



# Contents

|                       |     |
|-----------------------|-----|
| 1. Committees .....   | 3   |
| 2. Speakers .....     | 4   |
| 3. Program .....      | 5   |
| 4. Abstracts .....    | 6   |
| 5. Author Index ..... | 105 |

# Committees

# Speakers

**OTÁVIO CABRAL-MARQUES**

**ANDRÉ FONSECA**

**VINÍCIUS MARACAJÁ**

**ANA BENKO**

**ANDREA RIBEIRO-DOS-SANTOS**

**VALDIR BALBINO**

**MARILENE VAINSTEIN**

**AUGUSTO SHRANK**

**MATHEUS PASQUALI**

**FLÁVIA ABURJAILE**

**MARBELLA FONSÊCA**

**DANIEL LANZA**

**AUGUSTO MONTEIRO DE SOUZA**

**RODRIGO ROMÃO**

**GUILHERME OLIVEIRA**

**SÁVIO FARIAS**

**RENAN LEMES**

**DENYS SANTOS**

**ANDRÉ PERES**

**KHALLIL CHAIM**

**RENAN MATSUDA**

**TARCÍSIO COUTINHO**

**LUCYMARA LIMA**

**GILDERLANIO ARAUJO**

**SANDRO SOUZA**

**SHANTANU GUPTA**

**ARIANA SILVA SANTOS**

**PATRICK TERREMATE**

**ADRIÃO DUARTE DORIA NETO**

**BRUNO FÉLIX DE ARAÚJO**

**JOÃO PAULO LIMA**

**RODRIGO DALMOLIN**

# Program

# Abstract

**Disclaimer:**

The contents of the following pages were produced using author-supplied copy. No responsibility is assumed for any claims, instructions or methods: it is recommended that these are verified independently.

# Rare genetic diseases in Brazil: rates exceeding global estimates

Marcos Antonio Nobrega de Sousa<sup>1</sup>, Sávio Torres de Farias<sup>2</sup>

<sup>1</sup>Universidade Federal de Campina Grande, <sup>2</sup>Universidade Federal da Paraíba - UFPB

Reference allele frequencies derived predominantly from foreign populations may mask the burden of specific genetic diseases in Brazil, particularly rare diseases. To assess this underrepresentation in the Brazilian population, data from three public sources (SABE-609 exomes, SABE-WGS-1171 genomes, and BIPMed exomes) were combined with a curated panel of 40 autosomal recessive genes associated with rare diseases, resulting in a dataset of 3,560 alleles (~2,125 individuals). Based on Hardy-Weinberg equilibrium, the Brazilian prevalence rates derived from combined allele frequencies of pathogenic or likely pathogenic variants exceeded global rates (based on gnomAD) for: homocystinuria (1:14,455 vs. 1:909,235; 63×), Wilson disease (1:12,377 vs. 1:404,202; 33×), and phenylketonuria (1:5,989 vs. 1:113,394; 19×). An enrichment analysis (Katz log, Fisher's exact test) against gnomAD was also performed, increasing the number of robust variants from 12 to 40 and revealing 33 population-specific signals that were undetectable in isolation, such as ABCA4 rs61749409 (51.3×;  $p=1\times10^{-10}$ ) and CBS rs5742905 (10.8×;  $p=7\times10^{-8}$ ). ACMG benignity criteria (BA1/BS1/BS2) reclassified 116 VUS/conflicting variants in healthy elderly individuals (27% SMN1 frequency, including one homozygote; HFE with 26 aggregated homozygotes), thereby reducing diagnostic ambiguity regarding these genes in the Brazilian population. AlphaMissense pathogenicity scores independently validated the classifications (median 0.445 for P/LP—pathogenic/likely pathogenic—variants vs. 0.137 for VUS—variants of uncertain significance); 18 enriched missense alleles mapped to stable AlphaFold regions, supporting their functional relevance. These results align with Nunes et al. (2025), who analyzed 2,723 Brazilian genomes and identified 8.7 million previously unknown variants. The analyses demonstrate that Brazilian allele frequencies, when aggregated from complementary cohorts, reveal a disease burden exceeding global estimates—a finding crucial for prioritizing national policies on genomic medicine and newborn screening. It is emphasized that the coordinated use of public resources and bioinformatics tools was fundamental to achieving these results.

# Exploring the Regulatory Landscape of COVID-19: Short Tandem Repeats as Potential Determinants of Disease Severity via cis-eQTL Effects

Matheus Azevedo Bomfim<sup>1</sup>, João Luiz de Lemos Padilha Pitta<sup>2</sup>, Beatriz Souza Toscano de Melo<sup>2</sup>, Gabriel Freitas Araújo<sup>1</sup>, Laryssa Bandeira de Melo Silva<sup>1</sup>, Pablo Rafael Silveira Oliveira<sup>3</sup>, Marcus Villander Barros de Oliveira Sá<sup>4</sup>, Rodrigo Feliciano do Carmo<sup>5</sup>, Luydson Richardson Silva Vasconcelos<sup>6</sup>, Túlio de Lima Campos<sup>1</sup>

<sup>1</sup>Instituto Aggeu Magalhães - FIOCRUZ, <sup>2</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco,

<sup>3</sup>Universidade Federal da Bahia, <sup>4</sup>Real Hospital Português, Recife, Pernambuco, Brazil., <sup>5</sup>Laboratório Avançado de Diagnóstico e Estudos em Saúde e Ambiente (LADESA), Colegiado de Medicina, Universidade Federal do Vale do São Francisco (UNIVASF), Petrolina, Pernambuco, Brazil., <sup>6</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, Fundação Oswaldo Cruz, Fiocruz, Recife, Pernambuco, Brazil.

**Introduction:** Short Tandem Repeats (STRs) are repetitive DNA sequences (1–6 bp) that can influence phenotype by modulating gene expression as cis-expression quantitative trait loci (cis-eQTLs). Recent studies indicate that non-coding elements regulate host responses to SARS-CoV-2, the etiological agent of the COVID-19 pandemic, which has caused at least 7 million deaths worldwide. While previous work has focused mainly on viral STRs, we investigated non-coding regulatory STRs with potential cis-eQTL activity, which may modulate host response and influence COVID-19 severity. **Materials and Methods:** This cross-sectional study analyzed whole-genome sequencing (WGS) data from 168 unvaccinated Brazilian patients (<60 years) hospitalized with severe COVID-19, stratified into survivors and non-survivors. Genomic data were processed via BWA-GATK and STRs identified by STRling. To prioritize biologically relevant candidates, we integrated public single-cell RNA sequencing (scRNA-seq) data from human lung tissue (Delorey et al., 2021), retaining only lncRNAs expressed in lung cell populations. We then refined candidates by comparing the largest allele lengths between groups, retaining loci with non-overlapping distributions. Finally, variants were annotated using the UCSC Genome Browser, assessing cis-eQTL potential through transcription factor binding sites, histone marks, chromatin interactions, and repetitive element annotation. This framework enables identification of STRs within regulatory nodes potentially modulating gene expression in COVID-19 severity. **Results and Discussion:** We identified 7 STRs within lung-expressed long non-coding RNAs (lncRNAs). Six overlapped in genes with high lung expression or repression and H3K27ac enhancer marks, and five showed high-confidence transcription factors binding ( $p < 0.0001$ ) in highly variable repeat regions, consistent with regulatory STR signatures. Notably, 4 STRs overlapped with high-confidence cis-eQTL regions in GTEx (score = 1000), including two loci with high causal posterior probabilities (0.72–0.92). Among them, an STR within the lncRNA NEAT1, previously associated COVID-19 severity, emerges as a strong regulatory candidate. **Conclusion:** Evidence suggests that non-coding STR variation contributes to transcriptional dysregulation in severe COVID-19. Although experimental validation is required, especially for mononucleotide STRs, these results highlight STRs as functional regulatory variants in host response.

# **TANBAQ: Tumor Antigen & Biomarker with Cross-ReActivity Risk Quantification**

Laíze Barros Pantoja<sup>1</sup>, Pedro Lucas Alves da Silva<sup>2</sup>, Moisés Abraão Sousa de Oliveira<sup>3</sup>, Helenize Catarina Moreira Costalat<sup>2</sup>, Andre Luis Fonseca Faustino<sup>4</sup>

<sup>1</sup>Universidade federal do Pará, <sup>2</sup>Universidade Federal do Pará, <sup>3</sup>Universidade Federal do Rio Grande do Norte, <sup>4</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Instituto Metr pole Digital (IMD); Federal University of Rio Grande do Norte (UFRN)

T cell-based immunotherapy is one of the main strategies for cancer treatment; however, its effectiveness depends on the identification of tumor epitopes capable of eliciting a specific immune response without triggering cross-reactivity with healthy tissues. Although the increasing availability of immunological data has driven the development of computational approaches for tumor antigen prioritization, information on epitopes, HLA molecules, T cell receptors (TCRs), and different cancer types remains scattered across multiple databases. In this context, TANBAQ (Tumor Antigen & Biomarker with Cross-ReActivity Risk Quantification) was developed to integrate these data and incorporate in silico predictions relevant to immunotherapy. A computational pipeline was developed in R and Python to integrate and curate epitope, MHC assay, and T-cell assay data across eight cancer types, sourced from the IEDB and CEDAR databases. After removing duplicate records and entries lacking experimental evidence, 222,060 unique epitopes were retained. Gene mapping was performed using a three-step strategy that employed UniProt identifiers and protein names, successfully assigning gene symbols to 221,144 epitopes (99.59%), while 916 unresolved epitopes were excluded from further analyses. Peptides were converted into 9-amino-acid sliding windows (9-mers), resulting in 421,950 records corresponding to 228,534 unique 9-mers and 193 HLA alleles distributed across the eight cancer types. The consolidated database also incorporates in silico predictions of MHC presentation/binding, immunogenicity, and cross-reactivity risk, enabling a more comprehensive characterization of the immunotherapeutic potential of tumor epitopes. The integration of experimental data and in silico predictions into a single platform enhances the identification and prioritization of tumor epitopes with greater immunotherapeutic relevance. The expansion of TANBAQ to multiple cancer types demonstrates the potential of a modular architecture to support research in cancer immunology, precision medicine, and immunotherapy development. TANBAQ consolidates experimental immunological data and in silico predictions into a standardized and comprehensive platform. This approach facilitates the exploration and rational prioritization of candidate epitopes for experimental validation, contributing to the development of new immunotherapeutic strategies.

# **In silico analysis of flavonoid interactions with Sortase A from Staphylococcus aureus: identification of potential inhibitors of bacterial virulence**

Marina de Lourdes Ribeiro Alves Coelho<sup>1</sup>, Jennifer Gabrielly de Sousa Pereira<sup>2</sup>, Marcos Antonio Nobrega de Sousa<sup>3</sup>, Rafael Trindade Maia<sup>2</sup>

<sup>1</sup>Universidade Federal da Paraíba, <sup>2</sup>universidade federal da paraiba, <sup>3</sup>Universidade Federal de Campina Grande

Bacterial resistance demands new therapeutic targets, highlighting the enzyme Sortase A, essential for the virulence of Gram-positive bacteria such as *Staphylococcus aureus*. Flavonoids such as quercetin, rutin, fisetin, apigenin, and kaempferol have been investigated for antimicrobial potential. In this study, *in silico* analyses were performed to evaluate the interaction of these compounds with Sortase A, aiming to identify potential inhibitors and understand their interactions with the enzyme's active site. The three-dimensional protein structure was obtained from the Protein Data Bank (ID 1T2P) and prepared in PyMOL, selecting chain A due to the presence of catalytic residues His120, Cys184, and Arg197. Additional chains, water molecules, and non-essential components were removed, resulting in a suitable receptor for docking. Ligands were obtained from PubChem and prepared using Open Babel with hydrogen addition, Gasteiger charges, energy minimization, and conversion to PDBQT format. Docking was performed using AutoDock Vina, with the grid centered on the catalytic region (center\_x = -40.033; center\_y = -18.643; center\_z = -1.073), dimensions of 25×25×25 Å, and exhaustiveness set to 8. Poses were ranked by binding energy (kcal/mol), and the best conformation was analyzed in PyMOL. The best-performing compounds were subjected to predictive ADMET evaluation using ADMETlab 3.0. Results indicated affinity for Sortase A, with binding energies ranging from -6.769 to -8.387 kcal/mol. Apigenin showed the highest affinity (-8.387), followed by fisetin (-8.203), quercetin (-7.523), kaempferol (-7.465), and rutin (-6.769). The Sortase A-apigenin complex showed interactions with GLY167, PRO163, THR93, ALA92, and LEU169, suggesting stability, while the Sortase A-fisetin complex showed contacts with residues such as GLY167, VAL168, LEU169, and SER116. In ADMET screening, apigenin showed higher affinity but also alerts for hepatotoxicity (DILI), carcinogenicity, genotoxicity, and irritation. Fisetin showed a better toxicological profile, with lower risks for hERG, DILI, carcinogenicity, nephrotoxicity, and neurotoxicity. It is concluded that the evaluated flavonoids have potential as Sortase A inhibitors. Apigenin stood out for its higher affinity, while fisetin showed a better balance between activity and safety, making them promising candidates for future *in vitro* studies aimed at the development of new antimicrobial strategies.

# Detection and Characterization of Non-Retroviral Endogenous Viral Elements (EVEs) across Orchidaceae Genomes

Maria Clara Santos Souza<sup>1</sup>, Felipe Gomes da Silva<sup>1</sup>, Lavinia Santos Prates do Nascimento<sup>1</sup>, Lucas Barbosa de Amorim Conceição<sup>2</sup>, Eric Roberto Guimaraes Rocha Aguiar<sup>2</sup>, Cassio van den Berg<sup>1</sup>, Rogério Mercês Ferreira Santos<sup>1</sup>

<sup>1</sup>Universidade Estadual de Feira de Santana, <sup>2</sup>Universidade Estadual de Santa Cruz

**Background:** Orchids (Orchidaceae) represent one of the most diverse plant families, characterized by high genomic complexity, variable genome sizes (0.3 to 31 Gb), and polyploidy. Endogenous Viral Elements (EVEs) are viral sequences integrated into host genomes that act as "molecular fossils," offering insights into ancient plant-virus interactions and potentially serving as "immune memory" against modern pathogens. **Methods:** This study analyzed 26 genomic assemblies representing 26 orchid taxa, including species from the genera *Dendrobium*, *Phalaenopsis*, *Gastrodia*, and *Apostasia*. The host genomes ranged significantly in size, from 346.6 Mb (*Apostasia ramifera*) to 5.2 Gb (*Ophrys sphegodes*), with GC contents varying between 31% and 44.5%. EVE detection followed a multi-step bioinformatics pipeline: 1) Open Reading Frame (ORF) prediction using Orfipy; 2) Homology searches against a non-redundant viral protein reference database using DIAMOND (blastx) with an E-value threshold of  $\leq 10^{-5}$ ; 3) Removal of redundancy through CD-HIT (clustering at  $\geq 90\%$  similarity); and 4) Validation via online BLASTn/x. Functional domains were identified using HMMER, InterProtScan, and CD-Blast, while genomic mapping and flanking region analysis were performed with BEDTools. **Results:** Hundreds of curated non-retroviral EVEs were identified across the analyzed genomes. The highest abundances were found in *Anoectochilus roxburghii* (~64 EVEs), *Gastrodia elata* (~63 EVEs), and *Bletilla sinensis* (~38 EVEs). The detected elements primarily derive from RNA virus families, with Mitoviridae being the most prevalent (predominantly encoding RNA-dependent RNA polymerase - RdRP). Other significant integrations include Amalgaviridae (fusion proteins), Partitiviridae (capsid proteins and RdRP), and Rhabdoviridae (nucleocapsid proteins). Remnants of ssDNA viruses from the Genomoviridae family (Rep proteins) were also documented. Sequence identities ranged from 40% to 85%, indicating ancient integration events. **Conclusion:** These findings demonstrate that orchids are significant reservoirs of diverse ancient viral sequences. The identification of conserved domains suggests potential functional domestication, providing a framework for future studies on orchid evolution and innate immunity.

# **An Integrated and Low-Cost Bioinformatics Pipeline for Clinical Whole-Exome Analysis Applied to Autism Spectrum Disorder**

Ayron Luigi de Paiva<sup>1</sup>, Thales Henrique de Paiva<sup>1</sup>, Breno de Mello Silva<sup>1</sup>

<sup>1</sup>Universidade Federal de Ouro Preto

Autism Spectrum Disorder (ASD) has a polygenic architecture spanning more than 800 genes catalogued in the SFARI Gene database, and whole-exome sequencing (WES) is now a primary diagnostic strategy; yet the tertiary analysis that converts thousands of raw variants into actionable findings remains a costly technical barrier for most Brazilian laboratories. We developed and validated an integrated, low-cost system for automated clinical exome processing and analysis. The pipeline implements GATK Best Practices: quality assessment (FastQC), alignment to GRCh38 (BWA-MEM), sorting (SAMtools), duplicate marking and base quality recalibration (MarkDuplicates, BQSR), variant calling (HaplotypeCaller), type-specific filtering (VariantFiltration), and functional annotation (Funcotator), followed by Python (pandas) cross-matching against SFARI Gene and a custom prioritization algorithm (AFIS Score) that ranks candidates by gene-level ASD evidence and produces dual technical and plain-language clinical reports. Accuracy was benchmarked against the Genome in a Bottle v4.2.1 truth set for NA12878 (HG001) using hap.py, restricted to high-confidence regions intersected with the exome capture target on GRCh38. The pipeline achieved 99.0% precision and 78.6% recall for SNVs (F1=0.88) and 76.6% precision and 65.5% recall for INDELs (F1=0.71); SNV precision is comparable to commercial platforms, while the lower recall reflects modest sequencing depth and the use of fixed-threshold rather than machine-learning filtering. Applied to the reference sample, 15 variants mapped to SFARI-catalogued genes, demonstrating the prioritization module on real exome data. The system delivers reproducible analysis using exclusively open-source tools at negligible cost relative to commercial solutions, offering an accessible framework for ASD genetic analysis in research and clinical settings in Brazil.

# **In Silico Prospecting of Compounds from *Peumus boldus* Molina Essential Oil as Inhibitors of Aromatase (CYP19A1) and Lysyl Oxidase (LOX) Enzymes via Molecular Docking with AutoDock Vina and Pharmacokinetic and Toxicological Prediction**

Guilherme Matheus Chagas Ribeiro Monteiro<sup>1</sup>

<sup>1</sup>Universidade Federal do Pará

**Introduction and Justification** Cancer and chronic inflammatory diseases involve multiple biochemical pathways. The enzymes Aromatase (CYP19A1) and Lysyl Oxidase (LOX) play central roles in tumor progression, metastasis, and hormonal dysregulation, serving as validated therapeutic targets. Natural products, such as the essential oil of *Peumus boldus* Molina (boldo-do-chile), are rich in bioactive molecules, notably ascaridole (a peroxidated monoterpene). This research is justified by the need to search for new naturally derived inhibitors that can modulate these enzymes with fewer adverse effects, using computational (in silico) methods to optimize time and resources.

**General Objective** To evaluate in silico whether the substances present in the essential oil of *Peumus boldus* Molina act as potential inhibitors of the enzymes Aromatase (CYP19A1) and Lysyl Oxidase (LOX) through molecular docking using AutoDock Vina software.

**Methodology** The research is divided into experimental and computational stages:

**Extraction and Chemical Analysis:** Plant material will undergo hydrodistillation in a Clevenger apparatus for 3 hours to extract the essential oil. The chemical composition will be characterized by gas chromatography coupled with mass spectrometry (GC/MS).

**In Silico Preparation:** The 3D structures of the target enzymes (PDB 3EQM for Aromatase and 6N2W for LOX) will be prepared by removing water molecules and adding charges. Major ligands (especially ascaridole) will be modeled in 2D/3D and geometrically optimized.

**Molecular Docking:** Performed using AutoDock Vina. The method will be validated via redocking (targeting  $\text{RMSE} < 2.0 \text{ \AA}$ ). Binding energies and intermolecular interactions will be analyzed.

**Pharmacokinetic and Toxicological Prediction:** ADMET properties (Absorption, Distribution, Metabolism, Excretion, and Toxicity), as well as cytotoxicity and genotoxicity, will be evaluated using SwissADME and ADMETlab 2.0 platforms.

**Expected Results** Determination of the phytochemical profile of boldo-do-chile essential oil. Creation of a digital molecular library containing the identified compounds. Elucidation of binding modes and thermodynamic interactions between ligands and the Aromatase and LOX enzymes. Characterization of the predicted toxicological and pharmacokinetic profiles of promising compounds. Submission of at least one scientific article to a high-impact peer-reviewed journal.

# Software-Simulated Fragility in Virtual Reality: Grip Telemetry as a Behavioral Instrument for Laboratory Handling Training

João Anisio Marinho da Nobrega<sup>1</sup>, Alyson Matheus Carvalho de Souza<sup>1</sup>

<sup>1</sup>UFRN

**Introduction.** Handling biological material and laboratory glassware demands fine force control: slides, pipettes and test tubes simultaneously impose a slipping threshold and a crushing threshold within a narrow safe window. Virtual environments are increasingly adopted for bioscience training, yet controller-less hand tracking provides no tactile stimulation, producing the "superhuman grip" failure mode. Reproducing fragility through software alone, without haptic gloves, is therefore a requirement for low-cost simulators in biosafety, biological collection digitization and procedural training. **Objective.** To assess whether a multi-modal pseudo-haptic system induces perceptual discrimination of fragility levels and measurable motor recalibration, and to characterize the resulting telemetry as input for computational modeling of grasping behavior. **Methods.** On a Meta Quest 3 with controller-less hand tracking, four feedback layers were coupled to normalized grip intensity  $G$  between 0 and 1: a procedural Voronoi crack shader, a two-layer auditory engine (pitch-modulated stress tone plus a cracking sound above 60% of threshold), a diegetic pressure gauge on the robotic avatar, and Perlin-noise micro-vibration. A within-subjects study ( $N=37$  questionnaires; 53 sessions) required transporting three artifacts of increasing fragility (vase, teacup, test tube). Session logs captured grip-closure time series, breakage events and delivery times, analyzed with Friedman tests, Wilcoxon post-hoc with Bonferroni correction, and Spearman correlations. **Results.** Perceived fragility ( $\chi^2=20.93$ ), sense of care (18.99) and tension (22.38) increased monotonically ( $p<0.001$ ), as did responsibility ( $p=0.011$ ). Telemetry confirmed calibration: breakage rates of 19% (vase), 38% (teacup) and 60% (test tube;  $p=0.003$ ). Mean grip closure also rose monotonically with fragility ( $p=0.017$ ). IPQ Involvement correlated with value attribution ( $r_s=0.582$ ) and sense of care ( $r_s=0.529$ ) for the most fragile object. Two distinct motor regimes emerged, inhibited approach at high fragility and exploratory engagement at medium fragility, with negligible cybersickness. **Conclusion.** Software-only fragility simulation is feasible on consumer hardware and yields structured grip telemetry suitable for training predictive models of motor error and for adaptive per-user threshold calibration, moving bioscience training toward an instrumented, data-driven practice.

# BioRemPP: An Integrated Ecosystem of Tools for Multi-Omics Assessment of Bioremediation Potential

Douglas Felipe de Lima Silva<sup>1</sup>, Lucymara Fassarela Agnez-Lima<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Chemical contamination by priority pollutants remains a global concern, reflected in international regulatory frameworks. Functional, toxicological and regulatory evidence remain scattered across independent databases, hindering multi-omics analysis of these compounds. BioRemPP (Bioremediation Potential Profile) addresses this gap by unifying regulatory prioritisation with functional annotation of bioremediation potential through a compound centric approach, delivered as an ecosystem of four components sharing a curated evidence base consolidated in the BioRemPP core database: the Integrated Database, the Database Explorer (DBX), the analysis based Web Service (BWS) and a command line interface (CLI). The ecosystem serves both web interface users and programmatic pipeline operators. The Integrated Database consolidates regulatory-listed compounds, 20 canonical KEGG xenobiotic metabolism pathways, HADEG and toxCSM in a model built around KEGG Orthology (KO) and KEGG Compound (CPD) identifiers, assembled by a reproducible Snakemake pipeline. This compound centric design delivers functional and toxicological evidence directly anchored to regulatory priority. The design uses interoperable identifiers (KO, CPD, KEGG Reaction, EC, ChEBI, SMILES) for FAIR compliance and multi-omics interoperability. The v1.1.0 core release links 384 priority compounds across 12 chemical classes to 1,543 KOs, 994 EC numbers, 1,939 KEGG reaction identifiers and 206 curated enzyme activity families, anchored to nine regulatory agency codes (IARC, EPA, ATSDR, WFD, PSL, EPC, CONAMA); toxicological predictions from 31 endpoints in five categories (environmental, genomic, organic, nuclear receptor, stress response) cover 370 of these compounds. The DBX exposes five entity centric modules (Compounds, Genes/KOs, Pathways, Toxicity, Compound Classes) and eight guided analysis recipes for open access browsing without registration. The BWS processes user submitted KO profiles through eight analytical modules totalling 56 use cases. The CLI supports four operational modes (genomic, expression, reaction, enzymatic) covering genomic, transcriptomic, metabolomic and enzymatic inputs. All components are centralised in the GitHub organization (<https://github.com/BioRemPP>), with source code under Apache 2.0 and data under CC BY 4.0. BioRemPP makes this evidence accessible to researchers working in bacteria, fungi and plants, supporting bioremediation-focused screening prior to bench studies

# **ALMA: A Nextflow pipeline based on phylogenetic placement for LSU metabarcoding of Arbuscular Mycorrhizal Fungi**

Juliana Rayssa Barros Felix<sup>1</sup>, João Paulo Matos Santos Lima<sup>2</sup>, Bruno Tomio Goto<sup>1</sup>

<sup>1</sup>Universidade Federal Do Rio Grande Do Norte, <sup>2</sup>Universidade Federal do Rio Grande do Norte

The study of Arbuscular Mycorrhizal Fungi (AMF) communities from environmental DNA (eDNA) in soil, root, or leaf litter samples has been developed over the last decades through the employment of modular, manual, and user-dependent tools, representing a barrier of use for professionals with limited computational skills. Furthermore, several tools are limited to taxonomic classifications based solely on sequence clustering by similarity using generic thresholds. In contrast, methods based on phylogenetic placement emerge as a robust alternative, assigning taxonomic identities based on evolutionary relationships. To overcome these limitations, ALMA was developed, an optimized pipeline for metabarcoding analysis of the Glomeromycota rDNA LSU region, structured in Nextflow DSL2 and containerized via Singularity, ensuring reproducibility and portability. For the reference database, a previous dataset of 629 sequences was used and integrated with new data from NCBI GenBank, totaling 739 sequences of the large subunit (LSU) rDNA, covering all 9 orders, 25 families, and 57 described genera of Glomeromycota. The workflow comprises five main stages: (i) quality filtering and chimera removal (Mothur); (ii) sequence clustering (Mothur/VSEARCH); (iii) alignment against the reference database (MAFFT); (iv) phylogenetic placement (RAxML-EPA); and (v) taxonomic assignment by likelihood (GAPPA), in addition to an optional module (vi) for community ecological statistics. For tool validation, a dataset from Amazonian soils of Cropland and Native Forest was used, comparing the pipeline's efficiency before and after the reference database update. With the old database, 87 OTUs were recovered (51 in Cropland and 59 in Forest), displaying low taxonomic resolution, where less than half of the OTUs were classified at the species level, corresponding to only 41.2% (21/51) in Cropland and 33.9% (20/59) in Forest site. In contrast, ALMA detected 61 OTUs, 42 in Cropland and 39 in Forest, and when evaluating taxonomic resolution, species-level identification reached 70% (28/40) and 63.4% (26/41), respectively. Furthermore, ALMA was able to identify 13 species and 7 unique genera that had not been previously captured. With the application of ALMA, it was possible to achieve higher taxonomic accuracy and reduce inconclusiveness, besides enabling robust analyses accessible to researchers with different levels of experience in bioinformatics.

# **OASIS: development of an automated pipeline for retrieval and filtering of orthologous sequences from NCBI for phylogenetic analyses**

Rodrigo Olímpio Orvate Arruda<sup>1</sup>, João Pedro Lemos e Silva Rodrigues<sup>1</sup>, João Paulo Matos Lima Silva<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Obtaining orthologous sequences is a critical step in phylogenetics. Conventional searches via the BLAST web interface frequently introduce noise due to paralogs and isoforms, compromising multiple sequence alignments (MSA). Furthermore, the lack of automated tools for this workflow has become a critical bottleneck in light of changes to NCBI's data infrastructure for 2026. To address this need, this work presents OASIS (Ortholog Alignment & Similarity Screener). It is a command-line pipeline that retrieves orthologs from NCBI and applies dual filtering by identity and similarity, ensuring the generation of reliable datasets ready for evolutionary analyses. Developed in Bash, the pipeline integrates the NCBI Datasets CLI for sequence retrieval, keeping `efetch` as a fallback, and uses NCBI BLAST+ for local alignment, without requiring administrator privileges. The system incorporates an optional taxonomic filter for clade-based searches, a strategy that mitigates response truncation in numerous gene families, as well as a heuristic routine that groups sequences by organism to remove isoforms and redundancies before BLAST. A dynamic alert flags discrepancies in the count of retrieved sequences. The code was structured and documented with the support of generative artificial intelligence (Claude, Anthropic) and is available on GitHub, with versions for Google Colab and Docker/Singularity containers. OASIS demonstrated effectiveness in generating multifasta files (proteins and CDS) in an automated and reproducible manner. Among the limitations identified for future improvement, it should be noted that the NCBI Datasets CLI returns a maximum of 499 sequences per request in specific cases, a restriction mitigated—but not fully eliminated—by taxonomic fractioning. Additionally, heuristic isoform screening does not eliminate the need for manual curation prior to analysis software. It is concluded that OASIS is a robust, accessible solution aligned with NCBI's new infrastructure, providing rigorously filtered datasets for evolutionary studies.

# Integration of Germline and Somatic Variants with Machine Learning Models for Classification of Gastric Cancer in Young Patients

Daniel Henrique Ferreira Gomes<sup>1</sup>, Maria Clara Barros<sup>2</sup>, Júlio César da Silva Filho<sup>1</sup>, Gabrielle Germek Coelho Santos<sup>3</sup>, Ândrea Ribeiro-dos-Santos<sup>2</sup>, Sidney Emanuel Batista Dos Santos<sup>2</sup>, Jorge Estefano Santana de Souza<sup>1</sup>

<sup>1</sup>UFRN, <sup>2</sup>UFPA, <sup>3</sup>Universidade Federal do Rio Grande do Norte

Early-onset gastric cancer (EOGC), diagnosed before age 50, accounts for 10–15% of gastric cancer cases and exhibits a more aggressive phenotype and stronger genetic contribution than conventional disease. However, its germline molecular predisposition remains poorly characterized, particularly in underrepresented Brazilian populations. We developed and benchmarked machine learning models to identify germline variant profiles associated with EOGC using whole-exome sequencing from 375 individuals: 232 cases (95 from Pará, 13 from Rio Grande do Norte, and 124 Korean patients from public databases) and 143 cancer-free controls, including 110 Brazilian Amerindian individuals. Variants were called using DeepVariant and functionally annotated with population, clinical, somatic, and pathogenicity databases. Germline high-functional-impact (gHFI) variants were defined using DTreePred as the primary pathogenicity predictor. Fifteen classifiers were evaluated using three feature representations: gene-level variant counts, z-score-normalized counts, and individual binary variants. Individual-variant features yielded the highest predictive performance, with Random Forest and Extra-Trees achieving 0.97 accuracy and AUC-ROC of 1.00, while Logistic Regression achieved 0.97 accuracy and 0.99 AUC using only 12 interpretable gene-level features. Feature-importance analysis showed that gHFI-derived variables accounted for 73.2% of total importance in the Random Forest model, exceeding the contribution of individual functional gene sets; cancer hallmark and DNA repair/replication genes were the most informative predefined sets. In silico assessment of a top-ranked KCNJ18 variant showed substantial discordance among pathogenicity predictors, supporting the potential utility of DTreePred for candidate-variant prioritization. Overall, these findings indicate that germline high-functional-impact variant burden across multiple cancer-related pathways contains a strong machine-learnable signal distinguishing EOGC cases from controls. This study provides an initial framework for germline-based molecular stratification of EOGC incorporating underrepresented Brazilian populations and highlights the potential of interpretable machine learning to identify distributed genetic susceptibility signals. Somatic variant calling has been completed, with functional annotation and germline-somatic integration underway to further refine molecular stratification of young gastric cancer patients.

