

Alternative Cleavage and Polyadenylation during Colorectal Cancer Development

Adam R. Morris¹, Arnold Bos¹, Begoña Diosdado², Koos Rooijers¹, Ran Elkon¹, Anne S. Bolijn², Beatriz Carvalho², Gerrit A. Meijer², and Reuven Agami^{1,3}

Abstract

Purpose: Alternative cleavage and polyadenylation (APA) of mRNAs is a phenomenon that alters 3'-untranslated region length leading to altered posttranscriptional regulation of gene expression. Changing APA patterns have been shown to result in misregulation of genes involved in carcinogenesis; therefore, we hypothesized that altered APA contributes to progression of colorectal cancer, and that measurement of APA may lead to discovery of novel biomarkers.

Experimental Design: We used next-generation sequencing to directly measure global patterns of APA changes during colorectal carcinoma progression in 15 human patient samples. Results were validated in a larger cohort of 50 patients, including 5 normal/carcinoma pairs from individuals.

Results: We discovered numerous genes presenting progressive changes in APA. Genes undergoing untranslated region (3'UTR) shortening were enriched for functional groups such as cell-cycle and nucleic acid-binding and processing factors, and those undergoing 3'UTR lengthening or alternative 3'UTR usage were enriched for categories such as cell-cell adhesion and extracellular matrix. We found indications that APA changes result from differential processing of transcripts because of increased expression of cleavage and polyadenylation factors. Quantitative PCR analysis in a larger series of human patient samples, including matched pairs, confirmed APA changes in *DMKN*, *PDXK*, and *PPIE* genes.

Conclusions: Our results suggest that genes undergoing altered APA during human cancer progression may be useful novel biomarkers and potentially targeted for disease prevention and treatment. We propose that the strategy presented here may be broadly useful in discovery of novel biomarkers for other types of cancer and human disease. *Clin Cancer Res*; 18(19); 5256–66. ©2012 AACR.

Introduction

Colorectal cancer (CRC) is the second most common cause of cancer death in the Western world and consequently a major health care problem (1). Decreased CRC mortality can be achieved by identification of novel diagnostic, predictive, and prognostic biomarkers and the development of new therapies. To date, most of the studies concerning identification of biomarkers in CRC have mainly focused on known molecular pathways and aberrations (i.e., mutation, methylation changes, DNA copy number alterations, and microRNAs) contributing to CRC pathogenesis (2–6). Conducting unbiased genome wide studies based on novel

processes contributing to oncogene activation and tumor suppressor inactivation, rather than focusing on specific described pathways and molecular events, can lead to the identification of genes contributing to the initiation and progression of CRC, and consequently discovery of novel biomarkers (2).

Aberrant posttranscriptional regulation of gene expression has been shown to play an important role in the maintenance of cancer cells by modulating oncogene activation and tumor suppressor inactivation (4). Posttranscriptional regulation at the level of the mRNA is typically mediated by microRNAs (miRNA) and RNA-binding proteins that recognize and bind to elements within the 3' untranslated region (UTR) of an mRNA. The process of alternative cleavage and polyadenylation, or APA, can alter the expression of a gene by changing the length of the 3'UTR (UTR-APA) or producing an mRNA with a completely different 3'UTR [coding region APA (CR-APA)] (7). In UTR-APA 2 or more cleavage sites reside in one 3'UTR, whereas CR-APA results in a switch in a gene's 3' terminal exon. While the phenomenon of APA has been known for decades, only recently has the extent of APA been fully appreciated (7). The majority of human genes contain at least two polyA sites (PAS; ref. 8), and widespread APA has been shown in multiple organisms and systems (7). In

Authors' Affiliations: ¹Division of Gene Regulation, The Netherlands Cancer Institute; ²Department of Pathology VU Medical Centre, Amsterdam; and ³The Centre for Biomedical Genetics, The Netherlands

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

A.R. Morris and A. Bos contributed equally to the work.

Corresponding Author: Reuven Agami, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066 CX, Netherlands. Phone: 31-020-512-2079; Fax: 31-020-512-1999; E-mail: r.agami@nki.nl

doi: 10.1158/1078-0432.CCR-12-0543

©2012 American Association for Cancer Research.

Translational Relevance

Colorectal cancer is a major public health problem. Strategies to improve current diagnostics and survival were traditionally aimed at known molecular pathways and genomic aberrations. However, examining novel processes contributing to gene misregulation during carcinogenesis can uncover novel biomarkers and therapeutic targets. Interestingly, malignant transformation involves processes orchestrating gene regulation at the posttranscriptional level, often through the 3'-untranslated region (3'UTR) of mRNAs. Particularly, specific patterns of alternative cleavage and polyadenylation (APA), leading to alterations in 3'-UTRs length and content, have been shown to discriminate histologically indistinguishable tumors. Here, we used a genome wide 3' sequencing method to determine APA changes during progression from normal colon mucosa to colorectal carcinoma. Our results outline a strategy to efficiently identify genes undergoing altered APA in a disease state that can be used as biomarkers and can be potentially targeted for therapy.

general, proliferative cells, such as induced pluripotent stem cells and cancer cells, show a global shortening of 3'UTRs as compared with their less proliferative counterparts (9–14). As most regulation of gene expression mediated by 3'UTRs is repressive, it is generally assumed that a truncated 3'UTR will result in higher mRNA and/or protein levels. This assumption has proven valid, with measurements of mRNA and protein abundance during altered APA reflecting changes in 3'UTR length; for example, Mayer and Bartel showed that 3'UTR shortening of oncogene mRNAs in cancer cells leads to increased protein abundance (11).

APA patterns have the potential to be used as a diagnostic tool for distinguishing tumor types, as shown by Singh and colleagues (15) using mRNA expression arrays. In this study, the authors were able to define molecular signatures with which they could very accurately distinguish between leukemia/lymphoma cancer subtypes in a mouse model system. Importantly, the cancer subtypes were assigned as "histologically and immunophenotypically indistinguishable," thus showing the usefulness of this novel approach. The authors went on to show that molecular signatures based on APA could also be identified in previously published microarray data, indicating measurement of APA could be a powerful diagnostic tool in human patients.

Because of the great potential for gene regulation by APA, as well as the potential to use APA measurement as a diagnostic tool, we hypothesized that altered APA contributes to progression of CRC, and that measurement of APA may be used for discovery of novel biomarkers. Thus, we analyzed APA in different stages of CRC development, using a series of samples from normal colon mucosa, colorectal adenomas, and colorectal carcinomas. As the pathogenesis of CRC is considered a paradigm for multistep carcinogen-

esis, our results outline a strategy to efficiently identify genes with altered APA in various types of human cancer and other diseases that can be used as novel biomarkers and potentially targeted for disease prevention or treatment.

