

The Matrisome: *In Silico* Definition and *In Vivo* Characterization by Proteomics of Normal and Tumor Extracellular Matrices*[§]

Alexandra Naba‡ ††, Karl R. Clauser§††, Sebastian Hoersch¶||, Hui Liu‡, Steven A. Carr§**, and Richard O. Hynes†**

The extracellular matrix (ECM) is a complex meshwork of cross-linked proteins providing both biophysical and biochemical cues that are important regulators of cell proliferation, survival, differentiation, and migration. We present here a proteomic strategy developed to characterize the *in vivo* ECM composition of normal tissues and tumors using enrichment of protein extracts for ECM components and subsequent analysis by mass spectrometry. In parallel, we have developed a bioinformatic approach to predict the *in silico* “matrisome” defined as the ensemble of ECM proteins and associated factors. We report the characterization of the extracellular matrices of murine lung and colon, each comprising more than 100 ECM proteins and each presenting a characteristic signature. Moreover, using human tumor xenografts in mice, we show that both tumor cells and stromal cells contribute to the production of the tumor matrix and that tumors of differing metastatic potential differ in both the tumor- and the stroma-derived ECM components. The strategy we describe and illustrate here can be broadly applied and, to facilitate application of these methods by others, we provide resources including laboratory protocols, inventories of ECM domains and proteins, and instructions for bioinformatically deriving the human and mouse matrisome. *Molecular & Cellular Proteomics* 11: 10.1074/mcp.M111.014647, 1–18, 2012.

The extracellular matrix (ECM)¹ is a fundamental and important component of metazoan organisms providing architectural

From the †Howard Hughes Medical Institute, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; §Proteomics Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142; ¶Swanson Biotechnology Center, Bioinformatics and Computing Facility, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ||Bioinformatics Group, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany

✂ Author's Choice—Final version full access.

Received September 26, 2011, and in revised form, November 28, 2011

Published, MCP Papers in Press, December 9, 2011, DOI 10.1074/mcp.M111.014647

¹ The abbreviations used are: ECM, Extracellular matrix; LC-MS/MS, Liquid chromatography-Tandem Mass Spectrometry; OGE, Off-gel Electrophoresis.

support and anchorage for the cells. The ECM consists of a complex meshwork of highly cross-linked proteins and exists as interstitial forms within organs and as specialized forms, such as basement membranes underlying epithelia, vascular endothelium, and surrounding certain other tissues and cell types (e.g. neurons, muscles). Cells adhere to the ECM via transmembrane receptors, among which integrins are the most prominent (1, 2). These cell-matrix interactions result in the stimulation of various signaling pathways controlling proliferation and survival, differentiation, migration, etc. The composition of the ECM and the repertoire of ECM receptors determine the responses of the cells. The biophysical properties of the ECM (deformability or stiffness) have also been shown to modulate these cellular functions (3, 4). In addition to core ECM components (fibronectins, collagens, laminins, proteoglycans, etc.), the ECM serves as a reservoir for growth factors and cytokines and ECM-remodeling enzymes that collaborate with ECM proteins to signal to the cells (5, 6). Hence, the ECM provides not only biophysical cues but also biochemical cues that regulate cell behavior. In addition to being important for normal development, alterations of the ECM have been associated with various pathologies such as fibrosis, skeletal diseases, and cancer (7–9) and it has been emphasized recently that the ECM proteome needs better characterization (10).

The role of the ECM in cancer is of particular interest. Long-standing as well as recent data implicate tumor ECM as a significant contributor to tumor progression. Indeed, the ECM is a major component of the tumor microenvironment (11, 12) and classical pathology has shown that excessive deposition of ECM is a common feature of tumors with poor prognosis. More recently, gene expression screens have revealed that many genes encoding ECM components and ECM receptors are dysregulated during tumor progression (13–16). Finally, modifications of the extracellular matrix architecture and biophysical properties have been shown to influence tumor progression (6, 17, 18). Despite these clear indications that tumor ECM and the interactions of cells with it are very likely to play important roles in tumor progression, we do not have a good picture of ECM composition, origins and functions in tumors. One reason for this lies in the biochemical properties of ECM proteins (large size, insolubility, cross-linking, etc.) that have rendered very challenging at-

tempts to characterize systematically the composition of the ECM from tissues and tumors.

Thanks to the completion of the genomes of many species and to previous studies (19–21), it is now clear that vertebrate genomes contain hundreds of genes encoding ECM proteins. Specific features of ECM proteins have emerged from these studies, in particular their distinctive structures based on the repetition of conserved domains (22, 23). During the last few years, several attempts have been made at *in silico* predictions of the complement of ECM proteins (24–26). Furthermore, recent studies have begun to characterize experimentally the composition of the extracellular matrix of specific model systems such as retinal and vascular basement membranes (27–29), mammary gland (30, 31), and cartilage (32). However, there remains a pressing need for a better definition of the number and diversity of ECM proteins and even of what should be included in that definition. Limitations arise also from the lack of experimental reagents and approaches because of the biochemical intractability of ECM and the lack of an adequate library of antibodies or other probes to characterize ECM proteins *in situ*. Thus, deciphering the complexity of the extracellular matrix *in vivo* represents an important scientific challenge.

We describe here the development of proteomics-based methods coupled with a bioinformatic definition of the “matrisome” (ECM and ECM-associated proteins) to analyze the protein composition of the tissue extracellular matrix. We have successfully applied this strategy to characterize in detail the extracellular matrices both of normal murine tissues (lung and colon) and of melanoma tumors (nonmetastatic and metastatic), which each comprise well over 100 proteins. Moreover, we have applied this approach to understand the origins of tumor ECM proteins and have been able to show, using human into mouse xenograft models, that both tumor cells and stromal cells contribute in characteristic ways to the ECM of the tumor microenvironment. Furthermore, we show that both tumor and stromal cells contribute to significant changes in the extracellular matrices of tumors of differing metastatic potential. The strategy we describe and illustrate here can be broadly applied and we provide protocols and inventories of ECM domains and proteins to facilitate application of these methods by others.

EXPERIMENTAL PROCEDURES

Animal Models, Cell Culture and Human Tumor Cell Xenografts—Normal tissues were from 8- to 12-week-old FVB mice. Lungs were perfused by intracardiac injection with 3 ml of phosphate-buffered saline to remove blood. Colon segments were rinsed with phosphate-buffered saline to remove feces. A375 and MA2 human melanoma cell lines (16) were grown in HyClone high-glucose Dulbecco’s modified Eagle’s medium (Thermo Scientific) supplemented with 2 mM glutamine and 10% fetal bovine serum (Invitrogen, Carlsbad, CA). Eight-week-old NOD/SCID/IL2R γ null (Jackson Laboratory, West Grove, PA) male mice were anesthetized using isoflurane (Abbott Laboratories, North Chicago, IL) and 5×10^5 cells were injected subcutaneously into the left flank of the mouse. Animals were sacrificed 5 weeks

post-injection and the tumors were dissected, flash frozen and kept at -80°C .

Tissue Preparation and ECM Protein Enrichment—Sequential extractions of frozen samples of tissues or tumors were performed using the CNMCS (Cytosol/Nucleus/Membrane/Cytoskeleton) Compartmental Protein Extraction kit (Cytomol, Union City, CA) according to manufacturer’s instructions. In brief, frozen tissues (150–200 mg) or tumors (200–400 mg) were homogenized and extracted sequentially to remove (1) cytosolic proteins (2), nuclear proteins (3), membrane proteins (4), and cytoskeletal proteins leaving a final insoluble fraction enriched for ECM proteins. Fractions were separated on SDS-polyacrylamide gradient gels, transferred to nitrocellulose membranes and probed with antibodies to proteins characteristic of different subcellular compartments (see Fig. 1A and [Extended Experimental Procedures](#)).

Mass Spectrometry—ECM-enriched fractions were solubilized in urea, disulfide bonds reduced and alkylated, and proteins digested with PNGaseF, Lys-C, and trypsin. Solutions that began cloudy upon initial reconstitution were clear after overnight digestion. The resulting peptides were separated by off-gel electrophoresis (OGE) according to isoelectric point and by reversed-phase high-performance liquid chromatography followed by tandem mass spectrometry (MS/MS) on an LTQ Orbitrap mass spectrometer. Mass spectra were interpreted with SpectrumMill and annotated using the matrisome bioinformatics lists developed in this work. MS/MS spectra were searched against a UniProt database containing either mouse only or both mouse (53,448 entries) and human (78,369 entries) sequences; all 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. Peptides identified with a false discovery rate $< 2.5\%$ were assembled into identified proteins, and our *in silico* matrisome list was then used to categorize all of the identified proteins as being ECM derived or not. MS/MS spectra searches allowed for 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 (collagens and proteins containing collagen domains, emilins, etc.) and only within the expected GXP sequence motifs. [supplemental Tables S7 and S8](#) 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 [supplemental Tables S7 and S8](#), this is likely to indicate a site where N-linked glycosylation was removed by the PNGaseF treatment of the sample. Although 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 cannot be explicitly distinguished. Our automated database searching based interpretation of the MS/MS spectra did not attempt to detect any of the many known examples of crosslinking previously observed in collagen family proteins (33, 34) nor did our sample processing methods attempt to enrich for or deplete crosslinked peptides from the samples. Consequently, the spectra generated in this study may be a valuable resource to mine for sites and forms of collagen crosslinking. Additional detailed information can be found in the Extended Experimental Procedures. The raw LC-MS/MS data associated with this manuscript

may be downloaded from ProteomeCommons.org Tranche using the following hash:

onoKvMCC9umP0dW2GiZSGN/DVfz6tqoyRvsA16h351RjGJZiqjk-0wWGYQB9+zPpRj+ASAgKsd779N0Td4250FCPJ6jUAAAAAAAA-BFoQ==.

Bioinformatic Definition of Extracellular Matrix Proteins—The human and mouse proteomes were each screened for proteins containing domains characteristic of ECM proteins, ECM-affiliated proteins, ECM modifiers and secreted factors. Those lists were subsequently screened to eliminate proteins that shared one or more of the defining domains but were not ECM or ECM-associated proteins based on other criteria. Detailed information can be found in the [Extended Experimental Procedures](#). We have also deployed a webpage providing collection of resources (data files, sequence files) and further annotations on the bioinformatic pipeline developed for this study <http://web.mit.edu/hyneslab/matrisome/>.

Immunohistochemistry—Tumor samples were formalin-fixed and paraffin-embedded. Sections were dewaxed and rehydrated following standard procedures. Antigen retrieval was performed by incubating sections in boiling 10 mM sodium citrate buffer (pH6.0) for 20 min. Sections were then blocked with PBS containing 4% ovalbumin. Incubation with rabbit anti-HAPLN1 antibody (Sigma) was performed overnight at 4°C and secondary antibody incubation, 2 h at room temperature. Secondary goat-anti-rabbit antibody conjugated with Alexa-568 was from Invitrogen. Sections were counterstained with DAPI (4',6-diamidino-2-phenylindole) to visualize nuclei.

RESULTS

ECM Protein Enrichment From Tissues—Analysis of the protein composition of the extracellular matrix presents challenges due to the diversity, large size, insolubility and cross-linking of these proteins. By contrast, most other cellular components are soluble even at relatively low concentrations of salt or detergents. Therefore, we took advantage of the insolubility of ECM proteins to enrich for them while depleting other cellular components. We used a subcellular fractionation protocol to extract sequentially components from the cytosol, the nucleus, the membrane and the cytoskeleton and enrich for ECM proteins (see [Extended Experimental Procedures](#)). Fig. 1A shows the sequential extraction of proteins from the different cellular compartments, using diagnostic marker proteins for each compartment. ECM proteins such as fibronectin (as well as laminins and collagens, not shown) were not extracted during these intermediate steps and were found to be enriched in the final insoluble fraction.

