

# Large differences in global transcriptional regulatory programs of normal and tumor colon cells

Cordero *et al.*

RESEARCH ARTICLE

Open Access

# Large differences in global transcriptional regulatory programs of normal and tumor colon cells

David Cordero<sup>1,2,3†</sup>, Xavier Solé<sup>1,2,3†</sup>, Marta Crous-Bou<sup>1,2,3</sup>, Rebeca Sanz-Pamplona<sup>1,2,3</sup>, Laia Paré-Brunet<sup>1,2,3</sup>, Elisabet Guinó<sup>1,2,3</sup>, David Olivares<sup>1,2,3</sup>, Antonio Berenguer<sup>1,2,3</sup>, Cristina Santos<sup>2,4</sup>, Ramón Salazar<sup>2,4</sup>, Sebastiano Biondo<sup>5,6</sup> and Víctor Moreno<sup>1,2,3,6\*</sup>

## Abstract

**Background:** Dysregulation of transcriptional programs leads to cell malfunctioning and can have an impact in cancer development. Our study aims to characterize global differences between transcriptional regulatory programs of normal and tumor cells of the colon.

**Methods:** Affymetrix Human Genome U219 expression arrays were used to assess gene expression in 100 samples of colon tumor and their paired adjacent normal mucosa. Transcriptional networks were reconstructed using ARACNe algorithm using 1,000 bootstrap replicates consolidated into a consensus network. Networks were compared regarding topology parameters and identified well-connected clusters. Functional enrichment was performed with SIGORA method. ENCODE ChIP-Seq data curated in the *hmChIP* database was used for *in silico* validation of the most prominent transcription factors.

**Results:** The normal network contained 1,177 transcription factors, 5,466 target genes and 61,226 transcriptional interactions. A large loss of transcriptional interactions in the tumor network was observed (11,585; 81% reduction), which also contained fewer transcription factors (621; 47% reduction) and target genes (2,190; 60% reduction) than the normal network. Gene silencing was not a main determinant of this loss of regulatory activity, since the average gene expression was essentially conserved. Also, 91 transcription factors increased their connectivity in the tumor network. These genes revealed a tumor-specific emergent transcriptional regulatory program with significant functional enrichment related to colorectal cancer pathway. In addition, the analysis of clusters again identified subnetworks in the tumors enriched for cancer related pathways (immune response, Wnt signaling, DNA replication, cell adherence, apoptosis, DNA repair, among others). Also multiple metabolism pathways show differential clustering between the tumor and normal network.

**Conclusions:** These findings will allow a better understanding of the transcriptional regulatory programs altered in colon cancer and could be an invaluable methodology to identify potential hubs with a relevant role in the field of cancer diagnosis, prognosis and therapy.

**Keywords:** Colon cancer, Gene expression, Gene regulatory networks, Transcription factors, Transcriptional interactions

\* Correspondence: v.moreno@iconcologia.net

†Equal contributors

<sup>1</sup>Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), Av Gran Via 199-203, E-08907 L'Hospitalet de Llobregat, Barcelona, Spain

<sup>2</sup>Colorectal Cancer Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain

Full list of author information is available at the end of the article

## Background

Transcriptional regulation has an essential role for proper cell functioning. Gene regulatory programs establish and maintain specific cell states [1], ensure cell homeostasis and avoid metabolic disorders [2]. Genetic regulatory information encoded in DNA binding sites, such as enhancers and promoters, is interpreted by a network of transcription factors (TFs) [3]. Epigenetic events like DNA methylation or histone modifications are regulators of transcription [4,5] and non-coding RNAs such as siRNAs and miRNAs are also involved in gene expression regulation at the post-transcriptional level [6].

Identification of global regulatory perturbations that actively participate in the initiation and maintenance of the tumor state is one of the major challenges in cancer biology [7]. Important processes intimately related to the neoplastic process, such as development and cell differentiation, are widely mediated by gene regulation [8]. Dysregulation of signaling pathways has also been related with tumor growth and cancer progression [9]. Although specific tumor genetic alterations are well described and annotated [10], comprehensive studies are required to obtain more information about the transcriptional programs involved in tumor development. Thus, a global analysis of regulatory network perturbations still remains a fundamental challenge for cancer biology [7].

Recent bioinformatics developments make use of large-scale gene expression datasets to infer genome-wide gene regulatory networks (GRN) [11]. Although not as accurate as methods based on experimental procedures and usually requiring subsequent validation, this approach to computationally-infer regulatory networks can be useful to predict *in-vivo* functions of specific cell types [12]. Diverse methodological approaches to infer GRNs have been proposed, such as regression-based methods, correlation, information-theoretic approaches and Bayesian networks [13]. Among all those, the ARACNe algorithm for the reconstruction of GRNs has been successfully applied to reverse-engineer large-scale transcriptional networks in B-cell leukemia [14,15], neuroblastoma [16], T cell acute lymphoblastic leukemia [17] and prostate cancer [18]. These methodologies have also been applied to analyze and compare GRNs of several human tissues [19]. However, there are a limited number of studies about gene regulatory network inference in colon cancer cells, and these analyses were restricted to a small number of genes or used small sample sizes for the inference [20-23].

The aim of our study is to infer GRNs from transcriptional data obtained for a large sample of stage II colon tumor cells and paired adjacent pathologically normal mucosa, as well as to perform a comprehensive analysis of the changes in the transcriptional regulatory programs related to the tumor phenotype.

## Methods

### Patients and samples

One hundred patients with an incident diagnosis of colon cancer who were visited at the Bellvitge University Hospital (Barcelona, Spain) between January 1996 and December 2000 were included in the study. Cases were selected to define a homogenous series of patients with stage II, microsatellite-stable, pathology confirmed adenocarcinoma of the colon. All patients underwent radical surgery and had no signs of tumor cells when margins were examined. Fresh samples were collected and frozen by the pathologist from the surgical specimen. Adjacent mucosa was obtained from the proximal margin and was at least 10 cm distant from the tumor lesion. The Clinical Research Ethics Committee (CEIC) of the Bellvitge Hospital approved the study protocol, and all individuals provided written informed consent to participate and for genetic analyses to be done on their samples. The approval number is PR178/11. Additional information about the study and patient samples can be found at <http://www.colonomics.org>.

### Gene expression dataset

Total RNA was isolated from tissue samples of tumor and normal adjacent mucosa using Exiqon's miRCURY™ RNA Isolation Kit (Exiqon, Denmark), according to manufacturer's protocol. Extracted RNA was quantified by NanoDrop® ND-1000 Spectrophotometer (Nanodrop technologies, Wilmington, DE) and stored at -80°C. RNA quality was assessed with RNA 6000 Nano Assay (Agilent Technologies, Santa Clara, CA) following manufacturer's recommendations and was further confirmed by gel electrophoresis. RNA integrity numbers showed good quality (mean = 8.1 for tumors, and 7.5 for adjacent normal). RNA purity was measured with the ratio of absorbance at 260 nm and 280 nm (mean = 1.96, sd = 0.04), with no differences among tissue types.