Materials and Methods

Collection of human tissue samples

Fresh tissue samples from 26 colorectal carcinomas, 26 colorectal adenomas, and 15 healthy control colon mucosa were collected at the department of pathology of the University medical center (VUmc), Amsterdam, the Netherlands. All samples were examined by one experienced pathologist (GM). Colorectal carcinomas and adenomas were used when more than 70% of epithelial component was present in the examined tissue section. All healthy control samples were collected along with tumor resections, were cancer free and the tissue samples showed normal histology. Clinical information of all individuals is listed in Table 1. The study was carried out in accordance with the code for proper secondary use of human tissue in the Netherlands by the Federation of Medical Scientific Societies. TRIzol (Invitrogen) was used to isolate the total RNA of the samples following the manufacturer's protocol. Total RNA quantity was determined with a Nanodrop ND-1000 spectrophotometer (Isogen) and quality was assessed in a 1% agarose gel, stained with ethidium bromide.

Microsatellite markers status

Microsatellite status was established using the MSI Analysis System (MSI Multiplex System Version 1.2, Promega) containing the BAT-25, BAT-26, NR-21, NR-24, MONO-27 nearly monomorphic markers according to the manufacturer's instructions. The obtained PCR products were separated on an ABI 3130 DNA sequencer (Applied Biosystems) and analyzed by GeneScan 3100 (Applied Biosystems). Tumors were considered as microsatellite instable (MSI) when 2 or more markers showed length changes (instability).

Construction of 3'-Seq libraries

Libraries were prepared essentially as described previously (16). Briefly, polyadenylated mRNA was fragmented, reverse transcription was conducted with an anchored oligo(d)T primer containing an adaptor sequence, and second strand synthesis was conducted. This double-stranded cDNA was then ligated to another adaptor, size-selection was conducted, and PCR was conducted using primers targeting the adaptor sequences to generate the final sequencing libraries. The 3'-Seq libraries were sequenced on the Illumina HiSeq system. For detailed methods, see Supplementary Methods.

Analysis of 3'-Seq data

3'-seq analysis was conducted as described previously (16). Briefly, reads were aligned to the human genome (hg18) using Bowtie (17) allowing up to 2 mismatches in the first 28 bp seed region. To also obtain alignment of reads covering cleavage sites, and thus containing an untemplated

Table 1. Clinical and histopathologic features of 62 individuals and their respective tumors and/or normal mucosa

Sample ID	Gender	Age	Tumor type	Localization gastrointestinal tract	Histologic-type colorectal adenomas	Tumor size (mm)	Astler-Coller staging	Grade dysplasia CRC	Degree differentiation carcinomas	3'-seq/3'-QPCR
1	F	58	Adenoma	Rectum	Tubulovillous	43	—	Low	—	3'-seq
2	F	65	Adenoma	Rectum	Tubular	10	—	Low	—	3'-seq
3	M	51	Adenoma	Sigmoid	Tubular	19	—	Low	—	3'-seq
4	F	58	Adenoma	Coecum	Tubulovillous	20	—	High	—	3'-seq
5	M	65	Adenoma	Rectosigmoid	Tubular	18	—	Low	—	3'-seq
6	M	56	Adenoma	UN*	Tubulovillous	20	—	High	—	3'-seq
7	F	48	Adenoma	Sigmoid	Tubulovillous	28	—	Low	—	3'-QPCR
8	F	80	Adenoma	Sigmoid	Tubulovillous	12	—	High	—	3'-QPCR
9	F	59	Adenoma	Rectosigmoid	Tubulovillous	40	—	High	—	3'-QPCR
10	M	57	Adenoma	Rectum	Tubulovillous	10	—	High	—	3'-QPCR
11	M	70	Adenoma	Colon ascendens	Tubulovillous	30	—	High	—	3'-QPCR
12	M	82	Adenoma	Sigmoid	Tubular	15	—	High	—	3'-QPCR
13	M	75	Adenoma	Colon ascendens	Tubulovillous	13	—	Low	—	3'-QPCR
14	M	63	Adenoma	Rectosigmoid	Tubulovillous	19	—	Low	—	3'-QPCR
15	F	82	Adenoma	Colon ascendens	Tubular	50	—	Low	—	3'-QPCR
16	M	53	Adenoma	Rectum	Tubulovillous	20	—	High	—	3'-QPCR
17	F	64	Adenoma	Colon descendens	Tubulovillous	18	—	High	—	3'-QPCR
18	M	63	Adenoma	Rectum	Tubulovillous	15	—	High	—	3'-QPCR
19	F	78	Adenoma	Colon ascendens	Tubulovillous	40	—	Low	—	3'-QPCR
20	F	71	Adenoma	Rectum	Tubulovillous	UN	—	High	—	3'-QPCR
21	M	55	Adenoma	UN	Tubulovillous	35	—	Low	—	3'-QPCR
22	F	78	Adenoma	Sigmoid	Tubulovillous	UN	—	Low	—	3'-QPCR
23	M	65	Adenoma	Rectum	Tubulovillous	25	—	High	—	3'-QPCR
24	M	68	Adenoma	Rectosigmoid	Tubulovillous	22	—	Low	—	3'-QPCR
25	M	54	Adenoma	Sigmoid	Tubular	19	—	Low	—	3'-QPCR
26	M	58	Adenoma	Sigmoid	Tubulovillous	30	—	High	—	3'-QPCR
27	M	73	Carcinoma	Colon ascendens	—	65	C2	—	Moderate	3'-QPCR
28	F	85	Carcinoma	Colon ascendens	—	130	C3	—	Poor	3'-QPCR
29	M	81	Carcinoma	Colon ascendens	—	50	C2	—	UN	3'-QPCR
30	F	83	Carcinoma	Colon ascendens	—	45	B2	—	Moderate	3'-QPCR
31	F	85	Carcinoma	Coecum	—	30	B1	—	Moderate	3'-QPCR
32	F	65	Carcinoma	Coecum	—	35	B2	—	Moderate	3'-QPCR
33	M	85	Carcinoma	Colon ascendens	—	35	B1	—	Moderate	3'-QPCR
34	F	62	Carcinoma	Colon ascendens	—	90	C2	—	Moderate	3'-seq
35	M	76	Carcinoma	Colon ascendens	—	25	D	—	Poor	3'-seq
36	F	65	Carcinoma	Sigmoid	—	30	D	—	Moderate	3'-seq
37 ^a	M	67	Carcinoma	Rectum	—	53	A2	—	Moderate	3'-seq
38	F	84	Carcinoma	Sigmoid	—	60	B2	—	Moderate	Both
39	F	86	Carcinoma	Colon ascendens	—	65	B1	—	Moderate	3'-QPCR
40	M	65	Carcinoma	Sigmoid	—	60	B2	—	Moderate	3'-QPCR
41	M	66	Carcinoma	Rectum	—	60	D	—	Moderate	3'-QPCR
42	F	71	Carcinoma	Coecum	—	30	A	—	Good	3'-QPCR
43	F	80	Carcinoma	Coecum	—	50	C2	—	Poor	3'-QPCR
44	M	62	Carcinoma	Sigmoid	—	85	B2	—	Moderate	3'-QPCR
45	F	74	Carcinoma	Rectosigmoid	—	20	A1	—	Moderate	3'-QPCR
46	M	53	Carcinoma	Colon transversum	—	90	B2	—	UN	3'-QPCR
47	F	57	Carcinoma	Sigmoid	—	35	C1	—	Good	3'-QPCR
48	F	75	Carcinoma	Rectum	—	18	A1	—	Moderate	3'-QPCR
49	F	50	Normal mucosa	Sigmoid	—	—	—	—	—	3'-seq
50 ^a	M	56	Normal mucosa	Colon ascendens	—	—	—	—	—	Both
51 ^a	M	71	Normal mucosa	Coecum	—	—	—	—	—	Both
52	F	72	Normal mucosa	Coecum	—	—	—	—	—	3'-seq
53	F	91	Normal mucosa	Sigmoid	—	—	—	—	—	3'-QPCR

