Published in final edited form as:

Clin Exp Metastasis. 2005 ; 22(7): 593–603.

Metastasis predictive signature profiles pre-exist in normal tissues

Haiyan Yang, Nigel Crawford, Luanne Lukes, Richard Finney, Mindy Lancaster, and Kent W. Hunter.*

Laboratory of Population Genetics, National Cancer Institute, Bethesda, Maryland 20892

Abstract

Previous studies from our laboratory have demonstrated that metastatic propensity is significantly influenced by the genetic background upon which tumors arise. We have also established that human gene expression profiles predictive of metastasis are not only present in mouse tumors with both high and low metastatic capacity, but also correlate with genetic background. These results suggest that human metastasis-predictive gene expression signatures may be significantly driven by genetic background, rather than acquired somatic mutations. To test this hypothesis, gene expression profiling was performed on inbred mouse strains with significantly different metastatic efficiencies. Analysis of previously described human metastasis signature gene expression patterns in normal tissues permitted accurate categorization of high or low metastatic mouse genotypes. Furthermore, prospective identification of animals at high risk of metastasis was achieved by using mass spectrometry to characterize salivary peptide polymorphisms in a genetically heterogeneous population. These results strongly support the role of constitutional genetic variation in modulation of metastatic efficiency and suggest that predictive signature profiles could be developed from normal tissues in humans. The ability to identify those individuals at high risk of disseminated disease at the time of clinical manifestation of a primary cancer could have a significant impact on cancer management.

Introduction

Elucidation of the mechanisms governing metastases development are of great importance given that the majority cancer-related patient mortality is due to metastatic disease [1]. While it is recognized that metastasis occurs as a result of complex interactions between normal tissues and neoplastic cells, most contemporary studies primarily focus on how tumor-specific changes affect both the biological and clinical outcome. For example, a number of studies have identified metastasis suppressor genes, whose loss of function in primary tumors promotes metastatic progression [2]. The ability of a tumor to successfully form a metastasis, however, is potentially influenced by many complex and interacting factors that are not necessarily inherent to the primary tumor itself. Indeed, it is plausible that the metastatic capacity any cells shed by a primary tumor is significantly influenced by variability in the functional state of normal tissues that metastatic cells come into contact with while both in transit to and in residence at the site of secondary lesion formation. Therefore, the observed differential metastatic capacity evident in human tumorigenesis is a function not only of somatic events within tumor cells, but also from constitutional genetic heterogeneity affecting gene expression in transit and secondary sites. It has been demonstrated that constitutional genetic polymorphism in mice can influence almost any measurable trait [3], including gene expression, with steady state levels of mRNAs in tissues being significantly influenced by

*Corresponding Author: Kent W. Hunter, Laboratory of Population Genetics, CCR/NCI/NIH, Bldg 41 Rm 702, 41 Library Drive, Bethesda, MD 20892, tel: 301-435-8957, fax: 301-435-8963, email: hunterk@mail.nih.gov

genetic background [4–7]. Additionally, such studies have shown that it is possible to map genetic factors controlling transcript levels in numerous organisms.

These observations suggest that a complex process like metastasis could be significantly influenced by the genetic background upon which a tumor arises. To address this question, we have previously conducted a series of studies demonstrating that the genetic background upon which the polyoma middle-T (PyMT) transgene is carried has a significant impact on the metastatic efficiency of resulting mammary tumors. Specifically, when PyMT mice were crossed with inbred mice of different strains, the resulting F1 progeny demonstrated significantly different frequencies of macroscopic lung metastasis formation, ranging from ~10-fold fewer than the original FVB inbred background to ~2-fold more [8]. The variation in the number of observed metastases in the F1 progeny is most likely the result of a constitutional genetic polymorphism in the recipient animals since the tumors in all of the progeny result from polyoma middle-T transgene activation. In concordance with these observations, we have recently identified a gene in which hereditary variation significantly contributes to the modulation of metastatic efficiency in this system [9].

In addition, we have observed that the expression of a number of genes described as components of a metastasis predictive signature set [10] were differentially regulated between tumors of different genotypes and metastatic potential, thus largely replicating the published human metastasis signature profile [11,12]. Since all tumors in our model system were induced by the same oncogenic event, the induction of the PyMT transgene, we hypothesize that genetic polymorphism, rather than oncogenic mutation as previously suggested [13], regulates expression of constituents of such metastasis predicting gene signature profiles. Furthermore, if this hypothesis is correct, it would be possible to make two important predictions; first, that signature genes should be differentially expressed in normal tissues prior to tumor induction; and second, by profiling normal tissues prior to the occurrence of metastatic disease, it should theoretically be possible to prospectively predict metastatic risk given that the polymorphisms responsible for differential signature gene expression exist in germ line DNA.

To test this hypothesis we have examined the expression of a known [10] and a novel metastasis predictive gene signature profile in normal tissues within the PyMT mouse mammary tumor model to attempt to differentiate between high and low metastatic genotypes. Furthermore, we aim to explore the possibility of prospectively identifying animals at high risk of metastasis by analyzing normal tissues prior to metastasis development.

