Article

Identification of 42 Genes Linked to Stage II Colorectal Cancer Metastatic Relapse

Rabeah A. Al-Temaimi [1], Tuan Zea Tan [2], Makia J. Marafie [3], Jean Paul Thiery [4], Philip Quirke [5] and Fahd Al-Mulla [6],*

[1]   Human Genetics Unit, Department of Pathology, Faculty of Medicine, Kuwait University, Kuwait 13110, Kuwait; rabeah@hsc.edu.kw

[2]   Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore; csittz@nus.edu.sg

[3]   Kuwait Medical Genetics Center, Maternity Hospital, Kuwait 13001, Kuwait; mj_marafie@yahoo.com

[4]   Department of Biochemistry, National University of Singapore, Singapore 117599, Singapore; bchtjp@nus.edu.sg

[5]   Pathology and Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds LS9 7TF, UK; p.quirke@leeds.ac.uk

[6]   Molecular Pathology Unit, Department of Pathology, Faculty of Medicine, Kuwait University, Kuwait 13110, Kuwait; fahd@al-mulla.org

*   Correspondence: fahd@al-mulla.org; Tel.: +965-677-71040

Abstract: Colorectal cancer (CRC) is one of the leading causes of cancer mortality. Metastasis remains the primary cause of CRC death. Predicting the possibility of metastatic relapse in early-stage CRC is of paramount importance to target therapy for patients who really need it and spare those with low-potential of metastasis. Ninety-six stage II CRC cases were stratified using high-resolution array comparative genomic hybridization (aCGH) data based on a predictive survival algorithm and supervised clustering. All genes included within the resultant copy number aberrations were each interrogated independently at mRNA level using CRC expression datasets available from public repositories, which included 1820 colon cancers, and 167 normal colon tissues. Reduced mRNA expression driven by copy number losses and increased expression driven by copy number gains revealed 42 altered transcripts (29 reduced and 13 increased transcripts) associated with metastatic relapse, short disease-free or overall survival, and/or epithelial to mesenchymal transition (EMT). Resultant genes were classified based on gene ontology (GO), which identified four functional enrichment groups involved in growth regulation, genomic integrity, metabolism, and signal transduction pathways. The identified 42 genes may be useful for predicting metastatic relapse in stage II CRC. Further studies are necessary to validate these findings.

Keywords: colorectal cancer; microarray; stage II; copy number aberrations; disease free survival; gene expression; metastasis

---

1. Introduction

Microarray comparative genomic hybridization (aCGH) has been extensively used to profile colorectal cancer (CRC) for copy number aberrations. However, their direct relevance to prognosis and therapeutic interventions has remained elusive [1,2]. Undoubtedly, early cancer detection contributes to a better patient outcome. Fecal occult blood testing, fecal immunochemical/molecular testing, and colonoscopy-based screening have become wide-spread in developed countries as a preventive measure for the detection of CRC in its early stages. However, 10%–30% of early-stage CRC still relapse with metastases within two years of surgical treatment [3,4]. To address this critical issue, various

technologies such as aCGH, metaphase, and next-generation sequencing were employed in attempts targeted towards the identification of suitable prognostic biomarkers in early stage CRC. Array-based analysis of DNA copy number aberrations (CNAs), in particular, has proven to be very effective in identifying recurrent CNAs in CRC. It is usually argued that CNAs of recurrent nature may imply their evolutionary importance in driving CRC progression [5,6]. For this reason, large bodies of publications have attempted to ascertain early stage CRC patients at risk of metastatic relapse based on the genomic profiling of primary tumors. However, such attempts rarely identified reliable and replicable CNAs or genes predictive of survival. Several predictive gene signatures have been reported for CRC [7,8]. Yet, these studies suffer limitations such as small sample size and lack of adequate expression-based validation. This is not surprising given that most CNAs encompass large genomic aberrations that usually harbor hundreds of bystander genes among which only few genes may be potentially useful for prognostication [8]. Therefore, validation studies linking genomic aberrations at the DNA level to transcription have been limited [7]. The two questions that need to be addressed are; first, how can we identify key "driver" CNA events from the enormous amount of random "passenger" CNAs that have no functional significance? Second, how can we pinpoint the few metastasis suppressor genes residing within large areas of genomic deletions? Here, we attempted to address these questions. Firstly, we identified CNAs exclusively found in stage II metastatic CRC using stringent experimental and statistical algorithms. We then grouped the CNAs generated based on their significant effects on disease survival. The resultant list of potential metastasis-associated genes (958 genes) residing within CNA regions were each independently validated for aberrant expression in publicly available CRC cohorts' expression data. This validation identified 42 genes, potentially involved in metastasis. Twenty-nine metastasis suppressor genes whose copy number deletions resulted in diminished transcription, and reduced overall survival (OS) or disease-free survival (DFS), and 13 potentially metastasis-enhancing genes whose copy number gains and overexpression associated with metastatic relapse, short disease-free or overall survival or/and epithelial to mesenchymal transition (EMT). A large and randomized study focused on the 42 genes to confirm our data and validate their clinical utility is now warranted.

2. Results

2.1. CRC aCGH Profile

Table 1 illustrates the clinicopathological characteristics of 96 stage II CRC cases investigated in our study. The data show that pathological classification alone has limited efficacy in distinguishing subclasses of stage II CRC in terms of aggressiveness or prognosis.

Table 1. Clinicopathological characteristics of stage II CRC cohort.

Patients' Characteristics		Stage II Who Stayed Disease Free	Stage II with Local Recurrences	Stage II Who Relapsed with Distant Metastasis	*p* Value [a]
Mean age in years		64.4	75.5	75.4	0.004 [b]
Sex	Male	41	2	5	0.43
	Female	37	5	6	
Total		78	7	11	
Site	Right	19	2	1	0.16
	Left	28	1	9	
	Rectum	16	1	1	
	Unknown	15	3	0	
T-stage	T3	44	3	9	0.27
	T4	19	4	2	
	Unknown	15	0	0	
Differentiation	Well	10	1	1	0.8
	Moderate	56	5	9	
	Poor	5	1	0	
	Unknown	7	0	1	

Table 1. *Cont.*

Patients' Characteristics		Stage II Who Stayed Disease Free	Stage II with Local Recurrences	Stage II Who Relapsed with Distant Metastasis	*p* Value [a]
MMR status	MSI	14	1	0	0.45
	MSS	59	5	9	
	Unknown	5	1	2	
Follow-up	Mean DFS	9.5 years	3.9 years	3.08 years	

[a] Fisher's exact test; [b] one-way ANOVA. DFS is disease free survival; MSI is microsatellite unstable; MSS is microsatellite stable; MMR is mismatch repair.

To refine our search for genes involved in metastasis, we excluded seven cases that later presented with local recurrence. Significance testing for aberrant copy number (STAC) analysis focused on chromosomal CNAs of 89 CRC cases, which produced acceptable derivative of log ratio spread (DLRS) values below 0.5. DLRS is defined as the spread of the log ratio differences between consecutive probes along all chromosomes. Comparison using stringent criteria between CNAs found in 11 stage II CRC with metastatic disease and CNAs found in 78 stage II CRC cases that remained disease-free resulted in the identification of chromosomal CNAs significantly enriched in metastatic CRC (Figure 1a). The most frequent copy number gains (CNGs) were in chromosomes seven (56%), 8q (56%), 13q11-q34 (61%), and 20q11.1-q13.33 (79%). Other less frequent regional gains include; chromosomes 1q, 9p, and 17q. The most frequent copy number losses (CNLs) were in chromosome arms 1p (71%), 8p (72%), 17p (55%), 22q (60%), and chromosomes 14 (77%), 15 (66%), and 18 (80%). To narrow down genes with predictive potential, a second analytical approach that is dependent solely on survival data and not whether the case relapsed with metastases or not was utilized to point out CNAs shared by CRC cases with reduced DFS (Figure 1b). The two approaches were compared to compile a common CNA profile that may influence metastasis and DFS (Figure 1c). The most significant chromosomal gains that are consistently associated with metastatic CRC recurrence and poor survival were gains in chromosomes 1q, 9p, 13q11-q34, 17q, and 20q11.1-q13.33; and losses in chromosomes 5q12.1-q35.3, 8p12-p23.3, 9q33.1-q34.3, 11q23.3-q25, 14q11.2-q32.31, 15q21.1-q26.3, 18p11.31-q21.1, 20p12.1-p13, and 22q11.21-q13.31. There were 1099 genes located within significant CNAs that were further scrutinized for their existence in normal healthy individuals [9,10]. CNAs frequent in healthy normal genomes that overlapped 5%–100% with mapped CRC metastatic *vs.* non-metastatic chromosomal gains were excluded from further analysis. Interestingly, all CNAs, except for 8p chromosomal region harboring SPAG11A, had minimal overlap with common CNAs in normal healthy individuals. This elimination step resulted in 958 candidate genes for further investigation (Table S1). While some of the CNAs housed one to a few genes, larger CNAs contained many genes that may modify cancer progression or are more likely "innocent passenger" CNAs that are of no functional significance. In addition, CNA profiles segregated according to our cohort's clinicopathological characteristics also resulted in significant clustering into identifiable groups. Distinct CNAs were found between microsatellite stable (MSS) and microsatellite instable (MSI) CRC, and other clinicopathological characteristics (manuscripts submitted elsewhere). However, we chose to focus here on metastasis associated CNAs. Consequently, the expression level of each of the 958 genes resultant from metastatic CRC CNAs profile was interrogated independently using publicly available CRC datasets to validate their clinical relevance based on; their differential expression in normal and CRC samples, association with grade, microsatellite instability, stage, DFS, overall survival (OS), and EMT score.
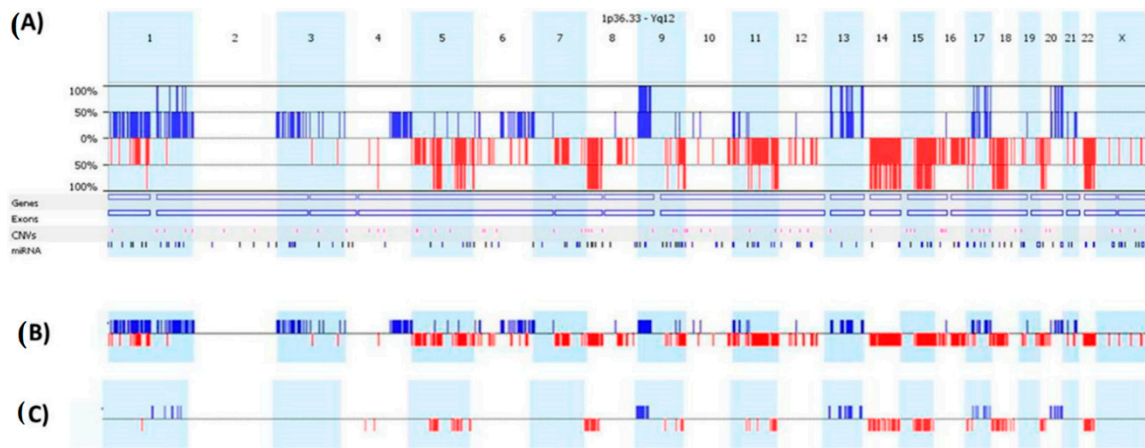
Figure 1. Copy number aberrations in CRC stage II using the first analytical approach STAC (**A**), which compares the chromosomal copy numbers found in metastatic CRC with non-metastatic CRC; and the second analytical approach; the survival predictive algorithm (**B**), which identifies chromosomal copy number gains associated with reduced disease-free survival; (**C**) shows the frequencies of the aberrations in the two approaches and identifies CNA common to both methods (100%). Each column represents a chromosome with blue bars indicating copy number gains and red bars copy number losses.