(Continued on the following page)

Table 1. Clinical and histopathologic features of 62 individuals and their respective tumors and/or normal mucosa (Cont'd)

Sample ID	Gender	Age	Tumor type	Localization gastrointestinal tract	Histologic-type colorectal adenomas	Tumor size (mm)	Astler-Coller staging	Grade dysplasia CRC	Degree differentiation carcinomas	3'-seq/3'-QPCR
54	M	61	Normal mucosa	Sigmoid	—	—	—	—	—	3'-QPCR
55	M	77	Normal mucosa	Sigmoid	—	—	—	—	—	3'-QPCR
56	M	71	Normal mucosa	Colon ascendens/ transversum	—	—	—	—	—	3'-QPCR
57	F	86	Normal mucosa	Colon	—	—	—	—	—	3'-QPCR
58 ^a	M	73	Normal mucosa	Colon descendens	—	—	—	—	—	3'-QPCR
59	F	91	Normal mucosa	Coecum	—	—	—	—	—	3'-QPCR
60	M	89	Normal mucosa	Coecum	—	—	—	—	—	3'-QPCR
61 ^a	M	73	Normal mucosa	Coecum	—	—	—	—	—	3'-QPCR
62	F	50	Normal mucosa	Colon	—	—	—	—	—	3'-QPCR

Abbreviation: UN, information unavailable.

^aSample with matched normal/carcinoma.

poly(A) tail at their 3'-end, mismatches are allowed after the seed region. Only uniquely mapped reads were used in subsequent analysis. Cleavage sites were detected at a nucleotide resolution by identifying mapped reads containing an untemplated stretch of As. To link the identified cleavage sites to genes, each annotated transcript 3'UTR was extended by 1,000 bp to cover novel cleavage sites. Transcripts with more than one 3'UTR cleavage site were tested for significant relationship between PAS usage and condition using χ^2 test. To determine if shift was toward proximal or distal cleavage site for each transcript in each condition, we calculated a weighted mean of PAS index, weighing each PAS according to its reads coverage. For more details, see Supplementary Methods.

Gene expression analysis

3'Seq allows one not only to identify PASs and quantify their usage but also to obtain estimates for gene expression levels. As an estimate for expression levels we summed reads from all of a gene's isoforms as measured by reads mapping to a 3'UTR PAS. Quantile normalization was applied before comparing expression levels between samples.

3'PCR

Five hundred nanogram of the total RNA was fragmented with fragmentation buffer (Ambion) for 5 minutes at 70°C. The fragmented total RNA was precipitated and the first strand of cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions with the P7-t25-vn oligo-dT primer described in Supplementary Methods.

qRT-PCR was conducted with standard SYBR Green protocols using P7-primer and the gene-specific forward primers listed in table S1. Reactions were conducted in triplicate and the median Ct value was used for further analysis. Significance between sample types for each gene was determined with a Welch *t* test, using the Benjamini-Hochberg method for multiple testing correction.

PANTHER analysis

Analysis of enrichment of functional groups was conducted at PANTHERdb.org (18) using the "Compare gene lists" options in the Gene Expression Data Analysis Research Tool (19). The various gene lists described in the Results section consist only of genes with altered APA with a *P* value of more than 0.05, with shorter or longer describing negative or positive Δ PAS_J values, respectively. A list of all genes for which we obtained data in each comparison was used as the reference. Only functional groups containing at least 4 genes with altered APA and *P* < 0.01 are presented.

Results

Global analysis of APA patterns in colorectal cancer by 3'Seq

Our first goal was to directly measure global patterns of APA during adenoma to carcinoma progression. To achieve this, we used 3'Seq, a method we recently developed to assay APA (16). The 3'Seq method is presented in Fig. 1A and outlined in the Materials and Methods section. We measured APA in human tissue samples from 5 colorectal carcinomas, 6 colorectal adenomas, and 4 normal colon mucosae. Sequencing statistics are presented in Supplementary Fig. S1. We obtained information on 7,035 cleavage sites, representing 6,306 genes, in the combined samples, with 2 or more cleavage sites detected in 656 genes. 3'Seq data can be directly visualized in the UCSC genome browser, as shown by the modified genome browser view presented in Fig. 1B.