# Conserved Neuroimmune Programs Underlie HPV Infection and Cervical Cancer Progression

Yohan Lucas Gonçalves Corrêa<sup>1</sup>, Adriel Leal Nóbile<sup>1</sup>, Júlia Nakanishi Usuda<sup>1</sup>, Roseane Galdioli Nava<sup>1</sup>, Anny Silva Adri<sup>1</sup>, Ronaldy Santana Santos<sup>1</sup>, Igor Salerno Filgueiras<sup>1</sup>, Otavio Cabral Marques<sup>2</sup>

<sup>1</sup>Universidade de São Paulo, <sup>2</sup>USP

Persistent infection with high-risk human papillomavirus (HPV) is the principal cause of cervical cancer, yet the multicellular signalling programmes that accompany disease progression remain incompletely defined. Here, we integrated Single-cell RNA sequencing, independent microarray and bulk RNA-sequencing meta-analyses, and TCGA-CESC tumor data to characterize the neuroimmune transcriptional architecture of HPV-associated cervical disease. Single-cell profiling revealed progressive remodeling of epithelial, stromal, endothelial and immune compartments across HPV infection, high-grade squamous intraepithelial lesions and cervical cancer. Although immune dysregulation predominated, lesion and tumor states showed coordinated enrichment of nervous-system-associated programs, including synapse organization, trans-synaptic signaling, axon development, neurogenesis and neurotransmitter-related signaling. These programs formed interconnected networks with cytokine production, leukocyte activation, chemotaxis, cell adhesion, apoptosis, and immune-effector regulation. Cross-platform integration identified a recurrent 12-gene neuroimmune signature associated with tumor-grade-related transcriptional variation, including CXCL12, P2RY6, P2RX5 and TACR1. Ligand-receptor inference further highlighted glutamatergic, adrenergic, noradrenergic, GABAergic and purinergic signaling circuits. Notably, these signals were distributed across immune, epithelial, stromal and endothelial populations rather than confined to a canonical neural-cell compartment. P2RX5 showed the strongest association with advanced tumor grade, whereas P2RY6, P2RY11 and GABRP displayed cell-type- and stage-dependent regulation. Together, our findings define cervical carcinogenesis as a distributed neuroimmune state in which HPV-associated immune evasion converges with ectopic neurotransmission-related programs, revealing candidate biomarkers and potentially actionable signaling axes in the cervical tumor microenvironment.

# **Development of a multi-epitope B- and T-cell subunit vaccine for the control of Mayaro Virus infection: an immunoinformatics and molecular modeling approach**

Thiago Lima do Nascimento<sup>1</sup>, José Tarcísio Domingos Coutinho<sup>2</sup>, Camila Franco Batista de Oliveira<sup>2</sup>, Ariana Silva Santos<sup>3</sup>

<sup>1</sup>Universidade Federal do Vale do São Francisco, <sup>2</sup>Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campus Princesa Isabel, BR-426 - Zona Rural, Princesa Isabel, Paraíba, Brasil.,

<sup>3</sup>UNIVERSIDADE FEDERAL VALE SSAO FRANCISCO

Mayaro virus (MAYV) is an emerging arbovirus responsible for outbreaks of Mayaro fever, mainly in the Amazon region, representing a significant public health problem due to the lack of licensed vaccines. Therefore, immunoinformatics-based strategies are alternatives to accelerate the development of vaccine candidates. The aim of this work was to develop in silico a multiepitope chimeric protein as a vaccine candidate against MAYV infection. 43 sequences of structural proteins (Nucleocapsid [C], Envelope [E2] and [E3], Protein [6K] and Protein [E1]) were retrieved from NCBI, aligned using Clustal Omega, and used to obtain the consensus sequence. T-cell epitopes (MHC I and II) were predicted using NetMHCpan and NetMHCIIpan software, and B-cell epitopes were predicted using BepiPred-3.0. Epitopes were selected based on antigenicity (Vaxijen), absence of allergenicity (AllerTOP), toxicity (ToxinPred), conservation among viral isolates, and population coverage greater than 70% (IEDB). Approved epitopes were superimposed on vaccine subunits in the construction of a sequence containing a (Met), Human  $\beta$ -defensin 2 (hBD-2), PADRE peptide (Pan-DR Epitope) in the N-Ter and 6 His-Tags in the C-Ter. Physicochemical properties were evaluated using ProtParam and ProteinSol. The 3D structure was modeled on the AlphaFold, ESMFold, and I-TASSER servers. The model chosen with the best structural conformation was from AlphaFold, refined in GalaxyRefine and validated with MolProbity, ProSA-web, and PROCHECK. Of the 417 epitopes identified with IC<sub>50</sub> <400 nM, 35 met the selection criteria, forming 11 vaccine subunits, named SBMV (SBMV-1 to SBMV-11). The chimeric subunit presented antigenicity (0.6955 threshold 5.0>, absence of allergenic and toxic potential, predicted solubility (0.691), instability index of 38.53, aliphatic index of 64.40, and GRAVY of -0.445. Structural validation revealed that 99.2% of the residues are located in the favorable regions of the Ramachandran plot, corroborated by MolProbity analyses and the ProSA-web Z-score. The results indicate that the chimera presents promising structural and immunological characteristics as a vaccine candidate against MAYV. In future stages, molecular docking will be performed with Toll-like receptor (TLR) immune receptors: TLR-2, TLR-3, TLR-6, and TLR-8, simulations of the immune response on the C-ImmSim server, as well as codon optimization for in silico cloning in the Escherichia coli expression vector.

# Differential co-expression analysis: A computational benchmark review in simulated data

Arthur Ribeiro dos Santos<sup>1</sup>, Sandro Jose de Souza<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>3</sup>

<sup>1</sup>Universidade Federal do Pará, <sup>2</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Federal University of Rio Grande do Norte (UFRN); DNA GTx Bioinformatics, <sup>3</sup>Universidade Federal do Rio Grande do Norte

System-level analysis of biological data enables new ways of observing relationships between biological elements by modeling them as networks, among which, differential co-expression (DC) analysis is particularly notable. Although this strategy offers valuable insights for a better understanding of biological mechanisms, DC analysis remains underutilized due to the lack of a "gold standard" tool and a scarcity of comparative studies. Therefore, this study aims to implement a computational benchmark of DC tools to assist in the selection and identification of recommended applications for these approaches. A literature review was conducted to identify and select tools for this computational benchmark, which were filtered based on availability, programming language, and primary development purpose. Analyses were isolated from machine processes using a Singularity virtual container and performed on a node equipped with two 2.0 GHz AMD EPYC 7713 CPUs (128 cores each), 512 GB of DDR4 RDIMM RAM, a 240 GB local SSD and the Rocky Linux 8.5 (Green Obsidian) OS. We applied sampling and randomization techniques in simulated datasets to identify false positives and estimate computational costs across tools. Our review cataloged and compared six DC tools: csdR, dcanr, discordant, diffcoexp, EBcoexpress and iDINGO. A total of 10,385 DC analyses were performed. Our results demonstrated that diffcoexp offers a robust, fast, and high-precision approach for the majority of DC analysis applications. Further, we suggest the use of iDINGO and discordant for analyses aimed at identifying highly significant differentially co-expressed genes and investigating gene-gene interaction behavior, respectively.

# **Bioinformatics in Genetics Education: Development and Application of Practical Classes in K-12 Education**

Naiara Jesus Porcino<sup>1</sup>, Hellen Karla Vital dos Santos<sup>2</sup>, Danilo dos Santos Lima<sup>2</sup>,  
Micaelly Vitória Lima Dos Santos<sup>2</sup>, Thiago Lima do Nascimento<sup>3</sup>, Iris Eduarda da Silva  
Caetano.<sup>2</sup>, Ariana Silva Santos<sup>2</sup>

<sup>1</sup>UNIVERSIDADE FEDERAL VALE SSAO FRANCISCO, <sup>2</sup>UNIVERSIDADE FEDERAL VALE SSAO FRANCISCO,  
<sup>3</sup>Universidade Federal do Vale do São Francisco

The abstract nature of genetics content makes learning it challenging in basic education. In this context, bioinformatics can be employed as an investigative and contextualized pedagogical practice in the classroom. Thus, this study aimed to promote the teaching of genetics in basic education through the application of bioinformatics tools in practical classes. The research was conducted with a second-year class from the Technical Course in Biotechnology at the Professor Paulo Batista Machado State Center for Professional Education in Senhor do Bonfim, Bahia. Genetics topics outlined in the National Common Curricular Base (BNCC) were selected to develop nine practical lesson plans utilizing free bioinformatics tools. A diagnostic questionnaire comprising four sections was also administered to assess students' prior knowledge of bioinformatics and genetics. Twenty-seven students participated in the practical classes. To date, four practical classes have been held. The "Biological Data Real Estate Game" (Banco Imobiliário dos Dados Biológicos) utilized the NCBI database to explore the organization and analysis of biological data. The "From DNA to Protein" class addressed the flow of genetic information using the Biomodel Transcription tool. The "Protein Modeling" and "Evolution Lottery" (Mega-Sena da Evolução) activities utilized SWISS-MODEL for the 3D visualization of proteins and to foster an understanding of molecular evolution concepts. Preliminary results indicate that 62.9% of the students were 16 years old, and 66.7% lived in urban areas. All students had internet access, although 44.4% had never used a desktop or laptop computer. Regarding prior knowledge of genetics and bioinformatics, 96.3% demonstrated an understanding of the relationship between DNA, RNA, and protein synthesis, while 44.4% correctly identified nucleic acids. Only 11.1% stated they were familiar with the term "bioinformatics," yet 40.7% associated it with computing, living organisms, and information technology. Furthermore, 96.3% acknowledged that computational tools aid in understanding genes and proteins, and 84.6% considered that digital resources make classes more interesting. The preliminary results reinforce the potential of bioinformatics as a tool for teaching genetics. Continuing the research—through the implementation of the remaining activities and the post-test questionnaire—will allow for an assessment of its effectiveness regarding student learning.

# **DOWNREGULATED MITOCHONDRIAL-ASSOCIATED GENES ARE POTENTIALLY INVOLVED IN CARDIOVASCULAR SYMPTOMS OF POST COVID-19 CONDITION**

Maria Clara da Costa Barros<sup>1</sup>, Leticia de Aguiar da Costa<sup>1</sup>, Felipe Gouvea<sup>1</sup>, Ândrea Ribeiro-dos-Santos<sup>1</sup>, André Ribeiro dos Santos<sup>1</sup>, Juarez Quaresma<sup>2</sup>

<sup>1</sup>Universidade Federal do Pará, <sup>2</sup>Universidade Estadual do Pará

Although most individuals recover from a COVID-19 infection within a few weeks, some individuals continue to experience prolonged symptoms that can significantly affect their daily functioning and quality of life, defined as Post COVID-19 Condition (PCC). The reasons some individuals develop PCC while others fully recover remain unclear, although dysregulated host immune responses and mitochondrial dysfunction may be involved, leading to oxidative stress and impaired cellular energy production, which can result in symptoms such as cognitive difficulties, shortness of breath and fatigue. Therefore, in order to investigate which mitochondrial-associated genes are transcriptionally involved in the cardiovascular symptoms (CS) of PCC, we reanalyzed the public RNA-Seq dataset GSE215865, selecting 30 individuals with CS (fatigue, chest and muscle pain), without previous coronary artery diseases and not taking the anticoagulant Warfarin as the case group, and 40 individuals without these requisites as the control group. For this purpose, raw reads were processed using the nf-core/rnaseq pipeline, transcript quantifications were imported with tximport, and count data were normalized and analyzed with DESeq2. Differentially expressed genes were identified using  $|\log_2FC| > 1$  and adjusted  $p < 0.05$ , and non-parametric Wilcoxon tests were applied to validate expression differences between groups, resulting in 29 robustly differentially expressed genes. To explore the biological processes associated with these genes, we performed gene set enrichment analysis using the clusterProfiler package. Among the enriched pathways, we identified three genes (SLC2A1, KCNMA1, PDE2A) predominantly related to vascular processes in the circulatory system, response to osmotic stress and regulation of inflammatory response, all potentially linked to the development of CS in PCC. Functionally, PDE2A, located in mitochondrial inner and outer membranes, encodes a phosphodiesterase that modulates cAMP levels, influencing mitochondrial morphology and dynamics, respiration and apoptosis. Likewise, specific isoforms of KCNMA1, a large-conductance  $Ca^{2+}$ -activated potassium channel, have been described in cardiomyocytes, where they contribute to the regulation of mitochondrial membrane potential. Together, these findings support the concept that mitochondrial functionality can be involved in the etiopathogenesis of PCC-related cardiovascular manifestations.

# Genomic Evidence Reveals Polyphyly in Cladophialophora and the Urgent Need for Taxonomic Revision of Chromoblastomycosis Agents

Ricardo Cunha de Oliveira<sup>1</sup>, João Ricardo Guerreiro Duarte<sup>1</sup>, Pablo Diego do Carmo Pinto<sup>1</sup>

<sup>1</sup>Universidade Federal do Pará

Chromoblastomycosis (CBM) is a neglected tropical disease caused by melanized fungi of the family Herpotrichiellaceae. Despite their clinical importance, the evolutionary relationships between the primary etiological agents *Fonsecaea*, *Cladophialophora* and rare etiological species remain unresolved due to limited genetic markers. To address this gap, we conducted a comprehensive phylogenomic analysis to elucidate the evolutionary relationships among the fungal genera responsible for CBM. Based on this, species belonging to the genera *Fonsecaea*, *Cladophialophora*, *Exophiala*, *Phialophora* and *Rhinocladiella* were included, alongside 22 genera from the kingdom Fungi as outgroups. Analyses were performed in R and Python. We obtained species names using taxize package and genomes using biomart. To assess genome completeness and identify orthologous gene sequences, we used BUSCO. After quality control of the 329 downloaded genomes, 84 remained. Single-copy orthologs were grouped, aligned using MAFFT, trimmed with trimAl, and concatenated into a supermatrix using AMAS. The supermatrix and partition file were submitted to IQ-TREE for phylogenetic analysis. We present a comprehensive phylogenomic analysis of the genera associated with CBM, utilizing 84 genomes and a supermatrix of 140 single-copy orthologous genes (176,418 nucleotide sites). Maximum Likelihood and coalescent-based (wASTRAL) analyses resolved the backbone of this family. Topology tests (AU test) confirmed the monophyly of Chaetothyriales ( $p=0.518$ ). However, Gene and Site Concordance Factors (gCF/sCF) revealed complex evolutionary dynamics. *Fonsecaea* and *Phialophora* monophyly is supported across gene trees (BS=100%, gCF~100%). However, we provide genomic evidence that *Cladophialophora* is polyphyletic, comprising at least three distinct lineages. Notably, *C. bantiana* and *C. psammophila* form a distinct clade, while *C. immunda* is sister to *Fonsecaea*. Furthermore, concordance and coalescent analyses highlighted substantial topological conflict regarding *Rhinocladiella*. Although *R. mackenziei* clustered with *Exophiala*, this relationship showed weak gene and site concordance (gCF=10.7%, sCF=34%) and high gene discordance (gDFP=80%). Our findings demonstrate that morphology-based taxonomy is biologically invalid for these fungi. The observed phylogenetic conflict likely reflects the urgent need for a genome-based taxonomic revision of CBM agents.

# **A MACHINE LEARNING FRAMEWORK FOR SEROGROUP CLASSIFICATION OF PATHOGENIC LEPTOSPIRA SPECIES BASED ON RFB LOCUS PROFILES**

Edson de Carvalho Ferreira Filho<sup>1</sup>, Pedro Melo Arruda<sup>1</sup>, Tetsu Sakamoto<sup>1</sup>

<sup>1</sup>UFRN

The bacterial genus *Leptospira* is notable for its generalist nature and high genetic diversity; it is currently classified via serological assays into over 300 serovars, grouped into approximately 30 serogroups. However, this system is complex and prone to inconsistencies, as cross-reactions between antigens can lead to ambiguous results. Furthermore, current serological tests—such as MAT and CAAT—are costly and labor-intensive, requiring work with live cultures and facing challenges regarding inter-laboratory standardization. To address these issues, genomic data from 721 pathogenic *Leptospira* samples were obtained from the NCBI RefSeq and BIGSdb (Institut Pasteur) databases to develop a machine learning-based serological prediction tool. The project focused on the *rfb* locus and was divided into two stages: the first classified the four main serological classes, while the second classified their respective serogroups. The resulting models performed excellently, achieving a perfect score in the first stage, while the subsequent stage yielded an average F1-score of 0.948. Feature importance analysis revealed that serogroup discrimination is driven by specific patterns of gene presence and absence. Consequently, the term "seroclass" was proposed to designate higher-order serological groupings, based on the observed strong genetic correlation between the *rfb* locus and shared antigenic characteristics. Finally, this approach offers a scalable and reproducible alternative to classical serological tests, with potential applications in epidemiological surveillance, outbreak investigation, and vaccine development for the *Leptospira* genus.

# VEGABrain: A Visualizer of Enriched Gene Analysis in development of human brain

Isaac Medeiros do Amaral Costa<sup>1</sup>, João Vitor Cavalcanti<sup>1</sup>, Diego Marques Coelho<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Transcriptomic analysis of human nervous system development is fundamental to understanding neurobiological processes and neurodevelopmental disorders. One of the most relevant datasets in this field is provided by the Allen Human Brain Atlas; however, its visual exploration relies on legacy platforms with limited usability. To address this bottleneck, we present VEGABrain, an interactive web application developed to optimize and facilitate the spatiotemporal visualization of gene expression throughout human cortical and subcortical development. The system processes data from the original atlas, summarized across five developmental time windows, integrating a lightweight architecture with a fast-rendering API. VEGABrain performs direct nomenclature mapping and alignment between the regions of the original atlas and the structures of the ggseg package, enabling precise spatial rendering within the Desikan-Killiany (DK) and FreeSurfer Automatic Subcortical Segmentation (ASEG) systems. The platform offers analyses at two spatial resolutions (microregions and macroregions) and provides three visualization modules: single-gene expression, user-customized gene lists, and enrichment of ontologies/biological pathways from the Human MSigDB collection (Hallmarks, BioCarta, KEGG, Reactome, and Gene Ontology). For the visualization of custom gene lists and genes involved in specific ontologies, a composite GSEA score is calculated using the GSVA package, allowing visualization of average expression within each specific spatial segment displayed by the tool. In summary, VEGABrain establishes itself as an accessible, fast, and intuitive tool, modernizing access to and interpretation of neurodevelopmental transcriptomic data for the scientific community.

# Performance Evaluation of Deep Learning Tools in the Prediction of Structures and Interactions Between Cysteine Proteases and Cystatins

João Vitor Santos de França<sup>1</sup>, João Paulo Matos Santos Lima<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

The study evaluates the performance of deep learning-based artificial intelligence tools (AlphaFold 2, AlphaFold 3, and Boltz) in predicting the structures and interactions of complexes formed by cysteine proteases and their endogenous inhibitors, cystatins. Using 16 crystallographic structures from the Protein Data Bank (PDB) as experimental references, the generated models were compared using quantitative quality and confidence metrics, specifically pLDDT, ipTM, and RMSD. Additionally, computational analyses of residue interaction networks were conducted to map non-covalent bonds at the binding interfaces. Preliminary results indicate that the resolution of the experimental structures directly affects prediction accuracy; reference structures yielded models with an overall average RMSD of 0.42 Å. While all tested tools demonstrated high global structural accuracy (average pLDDT > 94.9 and ipTM > 0.89 across all tools), predicting the exact interface interactions remains a challenge. The Boltz algorithm outperformed the others in specific structural contacts, retaining up to 31% of critical non-covalent bonds (such as  $\pi$ - $\pi$  stacking) at the binding interface. In contrast, AlphaFold 3 showed severe limitations, accurately capturing less than 10% of fine side-chain interactions. The study concludes that AI tools are highly viable for global structural studies of these proteins, but emphasizes the absolute need to integrate these models with complementary approaches, such as molecular dynamics simulations and docking, to overcome current predictive limitations at the binding interfaces.

# **Breast Cancer Malignancy Classification in Mammograms Using a Convolutional Neural Network**

Felipe Alves de Vasconcelos<sup>1</sup>, Robinson Luis de Souza Alves<sup>2</sup>, Francisco Sales de Lima Filho<sup>2</sup>, André Salles Cunha Peres<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte

Breast cancer is one of the most diagnosed neoplasms in women, showing growth trends in national incidence. The main strategy to mitigate the disease is early diagnosis through routine screening, which substantially increases treatment efficacy by detecting cellular proliferation in its early stages. Given the need to ensure high diagnostic precision and mitigate the occurrence of false negatives, computer vision models based on deep learning have become essential tools. This work proposes the development of a convolutional neural model for the automatic classification of mammographic findings into benign and malignant categories. The methodology utilized 1319 samples from the public CBIS-DDSM dataset, applying the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm for image preprocessing and contrast enhancement. For the classification stage, a model based on the DenseNet121 architecture was employed. To prevent overfitting, a cross-validation strategy was utilized, splitting 80% of the data for training and 20% for testing. The model achieved an accuracy of 85%, a precision of 85%, a recall of 86% and an AUC of 95%, which highlights its high generalization capacity and sensibility to false negative cases, consolidating its importance as a robust and reliable computational tool to assist specialists in the early diagnosis of breast cancer.

# Mechanism-Guided LLMs Generate Candidate Sets Enriched for Documented Drug-Excipient Incompatibilities

Matheus Dalmolin<sup>1</sup>, Raquel de Melo Barbosa<sup>2</sup>, César Viseras<sup>2</sup>, Marcelo A. C. Fernandes<sup>3</sup>

<sup>1</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil, <sup>2</sup>Universidad de Granada, <sup>3</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil Department of Computer Engineering and Automation, UFRN, Natal, Brazil

Drug-excipient incompatibility data are scarce, and candidate-generation systems require evaluation against external evidence. We asked whether candidate pairs generated by a mechanism-guided large language model (LLM) pipeline recovered relationships documented in PharmDE, a drug-excipient incompatibility database, more often than expected under a model-accessible Cartesian null and whether their mechanistic annotations agreed with PharmDE. Starting from 344 reported-incompatible parent pairs, the same seven-step chain-of-thought workflow was applied using four OpenRouter endpoints: deepseek/deepseek-v3.2, google/gemini-3.1-flash-lite-preview, google/gemini-3-flash-preview, and x-ai/grok-4.1-fast. Outputs were resolved to PubChem structures and uniformly post-processed, yielding 5,627 candidate records and 3,642 unique canonicalized active pharmaceutical ingredient (API)-excipient pairs. For each model, the null universe was the Cartesian product of APIs and excipients in its candidate-or-parent inventory ( $N=67,452-92,862$ ), and  $K$  included only PharmDE pairs in that universe ( $K=205-219$ ); the pooled universe comprised  $N=146,300$  combinations and  $K=233$  of the 434 canonical PharmDE pairs. One-sided upper-tail hypergeometric tests found 46 pooled PharmDE matches versus 5.80 expected (7.93-fold enrichment;  $p=1.63\times 10^{-27}$ ); all model-specific sets were enriched (6.41-8.18-fold; all  $p\leq 2.35\times 10^{-11}$ ). Among the 37 matched records annotated in both sources (36 unique pairs), diagonal concordance within mapped mechanism classes was 20/37 (54.1%), while broad concordance under an author-defined reconciliation rubric was 33/37 (89.2%; 32/36 after pair-level deduplication). Thus, LLM-generated candidates were enriched for externally documented pharmaceutical relationships under the defined null and provided a plausibility signal for hypothesis prioritization. PharmDE was not an explicit pipeline input, but pretraining exposure cannot be excluded. Moreover, the Cartesian null does not preserve API/excipient marginals or parent-pair clustering, and this external-reference comparison does not independently validate or experimentally confirm individual incompatibilities.

# **Pneumovirus RdRp Dynamics and Allosteric Inhibition: Integrating Quantum Mechanics, AI (Co)folding and Sequence Conservation**

Klaus Augustus Ramos Reiniger<sup>1</sup>, Katyanna Sales Bezerra<sup>1</sup>, Jonas Ivan Nobre Oliveira<sup>1</sup>, João Paulo Matos Lima Silva<sup>1</sup>, Umberto Laino Fulco<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Human Respiratory Syncytial Virus (RSV) and Human Metapneumovirus (hMPV) impose a severe global healthcare burden due to the lack of targeted antiviral therapeutics. Recently, MRK-1 was synthesized as a potent in vitro pneumovirus inhibitor targeting a novel allosteric site within the RNA-dependent RNA polymerase (RdRp). However, the lack of resolved co-crystal structures leaves its precise catalytic and inhibitory mechanisms unclear. Here, we combined quantum mechanical calculations using the Molecular Fractionation with Conjugate Caps (MFCC) scheme under Density Functional Theory (DFT) with positional mapping, sequence variability analysis (MSA, phylogeny, and Shannon entropy), and AI-based (co)folding predictions across distinct polymerase configurations and binders. This integrated approach mapped amino acid residues potentially implicated in allosteric communication and functional transitions. Energetic profiling revealed that ligand rearrangement within the binding pocket drives a conformational switch in interacting residues, accompanied by distinct shifts in binding energies between homologous, 100% conserved pairs: LEU1372\_RSV/LEU1297\_hMPV, ASN1369\_RSV/ASN1294\_hMPV, PHE1385\_RSV/PHE1310\_hMPV, VAL1384\_RSV/VAL1309\_hMPV, and ASP1380\_RSV/ASP1305\_hMPV. Residues with the strongest attractive energies map adjacent to or within Motif E of the polyribonucleotidyltransferase domain, which maintains the core structural scaffold, while interior residues (e.g., GLU1194\_hMPV) provide key stabilizing thermodynamic contributions. An oscillating Shannon entropy profile across L indicates that sequence diversity preserves critical enzymatic mechanisms and supports flexible region-driven conformational switches between functional states. Beyond the binding pocket, structural and sequence evidence highlights an unprecedented role for the conserved phosphoprotein tetramer interface in modulating substrate-bound RdRp architectures. Although cofolding predictions failed to place MRK-1 within its allosteric site, predicting previously unmodeled cryo-EM regions, particularly RdRp Motif F, provides deeper insight into elongation-prone architectures. Collectively, these findings reinforce MRK-1 as an allosteric inhibitor that stalls critical conformational transitions, refining pneumovirus RdRp topology and setting the stage for future workflows incorporating ensemble contact networks, post-translational modifications, and M2-1-containing complexes.

# **Lumos Networks: A Reproducible Python Ecosystem for Transcriptomics and Systems Biology**

Laís de Carvalho Gonçalves<sup>1</sup>, Djorkaeff Oliveira Fontinele<sup>1</sup>, Rodrigo Olímpio Orvate Arruda<sup>1</sup>, Shantanu Gupta<sup>1</sup>, João Paulo Matos Santos Lima<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

The transition from raw transcriptomic data to systemic biological interpretations represents a computational bottleneck. Conventional pipelines require switching between fragmented environments, increasing the learning curve and susceptibility to technical errors. This study developed and validated Lumos Networks, a unified open-source ecosystem in Python designed to democratize complex transcriptomic analyses and the modeling of regulatory networks. The platform was built as a multi-page web framework using Streamlit and containerized via Docker, ensuring environmental reproducibility. The processing methodology is divided into three sequential modules: DDEA (Diagonal Differential Expression Alley): Responsible for robust statistical modeling using the PyDESeq2 engine. Count data are modeled assuming a negative binomial distribution. The module applies empirical Bayesian variance estimation (shrinkage) to stabilize the fold change of genes with low counts, minimizing technical noise. Differential expression analysis is performed using Wald tests, with strict control of the false discovery rate (FDR) via the Benjamini-Hochberg method; APP (Arithmancy Pathway Profiler): Performs functional enrichment of the obtained molecular signatures using the GSEAPy framework. The module performs both Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA), calculating enrichment scores to identify activated or repressed biological pathways. The tool provides automated integration with curated ontology and interaction repositories, including KEGG, GO, STRING, and JASPAR; PrioriGraph (PG): Identifies regulatory causality. The module uses the NetworkX library to overlay differential expression data onto interactions validated by prior biological knowledge (TRRUST database). By constructing directed graphs, the algorithm applies advanced topological metrics, specifically eigenvector centrality and in/out-degree calculations, to mathematically identify Genetic Hubs and Master Regulators. Lumos Networks eliminates the need for distributed computing infrastructure and command-line proficiency. By unifying rigorous statistics (PyDESeq2) with prior-knowledge-driven graph theory, the tool enables researchers to rapidly identify critical regulatory hierarchies in complex pathologies, establishing a validated, fast, and fully reproducible workflow.

# **A Real-Time Integration Architecture for AI-Driven Mammography Triage and Clinical Workflow Monitoring**

Pedro Henrique Álvares Róbias<sup>1</sup>, Rayssa Beatriz Ribeiro Cavalcante<sup>1</sup>, César Rennó-Costa<sup>2</sup>, André Salles Cunha Peres<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

In diverse healthcare settings, clinical exams are often processed sequentially without risk-based prioritization (FIFO), delaying the identification of situations that require immediate clinical attention. This challenge is particularly critical in mammography screening, where the timely referral of high-risk findings directly impacts patient outcomes. This work presents the design, implementation, and end-to-end validation of an integration system that bridges automated exam ingestion, artificial intelligence-based triage, and real-time clinical monitoring. The developed solution operates as an automated interface that detects exams immediately after acquisition and seamlessly injects them into a classification pipeline. The system relies on a FastAPI backend that continuously monitors an input directory, simulating immediate image collection from a mammography unit. Through a modular interface, the system feeds these images into a deep learning model based on a residual architecture (originally adapted by NYU Langone Health using public pre-trained weights), which returns a binary critical or non-critical flag. The results automatically feed a Grafana dashboard designed to provide comprehensive operational metrics, including the number of exams performed per period (daily, weekly, and monthly), the volume of exams awaiting a report, the average time between image acquisition and referral confirmation, and the number of critical cases identified within a given timeframe. Simultaneously, a Prometheus metrics store continuously records these values, enabling a Grafana-managed alerting rule to notify staff when unresolved high-risk exams accumulate, ensuring that critical cases are prioritized for reporting. Focusing on architectural viability, the evaluation confirmed that the proof-of-concept operated seamlessly, demonstrating a continuous data flow with no communication bottlenecks between the acquisition, classification, and monitoring layers. These results demonstrate a functional, reproducible integration interface successfully connecting clinical workflow to real-time monitoring infrastructure.

