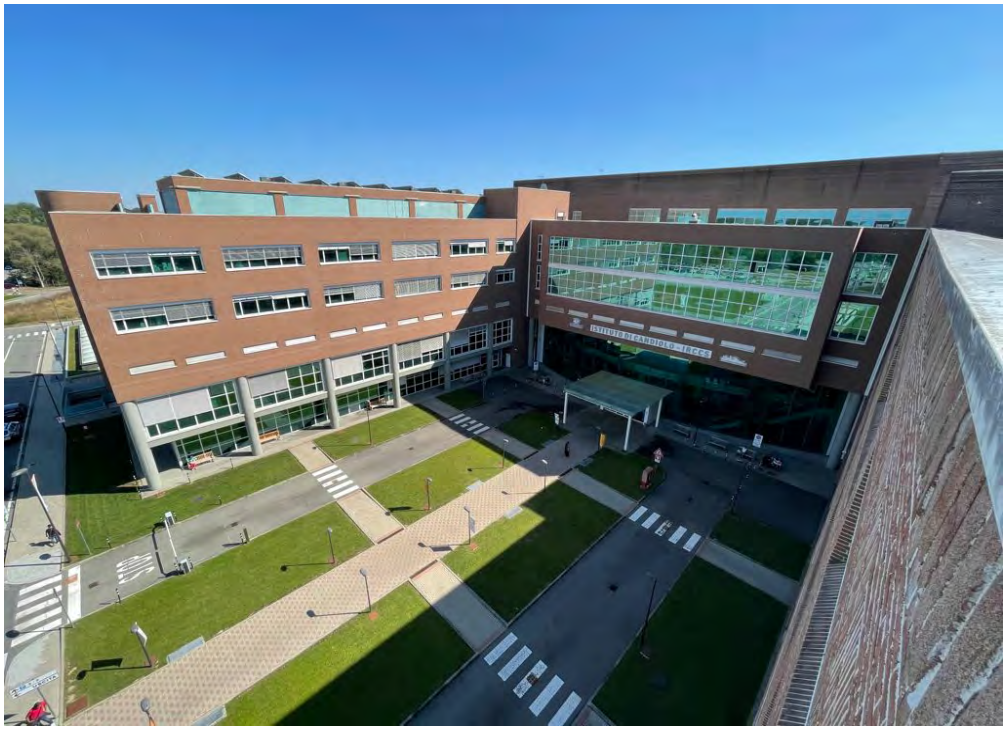
Italian Institute for Genomic Medicine

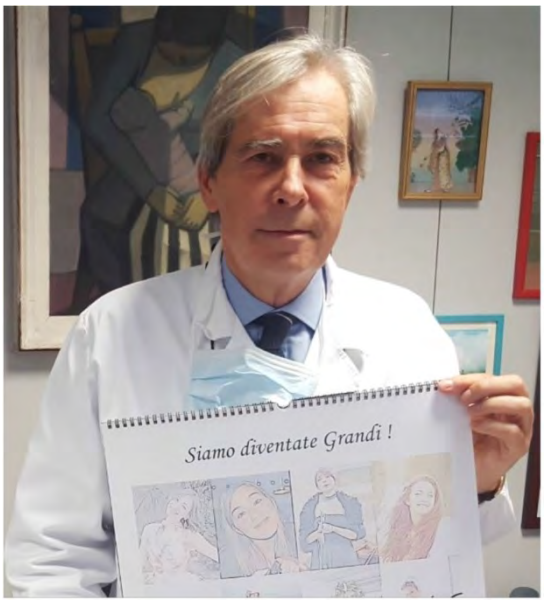
IIGM

Scientific Report

2020-2021



In Memoriam



Working with Prof. Basso was not easy.

He was as demanding with his collaborators as he deeply was with himself. However, his strong work ethics and his deep and contagious sense of duty always came along with a smile, or a joke, and frequently with the right encouragement at the right time, which pushed us not to give up even when the challenge was tough. An so, all of a sudden, working with him became a pleasure, an adventure, sometimes almost a game when, following some brilliant thoughts of his (he was a master in that), a naughty smile perhaps borrowed from decades of life with his young patients appeared on his face, and suddenly we were thrown into the reconstruction of a work "puzzle" where every single tile naturally found its place.

He was a master of life, not just a master of science. He often told anecdotes and first-hand life stories with his stunningly captivating and folkloristic Venetian speech, giving us little pearls of wisdom in brief moments of pure amusement, when he wandered around the offices and laboratory counters on a daily bases, dispensing everyone a word, an encouragement, or a reproach when needed, sometimes severe but always administered with a quiet smile.

We like to remember the man, not just the scientist. Albeit, his scientific life speaks for itself through his achievements in the field of pediatric oncology, and through the legacy he left, including new generations of clinicians and researchers trained by him.

We say goodbye to the Director, but we will never forget the man, his cheerfulness, his teachings, his commitment to work as well to life, and above all his goodness of heart in caring for each one of us, since he strongly believed the value of an institute lies in the people who make it up. He elicited in us all a sense of identity as an "Institute", especially when we all felt somewhat lost due to the many changes we experienced in recent years. It is with this identity as an Institute that now we face the new changes and challenges that still lie ahead, doing our best, as he always urged us to do.

Have a nice journey, Professor Basso.

Lavorare con il Prof. Basso non era facile.

Pretendeva molto dai suoi collaboratori, così come ha sempre preteso molto da se stesso. Ma la sua ferrea etica lavorativa ed il suo senso del dovere, fortemente contagiosi, erano sempre accompagnati da un sorriso, da una battuta di spirito, dall'incoraggiamento giusto al momento giusto che incitavano a non mollare anche quando la sfida sembrava insormontabile. Ed allora, lavorare con lui diventava un piacere, un'avventura, talvolta quasi un gioco quando a seguito di qualche trovata brillante, cosa in cui lui era maestro, sul suo volto compariva un sorrisetto monello, mutuato forse da decenni di vita con i suoi piccoli pazienti, ed all'improvviso ci si trovava catapultati nella ricostruzione di un "puzzle" lavorativo in cui alla fine ogni pezzo andava naturalmente al suo posto.

Era un maestro di vita, non soltanto un maestro di scienza. Raccontava spesso aneddoti e situazioni vissute in prima persona con quella sua parlata veneta coinvolgente e ricca di colore, regalandoci piccole perle di saggezza in brevi momenti di puro svago, girando per gli uffici ed i banconi da laboratorio ed elargendo a tutti, ma proprio tutti, una parola, un incoraggiamento, e non di rado anche qualche rimprovero, talvolta severo ma sempre detto col sorriso e di sicuro non per questo meno efficace.

Ci piace, qui, ricordare l'uomo, non soltanto lo scienziato. La sua vita scientifica parla da sola attraverso le sue conquiste nell'ambito dell'oncologia pediatrica, e per tramite di quello che all'oncologia pediatrica ha lasciato in eredità, tra cui le nuove generazioni di clinici e ricercatori da lui formate.

Salutiamo il direttore, ma non dimenticheremo l'uomo, la sua allegria, i suoi insegnamenti, il suo impegno, e soprattutto la sua bontà, il suo prendersi cura di ciascuno di noi, perché per lui il valore dell'istituto risiede nelle persone che lo compongono. Ci ha dato un senso di unione ed un'identità come "istituto" quando per i numerosi cambiamenti degli ultimi anni ci sentivamo tutti un po' smarriti. Ed è con questa identità come istituto che affronteremo i cambiamenti e le molte sfide che ancora ci aspettano, facendo del nostro meglio, come lui sempre ci esortava a fare.

Buon viaggio, Professore.

The **Italian Institute for Genomic Medicine (IIGM)** (formerly known as Human Genetics Foundation - Torino) is an operating body of the Compagnia di San Paolo since 2007.
The IIGM is a research center of excellence and training in human genomics, epigenomics, and immunology, and carries out its activities through a model of efficiency and transparency, aiming at maximizing resources for research activities, training, and high-level education.
The Institute shares scientific partnerships and collaborations with University of Turin (UniTo) and Polytechnic of Turin (PoliTo).

In December 2018 IIGM and the Piedmont Foundation for Oncology (FPO) – IRCCS created a joint research platform in the FPO-IRCCS building in Candiolo (To), where IIGM laboratories moved in July 2019.
IIGM has organized the new laboratories in accordance with new operational procedures and structural methods: the spaces on have been organized according to the criterion of "functionality and sharing", so to create more opportunities for interaction between research teams, favoring the sharing of scientific notions and the spring of new ideas and projects.
The scientific collaboration between IIGM and FPO-IRCCS will exploit research projects of excellent scientific value, allowing the achievement of the highest international standards in the field of translational and bio-medicine and oncology.

IIGM's research projects are financed by the Compagnia di San Paolo as well as external funders, and address cancer genomics and bioinformatics, epigenetic modifications related to malignant diseases, genomics instability and tumor immunity, immune-regulation, genomic epidemiology, quantitative and computational biology, immunotherapy. The IIGM's commitment to sustain young promising researchers in their early career, providing a stimulating and well equipped research environment, led to the set-up, in early 2021, of a new research group under the supervision of the young PI Dr. Carlo C. Campa, leader of the "Organelle morphogenesis, engineering, and targeting" Unit.

In 2020 a total of 56 articles with IIGM affiliation were published in peer reviewed journals for a total Impact Factor (IF)=387.939 (mean IF 6.927).
In 2021 IIGM's researchers published 65 papers for a total IF= 598.542 (mean IF 9.501)

IIGM has several scientific collaborations aimed at fostering the basic, translational and clinical research activities, through the activation of joint research projects involving both clinical and laboratory activities.

- The IIGM and “La Città della Salute e della Scienza” in Torino cooperate from years in ongoing projects with the goal to identify new molecular markers for the early diagnosis of tumors (collaboration with the Center for the Oncological Prevention of Piedmont). Since 2018 was set up and developed the project "Functional genomics applied to pediatric neoplasms: from mutations, to function, to therapy" (Sargen - Sargenita) funded by a grant from the Compagnia di San Paolo to the Italian Foundation for Pediatric Hematology and Oncology (FIEOP) and a grant by Fondazione Veronesi, which is carried on at the IIGM under the scientific supervision of Prof. Franca Fagioli (Ospedale Infantile Regina Margherita) with the scientific collaboration of Prof. Salvatore Oliviero and Dr. Matteo Cereda. The project aims to deepen the knowledge of the biological mechanisms underlying sarcomas, improve diagnosis, and design new potential therapeutic strategies through a functional genomics platform that integrates nucleic acids sequencing analyses, cell and animal models, and functional studies;
- The partnership with Politecnico di Torino sustains the following projects: "Statistical inference and computational biology" and "New algorithms for inference and optimization from large-scale biological data (INFERNET)” European project funded by the Horizon 2020 Marie Curie RISE program in which IIGM participates as coordinator;
- The partnership with Università Cattolica del Sacro Cuore and Fondazione Policlinico Universitario Agostino Gemelli IRCCS sustains the study of the following topics: 1. Development of therapeutic antibodies; 2. Development of new CAR T therapies; 3. Genomic study of tumors. Thanks to this collaboration, starting from 2020 the IIGM laboratories will host a new "Immunotherapy" research group that will focus on the production of monoclonal antibodies for therapies and/or CAR-T, whose Principal Investigator will be Dr. Tobias Longin Haas;
- The partnership with Department of Biology of the University of Rome "Tor Vergata" led to the creation, in early 2020, of a new research group in IIGM, the "Genomic instability and tumor immunity" Unit, whose Principal Investigator is Dr. Ilio Vitale.

The several collaborations the IIGM is involved in, both with Academia and Health Services, will foster the development of intellectual properties that could contribute to the establishment of new biotechnological companies and investments in the Piedmont area.

Operating structure

The Foundation carries out its activities through a streamlined but effective management structure, and through close collaboration with the consortium PR.I.S.MA. and the Compagnia di San Paolo itself.

This model of efficiency and transparency of management allows to maximize the resources for the core activities of research and service to the scientific community.

The *Board of Directors* consists of five members, including the *President*, who is the legal representative of the Foundation, with signing rights and being invested with all the powers for both ordinary and extraordinary administration of the Foundation.

The *Scientific Director* manages the ordinary administration delegated to him by the Board of Directors, and collaborates with the President coordinating the institutional activity according to the indications and deliberations of the Board itself.

The *operating office* coordinates and optimizes the planning and operational activities of the Foundation in order to make all internal activities effective, performing and functional for the general management of the institute itself.

The Foundation has a *Control Body*, which exercises the powers and functions provided as stated in art. 2403 *et seq*. of the Civil Code, and a *Supervisory Body*, which oversees the correct application of the principles provided for by the "Model of organization, management and control" pursuant to Legislative Decree 231, and has entrusted the task of voluntary auditing of the financial statements to Ernst & Young for the three-year period 2019-2020-2021.

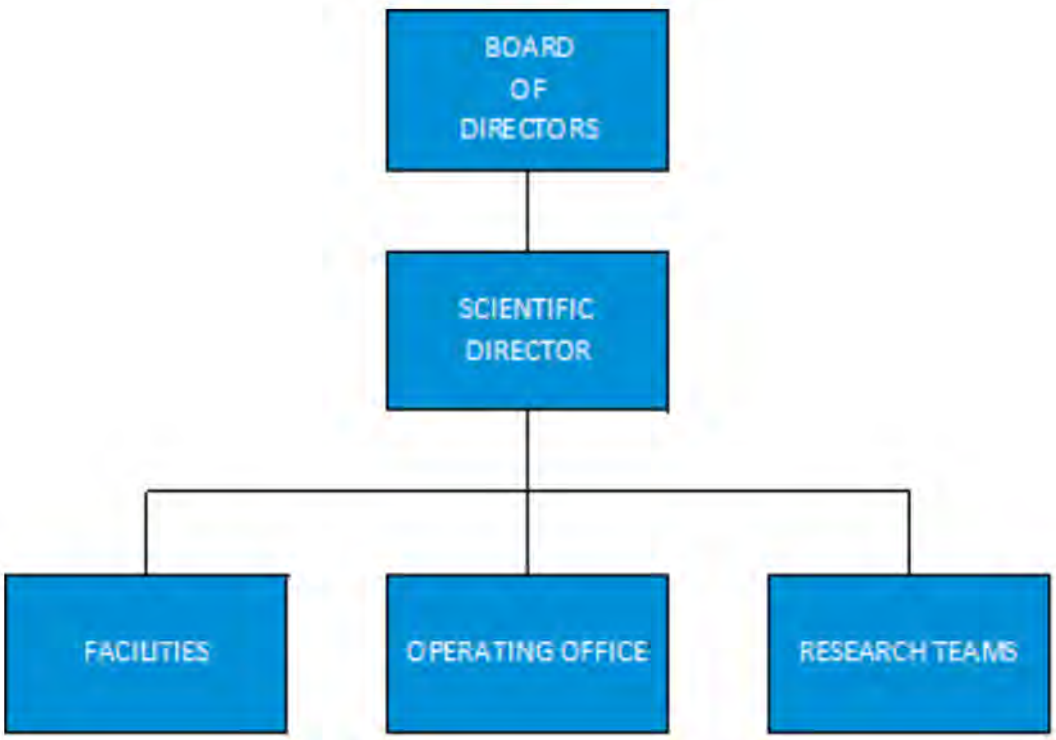
Dr. Alessandro Mazzantini
Chief Operating Officer

Dr. Sabrina Bertinetti
General Secretariat

Dr. Simonetta Guarrera, MSc.
Scientific Secr. & Lab. Managing

The IIGM pursues a policy of state-of the-art research providing its researchers with innovative and up-to-date instrumentation.

- The IIGM's **Metabolic Facility** hosts the Seahorse XF Analyzer, an *in vivo* metabolic analyzer which allows to study key cellular functions such as mitochondrial respiration and glycolysis. The instrument is operating under the supervision of Dr. B. Pardini, and is available under a "all-inclusive service" or a "collaborative" regimen;
- The **Flow Cytometry and Cell Sorting Facility**, run by Dr. D. Baev, is offering services to researchers from the IIGM, the Candiolo Cancer Institute FPO-IRCCS, the Academia, and other research institutes;
- The **Genomic Facility** was established in 2019 to provide access to the most advanced technologies for genomic analysis to researchers from IIGM, FPO-IRCCS Candiolo, University of Turin, and other institutions in Italy and abroad. The facility is operated by dedicated personnel (Dr. K. Gizzi, Dr. C. Parlato) and is available under a "all-inclusive service" or a "collaborative" regimen.



IIGM operational structure

Research and Innovation

IIGM Principal Investigators

Cancer Genomics and Bioinformatics Unit

Dr. Matteo CEREDA, PhD

International experiences:
King's College London, London, UK
IEO, Istituto Europeo di Oncologia, Milan, IT
MRC Laboratory of Molecular Biology, Cambridge, UK

Immunotherapy Unit

Dr. Tobias L. HAAS, PhD

Assistant Professor, Catholic University of the Sacred Heart, Rome

International experiences:
Istituto Oncologico del Mediterraneo, Fondazione IOM Ricerca, Catania, IT
Istituto Superiore di Sanità – ISS, Roma, IT
German Cancer Research Center (DKFZ), Heidelberg, Germany

Molecular Epidemiology and Exposome Unit

Dr. Alessio NACCARATI, PhD

International experiences:
Institute of Experimental Medicine (IEM), Czech Academy of Sciences, Prague, Czech Republic

Immuno-Regulation Unit

Dr. Luigia PACE, PhD

International experiences:
Curie Institut, Paris, France
HELMHOLTZ Centre for infection research, Hannover, Germany

Epigenomics Unit

Prof. Salvatore OLIVIERO, PhD

Full Professor in Molecular Biology, University of Turin

International experiences:
Albert Einstein College of Medicine, The Bronx, NY, USA
Harvard Medical School, Boston, USA
EMBL, Heidelberg, Germany

Statistical Inference and Computational Biology Unit

Prof. Andrea PAGNANI, PhD

Full Professor (L.240), Polytechnic University of Turin

International experiences:
University of Paris-Sud ORSAY, Paris, France.

Genomic Instability and Tumor Immunity Unit

Dr. Ilio VITALE, PhD

International experiences:
Gustave Roussy Cancer Campus (INSERM U848), Villejuif, France

Organelle Morphogenesis, Engineering and Targeting Unit

Dr. Carlo C. CAMPA, PhD

International experiences:
Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

INDEX

Collaborative project with Fondazione Italiana di Ematologia e Oncologia Pediatrica

Project: "Functional genomics applied to pediatric neoplasms: from mutations to biological function and therapy"

F. Fagioli (PI), S. Oliviero, M. Cereda, E. Tirtei, M. Del Giudice, S. Peirone, S. Perrone, G. Romano, F. Priante, C. Parlato, F. Varese, H. Dastsooz, S. Rapelli, M. Maldotti, F. Anselmi, A. Lauria

In recent years it has become possible to study the role of genetic and epigenetic alterations in the onset of cancer. Thanks to the close collaboration between the IIGM researchers, the University of Turin, and the Regina Margherita pediatric hospital in Turin, we identified somatic mutations, gene fusions, epigenetic alterations in patient tumors including sarcomas, hematopoietic tumors, breast and colon cancers.

Aims

1. Next generation sequencing (NGS) of biopsy samples from pediatric or young adult patients and related genomic and transcriptional analyses to characterize somatic alterations, identify possible therapeutic strategies and provide the clinician with an additional diagnostic tool

2. To study the molecular mechanisms that determine the onset and progression of sarcomas. These tumors are characterized by a low level of somatic mutations which indicate that epigenetic and developmental alterations play an important role in the genesis and evolution of this tumor. Our working hypothesis, supported by our preliminary data, is that epigenetic alterations play a key role in the onset and progression of the disease.

Results

In order to classify the samples to be analyzed and have a database of each patient we have created a new software for NGS data analysis with multiple and dynamic workflows with high reliability. This software introduces a novelty compared to the state of the art and we strongly believe that it represents a great advantage for many research groups. HaTSPiL is a pipeline based on the Python programming language, highly customizable and expandable. Thanks to the use of barcodes containing information relating to the sample to its processing, the software is able to automatically choose different workflows and parameters. Analysis of the gene alterations of sarcomas revealed a low mutation spectrum and some gene fusions, some of which are known. Starting from a cohort of ~ 30 sarcoma patients, we analyzed tumor mass expression profiles and compared them to surrounding healthy tissue. In our cohort of patients two patients respectively had 1) A translocation resulting in the production of a fusion protein between the ATRX and ERG genes (called ATRX-

ERG), 2) An internal tandem duplication (ITD) of approximately 100 nucleotides in the fifteenth exon of the BCOR gene (co-repressor of BCL6, called BCOR-ITD) that generates a mutated protein, but with the correct reading register. We investigated whether the overexpression of the two mutated proteins could be the cause of the onset of the tumor. We cloned the entire sequence of BCOR (both WT and IDT) and ATRX-ERG in expression vectors and we overexpressed them in human stem cells. After verifying that the corresponding protein is correctly produced we carried out a pilot ChIP-sequencing experiment to analyze if and where the different proteins bind nuclear DNA. Preliminary results showed that BCOR binds DNA in both WT and mutated form, but in different regions. In particular, mutated BCOR (and not its WT counterpart) binds inside the first intron of lncRNA-C00379, whose function is currently unknown. For the ATRX-ERG fusion protein, no DNA binding sequences have been found. Further experiments are required for further verification. We set up the *in vitro* culture of human organoids derived from patient-derived Induced Pluripotent Stem cells (IPS) to study the function of the two mutant proteins in an *in vitro* system that better reflects the complexity of human organs. Particular attention is given to the study of osteosarcomas and Ewing's sarcomas, for which the unit has identified a recurrence of genetic defects affecting genes linked to the mechanism of chromosome segregation during the mitotic process. These alterations could explain the high level of chromosomal instability typical of bone sarcomas. The unit is currently engaged in the in-depth analysis of these modifications using third-generation sequencing technology capable of mapping and identifying the variations in the RNA molecules. Preliminary results show that these patients, despite having a reduced mutational load, accumulate a high number of alterations in the RNA. These variations are dependent on the alternative splicing mechanism and give rise to altered RNAs capable of encoding for new epitopes. The study of these variations and their ability to generate aberrant proteins gives the possibility to increase the spectrum of possible immunotherapeutic targets, even in pediatric tumors and with few somatic mutations. The unit is engaged in the identification and classification of these alterations and their possible recurrence in patients with osteosarcoma. In addition, an IT structure was built to automate the processing of samples and the related data flow, including the scientific clinical report provided to the doctor.

Collaborations

- Prof. F. Fagioli (Ospedale Infantile Regina Margherita, Torino, Italy)

Funds and Grants

- Compagnia di San Paolo (PI: F.Fagioli)

Cancer Genomics and Bioinformatics Unit

Research Group

Matteo Cereda, PhD, head of Unit (IIGM)

Liliana Bracco, research fellow (IIGM)

Livia Caizzi, PhD research fellow (IIGM)

Marco Del Giudice, PhD, post-doctoral fellow (IIGM)

Federica Gaudino, PhD, post-doctoral fellow (IIGM) until 30/09/2020

Serena Peirone, PhD student (IIGM and INFN)

Sarah Perrone, PhD student (UniTo)

Francesca Priante, PhD student (IIGM and UniTo)

Alberto Rissone, research fellow (IIGM)

Greta Romano, PhD, post-doctoral fellow (IIGM) until 31/12/2020

Fabiola Varese, research fellow (IIGM) until 31/07/2021

The "Cancer Genomics and Bioinformatics" Unit studies the genomics of cancer using next generation sequencing techniques to characterize the molecular mechanisms underlying the onset and progression of the disease. With the growth of national and international sequencing projects, modern biology is facing new challenges caused by the massive production of genomic data. The main one is to extract relevant information from these data taking into account their intrinsic heterogeneity.

Our unit is positioned in this field and addresses a wide range of biological problems resulting from large-scale genomic, transcriptomic and epigenetic experiments. Through the use of the most recent bioinformatic, statistical and mathematical approaches, our research aims to disarticulate the complexity of genomic data in order to support a personalized medicine capable of being incorporated into clinical practice.

The group's activity focuses on identifying the somatic alterations, defining the clonal evolution of tumors, and identifying new molecular mechanisms that are points of clinical intervention. In particular, our interest focuses on the identification of new oncogenic alterations in tumors characterized by the absence of mutations, structural variations of the genome and transcriptome.

In this context there is the study of the transcriptional regulation of splicing proteins and their role in the onset and progression of cancer in order to discover new potential therapeutic targets. Alternative splicing drives proteome diversity. Alterations of the protein-RNA interaction can lead to a variety of diseases, including cancer. It has also been recently shown that in cancers the protein alterations produced as a result of defects in the splicing mechanism outnumber those produced by somatic mutations (neo-antigens), and for this reason their study can potentially increase the range of action of precision medicine.

Projects

Project 1: "Genomic, transcriptomic and epigenetic data analysis of tumors of different tissues"

Subproject "Deciphering alternative splicing regulation in cancer to identify new therapeutic targets"

M. Cereda (PI), M. Del Giudice, S. Peirone, F. Gaudino, F. Varese

Aims

1. To analyze genomic, transcriptomic and epigenetic data of different tumors in order to identify alterations in tumor transcripts
2. To identify the proteins responsible for the alteration of the splicing mechanism, and their regulators, and the damaged isoforms created, in order to select new therapeutic targets

Results

The group's activity has focused on studying the alteration of biological processes in cancer by applying the GSECA algorithm of transcriptomics Big Data anasyls development last year. By applying this algorithm in 19 types of cancer for which we have expression data for normal tumors and tissues on 16 biological processes that summarize the processing of genetic information, we have shown that RNA splicing is one of the most common sets of genes somatic altered (**Figure 1**).

Figure 1. GSECA analysis of significantly altered processes between tumors and normal counterparts.

We then focused on prostate cancers, investigating whether the deregulation of the factors was driven by other known oncogenes such as FOXA1, ERG, AR, PTEN, MYC and HOXB13. We

hypothesized that RBP deregulation could be controlled by these transcription factors. To test this hypothesis, we stratified the prostate tumor samples (primary and metastatic) based on the expression levels of each factor and applied GSECA. Overall, the expression of splicing proteins is significantly influenced by the expression levels of FOXA1 and AR in primary and metastatic tumors and to a greater extent than those of MYC. To assess which TFs might have a major impact on RBP regulation, we used a modeling approach based on linear regression and measured the relative importance of each transcription factor. FOXA1 expression was the most crucial regressor. GSECA showed a decrease in the expression of splicing factors at the highest expression levels and an increase in the same at low levels. We also investigated whether these transcription factors can influence the expression of splicing proteins by acting on specific genomic positions (**Figure 2**).

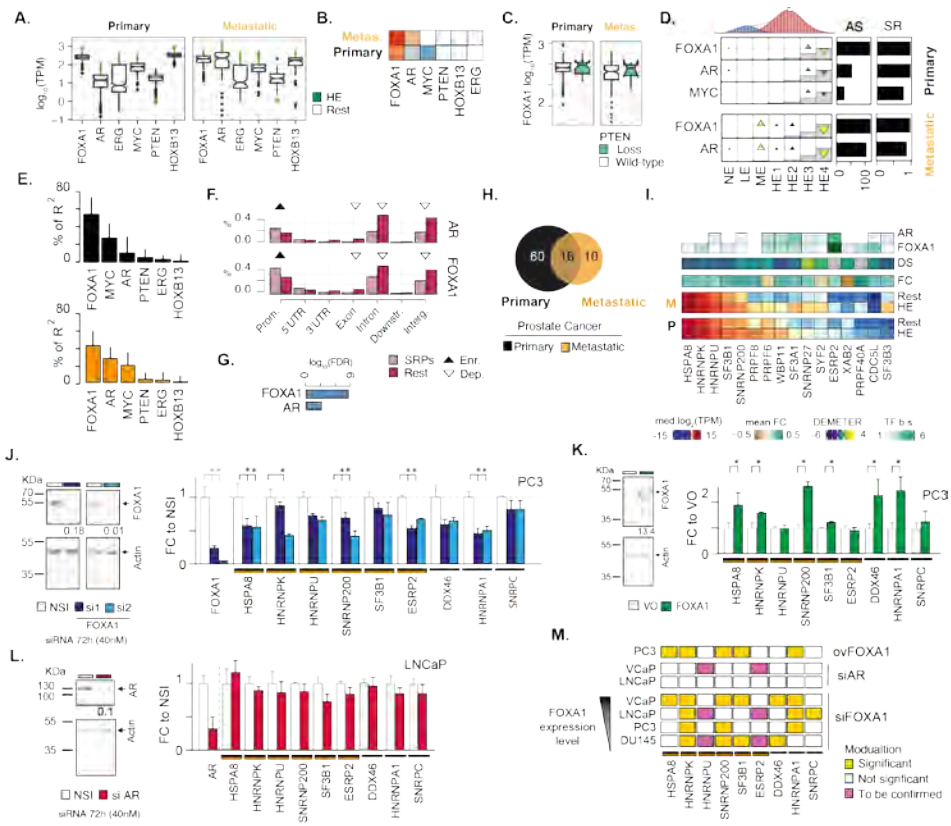


Figure 2. A. Distribution of the expression values of transcription factors. **B.** Contribution of the expression of transcription factors on the expression of splicing proteins. **C.** Contribution of PTEN expression to FOXA1 expression. **D.** GSECA results for the splice pathway. **E.** Relative contribution of the expression of transcription factors on the expression of splicing proteins. **F.** Enrichment of binding sites of transcription factors in the genomic regions of splice proteins. **G.** Enrichment of FOXA1 and AR binding sites in the promoters of splice proteins. **H.** Differentially expressed splicing proteins in prostate tumors. **I.** Expression of candidate splice proteins. **J, K, L.** *In vitro* validation of results: siRNA FOXA1 (J), overexpression FOXA1 (K) and siRNA AR (L). **M.** Results overview.

To do this, we used ChIP and ATAC data for 62 experiments in prostate cancer cell lines and 22 primary tumor samples. The results showed that the FOXA1 and AR peaks were significantly enriched in the promotor region of the splice proteins. Through differential expression analysis we identified 16 FOXA1 deregulated splicing factors, essential for cell line survival, and with FOXA1 binding sites in their promoter region. To validate these results, we measured the expression levels of 9 candidate proteins in vitro using: (1) two independent FOXA1 siRNA duplexes to reduce their expression in 4 prostate cancer cell lines; (2) a siRNA to reduce the expression of AR in VCaP and LNCaP cell lines; and (3) FOXA1 overexpression in PC3 lines. Preliminary results, performed in collaboration with Dr Rajan (Barts Cancer Institute, London, UK), showed a consistent impact of FOXA1 expression on splice proteins, particularly for HNRNPK. AR affected the expression of these to a minor and insignificant extent (**Figure 3**).

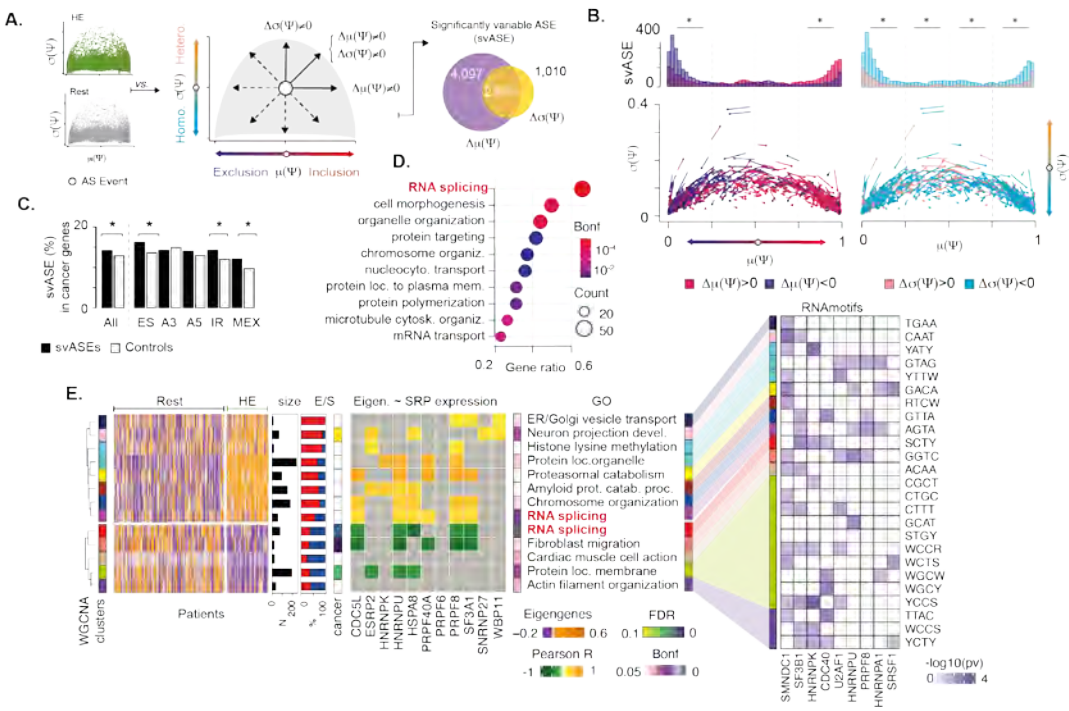


Figure 3. A. Analysis of alternative splice events in primary prostate cancers. **B.** Distributions of the mean and standard deviation of the level of inclusion. **C.** Rates of alternative splice events in cancer genes. **D.** Processes Affected by Alternative Splicing Events. **E.** Identification of binding motifs and associated splicing proteins in their respective biological processes.

To determine how FOXA1 overexpression affects alternative splicing, we analyzed the collection of 242,122 alternative splice events in prostate tumors available from the TCGA portal. We explored

the levels of inclusion of alternative events by identifying 7,121 events alternately excluded or included in FOXA1 overexpressing tumors.

We observed a greater and lesser inclusion of highly and lowly included events by FOXA1, suggesting its action in improving splicing fidelity. Furthermore, these events had a significant impact on cancer driver genes. We found that the genes linked to "RNA splicing" are the most significantly enriched for alternative events, suggesting a possible "feedback" dynamics of the splicing factors. FOXA1 controls the expression of splicing proteins by binding to regulatory regions known as enhancers, whose physical contact with the promoter regions of splicing factors activates their expression. Finally, we examined the binding sites of the splicing factors on their genomic sequences to discover the potential signature of their self-regulation. The preliminary results showed that our strategy could improve the detection of the binding motifs of RBPs whose expression is related to the inclusion levels of alternative splicing events (i.e. HRNRPU) (**Figure 4**).

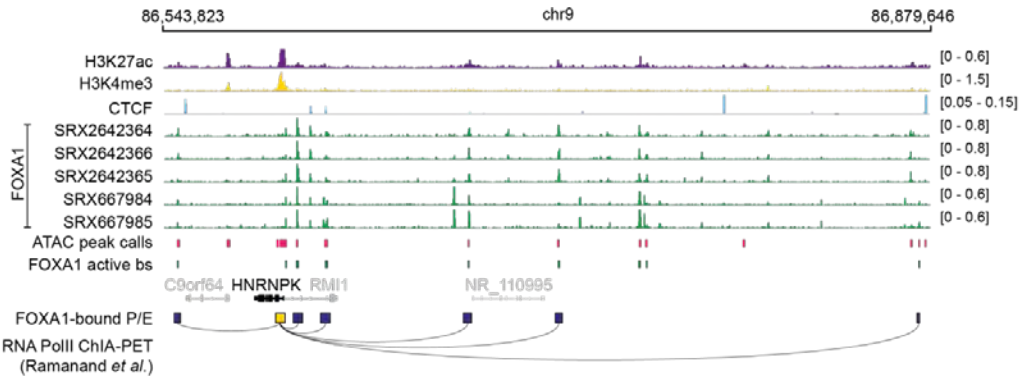


Figure 4. Genome view representing promoter and enhancer regions of HNRNPK bound by FOXA1 in VCaP cells. ChIP-seq density readout traces of H3K27ac, H3K4me3, CTCF (2 overlapping experiments), and FOXA1 in VCaP cells are shown along with ATAC-seq peaks, FOXA1 active binding sites, and RNA PolII ChIA-PET associations FOXA1-HNRNPK. The location of the HNRNPK regulatory regions considered along chromosome 9 is shown at the top of the panel.

We are now working to integrate these steps, formalize the algorithm and perform validation studies similar to GSECA. Based on these data, we tested alternative FOXA1-regulated events in vitro, starting with events contained in cancer genes and splice factors. Preliminary results showed a significant correlation of inclusion levels in prostate turmore cell lines and the primary prostate cancer dataset.

Taken together, these data show that a pioneer transcription factor, FOXA1, controls RBP expression and shapes the landscape of alternative splicing events in prostate cancer. This

recalibration of the alternative splicing process can in turn have an impact on patients' disease relapse. At the moment, the activity focuses on determining which splicing isoforms are more aggressive for patients and analyzing their specific regulatory mechanisms to provide new possible therapeutic strategies. Alternative splicing has been shown to be a therapeutic vulnerability for tumors caused by the MYC oncogene. Our results in prostate cancer propose for the first time the applicability of the same concepts to prostate cancers by providing new candidates of therapeutic interest.

Collaborations

- Prof. J. Ule (Francis Crick Institute, London, UK)
- Dott. P. Rajan (Barts Cancer Institute, Queen Mary University, London, UK)

Subproject "Artificial Intelligence for precision medicine"

M. Cereda (PI), M. Del Giudice, S. Peirone, S. Perrone, F. Priante, F. Varese

Aims

1. To Integrate artificial intelligence systems into research practice in order to improve the diagnosis of cancer subtypes, Identify new biomarkers, select new therapeutic targets

Results

Big Data, High Performance Computing (HPC), and Artificial Intelligence (AI) are becoming increasingly fundamental to healthcare projects. Although in their infancy, research efforts involving the combination of these three aspects are changing the way we look at medicine, from identifying disease risks and taking preventive measures, to diagnosing and personalizing therapeutic treatments. According to the recent report by the European Commission's JRC Science for Policy, venture capital investments in startups dedicated to applying AI in medical practice have reached about €5 billion per year and represent 13% of global investments in healthcare. These financial commitments reflect the interest of the public and private sectors in the application of artificial intelligence in healthcare to address challenges related to improving the quality and life expectancy

of the individual, the aging population and the support of healthcare professionals.

It is now becoming increasingly clear that introducing AI into medical practice will not sideline physicians but instead foster an increase in their strengths. Physicians, supported by AI models derived from Big Data generated from the real world, will be able to focus exclusively on the human elements of their profession, unique qualities that can drive computationally optimized diagnoses and therapeutic strategies.

In this landscape, precision medicine is the new approach to healthcare. In the last decade, the generation of genomic Big Data through genome sequencing, the collection and harmonization of clinical data, and the growth of the potential of bioinformatics have made it possible to recognize the genetic causes responsible for the diseases of individuals and to support the choice of the most effective interventions to improve our health, including personalized medicines (**Figure 5**).

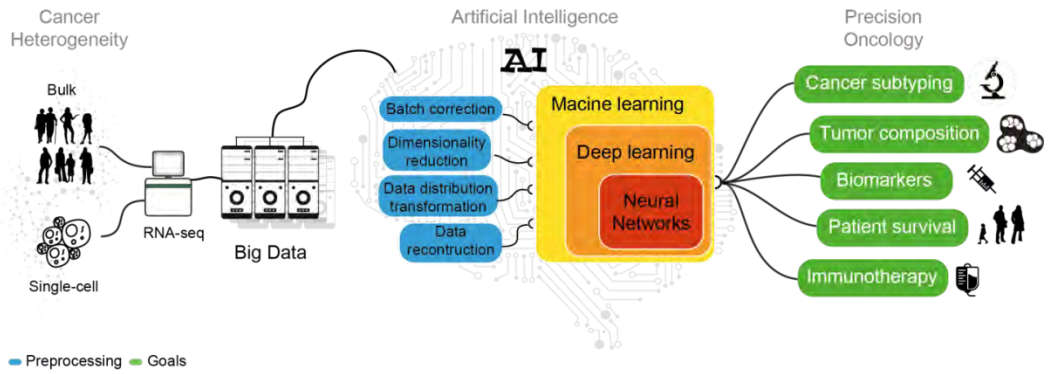


Figure 5. Diagram representing the analyses needed to decipher the heterogeneity present in the genomic data of cancer patients and arrive at effective precision oncology.

Subproject "Identification of alternative splicing dynamics and RNA methylation mechanisms in cancer through third generation sequencing"

M. Cereda (PI), M. Del Giudice, S. Peirone, F. Priante

Aims

2. To characterize the mechanisms of alternative splicing and RNA methylation in cancer through third generation or "long-read" sequencing technologies
3. To identify new therapeutic targets in pediatric sarcomas and prostate cancers

Results

Third generation, or "long-read" sequencing allows for the mapping of RNA molecules in their complete form as they are produced by cells. This technology provides new opportunities to study the complexity of alternative splicing and RNA methylation mechanisms in cancer. The goal of the research is to provide functional insights into alternative splicing (AS) and RNA methylation in pediatric sarcomas and prostate cancers using long-read sequencing. Together with shor-read sequencing, we aim to characterize the role and mechanisms by which proteins control these processes. Long-read sequencing data allows for precision mapping of alternative isoforms, poly-A tails, and RNA methylation modifications (e.g. m6A). This will make it possible to identify new isoforms that, when translated into new proteins, could become precision medicine targets.

We sequenced two pediatric osteosarcomas using the Oxford Nanopore (ON) direct mRNA protocol (~ 1 million molecules). After selecting highly reliable transcripts, we identified ~ 15,000 transcripts, of which ~ 7,500 had not previously been annotated, highlighting the potential to identify novel targets that could generate neoepitopes and drive cancer. A high fraction of these transcripts included new exons that overlapped with repeating elements (Alu and L1) (**Figure 6**).

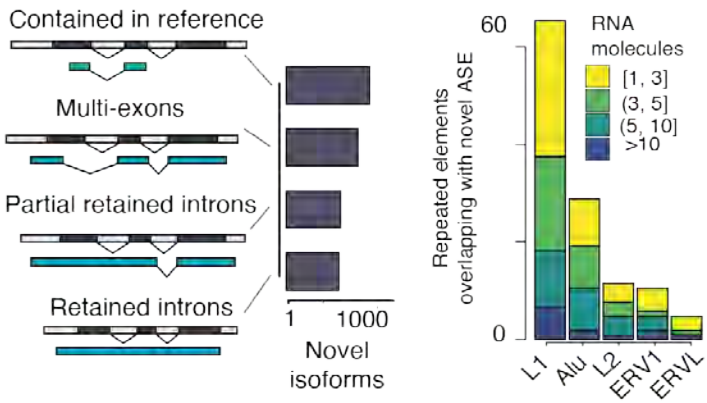


Figure 6. Types of isoforms (left) and number of transposable elements overlapping with new splice junctions (right) identified by long-read sequencing of RNA from osteosarcomas.