Development of a Proteomics-based Strategy to Characterize the Composition of ECM In Vivo—To analyze the composition of the ECM-enriched fractions obtained after depletion of other cellular components, we digested the proteins to peptides and employed a proteomics pipeline shown in Fig. 1B using liquid chromatography combined with tandem mass spectrometry (LC-MS/MS) to identify peptides and proteins (see [Extended Experimental Procedures](#)). Analysis by LC-MS/MS of ECM proteins enriched from murine lung and digested to peptides confirmed a significant enrichment for matrix proteins, with more than 75% of the total precursor ion intensity (the sum of MS1 precursor ion peak areas for all identified peptides) corresponding to proteins defined as

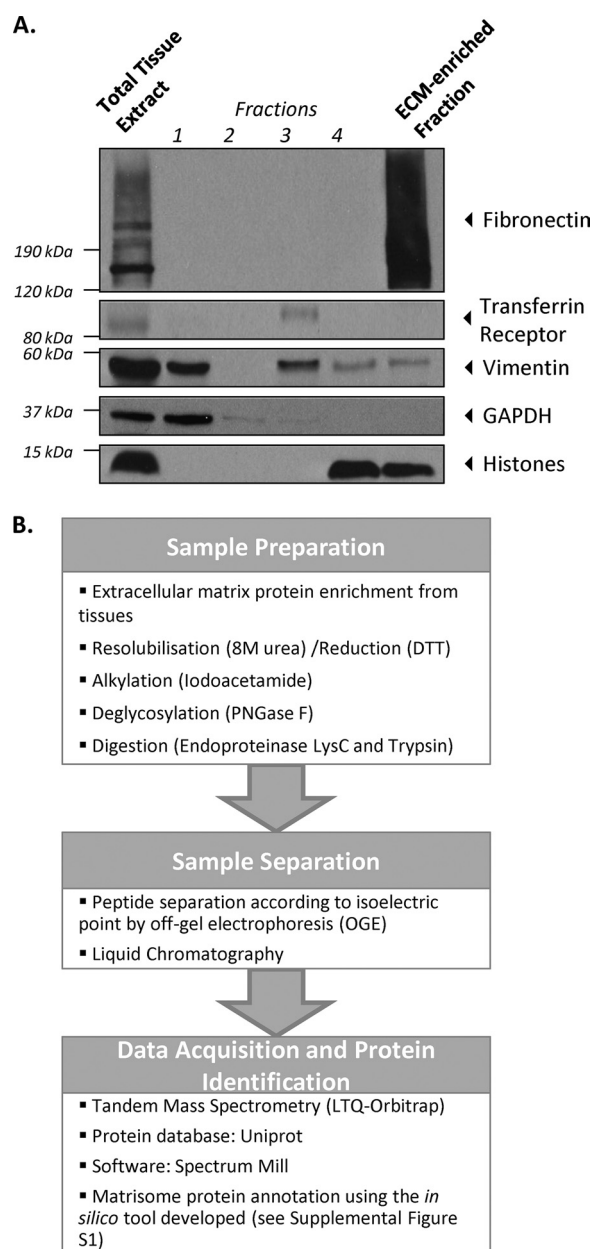
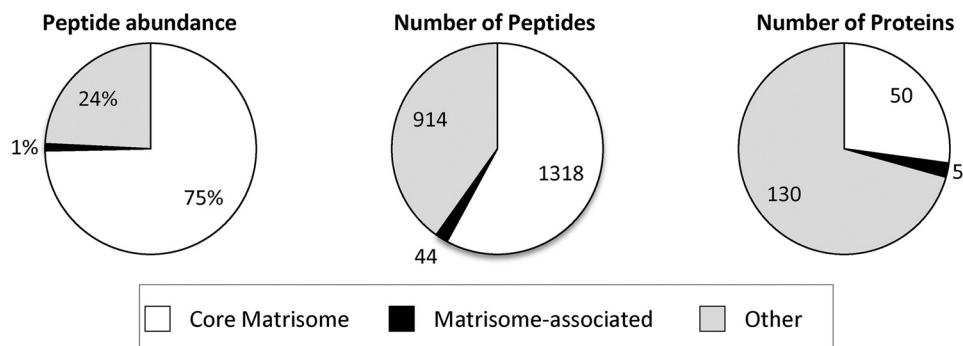


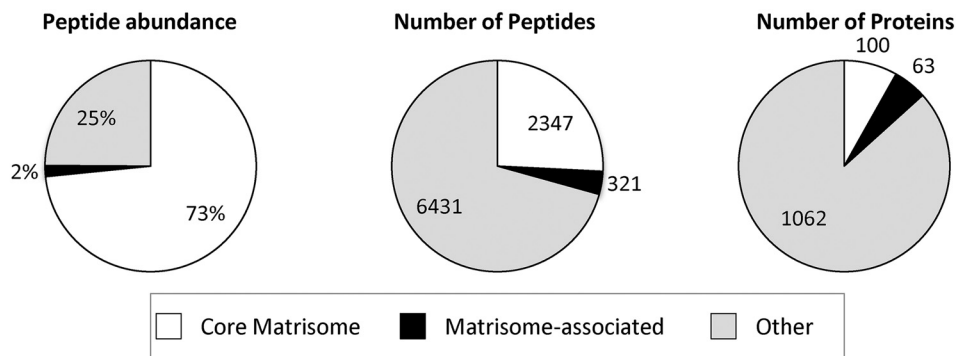
FIG. 1. Extracellular matrix enrichment protocol and proteomic analysis. A, ECM protein enrichment from total tissue sample. The extraction of intracellular components from [1] cytosolic, [2] nuclear, [3] membrane and [4] cytoskeletal fractions was monitored by immunoblotting for GAPDH (cytosol), histones (nucleus), the transferrin receptor (plasma membrane) and vimentin (cytoskeleton). The remaining insoluble fraction was enriched for ECM proteins (as shown in the fibronectin panel) and largely depleted for intracellular components. B, Proteomics workflow (see [Extended Experimental Procedures](#)).

ECM (Fig. 2A, left panel). To help measure the success of our enrichment strategy and focus downstream biological follow-up we sought to categorize the identified proteins as being ECM-derived or not. The categorization of each protein identified by mass spectrometry was initially performed using

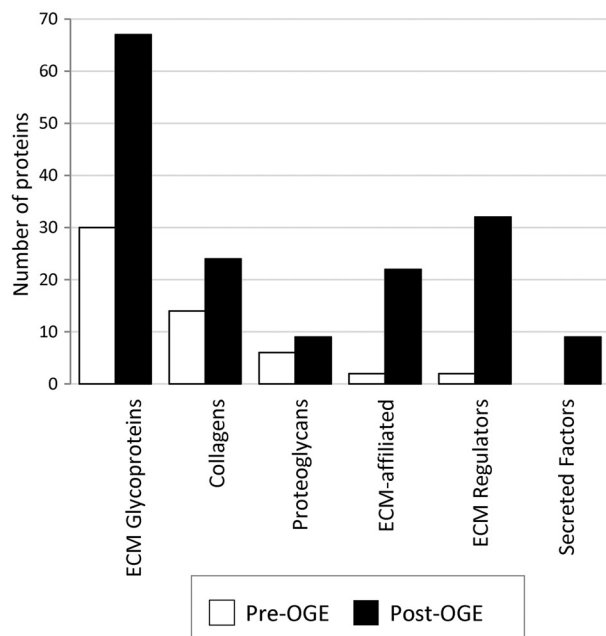
A. LC-MS/MS performed before peptide separation by off-gel electrophoresis



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



C.



D.

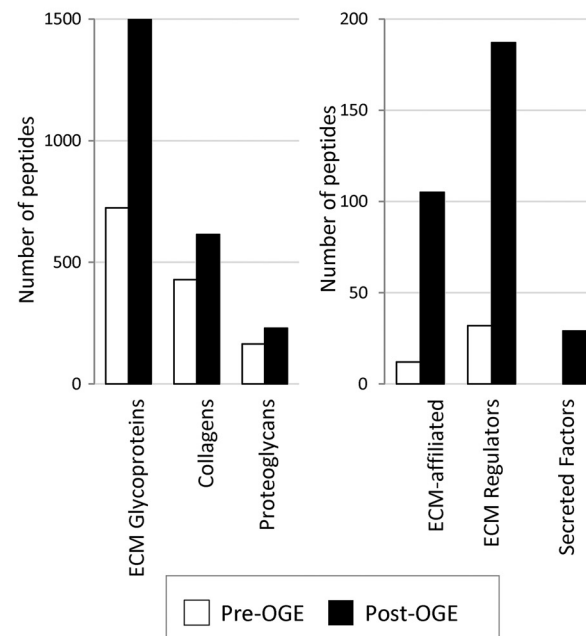


FIG. 2. **Characterization of the ECM of murine lung.** A, Characterization of the ECM-enriched fraction from lung by LC-MS/MS proteomics. The pie charts display the results from one murine lung sample processed through the proteomics workflow. Proteins represented by at least two peptides were included in the analysis. *Left panel* shows the peptide abundance by precursor-ion MS signal corresponding to different protein categories. *Middle panel*: distribution in terms of numbers of peptides. *Right panel*: distribution in terms of numbers of proteins. The “core matrisome” division comprises ECM glycoproteins, collagens and proteoglycans; the “matrisome-associated” protein division encompasses ECM-affiliated proteins, ECM regulators and Secreted factors (see discussion in text and Fig. 3). MS data are presented in [supplemental Table S1](#). B, Mass spectrometry results after peptide separation into 11 fractions by off-gel electrophoresis (OGE). The pie charts

TABLE I

Number of genes encoding the matrisome within the human and mouse genomes. The core matrisome comprises 278 genes in the human genome (274 in the mouse) and includes genes encoding ECM glycoproteins, collagens, and proteoglycans. The subsequent assignment to each category was based on presence of signature domains (for example, the decorin/asporin/biglycan domain is a proteoglycan-specific domain, whereas, the type 4 procollagen, C-terminal repeat is a collagen-specific domain) and on prior knowledge and data from the literature (see [Extended Experimental Procedures](#))

The matrisome-associated set of genes includes: 1) genes encoding ECM-affiliated proteins i.e. proteins that share either some architectural or biochemical similarities with ECM proteins (ficolins, mucins, etc.) or that are known to be associated with ECM proteins, either from prior information or from our own results (semaphorins, annexins, galectins, etc.), 2) ECM regulators, including ECM-remodeling enzymes such as transglutaminases, lysyl oxidases, matrix metalloproteinases, serpins etc. and their regulators and 3) genes encoding secreted factors such as growth factors, cytokines etc.

	Categories	Gene count corrected for pseudo/noncoding genes	
		Human	Mouse
Core matrisome	ECM	200	195
	Glycoproteins		
	Collagens	43	43
Matrisome-associated	Proteoglycans	35	36
	ECM-affiliated	176	164
	Proteins		
	ECM	250	297
	Regulators		
	Secreted	352	363
	Factors		

the Gene Ontology (GO) “Cellular Compartment” annotations. However, this annotation showed several clear limitations. For example, several cytosolic or cytoskeletal proteins involved in cell-matrix adhesion are mis-annotated as being part of the extracellular matrix (see [supplemental Table S1](#)); conversely, some known ECM proteins (thrombospondin 1, von Willibrand factor, agrin, etc.) are defined by vague terms such as “external side of the plasma membrane” or “cell surface” ([supplemental Table S1](#)). In addition more than 20 different GO categories correspond to the extracellular matrix (extracellular matrix, basal lamina, basement membrane, etc.) and yet UniProt identifiers for some known ECM proteins are not associated with any cellular compartments in the Gene Ontology database. Finally, and of importance to the study of human/mouse xenografts (see below), we noted conflicting annotations between human and mouse proteins. In order to interpret the mass spectrometric data we needed a better definition of which proteins should be considered as part of the ECM.

In Silico Definition of the Matrisome—Therefore, we developed a bioinformatic approach to predict within any genome the ensemble of genes encoding what we define as the “matrisome,” namely all those components constituting the extracellular matrix (the “core matrisome”) and those components associated with it (“matrisome-associated” proteins). One hallmark of ECM proteins is their domain-based structure (23). Exploiting this characteristic, we established a list of 55 diagnostic InterPro domains commonly found in ECM proteins (type I, II and III fibronectin domains, type I thrombospondin repeats, laminin G domain, etc.; Fig. 3A and [supplemental Table S2A](#)). This domain list was used to screen the UniProt protein database. We know that some of the domains used to select positively for ECM proteins are also found in transmembrane receptors and proteins involved in cell adhesion (growth factor receptors, integrins, etc) that do not belong to the ECM. These families of proteins also display a subset of specific domains (e.g. tyrosine kinase and phosphatase domains) and transmembrane domains incompatible with definition as “extracellular matrix” proteins. Therefore, a second step comprised a negative selection using 20 domains ([supplemental Table S2B](#)) and a transmembrane domain prediction (see [Extended Experimental Procedures](#) for details). This analysis was performed in parallel for both the mouse and human genomes and the respective murine and human matrisome lists were compared based on orthology. Manual curation of the matrisome lists also allowed us to add

a very few known ECM proteins that do not contain any known domains; for example, dermatopontin and dentin sialophosphoprotein ([supplemental Table S3](#)). Finally, knowledge-based annotation of these gene lists allowed us to define subcategories within the core matrisome; namely, ECM glycoproteins, collagens, and proteoglycans. The bioinformatics workflow developed is presented in Fig. 3B and defined sets of core matrisome proteins from both species are listed in [supplemental Table S3](#).

We also wished to characterize those proteins that are known to be associated with the ECM, but are not included in the core list of ECM proteins. To this end, we defined separate lists of domains commonly found in 1) ECM-affiliated proteins (proteins that share either some architectural similarities with ECM proteins or that are known to be associated with ECM proteins, or that repeatedly appeared in our proteomics

display the result of one murine lung sample processed through the proteomics workflow. Proteins represented by at least two peptides were included in the analysis. *Left* panel shows the peptide abundance by precursor-ion MS signal corresponding to the different divisions of the matrisome. *Middle* panel: distribution in terms of numbers of peptides. *Right* panel: distribution in terms of numbers of proteins. Note the increase in number of matrisome and matrisome-associated proteins detected in each category and the larger increase in nonECM proteins; that is, the increase in apparent “noise” relative to “signal” for ECM proteins. MS data are presented in [supplemental Table S4A](#), sample 1. *C*, Comparison of the number of core matrisome and matrisome-associated proteins detected before and after peptide separation by OGE. *D*, Comparison of the number of peptides belonging to the core matrisome and matrisome-associated proteins before and after OGE.

