SUPPLEMENTAL DATA

EXTENDED EXPERIMENTAL PROCEDURES

ECM protein enrichment and immunoblotting

100mg of frozen samples were homogenized using a polytron (Kinematica, Bohemia, NY) in 250μL of buffer C (HEPES pH7.9, MgCl2, KCl, EDTA, Sucrose, Glycerol, Sodium OrthoVanadate) of the CNMCS protein extraction kit. Lysates were then spun for 20 min. at 15,000 rpm at 4°C, the supernatant was saved (intermediate fraction 1, enriched for cytosolic proteins) and the pellet (containing proteins insoluble in buffer C), after a wash, was resuspended in 150μL of buffer N (HEPES pH7.9, MgCl2, NaCl, EDTA, Glycerol, Sodium OrthoVanadate and containing DNase and RNase) and incubated at 4°C for 20 minutes to solubilize nuclear proteins. Protein extract was spun 20 min. at 15,000 rpm at 4°C. The nuclear protein extraction step was repeated twice to allow maximum depletion of nuclear protein from protein extracts. The supernatants (intermediate fraction 2, enriched for nuclear proteins) were pooled and saved and the pellet, resuspended in 100μL of buffer M (HEPES (pH7.9), MgCl2, KCl, EDTA, Sucrose, Glycerol, Sodium deoxycholate, NP-40, Sodium OrthoVanadate), incubated at 4°C for 20 minutes and spun 20 min. at 15,000 rpm at 4°C. The supernatant (intermediate fraction 3, enriched for membrane proteins) was saved and the pellet was finally resuspended in 200μL of buffer CS (Pipes pH6.8, MgCl2, NaCl, EDTA, Sucrose, SDS, Sodium OrthoVanadate), incubated at room temperature for 20 minutes and spun 20 min. at 15,000 rpm. The pellet was finally resuspended in 150μL of buffer C, incubated at 4°C for 20 minutes and spun 20 min. at 15,000 rpm at 4°C. The supernatant from the extraction in buffer CS and from the second extraction in buffer C were pooled (intermediate fraction 4, enriched for cytoskeletal proteins) and saved and the remaining insoluble protein pellet (ECM-enriched fraction) was flash-frozen and kept at -80°C. A 20μL aliquot of total tissue extract and 50μL aliquots of intermediate fractions were mixed with an equal volume of Laemmli Buffer containing 100mM
dithiothreitol. The ECM-enriched, insoluble fraction was resuspended in 3X Laemmli Buffer containing 100mM dithiothreitol. 15µL of each samples were loaded on 4-20% SDS/polyacrylamide gradient gels (Invitrogen). Proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes (Millipore, Billerica, MA). Immunoblots were performed using the following antibodies: rabbit anti-fibronectin produced in our laboratory, mouse anti-transferrin receptor antibody (Invitrogen), mouse anti-vimentin antibody (Sigma, St Louis, MO), rabbit anti-collagen I, mouse anti-GAPDH antibody and rabbit anti-pan-histone antibody (Millipore, Billerica, MA). Following primary antibody incubation, the membranes were washed and incubated in the presence of HRP-conjugated goat anti-rabbit or goat anti-mouse secondary antibody (Jackson ImmunoResearch Laboratory, West Grove, VA). Finally, the membranes were washed and incubated in Western Lightning™ Chemiluminescence Reagent (PerkinElmer LAS, Inc.Boston, MA).

