Mesothelin (MSLN) may be the most “dramatic” of the tumor markers, being strongly overexpressed in nearly one-third of human malignancies. The biochemical cause is unclear. We previously ascribed this cancer-specific overexpression to an element, Canscript, residing around 50 bp 5′ of the transcription start site in cancer (Hucl, T., Brody, J. R., Gallmeier, E., Iacobu-zio-Donahue, C. A., Farrance, I. K., and Kern, S. E. (2007) Cancer Res. 67, 9055–9065). Herein, we found a Canscript promoter activity elevated over 100-fold in cancer cells. In addition to a highly conserved TEAD1 (TEA domain family member 1)-binding MCAT motif, nucleotide substitution revealed the consensus core sequence (WCYCCACCC) of an SP1-like motif in Canscript. The unknown transcription factor binding to the SP1-like motif may hold the key for the cancer specificity of Canscript. SP1, GLI1, and RUNX1, -2, and -3 appeared unlikely to be the direct transcription factors acting at the SP1-like motif, but KLF6 had some features of such a candidate. YAP1, a TEAD1-binding protein, appeared necessary, but not sufficient, for Canscript activity; knockdown of YAP1 by small interfering RNAs greatly reduced MSLN levels in MSLN-overexpressing cells, but overexpressing YAP1 in MSLN-negative cells did not induce MSLN expression. Canscript-like sequences were found in other genes up-regulated in pancreatic cancer; reporters driven by the sequences from FXYD3, MUC1, and TIMP1 had activities more than 2 times that of the control. This suggested that the cause of MSLN overexpression might also contribute mechanistically to the overexpression of other tumor markers.

Tumor-overexpressed and tumor-specific biomarkers could serve as sensitive diagnosis tools and specific targets for therapeutic purposes. Deciphering the differential transcription regulation of tumor markers in cancers may further reveal the pathways that drive the origin and malignancy of cancers. There were a few advanced attempts to discover cancer-related transcription factors by studying promoters of tumor markers. For example, RREB-1 (Ras-responsive transcriptional element-binding protein) was proposed as an up-regulator of calcitonin in thyroid cancer cells (2). Transcriptional studies of some widely applied tumor markers, such as AFP (3) and CEA (4), were stalled at the promoter-identification stages, whereas for most other markers, such as CA125 and CA19-9, exact promoters were not revealed.

MSLN is a 40-kDa glycosylphosphatidylinositol-linked membrane glycoprotein (5), the C-terminal end cleaved from a 71-kDa MSLN precursor protein by the endoprotease furin (6, 7). MSLN transcripts are only expressed in mesothelial cells under normal conditions (6). As a tumor marker, MSLN is overexpressed in 70% of ovarian epithelial tumors (8), 50% of lung cancer (9), and almost all of the ducal pancreatic adenocarcinomas (10, 11). Interestingly, in pancreatic precursor neoplasms, MSLN is only detected in the latest, highest risk stage (12).

As a cancer cell membrane marker having very limited expression in normal tissues, MSLN is an ideal target for therapeutic potential. For example, an anti-MSLN monoclonal antibody was covalently linked to Pseudomonas exotoxin A to form the toxic antibody SS1P (13), which was recently explored in two phase I clinical trials (14, 15).

Initial studies of the MSLN promoter were inconclusive (16). An exon/intron map of the MSLN gene was created by aligning a single clone of MSLN cDNA (HS335H7) with fragments of genomic sequence. An arbitrarily chosen 1850-bp genomic DNA fragment (MP1850) 5′ of the predicted exon 1 (Fig. 1A) was studied in the JU77 mesothelioma cell line. To further define MSLN-specific promoter and enhancer elements, we reported the differing control sequences active in (epithelial) carcinomas (1).

In carcinomas, exon1 (1) was much longer than predicted (16), providing evidence that MSLN utilized different promoters or TSSs2 in cancers. We reported a 20-bp sequence (Canscript, TCTCCACCCACACATTCCTG) that appeared responsible for MSLN expression in certain cancer cells (1) according to Fig. 4A of Hucl T. et al. (1), the text description of the Canscript sequence on page 9059 should be −65 to −46).

The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2–S8.

1 To whom correspondence should be addressed: CRB1 451, 1650 Orleans St., Baltimore, MD 21231. Tel: 410-614-3314; E-mail: sk@jhmi.edu.

2 The abbreviations used are: TSS, transcription start site; RACE, rapid amplification of cDNA ends; RLA, relative luciferase activity.
Structural Analysis of Canscript

The element of Canscript matches the conventional MCAT sequence (CATTCCCT) in humans to chickens (17). We prefer to specify an extra 5’ A in the MCAT element because this A is part of MCAT in promoters regulating many other genes (18, 19). We reported that the Canscript MCAT element (Fig. 1A) specifically recruited the TEAD1 transcription factor but not TEAD2 and TEAD3 in cancer cells (1). Because TEAD1 is ubiquitously expressed in almost all of the tissues and cell lines (supplementary Fig. S1), the MCAT element is not suspected to provide the cancer specificity of Canscript.

