

Supplementary materials to:

Intra-tumor heterogeneity impacts patient prognosis in localized colorectal cancer

by Marisa et al.

Supplementary Methods

Derivation of pseudo-FFPE profiles:

There is a gene-level bias between FF and FFPE expression profiles. Given a gene, let's denote Y its FFPE-based expression, and X its FF-based expression. One can fit a linear model ($Y = a.X + b$) to predict Y from X . Here, we wanted to build a CMS classifier for FFPE samples, however the CIT initial series, used as a training set, is based on FF samples. To overcome this difficulty, we calculated a pseudo-FFPE version of the FF samples. To do so, we selected 23 markers showing a linear systematic bias between FF and FFPE profiles, and derived a linear model for each of these 23 markers using the 90 paired FF/FFPE samples from the CIT extension series. We were thus able to transform the FF profiles from the CIT initial series in pseudo-FFPE profiles, for these 23 markers.

Calculus of WISP CMS proportions in publicly available transcriptome series:

To calculate WISP-based intratumor CMS proportions in new transcriptome series, one first needs to classify the tumors using the RF CMS classifier [1]. Then the CMS labels can be used to train WISP [2] using the `WISP.getPureCentro` function; this yields WISP-based CMS centroids. Note that this step will require that each of the four CMS is represented by a sufficient number of samples, this number

depending on the degree of intratumor heterogeneity. Lastly these WISP-based centroids are used to calculate the CMS intratumor proportions, using the WISP.getWeight function.

[1] <https://github.com/Sage-Bionetworks/CMSclassifier/blob/master/R/cmsClassifier.R>

[2] <https://github.com/cit-bioinfo/WISP>

Supplementary Materials

Fig. S1. Flow chart of the study

Fig. S2. Characterization of CMS as predicted by a Nanostring-based Random Forest classifier, on 1,779 samples from PETACC8 trial.

Fig. S3. CMS intra-sample heterogeneity (ITH) in 155 CC cell lines (GSE59857).

Fig. S4. Single cell RNA-seq analysis of Lovo and Mdst8 samples.

Fig. S5. Distribution of molecular and immune characteristics according to intra-sample CMS heterogeneity.

Table S1. Description of the 196 genes measured based on Nanostring technology. (excel file)

Table S2. Univariate and multivariate Cox models.

Table S3. P-values of the T-tests comparing tumors with RF CMS attribution > 50% to those with RF CMS attribution <50% for all continuous variables reported in Figure 1c.

Table S4. p-values of the T-tests comparing tumors with low ITH to those with high ITH for all continuous variables reported in Figure 4c.

Table S5. Comparison of ITH score, WISP-based major CMS, and WISP-based CMS1.CMS4 or CMS4.CMS1 combinations between metastatic and non-metastatic colon cancers from the public TCGA, GSE39582, GSE5851 and GSE72970 data.

Table S2

A. Disease free Survival univariate analysis

Characteristic	N	Summary Statistics ¹	HR ²	95% CI ²	p-value
Age	1779				
≤ 70		1594 (90%)	---	---	
>70		185 (10%)	1.28	1.00, 1.64	0.047
Gender	1779				
Female		760 (43%)	---	---	
Male		1019 (57%)	1.22	1.03, 1.45	0.020
WHO Score	1715				
0		1403 (82%)	---	---	
1-2		312 (18%)	1.38	1.13, 1.69	0.002
Sideness	1771				
distal		1067 (60%)	---	---	
proximal		704 (40%)	1.01	0.85, 1.19	>0.9
T stage	1778				
pT1-3		1401 (79%)	---	---	
pT4		377 (21%)	2.31	1.94, 2.75	<0.001
N stage	1779				
pN1		1115 (63%)	---	---	
pN2		664 (37%)	2.12	1.80, 2.50	<0.001
MMR status	1743				
dMMR		174 (10.0%)	---	---	
pMMR		1569 (90%)	1.21	0.90, 1.63	0.2
RAS status	1632				
NM		846 (52%)	---	---	
M		786 (48%)	1.42	1.19, 1.69	<0.001
BRAF status	1697				
NM		1508 (89%)	---	---	
M		189 (11%)	0.99	0.75, 1.30	>0.9
T-cell infiltration	1778				
High		592 (33%)	---	---	
Low		1186 (67%)	1.23	1.03, 1.47	0.025
Intra tumor heterogeneity Score	1779				
1		758 (43%)	---	---	
2		932 (52%)	1.34	1.12, 1.59	0.001
3		89 (5.0%)	1.75	1.24, 2.48	0.001
CMS combination	1779				
CMS2 pure		326 (18%)	---	---	
CMS1 pure		126 (7.1%)	0.99	0.65, 1.50	>0.9
CMS1.CMS2		38 (2.1%)	0.94	0.47, 1.87	0.9
CMS1.CMS3		63 (3.5%)	1.99	1.28, 3.09	0.002