RNA samples were hybridized onto the Affymetrix Human Genome U219 96-Array Plate platform (Affymetrix, Santa Clara, CA) following Affymetrix standard procedures. Annotation of the array was based on hg19 genome version. A blocked experimental design was implemented to avoid biases due to potential plate effects (i.e. all plates contained the same proportion of normal and tumor samples). After evaluating the quality of the 200 CEL files using Affymetrix standard quality parameters (e.g. level of background noise, labeling and hybridization efficiency, and RNA degradation), 4 arrays (two normal-tumor pairs) were excluded. Therefore, a final dataset of 196 arrays was used for subsequent analyses. Raw data were normalized together using the Robust Multi-array Average (RMA) algorithm [24] implemented in the *affy* package [25] of the Bioconductor suite (<http://bioconductor.org>). All other analyses were done with R 2.15.1 statistical

computing suite (<http://www.R-project.org>). A model-based clustering was applied to the full expression dataset in order to detect and remove non-expressed and saturated probe-sets from further analyses.

The complete gene expression dataset was uploaded to the National Center for Biotechnology Information's Gene Expression Omnibus Database with GEO series accession number GSE44076.

#### Transcription factor selection

The list of TFs used was built by merging two different sources of information. The first one was the manually-curated compilation of human TFs reported by [26]. More specifically, 1,391 TFs classified in Supplementary Information S3 as 'a', 'b' or 'other' were chosen. In order to generate a broader set of putative TF genes, the collection of curated TFs was complemented with an additional set of 1,415 genes that were associated with specific GO terms related to transcription. In particular, genes associated with GO terms (GO:0045449 - Regulation of transcription, GO:0030528 - Transcription regulator activity and GO:0001071 - Nucleic acid binding transcription factor activity) were chosen. The GO database release used was 2011-03-19 accessed from AmiGO version 1.8 [27]. This yielded a set of 2,806 unique TFs, which were represented by 7,811 Affymetrix probe-sets in the expression array that was used.

#### Inference, representation and analysis of transcriptional regulatory networks

Transcriptional regulatory networks were built using the ARACNe algorithm [15]. Prior to the ARACNe analysis, simulations were performed to model the optimal kernel width that allowed a proper mutual information (MI) estimation in our dataset. The null distribution of the MI was also empirically determined by simulation analysis in order to be able to further identify those significant correlations between TFs and their putative target genes. The significance p-value used as a threshold was  $1e-07$ . ARACNe2 algorithm was run with DPI tolerance set to 0 to remove potential indirect transcriptional interactions from both networks. Remaining parameters were used with their default values. For each network, 1000 bootstrap replicates were performed and summarized to obtain more robust and accurate consensus networks. Only the giant connected component of both networks was considered for downstream analyses. Network visualization, descriptive, simple parameters estimation and figures were performed with Cytoscape software version 2.8.2 [28]. Directed graphs were used to describe networks, in which a regulatory relationship between a TF and a target gene was represented by a directed edge (i.e. arrow) between these two connected nodes, being the origin of the edge the TF. Comprehensive network topology

analyses, along with the estimation of complex parameters, were carried out with the Network Analyzer Cytoscape plugin [29]. KEGG pathway enrichment analysis was performed with the SIGORA R package version 0.9.2 and default parameter values [30]. In the analysis of lost edges, a gene was considered to become silenced in the tumor if its average expression level was smaller than 4 and the  $\log_2$  fold change between the tumor and the normal expression values was smaller than -1 (i.e. a 2-fold change decrease in the tumor). The analysis of network clusters was performed with the MINE Cytoscape plugin [31]. Only clusters with more than 10 nodes were considered for detailed analysis. Somatic mutation data were obtained from the COSMIC database [10] using the following parameters: large intestine (tissue), all (subtissue), carcinoma (histology), all (subhistology). Only genes with a mutation frequency greater than 5% were considered for further analysis.

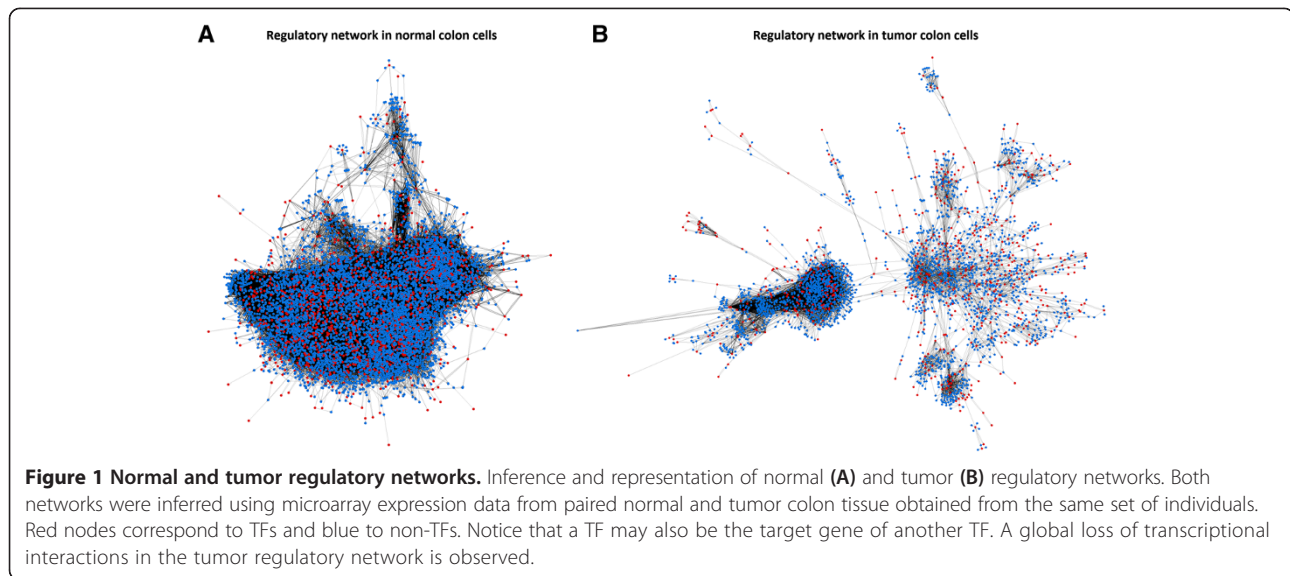
#### In-silico network validation

Gene annotation, (e.g. Ensembl gene id, chromosome, strand, start and end position) was retrieved through *BioMart* R/Bioconductor package [32]. For each gene, genomic sequence around the transcription start site  $\pm$  1 kb according to hg18 coordinates was obtained with the *BSgenome* R/Bioconductor package version 1.24.0. The validation analysis was performed using the *hmChIP* database, which contains ChIP-Seq and ChIP-on-chip data from ENCODE experiments that represent more than 10,000,000 protein-DNA interactions [33]. Only the interactions of TFs with at least more than 20 target genes in the normal tissue network were considered for validation. For each TF, the *hmChIP* database was queried providing a list of genomic regions corresponding to the regulatory sequences of their targets in the normal tissue network. Results were rank ordered based on the degree of overlap between the uploaded genomic regions and the peak lists collected by *hmChIP* database from ChIP-Seq and ChIP-on-chip ENCODE datasets. Enrichment ratios and significance p-values for the overlaps were provided by *hmChIP* tool. Benjamini and Hochberg false discovery rates were also reported by the tool to account for multiple testing.