**Domain-based definition of the in silico matrisome**

First, we defined lists of InterPro domains commonly found in i) known ECM glycoproteins and proteoglycans (55 domains, Figure 3), ii) ECM-affiliated proteins (6 domains; namely, syndecan, glypican, semaphorin or Sema, plexin/semaphorin/integrins or PSI, galectin carbohydrate recognition and annexin domains), iii) ECM regulators, including ECM-remodeling enzymes and their regulators (25 domains) and iv) secreted factors (39 domains), including growth factors, cytokines, etc. These domain lists were compiled independently based on previous knowledge, data from the literature and iterative query of UniProt to ensure efficient capture of known candidate proteins (Table S2A). Second, we defined lists of “excluding domains” whose presence excludes a protein from i) ECM glycoproteins and proteoglycans (20 domains), ii) ECM regulators (12 domains), or iii) secreted factors (17 domains). Because of the specificity of the defining domains designed to identify ECM-affiliated proteins, no “excluding domains” were necessary (Table S2B). These “inclusion” and “exclusion” domain lists were refined by iterative cycles to optimize efficiency of both capture and exclusion. It is worth noting that...
the presence of certain domains disqualifies a protein from being part of the matrisome, for example the tyrosine-protein kinase, catalytic domain (IPR020635) or the serine/threonine/tyrosine-protein kinase domain (IPR001245) (Table S2). On the other hand, some domains are excluding proteins for one category and defining domains for another. This is the case for the ADAM-TS Spacer 1 domain, that if present in a protein excludes it from belonging to the core matrisome but serves as a defining domain for ECM regulators.

**Bioinformatic procedures for deriving the in silico matrisome**

The downloadable InterPro index file `protein2ipr.dat` ([ftp://ftp.ebi.ac.uk/pub/databases/interpro/](ftp://ftp.ebi.ac.uk/pub/databases/interpro/), downloaded February 4, 2010), linking UniProt protein entries to InterPro domain information was searched independently for the presence of each set of “inclusion domains”. This was done in parallel for both human and murine protein databases [1]. The resulting lists were highly redundant from a gene perspective, since the UniProt protein database comprises sequences of both intact protein isoforms and fragments. Thus, the collection of UniProt accessions was made gene-centric. Because direct cross-referencing between the protein database UniProt and the gene database Entrez Gene [2] is incomplete, we chose a strategy that relied on GenPept [3] and Ensembl [4] as intermediary protein databases with the best cross-index coverage in both UniProt and Entrez Gene.


Ambiguous cases, arising either from assignment of multiple GenPept accessions to one UniProt accession, from disagreements between GenPept- and Ensembl-based results or from missing gene
name, were resolved by direct protein-to-genomic sequence comparison, using BLAT [5]. The resulting candidate gene lists were used to derive protein sequence data from RefSeq [6] and Ensembl (via the Entrez Gene gene2refseq and gene2ensembl index files) and UniProt (directly via the original accession numbers). UniProt accessions were then searched for the presence of “excluding domains” (via the data in the InterPro file protein2ipr.dat, see above). In the gene-centric representation of the matrisome lists, genes were demoted to non-member status if at least one of their UniProt members had an “excluding domain”. In addition, protein sequences were subjected to two transmembrane prediction programs, TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and Phobius (http://phobius.sbc.su.se/), the latter also predicting the presence of signal peptides within the protein sequence [7, 8]. Results obtained helped guide decisions on non-obvious candidate genes. With the exception of a few known transmembrane collagens, we considered the presence of a transmembrane domain as incompatible with the definition of core matrisome protein.

Finally, genes were assigned to one of two divisions: core matrisome or matrisome-associated and within these divisions to a category, namely, ECM glycoproteins, collagens or proteoglycans within the core matrisome division or ECM-affiliated proteins, ECM regulators or secreted factors within the matrisome-associated division.

A schematic representation of the complete bioinformatic pipeline is presented in Figure S1.

