Integrated Exon Level Expression Analysis of Driver Genes Explain Their Role in Colorectal Cancer

Mohammad Azhar Aziz[1]*, Sathish Periyasamy[2], Zeyad Al Yousef[3], Ibrahim AlAbdulkarim[1], Majed Al Otaibi[1], Abdulaziz Alfahed[1], Glowi Alasiri[1]

1 Medical Genomics, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, **2** Bioinformatics, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, **3** Surgery, National Guard Health Affairs, Riyadh, Saudi Arabia

Abstract

Integrated analysis of genomic and transcriptomic level changes holds promise for a better understanding of colorectal cancer (CRC) biology. There is a pertinent need to explain the functional effect of genome level changes by integrating the information at the transcript level. Using high resolution cytogenetics array, we had earlier identified driver genes by 'Genomic Identification of Significant Targets In Cancer (GISTIC)' analysis of paired tumour-normal samples from colorectal cancer patients. In this study, we analyze these driver genes at three levels using exon array data – gene, exon and network. Gene level analysis revealed a small subset to experience differential expression. These results were reinforced by carrying out separate differential expression analyses (SAM and LIMMA). ATP8B1 was found to be the novel gene associated with CRC that shows changes at cytogenetic, gene and exon levels. Splice index of 29 exons corresponding to 13 genes was found to be significantly altered in tumour samples. Driver genes were used to construct regulatory networks for tumour and normal groups. There were rearrangements in transcription factor genes suggesting the presence of regulatory switching. The regulatory pattern of AHR gene was found to have the most significant alteration. Our results integrate data with focus on driver genes resulting in highly enriched novel molecules that need further studies to establish their role in CRC.

Citation: Aziz MA, Periyasamy S, Al Yousef Z, AlAbdulkarim I, Al Otaibi M, et al. (2014) Integrated Exon Level Expression Analysis of Driver Genes Explain Their Role in Colorectal Cancer. PLoS ONE 9(10): e110134. doi:10.1371/journal.pone.0110134

Editor: Osman El-Maarri, University of Bonn, Institut of experimental hematology and transfusion medicine, Germany

Received February 20, 2014; **Accepted** September 16, 2014; **Published** October 21, 2014

Funding: This study was funded by King Abdullah International Medical Research Center through research grant RC10/083 awarded to MAA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: azharbiotech@yahoo.com

Introduction

There is a wealth of information at omics level that associates cytogenetics and gene expression changes leading to colorectal cancer (CRC). The integration of gene expression and copy number (CN) data to identify DNA CN alterations that induce changes in the expression levels of the associated genes is a common task in cancer studies [1–3]. The central dogma of molecular biology has thus been addressed at two important levels. There have been many reports providing evidence of changes at the genome level in the form of copy number aberration [4], single nucleotide polymorphisms, loss of heterozygosity that attempt to understand the molecular events associated with colorectal cancer. These somatic or hereditary changes have different mechanism of contributing to initiation and progression of CRC. Loss and gain of crucial chromosomal regions leading to deletion or amplification of cancer related genes has been very well established. The functional significance of these molecular events has been measured using different tools and algorithms. Genes targeted by somatic copy-number alterations (SCNAs), in particular, play central roles in oncogenesis and cancer therapy [5]. Several tools have been made available to assess the potential of genes that get affected by SCNAs in causing colorectal cancer. 'Genomic Identification of Significant Targets in Cancer' (GISTIC) tool has successfully been used in identifying 'driver SCNAs' based on the frequency and amplitude of observed events [6,7]. The second aspect of changes happening in tumour cells is at the transcription level. Differential expression analysis has been carried out to find out important genes playing a role in causing colorectal cancer. There could be several mechanisms by which the SCNA affected genes exert their effect at functional level. Amplifications and deletions in the genomic region are reflected in the transcript levels and could be detected by carrying out expression microarray based studies. By employing exon arrays, we gain extra dimension of the events happening at the exon level, which may lead to alternative splicing resulting in different gene isoforms. Alternative splicing is a crucial step in the generation of protein diversity and its misregulation is observed in many human cancer types [8].

The quest to explore the relationship between copy number changes and the expression level of affected genes/exons has received limited success owing to a number of reasons [9]. Technological improvements in the array design for cytogenetics as well as transcriptomics have improved the accuracy and precision of the data generated. Combining this with better analytical techniques and algorithms, possibilities of finding target genes responsible for causing colorectal cancer has further increased.

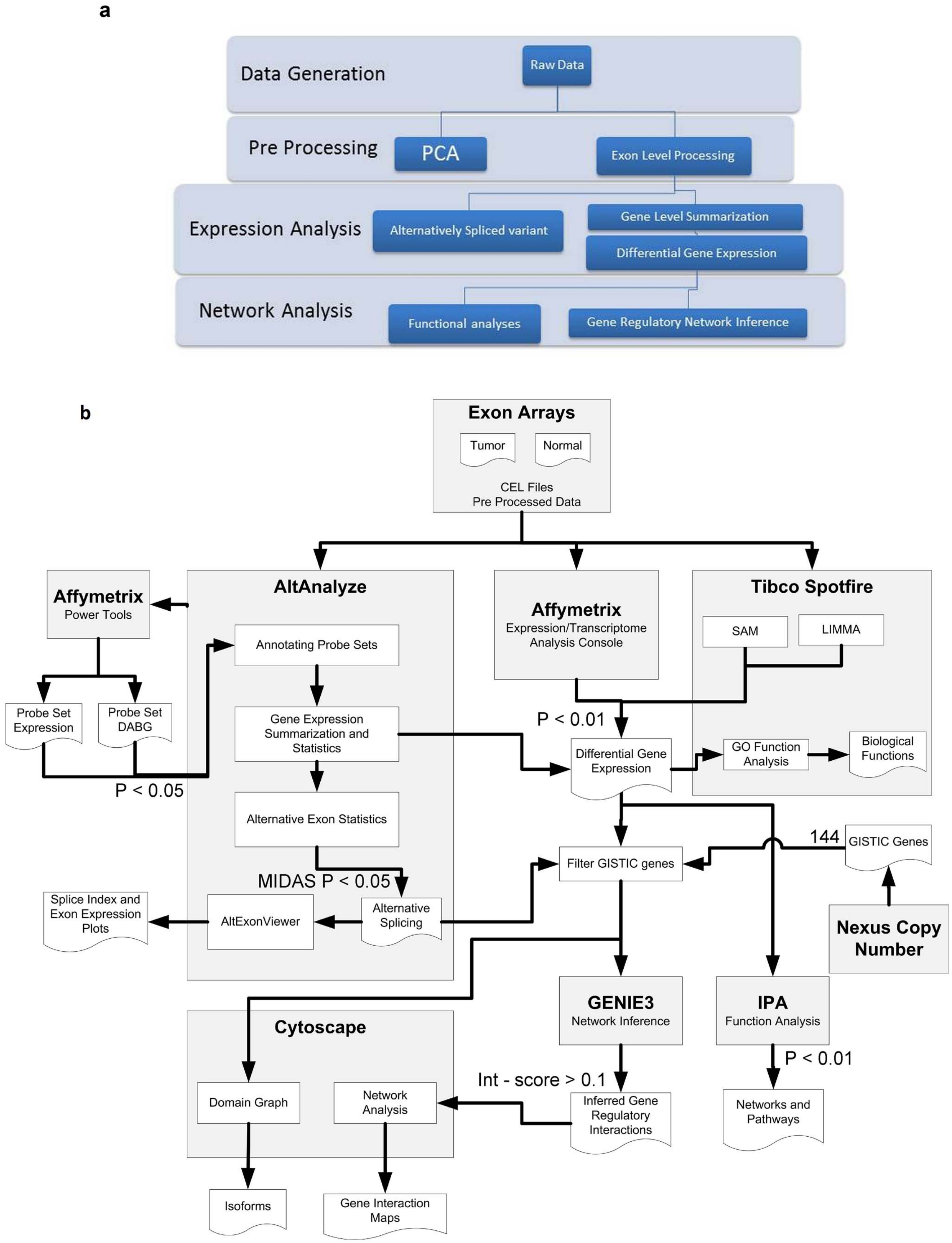
a
Data Generation
Raw Data
Pre Processing
PCA
Exon Level Processing
Expression Analysis
Alternatively Spliced variant
Gene Level Summarization
Differential Gene Expression
Network Analysis
Functional analyses
Gene Regulatory Network Inference
b
Exon Arrays
Tumor
Normal
CEL Files
Pre Processed Data
Affymetrix
Power Tools
Probe Set Expression
Probe Set DABG
P < 0.05
AltAnalyze
Annotating Probe Sets
Gene Expression Summarization and Statistics
Alternative Exon Statistics
MIDAS P < 0.05
Alternative Splicing
AltExonViewer
Splice Index and Exon Expression Plots
Affymetrix
Expression/Transcriptome Analysis Console
Tibco Spotfire
SAM
LIMMA
P < 0.01
Differential Gene Expression
GO Function Analysis
Biological Functions
144
GISTIC Genes
Filter GISTIC genes
Nexus Copy Number
GENIE3
Network Inference
IPA
Function Analysis
P < 0.01
Networks and Pathways
Int - score > 0.1
Inferred Gene Regulatory Interactions
Cytoscape
Domain Graph
Network Analysis
Isoforms
Gene Interaction Maps

Figure 1. Flow diagram for Analysis Strategy. A)The entire analyses is categorized into fours stages from 'Data Generation' to 'Network Analyses'. B) Analysis strategy using different programs is displayed in this diagram. There are three components of the analysis – Gene, Exon and Network handled by different programs. Gene level analyses are conducted using 'Affymetrix, Expression/Transcriptome analysis console' and 'Tibco Spotfire'. Exon level analysis is carried out by 'AltAnalyze' and 'Affymetrix power tools'. Network analyses employed 'GENIE3', 'IPA' and 'Cytoscape'. 'Nexus Copy Number' is a program used in earlier studies to eventually generate a list of 144 driver genes.
doi:10.1371/journal.pone.0110134.g001

Past few decades have seen a quest for finding novel genes that can serve as therapeutic targets or biomarkers. However, genes or proteins do not function alone but interact with each other to form networks or pathways so as to carry out biological functions [10]. Network-based approaches to finding biomarkers more closely represent *in vivo* molecular biology where a perturbation in one gene may affect many downstream genes. Cancer has thus been rightly addressed as a systems biology disease [11] as opposed to diseases caused by changes in few genes or mutations. Reconstructing gene regulatory networks in healthy and diseased tissues is therefore critical to understanding cancer phenotypes and devising effective therapeutics [12].

