

# Global secretome analysis identifies novel mediators of bone metastasis

Mario Andres Blanco<sup>1</sup>, Gary LeRoy<sup>1</sup>, Zia Khan<sup>2,3,5</sup>, Maša Alečković<sup>1</sup>, Barry M Zee<sup>1</sup>, Benjamin A Garcia<sup>1</sup>, Yibin Kang<sup>1,4</sup>

<sup>1</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA; <sup>2</sup>Department of Computer Science, Princeton University, Princeton, NJ 08540, USA; <sup>3</sup>Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA; <sup>4</sup>Genomic Instability and Tumor Progression Program, Cancer Institute of New Jersey, New Brunswick, NJ 08903, USA

**Bone is the one of the most common sites of distant metastasis of solid tumors. Secreted proteins are known to influence pathological interactions between metastatic cancer cells and the bone stroma. To comprehensively profile secreted proteins associated with bone metastasis, we used quantitative and non-quantitative mass spectrometry to globally analyze the secretomes of nine cell lines of varying bone metastatic ability from multiple species and cancer types. By comparing the secretomes of parental cells and their bone metastatic derivatives, we identified the secreted proteins that were uniquely associated with bone metastasis in these cell lines. We then incorporated bioinformatic analyses of large clinical metastasis datasets to obtain a list of candidate novel bone metastasis proteins of several functional classes that were strongly associated with both clinical and experimental bone metastasis. Functional validation of selected proteins indicated that *in vivo* bone metastasis can be promoted by high expression of (1) the salivary cystatins CST1, CST2, and CST4; (2) the plasminogen activators PLAT and PLAU; or (3) the collagen functionality proteins PLOD2 and COL6A1. Overall, our study has uncovered several new secreted mediators of bone metastasis and therefore demonstrated that secretome analysis is a powerful method for identification of novel biomarkers and candidate therapeutic targets.**

**Keywords:** cancer; metastasis; proteomics; secretome; bone

*Cell Research* (2012) 22:1339-1355. doi:10.1038/cr.2012.89; published online 12 June 2012

## Introduction

Bone metastasis is a common and devastating step in the progression of breast, lung, prostate, bladder, thyroid, and other cancers. In particular, more than 70% of metastatic breast cancer patients suffer from bone metastasis and must endure debilitating bone fractures, severe pain, and hypercalcemia [1, 2]. While treatment for cancer-induced bone complications does exist and improves the quality of life, the efficacy is limited and often does

not extend the survival of patients with bone metastasis. Therefore, the discovery of novel biomarkers and potential new therapeutic targets for bone metastasis is highly desirable.

Osteolytic bone metastasis has been described as being driven by a “vicious cycle” of tumor-stroma interactions [2]. Upon reaching the bone microenvironment, cancer cells activate the differentiation of bone-degrading osteoclasts either directly or indirectly by promoting the increased production of osteoclastogenesis cytokines by osteoblasts [2]. Differentiated osteoclasts dissolve mineralized bone, leading to the release of a large compendium of growth factors and cytokines, such as PDGF, FGF, and IGFs and TGF- $\beta$  [2-4], from the bone matrix, which in turn promote the growth of and continued pro-osteolytic signaling by cancer cells [5]. Notably, communication between cancer cells and bone stromal cells is typically mediated via secreted proteins. Accordingly, the majority

Correspondence: Yibin Kang

Tel: +01-609-258-8834; Fax: +01-609-258-2340

E-mail: ykang@princeton.edu

<sup>5</sup>Current address: Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

Received 25 December 2012; revised 21 March 2012; accepted 12 April 2012; published online 12 June 2012

of current bone metastasis therapeutic targets in use or in development – such as TGF- $\beta$ , Cathepsin K, RANKL, and DKK1 – are secreted proteins.

While the importance of secreted proteins in bone metastasis has long been appreciated, only via recent technological advances in mass spectrometry have global, proteome-scale secreted protein profiling experiments become possible. Analyses of the complete set of secreted proteins – otherwise known as the “secretome” – have been reported in various organisms, cell types, and pathologies, and are quickly gaining popularity [6, 7]. Not surprisingly, several studies have focused on analyses of cancer- or metastasis-specific changes in secretome profiles [8–22]. However, none of the studies to date has focused on secretomes relevant to the bone metastasis phenotype. Moreover, the majority of cancer and metastasis secretome studies have focused on mass spectrometry-based identification and cataloguing of secretomes, and none has included experimental, *in vivo* validation of the functional importance of identified novel mediators of cancer or metastasis.

Here we present a comprehensive analysis of the bone metastasis secretome, integrating both non-quantitative and quantitative, SILAC-based mass spectrometry analyses of the secretomes of three families of cell lines with relevance to bone metastasis. Specifically, we profiled parental and highly or lowly bone metastatic derivatives of the MDA-MB-231 (human breast cancer), 4T1 (mouse breast cancer), and TSU (human bladder cancer) cell lines, so as to help elucidate which secreted proteins are universally required for bone metastasis and which may depend on contexts such as species or cancer type. Following the identification of secreted, novel candidate bone metastasis proteins that are highly overrepresented in aggressively bone metastatic cell lines, we then analyzed the functional role of several proteins spanning multiple functional classes in promoting bone metastasis *in vivo*. Our analysis has functionally validated the novel role of several groups of clinically-important secreted proteins in promoting bone metastasis, including Cathepsin inhibitors (CST1, CST2, and CST4), collagen functionality proteins (PLOD2 and COL6A1), and plasminogen activators (PLAT and PLAU). These proteins may become potential biomarkers for the detection of bone metastasis, as well as therapeutic targets for treatment of bone metastasis.

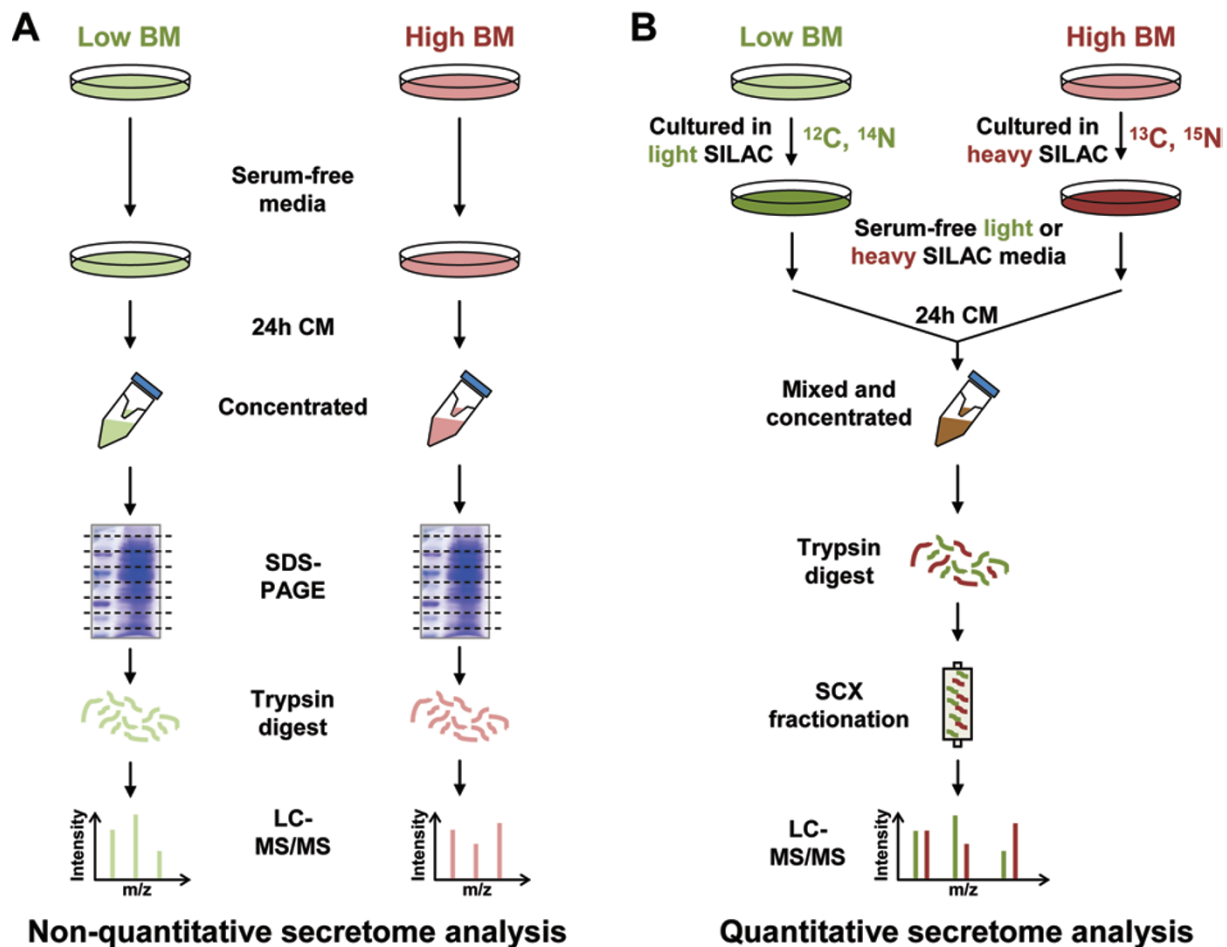
## Results

### *Non-quantitative bone metastasis secretome analysis*

For initial investigations into the bone metastasis secretome, we chose to profile secreted proteins in the

conditioned media (CM) from three families of cell lines of relevance to bone metastasis. Specifically, we analyzed (1) the MDA-MB-231 (human breast cancer)-derived sublines SCP4 and SCP6 (weakly bone metastatic), and SCP2 and 1833 (strongly bone metastatic) [23]; (2) the 4T1 (mouse breast cancer) family cell lines 4T1 and 4T1.2 (weakly and strongly bone metastatic, respectively) [24], and (3) the TSU-Pr1 (human bladder cancer) family of cell lines TSU-Pr1 and TSU-Pr1-B2 (weakly and strongly bone metastatic, respectively) [25]. Thus, our initial secretomic analysis spanned eight cell lines from two species and two cancer types, allowing for cross-species and cross-cancer type comparisons of secreted proteins from matched weakly and strongly bone metastatic derivative cell lines. To comprehensively determine the secreted proteins found in these eight cell lines, we harvested conditioned serum-free media from each cell line, separated the proteins via SDS-PAGE, and subjected in-gel trypsin-digested peptides to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis (Figure 1A). Peptides were searched against the Sequest database and 1 458 total unique proteins were identified across the eight cell lines, with the number of proteins identified per cell line varying from 103 to 722. Many proteins identified by this method are expected to be intracellular contaminants derived from the basal rates of apoptotic cells present in normal culturing conditions. To filter out likely intracellular contaminants and retain proteins most likely to be secreted by live cells, we used the SignalP [26, 27] and Gene Ontology (GO) [28, 29] databases to retain only proteins either experimentally observed to be extracellular or otherwise containing predicted cleavable signal peptides targeting them to the secretory pathway. This analysis yielded a total of 298 unique, secretory-predicted proteins across the eight cell lines, which is in line with previous quantities and percentages of secreted proteins found in comparable secretome analyses [8, 30–40].

