

1 **TITLE**

2 **Epithelial-to-mesenchymal transition lowers the cholesterol pathway, which**
3 **influences colon tumors differentiation.**

4

5 **RUNNING TITLE**

6 **Epithelial-to-mesenchymal transition regulates the cholesterol pathway and colon**
7 **cancer differentiation**

8

9

10 **AUTHORS' LIST**

11 Anaïs Aulas^{1*}, Maria Lucia Liberatoscioli^{1*}, Pascal Finetti¹, Olivier Cabaud¹, David J.
12 Birnbaum^{1,2}, Daniel Birnbaum^{1#}, François Bertucci^{1,3#} & Emilie Mamessier^{1#}

13

14 **AUTHORS' CONTRIBUTION and notes**

15 AA, FB, DB and EM designed the study. AA, MLL, and OC performed experiments. PF,
16 MLL and FB did the biocomputational analyses. DJB provided samples and clinical
17 rationale. FB, DJB, DB proofread the manuscript. AA, MLL and EM designed the figures
18 and wrote the manuscript. co-authorship: * AA and ML are co-first authors. # DB, FB
19 and EM are co-last authors.

20

21 **AFFILIATION**

22 ¹ Predictive Oncology Laboratory, Cancer Research Center of Marseille (CRCM),
23 Inserm U1068, CNRS UMR7258, Institut Paoli-Calmettes, Aix Marseille Université,
24 Marseille, France.

25 ² Department of Digestive Surgery, Hospital Nord, Aix-Marseille University, Chemin des
26 Bourrely, 13015, Marseille, France.

27 ³ Oncology Medical unit, Institut Paoli-Calmettes, Aix-Marseille University, Marseille,
28 France.

29

30 **CORRESPONDING AUTHOR**

31 emilie.mamessier@inserm.fr

32

33 **KEY WORDS**

34 Epithelial to mesenchymal transition, EMT, mesenchymal to epithelial transition, MET,
35 colorectal cancer, metastasis, prognosis, cholesterol, survival

36

37

38 **TRANSLATIONAL RELEVANCE (120-150 WORD)**

39 Metastases are the main cause of death of colorectal cancer (CRC) and the
40 major issue in CRC management. Understanding the chain of events that lead to
41 metastasis occurrence is urgent to identify new biomarkers of progression and/or new
42 potential targets. Epithelial to mesenchymal transition (EMT) is a major player in cancer
43 dissemination that remains non-targetable due to its broad involvement in tissue
44 homeostasis. Here, we used a relevant *in vitro* dynamic model to identify the pathways
45 triggered during EMT in CRC and how this might improve tumors classification,
46 prognosis and open therapeutic avenue for this deadly disease. The cholesterol
47 pathway surprisingly popped-out from this model and turned out to be a good prognosis
48 factor of disease-free survival for CRC. Altogether, our results showed that an active
49 EMT program lowers the cholesterol pathway, which further influence the differentiation
50 of colon tumor toward the most aggressive subtypes.

51

52

53 **ABSTRACT 185 words (current 184)**

54 Colorectal cancer (CRC) is the second cause of death worldwide. Up to 70% of
55 CRC patients will metastasize. Identify new biomarkers of progression to prevent/delay
56 disease evolution is urgent. Epithelial-to-mesenchymal transition (EMT) is a major
57 program engaged during metastasis. We aim at identifying a combination of new genes
58 that further identify signs of EMT in cancer tissues. We treated HT-29 cells grown in 3D
59 with an EMT Inducing cocktail, and collect them before during and after treatment. For
60 each condition, pan-transcriptomic analyses were done. Genes that were both induced
61 upon EMT induction and inhibited upon EMT release (mesenchymal-to-epithelial
62 transition or MET) were selected. We identified new genes for EMT-MET programs.
63 These genes were used to build a metagene that, when applied to a database of
64 transcriptomic data from primary colorectal tumors (n= 2,239), had an independent
65 prognosis value. Finally, we submitted this metagene to CMap and identified drugs that
66 might affect EMT-MET programs. Statins, well-known inhibitors of cholesterol synthesis,
67 were among them and effectively delayed MET. These data show that cholesterol and
68 EMT pathways are opposite regulators and impact differently tumor differentiation and
69 outcome.

70 **INTRODUCTION**

71 Colorectal cancer (CRC) is the third most common cancer and the second cause
72 of death worldwide (International Agency for Research on Cancer, WHO;
73 <http://gco.iarc.fr/today/fact-sheets-cancers>). However, the actual therapeutic
74 management proposed to CRC patients is not sufficient. For now, it consists in the
75 surgical removal of the primary tumor, followed by adjuvant chemotherapy in high-risk
76 cases. Metastases are the main cause of death of CRC patients, and thus the major
77 issue in CRC management. Up to 70% of CRC patients will relapse or metastasize
78 within 5 years following surgery of the primary tumor. The 5-year overall survival for
79 metastatic patients is 14% against 90% for non-metastatic patients [1].

80 Cells prone to metastasis undergo an epithelial-to-mesenchymal transition
81 (EMT), a biological process by which cells are re-programmed, lose their cell-to-cell
82 contacts, polarity and ultimately reshape their cytoskeleton to escape from the initial
83 tumor site [2]. Many works have searched for the initial events that trigger EMT in
84 cancer cells, based on the assumption that targeting EMT inducers might limit
85 metastasis [3, 4]. TGF β , which is a prominent growth factor at primary tumor sites [5, 6],
86 turned up to be a key inducer of the EMT program [7]. However, TGF β inhibition was
87 not successful because of the major involvement of this growth factor in many other
88 physiological functions [5, 6]. Moreover, EMT can be triggered in many ways that go
89 beyond the TGF β pathway alone [8, 9]. We choose to use a combination of different
90 factors known to be involved in the EMT process such as recombinant WNT5A [10, 11],
91 TGF β [12] and neutralizing antibodies against known inhibitors of EMT, such as SFRP1
92 [13, 14], DKK1 [15] and CDH1 [16, 17]. The combination of these factors reproducibly
93 induces an EMT phenotype in epithelial cancer cells cultivated *in vitro* within one week:
94 cells detach from their neighbors, adopt spindle shapes and a migratory potential.
95 Characteristic features of treated cells are a decreased expression of adhesion proteins

96 E-Cadherin and Occludin, whereas Fibronectin and Vimentin expression are enhanced
97 [18-20].

98 The identification of genes specifically activated during the EMT program has
99 been another goal, however difficult to achieve in complex samples such as human
100 tumors. *In vitro* approaches have thus often been privileged. A major common caveat
101 from these approaches was that the cells exposed to the different EMT inducing
102 factors were usually grown flat [6, 15, 21-23]. We know now that signaling pathways of
103 cells grown in 2D are different from those of tissues and that 3D cultures are more
104 relevant [24, 25]. The extensive study from cells grown in 2D most certainly contributed
105 to limit the identification of EMT regulatory factors to components forming the « tip of the
106 iceberg », leaving hidden other potential actors. This limitation led us to study how EMT
107 induction can promote pro-invasive features from colon cancer cells, with the aim of
108 revealing novel EMT-related molecular vulnerabilities, potentially useful as new
109 prognostic factors and/or novel targets.

110 For this, we used a 3D *in vitro* model and a timeline exposition to the EMT-
111 inducing cocktail to study reversible genes involved in EMT and Mesenchymal-to-
112 Epithelial Transition (MET) programs. From the genes identified, we build a metagene
113 and tested its prognostic value in a large database of CRC samples, annotated with
114 clinical data. Finally, still using these genes, we searched for FDA approved drugs that
115 might influence the EMT-MET programs and found that statins alter the execution of the
116 EMT program.

117 **MATERIAL& METHODS**

118 **Cell culture**

119 HT29 cells were obtained from the ATCC[®] Human Cell Lines (ATCC, Manassas, VA,
120 USA) and cultured using Dulbecco Modified Eagle Medium (DMEM, GIBCO, Waltham,
121 MA, USA) with 10% Fetal Calf Serum (FCS, Eurobio, Les Ulis, France) in 5% CO₂ at
122 37°C. Classical 2D cell cultures were seeded in Falcon[®] 6-well, whereas spheroids
123 culture were seeded in Cell Carrier Spheroid ULA 96-well Microplates™ (PerkinElmer,
124 Waltham, MA, USA). For EMT induction, cells were plated 2 days in regular media
125 before to be supplemented with StemXVivo[®] EMT Inducing Media Supplement (R&D,
126 Minneapolis, MN, USA) at 1X concentration for the indicated time. Cells were treated
127 with lovastatin and simvastatin (Sellekchem, Houston, TX, USA) at indicated
128 concentration and time.

129

130 **Western Blotting**

131 Cells were scraped and lysed in RIPA buffer (150 mM NaCl, 50 mM Tris, pH 7.4, 1%
132 Triton X-100, 0.1% SDS, 1% sodium deoxycholate) supplemented with HALT protease
133 inhibitor 1X (Thermo Fisher Scientific, Waltham, MA, USA), incubated for 10 min on ice,
134 and centrifuged at 13800 g. Supernatants were collected and quantified with Bradford
135 Protein Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). NuPAGE[®] LDS
136 Sample Buffer (Invitrogen, Carlsbad, CA, USA) was added to samples to 1X final
137 concentration. Samples were boiled at 95 °C for 5 min, loaded into
138 acrylamide/bisacrylamide gels and transferred to nitrocellulose membrane (GE
139 Healthcare, Chicago, IL, USA). Membranes were blocked using PBS 0.1% tween 5% fat
140 free milk for at least 30 min and incubated with primary antibodies diluted in PSB 5%
141 normal horse serum overnight at 4 °C (**Table S1**). The following day, membranes were
142 incubated with HRP coupled secondary antibodies diluted in PSB 5% normal horse
143 serum for 1 h at room temperature (**Table S1**). Blots were visualized by addition of ECL

144 Western Blotting Substrate (Thermo Fisher Scientific, Waltham, MA, USA) following
145 manufacturer's instructions. Blots were revealed using the imaging system ChemiDoc
146 MP (BioRad, Hercules, CA, USA).

147

148 **RNA extraction**

149 Total RNA was extracted from the HT29 spheroids using the RNeasy Mini Kit (Qiagen,
150 Hilden, Germany) following the manufacturer's instruction. The quality was tested using
151 an Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA).

152

153 **Gene expression profiling of treated and untreated HT29 spheroids**

154 Profiling was done using Affymetrix GeneChip HuGene 2.0 ST microarrays according to
155 manufacturer instructions (Affymetrix, Thermo Fisher, Waltham, MA, USA). Expression
156 data were analyzed by Robust Multichip Average (RMA) [26] in R using Bioconductor
157 and associated packages. Prior to analysis, log₂-transformed data were filtered out to
158 exclude poorly expressed genes, defined as genes having a median expression level
159 across all samples lower than the first quartile range. Data filtering resulted in 17,421
160 annotated genes.