## Results

### Massive loss of regulatory activity in tumor cells

A large loss of transcriptional interactions was found in the tumor regulatory network (Figure 1, Table 1). The tumor regulatory network contained 47% fewer TFs than the network of normal cells (621 vs. 1,177), as well as 60% fewer target genes (2,190 vs. 5,466). Most nodes disappeared in the tumor network because their expression was completely unrelated to other nodes. Furthermore, the number of direct transcriptional interactions was



reduced by 81% (11,585 in the tumor network vs. 61,226 in adjacent normal cells).

Notably, although the node overlap between both networks is large (81% of the tumor nodes are found in the normal network), only 19% of the interactions present in the tumor network are found in the normal network (Figure 2). To visualize both entire networks with Cytoscape [28] or another platform the network representations can be found online (Additional file 1). Additionally, specific TFs and their target genes (or vice versa) can be explored in the project website (<http://www.colonomics.org/regulatory-networks>).

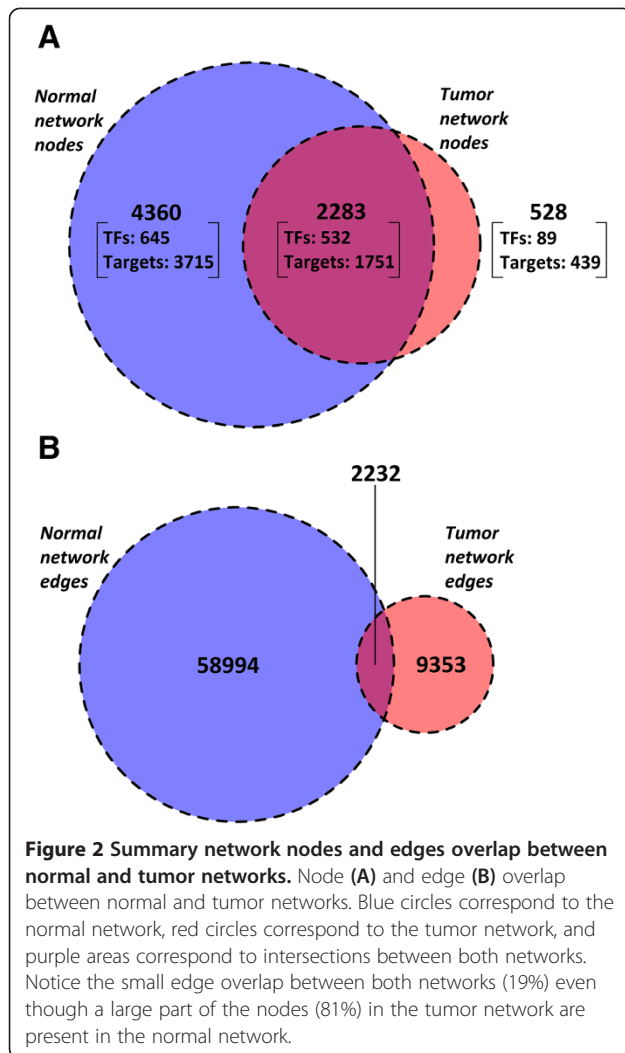
The vast majority of lost edges (76%) show a large decrease in MI but relatively small changes in gene expression (absolute  $\log_2$  fold change  $< 1$ , Figure 3). This

**Table 1 Networks descriptive parameters and topological features**

	Normal network	Tumor network	Ratio Tumor/Normal
<b>Descriptive parameters</b>			
Nodes	6,643	2,811	0.42
Transcription factors	1,177	621	0.53
Target genes	5,466	2,190	0.40
Edges	61,226	11,585	0.19
<b>Main topological features</b>			
Network diameter	12	17	1.42
Proportion of shortest paths	14%	4%	0.29
Characteristic path length	4.0	5.0	1.25
Average number of neighbors	16.9	7.6	0.45
Multi-edge node pairs	5,204	976	0.19

suggests that decreased connectivity in the tumor network was more related to transcriptional dysregulation than to gene silencing. Lost edges in the tumor network were classified into four groups according to their change in MI and gene expression change (Figure 4). Panels A-C contains examples of loss of interaction by either silencing of the TF and/or the target. These groups comprise a small proportion of lost edges (A: 80, 0.2%; B: 1,105, 2.1%; C: 923, 1.7%). Panel D shows a loss of interaction due to a decrease in the correlation, without evidence of TF or target silencing. Remarkably, most of the lost edges in the tumor network (50,882, 96%) belong to pattern D, where the loss of regulatory activity does not depend on major changes in average gene expression levels.

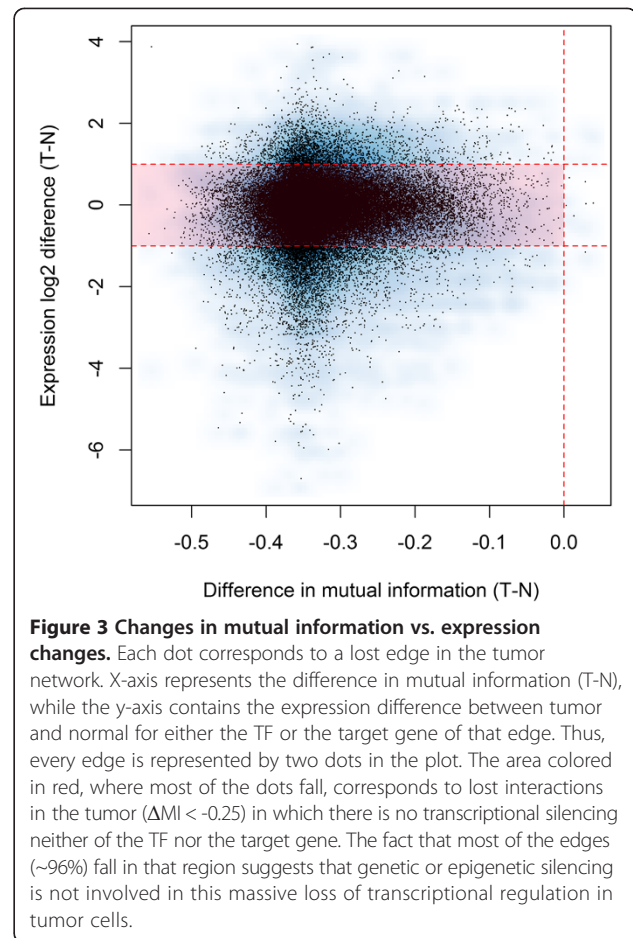
Loss of robustness in the tumor network was suggested by the comparison of the topological features of both networks, as shown in Table 1. Firstly, a larger distance between nodes in the tumor network was observed for different parameters, such as an increased network diameter, the characteristic path length or the decrease in average shortest paths. Secondly, a lower connectivity in the tumor network was identified according to the values of parameters related to neighborhood, such as the decrease in average number of neighbors and multi-edge node pairs. Furthermore, a characteristic of the tumor network not found on the normal was the existence of a small subset of low connected TFs with a remarkable contribution to minimal shortest paths (closeness centrality, see figure in Additional file 2). Although no significant functional enrichment was found for this set of TFs, these genes may have the potential ability to further disrupt the tumor network by breaking it into multiple disconnected components if some of their incoming or outgoing interactions are further lost. For a full set of figures



