

1 **Significant associations between driver gene mutations and**
2 **DNA methylation alterations across many cancer types**

3

4 Yun-Ching Chen¹, Valer Gotea¹, Gennady Margolin¹, Laura Elnitski^{1*}

5

6 ¹Genomic Functional Analysis Section, National Human Genome Research Institute,
7 National Institutes of Health, Bethesda, United States

8

9

10 *Corresponding author

11 E-mail: elnitski@mail.nih.gov (LE)

12

1 **Abstract**

2 Recent evidence shows that mutations in several driver genes can cause aberrant methylation
3 patterns, a hallmark of cancer. In light of these findings, we hypothesized that the landscapes of
4 tumor genomes and epigenomes are tightly interconnected. We measured this relationship using
5 principal component analyses and methylation-mutation associations applied at the nucleotide
6 level and with respect to genome-wide trends. We found a few mutated driver genes were
7 associated with genome-wide patterns of aberrant hypomethylation or CpG island
8 hypermethylation in specific cancer types. We identified associations between 737 mutated driver
9 genes and site-specific methylation changes. Moreover, using these mutation-methylation
10 associations, we were able to distinguish between two uterine and two thyroid cancer subtypes.
11 The driver gene mutation-associated methylation differences between the thyroid cancer subtypes
12 were linked to differential gene expression in JAK-STAT signaling, NADPH oxidation, and other
13 cancer-related pathways. These results establish that driver-gene mutations are associated with
14 methylation alterations capable of shaping regulatory network functions. In addition, the
15 methodology presented here can be used to subdivide tumors into more homogeneous subsets
16 corresponding to their underlying molecular characteristics, which could improve treatment
17 efficacy.

18

19 **Author summary**

20 Mutations that alter the function of driver genes by changing DNA nucleotides have been
21 recognized as a key player in cancer progression. Recent evidence showed that DNA methylation,
22 a molecular signature that is used for controlling gene expression and that consists of cytosine
23 residues with attached methyl groups in the context of CG dinucleotides, is also highly

1 dysregulated in cancer and contributes to carcinogenesis. However, whether those methylation
2 alterations correspond to mutated driver genes in cancer remains unclear. In this study, we
3 analyzed 4,302 tumors from 18 cancer types and demonstrated that driver gene mutations are
4 inherently connected with the aberrant DNA methylation landscape in cancer. We showed that
5 those driver gene-associated methylation patterns can classify heterogeneous tumors in a cancer
6 type into homogeneous subtypes and have the potential to influence the genes that contribute to
7 tumor growth. This finding could help us to better understand the fundamental connection between
8 driver gene mutations and DNA methylation alterations in cancer and to further improve the cancer
9 treatment.

10

11 **Introduction**

12 DNA methylation (DNAm) is highly dysregulated in cancers from many organs [1, 2] where it is
13 characterized by aberrant CpG island (CGI) hypermethylation and long-range blocks of
14 hypomethylation. Moreover, dysregulated DNAm at specific locations within the genome often
15 distinguishes heterogeneous tumors within cancer types into homogeneous subtypes [3, 4]. The
16 origin of these dramatic changes in the DNAm of tumor cells remains a puzzle. On the one hand,
17 DNAm alterations at particular CpG sites in tumors are associated with the aging process in normal
18 cells [5, 6]. This has led some researchers to propose that cell proliferation, which drives age-
19 associated DNAm errors in normal cells, is also responsible for aberrant DNAm in cancer [7]. On
20 the other hand, although these DNAm errors exhibit a linear association with the number of cell
21 divisions in normal cells, in cancer cells, they are not well correlated with the mRNA expression-
22 based mitotic index in many cancer types [7]. This suggests that in tumor cells, some factor other
23 than cell proliferation is shaping the DNAm landscape. Because tumors of the same molecular

1 subtype often harbor both dysregulated DNAm at particular locations in the genome and mutations
2 in driver genes [3, 8], we decided to investigate more broadly the connection between somatic
3 mutations and specific aberrant DNAm patterns.

4
5 Several genes that are known to play a defining role in a wide variety of cancer types [9, 10], such
6 as *TP53*, *IDH1*, *BRAF*, and *KRAS*, have been functionally linked to dysregulated DNA methylation.
7 For example, in hepatocellular carcinoma and esophageal squamous cell carcinoma, *TP53*
8 mutations are associated with extensive DNA hypomethylation [11, 12]. In glioblastoma, *IDH1*
9 mutations produce widespread CGI hypermethylation, termed the CpG island methylator
10 phenotype (CIMP), by inhibiting the TET-demethylation pathway [13, 14]. In colorectal cancer,
11 the BRAF V600E mutation results in DNA hypermethylation and CIMP development by
12 upregulating the transcriptional repressor MAFK, which recruits the DNA methyltransferase
13 DNMT3B to its targets at promoter CGIs [15]. Likewise, the KRAS G13D mutation upregulates
14 another transcriptional repressor, ZNF304, to establish a CIMP-intermediate pattern in colorectal
15 cancer [16].