CMS1.CMS4	65 (3.7%)	2.36	1.55, 3.61	<0.001
CMS2.CMS1	47 (2.6%)	1.51	0.88, 2.58	0.13
CMS2.CMS3	169 (9.5%)	1.19	0.83, 1.69	0.3
CMS2.CMS4	173 (9.7%)	1.40	1.00, 1.97	0.053
CMS3 pure	131 (7.4%)	1.18	0.80, 1.73	0.4
CMS3.CMS1	60 (3.4%)	0.78	0.42, 1.43	0.4
CMS3.CMS2	90 (5.1%)	1.41	0.93, 2.13	0.11
CMS3.CMS4	55 (3.1%)	2.43	1.58, 3.74	<0.001
CMS4 pure	175 (9.8%)	1.44	1.03, 2.01	0.034
CMS4.CMS1	66 (3.7%)	2.34	1.56, 3.51	<0.001
CMS4.CMS2	142 (8.0%)	1.55	1.09, 2.21	0.014
CMS4.CMS3	53 (3.0%)	1.87	1.17, 3.00	0.009
CMS low vs high risk*	1779			
Low risk CMS	1530 (86%)	---	---	
High risk CMS	249 (14%)	1.85	1.51, 2.27	<0.001

¹Statistics presented: n (%)

²HR = Hazard Ratio, CI = Confidence Interval

* high risk CMS groups CMS1-CMS4, CMS4-CMS1, CMS3-CMS4 and CMS1-CMS3

B. Overall Survival univariate analysis

Characteristic	N	Summary Statistics ¹	HR ²	95% CI ²	p-value
Age	1779				
≤ 70		1594 (90%)	---	---	
>70		185 (10%)	1.53	1.16, 2.00	0.002
Gender	1779				
Female		760 (43%)	---	---	
Male		1019 (57%)	1.34	1.10, 1.63	0.004
WHO Score	1715				
0		1403 (82%)	---	---	
1-2		312 (18%)	1.49	1.18, 1.87	<0.001
Sideness	1771				
distal		1067 (60%)	---	---	
proximal		704 (40%)	1.25	1.03, 1.51	0.026
T stage	1778				
pT1-3		1401 (79%)	---	---	
pT4		377 (21%)	2.49	2.04, 3.04	<0.001
N stage	1779				
pN1		1115 (63%)	---	---	
pN2		664 (37%)	2.30	1.90, 2.79	<0.001
MMR status	1743				
dMMR		174 (10.0%)	---	---	
pMMR		1569 (90%)	1.24	0.87, 1.76	0.2
RAS status	1632				
NM		846 (52%)	---	---	
M		786 (48%)	1.37	1.12, 1.68	0.002
BRAF status	1697				
NM		1508 (89%)	---	---	
M		189 (11%)	1.21	0.90, 1.63	0.2
T-cell infiltration	1778				
High		592 (33%)	---	---	
Low		1186 (67%)	1.10	0.89, 1.35	0.4
Intra tumor heterogeneity Score	1779				
1		758 (43%)	---	---	
2		932 (52%)	1.40	1.14, 1.71	0.001
3		89 (5.0%)	1.48	0.96, 2.28	0.073
CMS combination	1779				
CMS2 pure		326 (18%)	---	---	
CMS1 pure		126 (7.1%)	1.13	0.68, 1.88	0.6
CMS1.CMS2		38 (2.1%)	0.78	0.31, 1.95	0.6
CMS1.CMS3		63 (3.5%)	2.23	1.30, 3.81	0.004
CMS1.CMS4		65 (3.7%)	3.49	2.19, 5.56	<0.001
CMS2.CMS1		47 (2.6%)	1.33	0.66, 2.70	0.4
CMS2.CMS3		169 (9.5%)	1.52	1.00, 2.30	0.051

CMS2.CMS4		173 (9.7%)	1.76	1.18, 2.63	0.006
CMS3 pure		131 (7.4%)	1.52	0.96, 2.40	0.072
CMS3.CMS1		60 (3.4%)	1.07	0.54, 2.10	0.9
CMS3.CMS2		90 (5.1%)	1.82	1.13, 2.93	0.015
CMS3.CMS4		55 (3.1%)	2.10	1.20, 3.69	0.010
CMS4 pure		175 (9.8%)	1.85	1.24, 2.74	0.002
CMS4.CMS1		66 (3.7%)	3.22	2.02, 5.13	<0.001
CMS4.CMS2		142 (8.0%)	1.59	1.03, 2.46	0.038
CMS4.CMS3		53 (3.0%)	2.08	1.17, 3.71	0.012
CMS low vs high risk*	1779				
Low risk CMS		1530 (86%)	---	---	
High risk CMS		249 (14%)	1.93	1.52, 2.44	<0.001