# Characterization of rare germline variants in Brazilian cancer cohorts using DTreePred

Gabrielle Germek Coelho Santos<sup>1</sup>, Daniel Henrique Ferreira Gomes<sup>1</sup>, Julio Cesar da Silva Filho<sup>1</sup>, Beatriz Stransky<sup>1</sup>, Jorge Estefano Santana de Souza<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Genomic variants are DNA alterations that occur at the individual or population level. Understanding their origin and clinical implications is essential to precision oncology. The Brazilian population, shaped by a complex miscegenation, harbors a unique genomic background that may influence cancer susceptibility. Yet, Brazilian genomic data remain underrepresented in major databases, limiting functional interpretation and hindering precision medicine strategies. This study aimed to establish a methodology to identify and characterize rare germline genetic variants in Brazilian individuals using next-generation sequencing (NGS) data, investigating their potential association with cancer predisposition using DTreePred, a machine-learning and classic predictors based tool (Gomes et al., 2025). As a case study, we reanalyzed NGS data from a Brazilian cohort of 17 high-grade serous ovarian cancer (HGSOC) samples (Amorim et al., 2025). Reads were quality-controlled and aligned to the GRCh38 reference genome, followed by germline variant calling and functional annotation. Variant pathogenicity assessment was performed with DTreePred and other complementary predictors, such as CADD, MutPred2 and EVE. Driver/passenger prediction was conducted with CScape and Cancer Genome Interpreter (CGI) tools. From an initial set of 2,012 identified variants, 4 novel variants were reclassified as candidate pathogenic non-synonymous variants by DTreePred. These variants were absent from major genomic databases, like ABraOM, The 1000 Genomes Project and gnomAD, and lacked definitive clinical classification in resources like ClinVar and SIFT. Furthermore, the ensemble of complementary predictors mostly supported the results with high pathogenicity scores. CScape supported their candidate driver status, while CGI categorized 3 of the 4 variants as unclassified due to their novel, population-specific nature. These findings emphasize the potential of deep learning tools like DTreePred to enhance variant interpretation in underrepresented populations. By uncovering novel candidate driver variants, this study expands the catalog of Brazilian genomic diversity and establishes a framework for future functional studies in precision medicine.

# **Metagenomic Insights into the Microbial Diversity, Resistome, and Biosynthetic Potential of Mangrove and Restinga Ecosystems in Paraíba, Brazil**

Caio Gabriel Tavares Ferreira<sup>1</sup>, João Vitor Cavalcanti<sup>1</sup>, Gleison Medeiros de Azevedo<sup>1</sup>, Esther Maria Pacheco Rocha<sup>2</sup>, Augusto Schrank<sup>3</sup>, Marilene Henning Vainstei<sup>2</sup>, Valdir de Queiroz Balbino<sup>3</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidade Federal do Rio Grande do Sul,  
<sup>3</sup>Universidade Federal de Pernambuco

The MaRe Network is a project aimed at investigating different aspects of biodiversity in coastal ecosystems. In Northeastern Brazil, these environments are under increasing anthropogenic pressure, while their microbial communities remain underexplored in environmental studies. This study analyzed seven samples from mangrove and restinga environments in the state of Paraíba, collected at different sites in the municipalities of Mamanguape and Cabedelo. Metagenomic analysis was performed using the nf-core/mag v4.0.0 pipeline, orchestrated by the Nextflow workflow manager and executed using the Singularity profile. We obtained 1,549 bins, which were refined into 137 metagenome-assembled genomes (MAGs). Additional refinement with CAP3 significantly reduced the number of contigs by eliminating paralogous and repetitive regions. Restinga areas showed greater richness, alpha and beta diversity, and taxonomic abundance at the phylum level than mangrove areas, which exhibited a more specialized profile dominated by specific groups. Both environments showed similar functional profiles, with a predominance of COG category C (energy production and conversion). Regarding antibiotic resistance genes (ARGs), restinga samples showed a higher abundance of *adeF*, whereas the Barra de Mamanguape estuary displayed greater heterogeneity, with *vanW* (*vanI* cluster) predominating at Site 1 and *vanT* at Site 2. The heavy metal resistance gene *copR*, however, showed high relative abundance across all sites. Analysis of biosynthetic gene clusters (BGCs) revealed a predominance of terpene biosynthesis in restinga samples, whereas the estuary showed a higher prevalence of polyketide synthase (PKS) clusters. These findings indicate that restinga environments may support greater microbial adaptability to environmental stress through terpene biosynthesis and multidrug efflux (*adeF*). In contrast, the estuary appears to respond to metal and nutrient inputs through metabolically specialized pathways, glycopeptide resistance, and the production of complex secondary metabolites (PKS), suggesting distinct local anthropogenic pressures.

# **Marauder's GenoMap: An Integrated Pipeline for De Novo Assembly and Protein Family Discovery**

Djorkaeff Oliveira Fontinele<sup>1</sup>, Laís de Carvalho Gonçalves<sup>1</sup>, Rodrigo Olímpio Orvate Arruda<sup>1</sup>, João Pedro Lemos e Silva Rodrigues<sup>1</sup>, Shantanu Gupta<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

The development of new high-throughput sequencing technologies, such as NGS, has provided major breakthroughs in understanding previously unelucidated biological processes and components, such as genes and proteins. However, it has also led to a massive accumulation of raw data available. This is especially problematic for non-model species, where genome and transcriptomic reconstruction rely on de novo assembly. In this process, raw sequencing reads are assembled into longer continuous segments (contigs). After this reconstruction, the entire genome is scanned to identify coding sequence regions for an accurate and effective protein homology search. Although reliable, the algorithm requires significant computational power for assembly and command-line knowledge for its execution, limiting its use by researchers lacking technical familiarity or robust infrastructure. Given the vast amount of data aggregated in online databases, and aiming to facilitate access for multiple researchers, the development of user-friendly tools has become essential to efficiently analyze and interpret the enormous volume of online biological data. Marauder's GenoMap was built on a modular architecture, utilizing shell scripts that integrate third-party tools and featuring a graphical user interface (GUI). Its interface was constructed using the Python library PyQt6, alongside other libraries for data analysis and contig trimming. Its modular organization streamlines the workflow into four main steps: acquiring raw genetic data from the NCBI SRA database, assembling these files, probing for homologous proteins, and retrieving the results. Following workflow, the next step after the assembly is identifying homologous proteins by searching for conserved domains using profiles based on Hidden Markov Models (HMMs) by combining probabilistic alignment with heuristic and vector-parallel algorithms, this synergy enables exhaustive searches at BLAST-like speeds, significantly optimizing overall computational efficiency. Although the tool does not bypass the computational cost required for genome assembly, it offers a highly intuitive graphical interface. This accessibility to any researchers effectively eliminates the need for programming expertise, enabling the research for protein homology search in non-model species.

# **Integrated transcriptome and proteome profiling reveals extensive post-transcriptional regulation of the AreA-governed nitrogen response in *Paracoccidioides brasiliensis***

Thalison R Moreira<sup>1</sup>, Maria Eduarda Dourado Peixoto<sup>2</sup>, Laura de Sousa Lopes Moreira<sup>2</sup>, Orville Hernandez Ruiz<sup>3</sup>, Juliana Alves Parente-Rocha<sup>1</sup>, Célia Maria de Almeida Soares<sup>2</sup>, Clayton Luiz Borges<sup>2</sup>

<sup>1</sup>Universidade Federal de Goiás, <sup>2</sup>Universidade Federal de Goiás, <sup>3</sup>Universidad de Antioquia

Paracoccidioidomycosis, the most prevalent systemic mycosis in Latin America, is caused by the thermotrophic fungus *Paracoccidioides brasiliensis*. The fungus's ability to assimilate nitrogen within nutrient-restricted host niches is critical for both its growth and virulence. This process is governed by nitrogen catabolite repression (NCR) under the GATA master regulator AreA. The manner in which distinct nitrogen sources simultaneously remodel the fungus at the transcript and protein levels remains poorly defined. We investigated the growth of wild-type Pb18 under various nitrogen sources, in biological triplicates, against a rich-medium control (brain-heart infusion, BHI). The samples were profiled using paired RNA-seq (DESeq2) and label-free quantitative proteomics (data-independent acquisition, DIA-NN) under a single significance rule (adjusted  $p < 0.05$ ). Each source triggered an extensive transcriptional response, with 4,935 to 5,110 differentially expressed genes identified, as well as a far more focused proteomic shift, with approximately 860 differentially abundant proteins identified per source. This ~6:1 disparity indicates post-transcriptional buffering. The induction of AreA and NCR targets was observed, with proline demonstrating a specific activation of the proline-utilization pathway. In contrast, the urea/allantoin branch (Dur1,2; UreD; Dal1) was only observed at the protein level. Translation and oxidative phosphorylation were transcriptionally repressed, while sulfur metabolism and autophagy were induced. A striking finding was the poor correlation between transcript and protein fold-changes genome-wide (Spearman's  $\rho$  from  $-0.04$  to  $-0.12$ ). This discordance persisted in the Pro-versus-Gln contrast ( $\rho = -0.07$ ), suggesting true post-transcriptional regulation rather than a medium-richness effect. This notion is exemplified by the glutamate dehydrogenase genes *Gdh2* and *Gdh3*, which exhibited both transcriptional induction and protein diminution. Nonetheless, unsupervised sparse multivariate integration recovered a shared gene-protein axis on which the informative nitrogen signal lay at the protein layer. The nitrogen source effectively reprograms the AreA-governed metabolism in *P. brasiliensis*, predominantly through post-transcriptional control. This underscores the indispensability of proteome-level evidence for interpreting NCR and prioritizing candidate determinants of fungal growth and virulence.

# Achieving 30% Methionine Enrichment in Bean Phaseolin via $\Delta\Delta G$ -Guided Mutations, AlphaFold3, and High-Pressure Molecular Dynamics

Lucas Araújo de Queiroz<sup>1</sup>, Silva, J.R.P.<sup>2</sup>, Pegado, F. J.O.<sup>3</sup>, Ortega, J. M.<sup>2</sup>

<sup>1</sup>Universidade Estadual da Paraíba, <sup>2</sup>Universidade Federal de Minas Gerais, <sup>3</sup>Universidade Federal do Rio Grande do Norte

Beans are a vital source of dietary amino acids, though their inherently low methionine content usually requires complementary pairings, such as the classic combination with rice, to achieve nutritional completeness. To address this limitation, we developed an automated, high-throughput in silico pipeline for the targeted enrichment of proteins with specific amino acids. Our approach is based on iterative substitutions guided by predicted changes in folding free energy ( $\Delta\Delta G$ ) using the mCSM application. Here, we applied this methodology to phaseolin, a major reserve protein in common beans. We designed mutants with 10%, 20%, 30%, 40%, and 50% methionine enrichment by selecting optimal substitution sites ranked by decreasing  $\Delta\Delta G$  values, prioritizing stabilizing mutations. To evaluate structural integrity, the 3D structures of the variants were first modeled using a local AlphaFold3 server. Subsequently, the mutants underwent molecular dynamics (MD) simulations using NAMD3 under high pressure (1000 atm), a condition we previously demonstrated to accelerate conformational sampling, yielding insights comparable to much longer simulations at 1 atm within just 5 ns. Trajectory analyses revealed that phaseolin tolerates up to 30% methionine substitution while maintaining a globular structure indistinguishable from the wild-type protein. As a negative control, substitutions targeting the lowest (most destabilizing)  $\Delta\Delta G$  values led to rapid structural instability at merely 10% replacement. Ultimately, the application of this fast, bioinformatics-based pipeline highlights the biotechnological potential of engineered methionine-enriched phaseolin to enhance the nutritional profile of cultivated beans.

# **Conserved transcriptional signatures of oxidative stress and DNA damage: a meta-analysis in zebrafish models of Fetal Alcohol Spectrum Disorders**

Romério de Oliveira Lima Filho<sup>1</sup>, Caio Maia Farias de Andrade<sup>2</sup>, Ana Carolina Luchiar<sup>1</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidade Potiguar

Maternal alcohol consumption is a global public health concern that can cause a range of neurodevelopmental alterations collectively known as Fetal Alcohol Spectrum Disorder (FASD). Understanding of its mechanisms has advanced through experimental models, among which zebrafish is widely used to study embryonic ethanol effects, given its genetic homology to humans. Individual transcriptomic studies, however, vary in platform, developmental stage, and tissue analyzed, limiting robust, comparable signatures across studies. Here, we aimed to identify such cross-study signatures through a transcriptomic meta-analysis integrating four public zebrafish ethanol-exposure datasets, spanning distinct developmental stages (pre-gastrulation, 8 hpf, 7 dpf, and 13 dpf) and platforms (microarray and RNA-Seq). Differential expression matrices were extracted directly from the microarray datasets, while RNA-Seq data were processed with *nf-core/rnaseq* and *nf-core/differentialabundance*. Results from all four studies were mapped to human orthologs via *biomaRt* and integrated using a random-effects model (*metafor*). We identified 362 genes with significant combined effects (196 up-regulated, 166 down-regulated), with *IRF1* showing the strongest signal. Enrichment analyses (GO, KEGG) showed activated genes linked to antioxidant and detoxification responses, including metallothionein-mediated metal-ion response (copper, cadmium) and xenobiotic metabolism (cytochrome P450, retinol metabolism), reflecting oxidative stress from ethanol metabolism via *Adh1/Cyp2e1*. Suppressed genes were associated with muscle development and *PI3K-Akt/JAK-STAT* signaling. PPI network analysis organized these genes into 18 functional clusters, the largest linked to DNA damage response, ER stress, and mRNA destabilization, reinforcing the role of oxidative stress, genomic damage, and proteostatic dysfunction in the ethanol transcriptional response. Despite platform and stage heterogeneity, the meta-analysis revealed convergent genes and pathways tied to antioxidant response, xenobiotic metabolism, and DNA damage signaling, pointing to conserved mechanisms of ethanol-induced dysregulation. Next steps include functional validation of candidates, particularly metallothioneins and *IRF1*, in zebrafish loss- and gain-of-function models, followed by cross-validation in mammalian systems, to establish their causal role in FASD pathogenesis and translational relevance to human neurodevelopment.

# **Regulatory roles of lncRNAs in memory Breg cells in multiple sclerosis**

Caio César Maia<sup>1</sup>, Maria Bárbara Borges<sup>1</sup>, Vinícius Maracajá Coutinho<sup>2</sup>, Thaís Gaudencio do Rêgo<sup>3</sup>, Marcelo Tigre Moura<sup>3</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidad de Chile, <sup>3</sup>Universidade Federal da Paraíba

Non-coding RNAs (ncRNAs) comprise sequences that do not yield protein products but play essential regulatory roles. Among them, long non-coding RNAs (lncRNAs), molecules with more than 200 nucleotides, stand out as central regulators of gene expression. The dysregulation of these lncRNAs has been associated with diseases affecting the Central Nervous System, such as multiple sclerosis (MS) – an autoimmune and neurodegenerative condition characterized by inflammation, demyelination, and axonal loss. The pathogenesis of MS, previously understood under a T cell-centric paradigm, is now recognized for the fundamental role of B cells, with treatments associated with their depletion serving as a therapeutic alternative. The dysfunction of subpopulations, such as memory B cells (memB), has been implicated in the break of self-tolerance. However, adjacent molecular mechanisms, especially those involving lncRNAs, remain poorly elucidated. Thus, this work aimed to investigate the transcriptomic profile of memB cells defined by TIM-1 and TIGIT in MS through the analysis of public single-cell RNA sequencing data generated by the Smart-seq2 method. Differential gene expression analysis between samples from MS patients identified 15318 genes with altered expression, with lncRNAs being the second most abundant class after protein-coding genes. Among the 3879 upregulated protein-coding genes in patients, pro-inflammatory factors such as AHNK and KLF7 are prominent. Notably, the lncRNA-focused analysis revealed the upregulation of LINC01362, ROCK1P1, and LINC01725, associated with pro-inflammatory pathways like JAK-STAT and NF-κB, and the downregulation of lncRNAs with suppressive functions, such as CARD8-AS1 and LINC02446. Finally, functional enrichment analysis reinforced these findings, pointing to the dysregulation of immune signaling pathways and a decline in cell survival processes. In conclusion, the results characterize a transcriptomic dysfunction in regulatory memB cells and elucidate regulatory roles of lncRNAs, highlighting targets with potential for therapeutic investigations in MS.

# Inflammatory Signature of Chronic Schizophrenia: Integrating Bibliometrics and Meta-Analytical Evidence

Ronaldy Santana Santos<sup>1</sup>, Roseane Galdioli Nava<sup>2</sup>, Marina dos Santos Barreto<sup>3</sup>, Yohan Lucas Gonçalves Corrêa<sup>2</sup>, Adriel Leal Nóbile<sup>2</sup>, Júlia Nakanishi Usuda<sup>2</sup>, Anny Silva Adri<sup>2</sup>, Débora Gomes de Albuquerque Freitas<sup>4</sup>, Fernando Yuri Nery do Vale<sup>5</sup>, Amanda Rodriguez Valera<sup>6</sup>, Igor Salerno Filgueiras<sup>2</sup>, Luciana Bezerra de Souza Gianasi<sup>7</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>8</sup>, Otavio Cabral-Marques<sup>8</sup>

<sup>1</sup>University of São Paulo, <sup>2</sup>Universidade de São Paulo, <sup>3</sup>Federal University of Sergipe, <sup>4</sup>Interunit Postgraduate Program on Bioinformatics, University of São Paulo (USP), <sup>5</sup>RCSI Centre for Systems Medicine and Dept. of Physiology & Medical Physics, RCSI University of Medicine and Health Sciences, Dublin, Ireland, <sup>6</sup>Institute of Biomedical Sciences, University of São Paulo (ICB-USP), <sup>7</sup>Federal University of Southern and Southeastern Pará (Unifesspa), Brazil. Digital Metropolis Institute, Multidisciplinary Bioinformatics Center, Federal University of Rio Grande do Norte (UFRN), <sup>8</sup>Universidade Federal do Rio Grande do Norte

Psychoneuroimmunology highlights the role of immune dysregulation in schizophrenia. However, the inflammatory signature of its chronic phase remains inconsistent due to prolonged exposure to antipsychotics and metabolic burden. This study combines bibliometric mapping with a systematic review and meta-analysis to comprehensively evaluate peripheral cytokine profiles in chronic schizophrenia. Following PRISMA and MOOSE guidelines, we searched PubMed, Embase, Scopus, and Web of Science through May 6, 2025, for observational studies comparing blood-derived cytokines in adults with chronic schizophrenia (mean duration  $\geq 5$  years) versus healthy controls (HCs). Random-effects models calculated standardized mean differences (SMDs) using Hedges'  $g$ . Concurrently, bibliometric networks were mapped using metadata to analyze geographic trends, term co-occurrence, and thematic intersections. Of 4,504 records identified, 22 primary studies ( $N = 4,495$ ) were included. The bibliometric analysis revealed an exponential increase in publications after 2010, led by the United States and China, evidencing a strong network linking clozapine/olanzapine treatment and clinical symptoms to cytokine dynamics. The meta-analysis demonstrated a restructured systemic pro-inflammatory state, dominated by innate immunity and the acute-phase axis. Patient cohorts exhibited significant elevations in CRP (SMD = 0.51), IL-1 $\beta$  (1.03), IL-8 (0.17), and IL-10 (0.26), with a trend toward elevated IL-6, while adaptive immune markers (IFN- $\gamma$ , IL-2, TNF- $\alpha$ ) demonstrated no overall changes. "Leave-one-out" and subgroup analyses revealed untreated patients had significantly lower levels of acute-phase cytokines, underscoring the confounding metabolic impact of antipsychotics. Standardized correlations with clinical severity (Fisher's  $Z$ ) indicated that TNF- $\alpha$  and IL-8 correlated positively with symptom severity, while CRP and IL-6 correlated inversely, suggesting distinct clinical-immunological biotypes. Chronic schizophrenia is characterized by a restructured peripheral immune profile, strongly shaped by disease chronicity and antipsychotic-induced metabolic burden. Peripheral cytokines do not represent static diagnostic markers, but serve as promising clinical tools for the biological stratification of patients and targeted neuroimmunological interventions.

# **Integrative Boolean Network Modeling Reveals miR-34a Sequestration by lncRNAs MALAT1 and NEAT1 as the Core Driver of Chemoresistance in Ovarian Clear Cell Carcinoma**

Douglas Nunes da Silva<sup>1</sup>, Epitácio Dantas de Farias Filho<sup>2</sup>, Shantanu Gupta<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>2</sup>

<sup>1</sup>UFRN, <sup>2</sup>Universidade Federal do Rio Grande do Norte

Ovarian clear cell carcinoma (OCCC) is an aggressive histological subtype. It is characterized by high intrinsic chemoresistance to platinum-based agents and unfavorable clinical outcomes. Recurrent genetic alterations in ARID1A and PIK3CA are well documented. However, the stability of the resistant phenotype stems from complex transcriptomic plasticity and non-linear regulatory circuits. These circuits are sustained by competitive endogenous RNA (ceRNA) networks. In this study, we implemented an integrative systems biology framework to address this complexity. We combined multilevel RNA sequencing transcriptomic mapping with the construction and dynamic simulation of an asynchronous Boolean regulatory graph using GINSim software. The dataset comprised bulk RNA-seq from 11 OCCC samples and 6 normal ovarian controls from GTEx. Logical modeling quantified differential expression across three complementary layers: Differential Gene Expression (DGE), Differential Transcript Expression (DTE), and Differential Transcript Usage (DTU) via DESeq2 and DRIMSeq. MicroRNA-target interactions were incorporated strictly based on physical binding evidence from AGO2-CLIP-seq repositories. Finally, stochastic logical perturbations across 1,000 trajectories were evaluated. Dynamic simulations and topological analyses identified four positive feedback loops essential for sustaining cellular bistability states. Remarkably, *in silico* deactivation of miR-34a emerged as the sole topological event capable of completely collapsing the tumor resistance attractor. This highlights its role as a master regulatory hub and central effector of the p53-mediated DNA damage response pathway. Mechanistically, we demonstrated that malignant state persistence and apoptosis evasion depend on the hierarchical sequestration of miR-34a by the upregulated lncRNAs MALAT1 and NEAT1. These lncRNAs act as high-affinity molecular sponges to block the repression of key oncogenes such as MYC and STAT3. Virtual knockout of MALAT1 or NEAT1 disarticulated the pillars of bistability and restored platinum sensitivity. Additionally, we mapped the parallel, isolated miR-145/miR-200/ZNF217 axis involved in cellular plasticity. In conclusion, chemoresistance in OCCC is not a diffuse phenomenon. Rather, it is an emergent property locked by specific topological architectures. Strategic targeting of ceRNA hubs provides a novel blueprint for precision therapies aimed at overcoming platinum resistance.

# Use of Perturbation Models for Target Prioritization Across Cell Types in Alzheimer's Disease

Morais, B. P. R.<sup>1</sup>, Júlia A. de Amorim<sup>1</sup>, Marcos Romualdo Costa<sup>2</sup>, Diego Marques Coelho<sup>1</sup>

<sup>1</sup>UFRN, <sup>2</sup>Multiususer Bioinformatics Center, UFRN

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive and memory decline. It involves the accumulation of  $\beta$ -amyloid (A $\beta$ ) plaques, intracellular neurofibrillary tangles composed of hyperphosphorylated tau protein, and progressive synaptic loss. Current treatments for AD primarily focus on alleviating symptoms through acetylcholinesterase inhibitors and NMDA receptor antagonists. However, recent approaches have sought to modulate neuroinflammation, regulate metabolic pathways, and promote neuronal protection, thereby complementing existing therapeutic strategies. In this context, prioritizing therapeutic targets through cell type-specific gene regulatory network modeling is crucial to understanding the specificity of interventions and potential pharmacological side effects. This study screened candidate therapeutic targets through perturbation analysis of co-expression networks using single-cell transcriptomic (scRNA-seq) data from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD, middle temporal gyrus and dorsolateral prefrontal cortex), aiming to identify cell subtype-specific perturbation patterns across AD progression. A panel of 182 distinct candidate genes was established, comprising 76 GWAS-derived AD risk genes and approximately 110 druggable targets curated from National Library of Medicine clinical trials database. Gene regulatory networks were inferred via GENIE3, and perturbation scores were computed for each cell type across early (Braak I-III) and late (Braak IV-VI) disease stages using scRank with an antagonist and agonist binding model, while also evaluating APOE4 genotype and sex distributions across stages and regions, applying a significance threshold at the global top 2.5% of scores. A total of 43 candidate targets exhibited significant perturbation across 23 cell types. These targets were categorized as either cell-type selective (1-2 cell types, n=27) or non-selective (3+ cell types, n=16). Among the selective targets, 11 were exclusive to the early stage, 11 to the late stage, and 5 were shared across both, highlighting distinct temporal windows of cell-type vulnerability during AD progression. Ultimately, these findings demonstrate that single-cell RNAseq, stage-resolved perturbation analysis is a strategy for prioritizing safer and more precise therapeutic targets in AD.

# **In Silico Characterization and AlphaFold 3 Modeling of Dehydrins Across Three Plant Species**

Euller Barbosa Tenório Cavalcanti<sup>1</sup>, Raquel Silveira de Araújo<sup>1</sup>, Jakob Schlüter<sup>2</sup>, Rafael Lucas da Silva<sup>1</sup>, Saulo Rafael Mendes Penna<sup>1</sup>, Agnes Angélica Guedes de Barros<sup>1</sup>, Cínthia Carla Claudino Grangeiro Nunes<sup>1</sup>, Ana Maria Benko Iseppon<sup>1</sup>

<sup>1</sup>Universidade Federal de Pernambuco, <sup>2</sup>Freiherr-vom-Stein-Gymnasium Rösrath

Dehydrins (DHNs) are Late Embryogenesis Abundant (LEA) proteins involved in plant responses to abiotic stresses. This study aimed to identify and characterize DHNs in *Oropetium thomaeum* (Poaceae), *Eutrema salsugineum* (Brassicaceae), and *Calotropis procera* (Apocynaceae) using an in silico approach. Candidate proteins were identified using the Pfam dehydrin profile PF00257 through HMMER and validated by Batch CD-Search. A lineage-specific HMM profile was also constructed to refine candidate identification. Physicochemical properties, subcellular localization, conserved motifs, and sequence relationships were assessed using ProtParam, WoLF PSORT, MEME, and Neighbor-Joining (NJ) phenetic analysis, respectively. Structural modeling was performed using AlphaFold 3, with models generated in the absence and presence of Ca<sup>2+</sup>, Zn<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>3+</sup>, and Co<sup>2+</sup>. Eight candidate DHNs were identified across the three species, comprising three in *O. thomaeum* (OthDHN1-3), two in *C. procera* (CprDHN1-2), and three in *E. salsugineum* (EsaDHN1-3). The eight DHNs ranged from 102-265 amino acids and 12.78-26.93 kDa, with predicted pI values of 6.28-9.85 and characteristically highly negative GRAVY scores, ranging from -1.234 to -0.864. Subcellular localization analysis predicted a predominantly nuclear distribution for all DHN sequences. The motifs analysis identified the presence of typical DHNs canonical motifs, such as K, S, Y, F and truncated K motifs among the eight DHNs. In the phenetic analysis, the eight DHNs were compared with seven reference sequences from *Arabidopsis thaliana* to assist in sequence group delineation. The NJ tree revealed clustering associated with conserved DHN motif architectures, including KS, Kn, YK, YnSKn, and SKn configurations. Structural modeling of DHNs from the three species showed poorly defined conformations in the absence of ions, whereas models generated with ions displayed more defined predicted secondary structures. Notably, despite global disorder, all identified K segments were predicted as  $\alpha$ -helical regions with good (pLDDT 70-90) or great (pLDDT > 90) confidence, supporting the structural relevance of these conserved motifs. However, these predictions were insufficient to establish ion-specific binding or affinity. Overall, the results highlight the diversity of DHN motif architectures and physicochemical properties across the three species and support the use of lineage-specific HMM profiles for DHN discovery and characterization.

# Mining Biosynthetic Gene Clusters in *Helicobacter pylori* Genomes Associated with Premalignant and Malignant Gastric Phenotypes

Carlos Eugênio Costa de Lima<sup>1</sup>, Wilson Araújo da Silva Junior<sup>1</sup>

<sup>1</sup>Universidade de São Paulo

**Background:** Biosynthetic gene clusters (BGCs) encode microbial secondary metabolites with documented roles in tumor biology, yet BGC mining had never been applied to *Helicobacter pylori* in gastric carcinogenesis. Building on an unpublished pilot study by the same group on two metagenome-assembled *H. pylori* genomes, this work-in-progress scales the approach to a statistically powered comparison across phenotypes. **Methods:** We mined 554 preassembled NCBI *H. pylori* genomes (179 cancer, 375 precursor/chronic-gastritis) with nf-core/funcscan (antiSMASH, DeepBGC, GECCO). Lacking a cross-tool coordinate system, we built a tool-agnostic Gene Cluster Family (GCF) workflow combining hmalign-based backbone identity, for verified classes (43.1% of Jaccard-qualifying pairs), with Pfam-domain Jaccard similarity elsewhere, via DBSCAN (epsilon set by a sensitivity sweep). GCFs carried by  $\geq 3$  genomes were tested via Fisher's exact test, continent-stratified CMH, and multivariable logistic regression, with Benjamini-Hochberg correction across the full tested set; richness was also modeled via quasi-Poisson GLM with a phenotype-by-age interaction. **Results:** Of 2,076 filtered BGC predictions, 88.1% grouped into 135 non-singleton GCFs; the dominant GCF (554 BGCs, terpene-precursor locus) was phenotype-invariant ( $p \sim 1$ ). BGC richness per genome was modestly but robustly higher in cancer genomes (median 4 vs 3; Wilcoxon  $p = 5.96e-5$ ); GCF-level richness also differed significantly despite identical medians. Among 85 eligible GCFs, none reached significance after correction by Fisher's test, CMH, or logistic regression. A quasi-Poisson model showed a borderline cancer effect on BGC richness (IRR=1.119,  $p = 0.0487$ ), a robust age effect (IRR=1.005/year,  $p = 0.0047$ ), and no interaction between said effects ( $p = 0.864$ ). **Conclusion:** Results indicate a subtle, aggregate rise in biosynthetic richness in cancer genomes without a single dominant GCF driving it, consistent with a restricted, conserved *H. pylori* repertoire where phenotype effects manifest quantitatively rather than through discrete functional families. Backbone-domain identity was the dominant driver of GCF resolution. Next steps: correlating GCF repertoire with phylogeny, testing combinatorial multi-GCF malignancy signatures, extending backbone identity to GECCO and remaining DeepBGC classes, and functional validation of candidate BGCs in an animal model.

# **BIN1 Genotype-Associated Microglial-State Shifts May Influence Alzheimer's Disease eQTL Interpretation**

Maurício Matheus Araújo Silva Galvão<sup>1</sup>, Júlia A. de Amorim<sup>1</sup>, Marcos Romualdo Costa<sup>2</sup>

<sup>1</sup>UFRN, <sup>2</sup>Multiususer Bioinformatics Center, UFRN

The BIN1 locus contains genetic variants associated with increased risk of Alzheimer's disease (AD). Single-cell eQTL studies suggest that the AD-associated variant rs6733839 may be linked to increased BIN1 expression in microglia. Because BIN1 expression varies across microglial states, genotype-associated expression differences may partly reflect altered state representation rather than direct cis-regulation. We therefore hypothesized that genotype-associated shifts in microglial-state composition could influence microglial eQTL interpretation. We analyzed single-cell RNA-sequencing data from 32,657 microglial cells from 75 donors with AD-related neuropathology, including 11 homozygous risk-allele carriers and 64 non-recessive individuals. After normalization, 3,000 highly variable genes were selected, followed by principal-component analysis and construction of a 15-neighbor K-nearest-neighbor graph; cells were organized into 13 transcriptional states. Using Milo, a graph-based differential-abundance framework that models partially overlapping KNN neighborhoods, we tested 2,201 neighborhoods with edgeR quasi-likelihood models. Six recessive-genotype models were evaluated: three primary models (genotype-only, adjusted for age and sex, and adjusted for age, sex, and Braak stage) and three sensitivity models using age, AD neuropathologic change (ADNC), Braak stage, and Thal phase in different combinations. Robustness was assessed by leave-one-donor-out (LODO) analysis, repeating each primary model after excluding one donor at a time. Differential abundance testing identified 28 neighborhoods significantly enriched in recessive individuals across all three primary models. Of these, 25/28 mapped to activated-phagocytic/PVM or inflammatory-metabolic states, and 13 remained significant across all six specifications, indicating consistency across covariate choices. LODO analyses showed that effect direction was generally preserved, although the age/sex/Braak-adjusted model was more sensitive to sample composition. In summary, these findings support our hypothesis that the recessive BIN1 genotype is associated with shifts in microglial-state composition that may influence the interpretation of microglial eQTL signals, particularly through enrichment of activated-phagocytic/PVM and inflammatory-metabolic programs. However, donor-level sensitivity indicates that the magnitude of this association varies across individuals and requires independent replication.

