|  |  |  |
| --- | --- | --- |
| **Docente** | **Reperibilità** | **Argomento tesi – numero studenti**  |
| Daniela Scribano PhD, Cecilia Ambrosi PhD,  | Daniela Scribano PhD, daniela.scribano@uniroma1.it Cecilia Ambrosi PhD, cecilia.ambrosi@uniroma1.it | **Bacteriology labs at the Dept. of Public Health and Infectious Diseases****Project 1, acronym: AGRINTO (1-year project)**Title: **Identification of genetic resistance determinants in nosocomial pathogens responsible for respiratory infections.** **Description.** A total of 10 Gram-negative multidrug resistance bacterial isolates, belonging to the Bacteriology labs of the Dept. of Public Health and Infectious Diseases, will be analyzed for the presence of resistance genes, through whole genome sequencing (WGS). The presence of regulatory elements flanking resistance genes will be characterized. Finally, the antimicrobial susceptibility testing (AST) of isolates will be compared to their genetic content to evaluate if WGS can be used as a tool to infer antimicrobial susceptibility. Project phases* Phase I, bacteriology labs: extraction of bacterial DNA and evaluation of DNA quality and quantity,
* Phase II, Scientific Department of Army Medical Center of Rome: WGS,
* Phase III, bioinformatic analyses, AST and reporting.

**Project 2, acronym: AGITO (1-year project)**Title: **Characterization of urban water microbiome and resistome****Description:** Water samples will be collected and concentrated using the Lab Filtration Pump (EZ-Stream). Filters will be placed in a nutrient-rich liquid medium for bacterial recovery. DNA will be extracted from total recovered bacteria and subjected to next generation sequencing (NGS) to identify bacterial species and antimicrobial resistance genes. In agreement with the European surveillance of *waterborne disease*, the results of the project will increase the knowledge on microbial composition of urban water. Project phases:* phase I, bacteriology labs: water sample filtration, DNA extraction and evaluation of DNA quality and quantity,
* phase II, Scientific Department of Army Medical Center of Rome: NGS,
* phase III, bioinformatic analyses and reporting.

**Project 3, acronym: PATO-ESCAPE (1-year project)**Title: **Characterization of multidrug-resistant pathogens from Chronic Obstructive Pulmonary Disease (COPD) patients: role of epithelial adhesion molecules and of immune and inflammatory responses****Description**: Bacterial pathogens will be isolated from sputa of patients with acute exacerbation of COPD admitted to the Respiratory Rehabilitation Unit of the IRCCS San Raffaele Roma. After species identification, isolates will be subjected to antimicrobial susceptibility testing. MDR isolates will be selected for whole genome sequencing (WGS) to characterize bacterial resistome as well as bacterial virulome. Isolates will be subsequently analyzed for principal pathogenic phenotypes by using *in vitro* infection models. Sputa will be also used to evaluate the expression profiles of eukaryotic receptors (i.e. CEACAMs and PAFR) on the surface of bronchial epithelial cells to assess the risk of bacterial colonization / infection. Finally, cytofluorimetric assays will be performed to determine immune cell populations in both sputa and peripheral blood with particular attention to neutrophils, eosinophils, monocytes / macrophages and regulatory T cells. Project phases:* phase I, bacteriology labs: bacterial isolation, identification, AST, genomic DNA extraction and evaluation of DNA quality and quantity,
* phase II, Scientific Department of Army Medical Center of Rome and/or Bio-Fab Research srl: WGS,
* phase III, bioinformatic analyses,
* phase IV, bacteriology labs and Lab of cellular and molecular immunology (IRCCS San Raffaele): Western blot and cytofluorimetric analyses.

