“DNA methylation in metastatic thyroid cancer”

Helena Rodríguez Lloveras

Dra. Mireia Jordà Ramos
Grup de Tumors Endocrins
Programa de Medicina Predictiva i Personalitzada del Càncer, Institut de Recerca Germans Trias i Pujol (PMPPC-IGTP)

Dra. Teresa Mampel Astals
Departament de Biologia Molecular, Universitat de Barcelona (UB)

June 29th, 2017

Grau de Ciències Biomèdiques

Summary

Thyroid cancer is the most common endocrine malignancy. It has a high survival rate, but it decreases when the tumour is metastatic. Although there are good diagnostic tools for thyroid cancer, the tools to predict which tumours will become metastatic and/or have a higher risk for the patient are limited. It is known that epigenetics play an important role in thyroid cancer, but little is known about their implication in metastatic thyroid cancer.

Here we want to develop a prognosis tool able to predict if a tumour will develop metastasis or not based on the changes on DNA methylation.

To do so we performed a genome-wide DNA methylation analysis and some specific comparisons and analysed the differentially methylated sites in order to find the best signature capable of predicting the metastatic capacity or not of a given tumour. Furthermore, we tried to create a signature that could be applied clinically.

We proved that indeed there are specific differentially methylated sites in thyroid cancer depending on the prognosis of the tumour. We obtained two signatures that were able to predict if a tumour was metastatic or not, one consisting of 38 probes and associated with papillary carcinomas, and one of 156 probes independent of the histology. We used this last signature to analyse clinical data and to define a smaller signature so it can be used in the clinical routine.

To sum up, DNA methylation in thyroid cancer changes depending on the characteristics of the tumours, including whether it is metastatic or not, and some of these changes could be used as a biomarker.

Index

Keywords

T3: Triiodothyronine

T4: Thyroxine

TC: Thyroid Cancer

SEER: Surveillance, Epidemiology and End Results program

FNAB: Fine-Needle Aspiration Biopsy

RAI: Radioactive Iodine

WDTC: Well Differentiated Thyroid Carcinomas

PTC: Papillary Thyroid Carcinoma

FTC: Follicular Thyroid Carcinoma

PDTC: Poorly Differentiated Thyroid Carcinoma

ATC: Anaplastic Thyroid Carcinoma

fvPTC: follicular variant of Papillary Thyroid Carcinoma

DNMTs: DNA methyltransferases

CGI: CpG Islands

NT: Normal Tissue

TCGA: The Cancer Genome Atlas

THCA: Thyroid Carcinoma

STAD: Stomach Adenocarcinoma

LUAD: Lung Adenocarcinoma

COAD: Colon Adenocarcinoma

KIRC: Kidney Renal Clear Cell Carcinoma

SARC: Sarcoma

SKCM: Skin Cutaneous Melanoma

TSS: Transcription Start Site

1. Introduction

1.1. Thyroid gland

The thyroid is an endocrine gland located below the larynx, consisting of two lobes connected by an isthmus. The lobes measure about 50-60mm each, with a conjoined weight of about 30g, tough it tends to be bigger in women.

The lobes are composed of follicles, which consist of a simple layer of epithelium cells that extend to the lumen surrounded by a basement membrane. 95% of these epithelial cells are follicular cells, while the other 5% are the parafollicular or C cells, which produce calcitonin and help regulate calcium homeostasis.

However, the main function of the thyroid gland is the production of the thyroid hormones by the follicular cells. These cells can uptake iodine through the NIS/SLC5A5 channel in order to synthesize the thyroid hormones triiodothyronine (T3) and its precursor tetraiodothyronine or thyroxine (T4). These hormones have different effects throughout the body, regulating the endocrine system and controlling the metabolism, including the increase of the basal metabolic rate, maintenance of the normal body temperature or the increase of protein synthesis among others [1].

1.2. Thyroid cancer

Cancer is a group of diseases characterised by a malignant transformation and abnormal growth of cells which divide without control and can invade nearby tissues. It is a progressive process characterized by the accumulation of both genetic and epigenetic anomalies [2].

Next I will introduce you some of the most relevant clinical aspects of TC.

1.2.1. Epidemiology

Thyroid cancer (TC) is the most common endocrine malignancy, with an increasing incidence all over the world (Figure 1). It is thought that more than 60% of the population has thyroid nodules, but only a 8-16% of these will become malignant. Thus, TC only represents a 2.3% of all cancers worldwide [3].

According to the Surveillance, Epidemiology and End Results program (SEER), it affects more women than men (about 77% of all patients are female) and has a median diagnosis age of 51 years old. Fortunately, TC has a high survival rate: more than 98% after five years (99.9% for localised tumours and 98.0% for localised metastasis affecting the nodules). The remaining 4% develop distant metastasis, and have a survival rate after five years of 56.4% (Figure 2) [4].

Despite this, between 10-30% of the patients present persistence or recurrence of the disease, considered of bad prognosis. Part of this is caused by one of the main clinical problems with TC, the inappropriate diagnosis and treatment of the most aggressive tumours. The other clinical problem is the overdiagnosis of clinically irrelevant tumours. Recent studies have also shown a recent increase in death caused by TC, even if it is very low (Figure 2) [3].

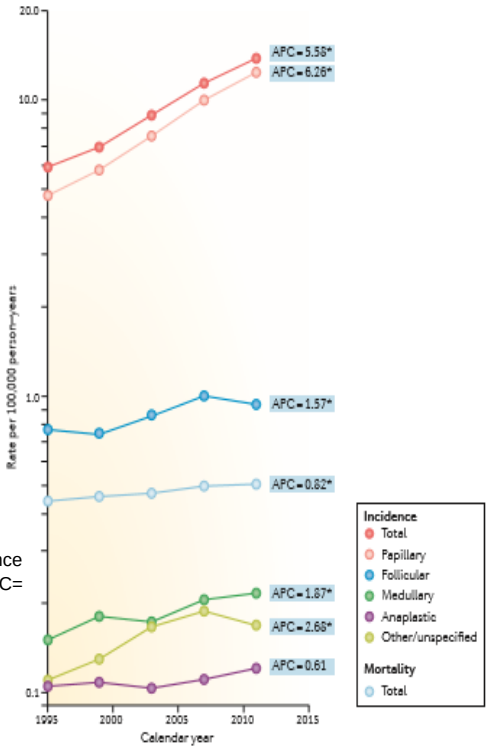


Figure 1. Increase of the total incidence of thyroid cancer, of its incidence depending on the histological types and its mortality from 1993-2012. APC= Annual Percentage Change. *pval<0.05. [3]

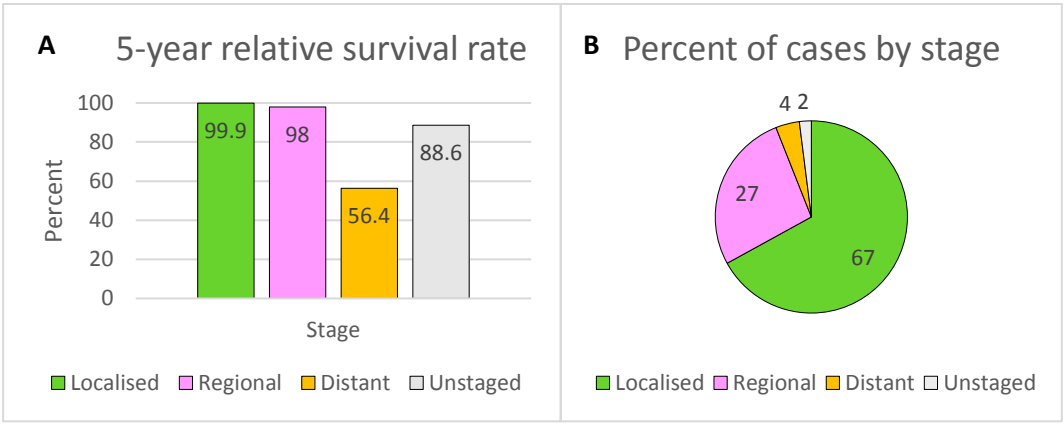


Figure 2. A: 5-year relative survival rate of thyroid cancer for localised (non metastatic), regional (nodules affected), distant (distant metastasis) or unstaged cases. **B:** Percent of cases of thyroid cancer depending on the stage. Modified from SEER [4]

1.2.2. Risk factors

A risk factor is any attribute, characteristic or exposure of an individual that increases the likelihood of developing a disease or injury [5]. In TC sex and age are considered risk factors, as it affects more women than men (like all thyroid diseases) and its incidence is higher at certain age (Figure 3) [4].

Another risk factor is the race, since despite the fact that the incidence increases in all races its rate and the histological type varies depending on the race. Other risk factors are the exposure to ionizing radiation, iodine deficiency, obesity or environmental pollutants, known to be carcinogenic agents [3].

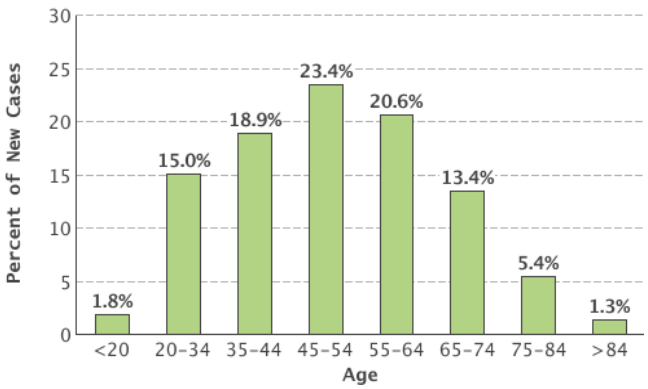


Figure 3. Percent of new cases depending on the age pf the patients from 2010 to 2014 [4].

There are also some genetic risk factors, such as *RET* mutations or the Gardner Syndrome [6].

1.2.3. Clinical management of thyroid cancer

After the detection of nodules by palpation or imaging techniques, TC is diagnosed with a Fine-Needle Aspiration Biopsy (FNAB). If the result is positive, the initial treatment is usually a total thyroidectomy, which may be accompanied by ganglia removal if the nodules are affected, and an adjuvant treatment with Radioactive Iodine (RAI). However, this total thyroidectomy may not be necessary if the tumour is not aggressive, especially since it has several side effects (together with RAI) like the need of hormone treatments for the rest of the patients' lives [7].

Another problem is that about 5% of the patients do not respond to RAI, and have a worse prognosis and few treatment options [8].

For these reasons, there is an urgent need to find new diagnosis and prognosis markers.

1.2.4. Histological types

About 98% of all tumours arise from follicular cells, while only a 2%, named medullary thyroid carcinomas, come from C cells. We will focus on those tumours that arise from follicular cells (Figure 4) [9].

The most common tumours are the Well Differentiated Thyroid Carcinomas (WDTC), still able to uptake iodine. This group include the Papillary Thyroid Carcinomas (PTC) and the Follicular Thyroid Carcinomas (FTC), representing 80% and 15% of all TC cases respectively (Figure 4). However, this incidence changes in some regions, for example in regions with poor iodine intake, where the FTC incidence may increase up to a 30-40% [10].

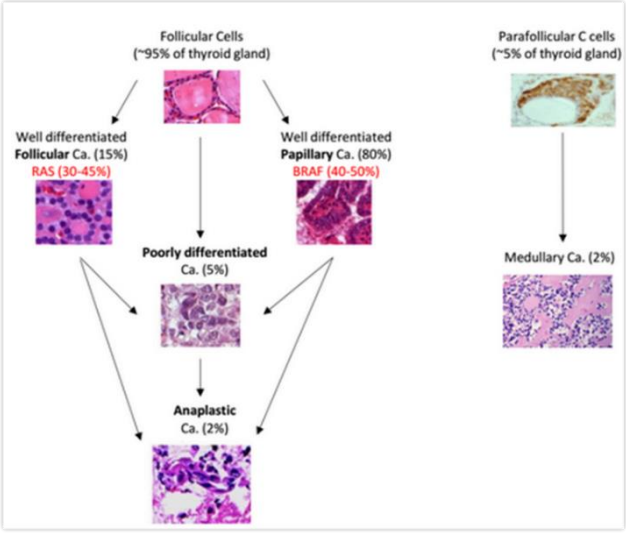


Figure 4. Different thyroid carcinomas and its frequency, with the most common mutation for PTC and FTC. Ca= carcinoma. Modified from Bose and Walts 2012[9].

Following these WDTC are the Poorly Differentiated Thyroid Carcinomas (PDTC), which represent a 5% of all TC cases and are a medium stage between WDTC and Anaplastic Thyroid Carcinomas (ATC). ATC is the most aggressive type of TC, with a medium life expectancy lower than 6 months and responsible of the third of all TC deaths, even if it only represents a 2% of all TC cases [11].

1.2.5. Metastatic thyroid cancer

Around 10% of the patients with thyroid cancer develop distant metastasis, which are the main cause of thyroid cancer-related deaths. PTCs tend to metastasise in local ganglia, while FTC often metastasise distant organs. Of these, the most frequent locations are lung, bone or both locations [12].

1.2.6. Genetic alterations in thyroid cancer

TC is one of the cancers with the lowest number of genetic alterations [13]. The two main driver mutations in TC affect either the MAPK or the PI3K-AKT pathways. The first is driven by *BRAF* mutations, being the most common a point mutation where a valine changes into a glutamic acid ($BRAF^{V600E}$), and is specific for PTC. Alteration in the later pathways are driven by *RAS* mutations, mutually exclusive with *BRAF* mutations, and specific for FTC and the follicular variant of PTC (fvPTC, representing 9-12% of all PTC).

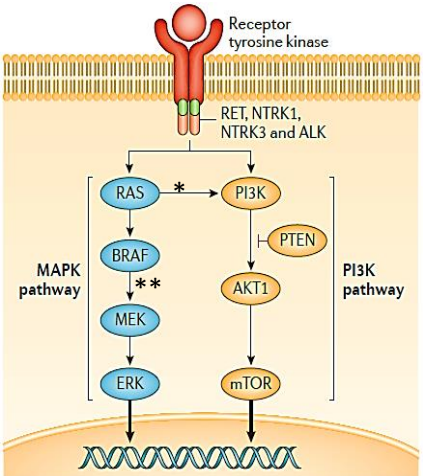


Figure 5. MAPK and PI3K-AKT pathways. * Enhanced in RAS mutated tumours. **Enhanced in BRAF mutated tumours.