# Analysis of Transcriptograms in Epithelial-Mesenchymal Transition (EMT)

Odilon Julio dos Santos<sup>1</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>, Rita M. C. de Almeida<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidade Federal do Rio Grande do Sul

Single-cell RNA sequencing (scRNA-seq) has revolutionized gene expression analysis. However, high dropout rates and stochastic noise frequently reduce the information retrieved from these experiments. The Epithelial-Mesenchymal Transition (EMT), fundamental to tumor progression and organismal development, is particularly challenging to characterize fully due to the existence of intermediate states. In this work, we demonstrate that projecting transcriptomic data onto protein-protein interaction (PPI) networks acts as a "biological low-pass filter," attenuating technical noise and increasing the statistical power of tests. We proposed and validated an innovative pipeline that integrates the Transcriptogram method with Principal Component Analysis (PCA). By applying a moving average over functionally ordered genes, we drastically increased the signal-to-noise ratio, enabling the inference of cellular trajectories. The method was applied to a public MCF10A cell dataset induced by TGF- $\beta$ 1, with rigorous batch effect correction based on biological controls. The results revealed that EMT is not merely a morphological change, but a coordinated systemic reprogramming. The approach allowed for the identification of critical modules that would remain hidden in conventional analyses: (i) a massive "Metabolic Switch," indicating the transition to Oxidative Phosphorylation to sustain invasion; (ii) a strategic cell cycle blockade; and (iii) a "Detoxification Shield" and chemoresistance, characterized by the endogenous activation of metallothioneins. We conclude that the combination of PPI network topology and dimensionality reduction offers superior resolution for dissecting cellular plasticity. The method not only validates classical markers but also exposes the hidden functional architecture of the transition, showing that it is not a single transition, but that cells can describe different trajectories, stopping at different stages of differentiation.

# **FOLR2+ Tissue-Resident Macrophages are shaped by spatial context and associated with lymphocyte profiles in colorectal cancer**

Igor Salerno Filgueiras<sup>1</sup>, Débora Gomes de Albuquerque Freitas<sup>2</sup>, Emmanuelle Nascimento Quagliato<sup>2</sup>, Kianni Hernandez Freitas<sup>2</sup>, Michel Mouawad<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>3</sup>, Otavio Cabral-Marques<sup>3</sup>, Jérôme Galon<sup>4</sup>, Rodrigo Nalio Ramos<sup>2</sup>

<sup>1</sup>Universidade de São Paulo, <sup>2</sup>D'Or Institute for Research and Education, IDOR, São Paulo, Brazil,

<sup>3</sup>Universidade Federal do Rio Grande do Norte, <sup>4</sup>INSERM, Sorbonne Université & Université de Paris, Equipe Labellisée Ligue Contre le Cancer, Centre de Recherche des Cordeliers, Laboratory of Integrative Cancer Immunology, Paris, France

The increasing incidence of colorectal cancer (CRC) among younger individuals, has intensified the need to understand the disease's progression mechanisms. Macrophages can exert both pro- and anti-tumoral functions, reflecting their heterogeneous origins, phenotypes, and spatial distribution within tissues. Here, we aimed to determine how FOLR2+ tissue-resident macrophages (TRMs) interact with the tumor and non-tumor microenvironment, and how these interactions relate to disease progression since these cells are already established before carcinogenesis. Using bioinformatics approaches and both original and public datasets, our study investigated FOLR2+ TRMs and TREM2+ monocyte-derived macrophages (MoDMs). In samples from 61 patients with CRC, we acquired multispectral immunohistochemistry data from histological slides, as well as RNA-seq data. Together, these were used to examine the spatial patterns of macrophage subsets and their interactions with CD3+ cells in non-tumoral regions, the invasive margin (IM), and the center of the tumoral lesion (CT). While TRMs are the most abundant subtypes in non-tumoral regions, their frequency in IM and CT is associated with the accumulation of T lymphocytes in these regions. By enriching lymphocyte signatures in the bulk RNA-seq of our samples, we observed that FOLR2+TRMs in IM and CT are associated with Tfh and B lymphocytes, cell types linked to tertiary lymphoid structures (TLS), whereas TREM2+MoDMs in CT are associated with Tregs. These associations were also linked to tumor extent (pT) and regional lymph node involvement (pN) staging. In CRC slides obtained through collaboration, interactions between FOLR2+TRMs and CD3+, CD4+, and CD20+ cells can be observed at the periphery of TLS. These results, supported by multiple techniques, reinforce the notion that macrophage and lymphocyte interactions in CRC are shaped by tissue context. This may influence tumor progression and, consequently, patient survival.

# Genomic characterization of a *Bacillus cereus* isolate with potential entomopathogenic activity

Beatriz Godoy Roncalli<sup>1</sup>, Marcelo Soares de Barros<sup>1</sup>, Beatriz Souza Toscano de Melo<sup>2</sup>, João Luiz de Lemos Padilha Pitta<sup>2</sup>, Rosemary Aparecida Roque<sup>3</sup>, Ricardo de Melo Katak<sup>4</sup>, Tatiany Patrícia Romão<sup>5</sup>

<sup>1</sup>Instituto Aggeu Magalhães, <sup>2</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, <sup>3</sup>Instituto Nacional de Pesquisa da Amazônia- INPA, <sup>4</sup>Instituto Nacional de Pesquisa da Amazônia-INPA, <sup>5</sup>Instituto Aggeu Magalhães Fiocruz/PE

The discovery of novel biocontrol agents against insect vectors relies on bioinformatic approaches integrating genome assembly, taxonomic classification, and functional annotation. This study investigated the taxonomic identity and entomopathogenic potential of bacterial isolate L49I, previously identified as promising in larvicidal assays. Genomic DNA was sequenced on an Illumina MiSeq platform, quality-checked using FastQC/MultiQC, assembled with Unicycler, and annotated with Prokka. *L. sphaericus* 2362 and *B. thuringiensis* serovar israelensis HD-789 were analyzed using the same pipeline for comparison. Taxonomic identification by rMLST, TYGS (GBDP), and GTDB-Tk consistently classified L49I as *Bacillus cereus* (ANI=97.24%). The draft genome comprised 5,888,068 bp across 66 contigs (N50=623,266 bp; GC=34.83%), using 6,001 CDS, seven rRNA genes, 83 tRNA genes, and one tmRNA. Putative pesticidal proteins were screened with BtToxin\_Digger, and biosynthetic gene clusters were identified with antiSMASH. The predicted proteome was then compared with 1,159 proteins from the Bacterial Pesticidal Protein Resource Center (BPPRC,v2.0.0) using local BLASTp, and candidate proteins were further characterized with InterProScan. Among six filtered loci, two were identified as high-priority candidates. IIICOIKO\_03788 showed 80.5% identity and 98% coverage to Spp1Aa1 (E-value=0), as well as additional similarity to Spp1Aa2. InterProScan identified cholesterol-dependent cytolysin domains, placing this protein in the same structural group as sphaericolysin and perfringolysin O. IIICOIKO\_03795 showed similarity to Tpp78/Tpp80 proteins with 31.86–34.22% identity, 85–92% coverage and E-values of approximately  $10^{-50}$ ). Its broad sequence coverage and highly significant E-values support its classification as a distant Tpp-family homolog, reinforcing the prediction generated by BtToxin\_Digger. InterProScan further identified insecticidal crystal toxin P42 and Ricin B-like lectin domains, an architecture characteristic of Tpp/Bin proteins associated with receptor recognition and pore formation. AntiSMASH detected biosynthetic gene clusters associated with known larvicidal metabolites, suggesting that the observed phenotype may involve Spp- and Tpp-related proteins rather than classical Cry-toxins. Overall, this workflow confirmed the taxonomic identity of L49I and identified two high-priority candidate pesticidal proteins for functional validation and further biocontrol research.

# GO-SYSTEMS: A Systems Classification Tool for Genes and Biological Processes

Thiago Jesus da Silva Xavier<sup>1</sup>, Júlia Nakanishi Usuda<sup>2</sup>, Adriel Leal Nóbile<sup>2</sup>, Roseane Galdioli Nava<sup>2</sup>, Ronaldy Santana Santos<sup>3</sup>, Yohan Lucas Gonçalves Corrêa<sup>2</sup>, Anny Silva Adri<sup>2</sup>, Amanda Rodriguez Valera<sup>4</sup>, Igor Salerno Filgueiras<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>, Otavio Cabral-Marques<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidade de São Paulo, <sup>3</sup>University of São Paulo, <sup>4</sup>Institute of Biomedical Sciences, University of São Paulo (ICB-USP)

The biological ontologies provide means to organize biological knowledge in a systematic, directional manner, with specific relations between terms. Further, many are collaborative and continuously updated and expanded, such as the Uberon and Gene Ontology (GO). The Uberon (Uberon Multi-species Anatomy Ontology) is an anatomical ontology which relates body parts, systems, and tissues within positional, morphofunctional and phylogenetic contexts. The Uberon is integrated with the GO, providing intercommunication between their terms. The GO is one of the largest and most used biological ontologies, bridging the relationships between genes and terms across three domains: biological processes (BPs), cellular components (CCs) and molecular functions (MFs). However, despite presenting cell-, tissue-, organ- and system-specific terms, the GO does not present an anatomical categorization of these, nor of genes. Here, we sought to systematically categorize GO terms and functionally related genes within fifteen main human anatomical systems, and to provide these results as a freely available tool. Fifteen human anatomical systems were selected from Uberon: respiratory, cardiovascular, circulatory, skeletal, musculoskeletal, endocrine, exocrine, digestive, nervous, excretory, genitourinary, reproductive, integumental, immune, and lymphoid systems. Sequentially, for each system, were obtained: the descendant anatomical terms, GO BPs related to the anatomical terms, descendant BP terms, and genes annotated to descendant BPs. Additionally, a manual curation of the BPs was performed to select those specific to humans, and their related genes, enhancing their accuracy for this species. The data was processed using R and RStudio. The system which presented most BPs and genes were the nervous and immune systems, possibly due to their extensive study, allied to the vast amount and diversity of functions specific to these. The sharing of genes was visibly larger than the sharing of BPs, in accordance with the pleiotropic nature of genes and the specific definition of BPs. The resulting tables and code are publicly available at <https://github.com/JNUsuda/uberon-go>. While our application focused on human systems, the GO-SYSTEMS tool can be used to classify terms specific to organs, tissues, and cells. Further, it can also be used with different organisms, and to classify other GO domains, the CCs and MFs, expanding its applications beyond the provided use case.

# **One-Class Approaches for Detecting Melanoma-Associated Transcriptomic Alterations: Comparison of Autoencoder and Classical Machine Learning Methods Across Independent Cohorts**

EMILLY YNIALIDE RODRIGUES DE LIMA<sup>1</sup>, Meryck Felipe Brito da Silva<sup>1</sup>, Rafael Souto<sup>1</sup>

<sup>1</sup>PUC-GO

Melanoma heterogeneity motivates computational strategies to identify transcriptomic alterations. One-class anomaly detection learns healthy profiles and identifies deviations without tumor samples during training. We compared an autoencoder, One-Class SVM (OC-SVM), Isolation Forest, and PCA-Mahalanobis across GEO cohorts GSE3189, GSE46517, GSE15605, GSE100050, GSE4587, and GSE7553. Models were trained exclusively on healthy controls; test samples were excluded from feature selection, normalization, fitting, and thresholding. Performance was assessed through cross-study and external testing in GSE7553 using ROC-AUC, AUPRC, balanced accuracy, and MCC, together with stage-stratified, differential expression, reproducibility, and functional enrichment analyses. In GSE7553, OC-SVM achieved ROC-AUC 0.982 (95% CI: 0.938–1.000), AUPRC 0.999, and balanced accuracy 0.920, outperforming the autoencoder (ROC-AUC 0.915, 95% CI: 0.799–1.000; AUPRC 0.994). Performance varied by disease stage: for primary melanoma, ROC-AUC was 0.786 for the autoencoder and 0.929 for OC-SVM, whereas for metastatic melanoma it reached 0.975 and 1.000, respectively. Cross-cohort analyses identified reproducible alterations involving PRAME, S100B, MITF, CDK2, and BIRC5. Functional enrichment highlighted interferon-gamma response, TNF- $\alpha$  signaling via NF- $\kappa$ B, inflammatory response, and epithelial-mesenchymal transition (FDR<0.10). One-class approaches identified melanoma-associated transcriptomic deviations in an independent cohort, with stage-dependent performance. Reproducible molecular alterations and enriched immune/inflammatory pathways support the biological relevance of the detected signals. OC-SVM outperformed the autoencoder, suggesting that classical anomaly-detection methods may outperform deep learning in this setting when transcriptomic sample sizes are limited.

# **METAGENOMIC CHARACTERIZATION OF A LICHEN ECOSYSTEM OF THE GENUS *Coenogonium* sp. IN THE ITAPIRACÓ PROTECTED AREA, SÃO LUÍS, MARANHÃO**

Marcos Davi de Sousa Campos<sup>1</sup>, Weneson Victor Diniz Sarges<sup>2</sup>, Emanuelle Costa Ramos<sup>2</sup>, José Isaías Pimentel Barros<sup>2</sup>, Anna Letícia Silva da Costa<sup>2</sup>, Hivana Patricia Melo Barbosa Dall'Agnol<sup>2</sup>, Tânia Keiko Shishido Joutsen<sup>3</sup>, Jouko Rikkinen<sup>3</sup>, Leonardo Teixeira Dall'Agnol<sup>2</sup>

<sup>1</sup>Universidade Federal do Maranhão, <sup>2</sup>Universidade Federal do Maranhão (UFMA), <sup>3</sup>Universidade de Helsinki

Lichens are symbiotic associations between fungi, algae, and/or cyanobacteria, giving rise to a vegetative structure called a lichen thallus, which exhibits different morphologies and can harbor communities of associated microorganisms that perform important functions for the maintenance of the lichen ecosystem. This study aims to characterize the structural and functional diversity of the microbial community associated with a lichen of the genus *Coenogonium* from the island of São Luís, MA, using metagenomic approaches. Specimen ITA05 was collected in the Itaipiracó Environmental Protection Area (APA) in São Luís, MA. After DNA extraction, the sample was sequenced on the AVITI platform and analyzed on the Kbase and Galaxy servers. Characterization using Kaiju indicated a predominance of bacteria (91%) in the sample, notably the phyla Pseudomonadota (52%), Actinomycetota (21%), and Acidobacteriota (11%). The remaining bacterial phyla accounted for 7% of the classified sequences, while eukaryotes represented 9% of the taxonomic composition. In comparison, GOTTCHA identified Pseudomonadota and Actinomycetota, but did not identify the Acidobacteriota phylum. Among the most abundant orders in Kaiju, Hypomicrobiales, Mycobacteriales, and Terriglobales stand out. antiSMASH identified 311 biosynthetic regions in bacteria and 254 in fungi, with a predominance of terpenes, NRPS, and NRPS-like regions. The regions were grouped using BiG-SCAPE, resulting in the identification of 129 biosynthetic gene clusters (BGCs) contained within 12 gene cluster families (GCFs) in bacteria and 72 BGCs in 11 GCFs in fungi, with terpenes, NRPS, and NRPS-LIKE remaining predominant. Furthermore, some classes were exclusive to bacteria, while fungi exhibited only classes shared with bacteria. Among the predominant classes, terpenes exhibited several regions associated with carotenoid biosynthesis, which play a role in photoprotection and antioxidant mechanisms, whereas NRPS products may be involved in the production of antimicrobial compounds and siderophores, which are involved in nutrient uptake. In fungi, BGCs of the T1PKS class may be related to the biosynthesis of metabolites such as usnic acid, which has antimicrobial potential, and anthranorine, which is associated with photoprotection. Community composition varies according to factors such as geographic location, environmental conditions, host and substrate, which contribute to the complexity of the lichen microbial consortium.

# **ABO SYSTEM AND COVID-19: EFFECT OF BLOOD TYPE DISTRIBUTION ON PANDEMIC OUTCOMES IN 123 COUNTRIES**

Késia de Carvalho Santos<sup>1</sup>, DANIEL JERÔNIMO DO NASCIMENTO<sup>1</sup>, LEONARDO CAPISTRANO FERREIRA<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

The variation in the distribution of human blood groups, governed by the A, B, and O alleles on chromosome 9, is remarkable among different ethnic and geographic groups. Recently, the literature has begun to investigate how this genetic polymorphism may modulate susceptibility to pathogens, including SARS-CoV-2. During the COVID-19 pandemic, these investigations intensified, and many studies observed that the risk of infection was strongly associated with the ABO locus. The present study examines the relationship between the prevalence of blood types and the epidemiological indicators of COVID-19 on a global scale, based on the hypothesis that blood type A confers greater vulnerability to the disease. The methodology was developed through the evaluation of associations using aggregated population data from 123 countries, obtained from Our World in Data, and genomic data from the 1000 Genomes Project, derived from 20 countries with available data for the analyses. Simple and multiple linear regression models were employed to isolate the effect of blood groups and their interactions with covariates. The percentage of blood type A was an independent predictor of susceptibility to infection, showing a positive association with an increase in cases per million, even after adjustment for multiple predictors ( $\beta = 1536.13$ ;  $p < 0.05$ ). It was also observed that populations with a higher standard deviation in the distribution of blood groups (lower genetic homogeneity) experienced more severe pandemic impacts, which exacerbated the number of infections ( $\beta = 1610.28$ ;  $p < 0.05$ ). The results indicated that secretor status acts as an effect modifier of ABO blood groups. The inclusion of this variable in the model altered the effect of blood groups on the analyzed outcomes, with statistically significant interactions ( $p < 0.05$ ). This effect was observed for blood type A in relation to the number of cases and for blood type O in relation to the case fatality rate. Taken together, the findings of this study deepen the understanding of how antigen expression in mucosal tissues and population genetic heterogeneity act as immunological determinants, providing a basis for more targeted public health interventions.

# Single-Sample Network Reconstruction and Differential Centrality Analysis Across Braak Stages in Alzheimer's Disease

Luísa Tannus Lima<sup>1</sup>, Júlia A. de Amorim<sup>2</sup>, Diego Marques Coelho<sup>2</sup>, Marcos Romualdo Costa<sup>3</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN, <sup>3</sup>Multiuser Bioinformatics Center, UFRN

Alzheimer's disease (AD) progression involves complex transcriptional and structural reorganization within vulnerable brain regions, driven by neuroinflammation, synaptic dysfunction, and neuronal loss. Differential expression analysis identifies genes with altered transcript abundance across disease stages, whereas network-based approaches reveal genes undergoing structural shifts in importance within molecular interactomes. Here, we integrated differential gene expression and differential network centrality analyses to characterize transcriptomic and topological changes across AD progression stages in Brodmann Area 36 (BM36), a temporal region affected early in AD, using Mount Sinai Brain Bank (MSBB) data. Samples were stratified by Braak stages into Low (Braak 0-2), Mid (Braak 3-4) and High (Braak 5-6) groups. To capture sample-specific network dynamics, individual gene interaction networks were reconstructed using the LIONESS framework, selecting 80 samples per group. Differential centrality analysis quantified shifts in gene network prominence across Braak stages, while differential expression analysis was performed across genes using DESeq2. Functional annotation of both analyses was conducted via Gene Set Enrichment Analysis (GSEA) targeting Gene Ontology Biological Process (BP) and Cellular Component (CC) domains. Genes upregulated across Braak stages showed significant enrichment in inflammatory and immune-related processes, reflecting progressive neuroinflammation. Conversely, downregulated genes were enriched in synaptic transmission and neuronal signaling components. Network analysis revealed a complementary topological dynamic, in which genes displaying increased differential centrality across stages were selectively enriched in excitatory synaptic membranes and vesicle machinery. Thus, while global synaptic transcripts showed reduced expression, a subset became increasingly central within the disease network. These findings suggest that transcriptional suppression of synaptic programs and topological reorganization of synaptic network architecture occur simultaneously during AD progression. Increased network centrality does not imply elevated expression, but highlights critical hubs or vulnerability points in a disrupted system. Integrating DESeq2 with LIONESS-derived differential centrality and GSEA offers a comprehensive view of AD pathology and may help prioritizing key synaptic hubs as potential biomarkers or therapeutic targets.

# Exploring antimicrobial resistance patterns across *Enterococcus* lineages.

Felipe Assunção Félix Ramos<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

*Enterococcus* species are ubiquitous commensals and major opportunistic pathogens notorious for acquiring multidrug resistance. Understanding the distribution of the antimicrobial resistome across different evolutionary lineages is crucial for mitigating nosocomial infections. In this study, we conducted a large-scale pan-genomic analysis of 13,389 publicly available *Enterococcus* genomes from the NCBI to map the evolutionary landscape of antimicrobial resistance (AMR). We constructed binary presence/absence matrices using genetic features annotated against the Comprehensive Antibiotic Resistance Database (CARD), a global reference catalog linking genomic data to molecular AMR phenotypes. To evaluate the relationship between evolutionary ancestry and resistance profiles, we employed advanced visual analytics, mapping established phylogenetic clades against phenotypic hierarchical clustering utilizing Ward's method. The analysis of the AMR profiles revealed three major, distinct resistome clusters. The first was predominantly composed of Clade I strains (e.g., *E. faecalis*), exhibiting a conserved, lineage-specific resistance signature. The second was heavily dominated by Clade II strains (e.g., *E. faecium*), reflecting a massive accumulation of resistance determinants characteristic of this highly adapted nosocomial group. Interestingly, the third cluster comprised a highly heterogeneous mix of diverse phylogenetic clades. This visual convergence of genetically distant species sharing identical resistance modules suggests that while classical phylogeny strongly dictates the baseline resistome in primary lineages, a distinct subset of the enterococcal pangenome is characterized by highly fluid resistance profiles. These mixed profiles transcend classical species barriers, underscoring the complex dynamics of AMR acquisition and shared environmental pressures.

# **XR-Score approach for chemical stress transcriptome meta-analysis in *Phthorimaea absoluta***

Vitor Quintela Sousa<sup>1</sup>

<sup>1</sup>Universidade Federal Rural de Pernambuco

The tomato pinworm, *Phthorimaea absoluta*, represents one of the major agricultural pests of tomato crops due to its high capacity to develop resistance to insecticides. This study performed a transcriptomic re-analysis of public RNA-Seq data to identify differentially expressed genes and conserved molecular pathways associated with chemical stress response. Public bioprojects from NCBI SRA (PRJNA1359366, PRJNA749726, PRJNA1270399) comprising 26 RNA-Seq libraries of larvae exposed to abamectin, emamectin benzoate, and spinosad were selected. Raw reads were processed with SRA Toolkit and Trim Galore!, and transcript abundances quantified using Salmon against the reference transcriptome. Using Python (Pandas/NumPy), Welch's t-test was conducted, and a Xenobiotic Response Score (Net XR-Score) adapted from Bono (2021) was calculated as an index of cross-treatment regulatory consistency (XR-Score = count<sub>up</sub> - count<sub>down</sub>) for active transcripts (log<sub>2</sub> FC | ≥ 1.0, P < 0.05). Functional annotation was performed using DIAMOND blastx against Swiss-Prot. Global partitioning of the transcriptome (64,591 transcripts) revealed high homeostatic stability, with 96.49% neutral (62,322), 1.24% repressed (803), and 2.27% induced (1,466) genes. Strict consistency filters identified core biomarkers systematically activated across all chemical classes: members of the cytochrome P450 family (e.g., CYP6a14, CYP4g15) and oxidative stress-responsive peroxidases maintained maximum induction (XR-Score = +3). Furthermore, an energetic trade-off characterized by the repression of digestive carboxylesterases and cuticular proteins was mapped (Net XR-Score = -3). In this study, we applied a meta-analysis framework to *P. absoluta*, an insect species whose molecular resources in public repositories are still expanding and remain limited, particularly regarding strictly paired control and exposed conditions from identical populations or experimental events. While these database constraints represent a methodological limitation, the implemented Net XR-Score successfully identified core genes consistently regulated across multiple insecticide treatments, shedding light on the core detoxification machinery and key metabolic pathways involved in pesticide tolerance.

# Use of BERT models in RNA-Seq analysis

gabrielle de vasconcelos borja<sup>1</sup>

<sup>1</sup>ufrn

RNA sequencing enables the characterization of gene expression patterns under different biological conditions, but its high dimensionality and the complex relationships among genes creates challenges for conventional statistical and machine-learning approaches. In this context, models based on the Transformer architecture, particularly BERT-derived models, have emerged as promising alternatives for representing and analyzing biological data. This ongoing research aims to investigate the application of BERT-based models to RNA-Seq analysis and bioinformatics predictions. The proposed workflow includes the preprocessing, normalization and study of biological tokenization strategies for bulk RNA-Seq datasets; the adaptation or fine-tuning of a pretrained Transformer model for a specific task, such as sample classification. Model explicability and interpretability strategies will also be explored to investigate whether the learned representations and attention patterns provide biologically meaningful insights. At its current stage, the project focuses on reviewing suitable architectures, selecting datasets, and evaluating the pipeline. The expected contribution is to assess the potential and limitations of BERT-based models for transcriptomic data analysis while developing an interdisciplinary framework connecting bioinformatics, molecular biology, and deep learning.

# Data Mining and Phylogenetic Analysis of the RAC1 Protein in Vertebrates

Ana Vivia dos Santos Chacon da Silva<sup>1</sup>, Rodrigo Olímpio Orvate Arruda<sup>1</sup>, João Paulo Matos Santos Lima<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Phylogenetic reconstruction is a fundamental step in understanding evolutionary relationships and patterns of amino acid conservation. It is essential for understanding the consequences and effects of mutations and polymorphisms on protein function and for predicting drug binding. The RAC1 protein, a small GTPase of the Rho family that plays an essential role in regulating the cellular cytoskeleton, motility, and cell division, is strongly associated with developmental processes and various human diseases, and is considered an important therapeutic target. However, there is a lack of comprehensive knowledge regarding the impacts of polymorphisms on the function of this protein and on drug binding in related complex diseases. Therefore, the objective of this study is to compare RAC1 sequences from different species within the Vertebrata taxon, relate the observed polymorphisms to structural modifications and the function of this protein, and compare them to those reported in different human populations. To this end, we first conducted a search for sequences homologous to human RAC1 in the NCBI and UniProt databases. The curation and filtering of the raw data were performed using the OASIS tool to remove redundant or uncharacterized sequences. Next, phylogenetic trees were reconstructed using the TaxOnTree program. The observed amino acid changes were compared with polymorphisms in the dbSNP database, and their effects were evaluated based on Alaphamissense predictions. The results demonstrated that the topologies of the reconstructed trees indicate a high degree of conservation in the sequence and structure of RAC1 throughout vertebrate evolution, implying strong mutational constraints on its critical functional regions, such as the Switch I and Switch II domains. Because RAC1 is a protein vital for cellular function, abrupt variations in its sequence tend to be eliminated by purifying selection; however, some of the observed polymorphisms are predicted to have a deleterious effect. These initial results indicate that some of the observed polymorphisms, although predicted to have no effect on function, may impact the binding of drugs that target RAC1 as their primary target.

# **Ricinus communis Defensins Binding to the Bacterial Outer Membrane: First Stages of Antimicrobial Action**

Giovanna Santos de Lima<sup>1</sup>, Madson Allan de Luna-Aragão<sup>2</sup>, Ana Maria Benko-Iseppon<sup>1</sup>, Denys Ewerton da Silva Santos<sup>1</sup>

<sup>1</sup>Universidade Federal de Pernambuco, <sup>2</sup>Universidade Federal de Minas Gerais

Plant defensins are a class of antimicrobial peptides (AMPs) that play a central role in the plant innate immune system, acting against infections. These cysteine-rich peptides are characterized by a predominantly cationic surface, with broad antimicrobial mechanisms, including membrane disruption and enzyme inhibition. The most conserved structural feature of plant defensins is the  $\gamma$ -core region, a disulfide-stabilized  $\beta$ -hairpin motif preserved across the defensin family despite extensive sequence divergence. In this context, our work aims to investigate the antimicrobial potential of novel *Ricinus communis* defensin (RcDef) models identified through multi-omics data mining in previous studies by our research group, using molecular dynamics simulations to evaluate their interactions with a *Pseudomonas aeruginosa* outer membrane model. To do so, classical molecular dynamics simulations were performed using GROMACS v2023.3 with the GROMOS 54A7 force field. The systems simulated included a homodimeric RcDef model, a configuration commonly associated with enhanced defensin function, and three  $\gamma$ -core peptides: the native  $\gamma$ -core, a consensus sequence derived from other RcDefs, and a mutated variant predicted to exhibit increased antimicrobial activity. The AMPs were initially positioned 2.5 nm above a membrane model composed of 128 penta-acylated lipid A molecules symmetrically distributed in an 8 × 8 arrangement in each leaflet. Following the initial equilibration, production simulations were conducted in the NPT ensemble for 200 ns with one RcDef copy and extended for an additional 300 ns after introducing a second copy, in order to assess concentration effects. GROMACS was employed to evaluate peptide structural stability through radius of gyration, hydrogen-bond number, and conformational sampling based on Ramachandran plots. SuAVE was used to assess the influence of RcDef on membrane properties, including thickness, curvature, and area per lipid. Although only limited membrane perturbation was observed within the 500 ns timescale, the simulations revealed potential insertion mechanisms, including penetration of hydrophobic amino acid side chains, particularly phenylalanine and leucine, and displacement of counterions from the membrane surface. Overall, these findings provide insights into RcDef potential antimicrobial mechanisms and highlight molecular dynamics as a valuable tool for unraveling defensin-membrane interactions.

# **From ecological restoration to bioprospecting: RNA-seq data reuse reveals antimicrobial peptide candidates and substrate-responsive genes in *Stryphnodendron pulcherrimum***

Mariana Costa Dias<sup>1</sup>, Eduarda Vianna Vilhena<sup>2</sup>, Marx Oliveira de Lima<sup>3</sup>, Diego Sotero de Barros Pinangé<sup>4</sup>

<sup>1</sup>Universidade Federal do Amazonas, <sup>2</sup>UFAM, <sup>3</sup>EMBRAPA Amazônia Ocidental, <sup>4</sup>UFAM/ITV

Antimicrobial peptides (AMPs) are bioactive molecules synthesized by a wide variety of living organisms that participate in the innate immune system and contribute to the defense against invading pathogens. Some bacteria and fungi have become resistant to several conventional drugs, therefore, AMPs have emerged as a promising alternative due to their diverse mechanisms of action and their lower propensity to induce resistance. In plants, AMP production is influenced by biotic and abiotic factors that can induce the expression of defense-related genes, including those encoding AMPs. In this context, *Stryphnodendron pulcherrimum* (Willd.) Hochr. (Fabaceae), exhibits proven antibacterial activity, which may be associated with AMP production. Accordingly, this study aimed to perform the *in silico* bioprospecting of AMPs in the transcriptome of *S. pulcherrimum*, as well as to evaluate the effect of different substrates on the expression of AMP-related genes. RNA-seq data from *S. pulcherrimum* were obtained from the Sequence Read Archive (SRA-NCBI). The reads were generated from leaf tissue samples of individuals grown on two substrate from Serra dos Carajás, Pará, Brazil: canga (a natural iron-rich substrate) and yellow waste substrate (mining residue). After quality control and adapter removal using FastQC and Trimmomatic, respectively, the reference transcriptome was aligned against natural AMPs from the APD6 database, followed by differential expression analysis using edgeR. The *in silico* analysis identified 2.326 transcripts as AMP candidates, corresponding to 365 genes. Differential expression of this set indicated that substrate was the main factor associated with variation in gene expression. A total of 42 genes and 62 isoforms were differentially expressed, including 22 upregulated and 20 downregulated genes in individuals grown on yellow waste substrate compared with those grown on canga. Among the upregulated genes, ILTI stood out as a protease inhibitor with previously reported antifungal activity, capable of inhibiting fungal growth. These data reveal a diversity of defense-related genes associated with AMPs in *S. pulcherrimum*, whose differential expression varied between the two substrates evaluated. The identification of candidates with known antimicrobial activities points to their biotechnological potential, enabling future studies involving structural and functional approaches.