With the availability of tools and techniques to capture the molecular changes happening at different stages of the central dogma with increased precision and accuracy, we are yet to develop an integrated comprehensive picture that could help us in finding better targets for colorectal cancer. In the present study, we aim to integrate the information from the cytogenetic analysis with the exon level expression data using paired normal-tumour samples from colorectal cancer patients. A set of driver genes suggested by GISTIC analysis (termed as 'driver genes' from now onwards) were queried at gene, exon and network level. Both DNA and RNA for cytogenetics and transcriptomic studies, respectively, were extracted from the same tissue in a single workflow to minimize variation. This study provides evidence for

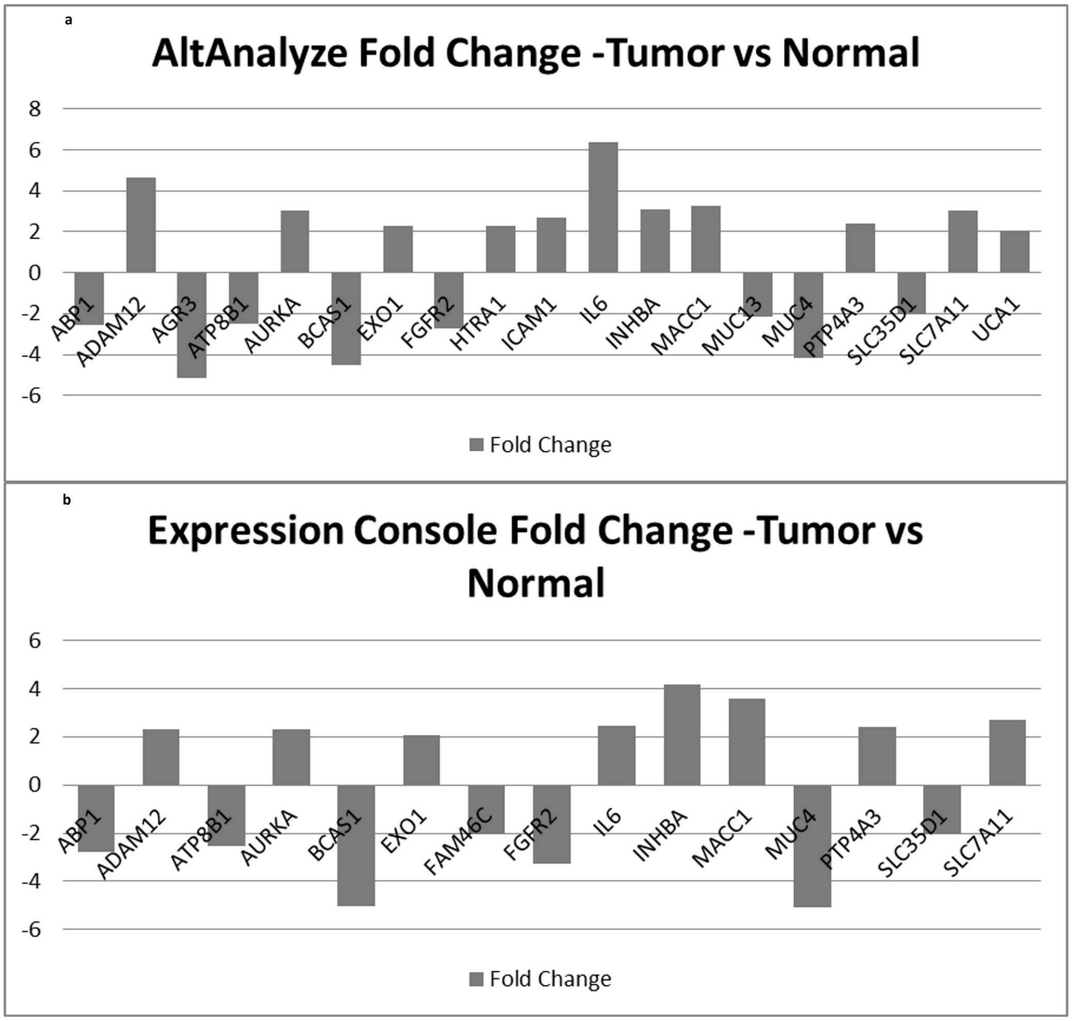

Figure 2. Significant change in expression value at gene level was observed in 20/144 genes. Two different algorithms were used to measure expression values from Exon array data to support the results. AltAnalyze (1a) and Expression Console (1b) show complimentary results with maximum changes observed in BCAS1, INHBA, IL6 and MUC4 genes.
doi:10.1371/journal.pone.0110134.g002

Table 1. Driver Genes from GISTIC analysis showing more than two fold change in expression value as calculated by two different programs – AltAnalyze and Expression Console.

Gene Symbol	GISTIC score	AltAnalyze	Expression Console
BCAS1[a,b]	5.323	−4.50771	−5.02
AURKA[a,b]	4.772	3.044088	2.31
ATP8B1[a]	2.326	−2.49709	−2.55
IL6[a,b]	2	6.353318	2.45
INHBA[a,c]	1.944	3.071911	4.18
PTP4A3[a,b,c]	1.892	2.366951	2.41
ABP1[a,b]	1.859	−2.54991	−2.78
AGR3	1.772	−5.12325	N/A
MACC1	1.772	3.272925	3.57
UCA1	1.155	2.058481	N/A
FGFR2[a,b]	0.892	−2.72455	−3.28
ADAM12[a,b]	0.871	4.646741	2.33
MUC13	0.704	−2.15979	N/A
MUC4[a,b,c]	0.66	−4.18466	−5.06
SLC7A11[a]	0.637	3.051363	2.72
ICAM1	0.532	2.655218	N/A
EXO1	0.405	2.268313	2.05
FAM46C[a]	0.271	−1.7528	−2.03
HTRA1	0.231	2.301062	N/A
SLC35D1[a]	0.217	−2.0238	−2.06

N/A = No annotation for these genes were found in the analysis program.

[a] Genes found to be differentially expressed in SAM/LIMMA analyses. CLDN7 and LOX genes were additional driver genes that were differentially expressed.

[b] Genes found to have high splice index values.

[c] Genes found to be eligible as biomarkers for colorectal cancer.

doi:10.1371/journal.pone.0110134.t001

explaining different possible mechanisms by which SCNA affected driver genes can exert their functional effect. A subset of driver genes were found to show gene level changes in expression. Most of these genes also indicated the exon level changes resulting in the formation of different isoforms. Network of GISTIC genes showed a clear shift in the transcription factors (TF) regulation. ATP8B1 gene was found to have novel association with colorectal cancer at cytogenetic, gene and exon level.

Materials and Methods

The study is approved by the ethics committee and Institutional Review Board (IRB) of King Abdullah International medical Research Center after due review process of the ethical aspects of the proposal. The necessary procedural and ethical consent forms were signed by each patient prior to sample collection.

Sample collection and RNA extraction

Sample collection was done as described previously [13]. The type and stage of all patient samples are provided in Table S1. The study was approved by institutional review board after the due process. Patients were consented and the records maintained in an approved manner. RNA was extracted from the same piece of

Table 2. Differentially regulated genes found to have incoherent expression levels and genomic changes.

