Multiregional transcriptomics identifies congruent consensus molecular subtypes with prognostic value beyond tumor heterogeneity in colorectal cancer.

Anita Sveen
anita.sveen@rr-research.no

Oslo University Hospital  https://orcid.org/0000-0001-8219-6251

Jonas Langerud
Oslo University Hospital

Ina Eilertsen
Oslo University Hospital

Seyed Moosavi
Oslo University Hospital/Institute for Cancer Research

Solveig Klokkerud
Oslo University Hospital

Ingeborg Backe
Oslo University Hospital

Merete Hektoen
Oslo University Hospital

Ole Sjo
Oslo University Hospital

Marine Jeanmougin
Oslo University Hospital

Sabine Tejpar
UZ Leuven

Arild Nesbakken
Oslo University Hospital

Ragnhild Lothe
Oslo University Hospital  https://orcid.org/0000-0002-1693-1032

Article
Keywords: 

Posted Date: September 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3290125/v1

License: ☑️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Yes there is potential Competing Interest. A.N., R.A.L. and A.S. are co-inventors of a patent application regarding the use of HSP90 inhibitors in relation to the consensus molecular subtypes of colorectal cancer (PCT/IB2018/000042). S.H.M., R.A.L. and A.S. are co-inventors of a patent application describing transcriptomic liver metastasis subtypes (LMS) of colorectal cancers (Attorney Docket No. INVEN-39613.101). Authors declare that they have no other competing interests.

Version of Record: A version of this preprint was published at Nature Communications on May 21st, 2024. See the published version at https://doi.org/10.1038/s41467-024-48706-2.
Abstract

Intra-tumor heterogeneity compromises the clinical value of transcriptomic classifications of colorectal cancer. We investigated the prognostic effect of transcriptomic heterogeneity and the potential for a classification less vulnerable to heterogeneity in a single-hospital series of 1,093 primary and metastatic tumor samples from 692 patients, including 2–4 multiregional samples from 98 primary tumors and primary-metastasis sets from 35 patients. Intra-tumor heterogeneity of the consensus molecular subtypes (CMS) was frequent (40%) and associated with poor patient survival independently of tumor microenvironment markers. Multiregional transcriptomics uncovered cancer cell-intrinsic and low-heterogeneity signals that recapitulated the two intrinsic subtypes (iCMS2/iCMS3) proposed by single-cell sequencing. Further subclassification resulted in four congruent CMSs defining good-prognostic and poor-prognostic subtypes. Congruent CMS explained a larger proportion of variation in patient survival than intra-tumor CMS heterogeneity. Evidence of plasticity was found by discordant phenotypes of matched primary and metastatic tumors (28%), even according to the two-state intrinsic classification. In conclusion, multiregional sampling reconciled the prognostic power of tumor classifications from single-cell and bulk transcriptomics in the context of intra-tumor heterogeneity, and phenotypic plasticity challenges the reconciliation of primary and metastatic subtypes.

Introduction

Tumor heterogeneity is a main cause of cancer progression and treatment failure\(^1\)\(^,\)\(^2\). Most solid tumors consist of multiple subclones with different genomic profiles, metastatic potentials and responses to treatment. Colorectal cancers (CRCs) commonly have polyclonal invasion, and genomic heterogeneity of the primary tumor is associated with frequent metastasis and a poor patient survival\(^3\)\(^,\)\(^4\). However, not all tumor subclones have an impact on cancer evolution\(^5\). Clonal selection operates on cellular phenotypes, not genotypes, and heterogeneity appears to be decoupled at the genomic and transcriptomic levels in CRC\(^6\). In fact, it has been proposed that most genomic intra-tumor variation of CRCs has no major phenotypic consequences\(^6\). This emphasizes the importance of phenotypic plasticity\(^7\), but the clinical relevance of intra-tumor heterogeneity is less well studied on the transcriptomic level in CRC.

CRC transcriptomes represent a collection of the four consensus molecular subtypes (CMS)\(^8\). This classification reflects tumor phenotypes and morphologies, and is associated with patient survival and drug sensitivities\(^9\)\(^–\)\(^11\). It is generally accepted that the CMS framework provides a useful starting point for further transcriptomic investigations of primary CRCs. However, the classification is vulnerable to intra-tumor heterogeneity, possibly to the point where individual tumors contain a mixture of all CMS classes at different proportions\(^10\)\(^,\)\(^12\). This compromises the biomarker value and the predictive power of CMS for clinical endpoints\(^13\). Single-cell RNA sequencing has illustrated that the diverse cell types of the tumor microenvironment contribute strongly to “bulk” tumor transcriptomes, as well as to intra-tumor heterogeneity and the definition of tumor subtypes\(^14\)\(^–\)\(^16\). Indeed, both the classification accuracy and the prognostic value of CMS are confounded by the tumor microenvironment\(^12\)\(^,\)\(^13\)\(^,\)\(^17\). Cancer cell-intrinsic
expression signals are also shaped by the microenvironment\textsuperscript{15}, but might be less vulnerable to tumor heterogeneity\textsuperscript{18}. Additional classification frameworks such as the CRC intrinsic subtypes (CRIS) and the two intrinsic CMS (iCMS) classes adhere to this rationale\textsuperscript{19,20}, but the potentially added clinical value of a cancer cell-intrinsic approach has yet to be defined. In this respect, phenotypic plasticity during metastasis and metastatic heterogeneity of the classifications are likely to be relevant, as demonstrated with the original CMS\textsuperscript{21}.

Single-cell transcriptomics is a powerful technology in the mapping of tumor heterogeneity. However, the high costs and technical and biological variability associated with single-cell analyses are challenges that currently limit the application to larger tumor series and the integration of datasets\textsuperscript{22}. We hypothesized that bulk transcriptomics of distinct tumor regions is a complementary approach in this setting, and that multiregional sampling can balance the needs to capture both intra-tumor and inter-tumor variation. This has previously been used to illustrate intra-tumor heterogeneity and sampling bias in CRC, although in a limited number of tumors (up to 25)\textsuperscript{6,23,24}. We analyzed 1,093 multiregional or single samples of primary tumors and/or liver metastases from 692 patients to investigate the prognostic impact of intra-tumor heterogeneity and the potential for a transcriptomic classification less vulnerable to tumor heterogeneity.

Results

Transcriptomic intra-tumor heterogeneity among multiregional samples

To get an initial overview of transcriptomic intra-tumor heterogeneity in CRC, we compared CMS classifications among 2–4 multiregional samples from each of 98 primary tumors (n = 286 samples; Fig. 1A, Supplementary Table S1 and Supplementary Figure S1). Intra-tumor CMS heterogeneity was found in 40% (seven tumors were undetermined due to unclassified samples) and reflected general transcriptomic heterogeneity, estimated as the maximum Euclidean distance of principal components (PC) 1–3 between any pair of samples per tumor (Fig. 1B). The level of heterogeneity increased with the number of samples per tumor (p = 2×10\textsuperscript{-4} by Kruskal-Wallis test; Supplementary Figure S2). Both CMS3 and CMS4 were enriched with heterogeneous tumors (Fig. 1C). CMS4 was most heterogeneous and mixed with other subtypes in 84% (n = 26 of 31 tumors with at least one CMS4 sample). The most common CMS combinations were CMS2/4 (n = 19, 49% of heterogeneous tumors) and CMS1/3 (n = 8, 21%), with CMS4 and CMS3 as the minor components, respectively. Combinations of CMS3/4 (n = 3, 8%) and CMS1/2 (n = 2, 5%) were rare.

