Artificial Intelligence-based model in identification the patterns of cancer cell and immune cell interaction during establishment of cancer metastasis and treatment-resistance

Purpose and Aims:

Vision and Goals: To develop and apply Artificial Intelligence (AI)/Machine Learning (ML)-based innovative prediction and decision-making models for disease prediction and designing tailored multi-module treatment of metastatic colorectal cancer.

- **Specific Aim**: Develop and apply Artificial Intelligence (AI)/Machine Learning (ML)-based innovative prediction model for risk stratification by using gene/protein expression signatures of tumor cells and tumor-associated immune cells

There is growing emphasis on developing clinical decision support systems based on AI-Machine Learning (ML) techniques. However, AI models in using ML techniques are not well-developed or applied for risk-stratification and tailored treatment for metastatic cancer. To address this urgent need, We wish to develop and apply AI-prediction models for the discovery of combinatorial therapies for metastatic cancer. **In this project, (i) we will first use the mRNA expression data of the samples from human cancer cells and immune cells isolated from the mono-culture and co-culture conditions in *in vitro* system to identify the unique gene signatures and ontology for cancer cells and immune cells. (ii) We will identify the patterns and specific functional pathways that cancer cells and immune cells recruit from each other during co-culture. (iii) Based on the bioinformatics analysis and clustering, we will** develop AI-predictive models which have ability for pattern recognition, analysis of complex bio-medical data, and molecular pathway-prediction. The long-term goal of our project is illustrated in (**Figure 1**).

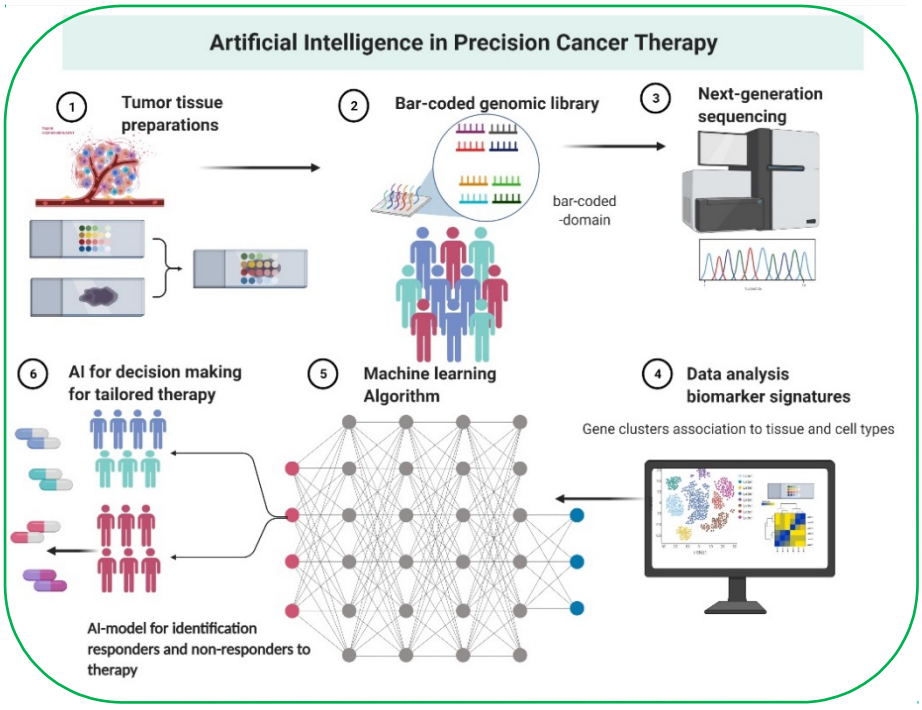

Figure 1. Schematic chart depicts the workflow of the study aims. The availability of both solid and liquid tumor biopsies will be subjected to next-generation sequencing to achieve fingerprint of individual patients. The clinical data and the biomolecular fingerprint – will be used to build a ML model**.** Using this ML approach, predicting the non-responsiveness to treatment, will aid in selecting metastatic cancer patients, who might potentially benefit from novel combinatorial therapy.

State-of-the-art

BOX 1 Background of colorectal cancer: Approximately 50% of colorectal cancer patients develop distant metastases (Van Cutsem et al., 2016). This disease has high prevalence, frequent development of metastases, and high costs of drugs, which constitutes a high economic burden on the health and social systems. Further, tumors rapidly acquire new mutations during chemotherapies. There is urgent need for develop prediction models based on the tumour mutational status integrated with clinical information to design effective and precision therapy.
AI-Machine Learning: Big data such as genomic, RNA-seq, proteomic, epigenetic, metabolomics, imaging and clinical data of primary colorectal tumors and metastatic lesions are too gigantic and impossible to be analyzed using the conventional tools and techniques. AI-ML is an emerging new tool for elucidating the molecular status of a tumor from pathological data. ML can provide scores of tumor marker protein expression in tissue slides. Deep learning CNN model has been used to successfully predict which genes in tumor tissue may harbor mutations (with AUC from 0.733 to 0.856) (Gupta et al., 2021). The ML-deep learning has already been used to perform versatile tasks, such as cancer diagnosis, at a level equal to or sometimes greater than that of expert physicians.

