

# Santiago Manuel Castro Dau

---

## General Information

Date of Birth 24th of May 1995  
Nationalities Mexico and United States of America  
Residence Zürich, Switzerland (student-B permit valid until 09/2024)  
Languages Spanish (native), English (fluent), German (B1)

## Bio

- Computational biologist with a chemical engineering background.
- Extensive interdisciplinary research experience.
- Technical expertise in mathematical modeling, machine learning and software development.
- Paassionate about science and a champion for curiosity-driven research.

## Education

**2020–2023 Master of Science** in Computational Biology and Bioinformatics, ETH Zürich  
Final average: 5.6. Master's thesis grade: 6.

**2014–2019 Bachelor of Engineering** in Chemical Engineering, Universidad Nacional Autónoma de México  
Admission rate 2014: 10%. Graduated with honors, top 7% of the cohort. Final average: 9.6/10.

**2018 Visiting Student**, University of Tokyo

## Experience

**2023–Now Software Engineer at the Food Systems Biotechnology group at ETH Zürich.**  
Duration 3 months at the moment of applying  
Description Developing microbiome analysis software tools for Qiime2, a package collection with reproducibility at its core.  
Supervisor Michal Ziemski

**2022–2023 Master Thesis at the IBM Research in the Artificial Intelligence for Single Cell Research group.**  
Duration 6 months  
Thesis Title *Modeling the Tumor Microenvironment with Graph Concept Learning*  
Description Designed and implemented an innovative machine learning framework, enabling interpretable predictions of tumor tissue images through concept and geometric-deep learning.  
Supervisors Pushpak Pati, Marianna Rapsomaniki and Mark Robinson

**2022 Research Internship at the Cortical Computation group, Institute of Neuroinformatics at the University of Zürich, and ETH Zürich.**  
Duration 3 months  
Description Modeled the learning process in biological neurons by representing the system with an artificial neural network and describing its dynamics through mixture model parameter inference.  
Supervisor Matthew Cook

**2021 Research Internship at the Computational Evolution group, D-BSSE, ETH Zürich.**  
Duration 3 months  
Description Conducted a statistical analysis of epidemiological data to investigate the potential applicability of a COVID-19 effective reproductive number estimator to data from countries outside the EU.  
Supervisor Jana Huisman

**2019–2020 Research Internship at the National Institute of Genomic Medicine in Mexico City.**  
 Duration 12 months  
 Description Conducted a statistical analysis of copy number mutations in breast cancer to explore their impact on the structure of the transcriptional network.  
 Supervisor Dr. Enrique Hernández Lemus

## Contributions to Repositories and Publications

- 2023 **Castro-Dau S.**, Martinelli A., Rapsomaniki M., Pait P. *Modeling the Tumor Microenvironment with Graph Concept Learning*. Paper in preparation.
- 2023 Contributed with pull requests to the packages q2-moshpit, q2-assembly, and ATHENA. Currently working on a contribution for PyTorch Geometric.
- 2020 Participated in the CTD<sup>2</sup> Pancancer Drug Activity DREAM challenge.

## Special Achievements

### 2023 Contributed Talk

Title *Modeling the Tumor Microenvironment with Graph Concept Learning*, BC2  
 Selected for an oral presentation at Basel Computational Biology Conference 2023, Switzerland's main event in the domain of Computational Biology and one of the major events of its kind in Europe.

### 2021 Winner of "ETH Week 2021: Health for Tomorrow"

An interdisciplinary, innovative, problem-solving challenge, where we completed with more than 50 ETH students to come up with an attractive solution for a real-world health-related problem.

### 2020 Recipient of "Jóvenes de Excelencia Citibanamex" scholarship

Scholarship for promising graduates looking to continue their studies at a recognized international institution.

### 2019 Contributed Talk

Title *Effect of Somatic Copy Number Alterations on Transcriptional Network Architectures*, INMEGEN  
 Selected for an oral presentation at Instituto Nacional de Medicina Genómica's 5th Student Encounter.

### 2018 Recipient of "Movilidad Internacional" scholarship

Scholarship to support my studies at the University of Tokyo, one of the world's top academic institutions.

## Technical Skills

- Programming languages: Python, R, git/GitHub, C++, SQL, Matlab, Bash,  $\text{\LaTeX}$ .
- Fluency in popular libraries and command line tools, e.g. PyTorch, scikit-learn, MLFlow, Snake-make, NetworkX, Qiime2, Tidyverse.
- Experience developing software in Python adhering to various standard technologies and best practices including CI/CD tools.
- First-hand experience with bash scripting, workflow management systems and high-performance computing environments.
- Experience with a variety of machine learning approaches, e.g. bayesian, concept, reinforcement and active learning, geometric deep-learning, and vision transformers.
- Familiarity with proteomic, genomic, transcriptomic, and spatial datasets.