16
17 Based on these findings, we hypothesized that the genomic and epigenetic landscapes were stable
18 and interdependent and therefore specific driver mutations would correlate with specific DNAm
19 patterns. Thus, in this study, we systematically evaluated mutation-methylation associations across
20 4,302 tumors from 18 cancer types, along with 727 normal tissue samples from The Cancer
21 Genome Atlas (TCGA). By investigating DNAm alterations associated with mutated driver genes
22 on both a genome-wide scale and a site-specific scale, we were able to show that i) mutated driver
23 genes are tightly associated with DNAm variation in cancer; ii) some associations are present

1 across cancer types, whereas others are cancer type-specific; iii) mutation-methylation associations
2 cannot be explained by cell proliferation alone; and iv) these associations can be used to classify
3 tumors into molecular subtypes and gain insight into functional alterations. Together, these results
4 establish that driver mutations and DNAm alterations are tightly coupled in tumor cells, and that
5 this coupling may affect important regulatory networks related to oncogenesis.

6

7 **Results**

8 **Association between driver gene mutations and methylation patterns in cancer**

9 To determine whether mutated driver genes were associated with methylation changes, first we
10 performed principal component analysis (PCA) on methylation data for each of 18 different cancer
11 types; within a given cancer type, tumor samples were projected onto the principal components
12 (PCs). Illumina Infinium human methylation 450K array data and somatic mutation data were
13 downloaded from TCGA (Table 1), and driver genes were predicted with MutSigCV [17] (see
14 Materials and Methods for details). For each cancer type, a driver gene was considered to be
15 associated with a PC if samples in which the gene was mutated were unevenly distributed toward
16 the positive or negative extremes of that PC ($q < 0.05$; two sided Wilcoxon rank-sum test). We
17 assessed each driver gene for the top five DNAm PCs; examples of PC1-associated driver genes
18 are shown in Fig 1A.

19

20 A PC-associated driver gene suggests that the mutated samples at one extreme display methylation
21 patterns distinct from the non-mutated samples at the other extreme. For instance, in stomach
22 adenocarcinoma (STAD), *TP53*-mutated samples were distributed toward the positive extreme of
23 PC1, whereas *ARID1A*- and *PIK3CA*-mutated samples were distributed toward the negative

1 extreme. Thus, distinct methylation patterns associated with PC1 separated the majority of *TP53*-
2 mutated STAD samples from *ARID1A*- and *PIK3CA*-mutated samples: *ARID1A*- and *PIK3CA*-
3 mutated samples were highly methylated at PC1-defining probes relative to *TP53*-mutated samples
4 (Fig 1B). Overall, we found a significant association between 159 driver genes and one or more
5 of the top five methylation PCs in 15 of 18 cancer types (the top three driver genes associated with
6 each PC shown in Fig 1C).

7
8 Mathematically, distinct PCs represent mutually orthogonal (uncorrelated) linear combinations of
9 probes, or different methylation patterns (such as the pattern associated with PC1 in STAD shown
10 in Fig 1B), with several top PCs usually capturing the majority of variance in the methylome. Thus,
11 frequent driver gene-PC associations in almost every cancer type suggest a tight connection
12 between driver-gene mutations and DNA methylation alterations in cancer.

13
14 Next, we investigated whether the mutation-methylation connection in cancer was limited to
15 certain CpG subsets—namely those occurring in CGIs, shores and shelves (SSs; the 4-kb regions
16 flanking the CGIs), or open-sea regions (i.e., outside of CGIs and SSs)—as the regulatory
17 functions of these CpG subsets often differ [18]. For example, DNAm in promoter CGIs often
18 causes gene silencing whereas DNAm in CGI shores is frequently altered and strongly correlated
19 to corresponding gene expression in cancer [19]. Thus, we repeated the analysis described above
20 for each subset of probes. We observed similar driver gene-PC associations across multiple cancer
21 types, indicating that the mutation-methylation connection is not limited to a particular CpG subset
22 (S1 Fig). In total, 14 of 18 cancer types harbored significant associations between driver gene
23 mutations and top five methylation PCs at CGIs, 14 of 18 at SSs, and 15 of 18 in open sea regions.

1 We then repeated the same analysis after stratifying probes by hypo- or hypermethylation status
2 and found that the results did not vary appreciably (S1 Fig). Of note, in this study we identified
3 hyper- and hypomethylated probes by comparing the methylation of tumor and control samples
4 ($q < 0.05$; Wilcoxon rank sum test).

