

Mikel  
HER-  
NAEZ

CIMA University of Navarra, Avda. Pio XII, 55, 31008, Pamplona, Spain  
☎ +34 (686) 872 682  
✉ [mhernaez@unav.es](mailto:mhernaez@unav.es)  
🌐 [mikelhernaez](https://mikelhernaez.com)

## POSITIONS

---

**Center for Applied Medical Research (CIMA) University of Navarra** SPAIN

Director, Computational Biology and Translational Genomics Program

Principal Investigator, Machine Learning for Biomedicine Lab

**Carl R. Woese Institute for Genomic Biology**

*University of Illinois, USA*

Affiliate Faculty

## CURRENT PROJECTS AND RESEARCH

---

### **Elucidating transcriptional rewiring on hematological malignancies via computational methods**

Funded by the Marie Curie Fellowship from the European Commission

- Development of computational methods to uncover dynamically-regulated GRNs.
- Elucidate transcriptional rewiring on hematological malignancies.

### **Reverse Engineering the Human Cancer Transcriptome with Applications to Prostate Cancer**

Funded by the American Department on Defense, in collaboration with Mayo Clinic (Dr. Wang, Dr. Weinshilboum)

- A method to elucidate mechanistic changes in gene regulatory networks based on drug response.
- Applications to patients with metastatic castrate-resistant prostate cancer and their response to Abiraterone Acetate/Prednisone.
- Funded by the US Department of Defense

### **Reducing the footprint of Genomic Information**

Funded by the Spanish National Research Plan

- Develop novel technologies to reduce the storage and transmission cost of next-generation sequencing data.
- Previously funded by the Mayo Clinic Grand Challenge project.

### **MPEG-G, a new international standard for genomic information representation**

under the Mitogen initiative ([mitogen.github.io](https://mitogen.github.io))

- Work under ISO (International Organization for Standardization) to develop a new compressed standard for genomic information representation.
- Funded by NIH and the Chan-Zuckerberg Initiative under the Human Cell Atlas.

### **Module-based analysis of the cancer transcriptome**

in collaboration with Stanford University and The Broad Institute (Prof. Gevaert and Prof. Pochet)

- A computational method for multi-omics data fusion to identify cancer driver genes.

## EDUCATION

---

### **Stanford University**

CA, USA

Postdoctoral researcher in Electrical Engineering

09/2013–12/2016

- o Developed compression tools for genomic data
- o Explored new approaches for performing alignment and variant calling on genomic data
- o Characterized long non-coding RNAs across human cancer using unsupervised learning
- o P.I.: Prof. Tsachy Weissman (tsachy@stanford.edu)
- o Collaborated with:
  - Prof. Euan Ashley, M.D. (euan@stanford.edu), Stanford Medical School
  - Prof. Olgica Milencovic (milencov@illinois.edu), Electrical and Computer Eng. Dept.
  - Prof. Olivier Gevaert (ogevaert@stanford.edu), Stanford Data Science in Biomedicine Dept.

### **TECNUN, University of Navarra**

Spain

PhD in Electrical Engineering, GPA: *Summa Cum Laude*

09/2010–12/2012

- o Topics: Conducted research in information theory and coding, communication systems and signal processing, with special focus on iterative channel codes (LDPC, turbo codes...)
- o Dissertation: Joint Network-Channel Coding Schemes for Relay Networks
  - Advisors: Prof. Pedro Crespo (pcrespo@tecnun.es) and Javier Del Ser (jdelser@tecnalia.es)

### **TECNUN, University of Navarra**

Spain

Telecommunications Engineering Master Degree

09/2003–02/2009

- o Ranking position: Top 10
- o Master thesis: Concatenated LDGM Codes for the Transmission of Correlated Sources over Gaussian Broadcast Channels (GPA: 10/10)
  - Advisor: Prof. Pedro M. Crespo (pcrespo@ceit.es)

### **Lulea Tekniska Universitet**

Sweden

Erasmus Program

08/2007–01/2008

## EXPERIENCE

---

### **Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign** IL, USA

Director of Computational Genomics

2017–2019

- o Advice leadership on the computational biology matters.
- o Co-PI of the Mayo Grand Challenge project: AI collaboration between Mayo Clinic and the University of Illinois. Lead PI of the genomic data compression front.
- o Develop statistical methods for RNA-Seq data in collaboration with the Carle College of Medicine and Stanford University.

### **Stanford University**

CA, USA

Group of Prof. Tsachy Weissman

2013–2016

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.

