Published in final edited form as:

Mol Carcinog. 2012 May ; 51(5): 411–421. doi:10.1002/mc.20804.

Right-side and left-side colon cancer follow different pathways to relapse

Kerry M. Bauer¹, Amanda B. Hummon^{1,*}, and Steven Buechler^{2,*}

¹Department of Chemistry & Biochemistry, University of Notre Dame

²Department of Applied and Computational Mathematics and Statistics, University of Notre Dame

Abstract

There is growing evidence that cancer of the ascending (right-side) colon is different from cancer of the descending (left-side) colon at the molecular level. Using microarray data from 102 right-side colon carcinomas and 95 left-side colon carcinomas we show that different pathways dominate progression to relapse in right-side and left-side colon cancer. Right-side tumors at a high risk for relapse exhibit elevated expression of cell cycle control genes and elevated Wnt signaling. On the other hand, relapse-prone left-side tumors show elevated expression of genes that promote stromal expansion and reduced expression of tumor suppressor genes that initiate Wnt signaling. Single gene prognostic biomarkers are found separately for right-side and left-side disease. In left-side tumors with low expression levels of NADPH oxidase 4 (*NOX4*) the 5-year relapse-free survival probability is 0.89 95%CI(0.80 – 0.99), and in tumors with elevated *NOX4* expression the probability is 0.51 95%CI(0.37 – 0.70). Right-side tumors with elevated expression levels of caudal type homeobox 2 (*CDX2*) have a 5-year relapse-free survival probability of 0.88 95%CI(0.80 – 0.96), and those with low *CDX2* expression have a corresponding probability of 0.39 95%CI(0.15 – 0.78). Both *NOX4* and *CDX2* are much less prognostic on the opposite sides. This newly identified role of *NOX4* in colon cancer is further investigated using the SW620 lymph-node metastasis colon adenocarcinoma cell line and RNA interference. We show that *NOX4* is expressed in the SW620 cell line and that application of *NOX4* siRNA causes a significant reduction in reactive oxidative species production.

Keywords

colonic neoplasms; gene expression profiling; disease-free survival

Introduction

There is growing evidence that right-side colon cancer (RCC) and left-side colon cancer (LCC) have significantly different histological and molecular characteristics. The right side of the colon extends from the cecum or ascending colon through the transverse colon, excluding the appendix. The left side of the colon begins at the left splenic flexure, includes

Corresponding author: Steven Buechler, Department of Applied and Computational Mathematics and Statistics, 262 Hurley Hall, University of Notre Dame, Notre Dame, IN 46556. Phone: 574-631-6233. Fax: 574-631-4822. buechler.1@nd.edu.

*Contributed equally to the project

the descending colon and ends with the sigmoid, but does not include the rectum [1]. RCC is more common in women than men, and has a poorer prognosis than LCC [1,2]. RCC and LCC tend to have different gross macroscopic pathology [3]. At the molecular level, a significant number of genes are differentially expressed between RCC and LCC [4], and patterns of loss of heterozygosity and promoter methylation vary by location [5].

Molecular biomarkers for relapse risk are of clinical interest, and lead to insight into how a tumor progresses to metastasis. In recent years, methods of stratifying breast cancer patients according to relapse risk have been developed using multi-gene measures of mRNA concentrations [6] [7]. These tests measure the expression levels of numerous genes in the primary tumor and partition tumors into a poor prognosis group that is likely to metastasize, and a good prognosis group that is likely to be relapse-free. This work was framed around the clinical problem of identifying those patients who can avoid adjuvant chemotherapy with no significant increased risk of metastasis. Analysis of the pathways containing the genes in these prognostic panels shows that acceleration of the cell cycle plays a dominant role in the progression of estrogen-receptor positive breast cancer to metastasis. Several authors have proposed similar prognostic signatures for colon cancer [8] [9] [10]; however, none of these tests has the population-wide prognostic power seen in the breast cancer tests, hence they are of questionable clinical value. One possible explanation for the failure to find an effective panel of genes is the existence of multiple disease subtypes that follow different pathways to progression. For example, the breast cancer prognostic panels are only significant in estrogen-receptor positive tumors.

Here, we present significant evidence that different processes dominate progression to relapse in LCC and RCC. Using a publicly available microarray database and a method of building survival models developed in [11], we show that genes that are most predictive of relapse in LCC are much less significant in RCC, and vice-versa. In particular, in the LCC samples, elevated expression of NADPH oxidase 4 (*NOX4*) is highly predictive of relapse, while *NOX4* is largely unexpressed in RCC. The NOX family of genes has been implicated in cancer development by reactive oxidative species (ROS) in several forms of cancer [12], but *NOX4* has not been previously implicated in colon cancer progression. A significant percentage of the RCC samples that relapse have low expression levels of caudal type homeobox 2 (*CDX2*), while *CDX2* has normal expression levels in most LCC relapse cases. Besides *NOX4* and *CDX2*, we find other genes that are strongly predictive of relapse on one side but not the other.

The pathways whose alterations lead to relapse of colon cancer are identified by survival analysis of the genes annotated to these pathways. Many relapse-prone right-side tumors show elevated expression of cell cycle control genes, while there are no such indications in LCC. Elevated expression of genes that promote stromal expansion are a characteristic of poor prognosis left-side tumors. Also, different sets of Wnt signaling genes are altered in LCC and RCC. We show that the relevant altered genes are significantly correlated to the biomarkers identified above.

In this paper, the role of *NOX4* in colon cancer is further investigated using the SW620 lymph-node metastasis colon adenocarcinoma cell line and RNA interference. We verify

that *NOX4* is expressed in the SW620 cell line and show that application of *NOX4* siRNA causes a significant reduction in ROS production.