6 **Driver gene mutations, methylation patterns, and cell proliferation are correlated**

7 Cell proliferation was proposed to be an important factor driving DNAm changes in cancer [7].
8 This was demonstrated with a set of 385 CpGs in the Illumina 450K array whose average
9 methylation levels approximated the mitotic rate in cancer, called epiTOC (epigenetic Timer Of
10 Cancer). The method has been proposed as an alternative to the mRNA expression-based mitotic
11 index [7]. We calculated the mitotic index for all tumor types using epiTOC or the mRNA
12 expression-based approach. We found these two indices were correlated with many of the top 5
13 PCs across cancer types ($q < 0.05$; Spearman correlation) (Fig 1C). In general, epiTOC correlated
14 with more PCs than the mRNA expression-based index. For example, STAD showed a significant
15 correlation between PC1 and epiTOC scores ($q = 0$) but not the mRNA expression-based index
16 ($q = 0.22$) (Fig 1B). We removed epiTOC-correlated methylation sites ($q < 0.05$; Pearson
17 correlation), and found that many methylation PC associated-driver genes remained in many
18 cancer types (S2 Fig), indicating that the mutation-methylation association cannot be totally
19 explained by epiTOC scores. We also used both methods to compare predicted cell proliferation
20 rates. To elucidate the relationship between cell proliferation rate and mutated driver genes, we
21 associated the presence of mutations in each driver gene with high/low score in each index. We
22 found that mutated driver genes associated with high and low cell proliferation rates estimated by
23 the two indices were inconsistent in most cancer types (Table 2). For example, *TP53*-mutated

1 tumors were correlated with a high cell proliferation rate in 10 cancer types and none were
2 correlated with low cell proliferation rate using the mRNA expression-based index. In contrast,
3 *TP53*-mutated tumors were correlated with low cell proliferation in HNSC and UCEC, using
4 *epiTOC*. Thus DNAm and mRNA-based mitotic indices are inconsistent and could be measuring
5 two different properties of these tumor cells. Despite their inconsistency, by using either index,
6 our results indicate that methylome patterns in cancer are correlated with the presence of mutated
7 driver genes and cell proliferation but the mutation-methylation association found in this study
8 cannot be explained by cell proliferation alone.

9
10 **Driver gene mutations, genome-wide CGI hypermethylation, and open sea hypomethylation**
11 **in tumors**

12 We next asked whether driver gene-associated methylation alterations corresponded to genome-
13 wide methylation patterns characteristic of cancer: i.e., widespread CGI hypermethylation and
14 huge hypomethylated blocks, primarily in open sea regions [1, 2, 20]. To answer this question, we
15 calculated the HyperZ and HypoZ indices for each sample [21]. A high HyperZ index indicates
16 aberrant hypermethylation in many CGIs for a given sample, whereas a high HypoZ index
17 indicates extensive open sea hypomethylation. The number of mutated driver genes that were
18 significantly associated with either high or low HyperZ and/or HypoZ indices is shown for all 18
19 cancer types in Table 3 ($q < 0.05$; Wilcoxon rank sum test); it varies from 67 driver genes in COAD
20 to 0 in PAAD, READ, and SKCM.

21
22 The complete list of mutated driver genes with significant associations to HyperZ and HypoZ is
23 shown in Table S1. Some known players appear on this list. For example, a high HyperZ index

1 was associated with *BRAF* in COAD and *IDH1* in GBM; both genes are linked to CIMP in cancer
2 [13, 22]. *NSD1*, which was associated with a high HypoZ index in HNSC, has also been linked to
3 hypomethylation in cancer [23]. The associations we detected in most cancer types underscore the
4 relationship between driver gene mutations and genome-wide methylation alterations commonly
5 observed in cancer, whereas only a few types lack these patterns.

6

7 **Driver gene mutations and site-specific methylation changes in tumors**

8 Next, we investigated whether the connection between driver gene mutations and methylation
9 alterations was methylation-site specific, in each cancer type. To do so, we calculated the
10 associations between every driver gene and every methylation array probe for all 18 cancer types,
11 testing whether the presence of mutations in a driver gene is associated with the high or low
12 methylation level at a given probe site. Across almost all cancer types, many more driver genes
13 were significantly associated with at least one probe than with the HyperZ and/or HypoZ indices,
14 after correcting for multiple testing ($q < 0.05$; Wilcoxon rank-sum test; Table 3). In total, 737
15 unique driver genes were implicated. The driver gene-methylation site associations were present
16 genome-wide. An example of the chromosomal distribution of driver gene-associated methylation
17 probes present in kidney renal clear cell carcinoma (KIRC) is shown in Fig S3. The numerous
18 gene-probe associations detected here suggest that driver gene-associated methylation changes
19 likely occur at certain CpG sites, potentially resulting from a site-targeting mechanism.

20

21 The number of probes associated with each driver gene varied greatly, ranging from fewer than 10
22 to tens of thousands (S2 Table). For each cancer type, a few (1–5) dominant driver genes accounted
23 for the majority of associations (Fig 2A), including known oncogenes and tumor suppressor genes