Microsatellite instability (MSI) status and KRAS/NRAS/BRAF\textsuperscript{V600E} mutation status were concordant among all multiregional samples from each tumor with CMS heterogeneity (Fig. 1A). MSI and BRAF\textsuperscript{V600E} mutations were strongly enriched among tumors with a major CMS1 component (MSI: 79%, odds ratio [OR] 55.6, 95% confidence interval [CI] 13.6–303.0, p = 9×10\textsuperscript{-13}; BRAF\textsuperscript{V600E}: 82%, OR 69.0, 95% CI 16.2-
394.4, \(p = 9 \times 10^{-14}\)), and not similarly frequent in CMS1-minor tumors (25% of the four tumors with < 50% CMS1 samples). KRAS/NRAS mutations were most frequent in CMS3 tumors (major or minor) without a CMS1 component (88% of the 16 tumors, OR 17.3, 95% CI 1.7-308.7, \(p = 0.005\)).

**Transcriptomic heterogeneity primarily driven by stromal infiltration**

Gene set enrichment analysis of a custom collection of gene sets relevant for CRC (\(n = 54\)) showed strong enrichment with mesenchymal-like and stromal features in tumors with heterogeneous compared to homogeneous CMS classifications (Fig. 1D and Supplementary Table S2). Results were similar in subgroup analyses of each of the CMS1-3 classes separately (Supplementary Figure S3 and Supplementary Table S3). Similar results were also found by enrichment testing of differentially expressed genes between homogeneous and heterogeneous tumors among biological processes in the Gene Ontology database (Supplementary Figure S4 and Supplementary Table S4). Sample-wise estimates of the abundance of cancer-associated fibroblasts were higher in heterogeneous tumors, but there was no difference in the abundance of cytotoxic lymphocytes (Supplementary Figure S5 and Supplementary Table S5). This highlighted the tumor stroma as a key component of intra-tumor transcriptomic heterogeneity, consistent with the frequent heterogeneity of CMS4.

In contrast, homogeneous tumors had strongest enrichment with signatures of cell cycle progression and regulation, as well as with MYC targets (Fig. 1D and Supplementary Figure S3). This was consistent with the large proportion of homogeneous tumors classified as CMS2 (48%). Notably, tumors homogeneous for CMS1 or CMS3 had no significant enrichments compared to heterogeneous tumors of the corresponding subtype, although signals were strongest for immune and metabolic processes, respectively (Supplementary Table S3). The unexpected subset of tumors homogeneous for CMS4 (\(n = 5\) microsatellite stable [MSS] tumors not exposed to treatment prior to sampling, two with \(BRAF^{V600E}\) or \(KRAS\) mutation) were enriched with signatures of extracellular matrix organization, the top of colonic crypts and inflammatory response (Supplementary Figure S3).

**Independent prognostic impact of intra-tumor CMS heterogeneity**

Intra-tumor CMS heterogeneity was not associated with any clinicopathological parameters (Supplementary Table S5) or 5-year relapse-free survival (RFS) in the multiregional sample set (\(p = 0.7\) from Cox proportional hazards analysis; \(p = 0.6\) from corresponding analysis of general transcriptomic heterogeneity as a continuous variable). To extend the analyses to a larger patient series, we used CMS as a framework for computational modeling of heterogeneity in single, bulk tissue samples from another 418 primary CRCs (Supplementary Table S1). Major and minor subtypes were assigned based on enrichment scores for template gene sets of each CMS class (Methods and Supplementary Figure S6). The major subtype of each tumor was largely concordant with assignments from the original random forest CMS classifier\(^8\), with an overall accuracy of 83% (Cohen's \(\kappa = 0.72\) and 0.75 for tumors analyzed on
Human Transcriptome 2.0 and Human Exon 1.0 ST arrays, respectively; Supplementary Figure S7). The majority (88%) of discordances were due to more frequent CMS2 classifications by the enrichment analyses. CMS heterogeneity was identified in 30% of the tumors, based on significant enrichments with more than one CMS class ($p < 0.05$ of singescore enrichments$^{25}$). This was less frequent than the heterogeneity observed among multiregional samples (OR 0.58, 95% CI 0.36–0.96, $p = 0.03$), and can likely be attributed to a combination of limited analytical discriminatory power and the indication that heterogeneity increased with the number of samples analyzed per tumor (Supplementary Figure S2B). The distribution of the most common CMS combinations was similar between the multiregional and single-sample tumor series, apart from a more frequent combination of CMS3 with CMS1 in favor of CMS2 among multiregional samples (Supplementary Figure S8), possibly related to the enrichment with MSI tumors in this series (Supplementary Table S1).

Analysis of the combined tumor series confirmed that intra-tumor CMS heterogeneity was associated with a high abundance of cancer-associated fibroblasts, but not with any clinicopathological parameter, except for frequent CMS heterogeneity among male patients (Supplementary Table S5). Survival analysis of patients treated by complete resection of stage I-III CRC and with determined CMS heterogeneity status ($n = 387$) showed a lower five-year RFS rate with heterogeneous (62.3%, 95% CI 54.2–71.5%) compared to homogeneous tumors (75.8%, 95% CI 70.7–81.2%; Fig. 2A). CMS heterogeneity retained prognostic value in a multivariable Cox proportional hazards model of all clinicopathological and molecular parameters, and was the only molecular marker with a significant prognostic association (hazard ratio, HR 1.5, 95% CI 1.0-2.2, $p = 0.05$; Supplementary Table S6). Notably, CMS heterogeneity explained a larger proportion of variation in five-year RFS (11%) than cancer-associated fibroblasts (5%; Fig. 2B and Supplementary Figure S9).

A stratified analysis of CMS heterogeneity according to the poor-prognostic CMS4 class (CMS4 versus CMS1-3) indicated that heterogeneous tumors with a CMS4 component (major or minor) were associated with the worst prognosis (Fig. 2C). Heterogeneous tumors without CMS4 (different combinations of CMS1-3) had non-significant associations to worse survival relative to homogenous tumors.

**Uniform intra-tumor activity of MSI-related and oncogenic processes**

The five CRIS classes derived from cancer cell-intrinsic expression signals$^{19}$ showed a similar frequency of intra-tumor heterogeneity among multiregional samples as CMS (43%, $p = 0.5$ from Fisher’s exact test compared to CMS; Supplementary Table S7), although there was no significant overlap of tumors with heterogeneity according to CMS and CRIS (OR 2.0, 95% CI 0.7–5.4, $p = 0.2$; Supplementary Figure S10). This indicated heterogeneity also within the epithelial cell compartment of CRCs and/or a stromal influence on the CRIS classification. To further investigate the basis for transcriptomic heterogeneity, we categorized protein-coding genes into three groups according to an intra-tumor heterogeneity score (ITH-score) representing intra-tumor relative to inter-tumor expression variation in the multiregional sample
The distribution of the ITH-scores was asymmetrical, with a heavy right-sided tail indicating a small subset of genes with high intra-tumor heterogeneity (ITH-high: 5% of genes). PC1 of tumor samples from principal components analysis (PCA) based on these ITH-high genes was most strongly correlated to single-sample enrichment scores of gene sets related to stromal and mesenchymal-like features (Fig. 3B). Similar gene set results were observed for PC1 of ITH-intermediate genes (48% of genes; Supplementary Figure S13), supporting that the majority of gene expression variation can be attributed to the stromal tumor component. In contrast, ITH-low genes (48%) were in a similar analysis associated with cancer cell-intrinsic features. PC1 of tumors based on ITH-low genes was strongly correlated to MSS/MSI-like signatures only, while PC2 was correlated to signatures of the cell cycle and proliferation (Fig. 3B). Notably, ITH-low genes showed less frequent gene set correlations with PC1 than PC2, while the opposite was observed for ITH-high and ITH-intermediate genes (OR 0.5, 95% CI 0.2–0.9, p = 0.02 comparing ITH-low and ITH-high genes; Fig. 3C). This suggested that genes with uniform expression across tumor regions (ITH-low) provided a more subtle tumor characterization based on the intrinsic features of cancer cells, compared to the dominating contribution from ITH-high genes and the tumor stroma.

