Gene expression profiling of NMU-induced rat mammary tumors: cross species comparison with human breast cancer

Maren M. Chan1, Xin Lu2, Faisal M. Merchant3, J. Dirk Iglehart4 and Penelope L. Miron1,5,*

1Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02115, USA, 2Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA, 3Department of Medicine, Massachusetts General Hospital, Boston, MA 02115, USA, 4Department of Surgery, Brigham and Women’s Hospital, Boston, MA 02115, USA and 5Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

*To whom correspondence should be addressed: Room 822, Smith Building, Boston, MA 02115, USA.
Tel: +1 617 632 5632; Fax: +1 617 632 3709.
Email: penelope_miron@dfci.harvard.edu

**Abbreviations:** DCIS, ductal carcinoma in situ; DMBA, 7,12-dimethylbenz[a]anthracene; KNN analysis, k-nearest neighbor analysis; MDS, Multi-Dimensional Scaling; NMU, N-methyl-N-nitrosourea; PCA, Principle Component Analysis.

**Introduction**

Breast cancer is the most frequent malignancy among women worldwide (1). It is a highly heterogeneous disease represented by tumors that have a diverse natural history, complex histology and a variable response to therapy. Although the molecular events that trigger breast cancer progression, including its initiation, promotion and progression to a fully malignant state, are not fully understood, many genetic alterations have been described (2-4). These changes primarily encompass mutations, chromosomal amplifications and deletions involving oncogenes and tumor suppressor genes. Altered expression of several key genes in breast cancer also form the basis for clinically useful subdivisions. These genes include the estrogen receptor and erbB2 (HER2/neu). A clearer understanding of the molecular events underlying this complex disease has been accomplished using gene expression profiling. This global approach provides evidence of the biologic diversity of the disease and has led to the classification of breast tumors into informative subtypes (5-7).

The induction of rodent mammary tumors following the administration of N-methyl-N-nitrosourea (NMU) or 7,12-dimethylbenz[a]anthracene (DMBA) is a widely used experimental animal model for investigating breast cancer in women (8-11). These carcinogen-induced tumors arise from terminal end buds (12,13), an analogous structure to the terminal duct lobular unit in humans, which is the proposed site of origin of ductal carcinoma in situ (DCIS) (9,14). Substantial evidence suggests that this animal model mimics human breast cancer. The rat tumor’s histopathology, origination from mammary ductal epithelial cells, and dependency on ovarian hormones for tumor development all correlate with human breast cancer (10,15-18). Comparative studies have also shown that chemically-induced mammary carcinomas, like their human counterparts, have altered expression of TGFβ, erbB2 and cyclin D1 (11). Furthermore, some human and most rat mammary tumors express estrogen and progesterone receptors (19,20).

Unlike human breast cancer, the majority of carcinogen-induced rat tumors are thought to have carcinogen-specific mutations in the H-ras oncogene (21). In particular an H-ras mutation at codon 12 has been identified in NMU-induced tumors, while DMBA-induced tumors harbor a mutation at codon 61. In the rat model these mutations have been implicated in tumor initiation based upon observations of ras oncogene activation preceding the onset of neoplasia (21-23). Nonetheless, this model has been used extensively to evaluate preventative and therapeutic agents for human breast cancer despite the near absence of ras mutations in human breast cancer. On a more global scale, little is known about the similarities in gene expression between human breast cancer and carcinogen-induced rat tumors.

While microarray technology has found powerful applications in human cancer class discovery and class prediction, its use in animal models has lagged behind. Recent reports describe the use of cDNA microarrays to profile genomic alterations in rat mammary gland cancer models. These studies have defined carcinogen-specific gene expression profiles (24,25) and described the modulation of differentially expressed genes by various chemopreventative agents (26,27).
Comparing animal model gene expression profiles with human cancer is challenging. Cross-species hybridization using non-human RNA hybridized onto a human array platform has been attempted with some success (28–33). Other investigators have directly compared lists of genes independently obtained from different platforms (34). To date, there are no published reports that have merged raw microarray data derived from different species prior to further statistical analysis.

To characterize the rat model of NMU-induced mammary carcinoma from a more global molecular perspective and to devise a technique to compare animal gene expression profiles with human microarray studies, we undertook a cross-species comparison between gene expression profiles derived from NMU-induced rat mammary tumors and human breast carcinomas. We analyzed NMU-induced rat mammary carcinomas and normal rat mammary tissue using rat Affymetrix RAE-230A microarrays. Tumors were characterized histopathologically using established classification schemes in order to correlate pathology with molecular signatures. The in situ nature and hormone status of the tumors were investigated in addition to the frequency and significance of H-ras mutation on gene expression. Analysis of over 15 000 genes demonstrated that the rat tumors are remarkably similar in their expression profiles, but differed greatly from the normal rat mammary gland. We compared rat expression profiles with those obtained from a human data set by merging the raw microarray data. Although >90% of the genes in the genome have homologs between human and rat, most of the genes have a different baseline and scale of variation between the two species, which make a ‘comparison’ between global gene expression profiles unfeasible before a standardization process is applied. We used a simple and intuitive approach to standardize rat and human expression data in order to achieve a global cross-species and cross-platform analysis, and validated the capacity of 2305 rat orthologues to recapitulate the molecular classification of human tumors. Moreover, the data suggest that rat mammary tumors share molecular signatures with low to intermediate grade, estrogen receptor positive human breast cancer.

