

Aberrant methylation patterns in colorectal cancer: a meta-analysis

Danielle Fernandes Durso^{1,2,*}, Maria Giulia Bacalini^{3,*}, Ítalo Faria do Valle^{4,5}, Chiara Pirazzini³, Massimiliano Bonafé¹, Gastone Castellani⁵, Ana Maria Caetano Faria⁶, Claudio Franceschi^{1,3,7,*}, Paolo Garagnani^{1,7,8,*}, Christine Nardini^{9,*}

¹Department of Experimental, Diagnostic and Specialty Medicine, Alma Mater Studiorum-University of Bologna, Bologna, Italy

²National Counsel of Technological and Scientific Development (CNPq), ministry of science technology and innovation (MCTI), Brasilia, Brazil

³IRCCS Institute of Neurological Sciences, Bologna, Italy

⁴CAPES Foundation, Ministry of Education of Brazil-Brasília (DF), Brazil

⁵Department of Physics and Astronomy, University of Bologna, Bologna, Italy

⁶Biochemistry and Immunology Department, Biological Sciences Institute, Federal University of Minas Gerais, Belo Horizonte, Brazil

⁷Interdepartmental Center "L. Galvani", University of Bologna, Bologna, Italy

⁸Applied Biomedical Research Center, S. Orsola-Malpighi Polyclinic, Bologna, Italy

⁹Personal Genomics S.r.l., Verona, Italy

*These authors contributed equally to this work

Correspondence to: Christine Nardini, **email:** christine.nardini.rsrc@gmail.com

Paolo Garagnani, **email:** paolo.garagnani2@unibo.it

Claudio Franceschi, **email:** claudio.franceschi@unibo.it

Keywords: DNA methylation, colorectal cancer, differential analysis, network analysis, infinium human methylation 450

Received: June 08, 2016

Accepted: December 27, 2016

Published: January 10, 2017

ABSTRACT

Colorectal cancer is among the leading causes of cancer death worldwide. Despite numerous molecular characterizations of the phenomenon, the exact dynamics of its onset and progression remain elusive. Colorectal cancer onset has been characterized by changes in DNA methylation profiles, that, owing to the stability of their patterns, are promising candidates to shed light on the molecular events laying at the base of this phenomenon.

To exploit this stability and reinforce it, we conducted a meta-analysis on publicly available DNA methylation datasets generated on: normal colorectal, adenoma (ADE) and adenocarcinoma (CRC) samples using the Illumina 450k array, in the systems medicine frame, searching for tumor gene epesignatures, to produce a carefully selected list of potential drivers, markers and targets of the disease. The analysis proceeds from a differential meta-analysis of the methylation profiles using an analytical pipeline recently developed by our group [1], through network reconstruction, topological and functional analyses, to finally highlight relevant epigenomic features. Our results show that genes already highlighted for their genetic or transcriptional alteration in colorectal cancer are also differentially methylated, reinforcing -regardless of the level of cellular control- their role in the complex of alterations involved in tumorigenesis.

These findings were finally validated in an independent cohort from The Cancer Genome Atlas (TCGA).

INTRODUCTION

Colorectal cancer is one of the main causes of death from cancer worldwide [2]. The development of the disease [3] is described as a progression from a pre-malignant lesion (adenomatous polyp or adenoma, ADE) arising in the normal colon epithelium, with the potential to further degenerate into a malignant lesion (colorectal adenocarcinoma, CRC) which in turn can spread to the surrounding tissues and systemically (metastasis, MET).

Evidences accumulated in the past two decades indicate that profound genetic and epigenetic changes occur in colon epithelial cells during colorectal tumorigenesis' onset and progression [4–6]. Microsatellite and chromosomal instability have been related to the increase in genomic mutations rates that contribute to the tumor evolution [7]. Among the epigenetic modifications that are deregulated in colorectal cancer, DNA methylation has attracted great attention, thanks, also, to the rise of cost-effective genome-wide profiling methods such as the Illumina Infinium HumanMethylation27 (HM27) and HumanMethylation450 (HM450) microarrays.

In humans, DNA methylation consists of the covalent addition of a methyl group at the 5-carbon of the cytosine ring and occurs mainly in CpG dinucleotides [8]. The distribution of CpG dinucleotides and their methylation status varies widely across the genome: in the bulk of the genome CpG dinucleotides are underrepresented and tend to be pervasively methylated, while regions of high CpG density, termed CpG islands, are often found at gene promoters' sites in a non-methylated status.

Colorectal cancer, like other solid tumors, is characterized by a profound remodeling of normal DNA methylation patterns [9, 10]. Widespread hypomethylation, up to one-third of the genes and of the bulk of the genome, has already been described three decades ago [11, 12]. More recently, Timp et al. found that large (hundreds of kb) hypomethylation blocks are a universal characteristic of colorectal cancers and other solid tumors [13]. Overall, DNA hypomethylation can contribute to tumor initiation and progression by promoting genomic instability and abnormal genes' activation [14]. In addition, aberrant DNA hypermethylation of specific CpG islands has also been observed to occur in colorectal cancer. The CpG island methylator phenotype (CIMP) was described in a subset of colorectal cancers for the first time in 1999 [15] and was subsequently refined as hypermethylation of the five genes *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1* [16]. CIMP-positive tumors can be inspected with specific assays (Methylight) and have fundamental clinicopathological differences compared to CIMP-negative cases [15, 17–20], indicating the profound clinical impact of these alterations. More recently, methylation array platforms (HM27 and HM450) have been largely

used to identify differentially methylated regions (DMRs) in adenoma and colorectal cancer [21–26], highlighting subtypes that are different in terms of clinical phenotype [27, 28] and response to treatment [10].

Despite the progresses made in characterizing the epigenomic landscape of colorectal cancer, little is still known on the pathways affected by aberrant methylation and that could be deregulated in the progression of the disease. To contribute to this research, we first selected a number of robust DMRs by aggregating 3 original studies on colorectal cancer-referred hereafter as the Naumov [25], Luo [26] and Timp [13] datasets - in a meta-analysis with the specific aim to extract robust results, secured by the independent reproducibility across different datasets. Finally, we used a fourth independent dataset, from the consortium The Cancer Genome Atlas (TCGA, [29]) to validate our previous findings.

