Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas

Genes & Cancer
2(8) 829–840
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1947601911431081
<http://ganc.sagepub.com>

Lloyd D. Graham^{1,*}, Susanne K. Pedersen^{2,*}, Glenn S. Brown¹, Thu Ho¹, Zena Kassir^{1,3}, Audrey T. Moynihan¹, Emma K. Vizgoft², Robert Dunne⁴, Letitia Pimlott⁵, Graeme P. Young⁵, Lawrence C. LaPointe², and Peter L. Molloy¹

Submitted 12-Aug-2011; accepted 02-Nov-2011

Abstract

An uncharacterized gene locus (Chr16:hCG_1815491), now named *colorectal neoplasia differentially expressed* (gene symbol *CRNDE*), is activated early in colorectal neoplasia. The locus is unrelated to any known protein-coding gene. Microarray analysis of 454 tissue specimens (discovery) and 68 previously untested specimens (validation) showed elevated expression of *CRNDE* in >90% of colorectal adenomas and adenocarcinomas. These findings were confirmed and extended by exon microarray studies and RT-PCR assays. *CRNDE* transcription start sites were identified in CaCo2 and HCT116 cells by 5'-RACE. The major transcript isoforms in colorectal cancer (CRC) cell lines and colorectal tissue are *CRNDE*-a, -b, -d, -e, -f, -h, and -j. Except for *CRNDE*-d, the known *CRNDE* splice variants are upregulated in neoplastic colorectal tissue; expression levels for *CRNDE*-h alone demonstrate a sensitivity of 95% and specificity of 96% for adenoma versus normal tissue. A quantitative RT-PCR assay measuring *CRNDE*-h RNA levels in plasma was (with a threshold of $2^{-\Delta C_t} = 2.8$) positive for 13 of 15 CRC patients (87%) but only 1 of 15 healthy individuals (7%). We conclude that individual *CRNDE* transcripts show promise as tissue and plasma biomarkers, potentially exhibiting high sensitivity and specificity for colorectal adenomas and cancers.

Keywords

RNA biomarker, colorectal cancer, colorectal adenoma, colorectal neoplasia, *CRNDE*

Introduction

Colorectal cancer (CRC) is among the most significant causes of cancer-related death in the developed world.¹ Early detection significantly improves the outcome for CRC patients,²⁻⁴ so public awareness campaigns and government screening programs are being introduced around the world to decrease the mortality and morbidity of this disease.⁵ Despite these measures and the clear opportunity for early intervention, many people still die from CRC each year. In 2008, the disease caused an estimated 609,051 deaths worldwide.⁶

Criteria have been established for cancer biomarker discovery and validation.⁷ Several studies have identified candidate gene biomarkers for colorectal neoplasia,^{8,9} but most of these have failed to meet expectations in subsequent validation studies. Elsewhere, we have described the discovery and validation of numerous RNA biomarkers for colorectal neoplasia, including gene transcripts that display high sensitivity and specificity for both adenocarcinomas and precancerous adenomas (LaPointe *et al.*, submitted manuscript).

One source of such biomarkers, locus hCG_1815491 on chromosome 16, is of particular interest, as it has negligible

expression in normal colorectal tissue but substantial expression in both colorectal adenomas and adenocarcinomas, suggesting that this gene is active early in neoplastic progression (LaPointe *et al.*, submitted manuscript). In recognition of these observations, the HUGO Gene Nomenclature Committee accepted our proposal to name the locus *colorectal neoplasia differentially expressed* (*CRNDE*). The gene has since been identified in an independent database-mining exercise as the most upregulated gene

Supplementary material for this article is available on the *Genes & Cancer* website at <http://ganc.sagepub.com/supplemental>.

¹CSIRO Food and Nutritional Sciences, Sydney, NSW, Australia

²Clinical Genomics Pty. Ltd., Sydney, NSW, Australia

³Garvan Institute of Medical Research, Sydney, NSW, Australia

⁴CSIRO Mathematical and Information Sciences, Sydney, NSW, Australia

⁵Flinders University, Adelaide, SA, Australia

*These authors contributed equally to this work.

Corresponding Author:

Lloyd D. Graham, Preventative Health National Research Flagship, CSIRO Food & Nutritional Sciences, P.O. Box 52, North Ryde, Sydney, NSW 1670, Australia
Email: lloyd.graham@csiro.au

Figure 1. Genomic locus for *CRNDE*. Exons (wide bars with colored fill) are shown approximately to scale; introns (black lines) are shown likewise but on a much smaller scale. Nucleotide numbering is for GRCh37/hg19. The region of greatest sequence conservation, here called gVC for “genomic Vertebrate Conserved” (stippled box), and transcribed extensions to exons (narrow bars with black fill) are depicted using the intron scale. The alternative exons E1A and E1B overlap by 9 nt. The hypothetical exons upstream of E1A and downstream of E6 correspond to speculative probe sets in the Affymetrix exon arrays. Numerical exon boundaries are provided in Supplementary Table S3. Transcript identification letters are based on those for hCG_1815491 in the Apr07 release of AceView, augmented by our own findings (e.g., In1T transcripts). While our nomenclature system would naturally identify transcript CRNDE-g and -h as CRNDE-g^{1B} and -g^{1A}, respectively, we have retained the former designations to maintain parity with AceView. Wherever the presence of a partial exon sequence in an AceView isoform is considered merely to reflect incomplete recovery of the actual sequence, the complete exon is shown. While the 2 major 3' boundary options for E6 are indicated, others are known but not shown (Suppl. Table S3). The current NCBI Reference Sequence (RefSeq) isoforms for *CRNDE* are identified on the left, while the original RefSeq for hCG_1815491 is shown at the bottom. Primer pairs (red arrows) are as listed in Supplementary Table S1B; any primer that spans an exon splice junction is shown as a bipartite arrow with a dotted line spanning the absent intronic sequence.

associated with CRC, showing a 42- and 16-fold increase (relative to normal mucosa) in adenoma and CRC, respectively.¹⁰

Since 2007, the NCBI AceView database has indicated at least 10 alternative RNA transcripts derived from the *CRNDE* locus (Fig. 1). Alternative splicing is now recognized to contribute to the pathogenesis of many diseases, including cancers.¹¹⁻¹³ While global splicing disorder is a feature of cancer cells in general,¹⁴ specific alternative splicing events have been observed to correlate with disease progression in CRC.¹⁵ In some cancers, the upregulation of

specific splice variants amounts to a “molecular signature” for neoplasia.^{16,17}

Here, we investigate the relationship of each *CRNDE* transcript to colorectal neoplasia. In addition, we define the boundaries and features of the *CRNDE* locus through exon microarray analyses, 5'-RACE, and transcript-specific RT-PCR assays. We compare *CRNDE* transcript abundance in colorectal and other cell lines, in neoplastic and nonneoplastic colorectal tissues, and in plasma from CRC patients and healthy individuals. From this, we show that certain *CRNDE* transcripts may have diagnostic utility.

