Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue

Received for publication, July 30, 2018, and in revised form, November 7, 2018 Published, Papers in Press, November 8, 2018, DOI 10.1074/jbc.RA118.005087

Nick A. van Huizen, Robert J. R. Coebergh van den Braak, Michael Doukas, Lennard J. M. Dekker, Jan N. M. IJzermans, and Theo M. Luider

From the Departments of Surgery, Pathology, and Neurology, Erasmus University Medical Center, P.O. Box 1738, 3015 GE Rotterdam, The Netherlands

Edited by Xiao-Fan Wang

Changes to extracellular matrix (ECM) structures are linked to tumor cell proliferation and metastasis. We previously reported that naturally occurring peptides of collagen type I are elevated in urine of patients with colorectal liver metastasis (CRLM). In the present study, we took an MS-based proteomic approach to identify specific collagen types that are up-regulated in CRLM tissues compared with healthy, adjacent liver tissues from the same patients. We found that 19 of 22 collagen-α chains are significantly up-regulated (p < 0.05) in CRLM tissues compared with the healthy tissues. At least four collagen-α chains were absent or had low expression in healthy colon and adjacent tissues, but were highly abundant in both colorectal cancer (CRC) and CRLM tissues. This expression pattern was also observed for six noncollagen colon-specific proteins, two of which (CDH17 and PPP1R1B/DARP-32) had not previously been linked to CRLM. Furthermore, we observed CRLM-associated up-regulation of 16 proteins (of 20 associated proteins identified) known to be required for collagen synthesis, indicating increased collagen production in CRLM. Immunohistochemistry validated that collagen type XII is significantly up-regulated in CRLM. The results of this study indicate that most collagen isoforms are up-regulated in CRLM compared with healthy tissues, most likely as a result of an increased collagen production in the metastatic cells. Our findings provide further insight into morphological changes in the ECM in CRLM and help explain the finding of tumor metastasis–associated proteins and peptides in urine, suggesting their utility as metastasis biomarkers.

It is well-known that the loss of the normal structure of the extracellular matrix (ECM), including the major component collagen, has been associated with enhanced tumor proliferation (2, 3). Tumor proliferation changes the function of collagen (4, 5), and it has been shown that these changes are dependent on changes of individual collagen levels, as shown by Nystrom et al. (6) for colorectal cancer (CRC). These changes have been suggested to play a role in chemoresistance and increased cell proliferation becoming a vicious circle. Naba et al. (7) showed differences of ECM proteins in colorectal liver metastasis (CRLM) compared with healthy liver in an underpowered sample set (n = 3). However, a sufficiently powered study regarding the concentration of specific collagen types in CRLM is lacking.

As a result of CRLM, it has been shown that an up-regulation of collagen can be observed in urine (8, 9) and plasma (10). Increased excretion of collagen in urine and plasma may be explained by a higher rate of tissue remodeling and turnover, as matrix metalloproteinases (MMPs) (11–13) are differently expressed in CRLM and CRC compared with healthy liver and colon tissue, respectively (14–16). However, the up-regulation in CRLM of proteins related to collagen production has not been shown.

We aim to study the changes in collagen induced by CRLM with MS in a well-powered data set. We determined which specific collagen types are altered, whether proteins in the collagen turnover pathway are altered, and whether we can define a specific protein footprint of the primary tumor in the metastasis.

Results

Patient characteristics

Patients with CRLM had a median age of 69.2 years (interquartile range (IQR) 60.5–74.7 years) and were mostly male (66.7%). Patients had a median of 1 (maximum 7) tumor, being moderately differentiated in 28 of 30 patients. The CRLM tumor had a median size of 2.4 cm (IQR 1.4–3.5 cm). The primary tumor was located in the colon (47%), rectum (43%), and sigmoid (10%). Patients with CRC had a median overall age of 71.1 years (IQR 54.8–80.9 years). The patients with liver fibrosis (FIB; ranging from moderate to severe fibrosis) had a median...
Collagen in colorectal liver metastasis

Table 1
Top 10 proteins with the highest absolute number of peptides down- or up-regulated in CRLM

