PHYSICIAN SCIENTIST
LA MEDICINA INCONTRA LA RICERCA
promosso da: AIRC Fondazione CARIPLO | con il supporto di: SIEMM Società Italiana di Medicina Molecolare | UNIVERSITÀ DEGLI STUDI DI MILANO

Principal Investigator	De Cecco Loris
Hosting institution	Fondazione I.R.C.C.S. Istituto Nazionale dei Tumori - Milano
Proposal title	Spatial Metagenomic and Tumor Microenvironment landscape in Head and Neck Cancer for AI-Driven Immunotherapy Prediction
Keywords	Immunotherapy; Genomics; Head and neck ca.; Clinical trials; Transcriptome/Transcriptomics
PhD project description	Immune checkpoint inhibitors are widely used for head and neck squamous cell carcinoma, but patient responses in clinical practice remain highly variable. This project, developed under the AIRC IG23573 grant (PI: Dr. De Cecco), aims to validate novel AI biomarkers and assess their predictive power in real-world settings. The intratumoral microbiota, the collection of microorganisms within a tumor, can significantly impact the effectiveness of immunotherapy by influencing the tumor microenvironment and modulating the host's immune response. The project integrates five key domains: clinical efficacy, resistance mechanisms, host/microbiota relationships, biomarker discovery, and personalized oncology strategies. Central to this effort is the use of spatial metatranscriptomics to decode the tumor microenvironment (TME), revealing immune-microbiota interactions and spatial gene expression gradients that may drive therapeutic outcomes and immune evasion. Hypothesis and Aims We hypothesize that spatial gene expression profiling will uncover actionable biomarkers and spatial immune-microbiota dynamics predictive of treatment response. Specific aims include: 1)Generating high-resolution spatial maps of TME architecture. 2)Identifying immune, stromal, and microbial gene expression signatures within intact tissues. 3)Comparing spatial features to reveal immune evasion mechanisms. 4)Integrating spatial and bulk/single-cell transcriptomics to build a multi-layered atlas for outcome stratification. Computational Framework A five-module pipeline will support the project: 1)High-resolution spatial transcriptomics using technologies like STOmics Stereo-seq and tools like Cell2location. 2)Multi-omics integration via VAEs, denoising autoencoders, and knowledge graphs. 3)Predictive modeling using deep multi-task networks, XGBoost, and interpretable AI. 4)Generative modeling with VAEs and GANs to simulate synthetic cohorts and phenotypes. Clinical validation and deployment under TRIPOD and CONSORT-AI standards, supporting integration into clinical systems through FHIR, dashboards, and privacy-preserving data sharing. This project seeks to advance precision immunotherapy through robust, spatially informed biomarker discovery and clinically actionable predictive tools. The experimental, computational, and clinical activities will be carried out at Fondazione IRCCS Istituto Nazionale dei Tumori (Milan).

Main topics of the lab	• Integrated experimental and computational framework • Multi-omics profiling • Spatial transcriptomics • Advanced computational analytics • Biomarker discovery • Translational precision oncology
Short description of the lab activity	Our lab utilizes an integrated experimental and computational framework for multi-omics profiling of FFPE tissues and liquid biopsies. We combine spatial transcriptomics, long-read methylation sequencing, and high-throughput RNAseq to capture comprehensive molecular insights. By unifying diverse data modalities through advanced analytics and adhering to FAIR data principles, our platform ensures reproducibility, interoperability, and robust integrative analysis. This approach supports biomarker discovery, disease stratification, and translational applications in precision oncology. 1. Spatially Resolved Transcriptomics StereoSeq OMNI (BGI) captures spatially barcoded transcripts from FFPE sections at 50 nm resolution, enabling single-cell mapping. Libraries undergo preparation, sequencing (DNBSEQ-T7RS), and mapping to spatial grids. Quality metrics (RIN, UMI counts, spatial correlation) are tracked to correct FFPE-induced artifacts and reveal tissue heterogeneity. 2. Methylation Profiling Using Oxford Nanopore's GridION R9.4, long-read DNA methylation is performed without PCR or bisulfite treatment, avoiding sequence-context bias. Basecalling (Guppy), alignment (minimap2), and methylation detection (Nanopolish) ensure deep coverage. POD5 tracking ensures reproducibility under FFPE conditions. 3. RNA Sequencing Bulk RNAseq (QuantSeq, Lexogen) and small RNAseq from liquid biopsies (QIAseq, Qiagen) are performed. Libraries are sequenced on Illumina platforms, quality-checked (FastQC), and aligned/quantified using Kangooroo and GeneGlobe. Outputs include standardized gene and miRNA profiles (aligned to miRBase). 4. Computational Infrastructure The team utilizes HPC resources, GPU acceleration, and scalable storage. Pipelines are containerized (Docker, Singularity) and orchestrated by Nextflow for reproducibility. Metadata versioning, access control, and encryption maintain data integrity and privacy. 5. Data Integration and Management Datasets (~28TB total) include spatial transcriptomics (8TB), methylation (15GB/sample), and RNA/miRNA (10GB/sample). All data are stored in standard formats (CSV, JSON, XML) and annotated with ontologies (Dublin Core, EDAM, ISA) to ensure machine-actionable integration and FAIR compliance. 6. Multi-Omics Analysis End-to-end pipelines integrate multi-omics for biomarker discovery and pathway modeling. Tools include version-controlled workflows, real-time dashboards, and ontology-linked clinical metadata for stratified analysis. Datasets are traceable and shared in open, non-proprietary formats. 7. Computational Modeling Advanced methods such as scVI, MOFA+, DCCA, and R-GCN reduce noise, integrate data, and uncover latent factors. Predictive models (Cox, RF, XGBoost, deep survival) are evaluated using SHAP, AUC, and C-index. Digital twins and disease trajectories are modeled with LSTMs and neural ODEs, following TRIPOD/CONSORT-AI standards. 8. Biomarker Discovery Tissue: Bulk and single-cell RNAseq with spatial mapping validate hypoxia and immune markers. Liquid biopsy: Serial ctDNA and miRNA profiling will assess treatment response. Integration:

