Transcriptome-based systematic identification of extracellular matrix proteins

Ri-ichiroh Manabe*†, Ko Tsutsui*†, Tomiko Yamada*, Mina Kimura*, Itsuko Nakano*†, Chisei Shimono*†, Noriko Sanzen*†, Yutaka Furutani*, Tomohiko Fukuda*, Yasuko Oguri*, Keiko Shimamoto*, Daiji Kiyozumi*†, Yuya Sato†, Yoshikazu Sado‡, Haruki Senoo§, Shohei Yamashina¶, Shiro Fukuda‖, Jun Kawai‖, Nobuo Sugiura**, Koji Kimata**, Yoshihide Hayashizaki‖, and Kiyotoshi Sekiguchi*†,††

*Sekiguchi Biomatrix Signaling Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, c/o Aichi Medical University, Nagakute, Aichi 480-1195, Japan; †Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan; ‡Shigei Medical Research Institute, Okayama, Okayama 701-0202, Japan; §Akita University School of Medicine, Akita, Akita 010-8543, Japan; ¶Kitasato University School of Medicine, Sagamihara, Kanagawa 228-8555, Japan; ‖RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045, Japan; and **Institute for Molecular Science of Medicine, Aichi Medical University, Nagakute, Aichi 480-1195, Japan

Communicated by Sen-itiroh Hakomori, Pacific Northwest Research Institute and University of Washington, Seattle, WA, April 25, 2008 (received for review January 2, 2008)

Extracellular matrix (ECM), which provides critical scaffolds for all adhesive cells, regulates proliferation, differentiation, and apoptosis. Different cell types employ customized ECMs, which are thought to play important roles in the generation of so-called niches that contribute to cell-specific functions. The molecular entities of these customized ECMs, however, have not been elucidated. Here, we describe a strategy for transcriptome-wide identification of ECM proteins based on computational screening of >60,000 full-length mouse cDNAs for secreted proteins, followed by *in vitro* functional assays. These assays screened the candidate proteins for ECM-assembling activities, interactions with other ECM molecules, modifications with glycosaminoglycans, and cell-adhesive activities, and were then complemented with immunohistochemical analysis. We identified 16 ECM proteins, of which seven were localized in basement membrane (BM) zones. The identification of these previously unknown BM proteins allowed us to construct a body map of BM proteins, which represents the comprehensive immunohistochemistry-based expression profiles of the tissue-specific customization of BMs.

basement membrane | body map | niche | cell adhesion | glycosaminoglycan

The extracellular environments provide cues for the determination of cell fates and functions. Extracellular matrix (ECM), a major constituent of the extracellular environment, is of particular interest because it modulates the activities of other extracellular factors, including soluble (e.g., growth factors) and insoluble ligands (e.g., cell–cell adhesion molecules) as well as physical stimuli. For example, ECM modulates the activities of growth factors and cytokines via interactions with these soluble ligands (1). ECM also transduces signals that influence cell–cell interactions (2) and growth factor signaling (3) through integrins and other cell surface receptors, thereby integrating these extracellular cues. ECM exhibits a marked degree of molecular diversity that is thought to be important for the generation of environmental niches occupied by individual cell types. However, it is currently very difficult to define the complement of proteins that constitute the ECM of a given tissue or cell type, because many ECM proteins have yet to be identified and no large-scale comparison of the spatiotemporal distribution of ECM proteins has been reported. Thus, a large-scale study to identify ECM proteins would be helpful for the identification of various complements of ECM proteins.

Genes encoding ECM proteins are estimated to represent 1.3–1.5% of the genes in mammalian genomes (http://www.pantherdb.org/genes/). Because there are ≈25,000 protein-encoding genes in a mammalian genome, there are estimated to be 300–400 ECM genes, one third of which have yet to be identified. Although recent innovations in analytical technology and the accumulation of various scientific resources offer some large-scale approaches for protein identification, such proteomic methods may not be applicable for ECM proteins because of their complex posttranslational modifications and the difficulties associated with isolating high-quality ECMs from individual tissues. In addition, there are no motifs or signatures that define ECM proteins, which precludes simple sequence-based screening techniques.

To overcome these difficulties, we developed an approach for identifying unknown ECM proteins that combines computational screening for secreted proteins from a mouse transcriptome database with *in vitro* functional screening and immunohistochemical analysis. This strategy led to the identification of 16 ECM proteins, including 7 basement membrane (BM) proteins. These findings prompted us to use immunohistochemistry to delineate the molecular composition of various BMs, a specialized subset of ECMs associated with epithelial, endothelial, muscle, and nerve cells. In addition, we have detailed the localization profiles of BM proteins in the epithelial BMs of developing molars.

Results

ECM proteins are secreted from cells and localized to ECMs as a result of their self-assembling activities, the intrinsic cell-mediated assembly of ECM, and/or binding affinities for other ECM components. Some ECM proteins are capable of promoting cell–substrate adhesion, whereas members of the proteoglycan subfamily of ECM proteins possess glycosaminoglycan (GAG) chains. By using these characteristics as hallmarks of ECM proteins, we developed a systematic screening procedure that consisted of three steps (Fig. 1). First, cDNAs from a mouse transcriptome database were computationally screened for clones encoding putative secreted proteins. Then, a series of *in vitro* functional screening assays were used to select putative ECM proteins based on their deposition to ECMs, interactions with known ECM molecules, promotion of cell adhesion, and GAG modifications. Finally, the localizations of the proteins were examined by using immunohistochemistry.