To determine the secreted proteins most likely to be essential for bone metastasis, we looked at which of these proteins, within a given cell family, were identified in conditioned media from the bone metastatic derivative cell line(s) and not identified in the corresponding weakly bone-metastatic subline(s) (Figure 2A). These three “bone metastasis secretome signatures” (BMSSs) consisted of 33, 28, and 27 proteins from the MDA-MB-231 (hereafter MDA231), TSU-Pr1 (hereafter TSU), and 4T1 cell line families (Supplementary information, Tables S1, S2, and S3). Manual inspection of these signatures revealed several proteins known to be of high importance in bone metastasis and bone biology, such as MMP1 and CTGF (MDA231 BMSS), OGN (4T1 BMSS) and DKK3



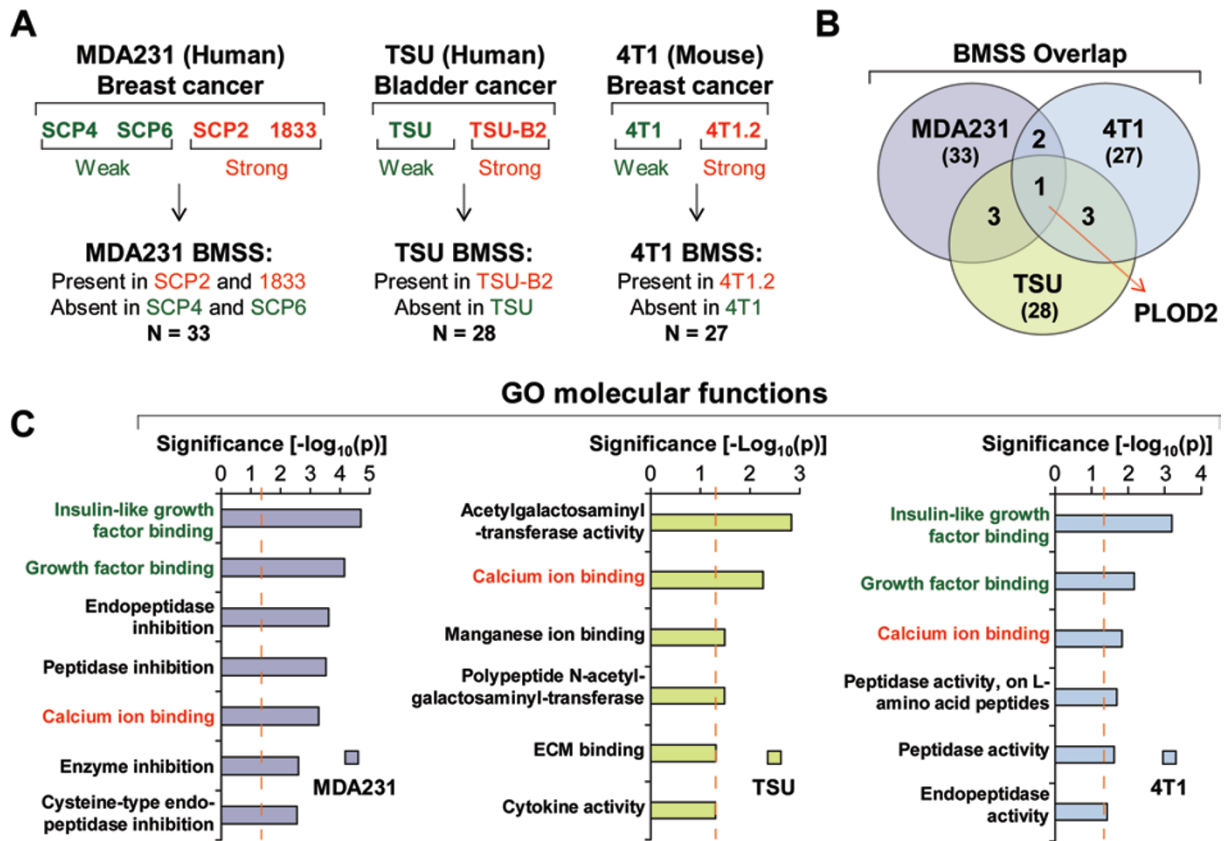
**Figure 1** Schematic overview of (A) non-quantitative and (B) quantitative secretome analysis approaches.

(TSU BMSS). For unbiased initial validation of the BMSSs, we performed GO analysis. GO cellular component analysis indicated that, in each BMSS, the most significantly overrepresented localization was “extracellular region” ( $P = 9.1E-12$ ,  $7.9E-7$ , and  $1.4E-7$  for MDA231, TSU, and 4T1 BMSSs, respectively), confirming the efficacy of the experimental and bioinformatic approaches employed to identify secreted proteins (Supplementary information, Figure S1). Furthermore, GO molecular function (MF) analysis confirmed that ontologies pertinent to tumorigenesis and metastasis, such as “growth factor binding” and, more specifically, “insulin-like growth factor binding” were enriched in multiple BMSSs (Figure 2C and Supplementary information, Table S7). Interestingly, more than half (4 of 7) of the significantly overrepresented MDA231 BMSS MFs were related to peptidase/enzyme inhibition. Notably, one MF ontology was significantly overrepresented in all three BMSSs – “calcium ion binding” (28.1% ( $P = 5.3E-4$ ), 25% ( $P = 0.006$ ), and 22.2% ( $P = 0.013$ ) of MDA231, TSU, and

4T1 BMSS proteins, respectively), which suggests the involvement of BMSS proteins in bone metastasis and bone biology (Figure 2C and Supplementary information, Table S7). While significant overlap was observed between the BMSSs in terms of categories of overrepresented proteins, especially between the two breast cancer (MDA231 and 4T1) BMSSs, less overlap was observed for individual proteins. However, one protein was present in all three BMSSs – Procollagen-lysine 2-oxoglutarate 5-dioxygenase 2 (PLOD2) (Figure 2B). PLOD2 catalyzes the hydroxylation of lysine residues on collagen-like peptides, enhancing their crosslinking potential. Notably, *PLOD2* mutations are associated with osteogenesis imperfecta (bone fragility) in humans [41].

#### *Quantitative, SILAC-based bone metastasis secretome analysis*

To complement our non-quantitative secretome analysis, we next extended our investigation to SILAC-based, quantitative proteomic approaches. Here we performed



**Figure 2** Non-quantitative secretome analysis. **(A)** Overview of cell lines used for the three bone metastasis secretome signatures (BMSS). **(B)** Overlap of non-quantitative BMSSs. **(C)** Non-quantitative BMSSs were searched via the Gene Ontology (GO) database for significantly overrepresented Molecular Functions (MFs). The dashed line indicates significance cutoff ( $P = 0.05$ ); green text indicates MFs significantly overrepresented in two of three BMSSs, and red text indicates overrepresentation in all three BMSSs.

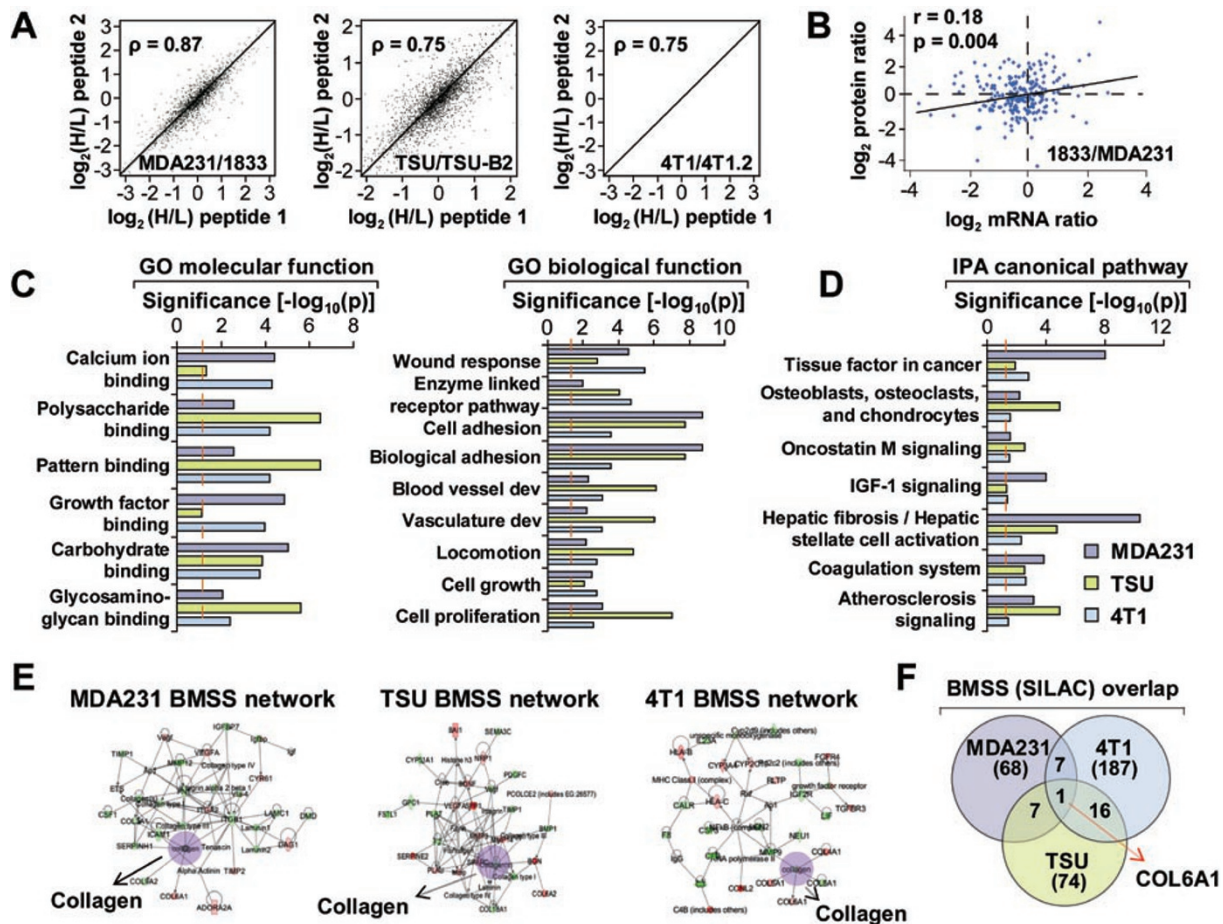
three pair-wise secretome comparisons: MDA231 (parental) vs. 1833, TSU vs. TSU-B2, and 4T1 vs. 4T1.2. As we only included one highly and one lowly bone metastatic MDA231 family cell line in this case, we chose to use the heterogeneous parental MDA231 cell line, rather than SCP4 or SCP6, as the representative weakly bone metastatic cell line so as to avoid biases associated with using a single clonally-derived cell line. We employed a different methodology for SILAC-based proteomics (Figure 1B), including trypsinization prior to fractionation and fractionation of tryptic peptides via strong cation exchange (SCX) chromatography. Here, we observed increases in total number of proteins identified (ranging from 2 013 to 3 426 per cell line), with the percentage of secretory-predicted proteins being roughly the same (384 to 742 per cell line). To examine the quality of SILAC peptide identification and quantification, we correlated  $\log_2$  SILAC ratios, heavy (H) vs. light (L), observed from independent measurements of two distinct tryptic

peptides, peptide 1 and peptide 2, from the same protein. In general, these ratios were highly correlated, indicating accurate protein expression measurements were derived from our analysis of the mass spectra collected by LC-MS/MS (Figure 3A). Quantitative proteomics allows for the comparison of gene expression to protein abundance. While overall correlation is expected between RNA and protein levels, the secreted protein fraction in particular is expected to have a significant degree of regulation at the post-transcriptional level, due to the possibility of differential protein expression, processing and secretion. Accordingly, while we observed a significant positive correlation between secreted protein abundance and RNA expression in MDA231 and 1833 cells ( $r = 0.18$ ,  $P = 0.004$ ; Figure 3B), clearly secreted protein abundance and RNA expression often do not correspond, underscoring the importance of proteomics-level investigations of the tumor cell secretome.