Materials and methods

Animals and husbandry

Transgenic DBA and NZB F₁ hybrids were generated by crossing two inbred strains, [FVB/N-TgN(MMTV-PyVT)^{634Mul} × inbred (DBA/2J or NZB/B1NJ)]. FVB F₁ transgenic mice were generated by crossing inbred FVB/N-TgN(MMTV-PyVT)^{634Mul} × FVB/NJ. Transgene positive DBA F₁ males were then bred to DBA/2J females to generate DBA N₂ backcross animals.

The AKXD RI panel consists of 25 inbred sublines with different combinations of the original parental genomes were derived from interbreeding AKR/J (a high metastatic genotype) and DBA/2J (a low metastatic genotype) as previously described [14]. The PyMT animal FVB/N-TgN(MMTV-PyVT)^{634Mul} was bred to 18 of the 25 AKXD sublines to generate 18 [PyMT × AKXD]F₁ sublines to provide genetic mapping information for metastasis efficiency modifier genes [15]. Transgene positive female animals were aged for 100 days, euthanized by carbon dioxide exposure and tissues harvested for analysis. Lungs were sectioned and metastases were quantitated using a Leica Q500IW image analysis system, as previously described [8].

Tissue sample collection

Mammary tumors were excised from female transgenic DBA F₁, NZB F₁ strains or AKXD sublines. Normal mammary gland, lung tissues and whole blood collected from non-transgenic FVB F1 or F1 progenies between [FVB/N × DBA/J2] or [FVB/N × NZB/B1NJ] were snap-frozen upon harvesting and stored at -80°C. These tissues were selected because they are the source of the primary tumor and the metastatic target organ in this model. Blood was also assayed due to its relatively easy clinical accessibility to model potential use in human populations. In some instances, tissues were placed in RNeasy Lysis Buffer (Qiagen) prior to freezing at -80°C, in accordance with the manufacturer's recommended protocol. Prior to RNA isolation, all tissues were pulverized while frozen solid on dry ice in an RNase free environment.

RNA extraction and cDNA synthesis

Total RNA extractions from tissue samples were carried out using TRIzol[®] Reagent (Life Technologies, Inc.) according to the standard protocol. Total RNA was prepared from whole blood using QIAamp RNA blood mini kit (Qiagen) per manufacturer's instruction. RNA quantity and quality were determined by the Agilent Technologies 2100 Bioanalyzer (Bio Sizing Software version A.02.01., Agilent Technologies) and/or the GeneQuant Pro (Amersham Biosciences Corp.). Samples containing high-quality total RNA with A₂₆₀/A₂₈₀ ratios between 1.8 and 2.1 were purified with the RNeasy Mini Kit (Qiagen). An on-column genomic DNA digestion was performed as part of this purification step using the RNase-Free DNase Kit (Qiagen).

Quantitative real-time PCR

First-strand cDNA was synthesized from total RNA by reverse transcription using the ThermoScript[™] RT-PCR System (Invitrogen), according to the manufacturer's protocol. Synthesized cDNA template was diluted 1:40 prior to use in quantitative real-time PCR (qPCR). Primers were either purchased from the Applied Biosystems Assays-On-Demand[™] Gene Expression product line or designed using Primer3 (Applied Biosystems, Foster City, CA). Reactions were prepared using Taqman[®] Universal PCR Mastermix (Applied Biosystems) or QuantiTect[™] SYBR[®] Green PCR Mastermix (Qiagen) per manufacturer's protocols, and were performed in triplicates on an ABI Prism[®] 7900HT Sequence Detection System. Target gene expression was normalized to that of *Gapdh* using Taqman[®] Rodent GAPDH (Applied Biosystems) internal control primers.

RNA extraction and processing for Affymetrix GeneChip[®] analysis

Purified total RNA for each strain used in Affymetrix GeneChip assays was processed as previously described [16]. Hybridizations were performed on Affymetrix Murine Genome Moe430 A and B GeneChip[®] Arrays. Microarrays were processed using an Agilent GeneArray Scanner with Affymetrix Microarray Suite version 5.0.0.032 software.

Analysis of Affymetrix GeneChip[®] data

Further analysis of raw microarray data, including normalization and statistical analysis was performed using BRB-Array Tools v3.2.0 [17,18]. Array data were filtered to exclude genes possessing < 20 % of expression data having at least a 1.5 -fold change in either direction from gene's median value, and p-value of the log-ratio variation < 0.01. Metastasis predictive gene signature expression profiles were then determined using the class prediction function of BRB ArrayTools. Models that prospectively predict sample class were created using the Compound Covariate Predictor [19] and Diagonal Linear Discriminant Analysis functions [20]. The models incorporated differentially expressed genes at the *P* < 0.05 significance level, as assessed by the random variance t-test [21]. The prediction error of each model was estimated

using leave-one-out cross-validation (LOOCV) as described by Simon *et al* [22]. The entire model building process was repeated for each LOOCV training set including the process of gene selection. Additionally, we evaluated whether the cross-validated error rate estimate for each model was significantly less than one would expect from random prediction. Class labels were randomly permuted, and the entire LOOCV process repeated. The significance level is reported as the proportion of random permutations that gave a cross-validated error rate no greater than the cross-validated error rate obtained with real data. One thousand random permutations were used.