Alu and L1 elements have been shown to facilitate the evolution of RNA processing (Attig J, *et al*.,Cell 2019). We are now investigating whether this phenomenon is also associated with cancer cells. We aim to increase the number of molecules sequenced and, therefore, the number of cells.

At the same time, we began to evaluate the effect of the transcription factor FOXA1 on RNA methylation in cellular models of metastatic prostate cancer. FOXA1 works as a pioneering transcription factor by facilitating access to chromatin for steroid hormone receptors, such as the androgen receptor and estrogen receptor. Although its contribution to DNA methylation processes is widely studied, its impact on RNA methylation processes remains largely unknown. N6-methyladenosine (m6A), the most abundant internal modification of RNA in eukaryotic cells, has gained increasing attention in recent years. The m6A modification affects multiple aspects of RNA metabolism, ranging from RNA processing, nuclear export, RNA translation to decay. Emerging evidence suggests that m6A methylation plays a critical role in cancer through various mechanisms. The critical role of m6A RNA methylation in cancer initiation and progression offers new possibilities for early diagnosis and treatment of cancers.

We recently demonstrated how FOXA1 regulates the expression of splicing proteins in primary and metastatic prostate tumors, validating the mechanism on cellular models. To obtain further information on the impact of FOXA1 on RNA regulation processes, we performed ON sequencing experiments on mRNA extracted from FOXA1-depleted PC3 prostate cell lines and related controls. Overall, we obtained ~ 3 million (M) full-length RNA molecules for the FOXA and non-silenced (NSI) siRNA samples, corresponding to ~ 1.6 M RNA molecules (~ 36,000 isoforms) with median length of ~ 1,700 bp. We then measured the fold change (FC) in the expression of transcription factors, known prostate tumor oncogenes, and FOXA1-controlled splicing factors, confirming the modulation of FOXA1 gene expression (**Figure 7**).

The same samples were subjected to Illumina sequencing, yielding comparable results with ON sequencing. From the ON sequencing data we were able to map the spectrum of full-length RNA isoforms present in cells, finding greater variability in the use of the isoforms (i.e. a greater number of multi-exons, retained intronic isoforms and included exons) in the absence of FOXA1. Importantly, we observed an overall increase in m6A levels in siFOXA1 samples compared to NSI. Furthermore, candidate splice proteins exhibit higher m6 A methylation levels in the absence of FOXA1.

These results suggest a possible role of m6A RNA methylation in the regulation of alternative splicing operated by FOXA1, whose expression could modulate the transcription of methyltransferases (e.g. METTL3, METTL4). Overall, m6A methylation may play a role in controlling FOXA1 AS rewiring.

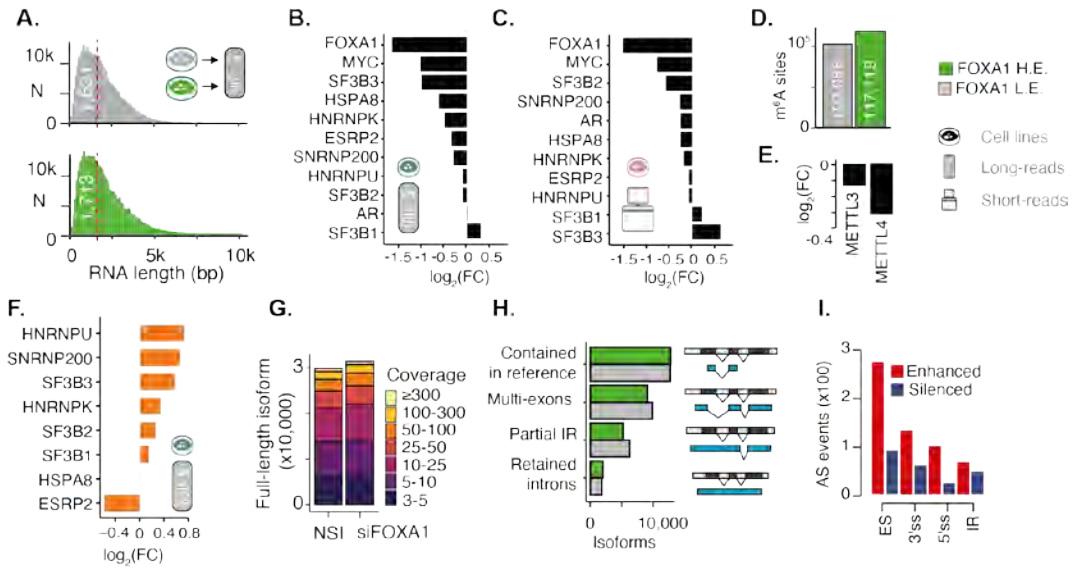


Figure 7. Results of ON sequencing of direct mRNA on FOXA1 and non-silenced (NSI) siRNA samples in PC3 cell lines. **A.** Distribution of the lengths of the sequenced RNAs. **B,C.** Fold-change of the expression of candidate genes measured by ON (B) and Illumina (C) sequencing. **D.** Number of methylated sites. **E.** Fold-change of methyltransferase. **F.** Fold-change of the number of methylated sites between siFOXA1 and NSI. **G.** Coverage of the isoforms. **H.** Variability of the type of isoforms identified. **I.** Alternative splice events.

Subproject "Development of algorithms for the characterization of molecular mechanisms responsible for the onset and progression of single cell cancer"

M. Cereda (PI), S. Peirone, S. Perrone

Aims

To identify new methodologies to extract relevant information on the origin and progression of diseases from single cell mass sequencing

Results

Through the application of artificial intelligence techniques, we have developed a new algorithm capable of identifying biological processes related to the onset and progression of tumors using tumor transcriptome sequencing data. The developed method performs better than reference algorithms for the scientific community. Applied to 5,941 samples of 14 cancers, our method correctly identified the alterations in the signaling pathways specific to each tumor. Furthermore, our results have allowed to highlight the role of the transcription factor PTEN in the modulation of immune processes. In particular, we have clarified that prostate tumors that show a loss of PTEN

have an immunosuppressive microenvironment, due to the activation of STAT3.

With the growth of high-throughput sequencing projects, modern biology is facing new possible limitations due to big data issues. One of the challenges is to extract relevant information from this high-volume data, taking into account their inherent heterogeneity. So far, genomic projections have profiled thousands of samples providing insights into the cell's transcriptome. However, unraveling the heterogeneity of these transcriptomic Big Data to identify faulty biological processes remains difficult. The application of NGS technologies at the single cell level has made it possible to study cellular heterogeneity, hitherto hidden in RNA sequencing experiments (RNA-seq) at the population level (bulk). Starting from GSECA, an algorithm developed by our unit to study bimodal behavior of RNA sequencing gene expression profiles in order to identify altered biological processes in heterogeneous patient cohorts, the unit applied the expression profile method single cell gene. By increasing the signal-to-noise ratio, GSECA is able to successfully manage the heterogeneity of thousands of samples and provides useful information on the clinical and biological patterns of a phenotype. In particular, RNA-seq data of 3,589 cells of Glioblastoma (GBM) samples from 4 patients were analyzed. The study of expression profiles through the use of GSECA demonstrates the existence, even at the single cell level, of two classes of gene expression: a group of poorly expressed genes (LE) and one of highly expressed genes (HE). To verify the presence of specific mechanisms underlying bimodality, a null model was devised for the generation of simulated cells, characterized by zero correlation between the expression of genes, to be used as a comparison for real cells (**Figure 8 A**).

In order to give a biological meaning to the existence of the two observed expression classes, the genes that populate them have been characterized (**Figure 8 B**). GBM samples have a stratified structure, which causes the cells found in its innermost parts to suffer from hypoxia (lack of oxygen). For this reason, particular attention has been paid to the study of genes whose expression is related to a hypoxic environment and to genes related to splicing, a process known to be influenced by hypoxia.

With this preliminary work we have shown that the "less is more" paradigm in the treatment of large heterogeneous RNA-seq data sets is also valid for single cell data, demonstrating that even with these types of data it improves the detection of processes biological alterations in the phenotype of interest.

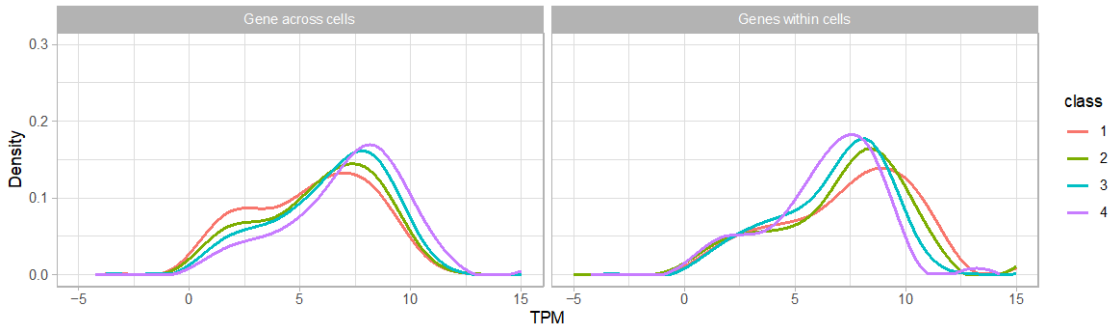


Figure 8A. Bimodal distribution of the gene expression values of a gene among cells (left) and of the genes of a cell (right) in the transcriptomic data of GBM cells. The categories correspond to discretizations of the number of cells and genes expressed in the cell, respectively.

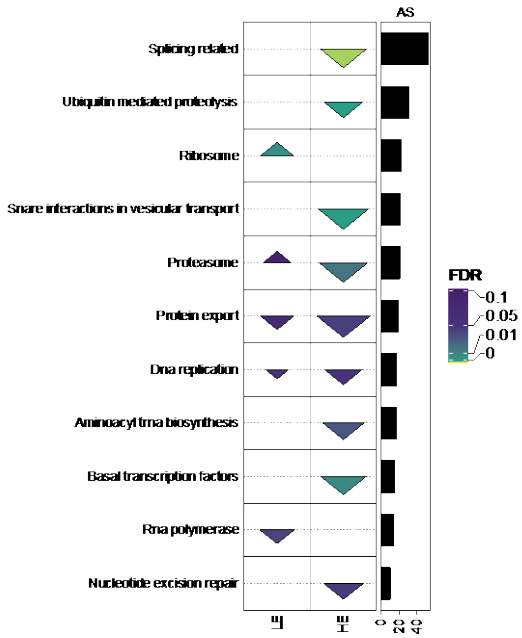


Figure 8B. Application of GSECA to GBM data. GSECA showing the altered expression signatures as a result of the onset of GBM.

Collaborations

- Prof. M. Caselle (Dipartimento di Fisica Teorica, Università degli Studi di Torino, Torino, Italy)
- Dott. Luca Persano (Fondazione Istituto di Ricerca Pediatrica, Padova, Italy)

Subproject "Analysis of heterogeneity and clonal evolution of tumors to identify new points of therapeutic intervention in neoplasms"

M. Cereda (PI), M. Del Giudice, S. Perrone, S. Peirone, G. Romano, F. Gaudino, F. Varese

Aims

1. To create a catalog of tumor evolution models that allow the characterization of cancer driver genes, the classification of clones and sub-clonal events responsible for drug resistance and the understanding of the interaction between clonal and sub-clonal events in resistance to drugs
2. To experimentally evaluate the predictions obtained on tumor samples and models through mass sequencing techniques
3. To identify and quantify the predicted acquired drug resistance effect within and between various cancers

Results

The study systematically evaluates the clonal evolution within and between different types of cancer using DNA and RNA NGS data from over 20,000 patients from sequencing consortia (TCGA). The unit has cataloged the tumor heterogeneity of these samples in a limited number of evolutionary models in order to offer information on the different selective advantages acquired by the tumor during its growth (**Figure 9**). It was therefore possible to group the tumor according to their evolution and categorize the cancer genes according to their selective advantage. Currently, the unit is validating the predictions obtained on sarcomas on sequenced osteosarcoma samples within the SARGEN project. The results converge towards clonal and subclonal alterations in the ritosin kinase receptor pathway of the pediatric patients analyzed, with notable and frequent alterations of the TP53 gene. The combined analysis of mutations, chromosomal alterations and fusions in the RNA made it possible to identify a substantial number of candidate genes for personalized therapies.

The unit is currently engaged in the identification of those pathogenic variants and which may be the subject of a new precision medicine.

The validations of the pathogenic variants are underway through the Oxford Nanopore third generation sequencing technology on cell lines derived from pediatric patients.

Collaborations

- Prof. F. Fagioli (Ospedale Infantile Regina Margherita, Torino, Italy)

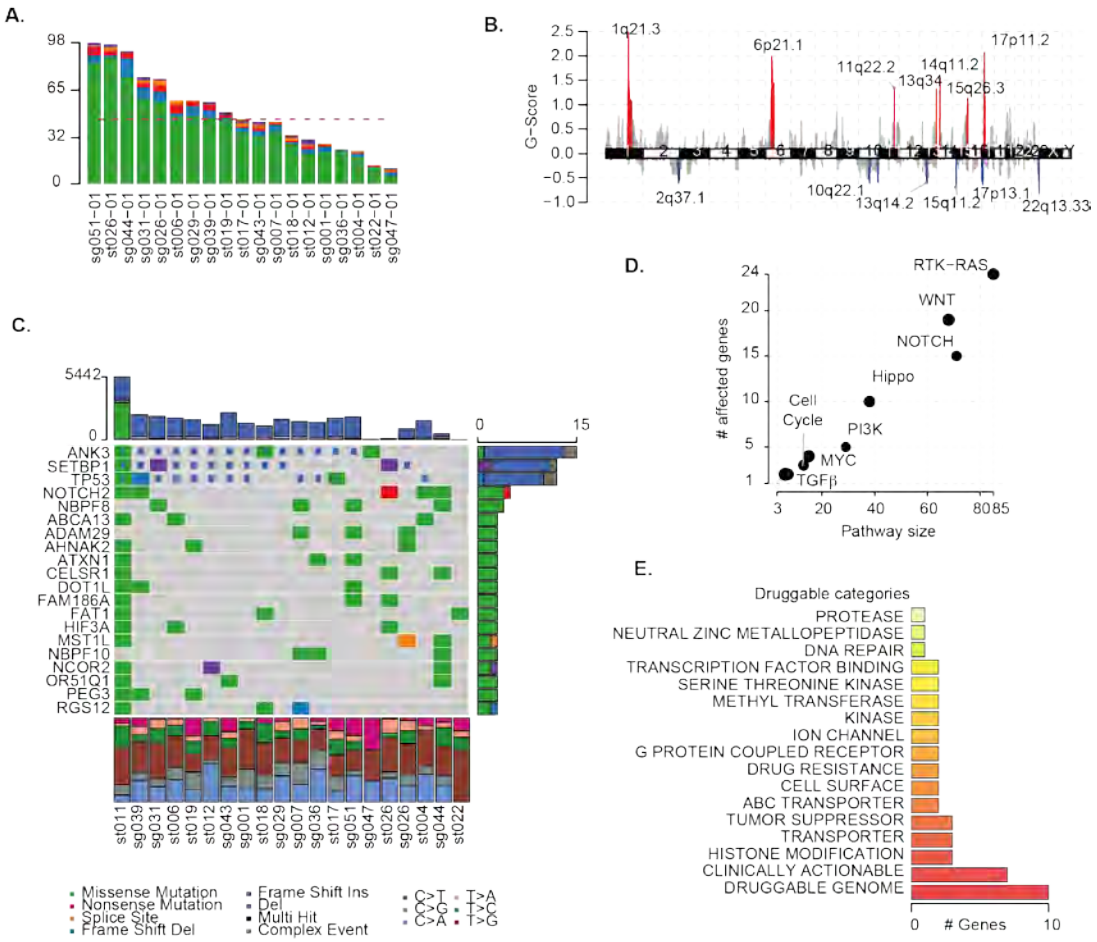


Figure 9. A. Number of somatic mutations in pediatric patients with osteosarcoma. **B.** Recurrent chromosomal changes in samples. **C.** Recurrently altered genes and their distribution in the analyzed osteosarcoma samples. **D.** Mainly altered pathways. **E.** Categories of genes subject to pharmacotherapy and possible targets in the analyzed samples.

Subproject "Characterization of the resistome of chemotherapy tolerant medulloblastomas through single cell RNA sequencing"

M. Cereda (PI), S. Perrone, S. Peirone

Aims

1. To characterize tumor evolution in therapy-resistant medulloblastomas
2. To identify the tumor clones responsible for resistance to therapy

Results

Several drug resistance mechanisms allow tumors to evade treatment. In this context, medulloblastoma (MB), the deadliest pediatric brain tumor, is highly heterogeneous. The identification of the cellular identities underlying resistance to MB therapy remains unknown. The unit will study the evolution of MB cells induced by chemotherapy through transcriptomic analysis at the level of a single cell and cell population. RNA-seq data from single cells and cell populations will be exploited to profile transcriptome expression levels, map alternative splice isoforms, identify allele-specific expressed mutations and infer copy number variants. The clonal evolution of MBs will be reconstructed by adapting the published approaches (Cereda M, *et al*., Nature Communications 2016). Clonal trajectories will be analyzed to identify carcinogenic events and drug resistant clones. Machine learning and network theory approaches will be used to prioritize the alterations responsible for acquired resistance and select those that generate new epitopes. Deregulated biological processes will be identified by applying model-based methods (Lauria A, *et al*., Nucleic Acids Res 2020). Data available from international consortia (MAGIC, St. Jude) will be integrated to corroborate the results. By applying machine learning and network theory analyzes, we expect to characterize the transcriptional landscape of MBs and to highlight the mechanism leading to MB drug resistance. In particular, we expect to obtain functional insights into the role of hypoxia in the rewiring of RNA processing and its impact on clonal evolution of MB. Currently, we have performed bulk sequencing of mRNA extracted from pediatric 14 MB. After applying rigorous quality controls, analysis of gene expression levels revealed that the MBs grouped according to their degree of hypoxia. Our preliminary results indicate that this condition has a significant impact on RNA processing.

Collaborations

- Dott L. Persano (Fondazione Istituto di Ricerca Pediatrica, Padova, Italy).

Subproject "Study of the immune microenvironment in breast cancer in order to evaluate its effects on therapeutic treatments and identify new vulnerabilities"

M. Cereda (PI), S. Peirone, S. Perrone

Aims

1. To analyze genomic and transcriptomic data of breast tumors in order to characterize the immune microenvironment
2. To Identify gene expression signatures and related markers that explain the type of microenvironment of each tumor
3. To experimentally validate the action of the identified markers and their contribution in immunotherapeutic treatments

Results

In vivo apoptosis of mesenchymal stromal cells (MSCs) plays a critical role in providing immunomodulation. Although the efferocytosis of apoptotic MSCs is a critical step in converting phagocytes into immunosuppressive cells, it is not known whether apoptotic MSCs directly impose any additional pressures on immune cells during apoptosis. To answer this question, in collaboration with Prof. Dazzi's laboratory at King's College London, we analyzed the transcriptomic profile of MSCs undergoing Fas-induced apoptosis. We detected a variety of immunomodulatory factors and chemokines regulated by the canonical NFκB pathway under activation of caspases. The secretome of apoptotic MSCs demonstrated robust chemotactic and immunosuppressive effects on primary human immune cells. The critical role of PGE2 was confirmed in a cohort of patients receiving MSC treatment for Crohn's disease. Both PGE2 and apoptosis levels were significantly correlated with their clinical responses to MSCs. The unit then analyzed transcriptomic data through statistical and artificial intelligence techniques for 982 tumor samples and 89 healthy counterparts from international consortia in order to evaluate the pro or anti-tumor activity of PTGS2.

The study identifies a novel mechanism by which caspase activation provides immunosuppressive activity independently of the classical pathway of efferocytosis.

Collaborations

- Prof. F. Dazzi (King's College London, London, UK)
- Prof.ssa F. Ciccarelli (Francis Crick Institute, London, UK)

Funds and Grants

- Ministero della Salute, Ricerca Finalizzata 2019, Progetto Giovani Ricercatori “Targeting alternative splicing neo-junctions as a novel source of neo-antigens in pediatric and adult tumors” 2021-2023 (PI: M.Cereda)
- Fondazione Cassa di Risparmio di Padova e Rovigo ”Deconvolution of medulloblastoma resistome through high-resolution tracking of drug-tolerant subclones” 2020-2022 (PI: L. Persano, Partner: M. Cereda)

Immunotherapy Unit

Research Group

Tobias Longin Haas, head of Unit (IIGM), assistant professor (UCSC)

Alessandro Abbati, fellow technician (IIGM) until 31/07/2021

Jolanda Magri, PhD student (UCSC)

Matteo Menotti, PhD, post-doctoral fellow (IIGM)

The Immunotherapy unit was established at IIGM at the beginning of 2020 and dedicates its activities by generating novel strategies to support the natural immunity in fighting and eradicating solid tumors.

In the last years it became evident that supporting the anti-tumor activities of the immune systems can provide a very effective weapon in the fight against cancer. Starting from this assumption, a new type of drugs has been developed, the immune checkpoint inhibitors, these drugs can unleash forces of T cells that are usually suppressed by the tumor or by the tumor microenvironment. These drugs resulted in remarkable clinical outcome in several solid tumor entities including lung cancer and melanoma. Currently many clinical trials are ongoing to evaluate the therapeutic efficacy against other solid tumors or in combination with other treatment schedules.

One of the main findings of these studies was that tumors with a high mutational load that have a immunologically active (or immunologically hot) microenvironment are more susceptible to this kind of treatment, while tumors with lower mutation rates, containing few infiltrated immune cells have an immunologically cold tumor microenvironment and do not respond as well to this treatment. The aim of this project is to turn an immunologically cold tumor (e.g. colorectal cancer (CRC)) into an immunologically hot one.

The main strategies to reach this aim include the generation of novel “bispecific T cell engagers” (BiTEs) and innovative fourth generation chimeric antigen receptors (CARs) targeting the tumor cells.

Projects

Project 1: "Immunotherapy"

Subproject "Generation and validation of a novel BiTEs antibodies"

T. L. Haas (PI), M. Menotti, A. Abbati

Before starting this project, we have generated antibodies showing high tumor selectivity for CRC, in collaboration with the Catholic University of Rome (UCSC). These mAbs are able to bind the cancer tissue of several patients while show no interactions to non-transformed healthy colon mucosa of the same patient as well as to other vital tissues. The hybridomas used for the production of those antibodies were 2 times subcloned using FACS assisted single cell laydown, mRNA was extracted, reverse transcribed and the variable regions (VH, Vk) of the tumor specific mAbs were amplified by PCR. The PCR products were cloned and sequenced by Sanger sequencing, the functional regions were identified in silico and artificial genes were designed and used for production of the BiTEs.

Aims

1. To design, manufacture and purify bispecific antibodies (BiTEs) capable of activating human or murine lymphocytes
2. To activate the immunological potential of T lymphocytes against tumor cells by using BiTEs
3. To test the ability of antibodies to activate T lymphocytes against tumor cells *in vitro* and *in vivo*

Results

Using recombinant DNA technologies we generate sequences that codify for BiTE antibodies, this subtype of antibodies are able to crosslink the endogenous T cell receptor complexes in the presence of tumor cells expressing a respective tumor associated antigen (TAA, **Figure 10**) leading to the activation of the lymphocyte. The BiTE designed and generated at IIGM consist of two monoclonal antibody (mAb) derived antigen-binding domains, that are expressed as single chain fragments of the variable chains (scFvs) linked by a flexible amino acid sequence. The T cells activation is obtained by the binding of the BiTE to the CD3 epsilon receptor (CD3e), within the T-cell receptor (TCR) complex. Once BiTE has bound both the tumor cell and the T cell via CD3e the lymphocyte will be activated independently of the major histocompatibility complex and the specificity of its TCR receptor, by the creation of an immunological synapse. This activation will cause the release by the lymphocyte of the enzymes perforin and granzyme B and the subsequent apoptotic death of the tumor cell.

In order to obtain a stronger T-cells activation we have also designed a second BiTE to be used in

combination with the previous one that has, instead of the CD3e scFv, another sequence that can bind and activate the CD28 co-receptor on T lymphocyte. Using this second BiTE will allow us to support the activation via CD3e and also to activate naïve T-cells, that are usually not activated by a single stimulation.

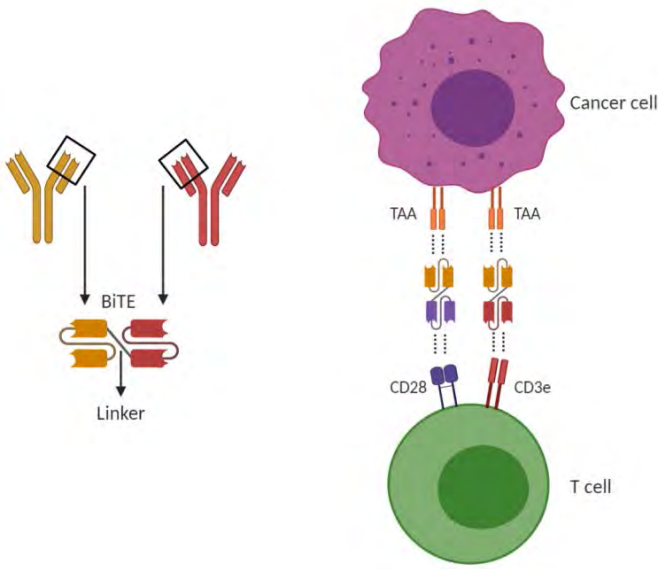


Figure 10. Structure of a BiTE and principle of it's function.

We tested the affinity of our BiTEs for both tumor antigen and T-cell receptors and evaluated the ability of our antibodies to activate lymphocytes in the presence of tumor cells; to do this we measured the expression of some surface markers that are normally over-expressed when T cells are active: CD69 and CD25. T cells express a higher amount of both markers when cultured with SA115 primary colon cancer cells and BiTE CD3 (**Figure 11 A**, center), the second BiTE CD28 alone does not affect lymphocyte activation (left), but amplifies the effect of BiTE CD3 when combined with it (right). As mentioned above, the ultimate goal of lymphocyte activation by BiTEs is the elimination of tumor cells. To assess this aspect we used CRC tumor cells engineered to express the luminescent protein luciferase, so doing we could quantify the amount of cells present in our co-culture simply by measuring the intensity of the luminescent signal given by this protein.

As can be seen in **Figure 11 B** the amount of cells expressing luciferase is extremely reduced in the presence of both BiTE, we can therefore deduce that lymphocytes, in addition to being activated in the presence of cancer cells, have also carried out their cytotoxic activity by eliminating the targets.

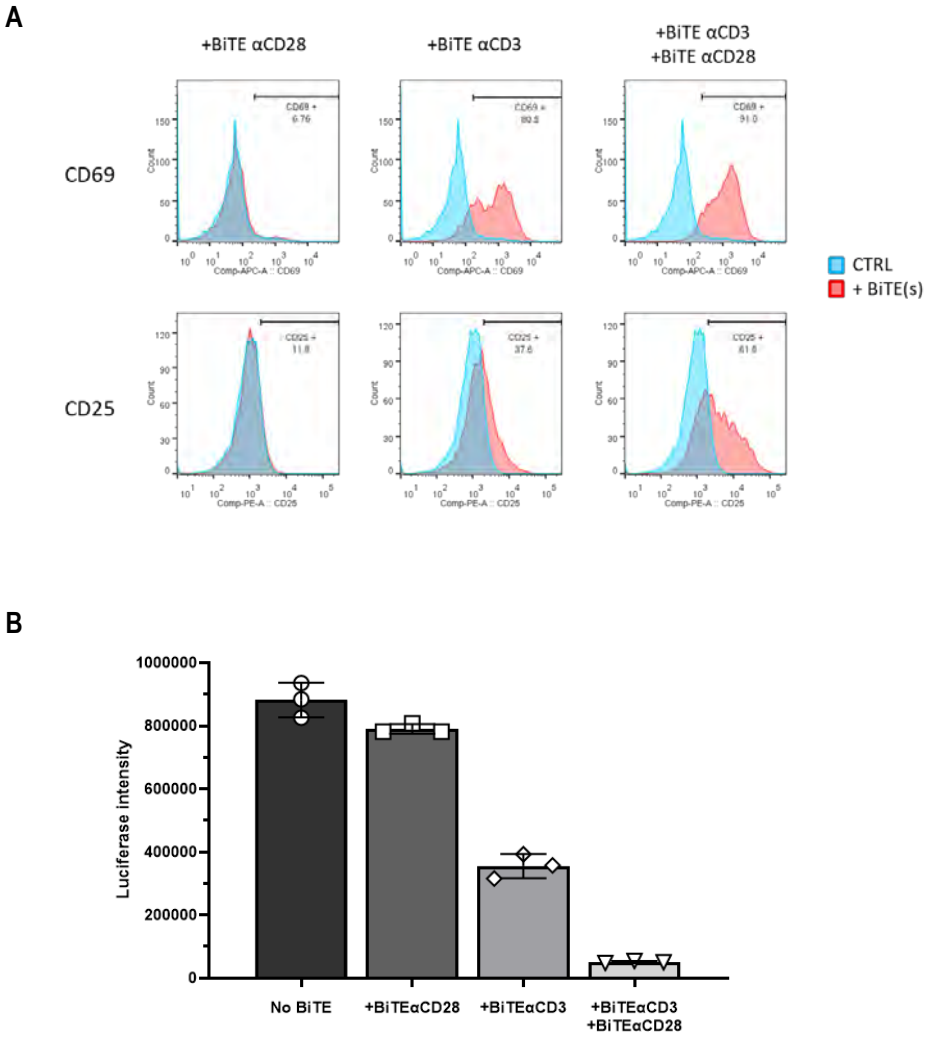


Figure 11. A. Analysis of BiTE binding capacity to tumor cells and lymphocytes. **B.** amount of luciferase-expressing cells in the presence of BiTE.

The next steps of this project will be to evaluate the ability of BiTEs to perform their function also in vivo using immunocompromised mice with subcutaneous tumors. In addition, to better exploit the ability of our antibodies to recognize the same tumor antigen both human and murine, we are currently designing two other BiTEs able to activate murine lymphocytes, we will use these antibodies to test their preclinical validity in an immuno-competent mouse model.

Subproject "Generation of 4th generation CARs targeting TAAs in solid tumors"

T. L. Haas (PI), M. Menotti, A. Abbati

This project is based on the recent finding that vaccination approaches targeting tumor-associated antigens in solid tumors can help to break the immunological tolerance. The use of this strategies has revealed anyhow a number of limitations, due to tumor heterogeneity, immune escape mechanisms and immunosuppression in the tumor microenvironment that currently impair cancer vaccines employment in clinics.

It was shown that adjuvants or intratumoral administration of immunomodulators can improve immunogenicity of cancer vaccines, even if only few adjuvants are licensed for clinical use and intratumoral injection is not always feasible, considering metastatic disease and patient compliance.

Adoptive cell transfer of CAR-T cells, which are able to recognize the target antigen through the engagement of an engineered T cell receptor in a MHC-independent manner, represents an innovative immunotherapeutic approach, as they help avoiding immune escape mechanisms and overcoming central T cell tolerance, leading to direct cancer cell lysis.

The idea behind this project is to create a 4th generation CAR-T cells that can enhance the normal effect of this therapy, adding to their ability to recognize tumor cells, the use of adjuvants in a single treatment which has the potential to overcome most of the limitations of immunotherapy in solid tumors. Moreover, such hypothesis will be tested in immunocompetent mice, allowing understanding of CAR-T cell therapy effects on the tumor microenvironment, unlike most studies in which immunodeficient mouse models are exploited for CAR-T cells engineered against human target antigens.

For this purpose we're currently producing CAR-T cells that, following their activation, not only cause the direct lysis of the target cell, but also activate the production of a modified version of a toll like receptor 5 (TLR5) ligand leading to the activation of the innate immune system.

Our hypothesis is that this construct will get expressed and secreted by CAR-T cells when they infiltrate the tumor mass and get in contact with the specific TAAs. In turn the TLR5 ligand will then activate innate immune cells present in the tumor microenvironment triggering the typical mechanisms of an initial immune response, such as inflammation and recruitment of other immune cells. Once activated, these cells will migrate into the lymph nodes to present new peptides derived from tumor antigens released following cancer cell lysis by CAR-T and this will result in the production of specific T cells against these antigens. These neo-synthesized T cells will subsequently migrate back into the tumor, as well as into possible metastases, contributing to their

elimination (**Figure 12**).

Given the fast-moving field of CAR T cell therapies, we are confident that in case we see efficacy *in vitro* and in animal models, this treatment regime can also be translated into the clinical setting.

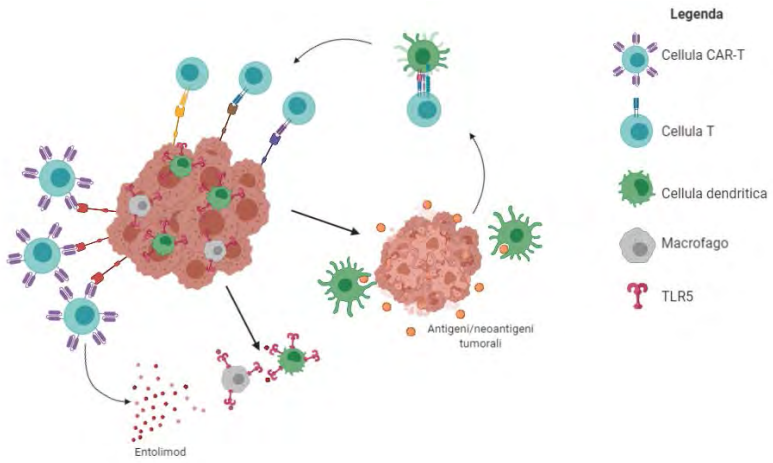


Figure 12. Scheme of the hypothesis

Aims

1. Generation of fourth generation CAR-T cells
2. Selection of the best promoter that allows secretion of Entolimod protein upon CAR-T cells activation
3. Validation of the antitumor efficacy of CAR-T cells in vitro and in vivo

Results

In order to achieve the production and secretion from T cells of the entolimod protein, ligand of TLR5, we designed several DNA constructs containing sequences of signal peptides for export from the cell membrane, or so-called "cell penetrating peptides" (CPPs) capable of passing through the cellular lipid membrane, together with entolimod. These constructs were cloned into a plasmid vector under a constitutive promoter. Of the seven constructs, three have the immunoglobulin K signal peptide (IgK SP) and the other four have different CPPs, namely transportan, an amino acid repeat of arginine (polyR) and a protein sequence belonging to homeoproteins termed SecPen (Secretin-Penetratin).

We transfected HEK293T cells with all constructs and the supernatants, containing entolimod, were used to control its production by a TLR5 stimulation assay using engineered cells capable of emitting

a signal upon TLR5 binding and activation. These results showed that each combination can stimulate TLR5 at different levels, in some cases comparable to the stimulation induced by flagellin, a bacterial protein that is a natural ligand of TLR5 (**Figure 13 A**).

By Western blot, however, we observed a different production and secretion in the supernatants of the protein between the various constructs, which is expressed in a higher molecular weight glycosylated form (a process dependent on the presence of the IgK SP) or in its native lower molecular weight form (in the presence of CPPs or in the "no gly" form deprived of sites subject to glycosylation) (**Figure 13 B**). From these experiments, it appears that the best combinations in terms of TLR5 production and stimulation activity are represented by the IC10, IC11, and IC22 constructs.

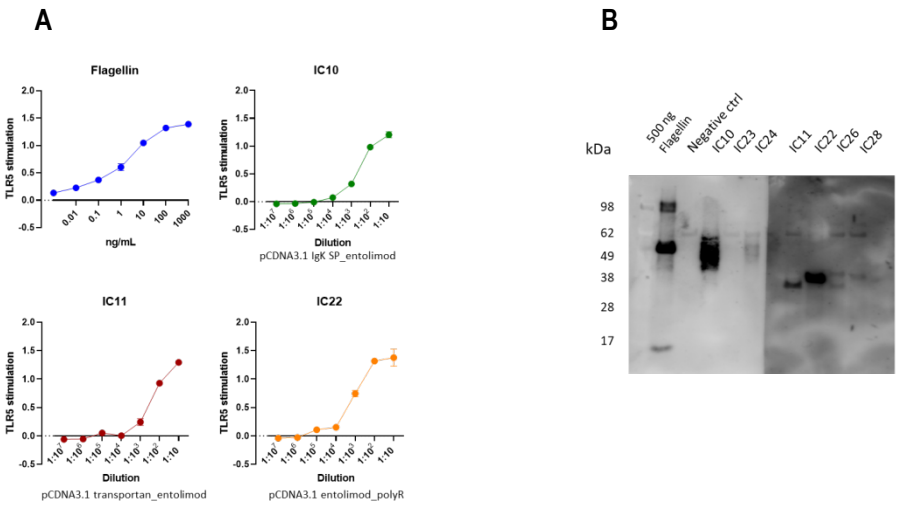


Figure 13. A. Activity and production of entolimod protein: graphs represent the activity of the top 3 constructs together with flagellin, used as a control for TLR5 stimulation. **B.** Western blot image indicates production and secretion in supernatants of entolimod.

The next step is to test the ability of an inducible promoter that can function only after CAR-T cell activation and drive entolimod production. Currently, a promoter containing 4 or 6 repeats of the NFAT transcription factor is being produced and tested.

On the other hand, the antigen chosen for the study of our hypothesis is the Epidermal Growth Factor Receptor (EGFR), which is found to be over expressed in several cancer types, including colo-rectal carcinomas. Given the need for an immunocompetent mouse model, in which to study the effects of CAR-T on the tumor microenvironment, we chose to express a truncated form of human EGFR, corresponding to the extracellular portion and lacking the tyrosine kinase domain, in MC38

cells of colorectal carcinoma syngeneic C57BL/6 mice, which will be used as target cells in subsequent experiments, thus having a model with an intact immune system but without the problem of tumor rejection.

The variable portion of the commercial antibody Cetuximab, able to recognize trEGFR, will be used to build part of the chimeric receptor expressed on T cells. This portion, called scFv, was initially produced by cloning it into a construct expressing the human Fc portion of immunoglobulin G, in order to test the antigen binding capacity. The antibody thus obtained is able to bind trEGFR on clones and not on control cells, at levels comparable to those of cetuximab and with good affinity. For this reason, the antibody was cloned into a vector containing a second-generation human-type CAR (CD8, 4-1BB, CD3z), currently being produced in human Jurkat T cells, to test the ability of CAR-Ts to bind to trEGFR, their activation and subsequent cytokine production.

All the elements analyzed and necessary for the realization of the fourth generation construct will then be inserted into a single lentiviral vector.

The CAR-Ts thus produced will be tested both in vitro and in vivo for their direct activity against tumor cells and for their ability to activate host immune defenses and provide long-lasting protection.

Collaborations

- Prof. R. De Maria (Università Cattolica del Sacro Cuore (UCSC), Fondazione Policlinico Universitario Agostino Gemelli (FPG), Rome, Italy)
- Dr. E. Ponterio (Fondazione Policlinico Universitario Agostino Gemelli (FPG), Rome, Italy)
- Dr. G. Sette (Fondazione Policlinico Universitario Agostino Gemelli (FPG), Rome, Italy)
- Dr. F. Guidi (Università Cattolica del Sacro Cuore (UCSC), Rome, Italy)

Project 2: “COVID”

T. L. Haas (PI), M. Menotti, A. Abbati

Aims

1. To evaluate the neutralization capacity of the antibodies produced following the anti-SARS-Cov-2 vaccine in patients previously infected by this virus comparing them with healthy vaccinated individuals
2. Selection of the best promoter that allows secretion of Entolimod protein upon CAR-T cells activatio To analyze the functionality of anti-SARS-Cov-2 antibodies produced by different donors on the main variants of the virus

Results

In collaboration with the Oncology Hospital of Candiolo, in the context of the "MemoryCoVax" project, sera were collected from 40 individuals previously recovered from SARS-Cov-2 infection (who tested positive to a molecular test at the time of infection) and from as many individuals who never tested positive and will be used as controls. Samples were collected at time zero, i.e., before the first dose of BNT162b2 vaccine (Pfizer-BioNTech), then following the first and second doses, and one, three, and six months after the last dose.

We assessed the total amount of antibodies produced by donors following infection and/or vaccination; to do so, we cloned the coding sequences for the Spike protein and the "Receptor Binding Domain" (RBD) of the SARS-Cov-2 virus into an expression vector used to produce these constructs on a large scale, which were subsequently purified by FPLC liquid chromatography (**Figure 14 A**). These two are surface proteins of the SARS-Cov-2 virus and thus are the ones against which most antibodies are normally produced.

The Spike and RBD proteins were used to be able to analyze, by ELISA assay, the relative amount of antibodies able to bind these two proteins. As can be seen in **Figure 14 B**, most of the former COVID patients (in red) developed antibodies against both proteins of the virus (RBD in the figure), in some cases in amounts comparable with those produced by healthy individuals after two doses of vaccine (yellow).

