

# Center for Science of Information

NSF Science and Technology Center

INVITED SPEAKER



**Idoia Ochoa**

Assistant Professor

Electrical and Computer Engineering  
University of Illinois at  
Urbana-Champaign

## Genomic data compression, processing, and analysis

### Abstract:

Recent technological advances have led to a drastic reduction in the cost of genome sequencing, resulting in the accumulation of an unprecedented amount of highly distributed and heterogenous genomic data. My research is dedicated to identifying and addressing the challenges arising in the context of such data. This undertaking - which combines tools from information and coding theory, statistics, and machine learning - includes the design and deployment of new algorithms for coping with the distribution and storage of the data, for facilitating its access and queries, and for improving the analysis and inference performed on it.

The talk will focus on some of my work in this area, geared toward alleviating the requirements for storage of the data, and improving its analysis. I will present some of our results on lossless compression of genomic data, as well as lossy compressors and denoisers of the quality scores that come with the sequencing data. In some cases lossy compression results not only in substantial storage savings, but also boosted performance when inference is done on the reconstructed data. Conversely, denoising the quality scores can result in substantial compression gains.

Tuesday, April 18th, 2017  
2:00pm EDT

LWSN 3102 (A&B)  
Purdue University



<http://www.soihub.org/>

Supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