Cancer-critical genes, defined by the Cancer Gene Census, were underrepresented in the ITH-high category (OR 0.3, 95% CI 0.03-1.0, p = 0.05; Supplementary Table S9). ITH-low cancer-critical genes were enriched in several pathways involved in CRC tumorigenesis, such as genomic instability (chromosomal and MSI), WNT signaling and the TP53 network (Fig. 3A and Supplementary Figure S14). ITH-high or ITH-intermediate cancer-critical genes showed no significant enrichments in a similar overrepresentation test of the “Wikipathway cancer” collection, suggesting that malignancy processes are not prone to intra-tumor heterogeneity on the transcriptomic level.

The ITH-scores were evaluated in a public single-cell RNA sequencing dataset of paired samples from the tumor core and border regions of six primary CRCs. This confirmed that ITH-high genes had a higher expression variation among cells from paired samples than ITH-low genes (p < 1×10^{-10} from Welch’s t-test; Supplementary Figure S15).

**Evolution of ITH-low subtypes in primary-metastasis comparisons**

ITH-low genes retained expression variation among tumors in the multiregional sample set, and had higher inter-tumor expression ranges (10-90th percentiles) than ITH-high (95% CI of the mean difference 0.56–0.59) or ITH-intermediate genes (95% CI 0.29–0.33; p < 1×10^{-15} from Welch's t-tests; Fig. 3A). To investigate the potential for transcriptomic classifications less prone to intra-tumor heterogeneity, we therefore performed subtype discovery by non-negative matrix factorization (NMF) of tumors based on the ITH-low genes. NMF across the full sample set (n = 704 samples from 516 primary tumors) at a predefined rank of k = 2 clusters resulted in subtypes (denoted k2) that were largely concordant with the
two iCMS classes previously derived from single-cell RNA sequencing of the malignant epithelial compartment of CRCs\(^2\) (classification accuracy 90\%, Cohen's \(\kappa = 0.80\); Fig. 4A). Subtype characteristics based on gene set enrichments were also highly similar between iCMS and k2, and both frameworks were primarily distinguished by MSI/MSS-like characteristics and immune signatures (Supplementary Figure S16). Both iCMS and k2 provided largely concordant intra-tumor classifications of multiregional primary tumor samples (82\% and 99\%, respectively; Fig. 4B). Collectively, this suggested that an average of three multiregional samples from each tumor could recapitulate the cancer cell-intrinsic subtypes from single-cell sequencing.

CRC liver metastases \((n = 304\) tumor samples from 179 patients) also showed concordant classifications between iCMS and the ITH-low k2 clusters (accuracy 83\%, Cohen's \(\kappa = 0.66\); Fig. 4A). The subtype distributions were similar among primary tumors and metastases in both frameworks (iCMS: \(p = 0.8\) and k2: \(p = 0.2\) from Pearson's chi-squared tests). Principal components analysis based on ITH-low genes or iCMS template genes showed no apparent distinctions according to tumor site (colorectum versus liver; Supplementary Figure S17), supporting that also ITH-low genes primarily have cancer cell-intrinsic expression and that both classifications are directly applicable to metastatic tumors. However, the frequency of intra-patient subtype heterogeneity of metastatic lesions \((n = 2–7\) lesions from each of 47 patients) was higher than intra-tumor heterogeneity of the primary tumor (iCMS: \(\chi^2 = 6.7, p = 0.008\) and k2: \(\chi^2 = 35.6, p = 3\times10^{-9}\), both with one degree of freedom; Fig. 4B). Furthermore, comparisons of patient-matched primary tumors and liver metastases \((n = 179\) samples from 35 patients) also showed evidence of phenotypic plasticity. Using iCMS for illustration, only 59\% of evaluable patients had fully concordant classifications (17 of 29 patients, six were not evaluable due to unclassified samples; Fig. 4C). Subtype switching between all or a majority of primary-metastasis samples was observed in eight patients (28\%). This occurred predominantly from iCMS2 primary tumors to iCMS3 liver metastases (six of eight patients, 75\%). There was no significant association between subtype switching and use of different analysis platforms (RNA sequencing versus Human Transcriptome 2.0 array: \(\chi^2 = 0.90, p = 0.6\), the numbers of samples/tumors per patient \((p = 0.5\) from Wilcoxon's test), diagnosis with synchronous versus metachronous metastases \((p > 0.9\) from Pearson's chi-squared test), or exposure to chemotherapy prior to sampling \((p = 0.9)\).

**Prognostic value of congruent CMS classification based on ITH-low genes**

Subtype discovery based on ITH-low genes was tested with different sets of samples and ITH-score thresholds to evaluate a possible impact on classification results (Methods). NMF at \(k = 4\) was identified as the best sample clustering across the complete primary tumor series (Supplementary Figures S18-S21), indicating potential for an intrinsic classification with a higher resolution than the two-state iCMS framework. The k4 clusters ranged in size from 10–53\% of samples and subdivided each of the k2 clusters, most prominently the cluster corresponding to iCMS3 (Fig. 5A-B; the iCMS framework was similarly split, Supplementary Figure S22). The iCMS3-corresponding cluster was split into one cluster with strong immune signals and one with high expression of genes encoding extracellular matrix...
remodeling proteins (FN1 and SPP1\textsuperscript{26}), while the largest and remaining cluster had high relative expression of genes involved in maintenance of the secretory intestinal stem cell niche (for example, REG4, TFF1, FCGBP and AGR2\textsuperscript{27-30}; analyzing ITH-low genes only; Supplementary Figure S23).

The k4 clusters were not independent of the original CMS ($X^2 = 589$, nine degrees of freedom, $p < 3 \times 10^{-16}$), and 67% of samples showed concordant classifications with CMS (classification accuracy 68%, Cohen's $\kappa = 0.52$; Fig. 5B). The strongest discordance was found for the cluster corresponding to the original CMS3, and this cluster was split between CMS1 and CMS3. Notably, samples with discordant classifications were located near the class boundaries in PCA (Supplementary Figure S24). Gene set enrichment analyses further demonstrated that each of the four sample clusters defined by ITH-low genes had similar characteristics to the corresponding CMS class (Supplementary Figure S25), and the k4 clusters were therefore termed congruent CMS (cCMS). Intra-tumor classification concordances of multiregional samples were higher for cCMS (77% of tumors) than for the original CMS (53%, OR 2.4, 95% CI 1.3–4.8) and CRIS frameworks (45%, OR 3.1, 95% CI 1.6–6.2; Fig. 4B), indicating stronger robustness to intra-tumor heterogeneity.