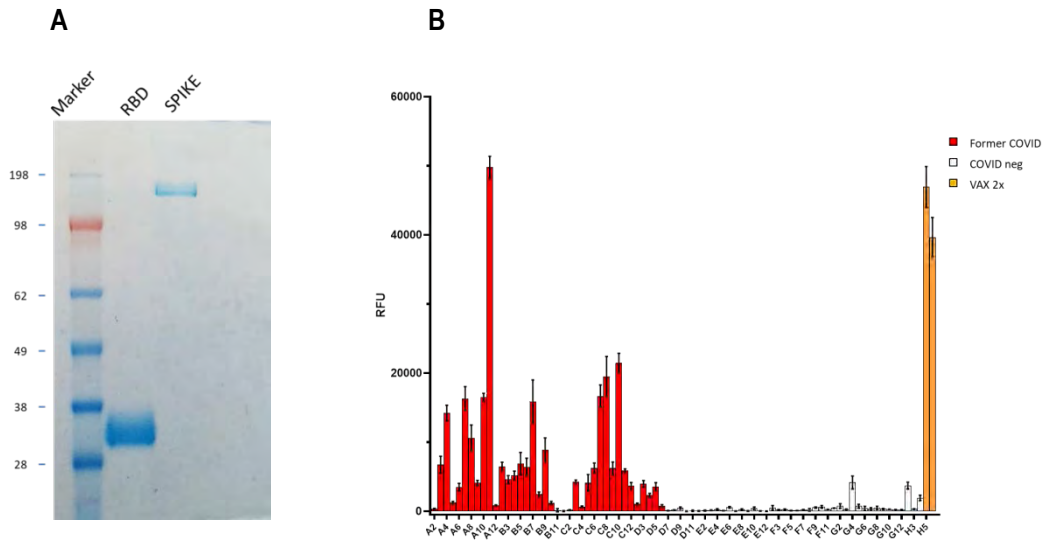


Figure 14. A. Evaluation of the amount and purity of RBD and Spike proteins by Coomassie Blue staining. **B.** Amount of anti-RBD antibodies present in the blood of individuals recruited for the study at time zero, before the first dose of vaccine.

In order to evaluate the potential of antibodies to neutralize viral particles, constructs have been devised for the production of artificial viruses that contain the genetic information for the expression of the luminescent protein luciferase and, on the surface of the viral capsid, express the Spike protein of Covid; These viruses are then able to infect cells expressing the human ACE2 (Angiotensin-converting enzyme 2) receptor, the target of the Spike protein, and thus induce the target cell to produce luciferase. We have therefore produced human Calu-1 cells that ectopically express ACE2 and can be detected by luminescence assays only if they are infected with Spike virus (**Figure 15**).

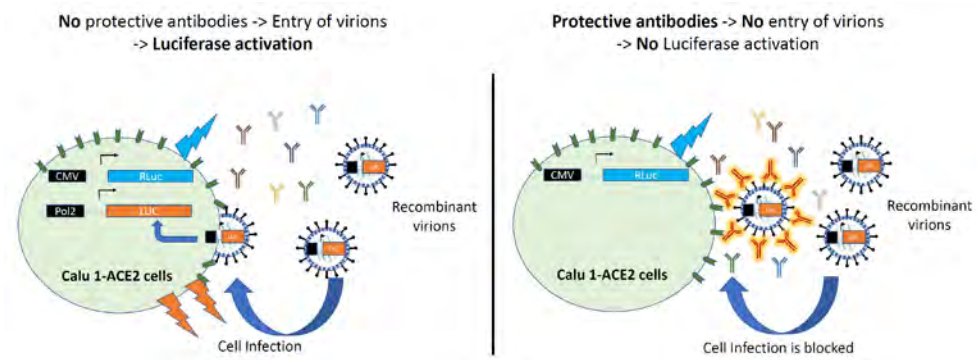


Figure 15. How the neutralization test works

Using this method we can evaluate, by culturing Spike virus and Calu-1(ACE2) cells in the presence of donor sera, whether antibodies have been produced in the sera themselves that can actively block the ability of the virus to bind to human cells, thereby blocking infection and consequently the production of luciferase.

To test the ability of donor antibodies to protect against infection even with virus variants, we cloned three most common variants of the Spike protein into a lentiviral vector.

The currently production-ready variants of the Spike protein are:

- Original Spike protein (originally identified in Wuhan, China);
- Protein Spike variant B.1.1.7 (first identified in the United Kingdom, also referred to as alpha);
- Variant Spike protein B.1.617.2 (first identified in India, also referred to as delta).

Collaborations

- Prof. R. De Maria (Università Cattolica del Sacro Cuore (UCSC), Fondazione Policlinico Universitario Agostino Gemelli (FPG), Rome, Italy)
- Prof. P.G. Pelicci (European Institute of Oncology IFOM-IEO, Milano, Italy)
- Dott. S. Pasqualato (European Institute of Oncology IFOM-IEO, Milano, Italy)
- Dr. G. Sette (Fondazione Policlinico Universitario Agostino Gemelli (FPG), Rome, Italy)

Funds and Grants

- Compagnia di San Paolo (PI: T.L. Haas)
- Intesa San Paolo foundation B/2018/0179 (PI: T.L. Haas)

Molecular Epidemiology and Exposome Unit

Research Group

Alessio Naccarati, PhD, senior researcher (IIGM), head of Unit

Francesca Cordero, researcher (UniTo) and visiting scientist (IIGM)

Carla Di Battista, student (UniTo)

Antonio Francavilla, PhD student (UniTo) and Fondazione Celiachia fellow

Giulia Francescato, research fellow (IIGM)

Amedeo Gagliardi, research fellow (IIGM)

Elton Jalis Herman, student (UniTo) graduated October 2021

Barbara Pardini, PhD, senior researcher (IIGM)

Giulia Beatrice Piaggeschi, PhD student (UniTo and IIGM)

Sonia Tarallo, PhD, junior researcher (IIGM)

Paolo Vineis, Unit honorary member (IIGM) and environmental epidemiology chair (Imperial College, London, UK

Laura Zanatto, student (UniTo) graduated april 2021

The Unit's research program is focused on the integration of environmental and individual (exposures, lifestyles, diseases and intermediate phenotypes) genomic and epigenomic data collected in large prospective studies. The research program has three main objectives:

(a) to examine and apply new laboratory techniques for the analysis of "**fingerprints" left on RNA, DNA, and proteins by environmental exposure** in human population;

(b) to identify new markers in blood and other body fluids that allow **early diagnosis of diseases** and possibly a more effective therapy;

(c) to use and apply cutting-edge technologies for the **assessment of environmental exposures**;

Since 2019, the unit is also engaged in the European Horizon 2020 project "Oncobiome" (A. Naccarati coordinates a WP), which pursues the following objectives::

1) to identify and validate specific microbial profiles linked to the onset of cancer, prognosis, response to therapy or other specific effects
2) to understand the functional relevance of gut commensal ecosystems associated with cancer
3) to integrate the results with other oncological aspects (clinical, genomic, immune, metabolomics)
4) to design complementary tests using integrated "molecular signatures" to foresee cancer onset and progression.

The Unit is among the very few in Italy and Europe which combines population research and the development of molecular tests to be used for early diagnosis and primary prevention of diseases like cancer. The development of high-throughput laboratory technologies and the collaboration with a big network of researchers has made the Unit an important node of the European Research system. The main impact on the regional healthcare system derives from the collaboration with the Piedmont Centre for Cancer Prevention, wherewith we have several research lines in common.

Projects

Project 1: "Research of biomarkers for primary and secondary prevention of tumors"

Subproject "Role of microRNA and gut microbiome as colorectal cancer biomarkers"

A. Naccarati (PI), S. Tarallo, B. Pardini, A. Francavilla, F. Cordero, G. Francescato, G. Piaggeschi

Aims

The aim of this project is to analyze miRNA and other small noncoding RNA (sncRNA) expression profiles using Next Generation Sequencing (NGS) technologies in different biological samples and to relate the obtained profiles with the risk of colorectal cancer or precancerous forms. The hypothesis is that specific miRNA profiles can be used to identify patients with cancer or precancerous lesions by differentiating them from healthy patients, and thus their usage in concomitance with the existing screening methods could improve early diagnosis. A further objective of the study is to examine the possible alterations of miRNA expression levels due to the influence of diet, lifestyles and other environmental factors, and to evaluate how much these potential alterations can affect colorectal cancer onset.

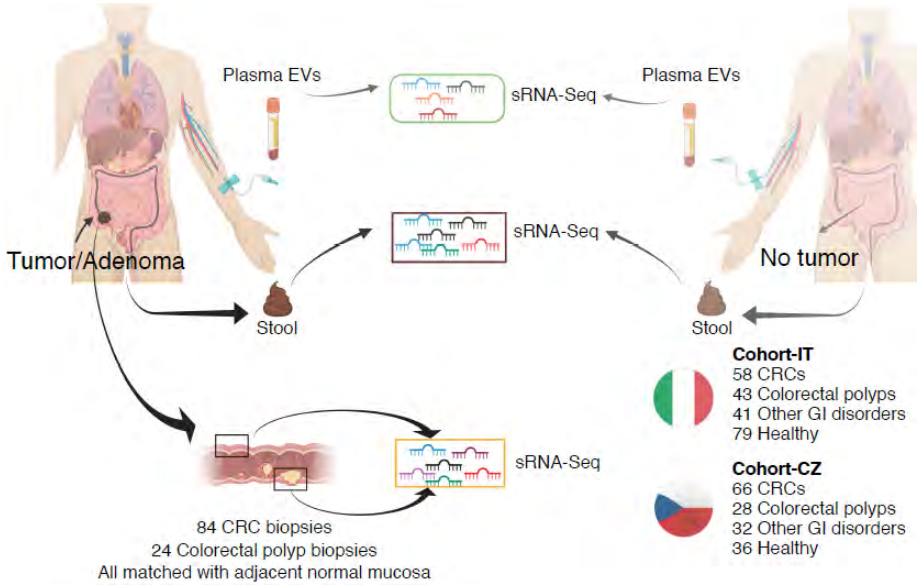
As part of the collaboration with UniTo, the S. Rita Clinic of Vercelli and the University of Trento (Resp. Dr Nicola Segata), two projects have been funded by Lega Italiana per la Lotta contro i Tumori (LILT 2015, 2018). In addition to the analysis of miRNAs and other sncRNAs, the project includes the study of the gut microbiome composition in stool samples from the same subjects. For this purpose, high resolution shotgun metagenomics methods and an integration between the data from the two types of NGS sequencing were performed. Both research areas will be developed in the frame of H2020 Oncobiome project funded by the European Community which started on January 1, 2019. A new AIRC IG project, of which A. Naccarati is PI, has been funded starting in January 2021 and involves collaboration with the Azienda Ospedaliera Universitaria Città della Salute e della Scienza di Torino for the collection and follow-up of new samples and further analysis to evaluate possible prognosis and survival markers.

Results

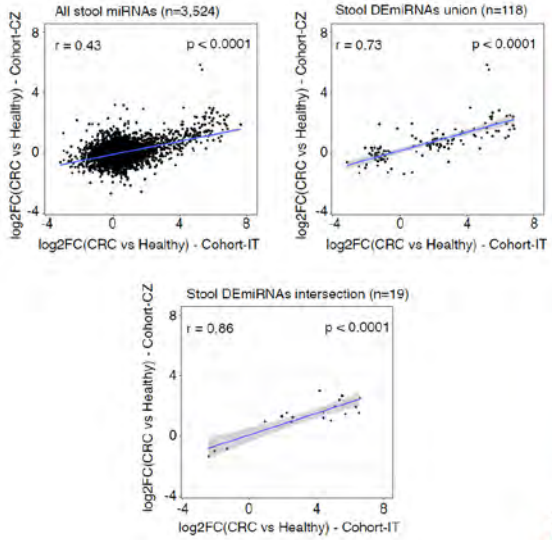
In our laboratory, a protocol for extracting RNA from stool samples and plasma exosomes has been developed for the optimal preparation of libraries for NGS sequencing. The discovery phase on 221 samples of subjects with different diagnoses (cancer, precancerous lesions, inflammations or negative colonoscopy) together with the related statistical analyses of the data obtained is concluded. At the same time, a collection of samples of subjects from the Czech Republic was set up at the time of colonoscopy to be used as a population for the validation of the signals obtained on

the Italian cohort. Over 160 samples from patients with CRC, precancerous lesions and healthy subjects were collected and processed. The study is summarized in **Figure 16 A**.

A.



B.



C.

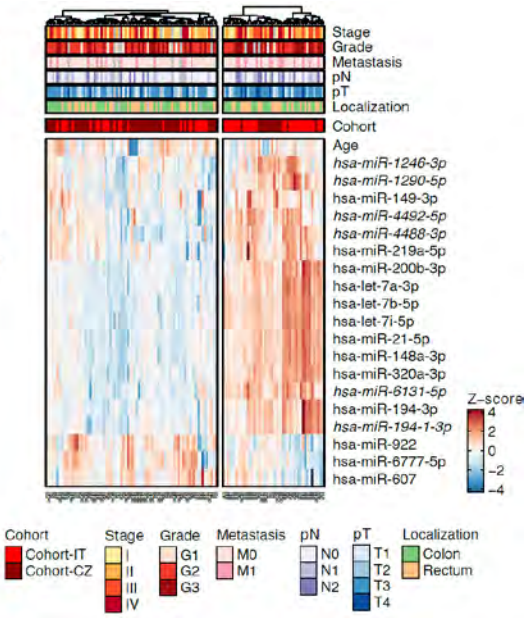


Figure 16. A. Study design. **B.** Scatterplots of stool miRNA log2Fold-change (log2FC) in the comparison between CRC and Healthy in Cohort-IT (x-axis) and Cohort-CZ (y-axis). Data are reported for DEmiRNAs in at least one cohort (top) or both (bottom) **C.** Heat map reporting the normalized expression levels of stool DEmiRNAs in samples from CRC patients of both cohorts. The column annotations reports following clinical data: tumor stage and grade, presence of metastasis, lymph node invasion status (pN), tumor size (pT), and tumor localization.

The results of microbial DNA sequencing in the stool samples of the same subjects have been published on Nature Medicine journal in two back-to-back articles that compared the data of over 900 fecal metagenomes obtained from our and other cohorts, or from available datasets with bioinformatics analyses and different approaches (Thomas AM, *et al.,* Nature Medicine 2019; Wirbel J, *et al.,* Nature Medicine 2019). We have also analyzed sncRNAs expression profiles in relation to metagenomics data in the feces of 80 subjects (adenomas and CRC). The comparison between bacterial expression profiles identified by analyzing shotgun metagenomics data and transcriptomics revealed a high correlation between the profiles. Moreover, we observed different bacterial species present in the different categories of subjects and importantly, a set of biomarkers composed of human-derived sncRNAs and microbial organisms allowed to achieve 88% of disease predictivity for the analyzed subjects (Tarallo S, *et al*., mSystems 2019).

We hypothesize that there is a close relationship between the composition of the gut microbiome, regulation by sncRNAs of microbial genes, and diet. An alteration of these components can be observed in CRC and precancerous lesions and can be further demonstrated by studying the plasma metabolomic profiles of the same subjects. To assess whether this hypothesis is correct, we undertook a collaboration with the IARC of Lyon (resp. Prof Scalbert), we analyzed the composition of plasma metabolites using liquid phase and gas chromatography coupled to mass spectrometry (LC-GC/MS-MS) in all subjects of the Italian and Czech cohort (divided into CRC, patients with adenoma and controls) for which metagenomic sequencing and sncRNA profiles are already available. Analysis of the results is ongoing.

The multi-omics approach, with a combination of state of the art technologies and integrated analyses applied to large scale fecal and plasma samples, will help us to understand how diet in interaction with the gut microbiome, influences host health and disease in order to define accurate and integrated diagnostic signatures for CRC prediction and prognosis.

Collaborations

- Dr. G. Ferrero (University of Turin, Turin, Italy)
- Dr. M. Trompetto, Dr. G. Gallo (S. Rita di Vercelli Clinic, Vercelli, Italy)
- Dr. P. Vodicka, (Department of Molecular Biology of Cancer, Institute of Exp. Medicine, Czech Academy of Sciences, Prague, Czech Rep.)
- Dr. N. Segata (Center for Integrative Biology, Cibio, University of Trento, Trento, Italy) for the microbiome study
- Dott. P. Racca (Azienda Ospedaliera Universitaria Città della Salute e della Scienza, Torino, Italy)
- Prof. A. Scalbert (International Agency for Research on Cancer, Lyon, France)

Subproject "Identification and comparison of miRNA expression profiles in plasma and feces and composition of the intestinal microbiome in subjects participating at the colon cancer screening program: how diet and lifestyle can alter the expression of miRNAs and the composition of the intestinal microbiota"

A. Naccarati (PI), S. Tarallo, A. Francavilla, G. Francescato, B. Pardini

Colorectal cancer (CRC) represents the third most frequent cancer in men and the second in women in the European Union. Population screening programs aim to reduce its incidence and mortality. To date, the gold standard for screening is colonoscopy considering its reliability and sensitivity, but it is invasive and expensive. However, this procedure can reduce the incidence of CRC by 60-90%.
Several European countries have introduced non-invasive screening programs such as fecal occult blood (FOBT), fecal immunochemical test (FIT) or sigmoidoscopy but these methods have limitations in specificity or sensitivity.

Aims

1. To evaluate the potential role of altered miRNA expression levels and the composition of the intestinal microbiota, as screening biomarkers for CRC. These specific "signatures" will be evaluated in plasma and feces using high-throughput methods as biomarkers for the screening of adenoma and CRC or in both, between subjects resulted positive (FIT+) and negative (FIT-) on the screening test FIT
2. To test the predictive power of the risk score, combining the information deriving from biomarkers, lifestyle data and the results obtained from the screening test able to determine the different subgroups and the different risk levels of the CRC

Results

The collection of FIT samples for microbiome analysis and the recruitment of FIT positive subjects who provided stool and plasma samples together with dietary and lifestyle data are still ongoing in collaboration with the Unit of Epidemiology, Screening and Cancer Registry of Turin. The project is funded by an AIRC IG of which Dr. Carlo Senore is PI and IIGM collaborator (the start of the project has been postponed by one year due to the ongoing pandemic). A dataset of clinical and anamnestic data was also prepared. Among the FIT+ subjects (1480), 340 samples of plasma and stool were used for the analysis of miRNA expression and gut microbiome composition. Preliminary analyses on the percentages of subjects at colonoscopy with advanced adenomas and CRC, reflected the expected results (25% and 4%, respectively). The subjects who were FIT- were offered a telephone interview on dietary and lifestyle habits. Among the 1817 contacted, we got the answers to the

questionnaire for 968 of them. To evaluate the potential impact of the screening test in following subjects at risk of tumor development, we started the collection of samples from subjects who had a FIT- results at the first round and that, after two years, repeated the test (**Figure 17**).

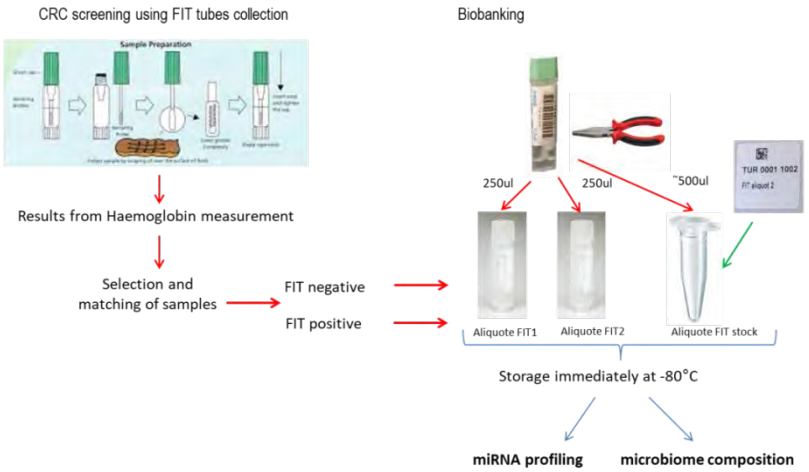


Figure 17. Organization of the study and biobanking of samples within the CRC screening in Torino.

As of June 2021, a total of 5040 samples were collected, of which 600 were from individuals at the second recruitment. An aliquot of FIT leftover sample for 1500 subjects was sent to the American National Cancer Institute to perform fecal metagenome analyses using shallow metagenomics techniques. At the same time, clinical, demographic, and food/lifestyle questionnaire data were processed to cross-reference with the results of the microbiome analyses. During the same period, the first miRNA sequencing analyses were also performed for a group of 45 FIT+ individuals for whom both a FIT sample, collected during screening, and a fecal sample were available after the FIT response. Of this group of individuals, some had a tumor lesion of varying degree or a polyp or were just false positives at colonoscopy. Analysis is currently ongoing.

Collaborations

- Dr. C. Senore, Dr. E. Riggi (Unità di Epidemiologia, Screening e Registro Tumori, Torino, Italy)
- Dr. R. Sinha (National Cancer Institute Division of Cancer Epidemiology & Genetics, Metabolic Epidemiology Branch) and Dr M. Gunter (the International Agency for Research on Cancer (IARC), Lyon, France)
- The project is also included in the "Collaborative International Network of Microbiome Cohorts Nested within National / Regional Colorectal Cancer Screening Programs" consortium led by IARC

Subproject "Study of the expression and composition of the intestinal microbiota in relation to different dietary habits"

S. Tarallo (PI), A. Naccarati, A. Francavilla, F. Cordero, B. Pardini

Aims

Although the relationship between diet, lifestyle and health status is ascertained, it is still difficult to correlate specific dietary patterns with health status. The diet can modulate the expression of microRNAs (miRNAs), small RNA molecules that regulate gene expression. Recent studies have shown that fecal miRNAs directly regulate the expression of specific bacterial genes and also microbial growth, making them essential for maintaining the balance of the intestinal microbiome.

The altered expression of miRNAs is involved in the development of several diseases including cancer. In a previous study, our group showed for the first time that some miRNAs, known in the literature for their altered levels in colon cancer, were differentially expressed in plasma and feces in relation to different dietary habits (Tarallo S, *et al*., Mutagenesis 2014). However, further studies are needed on a larger number of subjects to better understand possible variations in healthy populations.

Furthermore, several factors can positively or negatively influence the composition of the gut microbiome. These changes play an important role in the development of various chronic diseases. The main purpose of this study is to evaluate, through NGS methods, miRNA expression profiles and the gut microbiome composition, in plasma and feces of a large group of healthy volunteer subjects with different dietary habits (vegans, vegetarians and omnivores). For future "personalized medicine' approaches, miRNA expression profiles and the gut microbiome composition could be useful biomarkers to identify healthy nutritional status. This would improve the prevention and treatment of various diseases.

PI Sonia Tarallo received a Post-Doctoral fellowship (Years 2017 and 2018) from the Veronesi Foundation for this project.

Results

Plasma, serum, and stool samples together with questionnaires on diet and lifestyle were collected from 120 healthy volunteers with different dietary habits (vegans (VN), vegetarians (VT), and omnivores (O). All the information related to the anthropometric measurements (weight, height and abdominal circumference), serological data and those deriving from the questionnaires were included in a specific dataset. From the analyses of anthropometric data it has been observed that the BMI of O was significantly higher compared to those of VT and VN. Furthermore, from serological analyses, ferritin resulted significantly lower in VT and VN than in O, whereas vitamin

B12 resulted significantly lower in VN than in the VT that habitually integrated it.

Gut microbiome composition, preliminarily investigated using the 16S method, highlighted that subjects with a vegetarian and vegan diet had a more abundant percentage of *Prevotella* than the omnivores in which the *Bacteroides* predominate. Subsequently, in collaboration with the University of Naples, some samples were analyzed using the shotgun sequencing method for the validation of a study on the microbiome and diet. It has been observed that in diets rich in fiber the *P. copri* strain predominated, with an ability to metabolize carbohydrates, while in omnivores an increase in strains involved in the aminoacids biosynthesis is observed; the latter are known risk factors for the development of glucose intolerances and type 2 diabetes. The diversity at the level of *P. copri* strains is therefore modulated by the diet and presents a very varied repertoire of functions based on the genetic profile. Its importance has so far been underestimated (De Filippis F, *et al*., Cell Host Microbe 2019).

To date, all the samples of the study have been re-sequenced using the shotgun method allowing a greater depth of detection of the species present and the microbial functionality. Seventeen miRNAs correlated ($|rho|>0.22$, $adj.p<0.05$) with the estimated intake of nutrients, particularly animal proteins, phosphorus and, interestingly, lipids. In omnivores, higher *Prevotella* and *Roseburia* and lower *Bacteroides* abundances than in vegans and vegetarians were observed. Lipid metabolism-related miR-425-3p and miR-638 expression levels were associated with increased abundances of microbial species, such as *Roseburia*, SpCAG182, and *Akkermansia muciniphila*, specific of different diets. An integrated analysis identified 25 miRNAs, 25 taxa, and 7 dietary nutrients that clearly discriminated the three diets (**Figure 18 A, B**). The article containing the results of this study was published in the Gut journal (Tarallo *et al*, Gut. 2021).

From the analysis of the sequencing data regarding miRNA expression levels in stool samples we observed the presence of 49 differentially expressed miRNAs (DEmiRNAs) among vegans, vegetarians and omnivores ($adj.p<0.05$) and confirmed trends of expression levels of such miRNAs in vegans and vegetarians compared to an independent cohort of 45 omnivores. Two miRNAs related to lipid metabolism, miR-636 and miR-4739, were inversely correlated to the non-omnivorous diet duration, independently of subjects' age

For a better understanding of the role of the microbiota in determining the health status of the host in relation to diet, in collaboration with the IKEM Institute in Prague we are analyzing the metabolites and bacterial short-chain fatty acids (SCFA), in plasmas and serum of the analyzed subjects. In addition, the data obtained will be compared with those of omnivorous and vegan subjects from an independent cohort in the Czech Republic.

Finally, subjects from this study together with other healthy subjects from the CRC and celiac disease subprojects were used in a study to investigate the association between altered fecal miRNAs and age, sex, BMI and other lifestyle habits (i.e. cigarette, alcohol, coffee and physical activity consumption). A total of 151 miRNAs differentially expressed in at least one variable and 52 associated with more than two variables were identified. The greatest number of differentially expressed fecal miRNAs were seen to be associated with BMI (n=92) and cigarette smoking (n=84). The results suggest that fecal miRNA profiles reflect individual characteristics and lifestyle habits. The article describing the study and its findings has been recently published (Francavilla et al. Scientific report 2021)

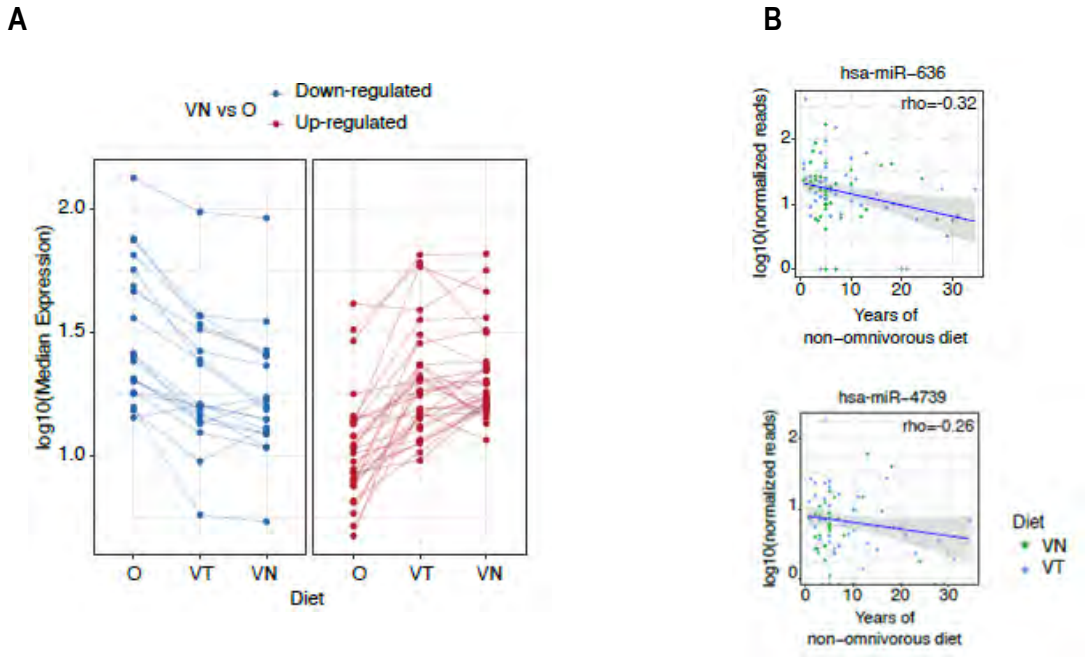


Figure 18. A. DEmiRNAs in the fecal samples showing a trend of expression between the different types of diet investigated. **B.** Scatter plots reporting the normalized expression levels of stool DEmiRNAs (miR-636 and miR-4739) whose expression decreased in relation with a non-omnivorous diet regime and its duration in time.

Collaborations

- Dr. N. Segata (Centre for Integrative Biology, Cibio, University of Trento, Trento, Italy)
- Dr. D. Ercolini (Dept. of Agricultural Sciences, University of Naples Federico II, Naples, Italy)
- Prof. Vittorio Krogh (Epidemiology and Prevention Unit, IRCCS Foundation National Cancer Institute, Milan, Italy)
- Dr. G. Ferrero (University of Turin, Turin, Italy)
- Dott. M. Cahova (Institute for Clinical and Experimental Medicine (IKEM), Rep. Ceca).

Subproject "Profiles of miRNA expression in faeces and plasma of subjects affected by celiac disease by Next-Generation-Sequencing"

A. Naccarati (PI), A. Francavilla (PI), S. Tarallo, B. Pardini, L. Zanatto, P. Vineis

Aims

Celiac disease (CD) occurs in about 1% of the world's population, although most people affected remain unaware of it throughout their lives. New biomarkers could be useful in the diagnosis and monitoring of CD. For this purpose, molecular markers based on new species of non-coding RNAs (ncRNAs) detectable in surrogate tissues, as well as the composition of the gut microbiome can represent an interesting research field. MicroRNAs (miRNAs) are small ~ 22 nucleotide long non-coding RNA molecules with an altered expression in many gastrointestinal diseases and therefore interesting to study as potential biomarkers. Dysbiosis, a recurrent feature in gastrointestinal diseases, including CD, is the pathogenic alteration of the composition of the intestinal microbiota which can disrupt intestinal homeostasis and promote inflammation. As a consequence, there is a growing interest in studying the interactions between microbiome and host. MiRNAs and other small RNAs seem to be involved in these interactions. For all these reasons, our study aims to analyze the miRNA and microbiome profiles in stool / plasma samples using Next Generation Sequencing (NGS) by comparing the effect of the gluten-free diet (GFD) between celiac and healthy subjects and in patients with a new diagnosis of CD before and after one year from the beginning of the GFD.

For this project, Antonio Francavilla received a three-year scholarship from the Italian Celiac Association (AIC).

Results

A total of 136 participants were recruited categorised as: i) untreated CD, recruited at diagnosis before starting the GFD (n=3), ii) CD treated, already on a GFD (n=63, further grouped according to serological levels of tTG2-Ab in those under (CD-ltTG) and over (CD-htTG) the <3.0 UA threshold), iii) NCGS individuals (n=2), and iv) healthy sex-/age-matched controls, not following any specific diet (n=68). A stool, plasma, and serum sample and questionnaires regarding their dietary and lifestyle habits were obtained for all individuals. For all of them, RNA extracted from stool samples was used for the miRNA sequencing experiments. Differential expression (DE) analysis revealed 100 miRNAs with significantly different expression levels in both groups of CD subjects on GFD diet compared to controls. **Figure 19 A** shows an example for three out of 100 DEmiRNAs identified. Enrichment

analyses of the target genes of dysregulated miRNAs have shown their involvement in inflammatory processes associated with celiac disease.

Also, metagenomics analysis on stool samples of the same cohort of patients have been performed and as result, 13 microbial species reported a different abundance between the two CD groups compared to controls. Interestingly, significant correlations were observed between the abundance of those microbial species and the DEmiRNAs levels (**Figure 19 B**). Significant correlations were also observed between DEmiRNAs levels or microbial species abundance and GFD years of duration (data not shown).

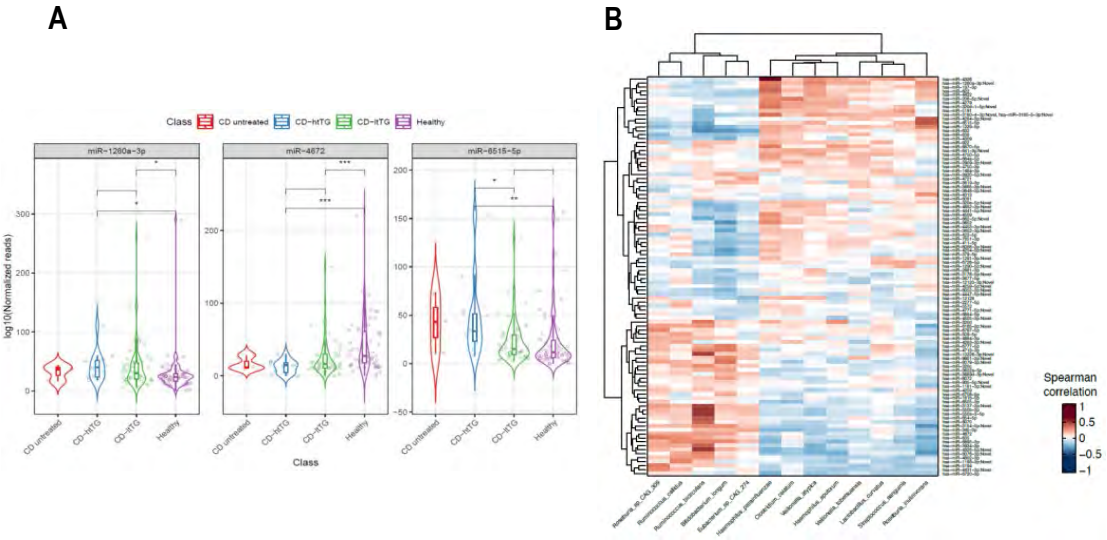


Figure 19. A. Violin plots reporting expression levels of selected DEmiRNAs showing a trend of expression (down-regulation for miR-1260a-3p and miR-6515-5p and up-regulation for miR-4672) going from CD untreated to CD-htTG, CD-ltTG and the healthy controls categories. **B.** Heatmap representing spearman correlation coefficients computed between the 13 microbial species reporting a different abundance among the categories and the DEmiRNAs.

Collaborations

- Dr. L. Crocellà, Dr. C. Guiotto, Prof. R. Rocca (Mauritian Order Hospital, Turin, Italy)
- Dr. D. Ribaldone, Dr. G. Caviglia (San Giovanni Antica Sede Hospital – SGAS, Turin, Italy)
- Prof. M. Bruno (University Hospital of the City of Health and Science, Turin, Italy)
- Dr. G. Ferrero (University of Turin, Turin, Italy)

Subproject "microRNAs and other small non-coding RNAs expression analysis in pediatric patients with suspicion of celiac disease"

A. Naccarati (PI), A. Francavilla (PI), S. Tarallo, B. Pardini

Aims

The diagnostic criteria for celiac disease (CD) in pediatric patients have recently been revised and updated. However, additional molecular markers based on small non-coding RNAs (sncRNAs) and microbiome composition could aid in the diagnosis and monitoring of this disease in its various forms and demonstrations. Based on of the previously discussed project concerning the analysis of sncRNAs and intestinal microbiome in adult celiac patients, a collaboration was recently born with the pediatric gastroenterology unit of the Regina Margherita Children's Hospital, coordinated by Dr. Pierluigi Calvo. A study was therefore organized based on the analysis of pediatric celiac patients with different forms and manifestations of this disease.

The project's specific aim will be to analyze sncRNAs and intestinal microbiome profiles in surrogate tissues of pediatricCD patients with different forms (classic or potential celiac disease), with or without symptoms, and see how these profiles undergo any variations due to adherence to the gluten-free diet.

Results

The ecruitment of patients within the categories of interest is in progress. 91 biological samples have been collected from patients with different forms of CD and healthy controls. For 83 of them, RNA was extracted from stool samples and subsequently used for sequencing experiments. The data generated will be integrated with clinical and dietary data obtained from the questionnaires completed for each participant.

Collaborations

- Dott. P. Calvo, Dott. I. Giraudo (Regina Margherita Children's Hospital, Turin, Italy)
- Dr. G. Ferrero (University of Turin, Turin, Italy)

Subproject "Evaluation of DNA methylation in leukocyte sub-populations associated with tobacco smoking exposure"

G. Piaggeschi (PI), S. Tarallo, F. Cordero, A. Naccarati

Aims

This project is the natural continuation of the work carried out by our group on the relationship between smoking habits and DNA epigenetic alterations. So far, the analyses were conducted on whole blood samples prospectively collected and stored in biological banks. This study aims to investigate more in-depth the epigenetic signals by analyzing the separate cell-types present in the blood. This will allow to understand if the signals detected are due to the contribution of a single or more cell-types and it will help to better understand the kinetics of smoke exposure markers to return to normal levels. The experimental design of the project is based on a preliminary phase of flow cytometry analyses (FACS) which will highlight the differences in the proportions of the main cell subpopulations among smokers and non-smokers. In parallel, the expression levels of membrane receptor (GPR15), known to be significantly elevated in smokers, will be evaluated in the different cell-types (Koks G**,** *et al*., Am. J. Pathol., 2015; Bauer M, *et al*., Clin Epigenetics, 2017). In addition, the levels of cotinine (a metabolite of nicotine) as a direct indicator of smoking exposure and, the concentration of C-Reactive Protein as inflammatory markers will be measured in plasma samples.. For this type of analyses, fresh blood samples have been collected from healthy volunteers with different smoking habits (current, former and never smokers).The data obtained from the first phase will allow the setting up of the analyses for the second phase of the project in which the different cell-types will be separated for a specific methylation analyses performed by bisulfite sequencing. This technique permits to sequence the genomic DNA converted with the sodium bisulfite in order to underline the DNA methylation patterns at single CpG level.

Results

Recruitment and samples collection started in April 2017 in collaboration with the Italian Blood Volunteers Association (AVIS) ended in December 2018 with 300 samples collected. For all samples, nine leukocyte subpopulations (T, T-helper, T-cytotoxic, NK, NKT, B, monocytes, neutrophils and eosinophils) were quantified by FACS with the relative measurement of GPR15 in each cell subtype. In smokers, there was a decrease in NK cells and an increase of B, T, T-helper and T-cytotoxic cells expressing GPR15 marker compared to non-smokers ($P < 0.05/18 = 2.8 \times 10^{-3}$) (**Figure 20 A**). The obtained results are comparable with those present in literature despite the fact that our cohort is made up of moderate smokers (<10 cigarettes/day).

For the second phase of the project, an additional 16 healthy subjects (8 smokers and 8 non-smokers) matched by sex and age were recruited. From their samples, the nine cell subpopulations were separated by FACS and the DNA was extracted.

The DNA extracted from whole blood and from two separate cell lines (monocytes and B cells) of 6 subjects (3 smokers/ 3 non-smokers) was sequenced for assessment of methylation levels at the Genecore facility of EMBL in Heidelberg. Samples were sequenced with two different techniques to analyze methylation levels: targeted and whole genome. Methylation profiles were analyzed by comparing smokers versus non-smokers, identifying 229 and 290 differentially methylated regions in monocytes and B cells, respectively (adj.p-value <0.001). These correspond to regions covering 137 genes for monocytes and 173 genes for B cells, respectively. Of note, 18 of these regions are in common between the two cell types. Further analyses comparing data from the two different sequencing techniques and further analyses comparing methylation levels between cell subtypes and levels measured in whole blood are underway.

In order to develop a bioinformatics pipeline to analyze the methylation sequencing data, an experimental and computational study has been conducted on 22 whole blood samples from the EPIC-Italy cohort to evaluate whether there were differences between the Illumina Bead-Chip array and the targeted sequencing technique to analyze the methylation levels. The obtained results demonstrate that the methylation levels of 1054 CpG sites analyzed with the two approaches correlated with high correlation coefficients (range 92-99%), demonstrating that the two techniques are comparable (**Figure 20 B**). Furthermore, within this subproject, MethylFASTQ, a tool that simulates the production of NGS methylation data has been developed to create the analysis pipeline and compared with other tools available in the literature (Piaggeschi G, *et al.*, 27th Euromicro Int. Conference on Parallel, Distributed and Network-Based Processing (PDP), 2019).

In collaboration with King's College London, the association between the smoking habits and the 35,651 immune cell subset frequencies was conducted on 358 healthy females from TwinsUK cohort. The results confirmed that active tobacco smoking is associated with increased frequencies of circulating CD8+ T cells expressing the CD25 activation marker. Moreover, we identified novel associations between smoking status and relative abundances of CD8+ CD25+ memory T cells, CD8+ memory T cells expressing the CCR4 chemokine receptor, and CD4+CD8+ (double-positive) CD25+ T cells. We also observed, in current smokers, a decrease in the relative frequencies of CD4+ T cells expressing the CD38 activation marker and an increase in class-switched memory B cell isotypes IgA, IgG, and IgE (**Figure 18 C**). Finally, using data from 135 former female smokers, we showed that the relative frequencies of immune traits associated with active smoking are usually

completely restored after smoking cessation, with the exception of subsets of CD8+ and CD8+ memory T cells, which persist partially altered (**Figure 18 D**). Our results are consistent with previous findings and provide further evidence on how tobacco smoking shapes leukocyte cell subsets proportion toward chronic inflammation (Piaggeschi G, *et al* Front Immunol 2021).

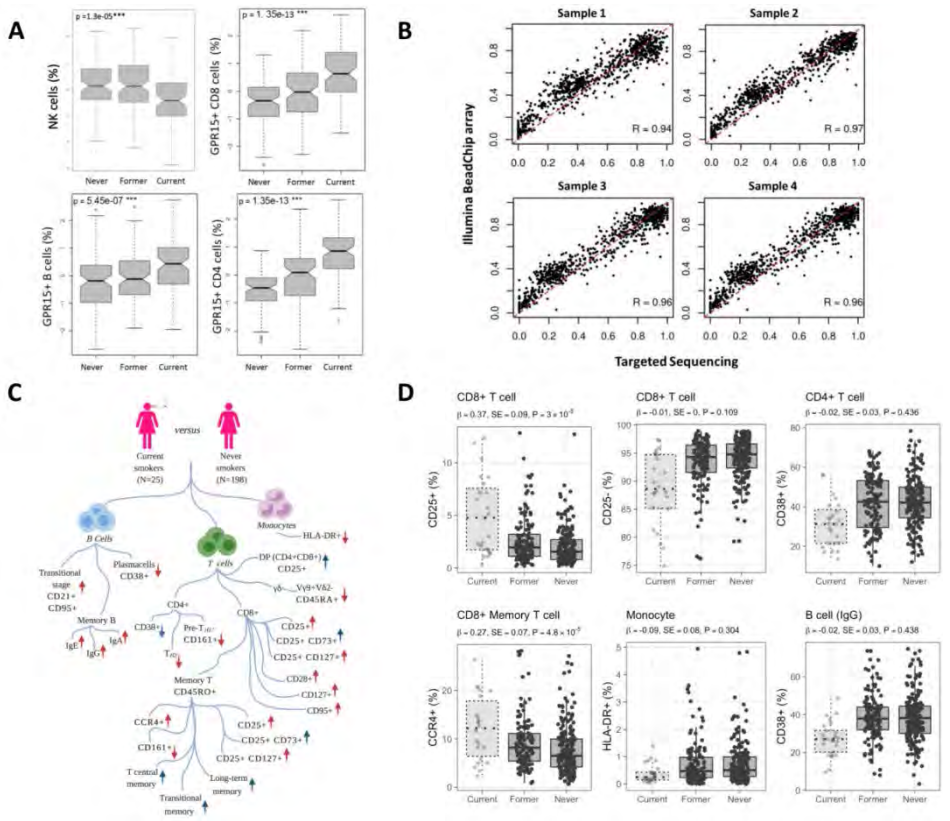


Figure 20. A. The box plots show the different percentage of the cell subtypes (y axis) in the smoking categories (x axis). **B.** An example of correlation between methylation levels obtained by the Illumina BeadChip array and those by Targeted Sequencing for four samples. **C**. Immune cells associated with active smoking. Arrows summarize the direction of association (positive/negative) and the pro-inflammatory (in red) or immunosuppressive (in blue) function. **D**.The box plots show the immune cell subsets partially (left) and completely (central and right) restored after smoking cessation.