Possibly, a TEAD1-binding cofactor could provide this specificity. Among known TEAD1-binding proteins, YAP1 (supplementary Fig. S1), an oncogenic transcription cofactor (20), was a candidate. YAP1 is phosphorylated and inactivated by the tumor-suppressive Hippo pathway. Overexpression of YAP1 or its Drosophila homologous Yki has been shown to result in tumor-like overgrowth of tissues in the mouse or fruit fly (21).

Alternately, a transcription factor binding to the SP1-like element may provide the cancer specificity. SP1 itself was not a strong candidate; for the central “G” of the SP1 consensus binding sequence (CCCCGGCCC) (22) does not match the A in the Canscript sequence (differences are underlined). SP1 belongs to the Krüppel-like family, which contains 21 DNA-binding transcription factors (23). All factors from this family bind to GC-rich DNA elements by three C_{2}H_{2} zinc finger domains. Among KLF family members, the KLF6/CPBP (core promoter-binding protein) binding sequence (CCCCACCCCA) (24), and that of hedgehog pathway effector GLI1, an oncoprotein (GACCAACCACA) (25), were most similar to the SP1-like element of Canscript. The confined SP1-like element and linker region (CACCAACA) also shared similarities with the RUNX protein family binding consensus sequence (AACCACA) (26). RUNX2, also, is a YAP1-binding protein (27).

There exists only one perfect copy of the Canscript sequence in the human genome, but similar sequences were potentially functional by which to drive the expression of other tumor markers. We explored the possibilities with structure-function studies of such Canscript-like sequences.

EXPERIMENTAL PROCEDURES

Cell Lines and Cell Culture—AsPc1, HeLa, RKO, HEK293, and MSTO-211H cell lines were obtained from the American Type Culture Collection. Cells were cultured in RPMI (AsPc1 only) or DMEM medium, supplemented with 10% FBS and 1% penicillin/streptomycin.

5’-Rapid Amplification of cDNA Ends (5’-RACE)—Total RNA was extracted from AsPc1 and MSTO-211H cells using TRIzol (Invitrogen, 15596-018), followed by 5’-RACE according to the SMARTer RACE protocol (Clontech, 634923). Briefly, first-strand cDNAs were synthesized by a template-switching extension that attached an adapter sequence at the 5’-end of the RNA template. The cDNA was amplified by PCR with 3’ gene-specific and 5’ adapter-specific primers. The MSLN-specific primer (CCCCCAACAGGGGTCGAGGCCCCTTGGCAAG) was located within exon 2. The PCR product was gel-purified and cloned into the TOPO TA vector (Invitrogen, K4500-01), followed by sequencing. The PCR primers for validating the transcript variants were as follows: reverse primer, CGAGGCTGAAGACGAGGAA; forward primer 1, TGTTCCCTTTTGACGGGC; forward primer 2, AGGGCTCA-GTGGCTGGAGG.

Western Blot—Cells were washed once in PBS and lysed by radioimmunoprecipitation assay buffer supplemented with protease inhibitor mixture (Roche Applied Science, 1183617001), followed by sonication. Protein concentrations were determined by a DC protein assay (Bio-Rad, 500-011). Samples were boiled with loading buffer and resolved by SDS-PAGE. Proteins were transferred onto a PVDF membrane and blotted with the following antibodies: anti-MSLN (Abcam, ab-3362); anti-TEAD1 (Santa Cruz Biotechnology, Inc. (Santa Cruz, CA), sc-81396); anti-TEAD2 (Santa Cruz Biotechnology, Inc., sc-32427); anti-YAP1 (Santa Cruz Biotechnology, Inc., sc-101199); anti-Sp1 (Santa Cruz Biotechnology, Inc., sc-420); anti-Taz (Avia, ARP-32540); anti-KLF6 (Santa Cruz Biotechnology, Inc., sc-7158); anti-GLI1 (Santa Cruz Biotechnology, Inc., sc-20687); anti-RUNX1 (Santa Cruz Biotechnology, Inc., sc-101146); anti-Rixn2 (Santa Cruz Biotechnology, Inc., sc-101145); anti-RUNX3 (Santa Cruz Biotechnology, Inc., sc-130014); anti-tubulin (Santa Cruz Biotechnology, Inc., sc-8035); anti-His tag (Santa Cruz Biotechnology, Inc., sc-8036). HRP-linked secondary antibodies were from Santa Cruz Biotechnology, Inc. (anti-mouse, sc-2005; anti-rabbit, sc-2004; anti-goat, sc-2020). Membranes were developed using Immobilon substrate (Millipore, WBKLS0500).