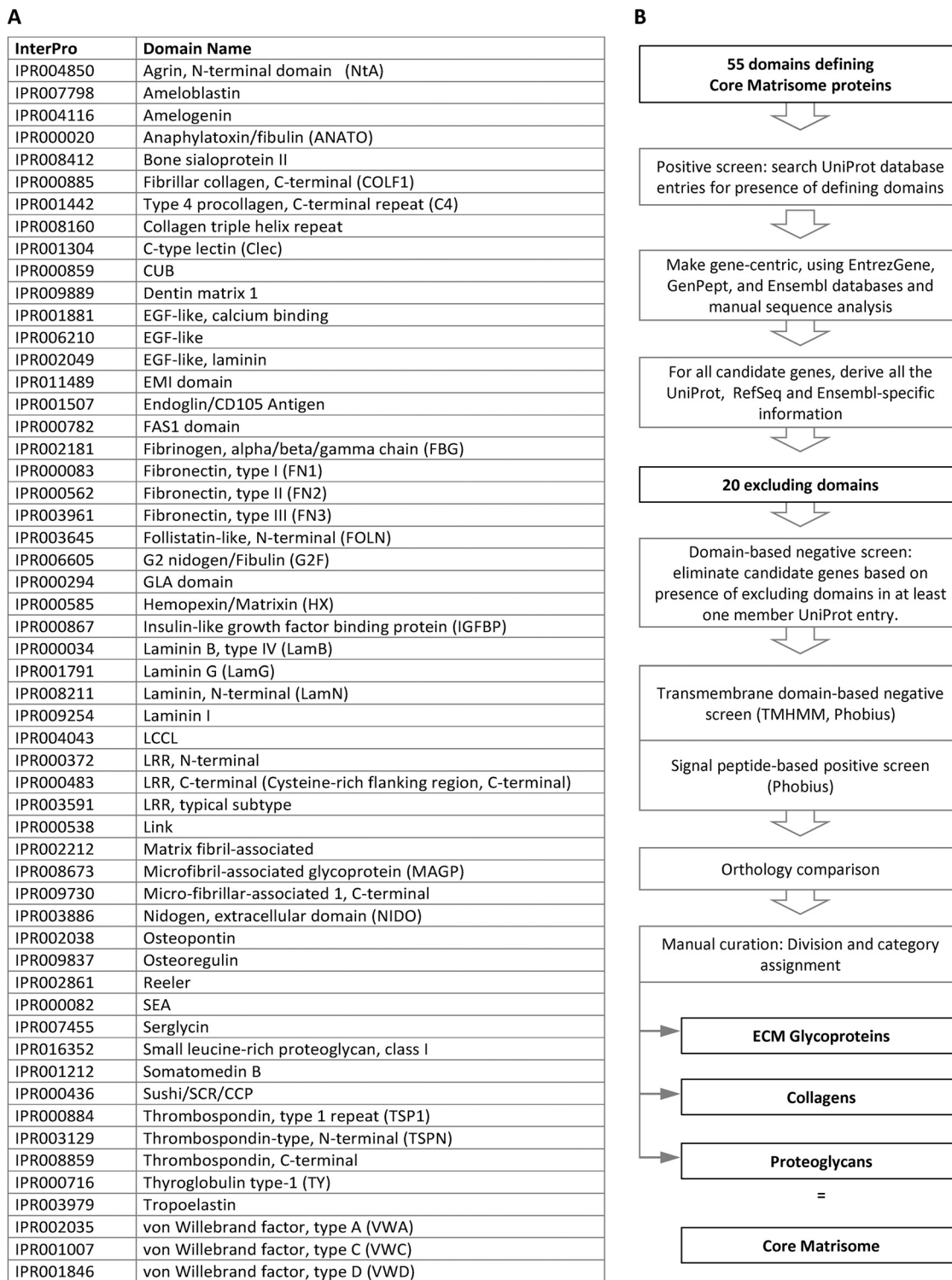


FIG. 3. **Bioinformatics pipeline used to define the *in silico* core matrisome.** A, List of the 55 InterPro domains used to define the core matrisome. This list was compiled based on prior knowledge, data from the literature and iterative query of UniProt as described in Extended Experimental Procedures (see also [supplemental Table S2A](#)). B, Schematic representation of the bioinformatic pipeline (see text, [Extended Experimental Procedures](#) and [supplemental Fig. S1](#)).

analyses of ECM-enriched fractions; see Tables I and [supplemental Table S2 and S4](#)); 2) ECM regulators (ECM-remodeling enzymes, crosslinkers, proteases, regulators etc.);

3) secreted factors, many of which are known to bind to ECM and others that may ([supplemental Table S2A](#)). As for the core matrisome list, we also defined lists of domains that excluded

mis-assigned proteins from these categories (supplemental Table S2B). Using similar bioinformatic pipelines as for the core matrisome, we defined three categories of “matrisome-associated” proteins: ECM-affiliated proteins, ECM regulators, and secreted factors (supplemental Fig. S1).

This bioinformatics approach allowed us to characterize the *in silico* human and mouse matrisomes (supplemental Table S3A and 3B respectively). The human core matrisome is composed of 278 genes corresponding to somewhat more than 1% of the proteome, the complete matrisome, with 1056 genes, accounting for ~4% of the human proteome (Table I and supplemental Table S3A). In addition to 43 known collagen genes, we conclude that the human genome encodes 200 ECM glycoproteins and 35 proteoglycans. As expected, the human and murine genomes encode very similar collections of matrisome proteins (Table I and supplemental Table S3B). In addition to providing a comprehensive *in silico* atlas of ECM and ECM-associated proteins, which includes all previously known ECM proteins, we report here the identification of potential novel ECM proteins that display architectures similar to ECM proteins but that have not previously been reported to be part of the ECM (highlighted in supplemental Table S3). These lists are intended to evolve as the *in vivo* characterization progresses and one might wish to add more proteins or domains to this bioinformatic pipeline in the future.

Characterization of the Extracellular Matrix of Murine Lung—The categorization of all the proteins identified by mass spectrometry in the ECM-enriched fraction, using our bioinformatic definition, revealed the presence of 55 matrisome proteins out of 184 proteins detected in the simple LC-MS/MS analysis (*i.e.* without peptide separation by OGE, see below; see Fig. 2A and supplemental Table S1A). The other 70% of proteins detected represent insoluble intracellular components such as proteins associated with the actin cytoskeleton or with intermediate filaments plus a few transmembrane proteins (supplemental Table S1A). However these “contaminating” proteins remain a minority fraction as they represent less than 25% of the precursor-ion MS intensity (Fig. 2A, left panel). The addition of an intermediate fractionation step by off-gel electrophoresis (OGE, separation of the peptide mixture into 11 fractions according to isoelectric point, see Extended Experimental Procedures), before submitting the peptides to LC-MS/MS, increased by a factor of 4 the total number of peptides detected (Fig. 2B middle panel and 2D), corresponding to sixfold more total proteins detected with two or more peptides per protein (Fig. 2B, right panel). However, this step led to only a threefold increase in the number of matrisome proteins, indicating that the analysis is approaching a plateau in the level of detection of matrisome proteins. Interestingly, this additional step increased by a factor of 15 the number of matrisome-associated proteins detected in the sample. Moreover, it was only after peptide separation by OGE that we detected the presence of growth factors in our samples (Fig. 2C and supplemental Table S4A; see Discussion).

To test the reproducibility of the pipeline, we processed a second sample and found a large overlap of the matrisome proteins detected in the two samples, which showed ~80% identity when considering proteins identified by two peptides (Fig. 4A and supplemental Table S4A). The overlap is particularly evident for the most abundant core matrisome proteins belonging to the collagen and proteoglycan categories (Fig. 4A, right panel). Therefore, to define the extracellular matrix composition of a tissue, we combined the data from two independent samples including only those proteins recovered in *both* samples represented by at least two peptides in at least one of the samples. By this definition, the extracellular matrix of murine lung comprises 143 total matrisome proteins: 92 core matrisome proteins, and 51 matrisome-associated proteins (supplemental Table S4B).

Tissue Specificity of Extracellular Matrix Composition—Using the same pipeline and criteria, we characterized the murine colon extracellular matrix (supplemental Table S4D). Replicate ECM-enriched colon samples overlapped by 85% (Fig. 4B, left panel and supplemental Table S4C). As for the ECM-enriched lung fractions, reproducibility is particularly evident for the most abundant core matrisome proteins belonging to the collagen and proteoglycan categories (Fig. 4B, right panel).

Comparison of the lung and colon extracellular matrices allowed definition for each of these tissues of a “matrisome signature”; *i.e.* the subset of matrisome proteins displaying tissue-specific expression (Table II). A core set of 84 matrisome proteins were consistently found in both lung and colon matrisomes (see Fig. 4C). The vast majority of collagens, proteoglycans and components of the basement membrane (laminins a-chains 2, 3, 4, 5, b-chains 2 and 3 and chain c1, nidogen, perlecan) were found in both tissues. In addition, these categories of shared proteins are abundant, as indicated by the peptide precursor ion abundance and number of peptides detected by MS (Table II).

In addition to this common set of 84 proteins, 59 matrisome proteins were detected exclusively in the lung and 22 matrisome proteins were detected only in the colon. These differentially expressed proteins belong mainly to the ECM glycoprotein, ECM-affiliated, ECM regulators and secreted factor categories (Fig. 4C). Some of these tissue-specific proteins conform in obvious ways with the organ’s physiology. For example, thrombospondin 1 (Thbs1) is expressed in the lung but not detected in the colon. This protein was found to be differentially expressed in total ECM preparations of lung *versus* colon by LC-MS/MS without peptide separation by OGE, indicating that it is an abundant component of the lung ECM. In addition, thrombospondin 1 has previously been shown to play a role in lung physiology (35). Conversely, mucin 2 (Muc2), a protein known to be secreted by the goblet cells of the colon is specifically found in the ECM of colon and not in the ECM of lung (36). Similarly, galectin 4 (Lgals4) has been shown to be expressed exclusively in the digestive tract

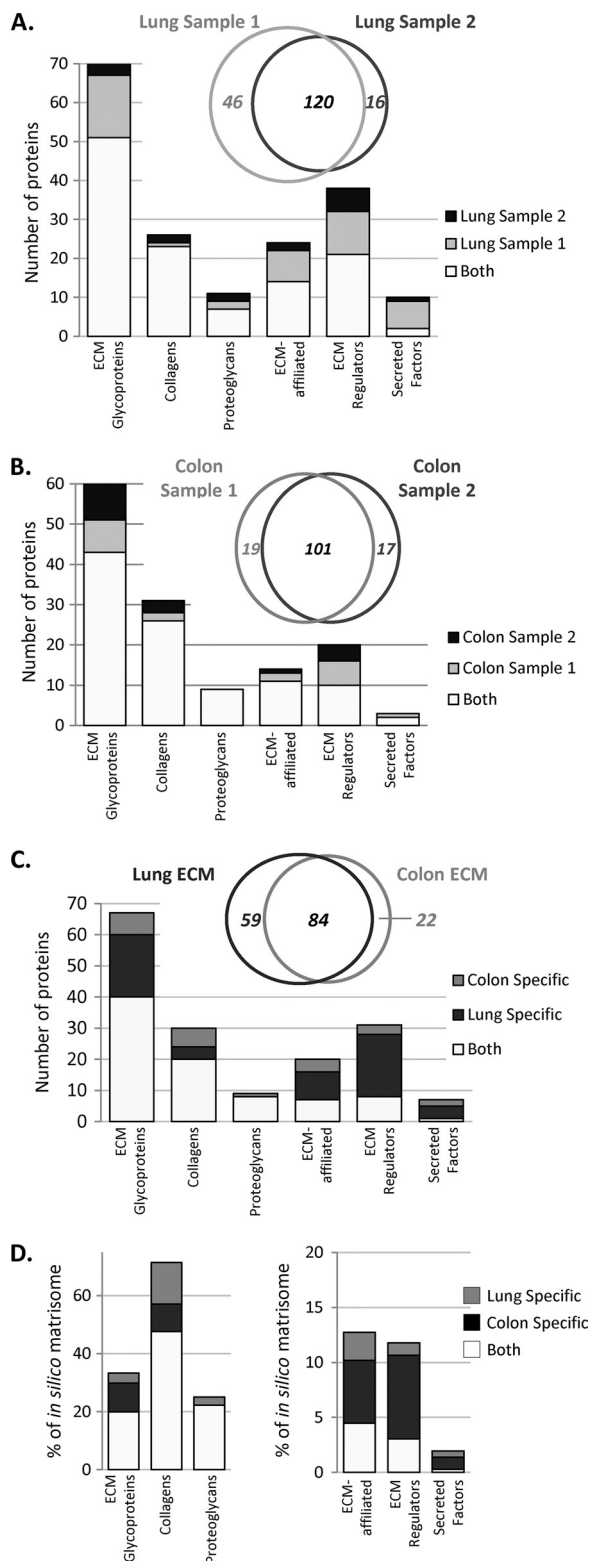


FIG. 4. Reproducibility of the proteomic analysis. *A*, Two lung samples were analyzed through the proteomics pipeline and inter-sample variability was assessed. Venn diagram shows reproducibility in terms of number of matrisome proteins identified by two peptides in each of two independent samples. Bar chart represents the matrisome

and is found only in the ECM of colon (37). This same strategy can be applied to various other tissues to characterize biochemically the complete tissue matrices in much greater detail than has been possible heretofore.