1 such as *TP53*, *PTEN*, and *PIK3CA*, and known CIMP-driving genes such as *BRAF*, *IDHI*, and
2 *KRAS* [15, 16, 24]. Dominant driver genes usually displayed both positive and negative
3 associations with methylation levels in a given cancer type (Fig 2B). However, there was typically
4 more of one type of association than the other. By definition, positive associations indicate higher
5 methylation levels in the presence of driver gene mutations among tumor samples, whereas
6 negative associations indicate lower methylation levels. Thus, positive associations would
7 correspond to hypermethylation (primarily in CGIs) if control samples displayed the low
8 methylation level at a given probe site, whereas negative associations would correspond to
9 hypomethylation (primarily in open-sea regions) if control samples displayed the high methylation
10 level. We found both positive and negative associations for many driver genes; however they can
11 be largely imbalanced. For example primarily positive associations occurred for the CIMP-driving
12 genes *BRAF* in COAD (5,166 positive vs. 103 negative associations) and *IDHI* in GBM (31,877
13 vs. 3,535). Other genes with predominantly positive associations included *RNF43* (4,784 vs. 173)
14 and *MACF1* (2,706 vs. 95) in COAD; *CASP8* (21,552 vs. 4,660) in HNSC; *PBRM1* (11,831 vs.
15 3,906), *SETD2* (6,443 vs. 2,440), and *BAP1* (5,162 vs. 1,240) in KIRC; *SETD2* (4,300 vs. 936) in
16 KIRP; *SPOP* (8,919 vs. 6,444) in PRAD; *PIK3CA* (58,843 vs. 2,153) in STAD; and *PTEN* (20,548
17 vs. 14,496) in UCEC. These genes were also associated with a high HyperZ index, suggesting that
18 they may play a role in genome-wide CGI hypermethylation in particular cancer types (S1 Table).
19
20 By contrast, genes that primarily displayed negative associations were often associated with a high
21 HypoZ index, suggesting that they may play a role in genome-wide open sea hypomethylation.
22 Examples were *NSD1* (1,423 positive vs. 72,475 negative associations) in HNSC, *CTNNB1* (1,065
23 vs. 23,869) in LIHC, and *STK11* (3,269 vs. 33,840) and *KEAP1* (4,122 vs. 31,348) in LUAD.

1 Interestingly, *RNF43* (31,905 vs. 23,267) was associated with both high HyperZ and HypoZ
2 indices in STAD, suggesting a dual role in genome-wide CGI hypermethylation and open sea
3 hypomethylation.

4
5 In short, a few driver genes were linked to genome-wide patterns of CGI hypermethylation and
6 open sea hypomethylation in particular cancer types, whereas many more driver genes were linked
7 to only a few probe sites that exhibited hyper- and hypomethylation in cancer (S2 Table).

8
9 **Consistency of mutated driver gene-associated methylation changes across cancer types**

10 Next, we asked whether mutated driver genes consistently displayed predominantly
11 positive/negative associations across multiple cancer types. We investigated the proportion of
12 positive and negative probe associations for 17 driver genes across 18 cancer types (Fig 3A).
13 Here, we only considered driver genes linked to extensive methylation changes (more than 1,000
14 probe associations per driver gene, for at least two cancer types). When compared to control
15 samples, positive associations often equated to hypermethylation (primarily in CGIs) in response
16 to mutations. Likewise, the negative associations often equated to hypomethylation (primarily in
17 open sea regions) (Fig 3B). *TP53* displayed predominantly negative associations in 9 of 18
18 cancer types, and no predominantly positive associations were observed in connection to this
19 gene in any cancer type, suggesting a tight connection between mutations in this gene and open
20 sea hypomethylation across multiple cancer types. Interestingly, those negatively associated
21 probes were shared across cancer types (from 62,951 probes shared across 2 types to 15 probes
22 shared across 7 types) (S4 Fig). In addition to *TP53*, *APC* and *CTNNB1* also displayed
23 predominantly negative associations in two different cancer types each.

1
2 By contrast, *IDHI* strongly favored positive associations in two cancer types, GBM and SKCM,
3 consistent with reports that mutated *IDHI* downregulates TET-dependent demethylation,
4 resulting in aberrant CGI hypermethylation [25]. *SETD2*, *PTEN*, *RNF43*, *HRAS*, *EPHA2*, and
5 *BAP1* were also linked to primarily positive associations in more than one cancer type,
6 suggesting that they may play a general role in CGI hypermethylation.

7
8 *BRAF*, which mediates CIMP in colorectal cancer, displayed a high proportion of positive
9 associations (0.98) in COAD, but low proportions in SKCM (0.02) and THCA (0.27). Its
10 negative associations in THCA corresponded to hypomethylation across CGIs, SSs, and open sea
11 regions (Fig 3B and S2 Table). This dramatic difference indicates that driver genes may be
12 associated with methylation patterns in a cancer type-specific manner. Such cancer type-specific
13 associations were also seen for *PIK3CA*, *KRAS*, *GATA3*, *CTCF*, *CDHI*, and *ARID1A*.

14 15 **Identification of molecular subtypes in thyroid carcinoma and uterine corpus endometrial** 16 **carcinoma based on driver gene-associated methylation patterns**

17 In previous studies, researchers have identified cancer subtypes in COAD and GBM by matching
18 mutational profiles to methylation patterns [13, 22]; here, we asked if the site-specific mutation-
19 methylation associations would separate THCA and UCEC tumors into subtypes. We focused on
20 the methylation patterns associated with the top three dominant driver genes which account for
21 the most probe associations in each cancer type (as shown in Fig 2A). In THCA, the top three
22 genes (*NRAS*, *HRAS*, or *BRAF*) were mutated in a mutually exclusive fashion. However, in

1 UCEC, mutations in *TP53* were nearly mutually exclusive with *PTEN* and *CTNNB1* mutations,
2 which co-occurred in many tumor samples.