161 Unsupervised analysis, applied to the 3,131 genes with standard deviation (SD)>0.25,
162 was performed with the Cluster program [27] using Pearson correlation and average
163 linkage clustering as parameters. Results were displayed using TreeView program [27].
164 Robustness of clusters was assessed with the R-package pvclust [28] with identical
165 clustering parameters and 100 bootstrap replications. Ontology analysis of the gene
166 clusters was based on GO biological processes of the Database for Annotation,
167 Visualization and Integrated Discovery (DAVID; <http://david.abcc.ncifcrf.gov/>). The
168 Consensus Molecular Subtype (CMS) classification [29] was applied to each cell culture
169 condition by using the CMScaller tool from Eide et al. [30], providing a CMS score for
170 each subtype (CMS1, CMS2, CMS3, and CMS4) to each condition.

171 Supervised analyses between the different cell culture conditions was applied by using
172 Gene Set Enrichment Analysis (GSEA) (<http://www.broadinstitute.org/gsea/>) in order to
173 identify the differential biological pathways. Analysis was focused on the 50-hallmark
174 gene sets from the MSigDB v7.0 Molecular Signatures Database ([https://www.gsea-
175 msigdb.org/gsea/msigdb](https://www.gsea-msigdb.org/gsea/msigdb)). We used the class differential metric to rank the filtered
176 genes, weighted enrichment statistic for computing enrichment score (ES) of each gene
177 set tested and 1000 permutations to evaluate significance as parameters for the GSEA.
178 Significant gene sets were defined by a False Discovery Rate (FDR)-corrected p-value
179 <0.01 . Metagenes from the significant gene sets were computed as the mean
180 expression of core genes defined by the GSEA analysis.

181

182 **Gene expression profiling of human cancer colon samples**

183 We built a large database gathering our own data set and 10 publicly available sets as
184 described [26]. These latter had been collected from the National Center for
185 Biotechnology Information (NCBI)/Genbank GEO, ArrayExpress and TCGA databases
186 (**Table S2**). Samples had been profiled using DNA microarrays (Affymetrix, Thermo
187 Fisher, Waltham, MA, USA) or Illumina RNA sequencing (United kingdom, Cambridge).
188 The whole data set included 2,239 primary colon cancer samples in the final analysis.
189 Their clinicopathological characteristics are summarized in **Table 1**. The pre-analytic
190 data processing and data analysis were done as previously described [26]. The CMS
191 classification and the metagenes identified by supervised analyses in the HT29 cells
192 were applied to each data set separately [29]. For these latter, the natural score of 0
193 was used as threshold to define a sample as “positive” (score >0) or “negative”
194 (score <0).

195

196 **Connectivity Map interrogation**

197 Finally, in order to identify potential therapies, we applied the Connectivity Map (CMap)
198 analysis [31]. This analysis contains a perturbation-driven gene expression dataset of
199 over a million of gene expression signatures associated with molecule compounds and
200 genetic perturbations treatment on human cells. CMap (<https://clue.io/cmap>) searched
201 for, among the L1000 signatures, the perturbagen's signatures that give significant
202 correlation with those genes included in our two interest metagenes given by
203 connectivity scores. The tested signatures were limited to those with a well-defined
204 mechanism of action of treatment compounds, corresponding to 10,847 signatures of
205 3,248 drugs through 57 cultured human cell lines. The significance threshold of the
206 normalized connectivity score (ncs) was adjusted p-value <0.001. Because each drug
207 has been used through multiple different culture conditions, we aggregated the
208 significant signatures by drug class, computed for each class the average of normalized
209 connectivity scores (average ncs) and assessed its statistical significance using one
210 sample Student t-test.

211

212 **Statistical analysis**

213 Correlations between the "EMT-CHL model"-based classes and the clinicopathological
214 variables were calculated with the Student's t-test for the continuous variables and the
215 Fisher's exact test for the binary variables. The survival endpoint was relapse-free
216 survival (RFS) calculated from the date of diagnosis until the date of relapse. The
217 follow-up was measured from the date of diagnosis to the date of last news for event-
218 free patients. Survival was calculated using the Kaplan-Meier method and curves were
219 compared with the log-rank test. Univariate and multivariate analyses were done using
220 Cox regression analysis (Wald test). The variables tested in univariate analysis included
221 patients' age (> vs. ≤50 years) and gender (male vs. female), anatomic location
222 (proximal vs. distal colon), pathological stage (2, 3, 4 vs. 1), pathological grade (2, 3 vs.
223 1), CMS molecular subtypes (CMS2, CMS3, CMS4 vs. CMS1), and the "EMT-CHL

224 model"-based classes ("high-risk" vs. "low-risk"). Multivariate analysis incorporated all
225 variables with a p-value inferior to 5% in univariate analysis. The likelihood ratio (LR)
226 tests were used to assess the prognostic information provided beyond that of each
227 metagene included in the Cox model, assuming a χ^2 distribution. Changes in the LR
228 values (LR- $\Delta\chi^2$) quantified the relative amount of information of one Cox model
229 compared with another. In order to build a prognostic Cox model for RFS combining our
230 two metagenes, we randomly split the 1,837 informative primary samples in two sets,
231 training and validation sets. The training set was composed of a quarter of the samples
232 (n=450) and allowed to build the model. This model was then applied to the validation
233 set composed of the 1,387 remaining samples, to estimate the robustness of the
234 predictor. All statistical tests were two-sided at the 5% level of significance. Statistical
235 analysis was done using the survival package (version 2.30) in the R software (version
236 3.5.2).

237 **RESULTS**

238 **The EMT-inducing cocktail induces reversible loose micro-tissues, consistent**
239 **with EMT and EMT-MET-like programs**

240 The EMT-inducing cocktail used for this study is a potent inducer of EMT [18, 20,
241 32-35]. However, growth in 2D on plastic dishes offers unnatural growth kinetics and
242 cells attachment, thus altering adhesion and related signaling pathways, and potentially
243 biasing and/or masking EMT-induced signals. To study EMT induction in a more
244 physiologic and dynamic model, we looked at the reversible effects of an EMT-inducing
245 cocktail on cells grown in 3D as spheroids. We induced EMT on several colon cancer
246 cell lines (HT29, HCT116, SW480). The most striking phenotype was observed with the
247 HT29 colon cancer cell line, and we choose to continue with them. We used three
248 conditions on HT29 spheroids: HT29 spheroids grown without any treatment (Baseline
249 condition), HT29 spheroids exposed to the EMT-inducing cocktail for 5 days to induce
250 EMT, and HT29 spheroids exposed to the EMT-inducing cocktail for 5 days, washed,
251 then incubated for an additional 4 days without any treatment to mimic a sequential
252 EMT followed by a MET process (**Figure 1A**).

253 In the absence of any specific treatment (Baseline), the spheroids continued to
254 grow as dense spheroids during the whole protocol (**Figure 1B, top row**). When the
255 EMT-inducing cocktail was added to the media (EMT condition) the spheroids became
256 loose, with round cells at the periphery of the spheroid core, already after 3 days of
257 treatment (**Figure 1B, middle row**). In the EMT-MET condition, the first 5 days were
258 similar to the EMT condition, then during the following 3 days, *i.e.* after the removal of
259 the EMT-inducing cocktail (MET-*like* program), the individualized round cells
260 progressively reintegrated the spheroids core to form a homogeneous and compact
261 spheroid (**Figure 1B, bottom row**).

262 We verified that the EMT-like phenotype induced by the EMT-inducing cocktail
263 actually induced known EMT factors. We studied key molecules of EMT at the protein

264 level by western blotting. We compared Baseline microtissues samples to EMT and
265 EMT-MET samples. The EMT-inducing cocktail induced the expression of specific
266 mesenchymal proteins such as Fibronectin, ZEB1 and Vimentin (**Figure 1C**). All these
267 markers went back to their respective baseline after removal of the EMT-inducing
268 cocktail (*i.e.* Baseline condition and EMT-MET are similar). Conversely and as
269 expected, the epithelial marker OCCLUDIN was downregulated during the EMT
270 process.

271 Thus the EMT-inducing cocktail induced reversible macroscopic changes,
272 phenotypically compatible with an EMT program induction.

273

274

275 **Transcriptomic analysis confirms the activation of EMT and EMT-MET programs** 276 **in spheroids treated with the EMT-inducing cocktail**

277 We surmised that the effect of the EMT-inducing cocktail was not limited to
278 known EMT markers. To identify broad biological pathways activated by the EMT-
279 inducing cocktail in treated spheroids, we performed DNA microarray analyses
280 (Affymetrix™ GeneChip Human Gene 2.0 ST Assay) from the three conditions
281 described above (Baseline, EMT, and EMT-MET, **Figure 1A**) and compared their
282 expression profiles. Analyses were done at day 5 for Baseline and EMT samples and at
283 day 7 for EMT-MET samples. Each condition was tested in triplicate. Unsupervised
284 hierarchical clustering revealed, as expected, three distinct and coherent clusters,
285 corresponding to each of the three conditions tested: Baseline, EMT, and EMT-MET
286 (**Figure 1D**). The EMT and EMT-MET conditions were closer to each other than to the
287 Baseline condition (**Figure 1D**). This is coherent with the fact that at day 7, spheroids
288 were not fully back to baseline yet (**Figure 1B**). Visual inspection of gene clusters
289 revealed two clusters, thereafter designated 1 and 2, that were deregulated from the
290 Baseline condition to the EMT condition and showed opposite deregulation from the

291 EMT condition to the EMT-MET condition. Cluster 1 was upregulated in EMT condition
292 and downregulated in EMT-MET condition: it included 198 genes with minimal Pearson
293 correlation equal to 0.74. In term of ontologies, the DAVID GO biological processes
294 associated with this cluster included cell migration ($p=1.15E-04$), extracellular matrix
295 organization ($p=1.78E-04$), regulation of cell adhesion ($p=7.87E-05$) and wound healing
296 pathways ($p=3.84E-03$). By contrast, Cluster 2 was downregulated in EMT condition
297 and upregulated in EMT-MET condition: it included 296 genes with minimal Pearson
298 correlation equal to 0.75. The GO biological processes represented in this cluster were
299 related to cholesterol biosynthetic process ($p=4.42E-20$), lipid metabolism ($1.6E-02$) and
300 metabolic pathways ($p=1.22E-02$) (**Figure 1D, Table S3**). Thus, transcriptional analysis
301 of the EMT and EMT-MET programs induced in the spheroids revealed a list of
302 candidates and associated molecular pathways whose expression was modulated and
303 reversible. Expected pathways related to cell migration and adhesion, known to be
304 central to this reversible phenomenon, validated the biological relevance of our *in vitro*
305 model that also revealed unexpected genes and pathways.