Collaborations

- Dr. D. Brusa (Universite Catholique de Louvain (UCL), Institute of Experimental and Clinical Research (IREC), Brussels, Belgium)
- Dr. M. Falchi and Dr A. Visconti (Computational Biology group, Dept. Twin Research & Genetic Epidemiology King's College, London, UK)
- Dr. S. Rolla (Dipartimento di Scienze Cliniche e Biologiche, University of Turin, Turin, Italy)

Subproject "DP3: Early detection and prognosis of prostate cancer: an integrated model between biology and technology"

A. Naccarati (PI), A. Gagliardi, B. Pardini, G. Francescato

Aims

The optimal management of prostate cancer (PC) is hindered by two major challenges: overcome the screening test based exclusively on the serum levels of PSA and the development of new strategies to differentiate between the indolent and the more aggressive forms, associated with a reduced survival rate and increased resistance to therapy. Several studies have showed that the PSA levels in blood have an excessively high ratio of false-positive and lead to unnecessary invasive biopsies. This is especially the case when serum PSA is in the range of 4-16 ng/mL and the test is not capable to differentiate among benign prostatic hypertrophy (BPH), pre-cancerous lesions and aggressive tumor. The "Fondazione Edo ed Elvio Tempia" (FEET), involved for years in prostate cancer research, in collaboration with IIGM, has developed a classifier based on expression levels of miRNAs circulating in blood that manage to improve the classification of prostate cancer (Mello-Grand M, *et al*., Carcinogenesis 2019). The rationale of this collaborative project is to identify a miRNA expression profile that could reflects the onset and progression of the disease. This profile could then be used as biomarker, in association with PSA test, and ameliorate the screening program and reduce the false-positive rate. In addition, we propose to evaluate the impact of introducing the analysis of prognostic genes and methylation markers into the prognosis definition pathway of prostate cancer.

An additional aim is to define a classifier based on RNA transcripts and circulating metabolites that could improve the staging and prognosis of the cancer. In this respect, RNA-seq on biopsies (FFPE or RNA later) will be perfomed on patients with a confirmed diagnosis of PC and the results will be analyzed in correlation with clinical outcome.

Results

The project, which officially started in March 2020, included in the initial stages the request for approval for the study design by the ethics committee for the Tempia Foundation and is currently in preparation the necessary documentation to obtain the approval of the ethics committee to start the study also at the Molinette Hospital in Turin. However, the recruitment and organization of the project have been affected by the effect of the Covid-19 pandemic. In the meantime, also due to delays in the start of patient recruitment, a collaboration has begun with the Institute of Cancer Research (University of Vienna), which has a large case series of patients and biological samples available and has been shown to participate in the study. As part of the collaboration, molecular analyses of

the samples provided are underway. IIGM is currently involved for transcriptomic analyses on FFPE samples.

Collaborations

- Dr. G. Chiorino (Fondazione Edo ed Elvio Tempia Onlus, Milan, Italy)
- Dr A. Zitella (Azienda Città della Salute di Torino, Turin, Italy)
- Dr A. Gsur (Institute of Cancer Research, Medical University of Vienna, Vienna, Austria)

Subproject "Evaluation of the impact of a Very Low Ketogenic Diet (VLKD) on the microbiome composition and epigenetic profiles of obese patients who are candidates for bariatric surgery"

A. Naccarati (PI),S. Tarallo (PI), B. Pardini

Aims

The effectiveness of nutritional interventions in reducing body weight are mediated by negative energy balance and other mechanisms. It was observed that the inability to respond satisfactorily to caloric restriction in terms of fat mass loss could depend on significant epigenomic and/or transcriptomic differences in adipose tissue (such as different methylation of specific loci that can differentiate responders and nonresponders to caloric restriction).

The composition of the gut microbiome was correlated with body weight and the probability of response to dietary treatment. The diet could represent an ideal bridge able to connect the gut microbiome and the host metabolism and modifying the state of health, through, at least in part, the regulation of epigenetic mechanisms.

The ketogenic diet, whose pathophysiological mechanism is based on a deficit of carbohydrates, which forces the body to produce energy substitutes (e.g. ketone bodies) from lipids, resulting in the consumption of adipose tissue storage, has a specific impact on the composition of the intestinal microbiome, while the effect on the epigenetics of subjects subjected to this diet is less detailed. In general, alterations in the gut microbiome and in the epigenetic pattern can be both a cause and a consequence of weight loss; therefore, even today it is still debated what is the exact causal relationship between these factors and weight loss.

In this context, the so-called "omics" analysis, i.e. at global level, such as transcriptomics (analysis of all human RNAs), metagenomics (analysis of microbial composition) and metabolomics (analysis of metabolites) represent in recent years a field of interest for research in precision medicine, as they could be source of accurate biomarkers. These biomarkers can be measured and evaluated as indicators of a disease state, or as a response to specific therapeutic or nutritional interventions. However, information on the profiles of these molecular markers in human samples, in relation to weight variation following ketogenic diets or bariatric surgery are so far scarce and based on small studies.The aim of the present project is to evaluate how ketogenic diet and subsequent bariatric surgery, may affect the gut microbiome and epigenetic profiles in fecal samples (both small non-coding RNA (sncRNA) and DNA methylation) in subjects with obesity grade III (BMI ≥ 40 kg/m2) and indication for bariatric surgery.

Results

The project has obtained the approval of the Ethics Committee of the City of Health of Turin. Unfortunately, due to COVID pandemics, recruitment has not yet been possible.

Collaborations

- Dr. A. De Francesco (Azienda Ospedaliera Universitaria Città della Salute e della Scienza, Torino, Italy)
- Dr.F. Rahimi (Azienda Ospedaliera Universitaria Città della Salute e della Scienza, Torino, Italy)
- Dr. (INSERM French Institute of Health and Medical Research, France)
- Dr.A. Balcerczyk (University of Lodz, Lodz, Poland)

Project 2: "Environmental Measurements and Molecular Footprints of Environmental Exposures (EXPOsOMICs)"

A. Naccarati (PI), S. Tarallo (PI), B. Pardini, F. Cordero, P. Vineis

Aims

The European project EXPOsOMICs (www.EXPOsOMICsproject.eu/publications /EXPOsOMICs), of which Prof Vineis was coordinator from the Imperial College London, had the objective of measuring the environmental exposures which we are subjected to in the everyday life (Vineis P, *et al.*, Int J Hyg Environ Health 2017). IIGM was responsible for the individual monitoring of pollutants based on 24-hour environmental measurements (for three days in a year) among 44 volunteers recruited within the EPIC cohort. The monitoring was carried out through a backpack that allows to measure exposure to ultrafine particles, physical activity and identify the geographical position. For each volunteer recruited, different types of biological samples were collected (blood, nasal swab, exfoliated cells of the mouth) and spirometry was performed.

In December 2018, still in the context of the EXPOsOMICs project, a new monitoring study was carried out to assess exposure to pollutants and chemicals, using silicone wristbands (an innovative mode of individual monitoring), in a population of students at the University of Turin. Students were assigned a bracelet that they wore for four consecutive days during their normal activities. In addition, they completed questionnaires on diet and lifestyle during these days and at the end of recruitment were collected biological samples (saliva, blood and urine) for omics analysis. In parallel, a similar study was conducted in France on 40 subjects.

Finally, since April 2020, a new collaboration with ASL of Salerno, coordinated by Dr. Luigi Montano, has started. It involves the recruitment of new subjects to study the relationship between external exposure (through the use of silicone bracelets and other personal or environmental devices) and internal exposure (on different biological matrices: plasma, saliva and seminal fluid) to organic and inorganic pollutants in areas of high/low pollution in the area of Terra dei Fuochi and Valle del Sele. The project has been approved by the local Ethics Committee and recruitment of the subjects is expected to begin in October 2021, if the COVID pandemic situation will allow it.

Results

The data on the environmental monitoring collected by the tools carried by the volunteers and in other 160 sites scattered throughout the city of Turin have been analyzed in collaboration with the European partners of Exposomics, as reported in numerous recent scientific publications (Van Nunen E, *et al.*, Environ Sci Technol 2017; Mostafavi N, *et al.*, Environ Int. 2018; Fiorito G, *et al.*, Environ Mol Mutagenesis 2018; Jeong A, *et al.*, Environ Int. 2018; Donaire-Gonzalez D, *et al.*,

Environ Int. 2019, Font-Ribera L, *et al*., Environ Int. 2019). Two new articles address the analysis of ultrathin particles (UFP) in relation to LUR models (Land Use Regression models) and personal exposure detected in the Exposomics study (Van Nunen E, *et al*., Environ Res. 2021) and always for UFP and PM2.5 (fine particles) in relation to lung function and blood pressure (Van Nunen E, *et al*., Environ Res. 2020).

Biological samples have been processed and statistical analyses of some "omics" markers (epigenetics, adduttomics, proteomics and metabolomics) in relation to environmental exposures are currently underway. The results obtained from this investigation suggest that personal exposure to PM2.5 is associated with the expression levels of some miRNAs, showing the potential of these circulating molecules as new biomarkers for health risk assessment of air pollution (Mancini *et al*, 2020). Results on the bracelet analyses found that both populations were exposed to OPEs (Organophosphates), PBDEs (Polybrominated diphenyl ethers), nBFRs (novel brominated flame retardants), and PAHs (Polycyclic aromatic hydrocarbons). The results of this study were published in a recent paper (Wang, *et al*., *Environ Pollut. 2020*). Further mass spectrometry analyses in urine and saliva are still being conducted on the same samples to see the relationship between external and internal pollutants. (**Figure 21**).

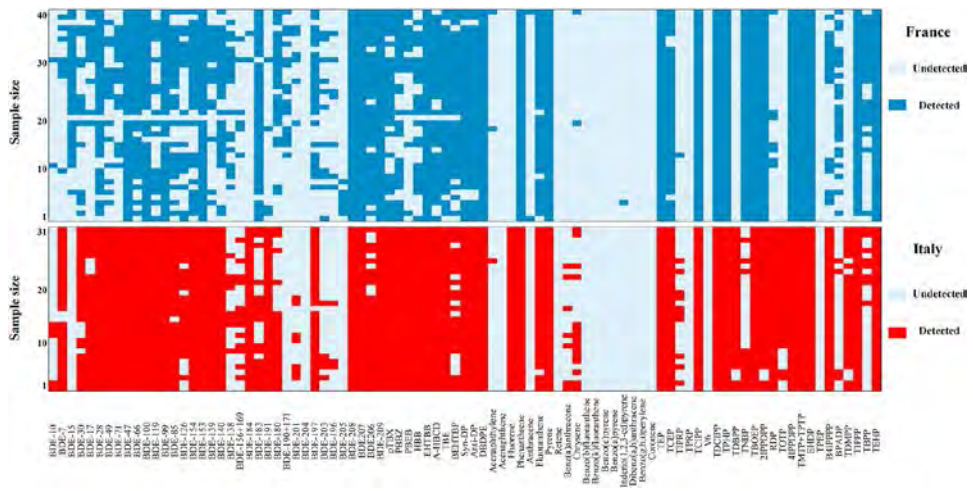


Figure 21. Detection map of all 92 SVOCs (including 39 PBDEs, 10 nBFRs, 25 OPEs, and 18 PAHs) tested in the wristbands collected from France (n ¼ 40) and Italy (n ¼ 31).

Collaborations

- Dr. M. Venier (Indiana University, USA)
- Dott. L. Montano (ASL di Salerno, Ospedale S. Francesco d'Assisi, Salerno, Italy).

Funds and Grants

- Compagnia di San Paolo (PI: A.G. Naccarati)
- EU Commission “Gut OncoMicrobiome Signatures (GOMS) associated with cancer incidence, prognosis and prediction of treatment response” (N. 210511940) (Partner: A.G. Naccarati)
- Lega Italiana per la Lotta contro i Tumori (LILT) "Validazione di biomarcatori clinici basati su RNA non codificante e microbioma intestinale per il cancro del colon-retto" (Partner: A.G. Naccarati).
- Grant Agency of the Czech Republic "Identification and comparison of plasma / stool miRNA signatures in colorectal cancer by next generation sequencing" (N. 17-16857S) (PI: A.G. Naccarati)
- Fondazione Celiachia (FC) and Associazione Italiana Celiachia (AIC) “MicroRNA profiling in stool and plasma of subjects affected by celiac disease by Next Generation Sequencing” (Fellowship 015_FC_2017) (PI: A. Francavilla).
- AIRC Associazione Italiana per la Ricerca sul Cancro "Combining faecal biomarkers to improve prediction of individual’s risk of pre-invasive and invasive colorectal lesions" (IG2019 N. 23473) (Partner: A.G. Naccarati)
- Compagnia di San Paolo – Fondazione Tempia (Partner: A.G. Naccarati)

Immuno-Regulation Unit

Research Group

Luigia Pace, Ph.D, head of Unit (IIGM)

Simona Aversano Stabile, PhD student (IIGM and UniTo)

Nadia Brasu, PhD student (IIGM and UniTo)

Marco Ceravolo, fellow (IIGM) until 31/07/2021

Carlo De Intinis, PhD, post-doctoral fellow (IIGM)

Ines Elia, PhD, post-doctoral fellow (IIGM)

Giuseppina Granato, PhD, post-doctoral fellow (IIGM) until 10/11/2020

Gaia Montacchiesi, PhD student (IIGM and UniTo)

Luca Petiti PhD, post-doctoral fellow (IIGM) until 11/09/2020

Valentina Russo, PhD student (IIGM and UniTo)

The immuno-regulation Unit was started at the beginning of 2018. The research activities of the laboratory aim at understanding the fundamental mechanisms underlying the development of the immune responses, especially the long-term immunological memory processes in the context of tumors and infections (**Figure 22**).

The research program is developed on 2 major axes:

1. defining **how T lymphocytes integrate complex environmental stimuli** and transform these signals into **specific epigenetic and differentiation programs**
2. to **determine the ontogenesis and function of the T lymphocyte** subpopulations **involved in the antitumor responses and infections**. These studies will allow acquiring new knowledge on the functional identity of T lymphocytes and on epigenetic mechanisms of memory, thus opening to new perspectives in the manipulation of the immune responses in the promising field of immunotherapy, as well as in the development of new vaccines.

The strategy of the laboratory is based on multidisciplinary approaches, at the crossroads of immunology, epigenetics, system and cancer biology, and bioinformatics. To this end, the research group uses a wide variety of methodologies, including immunological models relevant to the study of cancer and infection, genetic and genomic techniques, analysis of epigenetic markers and chromatin structure, new algorithms for bioinformatics analysis and modelling.

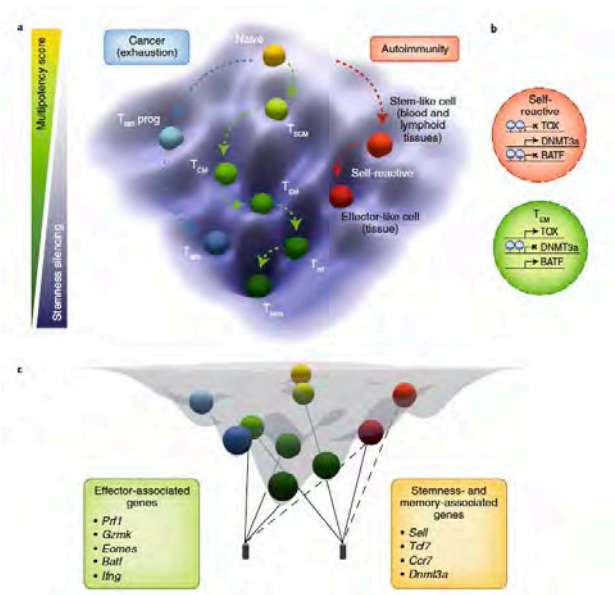


Figure 22. Epigenetic dynamics during T cell differentiation.
(L.Petiti, L.Pace, Nature Immunol. 2020).

a. Epigenetic landscapes shape the fate of CD8+ T cells in exhaustion (left), acute infection (center) and autoimmunity (right). T_{SCM}, stem cell memory T cell; T_{CM}, central memory T cell; T_{EM}, effector memory T cell; T_{eff}, effector T cell; T_{term}, terminally differentiated T cell; T_{exh} prog, exhausted T cell progenitor; T_{exh}, exhausted T cell.

b. DNA methylation footprints at key loci in self-reactive and differentiated effectors CD8+ T cells.

c. The Waddington model revisited. The phenotype of self-reactive beta cell–specific CD8+ T cells is orchestrated by effector and stemness/memory transcriptional programs (dashed lines).

Projects

Project 1: " Immuno-regulation"

Subproject "Analysis of nuclear dynamics of heterochromatin during T lymphocyte differentiation"

L. Pace (PI), G. Montacchiesi, I. Elia, C. De Intinis

The Unit's research activity has highlighted a novel epigenetic mechanism underlying effector T lymphocyte differentiation and memory. The study indicates that the activity of the Suv39h1 enzyme determines the formation of an epigenetic barrier during the terminal differentiation process of effector T lymphocytes. This epigenetic barrier is established by the deposition of the marker H3K9me3 (tri-methylation of histone H3), an epigenetic modification important for heterochromatin formation. Indeed, enrichment of this epigenetic marker inhibits reprogramming of effector lymphocytes, resulting in silencing of transcriptional programs related to stem and memory properties. As a result, effector lymphocytes undergo terminal differentiation and die.

Understanding the mechanisms of epigenetic regulation during effector and memory cell development has considerable implications in multiple fields of medicine, including cancer immunotherapy. Indeed, in cancer, lymphocytes differentiate at a dysfunctional stage, known as "exaustion," characterized by a poor functional responsiveness to tumor antigens. How the process of "exaustion" is established, what are the mechanisms that can inhibit this dysfunctional state are the subject of intense research activity.

Aims

1. To examine the role of heterochromatin factors during memory T cell differentiation
2. Analysis of heterochromatin factors during the anti-tumor immune response

Results

The study of the organization of heterochromatin in subpopulations of T lymphocytes was focused on the application of two multi-omics technologies: the analysis of ATAC-seq (Assay for Transposase Accessible Chromatin) for the study of chromatin structure, and the analysis of the expression profile of proteins and transcripts in each individual T cell, through the use of the technology "BD Rhapsody RNA and protein analysis system". Our study was focused on the analysis of CD8+ T lymphocytes during the anti-tumor response in mouse models.

The ATAC-seq technique is considered a fundamental methodology in the field of epigenetic research as it allows to map active regions of euchromatin throughout the genome. This technology

is based on the use of the enzyme Tn5 that, through a mechanism of "cut and sew", inserts adaptor sequences in regions of chromatin accessible to enzymatic activity. These regions are then detected following the construction of genomic libraries and the use of NGS (next generation sequencing).

In the context of immunological research, the study of lymphocyte populations involved in the anti-tumor response is often complicated by the low number of antigen-specific cells. Often, the manipulation of a low number of cells represents a limiting factor to analyze at molecular level and in detail, the epigenetic and nuclear dynamics of lymphocytes. Initially, we developed a method to isolate and map heterochromatin from a low cell number, using 30000 naive and effector CD8+ T cells (**Figure 23**).

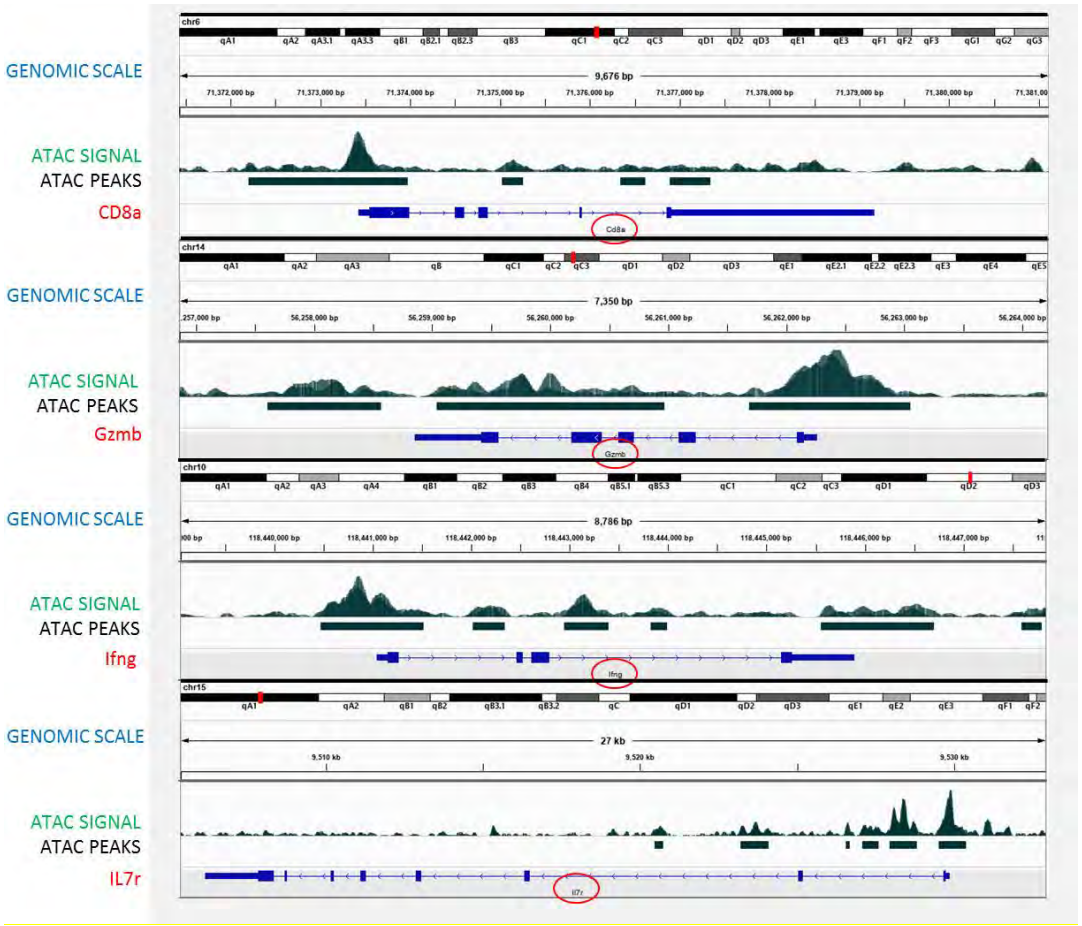


Figure 23. ATAC-seq profile of the open chromatin regions corresponding to genes activated in effector CD8 + T lymphocytes. The peaks indicate the accessible regions of the chromatin. Two technical replicates are shown.

Bioinformatic analysis correlated gene expression profiles with open and closed regions of chromatin identified with ATAC-seq. Analysis of regions of increased accessibility identified active or repressed genomic regions related to effector T lymphocyte differentiation.

New analyses of lymphocyte infiltrates in solid tumors allowed the identification of new lymphocyte subpopulations. Given the low number of cells recovered, we further adapted the ATAC-seq methodology to a few thousand cells. Genomic analysis of the identified lymphocyte subpopulations is ongoing.

The chromatin organization analysis has been extended to the identification of genes and genomic sites regulated by antigen-specific lymphocyte activation. ATAC-seq analysis was also complemented with gene expression profiling through the RNA-seq technique. The study showed a correlation between specific chromatin opening profiles and long-term differentiation of T effector memory cells. This population plays an important role in tumor rejection.

In order to analyze the profile of surface markers and gene expression of lymphocyte subpopulations, the Unit has developed a multi-parametric analysis protocol using BD Rhapsody technology. This innovative technology allows a simultaneous multi-omics analysis of gene expression in terms of both mRNA and protein, at the level of each individual cell (**Figure 24 A, B**). The BD Rhapsody uses a cartridge containing more than 20000 microwells, each of which can be occupied by a single cell, which in turn can bind a single bead. Each bead is bound to an oligonucleotide, with a recognition code, capable of capturing nucleic acids from the individual cell. It is possible to simultaneously analyze several tens of samples, each marked by a specific "Sample Tag" and up to 100 different antibodies.

In order to examine gene expression profiles and phenotype changes of T lymphocytes during anti-tumor response, we implemented the method using tumor infiltrating T lymphocytes isolated from lymph nodes. We generated 3 different types of NGS libraries: mRNA, proteomic, and "Sample Tag" that we combined together for a single multi-omics analysis (**Figure 24 C**).

New bioinformatics analyses have been conducted to assess gene expression profiles in correlation with surface markers, and new analysis pipelines have been developed for the integration of different omics (**Figure 25**).

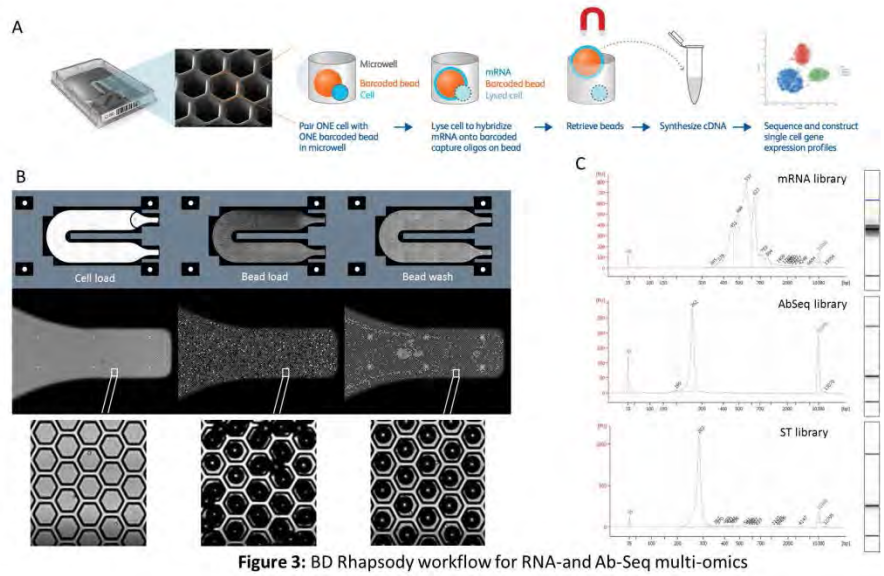


Figure 24. Workflow of the sample preparation method for the analysis of proteomics and single cell transcriptomics. **A,B.** Representative diagram of the various sample preparation stages. **C.** Analysis of the quality of the libraries obtained in the laboratory and analyzed through 2100 Bioanalyzer (Agilent).

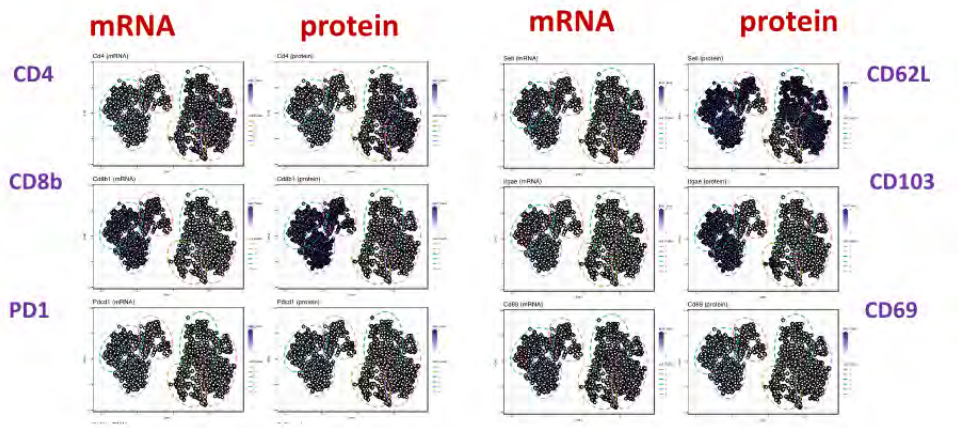
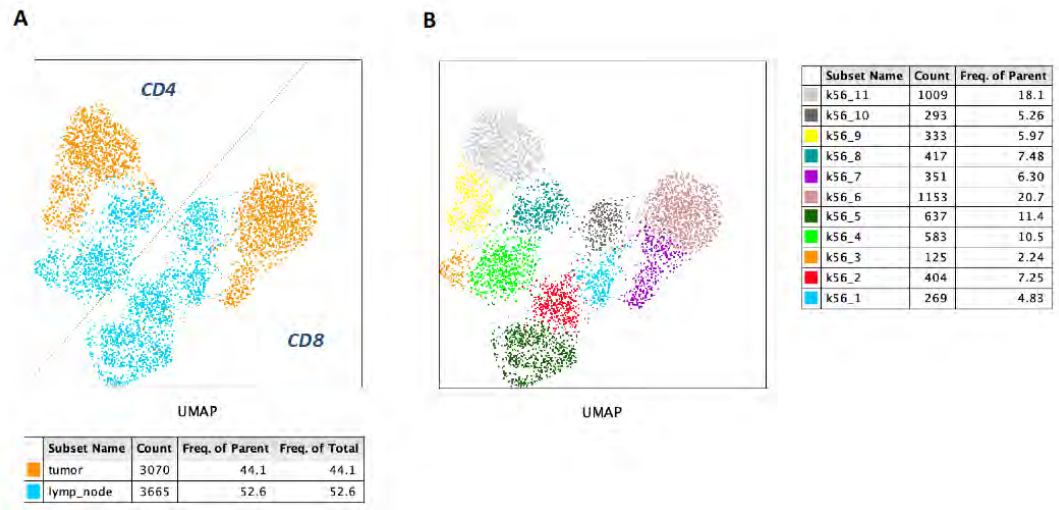


Figure 25. Analysis of the protein (Ab-Oligo) and mRNA molecule profiles by BD Rhapsody single cell multi-omics. About 1500 T cells isolated from a mouse lymphnodes are represented.

In order to examine the gene expression profiles and changes in the T lymphocytes during the anti-tumor responses, we have examined T cells in several experimental mouse tumor models. The study has also been extended to the analysis of the adaptive immune responses following different immunotherapy protocols (**Figure 26**).

These genomic data will be explored in greater detail in the coming months, improving bioinformatics analysis, and validating working hypotheses with functional validations.



Subset Name	Count	Freq. of Parent	Freq. of Total
tumor	3070	44.1	44.1
lymp_node	3665	52.6	52.6

Subset Name	Count	Freq. of Parent
k56_11	1009	18.1
k56_10	293	5.26
k56_9	333	5.97
k56_8	417	7.48
k56_7	351	6.30
k56_6	1153	20.7
k56_5	637	11.4
k56_4	583	10.5
k56_3	125	2.24
k56_2	404	7.25
k56_1	269	4.83

Figure 26. scRNA and Ab-Oligo-seq analysis of tumor infiltrating lymphocytes (TILs) and T cells from the tumor draining lymphonodes in a mouse model of adenocarcinoma. **A.** UMAP plot of about 7000 CD4+ and CD8+ T cells. **B.** The analysis has highlighted different new T cell subsets, especially among the TILs.

The study was also extended with analysis of the response to immunotherapy with anti-PD-1 antibodies and vaccines directed against tumor neoantigens. The efficacy of vaccines was assessed by both multiparametric cytometric analysis and single cell transcriptome analysis. Following a temporal kinetics, cell suspensions from lymph nodes of healthy mice and from draining tumors and lymph nodes were labeled with fluorochrome-conjugated antibodies and isolated at the IIGM cytometry platform.

Following bioinformatic analysis, new molecular factors specific to the different lymphocyte subpopulations were identified. In order to validate the relevance of the new markers, in recent months have been designed and developed models of efficient gene deletion through the CRISPR/CAS9 methodology.

Subproject "Study of **ontogenesis** *and heterogeneity of T lymphocytes during the immune response to tumor antigens and pathogens"*

L. Pace (PI), V. Russo, N. Brasu, C. De Intinis

The quality, efficacy, longevity, and tissue localization of T lymphocytes derive from the ability of naive T cells to diversify into several phenotypically distinct subpopulations with specific roles. The new single cell multi-omics technologies represent an important tool to study the complex heterogeneity of the immune system, both at phenotypic and functional level, at different time scales, even in different immunological contexts. The research Unit has set up a study on T lymphocytes isolated from mice infected with the pathogen *Liseria monocytogenes* using a new genomics methodology based on transcriptome sequencing at the single cell level (scRNA-seq). Bioinformatic analysis revealed a novel population of cycling T lymphocytes characterized by a unique gene transcript profile, including both genes involved in memory and effector functions. Understanding the origin, ontogeny, and molecular mechanisms underlying the development of these cyclic, multipotent intermediary cells has important implications in the field of cancer immunotherapy, especially for the treatment of cancer by modern genetic engineering methods with chimeric antigen receptor (CAR)-T cells, or in immunotherapy by adoptive transfer of tumor-infiltrating T lymphocytes. The identification of new immunological markers is therefore essential to isolate and analyze in detail the most relevant lymphocyte populations in the anti-tumor response. The aim of this study is therefore to identify new molecular targets to enhance the immune response against cancer and infections.

Aims

1. To develop a robust protocol for high-throughput sequencing of rare lymphocyte populations isolated from lymphatic organs and tissues;
2. To develop novel pipelines for bioinformatic analysis of scRNA-seq data;
3. Phenotype analysis of lymphocyte subpopulations in the anti-tumor response.

Results

The Unit has implemented the bioinformatics platform for the analysis of datasets, both public and produced in the laboratory, of single cell transcriptomics on T lymphocytes. The analyses allowed the unsupervised assignment of cells to different subpopulations, an allocation that can be easily compared with results obtained by cytofluorimetric analysis. Bioinformatics workflows were made flexible in order to adapt the parameters of each step to the internal variability of each sample, which can differ greatly between different experiments. Once the cell subpopulations were identified, the analysis activity was focused on the selection of the features (e.g. genes) most important to explain

the cell clustering obtained in each experiment. To thig goal, different methods were use, and several analysis pipelines were evaluated and integrated in our analysis workflow. Statistical methods and machine learning methods (e.g. extra tree classifiers) were used, manually setting thresholds regarding the number of comparisons between subgroups of cells in which each gene should be differentially enriched or expressed. This allowed the identification of specific genes for different subpopulations, such as CD4 T regulatory or CD8 T cells, memory, effector, cycling and exhausted. The analysis allowed to highlight new genes whose function has not been described so far in the literature. Subsequently, the attention has shifted to the search for sets of genes that could be modulated in a specific way in each lymphocyte subpopulation, in order to add an additional layer of information on the possible role in the context of the anti-tumor response. For this purpose, both reference gene expression profiles and sets of genes identified by clustering within individual experiments were used.

The presence of specific T cell subpopulations can be attributed to a common origin by clonal expansion following response to a tumor antigen. To this end, the investigation was further implemented with the identification of the T cell receptor (TCR), at the level of single T cell reactive against tumor epitopes generated following gene mutation. This methodology allowed to identify specific T cell clones for each cell subpopulation, thus validating the unsupervised clustering obtained in the previous steps, and showing how some lymphocyte subpopulations represent different states (e.g. metabolism, cell cycle) of the same T clone. The analysis of T clonotypes was deepened on about 4000 cells, derived from different donors, in order to assess the heterogeneity of the repertoire of each individual in correlation with the immunophenotype, and in response to treatment with checkpoint inhibitors. The trajectories that draw the process of gradual differentiation of lymphocyte subgroups were also analyzed by pseudotime techniques. The identification of T-cell subpopulations together with their phenotypic definition is a fundamental starting point for future experiments in which panels of genes of interest will be analyzed at a higher resolution, in order to isolate cellular subpopulations on which to perform further epigenetic analysis (e.g. scATAC-seq, ChIP-seq). In fact, it is expected that epigenetic analyses can provide in the near future a better understanding of the regulatory mechanisms underlying T cell activation, response and inactivation, identifying new molecular markers underlying these processes, and use them as possible targets for cellular reprogramming of exhausted T cells.

Project 2: "Study and monitoring of the immunological memory against SARS-CoV-2 after vaccination"

L. Pace (PI), V. Russo, N. Brasu, S. Aversano Stabile, M. Ceravolo, C. De Intinis, I. Elia

COVID-19, caused by infection with the SARS-CoV-2 virus, is a complex disease with a highly variable clinical course, in which the host's immune system can play both a protective and pathogenic role. In the last years, the high contagiousness of the SARS-CoV-2 virus has caused a pandemic disease spreads globally, with severe consequences in some of the subjects who contracted the infection. Understanding the mechanisms underlying the protective immune responses and long-term immunological memory specific to the SARS-CoV-2 virus are crucial and important aspects, in order to define the best therapeutic approaches to fight the virus and contains the infection later on with the reappearance of viral peaks. This study is focused on the analysis of the immune system's ability to develop short and long-term antigen-specific immunological memory, through the analysis of antigen specific CD4+ and CD8+ T cell responses against the SARS-CoV-2 virus, at the phenotypic, molecular and functional levels.

Aims

1. Identification of proteins and immunogenic peptides derived from SARS-CoV-2 recognized by activated CD4+ or CD8+ T cells (infection phase and memory).
2. To evaluate the presence and quality of immunological memory at different time points after infection, mediated by virus-specific CD4+ or CD8+ T cells in correlation with antibody titers
3. To determine gene expression profiles and new molecular markers, in correlation with the clonality of activated and memory antigen-specific T cells against SARS-CoV-2, using single cell multiomics approaches
4. To test whether specific CD4+ or CD8+ T-cell responses against SARS-CoV-2 are cross-reactive against endemic coronaviruses and novel viral variants;
5. To evaluate the long-term efficacy of vaccines

Results

In this project, we have collected several peripheral blood samples taken short- and long-term after recovery from COVID, or following administration of the Pfizer/Biontech vaccine.

The study was initially focused on analyzing the immunogenicity of SARS-CoV-2 structural antigens in terms of CD4+ or CD8+ T cell activation, using flow cytometry-based multiparameter analysis. The first phase of the study has highlighted the responses of different memory T cell subsets towards the more immunogenic antigens, both in terms of phenotype, inflammatory and cytotoxic activities

(**Figure 27**). In order to understand the differentiation and functional activity of these T lymphocyte subpopulations, we have analyzed the transcription profile of mRNAs a single cell level (scRNA-seq), combined with surface marker single cell proteomics (through the use of antibodies conjugated to oligonucleotides, scAb-seq) and scTRC-seq. The bioinformatics analysis has unravel the activation of unique gene signatures and specific surface markers for the most immunogenic antigens, in correlation with the intra- and inter-clonal heterogeneity of T cells determined with TCR sequencing. The results showed a different inflammatory response in subjects who contracted COVID six months after infection compared with healthy subjects, with the presence of different memory oligoclonal populations reactive against key SARS-Cov2 antigens.

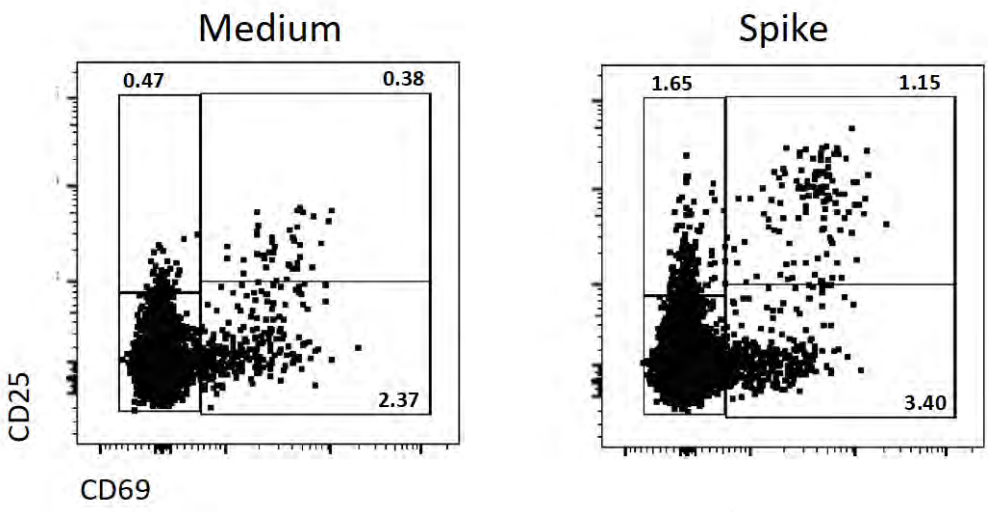


Figure 27. Flow cytometry analyses of CD8+ T cells, before (left) and after *ex vivo* activation with the Spike protein. The double positive CD69+ CD25+ CD8+ T cells are reactive against Spike

Collaborations

- Prof. A. Sapino (Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy)
- Dr. S. Amigorena (Curie Insitute, Paris, France)
- Dr. Christel Goudot (Curie Insitute, Paris, France)
- Prof. F. Fagioli (Ospedale Infantile Regina Margherita, Torino, Italy)

Funds and Grants

- Compagnia di San Paolo (PI L. Pace)
- Carrier Development Award Armenise-Harvard / Compagnia di San Paolo (PI: L. Pace)
- BD multi-omics (PI: L. Pace)
- AIRC (PI: L. Pace)
- Italian Ministry of Health (PI: L. Pace)

Epigenomics Unit

Research Group

Salvatore Oliviero, head of Unit (IIGM) and full Professor (UniTo)
Francesca Anselmi, PhD, technician (UniTo)
Carlo Castiglione, fellow (UniTo)
Hassan Dastsooz, PhD student (UniTo and IIGM)
Daniela Donna, PhD, tehcnician (UniTo)
Andrea Lauria, PhD student (UniTo and IIGM)
Corinne Magnetto, tehcnician (UniTo)
Mara Maldotti, post-doctoral fellow (UniTo)
Guohua Meng, PhD student (UniTo and IIGM)
Fatemeh Mirzadeh, PhD student (IIGM) until 30/06/2021
Ivan Molineris, researcher (UniTo)
Isabelle Polignano, PhD student (UniTo and IIGM)
Valentina Proserpio, PhD, post-doctoral fellow (IIGM)
Annalaura Tamburrini, PhD student (UniTo)
Claudia Vaccari, student (UniTo)

Our research unit combines molecular, cellular, and global genomic approaches to investigate epigenetic and epitranscriptomic modifications that occur in response to environmental or developmental signals and that are deregulated in tumor transformation. Several experimental tests have shown that the factors that bind DNA to specific sequences in collaboration with non-coding RNA and chromatin-modifying enzymes control epigenetic changes. The study of the interaction between these components is necessary to understand the rules that govern gene expression and cell fate. The work of the Unit is aimed at studying the mechanisms that determine these modifications and deciphering the rules to understand, and eventually influence, cell fate. Projects in progress in the laboratory address the study of epigenetic modifications that determine the immortality and pluripotency of stem cells and those involved in tumor transformation. These experiments allowed us to identify and characterize new molecular markers and cellular targets for anticancer therapies.