Figure 2. Affymetrix expression microarray data for colorectal tissue. **(A, B)** Discovery data for CRNDE probe sets 238021_s_at and 238022_at, respectively, in Affymetrix HG U133 A/B arrays. Both are scatter plots of RMA-normalized fluorescence intensity values, expressed against an arbitrary index to provide a horizontal offset for each array. Sample categories are normal tissue (N) ($n = 222$), inflammatory bowel disease (I) (colitis) ($n = 42$), adenoma (A) ($n = 29$), and colorectal cancer (CRC) ($n = 161$). **(C)** Box and whisker plots of RMA-normalized fluorescence intensity for CRNDE probe sets in exon microarrays hybridized with cDNA from colorectal tissue. Sample categories are normal tissue (N) ($n = 5$), adenoma (A) ($n = 5$), and colorectal cancer (CRC) ($n = 5$). For each probe set, the change in expression signal for neoplastic relative to normal tissue is reflected qualitatively by the shading of the disc at the probe set location, as follows: little or no change in white, modest increase in gray (pink in online color art), and large increase in black (red in online color art). Larger versions of the probe set box plots are presented in Supplementary Figure S1C.

Results

Elevated expression of CRNDE in neoplastic colorectal tissue. A biomarker discovery exercise using Affymetrix HG U133 array data (Santa Clara, CA) for colorectal tissue samples (222 normal, 42 colitis, 29 adenoma, 161 CRC) (LaPointe *et al.*, submitted manuscript) identified expression from Chr16 locus hCG_1815491, hereafter referred to

as *CRNDE*, as elevated relative to normal tissue in >90% of adenomas and adenocarcinomas (Fig. 2A and 2B and Table 1). These findings were subsequently validated in a previously untested cohort of tissue specimens (30 normal, 19 adenoma, 19 CRC) using a custom-made microarray containing the 2 HG U133 probe sets for *CRNDE* (Table 1 and Suppl. Fig. S1A and S1B). In Table 1, we note other

Table 1. Differential Expression of CRNDE in Colorectal Tissue Samples Assessed by Microarray and qPCR

Microarray								
Probe set	Location	Normal Samples	Adenoma Samples		Cancer Samples			
		Signal average ^a	Fold change ^b	P value ^c	Fold change ^b	P value ^c		
Data of others								
238021_s_at ^d	E6	5.2	2.6 ^d	0.022	—	—		
238021_s_at ^e	E6	—	14.2	—	—	—		
238021_s_at ^f	E6	—	—	—	13.8	<0.001		
Discovery data								
238022_at	E2, E4, E5	2.8	2.2	0.008	2.6	<0.001		
238021_s_at	E6	3.0	4.6	0.001	6.1	<0.001		
Validation data								
238022_at	E2, E4, E5	2.1	1.3	0.002	1.4	<0.001		
238021_s_at	E6	2.1	1.0	0.629	1.1	0.019		
Exon array data ^g								
3692505	Pre-E1A	2.3	0.8	0.354	0.9	0.615		
3692527	E1A	4.8	1.5	0.185	1.5	0.408		
3692526	E1B	5.8	0.8	0.480	1.1	0.880		
3692525	E2	2.2	2.8	0.022	2.6	0.006		
3692524	E4	1.6	8.9	0.012	14.3	0.041		
3692523	ln4 3′	2.0	5.0	0.039	4.3	0.084		
3692522	E5 5′	1.6	4.5	0.048	4.7	0.083		
3692521	E5 3′	1.7	14.9	0.005	37.7	0.011		
3692520	ln5	2.5	5.8	0.002	8.9	0.017		
3692519	E6 5′	1.4	1.6	0.031	1.8	0.002		
3692518	E6, mid	1.1	1.2	0.329	1.0	0.754		
3692517	E6 3′	1.0	1.1	0.181	1.0	0.953		
3692504	Post-E6	3.1	1.4	0.173	1.5	0.334		
qPCR								
CRNDE	Composition	Normal Samples	Adenoma Samples		Cancer Samples			
		ΔC _t average ^a	Fold change ^b	P value ^c	Fold change ^b	P value ^c		
a	E1A-E2	6.5	1.8	<0.001	1.4	0.004		
b	′E1-E2-E4-′E5-E6	25.9	12.4	<0.001	8.8	<0.001		
c	′E1-E2-E3-E4-E5-E6	33.8	>99 ^h	0.002	>99 ^h	<0.001		
d	′ln4E5ln5E6	1.6	0.3	0.076	0.3	0.072		
e	′ln4E5-E6	28.8	7.4	<0.001	5.9	<0.001		
f	′E1-E2-E4(ln4)E5-E6	24.8	5.6	<0.001	5.0	<0.001		
g	′E1B-E2-E4-E5-E6	31.9	15.5	<0.001	12.5	<0.001		
h	′E1A-E2-E4-E5-E6	23.3	9.9	<0.001	7.8	<0.001		
i	′E1-E2-E4-′E5-E6	30.6	19.8	<0.001	12.5	<0.001		
j	′ln3-E4-E5-E6	25.9	9.7	<0.001	5.2	<0.001		
qPCR ROC: Adenoma Versus Normal								
CRNDE	AUC ⁱ	Threshold ^j	Sensitivity, %	Specificity, %	AUC ⁱ	Threshold ^j	Sensitivity, %	Specificity, %
b	0.939	0.130	91	96	0.921	0.125	85	96
g	0.870	0.180	80	96	0.892	0.095	80	96
h	0.938	0.165	95	96	0.888	0.135	80	96

^aAverage for normal samples expressed as the mean probe fluorescence intensity (microarray) or mean value for ΔC_t (qPCR), where $\Delta C_t = C_{t, \text{CRNDE}} - C_{t, \text{HPRT1}}$ and the C_t for each sample is itself a mean value ($n = 3$).