306

307 **Spheroids treated with the EMT-inducing cocktail shift toward the mesenchymal** 308 **consensus molecular subtypes (CMS4)**

309 The CMS classification distinguishes four subtypes of CRC based on the tumor's
310 prominent biological profiles, each subtype having a defined prognosis: the CMS1
311 microsatellite instability/immune subtype (hyper-mutated, microsatellite unstable and
312 strong immune activation), the CMS2 canonical subtype (epithelial with marked WNT
313 and MYC signaling activation), the CMS3 metabolic subtype (epithelial and neat
314 metabolic dysregulation), and the CMS4 mesenchymal subtype (prominent TGF β
315 activation, stromal invasion and angiogenesis) [36]. We surmised that genes expressed
316 in the CMS4 mesenchymal class may be enriched in the EMT-inducing cocktail treated
317 spheroids, but neither in the Baseline nor in the EMT-MET spheroids.

318 The CMS classification applied to the 3 cell culture conditions (Baseline, EMT
319 and EMT-MET conditions) defined a CMS score for each subtype (CMS1, CMS2,
320 CMS3, and CMS4) and for each cell culture condition (Baseline, EMT and EMT-MET
321 conditions). As shown in **Figure 1E**, the prevalent score of Baseline samples
322 corresponded to the CMS3 class (mean CMS3 score = 0.332). This result is coherent
323 with a previous CMS classification of HT29 cells [37]. When compared with Baseline
324 conditions, EMT samples showed a decrease of the CMS3 score ($p=1.22E-02$) and an
325 increase of the CMS4 score ($p=9.58E-03$). Enrichment in CMS4 score (*i.e.*
326 mesenchymal genes) was concordant with the induction of EMT. After the EMT-
327 inducing cocktail removal (EMT-MET samples), the score for the CMS4 class
328 decreased, as expected, while the score for the CMS2 class increased compared to the
329 Baseline condition ($p=2.57E-02$). This result showed that a pro-epithelial program
330 (starting with a switch toward the CMS2 subtype) was re-activated once the EMT-
331 inducing cocktail was removed from the culture. Altogether, the results showed that
332 HT29 cells grown as spheroids and exposed to an EMT inducer activated a physiologic
333 pro-mesenchymal transcriptomic program (the shift toward the CMS4 mesenchymal
334 subtype) and that this process was reversible toward a pro-epithelial phenotype when
335 the EMT inducer was removed. This observation further validated the relevance of our
336 *in vitro* model.

337

338 **EMT genes, Bile Acid Metabolism and Cholesterol Homeostasis hallmark gene**
339 **sets are affected during the EMT and EMT-MET programs**

340 To identify the biological pathways reversibly altered during the EMT process, we
341 did a supervised analysis between the different cell culture conditions. Two
342 comparisons were done: EMT *versus* Baseline samples (EMT induction), and EMT-
343 MET *versus* EMT samples (MET induction). From these signatures, we used GSEA
344 applied to the 50-hallmark gene sets from the MSigDB Molecular Signatures Database.

345 Our objective was to identify and select gene sets that showed an opposite pattern of
346 variation in those two comparisons. Sixteen gene sets were differentially regulated at
347 least in one comparison (q-value<1%) (**Figure 2A, grey background**), and three of them
348 were significantly up- or downregulated in the EMT samples and returned to baseline
349 levels in the EMT-MET samples (**Figure 2A, white background**). The “*Epithelio-*
350 *Mesenchymal transition*” gene set was composed of 23 core genes upregulated in the
351 EMT compared to Baseline samples (q-value = 5.61E-03) and 39 core genes
352 downregulated in the EMT-MET compared to EMT samples (q-value < 1.00E-03) (**Table**
353 **S4, columns 3 and 10 respectively, orange background**). The *Bile Acid Metabolism*
354 gene set was composed of 33 core genes downregulated in the EMT compared to
355 Baseline samples (q-value = 9.16E-03) and 33 core genes upregulated in the EMT-MET
356 compared to EMT samples (q-value = 1.62E-03) (**Table S4, columns 3 and 10**
357 **respectively, green background**). The *Cholesterol Homeostasis* gene set was composed
358 of 31 core genes downregulated in the EMT samples compared to Baseline samples (q-
359 value <1.00E-03) and 28 core genes upregulated in the EMT-MET compared to EMT
360 samples (q-value= 2.64E-03) (**Table S4, columns 3 and 10 respectively, yellow**
361 **background**). Among the core genes: 14 genes were retained in the “*Epithelio-*
362 *Mesenchymal transition*” gene set, 17 in the “*Bile Acid Metabolism*” gene set, and 21 in
363 the “*Cholesterol Homeostasis*” gene set, representing a total of 52 genes (**Figure 2B,**
364 **Table S4 column 18**).

365 We built 3 independent metagenes from each of these three core gene sets. As
366 shown in **Figure 2C**, each metagene was robust enough to distinguish the HT29
367 samples according to their respective condition (Baseline, EMT, and EMT-MET): the
368 “*Epithelio-Mesenchymal transition*” metagene score was the highest in the EMT
369 samples compared to the Baseline samples (p-value=3.22E-06) and to the EMT-MET
370 samples (p-value=1.25E-05). The “*Bile Acid Metabolism*” and the “*Cholesterol*
371 *Homeostasis*” metagene scores were the lowest in the EMT samples compared to the

372 Baseline samples (p-value=2.08E-04 and p-value=1.70E-05 respectively) and to the
373 EMT-MET samples (p-value=2.08E-04 and p-value=1.70E-05, respectively). Altogether,
374 this procedure allowed us to identify 52 genes reversibly regulated during the EMT and
375 the EMT-MET programs. Those genes, when separated into three distinct metagenes
376 based on their hallmark ontology, could also individually distinguished an active from an
377 inactive EMT program in tumor cells.

378

379 ***In vitro* derived metagenes show independent prognostic value in colon cancer**

380 Given the known link between EMT and the metastatic process, we assessed the
381 prognostic value of each metagene (positive *versus* negative) in a large gene
382 expression database of untreated human colon cancer samples (**Table S2**). This base
383 included 2,239 clinically annotated primary colon tumors, including 1,837 with available
384 Recurrence-Free Survival (RFS) data. In univariate analysis, the “*Bile Acid Metabolism*”
385 metagene was not prognostic (**Figure 3A**). By contrast, the positive “*Epithelio-*
386 *Mesenchymal transition*” metagene (HR=1.37, [95CI 1.14-1.64]; p=6.28E-04, Wald test)
387 and negative “*Cholesterol Homeostasis*” metagene (HR=1.25, [95CI 1.04-1.49];
388 p=1.64E-02, Wald test) were associated with shorter RFS, and showed independent
389 prognostic value in multivariate analysis (**Figure 3B**). The **Figure 3C** shows the Kaplan-
390 Meier curves for RFS in four groups defined by each significant metagene: the 5-year
391 RFS was 77% (95%CI 73-82) in the “*Epithelio-Mesenchymal transition*”-
392 negative/“*Cholesterol Homeostasis*”-positive group, 71% (95%CI 67-77) “*Epithelio-*
393 *Mesenchymal transition*”-negative/“*Cholesterol Homeostasis*”-negative group, 70%
394 (95%CI 65-75) in the “*Epithelio-Mesenchymal transition*”-positive/“*Cholesterol*
395 *Homeostasis*”-positive group, and 65% (95%CI 60-70) in the “*Epithelio-Mesenchymal*
396 *transition*”-positive/“*Cholesterol Homeostasis*”-negative group (p=4.61E-04; log-rank
397 test). Such prognostic complementarity was confirmed using the likelihood ratio (LR)
398 test (**Table S5**): the “*Cholesterol Homeostasis*” metagene added prognostic information

399 to that provided by “*Epithelio-Mesenchymal transition*” metagene ($\Delta\text{LR-X2}=5.21$,
400 $p=2.24\text{E-}02$), and the “*Epithelio-Mesenchymal transition*” metagene added prognostic
401 information to that provided by the “*Cholesterol Homeostasis*” metagene ($\Delta\text{LR-}$
402 $X2=13.26$, $p=2.72\text{E-}04$).

403

404 ***The model combining the “Epithelial Mesenchymal Transition” and the***
405 ***“Cholesterol Homeostasis” metagenes is a robust prognostic factor for RFS in***
406 ***CRC***

407 Based on these results, we built a multigene model combining the “*Epithelial*
408 *Mesenchymal Transition*” (14 genes) and the “*Cholesterol Homeostasis*” (21 genes)
409 metagenes, thereafter designated “EMT-CHL model”. We randomly split the 1,837
410 informative primary samples in two sets: a training set composed of a quarter of the
411 samples ($n=450$) to build the model, and a validation set composed of the 1,387
412 remaining samples (**Table 2**) to test its robustness. As expected, in the training set, the
413 model showed a difference between the 5-year RFS of the “low-risk” class (75% [CI95
414 69-83]; $n=225$) and the “high-risk” class (66% [CI95 59-74]; $n=225$; $p=9.93\text{E-}03$, log-
415 rank test; (**Figure 3D, left panel**). Importantly, the model maintained its prognostic value
416 in the validation set, with 74% 5-year RFS in the “low-risk” class (CI95 70-78; $n=672$)
417 versus 68% in the “high-risk” class (CI95 64-72; $n=71$; $p=1.68\text{E-}23$, log-rank test)
418 (**Figure 3D, right panel**).

419 The correlation of the “EMT-CHL model”-based classes with clinicopathological
420 variables in the validation set are shown in **Table 2**. When compared with the “low-risk”
421 class, the “high-risk” class was associated with more female ($p=2.0\text{E-}03$, Fisher’s exact
422 test), more pathological stage 3 ($p=3.0\text{E-}04$, Fisher’s exact test), and more CMS1 and
423 CMS4 subtypes ($p=3.5\text{E-}33$, Fisher’s exact test). We then compared the prognostic
424 value of our “EMT-CHL model” with that of other clinicopathological factors in the
425 validation set in univariate and multivariate analyses (**Table 3**). The other variables

426 significant in univariate analysis were the pathological stage ($p=3.1E-10$) and the CMS
427 classification ($p=2.3E-03$, Wald test). In multivariate analysis, the pathological stage
428 and our multigene model ($p=4.0E-02$, Wald test) remained significant but not the CMS
429 classification. Altogether these analyses showed that our “EMT-CHL” model was a
430 robust and independent prognostic factor for RFS in primary colon cancer.