BRAF mutations lead to an over-activation of the MAPK pathway, while *RAS* mutations lead to an over-activation of the PI3K-AKT pathway (Figure 5) [7].

Other common mutations are *RET/PTC* and *RTK* rearrangements and the *PAX8/PPARγ* rearrangements [7].

However, TC is not only a genetic disease but also an epigenetic disease.

1.3. Epigenetics: DNA methylation

Epigenetics is the study of heritable changes in gene expression that are not explained by changes in the DNA sequence [22]. The epigenome is the combination of all epigenetic modifications within a cell which maintain chromatin architecture and a specific gene expression. These modifications include mainly histone modifications, non-coding RNAs and DNA methylation [23].

DNA methylation is the most studied epigenetic modification. It consists on the covalent addition of a methyl group ($-CH_3$) to the fifth carbon of the pyrimidine ring of a cytosine within a CG dinucleotide (CpG), transforming it into a 5-methylcytosine (5mC), symmetrically on both strands of the DNA. This process is performed by DNA methyltransferases (DNMTs). DNMT1 is involved in the maintenance of DNA methylation, while DNMT3A and DNMT3B are involved in both DNA methylation maintenance and *de novo* methylation [24].

DNA demethylation can also occur in different ways: passively by loss of DNA methylation during replication or actively through the removal of the 5mC by DNA repair mechanisms [25]. CpGs are also hotspots for mutations where the 5mC mutates into a thymine. Due to this, the frequency of CpGs in the genome is lower than expected (1% instead of the 4.41% expected) [23].

Between 70-80% of these CpGs are methylated, although there are some CpG-rich regions that remain unmethylated named CpG islands (GCIs) [26]. CGIs are regions of the genome of 200bp to 3000bp where the percentage of CpGs is superior to 50% and the ratio between observed CpGs and expected CpGs is superior to 0.6 [27]. The 2kb upstream and downstream regions from a CGI are known as CGI shores; the regions between 2kb and 4kb from the CGI are known as CGI shelves; and the regions which are more than 4kb away from the CGI are known as open sea [28]. Around 60-70% of all gene promoters are associated with CGIs.

The methylation of regulatory elements, such as promoters and enhancers, is considered a repression mark. DNA methylation can also regulate gene expression when it is found in the body region of the genes (exons and introns), promoting transcription [24].

This regulation of gene expression through DNA methylation affects many processes such as embryonic development, tissue-specific expression, imprinting, chromosome X inactivation in women, splicing and silencing of repetitive sequences. Aside from regulating gene expression, DNA methylation also helps to maintain the genome integrity [24].

DNA methylation may change during certain pathogenic processes, like cancer. When in the pathogenic process there is a loss of methylation compared to a healthy cell we call it hypomethylation, while when there is a gain of methylation we call it hypermethylation.

For this reason, and given the great stability of both DNA and DNA methylation, and its easy detection by numerous techniques, DNA methylation has a great potential as biomarker [29], and indeed some DNA methylation alterations are already used in some cancers [30].

1.3.1 DNA methylation in thyroid cancer

DNA methylation differs between a normal and a tumour cell. In cancer there is a global hypomethylation of the genome (Figure 6A), which may appear even before the tumour transformation [31]. This global hypomethylation can lead to loose of imprinting, genomic instability or alter gene expression (Figure 6B), increasing for example the expression of oncogenes. This global hypomethylation can be measured, for example, through the hypomethylation of Alu elements (i.e., repetitive sequences) [32].

However, in cancer there also are local hypermethylations which can affect some promoters (Figure 6C), repressing the transcription of tumour-suppressor genes or some tissue specific genes, like *NIS* on thyroid cancer [16].

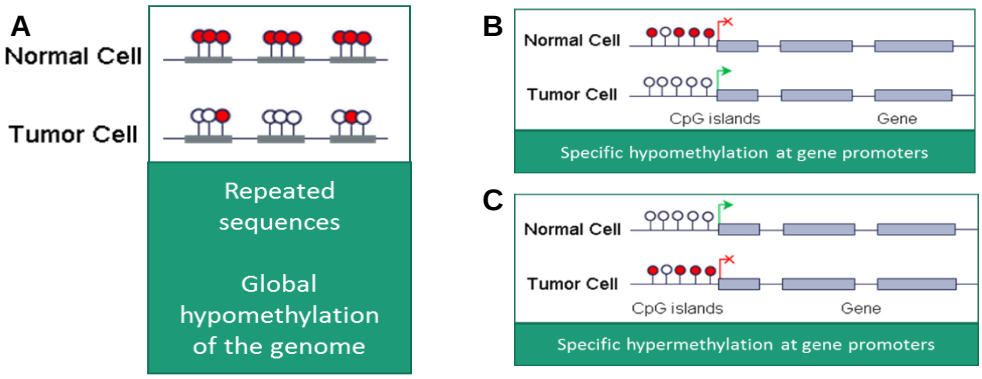


Figure 6. Changes in DNA methylation comparing Normal and Tumour cells. **A**: global hypomethylation of the genome. **B**: Specific hypomethylation of the genome at gene promoters which increases the expression of some genes. **C**: Specific hypermethylation of some gene promoters which reduces the expression of some genes. Modified from Atlas of Genetics and Cytogenetics in Oncology and Haematology [33].

TC is not an exception and there are many works identifying several differential methylated genes, like the hypermethylation of *PTEN* or *NIS* [16]. Moreover, in recent years some few genome-wide DNA methylation studies, including one from my research group, reported that DNA methylation profiles are associated with the histological type and the driver mutation [17], [18], [19], [20], [21].

2. Objectives

The purpose of this project is to study the changes on DNA methylation in the progression of thyroid cancer.

Our specific objectives are:

1. To characterize DNA methylation profiles in the progression of thyroid cancer.
2. To identify and study the DNA methylation alterations associated with the metastatic process.
3. To generate a new molecular tool that predicts if a tumour will develop metastasis.

3. Materials and methods

3.1. Materials

To analyse DNA methylation in TC we had three sets of samples:

- Discovery series: a set of 96 human samples consisting of 15 normal tissues (NT) paired with the tumours, 41 PTC, 23 FTC and 17 metastases (Table 1).
- Cell lines series: three human TC cell lines (TPC-1, HTH83 and BCPAP) [34].
- PDTC/ATC series: 17 human samples (3 NT, 8 ATC and 6 PDTC).

All tumours are macrocacinomas (>1cm).

Moreover, we used 7 different cancer types datasets from The Cancer Genome Atlas (TCGA): thyroid carcinoma (THCA) [21], stomach adenocarcinoma (STAD) [35], lung adenocarcinoma (LUAD) [36], colon adenocarcinoma (COAD) [37], kidney renal clear cell carcinoma (KIRC) [38], sarcoma (SARC, unpublished) and skin cutaneous melanoma (SKCM, unpublished).

Table 1. Clinical data of the discovery series

Parameter	PTC (n=41)	FTC (n=23)
Gender, n (%)		
Female	30 (73.17%)	20 (86.96%)
Male	10 (24.39%)	3 (13.04%)
Unknown	1 (2.44%)	0 (0%)
Mean age at diagnosis (max-min)	50 (22-82)	58 (21-78)
Prognosis		
Metastatic	21 (51.22%)	13(56.5%)
Low Risk	20 (48.78%)	10 (43.5%)
TNM Tumour stage, n (%)		
T stage		
T1	11 (26.83%)	3 (13.04%)
T2	15 (36.59%)	9 (39.13%)
T3	7 (17.07%)	8 (34.78%)
T4	7 (17.07%)	3 (13.04%)
Unknown	1 (2.44%)	0 (0%)
N stage		
N0	9 (21.95%)	10 (43.48%)
N1	23 (56.10%)	4 (17.39%)
Unknown	9 (21.95%)	9 (39.13%)
M stage		
M0	20 (48.78%)	10 (43.48%)
M1	21 (51.22%)	13 (56.52%)
Recurrence, n (%)		
Yes	5 (12.20%)	1 (4.35%)
No	35 (85.37%)	22 (95.65%)
Unknown	1 (2.44%)	0 (0%)
Persistence, n (%)		
Yes	11 (26.83%)	12 (52.17%)
No	22 (53.66%)	10 (43.48%)
Unknown	8 (19.51%)	1 (4.35%)

Metastasis location, n (%)		
Bone	5 (12.20%)	5 (21.74%)
Lung	6 (14.63%)	2 (8.70%)
Multi	6 (14.63%)	6 (26.09%)
None	20 (48.78%)	10 (43.48%)
Unknown	4 (9.76%)	0 (0%)
Mutation		
$BRAF^{V600E}$	12 (29.27%)	0 (0%)
BRAF, others	2 (4.88%)	0 (0%)
BRAFwt	2 (4.88%)	0 (0%)
RAS	5 (12.20%)	5 (21.74%)
WT2*	18 (43.90%)	16 (69.57%)
Unkown	2 (4.88%)	2 (8.70%)
Follow up		
Death, TC	7 (17.07%)	6 (26.09%)
Death, other causes**	3 (7.32%)	0 (0%)
Death, unknown causes	4 (9.76%)	2 (8.70%)
Active disease	3 (7.32%)	5 (21.74%)
Disease free	22 (53.66%)	10 (43.5%)
Unknown	4 (9.76%)	0 (0%)
Hypomethylated Alu elements* **		
Yes	3 (7.32%)	7 (30.43%)
No	25 (60.96%)	15 (65.2%)
Unknown	13 (31.71%)	1 (4.35%)

*Wild type for RAS and BRAF. **Other types of cancer or other diseases. ***We measured the hypomethylated Alu elements as reporters of the global hypomethylation of the genome.

3.2. Methods

3.2.1. Global methylation analysis

The methylation of the samples was assessed using the Infinium MethylationEPIC (or 850K) array (Illumina) [39]. This consists of a total of 867.926 probes, each specific for one CpG, covering all the genome. This array was used for the discovery and cell lines series. The PDTC/ATC series' methylation was assessed using the 450K array [40], an older version of the MethylationEPIC array. However, we will focus in the newest version.

The result of the array was analysed using RnBeads [41], to make a quality control and filter out those samples that could lead to confusing results (Figure 7). This analysis was performed for both arrays by the Genomics and Bioinformatic Unit of the PMPPC.

The final probes were classified depending on their CGI relation (i.e., distance from a CpG: island, shore, shelf or open sea) and the genomic compartment they belonged to (i.e., promoter, enhancer, body gene or intergenic).

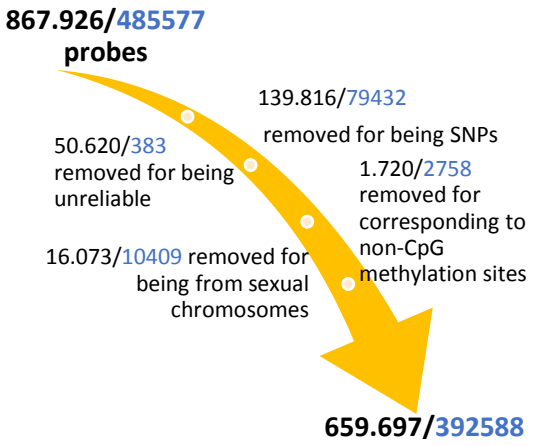


Figure 7. Filtering process of the probes from the 850K EpicMethylation array (black) and the 450K Methylation array (blue).

3.2.2. Data annotation and definitions

The genomic data (genomic position, flanking gene, relative position to the gene, CGI relation, etc.) for each probe was obtained from Illumina (MethylationEPIC BeadChip information) and the UCSC Genome Browser (https://genome.ucsc.edu/). Definitions used in this report: promoters were defined as the regions between -1500 and 500 base pairs from the Transcription Start Site (TSS); enhancers were defined according to ENCODE and FANTOM5; body gene included exons and introns; intergenic regions corresponded to those found outside a gene; CGI shores referred to 2kb from either side of a CGI; CGI shelves were defined as 2-4 kb from either side of the CGI; open sea referred to those regions outside a CGI, shore and shelf.

3.2.3. Hierarchical clustering analysis

For the cluster analyses we used a modified function heatmap.3 from the R package gplots, using the method Manhattan and the method Ward.D2 to create the clusters.

3.2.3.1. Unsupervised analysis

The unsupervised analyses were performed using the probes that showed the highest variability (based on the standard deviation) in the DNA methylation between all the samples. We considered those probes with a standard deviation superior to 0.2 (SD>0.2). We separated the probes on regulatory regions (enhancers and promoters), body gene and intergenic regions.

3.2.3.2. Differential DNA methylation analysis

The differential DNA methylation analyses were done with RnBeads [34] (by the Genomics and Bioinformatics Unit of the PMPPC), and we compared the samples between 2 groups (e.g., NT versus T). We considered those probes with an adjusted p-value < 0.05 and a medium difference in methylation > 0.2.Next, we performed hierarchical cluster analyses using the probes that were differentially methylated: a) in a specific histological type of TC (FTC or PTC) when compared to NT, b) in the comparison between Metastatic PTC and Low risk PTC, and c) in the comparison between Metastatic tumours and Low risk tumours.

We also performed several validation analysis using the cell lines and PDTC/ATC series.

3.2.4. Analysis of the clinical data

Differences in clinical data were evaluated by applying the Fisher's exact test, using Rstudio.

3.2.5. Feature selection

To reduce the number of probes necessary to predict if a tumour will become metastatic or not (i.e., if it is a low or a high-risk tumour) we used the SimpleLogistics method, from WEKA [42]. We used a 10-fold cross-validation method, and selected the 10 probes among the 34 that were statistically relevant with a higher effect size (i.e., with a higher influence in the results).

$$Effect\ size = \frac{mean1 - mean2}{SDpooled}$$

To manage big datasets during the whole project we used different tools from Galaxy (www.usegalaxy.org) [43].