^bFold change relative to normal samples. For our microarray data, fold change is given as $(2^{i_{\text{Neo}}})_{\text{mean}} / (2^{i_{\text{Norm}}})_{\text{mean}}$, where i_{Neo} and i_{Norm} are the RMA-normalized probe fluorescence intensities for neoplastic and normal samples, respectively. For qPCR data, fold change is given as $(2^{-\Delta\Delta C_t})_{\text{mean}} / (2^{-\Delta\Delta C_t})_{\text{mean}}$, where $\Delta\Delta C_t = C_{t, \text{CRNDE}} - C_{t, \text{HPRT1}} - C_{t, \text{REF}}$ (see above), REF refers to cancer sample TB_163_97 (reference sample), and Neo and Norm relate to neoplastic and normal samples, respectively.

^cBy t test using Z' values (as defined above) for microarray, with no need to correct for multiple hypothesis testing, and by Mann-Whitney U test (Gaussian approximation) using $2^{-\Delta\Delta C_t}$ values for qPCR, for which extreme outlier values (5 of 690 values) were not included.

^dData from GEO Profile GDS2609 for normal-appearing colonic mucosa (sampled >8 cm from tumor) from early-onset nonfamilial CRC patients compared with equivalent tissue from healthy individuals.²⁰ In the absence of any other preadenoma data, we have presented this in the adenoma columns.

^eData from Sabates-Bellver et al.¹⁸

^fData from Kaneda et al.¹⁹

^gThree probe sets located in intron 1 (Fig. 1) have been omitted because their signals showed little or no change in neoplastic tissue (fold change values = 0.97-1.27), and the changes lacked statistical significance ($P = 0.24-0.96$).

^hThe disproportionately large increase is merely a reflection of the fact that this isoform was undetectable in many normal tissue samples, which were therefore assigned a value of $C_t = 60$ (see Materials and Methods and Suppl. Text S1).

ⁱAUC = area under receiver operating characteristic (ROC) curve.

^jFold change (i.e., $2^{-\Delta\Delta C_t}$) threshold above which the test outcome was deemed to be positive.

published microarray expression studies in which signals from the same probe sets were substantially elevated in CRC,¹⁸ in colorectal adenoma,¹⁹ and even in macroscopically normal-looking colonic mucosa from CRC patients.²⁰ In those studies, the observations had not been investigated further.

CRNDE locus and transcript splice variants. The NCBI AceView database indicated the occurrence of at least 10 differently spliced transcripts arising from at least 5 exons at the CRNDE locus (Fig. 1). Because the microarrays used thus far contained probes targeting only the 3' end of the CRNDE locus, we examined its expression more comprehensively in a small set of colorectal tissues (5 normal, 5 adenoma, 5 CRC) using exon microarrays. This indicated substantially elevated expression across exons E2, E4, In4, E5, In5, and part of E6 in neoplastic samples relative to nonneoplastic control tissue ($P < 0.05$) (Fig. 2C and Table 1). Most of the differential expression observed for exons E1A and E6 was not statistically significant, and no elevation was observed for E1B or for the hypothetical exons immediately upstream and downstream of the CRNDE locus. Our initial reverse transcriptase PCR (RT-PCR) analyses therefore focused on the E2-E6 region.

Endpoint RT-PCRs using oligonucleotide primers targeting CRNDE exons E2 (forward) and E6 (reverse) (Fig. 1) provided a visual assessment of expression from the CRNDE locus in neoplastic and normal colorectal tissue in neoplastic and normal colorectal tissue (6 normal, 6 adenoma, 6 CRC). Three PCR products (480, 365, and 280 bp) were observed in all 12 neoplastic samples but were largely absent from the nonneoplastic controls (Fig. 3A). The different amplicon sizes indicated the presence of at least 3 different mRNA transcripts, with the smallest corresponding to that expected from the NCBI Reference Sequence (RefSeq) isoform for

Figure 3. CRNDE RT-PCR on colorectal tissue and plasma RNA. Primer pairs given in Supplementary Table S1B. (A) Endpoint PCR with generic E2-E6 primers ($n = 6$ for each sample category). (B) qPCR with E2-E6 primers with cDNA from normal tissue ($n = 30$), adenoma ($n = 21$), and CRC ($n = 20$). Plots show $2^{-\Delta\Delta C_t}$ relative to normal tissue C40_96N (reference sample). The asterisk identifies anomalous sample TB_152_00, which had been classified as normal but which behaved as a cancer specimen in a range of assays including conventional tests unrelated to CRNDE expression. (C) Box and whisker plots of qPCR data from transcript-specific primers ($n = 3$ for each assay), with the CRNDE splice variant letter shown at the top left of each panel. Data for CRNDE-d could not be plotted on the otherwise common vertical axis range of 0 to 4 and therefore are not shown here. N = normal tissue ($n = 28$); A = adenoma ($n = 21$); CRC = colorectal cancer ($n = 20$). Plots show $2^{-\Delta\Delta C_t}$ relative to cancer tissue TB_163_97 (reference sample). The symbol ^ denotes one or more values beyond the scale of the graph. (D) Box and whisker plots and ROC plots of qPCR data for CRNDE-h in RNA extracted from plasma. Fold change is plotted as $2^{-\Delta C_t}$, where $\Delta C_t = C_t - (C_t^{\text{Normals mean}})$ and C_t denotes a mean value ($n = 3$ for each assay) normalized as described in Materials and Methods (Suppl. Text S1). AUC = area under ROC curve; FC > = fold change (i.e., $2^{-\Delta C_t}$) threshold above which the test outcome was deemed to be positive, here $2^{-\Delta C_t} = 2.8$. Full scatter plots for C (including CRNDE-d data) and D are shown in Supplementary Figure S3A and S3B.

hCG_1815491, equivalent to CRNDE-b (Fig. 1). Subsequent analysis of a larger independent panel (24 normal, 15 adenoma, 14 CRC) gave a similar outcome (Suppl. Fig. S2).

