Multi-omics characterization markers of left and right-sided colorectal cancers


School, Dublin, Ireland

South Korea

Woodsmith

ABSTRACT

Right (R-C18.0,2,3) vs. left (L-C18.6,7) sided colorectal cancers are clinically distinguishable based on prognosis and response to certain therapies, yet limited data have emerged to explain these differences. Multi-omics analysis can further define disease specific subgroups. When assessing multiple signatures, pre-analytical quality of the tissue (such as ischemia time) and comparison to normal tissue controls become even more paramount to accurately reflect tumor molecular realities.

Primary tumors and paired normal tissues from colorectal patients were collected following a standardized SOP to minimize ischemia time. Comprehensive analyses were conducted using genomics (whole-genome sequencing - WGS), transcriptomics (RNA-seq), as well as annual clinical information. JADBio, AI machine learning platform, was then leveraged to perform an unbiased approach to identify simplified signatures which can then classify the right-left CRC cohorts. We have identified two novel homeobox transcription factors that almost perfectly capture between left and right CRC cohorts.

MACHINE LEARNING ANALYSIS OF LEFT-RIGHT

Supervised machine learning approach, using Jadibio (Gnosis) was used to classify left and right CRC based on the fold change in values between tumor and adjacent normal. The ROC and AUC (AUC = 0.995) show that the model can almost perfectly classify left and right-based tumors (Figure 6a).

When extracting features that have the largest impact on predictive performance, features 1 and 2 mostly influenced the clustering (Figure 6b; cumulative) and a UMAP analysis based all features shows clear separation between left and right CRC (Figure 6c).

DISTRIBUTIONS OF THE TWO NOVEL GENES & THEIR PROGNOSTIC AND/OR PREDICTIVE VALUE

Further detailed inspection of the extracted signatures reveal two novel genes that are bimodally expressed in CRC. Both genes are homebox transcription factors that define left and right CRC based on log (fold changes) from RNA-Seq data (Figure 7a).

A survival analysis based on these two novel genes, highlights that right CRC, as annotated by the homebox transcription factor 1, has a lower survival rate than left CRC (Figure 7b) indicating potential use of the targets as a prognostic and/or predictive markers.

CONCLUSION

Progress in precision medicine requires the inclusion of multi-omics data. High quality samples that are collected under standardized SOPs that account for preanalytical factors such ischemia time are critical in ensuring accurate assessment of multiple molecular signatures. Here using Indivumed’s CRC cohorts that were collected under standardized SOP, we were able to define distinct molecular markers between L and R CRC. In addition, two novel transcription factors were identified and can potentially play roles as clinical prognostic and predictive markers. Further validation will be necessary to determine their clinical utility.

RESULTS

INDIVUmed’S COLORRECTAL PATIENT COHORT OVERVIEW

474 cases with roughly equal distribution among gender, early and late stage, as well as tumor locations were selected (Figure 1). Multi-omics analysis (WGS and RNAseq) were assessed for both tumor and adjacent normal samples for each patient (Figure 2a).