Author contributions: R.-i.M. and K. Sekiguchi designed research; R.-i.M., K.T., T.Y., M.K., I.N., C.S., N. Sanzen, Y.F., T.F., Y.O., and K. Shimamoto performed research; D.K., Y. Sato, Y. Sado, S.F., J.K., N. Sugiura, K.K., and Y.H. contributed new reagents/analytic tools; R.-i.M., K.T., C.S., N. Sanzen, Y.F., T.F., H.S., S.Y., S.F., J.K., N. Sugiura, K.K., Y.H., and K. Sekiguchi analyzed data; and R.-i.M. and K. Sekiguchi wrote the paper.

The authors declare no conflict of interest.

††To whom correspondence should be addressed. E-mail: sekiguch@protein.osaka-u.ac.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803640105/DCSupplemental.

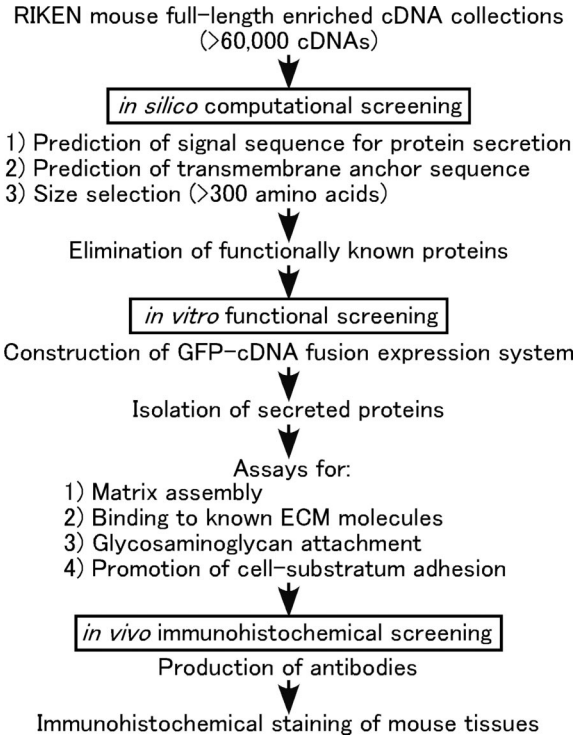

Fig. 1. The scheme for systematically screening for ECM proteins. For details, see the text and *Materials and Methods*.

***In Silico* Screening for Secreted Proteins.** Among the currently available cDNA collections, the cDNA collections from RIKEN are excellent resources because they include >90% of the predicted mouse gene products. Furthermore, >70% of the cDNAs from these collections are full-length (4), and can therefore be used to express full-length proteins for functional screening. We used the following criteria for computational screening of >60,000 cDNAs to identify secreted proteins: (*i*) the presence of a signal sequence at the N terminus; (*ii*) an absence of transmembrane domains; (*iii*) an ORF coding for at least 300 amino acid residues; and (*iv*) functionally unknown. These criteria are based on the fact that most known ECM genes encode proteins that are >300 aa in length, including an N-terminal secretion signal sequence. A total of 181 clones met the selection criteria.

To verify that the candidate proteins were secreted, we established a high-throughput expression system in which the full-length proteins encoded by candidate cDNAs were expressed in mammalian cells with C-terminal GFP tags, and their secretion into conditioned media was examined by Western blot analyses using anti-GFP mAb. Among the 181 clones, 146 clones were successfully expressed, and 93 of these proteins were secreted into the media (data not shown).

***In Vitro* Functional Screening for Candidate ECM Proteins.** The secreted proteins were subjected to *in vitro* high-throughput functional screening assays. The abilities of the candidate proteins to form pericellular deposits were examined by using *de novo* expression of GFP-tagged proteins in 293T cells or incubation of differentiated myoblasts or mouse embryonic fibroblasts (MEFs) in conditioned media containing the secreted proteins [Fig. 2*A* and supporting information (SI) Fig. S1]. To screen for interactions with known ECM molecules, conditioned medium containing the candidate proteins was placed in 96-well plates preadsorbed with 19 different ECM molecules, including various types of collagens, laminins, and GAGs as well as fibronectin. Bound proteins were quantified by using the fluorescent signals derived from the GFP tags (Fig. 2*B* and Fig. S2). To screen for GAG modifications, mobility shifts of the secreted proteins after treatment with enzymes that cleaved heparan sulfate and chondroitin sulfate chains were examined by Western blot analysis. The types of GAGs and their covalent attachments to the core proteins were further analyzed by using separate treatments with heparan sulfate- or chondroitin sulfate-cleaving enzymes, followed by Western blot analysis with anti-GFP mAb and various anti-stub mAbs that specifically recognized neoepitopes generated by the enzymes (Fig. 2*C* and Fig. S3). To screen for the promotion of cell–substrate adhesion, the secreted proteins were captured on 96-well plates via anti-GFP antibody and used as substrates in cell adhesion assays by using MG63 osteosarcoma, HT1080 fibrosarcoma, and HeLa cells (Fig. 2*D* and Fig. S4). These *in vitro* functional assays identified 24 proteins as putative ECM proteins of which 19 localized to ECMs, 14 bound to known ECM proteins, 4 contained GAG chains, and 11 promoted cell–substrate adhesion (summarized in Fig. 3).

To assess the reliability and specificity of our strategy, we performed pilot experiments in which 26 cDNAs encoding known ECM proteins were selected from RIKEN's cDNA collections and subjected to the *in vitro* assays. Each of these cDNAs conformed to the criteria for the computational screening and the encoded proteins were secreted after transfection. Twenty-three of these proteins were positive in at least one of the *in vitro* assays (Table S2), verifying the utility of our strategy.