To this aim we adopted a recently proposed analytical pipeline specifically tailored on the HM450 architecture [1], based on a multivariate approach that favors the selection of regions of adjacent CpG sites with concordant changes in DNA methylation levels [1]. The selected DMRs were further interpreted in the frame of *systems medicine* [30], using networks analysis, capable to capture the complex biological relationships that exists among the elements of a list of DMRs. From there, the identification of relevant molecules in a systemic context permits to highlight informative markers or drivers of the progression of the disease.

RESULTS AND DISCUSSION

The starting point of our meta-analysis is the collection of three original studies on colorectal cancer, referred as the Naumov [25], Luo [26] and Timp [13] datasets briefly described below.

In Naumov et al. [25] the identification of DMRs was achieved with the comparison between normal tissue N1 and CRC (N1xCRC), using three options of the Illumina Methylation Analyzer (IMA, [31]) to identify DMRs: a site level test (15667 DMRs), a region-level gene-based test (2954 DMRs) and a region-level island-based test (3084 DMRs). Luo et al. [26] used cluster analysis to identify distinct epiphenotypes in adenomas (ADE) and adenocarcinomas (CRC). Using this approach, they identified two methylator phenotypes in ADE (Adenoma-High and Adenoma-Low) and three methylator phenotypes in CRC (Methyl-High, Methyl-Intermediate and Methyl-Low). Finally, Timp et al. used the *Bumphunter* algorithm to identify large (median length on the order of hundreds of kb) DMRs distinctive of solid tumors, including colon cancer and normal tissues, thus demonstrating that large hypomethylated blocks are a universal feature of solid tumors appearing early in tumor progression.

Identification of DMRs between normal colorectal tissue, ADE and CRC

Table 1 shows the results of the differential analysis. The comparison N1xN2 outputted a limited number of DMRs (42, mapping on 57 genes) common to the Naumov and Luo datasets. Unsupervised hierarchical clustering using the selected DMRs showed that samples cluster primarily according to the dataset of origin (Supplementary Figure 1). We concluded that differences between tissues N1 and N2 are subtle and mainly cohort-dependent, and we did not take this comparison into further account.

In the N1xADE comparison we identified 2657 DMRs mapping on 2180 genes that were common to the Luo and Timp datasets. Unsupervised hierarchical clustering shows a clear separation of ADE versus N1 samples (Figure 1) with around half of the DMRs being hypermethylated in ADE. Notably, for these probes, ADE samples show two distinct patterns of intermediate and high hypermethylation, confirming the presence of the distinctive epiphenotypes previously described by Luo et al. [26].

N1xCRC results in 2185 DMRs mapping on 1902 genes largely overlapping with the previous findings by Naumov et al. (See Supplementary File S1). A direct comparison with the results of Timp and coworkers was not possible, as the authors' focus was on the identification

of large hypomethylated blocks and the lists of differentially methylated CpG islands were not reported. Among the identified DMRs 85% are hypermethylated in tumors. A clear separation between N1 and CRC samples was observed after unsupervised clustering analysis (Figure 2). Importantly, like in the N1xADE comparison, we confirm the epiphenotypes previously described by Luo et al., showing that CRC samples cluster in 3 groups. The first 2 groups are clearly distinct from N1 and are characterized respectively by intermediate and high levels of hypermethylation. In the third group, CRC samples have a methylation profile more similar to N1 with whom they cluster according to the dataset of origin. It is worth mentioning that, overall, samples from the same dataset tend to cluster together. This behavior can probably be ascribed to a batch effect in the generation and/or the processing of HM450 data, although we cannot exclude the presence of a biological component related for example to the method used to obtain the biological samples or to the different geographic origin of the patients.

Inference on colorectal cancer driver mechanisms—network analysis

To explore further the biological mechanisms associated to the differentially methylated data and infer potential epigenetic drivers of the tumor lesion, we derived

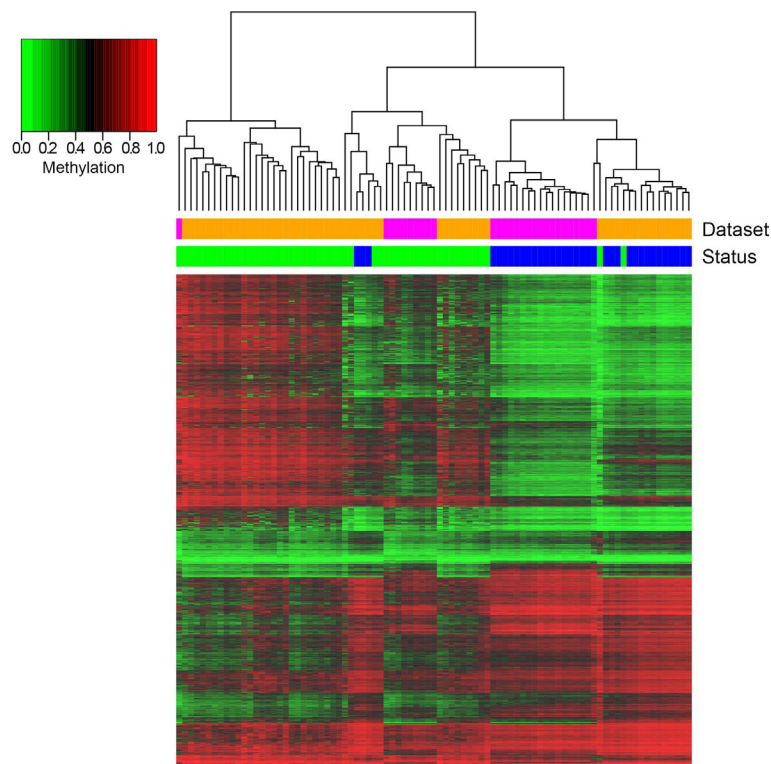


Figure 1: Hierarchical clustering of DMRs resulting from the comparison N1xADE and heatmap representation of their methylation values. Columns correspond to samples, rows correspond to DMRs (for graphical purposes only the top significantly differential CpG of each BOP is reported). Color bars indicate the status of the samples (blue: N1; green: ADE) and the dataset of origin (orange: Luo; magenta: Timp).