Gene	Genomic level	Gene level (AA, EC)	Network level
ABP1	Gain 53.33, Loss 13.33	−2.55, −2.78	TF in Normal = CTCFL TF Tumour = Unknown
AGR3	Gain 60, Loss 0	−5.12	TF in Normal =SMAD2 TF Tumour = Unknown
BCAS1	Gain 80, Loss 0	−4.5, −5	TF in Normal =SMAD4 TF Tumour = Unknown
HTRA1	Gain 6.7, Loss 26.7	2.3	TF in Normal =CTCFL TF Tumour =AHR

AA = Fold change value as calculated by AltAnalyze program.
EC = Fold change value as calculated by Expression Console program.
TF = Transcription Factor. Unknown is the TF that is not found in the driver genes.

doi:10.1371/journal.pone.0110134.t002

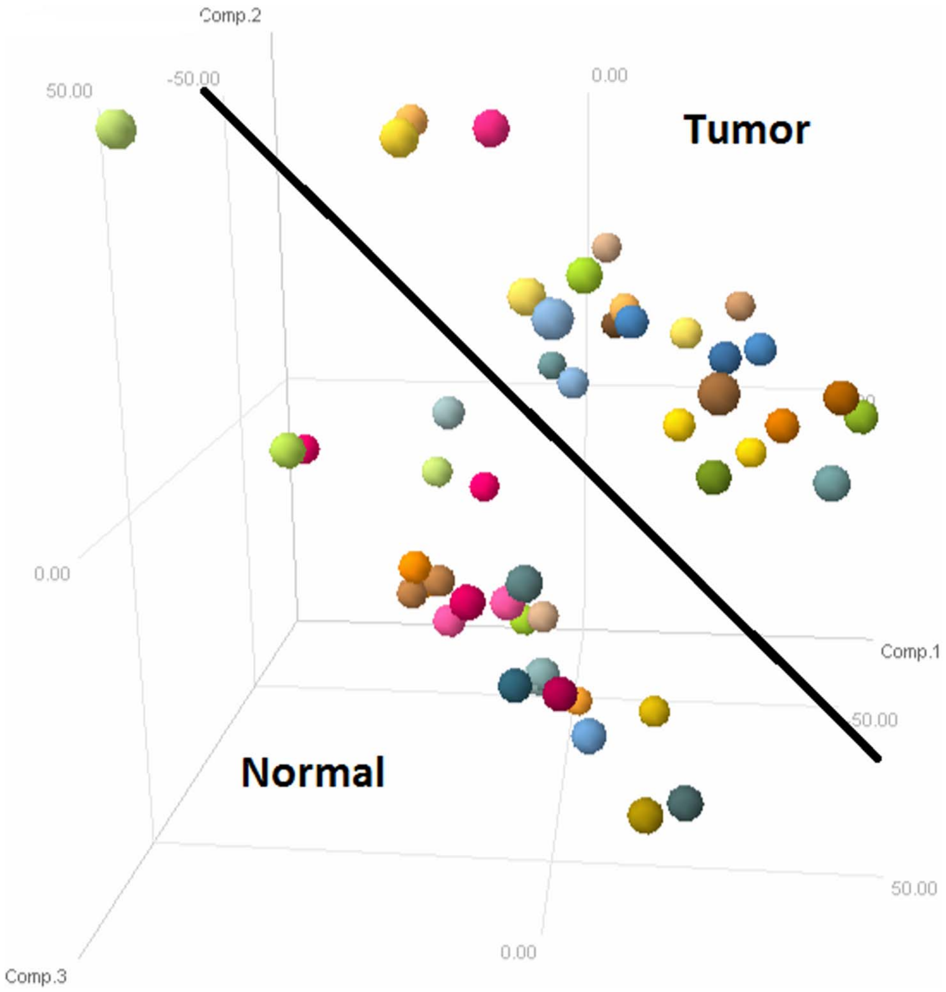

Figure 3. Principal Component Analysis of Exon array data from 32 patients. 60 samples from 32 patients were subjected to PCA and the outliers were removed. 4 normal and 7 tumour samples were removed from the final analysis.
doi:10.1371/journal.pone.0110134.g003

tissue that was used to extract DNA for cytogenetic studies in a single workflow. Maceherey Nagel trio prep kit (Germany) was used to extract DNA and RNA in the same protocol. Quality and quantity was checked using Nanodrop (Thermo Fisher Scientific, USA).

Exon microarray

GeneChip Human Exon 1.0 ST Arrays along with WT Terminal Labelling and Controls Kit and Hybridization, Wash, and Stain Kit were obtained from Affymetrix USA. Ambion WT Expression Kit was obtained by Ambion, USA. 31 Tumour and 29 normal samples from 32 patients were processed. The data was extracted using Expression Console software from Affymetrix, USA. Quality control was carried out using Principal Component Analysis (PCA) and Integromics biomarker suite (TIBCO spotfire). All the data is deposited in GEO database with an accession number GSE50421.

Data Analysis

Before carrying out any gene/exon level analysis, principal component analysis (PCA) was done to identify outliers. Data from 4 normal samples and 7 tumour samples was subsequently removed.

Data analysis specifically with driver genes

Gene level analysis. For checking expression levels of 144 driver gene list generated by GISTIC analysis (13), we employed two different softwares – Expression console and AltAnalyze [14]. Two different softwares were used to confirm our results using independent methods. Signal estimates were derived from the CEL files of 60 samples (29 normal and 31 tumour) using Robust Multi-Array Average (RMA) for normalizing the data. The core exon-level probe sets were used to summarize the gene expression levels.

The same list of 144 driver genes with expression values calculated using 'Altanalyze' was used for inference based (GENIE3) pathway/network analysis.

Exon level expression analysis. Altanalyze program was used to evaluate alternative splicing in driver genes. The raw data was filtered to remove probe sets that were considered to be non-expressed. A splicing score for filtered exons was calculated using splicing index method and exon/intron/splicing annotation were

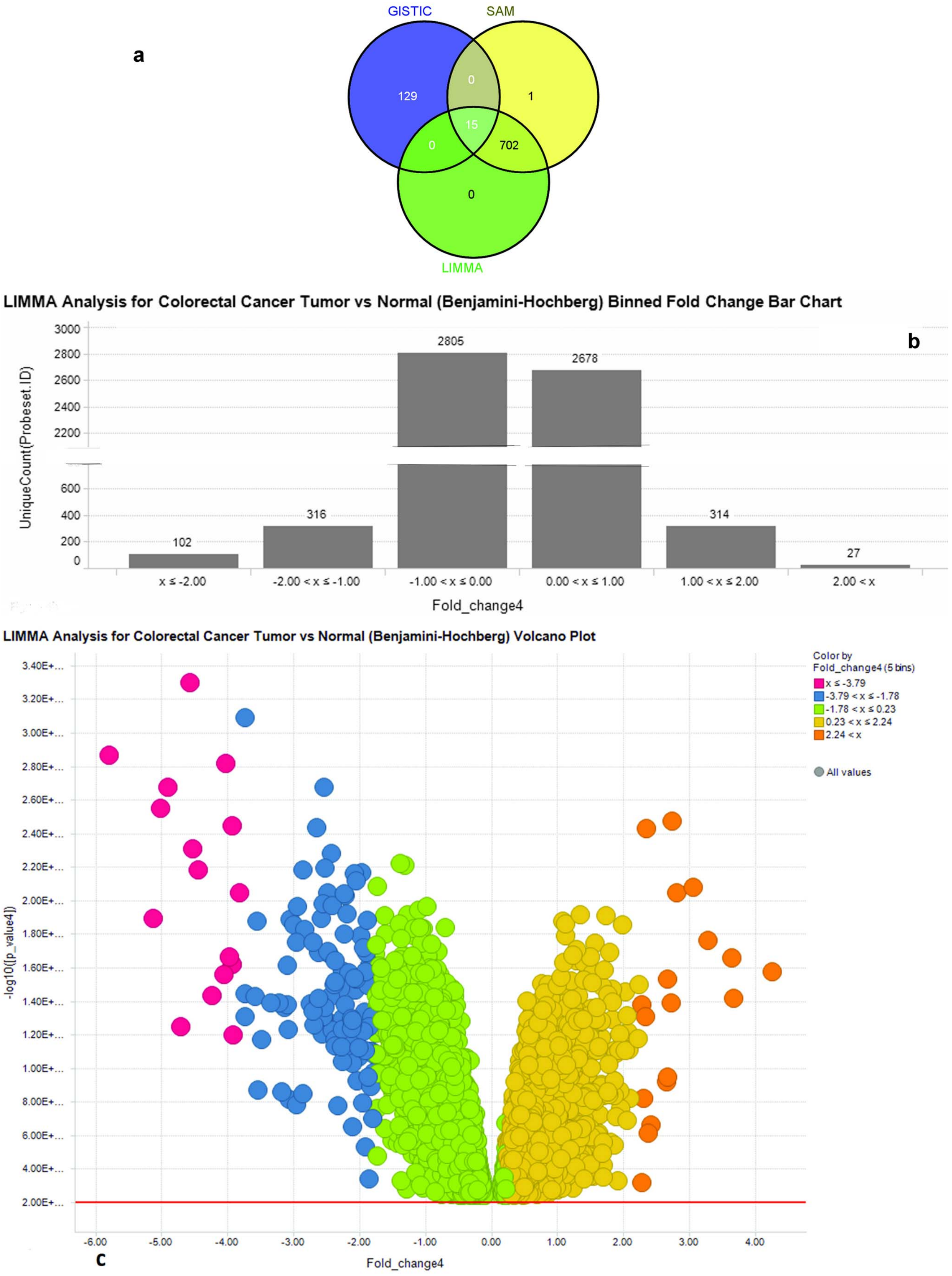
a
GISTIC
SAM
LIMMA
129
0
1
15
0
702
0
LIMMA Analysis for Colorectal Cancer Tumor vs Normal (Benjamini-Hochberg) Binned Fold Change Bar Chart
b
UniqueCount(Probeset.ID)
3000
2800
2600
2400
2200
600
400
200
0
102
316
2805
2678
314
27
x ≤ -2.00
-2.00 < x ≤ -1.00
-1.00 < x ≤ 0.00
0.00 < x ≤ 1.00
1.00 < x ≤ 2.00
2.00 < x
Fold_change4
LIMMA Analysis for Colorectal Cancer Tumor vs Normal (Benjamini-Hochberg) Volcano Plot
-log10([p_value4])
3.40E+...
3.20E+...
3.00E+...
2.80E+...
2.60E+...
2.40E+...
2.20E+...
2.00E+...
1.80E+...
1.60E+...
1.40E+...
1.20E+...
1.00E+...
8.00E+...
6.00E+...
4.00E+...
2.00E+...
-6.00
-5.00
-4.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
4.00
Fold_change4
c
Color by
Fold_change4 (5 bins)
x ≤ -3.79
-3.79 < x ≤ -1.78
-1.78 < x ≤ 0.23
0.23 < x ≤ 2.24
2.24 < x
All values

Figure 4. Differential Expression analysis of genes yielded genes common with GISTIC analysis. (a) Venn diagram of common genes among GISTIC, SAM and LIMMA analyses. All genes were annotated and compared using IPA 'compare' function. 43 genes from Integromics analyses were not mapped by IPA.15 genes are common among all three analyses. (b) Binned fold change bar chart of LIMMA analysis. Total 6242 genes were having adjusted p-value <0.01 of which 759 showed significant fold change (<-1 or>1). (c) Volcano plot showing highly significant genes (pink = downregulated, orange = upregulated) in terms of p- value and fold change.
doi:10.1371/journal.pone.0110134.g004

assigned to these results. A splice index p-value cut-off of <0.05 was used to filter alternative exon results. AltExonViewer – a component of Altanalyze and DomainGraph - a Cytoscape plugin were used to visualize splice index values and alternatively spliced exons. The splicing index (SI) value was calculated as described in [15]. Briefly, SI is $\log_2$ ratio of normalized intensities of tumour and normal samples. In our analysis 'sample 1' in the numerator was normal and 'sample 2' in the denominator was tumour.