- o Contributions:
  - Proposed, in collaboration with MIT and EPFL, a methodology for analysis of genomic data

compression on Variant Calling that set the bases for the Standardization process of genomic information

- Designed lossless and lossy compressors for genomic data
- Developed a denoiser to reduce noise present in genomic data

## **ENIGMEDIA**

Director of Research

*Spain-USA*

**03/2013 - 09/2013**

- o Supervisor: CEO & Founder Gerard Vidal (gerard@enigmedia.com)
- o Worked on encrypted real-time communications based on chaos-based stream-ciphers

## **NAUI SYSTEMS**

Consultant

*CA, USA*

**04/2013 - 06/2013**

- o Helped with the possibility of implementing a new coding solution for RAM memories

## **STANFORD UNIVERSITY-TECNUN**

Visiting Researcher

*USA-Spain*

**Summer 2012**

- o Supervisor: Golan Yona (golan.yona@stanford.edu)
- o Conducted research in biological relational databases under the BIOZON project (<http://biozon.org>)

## **CEIT**

Research Assistant in the Electrical Engineering Department

*Spain*

**2009-2012**

- o Supervisor: Prof. Pedro M. Crespo (pcrespo@ceit.es)
- o Proposed several practical coding schemes for relay channels using LDPC and Turbo codes
- o Set up of a point-to-point wireless communication system for pedagogic purposes

## **TECNUN, University of Navarra**

Collaborator Student in the Electrical Engineering Department

*Spain*

**2005**

- o Worked on Cadence circuit design and lay-out

## **TEACHING EXPERIENCE**

---

### **University of Illinois at Urbana-Champaign**

Lecturer

*CA, USA*

**01/2018 -**

- o BIOE 383: High-Throughput Genomic Data Analysis (Dept. Bioengineering)
- o ECE 365: Data Science and Engineering (Dept. Electrical and Computer Engineering)

### **TECNUN, University of Navarra**

Lecturer

*CA, USA*

**09/2012 - 01/2013**

- o Information Theory and Coding
- o Communication Systems
- o Fundamentals of Computers course

### **TECNUN, University of Navarra**

Teaching Assistant

*Spain*

**2011-2012**

- o Information Theory and Coding
- o Communication Systems

- Set up of a point-to-point wireless communication system for pedagogic purposes

## TECNUN, University of Navarra

Advisor of a Master Thesis

Spain  
2010–2011

- o Topic: Implementation of a Software Development Kit for Communications System.
- o Resulted in a IEEE publication.

## RESEARCH GRANTS

---

- Awarded a **Marie S. Curie Grant (PI)** from the European Union 2020
- Awarded an **American Department of Defense (DoD) grant (PI)** 2020
- Awarded an **Strategic Research Initiative (SRI) grant (PI)** from the University of Illinois at Urbana-Champaign (UIUC) 2018
- Awarded a **Chan Zuckerberg Initiative (CZI) grant (PI)**, under the Human Cell Atlas 2018
- Awarded a **Mayo Grand Challenge grant (co-PI)**, from Mayo Clinic 2017
- Awarded an **NIH grant**, under the BD2K initiative, in collaboration with the University of Illinois at Urbana-Champaign (UIUC) 2015

## PREPRINTS AND UNDER REVIEW

---

\*: Contacting author and/or (joint) first author

- o M. Goyal, G. Serrano, Ilan Shomorony, I. Ochoa, and **M. Hernaez**, *JIND: Joint Integration and Discrimination for Automated Single-Cell Annotation*, Bioarxiv, 2020.
- o J. Peng, U. V. Chembazhi, S. Bangru, I. M. Traniello, A. Kalsotra, I. Ochoa, and **M. Hernaez**, *SimiC: A Single Cell Gene Regulatory Network Inference method with Similarity Constraints*, Bioarxiv, 2020.
- o U. Chembazhi, S. Bangru, **M. Hernaez**, A. Kalsotra, *Cellular plasticity balances the metabolic and proliferation dynamics of a regenerating liver*, Bioarxiv, 2020.
- o C. Alberti, T. Paridaens, J. Voges, D. Naro, J. J. Ahmad, M. Ravasi, D. Renzi, P. Ribeca, G. Zoia, I. Ochoa, M. Mattavelli, J. Delgado, **M. Hernaez\***, *An introduction to MPEG-G, the new ISO standard for genomic information representation*, Biorxiv, 2018.