Materials and methods

Animals

This study was conducted in accordance with the Animal Welfare Act, the Guide for the Care and Use of Laboratory Animals, and the policies of the Dana-Farber Cancer Institute. Female Wistar-Furth rats (Harlan, Indianapolis, IN) were given a single intraperitoneal injection of NMU (Ash Stevens, Detroit, MN) (35 mg/kg body wt, n = 16) at 50 days of age. A subsequent group was injected with a higher dose (50 mg/kg body wt, n = 16); these animals were added to the study in order to increase the final sample size, and a higher dose of NMU was used to decrease the latency period before the detection of mammary tumors. NMU was dissolved immediately before use in 0.9% NaCl acidified to pH 4 with acetic acid. Rats were weighed weekly and palpated daily for mammary tumors. When the tumors reached 2 cm, they were harvested, sectioned and frozen in liquid nitrogen or O.C.T. compound (Tissue-Tek, Torrance, CA), or fixed in 10% Neutral Buffered Formalin (Fisher Diagnostic, Middletown, VA). The lungs, liver, spleen, kidney and lymph nodes were removed to examine if the tumours had metastasized. Normal mammary glands were obtained from virgin female Wistar-Furth rats (n = 11) ~11 weeks old—the average age of tumor-bearing animals at excision.

Immunohistochemistry

For pancytokeratin (AE3, Roche, Indianapolis, IN) and vimentin (Clone V9, Dako, Carpinteria, CA), formalin-fixed paraffin-embedded rat tumor sections were deparaffinized in xylene and rehydrated through graded ethanol. Antigen retrieval was performed using 10 mM sodium citrate buffer (pH 6.4) at 93°C for 30 min. Primary antibody (AE3 1:400, vimentin 1:200) was applied overnight at 4°C. Biotinylated goat anti-mouse IgG antibody at 1:250 (BA-9200; Vector Labs, Burlingame, CA) was applied for 30 min at room temperature, followed by 45 min in avidin-biotin horseradish peroxidase complex (PK-6100, Vectastain Elite ABC kit, Vector Labs). 3,3′-Diaminobenzidine substrate (D-5905; Sigma-Aldrich, St Louis, MO) was applied and sections were counterstained with hematoxylin. For collagen IV (RDI-PRO10760, Research Diagnostics Flanders, NJ) and laminin (L9393, Sigma-Aldrich) frozen sections (n = 5) were fixed in acetone for 5 min. Primary antibody was applied for 3 h (collagen IV 1:200, laminin 1:300) at room temperature. Staining proceeded as described above.

For the estrogen receptor (ER, Clone 1D5, Dako) and p63 (Clone 4A4, Dako), formalin-fixed, paraffin-embedded tissue sections were baked at 60°C for 2 h and deparaffinized and rehydrated through graded alcohol concentrations. Antigen retrieval was performed using citrate buffer at pH 6.1 (S1700, Target Retrieval solution; Dako) and steam (30 min for p63, 40 min for ER). Immunohistochemical staining was performed using the Universal DAKO EnVision System, Peroxidase (K1390, Dako). Sections were incubated in primary antibody 1:50 dilution at room temperature for 30–40 min followed by HRP labeled polymer. Detection was carried out using DAB and Mayer’s hematoxylin following by bluing in PBS (pH 7.4).

H-Ras mutation detection

DNA was extracted from tumors (n = 24) and normal mammary glands (n = 3) using ~25 μg of tissue digested overnight with proteinase K followed by heat inactivation (DNeasy Tissue Kit; Qiagen, Valencia, CA). DNA from normal rat liver was used as a negative control and pNMU-1 plasmid (41049; ATTIC, Manassas, VA) as a positive control. The method used to detect a G to A transition in the second nucleotide of codon 12 of the Hras1 gene was based upon a PCR technique described by Kumar and Barbacid (23) and modified by Aldaz et al. (35) for detection on etidium bromide-stained gels. In NMU-induced tumors, codon 12 G35→A35 is the mutation in H-ras reported in the literature. In brief, DNA samples (1 μg) were digested overnight with 2 U MluI (New England Biolabs, Beverly, MA) to enrich for the mutant allele. The mutant target sequence was amplified using the iCyler iQ PCR detection system (BioRad, Hercules, CA) and a reaction mixture of 25 μl Supermix (Biorad), 250 nM forward and reverse primers, 4 μl of DNA and water to make a final reaction volume of 50 μl. The primers used were 5′-TTGCCCT-ACTCATGGTGGGAG-3′ and 3′-CAGAAGGTTCCTCAGCGGGA-5′ (Qiagen), with the 3′ mismatched XmnI recognition site underlined. The thermal cycling program consisted of polymerase activation at 95°C for 1.5 min followed by 40 cycles of 20 s at 94°C (denaturation), 20 s at 55°C (annealing), and 12 s at 72°C (elongation) and a final extension at 73°C for 7 min. PCR products (10 μl) were digested with 20 U of XmnI (New England Biolabs) and electrophoresed in 3% agarose gels stained with etidium bromide. The XmnI digestion site, together with the H-ras mutation, produces a digested PCR product of 172 bp, while the undigested normal allele is 190 bp.