To generate BMSSs based on quantitative protein

abundance data, we considered secretory-predicted proteins upregulated at least 1.5-fold in bone-metastatic variants as compared to their matched, parental/non-bone metastatic cell lines. This analysis yielded quantitative BMSSs consisting of 68, 74, and 187 proteins for the MDA231, TSU, and 4T1 family comparisons, respectively (Supplementary information, Tables S4, S5, and S6). As with the non-quantitative secretome approach, BMSS overlap between the families of cell lines was large in terms of overrepresented pathways and categories of proteins. Again, “extracellular region” was the most highly overrepresented localization (Supplementary information, Figure S1B) and “calcium ion binding” proteins were significantly enriched in all three quantitative BMSSs (Figure 3C and Supplementary information,

Table S8). Other cancer/metastasis-related GO MFs or biological processes (BPs) enriched in all three quantitative BMSSs included “wound response,” “cell adhesion,” “cell proliferation,” and “growth factor binding” (Figure 3C and Supplementary information, Tables S8 and S9). With quantitative, fold-change BMSS data, we were able to further investigate and compare globally enriched BMSS pathways via Ingenuity Pathway Analysis (IPA). IPA also confirmed the biological relevance of all three quantitative BMSSs, with canonical pathways such as “osteoblasts, osteoclasts, and chondrocytes,” “IGF-1 signaling,” “tissue factor in cancer,” and “Oncostatin M,” each found to be significantly enriched in all three BMSSs (Figure 3D and Supplementary information, Table S10). BMSS IPA networks were found to be densely



**Figure 3** Quantitative secretome analysis. **(A)**  $\log_2$  SILAC heavy (H) vs. light (L) ratios were compared for two distinct tryptic peptides, peptide 1 and peptide 2, from the same protein. **(B)**  $\log_2$  secreted protein and mRNA abundance ratios were compared for 1833 and MDA-231 cells. **(C-E)** MDA231, TSU, and 4T1 BMSSs were subjected to GO database searching **(C)** and Ingenuity Pathway Analysis (IPA) **(D-E)**. The dashed lines indicate significance cutoffs ( $P = 0.05$ ) for GO Molecular and Biological functions and IPA Canonical Pathways. The most significant IPA networks are shown for each BMSS, with the “collagen” connectivity node being shaded in each case **(E)**. **(F)** Overlap of quantitative BMSSs.

interconnected, with each containing a “collagen” node consisting of 7, 9, and 5 connections in the MDA231, TSU, and 4T1 quantitative BMSSs, respectively (Figure 3E). Accordingly, one protein was present on all three BMSSs – Collagen 6A1 (COL6A1) (Figure 3F). Taken together, GO enrichment analyses and IPA strongly suggest that the three quantitative BMSSs do indeed consist of a high percentage of secreted proteins involved in cancer generally and bone metastasis/bone biology specifically.

#### *Validation of BMSSs in experimental and clinical bone metastasis datasets*

To further validate the BMSSs, we investigated to what extent transcriptomic overexpression of genes encoding BMSS proteins was observed in experimental and clinical bone metastasis microarray datasets. Despite the only moderate correlation between secreted protein abundance and mRNA expression, we reasoned that Gene Set Enrichment Analysis (GSEA) [42, 43] would potentially be robust enough to indicate whether BMSS gene sets are globally upregulated in bone metastasis microarray data, as GSEA is sensitive for detecting overall regulatory patterns while allowing moderate amounts of noise or even discordance in the analysis. For the first GSEA, we investigated whether the MDA231 BMSSs (both quantitative and non-quantitative) were enriched in microarray data from the Kang *et al.* study [23] comparing *in vivo*-derived highly and lowly bone metastatic sublines of MDA231 cells. Indeed, both MDA231 BMSSs were found to be significantly enriched in the list of genes ranked according to their degree of overexpression in highly vs. lowly bone metastatic cell lines (Figure 4A). We next expanded our GSEA validation of the BMSSs to clinical breast cancer patient datasets. Here we interrogated the MSK82 dataset [44], which contains microarray data from primary tumors of breast cancer patients with known metastatic outcome, including site of metastasis. We tested all BMSSs for enrichment in the list of genes ranked by their degree of overexpression in patients who suffered bone metastasis as compared to those who remained disease-free. Despite the small sample size of the bone metastatic patient group ( $n = 14$  for bone metastasis vs.  $n = 55$  for no metastasis) and the inherent noise in clinical primary tumor microarrays, all six BMSSs (three quantitative and three non-quantitative) were enriched, to varying degrees of significance, in the bone metastatic patient phenotype (Figure 4B and data not shown). Focusing on the quantitative BMSSs, as they provide larger gene sets more suitable in size for GSEA, enrichment of both breast cancer BMSSs was highly significant ( $P = 0.009$  and  $P < 0.001$  for MDA231

and 4T1 BMSSs, respectively), whereas the TSU (bladder cancer) BMSS showed a trend toward enrichment in the bone metastasis patient phenotype that fell short of statistical significance ( $P = 0.147$ , Figure 4B). This latter result was not unexpected given the cancer type difference. Thus, in line with and adding to IPA and GO analyses, experimental and clinical dataset GSEAs have globally validated the biological and clinical relevance of the BMSSs to the bone metastasis phenotype.

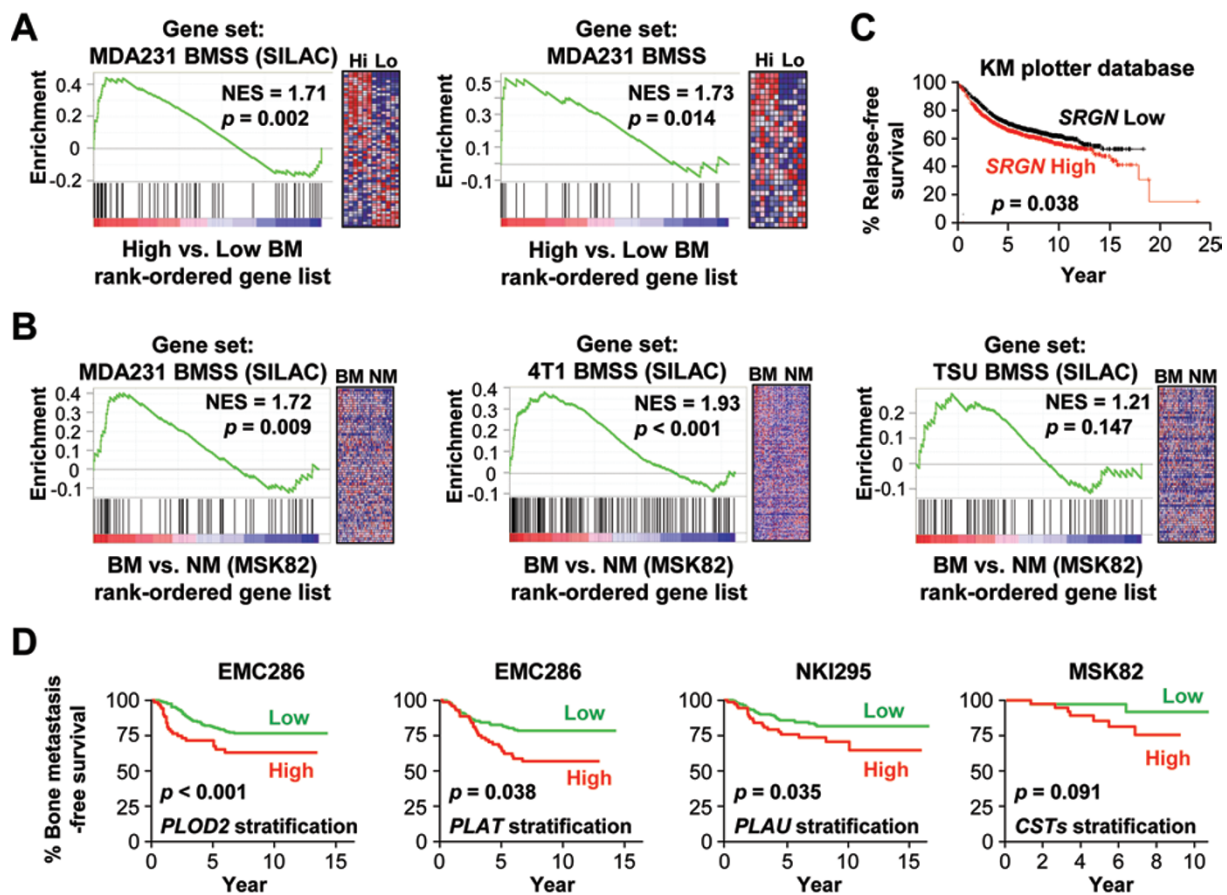
#### *Functional analysis of selected secreted proteins in mediating experimental bone metastasis*

Using our mass spectrometry-based secretome data, we next aimed to identify novel secreted mediators of bone metastasis via functional, *in vivo* experimental bone metastasis approaches. While we had a large list of candidate bone metastasis proteins, we focused on functionally intriguing proteins with gene expression patterns associated with either clinical bone metastasis or general metastasis. These candidate proteins spanned several functional classes. First, we considered the collagen functionality proteins PLOD2 and COL6A1, as they were the sole proteins on all the BMSSs from the non-quantitative and quantitative analyses, respectively. We also considered the plasminogen activators PLAT and PLAU, as each quantitative BMSS contained either PLAT or PLAU. The salivary cystatin family of proteins, CST1, CST2, and CST4, were also selected as they were strongly represented on both the non-quantitative (CST1, CST2, and CST4) and quantitative (CST1 and CST4) MDA231 BMSSs. Finally, we considered the secretory granule scaffold protein SRGN, as it was on both MDA231 BMSSs, showing a ~20-fold upregulation in 1833 as compared to parental (MDA231) cells, and is universally upregulated at the RNA level in bone metastatic variants of the MDA231 family, thus residing on the previously reported 43-gene pro-bone metastasis signature [23]. Notably, to our knowledge, no studies to date have evaluated the role of any of these proteins in breast or bladder to bone metastasis. Cox proportional hazards models indicated that expression of all selected genes except *SRGN* was significantly or near-significantly associated with either metastasis generally (*COL6A1* and *CST2*) or bone metastasis specifically (all others) in one of three large clinical breast cancer microarray datasets (Table 1). Kaplan-Meier plots indicated that patients expressing high levels *SRGN* showed significantly shorter relapse-free survivals than those expressing low *SRGN* (Figure 4C), and patients expressing high levels of *PLOD2*, *PLAT*, *PLAU*, or *CSTs* (averaged) had significantly or near-significantly shorter periods of bone metastasis-free survival than their lowly-expressing