**Proteomic analysis of ECM-enriched samples**

**ECM solubilization and digestion**

100-300 µg ECM-enriched pellets were solubilized and reduced in a solution of 8M urea, 100 mM ammonium bicarbonate, 10 mM dithiothreitol, pH 8 with vortexing at 37°C for 30 min. After cooling to room temperature, cysteines were alkylated by adding iodoacetamide to 25mM for 30 min. After diluting to 2M urea, 100 mM ammonium bicarbonate pH 8.0, samples were deglycosylated with 1000-2000 units of PNGaseF (New England BioLabs, Ipswich, MA) and vortexing at 37°C for 2 hours,
followed by digestion with Lys-C (Wako Chemicals USA, Inc., Richmond, VA), at a ratio of 1:100 enzyme:substrate, with vortexing at 37°C for 2 hours. Final digestion was done using trypsin (Sequencing Grade, Promega, Madison, WI), at a ratio of 1:50 enzyme:substrate, with vortexing at 37°C overnight, followed by a second aliquot of trypsin, at a ratio of 1:100 enzyme:substrate, and an additional 4 hours of incubation. Solutions that began cloudy upon initial reconstitution were clear after overnight digestion. Digests were acidified and desalted using 30mg HLB Oasis Cartridges (Waters Corp., Milford, MA) eluted with 50% acetonitrile, 0.1% trifluoroacetic acid (TFA), followed by concentration in a Speed-Vac.

**Peptide Fractionation by Off-gel Electrophoresis**

~50 µg samples of peptide digest were fractionated using an Agilent 3100 OFFGEL Fractionator (Agilent Technologies, Wilmington, DE) and 13 cm Immobiline Drystrips pH 3-10 (GE Healthcare BioSciences AB, Uppsala, Sweden, 17-6001-14). Fractionation was performed according to the Agilent instruction manual. Briefly, peptides were diluted in IPG buffer, pH 3-10 (GE Healthcare, 17-6000-87), containing 5% glycerol. 150 µL of peptide solution were loaded into each of 12 wells and focused for 20 kV hours with a maximum current of 50 µA and power of 200 mW (24-36 hours). Focused solutions were pipetted out of each well and the wells were re-extracted with 30% acetonitrile/0.1% TFA. Fractions 9 and 10 were typically combined, yielding 11 total fractions for subsequent LC-MS/MS analysis. Fractions were acidified with TFA, cleaned-up using stage tips, (Rappsilber et al., 2003), i.e., pipette tips packed with reversed-phase membrane disks (Empore C-18 #2215, 3M Corporation, St. Paul, MN), eluted with 50% acetonitrile, 0.1% TFA, and then concentrated in a Speed-Vac.

**Mass Spectrometry**

Tryptic digests were analyzed with an automated nano LC-MS/MS system, consisting of an Agilent 1100 nano-LC system (Agilent Technologies, Wilmington, DE) coupled to an LTQ-Orbitrap Fourier
transform mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a nanoflow ionization source (James A. Hill Instrument Services, Arlington, MA). Peptides were eluted from a 10 cm column (Picofrit 75 um ID, New Objectives) packed in-house with ReproSil-Pur C18-AQ 3 um reversed phase resin (Dr. Maisch, Ammerbuch Germany) using either a 120 or 133 min. gradient at a flow rate of 200 nl/min to yield ~20 s peak widths. Solvent A was 0.1% formic acid and solvent B was 90% acetonitrile/0.1% formic acid. The elution portion of the 120 min LC gradient was 3-6% solvent B in 2 min, 6-31% B in 75 min, 31-60% B in 13 min, 60-90% B in 1 min, and held at 90% B for 5 min. The elution portion of the 133 min LC gradient was 3-5% solvent B in 2 min, 5-35% B in 90 min, 35-90% B in 10 min, and held at 90% B for 9 min. Data-dependent LC-MS/MS spectra were acquired in ~3 s cycles; each cycle was of the following form: one full Orbitrap MS scan at 60,000 resolution followed by 8 MS/MS scans in the ion trap on the most abundant precursor ions using an isolation width of 3 m/z. Dynamic exclusion was enabled with a mass width of +/- 25 ppm, a repeat count of 1 and an exclusion duration of 45 sec. Charge-state screening was enabled along with monoisotopic precursor selection and non-peptide monoisotopic recognition to prevent triggering of MS/MS on precursor ions with unassigned charge or a charge state of 1. Normalized collision energy was set to 30 with an activation Q of 0.25 and activation time of 30 ms.