431

432 ***CMap analysis revealed inhibitors and antagonists that might affect the EMT and***
433 ***EMT-MET programs***

434 We then searched for drugs that might affect (promote or inhibit) the plasticity of
435 the EMT and/or EMT-MET programs. We submitted the 35 genes of the EMT-CHL
436 model to Connectivity Map (CMap), which reports the transcriptional responses of
437 human cells to a variety of chemical or genetic perturbations [31]. The resulting
438 connectivity scores reflect the level of agreement between the tested signature and the
439 L1000 profiles/signatures. Analysis was limited to the 10,847 signatures related to
440 treatment compounds with known mechanism of action through 57 cultured human cell
441 lines: 761 signatures had significant correlation (normalized connectivity score (ncs) *i.e.*
442 adjusted p -value <0.001) with the “EMT-CHL” model genes through 47 human cell lines
443 over 309 drugs compounds (representing 145 drug classes; **Table S6**). We aggregated
444 the significant signatures by drug class and computed the average ncs for each class:
445 98 drug classes were significant. **Table 4** shows the top 15 drug classes identified,
446 which includes antagonists/inhibitors of Dopamine receptor, EGFR, Serotonin receptor,
447 PDGFR, BCR-ABL, ABL, Histamine receptor, KIT, Adrenergic receptor, HMG-CoA
448 reductase (HMGCR), SRC, Calcium channel, Tricyclic antidepressant, Estrogen
449 receptor, and VEGFR. All connectivity scores were negatively correlated with our
450 signature, suggesting that these drugs might interfere with the EMT-MET process.

451

452 ***Blocking enzymes from the cholesterol pathway prevents the execution of the***
453 ***EMT-MET program***

454 We first confirmed that HMGCR and DHCR7 were downregulated during the
455 EMT process and came back to a control level during the MET process due to release
456 of the EMT-inducing treatment (**Figure S1**). From the candidates identified above, we
457 chose to test the HMGCR inhibitors, which directly target the cholesterol pathway. We
458 selected two statins, Lovastatin and Simvastatin, and applied them on 2-days-old
459 organoids (**Figure 4A**). Even at high concentration of each statin, we did not observe
460 any change in the phenotypical aspect of the microtissues when added alone (**Figure**
461 **4B, C** for *Simvastatin*) or in co-treatment with the EMT-inducing cocktail (*Data not*
462 *shown*) compared to the EMT-inducing factor alone. However, when statins were added
463 after a pre-treatment with the EMT-inducing cocktail, *i.e.* during the EMT-MET program
464 (**Figure 4D**), we observed that the treated spheroids maintained a loose EMT-like
465 phenotype instead of reversing to dense spheroids, as those observed with the Baseline
466 situation (**Figure 4E, F**). This result showed that inhibiting the cholesterol pathway after
467 EMT induction prevented the completion of the MET program, in the absence of EMT
468 inducing factors, highlighting the importance of the regulation of the cholesterol pathway
469 during the EMT-MET programs.

470

471 Discussion

472 EMT implication in oncogenesis and metastasis has been extensively studied
473 showing the role of many transcription (e.g. TWIST) and secreted factors (e.g.
474 TGF β). However, progress is necessary to better understand this implication and to
475 provide therapeutic targets [9, 22, 23, 38]. The novelty of our study rely on the
476 combination of the broad induction of EMT using an EMT-inducing cocktail, a 3D model
477 and the timeline approach of EMT-MET induction programs.

478 This allowed the identification of two sets of genes reversibly regulated during
479 either the induction of EMT or the EMT-MET program. The two key pathways retained
480 for our model comprised genes from the EMT pathway, as expected, but also more
481 surprisingly, genes from the cholesterol pathway. When combined as a multigene
482 model, they showed an independent prognostic value in a large dataset of patients with
483 primary colorectal cancer, highlighting the clinical relevance of our finding.

484 .

485 When analyzed in detail, we found that the 14 genes related to the EMT pathway
486 and reversibly modified in response to the EMT-inducing cocktail were genes induced
487 by TGF β (*PMEPA1*, *CTGF*, *TAGLN*, *CYR61* ...). They have also been involved in cell
488 migration (*TNC*, *IL8*, *TAGLN*, *FLNA*, *SERPINE1*, *MYL9*, *CYR61*, *GPC1*...) in various
489 cancers. At least four of them promoted colorectal liver metastasis through
490 enhancement of the EMT program (*THBS1*, *TAGLN*, *ITGA2*, *FLNA*). Several of them
491 code for surface glycoproteins or integrin-related molecules (*PMEPA1*, *TAGLN*,
492 *SERPINE1*, *THBS1*, *GPC1*). Some or a combination of these genes might be used as
493 biomarkers for EMT to identify CRC patients potentially more at risk to relapse [39-46].

494 The involvement of the cholesterol pathway inhibition (during EMT) and re-
495 activation (during EMT-MET) is a new and intriguing finding. The 21 genes of the
496 cholesterol metagene code for most (if not all) enzymes responsible for the synthesis of
497 cholesterol (highlighted in red in **Figure 5**). Cholesterol is essential for cell function and

498 viability. It is a component of the plasma membrane and lipid rafts, a signaling molecule
499 (ligand for estrogen-related receptor alpha (ESRRA)), and a precursor for steroid
500 hormones, Vitamin D and bile acids. The bile acid pathway was also found
501 downregulated during EMT and activated during the EMT-MET programs but did not
502 have prognostic value in our study. Many reports have suggested that cholesterol
503 pathway activation is important for stem cells maintenance in tissues but also for other
504 pro-metastatic function, such as proliferation, survival and resistance to treatment [47].
505 All these features are associated with pejorative prognosis. Cholesterol synthesis is
506 enhanced in cancer cells compared to normal cells [48, 49]. Additionally, patients with
507 high cholesterol levels have an increased risk for many cancers [48, 49]. But thus far,
508 clinical trials with HMGCR inhibitors (statins) have had mixed results [50, 51]. In our
509 case, activation of the cholesterol synthesis pathway tended to be of better prognosis.
510 Due to its multiple biological roles including in tissue homeostasis, cholesterol might
511 have several roles depending on conditions, cells type or timing. Changes in one of
512 these variables might reveal different functionalities. Compared to previous results, our
513 dynamic model enriched the “EMT field” by highlighting the role of cholesterol regulation
514 in relationship with the triggering of cell detachment signal and cell plasticity. Phenotypic
515 plasticity may provide cancer cells with increased adaptability and resistance, enabling
516 them to respond to a variety of external cues and physiological stresses, including EMT
517 signals. In addition to highlighting the reversibility of the cholesterol pathway activation
518 during EMT-MET programs, we also showed that *in vitro* blockade of the cholesterol
519 pathway altered the execution of the EMT-MET program. One hypothesis is that, in the
520 context of tumor cell motility and migration, EMT requires lowering the synthesis of
521 cholesterol to ensure the destabilization of lipid rafts and/or to allow more flexibility in
522 cell membrane for migration. In this line, in breast and hepatocellular carcinomas, CD44
523 affinity for cholesterol-rich lipid rafts and retention inside lipid rafts prevent migration [52-
524 56]. Also, statins induce disruption of lipid rafts, impairing tumor cell adhesion, but

525 enhance invasive and metastatic ability of cancer cells [52-56]. The molecular
526 mechanism involved remains however to be fully elucidated. Altogether, these results
527 suggest that beside or through their lipid lowering property, factors such as statins might
528 also influence complex cell biological programs, during oncogenesis such as EMT.

529 Similar findings (simultaneous opposite involvement of EMT and cholesterol
530 pathways) in pancreatic ductal adenocarcinoma (PDAC) revealed that cholesterol
531 pathway inhibition induces TGF β signaling and promotes basal differentiation [21]. More
532 specifically, the disruption of cholesterol biosynthesis by *Nsdhl* knockout (in mice) or
533 treatment with statins (in mice and in humans) induces the switch from a glandular to a
534 squamous PDAC subtype (mesenchymal subtype). In line with this, a previous study
535 suggested that human tumor cells belonging to the quasi-mesenchymal/squamous/
536 basal subtype of PDAC utilize transcriptionally-dominated programs to lose their
537 epithelial phenotype during EMT, while those belonging to the classical/pancreatic
538 progenitor/ADEX subtype rely on protein re-localization to lose their epithelial phenotype
539 during EMT [57]. This recent discovery now reveals that this glandular to squamous
540 PDAC switch can occur *via* the activation of SREBP1, which induces TGF β 1
541 expression, autocrine TGF β -SMAD2/3 signaling and EMT. In line with this major finding,
542 we showed that inducing EMT in our model can switch a cell line from a CMS3 subtype
543 to a CMS4 mesenchymal subtype using a transcriptionally-dominated program [36],
544 connecting this switch to a more pejorative prognosis. Accordingly, we found that the
545 CMS4 subtype [36] was enriched in the “high-risk” prognostic group, *i.e* “EMT high,
546 Cholesterol down”. It is thus possible that a general mechanism promoting
547 basal/squamous/mesenchymal differentiation with the transcriptional downregulation of
548 the cholesterol pathway and the upregulation of TGF β related genes occurs in multiple
549 tumor types (**Figure 5**). Moreover, it was reported that TGF β -induced mitotic defects in
550 proliferating cells are reversible upon its withdrawal, whereas the acquired genomic
551 abnormalities persist, leading to increased tumorigenic phenotypes. Tumor cells

552 differentiation might thus result from a different degree of EMT depending on several
553 factors, genetic and non-genetic such as the microenvironment and the various
554 treatments [58].

555 In parallel, and in line with this hypothesis, it would be of interest to identify
556 patients who received statins prior to their diagnosis of colorectal cancer and test if
557 these patient's tumors are more of the CMS4 subtype. As shown in the PDAC study,
558 this would suggest that the lowering of the cholesterol pathway can indeed influence the
559 subtype of the growing carcinoma, which would be a major discovery in the field. Our *in*
560 *vitro* data however suggested that cholesterol lowering effects, notably when using
561 statins, occurs only during a specific window to antagonize the MET program. We
562 cannot exclude that a feedback loop between EMT and cholesterol pathways might also
563 exist, which will add a new layer of complexity to this already intricate subject.