A quantitative RT-PCR (qPCR) analysis of colorectal tissue cDNA (30 normal, 21 adenoma, 20 cancer) with the E2-E6 primer pair confirmed elevated CRNDE expression in neoplastic colorectal tissue, in contrast to its low expression or absence in normal colorectal tissue (Fig. 3B). Transcript isoform-specific qPCR assays were then designed using features that were considered distinctive of the known CRNDE transcripts, such as characteristic splice junctions (Fig. 1), although many of the primer pairs could potentially amplify related variants that have not yet been identified. The results of these qPCR assays (Fig. 3C, Table 1, and Suppl. Fig. S3A) showed that, apart from CRNDE-d, all of the CRNDE splice variants were significantly elevated in expression ($P < 0.05$) in neoplastic colorectal tissue relative to normal colorectal specimens (28 normal, 21 adenoma, 20 CRC). ROC analyses for the ability of each transcript-specific qPCR assay to classify tissue phenotype (Table 1 and Suppl. Table S2) indicated that CRNDE-h could discriminate between adenoma and normal mucosa with a sensitivity (i.e., true-positive rate) and specificity (i.e., $1 - \text{false-positive rate}$) of 95% and 96%, respectively.

Identification of transcription start sites in cell lines. We refined our knowledge of the 5' end of the CRNDE locus by 5'-RACE, using E2- to E6-specific reverse primers (5'-RACE-1) on cDNA prepared from CaCo2 and HCT116 cells. Because the 2 cell lines yielded many amplicons of the same size, only the CaCo2 products were investigated in full (121 valid CRNDE sequences). Most of the CRNDE transcripts represented by intense 5'-RACE bands (major amplicons) commenced in the 3' end of E1A or E1B (Fig. 4A). However, approximately 16% of the transcripts represented by faint bands (minor amplicons) commenced with segments transcribed from intron 1 (In1) (Fig. 4B). The genomic sequence containing the transcriptional start sites within E1A/B (for both major and minor amplicons) is shown in Figure 4C, and plots showing all identified transcription start sites are available in Supplementary Figure S4A and S4B. Note that 5'-RACE-1 amplifications are intrinsically biased against transcripts lacking any of exons E2-E6, such as CRNDE-d, -e, and -j, and that amplicon length provides a further bias against the detection of relatively unspliced transcripts, such as CRNDE-d, -e, and possibly -f. Additional 5'-RACE and other amplifications targeting exons upstream of E1 or downstream of E6 did not yield any valid products (Suppl. Texts S2 and S3).

The 5' leader sequences of major transcripts were drawn from E1B about 1.5 times more often than from E1A, and the full-length E5 exon appeared to be approximately 18

times more common than the 5' truncated version, 'E5 (Suppl. Table S4). In1T family transcripts showed more diversity in the location of start sites for the transcribed intron segments (11 different initiation sites) than in where they ended (3 separate splice donor sites) (Suppl. Table S4). PCR using forward primers targeting transcribed In1 segments with reverse primers specific for E2 or E6 (Suppl. Table S1A and S1B) gave intense amplicon bands of the expected sizes with cDNA from HCT116 and CaCo2 cells (Suppl. Fig. S5A and S5B). When we performed identical amplifications on cDNA from colorectal tissue samples (11 normal, 6 adenoma, 18 CRC) (examples in Suppl. Fig. S5A and S5B), no amplification products were obtained, whereas control amplifications using the CRNDE E2-E6 primer pair gave intense product bands with most of the cDNA samples from neoplastic tissue (Suppl. Fig. S5C). Thus, although the In1T family transcripts are abundant in at least 2 CRC cell lines, they were not observed in normal or neoplastic colorectal tissue samples.

Transcript abundance profiles in cell lines and colorectal tissues. The relative abundance of CRNDE transcripts within each CRC or other cell line is shown in Figure 5A. Again, we caution that many of our qPCR primer pairs could potentially amplify isoforms that have not yet been identified. At face value, however, the dominant isoforms in CRC cell lines were CRNDE-a, -d, -e, -f, and -h, with appreciable expression also seen for CRNDE-b, -g, and -j. HT29, CaCo2, and HCT116 cells reproducibly exhibited high levels of CRNDE expression. In non-CRC cell lines, the most abundant transcripts were CRNDE-d and -f, with appreciable expression also seen for CRNDE-a, -h, and -j. MCF10A, a nontumorigenic mammary epithelial cell line with a relatively normal phenotype, reproducibly exhibited high levels of CRNDE expression. The near absence of CRNDE expression in LS174T CRC cells (Fig. 5A), in which the *wnt* pathway is constitutively active, means that CRNDE is unlikely to be a downstream target of *wnt* signaling. Furthermore, CRNDE expression did not change significantly upon inhibition of *wnt* signaling by the induction of a dominant-negative TCF4 (Suppl. Fig. S6).

In normal colorectal tissue ($n = 29$) (Fig. 5B), the transcript abundance profile was dominated by transcripts detected by the CRNDE-d primer pair (i.e., variants containing a 5' extension of E6 into In5) (Fig. 1). This signal appeared to decrease approximately 3-fold in neoplastic tissue (21 adenoma, 20 CRC), although statistically, the change was not quite significant ($P = 0.076$) (Table 1). In contrast, all of the other isoforms were upregulated (Table 1) so that the major transcripts in neoplastic colorectal tissue were CRNDE-a, -b, -d, -e, -f, -h, and -j (Fig. 5B). This coincides almost exactly with the list of isoforms showing high or appreciable expression in CRC cell lines. A survey

Figure 4. Transcription start sites in CaCo2 cells. **(A)** Major start sites, from intense amplicon bands obtained in 5'-RACE-I. **(B)** Minor start sites, from faint amplicon bands obtained in 5'-RACE-I. **(C)** Transcriptional start sites (major and minor combined) within the E1A (black uppercase, boxed with light fill, yellow fill in online color art) and E1B (dark gray uppercase italics, blue in online color art) regions are underlined, with initiation frequency reflected by nucleotide font size. Nucleotide numbering (horizontal axis in **A** and **B**) is for GRCh37/hg19, with nucleotides 54,962,872 to 54,962,336 shown in **C** (where intron 1 nucleotides are shown in gray lowercase).

of cDNA prepared from an independent panel of matched normal/CRC tissue samples, and confirmed to contain no genomic DNA ($n = 11$ pairs), again showed no significant change for CRNDE-d levels in CRC ($P > 0.05$ in paired or unpaired tests) (Suppl. Table S5). In contrast, CRNDE-b, -e, -f, -g, -h, and -i were significantly elevated in CRC relative to normal colorectal tissue from the same patient ($P < 0.05$ in paired test) (Suppl. Table S5). Once again, the major isoforms in neoplastic colorectal tissue were CRNDE-a, -b, -d, -e, -f, -h, and -j (not shown).