Materials and Methods

Statistical analysis

The language R <http://www.r-project.org/> was used for all statistical analyses. Survival models were fit with the R package *survival*. The microarray annotation package *hgu133plus2.db* in BioConductor <http://www.bioconductor.org/> was also used. The proportional hazard condition was verified with the `cox.zph` function. All p-values in survival models refer to the p-value of the logrank score of a Cox proportional hazard model (CPH). A CPH is considered statistically significant if the p-value of the logrank score is < 0.05 . The significance of a multivariate CPH compared to a CPH using a subset of the variables is measured using the log-likelihoods of the two models.

Microarray dataset of colon cancer samples

In this study we use the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) data series GSE14333 [13]. Samples are colorectal cancer specimens from the H. Lee Moffit Cancer Center in the United States and Royal Melbourne Hospital, Western Hospital, and Peter MacCallum Cancer Center in Australia. After preparation, the samples were hybridized to Affymetrix GeneChip® *hgu133plus2* arrays. In this paper, expression values are computed from the CEL files with *gcrma* [14]. The survival endpoint reported in GSE14333 is any relapse, distant or local. Since the third quartile of time to relapse in the dataset is 28 months, we censor the relapse data to 60 months for this study. No further follow-up was available on the Dukes stage D samples. The characteristics of the left-side and right-side tumors in the dataset are summarized in Table 1. More complete information about the patients is found in [13].

KEGG pathway analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>) identifies the component genes in selected pathways. The BioConductor package *hgu133plus2.db* is used to associate array probes with pathways.

Multistate survival models

In [11], we developed a method for defining survival models using gene expression data. This method is based on the principal that cancer in humans progresses through a series of discrete steps reflecting genetic alterations [15]. Informally, an array probe (gene) is called *multistate* if the probe's expression values naturally divide samples into two distinct subtypes, much like the bimodality of the estrogen receptor gene divides samples into ER+ and ER- subgroups. Just as with ER status, for a multistate probe p there is a threshold c , determined by statistical methods, such that the samples with expression values above c , denoted $p+$, form one component, and the samples with expression values below c , denoted $p-$, form the second component. In most multistate probes, one large component is approximately normally distributed, and the other smaller component is a tail to the right or

left. Many genes have nearly normal expression distributions, hence are not considered multistate.

In the multistate genes correlated with relapse, one of the components is highly enriched with poor prognosis patients. To further exploit the principle that a multistate probe represents distinct states, the expression vector for a multistate probe is replaced by a binary variable representing the two components. Here, the significance of a multistate probe in a survival model is measured by the p-value of a CPH using only the probe's binary variable.

Cell Culture

Colorectal cancer cell lines HCT 116, HT-29, SW480, SW620, DLD-1 and SW837 were purchased from the American Type Culture Collection (ATCC; Manassas, VA). SW620, SW480, DLD-1 and SW837 were maintained in RPMI 1640 medium (Invitrogen, Gaithersburg, MD) and HCT 116 and HT-29 were maintained in McCoy's 5A (Invitrogen, Gaithersburg, MD), all supplemented with 10% fetal bovine serum (Thermo Scientific, Pittsburgh, PA) and 2mM L-glutamine (Invitrogen, Gaithersburg, MD) and grown in 5% CO₂ at 37°C. Cell lines were used within 3 months after receipt or resuscitation of frozen aliquots thawed from liquid nitrogen. The provider assured the authentication of these cell lines by cytogenetic analysis.

NOX4 silencing by siRNA

At 50–60% confluence, SW620 cells were transfected with one of two siRNA oligonucleotides targeting the NOX4 transcript. The sequences are referred to in the text as siRNA NOX4_5 and siRNA NOX4_8 and correspond with the following sequences: 5'-CCAGGAGAUUGUUGGAUAATT-3' and 5'-GAGUUUCCAUAGGGAACUATT-3', respectively. The oligo AllStar negative control siRNA was used as a negative control and the oligo All Star Death siRNA was used as a positive control siRNA to assess transfection efficiency. All siRNAs used were obtained from Qiagen Inc. (Germantown, MD). For cell viability studies, each siRNA (2 pmol) was added to individual wells in a 96-well plate in 25 µL of serum-free RPMI and complexed with 0.5 µL Oligofectamine transfection reagent (Invitrogen, Gaithersburg, MD) in 25 µL of serum-free RPMI. The resulting mixture was allowed to complex for 30 min at ambient temperature. Next, SW620 cells (5000 cells/well) were added in 50 µL RPMI supplemented with 20% FBS to yield transfection mixtures consisting of 20 nM siRNA in RPMI with 10% FBS. The final mixture was incubated at ambient temperature for 45 min before being placed in an incubator in 5% CO₂ at 37°C. For analysis of mRNA and ROS production, transfections were performed as described above except that they were conducted in 6-well plates and all reagent amounts were scaled up 30-fold. For characterization studies, each siRNA was evaluated by comparing cell viability, mRNA or ROS levels with those found in cells transfected with negative-control siRNA. siRNA knockdown was validated using real-time PCR.

Cell viability assay

As a surrogate marker for cell viability, the reduction of resazurin to resorufin was measured in transfected cells using the Cell Titer-Blue Cell Viability assay (Promega, Madison, WI). Triplicate transfections for each siRNA duplex were set up in a 96-well plate at a

concentration of 5,000 cells/well. After 72 and 96 hr, 20 μ L of CellTiter-Blue reagent was placed in each well and incubated for 1 hr. Reduction of resazurin to resorufin, read as fluorescence emission (560Ex/590Em) was measured using a plate reader (Spectramax M5, Molecular Devices, Sunnyvale, CA). Viability of transfected cells was compared with cells transfected with the negative control siRNA. Additional cells transfected with Qiagen's All Star Death control were used as a positive control.

