Multiple factors are involved in the translation of functional genomic results into proteins for proteome research and target validation on tumoral tissues. In this report, genes were selected by using DNA microarrays on a panel of colorectal cancer (CRC) paired samples. A large number of up-regulated genes in colorectal cancer patients were investigated for cellular location, and those corresponding to membrane or extracellular proteins were used for a non-biased expression in Escherichia coli. We investigated different sources of cDNA clones for protein expression as well as the influence of the protein size and the different tags with respect to protein expression levels and solubility in E. coli. From 29 selected genes, 21 distinct proteins were finally expressed as soluble proteins with, at least, one different fusion protein. In addition, seven of these potential markers (ANXA3, BMP4, LCN2, SPARC, SPP1, MMP7, and MMP11) were tested for antibody production and/or validation. Six of the seven proteins (all except SPP1) were confirmed to be overexpressed in colorectal tumoral tissues by using immunoblotting and tissue microarray analysis. Although none of them could be associated to early stages of the tumor, two of them (LCN2 and MMP11) were clearly overexpressed in late Dukes' stages (B and C). This proteomic study reveals novel clues for the assembly of a robust and highly efficient high throughput system for the validation of genomic data. Moreover it illustrates the different difficulties and bottlenecks encountered for performing a quick conversion of genomic results into clinically useful proteins. Molecular & Cellular Proteomics 5:1471–1483, 2006.

Colorectal cancer (CRC) is the second most prevalent cancer in the western world. CRC develops over decades and involves multiple genetic events. From a genetic point of view, CRC is one of the best studied solid tumors (1). However, it is paradoxical that despite this level of knowledge and the existence of good screening procedures CRC continues to be a major cause of mortality in the developed countries. So the implementation of simpler, non-invasive methods for the early detection of CRC is necessary. These methods should be based on new biomarkers, preferably proteins or antibodies detectable in serum or plasma.

Traditional methods of identifying novel targets involved in cancer progression were based on studies of individual genes. Now the use of DNA microarrays permits the analysis of the expression of tens of thousands of genes simultaneously and rapidly (2, 3). Microarray analysis has been used for gene expression analysis of different neoplasms (4, 5) including CRC (6–9). However, only a few studies have pursued further insight into the function and/or importance of individual genes and their application to the proteome research of a tumor. Some of these genes have been proposed as candidate cancer biomarkers (10–12). More recently a number of proteomic studies have also addressed the identification of potential targets in CRC (13–15).

As a part of a comprehensive approach to study the proteome of CRC and to identify new biomarkers, we investigated the feasibility of expressing soluble proteins corresponding to up-regulated genes in cancer patients with surgically resected colon polyps and tumors. We used cDNA microarrays (CNIO Oncochip) (16) to identify differentially expressed genes in malignant versus normal samples isolated from individual patients with CRC.

However, to carry out these studies, it is necessary to design and standardize high throughput methods and strategies to translate genes into proteins (or protein fragments) and subsequently generate antibodies in a quick manner. To retrieve the necessary genes we chose two possible routes. The first alternative was the cDNA clones from the IMAGE Consortium that were printed on the Oncochip (16). The second was the full-length cDNA clones either from the Mammalian Gene Collection (MGC) (17, 18) or from other sources (i.e. donations, etc.). Regarding the expression system, Escherichia coli is still the most suitable host for massive production of recombinant proteins in a high throughput approach because of its fast growth, high production rates, and cost effectiveness. It is necessary to obtain as many soluble pro-
teins as possible because insoluble proteins are generally inadequate for antibody production, crystallization, or functionality studies in general. For this reason, we selected three fusion tags: His<sub>6</sub>, GST, and MBP; the latter two have been described previously as very efficient solubilizing agents in selected subsets of human proteins (19, 20). A limitation to the use of high throughput expression procedures is the use of conventional cloning procedures based on restriction and ligation of DNA fragments, which are not too efficient and are excessively time-consuming. Here we propose the use of recombinational cloning systems such as the Gateway cloning technology (21, 22). This system works in two recombination steps: the first step is the creation of an entry clone, and the second step is the subsequent generation of multiple destination vectors. These steps are usually very efficient and can be easily made in parallel for many different genes.

Another limitation in proteome research and target validation is the poor availability of antibodies against the gene products that are being identified in massive numbers either by DNA microarray analysis or by proteomic technologies such as two-dimensional PAGE or LC-MS/MS. In many cases there is no possibility to rapidly assess the actual value of new markers because the number of potential protein candidates exceeds by far that of existing antibodies. The use of antibodies for tissue profiling (i.e. in tissue microarrays) allows for a fast approach to generate protein expression data for normal and disease tissues on large numbers of individual patients (23). Only antibodies possess the sufficient specificity for a proper detection of the target proteins. As a result, the validation of some of those proteins that might become biomarkers for screening or prognostic value is considerably delayed. Moreover specific antibodies are needed for numerous functional assays, including ELISAs, localization studies, and “pulldown” experiments (24). Mouse-derived monoclonal antibodies continue to be the affinity reagent of choice in proteomic analyses, but their production remains restricted by antigen quality, high tissue culture requirements, and low throughput screening methods. Rabbit polyclonal antibodies are a faster and cheaper alternative, but the quality and specificity depend largely on the quality of the antigen. Other methods such as phage display that may become useful alternatives are currently being tested.

In this report, we describe the translation of potential gene targets into soluble proteins by using a recombinational cloning system and their evaluation by immunoblotting and tissue microarray. Important bottlenecks and limitations will be described in protein production and purification as well as in antibody production. As a proof of the usefulness of this approach for selecting potential biomarkers for CRC, seven proteins (annexin A3 (ANXA3), lipocalin 2 (LCN2), bone morphogenetic protein 4 (BMP4), osteonectin (secreted protein, acidic and rich in cysteine (SPARC)), osteopontin (secreted phosphoprotein 1 (SPP1)), and matrix metalloproteases 7 and 11 (MMP7 and MMP11) were characterized for CRC diagnosis and characterization.