Contact: Daniela Scribano PhD, daniela.scribano@uniroma1.it Cecilia Ambrosi PhD, cecilia.ambrosi@uniroma1.it |
| Massimo Reverberi | Massim.reverberi@uniroma1.it | Avviso di disponibilità di un posto di tirocinio (circa 2 mesi di impegno) a partire da subito presso il CREA DC di via CG Bertero 22 00156 Roma sotto la tutela del Dr. Alessandro GrottoliMetagenomics is one of the approaches that is gaining increasing popularity in large-scale genomics applications. The major application of metagenomics involve taxonomic and functional composition of microbial communities in areas such as environmental, agricultural, clinical, etc. The major difference between metagenomics and traditional single-genomic approaches is that metagenomics is not based on the single microbial colony isolations from complex matrices, with the many limitations of the technique, but provides an "omics" approach, with a holistic view of the microbial complexity analyzed, sequencing all the genes and genomes of a mixed community at once.The project involves the study and comparison of the lasts techniques for identification and determination of taxa present in the various samples and the design of a specific and accurate pipeline.It will be performed a comparative analysis of metagenomics data of sequencing (artificial and real) based on different approaches: shotgun sequencing, amplicon sequencing (Bacteria, Fungi). It is expected: knowledge of basic computer science applied to bioinformatics (use of commands on shell bash linux), use of various bioinformatics applications (toolkit blast, kraken2, etc) and practicality in the use of programming languages (Pyhton, R, Pearl, etc).massimo.reverberi@uniroma1.italessandro.grottoli@crea.gov.it |
| Paola Vittorioso | paola.vittorioso@uniroma1.it | Plants, as sessile organisms, have to survive to frequently changing environments presentingadverse or stressful conditions, which directly affect growth and development. Among abioticstress, low temperature is one of the most relevant factors limiting distribution of plant species. Toavoid damage due to freezing temperature, plants modify the polyunsaturated fatty acid content ofmembranes, accumulate stress-related proteins to prevent dehydration, and activate antioxidantenzymes, a response called cold acclimation. Although some key components of the cold signalingpathway have been deeply characterized, the complexity of this adaptative process is far from beingunveiled.Dof Affecting Germination 1 (DAG1) is a plant-specific transcription factor, which controls thebalance between the phytohormones GA and ABA during the seed-to-seedling transition. SinceABA plays a pivotal role in the stress response, we have assessed the involvement of DAG1 in theresponse to abiotic stress, through a phenotypic analysis of the tolerance of *dag1* KO mutant underdifferent environmental stresses. Interestingly, *dag1* mutant plants were significantly more tolerantto freezing temperatures respect to the wild type, suggesting that DAG1 may function as a repressorof this process.To deeply investigate the role of DAG1 in response to cold stress, we performed a genome-wideanalysis, through RNA-seq assay, by comparing the transcriptome of *dag1* and wild type coldtreatedplants with the untreated lines. A preliminary analysis of this study revealed a significantenrichment of the biological process (BP) GO categories related to lipid metabolism and membranetransport, in the cold-treated *dag1* sample respect to the cold-treated wild type or the untreated*dag1*. Accordingly, a preliminary biochemical analysis of lipid composition revealed an alteredfatty acid composition in both the cold-treated and untreated *dag1* samples compared to the wildtype ones.The project for the bachelor student thesis will concern a deep analysis of the RNA-seq data, inorder to identify Differentially Espressed Genes (DEGs) which are specifically cold-deregulated inthe *dag1*-mutant compared to the wild type, as well as DEGs which are specifically associated with*DAG1* inactivation. In addition, a Gene Ontology analysis focused on the BP and MF categoriesenriched in both the cold-treated and untreated *dag1* mutant respect to the wild type, will help inshedding light on DAG1 function in this adaptative response. |
