Matthieu Chaldebas | mchaldebas.github.io | Matthieu.Chaldebas@gmail.com | New York, NY

Machine Learning Scientist with doctoral research conducted at **The Rockefeller University (St. Giles Lab)**. Developer of 5ULTRA, an ML framework for functional variant prioritization that outperforms industry-standard benchmarks. Expert in managing massive genomic data flows (**490,000+ genomes**).

EDUCATION

Université Paris Cité | PhD Candidate, Bioinformatics — 2022 – 2026 (Expected)

Sup' Biotech | MS in Biotechnology Engineering — 2016 – 2021

UC San Diego (UCSD) | Exchange Semester: Immunology & Human Disease — 2018

EXPERIENCE

The Rockefeller University, New York, USA — 2021 – 2026
St. Giles Laboratory of Human Genetics of Infectious Diseases
PhD Research & Bioinformatics Analyst

- **Developed 5ULTRA**, a **Random Forest ML framework** for functional variant prioritization; **outperformed the industry-standard CADD score** in AUROC/AUPRC on test datasets. (Manuscript in prep for ***AJHG***).
- Led the functional prioritization of **490,640 UK Biobank (UKB)** whole genomes using 5ULTRA; **identified novel rare-variant associations** overlooked by the original consortium's approach. (Manuscript in prep).
- Orchestrated high-throughput pipelines to process and annotate **26,000+ WES/WGS datasets** using **High-Performance Computing** and **SLURM**, managing **multi-terabyte genomic data flows**.
- Validated ML predictions using **UKB Olink proteomics data (n=50,000)**; demonstrated directional accuracy and effect-size correlation for translational variants.
- Conducted **bulk and single-cell RNA-seq** analyses, enhancing transcriptomic insights across various projects; collaborate with a multidisciplinary team to design and implement custom analytical pipelines.

BioMérieux, Lyon, France — 6 Months – 2021
Data Scientist Intern

- Optimized **supervised Lasso ML models** to increase predictive accuracy for antibiotic resistance.
- Enhanced **cross-validation** to account for phylogenetic factors, improving model generalization.
- Implemented algorithms for **novelty detection** in heterogeneous datasets.

CEA, Paris Saclay, France — 4 Months – 2019
Data Engineer Intern

- Managed biological datasets, performed phylogenetic analysis, and created **3D protein models.**

LEADERSHIP

Sup'Biotech | Student Board Representative — 2018 – 2022

Tech My Wine Student Club | Vice-President — 2020 – 2021

SKILLS

Computation: Python, Snakemake, R, Bash, Docker, Git, UKB RAP, DNAnexus, HPC (SLURM).

Machine-Learning: Scikit-learn, TensorFlow, Model selection, Feature Engineering, AlphaGenome Integration.

Statistical Genetics: Burden Testing, pQTL/Proteomics, WGS/WES Discovery, scRNA-seq, 5'UTR Biology.

Languages: French (Native), English (Fluent - TOEIC 975/990).

Visa Status: J-1 (No 212e rule); 3 years H-1B eligibility remaining.