Introduction: Several independent groups have recently reported novel molecular subtypes in colorectal cancer (CRC). A systematic comparison of the reported subtypes is needed to determine common findings and facilitate translation into the clinic. A collaborative effort by the CRCSC was initiated to enable open data sharing and meta-analysis with the goal of establishing a consensus subtyping model.

Methods: CRCSC participants, representing more than 15 institutions, analyzed 30+ gene expression sets, including multiple platforms and sample preparation methods. Six previously published classifiers (with 3-6 subtypes each) were applied to a dataset comprising more than 5,000 samples, consisting primarily of stage II-III colon cancer. A central independent team (Sage Bionetworks) provided analysis concordance of subtype calls, and clinical/molecular/pathology annotation with methodology derived through consensus.

Results: Despite variation in the datasets and classification methodology in the separate published subtypes, there were 4 consensus molecular subtypes (CMS1-4) clearly defined by the concordance analysis, as indicated by a high degree of interconnectivity among the subtype calls from the 6 methods. CMS1 (∼15%) was enriched for MSI, right-side tumors, older age, females, hypermutation, BRAF mutation and immune pathway activation. Both CMS2 and 3 displayed epithelial markers, canonical WNT and MYC pathway activation. Highly proliferative CMS2 tumors (∼40%) were CIN, MSS, left-sided, TP53 mut with EGFR upregulation and had better survival rates. CMS3 tumors (∼10%) had low CIN status, higher proportion of KRAS/PIK3CA mutations and IGFBP2 overexpression. CMS4 tumors (∼20%) were characterized by mesenchymal/TGF-beta signaling upregulation, younger age at diagnosis, NOTCH3/VEGFR2 overexpression and worse survival outcomes. The remaining samples (∼15%) did not have a consensus assignment, which may be attributed to additional minor subtypes. Further refinement is anticipated.

Conclusion: This consortium represents the first effort to generate consensus cancer subtypes and a major step forward for precision medicine in CRC. By comparing multiple classifiers we were able to identify four biologically distinct subtypes of CRC with unique clinical and molecular markers, a remarkable overlap considering different methodologies and datasets used by individual groups. Our efforts on improving this consensus classification system and correlating subtypes with benefit from adjuvant chemotherapies/response to targeted agents will be presented.