¹Statistics presented: n (%)

²HR = Hazard Ratio, CI = Confidence Interval

* high risk CMS groups CMS1-CMS4, CMS4-CMS1, CMS3-CMS4 and CMS1-CMS3

C. Multivariate analysis of DFS including clinical, pathological, mutational and immunological variables (model 0)

Term	Hazard Ratio	95%CI	P-value
Age: >70y versus ≤70y	1.19	0.91 - 1.55	0.21
Gender: male versus female	1.16	0.97 - 1.4	0.10
WHO* score: 1-2 versus 0	1.41	1.14 - 1.74	0.0015
Bowel obstruction or perforation: yes versus no	1.25	1.01 - 1.54	0.04
T stage: T4 versus T1-T3	2.05	1.69 - 2.49	<0.0001
N stage: N2 versus N1	1.97	1.64 - 2.36	<0.0001
Grade: G1-G2 versus G3-G4	1.19	0.95 - 1.48	0.12
RAS status: RAS mutated versus RAS wildtype	1.42	1.18 - 1.7	0.0002
T-cell tumor infiltration: low and intermediate versus high infiltration	1.33	1.09 - 1.62	0.005

*World Health Organization †KRAS or NRAS

R² = 0.106; Likelihood ratio = 174.4; degree of freedom = 9; Harrell's C-index = 0.678

D. Multivariate analysis of DFS including clinical, pathological, mutational, immunological variables and CMS combination (model 1)

Term	Hazard Ratio	95%CI	P-value
Age: >70y versus ≤70y	1.22	0.93 - 1.59	0.15
Gender: male versus female	1.14	0.95 - 1.37	0.17
WHO* score: 1-2 versus 0	1.40	1.13 - 1.73	0.002
Bowel obstruction or perforation: yes versus no	1.26	1.02 - 1.55	0.035
T stage: T4 versus T1-T3	2.05	1.69 - 2.49	<0.0001
N stage: N2 versus N1	1.92	1.6 - 2.3	<0.0001
Grade: G1-G2 versus G3-G4	1.14	0.91 - 1.41	0.25
RAS status: RAS mutated versus RAS wildtype	1.40	1.17 - 1.68	0.0003
T-cell tumor infiltration: low and intermediate versus high infiltration	1.42	1.16 - 1.74	0.0006
CMS combination: high risk versus low risk CMS combination	1.74	1.38 - 2.19	<0.0001

* World Health Organization †KRAS or NRAS

R² = 0.118; Likelihood ratio = 194.7; degree of freedom = 10; Harrell's C-index = 0.685

Model 1 versus Model 0 Likelihood ratio test P<0.0001 (Chi-square test); C-indexes are statistically significantly different (p<0.03)

E. Multivariate OS analysis including clinical, pathological and mutational variables (model 2)

Term	Hazard Ratio	95%CI	P-value
Age:		1.11 -	
>70y versus ≤ 70y	1.48	1.98	0.008
Gender:		1.06 -	
male versus female	1.31	1.63	0.013
WHO* score:			
1-2 versus 0	1.52	1.2 - 1.93	0.0006
Bowel obstruction or perforation:			
yes versus no	1.17	0.91 - 1.5	0.22
T stage:		1.65 -	
T4 versus T1-T3	2.06	2.57	<0.0001
N stage:			
N2 versus N1	2.03	1.64 - 2.5	<0.0001
Grade:		1.08 -	
G1-G2 versus G3-G4	1.38	1.78	0.01
Sidedness:		0.86 -	
Proximal versus distal	1.07	1.33	0.54
RAS[†] status:		1.13 -	
RAS mutated versus RAS wildtype	1.39	1.72	0.002

* World Health Organization; [†] KRAS or NRAS

R² = 0.093; Likelihood ratio test = 150.5; degree of freedom = 9; Harrell's C-index = 0.691