<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene name</th>
<th>Fold change&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of peptides&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Down-regulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxymethyl-phosphate synthase (ammonia), mitochondrial</td>
<td>CPS1</td>
<td>&lt;0</td>
<td>135</td>
</tr>
<tr>
<td>Fatty acid synthase</td>
<td>FASN</td>
<td>&gt;0</td>
<td>87</td>
</tr>
<tr>
<td>Spectrin α chain, non-erythrocytic 1</td>
<td>SPTAN1</td>
<td>&gt;0</td>
<td>62</td>
</tr>
<tr>
<td>C-1-tetrahydrofolate synthase, cytoplasmic</td>
<td>MTHFD1</td>
<td>&gt;0</td>
<td>55</td>
</tr>
<tr>
<td>Glycogen debranching enzyme</td>
<td>AGL</td>
<td>&gt;0</td>
<td>46</td>
</tr>
<tr>
<td>Cytosolic 10-formyltetrahydrofolate dehydrogenase</td>
<td>ALDH1L1</td>
<td>&gt;0</td>
<td>44</td>
</tr>
<tr>
<td>Aldehyde oxidase</td>
<td>AOX1</td>
<td>&gt;0</td>
<td>40</td>
</tr>
<tr>
<td>Pyruvate carboxylase, mitochondrial</td>
<td>PC</td>
<td>&gt;0</td>
<td>40</td>
</tr>
<tr>
<td>3-Ketoacyl-CoA thiolase, mitochondrial</td>
<td>ACA2</td>
<td>&gt;0</td>
<td>39</td>
</tr>
<tr>
<td>Alcohol dehydrogenase 4</td>
<td>ADH4</td>
<td>&gt;0</td>
<td>39</td>
</tr>
<tr>
<td><strong>Up-regulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen-α-1(I) chain</td>
<td>COL1A1</td>
<td>&gt;0</td>
<td>11</td>
</tr>
<tr>
<td>Collagen-α-2(I) chain</td>
<td>COL1A2</td>
<td>&gt;0</td>
<td>10</td>
</tr>
<tr>
<td>Collagen-α-1(III) chain</td>
<td>COL3A1</td>
<td>&gt;0</td>
<td>16</td>
</tr>
<tr>
<td>Fibrillin-1</td>
<td>FBN1</td>
<td>&gt;0</td>
<td>1</td>
</tr>
<tr>
<td>Filamin-A</td>
<td>FBLN</td>
<td>&gt;0</td>
<td>3</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>FN1</td>
<td>&gt;0</td>
<td>1</td>
</tr>
<tr>
<td>Myosin-9</td>
<td>MYH9</td>
<td>&gt;0</td>
<td>2</td>
</tr>
<tr>
<td>Plectin</td>
<td>PLEC</td>
<td>&gt;0</td>
<td>5</td>
</tr>
<tr>
<td>Collagen-α-1(XII) chain</td>
<td>COL12A1</td>
<td>&gt;0</td>
<td>2</td>
</tr>
<tr>
<td>Collagen-α-2(V) chain</td>
<td>COL5A1</td>
<td>&gt;0</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup> The number of peptides with a fold change > 0 or < 0 and p < 0.05. The fold change is based on log<sub>10</sub> values.

<sup>b</sup> Number of identified peptides per protein. Gene names in boldface type are involved in liver processes (top half, down-regulated), gene names underlined are related to the ECM (bottom half, up-regulated).

Overall age of 59 years (IQR 58–59 years) and were all male. These patients developed fibrosis due to a combination of different causes: viral hepatitis (3 of 5), HCC (5 of 5), steatosis (1 of 5), and steatohepatitis (2 of 5). Additional information is found in supporting information S4.

Data assessment

Normalization—Normalized values obtained from healthy adjacent liver tissue (control) and CRLM tissue did not differ significantly (p = 0.58).

Permutation—In the true data set of control versus CRLM tissue, 5812 of the 20,635 peptides (28.1%) were tested significantly different. In the permuted data, an average of 631 (3%) peptides were significantly different. The true data set contained more significant peptides than the average of the permutation plus twice the S.D. (5812 ≈ 631; p < 0.05) of the permuted data.

Proteomics

Top 10 proteins down- and up-regulated—In total, 2817 proteins were identified in CRLM and control tissue (see supporting information S8). The 10 proteins with the highest absolute number of down- and up-regulated peptides are shown in Table 1. Many down-regulated proteins were present in liver-specific pathways (Table 1). Furthermore, seven of the top 10 up-regulated proteins in CRLM tissue originated from ECM.

