

# Idoia OCHOA

308 CSL, 1308 W Main St, Urbana, IL, 61801

☎ +1 (650) 739 5404

✉ idoia@illinois.edu

🌐 <http://idoia.ece.illinois.edu>

## POSITION

---

### University of Illinois at Urbana-Champaign

Assistant Professor

IL, USA

01/2017–Present

- Department of Electrical and Computer Engineering

## EDUCATION

---

### Stanford University

PhD candidate in Electrical Engineering

CA, USA

09/2010–09/2016

- Thesis: Genomic data compression and processing: theory, models, algorithms and experiments.
  - Advisor: Prof. Tsachy Weissman (tsachy@stanford.edu)

### Stanford University

Master in Electrical Engineering, *GPA: 4.031*

CA, USA

09/2010–12/2012

### TECNUN, University of Navarra

Telecommunications Engineering Degree

Spain

09/2003–02/2009

- Ranking position: 2 out of 51.
- Master thesis: Iterative Decoding Techniques for the Relay Channel (*GPA: 10/10*).
  - Advisor: Prof. Pedro M. Crespo (pcrespo@ceit.es)

### Lulea Tekniska Universitet

Erasmus Program

Sweden

08/2007–01/2008

## INTERESTS

---

Data Compression, Bioinformatics, Machine learning, Information Theory and Coding, Signal Processing.

## EXPERIENCE

---

### Stanford University

Group of Prof. Tsachy Weissman in the department of Electrical Engineering

CA, USA

2010–2016

I worked on the design and development of new algorithms to improve the distribution and storage of genomic data, to facilitate its access, and to boost the inferential power of analysis performed on it. My approach combines tools from information theory, statistics, and machine learning.

- Contributions:
  - Designed lossless and lossy compressors for genomic data
  - Developed a denoiser to reduce noise present in genomic data
  - Proposed a new method for cancer module discovery using gene expression data
  - Designed compression schemes for databases that enable efficient similarity queries in the compressed domain

**Brown Hill Productions**

Technical consultant

CA, USA  
2014–2015

- Supervisor: Producer Jonathan Dotan
- Technical consultant for the second season of HBO's TV show "Silicon Valley".

**Genapsys**

Internship

CA, USA  
Summer 2014

- Supervisor: Dariush Dabiri (dariush@genapsys.com)
- Design and implementation (in Matlab) of the base-calling algorithm for Genapsys' new sequencing technology.

**Google**

Internship

CA, USA  
Summer 2013

- Supervisor: Lynn Bos (lynnb@google.com)
- Design, implementation and verification of a flexible decimation filter for a digital microphone.
  - Architecture design in Matlab
  - Implementation of a C-model including fixed point conversion
  - C-to-RTL conversion using the tool Catapult
  - FPGA implementation

**CEIT, Centre of Studies and Technical Research of Gipuzkoa**

Research Assistant in the Electrical Engineering Department

Spain  
2009–2010

- Supervisor: Prof. Pedro M. Crespo (pcrespo@ceit.es)
- Conducted research in the relay channel, source-channel coding for non-uniform and cyclostationary sources, LDPC and Turbo codes.

Research Assistant in the Electrical Engineering Department

2008

- Supervisor: Prof. Andoni Irizar
- Design of a PLCP (Physical Layer Convergence Procedure) with VHDL, for a UWB (Ultra Wide Band) system.

**TECNUN, University of Navarra**

Collaborator Student in the Electrical Engineering Department

Spain  
2005

- Worked on Cadence circuit design and lay-out

**TEACHING EXPERIENCE**

---

**University of Illinois at Urbana-Champaign**

Assistant Professor

CA, USA  
2017–Present

- Probability with Engineering Applications (*ECE313 – MATH362, Spring 2017*)

**Stanford University**

Teaching Assistant

IL, USA  
2013–2015

- Information Theory (*EE376A – STATS376A, Winter 2015, Winter 2016*)
- Universal Schemes in Information Theory (*EE376C, Spring 2014*)
- Introduction to Statistical Signal Processing (*EE278B, Winter 2014*)

- Inference, Estimation and Information Processing (*EE378A, Spring 2013*)