3
4 We performed hierarchical clustering on the union of the 500 methylation probes most
5 significantly associated with mutations in each of the top three genes (Fig 4). In THCA, two
6 methylation subtypes emerged corresponding to *NRAS*- and *HRAS*-mutated tumors vs. *BRAF*-
7 mutated tumors (Fig 4A). Furthermore, we found tumors lacking any of the specified mutations
8 co-clustered within these patterns. The selected probes primarily fell in non-CGI positions (i.e.
9 SS and open sea regions); *BRAF* mutants displayed hypomethylation in open sea and some SS
10 regions, whereas *NRAS* and *HRAS* mutants displayed methylation levels similar to controls in
11 open sea and SSs regions, with little hypermethylation. Two methylation subtypes were also
12 identified in UCEC, this time corresponding to *TP53* vs. *PTEN* mutations (Fig 4B). *PTEN*-
13 mutated samples generally exhibited CGI hypermethylation, whereas *TP53*-mutated samples
14 generally exhibited normal methylation levels, with some hypomethylation in open sea regions.
15 Most UCEC samples with mutations in both *PTEN* and *CTNNB1* displayed greater levels of open
16 sea hypomethylation than samples with *PTEN* mutations alone, which has not been previously
17 reported, illustrating the connectivity between mutational profiles and DNA methylation in
18 cancer.

19

20 **Correlations between driver gene-associated methylation probe sites and corresponding** 21 **gene expression in two thyroid carcinoma subtypes**

22 Finally, we investigated whether driver gene-associated methylation patterns shape gene
23 regulatory networks. To investigate the three-way association between mutation, methylation,

1 and gene expression, we used THCA as our primary example; the mutually exclusive mutation
2 profile present in this type of cancer minimized the complexity of the associated methylation
3 patterns, facilitating the study of gene expression. We looked for genes whose aberrant
4 expression levels were correlated with aberrant methylation levels (each relative to controls). We
5 focused on CpG sites in promoter regions and gene bodies in *NRAS* and *HRAS* mutants (the
6 *NRAS-HRAS* group) vs. *BRAF* mutants (the *BRAF* group). These genes were subsequently sorted
7 into four different categories of expression change: (1) upregulation only in the *BRAF* group, (2)
8 upregulation only in the *NRAS-HRAS* group, (3) downregulation only in the *BRAF* group, and (4)
9 downregulation only in the *NRAS-HRAS* group. For each category, we reported the affected
10 genes, sorted by median difference in expression between mutants and controls (S3 Table), and
11 significantly enriched pathways (S4 Table).

12
13 In all, 1,565 genes were upregulated specifically in the *BRAF* group (S3 Table). Gene set
14 analysis showed that these genes were enriched in 97 canonical pathways primarily pertinent to
15 cell-cell communication/extracellular matrix gene sets and signaling pathways ($q < 0.05$;
16 hypergeometric test; S4 Table). Some upregulated genes were involved in many of the 97
17 pathways; for example, the 10 genes implicated in the most pathways were: *GRB2* (present in 34
18 out of 97 pathways), *RAC1* ($n=26$), *STAT1* ($n=25$), *LYN* ($n=24$), *VAV1* ($n=24$), *JAK1* ($n=22$),
19 *PTPN6* ($n=22$), *ITGB1* ($n=20$), *STAT3* ($n=18$), and *STAT5* ($n=18$). We noticed that genes in JAK
20 and STAT gene families (e.g. *STAT1*, *JAK1*, *STAT3*, and *STAT5*) were differentially regulated
21 between two THCA subtypes and implicated in many signaling cascades, including the KEGG
22 JAK-STAT signaling pathway (ranked 12th out of the 97 pathways; $q=4.7E-5$). In addition, *JAK3*
23 ($n=11$) and *STAT4* ($n=4$), were also upregulated and present in multiple pathways. Our results

1 show that this differential regulation of the JAK and STAT families may be shaped by
2 differences in DNA methylation (Fig 5A). Specifically, *STAT1* differential expression is
3 negatively correlated with methylation levels at the SS probe of the promoter CGI, whereas
4 *JAK3* differential expression is positively correlated with methylation levels at the 3' gene body
5 CGI and its north shore (Figs 5B and 5C). In addition, 9 of the top 15 differentially methylated
6 genes were involved in metastasis: *KLK6*, *KLK7*, *KLK11* [27], *CLDN10* [28], *B3GNT3* [29],
7 *RASGRF1* [30], *ST6GALNAC5* [31], *TACSTD2* (also known as *TROP2*) [32], and *CEACAM6*
8 [33] (S3 Table).

9
10 By contrast, 1,043 differentially methylated genes were downregulated specifically in the *BRAF*
11 group. Gene set analysis showed that these genes were enriched in five canonical pathways
12 pertinent to amino acid catabolism, triglyceride biosynthesis, glycerophospholipid metabolism,
13 and nucleic acid metabolism (S4 Table).