Collagen in CRLM—We then analyzed the number of differentially expressed collagen peptides and collagen-α chains. In CRLM tissue, 1137 collagen peptides were identified, of which 819 peptides (72%) were up-regulated and 55 (4.9%) were down-regulated. Overall, we observed an increased collagen level in CRLM versus control tissue (p < 0.00001, linear -fold change 1.67). Furthermore, 19 of the 22 collagen-α chains were significantly up-regulated (Table 2). COL12A1 was the most differentially expressed collagen alpha chain (2-fold change, 86% of identified peptides (n = 50) significantly up-regulated). An unsupervised clustering was performed based on all collagen peptides that separated both groups, except for four samples (CRLM-5, -9, -10, and -25). The four samples (CRLM-8, -12, -19, and -21) containing up to 50% control tissue were evenly clustered within the CRLM group (Fig. 1). CRLM-10 had much lower MS intensities for collagen peptides than the rest of the CRLM samples. However, in the normalization process, this sample remained within the normal range of samples.

Collagen turnover–related proteins—The observed up-regulation of collagen in CRLM may be due to changes in the normal turnover of collagen in the liver. To test this hypothesis, 65 different proteins included in collagen turnover pathways were retrieved from the "reactome pathway knowledgebase" (17, 18). Twenty-four of these 65 proteins were identified, 17 (~68%) of which were significantly up-regulated in CRLM tissue. Sixteen of the 17 up-regulated proteins are included in the collagen synthesis and/or fibril-forming pathways (Table 3). Supporting information S5 contains additional information.

CRC protein signature in CRLM—Next, we studied the presence of a correlation between collagen levels in CRLM and CRC tissue. For COL10A1, COL12A1, COL14A1, and COL15A1, an absence or low number of unique peptides was observed in healthy colon and healthy liver tissue, but a higher number of unique peptides were present in CRC and CRLM tissues (supporting information S6), being most apparent for COL12A1. Furthermore, the presence of proteins specific for colon tissue in CRLM and control tissues was studied. Elevated expression of 165 genes was found in colon tissue, and in CRLM tissue, 25 corresponding proteins were identified (Table 4). Supporting information S7 contains additional information. Six proteins matched the set criteria and were considered unique for colon tissue versus control liver tissue and were found to be significantly higher in CRLM tissue than control tissue.
Immunohistochemistry

MS data were cross-validated by IHC staining of COL12A1 in CRLM and control tissue. CRLM stroma tissue stained positive for COL12A1 (Fig. 2, a and b), except for three tissues (CRLM-19, -20, and -28; Fig. 2c). Control tissue did not show positive staining (Fig. 2d). The three negatively stained CRLM tissue sections did cluster with the rest of the CRLM samples by proteomics (Fig. 1). Based on summed peptide ranking, the