Genomic markers (MSI, $BRAF^{V600E}$) and tumor microenvironment markers (cancer-associated fibroblasts, cytotoxic lymphocytes) showed similar subtype associations in the cCMS and original CMS frameworks, with the exception that KRAS mutations were not skewed among cCMS classes (Supplementary Table S10). Consistent with the strong enrichment for MSI-like characteristics among ITH-low genes, MSI status was strongly skewed according to cCMS ($X^2 = 174$, three degrees of freedom, $p < 3 \times 10^{-16}$) and enriched in both cCMS1 and cCMS3 (OR 14.7, $p < 3 \times 10^{-16}$ and OR 2.9, $p = 4 \times 10^{-5}$, respectively; Fig. 5C). However, repeated subtype discovery of MSS tumors only (based on ITH-low genes; Supplementary Figure S18) largely recapitulated the cCMS classification (accuracy 90%, Cohen's $\kappa = 0.82$; Supplementary Figure S26), indicating that the transcriptomic MSI-like features of the ITH-low genes extended beyond the genomic phenotype.

Clinicopathological associations were also similar between the cCMS and original CMS frameworks, although patient age at diagnosis was skewed according to cCMS (older age with cCMS1 and younger with cCMS4: OR 4.5, $p = 1 \times 10^{-4}$; Supplementary Table S10). Survival analyses of patients with concordant intra-tumor classifications (no subtyping heterogeneity among multiregional samples) showed that cCMS had strong associations to five-year RFS and overall survival in stage I-III CRC ($n = 398$ patients; Fig. 5D and Supplementary Figure S27). Higher and lower RFS rates were observed with cCMS1 and cCMS4 tumors, respectively, relative to each of the other subtypes. These prognostic associations were consistent with cCMS1 consisting primarily of an immune-active subset of iCMS3 tumors, and cCMS4 of iCMS3 tumors (but also a proportion of iCMS2) with active extracellular matrix remodeling, which can promote immune suppression and metastasis\textsuperscript{15,26,31} (Supplementary Figures S22-S23). Notably, the original CMS classes had no prognostic value in this subset of patients (Supplementary Figure S28). The cCMS framework retained prognostic value when added to the multivariable survival model shown in Fig. 2B (Table 1 and Supplementary Table S11), and explained a
larger proportion of variation in five-year RFS (21%) than CMS heterogeneity (9%) and any other molecular variable (Fig. 5E).
Table 1
Multivariable survival analysis of clinicopathological and molecular features in patients with stage I-III CRC

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Five-year relapse-free survival&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>HR [95% CI]</td>
</tr>
<tr>
<td>Total</td>
<td>398 (100)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>204 (51)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>194 (49)</td>
<td>1.5 [1.0-2.2]</td>
</tr>
<tr>
<td>Age (continuous)</td>
<td>398 (100)</td>
<td>1.0 [1.0-1.1]</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>176 (44)</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>125 (31)</td>
<td>0.9 [0.6–1.5]</td>
</tr>
<tr>
<td>Rectum</td>
<td>97 (24)</td>
<td>1.0 [0.6–1.7]</td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>95 (24)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>176 (44)</td>
<td>1.2 [0.7–2.2]</td>
</tr>
<tr>
<td>III</td>
<td>127 (32)</td>
<td>2.1 [1.1–3.9]</td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>327 (82)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>64 (16)</td>
<td>1.2 [0.6–2.2]</td>
</tr>
<tr>
<td>Unknown</td>
<td>7 (2)</td>
<td>2.4 [0.7–8.1]</td>
</tr>
<tr>
<td>MSI status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSS</td>
<td>320 (80)</td>
<td></td>
</tr>
<tr>
<td>MSI</td>
<td>78 (20)</td>
<td>0.8 [0.3–1.7]</td>
</tr>
<tr>
<td>KRAS status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td>258 (65)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Excluding patients with synchronous tumors, heterogeneous cCMS classification, pre-surgical chemoradiation and residual tumor status 1 or 2. <sup>b</sup>Statistically significant p-values highlighted in bold. Results were similar in a stratified analysis according to the variable breaking the proportional hazard assumption (cox.zph p-value < 0.05; Supplementary Table S11). CAF, cancer-associated fibroblasts; CTL, cytotoxic lymphocytes; TNM, tumor-node-metastasis.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Five-year relapse-free survival&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>HR [95% CI] p-value</td>
</tr>
<tr>
<td>Mutation</td>
<td>140 (35)</td>
<td>1.1 [0.7–1.8] 0.6</td>
</tr>
<tr>
<td><strong>BRAF&lt;sup&gt;V600E&lt;/sup&gt;-status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td>328 (82)</td>
<td>reference</td>
</tr>
<tr>
<td>Mutation</td>
<td>70 (18)</td>
<td><strong>2.6 [1.3–5.3] 0.01</strong></td>
</tr>
<tr>
<td><strong>CTL-score (continuous)</strong></td>
<td>398 (100)</td>
<td>1.2 [0.5–3.2] 0.7</td>
</tr>
<tr>
<td><strong>CAF-score (continuous)</strong></td>
<td>398 (100)</td>
<td>1.1 [0.8–1.5] 0.7</td>
</tr>
<tr>
<td><strong>CMS heterogeneity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homogeneous</td>
<td>257 (65)</td>
<td>reference</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>115 (29)</td>
<td><strong>1.6 [1.1–2.4] 0.03</strong></td>
</tr>
<tr>
<td>Undetermined</td>
<td>26 (7)</td>
<td><strong>2.2 [1.0–4.7] 0.05</strong></td>
</tr>
<tr>
<td><strong>cCMS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cCMS1</td>
<td>70 (18)</td>
<td>reference</td>
</tr>
<tr>
<td>cCMS2</td>
<td>224 (55)</td>
<td><strong>3.7 [1.6–8.7] 0.002</strong></td>
</tr>
<tr>
<td>cCMS3</td>
<td>76 (19)</td>
<td><strong>2.3 [1.0–5.2] 0.04</strong></td>
</tr>
<tr>
<td>cCMS4</td>
<td>28 (7)</td>
<td><strong>4.3 [1.6–11.3] 0.003</strong></td>
</tr>
</tbody>
</table>

<sup>a</sup>Excluding patients with synchronous tumors, heterogeneous cCMS classification, pre-surgical chemoradiation and residual tumor status 1 or 2. <sup>b</sup>Statistically significant p-values highlighted in bold. Results were similar in a stratified analysis according to the variable breaking the proportional hazard assumption (cox.zph p-value < 0.05; Supplementary Table S11). CAF, cancer-associated fibroblasts; CTL, cytotoxic lymphocytes; TNM, tumor-node-metastasis.

**Discussion**

Multiregional tumor transcriptomics represents a feasible approach to balance the needs to capture both intra-tumor and inter-tumor gene expression variation. We report the largest published CRC series with multiregional transcriptomics, and used this to distinguish heterogeneous and uniform expression features across tumor regions, while retaining information of tumor subtypes (that is, variation across tumors). Three bulk samples per tumor could recapitulate cancer cell-intrinsic expression patterns and subtypes that were less vulnerable to intra-tumor heterogeneity. While single-cell RNA sequencing was needed to initially delineate these patterns and define the iCMS classification<sup>20</sup>, this study showed potential to expand on the knowledge and suggested a further substratification into four intrinsic subtypes. This resulted in a split predominantly of the subtype corresponding to iCMS3. However, the
split was not primarily defined by MSI status, as proposed with the refined IMF (intrinsic-microsatellite-fibrosis) classification. The four intrinsic subtypes rather converged on having the same discriminatory biological features as the original CMS, although it has previously been shown that the original CMS classifier is depleted of genes with uniform expression among tumor glands. Nonetheless, the convergence is consistent with the assumption that the tumor microenvironment is at least partly shaped by malignant epithelial cells and that the tumor epithelium can recapitulate the CMS classification. This was also the premise for the successful classification of diverse pre-clinical model systems according to CMS. Overall, this supports that CMS-related features provide a bona fide phenotypic stratification of CRCs, but the precise cellular interactions defining the subtypes with a rich microenvironment component are still to be uncovered. Spatial transcriptomics has potential to delineate such interactions, as recently shown with the detailed description of the interaction networks of immune and malignant cells according to tumor MSI status. In this context, the four congruent subtypes proposed in this study can be considered as an alternative CMS classification that is based on cancer cell-intrinsic template genes and therefore less vulnerable to intra-tumor heterogeneity. However, this interpretation does not fully account for the stronger prognostic power of the congruent subtypes.