14
15 The *NRAS-HRAS* group displayed 278 differentially methylated, upregulated genes. These genes
16 were enriched in three canonical pathways, relevant to the neuronal system, potassium channels,
17 and melanogenesis (S4 Table). We did not find a significant portion of differentially methylated
18 genes implicated in tumor progression among the top differentially expressed genes (defined by
19 median difference in expression between mutants and controls). However when we considered
20 differentially methylated genes within the top 17 that were highly transcribed, 6 out of 17 were
21 implicated in tumorigenesis: G protein alpha subunit (*GNAS*) [34], pyruvate dehydrogenase
22 kinase 4 (*PDK4*) [35], NADPH reductase (*NQO1*) [36], and three NADPH oxidases that produce
23 H₂O₂ for thyroid hormone synthesis: *DUOX2*, *DUOXA2*, and *DUOX1* [37] (S3 Table).

1
2 By contrast, 447 differentially methylated genes were downregulated specifically in the *NRAS*-
3 *HRAS* group. These genes were enriched in 166 canonical pathways; interestingly, 47 genes
4 overlapped the 97 pathways enriched in the *BRAF* upregulated group, including the JAK-STAT
5 signaling pathway (ranked 13th, $q = 3.5E-5$). Specifically, *STAT1*, *STAT3*, *STAT4*, and *JAK3*
6 were among the genes upregulated in the *BRAF* group but downregulated in the *NRAS-HRAS*
7 group (Fig 5A and S3 Table). This result demonstrates that methylation changes are indeed
8 associated with differential gene expression between *BRAF*-mutated and *NRAS*- and *HRAS*-
9 mutated samples in THCA.

10

11 **Discussion**

12 In this study, we demonstrated that driver gene mutations are tightly tied to the DNAm landscape
13 in multiple types of cancer. Furthermore, we showed that mutated driver genes are associated with
14 DNAm alterations in a reproducible, site-specific manner. In each cancer type, a few driver genes
15 dominate the site-specific associations, and some potentially contribute to CGI hypermethylation
16 and extensive hypomethylation, i.e., the hallmarks of cancer. However, we caution that these
17 findings do not equate to causality, but point to the highly interconnected nature of the genome
18 and epigenome.

19

20 Our findings are consistent with previous research on methylation in cancer; however, they also
21 contribute novel insights. For example, several driver genes that display primarily positive or
22 negative methylation probe associations in this study have been linked to CGI hypermethylation
23 or open sea hypomethylation, respectively. Driver genes associated with CGI hypermethylation in

1 both this study and past studies include *BRAF* in COAD [22], *IDHI* in GBM [13], *SETD2* in KIRC
2 [38], *PIK3CA* in STAD [3], and *PTEN* in UCEC [39], whereas genes associated with
3 hypomethylation include *TP53* in LIHC [12] and *NSD1* in HNSC [23]. In addition to these
4 examples, we identified novel driver genes that may contribute to CGI hypermethylation, such as
5 *CASP8* in HNSC and *SPOP* in PRAD, or to open sea hypomethylation, such as *CTNNB1* in LIHC
6 and *BRAF* in THCA. By illuminating the driver genes associated with widespread DNAm
7 alterations, as well as driver genes associated with more limited DNAm alterations, our
8 comprehensive analysis provides a detailed mutation-methylation map for many types of cancer.
9
10 Several mutated driver genes displayed consistent and widespread positive or negative associations
11 across cancers corresponding to extensive DNAm alterations, whereas others showed effects that
12 varied by cancer type. This discrepancy may be attributable to different underlying mechanisms.
13 For example, mutations in *IDHI* and *SETD2* directly affect the epigenetic landscape by inhibiting
14 TET-dependent demethylation and disturbing DNA methyltransferase targeting, respectively [14,
15 25, 38]. Both mechanisms cause DNA hypermethylation, in line with the corresponding primarily
16 positive associations observed in this study. By contrast, several documented mechanisms link
17 *TP53* loss to global hypomethylation, consistent with *TP53* mutations and hypomethylation in this
18 study [11, 12, 40, 41]. Moreover, probes negatively associated with *TP53* were shared across
19 cancer types, suggesting that similar CpG sites may be targeted in *TP53*-mutated tumors
20 independent of tissue types. In contrast to patterns of consistency across cancer types, *BRAF*
21 mutations displayed inconsistent methylation patterns between types. For example, in COAD,
22 BRAF V600E recruits the DNA methyltransferase to CGIs targets by stimulating the MEK/ERK
23 signaling pathway and upregulating the transcription repressor MAFG [15]. This is in line with

1 *BRAF*-associated, widespread CGI hypermethylation in COAD observed in this study. However,
2 in THCA, *BRAF*-mutated samples (260/266 of which harbored the V600E mutation) largely
3 displayed hypomethylation. Although no mechanistic explanation is yet available, it could be
4 either that the mutation does not upregulate MAFG in THCA or that MAFG is upregulated in both
5 types resulting CGI hypermethylation at a few binding sites but some other factor is responsible
6 for the extensive CGI hypermethylation in COAD.