564 References

565

- 566 1. Wang, J., et al., *Metastatic patterns and survival outcomes in patients with stage IV colon cancer: A*
567 *population-based analysis*. *Cancer Med*, 2020. **9**(1): p. 361-373.
- 568 2. Lamouille, S., J. Xu, and R. Derynck, *Molecular mechanisms of epithelial-mesenchymal transition*.
569 *Nat Rev Mol Cell Biol*, 2014. **15**(3): p. 178-96.
- 570 3. Batlle, E., et al., *The transcription factor snail is a repressor of E-cadherin gene expression in*
571 *epithelial tumour cells*. *Nat Cell Biol*, 2000. **2**(2): p. 84-9.
- 572 4. Cano, A., et al., *The transcription factor snail controls epithelial-mesenchymal transitions by*
573 *repressing E-cadherin expression*. *Nat Cell Biol*, 2000. **2**(2): p. 76-83.
- 574 5. Yoshida, G.J. and H. Saya, *Molecular pathology underlying the robustness of cancer stem cells*.
575 *Regen Ther*, 2021. **17**: p. 38-50.
- 576 6. Bragado, P., et al., *TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-*
577 *beta-RIII and p38alpha/beta signalling*. *Nat Cell Biol*, 2013. **15**(11): p. 1351-61.
- 578 7. Xu, J., S. Lamouille, and R. Derynck, *TGF-beta-induced epithelial to mesenchymal transition*. *Cell*
579 *Res*, 2009. **19**(2): p. 156-72.
- 580 8. Jayachandran, J., H. Srinivasan, and K.P. Mani, *Molecular mechanism involved in epithelial to*
581 *mesenchymal transition*. *Arch Biochem Biophys*, 2021: p. 108984.
- 582 9. Thiery, J.P. and J.P. Sleeman, *Complex networks orchestrate epithelial-mesenchymal transitions*. *Nat*
583 *Rev Mol Cell Biol*, 2006. **7**(2): p. 131-42.
- 584 10. Kanazawa, T., et al., *TLS/FUS-ERG fusion gene in acute lymphoblastic leukemia with*
585 *t(16;21)(p11;q22) and monitoring of minimal residual disease*. *Leuk Lymphoma*, 2005. **46**(12): p.
586 1833-5.
- 587 11. Dissanayake, S.K., et al., *The Wnt5A/protein kinase C pathway mediates motility in melanoma cells*
588 *via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal*
589 *transition*. *J Biol Chem*, 2007. **282**(23): p. 17259-71.
- 590 12. Chruscik, A., V. Gopalan, and A.K. Lam, *The clinical and biological roles of transforming growth*
591 *factor beta in colon cancer stem cells: A systematic review*. *Eur J Cell Biol*, 2018. **97**(1): p. 15-22.
- 592 13. Ren, J., et al., *SFRP1 inhibits epithelial-mesenchymal transition in A549 human lung*
593 *adenocarcinoma cell line*. *Cancer Biother Radiopharm*, 2013. **28**(7): p. 565-71.
- 594 14. Chung, M.T., et al., *SFRP1 and SFRP2 suppress the transformation and invasion abilities of cervical*
595 *cancer cells through Wnt signal pathway*. *Gynecol Oncol*, 2009. **112**(3): p. 646-53.
- 596 15. Duan, H., et al., *TET1 inhibits EMT of ovarian cancer cells through activating Wnt/beta-catenin*
597 *signaling inhibitors DKK1 and SFRP2*. *Gynecol Oncol*, 2017. **147**(2): p. 408-417.
- 598 16. Yue, B., et al., *LncRNA-ATB mediated E-cadherin repression promotes the progression of colon*
599 *cancer and predicts poor prognosis*. *J Gastroenterol Hepatol*, 2016. **31**(3): p. 595-603.
- 600 17. Hu, Y., et al., *Epigenetic suppression of E-cadherin expression by Snail2 during the metastasis of*
601 *colorectal cancer*. *Clin Epigenetics*, 2018. **10**(1): p. 154.
- 602 18. Scheel, C., et al., *Paracrine and autocrine signals induce and maintain mesenchymal and stem cell*
603 *states in the breast*. *Cell*, 2011. **145**(6): p. 926-40.
- 604 19. Tang, Y., et al., *Induction and analysis of epithelial to mesenchymal transition*. *J Vis Exp*, 2013(78).
- 605 20. Liu, Y.L., et al., *Assessing metastatic potential of breast cancer cells based on EGFR dynamics*. *Sci*
606 *Rep*, 2019. **9**(1): p. 3395.
- 607 21. Gabitova-Cornell, L., et al., *Cholesterol Pathway Inhibition Induces TGF-beta Signaling to Promote*
608 *Basal Differentiation in Pancreatic Cancer*. *Cancer Cell*, 2020. **38**(4): p. 567-583 e11.
- 609 22. Fazilaty, H., et al., *A gene regulatory network to control EMT programs in development and disease*.
610 *Nat Commun*, 2019. **10**(1): p. 5115.
- 611 23. Jia, D., et al., *Testing the gene expression classification of the EMT spectrum*. *Phys Biol*, 2019. **16**(2):
612 p. 025002.

- 613 24. Chitcholtan, K., et al., *Differences in growth properties of endometrial cancer in three dimensional*
614 *(3D) culture and 2D cell monolayer*. Exp Cell Res, 2013. **319**(1): p. 75-87.
- 615 25. Melissaridou, S., et al., *The effect of 2D and 3D cell cultures on treatment response, EMT profile and*
616 *stem cell features in head and neck cancer*. Cancer Cell Int, 2019. **19**: p. 16.
- 617 26. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array*
618 *probe level data*. Biostatistics, 2003. **4**(2): p. 249-64.
- 619 27. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad
620 Sci U S A, 1998. **95**(25): p. 14863-8.
- 621 28. Suzuki, R. and H. Shimodaira, *Pvclust: an R package for assessing the uncertainty in hierarchical*
622 *clustering*. Bioinformatics, 2006. **22**(12): p. 1540-2.
- 623 29. Bertucci, F., et al., *EndoPredict predicts for the response to neoadjuvant chemotherapy in ER-*
624 *positive, HER2-negative breast cancer*. Cancer Lett, 2014. **355**(1): p. 70-5.
- 625 30. Eide, P.W., et al., *CMScaller: an R package for consensus molecular subtyping of colorectal cancer*
626 *pre-clinical models*. Sci Rep, 2017. **7**(1): p. 16618.
- 627 31. Subramanian, A., et al., *A Next Generation Connectivity Map: L1000 Platform and the First*
628 *1,000,000 Profiles*. Cell, 2017. **171**(6): p. 1437-1452 e17.
- 629 32. Siddiqui, A., et al., *Thymidylate synthase is functionally associated with ZEB1 and contributes to the*
630 *epithelial-to-mesenchymal transition of cancer cells*. J Pathol, 2017. **242**(2): p. 221-233.
- 631 33. Druzhkova, I., et al., *E-Cadherin in Colorectal Cancer: Relation to Chemosensitivity*. Clin Colorectal
632 Cancer, 2019. **18**(1): p. e74-e86.
- 633 34. Brodaczewska, K.K., et al., *Metastatic renal cell carcinoma cells growing in 3D on polyDlysine or*
634 *laminin present a stemlike phenotype and drug resistance*. Oncol Rep, 2019. **42**(5): p. 1878-1892.
- 635 35. Dang, H.X., et al., *Long non-coding RNA LCAL62 / LINC00261 is associated with lung*
636 *adenocarcinoma prognosis*. Heliyon, 2020. **6**(3): p. e03521.
- 637 36. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer*. Nat Med, 2015. **21**(11): p.
638 1350-6.
- 639 37. Sveen, A., et al., *Colorectal Cancer Consensus Molecular Subtypes Translated to Preclinical Models*
640 *Uncover Potentially Targetable Cancer Cell Dependencies*. Clin Cancer Res, 2018. **24**(4): p. 794-806.
- 641 38. Lavin, D.P. and V.K. Tiwari, *Unresolved Complexity in the Gene Regulatory Network Underlying EMT*.
642 Front Oncol, 2020. **10**: p. 554.
- 643 39. Cheng, M., et al., *FLNA promotes chemoresistance of colorectal cancer through inducing epithelial-*
644 *mesenchymal transition and smad2 signaling pathway*. Am J Cancer Res, 2020. **10**(2): p. 403-423.
- 645 40. Takahashi, Y., et al., *Tumor-derived tenascin-C promotes the epithelial-mesenchymal transition in*
646 *colorectal cancer cells*. Anticancer Res, 2013. **33**(5): p. 1927-34.
- 647 41. Li, J., et al., *The clinical significance of circulating GPC1 positive exosomes and its regulative miRNAs*
648 *in colon cancer patients*. Oncotarget, 2017. **8**(60): p. 101189-101202.
- 649 42. Huang, X., et al., *Snail/FOKK1/Cyr61 Signaling Axis Regulates the Epithelial-Mesenchymal Transition*
650 *and Metastasis in Colorectal Cancer*. Cell Physiol Biochem, 2018. **47**(2): p. 590-603.
- 651 43. Jeong, D., et al., *Cyr61 expression is associated with prognosis in patients with colorectal cancer*.
652 BMC Cancer, 2014. **14**: p. 164.
- 653 44. Bie, Y., et al., *The Crucial Role of CXCL8 and Its Receptors in Colorectal Liver Metastasis*. Dis Markers,
654 2019. **2019**: p. 8023460.
- 655 45. Long, X., et al., *IL-8, a novel messenger to cross-link inflammation and tumor EMT via autocrine and*
656 *paracrine pathways (Review)*. Int J Oncol, 2016. **48**(1): p. 5-12.
- 657 46. Yang, J.D., L. Ma, and Z. Zhu, *SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma:*
658 *facilitates tumour cell proliferation, migration, and invasion by regulating EMT*. J Chemother, 2019.
659 **31**(7-8): p. 408-418.
- 660 47. McCorry, A.M., et al., *Epithelial-to-mesenchymal transition signature assessment in colorectal*
661 *cancer quantifies tumour stromal content rather than true transition*. J Pathol, 2018. **246**(4): p. 422-
662 426.
- 663 48. Vona, R., E. Iessi, and P. Matarrese, *Role of Cholesterol and Lipid Rafts in Cancer Signaling: A*
664 *Promising Therapeutic Opportunity?* Front Cell Dev Biol, 2021. **9**: p. 622908.

- 665 49. Beck, B. and C. Blanpain, *Unravelling cancer stem cell potential*. Nat Rev Cancer, 2013. **13**(10): p.
666 727-38.
- 667 50. Krens, L.L., et al., *Statin use is not associated with improved progression free survival in cetuximab*
668 *treated KRAS mutant metastatic colorectal cancer patients: results from the CAIRO2 study*. PLoS
669 One, 2014. **9**(11): p. e112201.
- 670 51. Shao, Y.Y., et al., *Statin Use Is Associated With Improved Prognosis of Colorectal Cancer in Taiwan*.
671 Clin Colorectal Cancer, 2015. **14**(3): p. 177-184 e4.
- 672 52. Yang, Z., et al., *Cholesterol inhibits hepatocellular carcinoma invasion and metastasis by promoting*
673 *CD44 localization in lipid rafts*. Cancer Lett, 2018. **429**: p. 66-77.
- 674 53. Babina, I.S., et al., *A novel mechanism of regulating breast cancer cell migration via palmitoylation-*
675 *dependent alterations in the lipid raft affiliation of CD44*. Breast Cancer Res, 2014. **16**(1): p. R19.
- 676 54. Donatello, S., et al., *Lipid raft association restricts CD44-ezrin interaction and promotion of breast*
677 *cancer cell migration*. Am J Pathol, 2012. **181**(6): p. 2172-87.
- 678 55. Murai, T., *The role of lipid rafts in cancer cell adhesion and migration*. Int J Cell Biol, 2012. **2012**: p.
679 763283.
- 680 56. Murai, T., et al., *Low cholesterol triggers membrane microdomain-dependent CD44 shedding and*
681 *suppresses tumor cell migration*. J Biol Chem, 2011. **286**(3): p. 1999-2007.
- 682 57. Aiello, N.M., et al., *EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration*. Dev
683 Cell, 2018. **45**(6): p. 681-695 e4.
- 684 58. Comaills, V., et al., *Genomic Instability Is Induced by Persistent Proliferation of Cells Undergoing*
685 *Epithelial-to-Mesenchymal Transition*. Cell Rep, 2016. **17**(10): p. 2632-2647.