Causal network analysis. For network analysis, knowledge and inference based approaches were used. For inference based approaches, GENIE3 [16,17] was used to generate gene regulatory networks for tumour and normal samples. To generate the network, driver genes were classified as TF genes and target genes. Although 1000 interactions were inferred for each of the group, an interaction score>0.1 was chosen as a cut-off value. Using this information the regulatory networks were independently inferred for tumour and normal samples using Cytoscape.

Data analysis with entire probeset

Integromics Biomarker Discovery Suite. In order to find differentially expressed genes in our dataset, without any prior bias, data from CEL files was analyzed using the Integromics software (TIBCO Spotfire, USA) pipeline for affymetrix exon 1.0 ST arrays. Quantile normalization was done after removing the outliers using PCA. Both Significance Analysis of Microarrays (SAM) and Linear Models for MicroArray data (LIMMA) analyses was carried out with a cut-off of 0.01 for adjusted p-value and fold change of>1 or <-1. Two different yet complimentary methodologies were used to make our results more confident. Gene ontology enrichment was carried out on a list of 760 differentially expressed genes obtained from LIMMA analysis.

Ingenuity pathway Analysis. List of 760 genes from SAM/LIMMA analysis done using Integromics was used to perform 'core' and 'biomarker' analyses. Core analysis was done as described before [13]. For biomarker analysis following filters were used: Consider only molecules where (species = Human) AND (tissues/cell lines = KM-12 OR HCT-116 OR RKO OR Colon Cancer Cell Lines not otherwise specified OR COLO205 OR HT29 OR HCC-2998 OR HCT-15 OR SW-480 OR Other Colon Cancer Cell Lines OR Tissues and Primary Cells not otherwise specified OR SW-620) AND (diseases = Cancer) AND ((biomarker applications = All Biomarker Applications) AND (biomarker diseases = colon cancer OR colon carcinoma OR colon neoplasm OR colorectal adenoma OR colorectal cancer OR colorectal carcinoma)).

Results

The analysis strategy leading to following results is illustrated in Figure 1a&b.

A small subset of genes identified by GISTIC show a significant change in expression level. We studied expression patterns of driver genes at gene level using AltAnalyze and Expression Console softwares. These analyses produced complimentary results and yielded a list of 20 genes that were found to have a significant fold change of greater than 2 and a p-value $<$ 0.01 [Table 1]. 9 genes experienced a down regulation. BCAS1 with highest GISTIC score of 5.323 was among the most significantly downregulated genes. 11 genes showed an upregulation with IL6 and INHBA showing the highest fold change [Figure 2a (AltAnalyze), 2b (Expression Console)]. Fold change values of all 144 driver genes are given in Table S2.

Three significantly down regulated genes (BCAS1, ABP1 and AGR3) were found in amplified regions of the genome whereas upregulated HTRA1 was found mostly in regions of loss. These genes experienced a change in their transcription factor in tumour and normal samples [Table 2].

Differential expression analysis using tumour-normal paired sample data of exon arrays from 32 patients was carried out. After removing outliers using PCA [Figure 3], 25 normal and 24 tumour samples were found suitable for further analyses. We employed non-parametric (SAM) and parametric (LIMMA) methods and found complimentary results. 6242 genes were differentially expressed with adjusted p-value of <0.01. 760 genes were found to be differentially regulated (fold change>1or <-1) of which 15 genes were common with driver genes from GISTIC analysis [Figure 4a-c and Figure S1]. BCAS1, AURKA, ATP8B1, IL6 and INHBA were the differentially regulated genes found to be among the top scorers in list of driver genes.

Significant changes in isoform expression is exhibited by genes identified by GISTIC analysis. Exon level analysis of 144 driver genes was carried out. 29 exons belonging to 13 genes were shown to have significant changes in isoform expression as reflected in their splice index scores [Table 3]. While exons E25-1 of MUC4 and E3-2 of PTP4A3 showed high negative SI values, exons E2-2 of IL6 and E21-2 of ADAM12 showed high positive SI values. Negative SI values indicate exons are enriched in tumour samples and are skipped or repressed in normal samples and vice versa for positive SI values. For MUC4 gene, exons E25-1, E2-2, E2-1, I4-6 and I3-6 recorded negative SI values and exon E8-1 recorded a positive SI value. Exons E2-2 and E4-3 of IL6 recorded positive SI values. [Figure 5 ai-ii & bi-ii and figure S2]. Exons in MYLK and ANK3 showed significant change in splicing pattern but did not show a significant fold change in gene expression value. Microarray Detection of Alternative Splicing (MiDAS) p-value range of 0.01–0.04 was observed in the filtered results.

Causal Network analysis displays switch in transcription factors in tumor samples. Causal Network analysis displays switch in transcription factors in tumour samples. The most significant influence of TF genes in tumour and normal samples was reflected in the alteration in number of directed edges [Table 4]. CHAF1A, AHR, PRPF4B, ZNF200, SMAD2, RUVBL1, SMAD4, TSHZ1, CTCFL and CEBPE were among the top influencers. The regulatory hierarchy between these groups changed with respect to TFs. While RUVBL1 transformed into a master regulator in tumours, TSHZ1 has lost its capacity to master regulate other genes in tumours [Figure 6a]. Significant rearrangement in modularity is also observed between these two groups. More target genes function as modules in tumours as