| Simona Giunta | Dipartimento di Biologia e Biotecnologie C.DarwinContact: simona.giunta@uniroma1.it | **PROJECT 1. Enriched Heatmaps using R to explore epigenetic changes in the human genome**The project, of the duration of 6-12 months, requires a student who is either already familiar with R and similar software, or particularly agile with bioinformatics approaches to explore and characterize changes in chromatin features across the human genome and under different experimental conditions. This project does not have a wet lab component and is amenable to remote work/supervision. This project is undertaken in English as part of a collaboration with scientists at Yale University (USA) with whom the student is expected to interact. **PROJECT 2. Mutivariate image analysis using ImageJ**The project, of the duration of 6-12 months, requires a student who is either already familiar with ImageJ or other imaging processing software, or particularly agile with image processing, acquisition and analysis approaches, with a particular plus for people interested in super-resolution microscopy. The project will explore and characterize changes in DNA markers under different experimental conditions, especially those of damage and endogenous cellular stress. This project has an optional wet lab component and is amenable to remote work/supervision.**PROJECT 3. The missing genome**The project, of the duration of 12 months, requires a student who is either already familiar with sequence analysis, WGS, GWAS, or particularly agile/interested in human genomics. The project will explore and characterize the latest human reference genome assembly and explore their relevance on our own, already obtained dataset. This project has no wet lab component and is amenable to remote work/supervision. This project is undertaken in English as part of a collaboration with several international scientists with whom the student is expected to interact.  |
| Maria Pia Conte | mariapia.conte@uniroma1.it | 1 studenteLinee di ricerca:-The pangenome structure of Escherichia coli patotypes: comparative genomic analysis and phylogenetic organisation.-Integrative metagenomic and culturomic approach to study temporal dynamics of skin microbiota inatopic dermatitis-Pangenome analysis of Burkholderia cepacia complex involved in healthcare-associated infections (HAIs) |
| Giuseppe Biondi Zoccai | gbiondizoccai@gmail.com | 1 studenteTitolo orientativo: Integration of preclinical and clinical evidence with umbrella reviews.In pratica il progetto si concentrerebbe sulla possibilità di condurre analisi semiautomatiche che permettano la sintesi quantitative di diverse fonti di dati, dagli studi clinici nell'uomo, agli studi negli animali, a quelli in vitro. |
| Serena Schippa | serena.schippa@uniroma1.it | Uno studente per volta:" Caratterizzazione del batterioma intestinale (fecale o mucosa-adeso) nell’ambito di diversi contesti patologici”. |
| Maria Elisa Crestoni | mariaelisa.crestoni@uniroma1.it | Uno-due studenti"Analisi statistica e bioinformatica per lo studio di metaboliti in matrici alimentari complesse con il supporto di banche dati dedicate" |
| Lucia Di Marcotullio | lucia.dimarcotullio@uniroma1.it | Our group has been studying for several years the molecular mechanisms which regulate the pathway of Hedgehog (Hh), a signal transduction pathway involved in the development processes, staminality and oncogenesis. Recently we discover a new Hedgehog pathway activator, the endoplasmic reticulum ERAP1 aminopeptidase, unrevealing an unexpected role in cancer for this protein which is becoming a promisful target for innovative therapeutic strategies. The student interested to a thesis stage in our lab will be involved in the analysis of transcriptomic data obtained following depletion or upregulation of ERAP1 in Hh-dependent cellular models. |
| Walter Adriani | walter.adriani@iss.it | 2 progetti di ricerca:"Ritmi circadiani e fenotipo comportamentale del ratto DAT ko" Su modello animale (ratto) di sintomi riconducibili al disturbo ossessivo-compulsivo, alla ADHD ed alle addictions - anche di natura puramente comportamentale (es. gioco d'azzardo patologico) - si prevede di svolgere letture ed analisi dati ottenuti con sistemi automatizzati, come [p.es](http://p.es). sensori di attività 24h/24 7gg/7 oppure test di comportamento operante a valenza neuro-economica ("gambling tasks" e simili). OPPURE "Sviluppo di una App per implementare il Public Goods Game"Il PGG è un 'gioco' che contempera la propensione a donare risorse economiche per il beneficio di tutti i partecipanti versus la tentazione di trattenere per sè queste stesse risorse (altruismo vs egoismo): si prevede di sviluppare una App online che possa permettere il reclutamento di volontari sani, al fine di monitorare e valutare le loro strategie di gioco.  |