**TECNUN, University of Navarra**

Teaching Assistant

Spain  
2003–2004

- Laboratory of Circuits and Physics I

## JOURNAL PAPERS

---

- K. Tatwawadi, M. Hernaez, **I. Ochoa**, and T. Weissman, *GTRAC: fast retrieval from compressed collections of genomic variants*, **Bioinformatics**, btw437, 2016.
- **I. Ochoa**, M. Hernaez, R. Goldfeder, T. Weissman and E. Ashley, *Effect of lossy compression of quality scores on variant calling*, **Briefings in Bioinformatics**, bbw011, 2016.
- S. Deorowicz, S. Grabowski, **I. Ochoa**, M. Hernaez and T. Weissman, *Comment on: "ERGC: An efficient referential genome compression algorithm"*, **Bioinformatics**, btw704, 2015.
- G. Malysa, M. Hernaez, **I. Ochoa**, M. Rao, K. Ganesan and T. Weissman, *QVZ: lossy compression of quality values*, **Bioinformatics**, btw330, 2015.
- **I. Ochoa**, M. Hernaez and T. Weissman, *Aligned genomic data compression via improved modeling*, **Journal of bioinformatics and computational biology**, Vol. 12, No. 6, 2014.
- A. Manolagos, **I. Ochoa**, K. Venkat, A. Goldsmith and O. Gevaert, *CaMoDi: a new method for cancer module discovery*, **BMC genomics**, Vol. 15, No. Sup. 10, 2014 (**Highly Accessed**).
- **I. Ochoa**, M. Hernaez and T. Weissman, *iDoComp: a compression scheme for assembled genomes*, **Bioinformatics**, btu698, 2014.
- **I. Ochoa**, H. Asnani, D. Bharadia, M. Chowdhury, T. Weissman and G. Yona, *QualComp: a new Lossy Compression Algorithm of Quality Scores based on Rate Distortion Theory*, **BMC Bioinformatics**, Vol. 14, No. 1, 2013 (**Highly Accessed**).
- **I. Ochoa**, P. Crespo and M. Hernaez, *LDPC Codes for Non-Uniform Memoryless Sources and Unequal Energy Allocation*, **IEEE Communications Letters**, Vol. 14, No. 9, 2010.
- **I. Ochoa**, P. Crespo, J. Del Ser and M. Hernaez, *Turbo Joint Source-Channel Coding of Non-Uniform Memoryless Sources in the Bandwidth-Limited Regime*, **IEEE Communications Letters**, Vol. 14, No. 4, 2010.

## CONFERENCE PAPERS

---

- R. Long, M. Hernaez, **I. Ochoa** and T. Weissman, *GeneComp, a new reference-based compressor for SAM files*, **Data Compression Conference (DCC)**, USA, 2017.
- **I. Ochoa**, A. No, M. Hernaez, and T. Weissman, *CROMqs: an infinitesimal successive refinement lossy compressor for the quality scores*, **IEEE Information Theory Workshop (ITW)**, UK, 2016.

- K. Tatwawadi, M. Hernaez, **I. Ochoa**, and T. Weissman, *GTRAC: fast retrieval from compressed collections of genomic variants*, **European Conference on Computational Biology (ECCB)**, Netherlands, 2016.
- **I. Ochoa**, M. Hernaez, R. Goldfeder, T. Weissman and E. Ashley, *Denoising of Quality Scores for Boosted Inference and Reduced Storage*, **Data Compression Conference (DCC)**, USA, 2016.
- M. Hernaez, **I. Ochoa** and T. Weissman, *A cluster-based approach to compression of Quality Scores*, **Data Compression Conference (DCC)**, USA, 2016.
- **I. Ochoa**, M. Hernaez and T. Weissman, *Aligned genomic data compression via improved modeling*, **GIW ISCB-Asia**, Japan, 2014.
- A. Manolakos, **I. Ochoa**, K. Venkat, A. Goldsmith and O. Gevaert, *CaMoDi: a new method for cancer module discovery*, **GIW ISCB-Asia**, Japan, 2014.
- **I. Ochoa**, A. Ingber and T. Weissman, *Compression Schemes for Similarity Queries*, **Data Compression Conference (DCC)**, USA, 2014.
- **I. Ochoa**, A. Ingber and T. Weissman, *Efficient Similarity Queries via Lossy Compression*, **Allerton Conference on Communication, Control, and Computing**, USA, 2013.
- B. G. Chern, **I. Ochoa**, A. Manolakos, A. No, K. Venkat and T. Weissman, *Reference Based Genome Compression*, **IEEE Information Theory Workshop (ITW)**, Switzerland, 2012.
- **I. Ochoa**, P. Crespo, J. Del Ser and M. Hernaez, *Turbo Joint Source-Channel Coding of Cycle-Stationary Sources in the Bandwidth-Limited Regime*, The 2nd International Conference on Mobile Lightweight Wireless Systems (MOBILIGHT), Spain, 2010.
- M. Hernaez, P. Crespo, J. Del Ser and **I. Ochoa**, *Codigos LDGM concatenados para la transmision de fuentes correlacionadas para canales de difusion*, XXIV National Assembly of the International Union of Radio Science (URSI), Spain, 2009.

