DECIPHERING CHROMOSOMAL INSTABILITY IN CONSENSUS MOLECULAR SUBTYPES (CMS) IN CRC: INSIGHTS FROM AN INTEGRATIVE MULTI-OMICS APPROACH

Indivumed Therapeutics

#409

Julia Bischof[1], David Church[2], Jonathan Woodsmith[1]
[1] Indivumed GmbH, Falkenried 88, 20251 Hamburg, Germany; [2] University of Oxford (Nuffield Department of Medicine), Old Road Campus, Oxford OX3 7BN, England

INTRODUCTION

CRC is a heterogeneous disease with different molecular subtypes, which can have a profound impact on treatment response and patient outcomes. CMS classification based on gene expression profiles has been shown to differentiate patients into clinically relevant sub-groups. However, deep multi-omics characterization of the CMS classes based on high quality patient data is lacking. Through the integration of multi-omics data, we aim to enhance our understanding of CRC's molecular heterogeneity and thereby facilitate the development of personalized therapeutic strategies.

COHORT OVERVIEW

We performed a multi-omics analysis of CRC patients to identify potential new features from CMS groups. 798 fresh-frozen, surgically resected CRC tumor and adjacent normal samples were used to analyze whole genome, RNA sequencing and mass spectrometry (whole proteome profiling) data.

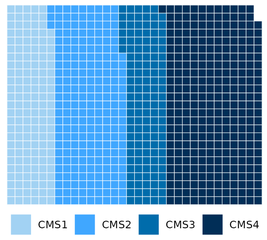

Figure 1: CMS classification on the entire CRC cohort was called using R Package CMSCaller [1] based on transcriptomics data. (CMS1: 18.4%, CMS2: 27.8%, CMS3: 16.3%, CMS4: 37.5%).

METHODS

Collection of samples: Tumor tissues were globally collected using a standardized protocol, minimizing the ischemia time until freezing in liquid nitrogen. To ensure the quality of the samples, all tissues were H&E stained and subjected to a pathological QC. Samples need to be invasive, have a tumor content of >= 30 % and Necrosis <= 30 %. Tumor-free normal tissues were processed.

NGS Sequencing data collection: After successful QC, ca.10 mg tissue material is taken for nucleic acid extraction and protein lysate preparation each, followed by a second QC.

For DNA and RNA, tissues are homogenized using the BeadBug system and DNA and RNA are extracted using the Qiagen AllPrep Universal Kit. RNAs need to have a RIN >= 4 and a DV200 >= 60 to be selected for library preparation. Libraries for whole genome sequencing (WGS) are prepared using the KAPA Hyper Prep Kit (PCR-free). For whole transcriptome sequencing, the TruSeq Stranded Total RNA Kit is used. Sequencing is performed on a NovaSeq6000 system.

NGS Data Processing: NGS data was aligned against Grch38 genome assembly. Identification and annotation of short genomic variations in normal sample was done using Haplotype Caller (genome analysis toolkit; GATK) [2]. WGS somatic variation were called using a consensus of Mutect2 [3], Strelka [4], Varscan [5] and Somatic Sniper [6] . RNA-Seq differential expression was based on normalized readcount data (TPM: transcripts per million).

Mass spectrometry (whole proteome) data collection: For whole proteome profiling, 5-10 mg of fresh-frozen tissue was lysed in 2 mL Precellys® CK14 tubes containing 1.4 mm ceramic beads and using a lysis buffer containing PhosSTOP™ and bead shaking using a Precellys® Evolution Homogenizer equipped with a Cryolys® cooling module. After overnight digest samples were acidified and subjected to peptide desalting using Waters HLB Oasis 30 mg 96-well plates. Peptides were desalted using Waters µElution plates, dried down and resolubilized.