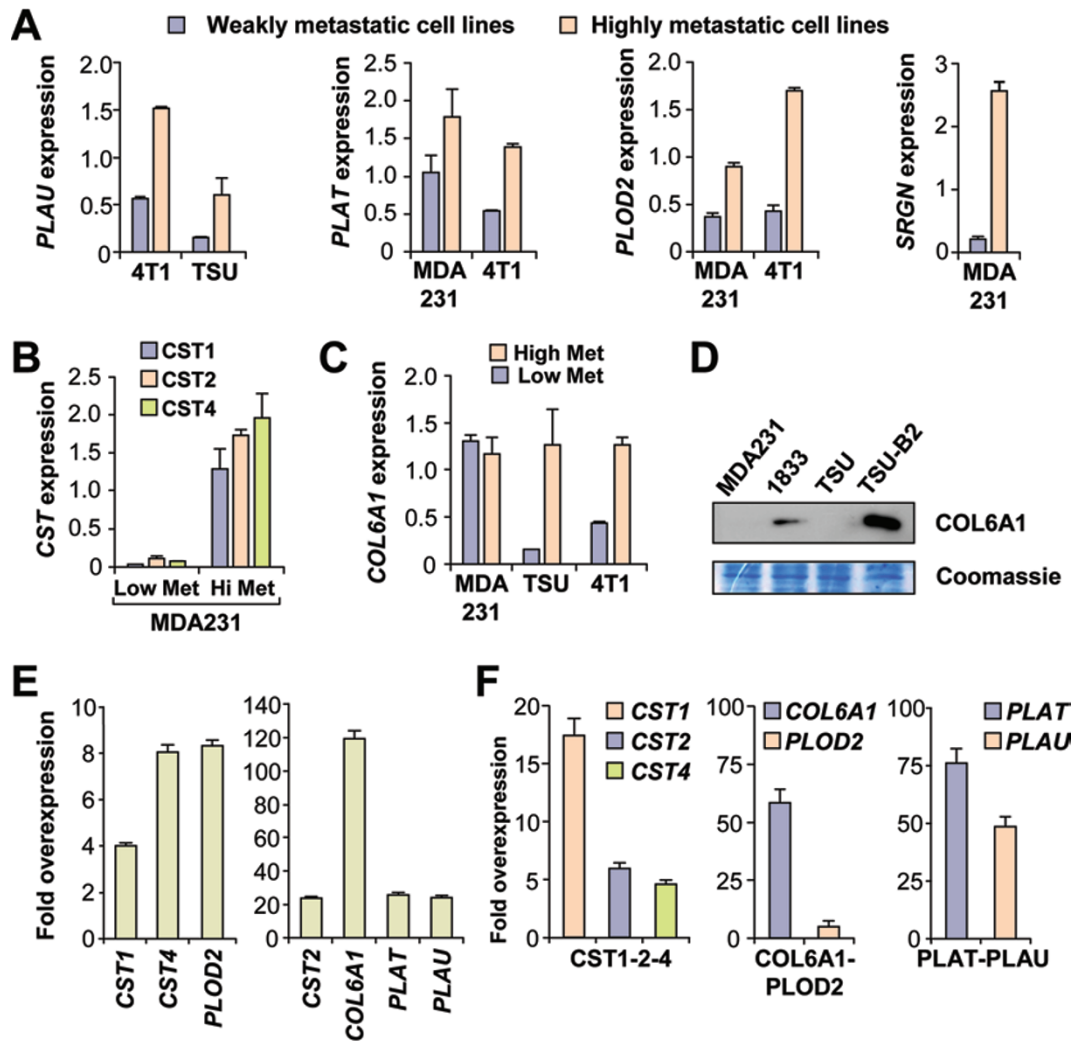
**Table 1** Genes encoding BMSS proteins are associated with bone metastasis and general metastasis

Gene	Dataset	Association	HR	95% CI	P-value
<i>CST1</i>	MSK82	Bone met	1.56	1.00-2.41	0.045
<i>CST2</i>	NKI295	Metastasis	11.63	2.65-51.04	0.001
<i>CST4</i>	MSK82	Bone met	2.13	0.92-4.91	0.076
<i>CST1/2/4 (AVG)</i>	MSK82	Bone met	2.09	1.02-4.28	0.043
<i>COL6A1</i>	MSK82	Metastasis	1.73	0.94-3.19	0.079
<i>PLOD2</i>	EMC286	Bone met	1.44	1.11-1.89	0.006
<i>PLAT</i>	EMC286	Bone met	1.32	1.12-1.56	0.001
<i>PLAU</i>	NKI295	Bone met	3.31	1.00-10.97	0.051

Cox proportional hazards ratios are shown for the association of the expression of selected BMSS genes with either general metastasis-free survival (*CST2* and *COL6A1*) or bone metastasis-free survival (all other genes) in either the MSK82, EMC286 or NKI295 clinical breast cancer datasets.



**Figure 4** Global and specific validation of BMSS proteins in experimental and clinical datasets. **(A)** Genes encoding proteins on the quantitative (“SILAC”) (left) and non-quantitative (right) MDA231 BMSSs were used as gene sets and tested for enrichment via Gene Set Enrichment Analysis (GSEA) in the list of genes ranked by degree of overexpression in 1833 as compared to MDA231 cell line microarray data. The panel to the right of each GSEA plot shows the corresponding heat map of the differential expression of BMSS genes sets in the 1833 and MDA231 microarray datasets. **(B)** The three quantitative (“SILAC”) BMSS gene sets were tested for enrichment in the list of genes ranked by degree of overexpression in primary breast cancer tumors from patients in the MSK82 dataset who developed bone metastasis as compared to patients who remained disease free. **(C)** Kaplan-Meier plot of relapse-free survival of breast cancer patients stratified by median primary tumor SRGN expression in the KM-Plotter database. Log-rank test P-value is displayed. **(D)** Kaplan-Meier plots of bone metastasis-free survival of breast cancer patients stratified by upper quartile primary tumor expression of the indicated genes in either the EMC286, NKI295, or MSK82 datasets, as indicated. Log-rank test P-values are displayed.



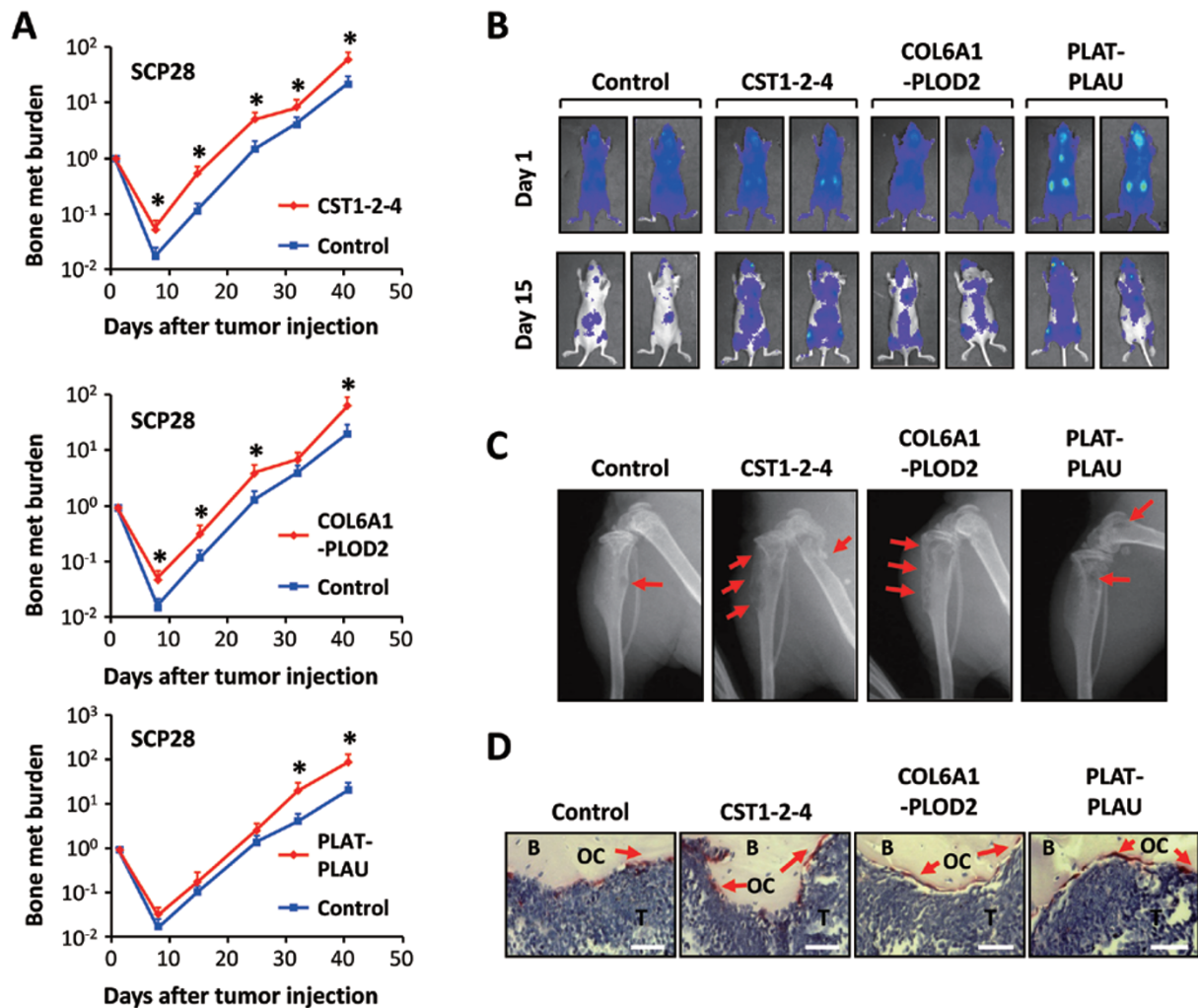
**Figure 5** Endogenous and exogenous expression of BMSS genes selected for experimental analysis. **(A-C)** Endogenous mRNA expression of all genes selected for functional analysis in selected MDA231, TSU, and 4T1 family cell lines. **(D)** Endogenous protein expression of COL6A1 in conditioned media from MDA231 and TSU family cell lines. **(E)** SCP28 cells were engineered to stably overexpress the indicated genes by retroviral infection followed by drug selection. Fold overexpression of the indicated genes relative to vector (control) cells was quantified via qRT-PCR. **(F)** The indicated combinations of high titer retrovirus were mixed and used to infect SCP28 cells. After roughly one month of culture without drug selection, RNA was harvested and fold overexpression of the indicated genes relative to vector (control) cells was quantified via qRT-PCR.

counterparts (Figure 4D). Finally, before initiating experimental evaluation of the role of these candidates in bone metastasis, we confirmed the upregulation of each gene in the relevant cell lines. Due to the large number of candidate genes overall, we attempted confirmation via qPCR first, proceeding to western blot confirmation only if upregulation in bone metastatic variant cell lines was not observed at the RNA level. Similar to mass spectrometry data, we observed upregulation of either *PLAU* or *PLAT* (or both) in MDA231, 4T1, or TSU bone metastatic variant cell lines (Figure 5A). Elevated ex-

pression of *PLOD2* was observed in MDA231 and 4T1 bone metastatic variants, and *SRGN* (Figure 5A) as well as all three cystatins displayed strong MDA231-specific bone metastatic variant patterns of upregulation (Figure 5B). Finally, *COL6A1* was found to be upregulated at the RNA level in 4T1.2 and TSU-B2 cells, as compared to 4T1 and TSU cells, and at the protein level in 1833 as compared to parental (MDA231) cells (Figures 5C and 5D).