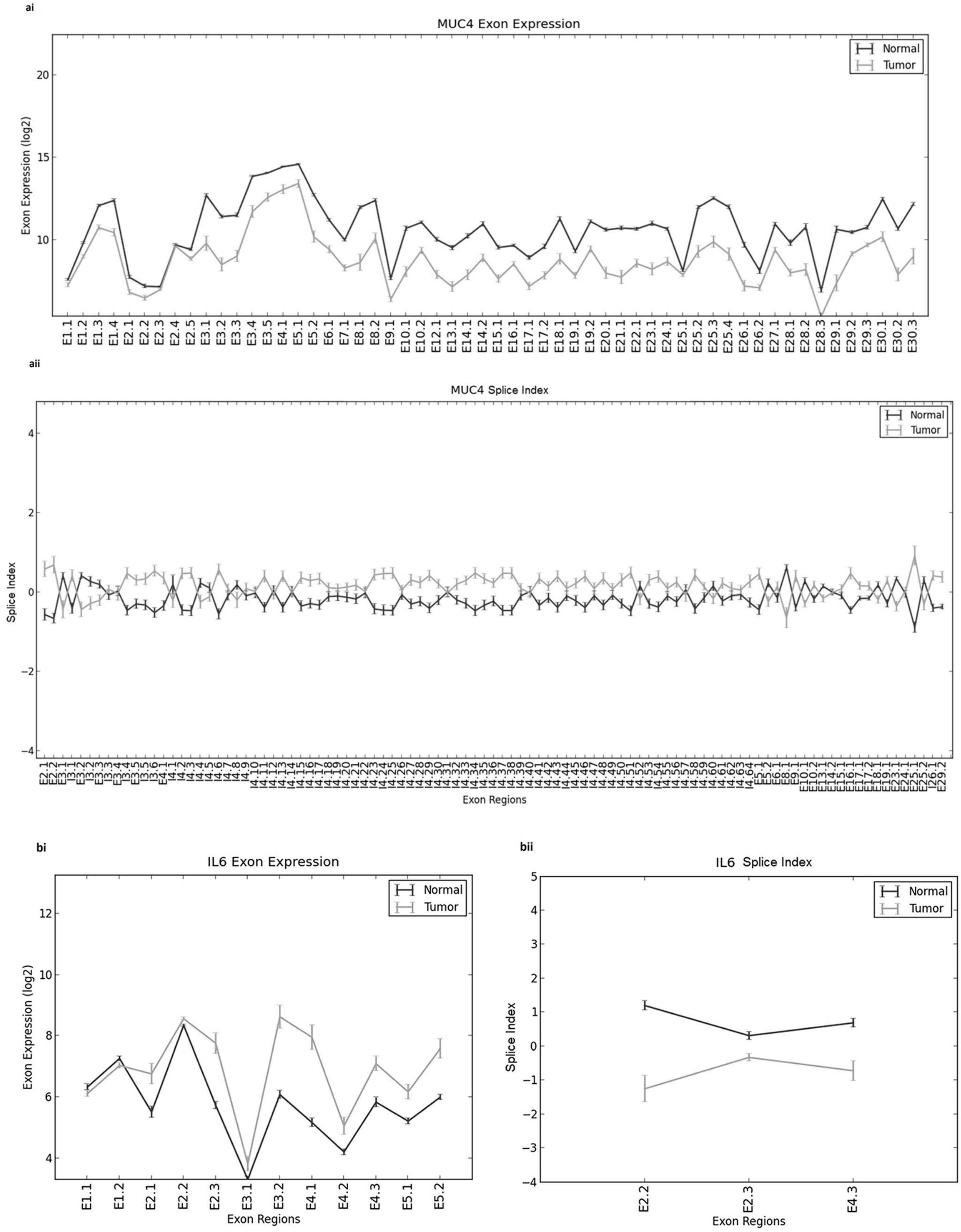
ai
MUC4 Exon Expression
Normal
Tumor
Exon Expression (log2)
20
15
10
E1.1 E1.2 E1.3 E1.4 E2.1 E2.2 E2.3 E2.4 E2.5 E3.1 E3.2 E3.3 E3.4 E3.5 E4.1 E5.1 E5.2 E6.1 E7.1 E8.1 E8.2 E9.1 E10.1 E10.2 E12.1 E13.1 E14.1 E14.2 E15.1 E16.1 E17.1 E17.2 E18.1 E19.1 E19.2 E20.1 E21.1 E22.1 E23.1 E24.1 E25.1 E25.2 E25.3 E25.4 E26.1 E26.2 E27.1 E28.1 E28.2 E28.3 E29.1 E29.2 E29.3 E30.1 E30.2 E30.3
aii
MUC4 Splice Index
Normal
Tumor
Splice Index
4
2
0
−2
−4
E2.1 E2.2 E3.1 I3.1 E3.2 I3.2 E3.3 I3.3 E3.4 I3.4 E3.5 I3.5 I3.6 E4.1 I4.1 I4.2 I4.3 I4.4 I4.5 I4.6 I4.7 I4.8 I4.9 I4.10 I4.11 I4.12 I4.13 I4.14 I4.15 I4.16 I4.17 I4.18 I4.19 I4.20 I4.21 I4.22 I4.23 I4.24 I4.25 I4.26 I4.27 I4.28 I4.29 I4.30 I4.31 I4.32 I4.33 I4.34 I4.35 I4.36 I4.37 I4.38 I4.39 I4.40 I4.41 I4.42 I4.43 I4.44 I4.45 I4.46 I4.47 I4.48 I4.49 I4.50 I4.51 I4.52 I4.53 I4.54 I4.55 I4.56 I4.57 I4.58 I4.59 I4.60 I4.61 I4.62 I4.63 I4.64 E5.1 E5.2 E6.1 E8.1 E9.1 E10.1 E10.2 E13.1 E14.2 E15.1 E16.1 E17.1 E17.2 E18.1 E19.1 E23.1 E24.1 E25.1 E25.2 I26.1 E29.2
Exon Regions
bi
IL6 Exon Expression
Normal
Tumor
Exon Expression (log2)
12
10
8
6
4
E1.1 E1.2 E2.1 E2.2 E2.3 E3.1 E3.2 E4.1 E4.2 E4.3 E5.1 E5.2
Exon Regions
bii
IL6 Splice Index
Normal
Tumor
Splice Index
5
4
3
2
1
0
−1
−2
−3
−4
E2.2 E2.3 E4.3
Exon Regions

Figure 5. Differential Expression at exon level was observed in thirteen GISTIC genes. Exon expression (i) and splice index (ii) values were mapped for both tumour and normal samples for twenty nine exons affecting thirteen genes. Exon 25-1 of MUC4 gene (a) shows highest negative splice index value (a ii) whereas exon 2-2 of IL6 (b) showed highest value of 2.44 (b ii). Exon expression and splice index values for rest 27 exons are provided as supplementary figure 1.
doi:10.1371/journal.pone.0110134.g005

observed in modules regulated by AHR, CHAF1A and PRPF4B. Further, there is significant regulatory crosstalk among the genes in the normal group [Figure 6b]. While the tumours have lost the scale-free properties in its regulatory interaction, normal group exhibit approximate scale-free out degree distributions, signifying the potential of TF to regulate host of target genes. The tumours do not show this type of regulation which indicate unidirectional feed forward regulatory mode.

Functional analysis of differentially expressed genes confirms their role in colorectal cancer and reveal important pathways and biomarkers. Gene Ontology enrichment of 760 differentially expressed genes obtained from SAM/LIMMA analyses showed cell division, mitosis and cell adhesion to be the most significant biological processes affected [figure S3]. Ingenuity pathway analysis of 760 genes showed cancer and gastrointestinal disease among the top functions followed by cellular movement growth and proliferation [Figure 7a]. NF-kB signalling, cell cycle G_2/M DNA damage checkpoint regulation, colorectal cancer metastasis were among the top scoring pathways followed by agranulocyte adhesion [figure 7b]. TGFB1 was among the top upstream regulators. Network analysis suggests MYC, MMP and IL6 genes as important nodes [Fig 7c-e]. Biomarker analysis reveals 28 molecules relevant for colorectal cancer [Table 5]. INHBA, CLDN7 and MUC4 were eligible as biomarkers and are common with GISTIC, SAM and LIMMA analyses results.

Table 3. Significant changes in exon level expression of driver genes.

Probeset	GeneID	Exon ID	Regulation Call	SI	SI p-value	MiDAS p-value
2712254	MUC4	E25-1	Upregulated	−1.80873	2.79E-08	0.021609
3118824	PTP4A3	E3-2	Upregulated	−1.61711	9.95E-12	0.012046
2712381	MUC4	E2-2	Upregulated	−1.3477	7.19E-07	0.026458
3310161	FGFR2	E11-1	Upregulated	−1.29699	4.55E-08	0.023093
3310118	FGFR2	E29-1	Upregulated	−1.27723	1.86E-09	0.01652
3910391	BCAS1	E4-3	Upregulated	−1.25187	3.35E-07	0.021612
3031671	ABP1	E11-1	Upregulated	−1.24918	1.22E-10	0.014292
3310159	FGFR2	E12-1	Upregulated	−1.17574	2.50E-07	0.026414
2712382	MUC4	E2-1	Upregulated	−1.15152	1.33E-05	0.037457
3290897	ANK3	E60-1	Upregulated	−1.14825	1.19E-04	0.041986
2712354	MUC4	I4-6	Upregulated	−1.12743	3.39E-07	0.028327
2712361	MUC4	I3-6	Upregulated	−1.05374	8.25E-07	0.03115
3290925	ANK3	E53-5	Upregulated	−1.05177	2.60E-05	0.040567
3268341	HTRA1	E1-8	Downregulated	1.007325	5.53E-10	0.016611
3310147	FGFR2	E17-1	Downregulated	1.086154	2.78E-06	0.031111
3268338	HTRA1	E1-5	Downregulated	1.086203	3.38E-11	0.013645
2388253	EXO1	E18-1	Downregulated	1.08806	3.11E-05	0.044937
3311901	ADAM12	E10-1	Downregulated	1.089706	1.70E-09	0.018986
3910807	AURKA	E1-2	Downregulated	1.105387	2.78E-07	0.026314
3820445	ICAM1	E3-3	Downregulated	1.222886	3.84E-07	0.027735
2712282	MUC4	E8-1	Downregulated	1.267574	1.63E-05	0.039593
2692532	MYLK	E13-1	Downregulated	1.273764	4.27E-07	0.030277
3268334	HTRA1	E1-1	Downregulated	1.288045	1.63E-12	0.011119
3310142	FGFR2	E20-1	Downregulated	1.403702	1.07E-09	0.015743
3268335	HTRA1	E1-2	Downregulated	1.404729	1.43E-12	0.011016
2992599	IL6	E4-3	Downregulated	1.406659	4.60E-05	0.044146
3311894	ADAM12	E11-1	Downregulated	1.420516	1.99E-10	0.016023
3311853	ADAM12	E21-2	Downregulated	2.016734	1.96E-09	0.018149
2992593	IL6	E2-2	Downregulated	2.448861	1.85E-07	0.027456

List of 29 exons corresponding to 13 genes are listed along with their respective splice index and MiDAS p values.
Probeset IDs are according to Affymetrix database
SI = Splice Index
MiDAS = Microarray Detection of Alternative Splicing
doi:10.1371/journal.pone.0110134.t003