Table 2

<table>
<thead>
<tr>
<th>Collagen types</th>
<th>No. of peptides</th>
<th>-Fold change</th>
<th>p value</th>
<th>-Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;0</td>
<td>&lt;0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P02452 (COL01A1)</td>
<td>286</td>
<td>239</td>
<td>1.84 × 10^{-09}</td>
<td>1.7</td>
</tr>
<tr>
<td>P08123 (COL01A2)</td>
<td>189</td>
<td>158</td>
<td>9.4 × 10^{-10}</td>
<td>1.7</td>
</tr>
<tr>
<td>P02458 (COL02A1)</td>
<td>5</td>
<td>1</td>
<td>2.4 × 10^{-03}</td>
<td>1.2</td>
</tr>
<tr>
<td>P02461 (COL03A1)</td>
<td>210</td>
<td>135</td>
<td>7.6 × 10^{-06}</td>
<td>1.5</td>
</tr>
<tr>
<td>P02462 (COL04A1)</td>
<td>5</td>
<td>1</td>
<td>7.0 × 10^{-07}</td>
<td>1.5</td>
</tr>
<tr>
<td>P08572 (COL04A2)</td>
<td>33</td>
<td>24</td>
<td>2.92 × 10^{-12}</td>
<td>2.0</td>
</tr>
<tr>
<td>P25320 (COL04A4)</td>
<td>1</td>
<td>1</td>
<td>5.19 × 10^{-06}</td>
<td>1.9</td>
</tr>
<tr>
<td>P20908 (COL05A1)</td>
<td>31</td>
<td>30</td>
<td>3.15 × 10^{-14}</td>
<td>2.0</td>
</tr>
<tr>
<td>P05997 (COL05A2)</td>
<td>50</td>
<td>49</td>
<td>7.73 × 10^{-16}</td>
<td>2.3</td>
</tr>
<tr>
<td>P25940 (COL05A3)</td>
<td>1</td>
<td>1</td>
<td>6.2 × 10^{-04}</td>
<td>1.6</td>
</tr>
<tr>
<td>P12109 (COL06A1)</td>
<td>35</td>
<td>10</td>
<td>2.8 × 10^{-02}</td>
<td>1.2</td>
</tr>
<tr>
<td>P12110 (COL06A2)</td>
<td>30</td>
<td>8</td>
<td>1.7 × 10^{-03}</td>
<td>1.3</td>
</tr>
<tr>
<td>P12111 (COL06A3)</td>
<td>96</td>
<td>39</td>
<td>6.5 × 10^{-04}</td>
<td>1.3</td>
</tr>
<tr>
<td>Q02388 (COL07A1)</td>
<td>2</td>
<td>2</td>
<td>9.7 × 10^{-11}</td>
<td>2.2</td>
</tr>
<tr>
<td>P27658 (COL08A1)</td>
<td>1</td>
<td>0</td>
<td>1.1 × 10^{-02}</td>
<td>0.7</td>
</tr>
<tr>
<td>Q03692 (COL09A1)</td>
<td>7</td>
<td>6</td>
<td>2.3 × 10^{-08}</td>
<td>1.7</td>
</tr>
<tr>
<td>P12107 (COL11A1)</td>
<td>5</td>
<td>4</td>
<td>3.5 × 10^{-14}</td>
<td>2.0</td>
</tr>
<tr>
<td>Q99715 (COL12A1)</td>
<td>58</td>
<td>50</td>
<td>6.9 × 10^{-14}</td>
<td>2.0</td>
</tr>
<tr>
<td>Q05707 (COL14A1)</td>
<td>46</td>
<td>29</td>
<td>8.8 × 10^{-07}</td>
<td>1.5</td>
</tr>
<tr>
<td>P90059 (COL15A1)</td>
<td>2</td>
<td>2</td>
<td>7.1 × 10^{-08}</td>
<td>1.8</td>
</tr>
<tr>
<td>Q07092 (COL16A1)</td>
<td>2</td>
<td>1</td>
<td>6.7 × 10^{-01}</td>
<td>1.0</td>
</tr>
<tr>
<td>P39060 (COL18A1)</td>
<td>14</td>
<td>1</td>
<td>6.2 × 10^{-01}</td>
<td>1.0</td>
</tr>
</tbody>
</table>

a The number of peptides identified per collagen-α chain.
b The number of peptides with a -fold change >0 or <0 and p < 0.05. The -fold change is based on log_{10} values.
c Indicates the p value at the protein level.
d Indicates the true -fold change at the protein level (not the log_{10} value).

These proteins are identified with one peptide, and their identification is considered as less reliable compared with proteins identified with two or more peptides.
COL12A1 values of the three negative IHC tissues were above 3 × S.D. of the control, and 0, 9, and 18 peptides, respectively, were identified per sample, indicating that by proteomics, these samples can be detected correctly.

The IHC staining of COL12A1 is present in the ECM. However, the staining was not equally distributed in the ECM of the CRLM tissue. Furthermore, studying tissue sections to determine IHC scoring revealed distortions in the CRLM morphology compared with the control tissue (Fig. 2). In addition, CRC and healthy colon tissue (control C) were stained. CRC tissues showed, similar to CRLM, staining of the ECM for COL12A1. Staining was also observed in the nuclei of epithelial cells, with repetitive staining. We could not validate the staining of the nuclei of the epithelial cells with another antibody (HPA009143) which stains COL12A1, also showed cross-reactivity. Both antibodies showed exactly the same staining for collagen.

### Discussion

In this study, we focused on the presence of collagen in CRLM, and we showed by using proteomics that we could distinguish all 30 CRLM samples correctly from normal liver tissue in a relative detailed protein level. Twenty-two of the 44 known collagen-α chains were observed in CRLM and/or control tissue (healthy adjacent liver tissue). The remaining 22 chains were not identified, possibly due to low concentration or absence in the liver.