**EXPERIMENTAL PROCEDURES**

**DNA Microarray Experiments**—Tissue samples from 22 CRC patients were provided by the Tissue Bank Network of the CNIO in collaboration with the following hospitals in Spain: Virgen de la Salud (Toledo), Cartagena (Murcia), Puerta de Hierro (Madrid), Clínico de San Carlos (Madrid), Ramón y Cajal (Madrid), and Alcorcón (Madrid). These frozen samples consisted of pairs of tumoral and normal mucosa sections. Clinical and pathologic data of the samples are shown in Table I. Total RNA was isolated from 50–100 mg of frozen biopsies as described previously (16, 25–27). To generate fluorescent cDNA, 30 µg of tumoral or normal RNA samples were labeled, respectively, with Cy5-dUTP and Cy3-dUTP (GE Healthcare). Hybridized slides were scanned with the Agilent G2565BA Microarray Scanner System, and images were then quantified using the GenePix Pro 4.0 application (Axon Instruments Inc.). After raw data normalization with the “Lowess” routine (28, 29), multiple testing to detect differentially expressed genes was carried out using the significance analysis of microarrays (SAM) method (30). One class response analysis was carried out, allowing for 2,000 random permutations.

**Prediction of Surface Extracellular Proteins**—Different bioinformatics algorithms were used for protein selection and classification. The Sosui system (classification and secondary structure prediction of membrane proteins, sosui.proteome.bio.tuat.ac.jp) was used to discriminate membrane from soluble proteins. To perform gene ontology analysis we used the Gene Ontology website (www.geneontology.org/) that uses controlled vocabularies (ontologies) describing gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner. Our search was achieved by tracking concepts such as “extra-
cellular, growth factor, membrane, and receptor."

cDNA Clones—Twenty-nine up-regulated genes were selected to express their corresponding protein products. Three sources of cDNA clones were used as starting material for cloning and expression of the activated genes: initially cDNA clones for the whole set of targets were obtained from the 40 K collection from the IMAGE Consortium (Research Genetics). The requisite for attempting expression was to contain at least 30% of the ORF. Another 24 full-length cDNA clones were obtained from the MGC and purchased from MRC Geneservices. SULF1 and thrombospondin 2 (THBS2) cDNA clones were a kind gift from Dr. Rosen (University of California San Francisco) and Dr. Bernstein (University of Washington), respectively. For three clones (COL5A2, ITGA2, and SLC2A1) we were unable to obtain a full-length sequence cDNA clone. All the clones were resequenced to verify the inserted sequence by using an automatic DNA sequencer.

**Primer Design and Cloning in the Gateway System**—For amplification of the coding regions, oligonucleotides were designed in such a way that the regions coding for the leader sequences of the proteins were removed to facilitate expression in E. coli. For identification of signal peptides we used SMART (Simple Modular Architecture Research Tool), which allows for a rapid identification and annotation of signaling domain sequences (smart.embl-heidelberg.de/). For a fast cloning of the genes into the Gateway system we used the directional pENTR/D-TOPO vector (Invitrogen). 5′-Oligonucleotides were designed to place the TOPO leader sequence (CACCC) immediately upstream of the ATG initial codon, resulting in primers with an average size of 28 bp. 3′-Oligonucleotides were designed either to contain the ORF stop codon in full-length clones or to introduce a novel stop codon for partial fragments. Oligonucleotides were purchased from Sigma Genosys. Coding regions were amplified from the cDNA clones by PCR with Vent DNA polymerase (New England Biolabs). For the cloning, 1 μl of the PCR product was added to 1 μl of the TOPO vector, left at room temperature for 5 min, and used to transform E. coli Top10 competent cells (Invitrogen). Every positive colony was fully sequenced to discard any mutation due to PCR amplification or TOPO recombination. Then positive colonies were used for plasmid DNA purification.

Destination plasmids pDEST17 (His_{6} tag) and pDEST15 (GST tag) were purchased from Invitrogen; plasmid pTH1, for MBP fusions, was a kind gift of Dr. Bernstein (University of Washington), respectively. For three clones (COL5A2, ITGA2, and SLC2A1) we were unable to obtain a full-length sequence cDNA clone. All the clones were resequenced to verify the inserted sequence by using an automatic DNA sequencer.

**Protein Expression and Solubility in E. coli**—Single colonies were grown at 37 °C in LB medium supplemented with 100 μg/ml ampicillin to midlog phase (A_{600} = 0.6–0.8) where expression was induced by the addition of 0.4 mM isopropyl 1-thio-β-d-galactopyranoside (Roche Applied Science). After a 2-h induction, cells were harvested by centrifugation and resuspended in 0.05 culture volumes of PBS. Three freeze/thaw cycles followed by sonication on ice were used to disrupt the cells, and the soluble and insoluble fractions were separated by centrifugation. The levels of expression and solubility were measured by Coomassie-stained SDS-PAGE and immunoblotting. For SDS-PAGE analysis, 10% SDS-polyacrylamide gels were run, and the resolved proteins were stained with Coomassie Blue G250 (Bio-Rad). For immunoblotting, proteins were transferred onto a Hybond-C nitrocellulose membrane (Amersham Biosciences) and blocked with 3% skimmed milk in PBS containing 0.05% Tween 20 for 1 h at room temperature. Then membranes were incubated with an anti-tag-specific antibody for 2 h at room temperature. Subsequently HRP- or alkaline phosphatase-labeled secondary antibody was added for 1 h at room temperature and visualized using ECL (Amersham Biosciences) or nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad) substrate, respectively.

**Purification of His_{6}, GST−, and MBP-tagged Proteins**—Proteins with His_{6} fusion were purified by affinity chromatography using a HiTrap™ column (Amersham Biosciences) on an ÄKTA Prime system (Amersham Biosciences) according to standard procedures. When insoluble His_{6} proteins were purified, an on-column refolding method was applied: 6 M guanidine hydrochloride, 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was used as binding buffer; 6 M urea, 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was the solubilization buffer; and 20 mM Tris-HCl, 0.5 M NaCl, 0.5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was the elution buffer.

For GST fusion proteins we used a GSTrap™ column (Amersham Biosciences). The binding buffer was 20 mM sodium phosphate, pH 7.3, with 0.15 M NaCl, and the elution buffer was 50 mM Tris-HCl, pH 8.0, with 10 mM reduced glutathione. In both cases, fusion proteins were pooled and dialyzed against PBS.

Soluble E. coli-derived MBP fusion proteins were purified by passing the extracts over an amyllose resin column (New England Biolabs) equilibrated in 20 mM Tris HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA and recovered by elution with 10 mM maltose in the same buffer. Fractions were analyzed on 10% SDS-polyacrylamide gels stained with Coomassie Brilliant Blue.