Plasmids—pGL3-Basic (E1751), pGL3-SV40 (E1761), and pRL-SV40 (E2231) plasmids were from Promega. All wild type, mutant, and derivative forms of a triplicated Canscript (Can3) sequence were inserted between the MluI and BglII cutting sites of the pGL3-Basic vector. A 20-bp (TGTCGAACTTTTC-GATCAGA) transcriptionally inactive hygromycin coding sequence were inserted between the MluI and BglII cutting sites of the pGL3-Basic vector. A 20-bp (TGTCGAACTTTTC-GATCAGA) transcriptionally inactive hygromycin coding sequence was used as negative control in Fig. 4C. The YAPI gene was cloned into pcDNA Zeo vector (21). RUNX1-1, -2, and -3 cDNA clones were purchased from Open BionSystems and inserted into pRK5-HA vector (from BD Pharmingen). KLF6 cDNA was obtained by RT-PCR from AsPc1 cells and inserted into the pRK5-HA vector. Plasmids were purified by the Plasmid Midi kit (Qiagen) followed by sequencing confirmation.

Transfection—Transient transfections were done by using Lipofectamine (Invitrogen, 18324-012) using the manufacturer’s instructions. We seeded 2–4 × 10^5 cells in each well of a 6-well plate 1 day before transfection. pGL3-Basic luciferase vector (0.2 µg) and pRL-SV40 Renilla vector (20 ng) were transfected in each well. The transfection medium was replaced by growth medium after 6 h.

Luciferase Reporter Assay—We used the dual luciferase reporter assay system (Promega, E1910). Cells were harvested 24 h after transfection. The cells were lysed by passive lysis buffer (300 µl/well). Each sample was measured in duplicate wells using 20 µl of cell lysate mixed with luciferase reagent II (LARII; 100 µl/well). After recording the luciferase luminescent signals, Stop and Glo reagent was added (100 µl/well), followed by the Renilla luminescence reading. Luminescence signals were detected by a PerkinElmer Microbeta Trilux plate reader. All luciferase readings were normalized using the Renilla readings of the same well. The average relative luciferase activity (RLA) was obtained from the duplicate wells. The
difference between duplicate wells usually was less than 5%. These averages were used to construct comprehensive surveys of serial structural modifications to the tested sequence. Each experiment was performed twice, on different days. Each luciferase survey presented was representative of the two independent experiments. Results from two independent experiments were similar, but transfection efficiency and the scale of results differed between replicates such that averaging the replicates was precluded. Due to the multiple comparisons, differences in scale, and limitation of only two replicates, the presentation of error bars would not be a valid test of confidence and is omitted.

**siRNA Sequences and Transfection**—The siRNAs were synthesized by Ambion. An irrelevant siRNA (FANCD2, Irr) was systematically used as a negative control throughout this paper. Except for the previously studied TEAD1 and TEAD2 genes (1), three non-overlapping siRNAs were designed for each target gene. An immunoblot was used to verify siRNA efficiencies.

The sense sequences were as follows: FANCD2 (Irr)si, CCAUGCUCUUAGUAGCGGATT; TEAD1 si, GCCCUGUUUCUAAUGUGGTT; TEAD2 si, GGACGCGAGAUUUGUGUACCTT; YAP si1, GUCAGACGAGUACUUAATT; YAP si2, GCCUUGAGUCUCACUCCCTT; YAP si3, CCAGAGACUCACAGGAGUU; SP1 si1, GCAACATGGGAATTATGAATT; SP1 si2, GGCAAGGCTTACATCTCATT; SP1 si3, GCAAGACCCCAAACACCTCTT; KLF6 si1, GAAGACUCAGUGACCAAAA; KLF6 si2, GCAGAGAGUUUACACCAA; KLF6 si3, GAGCUUUGUAAACCUUA. siRNA transfections were done in HeLa cells by Oligofectamine (Invitrogen, 12252011). The protocol was the same as for transient transfection except that 0.2 nM siRNA and 4 μL of oligofectamine were added into each well. The siRNA transfection efficiency in AsPc1 cells was very low, precluding the studies in AsPc1.

**Apoptosis Analysis**—The PE Annexin V apoptosis detection kit was used (BD Pharmingen, 559763). Briefly, cells were

---

**FIGURE 1. MSLN gene structure and Canscript constructs.** A, Canscript sequence, MSLN gene structure, and splicing variants. The most 5’ TSS (1) of MSLN is defined as +1. Exon 1 variant 1 comprises +1 to +438; exon 1 variant 2, +1770 to +1818; exon 2, +1936 to +2029; the ATG start codon, +1945; Canscript, −65 to −46; the MP1850 fragment (16), −1422 to +428. The arrowhead indicates the 5’-RACE primer location within exon 2. Exon 1 variant 1 only occurred in AsPc1 cells. B, vector maps of Canscript luciferase reporters used. Construct labels are given to the right of the luciferase element. pGL3-Basic and pGL3-SV40 are the negative and positive controls, respectively. LUC, luciferase gene; thick lines, MSLN gene sequences; thin solid lines, pGL3 backbone sequences; dashed lines, hygromycin sequences (Fig. 4C); thin dotted lines, eight Canscript-like sequences (Fig. 5).
washed in cold PBS twice, followed by resuspension in 1× binding buffer at a density of 1 × 10⁶/ml. Cells (in 100 µl) were mixed with Annexin V-PE (5 µl) and aminoactinomycin D (5 µl) and incubated in the dark for 15 min at room temperature before dilution (1× binding buffer, 400 µl). Cells were analyzed by a FACScan™ (BD Biosciences) flow cytometer.