2.2. Expression Analysis of Candidate Oncogenes

Approximately 307 genes were not represented on the Affymetrix U133Plus2 platform (Santa Clara, CA, USA), of which 166 genes had probes on U133A platform. Genes without a probe on either platform were mostly miRNA genes and were excluded. At first, two categorical variables were considered, DFS and OS to select CNAs' genes of altered expression. This was done to isolate genes of strong prognostic value. Interrogating gene expression changes driven by CNAs revealed 42 altered transcripts (29 reduced and 13 increased) associated with metastatic relapse, short DFS, or OS (Table 2).

Table 2. Cox regression for survival analysis of the 42 genes' mRNA expression in meta-cohort (Affymetrix U133A, *n* = 1820; or U133Plus2, *n* = 1436) for overall survival (OS), and disease-free survival (DFS). A negative regression coefficient implies a better prognosis for patients if they retain the function of a given gene (worse prognosis when the cancer underexpresses the gene). Conversely, a positive regression coefficient means that the hazard is higher for a given gene's overexpression and, thus, the prognosis worse.

Gene	Gene ID	aCGH CNA Event	%CNA Overlap with Normal	Cox's OS (*p*-Value)	Cox's DFS (*p*-Value)
ADRA1A	148	Loss	0.42	−0.40185, (0.152322)	−1.67582, (0.001948)
ADRA1D	146	Loss	1.15	−0.5968, (0.020507)	0.066504, (0.880385)
ADRB3	155	Loss	0	−1.21809, (0.002366)	−1.12818, (0.08601)
APOBEC3D	140564	Loss	0	−0.45803, (0.24853)	−1.56085, (0.0224)
BRF2	55290	Loss	0	−0.45526, (0.02013)	0.061887, (0.8515)
C20orf202	400831	Loss	0.89	−0.80141, (0.000481)	−0.19907, (0.628292)
CABIN1	23523	Loss	4.02	−0.47782, (0.0210)	−0.51601, (0.14935)
CACNA1I	8911	Loss	0	−0.69524, (0.00404)	−0.73386, (0.082677)
CSMD1	64478	Loss	0	−0.01578, (0.956357)	−1.30789, (0.01948)
DIO3	1735	Loss	1.76	−0.32561, (0.041777)	−0.00829, (0.973719)
EPHX2	2053	Loss	0.42	0.036582, (0.61411)	−0.30515, (0.0099)
FAM83F	113828	Loss	0	−0.13746, (0.287004)	−0.72925, (0.00085)
GP1BB	2812	Loss	0	−0.56827, (0.000246)	−0.19289, (0.495577)
KIAA1656	85371	Loss	0	−0.60241, (0.01922)	−1.02118, (0.032326)
LOC339593	339593	Loss	0.17	0.29021, (0.385267)	−1.12678, (0.047789)
MCM8	84515	Loss	0	0.043419, (0.614947)	−0.31948, (0.034203)

Table 2. *Cont.*

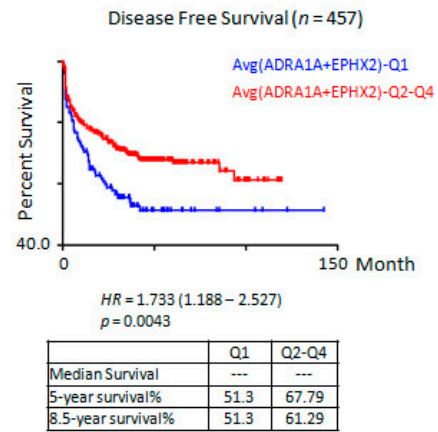| Gene	Gene ID	aCGH CNA Event	%CNA Overlap with Normal	Cox's OS (*p*-Value)	Cox's DFS (*p*-Value)
NAT1	9	Loss	0	0.028159, (0.713606)	−0.36337, (0.007197)
NAT2	10	Loss	0	−0.09059, (0.160865)	−0.30169, (0.003804)
HNF6	3175	Loss	0	−0.66924, (0.017955)	−0.51152, (0.31)
PCDHGA11	56105	Loss	0	−0.85799, (0.001447)	−0.577, (0.33928)
RAB11FIP1	80223	Loss	0	−0.38036, (5.55×10^{-05})	−0.02196, (0.87)
SPAG11A	653423	Loss	68.65	−0.81665, (0.001447)	−0.13298, (0.57)
SIRPD	128646	Loss	0.89	−0.89166, (0.014166)	−0.13392, (0.82)
TEX43	389320	Loss	0	−0.46611, (0.175471)	−0.19907, (0.01162)
TOP1P2	7152	Loss	0	−1.13965, (0.005692)	0.599602, (0.32)
WDR5	11091	Loss	0	−0.0007, (0.995962)	−0.52505, (0.024307)
ZNF366	167465	Loss	0.07	−0.76274, (0.01779)	−0.93429, (0.09208)
ZNF703	80139	Loss	0	−0.32601, (0.000744)	−0.14954, (0.366064)
ZNRF3	84133	Loss	0	−0.03445, (0.605797)	−0.33884, (0.002065)
ANXA2P2	304	Gain	0	0.478663, (0.003259)	0.997702, (0.000504)
CDC42BPA	8476	Gain	0	0.032966, (0.690294)	0.401718, (0.005473)
DOK5	55816	Gain	0.57	0.105512, (0.1)	0.577502, (0.002218)
DUSP14	11072	Gain	0.70	0.207902, (0.047604)	0.709569, (0.000159)
GLIS3	169792	Gain	1.25	0.0916, (0.210004)	0.397791, (0.000566)
ING1	3621	Gain	0	0.461934, (0.020587)	0.127394, (0.710857)
MPDZ	8777	Gain	0	0.079568, (0.406857)	0.657405, (3.98×10^{-5})
PITPNC1	26207	Gain	0	0.257758, (0.029048)	0.418362, (0.026674)
SCEL	8796	Gain	0	0.081094, (0.202236)	0.301113, (0.000357)
SEMG1	6406	Gain	0	0.098826, (0.043099)	0.153704, (0.061048)
SMU1	55234	Gain	0.73	0.36857, (0.008896)	0.016381, (0.942047)
USP32	84669	Gain	0	0.162429, (0.221638)	0.586012, (0.01236)
VLDLR	7436	Gain	0.14	−0.00629, (0.92868)	0.323607, (0.004972)

Our findings show that several candidate oncogenes of significant prognostic potential reside within regions of chromosomal gains and are overexpressed, most likely, under the influence of gene amplification within these regions. Moreover, our results pinpointed metastasis-suppressor genes of significant prognostic potential residing within chromosomal loss areas, and are underexpressed due to genetic deletions. To further refine our list for genes with the greatest prognostic power, we selected genes significantly associated with DFS or OS using a stringent criterion (*p*-value < 0.005), and compared their expression levels in tumor against normal colon tissues. A total of 19 genes had highly significant associations with DFS or OS, of which only 14 had significantly differential expression in normal colon *vs.* CRC tissues (Figure S1). Therefore, these genes have the greatest discriminatory application in clinical prognostication of stage II CRCs at risk of metastatic relapse. We next focused on evaluating the combined potential of a set of deleted genes in relation to their combined effects on DFS or OS. Our hypothesis was that double or more deletions of candidate tumor suppressor genes will significantly increase their effect on DFS and OS in CRC patients. Combination analysis results show that while single deleted gene reduced expression associated with a significant reduction in DFS or OS, their combined underexpression analysis in CRC strengthened their prognostic potential in association with reduced DFS/OS (Figure 2).