Measurement of CRNDE-h transcript in plasma. Among the better expressed transcripts, CRNDE-b and -h showed the greatest discrimination between normal and neoplastic colorectal tissue (Table 1). We focused on the latter because its actual level of expression in neoplastic tissue was higher (Table 1 and Fig. 5B) and because of preliminary PCR data showing the feasibility of its detection in plasma. We therefore measured by qPCR the level of CRNDE-h in the plasma of healthy individuals and CRC patients (Fig. 3D and Suppl. Fig. S3B). The mean value for the latter was 5.5

Figure 5. Relative abundance of transcripts within an individual cell line or tissue sample. Plots show $2^{-\Delta Ct}$, with CRNDE isoforms identified by letter along the horizontal axis. **(A)** Transcript-specific qPCR data for CRC cell lines (upper histogram) and non-CRC cell lines (lower histogram; noncancer cell lines are marked with an asterisk). The number of independent experiments for each cell line (n = number of bars) is given in the figure key. For CRC cell lines, values for the second HT29 and first LS174T bars have been scaled (divided by 3 and multiplied by 5, respectively) to facilitate co-plotting of data from different experiments, and the first and last CaCo2 bars show the mean values for repeated assays of 2 cDNA preparations (n = 8 and 6 assays, respectively); n = 3 for each assay. For non-CRC cell lines, values for some samples have again been scaled (the third PrEC bar divided by 20, the last 2 MCF10A bars divided by 10, and the last MRC5 bar multiplied by 5) to facilitate co-plotting. **(B)** Transcript-specific qPCR data for normal colorectal tissue (upper histogram; n = 28), colorectal adenoma (middle histogram; n = 21), and CRC (lower histogram; n = 20) samples; n = 3 for each assay. The transcripts detected within an individual RNA preparation are all shown using bars of the same color. Bars continuing beyond the displayed vertical scale are truncated with a horizontal line and their numerical value shown. In the upper histogram, asterisks identify conspicuous contributions by anomalous sample TB_152_00, which behaved as a cancer specimen in a range of assays (see legend to Fig. 3B). Primer pairs for both panels were as indicated in Supplementary Table S1B. Because qPCR data from non-DNase-treated cell/tissue RNA preparations had identical profiles to their DNase-treated counterparts, the figure incorporates both. For a discussion of this issue, see Supplementary Text S3.

times greater than that for the former, and the difference between the 2 data sets was statistically highly significant (P = 0.005 in a nonparametric Mann-Whitney U test). ROC analysis of the data indicated an optimal threshold of $2^{-\Delta Ct}$ = 2.8, with which the test result was positive for 13 of 15 cancer samples (sensitivity = 87%) but for only 1 of 15 normal controls (specificity = 93%) (Fig. 3D).

Discussion

Gene locus, alternative transcripts, and possible functions. Our expression array and RT-PCR data indicate that the CRNDE locus spans 6 exons, E1 to E6, located at nt 54,952,774 to 54,963,101 on Chr16 (Fig. 1 and Suppl. Texts S2 and S3). The promoter for E1-containing CRNDE

transcripts is expected to occur in the region immediately preceding the main transcription start sites identified by 5'-RACE (Fig. 4C), and the available data across a range of cell types for nucleosome occupancy, DNase I sensitivity, histone markings, and transcription factor binding (Suppl. Text S4) are consistent with this expectation. The promoter is therefore situated within a CpG island of 5.4 kb, which also includes the promoter of the *IRX5* gene transcribed from the opposite DNA strand (Suppl. Text S2). A separate promoter within In4 may drive the expression of CRNDE-d, the second most abundant CRNDE transcript in AceView (Fig. 1 and Suppl. Text S3).

In 5'-RACE, the most abundant transcripts proved to be E2-E4-E5-E6-containing messages headed by segments from the 3' end of E1A or E1B (Suppl. Table S4). These transcripts correspond closely to transcripts CRNDE-h and -g, respectively, which together account for approximately 40% of AceView accessions for this locus. The strong preference for full-length E5 over the 5' truncated form, 'E5, in major amplicons from 5'-RACE (ratio 18:1) (Suppl. Table S4) reflects the situation in AceView (ratio 24:1) and is embodied in the newly defined RefSeq entries for CRNDE (Fig. 1). The novel intron-headed transcripts that we detected in CaCo2 and HCT116 cells, but not in colorectal tissue, are discussed in Supplementary Text S2.

E2-E6 transcripts containing full-length E5, which predominated in 5'-RACE, possess an ORF that potentially encodes a polypeptide of 38 amino acids, which we call "ExP1" (the product of CRNDE-g and -h), whereas those containing the 5' truncated 'E5 exon potentially encode one of 84 residues ("ExP2", the product of CRNDE-b). Other CRNDE isoforms may encode other short polypeptides. None of the CRNDE polypeptides have significant homology to known proteins, and comparative genomics (Suppl. Text S4) indicates that key features for translation of ExP1/2 are poorly conserved in mammals more distantly related to us than the chimpanzee. On bioinformatics grounds, Cabili *et al.* classify CRNDE (there called XLOC_011950) as a large intergenic noncoding RNA (lincRNA), excluding it even from the marginal TUCP (transcripts of uncertain coding potential) category.²¹ Nevertheless, recent work has shown that short polypeptides expressed directly by "non-coding" RNAs act as important developmental regulators in *Drosophila*,²² so it remains possible that some CRNDE transcripts do express short yet functional peptides.