Cell Cycle Analysis—Cells (5 × 10⁶) were washed in PBS once and resuspended in 0.5 ml of PBS. Cells were mixed with 4.5 ml of ice-cold 70% ethanol for 2 h at −20 °C. Cells were washed in 5 ml of PBS again and resuspended in 1 ml of stain (0.1% Triton X-100, 0.2 mg/ml RNase A, and 20 µg/ml propidium iodide in PBS) for 30 min at room temperature. Cells were analyzed by a FACScan™ flow cytometer.

RESULTS

Detection of TSSs in Contrasting Cell Lines by 5′-RACE—Because the MSLN gene might utilize different promoters under tissue-specific and cancer-specific conditions, we adopted

Structural Analysis of Canscript

FIGURE 2. Canscript properties. The RLA of pGL3-Basic was defined as 1. A, location of Canscript in the MSLN promoter region. Different lengths of sequences from the MSLN promoter region (from −5, −48, −67, −98, −135, and −3716 to +411) (1) were inserted before the luciferase gene in the pGL3-Basic vector (Fig. 1B). RLA of B-67 was much higher than B-48 in AsPc1 (MSLN−) but not in RKO (MSLN+) and MSTO-211H (mesothelioma) cell lines. B, Canscript was triplicated and inserted before the luciferase gene in pGL3-Basic vector (pGL3-Can3; Fig. 1C). pGL3-Can3 had much higher activity than the pGL3-SV40 control in AsPc1 and HeLa (non-mesothelioma MSLN−) cell lines. C, Canscript as a promoter. Reversing Canscript in the pGL3-Can3 vector (creating pGL3-Can3R) left RLA unchanged. Moving Can3 to 3′ of the luciferase gene in the pGL3-Basic vector (creating pGL3-eCan3) greatly reduced RLA.

5′-RACE to explore the TSSs of the MSLN gene in tissue-informative cells (211H mesothelial/mesothelioma cells) and cancer-informative cells (AsPc1 pancreatic cancer cells). Using reverse primers located in exon 2, we confirmed the transcription start region in AsPc1 cells described in our report (1) and identified an alternate minor TSS at +1770 (Fig. 1A) having an alternate exon 1. In 211H cells, only the TSS at +1770 could be detected, supporting exclusive use of the more 5′ promoter in mesothelium. Both two-variant splicing forms were confirmed by RT-PCR (supplemental Fig. S8). Sequencing results showed adherence to classical splicing rules.

Canscript Nucleotide Structure/Function Studies—To focus upon Canscript transcriptional activity, the sequences flanking Canscript in the MSLN gene (pGL3-B67; Fig. 1B), including

APRIL 8, 2011 • VOLUME 286 • NUMBER 14

JOURNAL OF BIOLOGICAL CHEMISTRY 11963
presumptive minimal promoter activities, were stripped away. The activity of this pGL3-Canscript was reduced (data not shown) but was restored by using triplicate Canscript tandem copies (pGL3-Can3; Fig. 1B). We then surveyed the relative strength and position dependence of the element. pGL3-Can3 was over 100 times as active as the pGL3-Basic vector in AsPc1 cells (Fig. 2, A and B). The activity of pGL3-Can3 was much higher than pGL3-SV40 vector and comparable with the MSLN full promoter. Moreover, reversing the orientation of the Can3 sequence left its activity unchanged (Figs. 1B and 2C). Can3 was thus sufficient to drive luciferase expression in the promoter-less pGL3-Basic vector (Figs. 1B and 2B). On the contrary, relocation of Can3 to 3′ of the luciferase gene in pGL3-Basic vector eliminated almost all of the relative luciferase activity (Figs. 1B and 2C), suggesting that nearly all of its activity was as a promoter rather than as an enhancer.

To rigorously define the required sequence in Canscript, a nucleotide transitional substitution survey was conducted of the MCAT motif in pGL3-Can3. The results showed that all eight nucleotides (nucleotides 12–19) were essential (Fig. 3A). Any transition mutation in this area reduced the Can3 activity by more than 80%. As a control, the transitional A substitution of the G after the MCAT motif did not affect Can3 activity.

Similarly, all three nucleotide substitutions were surveyed at each position in the SP1-like region. Some substitutions eliminated almost all of the Can3 activity (Fig. 3, B and C). The consensus sequence of the SP1-like element generated by the nucleotide substitution survey revealed an 8-nt minimal core WCYCCACCC (Fig. 3C). This sequence contains a binding core (CACCC) for some members in the KLF protein family (23).