Clustering analysis

Clustering analysis of relative gene expression levels from normal mammary gland, as determined by qPCR was performed using the tree joining with complete linkage and Euclidean distance functions of the program Statistica ver 5.5.

Expression quantitative trait loci analysis

The Affymetrix .CEL files were normalized using the RMA method, averaged for each AKXD recombinant inbred strain, and loaded into the internet based quantitative trait loci (QTL) analysis package WebQTL (<http://webqtl.org/>) [23]. The database was then searched for probe sets for each of the metastasis predictive signature genes described in Ramaswamy *et al.* [10]. The interval mapping function of WebQTL was performed with each of the probe sets to identify regions of the genome that correlated with signal intensity levels of each probe set to identify potential expression QTLs (eQTLs).

Saliva collection

Saliva was collected from DBA N₂ animals at 60 days of age, an age at which the majority of animals have developed palpable tumors. Animals were fasted for three hours prior to sample collection, anesthetized using an intraperitoneal injection of Avertin (0.010 g/ml) and injected subcutaneously with pilocarpine (0.5 mg/ml in PBS) at a concentration of 0.5 mg/kg mouse weight. Saliva was collected with a sterile plastic pipette tip attached to a syringe with animals placed head down on an inclined platform. Saliva was then centrifuged to pellet debris, aliquoted into sterile eppendorf tubes, snap frozen and stored at -80°C.

Surface-enhanced laser desorption/ionization mass spectrometer

Following quantification of salivary protein concentration by BCA assay (Pierce), samples were diluted 1:1 in 9 M urea and 2% CHAPS, and vortex mixed for 30 minutes at 4°C. Ten microliters of each saliva protein sample (100 µg/ml) was then mixed with 90 µl of binding buffer (4.5 M urea, 1% CHAPS and 50 mM NaOAc, pH 4.5), and applied to individual spots on CM10 chips (Ciphergen Biosystems, Fremont, CA) in duplicate. The chips were then incubated at room temperature for 2 hours with agitation, and washed three times with binding buffer followed by a final rinse with deionized water. Following air drying, two 1 µl applications of matrix solution (saturated solution of α -cyano-4-hydroxycinnamic acid in 50% acetonitrile with 0.5% trifluoroacetic acid) were applied to each chip spot. Once dried, the chips were analyzed in a ProteinChip Biology System reader (PBS II, Ciphergen Biosystems). Mass spectral data over an m/z range of 2,000–30,000 were collected by averaging at least 70 laser shots from various regions of the spot surface, with a laser intensity of 160 and a detection sensitivity of 10. Saliva mass spectra were calibrated externally using the All-in-1 peptide standard (Ciphergen) per manufacture's instruction and normalized using total ion current. Peak identification (signal/noise ≥ 3) and alignment (cluster mass window at 0.3%) were performed using Ciphergen Biowizard 3.1.1 (Ciphergen).

Analysis of SELDI MS data

SELDI MS data were randomly divided into a training set containing spectra from 28 mice, and a test set containing spectra from 14 mice. Mice in the training set were classified as high- or low-metastatic phenotypes based on their metastatic indexes compared to the average metastatic index of the whole population. Intensity signals of protein profiles from each animal were obtained by averaging peak intensities of the duplicate spectra. Class prediction was performed as described above in analysis of Affymetrix GeneChip® Data using BRB ArrayTools. Models incorporated m/z values that were differentially expressed among m/z values at the 0.006 significance level as assessed by the random variance t-test. These m/z peak values were then applied to the test set to classify the metastatic phenotype (in relation to the average metastatic index of the training set) of the mouse using the compound covariate predictor and diagonal linear discriminant analysis algorithms.

Results

Quantitative PCR analysis of mammary tissue

To test the hypothesis that tumor-derived metastasis signature profile genes were also differentially expressed in normal tissues, qPCR was performed for 10 of the 17 metastasis signature genes defined by Ramaswamy et al. [10]. Nine out of ten of the genes exhibited differential expression between the FVB and NZB F₁ normal mammary tissue samples, (Figure 1a), while all 10 of the genes were differentially expressed in the FVB versus DBA F₁ comparison. In addition, the expression patterns of 9 out of 10 of the signature set genes were concordant between NZB and DBA F₁'s.

Evidence for differential expression of signature genes was also observed in lung tissues, though the magnitude of these differences was significantly reduced compared to the mammary tissue. Only two genes in the DBA and five genes in the NZB versus FVB comparisons differed by 1.5-fold or greater (Figure 1b). The remaining genes, however, demonstrated consistent, albeit modest, expression differences suggesting that the observed variation was likely real. As in the mammary qPCR experiments, most gene expression patterns were concordant between the NZB and DBA F₁ lung tissues. However, the expression patterns observed in the lung were significantly different from those seen in mammary tissue. Furthermore, both profiles differed in some respects from the published metastasis signature [10], which is not an entirely unexpected observation given the divergent cellular compositions of lung, mammary and tumor samples. It is therefore highly improbable that these dissimilar tissue types would have identical gene expression patterns.

Gene expression profiling was also performed on mRNA extracted from whole blood. Nine of the ten profile genes failed to amplify, which is not surprising given that the original gene expression profile was derived from solid tumors rather than cells of hematopoietic origin. One gene, *Pttg1*, was detectable, and displayed strain-specific expression levels similar to those observed in normal mammary tissue, with concordance observed between the DBA and NZB strains (Figure 1c).