For DIA LC-MS/MS measurements, 5 µg of peptides per sample were injected to a reversed phase column (nanoEase M/Z Peptide CSH C18 Column, 1.7 µm, 300 µm X 150 mm) on a Waters ACQUITY UPLC M-Class LC connected to a Thermo Scientific™ Orbitrap Q Exactive™ HF-X mass spectrometer equipped with an EASYspray source. The nonlinear LC gradient was 1 - 60 % solvent B in 45 minutes at 50°C and a flow rate of 5 µL/min. The DIA method consisting of one full range MS1 scan and 50 DIA segments was adapted from Bruderer et al. [7].

Tissue-specific spectral libraries were generated combining high-fractionated DDA and DIA measurements on a pool of tissue material and raw data processed using software DIA-NN version 1.8.1.

Bioinformatical analyses: Metrices to define chromosomal instability were determined using R package CINmetrics [8]. Aneuploidy was measured using ASCETS [9].

RESULTS – Characteristics of CMS subtypes

Our analysis revealed known distinct molecular characteristics across the four CMS subtypes, including

- unique driver mutations,
- signaling cascades, and
- immune cell infiltration and cancer associated fibroblast (CAF) profiles.
- The poor prognosis associated with CMS4 patients was corroborated by our clinical data.

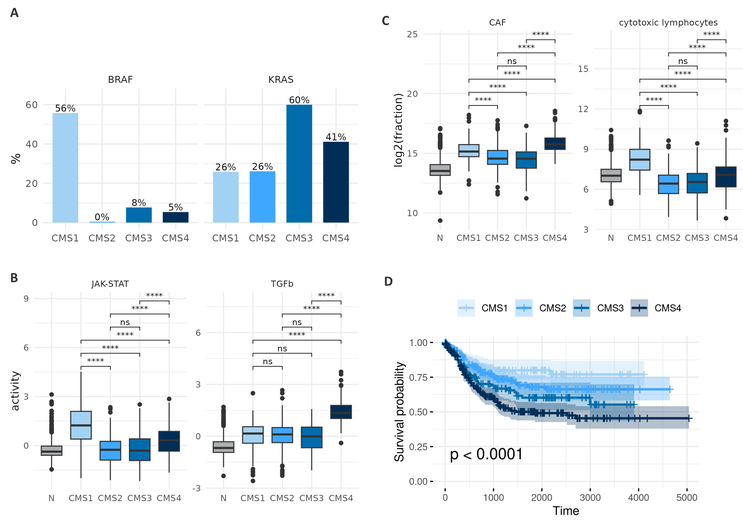

Figure 2: CRC CMS characteristics. **A:** unique driver mutations. BRAF V600E mutation is most frequent in CMS1, whereas KRAS mutations are most frequent in CMS3. **B:** JAK-STAT is driving immune evasion in CMS1 whereas TGFβ in CMS4 mainly affects epithelial to mesenchymal transition. **C:** Whereas CMS1 is characterized by high immune infiltration, in particular high abundance of cytotoxic lymphocytes, CMS4 can be classified by high density of stromal cells, e.g. cancer associated fibroblasts (CAF). D: Progression free survival. CMS1 is represented by good prognosis (but poor after recurrence), but CMS4 has the worst prognosis. CMS2 and CMS3 lie between CMS1 and CMS4.

REFERENCES

1. Eide, P. W. et al.: *CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models*. Sci Rep 7, 16618 (2017).
2. McKenna, A. et al.: *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data*. Genome Res. 2010, 20(9):1297-1303. PMID: 20644199
3. van der Auwera, G. et al.: *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, 2020, Incorporated. ISBN: 9781491975190
4. Kim, S., et al.: *Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods*, 2018. **15**(8): p. 591-594.
5. Koboldt, D. C., et al.: *VarScan: variant detection in massively parallel sequencing of individual and pooled samples*. Bioinformatics 2009. **25**(17): p. 2283-5.
6. Larson, D. E. et al.: *SomaticSniper: identification of somatic point mutations in whole genome sequencing data*. Bioinformatics, 2012, 28(3):311-317
7. Bruderer, R. et al.: *Analysis of 1508 plasma samples by capillary flow data-independent acquisition profiles proteomics of weight loss and maintenance*. MCP 18 (6), 1242-1254 (2019)
8. Oza, V. H. et al.: *CINmetrics: an R package for analyzing copy number aberrations as a measure of chromosomal instability*. Peer J, 2023. 11:e15244.
9. Spurr L. F., et al.: *Quantification of aneuploidy in targeted sequencing data using ASCETS*. Bioinformatics (2020)