Protein identification, quantitation, and distinction between mouse (stroma) and human (tumor) proteins.

All MS data was interpreted using the Spectrum Mill software package v4.0 beta (Agilent Technologies, Santa Clara, CA). Similar MS/MS spectra acquired on the same precursor m/z within +/- 60 sec were merged, MS/MS spectra with precursor charge >4 and poor quality MS/MS spectra, which failed the quality filter by not having a sequence tag length > 0 (i.e., minimum of two masses separated by the in-chain mass of an amino acid) were excluded from searching. MS/MS spectra were searched against a UniProt database containing either mouse only or both mouse (53,448 entries) and human (78,369
sequences; sequences (including isoforms and excluding fragments) were downloaded from the UniProt web site on June 30, 2010. To each database a set of common laboratory contaminant proteins (73 entries) was appended. Initial search parameters included: ESI linear ion-trap scoring parameters, trypsin enzyme specificity with a maximum of two missed cleavages, 35% minimum matched peak intensity, +/- 20 ppm precursor mass tolerance, +/-0.7 Da product mass tolerance, and carbamidomethylation of cysteines and possible carbamylation of N-termini as fixed/mix modifications. Allowed variable modifications were oxidized methionine, deamidation of asparagine, pyro-glutamic acid modification at N-terminal glutamine, and hydroxylation of proline with a precursor MH+ shift range of -18 to 97 Da. Hydroxyproline was only observed in the proteins known to have it (collagen and proteins containing collagen domains Emilin, etc) and only within the expected GXPG sequence motifs. Supplementary tables 7 and 8 containing the detailed peptide spectral matches might have some examples not in the expected motif when there is either a proline near the motif for which the spectrum could have had insufficient fragmentation to confidently localize the mass change to a particular residue, or a nearby methionine in the peptide and the spectrum had insufficient fragmentation to localize the mass change to oxidized Met or hydroxyproline. When the motif nX[ST] occurs in a peptide in Supplementary tables 7 and 8, this is likely to indicate a site where N-linked glycosylation was removed by the PNGaseF treatment of the sample. While a lowercase n indicates a gene-encoded asparagine residue detected in aspartic acid from, possible mechanisms of modification such as acid-catalyzed deamidation during sample processing versus enzymatic conversion during deglycosylation can not be explicitly distinguished. Identities interpreted for individual spectra were automatically designated as valid by applying the scoring threshold criteria provided below to all spectra derived from each sample in a two step process. First, protein mode was used, which requires 2 or more matched peptides per protein while allowing a range of medium to excellent scores for each peptide. Second, peptide mode was applied to the remaining spectra allowing excellent scoring peptides that are detected as the sole evidence for particular proteins. Protein mode thresholds: protein score >20, peptide (score, Scored
Percent Intensity, delta rank1 - rank2) peptide charge +2: (>7, >60%, > 1.5) peptide charge +3: (>8, >65%, > 1.5) peptide charge +4: (>11, >70%, > 2.5) peptide charge +2: (>6, >90%, > 1.5). Peptide mode thresholds: peptide charge +2 and +3 (>13, >70, > 2) peptide charge +4 (>15, >70, > 2) respectively. The above criteria yielded false discovery rates of 1.0 to 1.7% for each sample at the peptide-spectrum match level and 1.3 to 2.7 % at the distinct peptide level as estimated by target-decoy-based searches using reversed sequences. In calculating scores at the