Next, we determined changes in patterns of APA by analysis of the distribution of obtained reads over different cleavage sites for a gene (for details see Materials and Methods). We observed significant changes in APA (*P* < 0.05, χ^2 test) in 286 genes from normal mucosa to adenoma, 322 from adenoma to carcinoma, and 304 from normal mucosa to carcinoma. APA is known to mainly have 2

patterns: either the 2 cleavage sites reside in one 3'UTR, which has been termed UTR-APA, or there is a switch in a gene's 3' terminal exon, known as coding region APA (CR-APA; ref. 7). The ratio of observed UTR-APA versus CR-APA changes in each comparison was about 3:1 (216:70, 240:82, and 230:74; see Supplementary Fig. S2A). An example of each type of APA is presented in Fig. 1B. The mRNA of pyridoxal kinase (PDXK) displays UTR-APA and undergoes a shift from the distal cleavage site (relative to the coding sequence) to a more proximal cleavage site within a single 3'UTR in a single 3' terminal exon. In contrast, peptidyl-prolyl isomerase E (PPIE) is an example of a change in CR-APA, with a shift to an alternate 3' terminal exon changing both the coding sequence and 3'UTR. PPIE and PDXK are also examples of different genes that begin to undergo changes in APA from normal to adenoma during CRC progression (PPIE) and those that change later on from adenoma to carcinoma (PDXK).

Global mRNA shortening is observed during colorectal adenoma to carcinoma progression

Interestingly, global analysis of the genes that showed UTR-APA changes showed mRNA 3'-UTR shortening, or a shift toward the proximal PAS, from adenoma to carcinoma or from normal colon to carcinoma (158/240; $P = 1.1E-6$ and 139/230; $P = 1.9E-3$, respectively; 2-tailed binomial test; see Supplementary Fig. S2B). In contrast, no significant global shift toward shortening or lengthening was observed in the genes showing altered APA from normal mucosa to adenoma (105/216; n.s.). The proliferative nature of normal colon tissue in addition to the heterogeneity of human patient samples may explain this lack of difference. Thus, global induction of mRNA 3'-UTR shortening by UTR-APA is observed during colorectal adenoma to carcinoma progression. (Shortening vs. lengthening was determined on the basis of PAS_J scores, a metric which determines the center of mass of all reads mapping to cleavage sites within a

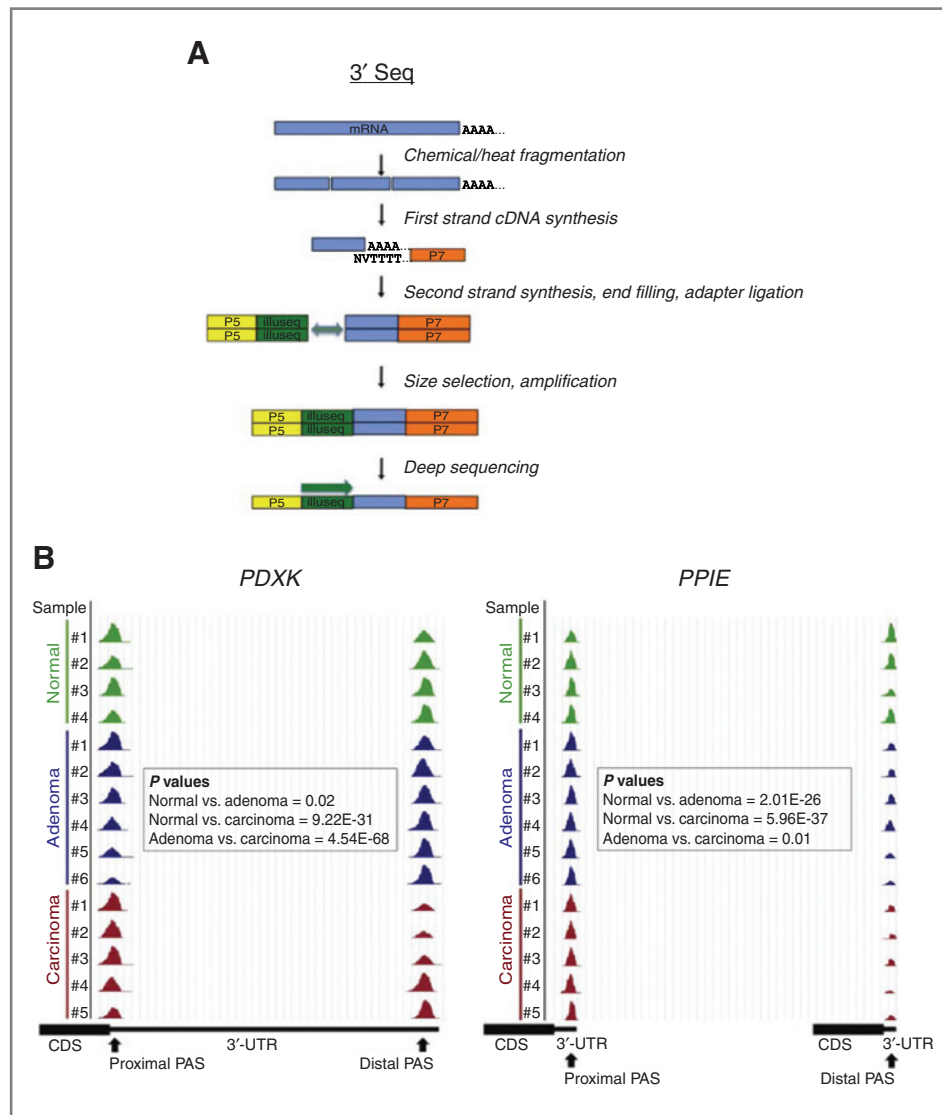


Figure 1. Quantifying 3' ends of mRNAs in colorectal cancer samples. A, overview of the 3' Seq method used in this study. For details, see Materials and Methods. B, modified UCSC genome browser view of the 3' ends of 2 genes undergoing altered alternative cleavage and polyadenylation (APA) during colorectal cancer progression. Peak height reflects the relative usage of a gene's proximal and distal polyA site (PAS; relative to the coding sequence) in each sample based on the reads in the 3' Seq data. Green, normal colon mucosa; blue, adenoma; red, carcinoma. The inset shows the significance of APA changes based on the χ^2 test. PDXK is an example of UTR-APA (each PAS in the same exon), whereas peptidylprolyl isomerase E (PPIE) is an example of coding region APA (PASs in different exons).

