

You are cordially invited to a talk in the **Edmond J. Safra Center for Bioinformatics Distinguished Speaker Series**.

The speaker is **Yaniv Erlich**, Assistant Professor of Computer Science, Columbia University and Core Member, New York Genome Center

**Title:** "Genetic Media"

**Time:** Wednesday, **December 21** 2016, at **11:15** sharp (refreshments from 11:00)

**Place:** Schreiber 309, School of Computer Science

**Host:** Dr. Oded Rechavi, [odedrechavi@gmail.com](mailto:odedrechavi@gmail.com), Department of Neurobiology, Life Sciences Faculty

**Abstract:** In the last decade, the human population has produced zettabytes ( $10^{21}$ ) of digital data. This creates immense opportunities and challenges for biology research. In this talk, I will present two research directions of my groups on the intersection between genetics and data, which we dub "genetic media".

Part I: I will speak about crowd sourcing massive genetic data using social media. We collected over 80 million profiles from the largest social-media website driven by genealogy. Using this data, we constructed a single family tree of 13-million individuals that spans tens of generations and every country in the Western world. We used this pedigree to analyze the genetic architecture of longevity and understand the effect of dominance and epistasis. I will speak about our on-going efforts to fuse genomes and social media phenotypes to this massive pedigree.

Part II: I will present our new study on using synthetic DNA as a medium for long-term data storage. Previous studies in leading journal have presented this concept but failed to show reliable data retrieval. Here, we report a storage strategy, called DNA Fountain, that is highly robust and approaches nearly doubles the information capacity per nucleotide compared to previous studies. To demonstrate its power, we stored a full computer operating system, movie, and other files with a total of 2.14Mbyte of data in DNA oligos and perfectly retrieved the information. We also tested a process that can allow  $2.18 \times 10^{15}$  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 molecules and obtained a perfect retrieval from a density of 215Petabyte/gram of DNA, orders of magnitudes higher than previous techniques.