Lastly, we analyzed each of the 42 genes' expression levels in relation to the clinicopathological characteristics of analyzed CRC samples from the GEO database, which included; tumor grade, cancer stage (stages I–IV), expression in MSS *vs.* MSI tumors, expression levels in tumors *vs.* normal colon tissues, and their relationship to EMT in cancers. GEO database clinicopathological correlation data for all of the 29 potentially metastasis-suppressor genes and the 13 potentially metastasis-enhancer genes are shown in (Tables S2–S9).
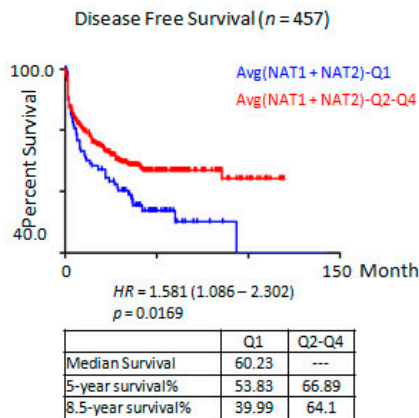
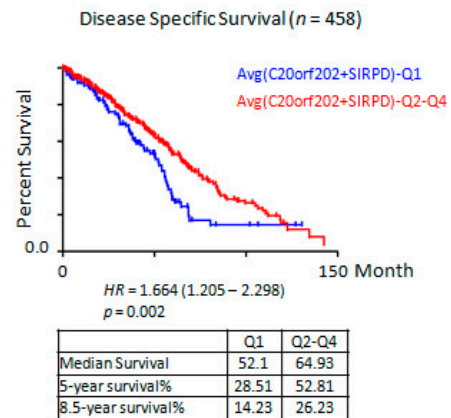**Combo 1: ZNF703 + ADRB3 + RAB11FIP1 + BRF2**

Disease Specific Survival (*n* = 458)

Avg(ZNF703+ADRB3+RAB11FIP1+BRF2)-Q1
Avg(ZNF703+ADRB3+RAB11FIP1+BRF2)-Q2-Q4

HR = 1.707 (1.242 − 2.347)
p = 0.001

	Q1	Q2-Q4
Median Survival	49.24	64.1
5-year survival%	34.18	51.55
8.5-year survival%	12.65	26.71

Combo 2: ADRA1A + EPHX2

Disease Free Survival (*n* = 457)

Avg(ADRA1A+EPHX2)-Q1
Avg(ADRA1A+EPHX2)-Q2-Q4

HR = 1.733 (1.188 − 2.527)
p = 0.0043

	Q1	Q2-Q4
Median Survival	---	---
5-year survival%	51.3	67.79
8.5-year survival%	51.3	61.29

Combo 3: NAT1 + NAT2

Disease Free Survival (*n* = 457)

Avg(NAT1 + NAT2)-Q1
Avg(NAT1 + NAT2)-Q2-Q4

HR = 1.581 (1.086 − 2.302)
p = 0.0169

	Q1	Q2-Q4
Median Survival	60.23	---
5-year survival%	53.83	66.89
8.5-year survival%	39.99	64.1

Combo 4: C20orf202 + SIRPD

Disease Specific Survival (*n* = 458)

Avg(C20orf202+SIRPD)-Q1
Avg(C20orf202+SIRPD)-Q2-Q4

HR = 1.664 (1.205 − 2.298)
p = 0.002

	Q1	Q2-Q4
Median Survival	52.1	64.93
5-year survival%	28.51	52.81
8.5-year survival%	14.23	26.23

Combo 5: CACNA1I + FAM83F

Disease Free Survival (*n* = 326)

Avg(CACNA1I + FAM83F)-Q1
Avg(CACNA1I + FAM83F)-Q2-Q4

HR = 2.077 (1.176 − 3.67)
p = 0.0118

	Q1	Q2-Q4
Median Survival	93.66	---
5-year survival%	65.92	78.8
8.5-year survival%	43.95	77.44

Figure 2. Five different combinations (Combo 1-5) of candidate tumor suppressor genes consistently consolidated their association with disease specific survival (OS), and disease free survival (DFS) in CRC samples' expression data. Q1–Q4 signify expression quartiles with Q1 being the lowest and Q4 the highest expression quartile.

3. Discussion

CRC represents a heterogeneous group of gastrointestinal neoplasia. Stage II CRC is the most clinically challenging stage as it is the critical point between remission and progression into aggressive stages. Stage II CRC has 20%–30% risk of metastasis, whereas stage III CRC has a 50%–80% risk of

distant metastasis [11]. The genetic complexity of stage II CRC is reflected by its extensive CNA profile that presents an additional limitation of sifting the "culprit" genes from the "innocent passenger" ones. The ongoing theory is that cancer metastases arise due to the acquisition of the metastatic phenotype, which appears to have an underlying metastatic genotype [12,13]. We generated genomic CNA profiles for stage II CRCs, with and without subsequent metastatic disease to identify CNAs associated with metastasis or reduced OS/DFS in stage II CRC. We then utilized publicly-available high-density oligonucleotide-based microarray expression data to study the effect of these CNAs. The use of well-defined expression profiles from sets of colorectal cancers may serve as reliable and focused validation cohorts. Our resultant genes were individually scrutinized for association with cancer biology. In addition, a stringent gene selection method was employed to exclude false positive CNAs that occur in normal colon samples, and exclude genes whose expression levels in CRC were comparable to normal colon tissues. All 42 genes' expression had significant association in predicting OS and/or DFS (Table 2). Functional annotation of each gene confirms that each of these genes partake in key processes of cancer progression and metastasis, providing insights into the genetic mechanisms driving metastatic CRC. The 42 genes were divided into genes deleted (29 genes) and genes amplified (13 genes) of which 26 deleted genes and 13 amplified genes had specific known functions. Each gene was literature-researched for reported evidence of known biological function(s) and involvement in cancer.

In the deleted genes list several genes had known tumor suppressor and anti-metastatic properties (Table S10). *CSMD1*, *FAM83F*, and *EPHX2* are known to be silenced, mutated, or deleted in gastrointestinal cancers though their precise functions remain unknown [14,15]. ADRA1A, APOBEC3D, CABIN1, DIO3, GP1BB, HNF6, MCM8, PCDHGA11, WDR5, ZNRF3, and ZNF366 all have functional attributes that are known to control and affect the transition into metastatic pathways. However, *CABIN1* did not significantly associate with our EMT potential analysis ($p = 0.8$), and associated with distinguishing MSS CRC ($p < 0.05$, Tables S4 and S5). *CABIN1* is a pro-apoptotic gene, which suggests that its reduced expression is an early marker for CRC's proliferation and genomic instability [16]. EMT fundamentally involves genomic and phenotypic changes of epithelial cells into mesenchymal cells, loss of epithelial cell-cell contact, loss of cell polarity, and acquisition of migratory characteristics facilitating metastasis [17]. *ADRA1A* and *ZNF366* have proliferation regulatory functions that limit tumor cells from forming tumors/metastases in response to stimulators [18,19]. Genomic instability and loss of epithelial genomic signatures are also a requirement for EMT. Three deleted signature genes; *APOBEC3D*, *MCM8* and *WDR5* have genomic stability, genomic fidelity, and epigenetic fidelity functions, respectively [20–22]. Downregulation of cell-cell contact to facilitate EMT and tumor mass blood transport is achieved by deletion of epithelial cell adhesion regulatory genes that include *GP1BB* and *PCDHGA11* in our resultant deleted genes list [23,24]. *DIO3* is a metabolic suppressor, which has cancer growth rate-limiting properties and may mitigate EMT [25]. Lastly, *ZNRF3* and *HNF6* are EMT and cell-differentiation regulators under normal conditions and are tumor suppressor genes inhibiting the EMT pathway [26,27]. Four deleted genes had contradictory functions to EMT inhibition but had a negative EMT score based on our analysis. These genes associated with other characteristics of CRC based on their expression analysis. *ADRA1D* was differentially expressed in stage IV compared to stages II and III. *NAT1* and *NAT2*, drug metabolism genes; whose expression was associated with stage I MSS CRC. *BRF2*, a pro-proliferation gene, had a positive EMT score and its expression was associated with MSS CRC. Activation of EMT was shown to reduce cell proliferation by targeting specific pro-proliferation genes [28,29]. It is possible that the loss of BRF2 functions in the primary tumor would result in the activation of alternative proliferation pathways in distant metastases during reversion of EMT. The remaining nine deleted genes were of unknown functions, albeit associating with several characteristics of CRC (Tables S2–S5), which warrants further functional studies of these genes.