## JOURNAL PAPERS AND CONFERENCE PROCEEDINGS

---

\*: Contacting author and/or (joint) first author

- o A. No, **M. Hernaez\***, I. Ochoa, CROMqs: An Infinitesimal SuccessiveRefinement Lossy Compressor for the Quality Scores, Journal of Bioinformatics and Computational Biology, 2020
- o Q. Meng, I. Ochoa, **M. Hernaez\***, *GPress: a framework for querying General Feature Format (GFF) files and expression files in a compressed form*, Bioinformatics, in press 2020.

- J. Voges, T. Paridaens, F. Muntefering, L. S.Mainzer, B. Bliss, M. Yang, I. Ochoa, J. Fostier, J. Ostermann, and **M. Hernaez\***. GABAC: an arithmetic coding solution for genomic data, *Bioinformatics*, 2019.
- O. Gevaert,..., **M. Hernaez**,..., N. Pochet. *Imaging-AMARETTO: a radiogenomics software tool for interpreting multi-omics networks for relevance to radiographic and histopathology imaging-derived biomarkers of clinical outcomes with application to studies of brain tumors*, <https://doi.org/10.1101/2020.10.06.327601>JCO Clinical Cancer Informatics, 2020.
- I. Fisher-Hwang, I. Ochoa, T. Weissman, **M. Hernaez\***, *Denoising of Aligned Genomic Data*, **Nature Scientific Reports**, 2019.
- **M. Hernaez\***, C. Blatti, O. Gevaert, *Comparison of single gene and module-based methods for modeling gene regulatory networks*, **Bioinformatics**, 2019.
- H. Zheng, K. Brennan, **M. Hernaez**, O. Gevaert, *Benchmark of lncRNA Quantification for RNA-Seq of Cancer Samples*, **GigaScience**, 2019.
- **M. Hernaez\***, D. Pavlichin, and T. Weissman, I. Ochoa, *Genomic Data Compression*, **Annual Review of Biomedical Data Science**, 2019.
- I. Ochoa, H. Li, F. Baumgarte, C. Hergenrother, J. Voges, **M. Hernaez\*** *AliCo: A New Efficient Representation for SAM Files*, **Data Compression Conference (DCC)**, 2019.
- J. Seimetz, W. Arif, S. Bangru, **M. Hernaez**, A. Kalsotra, *Cell-type specific polysome profiling from mammalian tissues*, **Methods**, 155, 2019.
- K. Kendig, S. Baheti, M. Bockol , T. Drucker, S. Hart, J. Heldenbrand , **M. Hernaez**, M. Hudson, M. Kalmbach, E. Klee , N. Mattson, C. Ross, M. Taschuk, E. Wieben, M. Wiefert, D. Wildman and L. Mainzer, *Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy*, **Frontiers in Genetics**, 2019.
- S. Chandak, I. Ochoa, T. Weissman, **M. Hernaez\***, *SPRING: a next-generation compressor for FASTQ data*, **Bioinformatics**, 2018.
- L. Rogusky, I. Ochoa, **M. Hernaez**, S. Deorowicz, *FaStore - a space-saving solution for raw sequencing data*, **Bioinformatics**, 2018.
- J. Voges, J. Oesterman, **M. Hernaez\***, *CALQ: compression of quality values of aligned sequencing data*, **Bioinformatics**, 2018.
- R Long, **M Hernaez\***, I Ochoa, T Weissman, *GeneComp, a new reference-based compressor for SAM files*, *Data Compression Conference (DCC)*, 2017
- K. Tatwawadi, **M. Hernaez\***, I. Ochoa, and T. Weissman, *GTRAC: fast retrieval from compressed*

*collections of genomic variants*, **Bioinformatics**, btw437, 2016.