Table 1: DMRs identified in each comparison

Selected BOPs	N1xN2	N1xADE	N1xCRC
Naumov	7868	-	10062
Luo	277	13426	5011
Timp	-	3210	6282
Shared BOPs	42	2657	2185
Shared genes	57	2180	1902
(% hypermethylated DMRs)	83	55	85

The number of DMRs identified in each dataset is reported along with the number of shared DMRs among all datasets for the same comparison and the corresponding number of genes. The percentage of hypermethylated DMRs among the shared ones is also reported.

for each comparison the corresponding network (Figure 3, Supplementary File 3), and further analyzed each of those in three steps.

The first step highlights the hubs of the network, to identify biological key players [32]. Owing to the common cellular type used by the network reconstruction software we identified, as expected, a set of common nodes/genes/proteins (141), and in particular a set of hubs common to the N1xADE and N1xCRC networks (Supplementary

File 5). Among them we highlight tumor suppressor p53 (*TP53*), cyclin-dependent kinase inhibitor 1A (*CDKN1A*) catenin beta-1 (*CTNNB1*) and cyclin D1 (*CCND1*), crucially involved in cell proliferation and differentiation, relevant in both tumor onset and progression.

The second analysis aims at the identification of distinctive functions characterizing each phase. For this, enrichment analysis on the nodes unique to each network (134 and 116 for N1xCRC and N1xADE, respectively)

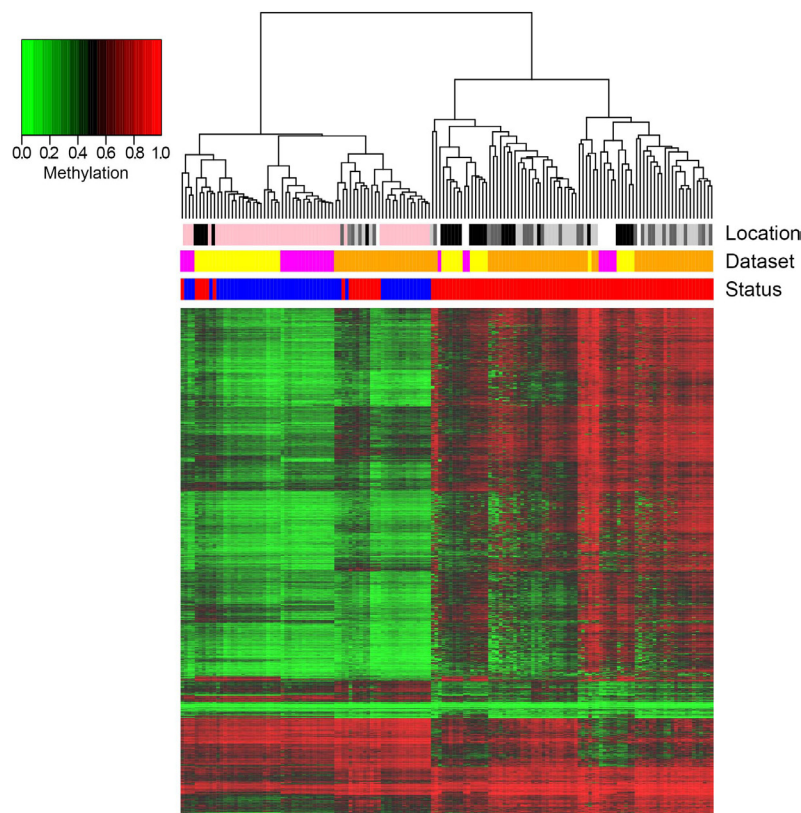


Figure 2: Hierarchical clustering of DMRs resulting from the comparison N1xCRC and heatmap representation of their methylation values. Columns correspond to samples, rows correspond to DMRs (for graphical purposes only the top significantly differential CpG of each BOP is reported). Color bars indicate the status of the samples (blue: N1; red: CRC), the dataset of origin (yellow: Naumov; orange: Luo; magenta: Timp) and the localization of the tumor (white: unknown; light grey: distal; grey: transverse; dark grey: proximal; black: rectal; pink: normal colorectal mucosa).

was performed (Supplementary File 5). No conclusive information could be derived from this analysis, both phases share, expectedly, a number of cancer related pathways, in addition to functions associated to early structural modification of the tissue and to the emergence of mesenchymal phenotypes (focal adhesion, WNT). This triggers the necessity of a third step of analysis, focusing on the differentially methylated hubs (DMH), distinctive of each phase.

DMHs represent 13.8 and 18.2% respectively of the total hubs in the N1xADE and N1xCRC networks, respectively (Supplementary File 4). This mildly increasing trend is in line with the numerous studies reporting a growing number of aberrant methylation features in cancers progression [33, 34]. In particular, the methylation levels in key regulator genes increases from normal to proliferative profiles, culminating with the methylation of hubs in N1xCRC involved in tumor progression, e.g. *NOTCH1* and *SOX2*. Following PANTHER protein class classification, we observed that most of the identified DM hub genes codify for transcription factors, signaling molecules, and cell adhesion proteins (Supplementary File 4).

In the following we describe the functions associated to some of the better known of such hubs on the basis of available literature (Table 2).

N1xADE

AXIN2 is an important gene involved in the WNT signaling [35]. The gene Keratin 18, Type I (*KRT18*) encodes for a structural protein expressed in epithelial

tissues and has been already suggested as a colorectal cancer marker [36, 37], hence confirming its importance and suggesting a mechanism of epigenetic aberration in colorectal tumors.

The hypermethylated gene V-Rel Avian Reticuloendotheliosis Viral Oncogene Homolog A (*RELA*) complexes with Nuclear Factor Of Kappa Light Polypeptide Gene Enhancer In B-Cells 1 (*NFKB1*) and forms the NFKB transcription factor. Aberrant NFKB signaling plays a role in colorectal cancer initiation and progression [38] as well as corollary processes including inflammation, immunity, differentiation, cell growth and apoptosis.