All amplified genes had known functions, except for *ANXA2P2*, which is a pseudogene for the mesenchymal marker Annexin A2. Most amplified genes had positive association with EMT except

for *SEMG1*, *SMU1* and *ING1*. *SMU1* and *ING1* regulate genomic integrity and cell growth and their function could be altered by aberrant expression or gene structural changes that resulted in their association with MSS CRC [30,31]. Similarly, *SEMG1* expression, a negative regulator of calcium import, is associated with MSS CRC. Metabolic advantage is a feature of metastatic tumors and might be mediated by overexpression of *VLDLR*, *USP32*, and *PITPINC1* [32–34]. Although *PITPINC1* had a non-significant EMT score, its expression was associated with MSS CRC. Five amplified genes were involved in multiple signaling cascades that can have pleiotropic effects. Those genes were *DOK5*, *DUSP14*, *GLIS3*, *PITPNC1*, and *SEMG1*. These genes are partly characterized for their involvement in cancer and would require further studies. Two amplified genes had contradictory functions to being overexpressed in metastatic cancer, *SCEL* (a keratinocyte differentiation factor), and *MPDZ* (a cell adhesion factor). However, both genes are involved in recruitment and assembly of proteins to cellular surfaces promoting protein-protein interactions. Aberrant expression of these proteins may be involved in the formation of invadopodia, which facilitates invasion of the local extracellular matrix and basement membranes [35,36]. Lastly, it should be noted that our study here is limited to estimating copy number aberrations; it does not discern sequence mutations or epigenetic changes.

In conclusion, our study characterized the personal nature of stage II CRC CNAs highlighting key genes involved in promoting CRC metastasis. Resultant genes were mostly involved in cell growth regulation, metabolic regulation, and signaling pathways (Figure 3). We have identified some novel candidate genes of possible prognostic value for both overall survival and DFS. As well as novel genes that can be used for CRC classification. These genes can offer new targets for diagnostic and therapeutic designs. We anticipate validating our findings in separate retrospective and prospective studies for their efficiency in predicting metastatic stage II CRC.
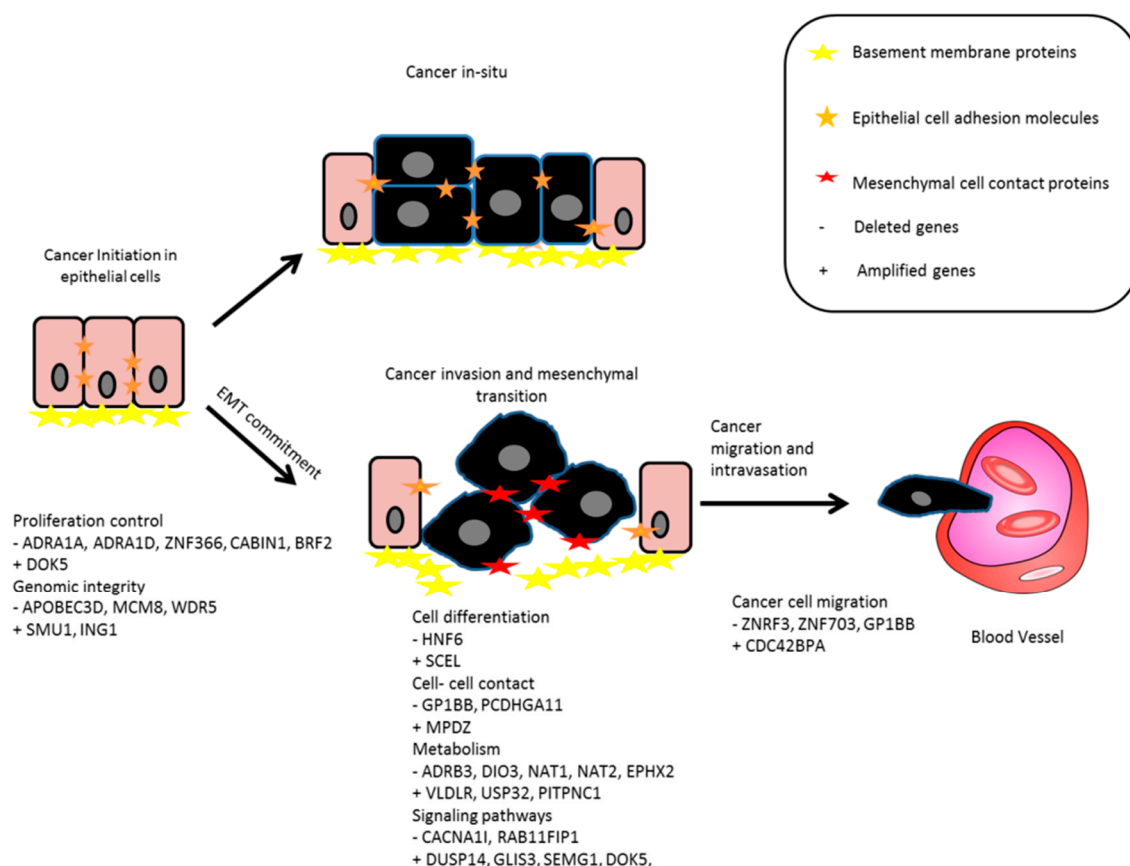

Figure 3. Proposed simple diagram depicting the functional involvement of signature genes in cancer progression, EMT, invasion, and metastasis. The arrows indicate the transition and progression of tumor cells from one stage into the next.

4. Materials and Methods

4.1. CRC Samples

DNA extracted from formalin-fixed paraffin-embedded (FFPE) tissues from 96 patients with sporadic early stage II CRC were used for genomic profiling. Genomic DNA was isolated from microdissected FFPE CRC tissues as described previously [37].

4.2. Microsatellite Instability Analysis

Microsatellite fragment analysis was performed on FFPE-extracted DNA using MSI Analysis System Version 1.2 kit (Promega, Madison, WI, USA). Spectral calibration was carried out on the Applied Biosystems 3130 genetic analyzer using the Powerplex Matrix Standards 3100/3130 kit (Promega). Promega's MSI Analysis System includes fluorescently labeled primers for amplification of seven markers; five mononucleotide repeat markers (Bat-25, BAT-26, NR-21, NR-24, and MONO-27), and two pentanucleotide repeat markers (Penta C and D). Mononucleotide markers are used to determine the MSI status, whereas pentanucleotide markers are used to detect potential sample mix-up by validating that tumor and matching normal samples are from the same individual. DNA concentrations of 10–20 ng from normal and tumor samples were used for the fluorescent PCR-based assay. Microsatellite instability was determined by comparing allelic profiles of microsatellite markers generated by amplification of normal and tumor DNA. Internal lane size standard ILS600 was added to amplified samples to ensure accurate sizing of alleles. A loading cocktail was prepared by mixing the ILS600-PCR product with highly-deionized formamide, and denatured prior to loading onto the 3130 Genetic Analyzer for capillary electrophoresis. The sample's fragment separation output data were analyzed using GeneMapper software version 4.0 (Applied Biosystems, Foster City, CA, USA). CRC samples were classified as MSS if no marker showed any length variation compared with its matching normal colon mucosa. When two or more of the markers showed length mutation in CRC compared with its matching normal mucosa, the CRC sample was labeled as MSI-high.

4.3. Genomic Landscaping Using aCGH

Array CGH was carried out on 96 CRC samples following our standard published protocol [38]. In summary, 2 µg of tumor DNA and sex-matched pooled reference DNA (Promega, Madison, WI, USA) were sonicated in a water bath. The universal linkage system (ULS) Cy3 and Cy5 (Agilent Technologies, Santa Clara, CA, USA) dyes were used to label DNA according to the manufacturer's protocol. Differentially-labeled DNA was purified by Agilent KREApure columns (Agilent Technologies, Santa Clara, CA, USA). Purified labeled tumor and reference samples were hybridized onto Human Genome CGH arrays 244A slides (Agilent Technologies, Santa Clara, CA, USA) in SureHyb chambers (Agilent Technologies, Santa Clara, CA, USA) for 40 h at 60 °C. Washing and scanning were carried out according to manufacturer's protocol. Slides were scanned immediately on an Agilent microarray scanner at 5 µm resolution to minimize the impact of environmental factors on signal intensities. Data was extracted from microarray image files using Feature Extraction software (Version 9.5, Agilent Technologies, Santa Clara, CA, USA) and analyzed using Nexus Copy Number™ software (BioDiscovery, El Segundo, CA, USA). Quality values generated from data analysis ranged between 0.05–0.4, and to minimize false positive calls and random copy number variations BioDiscovery's Fast Adaptive State Segmentation Technique (FASST2) algorithm with a stringent significance threshold of 5.0×10^{-6} was used to determine copy number states.

4.4. Data Clustering and Statistical Analysis

Traditional means of classifying the importance of cancer-related copy number gain or loss include the frequency of their occurrence in different patients. However, cancer genomes are highly complex and frequently harbor random "passenger" copy number aberrations of no functional significance.

