In silico data filtering to identify new angiogenesis targets from a large in vitro gene profiling data set

MARY E. GERRITSEN,1 ROBERT SORIANO,4 SUYA YANG,1 GLADYS INGLE,1 CONSTANCE ZLOT,1 KAREN TOY,4 JANE WINER,4 APARNA DRAKSHARAPU,2 FRANKLIN PEALE,2 THOMAS D. WU,3 AND P. MICKEY WILLIAMS4

Departments of 1Cardiovascular Research, 2Pathology, 3Bioinformatics and 4Molecular Biology Biology, Genentech, South San Francisco, California 94080

Received 28 March 2002; accepted in final form 8 May 2002

Gerritsen, Mary E., Robert Soriano, Suya Yang, Gladys Ingle, Constance Zlot, Karen Toy, Jane Winer, Aparna Draksharapu, Franklin Peale, Thomas D. Wu, and P. Mickey Williams. In silico data filtering to identify new angiogenesis targets from a large in vitro gene profiling data set. Physiol Genomics 10: 13–20, 2002. First published May 15, 2002; 10.1152/physiolgenomics.00035.2002.—The objective of this study was to use gene expression data from well-defined cell culture models, in combination with expression data from diagnostic samples of human diseased tissues, to identify potential therapeutic targets and markers of disease. Using Affymetrix oligonucleotide array technology, we identified a common profile of genes upregulated during endothelial morphogenesis into tubelike structures in three in vitro models of angiogenesis. Rigorous data selection criteria were used to identify a list of over 1,000 genes whose expression was increased more than twofold over baseline at either 4, 8, 24, 40 or 50 h. To further refine and prioritize this list, we used standard bioinformatic algorithms to identify potential transmembrane and secreted proteins. We then overlapped this gene set with genes upregulated in colon tumors vs. normal colon, resulting in a subset of 128 genes in common with our endothelial list. We removed from this list those genes expressed in 6 different colon tumor lines, resulting in a list of 24 putative, vascular-specific angiogenesis-associated genes. Three genes, gp34, stanniocalcin-1 (STC-1), and GA733-1, were expressed at levels 10-fold or more in colon tumors compared with normal mucosa. We validated the vascular-specific expression of one of these genes, STC-1, by in situ hybridization. The ability to combine in vitro and in vivo data sets should permit one to identify putative angiogenesis target genes in various tumors, chronic inflammation, and other disorders where therapeutic manipulation of angiogenesis is a desirable treatment modality.

A UNIQUE FEATURE of isolated cultured endothelial cells is their ability, in vitro, to differentiate into an interconnected network of tubelike structures when exposed to a suitable extracellular matrix and combination of growth factors or other stimuli. Previous studies by other investigators have demonstrated that genes expressed by the endothelium during this process play critical roles in the normal angiogenic process. For example, Hla and Macig (8) identified the novel G protein-coupled receptor, edg-1, as gene whose expression is significantly upregulated in cultured endothelial cells during reorganization into tubelike structures in response to phorbol myristate acetate (PMA). Years later, edg-1 was shown to be the receptor for sphingosine-1-phosphate (16, 23) and was shown to be necessary for proper endothelial migration and maturation in vivo (18). Similarly, the matrix metalloproteases MMP1, MMP2, and MMP9 are upregulated in various in vitro models of angiogenesis, and all three of these proteases have roles in normal and pathological angiogenesis (22, 24).