The *in vitro* assays revealed a marked degree of functional diversity among the selected proteins. Among the 19 ECM-assembling proteins, 18 were immobilized in the ECM of either 293T cells or differentiated myoblasts, whereas only three were capable of assembling into both ECMs (Fig. 3 and Fig. S1). A number of the proteins were found to bind with heparin (e.g., ECM482/eratin), although some of them displayed more specific binding to dermatan sulfate (e.g., ECM331/coffeecrisp) or chondroitin sulfate E (e.g., ECM392/ependolin) (Fig. 2*B*). Some of the proteins exhibited affinities for collagens with different binding spectra (e.g., ECM482/eratin). Among the 11 cell-adhesive proteins, seven showed some degree of cell-type specificity; for example, ECM482/eratin mediated the adhesion of MG63 and HT1080 cells, but not HeLa cells, whereas ECM742/cradin mediated the adhesion of HeLa and HT1080 cells, but not MG63 cells (Fig. 2*D* and Fig. S4). Five of the cell-adhesive proteins contained an RGD motif (Fig. 3), and at least two of them [ECM742/cradin and ECM661/MAEG (5)] promoted cell adhesion in an RGD-dependent manner (Fig. S4). These differences are indicative of the functional diversity of the selected proteins and support the use of the complementary *in vitro* screening assays.

Immunohistochemical Localization of the Candidate Proteins. To confirm that the selected proteins were indeed ECM proteins, we immunohistochemically examined their tissue localizations. Among the 24 candidate proteins, at least 16 were confirmed to be ECM proteins based on their localizations in ECM structures (Fig. 2*E* and Fig. S5). These proteins displayed clear tissue-specific expression patterns; three were localized in the periosteum, three in the perichondria, four in tendons and/or ligaments, seven in cartilage, and seven at the BM zones of various cell types (Fig. 3 and Fig. S5). Antibodies against six other proteins gave restricted staining patterns that were not indicative of ECM proteins (Fig. S6).

More than half of the proteins thus identified were encoded by the genes that are currently only represented by mammalian ESTs in public databases. Many of the proteins contained consensus domains frequently found in ECM proteins. Fibronectin type III repeats were found in four proteins; von Willebrand factor type A and type C domains, MAM domains, and multiple