We next studied the two-nucleotide (AC) linker between the SP1-like and MCAT motifs (Fig. 1A). Upon surveying a series of nucleotide substitutions and expansions of this linker, Can3 activity was retained in five of six substitutions (Fig. 4A), whereas the AC to GC substitution removed most Can3 activity. On the other hand, the expansion of this linker (AC) to six nucleotides (AGTGTC) or deletion of these two nucleotides did not affect the Can3-luc activity (Fig. 4B). When we further expanded the linker by inserting a transcriptionally inactive sequence (from the hygromycin coding sequence) between the A and the C (Fig. 1B), Can3 retained some activity even using
Canscript-like Sequences in Other Genes Up-regulated in Pancreatic Cancer—“Stretching” of the linker length between SP1-like and MCAT elements retained certain Canscript activity, and stretched “Canscript-like” sequences might exist in the promoter regions of other genes. We searched for such Canscript-like sequences by screening upstream sequences (5 kb upstream of each ATG start codon) of the top 52 overexpressed pancreatic neoplasia-associated genes from an unbiased meta-analysis (29). The MCAT sequence “CATTCCT” was required in the search. Only a one-nucleotide mismatch or insertion/deletion was permitted in the SP1-like region in accordance with the Canscript consensus sequence (only the nucleotide substitutions shown in Fig. 3C were allowed). We discovered seven additional sequences (in six genes; Table 1) having linker lengths of less than 40 bp. We then replaced the Canscript sequence with these fragments in the context of the MSLN promoter region (pGL3-Canscript-like vector; Fig. 1B) and tested their reporter activities in AsPc1, HeLa, RKO, and HEK293 cells. The pGL3-Canscript-like vectors incorporating sequences from FXYD3, MUC1, and TIMP1 had activities at least twice that of the matched empty vector in all four cell lines (Fig. 5A). To further validate this observation, we selected the four strongest sequences by performing seven independent transfections in AsPc1 cells. Each test sequence was compared with the Canscript-like control plasmid. The significant level with a conservative Bonferroni correction for four comparisons was 0.0125 (0.05/4). The levels of four sequences were greater than the control with a p value of 0.0078 (from sign and binomial test, statistically significant at the 0.0125 level). MUC1 was the most active, being more than 4-fold higher in AsPc1 cells (Fig. 5B). As a control, we also screened the upstream sequence (5 kb upstream of the ATG start codon) of 55 random genes and discovered eight stretched Canscript-like sequences (in five genes; supplemental Table S1) having linker lengths less than 40 bp. Thus, Canscript-like sequences did not appear with higher prevalence in genes up-regulated in pancreatic cancer, but some of these Canscript-like sequences were active.

YAP1 Is Required but Not Sufficient for Canscript Activity—To test YAP1 involvement in regulating Canscript activity, we first confirmed that YAP1 protein was associated with this TEAD1-MCAT complex in the MSLN promoter region by a ChIP assay (supplemental Fig. S7). Next we knocked down YAP1 by siRNAs in HeLa cells. As with the TEAD1 knockdown control, YAP1 knockdown dramatically reduced the endogenous MSLN expression level (Fig. 6A). Knockdown of YAP1 also reduced Can3 reporter activity (Fig. 6B), whereas the activity of control pGL3-SV40 was not affected (supplemental Fig. S2A). On the other hand, overexpression of ectopic YAP1 in RKO and 293 cells neither increased Can3-Luc activity (Fig. 6, C and D) nor boosted MSLN expression in MSLN-negative cell lines (data not shown). It was reported that unphosphorylated YAP1 is required but not sufficient for Canscript activity.

Structural Analysis of Canscript

FIGURE 4. The AC linker between the SP1-like and MCAT elements in Canscript. Experiments were done in AsPc1 cells. The RLA of pGL3-Basic was defined as 1. A, six possible point mutations of the two linker nucleotides were tested in pGL3-Can3. Only L-GC had less than half of the RLA of the wild type, showing an intolerance of G immediately after the SP1-like element. L, linker; L-AC, pGL3-Can3 (wild-type). B, position effects of the linker. C, length effects of the linker. Different spacer lengths (5, 10, 15, 20) of nucleotides taken from the hygromycin coding sequence were inserted within the AC linker in pGL3-67-94 (Fig. 18). The 20-nt spacer itself (pGL3-20; Fig. 18) did not have independent transcriptional activity.

D, position effect of the SP1-like and the MCAT elements. Reversing the order of the SP1-like and MCAT motifs around the AC linker (ACATTCCTGACTCTC-CACCC) in pGL3-67-94 reduced the RLA.
“active” YAP1 in the nucleus was more important than overall YAP1 protein levels (30). To address this, we overexpressed the constitutively active YAP1S127A mutant (30) in the same cell lines and again did not observe up-regulation of MSLN expression (data not shown). We also noticed similar ratios of phosphorylated to total YAP1 protein in the lysate of untransfected MSLN/H11001 and MSLN/H11002 cells (data not shown), suggesting that YAP1 may be similarly activated in these cells.