Projects

Project 1: "Epigenetic modifications involved in cellular transformation"

Subproject "Identification and characterization of non-coding RNAs in gene regulation"

S. Oliviero (PI), F. Anselmi

Understanding the heterogeneity of RNA structure is a major challenge in identifying actual RNA structures. RNA structural analysis by chemical analysis methods is powerful but suffers from the inherent limitation of providing only an average measurement of the basic reactivities of all coexisting conformations that are simultaneously sampled from an RNA species in a biological sample.

Aims

1. To study alternative coexisting in vivo conformations of cellular RNAs

Results

We wrote DRACO, an algorithm for deconvolution of coexisting RNA conformations in vivo in cells using data from mutational profiling experiments (Figure 9). In addition to cellular RNA analysis, I also analyzed the SARS-CoV-2 genome using dimethyl sulfate mutational profiling and sequencing (DMS-MaPseq). DRACO identifies multiple regions that fold into two mutually exclusive conformations, including a conserved structural switch in the 3' untranslated region. This work could pave the way for dissection of RNA structural heterogeneity.

DRACO will enable exploration of RNA structuring at unprecedented resolution and identification of transient and dynamic features of cellular transcriptomes.

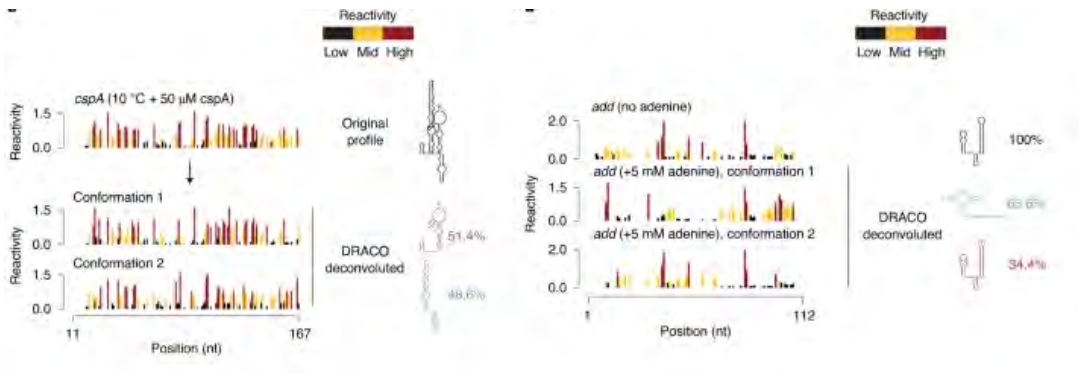


Figure 28. *In vitro* validation of DRACO. Original DMS-MaPseq and deconvoluted DRACO profiles for cspA 5' UTR (left) and V. vulnificus add riboswitch, in the absence or presence of 5 mM adenine (right). Schematic representation of the structures is shown along with estimated relative abundances.

Collaborations

- Prof. D. Incarnato (Groningen University, Groningen, The netherlands).

Subproject "Metabolic control of DNA methylation in naïve pluripotent cells"

S. Oliviero (PI), S. Rapelli, A. Lauria

Epiblast and embryonic stem (ESC) cells give rise to all adult cells. This evolutionary plasticity is associated with genome hypomethylation. After fertilization, the zygotic genome is demethylated to allow the differentiation of embryonic cells into different tissues. DNA methylation occurs on the 5 carbon of cytosine (5mC) and is catalyzed by DNA methyltransferases (DNMTs).

The Ten-Eleven translocation proteins (Tets) promote the oxidation of 5mC to hydroxymethylcytosine (h5mC)1,2. Further Tets-mediated oxidation steps lead to the conversion of h5mC to unmodified cytosine. Both Dnmts and Tets are dynamically expressed during early development, leading to a local minimum of 5mC at the preimplantation blastocyst stage at embryonic day 3.5 (E3.5).

Aims

1. To study how expression of Dnmts and Tets is controlled in the early embryo

Results

Regulation: Our experiments showed that LIF-Stat3 signaling induces genomic hypomethylation via metabolic reconfiguration. Stat3-/- ESCs show a decrease in α-ketoglutarate production from glutamine, leading to an increase in Dnmt3a and Dnmt3b expression and DNA methylation. Notably, genome-wide methylation is dynamically controlled through modulation of α-ketoglutarate availability or activation of Stat3 in mitochondria. Alpha-ketoglutarate links metabolism to the epigenome by reducing the expression of Otx2 and its targets Dnmt3a and Dnmt3b (**Figure 29**). Genetic inactivation of Otx2 or Dnmt3a and Dnmt3b results in genomic hypomethylation even in the absence of active LIF-Stat3. Stat3-/- ESCs show increased methylation in imprinting control regions and altered expression of related transcripts. Single cell analyses of Stat3-/- embryos confirmed deregulated expression of Otx2, Dnmt3a, and Dnmt3b. Several tumors show hyperactivation of Stat3 and abnormal DNA methylation; therefore, the molecular module we describe could be exploited in pathological conditions

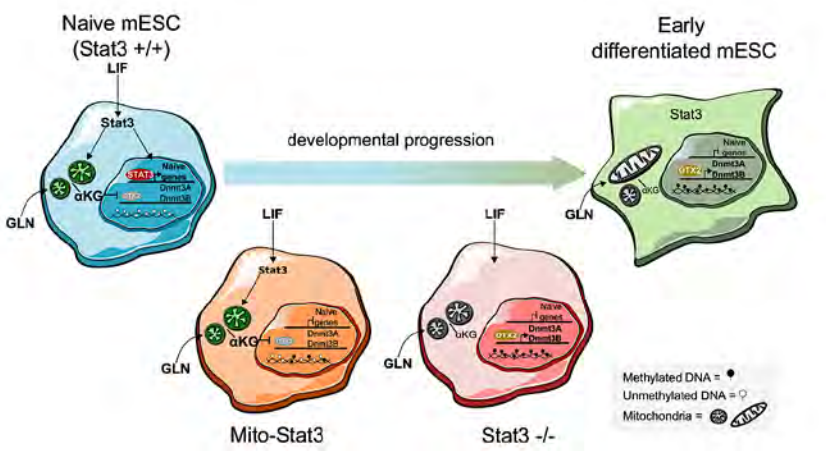


Figure 29. Model deduced from experimental data in which mitochondrial Stat3 induces the release of α–ketoglutarate (αKG) that in turn inhibits the expression of DNMTs blocking their regulators via Otx2.

Function: Proper DNA methylation during the early stages of development is essential to determine specific expression and target cells to their specification. However, the molecular targets as well as the mechanisms that determine the specificity of the methylation mechanism during differentiation are not fully elucidated.

The role of DNA methylases in the differentiation of embryonic stem cells was made by differentiation of embryonic stem cells to neuroectoderm and endomesoderm. The de novo DNMTs DNMT3A and DNMT3B are strongly induced in the transition from embryonic stem cells to epiblasts. Their level of expression rises considerably from ESCs to Epiblasts with an increase of DNA methylation. DNMT3B knockout shows a strong impairment of DNA methylation in epiblasts.

DNMT3A is mainly expressed in neuroectoderm cells while DNMT3B is mostly expressed in the mesendoderm. Interestingly, the knockout of DNMT3A or DNMT3B show quite different phenotypes, while DNMT3A KO are impaired in neural development DNMT3B KO do not appear to play a role in this developmental pathway. Most of the alterations are observed in intragenic and intergenic regulatory regions.

To elucidate the mechanisms underlying the regulation of DNA methylation we generated homozygous DNMT3A and DNMT3B knockout cell lines from the E14 mouse embryonic stem cell (ESCs) line. To study the early stages of development and the impact of gene knockout in vitro, ESC DNMT3A - / - (3AKO), DNMT3B - / - (3BKO) and wild type (WT) were differentiated into three-dimensional structures called embryoid bodies (EB). The ESCs within the EBs undergo differentiation along the three germinal ligneous ones (**Figure 30**). This behavior is similar to what happens in vivo: in fact, after 3 days of differentiation, the EBs resemble the early post-implantation embryo (E5.5), where most of the ICM is composed of epiblast stem cells and an overall increase in DNA methylation.

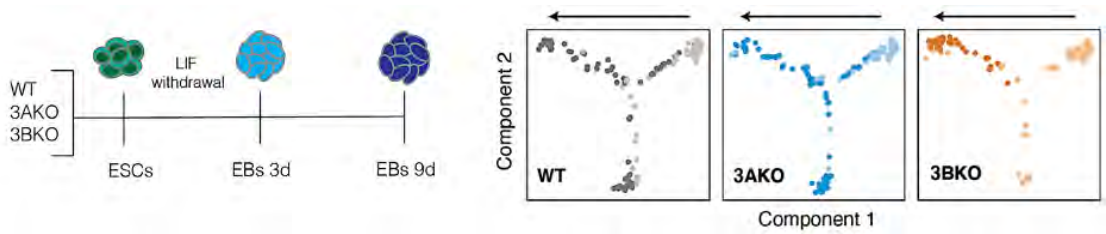


Figure 30. Knockout experiments demonstrate that DNMTs play different roles on cell development. As shown in the right panels DNMT3B knockout is required for early cell specification while DNMT3A function is not required.

Our experiments show that at this stage of differentiation DNMT3B-dependent DNA methylation that methylates regulatory regions of key developmental transcription factors and results in an epigenetic priming that ensures gene regulation in later stages. Using a combination of in vitro stem cell differentiation, loss of function experiments and RNA sequencing genomic DNA methylation analyzes we have dissected the specific role of DNMT3B in cell specification.

Subproject "Circular RNA biogenesis and splicing fidelity to support cancer cell fitness"

S. Oliviero (PI), M. Maldotti, I. Polignano, V. Proserpio

In eukaryotic cells, most genes harbor intronic sequences that are removed during RNA splicing and transcript maturation. This process is regulated by the spliceosome, which comprises small noncoding RNAs, spliceosome core proteins, and other ancillary factors. Under normal physiological conditions, proper regulation of splicing provides the cell with the opportunity to control gene expression in the absence of genetic alterations. By expressing alternative isoforms of the same gene, the cell can regulate both the inclusion/exclusion of specific protein and/or RNA domains thereby influencing both the fate and function of transcription. Regulation of alternative splicing plays a central role in development, cellular differentiation, as well as cellular response to external or internal stimuli. However, the very phenotypic plasticity afforded by the splicing machinery can act against the cell and the organism and confer competitive advantages to cells in pathological conditions such as cancer. Indeed, alterations in the expression of specific isoforms of certain genes and splicing factors themselves can promote cell proliferation.

Aims

1. To identify splicing factors important for prostate cancer eligibility

Results

We performed pooled shRNA screenings in vitro and in vivo. Our screenings identified heterogeneous nuclear ribonucleoprotein M (HNRNPM) as a regulator of prostate cancer cell growth. RNA and eCLIP sequencing identified binding of HNRNPM to transcripts of key homeostatic genes. Binding of HNRNPM to its targets prevents aberrant exon inclusion and backslicing events. In both linear and circular transcripts with misplicing, HNRNPM preferentially binds to GU-rich elements in long flanking proximal introns. Mimicry of HNRNPM-dependent linear splicing events using antisense splicing-switching oligonucleotides was sufficient to inhibit prostate cancer cell growth. This suggests that the dependence of PCa on HNRNPM is likely the result of splicing error of key homeostatic coding and noncoding genes. Our results were further confirmed in other solid tumors. Our data reveal a role for HNRNPM in supporting tumor cell fitness. Thus, inhibition of HNRNPM activity is a potential therapeutic strategy to suppress the growth of PCa and other solid tumors. (**Figure 31**).

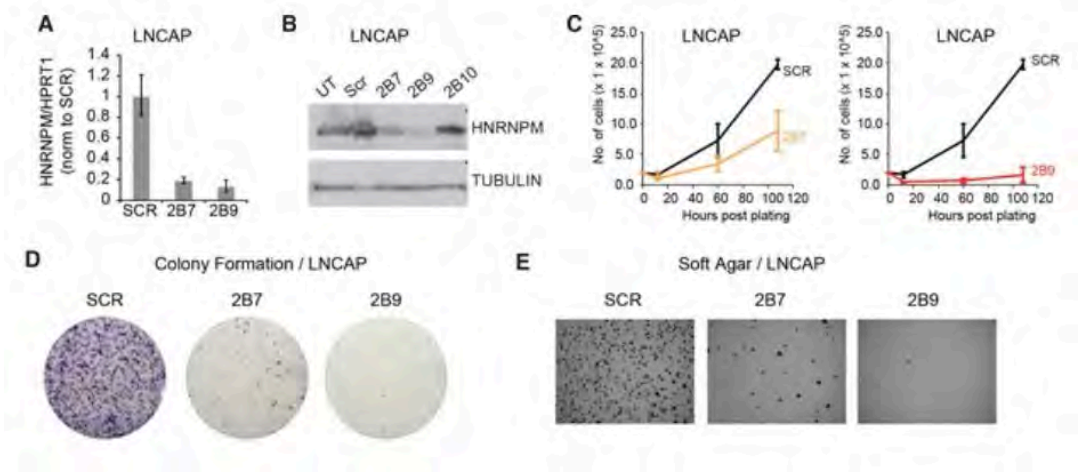


Figure 31. HNRNPM inhibits prostate cancer cell growth in vitro and in vivo. HNRNPM RNA (A) and protein levels (B) to expression of scrambled (Scr) or HNRNPM-specific shRNAs (2B7, 2B9, 2B10) in LNCAP cells. (C) Cell proliferation assays of LNCAP cells expressing shRNAs encoded or specific for HNRNPM (2B7 and 2B9). Mean and standard deviation of two biological replicates are shown. (D) Colony formation assays and (E) anchorage-independent growth (agar assays) of LNCAP cells expressing shRNAs encoded by or specific for HNRNPM (2B7 and 2B9).

Subproject "Role of BCOR internal tandem duplication (BCOR ITD) in tumorigenesis and stemness"

S. Oliviero (PI), H. Dastsooz, V. Proserpio

BCOR (BCL-6 interacting corepressor) is a B-cell/lymphoma 6 (BCL6) interacting corepressor, a POZ/zinc finger transcriptional repressor that is essential for germinal center formation and may have effects on apoptosis. BCOR ITDs have recently been found in tumors, such as clear cell sarcoma of the kidney. Therefore, we hypothesized that overexpression of this ITD BCOR may have some effects on tumorigenesis.

Aims

1. To identify molecular targets of the BCOR mutant

Results

We generated a stable cell line expressing the BCOR-ITD identified in one of our patients. At first, the above construct was subcloned into an inducible vector. Then, we decided to delete the endogenous BCOR gene in HEK cells by CRISPR/Cas9 system and transfected specific gRNA into

HEK cells together with Cas9GFP. After transfection, cells were evaluated for deletion of the affected region of the BCOR gene by PCR using internal and external primers, real-time PCR, and Sanger sequencing. After confirmation of BCOR deletion, we sorted cells using a cell sorter to obtain single cells in a 96-well plate. Then, we monitored the sorted cells for the presence of cell colonies. When the colonies were ready, we evaluated them for homozygous BCOR knockout. To prepare the stable cell line, we treated the cells with puromycin. After 3 weeks, we were able to achieve the stable cell line, which was confirmed by overexpression of BCOR in RNA and protein levels. Now we are performing other experiments on these cells such as Western blotting, chromatin immunoprecipitation with biotin, RNA sequencing.

Since our aim is to investigate the role of BCOR-ITD during differentiation of human embryonic stem cells into skeletal myogenic progenitors and in transdifferentiation of fibroblasts into skeletal muscle cells by overexpression of MyoD, we performed knock-in by CRISPR/Cas9 for BCOR-ITD in these cells with wild type gene destruction. We provided several gRNAs along with the template with this BCOR-ITD knock-in in these cells. Evaluation of H9-BCOR-ITD clones is ongoing, we can then investigate the role of duplication during differentiation of human embryonic stem cells into skeletal muscle cells and in the transdifferentiation model of fibroblasts into skeletal muscle cells.

Project 2: "Analysis of gene alterations in pediatric sarcomas"

S. Oliviero (PI), H. Dastsooz, V. Proserpio

In recent years it has been possible to study the role of genetic and epigenetic alterations in the onset of cancer. Our Unit, as part of a collaboration with researchers from IIGM, the University of Turin, and the pediatric hospital of Turin, identified somatic mutations, gene fusions, epigenetic alterations in tumors of patients including sarcomas, and hematopoietic, breast, and colon tumors.

Aims

1. Functional analysis of genomic alterations and gene fusions in pediatric and young adult sarcomas
2. Study of the molecular mechanisms that determine the onset and progression of sarcomas

These tumors are characterized by a low level of somatic mutations that indicate that epigenetic and developmental alterations play an important role in the genesis and evolution of this tumor.

Results

Using gene fusion analyses, we identified the fusion SS18-SSX1 in a synovial sarcoma involving the SS18 component of the mammalian SWI / SNF chromatin remodeling (BAF) and the transcription factor SSX1 and an ATRX-ERG fusion involving the chromatin remodeling factor ATRX and the ERG gene in an osteosarcoma and other alterations in genes involved in epigenetic regulators. In particular we showed a rare duplication of the last exon of the epigenetic factor BCOR. We analyzed both the genomic alterations and the transcriptional picture in primary tumors and in tumors induced in immunosuppressed mice. Our data confirm the presence of gene alterations also in mouse models and the regulatory framework in these tumors shows alterations of regulatory pathway involved in cell proliferation. We generated mouse embryonic cell lines carrying the BCOR with the duplicated exon and ongoing experiments aims at the understanding of the role of this duplication in the tumor onset.

Project 3: "Identification and characterization of non-coding RNA in gene regulation"

S. Oliviero (PI), M. Maldotti, A. Lauria, A. Tamburrini, I. Molineris

Most cellular RNAs are not translated into proteins, but perform their biological functions directly as RNA. They can also be post-transcriptionally modified and these modifications play a role in gene regulation.

The modification of adenine at position 6 (N6-methyladenosine, m6A) is deposited by the nuclear heterodimeric complex composed of methyltransferase-like-3 (METTL3) and 14 (METTL14), N6-methyladenosine is the most abundant modification of RNA transcripts coding and non-coding polymerase II and the one with the greatest impact on their dynamic regulation. m6A can control any aspect of mRNA post-transcriptional regulation, including splicing, export, stability, and translation.

Aims

1. To study the role of non-coding RNAs in the regulation of epigenetic and transcriptional modifications during cell differentiation and transformation
2. To identify interactions of lncRNA with epigenetic regulators

Results

Circ-ZNF609 circular RNA is an interesting example to study because high levels of expression are related to proliferative conditions in human myoblasts and rhabdomyosarcoma tumors, while its depletion decreases proliferation in both systems. circ-ZNF609 contains an open reading frame (ORF) and can be translated into two proteins using two alternative initiation codons, in a splice-dependent and cap-independent manner via an as yet unidentified mechanism.

We have shown that, in addition to their role in circRNA biogenesis, m6A modifications on circ-ZNF609 play a relevant role in its cap-independent translation. In particular, we identified the YTHDF3 and eIF4G2 proteins as important factors to mediate this process (**Figure 32**).

Collectively, these data reveal the important contribution of m6A modifications to circRNA biogenesis and function.

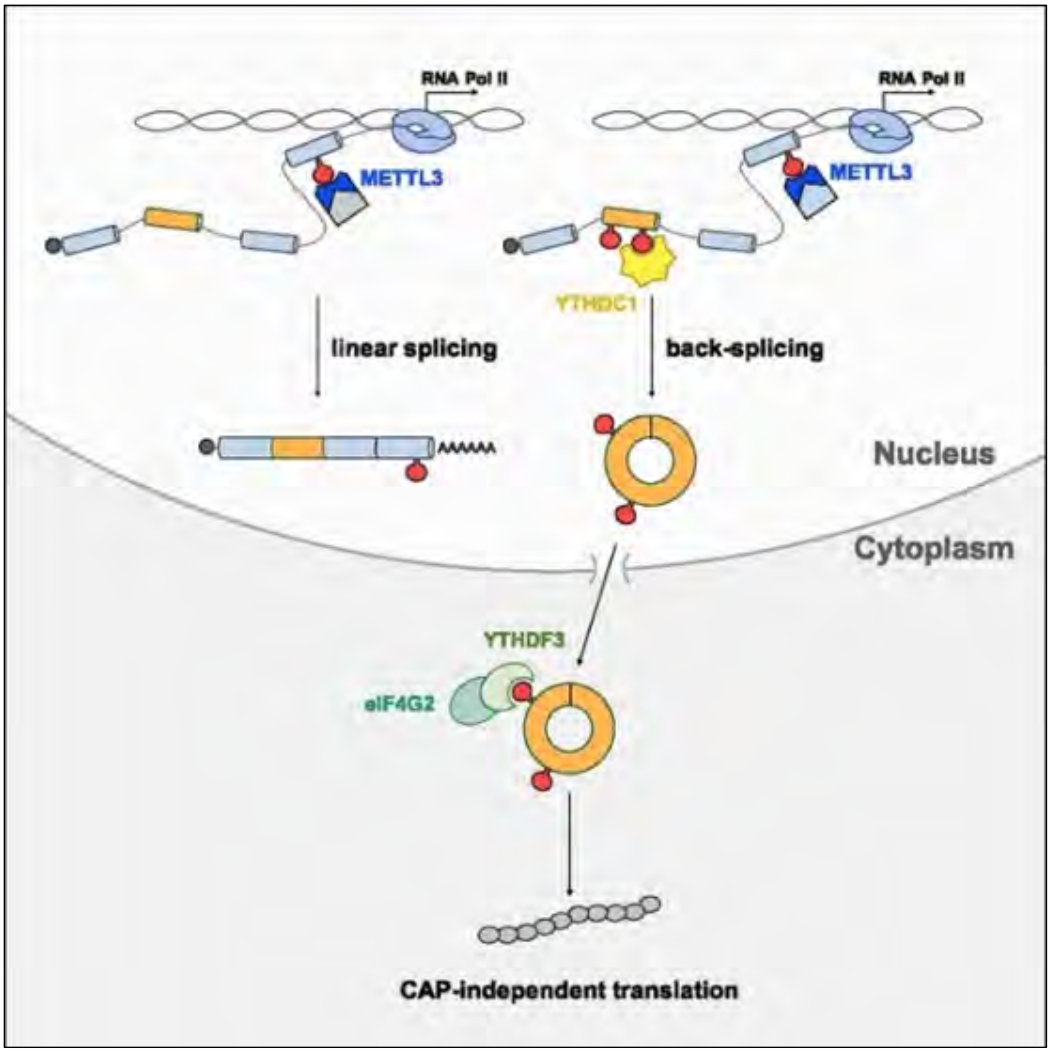


Figure 32. Model deduced from experimental data in which the METTL3 enzyme introduces the m6A modification

Collaborations

- Prof. G. Martello (University of Padova, Padova, Italy)
- Prof. I. Bozzoni (La Sapienza University, Rome, Italy)
- Prof. M. Studer (CNRS, Nice, France)
- Prof. F. Neri (Leibniz Institute of Aging, Jena, Germany)
- Prof. D. Incarnato (University of Groningen, Groningen, the Netherlands)

Services carried out by the Unit

The Epigenomics group is responsible for the genomics platform carried out in co-management between IIGM and the University of Turin for the NGS sequencing service which includes analysis using numerous genomic techniques: whole-genome-seq, exome-seq, WGBS, MAB-seq, DECAP-seq, RNA-seq, scRNA-seq, CIRP-seq, 2-OMe-seq.

Funds and Grants

- Compagnia di San Paolo (PI: S. Oliviero)

Statistical Inference and Computational Biology Unit

Research Group

Andrea Pagnani, PhD, head of Unit (IIGM) and full Professor (PoliTo)

Carla Bosia, PhD, post-doctoral fellow (PoliTo and IIGM)

Alfredo Braunstein, PhD, associate Professor (PoliTo and IIGM)

Carlo Cosimo Campa, PhD, junior researcher (PoliTo and IIGM) until 28/02/2121

Luca Dall'Asta, PhD, associate Professor (PoliTo)

Andrea De Martino, PhD (CNR)

Elsi Ferro, PhD student (PoliTo)

Andrea Gamba, PhD, associate Professor (PoliTo and IIGM)

Anna Paola Muntoni, PhD, fellow (IIGM)

Andrea Piras, fellow (IIGM)

Luca Sesta, PhD student (PoliTo)

Simone Tealdi, fellow (PoliTo and IIGM) until 30/10/2121

Guido Uguzzoni, PhD, post-doctoral fellow (PoliTo and IIGM)

The core activity of the unit is the development of computational techniques inspired by statistical mechanics for problems of inference, optimization, and for applications dedicated to the analysis of experimental databases and in particular to problems of inference in complex biological systems.

The search for specific techniques for the analysis of large amounts of experimental data obtained through modern sequencing and gene expression tools poses formidable problems from the computational point of view. Such problems are often intractable using standard algorithmic techniques. However, in recent years very efficient inference and optimization techniques inspired by the statistical physics of disordered systems have been developed. The first applications to inference problems in biology should be viewed as proofs of concept of the possibility to make these techniques available to the community working in the computational biology field.

These new computational techniques are of great interest for biological research: from the quantitative analysis of large-scale biological databases (gene expression, DNA-copy number, evolutionary variability of homologous protein sequences in different species), to the inference of complex interaction networks (protein-protein, cellular signal transduction, and interactions for the regulation of gene expression). These techniques are also promising when treating inference problems in the analysis of the genetic basis of cancer.

A consistent effort has been devoted to the planning for the experimental activities of the Unit. Particular attention has been paid to the organization of the molecular biology section, and to the implementation of last generation genetic engineering techniques (CRISPR-Cas).

Projects

Project 1: "Cancer and Metabolism"

The activity of the Unit develops along five lines: (1) the analysis of the role of microRNAs in cell differentiation; (2) the genome-wide analysis of cell metabolism, especially of the transition to the carbon overflow regime, by means of machine learning methods; (3) the quantitative study of trade-offs between growth and lag in cancer cell populations; (4) the mathematical characterization of regulatory and ecological networks, (5) the theoretical modeling of the molecular sorting process.

Subproject "Role of microRNAs in cellular differentiation"

A. De Martino (PI), A. Pagnani, C. Bosia, E. Ferro

Aims and Results

Pluripotent embryonic stem cells (ESCs) are capable of generating the full spectrum of cells in an organism, distinguished by different states of expression. In many organisms (such as Drosophila), the initiation of the differentiation process is given by the gradient of an exogenous morphogen, the level of which drives different cells to different cell fates. In these cases, the physical mechanism that enables this process is the ability of the cells to estimate the level of morphogen (via receptors) in a surprisingly precise manner. In other organisms, however, differentiation does not appear to be driven by exogenous factors. This is the case of human ESCs, which are able to generate different cell types without apparently using morphogens. An open problem of considerable importance for developmental biology concerns precisely the identification of the mechanisms that drive differentiation in these cases. In collaboration with the group of P.A. Sharp and S. Garg (MIT) we have investigated the role of a particular mechanism active in human ESCs, related to microRNAs (miRNAs). Briefly, using experiments and mathematical models in conjunction, we have shown that different cellular states can be induced endogenously by intrinsic variability in miRNA expression levels from cell to cell. Such variability naturally leads to a modulation of the expression levels of genes controlling pluripotency, which in turn, due to the interaction network that binds miRNAs to their targets, is organized into several distinguishable "states", corresponding to different cell types. Ongoing studies will allow to delineate the limits of this mechanism (e.g. answering the question of how many different cell states can be generated at most using only pluripotency genes as markers), and thus to understand if human embryonic development is optimized to generate the maximum number of cell types allowed or if instead the molecular mechanisms aim at the robustness of a reduced number of types.

Subproject "Theoretical modeling of the molecular sorting process"

A. De Martino (PI), A. Pagnani, C. Bosia, E. Ferro, A.P. Muntoni

Aims and Results

Current understanding of how cellular metabolism works on a genome-wide scale relies almost exclusively on the use of detailed mathematical models of the metabolic reaction network that assume that cellular activity optimizes a well-defined, context-dependent objective function. In this way, one can derive the optimal states of a metabolic network and make predictions about its functionality. This approach, now fundamental to reactor biotechnology, has shown many limitations in recent years when attempts have been made to apply it to biomedical contexts. The core of the problem is the optimality assumption. In fact, it seems increasingly clear that cellular activities impose severe constraints on metabolism, and that often what is observed is not the result of an optimization but rather of a series of trade-offs that balance the benefits of using a particular pathway with the relative costs (e.g. in terms of energy). Motivated by these problems, we attempted an inverse approach to the metabolic modeling problem: starting with data, we make use of machine learning techniques to infer the states of the metabolic network without making functional assumptions about optimality. In the first work we focused on the study of the metabolism of the bacterium Escherichia coli, aiming to obtain a detailed description of the transition that leads from a predominantly respiratory regime to a predominantly fermentative one. This transition mimics the transition, known as the "Warburg effect", that takes place in cancer cells. In the future we aim to use the methods developed for bacteria to study the same transition in the case of tumors.

Subproject "Quantitative study of trade-offs between growth and lag in cancer cell populations"

A. Pagnani (PI), A. Gamba, L. Dall'Asta

Aims and Results

Growth rate is an elementary measure of the fitness of an exponentially growing population of cells (bacteria, tumors), and helps determine energy and molecular resource allocation, expression levels, etc. However, the growth rate is in turn influenced by environmental factors as well as by the very factors it helps to determine. This situation generates a number of trade-offs in population growth, which have been the subject of much attention in recent years. Well-known examples of these trade-offs are the relationships existing between growth rate and adaptation time in a given environment (lag time), which have been well studied especially in bacterial populations. Beyond scientific

interest, these relationships directly reflect the complex relationship between metabolism and gene expression in individual cells. By integrating experiments and mathematical models, we have characterized a series of such trade-offs within cancer cell populations (Jurkat, K562), using inoculum density as a control parameter to regulate growth rate. This study led to the identification of a number of novel statistical regularities potentially relevant to the understanding of tumor development (related, for example, to the microscopic origin of the relationship between initial density and subsequent growth rate), the in-depth analysis of which will occupy a substantial part of our future work.

Subproject "Mathematical characterization of regulatory and ecological networks"

A. De Martino (PI), A. Pagnani, A.P. Muntoni

We have been working to improve the mathematical characterization of general models of regulatory networks (gene circuits), in order to refine currently available results and allow better comparison with experimental data. In particular, we have focused on a class of models (developed by Bialek and collaborators since 2008) that provide an ideal benchmark for the optimal behavior of minimal gene circuits (e.g. Transcription Factor-Transduced Protein). Such models are complicated to study computationally due to the high dimension of the parameter space in which they are defined. Mathematical solutions are therefore essential. Until now, these solutions were known only in a very particular limit, namely, one in which stochasticity in the behavior of the different components can be completely neglected. By exploiting a technique borrowed from physics, we were able to obtain a more general solution that takes such stochasticity into account. The results obtained improve the understanding of the role of stochastic effects for the robustness and reproducibility of the expression levels. In another area, we instead studied a class of ecological network models (known as MacArthur models) in order to apply them to microbiome characterizations.

Aims

1. Phenotype-related exploration and exploitation of cell populations characterized by heterogeneous fitness distribution

Results

In the quantitative study of cellular metabolism, we attempted to identify the physical principle that drives the dysregulation of energy metabolism in the proliferative regime (one of the hallmarks of cancer). To this end, we developed a mathematical model that integrates the biosynthetic costs

associated with growth and replication with those associated with the need to efficiently use available nutrients. In particular, we showed that the Warburg effect emerges from the balance between the different energy constraints imposed by growth on these two aspects. The obtained system was implemented in a genome-scale metabolism model, finding quantitative agreement with proteomics and fluxomics data in bacteria. This approach based on a general physical principle, allows to develop similar models for cancer cells, and this is the direction in which we are currently moving.

We have also derived the first transcriptomic resolution map of effective RNA interactions induced by microRNAs. To this end, we have integrated datasets of miRNA-target interactions from human cells with a mathematical model that allows us to estimate the intensity of effective interactions. Among the various aspects considered, we were able to show that some fundamental features of RNA crosstalk are uniquely related to the interaction pattern between RNA and miRNA. In this sense, these features thus appear to have been favored by natural selection. The observed large-scale effects, such as stabilization of expression levels or extended responsiveness to perturbations, suggest that these features are essential for cellular regulation despite being mediated by typically very weak interactions.

Subproject "Biophysical modeling of molecular sorting"

A. Pagnani (PI), A. Gamba, L. Dall'Asta

Aims and Results

The main goal of the project is the mathematical modeling of molecular sorting mechanism that allows protein and lipids to be specifically localized in nanovesicles and routed back and forth from the membrane to the nucleus. This mechanism is essential for the maintenance of the correct chemical identity of the various regions of the cell membrane. It is believed that the understanding of the protein sorting process at the systemic level and the identification of its main control parameters will allow in the future to better understand and counteract diseases related to the misregulation of protein trafficking, including, in particular, cancer.

With this in mind, a physical theory has been developed that describes the diffusion of molecules across the membrane to sorting centers that are augmented by the uptake of specific proteins, and that are extracted when they reach a particular critical size.

The following parameters are involved in the model, shown schematically in **Figure 33**:

- k_I: intake vesicular flux assuming an “infinite” molecular reservoir
- k_D: molecular diffusivity in the vesicle
- $g^{\#nn}$ adimensional aggregation coefficient
- (*#nn* is the number of nearby molecules at time t
- k_E: vesicular out take flux

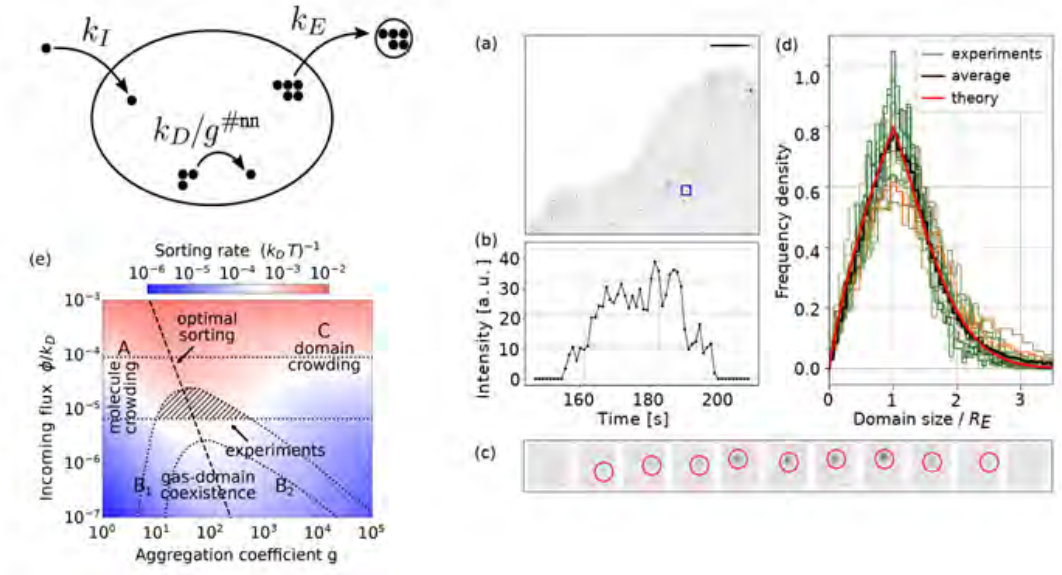


Figure 33. Minimal model of molecular sortingSorting.

The theoretical predictions of the model were compared with the results of computer simulations and the results of experimental observations.

The process turns out to be controlled by two fundamental parameters: the aggregation coefficient, g, and the flux of input molecules, kI. The efficiency of the sorting process turns out to be optimal for intermediate values of the aggregation coefficient, g. In this optimal regime, the molecule density is minimal and the process obeys simple scaling laws. A phase diagram of the process is shown in **Figure 33** (panel e), indicating the optimal region of control parameter values.

Quantitative measurements in primary endothelial cells are compatible with the model prediction *in vivo*. In **Figure 33**, we show some phenomenological details: in panel (a) we show Total Internal Reflection Fluorimetry micrography (scale 10 μm); (b) a temporal series of the fluorescence intensity in the region of interest; (c) time series with 10 s (1 pixel = 180nm) temporal resolution; (d) empirical density distribution of the domain radius overlaid to the theoretical prediction.

Subproject "Growth Rates and Metabolism"

A. De Martino (PI), A. Pagnani, A. Braunstein

Aims and Results

The coupling between the physiology of cell growth and cellular composition has been actively investigated since the 1940s. Such interdependence is perhaps best expressed in a quantitative way by the 'growth laws' that relate the proteins, DNA and RNA content of proliferating bacteria to their growth rate. Many such laws have been experimentally characterized and many more are currently being probed. While more work is required to fully unravel the biological underpinnings and limits of validity of growth laws, the emerged scenario suggests that proteome organization in bacteria is actively regulated in response to the growth conditions.

Recent experiments have validated the picture according to which, as the growth rate changes, bacteria adjust the relative amounts of ribosome-affiliated proteins, nutrient scavenging proteins and metabolic enzymes, so as to optimize their growth performance and energy production strategy. Revealingly, the same strategies appear to characterize energy production in proliferating cell types other than bacteria, most notably in tumors. Several phenomenological models explain the origin of different growth laws at coarse-grained levels. In contrast, full-fledged genome-scale approaches probing such relationships *in silico* are far less developed.

Constraint-based models (CBMs) are powerful mathematical tools that are used to examine metabolic networks at genome scale. Starting from a non-equilibrium steady state assumption for metabolic fluxes, CBMs define the space of feasible reaction profiles as those satisfying various physical and biochemical constraints. Despite many efforts, several aspects of the phenomenology of growth laws are currently not reproduced by standard CBMs. It has now become clear that factors such as finite pathway capacity, spatial constraints on enzyme distributions or costs associated to gene expression and protein synthesis may be the missing ingredients. For instance, CBMs accounting for effects due to molecular crowding have been successfully employed to recover aspects of cellular energetics that are specific of proliferating cells or tissues with high energetic demands More recently, a novel CBM called Constrained Allocation Flux Balance Analysis has been proposed, in which regulation is accounted for effectively through a single additional global constraint on fluxes that encodes for the relative adjustment of proteome sectors at different growth rates.

Project 2: "Optimization method of chemico-physical protein properties"

Subproject "Biophysical modeling of molecular sorting"

A. Pagnani (PI), A.P. Muntoni, L. Sesta, G. Uguzzoni

Aims and Results

Antibodies and other binding proteins are the subject of intense interest in chemical and pharmaceutical research, with an expanding market of products and services. The present project relates to the general problem of the silico design of proteins with high affinity with a specific target molecule. Despite the intense research work from an experimental and theoretical point of view, the problem of the prediction of aminoacid sequences that generate a specific behavior remains a difficult problem. This is particularly true for applications that require a high performance in terms of specificity, as in the case of industrial enzymes, or proteins with therapeutic purposes such as monoclonal antibodies. The experimental process of direct evolution is a powerful approach to protein engineering, as evidenced by the recent 2018 Nobel Prize in chemistry to Frances H. Arnold, George P. Smith and Gregory P. Winter for their research in this research field. Since its discovery in the early 1990s this approach has been applied with a high technological impact to the optimization of enzymes and antibodies for chemical synthesis. In the last decade, thanks to the spread of a new type of experimental biochemistry analysis techniques called "deep mutational scanning", they have combined large-scale selection of a combinatorial library of mutants specific to directed evolution, with recent sequencing techniques high-throughput", with the aim of quantifying the functionality of a number of the order of 10^6 variants of a given protein (**Figure 34**).

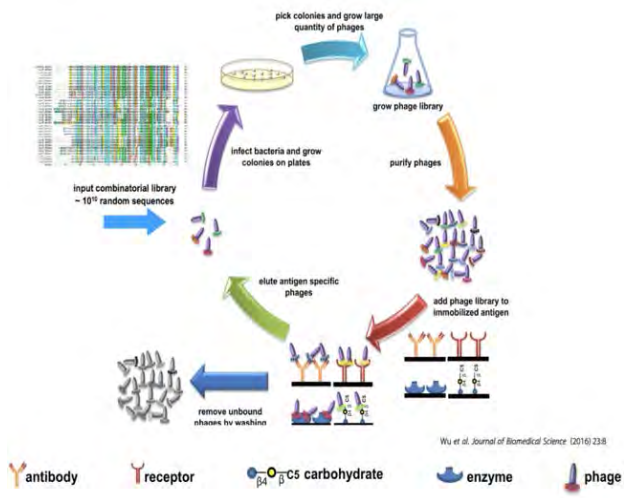


Figure 34. Typical set up for phage display directed evolution experiments.

