

Unraveling genetic heterogeneity of malignant and healthy cells in Waldenstrom Macroglobulinemia through integration of single-cell omics and whole-genome sequencing

Christos Vlachos¹, Ourania Theologi¹, Maria Sakkou¹, Ioannis Kollias¹, Nefeli Mavrianou-Koutsoukou¹, Christine Liacos¹, Alexandra Papadimou¹, Kostantina Taouxi¹, Katerina Chrisostomidou¹, Meletios A. Dimopoulos¹, Efstathios Kastritis¹, Tina Bagratuni¹

1. Department of Clinical Therapeutics, School of Medicine, National and Kapodistrian University of Athens, Greece

Waldenstrom Macroglobulinemia (WM) is a lymphoproliferative disorder characterized by high levels of Immunoglobulin M (IgM) produced by lymphoplasmacytoid cells which have infiltrated the bone marrow. Even though WM is mainly considered a clonal disease, significant heterogeneity within the clones has been observed.

To date, the genetic background of healthy and malignant B-cells in WM, as well as how well we can isolate one state from another in diseased patients remains largely unknown. Our study cohort of >20 samples diagnosed with WM constitutes a unique resource of data in order to elucidate genetic heterogeneity and clonality in WM at the single-cell level. We leverage the 10X Genomics technology to perform paired single-cell gene expression (GEX) and B-Cell Receptor (BCR) sequencing from bone marrow aspirates, as well as whole genome sequencing (WGS) of CD19+ B-cells from matched donor samples.

We first show that, even though WM clonal cells are highly pure at the stage of diagnosis, variation of disease-associated cancer cells across all patients is very high compared to the gene expression variation of their healthy cells.

Further, we use whole-genome, single-cell GEX and BCR sequencing to elucidate the genetic variation that differentiates malignant from healthy B-cells. To this aim we first perform copy number variation (CNV) analyses on B-cells, from both single-cell and WGS assays, to evaluate how well single-cell approaches can recapitulate the CNV architecture of tumor cells, estimate tumor purity and isolate malignant from normal cells. We complement this approach by performing immunoglobulin light chain kappa/lambda ratio analyses of the gene expression assay. Using the clonotypes derived from the BCR sequencing information as ground truth we show that the kappa/lambda ratio of B-cell light chains can very accurately separate the normal from malignant B-cells, demonstrating a very high correlation (~95%) with the BCR clonality information. In addition, we employ machine learning and deep learning approaches, such as logistic regression, and multi-layer neural networks to build a classifier using the BCR information for training. The classifiers are then tested on GEX data showing very good performance at identifying cancer cells in a sample-specific manner. Finally, through the identification of differentially expressed genes (DEGs) and Gene Set Enrichment analyses (GSEA) we aim to discover WM cancer specific

genes and pathways to better understand the underlying genetic factors participating in the mechanisms involved in the disease.

We are developing a comprehensive WM single-cell and WGS dataset which we are leveraging in order to assess and develop novel bioinformatics methodologies to accurately discriminate neoplastic from healthy cells in highly pure tumors. We demonstrate the power of the high-throughput 10X GEX+BCR assay and WGS to study genetic heterogeneity across WM patients, offering a unique resource to the community.