4. Results

4.1. Identification of differentially methylated sites in well-differentiated thyroid cancer

Since all the previous genome-wide studies mostly used tumours with a good prognosis, we performed a new genome-wide DNA methylation study using tumours of bad prognosis. From the differential DNA methylation analysis, we identified numerous differentially methylated sites (Supplementary Figure 1). We focused on those

specifically associated with PTC or FTC in comparison with NT, both in general and depending on their prognosis (low risk non-metastatic, and metastatic) (Figure 8).

Figure 8. A: Hypermethylated and hypomethylated probes when comparing PTC and FTC versus normal tissue. **B**: Hypermethylated and hypomethylated probes when comparing Low risk non-metastatic (Low Risk) or Metastatic (Met) Papillary or Follicular (PTC or FTC) samples versus Normal tissue. **C**: Hypermethylated and hypomethylated probes when comparing Metastatic versus Low risk

We observed how there are more hypomethylations than hypermethylations in most cases. Also, there are more DNA methylation changes in FTC than in PTC, and in metastatic than in low risk tumours, especially for FTC.

We also analysed the CGI relation and genomic compartment of the differentially methylated sites. Shown in Figure 9A –B there are the more important comparisons with NT according to their methylation state, while the whole set can be found in Figure 9C.

In figure 10A we can observe how most hypomethylations take place away from CGIs, while most DNA methylation alterations in CGIs are hypermethylations. We can also see that the distributions of the hypomethylations in relation to the CGIs is more similar between the different comparisons than the hypermethylations. On the other hand, there are more hypermethylations in CGI and CGI shores in FTC than in PTC, and also in metastatic tumours, both PTC and FTC, that in low risk tumours.

When looking at the DNA methylation alterations in the different genomic compartments there are not as many changes in comparison to the distribution in the

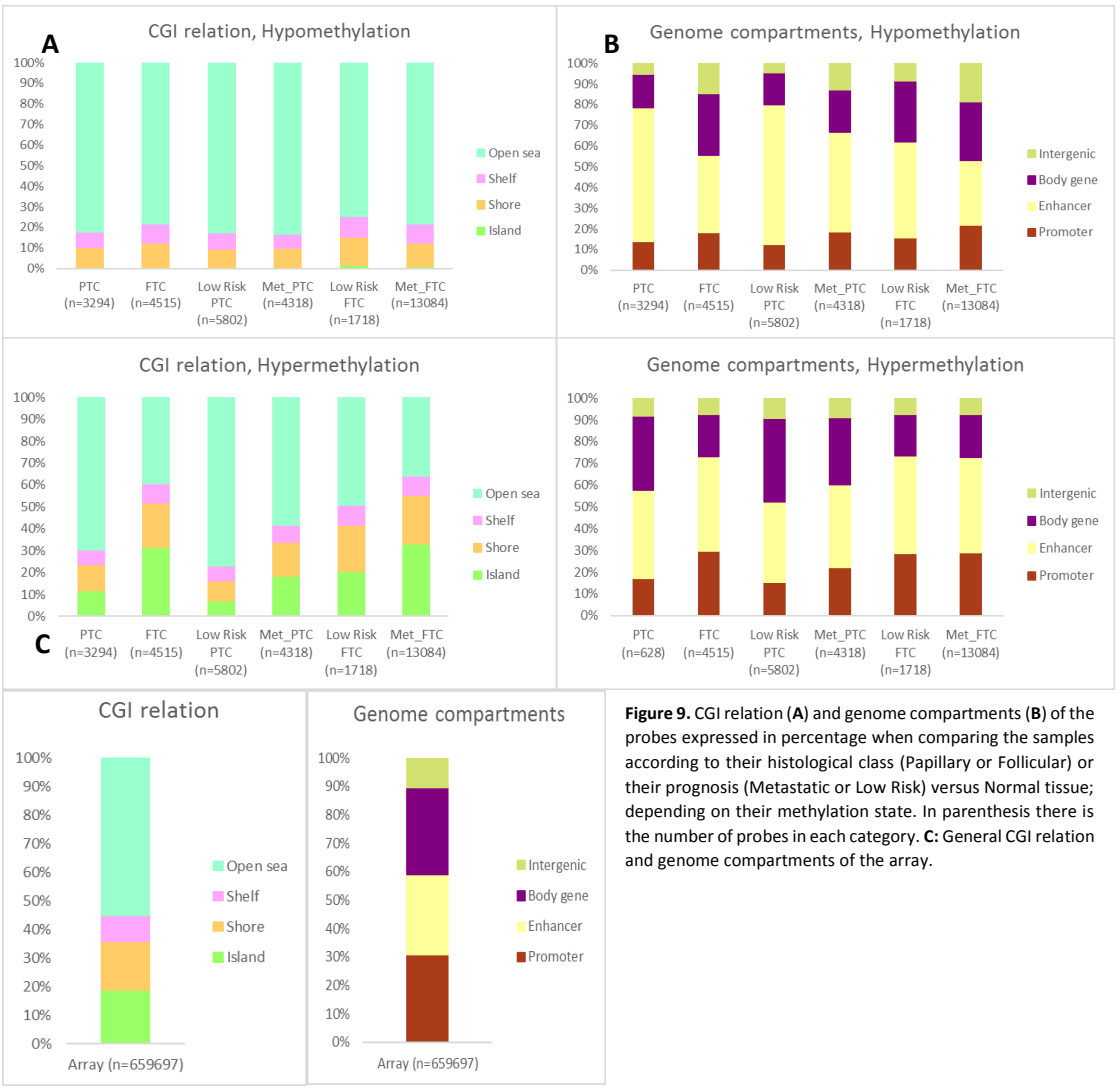


Figure 9. CGI relation (**A**) and genome compartments (**B**) of the probes expressed in percentage when comparing the samples according to their histological class (Papillary or Follicular) or their prognosis (Metastatic or Low Risk) versus Normal tissue; depending on their methylation state. In parenthesis there is the number of probes in each category. **C:** General CGI relation and genome compartments of the array.

array as occurred with the CGI relation. However, most hypomethylations are in regulatory regions, mainly in enhancers, while the hypermethylations have a distribution more similar to the array (Figure 9B).

4.2. DNA methylation profiles of promoters and enhancers reflect the histology and mutational state of *BRAF* and *RAS*

We classified the probes according to their genomic context (regulatory elements including promoters and enhancers, gene body and intergenic regions) and performed unsupervised analyses using the probes that showed the highest variability between all the samples. With the regulatory regions (Figure 10) we observed an association of the methylation profile with the driver mutation and the histology creating two main clusters: a first cluster including most normal tissues and a second one including most tumours.

This second cluster was divided in two subculsters: one mainly including the follicular tumours with RAS mutations and a second one mainly including the papillary tumours with BRAF mutations. However, the methylation profile was not associated with the prognosis. Additionally, we saw no difference between the clustering of the probes depending on whether they belonged to enhancers or promoters (left dendogram in the Figure 10).

When analysing the body and intergenic regions, the methylation profile of the samples was not associated with the driver mutations nor the histology nor the prognosis (Supplementary Figure 2).

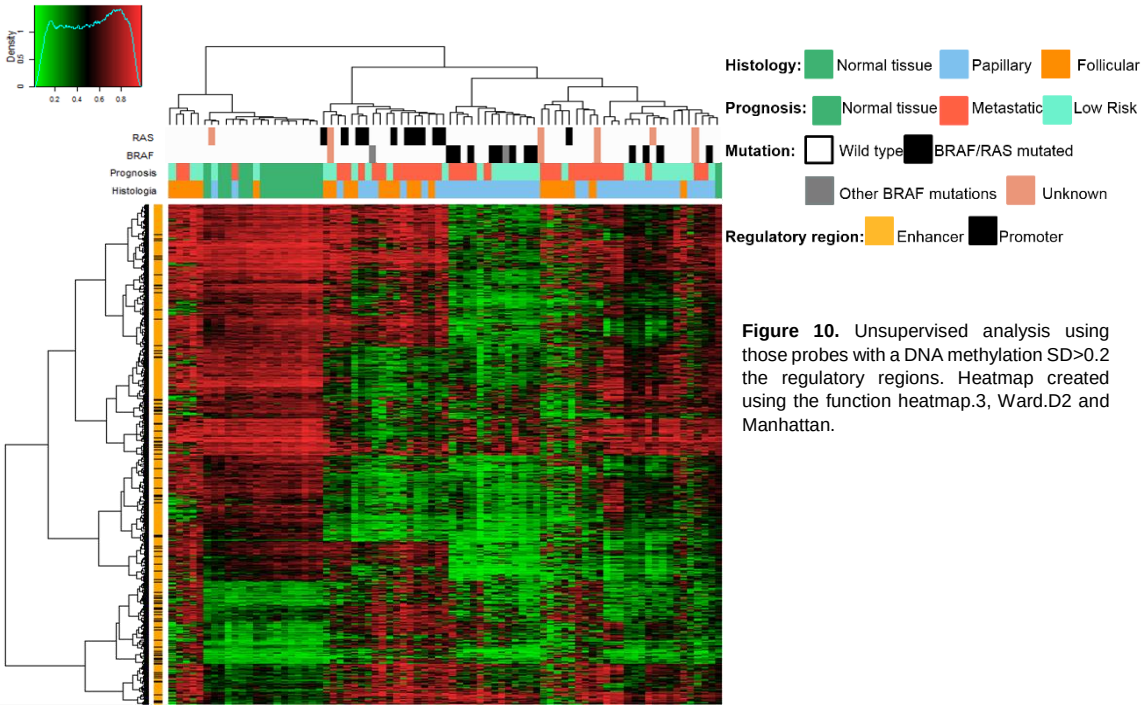


Figure 10. Unsupervised analysis using those probes with a DNA methylation SD>0.2 the regulatory regions. Heatmap created using the function heatmap.3, Ward.D2 and Manhattan.

4.3. The differentially methylated sites specifically associated with FTC or PTC are not related to the prognosis

We also analysed those CpGs that were differentially methylated and specifically associated with the main histological types of TC, FTC and PTC. In the FTC heatmap we observed one cluster with the NT and another with the tumours. Although there is a sub-cluster with most low risk tumours it groups with another sub-cluster of metastatic tumours. In the PTC heatmap, we also observed two clusters, one with the NT and the RAS-mutated tumours and one with the rest of tumours, including all the BRAF-mutated ones (Figure 11).

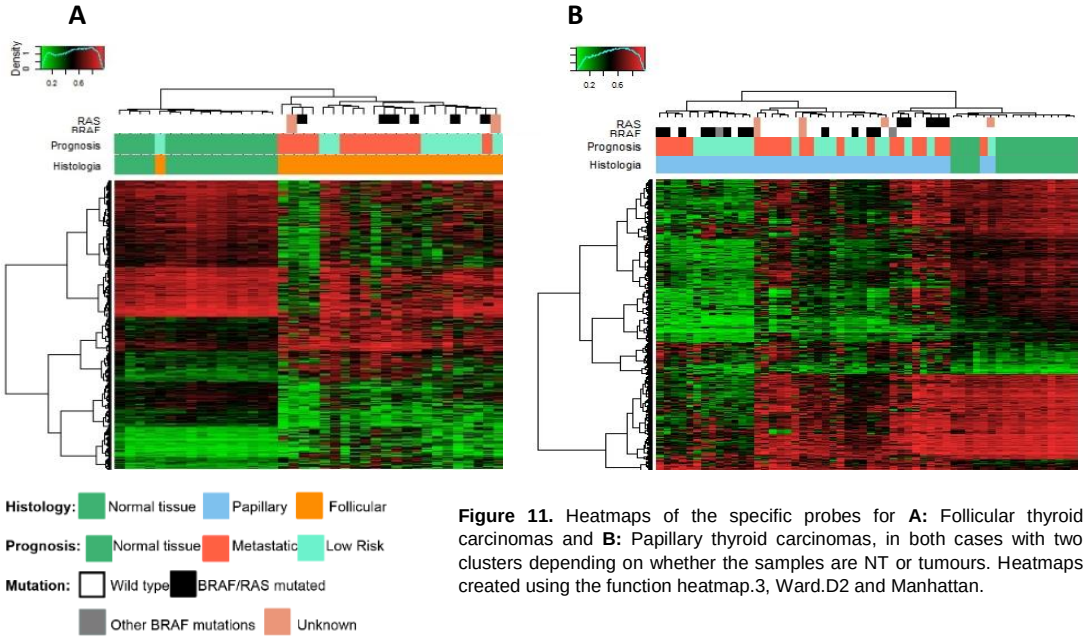


Figure 11. Heatmaps of the specific probes for **A:** Follicular thyroid carcinomas and **B:** Papillary thyroid carcinomas, in both cases with two clusters depending on whether the samples are NT or tumours. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

4.4. Identification of a DNA methylation signature associated with metastatic tumours

From the differential DNA methylation analysis, we had several specific comparisons, but we decided to focus on those related with the metastatic process: Metastatic versus Non-metastatic tumours (or Low risk), consisting of 156 probes; and Metastatic PTC versus Low risk PTC, consisting of 38 probes. It is important to remark that 15 of these probes were common within the two comparisons (Figure 12). When we compared the Metastatic FTC with the Low risk FTC we obtained only one CpG that was differentially methylated, probably due to the low number of tumours, and we did not use it in further analyses. Interestingly, it was included in the 156-CpG signature.

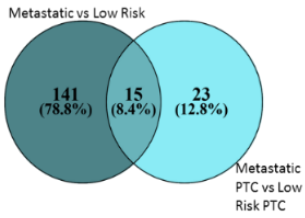


Figure 12. Venn diagram showing the common probes between the comparisons Metastatic versus Low risk and Metastatic PTC versus Low risk PTC.

For both comparisons, we performed two hierarchical clusters, one including metastatic tissues (Figure 13) and one without them (Figure 14).