A recent bioinformatics search for sources of human linc RNAs identified some 3,300 candidate regions, including the CRNDE locus.²³ CRNDE transcripts (called "lincIRX5" by Khalil *et al.*) expressed in human fibroblast and HeLa cell lines were found to associate with the chromatin-modifying polycomb repressive complex 2 (PRC2) and with CoREST, potentially directing these complexes to silence the transcription of particular sets of genes.²³ siRNA-mediated knockdown of CRNDE did indeed affect the

expression of many genes²³; our analysis of these authors' microarray data indicated that 1,128 genes were downregulated (with CRNDE itself showing a 7-fold decrease) and 862 genes were upregulated. An Ingenuity Pathway Analysis (Ingenuity Software, Redwood, CA) of this data set (Suppl. Table S6) rated cancer, cell death, and cell cycle as top-scoring functions (lowest *P* values of 4.5×10^{-7} , 2.3×10^{-5} , and 1.5×10^{-4} , respectively). Because CRNDE transcripts appear to be important regulators of gene expression, an increase in transcription from this locus could constitute an independent early driver of neoplasia, or it may reflect oncogenic changes in genes and pathways for which *CRNDE* is a downstream target. We note, however, that CRNDE expression does not seem to be dependent upon *wnt* signaling (Suppl. Fig. S6).

CRNDE expression in normal and neoplastic colorectal tissue. We have presented expression microarray data showing that an increase in CRNDE expression is an early event in colorectal neoplasia, its transcription being elevated in >90% of colorectal adenomas and adenocarcinomas. The increase is also evident in published microarray studies^{18,19} and in searches of public databases (see below). Overall, we found the expression of CRNDE in normal colorectal tissue to be very low and not much elevated in nonneoplastic diseases such as inflammatory bowel disease (Fig. 2A and 2B). The neoplasia-related increase in CRNDE expression was confirmed and extended for colorectal tissue using RT-PCR and exon expression array studies.

Transcript-specific qPCR revealed that most of the 10 known CRNDE splice variants, namely, CRNDE-a to -c and CRNDE-e to -j, have elevated expression in neoplastic colorectal tissues. CRNDE-d, the dominant isoform in normal colorectal tissue, is the exception. This species is unusual in being assigned the most 3' transcription start site of the known transcripts and in being the only fully unspliced transcript known to arise from the locus. We should, however, bear in mind that any transcript containing a significant 5' extension of exon E6 into intron In5 will be detected by our CRNDE-d primer pair. The CRNDE-d signal is lower in neoplastic tissue (Fig. 5B and Suppl. Table S5), although in our experiments, the decrease was not statistically significant. The extent of CRNDE-d expression in normal tissue did not result in an appreciable signal on expression microarrays, and it is clear that any neoplasia-related decrease in CRNDE-d was inconsequential in comparison to the combined increase in the other CRNDE transcripts (Fig. 2A). This is discussed further below and in Supplementary Text S3.

While very low in all normal colorectal tissue, the expression of fully or partially spliced CRNDE transcripts (i.e., all known isoforms except CRNDE-d) is higher in the distal than in the proximal region of the colorectum (Fig. 3B and Suppl. Figs. S2 and S7), a pattern reported

previously for many genes.²⁴ The Unigene EST Profile database indicates that the *CRNDE* locus is significantly expressed in a range of normal tissues (ear, lymph node, etc.) and that its expression appears to decrease throughout development (Suppl. Table S7A). This is consistent with recent observations that a transcript from the mouse locus orthologous to *CRNDE* (called “linc1399” by Guttman *et al.*) is implicated in maintaining the pluripotency of embryonic stem cells.²⁵ While *CRNDE* is highly expressed in human induced pluripotent stem cells, its levels actually increase during neuronal differentiation; however, *CRNDE* appears not to be expressed in differentiated neurons, as it is detectable in fetal but not in normal adult brain tissue.²⁶

In disease, we note that *CRNDE* expression is also elevated in some non-colorectal cancers (Suppl. Table S7A). For example, *CRNDE* expression is reported to be upregulated 21-fold in hepatocarcinoma,²⁷ and the NextBio database shows that even greater increases have been seen in some cancers of the blood and brain (Suppl. Table S7B). The same database records elevations of up to 20- and 43-fold for *CRNDE* transcription in colorectal adenomas and CRC, respectively (Suppl. Table S7B and S7C).

Diagnostic potential of *CRNDE* transcripts and future work. With the exception of *CRNDE*-d, the known *CRNDE* splice variants are upregulated both in colorectal adenomas and cancers. The former observation is particularly important, as molecular markers for adenomas have received little attention in the literature. Because the potential of a colorectal adenoma for malignancy can be ended by polypectomy, biomarkers useful in clinical screening for adenomas could afford not just an early detection of CRC but a means of actually preventing it.²⁸ Even single-transcript assays for *CRNDE* expression can usefully distinguish neoplastic from nonneoplastic colorectal tissue. For example, ROC analysis of the level of *CRNDE*-h (one of the more abundant isoforms in neoplastic tissue) (Fig. 5B) in colorectal tissues showed that it afforded 80% and 95% sensitivity for cancer and adenoma, respectively, with a specificity value of 96% for both (Table 1).

The observation that *CRNDE* transcripts are enriched 2.3-fold (relative to cellular mRNA) in SW480-derived microvesicles²⁹ suggests that *CRNDE* transcripts may be present in exosomes shed from CRC cells and thus could potentially be detectable in blood fractions.³⁰⁻³² Indeed, in a preliminary survey of plasma samples, *CRNDE*-h afforded a sensitivity and specificity of 87% and 93%, respectively, for the presence of CRC (Fig. 3D), with a positive test result being returned by 13 of 15 cancer patient samples (87%) but only 1 of 15 normal controls (7%). The presence of such a biomarker for colorectal neoplasia (including adenomas) in plasma samples raises the hope that, with optimization, *CRNDE* transcripts may have clinical utility in screening for and diagnosing the disease. In this, it is not alone; highly promising RNA biomarker candidates such as CCAT1 have

recently emerged from other studies.³³ In view of the current need for better, cheaper, and less invasive tools for the prevention and detection of CRC, such candidate RNA biomarkers should be targeted for further assay development and subjected to clinical testing in larger cohorts. In addition to these priorities for *CRNDE*, we plan to investigate the biological function of the gene (including the possibility of *CRNDE* expression at the protein level) and to explore the role of nonexonic transcripts from this locus. Such activities are a timely expression of the growing interest in lincRNAs as diagnostic/prognostic indicators for cancer and reflect their potential as a new class of therapeutic targets.³⁴