Gene expression analysis

Total RNA was extracted from HCT 116, HT-29, SW480, SW620, DLD-1 and SW837 cells 48 hr following seeding in addition to SW620 cells transfected with siRNA NOX4_5, siRNA NOX4_8 and control siRNA 48 hr post-transfection using RNeasy Mini kit (Qiagen, Germantown, MD), following the enclosed animal cell protocol and homogenizing via 20 gauge needles. Normal human colon RNA isolated postmortem from a donor was purchased from Ambion (Applied Biosystems, Foster City, CA). Nucleic acid quantity, quality and purity were determined using a Nanodrop 2000 UV-VIS spectrophotometer (Nanodrop, Rockland, DE). cDNA was generated using the High-Capacity Reverse Transcriptase cDNA kit (Applied Biosystems, Foster City, CA) and 1.0 μ g of total RNA according to the manufacturer's instructions. Quantitative PCR reactions were performed using the following primer sequences (Operon, Huntsville, AL): hypoxanthine phosphoribosyltransferase 1 (HPRT1), HPRT1 For 5'-GCCATGAAGCAGGACTCTAAAGA-3' and HPRT1 Rev 5'-TTGGCATAACACAGCTGATTGAT-3'; NADPH oxidase 4 (NOX4), NOX4 For 5'-ATGTCAGTTGCTGCATTCCTAA-3' and NOX4 Rev 5'-TCACTCAATAGTGCTGTGGTTT-3'; prostate cancer susceptibility candidate (PRAC) PRAC For 5'-CCATTTCTCAGATCAAGGAC and PRAC Rev 5'-GTCTCGCCAGTAGATGTTT. Quantitative PCR was performed with a real-time PCR system, StepOnePlus (Applied Biosystems, Foster City, CA). Reactions were conducted with 300 ng of cDNA, in a final volume of 25 μ L. The PCR mixture contained SYBR Green (Applied Biosystems, Foster City, CA) and 0.6 nmol of each primer (forward and reverse). The levels of transcripts were quantified using the comparative C_T method relative to levels of hypoxanthine phosphoribosyltransferase (HPRT1). All samples were analyzed in triplicate wells with the median of each measurement used for C_T calculations.

Measurement of Superoxide Production

SW620 cells were assayed for superoxide production 48 hr post transfection with siRNA NOX4_5, siRNA NOX4_8 or negative control siRNA. Transfected cells were washed with phosphate-buffered saline (Invitrogen, Gaithersburg, MD) and collected into a centrifuge tube. Superoxide production was measured by chemiluminescence with DIOGENES (National Diagnostics, Atlanta, GA), a superoxide-specific, luminol-based detection system, according to the manufacturer's instructions. Measurements were performed in 96-well microtiter chemiluminescence plates (5×10^4 cells per well). The total integrated light units recorded (Spectramax M5, Molecular Devices, Sunnyvale, CA) from siRNA NOX4 transfected cells were compared to those recorded in cells transfected with negative-control siRNA.

Results

The focus of this paper is relapse in colon cancer, so we restrict our attention to the right-side and left-side samples in GSE14333 with Dukes stage A, B or C.

Different pathways dominate progression to relapse in LCC and RCC

The pathways or biological processes that have a significant role in progression to relapse in RCC and LCC are identified through analysis of the component genes in the KEGG pathway annotation. On the hgu133plus2 array platform there are 10,300 probes annotated to a KEGG pathway. To focus on probes that vary significantly we restrict the analysis to the 1808 probes that have an interquartile range > 1 on the samples in GSE14333 having Dukes stage A, B or C. For each such probe, we calculate a CPH for relapse on the left-side tumors with the probe's expression vector as the only variable, and do similarly for the right-side tumors. The following pathways (with associated KEGG code) contain multiple probes that are highly significant in a CPH on the right side or left side: focal adhesion (04510), Wnt signaling (04310), p53 signaling (04115), cell cycle progression (04110), complement and coagulation cascade (04610), TGF- β signaling (04350) and insulin signaling (04910). On the left side 81 probes are connected to one of these pathways and significant in a CPH model for relapse; on the right side there are 109 such probes. Table 2 reports the most significant genes on the right side and the most significant genes on the left side along with the pathways containing these genes.

In the relapse-prone left-side tumors, there is significant elevation of the following genes involved in focal adhesion, cell migration and growth of the extracellular matrix: integrin-binding sialoprotein (*IBSP*), fibronectin (*FNI*), thrombospondin 1 (*THBS1*), v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma) (*AKT3*), and several species of collagen.

Among the genes most significantly predictive of relapse on the left side is wingless-type MMTV integration site family, member 5A (*WNT5A*), which is down regulated in the samples that will relapse. Secreted frizzled-related protein 2 (*SFRP2*), which competes with the Wnt proteins for the Frizzled receptor, is up regulated. Also, the frizzled receptor, frizzled homolog 3 (*FZD3*), is down regulated in the relapse cases. These expression changes point to a reduction in degradation of β -catenin by Wnt signaling in the left-side tumors that eventually relapse. In the poor prognosis right-side tumors there are signs of altered Wnt signaling, however through different sets of genes. Expression of the above-mentioned genes appears to be normal, however, Axin2 is down regulated in the relapse cases on the right side, reducing transcriptional inhibition by β -catenin. Moreover, cyclin D1 (*CCND1*), which is transcribed by β -catenin, is up regulated in many relapse cases. The most striking feature of relapse on the right side is elevated expression of cyclin-dependent kinase inhibitor 2B (p15, *CDKN2B*), growth arrest and DNA-damage-inducible, alpha (*GADD45A*) in the relapse cases, as well as *CCND1*. This points to a cohort of highly expressed genes associated with proliferation in the right-side tumors, of which there is no such indication on the left side. Genes involved in p53 signaling are altered on both sides, although more genes are altered on the right side. There are 30 probes from the selected pathways significantly implicated in relapse on both sides. These common probes are largely

involved in focal adhesion, plus alteration of cell adhesion and migration by serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1 (*SERPINE1*), plasminogen activator, urokinase (*PLAU*) and plasminogen activator, urokinase receptor (*PLAUR*).

Single multistate genes are strongly predictive of relapse in left-side and right-side tumors and encapsulate pathway activity

The multistate methodology is applied separately to the left-side tumors and the right-side tumors to identify multistate probes that are significantly predictive of relapse. These genes also act as biomarkers for the pathways to relapse described in the preceding section.