Downloaded from <http://aacrjournals.org/clinccancerres/article-pdf/18/19/5260/2006257/5260.pdf> by guest on 22 May 2024

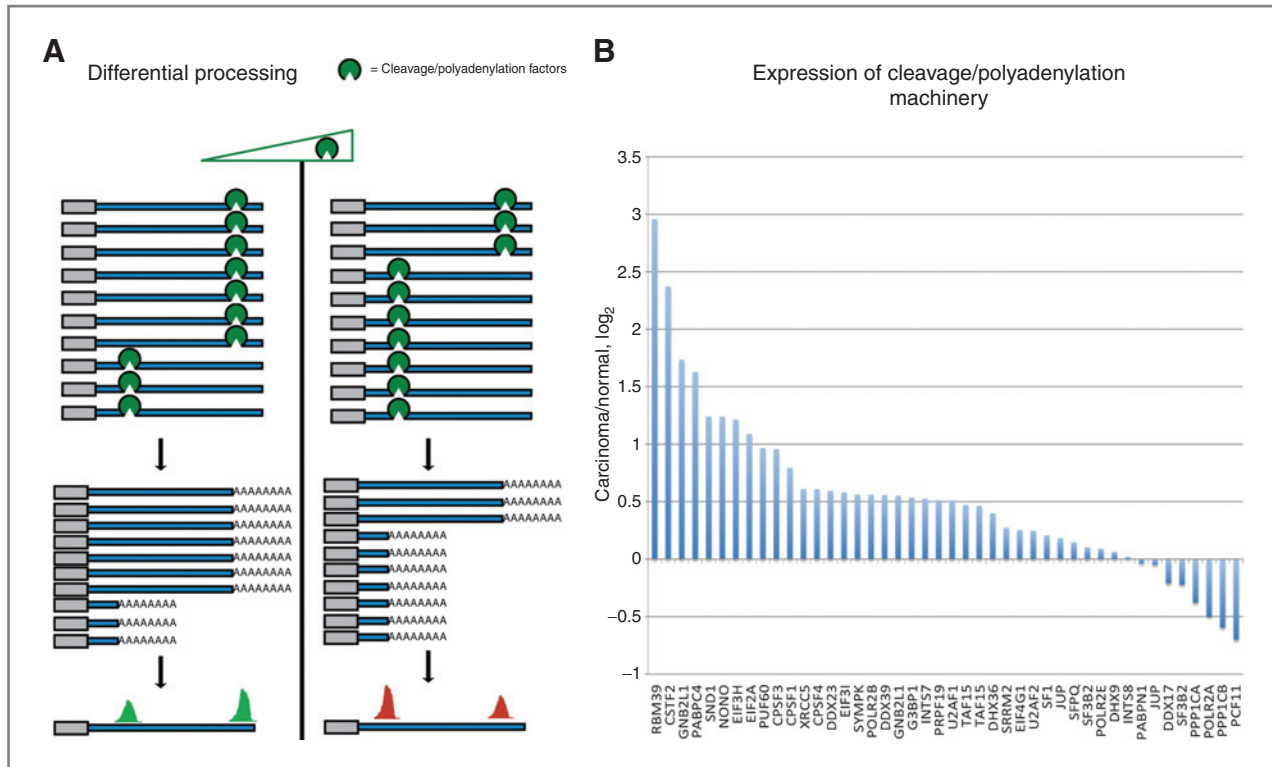


Figure 2. Expression of cleavage and polyadenylation factors during colorectal cancer progression. A, differences in abundance or content of the cleavage and polyadenylation complex results in differential processing. Note that the proximal cleavage sites, which generally are associated with weaker polyadenylation signals, are increasingly used upon increasing concentration of 3' end processing factors. B, increased expression of mRNA 3' end processing factors indicates differential processing. Change in mRNA abundance is shown on the y-axis, and genes comprising the cleavage and polyadenylation complex (10, 20) for which we obtained data are shown on the x-axis.

3'UTR, such that a greater PAS_J value indicates more usage of a distal PAS. See Supplementary Table S2 for global PAS_J values and Supplementary Methods for details on the PAS_J metric.)

APA changes may be driven by increased expression of cleavage and polyadenylation factors

It has been proposed that changes in PAS usage can result from differential cleavage and polyadenylation efficiencies between 2 conditions (Fig. 2A), which can theoretically be caused by an increase in abundance of cleavage and polyadenylation factors, structural changes of the mRNA, or an altered rate of transcription (7). Inspecting our data, we found increased expression of cleavage and polyadenylation factors during CRC progression. We used 3'Seq data to quantify the expression of the genes encoding the cleavage and polyadenylation machinery during CRC progression by summing reads from all of a gene's isoforms. Using the cleavage/polyadenylation gene list generated by Ji and colleagues (10, 20), we found that of the 44 genes for which we obtained expression data, 36 were more highly expressed (mean log₂ expression = 0.53; range = -0.70 to 2.96) in carcinoma than normal mucosa, demonstrating a significant increase ($P = 1.3E-5$; binomial test, 1 tailed) in expression of this set of genes during CRC progression (Fig. 2B). As increased abundance of 3' end processing factors is pre-

dicted to induce cleavage at inefficient proximal PASs, leading to 3'UTR shortening, this finding indicates that differential processing is at least partially responsible for the perceived APA changes during CRC progression.

Functional enrichment of genes displaying altered APA

Next, we examined whether there are enriched functional groups among the genes displaying altered APA ($P < 0.05$ in the 3'Seq analysis). We used the online tool PANTHER (18, 19) to search for overrepresented functional groups in genes displaying 3'UTR shortening or lengthening by UTR-APA in the transition from normal tissue to adenoma, adenoma to carcinoma, and normal tissue to carcinoma, or in genes displaying CR-APA changes in either direction in the above comparisons (as CR-APA results in a different 3' UTR rather than just a shorter or longer one, we ignored the direction of the shift). The results of this comparison are shown in Table 2. Genes displaying CR-APA changes are mostly enriched for categories involving cell-cell adhesion, indicating a role for misregulation of genes that mediate cell-cell interactions in CRC development. In general, genes that display 3'UTR shortening by UTR-APA are enriched for the categories of cell cycle, DNA and RNA binding, and RNA splicing, whereas those that display 3'UTR lengthening are enriched for categories such as cell-cell adhesion and extracellular matrix.