The method proposes an innovative methodology for the evaluation of protein sequences starting from "deep mutational scanning" experiments for a rational selection of mutants with an optimal affinity of binding towards a specific target. The ultimate aim of the technique is to generate, based on a predefined target, a library of proteins with increased biochemical characteristics compared to the natural counterpart. A typical example of functional optimization implemented by the method we

propose is represented by the binding affinity with target molecules that are important from a biological and/or therapeutic point of view. The originality of the approach proposed here is twofold: (i) an accurate statistical modeling of the different phases of the directed evolution process; (ii) the formulation of the functional relationship between the peptide sequence and the energy of binding with the target of interest based on a cost function that takes into account the epistatic between the aminoacids of the sequence.

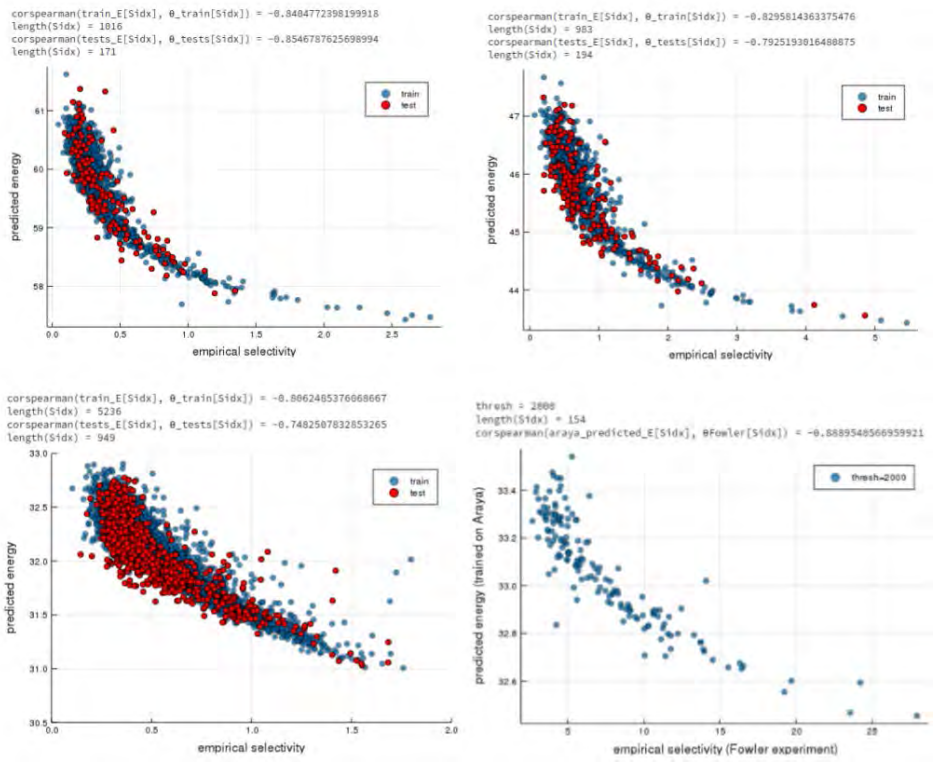


Figure 35. Predicted vs. empirical selectivities over a train and test-set for different directed evolution experiments (see text).

As an example of prediction of binding selectivity of mutant antibodies on data deriving from a DMS experiment performed by phage display, we consider a test of the model on the data published by Boyer S, *et al.,* in "Hierarchy and extremes in selections from pools of randomized proteins - PNAS 2016". The DMS experiment reported is aimed at analyzing a library of antibodies against a neutral synthetic polymer, polyvinylpyrrolidone (PVP). The experiment was conducted, in this case, by carrying out three rounds of amplification and selection through phage display. The initial library was

realized by mutagenesis at saturation of an anti-PVP antibody on four consecutive aminoacids of the region that determines complementarity 3 (CDR3). The dataset derived from these experiments was randomly divided into a training set for the model and a set of analyzes in which the model's binding energy was compared as shown in **Figure 35**.

The developed method is an innovative computational strategy that starting from sequencing data of screening experiments produces a statistical model of the genotype-phenotype relationship. The method has the dual purpose of: (i) produce an accurate statistical description of the time series of the differential composition of library mutants; (ii) predict single sequences or libraries rationally designed for enhanced biochemical activity against the chosen target. The novelty of the method is in its unsupervised feature and the inclusion of epistatic terms in modeling mutant-specific selection. We tested the method on five large scale mutational scanning experiments obtaining in all cases accurate predictions of the mutational effect on fitness. The inferred fitness landscape is robust to experimental noise and shows high generalization power. As a by-product of the method, the set of inferred model parameters can be used for structural predictions.

Subproject "Genetic Engineering of the CRISP3/Cas system"

A. Pagnani (PI), C.C. Campa, G. Uguzzoni, A.P. Muntoni

Aims

Among the main question of cell physiology is about the ability of cells to perform a variety of tasks using a limited set of genes. Thanks to a sophisticated regulatory mechanism, small DNA fragments orchestrate functional interactions among different genes that ultimately define the cell behavior. Modern "omics" technology (e.g. transcriptomics, proteomics, etc.) provide fundamental tools to dissect the topology and the functionality of this interaction network.

Recently the CRISPR/Cas technology allowed to unveil the functional role of many genes involved in complex human pathologies. However, one of the main limits of this technique is related with its difficulty to act simultaneously on many genes. Overcoming this limitation would pave the road to large-scale engineering of complex cellular regulatory programs. Coupling synthetic biology techniques with statistical inference driven modeling, we aim at designing new CRISPR/Cas variants with improved efficacy.

Results

Among the large number of CRISPR/Cas systems available, we selected the class II – type V RNA driven endonuclease CRISPR/Cas12a. This system induces gene modification based on a miniaturized CRISPR model that allows for very precise target modifications. By using a molecular modeling approach, we identified a central region of the structure of the Cas12a protein, that is responsible for the target DNA recognition. This region comprises a part of the recognition (REC) site, and a part of the nuclease region (NUC). These two segments of the Cas12a protein, generate a binding pocket for the target DNA and the RNA guide. It is on this region that we are going to introduce mutations to generate an improved Cas12a variant (**Figure 36**).

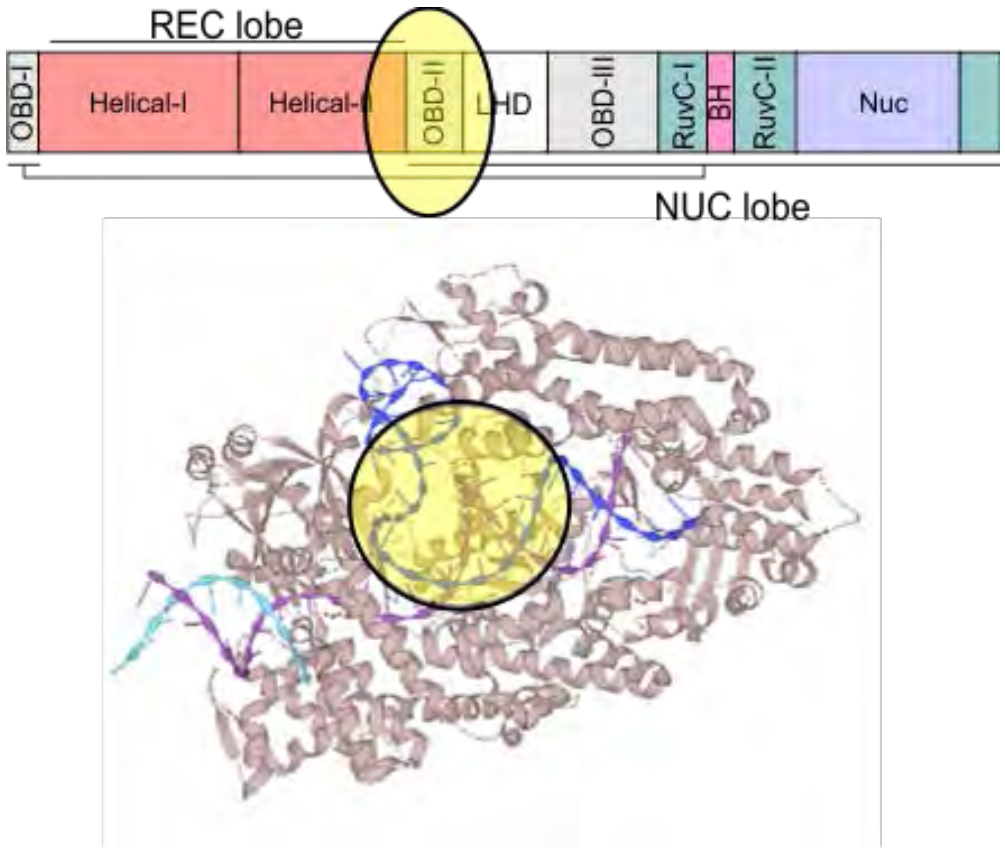


Figure 36. (Top) Domain architecture of Cas12. The yellow ellipse indicates the region of interest for the controlling the enzyme efficacy. (Bottom) Crystal structure of a Cas12a (pink) in complex with an RNA guide (blue), and the target DNA (violet and cyan). Again, the relevant sector is highlighted in yellow.

Project 3: "Quantitative biology and modelling"

C. Bosia (PI), E. Ferro, S. Tealdi, A. Pagnani, A. De Martino

Aims and Results

As quantitative biology and modeling Unit, we follow three parallel lines of research. The first one concerns the study of the interactions between different microRNA-mediated messenger RNAs (miRNAs), small non-coding RNA molecules with regulatory functions, and cellular noise, both intrinsic and extrinsic. The second one aims to establish links between cell physiology, cell population growth and intracellular resource partitioning strategies, transferring a quantitative approach developed in microbiology to cancer biology. The third line of research addresses vesicular trafficking, i.e. the biological process by which different vesicles transport material between the various cellular compartments and between the cell and the surrounding environment. For all projects, discussed below, the cellular and molecular biology experiments are supported by mathematical models. The experimental part involves measurements of cytofluorimetry and microscopy, both fluorescence and confocal. These projects are complemented by various collaborations with other experimental groups.

Subproject "Role of noise in the miRNA-target interaction"

Bimodal distributions in gene expression data are observed in different situations: from cancer cells to stem cells. The two peaks of the distribution are usually associated with different physiological states of the system. From a theoretical point of view, a common belief is that bimodality is linked to bistability, or to the presence of different stationary states in the absence of noise. However, in some biological systems, bimodality may be due to pure stochastic effects. Environmental fluctuations, henceforth called extrinsic noise, can be a source of noise in molecular networks. Together with the intrinsic fluctuations due to the probabilistic nature of chemical reactions, extrinsic noise shapes gene expression and can lead to cellular differentiation. Both theoretical and *in vitro* studies have hypothesized that miRNAs, in particular stoichiometric conditions, can induce bimodality in the expression of their targets simply because of stochastic effects linked to their peculiar titrative interaction and not due to bistable systems. The project concerns the study of the role of extrinsic noise in miRNA-mediated bimodal gene expression, both *in silico* (Gillespie simulations) and *in vitro* (transfections of synthetic constructs in immortalized cells), and with stochastic models of gene regulation developed by our unit.

In collaboration with Prof. Salil Garg (Koch Institute for Integrative Cancer Medicine at MIT, Cambridge, US), we showed how noise at the level of microRNAs pools can induce transitions

between different pre-differentiation states in embryonic stem cells. The study of recent literature on this topic led to the presentation of a review.

The temporal aspect related to these types of interactions is of increasing interest. In collaboration with Prof. Alejandra Ventura (Universidad de Buenos Aires and CONICET) we are developing the theoretical framework necessary for the interpretation of data produced by the unit with the support of an optogenetic platform. The expression levels of an miRNA of interest, responsive to light pulses of a particular wavelength, will be varied by a time-varying light pulse. We will then study the temporal responses of the targets of this miRNA and the physiological repercussions on the cells under investigation in response to different input signals.

With the help of analytical and numerical predictions we have shown how extrinsic noise in miRNA expression can induce bimodal distributions in targets (**Figure 37**) defining a wider bimodality region, in terms of interaction strength miRNA target and transcription rate, compared to the case of simple intrinsic noise (**Figure 38**).

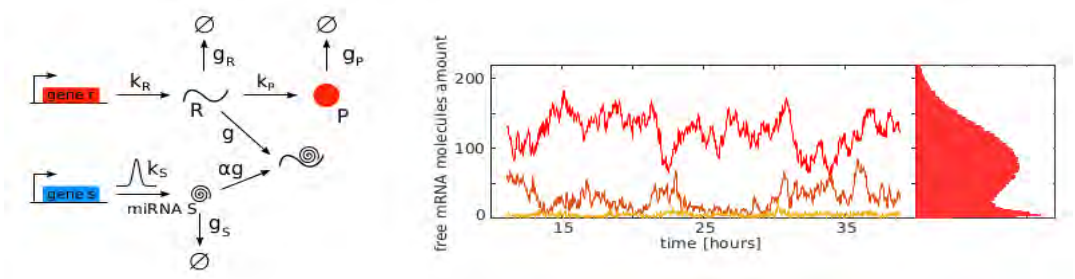


Figure 37. (Left Panel) The reference circuits (with and without extrinsic noise on the miRNA transcription rate) are represented. k_R and k_S are the target mRNA and miRNA transcription rates, g_R and g_S are respectively the mRNA and miRNA degradation rates. k_P is the protein translation rate and g_P is its degradation rate. g is the miRNA-target interaction strength and α is the fraction of miRNAs that are not recycled after binding to the mRNA. (Right Panel) trajectories for the mRNA, corresponding to the model on the left. The steady-state distributions of the number of free mRNA molecules are bimodal.

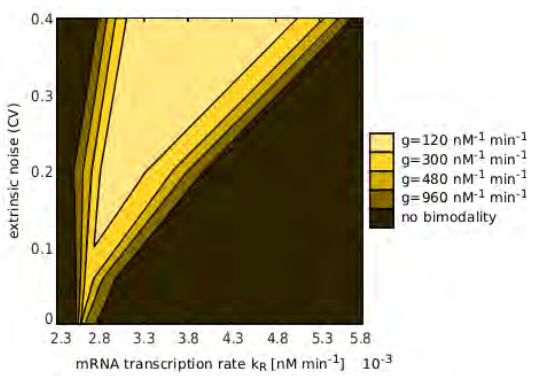


Figure 38. Phase diagram for bimodality in the free mRNA molecules distribution. On the *x* axis there is the target transcription rate k_R, on the *y* axis the extrinsic noise on the miRNA transcription rate. The color map indicates the presence of bimodality for different values of the miRNA-target interaction strength g. The presence of bimodality was computed by running several Gillespie's simulations for fixed sets of parameters and sampling targets' probability distributions.

Subproject "Quantitative approach to the physiology of cancer"

The objective of the project is to transfer to the cancer biology a quantitative approach recently developed in microbiology, where quantitative empirical relationships between cell composition and the rate of population growth have emerged. These relationships can be unified in a phenomenological theory of cell growth that assumes a simple three-component partitioning of the proteome: a fundamental part of the proteome, always expressed in all growth conditions, a part linked to ribosomes and a third sector (which includes metabolic proteins) which changes in a coordinated manner with the ribosomal content. A fundamental result of these studies is the discovery that global changes in the transcriptional program can be explained and predicted quantitatively starting from these physiological laws. The working hypothesis underlying the project is that the growth physiology of tumor cells can be controlled by simple quantitative rules dictated by the internal allocation of resources. Following the growth of tumor cells under different conditions, we were able to quantify the dependence of the population growth rate on the initial cell density and to define the optimal experimental conditions for rapidly adaptable populations (**Figures 39, 40**). This first result is preliminary to the definition of growth laws. In this case, support for cell biology is supported by mathematical models. The research continues in close collaboration with the group of Dr. Francesca Ceroni (Department of Chemical Engineering, Imperial College London).

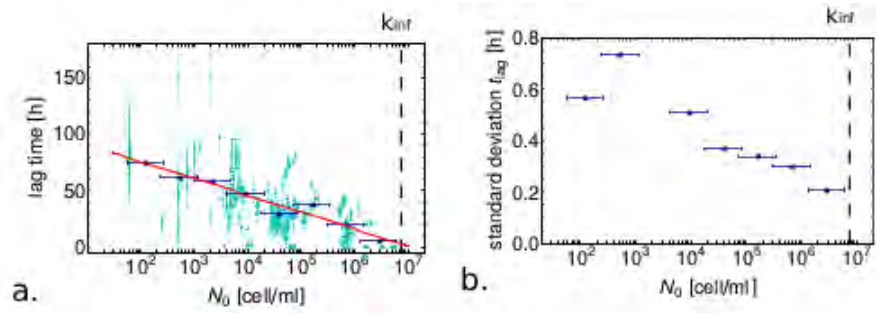


Figure 39. Panel a: Lag time as a function of inoculum size. Panel b: Estimated empirical fluctuations (standard deviation) of the lag time as a function of the inoculum size.

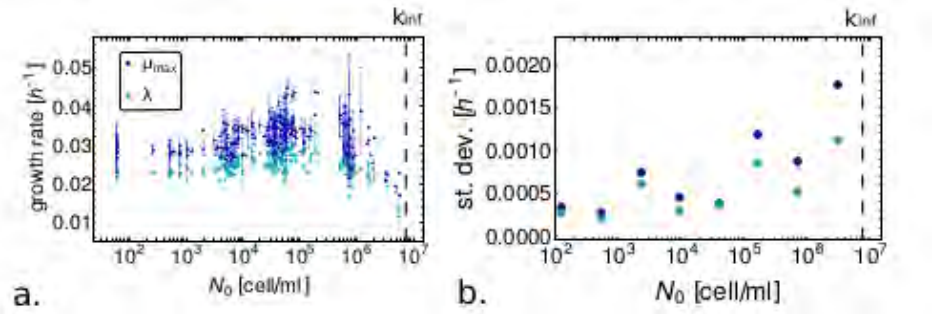


Figure 40. a. Estimated growth rate as a function of the inoculum size. **b.** Estimated fluctuation (standard deviation) of the growth rate as a function of the inoculum size.

Subproject "Intracellular Vesicular Traffic".

Vesicular trafficking is the biological process by which different vesicles transport material between cellular compartments and between a cell and its environment. These vesicles are characterized by different combinations of proteins that belong to the RAB family and whose spatial and temporal organization determines the efficiency of intracellular logistics. The project, in collaboration with the groups of Dr. Carlo Cosimo Campa (IIGM, Unit of Morphogenesis, Engineering and Targeting of Organelles) and Prof. Emilio Hirsch (Department of Molecular Biotechnology, University of Turin), concerns the study of the formation of membrane domains with a specific biochemical composition in early endosomes. The study has so far combined time-lapse microscopy experiments, quantitative image analysis and mathematical modeling. We are currently investigating the role of competition to define dynamic molecular patterns on endosomal membranes.

At the end of 2020 we concluded the project in collaboration with the group of Prof. Guido Serini (IRCCS and Department of Oncology, University of Turin). The collaboration focused on the characterization of macroscopic properties (such as size, velocity and direction of movement) of early and late endosomes under conditions of constitutive expression or overexpression of different proteins related to vesicular movement. Our work showed how different molecular motor systems drive differential transport and subcellular distribution of endosomes in mammalian cells. Again, the study included time-lapse microscopy experiments and quantitative image analysis.

Collaborations

International collaboration in the framework of INFERNET (RISE H2020 initiative)

- Prof. R. Mulet (University of Havana, Cuba)
- Prof. M. Reale (UHGS, Buenos-Aires, Argentina)
- Prof. L. Romanelli (UHGS, Buenos-Aires, Argentina)
- Prof. A. Fendrik (UHGS, Buenos-Aires, Argentina)
- Dr. J. De Cossio-Diaz (Centro de immunologia molecular, Cuba)

Funds and Grants

- INFERNET “New algorithms for inference and optimization of large scale biological data” European H2020 RISE initiative financed by the European commission (Coordinator and PI: A. Pagnani).

Genomic Instability and Tumor Immunity Unit

Research Group

Ilio Vitale, Ph.D, head of Unit (IIGM)

Ginevra Campia, student (UniTo)

Andrea Guarracino, PhD student (IIGM and Univ. "Tor Vergata")

Gwenola Manic, PhD, post-doctoral fellow (IIGM)

Luca Mattiello, PhD, post-doctoral fellow (IIGM)

Sara Soliman Abdel Rehim, PhD. student (IIGM and Univ. "Tor Vergata")

The main interest of the Unit is to unveil the link between cancer stemness, genetic instability and tumor immunity with the objective of developing anticancer strategies of precision immunotherapy.

We focus on three major topics:

1. the mechanisms for the preservation of genomic stability in cancer cells and cancer stem cells (CSCs);
2. the impact of genomic instability and cancer stemness on the immunogenic potential of tumors and their response to (immuno)therapy;
3. CSCs immunogenicity and immune escape.

Our hypothesis is that genetic instability and epigenetic deregulation can increase the visibility of CSCs to the immune system, making tumors more susceptible to immunosurveillance by cytotoxic T cells.

The identification of the molecular mechanisms regulating genomic stability and governing tumor immunity can guide the design of novel immunotherapeutic regimens.

Projects

Project: "Genomic instability and tumor immunity"

Subproject "Genomic instability and immunogenicity of CSCs"

I. Vitale (PI), G. Manic, A. Guarracino, L. Mattiello, S. Soliman

Aims

1. To investigate the immunogenic potential and the mechanisms regulating genomic stability in CSCs with distinct ploidy
2. To assess the mutanome, transcriptome, secretome and signalome of CSCs with different ploidy
3. To develop novel strategies of precision immunotherapy directed against CSCs

Results

In 2020, we characterized the immunogenic profile of diploid and tetraploid CSC clones generated from a panel of colorectal cancer patient-derived CSCs. Taking advantage of dedicated microfluidic devices in which peripheral blood mononuclear cells (PBMCs) are confronted with one single CSC clone (classic devices) or, simultaneously, with two clones with different ploidy (devices in

competition), we demonstrated that tetraploid CSC clones have an increased capability to chemoattract T cells, as compared to diploid CSC clones (**Figure 41**).

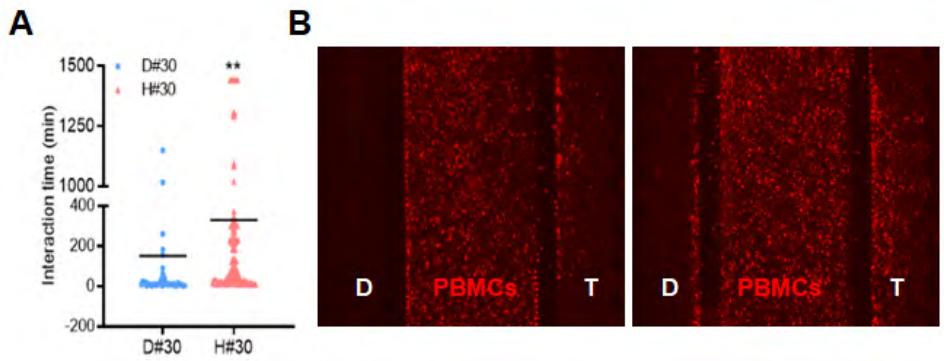


Figure 41. Assessment of the *in vitro* immunogenicity ot diploid (D) and tetraploid (H or T) clones through classic microfluidic devices (**A**) or microfluidic devices in competition (**B**). PBMCs: peripheral blood mononuclear cells.

In following *in vitro* experiments, we confirmed the increased immunogenic potential of tetraploid CSC clones by assessing the proliferation of T cells (**Figure 42 A**), the level of secreted interferon gamma (IFNγ) by T cells (**Figure 42 B**), and the phagocytosis potential of dendritic cells (**Figure 42 C**) upon co-cultivation of these immune cells with diploid or tetraploid CSC clones.

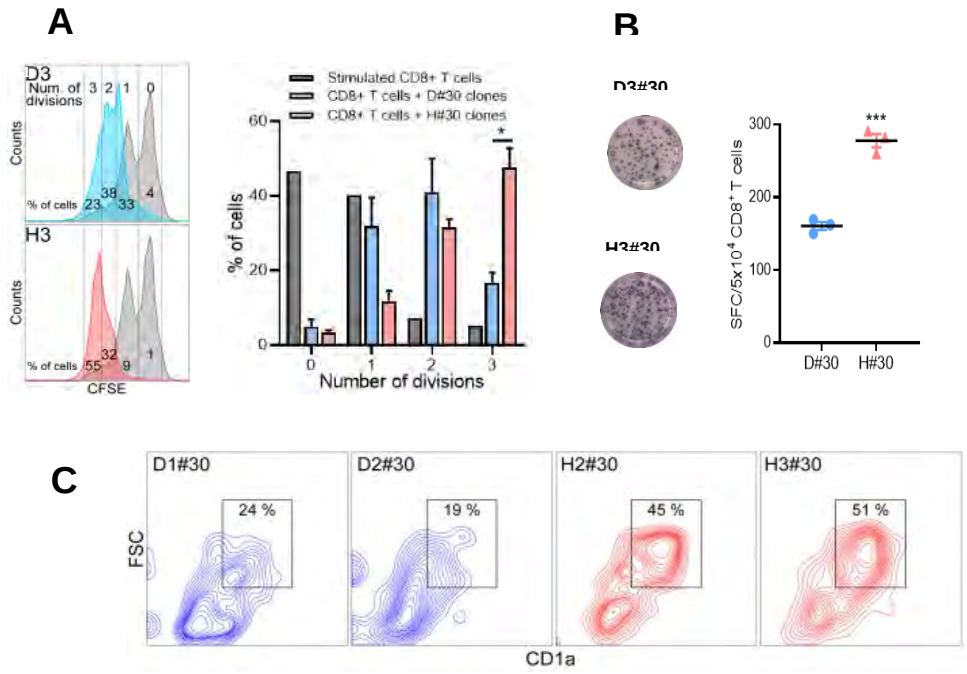


Figure 42. Evaluation of the *in vitro* immunogenicity ot diploid (D) and tetraploid (H) clones through: **A.** T cell proliferation assays; **B.** ELISpot; **C.** dendritic cell phagocytosis

To evaluate CSC immunogenicity *in vivo*, we used a humanized animal model based on the intraperitoneal injection of PBMCs in NSG immunodeficient mice. In these experiments, we provided evidence that tetraploid CSC clones grow less efficiently than diploid CSC clones when xenotransplanted into the flank of humanized mice (**Figure 43 A**). In following *ex vivo* analyses, we demonstrated that the infiltrate of tumors generated by tetraploid CSC clones xenotransplanted in humanized mice presents a higher percentage of $CD45^+$ cells and in particular $CD3^+CD4^+$ e $CD3^+CD8^+$ than that of diploid tumors generated by CSC clones, and this was accompanied by a higher fraction of functional and exhausted immune cells (**Figure 43 B**).

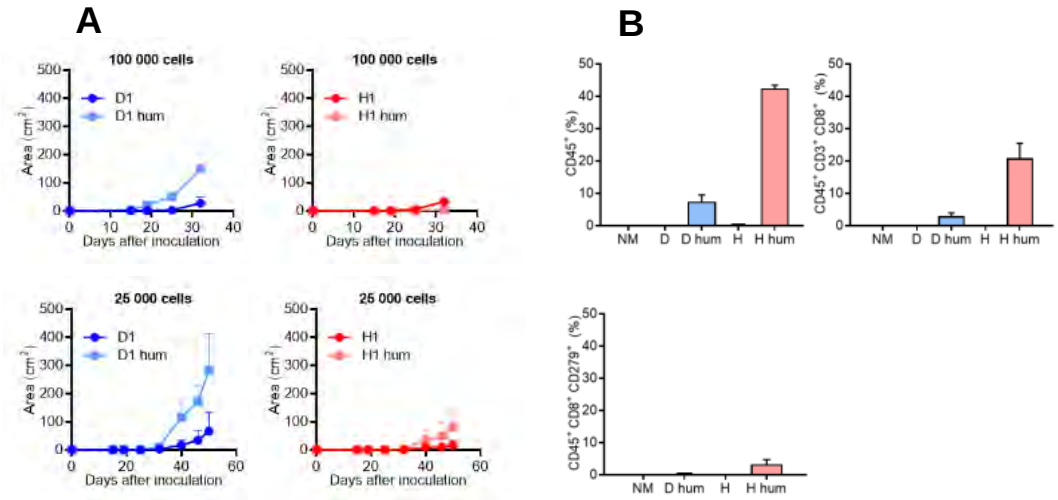


Figure 43. Evaluation of the *in vivo* immunogenic potential ot diploid (D) and tetraploid (H) clones using humanized mice models (**A**, **B**).

We then started an extensive characterization of the immunogenic profile of CRC-SC through omics studies including WES, RNASeq, RPPA and secretome in order (*i*) to characterize the mutational, transcriptomic and antigenic landscape of diploid and tetraploid clones and (*ii*) to identify the signaling pathways specifically deregulated or activated in CSCs with distinct ploidy.

As a first step, we inferred the HLA genotype of the three parental CRC-SC lines and the pseudo-tetraploid and diploid CRC-SCs clones through RNASeq and WES (**Table 1**).

# CRC-SC line	HLA-A,B,C haplotype
# 17	A*03:01 A*11:01 B*13:02 B*35:01 C*06:02 C*04:01
# 30	A*02:01 A*02:01 B*13:02 B*13:02 C*06:02 C*06:02
# 32	A*24:02 A*26:01 B*35:02 B*38:01 C*04:01 C*12:03

Table 1. HLA-A,B,C Haplotype of CRC-SC lines #17, #30 and #32

We then characterized the CRC-SCs secretome through the Proteome Profiler Array, an assay that allows quantification of cytokines and chemokines in cellular supernatants. In these assays, we observed that at least half of the pseudo-tetraploid clones exhibit an increase in two factors: (1) insulin like growth factor binding protein 2 (IGFBP2), a protein localized in the extracellular matrix that has immunomodulatory roles and is investigated as a tumor-associated antigens, and (2) C-X-C Motif Chemokine Ligand 8 (CXCL-8), a chemokine involved in the release of damage-associated molecular patterns (DAMPs) during immunogenic cell death, conferring immunogenicity to dying tumor cell The relevance of these targets to the immune response directed against tetraploid cells is being investigated. Additional targets will be identified by genomic, transcriptomic, and proteomic analysis.s (**Figure 44).**

The relevance of these targets to the immune response directed against tetraploid cells is being investigated. Additional targets will be identified by genomic, transcriptomic, and proteomic analysis.

Figure 44. Analysis of CSCs secretome. *In vitro* culture supernatants of pseudo-tetraploid (H#30) or pseudo-diploid (D#30) CRC-SC clones were recovered and subjected to Proteome Profiler Array. Quantification of pixels corresponding to the spot intensity of all cytokines and chemokines was performed with Image-J software.

Collaborations

- Dr. A. Sistigu (Università Cattolica del Sacro Cuore, Rome, Italy)
- Dr. M. Cereda (IIGM, Candiolo Research Institute, Candiolo, Italy)
- Dr. C. Bonini (IRCCS Ospedale San Raffaele, Milan, Italy)
- Dr. L. Galluzzi (Weill Cornell Medical College, New York, NY, USA)
- Dr. G. Kroemer (Centre de Recherche des Cordeliers, INSERM U 1138, Paris, France)
- Dott. I. Caruana (Ospedale Pediatrico Bambino Gesù, Rome, Italy).

Subproject "DNA replication and chromosome segregation as targets for the design of novel antitumor immunotherapies against cancer"

I. Vitale (PI), G. Manic, L. Mattiello, S. Soliman, A. Guarracino

Aims

1. To investigate the response to replication and mitotic stress in cancer cells and CSCs
2. To identify the intracellular and extracellular signals linked to the response to replication and mitotic stress in cancer cells and CSCs
3. To develop novel strategies of precision immunotherapy directed against CSCs

Results

The response to replication stress as a target for eradicating CSCs

We extensively investigated the DNA replication process in a large panel of CSCs derived from colorectal cancer patients, showing that colorectal CSCs have a heterogeneous level of replication stress (RS) despite showing a common proficiency in the response to RS. As a consequence, colorectal CSCs display distinct sensitivity to the abrogation of the ATR-CHK1 axis - the cascade of the DNA damage response specifically governing the cellular response to RS - and to RS inducers that are conventionally used in the clinics to treat colorectal cancer patients, including fluorouracil (5-FU), oxaliplatin ed irinotecan.

We have also unveiled a non-genetic mechanism of intrinsic and acquired resistance of CSCs to these therapeutic agents based on the increase in the levels of PARP1, a factor involved in the preservation of genomic stability and in the DNA damage response (**Figure 45**).

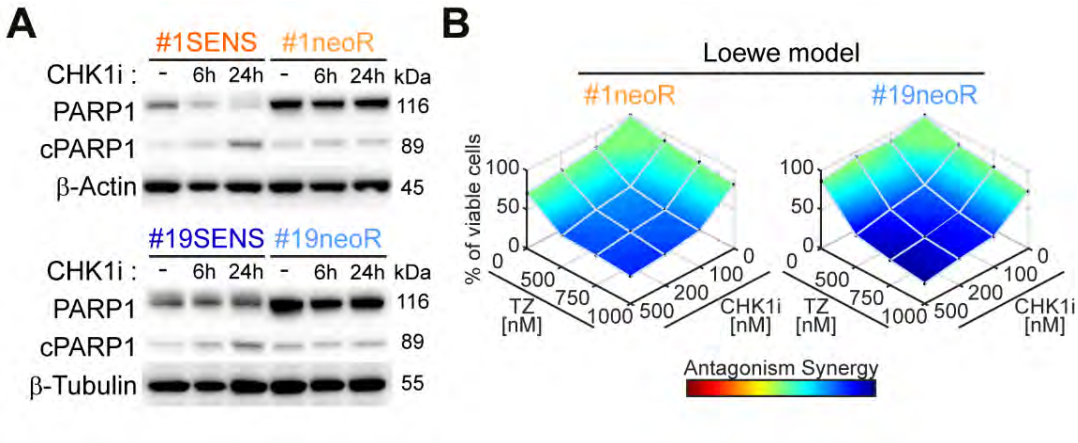


Figure 45. A. Western blot analysis of the expression levels of PARP1 in CSCs sensitive (SENS) o resistant (neoR) to CHK1 inhibitors. **B.** Assessment of the synergy between PARP1 (talazoparib, TZ) and CHK1 (CHK1i) inhibitors in killing CSCs.

In following experiments, we found that pharmacological inhibitors of PARP1 sensitize colorectal CSCs to the inhibition of the ATR-CHK1 axis, as demonstrated by the decrease in cell vitality and in clonogenic potential of CSCs exposed to CHK1 inhibitors in combination with PARP1 inhibitors (**Figure 45 A**). Such sensitization has been confirmed *in vivo* by assessing tumor growth of CSCs xenotransplanted into the flank of immunodeficient NSG mice upon treatment of mice with PARP1 and CHK1 inhibitors in monotherapy or combined therapy (**Figure 45 B**).

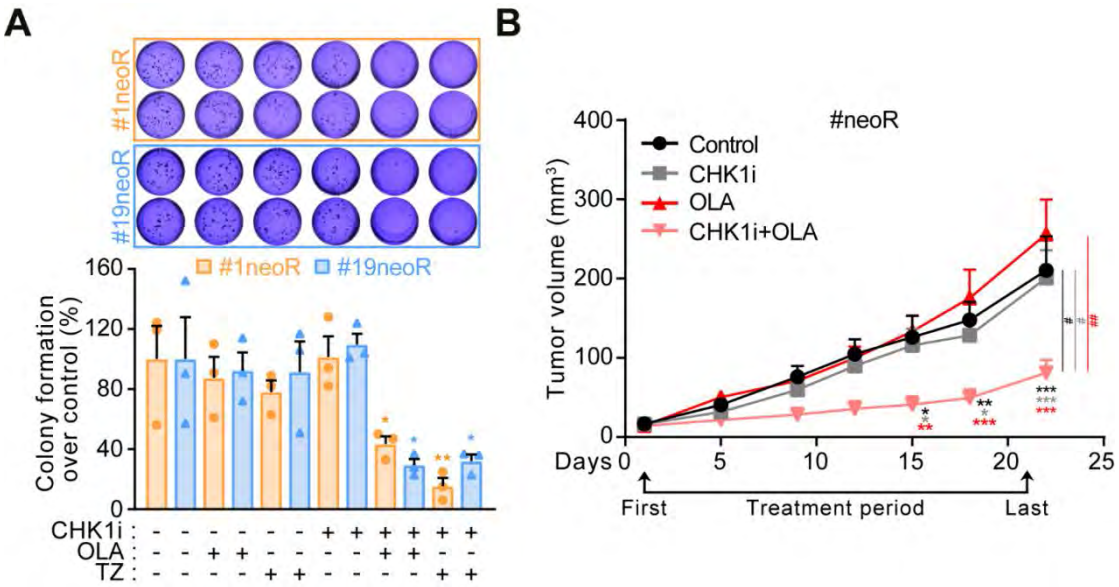


Figure 45. A. Evaluation of the clonogenic potential of CSCs resistant to the inhibition of CHK1 (neoR). **B.** Analysis of tumor growth of neoR CSCs xenotransplanted into the flank immunodeficient NSG mice () upon treatment with inhibitors of PARP1 (olaparib, OLA; talazoparib, TZ) and/or CHK1 (CHKi).

To characterize the functional impact of this increase, we analyzed basal levels of replicative stress in CRC-SCs observing that cells resistant to CHK1 inhibition and overexpressing PARP1 (neoR) have lower basal replicative stress, assessed by specific markers such as phosphorylation (p) of RPA32 and the presence of single-stranded (ss) DNA (**Figure 46 A**). Furthermore, in these cells, treatment with CHK1 inhibitors does not cause the increase in replicative stress, as demonstrated by the lack of increase in the markers pRPA32, pATM, pATR, and γH2AX compared with cells sensitive to CHK1 inhibition (SENS) (**Figures 46 A, B**).

Experiments aimed at analyzing the role of PARP1 in decreasing replicative stress and the potential contribution of other DNA damage response players such as MRE11 and RAD51 (which in preliminary analyses appear to cooperate for CSC survival) are ongoing.

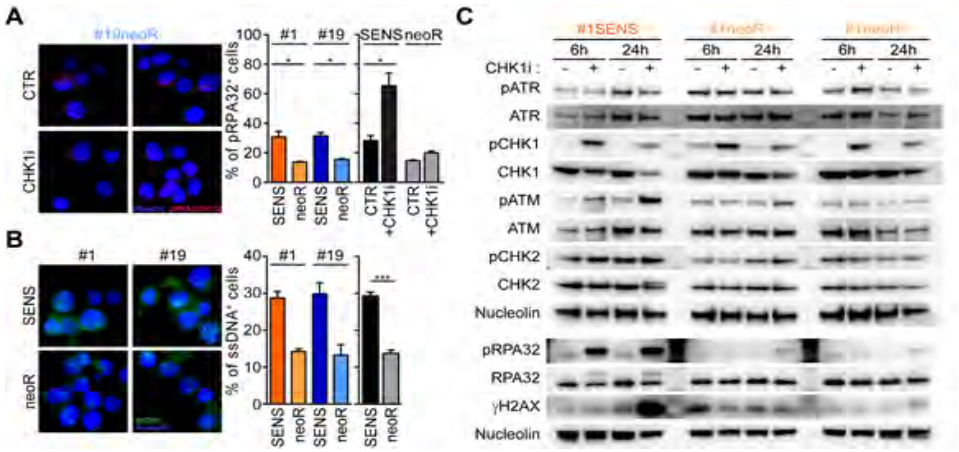


Figure 46. Basal levels of replication stress in CRC-SCs. Immunofluorescence (**A, B**) or western-blot (**C**) analysis of CRC-SCs sensitive (SENS) or resistant (neoR) to CHK1 inhibition after labeling with antibodies directed against the phosphorylated form of RPA32 (pRPA32) (A), anti-BrdU (B) or the phosphorylated form (p) and total ATM, ATR, CHK1, CHK2, RPA32 or γH2AX (C). In **C**, nucleolin was used to control charge.

The spindle assembly checkpoint as a target for eradicating CSCs

This subproject is aimed at characterizing mitosis regulation in CSCs and, specifically, is aimed at evaluating how and whether the deregulation of the "spindle assembly checkpoint" (SAC, the checkpoint monitoring the correct segregation of sister chromatids during mitosis) affects genomic stability, therapeutic response, immunogenic potential, and aggressiveness of CSCs. To this aim, we have developed a novel method to determine the activity of the SAC in a large number of CSCs and cancer cells. This method is based on the SAC score, a novel parameter we introduced that is calculated by analyzing the cellular response to antimitotic therapies through specific markers of cell divisions, and considering as endpoints: mitotic arrest, "mitotic slippage" and tetraploidization. Using the SAC score we could demonstrate that colorectal CSCs have a heterogenous SAC functionality (**Figure 47**). Importantly, we provided evidence that a considerable fraction of colorectal CSCs displays a weak SAC, as shown by low SAC score values.

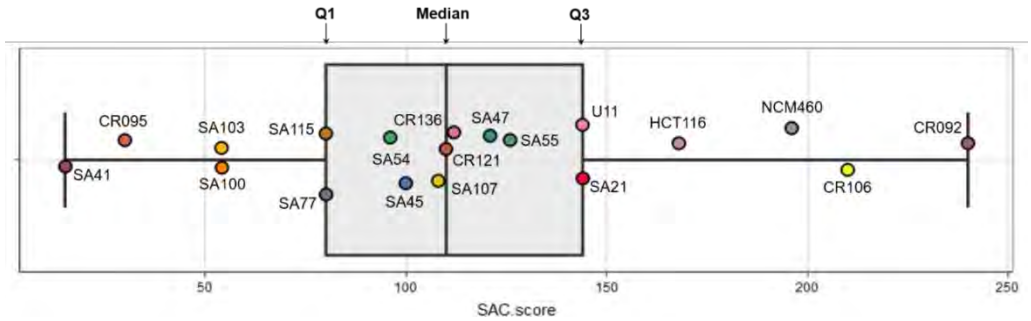


Figure 47. Determination of the activity of the SAC in colorectal CSCs, cancer cells (HCT116) and normal cells (NCM640) using the "SAC score", a novel parameter calculated using specific cell division markers.

We then launched a series of experiments aimed (*i*) at analyzing the impact of SAC deregulation on chromosome stability, therapeutic response of CSCs, and (*ii*) at identifying the molecular pathways involved in, and the signals emitted by cancer cells and CSCs in response to, SAC deregulation and mitotic stress.

To accurately quantify this decrease in SAC strength, we compared CRC-SCs exhibiting a weakened SAC with a SAC functionality value calculated by referring to the pool of cells exhibiting an extremely strong and efficient SAC. Through this analysis, we demonstrated that SAC remains partially functional even in CRC-SC #3, which has lower levels of SAC score and exhibits a decrease in SAC of approximately 40% (**Figure 48**). Experiments aimed at analyzing the impact of SAC deregulation in the chromosomal stability and response to therapy of CSCs are ongoing.