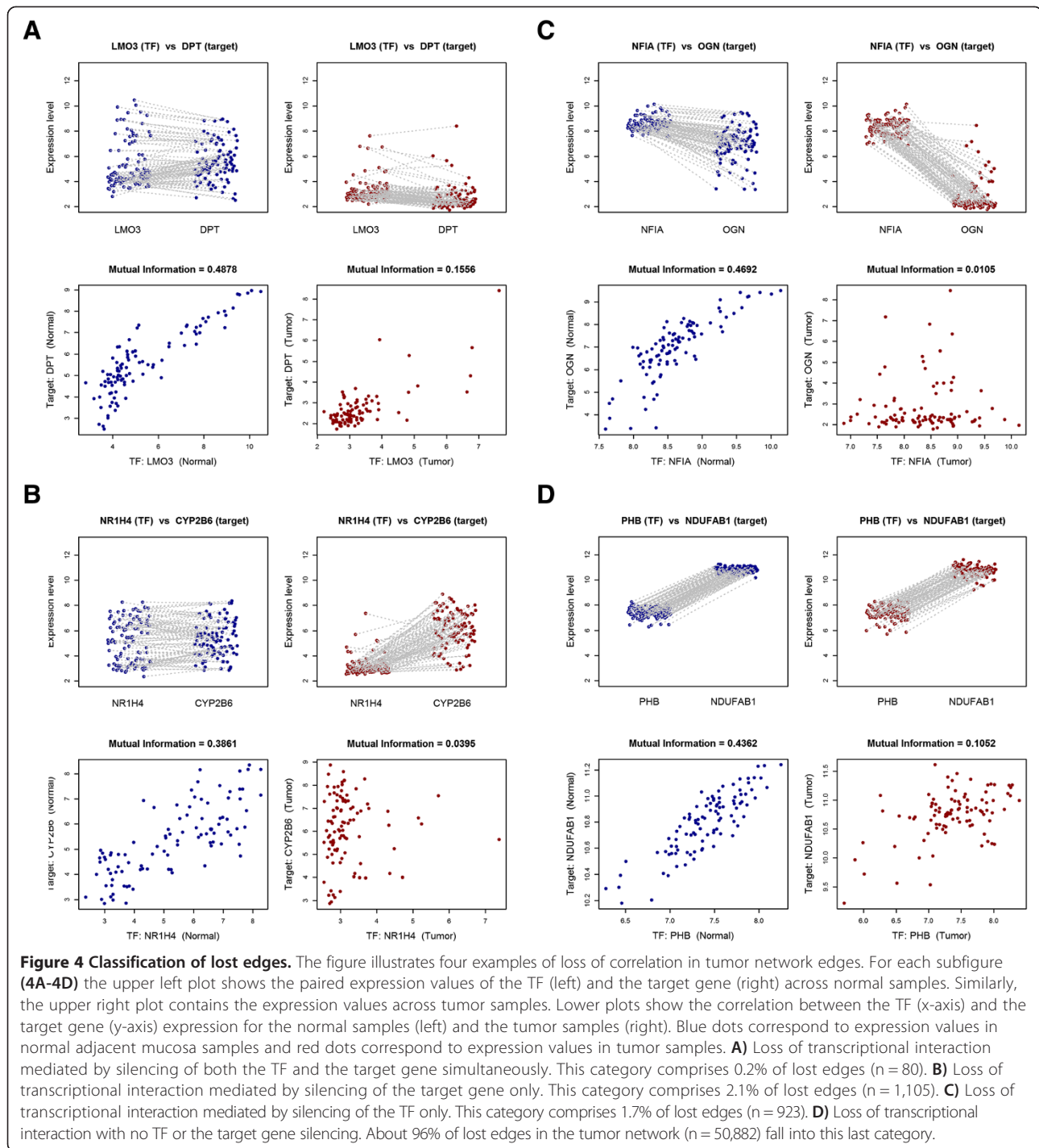
of other topological features comparing the networks see Additional file 2.

### Gain of regulatory activity in tumor cells

Although the tumor network shows a large loss of transcriptional interactions, there are also specific TFs that largely increase their number of target interactions in the tumor network. A total of 91 TFs with increased activity (i.e. out-degree) and 235 up-regulated (i.e. in-degree) target genes were identified in the tumor network. The analysis of gained edges suggests a stronger role of the TFs compared to the targets. Specifically, the 91 TFs with increased activity revealed 2,224 new edges in the tumor network (24 on average, median = 12) while the 235 up-regulated targets only comprise 1,292 new transcriptional interactions (5 on average, median = 4). TFs and target genes that most increase their connectivity in the tumor network are shown in Table 2 (see complete lists in Additional file 3). KEGG pathways [34] enrichment analysis of this set of genes using the SIGORA method [30]



revealed that the *Colorectal cancer* pathway (map05210) was significantly overrepresented among these TFs with increased activity (p-value =  $8.9e-9$ ). This pathway includes well-known cancer-related genes such as *FOS*, *TGFB3* and *TGFB1* that increased connectivity in the tumor network. In order to evaluate if this gain of regulatory activity in colon tumor cells may be related to somatic mutations we studied the degree distribution (as indicator of regulatory activity) for TFs and target genes, classified as frequently mutated (if present in COSMIC database) or not [10]. We have found that regulatory activity is independent of mutations for TFs. However, target genes included in COSMIC database showed a significant larger regulatory control than other non-mutated genes in tumors (mean in-degree 4.5 in non mutated and 7.7 in mutated,  $p = 0.000021$ ). These differences were not observed in the normal network (mean in-degree 11.3 in non mutated and 12.6 in mutated,  $p = 0.16$ ), indicating that mutated genes tend to lose less regulation or even increase it, since these differences were also true for targets that increased connectivity. Examples of mutated target genes that increase connectivity are *CDH11*, *CFH*, *COL3A1*, *COL6A3* and *COL5A2* (complete list in Additional file 4). These genes are



mutated with frequency greater than 5% and show in the tumor network a large increment of regulatory activity.

