Post-Doc position (AI-ML/Molecular-Histopathology)

SANOFI/UEVE-IBISC EA4526

March 2022

1 Introduction

Tumor and immune micro-environment is a recognized characteristic of cancer cells [1].

These two underlying cellular and molecular contexts influence many other biological processes, including cell signaling or the epigenetic control of gene expression.

The quantitative characterization of tumour-infiltrating lymphocytes (TILs), for example, is of growing importance in precision medicine. With the growth of cancer immunotherapy, these characterizations are likely to have increasing clinical importance, as understanding each patient's immune response becomes more important. Recent studies further suggest that the spatial context and nature of cellular heterogeneity within the tumor microenvironment, in terms of immune infiltrate in the tumor center and invasive margin, is important in cancer prognosis. Prognostic factors, including immunoscore, which quantify spatial densities of TIL in different tumor regions, have high prognostic value. Therefore, assessments of tumor-associated lymphocytes are increasingly important both in the clinical evaluation of pathology slides and in translational research on the role of these lymphocyte populations. Our goal is to better understand cancer heterogeneity in solid tumors from the perspective of tumor/immune infiltration and to assess how nutrient and oxygen gradients influence tumor cell fate and drug resistance. . Ultimately, understanding these processes would help find relevant patient subgroups with actionable treatment outcomes (i.e., survival, treatment response, clinical trial recruitment).

2 Project overview and objectives

Cancer is a complex disease involving multiple genetic and epigenetic changes that continually evolve during disease progression. To survive and proliferate, cancer cell populations use adaptive evolutionary strategies based on heterogeneity and survival of the fittest cells [3].

The high plasticity of cancer cells leads to the rewiring of signaling pathways and metabolic networks, all in response to changes in their micro-environmental conditions. For all these reasons, the mathematical modeling of the evolution of cancer must include several biological scales: molecular, cellular and tissue. In this project, we represent cancer cells as

distributions on multiple spatial and anatomical dimensions and study their evolution using high-dimensional partial differential equation models.

3 Scope of the work

The recent approval of immuno-oncology therapies holds great promise to address the need, but resistance and insufficient biomarkers to predict response [5] underscore the need for better characterization of the disease. immunosuppression in the tumor microenvironment (TME). Single-cell transcriptomics, multiplex (low and high plex technologies) have revealed intra-tumor heterogeneity (ITH) in many types of cancer, identifying cell populations that promote drug resistance, predict metastatic risk and mediate cancer. plasticity [21, 22, 20]. However, recent advances allow simultaneous capture of the locations of dozens of cell types in the TME, which is essential for understanding tumour-stroma crosstalk [19, 2, 16, 17]. The orthogonal integration of high-dimensional single-cell spatial data from normal and diseased tissues as well as standard/routine digital pathology data should therefore facilitate the dissection of TME cell communication.

Epithelial cancers represent approximately 90% of human malignancies. Here, we combine single-cell and spatial transcriptomics (Nanostring, GeoMX) with single-cell resolution multiplexed protein imaging of a series of primary human solid tumors, as well as matched normal tissues and routine H&E digital pathology data. These data will be complemented by public data-sets using similar technologies (MIBI, 10×Genomics). Among these, NSCLC and mesothelioma will serve as a prototype to study the role of TME in deep patient phenotyping using multi-modal spatial tissue imaging data. In case of positive results, the methodology will be applied to other solid tumors of strategic interest for internal immuno-oncology pipelines (i.e. breast, pancreas, liver and bladder). The selection of the indication, the target population for each indication should be further discussed with our internal commercial collaborator for this project.

In this project, various AI methods [4] – including deep learning – will be used to quantify tumor spatial heterogeneity at different scales (blood vessel distribution, invasion fronts, cell clusters and metabolic marker distribution) from H&E stained histological sections5, tissue imaging data-sets by multiplexed immunofluorescence and mass cytometry. Ultimately, these methods will feed the mathematical model with the information [9] (initial data and parameters) needed to predict tumor progression and response [10] during targeted therapy [11, 12, 13, 14]. Particular emphasis will be placed on the explainability of innovative AI approaches, allowing the development of a creative framework [15], by involving our biomedical partners) [18].

4 Task list

To be defined upon agreement with our collaborators. However, an initial plan would be:

4.1 Preliminary Phases

1. Scoping: Write the research protocol and datasets specification
2. Data Access: Activate identified business partners and install/enable the infrastructure to support at the long run the RD activities.
3. Knowledge Engine Priming: Review literature and preprocess public and internal datasets. Knowledge synthesis to be integrated in analysis or used in post hoc annotation of targets relevant to the current project and that are complementary to ongoing activities.
4. Data Generation: Determine possible new data modalities to be generated to enrich current datasets and that are needed for the development.