In a previous study by our laboratory, we identified over 80 genes expressed by endothelial cells differentiating into tubelike structures in three-dimensional collagen gels (13). Here, we have applied stringent screens to data from Affymetrix oligonucleotide arrays to identify sets of genes whose expression is upregulated in vitro models of endothelial cell differentiation, as well as subsets of these genes that are also associated with tumor vasculature. This initially involved examining changes in gene expression in three different in vitro systems that capture many of the steps associated with angiogenesis, including endothelial activation, matrix degradation, cell migration, capillary network formation, and lumen formation. This made use of endothelial cells cultured in different matrices and with different combinations of growth factors. We identified a set of 1,038 genes that were upregulated during endothelial differentiation into capillary-like tubes in all three of the in vitro models. To identify targets amenable to antibody-based therapies, we focused on 397 genes that were selected based on algorithms to identify transmembrane and secreted proteins. To filter this large number of genes further, we examined a preexisting gene expression database of colon tumors and colon tumor cell lines. Our intent was to identify genes that might be associated with angiogenic tumor stroma. We first generated a list of genes that were upregulated in colon tumors compared with normal colon mucosa. We identified 128 genes that...
were in the endothelial transmembrane/secreted protein group and were upregulated in tumors. From this subset, those genes expressed in colon tumor epithelial cells were subtracted; yielding a candidate list of 24 putative endothelial-derived angiogenesis associated genes. Three of these genes, i.e., gp34, stanniocalcin-1, and gastrointestinal tumor-associated antigen GA733-1, were expressed at levels 10-fold or more in colon tumor tissue compared with normal mucosa. The association of one of these genes, stanniocalcin-1, with angiogenesis was further validated by demonstrating its upregulation in a rat corneal pocket model of angiogenesis and by selective vascular-associated expression in colon tumors by in situ hybridization.

Thus by applying a conservative analysis, common gene profiles from well-defined cell culture models can be filtered by examining data sets of disease tissue. The intersection of data from these various models enables the rapid identification and prioritization of candidate genes for further target validation. We believe this is a logical way to proceed using the power of nucleotide arrays to analyze precise in vitro biological models and using these results to find correlation in silico with less precise human disease tissue data sets.

METHODS

Cell culture. Human umbilical vein endothelial cells (HUVEC) were obtained from Clonetics (San Diego, CA) and grown in Clonetics EGM medium supplemented with 10% fetal bovine serum (FBS) and endothelial cell growth supplements provided by the manufacturer. Cells from passages 4–7 were used through the study.

Collagen gels. Three-dimensional collagen gels were prepared as described previously (26). Briefly, collagen gels were formed by mixing together ice-cold gelation solution [10× medium 199 (M199), H2O, 0.53 M NaHCO3, 200 mM L-glutamine, type I collagen, 0.1 M NaOH; 100:27.2:50:10:750:62.5 by volume] and cells in 1× basal medium (see below) at a concentration of 3×10^6 cells/ml at a ratio of 4 vol gelation solution to 1 vol of cells. Ten milliliters of cell mixture was added to 10-cm dishes. After gelation at 37°C for 30 min, the gels were overlaid with 1× basal medium (BM) consisting of M199 supplemented with 1% FBS, 1× ITS (insulin, transferrin, selenium), 2 mM L-glutamine, 50 mg/ml ascorbic acid, 26.5 mM NaHCO3, 100 U/ml penicillin and 100 U/ml streptomycin. Tube formation was elicited by the inclusion in the culture media of either a mixture of phorbol myristate acetate (PMA, 50 nM), vascular endothelial cell growth factor (40 ng/ml) and basic fibroblast growth factor (40 ng/ml) (“PMA growth factor mix”) or hepatocyte growth factor (40 ng/ml) and vascular endothelial cell growth factor (40 ng/ml) (“HGF+VEGF mix”). Endothelial tube formation in the presence of these two combinations has been previously described in detail by our laboratory (25, 26).

Fibrin gels. HUVEC (4×10^5 cells/ml) were suspended in M199 containing 1% FBS (Hyclone) and human fibrinogen (2.5 mg/ml). Thrombin (50 U/ml) was then added to the fibrinogen suspension at a ratio of 1 part thrombin solution per 30 parts fibrinogen suspension. The solution was then layered onto 10-cm tissue culture plates (total volume: 15 ml/plate) and allowed to solidify at 37°C for 20 min. Tissue culture media [10 ml of BM containing PMA (50 nM), bFGF (40 ng/ml), and VEGF (40 ng/ml)] was then added, and the cells incubated at 37°C in 5% CO2 in air for the indicated period of time.

RNA preparation. Total RNA was extracted from endothelial cells incubated 0, 4, 8, 24, 40, and 50 h in the different matrices and media combinations using a TRIzol extraction followed by a second purification using RNAeasy Mini Kit (Qiagen). In the in vivo experiment, RNA was isolated from rat corneas using TRIzol reagent as previously described (13).