Table 2. Enriched functional groups in UTR and CR APA using PANTHER analysis

	<i>P</i> value	Genes	Category
UTR APA			
Adenoma versus normal			
<i>Shorter</i>			
Spermatogenesis	4.92E-04	8	BP
RNA splicing factor activity, transesterification mechanism	1.22E-03	8	MF
RNA binding	4.93E-03	10	MF
Ribonucleoprotein complex	6.96E-04	6	CC
DNA binding protein	4.14E-04	11	PPC
Replication origin binding protein	9.77E-04	4	PPC
Nucleic acid binding	3.79E-03	19	PPC
<i>Longer</i>			
Integrin signaling pathway	6.15E-03	6	P
Cell-cell adhesion	3.01E-03	11	BP
Receptor activity	4.18E-03	18	MF
Carcinoma versus adenoma			
<i>Longer</i>			
Spermatogenesis	3.87E-03	6	BP
Carcinoma versus normal			
<i>Shorter</i>			
Cell cycle	6.89E-03	26	BP
RNA splicing factor activity, transesterification mechanism	2.12E-03	9	MF
Protein phosphatase	7.48E-03	5	PPC
<i>Longer</i>			
Cell-cell adhesion	2.40E-03	10	BP
Complement activation	6.70E-03	4	BP
Extracellular matrix structural constituent	2.16E-03	5	MF
Extracellular matrix	2.50E-03	8	CC
Extracellular region	2.61E-03	8	CC
Membrane-bound signaling molecule	4.17E-03	4	PPC
CR-APA			
Adenoma versus normal			
Ectoderm development	3.20E-03	12	BP
Methyltransferase activity	5.23E-03	4	MF
Methyltransferase	4.92E-03	4	PPC
Carcinoma versus adenoma			
Cell adhesion	3.06E-03	13	BP
Cell-cell adhesion	3.34E-03	9	BP
Carcinoma versus normal			
Cell adhesion	4.06E-03	12	BP
Cellular process	7.26E-03	35	BP
Neurologic system process	8.87E-03	14	BP
Binding	8.40E-03	38	MF

NOTE: Shorter or longer refers to the first tissue type in relation to the second, and genes refers to the number of genes with altered APA in that group. Only functional groups with at least 4 genes and $P < 0.01$ are shown.

Abbreviations: P, pathway; BP, biological process; MF, molecular function; CC, gene ontology cellular component; PPC, PANTHER protein class.

Under the assumption that UTR length is inversely correlated with expression, our functional enrichment analysis seems to indicate that changes in APA patterns induce cell cycle and nucleic acid binding and processing genes during cancer development, while repressing genes mediating the

interaction of cells with their external environment. We went on to compare the extent of APA changes with changes in mRNA abundance, which revealed no global correlation (Supplementary Fig. S3). Thus, it seems that both UTR-APA and CR-APA are altered during CRC

development but that this does not lead to global directional change in mRNA abundance. This finding does not, however, address regulation of translation, which is known to be a mechanism of both miRNA (21) and RBP (22) mediated repression.

APA patterns as novel biomarkers

To test the potential of using APA patterns as novel biomarkers, we used 3'PCR, a QPCR-based method to measure APA (Fig. 3A and Materials and Methods). This method allows direct and accurate measurement of the