RNA extraction and Affymetrix RAE-230A chip hybridization

RNA was extracted from frozen rat mammary tumors (n = 24) and normal mammary glands (n = 11) using 30–40 μg of bulk tissue with the RNeasy Mini Kit (74104; Qiagen). The normal rat mammary gland has a significant amount of adipose tissue; therefore an additional chloroform purification and ethanol precipitation step was performed prior to RNA purification steps. Production of biotinylated cRNA and hybridization to the rat RAE-230A oligonucleotide microarray (Affymetrix, http://www.affymetrix.com) was performed at the Dana Farber Cancer Center Microarray Core Facility. An overview of the gene profiling procedure is available at http://chdp.dfci.harvard.edu/lab/services.php.

Briefly, a quality control analysis is performed on the RNA prior to gene profiling. RNA contamination and concentration is assessed by spectrophotometric analysis, and RNA integrity is measured using a microfluidics instrument (Agilent Technologies, Palo Alto, CA). cRNA is synthesized using biotinylated CTP and UTP and hybridized to the microarray chip overnight at 45°C in the presence of several biotinylated control oligonucleotides. The chips are then washed in a fluids instrument to remove unbound cRNA. Bound cRNA is fluorescently labeled using phycoerythrin-conjugated streptavidin (SAPE, Bioscience, San Diego, CA). Each bound cRNA is excited using a confocal laser scanner and the intensity measurements transferred to an Oracle database for further statistical analysis.

Microarray data and statistical analysis

Raw expression values from the Affymetrix RAE-230A chip were analyzed using the DNA-Chip Analyzer (dChip 1.3) software (36). Raw intensity values were normalized based on the rank invariant probe sets, and expression values were then calculated by model based expression (36). To confirm the reproducibility of bulk tumor sampling, we performed unsupervised hierarchical
clustering, from duplicate sections taken from the same tumor \((n = 3)\). Duplicates always clustered together within the resulting dendogram (data not shown). Thereafter, a set of highly variable genes was selected with a coefficient of variation between 0.5 and 10 and a presence call in at least 20% of all samples. The resulting 813 probe sets were used for unsupervised hierarchical clustering of rat tumors and normal glands.

Supervised analysis was performed to find genes that differed significantly between normal rat mammary glands \((n = 11)\) and tumors \((n = 24)\). By two-sample \(t\)-statistics using fully annotated genes, 3003 significant probe sets were obtained with an adjusted \(P\)-value, under false discovery rate control of \(<0.01\) (37,38). From this probe list, 613 probe sets were found to have a fold change of \(>2.0\) or \(<0.5\) when comparing normal glands with mammary tumors.

Cross-species gene expression profile comparison

Normalized expression data from human breast carcinomas \((n = 89)\) and normal human tissue \((n = 7)\) hybridized to the HG-U95av2 microarray chip \((\text{Affymetrix})\) was previously obtained in our lab (39,40); This data set was also normalized by dChip algorithms in a manner identical to the normalization procedures used for the rat expression data. Human orthologs of rat genes on the RAE 230A microarray gene chip were found using homologous Locus Link gene information from the GeneBank database. The homology information of the Locus Link annotations is provided by NCBI. These annotations are based on a combination of sequence similarity and conserved synteny (a stretch of chromosome in which the gene order is conserved across species).

The human-orthologs may vary as NCBI and other contributing databases are updated, but since many of the contributing databases from which the Locus Link information is collected are curated databases, and the Locus Link records are also reviewed by NCBI staff, we believe that the term Locus Link is a solid and reliable source of information about the orthologous genes.

Each human ortholog was then matched with its probe on the U95av2 Affymetrix chip. When more than one probe set corresponded to the same locus link, we arbitrarily chose the first probe set to appear on the chip. The probes from the human orthologs of the rat genes were used for merging datasets.

To determine whether the rat and human orthologs depict the heterogeneity of gene expression in human breast cancer, unsupervised hierarchical clustering of the 89 human carcinomas and 7 human breast tissue samples was performed using highly variable orthologs with the filtering criteria that the coefficient of variation was specified between 0.5 and 10 with a presence call in at least 20% of samples. This provided 628 genes for clustering analysis.