When we analysed all the samples together, including the metastasis, we observed how the samples were grouped in two clusters: one for the metastatic samples (tumours and metastasis) and one that was divided into two sub-clusters, one with the NT and one with the low risk tumours, indicating that the DNA methylation profile of this signature is similar between NT and low risk tumours. When we represented the metastasis with the same colour as their primary tumour we observed how most of the

metastasis were paired with their primary tumour, or at least were very close (Figure 14), indicating that the DNA methylation profile of the metastasis was similar to the matched primary tumour.

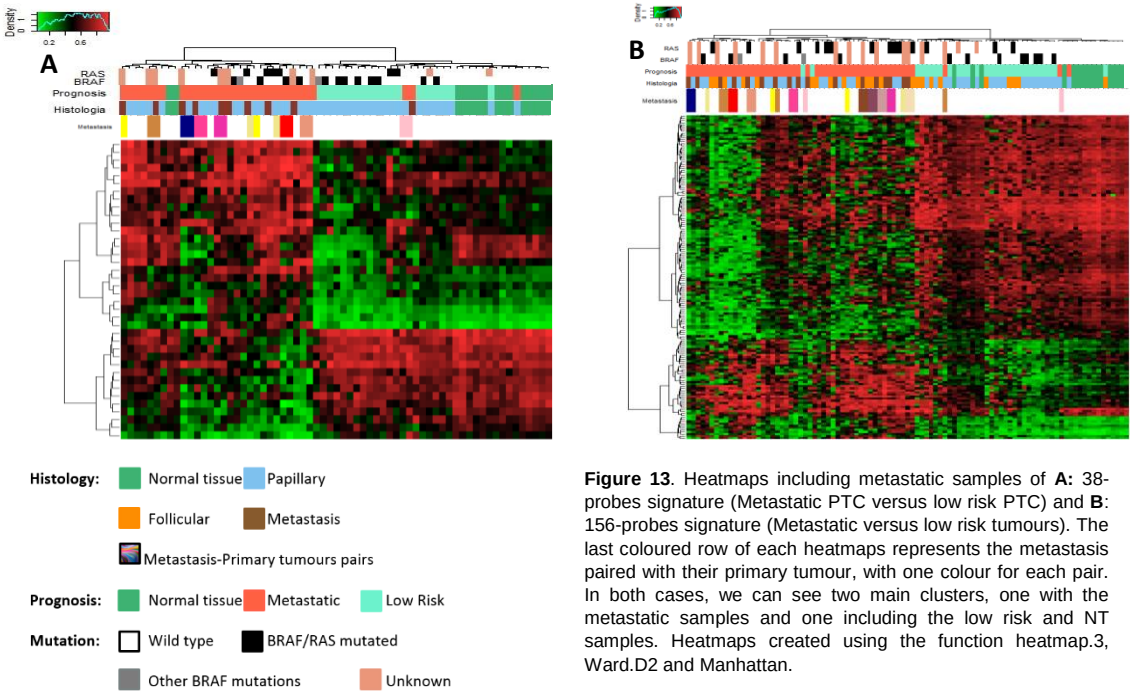


Figure 13. Heatmaps including metastatic samples of **A:** 38-probes signature (Metastatic PTC versus low risk PTC) and **B**: 156-probes signature (Metastatic versus low risk tumours). The last coloured row of each heatmaps represents the metastasis paired with their primary tumour, with one colour for each pair. In both cases, we can see two main clusters, one with the metastatic samples and one including the low risk and NT samples. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

Then we performed the hierarchical cluster analysis without the metastatic tissues, and obtained similar results: two clusters, one for the metastatic tumours and one that was divided into a NT, a low risk tumours and a metastatic tumours sub-clusters.

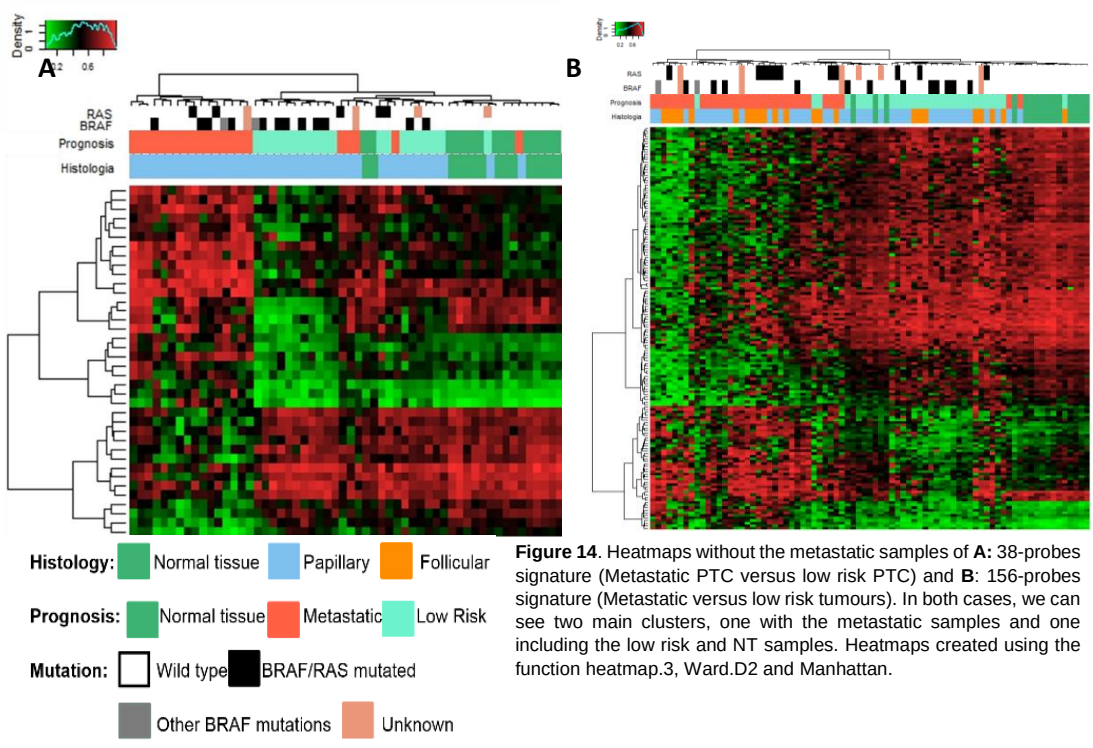


Figure 14. Heatmaps without the metastatic samples of **A:** 38-probes signature (Metastatic PTC versus low risk PTC) and **B**: 156-probes signature (Metastatic versus low risk tumours). In both cases, we can see two main clusters, one with the metastatic samples and one including the low risk and NT samples. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

Neither the 38- nor the 156-CpGs signatures were associated with the driver mutations or the histology.

4.5. Identification of a sub-group of metastatic tumours with worse prognosis features

We then added the cell lines series to the discovery series. We observed how the three of them clustered together in a sub-cluster of metastatic tumours, concretely with a subgroup of lowly methylated metastatic tumours (Figure 15). Since the cell lines are dedifferentiated this could indicate that those tumours had a worse prognosis.

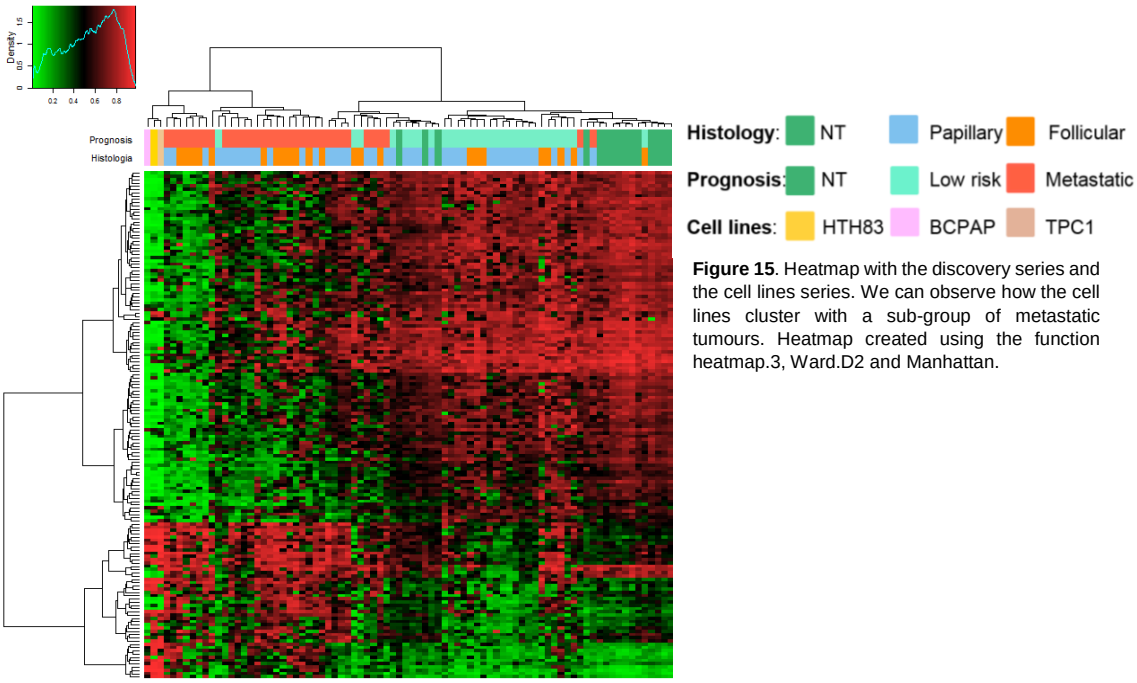


Figure 15. Heatmap with the discovery series and the cell lines series. We can observe how the cell lines cluster with a sub-group of metastatic tumours. Heatmap created using the function heatmap.3, Ward.D2 and Manhattan.

When we analysed the clinical data of this subset of metastatic tumours we observed that they tended to have some worse prognosis features (Supplementary figure 3), thus we compared this group with the rest of the metastatic samples and performed a Fisher test, but we did not obtain any significant result (Table 2).

		Bad n(%)	Worst n(%)	p-value
BRAF	BRAF_O	0 (0%)	1 (16.67%)	0.1188
	BRAFV600E	3 (12%)	1 (16.67%)	
	WT	22 (88%)	4 (66.67%)	
RAS	RASmut	7 (28%)	1 (16.67%)	1
	WT	18 (72%)	5 (83.33%)	
Nodules	Affected	14 (70%)	5 (83.33%)	1
	Not affected	6 (30%)	1 (16.67%)	
Metastasis_Location	Bone	8 (33.33%)	2 (33.33%)	0.2239
	Lung	8 (33.33%)	0 (0%)	
	Multi	8 (33.33%)	4 (66.67%)	
Persistence_Recurrence	Persistence	18 (81.81%)	5 (71.43%)	1
	Recurrence	4 (18.18%)	2 (28.57%)	
Follow_up	Active Disease	7 (35%)	2 (40%)	1
	Death, other	1 (5%)	0 (0%)	
	Death, TC	10 (50%)	3 (60%)	
	Disease free	2 (10%)	0 (0%)	
QUAlu	Hypomethylated	5 (25%)	3 (50%)	0.3301
	Not Hypomethylated	15 (75%)	3 (50%)	
Histo	FTC	10 (37.04%)	3 (42.86%)	1
	PTC	17 (62.96%)	4 (57.14%)	

Table 2. Table showing those clinical characteristics that are significant according to the Fisher test. Bad represents the metastatic tumours, while Worst represents the subgroup with, apparently, worse prognosis features. Pval: 0.05

4.6. Validation analysis

We used data from TCGA to validate our signature. However, TCGA used the Methylation450K array which covers only 53 probes from the 156. First we checked that these 53 CpGs were still able to distinguish in our discovery series the metastatic tumours from the non-metastatic ones (Supplementary figure 4). Next we performed the heatmap with TCGA series. Results did not show the expected clustering of NT and low risk tumours on one side and metastatic samples on the other (Figure 16A). We also used this 53-CpGs signature with the PDTC/ATC series and obtained the expected clustering. In this case all tumours are considered of bad prognosis and they all grouped together but subdivided depending on the metastatic capacity, with the NT in a different cluster (Figure 16B).

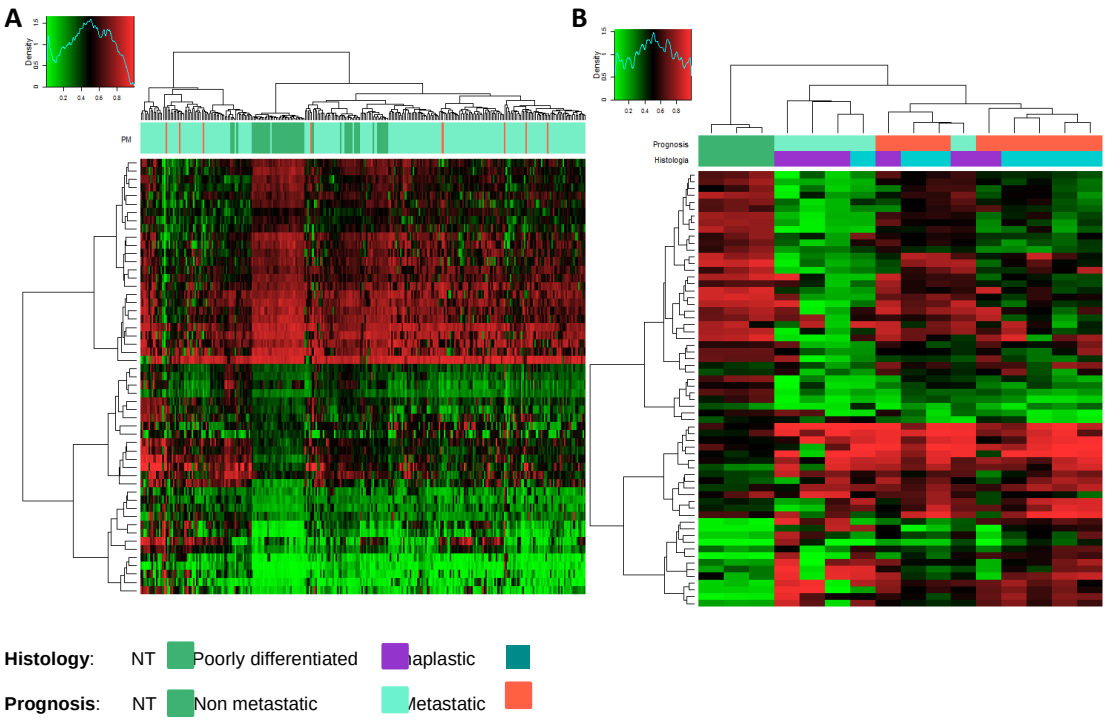


Figure 16. A: Heatmap with the TCGA data of TC, were we cannot see the expected cluster of bad prognosis samples. **B:** Heatmap with the PDTC/ATC series, with one cluster with the NT and one with two sub-clusters, one with the non-metastatic and one with the metastatic tumours. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

4.7. Reduction of the 156-CpGs signatures to a 10-CpG signature

In order to reduce the number of probes that distinguish metastatic from non-metastatic tumours we used the SimpleLogistics tool from Weka. We obtained a series of 34 probes (Supplementary Figure 2). We also calculated the effect size of all the 156 probes (Supplementary Figure 3). We then kept the first ten probes among the 34 from Simple Logistics with a higher effect size. When we perform a heatmap using these 10 probes, tumours are still clustering based on the metastatic capacity, that is, there are 2 clusters, one including Low Risk tumours and one including Metastatic tumours (Figure 17).