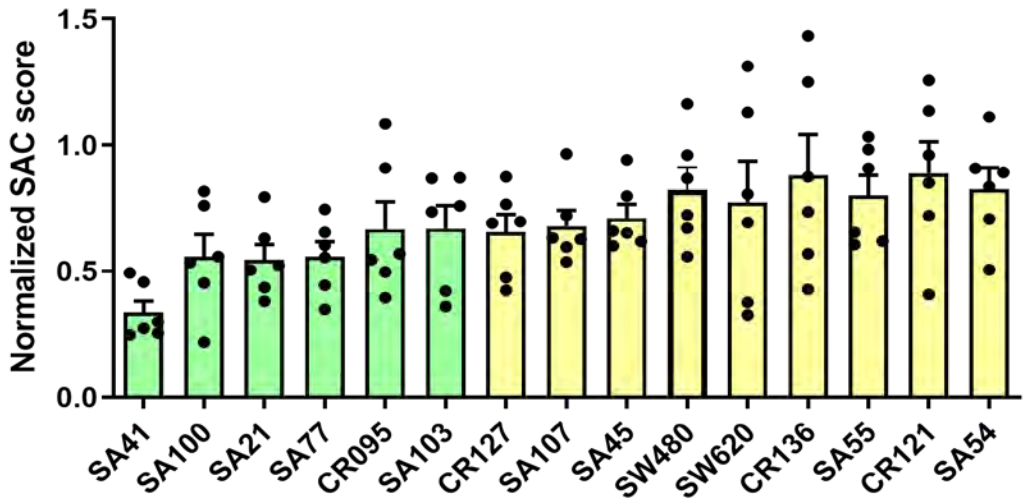


Figure 48. Percentage of SAC weakening in CRC-SCs. The cells shown in the figure and which we previously demonstrated to have a weakened SAC were compared to a panel of tumor cells and CRC-SCs that exhibit a very robust SAC, as assessed by SAC score calculation (see above).

Collaborations

- Prof. R. De Maria (Università Cattolica del Sacro Cuore, Rome, Italy)
- Dr. A. Sistigu (Università Cattolica del Sacro Cuore, Rome, Italy)
- Dr.L. Trusolino (Candiolo Research Institute, Candiolo, Italy)
- Dr.M. Cereda (IIGM, Candiolo Research Institute, Candiolo, Italy)
- Dr.L. Galluzzi (Weill Cornell Medical College, New York, NY, USA)
- Dr.O. Keep (Centre de Recherche des Cordeliers, INSERM U 1138, Paris, France)

Subproject "Cancer stemness and immunogenic therapies"

I. Vitale (PI), G. Manic, A. Guarracino, L. Mattiello, S. Soliman

Aims

1. To investigate the impact of agents inducing immunogenic cell dearth (ICD) on the generation and survival of CSCs
2. To identify the signaling pathways and epigenetic factors involved in the induction of CSCs upon therapies with ICD inducers
3. To develop dedicated therapies to sensitize CSCs to ICD inducers

Results

Our preliminary evidence indicates that CSCs display an immunoprivileged nature and a unique resistance to antitumor immunotherapies. To analyze the impact of ICD inducers on the appearance of CSCs, we performed qRT-PCR assessing the expression levels of several reprogramming factors in murine cancer cells treated with ICD inducer (doxorubicin and oxaliplatin) or non-ICD inducers (cisplatin). We observed that ICD inducers favor a complete transcriptional rewiring toward pluripotency in cancer cells by enhancing the expression of the entire panel of reprogramming factors analyzed (**Figure 49 A**). On the contrary, the non-ICD drug cisplatin affected the expression of only few transcriptional factors.

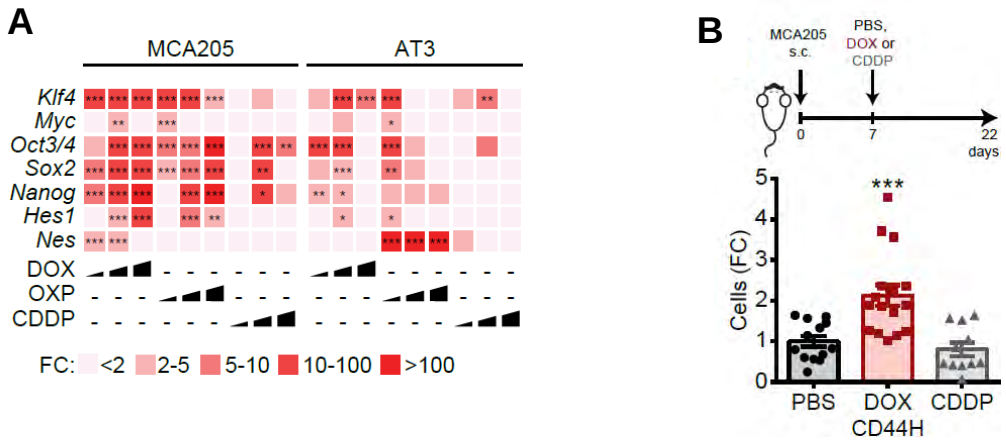


Figure 49. A. Quantification by qRT-PCR of the expression levels of the illustrated reprogramming factors in MCA205 mouse fibrosarcoma and AT3 mouse breast carcinoma cells left untreated or administered with doxorubicin (DOX: 0.25, 2.5, 25 μM), oxaliplatin (OXP: 3, 30, 300 μM), or cisplatin (CDDP: 1.5, 15, 150 μM). Data are mean fold change (FC) over untreated condition after intrasample normalization to the expression levels of Ppia. **B.** *Ex vivo* flow cytometric analysis of CSC surface markers expression in MCA205 cells grown in C57BL/6J mice treated intratumorally with vehicle (PBS) or 2.9 mg/kg DOX or 2.5 mg/kg CDDP. Data are mean FC±SEM over PBS treatment.

To explore the *in vivo* selection and appearance of CSCs upon ICD inducer-based therapy, we locally treated MCA205 mouse fibrosarcoma growing in syngeneic immunocompetent mice with doxorubicin or cisplatin, and evaluated CSC enrichment in recollected xenografts 15 days post-treatment (i.e., when tumors start escaping growth control). We found a twofold increase of CD44H cells (which represent the fraction of CSCs) upon treatment with doxorubicin but not cisplatin (**Figure 49 B**).

Next steps are (*i*) characterizing the drug-response and the tumorigenic, invasive and immunogenic potential of CSCs induced by ICD-based therapies and (*ii*) investigating the mechanism of chromatin remodeling of cancer cells subjected to therapies inducing ICD.

We started an extensive characterization of ICD-induced CSCs (ICD-CSCs) by assessing parameters such as sensitivity to therapy and self-renewal capacity. To this end, we isolated via FACS two fractions of ICD-CSCs (CD44H and CD44L cells) and subsequently administered ICD-inducing agents such as doxorubicin or oxaliplatin to these cells. In this analysis, we observed that CD44H cells (but not CD44L) had increased resistance to ICD-inducing chemotherapy compared with their parental counterparts (**Figure 50 A**). The same results were obtained in vivo after transplantation of ICD-CSCs into immunocompetent mice and treatment with ICD inducers (**Figure 50 B**). Finally, we evaluated the in vitro evolution of ICD-CSCs, demonstrating that both subsets are able to regenerate the phenotypic complexity of the parental cells (**Figure 50 C**)

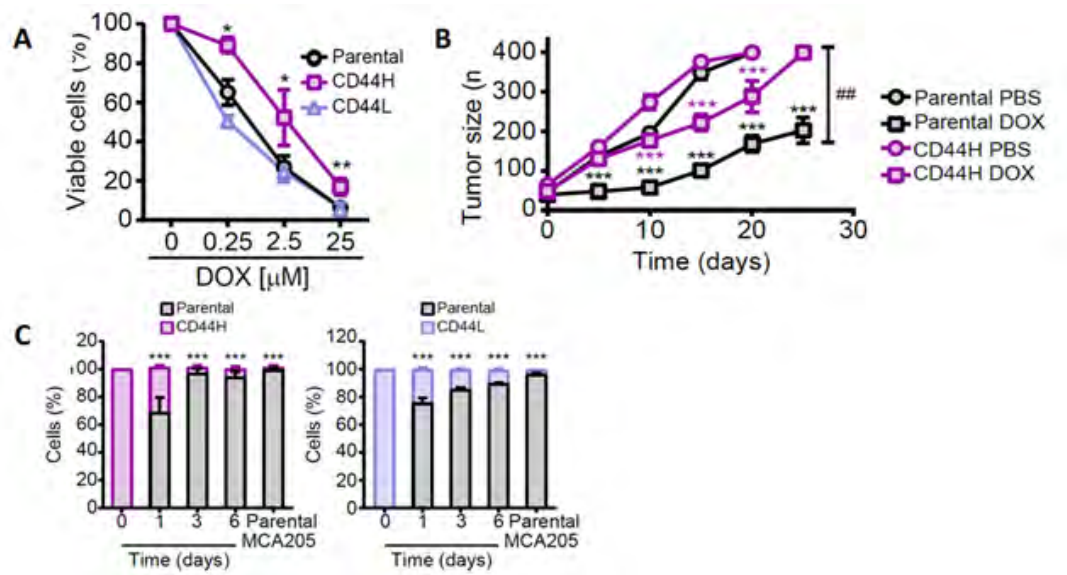


Figure 50. Characterization of CSCs selected by immunogenic cell death inducing agents (ICD-CSCs). **A.** Evaluation of proliferation and viability by CellTiter-Glo® assay. B. *in vivo* tumor growth after xenograft of immunodeficient mice C. regeneration capacity by cytofluorimetry analysis of ICD-CSC fractions isolated via FACS (CD44H and CD44L) and parental cells after treatment with doxorubicin (B) or oxaliplatin (C).

Experiments aimed at analyzing the tumorigenic and metastatic potential of CSCs and their ability to immune escape along with experiments aimed at elucidating the mechanisms of tumor cell remodeling undergoing ICD inducer-based therapy are ongoing.

Collaborations

- Dr. A. Sistigu (Università Cattolica del Sacro Cuore, Rome, Italy)
- Dr. L. Galluzzi (Weill Cornell Medical College, New York, NY, USA)
- Dr. G. Kroemer (Centre de Recherche des Cordeliers, INSERM U 1138, Paris, France)

Funds and Grants

- Compagnia di San Paolo (PI: I. Vitale)
- AIRC Individual Grant – IG 20417 (PI: I. Vitale)
- AIRC, Start-Up 2016 #18418 (Partner: I. Vitale)

Organelle Morphogenesis, Engineering and Targeting Unit

Research Group

Carlo Cosimo Campa, head of Unit (IIGM)

Giampaolo Placidi, PhD student (PoliTo and IIGM)

Simone Tealdi, PhD student (PoliTo and IIGM)

The "Organelle Morphogenesis, Engineering and Targeting" unit develops new technologies and new strategies for the enhancement of drug therapies.

Organelles are cellular structures dedicated to the performance of specific tasks within the cell. Similar to the organs that make up the human body, each organelle is home to a limited number of biological processes that ultimately determine its function. Each organelle is distinguished by a specific set of molecules that, similar to an identity card, allows its correct recognition.

Thanks to this notion, it was discovered that within the cell there are different organelles with different and divergent functions. Some of these are responsible for the synthesis of molecules, others provide energy to the cellular apparatus and others behave as real centers of storage and destruction of molecules. Exploiting the high level of specialization of organelles to enhance drug therapies is the object of our research.

Our program has three main objectives: a) to develop and apply new methods for the analysis of morphological and molecular processes that characterize organelles, b) to design and build organelles with new and increased functionalities, c) to explore new approaches to target drugs to specific organelles, thus improving their therapeutic efficacy.

The activities of the research unit include the use of the most advanced genomic engineering techniques in order to characterize the molecular interactions that confer greater efficacy to drug treatment. In parallel, protein engineering methodologies are applied to the development of new technologies in order to monitor and modify the biochemical composition of organelles.

Finally, various fluorescence microscopy techniques are used to monitor, at high spatio-temporal resolution, the impact of drug treatment on the morphology and biochemical composition of organelles.

Projects

Project 1: "Analysis of Rab molecular switches during drug treatment of brain tumors"

Subproject "Analysis of the dynamics of Rab molecular switches on endosomal membranes"

C.C. Campa (PI)

Aims

1. To analyze the biochemical activation kinetics of molecular switches on endosomes.

Results

The activity of the Unit has focused on the analysis of the dynamics of molecular switches belonging to the Rab family of proteins. Rab are a family of proteins that control the rate of degradation of molecules involved in the growth and proliferation of cancer cells. Rab proteins act on the surface of endosomes, specific cellular organelles that are responsible for storing cellular material pending their elimination (**Figure 51**).

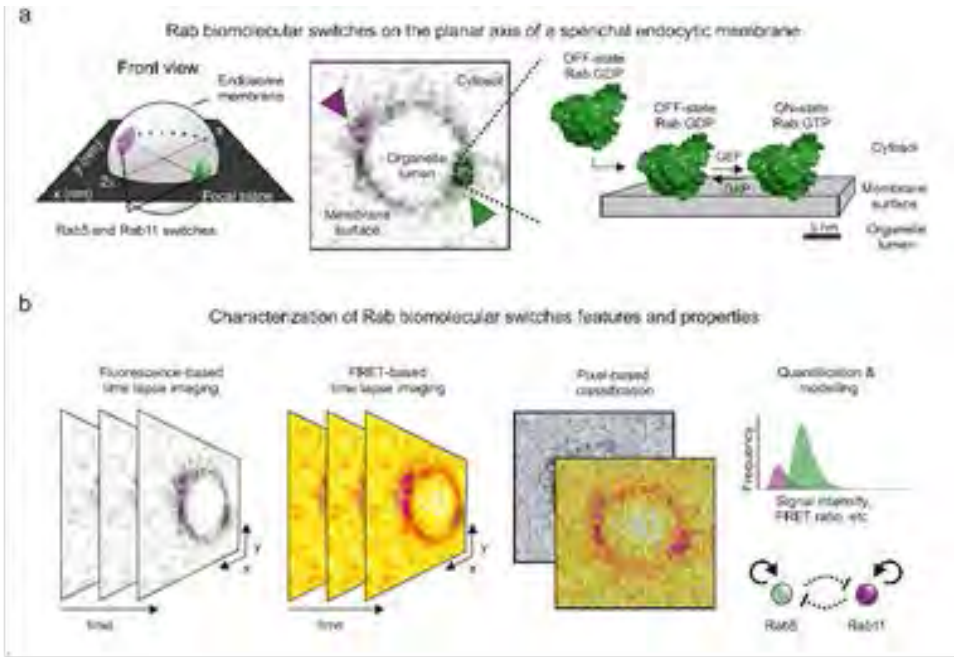


Figure 51. A. Illustration of the localization and activation mechanism of Rab proteins on the endosome. **B.** Aanalysis procedure to separate the localization signal from the FRET activation signal.

The storage process requires the compartmentalization, on the surface of the endosome, of Rab5 and Rab11 proteins. Understanding how these two Rab proteins organize themselves is therefore of

fundamental importance to improve the ability of each cell to destroy the main molecules involved in tumor proliferation. At present, there are no methods able to quantitatively analyze this process. Therefore, in collaboration with the Statistical Inference and Quantitative Biology Unit of IIGM, we have developed a method to monitor both the spatial segregation of Rab proteins and their biochemical activity.

Using this approach, we observed that Rab5 and Rab11 molecules are distributed in clusters on the surface of endosomes and show a well-defined spatiotemporal pattern. We found that genetic modification induced by overexpression of ZFYVE26, a gene responsible for processing cellular growth factors, significantly alters both the spatial and temporal organization of Rab proteins, creating areas of extensive overlap (**Figure 52**).

This work defines a mechanism by which alterations in ZFYVE26 contribute to cell proliferation.

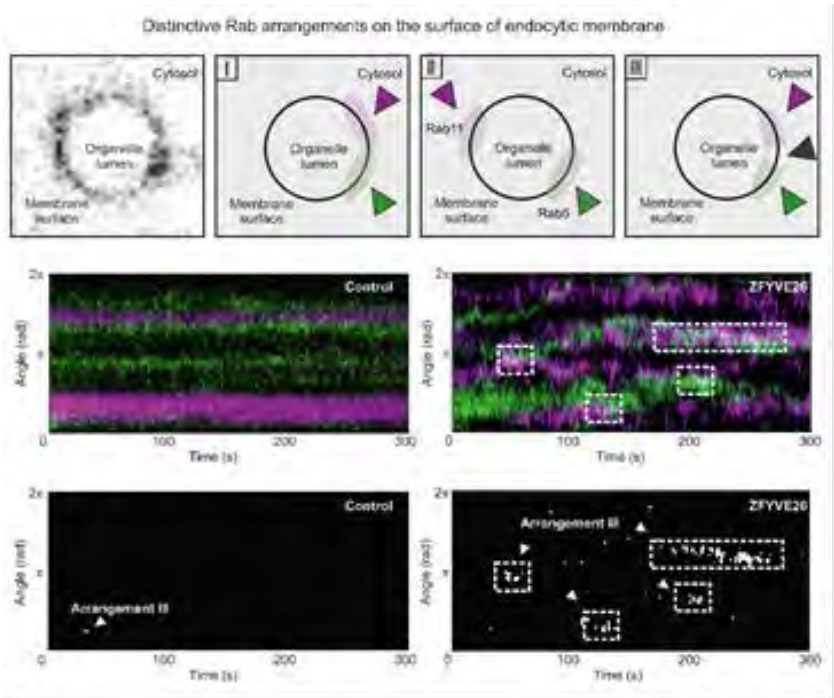


Figure 52. Spatiotemporal distribution of Rab5 (green) and Rab11 (purple) molecules observed on the surface of endosomes under conditions of normal expression or overexpression of the ZFYVE26 gene.

Collaborations

- Dr. C. Bosia and Dr. E. Ferro (Unità di Inferenza statistica e Biologia Computazionale, IIGM, Candiolo (To), Italy).

Subproject "Modeling of factors controlling the dynamism of Rab molecular switches on endosomal membranes"

C.C. Campa (PI)

Aims

1. To develop a computational model that predicts, from a set of images, the spatial and temporal dynamics of the small GTPases that make up the membrane of endosomes.endosomes.

Results

In order to give a broader meaning to our experimental observations, the group started a collaboration with the Statistical Inference and Quantitative Biology unit of IIGM to build a computational model that predicts the behavior of the molecules that constitute the surface of the endosome from a set of microscope images.

Thus, we constructed a computational model in which proteins of the small GTPase family, such as Rab5 and Rab11, and one of their effectors, the ZFYVE26 protein, are interdependent (**Figure 53**). This mutual dependence allowed us to obtain a more realistic computational representation of the biological process, in which the relative dosage between the various proteins plays a major role.

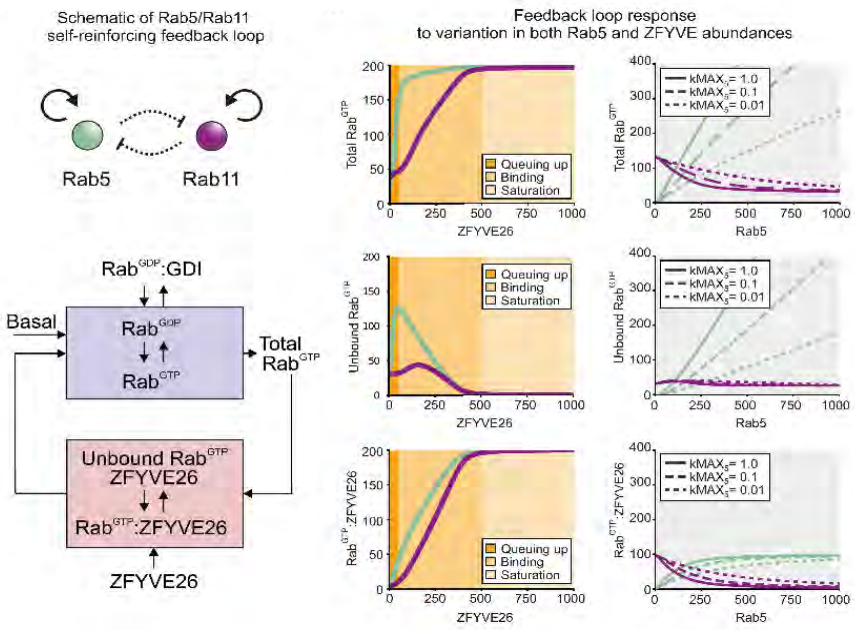


Figure 53. Schematic and related biochemical parameters on which the computational model is based.

Using this model we were able to predict the temporal profiles of the small GTPases Rab5 and Rab11 from single microscopy images.

This work defines new strategies useful for the design of synthetic systems in which molecular dynamics are controlled at the nanoscale, an essential feature to improve the efficacy of drugs once they enter the cell.

Collaborations

- Dr. C. Bosia and Dr. E. Ferro (Unità di Inferenza statistica e Biologia Computazionale, IIGM, Candiolo (To), Italy).

Subproject "Design of a high-throughput screening platform for the identification of inhibitors of the small GTPase family"

C.C. Campa (PI), G. Placidi

Aims

1. To design and validate a platform for the identification of inhibitors of cell signaling pathways downstream of small GTPases

Results

In order to translate the knowledge obtained towards new drug therapies, the group has initiated a project that aims to identify new inhibitors for small GTPases, a family of proteins that has been seen to be crucial in the onset of several genetic diseases.

The small GTPase family includes a group of genes with similar structure and sequence but different biological function ranging from proliferation to cell migration. These genes include Ras and Rac, known oncogenes, and the Rab proteins. The latter have been seen to be involved in numerous neurodegenerative diseases such as spastic paraplegias and Parkinson's disease.

Taking advantage of the numerous structural similarities between the members that compose the family of small GTPases, we have designed a sensor capable of identifying inhibitors for one of its best known members, the Rab11 protein (**Figure 54**).

We engineered a molecule belonging to the fatty acid binding protein family to emit green fluorescent light in response to Rab11 activation. Furthermore, we implemented our sensor with a feature that

allows it to generate light only upon binding to a specific effector, i.e. a molecule activated upon direct binding to Rab11 (**Figure 55**).

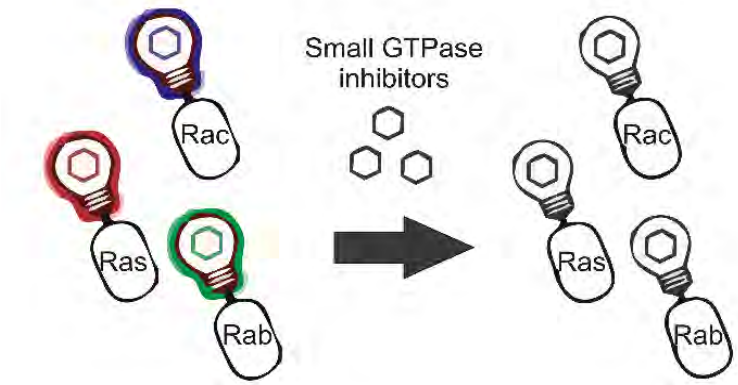


Figure 54. Schematic of the mechanism underlying the high-throughput screening platform for identifying inhibitors for the small GTPase family.

Figure 55. Cells in which the sensor for Rab11 was inserted. On the left are cells with the Rab11 protein active, on the right are cells in which the Rab11 protein has been turned off.

This design is highly innovative as it allows to identify new molecules capable of inhibiting specific cellular functions activated by small GTPases. The first prototypes analyzed showed a high sensitivity in discriminating between a complete or partial inhibition of the small GTPase Rab11

Subproject "Engineering platforms for massive disruption of gene networks"

C.C. Campa (PI), - Project in collaboration with the Statistical Inference and Computational Biology Unit (G. Uguzzoni, A.P. Muntoni, A. Pagnani)

Aims

1. To implement methods to improve the generation of cell lines with "ad hoc" mutations in order to reduce the time and cost of multi-gene editing of cellular genomes

Results

Recently, CRISPR/Cas genetic engineering technology has made it possible to unravel the role of many genes underlying human cellular function and disease. However, this technology has been ineffective under conditions where multiple genes are required to be edited simultaneously. As a result, the development of more powerful CRISPR/Cas systems will enable large-scale engineering of complex cellular programs.

Among the large number of recently discovered CRISPR/Cas genome editing systems, we selected the class II type V RNA-guided endonuclease named CRISPR-Cas12a. The CRISPR/ Cas12a system is a miniaturized CRISPR-based genome editing system that enables gene editing with a high degree of precision. Using a molecular modeling approach, we have identified a region in the middle of the Cas12a protein structure that is critical in DNA target binding.

This region comprises a portion of both the Recognition (REC) and Nuclease (NUC) lobes. These two portions of the Cas12a protein generate a binding pocket for both target DNA and guide RNAs necessary for the targeting of the Cas12a enzyme on genes. Taken together, these data identify the protein region on which to focus our efforts in order to generate a new, more efficient version of the Cas12a protein.

We intend to realize a technology that can be used for all those applications of industrial nature that require the rapid and efficient generation of a large number of cells in which the gene network has been massively redesigned. In this context we can cite as examples cells capable of recognizing and destroying human tumors and cells capable of producing molecules useful for therapeutic purposes.

Funds and Grants

- Compagnia di San Paolo
- Italian Association for Cancer Research (AIRC), Project "Mixing of Rab GTPase membrane domains reduces glioblastoma response to immunotherapeutic agents" (PI: C. C. Campa)

FACILITIES

Flow Cytometry and Cell Sorting Facility

Dr. Denis BAEV, PhD.

The mission of the IIGM's Flow cytometry and cells sorting Facility is to support the research projects by providing to IIGM, FPO-IRCCS researchers the access to modern multiparametric methods of rapid cell analysis and single cell manipulations.

The facility, run by Dr. D. Baev, was established in 2017 as a shared resource facility for the IIGM (former HuGeF) and UniTo MBC users. Originally the facility had one flow cytometer (BD FACS Canto, 2 laser- 6 FL channels) and the cells sorter (BD Aria III, 2 lasers, 6 FL channels). In 2018, thanks from the generous support from the Compagnia di San Paolo, the IIGM FCCSF acquired 2 analyzers, the BD FACS Celesta (3 lasers, 12 FL channels) and BD LSR Fortessa X20 (3 lasers, 14 FL channels, HTS option) and performed the complete upgrade of the existing Aria III sorter to 3 laser, 14 FL channel configurations, similar to the original optical layout of the X20 instrument.

In 2020 the BD Fortessa X20 instrument was upgraded to 4 laser 18 color configuration and the Beckman Coulter Cytoflex (2 lasers, 6FL channels, 96 well plate option) was adopted as an instrument for the entry-level analyses.

Currently, the IIGM FCCSF provides with the full spectra of flow cytometry and cell sorting services for the IIGM, FPO-IRCCS Candiolo and external customers. However, due to the COVID-19 –related restrictions the number of services provided to external users was limited.

The facility is well equipped with the following instruments:

Beckman Coulter CytoFLEX

Lasers and detectors	FS, SS, 488 -4FL, 633 -2FL
Maximum speed (events/second)	10000
Sensitivity	25 MESF
Resolution, nanometers	250
Flow rate	10-30-60ul/min (presets), up-to 200ul/min (custom)
Fluidics type	Peristaltic pump
HTS option	Yes, 96 well plate integrated loader
Software	CytExpert
Availability for unassisted usage	Yes, after training (IIGM/IRCC users only)

BD FACS Celesta BVR

Lasers and detectors	FS, SS, 405nm – 5FLs (BV421, BV510, BV605, BV650, BV786), 488nm – 4 FLs (FITC, PE, PE-TR, PerCP-Cy5.5), 633nm -3FLs (APC, AlexaFluor700, APC-Cy7)
Maximum speed (events/second)	25000
Sensitivity	25MESF (FITC), 15MESF (PE)
Resolution, nanometers	>250
Flow rate	10-30-60ul/min (presets), up-to 100ul/min (custom)
Fluidics type	Pressurized
HTS option	No
Software	BD DiVa 8.0.1.1 Celesta version
Availability for unassisted usage	Yes, after training (IIGM/IRCC users only)

BD Fortessa X-20

Lasers and detectors	FS, SS, 405nm – 6FLs (BV421, BV510, BV605, BV650, BV711, BV786), 488nm – 5 FLs (FITC, PE, PE-TR, PerCP-Cy5.5, PE-Cy7), 633nm - 3FLs (APC, AlexaFluor700, APC-Cy7), 355nm- 3/4FLs (BUV, flexible configuration)
Maximum speed (events/second)	50000
Sensitivity	30 MESF
Resolution, nanometers	>250
Flow rate	10-30-60ul/min (presets), up-to 200ul/min (custom)
Fluidics type	Pressurized, BD FACSFlow supply system
HTS option	BD HTS option 96/384 well plates
Software	BD DiVa 8.0.2
Availability for unassisted usage	Yes, after training (IIGM/IRCC users only)

BD FACS Aria III high-speed cell sorter with ACDU

Lasers and detectors	FS, SS, 405nm – 6FLs (BV421, BV510, BV605, BV650, BV711, BV786), 488nm – 5 FLs (FITC, PE, PE-TR, PerCP-Cy5.5, PE-Cy7), 633nm - 3FLs (APC, AlexaFluor700, APC-Cy7)
Maximum speed (events/second)	35000
Sensitivity	50 MESF
Sorting rate	Up to 30000/sec
Sorting options - tubes	2-way (15ml/5ml), 4-way (5ml/Eppendorfs)
Sorting options- plates	6/12/24/48/96/384 plates –continuous and single-cell sorting, Terumo plates, slides – single cell sorting
Nozzles	70, 85, 100, 130 micron
Sample handling	Rotating agitation, thermal control (from +4C to +37C)
Software	BD DiVa 8.0.2
Availability for unassisted usage	No, can be used by IIGM FCCSF staff only

IIGM's Genomic Platform

The IIGM's Genomic Platform was established in 2019 with the goal to provide IIGM researchers, the scientific communities of the FPO-IRCCS Candiolo Cancer Institute, of the University of Turin, and of other institutions in Italy and abroad, with the access to the most advanced technologies for genomic analysis, supporting researchers in their projects.

All instruments were acquired thanks to the contribution of Compagnia di San Paolo.

Dr. K. Gizzi, BSc.

Dr. C. Parlato, BSc.

Dr. S. Guarrera, MSc.

Genomic Facility: sequencing and microarray analyses

The activities of the Genomic Platform include consulting and analyses planning with the users, and laboratory analyses, and are offered as all-inclusive services or as scientific collaboration.The Facility is run by Dr. K. Gizzi and Dr. C. Parlato with the support of Dr. S. Guarrera, and is well equipped with a state-of-the-art NovaSeq6000 massive parallel sequencer (Illumina Inc) which allows a modular and scalable approach to sequencing with the availability of several sequencing formats, providing both the necessary high throughput processing for big-scale sequencing projects, as well as the flexibility to accommodate small and medium size projects at relatively moderate costs. A GridION (Oxford Nanopore Technologies) sequencer for proofreading long-reads sequencing is also available.

Being equipped with a high resolution iScan microarray scanner (Illumina), and with three automatic liquid handlers for the preparation of samples and microarrays, the Genomic Facility provides whole-genome genotyping analyses (microarray genotyping assays) and methylation analyses services (microarray methylation analysis) as well.

Dr. M. Cereda, PhD.

Bioinformatic support

The Genomic Facility is supported by the IIGM's Bioinformatics section, coordinated by Dr. M. Cereda, which provides support for primary data processing and data storage and management until delivery to the user.

Genomic Facility equipment

NovaSeq 6000 Sequencing System

High throughput sequencing

GridION | Oxford Nanopore Technologies

Proofreading long-reads sequencing

iScan microarray scanner

Freedom EVO liquid handlers

Metabolic Facility

Dr. B. Pardini, PhD.

The Metabolic Facility of IIGM was set up with the acquisition of a Seahorse XF Analyzer (Agilent) thanks to the kind support of Compagnia di San Paolo..

The instrument allows to measure oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) of live cells in a multi-well plate, interrogating key cellular functions such as mitochondrial respiration and glycolysis, thus allowing the investigation of several metabolic processes.

Dr. S. Guarrera, MSc.

The facility is supervised by Dr. B. Pardini with the support of Dr. S. Guarrera, and can be accessed in terms of:

- Scientific collaboration regimen: the facility is open to experienced researchers who wish to set up analyses by themselves, upon agreements with IIGM's facility personnel
- Service regimen: for non-experienced users who wish to take advantage of an all-inclusive service.

Seahorse XF Analyzer

PUBLICATIONS 2020

1. **Genetic variations in microRNA binding sites of solute carrier transporter genes as predictors of clinical outcome in colorectal cancer**

   Bendova P, **Pardini B**, Susova S, Rosendorf J, Levy M, Skrobanek P, Buchler T, Kral J, Liska V, Vodickova L, Landi S, Soucek P, **Naccarati A**, Vodicka P, Vymetalkova V.
   Carcinogenesis. epub Dec.2020, 2021 Apr 17;42(3):378-394 IF (2019): 4.603

2. **The Inhibitory Role of miR-486-5p on CSC Phenotype Has Diagnostic and Prognostic Potential in Colorectal Cancer**

   Pisano A, Griñan-Lison C, Farace C, Fiorito G, Fenu G, Jiménez G, Scognamillo F, Peña-Martin J, **Naccarati A**, Pröll J, Atzmüller S, **Pardini B**, Attene F, Ibba G, Solinas MG, Bernhard D, Marchal JA, Madeddu R.
   Cancers (Basel). 2020 Nov 19;12(11):3432. IF (2019): 6.126

3. **A Holistic Perspective: Exosomes Shuttle between Nerves and Immune Cells in the Tumor Microenvironment**

   Dragomir MP, Moisoiu V, Manaila R, **Pardini B**, Knutsen E, Anfossi S, Amit M, Calin GA.
   J Clin Med. 2020 Oct 31;9(11):3529. IF (2019): 3.303

4. **The Long Noncoding RNA CCAT2 Induces Chromosomal Instability Through BOP1-AURKB Signaling**

   Chen B, Dragomir MP, Fabris L, Bayraktar R, Knutsen E, Liu X, Tang C, Li Y, Shimura T, Ivkovic TC, Cruz De Los Santos M, Anfossi S, Shimizu M, Shah MY, Ling H, Shen P, Multani AS, **Pardini B**, Burks JK, Katayama H, Reineke LC, Huo L, Syed M, Song S, Ferracin M, Oki E, Fromm B, Ivan C, Bhuvaneshwar K, Gusev Y, Mimori K, Menter D, Sen S, Matsuyama T, Uetake H, Vasilescu C, Kopetz S, Parker-Thornburg J, Taguchi A, Hanash SM, Girnita L, Slaby O, Goel A, Varani G, Gagea M, Li C, Ajani JA, Calin GA
   Gastroenterology. 2020 Dec;159(6):2146-2162.e33 IF (2019): 17.373

5. **Stochastic Epigenetic Mutations Are Associated with Risk of Breast Cancer, Lung Cancer, and Mature B-cell Neoplasms**

   Gagliardi A, Dugué PA, Nøst TH, Southey MC, Buchanan DD, Schmidt DF, Makalic E, Hodge AM, English DR, Doo NW, Hopper JL, Severi G, Baglietto L, **Naccarati A**, **Tarallo S**, **Pace L**, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Lund E, Giles GG, **Pardini B**, Sandanger TM, Milne RL, **Vineis P**, Polidoro S, Fiorito G.
   Cancer Epidemiol Biomarkers Prev. 2020 Oct. IF (2019): 4.344

6. **Small Non-Coding RNA Profiling in Plasma Extracellular Vesicles of Bladder Cancer Patients by Next-Generation Sequencing: Expression Levels of miR-126-3p and piR-5936 Increase with Higher Histologic Grades**

Sabo AA, Birolo G, **Naccarati A**, Dragomir MP, Aneli S, Allione A, Oderda M, Allasia M, Gontero P, Sacerdote C, **Vineis P**, Matullo G, **Pardini B**.
Cancers (Basel). 2020 Jun 9;12(6):1507 IF (2019): 6.126

7. **Expression quantitative trait loci in ABC transporters are associated with survival in 5-FU treated colorectal cancer patients**

Vymetalkova V, Rosa F, Susova S, Bendova P, Levy M, Buchler T, Kral J, Bartu L, Vodickova L, Hughes DJ, Soucek P, **Naccarati A**, Kumar R, Vodicka P, **Pardini B**.
Mutagenesis. 2020 Jul 11;35(3):273-281. IF (2019): 3.379

8. **Exosomal microRNAs and other non-coding RNAs as colorectal cancer biomarkers: a review**

Francavilla A, Turoczi S, **Tarallo S**, Vodicka P, **Pardini B**, **Naccarati A.**
Mutagenesis. 2020 Jul 11;35(3):243-260 IF (2019): 3.379

9. **DNA repair and cancer in colon and rectum: Novel players in genetic susceptibility**

Pardini B, Corrado A, Paolicchi E, Cugliari G, Berndt SI, Bezieau S, Bien SA, Brenner H, Caan BJ, Campbell PT, Casey G, Chan AT, Chang-Claude J, Cotterchio M, Gala M, Gallinger SJ, Haile RW, Harrison TA, Hayes RB, Hoffmeister M, Hopper JL, Hsu L, Huyghe J, Jenkins MA, Le Marchand L, Lin Y, Lindor NM, Nan H, Newcomb PA, Ogino S, Potter JD, Schoen RE, Slattery ML, White E, Vodickova L, Vymetalkova V, Vodicka P, Gemignani F, Peters U, **Naccarati A.**
Int J Cancer. 2020 Jan 15;146(2):363-372. IF (2019): 5.145

10. **Short-term personal and outdoor exposure to ultrafine and fine particulate air pollution in association with blood pressure and lung function in healthy adults**

Van Nunen E, Hoek G, Tsai MY, Probst-Hensch N, Imboden M, Jeong A, **Naccarati A**, **Tarallo S**, Raffaele D, Nieuwenhuijsen M, Vlaanderen J, Gulliver J, Amaral AFS, **Vineis P**, Vermeulen R.
Environ Res. 2021 Mar;194:110579. IF (2019): 5.715

11. **The use of silicone wristbands to evaluate personal exposure to semi-volatile organic chemicals (SVOCs) in France and Italy**

Wang S, Romanak KA, **Tarallo S**, **Francavilla A**, Viviani M, **Vineis P**, Rothwell JA, Mancini FR, **Cordero F**, **Naccarati A**, Severi G, Venier M.
Environ Pollut. 2020 Dec. Environ Pollut. 2020 Dec;267:115490. IF (2019): 6.793

12. **Trial watch: the gut microbiota as a tool to boost the clinical efficacy of anticancer immunotherapy**

Daillère R, Derosa L, Bonvalet M, Segata N, Routy B, Gariboldi M, Budinská E, De Vries IJM, **Naccarati A**, Zitvogel V, Caldas C, Engstrand L, Loilbl S, Fieschi J, Heinzerling L, Kroemer G, Zitvogel L.
Oncoimmunology. 2020 Jun 3;9(1):1774298. IF (2019): 5.869

13. **DNA Mismatch Repair Gene Variants in Sporadic Solid Cancers**

Caja F, Vodickova L, Kral J, Vymetalkova V, **Naccarati A**, Vodicka P.
Int J Mol Sci. 2020 Aug 3;21(15):5561. IF (2019): 4.556

14. **microRNA expression profiles and personal monitoring of exposure to particulate matter**

Mancini FR, Laine JE, **Tarallo S**, Vlaanderen J, Vermeulen R, van Nunen E, Hoek G, Probst-Hensch N, Imboden M, Jeong A, Gulliver J, Chadeau-Hyam M, Nieuwenhuijsen M, de Kok TM, Piepers J, Krauskopf J, Kleinjans JCS, **Vineis P**, **Naccarati A**.
Environ Pollut. 2020 Aug;263(Pt B):114392. IF (2019): 6.793

15. **Epistatic effect of TLR3 and cGAS-STING-IKKε-TBK1-IFN signaling variants on colorectal cancer risk**

Catalano C, da Silva Filho MI, Frank C, Lu S, Jiraskova K, Vymetalkova V, Levy M, Liska V, Vycital O, **Naccarati A**, Vodickova L, Hemminki K, Vodicka P, Weber ANR, Försti A.
Cancer Med. 2020 Feb;9(4):1473-1484. IF (2019): 3.491

16. **Agnostic Cys34-albumin adductomics and DNA methylation: Implication of N-acetylcysteine in lung carcinogenesis years before diagnosis**

Dagnino S, Bodinier B, Grigoryan H, Rappaport SM, Karimi M, Guida F, Polidoro S, Edmands WB, **Naccarati A**, Fiorito G, Sacerdote C, Krogh V, Vermeulen R, **Vineis P,** Chadeau-Hyam M.
Int J Cancer. 2020 Jun 15;146(12):3294-3303. IF (2019): 5.145

17. **Consumption of Fish and Long-chain n-3 Polyunsaturated Fatty Acids Is Associated With Reduced Risk of Colorectal Cancer in a Large European Cohort**

Aglago EK, Huybrechts I, Murphy N, Casagrande C, Nicolas G, Pischon T, Fedirko V, Severi G, Boutron-Ruault MC, Fournier A, Katzke V, Kühn T, Olsen A, Tjønneland A, Dahm CC, Overvad K, Lasheras C, Agudo A, Sánchez MJ, Amiano P, Huerta JM, Ardanaz E, Perez-Cornago A, Trichopoulou A, Karakatsani A, Martimianaki G, Palli D, Pala V, Tumino R, **Naccarati A**, Panico S, Bueno-de-Mesquita B, May A, Derksen JWG, Hellstrand S, Ohlsson B, Wennberg M, Van Guelpen B, Skeie G, Brustad M, Weiderpass E, Cross AJ, Ward H, Riboli E, Norat T, Chajes V, Gunter MJ.
Clin Gastroenterol Hepatol. 2020 Mar;18(3):654-666.e6. IF (2019): 8.549

18. **Age-related DNA methylation changes are sex-specific: a comprehensive assessment**