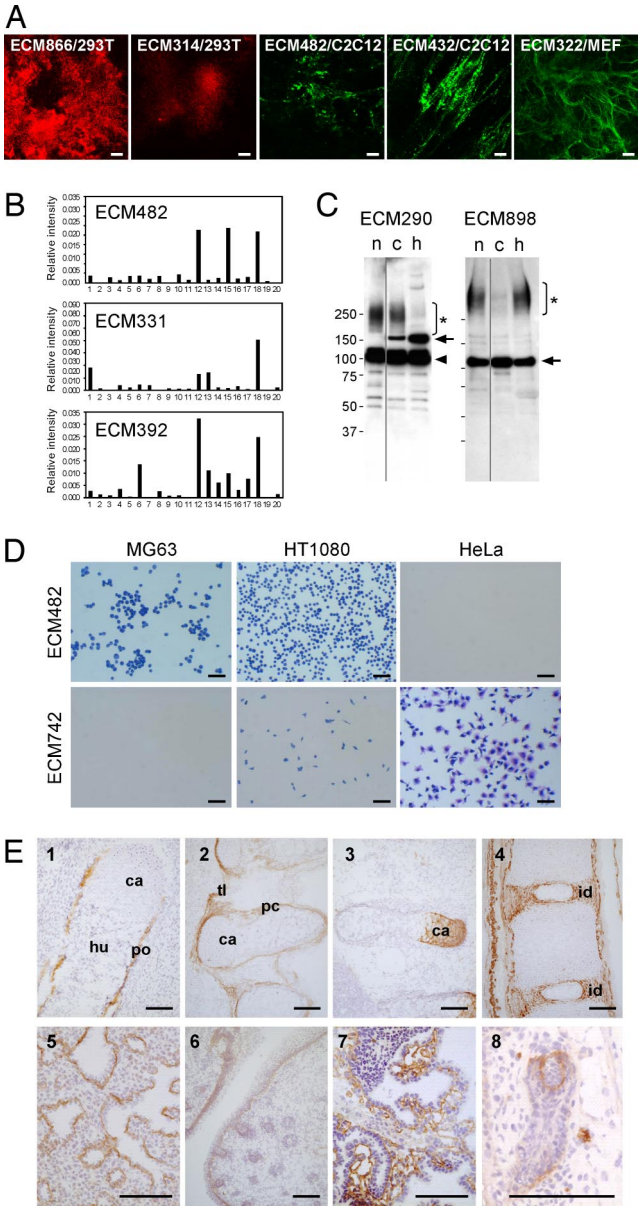

Fig. 2. Representative results obtained in the screening assays. (*A*) Pericellular deposits were examined by using the *de novo* expression of GFP-tagged candidate proteins in 293T cells and incubation of differentiated C2C12 myoblasts or MEFs in conditioned media containing proteins secreted from transfected 293F cells. (Scale bars, 10 μm.) (*B*) The ECM molecules used in the solid-phase binding assays were as follows: 1, collagen I; 2, collagen II; 3, collagen III; 4, collagen IV; 5, collagen V; 6, collagen VI; 7, gelatin; 8, fibronectin; 9, laminin-111; 10, laminin-211/221; 11, laminin-511/521; 12, heparin; 13, heparan sulfate; 14, chondroitin sulfate A; 15, dermatan sulfate; 16, chondroitin sulfate C; 17, chondroitin sulfate D; 18, chondroitin sulfate E; 19, hyaluronic acid; and 20, BSA (negative control). The amount of bound GFP-fusion protein was determined by measuring the intensity of GFP fluorescence. (*C*) Attachment of GAG chains was examined by incubating GFP-fusion proteins in conditioned media from transfected 293F (ECM290) or COS (ECM898) cells with heparinase and heparitinase (*h*), chondroitinase ABC (*c*), or buffer alone (*n*), followed by Western blot analyses using anti-GFP mAb to detect mobility shifts of the candidate proteins. After incubation with GAG-degrading enzymes, the smeared bands (asterisks) disappeared (or diminished) with a concomitant appearance (or increase) of the faster migrating bands (arrows). A putative proteolytic fragment of the ECM290-GFP fusion protein is indicated by an arrowhead. (*D*) The cell adhesion-promoting activities were examined by incubating MG63, HT1080, and HeLa cells in 96-well plates containing immobilized GFP-tagged candidate proteins. Cells adhering to the substrates were visualized by staining with Diff-Quick. (Scale bars, 100 μm.) (*E*) Immunohistochemical localizations of the candidate proteins. Sagittal sections of E16.5 mouse embryos were stained with affinity-purified antibodies. The following representative localization data are shown: ECM201 (*1*), periosteum of the humerus; ECM322 (*2*), ligaments and associated perichondrium surrounding the spinal cartilage; ECM314 (*3*), rib cartilage; ECM306 (*4*), intervertebral disks in the spinal cord; ECM661 (*5*), lung epithelium; ECM392 (*6*), lip epithelium; ECM270 (*7*), choroid plexus epithelium and associated blood vessels; ECM290 (*8*), hair follicle. ca, cartilage; pc, perichondrium; po, periosteum; tl, tendon or ligament; id, intervertebral disk; hu, humerus. (Scale bars, 100 μm.)

small leucine-rich repeats were found in two proteins, respectively (Fig. 3). One protein (ECM290/nectican) showed similarity with agrin in that both proteins shared a domain structure characterized by the presence of three laminin G-like domains interspersed with EGF-like domains. It should be noted that the frequency of such ECM domains in the proteins that were selected with the *in silico* screening but failed in the *in vitro* screening assays was significantly lower than the frequency in the proteins that passed the *in vitro* assays (Table S3). Moreover, three proteins without any known domains (ECM392/ependolin, ECM314/emprin, and ECM517/RAINB2) were also identified as ECM proteins, underscoring the advantages of our function-based and localization-based screening strategies compared with screening based solely on the primary protein structure.

Comprehensive Profiling of BM Proteins in Epithelial BMs of Developing Teeth. Notably, 7 of the 16 newly identified ECM proteins were detected, at least in part, in BM zones. BMs are thin sheets of ECM with limited protein compositions, which primarily include laminins, type IV collagens, nidogens, and perlecan (7). Currently, 46 genes that encode BM proteins have been identified, including the seven genes identified in the present study (Table S4). To gain insight into how extracellular environments are customized for individual cell types, we set out to survey the comprehensive expression profiles of all known BM proteins in mouse embryos. We produced antibodies against 19 known BM proteins, including individual laminin subunits, to complement the antibodies already available in our laboratories or commercially, and labeled sagittal and frontal sections of embryonic day (E) 16.5 mouse embryos with 38 different antibodies against BM proteins, covering >80% of the BM proteins identified to date (Table S4). The resulting immunohistochemical data have been compiled into an image-based database (http://www.matrixome.com/bm), which we refer to as the "body map" of mouse BMs. Representative datasets focusing on the epithelial BMs in developing teeth are shown in Fig. 4*A* and Fig. S7.

Teeth develop through reciprocal interactions between the oral epithelium and the underlying mesenchyme (8). BMs not only serve as a physical barrier that separates the epithelium and the mesenchyme, but also generate signals that regulate the proliferation and/or differentiation of both epithelial and mesenchymal cells (9). This survey of BM protein localization revealed a remarkable degree of molecular complexity and regional customization of BMs in the developing molar. Among the 38 BM proteins examined, 30 were detected in BM zones underlying the tooth germ epithelia; 13 of these proteins were uniformly expressed throughout the BMs of the tooth germs, whereas 17 exhibited regionally restricted localizations, resulting in distinct protein compositions among the different BMs (Table S5).

In enamel epithelia, the laminin α2 and γ3 chains were predominantly expressed in the BM of the outer enamel epithelium, whereas netrin-1, Fras1, QBRICK/Frem1, ECM306/WARP, ECM661/MAEG, and ECM392/ependolin were preferentially expressed in the BM of the inner enamel epithelium (IEE). The latter subset of proteins showed distinctive localization patterns, which divided the IEE into three zones, namely, the apical (including the cervical loop), middle, and basal

