RESEARCH

Association between DNA methylation profile and malignancy in follicular-patterned thyroid neoplasms

Ornella Affinito1,*, Paolo Salerno1,*, Alfonso D’Alessio1, Mariella Cuomo1, Ermanno Florio1, Francesca Carlomagno1,2, Agnese Proietti3, Riccardo Giannini4, Fulvio Basolo4, Lorenzo Chiariotti1,2, Sergio Cocozza1 and Massimo Santoro1

1Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università Federico II, Napoli, Italy
2Istituto di Endocrinologia ed Oncologia Sperimentale, Consiglio Nazionale delle Ricerche, Napoli, Italy
3Division of Anatomical Pathology, University Hospital of Pisa (AO-UP), Pisa, Italy
4Department of Surgical, Medical, Molecular Pathology and Critical Area (Anatomical Pathology Section), University of Pisa, Pisa, Italy

Correspondence should be addressed to L Chiariotti or S Cocozza: lorenzo.chiariotti@unina.it or sergio.cocozza@unina.it

*(O Affinito and P Salerno contributed equally to this work)

Abstract

Molecular differentiation between benign (follicular thyroid adenoma (FTA)) and malignant (follicular thyroid carcinoma (FTC)) thyroid neoplasms is challenging. Here, we explored the genome-wide DNA methylation profile of FTA (n.10) and FTC (n.11) compared to normal thyroid (NT) (n.7) tissues. FTC featured 3564 differentially methylated CpGs (DMCpG), most (84%) of them hypermethylated, with respect to normal controls. At the principal component analysis (PCA), the methylation profile of FTA occupied an intermediate position between FTC and normal tissue. A large fraction (n. 2385) of FTC-associated DMCpG was related (intragenic or within 1500 bp from the transcription start site) to annotated genes (n. 1786). FTC-hypermethylated genes were enriched for targets of the Polycomb transcriptional repressor complex and the specific histone H3 marks (H3K4me2/me3-H3K27me3) found in chromatin domains known as ‘bivalent’. Transcriptome profiling by RNAseq showed that 7.9% of the DMCpGs-associated genes were differentially expressed in FTC compared to NT, suggesting that altered DNA methylation may contribute to their altered expression. Overall, this study suggests that perturbed DNA methylation, in particular hypermethylation, is a component of the molecular mechanisms leading to the formation of FTC and that DNA methylation profiling may help differentiating FTCs from their benign counterpart.

Key Words
- DNA methylation
- follicular thyroid neoplasms
- integrative analysis
- polycomb genes
- bivalent chromatin domains

Introduction

Well-differentiated papillary carcinoma (PTC) (about 85%) and follicular carcinoma (FTC) (about 10%) account for the vast majority of thyroid cancers (Nikiforova & Nikiforov 2009, Fagin & Wells 2016, Giordano 2018). Non-autonomous benign follicular thyroid adenomas (FTAs) present as solitary nodules lacking capsular or lymphovascular invasion. The differential diagnosis between FTC and PTC may be challenging (Hegedüs 2004, Sobrinho-Simões et al. 2011, Fagin & Wells 2016, Giordano 2018).
FTA and FTC share similar genetic lesions suggesting that these two neoplasms may be representative of different steps of a common biological process (Zhu et al. 2003, Arora et al. 2008, Nikiforova & Nikiforov 2009, Nikiforov et al. 2009, Fagin & Wells 2016, Jung et al. 2016, Yoo et al. 2016, Giordano 2018). Recent gene expression profiling by mRNA and miRNA arrays has confirmed this hypothesis (Dom et al. 2018). The driver mutations that are more frequently detected in FTC involve missense mutations in the H/K/N-RAF genes (up to 53% of the cases) and, less commonly, the PAX8-PPARG fusion (19–63% of the cases). Albeit with a reduced prevalence, RAS mutations (up to 27%) and PAX8-PPARG fusion (up to 55%) can also be found in FTA (Arora et al. 2008). Somatic mutations of PTEN and PI3K are found in about 10% FTC (Xing 2010). TERT promoter mutations are found in FTC featuring an aggressive disease course (Wang et al. 2014).

