AHMED DEBIT, Ph.D.

✉ ahmedebit@gmail.com ● ☎ +33 (0)788196974 ● Google Scholar ● LinkedIn ● ahmedebit.com

SUMMARY

Skilled computational biologist with a strong scientific and technical background, including experience in genomics/transcriptomics, data analysis, statistics, and machine learning.

WORK EXPERIENCE

Institut de Biologie ENS (IBENS)

Postdoctoral Research Fellow *November 2020-Present*

My primary focus is on **identifying and characterizing non-coding transcripts**, specifically long non-coding RNAs (lncRNAs), in marine species such as microplanktonic organisms and macroalgae. I'm particularly interested in exploring the epigenetic regulation role of lincRNAs and lncNATs in these species, including their synteny and functional conservation. My research involves developing a **machine learning meta-model pipeline tool (votingLNC)** to predict the coding potential of RNA transcripts, as well as conducting bioinformatics analysis (including **assembly and annotation**), metagenomics, and meta-transcriptomics. Additionally, I've created a **database called lncplankton (www.lncplankton.bio.ens.psl.eu),** which integrates long non-coding RNAs from marine protists. I'm mainly working with the Marine Microbial Eukaryote Transcriptome Sequencing Project (**MMETSP**), **Tara Oceans**, and **Phaeoexplorer data** (in **collaboration** with Roscoff Marine Station). I contributed to several research projects, presented results at international conferences, manuscript submissions, and took part in grant writing. I mostly used R, bash, Python, and SQL.

University of Liege

Research Fellow - GIGA-Research, Unit of Human Genetics *February 2017-March 2020*

Co-supervised by the director of the department of genetics and the head of BIO3 Unit, leveraged the integration of sequencing and biological data coming from various sources (**RNA-seq**, and **miRNA** expression) to solve questions related to **human cancer**. I contributed to **several projects**. I used circulating miRNAs data to identify **biomarker for Breast Cancer (BC) screening** in Rwandan population. I designed an optimized machine learning pipeline aiming at discovering **short molecular biomarker signatures** which can be used in clinics. I designed a flexible RNA-seq analysis pipeline including steps: QC and cleaning, mapping, **UMI deduplication** (in collaboration with **Diagenode**) summarization and visualization. I applied gene co-expression network approaches on RNA-seq data to study gene module preservation and information flow. My research projects have led me to work with

RNA-seq data generated using different library kits: Ovation SOLO, CATS Diagenode, BGI, & Illumina. and **miRNAs** generated with qPCR.

Research Fellow - GIGA-Research, BIO3 Unit *November 2015 - January 2017*

Under the supervision of Prof. Kristel Van Steen, head of the BIO3-Medical Genomics Unit, I utilized **prior biological knowledge of RNA-seq** to guide gene-gene interaction for **epistasis analysis**. I developed a pipeline for differential expression analysis that incorporates various **normalization methods** for RNA-seq. By examining patterns of expression changes within gene sets, we can gain valuable biological insights. Furthermore, a **co-expression network** was inferred using the WGCNA method, and **module information** was integrated as part of gene-gene interaction pipeline.

EDUCATION

University of Liege

Ph.D., Bioinformatics (Cancer genomics and Machine learning) *2020*

Thesis: An in-depth study of random forests methodologies for short biomarker signature discovery

Advisors: Prof. Kristel Van Steen and Vincent Bours

Master Bioinformatics and Modeling *2015*

Topic: Development and deployment of a machine learning-driven tool for automated quantification and classification of follicles in high-resolution histological images (Cytomine).

Advisors: Prof. Pierre Geurts

University of Tizi-Ouzou (Algeria)

Computer Science Engineering *2008*

Topic: Design and Implementation of a Multi-function Communication Platform

SELECTED PUBLICATIONS

Assessing Random Forest self-reproducibility for optimal short biomarker signature discovery. **Briefings in Bioinformatics** 2025

Evolutionary genomics of the emergence of brown algae as key components of coastal ecosystems. **Cell** 2024

LncPlankton V1.0: a comprehensive collection of plankton long non-coding RNAs. **bioRxiv** 2023

Differential expression patterns of long noncoding RNAs in a pleiomorphic diatom and relation to hyposalinity. **Scientific Reports** 2023

Differences in plasma microRNA content impair microRNA-based signature for breast cancer diagnosis in cohorts recruited from heterogeneous environmental sites. **Scientific Reports** 2021

Prevalence of histological characteristics of breast cancer in Rwanda in relation to age and tumor stages. **Hormones and Cancer** 2020

SOFTWARE AND TOOLS

votingLNC: LncRNA identification through majority voting ensemble learning. **Figshare** 2023. Software. https://doi.org/10.6084/m9.figshare.24799632.v2

stabFS: a stable feature selection based on random forest and stability measures. **Figshare** 2023. Software. https://doi.org/10.6084/m9.figshare.24878646.v1

MEMBERSHIPS

Since 2023: Societé Phycologique de France SPF

Since 2019: International Genetic Epidemiology Society IGES

Since 2019: Belgian Society of Human Genetics BeSHG

Since 2017: European/American Society of Human Genetics ESHG/ASHG

Since 2016: European Cooperation in Science and Technology COST

SKILLS

Programming & markup languages: R, R Shiny, Bash, Python, Java, SQL, PHP, XML, HTML/CSS, JavaScript, LaTeX

Bioinformatics API: Ensembl, TCGA, GenePattern, R Plumber (for APIs)

Frameworks: Maven, Hibernate, Spring

Databases: MySQL, SQLite

High performance computing: Unix environment, HPC Cluster, Condor, Slurm, Parallel computing

Versioning: GitLab, Github, sourceTree