Application of the multistate methodology to the left-side tumors with Dukes stage A, B or C identifies 219773_at (*NOX4*) as one of the most significant predictors of relapse. The distribution of *NOX4* in the left-side tumors shows a large component with low mean expression and narrow variance, and a right tail of elevated expression (Figure 1(A)). The multistate methodology divides the two components at the expression value 3.0. As the boxplot in Figure 1(B) shows, there is a strong correlation between elevated *NOX4* expression and relapse within 5 years. *NOX4* expression is summarized as a binary variable by assigning 1 to every sample with expression level above the cut value (the *NOX4*+ component), and 0 for the other samples (*NOX4*–). The Kaplan-Meier survival curve for the binary *NOX4* variable (Figure 1(C)) shows the prognostic power of this variable in the left-side tumors. The corresponding plot for the right-side tumors is reported in Figure 1(D). *NOX4* has a distinctly lower significance on the right side.

The rate of relapse in colon cancer varies significantly by tumor stage (Table 1). We show that the prognostic power of *NOX4* significantly improves on the power of tumor stage as follows. On the left-side samples, Dukes Stage is a significant predictor of relapse with a p-value of 0.004. A CPH model with variables for both Dukes Stage and *NOX4*, significantly improves on a model with Dukes Stage alone with a p-value of 0.006. Adjuvant chemotherapy is known to improve 5-year survival in colorectal cancer in a large clinical trial [16]. On the left-side samples in this study, the patients who received chemotherapy have significantly better 5-year survival ($p = 0.004$). However, on the *NOX4*– samples; i.e., the good prognosis group, chemotherapy does not improve survival. This supports the use of *NOX4* in deciding which colon cancer patients can safely avoid chemotherapy.

Low expression of the probe 206387_at (*CDX2*) is highly predictive of relapse in the right-side tumors (Figure 2(B)). Figure 2(A) illustrates the partition of the right-side samples into low (*CDX2*–) and high (*CDX2*+) components. The survival characteristics of the binary *CDX2* variable in the right-side tumors and the left-side tumors are reported in Figure 2(C,D). On the right-side samples, *CDX2* significantly improves on prognostic power of Dukes Stage with a p-value of 0.0001. Contrary to the behavior on the left side, chemotherapy does not increase the probability of survival in the full set of right-side tumors.

In addition to *NOX4*, matrix metalloproteinase 3 (stromelysin 1, progelatinase) (*MMP3*) is a multistate gene that is predictive of relapse in left-side tumors. In this case, the low

component is enriched with relapse cases. A CPH with a binary variable representing the *MMP3*+/- components has a p-value 3.86×10^{-6} . Although 2/3 of the *MMP3*- samples are in *NOX4*+, *MMP3* provides independent information about relapse in the left-side samples. In the right-side tumors, the high component of family with sequence similarity 69, member A (*FAM69A*) is enriched with relapse cases not identified by *CDX2*.

In addition to separating the population by prognosis, the multistate genes capture the pathogenic effects of the genes listed in Table 2, hence, the pathways containing these genes. For example, the most significant gene in the left-side analysis of pathway genes is *IBSP*. The *NOX4*+ component, in addition to containing the samples with elevated *NOX4* expression contain the samples with elevated *IBSP* expression. Assessing the relationship quantitatively, a t-test for the mean expression value of *IBSP* in *NOX4*+ versus *NOX4*- has a p-value of 1.38×10^{-5} . The multistate gene *MMP3* similarly represents the next most significant gene, *WNT5A*. In Table 2, for each probe, we identify the multistate gene that separates the gene's expression into high and low components in a statistically significant manner. On the right side, *CDX2* effectively represents most of the probes listed in Table 2; however, some the genes are better represented by *FAM69A*.

***NOX4* mRNA interference with siRNA in SW620 significantly reduces superoxide production levels without affecting cell viability**

We surveyed the mRNA expression levels of *NOX4* in several human colon cancer cell lines. We found cell lines derived from primary adenocarcinoma or carcinoma colon tumors (HCT 116, HT-29 and SW480) to have *NOX4* expression levels below or comparable to normal colon *NOX4* levels. However, we found *NOX4* expression to be greatly elevated compared to normal colon in the adenocarcinoma cell line derived from the lymph metastatic site (SW620) as reported in Figure 3(A).

To investigate *NOX4* function in metastatic SW620 cells we examined the affect of *NOX4* on cell viability. *NOX4* expression was silenced using RNA interference by transfecting SW620 cells with oligonucleotides targeting the *NOX4* transcript (Figure 3(B)). We observed similar cell viability between *NOX4* silenced cells and negative control cells as reported in Figure 3(C). Therefore, targeted *NOX4* knockdown does not seem to affect cell viability. This result confirms previous studies reporting that suppressing NAD(P)H oxidases in colon cancer cells does not affect cell viability [17].

To further investigate *NOX4* function in SW620 cells, we assayed superoxide production by the chemiluminescent method. We found that *NOX4* inhibition with RNA interference in SW620 cells is associated with a decrease in superoxide producing activities of the cells as indicated by reduced ROS production (Figure 3(D)).

The locations of the sources of colon cancer cell lines are hypothesized

The results in this paper and others suggest that RCC and LCC should be studied as separate diseases. While there are numerous colon cancer cell lines, rarely is the location of the tissue source reported. One exception is DLD-1, where the original manuscript [18] clearly indicates that the tissue sample originated on the left side. The gene prostate cancer

susceptibility candidate (*PRAC*) is specifically expressed in the prostate, rectum and distal colon [19]. Figure 4(A) shows expression of a probe for *PRAC* in the right-side and left-side tumors from GSE14333. *PRAC* is highly expressed in almost all left-side tumors and expressed at a low baseline level, or unexpressed, in the right-side samples. Figure 4(B) reports the expression level of *PRAC* in several human colorectal cancer cell lines. The cell lines of unknown tumor origin were compared to DLD-1 (distal colon) cell lines to distinguish between descending and ascending colon derivation. SW460 and SW620 express *PRAC* at levels greater than both DLD-1 and SW837 while HT-29 and HCT 116 have almost non-detectable *PRAC* expression levels. This is significant evidence that DLD-1, SW480 and SW620 are left-side cell lines, and HT-29 and HCT 116 are right-side cell lines. In future research we will develop a panel of genes whose expression patterns predicts the location of the source of a colon tissue sample with a high degree of accuracy.