**Protein Identification by MALDI-MS**—Identity of the recombinant proteins was confirmed by MALDI-TOF mass spectrometry analysis. The protein spots were excised manually and transferred into siliconized 0.5-ml tubes. The gel pieces were washed twice with 50% acetonitrile. Then the gel fragments were placed at 56 °C for 45 min in 10 mM DTT, 55 mM iodoacetamide in 25 mM ammonium bicarbonate in the dark. Approximately 10 μl of 0.1 μg/μl modified trypsin (Promega) in 25 mM ammonium bicarbonate was added to the gel fragments and incubated overnight at 37 °C. After the supernatant was transferred to an Eppendorf tube, 20 μl of 50% acetonitrile, 0.1% trifluoroacetic acid was added, and the peptides were further extracted from the gel piece by sonication for 5 min and dried down. Peptides were reconstituted in 10 μl of 33% acetonitrile, 0.1% trifluoroacetic acid.

For MS analysis, a MALDI-TOF mass spectrometer (Autoflex, Bruker Daltonics) was used in positive ion reflector mode. The ion acceleration voltage was 20 kV. Each spectrum was internally calibrated with the masses of two trypsin autolysis products. For peptide mass fingerprint identification, the tryptic peptide mass maps were transferred through the MS BioloTools™ program (Bruker Daltonics) as inputs to search Swiss-Prot using Mascot software (Matrix Science). Up to one missed tryptic cleavage was considered, and a mass accuracy of 50 ppm was used for all tryptic mass searches.

**Antibody Production—BMP4**—clone ALB190F) and LCN2 (clone HAT265B)-specific monoclonal antibodies were produced as described previously (31) using partial-length MBP-BMP4 and full-length GST-LCN2 antigens, respectively. Hybridoma supernatants were screened initially by ELISA, immunoblotting on transfected cells and cell line and human tissue extracts. Finally they were tested by immunohistochemistry on bone marrow and spleen. Polyclonal antibodies against ANXA3 were produced by immunizing New Zealand White rabbits with the His fusion of the protein.

**Immunoblotting Analysis of Tissue Samples**—Frozen tissues were washed twice with chilled PBS, and proteins were extracted by son-
Validation of Genomic Targets in Colorectal Cancer

**Table II**

Summary list of the 29 selected target proteins with information regarding the availability of partial- and full-length sequence clones.

| Protein name          | Gene symbol | Molecular mass | MBP fusion proteins | GST fusion proteins | His<sub>6</sub> fusion proteins |
|-----------------------|-------------|----------------|---------------------|---------------------|-------------------------------|
|                       |             | Full length    | Partial length      |                     |                               |
| Collagen type V α 2   | COL5A2      | NA             | BC                  | –                   | –                             |
| Integrin α 2          | ITGA2       | NA             | BC                  | –                   | –                             |
| Protein-tyrosine kinase | PTK7      | 118,391        | BC                  | +                   | +                             |
| Ephrin B3             | EPHB3       | 106,520        | BC                  | +                   | +                             |
| Sulfatase 1           | SULF1       | 101,026        | BC                  | –                   | +                             |
| Solute carrier family 26 member 3 | SLC26A3 | 84,671        | BC                  | –                   | +                             |
| Solute carrier family 2 member 1 | SLC2A1 | NA             | BC                  | –                   | +                             |
| Angiotensinogen precursor | AGT        | 53,154        | BC                  | +                   | +                             |
| Carboxic anhydrase IX | CA9         | 49,698         | 15,653              | –                   | –                             |
| Plasminogen activator | PLAU        | 48,525         | BC                  | +                   | +                             |
| Thrombospinon 2       | THBS2       | 47,898         | +                   | +                   | +                             |
| Bone morphogenetic protein 4 | BMP4  | 44,048        | 23,867              | +, + (F, P)         | +, +                          |
| Selenoprotein P 1     | SEPP1       | 42,705         | BC                  | +                   | +                             |
| Annexin 3             | ANXA3       | 36,375         | BC                  | +                   | +                             |
| Osteopontin           | SPPI        | 35,422         | 21,523              | +, +                 | +, +                          |
| Osteonectin           | SPARC       | 34,632         | BC                  | +                   | +                             |
| Insulin-like growth factor-binding protein 3 | IGFBP3 | 31,674         | BC                  | –                   | –                             |
| Matrix metalloproteinase 7 | MMP7   | 29,677       | 19,429              | +, +                 | +, +                          |
| Collagen triple helix repeat-containing 1 | CTHRC1 | 26,224       | 11,813              | +, +                 | +, +                          |
| Matrix metalloproteinase 11 | MMP11 | 23,628       | 13,956              | –, +                 | –, +                          |
| Tissue inhibitor of metalloproteinase 1 | TIMP1 | 23,171         | BC                  | +                   | +                             |
| Lipocalin 2           | LCN2        | 22,588         | 15,488              | +, +                 | +, +                          |
| Teratocarcinoma-derived growth factor 1 | TDGF1 | 21,168         | BC                  | +                   | +                             |
| Interferon-induced transmembrane protein 3 | IFITM3 | 14,632       | 14,148              | –, +                 | –, +                          |
| Interferon-induced transmembrane protein 2 | IFITM2 | 14,546       | 12,452              | +, +                 | –, +                          |
| Interferon-induced transmembrane protein 1 | IFITM1 | 13,964       | 14,364              | –                   | –                             |
| GRO3 oncogene         | CXCL3       | 11,342         | BC                  | +                   | +                             |
| GRO1 oncogene         | CXCL1       | 7,864          | BC                  | +                   | +                             |
| Hypoxia-inducible protein 2 | HIG2   | 6,950          | BC                  | +                   | +                             |
| Total                 |             | 29             | 25                  | 11                  | 22                            |

26 distinct available proteins: 19 F, 18 F, 21 F, 9 P, 9 P

LCN2 was used undiluted, -ANXA3 was used at 1:1,000 dilution, -SPARC was used at 1:1,000 dilution, -SPPI was used at 1:10,000 dilution, -MMP7 was used at 1:2,000 dilution, and -MMP11 was used at 1:200 dilution. Anti-tubulin mouse antibody (Sigma) was used as a loading control at 1:5,000 dilution. Following three washes in PBS containing 0.05% Tween 20, membranes were incubated with a 1:1,000 dilution of either rabbit anti-mouse IgG HRP conjugate (DakoCytomation), or anti-rabbit IgG HRP conjugate (Sigma) for 90 min. Antibody binding was detected using ECL reagent or SuperSignal Femto (Pierce).