| Valerio Fulci | valerio.fulci@uniroma1.it | 1 studente:Development of a bisulfite RNA-seq analysis pipeline to assess the 5mC profile in mammalian small RNAs.Recently, a great deal of attention has been paid to so called "epitranscriptomic" modifications, i.e. covalent modifications of cell RNA molecules. It has been shown that mammalian mRNAs and non-coding RNAs harbour several modifications such as 1-methyl-Adenosine (m1A), 6-methyl-Adenosine (m6A) and 5-methyl-cytosine (m5C). m5C can be easily detected by bisulfite treatment which results in the conversion of non-methylated C residues into U residues, which are subsequently sequenced as T residues. 5mC residues are not converted. The pipeline we are currently developing aims to focus on the small RNA fraction in which 5mC has been poorly investigated thus far. The pipeline will consist of a dedicated alignment strategy taking advantage of custom PERL/python scripts for parsing and processing of aligned reads. R statistical software will be used to generate publication quality pictures and statistical testing to call significantly differentially 5-methylated C residues across multiple samples. |
| Matteo Pallocca | matteo.pallocca@gmail.com |

|  |  |
| --- | --- |
| Title | Background and Task |
| Project 1. A coverage check-package for clinical-based Whole Exome Sequencing | **Background**Wide-target genomic NGS experiments are becoming a de facto clinical tool for some specific situations, e.g. when the histology of the tumor is unknown and/or when there’s need to unveil possible alternative drugs for the patient.**Task**Given a BAM file of a WES experiment and a specific list of genes of interest (e.g. Actionable genes), provide a minimum/median coverage for each one and represent it in a graphical way. This process is extremely computation-intensive so it requires some optimization. **Tools**Bash, Python/R, Web technologies (extra) |
| Project 2. Development of an enhancer-ranking algorithm for in vivo cancer epigenetics | **Background**Translational cancer epigenetics projects involve large-scale profiling of many patients with high throughput techniques such as ChIP-seq and ATAC-seq in order to profile active areas of the genome such as enhancers and super-enhancers. **Task**Develop an algorithm in order to rank and call in a standardized/fully automated way peaks and regions among a patient’s population, by their own Ranking and Sharing indexes. This tool will help to shear light among the strong inter-sample heterogeneity at the epigenetic level, very often being the primary cause of pharmaco-resistance in tumors. **Tools**Python, Bash, Web technologies (extra). |
| Project 3. Development of a framework for ImmunoInformatics analysis | **Background**Many new drugs have been approved in order to unleash the immune system against solid tumors, such as the new wave of Immuno-Checkpoint Inhibitors (ICI). In spite of these major breakthroughs, responses to immunotherapy vary from long term survival to early non-response, and biomarkers in current regulatory protocols only marginally increases the percentage of responders. **Task**This project aims to develop a Immuno-Informatic Framework, in order to integrate several HT techniques (RNA-seq, WES, TCR-seq, HLA-seq) to find possible novel biomarkers for Cancer Immunotherapy in Non-Small-Cell Lung Cancer. **Tools**Python, Bash, Web technologies (extra). |
| Project 4. Automated Variant Annotation | **Background**Variant/Mutation annotation is a process that involves the integration of several databases, for clinical significance (e..g ClinVar), somatic tumor evidence (COSMIC), pathogenicity prediction (e.g PolyPhen). Can be cumbersome and, as usual, could use some automatization. **Task**Starting from Basic Scripts calling the ANNOVAR package, create a web page that takes as an input a VCF, calls Annotation with ANNOVAR and displays a series of plots or statistics, and enables the download of the output file in TSV/XLSX format **Tools**Web technologies, Bash, Python.  |

 |
| Stefano Pascarella | stefano.pascarella@uniroma1.it | Uno o due studenti:- "Development of a pipeline for inverse docking".- "Analysis of the taxonomic distribution of bacterial genes involved in the Pyridoxal 5'-phosphate metabolism". |
| Marco D’Abramo | marco.dabramo@uniroma1.it | Un paio di studenti:Point mutations in kinases: structural/dynamical effectsThe project concerns the use of bioinformatics approaches to find and understand the biophysical effects of specific mutations in protein kinases.  |