Discussion

In this paper we show that progression of colon cancer is dominated by different processes in left-side and right-side tumors in GSE14333. This adds to the growing evidence that left-side and right-side colon cancer should be treated as separate diseases. Even if many of the same processes lead to neoplasm, the tumors on the left side and right side have different macroscopic properties, different dominant pathways to relapse, and hence may require different treatment methods.

Sporadic colon cancer is a highly heterogeneous disease having subtypes identified by numerous phenotypic and genotypic traits that play a role in tumorigenesis. Among these traits are microsatellite instability (MSI) status, flat versus polypoid tumor type, adenoma-carcinoma or serrated pathway [20], and the mutation statuses of v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*), tumor protein p53 (*TP53*), and adenomatous polyposis coli (*APC*). Based on the accumulated evidence, tumor location must be added to this list. While these traits are known to influence tumorigenesis, their roles in metastasis may not be understood. Moreover, these roles may be confounded by tumor location or other factors. For example, LCC patients with a *KRAS* mutation have poorer prognosis than the wild type patients, however in RCC *KRAS* mutation status does not affect prognosis [21]. Our approach to the study of metastasis in colon cancer was to identify predictive and prognostic gene expression patterns as independently as possible from other traits. While we found that tumor location is a fundamental factor affecting the pathways to metastasis, the relationships between these metastatic pathways and other traits is a problem for future studies. The results presented here suggest that some of the multiple phenotypes mentioned above share some critical metastatic pathway components, however this too requires deeper analysis with a significant number of patient samples from each phenotype.

NOX4 has been shown to play a role in myofibroblast differentiation and fibrinogenesis in lung mesenchymal cells [22]. Transforming growth factor beta 1 (*TGFB1*) induces *NOX4* expression, and H₂O₂ produced by *NOX4* is necessary for *TGFB1*-induced myofibroblast differentiation, extracellular matrix production and contractility. Moreover, suppression of *NOX4* with siRNA reduces fibrinogenesis in two murine models of lung injury. In LCC *NOX4* expression is coupled with elevated expression of fibronectin and several species of

collagen. Combined, these are strong indications that stromal expansion plays a primary role in metastasis in LCC.

Overproduction of reactive oxidative species (ROS) has long been recognized as a risk factor in carcinogenesis. The NOX family of genes has been implicated in ROS-induced cancer development in several forms of cancer [12]. We examined expression of *NOX4* in four cell lines: HT-29, HCT 116, SW480 and SW620, the patient-matched lymph-node metastasis to SW480 (Dukes stage B). In a pancreatic cancer cell line, inhibition of *NOX4* activates apoptosis via the AKT-ASK1 cell survival pathway [23]. In our study, *NOX4* inhibition in SW620 shows no decrease in cell viability; however, reduction of *NOX4* expression via siRNA-mediation corresponds to a significant reduction in ROS production in the SW620 cells. This finding suggests that *NOX4* is a novel source of ROS production in metastatic and pre-metastatic colon cancer. If *NOX4* exerts cancer-promoting effects, it is most likely at more advanced tumor stages, as *NOX4* expression is comparable to normal colon levels in primary adenocarcinoma and carcinoma derived cell lines and above normal colon levels in the metastatic cell line (Figure 3(A)). In renal cell carcinoma, *NOX4* is critical for HIF2- α transcriptional activity [24], in particular, inhibition of *NOX4* decreases HIF2- α production. In the left-side samples in the microarray data there was no change in HIF2- α expression between the NOX4⁻ and NOX4⁺ component. In fact, the only change in hypoxia-related gene expression was a small decrease in HIG1 hypoxia inducible domain family, member 1A and HIG1 hypoxia inducible domain family, member 2A expression in NOX4⁺ over NOX4⁻.

Low expression of *MMP3* is implicated in left-side disease progression. The matrix metalloproteinases are a family of extracellular proteinases that mediate many of the changes in the tumor microenvironment during cancer progression [25]. The genes correlated to *MMP3* in expression (see Table 2) point to a significant role for Wnt signaling in left-side disease progression.

In RCC, low expression of *CDX2* is strongly associated with relapse. *CDX2* acts as a transcription factor, initially expressed during embryogenesis in the development of the small intestine and colon, and regulating a diverse range of functions from proliferation, cell-cycle arrest, differentiation, and apoptosis [26]. In healthy adult colon tissue, *CDX2* is expressed throughout the colon and regulated post-translationally through phosphorylation. Expression patterns of *CDX2* are altered after carcinogenesis. Analysis of 65 colorectal tumors mapping *CDX2* expression throughout the colon and rectum found significantly lower expression of *CDX2* in 37 right-sided, poorly differentiated tumors as compared to 28 left-sided tumors [27]. Methylation of the *CDX2* promoter has been proposed as a mechanism for down-regulation in colorectal carcinomas [28]. *CDX2* inhibits the Wnt signaling pathway, through reduction of the tyrosine phosphorylation of β -catenin, resulting in decreased T-cell factor signaling and cell proliferation [29]. With the reduced expression of *CDX2* that accompanies carcinoma, it has been described as a tumor-suppressor. In addition, *CDX2* regulates E-cadherin trafficking to the cell membrane [30].

FAM69A is located at 1p22.1, a genomic region that is preferentially deleted in microsatellite stable colon tumors [31]. Expression patterns of the genes in this region do not

show signs of deletion in the microarray data used here. The mechanism by which *FAM69A* expression is correlated with relapse risk remains an open problem for future study.