Previously, Naba et al. (7) showed the presence of specific collagen-α chains in colon, CRC, liver, and CRLM tissue in a small sample set (n = 3) including a pooled control (n = 2). Naba et al. (7) observed more collagen types, although they performed additional enrichment and extensive fractionation. The authors indicated that validation needs to be performed in a larger sample set, which was confirmed by our power calculation with an estimation of (minimum) 25 samples/group. In the present study, we analyzed matched control and CRLM tissue from 30 individuals, providing a well-powered data set. An up-regulation of collagen peptides (p < 0.0001) in CRLM tissue compared with control tissue was observed as well as up-regulation of individual collagen-α chains. Three collagen-α chains (COL8A1, COL16A1, and COL18A1) were not significantly affected. COL18A1 is mainly involved in the development of the eye (19), central nervous system, and liver structures (20, 21). Musso et al. (22) demonstrated that COL18A1 production is not increased in CRLM compared with liver tissue and therefore may be considered as a negative control. The low number (two or less) of identified peptides for COL4A4, COL5A3, COL7A1, COL8A1, COL15A1, and COL16A1 indicate relatively low levels in CRLM.

The increase in collagen is accompanied by an increase in proteins related to collagen synthesis, of which several have been described separately to stimulate tumor proliferation: P4HA1 (23, 24), PCOLCE (24), PLID1 (24), PLID2 (25), PLOD3 (24), and SERPINH1 (26). The number of proteins identified from the collagen degradation pathway was not sufficient to draw a conclusion regarding down- or up-regulation. Only one collagenase was identified, MMP-9 (-fold change 1.44). MMP-9 cleaves collagen types IV, V, XI, and XIV (11–13). Other MMPs might also be up-regulated (14, 15), although they were not identified.

COL12A1 MS data were cross-validated with IHC, and both techniques showed significant up-regulation of COL12A1. During scoring of the tissues, no abnormalities were observed in healthy adjacent liver tissue. Nevertheless, the presence of molecular abnormalities cannot be ruled out with certainty. We tried to exclude direct tumor effects by taking control tissue that was at least at a distance of 1 cm away from the tumor. Many articles referring to premetastatic niches have been published (27–30). A premetastatic niche is formed by interaction of the potential metastasis site with proteins excreted from the primary tumor, creating a niche that is favorable for the growth of a metastasis. We were not able to visualize COL12A1 in three CRLM tissue sections with IHC, although increased COL12A1 levels were detected with MS. In CRC, COL12A1 was mainly present in stroma, analogous to CRLM.

However, in colon tissue, it was surprisingly observed inside the nuclei of epithelial cells, with repetitive staining. We could not validate the staining of the nuclei of the epithelial cells with another antibody (HPA009143). This indicates that staining of the nuclei is most likely false positive. This antibody (HPA009143), which stains COL12A1, also showed cross-reactivity. Both antibodies showed exactly the same staining for collagen.
Collagen in colorectal liver metastasis

Table 4
Colon-specific proteins are up-regulated in CRLM

These 25 proteins were identified with proteomics and, according to the Human Protein Atlas, are elevated in colon tissue compared with all other tissue types.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene*</th>
<th>-Fold change*</th>
<th>No. of. peptides†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Villin 1</td>
<td>VILI</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Polymeric immunoglobulin receptor</td>
<td>PIGR</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Fatty acid–binding protein 1</td>
<td>FABP1</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Cadherin 17</td>
<td>CDH17</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Keratin, type I</td>
<td>KRT20</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Carbonic anhydrase II</td>
<td>CA2</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Galectin-4</td>
<td>LGALS4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Carbonic anhydrase I</td>
<td>CA1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>UDP-glucuronosyltransferase 1–8</td>
<td>UGT1A8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Carcinoembryonic antigen-related cell adhesion molecule 5</td>
<td>CEACAM5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>IgGFC-binding protein</td>
<td>FCGBP</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Meprin A subunit α</td>
<td>MEP1A</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Glycoprotein A33</td>
<td>GPA33</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>UDP-glucuronosyltransferase 2B15</td>
<td>UGT2B15</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Mucin 13</td>
<td>MUC13</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Protein phosphatase 1 regulatory subunit 1B</td>
<td>PPP1R1B/DARPP32</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Epidermal growth factor receptor kinase substrate 8-like protein 3</td>
<td>EPSL3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Protein FAM3D</td>
<td>FAN3D</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Carcinoembryonic antigen-related cell adhesion molecule 6</td>
<td>CEACAM6</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Left-right determination factor 1</td>
<td>LEFTY1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sulfratransferase family cytosolic 1B member 1</td>
<td>SULT1B1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Carcinoembryonic antigen-related cell adhesion molecule 7</td>
<td>CEACAM7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Carcinoembryonic antigen-related cell adhesion molecule 8</td>
<td>CEACAM8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Protein-arginine deiminase type-2</td>
<td>PAD2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>UDP-glucuronosyltransferase 1–10</td>
<td>UGT1A10</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