F. Multivariate OS analysis including clinical, pathological, mutational variables and combination of CMS (model 3)

Term	Hazard Ratio	95%CI	P-value
Age:		1.15 -	
>70y versus ≤ 70y	1.54	2.06	<0.004
Gender:			
male versus female	1.29	1.04 - 1.6	0.021
WHO* score:		1.19 -	
1-2 versus 0	1.51	1.92	0.0007
Bowel obstruction or perforation:			
yes versus no	1.17	0.91 - 1.5	0.23
T stage:		1.64 -	
T4 versus T1-T3	2.05	2.56	0.0000
N stage:		1.61 -	
N2 versus N1	1.99	2.46	0.0000
Grade:			
G1-G2 versus G3-G4	1.33	1.03 - 1.7	0.029
Sidedness:		0.82 -	
Proximal versus distal	1.02	1.27	0.88
RAS[†] status:		1.13 -	
RAS mutated versus RAS wildtype	1.40	1.73	0.002
CMS combination:			
high risk versus low risk CMS combination	1.76	1.33 - 2.34	0.0001

* World Health Organization [†]KRAS or NRAS

R² = 0.101; Likelihood ratio = 164.2; degree of freedom = 10; Harrell's C-index = 0.695

Model 3 versus Model 2 Likelihood ratio test P<0.0003(Chi-square test)

Table S3

p-values of the T-tests comparing tumors with RF CMS attribution > 50% to those with RF CMS attribution <50% for all continuous variables reported in Figure 1c, within each RF predicted CMS:

RF pred.	CMS1	CMS2	CMS3	CMS4
T.cells	0.56	5e-03	0.06	0.80
CD8.T.cells	0.17	0.39	0.39	0.10
Cytotoxic.lymphocytes	1e-05	1e-03	0.83	0.63
NK.cells	0.03	0.04	0.64	0.51
B.lineage	0.38	0.02	0.06	0.79
Monocytic.lineage	0.31	2e-04	0.19	9e-04
Myeloid.dendritic.cells	0.23	3e-02	0.86	0.15
Neutrophils	8e-03	5e-04	0.41	0.44
Endothelial.cells	0.05	2e-03	0.43	9e-03
Fibroblasts	0.25	4e-03	0.04	1e-11
ImmunoScore.v2	0.13	2e-02	0.29	0.47
EMT	0.04	0.19	0.71	1e-03
TGFB	0.25	2e-07	0.43	7e-04
Angiogenesis	0.27	7e-03	0.23	2e-07

Table S4

p-values of the T-tests comparing tumors with low ITH to those with high ITH for all continuous variables reported in Figure 4c, within each WISP-based major CMS:

ITH	CMS1	CMS2	CMS3	CMS4
CD8 T cells	0.2	0.01	0.3	0.02
Cytotoxic lymphocytes	6e-08	8e-04	0.07	0.02
NK cells	9e-04	0.008	0.3	0.2
B lineage	0.7	7e-08	0.3	2e-04
Monocytic lineage	0.8	2e-13	0.004	2e-07
Myeloid dendritic cells	0.01	0.05	0.2	0.9
Neutrophils	0.3	7e-06	0.1	0.4
Endothelial cells	2e-04	5e-08	0.01	2e-06
Fibroblasts	6e-05	1e-17	5e-06	2e-17
ImmunoScore-like	0.03	1e-06	0.6	0.002
EMT	3e-04	0.5	0.001	0.07
TGFB	0.001	2e-11	0.009	5e-08
Angiogenesis	2e-04	2e-07	3e-04	7e-12

Table S5

A: distribution of ITH scores in metastatic (M+) and non-metastatic (M-) colon cancer

ITH score	1	2	3
M+	138	192	18
M-	384	478	40

X-squared test, p value ns

B: distribution of WISP-based major CMS in metastatic (M+) and non-metastatic (M-) colon cancer

CMS	CMS1	CMS2	CMS3	CMS4
M+	128	392	248	134
M-	19	118	58	153

X-squared test, p value < 10⁻¹⁶

C: distribution of WISP-based CMS1.CMS4 or CMS4.CMS1 combinations versus other combinations in metastatic (M+) and non-metastatic (M-) colon cancer