686

687 **Acknowledgements**

688 This work has been supported by Inserm/Institut Paoli-Calmettes and grants from the AP-
689 HM AORC Junior 2018, Canceropole PACA 2018, GIRCI Méditerranéenne 2019. A.A. is
690 supported by a postdoctoral fellowship from the "Fondation ARC" (n°PDF20180507565)
691 and the "Fondation de France" (n°00107936). We want to thank C. Lachaud for sharing
692 the chemidoc equipment.

693

694

695 **Figure legends**

696

697 **Figure 1: Characterization and validation of the EMT induced model**

698 Cells were grown in 3D in ULA plate for 2 days (Day-2). At day 0, the EMT and EMT-MET
699 samples were treated with the EMT-inducing cocktail for 5 days. For EMT-MET samples,
700 the spheroids were washed off the EMT-inducing cocktail and grown for 4 more days in
701 normal medium. **A)** Experimental setting. **B)** Representative photos of the spheroids for
702 the baseline, EMT, EMT-MET conditions, at days 0, 3, 5, 7 and 9. **C)** Western Blots of
703 proteins involved in EMT: FIBRONECTIN, ZEB1, VIMENTIN, OCCLUDIN. ACTIN was
704 used as loading control. **D)** Hierarchical clustering of triplicate samples of three conditions
705 i.e. baseline, EMT and EMT-MET and 3,131 genes based on mRNA expression levels with
706 standard deviation (SD)>0.25. Each row represents a gene and each column represents a
707 sample. Robust clusters defined by pvclust with an approximately unbiased p-value>100%
708 and identified with (an asterisk and) the associated Pearson correlation. **E)** The
709 Consensus Molecular Subtype (CMS) classification was applied to mRNA expression
710 profile of each sample by using the CMScaller tool from Eide et al. [21], providing a score
711 for the four CMS subtypes (CMS1, CMS2, CMS3, and CMS4). Box plot comparing the four
712 CMS scores according each condition (baseline, EMT or EMT-MET). The significant
713 differences defined by one-way ANOVA post hoc Tukey test are visualized with an arrow
714 and p-value are indicated as follows. * : $p < 0.05$, **: $p < 0.01$

715

716 **Figure 2. Epithelial to Mesenchymal transition pathways, Cholesterol homeostasis**
717 **and bile acid metabolism are inversely regulated by the addition and the release of**
718 **the EMT-inducing cocktail**

719 **A)** Sixteen gene-sets out from the MSigDB Molecular Signatures Database (GSEA) were
720 differentially regulated in at least one comparison (q-value<1%): Baseline vs EMT or EMT

721 vs EMT-MET. The 3 gene sets inversely regulated in the two conditions were “*Epithelial*
722 *Mesenchymal Transition*”, “*Bile Acid Metabolism*”; “*Cholesterol Homeostasis*” (white
723 background). **B)** Heatmap of metagenes in the nine samples ordered by condition i.e.
724 Baseline, MT and EMT-MET conditions. Metagene of the three gene sets that were
725 significantly and oppositely deregulated in both comparisons were built using the core
726 genes defined in GSEA analysis: “*Epithelial Mesenchymal Transition*” (14 genes); “*Bile*
727 *Acid Metabolism*” (17 genes); “*Cholesterol Homeostasis*” (21 genes). **C)** Box plot
728 comparing the three individual metagenes according each condition (baseline, EMT or
729 EMT-MET). P-values (ANOVA post hoc Tukey test) are indicated as follows *** $p < 0.001$.

730

731

732 **Figure 3: Epithelial to Mesenchymal transition pathways and Cholesterol**
733 **homeostasis have synergic prognostic value in patients.**

734 Patients were classified as positive or negative according to their metagenes scores. **A)**
735 Kaplan-Meier curves for recurrence free survival (RFS) of 1887 CRC patients assessed
736 using the log-rank test, for the 3 metagenes: “*Epithelial Mesenchymal Transition*”, “*Bile*
737 *Acid Metabolism*” and “*Cholesterol Homeostasis*”. **B)** Forest plot showing the hazard ratio
738 for survival events of each GSEA metagene and recurrence-free survival in CRC patients
739 in univariate and multivariate analysis. A ratio greater than one indicates a poor prognosis
740 and a ratio lower than one indicates good prognosis. The black squares correspond to
741 significant genes and the grey ones to non-significant and the lines defined the 95%
742 confidence interval. **C)** Kaplan-Meier curves for RFS according to four groups: *Epithelio-*
743 *Mesenchymal transition*”-negative/“*Cholesterol Homeostasis*”-positive” group (green,
744 $n=452$), “*Epithelio-Mesenchymal transition*”-negative/“*Cholesterol Homeostasis*”-negative
745 group (black, $n=445$), “*Epithelio-Mesenchymal transition*”-positive/“*Cholesterol*
746 *Homeostasis*”-positive group (blue, $n=481$), and “*Epithelio-Mesenchymal transition*”-

747 positive/“*Cholesterol Homeostasis*”-negative group (red, n=459). **D)** Kaplan-Meier curves
748 for survival model derived from the fusion of and Cholesterol Homeostasis metagenes
749 (“*Epithelial Mesenchymal Transition*” and “*Cholesterol Homeostasis*”).) in the training
750 (n=450) and the validation (n=1387) sets.

751

752 **Figure 4: Morphological changes induced with statins depending on the EMT-MET**
753 **status**

754 For these experiments, cells were grown in ULA plate for 2 days (Day-2) as dense
755 spheroids. Two setting were then used. **A-C)** First setting: From day 0 to day 5, spheroids
756 were either cultured with regular media (baseline), or co-treated with the EMT-inducing
757 cocktail plus DMSO or the EMT-inducing cocktail plus simvastatin. **A)** Experimental
758 setting. **B)** Representative photos of the spheroids in baseline, DMSO and Simvastatin
759 conditions. **C)** Quantification of the spheroid size in arbitrary unit (no significant statistical
760 differences).

761 **D-F)** Second setting: From day 0 to day 5, spheroids were treated with the EMT-inducing
762 cocktail. Then, the EMT-inducing cocktail was washed out and replaced with medium,
763 DMSO (0.5 %), Ethanol (0.5 %), Simvastatin (50 μ M), or Lovastatin (50 μ M) for 5 days. **D)**
764 Experimental setting. **E)** Representative photos of the spheroids for each condition. **F)**
765 Quantification of the spheroid size, in arbitrary unit. Mean values (+/- SEM) are
766 represented as histograms.

767

768 **Figure 5: Graphical summary of the potential relationship between EMT activation**
769 **and the cholesterol pathway.**

770 The mevalonate pathway leads to cholesterol synthesis. Cholesterol is involved in multiple
771 biological homeostatic mechanisms (Bile acid, steroid hormones and vitamin D precursor,
772 Membrane stability and rigidity, Lipid raft organization ...). Cholesterol inhibits the

773 synthesis of SREBP1, a transcription factor for the TGF β 1 signaling. TGF β 1 signaling is a
774 key feature of the CMS4 mesenchymal class. Our data suggest that low cholesterol in a
775 pro-EMT environment might favor the differentiation of CMS4 mesenchymal tumors, while
776 high cholesterol favored more epithelial tumors. Genes down regulated in our Cholesterol
777 Homeostasis metagene are written in red. Statins are inhibitors of the HMGCR enzyme.
778
779

780

781 **Supplementary Figure 1: Cholesterol homeostasis regulator HMGCR and DHCR7 are**
782 **down regulated by the EMT-inducing cocktail treatment.**

783 Cells were grown in 3D in ULA plate for 2 days (Day-2). At day 0, the EMT and EMT-MET
784 samples were treated with the EMT-inducing cocktail for 5 days. For EMT-MET samples,
785 the spheroids were washed off the EMT-inducing cocktail and grown for 4 more days in
786 normal medium. Cells are collected and lysed with RIPA buffer before to be subjected to a
787 Western Blot analysis. FIBRONECTINE (FIBRO.) is a mesenchymal marker,
788 CYTOKERATINES (pan CTK) are mostly epithelial markers, both to verify EMT induction
789 by the EMT-inducing cocktail. DHCR7 and HMGCR are regulators of the cholesterol
790 homeostasis pathway. Actin is the loading control.