FTC features more common chromosome copy number variations than FTA (Roque et al. 2003, Jung et al. 2016, Yoo et al. 2016). Several additional genetic lesions have been identified in FTA and/or FTC (Jung et al. 2016, Karunamurthy et al. 2016, Yoo et al. 2016).

Neoplasms may exhibit alterations of DNA methylation pattern (Baylin & Jones 2011, Dawson & Kouzarides 2012, Flavahan et al. 2017). DNA methylation involves the 5-carbon of cytosine residues (5-methylcytosine, 5mC) that typically map in the context of CpG dinucleotides. CpG-rich regions (CpG islands, CGIs) are usually found near the transcription start site (TSS) of gene promoters. CGIs-adjacent DNA regions are named ‘shores’ (1 to 2000bp from CGI) or ‘shelves’ (2001 to 4000bp from CGI); ‘open sea’ is located >4000bp from CGIs. In cancer cells, CGIs typically become hypermethylated, while CGPs outside CGIs more commonly become hypomethylated (Baylin & Jones 2011, Dawson & Kouzarides 2012, Chen et al. 2017, Flavahan et al. 2017). DNA methylation of CGIs mapping in gene promoters is able to silence their expression. Thus, CGI methylation may silence anti-proliferation or pro-differentiation genes, thereby acting as a mechanism, alternative to direct mutations, to inactivate tumor suppressors (Ohm et al. 2007, Flavahan et al. 2017). Genes that are hypermethylated in different cancer types are enriched for those that, in embryonal stem cells, are simultaneously marked by the repressive histone H3 mark, H3K27me3 (tri-methylation of lysine 27 of histone H3), which is mediated by the Polycomb group (PcG) of transcriptional repressors, and the active transcription marks, H3K4me2/me3 (di- or tri-methylation of lysine 4 of histone H3). Such chromatin arrangement is called ‘bivalent’ and is thought to maintain under-check the transcription of genes that contribute to lineage specification during development. Indeed, the ‘bivalent’ chromatin state sets low gene expression levels allowing, at the same time, prompt gene expression switch-on during development (Schlesinger et al. 2007, Widschwendter et al. 2007, Ben-Porath et al. 2008, Meissner et al. 2008, Kim et al. 2010, Easwaran et al. 2012).

Several studies have explored changes in DNA methylation pattern in thyroid neoplasms. The TCGA (Cancer Genome Atlas) study reported the methylation profile of a large cohort of PTC cases and identified four distinct tumor clusters termed ‘classical 1’ and ‘classical 2’ (mainly featuring classical PTC morphology (CVPTC)) and ‘follicular’ and ‘CpG island methylated cluster’ (mainly featuring a follicular architecture (FVPTC)) (Cancer Genome Atlas Research Network 2014). A global DNA hypomethylation profile was identified in PTC in several additional studies (Ellis et al. 2014, White et al. 2016, Beltrami et al. 2017), with FVPTC featuring a more variable methylation profile (Ellis et al. 2014). Chen and coworkers showed that BRAF-mutant CVPTC samples primarily featured DNA hypomethylation in open sea and shore/shelf regions (Chen et al. 2017). DNA methylation profile of FTC has also been explored. Mancikova and coworkers, using 27K arrays, were the first to report that FTCs featured significant DNA hypermethylation and that they clustered together with FVPTCs and separately from FTAs, normal samples and the majority of CVPTCs (Mancikova et al. 2014). By using 450K arrays, Bisarro dos Reis confirmed that while DNA hypomethylation was predominant in PTC, hypermethylation was present in FTC and benign thyroid lesions. However, benign lesions showed significantly less numerous differentially methylated CpGs (222 hypomethylated and 1531 hypermethylated) with respect to FTC (1475 hypomethylated and 4100 hypermethylated) (Bisarro Dos Reis et al. 2017). Importantly, the same study also suggested prognostic implications of the DNA methylation profile, with massive hypomethylation in aggressive (PDTC/ATC) thyroid cancer subtypes (Bisarro Dos Reis et al. 2017). Recently, Yim et al. (2019) showed that a DNA methylation-based molecular method, named Diagnostic DNA Methylation Signature approach (DDMS), features high sensitivity/specificity in the identification of malignant and benign thyroid nodules.