To minimize interference from these passenger events, a systematic statistical approach termed significance testing for aberrant copy number (STAC) was employed [39]. The STAC-based algorithm is a robust method, which identifies a set of aberrations that are stacked on top of each other from different samples such that it would not occur by chance. To find these events, aberrations (usually narrow regions) were permutated in each arm of each chromosome and the likelihood for an event occurring at any location at a particular frequency was calibrated. Here, we used STAC to compare CNAs from 11 stage II primary CRC cases with confirmed distant metastases based on a 10-year follow up, and 78 stage II primary CRC relapse-free cases (Table 1). We defined metastatic CRC as distant cancer recurrence away from the primary site. Disease/relapse-free status is defined as no distant metastases (liver, brain, or bone) or local recurrence. We set the likelihood frequency stringently at 35% and a significance *p*-value of <0.05 [39]. To further refine the CNA to include clinically-relevant aberrations, we reanalyzed the data using a survival predictive power statistical approach (SPPS). SPPS correlates CNA with survival time. In this analysis samples with similar CNA regions are grouped (In-Group) and their mean survival calculated and compared to samples without the named CNA (Out-Group). The mean survival times were then compared using the log rank test. This approach highlights CNAs that influence survival time and may be more clinically relevant. Such permutations are not considered in the STAC analysis we used; therefore, the two methods may be considered complementary. Results from both approaches were compared to narrow down the chromosomal regions and resident genes associated with CRC metastasis and disease-free status.

4.5. Data Preprocessing of Affymetrix Microarray Gene Expression

Gene expression microarray data of colon cancer on U133A or U133Plus2 platform (Affymetrix, Santa Clara, CA, USA) were downloaded from Gene Omnibus (GEO), including synchronous and metachronous liver metastases from CRC (GSE10961, *n* = 18), primary colorectal tumors (GSE13067, *n* = 74), primary CRCs (GSE13294, *n* = 155), primary CRCs (GSE14333, *n* = 290), colon adenomas and CRCs (GSE15960, *n* = 12, normal = 6), CRCs (GSE17536, *n* = 177 of which 144 are stage II and III), metastatic CRCs (GSE17537, *n* = 55), stage II CRCs (GSE18088, *n* = 53), stage II and III CRCs (GSE18105, *n* = 77, normal = 34), colon adenomas and CRCs (GSE20916, *n* = 101, normal = 44), CRCs (GSE23878, *n* = 35, normal = 24), MSI CRCs (GSE24514, *n* = 34, normal = 15), MSI CRCs (GSE26682, *n* = 331), stage II and III CRCs (GSE31595, *n* = 37), primary stage II CRCs (GSE33113, *n* = 90), primary CRC tumors (GSE35896, *n* = 62), serrated and conventional colorectal adenocarcinoma tumors (GSE4045, *n* = 37), metastatic CRCs (GSE5851, *n* = 80), colorectal adenomas (GSE8671, *n* = 32, normal = 32), and early stage CRC tumors (GSE9348, *n* = 70, normal = 12). Robust Multichip Average normalization was performed on each dataset using R version 2.15.3, Bioconductor Affy package version 1.38.1 (Affymetrix, Santa Clara, CA, USA). The normalized data was compiled and subsequently standardized using ComBat to remove batch effects [40]. The standardized data yielded a meta-cohort of 1820 colon carcinoma, and 167 normal colon tissues. Note that some of the genes are only available on Affymetrix U133Plus2 platform (*n* = 1436), a subset of the meta-cohort. To focus on the effect of tumor suppressor genes, we co-analyzed different combinations of copy number deleted genes in relation to their effect on OS and DFS. Deleted gene combinations depended on their co-presence of their probes in any of the Affymetrix platforms used, and their co-underexpression in a given CRC sample. We stratified expression levels into quartiles (Q) where the first quartile (Q1) is the expression level at the 25th percentile; second quartile is the median or the 50th percentile; the third quartile is the 75th percentile; and, the last quartile (Q4) is the expression levels of the highest 25 percentile.

4.6. Estimation of Epithelial-Mesenchymal Transition Score

Derivation of EMT signature and estimation of EMT score were described previously [41]. Briefly, a curated EMT gene set [42], after removing basal keratins genes [43,44], was used as an initial selection of epithelial and mesenchymal cell lines. The first step was to establish an EMT signature using binary regression (BinReg) 2.0 [45] comparing profiles of colon cell lines with low or

high EMT enrichment score. Briefly, BinReg uses a Bayesian statistical analysis to fit a binary probit regression model on training data given a set of genes that are most correlated with the EMT phenotype. The regression coefficients of these genes indicate the discriminating power of the gene and are weights for the overall meta-gene profile. The overall meta-gene profile is subsequently used for comparison and predicts the status of the phenotype of the new sample or dataset. In the second step, the BinReg colon cancer EMT signature was applied to predict the EMT status of colon cancer tumor or cell lines. In the third step the extreme 25% samples with the highest probabilities for epithelial or mesenchymal phenotype were used to obtain the epithelial or mesenchymal specific gene list for the colon cancer tumors or cell lines using Significance Analysis of Microarray (SAM) *q*-value = 0 and receiver-operating characteristics curve (ROC) value of 0.85. SAM computed the expression fold change of the genes based on the EMT phenotype, and assessed the significance by comparing the observed fold change against the background expression fold change estimated from 1000 permutations of the samples' EMT phenotypes. On the other hand, ROC method constructs a ROC curve assessing the predictive power of a gene with respect to the EMT phenotypes. In the fourth step, single-sample GSEA (ssGSEA) was employed to compute the enrichment score of a tumor or cell line based on the expression of the colon cancer tumor- or cell line-specific epithelial or mesenchymal signature genes [46]. In ssGSEA, a Kolmogorov-Smirnov-based method, the enrichment of a signature is estimated by comparing the empirical cumulative distribution function of genes defined by the signature, *vs.* the genes not in the signature. EMT score is defined as the normalized subtraction of the mesenchymal from epithelial enrichment scores. The EMT score is an estimate for the cell line status as epithelial or mesenchymal phenotype with −1.0 indicates fully epithelial and +1.0 fully mesenchymal.

4.7. Statistical Analysis

Statistical significance evaluation by Mann-Whitney and Spearman correlation tests were computed using Matlab® R2012a (MathWorks, Natick, MA, USA). Dot plot and Kaplan-Meier analysis were done using GraphPad Prism version 5.04 (GraphPad, La Jolla, CA, USA) or SPSS V6.5 (IBM, New York City, NY, USA).

Author Contributions: Fahd Al-Mulla and Jean Paul Thiery conceived and designed the experiments; Fahd Al-Mulla and Rabeah A. Al-Temaimi performed the experiments; Philip Quirke and Tuan Zea Tan analyzed the data; Makia J. Marafie contributed reagents/materials/analysis tools; Rabeah A. Al-Temaimi wrote the paper and gene literature reviews.

Abbreviations

aCGH	Array comparative genomic hybridization
CRC	Colorectal Cancer
CAN	Copy number aberration
OS	Overall survival
DFS	Disease free survival
EMT	Epithelial mesenchymal transition
MSI	Microsatellite instable
MSS	Microsatellite stable
MMR	Mismatch repair

STAC	Significance testing for aberrant copy number
DLRS	Derivative of log ratio spread
CNG	Copy number gain
CNL	Copy number loss
FFPE	Formalin-fixed paraffin embedded
FASST2	Fast Adaptive State Segmentation Technique
SPPS	survival predictive power statistical
BinReg	Binary regression
SAM	Significance analysis of microarray
ROC	Receiver-operating characteristics curve

References

1. Dotan, E.; Cohen, S.J. Challenges in the management of stage II colon cancer. *Semin. Oncol.* **2011**, *38*, 511–520. [CrossRef] [PubMed]

2. Ogino, S.; Chan, A.T.; Fuchs, C.S.; Giovannucci, E. Molecular pathological epidemiology of colorectal neoplasia: An emerging transdisciplinary and interdisciplinary field. *Gut* **2011**, *60*, 397–411. [CrossRef] [PubMed]

3. Kahlenberg, M.S.; Sullivan, J.M.; Witmer, D.D.; Petrelli, N.J. Molecular prognostics in colorectal cancer. *Surg. Oncol.* **2003**, *12*, 173–186. [CrossRef]

4. Compton, C.C. Colorectal carcinoma: Diagnostic, prognostic, and molecular features. *Mod. Pathol.* **2003**, *16*, 376–388. [CrossRef] [PubMed]

5. Berg, M.; Nordgaard, O.; Korner, H.; Oltedal, S.; Smaaland, R.; Soreide, J.A.; Soreide, K. Molecular subtypes in stage II–III colon cancer defined by genomic instability: Early recurrence-risk associated with a high copy-number variation and loss of *RUNX3* and *CDKN2A*. *PLoS ONE* **2015**, *10*, e0122391. [CrossRef] [PubMed]

6. Liu, P.; Carvalho, C.M.; Hastings, P.J.; Lupski, J.R. Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.* **2012**, *22*, 211–220. [CrossRef] [PubMed]

7. Sanz-Pamplona, R.; Berenguer, A.; Cordero, D.; Riccadonna, S.; Sole, X.; Crous-Bou, M.; Guino, E.; Sanjuan, X.; Biondo, S.; Soriano, A.; *et al.* Clinical value of prognosis gene expression signatures in colorectal cancer: A systematic review. *PLoS ONE* **2012**, *7*, e48877. [CrossRef] [PubMed]

8. Shi, M.; Beauchamp, R.D.; Zhang, B. A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients. *PLoS ONE* **2012**, *7*, e41292. [CrossRef] [PubMed]

9. Wellcome Trust Sanger Institute. Available online: http://www.sanger.ac.uk/research/areas/humangenetics/CNA/ (accessed on 20 June 2014).