In contrast to the original CMS classification, the congruent CMS classes provided substantial prognostic value beyond both intra-tumor heterogeneity and the tumor microenvironment components in patients with non-metastatic cancer. However, the two prognostic subtypes (cCMS1 and cCMS4) constituted only one-fourth of the tumors in total. Both of the prognostic subtypes were dominated by tumors corresponding to iCMS3, but included only approximately half of all iCMS3 tumors. This is largely consistent with the original publication showing that the binary iCMS classification is not prognostic. A poor patient survival was found with fibrotic iCMS3 tumors only, and this subtype constituted approximately 30% of iCMS3 tumors and 14% in total. Notably, the proposed cCMS classification additionally identified a subset of mostly iCMS3 tumors with a favorable prognostic association, independently of MSI status. This further supports that substratification of iCMS is needed in the evaluation of patient prognosis, and the proposed cCMS might reconcile the single cell-derived iCMS and the original bulk transcriptomics-derived CMS for this purpose. Application of cCMS to additional tumors is not dependent on multiregional sampling and can be done based on the ITH-low genes. This facilitates additional translational studies, which are needed to support the proposed prognostic value and to explore additional clinical relevance, for example, by associations with drug sensitivities.

Our work also showed prognostic relevance of intra-tumor heterogeneity. These results are highly similar to a previous study based on computational deconvolution of intra-tumor CMS heterogeneity in single samples of stage III colon cancers, and supported a poor prognosis with a minor CMS4 component in particular. Notably, the CMS combinations frequently observed by multiregional sampling, or estimated by computational enrichments, were similar to results from single-cell sequencing of a smaller tumor series. Nonetheless, tumors analyzed by the largest number of multiregional samples were frequently scored as heterogeneous, and it is likely that CMS heterogeneity is underestimated in studies based on bulk transcriptomics. Even small tumor subclones can have clinical relevance with respect to
development of resistance during treatment\textsuperscript{36}, but we cannot conclude on the lower limit of what can be considered prognostically relevant transcriptomic heterogeneity, or on the number of samples needed to detect this. According to the “big bang” model of CRC development, invasive cancers have spatially intermixed subclones on the genomic level\textsuperscript{37}. This supports the potential to capture heterogeneity with a small number of samples, although a potential caveat is that such clonal intermixing is not necessarily reflected on the transcriptomic and phenotypic levels.

Transcriptomic subtypes based on cancer cell-intrinsic signals have the presumed advantage of being applicable to both primary and metastatic tumors, without the need to adapt the classification approach. This was supported by PCA based on the ITH-low genes detected in this study and on the iCMS template genes, both showing intermingling of primary tumors and liver metastasis, which is in contrast to results based on unselected genes\textsuperscript{38}. In further support of the appropriateness of intrinsic classifications for metastatic tumors, we did not observe any subtype depletions or shift in the distribution of iCMS classes between primary tumors and liver metastases. This was unexpected based on the strong depletion of the original CMS1 and CMS3 classes among metastases\textsuperscript{21}, which would suggest a depletion also of iCMS3. Nonetheless, subtype switching of iCMS between matched primary and metastatic tumors was observed in almost a third of patients. This was particularly remarkable since iCMS is only a two-state classification. Switches of cancer cell-intrinsic classes can be due to either clonal evolution or transition of differentiation states. Clonal evolution and selection of the minor clone is a possible explanation based on the non-exclusivity of iCMS classes among cells in each primary tumor\textsuperscript{20}. However, phenotypic plasticity and cellular differentiation and dedifferentiation might be an essential trait for cancer metastasis\textsuperscript{39}. The dynamic cellular states observed in models of CRC metastasis\textsuperscript{40} open up the possibility for cells to even transition between iCMS classes during metastasis and to eventually end up in their original iCMS in established metastatic tumors. According to this view, the heterogeneity is dependent on the timing of sampling and would therefore be underestimated in our study. The most frequently observed switch from iCMS2 primary tumors to iCMS3 liver metastases is consistent with dedifferentiation from an LGR5-positive stem cell\textsuperscript{20}, although our study was not sufficiently powered to confirm this predilection. Nonetheless, the profound phenotypic plasticity observed in at least a subset of patients challenges the potential reconciliation of subtyping schemes of primary and metastatic tumors, also of the congruent CMS proposed here. This supports the need for a de novo classification of metastases based on their in situ cellular states\textsuperscript{38}.

In conclusion, we describe transcriptomic features with prognostic value independently of the tumor microenvironment and in the context of intra-tumor heterogeneity of CRC. Multiregional transcriptomics captured cancer cell-intrinsic features with low intra-tumor heterogeneity, and identified congruent CMS classes that appeared to reconcile the prognostic potential of current classifications derived from single cell and bulk transcriptomics. However, evidence of phenotypic plasticity during metastasis, even with a two-state cancer cell-intrinsic classification, indicated that reconciliation of primary and metastatic subtyping frameworks is challenging.
Methods

Patient material

A total of 1,093 fresh frozen primary tumor and liver metastasis samples from 692 patients treated surgically for primary and/or metastatic CRC at Oslo University Hospital were analyzed for gene expression in the study. Two to four multiregional samples (mean of 2.9) were taken from spatially distinct areas of each of 98 primary tumors from 96 patients treated in 2015 and 2016 ($n = 286$ samples; Supplementary Table S1). Multiregional samples were not taken unless clearly spatially separated. There was no association between tumor size and the number of samples from each tumor ($p = 0.4$ by Kruskal-Wallis test; Supplementary Figure S2A). RNA and DNA were extracted using the Qiagen AllPrep DNA/RNA/miRNA Universal Kit or DNA/RNA Mini Kit in accordance with the manufacturer’s protocol (Qiagen GmbH, Hilden, Germany). Molecular data from single primary tumor samples of an additional 418 patients treated between 2005 and 2013 have previously been published (Supplementary Table S1). Liver metastasis samples ($n = 338$) were collected from 191 patients treated by hepatic resection between 2013 and 2018, and molecular data have previously been published for the majority ($n = 280$ samples from 1–7 liver lesions of each of 171 patients)\(^3\). Patient-matched sets of primary tumor and metastasis samples were available from 35 patients (total $n = 179$ samples). The primary tumor from 22 of these patients ($n = 51$ samples) were included for longitudinal comparisons only and not otherwise analyzed in the study. Twenty-one (60\%) of the patients with primary-metastasis samples had synchronous metastatic disease (liver metastases diagnosed within six months of the primary tumor), and 14 (40\%) had metachronous metastases. Eighteen (51\%) received neoadjuvant chemotherapy for the sampled metastases, eight (23\%) had previously received chemotherapy for primary and/or metastatic CRC, and nine (26\%) were chemo naive at the time of sampling.