For our initial functional analyses, we focused on the collagen functionality proteins *PLOD2* and *COL6A1*, as





**Figure 6** Salivary cystatins, COL6A1 and PLOD2, and plasminogen activators promote bone metastasis *in vivo*. **(A)** Control and CST1-2-4 (top), COL6A1-PLOD2 (middle) and PLAT-PLAU (bottom) overexpressing SCP28 cells were xenografted into nude mice for experimental bone metastasis assays. Metastatic burden of hindlimb bone metastasis was quantified by weekly bioluminescent imaging (BLI). \* $P < 0.05$ . **(B)** BLI, **(C)** radiographic, and **(D)** TRAP staining (for mature osteoclasts) images are shown for representative mice from each experimental group. T = tumor, B = bone, OC = osteoclast. Scale bar, 50  $\mu\text{m}$  in **D**.

they were found on all three BMSSs, and on the secretory scaffold protein SRGN, as it was a highly-intriguing candidate specific to but robust in the MDA231 BMSSs. To test the importance of *COL6A1*, *PLOD2*, and *SRGN* in bone metastasis, we stably knocked down each gene in SCP2 cells, as aggressively bone metastatic variants of the MDA231 family of cells are the most suitable for bone metastasis loss-of-function experiments. We achieved roughly 95% efficiency of *SRGN* knockdown (KD), whereas KDs of *COL6A1* and *PLOD2* were effective (> 80% silencing), but did not completely ablate gene expression (Supplementary information, Figure S2A). We then used these cell lines, as well as vector

control cells, for experimental bone metastasis assays. Following intracardiac injections into nude mice, we quantified bone metastasis progression over time via *in vivo* bioluminescent imaging (BLI) over the course of six weeks. However, we did not observe significant changes in bone metastasis burden via BLI, nor did we observe significant differences in the number or extent of bone lesions via X-ray analysis for any of these genes as compared to control cells (Supplementary information, Figure S2C-S2E). Interestingly, while knockdown of these genes, individually, did not significantly affect bone metastasis progression, double knockdown of *SRGN* and *PLOD2* (Supplementary information, Figure S2B) did

lead to modest, but significant, reductions in bone metastatic burden (Supplementary information, Figure S2C-S2E).

Extending the reasoning that manipulation of single secreted proteins may not frequently affect bone metastasis progression, we focused our analysis instead on combinatorial manipulation of multiple proteins of the same functional class on our list of candidates. Thus, we sought to investigate the combinatorial role of (1) *PLOD2* and *COL6A1*, (2) *CST1*, *CST2*, and *CST4*, and (3) *PLAT* and *PLAU* in bone metastasis. Here we chose overexpression rather than knockdown due to the option of stably manipulating the expression of multiple genes simultaneously with relative ease. Furthermore, we suspected significant functional redundancy among proteins in the bone metastasis secretome, and gain of function (i.e. overexpression) phenotypes are less likely to be affected by redundancy than loss of function (i.e. knockdown) phenotypes are. Accordingly, stable, retrovirus-based overexpression of each of the seven selected genes was performed, individually at first, and anywhere between moderate (approximately four-fold) and extremely high (> 100-fold) stable overexpression was achieved for each gene in SCP28 cells (Figure 5E). Next, we pooled the appropriate combinations of high-titer retrovirus and combinatorially infected SCP28 cells to make stable cell lines overexpressing either (1) *PLOD2* and *COL6A1*, (2) *CST1*, *CST2*, and *CST4*, (3) *PLAT* and *PLAU*, or (4) vector (control). While we were not able to drug select these lines for overexpression of individual genes, as all constructs had the same resistance marker gene, we infected with very high titer, cultured the cells *in vitro* for approximately one month (approaching the duration of an *in vivo* experimental bone metastasis experiment), and then tested for maintenance of overexpression. We found that high expression of all constructs (comparable to levels observed in individually drug-selected overexpression cell lines) was indeed maintained (Figure 5F). We then used these three combinatorial overexpression cell lines, as well as control cells, for intracardiac injection of nude mice followed by quantification of bone metastatic burden via BLI over six weeks. In stark contrast to the results with manipulation of individual genes, here we observed that overexpression of each combination of secreted proteins led to significant increases in bone metastasis as compared to the control cell line (Figure 6A and 6B), thus confirming these BMSS proteins as novel mediators of bone metastasis. The strongest bone metastasis phenotype was observed in the SCP28-CST1-CST2-CST4 cell line, with a nearly five-fold increase in bone metastatic burden being observed at day 15, and no less than a two-fold increase being observed at any

other time points. Interestingly, the kinetics of bone metastasis progression varied across these cell lines. The SCP28-COL6A1-PLOD2 and SCP28-CST1-CST2-CST4 overexpressing cell lines had marked increases in bone metastasis at the first time point, and these increases were for the most part maintained over the course of the experiment. In contrast, the SCP28-PLAT-PLAU cell line behaved similarly to control cells until the later stages of the experiment, at which point outgrowth of SCP28-PLAT-PLAU bone metastatic lesions proceeded more rapidly than the control cells. All cell lines overexpressing candidate bone metastasis proteins showed an increased number and severity of bone metastatic lesions by X-ray as compared to control cells (Figure 6C). For histological analyses, we performed TRAP (tartrate-resistant acid phosphatase) staining to identify metastatic lesion-associated differentiated osteoclasts, as osteolytic metastases often promote osteoclast activity in the bone tissue. In all cases we observed large bone macrometastases with many TRAP-positive, differentiated osteoclasts in close proximity, although no statistically significant differences in TRAP-positive cell numbers were observed between the different experimental groups (Figure 6D). In conclusion, these experiments have validated the role of the salivary cystatins *CST1*, *CST2*, and *CST4*, the collagen functionality proteins *PLOD2* and *COL6A1*, and plasminogen activators *PLAT* and *PLAU* in promoting bone metastasis *in vivo*.

## Discussion

Proteome-level methodology provides significant advantages over more traditional genomic and transcriptomic analyses. While it is possible to approximate the abundance of proteins in the secreted fraction from microarray data [45], extrapolating proteomic data from transcriptomic data is problematic due to the prevalence of post transcriptional regulation, especially in the context of secreted proteins. In particular, our study has demonstrated that, in the secreted protein fraction, quantitative protein and mRNA measurements only loosely correlate, thus highlighting the importance of protein-level investigations. Therefore, with the maturation of proteomics technologies, secretome analysis has emerged as an attractive method for discovery of novel biomarkers and potential therapeutic targets.

While the earliest secretome analyses were performed in bacteria [46], investigations into the mammalian secretome have also become prevalent. The first cancer secretome analysis identified 145 proteins that were differentially expressed between tumorigenic and non-tumorigenic pancreatic epithelial cells, and the first

metastasis secretome study found 27 secreted proteins to be differentially expressed between metastatic and non-metastatic variants of the MDA-MB-435 melanoma cancer cell line [47]. The first breast cancer metastasis secretome study profiled the secretomes of the MCF10A metastasis progression series, finding proteins such as alpha-1-antichymotrypsin and galectin-3-binding protein to be upregulated in conditioned media from the metastatic variants [48]. More recently, many secretome analyses of cell lines representing several different cancer types have been reported [8, 30-40]. Most secretome studies have used methodology similar to ours in terms of initial protein harvesting conditions and then have typically used either concentrating columns or total protein precipitation methods for secreted protein isolation. The most common downstream fractionation methods include either 1D or 2D gel electrophoresis of proteins or SCX of peptides. Ultimately, as also found in our analyses, cell line-based secretome approaches commonly report identification of between one to several hundred proteins in conditioned media, with between 20 and 40 percent predicted to be secreted.

All secretome studies have faced the challenge of implementing bioinformatics-based filters to remove from their analyses all non-secreted proteins (from ruptured, apoptotic cells) identified in the conditioned media of cell lines. Such methodology typically involves interrogating either primary sequence-based secretory pathway prediction algorithms, or curated, empirical subcellular localization databases (or a combination of both approaches). Neither method is fully accurate – targeting to the secretory pathway may not in fact lead to secretion, and, conversely, lack of empirical observation of extracellular localization does not preclude secretion. Here we considered proteins to be secreted if they were either predicted by sequence or observed by observation to be secreted. Thus, we aimed to reduce false negatives, as we had downstream biological and clinical filters to further enrich for proteins of high interest.

Mass spectrometry-based proteomics studies also face the question of whether and how to quantify protein abundance. Here we have performed both label-free, non-quantitative analyses as well as quantitative SILAC-based analyses. The former approach was utilized for several secretomic analyses due to its high time and cost efficiency; the latter approach was used for three selected high interest comparisons. Non-quantitative methodology provides a natural, simple cutoff – the limit of detection – for determining differential abundance of a protein between samples of interest. However, this relatively arbitrary cutoff can give rise to both false positives and negatives. Quantitative, SILAC-based proteomic ap-

proaches do not feature this arbitrary cutoff, and will thus determine a larger number of differentially expressed proteins (depending on the fold-change threshold used for significance). In both analyses, however, only relatively abundant proteins will be identified when using standard fractionation approaches. Interestingly, in our hands, SCX fractionation of peptides (i.e. MudPIT [49]), as used for SILAC-based analyses, allowed for identification of greater numbers of proteins than did fractionation of proteins via 1D gel electrophoresis.

In this study, we have presented what is, to our knowledge, the first analysis of the bone metastasis secretome. Using both quantitative and non-quantitative proteomics approaches, we have globally profiled the secreted protein fraction from a total of nine cell lines of relevance to bone metastasis spanning two species (mouse and human) and two cancer types (breast and bladder). In general, significant overlap at the pathway level was observed between the BMSSs from the three different cell line families profiled. Pathways and proteins with MFs related to bone biology and bone metastasis were overrepresented in all BMSSs, and global enrichment of the BMSSs in clinical bone metastasis datasets was observed. Thus, we conclude that the identified BMSSs are each appropriately reflecting protein-level changes pertinent to bone metastasis in the secreted fraction. We further observed, however, less overlap among the BMSSs at the level of individual proteins. We therefore hypothesize that general secreted protein pathways and mechanisms underlying bone metastasis are conserved across mouse and humans and breast and bladder cancer, but that the specific proteins mediating this process do vary significantly by species and cancer type.