By a comprehensive methylation analysis with the high-density human methylation 450k array, here we confirm that FTC is characterized by a global DNA hypermethylation profile and identify a methylation profile that may help distinguishing FTC from FTA and normal thyroid tissue.
Materials and methods

Tissue samples

Formalin-fixed paraffin embedded (FFPE) samples were retrospectively collected from the archive of Department of Surgical, Medical, Molecular Pathology and Critical Area (Anatomical Pathology Section), University of Pisa. All FTC samples, with sufficient available tissue, present in the archive were included in the study; FTA samples were randomly selected from those available in the archive. Normal samples (NT) were from the contralateral lobe of patients affected by FTC (Table 1). This retrospective study was conducted anonymously and on archival material; therefore, ethical committee approval was not required. The study conforms to the Principles of the Helsinki Declaration of 1975. Both informed and surgical consent were achieved 1 day before the operation.

Nucleic acids extraction and RAS and PPARG genotyping

For RAS gene analysis, genomic DNA was purified from unstained tissue slices using the QIaAmp DNA Mini kit (Qiagen), following the manufacturer's instructions and quantified by the ND-1000 spectrophotometer (NanoDrop, Thermo Fisher Scientific). DNAs were analyzed for the presence of somatic mutations in N-RAS (exon 3), K-RAS (exon 2) and H-RAS (exon 3) genes by high resolution melting PCR. Briefly, the reaction was prepared by using the Hot StarTaq Master Mix (Qiagen), a DNA-binding dye (EvaGreen, Biotium Inc., Fremont, CA, USA) and 100 ng of each DNA; the primer pairs used for the RAS genotyping were previously described (Macerola et al. 2018). Samples with altered melting curves were purified and analyzed by direct sequencing (AbiPrism 3130 Genetic Analyzer, Applied Biosystems) using the same PCR primer pairs. For PPARG analysis, total RNA was isolated from FFPE thyroid tissues with the RNeasy FFPE Kit (Qiagen) according to the manufacturer's instructions. RNAs (500 ng) were retro-transcribed by using the RevertAid First Strand cDNA Synthesis Kit (Sigma-Aldrich). PAX8/PPARG rearrangements were detected by real-time PCR on a Rotor Gene Q instrument (Qiagen) with primers designed to flank the fusion point according to a previously described procedure (Nikiforov et al. 2009).

Genome-wide DNA methylation profiling and data processing

Microarray analysis was outsourced to Genomix4life S.R.L. (Baronissi, Salerno, Italy). DNA suitability for the analysis was assessed with the FFPE QC Kit (Illumina). Samples that passed the QC test were processed by the Infinium HD FFPE Restore Kit (Illumina) according to manufacturer's instruction. Restored DNA (500 ng) was bisulfite converted with the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer’s recommendations. Bisulfite converted DNA (200 ng) was used for the analysis of whole-genome methylation using the HumanMethylation 450K BeadChip (Illumina); this array contains 485,577 probes covering 21,231 (99%) RefSeq genes. Briefly, bisulfite-converted DNA was whole-genome amplified for 20 h followed by end-point fragmentation. Fragmented DNA was precipitated, denatured and hybridized to the BeadChips for 20 h at 48°C. The BeadChips were washed and the hybridized primers were extended and labeled before scanning by the Illumina iScan system. The GenomeStudio software (version 2011.1; Illumina) was used for the extraction of DNA methylation signals from scanned arrays. Methylation values of each CpG site in each sample were measured as β-values, defined as the ratio of the methylated probe intensity to the sum of both methylated and unmethylated probes; thus, β-values range from 0 (unmethylated) to 1 (methylated).

To exclude technical and gender-specific biases, we excluded probes with a detection P value >0.01 (n. 63,217), with missing data in all samples (n. 5620), associated with chromosome X or Y (n. 9485) or to SNPs (n. 60), containing CHG or CHHG (H = A, T or C) (n. 2035) and cross-hybridizing (n. 25,920) (Chen et al. 2013). Constitutively methylated (β value ≥0.8; n. 17,091) or unmethylated (β value ≤0.2; n. 67,897) probes in all samples were also excluded. The remaining autosomal 294,252 CpG probes were used in the analysis. Differentially methylated CpGs (DMCpG) were identified by comparing FTC or FTA with NT using the limma package, adopting an adjusted (FDR) P value <0.05 and mean delta beta (Δβ) ≤−0.2 or ≥+0.2. Data will be made available upon request. A heatmap was generated by using the R gplots package by applying hierarchical clustering with complete linkage and using the Euclidean distance as similarity measure.

