Allegato 1 – Progetto National Center for Gene Therapy and Drugs based on RNA Technology

Spoke # 7: Biocomputing. Spoke leader UNIBA

State of the field and unmet needs. The extraordinary technological advances that led to the development of novel, massively parallel and increasingly cost-effective high-throughput technologies (e.g. massive nucleotide sequencing, single cell analysis) have revolutionized biomedical research leading Biology in the era of "Big Data". The unprecedented ability to profile biological systems in a quantitative and accurate manner, however, is creating a pressing need to bridge the gap between high-throughput technological development and our ability for managing, analyzing, and integrating biological big data. Massively parallel assays are now routinely applied to investigate the mutational, transcriptional, epigenetic, and proteomic landscapes of healthy and pathological samples. Moreover, recent advances in single-cell techniques offer the unprecedented opportunity to disentangle tissue complexity and investigate cell identities and functions within tissues. Several are the gaps to be filled to optimize the capacity to transform collections of multi-layered data and patients' characteristics into models of cellular functions and diseases, and then into effective products for the healthcare market as, for example, new RNA-based therapeutic products and innovative approaches for gene therapy. Concerning the exploitation of DNA- and RNA-based therapeutic approaches, no efficient algorithms for designing reliable guide RNAs for large-scale CRISPR-Cas and RNA-based approaches are available and the direct detection of nucleotide chemical modification through the analysis of third generation sequencing platforms is still at infancy. The prediction of 3D RNA folding is also still quite an immature field as computational modeling for RNA still suffers from significant limitations related to accurate conformational sampling and adequate potentials for extensive and reliable simulations. In addition, the difficulty of experimentally obtaining the structure of RNA and its complexes with biomolecules of therapeutic of interest, in particular, through the techniques of X-ray spectrometry or nuclear magnetic resonance (NMR), is still today one of the most limiting factors in the design and development of new RNA drugs. The improvement of computational methodologies for the accurate prediction of the RNA structure, and its complexes, represents today one of the most promising approaches to respond to this need in timescales that are competitive with those of drug development. Finally, a very relevant largely unmet issue is the development of computational approaches for AI-based drug discovery and multidimensional data integration (e.g. study of intracellular effects of RNA drugs exploiting 'omics and pathway analysis) cross-correlating molecular, clinical and personal lifestyle data. This requires a big effort in data and methods FAIRification to guarantee robustness and reproducibility of predicted inferences.

How the Spoke will contribute to advance in this scenario. Spoke #7 will provide a dedicated infrastructure, fully integrated in the framework developed by the National Center for HPC, Big Data and Quantum Computing, with a state of the art ecosystem of data and tools supporting the activities of the vertical Spokes. This integrated bioinformatic environment will play a dual role: on the one hand, providing data-driven hypothesis generating insights and testable hypotheses on novel classes of targets (such as unannotated genes or novel transcriptional isoforms), and on the other hand, our team will intervene in the fine tailoring of an RNA-based therapeutics, pinpointing to what ideal RNA sequence to be used as therapeutics (in siRNA, ASO and mRNA applications) in constant iteration between experimental and biocomputing predictions. The work will develop in 4 parallel WP. In WP7.1, a rich portfolio of data and tool resources will be implemented or suitably developed to establish a cutting-edge work environment in the computational infrastructure also implementing new algorithms for sequence and structural data analysis, to support the research activities of all

other Spokes. These will include tools for genomic, transcriptomics, epitranscriptomics, epigenetic and proteomic analyses also at single cell resolution. WP7.2 will be devoted to the identification of neo-antigens for mRNA vaccinology carried out through large-scale epitranscriptomics profiling. An integrated computational pipeline will be created to identify Alternative Splicing (AS) from RNA-seq and infer integrating information of epitranscriptomic modifications detectable by Nanopore data analysis. AS and RNA editing (RE) provide novel potential targets for vaccines that may vastly outnumber mutational cancer neoantigens and may represent a largely unexplored pool of antigens in non-tumoral disease settings, including autoimmunity. WP7.2 also plans to develop a multiomics platform for the prediction of tumor neoantigens. The goal of WP7.3 is the development and application of informatics/computational methodologies to support the design and industrial development of RNA drugs in different therapeutic areas of the vertical Spokes. This will include a reiterated computational and experimental validation interplay aiming at the accurate prediction of the tertiary structure of native and chemically modified RNA and of their target. This can be achieved by optimizing Molecular Dynamics simulations, able to explore the conformational space in solution of native and chemically modified RNA and the interaction with small organic molecules, proteins, oligonucleotides and membrane systems. The integration with docking simulation data and Cryo-EM maps remarkably improves the accuracy and resolution of inferred functional-structural relationships. WP7.4 is dedicated to the development and application of data integration strategies focused on the reconstruction of regulatory networks also at single-cell resolution, the development of prioritization strategies for drug targets and actionable disease features, to patient stratification and identification of disease vulnerabilities.

The activities of Spoke 7 will be fully integrated with those of the other vertical and horizontal Spokes. In particular, we plan to launch an internal call for “on-demand bioinformatics services” and “web-based bioinformatics platforms” to make the NC bioinformatics ecosystem fully compliant to project needs. Some relevant interactions have been already planned, e.g. with Spoke #6 for designing programmable RNA editing enzymes, with Spokes #2 and #3 for their validation in cancer and neurodegeneration, and with Spoke #10 for optimizing biosafety in cell-based immunotherapy approaches. Concerning 73 the planned interactions with private companies a collaborative action with Intesa San Paolo Innovation Center has been planned for implementing a smart IT platform for genomic surveillance during viral outbreaks. The private Partner Intesa Sanpaolo will support this Spoke with Intesa Sanpaolo Innovation Center (ISPIC), an Intesa Sanpaolo subsidiary born with the mission of exploring and assimilating new future business models, with a strong focus on Circular Economy and Digital Transformation. In particular ISPIC has developed with the University of Milano and the Sacco Hospital a smart IT platform for SaRS-CoV-2 genomic surveillance. The idea is to implement the performances of this platform for other possible viral outbreaks and open it to the scientific community.