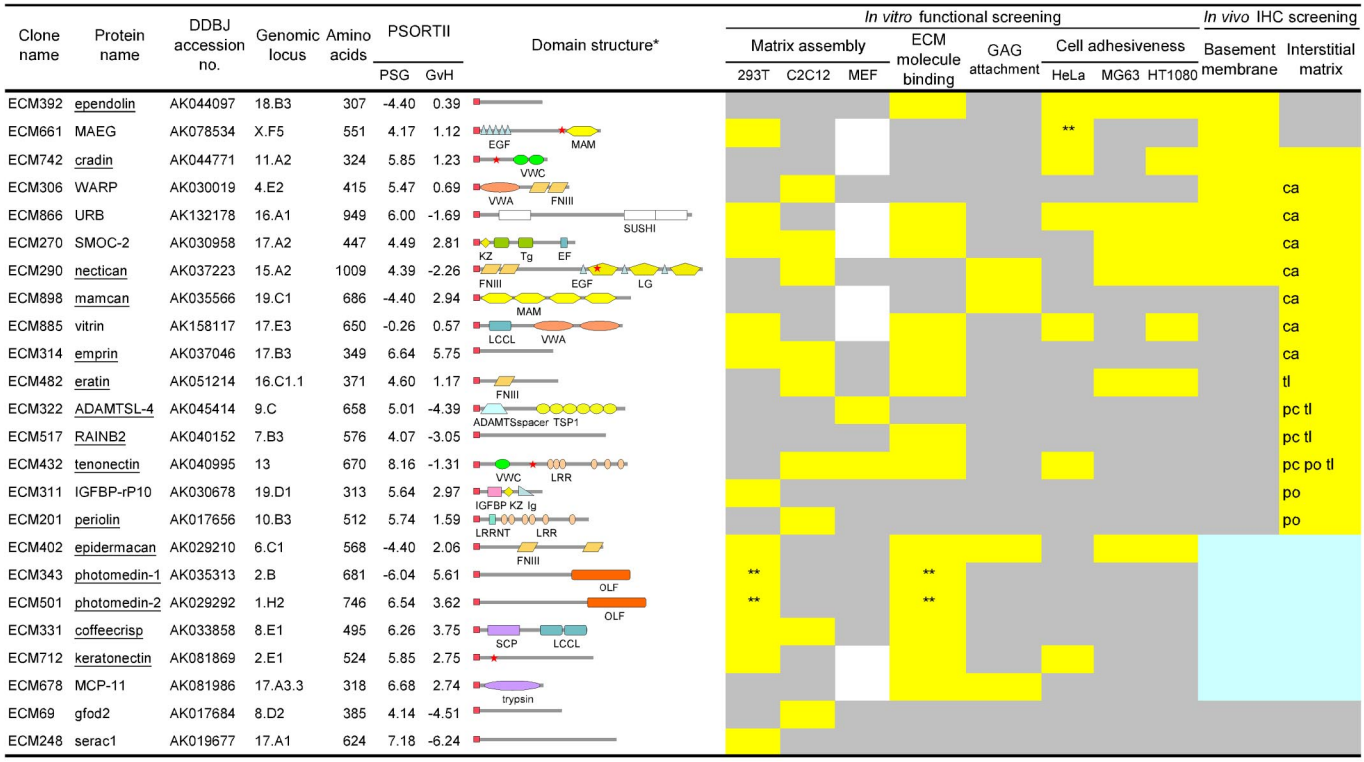

Fig. 3. A summary of the screening results. Accession numbers, genomic loci in mice, numbers of amino acids, PSG and GvH scores, and the domain structures of the putative ECM proteins encoded by the clones isolated with the *in silico* and *in vitro* screening assays are indicated. Proteins that were given new names in this study are underlined (see also Table S1). Results obtained from *in vitro* screening and *in vivo* IHC screening are also indicated on the right. Labeling in yellow and gray denotes positive and negative results, respectively. Labeling in white (blank) denotes ''not determined.'' Proteins localized at atypical ECM structures are blue. Positive staining of the interstitial matrices of cartilage (ca), tendon/ligament (tl), perichondrium (pc), and periosteum (po) are indicated. *, The predicted protein structures based on the arrangement of the Pfam domains. FNIII, fibronectin type III domain; LG, laminin G domain; VWA, von Willebrand factor type A domain; VWC, von Willebrand factor type C domain; KZ, KAZAL-type serine protease inhibitor domain; Tg, thyroglobulin type 1 repeat; EF, EF-hand calcium-binding domain; TSP1, thrombospondin type 1 domain; LRR, leucine-rich repeat; IGFBP, insulin-like growth factor binding protein; Ig, Ig domain; LRRNT, leucine-rich repeat N-terminal domain; OLF, olfactomedin-like domain. RGD motifs are denoted with *red stars*. **, The results have been published (5, 6).

(including the cusp and intercusp) zones (Fig. 4*B*). ECM306/WARP was highly expressed in the basal zone, whereas netrin-1 was enriched in the intercuspal region within the basal zone. In contrast, ECM661/MAEG was preferentially expressed in the apical and middle zones, but was only faintly detected in the basal zone. This survey also revealed that these proteins were asymmetrically distributed between the buccal and lingual sides of the outer enamel and dental lamina epithelia. Laminin $\alpha 2$ and $\gamma 3$, Fras1, QBRICK/Frem1, and nephronectin were preferentially expressed in the BM on the buccal side of these epithelia, whereas ECM661/MAEG was preferentially expressed on the lingual side. Taken together, these localization profiles strongly suggest that BMs are regionally customized and reveal compositional gradients of BM proteins along the apical–basal axis and the lingual–buccal axis in tooth germ epithelia.

Discussion

We have described a transcriptome-wide screening method for ECM proteins that combines *in silico*, *in vitro*, and immunohistochemistry screening assays; this technique identified 16 ECM proteins. For our screening strategy, the use of cDNA collections and an associated sequence database (FANTOM) (4), rather than regular cDNA libraries, was of fundamental importance. All of the cDNAs in the collections had already been cloned and fully sequenced, making it possible to computationally isolate putative secreted proteins without any redundant and/or functionally known proteins. This allowed us to effectively narrow down the number of the candidates from >60,000 to 181 clones. In addition, the availability of full-length cDNAs made it possible to construct a recombinant expression system with complete ORFs containing the intrinsic signal sequences and C-terminal GFP tags, thereby facilitating one-step screening of the secretion of the candidate proteins and high-throughput functional assays.