4.2 Analysis Phases

1. Data Preprocessing: Curate clinical data, preprocess molecular data (e.g. harmonization across datasets) and histology data (e.g. matter detection, tiling, etc.).
2. Disease Subtype Identification: Perform unsupervised or guided clustering of patients with multimodal data. Characterization of clusters and prioritization.
3. Subgroup Characterization: Build predictive models from multimodal data and extract biologically relevant features for subtypes.

4.3 Discovery Phases

1. Target Prioritization: Prioritize potential targets based on strength of association with subtype features and mechanistic relations with disease phenotype
2. Target Discovery: Apply Knowledge Engine to identify possible biomarkers of identified subtypes from characterization multimodal features.

5 Project Schedule

The planned duration is 12 months. The key phases and milestones are listed above. A more detailed version could be discussed with our collaborators.

The project it is a collaboration between DDS/AIDA represented by the BIDP (Biomedical Imaging and Digital Pathology) team located in Chilly-Mazarin (FR), Translational Sciences Molecular Histology and Digital Pathology laboratory located in Vitry-sur-Seine (FR), and IBISC laboratory from université Paris-Saclay.

References

[1] Hanahan, D. Hallmarks of Cancer: New Dimensions. Cancer Discov 12, 31–46 (2022).

[2] Angelo, M., Bendall, S., Finck, R. *et al.* Multiplexed ion beam imaging of human breast tumors. Nat Med 20, 436–442 (2014). https://doi.org/10.1038/nm.3488

[3] Barnett, A. J. *et al.* A case-based interpretable deep learning model for classification of mass lesions in digital mammography. Nat Mach Intell 1–10 (2021) doi:10.1038/s42256-021-00423-x.

[4] Berry, S. *et al.* Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. Science 372, eaba2609 (2021).

[5] Keenan, S.R., Currie, P.D. (2019) The Developmental Phases of Zebrafish Myogenesis. Journal of developmental biology. 7(2).

[6] Partel, G. & Wählby, C. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. Febs J 288, 1859–1870 (2021).

[7] 5. Cao, R. *et al.* Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. Theranostics 10, 11080–11091 (2020).

[8] Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer 1–13 (2021) doi:10.1038/s41568-021-00408-3.

[9] Levy-Jurgenson, A., Tekpli, X., Kristensen, V. N. & Yakhini, Z. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. Sci Rep-uk 10, 18802 (2020).

[10] Grünwald, B. T. *et al.* Spatially confined sub-tumor microenvironments in pancreatic cancer. Cell (2021) doi:10.1016/j.cell.2021.09.022.

[11] Lu, S., Fürth, D. & Gillis, J. Integrative analysis methods for spatial transcriptomics. Nat Methods 1–2 (2021) doi:10.1038/s41592-021-01272-7.

[12] Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods 1–10 (2021) doi:10.1038/s41592-021-01255-8.

[13] Zhao, T. *et al.* Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. Nature 1–7 (2021) doi:10.1038/s41586-021-04217-4.

[14] Vu, T. *et al.* Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis. Nat Commun 13, 169 (2022).

[15] Zheng, Y. *et al.* Encoding histopathology whole slide images with location-aware graphs for diagnostically relevant regions retrieval. Med Image Anal 76, 102308 (2021).

[16] Salmén, F., Ståhl, P.L., Mollbrink, A. *et al.* Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. Nat Protoc 13, 2501–2534 (2018). https://doi.org/10.1038/s41596-018-0045-2

[17] Thrane, K., Eriksson, H., Maaskola, J., Hansson, J., & Lundeberg, J. (2018). Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. Cancer research, 78(20), 5970–5979. https://doi.org/10.1158/0008-5472.CAN-18-0747

[18] Wang, Y. *et al.* Predicting molecular phenotypes from histopathology images: a transcriptome-wide expression-morphology analysis in breast cancer. Cancer Res 81, canres.0482.2021 (2021).

[19] Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S. R., Kurian, A., Van Valen, D., West, R., Bendall, S. C., and Angelo, M. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell, 174(6), 1373–1387.e19. https://doi.org/10.1016/j.cell.2018.08.039

[20] Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., 2nd, Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G., Kazer, S. W., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (New York, N.Y.), 352(6282), 189–196. https://doi.org/10.1126/science.aad0501

[21] Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., Richman, A. R., Silverbush, D., Shaw, M. L., Hebert, C. M., Dewitt, J., Gritsch, S., Perez, E. M., Gonzalez Castro, L. N., Lan, X., Druck, N., Rodman, C., Dionne, D., Kaplan, A., Bertalan, M. S., ... Suvà, M. L. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell, 178(4), 835–849.e21. https://doi.org/10.1016/j.cell.2019.06.024

[22] Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., & Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell, 171(7), 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044