MARCAL COMAJOAN CARA

Berkeley, CA | +1 650 ### #### | comajoan@berkeley.edu

linkedin.com/in/marcal-comajoan
github.com/salcc
marcal.cc

EDUCATION

University of California, Berkeley | PhD in Computer Science Berkeley, CA | Aug 2025 - May 2030 (expected)
Researcher in the Berkeley AI Research (BAIR) Lab. Supported by the "la Caixa" Foundation Fellowship.

- Focus: Deep learning and reinforcement learning methods for autonomous scientific discovery and to advance genetics.

Polytechnic University of Catalonia (UPC) | BSE in Data Science and Engineering Barcelona, Spain | Sep 2020 - Jul 2024
Graduated as Valedictorian. Exchange semesters at **ETH Zurich** (Fall 2023) and at **Stanford University** (Spring 2024).

RELEVANT EXPERIENCE

Stanford University | Research Assistant Stanford, CA | Jan 2023 - Present
Advisor: Prof. Alexander Ioannidis. Research on deep learning, population genetics, and precision health.

- Trained a multimodal foundation model on ∼500k samples from the UK Biobank, fusing genome variation, clinical traits, and images; raised AUCs for challenging conditions like glaucoma (+2.1%) and cataract (+2.2%); enabled retinal image synthesis with a conditioned diffusion model, with a ResNet trained only on 50k synthetic images reaching AUC 85% on real data.
- Achieved state-of-the-art results on a wide range of tabular data benchmarks by developing a large hypernetwork model meta-trained on over 1.8k tables, and enhanced with tree embeddings, latent retrieval, ensembling, and transfer learning.
- Reduced genealogical prediction error over competing methods by -38.6% (TMRCA RMSE) by replacing HMM/MCMC-based models with an end-to-end Transformer model for ancestral recombination graph (ARG) inference.
- Accelerated biobank-scale genomics input and output routines by implementing core features in the open-source snputils library, achieving a 99.91% speed-up in genotype data loading compared to existing solutions.
- Improved disease risk prediction in data-scarce populations, by +10.9% (Nigeria), +6.8% (Sri Lanka), and +4.1% (Hawaiian) in average AUC over the best supervised baselines by proposing a semi-supervised domain adaptation method for genetics.
- Mentored 4 undergraduates across 2 research projects about design of feature encodings and evaluation of biomedical AI models.

Machine Learning for Science (ML4Sci.org) | Research Assistant Remote | Apr 2023 - Nov 2023
Advisor: Prof. Sergei Gleyzer. Research on quantum deep learning, as part of Google Summer of Code 2023.

- Integrated variational quantum circuits into the Vision Transformer architecture and matched classical accuracy.
- Cut training cost from >100 h in prior work to 39 min, while scaling to >930k images with an optimized JAX/Flax pipeline.

Polytechnic University of Catalonia (UPC) | Research Assistant Barcelona, Spain | Jul 2022 - Dec 2022
Advisor: Prof. Jordi Cortadella. Research on automatic chip floor-planning.

- Implemented algorithms based on spectral graph theory, force-based algorithms, and nonlinear optimization.

SELECTED PUBLICATIONS

(*equal contribution) Google Scholar: scholar.google.com/citations?user=hnYw3tMAAAAJ

- **M Comajoan Cara**, ..., A Ioannidis. "SNP2ARG: Inferring Ancestral Recombination Graphs with Deep Learning". Meeting of the Society for Molecular Biology & Evolution (SMBE) 2025 (oral).
- **M Comajoan Cara***, M Geleta*, C Thomassin*, D Bonet*, ..., A Ioannidis. "PM1: A Foundation Model Fusing Genotype, Phenotype, and Image for Precision Medicine". International Conference on Machine Learning (ICML) 2025 Workshop on Multimodal Foundation Models and Large Language Models for Life Sciences (poster).
- **M Comajoan Cara***, D Bonet*, ..., A Ioannidis. "iLTM: Integrated Large Tabular Model". International Conference on Learning Representations (ICLR) 2025 Workshop on Weight Space Learning (spotlight, oral).
- **M Comajoan Cara***, D Bonet*, M Barrabés*, ..., A Ioannidis. "snputils: A Python Library for Processing Genetic Variation and Population Structure". Meeting of the American Society of Human Genetics (ASHG) 2024 (poster).
- **M Comajoan Cara**, ..., A Ioannidis. "PopGenAdapt: Semi-Supervised Domain Adaptation for Genotype-to-Phenotype Prediction in Underrepresented Populations". Pacific Symposium on Biocomputing (PSB) 2024 (oral).
- **M Comajoan Cara**, ..., S Gleyzer. "Quantum Vision Transformers for Quark-Gluon Classification". Axioms 2024.

LANGUAGES

- English (fluent, TOEFL iBT 115/120)
- Spanish (native)
- Catalan (native)
- French (basic, DELF A2)

SKILLS & TECHNOLOGIES

- Programming in Python, C++, R, MATLAB, JavaScript, and SQL.
- Machine learning with PyTorch, JAX/Flax, and scikit-learn/Pandas/NumPy.
- Web development with React, Bootstrap, Node, and Flask.
- Other: Git, GitHub, Wandb, LaTeX, Slurm, GNU/Linux, macOS, Windows.