Among the 24 candidates isolated by using the *in vitro* assays, 22 were analyzed by immunohistochemistry and the ECM localizations of 16 of these were immunohistochemically verified, resulting in a 72% screening efficiency with the *in vitro* assays. Although this value was slightly lower than that obtained in pilot experiments using known ECM proteins (88%), the *in vitro* assays clearly enriched the candidate pool as only ≈14% of secreted proteins are ECM/adhesion proteins (10). Although our *in vitro* assays efficiently isolated ECM protein candidates, we noticed that a few known soluble proteins that were used as negative controls (i.e., fibroblast growth factor-4 and transferrin) were capable of assembling into ECMs, binding to known ECM proteins, or promoting cell attachment (data not shown), underscoring the importance of our immunohistochemical verification of the identified ECM protein candidates.

We screened >60,000 RIKEN cDNAs and identified 16 new ECM proteins. Screening other cDNA collections should identify other unknown ECM proteins. The RIKEN cDNA collections have now been expanded to >120,000 entries (11), which include unpublished clones; this has allowed us to identify >30 additional candidate cDNAs by using our *in vitro* assays (R.M., unpublished data). A cDNA collection containing longer cDNAs would be another attractive source, because the cDNAs com-

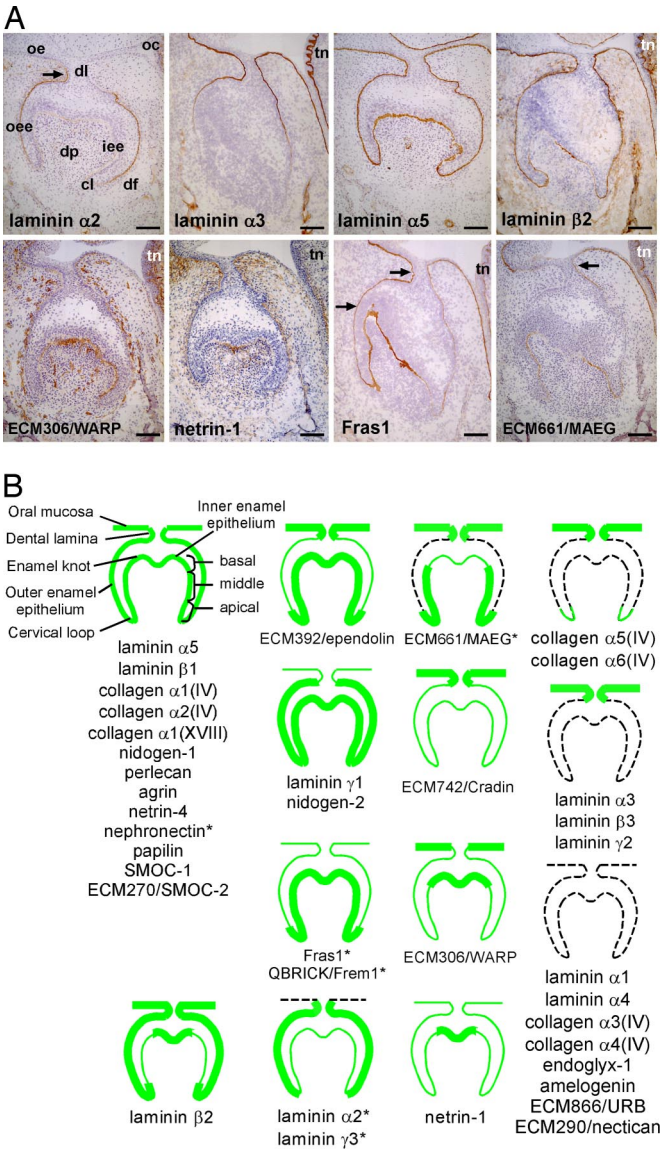

Fig. 4. Customized epithelial BMs in a developing molar. The localizations of 38 BM proteins in the bell stage of the first mandibular molar were determined immunohistochemically by using frontal sections of E16.5 mouse heads. (*A*) Representative staining patterns of the BM proteins showing distinctive localizations in the BMs of different parts of the developing molar (buccal to the left and lingual to the right). The arrows indicate asymmetries in the staining intensities found in individual epithelial BMs. oe, oral epithelium; oee, outer enamel epithelium; iee, inner enamel epithelium; dl, dental lamina; cl, cervical loop; tn, tongue; oc, oral cavity; df, dental follicle; and dp, dental papilla. (Scale bars, 100 μm.) (*B*) A schematic illustration of the regional expression patterns of individual BM proteins. High, moderate, and undetectable levels of expression in individual epithelial cell layers of the molar are illustrated as bold, thin, and dotted lines, respectively. Asterisks indicate proteins that showed asymmetric localization patterns in individual BMs. Three zones (apical, middle, and basal) of the IEE divided by the differential distributions of the BM proteins are indicated.

piled in the current RIKEN collections encode proteins that are, on average, 200–300 aa in length, and many of the cDNAs encoding known large ECM proteins (>1,500 aa) are truncated or missing. High-throughput cloning of longer full-length cDNAs (i.e., >8 kb) is technically challenging, although it has now become possible (12). Using our screening strategy with longer cDNAs should allow us to identify large ECM proteins.

ECMs for different cell and tissue types are known to differ in their compositions, providing cells with customized extracellular environments or niches (13), in which they can proliferate or differentiate. The molecular components of these customized ECMs, however, have not been systematically explored. Our comprehensive survey of the localization profiles of BM proteins has clarified the protein compositions of some individual BMs, providing a framework for understanding the environmental cues for fate determination in a diverse range of cell types. For example, we found a compositional gradient along the apical–basal axis in the BM of the IEE, which included graded expression levels of netrin-1, ECM306/WARP, and ECM661/MAEG. Interestingly, there are similar gradients in the proliferative potential of the IEE (14) and in the differentiation status of odontoblasts (15) located adjacent to the BM of the IEE.