10. Database of Genomic Variants. Available online: http://dgv.tcag.ca/dgv/app/chromosome?ref=NCBI36/hg18 (accessed on 11 July 2014).

11. Markle, B.; May, E.J.; Majumdar, A.P. Do nutraceutics play a role in the prevention and treatment of colorectal cancer? *Cancer Metastasis Rev.* **2010**, *29*, 395–404. [CrossRef] [PubMed]

12. Jones, S.; Chen, W.D.; Parmigiani, G.; Diehl, F.; Beerenwinkel, N.; Antal, T.; Traulsen, A.; Nowak, M.A.; Siegel, C.; Velculescu, V.E.; *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4283–4288. [CrossRef] [PubMed]

13. Kim, S. New and emerging factors in tumorigenesis: An overview. *Cancer Manag. Res.* **2015**, *7*, 225–239. [CrossRef] [PubMed]

14. Shull, A.Y.; Clendenning, M.L.; Ghoshal-Gupta, S.; Farrell, C.L.; Vangapandu, H.V.; Dudas, L.; Wilkerson, B.J.; Buckhaults, P.J. Somatic mutations, allele loss, and DNA methylation of the Cub and Sushi Multiple Domains 1 (*CSMD1*) gene reveals association with early age of diagnosis in colorectal cancer patients. *PLoS ONE* **2013**, *8*, e58731. [CrossRef] [PubMed]

15. Udali, S.; Guarini, P.; Ruzzenente, A.; Ferrarini, A.; Guglielmi, A.; Lotto, V.; Tononi, P.; Pattini, P.; Moruzzi, S.; Campagnaro, T.; *et al.* DNA methylation and gene expression profiles show novel regulatory pathways in hepatocellular carcinoma. *Clin. Epigenet.* **2015**, *7*, 43. [CrossRef] [PubMed]

16. Yi, J.K.; Kim, H.J.; Yu, D.H.; Park, S.J.; Shin, M.J.; Yuh, H.S.; Bae, K.B.; Ji, Y.R.; Kim, N.R.; Park, S.J.; *et al.* Regulation of inflammatory responses and fibroblast-like synoviocyte apoptosis by calcineurin-binding protein 1 in mice with collagen-induced arthritis. *Arthritis Rheum.* **2012**, *64*, 2191–2200. [CrossRef] [PubMed]

17. Tsai, J.H.; Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* **2013**, *27*, 2192–2206. [CrossRef] [PubMed]

18. Collette, K.M.; Zhou, X.D.; Amoth, H.M.; Lyons, M.J.; Papay, R.S.; Sens, D.A.; Perez, D.M.; Doze, V.A. Long-term α1B-adrenergic receptor activation shortens lifespan, while α1A-adrenergic receptor stimulation prolongs lifespan in association with decreased cancer incidence. *Age* **2014**, *36*, 9675. [CrossRef] [PubMed]

19. Ansems, M.; Hontelez, S.; Karthaus, N.; Span, P.N.; Adema, G.J. Crosstalk and DC-script: Expanding nuclear receptor modulation. *Biochim. Biophys. Acta* **2010**, *1806*, 193–199. [CrossRef] [PubMed]

20. Hulme, A.E.; Bogerd, H.P.; Cullen, B.R.; Moran, J.V. Selective inhibition of Alu retrotransposition by *APOBEC3G*. *Gene* **2007**, *390*, 199–205. [CrossRef] [PubMed]

21. Park, J.; Long, D.T.; Lee, K.Y.; Abbas, T.; Shibata, E.; Negishi, M.; Luo, Y.; Schimenti, J.C.; Gambus, A.; Walter, J.C.; *et al.* The MCM8-MCM9 complex promotes RAD51 recruitment at DNA damage sites to facilitate homologous recombination. *Mol. Cell. Biol.* **2013**, *33*, 1632–1644. [CrossRef] [PubMed]

22. Wu, M.Z.; Tsai, Y.P.; Yang, M.H.; Huang, C.H.; Chang, S.Y.; Chang, C.C.; Teng, S.C.; Wu, K.J. Interplay between *HDAC3* and *WDR5* is essential for hypoxia-induced epithelial-mesenchymal transition. *Mol. Cell.* **2011**, *43*, 811–822. [CrossRef] [PubMed]

23. Perrault, C.; Mangin, P.; Santer, M.; Baas, M.J.; Moog, S.; Cranmer, S.L.; Pikovski, I.; Williamson, D.; Jackson, S.P.; Cazenave, J.P.; *et al.* Role of the intracellular domains of GPIB in controlling the adhesive properties of the platelet GPIB/V/IX complex. *Blood* **2003**, *101*, 3477–3484. [CrossRef] [PubMed]

24. Waha, A.; Guntner, S.; Huang, T.H.; Yan, P.S.; Arslan, B.; Pietsch, T.; Wiestler, O.D.; Waha, A. Epigenetic silencing of the protocadherin family member *PCDH-γ-A11* in astrocytomas. *Neoplasia* **2005**, *7*, 193–199. [CrossRef] [PubMed]

25. Sibilio, A.; Ambrosio, R.; Bonelli, C.; de Stefano, M.A.; Torre, V.; Dentice, M.; Salvatore, D. Deiodination in cancer growth: The role of type III deiodinase. *Minerv. Endocrinol.* **2012**, *37*, 315–327.

26. Zhou, Y.; Lan, J.; Wang, W.; Shi, Q.; Lan, Y.; Cheng, Z.; Guan, H. *ZNRF3* acts as a tumour suppressor by the Wnt signalling pathway in human gastric adenocarcinoma. *J. Mol. Histol.* **2013**, *44*, 555–563. [CrossRef] [PubMed]

27. Wang, K.; Holterman, A.X. Pathophysiologic role of hepatocyte nuclear factor 6. *Cell Signal.* **2012**, *24*, 9–16. [CrossRef] [PubMed]

28. Vega, S.; Morales, A.V.; Ocana, O.H.; Valdes, F.; Fabregat, I.; Nieto, M.A. Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **2004**, *18*, 1131–1143. [CrossRef] [PubMed]

29. Tsai, J.H.; Donaher, J.L.; Murphy, D.A.; Chau, S.; Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **2012**, *22*, 725–736. [CrossRef] [PubMed]

30. Ren, L.; Liu, Y.; Guo, L.; Wang, H.; Ma, L.; Zeng, M.; Shao, X.; Yang, C.; Tang, Y.; Wang, L.; *et al.* Loss of SMU1 function de-represses DNA replication and over-activates ATR-dependent replication checkpoint. *Biochem. Biophys. Res. Commun.* **2013**, *436*, 192–198. [CrossRef] [PubMed]

31. Han, X.; Chen, Y.; Yao, N.; Liu, H.; Wang, Z. MicroRNA *LET-7B* suppresses human gastric cancer malignancy by targeting *ING1*. *Cancer Gene Ther.* **2015**, *22*, 122–129. [CrossRef] [PubMed]

32. Chen, T.; Wu, F.; Chen, F.M.; Tian, J.; Qu, S. Variations of very low-density lipoprotein receptor subtype expression in gastrointestinal adenocarcinoma cells with various differentiations. *World J. Gastroenterol.* **2005**, *11*, 2817–2821. [CrossRef] [PubMed]

33. Akhavantabasi, S.; Akman, H.B.; Sapmaz, A.; Keller, J.; Petty, E.M.; Erson, A.E. USP32 is an active, membrane-bound ubiquitin protease overexpressed in breast cancers. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **2010**, *21*, 388–397. [CrossRef] [PubMed]

34. Garner, K.; Hunt, A.N.; Koster, G.; Somerharju, P.; Groves, E.; Li, M.; Raghu, P.; Holic, R.; Cockcroft, S. Phosphatidylinositol transfer protein, cytoplasmic 1 (*PITPNC1*) binds and transfers phosphatidic acid. *J. Biol. Chem.* **2012**, *287*, 32263–32276. [CrossRef] [PubMed]

35. Eckert, M.A.; Lwin, T.M.; Chang, A.T.; Kim, J.; Danis, E.; Ohno-Machado, L.; Yang, J. *TWIST1*-induced invadopodia formation promotes tumor metastasis. *Cancer Cell.* **2011**, *19*, 372–386. [CrossRef] [PubMed]

36. Pignatelli, J.; Tumbarello, D.A.; Schmidt, R.P.; Turner, C.E. Hic-5 promotes invadopodia formation and invasion during TGF-β-induced epithelial-mesenchymal transition. *J. Cell Biol.* **2012**, *197*, 421–437. [CrossRef] [PubMed]

37. Bosso, M.; Al-Mulla, F. Whole genome amplification of DNA extracted from ffpe tissues. *Methods Mol. Biol.* **2011**, *724*, 161–180. [PubMed]

38. Al-Mulla, F. Microarray-based CGH and copy number analysis of FFPE samples. *Methods Mol. Biol.* **2011**, *724*, 131–145. [PubMed]

39. Diskin, S.J.; Eck, T.; Greshock, J.; Mosse, Y.P.; Naylor, T.; Stoeckert, C.J., Jr.; Weber, B.L.; Maris, J.M.; Grant, G.R. STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res.* **2006**, *16*, 1149–1158. [CrossRef] [PubMed]

40. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* **2007**, *8*, 118–127. [CrossRef] [PubMed]

41. Tan, T.Z.; Miow, Q.H.; Miki, Y.; Noda, T.; Mori, S.; Huang, R.Y.; Thiery, J.P. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **2014**, *6*, 1279–1293. [CrossRef] [PubMed]

42. Lee, J.M.; Dedhar, S.; Kalluri, R.; Thompson, E.W. The epithelial-mesenchymal transition: New insights in signaling, development, and disease. *J. Cell Biol.* **2006**, *172*, 973–981. [CrossRef] [PubMed]

43. Schweizer, J.; Bowden, P.E.; Coulombe, P.A.; Langbein, L.; Lane, E.B.; Magin, T.M.; Maltais, L.; Omary, M.B.; Parry, D.A.; Rogers, M.A.; *et al.* New consensus nomenclature for mammalian keratins. *J. Cell Biol.* **2006**, *174*, 169–174. [CrossRef] [PubMed]

44. Moll, R.; Divo, M.; Langbein, L. The human keratins: Biology and pathology. *Histochem. Cell Biol.* **2008**, *129*, 705–733. [CrossRef] [PubMed]

45. Gatza, M.L.; Lucas, J.E.; Barry, W.T.; Kim, J.W.; Wang, Q.; Crawford, M.D.; Datto, M.B.; Kelley, M.; Mathey-Prevot, B.; Potti, A.; *et al.* A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6994–6999. [CrossRef] [PubMed]

46. Verhaak, R.G.; Tamayo, P.; Yang, J.Y.; Hubbard, D.; Zhang, H.; Creighton, C.J.; Fereday, S.; Lawrence, M.; Carter, S.L.; Mermel, C.H.; *et al.* Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Investig.* **2013**, *123*, 517–525. [CrossRef] [PubMed]

47. Ballou, L.M.; Cross, M.E.; Huang, S.; McReynolds, E.M.; Zhang, B.X.; Lin, R.Z. Differential regulation of the phosphatidylinositol 3-kinase/akt and p70 s6 kinase pathways by the α(1a)-adrenergic receptor in rat-1 fibroblasts. *J. Biol. Chem.* **2000**, *275*, 4803–4809. [CrossRef] [PubMed]

48. Kassahun, W.T.; Gunl, B.; Jonas, S.; Ungemach, F.R.; Abraham, G. Altered liver α1-adrenoceptor density and phospholipase c activity in the human hepatocellular carcinoma. *Eur. J. Pharmacol.* **2011**, *670*, 92–95. [CrossRef] [PubMed]

49. Keffel, S.; Alexandrov, A.; Goepel, M.; Michel, M.C. α(1)-adrenoceptor subtypes differentially couple to growth promotion and inhibition in chinese hamster ovary cells. *Biochem. Biophys. Res. Commun.* **2000**, *272*, 906–911. [CrossRef] [PubMed]

50. Gonzalez-Cabrera, P.J.; Shi, T.; Yun, J.; McCune, D.F.; Rorabaugh, B.R.; Perez, D.M. Differential regulation of the cell cycle by α1-adrenergic receptor subtypes. *Endocrinology* **2004**, *145*, 5157–5167. [CrossRef] [PubMed]

51. Vinci, M.C.; Bellik, L.; Filippi, S.; Ledda, F.; Parenti, A. Trophic effects induced by α1d-adrenoceptors on endothelial cells are potentiated by hypoxia. *Am. J. Physiol. Heart Circ. Physiol.* **2007**, *293*, H2140–H2147. [CrossRef] [PubMed]

52. Shibata, K.; Katsuma, S.; Koshimizu, T.; Shinoura, H.; Hirasawa, A.; Tanoue, A.; Tsujimoto, G. A 1-adrenergic receptor subtypes differentially control the cell cycle of transfected cho cells through a camp-dependent mechanism involving p27kip1. *J. Biol. Chem.* **2003**, *278*, 672–678. [CrossRef] [PubMed]

53. Babol, K.; Przybylowska, K.; Lukaszek, M.; Pertynski, T.; Blasiak, J. An association between the trp64arg polymorphism in the beta3-adrenergic receptor gene and endometrial cancer and obesity. *J. Exp. Clin. Cancer Res.* **2004**, *23*, 669–674. [PubMed]

54. Guay, S.P.; Brisson, D.; Lamarche, B.; Biron, S.; Lescelleur, O.; Biertho, L.; Marceau, S.; Vohl, M.C.; Gaudet, D.; Bouchard, L. Adrb3 gene promoter DNA methylation in blood and visceral adipose tissue is associated with metabolic disturbances in men. *Epigenomics* **2014**, *6*, 33–43. [CrossRef] [PubMed]

55. Lackey, L.; Law, E.K.; Brown, W.L.; Harris, R.S. Subcellular localization of the apobec3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle* **2013**, *12*, 762–772. [CrossRef] [PubMed]

56. Stenglein, M.D.; Harris, R.S. Apobec3b and apobec3f inhibit l1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.* **2006**, *281*, 16837–16841. [CrossRef] [PubMed]

57. Lu, M.; Tian, H.; Yue, W.; Li, L.; Li, S.; Qi, L.; Hu, W.; Gao, C.; Si, L. Overexpression of tfiib-related factor 2 is significantly correlated with tumor angiogenesis and poor survival in patients with esophageal squamous cell cancer. *Med. Oncol.* **2013**, *30*, 553. [CrossRef] [PubMed]

58. Tian, Y.; Lu, M.; Yue, W.; Li, L.; Li, S.; Gao, C.; Si, L.; Qi, L.; Hu, W.; Tian, H. Tfiib-related factor 2 is associated with poor prognosis of nonsmall cell lung cancer patients through promoting tumor epithelial-mesenchymal transition. *BioMed Res. Int.* **2014**, *2014*, 530786. [CrossRef] [PubMed]

59. Yu, D.H.; Yi, J.K.; Park, S.J.; Kim, M.O.; Kim, H.J.; Yuh, H.S.; Bae, K.B.; Ji, Y.R.; Lee, H.S.; Lee, S.G.; *et al.* Tissue-specific expression of human calcineurin-binding protein 1 in mouse synovial tissue can suppress inflammatory arthritis. *J. Interferon Cytokine Res.* **2012**, *32*, 6–11. [CrossRef] [PubMed]

60. Talavera, K.; Nilius, B. Biophysics and structure-function relationship of t-type ca2+ channels. *Cell Calcium* **2006**, *40*, 97–114. [CrossRef] [PubMed]

61. Farrell, C.; Crimm, H.; Meeh, P.; Croshaw, R.; Barbar, T.; Vandersteenhoven, J.J.; Butler, W.; Buckhaults, P. Somatic mutations to csmd1 in colorectal adenocarcinomas. *Cancer Biol. Ther.* **2008**, *7*, 609–613. [CrossRef] [PubMed]

62. Dentice, M.; Ambrosio, R.; Salvatore, D. Role of type 3 deiodinase in cancer. *Exp. Opin. Ther. Targets* **2009**, *13*, 1363–1373. [CrossRef] [PubMed]

63. Newman, J.W.; Morisseau, C.; Harris, T.R.; Hammock, B.D. The soluble epoxide hydrolase encoded by epxh2 is a bifunctional enzyme with novel lipid phosphate phosphatase activity. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1558–1563. [CrossRef] [PubMed]

64. Diede, S.J.; Yao, Z.; Keyes, C.C.; Tyler, A.E.; Dey, J.; Hackett, C.S.; Elsaesser, K.; Kemp, C.J.; Neiman, P.E.; Weiss, W.A.; *et al.* Fundamental differences in promoter cpg island DNA hypermethylation between human cancer and genetically engineered mouse models of cancer. *Epigenetics* **2013**, *8*, 1254–1260. [CrossRef] [PubMed]

65. Nishimura, K.; Ishiai, M.; Horikawa, K.; Fukagawa, T.; Takata, M.; Takisawa, H.; Kanemaki, M.T. Mcm8 and mcm9 form a complex that functions in homologous recombination repair induced by DNA interstrand crosslinks. *Mol. Cell* **2012**, *47*, 511–522. [CrossRef] [PubMed]

66. Lilla, C.; Verla-Tebit, E.; Risch, A.; Jager, B.; Hoffmeister, M.; Brenner, H.; Chang-Claude, J. Effect of nat1 and nat2 genetic polymorphisms on colorectal cancer risk associated with exposure to tobacco smoke and meat consumption. *Cancer Epidemiol. Biomark. Prev.* **2006**, *15*, 99–107. [CrossRef] [PubMed]

67. Johansson, I.; Nilsson, C.; Berglund, P.; Lauss, M.; Ringner, M.; Olsson, H.; Luts, L.; Sim, E.; Thorstensson, S.; Fjallskog, M.L.; *et al.* Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. *Breast Cancer Res.* **2012**, *14*, R31. [CrossRef] [PubMed]

68. Cai, J.; Zhao, Y.; Zhu, C.L.; Li, J.; Huang, Z.H. The association of nat1 polymorphisms and colorectal carcinoma risk: Evidence from 20,000 subjects. *Mol. Biol. Rep.* **2012**, *39*, 7497–7503. [CrossRef] [PubMed]