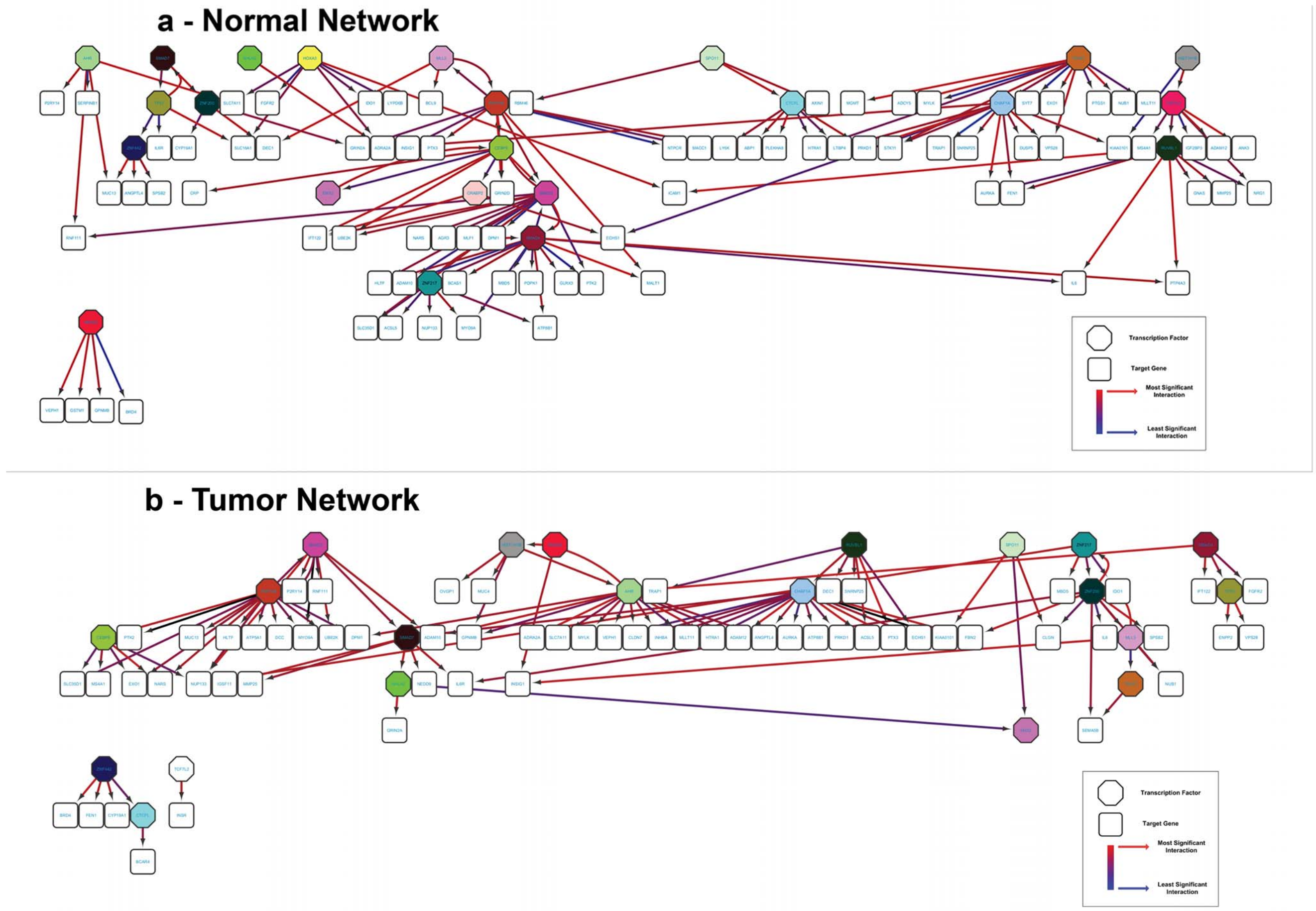
a - Normal Network
Transcription Factor
Target Gene
Most Significant Interaction
Least Significant Interaction
b - Tumor Network
Transcription Factor
Target Gene
Most Significant Interaction
Least Significant Interaction

Figure 6. Gene Regulatory Network Inference diagram for Tumor and Normal Groups. A hierarchical network topology is used to visualize the degrees of interaction between transcription factor genes and target genes. (a) The inferred network for tumour group showing RUVBL1 as master regulator. (b) The inferred network for normal group showing TSHZ1as master regulator.
doi:10.1371/journal.pone.0110134.g006

Discussion

In this study we attempt to understand the transcription level changes in driver genes affected by SCNAs in colorectal cancer. We queried these genes from three perspectives using gene/exon/network level analysis tools. Our integrated analysis at genomic/transcriptomic levels resulted in finding genes of high priority that can be experimentally studied to establish their role in colorectal cancer. Functional significance of differentially expressed genes confirmed the outcome of our analyses.

Due to the unavailability of high resolution cytogenetic arrays large size of chromosomal regions were implicated in causing colorectal cancer through copy number changes. The commercial SNP genotyping arrays focus on variants that are present in 5% or more of the population and feature a limited number of CNV probes. Therefore, sub microscopic structural variants are poorly captured by available SNP genotyping arrays that were designed to evaluate SNPs. The recent introduction of the Affymetrix CytoScan HD Array (CNV-targeted array), which is based on the validated Genome-Wide Human SNP Array 6.0 and contains more than 2.6 million markers for copy number variants and approximately 750,000 SNPs, has enabled the detection of copy number aberrations with high resolution across the genome [18]. Data from this platform was used to obtain the list of driver genes which could thus be considered to be most precise and accurate. Earlier, results from different groups often lead to a high level of discordance with an overlap of <5% in some cases. GISTIC analysis was able to address this issue and differentiate between passenger and driver mutations with a high level of precision and accuracy [6]. The correlation between expression and CN data is very complex [19] and is very much affected by the type of platform used for generating the data as well as the analysis strategies. Our analysis strategy aimed at reducing the confounding factors. We extracted DNA/RNA from the same piece of tissue in a single protocol [20]. We used matched paired tumour-normal control which is arguably the best way to do a comparative study [21]. Several approaches have been employed for cancer gene prioritization by integrative analysis [22,23]. We employed a modified two-step approach by filtering the cancer genes using GISTIC and then carried out exon expression analysis.

Gene Level. The effect of chromosomal changes is not always direct. Global amplification at genomic level would result into higher level expression of selected genes [9]. Our gene level analysis showed only 20 of 144 driver genes to experience significant change at transcription level. 5 of these genes were

Table 4. Transcription factor genes showing significant change in their effect as represented by the change in number of outbound and inbound edges in tumour and normal samples.

Transcription Factor	In bound edges (Normal, Tumour)	Out bound edges (Normal, Tumour)
CHAF1A	***1,1***	***16,12***
AHR	***0,2***	***5,14***
PRPF4B	***2,1***	***11,14***
ZNF200	***1,2***	***3,9***
SMAD2	***3,1***	***13,8***
RUVBL1	***1,0***	***8,7***
CEBPE	***2,1***	***7,5***
SMAD7	1,2	2,5
HIST1H1B	0,1	2,4
MLL3	1,1	3,4
SMAD4	***2,0***	***14,4***
ZNF217	1,1	6,4
ZNF442	1,0	3,4
SPO11	0,0	3,3
NHLH2	0,1	1,2
TP53	1,1	4,2
CTCFL	***1,1***	***6,1***
SMAD3	0,0	4,1
TCF7L2	1,0	6,1
TSHZ1	***0,1***	***12,1***

doi:10.1371/journal.pone.0110134.t004

listed among the top scoring driver genes. Our results correlate focal amplifications with increased expression levels and have been reported earlier using array CGH (aCGH) for FGFR2, GNAS and AURKA genes [24]. With an average amplicon size of 4.56 Mb, it could be misleading to report all affected genes/non-coding regions to be associated with CRC. FAM46C, EGFR2 and IL6 genes from our analysis have also been listed in the updated cancer gene census [5,25]. BCAS1 gene that scored the highest in the GISTIC analysis was earlier reported to undergo alternative splicing and downregulation [26] which is consistent with our results. Upregulation of human SLC7A11 mRNA in stromal fibroblast cells from liver metastases is associated with metastatic colorectal cancer in human [27]. PTP4A3 reported to be significantly upregulated in this study has recently been implicated in colon tumorigenesis [28]. ATP8B1 is the gene that has not yet been reported to have any association with colorectal cancer and would be an important molecule for further studies.

Our differential expression analysis of tumour and normal exon array datasets using two independent tools (SAM and LIMMA) showed an overlap of 15 genes with the driver genes. Analysis of> 44,000 probes resulting in differentially expressed genes provide an unbiased approach and lends further confidence in the list of overlapping genes. Both LIMMA and SAM approaches generated same results at the functional level as evidenced by Ingenuity Pathway Analysis (IPA).

ABP1, AGR3 and BCAS1 showed downregulation despite amplification at the genome level. This is supported by earlier studies for BCAS1in MCF7 cell line [29]. We observed that the influence of SMAD4 transcription factor was lost in tumour cells and may be responsible for this observation. Similar changes in transcription factors was observed for two other genes indicating the switching behaviour as discussed below.

Exon Level. Exon level analysis for measuring gene expression changes is challenging but rewarding. Even the improved algorithms have limitations in providing absolute quantification of the transcript levels. Earlier methods have found exon level data to be more informative about the nature and level of transcripts [30]. Now with the knowledge that more than 90% of all genes undergo alternative splicing to produce more than one transcript for a gene, the potential of exon level data is being realized more than ever [31–33]. Gene centric approach for carrying out integrated analysis using aCGH and exon array data has yielded more of confirmatory results and lack the use of full potential of whole genome arrays [34]. FGFR2 gene was shown to be amplified and upregulated which is misleading due to the truncated form of FGFR2. Wt FGFR2 gene was not measured and hence could not be compared with our study [34].