With respect to IEE proliferation, the enamel knot cells, which reside at the cusp region of IEE and are known as signaling centers for tooth morphogenesis, do not divide, but stimulate the proliferation of nearby epithelial cells that form the cervical loops. The terminal differentiation of odontoblasts always starts from the tips of the cusps, where the enamel knot cells reside, and proceeds in a cervical or intercuspal direction. Such regional mitotic and differentiation activities are thought to be regulated by gradients of morphogen activities, including those of sonic hedgehog (Shh) and bone morphogenetic proteins (16). It has been shown that ECM molecules, particularly heparan sulfate proteoglycans, contribute to the formation of gradients of these diffusible molecules (16). Although perlecan and agrin—major heparan sulfate proteoglycans expressed in tooth germ epithelial BMs—showed no detectable graded expression patterns in the BMs of the IEE, the ECM proteins capable of binding to perlecan and/or heparan sulfate [e.g., ECM306/WARP (17) and netrin-1 (18), both of which showed graded expression patterns within the IEE] may play a role in the formation of morphogen gradients. In addition, netrin-1 was reported to cooperatively regulate axon guidance with Shh (19), suggesting that it is involved in the proliferation of the IEE through the regulation of Shh activity. Thus, locally restricted ECM components together with constitutively expressed ECM proteins and diffusible factors may orchestrate the extracellular environmental cues for cell fate determination and tissue patterning.

In summary, our systematic identification of new ECM proteins has paved the way for comprehensive profiling of the molecular compositions of ECMs that regulate cell fate and behavior. The body-map database of mouse BMs provides the first bird's-eye view of the customization of BMs in different tissues and will contribute to our understanding of the specificity of extracellular environments for individual cells. We propose that the term "matrixome" be used for the subset of the proteome that constitutes these customized microenvironments. Defining the matrixomes of individual cell types, particularly those of stem cells, should allow these customized environments or niches to be reconstituted, which will facilitate the *in vitro* manipulation of stem cells for regeneration medicine and tissue engineering.

Materials and Methods

Computational Screening. cDNAs encoding peptide chains of at least 300 aa were selected on the basis of their sizes. Signal peptides were predicted by using PSORT II (20), which calculates the probability of the presence of a signal sequence (given as a PSG score) and a cleavage site (given as a GvH score). ORFs with PSG scores >4.0 or with GvH scores > −2.1 were chosen for further analysis. ORFs with potential transmembrane regions were identified by using SOSUI (21) and were excluded from further analysis. ORFs coding for functionally known proteins or proteins considered to be mouse orthologs or alternatively spliced variants were identified by using FASTA (22) and omitted from further analysis. Known domains and motifs were identified by using InterPro (23).

Screening for ECM Deposition. 293T cells were plated in 96-well glass-bottomed plates precoated with fibronectin (10 μg/ml) and transfected with

the GFP fusion plasmids. After incubation for 3–4 days, the cells were fixed with 2% formaldehyde and labeled with anti-GFP mAb (Santa Cruz Biotechnology) under nonpermeable conditions followed by rhodamine-conjugated secondary antibodies. Alternatively, C2C12 myoblasts plated in 96-well glass-bottomed plates precoated with 0.1% gelatin were cultured in DMEM supplemented with 1% horse serum and insulin (0.4 unit/ml) to induce differentiation. At 4–5 days after the induction of differentiation, the medium was replaced with conditioned medium from transfected 293F cells (Invitrogen), and the cells were cultured for 1 day. Confluent monolayer cultures of MEFs were fed with a 1:1 mixture of complete medium and conditioned medium from transfected 293F cells and cultured for 3 days. Samples were visually screened for the deposition of GFP-fusion proteins in ECM-like structures by using an LSM5 PASCAL confocal microscope (Carl Zeiss).

Screening for ECM Molecule-Binding Activity. Black 96-well plates were coated with ECM proteins or GAGs at 4°C overnight. The GAG chains were conjugated to dipalmitoyl phosphatidylethanolamine (24) to ensure their adsorption to the plastic surface. The plates were blocked with 3% BSA and then incubated with conditioned media from transfected 293T or 293F cells overnight at 4°C. After washing the plates, the amount of bound GFP-fusion protein was determined by measuring the intensity of GFP fluorescence with a Fluoroskan Ascent fluorometer (Thermo Scientific). The mean background fluorescence intensity ($n = 2$) after subtraction of the intensity of sGFP (the negative control) was usually <0.001; therefore, we set the threshold value for positivity in the assays to 0.005. At least two independent experiments were performed and representative data are shown. When the mean background intensity was >0.001, the threshold was set to a value that was 5 times higher than the mean value.

Screening for Cell Adhesion-Promoting Activity. To immobilize the GFP-fusion proteins, conditioned media containing the GFP-fusion proteins were added to anti-GFP polyclonal antibody-coated 96-well plates and incubated at 4°C overnight. The plates were blocked with 1% BSA, and seeded with MG63, HT1080, or HeLa cells suspended in DMEM containing 0.1% BSA at a density of 3×10^3 cells per well. After incubation for 40 min, the cells were washed twice with PBS, fixed with 3.7% formaldehyde, and stained with Diff-Quik (Sysmex). Under these conditions, no background activity was detected irrespective of the examined cell types.