# **Proteome-wide discovery of 'hidden' allergens in Pacific white shrimp (*Penaeus vannamei*): In silico characterization of uncharacterized candidates with high cross-reactivity potential.**

Daniel Pacheco Penha de Oliveira<sup>1</sup>, Maria Karolayne da Silva<sup>1</sup>, Alana de Medeiros Dantas<sup>1</sup>, Gabriel Christian de Farias Moraes<sup>1</sup>, Gabriel Vinícius Rolim Silva<sup>1</sup>, Guilherme Bastos Alves<sup>1</sup>, Jonas Galileu Ferreira de Aquino<sup>1</sup>, Letícia Maria Azevedo Martins<sup>1</sup>, Renata Maria de Araújo<sup>1</sup>, Zayra Jasminne Dias de Lima<sup>1</sup>, Jonas Ivan Nobre Oliveira<sup>1</sup>, Umberto Laino Fulco<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

*Litopenaeus vannamei* (Pacific white shrimp) is the most commercially important crustacean species worldwide and a leading cause of IgE-mediated food allergy. Although four canonical allergens (Lit v 1–4) have been characterized, the allergenic potential of most proteins encoded in its reference proteome remains unexplored. This study aimed to perform a proteome-wide in silico allergenicity screening of *L. vannamei* and to characterize the structural, functional, and immunogenic properties of novel uncharacterized candidate allergens. The complete reference proteome (NCBI RefSeq assembly GCF\_042767895.1, 43,631 annotated sequences) was analyzed through a multi-tier computational pipeline, combining quality filtering, similarity screening, and independent allergenicity prediction approaches. Priority uncharacterized proteins underwent structural characterization, protein-protein interaction network analysis, T-cell and B-cell epitope mapping, and global HLA population coverage analysis. After quality filtering, 35,993 sequences were retained, and similarity screening identified 710 proteins meeting allergen-similarity criteria, of which 705 proceeded to downstream analysis. Of these, 600 (85.1%) showed computational evidence of allergenicity, including 283 with strong and 317 with weak evidence, whereas 105 showed no evidence. An independent prediction approach classified 158 proteins as probable allergens. Concordance between the two prediction strategies prioritized five uncharacterized proteins, which showed  $\geq 35\%$  sequence identity to established mite, cockroach, and pollen allergens. Epitope mapping of the two highest-priority candidates (XP\_069987405.1 and XP\_069969538.1) identified, respectively, eight CTL, two HTL, and three LBL epitopes for the first, and 52 CTL, five HTL, and 16 LBL epitopes for the second. Estimated global population coverages were 67.99% (CTL) and 10.73% (HTL) for XP\_069987405.1, and 98.55% (CTL) and 53.99% (HTL) for XP\_069969538.1, considering HLA-DRB alleles for HTL coverage. This study expands the predicted allergen repertoire of *L. vannamei* beyond Lit v 1–4 and prioritizes five previously uncharacterized candidate allergens as promising targets for experimental validation of IgE binding, cross-reactivity, and clinical relevance.

# Automated Laterality Detection in Mammography Reports Using NLP and Classical Machine Learning Models

Rayssa Beatriz Ribeiro Cavalcante<sup>1</sup>, Pedro Henrique Álvares Róbias<sup>1</sup>, César Rennó-Costa<sup>2</sup>, André Salles Cunha Peres<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

The structuring of healthcare data related to mammography and breast cancer has gained increasing relevance, driven by the need for high-quality datasets to train artificial intelligence models. While Large Language Models (LLMs) are currently extensively used for clinical report structuring, they present significant operational and technical challenges. LLMs incur high computational costs, demand massive datasets for fine-tuning, and are inherently prone to hallucinations. Although smaller LLMs can be deployed locally to prevent sensitive patient data leakage, they frequently misinterpret specific mammography jargon and suffer from attention degradation with large prompts, for instance, in few-shot strategies. Conversely, mammography reports are typically short, direct, and structurally consistent, which heavily favors the use of classical machine learning and natural language processing (NLP) techniques. To address this, the present study developed and compared two low-cost, locally executable computational strategies to identify the laterality of high-risk findings in Portuguese mammography reports. The first strategy consisted of a rule-based NLP pipeline utilizing the NLTK library and a custom clinical dictionary. The second approach employed contextual embedding vectors generated by the BioBERTpt model—fine-tuned on real patient reports from a partner oncology center—coupled with classical supervised machine learning models (Support Vector Machine, K-Nearest Neighbors, and Random Forest). A ground-truth dataset of 500 structured reports (obtained from Liga Norte Riograndense Contra o Câncer) was built, and both strategies were evaluated using a k-fold cross-validation approach. Among the classical models, SVM yielded the highest accuracy at 88.61%, followed by Random Forest (86.60%) and KNN (85.20%). Under the same evaluation framework, the rule-based NLP pipeline achieved a competitive accuracy of 83.96%. These results demonstrate that classical computational models are highly viable, secure, and resource-efficient alternatives for automating data extraction from standardized clinical texts, proving to be an excellent and highly effective strategy for the automated structuring of medical reports.

# **CTNNB1 mutational status shapes distinct metabolic and inflammatory transcriptional programs in hepatocellular carcinoma**

Raylam Wendel da Silva<sup>1</sup>, Epitácio Dantas de Farias Filho<sup>1</sup>, Beatriz Stransky<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Hepatocellular carcinoma (HCC) exhibits substantial molecular heterogeneity, with subtypes associated with distinct transcriptomic and functional programs. Among them, tumors harboring CTNNB1 mutations have been associated with activation of the Wnt/ $\beta$ -catenin pathway and metabolic features characteristic of more differentiated hepatocellular tumors. In this study, we investigated the transcriptomic and functional profiles associated with CTNNB1 mutational status in the TCGA-LIHC cohort, aiming to understand how these alterations relate to the functional heterogeneity of HCC. Gene expression and somatic mutation data from 377 HCC patients were integrated to compare the CTNNB1-mutated and non-mutated tumors. Differential expression analysis was performed using DESeq2, followed by pathway enrichment and gene set enrichment analyses to identify biological programs associated with each molecular group. The analyses revealed a distinct pattern between the groups, with CTNNB1 mutated tumors showing increased expression of Wnt/ $\beta$ -catenin-related genes, including AXIN2, SP5, TCF7, LEF1, TBX3, GLUL and DKK1. Functional analyses further revealed enrichment of programs related to bile acid and fatty acid metabolism in mutated tumors, whereas non-mutated tumors exhibited enrichment of inflammatory processes, hypoxia, and tissue remodeling. These metabolic programs are particularly relevant in HCC because they are closely linked to hepatocyte-specific functions, suggesting that CTNNB1-mutated tumors retain a distinct liver-related metabolic transcriptional profile. Overall, these findings indicate that CTNNB1 mutational status is associated with distinct functional profiles in HCC. This molecular stratification contributes to the characterization of HCC heterogeneity and highlights the value of integrative transcriptomic and genomic analyses for defining biologically distinct tumor subgroups.

# Sauron: A Versatile Tool for Scalable Residue Interaction Networks (RINs) Analysis

João Paulo Matos Santos Lima<sup>1</sup>, Arthur Henrique Dantas de Azevedo<sup>1</sup>, Laura Shimohara Bradaschia<sup>2</sup>, Gustavo Antônio de Souza<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Residue Interaction Networks (RINs) are a powerful tool in bioinformatics analysis, given that they represent 3D protein structures as graph networks in which amino acid residues act as nodes and non-covalent interactions serve as edges. Consequently, RINs are essential for identifying catalytic sites, assessing mutation effects, evaluating structural stability, and analyzing molecular dynamics simulations. Thus, we introduce Sauron, a versatile RINs calculator that streamlines network generation from multiple input types, including PDB and mmCIF files or direct database fetching via RCSB PDB and AlphaFold UniProt IDs. Sauron boasts four distinct calculation methodologies—Standard distance/angle cutoffs, Voronoi tessellation, ProDy InSty atomic-level interaction detector, and RinPy elastic network models—allowing researchers to customize analytical sensitivity to their specific workflow. The tool identifies, maps, and displays key intramolecular interactions, including hydrogen bonds, salt bridges, disulfide bonds, van der Waals contacts,  $\pi$ -stacking and cation- $\pi$  interactions. Additionally, users can explore network topology interactively through an advanced 3D web dashboard featuring a force-directed graph synchronized with a PDBe Molstar structural viewer, alongside a primary sequence viewer and degree distribution plots. Finally, Sauron standardizes downstream analysis by outputting comprehensive node and edge data files, full topological metrics (e.g., Degree, Betweenness, and Eigenvector centrality), and ranked summary tables. Therefore, by combining calculation methods versatility, rich interactive visualization, and structured output files, Sauron provides an end-to-end platform for scalable topological profiling in bioinformatics workflows.

# Integration of GWAS and gene-based analysis to investigate genetic signals associated with breast cancer

João Luiz de Lemos Padilha Pitta<sup>1</sup>, Beatriz Souza Toscano de Melo<sup>1</sup>, Matheus Azevedo Bomfim<sup>2</sup>, Laryssa Bandeira de Melo Silva<sup>2</sup>, Luydson Richardson Silva Vasconcelos<sup>3</sup>, Túlio de Lima Campos<sup>2</sup>

<sup>1</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, <sup>2</sup>Instituto Aggeu Magalhães - FIOCRUZ, <sup>3</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, Fundação Oswaldo Cruz, Fiocruz, Recife, Pernambuco, Brazil.

Genome-wide association studies (GWAS) identify genetic variants associated with complex phenotypes, whereas gene-based approaches aggregate evidence across multiple variants and may improve signal interpretation. This study integrated variant- and gene-level analyses to investigate breast cancer-associated signals and assess the potential impact of using external controls. Exome data from 459 breast cancer cases (dbGaP-phs000822.v1.p1) were compared with 459 public controls from the 1000 Genomes Project, matched for genetic ancestry using PCAmatchR. Variant-level association testing was performed using Firth logistic regression, followed by functional annotation with the Variant Effect Predictor (VEP). Gene-based analysis was conducted using MAGMA, with independent breast cancer GWAS data from the Million Veteran Program (MVP) used as a reference for pipeline evaluation. In the MVP cohort, *FGFR2* ( $p=1.96\times10^{-9}$ ) and *TOX3* ( $p=7.09\times10^{-8}$ ) exceeded the Bonferroni-corrected threshold ( $p<2.6\times10^{-6}$ ; 18,996 genes tested), reproducing established breast cancer susceptibility loci and supporting technical validation of the gene-level pipeline. In the study cohort ( $p<4.1\times10^{-6}$ ; 12,157 genes tested), the strongest signals involved *ALPPL2*, olfactory receptor genes, immunoglobulin regions, and *ARL17B/LRRC37A*, mainly located in paralogous, multicopy, or structurally complex regions prone to mapping errors in exome data. A sensitivity analysis combining ENCODE blacklist filtering with the exclusion of known problematic gene families and structurally complex regions removed the initial top signals; however, new top-ranked genes again emerged in paralogous or multicopy regions. Only 23.93% of variants were shared between cases and controls, and genomic inflation persisted after ancestry matching ( $\lambda_{GC}=1.24$ ), indicating residual confounding consistent with batch effects arising from technical differences in sequencing, coverage, and variant calling between datasets. The combined GWAS-MAGMA framework recovered established breast cancer susceptibility signals while revealing batch effects and technical artifacts that could otherwise be misinterpreted as biological associations. These findings show that ancestry matching alone may not fully overcome heterogeneity between independently generated genomic datasets and highlight the value of integrating variant-level, gene-level, and quality-control evidence for robust association analysis.

# Comparative Evolutionary Analyses of GATM and SLC6A8 and Their Functional Integration into Creatine Metabolism

Pierina Sol Elzegbe<sup>1</sup>, Ana Vivian dos Santos Chacon da Silva<sup>2</sup>, João Paulo Matos Lima Silva<sup>2</sup>

<sup>1</sup>BioinfoMetrópole, Instituto Metrópole Digital (IMD), Universidade Federal do Rio Grande do Norte (UFRN), <sup>2</sup>Universidade Federal do Rio Grande do Norte

Creatine is an essential component of energy metabolism, as it participates in the creatine/phosphocreatine system, which helps maintain ATP availability in tissues with high energy demands, such as muscle, heart, and the central nervous system. Within this system, the GATM gene encodes glycine amidinotransferase, the enzyme that catalyzes the first step of creatine biosynthesis, whereas SLC6A8 encodes the creatine transporter responsible for creatine uptake into cells. Given the functional importance of both genes, this study aimed to characterize their evolutionary distribution and conservation to provide insights into their origin and evolution across taxonomic groups and to relate these patterns to their roles in creatine metabolism. For the analysis, the TaxOnTree tool was used to retrieve proteins homologous to GATM and SLC6A8 via BLAST searches and to generate phylogenetic trees from sequence alignments. Different identity thresholds (50% and 75%) and maximum numbers of retrieved sequences (200 and 800) were evaluated. Taxonomic distributions were examined at different taxonomic levels to characterize the distribution of homologs and identify phylogenetic clustering patterns based on sequence identity. The analyses showed that, for both genes, lowering the identity threshold increased the taxonomic diversity of the retrieved sequences, whereas higher thresholds resulted in a progressive concentration of sequences within Vertebrata. This distribution, together with the conservation observed in the phylogenetic trees, allowed us to relate phylogenetic patterns to the functions of GATM and SLC6A8 in the creatine system, which is involved in the synthesis and transport of creatine to tissues with high energy demands. Overall, these results provide a basis for reconstructing the evolutionary history of GATM and SLC6A8 and integrating this information with the functions of these proteins in tissues such as the kidney, liver, muscle, and central nervous system, contributing to a better understanding of the evolution of the creatine synthesis and transport system.

# **Uma Metodologia de Calibração da Confiança no Modelo Human-in-the-Loop para IA Clínica Usando Códigos de Justificativa Padronizados e Limiares Adaptativos de Quebra**

Luana Ferraz Alvarenga<sup>1</sup>, Andre Moreira<sup>1</sup>, Caio Justino<sup>1</sup>, Andre Peres<sup>1</sup>, Renan Moiolli<sup>1</sup>,  
César Rennó-Costa<sup>1</sup>

<sup>1</sup>UFRN

As artificial intelligence (AI) becomes deeply integrated into high-stakes clinical decision-making, while its opacity poses a profound regulatory challenge, leaving clinical practice polarized between clinician alert fatigue and dangerous automation bias. To resolve this impasse, this work explores the isomorphism between financial risk models and biomedical decision support systems, proposing a foundational roadmap that translates mature financial transparency frameworks across the US, EU, Brazil, and China into healthcare. The proposed architecture operates across three interconnected levels: at the macro level, it adapts multi-jurisdictional financial transparency models to establish clinical governance and standardized reason codes. At the meso level, it optimizes the user experience through Cognitive Forcing Functions and Dynamic Authority Transitions (alternating between Human-Leader and AI-Leader states), with transitions critically governed by an Adaptive Breach Threshold Indicator that modulates real-time sensitivity based on patient acuity, algorithmic confidence, and clinician workload. At the micro level, it introduces a computational translation layer that utilizes clustering, thresholding, and pruning algorithms to compress high-dimensional statistical features (such as SHAP values) into semantically rigorous medical concepts, anchored in established ontologies such as SNOMED CT. By comparatively analyzing regulatory frameworks, our findings demonstrate the feasibility of transposing mature financial governance standards into clinical AI. Finally, our work resolves medico-legal ambiguities surrounding civil liability, outlines Basel-style Clinical Use Tests for national health data infrastructures, and achieves essential sociotechnical integration - preserving meaningful human oversight without imposing prohibitive operational burdens on medical practitioners.

# CHARACTERIZATION OF THE MICROBIOME OF FILAMENTOUS LICHENS OF THE GENUS *Coenogonium* FROM THE AMAZON REGION OF MARANHÃO

Weneson Victor Diniz Sarges<sup>1</sup>, Emanuelle Costa Ramos<sup>2</sup>, Anna Letícia Silva da Costa<sup>2</sup>, José Isaías Pimentel Barros<sup>2</sup>, Nathalia Silva Batista<sup>1</sup>, Marcos Davi de Sousa Campos<sup>1</sup>, Hivana Patricia Melo Barbosa Dall'Agnol<sup>2</sup>, Tânia Keiko Shishido Joutsen<sup>3</sup>, Jouko Rikkinen<sup>3</sup>, Leonardo Teixeira Dall'Agnol<sup>2</sup>

<sup>1</sup>Universidade Federal do Maranhão, <sup>2</sup>Universidade Federal do Maranhão (UFMA), <sup>3</sup>Universidade de Helsinki

The genus *Coenogonium* is a genus of chlorolichens with a crustose to filamentous morphology, consisting of an Ascomycota mycobiont and an algal photobiont of the genus *Trentepohlia*. Of the 130 species currently known, 52 occur in Brazil. Although taxonomic studies exist, there are no records of research on the microbial community of these lichens. Thus, this study aims to characterize the bacterial structural diversity of *Coenogonium* lichens. Samples were collected in four municipalities in the Maranhão Amazon: São Luís, Morros, Anajatuba, and São Vicente Férrer. DNA was extracted, and the 16S rRNA gene was sequenced using the Illumina MiSeq platform. Data quality was assessed using FastQC v. 0.12.0, and the data were analyzed using QIIME2 v. 2025.10, which removed low-quality sequences (Phred  $\geq 25$ ), applied filters, and calculated diversity metrics. Taxonomy was assigned using the SILVA database, with artifacts exported to RStudio for data plotting. A total of 2,685,181 reads were obtained and grouped into 2,745 ASVs. The Shannon index showed that ANA15 exhibited the highest diversity and MO2 the lowest. As for the Simpson index, the highest values were observed in ANA15 and the lowest in ITA05, indicating high diversity among the samples and high evenness. Beta diversity, as measured by the Bray-Curtis index, showed that ANA2 and ANA15 have similar bacterial compositions, with CNC1 being close to them, while ITA05 and IFMA2 are dissimilar to them and have a composition similar to each other. MO2 was found to be distinct from all other samples, indicating a distinct taxonomic group. The taxonomic composition was dominated by the phyla Proteobacteria ( $41.56 \pm 14.35$ ), Actinobacteria ( $25.33 \pm 8.81$ ), Acidobacteria ( $13.77 \pm 5.66$ ), Verrucomicrobia ( $4.78 \pm 5.00$ ), and Chloroflexi ( $3.69 \pm 3.83$ ). Alphaproteobacteria ( $39.40 \pm 14.47$ ), Actinobacteria ( $18.00 \pm 6.61$ ), and Acidobacteria ( $11.28 \pm 7.97$ ) were the most abundant classes. Among the genera, *Lichenibacterium* was dominant ( $9.71 \pm 8.84$ ). This genus was first described based on bacteria associated with lichens, with some of its representatives being endophytic lichens. These results indicate a diverse bacterial community and highlight the structural richness of lichens in Maranhão, a region where research remains scarce. This opens new avenues for studies on lichen microbiomes and how they can be used as tools to assess the physiology of symbiosis and environmental quality.

# **An automated genome-mining pipeline for RiPP discovery identifies four biosynthetic systems in an Amazonian *Bacillus* sp.**

Marx Oliveira de Lima<sup>1</sup>, Jaqueline Ferreira Gomes<sup>2</sup>, Anderson Nogueira Barbosa<sup>2</sup>, Sabrina Sinara Portela de Sousa<sup>2</sup>, Sérgio Dantas de Oliveira Júnior<sup>2</sup>, Douglas de Moraes Couceiro<sup>2</sup>, Lucas de Jesus Patrício<sup>2</sup>, Gilvan Ferreira da Silva<sup>2</sup>

<sup>1</sup>EMBRAPA Amazônia Ocidental, <sup>2</sup>EMBRAPA - Amazônia Ocidental

Ribosomally synthesized and post-translationally modified peptides (RiPPs) comprise a diverse group of bioactive natural products whose genomic discovery requires integration of biosynthetic context, precursor recognition and sequence evidence. This work aimed to develop an automated and reproducible Snakemake pipeline for serial analysis of microbial genomes, designed to identify RiPP-associated biosynthetic gene clusters (BGCs) and prioritize putative precursor peptides for downstream characterization. The workflow integrates antiSMASH for BGC detection, RRE-Finder for RiPP recognition elements, Prodigal and orfipy for recovery of small ORF-derived peptides, similarity searches against curated RiPP references, Macrel and ampir for antimicrobial prediction, and AMRFinderPlus, CARD/RGI and VFDB for preliminary biosafety screening. Application of the pipeline to BAC-JFG180, a *Bacillus* sp. isolated from soil in Novo Airão, Amazonas, Brazil, identified four RiPP-associated biosynthetic regions, including RiPP-like, class I and class III lanthipeptide, and azole-containing RiPP systems. A broad nonredundant catalog of 98,176 peptides (20–250 aa) was generated as a discovery space and integrated with BGC context, RRE support and RiPP-reference similarity. Candidates were classified into six decreasing evidence zones (P1–P6), ranging from characterized-reference similarity plus compatible biosynthetic context to exploratory or currently unsupported sequences. No P1 candidates were detected, whereas 751, 2,282, 187, 6,538, and 88,418 sequences were assigned to P2–P6, respectively. Overall, 3,116 sequences showed RiPP-context support, including 788 with combined RiPP-BGC and RRE evidence, while 88 showed significant similarity to RiPP references. Integration with antimicrobial predictions retained 942 RiPP-prioritized candidates, including 54 without a hemolysis warning. Genome-wide biosafety screening detected 17 AMRFinderPlus loci, 14 strict CARD/RGI hits and 15 VFDB core loci, indicating the need for manual review. These results demonstrate the operational capacity of the automated workflow to screen microbial genomes serially and reduce large genomic search spaces to traceable RiPP biosynthetic systems and ranked putative precursor candidates for structural and experimental characterization.

# **Multi-objective evolutionary gene selection reveals compact molecular signatures for medulloblastoma classification**

Vinícius Pereira de Pontes<sup>1</sup>, Raylam Wendel da Silva<sup>1</sup>, Matheus Dalmolin<sup>2</sup>, Raquel de Melo Barbosa<sup>3</sup>, Marcelo A. C. Fernandes<sup>4</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil, <sup>3</sup>Universidade de Granada, <sup>4</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil Department of Computer Engineering and Automation, UFRN, Natal, Brazil

Medulloblastoma (MB) is a molecularly heterogeneous pediatric brain tumor classified into four major molecular subgroups: WNT, SHH, Group 3, and Group 4, which exhibit distinct biological characteristics and clinical outcomes. Accurate subgroup classification is therefore essential for molecular characterization and risk stratification. Although machine learning approaches can achieve high predictive performance using gene expression data, their dependence on large numbers of genes may limit model interpretability and the identification of concise molecular signatures. In this study, we developed an explainable-by-design framework integrating the Non-dominated Sorting Genetic Algorithm II (NSGA-II) with XGBoost to simultaneously optimize molecular subgroup classification and gene set sparsity. Gene expression data from the public GSE85217 cohort were analyzed, comprising 612 samples after preprocessing. Candidate gene spaces containing 1,000 and 2,000 genes were initially defined using univariate feature filtering, followed by multi-objective evolutionary optimization to identify compact gene subsets that preserved classification performance. The proposed approach identified highly compact signatures containing 7 and 14 genes, respectively, while preserving strong classification performance across the four molecular subgroups. SHH and WNT tumors were consistently distinguished, whereas the main classification overlap occurred between Groups 3 and 4. The selected signatures included genes involved in neurodevelopmental regulation, metabolism, and cellular signaling, with PDLIM4 and FIGNL2 recurring across both configurations. Overall, these findings demonstrate that multi-objective evolutionary gene selection can substantially reduce transcriptomic dimensionality while preserving accurate medulloblastoma subgroup classification. This explainable-by-design strategy provides compact and biologically interpretable molecular signatures with potential applications in biomarker discovery and translational neuro-oncology.

# Region-Specific regulon in microglial cell profile in Parkinson's Disease

Marcella Vitória Belém Souza<sup>1</sup>, Gustavo Barra Matos<sup>1</sup>, Andre Luis Fonseca Faustino<sup>2</sup>, Gilderlanio Santana de Araújo<sup>1</sup>

<sup>1</sup>Universidade Federal do Pará, <sup>2</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Instituto Metrópole Digital (IMD); Federal University of Rio Grande do Norte (UFRN)

Parkinson's disease (PD) is a neurodegenerative disorder characterized by dopaminergic neuron loss, resulting in motor and cognitive impairments. Although the etiology is unknown, the role of glial cells in neurodegeneration has gained attention. Furthermore, transcription factors (TFs), genes that regulate gene expression, have been linked to the development of various complex diseases, such as neurodegenerative disorders, as well as to cell regulation and differentiation. Our study aimed to understand the microglial regulatory microenvironment in PD across brain regions. Thus, we standardized a pySCENIC (v.0.12.1) workflow on single-nucleus RNA-seq data from 101 individuals (80 with PD) across 5 brain regions: putamen (PUT), substantia nigra (SN), middle temporal gyrus (MTG), cingulate gyrus (CG), and cerebellum (CBE). Scanpy was performed for quality control, normalization, Leiden clustering (res. 0.25), and Harmony batch correction to isolate microglial cells. Also, we performed a regulatory network (RN) analysis for each brain region. We retained 131,811 microglial cells and identified three predominant cellular states: homeostatic, reactive (Mic-REAT), and phagocytic (Mic-FAG). From RN analysis for each brain region, a total of 1,779 TFs were identified, with 97 shared across all regions. For each region, we performed a Differential AUC analysis and a cell type-specific regulators score to identify the most specific markers for Mic-FAG/Mic-REAT regulons. Preliminary analysis, we identified 4 significant regulons using Wilcoxon test ( $p_{adj} < 0.05$ ,  $\log FC > 0.2$ , scores  $> 0$  and  $pct\_group > 0.5$ ) common to SN, PUT and CG regions for these states (ATF3(+), FOS(+), THRA(+), JUN(+)) and 66 region-specific TFs, except CBE (SN=16, MTG=20, CG=13, PUT=10). ATF3 and FOS regulons suggest it acts as a core network regulon in cortical regions, despite its targetome being regionally distinct (overlap coef.  $< 0.1$ ). These regulons reinforce a central micro-regulatory network driving microglial adaptation to environmental damage associated with communication, signaling, and stimulus response. In SN and CG, regulons are linked to the pro-inflammatory JUN-FOS complex. These findings indicate region-specific microglial responses to the degenerative microenvironment across brain regions, establishing a region-specific regulatory profile for Mic-FAG/Mic-REAT in PD development. Although preliminary, these findings represent a foundational step toward understanding PD pathology.

# Taxonomic Characterization and Functional Signature of Gut Dysbiosis in Inflammatory Bowel Disease

Bianca Cristiane Ferreira Santiago<sup>1</sup>, João Vitor Cavalcanti<sup>1</sup>, Gleison Medeiros de Azevedo<sup>1</sup>, Jacinta Fernanda Novais Cezar<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

The gut microbiota is essential for host homeostasis, and its disruption (dysbiosis) characterizes Inflammatory Bowel Disease (IBD): Crohn's Disease (CD; transmural GI inflammation) and Ulcerative Colitis (UC; colonic mucosa). Given delayed diagnosis, this study contrasted taxonomic and functional profiles in CD and UC via metagenomics to identify biomarkers and disease subgroups. We analyzed data from 278 stool samples from the iHMP project (PRJNA389280; 145 CD, 78 UC, 55 nonIBD controls). Diversity (Shannon Index) and taxonomy were processed using Euryale pipeline; functional pathways were quantified via HUMAnN 3 pipeline. Multivariate linear models in MaAsLin 2 (transform = "LOG", TSS normalization) evaluated adjusted differential abundance across 4 comparisons: IBD vs nonIBD (macro signal), CD vs nonIBD, UC vs nonIBD, and CD vs UC (differential diagnosis), adopting  $q < 0.05$  (Benjamini-Hochberg FDR). Diversity was significantly lower in CD, while UC preserved diversity but showed marked abundance reduction versus controls. Dominant phyla were Actinomycetota, Bacillota, Bacteroidota, and Pseudomonadota. Patient-increased species (mainly UC) included *Waddlia chondrophila*, *Vibrio vulnificus*, and *Staphylococcus*, whereas species of genera such as *Roseburia* and *Clostridium* (mainly CD) were reduced. Functionally, global analysis (IBD vs nonIBD) revealed 35 altered pathways, 34 with negative coefficients ( $\beta < 0$ ), showing that IBD functional collapse stems from the loss of symbiotic pathways, such as SCFA/butyrate (P163-PWY) and quinone/Vitamin K2 biosynthesis (PWY-7371, PWY-7992), rather than pathogenicity gain. The only enriched pathway ( $\beta > 0$ ) in the global inflamed state was the anaerobic degradation of L-arginine (PWY-8187;  $\beta = +0.5205$ ). Stratification yielded 19 pathways in CD vs nonIBD, 16 in UC vs nonIBD, and 46 in CD vs UC. The strict intersection of all 4 comparisons isolated a core set of 9 pathways driving dysbiosis and phenotype differentiation. Preliminary findings suggest that CD and UC exhibit distinct dysbiotic profiles: CD via severe diversity loss and UC via reduced abundance. Functionally, IBD is marked by massive loss of beneficial pathways, with a restricted core useful for differential diagnosis and targeted therapies.

# Mapping Prognostic Vulnerabilities in Acute Myeloid Leukemia via Gene Network Analysis

Lorenzo Santos Furtado<sup>1</sup>, Diego Marques Coelho<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Acute Myeloid Leukemia (AML) is a heterogeneous hematologic neoplasm with clinical outcomes often diverging from what is expected based on traditional targets, highlighting the need for studies that characterize the affected molecular interactions. To better describe the personalized molecular rewiring driving AML pathogenesis, we utilized bulk RNA-sequencing data harmonized from five different consortiums to construct sample-specific gene co-expression networks by applying the LIONESS algorithm integrated with experimental protein-protein interaction evidence. We were able to compare the differentially interacting genes (DIGs) that gained connections for each of the 17 clusters, based on consensus clustering of batch-adjusted RNA-seq from 1224 AML patients, NPM1.5 (1722), NPM1.1 (1521), PML-RARA (1384), KMT2A.1 (1299), NPM1.2 (1273), CBFB-MYH11 (1182), CEBPA (942), AML\_MRC.2 (837), RUNX1-RUNX1T1 (681), AML\_MRC.3 (509), AML\_MRC.4 (500), NPM1.4 (437), AML\_MRC.5 (415), AML\_MRC.1 (77), NPM1.3 (72), KMT2A.2 (21), NUP98 (19) and subsequently investigate the biological processes affected by the interactions. The results from the enrichment analysis revealed affected immune system pathways such as T cell receptor signaling pathway (AML\_MRC.2), B cell receptor signaling pathway (AML\_MRC.5, NPM1.5), and hematopoietic cell lineage (AML\_MRC.2, CBFB-MYH11, NPM1.1), alongside altered signal transduction cascades including the MAPK (CBFB-MYH11, NPM1.2), Rap1 (NPM1.2, CBFB-MYH11), and JAK-STAT (AML\_MRC.5). Furthermore, we performed a cluster-specific spearman correlation between genes that gained connections within each cluster and the available clinical metadata. Applying an FDR threshold of 0.01, we found a negative correlation between survival time and the genes TGOLN2 (NPM1.4, NPM1.5) and KDM3B (AML\_MRC.4), as well as a positive correlation for HOXA10 (CBFB-MYH11, CEBPA, PML-RARA, RUNX1-RUNX1T1) and CPNE8 (CBFB-MYH11, CEBPA). Our data indicates that TGOLN2 and KDM3B network gains as negative prognostic factors in NPM1 and AML\_MRC subsets revealing vulnerabilities that could be exploited therapeutically. On the other hand, the prominence of HOXA10 and CPNE8 correlates with a protective clinical effect in core binding factor (CBF) and CEBPA-mutated leukemias. Together, these findings demonstrate that interactome dynamics can refine risk stratification and pinpoint subtype-specific vulnerabilities beyond classical mutational profiling.