Conclusions and impact. *CRNDE*, a novel locus unrelated to any known protein-coding gene, has been characterized (Fig. 1) and its transcription start sites in CRC cell lines identified (Fig. 4). Because *CRNDE* expresses at least 10 splice variants (Fig. 1), the transcript isoforms in CRC cell lines and colorectal tissue were quantified and the most abundant species identified as *CRNDE*-a, -b, -d, -e, -f, -h, and -j (Fig. 5). With the notable exception of *CRNDE*-d, *CRNDE* transcripts are upregulated at the earliest stages of colorectal neoplasia (Figs. 2 and 3 and Table 1), showing elevated expression in >90% of colorectal adenomas and adenocarcinomas. Some transcripts appear to have strong diagnostic potential as tissue and plasma biomarkers for colorectal neoplasia, including precancerous adenomas (Fig. 3C and 3D and Suppl. Table S2). For example, expression levels for *CRNDE*-h demonstrate a sensitivity of 95% and specificity of 96% for adenoma versus normal tissue (Suppl. Table S2 and Table 1), while a qPCR assay measuring *CRNDE*-h RNA levels in plasma was characterized by 87% sensitivity and 93% specificity for CRC patients versus healthy individuals (Fig. 3D). Because the potential of a colorectal adenoma for malignancy can be ended by polypectomy, the use of biomarkers such as this in clinical screening for adenomas could afford not just an early detection of CRC but also a means of actually preventing it.

Materials and Methods

Specimen collection and RNA preparation. All colorectal tissue specimens were classified as normal, adenoma, or adenocarcinoma (CRC) on the basis of histological assessment by an expert pathologist. In the discovery cohort (222 normal, 42 colitis, 29 adenoma, 161 CRC), 84% of the normal tissue samples came from individuals who were apparently free of any colorectal neoplastic disease (Suppl. Text S1). Colorectal tissue specimens for the validation microarrays (30 normal, 19 adenoma, 19 CRC) and for other experiments (including RT-PCR) came from colonoscopy patients via a tertiary referral hospital tissue bank in Adelaide, Australia (Suppl. Text S1). All of the normal samples in the validation cohort came from individuals who were apparently free of

synchronous colorectal neoplasia (Suppl. Text S1). A small independent panel of matched normal/CRC tissue samples ($n = 14$ pairs) was also sourced from the Adelaide center (Suppl. Text S1). Cell line provenance is described in Supplementary Text S1. Plasma specimens were collected under contract by ProteoGenex (Culver City, CA) from consenting individuals who had undergone colonoscopic screening some 4 to 6 days earlier, a procedure that did not involve tissue biopsy or polypectomy. Patient classification and plasma preparation are described in Supplementary Text S1; plasma was stored at -80°C within 4 hours of blood collection.

Expression microarrays. RNA extracted from colorectal tissue specimens obtained from consenting patients (222 normal, 42 colitis, 29 adenoma, 161 CRC; discovery cohort) was amplified using oligo-dT priming, and RNA expression profiles were analyzed using HG U133A/B gene chips (Affymetrix) by GeneLogic Inc. (Gaithersburg, MD). A custom microarray (Adenoma Biomarker Gene Chip), manufactured for us by Affymetrix, was used to measure RNA expression in a previously untested tissue specimen cohort (30 normal, 19 adenoma, 19 CRC; validation cohort); the data set is available via Gene Expression Omnibus (GEO) accession number GSE24713. More detailed information for the 2 cohorts, including the identity of the GeneLogic array data sets, is given elsewhere (LaPointe *et al.*, submitted manuscript). Human Exon 1.0 ST GeneChips (Affymetrix) were used with a subset of the validation cohort (5 normal, 5 adenoma, 5 CRC). For the latter 2 types of microarray, cDNA was prepared by random-primer amplification using the Affymetrix WT target labeling and control kit, and the arrays were processed according to the manufacturer's instructions.³⁵

RT-PCR of cell line and colorectal tissue RNA, including 5'-RACE. Total RNA from cell lines and colorectal tissue specimens was converted to cDNA, which then served as a template for endpoint and quantitative PCR (qPCR) amplifications; the procedures (including quality control and data normalization steps) are described in Supplementary Text S1, with primer information given in Supplementary Table S1A and S1B. 5'-RACE methods are also described in Supplementary Text S1; the main set of experiments, 5'-RACE-1, relied upon nested reverse primers targeting sites within exons E2-E6, while known and potential upstream exons were explored in a supplementary set of experiments, 5'-RACE-2 (Suppl. Table S1A and S1C).

qPCR assay for CRNDE-h transcript in plasma. RNA was extracted from 2 mL plasma (Suppl. Text S1) (15 normal, 15 CRC) that had been spiked with Armored RNA (arm-RNA) Enterovirus (Asuragen, Austin, TX) and converted to cDNA. CRNDE-h levels were measured using 15 μL of cDNA in a total qPCR reaction volume of 150 μL

containing 450 nM of each CRNDE-h transcript-specific primer (Suppl. Table S1A and S1B) and 125 nM of a fluorescently labeled hydrolysis probe. The procedures (including quality control and data normalization steps) are described in Supplementary Text S1.

Statistics and bioinformatics. Discovery, validation, and exon microarray expression data were analyzed as described elsewhere.²⁴ Receiver operating characteristic (ROC) analyses were performed using GraphPad Prism v.5.0a for Mac OS X (GraphPad Software, San Diego, CA). Statistical analyses for matched normal/CRC tissue pairs were performed using GraphPad Instat v.3.01. Genome and transcriptome analyses were based on UCSC Genome Browser³⁶ (mainly NCBI36/hg18) and NCBI AceView (Apr07); details are given in Supplementary Text S4. Microarray expression data sets for siRNA-mediated knockdown of CRNDE transcription²³ (GSM408726, GSM408729, GSM408731, and GSM452277-9) were downloaded from the GEO repository, RMA normalized, and compared using Partek Genomics Suite 6.5 (Partek Inc., St. Louis, MO) with a false detection rate <0.1 ; the output was then subjected to a functional analysis by Ingenuity Pathway Analysis (IPA) v8.6 (Ingenuity Software), content version 3002 (April 2010).