	CMS1.CMS4 or CMS4.CMS1	Other combinations
M+	27	321
M-	42	870

X-squared test p value = 0.039

FIGURE S1

Study flow chart

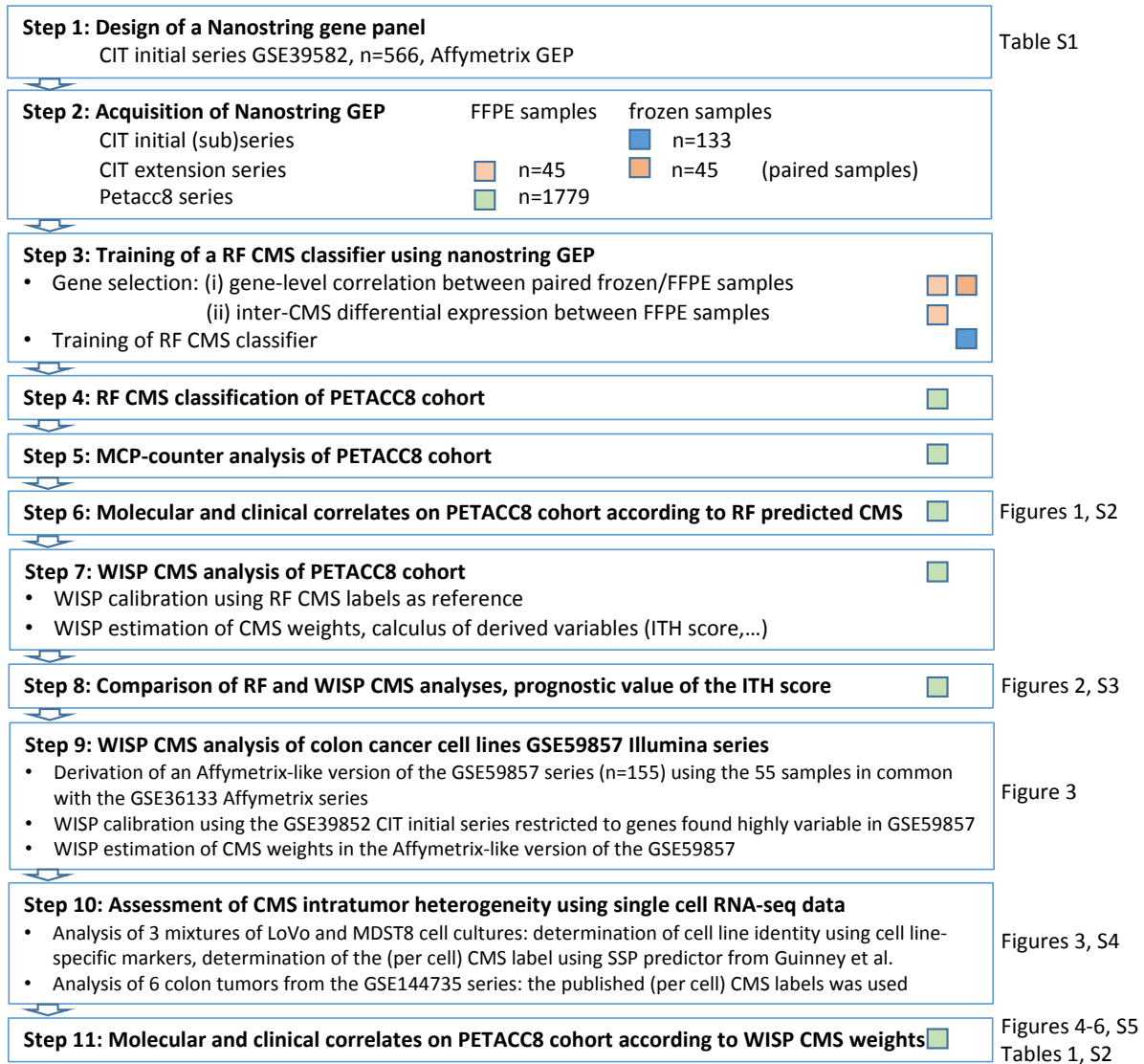
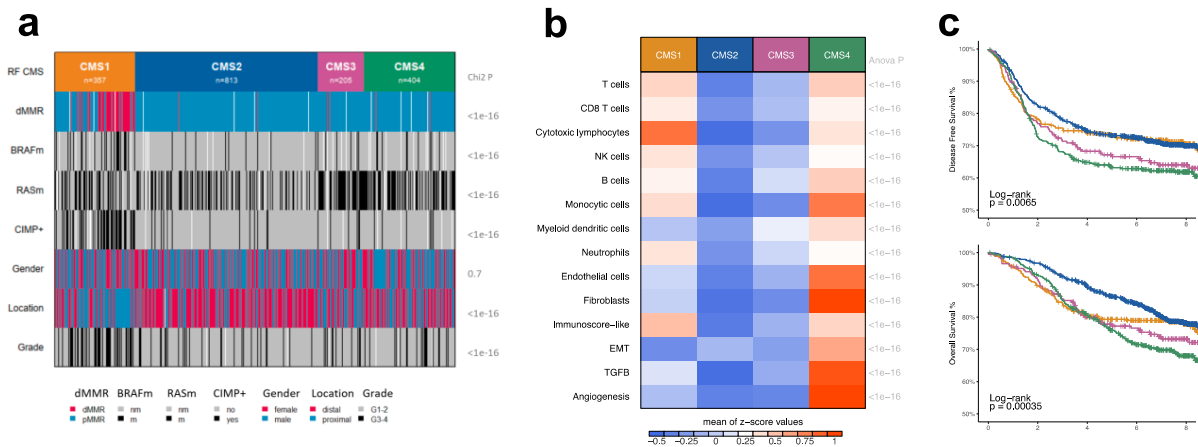
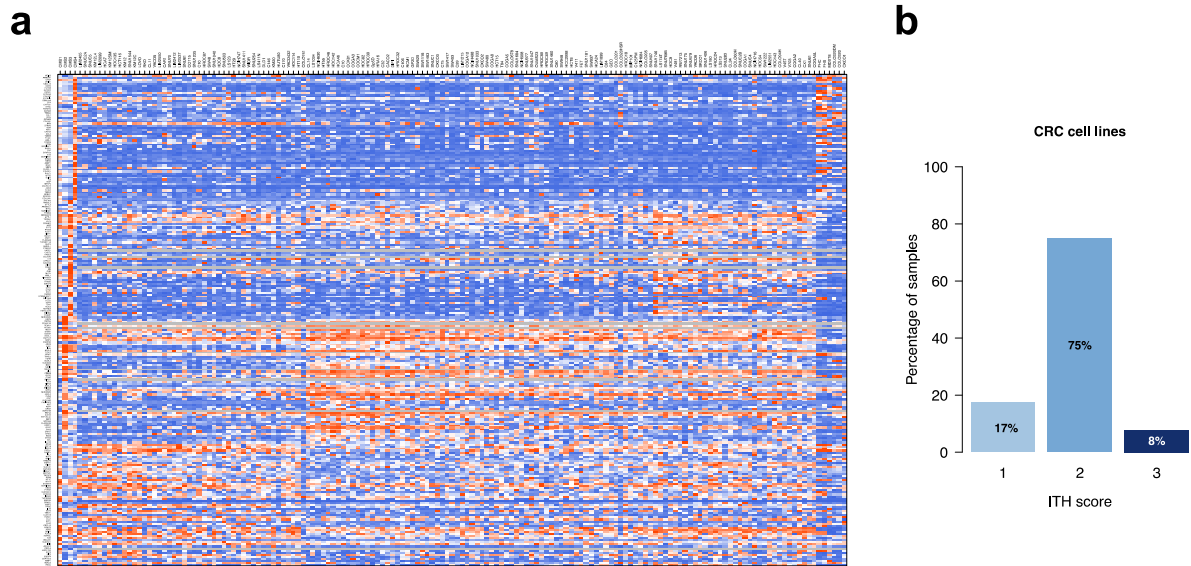