**In-silico network validation with experimental data**

Public ChIP-Seq and ChIP-on-chip datasets mainly from the ENCODE project [35] and compiled in the *hmChIP* database were used [33]. In order to avoid biases derived

from tumor-specific interactions, only TFs from our normal regulatory network with available datasets from ChIP-Seq or ChIP-on-chip experiments were initially selected for validation. TFs with less than 20 targets in the normal network or showing less than 500 peaks in *hmChIP* database were filtered out to avoid focusing on tissue-specific regulations. Finally 16 TFs and their 1,443 putative target genes were selected for validation. Remarkably, though

**Table 2 Nodes with increased activity**

*TFs that most increase their activity in tumors*

Transcription factor	Targets in Normal network	Targets in Tumor network	Gained interactions	Ratio T/N
<i>SNAI2</i>	1	119	118	119.0
<i>MMP14</i>	10	121	111	12.1
<i>AEBP1</i>	103	186	83	1.8
<i>BASP1</i>	43	123	80	2.9
<i>HCLS1</i>	91	170	79	1.9
<i>TFEC</i>	6	84	78	14.0
<i>DKK3</i>	41	112	71	2.7
<i>COL1A1</i>	62	131	69	2.1
<i>CD86</i>	74	141	67	1.9
<i>MAFB</i>	125	189	64	1.5
<i>NOTCH3</i>	18	82	64	4.6
<i>GLI2</i>	37	100	63	2.7
<i>TGFB1</i>	1	61	60	61
<i>GREM1</i>	14	70	56	5.0
<i>HOPX</i>	46	102	56	2.2

*Most up-regulated targets in Tumors*

Target gene	Targets in-degree in Normal	Targets in-degree in Tumor	Gained interactions	Ratio T/N
<i>NNMT</i>	3	32	29	10.7
<i>CDH11</i>	1	24	23	24.0
<i>RAB31</i>	20	42	22	2.1
<i>MXRA8</i>	3	23	20	7.7
<i>RFTN1</i>	8	28	20	3.5
<i>CFH</i>	3	20	17	6.7
<i>COL3A1</i>	14	31	17	2.2
<i>EMILIN1</i>	12	28	16	2.3
<i>ENTPD1</i>	12	28	16	2.3
<i>MRC2</i>	7	23	16	3.3
<i>STAU1</i>	1	17	16	17.0
<i>AXL</i>	9	24	15	2.7
<i>OLFML2B</i>	10	25	15	2.5
<i>VCAM1</i>	1	15	14	15.0
<i>COL6A3</i>	12	25	13	2.1

The table lists the top 15 TFs and target genes that most increase their activity in the tumor network, sorted by the number of gained interactions. Only nodes that appeared in both networks were considered. See complete lists in Additional file 3.

the experimental datasets were not restricted to colon tissue, 6 out of the 16 TFs (38%) showed significant overrepresentation (enrichment ratio > 1). One additional TF showed marginally significant overrepresentation in the experimental data collected in the *hmChIP* database, as shown in Table 3. This result reinforces the robustness of our inferred

networks, which seem to be reasonably capturing transcriptional relationships between TFs and their target genes.

### Functional analysis of node clusters

It is known that functionally related genes tend to cluster together in network-defined biological systems (e.g. protein-protein interaction, transcriptional, or co-expression networks). Therefore, we aimed to detect clusters of genes in both the normal and tumor network to identify tumor-specific highly interconnected sub-networks, potentially enriched in relevant biological pathways. The network cluster analysis revealed 42 clusters in the normal network with more than 10 nodes. These included 953 highly interconnected genes. The tumor network included 29 clusters with 871 nodes. The distribution of nodes among clusters was similar for both networks. The list of clusters and enriched pathways (identified by SIGORA method) can be found in Additional file 5. Although most of the clusters in the tumor network were enriched in functions already present in the normal network, some clusters showed tumor-specific significant enrichments in functions with a potential role in tumor development (Table 4). More specifically, clusters 3 and 19 showed an overrepresentation of immune response pathways (e.g., Chemokine signaling pathway, Toll-like receptor signaling pathway, Cytokine-cytokine receptor interaction), and cluster 4 showed enrichment in Wnt signaling proteins. Other clusters, such as 11 and 18, also included significant enrichment of potentially relevant processes such as cell proliferation (e.g. MAPK pathway) or apoptosis, respectively.

### Discussion

In this study we have reverse-engineered the transcriptional regulatory networks of both pathologically normal and tumor colon cells obtained from the same set of patients. Using a large-scale gene expression microarray dataset, the ARACNe algorithm was applied to both tissue types independently. ARACNe gives preference to identify direct transcriptional regulatory interactions between TFs and their target genes. When both networks are compared, the most outstanding feature is the considerable loss of transcriptional interactions found in tumor cells (81%), with a global significant decrease in TFs (47%), target genes (60%). The fact that both normal and tumor samples belong to the same set of individuals, as well as the carefully performed experimental design to prevent biases between tissue types, strongly suggests that these large differences between networks are mainly due to the tumor phenotype.

Most of the TFs and target genes involved in disrupted interactions in the tumor network still maintain their expression levels, while only a minor proportion of lost edges may be explained by a complete loss of expression



**Table 3 In-silico network validation**

Transcription factor (Gene Symbol)	# Targets (In normal network)	# Peaks (In hmChIP DB)	Enrichment ratio	p-value	FDR
<i>TCF4</i>	408	46,018	<b>1.82</b>	2.0e-07	3.7e-06
<i>NR3C1</i>	246	24,967	0.60	0.12	-
<i>PBX3</i>	186	39,691	0.40	0.0063	0.019
<i>HNF4A</i>	103	32,083	<b>2.71</b>	0.00027	0.0016
<i>TCF12</i>	67	54,191	<b>3.33</b>	2.0e-06	1.8e-05
<i>RBL2</i>	55	16,395	<b>2.33</b>	0.0050	0.018
<i>SUZ12</i>	50	8,742	0.62	0.12	-
<i>ESRRA</i>	42	3,284	1.50	0.37	-
<i>FOXP2</i>	42	44,482	<b>2.00</b>	0.043	0.11
<i>MAX</i>	41	16,467	1.80	0.12	-
<i>CDX2</i>	40	24,460	1.38	0.38	-
<i>SRF</i>	39	35,784	1.91	0.052	0.12
<i>STAT1</i>	35	2,804	<b>3.20</b>	0.00097	0.0044
<i>FOXA1</i>	32	21,540	0.55	0.062	0.12
<i>NFYB</i>	31	4,630	1.20	1	-
<i>RAD21</i>	26	33,302	1.40	0.50	-

Results provided by hmChIP tool containing CHIP-Seq and CHIP-chip ENCODE experiments [33]. TFs are ordered according to the number of target genes in the normal network. Cells with enrichment ratio in bold highlight significantly overrepresented TFs.

of one or both interactors. This expression silencing may be attributed either to genomic (e.g. DNA deletions, somatic mutations in promoter regions that hinder TF binding, transcript-truncating alterations, etc.) or epigenomic mechanisms (e.g. miRNA-associated transcript degradation, promoter hypermethylation, alterations in chromatin activation and repression marks, etc). On the other hand, disrupted interactions involving TFs and target genes that maintain expression levels in normal and tumor cells may be attributed to multiple reasons: presence or absence of a third-party molecule that could be acting as a post-translational modulator of the TF activity (i.e. phosphorylation, acetylation, ubiquitination) [36], alteration of key co-factors [1], or alterations in promoter regions that could create new TF-binding sites in target genes [37,38]. The small set of genes involved in the loss of interactions through TFs or target gene silencing (~4%) is more likely to belong to currently known altered colon cancer pathways as the Wnt signaling and others, due to apparent under-expression. However, the vast majority of lost edges would not be easy to identify just by exploring the expression values of their TFs or targets genes. We think new and interesting undescribed mechanisms for molecular biology of colon cancer might be related to this gene deregulation without average gene expression change. A potential limitation may be the tumor cellular heterogeneity that could also be contributing to the observed loss of connectivity. While normal mucosa is a relatively homogeneous tissue among subjects, tumors are more heterogeneous, with diverse

predominant cellular clones (epithelial, stromal and derived from the immune system). This could result in an apparent global loss of correlation if diverse transcriptional networks were mixed in the tumor.