Screening for GAG Modifications. Recombinant proteins in conditioned media from transfected 293F or COS cells were treated with a mixture of heparitinase (0.005 unit/ml), heparinase (0.005 unit/ml), and chondroitinase ABC (0.25 unit/ml) in 50 mM sodium acetate (pH 7.0), 4 mM $CaCl_2$, and 0.2 mM Pefabloc (Roche) or with buffer alone at 37°C for 2 h. The samples were separated by SDS/PAGE under reducing conditions, and subjected to Western blot analysis with anti-GFP mAb. Proteins that migrated faster after treatment with the GAG-cleaving enzymes were treated separately with a mixture of heparinase and heparitinase or chondroitinase ABC to determine which types of GAG chains were covalently attached to the proteins.

Immunohistochemistry. Fresh-frozen sections of whole E16.5 and newborn mice or adult ICR mouse tissues were immunohistochemically stained as described elsewhere (10) with optimized fixation and pretreatment protocols for the individual primary antibodies (Table S6). The specificities of the immunohistochemical analyses were verified by using the following two criteria: (*i*) the immunoreactivities were abrogated by absorption with antigenic fragments and/or full-length recombinant proteins, and (*ii*) two separate antibodies directed against different regions of the candidate proteins produced essentially identical staining patterns. Images were captured by using a Nikon DXM1200 CCD camera fitted to an ECLIPSE E800M microscope (Nikon) and their tonal ranges were adjusted by using Adobe Photoshop.

Animals, materials, cell culture, construction of mammalian expression plasmids, screening for protein secretion, detection of neoepitopes generated by cleavage of GAG chains, and antibody production are available in the *SI Materials and Methods*. A tutorial for viewing the ''Mouse Basement Membrane Body Map'' database is also available in the *SI Appendix*.

ACKNOWLEDGMENTS. We thank T. Emoto, H. Fujiwara, and Y. Hayashi for antigen production; N. Sugimoto, C. N. Weber, T. Hasegawa, M. Kondo, and Y. Iwase for digital image construction; A. Ueda, M. Mutoh, and A. Morimoto for webpage construction; K.M. Yamada and M. Okuhara for helpful comments; T. Nishida and K. Seki for providing tissue samples; H. Nakamura and S. Futaki for web server maintenance; and M. Saito for critical discussions. This study was supported in part by New Energy and Industrial Technology Development Organization of Japan Research Contract 06001294-0 and by a grant from the RIKEN Genome Exploration Research Project.

1. Schönherr E, Hausser H-J (2000) Extracellular matrix and cytokines: A functional unit. *Dev Immunol* 7:89–101.
2. Chen X, Gumbiner B-M (2006) Crosstalk between different adhesion molecules. *Curr Opin Cell Biol* 18:572–578.
3. Comoglio PM, Boccaccio C, Trusolino L (2003) Interactions between growth factor receptors and adhesion molecules: Breaking the rules. *Curr Opin Cell Biol* 15:565–571.
4. Okazaki Y, *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563–573.
5. Osada A, *et al.* (2005) Expression of MAEG, a novel basement membrane protein, in mouse hair follicle morphogenesis. *Exp Cell Res* 30:148–159.
6. Furutani Y, *et al.* (2005) Identification and characterization of photomedins: Novel olfactomedin-domain-containing proteins with chondroitin sulphate-E-binding activity. *Biochem J* 389:675–684.
7. Erickson A-C, Couchman J-R (2000) Still more complexity in mammalian basement membranes. *J Histochem Cytochem* 48:1291–1306.
8. Thesleff I (2003) Epithelial-mesenchymal signalling regulating tooth morphogenesis. *J Cell Sci* 116:1647–1648.
9. Fukumoto S, Yamada Y (2005) Extracellular matrix regulates tooth morphogenesis. *Connect Tissue Res* 46:220–226.
10. Grimmond S-M, *et al.* (2003) The mouse secretome: Functional classification of the proteins secreted into the extracellular environment. *Genome Res* 13:1350–1359.
11. Maeda N, *et al.* (2006) Transcript annotation in FANTOM3: Mouse gene catalog based on physical cDNAs. *PLoS Genet* 2:498–503.
12. Carninci P, Shiraki T, Mizuno Y, Muramatsu M, Hayashizaki Y (2002) Extra-long first-strand cDNA synthesis. *Biotechniques* 32:984–985.
13. Scadden D-T (2006) The stem-cell niche as an entity of action. *Nature* 441:1075–1079.
14. Thesleff I, Keranen S, Jernvall J (2001) Enamel knots as signaling centers linking tooth morphogenesis and odontoblast differentiation. *Adv Dent Res* 15:14–18.
15. Lesot H, *et al.* (2001) Epigenetic signals during odontoblast differentiation. *Adv Dent Res* 15:8–13.
16. Tabata T, Takei Y (2004) Morphogens, their identification and regulation. *Development (Cambridge, U.K.)* 131:703–712.
17. Allen J-M, *et al.* (2006) WARP is a novel multimeric component of the chondrocyte pericellular matrix that interacts with perlecan. *J Biol Chem* 281:7341–7349.
18. Kappler J, *et al.* (2000) Glycosaminoglycan-binding properties and secondary structure of the C-terminus of netrin-1. *Biochem Biophys Res Commun* 271:287–291.
19. Charron F, *et al.* (2003) The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance. *Cell* 113:11–23.
20. Nakai K, Horton P (1999) PSORT: A program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem Sci* 24:34–35.
21. Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14:378–379.
22. Pearson W-R (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 132:185–219.
23. Apweiler R, *et al.* (2000) InterPro-an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* 16:1145–1150.
24. Sugiura N, *et al.* (1993) Preparation of lipid-derivatized glycosaminoglycans to probe a regulatory function of the carbohydrate moieties of proteoglycans in cell-matrix interaction. *J Biol Chem* 268:15779–15787.