Tissue Microarray Preparation and Analysis—CRC tumoral samples corresponding to the three first stages of Dukes’ classification (A, B, and C), their normal mucosa counterparts, and tissue controls such as tonsil, breast, liver, lung, pancreas, and placenta were used for the preparation of the tissue microarrays. Neoplastic biopsies from two groups of 48 and 49 CRC patients and their non-tumoral mucosal counterparts were collected after tumor resection. Samples were collected and made anonymous by the Tumor Bank Network (CNIO). They were handled according to the ethical and legal standards. Diagnostic paraffin blocks were selected on the basis of the availability of suitable formalin-fixed paraffin-embedded tissue, containing enough remaining tissue for a minimum of 60 sections. Histological confirmation of CRC was achieved in all cases by central review using standard tissue sections, and the most tumor-rich areas were marked in the paraffin blocks. We used a tissue arrayer device (Beecher Instruments) to construct two microarray blocks containing a total of 97 samples, including replicates for some of the tumors and representing different locations of the neoplasia along the colonic region. Protein expression was graded 1–3 in the microarrays (1, <10%; 2, 10–50%; 3, >50% of stained tumor specific location). Two selected 1-mm-diameter cylinders from two different areas were included in each case along with several controls to ensure the quality, reproducibility, and homogenous staining of the slides. The arrays were incubated with antibodies against ANXA3, BMP4, and LCN2 at dilu-
tions of 1:500, 1:25, and 1:2, respectively. Anti-SPARC (1:25 dilution) was from Novocastra, anti-SPP1 (1:750) was from Abcam, and anti-MMP7 (1:200) and anti-MMP11 (1:50 dilution) were from Chemicon. K67 (1:50 dilution, DakoCytomation) was included as a control for proliferation. Specific binding was followed by incubation with prediluted commercial anti-mouse/rabbit IgG conjugated with biotin. Visualization of specific interactions was monitored by using the EnVision HRP system (Dako) following the manufacturer’s instructions.

RESULTS

Identification of Up-regulated Genes in CRC by DNA Microarray Analysis—Twenty-two paired samples of CRC patients were analyzed for differential gene expression by using a cDNA microarray (CNIO Oncochip). After one class SAM analysis, a total of 1,182 probes were found to be up-regulated at a delta value threshold corresponding to a q-value <1% (data not shown). At this lower limit, because the q-value established the lowest false discovery rate, the number of false positives in this set of selected genes was lower than 12. In addition, we established an upper limit of 0.6 in the 22-patient averaged log₂Ratio to filter out genes showing low global transcription increases in the tumors. In this way, 371 probes showed more than 1.5-fold change in CRC tumoral samples. The 371 probes matched 337 known genes and 13 hypothetical proteins (Supplemental Table S1).

Prediction of Surface and Extracellular Proteins as Putative Targets—Further selection criteria were based on subcellular location and gene ontology characteristics of the gene products. We preferentially selected those genes between the 100 most up-regulated genes among the 337 found genes with only a few exceptions (SLC2A1, HIG2, CA9, EPHB3, and ITGA2) that were placed between positions 100 and 200 (Supplemental Table S1). After bioinformatic analysis, a total of 29 proteins were selected for expression (Table II). Only proteins with cell surface accessibility or extracellular location were selected because these proteins would be accessible for antibody-mediated diagnosis or therapy. The range of molecular masses varied between 129 kDa for THBS2 down to 6.9 kDa for hypoxia-inducible gene 2 (HIG2). Many of those proteins do not have commercially available antibodies against them.

Expression of Partial-length cDNA Clones with Three Different Tags—Collecting and validating the cDNA clones for every gene was a time-consuming and lengthy process. The initial idea was to use the same cDNA clones from the IMAGE collection used for printing the Oncochip microarray as the source of ORFs for protein expression. Unfortunately only 12 of 29 clones covered at least 30% of the ORF sequence length. Only nine of 12 could be used for gene expression; the other three (ANX3, IFITM1, and TIMP1) consistently showed deletions after PCR amplification and cloning into pENTR/D-TOPO vector. A total of 27 destination clones were prepared after recombination of those nine entry clones with the three destination vectors. The efficiency of recombination was lower and more variable when E. coli BL21(DE3) cells were used for direct transformation as compared with DH5α cells: 70% versus >95% at the first attempt. However, for high throughput purposes, the overall time savings justifies direct transformation of BL21(DE3). The results for protein expression and solubility in E. coli are summarized in Table II and shown as Coomassie-stained gels (Fig. 1A) and immunoblotting analysis (Fig. 1B).

Soluble products were obtained for seven of nine MBP fusion proteins (78%), six GST fusion proteins (67%), and two His-tagged proteins (22%). These results confirm previous findings that MBP is a more efficient solubilizing agent for expression and recovery of soluble proteins in E. coli than GST or His₉ tags. The low efficiency of His₉ was surprising. No clone was soluble with the three tags, but all of them were correctly expressed and solubilized with at least one tag. In addition, BMP4, IFITM2, IFITM3, MMP11, and SPP1 were soluble with two tags. In summary, by combining the three tags all these protein fragments could be recovered in a soluble form.

Expression of Full-length cDNA Clones with the Three Tags—To get full-length ORFs, we used either the MGC (24 of 29 clones (82%)) or donations from other laboratories. Three cDNAs (COL5A2, ITGA2, and SLC2A1) were not found. A total of 26 clones containing full-length cDNAs were used for expression. After removal of the leader sequences, coding regions were PCR-amplified and cloned into the pENTR/D-TOPO vector with a variable efficiency between 15 and 100%. No entry clones were obtained for three clones (CA9, IGFBP3, and IFITM1). The entry clones were subsequently recombined with the same three destination vectors as above. Data for expression of the proteins in E. coli BL21(DE3) and solubility properties are summarized in Table III. Coomassie Blue and immunoblotting analyses are shown in Fig. 2, A and B, respectively. No destination clones were obtained for many of the large proteins (>50 kDa), suggesting an increasing difficulty for recombination. The efficiency of cloning ORF PCR products up to 2 kb is about 90% for the Gateway system but decreases considerably for larger clones (32).