Preparation of cRNA. The methods for preparation of cRNA and hybridization/scanning of the arrays were provided by Affymetrix (Santa Clara, CA). Samples were hybridized to the arrays at 40°C for 19 h in a rotisserie hybridization oven set at 60 rpm. Arrays were washed, stained twice in the Affymetrix Fluidics station, and then analyzed in the Affymetrix Scanner (following the Affymetrix protocol). In the study described herein, the HUFL 6800 chip was used, which contains oligonucleotide sets for 6,800 known human genes (although some genes were represented by more than one probe set on the microarray).

Data analysis. Data analysis was performed using the Affymetrix GeneChip Analysis (v3.2) software. Pairwise comparisons were made using time 0 chips as a baseline. Three replicate samples were analyzed for each experimental condition and time. Hence, there were three time 0 samples for each treatment and three replicates of each successive time point. Therefore, a 3×3 comparison was performed for each time point compared against each time 0 point. This resulted in nine comparisons. We next utilized the Affymetrix Data Mining Tool software to examine the qualitative comparison parameter of “increase” (or “marginal increase”) to find genes that increased in expression compared with the time 0 samples. Each of the nine pairwise analysis results was pivoted to display each of the nine pairwise comparisons vs. each gene analyzed. These results were exported into Microsoft Excel, where the data was sorted, placing those genes with the greatest percent agreement in expression for “increase” calls in the nine comparisons on top and listing genes in descending order of agreement. We used a Mann-Whitney rank order test to determine what percent agreement was necessary for confidence (12), and we used a cutoff of 88%; i.e., in a 3×3 comparison only those genes called increased in at least eight of nine pairwise comparisons were considered to increase in their expression. It is also important to note that this is considered to be a fairly stringent approach, as the Affymetrix Data Mining Tool software will opt to call a gene’s expression as unchanged unless several criteria have been satisfied (see Affymetrix DataMining Tool Guide for details). The magnitude of the change in gene expression (fold change) was reported for each comparison, which was called as an increase or decrease in expression. These values were averaged to obtain and average fold change for each gene, which was determined to be increased in expression. This very conservative analytical approach thus limited the number of false-positive gene identifications.

Colon tumor and normal mucosal samples. Total RNA from six different colon adenocarcinoma samples as well as six normal colon mucosal samples were prepared for Affymetrix array analysis. Colon tumor samples and normal mucosal samples were provided by the University of Leeds and the University of Sheffield and were obtained with full Institutional Review Board approval. Total RNA was also isolated from a number of colon tumor cell lines: HT29, HCC2998, HCT116, SW620, and Colo 205.

Rat corneal angiogenesis. Angiogenesis elicited in the cornea using Hydron pellets containing bFGF and VEGF was
used to confirm the increased expression of one of the genes in an in vivo model of angiogenesis. Hydron pellets containing the combination of bFGF and VEGF (200 ng/ml each) were implanted into the corneas of 600-g male Sprague-Dawley rats. At day 7, the animals were euthanized. The contralateral cornea was collected as the control. Total RNA was isolated from the corneas and used for real-time PCR (Taqman) as described below.

**Real-time PCR (Taqman).** Real-time PCR was performed as described previously (13) using the Taqman model 7700 sequence detector (ABI-Perkin Elmer). Expression levels for each gene were normalized to GAPDH, which was unaffected in the different treatment groups. Results are expressed as the mean of duplicate determinations. Table 1 provides the sequences for the Taqman probes and primers used in this study. Table 2 provides the clinical information for samples used in this study.

**In situ hybridization of tissue specimens.** Formalin-fixed, paraffin-embedded colon tumor tissue (see above) was used for in situ mRNA expression. In vitro transcription and 32P labeling of sense and antisense riboprobes was performed as described previously (13). Stanniocalcin-1 sequences were PCR-amplified from plasmid DNA using gene-specific primers that encoded T3 or T7 RNA polymerase initiation sites. Sense and antisense riboprobes were prepared by in vitro transcription from the PCR-amplified templates and diluted in hybridization buffer to a specific activity of 1 × 10^6 cpm/ml. Tissue sections 5 μm thick were deparaffinized, deproteinized in 4 μg/ml of proteinase K for 30 min at 37°C, hybridized at 55°C overnight, then washed at high stringency (55°C in 0.1× SSC for 2 h). Glass slides were dipped in NBT2 nuclear track emulsion (Eastman Kodak), exposed in sealed plastic slide boxes containing desiccant for 4 wk at 4°C, developed, and counterstained with hematoxylin and eosin.