In choosing to evaluate candidate novel bone metastasis proteins, we first considered clinically relevant proteins that were consistently upregulated in cross species and/or cross cancer BMSSs. However, we also considered the most intriguing candidates that were specific to the MDA231 (human breast cancer) BMSS, as our focus is on human disease and as the MDA231 organotropic metastatic derivatives are the most widely used and well-validated system for functional bone metastasis investigations. Our initial finding, that genetic alteration of a single secreted protein was typically insufficient to significantly affect bone metastatic progression, was not unexpected, as bone metastasis is considered to be a multigenic phenotype [23]. Thus, we chose to focus our analysis on classes of functionally related, clinically relevant proteins identified in the three BMSSs. Specifically, we have shown that combinatorial overexpression of the following groups of proteins significantly promoted experimental bone metastasis progression,

thus confirming them as novel secreted mediators of bone metastasis: (1) two collagen functionality proteins (PLOD2 and COL6A1); (2) two plasminogen activators (PLAT and PLAU); and (3) three salivary cystatin family members (CST1, CST2, and CST4). We note that there is considerable evidence that the observed co-overexpression of these proteins in cell line-based bone metastasis models is conserved in the clinical context. For example, the three *Cystatin* genes investigated have very tightly correlated gene expression patterns in MSK82 clinical samples [44], ranging from  $r = 0.61$  ( $P = 7.5 \times 10^{-10}$ ) to  $r = 0.74$  ( $P = 1.3 \times 10^{-15}$ ) which may in part be due to the fact that these genes arose from duplication events and share highly similar promoter regions. In this dataset, *COL6A1* and *PLOD2* also show evidence of co-overexpression, as the Pearson correlation coefficient of their gene expression patterns is  $r = 0.21$  ( $P = 0.055$ ). The expression patterns of *PLAT* and *PLAU* are also positively correlated in this dataset, although the data do not reach statistical significance. However, investigating the larger NKI295 dataset [50] ( $n = 295$  patients), the Pearson correlation coefficient of *PLAT* and *PLAU* gene expression patterns is statistically significant, at  $r = 0.13$  ( $P = 0.028$ ). Thus, the synergistic pro-metastatic effect of BMSS protein combinations that we observed in our mouse model studies may also occur in clinical bone metastasis.

Of considerable biological and therapeutic interest now are the mechanisms by which these secreted proteins facilitate bone metastasis. Collagen is crucial to the bone matrix, but the most common bone collagen is Collagen type 1, with far less being known about COL6A1, though its mutations are linked to muscular-skeletal defects such as Bethlam myopathy and Ulrich Congenital Muscular Dystrophy [51-54]. Interestingly, a recent secretome analysis identified COL6A1 as a strongly upregulated protein in the conditioned media of metastatic, as compared to non-metastatic lung cancer cell lines [55]. PLOD2 hydroxylates collagen-like peptides, leading to their increased crosslinking potential. Little mechanistic work has been reported on the effect of PLOD2-based crosslinking in bone tissue; however, mutations in *PLOD2* are the cause of Bruck syndrome, a congenital osteogenesis-imperfecta-like skeletal disorder caused by insufficient/improper collagen crosslinking [41]. Notably, high extracellular matrix (ECM) tension and stiffness, which is affected by the amount and type of crosslinked collagen polymers, has recently been shown to have a powerful impact on metastasis [56] as well as the differentiation of mesenchymal stem cells into the osteoblast lineage [57]. Thus, it is possible that PLOD2 and COL6A1-overexpressing cancer cells could be influencing the local bone ECM so as to alter the normal bone

homeostasis, and to in turn foster a bone microenvironment hospitable for survival and progression of metastatic lesions. Interestingly, Lysyl Oxidase (LOX), a different collagen-modifying enzyme that also increases collagen crosslinking, has also been shown to influence metastatic progression. High expression of LOX – either by stromal cells [56] or by cancer cells [58] – dramatically promotes local and distant metastasis in breast cancer. Thus, it is possible that increased local ECM stiffness – whether achieved via PLOD2 or LOX – can be viewed as a general mechanism for promoting cancer cell invasion and colonization in multiple tissue microenvironments.

PLAT and PLAU are serine proteases which activate plasmin and allow it to degrade ECM components such as fibronectin and laminin, and activate zymogen-state MMPs. While PLAT and PLAU have been implicated in osteoclast-mediated bone resorption [59, 60], cancer cells are not thought to be able to directly resorb bone. Thus, whether tumor cell-derived PLAT/PLAU can enhance bone resorption indirectly remains to be seen. A further question is whether PLAT and PLAU function redundantly or have unique roles in promoting bone metastasis. The salivary cystatins – CST1, CST2, and CST4 – are perhaps the most surprising of the identified novel mediators of bone metastasis. Cystatins are cysteine protease inhibitors, with cystatin-C (CST3) potently inhibiting pro-metastatic Cathepsin B [61] and thus being considered to be a potential metastasis suppressor. The salivary cystatins, however, have not been shown to strongly antagonize Cathepsin B [62], and instead have been hypothesized to function in salivary glands as inhibitors of harmful proteinases expressed by pathogens [63]. Additionally, the salivary cystatins have been suggested to have high affinity for hydroxyapatite, the main mineralized component of bone tissue [64]. Thus, whether salivary cystatin production functions in bone metastasis via facilitating evasion of the immune system, or homing to the bone tissue, or by a different mechanism altogether, remains to be determined.

Finally, it is also of interest to determine the *in vivo* source of production of these secreted proteins of interest. Here we have profiled tumor cell line conditioned media *in vitro*, but we envision the possibility that, *in vivo*, production of pro-bone metastatic secreted proteins could be achieved by either tumor or tumor-associated stromal cells, as secreted proteins are accessible to all cells in proximity to the secretion source. Thus, it would be interesting to determine whether knockout mice lacking key secreted bone metastasis proteins would have reduced bone metastatic burdens in *in vivo* bone metastasis assays due to deficiencies in the stromally-produced protein of interest.

Overall, we have aimed to present an analysis that has not only identified novel mediators of bone metastasis, but also has implications for the treatment of clinical bone metastasis. In general, proteins identified by secretome studies have two main potential clinical applications: (1) biomarker identification, and (2) candidate therapeutic target identification. Here we have provided groundwork that can be used by future studies to determine whether several of the novel bone metastasis proteins identified herein can be further developed as novel biomarkers and therapeutic targets for bone metastasis.

## Materials and Methods

### *Cell culture and conditioned media preparation*

All cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% Fetal Bovine Serum, fungizone, and antibiotics unless otherwise noted. H29 cells were cultured with doxycycline when not actively producing virus. For quantitative SILAC labeling analyses, matched pairs of highly or lowly bone metastatic cell lines were cultured in DMEM without arginine or lysine supplemented with 10% dialyzed FBS (Invitrogen) and either normal or heavy arginine (13C6, 15N4; +10 Da) (Thermo Scientific Pierce) and lysine (13C6, 15N2; +8 Da) (Sigma). After 10 days of culture, cells were counted and equal numbers of cells were split to 15 cm dishes at roughly 50% confluence. Once cell lines reached ~70% confluence, two 15-cm dishes of each cell line were washed 3 times with PBS to remove serum proteins and incubated for 24 h in 15 ml of serum-free DMEM. Conditioned media was then removed, mixed (SILAC protocol only), filtered with 0.45  $\mu\text{m}$  filters, and maximally concentrated (75 to 100-fold) via room temperature centrifugation at 4 000 RPM in 3 kDa molecular weight cutoff concentrating columns (Millipore). Concentrated conditioned media samples were stored at  $-80^\circ$  until processing for mass spectrometry. For non-quantitative analyses, the procedure was carried out in the same fashion but with regular complete media and with samples processed separately, in parallel throughout (i.e. conditioned media were never mixed).

### *Sample preparation and mass spectrometry*

For non-quantitative mass spectrometry preparation and analysis, samples were prepared as previously described [65]. Briefly, concentrated conditioned media samples were run on a 10% SDS-polyacrylamide gels and cut into 10 gel slices per sample spanning approximately the entire molecular weight range. Solubilized gel slices were treated with dithiothreitol (DTT) and 55 mM iodoacetamide prior to in-gel digestion with sequencing grade trypsin at a 20:1 protein:enzyme ratio for 12 h at room temperature [66]. All peptide digests were desalted using homemade C18 STAGE tips [67] and diluted in 10  $\mu\text{l}$  of 0.1% acetic acid for MS analysis. For quantitative (SILAC-based) mass spectrometry preparation, samples were treated with DTT and iodoacetamide and subjected to in-solution trypsin digest overnight prior to desalting and subsequent SCX fractionation as described previously [68].

Mass spectrometry for non-quantitative samples was performed as previously described [65]. Briefly, nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed on

a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer (Thermo Electron, San Jose, CA) coupled to an Agilent 1200 Series binary HPLC pump (Agilent Technologies, Palo Alto, CA) and an Eksigent AS2 autosampler (Eksigent Technologies). Peptides samples (2  $\mu\text{l}$ ) were loaded via autosampler on 75- $\mu\text{m}$  inner diameter-fused silica capillary columns constructed with an integrated electrospray tip that was packed with C18 reversed phase resin (Magic C18, 5- $\mu\text{m}$  particles, 200- $\text{\AA}$  pore size; Michrom BioResources, Auburn, CA). Peptides were separated by reversed phase liquid chromatography using a gradient of Buffer A (0.1 M acetic acid) and Buffer B (70 acetonitrile in 0.1 M acetic acid) at a flow rate of ~200 nl/min for 70 min. The Orbitrap instrument was operated in data-dependent mode using a resolution setting of 30 000 to obtain a full MS spectrum followed by 10 MS/MS spectra obtained in the ion trap. Peptides selected for MS/MS interrogation were then placed on an exclusion list for 30 s to limit duplicate spectra. The MS scans were collected with an automatic gain control target value of  $5 \times 10^5$  and maximum injection time of 100 ms over a mass range of 300-1 650  $m/z$ . MS/MS scans were collected using an automatic gain control value of  $4 \times 10^4$  and a threshold energy of 35% for collision-activated dissociation.

For quantitative (SILAC-based) analyses, mass spectrometry was performed as previously described [68]. Briefly, capillary nano LC-MS/MS of each SCX fraction was performed using an LTQ-Orbitrap hybrid mass spectrometer interfaced with an Eksigent nano HPLC system. Samples were loaded onto a silica microcapillary C18 resin using an Eksigent autosampler. Samples were separated over a three-hour gradient run at a 300 nl/min flow rate. The data-dependent MS/MS scans were done using the top seven most abundant ions and collision-induced dissociation (CID) for fragmentation.

### *Mass spectrometry data analysis*

Non-quantitative mass spectrometry data were analyzed as previously described [65]. Briefly, all MS/MS spectra were processed through the Bioworks 2.0 program utilizing the Sequest algorithm. Parameters for MS/MS database searching included a precursor tolerance of 0.1 Da and fragment tolerance of 0.5 Da. Searches were performed using the appropriate enzyme (with up to 3 missed cleavages). A static modification of 57 Da for cysteine iodoacetamide treatment and a dynamic modification of 16 Da for methionine oxidation were used in the database searches. The Scaffold program (version Scaffold\_2.1.03, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted at greater than 95% probability as specified by the Peptide Prophet algorithm [69] or better than 0.01 peptide probability by the Sequest algorithm. False discovery rates were estimated to be 1% by searching a reverse database as previously described [70].