protein level and reporting the identified proteins, redundancy is addressed in the following manner: the protein score is the sum of the scores of distinct peptides. A distinct peptide is the single highest scoring instance of a peptide detected through an MS/MS spectrum. MS/MS spectra for a particular peptide may have been recorded multiple times, (i.e. as different precursor charge states, isolated from adjacent OGE fractions, modified by deamidation at Asn or oxidation of Met) but are still counted as a single distinct peptide. When a peptide sequence >8 residues long is contained in multiple protein entries in the sequence database, the proteins are grouped together and the highest scoring one and its accession number are reported. In some cases when the protein sequences are grouped in this manner there are distinct peptides which uniquely represent a lower scoring member of the group (isoforms, family members, and different species i.e. mouse vs human). Each of these instances spawns a subgroup and multiple subgroups are reported and counted towards the total number of proteins and in Table S5 they are given related protein subgroup numbers (e.g., two fibrillin-1 proteins were identified, the murine form and the human form and are listed as subgroup members 15.1 and 15.2 in the A375 samples, Table S5A). Our in silico matrisome list was then used to categorize all of the identified protein subgroups as being ECM derived or not. The reporting of the number of peptides contributing to each subgroup can be altered by enabling the subgroup-specific option in Spectrum Mill. This was done in Figure S3 to report separately the species-specific peptides, the peptides common to both human and mouse, and the total of common and species-specific peptides.
Relative abundances of proteins were determined using extracted ion chromatograms (XIC’s) for each peptide precursor ion in the intervening high resolution FT-MS scans of the LC-MS/MS runs. An individual protein’s abundance was calculated as the sum of the ion current measured for all quantifiable peptide precursor ions with MS/MS spectra confidently assigned to that protein. Peptides were considered not quantifiable if they were shared across multiple subgroups of a protein or the precursor ions had a poorly defined isotope cluster (i.e. the subgroup-specific and exclude poor isotope quality precursor XIC’s filters in Spectrum Mill were enabled). Proteins were considered quantifiable if they were represented in two independent samples and represented by at least 2 distinct peptides in one of the two samples. The peak area for the XIC of each precursor ion subjected to MS/MS was calculated automatically by the Spectrum Mill software in the intervening high-resolution MS1 scans of the LC-MS/MS runs using narrow windows around each individual member of the isotope cluster. Peak widths in both the time and $m/z$ domains were dynamically determined based on MS scan resolution, precursor charge and $m/z$, subject to quality metrics on the relative distribution of the peaks in the isotope cluster vs theoretical. Although the determined protein ratios are generally reliable to within a factor of 2-fold of the actual ratio, numerous experimental factors contribute to variability in the determined abundance for a protein. These factors may include incomplete digestion of the protein; widely varying response of individual peptides due to inherent variability in ionization efficiency as well as interference/suppression by other components eluting at the same time as the peptide of interest, differences in instrument sensitivity over the mass range analyzed, and inadequate sampling of the chromatographic peak between MS/MS scans.
REFERENCES


SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure S1: Bioinformatic pipeline developed to characterize the \textit{in silico} matrisome; \textit{related to Figure 3 and Table 1.}

For details on the bioinformatic pipeline, see Extended Experimental Procedures.

The dotted arrows indicate manual curation of proteins leading to their reassignment to the ECM-affiliated category. Among these are proteins (or families of proteins) presenting at least one of the 55 domains used to define the core matrisome and none of the 20 excluding domains but which either contain a transmembrane domain or could not be considered as “core matrisome” proteins and were thus reassigned to the ECM-affiliated category of the matrisome-associated division. Examples of these proteins are C-lectins, collectins, ficolins, intelectins, C1q and tumor necrosis-factor-related proteins and mucins.

Supplemental Figure S2: ECM of Normal Murine Colon; \textit{related to Figure 4 and Tables S1B and S4C.}

\textbf{A.} ECM protein enrichment from total tissue sample. The extraction of intracellular components from [1] cytosolic, [2] nuclear, [3] membrane and [4] cytoskeletal fractions was monitored by immunoblotting for histones (nucleus), the transferrin receptor (plasma membrane) and vimentin (cytoskeleton). The remaining insoluble fraction was enriched for ECM proteins (as shown in the collagen I panel) and largely depleted for intracellular components.

\textbf{B.} Characterization of the ECM-enriched fraction from one colon sample by LC-MS/MS. Pie charts display the results from one murine colon sample processed through the proteomics workflow. Proteins represented by at least 2 peptides were included in the analysis. Left panel shows the peptide abundance by precursor-ion MS signal. Middle panel: distribution in terms of numbers of peptides. Right panel:
distribution in terms of numbers of proteins. The core matrisome division comprises the ECM
glycoproteins, collagens and proteoglycans, the matrisome-associated protein division encompasses
ECM-affiliated proteins, ECM regulators and secreted factors. MS data corresponding to this panel are
presented in Table S1B.

C. Mass spectrometry results after peptide separation by off-gel electrophoresis. Pie charts display the
result of one murine colon sample processed through the proteomics workflow. Proteins represented by
at least 2 peptides were included in the analysis. Left panel shows the peptide abundance by precursor
ion MS signal. Middle panel: distribution in terms of numbers of peptides. Right panel: distribution in
terms of numbers of proteins. MS data are presented in Table S4B.

Supplemental Figure S3: Protein sequence alignment of murine and human fibrillin-1; related to
Table 3 and Figure 5.

A. MS data from ECM-enriched fraction extracted from the MA2 tumor sample 1 (see Table S5B).
Numbers of non-distinguishing and species-specific peptides are indicated.

B. Sequence alignment of murine and human fibrillin-1. Peptides identified by LC-MS/MS in the MA2
tumor sample 1 are highlighted in yellow or orange if the peptide is identical in the two species, in blue
if the peptide matches the murine sequence and in pink if the peptide matches the human protein
sequence. The identification of peptides specific to both species indicates that fibrillin-1 is secreted by
both tumor and stromal cells.
SUPPLEMENTAL TABLE LEGENDS

Supplemental Table 1: Mass spectrometric analysis of ECM-enriched fraction from murine lung and colon before peptide separation by off-gel electrophoresis; related to Figure 2A and Figure S2A.

A. Mass spectrometric analysis of ECM-enriched fraction from murine lung (sample 1) before peptide separation by off-gel electrophoresis; related to Figure 2A.

The table shows all proteins identified by at least two peptides. This table also indicates the cellular compartment annotation based on Gene Ontology (column H) and the designations based on our matrisome categorization (column I).

B. Mass spectrometric analysis of ECM-enriched fraction from murine colon (sample 1) before peptide separation by off-gel electrophoresis, related to Figure S2A.

The table shows all proteins identified by at least two peptides. This table also indicates the cellular compartment annotation based on Gene Ontology; note that proteins are found in multiple GO categories (some of them incorrect) and that many have no GO annotations. Proteins listed as Other are not included in the matrisome – many are obviously contaminants from other cellular compartments.

Supplemental Table S2: List of defining and excluding domains used to predict the in silico matrisome; related to Figure 3.

A. List of defining domains for each category of matrisome proteins (core matrisome, ECM-affiliated proteins, ECM regulators and secreted factors).

B. List of excluding domains for each category of matrisome proteins (core matrisome, ECM regulators and secreted factors).

InterPro domain accession numbers and their equivalent in the SMART (http://smart.embl-heidelberg.de/) and/or Pfam (http://pfam.sanger.ac.uk/) databases are indicated.
Supplemental Table S3: The human (A) and murine (B) in silico matrisomes; related to Table 1. 
(*) indicates pseudogenes.
(**) indicate novel ECM glycoproteins
(°) indicates genes added manually (only 3 core matrisome proteins)

Supplemental Table S4: Complete MS data set of ECM-enriched fraction from murine lung and colon after peptide separation by off-gel electrophoresis; related to Figures 2B, 4 and Table 2.
A. Complete MS data set from two independent lung samples, related to Figure 2B and Figure 4A.
B. The lung ECM comprises proteins that were found in the two independent samples analyzed with at least two peptides in one of the two samples (see also Table 2).
C. Complete MS data set from two independent colon samples, related to Figure S2 and Figure 4B.
D. The colon ECM comprises proteins that were found in the two independent samples analyzed and with at least two peptides in one of the two samples (see also Table 2).