Due to intrinsic species differences and also the probe sequence sensitivities, most of the human and rat homolog genes have a different mean and scale of variation across samples in this study. Consequently, the global gene expression profiles are dramatically different between human and rat (data not shown), which means that any cross-species comparison is not feasible before a standardization procedure is applied to remove this systematic difference. The gene expression data were standardized within the normalized human and rat datasets separately as follows. First, low expression values were truncated to value 1 and the entire dataset was log transformed. The mean probe intensity across all samples was subtracted from the probe expression level of each sample and divided by the standard deviation. Standardized human and rat datasets were then combined for further statistical analyses. This standardization procedure equalizes the baseline and variation scale of genes across all samples. This provided 2305 unique rat-human orthologs. Principle Component Analysis (PCA) was used to determine the extent of variability in gene expression and to compare the degree of heterogeneity in rat tumors and human breast cancer (42,43). In order to identify the subclass of human breast cancer with which NMU-induced tumors share similar molecular expression profiles, we used the voting form of the k-nearest neighbor (KNN) analysis to assign a human breast cancer characteristic to each rat tumor. The estrogen receptor status and tumor histologic grade were the two characteristics assigned separately to each rat tumor. Since the sample size is limited, our conventional approach to randomly break any ties that occurred when using the KNN algorithm may skew the ultimate assignments in the classification result. In order to address this, we also assigned a 0.5 count to both of the possible assignments when a tie was created. Use of this alternative 'equal assignment' approach did not significantly change the data and led to the same conclusions reported in the results. We, therefore, chose to use the first method in our analysis.

Results

Histopathologic characterization

A total of 24 NMU-induced tumors were examined and classified histologically according to the criteria outlined by Russo and Russo (17) for chemically induced rat mammary tumors. The majority of tumors were identified as \textit{in situ} ductal carcinomas with papillary and/or cribriform characteristics (15/24) (Figure 1A, panel a). Three tumors were tubular carcinomas and one was a benign fibroadenoma. Five of the tumors were found to have a mixed phenotype, in which the tumor had areas similar to tubular carcinomas (2/24) or fibroadenomas (3/24) constituting at least 50% of the cross-sectional area. These data are summarized in Table I. The normal mammary gland consisted primarily of adipose tissue with rare mammary ducts dispersed throughout (Figure 1A, panel b).

We further characterized the tumors by a number of immunohistochemical studies. Using stains for laminin, (Figure 1A, panel c) and collagen IV (data not shown), we demonstrated a continuous basement membrane surrounding nests of neoplastic cells. Cells within the basement membrane stained intensely for epithelial-type cytokeratins, and were negative for vimentin (Figure 1B, panel d). Cells outside the basement membrane displayed the reverse immunophenotype; these cells were uniformly vimentin-positive and lacked epithelial markers (Figure 1B, panel e). Most duct-like structures demonstrated a basal cell layer with discontinuous staining of p63-positive cells around nests of luminal epithelial cells (Figure 1B, panel f). This p63 staining pattern is similar to that seen in human DCIS (46) and is evidence of the non-invasive behavior of the rat tumor. There were also isolated areas with densely populated p63 cells that were not surrounding a luminal cell layer. As a final test for the absence of invasion, animals were necropsied at the time of killing and peripheral organs examined. We found no evidence of metastatic foci in distant organs. When stained for hormone receptors, all tumors were found to have at least 10% ER-positive cells (Figure 1B, panel g). Therefore, these NMU-induced tumors could be described as non-metastatic, non-invasive, and ER-positive with a composition of differentiated epithelial and myoepithelial cell lineages.

Gene expression profiling of rat mammary tumors

As a first step towards the molecular characterization of these tumors, we compared the genomic profiles of the
24 NMU-induced tumors with those of 11 normal rat mammary glands. The overall similarity of gene expression patterns between normal mammary glands and NMU-induced tumors was evaluated using unsupervised hierarchical clustering and a total of 813 genes, which satisfied criteria outlined in the Materials and methods section. Two distinct clusters clearly distinguished the gene expression profile of normal glands from that of tumors (Figure 2). Of the genes differentially expressed in NMU-induced tumors, the majority was down-regulated in tumor tissue (76%). Most rat tumors (22/24) populated a uniform single large cluster with shallow vertical branches, suggesting a high degree of similarity within the tumors. We are comparing tumors composed primarily of epithelial cells with normal glands that consist of adipose and stromal components. Therefore, it is likely that the proportion of epithelial and stromal cells may contribute to the observed expression differences between carcinomas and normal tissue. However, in support of the notion that not all of the changes that we identified are due to changes in cellular composition, Wang et al. (26) demonstrated that the differential expression of five selected genes using both bulk tissue and isolated epithelial cells derived from rat mammary normal and tumor tissue resulted in identical differential expression patterns. This result indicated that the lower percentage of rat mammary epithelial cells in normal mammary glands versus those from mammary tumors did not affect the expression data obtained in their study.

To gain insight into the major biological processes contributing to tumorigenesis in this model and its correlation with human breast cancer, we compared normal and tumor tissue, and classified a subset of genes with >2-fold and <0.5-fold expression changes (613 genes) according to a predominant functional term. This was accomplished by using data files from the NCBI (National Cancer for Biotechnology Information) and Gene Ontology® Consortium in dChip together with the PubMed database. We found that the biological functions of the genes were diverse and fell into twelve functional classifications for both upregulated and downregulated genes (Table II). Predominant categories included metabolism, signal transduction, growth and cell cycle, immune response and poorly characterized genes (miscellaneous). A complete list of genes and their functional classification can be found in supplementary data, Tables I and II.

