Poster #3542

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

Authors: Marshall JL[1] • Yoshino T[2] • Rha SY[3] • Church D[4] • Coutinho A[5] • Sampaio C[5] • Gallagher D[6] • Garcia-Foncillas J[7] • Kerr D[4] Woodsmith JW[8] • Von Der Heyde S[8] • Bischof J[8] • Juhl H[8]

Georgetown University Medical Center, Washington, DC[1], National Cancer Center Hospital East, Kashiwa, Japan[2], Yonsei University, Seoul, South Korea [3], Oxford University Medical School, Oxford, UK [4], Clínica AMO, Salvador, BA, Brasil [5], Cancer Genetics Service, Mater Hospital School, Dublin, Ireland [6], IIS-Fundación Jimenez Diaz-UAM, Madrid, Spain [7], Indivumed GmbH, Hamburg, Germany [8]

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-analytic quality of the tissues (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 analysis 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 than can classify the left-right CRC cohorts. We have identified two novel homeobox transcription factors that almost perfectly bifurcates L&R CRC based on fold changes in RNA-seq expression.

RESULTS

INDIVUMED'S COLORECTAL 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 2&3).

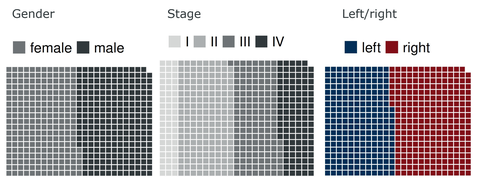

Figure 1: Demographic & stage distribution of left-sided and right-sided Indivumed collected colorectal cancer cohorts.

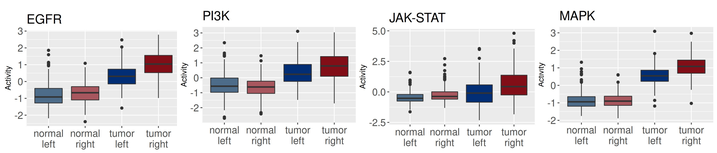

Figure 2: Differences in activation signaling pathway is noted between left-sided and right-sided colorectal cancer. As previously reported, higher signaling cascade activation is predominately observed in right-sided colorectal cancers. As expected, Indivumed CRC cohorts depict similar pattern of the high activation of EGFR, PI3K, JAK-STAT and MAPK pathways by right-sided CRC cohorts (signalling scores obtained from PROGENY).

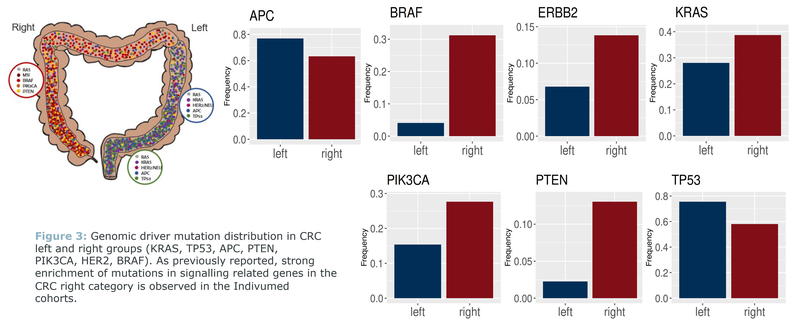

Figure 3: Genomic driver mutation distribution in CRC left and right groups (KRAS, TP53, APC, PTEN, PIK3CA, HER2, BRAF). As previously reported, strong enrichment of mutations in signalling related genes in the CRC right category is observed in the Indivumed cohorts.

CLINICALLY RELEVANT BIOMARKERS EXPRESSION PATTERN

Tumor mutational burden (TMB), microsatellite instability (MSI), consensus molecular subtypes (CMS) distribution patterns in left-sided and right-sided were assessed using RNASeq data. TMB for the right-sided CRC cases are significantly more mutated than the left-sided cohorts (Figure 4a. p = 2.9e-05). The right-sided CRC cohorts also have high MSI expression than the left-sided cohorts (4b). CMS1 & CMS3 subtypes are predominately to be in right-sided where as CMS2 and CM4 are left-sided CRC cohorts (4c).

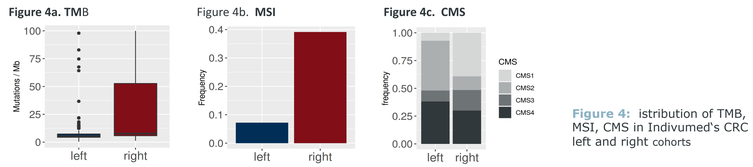

Figure 4: istribution of TMB, MSI, CMS in Indivumed's CRC left and right cohorts

IMMUNE CELLS INFLITRATION PROFILE

Tumor-infiltrating immune cells were quantified from transcriptomics data. Higher levels of cytotoxic lymphocytes including NK and CD8+ T cells are noted in the right-sided CRC cohorts. In addition, lower levels of NK cells infiltration were noted in the tumor compared to normal adjacent for both left and right-sided CRC cohorts (Figure 5).

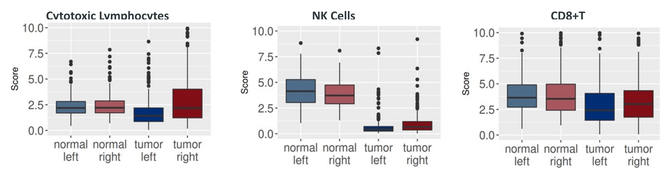

Figure 5: Distribution of immune infiltration cell types in CRC left and right groups. Boxplots of tumor with adjacent normal tissues are shown.

MACHINE LEARNING ANALYSIS OF LEFT-RIGHT

Supervised machine learning approach, using JadBio (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).

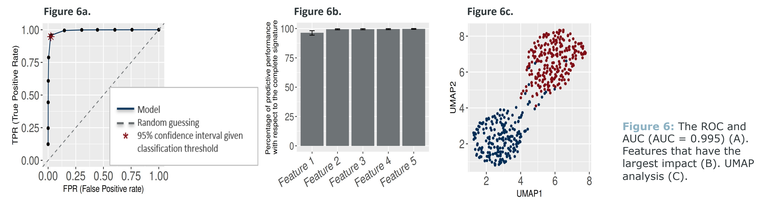

Figure 6: The ROC and AUC (AUC = 0.995) (A). Features that have the largest impact (B). UMAP analysis (C).

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 homeobox transcription factors that define left and right CRC based on log (fold changes) from RNA-Seq data (Figure 7a).

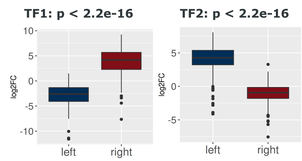

Figure 7a: Expression distributions of the two novel genes (TF1 & TF2)

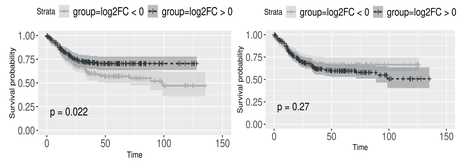

Figure 7b: (Left TF1, Right TF2) Based on survival analysis shows the two novel genes have the prognostic predictive value

A survival analysis based on these two novel genes, highlights that right CRC, as annotated by the homeobox 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 R and L CRC. In addition, two novel transcription factors were identified and can potentially play role as clinical prognostic and predictive markers. Further validation will be necessary to determine their clinical utility.