* Genes in boldface type are considered unique for colon tissue.
† Number of peptides that are significantly different (p < 0.05) and have a -fold change >0 or <0. The -fold change is based on log10 values.
‡ Total number of peptides identified for this protein.

Figure 2. IHC staining of COL12A1. a, CRLM-1, IHC score 2; b, CRLM-8, IHC score 1.5; c, CRLM-20, IHC score 0; d, Control-8, IHC score 0; e, IHC scores for COL12A1 in control and CRLM tissue: 0, no staining; 0.5, small focal staining; 1, focal staining; 1.5, few focal staining areas; 2, several focal staining areas; 2.5, staining >40% and <50%; 3, >50% staining; f, FIB-1, IHC score 0; g, Control C-1, antibody NBP1-88062; h, Control C-1, antibody HPA009143, healthy tissue; i, CRC-5, IHC score: 1.
Collagen in colorectal liver metastasis

structures, and they showed different additional cross-reactivity (nuclei endothelial cell and smooth muscle tissue).

The shotgun proteomics data allowed us to search for specific proteins produced by CRLM, which originate from the primary colon tumor tissue. First, we looked into collagen types that were expressed higher in CRC with respect to colon tissue and were also more abundant in CRLM with respect to liver tissue. Four collagen types matched these criteria, COL10A1, COL12A1, COL14A1, and COL15A1, and the abundance was strongest for COL12A1. We hypothesize that these four collagen types visible in CRLM are a sign of origin of the primary tumor and highly specific for CRLM. We cannot exclude the possibility that fibrotic tissue is triggered by the metastatic cells to produce, for example, COL12A1; however, we observed in five fibrotic tissue sections that COL12A1 is not present in just fibrotic tissue itself. Collagen type X (31) and type XII (32) are described as being up-regulated in CRC; however, collagen type XIV and XV are not described. Collagen type X, XII, XIV, and XV are not described in literature in relation with CRLM.

Colon-specific proteins were selected from the Protein Atlas, colon and after applying the selection criteria, six proteins remained. These remaining proteins (CDH17, KRT20, CEACAM5, GPA33, MUC13, and PPP1R1B/DARP-32) are colon-specific and were significantly present in the CRLM tissue. Previous reports state that KRT20 (33), MUC13 (7, 34), CEACAM5 (35), and GPA33 (36, 37) are differently expressed in CRLM tissue, whereas CDH17 and PPP1R1B/DARP-32 have not been described previously. It is likely that CDH17 is involved in cell organization and stimulates tumor proliferation (38), whereas PPP1R1B/DARP-32 is highly expressed in CRC tissue and is a predictor for metastasis (39). Furthermore, six of the 10 top down-regulated proteins are related to liver-specific processes. The relative loss of proteins involved in liver processes in the tumor relates to a known decrease of hepatocytes in tumor tissue (40). The top 10 up-regulated proteins included collagen and two other ECM proteins, fibronectin-1 (FN1) and fibrillin-1 (FN1), suggesting that the ECM composition strongly deviates from normal.

We conclude that collagen is up-regulated in CRLM compared with control tissue and that specific collagen types (COL10A1, COL12A1, COL14A1, and COL15A1) from the primary tumor can be detected in the CRLM tissue as well. The collagen changes found in CRC and CRLM may reflect changes of the ECM related to tumor proliferation and seeding into the liver. These data may help to define new biomarkers to be used in CRLM detection after treatment of the primary tumor or its metastases.