- C. Alberti, N. Daniels, **M. Hernaez\***, J. Voges, R. L. Goldfeder, A. A. Hernandez-Lopez, M. Mattavelli, B. Berger, *An Evaluation Framework for Lossy Compression of Genome Sequencing Quality Values*, **Data Compression Conference (DCC)**, 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)**, 2016.
- **M. Hernaez\***, I. Ochoa and T. Weissman, *A cluster-based approach to compression of Quality Scores*, **Data Compression Conference (DCC)**, 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**, 2016.
- S. Deorowicz, S. Grabowski, I. Ochoa, **M. Hernaez** and T. Weissman, *Comment on: "ERGC: An efficient referential genome compression algorithm"*, **Bioinformatics**, btv704, 2015.
- G. Malysa, **M. Hernaez\***, I. Ochoa, M. Rao, K. Ganesan and T. Weissman, *QVZ: lossy compression of quality values*, **Bioinformatics**, btv330, 2015.
- I. Alustiza, **M. Hernaez**, P. Crespo, *Design of a new scheme for multi-hop wireless networks using decode-and-forward strategy*, **EURASIP Journal on Wireless Communications and Networking**, (1), 1-8, 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.
- I. Ochoa, **M. Hernaez\*** and T. Weissman, *iDoComp: a compression scheme for assembled genomes*, **Bioinformatics**, btu698, 2014.
- **M. Hernaez\***, P.M. Crespo, J. Del Ser, *On the Design of a Novel Joint Network-Channel Coding Scheme for the Multiple Access Relay Channel*, **IEEE Journal on Selected Areas in Communications**, Vol. 31, No. 8, 1157-1167, August 2013.
- **M. Hernaez\***, P.M. Crespo, J. Del Ser, *A Decode-and-Forward Scheme for Multihop Wireless Networks*, **IEEE Vehicular Technology Conference (VTC2013-Fall)**, Las Vegas, Sept. 2013
- I. Alustiza, **M. Hernaez**, X. Insasusti and P.M. Crespo, *Teaching Information Theory via a Simulation Tool for Communications Systems*, **IEEE Collaborative Learning & New Pedagogic Approaches in Engineering Education (IEEE EDUCON)**, Berlin (Germany), March 2013.
- **M. Hernaez\***, P.M. Crespo, J. Del Ser *A Flexible Channel Coding Approach for Short-Length Codewords*, **IEEE Communications Letters**, Vol. 16, No. 9, 1508-1511, September 2012.
- **M. Hernaez\***, P.M. Crespo, *A novel Scheme for Message-Forwarding in Ad-Hoc Wireless Net-*

works, **IEEE Vehicular Technology Conference (VTC2011-Spring)**, Budapest (Hungary), May 2011

- **M. Hernaez\***, P.M. Crespo, J. del Ser, *Joint Non-Binary LDPC-BICM and Network Coding with Iterative Decoding for the Multiple Access Relay Channel*, **IEEE Vehicular Technology Conference (VTC2011-Spring)**, Budapest (Hungary), May 2011
- 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.
- **M. Hernaez\***, P. M. Crespo, J. Del Ser, J. Garcia-Frias, *Serially-Concatenated LDGM Codes for Correlated Sources over Gaussian Broadcast Channels*, **IEEE Communications Letters**, Vol 13, No. 10, 788-790, October 2009.
- 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.

## SCHOLARSHIPS AND AWARDS

---

Awarded an **NIH grant**, under the BD2K initiative, in collaboration with the University of Illinois at Urbana-Champaign (UIUC) 2015

Postdoctoral research funded by the **Stanford Data Science Initiative** 2015-2016

Enigmmedia named the best new company of the Basque Country (Spain) 2013

**University of Navarra Fellowship** for graduate studies 2009–2011

Master Thesis funded by **Telefonica Fellowship**. 2008

## US PATENTS

---

◦ I. Ochoa and **M. Hernaez**, *A Universal Compressor for Genomic Re-Sequencing Data*, Provisional US patent filed by Stanford's OTL - The Office of Technology Licensing, June 2014.

◦ S.Chandak, K. Tatwawadi, T. Weissman, I. Ochoa, and **M. Hernaez**, *Systems and Methods for Compressing Genetic Sequencing Data*, Submitted to Stanford's OTL - The Office of Technology Licensing - and UIUC's OTM - Office of Technology Management, 2018.

## SERVICE ACTIVITIES

---

### Workshop Organization Committees:

- Chair of special session on "Omics Data Compression and Storage: Present and Future" at ISMB/ECCB, Basel, 2019 (acceptance rate 15%).
- Chair of special session on "Omics Data Compression and Storage: Present and Future" at ISMB (International Society for Computational Biology), Chicago, 2018 (acceptance rate 20%).
- Chair of special session on "Bioinformatics" at the 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton), October, 2018.
- Chair of special session on "Bioinformatics" at the 55th Annual Allerton Conference on Commu-

nication, Control, and Computing (Allerton), October, 2017.

**Professional Organizations:**

- International Society of Computational Biology (ISCB): Member
- Stanford Compression Forum: Organizer of the first and second edition (2015 - 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): Active member and grantee (2014 - 2016)
- Center for Science of Information (CSol), NSF Science and Technology Center: Active member and grantee (2013-2015)

ADDITIONAL INFORMATION\_\_\_\_\_

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