs the costly and time-consuming process of writing and reading the stored data.

**Image credit:** Me.

#### Reference:

1. Church, G. M., Gao, Y., & Kosuri, S. **Next-Generation Digital Information Storage in DNA**. *Science [Science Express]*. Published online August 16, 2012. Also see [online supplement](#).
2. LHC Factoids. Cosmic Variance.



1 Comment

### COMMENTS

September 02, 2012 | 10:25 AM  
great research

Posted By: **SEO India**