The network of tumor cells also showed the emergence of a new set of transcriptional interactions that may have an essential role in tumor development and the acquisition of new cellular abilities. Recent studies have demonstrated that the activation of a small regulatory module is necessary and sufficient to initiate and maintain an aberrant phenotypic state in brain tumors [16]. Therefore, network inference approaches could prove effectively useful to uncover new modules and the master regulators that orchestrate malignant transformation. Among the TFs ranked at the top of the list of increased connectivity, our analysis identified colorectal cancer related genes: two oncogenes (*MAFB* [39] and *GLI2* [40]), proliferation-related genes (*NOTCH3* [41] and *TGFB1* [42]), epithelial-mesenchymal transition (*SNAI2* [43]) and the Wnt signaling genes *SFRP4*, *TWIST1*, *SMARCA4* and *DKK3*, potentially involved in colorectal cancer angiogenesis [44]. One remarkable gene with increased activity in the tumor network was *GREM1*. This gene encodes a member of the bone morphogenic protein antagonist family and may play a role in regulating organogenesis, body patterning and tissue differentiation. Interestingly, *GREM1* has been previously related with a locus strongly associated with increased colorectal cancer risk [45]. Moreover, increased expression of *GREM1* has also been recently found in

**Table 4 Emergent network clusters in Tumors**

Tumor cluster*	Number of genes	Pathway	Adjusted P-value <sup>§</sup>
1	120	Vascular smooth muscle contraction	1.1e-09
2	112	GnRH signaling pathway	5.9e-04
2	112	Staphylococcus aureus infection	4.8e-02
3	70	Chemokine signaling pathway	8.1e-08
3	70	Toll-like receptor signaling pathway	3.1e-07
3	70	Ether lipid metabolism	9.5e-04
4	51	Glycosphingolipid biosynthesis - ganglio series	1.6e-03
4	51	Wnt signaling pathway	1.7e-03
4	51	GnRH signaling pathway	1.3e-02
5	70	Adherens junction	1.9e-04
5	70	Chemokine signaling pathway	4.1e-02
7	44	Tight junction	5.6e-05
7	44	Tryptophan metabolism	2.4e-04
7	44	Glycosaminoglycan biosynthesis - chondroitin sulfate	4.7e-04
8	27	Adherens junction	4.0e-03
9	16	Protein digestion and absorption	4.4e-07
9	16	Adherens junction	5.9e-03
11	16	MAPK signaling pathway	2.1e-15
11	16	Prion diseases	2.4e-03
13	24	Beta-Alanine metabolism	4.4e-04
13	24	NOD-like receptor signaling pathway	9.8e-03
16	32	Glycosaminoglycan biosynthesis - chondroitin sulfate	4.5e-08
18	14	Apoptosis	2.2e-06
18	14	Nucleotide excision repair	1.0e-03
19	14	Cytokine-cytokine receptor interaction	1.4e-02
21	13	Butanoate metabolism	5.6e-05
21	13	Amino sugar and nucleotide sugar metabolism	3.4e-03
22	12	Glutathione metabolism	3.4e-04
23	18	DNA replication	6.7e-06
25	32	Vascular smooth muscle contraction	3.7e-06
28	12	DNA replication	9.6e-05

\*Only clusters with significant enriched functions in tumors not already present in normal are shown.

<sup>§</sup>P-value for functional enrichment derived from SIGORA method.

colorectal polyps [46], as well as in the dysplasia to carcinoma transition in colon tumors [47]. Therefore our results suggest that *GREMI* may be mediating its tumorigenic effect by the activation of a large transcriptional program.

Furthermore, encouraging results were obtained in the study of the relationship of somatic mutations in colorectal tumors in the set of relevant genes identified through our network approach. Though frequent mutation was independent of regulatory activity for TFs, we observed an association for target genes, with larger regulatory activity among mutated genes. Though this was a correlation analysis using external data from COSMIC database (we do not know if our tumors were actually mutated), it is suggestive that mutated genes trigger a regulatory control in the tumor. The presence of mutations combined with the alteration in their transcriptional regulatory connectivity postulate these genes as strong candidates to be involved in the pathogenesis of colon cancer, and even other type of tumors.

The analysis of network clusters has identified relevant sub-networks of highly connected genes specific of tumors. The regulatory network of normal cells is large and compact. Only 42 clusters have been identified with more than 10 genes. These clusters only account for 14% of the network genes, indicating that there is extensive regulation, but relatively low modularity. The tumor cell, however, has revealed 29 clusters that include 30% of their genes. This is consistent with a more modular organization of the regulatory machinery, which is also evident from the network representation (Figure 1). The functional analysis of these clusters has shown significant enrichment of known tumor-specific pathways: immune response, Wnt signaling, DNA replication, cell adherence, apoptosis, DNA repair, among others (Table 4). Some specific metabolism pathways appear also specifically captured by this analysis of sub-networks, which may be candidate for intervention: glycosphingolipid biosynthesis, tryptophan metabolism, glycosaminoglycan biosynthesis (chondroitin sulfate), beta-alanine metabolism, butanoate metabolism, glutathione metabolism. Obviously, all these functions are present in the normal cell, but they seem enhanced at the transcriptional level in the tumor, in such a way that a large cluster of related genes appear as a relevant entity. In this analysis we have generally focused on the gain of activity in the tumor network rather than on the lost interactions, given the massive loss of tumor network interactions that difficult to detect enriched functions. Despite this intrinsic limitation, we want to emphasize that the transcriptional loss found may influence the emergence of new functionality in the tumor cells. This finding may have a potential impact on the future of cancer molecular biology at level of further experiments and their corresponding biological interpretations.

The inference of GRNs has already been successfully applied to other malignances such as leukemia [14], breast cancer [48,49] or ovarian tumors [50], with relevant findings regarding breast cancer metastasis prognostic markers

or prioritization of druggable gene targets for ovarian cancer. In colorectal cancer some researchers have also explored the reconstruction of GRNs, but with limited approaches to one transcription factor [23] or only tumor tissue [21,22]. To our knowledge, this is the first study in colon cancer that has simultaneously inferred networks for both tumor and adjacent normal cells obtained from the same set of individuals with a consistent methodology that makes both networks totally comparable.