Experimental Coverage of the In Silico Matrisome—Comparison of the proteins detected *in vivo* with the predicted *in silico* matrisome shows that ~33% of the 274 core matrisome proteins and 11% of the 1098 total murine matrisome proteins (core plus associated) characterized *in silico* were detected at the protein level in the lung (Fig. 4D). Considering both lung and colon together, we detected 8 of the 36 predicted proteoglycans encoded by the murine genome (22%). The number of collagens found to be expressed in either the lung or the colon reached 70% of the total predicted *in silico* (Fig. 4D). We suspect that the other proteoglycans and collagens not detected will be found in other tissue types (see Discussion). Similar conclusions can be drawn for the ECM glycoproteins of the core matrisome. The representation of the predicted ECM-associated proteins in our experimental samples is significantly lower; we detect fewer than 15% of the predicted matrisome-associated proteins (Fig. 4D). Low abundance and tissue-specificity of many of these proteins, no doubt, contribute to the lower coverage of these sectors of the *in silico* lists. However, it is also possible that greater solubility might account for the fact that they are not detected in our samples (see Discussion).

The Composition of the Tumor ECM Changes With the Tumor's Metastatic Potential—Having developed this strategy, we wished to apply it to begin to tackle several questions about the tumor microenvironment. As discussed in the Introduction, changes in the ECM of tumors are of considerable interest. One challenging question is to determine whether the composition of tumor extracellular matrix differs in tumors with different metastatic ability. Accordingly, we grew subcutaneous tumors by injection into NOD/SCID/IL2R γ mice of A375 human melanoma cells (poorly metastatic) or their highly metastatic derivatives MA2 (16). The tumors were dissected 5

proteins detected in the two independent samples (see [supplemental Table S4A](#)). *B*, Two colon samples were analyzed through the proteomics pipeline and intersample variability was assessed. Venn diagram shows reproducibility in terms of number of matrisome proteins identified by two peptides in each of two independent samples. Bar chart represents the matrisome proteins detected in two independent samples (see [supplemental Table S4C](#)). *C*, Bar chart represents the overlap and tissue specificity for the lung and colon within each category of matrisome proteins. Venn diagram shows the number of matrisome proteins overlapping between the lung and colon and the number of tissue-specific matrisome proteins. Here, proteins are included only if they are found in *both* samples from a given tissue and represented by two peptides in at least one of them. *D*, *Left* panel: Comparison of the *in vivo* matrices of lung and colon core matrisome sets with the predicted *in silico* core matrisome. *Right* panel: Comparison of the *in vivo* lung and colon matrisome-associated sets of proteins with the predicted *in silico* matrisome-associated set of genes.

TABLE II

Tissue specificity of the extracellular matrix. Color code represents the number of peptides and total peptide precursor ion abundance of each protein in the two independent samples used to define each of the lung and colon matrices. Values used to generate the table were extracted from columns *F* (number of peptides) and *I* (peptide abundance) of [supplemental Table S4B](#) (lung) and [supplemental Table S4D](#) (colon). Dash (–) indicates that the protein was not found in either sample from a given tissue. Proteins included in this table were found in two independent samples with at least two peptides in one of the two samples. Uniprot accession numbers are given in [supplemental Table S4](#)

Core Matrissome						Matrissome-associated											
ECM Glycoproteins						Collagens						ECM-affiliated Proteins					
Entrez Gene Symbol	Protein MW (Da)	Lung		Colon		Entrez Gene Symbol	Protein MW (Da)	Lung		Colon		Entrez Gene Symbol	Protein MW (Da)	Lung		Colon	
		Number of Peptides	Peptide Abundance	Number of Peptides	Peptide Abundance			Number of Peptides	Peptide Abundance	Number of Peptides	Peptide Abundance			Number of Peptides	Peptide Abundance	Number of Peptides	Peptide Abundance
Abi3bp	98899	13	8.4E+07	8	7.1E+07	Col1a1	140079	83	1.9E+11	90	4.4E+11	Anxa1	41565	5	4.4E+07	7	8.1E+07
Agrn	213574	41	3.4E+09	4	1.6E+07	Col1a2	131090	69	1.7E+11	76	2.6E+11	Anxa2	40009	19	8.3E+08	13	1.9E+08
Bmper	81891	8	6.2E+07	-	-	Col3a1	141218	87	1.6E+11	-	-	Anxa5	36857	5	4.5E+07	-	-
Colq	49730	5	2.3E+07	4	2.4E+08	Col4a1	162840	-	-	21	1.3E+10	Anxa6	76764	9	3.4E+07	8	3.6E+07
Dmbt1	233356	-	-	6	2.0E+07	Col4a2	169542	43	6.0E+10	30	1.5E+10	Anxa7	51258	-	-	4	1.6E+07
Dpt	25655	5	5.2E+08	5	8.4E+08	Col4a4	167170	28	4.3E+09	14	1.2E+08	Anxa9	39500	3	6.6E+06	-	-
Ecm1	65496	12	2.3E+08	4	2.6E+07	Col4a5	164154	31	5.9E+10	-	-	Anxa11	55501	-	-	9	3.7E+07
Efemp1	58486	15	4.8E+08	7	1.2E+08	Col4a6	166363	-	-	15	1.8E+08	Cspg4	254278	9	2.1E+07	5	1.2E+07
Efemp2	53042	3	2.1E+07	-	-	Col5a1	185380	18	3.9E+09	24	8.7E+09	Clec14a	51399	5	3.0E+07	-	0.0E+00
Eln	73157	18	4.0E+09	16	2.5E+09	Col5a2	147065	36	4.1E+09	43	9.2E+09	Colec12	82895	8	5.7E+07	-	-
Emid1	47342	7	1.8E+08	-	-	Col5a3	173730	14	4.6E+08	20	1.0E+09	Frem1	247578	14	6.7E+07	-	-
Emilin1	110148	49	6.7E+09	40	2.0E+09	Col6a1	110651	43	1.2E+10	43	1.6E+10	Frem2	353501	26	1.2E+08	6	2.6E+08
Emilin2	119811	3	3.6E+06	3	4.2E+06	Col6a2	112496	38	1.4E+10	47	1.4E+10	Itih1	103092	4	-	11	1.3E+08
Fbln1	83292	4	1.6E+07	-	-	Col6a3	289383	120	4.3E+10	106	4.1E+10	Itih2	107300	4	9.8E+06	9	6.8E+07
Fbln2	139045	4	4.4E+07	3	2.3E+08	Col6a4	253385	-	-	68	1.7E+09	Itih3	100681	-	-	6	7.1E+07
Fbln5	53599	12	4.4E+09	5	2.2E+08	Col7a1	297276	23	4.1E+08	12	3.6E+08	Itih4	105813	3	5.0E+06	-	-
Fbn1	334014	96	1.7E+10	48	1.6E+09	Col9a3	64807	-	-	5	2.9E+07	Itih5	108179	9	4.3E+07	10	6.9E+07
Fga	62847	27	5.4E+09	12	3.1E+08	Col10a1	67854	5	2.5E+09	5	1.1E+09	Lox	48360	6	1.4E+08	-	-
Fgb	56446	39	1.3E+10	18	3.8E+08	Col12a1	342428	44	5.3E+08	85	4.3E+09	Lox1	68165	19	3.6E+09	12	1.0E+09
Fgg	52042	28	8.1E+09	16	4.9E+08	Col13a1	73232	-	-	9	2.5E+08	Mmp9	82732	2	3.0E+06	-	-
Fn1	267614	106	1.5E+10	79	7.3E+09	Col14a1	195230	16	8.7E+07	31	4.2E+08	Plg	94478	6	3.8E+07	-	-
Fras1	454888	9	3.4E+07	-	-	Col15a1	141948	10	1.3E+08	7	8.5E+07	Plod1	85123	5	1.5E+07	-	-
Hmnc1	609323	24	7.1E+07	-	-	Col16a1	158650	12	4.7E+08	13	1.1E+09	Plod3	86450	5	1.9E+07	-	-
Hmnc2	556151	-	-	35	1.5E+08	Col18a1	184217	14	1.3E+09	11	8.4E+07	Pzp	168327	11	3.1E+07	3	2.6E+06
Igfbp6	27403	2	1.3E+07	-	-	Col23a1	52896	9	1.5E+08	-	-	Serpina1a	47350	6	4.9E+07	-	-
Igfbp7	31084	9	8.5E+07	3	5.0E+06	Col24a1	177524	-	-	8	4.4E+07	Serpina3k	48228	7	2.6E+07	-	-
Lama1	348588	3	7.4E+06	9	2.5E+07	Col25a1	66880	8	7.9E+07	-	-	Serpinc1	53693	3	1.7E+07	-	-
Lama2	352981	52	1.2E+09	37	5.3E+08	Col27a1	188194	7	1.0E+09	6	1.8E+08	Serpinf2	56376	7	4.5E+07	-	-
Lama3	375996	110	1.0E+10	10	3.1E+07	Col28a1	72083	7	3.9E+08	5	3.3E+08	Serping1	57046	3	4.3E+07	-	-
Lama4	205620	55	3.1E+09	35	7.4E+08	Gm7455	292687	31	6.6E+08	41	9.8E+08	Serpinh1	47825	16	3.2E+08	13	1.4E+08
Lama5	417507	128	2.6E+10	76	2.7E+09							Tgm2	79311	39	1.0E+10	30	4.2E+09
Lamb1-1	210837	-	-	28	5.5E+08							Tgm3	78844	-	-	3	6.9E+06
Lamb2	204830	90	2.0E+10	46	2.5E+09												
Lamb3	133900	57	4.8E+09	4	7.4E+06												
Lamc1	184066	88	1.8E+10	40	2.8E+09												
Lamc2	135104	31	2.8E+09	-	-	Aspn	44132	10	3.4E+08	12	1.1E+09						
Ltbp1	196246	27	4.3E+08	7	8.3E+07	Bgn	43187	16	2.8E+09	11	1.4E+09						
Ltbp2	206157	6	2.0E+07	-	-	Dcn	41243	14	1.7E+09	14	2.0E+09						
Ltbp4	180949	28	5.7E+08	10	7.6E+07	Hapln1	42332	-	-	3	1.7E+07						
Matn2	111766	-	-	3	1.7E+07	Hspg2	480853	162	7.1E+10	115	1.5E+10						
Matn4	71851	3	7.0E+06	-	-	Lum	39786	19	2.0E+09	15	7.0E+08						
Mfap2	22685	4	1.5E+08	-	-	Ogn	35331	11	6.4E+08	11	2.0E+08						
Mfap4	30386	8	2.1E+09	5	1.4E+08	Prelp	44707	15	1.0E+09	9	4.4E+08						
Mfap5	20101	2	8.0E+07	5	4.5E+07	Vcan	369911	5	1.7E+07	8	7.0E+07						
Mfge8	53402	7	7.4E+07	-	-												
Mmrrn1	138614	7	1.3E+08	-	-												
Mmrrn2	107087	13	1.2E+08	-	-												
Nid1	140449	49	1.7E+10	29	2.7E+09												
Npnt	70077	24	8.3E+09	-	-												
Ntn1	71240	4	2.5E+07	-	-												
Papln	146609	8	7.8E+07	5	5.0E+07												
Pcolce2	47280	4	2.3E+07	-	-												
Postn	91600	48	1.5E+10	44	8.4E+09												
Pxdn	168830	21	1.3E+08	4	8.0E+06												
Spon1	94621	5	4.4E+07	-	-												
Srp2	55326	-	-	2	1.7E+06												
Tgfb1	76186	25	7.7E+08	32	4.4E+09												
Thbs1	134673	8	4.6E+07	-	-												
Thbs4	110009	-	-	5	7.0E+07												
Tinag	56847	10	2.7E+08	-	-												
Tinagl1	55111	16	1.5E+09	13	8.6E+08												
Tnc	238530	-	-	52	1.7E+09												
Tnxb	455363	55	8.3E+08	81	2.0E+09												
Vtn	56753	9	5.9E+08	2	1.2E+07												
Vwa1	45783	5	7.2E+07	8	1.7E+08												
Vwa5a	88844	6	1.5E+07	8	1.4E+07												
Vwf	323518	47	1.4E+09	5	1.7E+07												
ECM Regulators																	
Adam10	87101	7	2.4E+07	-	-	Adam19	104849	2	4.5E+07	-	-	Adamts17	126041	5	9.3E+07	-	-
Adamts1	198848	8	1.1E+08	5	1.5E+08	Adamts18	198848	8	1.1E+08	5	1.5E+08	Adamts19	56099	6	2.7E+07	-	-
Adamts5	56099	6	2.7E+07	-	-	Ambp	41135	3	9.8E+06	-	-	Cpn2	62584	2	5.2E+06	-	-
Sftpa1	27588	8	1.0E+09	-	-	Ctsb	39254	2	8.7E+06	-	-	Elane	30210	5	3.5E+07	-	-
Sftpb	44273	3	1.7E+07	6	3.2E+08	F13a1	84728	8	1.8E+07	8	3.1E+07	Htra1	53124	3	1.8E+07	-	-
Sftpd	39092	5	7.1E+07	-	-	Itih1	103092	-	-	11	1.3E+08	Itih2	107300	4	9.8E+06	9	6.8E+07
Secreted Factors																	
Egfl7	32200	6	7.2E+07	-	-	Itih3	100681	-	-	6	7.1E+07	Itih4	105813	3	5.0E+06	-	-
Hcfc1	217929	-	-	-	-	Itih5	108179	9	4.3E+07	10	6.9E+07	Lox	48360	6	1.4E+08	-	-
Il16	144062	3	8.1E+06	2	1.3E+07	Lox1	68165	19	3.6E+09	12	1.0E+09	Mmp9	82732	2	3.0E+06	-	-
InsI5	18172	-	-	2	9.8E+06	Plg	94478	6	3.8E+07	-	-	Plod1	85123	5	1.5E+07	-	-
P4	12431	2	9.0E+07	-	-	Plod3	86450	5	1.9E+07	-	-	Pzp	168327	11	3.1E+07	3	2.6E+06
S100a11	12203	2	2.9E+07	-	-	Serpina1a	47350	6	4.9E+07	-	-	Serpina1a	47350	6	4.9E+07	-	-
S100a13	12107	2	1.2E+07	-	-	Serpina3k	48228	7	2.6E+07	-	-	Serpinc1	53693	3	1.7E+07	-	-
						Serpinc1	53693	3	1.7E+07	-	-	Serpinf2	56376	7	4.5E+07	-	-
						Serpinf2	56376	7	4.5E+07	-	-	Serping1	57046	3	4.3E+07	-	-
						Serping1	57046	3	4.3E+07	-	-	Serpinh1	47825	16	3.2E+08	13	1.4E+08