| Maria Carafa | Maria.carafa@uniroma1.it | Uno studente:The research activity of the Nanomedicine\_Lab is mainly involved in preparation and characterization of innovative drug delivery systems with application in therapeutic, diagnostic and theranostic fields, by different administration routes. The lab activity focus on the application of different preparation techniques of liposomes, niosomes, nanobubbles and nanoemulsions and their physical-chemical characterization. |
| Laura Ottini | Laura.ottini@uniroma1.it | Analysis of matched germline and tumor profiles in male breast cancer for new molecular biomarker discovery.Male breast cancer is a rare and poorly studied disease. To date, clinical management has been considered similar to female breast cancer management. Increasing evidence indicates that breast cancer in men and women may behave differently. Thus, there is need to obtain further data on the genetics and biology of this rare disease and how to improve the appropriateness of the clinical management of men with breast cancer.In this project we propose to analyze and integrate data of matched germline and tumor profiles obtained by analyzing a series of male breast cancer samples using Next Generation Sequencing (NGS) methods. The integration of genomic data with matched cancer transcriptomic data will allow to associate genomic alterations with gene expression. The results of the study will allow to identify differentially expressed gene sets/pathways in distinct molecular subtypes of male breast cancer with possible prognostic and therapeutic relevance. |
| Michele Signore eRomina Alfonsi | michele.signore@iss.it | RPPA unit, Proteomics area, Core Facilities, Istituto Superiore di Sanità[https://corefacilities.iss.it/dw/doku.php?id=aree:proteomica:rppa:start]Michele Signore, PhD [https://orcid.org/0000-0002-0262-842X]Romina Alfonsi, PhD [https://orcid.org/0000-0001-5100-5491]-- Project 01 --Title: Implementation of a pipeline for pathway enrichment analysis of RPPA data.Background: Reverse-Phase Protein microArrays (RPPA) are an antibody-based technology aimed at biasedstudying of the (phospho-)proteome in biological samples [PMID: 17892368]. The exquisite sensitivity andsample throughput of RPPA as well as its minimal sample amount requirements, have made the RPPA an idealplatform for pathway analysis by the The Cancer Genome Atlas (TCGA) consortium [PMID: 31201206].Although the RPPA is a powerful technique to investigate the functional proteome of e.g. cancer cells, theinterpretation of RPPA results is still based on the curated knowledge gathered by individual inverstigatorsapplying RPPA within their experimental design. In addition, since i) RPPA is based on relative quantification of analytes, ii) the amount of activation thershold having a biologically meaningful effect is context-specificand iii) each antibody has its own binding curve, making it impossible to compare levels of analytes withoutdata scaling, conversion of RPPA data to a set of useable and aknowledged results still represents a challenge.Despite several gene-based tools exist for pathway enrichment analysis, direct application to RPPA data is notstraightforward and the absence of a full coverage of the proteome in a typical RPPA dataset precludes thepossibility to interrogate available RNAseq- or mass spectrometry-based bioinformatic tools. Aims: Definition of a Python- and R-based pipeline for fetching approved phosphorylation and functional data fromPathway Commons database and generation of a data report template.Methods: Implementation of a pre-existing set of R and Perl codes for fetching and parsing PC2 data as well as applying hypergeometric test to specific RPPA datasets. Update of Perl-based and, whether possible, of R-based codes, to a Python 3 format. Generation of an R Markdown document template for reporting RPPA dataenrichment pipeline results.Base: Istituto Superiore di Sanità, Via Giano della Bella 34, 00162, Rome-- Project 02 --Title: Rational integration of RPPA and RNAseq data.Background: Reverse-Phase Protein microArrays (RPPA) are an antibody-based technology aimed at biasedstudying of the (phospho-)proteome in biological samples [PMID: 17892368]. The exquisite sensitivity andsample throughput of RPPA as well as its minimal sample amount requirements, have made the RPPA an idealplatform for pathway analysis by the The Cancer Genome Atlas (TCGA) consortium [PMID: 31201206].RNAseq is a NGS-based technology mainly used