In our hands, none of the proteins above 50 kDa, except SULF1, were soluble with any of the different tags when expressed in E. coli. MBP fusion proteins were soluble more often than fusions with GST or His₉. Interestingly all the proteins below 42.7 kDa (corresponding to SEPP1) were soluble when fused to MBP (13 of 19 (68%)). AGT and PLAU expression seems to correspond to the MBP tag alone. For GST, 5 of 18 (28%) were soluble. AGT and TIMP1 products seem to be mainly GST. In the case of His₉, solubilization was observed in 7 of 22 cases (32%), mostly for proteins with molecular masses below 40 kDa.

Purification of Fusion Proteins and MALDI-TOF Analysis—Fusion proteins with the three tags were purified for immunization purposes. Results are shown in Fig. 3. Soluble MBP fusion proteins were purified by using amylose resins in a batch procedure. A characteristic of those elutions was the
recurrent presence of minor bands underneath the correct one that probably arise from degradation fragments of the MBP. Although this fact might be disturbing for other uses of the protein, it is not the case for the preparation of antibodies as all of them are derived from the fusion proteins. Similar findings were obtained for purified GST-tagged proteins, which also showed the presence of several lower molecular weight bands. With respect to the His\(_6\) fusion proteins, five His\(_6\)-tagged proteins were purified under denaturing conditions followed by a refolding on-column. In general, although His fusions were recovered as soluble proteins in only a few cases, they usually yielded purer proteins in a single step.

Every fusion protein (partial- and full-length constructions) prepared with any of the three tags was subsequently analyzed by mass spectrometry to confirm its identity. Nineteen of 20 analyzed full-length expressed proteins were correctly identified as expected except for one case (SLC26A3) that corresponded to a non-visible band product. The whole set of partial-length products (nine of nine) were also positively validated.

Validation Studies by Immunoblotting—Seven proteins (ANXA3, BMP4, LCN2, SPARC, SPP1, MMP7, and MMP11) were initially selected as potential CRC markers for their novelty and relationship to some interesting pathways in tumor development and progression. To check whether those proteins were truly overexpressed in the samples, protein extracts from normal and tumoral tissues from six patients representing stages A–C of CRC were resolved by SDS-PAGE, blotted onto nitrocellulose membranes, and incubated with these antibodies. Fig. 4 shows the results obtained. ANXA3, MMP7, and SPARC antibodies were particularly efficient in tumoral versus normal differentiation with no apparent preference for either early or advanced stages. BMP4 recognized preferentially A and C stages, although normal samples also showed significant expression. A trend was observed for LCN2 with respect to the progression stage showing an increase in the expression for the more advanced stages corresponding to Dukes’ B and C series. MMP11 allowed for a significant discrimination between normal and tumoral samples, especially in stage C samples. The only molecule that could not be validated by immunoblotting as being differentially expressed in carcinoma samples was SPP1, which showed an erratic pattern along the whole series of samples. Tubulin was used as a positive control for normalization.
Validation of Genomic Targets in Colorectal Cancer

| Protein name                   | Gene symbol | Ontology                  | Full or partial length | Molecular mass | MBP fusion proteins | GST fusion proteins | His6 fusion proteins | Mass spectrometry analysis |
|-------------------------------|-------------|---------------------------|------------------------|----------------|---------------------|---------------------|----------------------|---------------------------|
| Protein-tyrosine kinase       | PTK7        | Signal transduction       | F P                    | 118,391        | – –                 | U U                 | – –                  | Da                        |
| Ephrin B3                     | EPHB3       | Angiogenesis              | F P                    | 106,520        | ++ 0                | ++ 0               | 0                    | +                        |
| Sulfatase 1                   | SULF1       | Apoptosis                 | F P                    | 101,026        | U U                 | 0                  | ++                   | +                        |
| Solute carrier family 26      | SLC26A3     | Sulfate transporter       | F P                    | 84,671         | U U                 | U U                | –                    | –                        |
| Angiotensinogen precursor 3   | AGT         | Blood pressure            | P P                    | 53,154         | * * *               | *                  | *                    | +                        |
| Carbonic anhydrase IX         | CA9         | Hypoxia marker            | F P                    | 49,698         | U U U U U U U       |                    |                      |                           |
| Plasminogen activator         | PLAU        | Growth factor             | F P                    | 48,525         | + + ++ + + + + +    |                     |                      |                           |
| Thrombospondin 2              | THBS2       | Angiogenesis              | F P                    | 47,898         | – – ++ ++ + + +     |                    |                      |                           |
| Bone morphogenetic protein 4  | BMP4        | Growth factor, cytokine   | F P                    | 44,048         | ++ 0 ++ ++ + + +    |                    |                      |                           |
| Selenoprotein P 1             | SEPP1       | Extracellular matrix      | F P                    | 23,867         | +++ ++ ++ ++ ++ +    |                    |                      |                           |
| Annexin 3                     | ANXA3       | Calcium ion binding       | F P                    | 36,375         | ++ ++ ++ ++ ++ +     |                    |                      |                           |
| Osteopontin                   | SP1         | Growth factor             | F P                    | 35,422         | – –                  | – –                | – –                  |                           |
| Osteonectin                   | SPARC       | Growth factor             | F P                    | 21,523         | +++ ++ ++ ++ ++ ++   |                    |                      |                           |
| Insulin-like growth factor-    | IGFBP3      | Growth factor             | F P                    | 34,632         | + + ++ – – – – –     |                    |                      |                           |
| binding protein 3              |             |                           | F P                    | 31,674         | U U U U U U U       |                    |                      |                           |
| Matrix metalloproteinase 7    | MMP7        | Extracellular matrix      | P P                    | 29,677         | +++ ++ ++ ++ ++ ++ 0 |                    |                      |                           |
| Collagen triple helix repeat-containing 1 | CTHRC1 | Extracellular matrix | F P | 19,429 | +++ ++ ++ ++ ++ 0 | * | * | + |
| Matrix metalloproteinase 11   | MMP11       | Extracellular matrix      | F P                    | 26,224         | +++ ++ ++ ++ ++ 0    |                    |                      |                           |
| Tissue inhibitor of metalloproteinase 1 | TIMP1 | Growth factor             | F P                    | 11,813         | + + ++ ++ ++ ++ 0    |                    |                      |                           |
| Lipocarnin 2                  | LCN2        | Oncogene                  | F P                    | 23,628         | U U U U U U ++ ++ ++ |                    |                      |                           |
| Teratocarcinoma-derived       | TDGF1       | Growth factor             | F P                    | 15,362         | ++ ++ ++ ++ ++ ++ ++ |                    |                      |                           |
| growth factor 1               | IFITM3      | Cell cycle                | F P                    | 15,488         | ++ ++ ++ ++ ++ ++ ++ |                    |                      |                           |
| Interferon-induced transmembrane 1 | IFITM2 | Cell cycle                | F P                    | 14,364         | U U U U U U U U U U |                    |                      |                           |
| Interferon-induced transmembrane 2 | IFITM1 | Cell cycle                | F P                    | 14,342         | ++ ++ ++ ++ ++ ++ ++ |                    |                      |                           |
| GRO3 oncogene                 | CXCL3       | Growth factor, cytokine   | F P                    | 13,964         | U U U U U U U U U   |                    |                      |                           |
| GRO1 oncogene                 | CXCL1       | Growth factor, cytokine   | F P                    | 11,342         | ++ ++ ++ ++ ++ ++ ++ |                    |                      |                           |
| Hypoxia-inducible protein 2   | HIG2        | Hypoxia marker            | F P                    | 7,864          | ++ ++ ++ ++ ++ ++ ++ |                    |                      |                           |