weeks later, and the tumor ECM was enriched using the protocols outlined above. As before, we define the tumor ECM as the ensemble of ECM proteins and ECM-associated proteins found in two independent samples ([supplemental Table S5](#)). The analysis of the composition of matrix from A375 or MA2 tumors revealed that the majority of the matrix proteins detected were expressed by both tumor types and in comparable proportions (fibronectin, LTBP3 and 4, periostin, etc.) (Table III). In addition, this study also revealed the differential expression of matrix proteins. Basement membranes are specialized forms of extracellular matrix of defined composition that are found at the interfaces between epithelia and connective tissues and underlying the endothelial cells of the blood vessel wall. In this comparative study, most of the basement membrane components (type IV and XV collagens, laminins, HSPG2 or perlecan, nidogen 2) were found to be expressed in the same proportions by the two tumor types although we could also detect differential expression of collagens XVIII and IVa3 expressed by the poorly metastatic melanoma whereas laminin b2 and nidogen 1 were only detected in the ECM of metastatic melanoma (Table III and [supplemental Table S5](#)).

Lysyl oxidases are enzymes capable of cross-linking collagens and elastin (38). Both tumor types expressed lysyl oxidase-like 1 (LOXL1) in the same proportions, whereas lysyl oxidase-like 2 (LOXL2) was found to be expressed by the A375 but not detected in the MA2 tumors. Conversely, LOXL3 and LOXL4 were not detected in A375 tumors but were present in the MA2 tumors. Emilins are ECM proteins that bind elastin and ECM fibrils (39). In our study, we observed that, whereas both tumor types expressed Emilin 1, only the MA2, highly metastatic tumors expressed also the two other related proteins Emilin 2 and Emilin 3. Interestingly, elastin was found in both tumor types but showed a 10-fold increase in the ECM of MA2 as compared with that of A375 (Table III).

Most of the collagens were found in both tumor types in the same proportions (collagens I, III, V), whereas some were detected in only one of the two tumor types (collagens IV chain a3, IX chain a1, XVIII in A375 tumors and collagens VI chain a6, VIII chains a1 and a2, X, XIV, XXVII, XXVIII in MA2 tumors).

Although we do not yet fully understand the functional significance of these examples, we report here for the first time a detailed composition of melanoma extracellular matrices, which can lead to further analyses.

Understanding the Origins of the Tumor Extracellular Matrix—Another challenging question when studying the tumor microenvironment is to understand the origin of the tumor ECM; that is, whether the tumor ECM is produced and secreted by the tumor cells themselves, by the stromal cells or by both compartments. To address this question, we pursued the analysis of the melanoma xenografts described above, by identifying for each protein its origin: is the protein secreted by the human tumor cells or by the murine stroma? The murine

sequence of a given protein is, in most cases, sufficiently different from its human ortholog to be distinguished by proteomic analyses ([supplemental Fig. S3](#)). The mass spectrometric analysis allowed us to distinguish the human proteins from their murine counterparts. The origin of a protein could not be determined in only a very few cases (protein indicated as “indistinguishable” in column D of [supplemental Table S6](#) such as S100A10 in the A375 ECM or SRPX in the MA2 matrix). In order to be able to identify without ambiguity the origin of each protein, we required that proteins needed to be detected in two independent samples with at least two species-specific peptides in one of them. Using this strategy, we identified for each tumor type a set of matrix proteins exclusively secreted by the (human) tumor cells, and another set exclusively secreted by the (murine) stromal cells (Fig. 5 and [supplemental Table S6](#)). Extracellular matrix proteins and proteases found in plasma, fibrinogen (Fg) subunits, plasminogen (Plg), thrombin (F2) and Factor XIIIa1, were all, in agreement with their function, found to be expressed exclusively from the mouse genome. In addition, we found proteins such as fibrillin-1, fibronectin etc. that are secreted by both the tumor cells and the stromal cells. Among these latter proteins are several basement membrane components: collagen IV, a1, a2, a3, a5 chains, collagen XVIII, perlecan (HSPG2), and nidogen 1.

To analyze further the relative contributions of the two cellular compartments to the secretion of a given protein, we developed a semiquantitative measure. In brief, we calculated the ratio of the observed precursor-ion MS intensities for sequence-distinguishable homologous peptides from the mouse *versus* human proteins ([supplemental Fig. S3](#) and [supplemental Table S6](#)). We consider a ratio above five to indicate that the contribution of the tumor cells to the secretion of the protein of interest is predominant (indicated in orange in Fig. 5 and [supplemental Table S6C](#)). Conversely, a ratio under 0.2 indicates a greater contribution from the stroma (Fig. 5 and [supplemental Table S6C](#), proteins highlighted in green). Finally, some proteins are secreted in significant proportions by both compartments—ratio between 0.2 and 5 (Fig. 5 and [supplemental Table S6C](#), proteins highlighted in yellow). We conclude from these data that both tumor cells and stromal cells contribute, although differentially, to the secretion of proteins making up the tumor ECM.

Our results are in agreement with recent data from the literature. For example, periostin, a known ECM glycoprotein involved in tumor progression and metastasis formation was expressed exclusively by the stroma in A375 tumors, whereas its expression was also induced in the tumor cells in MA2 tumors. This result conforms with a gene expression study by Tilman *et al.* who showed that periostin is not expressed by normal skin but becomes expressed by melanoma cells and the stroma during tumor progression and metastasis formation (40). Conversely, the basement membrane components, laminin chains a5 and c1, were only expressed by the tumor

TABLE III

The tumor ECM changes with tumor's metastatic potential. Color code represents the number of peptides and total peptide precursor ion abundance of each protein in the two independent samples used to define the ECM compositions of non-metastatic (A375) and metastatic (MA2) human melanoma xenografts. Values used to generate the table were extracted from columns I (number of unique peptides) and J (peptide abundance) of [supplemental Table S5B](#) (A375) and [supplemental Table S5D](#) (MA2). Dash (–) indicates that the protein was not detected in either sample from a given tissue. Proteins included in this table were found in two independent samples with at least two peptides in one of the two samples. Uniprot accession numbers are given in [supplemental Table S5](#)

Core Matrisome						Matrisome-associated											
ECM Glycoproteins						Collagens						ECM-affiliated Proteins					
Entrez Gene Symbol	Protein MW (Da)	A375		MA2		Entrez Gene Symbol	Protein MW (Da)	A375		MA2		Entrez Gene Symbol	Protein MW (Da)	A375		MA2	
		Number of Peptides	Peptide Abundance	Number of Peptides	Peptide Abundance			Number of Peptides	Peptide Abundance	Number of Peptides	Peptide Abundance			Number of Peptides	Peptide Abundance		
AEBP1	133177	4	3E+07	-	-	Col1a1	140079	80	1.2E+11	75	5.2E+11	ANXA1	40407	6	2.6E+08	5	2.8E+07
C15orf44	59297	5	2E+07	-	-	Col1a2	131090	61	1.1E+11	69	3.3E+11	ANXA2	41749	17	3.8E+08	-	-
Cilp	135749	-	-	6	2.0E+07	Col3a1	141218	60	3.0E+10	78	1.8E+11	ANXA5	37459	2	1.1E+07	-	-
Dpt	25655	2	1E+08	4	5.3E+07	Col4a1	162840	15	1.4E+09	15	1.1E+10	Anxa6	76764	4	1.1E+07	-	-
Ecm1	65624	8	1E+08	-	-	Col4a2	169542	16	1.6E+09	25	8.7E+09	CSPG4	252714	-	-	11	3.2E+07
Efemp1	58486	3	2E+07	-	-	Col4a3	164400	10	4.2E+07	-	-	LGALS1	16595	9	5.2E+09	5	1.6E+08
Eln	73173	7	7E+07	9	1.9E+08	COL4A4	334471	6	7.1E+07	4	4.7E+07	LGALS3	27747	4	5.0E+07	-	-
EMILIN1	109787	35	3E+09	43	6.5E+09	Col5a1	185380	12	1.9E+09	22	4.7E+09	LMAN1	59359	6	1.0E+08	-	-
Emilin2	119817	-	-	5	8.0E+06	Col5a2	147065	29	1.9E+09	38	8.6E+09	ECM Regulators					
EMILIN3	85482	-	-	4	1.0E+07	COL5A3	174243	11	5.0E+08	13	7.1E+08	ADAMTSL1	201182	7	2.4E+08	-	-
Fbln2	133346	5	2E+08	-	-	COL6A1	111109	32	7.3E+09	29	2.2E+09	CD109	163970	7	6.8E+07	-	-
Fbn1	334014	50	2E+09	18	7.6E+08	COL6A2	111216	20	5.7E+09	24	1.7E+09	Ctsb	39254	6	4.8E+07	-	-
Fga	62847	18	5E+09	16	2.0E+09	COL6A3	346811	112	3.1E+10	80	8.4E+09	F13a1	84728	11	3.4E+08	5	2.5E+07
Fgb	56446	37	1E+10	25	2.2E+09	Col6a6	249109	-	-	9	3.9E+07	F2	72703	19	1.0E+09	-	-
Fgg	52042	22	1E+10	14	1.3E+09	COL7A1	297625	4	6.6E+07	6	2.6E+07	HTRA1	49549	3	2.2E+07	-	-
Fn1	267614	86	5E+10	71	2.8E+10	COL8A1	74491	-	-	5	2.5E+07	Itih1	103092	-	-	6	8.2E+07
HMCN1	625287	-	-	26	4.5E+07	COL8A2	68855	-	-	5	4.3E+07	Itih2	107300	-	-	5	1.9E+07
Lama2	352981	5	3E+07	10	1.3E+07	Col9a1	93797	4	5.1E+07	-	-	Itih4	105813	3	1.8E+07	-	-
LAMA4	206628	9	8E+07	14	1.0E+08	Col10a1	67854	-	-	4	3.5E+09	LEPREL2	84468	4	3.7E+07	-	-
LAMAS	413827	41	3E+08	8	1.8E+07	Col12a1	342428	111	1.6E+10	70	2.5E+09	Lox1	68165	3	2.2E+07	6	2.6E+07
LAMB1	206836	13	2E+08	8	5.8E+07	Col13a1	74707	6	8.2E+08	6	2.0E+07	LOXL2	90270	6	7.5E+07	-	-
Lamb2	204830	-	-	14	4.9E+07	Col14a1	195230	-	-	6	4.3E+07	LOXL3	86769	-	-	3	1.7E+07
LAMC1	184877	15	2E+08	16	2.7E+08	Col15a1	141948	7	1.1E+08	7	4.4E+06	LOXL4	87915	-	-	7	6.5E+07
LTBP1	196760	5	1E+07	4	2.7E+06	Col16a1	158650	8	1.1E+08	9	2.5E+08	P4HA1	62735	4	4.6E+07	-	-
Ltbp2	206157	3	5E+07	-	-	Col18a1	184217	12	1.0E+08	-	-	Plg	94478	7	1.5E+08	-	-
LTBP3	148068	3	2E+07	2	8.3E+07	COL19A1	117458	5	4.4E+07	4	6.8E+07	PLOD1	85541	13	7.4E+08	3	1.1E+07
LTBP4	175750	9	2E+08	6	3.3E+08	COL22A1	163610	4	2.1E+07	5	5.2E+07	PLOD2	88848	5	5.8E+07	-	-
Mfap2	22685	2	1E+07	-	-	Col27a1	188194	-	-	4	1.5E+07	PLOD3	86776	14	3.5E+08	-	-
NID1	140666	-	-	12	1.2E+08	Col28a1	120682	-	-	3	8.3E+08	Pzp	168327	6	5.2E+07	4	4.7E+06
Nid2	157712	10	3E+08	4	2.7E+07	Proteoglycans						Serpina1a	47350	4	1.6E+07	-	-
Postn	91600	26	9E+08	25	8.7E+08	Aspn	44132	10	9.4E+08	3	1.2E+07	Serpina3k	48228	3	4.9E+07	-	-
PXDN	169432	-	-	7	1.5E+07	Bgn	43187	13	4.0E+09	8	1.5E+08	SERPINB1	44763	2	4.8E+06	-	-
SRPX	54376	4	2E+07	2	6.8E+06	Dcn	41243	7	1.5E+08	3	6.1E+07	SERPINE2	46152	9	9.3E+07	9	4.9E+07
TGFB1	76937	34	2E+10	29	6.3E+09	HAPLN1	42421	-	-	6	5.9E+07	SERPINF1	48453	4	4.3E+07	-	-
Thbs1	134673	4	3E+07	-	-	HSPG2	480990	78	3.1E+09	61	1.9E+09	Serpinf2	56376	7	1.4E+08	-	-
THSD4	117408	12	3E+08	3	3.7E+06	Lum	39786	9	8.8E+08	-	-	Serpinh1	47825	11	7.4E+08	5	4.0E+07
TINAGL1	55422	12	2E+08	6	4.8E+07	Prelp	44707	10	1.7E+08	-	-	Tgm2	79311	16	4.7E+08	13	2.2E+08
TNC	247969	76	2E+10	58	5.9E+09	Secreted Factors						ANGPTL4	47262	3	4.7E+07	-	-
Tnn	176132	9	1E+08	-	-	HCFC1	212427	9	4.6E+07	-	-	S100A10	12571	2	1.2E+08	-	-
Tnxb	455363	6	2E+07	4	1.4E+08	S100A11	13109	2	4.8E+07	-	-	S100A13	12725	2	2.6E+07	-	-
Vtn	56723	9	3E+09	9	5.8E+08	S100A4	13211	2	1.2E+08	-	-	S100A6	11491	2	1.6E+07	-	-
VWA1	48370	3	3E+07	-	-	TGFB1	46542	3	5.9E+07	-	-						

FIG. 5. The tumor extracellular matrix is secreted by both tumor cells and stromal cells and differs with metastatic potential.