FIGURE S2



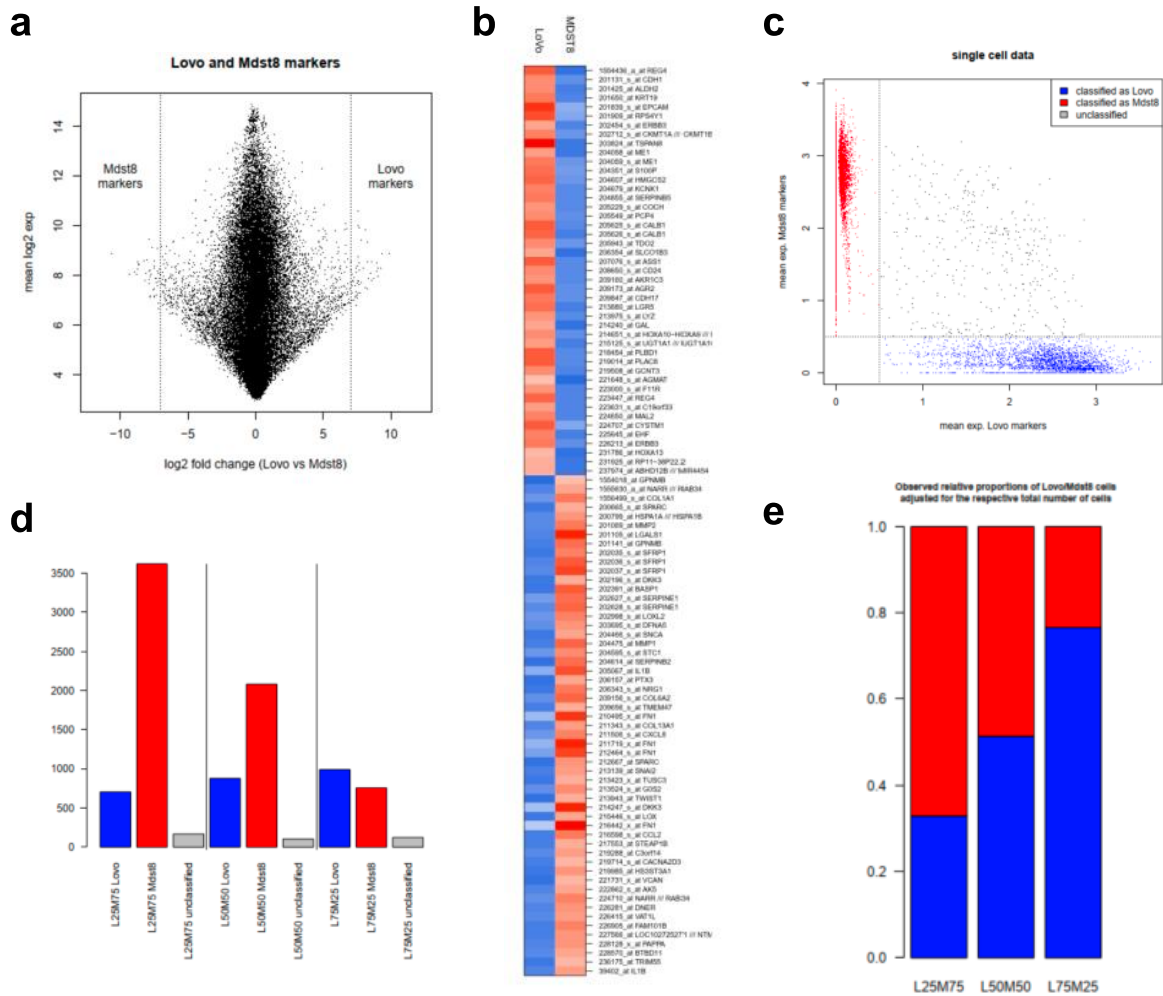
Extended data Fig. S2. Characterization of CMS as predicted by a Nanostring-based Random Forest classifier on 1,779 samples from PETACC8 trial. a. Samples (columns) are grouped according to their predicted CMS, and annotated for the following parameters: dMMR – red: yes, blue: no, white: unknown; BRAFm – black: yes, grey: no, white: unknown; RASm – black: yes, grey: no, white: unknown; CIMP+ – black: yes, grey: no, white: unknown; Gender – pink: female, blue: male; Location – blue: proximal, red: distal; Grade – grey: G1/G2, black: G3/G4, white: unknown. **b.** Heatmap annotating CMS for 8 immune and 2 stromal cell population infiltration based on MCP-counter estimates and 3 pathways (mean scores per CMS). Kaplan-Meier curves of DFS (**c**) and OS (**d**) are shown according to the four CMS.

FIGURE S3



Extended data Fig. S3. CMS intra-sample heterogeneity (ITH) in 155 CC cell lines (GSE59857). **a.** Heatmap showing the expression profile of the markers of the four CMS across the CMS1..4 centroids (first columns, starting from the left) and the 155 cell lines (remaining columns). **b.** Histograms representing the proportion of cell lines with low (1), intermediate (2) and high(3) CMS ITH score.