	Multi-omic data will be fused to build composite predictors to stratify responders, guiding biomarker-driven trials. 9. Mechanistic Insights We expect to enhance immune infiltration, improving response to PD-L1 blockade. Spatial profiling will uncover resistance drivers such as MDSCs, TAMs, and CAFs, informing rational combination strategies. 10. Translational Impact Findings may support trials with novel agents. Real-time omics will underpin adaptive trial designs, reducing decision timelines. Health economics and QoL assessments will compare TPE-based regimens to standard care, potentially shaping regulatory approvals and guideline updates for underserved HNSCC populations.
Main research area	Computational biology
Group composition	Our laboratory combines comprehensive experimental and computational expertise to support end-to-end multi-omics research, from sample processing to advanced data analysis. The team comprises 2 postdoctoral researchers (one in experimental biology, one in bioinformatics), 3 PhD students, 4 bioinformaticians/computational scientists, 4 research fellows in biology/biotechnology, and 1 lab technician. This diverse skill set enables the lab to independently execute complex projects across multiple disciplines. On the experimental side, we perform spatial transcriptomics, long-read methylation sequencing, and bulk/small RNA sequencing. Our biologists and technician handle sample preparation, library construction, and quality control. The team follows standardized protocols and ensures that all procedures are optimized for FFPE and liquid biopsy samples. Computationally, the bioinformatics team leads data curation, metadata annotation, and pipeline development. They integrate public and proprietary datasets, apply rigorous quality control and normalization procedures, and manage FAIR-compliant metadata standards. Signature and model development and AI based analyses are implemented in reproducible workflows. Our infrastructure supports scalable computing via containerized pipelines, compatible with cloud platforms (e.g., ELIXIR, EOSC). Machine learning specialists develop and benchmark predictive models (lasso-Cox, XGBoost, Random Forest), validate them using survival metrics (C-index, AUC), and deploy them as interactive tools (R Shiny, Streamlit). The team ensures dissemination through detailed documentation, tutorials, and version-controlled repositories. All data and software are deposited in public repositories (Zenodo, EMBL-EBI) with persistent identifiers and full provenance tracking. Close collaboration with clinicians enables real-world validation and translation of models into clinical workflows. Our group actively contributes to publications, training materials, and methodological sections of project reports, ensuring transparency, interoperability, and scalability of all outputs. The Medical Oncology Unit – Head and Neck Cancer provides comprehensive, multidisciplinary care for patients with cervicofacial tumors, following national and international guidelines. It is actively involved in clinical research and offers access to innovative therapies, including immunotherapy and targeted agents. Multidisciplinary teamwork ensures personalized treatment planning. The Unit is also committed to education, hosting international conferences and

	producing patient-oriented materials, contributing to high clinical trial enrollment and scientific dissemination.
Institutional page link	https://www.istitutotumori.mi.it/en/home
Lab website link	
Social media link	
Lab bibliography	Immune-based classification of HPV-associated oropharyngeal cancer with implications for biomarker-driven treatment de-intensification. Zeng PYF, Cecchini MJ, Barrett JW, Shammas-Toma M, De Cecco L, Serafini MS, Cavalieri S, Licitra L, Hoebers F, Brakenhoff RH, Leemans CR, Scheckenbach K, Poli T, Wang X, Liu X, Laxague F, Prisman E, Poh C, Bose P, Dort JC, Shaikh MH, Ryan SEB, Dawson A, Khan MI, Howlett CJ, Stecho W, Plantinga P, Daniela da Silva S, Hier M, Khan H, MacNeil D, Mendez A, Yoo J, Fung K, Lang P, Winquist E, Palma DA, Ziai H, Amelio AL, Li SS, Boutros PC, Mymryk JS, Nichols AC EBioMedicine 2022 Dec; 86: 104373