We observed that the cell numbers were reduced when HeLa cells were subjected to YAP1 knockdown (as compared with control siRNAs; data not shown). Cell cycle analysis failed to reveal a significant shift of the cell cycle profile in YAP1 siRNA knockdown cells (supplemental Fig. S2B). Instead, an apoptosis assay confirmed the antiapoptotic function of YAP1 (Fig. 6E) (21) because the apoptotic cell population greatly increased when YAP1 was knocked down. In contrast, cell number changes did not occur upon YAP1 overexpression in RKO or 293 cells.

Taz, a YAP1-homologous protein in mammals (31), is a candidate oncoprotein (32). We detected weak expression of Taz only in AsPc1 cells, but not in HeLa, RKO, and HEK293 cells (supplemental Fig. S1). Taz thus did not appear to be required for Canscript activity in HeLa cells. Its role in AsPc1 cells remains unexplored.

**The SP1-like Motif Is Not an SP1 Response Element**—The SP1-like element was thought to hold the key for the cancer specificity of Canscript (1). Attracted by its name, we tested and excluded SP1 as a candidate. There is one nucleotide difference between conventional SP1 binding site (CCCCGCCC) (22) and our SP1-like consensus sequence (CYCCACCC), but the nucleotide substitution pattern did not support the SP1 as a target; the A → G switch almost eliminated the Can3-luc activity (Fig. 3B). Sp1 was ubiquitously expressed in almost all of the cell lines (supplemental Fig. S1). MSLN protein expression was minimally affected by SP1 knockdown in HeLa cells (supplemental Fig. S3).

Interestingly, an SP1-binding sequence (TCTCCGCCC) in the SV40 promoter region of the plasmid (33) was known as pivotal for SV40 promoter activity. This made it impossible to study the SP1 effects on Canscript activity in the context of the
pGL3-SV40 backbone (1). For the same reason, the SV40 promoter-driven pRL-SV40 Renilla vector could not be used as an internal control for normalization of transfection during the SP1 knockdown assay.

**KLF6 May Be Required but Not Sufficient to Augment Canscript Activity**—Among the available DNA binding reports concerning the Krüppel-like family, the KLF6 consensus binding sequence (CCACCCA) optimally matched the SP1-like element. We found that the expression level of endogenous KLF6 protein was consistent with the MSLN expression pattern in different cell lines (supplemental Fig. S1). Knockdown of KLF6 (KLF6 si-2 and si-3) reduced the endogenous MSLN expression in HeLa cells (Fig. 7A). pGL3-Can3 reporter activity could also be reduced by knockdown of KLF6, and the combination of TEAD1 and KLF6 knockdown could not achieve further suppression (Fig. 7B). Overexpression of HA-tagged wild-type KLF6 in HEK293 cells (Fig. 7C) and in RKO cells, however, did not turn on endogenous MSLN expression (data not shown). Overexpression of HA-tagged KLF6 increased pGL3-Can3 activity at most slightly, as did the coexpression of both YAP1 and KLF6 (Fig. 7D). We tested whether KLF6 could directly bind to Canscript in vitro by EMSA. The biotin-labeled Canscript-containing oligonucleotide identified a DNA-protein complex (supplemental Fig. S4). A portion of this complex was supershifted by the positive control anti-TEAD1 antibody but not by the anti-KLF6 antibody (supplemental Fig. S4). ChIP assay also failed to detect a KLF6-Canscript complex in the MSLN promoter region when using the same anti-KLF6 antibody (supplemental Fig. S7).

**GLI1 Is Not a Strong Candidate**—We also investigated several oncogenic transcription factors. Among them, the GLI1-binding sequence (GACCACCCA) was similar to the SP1-like element, although the first G in the GLI1-binding sequence was not tolerated in our SP1-like consensus sequence (Fig. 3, B and C). Our immunoblot failed to detect the full-length GLI1 protein in AsPc1, HeLa, RKO, and HEK 293 cells (data not shown).

**RUNX Family Genes Were Not Sufficient for Canscript Activity**—Chimeric oncoprotein RUNX1 belongs to the RUNX protein family, whose binding consensus sequence (AAC-CACA) overlaps with the SP1-like element and the AC linker, although the unmatched A was not tolerated in the SP1-like consensus sequence (Fig. 3, B and C). The RUNX1 expression level was consistent with the MSLN expression pattern in different cell lines (supplemental Figs. S1 and S5A), but overexpression of each gene (RUNX1, -2, and -3) in MSLN-
negative cell lines (Fig. S5B) did not turn on endogenous MSLN expression (data not shown). Overexpression of RUNX proteins only had a marginal enhancement on Can3 activity (Fig. S5, B and C). These marginal effects were possibly contributed by the backbone of the pGL3-Basic vector (Fig. S5D) because three consensus RUNX-binding sequences (AACCACA) exist in the pGL3-Basic vector.