Correlation of metastasis efficiency QTLs with eQTLs

Differential expression of metastasis predictive expression signature genes provides strong, yet not definitive, evidence that genetic background influences both metastatic and prognostic signatures. Co-localization of the metastasis efficiency modifying genes and any genetic factors controlling expression levels of individual members of the signature profile would provide additional support of this hypothesis. Therefore, gene expression profiling was performed on three independent tumor samples from each of the 18 substrains of the AKXD recombinant inbred mouse panel used in previous studies to map the metastasis efficiency

modifier locus *Mtes1* [15]. Due to the lack of statistical power associated with only 18 genotypes in the sample set, a statistically significant genetic factor influencing expression levels was found with only one gene, *Coll1a1* (see Figure 2). Suggestive genetic factors were observed for 13 genes, while three additional genes did not produce either suggestive or significant results. However, of the 14 genes exhibiting suggestive or significant results, 9 were linked to chromosomes associated with metastatic efficiency modifier loci [chrs 4, 9, 11, 17; [15,24] and unpublished results], consistent with a possible link between the metastasis and gene expression modifying loci.

Identification and analysis of mouse metastasis gene signature set

Although normal tissue data is consistent with the hypothesis that genetic background influences both metastatic capacity and the gene signature profile, one may argue that expression of the human metastasis predictive gene set may be irrelevant in regards to metastatic efficacy in our mouse model. Therefore, we generated a mouse metastasis gene signature profile to test whether these phenomena could be observed using mouse genes directly linked to metastatic propensity. The AKXD recombinant inbred (RI) metastasis and gene expression data were therefore used to generate a mouse metastasis predictive gene expression profile, analogous to the previously describe human signature profiles. The Affymetrix .CEL data was loaded into BRB ArrayTools and normalized by the RMA method. The average density of pulmonary metastases in each of the 18 AKXD RI sublines was then analyzed using the Class Prediction tool to generate a signature profile. A 58 probe-set classifier that predicted metastasis with 77% accuracy was identified from the 18 segregating genotypes (see Table 1).

Quantitative RT-PCR analysis was performed to determine whether predictive genes were differentially expressed in normal tissue. Primers were therefore designed against each gene in the predictive profile. Thirty eight of the 58 primer pairs selectively amplified cDNA but not genomic DNA and were used for qPCR experiments (see Table 1). The relative abundance of each of the genes of the mouse metastasis classifier was quantified in a set of normal FVB ($n=5$) or [FVB $\times$ DBA]F₁ ($n=4$) mammary glands, and used in an unsupervised clustering analysis to determine whether the high metastatic FVB genotype would cluster separately from the low metastatic [FVB $\times$ DBA]F₁ genotype. Consistent with our hypothesis, expression of the mouse metastasis gene signature set in normal tissues segregated the high and low metastatic genotypes (see Figure 3).

Metastasis prediction in a genetically segregating population

These data are consistent with the possibility that the metastasis predictive gene profile is at least partially the result of germline polymorphism rather than somatic mutation, and the gene expression modifiers and metastasis efficiency modifiers may be related. If true, this suggests that metastasis signature profiles may be functioning as a surrogate for genotype, with each expression profile member representing one or more polymorphic sites in a genome-wide scan, analogous to a complex trait mapping study. Since the polymorphisms underlying a significant fraction of the gene expression differences are present in germline DNA, this would suggest that it should be possible to use other technologies that permit genome-wide screens of polymorphisms to predict metastatic capacity. In addition, since germ line polymorphisms are present in all tissues of the body, it should be possible to perform the assays in tissues other than tumor tissue.

To test these possibilities, mass spectrometric profiling of mouse saliva was performed. Examination of microarray data from high and low metastatic genotypes revealed that significant differences were observed in the expression of a number of different salivary protein components ([12]; see Figure 4). In addition, a literature search revealed that known

polymorphisms in salivary gland proteins largely segregated mouse strains into known high- and low-metastatic classes [25–27]. Saliva was therefore collected from FVB and DBA F1 animals and assayed by Ciphergen SELDI-TOF to determine whether the two genotypes would cluster independently. One hundred and thirty eight peptide peaks from the saliva profiles were used in an unsupervised clustering analysis. As expected, the two genotypes segregated with only one single error (Figure 5). Furthermore, transgene-positive samples did not segregate separately from transgene-negative samples in either the FVB or DBA F1 samples, arguing against the possibility that the presence of tumor cells might influence a salivary peptide predictive profile.

To extend these result to a genetically heterogeneous population more analogous to human cancer populations, a backcross was performed to generate PyMT-positive N₂ animals. Saliva was collected at 60 days of age, and saliva protein profiles analyzed in duplicate on Ciphergen CM10 chips. Animals were held until 100 days of age, when their metastatic burden was determined by measurement of pulmonary metastasis density on coronal H & E sections. The number of metastases in the N₂ population ranged from 0–37 per cm² of lung tissue.

