

Advancing Multi-Cancer Early Detection: High-performance cell-free RNA profiling with the Flomics liquid biopsy platform

Phil Sanders¹, Lucila González¹, Laura Sanchez¹, Beatriz Neto¹, Julien Lagarde¹, Marc Weber¹, Pablo Monteagudo¹, Cristina Tuñí¹, Giovanni Asole¹, Elena Cristina Rusu¹, Sol García², Pilar Simón-Extermera², Sonia Panadero³, María Concepción Romero³, Purificación Catalina³, João Curado¹.

¹ Flomics Biotech, Roc Boronat 31, Barcelona 08005, Spain. ² Health and Biomedicine, Leitat Technological Center, Barcelona Science Park, Barcelona 08028, Spain. ³ Andalusian Public Health System Biobank, Coordinating Node, Granada 18016, Spain

Introduction

Cell-free RNA (cfRNA) is a rich source of potential biomarkers in liquid biopsies as it is actively secreted into biofluids, including blood, by both healthy and diseased cells.

cfRNA analysis of a liquid biopsy therefore provides insight into the health of an individual.

cfRNA is well suited to early cancer detection, and its advantages include:

- **Earliest detection:** Detectable RNA changes often occur before the detection of DNA mutations and symptoms, which is crucial for the detection of small and early cancers.
- **Tissue specificity:** RNA varies by tissue, facilitating cancer tissue of origin identification.
- **Personalized Medicine:** RNA provides functional information about the dysregulated signalling pathways in the cancer and guides treatment.
- **Dynamic molecule:** Recurring measurement of RNA levels allows the establishment of a baseline signal which is key for Minimal Residual Disease (MRD) detection and monitoring.

At Flomics Biotech we have developed a high quality cfRNA-Seq platform that profiles human plasma cfRNA in a robust and reproducible manner. We are currently applying this platform in the LiquiDx pre-clinical study with the goal of developing a cfRNA-based multi-cancer early detection test.

Flomics cell-free RNA-Seq platform

In this study we apply our cfRNA-Seq platform (Figure 1 - Left) to plasma samples from a cohort of over 1,000 donors (Figure 1 - Right). 1 ml of plasma from each donor was processed and sequenced at a read depth of between 30 and 50 million paired end reads per library.