Exon level studies in colorectal cancer have been few and carry several limitations in terms of data analysis. Many of these studies compared tumour and normal from different sources [35]. Our results show a subset of differentially expressed genes to experience change in splicing pattern. 29 exons belonging to 13 genes showed significant splice index (SI) and MiDAS p-values. Both the values are strong indicators to measure alternative splicing. MUC4 is very well known to undergo alternative splicing and cause cancer [36] but the alternative splicing of IL6 is novel in association with CRC. ADAM12 that scored a high SI value has been implicated earlier in lung cancer [37]. 5 genes (ACTN1, CALD1, SLC3A2,

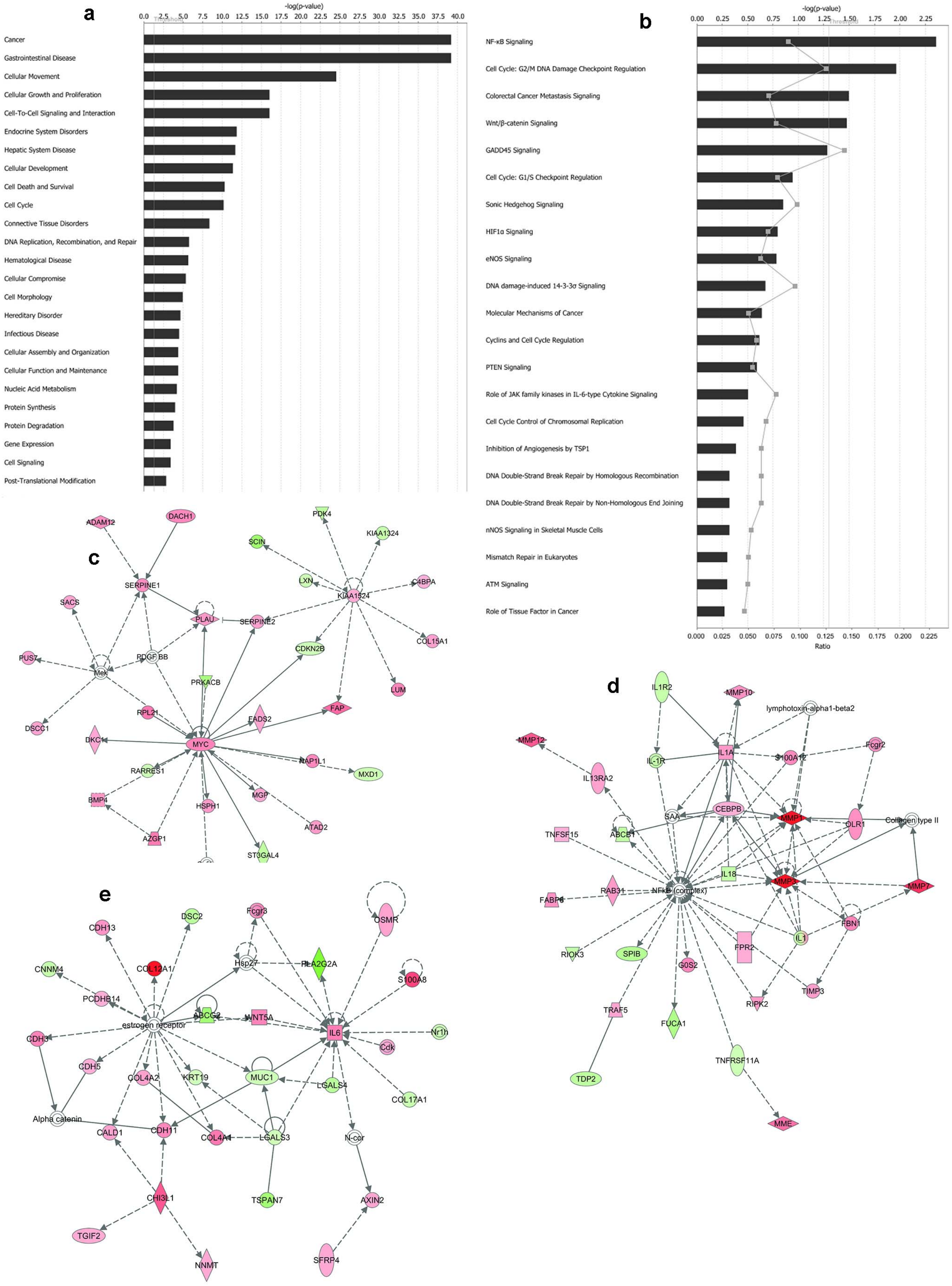
a
-log(p-value)
0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0 27.5 30.0 32.5 35.0 37.5 40.0
Threshold
Cancer
Gastrointestinal Disease
Cellular Movement
Cellular Growth and Proliferation
Cell-To-Cell Signaling and Interaction
Endocrine System Disorders
Hepatic System Disease
Cellular Development
Cell Death and Survival
Cell Cycle
Connective Tissue Disorders
DNA Replication, Recombination, and Repair
Hematological Disease
Cellular Compromise
Cell Morphology
Hereditary Disorder
Infectious Disease
Cellular Assembly and Organization
Cellular Function and Maintenance
Nucleic Acid Metabolism
Protein Synthesis
Protein Degradation
Gene Expression
Cell Signaling
Post-Translational Modification
b
-log(p-value)
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.25
Threshold
NF-κB Signaling
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
Colorectal Cancer Metastasis Signaling
Wnt/β-catenin Signaling
GADD45 Signaling
Cell Cycle: G1/S Checkpoint Regulation
Sonic Hedgehog Signaling
HIF1α Signaling
eNOS Signaling
DNA damage-induced 14-3-3σ Signaling
Molecular Mechanisms of Cancer
Cyclins and Cell Cycle Regulation
PTEN Signaling
Role of JAK family kinases in IL-6-type Cytokine Signaling
Cell Cycle Control of Chromosomal Replication
Inhibition of Angiogenesis by TSP1
DNA Double-Strand Break Repair by Homologous Recombination
DNA Double-Strand Break Repair by Non-Homologous End Joining
nNOS Signaling in Skeletal Muscle Cells
Mismatch Repair in Eukaryotes
ATM Signaling
Role of Tissue Factor in Cancer
0.000 0.025 0.050 0.075 0.100 0.125 0.150 0.175 0.200 0.225
Ratio
c
ADAM12
DACH1
PDK4
KIAA1324
SCIN
LXN
C4BPA
SERPINE1
KIAA1524
SACS
PLAU
SERPINE2
COL15A1
PUS7
PDGF BB
Mek
CDKN2B
PRKACB
LUM
RPL21
FADS2
FAP
DSCC1
DKC1
MYC
NAP1L1
MXD1
RARRES1
BMP4
MGP
HSPH1
ATAD2
AZGP1
ST3GAL4
d
IL1R2
MMP10
lymphotoxin-alpha1-beta2
MMP12
IL1A
S100A12
Fcgr2
IL-1R
IL13RA2
SAA
CEBPB
MMP1
OLR1
Collagen type II
TNFSF15
ABCB1
IL18
MMP3
MMP7
FABP6
RAB31
NFkB (complex)
FBN1
IL1
RIOK3
SPIB
G0S2
FPR2
TIMP3
RIPK2
TRAF5
FUCA1
TDP2
TNFRSF11A
MME
e
DSC2
Fcgr3
OSMR
CDH13
CNNM4
COL12A1
Hsp27
PLA2G2A
S100A8
PCDHB14
ABCG2
WNT5A
estrogen receptor
IL6
Nr1h
CDH3
Cdk
CDH5
COL4A2
KRT19
MUC1
LGALS4
COL17A1
Alpha catenin
CALD1
CDH11
COL4A1
LGALS3
N-cor
CHI3L1
TSPAN7
AXIN2
TGIF2
NNMT
SFRP4

Figure 7. Core Analysis of Differentially Expressed Genes using IPA. Core analysis using IPA was carried out using set of 760 genes that were differentially expressed in tumour samples. Important biological functions (a) pathways (b) and networks (c-e) were revealed by this analysis.
doi:10.1371/journal.pone.0110134.g007

CTTN and FN1) reported earlier as differentially spliced [38] have been found in the our study as well which used a different analysis strategy.

ANK3 is known to use alternative transcription start sites in colorectal cancer [8] and could also be the mechanism for other gene MYLK for which we did not see a significant change in gene level expression.

Network Level. Pathways analysis has been used to measure the relevance of the genes affected by CNAs by creating networks among them [1]. However, these networks are limited in their useful interpretation owing to the absence of directionality. Our causal network analysis provides more useful information on the genes involved in these networks. We observed a significant difference in the number of target genes between tumour and normal. In case of genes found in the amplified region but were downregulated we observed a loss of the transcription factor regulation, whereas in the upregulated gene found in the deleted region there was a switch in the transcription factor. From the list of driver genes we chose the transcription factors and studied their change in behaviour in tumour samples. AHR has been established as a tumour suppressor gene in colon and other cancers [39]. This study further explains the enhanced role of AHR by the increased number of outbound (target) genes in tumour. Our study provides evidence that TSHZ1 loses its role as master regulator in normal cells while RUVBL1 assumes that role. These provide interesting opportunities for mechanistic studies of network/pathways affected in CRC. It has been envisaged through integrated studies that many different genomic alterations potentially dys-regulate the same pathways in complex diseases [40]. Further studies of the regulatory level changes in this study will be able to establish this concept in CRC.

Functional role of the differentially expressed genes and the identification of MYC, MMP and IL6 as important nodes in the affected networks provide leads that need to be validated to

Table 5. Biomarker molecules among the differentially expressed genes.