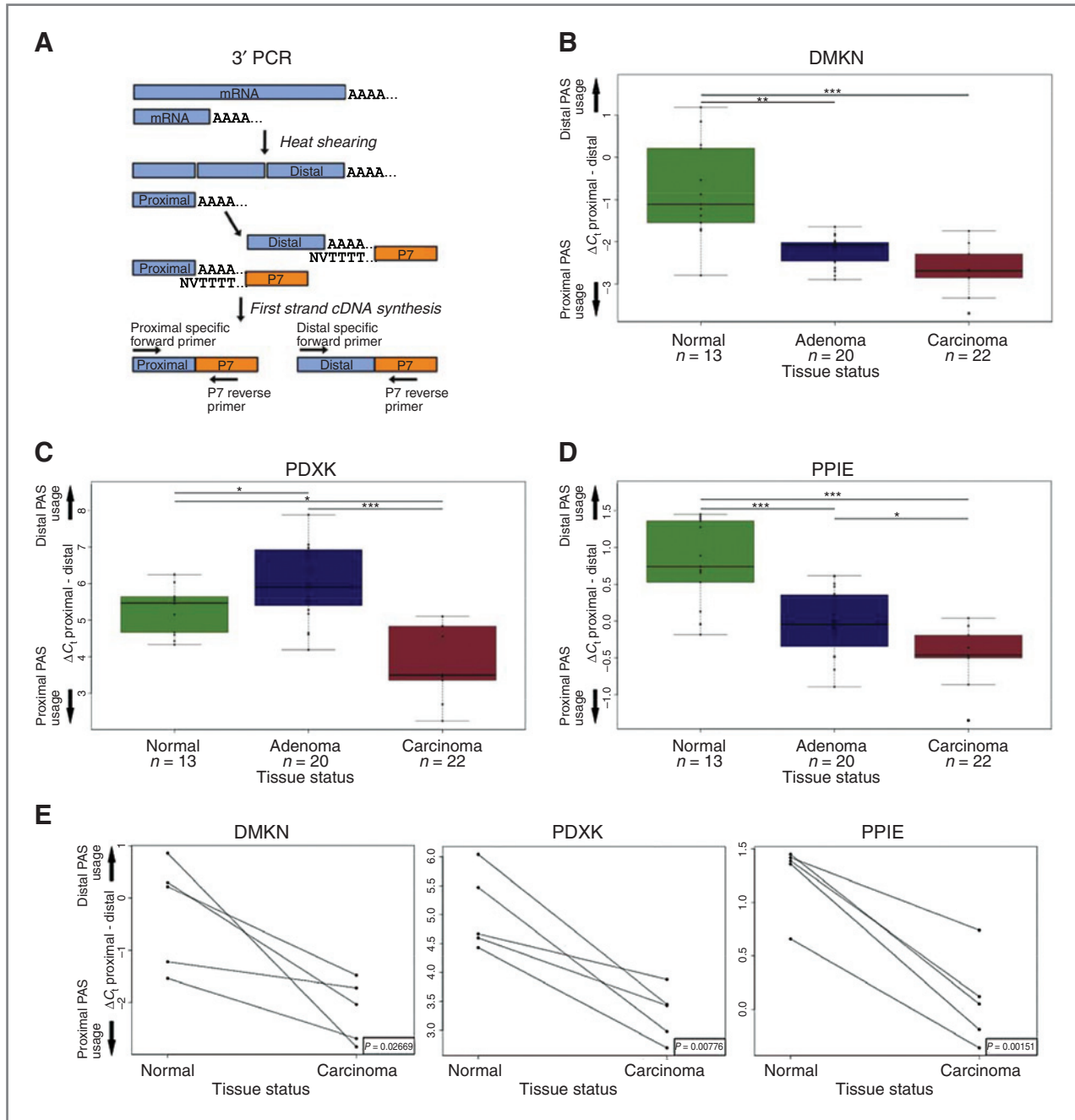


Figure 3. Alternative cleavage and polyadenylation as a diagnostic tool. A, overview of the 3'PCR method used in this study. Proximal and distal isoforms are independently measured. For details, see Materials and Methods. B–D, results of 3'PCR from multiple patient samples including 13 normal mucosa, 20 adenomas, and 22 carcinomas. (N) and green, normal mucosa; (A) and blue, adenoma; (C) and red, carcinoma. Each individual patient sample is represented by a black dot. The y-axis represents ΔC_T (proximal PAS – distal PAS), thus a smaller value indicates increased usage of the proximal polyA sites (PAS) in relation to the distal. *, $P < 0.05$; **, $P < 10^{-2}$; ***, $P < 10^{-3}$ (Welch t test, Benjamini–Hochberg corrected for multiple testing). E, results of the 3'–PCR method for matched sets of normal carcinoma. Each individual sample is represented by a black dot, and each line connects the normal tissue sample to the carcinoma sample from the same patient. P values (paired t test) are displayed in lower right corner.

relative abundance of each mRNA form, rather than relying on a subtractive analysis that would be required using typical RT-QPCR techniques. Using this technique we assayed APA patterns of genes that showed significant and substantial APA changes in the 3'Seq experiment in 13 normal colon mucosa, 20 colorectal adenomas, and 22 colorectal carcinomas.

Figure 3 shows that APA patterns show promise as novel disease biomarkers. The genes *dermokine* (*DMKN*) and *PPIE* show significant APA changes between normal colon mucosa and adenoma ($P < 0.0014$ and $P < 0.00062$, respectively; Welch *t* test with Benjamini–Hochberg correction) or between normal mucosa and carcinoma ($P < 0.00040$ and $P < 1.5E-5$, respectively) with a shift toward increased usage of the proximal (relative to the coding region) PAS in relation to the distal. In addition, the APA shifts of *PPIE* and *PDXK* are significant from colorectal adenoma to carcinoma ($P < 0.013$ and $P < 0.00016$, respectively). When analyzing APA patterns within individual patients, using matched normal mucosa and colorectal carcinoma, we found that the reported APA shifts also occurred consistently and significantly within individual patients (Fig. 3E). These results confirm the 3'Seq findings in both a larger cohort of human patients and in individual patients, and indicate that our strategy could be valuable in discovering novel biomarkers.

Adenoma samples could be further subdivided by the existence of chromosomal aberrations (gains of 8q, 13q, and 20q), and the carcinomas subdivided as microsatellite stable or unstable. Supplementary Fig. S4 shows that analysis of each subgroup individually yields similar results as the combined samples, even though the combined data do not represent the percentages of each type of adenoma and carcinoma in the general patient population (5). Although the results when comparing tumor subgroups are not statistically significant, visual inspection of the data suggests the potential to distinguish tumor subtypes with this kind of analysis. Notably, comparison of abundance of either the proximal or distal APA isoforms to that of a control gene was less effective in distinguishing tissue types than direct comparison of the isoforms (Supplementary Fig. S5). Thus, it appears to be the balance between isoforms, rather than the abundance of one isoform or the other, which is associated with CRC progression, demonstrating the utility of measuring APA patterns rather than simple mRNA abundance (as is *modus operandi* in analysis of gene expression data and signature/biomarker development). Overall, these results indicate that measurement of APA patterns is promising in discovery and implementation of novel biomarkers for cancer and other diseases.

Discussion

In this study, we conducted global direct measurement of 3' ends of mRNAs during human CRC progression and identified changing patterns of APA in genes involved in cell cycle, RNA and DNA binding, and interaction with the extracellular environment. As cancer progression involves alterations in both cellular proliferation and phenotype, it is

likely that many of the APA changes described in this study contribute to these aspects of cancer progression. QPCR-based measurements of APA in a selected group of genes in an independent series of CRC patient samples confirmed the findings of the global analysis, including in paired samples, and reinforce the validity of using this strategy to discover novel biomarkers.

APA has typically been measured either by northern blot or by microarrays using a "subtractive" analysis approach, but the potential to study APA has recently improved substantially with the availability of next-generation sequencing (NGS) technology. A number of NGS-based methods to directly measure APA patterns have been developed, including the one used in the present study (13, 23, 24). This has advantages over traditional microarrays; one, common to most NGS experiments, is the ability to assay sequences that were not identified *a priori*, that is, spotted on a microarray. In terms of APA, NGS technology has an additional advantage over microarrays in that it allows direct measurement of usage of each PAS rather than relying on comparing probes within different regions of a 3'UTR. Measurement of APA by microarray is also less comprehensive, as one is limited to the genes which contain probes in appropriate sites in the 3'UTR.

There are a number of aspects that set this study of APA apart from others. To our knowledge, this is the first study to directly measure global APA during cancer progression in a group of human patient samples and validate these APA patterns in an independent series of human patients. Other studies that globally measured APA in cancer were conducted with cancer cell lines rather than human patient samples (24), and previous studies that assayed patient samples were based on indirect measurements using microarrays and were not validated in another independent series (15). By combining a large number of human patient samples, NGS-based technology, and a novel QPCR-based technique, we were able to identify and confirm substantial and significant APA changes during human colorectal adenoma to carcinoma progression.

Our approach measures and distinguishes UTR-APA, where the alternative cleavage sites reside in one 3'UTR exon, from CR-APA, where alternative cleavage sites reside in different 3'UTR exons. The potential effect of altered UTR-APA on gene expression is easier to interpret than that of CR-APA, as the latter frequently involves changes in protein sequence in addition to an entire switch of 3'UTRs. As the majority of UTR-APA changes are likely to induce gene expression upon mRNA shortening, its effect has been investigated in many studies (7). However, our data indicate that CR-APA changes are also quite common and also deserve research attention. Our results provide a resource for additional experimental and computational studies to assess the usefulness of UTR-APA and CR-APA patterns as potential biomarkers for cancer and as targets of therapeutics.

We found evidence that APA changes were the result of increased expression of the cleavage and polyadenylation machinery. Another potential mechanism leading to perceived APA changes is differential degradation of isoforms.

For example, if a gene has 2 APA isoforms, one containing a miRNA target site and one lacking it, an increase in abundance or activity of the miRNA would lead to enhanced degradation of the miRNA site-containing transcripts, which would be perceived as an APA shift toward the isoform lacking the APA site. Indeed, we saw some indication of differential degradation during CRC progression (data not shown), and this mechanism will be further explored in future studies. Importantly, knowledge of the relative contribution of differential processing and degradation to perceived APA is not required for the use of APA patterns as biomarkers, as our data show that comparison of the isoforms is sufficient.