Experimental procedures

Experimental design and statistical rationale

This study was approved by the Erasmus MC review board (MEC-2007-088), and we have worked according to the Declaration of Helsinki. We identified collagen types that are differentially present in paired (n = 30) CRLM tissue compared with normal adjacent liver tissue (control). The presence of colon- or CRC-specific proteins in CRLM was studied by comparing CRLM, control, healthy adjacent colon (n = 5), and CRC tissue (n = 5). Fibrotic tissue (n = 5) was analyzed to determine whether COL12A1 up-regulation is a general liver process or caused by metastatic cells. CRLM, control liver, CRC, and colon tissue were analyzed by IHC (COL12A1 antibody: NB1P-88062) and MS. Fibrotic tissue was analyzed by IHC (COL12A1 antibody: NB1P-88062). Colon tissue was additionally stained with a second COL12A1 antibody (HPA009143) to verify the unexpected staining of epithelial nuclei.

Sample sizes of the CRLM and control tissue were based on a power analysis (α = 0.05, β = 0.20). Mean and S.D. used for the power analysis were calculated on the overall data of log-transformed significant up-regulated collagen peptides in CRLM and control tissue of five patients (control mean = 5.00, CRLM mean = 5.94, pooled S.D. = 1.18). The power analysis depicted a sample size larger than 25 samples/group to determine the observed differences in the subset. No replicate measurements were performed; we assumed the biological variation to be much larger than the technical variation, as has also been described in literature for the technique used (41).

Quality of the LC-MS/MS runs was monitored by measuring samples first on a test system to identify incomplete digestions and determine the presence of other components influencing the chromatography. Furthermore, after every 10 samples, a quality control sample was measured containing a set of peptides spread over the whole gradient to monitor possible retention time shifts or loss of intensity in the mass spectrometer. The peptides showed a retention time variation that remained within 0.2 min.

Data were assessed by analysis of the normalization by a t test, and background was assessed by permutation testing (see “Data assessment”). Proteins included in collagen turnover pathways were analyzed. Statistics are described under “Data analysis” and “Statistical analysis.” Mass spectrometry data were orthogonally cross-validated by IHC on the most distinctive collagen type.

Sample selection criteria

All samples were provided by the Department of Pathology, Erasmus MC, and examined (by a gastrointestinal pathologist) for the presence of healthy and tumor tissue in the same section. Four CRLM sections were included that contained up to 50% control tissue (CRLM-8, -12, -19, and -21); the other sections were free of control tissue. Control liver sections were demonstrated to be free of tumor tissue by standard histological examination and were taken at a minimum distance of 1 cm away from the tumor. All CRC sections included contained nontumorous colon tissue (by standard histological examination). All samples were included in the data analysis. Tissue sections and patient information were handled according to the Federation of Dutch Medical Societies (47). Therefore, no approval of the Medical Research and Ethics Committee was required.

Chemicals

Ultra-HPLC grade nano-LC solvents were obtained from Biosolve (Valkenswaard, The Netherlands). Two COL12A1 antibodies (NB1P-88062 and HPA009143) were obtained from Bio-Connect BV (Huissen, The Netherlands). EnVision Detec-
Collagen in colorectal liver metastasis

Sample treatment for MS

Tissue sections for MS analysis were placed in an Eppendorf cup; one tissue section was analyzed per sample, irrespective of the size of the resection material. Injection volumes were normalized based on the UV absorbance measured on a test system. The tissue sections were deparaffinized and rehydrated, followed by removal of formalin cross-links. Formalin cross-links were removed by incubation with a solution of Tris and RapiGest (42), followed by reduction, alkylation, and overnight trypsin digestion. Trypsin cleaves exclusively C-terminal to lysine or arginine unless proline is located at the C-terminal side (43). The full protocol for sample treatment for MS analysis and nano-LC and MS settings are available in supporting information S1. Mass spectrometry data were made publicly accessible via the PRIDE archive, accession number PXD008383.