# EXPLORING MACHINE LEARNING FOR IBS CLASSIFICATION USING GUT MICROBIOTA

João Victor Villas Boas Spelta<sup>1</sup>, Tetsu Sakamoto<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Irritable Bowel Syndrome (IBS) is a common functional gastrointestinal disorder characterized by chronic abdominal pain and altered bowel habits that substantially affect quality of life. Diagnosis is primarily clinical and relies on symptom-based criteria, while overlapping gastrointestinal conditions may require exclusion, contributing to diagnostic complexity and highlighting the need for objective diagnostic markers. IBS is also frequently associated with psychological comorbidities, underscoring the complex interactions involving the gut microbiota and the gut-brain axis. However, the contribution of gut microbial composition to IBS remains incompletely understood. Here, we used publicly available 16S rRNA gene sequencing data from the European Nucleotide Archive (BioProject PRJEB44533) to evaluate whether gut microbial composition can distinguish IBS patients from healthy controls. The dataset comprised 121 fecal samples from 30 healthy controls and 91 IBS patients, with 360 bacterial and archaeal taxonomic features detected. Microbial abundance data were transformed using the centered log-ratio (CLR) transformation to account for their compositional nature. Eleven supervised classifiers were evaluated using nested cross-validation, with 5-fold outer and 3-fold inner loops, grid-search hyperparameter tuning, and ROC-AUC as the primary performance metric. Microbiota-only models achieved a maximum mean ROC-AUC of  $0.851 \pm 0.044$  with AdaBoost, followed by XGBoost ( $0.828 \pm 0.121$ ) and Random Forest ( $0.792 \pm 0.156$ ), whereas distance-based and simpler classifiers showed lower performance. Permutation importance highlighted *Collinsella*, *Bifidobacterium*, and *Adlercreutzia*, along with uncharacterized members of the Ruminococcaceae and Barnesiellaceae families, as important features for classification. Overall, these findings suggest that gut microbial composition contains a classifier-dependent signal for distinguishing IBS from healthy controls and support further investigation of microbial signatures as potential markers for IBS classification.

# **PepMem-AI: a machine-learning pipeline for peptide-membrane interaction scoring and antimicrobial activity prioritization of *Tityus stigmurus* analogs**

Gabriel Bezerra Motta Câmara<sup>1</sup>, Giovanna A. P. Soares<sup>2</sup>, Allanny Alves Furtado<sup>1</sup>, Adriana Marina e Silva Parente<sup>1</sup>, Matheus de Freitas Fernandes-Pedrosa<sup>1</sup>, Raquel de Melo Barbosa<sup>3</sup>, Marcelo A. C. Fernandes<sup>4</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil,

<sup>3</sup>Universidade de Granada, <sup>4</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil Department of Computer Engineering and Automation, UFRN, Natal, Brazil

Antimicrobial peptides (AMPs) from scorpion venom are promising leads against multidrug-resistant pathogens, but experimental screening of analogs remains time-consuming and resource-intensive. We present PepMem-AI, a bioinformatics pipeline for prioritizing peptide-target pairs as part of a bioprospecting study of peptides derived from the native peptides Stigmurin/TsAP-2 of the northeastern Brazilian scorpion *Tityus stigmurus*. Each pair is represented by sequence descriptors (net charge, mean Kyte-Doolittle hydrophobicity, and Eisenberg hydrophobic moment) and computational membrane profiles assigned to each target, including surface charge and LPS-, peptidoglycan-, cholesterol-, and ergosterol-related properties. We introduce the Peptide-Membrane Interaction Index (PMI), an interpretable score integrating peptide-membrane charge complementarity, hydrophobic compatibility based on a typed membrane hydrophobicity proxy, peptide amphiphilicity, and a penalty associated with cholesterol- or ergosterol-rich environments. Using 90 peptide-target pairs with experimentally determined minimum inhibitory concentration (MIC) values across 10 peptides and bacterial, fungal, and reference targets, with high activity operationally defined as  $\text{MIC} \leq 3.4 \mu\text{M}$ , we trained Random Forest classifiers in two settings: a baseline using 11 classical features plus PMI and a multimodal model combining the same features with ESM-2 protein language model (PLM) embeddings. Generalization to unseen peptides was assessed by leave-one-peptide-out (LOPO) cross-validation, concatenating raw out-of-fold predictions across folds for pooled evaluation. The baseline achieved a pooled LOPO ROC-AUC of 0.854 and 76.7% accuracy, whereas the multimodal model achieved 0.843 and 80.0%, respectively. Isotonic calibration was subsequently fitted to raw OOF probabilities for probabilistic interpretation at inference and was not used to compute the reported LOPO metrics. These results show that interpretable physicochemical and target-associated descriptors were competitive with ESM-2 embeddings, which did not improve ROC-AUC in this dataset, supporting PepMem-AI as a prioritization framework for selecting Stigmurin/TsAP-2 candidates for experimental testing.

# Pan-Cancer Identification of Candidate Tumor Antigens Based on Malignant Cell Specificity

Thayana Vieira Tavares<sup>1</sup>, Pedro Lucas Alves da Silva<sup>2</sup>, Laíze Barros Pantoja<sup>3</sup>, Jorge Estefano de Santana Souza<sup>4</sup>, Andre Luis Fonseca Faustino<sup>5</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Laboratory of Bioinformatics, Artificial Intelligence, and Technological Innovation. Federal University of Rio Grande do Norte., <sup>3</sup>Universidade federal do Pará, <sup>4</sup>Digital Metropolis Institute (IMD), UFRN, Natal, Brazil, <sup>5</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Instituto Metrópole Digital (IMD); Federal University of Rio Grande do Norte (UFRN)

The discovery of tumor-specific antigens is essential for developing precise cancer immunotherapies but remains challenging due to tumor heterogeneity. Single-cell RNA sequencing (scRNA-seq) enables gene expression profiling at single-cell resolution, allowing the identification of genes specifically expressed in malignant cells and the detection of potential false-positive antigen genes in non-malignant populations. In this study, candidate tumor antigen genes were identified and prioritized using a pan-cancer scRNA-seq dataset comprising breast, liver, renal, and skin cancers, including 90,270 malignant and 265,671 non-malignant cells. Gene expression specificity was evaluated across 21 cell types using criteria adapted from the Human Protein Atlas. Candidate antigens were further assessed according to antigen type, recurrence across cancer types, and validated T-cell epitopes. A total of 22,779 genes were evaluated, resulting in 226 candidate tumor antigen genes. Among these, CYP2J2 and MYO5B were detected in all nine tumor types analyzed, highlighting recurrent expression in malignant cells. CYP2J2 has been associated with prognosis and malignant cell proliferation in liver and renal cancers, while MYO5B has been linked to prognosis in renal and pancreatic cancers. AGR2 was another notable candidate, detected in seven tumor types and associated with poor prognosis across multiple cancers. AGR2 is also being investigated as a therapeutic target through an NCI/NIH-funded vaccine project for pancreatic ductal adenocarcinoma. Importantly, the specificity analysis showed that some established Cancer-Testis Antigens (CTAs) were specifically expressed in non-malignant populations, highlighting limitations of conventional tumor markers and the value of single-cell analysis for antigen discovery. To refine candidate prioritization, immunopeptidomic evidence was integrated through previously validated T-cell epitopes. This analysis supported 113 of the 226 candidate antigens, including epitopes from CYP2J2, MYO5B, and AGR2. These 113 candidates were grouped into 11 expression modules across 24 malignant subpopulations, revealing distinct expression patterns within tumors and reinforcing malignant cell heterogeneity. Future studies should validate candidates spatially, assess normal tissue expression, and examine associations with histopathological features.

# **A High-Plasticity State Defines a Proliferative Malignant Program Preferentially Represented in Recurrent Medulloblastoma**

João Vitor Almeida da Costa<sup>1</sup>, Gustavo Lovatto Michaelsen<sup>1</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Medulloblastoma (MB), the most common malignant pediatric brain tumor, comprises four major molecular subgroups — WNT, SHH, Group 3, and Group 4, with distinct biology and clinical outcomes. Recently, a High-Plasticity State (HPS) has emerged as a proliferative malignant state potentially relevant to tumor adaptation and recurrence. Here, we characterized HPS across malignant cells from 38 MB tumors and evaluated its molecular identity, analytical robustness, and association with recurrent disease. HPS was strongly coupled to proliferation and displayed coordinated activation of DNA replication, cell-cycle checkpoint, genome-maintenance, chromosome-segregation, and mitotic programs, together with reduced neuronal-identity-associated expression. Tumor-level pseudobulk analysis, complemented by categorical cell-cycle phase control, within-cluster comparisons, and MAST, consistently preserved these major molecular axes, indicating that the HPS program extends beyond coarse differences in cell-cycle composition. Recurrent tumors showed increased representation of HPS, including within the proliferative compartment, while the core transcriptional direction was retained. Surfaceome-focused analyses further identified a restricted set of HPS-associated candidate surface markers with complementary RNA-level coverage and specificity profiles. Together, these findings define HPS as a robust proliferation-associated malignant state enriched in recurrent MB and provide a focused framework for investigating plasticity-associated biology and potential therapeutic vulnerabilities.

# **Bridging Molecular Biology and Traditional Chinese Medicine Through Functional Biological Programs**

Leonardo Luiz de Sena Fernandes<sup>1</sup>, Deisy Morselli Gysi<sup>1</sup>

<sup>1</sup>Universidade Federal do Paraná

Traditional Chinese Medicine (TCM) models health and disease through systemic, phenotype-centered frameworks, whereas contemporary molecular biology commonly relies on bottom-up approaches, inferring biological organization from genes and molecular interactions. This difference in the level of abstraction may limit direct comparisons between TCM concepts and molecular datasets. Here, we evaluated whether representing tissue biology at an intermediate functional level could reveal biologically coherent relationships that are less evident at the gene level. We analyzed the RNA expression data from 40 human tissues in the Human Protein Atlas (20,162 genes) database using 50 curated MSigDB Hallmark gene sets, which summarize broad and coordinated biological programs. GSVA was applied to log2-transformed TPM values to generate a 50-dimensional Hallmark activity profile for each tissue. Spearman correlation between tissue profiles quantified functional similarity, followed by average-linkage hierarchical clustering. Correlations ranged from  $-0.832$  to  $0.927$ . Hallmark-based clustering revealed functional relationships not restricted to anatomical or gene expression similarity. For example, heart muscle and tongue showed high functional similarity ( $\rho = 0.840$ ) and clustered within an oxidative/metabolic module together with kidney, adrenal and thyroid, driven by shared oxidative phosphorylation and metabolic programs. Conversely, cerebral cortex clustered with reproductive and epithelial tissues rather than forming an independent neural group. Five biologically coherent tissue groups were identified, characterized by immune/proliferative, metabolic/digestive, barrier/inflammatory, oxidative/metabolic, and reproductive/epithelial programs. In contrast, previous gene-level similarity analysis resulted in 33 of 40 tissues collapsing into a single cluster. These findings indicate that molecular similarity is strongly dependent on the biological level at which data are represented. Rather than directly mapping a phenotype-centered system such as TCM onto individual genes or highly specific molecular pathways, functional representations such as Hallmarks may provide a more appropriate intermediate layer for identifying systemic biological patterns. This hierarchical strategy may facilitate subsequent integration of molecular evidence with TCM frameworks while reducing the risk of imposing macro-level concepts onto inherently bottom-up molecular structures.

# Passage-level retrieval evaluation in a RAG pipeline for antimicrobial prescribing support: an approach to antimicrobial resistance stewardship

Jean Paes Landim de Lucena<sup>1</sup>, Mariana Timbó de Oliveira<sup>1</sup>, Patrick Terrematte<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Antimicrobial resistance is driven largely by inappropriate prescribing, making guideline-grounded decision support systems a direct lever on prescription quality. Retrieval-augmented generation (RAG) promotes this, but is usually evaluated by answer accuracy or document-level retrieval metrics. Neither shows whether the passage carrying the recommended regimen reached the generator, which prevents attributing a failure to the responsible module. This present work aims to measure passage-level retrieval quality against a human reference annotation and whether it diverges from document-level evaluation enough to change component choice in a guideline-grounded prescribing support system. A corpus of 20 official clinical guidelines about infectious diseases were used as knowledge base and 135 multiple-choice questions from REVALIDA (Brazilian medical licensing exam) were curated for benchmarking, and from these questions a stratified subsample of 45 questions were manually annotated for the reference passage, anchored to character offsets rather than passage identifiers, keeping it valid across segmentations. Across 44 questions, 87 annotations were recorded; 6 questions found no supporting passage, delimiting corpus coverage. On the remaining 38, dense retrieval, hybrid retrieval (dense + BM25) and hybrid with cross-encoder reranking were compared on document-level hit rate, Recall@5, gold passage coverage, MRR and nDCG@5, with bootstrap 95% CI (10,000 resamples) and exact paired McNemar tests. The following approaches: dense, hybrid and hybrid+reranker, were evaluated using LLM retrieval metrics that reached respectively: document-level hit rate 73.7%, 78.9% and 86.8%; Recall@5 52.6% (95% CI 36.8-68.4), 71.1% (55.3-84.2) and 60.5% (44.7-76.3); coverage 36.6%, 54.6% and 45.2%; MRR 0.354, 0.454 and 0.382; nDCG@5 0.318, 0.447 and 0.352. Reranking achieved the highest document-level hit rate but a Recall@5 10.6% below plain hybrid: it brings the right document and the wrong paragraph. Paired, the gain of hybrid over dense was supported (7 discordant, all favorable,  $p=0.016$ ); for reranking it was not ( $p=0.219$ ). Therefore, document-level evaluation overestimates the context actually delivered to the generator and may favor the wrong component. For antimicrobial prescribing support, where the regimen lives in one specific passage, this is the difference between citing the right guideline and recommending the right drug.

# TRANSCRIPTOMICS REVEALS HEMOGLOBIN ADAPTATIONS TO MODERATE ALTITUDE IN SOUTH AMERICAN ANURANS

Leandro Amaro da Silva Felix<sup>1</sup>, Lucas Yago Melo Ferreira<sup>2</sup>, João Pedro Nunes Santos<sup>2</sup>,  
Eric Roberto Guimarães Rocha Aguiar<sup>3</sup>, Mariana Leivas Muller Hoff<sup>1</sup>

<sup>1</sup>Universidade Federal de Pernambuco, <sup>2</sup>State University of Santa Cruz, <sup>3</sup>Federal University of Southern  
Bahia

The partial pressure of oxygen (O<sub>2</sub>) in the atmosphere decreases with increasing altitude. Therefore, high altitude inhabiting animals require physiological adaptations to lead with environmental hypobaric hypoxia. Hemoglobin (Hb), is an essential O<sub>2</sub> transport molecule in all vertebrate groups. This respiratory pigment is a tetrameric protein formed by different subunits, each presenting different oxygen affinities, which have been reported as part of the high altitude adaptation in vertebrates, including anurans. To test whether South American anurans hemoglobin subunits show adaptations to moderate altitude's O<sub>2</sub> partial pressures (PO<sub>2</sub>), we conducted a transcriptomic analysis of RNA-sequencing (RNA-seq) liver samples of four species. publicly available at the National Center for Biotechnology Information website (PRJNA625657). *Chimerella mariaelenae* and *Nymphargus posadae* occur between 1000 and 2800 m above sea level and constitute the moderate altitude group (MAG). *Pseudis minuta* and *Sphaenorhynchus lacteus* represented the low altitude group (LAG). FastQC quality check and Trimmomatic pre-processing were performed. Trimmed reads were used for Trinity de novo assembly as well as for Salmon mapping. Assembly redundancy was reduced with Cd-hit, quality control was checked with BUSCO, the contigs coding region sequences were predicted by TransDecoder and annotated as *Xenopus Tropicalis* proteins (UP000008143) using BlastP. After annotation, Salmon output was used for differentially expressed genes (DGE) analysis in EdgeR. We also aligned the obtained Hb alpha (Hba) amino acid sequences with sequences of *Homo sapiens*, *Telmatobius peruvianus* and *X. tropicalis* and by MAFFT. Despite the lack of DEG, we observed Ser131Ala substitution in Hba, which is related to decreased chloride sensibility and higher O<sub>2</sub> affinity, in MAG and in *P. minuta*. This later species is semi-aquatic and this substitution may be related to aquatic respiration. Surprisingly, the ratio of larval/adult Hba is higher in LAG (~0,7) than in MAG (<0,25), showing that other factors, in addition to altitude, influence the expression of hemoglobins in these species. Our transcriptomic analyses showed higher Hb affinity based on protein sequence in MAG and different expression patterns of hemoglobin subunits expression, which, together, may be part of the adaptations to low PO<sub>2</sub> posed by moderate altitude in *C. mariaelenae* and *N. posadae*.

# **Predicting adverse events for risk stratification of chemotherapy based stem cell mobilization in multiple myeloma**

Friedrich Schwarz<sup>1</sup>, Enver Aydilek<sup>2</sup>, Nils Brökers<sup>2</sup>

<sup>1</sup>University of Goettingen, <sup>2</sup>University Medical Center Goettingen

Autologous stem-cell transplantation is a fundamental therapy for multiple myeloma. Although inpatient chemo-based stem-cell mobilization (SCM) is standard care in Germany, outpatient approaches could ease healthcare constraints. We analyzed 109 myeloma patients undergoing SCM and collection at the University Medical Center Göttingen for safety. We then trained machine learning models to predict adverse events (AEs) requiring hospitalization and to forecast AE onset timing for optimized ward management. In our cohort, 97% achieved successful collection, but 69% experienced severe AEs necessitating hospitalization. Simulations suggest a risk-stratified outpatient protocol could cut bed usage by at least one third without compromising safety. Classification models accurately predicted some AE types (e.g., elevated creatinine, ROC-AUC 1.0), though neutropenic fever remained challenging (ROC-AUC 0.67). Regression models forecast AE onset with a mean error of just over one day. These results outline a data-driven roadmap for safely adopting outpatient SCM and optimizing resource allocation in clinical practice.

# **Computational Investigation of the I308V Mutation in the GAUT9 Enzyme of *Manihot esculenta* with Foliar Subsessility and Disease Resistance**

Nycolas Henrique Moraes de Souza<sup>1</sup>, Serli de Oliveira Cabral<sup>2</sup>, André da Silva Xavier<sup>3</sup>, Danilo Fernandes Coelho<sup>1</sup>

<sup>1</sup>UNIVERSIDADE FEDERAL DE PERNAMBUCO, <sup>2</sup>UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO,  
<sup>3</sup>UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

Cassava (*Manihot esculenta*) exhibits spontaneous mutants with foliar subsessility, a trait characterized by a pronounced reduction in the leaf petiole that introduces interesting agronomic features, such as disease resistance, lower leaf temperature, chlorophyll accumulation, and delayed leaf abscission. Despite being advantageous, the genetic mechanisms and pleiotropic effects related to this mutation remain poorly understood. This study investigated, through computational simulations, the potential impact of the I308V mutation on the structural dynamics of the Galacturonosyltransferase 9 (GAUT9) enzyme, which is encoded by a candidate gene for the manifestation of foliar subsessility in cassava. The 3D structures of both native and mutant variants, each containing the Mg<sup>2+</sup> ion in the active site, were predicted using AlphaFold2. The energetic impact of the mutation was estimated using FoldX and DynaMut2 softwares, narrowing the structures down to the catalytic domain (residues 77-561) to focus on the active site behavior. Initially, 100 ns Molecular Dynamics (MD) simulations were performed using the GROMACS 2025.4 package with the CHARMM36 force field. For a more robust assessment, ongoing 1  $\mu$ s triplicates are being conducted to analyze structural stability, flexibility, and conformational population, using VMD software for visual inspections. Thermodynamic predictions indicated that the mutation is destabilizing, and preliminary 100 ns results revealed a loss of local interactions within the active site of the mutant variant. Although RMSD, radius of gyration, and secondary structure map analyses show that both proteins maintain their global folding, PCA and free energy surface (FEL) analyses demonstrated that the mutant possesses a more restricted conformational dynamics than the native type. In conclusion, the I308V mutation does not compromise the global folding of GAUT9, but acts locally by reducing hydrophobic interactions and restricting the enzyme's conformational dynamics, pointing to a possible allosteric modulation associated with foliar subsessility.

# **Identification of putative lncRNAs expressed by two strains of *Leishmania infantum* (BOS1FL1 and BCN150) during in vitro infection of canine macrophages**

Bruno Mendes Ceneviva<sup>1</sup>, Mariana Cordeiro Almeida<sup>1</sup>, Matheus Fujimura Soares<sup>1</sup>,  
Juliana De Souza Felix<sup>1</sup>, Flávia Lombardi Lopes<sup>1</sup>

<sup>1</sup>Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP

**Introduction:** *Leishmania infantum* is the primary causative agent of visceral leishmaniasis in the Americas. Long non-coding RNAs (lncRNAs) play regulatory roles during host-parasite interactions, yet their role in *L. infantum* remains poorly understood. **Objectives:** This study aimed to analyze putative lncRNAs in two *L. infantum* strains (BOS1FL1 and BCN150) during in vitro infection of canine macrophages (DH82) to clarify strain-specific and shared infection mechanisms. **Materials and Methods:** RNA-Seq data from DH82 macrophages infected for 4h (PRJNA701880) were analyzed. Non-coding potential was evaluated via FEELnc, CPC2, and BASiNET, followed by Pfam and Infernal filtering. Host-pathogen interaction was assessed by predicting lncRNA binding potential against canine mRNAs using IntaRNA, followed by g:Profiler functional enrichment analysis of target mRNAs. **Results:** Bioinformatic prediction identified >1,200 candidate lncRNAs per strain with high consensus across tools (1,444 for BOS1FL1; 1,273 for BCN150). Candidate lncRNAs spanned all 36 chromosomes with a conserved density pattern (peak on LinJ.31; lowest on LinJ.03) and a similar size profile skewed toward 200–1,200 nt. Functional profiling of canine target mRNAs revealed significant enrichment across key KEGG pathways for both strains, including PI3K-Akt signaling, Hematopoietic cell lineage, Complement and coagulation cascades, Regulation of actin cytoskeleton, and ECM-receptor interaction. **Discussion:** High tool consensus and stringent filtering demonstrate a methodology for lncRNA identification in *Leishmania*, aligning with previous *L. braziliensis* findings. Conserved chromosomal distribution, size profiles, and target pathway enrichments suggest that lncRNAs may act as regulatory elements in canine leishmaniasis caused by *L. infantum*. In addition, binding predictions between parasite lncRNAs and host mRNAs point toward cross-kingdom regulation, suggesting parasite lncRNAs may modulate host post-transcriptional networks to subvert immune defenses. **Conclusion:** This study establishes a catalog of candidate lncRNAs in *L. infantum* during infection. By demonstrating parasite lncRNA binding potential toward host mRNAs, our findings suggests a novel mechanism of host-pathogen crosstalk where the parasite may actively rewire host macrophage function, advancing leishmaniasis pathogenesis knowledge and highlighting new therapeutic targets.

# HumanPopGen APP/API: An Architecture for Storing and Querying Human Genetic Diversity Data

Rebeca Costa dos Reis Galeno<sup>1</sup>, Lucas Brito da Silva<sup>2</sup>, Layana Cristine Carvalho Fernandes<sup>2</sup>, Gilderlanio Santana de Araújo<sup>2</sup>, Helber Gonzales Almeida Palheta<sup>2</sup>

<sup>1</sup>Universidade Federal Rural da Amazônia, <sup>2</sup>Universidade Federal do Pará

Population genomics aims to study DNA sequences from multiple individuals in a large scale. Bioinformatics advances have revolutionized the understanding of human genetic variation, becoming essential for variant interpretation and biomedical research. However, the large volume of data poses challenges for efficient and interactive storage, distribution, and querying, especially in environments with limited computational infrastructure. This study aimed to develop the HumanPopGen APP/API, a platform based on the DuckDB-Parquet architecture for distributing and querying population genomics data. We investigated whether columnar storage and analytical databases can provide an efficient alternative. To do so, data were derived from genome sequences from 4,150 samples from the Human Genome Diversity Project and the 1000 Genomes Project. The data were processed through pipelines developed in Python, normalized, and converted from the .json format to Apache Parquet, organized into chromosomal partitions, and queried using the DuckDB analytical database. In addition, the platform provides an API developed in FastAPI that enables queries by rsID, genomic coordinates using the hg38 reference genome, continental and subcontinental allele frequencies, and structured .json representations. The HumanPopGen APP/API also offers a web interface and a Progressive Web Application (PWA), allowing researchers to access the platform without the need for specialized database software, while incorporating a functional annotation layer based on AlphaGenome predictions for regulatory effects. Converting the data from .json to Apache Parquet reduced the dataset from approximately 524 GB to 30 GB, and the variant index, based on rsIDs, contains more than 148 million records distributed across 5,722 parquet files. The results demonstrate that the proposed strategy significantly reduces data storage requirements and improves query efficiency without requiring distributed infrastructure or large conventional database systems. Therefore, the HumanPopGen APP/API constitutes a portable and accessible solution for managing large-scale population genomics datasets, contributing to broader access to information and supporting research in bioinformatics and human genetics.

# Identification of Consistent and Functionally Reproducible Transcriptomic Signatures in Triple-Negative Breast Cancer through Comparison of Independent RNA-seq Cohorts

Gabriel Freitas Araújo<sup>1</sup>, Túlio de Lima Campos<sup>1</sup>, Luydson Richardson Silva Vasconcelos<sup>2</sup>

<sup>1</sup>Instituto Aggeu Magalhães - FIOCRUZ, <sup>2</sup>Instituto Aggeu Magalhães FIOCRUZ/PE

Triple-negative breast cancer (TNBC) represents the most aggressive breast cancer subtype and exhibits high molecular heterogeneity, hindering the identification of reproducible biomarkers and limiting the availability of therapeutic targets. Although several studies have described differentially expressed genes in this subtype, few have investigated the consistency of these alterations across different cohorts. Therefore, this study aimed to identify consistent transcriptomic signatures and characterize the biological processes altered in TNBC. For this purpose, tumor breast tissue samples were compared with non-tumor breast tissue samples across three independent RNA-seq cohorts obtained from public repositories, using the edgeR package to identify shared differentially expressed genes with consistent directions of regulation, followed by functional enrichment analyses of Gene Ontology (GO) terms using PANTHER and protein-protein interaction (PPI) analysis using STRING. Overall, 107 genes were identified as shared among the cohorts, of which 90 exhibited consistent transcriptional patterns, constituting a consensus TNBC signature. Upregulated genes (n = 34) were enriched in processes related to the cell cycle, cell division, and chromatin organization, highlighting CDC20, PLK1, and TPX2. In contrast, downregulated genes (n = 56) were enriched in lipid metabolism, cilium organization, and cellular projection formation, including PLIN1, LIPE, DYNC2H1, and DYNC2LI1. PPI interaction networks revealed distinct functional modules corresponding to these biological processes. Among the main findings, the consistent downregulation of genes simultaneously associated with cilium organization and lipid metabolism was observed across all three cohorts, suggesting a possible functional connection between these processes, which remains poorly explored in TNBC. These results demonstrate that the integration of independent cohorts enables the refinement of biologically reproducible transcriptomic signatures, providing priority candidates for future functional investigations and molecular biomarker development. Furthermore, this consistency was observed across cohorts from geographically distinct populations, reinforcing that the identified biological processes represent robust molecular alterations in TNBC rather than artifacts specific to a particular population.

# Virtual screening of curcumin in cutaneous carcinogenesis: docking with reference ligands and ADMET profile

jeniffer gabrielly de souza pereira<sup>1</sup>, Marina de Lourdes Ribeiro Alves Coelho<sup>2</sup>, Marcos Antonio Nobrega de Sousa<sup>3</sup>, Rafael Trindade Maia<sup>1</sup>

<sup>1</sup>universidade federal da paraiba, <sup>2</sup>Universidade Federal da Paraiba, <sup>3</sup>Universidade Federal de Campina Grande

Cutaneous carcinogenesis is driven by dysregulated proliferation, inflammation, apoptosis and oncogenic signaling, supporting the screening of natural compounds with multitarget potential. Curcumin, a polyphenol from *Curcuma longa*, has recognized bioactivity; however, its interaction with proteins in skin tumor progression requires integrated computational evaluation. This study assessed the predictive affinity of curcumin against HSP90, IKKB, SMO, COX2 and AKT1, comparing its performance with reference ligands. Protein structures were initially searched in the Protein Data Bank; however, some targets showed chain discontinuities, leading to the use of AlphaFold models refined with GalaxyRefine. Structural cleaning was performed in PyMOL by removing water molecules, ligands and non-essential components. Curcumin and the controls geldanamycin, BMS-345541, vismodegib, celecoxib and MK-2206 were prepared in Open Babel by hydrogen addition, Gasteiger charges, minimization and conversion to PDBQT. Molecular docking was conducted in AutoDock Vina using blind docking and equal search parameters for curcumin and controls, enabling direct comparison of predicted binding energies. ADMET properties were assessed in ADMETlab 3.0. Curcumin showed affinity for all targets, with binding energies of -6.810 kcal/mol for HSP90, -6.724 for IKKB, -6.690 for SMO, -6.574 for COX2 and -6.409 kcal/mol for AKT1. Compared with controls, curcumin showed more favorable energy than geldanamycin in HSP90 (-6.810 versus -6.302 kcal/mol) and BMS-345541 in IKKB (-6.724 versus -6.518 kcal/mol). For SMO, COX2 and AKT1, reference ligands showed higher affinity: vismodegib (-7.131), celecoxib (-6.996) and MK-2206 (-7.573 kcal/mol), validating the discriminatory behavior of the docking model. These findings indicate HSP90 and IKKB as the most promising targets for curcumin, while SMO and COX2 remain relevant secondary targets. ADMET analysis showed favorable physicochemical properties, including molecular weight of 368.13 g/mol, TPSA of 93.06, logP of 2.147, no Lipinski violations and no PAINS alerts. Nevertheless, predicted limitations included permeability, high plasma protein binding, short half-life, skin sensitization, eye irritation and respiratory toxicity alerts. In conclusion, curcumin showed competitive performance against reference ligands, especially for HSP90 and IKKB, supporting its multitarget potential and justifying molecular dynamics and experimental validation studies.

# **RegVAE-Bov: Regulatory Variational Autoencoder for Multi-Omics Integration of the Bovine Mammary Gland under Heat Stress**

Giovanna A. P. Soares<sup>1</sup>, Matheus Dalmolin<sup>2</sup>, Raylam Wendel da Silva<sup>3</sup>, Marcelo A. C. Fernandes<sup>4</sup>

<sup>1</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil, <sup>2</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil, <sup>3</sup>Universidade Federal do Rio Grande do Norte, <sup>4</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil Department of Computer Engineering and Automation, UFRN, Natal, Brazil

This work proposes RegVAE-Bov, a regulatory-weighted variational autoencoder that integrates single-nucleus RNA-seq, bulk RNA-seq, ATAC-seq and bisulfite sequencing of the bovine mammary gland into a shared ten-dimensional latent space. By treating epigenomic evidence as a reconstruction prior rather than an input feature, this approach keeps chromatin and methylation patterns out of the encoder and preserves them for independent validation. Two public datasets serve as benchmarks: GSE317482, with 26,243 nuclei annotated into eight cell types across thermoneutral, heat-stress and pair-fed conditions, and GSE249950, with 18 bulk MAC-T samples across six conditions, complemented by 681,895 ATAC-seq peaks and single-CpG methylation from 16 samples. The encoder maps 2,000 harmonised highly variable genes through 256 and 128 units to ten latent dimensions and is trained exclusively on the nuclei; the bulk samples are projected afterwards with the frozen encoder, so their latent position is a prediction rather than a memorised assignment. In addition, two complementary mechanisms are proposed: a binary regulatory mask, which distinguishes genes without epigenomic coverage from genes without regulation, and a learnable gate, which optimises the contribution of each epigenomic layer jointly with the network. An orthogonal decomposition of the transcriptomic drift attributes 87% to cell-state reprogramming and 13% to compositional remodelling, a partition that is deterministic and independent of the model. FKBP5 ranks among the three strongest drifts in six of eight cell types, and endothelial NFKBIA is the only candidate supported by three independent omics layers. Basal bulk samples project between the luminal secretory and progenitor populations without supervision, and recover to 88% of baseline within six hours ( $d = 0.204 \pm 0.046$  over five initialisations). The gate converges to negative weights on all three epigenomic signals, indicating that the epigenetically primed heat-shock response is predominantly stromal rather than luminal. By integrating deep generative modelling with epigenomic priors, this framework establishes a scalable basis for projecting bulk transcriptomes onto single-cell manifolds and for quantifying the reversibility of environmental stress responses.