A, Proteins secreted by A375 tumors and not by MA2 tumors. B, Proteins secreted by MA2 tumors and not by A375 tumors. C, Proteins expressed by both tumor types and by the same compartment in the two tumor types. D, Proteins expressed by both tumor types but by different compartments. Proteins are sorted by tumor type and by their origins: tumor (red), stroma (blue), or both (yellow: similar abundance of the human and mouse proteins, orange: human form is at least 5 times more abundant than the mouse form, green: the mouse form is at least 5 times more abundant than the human form in the two samples, ratios were calculated using the values indicated in column Y and values are given in column Z of supplemental Table S6B and 6D) and by category. UniProt accession numbers are given in supplemental Table S6 (columns D and E). The number of shared peptides (common to human and mouse sequences) and of human- and mouse-specific peptides is given for each protein (data were extracted from columns V and W of supplemental Table S6B and 6D). BM indicates basement membrane proteins, H indicates proteins involved in hemostasis (i.e. plasma-derived - as expected, all of these are murine). Asterisks (*) indicate that, in addition to the identification of human- and mouse-specific peptides, the entries had a few peptides observed corresponding to distinct isoforms involving point mutations or exon deletions. For simplicity the isoforms are combined here. For detailed information, see Extended Experimental Procedure and supplemental Table S6.

A. Proteins expressed by non-metastatic melanoma

Entrez Gene Symbol	Note	A375		
		Shared Peptides	Tumor (human)	Stroma (mouse)
ADAMTSL1		1	6	-
ANGPTL4		0	3	-
ANXA2		13	4	-
CD109		0	7	-
HCFC1		5	4	-
HTRA1		1	2	-
LEPREL2		0	4	-
LGALS3		2	2	-
LMAN1		0	6	-
LOXL2		1	5	-
P4HA1		1	3	-
PLD2		2	3	-
S100A13		0	2	-
S100A4		0	2	-
S100A6		0	2	-
SERPINB1		0	2	-
SRPX		1	3	-
TGFB1		1	2	-
THSD4		5	7	-
BGN		10	2	3
PLD3		4	10	3
COL4A4	BM	0	3	3
CTSB		0	3	3
SERPINF1		0	2	2
COL18A1	BM	0	5	7
ECM1		0	3	5
C15orf44		n/a	-	5
Efemp1		1	-	2
F2	H	4	-	15
Fbln2		1	-	4
Itih4		0	-	3
Itih2		1	-	2
Lum		2	-	7
Nid2	BM	3	-	7
Plg	H	0	-	7
Serpina1a		0	-	4
Serpina3k		0	-	3
Serpinf2		0	-	7
Thbs1		2	-	2
Tnn		2	-	7
Vwa1	H	0	-	2

B. Proteins expressed by metastatic melanoma

Entrez Gene Symbol	Note	MA2		
		Shared Peptides	Tumor (human)	Stroma (mouse)
COL8A1		2	3	-
COL8A2		0	3	-
CSPG4		1	10	-
EMILIN3		0	4	-
HAPLN1		4	2	-
HMCN1		4	22	-
LOXL4		1	6	-
PXDN		2	5	-
SERPINE2		3	6	-
COL10A1		0	2	2
COL11A2		3	4	3
LAMB2	BM	6	3	8
Clp		2	-	4
Col28a1		0	-	3
Col5a1		18	-	4
Emilin2		1	-	4
Itih1		0	-	6
Itih2		2	-	3

Origin of the tumor ECM

- Tumor cells
- Both, more from tumor cells
- Both equally
- Both, more from the stroma
- Stroma
- Not detected

C. Proteins expressed by the same compartment in A375 and MA2 tumors

Entrez Gene Symbol	Note	A375			MA2		
		Shared Peptides	Tumor (human)	Stroma (mouse)	Shared Peptides	Tumor (human)	Stroma (mouse)
COL19A1		1	4	-	1	3	-
COL22A1		0	4	-	0	5	-
COL7A1		1	3	-	0	4	-
LAMA4	BM	4	5	-	5	9	-
LAMB1	BM	7	6	-	4	4	-
LGALS1		2	7	-	2	3	-
LTBP1		2	3	-	1	3	-
LTBP3		0	3	-	0	2	-
LTBP4		3	6	-	2	4	-
PLD1		4	9	-	1	2	-
TINAGL1		3	9	-	2	4	-
COL4A3	BM	0	3	7	0	4	6
COL6A3	*	17	95	74	23	62	79
FN1	*	77	6	11	65	9	9
NID1	BM	2	7	8	3	9	5
SERPINH1		5	4	6	2	3	2
TGFB1		12	22	8	9	20	13
Col1a1	*	35	15	45	40	10	48
COL1A2		9	26	52	10	27	59
COL3A1	*	6	22	39	11	25	62
COL5A2		11	6	18	19	13	19
Aspn		3	-	7	1	-	2
Col5a3		0	-	9	0	-	13
Col14a1		2	-	5	1	-	5
Dcn		0	-	7	0	-	3
Dpt		0	-	2	2	-	2
Ein		0	-	7	0	-	9
F13a1	H	0	-	11	0	-	5
Fga	H	0	-	18	0	-	16
Fgb	H	11	-	26	7	-	18
Fgg	H	6	-	16	3	-	11
Lama2	BM	2	-	3	4	-	6
Pzp		0	-	6	0	-	4

D. Proteins expressed by a different compartment in A375 and MA2 tumors

Entrez Gene Symbol	Note	A375			MA2		
		Shared Peptides	Tumor (human)	Stroma (mouse)	Shared Peptides	Tumor (human)	Stroma (mouse)
Increased stromal contribution in metastatic MA2 tumors							
EMILIN1		7	28	-	9	34	15
COL11A1		5	0	-	0	5	2
LAMC1	BM	6	9	-	6	10	8
LAMA5	BM	5	36	-	2	-	6
COL13A1		0	3	3	0	-	4
COL6A1		6	26	21	4	15	25
COL6A2		7	13	13	8	8	16
TNC	*	14	62	51	12	53	31
TGM2		2	8	14	1	-	12
Increased tumor contribution in metastatic MA2 tumors							
ANXA1		2	4	3	1	3	-
COL4A1	BM	6	4	9	7	5	8
COL4A2	BM	5	9	11	7	18	17
COL4A5	BM	2	3	5	2	3	2
HSPG2	BM*	18	60	40	11	51	39
COL12A1		52	3	59	33	30	37
FBN1		30	15	20	11	7	5
Vtn		0	-	9	0	2	7
Col2a1		6	-	5	6	3	3
Col15a1	BM	0	-	7	0	4	3
Col16a1		3	-	5	4	5	5
Postn		5	-	21	8	4	-
Tnxb		1	-	5	1	3	-

cells in the A375 tumors and were secreted by both compartments (laminin c1 chain) or exclusively secreted by the stroma (laminin a5 chain) in the MA2 tumor. Of interest,

laminin 511, composed of the a5, b1 and c1 chains has been shown recently to promote melanoma cell migration and invasion (41).

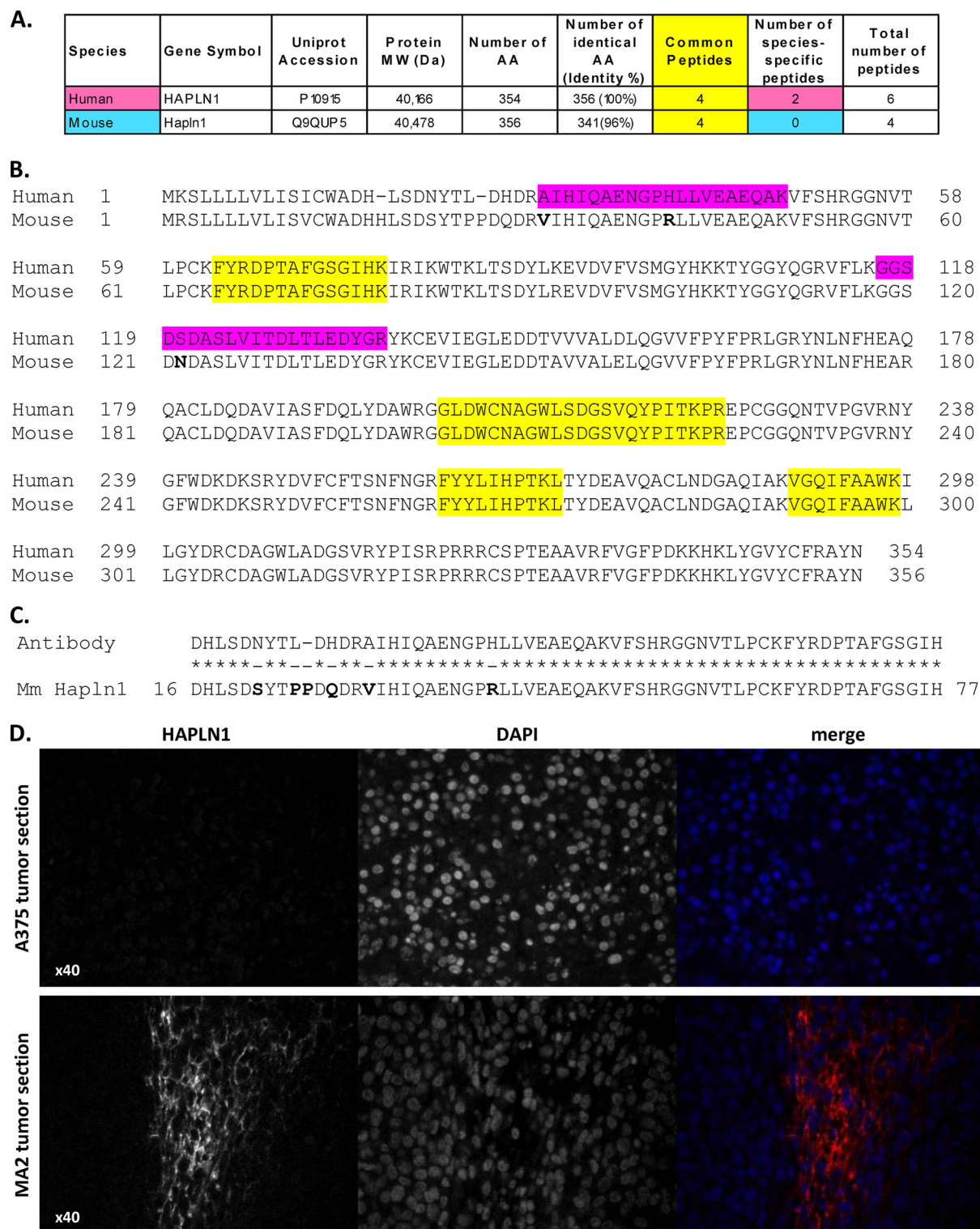


FIG. 6. Hyaluronan and proteoglycan link protein-1 (HAPLN1) is overexpressed in metastatic melanoma and secreted by the tumor cells. **A**, Numbers of nondistinguishing, species-specific and total peptides identified in the two metastatic tumor samples. In addition to peptides common to both the human and murine proteins, two human-specific peptides and no murine-specific peptide were detected, indicating that HAPLN1 is secreted by the tumor cells. Moreover, the two proteins are 96% identical in sequence so there are only a few diagnostic peptides. **B**, Sequence alignment of human and murine HAPLN1. Peptides identified by LC-MS/MS are highlighted in yellow if the peptide is identical in the two species and in pink if the peptide matches the human protein sequence. The identification of human-specific peptides and not closely related mouse peptides suggests that HAPLN1 is secreted exclusively by the tumor cells. **C**, Alignment of the

Our study is the first to report the detailed composition and origin of the tumor extracellular matrix at the protein level and highlights many proteins of potential importance for metastatic disease. Therefore, we conclude that the matrisome components secreted by the tumor cells vary with the metastatic potential of the tumor cells. In addition, the matrisome components secreted by the stromal cells also change in response to the metastatic potential of the tumor cells, indicating significant cross-talk between the tumor cells and the stromal cells.