From the spectra of a training set consisting of 28 animals, a predictive peptide profile was developed that would classify animals as either above or below the median metastasis value of the training set. Four hundred and sixty nine peptide peaks were analyzed to determine a predictive peptide set. Cross-validation was performed to identify a predictive profile consisting peptide peaks with m/z values of 2153.99, 2481.01, 2770.25, 2524.64, 2567.03 that could correctly categorize training set samples with ~80% accuracy, and a sensitivity and specificity of 77–85% (Table 1). An additional 14 samples were subsequently classified using the predictors and the predicted assignment compared to the experimental metastasis data. Approximately 80% of the samples (11/14) of the test set were correctly classified (Table 2).

Discussion

The recent observation that gene expression profiles from bulk tumors can define high and low metastatic propensities prior to the formation of overt clinical secondary lesions [10,28,29] suggests that a significant proportion of cells in a tumor must exhibit the predictive gene expression profile. Thus, it has been suggested that metastatic potential is encoded early in tumorigenesis, most likely by the oncogenic mutations themselves [13]. However, these results are inconsistent with the commonly accepted progressive theory of metastasis [30,31], which predicts that only a subpopulation of a tumor obtains full metastatic competency. We argue that our previous genetic and expression studies [8,12,15] explain this paradox. It would therefore follow that the genetic background from which a tumor arose might significantly influence the metastatic efficiency of a primary tumor, and be a major determinant of the prognostic expression profile [11,32]. If this theory was pursued to its logical conclusion, it should therefore be possible to identify those individuals at high risk using differential expression patterns in normal tissues.

To test this hypothesis, we have examined expression of the proposed metastasis predictor gene set in normal tissues. Consistent with our hypothesis, the data presented here demonstrate that a significant fraction of the predictive metastatic signature genes proposed by Ramaswamy et al. [10] are differentially expressed in normal tissues between high- (FVB) and low- (DBA, NZB F₁) metastatic genotypes prior to the induction of malignant disease. These data, in addition to other studies demonstrating the significant impact of genetic background on gene expression [4–7], suggest that tumor metastatic propensity and predictive gene expression patterns are significantly influenced by constitutional polymorphisms rather than solely by oncogenic mutations. These data are also entirely consistent with the recent report by Malins et al. [33] who demonstrated that it was possible to distinguish metastatic and non-metastatic

prostate cancer patients based on DNA structure in normal prostate epithelium, suggesting that predisposition to metastasis is constitutively encoded.

More importantly, these observations suggest that prospective identification of patients at higher risk of developing disseminated disease using readily available samples instead of the more difficult to obtain tumor tissue might well be possible. In support of this hypothesis, we have demonstrated that it is possible to use saliva peptide polymorphisms to prospectively identify individuals at high risk for metastatic disease in a genetically segregating population. This approach has enabled the protein expression pattern itself, independent of the identity of the proteins or peptides, to be employed as a discriminator of metastatic propensity. Although significant improvements in the sensitivity and specificity of the assays are obviously required to achieve the level of accuracy required for clinical use, the ability to correctly categorize ~80% of the samples based on a relatively small training set suggests this methodology, or some variant of it, might be of significant value for clinical prognosis.

To achieve the required level of accuracy, a variety of strategies are currently being pursued. Foremost is the use of more sensitive instrumentation to increase the resolution, sensitivity, reproducibility and the number of the data points used to derive signature gene profiles. Increased sensitivity will likely permit the identification of more reliable and informative markers than is currently possible with a low resolution methodology such as SELDI-TOF. Second, a much larger cohort of animals or individuals will likely be required to generate a clinically usable predictor than were used in this proof-of-principle study. Significant variability in the metastasis frequency exists in both high and low metastatic genotypes, resulting in a small fraction of animals that are highly predisposed to metastasis phenotypically resembling low metastatic strains and vice versa. Inclusion of misclassified individuals in the training set would therefore reduce the accuracy of the predictor and likely is partially responsible for the 20% inaccuracy rate observed in the current study. Significant increases in the number of samples used to generate the predictor should reduce the confounding effects of misclassified samples, thus improving the overall predictive value. Third, analysis of tissues other than saliva might provide a better prognostic profile. While perhaps the most easily available non-invasive human tissue, salivary secretions might not possess adequate complexity to provide sufficient marker variability to enable accurate prognosis in the more genetically complex human population. Subsequently, blood or serum samples may provide a richer source of polymorphisms than saliva for more accurate clinical diagnosis.

The ability to prospectively identify those patients at increased risk of tumor dissemination also raises the potential for pre-emptive use of anti-metastatic therapies. Identifying patients with a higher likelihood of developing metastases prior to detectable secondary disease and enrolling them in metastasis-prevention regimens might significantly reduce the burden of secondary lesions. Preliminary evidence of the feasibility of such a strategy is currently ongoing in our laboratory, and by using a small molecule agent, caffeine, we have demonstrated the ability to significantly reduce the efficiency of pulmonary colonization in the PyMT animal [16]. If these studies can be translated into human populations, similar strategies and agents might eventually lessen metastasis-associated mortality and morbidity. Alternatively, identifying high risk patients would also permit better monitoring to enable earlier therapeutic intervention against metastatic disease.