The presence of an H-ras mutation fails to confer a unique molecular profile

A mutation of the H-ras gene in codon 12 has been implicated in the pathogenesis of NMU-induced tumors (21,22). Other mutations in H-ras and mutations in K-ras have not been detected in NMU-induced rat tumors (47), while mutations in N-ras have not been investigated. We tested whether tumors bearing the G35-A35 mutation in codon 12 possess a molecular signature different from that of tumors without the mutation. Twenty-five percent of tumors were found to have H-ras mutations by PCR-based, mobility shift differences. In the unsupervised hierarchical clustering dendogram, tumors with the H-ras mutation do not segregate or show a higher degree of similarity in the global gene expression pattern compared with wild-type tumors (Figure 2). In addition, supervised analysis using a two-sampled t-test under false discovery rate control failed to detect any differentially expressed genes. Therefore, the presence of an H-ras mutation does not appear to appreciably modify global gene expression in NMU-induced tumors.

Cross-species analysis merging raw data from human and rat microarrays

To directly examine the relationship of gene expression in NMU-induced mammary tumors and human breast cancer, we merged the microarray hybridization intensity measurements obtained from rat tumors and from a data set derived from human breast carcinomas (39,40). Rat and human orthologs were identified using available information in the GenBank database. Of the 15 866 probe sets on the RAE-230A gene chip, 4008 rat probe sets were found to have a known human ortholog. After removing probe sets corresponding to the same gene, 2305 orthologous probe sets were identified.

Because the number of orthologous genes represents only 20% of the genes present on the U95av Affymetrix gene chip, we questioned whether this subset of genes could adequately depict the heterogeneity of gene expression in human breast cancer. In a previous study by Wang et al. (40), 89 breast cancers and 7 normal breast tissue samples formed four distinct subclasses after hierarchical clustering using 672 probe sets generated under stringent filtering criteria. We used 628 highly variable orthologs to cluster the same cohort of human tissue samples and recapitulated the predominant hierarchical relationships between breast cancer subclasses. For instance, most tissue samples within a previously defined subgroup, as indicated by the colored bars below the clustering tree, continue to lie adjacent to each other (Figure 3). The clustering of normal human breast tissue within the previously defined IIA subclass (Figure 3, pink bars), which is predominately populated by low-grade ER-positive tumors, persists when clustering is performed using only unique orthologous genes. Therefore, this subset of orthologous probe sets contains enough information to recapitulate the predominant hierarchical relationships between breast cancer subclasses. These 2305 unique orthologs were used in all subsequent cross-species analysis. Similar results were obtained using all 4008 rat probes, i.e. without removing redundant probes (data not shown).

Cross-platform analyses of human microarray datasets is evolving (48,49) and this study adds complexity by comparing expression profiles of different species. We used several techniques to compare rat and human tumors. In order to see whether human cancer and rat mammary tumors have comparable gene expression profiles, an overall survey of gene
expression patterns across all rat and human samples was performed using MDS. MDS is an unsupervised method of gene expression analysis, which projects the spatial relationship between samples into low dimensional space; the distance between samples reflects their approximate degree of correlation measured by Pearson correlation coefficient (42). MDS was used to visualize the similarities between rat and human tissues (Figure 4A). Several features of gene expression patterns are evident. Normal human breast tissue and normal rat mammary glands are remarkably different. Human tumors and normal breast tissue demonstrate great variation in their gene expression patterns. Gene expression in the rat mammary tumors intersects the gene space of human tumors but not that of the normal rat mammary gland. These observations

Table 1. Histologic classification of 24 NMU-induced tumors

<table>
<thead>
<tr>
<th>Tumor classification</th>
<th>Number of tumors</th>
<th>Tumor identification number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papillary cribriform DCIS</td>
<td>15/24</td>
<td>2,4,6,8,9,14,19,20,21,22,25L, 25R, 26,28,29 (1)</td>
</tr>
<tr>
<td>Tubular</td>
<td>3/24</td>
<td>13,15,27</td>
</tr>
<tr>
<td>Fibroadenoma</td>
<td>1/24</td>
<td>29 (2)</td>
</tr>
<tr>
<td>Mixed papillary cribriform/tubular&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2/24</td>
<td>17,32</td>
</tr>
<tr>
<td>Mixed papillary cribriform/fibroadenoma&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3/24</td>
<td>10,11,18</td>
</tr>
</tbody>
</table>

<sup>a</sup>According to Russo and Russo (17).
<sup>b</sup>Referred to in Figure 2.
<sup>c</sup>At least 50% of tumor cross-section with tubular or fibroadenoma histology.

Fig. 2. Unsupervised hierarchical clustering of 24 rat NMU-induced tumors and 11 normal rat mammary glands using 813 probes sets (genes) with high variation across the sample set. Rat tissue samples of similar gene expression are clustered at the top of the dendogram. Genes are ordered by similarity to the left of the dendogram. Relative gene expression levels are color-coded as follows; white, mean expression level; red, expression above the mean; blue, expression below the mean. The dendogram shows two distinct clusters representing rat normal mammary glands (green bars) and NMU-induced tumors (purple bars). The red " identifies tumors with H-ras mutations.
suggest that despite differences in gene expression between these heterogeneous tissues, human breast carcinomas and NMU-induced tumors appear to be closely related.