The DMCpG distribution among CpG islands (CGI), shores, shelves or open sea as well as the occurrence in enhancers, gene regulatory regions, gene bodies or intergenic regions was based on Illumina annotations.

Datasets used in this study

Results were compared to three DNA methylation microarray datasets available in GEO
RNA sequencing

Transcriptome sequencing of FTC (n. 8) and NT (n. 6) samples was outsourced to Genomix4life S.R.L. RNA quality was assessed with the TapeStation 4200 instrument (Agilent Technologies). On average, RIN values of 2.14 (range 1.6 to 2.7) were obtained. Indexed libraries were prepared, with the TruSeq Stranded totalRNA Sample Prep Kit (Illumina), from 500ng of purified RNA by applying the procedure optimized for degraded RNAs and omitting the fragmentation step, as recommended by the manufacturer's protocol. Libraries were quantified by the Agilent 2100 Bioanalyzer and Qubit fluorometer

Table 1  Clinicopathological features of the tissue samples analyzed in this study.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Tumor size (cm)</th>
<th>Genetic alteration&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Invasion</th>
<th>Capsule invasion</th>
<th>Multifocal disease</th>
<th>Tumor stage&lt;sup&gt;c,d&lt;/sup&gt;</th>
<th>Transcriptome analysis</th>
<th>Tissue analyzed&lt;sup&gt;e&lt;/sup&gt; (screening method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T15</td>
<td>5</td>
<td>RAS WT; PPARG WT</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T17</td>
<td>3.2</td>
<td>RAS WT; PPARG WT</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T19</td>
<td>3.4</td>
<td>N-RAS p.Q61L</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T21</td>
<td>2.3</td>
<td>RAS WT; PPARG WT</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T38</td>
<td>1.3</td>
<td>NA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T39</td>
<td>2.8</td>
<td>RAS WT; PPARG NA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T40</td>
<td>6</td>
<td>RAS WT&lt;sup&gt;b&lt;/sup&gt;; PPARG NA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T41</td>
<td>3</td>
<td>RAS WT; PPARG NA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T42</td>
<td>1.8</td>
<td>RAS WT; PPARG WT</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T43</td>
<td>2.2</td>
<td>RAS WT; PPARG NA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>FTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>2.3</td>
<td>PAX8-PPARG</td>
<td>Minimally invasive</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>Yes</td>
<td>Transcriptome</td>
</tr>
<tr>
<td>T4</td>
<td>4</td>
<td>RAS WT; PPARG WT</td>
<td>Widely invasive</td>
<td>No</td>
<td>Yes</td>
<td>pT2(m)N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T11</td>
<td>1.4</td>
<td>N-RAS p.Q61L</td>
<td>Minimally invasive</td>
<td>No</td>
<td>No</td>
<td>pT1N0M0</td>
<td>Yes</td>
<td>Methylome</td>
</tr>
<tr>
<td>T16</td>
<td>2.4</td>
<td>N-RAS p.Q61K</td>
<td>Minimally invasive</td>
<td>No</td>
<td>No</td>
<td>pT2N0M0</td>
<td>Yes</td>
<td>Methylome and transcriptome</td>
</tr>
<tr>
<td>T24</td>
<td>3.2</td>
<td>RAS WT; PPARG WT</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>No</td>
<td>pT3N0M0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>T25</td>
<td>10</td>
<td>RAS WT; PPARG WT</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>No</td>
<td>pT3N1bM0</td>
<td>No</td>
<td>Methylome</td>
</tr>
<tr>
<td>T30</td>
<td>1.8</td>
<td>K-RAS p.G13S</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>Yes</td>
<td>pT3(m)N0M0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>T31</td>
<td>6</td>
<td>N-RAS p.Q61L</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>Yes</td>
<td>pT3(m)N0M0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>T35</td>
<td>8</td>
<td>RAS NA; PPARG WT</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>No</td>
<td>pT3N0M0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T36</td>
<td>4</td>
<td>RAS WT; PPARG WT</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>Yes</td>
<td>pT3(m)N0M0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>T37</td>
<td>7</td>
<td>N-RAS p.Q61R</td>
<td>Widely invasive</td>
<td>Yes</td>
<td>Yes</td>
<td>pT3(m)N0M0</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

<sup>a</sup>NA, not assessed; Mutational status of RAS genes was searched for the following codons: K-RAS G12, K-RAS G13, H-RAS Q61 and N-RAS Q61. <sup>b</sup> K-RAS G12 and K-RAS G13: not assessed. <sup>c</sup>(m), multifocal. <sup>d</sup>TNM Classification of Malignant Tumours, 7th Edition (Sobin et al. 2010).