Finally, it is also important to point out that while the outcome of this study might be extrapolated to have implications for the seed and soil hypothesis proposed by Paget [34], there are a number of important distinctions. To date most studies examining metastasis-related genes have focused on factors intrinsic to the tumor cell (seed) or cells surrounding the tumor cell (soil) in the secondary site. Since normal tissues and autochthonous tumors have the same constitutional genetic information, genetic variation that modifies metastatic efficiency can

operate in the tumor cell, the stroma of the secondary site, in intermediate tissues during dissemination, or any combination of the above simultaneously. These data suggests that genetic heterogeneity influences both seed and soil between individuals, not somatic changes within the “seed” or the interaction of the “seed” with the different “soils” that exist within individuals. Thus, at present, this study addresses the role of genetic constitutional polymorphism on metastatic efficiency at the level of the whole organism, though individual contributions of the “seed” and “soil” are likely to be critical determinants in the efficiency of metastatic dissemination.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH and the National Cancer Institute.

References

1. Liotta, LA.; Stetler-Stevenson, WG. Principles of molecular cell biology of cancer: Cancer metastasis. In: DeVita, SHV.; Rosenberg, SA., editors. Cancer: Principles & Practice of Oncology. 4. Philadelphia, PA: J.B. Lippincott Co; 1993. p. 134-149.
2. Steeg PS. Metastasis suppressors alter the signal transduction of cancer cells. *Nat Rev Cancer* 2003;3:55–63. [PubMed: 12509767]
3. Frankel WN. Taking stock of complex trait genetics in mice. *Trends Genet* 1995;11:471–477. [PubMed: 8533162]
4. Schadt EE, Monks SA, Drake TA, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003;422:297–302. [PubMed: 12646919]
5. Chesler EJ, Lu L, Shou S, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 2005;37:233–42. [PubMed: 15711545]
6. Bystrykh L, Weersing E, Dontje B, et al. Uncovering regulatory pathways that affect hematopoietic stem cell function using ‘genetical genomics’. *Nat Genet* 2005;37:225–32. [PubMed: 15711547]
7. Hubner N, Wallace CA, Zimdahl H, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet* 2005;37:243–53. [PubMed: 15711544]
8. Lifsted T, Le Voyer T, Williams M, et al. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int J Cancer* 1998;77:640–4. [PubMed: 9679770]
9. Park YG, Zhao X, Lesueur F, et al. Sip1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1. *Nat Genet* 2005;37:1055–1062. [PubMed: 16142231]
10. Ramaswamy S, Ross KN, Lander ES, et al. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54. [PubMed: 12469122]
11. Hunter KW, Welch DR, Liu ET. Genetic background is an important determinant of metastatic potential. *Nat Genet* 2003;34:23–24. [PubMed: 12721549]
12. Qiu TH, Chandramouli GV, Hunter KW, et al. Global expression profiling identifies signatures of tumor virulence in MMTV-PyMT-transgenic mice: correlation to human disease. *Cancer Res* 2004;64:5973–81. [PubMed: 15342376]
13. Bernards R, Weinberg RA. A progression puzzle. *Nature* 2002;418:823. [PubMed: 12192390]
14. Bedigian HG, Copeland NG, Jenkins NA, et al. Emv-13 (Akv-3): a noninducible endogenous ecotropic provirus of AKR/J mice. *J Virol* 1983;46:490–7. [PubMed: 6188864]
15. Hunter KW, Broman KW, Voyer TL, et al. Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene Brms1. *Cancer Res* 2001;61:8866–72. [PubMed: 11751410]
16. Yang H, Rouse J, Lukes L, et al. Caffeine suppresses metastasis in a transgenic mouse model: a prototype molecule for prophylaxis of metastasis. *Clin Exp Metastasis* 2005;21:719–735. [PubMed: 16035617]
17. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;344:539–48. [PubMed: 11207349]

18. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35. [PubMed: 10700174]
19. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol* 2002;9:505–11. [PubMed: 12162889]
20. Dudoit S, Fridlyand F, Speed T. Comparison of discrimination methods for classification of tumors using DNA microarrays. *Journal of the American Statistical Association* 2002;97:77–87.
21. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 2003;19:2448–55. [PubMed: 14668230]
22. Simon R, Radmacher MD, Dobbin K, et al. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003;95:14–8. [PubMed: 12509396]
23. Wang J, Williams RW, Manly KF. WebQTL: web-based complex trait analysis. *Neuroinformatics* 2003;1:299–308. [PubMed: 15043217]
24. Lancaster M, Rouse J, Hunter K. Modifiers for mammary tumor latency, progression and metastasis are present on mouse chromosomes 7, 9 and 17. *Mamm Genome* 2005;16:120–126. [PubMed: 15859357]
25. Owerbach D, Hjorth JP. Inheritance of a parotid secretory protein in mice and its use in determining salivary amylase quantitative variants. *Genetics* 1980;95:129–41. [PubMed: 6159249]
26. Nielsen JT, Sick K. Genetic polymorphism of amylase isoenzymes in feral populations of the house mouse. *Hereditas* 1975;79:279–86. [PubMed: 1141004]
27. Kaplan RD, Chapman V, Ruddle FH. Electrophoretic variation of alpha-amylase in two inbred strains of *Mus musculus*. *J Hered* 1973;64:155–7. [PubMed: 4745489]
28. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6. [PubMed: 11823860]
29. Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 2003;9:416–23. [PubMed: 12640447]
30. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23–28. [PubMed: 959840]
31. Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977;197:893–5. [PubMed: 887927]
32. Hunter KW. Allelic diversity of the host genetic background as a determinant in tumor metastatic dissemination. *Cancer Letters* 2003;200:97–105. [PubMed: 14568162]
33. Malins DC, Gilman NK, Green VM, et al. Metastatic cancer DNA phenotype identified in normal tissues surrounding metastasizing prostate carcinomas. *Proc Natl Acad Sci U S A* 2004;101:11428–31. [PubMed: 15280533]
34. Paget S. The distribution of secondary growths in cancer of the breast. *Lancet* 1889;1:571–3.