DISCUSSION

We found that the TSS in 211H mesothelioma cells was located 1380 bp downstream of the TSS reported in AsPc1 cells (1). The shorter transcript may be driven by a tissue-specific promoter. The upstream TSS in AsPc1 cells may be cancer-specific.

Canscript resembled more a promoter than an enhancer. The Canscript sequence, being located 11002-11046 bp 5′ of the upstream TSS in AsPc1 cells, is consistent with a pattern of promoter regulation. Reversing the orientation did not change the transcription activity, whereas relocation of Canscript resulted in loss of its activity (Fig. 2C). Human promoters, especially highly active promoters, are often bidirectional (34, 35), as is Canscript. This is consistent with a report in which a triplicated shorter 18-bp Canscript (from 11002-11045, CCACCCA-CACATTCCTGG) had a promoter-like activity 7 times higher than the pGL4-Basic vector (36). The 18-bp Canscript had much less activity than the 20-bp Canscript because the first three nucleotides, TCT, were missing (supplemental Fig. S6).

The cancer specificity (Fig. 2, A and B) and the promoter-like activity of Canscript imply that this 20-nt sequence may regulate MSLN transcription in many cancer cell lines.

Our nucleotide transition screen within the MCAT element confirmed that all eight nucleotides were functional, as in the consensus MCAT sequence (Fig. 3A). Yet, an unidentified transcription factor binding to the SP1-like element may be responsible for the cancer specificity of Canscript. For example, the flanking sequence of the MCAT motif played a critical role for selective expression of cardiac troponin T in striated muscles (19), and modification of its 5′-flanking MyoD-like binding site eliminated its tissue specificity.

To identify this unknown transcription factor, we examined KLF6, whose consensus DNA-binding sequence was very similar to the SP1-like element. The expression pattern of endogenous KLF6 was consistent with the MSLN expression pattern in various cell lines. Knocking down the expression of KLF6 in HeLa cells reduced the MSLN expression level (Fig. 7A). However, overexpression of KLF6 protein did not turn on MSLN expression in 293 and RKO cells, and we failed to detect the direct binding of KLF6 to the Canscript sequence by EMSA (supplemental Fig. S4). Based on the mixed evidence, KLF6 remains a candidate, but its possible role in MSLN regulation is not clearly established. GLI1 and RUNX family members, however, are not likely to be involved in MSLN transcription regulation.

Because the MCAT-binding TEAD1 cofactor could in theory contribute to cancer specificity if it were activated by a cancerspecific signal, such as from the Hippo-YAP1 pathway, we investigated the role of YAP1 in regulating Canscript activity. Knocking down YAP1 expression in HeLa cells dramatically
reduced endogenous MSLN expression and suppressed most of the Canscript reporter activity. Overexpression of wild-type YAP1 or its constitutively active mutant in RKO and HEK293 cells did not turn on MSLN expression, indicating that YAP1 may be necessary but not sufficient for MSLN overexpression in certain cancers. We did not find that the cancer specificity of Canscript was explained by YAP1.

The AC linker between SP1-like and MCAT elements is very flexible. Most nucleotide substitutions were tolerated (Fig. 4C) and MCAT was explained by YAP1. The linker could be deleted without affecting the pGL3-Canscript-like sequences from the 5’ region of FXYD3, MUC1, and TIMP1 genes displayed increased transcription activity. Canscript-like motifs may contribute to the overexpression of FXYD3, MUC1, and TIMP1 genes in pancreatic cancer and perhaps to the overexpressed “marker” genes in other cancer types as well.

Acknowledgment—We thank Dr. Duojia Pan for generously providing the pcDNA Zeo-YAP1 wild-type and mutant vectors.