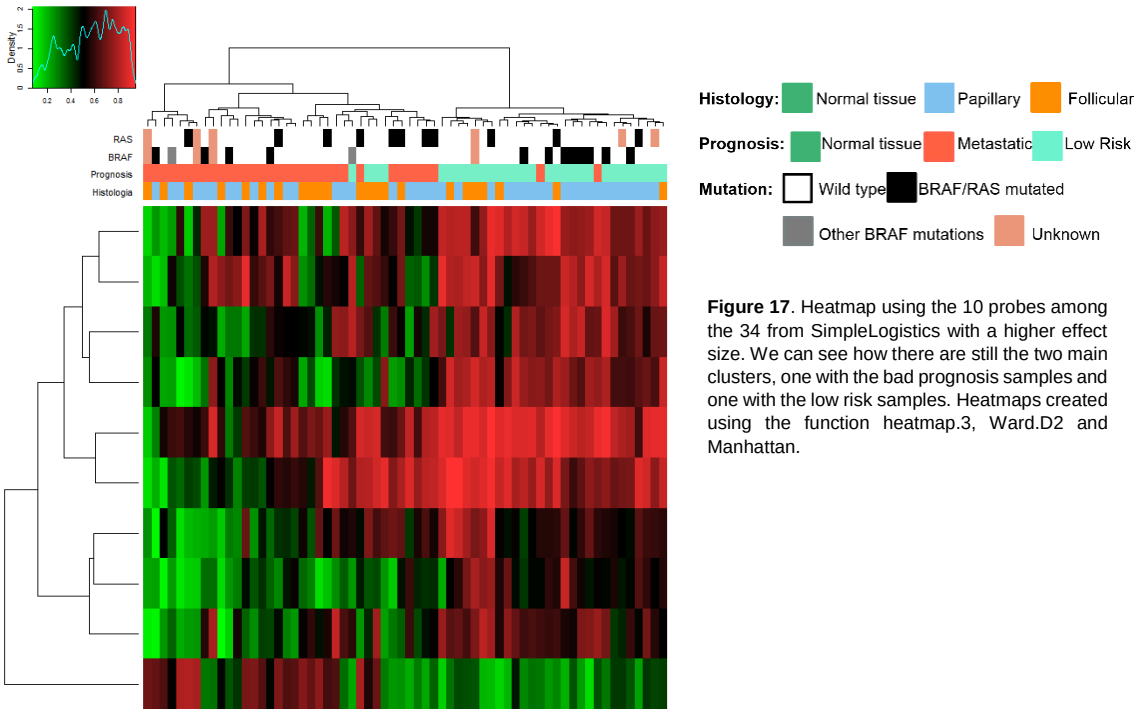


Figure 17. Heatmap using the 10 probes among the 34 from SimpleLogistics with a higher effect size. We can see how there are still the two main clusters, one with the bad prognosis samples and one with the low risk samples. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

4.8. The 156-CpGs signature is thyroid specific

Our final analysis was done using different cancers' data from TCGA and the 53 probes from the 450K array. In none of the cases the clustering was associated with the prognosis, only the NT seemed to group together, and not in all the cases. In Figure 18 you can see one example, COAD; the rest can be found in the Supplementary Figure 4.

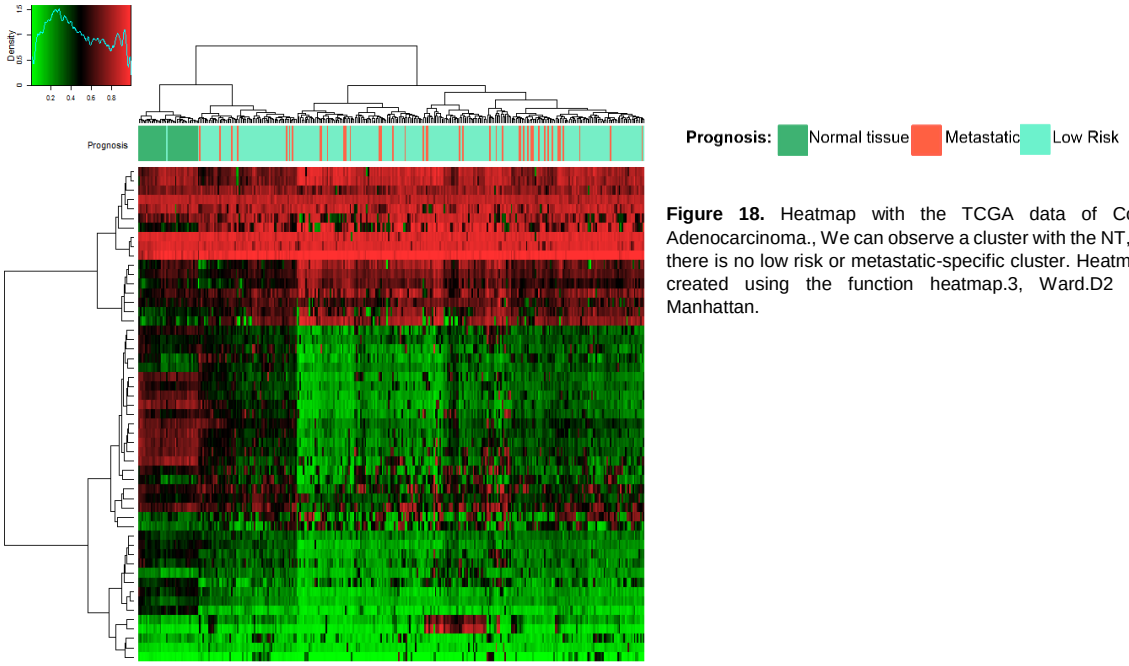


Figure 18. Heatmap with the TCGA data of Colon Adenocarcinoma., We can observe a cluster with the NT, but there is no low risk or metastatic-specific cluster. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

5. Discussion

The incidence and mortality of TC has increased in the recent years [3]. The increase of the incidence (mostly of WDTC) can be due to the more efficient diagnosis tools or, according to some authors, also to some environmental (e.g., dietary) or genetic factors [14]. The increase in mortality, tough low, is partly caused by the inappropriate diagnosis and treatment of the most aggressive tumours. In fact, while the survival rate of WDTC has increased in the recent years, there has not been any increase in the survival of metastatic WDTC over the past two decades [44]. These aggressive tumours tend to be poorly diagnosed, and its outcome is usually incorrectly predicted.

This, together with the opinion of some authors that the current treatment for TC (i.e., total thyroidectomy and RAI) may be unnecessary for the less aggressive tumours creates a need for new prognosis markers of TC.

DNA methylation has proven to be a good biomarker in many cancers [30], but until now there are not many studies on DNA methylation in TC, and none of them focuses on prognosis.

As we saw on the first global analysis, metastatic tumours have more changes in DNA methylation than low risk tumours when comparing them with NT. We also saw differences on DNA methylation when comparing metastatic versus low risk tumours, suggesting that DNA methylation is involved in the metastatic capacity of primary tumours and that some changes may be used as prognosis markers.

In the supervised analyses, our results revealed that the DNA methylation profile of regulatory elements were associated with the histology and the mutation in BRAF/RAS, but not that of intergenic regions and body genes. These results concurred with previous studies [17], [18], [19], [20], [21].

Even if the global methylation profiles did not seem to correlate with the metastatic capacity of the tumours we did find some differentially methylated CpGs between metastatic and none metastatic tumours. These CpGs were 38 when comparing only papillary tumours or 156 when comparing metastatic versus non metastatic rumours regardless their histological type. Moreover, when we added the metastatic samples to the primary tumours in the clustering analysis most of them paired together, suggesting that the DNA methylation of the 156-CpGs signature does not change during the metastatic process. Furthermore, although most metastatic were present at the diagnosis of the TC, some tumours developed metastasis later and clustered with the previous ones, suggesting the predictive potential of this signature.

In these last two analyses (metastatic PTC versus low risk PTC and metastatic versus low risk) that the NT samples tended to cluster with the low risk tumour samples

proving that low risk tumours were more reminiscent of NT than metastatic tumours, as we saw in the first global DNA-methylation analysis.

Since the clustering in relation to the prognosis was more accurate in the 156-probes signature we decided to focus on it, and moved to analyse the clinical data of the samples.

We tried to validate our results using TCGA's data of TC, but we obtained negative results. This can be due to the lack of metastatic samples among the TCGA samples (8 metastatic tumours out of 493 tumours) and the fact that there is a huge variety among the non-metastatic samples from TCGA, while our non-metastatic samples were of extremely good prognosis. Another reason could be that the methylation array used by TCGA is different, or that the samples were not embedded in paraffin, like ours. Nevertheless we could validate our results with the PDTC/ATC series, even if this were also analysed with the 450K array. The tumours from this series are all considered of bad prognosis, but despite this we could separate metastatic from not metastatic tumours (what we called low risk and metastatic during the analyses with the discovery series). Also, the NT were in a different clustering instead than with the non-metastatic.

However, further validation is needed and at the moment there is not any publicly available DNA methylation data for metastatic thyroid tumours, so we are now recollecting more samples to further validate this.

Interestingly, in the clustering using the 156 signature, we observed a group of metastatic tumours that were highly unmethylated with a similar DNA methylation profile to the cell lines, suggesting that these tumours could have a worse prognosis. This is because the cell lines of TC dedifferentiate in culture.

To confirm these preliminary results we analysed some clinical data. We observed some trends suggesting a worse prognosis for these tumours, such as a higher hypomethylation of Alu elements, but the differences were not statistically significant probably due to the low number of samples in this subgroup is small, thus a validation analysis is needed.

With the clinical implementation as objective, we used WEKA to create a signature with a lower number of probes but equally efficient. We analysed them using SimpleLogistics and a 10-fold cross-validation. We obtained a total of 34 probes, but we decided to keep only those 10 with a higher effect size, so as to avoid redundancy. However, this 10-CpGs signature need further validation, and the CpGs will be set up for bisulphite sequencing to analyse them in a bigger set of samples.

Finally we tried to see if this signature could be applied to other cancers, but the result was negative, suggesting that it is a TC-specific signature.

Aside from all this other members of the laboratory are analysing the 156 CpGs, especially those located in promoters (27%) to assess the functional implications of their differences in DNA methylation. In particular, they are focusing on ELTD1 and FOXD3, a receptor involved in the regulation of angiogenesis and a transcription factor respectively [45], [46].

6. Conclusions

- Thyroid tumours tend to have more hypomethylations than hypermethylations when comparing them to NT. These changes in methylation are more abundant in FTCs than PTCs, and in metastatic tumours than low risk tumours.
- The profiles of the CpGs with the highest variability in DNA methylation located in regulatory regions, but not in intergenic regions and gene bodies, are specifically associated with the histology and mutation in *BRAF* and *RAS*.
- We identified a 38-CpGs that distinguish between metastatic and non-metastatic papillary thyroid carcinomas.
- We identified a 156-CpGs signature that distinguish non-metastatic from metastatic thyroid carcinomas independently of the histology and the driver mutation. It is thyroid cancer-specific and does not change during the metastatic process.
- Preliminary results based on the 156-CpGs signature suggest the existence of a subgroup of metastatic tumours with a worse prognosis.
- The 156-probes signature can be reduced to a 10-probes signature, a more clinically manageable number.

7. Bibliography

1. Tortora G.J., Derrickson, B., 2012. Principles of anatomy & physiology, 13th edition. John Wiley & Sons, Inc.

2. Toyota M., Suzuki H., 2010. “Epigenetic drivers of genetic alterations”. *Advances in genetics* doi:10.1016/B978-0-12-380866-0.60011-3.

3. Kitahara C.M., Sosa J.A., 2016. “The changing incidence of thyroid cancer”. *Nature Reviews* 12: 646-52.

4. National Cancer Institute: Surveillance, Epidemiology and End Results Program: https://seer.cancer.gov/statfacts/html/thyro.html

5. World Health Organization: http://www.who.int/topics/risk_factors/en/.

6. Gara et al, 2015. “Germline HABP2 Mutation Causing Familial Non-medullary Thyroid Cancer”. *New England Journal of Medicine* 373: 448-55.

7. Xing M., 2013. “Molecular pathogenesis and mechanisms of thyroid cancer”. *Nature Reviews* 13: 184-99.

8. Ricarte-Filho, J.C., et al, 2009. “Mutational Profile of Advanced Primary and Metastatic Radioactive Iodine-Refractory Thyroid Cancers Reveals Distinct Pathogenetic Roles for BRAF, PIK3CA, and AKT1”. *Cancer Research* 69: 4885-93.