# **Transcriptomic alterations in mature oligodendrocytes following ethanol exposure: Insights into Fetal Alcohol Spectrum Disorder**

Caio Maia Farias de Andrade<sup>1</sup>, Romério de Oliveira Lima Filho<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>2</sup>

<sup>1</sup>Universidade Potiguar, <sup>2</sup>Universidade Federal do Rio Grande do Norte

Fetal Alcohol Spectrum Disorder (FASD) encompasses a range of neurodevelopmental alterations caused by alcohol consumption during pregnancy, representing a major public health concern. Among these alterations, white matter impairment suggests oligodendrocyte dysfunction, highlighting their potential role in FASD pathogenesis. Previous studies have shown that alcohol disrupts myelin formation and oligodendrocyte function, however, the molecular mechanisms underlying these effects remain poorly understood. High-throughput RNA sequencing combined with reproducible bioinformatics workflows provides an effective approach to characterize transcriptomic alterations and identify biological pathways associated with alcohol-induced cellular dysfunction. In this study, we performed a differential gene expression analysis using publicly available RNA-Seq data (PRJNA1000573) from a *Mus musculus* model to investigate the effects of embryonic alcohol exposure on the transcriptome of mature oligodendrocytes. Cortical oligodendrocytes isolated from postnatal day 5–7 mice were cultured and differentiated *in vitro* for three days, followed by exposure to moderate (10 mM) or high (30 mM) ethanol concentrations for an additional three days. Raw sequencing data were processed using the nf-core RNA-seq and Differential Abundance pipelines implemented in Nextflow, including quality control, read trimming, alignment, transcript quantification, normalization, and gene count generation. Differential expression analysis was performed using DESeq2, while functional enrichment was assessed through Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) using the clusterProfiler package in R. Genes with an adjusted p-value < 0.05 were considered significantly differentially expressed. High alcohol exposure induced a markedly stronger transcriptional response, with 956 differentially expressed genes (197 upregulated and 759 downregulated), whereas moderate exposure resulted in 43 genes (36 upregulated and 7 downregulated). Functional enrichment revealed significant alterations in pathways related to protein translation, cell cycle regulation, integrin signaling, inflammatory response, cytoskeleton organization, and myelination. These results demonstrate that ethanol induces concentration-dependent transcriptomic changes in mature oligodendrocytes providing new insights into the molecular pathways potentially involved in FASD pathogenesis.

# Systems biology approach uncovering a network model of EWSR1::FLI1 effect in Ewing sarcoma

Marialva Sinigaglia<sup>1</sup>, Shantanu Gupta<sup>2</sup>, André T. Brunetto<sup>1</sup>, José Carlos Merino Mombach<sup>3</sup>, Daner Acunha Silveira<sup>1</sup>

<sup>1</sup>Instituto do Câncer Infantil, <sup>2</sup>Universidade Federal do Rio Grande do Norte, <sup>3</sup>Universidade Federal de Santa Maria

Ewing sarcoma (ES) is a rare and aggressive pediatric malignancy driven primarily by the oncogenic fusion protein EWSR1::FLI1. This transcription factor profoundly alters gene regulation, promoting tumor progression and cellular plasticity, although its broader regulatory network remains incompletely understood. In this study, we aim to construct and simulate a computational model integrating key pathways implicated in ES, with a particular focus on the regulatory influence of EWSR1::FLI1 on downstream network components. We have performed an extensive bibliographical search in order to select the most important pathways used as therapeutic targets in ES. The collected data were translated into a molecular interaction network. Subsequently, the EWSR1::FLI1 targets were mapped and included in the network. The activation and inactivation dynamics of the nodes were modeled by assigning logical operators to each component, representing the regulatory effects of upstream molecules. The model was evaluated by comparing the simulated biological responses of its components with experimental observations reported in the literature, using *in silico* gain-of-function (GoF) and loss-of-function (LoF) perturbations. The pathways included in the network comprise CXCL12/CXCR4, WNT/ $\beta$ -catenin, SHH, Notch, Hippo, IL6/JAK/STAT3, and TNF signaling pathways. A total of 13 EWSR1::FLI1 target genes were incorporated into the network, resulting in a model with 66 nodes and 173 edges. Simulations under wild-type conditions revealed a high number of proliferative states, consistent with the aggressive behavior observed in ES. Pathway-specific perturbation analyses showed overall consistency between model predictions and experimental observations reported in the literature. Furthermore, gain-of-function (GoF) simulations of EWSR1::FLI1 led to an increased number of proliferative states, reinforcing its central role in promoting tumor aggressiveness. Preliminary simulations of p53 GoF resulted in the induction of apoptotic states, in agreement with current literature on ES. Future work will focus on simulating the effects of known therapeutic agents targeting network components, with the aim of identifying potential synergistic effects.

# Seq2Image-DTI: Tokenization-Enhanced Molecular Image Representation for Drug-Target Interaction Prediction

Hannah Isabele Da Silva Marques<sup>1</sup>, Giovanna A. P. Soares<sup>2</sup>, Matheus Dalmolin<sup>3</sup>,  
Raquel De M. Barbosa<sup>4</sup>, Marcelo A. C. Fernandes<sup>5</sup>

<sup>1</sup>Universidade Federal do Rio Grande Do Norte, <sup>2</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil, <sup>3</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil, <sup>4</sup>Faculty of Pharmacy, Granada University, Granada, Spain, <sup>5</sup>InovAI Lab, nPITI/IMD, UFRN, Natal, RN, Brazil Bioinformatics Multidisciplinary Environment (BioME), IMD, UFRN, Natal, Brazil Department of Computer Engineering and Automation, UFRN, Natal, Brazil

Drug-target interaction (DTI) prediction is an important computational task in drug discovery, and the quality of molecular and protein representations directly affects predictive performance. Image-based approaches such as MPS2IT-DTI have shown that molecular SMILES strings and protein sequences can be transformed into two-dimensional representations and processed by convolutional neural networks, providing an alternative to sequence models based on learned embeddings. Building on this concept, we present Seq2Image-DTI, an extension of the sequence-to-image framework that introduces a symbol-aware SMILES tokenization stage before molecular image generation. In the original character-based representation, chemically meaningful multi-character symbols may be separated during sequence processing. Seq2Image-DTI instead tokenizes the SMILES string into meaningful molecular symbols and then generates a k-mer-based image representation from the resulting token sequence. Protein sequences are transformed into images using the established sequence-to-image strategy. The molecular and protein images are processed by independent CNN branches, whose extracted features are concatenated and passed through fully connected layers to predict drug-target interactions. The framework was evaluated on the Davis and KIBA benchmark datasets using five-fold cross-validation and compared with established DTI methods. The results demonstrate that the proposed tokenization strategy can be incorporated into the image-based DTI pipeline while preserving its compact representation and predictive capability, with particularly strong binary classification performance on the Davis dataset. Seq2Image-DTI therefore provides a refinement of image-based molecular representation by explicitly preserving SMILES-level chemical symbols before visual encoding.

# Integration of Molecular Data for Identifying Biomarkers of Monogenic Diabetes (MODY3)

Joana Menezes de Faria<sup>1</sup>, DANIEL JERÔNIMO DO NASCIMENTO<sup>1</sup>, LEONARDO CAPISTRANO FERREIRA<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Maturity Onset Diabetes of the Young (MODY) is a rare form of monogenic diabetes, frequently underdiagnosed due to its clinical similarity to type 1 and type 2 diabetes. Its most common subtype, MODY3, is caused by pathogenic variants in HNF1A, a transcription factor essential for gene regulation in pancreatic beta cells and hepatocytes. Once genetic sequencing (the diagnostic gold standard) has high cost and limited availability, complementary biomarkers are needed to support diagnostic screening. We hypothesized that genes regulated by HNF1A, together with their downstream circulating proteins and metabolites, could serve as potential biomarkers for MODY3. We followed a sequential analytical pipeline: (1) ChIP-seq data from hepatocytes (Ng et al., 2024) identified 3,121 target genes regulated by HNF1A; (2) cross-referencing these targets with The Human Protein Atlas showed that 723 of them encode proteins detectable in blood by immunoassay or mass spectrometry; (3) querying the GWAS Catalog database for protein quantitative trait loci (pQTL) identified 83 distinct proteins in blood associated with variants in HNF1A, with C-reactive protein (CRP) and gamma-glutamyl transferase (GGT) as the most associated, corresponding to approximately 29.8% and 11.2% of the total hits, respectively; (4) the corresponding metabolite QTL GWAS Catalog query identified 94 distinct circulating metabolites (190 hits), predominantly lipid-related compounds, led by total cholesterol (20 hits, 10.5% of hits) and polyunsaturated fatty acids/linoleic acid (7 hits each, 3.7%). The recurrent association between HNF1A and the circulating proteins CRP and GGT reproduces findings already described in the literature, supporting the validity of this approach, while the lipid-related metabolite signature points to additional, less-explored candidates. We conclude that this stepwise integration of tissue-specific gene regulation with pQTL/metQTL data offers a promising strategy for prioritizing complementary biomarkers for the molecular diagnosis of MODY3, with the potential to reduce costs and underdiagnosis. As future perspectives, we aim to prioritize and experimentally validate the candidate biomarkers identified in this study.

# **Amazon AquaBio: A Web Platform for the Dissemination and Comparative Analysis of Amazonian Fish Mitogenomes**

Júlio César da Silva Filho<sup>1</sup>, Larissa Martins Brito e Silva<sup>1</sup>, Gleison Medeiros de Azevedo<sup>2</sup>, Daniel Henrique Ferreira Gomes<sup>1</sup>, Sávio Lucas de Matos Guerreiro<sup>3</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>2</sup>, Beatriz Stransky<sup>2</sup>, Jorge Estefano Santana de Souza<sup>1</sup>

<sup>1</sup>UFRN, <sup>2</sup>Universidade Federal do Rio Grande do Norte, <sup>3</sup>UFRA

The Amazon Basin harbors more than 2,500 fish species, yet this exceptional biodiversity remains underrepresented in public genomic databases. Moreover, available genomic data are often underutilized, limiting their potential to support biodiversity assessment, management, and conservation. To address this gap, we reconstructed 100 mitochondrial genomes (mitogenomes) from 34 Amazonian fish species using publicly available raw sequencing data retrieved from the NCBI. We implemented an automated and reproducible bioinformatics workflow for this process. The pipeline integrates de novo assembly with NOVOPlasty and comparative validation with Unicycler. Gene annotation was performed using MitoAnnotator/MitoFish and cross-validated with MITOS2, alongside additional tools for quality control and downstream analyses. This approach generated 64 previously unreported mitogenomes. Notably, 22 of these species had no prior mitogenomic records in GenBank. Furthermore, we identified potentially heteroplasmic variants in 28 species. To facilitate data access and exploration, we developed Amazon AquaBio, a web platform built with React and Vite and designed according to FAIR principles. The platform integrates processed genomic data, methodological information, and comparative analyses into an accessible, visually oriented environment. It serves researchers, students, taxonomists, and professionals working in aquaculture, biodiversity, and conservation. Amazon AquaBio provides comparative tools for side-by-side analysis of mitochondrial genes. It also features dedicated modules for circularized contigs, relative synonymous codon usage (RSCU), tandem repeats, and synteny. Additional visual modules cover sequencing coverage, AT skew, tRNAs, overlapping regions, and intergenic spacers. By transforming underutilized public sequencing data into structured and accessible genomic resources, Amazon AquaBio supports comparative mitogenomic research while promoting scientific dissemination and data reuse. Ultimately, the platform provides a foundation for future taxonomic, phylogenetic, and biodiversity studies, as well as for the development of molecular approaches for species identification, monitoring, and conservation of Amazonian fish diversity.

# Genetic variants on the sex chromosomes and the risk of leprosy

DANIEL JERÔNIMO DO NASCIMENTO<sup>1</sup>, LEONARDO CAPISTRANO FERREIRA<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte

Leprosy is an infectious disease marked by acute inflammatory reactions and male predominance. Although this disparity suggests sex-linked genetic factors, genome-wide association studies have focused mainly on autosomes, leaving the sex chromosomes and pseudoautosomal regions (PARs) underexplored. We investigated whether X, PAR, and Y variants influence susceptibility, clinical forms, and reactional phenotypes in a Brazilian cohort of 770 unrelated individuals (365 females and 405 males). After filtering, 991 X, 574 PAR, and 488 Y variants were analyzed with PLINK 2.0 and XWAS. X, PAR, and Y variants were tested by logistic regression under different X-inactivation assumptions, standard logistic regression, and Firth regression, respectively. Y haplogroups were inferred with Yhaplo, Y-Lineage, and Snappy, and functional effects were assessed with GTEx and AlphaGenome. Variant rs1013150 (X) was associated with type 2 reaction under a random X-inactivation model (OR = 1.74; P = 0.0000492). In males, allele A was more frequent in cases than controls (45.05% vs. 19.66%; OR = 3.33; P = 0.000133). In females, allele A was also more frequent in cases (40.00% vs. 31.72%), but neither the allelic (OR = 1.43; P = 0.205) nor global genotypic test (P = 0.319) was significant. Variant rs6644250, identified in the PAR, was associated with reduced reaction risk. Allele A was less frequent in cases than controls in females (25.94% vs. 37.59%; OR = 0.58; P = 0.0036) and males (29.11% vs. 40.09%; OR = 0.61; P = 0.0048), with significant global genotypic differences in both sexes (P = 0.0082 and P = 0.0046). No individual Y variant reached significance. Haplogroup R was associated with increased odds of paucibacillary (OR = 2.64; P = 0.0256), multibacillary (OR = 2.14; P = 0.0366), and overall leprosy (OR = 2.24; P = 0.0248), whereas haplogroup I was associated with reduced odds of multibacillary (OR = 0.31; P = 0.0294) and overall leprosy (OR = 0.35; P = 0.0404). Functional analyses indicated regulatory effects of rs1013150 at the TLR7/TLR8 locus and of rs6644250 on genes associated with immune and hematopoietic pathways. In conclusion, the findings suggest that specific X-chromosomal and PAR loci are part of the genetic architecture of leprosy reactions. The associations observed for Y-chromosome haplogroups suggest that paternal lineages may contribute to the higher disease prevalence among men, highlighting the role of non-autosomal regions in sex-biased outcomes.

# Single-Sample Network Topology Unveils Pathogenic Hubs in Sjögren's Disease

João Vitor Cavalcanti<sup>1</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>, Diego Marques Coelho<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>UFRN

Sjögren's disease (SD) is a systemic autoimmune condition characterized by extensive clinical and biological heterogeneity, complicating the development of targeted therapies. To better describe the personalized molecular rewiring driving SD pathogenesis, we utilized bulk RNA-sequencing data from whole blood samples of the PRECISESADS IMI consortium to construct sample-specific gene interaction networks by applying the LIONESS algorithm integrated with experimental protein-protein interaction evidence. Comparing SD patients against healthy controls, we identified an absolute total of 647 Differentially Interacting Genes (DIGs) representing structural shifts in network connectivity. Topological clustering of these DIGs revealed four major functional macro-modules underlying disease pathology: defense response to virus, B cell activation, DNA replication, and positive regulation of protein catabolic process. Central topological hubs, such as ISG15, OASL, UBE2L6, and LGALS3BP, were identified as drivers of this reorganization, a fact supported by integrating network topology with single-sample Gene Set Enrichment Analysis (ssGSEA), which demonstrated that the interaction degree of these hubs exhibits strong positive correlations with the functional activity of their respective pathogenic pathways. To translate these structural findings into therapeutic opportunities, we mapped the first-order interacting neighbors of the core disease hubs to clinically approved pharmacological agents. This approach revealed targetable nodes, including the endosomal receptor TLR7, the lysosomal protease CTSL, the apoptosis regulator BCL2, the plasma cell marker SLAMF7, and receptor tyrosine kinases such as AXL, FLT3, and EPHB2. Thus, our single-sample network framework provides a computational mechanism to uncover pathogenesis related modules and prioritize targeted drug repurposing candidates in SD.

# Systemic Integrative Analysis and Characterization of Interferon-Regulated Genes in Major Depressive Disorder

Amanda Rodriguez Valera<sup>1</sup>, Roseane Galdioli Nava<sup>2</sup>, Yohan Lucas Gonçalves Corrêa<sup>2</sup>, Adriel Leal Nóbile<sup>2</sup>, Ronaldy Santana Santos<sup>2</sup>, Júlia Nakanishi Usuda<sup>2</sup>, Anny Silva Adri<sup>2</sup>, Débora Gomes de Albuquerque Freitas<sup>3</sup>, Fernando Yuri Nery do Vale<sup>4</sup>, Luciana Bezerra de Souza Gianasi<sup>5</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>6</sup>, Igor Salerno Figueiras<sup>2</sup>, Haroldo Dutra Dias<sup>7</sup>, Otavio Cabral-Marques<sup>6</sup>

<sup>1</sup>Institute of Biomedical Sciences, University of São Paulo (ICB-USP), <sup>2</sup>Universidade de São Paulo, <sup>3</sup>Interunit Postgraduate Program on Bioinformatics, University of São Paulo (USP), <sup>4</sup>RCSI Centre for Systems Medicine and Dept. of Physiology & Medical Physics, RCSI University of Medicine and Health Sciences, Dublin, Ireland, <sup>5</sup>Federal University of Southern and Southeastern Pará (Unifesspa), Brazil. Digital Metropolis Institute, Multidisciplinary Bioinformatics Center, Federal University of Rio Grande do Norte (UFRN), <sup>6</sup>Universidade Federal do Rio Grande do Norte, <sup>7</sup>Harvard Medical School

**Background:** Major depressive disorder (MDD) is a prevalent and heterogeneous psychiatric disorder whose pathophysiology extends beyond the monoaminergic hypothesis. Increasing evidence supports the involvement of immune and inflammatory mechanisms, particularly interferons (IFNs) and interferon-regulated genes (IRGs), which may link peripheral immune activation to central nervous system dysfunction. However, most transcriptomic studies evaluate isolated biomarkers rather than integrated gene networks. **Objective:** To identify IFN-associated differentially expressed genes (DEGs) in patients with MDD through an integrative analysis of publicly available transcriptomic datasets. **Methods:** Five peripheral blood microarray datasets (GSE19738, GSE32280, GSE52790, GSE98793, and GSE39653) were retrieved from the Gene Expression Omnibus (GEO). After quality assessment and batch effect correction using ComBat, differential expression analysis was performed with the limma package in R. DEGs were defined as  $p < 0.05$  and  $|\log_{2}FC| > 0.1$ . Gene Ontology Biological Process enrichment was conducted using clusterProfiler, followed by classification into immune, nervous, and neuroimmune categories. BP-gene interaction networks were generated with Appyter, and DEGs were subsequently filtered against the Interferome database to identify IRGs. **Results:** The integrated analysis identified a neuroimmune molecular signature associated with MDD. Enriched biological processes included immune activation, antiviral response, cytokine signaling, leukocyte regulation, synaptic organization, neuronal homeostasis, and neuroimmune communication. Network analysis revealed extensive interactions between immune and neuronal pathways through shared DEGs. Key IFN-related genes, including IRF1, SOCS1, SOCS3, and IDO1, emerged as central components, suggesting that interferon-mediated immune activation and kynurenine metabolism may contribute to neuronal dysfunction in MDD. **Conclusion:** This integrative transcriptomic analysis supports the neuroimmune hypothesis of depression by identifying IFN-related genes potentially involved in MDD pathophysiology. Integration with the Interferome database and future bulk and single-cell RNA sequencing analyses may further define inflammatory molecular subgroups and identify biomarkers for precision medicine approaches in MDD.

# **Immune Dysregulation in Major Depressive Disorder and the Effects of Antidepressant Treatment: A Systematic Review and Meta-Analysis of Peripheral Cytokines**

Roseane Galdioli Nava<sup>1</sup>, Ronaldy Santana Santos<sup>2</sup>, Yohan Lucas Gonçalves Corrêa<sup>2</sup>, Adriel Leal Nóbile<sup>2</sup>, Júlia Nakanishi Usuda<sup>2</sup>, Anny Silva Adri<sup>2</sup>, Débora Gomes de Albuquerque Freitas<sup>1</sup>, Fernando Yuri Nery do Vale<sup>3</sup>, Amanda Rodriguez Valera<sup>4</sup>, Luciana Bezerra de Souza Gianasi<sup>5</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>6</sup>, Igor Salerno Filgueiras<sup>2</sup>, Haroldo Dutra Dias<sup>7</sup>, Otavio Cabral-Marques<sup>6</sup>

<sup>1</sup>Interunit Postgraduate Program on Bioinformatics, University of São Paulo (USP), <sup>2</sup>Universidade de São Paulo, <sup>3</sup>RCSI Centre for Systems Medicine and Dept. of Physiology & Medical Physics, RCSI University of Medicine and Health Sciences, Dublin, Ireland, <sup>4</sup>Institute of Biomedical Sciences, University of São Paulo (ICB-USP), <sup>5</sup>Federal University of Southern and Southeastern Pará (Unifesspa), Brazil. Digital Metropolis Institute, Multidisciplinary Bioinformatics Center, Federal University of Rio Grande do Norte (UFRN), <sup>6</sup>Universidade Federal do Rio Grande do Norte, <sup>7</sup>Harvard Medical School

Immune dysregulation is increasingly recognized as a biological feature of major depressive disorder (MDD), but whether antidepressant treatment normalizes peripheral cytokine alterations remains unclear. We conducted a systematic review and random-effects meta-analysis of clinical studies reporting circulating cytokine concentrations in adults with MDD. Meta-analyses evaluated baseline differences between patients with MDD and healthy controls (HC) and within-group changes before and after antidepressant treatment. Of 1,599 records identified, 27 studies met the inclusion criteria; 22 contributed to the MDD-versus-HC comparison and 24 to the pre-post-treatment analysis. At baseline, patients with MDD showed significantly higher concentrations of TNF- $\alpha$ , IL-1 $\beta$ , IL-6, IL-8, IL-10, and IL-13. Conversely, IFN- $\gamma$ , IL-2, IL-4, and IL-5 did not differ significantly between groups. Following antidepressant treatment, IL-4 concentrations decreased, whereas IL-5 concentrations increased. No statistically significant treatment-associated changes were observed for the remaining cytokines. Most pooled estimates exhibited substantial heterogeneity, indicating considerable variability across patient populations, clinical characteristics, treatment protocols, and analytical methods. Overall, MDD was associated with heterogeneous peripheral immune alterations involving both inflammatory and counter-regulatory responses. These alterations were not consistently reversed by standard antidepressant treatment, providing no evidence for a uniform anti-inflammatory effect. Future studies integrating multidimensional immune profiling with clinical and neurobiological measures are needed to identify reproducible neuroimmune signatures and support biological stratification in precision psychiatry.

# **In silico identification and characterization of potential effectors of *Colletotrichum gloeosporioides* in the interaction with *Mangifera indica***

Danilo dos Santos Lima<sup>1</sup>, Natasha dos Santos Lopes<sup>2</sup>, José Tarcísio Domingos Coutinho<sup>3</sup>, Raner José Santana Silva<sup>2</sup>, Ariana Silva Santos<sup>1</sup>

<sup>1</sup>UNIVERSIDADE FEDERAL VALE SSAO FRANCISCO, <sup>2</sup>Universidade Estadual de Santa Cruz, <sup>3</sup>Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campus Princesa Isabel, BR-426 - Zona Rural, Princesa Isabel, Paraíba, Brasil.

Anthrachnose, caused by the fungus *Colletotrichum gloeosporioides*, is one of the major diseases affecting mango trees (*Mangifera indica*), causing significant losses in fruit production and quality. Effector proteins play a central role in the fungus-host interaction. The objective of this study was to identify and characterize candidate *C. gloeosporioides* effector proteins by constructing an in silico database based on public transcriptomic data regarding the interaction with *M. indica*. To this end, the transcript data available on NCBI (BioProject PRJNA575336) were curated. The expressed transcripts were subjected to quality control (FastQC), filtering (Trimmomatic), and quantification (Salmon Quant), considering TPM  $\geq 1$ . The selected sequences were subjected to coding region prediction using TransDecoder, analyzed for the presence of a signal peptide using SignalP and the absence of transmembrane domains using DeepTMHMM, aiming to filter for proteins with the potential to be secreted. Subsequently, these proteins were subjected to effector prediction using EffectorP. The pipeline employed processed 2,433,700 reads from the *C. gloeosporioides*-*M. indica* interaction, which were aligned to the fungal reference genome (GCF\_011800055.1), resulting in the identification of 12,088 transcripts with TPM  $\geq 1$ . The proteins encoded by these transcripts were recovered and characterized regarding their secretory potential and function as effectors. Of the total predicted proteins, 5.44% possessed a signal peptide, and 66.36% of these were classified as secreted proteins. Among the secreted proteins, 67 (15.37%) were predicted to be effectors, with 48 (71.64%) classified as apoplastic and 19 (28.36%) as cytoplasmic, indicating a predominance of interaction mechanisms within the extracellular environment during infection. Considering the 12,088 identified transcripts, effectors accounted for approximately 0.55% of this set, demonstrating the pipeline's ability to progressively narrow down thousands of transcripts to a highly specific set of candidate proteins associated with pathogenicity. Currently, the database compiles 67 candidate effector proteins including information on expression, binding, and site of action serving as a resource for studies on virulence and fungus-host interactions. Upon completion of functional annotations, it will be made available in a public repository for use in bioinformatics, plant pathology, and biotechnology.

# **Integrative transcriptomic analysis reveals distinct molecular signatures across clinical stages of liver disease and hepatocellular carcinoma**

Laryssa Bandeira de Melo Silva<sup>1</sup>, Matheus Azevedo Bomfim<sup>1</sup>, Beatriz Souza Toscano de Melo<sup>2</sup>, João Luiz de Lemos Padilha Pitta<sup>2</sup>, João Pacifico Bezerra Neto<sup>3</sup>, Túlio de Lima Campos<sup>1</sup>, Luydson Richardson Silva Vasconcelos<sup>4</sup>

<sup>1</sup>Instituto Aggeu Magalhães - FIOCRUZ, <sup>2</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, <sup>3</sup>University of Pernambuco, Recife, Pernambuco, Brazil, <sup>4</sup>Instituto Aggeu Magalhães, Fiocruz Pernambuco, Fundação Oswaldo Cruz, Fiocruz, Recife, Pernambuco, Brazil.

Hepatocellular carcinoma (HCC) is the most common primary malignant tumor of the liver and is generally associated with chronic liver diseases, including chronic hepatitis and cirrhosis. Despite advances in molecular characterization, transcriptomic alterations across different clinical conditions remain poorly understood. This study aimed to characterize molecular signatures across liver disease groups through an integrative transcriptomic approach. Public RNA-seq data from the GSE114564 dataset, comprising five samples per group (control, chronic hepatitis, cirrhosis, early-stage HCC, and advanced-stage HCC), were analyzed. Following quality assessment (FastQC), genome alignment (GRCh38) via HISAT2, and gene quantification with featureCounts, differential expression analyses were performed using edgeR ( $FDR < 0.05$ ;  $|\log_2FC| > 1$ ), followed by Gene Ontology enrichment analysis (clusterProfiler) and weighted gene co-expression network analysis (WGCNA). The comparison between cirrhosis and early-stage HCC yielded the highest number of differentially expressed genes (1,667), whereas other comparisons showed less pronounced alterations. Among the key differentially expressed genes, AKR1B10 and SPINK1 showed higher expression in chronic hepatitis and cirrhosis, indicating molecular features associated with hepatocarcinogenesis, while AFP and PEG10 were highlighted in advanced-stage HCC, reinforcing their association with more aggressive tumor phenotypes. Functional enrichment analysis revealed immune response-related processes in comparisons involving chronic hepatitis, and processes associated with the cell cycle and lymphocyte activation in the comparison between cirrhosis and early-stage HCC. WGCNA identified module ME7 as the most strongly associated with clinical groups ( $r = 0.79$ ;  $p < 0.001$ ), enriched for cell cycle-related processes. In this module, ANLN, TOP2A, and KIF11 were identified as hub genes and differentially expressed genes, exhibiting high intramodular connectivity and elevated expression in HCC samples, particularly in advanced-stage HCC samples. These findings demonstrate distinct transcriptomic signatures among the clinical groups evaluated, highlighting immune response and cell cycle regulation as potential mechanisms involved in hepatocarcinogenesis, while also providing novel insights into molecular mechanisms driving HCC progression and supporting the discovery of clinically relevant biomarkers for future functional validation studies.

# **Bibliometric analysis of the use of Bioinformatics in Basic Education**

Iris Eduarda da Silva Caetano.<sup>1</sup>, Micaelly Vitória Lima Dos Santos<sup>2</sup>, José Tarcísio Domingos Coutinho<sup>3</sup>, Ariana Silva Santos<sup>1</sup>

<sup>1</sup>UNIVERSIDADE FEDERAL VALE SSAO FRANCISCO, <sup>2</sup>UNIVERSIDADE FEDERAL DO VALE SÃO FRANCISCO,

<sup>3</sup>Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campus Princesa Isabel, BR-426 - Zona Rural, Princesa Isabel, Paraíba, Brasil.

Bioinformatics is expanding its scope beyond scientific research and emerging as an educational strategy to integrate biological and computational knowledge, fostering active learning. Despite this potential, analyses characterizing its integration into Basic Education remain scarce. This study aimed to analyze the bibliometric landscape of scientific production on Bioinformatics in Basic Education. Searches were conducted in the PubMed, Scopus, WOS, and ERIC databases using structured search strategies for bioinformatics and basic education. After removing duplicates and applying eligibility criteria using StArt software, the included studies underwent bibliometric analysis in the R environment (Bibliometrix) and VOSviewer. The 483 studies identified, 29 were deemed eligible for the bibliometric analysis. Published between 2006 and 2026, these works employed bioinformatics as a classroom teaching tool, indicating that its use in basic education remains an emerging field. An increase in publications was observed primarily from 2022 onwards, with the highest numbers recorded in 2023 and 2025. Research output was concentrated mainly in the USA (41.38%) and Brazil (24.14%). Although Brazil ranked second in terms of publications, its output was dispersed across various institutions, without the consolidation of a central research hub. Collaboration networks revealed low levels of internationalization and limited cooperation among authors, institutions, and countries, indicating that the subject matter remains concentrated within a few research groups. Co-occurrence analysis revealed three central axes of investigation—teaching and learning; bioinformatics/computational biology; and genetics, molecular biology, and STEM education—highlighting that bioinformatics has been predominantly employed as a pedagogical strategy for teaching biological subject matter. The studies were published in high-impact international journals (IF: 1.0–8.2), demonstrating the recognition of the interface between bioinformatics and basic education by top-tier journals. Although scientific output remains limited, its publication in high-impact international journals and the leading roles played by the USA and Brazil highlight the growing recognition of bioinformatics as a strategy for science education. However, the low level of international collaboration reveals opportunities to strengthen research networks and expand studies across different educational contexts.