Hyaluronan and Proteoglycan Link Protein-1 (HAPLN1) is Overexpressed in Metastatic Melanoma and is Secreted Exclusively by the Tumor Cells—To illustrate the potential of this approach, we considered the example of hyaluronan and proteoglycan link protein-1 (HAPLN1), which was found to be expressed exclusively in the metastatic tumors and specifically secreted by the tumor cells (Table III, Fig. 5, and [supplemental Table S5B](#)). The function of HAPLN1 is to link proteoglycans to hyaluronic acid thus participating in the architecture of the extracellular matrix (42). Interestingly, this gene was previously found to be up-regulated in gene expression arrays comparing MA2 tumors to A375 tumors (43) and is often up-regulated in highly metastatic cell lines or human tumors (44, 45).

Additionally, the Human Protein Atlas database indicates an increased expression of HAPLN1 in malignant melanoma (<http://www.proteinatlas.org/ENSG00000145681/cancer/malignant+melanoma>) (46).

Our proteomic analysis identified the origin of HAPLN1 as being secreted by the tumor cells and not by the stromal cells, since human-specific but not murine-specific peptides were detected (Figs. 6A, 6B and [supplemental Table S6B](#)). We confirmed the expression of HAPLN1 in metastatic and not in nonmetastatic tumors by immunohistochemistry (Fig. 6D). The fragment used to generate the antibody is 89% identical with the corresponding murine sequence (Fig. 6C), which renders ambiguous the identification of the origin of HAPLN1 by immunohistochemistry alone. This example demonstrates that our proteomics strategy is a method of choice to identify without ambiguity the origin of a given protein of tumor ECM.

DISCUSSION

We report here the development of proteomic methods to characterize in detail the biochemical composition of the extracellular matrix from normal and tumor tissues. In addition to the development of this experimental strategy it was essential to develop a systematic and objective bioinformatic definition of those proteins that should be considered as part of, or associated with, the ECM—we call this set of proteins the

“matrisome” (Fig. 7). This detailed inventory of ECM and ECM-associated proteins serves as a basis for analyses of the tissue specificity of ECM composition and of changes that occur during physiological or pathological processes, exemplified here by analyses of cancer. These methods can now be applied to diverse biological questions concerning ECM biology and pathology and we present here detailed protocols to allow their exploitation by others.

In Silico Definition of the Matrisome—Taking advantage of the characteristic domain-based organization of ECM proteins, we established a bioinformatic pipeline aimed at identifying all ECM proteins encoded within the genome, based on the presence and absence of diagnostic domains. In parallel with the definition of this “core” group of extracellular components (structural and fibrillar glycoproteins and proteoglycans), we also defined, using similar methodology, an inclusive list of matrisome-associated proteins, comprising 1) ECM-affiliated proteins (affiliated either structurally or physically with the core matrisome), 2) ECM-remodeling enzymes and iii) secreted factors such as growth factors and cytokines, known or suspected to bind to ECM. In addition to providing an atlas of the matrisome, these lists were incorporated into the proteomic pipeline to annotate automatically the mass spectrometric output (see below and Fig. 7).

The compilation of matrisome-diagnostic domains and our inventory of 1056 matrisome genes in the human genome (accounting for 4% of the proteome), including 278 core matrisome genes (1% of the proteome) represents the most comprehensive definition of extracellular matrix components to date and provides a framework for future analyses of the ECM of normal and diseased tissues. Indeed, the list of core matrisome proteins should serve as a more comprehensive alternative to GO annotation of ECM proteins, which is inadequate in several ways. This list of 278 core matrisome proteins includes all previously known ECM proteins as well as a number of previously unknown proteins (see [supplemental Table S3](#)). Of note, this systematic approach would not allow the identification of proteins that lack ECM-diagnostic domains and that are not known to be part of the extracellular matrix. However, the domain and protein lists are intended to evolve as the *in vivo* characterization of the extracellular matrix progresses and knowledge in the field advances; one could readily add to the bioinformatic pipeline new domains or criteria, thereby including additional proteins in the lists. It is worth noting that *all* proteins detected by the mass spectrometry but not included in the lists of bioinformatically defined matrisome proteins are included in the supplemental Tables and available for search by additional bioinformatic analyses.

sequence of the human protein fragment (amino acids 16 to 75) used to generate the anti-HAPLN1 antibody with the murine protein sequence (amino acids 16 to 77), which is 89% identical. D, HAPLN1 is up-regulated in MA2 tumors. Immunohistochemistry using anti-HAPLN1 antibody of A375 (*upper panel*) and MA2 (*lower panel*) tumor sections. *Left panels* show HAPLN1 staining, *middle panel* DAPI staining of the nuclei and *right panels* merge images.

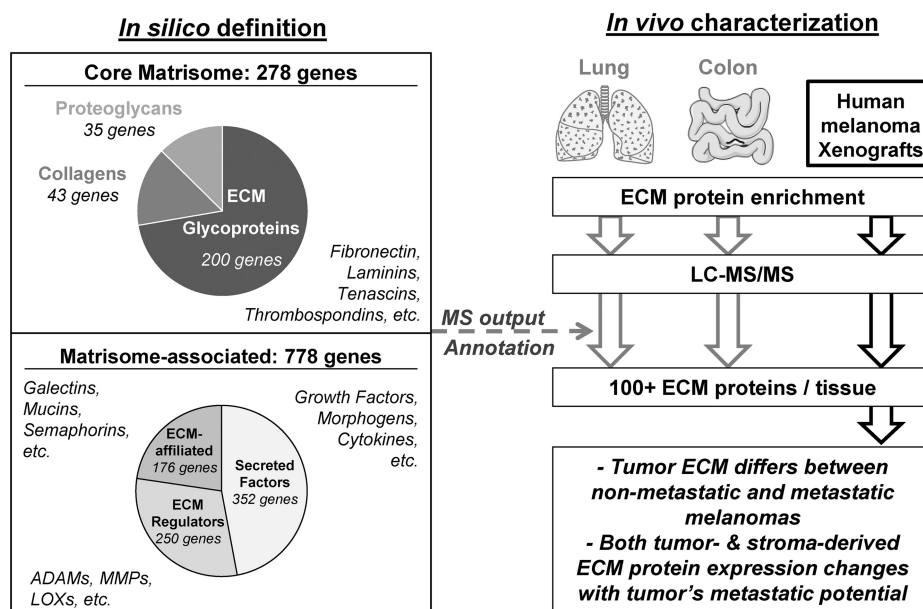


FIG. 7. The matrisome: *in silico* definition and *in vivo* characterization by proteomics of normal and tumor extracellular matrices. We propose here to update and broaden the use of the term “matrisome”—first used by Martin and collaborators in 1984 to define supramolecular basement membrane complex (including collagen IV, laminin, nidogen, and proteoglycan) (57)—to define the universe of ECM proteins and their complexes that can contribute to ECM structures. The domain-based bioinformatic analyses defined a set of extracellular matrix proteins (core matrisome) and an additional set of matrisome-associated proteins within the human and mouse genomes. This *in silico* definition of the matrisome was further used to annotate the proteomic output. The combination of a robust definition of matrisomal proteins and a high throughput proteomics pipeline led to the identification of more than 100 extracellular matrix proteins within any given tissue or tumor.

The Composition of Extracellular Matrices In Vivo—To study the global composition of the extracellular matrix from tissues and tumors, we employed a protocol originally designed to extract proteins from several intracellular and membranous subcellular compartments. This protocol largely depletes intracellular proteins and in turn enriches for insoluble ECM and ECM-associated proteins. Hence, this strategy focuses on the insoluble components present in the ECM and proteins tightly associated with these components. The enrichment for ECM components was confirmed by the mass spectrometric analysis of the composition of protein extracts. It is important to note that this protocol will miss proteins loosely bound to the ECM and eluted by the sequential extractions. However, such proteins would be detectable in the extracted fractions using targeted proteomics (see below).

Our analyses of the lung and colon showed that the ECM of a given tissue comprises between 100 and 200 proteins, including a set of tissue-specific proteins that represent 10% to 30% of the total. Considering both the lung and colon together, we detected 67 of the 195 murine ECM glycoproteins predicted bioinformatically (representing a third of the predicted ECM glycoproteins), 30 of the 43 collagen subunits and eight of the 36 proteoglycans encoded by the murine genome. We suspect that those matrisome proteins not detected will be found in other tissue types or at different developmental stages; for example, some proteoglycans, such as aggrecan or fibromodulin, and some collagens (collagen II and collagen IX) are known to be cartilage-specific (33, 47)

and many ECM proteins appear to be selectively or exclusively expressed in the nervous system (48). Thus, our coupled mass spectrometric and bioinformatic approaches have allowed a more extensive characterization of the ECM of tissues than has been possible previously.

The proportions of the predicted ECM-associated protein categories found in our experimental samples are significantly lower, and most of them were only detected after peptide separation by isoelectric focusing. ECM modification enzymes and secreted growth factors are expected to be present at lower molar ratios than structural proteins and therefore to be less well represented in the data sets. Greater solubility (looser binding to ECM) might also explain in part the fact that fewer are detected in our data sets to date. Finally, the list of ECM-affiliated proteins may well be somewhat inflated, since we included in our lists entire families of proteins if *any* members of the family were identified in any of our experimental samples (examples include annexins, galectins, mucins, etc.), and this will increase the probability that individual proteins may not be found in particular samples. The purpose was to avoid ignoring potential matrisome components but this inclusive nature of the ECM-associated category inevitably increases the probability that such proteins will not be found in particular samples. Some of these proteins may indeed not be ECM-associated.

We anticipate that further analyses will allow detection of specific subsets of proteins present at lower levels in the ECM-enriched fractions. Targeted proteomic methods (such as Accurate Inclusion Mass Screening or Multiple Reaction

Monitoring, etc.) are designed to detect specific peptides in samples and are particularly useful to detect low-abundance proteins within a complex mixture (49–51). Using as resources the *in silico* and experimentally determined matrices reported here, one can now select for each matrixome protein, signature peptides particularly likely to be detected by MS (49, 52) and use this peptide list for enhanced detection of lower abundance proteins of interest. Such approaches based on our lists of candidate proteins will, no doubt, allow the identification of more low-abundance proteins such as growth factors. Similar methods could also be used to detect alterations in solubility of particular proteins by analysis of the extracted fractions.

Identifying the Origin of the Tumor Extracellular Matrix—Recent research in cancer biology has demonstrated that tumor cells not only need to accumulate genetic alterations allowing them to proliferate but need also to be surrounded by a locally permissive microenvironment, which markedly affects tumor progression (53, 54). In this context, the surrounding stroma, including the tumor extracellular matrix, is of great interest. Previous proteomic analyses of tumors have not focused on the ECM, although ECM proteins were detected among the tumor proteins analyzed (55, 56).

We demonstrate here that our methodology can be applied to characterize specifically and in detail the ECM of tumors. Moreover, mass spectrometry is a method of choice to distinguish human from murine protein sequences and thus can address the question of the tumor or stromal origins of tumor extracellular matrix in human into mouse xenograft systems. Our data prove that the origin of the tumor extracellular matrix is dual and comes from both the tumor cells and the stromal cells, as might well have been expected. However, our analyses provide semi-quantitative estimates of the origins (tumor cell or stromal cell) for over a hundred ECM proteins, which has not been feasible before. Furthermore, we show that the matrixome components secreted by the tumor cells vary significantly with the metastatic potential of the tumor cells. Moreover, the matrixome components secreted by the stromal cells also change in response to the metastatic potential of the tumor cells, indicating significant cross-talk between the tumor cells and the stromal cells. This study is the first to report the detailed composition and origin of the tumor extracellular matrix at the protein level. It will be of interest to investigate the significance of these changes for tumor progression.

In conclusion, the strategy presented offers a new perspective to the analysis of the extracellular matrix. Moreover, the methods described here can reveal the presence and origins of novel or unsuspected proteins within the ECM and, thus, novel molecular mechanisms by which the ECM could influence cancer progression and metastasis formation. In addition to revealing novel molecular players whose functional contributions can then be investigated, this approach has the potential to identify biomarkers that can serve as prognostic

and diagnostic tools. The methods described here are generalizable and portable and similar strategies could readily be applied to many other physiological (e.g. development, angiogenesis) and pathological (e.g. fibrosis, wound healing, genetic disease) processes that involve important changes in ECM.

Acknowledgments—We thank Denise Crowley for histology, Tonje Steigedal for assistance in developing the list of secreted factors, Myriam Labelle, John Lamar, and Christopher Turner for helpful critiques of the manuscript and Charlie Whittaker and the members of the Hynes Lab for many helpful discussions.

* This work was supported by grants from the National Cancer Institute (U54 CA126515); the Koch Institute, the Broad Institute of MIT and Harvard and the Howard Hughes Medical Institute, of which ROH is an Investigator and AN a Postdoctoral Associate.

§ This article contains [supplemental Extended Experimental Procedures, Figs. S1 to S3 and Tables S1 to S8](#).

** To whom correspondence should be addressed: ROH, Koch Institute for Integrative Cancer Research at MIT, Room 76-361D, Cambridge, MA 02139. Tel.: 617-253-6422; Fax: 617-253-8357; E-mail: rohynes@mit.edu. Steven A. Carr, Broad Institute of MIT and Harvard - Proteomics Platform, Cambridge, MA 02139. Tel.: 617-714-7630; Fax: 617-252-1902; E-mail: scarr@broadinstitute.org

‡‡ These authors contributed equally to this work.

