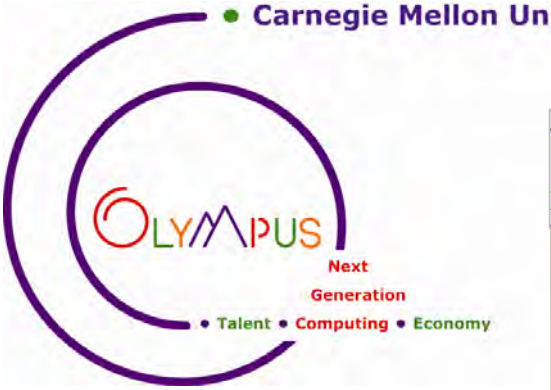

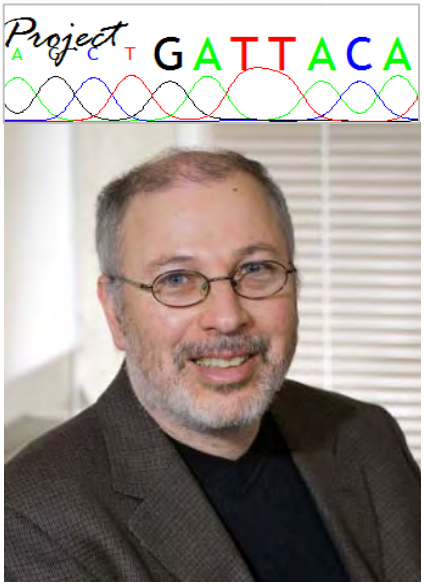

Roni Rosenfeld
Professor, School of Computer Science
Carnegie Mellon University
roni@cs.cmu.edu
(412) 268-7678
http://www.cs.cmu.edu/~roni
http://www.cs.cmu.edu/~gattaca/

Vaccine Design in the Computer Lab

In the 1990s, it took $10 billion and 10 years to sequence the human genome. A few years later, it took only $50 million and a fraction of the time to sequence the comparable-size chimpanzee genome. In 2007, a similar-sized project cost $1 million and lasted less than a year. NIH has recently set a goal of $1,000 per human genome. As affordable, routine genome sequencing is quickly becoming a reality, the amount of sequenced DNA will continue to grow exponentially for at least another decade or two. Existing tools for modeling bio-sequences were not designed for the 1010 nucleotides/Amino Acids currently available, nor the 10^{12}--10^{15} nts/AAs anticipated in the foreseeable future. The development of new algorithms, visualization tools, and predictive models will change not only medical practice, via personalized diagnostics and treatment, but also the nature and pace of biomedical investigation, drug discovery, and public health decision making.

This is what **project GATTACA** sets out to do. We focus on RNA viruses and other fast evolving pathogens. We build large scale computational models of important viral proteins (like HIV's Env and RT, or Influenza's hemagglutinin and neuraminidase). We then develop algorithms to infer molecular correlates of important viral properties such as drug-resistance, pathogenicity, antigenicity, immunogenicity, virulence, infectivity, neutralizability, etc. Our algorithms make concrete predictions that can be verified experimentally. We also build models of viral molecular evolution, attempting to anticipate the future behavior and properties of the virus. In parallel, we design and build visualization tools to support interactive exploration of very large biosequence alignments.