All patients were treated according to national standard guidelines. The study has been approved by the Regional Committee for Medical and Health Research Ethics, South Eastern Norway (REC numbers 1.2005.1629;2010/1805). All patients provided written informed consent, and the study was conducted in accordance with the Declaration of Helsinki.

Processed single-cell RNA sequencing data and metadata for totally 17,678 cells from twelve paired samples of the tumor core and tumor border regions of each of six primary CRCs were downloaded from NCBI’s Gene Expression Omnibus (GEO) with accession number GSE144735\(^1\) and used for validation analyses.

MSI and mutation analyses

MSI status of the multiregional primary tumor series was determined by PCR-based analyses of mononucleotide repeat markers using the Promega MSI Analysis System in accordance with the manufacturer’s protocol (Promega, Madison, WI, USA). Sequencing of mutational hotspots in \textit{KRAS} and \textit{NRAS} exons 2–4, as well as \textit{BRAF} exon 15 (including codon 600) was performed as previously
described\textsuperscript{41}. One randomly selected sample per tumor and all samples from tumors with discordant CMS classifications were analyzed (n = 158 samples).

Tumor content has been confirmed in the multiregional samples by deep sequencing of a custom panel of twenty genes, using matched normal colonic mucosa samples as reference. Homogenous somatic single nucleotide variants or short insertion or deletions in \textit{APC}, \textit{TP53}, \textit{KRAS}, \textit{NRAS}, \textit{BRAF}, \textit{PIK3CA} and/or \textit{FBXW7} (same mutation present in all samples per tumor) were found in all tumors except one that was not scored, all with a mutant allele fraction above 5% (the data and additional details will be published elsewhere).

\section*{Gene expression profiling and data processing}

All in-house tumor samples have been analyzed for gene expression on high-resolution platforms (n = 1,093; Supplementary Figure S1). Multiregional primary tumor samples were analyzed on Affymetrix Human Transcriptome 2.0 arrays (HTA), using 100 ng of total RNA as input and following the manufacturer’s protocol (Thermo Fisher Scientific, Waltham, MA, USA). The extended single-sample primary tumor set has previously been analyzed on HTA (n = 217) or Affymetrix GeneChip Human Exon 1.0 ST arrays (HuEx; n = 201)\textsuperscript{9}. Patient-matched primary-metastasis samples were analyzed on HTA (n = 23 patients and 116 samples) or by total RNA sequencing (n = 12 patients and 63 samples). The remaining liver metastases samples have been analyzed on HTA arrays\textsuperscript{38}. RNA sequencing was performed in 2×101 base-pair paired-end mode on the Illumina HiSeq 4000 platform (Illumina, San Diego, CA, USA) at the Oslo University Hospital Genomics Core Facility to a median depth of 52.6×10\textsuperscript{6} uniquely mapped read pairs per sample (10-90th percentile 40.5×10\textsuperscript{6}-71.6×10\textsuperscript{6}). Sample preparation was performed with ribosomal RNA depletion using the Ribo-Zero Gold rRNA removal kit and sequence library generation with the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina).

Raw intensity data CEL-files from microarray experiments were processed in five separate datasets (multiregional primary tumor samples, primary single-sample HTA, primary single-sample HuEx, all liver metastasis samples, all patient-matched primary-metastasis samples; Supplementary Figure S1) according to the robust multi-array average approach\textsuperscript{42} using the function justRMA in the R package affy (v1.64.0)\textsuperscript{43} and custom CDF files from Brainarray (hta20hsgencodergcdf_23.0.0 and huex10sthsgencodergcdf_23.0.0). A batch effect from different lot numbers of the GeneChip™ WT Plus Reagent Kit was corrected among multiregional primary tumor samples with ComBat in the R package sva (v.3.36.0)\textsuperscript{44} using default parameters. Gene annotations were retrieved from GENCODE using the gencode.v29.annotation.gtf file. Only protein-coding genes were retained and genes on the Y chromosome were excluded. Entrez IDs were obtained using the R package org.Hs.eg.db (v.3.10.0) and gene symbols were updated with checkGeneSymbols in HGNChelper (v.0.8.1).

Raw RNA sequencing reads were processed in a bioinformatics pipeline implemented with Snakemake (v.6.6.1) and using Python (v.3.9.5), Java (v.11.0.2) and PyYAML (v.5.4.1). The pipeline has previously been described and included adapter trimming with Trimmomatic (v.0.38), read alignment to the human
reference genome GRCh38.p13 (v.41) using STAR (v.2.7.6a) with 2-pass mapping and the feature annotation file gencode.v41.annotation.gtf, quantification of reads mapping to protein-coding genes using the HTseq-count tool (v.2.0.2), and normalization of gene expression levels by estimation of transcripts per million (TPM) for non-overlapping exonic gene lengths\(^{45}\). The TPM values were log2-transformed after adding a pseudocount of 1.

**Gene expression classification and enrichment analyses**

Tumor samples were classified according to CMS with the R package CMSclassifier (v.1.0.0)\(^8\) and using the function classifyCMS.RF with a custom posterior probability threshold of 0.4. The threshold was adjusted in the multiregional primary tumor set to lower the number of unclassified samples while retaining proportionality in the number of tumors with homogeneous and heterogeneous CMS classifications (Supplementary Figure S29). CRIS classifications were assigned with the function cris_classifier in the R package CRISclassifier (v.1.0.0)\(^{19}\), using the inverse of log2-transformed gene expression data and default parameters (false discovery rate [FDR] < 0.2). iCMS classification was performed using the approach and gene template described in the original publication\(^{20}\). Gene expression matrices on log2-scale were normalized to z-scores using ematAdjust and classified with the nearest template prediction approach using the function ntp and an FDR-threshold < 0.05 in the R package CMScaller (v.2.0.1)\(^{32}\).

Differential gene expression analyses were performed with limma as implemented in CMScaller and with \(p\)-value adjustment by the Benjamini-Hochberg procedure. Tumor-infiltrating cancer-associated fibroblasts and cytotoxic lymphocytes were estimated using the R package MCPcounter (v.1.2.0)\(^{46}\) on a combined and batch corrected gene expression dataset of all primary tumor samples \((n = 704)\). Gene set enrichment analyses were performed with the R package topGO (v.2.38.1) using fisher statistics and the weight01 algorithm, as well as with the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt)\(^{47}\) using default settings, over-representation analysis in the Wikipathway cancer database, FDR < 0.05 and the complete list of protein-coding genes as reference. Sample group comparisons were performed with the subCamera function in CMScaller on a custom gene set collection relevant for CRC \((n = 54;\) Supplementary Table S2) and with FDR adjustment of \(p\)-values according to the Benjamini-Hochberg procedure. One random sample from each tumor in the multiregional sample set was selected for comparisons according to CMS heterogeneity (the analysis was repeated across all multiregional samples and showed highly similar results; Supplementary Table S2).