7
8 It is unclear what links site-specific methylation alterations to driver gene mutations. A model of
9 DNAm targeting proposed by Struhl [42] provides one potential mechanistic explanation. In this
10 model, DNA methylation alterations are positioned and maintained by transcriptional circuitry that
11 is aberrantly regulated as the result of driver gene mutations. This model was supported by the
12 recent finding that *BRAF* V600E and *KRAS* G13D mutations in COAD upregulate the
13 transcription factors MAFG and ZNF304, respectively, resulting in targeted promoter CGI
14 hypermethylation near TF binding sites [15, 16]. However, in COAD, the *BRAF*-associated
15 widespread CGI hypermethylation is not restricted to the published TF binding sites, making it
16 unclear whether this model can be generalized to explain all driver gene-probe associations in
17 COAD. On the other hand, several biological processes that can alter DNAm at specific sites have
18 been documented recently. For example, cellular oxidative stress can produce hypermethylation
19 in promoters of low-expression genes [43]. Hypoxia can reduce TET activity leading to
20 hypermethylation at corresponding sites [44]. Cell proliferation accumulates aberrant DNAm in
21 promoters of polycomb group target CpGs (pcgt) [7]. In this study, we showed that driver gene
22 mutations correlate with the pcgt-derived mitotic index in many cancer types. The same connection

1 between driver gene mutations and DNAm-altered sites may be established via other DNAm-
2 altering processes as well.

3
4 Because mutation-methylation patterns are likely to reflect important oncogenic characteristics,
5 using these patterns to separate tumors into molecular subtypes could potentially aid treatment
6 selection. In a proof of concept portion of this study, we successfully identified molecular subtypes
7 in THCA and UCEC based on the dominant driver gene-associated methylation patterns. These
8 subtypes agreed with previous reports of subtypes defined by gene expression analyses [45, 46].
9 Though we only attempted to identify subtypes in two cancer types, these results indicate that our
10 mutation-methylation-based approach could be useful for identifying molecular subtypes in other
11 cancer types as well.

12
13 The mutually exclusive mutations in *NRAS*, *HRAS*, and *BRAF* in THCA tumors have been
14 interpreted to mean that these mutations have interchangeable effects on activating MAPK
15 signaling, the main cancer-driving event in papillary thyroid carcinomas [46]. Our analysis,
16 however, highlights substantial differences in DNAm between *BRAF*-mutated vs. *NRAS*- and
17 *HRAS*-mutated THCA tumors; moreover, the differences in DNAm appear to profoundly shape
18 gene expression profiles, which may contribute to thyroid tumorigenesis. In this study, DNAm
19 alterations in *BRAF*-mutated tumors were correlated with the activation of JAK-STAT signaling
20 and metastasis, whereas DNAm alterations in *NRAS*- and *HRAS*-mutated tumors were linked to
21 potential DNA damage by H₂O₂ overproduction [37], activation of G-protein signaling by *GNAS*
22 overexpression [34, 47], and activation of mTOR signaling by *PDK4* overexpression [46]. These

1 differences in the molecular processes linked to different driver gene mutations may contribute to
2 distinct pathways of tumorigenesis, yielding different prognoses and clinical phenotypes.
3
4 This work has several limitations. First, all samples with mutations in the same gene were
5 classified together. For example, although the majority of *BRAF*-mutated samples carried the
6 V600E mutation (25 out of 34 *BRAF*-mutated tumors with BRAF V600E in COAD, 167/195 in
7 SKCM, and 260/266 in THCA), this group also included a few non-V600E mutations. We took
8 this approach because MutSigCV only reports driver genes but not mutations. However, different
9 mutations in the same gene may be linked to different methylation patterns, introducing noise
10 into our analysis and lowering our statistical power to detect mutation-methylation associations.
11 Second, we examined the individual associations between driver genes and methylation sites.
12 However, the combinatorial effect of driver gene mutations on methylation could exist, since
13 several driver gene mutations can typically co-occur in a given tumor. Third, we focused only on
14 MutSigCV-reported driver genes and were limited to the information present in TCGA data.
15 Although MutSigCV is one of the most reliable driver gene detection tools available, limitations
16 associated with the detection algorithm—paired with the limitations imposed by the number of
17 tumor samples available in TCGA—may have led us to miss methylation-altering mutations that
18 occurred in unknown driver genes. Finally, tumor purity is a potential confounding factor in
19 analyses of cancer data. Although we were not able to exclude the possibility of confounding, the
20 mutation-methylation associations reported here were seen in cancer types where most TCGA
21 samples (>80%) were predicted with high purity (>70%), including GBM, KIRP, THCA, and
22 UCEC [48]. Therefore, it seems unlikely that confounding by tumor purity level was extensive.
23