**RESULTS AND DISCUSSION**

Three different in vitro models of endothelial differentiation were selected to identify a common set of genes required for endothelial activation, migration, lumen formation, and branching morphogenesis. Two of the models employed a matrix of type I collagen, while the third model used a gel composed of fibrin. The selection of two different matrix substrates was important, since the integrins required in these two models differ. In the collagen gel model, tube formation is highly and selectively dependent on the integrin αβ1 (7), whereas in the fibrin gel model, tube formation is dependent on the integrins αβ3 and αβ1 (3). In the collagen model, previous studies by our laboratory as well as others demonstrated that neither bFGF nor VEGF, alone or in combination, were sufficient to promote survival and tube formation (7, 10, 26). However, PMA alone, or more potently, in combination with bFGF and VEGF, promoted endothelial survival and differentiation. Therefore, one comparison made in this study was the gene expression profile elicited by PMA, bFGF, and VEGF in type I collagen gels vs. fibrin gels. More recently, we found that HGF, in combination with VEGF, also supported differentiation of endothelial cells in collagen gels into tubelike structures (25). Thus comparison of the genes expressed by HGF + VEGF-driven tube formation vs. PMA + bFGF + VEGF-driven tube formation in collagen gels provided an additional method of identifying a common set of genes.

The experimental design included three replicates derived from three independent HUVEC cell lines (i.e., derived from different patients) and five different time points. All samples were compared with the same sets of endothelial cells at time 0, i.e., cells in monolayer just prior to trypsinization and suspension in the collagen or fibrin gels. To improve accuracy and minimize variation, care was taken to en-

**Table 1. Sequences used for real-time PCR analyses (Taqman)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC1</td>
<td>GGGTGGCGGCTCAAAAT</td>
<td>GCCAACCTGTAAGGGCCTGT</td>
<td>CAGCTGAAGTGGTCCGCTGCCTC</td>
</tr>
<tr>
<td>GAPDH</td>
<td>TGTTCCAGTAGCTCTACCACG</td>
<td>GACGCCAGTAGACTCCACGAC</td>
<td>ACCCCATCACCATCTTCCAGGAGCGAGA</td>
</tr>
<tr>
<td>Rat STC-1</td>
<td>GGTTGCGCGCTCAAAAT</td>
<td>GCCAACCTGTAAGGGCAGTGT</td>
<td>CAGCTGAAGTGGTCCGCTGCTC</td>
</tr>
<tr>
<td>Rat GAPDH</td>
<td>TGTTCCAGTAGCTCTACCACG</td>
<td>GACGCCAGTAGACTCCACGAC</td>
<td>ACCCCATCACCATCTTCCAGGAGCGAGA</td>
</tr>
</tbody>
</table>