*a* Expression levels given as: +++ = strong band on SDS-PAGE gel; ++ = medium; + = visible band; – = no visible band; * = incorrect size.

*b* Solubility given as: +++ = most of the protein in soluble fraction; ++ = roughly 50% in soluble fraction; + = minority in soluble fraction; 0 = nothing in soluble fraction; – = no expression.

---

*Table III: Expression and solubility levels of 26 target proteins when cloned as three different gene fusions.*

No entry clones were obtained for three genes (CA9, IFITM1, and IGFBP3). From the remaining 23 genes with verified expression clones, tagged products matching 21 different proteins were expressed (full- and/or partial-length sequence). NA, no clone available; F, full-length cDNA sequence; P, partial-length cDNA sequence; E, expression; S, solubility; U, unsuccessful cloning attempt.
Applications of the Selected Proteins in CRC Diagnosis—

Protein expression in tumoral tissues was confirmed by immunohistochemistry on two distinct CRC-specific tissue microarrays containing 97 different paired human tissues (Fig. 5). Annexin A3 staining showed a clear preference for the tumoral tissues; an increase in the intensity of the signal was noticed for two-thirds of the cases (63%). In most cases there was a specific reactivity with epithelial cells surrounding the crypts and villi (51%), whereas the remaining tissues exhibited staining of stromal cells. Although specific staining for ANXA3 was detected in the cytoplasm, the clearest reaction was observed in the cellular membrane. Normal samples barely showed annexin A3 expression in epithelial cells.

BMP4 was detected throughout the cytoplasm of tumoral cells as well as in the extracellular matrix surrounding the mucosal cells, displaying a scattered, sometimes granular

Fig. 2. Expression-solubility profile of the selected full-length cDNAs fused to three different affinity tags: MBP, GST, and His6. Growing and expression conditions were the same as those described for partial-length proteins. A, Coomassie Blue-stained gels of soluble (S) and insoluble (I) fractions of the recombinant full-length protein named at the top of each two lanes. B, detection of fusion proteins by immunoblotting with antibodies against the corresponding tags: MBP, upper panel; GST, middle panel; and His6, lower panel. Marker sizes are given in the left side of each panel. The expected size of each protein is indicated by an arrow on the right side of the S lanes.
staining in agreement with the condition of BMP4 as a secreted protein. In a minor percentage of the cells, nuclei were also marked; this is in accordance with the reported translocation of BMP4 for the activation of the Smad pathway. Normal tissues gave a weak staining confined to some areas in the stroma and not in the epithelial cells.

Within CRC samples, there were a high number of epithelial cells expressing LCN2. The staining was cytoplasmic with a clear reinforcement of the membrane signal. The LCN2-positive cells were homogeneously distributed in the luminal area of the microvilli with a distinctive staining of their apical epithelium. Expression of LCN2 was markedly detected in the sigmoid colon (78% of the cases) and sigmoid-rectum (83%) where 57% of the positive cases were intermediately or strongly stained. With respect to other colon locations (such as cecum; ascending, transverse, and descending colon; and rectum), the reactivity and intensity varied among the cases, ~50% of which stained positively. No LCN2 protein expression was observed in the normal mucosa tissues tested.

In the case of MMP7, the staining of the tumoral cells was cytoplasmic, with a strong signaling marking the Golgi apparatus, and preferentially located in the apical regions of the neoplasia, next to the lumen, in focal locations at the end of the crypts or in areas of incipient metastasis. Apart from these focal nodes, MMP7 antibody stained the stroma surrounding the epithelial cells to some extent even in normal mucosa, giving the average tissue slice a weak background. Associated to MMP7 and sharing a similar pattern, SPP1 staining appeared in only a small percentage of the cases (13%) and usually was associated with focal expression near the cavities of the glands or blood capillaries. The signal in positive tumoral cells is visible throughout the cytoplasm, and because SPP1 is a secreted protein, a faint expression is also found in the extracellular matrix even in normal tissues in some cases.

MMP11 showed a juxtatumoral staining of stromal cells in CRC tissues that is probably due to the role of MMP11 in invasiveness. No staining was observed in normal tissues. A similar observation was made for SPARC for which the stroma subjacent to the cells was clearly stained in neoplastic samples, was substantially reduced in dysplasias, but was not at all present in hyperplasias and normal tissues. The SPARC labeling in juxtatumoral stromal cells developed as a strong and homogeneous cytoplasmic staining.

As a marker for proliferation, we used Ki67. A strong staining of the carcinoma cells was observed in the tumoral tissues.
DISCUSSION

After gene expression analysis of CRC samples using DNA microarrays, a collection of up-regulated genes was investigated for their ability as biomarkers for CRC. There were several objectives in this study: first, to investigate the possibilities to translate as fast as possible these genomic targets into soluble proteins; second, to test the recombinant proteins for the preparation of monoclonal and/or polyclonal antibodies; third, to check these antibodies for the validation of the genomic results; and fourth, to investigate and characterize these potential new markers for CRC. The selection of up-regulated CRC genes was non-biased, that is, independent of gene size, location, and biochemical function. We did not exclude large molecular weight proteins, integral membrane proteins, or secreted proteins. On the contrary, we specifically focused on these (membrane and secreted) proteins as more likely to be biomarkers of interest and targets for antibody diagnostics and therapy.

