**Analysis of combinatorial cis-regulation**

**Dr. Barak Cohen**

*Alvin Goldfarb Distinguished Professor of Computational Biology, Professor of Genetics*

The majority of signals from human genetic studies reside in non-coding DNA. Presumably, a large fraction of these variants exert their effects by influencing the functions of cis-regulatory elements. An important goal is to understand the sequence features that comprise cis-regulatory elements to the point where we can predict the effects of sequence variants on gene expression. Transcription Factor Binding Sites (TFBS) are key components of cis-regulatory elements, however, because TFBS are short, degenerate sequences they occur millions of times in mammalian genomes. This leads to a major problem: The vast majority of sequences that match TFBS are non-functional. My lab is studying the sequence determinants that specify active from inactive instances of TFBS. Some of these features include general sequence properties such as the local GC-content or intrinsic shape of DNA, while other features specify the cooperative interactions between TFs. We aim to develop quantitative models that discriminate active from inactive occurrences of TFBS and to use these models to understand the impact of genetic variation in non-coding DNA.