N1xCRC

All the DM hub genes in the N1xCRC comparison are hypermethylated. They include genes that codify for transcriptional factors (*TP73* and *SOX2*) and signaling molecules (*NOTCH1*, *NFKB1A* and *EGFR*). All these genes have been known to act in deregulating important pathways driving cells to tumorigenesis, although the exact correlation between the methylation and the biological effect remains to be elucidated [33]. An important hypermethylated hub is represented by the gene that expresses the insulin-like growth factor-binding protein 3 (*IGFBP3*), which belongs to the IGFBPs protein class, deeply implicated in colorectal tumorigenesis [39]. Geoges et al. [40], have shown that both *IGFBP3* and *IGFBP7* are upregulated in 68 human CRC samples, an observation further refined by Hinoue et al. [41] describing a CIMP-specific epigenetic silencing of IGF-

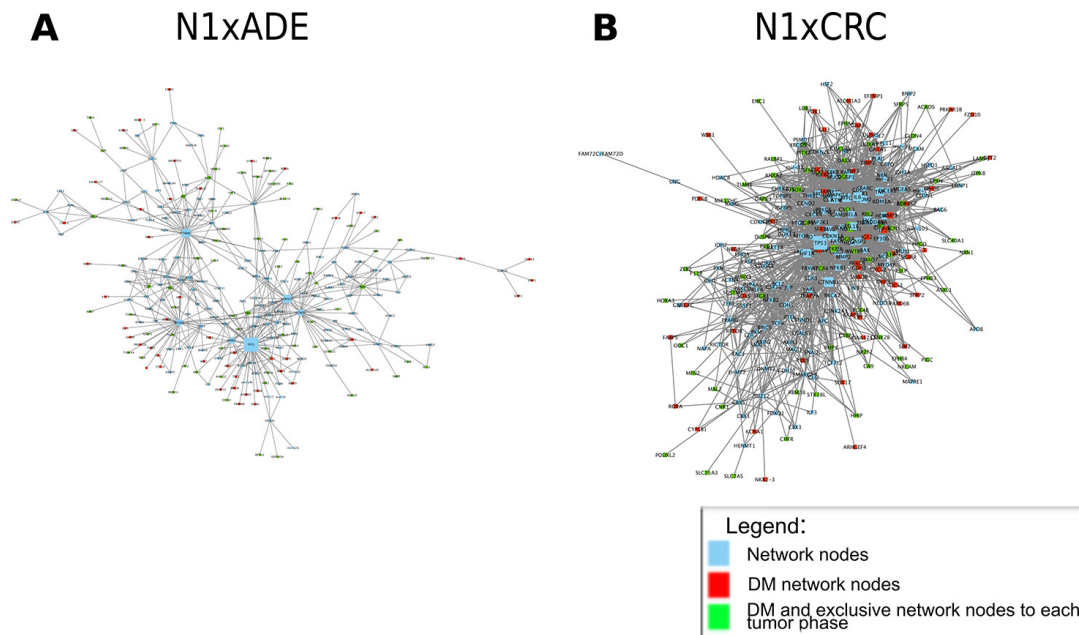


Figure 3: Network analysis. Networks were derived with the IPA software by using differentially methylated genes in the comparisons of cancer-free patient normal tissues (N1) with ADE and CRC. Panels A and B represent networks characterized by {nodes, edges} as follows A: {257, 500}; B: {275, 1994}.

Table 2: The table shows the DMH in the N1xADE and N1xCRC networks ranked by increasing degree

N1xADE	N1xCRC
<i>SMAD2</i>	<i>IGFBP3</i>
<i>H2AFY</i>	<u><i>SOX2</i></u>
<u><i>KRT18</i></u>	<i>SMAD7</i>
<i>TP73</i>	<i>IGF2</i>
<i>AXIN2</i>	<i>ETS1</i>
<i>RELA</i>	<u><i>NOTCH1</i></u>
<i>RBI</i>	<i>MMP9</i>
<i>BAX</i>	<i>TP73</i>
	<i>NFKBIA</i>
	<i>EGFR</i>

Hypermethylated genes are displayed in bold and underlined genes are phase-specific.

binding protein 7 (IGFBP7). Furthermore, Wajapeyee et al. [55] discussed the connection between IGFBP7 and the expression in CIMP-positive tumors of the mutated *BRAFV600E* (with *BRAF* being a DM hub in our MET network) implicated in oncogene-induced senescence in melanomas and colorectal cancers [42].

Altogether, the DM hub genes relate to common functions including apoptosis, TP53 cell signaling, hepatitis and differentiation as well as alterations of the he NFKB signaling. The NFKB family of transcription factors is pleiotropic and expressed in numerous cell types, it is known to play important roles in the immune response and is increasingly recognized as a crucial player in numerous steps of cancer initiation and progression where it cooperates with multiple other signaling molecules and pathways. Overall, pathological deregulations of the NFKB signaling are linked to inflammatory and autoimmune diseases [43] and in solid tumors NFKB acts as survival factor for transformed cells which would otherwise become apoptotic or senescent [44].

At the base of such functional alterations, genetic polymorphisms in numerous genes of the NFKB signaling pathway are a well described cause of increased risk of

colorectal cancer in diverse populations around the world, including the Malaysian [45, 46] and Swedish [47], Danes [48], Swedish and Chinese populations [49, 50].

The obvious and yet unclear tight relation between genomic, epigenomic, transcriptional and functional alterations has already been discussed by Weisenberger et al. in the specific context of colorectal cancer [51]. Here we add a piece of evidence to this complex landscape with the observation that *NFKBIA* presents CpG sites with stable and reproducible hypermethylation profiles across all the CRC datasets (Figure 4, line plots from other DMH in Supplementary File 6).

Independent validation: the TCGA cohort

We validated the relevance of the identified molecules (DMR and DMH) in an independent cohort from the TCGA repository including colon (COAD) and rectum tumor samples (READ).