Preliminary and previous results

(i). In collaboration with our research partners from USA and China, we developed **ML-algorithms** of multitude-biomarkers panels for stratification and subtyping cancer stages and predicting/monitoring treatment outcome. **Using small quantities of liquid biopsies urine samples** collected prior and after surgery and chemotherapies, we were able to use this ML model to subtype and predict patient outcome (Johnson et al., BMC Medicine, 2020, Guo et al., Frontiers in Cell and Developmental Biology, 2020; Guo et al., Clinical Translational Medicine, 2021).
(ii). In collaboration with our partners within the EU REVERT consortium, we have been building REVERT databank (REVERT-DB) by collecting multiple clinical and molecular parameters including gene mutations, epigenetic changes and gene expression profiling signatures of colorectal cancer. REVERT-DB will be integrated to train the machine learning algorithms. Our partners' labs have demonstrated the potential of a ML approach based on Multiple Kernel Learning (MKL) (Ferroni P, et al. Med. Decision Making. 2017; Ferroni P, et al. Cancers, 2019). This ML approach that combines Support Vector Machines (SVM) algorithm and Random Optimization (RO) is capable of exploiting significant patterns in routinely collected demographic, clinical and biochemical data and allows the design of a clinical decision support system (DSS).

Project description

- **Specific Aim**: Develop and apply Artificial Intelligence (AI), Machine Learning (ML)-based innovative prediction model for risk stratification by using gene/protein expression signatures of tumor cells and tumor-associated immune cells.

  Experimental approaches: We have performed next-generation sequencing, single cell sequencing and integrated molecular parameters including gene mutations, epigenetic changes and gene expression profiling signatures of the human prostate cancer cells and cancer associated cells. The analysis of cancer patients and tumor-derived cell lines to genomic and proteomic analysis is described in our previous publications (Semenas et al., Molecular Oncology, 2020). For this thesis project, **Data collection, analysis, hierarchical sorting and classification** by using the mRNA expression data of the samples from human cancer cells and immune cells isolated from the mono-culture and co-culture conditions in *in vitro* system to identify the unique gene signatures and ontology for cancer cells and immune cells. Bioinformatics analysis including hierarchical sorting and classification of the biomolecular signature/fingerprint of individual sample will be performed with the approach described in Semenas et al., Molecular Oncology, 2020 as a starting point for the student. Unsupervised hierarchical clustering can be performed to compare the clusters of primary tumor cells and metastatic cells. Gene ontology enrichment analysis can be further performed to identify the differential expression between subtypes of cells. One aim is to generate a major gene signature involved in survival, proliferation and invasion. The mRNA expression profiles show the existence of distinct malignant subpopulations exhibited different expression pattern. In addition, the two important pieces of information – the unique biomolecular fingerprint and functional pathways– can be used to build

cancer-immune cell recoginition ML model based on the already available RISK model, possibly capable of identifying those patients who may not benefit from first line therapy. Our model will be used for predicting risk of cancer metastasis and treatment response.

Significance and scientific novelty

- Joint force with EU network to improve Research and Development on AI-health
- Access to European network of computer technologies and biobanks data derived from retrospective and currently running clinical studies
- Innovative and cost-effective personalized risk-stratification and targeted therapy using AI models
- AI-based decision support system to contribute AI-medicine and health in Sweden

Time-lines and milestones

The total project period for a Mater thesis student and pre-PhD student is estimated to be 6-10 months. The Master-degree work in Machine Learning and Statistics is held at Linköping University. As a joint effort from both Umeå University and Linköping University.

Equipment and Facilities

We have multiple centers and core facilities at Department of Molecular Biology, MIMS, UCMR, ICE Lab and Metabolomics Center at Umeå University and Hospital. We have the most advanced bioinformatics, computer sciences and infrastructures. We have access to the sequencing facility to perform sequencing of RNAtagSeq libraries. We have also access to SciLifeLab platforms for single cells sequencing analysis. We have our large infrastructures and REVERT Biobank and computer sciences center in Rome, Italy and Madrid in Spain.

Internationellt och nationellt samarbete (International and national collaboration)

As the Chairman of the Medical Faculty Council for Artificial Intelligence and Autonomic systems, Umeå, I lead the AI work at the Medical faculty. I am also the Member of Steering Board for Medical Faculty Biobank Infrastructure PREDICT. As a board member of Umeå University Artificial Intelligence Council (Rådet för AI) since jan, 2020, my major role is to make strategic plan and implementation of the AI research and innovation at Umeå University.

International collaboration partners: 1) EU-REVERT Consortium led by SAN RAFFAELE ROMA SRL, Rome, Italy. There are 20 partners including IT and computer sciences across EU that are involved in this consortium. 2) Clinical Laboratories, San Francisco General Hospital, San Francisco, California, USA; 3) IT and Computational companies in Santa Clara, California, USA; 4) Dept. of Biosciences and Informatics, Keio University, Japan; 5) University of Nottingham; 6) Kazan Federal State University, Russia; 7) Weill Cornell Medical Center.

Ethical aspects of the project

In the proposed study, the data security and privacy in compliance with all National and European regulations will be applied, not only for patients, but also for physicians and healthcare professionals, as well as for ICT (Information and Communication Technologies) infrastructure developers. We have previously obtained ethical permissions for using human tissues including freshly isolated and achieved clinical samples and clinical information. We will strictly follow the ethical guidelines in using both human tissues samples and clinical data.

References:

1. Ferroni P, et al. *Disease Markers*. 2017. https://doi.org/10.1155/2017/8781379.
2. Ferroni P, et al. *Cancers* (Basel). 2019. *11*(3), 328.
3. Gupta R et al., *Molecular Diversity*, 2021, doi.org/10.1007/s11030-021-10217-3.
4. Semenas et al., Molecular Oncology, 2020, 15(4):968-986.
5. Van Cutsem E, et al. Ann Oncol. 2016, 27:1386–1422.