RESULTS – Survival associated patterns of CIN and aneuploidy in CMS2 and CMS4

We observed high global chromosomal instability (CIN) profiles for both CMS2 and CMS4 and identified recurrent global aneuploidy patterns within these subtypes that link alterations of known cancer genes to these large-scale structural events, for example SMAD4 and CLC11.

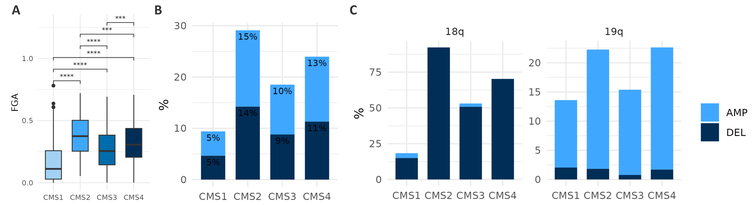

Figure 3: Chromosomal instability and aneuploidy in CMS subtypes. **A:** We found higher levels of CIN in CMS2 compared to CMS4, which appears to be predominantly driven by numerical, rather than structural CIN (only numerical CIN shown; FGA: fraction genome altered). **B:** Aneuploidy events, which span more than 90% of the chromosome arm are most frequent in CMS2 and CMS4. **C:** Similar patterns can be identified for chromosome arms 18q and 19q, where 18q is mostly driven by deletions and 19q by amplifications-

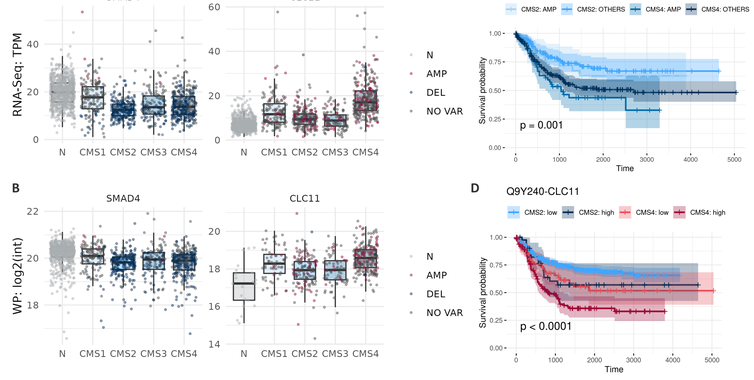

Figure 4: In CMS2, SMAD4 is commonly impacted by deletion on chromosome arm 18q, while CLC11 is more influenced by amplifications on 19q respectively, suggesting potential values as CRC CMS biomarkers. **A** and **B**: SMAD4 (18q) and CLC11 (19q) gene expression and protein intensities in CMS groups. CMS2 and 4 are characterized by more deletions on chromosome arm 18q, than CMS1 or CMS3, which is reflected by lower SMAD4 gene expression and protein intensities in CMS2/4 (blue dots). For CLC11 higher gene expression can be seen in CMS4 patients, specifically with chromosome arm amplifications on 19q (red dots). **C** and **D**: Progression free survival is impacted by amplifications on chromosome arm 19q. For both groups, CMS2 and CMS4, patients with amplifications on 19q show worse survival; for CMS4 worse than CMS2 (**C**). Specifically, survival is worse in patients with high expression of CLC11 (on 19q) (**D**).

CONCLUSION

Our integrative multi-omics analysis offers a comprehensive understanding of CRC molecular subtypes, which can inform personalized treatment strategies. Our findings underscore the importance of considering not only specific driver mutations but also large-scale structural rearrangements in the context of CMS subtypes, which can predict clinical outcome and help shaping the future of precision medicine.