Figure 1.

a) Quantitative PCR expression profiles comparing the relative expression of NZB or DBA normal mammary tissue versus FVB normal mammary tissue. **b)** Quantitative PCR expression profiles comparing the relative expression of NZB or DBA normal lung tissue versus FVB normal lung tissue. **c)** Quantitative PCR expression profile of *Pttg1* gene in whole blood.

Figure 2. Genome-wide eQTL mapping of genetic factors influencing *Coll1a1* steady state mRNA levels in AKXD mammary tumors. The mouse chromosomes are laid head-to-tail along the x-axis. The blue line represents the statistical probability of an eQTL residing at any position in the genome. The dashed green line represents the genome-wide suggestive threshold and the dashed blue line the significant threshold. The red line represents the additive effect of the eQTLs. Yellow histograms represent the most probable location of the eQTL, based on bootstrap analysis.

Figure 3. Unsupervised clustering of high metastatic FVB and low metastatic DBA F₁ animals based on the relative gene expression profiles of the mouse metastasis predictive gene expression set. The individual animals and their genotypes are indicated on the left.

Figure 4.
Relative expression level of salivary gland components in high and low metastatic genotypes.

Table 1

Mouse metastasis predictive probe sets and qRT-PCR Primers

Probe set	Gene symbol	qRT-PCR Primers	Reverse	Relative expression in high vs. low metastatic tumors
1 1449603_at	Al594671	AGACAAAGCAGAATGTTTATTTAAAGT	TCTCATTTCAACAGCAATTAGGT	d
2 1416077_at	Adm	CTTGTGTCCAGCCAAAGGT	CACACACGGAAACCAACAACAT	d
3 1427820_at		TTTTGTGGGTGTGAATTTGAAAG	AATCTGTGAGAGTGCTTCAT	d
4 1437568_at	Mmpl16	CGGACCAGCTTACAGAGTCA	TTGGCAAGTCACATCATCGT	d
5 1438824_at	Slc20a1	TGCGACCTTTCATGAGTCAC	TTCTTCTCTGAATGCCCTCT	d
6 1420352_at	4733401N09Rik	GGTGGATGACCACTGGGCTAC	GGCTGCTCATCTTCAGATCC	d
7 1450788_at	Saa1	GCGAGCTACACGTGACATGA	ATGCTGTGGGCTTCTTGGT	u
8 1427063_at	5330417C22Rik	AAGAGAGGACGTGGAGGAT	GTAGTGTGCCAGGTGAGGA	u
9 1424581_at	AW240854	CCCTACCAACAGAGCAGAC	ATTTGATGGTGGCCAAAAGA	u
10 1418078_at	Psme3	GCGAAGGTCAAACCCATAGA	TAAAAGCTGGGGGCTAGTG	u
11 1422651_at	Acrp30	TATCGCTCAGCGTTCAGTGT	TGAAGAGGCTCACCTTCACA	u
12 1423049_at	Tpm1	GTATGAAAGGTGGCCCGTA	CCTGAGCTCCAGTGACTTC	u
13 1418454_at	Mifap5-pending	CGACTTCTCCCTGTGAAA	CGTTACCAGCACCAAGTAAA	u
14 1449106_at	Gpx3	CACGTATCTGCCTCTTGAA	GGCTGGTGTGTGTAGTGTG	u
15 1418091_at	Car3	TGACCGAACTACGGGAGAG	CCCACCTAGAGGAAGAAAGG	u
16 1449434_at	Car3	TCGGGTATGGAGAGATATG	CTGGCTTATGGGTGTGTTT	u
17 1421001_at	Car6	AAGCCAAATTCCTGAACGTC	CAAGGAGACCCACATTCAG	u
18 1418475_at	Scnn1b	TAGGTTCAGATGCTGGTTCC	GAGGTCAACCAATGCCAACT	u
19 1418021_at	C4	GATCACCAAGTCTCTGCATT	TGAAGGCATCTCTCAATCC	u
20 1434676_at	AA516943	GGCTTGCTTATGGACTTCA	ACACTGCAAGGAAGCCATCT	u
21 1448485_at	Ggt1	GGGTACCCAGATCACCATAT	GGTGGCGTGAATCTGTGTAT	u
22 1448183_at	Hif1a	CAGCACGACTGATTTTCTCC	GCTGCCAAAATCCAGAGAGA	u
23 1449615_s_at	Hdlbp	ATCACTGGCACCAAGAAGG	ACGGTTAACACTGGGTGGAG	u
24 1452254_at	Mtmr9	AGCCAAGGTCAATGTCTGA	TACAGGGATTCGCACTAC	u
25 1426208_x_at	Plagl1	AAGTCTCACGGGAAGAAA	TCTTGGTGTGACGAGTGAGG	u
26 1434280_at	Xbp1	TGCAGCTTGCAGTTTGATT	GCCAGCAGCAACTATAACC	u
27 1425614_x_at	H2-D1	ATCATTTGCTGTCCGGTTGT	TCTGTACCAAGTCCACACC	u
28 1415997_at	Txnip	CCTAGTGATTGGCAGCAGGT	TGAGACTGTCTCCACATCGTC	u
29 1453839_at	1200009H11Rik	GTCTGGAAGGGAAGGCTGT	GTCCAAGCTACATGGGCAT	u
30 1416168_at	Serpinf1	AGCTCACCCAAAGTGGAAAC	GGTCCAGGATTTCTGCCTAT	u
31 1425896_a_at	Fbn1	GAACTGGGCGATAACCTGAA	CCATTCAAGAGTGGTGTCTCA	u
32 1427797_s_at				u
33 1460256_at	Car3	TCAAAGAGCCCATGACTGTG	TTCAAGGGCAGAGTTTCTTG	u
34 1418450_at	Islr	TAAACGCAATTGCCGTGCTC	CGTGGTACCTCTTTCAGA	u
35 1452544_x_at	H2-D1	CGGGCTCAAAGTGAGAACT	TGCAGATGTCCACCAATGTT	u
36 1416625_at	Serpinl1	TCTGCCATGAACACGAGAC	TACCAAGCCTGTCTCAACC	u
37 1450995_at	Folr1	GGCCCTGAGGACAAATTTACA	CAAGTTCGGGGAACACTCAT	u
38 1415812_at	Gsn	CATCGAGACAGATCCAGCAA	GGCACTGATTGGTGACATG	u
39 1416203_at	Aqp1	CAGAGTGCAGTCAAGCTCAGG	CTAGGGAACGAGACGAGTG	u
40 1450798_at	Tnxb	GTGTCTTTTCTGCCCGTGAT	CAGCTTCATTTCCTGTAAGG	u
41 1452336_at	BC027382	CTGTGGCAGTGAAGTCTCA	TTCCACCACCAATCAGTCTCC	u
42 1418018_at	Cpd	TGCTATGCTCCATCAAGTC	GGAGATATGCAAAACCGTGCT	u
43 1439364_a_at	Mmp2	CTATTCTGCAGCATTG	CAGACTTCTTCTCCAACTT	u
44 1437733_at	BC010348	CTGGTAACGGGTTTCTGCAT	ACCATGGATCACCTGACGTT	u