Symbol	Fold Change	p-value	Location	Type(s)
ANGPT2	2.055	6.89E-14	Extracellular Space	growth factor
CD34	1.104	4.27E-10	Plasma Membrane	other
CD44	1.256	9.28E-14	Plasma Membrane	enzyme
CDH13	1.338	4.13E-11	Plasma Membrane	other
CLDN7	−1.030	1.81E-09	Plasma Membrane	other
CXCL10	1.073	8.50E-03	Extracellular Space	cytokine
EREG	1.352	2.39E-05	Extracellular Space	growth factor
GDF15	1.191	6.33E-06	Extracellular Space	growth factor
IGF2	1.398	3.85E-04	Extracellular Space	growth factor
IL8	3.641	2.55E-17	Extracellular Space	cytokine
INHBA	2.225	1.71E-12	Extracellular Space	growth factor
KRT20	−2.032	4.90E-12	Cytoplasm	other
MET	1.049	1.87E-07	Plasma Membrane	kinase
MKI67	1.015	1.58E-10	Nucleus	other
MMP2	1.764	2.31E-09	Extracellular Space	peptidase
MUC1	−1.070	1.65E-05	Plasma Membrane	transcription regulator
MUC4	−2.264	3.71E-11	Extracellular Space	growth factor
OSMR	1.166	4.95E-07	Plasma Membrane	transmembrane receptor
PDCD4	−1.572	9.64E-13	Nucleus	other
PDGFRB	1.090	1.47E-10	Plasma Membrane	kinase
PDPN	1.136	5.95E-11	Plasma Membrane	transporter
PTGS2	2.303	5.99E-09	Cytoplasm	enzyme
PTP4A3	1.132	2.47E-12	Plasma Membrane	phosphatase
SERPINE1	1.418	4.29E-09	Extracellular Space	other
SPP1	2.648	6.37E-10	Extracellular Space	cytokine
TIMP1	1.938	9.22E-13	Extracellular Space	other
TNS4	1.486	3.15E-07	Cytoplasm	other
TSPAN8	−1.031	2.83E-09	Plasma Membrane	other

Fold change and p-values were calculated using Integromics biomarker suite.
doi:10.1371/journal.pone.0110134.t005

establish their association with CRC. Biomarkers, especially MUC4 will be an important molecule to study mechanistically and establish their use in clinical studies. This study provides wealth of analyzed data and an enriched list of genes that can serve as potential clues to understand the biology of colorectal cancer.

Supporting Information

Figure S1 Significance of Microarray analysis. SAM analysis was performed using Integromics biomarker discovery suites on all samples. The results were complimentary to LIMMA analysis as reflected in the number of differentially expressed genes.
(TIF)

Figure S2 Splice index and exon expression plots for remaining 11 genes. Splice index and exon expression values of all 13 genes that were found significant among the driver genes were plotted. Comparison of 'Normal' and 'Tumor' samples is depicted to observe the change in splice index as well the expression pattern at exon level.
(DOCX)

Figure S3 Biological processes enrichment chart for differentially expressed genes. This bar plot shows the differentially expressed genes (as obtained from Integromics) are enriched in three functions viz., cell division, mitosis and cell adhesion.
(PNG)

Table S1 Types and stages of all the patient samples used in the study.
(DOCX)

Table S2 Fold change values of all 144 GISTIC genes as calculated by AltAnalyze Program.
(TXT)

Acknowledgments

The help extended by Drs. Balwi and Hussein in providing facilities for conducting experiments and analyses is acknowledged.

Author Contributions

Conceived and designed the experiments: MAA. Performed the experiments: MAO AA GA. Analyzed the data: MAA SP. Contributed reagents/materials/analysis tools: ZAY IA. Wrote the paper: MAA.

References

1. Fan B, Dachrut S, Coral H, Yuen ST, Chu KM, et al. (2012) Integration of DNA copy number alterations and transcriptional expression analysis in human gastric cancer. PLoS One 7: e29824.
2. Lu T-P, Lai L-C, Tsai M-H, Chen P-C, Hsu C-P, et al. (2011) Integrated Analyses of Copy Number Variations and Gene Expression in Lung Adenocarcinoma. PLoS One 6: e24829.
3. Louhimo R, Lepikhova T, Monni O, Hautaniemi S (2012) Comparative analysis of algorithms for integration of copy number and expression data. Nat Methods 9: 351–355.
4. Yoshida T, Kobayashi T, Itoda M, Muto T, Miyaguchi K, et al. (2010) Clinical omics analysis of colorectal cancer incorporating copy number aberrations and gene expression data. Cancer Inform 9: 147–161.
5. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. Nat Rev Cancer 10: 59–64.
6. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A 104: 20007–20012.
7. Mermel C, Schumacher S, Hill B, Meyerson M, Beroukhim R, et al. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biology 12: R41.
8. Thorsen K, Mansilla F, Schepeler T, Øster B, Rasmussen MH, et al. (2011) Alternative Splicing of SLC39A14 in Colorectal Cancer is Regulated by the Wnt Pathway. Molecular & Cellular Proteomics 10.
9. Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, et al. (2002) Silence of chromosomal amplifications in colon cancer. Cancer Res 62: 1134–1138.
10. Wen Z, Liu ZP, Liu Z, Zhang Y, Chen L (2013) An integrated approach to identify causal network modules of complex diseases with application to colorectal cancer. J Am Med Inform Assoc 20: 659–667.
11. Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J (2006) Cancer: A Systems Biology disease. Biosystems 83: 81–90.
12. Madhamshettiwar P, Maetschke S, Davis M, Reverter A, Ragan M (2012) Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. Genome Med 4: 1–16.
13. Hisham ElDai SP, Al Qarni S, Al Rodayyan M, Mustafa S, Deeb A, et al. (2013) Novel Genes Associated with Colorectal Cancer Are Revealed By High Resolution Cytogenetic Analysis in a Patient Specific Manner. PLOS ONE 8: e76251.
14. Emig D, Salomonis N, Baumbach J, Lengauer T, Conklin BR, et al. (2010) AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. Nucleic Acids Res 38: W755–W762.
15. Srinivasan K, Shiue L, Hayes JD, Centers R, Fitzwater S, et al. (2005) Detection and measurement of alternative splicing using splicing-sensitive microarrays. Methods 37: 345–359.
16. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, et al. (2012) Wisdom of crowds for robust gene network inference. Nat Meth 9: 796–804.
17. Huynh-Thu VA, Irrthum A, Wehenkel L and Geurts P (2010) Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLoS ONE 5: e12776.
18. Chen W, Yuan L, Cai Y, Chen X, Chi Y, et al. (2013) Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiency. Genomics 102: 27–34.
19. Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, et al. (2006) Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 66: 2129–2137.
20. Mathieson W, Thomas GA (2013) Simultaneously extracting DNA, RNA, and protein using kits: is sample quantity or quality prejudiced? Anal Biochem 433: 10–18.
21. Mojica W and Hawthorn L (2010) Normal colon epithelium: a dataset for the analysis of gene expression and alternative splicing events in colon disease. BMC Genomics 11: 5.
22. Lahti L, Schafer M, Klein HU, Bicciato S, Dugas M (2013) Cancer gene prioritization by integrative analysis of mRNA expression and DNA copy number data: a comparative review. Brief Bioinform 14: 27–35.
23. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, et al. (2013) An integrative framework identifies alternative splicing events in colorectal cancer development. Mol Oncol.
24. Camps J, Nguyen QT, Padilla-Nash HM, Knutsen T, McNeil NE, et al. (2009) Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. Genes Chromosomes Cancer 48: 1002–1017.
25. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. Nat Rev Cancer 4: 177–183.
26. Correa RG, de Carvalho AF, Pinheiro NA, Simpson AJ, de Souza SJ (2000) NABC1 (BCAS1): alternative splicing and downregulation in colorectal tumors. Genomics 65: 299–302.
27. Nakagawa H, Liyanarachchi S, Davuluri RV, Auer H, Martin EW Jr, et al. (2004) Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles. Oncogene 23: 7366–7377.
28. Zimmerman MW, Homanics GE, Lazo JS (2013) Targeted Deletion of the Metastasis-Associated Phosphatase *Ptp4a3* (PRL-3) Suppresses Murine Colon Cancer. PLoS One 8: e58300.
29. Collins C, Rommens JM, Kowbel D, Godfrey T, Tanner M, et al. (1998) Positional cloning of ZNF217 and NABC1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. Proceedings of the National Academy of Sciences 95: 8703–8708.
30. Bemmo A, Benovoy D, Kwan T, Gaffney DJ, Jensen RV, et al. (2008) Gene expression and isoform variation analysis using Affymetrix Exon Arrays. BMC Genomics 9: 529.
31. Consiglio A, Carella M, De Caro G, Delle Foglie G, Giovannelli C, et al. (2012) BEAT: Bioinformatics Exon Array Tool to store, analyze and visualize Affymetrix GeneChip Human Exon Array data from disease experiments. BMC Bioinformatics 13 Suppl 4: S21.
32. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413–1415.
33. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470–476.
34. Furuta K, Arao T, Sakai K, Kimura H, Nagai T, et al. (2012) Integrated analysis of whole genome exon array and array-comparative genomic hybridization in gastric and colorectal cancer cells. Cancer Sci 103: 221–227.

35. Piepoli A, Palmieri O, Maglietta R, Panza A, Cattaneo E, et al. (2012) The expression of leucine-rich repeat gene family members in colorectal cancer. Exp Biol Med (Maywood) 237: 1123–1128.
36. Singh AP, Chaturvedi P, Batra SK (2007) Emerging roles of MUC4 in cancer: a novel target for diagnosis and therapy. Cancer Res 67: 433–436.
37. Xi L, Feber A, Gupta V, Wu M, Bergemann AD, et al. (2008) Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer. Nucleic Acids Res 36: 6535–6547.
38. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, et al. (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 7: 325.
39. Zudaire E, Cuesta N, Murty V, Woodson K, Adams L, et al. (2008) The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. The Journal of Clinical Investigation 118: 640–650.
40. Kim YA, Wuchty S, Przytycka TM (2011) Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput Biol 7: e1001095.