The initial idea was to use for protein expression the cDNA clones used regularly for the printing of DNA chips. However, these clones consisted of long untranslated regions, and none of these “chip” clones contained a full-length cDNA of the coding region. In many cases the ORF did not even reach the 30% that we had set up as our lower limit for expression experiments. This process of sequence verification for each clone already consumed a significant amount of time. As an alternative, full-length clones were obtained from the MGC repository or from colleagues after intensive search. For 30 genes the whole collection process was lengthy and time-consuming, making the retrieval of cDNA clones one of the bottlenecks of the whole process. Availability of public repositories containing full-length and sequence-verified ORF clones in recombinational cloning systems such as Gateway or Creator would be extremely beneficial for a high throughput expression project (33).

To speed up the cloning process, we selected TOPO-Gateway vectors in which the topoisomerase enzyme has been attached to the free ends of the vector. TOPO vectors present several advantages over pDONR vectors. They need shorter oligonucleotides for recombination; therefore, fewer foreign residues are introduced, and the cloning efficiency is higher. About 90% of pENTR constructions were generated at the first attempt. In the remaining 10% of the clones, small deletions were observed that obliged us to sequence every clone, slowing down the process. For instance, the entry clone for IFITM1 showed an internal 5-nucleotide deletion, which aborted the cloning process at an early stage. Another limitation of the Gateway system is the need to perform up to three transformation steps in *E. coli*, one for the entry clone and two more for the destination vectors, first in DH5 and then in BL21 strains. With our procedure, we skipped one step by transforming *E. coli* BL21 directly with the LR Clonase prod-

![Fig. 5. Tissue microarray analysis of the selected targets in colorectal cancer.](image)

Selected images of colorectal cancer tissue microarray elements representing immunohistochemical staining of the selected seven proteins (ANXA3, BMP4, LCN2, MMP7, MMP11, SPARC, and SPP1) at different magnifications (10× and 40×) in tumoral tissues and normal tissues are shown. Control Ki67 preferentially marked those cells undergoing proliferation, a distinctive label for neoplastic tissues. A hematoxylin/eosin (H/E) staining of section controls is also included.
uct. Ninety-five percent of the attempted transformations succeeded at the first round, and the rest were obtained in a second attempt.

At the present time, it is impossible to predict the results of protein expression. There were some previous attempts to predict expression probabilities based on the characteristics of the protein family and the Pfam domains with little success (34). In the present study, our observation is that, by using this high throughput approach, only proteins below 50 kDa were expressed as soluble proteins with at least one of the tags; this constitutes a severe limitation for these genomic projects. MBP tag was the best choice for an increased solubility of the fusion proteins. The solubility decreased significantly in proportion to the size of the protein. The fact that MBP fusion proteins are soluble does not mean that they are correctly folded or functionally active (35, 36). However, it is evident that for many applications such as antibody production these soluble proteins represent a great advantage due to easier handling and a lower toxicity for the host animals. Moreover the possibility of developing antibodies against natural conformational epitopes is much higher using soluble proteins as shown by the reactivity obtained by the monoclonal antibodies developed against LCN2 and BMP4.

When proteins were fused to GST or His6 tag only 27.7 and 13.6% were recovered in a soluble form, respectively. In a previous study (20), a similar approach was followed to express 27 small human protein (below 20 kDa), obtaining similar solubility results with MBP and thioredoxin tags. However, thioredoxin is not appropriate for affinity purification and would require a second affinity tag for purification. Braun et al. (19) expressed and purified 32 human proteins using the same three tags plus calmodulin-binding protein. They described similar results of protein recovery for MBP and GST and did not provide details about solubility of the proteins. We believe that our procedure still needs some improvements in automation and especially in the purification steps, making it more amenable to high throughput, by using chemical lysis and microplate batch purification procedures as reported previously (19).

In summary, from the initially selected 29 genes, only 21 distinct proteins were finally expressed with, at least, one different fusion protein. With respect to the full- or partial-length sequence option, seven proteins were achieved in both formats, and 12 of them were obtained as complete sequence products, whereas the remaining two could only be obtained as protein fragments. According to these considerations and limitations, either new tags should be developed for a more efficient recovery of soluble proteins, or alternative eukaryotic systems should be optimized for high throughput approaches (37).

Recent advances in proteomic research underscore the increasing need for high affinity antibodies, which are still generated with lengthy, low throughput antibody production techniques. Other alternative routes such as phage display are currently being tested. Regarding antibody production, in our hands, no significant differences in time were observed between preparing monoclonal or polyclonal antibodies. Moreover the performance of the polyclonal antibody against ANXA3 equals or outperforms the use of monoclonal antibodies for BMP4 or LCN2. Therefore, the final method of choice should be based in terms of availability or accessibility to antibody production facilities if pure antigens are on hand. Probably for a first screening, polyclonal antibodies may represent a cheaper alternative.

To test the usefulness of this approach for marker validation, we chose seven proteins for antibody production and/or validation. These seven potential markers were selected for their novelty and/or relationship to some interesting pathways in CRC tumor development, invasiveness, and progression. ANXA3 encodes a member of the annexin family. Members of this calcium-dependent phospholipid-binding protein family play a role in the regulation of cellular growth and in signal transduction pathways. There are no previous reports of association of this protein to tumoral processes. Lipocalin 2, also known as neutrophile gelatinase-associated lipocalin, is a 25-kDa protein stored in specific granules of the human neutrophile (38). It might function as a modulator of inflammation. The LCN2 gene is highly homologous to the mouse oncogene 24p3. Potential cis-acting elements include binding sites for transcription factors GATA-1 and PU.1 and NFκB. LCN2 was found to be highly expressed after malignant transformation of the breast, lung, colon, and pancreatic epithelia by immunohistochemistry (39). More recently, it has been described that LCN2 decreases the invasiveness and metastasis of Ras-transformed cells (40). BMP4 is a member of the bone morphogenetic protein family, which is part of the transforming growth factor-β superfamily. This superfamily includes large families of growth and differentiation factors. Recent experiments revealed that the oncogenic allele of β-catenin is required for BMP4 expression and secretion by human cancer cells and that BMP4 is overexpressed and secreted by human colon cancer cells with mutant adenomatous polyposis coli (APC) genes. These data identify the presence of regulatory interactions between the Wnt and BMP signaling pathways in cancer pathogenesis, providing an intriguing connection between the sporadic and inherited forms of a common human malignancy (41). BMP4 also promotes melanoma cell invasion and migration (42). SPARC, also called osteonectin, is a multifunctional matricellular glycoprotein that has been associated with impaired tumor growth, antiangiogenic properties, apoptosis induction, and changes in the extracellular matrix (43). SPP1, also called osteopontin, is overexpressed in a wide number of tumors, inducing various cellular signaling events leading to the activation of various kinases, urokinase plasminogen activator, and matrix metalloproteases (44). It plays a crucial role in tumor progression. MMP7 and MMP11 are members of the MMP family, which is involved in the breakdown of extracellular matrix in normal physiological processes, such as em-
bryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. MMP7 is elevated in several human cancers and seems to be related to CRC liver metastases (45). MMP11 protein is involved in the pathway of colorectal cancer development in females, distal locations, infiltrative growth patterns, and microsatellite stability (46).