Our analysis suggests that disease progression in RCC is dominated by elevated expression of cell cycle control genes and elevated Wnt signaling. Up regulation of *CCND1*, which is transcribed by β -catenin, is accompanied by increased expression of the pro-apoptotic gene *GADD45A* and elevation of the growth arrest gene *CDKN2B*. Thus, the tumors that have not yet metastasized may be in a cycle of rapid mitosis and apoptosis. Inhibition of one of these genes by a subsequent genetic change could lead to unrestricted growth.

There are also signs of Wnt signaling activation in poor prognosis LCC, although through a different route. Initiation of Wnt signaling is reduced by lower expression of *WNT5A* and the frizzled receptor, *FZD3*, and elevated expression of *SFRP2*, which competes with the Wnt proteins for the frizzled receptor. These expression changes suggest that β -catenin stabilization and subsequent degradation is inhibited in these samples. That there are metastatic subtypes involving Wnt signaling in both RCC and LCC is understandable given the high frequency of genetic alterations of *APC* in colon cancer [32].

Inherited colon cancers such as familial adenomatous polyposis and hereditary non-polyposis colon cancer form important subgroups of the disease. Based in specific germline mutations, we expect these diseases to follow different pathways from the sporadic cases that are in the majority. The pathways discussed here dominate the general population of colon cancer cases but are not meant to be comprehensive.

It has previously been proposed that the differences in survival between RCC and LCC could be the results of any number of causes, for example difference in time of detection, embryologic origin, exposure to fecal matter or genetics [1]. This paper shows that regardless of the underlying cause, different mechanisms dominate progression of RCC and LCC, providing further evidence that they should be treated as different diseases. The prominent role of *NOX4* as a prognostic biomarker in LCC makes it an important target for future research in cancer biology and therapeutics.

Abbreviations

NOX4	NADPH oxidase 4
CDX2	caudal type homeobox 2
RCC	right-side colon cancer
LCC	left-side colon cancer
ROS	reactive oxidative species
CPH	Cox proportional hazard model
PCR	polymerase chain reaction
HPRT1	hypoxanthine phosphoribosyltransferase 1
PRAC	prostate cancer susceptibility candidate

WNT5A	wingless-type MMTV integration site family, member 5A
SFRP2	secreted frizzled-related protein 2
FZD3	frizzled receptor, frizzled homolog 3
CCND1	cyclin D1
CDKN2B	cyclin-dependent kinase inhibitor 2B
GADD45A	growth arrest and DNA-damage-inducible, alpha
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1
PLAU	plasminogen activator, urokinase
PLAUR	plasminogen activator, urokinase receptor
MMP3	matrix metalloproteinase 3 (stromelysin 1, progelatinase)
FAM69A	family with sequence similarity 69, member A
IBSP	integrin-binding sialoprotein
THBS1	thrombospondin 1
FN1	fibronectin
AKT3	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)

References

1. Meguid RA, Slidell MB, Wolfgang CL, Chang DC, Ahuja N. Is there a difference in survival between right- versus left-sided colon cancers? *Ann Surg Oncol*. 2008; 15(9):2388–2394. [PubMed: 18622647]
2. Benedix F, Kube R, Meyer F, et al. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. *Dis Colon Rectum*. 2010; 53(1):57–64. [PubMed: 20010352]
3. Papagiorgis P, Oikonomakis I, Karapanagiotou I, Wexner SD, Nikiteas N. The impact of tumor location on the histopathologic expression of colorectal cancer. *J BUON*. 2006; 11(3):317–321. [PubMed: 17309156]
4. Glebov OK, Rodriguez LM, Nakahara K, et al. Distinguishing right from left colon by the pattern of gene expression. *Cancer Epidemiol Biomarkers Prev*. 2003; 12(8):755–762. [PubMed: 12917207]
5. Azzoni C, Bottarelli L, Campanini N, et al. Distinct molecular patterns based on proximal and distal sporadic colorectal cancer: arguments for different mechanisms in the tumorigenesis. *Int J Colorectal Dis*. 2007; 22(2):115–126. [PubMed: 17021745]
6. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004; 351(27):2817–2826. [PubMed: 15591335]
7. Van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002; 347(25):1999–2009. [PubMed: 12490681]
8. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol*. 2004; 22(9):1564–1571. [PubMed: 15051756]
9. Barrier A, Lemoine A, Boelle P-Y, et al. Colon cancer prognosis prediction by gene expression profiling. *Oncogene*. 2005; 24(40):6155–6164. [PubMed: 16091735]