	Probe set	Gene symbol	qRT-PCR Primers	Reverse	Relative expression in high vs. low metastatic tumors
			Forward		
45	1451721_a_at	H2-Ab1	CTTGAAACAGCCCAATGTCG	AGGTCCAGTCCCATTCCTA	*
46	1427884_at	Col3a1	TTCTGCTGTGCTGGTGAAC	GGCATGCCAGTCAATAATTTC	*
47	1426550_at	B830021E24Rik	ACTCTGGATCCTGGAGCAAA	ACGTGCCCATCCTGAGTTAC	*
48	1416978_at	Fcgrt	CACAGCCTCTCCTGTTGGAC	GCAACAGGTCAACAGAGTCA	*
49	1416808_at	Nid1	TTTTTCTTGGGAACAGAGC	GGGTGCATGAAAGAGTCACA	*
50	1423407_a_at	Fbln2	ACCATCTCCTGACCATCAC	GGGCAAAAGTGGTGAAGAAG	*
51	1432344_a_at	Aplp2	TTCAAAGTGTCTTGTGGTGAA	GACATGCCCTCTTGACTAACG	*
52	1417963_at	Pltp	AAATCAGTCTGCGCTGGAGT	TTGGCAAGTGGAGTTCAG	*
53	1432466_a_at	Apoe	CAGAAAGCCTAGCTGTGA	AGCGCCTAGGTGGCAGT	*
54	1420924_at	Timp2	ACATGTCCCTCTGTTTCTGC	TCCATTTCGCTGAAAGTCTGTG	*
55	1422973_a_at	Thrsp	ACGGAGCCCTGATCTCTAT	GGCTTCTAGGTCCAGCTCCT	*
56	1434279_at	Xbp1	GGTTGTGCCCAAAACAATTC	TGGAAACAGGGGAAACTGTGT	*
57	1415784_at	Vps35	ATGCCCTGGACAAAAACAAGC	ATCTGGCCACCTTCTTCTCT	*
58	1448792_a_at	Cyp2f2	AAGACGCCTCAGGAGTTCAA	ATGTCTCTCAGGATCCACCAG	u

* = primer pairs that specifically amplify cDNA and were used in clustering

d = down regulated in the high metastatic tumors when compared to the low metastatic tumors

u = up regulated in the high metastatic tumors when compared to the low metastatic tumors

Table 2

Performance of salivary peptide metastasis prognostic tests

Classifier Method	Peptides in classifier	Performance during cross validation	Experimental samples Correctly Classified
Compound Covariate	5	78% (21/27)	79% (11/14)
Linear Discriminant	5	81% (22/27)	79% (11/14)