Quantitative (SILAC-based) mass spectrometry was performed as previously described [68, 71]. Briefly, Arg-10 and Lys-8 labeled peptides were quantified using area under extracted ion chromatograms (XICs). XICs were found and paired using the previously described methods [71]. The ratio of the areas under the paired XICs was reported as the ratio between heavy and light versions of peptides. MS/MS spectra were searched using IPI version 3.59 mouse protein sequence database for the 4T1 vs. 4T1.2 comparison and the IPI version 3.62 human protein sequence database for the MB-231 vs. 1833 and TSU vs. TSU-B2 comparisons. MS/MS

database search and quantification were conducted using a 10 ppm precursor mass window using publicly accessible and open source software PVIEW (<http://compbio.cs.princeton.edu/pview>). Up to one missed cleavage was allowed for database search. MS/MS spectra were assigned an amino acid sequence using a high confidence < 1% false discovery rate (FDR). For proteins quantified by multiple peptides, the median ratio of all of the peptides was assigned to the protein. Protein expression ratios were normalized using all detected XIC pairs so the median of their logarithm was zero, correcting for any unequal loading of light and heavy sample.

### Cloning

All genes to be overexpressed were PCR amplified from cDNA and inserted into the pRVPTOpuro retroviral plasmid using either 5'*Bam*HI and 3' *Eco*RI (for *COL6A1*, and *SRGN*), 5' *Hind*III and 3' NotI (for *PLOD2*, *PLAT*, *CST1*, *CST2*, and *CST4*), or 5' *Hind*III and 3' *Xho*I (*PLAU*).

### Generation of stable cell lines

Stable shRNA-mediated knockdown was achieved with the pSuper-Retro system (OligoEngine) targeting the sequences 5'-GCTGTGCTTACTAGAAA-3', 5'-GCAAGTGCCTTAAGTCAA-3', and 5'-CGAAAGTGATGCTTTCCAT-3' for *COL6A1*, *PLOD2*, and *SRGN*, respectively. Stable overexpression of all genes was achieved using pRVPTOpuro retroviruses. Retroviral vectors were transfected into the packaging cell line H29. After 48 h viruses were collected, filtered, concentrated, and used to infect target cells in the presence of 5 µg/ml polybrene. The infected cells were selected with 0.8 µg/ml puromycin or 0.5 mg/ml hygromycin. In the case of combinatorial overexpression, concentrated virus was mixed and cells were not subjected to puromycin selection.

### Western blot

For western blot analysis, either IP or whole cell lysate samples were subjected to SDS-PAGE and subsequently transferred to Immobilon PVDF membranes (Millipore). Membranes were blocked in 5% milk for two hours prior to primary antibody incubation for two hours (antibody was diluted 1:1 000 in 5% milk). Human anti-COL6A1 (Santa Cruz SC 20649) was used. Coomassie blue staining of the membrane was used to assess equal loading of conditioned media samples.

### Statistics and bioinformatics

In considering whether proteins identified by mass spectrometry of conditioned media should in fact be considered secreted, amino acid sequences from all proteins identified were first submitted to the SignalP 3.0 server online database [26, 27] to identify predicted cleavable signal peptides. Submissions were searched within the eukaryotes organism group using the hidden Markov models method, and sequences scoring positive by the SignalP D-score within the N-terminal 70 amino acids were considered to be secretory-predicted. Secondly, all identified proteins were searched via the DAVID v. 6.7 [28, 29] website for GO cellular compartment (i.e. subcellular localization) annotations, and all any proteins annotated as being extracellular were considered to be secreted, even if they were not predicted as such by SignalP.

GSEA v2.0 [42, 72] was used to determine enrichment of bone metastasis signatures in normalized microarray data of published

experimental and clinical datasets [23, 44, 50, 73]. Clinical datasets were separated into two phenotypes for GSEA: those who developed bone metastasis and those who did not develop any metastasis, according to clinical, site-specific metastasis annotations [74]. For gene list ranking, multiple probe matches for the same gene were collapsed into one value, with the highest probe reading being used in each case. Only probes with matches to gene symbols were used. Genes were ranked using the provided signal-to-noise ranking statistic and GSEA was run using a weighted statistic and evaluated for statistical significance by comparison to results obtained using 1 000 random permutations of each gene set. For all other GSEA parameters, default settings were used.

For GO enrichment analysis, bone metastasis signature genes were loaded into the DAVID v. 6.7 [28, 29] website and submitted to the functional annotation tool. MF and Biological Function ontological databases were interrogated. For IPA (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)), signatures were loaded and analyzed according to default settings, with indirect network connections not being included.

For clinical dataset survival analyses, association between gene expression and bone metastasis was assessed by univariate Cox proportional hazards models or by Kaplan-Meier plotting followed by significance evaluation via log-rank test. In both cases, only patients having either bone metastases only (i.e. without concomitant metastases at other organs) or no metastasis at all were considered. For association between gene expression and non-site-specific metastasis, all patients – metastatic or non-metastatic – were considered and evaluated via univariate Cox proportional hazards models or via Kaplan-Meier plotting using the online KM-Plotter database [75].

For other statistical analyses, results were reported as average ± S.E. Two-sided independent Student's *t* tests without equal variance assumption or were performed to analyze the data with *P* < 0.05 considered as statistically significant.

### Tumor xenografts, bioluminescence, radiographic, and histological analyses

All animal work was performed in accordance with the guidelines of the Institutional Animal Care and Use Committee of Princeton University under approved protocols.  $1 \times 10^5$  cells were washed in PBS and injected intracardiac into female athymic Ncr-nu/nu mice to study bone metastasis activity as described previously [23]. Noninvasive bioluminescent imaging was performed to quantify the metastasis burden in using an IVIS 200 Imaging System (Caliper Life Sciences). Analysis was performed with Living Image software (Xenogen) by measuring photon flux of the region of interest. Data were normalized to the signal obtained immediately after injection (Day 1). X-Rays were obtained using 35 kV for 15 s. Bone isolation, fixation, decalcification, and subsequent TRAP staining was performed as previously described [76].

### Reverse transcription and quantitative PCR

Total RNA was isolated using the RNeasy kit (Qiagen) and reverse-transcribed with the Superscript III kit (Invitrogen) following the manufacturer's instructions. Quantitative PCR was performed using the SYBR Green PCR Master Mix (Applied Biosystems) with the ABI Prism 7900HT thermocycler (Applied Biosystems). The following primers were used for q-PCR: *hSRGN* forward: 5'-GGCTTGCCTGGCTCTTGCCC-3', reverse: 5'-TT-

GGTACCTGGCTCTCCGCGT-3'; *hCOL6A1* forward: 5'-CAGG-GACCCCCAGGACACCAA-3'; reverse: 5'-GGCCGCACTTG-CATTACACAGC-3'; *hPLOD2* forward: 5'-GCGTTCTCTTCGTC-CTCATCA-3', reverse: 5'-TGAAGCTCCAGCCTTTTCGTG-3'; *hCST1* forward: 5'-GCGGGTACTAAGAGCCAGGCAACA-3', reverse: 5'-GCACAGGTGTCCAAGTTGGGCTG-3'; *hCST2* forward: 5'-CCTGCTGCGGGTGTACGAG-3', reverse: 5'-GCA-CAGGTGTCCAAGTTGGGCT-3'; *hCST4* forward: 5'-CCGCT-GCAGGTGCTGCGAG-3', reverse: 5'-ATGGTGC GGCCAC-CTCTACG-3'; *hPLAT* forward: 5'-TTCGTTTCGCCAGCCA-GGAA-3', reverse: 5'-GCGCAGCCATGACTGATGTTGCT-3'; *hPLAU* forward: 5'-TCGTGAGCGACTCCAAAGGCA-3', reverse: 5'-GGCAGTTGCACCAGTGAATGTTGG-3'; *hGAPDH* forward: 5'-GAAGGTGAAGGTCCGAGTC-3', reverse: 5'-GAA-GATGGTGATGGGATTTC-3'; *mSRGN* forward: 5'-TCCTG-GCTCTCGCCTTCGTCC-3', reverse: 5'-TTCGGTTTGCAGCG-GACCCAC-3'; *mCOL6A1* forward: 5'-CACCTACACCGACT-GCGCCAT-3', reverse: 5'-CCCCTCGGTACCACGATCAA-3'; *mPLOD2* forward: 5'-GCGGCTGGCAGACAAGTACCC-3', reverse: 5'-AGACGGCTGATGTACGGGCA-3'; *mPLAT* forward: 5'-ATGCTGTGCGCTGGAGACACC-3', reverse: 5'-ATGCA-CACCAGAGGGCCTCCC-3'; *mPLAU* forward: 5'-GTGTGCT-GCGGACCCAGAGTG-3', reverse: 5'-CTCAGAGTTGGGCG-GCCTTCG-3'; *mGAPDH* forward: 5'-TCCCCTCTTCCAC-CTTCGATGC-3', reverse: 5'-GGGTCTGGGATGGAAGTGGT-GAGG-3'.

## Acknowledgments

We thank Saw Kyin and Henry Shwe for mass spectrometry assistance and Y Wei, N Sethi, and B Ell for technical assistance and helpful discussions. We also thank E Williams (Monash Institute of Medical Research) and RL Anderson (Peter MacCallum Cancer Centre) for the TSU-PR1 and 4T1.2 cell line series, respectively. This work was supported by grants from the National Institutes of Health R01CA134519 and R01CA141062 (Y Kang), the Department of Defense, Brewster Foundation and the Champalimaud Foundation. BA Garcia is supported by an NSF grant (CBET-0941143), an NJCCR SEED grant and a grant supported by award number DP2OD007447 from the Office of the Director, National Institutes of Health. MA Blanco was a recipient of an NRSA pre-doctoral fellowship from the National Institutes of Health. Barry Zee is a recipient of an NSF pre-doctoral fellowship.