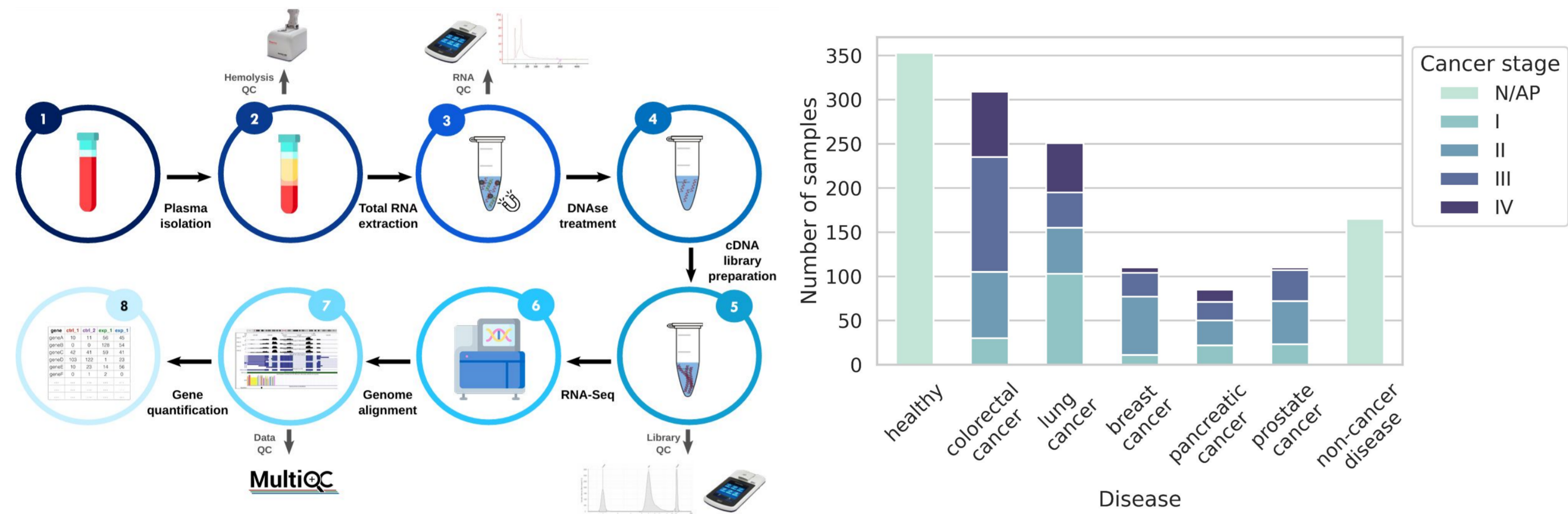


Figure 1. Left: Summary of the Flomics cfRNA-Seq platform. Data generated by the platform is analysed using a combination of gold-standard bioinformatics tools and advanced machine learning methods to identify cancer-specific biomarker signatures and develop predictive machine learning models for early cancer detection. QC = quality control. **Right:** Sample distribution across the different diseases. For each cancer type sample distribution across cancer stages is indicated. "non-cancer disease": patients with non-cancer diseases of the same organs as the cancer types.

High-quality cancer biomarker detection

The Flomics cfRNA-Seq platform generates high quality data rich in potential biomarkers, with over 6000 genes detected in 80% of samples (Figure 2 - top left). The genes are distributed across a broad range of biotypes (Figure 2 - bottom). Many oncogenes and tumour suppressor genes are detected (Figure 2 - top right), showing the potential of the Flomics cfRNA-Seq platform to detect cancer-related signals.