In addition to providing valuable information on potential biomarkers, the changes in APA patterns measured by this technique undoubtedly reflect physiological changes that contribute to tumorigenesis. Further investigation of these changes at the global and individual gene levels will likely reveal new genes and pathways that can be targeted in cancer treatment and prevention. For example, one of the genes revealed by this study, PPIE, is known to regulate the function of MLL, a transcriptional coactivator with roles in hematopoiesis and leukemia (25). Through its proline isomerase activity, it can change the conformation of MLL from a transcriptional activator to a repressor, affecting a whole repertoire of developmentally regulatory genes.

Another of the identified genes is DMKN. This gene is known to have several isoforms, derived from multiple transcription start sites, alternative exons, and PAS sites (26). The β/χ -isoforms have been shown to be secreted markers for early stage colorectal carcinoma (27), with the χ -isoform making use of an earlier PAS site. In our data, we observed an APA shift toward a proximal PAS site in adenoma compared with normal tissue. Therefore, although previous studies indicate alternative splicing/polyadenylation is a marker for early stage colorectal carcinoma, our data indicate this event may occur even earlier in cancer progression.

Finally, PDXK codes for a gene critical in the metabolism of vitamin B6, phosphorylating the immature vitamin B6 to make its bioactive form. Although this bioactive form of vitamin B6 has been associated with prevention of DNA lesions (28) and a reduced risk of CRC (29), the *PDXK* gene itself is still largely unknown in the field of cancer. Notably, another 3'Seq dataset from our lab from breast cancer patients also showed a highly significant APA shift for PDXK (data not shown), suggesting a more general role for this gene.

Future studies will elucidate the role of APA regulation in driving tumorigenesis through the discussed genes, as well as others. Moreover, Singh and colleagues showed that

ovarian cancer cells that are resistant or sensitive to cisplatin show distinct APA patterns (15), providing further support for the usefulness of this type of approach for improving cancer therapy.

The strategy presented here can theoretically be used to identify APA-based biomarkers for many types of cancer and other diseases. The process of first identifying potential candidates through a global analysis and then validating some of the most altered genes by QPCR in a larger cohort can be used to quickly and efficiently identify novel biomarkers. Exploiting QPCR technology provides the added advantage of requiring very little starting material, which is particularly relevant with human patient samples where the amount of material available is often limited. We present here a proof of principle of this process and provide 3 examples.

In conclusion, we directly measured global patterns of APA during human CRC progression and went on to show in an independent series of human patient samples that these APA patterns can be valuable as novel biomarkers. We believe that our strategy can be used in many human cancers and other diseases to identify genes that are novel biomarkers and can potentially be targeted in disease prevention and treatment.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: R. Agami

Development of methodology: A.R. Morris, A. Bos

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.R. Morris, A. Bos, B. Diosdado, A.S. Bolijn, B. Carvalho, G.A. Meijer

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.R. Morris, A. Bos, K. Rooijers, R. Elkon

Writing, review, and/or revision of the manuscript: A.R. Morris, A. Bos, B. Diosdado, K. Rooijers, A.S. Bolijn, B. Carvalho, G.A. Meijer, R. Agami

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): B. Diosdado, A.S. Bolijn

Study supervision: R. Agami

Acknowledgments

The authors thank all the members of the Agami laboratory for technical help and discussions. The authors thank Arno Velds and Ron Kerkhoven at the NKI Central genomics facility for deep sequencing our samples.

Grant Support

This work was supported by ERC (European Research Council), VICI-NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), and CBG (Centre for Biomedical Genetics) to R. Agami.

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.

Received February 15, 2012; revised July 31, 2012; accepted August 2, 2012; published OnlineFirst August 8, 2012.

References

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 2009;9:489–99.
- van Engeland M, Derks S, Smits KM, Meijer GA, Herman JG. Colorectal cancer epigenetics: complex simplicity. *J Clin Oncol* 2011;29:1382–91.
- Lee SK, Calin GA. Non-coding RNAs and cancer: new paradigms in oncology. *Discov Med* 2011;11:245–54.

5. Hermesen M, Postma C, Baak J, Weiss M, Rapallo A, Sciotto A, et al., Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology* 2002;123:1109–19.
6. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759–67.
7. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* 2011;43:853–66.
8. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 2005;33:201–12.
9. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 2009;106:7028–33.
10. Ji Z, Tian B. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* 2009;4:e8419.
11. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009;138:673–84.
12. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 2008;320:1643–7.
13. Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 2011;17:761–72.
14. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al., Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008;456:470–6.
15. Singh P, Alley TL, Wright SM, Kamdar S, Schott W, Wilpan RY, et al., Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res* 2009;69:9422–30.
16. Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, et al., The poly(a)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 2012;149:538–53.
17. Langmead B, Schatz MC, Lin J, Pop M, Salzberg SL. Searching for SNPs with cloud computing. *Genome Biol* 2009;10:R134.
18. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, et al., PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 2003;31:334–41.
19. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, et al., Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* 2006;34(Web Server issue):W645–50.
20. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, et al., Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 2009;33:365–76.
21. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 2010;79:351–79.
22. Abaza I, Gebauer F. Trading translation with RNA-binding proteins. *RNA* 2008;14:404–9.
23. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, et al., Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 2010;143:1018–29.
24. Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, et al., Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* 2011;21:741–7.
25. Park S, Osmers U, Raman G, Schwantes RH, Diaz MO, Bushweller JH. The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. *Biochemistry* 2010;49:6576–86.
26. Naso MF, Liang B, Huang CC, Song XY, Shahied-Arruda L, Belkowski SM, et al. Dermokine: an extensively differentially spliced gene expressed in epithelial cells. *J Invest Dermatol* 2007;127:1622–31.
27. Tagi T, Matsui T, Kikuchi S, Hoshi S, Ochiai T, Kokuba Y, et al., Dermokine as a novel biomarker for early-stage colorectal cancer. *J Gastroenterol* 2010;45:1201–11.
28. Kanellis P, Gagliardi M, Banath JP, Szilard RK, Nakada S, Galicia S, et al., A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. *PLoS Genet* 2007;3:e134.
29. Larsson SC, Orsini N, Wolk A. Vitamin B6 and risk of colorectal cancer: a meta-analysis of prospective studies. *JAMA* 2010;303:1077–83.