Data analysis

MGF peak list files were extracted from raw files by ProteoWizard (version 3.0.9166). MGF files were searched using the Mascot search engine (version 2.3.2, Matrix Science Inc., London, UK) and the UniProt/Swiss-Prot database (20,194 entries), as described by Singh et al. (44). In short, the database search contained a mass tolerance of 10 ppm for the peptide mass and 0.5 Da for the fragment mass. A maximum of 4 missed cleavages was allowed. Hydroxylation (+16 Da) of proline, lysine, and methionine were included as variable modifications, and carbamidomethylation (+57 Da) of cysteine as a fixed modification. Mascot search results were further analyzed by Scaffold (version 4.6.2, Proteome Software, Portland, OR). In Scaffold, protein confidence levels were set to a 1% false discovery rate (FDR), at least 1 peptide/protein, and a 1% FDR at the peptide level. FDRs were estimated by inclusion of a decoy database search generated by Mascot. Subsequently, the Scaffold data were imported into Progenesis QI (version 4, Nonlinear Dynamics, Newcastle-upon-Tyne, UK) to align the LC runs. Features not matching the selection criteria: charge between +2 and +5, and at least two isotopes were excluded from further analysis. Identifications from Scaffold were imported via Scaffold into Progenesis QI, followed by export of the normalized abundance to Excel 2010 (Microsoft, Redmond, WA). Duplicate feature intensities were summed. Data were further processed with Excel, GraphPad Prism (version 5.01, GraphPad Software, Inc., La Jolla, CA), and R (version 3.3.1, R Consortium, Vienna, Austria). Prior to log_{10} transformation, a relatively small value of “10” was added, to include missing values for further data analysis. The log_{10}-transformed data were used for statistics; thereby, the data were assumed to be normally distributed after log transformation. Protein significance was determined by ranked peptide values and summation of all peptides per protein, followed by an unequal variance t test. In all following comparisons, an unequal-variance t test was used unless stated otherwise. p values < 0.05 were considered significant unless stated otherwise.

Data assessment

CRLM and control MS data were assessed; this included the analysis of normalization for injection volumes, permutation testing, and intra- and intercollagen-α chain correlations. To exclude a bias a priori in the data, normalization was tested by summation of all individual peptide values per patient, followed by testing for significance between CRLM and control tissue. To exclude the possibility that results were found by chance in large data sets, permutation testing (n = 1000 iterations) was performed using R. The R script has been added to supporting information S2. Briefly, the data were randomly divided into two groups at the peptide level; significant differences between the two groups were determined using the Wilcoxon signed-rank test. Significant p values were summed per permutation, and the log_{10} was taken. It was assumed that the log_{10} summed significant p values were normally distributed. If the true data set value was greater than the average value of the permutation test plus twice the S.D. (p < 0.05), then a significant difference was assumed.

Statistical analysis

An unsupervised cluster analysis was performed with R on ranked peptides. The R code is available in the supporting information S3. The observed up-regulation of collagen can be due to a disturbance in its normal turnover. To test this hypothesis, proteins included in collagen turnover pathways were retrieved from the reactome pathway knowledgebase (17, 18). Three pathways present in the reactome pathway knowledgebase are related to collagen production and degradation: “collagen biosynthesis and modifying enzymes,” “assembly of collagen fibrils and other multimetric structures,” and “collagen degradation.” Although not included in the reactome pathway knowledgebase, Xaa-Pro dipeptidase (PEPD) is involved in collagen degradation. PEPD cleaves Xaa-Pro and Xaa-Hyp, but not Pro-Pro, and is essential for the final degradation of collagen. PEPD was added to the retrieved list of proteins from the reactome pathway knowledgebase.
Collagen in colorectal liver metastasis

The CRC protein signature in CRLM was determined at two levels: collagen-specific and the general proteome. At the collagen level, the total unique peptide count was used as a measure of abundance. Proteins were considered as “hitchhiked” from the colon to CRLM if they matched the following criteria: 1) they were marked in the Human Protein Atlas (www.proteinatlas.org)³ (45) as tissue-enriched, group-enriched, or tissue-enhanced in colon tissue; 2) they had an intensity-based absolute quantification (iBAQ) value at the protein level for colon tissue but not for liver tissue; and 3) the RNA expression level was 100-fold lower in the liver versus colon tissue, and five peptides or more were identified per protein. iBAQ and RNA expression levels were taken from https://www.proteomicsdb.org³ (46).

Author contributions—The study was designed by L. J. M. D., J. N. M. I., N. A. v. H., R. R. J. C. v. d. B., and T. M. L. Experiments were performed by N. A. v. H., and acquired data were analyzed by N. A. v. H. and M. D. All authors were involved in writing of the manuscript and approved the submitted manuscript.

Acknowledgments—We thank the Erasmus MC Cancer Computational Biology Center for access to their IT-infrastructure and software, which was used for the computations and data analysis in this study.

References

³ Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.