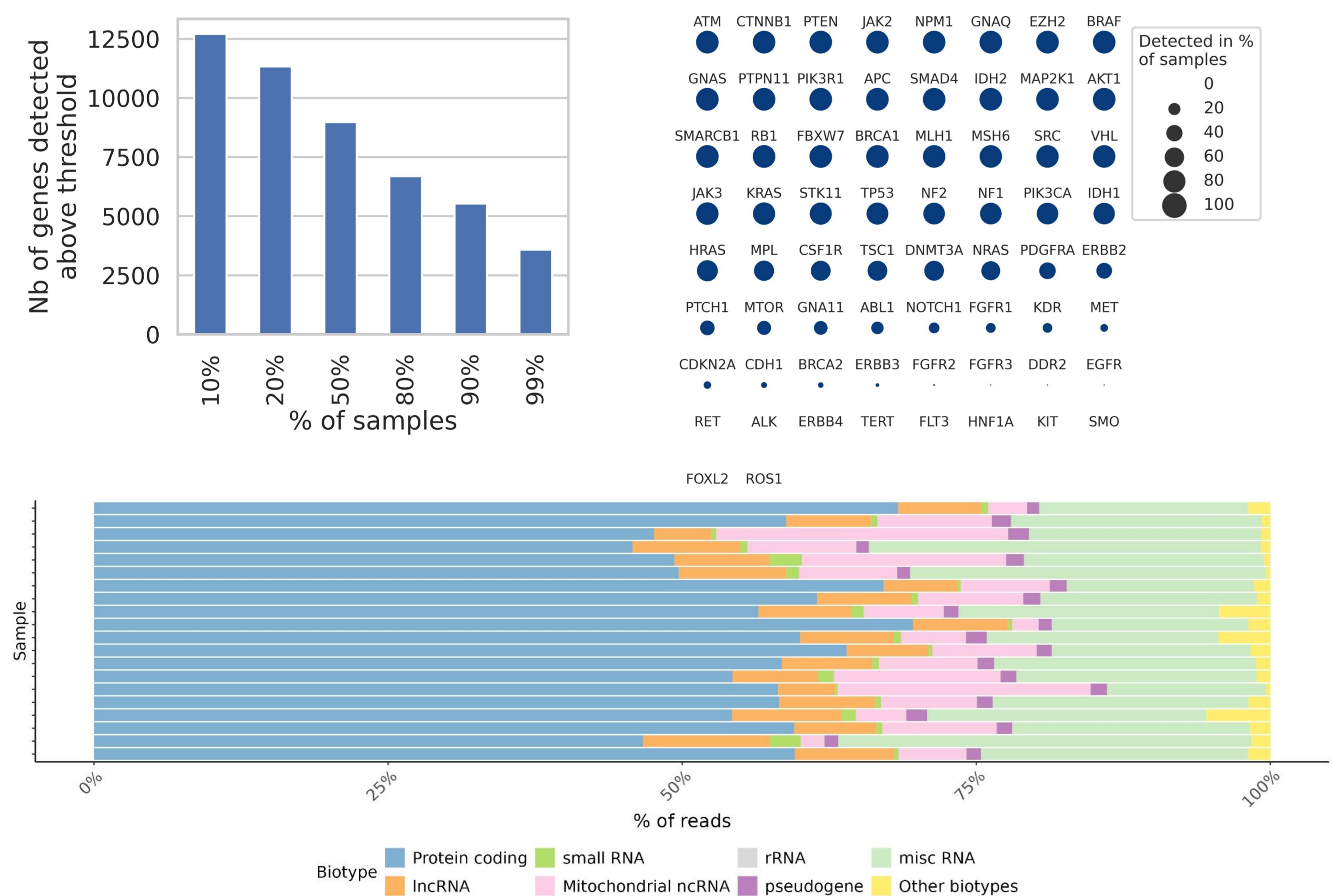


Figure 2. Top left: Number of genes detected >1 TPM in the indicated percentage of samples. **Top right:** For a panel of 65 oncogenes and tumour suppressor genes, the percentage of samples (disk size) in which the indicated gene is detected with >1 TPM. **Bottom:** Gene biotype distribution in 20 randomly selected samples (y axis). The x axis represents the percentage of total reads in a sample mapping to each biotype. The distribution of reads across biotypes displayed here is representative of the full data set. "Small RNA" combines all small RNA biotypes, e.g. miRNA, snRNA, and snoRNA. "Mitochondrial ncRNA" combines the mitochondrial rRNA and tRNA biotypes.

Conclusions

- Flomics has developed a world class cfRNA-Seq platform that will pioneer the use of cfRNA in clinical practice.
- We have developed a highly promising test for detecting cancer and identifying the cancer tissue-of-origin.
- Our technology maintains high performance across all cancer stages including stage I, illustrating the benefits of using cfRNA for cancer early detection.

Identifying a cancer biomarker signature

Healthy and cancer samples have distinct cfRNA expression profiles that can be used to separate them (Figure 3 - Left). Differential gene expression analysis identifies 36 significantly upregulated and 5 significantly downregulated genes in the cancer sample group to give a 41 gene cancer biomarker signature (Figure 3 - Middle). Gene set enrichment analysis of the data set reveals the enrichment of genes associated with various cellular processes and signalling pathways implicated in cancer (Figure 3 - Right). This supports the validity of using cfRNA to identify cancer biomarker signatures.