(Thermo Fisher) and pooled to a final concentration of 2nM, in order to make each index-tagged sample present in equimolar amounts. The pooled samples were subjected to cluster generation and sequencing using the Illumina HiSeq 2500 System (Illumina) in a 2 x 100 paired-end format at a final concentration of 8pM. Raw reads underwent quality control analysis using FastQC (http:// www.bioinformatics.babraham.ac.uk/projects/fastqc/), mapped to the human genome (hg19) by TopHat (version 2.1.1) (Kim et al. 2013) and counted (Ensembl database) using HTSeq package (version 0.6.1) (Anders et al. 2015). Gene expression was measured as counts-per-million (CPM). Genes with a CPM value ≤0.5 in NT samples were filtered-out. Finally, differential expression was assessed by R/Bioconductor package DESeq2 (version 1.8.2) (Love et al. 2014). Genes showing an absolute logFC value ≤−1 or ≥1 and an adjusted (FDR) P value ≤0.05 were defined as differentially expressed. For those genes (n. 5) that were associated, at the same time, with both hyper- and hypomethylated CpGs, the gene methylation status was assigned either based only on the status of the probes mapping in the regulatory regions (TSS200, TSS1500 and/or 5′UTR) or, when no CpGs mapped in regulatory regions, based on the status of the majority of the CpGs.

**Functional enrichment analysis**

Functional annotation of differentially methylated genes and differentially expressed genes was performed through the Molecular Signatures Database (MSigDB; https:// software.broadinstitute.org/gsea/msigdb/) using Chemical and Genetic Perturbations (CGP), GO biological processes (BP), GO molecular functions (MF) and oncogenic signatures (C6) gene sets. Lists of Polycomb target genes (PrGs) and of genes marked by bivalent chromatin were available in Supplemental materials of the articles by Ben-Porath et al. (2008) and Ku et al. (2008), respectively. Multi-cluster functional enrichment analysis was performed by ToppCluster tool (Kaimal et al. 2010, https://toppcluster. cchmc.org/); in particular, connectivity networks of common functional features were explored for the following gene clusters: downregulated, upregulated, hypomethylated, hypermethylated/bivalent chromatin marked and hypermethylated/not bivalent chromatin marked.

**Statistical analysis**

All analyses were performed using R statistical package (version 3.2.5) (www.cran.r-project.org). Statistical tests were performed using an α significance level of 0.05.

**Results**

**DNA methylation profiles of follicular thyroid neoplasms**

DNA methylation profiles of 11 FTCs, 10 FTAs and 7 control NT samples were obtained by the Illumina Infinium HumanMethylation 450K BeadChip. Clinicopathological characteristics of the tissue samples, as well as the presence of RAS or PAX8/PPARG lesions, are reported in Table 1. For each CpG site, methylation levels were expressed as β values and methylation differences between groups were statistically assessed by linear model (limma software package). CpGs were considered differentially methylated (hereafter referred to as DMCpGs) in tumor samples with respect to NTs when differential β values (Δβ) were ≥0.2 or ≤−0.2 (hypermethylated, respectively) and with P value ≤0.05 after FDR correction.

FTCs exhibited a profoundly altered DNA methylation profile with respect to NT controls, mainly featuring hypermethylated DMCpGs. In particular, 3564 DMCpGs (2979 hypermethylated and 585 hypomethylated) were identified in FTCs. Instead, with the above described cutoff values, virtually no DMCpG was identified in FTAs. A list of the DMCpGs is provided in supplementary data (Supplementary Table 1, see section on supplementary data given at the end of this article). Unsupervised hierarchical clustering of samples using FTC-associated DMCpGs identified two main groups, one including 9 out of 11 FTCs and 1 FTA and the other including the 2 remaining FTCs, 9 out of 10 FTAs and all NT samples (Fig. 1).