# NetPerturb: A Scalable Nextflow Pipeline for Gene Network Inference and Perturbation in scRNA-seq

Leonardo Gabriel de Queiroz Amorim<sup>1</sup>, Júlia A. de Amorim<sup>1</sup>, João V. F. Cavalcante<sup>1</sup>,  
Morais, B. P. R.<sup>1</sup>, Diego Marques Coelho<sup>1</sup>

<sup>1</sup>UFRN

Gene network inference and in silico perturbation analysis from standard single-cell RNA-sequencing (scRNA-seq) data are useful for understanding molecular dynamics and cellular states in complex biological systems. However, the increasing volume of experimental data imposes significant scalability challenges on existing computational tools. To address these limitations, we present NetPerturb, a scalable, efficient, and reproducible Nextflow pipeline that generates scRNAseq-based networks and prioritizes cell types based on target in-silico perturbation. NetPerturb automates gene network inference and calculates single-cell perturbation scores using the scRank algorithm, which infers network perturbation by constructing gene networks and scores from a sum of how far the target gene and its nearby neighbours shift between the original and the perturbed networks. Its workflow integrates: statistical sub-sampling of cell identities to optimize computational load; network inference centered natively on scRank (or with GENIE3 and scTenifoldNet available as alternative methods); and the quantification of specific target gene scores across defined cellular states. Finally, the pipeline merges these parallelized analyses into a unified output, generating tables that detail perturbation scores for each cell type relative to the analyzed target. Downstream biological interpretation was created using Quarto-based (.qmd) interactive dashboard directly into the Nextflow execution. This dashboard allows users to explore the results through data tables, regulatory subnetworks, expression heatmaps, and functional enrichment analyses. For validation of the pipeline's performance and reproducibility across different biological contexts, we applied NetPerturb to four benchmark scRNA-seq datasets ranging from mouse organoids and bone marrow cells to human cancer models. Our pipeline successfully reproduced the biological findings from the original publications and expanded upon them by delivering an integrated functional analysis. Ultimately, NetPerturb provides a user-friendly framework that accelerates target prioritization, enabling systemic, large-scale analysis of gene perturbations across complex biological systems.

# Alternative Splicing in Colorectal Cancer: Mapping Public Neoantigen Candidates for Precision Vaccine Development.

Helenize Catarina Moreira Costalat<sup>1</sup>, Pedro Lucas Alves da Silva<sup>1</sup>, Gleison Medeiros de Azevedo<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>2</sup>, Ândrea Ribeiro-dos-Santos<sup>1</sup>, André Luis Fonseca Faustino<sup>2</sup>

<sup>1</sup>Universidade Federal do Pará, <sup>2</sup>Universidade Federal do Rio Grande do Norte

Neoantigens derived from alternative splicing are promising targets for immunotherapies and vaccines in colorectal cancer (CRC). Since tumor-specific, these aberrant isoforms may be shared among patients, an advantage over conventional neoantigens, which arise from private somatic mutations restricted to a single individual, enabling broad-spectrum therapies. This study tested the hypothesis that alternative splicing generates shared neoantigen candidates by mapping common profiles across cohorts. Non-canonical splicing was analyzed with IsoformSwitchAnalyzeR on 66 RNA-seq samples (GEO), selecting putative unique isoforms, while OptiType performed HLA genotyping. Presentation and immunogenic potential of 9-mer peptides from these isoforms ( $p < 0.05$ ,  $TPM > 1$ ) were predicted using a combined score, defined as the mean of the MHCflurry presentation score and the DeepImmuno score. Transcriptional landscape characterization revealed a conserved core of ~3,000 isoforms shared across all groups, alongside cohort-specific subsets largest in the Asian (ASN) cohort (~1,400 unique isoforms). Alternative splicing event mapping showed the ASN cohort depleted in alternative termination sites and enriched in intron retention, opposite to the European/North American (EUR/EUA) cohort, enriched in alternative transcription start sites. We also identified 93 lncRNAs potentially undergoing non-canonical translation in the CRC microenvironment, suggesting non-coding RNAs as an underexplored reservoir of antigenic diversity. Unique isoforms of coding genes were evaluated for affinity and immunogenicity against prevalent HLA alleles (HLA-A02:01, HLA-B51:01, HLA-C\*07:02), yielding a combined score per peptide-allele pair. Peptides above the third quartile (score  $< 0.75$ ) were ranked by allelic promiscuity, generating a panel of 20 neoantigen candidates. Unique isoforms of UCKL1, ITGB7, and HEXDC stood out for broader allelic recognition, with mean scores of 0.9419, 0.9809, and 0.9714, respectively. Notably, these three candidates are shared across all three population groups (ASN, EUR, EUA), reflecting broad allelic recognition across the HLA alleles most frequent in these populations. These results highlight alternative splicing as a strategic source of shared neoantigens, with translational potential for broad-spectrum therapeutic vaccines in CRC. Future work will expand predictions to a broader range of HLA alleles, ensuring greater global coverage and population equity.

# Comparative Genome Mining Reveals Conserved and Species-Associated Biosynthetic Gene Cluster Repertoires in Petroleum-Associated *Bacillus* Strains.

Jéssica Maria da Silva Pinheiro<sup>1</sup>, Clara Detúlio Martins Cavalcante<sup>2</sup>, Júlia Firme Freitas<sup>2</sup>, Lucymara Fassarela Agnez-Lima<sup>2</sup>, Renata Mendonça Araújo<sup>2</sup>

<sup>1</sup>Universidade federal do Rio Grande do Norte, <sup>2</sup>Universidade Federal do Rio Grande do Norte

The rapid emergence of antimicrobial resistance (AMR) has intensified the search for new antimicrobial compounds. Natural products from microbial sources represent a valuable reservoir of chemically diverse molecules. Comparative genome mining enables the identification of microbial biosynthetic potential through biosynthetic gene cluster (BGC) analysis. This study compared the biosynthetic potential of four petroleum-associated *Bacillus* strains from two closely related species, *Bacillus amyloliquefaciens* (BG1 and BD162) and *Bacillus paralicheniformis* (BD129 and BD165), to identify conserved and species-associated biosynthetic pathways with potential relevance for antimicrobial natural product discovery. BGCs were predicted using antiSMASH v8, grouped into Gene Cluster Families (GCFs) using BiG-SCAPE (cutoff 0.5), visualized in Cytoscape, and compared against the MIBiG database through the KnownClusterBlast module. A total of 66 BGCs were organized into 26 GCFs. Four GCFs were conserved across all four strains. Two terpene-associated GCFs showed no significant KnownClusterBlast hit, whereas a T3PKS-associated GCF showed 5% similarity to the closest MIBiG reference. These findings suggest conserved but poorly characterized biosynthetic potential. The remaining conserved GCF included BGCs with significant similarity to the bacillibactin biosynthetic cluster. Comparative analysis of the GCF repertoires of *B. amyloliquefaciens* and *B. paralicheniformis* revealed predominantly distinct biosynthetic repertoires between the two species. In *B. amyloliquefaciens*, species-associated GCFs were linked to surfactin, bacillaene, macrolactin H, mersacidin, bacilysin and diffidin, as well as to an uncharacterized phosphonate-associated GCF. In contrast, *B. paralicheniformis* was characterized by GCFs associated with bacitracin, lichenysin, pulcherriminic acid and schizokinen. Fengycin-related GCFs showed a complex distribution, occurring as species-associated GCFs and as GCFs shared among three of the four strains. One of these GCFs grouped BGCs annotated as both fengycin and plipastatin by KnownClusterBlast, reflecting differences between similarity-based clustering and product annotation. These findings demonstrate that integrating antiSMASH and BiG-SCAPE reveals conserved and species-associated biosynthetic potential in petroleum-associated *Bacillus* strains, providing a framework to prioritize biosynthetic pathways for metabolomic validation.

# Multi-Omics Integration Reveals Distinct Factors Related to Kidney Cancer Progression, Immune Activity and Patient Survival

Epitácio Dantas de Farias Filho<sup>1</sup>, Juliana De Souza Felix<sup>2</sup>, Iara Dantas de Souza<sup>3</sup>, Alexandre R. Paschoal<sup>4</sup>, Andre Luis Fonseca Faustino<sup>5</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>1</sup>

<sup>1</sup>Universidade Federal do Rio Grande do Norte, <sup>2</sup>Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, <sup>3</sup>Ohio State University, <sup>4</sup>Universidade Tecnológica Federal do Paraná, <sup>5</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Instituto Metrópole Digital (IMD); Federal University of Rio Grande do Norte (UFRN)

Clear Cell Renal Cell Carcinoma (ccRCC) is the most common subtype of renal cell carcinoma, accounting for nearly 90% of RCC cases. Although several molecular alterations, including chromosome 3 deletion, VHL, PBRM1 and BAP1 mutations, and dysregulation of regulatory elements have been described in ccRCC, single-omics studies cannot fully explain its molecular heterogeneity. Integrative multi-omics approaches have therefore become increasingly relevant for characterising coordinated molecular events and their effects on phenotypes. This study aimed to identify integrated molecular factors and investigate their biological processes and association with patient survival in ccRCC using a multi-omics approach. We applied the Multi-Omics Factor Analysis (MOFA) framework based on factorial analysis and Bayesian inference to integrate multimodal data (genomics, transcriptomics, epigenomics, and proteomics) from the TCGA-KIRC (N = 249) project. As a result, we identified 15 integrated factors that account for the largest proportion of variability across the samples. In particular, Factor 1 alone explained approximately 48% of the total variability, largely defined by the transcriptomics, epigenomics, and proteomics data. We also observed that Factor 2 is mainly associated with the epigenomic data, whereas Factor 3 was predominantly driven by transcriptomic data. Factors 1 and 2 were significantly associated with disease stage and patient outcome (adjusted p-value <  $1 \times 10^{-4}$ ). Each factor showed significant enrichment (FDR < 0.05) for specific biological activities related to T-cell differentiation, lymphocyte differentiation, renal cell development, the NF-kappa-B signalling pathway, glycine/pyrimidine/pyruvate metabolism, extracellular matrix alterations, and fibrinolysis. Within Factor 1, canonical ccRCC molecular features were identified, such as BAP1 and PBRM1, as well as new features, including PAI-1 overabundance, 7q31.1 amplification, cg0491556 hypomethylation, and hsa-miR-204 downregulation. In conclusion, the identification of novel candidate markers demonstrates how distinct molecular alterations can collectively modulate cancer progression and poor outcome.

# **Psychoneuroimmunology of Viral Hepatitis and Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis**

Adriel Leal Nóbile<sup>1</sup>, Roseane Galdioli Nava<sup>1</sup>, Ronaldy Santana Santos<sup>1</sup>, Júlia Nakanishi Usuda<sup>1</sup>, Anny Silva Adri<sup>1</sup>, Yohan Lucas Gonçalves Corrêa<sup>1</sup>, Lena F Schimke Marques<sup>1</sup>, Igor Salerno Filgueiras<sup>1</sup>, Haroldo Dutra Dias<sup>2</sup>, Rodrigo Juliani Siqueira Dalmolin<sup>3</sup>, Otavio Cabral-Marques<sup>3</sup>

<sup>1</sup>Universidade de São Paulo, <sup>2</sup>Harvard Medical School, <sup>3</sup>Universidade Federal do Rio Grande do Norte

Hepatocellular carcinoma (HCC) and viral hepatitis (VH) are recognized primarily as hepatic diseases, yet they impose a substantial neuropsychiatric disorders (NPDs) encompassing depression, anxiety, cognitive impairment, and fatigue. Whether these manifestations reflect psychological reactions or biologically grounded psychoneuroimmune dysfunction remains poorly understood. Here, we present a pre-registered systematic review and meta-analysis (PROSPERO 2026 CRD420251081757) synthesizing evidence from 24 studies combining psychosocial outcomes with immune dysregulation markers across HCC and VH populations. Bibliometric and network analyses identified a growing psychoneuroimmunology research hub linking cytokine signaling, microglial activation, blood-brain barrier (BBB) disruption, and gut-liver-brain axis dysregulation to the neuropsychiatric phenotype of these diseases. Meta-analysis revealed a pooled hazard ratio (HR) of 1.92 (95% CI: 1.09–3.39) for adverse outcomes in HCC and a pooled HR of 2.28 (95% CI: 1.48–3.52) for depression associated to genetic variants in VH. TNF- $\alpha$ , but not IL-6, showed a tendency toward lower levels in VH patients with versus without NPDs, suggesting differential inflammatory roles. The literature, however, remains fragmented by heterogeneous designs and a paucity of longitudinal data. A psychoneuroimmunological approach, integrating immune, endocrine, neural, and psychosocial dimensions, is essential for capturing the full systemic impact of HCC and VH and should inform future mechanistic research and multidisciplinary clinical strategies.

# Meta-analysis of public proteomic datasets for the identification of protein signatures and targetable antigens

Tayna da Silva Fiuza<sup>1</sup>, Karla Cristina Tabosa Machado<sup>2</sup>, Andre Luis Fonseca Faustino<sup>3</sup>, Sandro Jose de Souza<sup>4</sup>, Gustavo Antonio de Souza<sup>5</sup>

<sup>1</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Federal University of Rio Grande do Norte (UFRN); Bioinformatics Core Facility (NBI), Aggeu Magalhães Institute (IAM), Fiocruz Pernambuco, Oswaldo Cruz Foundation, Fiocruz, <sup>2</sup>Federal University of Paraiba (CCHSA UFPB), <sup>3</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Instituto Metrópole Digital (IMD); Federal University of Rio Grande do Norte (UFRN), <sup>4</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Federal University of Rio Grande do Norte (UFRN); DNA GTx Bioinformatics, <sup>5</sup>Bioinformatics Multidisciplinary Environment (BioME UFRN); Department of Biochemistry (DBQ/CB UFRN); Federal University of Rio Grande do Norte (UFRN); DNA GTx Bioinformatic

Integrating mass spectrometry-based (MS) proteomics data has emerged as a powerful approach for identifying biomarkers or targetable antigens across multiple tumor types. Targeted and immune therapies may improve cancer outcomes, particularly if a minimal set of markers combines broad multi-cancer applicability, tumor specificity, and high antigenicity. Here, we sought to identify multi-cancer marker candidates and multi-binding antigens by integrating 15 public MS-based proteomic datasets from eight cancer types: prostate, breast, ovary, lung, colon, kidney, lymphoma, melanoma, and healthy tissues. Protein abundances were log2-transformed, imputed, normalized and classified into tiers. Cancer-exclusive transmembrane proteins were subject to class I (27 alleles) and class II (26 alleles) HLA binding affinity and B-cell linear epitope prediction. Next, we applied a minimal-set approach to identify the smallest antigen sets maximizing either HLA-allele diversity and B-cell linear epitope count. Of 17,200 identified proteins, 2,047 were cancer-enriched (high/medium in cancer, low in healthy) and 890 were cancer-exclusive (high/medium in cancer, absent in healthy), organized in 191 and 65 tissue signatures, respectively. Of the 209 transmembrane proteins in the cancer-exclusive group, 193 contained at least one antigenic epitope. The minimal-set was applied to all eligible signature sets. For instance, an 11-protein signature overabundant across all cancer types containing five transmembrane antigenic-predicted proteins went through the minimal-set approach at protein and peptide level; the selection prioritized proteins TYR and TYRP1 in both cases. These proteins were predicted to bind to 27 HLA I- and 18 HLA II- alleles and 21 B-cell linear epitopes. Peptide-level prioritization further reduced it to 10 segments (eight from TYRP1 and two from TYR) while preserving HLA diversity with two B-cell epitopes containing only 37.6% and 7.56% of the original proteins. Our data demonstrate the potential to use a meta-analysis of public proteomics datasets in order to elucidate candidate antigenic biomarkers that are conserved and overabundant in multiple cancer types and, consequently, are excellent targets for further immunogenic research.

# Author Index

## A

### Adri

Anny Silva, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)

### Agnez-Lima

Lucymara Fassarella, [14](#), [100](#)

### Aguiar

Eric Roberto Guimaraes Rocha, [10](#)

Eric Roberto Guimarães Rocha, [78](#)

### Almeida

Mariana Cordeiro, [81](#)

Rita M. C. de, [45](#)

### Alvarenga

Luana Ferraz, [65](#)

### Alves

Guilherme Bastos, [59](#)

Robinson Luis de Souza, [27](#)

### Amorim

Júlia A. de, [41](#), [44](#), [52](#), [98](#)

Leonardo Gabriel de Queiroz, [98](#)

### Andrade

Caio Maia Farias de, [37](#), [86](#)

### Aquino

Jonas Galileu Ferreira de, [59](#)

### Araújo

Gabriel Freitas, [7](#), [83](#)

Gilderlanio Santana de, [69](#), [82](#)

Raquel Silveira de, [42](#)

Renata Maria de, [59](#)

Renata Mendonça, [100](#)

### Arruda

Pedro Melo, [24](#)

Rodrigo Olímpio Orvate, [16](#), [30](#), [34](#), [56](#)

### Aydilek

Enver, [79](#)

### Azevedo

Arthur Henrique Dantas de, [62](#)

Gleison Medeiros de, [33](#), [70](#), [90](#), [99](#)

## B

### Balbino

Valdir de Queiroz, [33](#)

### Barbosa

Anderson Nogueira, [67](#)

Raquel De M., [88](#)

Raquel de Melo, [28](#), [68](#), [73](#)

### Barreto

Marina dos Santos, [39](#)

### Barros

Agnes Angélica Guedes de, [42](#)

José Isaías Pimentel, [50](#), [66](#)

Marcelo Soares de, [47](#)

Maria Clara, [17](#)

Maria Clara da Costa, [22](#)

### Batista

Nathalia Silva, [66](#)

### Benko-Iseppon

Ana Maria, [57](#)

### Berg

Cassio van den, [10](#)

### Bezerra

Katyanna Sales, [29](#)

### Bomfim

Matheus Azevedo, [7](#), [63](#), [96](#)

### Borges

Clayton Luiz, [35](#)

Maria Bárbara, [38](#)

### Bradaschia

Laura Shimohara, [62](#)

### Brunetto

André T., [87](#)

### Brökers

Nils, [79](#)

### borja

gabrielle de vasconcelos, [55](#)

## C

### Cabral

Serli de Oliveira, [80](#)

### Cabral-Marques

Otavio, [39](#), [46](#), [48](#), [93](#), [94](#), [102](#)

### Caetano.

Iris Eduarda da Silva, [21](#), [97](#)

### Campos

Marcos Davi de Sousa, [50](#), [66](#)

Túlio de Lima, [7](#), [63](#), [83](#), [96](#)

### Carmo

Rodrigo Feliciano do, [7](#)

### Cavalcante

Clara Detúlio Martins, [100](#)

João V. F., [98](#)

Rayssa Beatriz Ribeiro, [31](#), [60](#)

### Cavalcanti

Euller Barbosa Tenório, [42](#)

João Vitor, [25](#), [33](#), [70](#), [92](#)

### Ceneviva

Bruno Mendes, [81](#)

### Cezar

Jacinta Fernanda Novais, [70](#)

### Coelho

Danilo Fernandes, [80](#)

Diego Marques, [25](#), [41](#), [52](#), [71](#), [92](#), [98](#)

Marina de Lourdes Ribeiro Alves, [9](#), [84](#)

### Conceição

Lucas Barbosa de Amorim, [10](#)

### Corrêa

Yohan Lucas Gonçalves, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)

### Costa

Anna Letícia Silva da, [50](#), [66](#)

Isaac Medeiros do Amaral, [25](#)

João Vitor Almeida da, [75](#)

Leticia de Aguiar da, [22](#)

Marcos Romualdo, [41](#), [44](#), [52](#)

### Costalat

Helenize Catarina Moreira, [8](#), [99](#)

### Couceiro

Douglas de Moraes, [67](#)

### Coutinho

José Tarcísio Domingos, [19](#), [95](#), [97](#)

Vinícius Maracajá, [38](#)

### Câmara

Gabriel Bezerra Motta, [73](#)

## D

### Dall’Agnol

Hivana Patricia Melo Barbosa, [50](#), [66](#)

Leonardo Teixeira, [50](#), [66](#)

### Dalmolin

Matheus, [28](#), [68](#), [85](#), [88](#)

Rodrigo Juliani Siqueira, [20](#), [33](#), [37](#), [39](#), [40](#), [45](#), [46](#),

[48](#), [70](#), [75](#), [86](#), [90](#), [92](#), [93](#), [94](#), [99](#), [101](#), [102](#)

### Dantas

Alana de Medeiros, [59](#)

### Dias

Haroldo Dutra, [93](#), [94](#), [102](#)

Mariana Costa, [58](#)

### Duarte

João Ricardo Guerreiro, [23](#)

## E

### Elzegbe

Pierina Sol, [64](#)

## F

### FERREIRA

LEONARDO CAPISTRANO, [51](#), [89](#), [91](#)

### Faria

Joana Menezes de, [89](#)

### Farias

Sávio Torres de, [6](#)

### Faustino

Andre Luis Fonseca, [8](#), [69](#), [74](#), [101](#), [103](#)

André Luis Fonseca, [99](#)

### Felix

Juliana De Souza, [81](#), [101](#)

Juliana Rayssa Barros, [15](#)

Leandro Amaro da Silva, [78](#)

### Fernandes

Layana Cristine Carvalho, [82](#)

Leonardo Luiz de Sena, [76](#)

Marcelo A. C., [28](#), [68](#), [73](#), [85](#), [88](#)

### Fernandes-Pedrosa

Matheus de Freitas, [73](#)

### Ferreira

Caio Gabriel Tavares, [33](#)

Lucas Yago Melo, [78](#)

### Filgueiras

Igor Salerno, [18](#), [39](#), [46](#), [48](#), [93](#), [94](#), [102](#)

### Filho

Edson de Carvalho Ferreira, [24](#)

Epitácio Dantas de Farias, [40](#), [61](#), [101](#)

Francisco Sales de Lima, [27](#)

Julio Cesar da Silva, [32](#)

Júlio César da Silva, [17](#), [90](#)

Romério de Oliveira Lima, [37](#), [86](#)

### Fiuza

Tayna da Silva, [103](#)

### Fontinele

Djorkaeff Oliveira, [30](#), [34](#)

### França

João Vitor Santos de, [26](#)

### Freitas

Débora Gomes de Albuquerque, [39](#), [46](#), [93](#), [94](#)

Júlia Firme, [100](#)

Kianni Hernandez, [46](#)

### Fulco

Umberto Laino, [29](#), [59](#)

### Furtado

Allanny Alves, [73](#)

Lorenzo Santos, [71](#)

## G

### Galeno

Rebeca Costa dos Reis, [82](#)

### Galon

Jérôme, [46](#)

### Galvão

Maurício Matheus Araújo Silva, [44](#)

### Gianasi

Luciana Bezerra de Souza, [39](#), [93](#), [94](#)

### Gomes

Daniel Henrique Ferreira, [17](#), [32](#), [90](#)

Jaqueline Ferreira, [67](#)

### Gonçalves

Laís de Carvalho, [30](#), [34](#)

### Goto

Bruno Tomio, [15](#)

### Gouvea

Felipe, [22](#)

### Guerreiro

Sávio Lucas de Matos, [90](#)

### Gupta

Shantanu, [30](#), [34](#), [40](#), [87](#)

### Gysi

Deisy Morselli, [76](#)

## H

### Hoff

Mariana Leivas Muller, [78](#)

## I

### Iseppon

Ana Maria Benko, [42](#)

## J

### J.O.

Pegado, F., [36](#)

### J.R.P.

Silva,, [36](#)

### Joutsen

Tânia Keiko Shishido, [50](#), [66](#)

### Junior

Wilson Araújo da Silva, [43](#)

### Justino

Caio, [65](#)

### Júnior

Sérgio Dantas de Oliveira, [67](#)

## K

### Katak

Ricardo de Melo, [47](#)

## L

### LIMA

EMILLY YNIALIDE RODRIGUES DE, [49](#)

### Lima

Carlos Eugênio Costa de, [43](#)

Danilo dos Santos, [21](#), [95](#)

Giovanna Santos de, [57](#)

João Paulo Matos Santos, [15](#), [26](#), [30](#), [56](#), [62](#)

Luísa Tannus, [52](#)

Marx Oliveira de, [58](#), [67](#)

Zayra Jasminne Dias de, [59](#)

### Lopes

Flávia Lombardi, [81](#)

Natasha dos Santos, [95](#)

### Lucena

Jean Paes Landim de, [77](#)

### Luchiari

Ana Carolina, [37](#)

### Luna-Aragão

Madson Allan de, [57](#)

## M

### M.

Ortega, J., [36](#)

### Machado

Karla Cristina Tabosa, [103](#)

### Maia

Caio César, [38](#)

Rafael Trindade, [9](#), [84](#)

### Marques

Hannah Isabele Da Silva, [88](#)

Lena F Schimke, [102](#)

Otávio Cabral, [18](#)

### Martins

Letícia Maria Azevedo, [59](#)

### Matos

Gustavo Barra, [69](#)

### Melo

Beatriz Souza Toscano de, [7](#), [47](#), [63](#), [96](#)

### Michaelson

Gustavo Lovatto, [75](#)

### Moioli

Renan, [65](#)

### Mombach

José Carlos Merino, [87](#)

### Monteiro

Guilherme Matheus Chagas Ribeiro, [12](#)

### Morais

Gabriel Christian de Farias, [59](#)

### Moreira

Andre, [65](#)

Laura de Sousa Lopes, [35](#)

Thalison R, [35](#)

### Mouawad

Michel, [46](#)

### Moura

Marcelo Tigre, [38](#)

## N

### NASCIMENTO

DANIEL JERÔNIMO DO, [51](#), [89](#), [91](#)

### Nascimento

Lavinia Santos Prates do, [10](#)

Thiago Lima do, [19](#), [21](#)

### Nava

Roseane Galdioli, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)

### Neto

João Pacifico Bezerra, [96](#)

### Nobrega

João Anisio Marinho da, [13](#)

### Nunes

Cínthia Carla Claudino Grangeiro, [42](#)

### Nóbile

Adriel Leal, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)

## O

### Oliveira

Camila Franco Batista de, [19](#)

Daniel Pacheco Penha de, [59](#)

Jonas Ivan Nobre, [29](#), [59](#)

Mariana Timbó de, [77](#)

Moisés Abraão Sousa de, [8](#)

Pablo Rafael Silveira, [7](#)

Ricardo Cunha de, [23](#)

## P

### Paiva

Ayron Luigi de, [11](#)

Thales Henrique de, [11](#)

### Palheta

Helber Gonzales Almeida, [82](#)

### Pantoja

Laíze Barros, [8](#), [74](#)

### Parente

Adriana Marina e Silva, [73](#)

### Parente-Rocha

Juliana Alves, [35](#)

### Paschoal

Alexandre R., [101](#)

### Patrício

Lucas de Jesus, [67](#)

### Peixoto

Maria Eduarda Dourado, [35](#)

### Penna

Saulo Rafael Mendes, [42](#)

### Peres

Andre, [65](#)

André Salles Cunha, [27](#), [31](#), [60](#)

### Pinangé

Diego Sotero de Barros, [58](#)

### Pinheiro

Jéssica Maria da Silva, [100](#)

### Pinto

Pablo Diego do Carmo, [23](#)

### Pitta

João Luiz de Lemos Padilha, [7](#), [47](#), [63](#), [96](#)

### Pontes

Vinícius Pereira de, [68](#)

### Porcino

Naiara Jesus, [21](#)

### pereira

jeniffer gabrielly de sousa, [9](#), [84](#)

## Q

### Quagliato

Emmanuelle Nascimento, [46](#)

### Quaresma

Juarez, [22](#)

### Queiroz

Lucas Araújo de, [36](#)

## R

### R.

Morais, B. P., [41](#), [98](#)

### Ramos

Emanuelle Costa, [50](#), [66](#)  
Felipe Assunção Félix, [53](#)  
Rodrigo Nalio, [46](#)

### Reiniger

Klaus Augustus Ramos, [29](#)

### Rennó-Costa

César, [31](#), [60](#), [65](#)

### Ribeiro-dos-Santos

Ândrea, [17](#), [22](#), [99](#)

### Rikkinen

Jouko, [50](#), [66](#)

### Rocha

Esther Maria Pacheco, [33](#)

### Rodrigues

João Pedro Lemos e Silva, [16](#), [34](#)

### Romão

Tatiany Patrícia, [47](#)

### Roncalli

Beatriz Godoy, [47](#)

### Roque

Rosemary Aparecida, [47](#)

### Ruiz

Orville Hernandez, [35](#)

### Rêgo

Thaís Gaudencio do, [38](#)

### Róbias

Pedro Henrique Álvares, [31](#), [60](#)

## S

### Sakamoto

Tetsu, [24](#), [72](#)

### Santiago

Bianca Cristiane Ferreira, [70](#)

### SantoS

João Pedro Nunes, [78](#)

### Santos

André Ribeiro dos, [22](#)  
Ariana Silva, [19](#), [21](#), [95](#), [97](#)  
Arthur Ribeiro dos, [20](#)  
Denys Ewerton da Silva, [57](#)  
Gabrielle Germek Coelho, [17](#), [32](#)  
Hellen Karla Vital dos, [21](#)  
Késia de Carvalho, [51](#)  
Micaelly Vitória Lima Dos, [21](#), [97](#)  
Odilon Julio dos, [45](#)  
Rogério Mercês Ferreira, [10](#)  
Ronaldy Santana, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)  
Sidney Emanuel Batista Dos, [17](#)

### Sarges

Weneson Victor Diniz, [50](#), [66](#)

### Schlüter

Jakob, [42](#)

### Schrank

Augusto, [33](#)

### Schwarz

Friedrich, [79](#)

### Silva

Ana Vivia dos Santos Chacon da, [56](#), [64](#)  
Breno de Mello, [11](#)  
Douglas Felipe de Lima, [14](#)  
Douglas Nunes da, [40](#)  
Felipe Gomes da, [10](#)  
Gabriel Vinícius Rolim, [59](#)  
Gilvan Ferreira da, [67](#)  
João Paulo Matos Lima, [16](#), [29](#), [64](#)  
Larissa Martins Brito e, [90](#)  
Laryssa Bandeira de Melo, [7](#), [63](#), [96](#)  
Lucas Brito da, [82](#)  
Maria Karolayne da, [59](#)  
Meryck Felipe Brito da, [49](#)  
Pedro Lucas Alves da, [8](#), [74](#), [99](#)  
Rafael Lucas da, [42](#)  
Raner José Santana, [95](#)  
Raylam Wendel da, [61](#), [68](#), [85](#)

### Silveira

Daner Acunha, [87](#)

### Sinigaglia

Marialva, [87](#)

### Soares

Célia Maria de Almeida, [35](#)  
Giovanna A. P., [73](#), [85](#), [88](#)

Matheus Fujimura, [81](#)

### **Sousa**

Marcos Antonio Nobrega de, [6](#), [9](#), [84](#)

Sabrina Sinara Portela de, [67](#)

Vitor Quintela, [54](#)

### **Souto**

Rafael, [49](#)

### **Souza**

Alyson Matheus Carvalho de, [13](#)

Gustavo Antonio de, [103](#)

Gustavo Antônio de, [62](#)

Iara Dantas de, [101](#)

Jorge Estefano Santana de, [17](#), [32](#), [90](#)

Jorge Estefano de Santana, [74](#)

Marcella Vitória Belém, [69](#)

Maria Clara Santos, [10](#)

Nycolas Henrique Moraes de, [80](#)

Sandro Jose de, [20](#), [103](#)

### **Spelta**

João Victor Villas Boas, [72](#)

### **Stransky**

Beatriz, [32](#), [61](#), [90](#)

### **Sá**

Marcus Villander Barros de Oliveira, [7](#)

## **T**

### **Tavares**

Thayana Vieira, [74](#)

### **Terrematte**

Patrick, [77](#)

## **U**

### **Usuda**

Júlia Nakanishi, [18](#), [39](#), [48](#), [93](#), [94](#), [102](#)

## **V**

### **Vainstei**

Marilene Henning, [33](#)

### **Vale**

Fernando Yuri Nery do, [39](#), [93](#), [94](#)

### **Valera**

Amanda Rodriguez, [39](#), [48](#), [93](#), [94](#)

### **Vasconcelos**

Felipe Alves de, [27](#)

Luydson Richardson Silva, [7](#), [63](#), [83](#), [96](#)

### **Vilhena**

Eduarda Vianna, [58](#)

### **Viseras**

César, [28](#)

## **X**

### **Xavier**

André da Silva, [80](#)

Thiago Jesus da Silva, [48](#)