We are aware that computational approaches of network reverse-engineering may suffer from intrinsic limitations. Therefore, we attempted a validation of the network to reinforce the validity of our study. An initial attempt to *in-silico* identify expected TF binding sites in targets was rejected because of the limited number and relative quality of the available TF positional weight matrices both in JASPAR [51] and TRANSFAC Public [52] databases. Other approach to validate the inferred regulatory networks would be to replicate our results in another colon cancer dataset. This has not been possible due to the lack of proper datasets to replicate the findings. The ARACNe's authors emphasize in their papers that a hundred samples is the minimum sample size required to infer transcriptional networks with proper accuracy and they specifically discourage users to apply their algorithm on small datasets [15,53]. The TCGA project [54] only provides 23 normal-tumor colon pairs available and we were unable to find a dataset with a more than 50 samples available after an exhaustive search in the most comprehensive public gene expression databases (GEO and ArrayExpress). Over the last decade, ChIP-on-chip and especially ChIP-Seq assays have become gold standard techniques for large-scale protein-DNA interaction identification. Therefore, ChIP-Seq and ChIP-on-chip datasets from the ENCODE project were used to validate interactions inferred by ARACNe. Since we restricted the potential set of TFs to be validated to those that had more than 20 interactions in the normal network and more than 500 experimentally observed peaks, only a very small part of the network could be tested. However, the obtained results were encouraging since 6 of the 16 tested TFs showed a good level of agreement. The large differences between the number of experimentally detected peaks and the number of inferred target genes for each one of the TFs may suggest a high rate of false negative interactions in our inferred networks, though it is not easy to interpret ChIP data, that provides may peaks that are not necessarily related to direct transcription interactions [55]. Failure in the validation of some TFs might also be partially influenced by the failure of the algorithm to completely remove indirect associations from the network due to high order interactions. In this direction, an extension of the ARACNe algorithm (hARACNe) specifically designed to deal with n-order

interactions has been recently released, showing a significant increase in the quality and robustness of the inferred network [56]. Network deconvolution solutions over correlation-based networks have also proven to be successful for this purpose [57]. Due that the large heterogeneity of cell line tissues explored in the ENCODE project, we positively consider the overall observed level of agreement (38%), which is in the same range as previous studies found for other inferred transcriptional networks [14].

## Conclusion

The inference of direct transcriptional networks at the whole-genome level has allowed us to detect a predominant loss of transcriptional activity in colon tumor cells, which has not been described before to the best of our knowledge. However, some specific TFs and biological processes related to colon cancer also increased the connectivity and became hubs in the dysregulated tumor network. These findings will allow a better comprehension of the transcriptional regulatory programs altered in colon cancer and could be an invaluable methodology to identify potential hubs with a relevant role in the field of cancer diagnosis, prognosis and therapy.

## Additional files

**Additional file 1:** The two networks representation.

**Additional file 2:** Complex networks parameters.

**Additional file 3:** Full list of nodes that increase their activity.

**Additional file 4:** Genes with altered activity and mutations in COSMIC database.

**Additional file 5:** Clusters enrichment analysis.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Conceived the study: DC, XS, VM. Performed analysis: DC, XS, MCB, RSP, LPB, EG, DO, AB, VM. Recruited patients: CS, RS, SB. Wrote the manuscript: DC, XS, VM. Discussed the manuscript: MCB, RSP, LPB, EG, DO, AB, CS, RS, SB. All authors read and approved the final manuscript.

## Acknowledgments

The authors would like to thank Ferran Martínez, Adrià Closa, Carmen Atencia, Pilar Medina and Isabel Padrol for their technical assistance. This work was supported by the Catalan Institute of Oncology and the Private Foundation of the Biomedical Research Institute of Bellvitge (IDIBELL), the Instituto de Salud Carlos III grants PI08-1635, P509-1037, PI11-1439, PIE13/00022 and CIBERESP CB06/02/2005 and the "Acción Transversal del Cáncer", the European Commission grant FP7-COOP-Health-2007-B HiPerDART, the Catalan Government DURSI grant 2009SGR1489, the Fundación Privada Olga Torres (FOT), and the AECC (Spanish Association Against Cancer) Scientific Foundation. The "Xarxa de Bancs de Tumors de Catalunya" sponsored by "Pla Director d'Oncologia de Catalunya (XBTC)" helped with sample collection.

## Author details

<sup>1</sup>Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), Av Gran Via 199-203,

E-08907 L'Hospitalet de Llobregat, Barcelona, Spain. <sup>2</sup>Colorectal Cancer Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. <sup>3</sup>Biomedical Research Centre Network for Epidemiology and Public Health (CIBERESP), Barcelona, Spain. <sup>4</sup>Department of Medical Oncology, Catalan Institute of Oncology (ICO), Barcelona, Spain. <sup>5</sup>Department of General and Digestive Surgery, Colorectal Unit, Bellvitge University Hospital (HUB - IDIBELL), Barcelona, Spain. <sup>6</sup>Department of Clinical Sciences, School of Medicine, University of Barcelona (UB), Barcelona, Spain.