β values relative to the 3564 FTC-associated DMCpGs identified in this study were compared with those available in GEO (Gene Expression Omnibus) from previously published DNA methylation analyses of normal thyroid tissues, FTAs and FTCs (Mancikova et al. 2014, Timp et al. 2014, Bisarro Dos Reis et al. 2017). Though an overall good correlation was found (Pearson’s correlation index, r, ranking from 0.86 to 0.92 for NT, from 0.86 to 0.90 for FTAs and from 0.72 to 0.85 for FTCs), some β values featured a less good correlation in FTCs. This may, at least in part, depend on the fact that CpGs that scored differentially methylated (FTC vs NT) in the present study were selected for comparison (Supplementary Fig. 1). In addition, we selected the top 20 (those with larger differential values from FTC and NT) FTC-associated DMCpGs and individually checked their methylation status in the two previous studies that used the same microarray (Timp et al. 2014, Bisarro Dos Reis et al. 2017). These comparisons show a good concordance, being the
differences always in the same direction and, in most of the cases, within the statistical threshold (FDR ≤ 0.05) (Supplementary Fig. 2).

Mancikova et al. have noted that, when considered separately, RAS-positive and RAS-negative FTC samples featured a different number of DMCpGs (Mancikova et al. 2014). The clustering reported in Fig. 1 did not show separation of the FTC samples according to their mutational status. It is worth mentioning, however, that it was generated using the 3564 DMCpGs identified in the FTC group as a whole, for example, regardless of the presence of specific oncogenic lesions. When RAS-positive and RAS-negative FTCs samples were separately analyzed, a larger number of DMCpGs was identified in RAS-positive ones, though, probably because of the small sample sets, these additional RAS-specific DMCpGs did not reach statistical significance in the pairwise comparison between the two FTC groups (data not shown).

To explore differences between FTCs and FTAs, a principal component analysis (PCA) was performed using β values relative to the 3564 FTC-associated DMCpGs. The PCA biplot showed that FTCs clustered distantly from NTs, while most FTAs grouped in an intermediate position between FTCs and NTs (Fig. 2A). Similar distribution was obtained when all the explored 294,252 CpG probes were considered (data not shown). This finding suggested that FTAs had a DNA methylation profile intermediate between FTCs and NTs. To better visualize the contribution of each CpG to the position of FTA samples in the PCA biplot, we plotted the average Δβ values obtained in FTAs and FTCs for each of the 3564 DMCpGs (Fig. 2B). At limma analysis, mean Δβ values of most of the probes (n. 3342) in FTAs hold an intermediate position (e.g. larger than 0 but smaller than that found in FTC) between those of FTCs and NTs. Similarly, Δβ values of each individual FTA, in most of the cases, occupied an intermediate position between
average FTCs and NTs values (Supplementary Fig. 3). The reduced $\Delta\beta$ value combined with the small sample size probably accounted for the lack of significance of the FTA-NT comparison (Supplementary Table 1). Similarly, Yim et al. (2019) also noted that benign thyroid nodules frequently feature an epigenetic pattern intermediate between normal thyroids and malignant nodules.

**Features of the differentially methylated CpGs in follicular thyroid carcinoma**

We classified the 3564 FTC-associated DMCPGs based on the occurrence in CpG islands (CGI), ‘shores’, ‘shelves’, or ‘open sea’ as well as the occurrence in enhancers, gene regulatory regions (e.g. within 1500bp upstream the transcriptional start site-TSS or within the 5’-UTR), gene bodies or intergenic regions (Supplementary Table 1). While hypermethylated DMCPGs were statistically enriched (40%) in CGIs and poorly represented (27%) in open sea (Pearson’s Chi-squared test; $P$ value $\leq 2.2 \times 10^{-16}$), the hypomethylated ones were mostly found (59%) in open sea and poorly represented (4%) in CGIs (Pearson’s Chi-squared test; $P$ value $\leq 2.2 \times 10^{-16}$) (Fig. 3). Moreover, a significant enrichment was found for both hypermethylated and hypomethylated enhancer-associated DMCPGs with respect to their representation in the array (Pearson’s Chi-squared test; $P$ value $\leq 2.2 \times 10^{-16}$ and $P$ value $\leq 5 \times 10^{-8}$, respectively). Instead, no significant enrichment for gene-associated or intergenic DMCPGs was noted (Fig. 3).