**Table 2. Clinical information for the colon tumor and normal mucosal samples included in this study**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pathology</th>
<th>Age, yr</th>
<th>Sex</th>
<th>Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>861</td>
<td>adenocarcinoma</td>
<td>64</td>
<td>M</td>
<td>&gt;50% adenocarcinoma; 20% necrosis</td>
</tr>
<tr>
<td>722</td>
<td>adenocarcinoma</td>
<td>69</td>
<td>M</td>
<td>&gt;60% adenocarcinoma</td>
</tr>
<tr>
<td>881</td>
<td>adenocarcinoma</td>
<td>46</td>
<td>F</td>
<td>10–50% adenocarcinoma; 10% necrosis</td>
</tr>
<tr>
<td>981</td>
<td>adenocarcinoma</td>
<td>56</td>
<td>F</td>
<td>&lt;10% necrosis</td>
</tr>
<tr>
<td>992</td>
<td>adenocarcinoma</td>
<td>76</td>
<td>M</td>
<td>abundant inflamed stroma, no necrosis</td>
</tr>
<tr>
<td>992</td>
<td>adenocarcinoma</td>
<td>57</td>
<td>F</td>
<td>&lt;10% necrosis</td>
</tr>
<tr>
<td>L3</td>
<td>normal mucosa</td>
<td>UN</td>
<td>UN</td>
<td>50% adenocarcinoma; 20% necrosis</td>
</tr>
<tr>
<td>L5</td>
<td>normal mucosa</td>
<td>UN</td>
<td>UN</td>
<td>&gt;50% adenocarcinoma; 20% necrosis</td>
</tr>
<tr>
<td>973</td>
<td>normal mucosa</td>
<td>62</td>
<td>F</td>
<td>&gt;60% adenocarcinoma</td>
</tr>
<tr>
<td>974</td>
<td>normal mucosa</td>
<td>72</td>
<td>UN</td>
<td>10–50% adenocarcinoma; 10% necrosis</td>
</tr>
<tr>
<td>987</td>
<td>normal mucosa</td>
<td>74</td>
<td>M</td>
<td>&lt;10% necrosis</td>
</tr>
<tr>
<td>988</td>
<td>normal mucosa</td>
<td>65</td>
<td>M</td>
<td>abundant inflamed stroma, no necrosis</td>
</tr>
</tbody>
</table>

M, male; F, female; UN, unknown.
sure that identical lot numbers of serum, growth factors, plastic ware, collagen, thrombin, and fibrin were used in all of the experiments. All cells used were at passage 5 and were maintained on an identical feeding and split cycle. The use of multiple samples and several time points, as well as the care taken to make samples as similar as possible, enabled identification of more than a thousand genes whose expression was upregulated at least twofold at one of the time points compared with time 0 and which were identified as increased in at least eight of nine pairwise comparisons. The validity of this cutoff was assessed by independent determination of difference calls by RT-PCR analysis of over 100 genes; and 100% of the genes were confirmed (data not shown). This stringent analysis therefore results in very few, if any, false positives. This approach does allow for false negatives (that is, genes whose expression did change may be missed).

Table 3 summarizes the results obtained from this initial analysis of the data. Each in vitro model of tube formation resulted in the identification of well over a thousand genes whose expression was upregulated more than twofold at some time (4, 8, 24, 40, or 50 h). Such a list of genes is too large to analyze experimentally, and so we sought additional methods to reduce the complexity of the data set with the ultimate goal of identifying genes with a role in human disease. The first step was to compile lists of upregulated genes for each tissue culture model. We then asked which of these genes were upregulated in common among the three models. This yielded a data set of 1,038 different genes or, more precisely, accession numbers. Several genes were represented multiple times due to their representation by different accession numbers. For example, cyclooxygenase-2 came up multiple times, as did caldesmon-1 and CGRP1. Thus the number 1,038 is an overestimate of the total number of independent genes identified.

Using the DataMining software, GeneSpring (Silicon Genetics), we generated self-organizing maps (i.e., SOM) of gene expression. These maps group genes based on similarity of expression profiles over the time course. Representative clusters (Fig. 1) demonstrated remarkably similar patterns of gene expression in the three models. The genes identified in the

Table 3. Total numbers of genes (i.e., probe sets) upregulated more than twofold compared with 0 h in at least 8/9 of the pairwise comparisons.

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>PMA Collagen</th>
<th>PMA Fibrin</th>
<th>HGF + VEGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 h</td>
<td>877</td>
<td>968</td>
<td>912</td>
</tr>
<tr>
<td>8 h</td>
<td>872</td>
<td>978</td>
<td>841</td>
</tr>
<tr>
<td>24 h</td>
<td>813</td>
<td>1,111</td>
<td>867</td>
</tr>
<tr>
<td>40 h</td>
<td>853</td>
<td>1,137</td>
<td>613</td>
</tr>
<tr>
<td>50 h</td>
<td>785</td>
<td>1,032</td>
<td>805</td>
</tr>
</tbody>
</table>

Some genes were represented more than once on the microarray by different probe sets.