Supplemental Table S5: Complete MS data set of ECM-enriched fraction from A375 and MA2 human melanoma tumor xenografts; related to Table 3.
A. Complete MS data set from two independent human A375 melanomas grown subcutaneously in mice.
B. The ECM of A375 tumors comprises proteins that were found in the two independent samples analyzed with at least two peptides in one of the two samples
C. Complete MS data set from two independent human MA2 melanomas grown subcutaneously in mice.
D. The ECM of MA2 tumors comprises proteins that were found in the two independent samples analyzed with at least two peptides in one of the two samples.
Supplemental Table S6: Complete MS data set of ECM-enriched fraction from A375 and MA2 human melanoma tumor xenografts taking into account the origin of matrisome proteins; related to Figure 5.

A. Complete MS data set from two independent human A375 melanomas grown subcutaneously in mice including the origin (human or mouse, column I) of the proteins.

B. The ECM of A375 tumors comprises proteins that were found in the two independent samples analyzed with at least two peptides in one of the two samples. Columns V to AA present species-specific numerical data.

C. Complete MS data set from two independent human MA2 melanomas grown subcutaneously in mice including the origin (human or mouse, column I) of the proteins.

D. The ECM of MA2 tumors comprises proteins that were found in the two independent samples analyzed with at least two peptides in one of the two samples. Columns V to AA present species-specific numerical data.

Supplemental Table S7: Detailed list of all of the confidently identified peptide spectrum matches (PSMs) from the LC-MS/MS runs of the 11 fractions resulting from off-gel electrophoresis of each of the 4 normal tissue samples.

A. Murine lung, sample 1.
B. Murine lung, sample 2.
C. Murine colon, sample .1
D. Murine colon, sample 2.

Supplemental Table S8: Detailed list of all of the confidently identified peptide spectrum matches (PSMs) from the LC-MS/MS runs of the 11 fractions resulting from off-gel electrophoresis of each of the 4 xenograft tumor samples.

A. A375 tumor, sample 1.
B. A375 tumor, sample 2.
C. MA2 tumor, sample 1.
D. MA2 tumor, sample 2.
Make gene-centric, using EntrezGene, GenPept, and Ensembl databases and manual sequence analysis.

For all candidate genes, derive all the UniProt, RefSeq, and Ensembl-specific information.

Positive screen: search UniProt database entries for presence of defining domains.

Domain-based negative screen: eliminate candidate genes based on presence of excluding domains in at least one member UniProt entry.

Transmembrane domain-based negative screen: (TMHMM, Phobius).

Signal peptide-based positive screen (Phobius).

Orthology comparison.

Manual curation: Division and category assignment.

ECM Glycoproteins
Collagens
Proteoglycans
ECM-affiliated Proteins
ECM regulators
Secreted Factors
Core Matrisome
Matrisome-associated
Supplemental Figure S2

A.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Total Tissue Extract</th>
<th>Fractions 1</th>
<th>Fractions 2</th>
<th>Fractions 3</th>
<th>Fractions 4</th>
<th>ECM-enriched Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>190 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>190 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. LC-MS/MS before peptide separation by off-gel electrophoresis

Peptide Abundance:
- 32% Core Matrisome
- 67% Matrisome-associated
- 1% Other

Number of Peptides:
- 962 Total
- 44 ECM-enriched

Number of Proteins:
- 157 Total
- 44 ECM-enriched
- 7 Other

C. LC-MS/MS after peptide separation by off-gel electrophoresis

Peptide Abundance:
- 26% Core Matrisome
- 73% Matrisome-associated
- 1% Other

Number of Peptides:
- 1681 Total
- 187 ECM-enriched

Number of Proteins:
- 88 Total
- 32 ECM-enriched
- 1111 Other