791

792

793

794

795

796

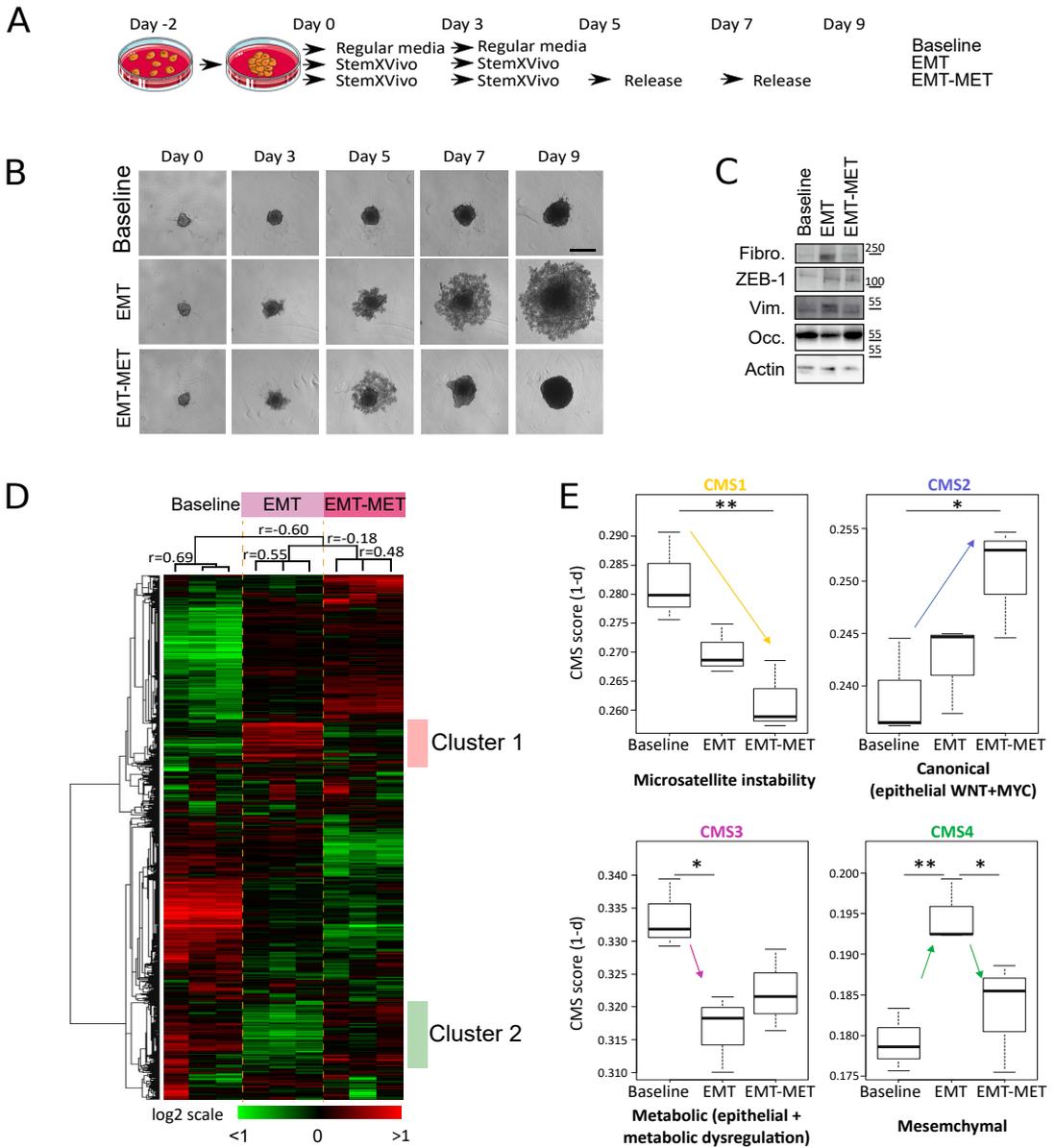


Figure 1

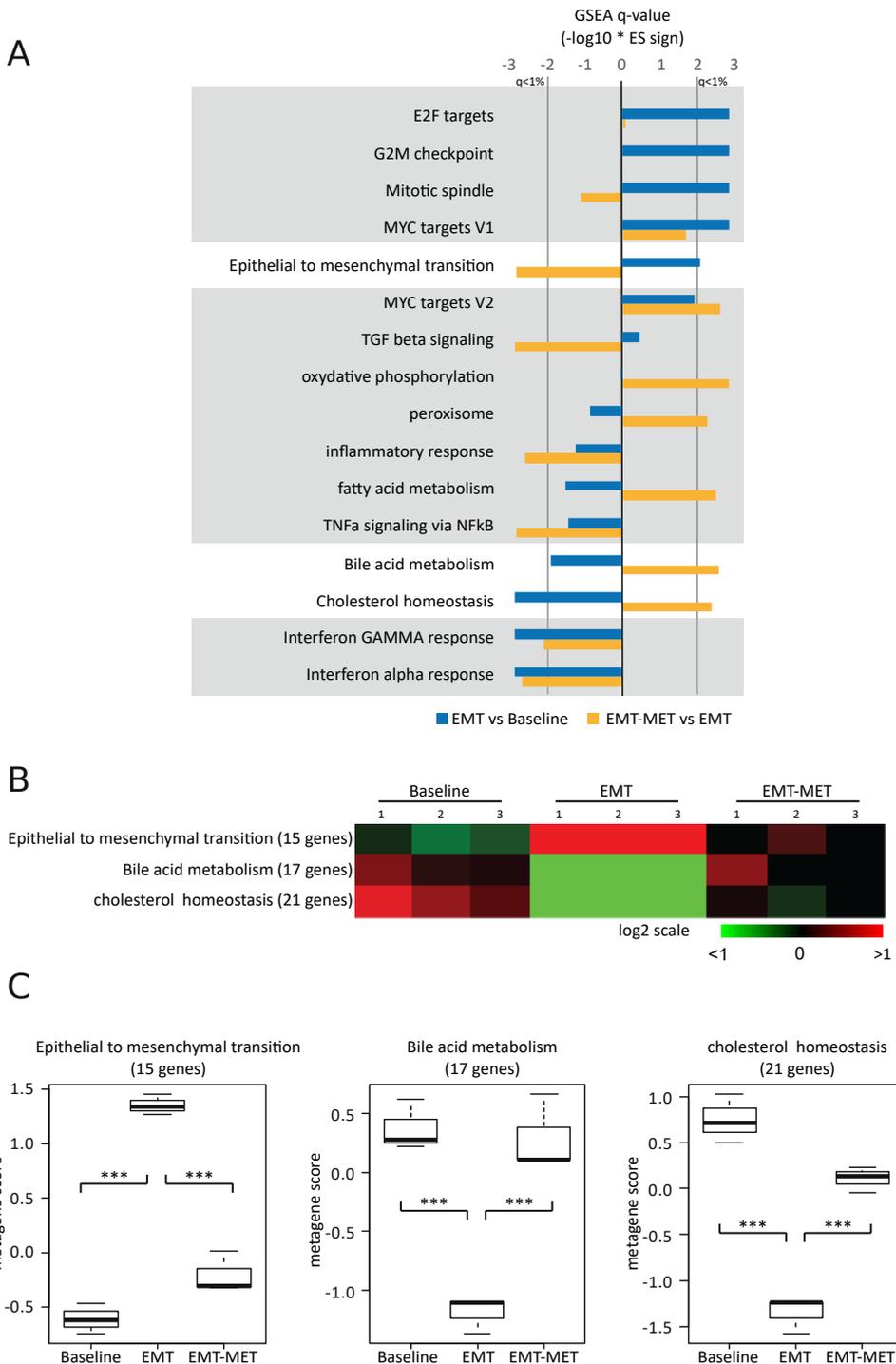


Figure 2

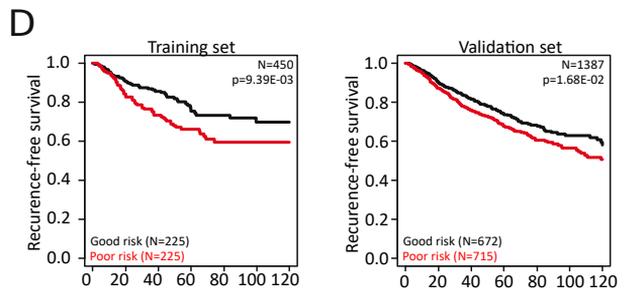
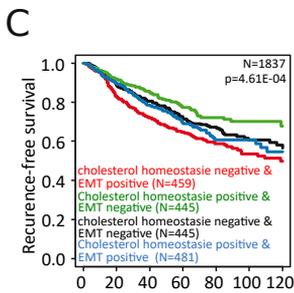
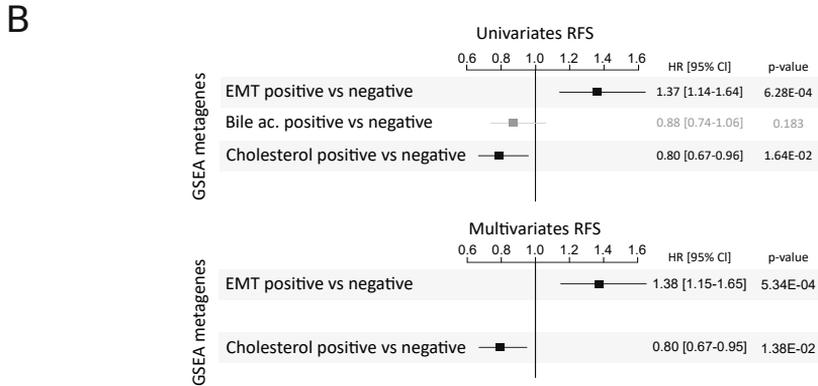
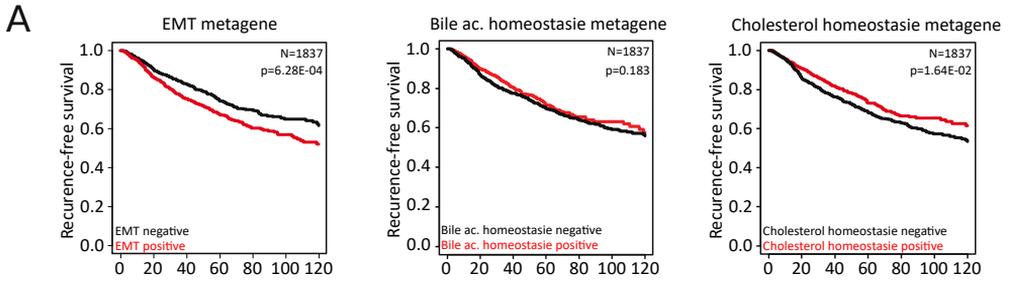