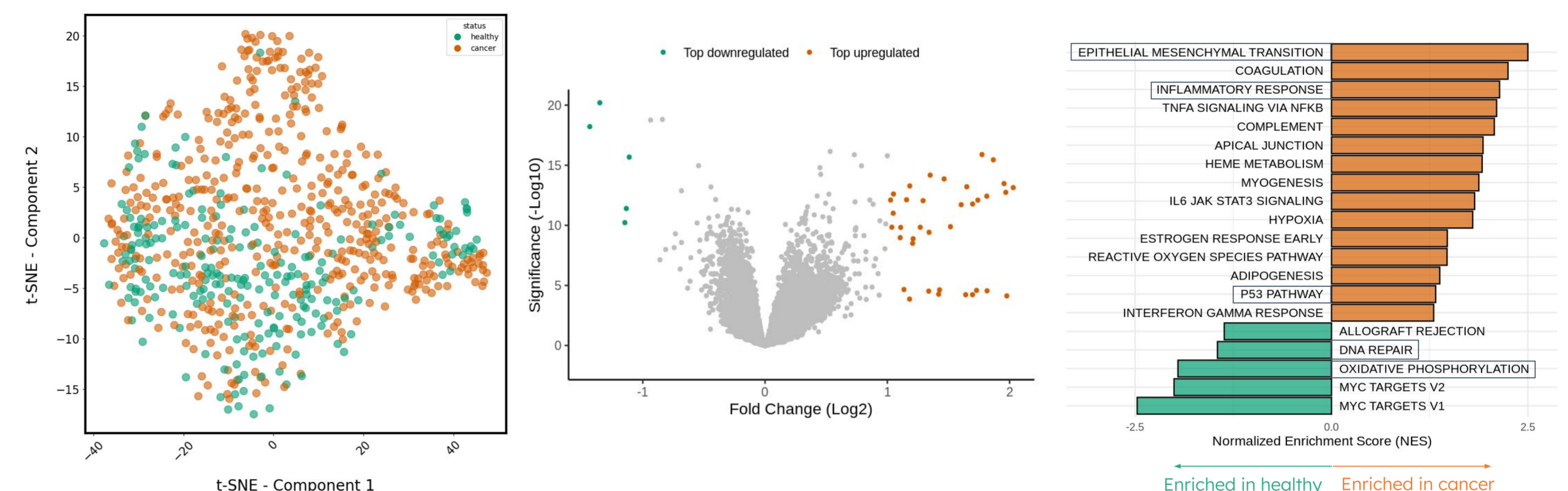


Figure 3. Left: cfRNA expression-based t-SNE representation of 711 samples. Green: healthy samples. Orange: cancer samples. Samples close to each other in t-SNE space have similar expression profiles. **Middle:** Cancer vs healthy sample group differential expression analysis. Each dot represents a gene. Orange and green dots represent significantly up- and down-regulated genes respectively in cancer samples. **Right:** Gene Set Enrichment Analysis of the data set. Positive and negative NES reflect enrichment of genes associated with the indicated cellular processes and signalling pathways in the cancer and healthy sample sets respectively. The boxes highlight processes and signaling pathways that are strongly implicated in cancer.

High-performance stage I cancer detection

We leverage differences in cfRNA expression between cancer patients and healthy individuals to build a machine learning model to detect cancer. Training of the model is ongoing, however using currently available data (Figure 4 - Left) we have developed a binary classifier that predicts cancer with promising results (Figure 4 - Middle left). For the prediction of specific cancer types, the model performs particularly well for lung and prostate cancer (Figure 4 - Middle right). The model displays a high performance level for the prediction of all cancer stages including stage I (Figure 4 - Right).

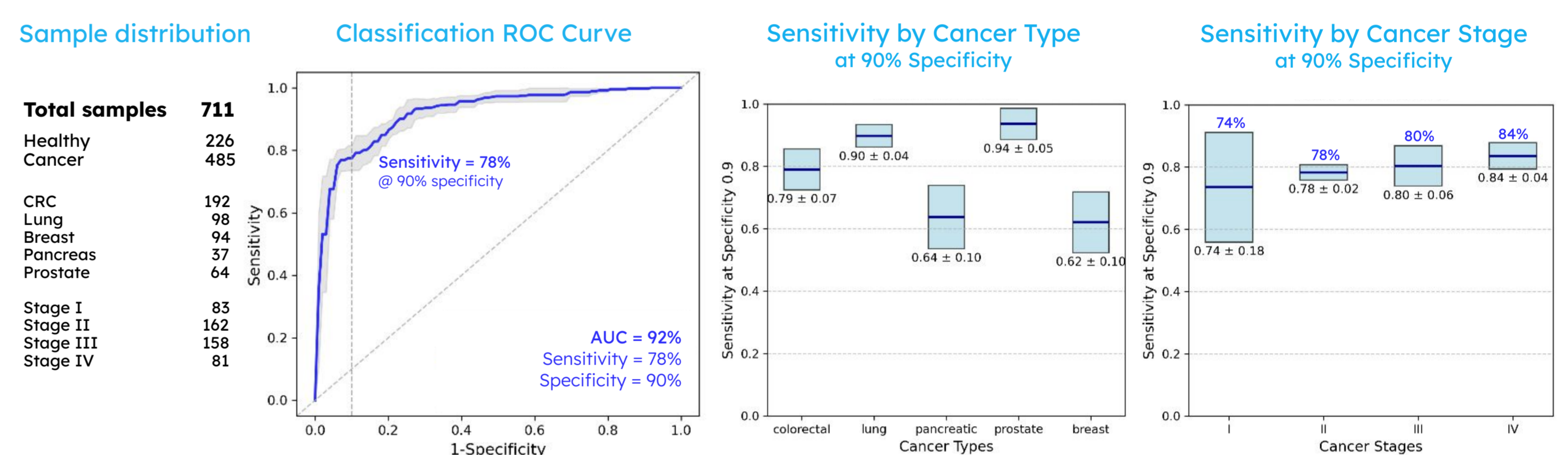


Figure 4. Left: Sample distribution across cancer types and stages. **Middle left:** Performance of our preliminary cancer vs healthy machine learning (ML) binary classifier, using the Extreme Gradient Boosting method. ROC curves (Sensitivity v.s. 1-Specificity) and Areas Under the Curve (AUCs) for the 4-folds cross-validation. **Middle right:** Performance of the ML binary classifier in predicting individual cancer types at 90% specificity. **Right:** Performance of the ML binary classifier in predicting different stages of cancer at 90% specificity.