Fig. 1. Representative patterns of gene expression in three independent models of endothelial tubulogenesis. Each frame (i.e., A–D) shows the patterns of expression of a cluster of genes [based on self-organizing maps (SOM), see below] in the three models, from left to right: HGF + VEGF in collagen gel (“HGF-Coll_FL”), PMA + VEGF + bFGF in collagen gel (“PMAcol_FL”), PMA + VEGF + bFGF in fibrin gel (“PMAfib_FL”). SOM were generated from the 1,038 genes identified as upregulated and common to the three models of endothelial differentiation into tubelike structures. Shown are 4 clusters demonstrating representative patterns of expression (out of a total of 24). Data are expressed as relative intensity (average difference values for a specific gene normalized across the chip). Each line represents the time course (0, 4, 8, 24, 20, 50 h) of expression for an individual gene. Supplemental Table 4 (online supplement, available at the Physiological Genomics web site) provides lists of those genes represented in frames A, B, C, and D, respectively.
groups shown in Fig. 1 are listed in Supplemental Table 4, published online at the Physiological Genomics web site. As observed earlier by our group (13), gene expression during tube formation falls into several generalizable patterns that we termed "early transient," "delayed transient," and "stably induced."

A hierarchical cluster of the 1,038 genes in the PMA + bFGF + VEGF collagen model is shown in Fig. 2 and correlated with temporal morphological events in the cell culture model as depicted in cartoon form in Fig. 2. Comparison of the data in Figs. 1 and 2 and Supplemental Table 4 reveals that early events, such as cell activation and cell membrane ruffling, are coincident with the upregulation of a number of genes associated with endothelial activation and chemotaxis (interleukin-8), membrane turnover (phospholipase A2), cytoskeletal rearrangements (gelsolin, various rab and ras family members), and enzymatic digestion (matrix metalloproteinases, serine proteases). In contrast, the genes whose expression is upregulated at later time points appear to represent processes associated with vessel maturation (junction formation, matrix proteins, pericyte recruitment; e.g., cadherins, collagens, laminins, nidogens, integrins). Thus genes of unknown function that cocluster with these genes may be involved in related processes.

One criterion that could be applied to validate angiogenesis related genes is that the expression of those genes should be upregulated in vivo at sites of angiogenesis, such as in tumors. This could be evaluated on a one-by-one basis, but with a list as long as 1,038, such a task would be overwhelming. We initially trimmed the number of candidates to those amenable to either protein- or antibody-based manipulation (and also potential targets for protein-based therapeutics) by selecting only those genes that are likely to be either transmembrane proteins or secreted proteins. The sequences of the list of 1,038 genes were thus analyzed by several different programs, which predict signal sequences and transmembrane domains, with the resultant identification of 397 genes. (It should be noted that...

---

Supplementary material (Tables 4 and 5) to this article is available online at http://physiolgenomics.physiology.org/cgi/content/full/10/1/13/DC1.
these programs can also identify intracellular proteins which meet signal sequence or hydropathy plot defaults and may not identify secreted proteins that do not have a classic signal sequence.) We then used an in silico approach to identify genes upregulated in colon carcinomas using a gene expression database generated at Genentech. This database contained expression data from multiple colon tumor samples, as well as colon tumor-derived cell lines. We reasoned that one way to identify endothelial-specific genes that were important in angiogenesis would be that 1) the gene would be expressed in colon tumors, 2) the expression of the gene would be greater in tumors than in normal colon mucosa, and 3) the gene would not be expressed in colon tumor lines. Because of the heterogeneity in tumor samples and regions of tumors that were analyzed, we only selected those tumor data sets in which the endothelial cell marker, CD31 (PECAM), and the angiogenesis marker, VEGF, were detected (a total of 6 tumors). We then filtered this database using a non-parametric Wilcoxon two sample rank test with a P value of 0.2 to identify genes that were differentially expressed in colon tumors vs. normal mucosal tissue (2,279 genes). This list was then placed in a Pearson hierarchical cluster, and only the genes increased in tumors were selected (1,236 genes). This list was overlapped with our list of 397 transmembrane and secreted angiogenesis associated genes, generating a list of 128 common genes (Supplemental Table 5). The list of 128 genes contained several that have been previously shown to play roles in angiogenesis, including interleukin-8 (2, 14), Thy-1 (17), and endothelial monocyte activating polypeptide II (4, 5, 27). To further enrich this list for “stromal-derived” genes, i.e., genes expressed by cells other than the tumor cell (e.g., endothelial cells, pericytes, fibroblasts, inflammatory cells), we subtracted those genes expressed in 100% of the five colon epithelial tumor cell lines evaluated. The final list of angiogenesis associated, endothelial/stromal cell-specific genes comprised 24 genes (Fig. 3). The three most highly upregulated angiogenesis-associated genes in colon tumor tissue compared with normal mucosa were gp34, stanniocalcin-1 (STC-1), and the gastrointestinal tumor-associated antigen GA733-1 (Fig. 4) and thus may represent three new potential therapeutic targets. D90224 (gp34, also known as OX40 ligand) is a membrane-bound TNF superfamily member thought to play a role in T-cell costimulatory
responses (15) and T-cell interactions with endothelial cells (11). Association with colon carcinomas or angiogenesis has not been previously described. U25997 (STC-1) is a secreted homodimeric glycoprotein with poorly understood function in mammals. It is a homolog of a fish protein, also known as STC-1, which regulates calcium and phosphate uptake and excretion across the gills and kidneys of bony fishes, and may also exert calcitropic actions in humans (6, 9, 19–21).