10. Webber EM, Lin JS, Whitlock EP. Oncotype DX tumor gene expression profiling in stage II colon cancer Application: prognostic, risk prediction. *PLoS Curr.* 2010;2.
11. Buechler S. Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer. *BMC Cancer.* 2009; 9(1):243. [PubMed: 19619298]
12. Kamata T. Roles of Nox1 and other Nox isoforms in cancer development. *Cancer Sci.* 2009; 100(8):1382–1388. [PubMed: 19493276]
13. Jorissen RN, Lipton L, Gibbs P, et al. DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers. *Clinical cancer research.* 2008; 14(24):8061–8069. [PubMed: 19088021]
14. Wu Z, Irizarry R, Gentleman R, Martinez-Murillo F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association.* 2004
15. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000; 100(1):57–70. [PubMed: 10647931]
16. Quasar, Collaboration Group; Gray, R.; Barnwell, J., et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet.* 2007; 370(9604):2020–2029. [PubMed: 18083404]
17. Geiszt M, Lekstrom K, Brenner S, et al. NAD(P)H oxidase 1, a product of differentiated colon epithelial cells, can partially replace glycoprotein 91phox in the regulated production of superoxide by phagocytes. *J Immunol.* 2003; 171(1):299–306. [PubMed: 12817011]
18. Dexter DL, Spremulli EN, Fligel Z, et al. Heterogeneity of cancer cells from a single human colon carcinoma. *Am J Med.* 1981; 71(6):949–956. [PubMed: 7315857]
19. Liu XF, Olsson P, Wolfgang CD, et al. PRAC: A novel small nuclear protein that is specifically expressed in human prostate and colon. *Prostate.* 2001; 47(2):125–131. [PubMed: 11340635]
20. Young J, Jass JR. The case for a genetic predisposition to serrated neoplasia in the colorectum: hypothesis and review of the literature. *Cancer epidemiology, biomarkers & prevention.* 2006; 15(10):1778–1784.
21. Bleeker WA, Hayes VM, Karrenbeld A, et al. Impact of KRAS and TP53 mutations on survival in patients with left- and right-sided Dukes' C colon cancer. *The American journal of gastroenterology.* 2000; 95(10):2953–2957. [PubMed: 11051374]
22. Hecker L, Vittal R, Jones T, et al. NADPH oxidase-4 mediates myofibroblast activation and fibrogenic responses to lung injury. *Nature medicine.* 2009; 15(9):1077–1081.
23. Mochizuki T, Furuta S, Mitsushita J, et al. Inhibition of NADPH oxidase 4 activates apoptosis via the AKT/apoptosis signal-regulating kinase 1 pathway in pancreatic cancer PANC-1 cells. *Oncogene.* 2006; 25(26):3699–3707. [PubMed: 16532036]
24. Maranchie JK, Zhan Y. Nox4 is critical for hypoxia-inducible factor 2-alpha transcriptional activity in von Hippel-Lindau-deficient renal cell carcinoma. *Cancer Res.* 2005; 65(20):9190–9193. [PubMed: 16230378]
25. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell.* 2010; 141(1):52–67. [PubMed: 20371345]
26. Guo RJ, Suh ER, Lynch JP. The role of Cdx proteins in intestinal development and cancer. *Cancer Biol Ther.* 2004; 3(7):593–601. [PubMed: 15136761]
27. Rozek LS, Lipkin SM, Fearon ER, et al. CDX2 polymorphisms, RNA expression, and risk of colorectal cancer. *Cancer Res.* 2005; 65(13):5488–5492. [PubMed: 15994917]
28. Kawai H, Tomii K, Toyooka S, et al. Promoter methylation downregulates CDX2 expression in colorectal carcinomas. *Oncol Rep.* 2005; 13(3):547–551. [PubMed: 15706431]
29. Guo RJ, Funakoshi S, Lee HH, Kong J, Lynch JP. The intestine-specific transcription factor Cdx2 inhibits beta-catenin/TCF transcriptional activity by disrupting the beta-catenin-TCF protein complex. *Carcinogenesis.* 2010; 31(2):159–166. [PubMed: 19734199]
30. Funakoshi S, Kong J, Crissey MA, Dang L, Dang D, Lynch JP. Intestine-specific transcription factor Cdx2 induces E-cadherin function by enhancing the trafficking of E-cadherin to the cell membrane. *Am J Physiol Gastrointest Liver Physiol.* 2010; 299(5):G1054–G1067. [PubMed: 20671195]
31. Camps J, Armengol G, del Rey J, et al. Genome-wide differences between microsatellite stable and unstable colorectal tumors. *Carcinogenesis.* 2006; 27(3):419–428. [PubMed: 16272173]

32. Lea IA, Jackson MA, Li X, Bailey S, Peddada SD, Dunnick JK. Genetic pathways and mutation profiles of human cancers: site- and exposure-specific patterns. *Carcinogenesis*. 2007; 28(9):1851–1858. [PubMed: 17693665]

Figure 1.

Elevated NOX4 expression is a significant predictor of relapse in left-side colon cancer. (A) The density distribution of NOX4 expression in the left-side tumors on GSE14333 with Dukes Stage A, B or C shows a large component with low baseline expression and a tail of elevated expression values. Individual expression values are indicated with has marks at the lower edge. The multistate methodology divides the samples at an expression value of 3.0; the samples with expression below 3.0 are in the NOX4- component, and those with expression values above 3.0 are in NOX4+. (C) The Kaplan-Meier curves for the NOX4- and NOX4+ components plot the expected survival probabilities for the components in the left-side tumors. The 5-year expected survival probability for NOX4- is 0.89 95%CI(0.80 – 0.99) and for NOX4+ it is 0.51 95%CI(0.37 – 0.70). NOX4- contains 53 samples and NOX4+ contains 42 samples. (D) The corresponding Kaplan-Meier plots for the NOX4

components in the right-side tumors shows a distinctly lower connection with relapse than on the left side. The 5-year expected survival for NOX4⁻ on the right side is 0.82 95%CI (0.73 – 0.93) and for NOX4⁺ it is 0.73 95%CI(0.56 – 0.95).

Figure 2.

Low CDX2 expression is a significant predictor of relapse in right-side colon cancer. (A) The density distribution of CDX2 expression in the right-side tumors on GSE14333 with Dukes Stage A, B or C follows a bimodal distribution. The multistate methodology divides the samples at an expression value of 4.76; the samples with expression below 4.76 are in the CDX2- component, and those with expression values above 4.76 are in CDX2+. (C) The Kaplan-Meier curves for the CDX2+ and CDX2- components plot the expected survival probabilities for the components in the right-side tumors. The 5-year expected survival probability for CDX2+ is 0.88 95%CI(0.80 – 0.96) and for CDX2- it is 0.39 95%CI(0.15 – 0.78). (D) The corresponding Kaplan-Meier plots for the CDX2 components in the left-side tumors shows a distinctly lower connection with relapse than on the right

side. The 5-year expected survival for CDX2+ on the left side is 0.75 95CI%(0.66 – 0.86) and for CDX2– it is 0.35 95%CI(0.12 – 1.0).

Figure 3.