Human breast cancer is a heterogeneous disease with regard to natural history, treatment response, and gene expression. In contrast, rat tumors appear a more homogeneous disease based on the high degree of intra-tumor similarity observed using both unsupervised hierarchical clustering (Figure 2) and MDS analysis (Figure 4A). To further corroborate this observation, we used PCA, an analysis tool that describes the variability within a sample set (42). By plotting the gene expression profiles for rat and human carcinomas using the first two principle components, Figure 4B shows that human tumors occupy a larger portion of the available 2D space suggesting high variability in gene expression. The first two principal components were used because they captured 33% of the total variation, while the first three principal components captured 38%, a difference of only 5%. In contrast to the pattern displayed by the human tumors, the rat tumors form a tight cluster in the center of the diagram. PCA supports the notion that NMU-induced mammary tumors are indeed more homogeneous than human carcinomas at the level of gene expression.

Though NMU-induced tumors and human breast cancers have similar gene expression profiles, rat tumors appear more homogeneous than human breast cancer. This implies that the rat model reflects a particular subclass of human breast cancer, although we cannot rule out the possibility that the smaller within-group variance in rat tumors, when compared with human tumors, may reflect the greater underlying degree of genetic homogeneity in the rat strain used than in the human population from which the tumor samples were drawn. Nonetheless, we sought the genetic correlation of NMU-induced tumors with a specific class of human breast cancer. Several studies have used microarray technology to define the molecular classification of human breast cancer and a common finding is the predominant partitioning of breast cancer based upon estrogen receptor status (5,6,50,51) and tumor grade (40,50,52). Therefore, using KNN analysis (\(K = 5\)), the estrogen receptor status and grade of the five human breast cancers with gene expression profiles most similar to a single NMU-induced tumor was determined. This analysis was repeated for each of the 24 rat tumors. According to the voting among the 5 nearest neighbors, all the NMU-induced tumors were most similar to ER-positive human carcinomas (\(P\)-value = 0.0454) and the majority had gene expression profiles similar to low to intermediate grade (\(P\)-value = 0.0002) human breast carcinomas. Specifically, the KNN analysis showed 20 out of 24 tumors to be Grade II with a \(P\)-value of \(1.205 \times 10^{-13}\); Out of 24 tumors 3 were Grade I with a \(P\)-value of 0.5680, and 1 out of 24 tumors were Grade III with a \(P\)-value of 0.0082. These data are represented as a bar graph (shown in Figure 4C) and illustrate the resultant distribution of human tumor subtypes that were found to be most similar to those of the rat tumors. This supervised analysis suggests that NMU-induced tumors share molecular signatures with lower grade ER-positive human breast cancer.

Discussion

In the present study, we used a combination of immunohistochemical analyses and gene expression profiling to

<table>
<thead>
<tr>
<th>Functional classification</th>
<th>Upregulated genes (%)</th>
<th>Downregulated genes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cell structure</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Immune response</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Transport</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Extra cellular matrix</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Matrix metalloproteinases</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Hormone-related</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>20</td>
<td>5</td>
</tr>
</tbody>
</table>

A total of 178 genes were upregulated and a total of 435 genes were downregulated in NMU-induced tumors. The distribution of functional classifications for each set of genes is expressed as a percent of the total significant genes.

Fig. 3. Cluster analysis of human breast cancers. Simplified dendograms from unsupervised hierarchical clustering of human carcinomas and normal breast tissue using: (A) the highly variable 672 genes from Wang et al. (40) and (B) a set of highly variable rat and human orthologs (628 genes) are compared. Wang et al. described sub-clusters of human invasive breast cancer defined by gene expression. Sub-clusters are depicted in the colored bars below the dendograms; grey is IA, blue is IB, pink is IIA, and green is IIB. Tumor samples are labeled with a ‘B’ and normal breast tissue with ‘N’.

Microarray analysis of NMU-induced rat mammary tumors

Table II. Functional classification of genes discriminating the normal rat mammary gland from NMU-induced tumors

Downloaded from https://academic.oup.com/carcin/article-abstract/26/8/1343/2390919 by guest on 03 February 2019
characterize NMU-induced mammary tumors in the rat. By immunohistochemical staining, the majority of tumors were non-invasive, non-metastatic and ER-positive ductal carcinomas. The finding of focal proliferation of p63 positive cells not confined to the basal cell layer suggested possible mixed myoepithelial/epithelial tumors. Genes differentially expressed between the rat normal gland and tumors were identified and are available in the supplementary data. We found that H-ras mutations are restricted to a minority of tumors and do not confer a unique molecular signature, implicating alternate or redundant pathways in the NMU model of tumorigenesis. A cross-species microarray analysis identified a set of highly variable orthologs that were able to recapitulate previously defined human subclusters. By comparing gene expression profiles with those of human carcinomas, NMU-induced mammary tumors are more homogeneous than their human counterparts. Furthermore, the gene expression profiles of NMU tumors are most similar to ER-positive, low to intermediate grade breast carcinomas.

