Tiago Maié — Curriculum vitæ

Sigmundstr. 20 – 52070 Aachen – Germany
☎ +351 96 94 84 155 • ✉ tiagomaie@hotmail.com • tiagomaie.github.io
in tiagomaie • Ph0bi4 • tiagomaie • UfSvF34AAAAJ • Tiago-Maie-2

Ph.D. dissertation

Title: *Small epigenetic signatures for classification and deconvolution*

Supervisor: Ivan G. Costa

Dissertation defense date: 27th June 2024

Keywords: *Machine learning, DNA methylation, Epigenetic signatures, Cell type classification, Cellular deconvolution, Biomarker selection, Clinical application*

Description: My project focused on epigenomics - in particular DNA methylation in CG dinucleotides (CpGs) - and machine learning. Here, I developed a novel method focused on the idea of using the minimum amount of data possible to produce a working classifier and deconvolution tool. The premise being that, if we can produce a model that only uses the minimum amount of information possible, then, the markers to be tested with targeted approaches would be the minimum amount of markers necessary to produce results. This kind of approach is more translatable to a clinical setting. Furthermore, the biological system being used, the methylation (or lack thereof of) CpGs, is highly suitable to define small and targeted signatures that, as shown in the dissertation, are enough to classify cell-types and deconvolute mixed-samples with a performance comparable to state-of-the-art models using orders of magnitude more data.

Education

Ph.D. in Bioinformatics — **Aachen, Germany**
Institute for Computational Genomics, RWTH University Hospital — *January 2018–June 2024*

M.Sc., Bioinformatics and Computational Biology — **Lisbon, Portugal / Norwich, United Kingdom**
University of Lisbon / John Innes Centre — *October 2011–February 2015*
Specialization in Computational Biology

B.Sc., Biology — **Lisbon, Portugal**
University of Lisbon — *September 2008–July 2011*
Specialization in Molecular Biology and Genetics

Computer skills

Programming languages

Expert: R, Python

Proficient: Bash, Java

Competent: Javascript, SQL(ite), GDScript

Basic knowledge: C, C^{++}, C$^{\#}$, HTML, Matlab

Techstacks

R: tidymodels, glmnet, caret, limma, DESeq2, Seurat, minfi, sesame, nnls, Rfast, tidyverse, data.table, devtools, and more.

Python: scikit-learn, Scanpy, scipy, NumPy, pandas, TensorFlow, anndata, seaborn, matplotlib, plotly, tkinter, and more.

Bioinformatics misc.: Nextflow, SAMtools, BWA, GATK, IGV, Adegenet, Qualimap, ImageJ, and more.

Authored open-source software

CimpleG [link]: an R package to find (small) CpG signatures. CimpleG allows the user to find DNA methylation (DNAm) biomarkers using the least amount of information possible, a one-CpG signature per cell type. Furthermore, it allows the user to use these signatures for cell type deconvolution of DNAm mixes.

PlotPSMC [link]: a python program to easily plot PSMC curves.

SINS4fun [link]: a javascript program where you can visualize how a basic simulation in SINS would look like.

Co-authored open-source software

SINS (Simulating INdividuals in Space) [link]: A Java program to simulate individuals (as well as their associated genetic data) migrating through an environment, with possible habitat loss and fragmentation.

Open-source project contributions

scrna_seurat_pipeline [link]: Part of the core development team.

bulk_RNASeq [link]: Part of the core development team.

Adegenet [link]: Fixed an error when trying to acquire summary statistics from simulated data that are fully homozygous.

minfi [link]: Fix to allow for a better control of an input reading function.

tidymodels/parsnip [link]: Fix to the argument calls for the parsnip XGBoost interface.

TCGAbiolinks [link]: Fix to the behaviour of a pipeline moving files to incorrect locations.

Bioinformatics analysis contributions

I have contributed to many scientific publications, including analysis with different data modalities such as bulk-RNAseq, scRNA-seq, Visium, WGS, DNAm arrays, WGBS, and others.
For a complete list of my publications, please check my **Google scholar profile [link]**.

General

OS: Unix, MOSIX[1], Windows

Typography: LaTeX

Office Suites: Google Workspaces, Microsoft Office, Libre Office, Prezi

Miscellaneous: Git (github, bitbucket, gitlab), Make, Tmux, (Neo)Vim, Slurm, Docker, Vagrant

Experience

Vocational

Post-Doctoral Researcher **Aachen, Germany**
Institute for Computational Genomics, RWTH University Hospital *June 2024–Present*
Single cell RNA-sequencing, Spatial single cell transcriptomics, Cell-cell communication, Biomarker detection

Data Management Officer **Aachen, Germany**
Institute for Computational Genomics, RWTH University Hospital *May 2024–Present*
Responsible for the data management and organization of our High-Performance Computing (HPC) infrastructure

Research Fellow **Lisbon, Portugal**
Population and Conservation Genetics - Instituto Gulbenkian de Ciência (IGC) *January 2016–December 2017*
Comparative landscape evolutionary genomics analysis of West-African primates