Our results have proven that there is a good correlation between the transcriptomic analysis and the protein expression data for the four genes tested. There was a clear overexpression in most cases; overexpression was particularly significant in the cases of ANXA3, SPARC, MMP7, and LCN2 for making a distinction between tumoral and normal tissues. Remarkably about 73% of all the cases showed LCN2 reactivity (they were more strongly staining Dukes’ B stages). MMP11 also showed a tendency to be found in late stages. The staining pattern of some of these partners, mainly in the stromal tissue, stresses the importance of considering the whole tissue for evaluations because some proteins are differentially expressed in stromal cells surrounding and adjacent to regions of diseased epithelium that correlate with tumor progression (47).

In summary, we have demonstrated the feasibility of this approach for a relatively quick expression and validation of a high number of genomic targets. We have identified some major bottlenecks and suggested ways to overcome them. Although this preliminary characterization was only restricted to seven targets, it is clear that approaches like this may yield relevant biological information about the neoplastic processes and lead to the characterization of potentially interesting markers in a quite straightforward manner for early diagnosis or individual prognosis.

Acknowledgments—we thank Manuel Morente from the CNIO Tissue Bank Network and Jose Solano from the Pathology Department of the Hospital of Cartagena for cooperation in the collection of CRC samples. We thank A. Nuñez, S. Romero, F. Fernandez, and P. Alfonso for technical assistance and the staff of the monoclonal antibody unit of the CNIO for assistance in the preparation of the different antibodies. We also thank the technical staff of the histology and immunohistochemistry unit of the CNIO for assistance in the preparation of the tissue microarrays and immunohistochemical staining.

* This work was supported in part by Spanish Ministry of Science Grant BIO2003-01481. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

REFERENCES

1. Fearon, E. R. (1996) Colorectal cancer: molecular genetic studies and their future clinical applications. *Med. Pediatr. Oncol. Suppl.* 1, 35–40.
2. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
3. Lockhart, D. J., Dong, H., Byrne, M. C., Follette, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675–1680.
4. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J. Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.
5. Defilsi, J., Peniland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14, 457–460.
6. Williams, N. S., Gaynor, R. B., Scoggin, S., Verma, U., Gokaslan, T., Simmang, C., Fleming, J., Tavana, D., Frenkel, E., and Becerra, C. (2003) Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin. Cancer Res.* 9, 931–946.
7. Zou, T. T., Selaru, F. M., Xu, Y., Shustova, V., Yin, J., Mori, Y., Shibata, D., Sato, F., Wang, S., Oliaru, A., Deacu, E., Liu, T. C., Abraham, J. M., and Meltzer, S. J. (2002) Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* 21, 4855–4862.
8. Birkenkamp-Demtroder, K., Christensen, L. L., Olesen, S. H., Frederiksen, C. M., Laiho, P., Aaltonen, L. A., Lauberg, S., Sorenson, F. B., Hagemann, R., and Omhoff, T. F. (2002) Gene expression in colorectal cancer. *Cancer Res.* 62, 4352–4363.
9. Notterman, D. A., Alon, U., Sierk, A. J., and Levine, A. J. (2001) Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res.* 61, 3124–3130.
10. Walsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A., and Hampton, G. M. (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1176–1181.
11. Walsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2001) Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* 61, 5974–5978.
12. Hellstrom, I., Raycraft, J., Hayden-Ledbetter, M., Ledbetter, J. A., Schummer, M., McIntosh, M., Drescher, C., Urban, N., and Hellstrom, K. E. (2003) The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res.* 63, 3695–3700.
13. Alfonso, P., Nuñez, A., Madoz-Gurpide, J., Lombardia, L., Sanchez, L., and Casal, J. I. (2005) Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. *Proteomics* 5, 2602–2611.
14. Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B., Levy, S. E., Coffey, R. J., and Caprioli, R. M. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 4, 793–811.
15. Olesen, S. H., Christensen, L. L., Sorenson, F. B., Cabezon, T., Lauberg, S., Omhoff, T. F., and Birkenkamp-Demtroder, K. (2005) Human FK506 binding protein 65 is associated with colorectal cancer. *Mol. Cell. Proteomics* 4, 534–544.
16. Tracey, L., Villuendas, R., Ortiz, P., Dopazo, A., Sptier, I., Lombardia, L., Rodriguez-Peralto, J. L., Fernandez-Herrera, J., Hernandez, A., Fraga, J., Dominguez, O., Herrero, J., Alonso, M. A., Dopazo, J., and Pins, M. A.
(2002) Identification of genes involved in resistance to interferon-α in cutaneous T-cell lymphoma. Am. J. Pathol. 161, 1825–1837.

17. Strausberg, R. L., Feingold, E. A., Klausner, R. D., and Collins, F. S. (1999) The mammalian gene collection. Science 286, 455–457.

18. Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaeffer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F. D., Liang, Y. S., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodríguez, A. C., Crimwood, J., Schmutz, J. M., Myers, R. M., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Diatchenko, S. F., Zeeberg, B., Buetow, K. H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wetter, K., Werthagen, U., Xu, L. L., Hobé, S., and Ponten, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol. Cell. Proteomics 4, 1920–1932.

19. Tyers, M., and Mann, M. (2003) From genomics to proteomics. Mol. Cell. Proteomics 2, 318–324.

20. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

21. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

22. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

23. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

24. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

25. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

26. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

27. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

28. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

29. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

30. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

31. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

32. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

33. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

34. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

35. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

36. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

37. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

38. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

39. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

40. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

41. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

42. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

43. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

44. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.

45. Calito, A., Meyer, C., Jessen, K., and Mann, M. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Expr. Purif. 23, 111–119.