In particular, since the TCGA dataset offered information on the age of the subjects (not available in Luo nor Naumov and only partially in Timp) we performed the analysis corrected by age, to guarantee that

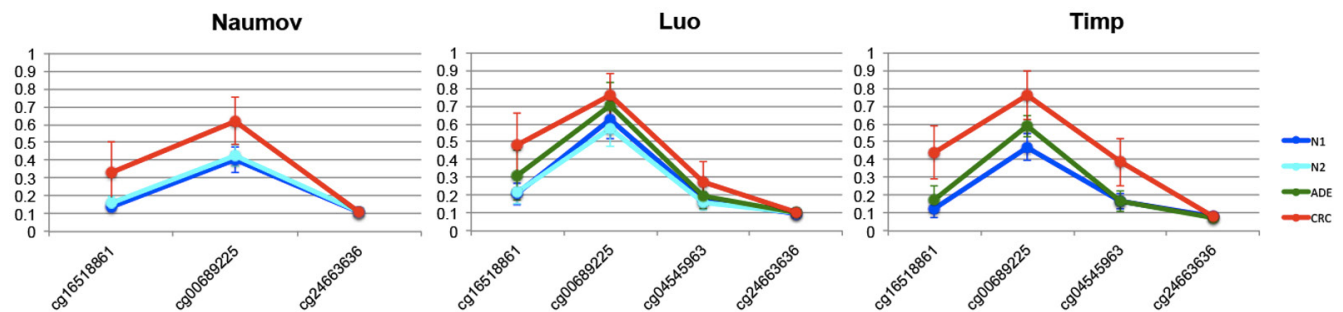


Figure 4: Comparison of methylation profiles of NFKBIA gene of N1xCRC. The lines show mean methylation values and standard deviation for each CpG probe within the shore of chr14:35873047-35873990 island in the NFKBIA gene for the following datasets: Naumov (information on CpG cg04545963 was not available), Luo and Timp.

the results of the differential analysis were not affected by this parameter, known to impact on methylation [52].

From the differential analysis (q -value < 0.001) we obtained a list of 14578 DMRs, mapping to 9156 genes. Of the 2185 DMRs resulting from the N1xCRC GEO meta-analysis, 2030 were confirmed in the TCGA analysis. Further, 122/125 network DM nodes and all 10/10 DMH were validated in the TCGA results (Supplementary File 7).

As the great majority of DMRs emerging from the meta-analysis was confirmed in the TCGA cohort, we can infer that the CRC-specific DMRs that we identified are largely independent from the age of the subjects.

As an additional test we considered the DNA methylation levels of the 10 DM hub genes (corresponding to 12 BOPs) with respect to the age of the TCGA subjects and no interaction between disease status and age was evident (Supplementary File 8).

To deepen our understanding on the biological meaning and potential clinical role of the identified features (DMR and DMH), we used data mining techniques to explore how samples tend to cluster when similarity is guided by all DMR features or only by DMH.

The DAPC analysis (Figure 5) shows that DMRs can not only separate, as expected, tumoral from normal tissue, but also permit to identify largely overlapping but still distinct clusters referring to the tumor stage.

Interestingly, this distinction is not only preserved, but mildly improved when guided only by the 10 DMH. This indicates that the additional systemic analysis run on top of the differential analysis allows to isolate a handful

of features whose biological meaning is indeed able to capture the essence of the full list of DMR, and further to remove some of the noise, allowing a clearer distinction of the four stages.

Finally, owing to the larger number of clinicopathological information available in the TCGA dataset (compared to GEO) we also evaluated the DNA methylation of the 10 DM hub genes according to the mutational status of *BRAF* and *KRAS* genes (Supplementary File 9). No statistically significant differences were found when DNA methylation of these CpG sites was analyzed according to the mutational status of *BRAF* and *KRAS* genes, despite a trend towards higher methylation levels for some of the hubs (for example, *SOX2* and *IGF2*) in patients carrying mutations in *BRAF*.

Overall, in this study, we observe and confirm the high reproducibility of methylation data, and hence the importance of this approach in identifying stable markers, drivers, targets of the disease. Finally, we highlight an exemplar case of methylated key molecule, emerging from the systemic approach we used, embodied by *NFKBIA*, a well-known genetic risk factor for colorectal cancer, here also emerging as in important epigenotype.

MATERIAL AND METHODS

Datasets

Infinium 450k datasets were downloaded from the Gene Expression Omnibus–GEO repository [53] using *GPL13534* (identifier of the Infinium 450k platform) and

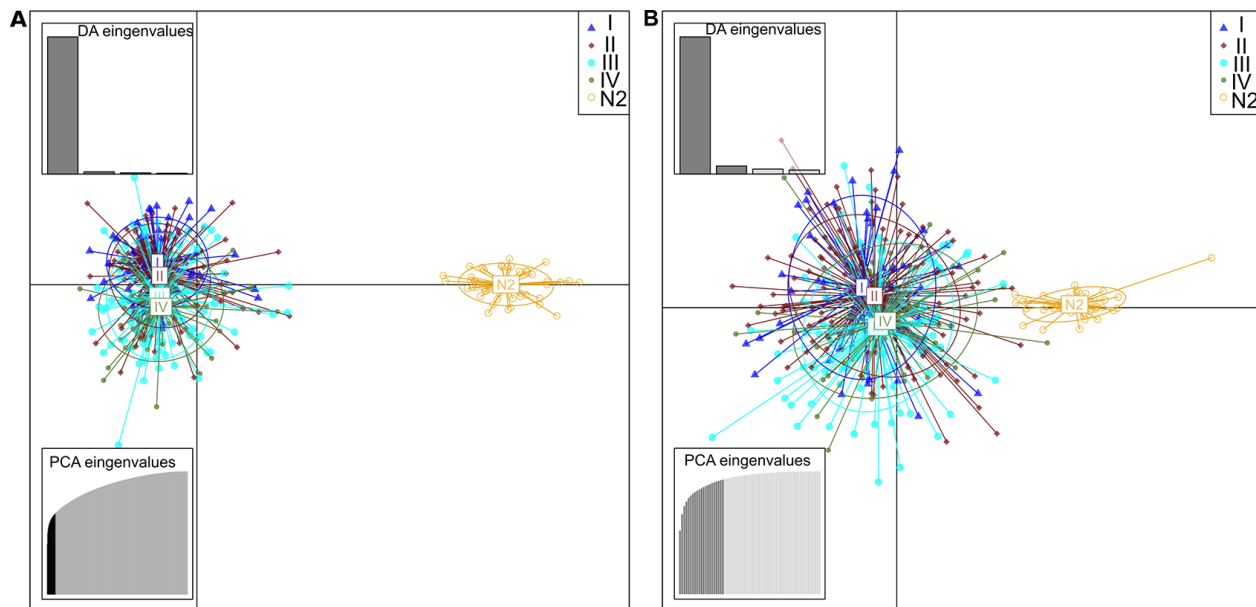


Figure 5: Scatterplots resulted from DAPC analysis of TCGA data. These scatterplots show the first two principal components of the DAPC of data simulated according to hierarchical islands model. Clusters are shown by different colours and ellipses, while dots represent individual samples–N2 represent normal samples from affected individuals, and the groups I to IV are the corresponding I–IV stage CRC samples. The chart A is referring to DM CpG that emerged from previous analysis of N1xCRC datasets and the chart B were performed with CpGs related to the CRC networks DMH.