Several investigators have addressed the histologic classification of carcinogen-induced tumors (17,53,54). Consistent with prior studies, our NMU-induced tumors are non-metastatic, in situ, ER-positive, epithelial cancers; similar to human DCIS. A unique aspect of this study is the description of p63 expression in rat mammary tumors. p63 is a marker of human myoepithelial cells and a member of the p53 gene family (46,55,56). In humans, p63 stains the basal cell layer of myoepithelial cells in normal ducts. In human DCIS, the myoepithelial layer stains discontinuously for p63, while staining is frequently absent in invasive lesions (46,55,56). Similar to human DCIS, the basal cell layer of ducts in NMU-induced tumors stains positive for p63, suggesting a lesion analogous to human DCIS. However, areas of p63 positive cells not directly lining ducts were observed. Though p63 is a relatively new immunohistochemical marker of myoepithelial cells, recent studies have demonstrated p63 expression in cells of myoepithelial tumors including benign and malignant adenomyoepitheliomas, epithelial-myoepithelial carcinomas, adenoid-cystic carcinomas, and sclerosing papillomas (46,57).

Fig. 4. A comparison between human and rat gene expression profiles using various statistical analyses. (A) Multi-dimensional scaling analysis (MDS) of rat and human normal and tumor tissue. MDS using the 2305 unique orthologs shows large disparity in gene expression between normal rat mammary glands and normal human breast tissue. A higher degree of similarity in gene expression between the rat and human tumors is demonstrated by the closer proximity of both species’ tumors compared with their normal glands. Red circles are normal rat mammary glands. Black circles are normal human breast tissue. Red stars are NMU-induced tumors. Black stars are human breast carcinomas. (B) Principle Component Analysis (PCA) of rat and human tumors. Using the same 2305 orthologs referred to in Figure 4A, the variability of gene expression in rat tumors and human breast cancer are compared by PCA. The x- and y-axes show the range of variability of the first two principle components (PC1 and PC2) across the entire sample set. The first two principle components for each NMU-induced tumors (red stars) and each human carcinomas (black stars) are plotted. (C) K-nearest neighbor (KNN) analysis of NMU-induced tumors. The estrogen receptor status and grade of the five human breast cancers with gene expression profiles most similar to a single NMU-induced tumor was determined by KNN analysis (see Materials and methods) and the combined data represented as a bar graph. One hundred percent of rat carcinogen-induced tumors had gene expression profiles similar to ER-positive breast cancer. The majority of rat tumors had gene expression profiles similar to intermediate grade (83%, Grade II) human breast carcinomas. I = Grade I (low), II = Grade II (intermediate), III = Grade III (high).
Our findings demonstrate focal areas of myoepithelial cells and suggest that NMU-induced tumors may constitute a mixed cellular lineage.

Previous studies of gene expression profiling in carcinogen-induced animal models of mammary tumorigenesis have been performed on a limited number of tumors and genes. Wang et al. (26) used cDNA library screening and cDNA microarray analyses to compare gene expression in six NMU-induced rat mammary tumors and three normal glands. One hundred and sixty differentially expressed genes were identified using a rat ovary cDNA library in which only 10% of the library was screened, and several sequences showed limited homology to rat, mouse or human genes. The cDNA microarray analysis found 41 differentially expressed genes from an array containing 588 genes. Shan et al. (25) found 68 known genes to be differentially expressed between PhIP (2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine, n = 5) and DMBA-induced tumors (n = 4) combined with normal mammary glands using a more extensive cDNA microarray with 6900 cDNA clones. Several of these genes were also differentially expressed in our study including stearyl-CoA desaturase 2, cyclin D1, lipoprotein lipase, aquaporin 1 and alpha 1 type V collagen.

In our study, 613 genes were identified showing altered expression in NMU-induced tumors compared with the normal rat mammary gland from an oligonucleotide microarray containing 15 866 probe sets. Significantly larger portions of these genes were down regulated (435/613) in NMU-induced tumors. Clustering of known genes with reported functions revealed a predominance of metabolism-related genes and genes involved in cell cycle, signal transduction, transport, and immune response. A smaller portion of genes were involved in other biological processes involved in cell adhesion, extracellular matrix proteins, and transcription factors. Collectively, this wide distribution of functionally diverse genes supports the hypothesis that global genomic changes are more likely associated with tumor development than the alteration of a single genetic expression event. Conversely, the data may also be a reflection of cellular composition, as we are comparing epithelial-rich tumors with normal glands that contain more adipose and stromal tissue. The substantial metabolic differences between epithelial cells and adipocytes likely account for the association with the identified ‘metabolic’ genes.

A mutation in the H-ras gene is implicated in the pathogenesis of NMU-induced tumors (21,22); however, the role of the H-ras mutation as a predominant oncogenic event driving tumorigenesis has been questioned. Reports of H-ras mutations in preneoplastic lesions and unexposed mammary glands suggest that the mutation is not necessary or sufficient for oncogenic transformation (58–60). In transgenic rats carrying the c-Ha-ras oncogene, susceptibility to carcinogen-induced tumorigenesis is independent of transgene or endogenous H-ras mutations (61,62). In this study, only 25% of tumors harbored an H-ras mutation. The expression of the H-ras gene by microarray analysis was not altered in NMU-induced tumors compared with normal rat mammary glands and H-ras mutations did not confer a unique expression profile by supervised analysis. Although the gene expression effects of this mutation do not specify a particular global expression profile, unaccounted for alterations in the H-ras pathway may lead to indistinguishable profiles. Our data and that cited above lend support to the notion that H-ras mutation does not uniquely play a major role in the development of NMU-induced rat mammary tumors.