NOS Alive 2015 Research Fellowship in Biodiversity **Toulouse, France / Lisbon, Portugal**
Population and Conservation Genetics - Instituto Gulbenkian de Ciência (IGC) *November 2014–November 2015*
Software development and improvement of the SINS population genetics software

Visiting Researcher **Norwich, United Kingdom**
Computational & Systems Biology - John Innes Centre (JIC) *November 2012–August 2014*
Systems Biology research focused on the development of the Master thesis

Research Trainee **Lisbon, Portugal**
Cell Biophysics and Development group - Instituto Gulbenkian de Ciencia (IGC) *October 2011–June 2012*
Research on the pollen tube development, focus on GLRs. Applied the necessary techniques for genotyping *A. thaliana* seeds

Research Trainee **Lisbon, Portugal**
Center for Biodiversity, Functional & Integrative Genomics (BioFIG) *February 2011–September 2011*
Research on human hearing impairment, focus on the genes TMC1 and TMPRSS3. Applied different molecular biology techniques, such as DNA extractions from human blood samples, PCR, electrophoresis, purification, and sequence analysis

Organizational experience

ByteMAL Conference 2023 **Aachen, Germany**
RWTH Aachen University Hospital *January 2023–May 2023*
Member of the organizing committee

Supervision experience

Master student supervision **Aachen, Germany**
Supervised the M.Sc. thesis projects of Fabian Schlebusch and Myriam Erz *July 2020–June 2023*

Teaching experience

Bioinformatics In R **Aachen, Germany**
Teaching Assistant (editions: '18, '19, '20, '21, '22, '23, and '24) *November 2018–November 2024*

---

[1] for High Performance Computing

INTRICARE - Intro to bioinformatics **Aachen, Germany**
Teaching Assistant *September 20th 2019*

Conferences/Symposiums

Conference and Poster presentation **Lyon, France**
Centre de Congrès de Lyon *July 23rd 2023–July 27th 2023*
ISMB/ECCB 2023 (Intelligent Systems for Molecular Biology and European Conference on Computational Biology)

Conference and Poster presentation **Sitges, Spain**
Meliá Sitges Hotel *September 28th 2022–September 21st 2022*
ECCB 2022 (European Conference on Computational Biology)

Conference and Poster presentation **Basel, Switzerland**
Congress Center Basel *July 21st 2019–July 25th 2019*
ISMB/ECCB 2019 (Intelligent Systems for Molecular Biology and European Conference on Computational Biology)

Conference and Oral presentation **Braga, Portugal**
University of Minho *February 22nd 2017–February 24th 2017*
Bioinformatics Open Days 2017

Conference and Poster presentation **Vairão, Portugal**
CIBIO-InBIO *June 1st 2015–June 2nd 2015*
Trends in Biodiversity and Evolution 2015 - GLOBAL BIODIVERSITY CHANGE: FROM GENES TO ECOSYSTEMS

Conference **London, United Kingdom**
The Royal Society *November 7th 2013–November 8th 2013*
The Genetics Society Autumn Meeting 2013 - From Genes to Shape

Training attended

Health Data Challenge: Matrix factorization and deconvolution (...) **Aussois, France**
Univ. Grenoble Alpes *November 25th 2019–November 29th 2019*

Advanced Statistics and Data Mining Summer School **Madrid, Spain**
Univ. Politécnica de Madrid *June 25th 2018–July 6th 2018*

- Feature Subset Selection
- Supervised Pattern Recognition
- Unsupervised Pattern Recognition
- Neural Networks and Deep Learning
- Support Vector Machines and Regularized Learning
- Bayesian Networks

ELIXIR-EXCELERATE - Train the Trainer Pilot Course **Lisbon, Portugal**
Instituto Gulbenkian de Ciência (IGC) *January 19th 2017–January 20th 2017*

Bioinformatics using Python for Biomedical Researchers **Lisbon, Portugal**
Instituto Gulbenkian de Ciência (IGC) *July 20th 2015–July 24th 2015*

Population Genetics and Demographic History: model-based approaches **Lisbon, Portugal**
Instituto Gulbenkian de Ciência (IGC) *July 6th 2015–July 9th 2015*

Automated and reproducible analysis of NGS data **Lisbon, Portugal**
Instituto Gulbenkian de Ciência (IGC) *May 11th 2015–May 15th 2015*

EMBO Practical Course on 'Multi-level Modelling of Morphogenesis' **Norwich, United Kingdom**
John Innes Centre (JIC) *July 14th 2013–July 26th 2013*

Workshop in Molecular Biology Laboratory Techniques **Lisbon, Portugal**
BioFIG *February 7th 2011–February 11th 2011*

Awards

Vocational

Best talk award - Bioinformatics Open Days 2017 **Braga, Portugal**
University of Minho *February 22nd 2017–February 24th 2017*
Title: SINS - Simulating INdividuals in Space

NOS Alive 2015 Research Fellowship in Biodiversity **Lisbon, Portugal**
Instituto Gulbenkian de Ciência (IGC) *November 2014–November 2015*

Languages

Portuguese: Native
English: Near-native
German: Elementary proficiency
Spanish: Limited working proficiency
French: Limited working proficiency