FIGURE S4

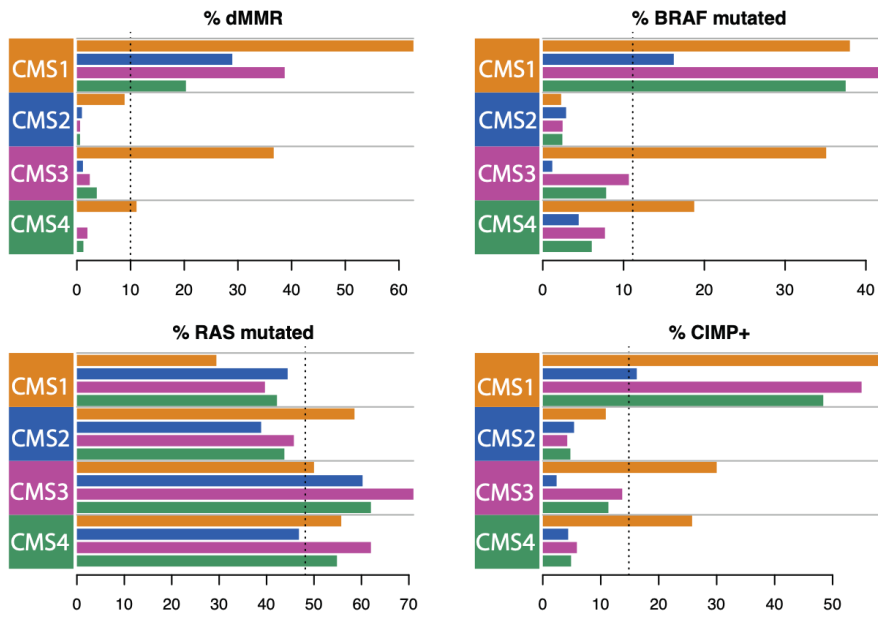


Extended data Fig. S4. Single cell RNA-seq analysis of Lovo and Mdst8 samples. **a.** Selection of Lovo and Mdst8 markers based on bulk expression profiles of these two cell lines (GSE36133). The x axis represents the log₂ fold change of expression between the (bulk) transcriptome profiles; the y axis represents the mean log₂ transcriptome profiles; each dot represents a gene. Affymetrix probesets with a log₂ fold change above 7 (respectively below -7) were selected as Lovo (respectively Mdst8) specific markers. **b.** Heatmap representing the expression of the selected markers in Lovo and Mdst8 bulk profiles (red: high expression, blue: low expression). **c.** Assignment of single cells to Lovo or Mdst8 based

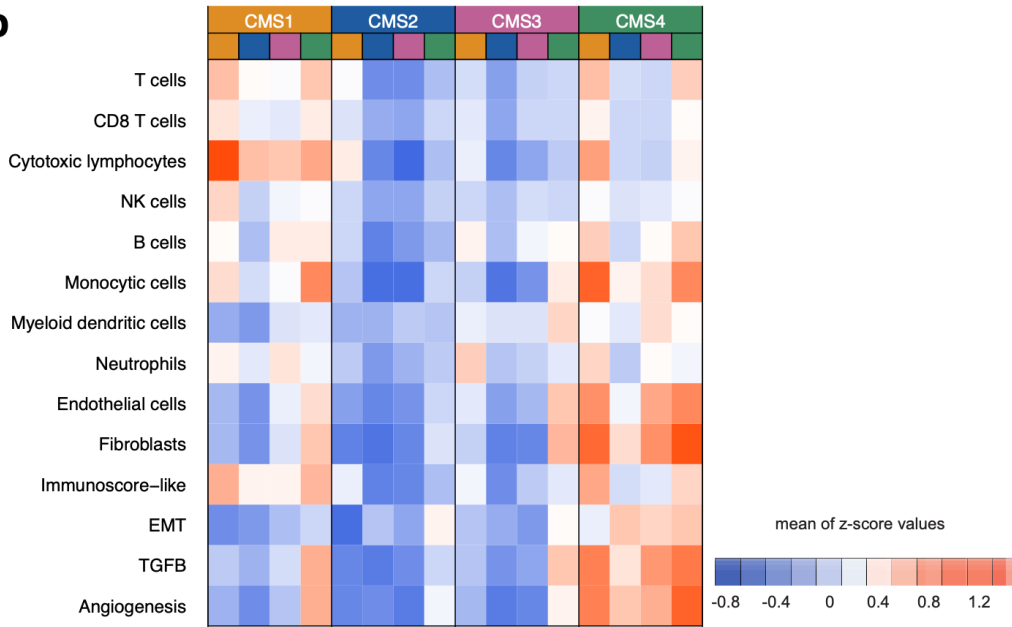
on their mean expression of Lovo (x-axis) and Mdst8 (y-axis) markers. Cells showing both a mean expression of Lovo (respectively Mdst8) markers above 0.5 and a mean expression of Mdst8 (respectively Lovo) markers below 0.5 were classified as Lovo (respectively Mdst8) cells, represented in blue (respectively red). Other cells were not classified. **d.** Number of cells classified as Lovo or Mdst8, or unclassified, within each of the three mixtures of Lovo and Mdst8. **e.** Observed relative proportion of Lovo (blue) and Mdst8 (red) cells in the three mixtures of Lovo and Mdst8, adjusted for the total number of Lovo and Mdst8 cells.

FIGURE S5

a



b



Extended data Fig. S5. Distribution of molecular and immune characteristics according to intra-sample CMS heterogeneity. **a.** Barplots showing the percent of samples with dMMR (deficient Mismatch repair), BRAF mutated, RAS mutated (KRAS/NRAS), CIMP+ (CpG Island Methylator Phenotype), within the 16 subgroups defined by combining WISP major CMS and WISP minor CMS. The dotted line represents the overall mean frequency. **b.** Heatmap annotating the 16 above defined subgroups for 8 immune and 2 stromal cell populations infiltration (based on MCP-counter estimates), a surrogate of the Immunoscore® (Immunoscore-like) and 3 pathways. CMS color codes: CMS1 (orange), CMS2 (blue), CMS3 (purple), CMS4 (green).