9. Bose S., Walts A.E., 2012. “Thyroid fine needle aspiration: A post-Betshada Update”. *Advanced anatomy and pathology* 19: 160-69.

10. Scopa, C.D., 2004. “Histopathology of Thyroid Tumours. An overview”. *Hormones* 3: 100-10.

11. Adeniran, A.J., Chieng, D., 2016. Variants of Papillary Thyroid Carcinoma, in: Common Diagnostic Pitfalls in Thyroid Cytopathology. Springer International Publishing.

12. Durante, C., et al, 2006. “Long-Term Outcome of 444 Patients with Distant Metastases from Papillary and Follicular Thyroid Carcinoma: Benefits and Limits of Radioiodine Therapy”. *The journal of clinical endocrinology and metabolism* 91: 2892-99.

13. Riesco-Eizaguirre, Garcilaso, Pilar Santisteban, 2016. “ENDOCRINE TUMOURS: Advances in the Molecular Pathogenesis of Thyroid Cancer: Lessons from the Cancer Genome.” *European Journal of Endocrinology* 175: 203–17.

14. Nikiforov Y.E., Nikiforova M.N., 2011. “Molecular genetics and diagnosis of thyroid cancer”. *Nature Reviews Endocrinology* 7: 569-80.

15. Boufraquech M., Kluobo-Gwiezdzinska J., Kebebew E., 2016. “MicroRNAs in the thyroid”. *Best practice and research Clinical Endocrinology & Metabolism*, doi: 10.1016/j.beem.2016.10.001.

16. Faam et al, 2015. “Epigenetic modifications in human thyroid cancer”. *Biomedical reports* 3: 3-8.

17. Mancikova, V. et al, 2014. “DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival”. *International Journal of Cancer* 135: 598-610.

18. Ellis et al, 2013. “Genome-Wide Methylation Patterns in Papillary Thyroid Cancer Are Distinct Based on Histological Subtype and Tumour Genotype”. The journal of clinical endocrinology and metabolism” 99: 329-337.

19. Kikuchi et al, 2013. “Aberrantly methylated genes in human papillary thyroid cancer and their association with BRAF/RAS mutation” *Frontiers in genetics* 4: Article 271.

20. White, M.G. et al, 2016. “Epigenetic Alterations and Canonical Pathway Disruption in Papillary Thyroid Cancer: A Genome-wide Methylation Analysis”. Annals of Surgical Oncology 23: 2302–09.

21. TCGA, 2014. “Integrated Genomic Characterization of Papillary Thyroid Carcinoma”. *Cell* 159: 676-90.

22. Bird, A., 2007. “Perceptions of epigenetics”. *Nature* 447: 396-398.

23. Portela A., Esteller M., 2010. “Epigenetic modifications and human disease”. *Nature: Biotechnology* 28:1057-68.

24. Jones P.A., 2012 “Functions of DNA methylation: islands, start sites, gene bodies and beyond”. *Nature Reviews: Genetics* 13: 484–92.

25. Wu X., Zhang Y., 2017. “TET-mediated active DNA demethylation: mechanism, function and beyond”. *Nature Reviews: Genetics.* doi: 10.1038/nrg.2017.33.

26. Kohli R.M., Zhang, Y., 2013. “TET enzymes, TDG and the dynamics of DNA methylation”. *Nature* 502: 472-79.

27. Deaton A.M., Bird A., 2011. “CpG islands and the regulation of transcription”. *Genes and development* 25:1010-22.

28. Timp W. et al, 2014. “Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumours”. *Genome medicine* 6:61.

29. Esteller M., 2008. “Epigenetics in cancer”. *The New England Journal of Medicine* 358:1148-59.

30. Moran, S. et al, 2016. “Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis”. *The lancet Oncology* 10: 1386-95.

31. Sheaffer K.L. et al, 2016. “DNA hypomethylation contributes to genomic instability and intestinal cancer initiation”. *Cancer Prevention Research* doi:10.1158/1940-6207.

22. Buj et al, 2016. “Quantification of Unmethylated Alu (QUAlu): a tool to assess global hypomethylation in routine clinical samples”. *Oncotarget* 7: 10536-46.

33. Atlas of Genetics and Cytogenetics in Oncology and Haematology: http://atlasgeneticsoncology.org/Deep/DNAMethylationID20127.html

34. Saiselet M., Floor S. et al, 2012. “Thyroid cancer cell lines: an overview”. *Frontiers in Endocrinology* 3: article 133.

35. TCGA, 2014. “Comprehensive molecular characterization of gastric adenocarcinoma”. *Nature* 513: 202-09.

36. TCGA, 2014. “Comprehensive molecular characterization of lung adenocarcinoma”. *Nature* 511: 543-50.

37. TCGA, 2012. “Comprehensive molecular characterization of human colon rectal cancer”. *Nature* 487: 330-7.

38. TCGA, 2013. “Comprehensive molecular characterization of clear cell renal cell carcinoma”. *Nature* 499: 43-9.

39. Pidsley et al, 2016. “Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling” *Genome Biology* 17: 208-25.

40. Bibikova, M., et al, 2009. “Genome-wide DNA methylation profiling using Infinium ® assay”. *Epigenomics* 1: 177-200.

41. Assenov, Y. et al, 2014. “Comprehensive analysis of DNA methylation data with RnBeads”. Nature methods 11: 1138-45.

42. Frank E. et al, 2004. “Data mining in bioinformatics using Weka”. *Bioinformatics applications note* 20: 2479-81.

43. Galaxy: https://usegalaxy.org/.

44. Goffredo P et al, 2013. “Differentiated Thyroid cancer presenting with distant metastases: A population analysis over two decades”. *World journal of surgery* 37: 1599-1605.

45. Favara D.M., Banham A.H., Harris A.L., 2014. A review of ELTD1, a pro-angiogenic adhesion GPCR “”. *Biochemical society transactions* 6: 1658-64.

46. Yin H., Meng T. et al, 2017. “FOXD3 regulates anaplastic thyroid cancer progression”.*Oncotarget* 8:33644-51.

8. Supplementary Figures

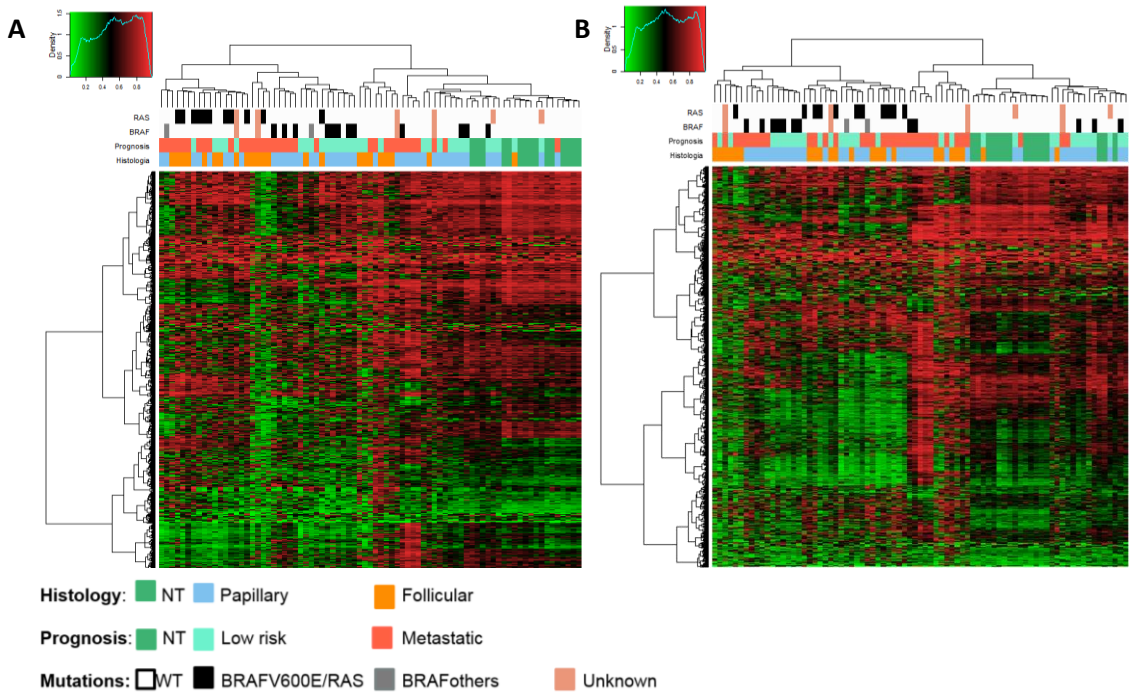
A

DIFFERENTIAL METHYLATED SITES	Normal = NT (15)	Tumour (67)	Metastatasis = MET (1	PTC (42)	FTC (22)	ATCPDTC = Bad_ATCI	Bad (37)	Good (30)	Good PTC (20)	Bad PTC (22)	Good FTC (10)	Bad FTC (12)	Distal (11)	Proximal (6)
Normal = NT (15)														
Tumour (67)	3628													
Metastatasis = MET (17)	17166	0												
PTC (42)	3922		0											
FTC (22)	6602		9	1007										
ATCPDTC = Bad_ATCPDTC (3)	35260		0	227	4108									
Bad (37)														
Good (30)							156							
Good PTC (20)	6343		1938			12001								
Bad PTC (22)	6493		0			3			38					
Good FTC (10)	3323		2			0			1028	0				
Bad FTC (12)	17069		0			174			5797	0	1			
Distal (11)							0							
Proximal (6)							0						0	

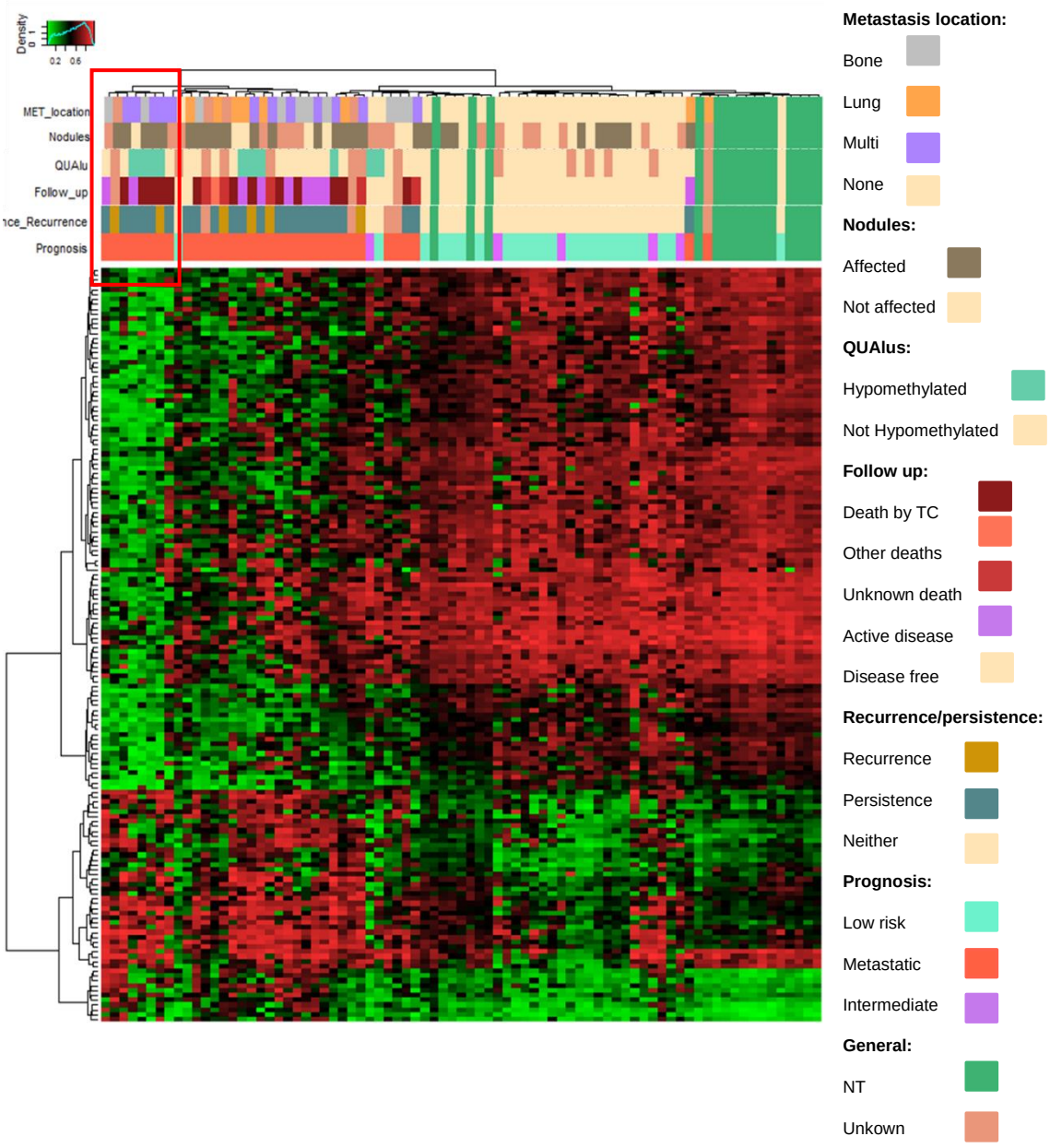
B

HYPERMETHYLATIONS HYPOMETHYLATIONS	Normal = NT (15)	Tumour (67)	Metastatasis = MET (1	PTC (42)	FTC (22)	ATCPDTC = Bad_ATCI	Bad (37)	Good (30)	Good PTC (20)	Bad PTC (22)	Good FTC (10)	Bad FTC (12)	Distal (11)	Proximal (6)
Normal = NT (15)		881	4747	628	2087	21218			541	2175	1605	3985		
Tumour (67)	2747		0											
Metastatasis = MET (17)	12419	0		0	0	0			749	0	0	0		
PTC (42)	3294		0		828	166								
FTC (22)	4515		9	179		2921								
ATCPDTC = Bad_ATCPDTC (3)	14042		0	61	1187				9555	2	0	152		
Bad (37)								48					0	0
Good (30)							108							
Good PTC (20)	5802		1198			2446				24	952	2713		
Bad PTC (22)	4318		0			1			14		0	0		
Good FTC (10)	1718		2			0			76	0		0		
Bad FTC (12)	13084		0			22			3084	0	1			
Distal (11)							0							0
Proximal (6)							0						0	