Acknowledgments

The authors thank the colorectal surgeons (especially David Watchow and Paul Hollington) at Flinders Medical Centre and Repatriation General Hospital for collection and storage of specimens. They also thank Deb Shapira (CSIRO) for some cell line RNA isolations and qPCR assays. They are grateful to Aaron Statham (Garvan Institute of Medical Research, Sydney) for recognizing lincIRX5 as CRNDE. They thank Hans Clevers' laboratory (Utrecht, the Netherlands) for the kind gift of LS174T cells engineered to express an inducible dominant-negative TCF4. They also thank Chris McLaughlin (CSIRO) for qPCR of CRNDE expression in this cell line.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: In relation to Clinical Genomics Pty. Ltd., S.K.P. and L.C.L. are employees with ownership interests, E.K.V. is a former employee, G.P.Y. is a paid consultant, and R.D. and P.L.M. received commercial research support from the company.

Funding

This work was supported by Clinical Genomics Pty. Ltd., a company involved in the discovery and commercialization of biomarkers for colorectal cancer; Flinders University of South Australia; and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia.

References

- Stein U, Schlag PM. Clinical, biological, and molecular aspects of metastasis in colorectal cancer. *Recent Results Cancer Res.* 2007;176:61-80.
- Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev.* 2007;(1):CD001216.
- Cunningham D, Atkin W, Lenz H-J, *et al.* Colorectal cancer. *Lancet.* 2010;375:1030-47.
- Atkin WS, Edwards R, Kralj-Hans I, *et al.* UK Flexible Sigmoidoscopy Trial Investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet.* 2010;375:1624-33.
- Young GP. Population-based screening for colorectal cancer: Australian research and implementation. *J Gastroenterol Hepatol.* 2009;24(Suppl 3):S33-42.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2: cancer incidence and mortality worldwide. IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Cited 29 Sep 2011. Available from: <http://globocan.iarc.fr>.
- Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer.* 2004;4:309-12.
- Hundt S, Haug U, Brenner H. Blood markers for early detection of colorectal cancer: a systematic review. *Cancer Epidemiol Biomarkers Prev.* 2007;16:1935-53.
- Chan SK, Griffith OL, Tai IT, Jones SJM. Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers. *Cancer Epidemiol Biomarkers Prev.* 2008;17:543-52.
- Nagaraj SH, Reverter A. A Boolean-based systems biology approach to predict novel genes associated with cancer: application to colorectal cancer. *BMC Syst Biol.* 2011;5:e35.
- Srebrow A, Kornblihtt AR. The connection between splicing and cancer. *J Cell Sci.* 2006;119:2635-41.
- Thorsen K, Sørensen KD, Brems-Eskildsen AS, *et al.* Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics.* 2008;7:1214-24.
- Venables JP, Klinck R, Koh C, *et al.* Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol.* 2009;16:670-6.
- Ritchie W, Granjeaud S, Puthier D, Gautheret D. Entropy measures quantify global splicing disorders in cancer. *PLoS Comput Biol.* 2008;4:e1000011.
- Wittig BM, Goebel R, Weg-Remers S, *et al.* Stage-specific alternative splicing of CD44 and alpha-6 beta-1 integrin in colorectal tumorigenesis. *Exp Mol Pathol.* 2001;70:96-102.
- Li HR, Wang-Rodriguez J, Nair TM, *et al.* Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. *Cancer Res.* 2006;66:4079-88.
- André F, Michiels S, Dessen P, *et al.* Exonic expression profiling of breast cancer and benign lesions: a retrospective analysis. *Lancet Oncol.* 2009;10:381-90.
- Sabates-Bellver J, Van der Flier LG, de Palo M, *et al.* Transcriptome profile of human colorectal adenomas. *Mol Cancer Res.* 2007;5:1263-75.
- Kaneda H, Arai T, Tanaka K, *et al.* FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth. *Cancer Res.* 2010;70:2053-63.
- Hong Y, Ho KS, Eu KW, Cheah PY. A susceptibility gene set for early onset colorectal cancer that integrates diverse signaling pathways: implication for tumorigenesis. *Clin Cancer Res.* 2007;13:1107-14.
- Cabili MN, Trapnell C, Goff L, *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011;25:1915-27.
- Kondo T, Plaza S, Zanet J, *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science.* 2010;329:336-9.
- Khalil AM, Guttman M, Huarte M, *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A.* 2009;106:11667-72.
- LaPointe LC, Dunne R, Brown GS, *et al.* Map of differential transcript expression in the normal human large intestine. *Physiol Genomics.* 2008;33:50-64.
- Guttman M, Donaghey J, Carey BW, *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* 2011;477:295-300.
- Lin M, Pedrosa E, Shah A, *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One.* 2011;6:e23356.
- Chang Q, Chen J, Beezhold KJ, Castranova V, Shi X, Chen F. JNK1 activation predicts the prognostic outcome of the human hepatocellular carcinoma. *Mol Cancer.* 2009;8:e64.
- Levin B, Lieberman D, McFarland B, *et al.* Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology.* 2008;134:1570-95.
- Hong BS, Cho JH, Kim H, *et al.* Colorectal cancer cell-derived microvesicles are enriched in cell cycle-related mRNAs that promote proliferation of endothelial cells. *BMC Genomics.* 2009;10:e556.
- Rabinowitz G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal microRNA: a diagnostic marker for lung cancer. *Clin Lung Cancer.* 2009;10:42-6.
- Wittmann J, Jäck HM. Serum microRNAs as powerful cancer biomarkers. *Biochim Biophys Acta.* 2010;1806:200-7.
- Nilsson J, Skog J, Nordstrand A, *et al.* Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. *Br J Cancer.* 2009;100:1603-7.
- Nissan A, Stojadinovic A, Mitrani-Rosenbaum S, *et al.* Colon cancer associated transcript-1 (CCAT1): a novel RNA expressed in malignant and pre-malignant human tissues. *Int J Cancer.* In press.
- Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* 2011;71:3-7.
- Pradervand S, Paillusson A, Thomas J, *et al.* Affymetrix Whole-Transcript Human Gene 1.0 ST array is highly concordant with standard 3' expression arrays. *Biotechniques.* 2008;44:759-62.
- Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC known genes. *Bioinformatics.* 2006;22:1036-46.