**Intra-tumor transcriptomic heterogeneity**

For tumors with multiregional samples, intra-tumor heterogeneity was evaluated as discordant sample classifications within subtyping frameworks (the subtype representing \(\geq 50\%\) of samples per tumor was considered the major component) and by general transcriptomic heterogeneity. The latter was estimated as the maximum Euclidean distance of PC1-PC3 for any pair of samples from each tumor.
For primary tumors with single samples, intra-tumor CMS heterogeneity was estimated based on enrichment scores for each CMS class (Supplementary Figure S6 shows an overview of the approach). The single-sample HTA and HuEx datasets were analyzed separately. First, gene sets for each of the four CMS classes were identified by differential gene expression analyses comparing tumors in each class with the rest using limma (Benjamini-Hochberg adjusted p-value < 0.001 and log2 fold change > |1.0|; Supplementary Tables S12 and S13). Second, enrichment scores for each CMS-specific gene set in each sample were obtained using the gsva function in the R package GSVA (v.1.34.0)\textsuperscript{48} for up-regulated genes only, and with the functions simpleScore and rankGenes in the R package singscore (v.1.6.0)\textsuperscript{25} for up- and down-regulated genes combined. The CMS enrichment scores were evaluated in a similar analysis of the multiregional sample set, and the strongest correlations to the posterior probabilities from the original random forest CMS classifier were found for the singscore enrichments (Spearman’s p > 0.8; Supplementary Figure S30). Singscore also provided functions to evaluate statistical significance (generateNull and getPvals) and was selected for further analyses. Single-sample tumors were considered unclassified if none of the four CMS enrichment scores were significant, and classified with intra-tumor CMS heterogeneity if more than one was significant (p < 0.05).

**Gene-wise intra-tumor heterogeneity**

Intra-tumor heterogeneity of the expression level of each protein-coding gene (n = 18,823) was estimated in the multiregional sample set using a previously published method\textsuperscript{49}. In brief, a linear mixed effects model was fitted for each gene across all samples from all tumors using the function lmer in the R package lme4 (v.1.1–29)\textsuperscript{50} and with “tumor” as the random effect. Intra-class correlation coefficients (ICCs) were calculated for each model (gene) using the function icc in the R package performance (v.0.10.4)\textsuperscript{51}:

$$\text{ICC} = \frac{\sigma^2_i}{\sigma^2_i + \sigma^2_c}$$

Here, $\sigma^2_i$ is the random effects variance, that is, the variance explained by the grouping structure (tumor) and $\sigma^2_c$ is the residual variance. An ITH-score for each gene was calculated as:

$$ITH_{gene} = 1 - ICC_{gene}$$

Genes with low expression variation across the data set (10-90th percentile range < 1; n = 15,585 genes) were considered non-informative and filtered out, retaining 3,238 genes (17.2\%) for analyses (Supplementary Figure S11 and Supplementary Table S8). Genes were categorized according to the ITH-score using the previously published thresholds in four categories\textsuperscript{49}, or custom thresholds in the three categories ITH-low, ITH-intermediate and ITH-high (Supplementary Figure S12 and Supplementary Table S9). The two different thresholds to score ITH-low genes were compared in gene set enrichment analyses and showed largely concordant results (Supplementary Figure S31). The custom threshold retained the largest number of ITH-low genes and was used for further analyses.
Tumor classification based on ITH-low genes

Subtype discovery based on ITH-low genes was performed by the NMF approach implemented in the R package NMF\textsuperscript{52} using the brunet method\textsuperscript{53}, predefined ranks 2–10 and nrun = 100 on the inverse of log2-transformed gene expression data. The cluster number (k) preceding the first, large drop in the silhouette width and cophenetic score was selected as the optimal number of clusters. To evaluate a potential impact of the use of different gene expression platforms and the inclusion of multiregional samples for a subset of tumors, NMF was run both for the complete primary tumor sample set (n = 704 samples from 516 tumors) and for single, randomly selected samples from each of the primary tumors analyzed on HTA (n = 315). This resulted in k = 5 and k = 4 optimal sample clusters, respectively (Supplementary Figure S18). There was a near perfect concordance in sample clustering between the two runs at k = 4 (considering overlapping samples between the two sets only; accuracy 97%, Cohen's κ = 0.96; Supplementary Figure S19A). In the full sample set, the largest sample cluster from k = 4 was subdivided into two clusters at k = 5 (Supplementary Figure S19B), but gene set enrichment analyses showed little discrimination between the two clusters (Supplementary Figure S20). The full tumor series at k = 4 was therefore used for further analyses, to strengthen the biological and statistical rigor. NMF classification was also tested using ITH-low genes defined by the previously published scoring threshold as a template (ITH-score 0-0.2; n = 396 genes)\textsuperscript{49}. This resulted in only two sample clusters differentiated mainly based on MSI/MSS-like gene expression characteristics (Supplementary Figure S21). Classification of liver metastases by NMF was also based on genes identified as ITH-low in primary tumors. Alluvial diagrams were plotted using the R package ggalluvial (v.0.12.4).

Statistical analyses

All statistical analyses were performed in R v.4.2.2\textsuperscript{54}. Principal components analysis was performed with the prcomp function in the package stats (v.4.2.2) based on the genes with highest cross-sample variance (n = 1,000). Pearson's correlations were calculated and visualized using the functions cor, cor.mtest and corrplot in the R package corrplot (v.0.92), and with conf.level = 0.95. Odds ratios and 95% CIs were estimated with Fishers' Exact Test (fisher.test), and were together with Pearson's chi-squared test (chisq.test) and Welch's two sample t-test (t.test) used to evaluate associations between clinicopathological parameters and sample groups according to molecular characteristics. Cohen's κ was estimated with the function confusionMatrix in the package caret (v.6.0–93). Two-sided p-values, or adjusted p-values as specified, below 0.05 were considered significant.

Survival analyses were performed for patients with stage I-III CRC treated by complete tumor resection (residual tumor status R0) and with no pre-surgical chemoradiation or synchronous tumors. Five-year RFS was the primary endpoint and estimated as time from surgery to relapse or death from any cause. Patients with no events were censored after five years or at last follow-up. Overall survival was evaluated as the time from surgery to death from any cause. Multivariable and univariable Cox proportional hazards models were estimated using the coxph function in the survival package (v.3.4-0) with p-values from Wald tests. The proportional hazards assumption was assessed for all models using the cox.zph function in the survival package (v.3.4-0).
function, and all variables met the assumption, except for patient age or KRAS mutation status in multivariate models including gene expression subtypes. Stratification of models according to these variables did not have a strong impact on the results (Supplementary Tables S6 and S11). Kaplan-Meier plots were generated with the ggsurvplot function in the survminer package (v.0.4.9), with p-values from Wald test. The proportion of explained variation in five-year RFS by each variable in multivariable models was calculated using rsq in the survMisc package (v.0.5.6)\(^5\), and bootstrapped with 5,000 iterations and sampling with replacement.

**Declarations**

**Data availability**

Gene expression data of multiregional samples from primary CRCs analyzed on Human Transcriptome 2.0 arrays have been deposited to GEO under accession number GSE241101 (n = 286). The extended single-sample primary tumor set has previously been deposited under accession numbers GSE24550, GSE29638, GSE69182, GSE79959, GSE139170, and GSE96528. The liver metastases samples have previously been deposited under accession number GSE159216. The remaining in-house gene expression data from patient-matched primary-metastasis samples will be published elsewhere. Public single-cell RNA sequencing data used for validation analyses were downloaded from GEO with accession number GSE144735.

**Code availability**

All data processing and analyses were performed with published software packages and computer code, and have been described and cited in the Results and/or Methods. No custom code was developed in the study.