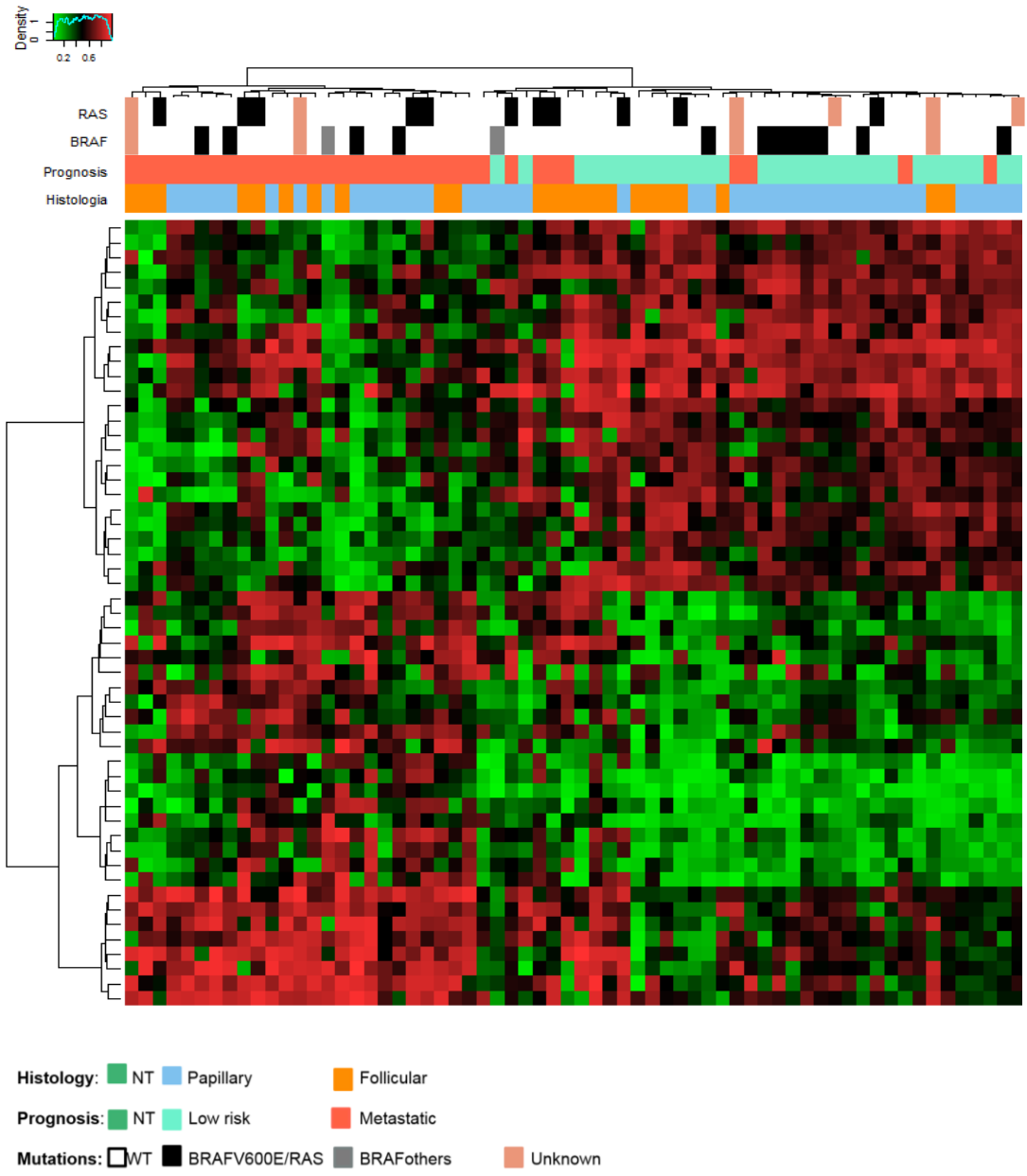
Supplementary figure 1. A: Differentially methylated sites when comparing the samples depending on their histology (PTC: papillary thyroid cancer; FTC: follicular thyroid cancer; ATC/PDTC: anaplastic/poorly differentiated thyroid cancer), prognosis (good or bad), and location (distal or proximal metastasis). **B**: Methylation state (hyper- or hypomethylation) of the differential methylated sites from A. The number of samples in each group is indicated in brackets).



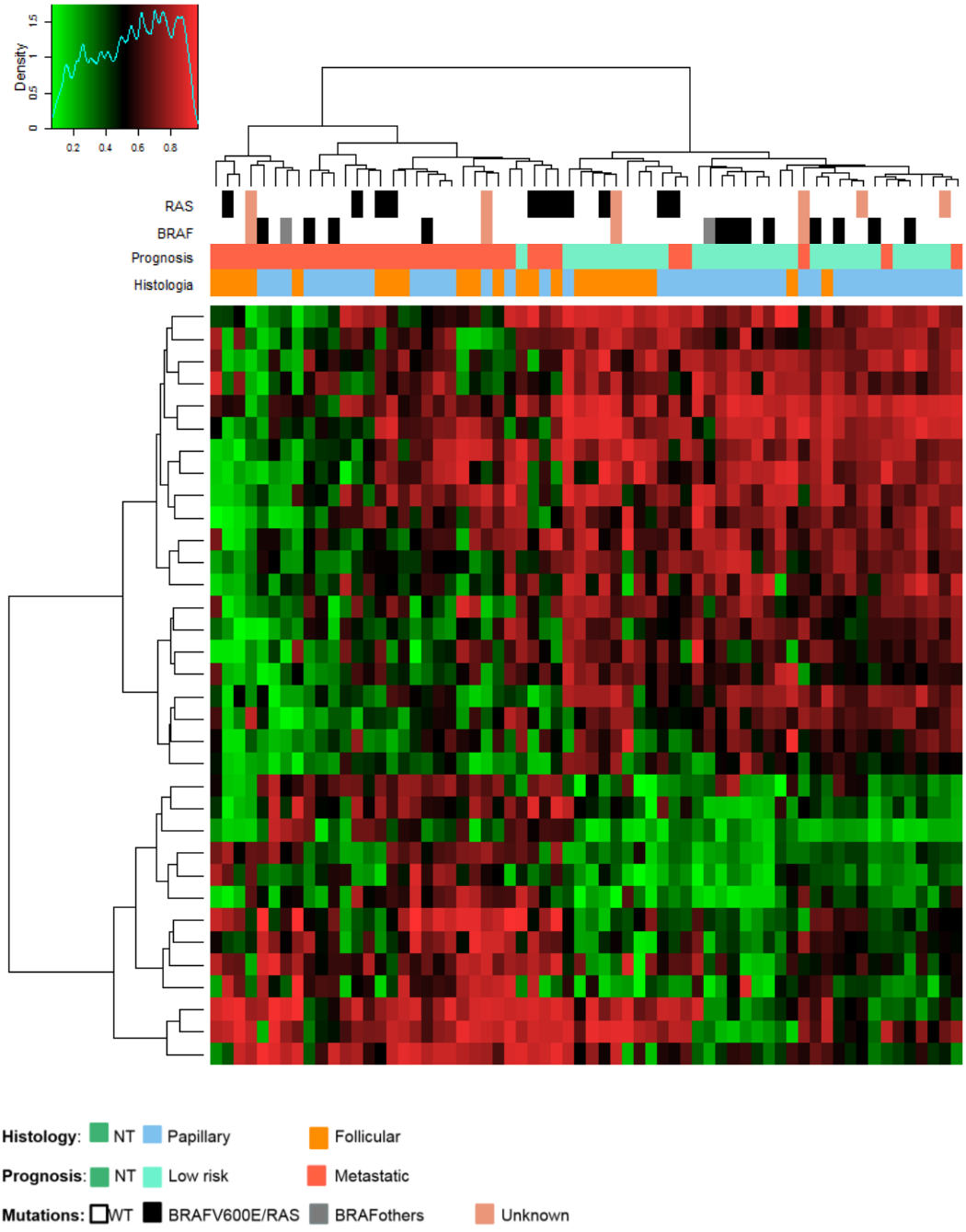
Supplementary Figure 2. Unsupervised analysis using those probes with a DNA methylation SD>0.2 for the intergenic (**A**) and gene body regions (**B**). Heatmap created using the function heatmap.3, Ward.D2 and Manhattan.



Supplementary figure 3. Heatmap representing the clinical data of the samples and using the 156-probes signature (Metastatic vs Low risk). We can see a sub-cluster among the metastatic tumour with worse prognosis characteristics (Red square). We added a new prognosis type, intermediate, which includes those low risk tumours graded T_{III} or T_{IV}. In all rows we can see the best possibility in pale yellow and the unknown cases in salmon. Heatmap created using the function heatmap.3, Ward.D2 and Manhattan.



Supplementary figure 4. Heatmap with the 53 CpGs from the 156-CpGs signature found in the 450K array. Heatmap created using the function heatmap.3, Ward.D2 and Manhattan.



Supplementary Figure 5. Heatmap created using the 34 CpGs differentially methylated when comparing metastatic versus non metastatic tumours with a higher impact when analysing the 156-CpGs signature with SimpleLogistics. We can observe two main clusters, one with most metastatic tumours and another with most of the non-metastatic ones. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.

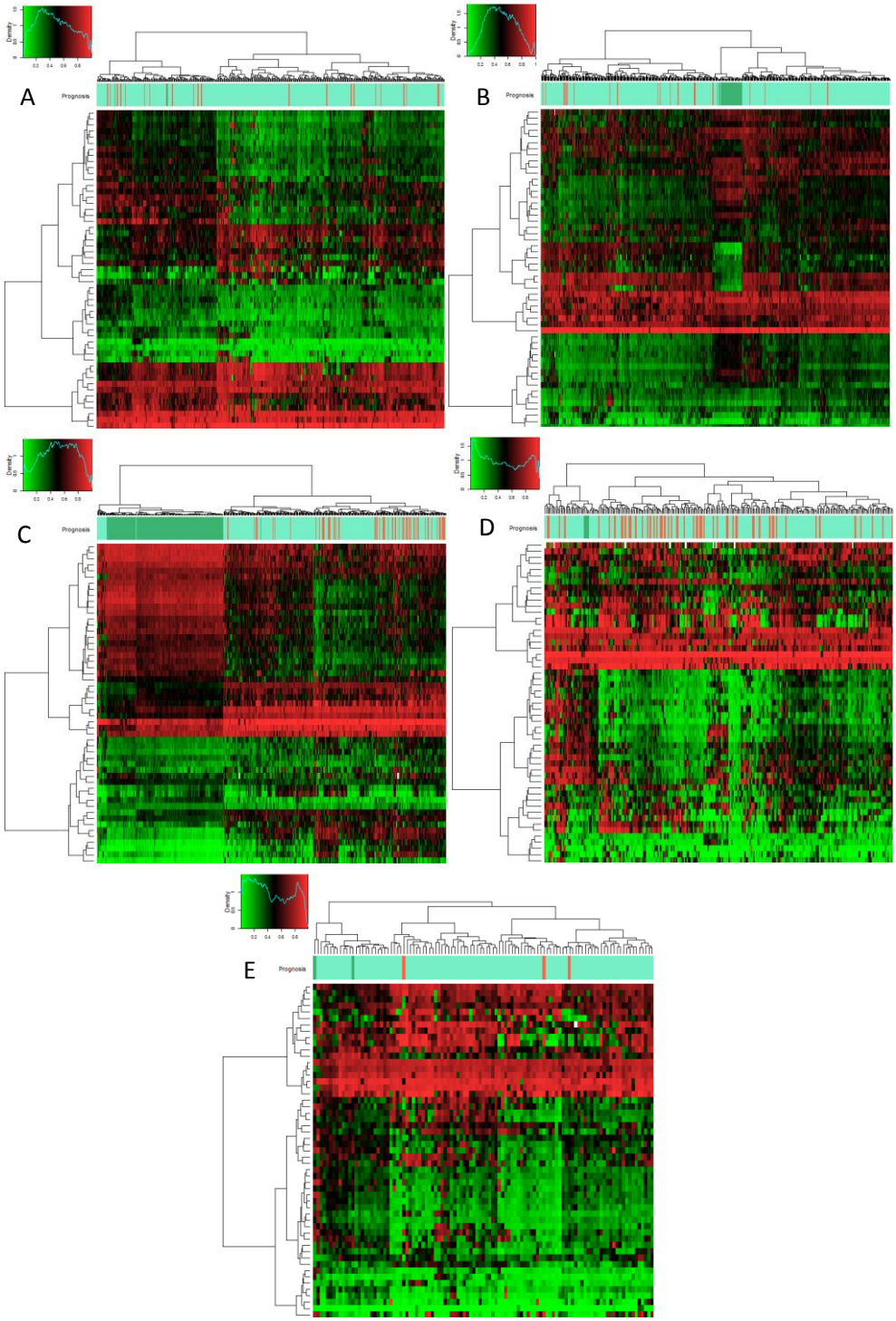
	CPG	Good (n=30)		Bad (n=37)				
rank	id	Media	dt	Media	dt	dif	efect size	sd pooled
1	cg25001805	0.490	0.133	0.694	0.128	0.20	1.57	0.130
2	cg01378044	0.698	0.140	0.424	0.199	-0.27	1.56	0.175
3	cg23450358	0.493	0.146	0.247	0.170	-0.25	1.54	0.160
4	cg16326851	0.605	0.144	0.381	0.154	-0.22	1.50	0.150
5	cg20009810	0.775	0.085	0.523	0.213	-0.25	1.50	0.168
6	cg03478630	0.735	0.084	0.531	0.169	-0.20	1.49	0.137
7	cg18916149	0.703	0.115	0.433	0.222	-0.27	1.48	0.182
8	cg11328079	0.490	0.154	0.267	0.148	-0.22	1.47	0.151
9	cg18157275	0.519	0.157	0.313	0.127	-0.21	1.46	0.141
10	cg10589286	0.660	0.110	0.431	0.187	-0.23	1.46	0.157
11	cg18305394	0.808	0.090	0.548	0.228	-0.26	1.45	0.180
12	cg11925561	0.850	0.082	0.644	0.180	-0.21	1.43	0.144
13	cg17699869	0.744	0.094	0.515	0.200	-0.23	1.42	0.161
14	cg13985307	0.599	0.123	0.369	0.191	-0.23	1.41	0.164
15	cg04562441	0.539	0.154	0.334	0.140	-0.20	1.40	0.147
16	cg09918581	0.669	0.113	0.453	0.182	-0.22	1.39	0.155
17	cg25175009	0.657	0.124	0.426	0.195	-0.23	1.38	0.167
18	cg13453847	0.712	0.110	0.487	0.198	-0.23	1.37	0.164
19	cg25743622	0.681	0.146	0.470	0.162	-0.21	1.36	0.155
20	cg18279094	0.171	0.129	0.410	0.206	0.24	1.36	0.176
21	cg22424903	0.318	0.125	0.546	0.198	0.23	1.35	0.169
22	cg05304729	0.587	0.130	0.380	0.170	-0.21	1.35	0.153
23	cg01611504	0.556	0.108	0.353	0.179	-0.20	1.35	0.151
24	cg09262300	0.720	0.106	0.513	0.187	-0.21	1.33	0.156
25	cg04206637	0.618	0.144	0.386	0.198	-0.23	1.32	0.176
26	cg06626892	0.534	0.118	0.332	0.177	-0.20	1.31	0.154
27	cg07826516	0.633	0.132	0.421	0.183	-0.21	1.30	0.162
28	cg15325961	0.625	0.124	0.421	0.178	-0.20	1.30	0.156
29	cg09783969	0.645	0.133	0.435	0.180	-0.21	1.30	0.161
30	cg20997704	0.783	0.100	0.546	0.229	-0.24	1.30	0.183
31	cg02762107	0.688	0.113	0.460	0.216	-0.23	1.29	0.177
32	cg25075684	0.616	0.130	0.396	0.198	-0.22	1.29	0.171
33	cg27567836	0.559	0.157	0.357	0.159	-0.20	1.28	0.158
34	cg27039467	0.536	0.164	0.750	0.171	0.21	1.28	0.167
35	cg26998150	0.632	0.099	0.430	0.194	-0.20	1.28	0.159
36	cg03974619	0.621	0.131	0.394	0.208	-0.23	1.27	0.178
37	cg17659636	0.552	0.148	0.348	0.170	-0.20	1.27	0.160
38	cg14672100	0.741	0.106	0.528	0.205	-0.21	1.27	0.168
39	cg23117999	0.642	0.146	0.429	0.183	-0.21	1.27	0.168
40	cg16768018	0.149	0.088	0.349	0.198	0.20	1.26	0.159
41	cg00237606	0.778	0.095	0.562	0.214	-0.22	1.26	0.171
42	cg26976756	0.813	0.099	0.612	0.194	-0.20	1.26	0.159
43	cg11083325	0.460	0.204	0.686	0.157	0.23	1.26	0.180
44	cg15366555	0.506	0.192	0.718	0.147	0.21	1.26	0.169