Figure 3

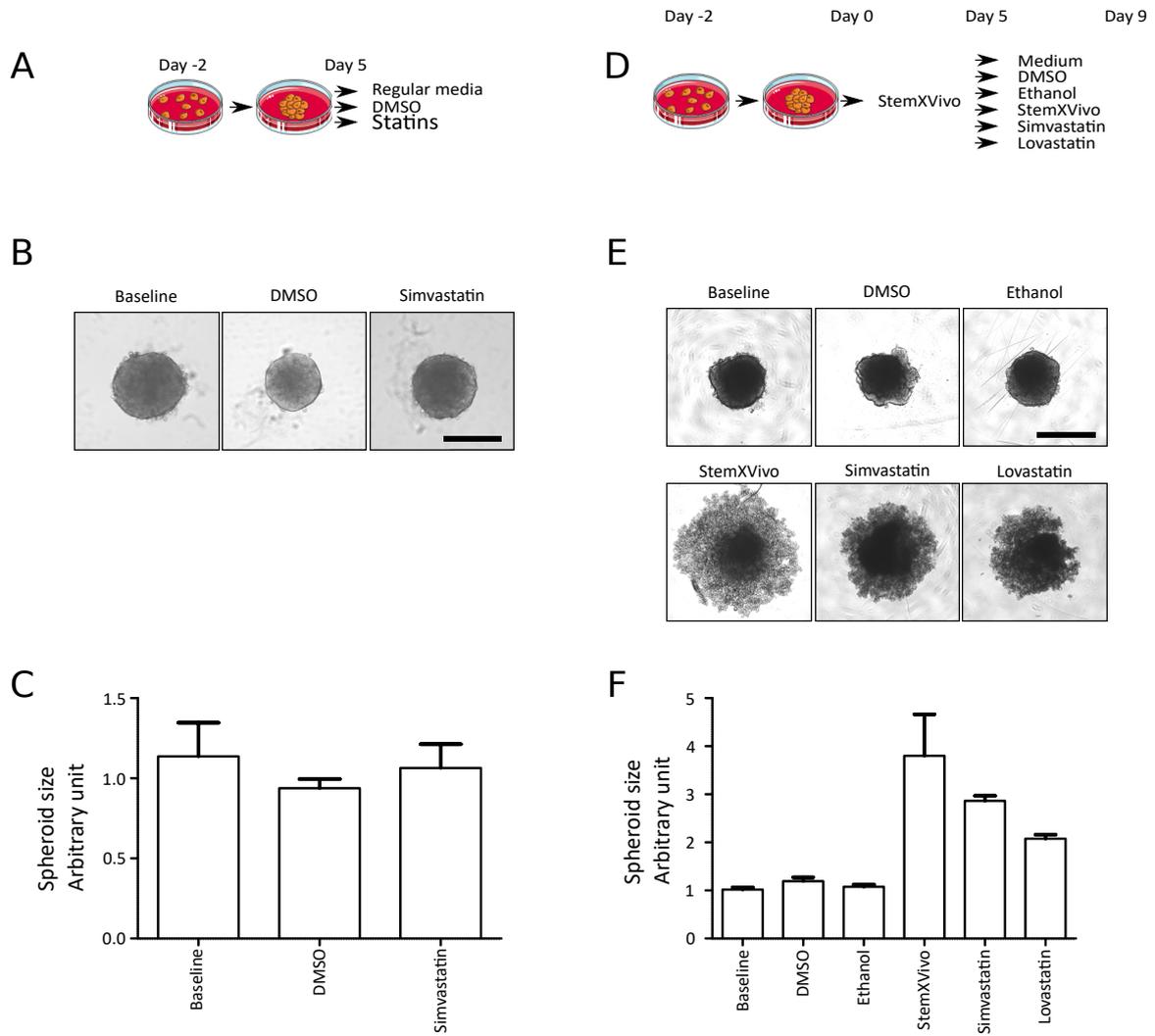


Figure 4

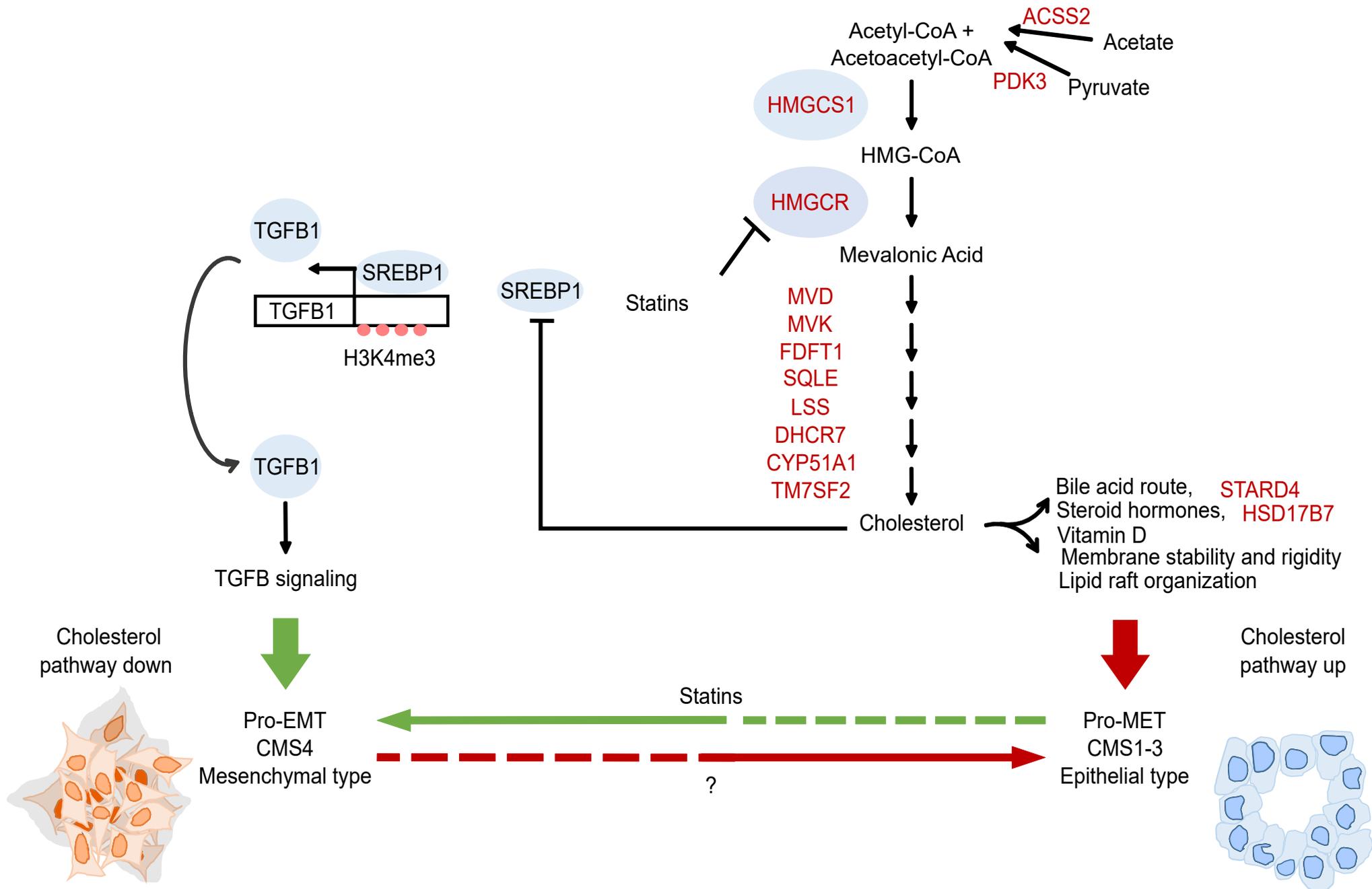


Figure 5

Table 1: (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.
 Clinicopathological characteristics of the 2,239 primary colon cancer samples

Characteristics		N (%)
Age_Diag50		
	<=50	228 (12%)
	>50	1702 (88%)
Sex		
	female	954 (47%)
	male	1069 (53%)
Location		
	distal	840 (50%)
	proximal	854 (50%)
pStage		
	1	167 (10%)
	2	667 (41%)
	3	548 (34%)
	4	237 (15%)
Grade		
	1	26 (6%)
	2	353 (77%)
	3	77 (17%)
CMS		
	CMS1	389 (20%)
	CMS2	640 (32%)
	CMS3	343 (17%)
	CMS4	604 (31%)
Follow-up median, months (range)		43 (1-212)
5y-RFS		71% [68-73]

Table 2: Correlations of the "EMT-CHL model"-based classes with clinicopathological variables

EMT' + 'CHL hom.' Cox model (Validation set)				
Characteristic	N	low-risk	high-risk	p-value
Age_Diag50				0.0507
<=50	139	25 (8%)	114 (13%)	
>50	1064	277 (92%)	787 (87%)	
Sex				0.00201
female	603	127 (40%)	476 (50%)	
male	668	192 (60%)	476 (50%)	
Location_2K				0.921
distal	517	130 (48%)	387 (49%)	
proximal	543	139 (52%)	404 (51%)	
pStage				3.00E-04
1	123	34 (15%)	89 (12%)	
2	466	130 (56%)	336 (44%)	
3	399	67 (29%)	332 (44%)	
4	0	0 (0%)	0 (0%)	
Grade				0.876
1	19	5 (7%)	14 (6%)	
2	259	59 (78%)	200 (80%)	
3	47	12 (16%)	35 (14%)	
CMS				3.50E-33
CMS1	269	52 (19%)	217 (23%)	
CMS2	401	138 (51%)	263 (28%)	
CMS3	212	79 (29%)	133 (14%)	
CMS4	343	3 (1%)	340 (36%)	
Follow-up median, months (range)	1387	42 (1-192)	42 (1-203)	0.387
5y-RFS	1387	74% [70-78]	68% [64-72]	1.68E-02

Table 3: Uni- and multivariate analyses for RFS in the validation set

RFS		Univariate			Multivariate		
		N	HR [95%CI]	p-value	N	HR [95%CI]	p-value
Age_Diag50	>50 vs. <=50	1203	0.98 [0.70-1.37]	0.894			
Sex	male vs. female	1271	1.09 [0.87-1.35]	0.458			
Location	proximal vs. distal	1060	1.00 [0.79-1.26]	0.998			
pStage	2 vs. 1	988	3.48 [1.41- 8.62]	3.09E-10	861	2.53 [1.01-6.33]	4.78E-02
	3 vs. 1		7.74 [3.16-18.93]		861	5.77 [2.35-14.20]	1.36E-04
	4 vs. 1		NA [NA- NA]		861	<NA> [NA-NA]	<NA>
Grade	2 vs. 1	325	1.24 [0.39-3.97]	0.569			
	3 vs. 1		1.70 [0.47-6.12]				
CMS	CMS2 vs. CMS1	1225	0.75 [0.54-1.03]	2.29E-03	861	0.69 [0.43-1.10]	0.121
	CMS3 vs. CMS1		0.91 [0.64-1.30]		861	1.13 [0.68-1.90]	0.636
	CMS4 vs. CMS1		1.28 [0.94-1.73]		861	1.31 [0.87-1.97]	0.196
EMT-CHL model	high- vs. low-risk	1387	1.35 [1.04-1.75]	2.51E-02	861	1.7 [1.02-2.83]	4.02E-02

Table 4:

Top 15 drug classes identified with connectivity map (L1000), 14 EMT vs. 21g CHL GSEA core genes

Drug class (moa)	Number of significant pertubagen signatures	Normalised connectivity score (ncs): average ncs	ncs average 95% CI	reproducibility of ncs per drug class (t-test)		
				t	p.value	score
Dopamine receptor antagonist	93	-1.64	[-1.67,-1.61]	-98.47	4.83E-95	-94.316
EGFR inhibitor	80	-1.65	[-1.74,-1.55]	-33.31	2.86E-48	-47.544
Serotonin receptor antagonist	44	-1.63	[-1.68,-1.58]	-62.43	7.92E-44	-43.101
PDGFR inhibitor	37	-1.65	[-1.71,-1.59]	-57.98	3.75E-37	-36.426
Bcr-Abl inhibitor	33	-1.67	[-1.72,-1.62]	-63.64	2.85E-35	-34.545
Abl inhibitor	33	-1.67	[-1.72,-1.62]	-63.64	2.85E-35	-34.545
Histamine receptor antagonist	33	-1.61	[-1.66,-1.56]	-62.66	4.67E-35	-34.331
KIT inhibitor	34	-1.66	[-1.72,-1.6]	-56.25	2.34E-34	-33.631
Adrenergic receptor antagonist	30	-1.61	[-1.66,-1.55]	-58.08	1.46E-31	-30.836
HMGCR inhibitor / statin	31	-1.67	[-1.74,-1.61]	-53.31	2.79E-31	-30.554
Src inhibitor	30	-1.71	[-1.77,-1.64]	-55.4	5.65E-31	-30.248
Calcium channel blocker	24	-1.66	[-1.72,-1.61]	-61.46	5.13E-27	-26.29
Tricyclic antidepressant	26	-1.7	[-1.77,-1.63]	-48.88	2.45E-26	-25.611
Estrogen receptor antagonist	25	-1.68	[-1.75,-1.62]	-52.34	2.99E-26	-25.524
VEGFR inhibitor	50	-1.63	[-1.78,-1.47]	-21.1	3.04E-26	-25.517