ETH zürich

ICB seminar series 2018/19
chairman: Prof. Dr. Chih-Jen Shih

GENETIC MEDIA

Prof. Dr. Yaniv Erlich

Chief Science Officer of MyHeritage.com and
Associate Professor of Computer Science and
Computational Biology at Columbia University

ETH Hönggerberg, HCI J 7
Wednesday, 17/04/2019, 17.00 h

The Seminar will be followed by an Apéro

Abstract: This talk will explore the interplay between DNA and big data. In the first part, will talk about storing data in DNA. DNA is an attractive medium to store digital information and I will report our method to record information using a strategy called DNA Fountain. Using our approach, we stored 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. 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. In the second part, I will present a new study about the limitation of genetic privacy. Consumer genomics databases have reached the scale of millions of individuals. Recently, law enforcement authorities have exploited some of these databases to identify suspects via distant familial relatives. Using genomic data of 1.28 million individuals tested with consumer genomics, we investigated the power of this technique. We project that over 60% of the searches for individuals of European-descent will result in a third cousin or closer match, which can allow their identification using demographic identifiers. Moreover, the technique could implicate nearly any US-individual of European-descent in the near future. We demonstrate that the technique can also identify research participants of a public sequencing project. Based on these results, we propose a potential mitigation strategy and policy implications to human subject research.

Bio: Dr. Yaniv Erlich is the Chief Science Officer of MyHeritage.com and an Associate Professor of Computer Science and Computational Biology at Columbia University (leave of absence). Prior to these positions, he was a Fellow at the Whitehead Institute, MIT. Dr. Erlich received his bachelor's degree from Tel-Aviv University, Israel (2006) and a PhD from the Watson School of Biological Sciences at Cold Spring Harbor Laboratory (2010). Dr. Erlich's research interests are computational human genetics. Dr. Erlich is a TEDMED speaker (2018), the recipient of DARPA's Young Faculty Award (2017), the Burroughs Wellcome Career Award (2013), Harold M. Weintraub award (2010), the IEEE/ACM-CS HPC award (2008), and he was selected as one of 2010 Tomorrow's PIs team of Genome Technology.