We had previously identified STC-1 as a highly upregulated gene during endothelial tubulogenesis using a different gene expression analysis (13). The function of STC-1 in angiogenesis, however, is unknown. J04152 (GA733-1) was initially defined as an antigen recognized by a monoclonal antibody to a tumor-associated antigen, gastrointestinal tumor-associated antigen GA733-1. The product of GA733-1, also known as TROP2, exhibits 49% homology to Ep-CAM, an epithelial cell homophilic cell adhesion molecule with proposed roles in epithelial morphogenesis. Both Ep-CAM and TROP2 have been previously reported to be highly expressed in colon carcinomas (1).

We selected one gene, STC-1, for further validation. To evaluate STC-1 expression in an independent model of angiogenesis, mRNA levels for rat STC-1 were determined in corneas from eyes containing either Hydron pellets containing excipient (control) or Hydron pellets containing 200 ng/ml VEGF. The corneas were harvested at day 6, a time point at which numerous small vessels growing up from the choroidal circulation can be visualized in the VEGF eye, but not the control eye (25). The ratio of STC-1 to GAPDH (a housekeeping gene) in control corneas was 3.4 ± 0.6 and was dramatically higher in the VEGF-treated corneas (98.3 ± 20.6). Figure 5 shows the focal, intense expression of STC-1 in small vessels present in three different colon adenocarcinomas. Note the lack of expression in the adjacent tumor cells and in the colonic epithelial cells. In contrast, no detectable STC-1 signal was observed in the vasculature of normal mucosa (e.g., in more normal regions of Fig. 5I).

In summary, many different methods have been developed to analyze gene expression data in an attempt to derive functionally related data from a single experimen-
tional sample set. An alternative, or perhaps adjunct method, to further refine these analyses is to perform multiple overlapping in silico analyses with data sets derived from related samples of interest. Furthermore, the well-controlled relatively precise biology of an in vitro model can be a powerful starting point for array analysis. In comparison, the complexity of normal and disease samples requires large sample sizes and less stringent analysis methods to sift through the noise and find genes of biological interest. The validity of this approach was demonstrated in the present study where a list of 1,038 candidate genes were identified which were common to three related models of in vitro endothelial tube formation. Greater than 50% of these genes had never been previously associated with endothelial differentiation or angiogenesis. Further filters, utilizing in silico comparison to tissue data sets, enabled the identification of a smaller subset of genes, which would be potential targets for angiogenesis-related disease therapy. This method therefore provides a powerful and relatively rapid means of identifying new genes with potentially important roles in angiogenesis. Similar analyses could be applied using data sets from different types of tumors (e.g., renal cell carcinomas, lung tumors, breast tumors) or other disease tissues.

We gratefully acknowledge Dr. Kenneth Hillan, Dept. of Pathology at Genentech, for help in preparation of Taqman probes and primers, and we thank Gretchen Frantz for help with the in situ studies.

REFERENCES

10. Ian N, Mahooti S, and Madri JA. Distinct signal transduction pathways are utilized during the tube formation and sur-