Received: 30 April 2014 Accepted: 17 September 2014

Published: 24 September 2014

## References

- Lee TI, Young RA: Transcriptional regulation and its misregulation in disease. *Cell* 2013, **152**(6):1237–1251.
- Desvergne B, Michalik L, Wahli W: Transcriptional regulation of metabolism. *Physiol Rev* 2006, **86**(2):465–514.
- Kadonaga JT: Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 2004, **116**(2):247–257.
- Bannister AJ, Kouzarides T: Regulation of chromatin by histone modifications. *Cell Res* 2011, **21**(3):381–395.
- Choy MK, Movassagh M, Goh HG, Bennett MR, Down TA, Foo RS: Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* 2010, **11**:519.
- Lu J, Clark AG: Impact of microRNA regulation on variation in human gene expression. *Genome Res* 2012, **22**(7):1243–1254.
- Goodarzi H, Elemento O, Tavazoie S: Revealing global regulatory perturbations across human cancers. *Mol Cell* 2009, **36**(5):900–911.
- Ben-Tabou de-Leon S, Davidson EH: Gene regulation: gene control network in development. *Annu Rev Biophys Biomol Struct* 2007, **36**:191.
- Anastas JN, Moon RT: WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* 2013, **13**(1):11–26.
- Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA: COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 2011, **39**(Database issue):D945–950.
- Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D: How to infer gene networks from expression profiles. *Mol Syst Biol* 2007, **3**:78.
- Deng Y, Johnson DR, Guan X, Ang CY, Ai J, Perkins EJ: In vitro gene regulatory networks predict in vivo function of liver. *BMC Syst Biol* 2010, **4**:153.
- Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G: Wisdom of crowds for robust gene network inference. *Nat Methods* 2012, **9**(8):796–804.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005, **37**(4):382–390.
- Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* 2006, **7** Suppl 1:S7.
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A, Iavarone A: The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 2010, **463**(7279):318–325.
- Della Gatta G, Palomero T, Perez-Garcia A, Ambesi-Impiombato A, Bansal M, Carpenter ZW, De Keersmaecker K, Sole X, Xu L, Paietta E, Racevskis J, Wiernik PH, Rowe JM, Meijerink JP, Califano A, Ferrando AA: Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat Med* 2012, **18**(3):436–440.
- Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, Eastham JA, Gopalan A, Pienta KJ, Shen MM, Califano A, Abate-Shen C: Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 2014, **25**(5):638–651.
- Li J, Hua X, Haubrock M, Wang J, Wingender E: The architecture of the gene regulatory networks of different tissues. *Bioinformatics* 2012, **28**(18):i509–i514.
- Fu J, Tang W, Du P, Wang G, Chen W, Li J, Zhu Y, Gao J, Cui L: Identifying microRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. *BMC Syst Biol* 2012, **6**:68.
- Vineetha S, Chandra Shekara Bhat C, Idicula SM: Gene regulatory network from microarray data of colon cancer patients using TSK-type recurrent neural fuzzy network. *Gene* 2012, **506**(2):408–416.
- Wang X, Gotoh O: Inference of cancer-specific gene regulatory networks using soft computing rules. *Gene Regul Syst Biol* 2010, **4**:19–34.
- Weltmeier F, Borlak J: A high resolution genome-wide scan of HNF4alpha recognition sites infers a regulatory gene network in colon cancer. *PLoS One* 2011, **6**(7):e21667.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Bioinformatics* 2003, **4**(2):249–264.
- Gautier L, Cope L, Bolstad BM, Irizarry RA: affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004, **20**(3):307–315.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 2009, **10**(4):252–263.
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S: AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009, **25**(2):288–289.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, **13**(11):2498–2504.
- Doncheva NT, Assenov Y, Domingues FS, Albrecht M: Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc* 2012, **7**(4):670–685.
- Foroushani AB, Brinkman FS, Lynn DJ: Pathway-GPS and SIGORA: identifying relevant pathways based on the over-representation of their gene-pair signatures. *PeerJ* 2013, **1**:e229.
- Rhissorakrai K, Gunsalus KC: MINE: Module identification in networks. *BMC bioinformatics* 2011, **12**:192.
- Durinck S, Spellman PT, Birney E, Huber W: Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009, **4**(8):1184–1191.
- Chen L, Wu G, Ji H: hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics* 2011, **27**(10):1447–1448.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012, **40**(Database issue):D109–114.
- Encode Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M: An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, **489**(7414):57–74.
- Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, Klein U, Dalla-Favera R, Califano A: Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol* 2009, **27**(9):829–839.
- Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schandendorff D, Kumar R: TERT promoter mutations in familial and sporadic melanoma. *Science* 2013, **339**(6122):959–961.
- Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA: Highly recurrent TERT promoter mutations in human melanoma. *Science* 2013, **339**(6122):957–959.
- Suzuki A, Iida S, Kato-Uranishi M, Tajima E, Zhan F, Hanamura I, Huang Y, Ogura T, Takahashi S, Ueda R, Barlogie B, Shaughnessy J Jr, Esumi H: ARK5 is transcriptionally regulated by the Large-MAF family and mediates IGF-1-induced cell invasion in multiple myeloma: ARK5 as a new molecular determinant of malignant multiple myeloma. *Oncogene* 2005, **24**(46):6936–6944.
- Ruiz i Altaba A: Hedgehog signaling and the Gli code in stem cells, cancer, and metastases. *Sci Signal* 2011, **4**(200):pt9.
- Kato M: Notch signaling in gastrointestinal tract (review). *Int J Oncol* 2007, **30**(1):247–251.
- Biassi F, Tessitore L, Zanetti D, Cutrin JC, Zingaro B, Chiarpotto E, Zarkovic N, Serviddio G, Poli G: Associated changes of lipid peroxidation and transforming growth factor beta1 levels in human colon cancer during tumour progression. *Gut* 2002, **50**(3):361–367.
- Wang Y, Ngo VN, Marani M, Yang Y, Wright G, Staudt LM, Downward J: Critical role for transcriptional repressor Snail2 in transformation by

- oncogenic RAS in colorectal carcinoma cells. *Oncogene* 2010, **29**(33):4658–4670.
44. Zitt M, Untergasser G, Amberger A, Moser P, Stadlmann S, Muller HM, Muhlmann G, Perathoner A, Margreiter R, Gunsilius E, Ofner D: **Dickkopf-3 as a new potential marker for neoangiogenesis in colorectal cancer: expression in cancer tissue and adjacent non-cancerous tissue.** *Dis Markers* 2008, **24**(2):101–109.
  45. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinimann K, Lubbe S, Domingo E, Barclay E, Martin L, Gorman M, Chandler I, Vijaykrishnan J, Wood W, Papaemmanuil E, Penegar S, Qureshi M, Farrington S, Tenesa A, Cazier JB, Kerr D, Gray R, Peto J, Dunlop M, et al: **Common genetic variants at the CRAC1 (HMP5) locus on chromosome 15q13.3 influence colorectal cancer risk.** *Nat Genet* 2008, **40**(1):26–28.
  46. Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, East J, Taylor J, Thomas H, Tomlinson I: **Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1.** *Nat Genet* 2012, **44**(6):699–703.
  47. Galamb O, Wichmann B, Sipos F, Spisak S, Krenacs T, Toth K, Leiszter K, Kalmar A, Tulassay Z, Molnar B: **Dysplasia-carcinoma transition specific transcripts in colonic biopsy samples.** *PLoS One* 2012, **7**(11):e48547.
  48. Ahmad FK, Deris S, Othman NH: **The inference of breast cancer metastasis through gene regulatory networks.** *J Biomed Inform* 2012, **45**(2):350–362.
  49. Demicheli R, Coradini D: **Gene regulatory networks: a new conceptual framework to analyse breast cancer behaviour.** *Ann Oncol* 2011, **22**(6):1259–1265.
  50. Madhamshettiwar PB, Maetschke SR, Davis MJ, Reverter A, Ragan MA: **Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets.** *Genome Med* 2012, **4**(5):41.
  51. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B: **JASPAR: an open-access database for eukaryotic transcription factor binding profiles.** *Nucleic Acids Res* 2004, **32**(Database issue):D91–94.
  52. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: **TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006, **34**(Database issue):D108–110.
  53. Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A: **Reverse engineering cellular networks.** *Nat Protoc* 2006, **1**(2):662–671.
  54. Cancer Genome Atlas Network: **Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012, **487**(7407):330–337.
  55. Levitsky VG, Kulakovskiy IV, Ershov NI, Oschepkov DY, Makeev VJ, Hodgman TC, Merkulova TI: **Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data.** *BMC Genomics* 2014, **15**(1):80.
  56. Jang IS, Margolin A, Califano A: **hARACNe: improving the accuracy of regulatory model reverse engineering via higher-order data processing inequality tests.** *Interface Focus* 2013, **3**(4):20130011.
  57. Feizi S, Marbach D, Medard M, Kellis M: **Network deconvolution as a general method to distinguish direct dependencies in networks.** *Nat Biotechnol* 2013, **31**(8):726–733.

doi:10.1186/1471-2407-14-708

**Cite this article as:** Cordero et al.: Large differences in global transcriptional regulatory programs of normal and tumor colon cells. *BMC Cancer* 2014 **14**:708.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