for, but not limited to, transcriptome profiling and, ifcompared to conventional cDNA microarrays, provides more accurate estimates of isoform abundance over a wider dynamic range, thus allowing accurate measurements down to single cells [PMID: 18516045, 31792409].RPPA and gene expression data have been successfully integrated with other molecular information available for specific TCGA datasets [PMID: 23000897, 25079552, 29100075], for the Pan-Cancer dataset [PMID: 24871328] and, more recently, in the context of the Cancer Cell-Line Encyclopedia (CCLE) [PMID: 29293502].Topological integration of RPPA data with other '-omics' has been shown to provide pathway-based information for prediction of patient survival in breast cancer [PMID: 31296204] and useful perspectives have beenproposed for intuitive and synthetic representation of integrated RPPA results [PMID: 26185419]. Recently, efforts to dissect specific pathways at proteogenomic and Pan-Cancer level, have further confirmed thepower of pathway-based molecular correlates in predicting cancer patient survival [PMID: 28528867].Nonetheless, to date an established methodology for collating RNAseq and RPPA data is not available and most of the published studies employ manually curated pathway definition [PMID: 29625050].Aims: Pathway-based collation of RPPA and RNAseq dataset and network and/or pathway glyph representation of results.Methods: Annotation of the full panel of antibodies validated by RPPA@ISS with Pathway Commons 2 official pathways. Definition of a minimal pathway set to be integrated with RNAseq data, based on the number and role of represented RPPA analytes. Definition of a data integration algorithm that takes into account the functional role of phospho-proteins and the relative abundance of corresponding transcripts in differentsample groups. Representation of integrated results by using synthetic pathway visualization (e.g. Figure 2from PMID: 29625050).Base: Istituto Superiore di Sanità, Via Giano della Bella 34, 00162, Rome / Istituti Fisioterapici Ospitalieri,Via Elio Chianesi, 53, 00144 Roma RM-- Project 03 --Title: Definition of a pipeline for analysis of ImmunoSeq dataBackground: Reverse-Phase Protein microArrays (RPPA) are an antibody-based technology aimed at biasedstudying of the (phospho-)proteome in biological samples [PMID: 17892368]. RPPA holds an exquisite sensitivity, high sample throughput and minimal sample amount requirements, all of which have allowed itsusage within The Cancer Genome Atlas (TCGA) consortium [PMID: 31201206]. The RPPA protocol is based on the sequential application of diverse techniques, involving sample printing, total protein quantification, antibody staining, image analysis and generation of normalized data, data QC and analysis [https://doi.org/10.1016/B978-0-12-809633-8.12272-1]. The complexity of such experimental protocol, exposesthe RPPA analysis to multiple sources of systematical as well as random errors that inevitably lead to poorinter-experimental reproducibility. The application of batch standard lysate curves as well as eliminationof the lot-to-lot variability in the materials used, minimizes the introduction of errors but absolute, reproducible quantification of (phospho-)protein content is not feasible by using RPPA. Immune-Detection bysequencing (IDE-seq) [PMID: 29921844] is part of a novel set of antibody-based technologies that allow multimodal analyzequantification in single cells [PMID: 31011186] and are based on antibody barcoding [PMID: 28759029].Barcoding of the antibody repertoire available at the RPPA@ISS and conversion of the RPPA technology to NGS, would allow multiplexed, absolute quantification of RPPA analytes by means of a shorter, simple protocol.The generation of a single, multi-dimensional NGS sample with simultaneous barcoding of the sample and antibody spaces, requires the application of ad-hoc pipelines for raw data QC and deconvolution.Aims: Implementation of existing NGS data handling and analysis for application with IDE-seq data and generationof a ready-to-use toolkit for automated production of results starting from raw NGS data.Methods: Selection of a list of barcode sequences including NGS adaptors, to be applied to our antibody repertoire.Generation of a python-based pipeline for reading FASTQ files, trimming, count and deconvolve reads based on sourceAntibody and sample code tables.Base: Istituto Superiore di Sanità, Via Giano della Bella 34, 00162, Rome / Istituti Fisioterapici Ospitalieri,Via Elio Chianesi, 53, 00144 Roma RM |