REFERENCES

1. Hucl, T., Brody, J. R., Gallmeier, E., Iacobuzio-Donahue, C. A., Farrance, I. K., and Kern, S. E. (2007) Cancer Res. 67, 9055–9065
2. Thiagalingam, A., De Bustros, A., Borges, M., Jasti, R., Compton, D., Diamond, L., Mabry, M., Ball, D. W., Baylin, S. B., and Nelkin, B. D. (1996) Mol. Cell Biol. 16, 5335–5345
3. Watanabe, K., Saito, A., and Tamaoki, T. (1987) J. Biol. Chem. 262, 4812–4818
4. Hauck, W., Nédellec, P., Turbide, C., Stanners, C. P., Barnett, T. R., and Sartorelli, V. (1994) Eur. J. Biochem. 223, 529–541
5. Hassan, R., Bera, T., and Pasanth, I. (2004) Clin. Cancer Res. 10, 3937–3942
6. Chang, K., and Pasanth, I. (1996) Proc. Natl. Acad. Sci. U.S.A. 93, 136–140
7. Sathyanarayana, B. K., Hahn, Y., Patankar, M. S., Pasanth, I., and Lee, B. (2009) BMC Struct. Biol. 9, 1
8. Murumina, Z. E., Strong, T. V., and Shaw, D. R. (2004) BMC Cancer 4, 19
9. Frierson, H. F., Jr., Moskaluk, C. A., Powell, S. M., Zhang, H., Cerilli, L. A., Stoler, M. H., Cathro, H., and Hampton, G. M. (2003) Hum. Pathol. 34, 605–609
10. Argani, P., Iacobuzio-Donahue, C., Ryu, B., Rosty, C., Goggins, M., Wientz, R. E., Murugesan, S. R., Leach, S. D., Jaffe, E., Yeo, C. J., Cameron, J. L., Kern, S. E., and Hruban, R. H. (2001) Clin. Cancer Res. 7, 3862–3868
11. Ryu, B., Jones, J., Blades, N. J., Parmigiani, G., Hollingsworth, M. A., Hruban, R. H., and Kern, S. E. (2002) Cancer Res. 62, 819–826
12. Maitra, A., Adsay, N. V., Argani, P., Iacobuzio-Donahue, C., De Marzo, A., Cameron, J. L., Yeo, C. J., and Hruban, R. H. (2003) Mod. Pathol. 16, 902–912
13. Zhang, Y., Xiang, L., Hassan, R., Paik, C. H., Carrasquillo, J. A., Jing, B. S., Le, N., Ho, M., and Pasanth, I. (2006) Clin. Cancer Res. 12, 4695–4701
14. Hassan, R., Bullock, S., Premkumar, A., Kreitman, R. J., Kindler, H., Willingham, M. C., and Pasanth, I. (2007) Clin. Cancer Res. 13, 5144–5149
15. Kreitman, R. J., Hassan, R., Fitzgerald, D. J., and Pasanth, I. (2009) Clin. Cancer Res. 15, 5274–5279
16. Urvin, D., and Lake, R. A. (2000) Mol. Cell Biol. Res. Commun. 3, 26–32
17. Yoshida, T. (2008) Arterioscler. Thromb. Vasc. Biol. 28, 8–17
18. Zhao, P., Caretti, G., Mitchell, S., McKeehan, W. L., Boskey, A. L., Pachman, L. M., Sartorelli, V., and Hoffman, E. P. (2006) J. Biol. Chem. 281, 429–438
19. Larkin, S. B., Farrant, I. K., and Ordahl, C. P. (1996) Mol. Cell Biol. 16, 3742–3755
20. Kitagawa, M. (2007) Biochem. Biophys. Res. Commun. 361, 1022–1026
21. Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S. A., Gayyed, M. F., Anderson, R. A., Maitra, A., and Pan, D. (2007) Cell 130, 1120–1133
22. Berg, J. M. (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 11109–11110
23. Kaczynski, J., Cook, T., and Urrutia, R. (2003) Genome Biol. 4, 206
24. Cook, T., Gebelein, B., and Urrutia, R. (1999) Ann. N.Y. Acad. Sci. 880, 94–102
25. Kinzler, K. W., and Vogelstein, B. (1990) Mol. Cell Biol. 10, 634–642
26. Taniuchi, I., and Littman, D. R. (2004) Oncogene 23, 4341–4345
27. Vitolo, M. I., Anglin, I. E., Mahoney, W. M., Jr., Renoud, K. I., Gartenhaus, R. B., Bachman, K. E., and Passaniti, A. (2007) Cancer Biol. Ther. 6, 856–863
28. Mcgarvey, K. M., Fahrner, J. A., Greene, E., Martens, J., Jenuwein, T., and Baylin, S. B. (2006) Cancer Res. 66, 3541–3549
29. Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadrin, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S. B., Telikicherla, D., Selvan, L. D., Goel, R., Mathivanan, S., Marimuthu, A., Kashyap, M., Vizza, R. F., Mayer, R. J., Decaprio, J. A., Srivastava, S., Hanash, S. M., Hruban, R. H., and Pandey, A. (2009) PLoS Med. 6, e1000046
30. Zhao, B., Wei, X., Li, W., Udan, R. S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J., Li, L., Zheng, P., Ye, K., Chinnaiyan, A., Balak, L. Z., and Guan, K. L. (2007) Genes Dev. 21, 2747–2761
31. Kania, F., Marignani, P. A., Barbassova, D., Yagi, R., Hall, R. A., Donowitz, M., Hisamintno, A., Fujiwara, T., Ito, Y., Cantley, L. C., and Yaffe, M. B. (2000) EMBO J. 19, 6778–6791
32. Chan, S. W., Lim, C. J., Guo, K., Ng, C. P., Lee, I., Hunziker, W., Zeng, Q., and Wong, W. (2008) Cancer Res. 68, 2592–2598
33. Byrne, B. J., Davis, M. S., Yamaguchi, J., Bergsma, D. J., and Subramanian, K. N. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 721–725
34. Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008) Science 322, 1849–1851
35. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W. (2008) Science 322, 1855–1857
36. Huang, Y. H., Cozzitorto, J. A., Richards, N. G., Etoukhy, A. A., Yeo, C. I., Langer, R., Anderson, D. G., Brody, J. R., and Sawicki, J. A. (2010) Cancer Biol. Ther. 10, 878–884