With the increasing use of microarrays to study gene expression in disease and continued reliance on animal models, the ability to compare array results across species will be helpful for translational research. Few reports address this issue, although recent studies on cross-species hybridizations to human arrays have been described (28–30). Despite adequate reproducibility, there is evidence that the specificity of cross-species hybridization may be an obstacle for accurate interpretation of results. Chismar et al. (28) detected a presence call of only 29% for non-human primate hybridizations compared with 46% for human hybridization to Affymetrix U95av2 gene chips.

In this study, raw data from species-specific microarray chips for rat and humans were compared. Known orthologs were identified using information in the GenBank database and the raw hybridization intensities for these genes standardized to remove systematic differences between species. Separate species arrays were then combined for further analysis. One drawback of this analysis, in addition to having to account for the differences in baseline and scale of variation between homologous human and rat genes (addressed in the Materials and methods) is the availability of proven rat to human orthologous genes. Despite this limitation, we found 2305 unique orthologous probe sets out of ~5000 non-EST genes on the RAE230a rat chip and ~10 000 genes on the U95av human chip excluding multiple probe sets that match the same gene. Using the unique 2305 orthologs, we were able to approximate the clustering of human tumors by Wang et al. (40). It is probable that our cross-species array comparison overlooks genes that are important in the pathogenesis of breast cancer in these two species. Nonetheless, because only well-established orthologs are used and hybridization is done within and not between species, our analysis may be more accurate than hybridization of non-human cRNA to human microarrays. One cannot predict the fidelity of cross-species hybridization considering that very few genes can be matched between rat and human gene chip by comparing the probe sequence alone (data not shown). As we continue to expand our understanding of non-human genomes our identification of orthologous genes will only increase.

Our primary purpose for comparing tumors induced by NMU in rats and human breast cancer is to define a subset of human tumors that this rat model most closely emulates. Using KNN analysis, we acquired the ER status and grade of the human breast cancers with gene expression profiles most similar to an NMU-induced tumor. After analyzing all rat tumors, we found that NMU-induced rat tumors have a molecular gene expression pattern most similar to ER-positive, low to intermediate grade breast cancer. Experimental studies using this model support these findings. Hormonal dependence of these tumors is generally accepted for both induction and early growth. This has been supported by observations of tumor suppression in ovariectomized animals (63,64) and those pretreated with estradiol, progesterone and tamoxifen (9,65). Furthermore, carcinogen-induced tumors have previously been shown to be ER-positive by steroid binding assays and immunohistochemical assays (66–68). The human breast cancers profiled by Wang et al. (40) were all invasive carcinomas. The results reported here show that NMU-induced rat mammary tumors are all non-invasive carcinomas. Nevertheless, the rat mammary cancers appear closely related to human...
invasive carcinomas in general, and in particular, to lower grade, ER-positive human tumors. This is not surprising in view of the phenotypic, single gene and global gene similarities of the non-invasive and invasive compartments of individual human breast carcinomas (52, 69, 70).

In conclusion, primary NMU-induced rat mammary tumors are ER-positive, low to intermediate grade, non-invasive breast carcinomas. Only a minority harbor mutations in H-ras. Even those tumors with activating ras mutations could not be distinguished from tumors with a wild-type H-ras by analysis of global gene expression. It becomes harder to argue that the presence of an H-ras mutation, although rarely found in human breast cancer, qualifies for the comparison of carcinogen-induced rat tumors with human breast cancer. According to gene expression profiling, these rat tumors are homogeneous. Using orthologous genes and statistical analyses raw microarray data from different species were compared. The approach described in this study provides a means to directly correlate gene expression in animal models to human cancer and disease states. Using this approach, the study found that the rat tumors share a global molecular phenotype with invasive, low-to-intermediate grade, ER-positive human breast cancer. The fact that the rat tumors are non-invasive may be the most obvious difference between the two tumor systems. Previous work has shown that non-invasive rat tumors will progress to invasive behavior with repeated transplantation (71). Now that we have identified the similarity between NMU-induced tumors and ER-positive, low-grade breast cancer, this model can be used to study the transition of this subset of breast cancers from an in situ lesion to invasive cancer.

Supplementary material

Supplementary material can be found at: http://www.carcin.oxfordjournals.org/

Acknowledgements

This work was supported by the Womens’ Cancers Program, Dana Farber Cancer Institute; Breast Cancer Research Foundation; Dana Farber/Harvard SPOR in Breast Cancer from the NCI (NIH 1R01HG022341); and Ruth L. Kirschstein National Research Service Award (T32 CA09535-18).

Conflict of Interest Statement: None declared.

References