REFERENCES

1. Hynes, R. O. (2002) Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687
2. van der Flier, A., and Sonnenberg, A. (2001) Function and interactions of integrins. *Cell Tissue Res.* **305**, 285–298
3. Frantz, C., Stewart, K. M., and Weaver, V. M. (2010) The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200
4. Schwartz, M.A. (2010) Integrins and extracellular matrix in mechanotransduction. In *Extracellular Matrix Biology*, Hynes R.O., and Yamada, K.M. eds. *Cold Spring Harb Perspect. Biol.* doi: 10.1101/cshperspect.a005066
5. Hynes, R. O. (2009) The extracellular matrix: not just pretty fibrils. *Science*. **326**, 1216–1219
6. Mott, J. D., and Werb, Z. (2004) Regulation of matrix biology by matrix metalloproteinases. *Curr. Opin. Cell Biol.* **16**, 558–564
7. Nelson, C. M., and Bissell, M. J. (2006) Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu. Rev. Cell Dev. Biol.* **22**, 287–309
8. Aszódi, A., Legate, K. R., Nakchbandi, I., and Fässler, R. (2006) What mouse mutants teach us about extracellular matrix function. *Annu. Rev. Cell Dev. Biol.* **22**, 591–621
9. Bateman, J. F., Boot-Handford, R. P., and Lamandé, S. R. (2009) Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. *Nat. Rev. Genet.* **10**, 173–183
10. Wilson, R. (2010) The extracellular matrix: an underexplored but important proteome. *Expert Rev. Proteomics* **7**, 803–806
11. Cretu, A., and Brooks, P. C. (2007) Impact of the non-cellular tumor microenvironment on metastasis: potential therapeutic and imaging opportunities. *J. Cell. Physiol.* **213**, 391–402
12. van Kempen, L. C., Ruiter, D. J., van Muijen, G. N., and Coussens, L. M. (2003) The tumor microenvironment: a critical determinant of neoplastic evolution. *Eur. J. Cell Biol.* **82**, 539–548
13. Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54
14. Wong, S. Y., Crowley, D., Bronson, R. T., and Hynes, R. O. (2008) Analyses of the role of endogenous SPARC in mouse models of prostate and breast cancer. *Clin. Exp. Metastasis* **25**, 109–118
15. Eckhardt, B. L., Parker, B. S., van Laar, R. K., Restall, C. M., Natoli, A. L., Tavaría, M. D., Stanley, K. L., Sloan, E. K., Moseley, J. M., and Anderson,

- R. L. (2005) Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. *Mol. Cancer Res.* **3**, 1–13
16. Xu, L., Begum, S., Hearn, J. D., and Hynes, R. O. (2006) GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9023–9028
 17. Erler, J. T., and Weaver, V. M. (2009) Three-dimensional context regulation of metastasis. *Clin. Exp. Metastasis* **26**, 35–49
 18. Cukierman, E., and Bassi, D. E. (2010) Physico-mechanical aspects of extracellular matrix influences on tumorigenic behaviors. *Semin. Cancer Biol.* **20**, 139–145
 19. Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., McClay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V. D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., Bellé, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Genevière, A. M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., McCafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K. F., Brandhorst, B. P., Whittle, J., Berney, K., Bottjer, D. J., Calestani, C., Peterson, K., Chow, E., Yuan, Q. A., Elhaik, E., Graur, D., Reese, J. T., Bosdet, I., Heesun, S., Marra, M. A., Schein, J., Anderson, M. K., Brockton, V., Buckley, K. M., Cohen, A. H., Fugmann, S. D., Hibino, T., Loza-Coll, M., Majeske, A. J., Messier, C., Nair, S. V., Pancer, Z., Terwilliger, D. P., Agca, C., Arboleda, E., Chen, N., Churcher, A. M., Hallböök, F., Humphrey, G. W., Idris, M. M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R. P., Raible, F., Rowe, M., Taylor, J. S., Tessmar-Raible, K., Wang, D., Wilson, K. H., Yaguchi, S., Gaasterland, T., Galindo, B. E., Gunaratne, H. J., Juliano, C., Kinukawa, M., Moy, G. W., Neill, A. T., Nomura, M., Raisch, M., Reade, A., Roux, M. M., Song, J. L., Su, Y. H., Townley, I. K., Voronina, E., Wong, J. L., Amore, G., Branno, M., Brown, E. R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Röttinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayake, A., Wu, S. Y., Xu, R., Brown, C. T., Chen, L., Gray, R. F., Lee, P. Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L. V., Okwuonu, G., Parker, D., Pu, L. L., Thorn, R., and Wright, R. (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952
 20. Whittaker, C. A., Bergeron, K. F., Whittle, J., Brandhorst, B. P., Burke, R. D., and Hynes, R. O. (2006) The echinoderm adhesome. *Dev. Biol.* **300**, 252–266
 21. Hynes, R. O., and Zhao, Q. (2000) The evolution of cell adhesion. *J. Cell Biol.* **150**, F89–96
 22. Engel, J. (1996) Domain organizations of modular extracellular matrix proteins and their evolution. *Matrix Biol.* **15**, 295–299
 23. Hohenester, E., and Engel, J. (2002) Domain structure and organisation in extracellular matrix proteins. *Matrix Biol.* **21**, 115–128
 24. Jung, J., Ryu, T., Hwang, Y., Lee, E., and Lee, D. (2010) Prediction of extracellular matrix proteins based on distinctive sequence and domain characteristics. *J. Comput. Biol.* **17**, 97–105
 25. Manabe, R., Tsutsui, K., Yamada, T., Kimura, M., Nakano, I., Shimono, C., Sanzen, N., Furutani, Y., Fukuda, T., Ogur, Y., Shimamoto, K., Kiyozumi, D., Sato, Y., Sado, Y., Senoo, H., Yamashina, S., Fukuda, S., Kawai, J., Sugiura, N., Kimata, K., Hayashizaki, Y., and Sekiguchi, K. (2008) Transcriptome-based systematic identification of extracellular matrix proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12849–12854
 26. Adams, J. C., and Engel, J. (2007) Bioinformatic analysis of adhesion proteins. *Methods Mol. Biol.* **370**, 147–172
 27. Didangelos, A., Yin, X., Mandal, K., Baumert, M., Jahangiri, M., and Mayr, M. (2010) Proteomics characterization of extracellular space components in the human aorta. *Mol. Cell. Proteomics* **9**, 2048–2062
 28. Balasubramani, M., Schreiber, E. M., Candiello, J., Balasubramani, G. K., Kurtz, J., and Halfter, W. (2010) Molecular interactions in the retinal basement membrane system: A proteomic approach. *Matrix Biol.* **6**, 471–483
 29. Didangelos, A., Yin, X., Mandal, K., Saje, A., Smith, A., Xu, Q., Jahangiri, M., and Mayr, M. (2011) Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. *Mol. Cell. Proteomics* **10**, M111.008128
 30. Hattar, R., Maller, O., McDaniel, S., Hansen, K. C., Hedman, K. J., Lyons, T. R., Lucia, S., Wilson, R. S., Jr., and Schedin, P. (2009) Tamoxifen induces pleiotrophic changes in mammary stroma resulting in extracellular matrix that suppresses transformed phenotypes. *Breast Cancer Res.* **11**, R5
 31. Hansen, K. C., Kiemele, L., Maller, O., O'Brien, J., Shankar, A., Fornetti, J., and Schedin, P. (2009) An in-solution ultrasonication-assisted digestion method for improved extracellular matrix proteome coverage. *Mol. Cell. Proteomics* **8**, 1648–1657
 32. Wilson, R., Diseberg, A. F., Gordon, L., Zivkovic, S., Tatarczuch, L., Mackie, E. J., Gorman, J. J., and Bateman, J. F. (2010) Comprehensive profiling of cartilage extracellular matrix formation and maturation using sequential extraction and label-free quantitative proteomics. *Mol. Cell. Proteomics* **9**, 1296–1313
 33. Eyre, D. R. (1991) The collagens of articular cartilage. *Semin. Arthritis Rheum.* **21**, 2–11
 34. Robins, S. P. (2007) Biochemistry and functional significance of collagen cross-linking. *Biochem. Soc. Trans* **35**, 849–852
 35. Lawler, J., Sunday, M., Thibert, V., Duquette, M., George, E. L., Rayburn, H., and Hynes, R. O. (1998) Thrombospondin-1 is required for normal murine pulmonary homeostasis and its absence causes pneumonia. *J. Clin. Invest.* **101**, 982–992
 36. Allen, A., Hutton, D. A., and Pearson, J. P. (1998) The MUC2 gene product: a human intestinal mucin. *Int. J. Biochem. Cell Biol.* **30**, 797–801
 37. Hufleit, M. E., and Leffler, H. (2004) Galectin-4 in normal tissues and cancer. *Glycoconj. J.* **20**, 247–255
 38. Mäki, J. M. (2009) Lysyl oxidases in mammalian development and certain pathological conditions. *Histol. Histopathol.* **24**, 651–660
 39. Colombatti, A., Doliana, R., Bot, S., Canton, A., Mongiat, M., Mungiguerra, G., Paron-Cilli, S., and Spessotto, P. (2000) The EMILIN protein family. *Matrix Biol.* **19**, 289–301
 40. Tilman, G., Mattiussi, M., Brasseur, F., van Baren, N., and Decottignies, A. (2007) Human periostin gene expression in normal tissues, tumors and melanoma: evidences for periostin production by both stromal and melanoma cells. *Mol. Cancer.* **6**, 80–92
 41. Oikawa, Y., Hansson, J., Sasaki, T., Rousselle, P., Domogatskaya, A., Rodin, S., Tryggvason, K., and Patarroyo, M. (2011) Melanoma cells produce multiple laminin isoforms and strongly migrate on $\alpha 5$ laminin (s) via several integrin receptors. *Exp. Cell Res.* **317**, 1119–1133
 42. Spicer, A. P., Joo, A., and Bowling, R. A., Jr. (2003) A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes: the missing links. *J. Biol. Chem.* **278**, 21083–21091
 43. Xu, L., Shen, S. S., Hoshida, Y., Subramanian, A., Ross, K., Brunet, J. P., Wagner, S. N., Ramaswamy, S., Mesirov, J. P., and Hynes, R. O. (2008) Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. *Mol. Cancer Res.* **6**, 760–769
 44. Yau, C., Esserman, L., Moore, D. H., Waldman, F., Sninsky, J., and Benz, C. C. (2010) A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res.* **12**, R85
 45. Zhuang, Z., Jian, P., Longjiang, L., Bo, H., and Wenlin, X. (2010) Oral cancer cells with different potential of lymphatic metastasis displayed distinct biologic behaviors and gene expression profiles. *J. Oral Pathol. Med.* **39**,

168–175

46. Uhlen, M., Bjorling, E., Agaton, C., Szgyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., *et al.* (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932
47. Roughley, P. J., and Lee, E. R. (1994) Cartilage proteoglycans: structure and potential functions. *Microsc. Res. Tech.* **28**, 385–397
48. Barros, C. S., Franco, S. J., and Müller, U. (2011) Extracellular Matrix: Functions in the Nervous System. In *Extracellular Matrix Biology*, Hynes, R. O., and Yamada, K. M. eds. *Cold Spring Harb Perspect Biol.* doi: 10.1101/cshperspect.a005108
49. Jaffe, J. D., Keshishian, H., Chang, B., Addona, T. A., Gillette, M. A., and Carr, S. A. (2008) Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. *Mol. Cell. Proteomics* **7**, 1952–1962
50. Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A. J. L., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Hiltke, T., Jackson, A., Johansen, E. B., Kinsinger, C. R., Li, J., Mesri, M., Neubert, T. A., Niles, R. K., Pulsipher, T. C., Ransohoff, D., Rodriguez, H., Rudnick, P. A., Smith, D., Tabb, D. L., Tegeler, T. J., Variyath, A. M., Vega-Montoto, L. J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J. R., Zhao, L., Anderson, N. L., Fisher, S. J., Liebler, D. C., Paulovich, A. G., Regnier, F. E., Tempst, P., and Carr, S. A. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641
51. Keshishian, H., Addona, T., Burgess, M., Mani, D. R., Shi, X., Kuhn, E., Sabatine, M. S., Gerszten, R. E., and Carr, S. A. (2009) Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **8**, 2339–2349
52. Fusaro, V. A., Mani, D. R., Mesirov, J. P., and Carr, S. A. (2009) Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.* **27**, 190–198
53. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646–674
54. Egeblad, M., Nakasone, E. S., and Werb, Z. (2010) Tumors as organs: complex tissues that interface with the entire organism. *Dev. Cell* **18**, 884–901
55. Zanivan, S., Gnad, F., Wickström, S. A., Geiger, T., Macek, B., Cox, J., Fässler, R., and Mann, M. (2008) Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J. Proteome Res.* **7**, 5314–5326
56. Sutton, C. W., Rustogi, N., Gurkan, C., Scally, A., Loizidou, M. A., Hadjisavvas, A., and Kyriacou, K. (2010) Quantitative proteomic profiling of matched normal and tumor breast tissues. *J. Proteome Res.* **9**, 3891–3902
57. Martin, G. R., Kleinman, H. K., Terranova, V. P., Ledbetter, S., and Hassell, J. R. (1984) The regulation of basement membrane formation and cell-matrix interactions by defined supramolecular complexes. *Ciba Found Symp.* **108**, 197–212