colorectal cancer as search terms. As to January 1st 2016 we identified, based on these terms, 3 studies from colonic fresh-frozen tissues (Table 3): Naumov et al. [25] analysed 22 CRC samples, their matched healthy mucosa samples (N1) and 19 cancer-unrelated normal colon tissues (N2); Luo et al. [26] studied 41 normal colon tissues (N1), 42 colon adenomas (ADE) and 64 colorectal adenocarcinoma (CRC); Timp et al. [13] performed an epigenetic analysis of different cancer types, including colorectal cancer, and in particular measured DNA methylation in 18 normal colon tissues (N1), 10 adenoma (ADE), 9 colorectal adenocarcinoma (CRC) and 16 metastases (MET). As our focus is on meta-analysis, to guarantee robustness of the results we only included ADE and CRC samples for which more than one data-set is available. This leaves out of the picture the MET phase, for which one dataset only is available in Timp, while the heterogeneity of this phase (intense proliferation and different metastases localizations including liver and lung) strongly demands the natural filtering of the meta-analysis.

In order to validate our findings we used an independent Infinium 450k dataset from The Cancer Genome Atlas (TCGA) project [29], including data from colorectal adenocarcinoma (COAD) and rectum adenocarcinoma (READ). Methylation data, demographic and clinical information of 418 samples were downloaded from the TCGA data portal (<http://firebrowse.org>, [54]): they included 323 colon (285 tumor samples, 38 of which had also a normal counterpart) and 95 rectum samples (88 tumor samples, 7 of which had also a normal counterpart) [55].

Differential analysis

DMRs were defined and selected according to the multivariate approach proposed by Bacalini et al. [1] to compare the methylation of blocks of 3 or more adjacent CpG probes (Block Of Probes, BOPs) between groups. In the present meta-analysis, the focus is on BOPs localized in CpG islands and CpG islands-surrounding sequences (shores and shelves) that map on genic regions. The rationale underlying this approach is that, compared to punctual changes in DNA methylation of a single CpGs, concomitant alterations in adjacent CpGs are more likely to modify the chromatin structure and hence gene expression and, more in general, to affect biological functions [1]. Correction for multiple hypotheses comparison was done by Benjamini-Hochberg (False discovery rate–FDR, q -value < 0.001, [56]).

To guarantee robustness while avoiding loss of information we requested that the retained candidates show unanimous statistical significance across all datasets, without further limitations on the absolute value of the differential methylation (delta), owing to the recent findings on the relevance of minor changes in DNA methylation levels in terms of phenotypical consequences [57]. Delta distributions are nevertheless

provided in Supplementary Figures 2 and 3 showing that less than 2% of the samples have delta below 0.05.

Three pairwise comparisons (N1xN2, N1xADE and N1xCRC) were run. In the TCGA validation cohort, the 45 N2 samples were compared with 373 CRC samples. In this last comparison, associations were corrected for the age of the subjects, an information that was not available for the 3 GEO datasets. In the following, if not explicitly stated, DMRs always refer to significantly differentially methylated BOPs.

Network analysis

For each comparison, we identified robust results by selecting across all 3 datasets only common and concordant (coherently hyper- or hypo-methylated) DMRs (Supplementary File 1).

Network analysis was adopted to gain understanding on the biological interconnections occurring among the DMRs identified in each meta-list: DMRs genes were set as nodes of the network, and connecting biochemical relations were reconstructed importing the gene lists into Ingenuity Pathway Analysis (IPA, Qiagen v.1-04 [58], (Supplementary File 2). To improve the reconstruction, IPA adds external (non-differentially methylated) items including genes and genes products, mRNA, miRNA, proteins, from here on interchangeably named nodes. These additional nodes are extracted from IPA internal data-base and filtered by the *Tissues & Cell Lines* parameter here set to *large intestine* and *colon cancer cells*. To guarantee proper reconstruction, Qiagen recommends to output from each meaningful list of molecules a maximum of 10 networks with a maximum of 70 nodes each. Successively, all networks sharing nodes (interconnected) were merged in one large interconnected map, leading to one network per comparison.

From there, we first explored the networks' nodes based on their topological characteristics, known to match biological relevance [59]. In particular, topological analyses were performed on nodes degree, i.e. the number of connections stemming from a node/gene [60]. Degree is an intuitive measure to determine the importance of a node, indicating that the more the node is connected, the more genes/proteins/molecules it interacts with and the more pathways it is involved in. Nodes with degree above a certain threshold (here 80th percentile [61]) are defined as *hubs*. For such topological analyses, the network generated by IPA was input into Cytoscape v 3.3.0 [62], a network analyzer software (Supplementary File 3).

We further explored network genes' meaning with functional enrichment analysis for KEGG pathways with WebGestalt online tool [63] on a selection of nodes belonging exclusively to each tumor stage i.e. by excluding the nodes shared by all networks (Supplementary File 2). This selected list of genes was manually curated first with PANTHER [64] to identify the protein class and the general

Table 3: GEO datasets used in the meta-analysis

	Naumov	Luo	Timp
GEO ID	GSE42752	GSE48684	GSE53051
N1	20	17	18
N2	21	24	–
ADE	–	42	10
CRC	22	64	9
Total samples	63	147	53

N1: Normal colorectal tissue from cancer-free subjects; N2: Normal colorectal tissue from subjects with colorectal cancer; ADE: Colorectal samples from adenoma lesions; CRC: Colorectal samples from CRC patients.

functional importance of the hubs (See Supplementary File 4) and then from literature to highlight key regulators of the disease and discuss potential mechanisms driver of tumor onset and progression.