FTC-associated differentially methylated genes (DMG) (n. 1786) were defined as those bearing DMCPGs either within their gene body or regulatory region (Supplementary Table 1). DMGs functional annotation was performed by Gene Set Enrichment Analysis (GSEA). FTC-associated hypermethylated DMGs were enriched for genes possessing trimethylated H3K27 (histone H3 trimethylated lysine 27, H3K27me3) repressive histone mark as well as for genes that are regulated by the Polycomb group (PcG) of transcriptional repressors (Fig. 4A and Supplementary Table 2) (Schlesinger et al. 2007, Widschwendter et al. 2007, Ben-Porath et al. 2008, Meissner et al. 2008, Kim et al. 2010, Easwaran et al. 2012). Chromatin featuring simultaneously repressive H3K27me3 and activatory H3K4me2/me3 transcription marks is named ‘bivalent’. Noteworthy, about 25% of FTC-associated hypermethylated DMGs are marked by bivalent chromatin modifications in embryonic cells (Ku et al. 2008) (Fig. 4B). Both PcG targets and bivalent chromatin-marked genes were further enriched when only genes differentially methylated in CGIs mapping in their regulatory regions were considered (Fig. 4).

**Integration of epigenome and transcriptome in follicular thyroid carcinoma**

Transcriptome profiling of 8 FTCs (out of the 11 FTCs whose methylome was analyzed) and 6 NTs (including 5 whose methylome was also analyzed) was performed by RNAseq. Out of the 16,702 genes whose expression was measurable, 776 were differentially expressed in FTC compared to NT (361 downregulated and 415 upregulated) with a Log fold-change (LogFC) $\geq 1$ and adjusted (FDR) $P$ value $\leq 0.05$. A list of these differentially expressed genes...
is reported in Supplementary Table 3. At the Gene Set Enrichment Analysis, FTC-deregulated genes showed large overlaps with those found differentially expressed in previous studies both in thyroid cancer subtypes (PTC, ATC and PDTC) (Delys et al. 2007, Rodrigues et al. 2007) and in other cancer types (Sweet-Cordero et al. 2005, Vecchi et al. 2007, Liao et al. 2008, Smid et al. 2008) (Supplementary Table 4).

Thus, FTC methylome and transcriptome were integrated. Out of a total of 1786 FTC-associated DMGs, 1172 were measurable at the RNAseq and 93 of them (7.9%) featured a differential expression in FTC with respect to NT, including 34 downregulated and 59 upregulated. In particular, methylation and expression were negatively correlated in the case of 52 DMGs (30 hypermethylated/downregulated and 22 hypomethylated/upregulated) (Supplementary Table 5). Finally, we divided FTC-associated differentially methylated and differentially expressed genes in five groups (downregulated, upregulated, hypomethylated, hypermethylated/bivalent chromatin-marked and hypermethylated/not bivalent chromatin-marked) and performed multi-cluster functional enrichment analysis by the ToppCluster tool (Kaimal et al. 2010). Genes characterized by hypermethylation (particularly those marked by bivalent histone modifications) and genes with downregulated expression belong to common functional categories (Fig. 5). These data suggest that in FTC, notwithstanding the relatively small fraction of genes simultaneously affected by DNA methylation and altered expression, gene methylation...
and gene repression may independently concur to deregulate specific biological functions.

**Discussion**

Here, we have studied DNA methylome of FTC compared to FTA and normal thyroid. Our study has some limitations, including the small number of samples and its retrospective design. In particular, the small sample size precluded the possibility of driving correlations between altered DNA methylation and oncogenic driver lesions as well as FTC histological subtypes. Notwithstanding these limitations, our results confirm that aberrant DNA methylation is associated with thyroid tumorigenesis. Moreover, according to previous studies, our data also indicate that the methylation pattern of FTC, in particular, is characterized by predominant DNA hypermethylation (Mancikova et al. 2014, Bisarro Dos Reis et al. 2017). Similar to these previous reports, our results also show that, on average, FTAs feature more modest altered DNA methylation with respect to FTC. By focusing on FTC-associated DMCPGs, we found that, in FTAs, their methylation levels distribute in an intermediate position between normal samples and FTCs. As recently demonstrated by global expression profiling (Dom et al. 2018), these findings are consistent with the possibility that FTA may represent a biological intermediate between normal tissue and FTC (Arora et al. 2008, Sobrinho-Simões et al. 2011, Giordano 2018).