Yusipov I, Bacalini MG, Kalyakulina A, Krivonosov M, Pirazzini C, Gensous N, Ravaioli F, Milazzo M, Giuliani C, Vedunova M, Fiorito G, **Gagliardi A**, Polidoro S, Garagnani P, Ivanchenko M, Franceschi C.
Aging (Albany NY). 2020 Dec 3;12(23):24057-24080. IF (2019): 4.831

19. **Comparison of smoking reduction with improvement of social conditions in early life: simulation in a British cohort**

Lepage B, Colineaux H, Kelly-Irving M, **Vineis P**, Delpierre C, Lang T.
Int J Epidemiol. epub Dec.2020, 2021 Jul 9;50(3):797-808 IF (2019): 7.707

20. **The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk**

Berger E, Maitre N, Romana Mancini F, Baglietto L, Perduca V, Colineaux H, Sieri S, Panico S, Sacerdote C, Tumino R, **Vineis P**, Boutron-Ruault MC, Severi G, Castagné R, Delpierre C.
BMC Cancer. 2020 Nov 23;20(1):1138. IF (2019): 3.150

21. **Patterning of educational attainment across inflammatory markers: Findings from a multi-cohort study**

Maurel M, Castagné R, Berger E, Bochud M, Chadeau-Hyam M, Fraga S, Gandini M, Hutri-Kähönen N, Jalkanen S, Kivimäki M, Marmot M, McCrory C, Preisig M, Raitakari O, Ricceri F, Salmi M, Steptoe A, **Vineis P**, Delpierre C, Kelly-Irving M.
Brain Behav Immun. 2020 Nov;90:303-310. IF (2019): 6.633

22. **A multi-omics approach to investigate the inflammatory response to life course socioeconomic position**

Castagné R, Kelly-Irving M, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Hebels DG, Kleinjans JC, de Kok TM, Georgiadis P, Kyrtopoulos SA, Vermeulen R, Stringhini S, **Vineis P**, Chadeau-Hyam M, Delpierre C.
Epigenomics. 2020 Aug;12(15):1287-1302. IF (2019): 4.112

23. **A multi-omic analysis of birthweight in newborn cord blood reveals new underlying mechanisms related to cholesterol metabolism**

Alfano R, Chadeau-Hyam M, Ghantous A, Keski-Rahkonen P, Chatzi L, Perez AE, Herceg Z, Kogevinas M, de Kok TM, Nawrot TS, Novoloaca A, Patel CJ, Pizzi C, Robinot N, Rusconi F, Scalbert A, Sunyer J, Vermeulen R, Vrijheid M, **Vineis P**, Robinson O, Plusquin M.
Metabolism. 2020 Sep;110:154292. IF (2019): 6.159

24. **Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study**

Freisling H, Viallon V, Lennon H, Bagnardi V, Ricci C, Butterworth AS, Sweeting M, Muller D, Romieu I, Bazelle P, Kvaskoff M, Arveux P, Severi G, Bamia C, Kühn T, Kaaks R, Bergmann M, Boeing H, Tjønneland A, Olsen A, Overvad K, Dahm CC, Menéndez V, Agudo A, Sánchez MJ, Amiano P, Santiuste C, Gurrea AB, Tong TYN, Schmidt JA, Tzoulaki I, Tsilidis KK, Ward H, Palli D, Agnoli C, Tumino R, Ricceri F, Panico S, Picavet HSJ, Bakker M, Monninkhof E, Nilsson P, Manjer J, Rolandsson O, Thysell E, Weiderpass E, Jenab M, Riboli E, **Vineis P**, Danesh J, Wareham NJ, Gunter MJ, Ferrari P.
BMC Med. 2020 Jan 10;18(1):5. IF (2019): 6.782

25. **Determinants of accelerated metabolomic and epigenetic aging in a UK cohort**

Robinson O, Chadeau Hyam M, Karaman I, Climaco Pinto R, Ala-Korpela M, Handakas E, Fiorito G, Gao H, Heard A, Jarvelin MR, Lewis M, Pazoki R, Polidoro S, Tzoulaki I, Wielscher M, Elliott P, **Vineis P**.
Aging Cell. 2020 Jun;19(6):e13149. IF (2019): 7.238

26. **Epigenetic Clocks and Allostatic Load Reveal Potential Sex-Specific Drivers of Biological Aging**

McCrory C, Fiorito G, McLoughlin S, Polidoro S, Cheallaigh CN, Bourke N, Karisola P, Alenius H, **Vineis P**, Layte R, Kenny RA.
J Gerontol A Biol Sci Med Sci. 2020 Feb 14;75(3):495-503. IF (2019): 5.236

27. **A nutrient-wide association study for risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition and the Netherlands Cohort Study**

Papadimitriou N, Muller D, van den Brandt PA, Geybels M, Patel CJ, Gunter MJ, Lopez DS, Key TJ, Perez-Cornago A, Ferrari P, **Vineis P**, Weiderpass E, Boeing H, Agudo A, Sánchez MJ, Overvad K, Kühn T, Fortner RT, Palli D, Drake I, Bjartell A, Santiuste C, Bueno-de-Mesquita BH, Krogh V, Tjønneland A, Lauritzen DF, Gurrea AB, Quirós JR, Stattin P, Trichopoulou A, Martimianaki G, Karakatsani A, Thysell E, Johansson I, Ricceri F, Tumino R, Larrañaga N, Khaw KT, Riboli E, Tzoulaki I, Tsilidis KK
Eur J Nutr. 2020 Oct;59(7):2929-2937. IF (2019): 4.664

28. **GrimAge outperforms other epigenetic clocks in the prediction of age-related clinical phenotypes and all-cause mortality**

McCrory C, Fiorito G, Hernandez B, Polidoro S, O'Halloran AM, Hever A, Ni Cheallaigh C, Lu AT, Horvath S, **Vineis P**, Kenny RA.
J Gerontol A Biol Sci Med Sci. epub Nov 2020, 2021 Apr 30;76(5):741-749. IF (2019): 5.236

29. **Genome-Wide DNA Methylation in Peripheral Blood and Long-Term Exposure to Source-Specific Transportation Noise and Air Pollution: The SAPALDIA Study**

Eze IC, Jeong A, Schaffner E, Rezwan FI, Ghantous A, Foraster M, Vienneau D, Kronenberg F, Herceg Z, **Vineis P**, Brink M, Wunderli JM, Schindler C, Cajochen C, Röösli M, Holloway JW, Imboden M, Probst-Hensch N.
Environ Health Perspect. 2020 Jun;128(6):67003 IF (2019): 8.382

30. **Modulation of circRNA Metabolism by m(6)A Modification**

Di Timoteo G, Dattilo D, Centrón-Broco A, Colantoni A, Guarnacci M, Rossi F, Incarnato D, **Oliviero S**, Fatica A, Morlando M, Bozzoni I.
Cell Rep. 2020 May 12;31(6):107641 IF (2019): 8.109

31. **The transcriptional regulator ZNF398 mediates pluripotency and epithelial character downstream of TGF-beta in human PSCs**

Zorzan I, Pellegrini M, Arboit M, Incarnato D, **Maldotti M**, Forcato M, Tagliazucchi GM, Carbognin E, Montagner M, **Oliviero S**, Martello G.
Nat Commun. 2020 May 12;11(1):2364 IF (2019): 12.121

32. **NR2F1 regulates regional progenitor dynamics in the mouse neocortex and cortical gyrification in BBSOAS patients**

Bertacchi M, Romano AL, Loubat A, Tran Mau-Them F, Willems M, Faivre L, Khau van Kien P, Perrin L, Devillard F, Sorlin A, Kuentz P, Philippe C, Garde A, Neri F, Di Giaimo R, **Oliviero S**, Cappello S, D'Incerti L, Frassoni C, Studer M.
EMBO J. 2020 Jul 1;39(13):e10416 IF (2019): 9.889

33. **Identification of altered biological processes in heterogeneous RNA-sequencing data by discretization of expression profiles**

Lauria A, **Peirone S**, **Del Giudice M**, **Priante F**, Rajan P, Caselle M, **Oliviero S**, **Cereda M**.
Nucleic Acids Res. 2020 Feb 28;48(4):1730-1747 IF (2019): 11.501

34. **Omic approaches to pediatric bone sarcomas**

Tirtei E, **Cereda M**, De Luna E, Quarello P, Asaftei SD, Fagioli F.
Pediatr Blood Cancer. 2020 Feb;67(2):e28072 IF (2019): 2.355

35. **The immune system view of the coronavirus SARS-CoV-2**

Celardo I, **Pace L**, Cifaldi L, Gaudio C, Barnaba V.
Biol Direct. 2020 Dec 29;15(1):30 IF (2019): 2.193

36. **Type I interferons induce peripheral T regulatory cell differentiation under tolerogenic conditions**

Vitale S, **Russo V**, Dettori B, Palombi C, **Baev D**, Proietti E, Le Bon A, Belardelli F, **Pace L**.
Int Immunol. 2020 Aug 25;dxaa058 IF (2019): 3.519

37. **Epigenetics of T cell fate decision**

Pace L, Amigorena S.
Curr Opin Immunol. 2020 Apr;63:43-50 IF (2019): 7.290

38. **The persistence of stemness**

Petiti L, **Pace L**.
Nat Immunol. 2020 May;21(5):492-494 IF (2019): 20.479

39. **GATA Factor-Mediated Gene Regulation in Human Erythropoiesis**

Romano O, **Petiti L**, Felix T, Meneghini V, Portafax M, Antoniani C, Amendola M, Bicciato S, Peano C, Miccio A.
iScience. 2020 Apr 24;23(4):101018 IF (2019): 4.447

40. **Identification of Targets to Redirect CAR T Cells in Glioblastoma and Colorectal Cancer: An Arduous Venture**

Ponterio E, **De Maria R**, **Haas TL.**
Front Immunol. 2020 Sep 30;11:565631 IF (2019): 6.429

41. **Distinct retrograde microtubule motor sets drive early and late endosome transport**

Villari G, **Enrico Bena C**, **Del Giudice M**, Sandri C, Gioelli N, Camillo C, Fiorio-Pla A, **Bosia C**, Serini G.
EMBO J. 2020 Dec 15;39(24):e103661 IF (2019): 9.960

42. **MicroRNAs organize intrinsic variation into stem cell states**

Chakraborty M, Hu S, Visness E**, Del Giudice M**, **De Martino A**, **Bosia C**, Sharp PA, Garg S.
Proc Natl Acad Sci U S A. 2020 Mar 24;117(12):6942-6950 IF (2019): 9.412

43. **MicroRNA-mediated noise processing in cells: a fight or a game?**

Ferro E, **Enrico Bena C**, Grigolon S, **Bosia C**.
Comput Struct Biotechnol J. 2020 Mar 10;18:642-649 IF (2019): 6.018

44. **Aligning biological sequences by exploiting residue conservation and coevolution**

Muntoni AP, **Pagnani A**, Weigt M, Zamponi F.
Phys Rev E. 2020 Dec;102(6-1):062409. IF (2019): 2.300

45. **Clostridium cellulovorans metabolism of cellulose as studied by comparative proteomic approach**

Usai G, Cirrincione S, Re A, Manfredi M, **Pagnani A**, Pessione E, Mazzoli R.
J Proteomics. 2020 Mar 30;216:103667 IF (2019): 3.530

46. **Consensus guidelines for the definition, detection and interpretation of immunogenic cell death**

Galluzzi L, **Vitale I**, Warren S, Adjemian S, Agostinis P, Martinez AB, *et al*.
J Immunother Cancer. 2020 Mar;8(1):e000337 IF (2019): 8.660

47. **Immunological impact of cell death signaling driven by radiation on the tumor microenvironment**

Rodriguez-Ruiz ME, **Vitale I**, Harrington KJ, Melero I, Galluzzi L.
Nat Immunol. 2020 Feb;21(2):120-134 IF (2019): 20.479

48. **Cytofluorometric assessment of dendritic cell- mediated uptake of cancer cell apoptotic bodies**

Musella M, **Manic G**, Galassi C, **Vitale I**, Sistigu A.
Methods Enzymol. 2020;632:39-54. IF (2019): 1.394

49. **Assessment of IFN-γ and granzyme-B production by in "sitro" technology**

Galassi C, **Manic G**, Musella M, Sistigu A, **Vitale I**.
Methods Enzymol. 2020. IF (2019): 1.394

50. **Multicohort and cross-platform validation of a prognostic Wnt signature in colorectal cancer**

Goeman F, De Nicola F, Amoreo CA, Scalera S, Marinelli D, Sperati F, Mazzotta M, Terrenato I, Pallocca M, Ciuffreda L, Sperandio E, Barba M, Pizzuti L, Sergi D, Amodio A, Paoletti G, Krasniqi E, Vici P, Casini B, Gallo E, Buglioni S, Diodoro MG, Pescarmona E, **Vitale I**, **De Maria R**, Grazi GL, Ciliberto G, Fanciulli M, Maugeri-Saccà M.
Clin Transl Med. 2020 Dec;10(8):e199 IF (2019): 7.919

51. **Tuning Cancer Fate: Tumor Microenvironment's Role in Cancer Stem Cell Quiescence and Reawakening**

Sistigu A, Musella M, Galassi C, **Vitale I**, **De Maria R**.
Front Immunol. 2020 Oct 21;11:2166 IF (2019): 6.429

52. **Caspase 2 and p53 Reunited in Tumor Control**

Petroni G, **Vitale I**, Galluzzi L.
Trends Cell Biol. 2020 Dec;30(12):917-918. IF (2019): 16.041

53. **Relapses and treatment-related events contributed equally to poor prognosis in children with ABL-class fusion positive B-cell acute lymphoblastic leukemia treated according to AIEOP-BFM protocols**

Cario G, Leoni V, Conter V, Attarbaschi A, Zaliova M, Sramkova L, Cazzaniga G, Fazio G, Sutton R, Elitzur S, Izraeli S, Lauten M, Locatelli F, **Basso G**, Buldini B, Bergmann AK, Lentes J, Steinemann D, Göhring G, Schlegelberger B, Haas OA, Schewe D, Buchmann S, Moericke A, White D, Revesz T, Stanulla M, Mann G, Bodmer N, Arad-Cohen N, Zuna J, Valsecchi MG, Zimmermann M, Schrappe M, Biondi A.
Haematologica. 2020 Jul;105(7):1887-1894. IF (2019): 7.116

54. **Randomized post-induction and delayed intensification therapy in high-risk pediatric acute lymphoblastic leukemia: long-term results of the international AIEOP-BFM ALL 2000 trial**

Attarbaschi A, Mann G, Zimmermann M, Bader P, Barisone E, **Basso G**, Biondi A, Cario G, Cazzaniga G, Colombini A, Flotho C, Kuhlen M, Lang P, Lauten M, Linderkamp C, Locatelli F, Lo Nigro L, Möricke A, Niggli F, Panzer-Grümayer R, Parasole R, Peters C, Caterina Putti M, Rizzari C, Suttorp M, Valsecchi MG, Conter V, Schrappe M; on behalf of the AIEOP-BFM (Associazione

Italiana di Ematologia e Oncologia Pediatrica & Berlin-Frankfurt-Münster) Study Group.
Leukemia. 2020 Jun;34(6):1694-1700. IF (2019): 8.665

55. Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation

Fang C, Wang Z, Han C, Safgren SL, Helmin KA, Adelman ER, Serafin V, **Basso G**, Eagen KP, Gaspar- Maia A, Figueroa ME, Singer BD, Ratan A, Ntziachristos P, Zang C.
Genome Biol. 2020 Sep 15;21(1):247. IF (2019): 10.806

56. Prognostic value of minimal residual disease measured by flow-cytometry in two cohorts of infants with acute lymphoblastic leukemia treated according to either MLL-Baby or Interfant protocols

Popov A, Buldini B, De Lorenzo P, Disarò S, Verzhbitskaya T, Movchan L, Giarin E, Shorikov E, Di Meglio A, Tsaur G, Parasole R, Miakova N, Boichenko E, Kondratchik K, Aleinikova O, Karachunskiy A, Roumiantsev A, Locatelli F, Biondi A, Pieters R, Valsecchi MG, Fechina L, **Basso G**.
Leukemia. 2020 Nov;34(11):3042-3046. IF (2019): 8.665

PUBLICATIONS 2021

1. **Metabolic perturbations prior to hepatocellular carcinoma diagnosis: Findings from a prospective observational cohort study**

   Stepien M, Keski-Rahkonen P, Kiss A, Robinot N, Duarte-Salles T, Murphy N, Perlemuter G, Viallon V, Tjønneland A, Rostgaard-Hansen AL, Dahm CC, Overvad K, Boutron-Ruault MC, Mancini FR, Mahamat-Saleh Y, Aleksandrova K, Kaaks R, Kühn T, Trichopoulou A, Karakatsani A, Panico S, Tumino R, Palli D, Tagliabue G, **Naccarati A**, Vermeulen RCH, Bueno-de-Mesquita HB, Weiderpass E, Skeie G, Ramón Quirós J, Ardanaz E, Mokoroa O, Sala N, Sánchez MJ, Huerta JM, Winkvist A, Harlid S, Ohlsson B, Sjöberg K, Schmidt JA, Wareham N, Khaw KT, Ferrari P, Rothwell JA, Gunter M, Riboli E, Scalbert A, Jenab M.
   Int J Cancer. 2021 Feb 1;148(3):609-625. IF (2020): 7.396

2. **Unsupervised Inference of Protein Fitness Landscape from Deep Mutational Scan**

   Fernandez-de-Cossio-Diaz J, Uguzzoni G, **Pagnani A**.
   Mol Biol Evol. 2021 Jan 4;38(1):318-328. IF (2020): 16.240

3. **Soluble Receptor for Advanced Glycation End-products (sRAGE) and Colorectal Cancer Risk: A Case-Control Study Nested within a European Prospective Cohort**

   Aglago EK, Rinaldi S, Freisling H, Jiao L, Hughes DJ, Fedirko V, Schalkwijk CG, Weiderpass E, Dahm CC, Overvad K, Eriksen AK, Kyrø C, Boutron-Ruault MC, Rothwell JA, Severi G, Katzke V, Kühn T, Schulze MB, Aleksandrova K, Masala G, Krogh V, Panico S, Tumino R, **Naccarati A**, Bueno-de-Mesquita B, van Gils CH, Sandanger TM, Gram IT, Skeie G, Quirós JR, Jakszyn P, Sánchez MJ, Amiano P, Huerta JM, Ardanaz E, Johansson I, Harlid S, Perez-Cornago A, Mayén AL, Cordova R, Gunter MJ, Vineis P, Cross AJ, Riboli E, Jenab M.
   Cancer Epidemiol Biomarkers Prev. 2021 Jan;30(1):182-192. IF (2020): 4.254

4. **GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality**

   McCrory C, Fiorito G, Hernandez B, Polidoro S, O'Halloran AM, Hever A, Ni Cheallaigh C, Lu AT, Horvath S, **Vineis P**, Kenny RA.
   J Gerontol A Biol Sci Med Sci. 2021 Apr 30;76(5):741-749. IF (2020): 6.053

5. **RAS-protein activation but not mutation status is an outcome predictor and unifying therapeutic target for high-risk acute lymphoblastic leukemia**

   Koschut D, Ray D, Li Z, Giarin E, Groet J, Alić I, Kham SK, Chng WJ, Ariffin H, Weinstock DM, Yeoh AE, **Basso G**, Nižetić D.
   Oncogene. 2021 Jan;40(4):746-762. IF (2020): 1.880

6. **Short-term personal and outdoor exposure to ultrafine and fine particulate air pollution in association with blood pressure and lung function in healthy adults**

   van Nunen E, Hoek G, Tsai MY, Probst-Hensch N, Imboden M, Jeong A, **Naccarati A**, **Tarallo S**,

Raffaele D, Nieuwenhuijsen M, Vlaanderen J, Gulliver J, Amaral AFS, **Vineis P**, Vermeulen R.
Environ Res. 2021 Mar;194:110579. IF (2020): 6.498

7. **Genetic variations in microRNA-binding sites of solute carrier transporter genes as predictors of clinical outcome in colorectal cancer**

Bendova P, **Pardini B**, Susova S, Rosendorf J, Levy M, Skrobanek P, Buchler T, Kral J, Liska V, Vodickova L, Landi S, Soucek P, **Naccarati A**, Vodicka P, Vymetalkova V.
Carcinogenesis. 2021 Apr 17;42(3):378-394. IF (2020): 4.944

8. **Solving the α-helix correspondence problem at medium-resolution Cryo-EM maps through modeling and 3D matching**

Behkamal B, Naghibzadeh M, **Pagnani A**, Saberi MR, Al Nasr K.
J Mol Graph Model. 2021 Mar;103:107815. IF (2020): 2.518

9. **Comparison of smoking reduction with improvement of social conditions in early life: simulation in a British cohort**

Lepage B, Colineaux H, Kelly-Irving M, **Vineis P**, Delpierre C, Lang T.
Int J Epidemiol. 2021 Jul 9;50(3):797-808. IF (2020): 7.196

10. **RAB11-Mediated Trafficking and Human Cancers: An Updated Review**

Ferro E, **Bosia C**, **Campa CC**.
Biology (Basel). 2021 Jan 4;10(1):26. IF (2020): 5.079

11. **Targeting Cancer Heterogeneity with Immune Responses Driven by Oncolytic Peptides**

Vitale I, Yamazaki T, Wennerberg E, Sveinbjørnsson B, Rekdal Ø, Demaria S, Galluzzi L.
Trends Cancer. 2021 Jun;7(6):557-572. IF (2020): 14.226

12. **Intake of Natural Compounds and Circulating microRNA Expression Levels: Their Relationship Investigated in Healthy Subjects With Different Dietary Habits**

Ferrero G, Carpi S, Polini B, **Pardini B**, Nieri P, Impeduglia A, Grioni S, **Tarallo S**, **Naccarati A**.
Front Pharmacol. 2021 Jan 14;11:619200. IF (2020): 6.244

13. **Metabolic control of DNA methylation in naive pluripotent cells**

Betto RM, Diamante L, Perrera V, Audano M, **Rapelli S**, **Lauria A**, Incarnato D, Arboit M, Pedretti S, Rigoni G, Guerineau V, Touboul D, Stirparo GG, Lohoff T, Boroviak T, Grumati P, Soriano ME, Nichols J, Mitro N, **Oliviero S**, Martello G.
Nat Genet. 2021 Feb;53(2):215-229. IF (2020): 38.330

14. **SCIRT lncRNA Blocks the Shot of Breast Cancer Cells Self-Renewal Mechanism**

Pardini B, Dragomir MP.
Cancer Res. 2021 Feb 1;81(3):535-536. IF (2020): 12.701

15. **Control of replication stress and mitosis in colorectal cancer stem cells through the interplay of PARP1, MRE11 and RAD51**

Manic G, Musella M, Corradi F, Sistigu A, Vitale S, **Soliman Abdel Rehim S**, **Mattiello L**, Malacaria E, Galassi C, Signore M, Pallocca M, Scalera S, Goeman F, De Nicola F, **Guarracino A**, Pennisi R, Antonangeli F, Sperati F, Baiocchi M, Biffoni M, Fanciulli M, Maugeri-Saccà M, Franchitto A, Pichierri P, **De Maria R**, **Vitale I**.
Cell Death Differ. 2021 Jul;28(7):2060-2082. IF (2020): 15.828

16. **Adipocyte-derived extracellular vesicles regulate survival and function of pancreatic β cells**

Gesmundo I, **Pardini B**, Gargantini E, Gamba G, Birolo G, Fanciulli A, Banfi D, Congiusta N, Favaro E, Deregibus MC, Togliatto G, Zocaro G, Brizzi MF, Luque RM, Castaño JP, Bocchiotti MA, Arolfo S, Bruno S, Nano R, Morino M, Piemonti L, Ong H, Matullo G, Falcón-Pérez JM, Ghigo E, Camussi G, Granata R.
JCI Insight. 2021 Mar 8;6(5):e141962. IF (2020): 8.315

17. **Molecular dynamics of targeting CD38 in multiple myeloma**

Malavasi F, Faini AC, Morandi F, Castella B, Incarnato D, **Oliviero S**, Horenstein AL, Massaia M, van de Donk NWCJ, Richardson PG.
Br J Haematol. 2021 May;193(3):581-591. IF (2020): 1.050

18. **Precision Medicine in Osteosarcoma: MATCH Trial and Beyond**

Tirtei E, Campello A, Asaftei SD, Mareschi K, **Cereda M**, Fagioli F.
Cells. 2021 Jan 31;10(2):281. IF (2020): 6.600

19. **Oncosuppressive functions of PIDD1 in response to centrosome amplification**

Vitale I, **Manic G**, Galluzzi L.
Cell Death Dis. 2021 Feb 11;12(2):175. IF (2020): 8.469

20. **Intratumoral heterogeneity in cancer progression and response to immunotherapy**

Vitale I, Shema E, Loi S, Galluzzi L.
Nat Med. 2021 Feb;27(2):212-224. IF (2020): 53.440

21. **Relative Information Gain: Shannon entropy-based measure of the relative structural conservation in RNA alignments**

Pietrosanto M, Adinolfi M, **Guarracino A**, Ferrè F, Ausiello G, **Vitale I**, Helmer-Citterich M.
NAR Genom Bioinform. 2021 Feb 15;3(1):lqab007. IF (2020): N.A.

22. **Genome-scale deconvolution of RNA structure ensembles**

Morandi E, Manfredonia I, Simon LM, Anselmi F, van Hemert MJ, **Oliviero S**, Incarnato D.
Nat Methods. 2021 Mar;18(3):249-252. IF (2020): 28.547

23. **Guidelines for the use and interpretation of assays for monitoring autophagy (4th edition)(1)**

Klionsky DJ, Abdel-Aziz AK, Abdelfatah S, Abdellatif M, Abdoli A, Abel S, Abeliovich H, Abildgaard MH, Abudu YP, Acevedo-Arozena A, Adamopoulos IE, Adeli K, *et al.* (**Vitale I**)
Autophagy. 2021 Jan;17(1):1-382. IF (2020): 16.016

24. **Droplet Digital PCR Improves IG-/TR-based MRD Risk Definition in Childhood B-cell Precursor Acute Lymphoblastic Leukemia**

Della Starza I, Nunes V, Lovisa F, Silvestri D, Cavalli M, Garofalo A, Campeggio M, De Novi LA, Soscia R, Oggioni C, Mussolin L, Biondi A, Guarini A, Valsecchi MG, Conter V, Biffi A, **Basso G**, Foà R, Cazzaniga G.
Hemasphere. 2021 Feb 24;5(3):e543. IF (2020): 0.310

25. **Long Noncoding RNAs in Human Stemness and Differentiation**

Mirzadeh Azad F, **Polignano IL**, **Proserpio V**, **Oliviero S**.
Trends Cell Biol. 2021 Jul;31(7):542-555. IF (2020): 20.808

26. **Optimality in Self-Organized Molecular Sorting**

Zamparo M, Valdembri D, Serini G, Kolokolov IV, Lebedev VV, **Dall'Asta L**, **Gamba A**.
Phys Rev Lett. 2021 Feb 26;126(8):088101. IF (2020): 9.161

27. **Initial cell density encodes proliferative potential in cancer cell populations**

Enrico Bena C, Del Giudice M, Grob A, Gueudré T, Miotto M, Gialama D, Osella M, Turco E, Ceroni F, De Martino A, **Bosia C**.
Sci Rep. 2021 Mar 17;11(1):6101. IF (2020): 4.380

28. **Personalized therapeutic strategies in HER2-driven gastric cancer**

Ughetto S, Migliore C, Pietrantonio F, Apicella M, Petrelli A, D'Errico L, Durando S, Moya-Rull D, Bellomo SE, Rizzolio S, Capelôa T, Ribisi S, Degiuli M, Reddavid R, Rapa I, Fumagalli U, De Pascale S, Ribero D, Baronchelli C, Sgroi G, Rausa E, Baiocchi GL, Molfino S, Manenti S, Bencivenga M, Sacco M, Castelli C, Siena S, Sartore-Bianchi A, Tosi F, Morano F, Raimondi A, Prisciandaro M, Gloghini A, Marsoni S, Sottile A, Sarotto I, Sapino A, Marchiò C, Cassoni P, **Guarrera S**, Corso S, Giordano S.
Gastric Cancer. 2021 Jul;24(4):897-912. IF (2020): 7.370

29. **Immune Trait Shifts in Association With Tobacco Smoking: A Study in Healthy Women**

Piaggeschi G, Rolla S, Rossi N, Brusa D, **Naccarati A**, Couvreur S, Spector TD, Roederer M, Mangino M, **Cordero F**, Falchi M, Visconti A.
Front Immunol. 2021 Mar 9;12:637974. IF (2020): 7.561

30. **The Non-Coding RNA Journal Club: Highlights on Recent Papers-8**

Tarallo S, **Pardini B**, Fox AH, Ingram H, Taccioli C, Taube JH, Mani SA.
Noncoding RNA. 2021 Mar 19;7(1):23. IF (2020): N.A.

31. **CD56, HLA-DR, and CD45 recognize a subtype of childhood AML harboring CBFA2T3-GLIS2 fusion transcript**

Zangrando A, Cavagnero F, Scarparo P, Varotto E, Francescato S, Tregnago C, Cuccurullo R, Fagioli F, Nigro LL, Masetti R, Putti MC, Rizzari C, Santoro N, Pession A, Pigazzi M, Locatelli F, **Basso G**, Buldini B.
Cytometry A. 2021 Aug;99(8):844-850. IF (2020): 4.355

32. **Artificial Intelligence in Bulk and Single-Cell RNA-Sequencing Data to Foster Precision Oncology**

Del Giudice M, **Peirone S**, **Perrone S**, **Priante F**, **Varese F**, Tirtei E, Fagioli F, **Cereda M**.
Int J Mol Sci. 2021 Apr 27;22(9):4563. IF (2020): 5.924

33. **The Targeting of MRE11 or RAD51 Sensitizes Colorectal Cancer Stem Cells to CHK1 Inhibition**

Mattiello L, **Soliman Abdel Rehim S**, Musella M, Sistigu A, **Guarracino A**, Vitale S, Corradi F, Galassi C, Sperati F, **Manic G**, **De Maria R**, **Vitale I**.
Cancers (Basel). 2021 Apr 19;13(8):1957. IF (2020): 6.639

34. **Identification of novel circulating microRNAs in advanced heart failure by next-generation sequencing**

Galluzzo A, Gallo S, **Pardini B**, Birolo G, Fariselli P, Boretto P, Vitacolonna A, Peraldo-Neia C, Spilinga M, Volpe A, Celentani D, Pidello S, Bonzano A, Matullo G, Giustetto C, Bergerone S, Crepaldi T.
ESC Heart Fail. 2021 Aug;8(4):2907-2919. IF (2020): 4.411

35. **Temporal and Epigenetic Control of Plasticity and Fate Decision during CD8(+) T-Cell Memory Differentiation**

Pace L.
Cold Spring Harb Perspect Biol. 2021 Dec 1;13(12):a037754. IF (2020): 10.005

36. **Application of the Euro Clonality next-generation sequencing-based marker screening approach to detect immunoglobulin heavy chain rearrangements in mantle cell lymphoma patients: first data from the Fondazione Italiana Linfomi MCL0208 trial**

Genuardi E, Romano G, Beccuti M, Alessandria B, Mannina D, Califano C, Rota Scalabrini D, Cortelazzo S, Ladetto M, Ferrero S, Calogero RA, **Cordero F**.
Br J Haematol. 2021 Jul;194(2):378-381. IF (2020): 1.050

37. **Expectation propagation on the diluted Bayesian classifier**

Braunstein A, Gueudré T, **Pagnani A**, Pieropan M.
Phys Rev E. 2021 Apr;103(4-1):043301. IF (2020): 2.529

38. **BRIO: a web server for RNA sequence and structure motif scan**

Guarracino A, Pepe G, Ballesio F, Adinolfi M, Pietrosanto M, Sangiovanni E, **Vitale I**, Ausiello G, Helmer-Citterich M.
Nucleic Acids Res. 2021 Jul 2;49(W1):W67-W71. IF (2020): 16.971

39. **New DNA Methylation Signals for Malignant Pleural Mesothelioma Risk Assessment**

Cugliari G, Allione A, Russo A, Catalano C, Casalone E, **Guarrera S**, Grosso F, Ferrante D, Sculco M, La Vecchia M, Pirazzini C, Libener R, Mirabelli D, Magnani C, Dianzani I, Matullo G.
Cancers (Basel). 2021 May 27;13(11):2636. IF (2020): 6.639

40. **HNRNPM controls circRNA biogenesis and splicing fidelity to sustain cancer cell fitness**

Ho JS, Di Tullio F, Schwarz M, Low D, Incarnato D, Gay F, Tabaglio T, Zhang J, Wollmann H, Chen L, An O, Chan THM, Hall Hickman A, Zheng S, Roudko V, Chen S, Karz A, Ahmed M, He HH, Greenbaum BD, **Oliviero S**, Serresi M, Gargiulo G, Mann KM, Hernando E, Mulholland D, Marazzi I, Wee DKB, Guccione E.
Elife. 2021 Jun 2;10:e59654. IF (2020): 8.146

41. **Hair follicle stem cell progeny heal blisters while pausing skin development**

Fujimura Y, Watanabe M, Ohno K, Kobayashi Y, Takashima S, Nakamura H, Kosumi H, Wang Y, Mai Y, **Lauria A**, **Proserpio V**, Ujiie H, Iwata H, Nishie W, Nagayama M, **Oliviero S**, Donati G, Shimizu H, Natsuga K.
EMBO Rep. 2021 Jul 5;22(7):e50882. IF (2020): 8.807

42. **Genetic variations in 3'UTRs of SMUG1 and NEIL2 genes modulate breast cancer risk, survival and therapy response**

Cumova A, Vymetalkova V, Opattova A, Bouskova V, **Pardini B**, Kopeckova K, Kozevnikovova R, Lickova K, Ambrus M, Vodickova L, **Naccarati A**, Soucek P, Vodicka P.
Mutagenesis. 2021 Aug 27;36(4):269-279. IF (2020): 3.000

43. **Physics of compartmentalization: How phase separation and signaling shape membrane and organelle identity**

Floris E, **Piras A**, **Dall'Asta L**, **Gamba A**, Hirsch E, **Campa CC**.
Comput Struct Biotechnol J. 2021 May 24;19:3225-3233. IF (2020): 7.271

44. **Mutational landscape of plasma cell-free DNA identifies molecular features associated with therapeutic response in patients with colon cancer. A pilot study**

Cervena K, **Pardini B**, Urbanova M, Vodenkova S, Eva P, Veskrnova V, Levy M, Buchler T, Mokrejs M, **Naccarati A**, Vodicka P, Vymetalkova V.
Mutagenesis. 2021 Oct 6;36(5):358-368. IF (2020): 3.000

45. **Expression and Function of C1orf132 Long-Noncoding RNA in Breast Cancer Cell Lines and Tissues**

Shafaroudi AM, Sharifi-Zarchi A, Rahmani S, Nafissi N, Mowla SJ, **Lauria A**, **Oliviero S**, Matin MM.
Int J Mol Sci. 2021 Jun 23;22(13):6768. IF (2020): 5.924

46. **Two Worlds Colliding: The Interplay Between Natural Compounds and Non-Coding Transcripts in Cancer Therapy**

Sabo AA, Dudau M, Constantin GL, Pop TC, Geilfus CM, **Naccarati A**, Dragomir MP.
Front Pharmacol. 2021 Jul 6;12:652074. IF (2020): 6.244

47. **Epidemic mitigation by statistical inference from contact tracing data**

Baker A, Biazzo I, **Braunstein A**, Catania G, **Dall'Asta L**, Ingrosso A, Krzakala F, Mazza F, Mézard M, **Muntoni AP**, Refinetti M, Sarao Mannelli S, Zdeborová L.
Proc Natl Acad Sci U S A. 2021 Aug 10;118(32):e2106548118. IF (2020): 11.205

48. **Inflammation-Related Marker Profiling of Dietary Patterns and All-cause Mortality in the Melbourne Collaborative Cohort Study**

Li SX, Hodge AM, MacInnis RJ, Bassett JK, Ueland PM, Midttun Ø, Ulvik A, Rinaldi S, Meyer K, Navionis AS, Shivappa N, Hébert JR, Flicker L, Severi G, Jayasekara H, English DR, **Vineis P**, Southey MC, Milne RL, Giles GG, Dugué PA.
J Nutr. 2021 Oct 1;151(10):2908-2916. IF (2020): 4.798

49. **Editorial: Non-Coding RNAs as Mediators of the Activity of Natural Compounds**

Scoditti E, **Naccarati A**, Carpi S, Polini B, Ebada SS, Nieri P.
Front Pharmacol. 2021 Aug 19;12:751956. IF (2020): 6.244

50. **Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation**

Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E, Lawson DJ, Burrows K, Suderman M, Bretherick AD, Richardson TG, Klughammer J, Iotchkova V, Sharp G, Al Khleifat A, Shatunov A, Iacoangeli A, McArdle WL, Ho KM, Kumar A, Söderhäll C, Soriano-Tárraga C, Giralt-Steinhauer E, Kazmi N, Mason D, McRae AF, Corcoran DL, Sugden K, Kasela S, Cardona A, Day FR, Cugliari G, Viberti C, **Guarrera S**, Lerro M, Gupta R, Bollepalli S, Mandaviya P, Zeng Y, Clarke TK, Walker RM, Schmoll V, Czamara D, Ruiz-Arenas C, *et al.*
Nat Genet. 2021 Sep;53(9):1311-1321. IF (2020): 38.330

51. **SERS liquid biopsy: An emerging tool for medical diagnosis**

Moisoiu V, Iancu SD, Stefancu A, Moisoiu T, **Pardini B**, Dragomir MP, Crisan N, Avram L, Crisan D, Andras I, Fodor D, Leopold LF, Socaciu C, Bálint Z, Tomuleasa C, Elec F, Leopold N.
Colloids Surf B Biointerfaces. 2021 Dec;208:112064. IF (2020): 5.268

52. **Sparse generative modeling via parameter reduction of Boltzmann machines: Application to protein-sequence families**

Barrat-Charlaix P, **Muntoni AP**, Shimagaki K, Weigt M, Zamponi F.
Phys Rev E. 2021 Aug;104(2-1):024407. IF (2020): 2.529

53. **Analysis of MicroRNA Expression Changes During the Course of Therapy In Rectal Cancer Patients**

Cervena K, Novosadova V, **Pardini B**, **Naccarati A**, Opattova A, Horak J, Vodenkova S, Buchler T, Skrobanek P, Levy M, Vodicka P, Vymetalkova V.
Front Oncol. 2021 Sep 2;11:702258. IF (2020): 6.244

54. **Genetic and Epigenetic Characterization of a Discordant KMT2A/AFF1-Rearranged Infant Monozygotic Twin Pair**

Russo A, Viberti C, Mareschi K, Casalone E, **Guarrera S**, Birolo G, Cazzaniga G, Corral L, Trentin L, **Basso G**, Fagioli F, Matullo G.
Int J Mol Sci. 2021 Sep 9;22(18):9740. IF (2020): 5.924

55. **Efficient generative modeling of protein sequences using simple autoregressive models**

Trinquier J, **Uguzzoni G**, **Pagnani A**, Zamponi F, Weigt M.
Nat Commun. 2021 Oct 4;12(1):5800. IF (2020): 14.919

56. **Molecular Characterization of Prostate Cancers in the Precision Medicine Era**

Giunta EF, Annaratone L, Bollito E, Porpiglia F, **Cereda M**, Banna GL, Mosca A, Marchiò C, Rescigno P.
Cancers (Basel). 2021 Sep 24;13(19):4771. IF (2020): 6.639

57. **Faecal miRNA profiles associated with age, sex, BMI, and lifestyle habits in healthy individuals**

Francavilla A, **Gagliardi A**, **Piaggeschi G**, **Tarallo S**, **Cordero F**, Pensa RG, Impeduglia A, Caviglia GP, Ribaldone DG, Gallo G, Grioni S, Ferrero G, **Pardini B**, **Naccarati A**.
Sci Rep. 2021 Oct 19;11(1):20645. IF (2020): 4.380

58. **AMaLa: Analysis of Directed Evolution Experiments via Annealed Mutational Approximated Landscape**

Sesta L, **Uguzzoni G**, Fernandez-de-Cossio-Diaz J, **Pagnani A**.
Int J Mol Sci. 2021 Oct 9;22(20):10908. IF (2020): 5.924

59. **adabmDCA: adaptive Boltzmann machine learning for biological sequences**

Muntoni AP, **Pagnani A**, Weigt M, Zamponi F.
BMC Bioinformatics. 2021 Oct 29;22(1):528. IF (2020): 3.169

60. **Deliver on Time or Pay the Fine: Scheduling in Membrane Trafficking**

Placidi G, **Campa CC**.
Int J Mol Sci. 2021 Oct 29;22(21):11773. IF (2020): 5.924

61. **MicroRNA 146a is associated with diabetic complications in type 1 diabetic patients from the EURODIAB PCS**

Barutta F, Corbetta B, Bellini S, **Guarrera S**, Matullo G, Scandella M, Schalkwijk C, Stehouwer CD, Chaturvedi N, Soedamah-Muthu SS, Durazzo M, Gruden G.
J Transl Med. 2021 Nov 25;19(1):475. IF (2020): 5.531

62. **Autophagy: The Potential Link between SARS-CoV-2 and Cancer**

Habibzadeh P, **Dastsooz H**, Eshraghi M, Łos MJ, Klionsky DJ, Ghavami S.
Cancers (Basel). 2021 Nov 16;13(22):5721. IF (2020): 6.639

63. **Using epigenetic modifiers to target cancer stem cell immunoevasion**

Galassi C, **Vitale I**, Galluzzi L.
Cancer Cell. 2021 Dec 13;39(12):1573-1575. IF (2020): 31.743

64. **Three-Dimensional Graph Matching to Identify Secondary Structure Correspondence of Medium-Resolution Cryo-EM Density Maps**

Behkamal B, Naghibzadeh M, Saberi MR, Tehranizadeh ZA, **Pagnani A**, Al Nasr K.
Biomolecules. 2021 Nov 26;11(12):1773. IF (2020): 4.879

65. **Genome-Wide Analysis of Smad7-Mediated Transcription in Mouse Embryonic Stem Cells**

Meng G, **Lauria A**, **Maldotti M**, **Anselmi F**, **Polignano IL**, **Rapelli S**, Donna D, **Oliviero S**.
Int J Mol Sci. 2021 Dec 18;22(24):13598. IF (2020): 5.924