Finally, a list of the topologically relevant DMRs (DMH, i.e. differentially methylated hubs) in each network was filtered to select the genes exclusive to each tumor stage, leading to the final discussion on genes of topological and functional relevance, characterized by an aberrant methylation state in the disease.

Statistical analyses

Statistical analyses were performed using the computing environment R. Discriminant Analysis of Principal Componentes (DAPC) was performed using the R package *adegenet* [65].

ACKNOWLEDGMENTS AND FUNDING

We are grateful to all unknown individuals who participated in each study used in this work. This work was supported by the European Union's Seventh Framework Programme (grant agreement 602757 "HUMAN"), by the European Union's H2020 Programme (grant agreement 634821 "PROPAG-AGEING") and by the National Counsel of Technological and Scientific Development (CNPq) Science Technology and Innovation (MCTI) (n. 200891/2014-6).

CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

REFERENCES

- Bacalini G, Boattini A, Gentilini D, Giampieri E, Giuliani C, Fontanesi E, Remondini D, Capri M, Del A, Luiselli D, Vitale G, Mari D, Castellani G, et al. A meta-analysis on age associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging* (Albany NY). 2015; 7:97–109. doi: 10.18632/aging.100718.
- World Health Organization [Internet]. [cited 2014 Apr 20]. Available from: <http://www.who.int/mediacentre/factsheets/fs297/en/>.
- Muto T, Bussey HJR, Morson BC. The evolution of cancer of the colon and rectum. *Cancer*. 1975; 36:2251–70.
- Valinluck V, Grady W. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol*. 2011; 8:686–700.
- Goel A, Boland CR. Epigenetics of Colorectal Cancer. *Gastroenterology*. Elsevier Inc.; 2012; 143:1442–60.
- Garagnani P, Franceschi C. Colorectal Cancer Microenvironment: Among Nutrition, Gut Microbiota, Inflammation and Epigenetics. *Curr Pharm Des*. 2013; 19:1–14.
- Burrell RA, Mcgranahan N, Bartek J, Swanton C. The causes and consequences of genetic. 2013;
- Ulrey CL. The impact of metabolism on DNA methylation. *Hum Mol Genet*. 2005; 14:139–47.
- Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007; 128:683–92.
- Okugawa Y, Grady WM, Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology*. 2015; 149:1204–25.
- Feinberg A, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983; 301:89–92.
- Gama-Sosa M, Slagel V, Trewyn R, Oxenhandler R, Kuo K, Gehrke C, et al. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 1983; 11:6883–94.
- Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, Feinberg AP, Irazarry RA. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med*. 2014; 6:61.
- Gonzalo S. Epigenetic alterations in aging. *J Appl Physiol*. 2010; 109:586–97.
- Toyota M, Issa J. The role of DNA hypermethylation in human neoplasia. *Electrophoresis*. 2000; 21:329–33.
- Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with

- BRAF mutation in colorectal cancer. *Nat Genet.* 2006; 38:787–93.
17. van Rijnsoever M, Grieu F, Elsaleh H, Joseph D, Iacopetta B. Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. *Gut.* 2002; 51:797–802.
 18. Hawkins N, Norrie M, Cheong K, Mokany E, Ku S-LL, Meagher A, O'Connor T, Ward R. CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability. *Gastroenterology.* 2002; 122:1376–87.
 19. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology.* 2005; 129:837–45.
 20. Issa J. CpG island methylator phenotype in cancer. *Nat Rev Cancer.* 4:988–93.
 21. Hinoue T, Weisenberger DJ, Lange CPE, Shen H, Byun H, Berg D Van Den, Malik S, Pan F, Noushmehr H, Dijk CMV, Tollenaar RAEM, Laird PW. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* 2012; 22:271–82.
 22. Kibriya MG, Raza M, Jasmine F, Roy S, Paul-Brutus R, Rahaman R, Dodsworth C, Rakibuz-Zaman M, Kamal M, Ahsan H. A genome-wide DNA methylation study in colorectal carcinoma. *BMC Med Genomics.* 2011; 4:1–16.
 23. Kim YH, Lee HC, Kim SY, Yeom Y Il, Ryu KJ, Min BH, Kim DH, Son HJ, Rhee PL, Kim JJ, Rhee JC, Kim HC, Chun HK, et al. Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations. *Ann Surg Oncol.* 2011; 18:2338–47.
 24. Barat A, Ruskin HJ. Comparative correlation structure of colon cancer locus specific methylation: characterisation of patient profiles and potential markers across 3 array-based datasets. *J Cancer.* 2015; 6:795–811.
 25. Naumov VA, Generozov EV, Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov SV, Alekseev MV, Shelygin YA, Govorun VM. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium Human Methylation 450 BeadChips. *Epigenetics.* 2013; 8:921–34.
 26. Luo Y, Wong C, Kaz AM, Dzieciatkowski S, Kelly T, Lutterbaugh JD, Shrubsole MJ, Zheng W, Sanford D. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. *Gastroenterology.* 2014; 147:418–29.
 27. Ren Z, Wang W, Li J. Identifying molecular subtypes in human colon cancer using gene expression and DNA methylation microarray data. *Int J Oncol.* 2016; 48:690–702.
 28. Zhou F, Huang X, Zhang Z, Chen Y, Liu X, Xing J, He X. Functional polymorphisms of ITGB1 are associated with clinical outcome of Chinese patients with resected colorectal cancer. *Cancer Chemother Pharmacol.* 2015; 75:1207–15.
 29. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* [Internet]. 2013; 45:1113–20. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3919969&tool=pmcentrez&rendertype=abstract>.
 30. Heinzel A, Perco P, Mayer G, Oberbauer R, Lukas A, Mayer B. From molecular signatures to predictive biomarkers: modeling disease pathophysiology and drug mechanism of action. *Front cell Dev Biol.* 2014; 2:1–11.
 31. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S. IMA: An R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics.* 2012; 28:729–30.
 32. Barabási A, Oltvai ZN. Network biology : understanding the cell' s functional organization. *Nat Genet.* 2004; 5:101–13.
 33. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet.* 2007; 8:286–298.
 34. Heyn H, Li N, Ferreira HHJ, Moran S, Pisano DG, Gomez A, Diez J. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci USA.* 2012; 109:10522–10527.
 35. Jho E, Zhang T, Domon C, Joo CK, Freund JN, Costantini F. Wnt/ β -Catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol Cell Biol.* 2002; 22:1172–83.
 36. Drew JE, Farquharson AJ, Mayer CD, Vase HF, Coates PJ, Steele RJ, Carey FA. Predictive Gene Signatures: Molecular Markers Distinguishing Colon Adenomatous Polyp and Carcinoma. *PLoS One.* 2014; 9:e113071.
 37. Kwaa R, Looi K, Omary MB. Absence of keratins 8 and 18 in rodent epithelial cell lines associates with keratin gene mutation and DNA methylation: Cell line selective effects on cell invasion. *Exp Cell Res.* 2015; 335:12–22.
 38. Thoms HC, Dunlop MG, Stark LA. p38-mediated inactivation of cyclin D1/cyclin-dependent kinase 4 stimulates nucleolar translocation of RelA and apoptosis in colorectal cancer cells. *Cancer Res.* 2007; 67:1660–9.
 39. Vigneri PG, Tirrò E, Pennisi MS, Massimino M, Stella S, Romano C, Manzella L. The Insulin/IGF System in Colorectal Cancer Development and Resistance to Therapy. *Front Oncol.* 2015; 5:230.
 40. Georges R, Adwan H, Hamdi H, Hielscher T, Linnemann U, Berger M. The insulin-like growth factor binding proteins 3 and 7 are associated with colorectal cancer and liver metastasis. *Cancer Biol Ther.* 2011; 12:69–79.
 41. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, Whitehall VL, Leggett BA, Laird PW. Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. *PLoS One.* 2009; 4:e8357.
 42. Wajapeyee N, Serra RW, Zhu X, Mahalingam M, Green MR. Oncogenic BRAF Induces Senescence and Apoptosis through Pathways Mediated by the Secreted Protein IGFBP7. *Cell.* 2008; 132:363–74.