By integrating methylation and expression data, we show that the overall proportion of differentially methylated genes found to be transcriptionally deregulated was 7.9%. These findings are consistent with previously reported data (Mancikova et al. 2014) and suggest that although enrichment was found for CGIs (in the case of hypermethylated DMCPGs) and for enhancers (in the case of both hypermethylated and hypomethylated DMCPGs), methylation of these sites may not cause a detectable change of the expression level.

To further explore the functional significance of DNA methylation in FTC, we performed an enrichment analysis.
of differentially methylated genes. This analysis showed that, similar to other cancer types, a large (25%) fraction of the hypermethylated genes in FTC are marked by bivalent H3 (K4/K27) histone modifications in embryonic cells (Schlesinger et al. 2007, Widschwendter et al. 2007, Ben-Porath et al. 2008, Meissner et al. 2008, Kim et al. 2010, Easwaran et al. 2012). In cancer, hypermethylation of ‘bivalent’ genes may permanently preclude their activation. H3K27 methylation is orchestrated by the Polycomb group (PcG) of transcriptional repressors. Thus, these observations suggest that PcG- and bivalent chromatin-based gene regulation may play a role in FTC pathogenesis. On the other hand, they also provide a reasonable explanation for the lack of a more robust overlap between hypermethylation and downregulated expression in FTC; indeed, bivalent genes are expected to be already repressed in normal cells and therefore their methylation in cancer cells would not result in a detectable downregulated expression. Finally, network analysis showed that though the individual genes affected by the altered methylation and expression may be different, the two phenomena may converge to a common group of molecular functions.

Methods exploiting gene expression profiling and/or detection of driver mutations have been developed to refine the molecular differentiation between benign and malignant thyroid tumors (Nikiforov et al. 2015, 2018, Kloos 2017). The results of the present study suggest that DNA methylation profiling of specific CpG sites may be exploited to complement the other methods in the differential diagnosis of FTAs from FTCs. Moreover, FTC can present with different variants, the most common being classical (CVPTC) and follicular (FVPTC) ones. Architectural and genetic similarities between FTC and FVPTC have suggested that these two tumors may bear common pathogenetic mechanisms (Giordano 2016, 2018). Accordingly, Mancikova and coworkers have reported that FTC histomelone quite closely resembles that of FVPTC (Mancikova et al. 2014). Thus, DNA methylation pattern reported in the present and previous studies may also be exploited for the differential diagnosis of FTAs from FVPTCs.

Supplementary data
This is linked to the online version of the paper at https://doi.org/10.1530/ERC-18-0308.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This study was supported by the Epigenomics Flagship Project SP6.6 – EPiGEN, C.N.R. Italy.

Author contribution statement
O Affinito and P Salerno performed data extraction and analysis. A D’Alessio performed the external dataset analysis. E Florio, M Cuomo and F Carломagno were responsible for DNA and RNA analyses. A Piroietti and F Basolo collected samples and performed the surgical pathology analysis. R Giannini performed DNA and RNA extractions and mutation analysis. L Chiarotti coordinated the methylome and the transcriptome screening. S Cocozza coordinated all the bioinformatic analysis and wrote the manuscript. M Santoro participated in data analysis and manuscript writing.

References


Flavahan WA, Gaskell E & Bernstein BE 2017 Epigenetic plasticity and embryonic stem and cancer cell transcription programs. Molecular and Cellular Biology 37 6348. (https://doi.org/10.1126/science.aal2380)


Xing M 2010 Genetic alterations in the phosphatidylinositol-3 kinase/Akt pathway in thyroid cancer. Thyroid 20 697–706. (https://doi.org/10.1089/thy.2010.1646)


Received in final form 6 February 2019
Accepted 11 February 2019
Accepted Preprint published online 11 February 2019