45	cg19387132	0.672	0.119	0.468	0.190	-0.20	1.25	0.163
46	cg13210663	0.540	0.147	0.327	0.187	-0.21	1.25	0.170
47	cg11601663	0.662	0.116	0.453	0.199	-0.21	1.25	0.167
48	cg24749265	0.730	0.104	0.523	0.203	-0.21	1.25	0.166
49	cg22836826	0.672	0.133	0.446	0.214	-0.23	1.24	0.182
50	cg20804418	0.648	0.133	0.427	0.210	-0.22	1.23	0.180
51	cg08241694	0.754	0.100	0.546	0.209	-0.21	1.23	0.169
52	cg16493240	0.628	0.139	0.413	0.200	-0.21	1.22	0.176
53	cg05133853	0.604	0.147	0.372	0.220	-0.23	1.21	0.191
54	cg07530798	0.269	0.109	0.474	0.206	0.20	1.21	0.170
55	cg22438338	0.693	0.109	0.486	0.210	-0.21	1.20	0.173
56	cg19104050	0.683	0.125	0.480	0.198	-0.20	1.20	0.169
57	cg01664151	0.668	0.126	0.466	0.197	-0.20	1.20	0.169
58	cg11963062	0.280	0.126	0.496	0.216	0.22	1.19	0.181
59	cg11412468	0.645	0.174	0.419	0.202	-0.23	1.19	0.190
60	cg19491717	0.767	0.128	0.518	0.256	-0.25	1.19	0.209
61	cg08584061	0.729	0.084	0.521	0.223	-0.21	1.19	0.175
62	cg20925031	0.420	0.152	0.635	0.202	0.22	1.19	0.181
63	cg16681908	0.601	0.164	0.400	0.174	-0.20	1.18	0.170
64	cg26785499	0.671	0.138	0.468	0.195	-0.20	1.18	0.172
65	cg19291618	0.548	0.152	0.347	0.184	-0.20	1.18	0.170
66	cg15832062	0.710	0.114	0.491	0.228	-0.22	1.18	0.186
67	cg06437651	0.611	0.142	0.410	0.192	-0.20	1.17	0.171
68	cg04690464	0.630	0.124	0.427	0.205	-0.20	1.17	0.174
69	cg10092198	0.716	0.109	0.510	0.218	-0.21	1.16	0.177
70	cg16689883	0.759	0.102	0.551	0.222	-0.21	1.16	0.179
71	cg21445398	0.639	0.180	0.421	0.194	-0.22	1.16	0.188
72	cg04980805	0.628	0.135	0.424	0.203	-0.20	1.16	0.176
73	cg24398933	0.515	0.190	0.746	0.206	0.23	1.16	0.199
74	cg22540067	0.802	0.092	0.591	0.230	-0.21	1.16	0.182
75	cg20185461	0.225	0.124	0.428	0.209	0.20	1.15	0.177
76	cg19537600	0.702	0.158	0.494	0.198	-0.21	1.15	0.181
77	cg05053811	0.470	0.149	0.266	0.198	-0.20	1.15	0.178
78	cg20069378	0.595	0.146	0.387	0.207	-0.21	1.14	0.182
79	cg00086247	0.595	0.152	0.389	0.199	-0.21	1.14	0.180
80	cg05647867	0.736	0.097	0.528	0.228	-0.21	1.14	0.181
81	cg14275423	0.668	0.126	0.453	0.226	-0.21	1.14	0.188
82	cg03287339	0.322	0.187	0.563	0.228	0.24	1.14	0.211
83	cg26139866	0.730	0.153	0.524	0.201	-0.21	1.14	0.181
84	cg06603761	0.733	0.134	0.507	0.238	-0.23	1.14	0.198
85	cg12906989	0.804	0.131	0.593	0.221	-0.21	1.14	0.186
86	cg04768709	0.767	0.137	0.552	0.222	-0.21	1.14	0.189
87	cg16455187	0.739	0.098	0.529	0.236	-0.21	1.12	0.187
88	cg19386774	0.717	0.103	0.515	0.223	-0.20	1.12	0.180
89	cg14035553	0.667	0.145	0.458	0.214	-0.21	1.12	0.187
90	cg02442623	0.802	0.140	0.582	0.233	-0.22	1.12	0.197

91	cg04027576	0.681	0.192	0.472	0.184	-0.21	1.12	0.188
92	cg16122778	0.726	0.112	0.517	0.231	-0.21	1.12	0.188
93	cg24364593	0.666	0.175	0.443	0.218	-0.22	1.11	0.200
94	cg11174855	0.220	0.154	0.426	0.208	0.21	1.11	0.186
95	cg06587623	0.747	0.141	0.532	0.228	-0.22	1.11	0.194
96	cg21939911	0.702	0.127	0.499	0.219	-0.20	1.11	0.184
97	cg15819924	0.562	0.157	0.357	0.205	-0.20	1.10	0.185
98	cg17738861	0.360	0.199	0.584	0.207	0.22	1.10	0.204
99	cg04783185	0.772	0.107	0.568	0.228	-0.20	1.10	0.184
100	cg04314280	0.565	0.236	0.787	0.170	0.22	1.10	0.202
101	cg14861920	0.781	0.095	0.569	0.246	-0.21	1.09	0.194
102	cg13811240	0.319	0.135	0.521	0.216	0.20	1.09	0.185
103	cg09551984	0.773	0.114	0.566	0.234	-0.21	1.09	0.190
104	cg18621766	0.645	0.184	0.430	0.209	-0.22	1.08	0.198
105	cg03290223	0.748	0.110	0.547	0.229	-0.20	1.08	0.186
106	cg15446845	0.614	0.150	0.376	0.263	-0.24	1.08	0.220
107	cg18549952	0.521	0.183	0.722	0.189	0.20	1.08	0.186
108	cg16509509	0.505	0.188	0.297	0.198	-0.21	1.08	0.194
109	cg07178994	0.626	0.150	0.422	0.217	-0.20	1.08	0.190
110	cg03916694	0.473	0.214	0.678	0.169	0.20	1.07	0.191
111	cg23368579	0.603	0.137	0.377	0.255	-0.23	1.07	0.211
112	cg06915574	0.391	0.189	0.596	0.195	0.21	1.07	0.192
113	cg18780425	0.300	0.185	0.525	0.229	0.22	1.07	0.211
114	cg26324258	0.613	0.177	0.407	0.207	-0.21	1.06	0.194
115	cg17406248	0.477	0.216	0.711	0.226	0.23	1.06	0.221
116	cg27534912	0.689	0.162	0.483	0.219	-0.21	1.05	0.196
117	cg27334056	0.598	0.197	0.390	0.197	-0.21	1.05	0.197
118	cg15084543	0.237	0.137	0.439	0.228	0.20	1.04	0.193
119	cg09696091	0.627	0.173	0.425	0.211	-0.20	1.04	0.195
120	cg01252526	0.380	0.217	0.600	0.210	0.22	1.03	0.213
121	cg10780897	0.657	0.200	0.446	0.208	-0.21	1.03	0.205
122	cg01653189	0.678	0.153	0.473	0.230	-0.21	1.03	0.199
123	cg07340894	0.381	0.192	0.583	0.201	0.20	1.03	0.197
124	cg15858166	0.804	0.143	0.603	0.230	-0.20	1.03	0.196
125	cg04770195	0.525	0.224	0.743	0.204	0.22	1.02	0.213
126	cg24690071	0.250	0.129	0.456	0.244	0.21	1.02	0.201
127	cg14601733	0.527	0.187	0.325	0.207	-0.20	1.02	0.198
128	cg19839825	0.508	0.188	0.710	0.206	0.20	1.02	0.198
129	cg08264839	0.529	0.176	0.320	0.226	-0.21	1.02	0.206
130	cg07497602	0.673	0.167	0.461	0.237	-0.21	1.02	0.209
131	cg01586506	0.518	0.225	0.736	0.210	0.22	1.01	0.217
132	cg10776919	0.219	0.135	0.423	0.244	0.20	1.01	0.203
133	cg14096615	0.705	0.209	0.484	0.229	-0.22	1.00	0.220
134	cg14793063	0.386	0.188	0.612	0.252	0.23	1.00	0.226
135	cg25535351	0.794	0.120	0.594	0.249	-0.20	0.99	0.202
136	cg22151881	0.367	0.182	0.570	0.223	0.20	0.99	0.206

137	cg20876760	0.286	0.191	0.514	0.258	0.23	0.99	0.231
138	cg06593940	0.760	0.144	0.556	0.247	-0.20	0.98	0.207
139	cg19058865	0.323	0.168	0.528	0.235	0.20	0.98	0.208
140	cg14231073	0.424	0.229	0.626	0.190	0.20	0.97	0.208
141	cg04360793	0.314	0.178	0.525	0.244	0.21	0.97	0.217
142	cg05355757	0.699	0.190	0.494	0.228	-0.20	0.96	0.212
143	cg06437931	0.754	0.155	0.552	0.245	-0.20	0.96	0.210
144	cg15438734	0.543	0.270	0.765	0.196	0.22	0.96	0.232
145	cg15983026	0.344	0.209	0.548	0.217	0.20	0.96	0.213
146	cg21618157	0.351	0.194	0.554	0.227	0.20	0.95	0.213
147	cg05198244	0.673	0.181	0.471	0.239	-0.20	0.94	0.215
148	cg12341429	0.429	0.251	0.641	0.205	0.21	0.93	0.227
149	cg19530424	0.343	0.175	0.547	0.248	0.20	0.93	0.219
150	cg16573636	0.568	0.267	0.771	0.169	0.20	0.93	0.218
151	cg04871469	0.467	0.175	0.671	0.248	0.20	0.93	0.219
152	cg16596367	0.347	0.229	0.572	0.267	0.22	0.90	0.251
153	cg21851534	0.439	0.213	0.643	0.240	0.20	0.89	0.228
154	cg01431482	0.362	0.196	0.563	0.245	0.20	0.89	0.225
155	cg26404722	0.445	0.214	0.649	0.241	0.20	0.89	0.229
156	cg11448675	0.381	0.251	0.589	0.237	0.21	0.86	0.243

Supplementary table 1. Effect size of the CpGs of the 156-CpGs signature, from higher to less effect (done with the statistician of the group).



Supplementary figure 6. Heatmap with the TCGA data of stomach adenocarcinoma (A), lung adenocarcinoma (B), kidney renal clear cell carcinoma (C), sarcoma (D) and skin cutaneous melanoma (E). We cannot observe in any case a low risk or metastatic-specific cluster. Heatmaps created using the function heatmap.3, Ward.D2 and Manhattan.