43. Hoesel B, Schmid JA. The complexity of NF- κ B signaling in inflammation and cancer. *Mol Cancer*. 2013;12:86.
44. Hagemann T, Lawrence T, McNeish I, Charles KA, Kulbe H, Thompson RG, Robinson SC, Balkwill FR. "Re-educating" tumor-associated macrophages by targeting NF- κ B. *J Exp Med*. 2008; 205:1261–8.
45. Mohd Suzairi M, Tan S, Ahmad Aizat A, Mohd Aminudin M, Siti Nurfatimah, MS Andee D, Ankathil R. The functional -94 insertion/deletion ATTG polymorphism in the promoter region of NFKB1 gene increases the risk of sporadic colorectal cancer. *Cancer Epidemiol*. 2013; 37:634–638.
46. Tan SC, Suzairi MSIM, Aizat AAA, Aminudin MM, Nurfatimah MSS, Bhavaraju VMK, Biswal BM, Ankathil R. Gender-specific association of NFKB1A promoter polymorphisms with the risk of sporadic colorectal cancer. *Med Oncol*. 2013; 30:1–6.
47. Lewandera A, Reddy Butchia A, Gao J, Heb L, Lindblom A, Armand G, Carstensen J, Zhangf K, Group TSL-RCCS, Sun X. Polymorphism in the promoter region of the NFKB1 gene increases the risk of sporadic colorectal cancer in Swedish but not in Chinese populations. *Scand J Gastroenterol*. 2007; 42:1332–8.
48. Andersen V, Christensen J, Overvad K, Tjønneland A, Vogel U. Polymorphisms in NFKB, PXR, LXR and risk of colorectal cancer in a prospective study of Danes. *BMC Cancer*. 2010; 10:484.
49. Song S, Chen D, Lu J, Liao J, Luo Y, Yang Z, Fu X, Fan X, Wei Y, Yang L, Wang L, Wang J. NF κ B1 and NF κ B1A polymorphisms are associated with increased risk for sporadic colorectal cancer in a southern Chinese population. *PLoS One*. 2011;6:e21726.
50. Gao J, Pfeifer D, He L, Qiao F, Zhang Z, Arbman G, Wang Z, Jia C, Carstensen J, Sun X. Association of NFKB1A polymorphism with colorectal cancer risk and prognosis in Swedish and Chinese populations. *Scand J Gastroenterol*. 2007; 42:345–50.
51. Weisenberger DJ. Characterizing DNA methylation alterations from the cancer genome atlas. *J Clin Invest*. 2014; 124:17–23.
52. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:R115.
53. GEO - Gene Expression Omnibus [Internet]. [cited 2016 Apr 12]. Available from: <http://www.ncbi.nlm.nih.gov/geo/>.
54. Firebrowse [Internet]. [cited 2016 Jun 11]. Available from: <http://firebrowse.org/>
55. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editors. *New York NS 2010. AJCC cancer staging manual (8th ed)*. 2016.
56. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 1995; 57:289–300. 289–300.
57. Leenen FAD, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin Epigenetics*. *Clinical Epigenetics*; 2016; 8:92.
58. IPA - Ingenuity Pathway Analysis [Internet]. [cited 2016 Apr 20]. Available from: <http://www.ingenuity.com/>
59. Lee Y, Li H, Li J, Rebman E, Achour I, Regan KE, Gamazon ER, Chen JL, Yang XH, Cox NJ, Lussier YA. Network models of genome-wide association studies uncover the topological centrality of protein interactions in complex diseases. *J Am Med Informatics Assoc*. 2013; 20:619–29.
60. Borgatti SP. Centrality and network flow. *Soc Networks*. 2005; 27:55–71.
61. Wu G, Zhu L, Dent JE, Nardini C. A comprehensive molecular interaction map for rheumatoid arthritis. *PLoS One*. 2010; 5:e10135.
62. 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; 13:2498–504.
63. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res*. 2013; 41:77–83.
64. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H, Thomas PD. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res*. 2005; 33:284–8.
65. Jombart T, Caitlin C. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. *Imp Coll London - MRC Cent Outbreak Anal Model*. 2015;1–43.