NOX4 expression levels and phenotypic affect in colon cancer cell lines. (A) Expression of endogenous NOX4 in human colon cancer cell lines. RTPCR was performed using RNA from several human colon cancer cell lines (DLD-1, SW480, SW620, HT-29 and HCT 116) with specific primers for NOX4. NOX4 mRNA levels were quantified using the comparative C_T method relative to levels of hypoxanthine phosphoribosyltransferase (HPRT1). The fold change in expression (log₁₀) was normalized to normal colon NOX4 mRNA levels. Three experiments were conducted in triplicate. Bars represent the median fold change (log₁₀) and the error bars represent the standard deviation. (B) RTPCR results showing transfection with NOX4 siRNA reduces NOX4 mRNA levels in SW620 cells compared with control AllStar negative siRNA. A positive control AllStar Death siRNA was also used to validate transfection efficiency (data not shown). (C) siRNA suppression of NOX4 does not affect SW620 cell viability. Cell viability was normalized to AllStar negative transfected SW620 cells. NOX4 siRNA and AllStar negative siRNA transfected SW620 cell have similar cell viability. Assays were performed in triplicate. Similar results were obtained in two separate experiments. (D) Targeted NOX4 knockdown decreases superoxide production in SW620 cells. SW620 cells were transfected with NOX4 siRNA and assayed for superoxide production by the chemiluminescent method. Superoxide producing activity of AllStar negative siRNA transfected cells are set as 100%. Each bar represents the mean data from 2 independent transfections, with error bars representing the standard deviation for percentage of activity.

Figure 4.

(A) Expression levels of *PRAC* (230784_at) by tumor location in the GSE14333 microarray data. *PRAC* is expressed at a high level in most left-side tumors and at a baseline level in most right-side tumors. (B) Expression of endogenous *PRAC* in human colon cancer cell lines. RTPCR was performed using RNA from several human colon cancer cell lines (DLD-1, SW837, SW480, SW620, HT-29 and HCT 116) with specific primers for *PRAC*. *PRAC* mRNA levels were quantified using the comparative C_T method relative to levels of hypoxanthine phosphoribosyltransferase (*HPRT1*). The fold change in expression ($\log_{10}$) was normalized to DLD-1 mRNA levels. Two experiments were conducted in triplicate. Bars represent the median fold change ($\log_{10}$) and the error bars represent the standard deviation.

Table 1

Characteristics of the LCC and RCC patients in GSE14333

	No.	gender (M/F)	Dukes stage	relapse (no/yes/NA)	Chemo (no/yes)
left	122	77/45	A	15/1/2	16/2
			B	31/6/0	29/8
			C	22/18/0	10/30
			D	-	-
right	125	59/66	A	16/0/1	17/0
			B	40/4/0	33/11
			C	27/14/0	13/28
			D	-	-

Table 2

Genes and associated pathways most significantly implicated in relapse in left side and right side colon cancer with Dukes stage A, B or C

probe	gene	CPH p-value ^a	direction in relapse cases	pathways affected ^b	multistate marker
left side					
236028_at	IBSP	2.7×10^{-5}	UP	FA	NOX4
210095_s_at	IGFBP3	1.0×10^{-4}	UP	P53	NOX4
213425_at	WNT5A	2.5×10^{-4}	DOWN	WNT	MMP3
223121_s_at	SFRP2	3.1×10^{-4}	UP	WNT	NOX4
229271_x_at	COL11A1	7.1×10^{-4}	UP	FA	NOX4
216442_x_at	FN1	7.3×10^{-4}	UP	FA	NOX4
220088_at	C5AR1	1.4×10^{-3}	UP	CCC	NOX4
201109_s_at	THBS1	1.9×10^{-3}	UP	P53, TGFB, FA	NOX4
202627_s_at	SERPINE1	2.7×10^{-3}	UP	P53, CCC	NOX4
212607_at	AKT3	2.9×10^{-3}	UP	FA, INS	NOX4
221729_at	COL5A2	3.3×10^{-3}	UP	FA	NOX4
203083_at	THBS2	3.6×10^{-3}	UP	TGFB, FA	NOX4
204315_s_at	GTSE1	5.8×10^{-3}	DOWN	P53	MMP3
210511_s_at	INHBA	6.2×10^{-3}	UP	TGFB	NOX4
202310_s_at	COL1A1	6.4×10^{-3}	UP	FA	NOX4
202833_s_at	SERPINA1	6.5×10^{-3}	DOWN	CCC	MMP3
right side					
202267_at	LAMC2	3.7×10^{-7}	UP	FA	CDX2
236313_at	CDKN2B	8.9×10^{-6}	UP	CC, TGFB	CDX2
203725_at	GADD45A	1.7×10^{-5}	UP	CC, P53	CDX2
204420_at	FOSL1	2.0×10^{-5}	UP	WNT	FAM69A
202628_s_at	SERPINE1	8.7×10^{-5}	UP	P53, CCC	CDX2
203323_at	CAV2	1.6×10^{-4}	UP	FA	CDX2

probe	gene	CPH p-value ^a	direction in relapse cases	pathways effected ^b	multistate marker
left side					
201124_at	ITGB5	1.7×10^{-4}	UP	FA	FAM69A
213792_s_at	INSR	1.8×10^{-4}	UP	INS	CDX2
202627_s_at	SERPINE1	1.9×10^{-4}	UP	P53, CCC	CDX2
203726_s_at	LAMA3	2.2×10^{-4}	UP	FA	CDX2
208711_s_at	CCND1	2.4×10^{-4}	UP	CC, P53, WNT, FA	FAM69A
208613_s_at	FLNB	3.3×10^{-4}	UP	FA	FAM69A
201925_s_at	CD55	3.4×10^{-4}	UP	CCC	CDX2
214866_at	PLAUR	3.4×10^{-4}	UP	CCC	FAM69A
204714_s_at	F5	4.7×10^{-4}	UP	CCC	CDX2
204363_at	F3	5.4×10^{-4}	UP	CCC	CDX2

^aThis is the logrank score p-value of a Cox proportional hazard model in which the only variable is the expression vector for the probe.

^bCC=cell cycle, CCC=complement and coagulation cascades, FA=focal adhesion, INS=insulin signaling, P53=p53 signaling, TGFB=TGFβ signaling, WNT=Wnt signaling