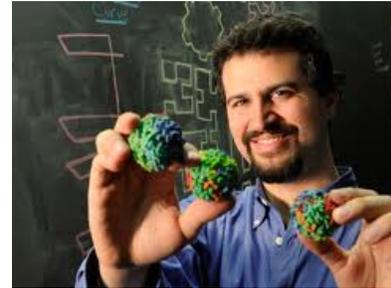


**Research Seminar by
Erez Lieberman Aiden
Fellow, Harvard Society of Fellows
and Visiting Faculty at Google
Monday, March 19, 2012
Union South (TITU)
Noon-1:00 pm**



Title: Zooming Out: New Tools for Probing the Historical Record and the Human Genome

Abstract: New structures often emerge when we explore a known phenomenon from a more global vantage point. For instance, any given book can be read and comprehended. But what happens when we try to read all the books at once? Or: the local structure of DNA is a double helix. But if DNA did not fold further, the human genome - which is two meters long - could never fit inside the nucleus of a cell. How does it fold? This talk will focus on the extraordinary potential of technologies that enable us to zoom out, in the process transforming familiar concepts, like the contents of a book or the shape of DNA, into new research horizons.

First, I will describe efforts, together with my collaborator Jean-Baptiste Michel and Google, to create tools for the quantitative analysis of a significant portion of the historical record. We began by constructing a reliable corpus of digitized texts containing about 4% of all books ever printed. Analysis of this corpus enables us to investigate cultural trends quantitatively. 'Culturomics' provides insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology. The Google Ngram Viewer, a simple web-based tool we released for the analysis of this corpus, was used over a million times in the first 24 hours. Culturomics extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities.

In the second half of my talk, I will describe Hi-C, a novel technology for probing the three-dimensional architecture of whole genomes. Developed together with collaborators at the Broad Institute and UMass Medical School, Hi-C couples proximity-dependent DNA ligation and massively parallel sequencing. My lab employs Hi-C to construct spatial proximity maps of the human genome. Hi-C maps have revealed that active and inactive portions of the human genome are spatially segregated, i.e., that cells employ a sort of 'regulatory origami' as they turn genes on and off. At the megabase scale, the genomic fold is consistent with a fractal globule, a knot-free conformation that enables maximally dense packing while preserving the ability to easily fold and unfold any genomic locus.

Erez Lieberman Aiden is a fellow at the Harvard Society of Fellows and Visiting Faculty at Google. His work integrates mathematical and physical theory with the invention of new technologies.

He recently invented a method for three-dimensional genome sequencing; he subsequently led the team that, in 2009, reported the first three dimensional map of the human genome. Together with collaborator Jean-Baptiste Michel, he developed culturomics, a quantitative approach to the study of history and culture that relies on computational analysis of a significant fraction of the historical record. This work led to the creation of the Google Ngram Viewer, a supplemental website that was visited over a million times in the 24 hours after its launch.

Erez's research has won numerous awards, including a \$2.5M NIH New Innovator Award; the GE & Science Prize for Young Life Scientists; the Lemelson-MIT prize for the best student inventor at MIT; the American Physical Society's Award for the Best Doctoral Dissertation in Biological Physics; recognition for one of the top 20 "Biotech Breakthroughs that will Change Medicine", by Popular Mechanics; and membership in Technology Review's 2009 TR35, recognizing the top 35 innovators under 35. His last three research articles have all appeared on the cover of Nature and Science. His work has also been featured on the front page of the New York Times, the Boston Globe, and the Wall Street Journal.