69. Zhuo, W.; Zhang, L.; Qiu, Z.; Cai, L.; Zhu, B.; Chen, Z. Association of nat2 polymorphisms with risk of colorectal adenomas: Evidence from 3,197 cases and 4,681 controls. *Exp. Ther. Med.* **2012**, *4*, 895–900. [CrossRef] [PubMed]

70. Jacquemin, P.; Durviaux, S.M.; Jensen, J.; Godfraind, C.; Gradwohl, G.; Guillemot, F.; Madsen, O.D.; Carmeliet, P.; Dewerchin, M.; Collen, D.; *et al.* Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene *NGN3*. *Mol. Cell. Biol.* **2000**, *20*, 4445–4454. [CrossRef] [PubMed]

71. Prevot, P.P.; Simion, A.; Grimont, A.; Colletti, M.; Khalaileh, A.; Van den Steen, G.; Sempoux, C.; Xu, X.; Roelants, V.; Hald, J.; *et al.* Role of the ductal transcription factors HNF6 and Sox9 in pancreatic acinar-to-ductal metaplasia. *Gut* **2012**, *61*, 1723–1732. [CrossRef] [PubMed]

72. Lehner, F.; Kulik, U.; Klempnauer, J.; Borlak, J. The hepatocyte nuclear factor 6 (hnf6) and foxa2 are key regulators in colorectal liver metastases. *FASEB J.* **2007**, *21*, 1445–1462. [CrossRef] [PubMed]

73. Cullis, D.N.; Philip, B.; Baleja, J.D.; Feig, L.A. Rab11-Fip2, an adaptor protein connecting cellular components involved in internalization and recycling of epidermal growth factor receptors. *J. Biol. Chem.* **2002**, *277*, 49158–49166. [CrossRef] [PubMed]

74. Carson, B.P.; Del Bas, J.M.; Moreno-Navarrete, J.M.; Fernandez-Real, J.M.; Mora, S. The rab11 effector protein fip1 regulates adiponectin trafficking and secretion. *PLoS ONE* **2013**, *8*, e74687. [CrossRef] [PubMed]

75. Gori, F.; Friedman, L.; Demay, M.B. Wdr5, a novel wd repeat protein, regulates osteoblast and chondrocyte differentiation in vivo. *J. Musculoskelet. Neuronal Interact.* **2005**, *5*, 338–339. [PubMed]

76. Ang, Y.S.; Tsai, S.Y.; Lee, D.F.; Monk, J.; Su, J.; Ratnakumar, K.; Ding, J.; Ge, Y.; Darr, H.; Chang, B.; *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **2011**, *145*, 183–197. [CrossRef] [PubMed]

77. Lopez-Garcia, J.; Periyasamy, M.; Thomas, R.S.; Christian, M.; Leao, M.; Jat, P.; Kindle, K.B.; Heery, D.M.; Parker, M.G.; Buluwela, L.; *et al.* Znf366 is an estrogen receptor corepressor that acts through ctbp and histone deacetylases. *Nucleic Acids Res.* **2006**, *34*, 6126–6136. [CrossRef] [PubMed]

78. Ansems, M.; Karthaus, N.; Hontelez, S.; Aalders, T.; Looman, M.W.; Verhaegh, G.W.; Schalken, J.A.; Adema, G.J. Dc-script: Ar and vdr regulator lost upon transformation of prostate epithelial cells. *Prostate* **2012**, *72*, 1708–1717. [CrossRef] [PubMed]

79. Sircoulomb, F.; Nicolas, N.; Ferrari, A.; Finetti, P.; Bekhouche, I.; Rousselet, E.; Lonigro, A.; Adelaide, J.; Baudelet, E.; Esteyries, S.; *et al.* Znf703 gene amplification at 8p12 specifies luminal b breast cancer. *EMBO Mol. Med.* **2011**, *3*, 153–166. [CrossRef] [PubMed]

80. Ma, F.; Bi, L.; Yang, G.; Zhang, M.; Liu, C.; Zhao, Y.; Wang, Y.; Wang, J.; Bai, Y.; Zhang, Y. Znf703 promotes tumor cell proliferation and invasion and predicts poor prognosis in patients with colorectal cancer. *Oncol. Rep.* **2014**, *32*, 1071–1077. [CrossRef] [PubMed]

81. Warnatz, H.J.; Schmidt, D.; Manke, T.; Piccini, I.; Sultan, M.; Borodina, T.; Balzereit, D.; Wruck, W.; Soldatov, A.; Vingron, M.; *et al.* The btb and cnc homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. *J. Biol. Chem.* **2011**, *286*, 23521–23532. [CrossRef] [PubMed]

82. Wilkinson, S.; Paterson, H.F.; Marshall, C.J. Cdc42-mrck and rho-rock signalling cooperate in myosin phosphorylation and cell invasion. *Nat. Cell Biol.* **2005**, *7*, 255–261. [CrossRef] [PubMed]

83. Shi, L.; Yue, J.; You, Y.; Yin, B.; Gong, Y.; Xu, C.; Qiang, B.; Yuan, J.; Liu, Y.; Peng, X. Dok5 is substrate of trkb and trkc receptors and involved in neurotrophin induced mapk activation. *Cell Signal.* **2006**, *18*, 1995–2003. [CrossRef] [PubMed]

84. Pan, Y.; Zhang, J.; Liu, W.; Shu, P.; Yin, B.; Yuan, J.; Qiang, B.; Peng, X. Dok5 is involved in the signaling pathway of neurotrophin-3 against trkc-induced apoptosis. *Neuro Lett.* **2013**, *553*, 46–51. [CrossRef] [PubMed]

85. Zheng, H.; Li, Q.; Chen, R.; Zhang, J.; Ran, Y.; He, X.; Li, S.; Shu, H.B. The dual-specificity phosphatase dusp14 negatively regulates tumor necrosis factor- and interleukin-1-induced nuclear factor-kappab activation by dephosphorylating the protein kinase tak1. *J. Biol. Chem.* **2013**, *288*, 819–825. [CrossRef] [PubMed]

86. Yang, C.Y.; Li, J.P.; Chiu, L.L.; Lan, J.L.; Chen, D.Y.; Chuang, H.C.; Huang, C.Y.; Tan, T.H. Dual-specificity phosphatase 14 (DUSP14/MKP6) negatively regulates tcr signaling by inhibiting tab1 activation. *J. Immunol.* **2014**, *192*, 1547–1557. [CrossRef] [PubMed]

87. Huang, Y.; Prasad, M.; Lemon, W.J.; Hampel, H.; Wright, F.A.; Kornacker, K.; LiVolsi, V.; Frankel, W.; Kloos, R.T.; Eng, C.; *et al.* Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 15044–15049. [CrossRef] [PubMed]

88. Lichti-Kaiser, K.; ZeRuth, G.; Kang, H.S.; Vasanth, S.; Jetten, A.M. Gli-similar proteins: Their mechanisms of action, physiological functions, and roles in disease. *Vitam. Horm.* **2012**, *88*, 141–171. [PubMed]

89. Adachi, M.; Hamazaki, Y.; Kobayashi, Y.; Itoh, M.; Tsukita, S.; Furuse, M. Similar and distinct properties of mupp1 and patj, two homologous PDZ domain-containing tight-junction proteins. *Mol. Cell. Biol.* **2009**, *29*, 2372–2389. [CrossRef] [PubMed]

90. Ernkvist, M.; Luna Persson, N.; Audebert, S.; Lecine, P.; Sinha, I.; Liu, M.; Schlueter, M.; Horowitz, A.; Aase, K.; Weide, T.; *et al.* The amot/patj/syx signaling complex spatially controls rhoa gtpase activity in migrating endothelial cells. *Blood* **2009**, *113*, 244–253. [CrossRef] [PubMed]

91. Hamazaki, Y.; Itoh, M.; Sasaki, H.; Furuse, M.; Tsukita, S. Multi-pdz domain protein 1 (mupp1) is concentrated at tight junctions through its possible interaction with claudin-1 and junctional adhesion molecule. *J. Biol. Chem.* **2002**, *277*, 455–461. [CrossRef] [PubMed]

92. Png, K.J.; Halberg, N.; Yoshida, M.; Tavazoie, S.F. A microrna regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature* **2012**, *481*, 190–194. [CrossRef] [PubMed]

93. Fonseca-Sanchez, M.A.; Perez-Plasencia, C.; Fernandez-Retana, J.; Arechaga-Ocampo, E.; Marchat, L.A.; Rodriguez-Cuevas, S.; Bautista-Pina, V.; Arellano-Anaya, Z.E.; Flores-Perez, A.; Diaz-Chavez, J.; *et al.* Microrna-18b is upregulated in breast cancer and modulates genes involved in cell migration. *Oncol. Rep.* **2013**, *30*, 2399–2410. [PubMed]

94. Chen, L.S.; Wei, J.B.; Zhou, Y.C.; Zhang, S.; Liang, J.L.; Cao, Y.F.; Tang, Z.J.; Zhang, X.L.; Gao, F. Genetic alterations and expression of inhibitor of growth 1 in human sporadic colorectal cancer. *World J. Gastroenterol.* **2005**, *11*, 6120–6124. [CrossRef] [PubMed]