Cancer tissue-of-origin identification

Combining early stage cancer detection with accurate prediction of cancer tissue-of-origin (CTOO) will improve patient diagnosis and prognosis. The information that we extract from cfRNA profiles can be exploited to pinpoint CTOO and so we are developing a multi-class machine learning classifier model to predict the CTOO of cancer patient samples. The model currently classifies the CTOO in 67% of samples while the remaining 33% are unclassified (Figure 5 - left). For the classified samples, the classifier determines CTOO with high accuracy, achieving 74% and 93% mean accuracy for the most probable (top-1) and the two most probable (top-2) tissues, respectively (Figure 5 - right). This approach also has great potential in identifying the tissue-of-origin in cancers of unknown primary.

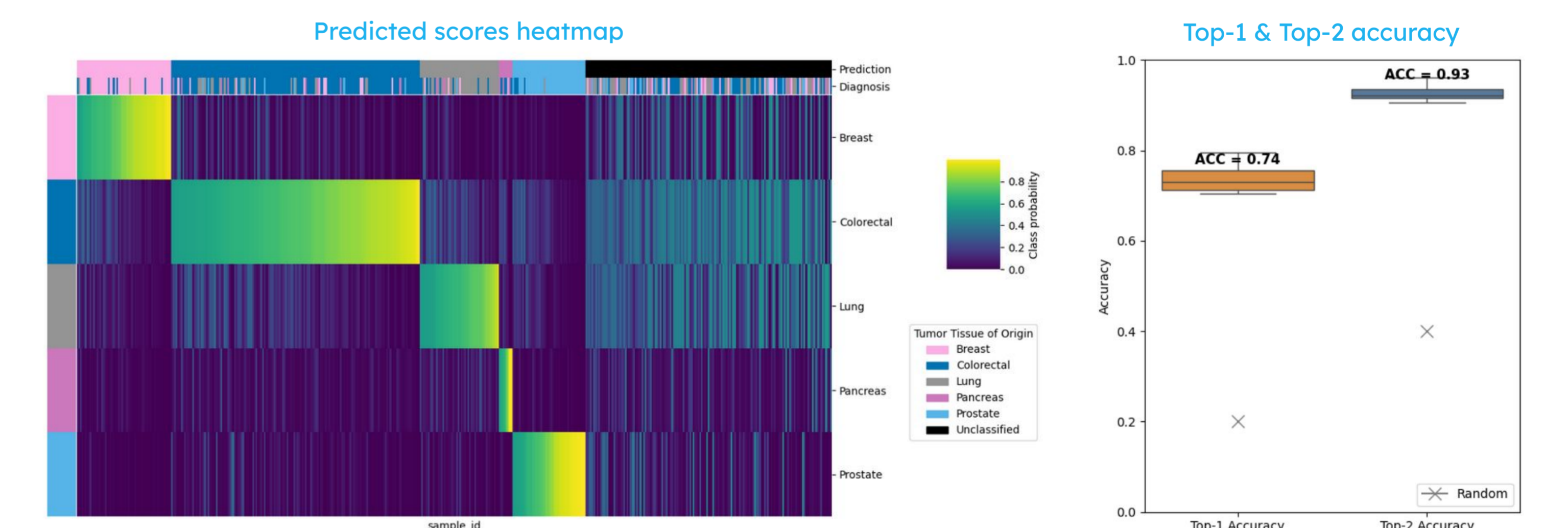


Figure 5. Left: Predicted scores of the model for each sample (x axis) and CTOO (y axis) on the validation set using 4-folds cross-validation (CV). Predicted CTOO (top row) was assigned to the tissue with highest probability. Samples with maximum probability < 0.55 for all tissues were considered as unclassified. Diagnosis (second row) indicates the actual diagnosis of the patient. **Right:** Accuracy of the CTOO multi-class ML classifier model in predicting the top-1 and top-2 most probable tissues of origin of cancer samples in the validation set using 4-folds CV. The unclassified samples are excluded from this analysis. The grey crosses indicate the expected performance of the model if guessing at random.

Working with you!



- Flomics is always looking for:
- Collaborations
 - Opportunities to provide our technology and knowhow as a service
 - Clinical expertise including sample provision

Contact us at info@flomics.com

Acknowledgements

The 'LiquiDx' project (reference CPP2021-008897) is financed by the Spanish Ministry of Science, Innovation and Universities (MCIN/AEI/10.13039/501100011033) and by the European Union "NextGenerationEU/PRTR".