## CONFERENCE POSTERS

---

- P. Li, M. Hernaez, **I. Ochoa**, and O. Milenkovic, *Micro: Microbial Database Compression Via Community Detection and Reference Discovery*, **Conference on Research in Computational Molecular Biology (RECOMB)**, San Diego, 2016.
- **I. Ochoa**, M. Hernaez, R. Goldfeder, T. Weissman and E. Ashley, *Denoising of Quality Scores for Boosted Inference and Reduced Storage*, **Conference on Research in Computational Molecular Biology (RECOMB)**, San Diego, 2016.
- K. Tatwawadi, M. Hernaez, **I. Ochoa**, and T. Weissman, *Fast retrieval from compressed collections of genomic variants*, **Conference on Research in Computational Molecular Biology (RECOMB)**, San Diego, 2016.

## INVITED TALKS

---

- **Current challenges related to Genomic data processing and analysis**, Center for Science of Information (CSol), Purdue University, Indiana, USA, April 2017.
- **Trends and Methods in Genomic Data Compression**, International Society for Computational Biology (ISMB), Orlando, USA, July 2016.
- **Genomic data compression, processing, and analysis**, University of Illinois at Urbana-Champaign, USA, April 2016.
- **Genomic data compression, processing, and analysis**, Princeton University, USA, March 2016.
- **Genomic data compression, processing, and analysis**, Washington University in St. Louis, USA, March 2016.
- **Genomic data compression, processing, and analysis**, University of California, Berkeley, USA, February 2016.
- **Genomic data compression, processing, and analysis**, Stanford Compression Forum, USA, February 2016.
- **Genomic data compression**, Information Theory and Applications (ITA) workshop, San Diego, USA, February 2016.
- **Current trends on genomic data compression**, VA research seminar, Palo Alto, USA, November 2015.
- **Genomic Data Compression and Processing**, EECS Rising Stars, Massachusetts Institute of Technology (MIT), USA, November 2015.
- **An Overview of Genomic Data Compression**, Center for Science of Information (CSol), Virtual Brown Bag Research Discussion Series, September 2015.
- **Compression schemes for similarity queries**, Stanford Compression Forum, Technical Talk, Stanford, USA, January 2015.
- **Efficient similarity queries via lossy compression**, New Directions in the Science of Information, Poster Presentation, UC Berkeley, USA, November 2014 (**second place winner**).

## SCHOLARSHIPS AND AWARDS

---

Awarded an <b>NIH grant</b> , under the BD2K initiative, in collaboration with the University of Illinois at Urbana-Champaign (UIUC)	2015
Selected to give a talk as a participant of the <b>MIT EECS Rising Starts Workshop</b>	2015
Richard and Naomi Horowitz Fellow, <b>Stanford Graduate Fellowship</b>	2013–2015
<b>Basque Government Fellowship</b> for graduate studies	2013–2016
Ranked 23 <sup>rd</sup> out of 147 candidates in the <b>EE Qualifying Examination</b> at Stanford	2010
<b>La Caixa Fellowship</b> Program to extend studies in the USA	2010–2012
Master Thesis funded by <b>Telefonica Fellowship</b>	2008

## US PATENTS

---

- A. Manolakos, **I. Ochoa**, K. Venkat, A. Goldsmith and O. Gevaert, *Cancer module discovery using gene expression data*, Submitted to Stanford's OTL - The Office of Technology Licensing, September 2014.
- **I. Ochoa** and M. Hernaez, *A Universal Compressor for Genomic Re-Sequencing Data*, Provisional US patent filled by Stanford's OTL - The Office of Technology Licensing, June 2014.

## ADDITIONAL INFORMATION

---

### **Professional Organizations:**

- Collins Scholar, University of Illinois at Urbana-Champaign (2017 - Present)
- Stanford Compression Forum: organizer of the second edition (February 2016)
- International Organization for Standardization (ISO): Active participant in the initiative to define and establish a compression standard for genomic data (under the MPEG working group).
- Stanford Data Science Initiative (SDSI) (2014 - 2016)
- Center for Science of Information (CSol), NSF Science and Technology Center (2010 - Present)

**Reviewer:** Bioinformatics, Nature Technical Reports, Nature Biotechnology, BMC Bioinformatics, IEEE Communications Letters, several conference proceedings.

**Languages:** Native: Spanish, Proficiency: English, Low-Intermediate: German, French and Basque.

**Computer skills:** Programming Languages: C/C++, Python, Applications: R, MatLab, LATEX, MS Office, CVX, Java Operating Systems: Linux, UNIX, Windows.

**Associations:** President, Vicepresident and Financial Officer of Iberia - Spanish Association at Stanford (2010 - 2016).

## REFERENCES

---

Available upon request.