## References

- Mundy GR. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nat Rev Cancer* 2002; **2**:584-593.
- Weilbaecher KN, Guise TA, McCauley LK. Cancer to bone: a fatal attraction. *Nat Rev Cancer* 2011; **11**:411-425.
- Juarez P, Guise TA. TGF-beta in cancer and bone: implications for treatment of bone metastases. *Bone* 2011; **48**:23-29.
- Korpala M, Yan J, Lu X, Xu S, Lerit DA, Kang Y. Imaging transforming growth factor-beta signaling dynamics and therapeutic response in breast cancer bone metastasis. *Nat Med* 2009; **15**:960-966.
- Guise T. Examining the metastatic niche: targeting the microenvironment. *Semin Oncol* 2010; **37** Suppl 2:S2-S14.
- Makridakis M, Vlahou A. Secretome proteomics for discovery of cancer biomarkers. *J Proteomics*; **73**:2291-2305.
- Sethi N, Kang Y. Unravelling the complexity of metastasis – molecular understanding and targeted therapies. *Nat Rev Cancer* 2011; **11**:735-748.
- Luo X, Liu Y, Wang R, Hu H, Zeng R, Chen H. A high-quality secretome of A549 cells aided the discovery of C4b-binding protein as a novel serum biomarker for non-small cell lung cancer. *J Proteomics* 2011; **74**:528-538.
- Tang CE, Guan YJ, Yi B, *et al.* Identification of the amyloid beta-protein precursor and cystatin C as novel epidermal growth factor receptor regulated secretory proteins in nasopharyngeal carcinoma by proteomics. *J Proteome Res* 2011; **9**:6101-6111.
- Schiarea S, Solinas G, Allavena P, *et al.* Secretome analysis of multiple pancreatic cancer cell lines reveals perturbations of key functional networks. *J Proteome Res* 2011; **9**:4376-4392.
- Wang CI, Wang CL, Wang CW, *et al.* Importin subunit alpha-2 is identified as a potential biomarker for non-small cell lung cancer by integration of the cancer cell secretome and tissue transcriptome. *Int J Cancer* 2011; **128**:2364-2372.
- Chang KP, Wu CC, Chen HC, *et al.* Identification of candidate nasopharyngeal carcinoma serum biomarkers by cancer cell secretome and tissue transcriptome analysis: potential usage of cystatin A for predicting nodal stage and poor prognosis. *Proteomics* 2010; **10**:2644-2660.
- Kerr BA, Miocinovic R, Smith AK, Klein EA, Byzova TV. Comparison of tumor and microenvironment secretomes in plasma and in platelets during prostate cancer growth in a xenograft model. *Neoplasia* 2010; **12**:388-396.
- Makridakis M, Roubelakis MG, Bitsika V, *et al.* Analysis of secreted proteins for the study of bladder cancer cell aggressiveness. *J Proteome Res* 2010; **9**:3243-3259.
- Xu BJ, Yan W, Jovanovic B, *et al.* Quantitative analysis of the secretome of TGF-beta signaling-deficient mammary fibroblasts. *Proteomics* 2010; **10**:2458-2470.
- Wu CC, Hsu CW, Chen CD, *et al.* Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol Cell Proteomics* 2010; **9**:1100-1117.
- Srisomsap C, Sawangareetrakul P, Subhasitanont P, *et al.* Proteomic studies of cholangiocarcinoma and hepatocellular carcinoma cell secretomes. *J Biomed Biotechnol* 2010; **2010**:437143.
- Xue H, Lu B, Zhang J, *et al.* Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J Proteome Res* 2010; **9**:545-555.
- Chang YH, Wu CC, Chang KP, Yu JS, Chang YC, Liao PC. Cell secretome analysis using hollow fiber culture system leads to the discovery of CLIC1 protein as a novel plasma marker for nasopharyngeal carcinoma. *J Proteome Res* 2009; **8**:5465-5474.
- Wang CL, Wang CI, Liao PC, *et al.* Discovery of retinoblastoma-associated binding protein 46 as a novel prognostic

- marker for distant metastasis in nonsmall cell lung cancer by combined analysis of cancer cell secretome and pleural effusion proteome. *J Proteome Res* 2009; **8**:4428-4440.
- 21 Ji H, Greening DW, Kapp EA, Moritz RL, Simpson RJ. Secretome-based proteomics reveals sulindac-modulated proteins released from colon cancer cells. *Proteomics Clin Appl* 2009; **3**:433-451.
- 22 Mathias RA, Wang B, Ji H, *et al.* Secretome-based proteomic profiling of Ras-transformed MDCK cells reveals extracellular modulators of epithelial-mesenchymal transition. *J Proteome Res* 2009; **8**:2827-2837.
- 23 Kang Y, Siegel PM, Shu W, *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003; **3**:537-549.
- 24 Lelekakis M, Moseley JM, Martin TJ, *et al.* A novel orthotopic model of breast cancer metastasis to bone. *Clin Exp Metastasis* 1999; **17**:163-170.
- 25 Chaffer CL, Dopheide B, McCulloch DR, *et al.* Upregulated MT1-MMP/TIMP-2 axis in the TSU-Pr1-B1/B2 model of metastatic progression in transitional cell carcinoma of the bladder. *Clin Exp Metastasis* 2005; **22**:115-125.
- 26 Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007; **2**:953-971.
- 27 Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795.
- 28 Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**:44-57.
- 29 Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**:1-13.
- 30 Karagiannis GS, Petraki C, Prassas I, *et al.* Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget* 2012; **3**:267-285.
- 31 Loei H, Tan HT, Lim TK, *et al.* Mining the gastric cancer secretome: identification of GRN as a potential diagnostic marker for early gastric cancer. *J Proteome Res* 2012; **11**:1759-1772.
- 32 Zhao T, Zeng X, Bateman NW, *et al.* Relative quantitation of proteins in expressed prostatic secretion with a stable isotope labeled secretome standard. *J Proteome Res* 2012; **11**:1089-1099.
- 33 Pocsfalvi G, Votta G, De Vincenzo A, *et al.* Analysis of secretome changes uncovers an autocrine/paracrine component in the modulation of cell proliferation and motility by c-Myc. *J Proteome Res* 2011; **10**:5326-5337.
- 34 Yao L, Zhang Y, Chen K, Hu X, Xu LX. Discovery of IL-18 as a novel secreted protein contributing to doxorubicin resistance by comparative secretome analysis of MCF-7 and MCF-7/Dox. *PLoS One* 2011; **6**:e24684.
- 35 Ji H, Goode RJ, Vaillant F, *et al.* Proteomic profiling of secretome and adherent plasma membranes from distinct mammary epithelial cell subpopulations. *Proteomics* 2011; **11**:4029-4039.
- 36 Yu CJ, Wang CL, Wang CI, *et al.* Comprehensive proteome analysis of malignant pleural effusion for lung cancer biomarker discovery by using multidimensional protein identification technology. *J Proteome Res* 2011; **10**:4671-4682.
- 37 Chang YT, Wu CC, Shyr YM, *et al.* Secretome-based identification of ULBP2 as a novel serum marker for pancreatic cancer detection. *PLoS One* 2011; **6**:e20029.
- 38 Polisetty RV, Gupta MK, Nair SC, *et al.* Glioblastoma cell secretome: analysis of three glioblastoma cell lines reveal 148 non-redundant proteins. *J Proteomics* 2011; **74**:1918-1925.
- 39 Ralhan R, Masui O, Desouza LV, Matta A, Macha M, Siu KW. Identification of proteins secreted by head and neck cancer cell lines using LC-MS/MS: strategy for discovery of candidate serological biomarkers. *Proteomics* 2011; **11**:2363-2376.
- 40 Formolo CA, Williams R, Gordish-Dressman H, MacDonald TJ, Lee NH, Hathout Y. Secretome signature of invasive glioblastoma multiforme. *J Proteome Res* 2011; **10**:3149-3159.
- 41 Ha-Vinh R, Alanay Y, Bank RA, *et al.* Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *Am J Med Genet A* 2004; **131**:115-120.
- 42 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**:15545-15550.
- 43 Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 2007; **23**:3251-3253.
- 44 Minn AJ, Gupta GP, Siegel PM, *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* 2005; **436**:518-524.
- 45 Dombkowski AA, Cukovic D, Novak RF. Secretome analysis of microarray data reveals extracellular events associated with proliferative potential in a cell line model of breast disease. *Cancer Lett* 2006; **241**:49-58.
- 46 Antelmann H, Tjalsma H, Voigt B, *et al.* A proteomic view on genome-based signal peptide predictions. *Genome Res* 2001; **11**:1484-1502.
- 47 Kreunin P, Urquidi V, Lubman DM, Goodison S. Identification of metastasis-associated proteins in a human tumor metastasis model using the mass-mapping technique. *Proteomics* 2004; **4**:2754-2765.
- 48 Mbeunkui F, Metge BJ, Shevde LA, Pannell LK. Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. *J Proteome Res* 2007; **6**:2993-3002.
- 49 Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001; **19**:242-247.
- 50 van de Vijver MJ, He YD, van't Veer LJ, *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; **347**:1999-2009.
- 51 Lamande SR, Morgelin M, Selan C, Jobsis GJ, Baas F, Bateman JF. Kinked collagen VI tetramers and reduced microfibril formation as a result of Bethlem myopathy and introduced triple helical glycine mutations. *J Biol Chem* 2002; **277**:1949-1956.
- 52 Lamande SR, Bateman JF, Hutchison W, *et al.* Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional ha-



- plainsufficiency. *Hum Mol Genet* 1998; **7**:981-989.
- 53 Lampe AK, Flanigan KM, Bushby KM. Collagen Type VI-Related Disorders. 1993.
- 54 Bonaldo P, Braghetta P, Zanetti M, Piccolo S, Volpin D, Bresnan GM. Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet* 1998; **7**:2135-2140.
- 55 Chiu KH, Chang YH, Wu YS, Lee SH, Liao PC. Quantitative secretome analysis reveals that COL6A1 is a metastasis-associated protein using stacking gel-aided purification combined with iTRAQ labeling. *J Proteome Res* 2011; **10**:1110-1125.
- 56 Levental KR, Yu H, Kass L, *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* 2009; **139**:891-906.
- 57 Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. *Cell* 2006; **126**:677-689.
- 58 Erler JT, Bennewith KL, Cox TR, *et al.* Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. *Cancer Cell* 2009; **15**:35-44.
- 59 Furlan F, Galbiati C, Jorgensen NR, *et al.* Urokinase plasminogen activator receptor affects bone homeostasis by regulating osteoblast and osteoclast function. *J Bone Miner Res* 2007; **22**:1387-1396.
- 60 Tumber A, Papaioannou S, Breckon J, Meikle MC, Reynolds JJ, Hill PA. The effects of serine proteinase inhibitors on bone resorption *in vitro*. *J Endocrinol* 2003; **178**:437-447.
- 61 Brzin J, Popovic T, Turk V, Borchart U, Machleidt W. Human cystatin, a new protein inhibitor of cysteine proteinases. *Biochem Biophys Res Commun* 1984; **118**:103-109.
- 62 Turk V, Stoka V, Turk D. Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* 2008; **13**:5406-5420.
- 63 Blankenvoorde MF, Henskens YM, van't Hof W, Veerman EC, Nieuw Amerongen AV. Inhibition of the growth and cysteine proteinase activity of *Porphyromonas gingivalis* by human salivary cystatin S and chicken cystatin. *Biol Chem* 1996; **377**:847-850.
- 64 Jensen JL, Lamkin MS, Oppenheim FG. Adsorption of human salivary proteins to hydroxyapatite: a comparison between whole saliva and glandular salivary secretions. *J Dent Res* 1992; **71**:1569-1576.
- 65 LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, Garcia BA. Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. *Mol Cell Proteomics* 2009; **8**:2432-2442.
- 66 Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 1996; **68**:850-858.
- 67 Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 2003; **75**:663-670.
- 68 Korpál M, Ell BJ, Buffa FM, *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 2011; **17**:1101-1108.
- 69 Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002; **74**:5383-5392.
- 70 Plazas-Mayorca MD, Zee BM, Young NL, *et al.* One-pot shotgun quantitative mass spectrometry characterization of histones. *J Proteome Res* 2009; **8**:5367-5374.
- 71 Khan Z, Bloom JS, Garcia BA, Singh M, Kruglyak L. Protein quantification across hundreds of experimental conditions. *Proc Natl Acad Sci USA* 2009; **106**:15544-15548.
- 72 Mootha VK, Lindgren CM, Eriksson KF, *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003; **34**:267-273.
- 73 Wang Y, Klijn JG, Zhang Y, *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005; **365**:671-679.
- 74 Bos PD, Zhang XH, Nadal C, *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* 2009; **459**:1005-1009.
- 75 Györfy B, Lanczky A, Eklund AC, *et al.* An online survival analysis tool to rapidly assess the effect of 22 277 genes on breast cancer prognosis using microarray data of 1 809 patients. *Breast Cancer Res Treat* 2010; **123**:725-731.
- 76 Sethi N, Dai X, Winter CG, Kang Y. Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells. *Cancer Cell* 2011; **19**:192-205.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)