**Acknowledgements**

The study was financially supported by the South-Eastern Norway Regional Health Authority (project number 2019042 to A.S. and 2017102 and 2016123 to R.A.L.), the Research Council of Norway (project number 287899 to A.S. and 250993 [FRIPRO Toppforsk] to R.A.L.), and the Norwegian Cancer Society (project number 208336-2019 to A.S. and 182759-2016 to R.A.L.). RNA sequencing and library generation was performed at the Oslo University Hospital Genomics Core Facility.

**Author contributions**

manuscript draft: J.L. and A.S. Study supervision: A.S. All authors were involved in revision of the manuscript and have approved the final version.

Competing interests statement

A.N., R.A.L. and A.S. are co-inventors of a patent application regarding the use of HSP90 inhibitors in relation to the consensus molecular subtypes of colorectal cancer (PCT/IB2018/000042). S.H.M., R.A.L. and A.S. are co-inventors of a patent application describing transcriptomic liver metastasis subtypes (LMS) of colorectal cancers (Attorney Docket No. INVEN-39613.101). Authors declare that they have no other competing interests.

References


### Figures

#### A

Homogeneous (n=52, 53%)

![Graph A]

Heterogeneous (n=39, 40%)

![Graph B]

Undetermined (n=7, 7%)

#### B

![Graph C]

#### C

![Graph D]

#### D

- Cancer mesenchymal transition
- Stromal infiltration
- Epithelial-mesenchymal transition
- Extracellular matrix organization
- TGFβ-response (fibroblasts)
- Immune infiltration
- Angiogenesis
- KRAS signaling
- Coagulation
- Cell cycle (mitotic)
- MYC targets
- Proteasome
- Cancer stem cell (LGR5)
- G2/M checkpoint
- Cancer stem cell (EPHB2)
Figure 1

Landscape of CMS heterogeneity among multiregional samples of primary CRCs

A) CMS classification of 2-4 multiregional samples from each of 98 primary tumors ordered according to intra-tumor classification heterogeneity. Each column represents one tumor. Spatially separated samples are annotated t1-t4 and colored according to CMS. MSI, RAS/BRAF and Location indicate the MSI status, mutation status (for \(\text{BRAF}^{V600E}\), \(\text{KRAS}\) and \(\text{NRAS}\)) and location of each tumor in the large bowel. One tumor exposed to pre-operative radiation is marked with an asterisk (*). Two patients with synchronous tumors are marked with daggers (†). B) Boxplot of the general transcriptomic heterogeneity of each tumor, estimated as the maximum Euclidean distance of PC1-PC3 between any pairs of samples per tumor and plotted according to CMS heterogeneity. \(P\)-value is from Welch’s t-test. C) Frequency of intra-tumor heterogeneity of each CMS class. The OR and 95% CI for enrichment of heterogeneous tumors are showed above each class. D) Bar plot of gene sets sorted according to \(p\)-values from gene set enrichment analysis of tumors with heterogeneous versus homogeneous CMS classifications (one random sample per tumor, \(n = 98\)). Enrichment in heterogeneous or homogeneous tumors is indicated by the color code.
Figure 2

Prognostic value of intra-tumor CMS heterogeneity

A) Relapse-free survival according to intra-tumor CMS heterogeneity among patients treated by complete resection of stage I-III CRC. Patients with synchronous tumors, pre-surgical radiation treatment and undetermined CMS heterogeneity were excluded from all analyses. B) Explained variation in five-year relapse-free survival by each variable in a multivariable Cox proportional hazards model, estimated as the percentage of the full model. C) Survival according to intra-tumor CMS heterogeneity and stratified by...
CMS4 classification. Hazard ratios (HR) are from Cox proportional hazards analyses and \( p \)-values from Wald tests. CAF, cancer-associated fibroblasts; CTL, cytotoxic lymphocytes; TN stage, tumor-node stage.

Figure 3

Gene categories and enrichments according to intra-tumor heterogeneity

A) Upper part shows the proportion of genes (filtered to include only genes with inter-tumor 10-90th percentile expression range > 1) categorized as either ITH-low \( (n = 1,540) \), ITH-intermediate \( (n = 1,549) \) or ITH-high \( (n = 149) \) in the multiregional primary tumor set \( (n = 286 \) samples). Lower part shows the density distribution of the corresponding ITH-scores, with dashed lines indicating thresholds for the three gene categories. The forest plot shows the ITH-score of genes designated as CRC-associated in the Cancer Gene Census \( (n = 13 \) genes passing the inter-tumor expression range filter). The genes are sorted according to involvement in pathways over-represented among ITH-low genes (Wikipathways cancer; Supplementary Figure S14). B) Density plots (upper: all gene sets) and bar plots (lower: 15 top-ranked gene sets) of Spearman's correlation coefficients (absolute values) between PC1 or PC2 of ITH-high (right) or ITH-low (left) genes and single-sample enrichment scores of significantly correlated gene sets \( (p < 0.05) \). C) Number of gene sets (of totally 54) with significant Spearman's correlations \( (p < 0.05) \) to PC1 and PC2 of ITH-high and ITH-low genes. CIN, chromosomal instability.
Figure 4

Classification of primary tumors and liver metastases based on ITH-low and cancer cell-intrinsic genes

A) Alluvial diagrams of classification concordances between iCMS and the k2 clusters identified based on ITH-low genes among primary CRCs and liver metastases. The sample overlap between classes in iCMS and k2 is indicated relative to the total number per subtype. B) Proportion of primary tumors with homogeneous and heterogeneous intra-tumor classifications of multiregional samples \( (n = 286 \text{ samples}) \) according to the indicated frameworks \( (k2 \text{ and } k4 \text{ are based on the ITH-low genes}) \). Proportion of patients with inter-metastatic heterogeneity among liver lesions \( (2-7 \text{ distinct lesions per patient, total } n = 143 \text{ metastases}) \) according to iCMS and the k2 clusters is also indicated. C) iCMS classifications of matched primary tumors and liver metastases from 35 patients \( (n = 179 \text{ samples}) \). Each column represents one patient. For primary tumors, each square represents one multiregional sample numbered with lower case t. For liver metastases, each square represents one tumor numbered with upper case T separately for each resection, and diagonal lines indicate multiregional samples. Indicated for each patient is the platform used for gene expression analysis, diagnosis with synchronous versus metachronous metastases and exposure to chemotherapy prior to sampling.
Figure 5

Prognostic value of the proposed congruent CMS framework

A) Proportion of the four congruent (cCMS) classes among primary CRCs \( (n = 704 \) samples from 516 tumors). B) Alluvial diagrams of classification concordances between k2 and cCMS (left; representing k2 and k4 classifications based on ITH-low genes), as well as cCMS and CMS (right). The sample overlap between classes is indicated relative to the total number per subtype and illustrated by dashed lines for k2 and CMS. C) Proportion of MSI tumors across the cCMS classes. D) Relapse-free survival according to cCMS among patients treated by complete resection of stage I-III CRC \( (n = 398) \). Patients with heterogeneous intra-tumor cCMS classifications, synchronous tumors, or pre-surgical radiation treatment were excluded from analyses. Hazard ratios (HR) are from Cox proportional hazards analyses and \( p \)-values from Wald tests. E) Explained variation in five-year relapse-free survival \( (n = 398 \) patients)
by each variable in a multivariable Cox proportional hazards model, estimated as the percentage of the full model. CAF, cancer-associated fibroblasts; CTL, cytotoxic lymphocytes; TN stage, tumor-node stage.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTablesLangerudetal.xlsx](#)
- [SupplementaryFiguresLangerudetal.docx](#)