1 **Conclusions**

2 This pan-cancer analysis provides the strongest evidence to date for a widespread connection
3 between genomic and epigenomic alterations in cancer. The mutation-methylation relationships
4 described here could potentially be used to identify tumor subtypes, thus aiding in prognosis and
5 treatment decisions. In addition, in the future, further analysis of methylation and expression data
6 may identify driver gene mutation-induced methylation alterations that dysregulate
7 genes/pathways that promote tumor growth. Importantly, such dysregulation could potentially be
8 corrected by treating patients with agents that influence the DNA methylation landscape.
9 Demethylating agents such as 5-aza-2'-deoxycytidine, for example, have been used to reactivate
10 epigenetically silenced tumor suppressor genes and also to decrease overexpression of oncogenes
11 [49, 50]. By contrast, the methyl donor S-adenosylmethionine has been shown to downregulate
12 the oncogenes *c-MYC* and *HRAS*, inhibiting cancer cell growth [51]. In summary, in light of the
13 connection between driver gene mutations and DNA methylation shown here, it will be
14 important to further study how coordinated genomic and epigenomic alterations result in the
15 hallmarks of cancer. A better understanding of the molecular mechanisms underlying cancer may
16 help us identify factors that accelerate tumor onset, predict biomarkers for early diagnosis, and
17 assess new therapeutic targets.

18

19 **Materials and Methods**

20 We analyzed samples that had both somatic mutation data and DNA methylation data available.
21 These samples represented 18 cancer types: bladder urothelial carcinoma (BLCA), breast
22 invasive carcinoma (BRCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM),
23 head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC),

1 kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung
2 adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma
3 (PAAD), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), skin cutaneous
4 melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT),
5 thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). The number of
6 tumors and controls (normal tissue samples adjacent to tumors or healthy donors) for each cancer
7 type is listed in Table 1.

8

9 **Data preprocessing**

10 Exome-sequenced level 2 somatic mutation data were downloaded from TCGA's data portal
11 (<https://tcga-data.nci.nih.gov/tcga/>) on February 1, 2015.

12

13 TCGA level 3 DNA methylation array-based data (Illumina Infinium HumanMethylation450
14 BeadChip array) were downloaded from the UCSC Cancer Genomics Browser ([https://genome-](https://genome-cancer.ucsc.edu)
15 [cancer.ucsc.edu](https://genome-cancer.ucsc.edu)) on October 26, 2015. DNA methylation levels were measured with β values.

16 We normalized β values for type I and II probes using the β mixture quantile (BMIQ) method
17 [52]. The following types of probes were removed from the analysis: (1) probes on the X and Y
18 chromosomes, (2) cross-reactive probes [53], (3) probes near single nucleotide polymorphisms
19 (SNPs), and (4) probes with missing rates $\geq 90\%$ across all samples for a given cancer type. A
20 final set of 314,421 probes was analyzed for each cancer type.

21

1 Finally, TCGA level-3 gene expression data measured by log transformed (base 2) RSEM-
2 normalized RNA-Seq (Illumina HiSeq) counts were downloaded from the UCSC Cancer
3 Genomics Browser (<https://genome-cancer.ucsc.edu>) on November 4, 2015.

4 5 **Driver genes**

6 We defined driver genes as those reported by MutSigCV2 [17] for each of the 18 cancer types;
7 these data were downloaded from the Broad Institute of MIT & Harvard (<http://firebrowse.org>).
8 Specifically, we analyzed genes that had reported q-values < 0.05 and that were mutated in at
9 least five samples from each cancer type. The number of driver genes for the 18 cancer types is
10 summarized in Table 1.

11 12 **Hyper- and hypomethylated probes**

13 For each probe, we compared the distribution of methylation levels among tumor samples with
14 that among control samples using one-sided Wilcoxon rank sum tests, one for each direction,
15 stratified by cancer type. Each significant probe ($q < 0.05$) was classified as either
16 hypermethylated (methylation levels in tumor samples were greater than in control samples) or
17 hypomethylated (methylation levels in tumor samples were less than in control samples).

18 19 **Principal component analysis**

20 PCA was performed using the R package ‘pcaMethods,’ and missing values were imputed by
21 probabilistic PCA. The top five PCs were computed within each cancer type for all probes and
22 subsets of probes, including probes in CGIs, the SSs around CGIs (the 4-kb regions flanking
23 CGIs), and open sea regions (CpGs outside CGIs and SSs). The probe sets (CGIs, SSs, and open

1 sea regions) were further stratified by hyper- and hypomethylation status. CGIs, SSs, and open
2 sea regions were defined in the Illumina 450K array annotation file.

3

4 **HyperZ and HypoZ indices**

5 HyperZ and HypoZ indices were computed for each tumor sample within a cancer type. The
6 HyperZ and HypoZ indices were introduced by Yang et al. [21] to measure the level of overall
7 CGI hypermethylation and open sea hypomethylation, respectively.

8

9 **Association between driver gene mutations and DNA methylation**

10 We analyzed somatic mutations at the gene level. A driver gene was classified as either mutated
11 (any mutations) or not mutated (no mutations) for each tumor sample. Associations between
12 driver gene mutations and methylation were tested using the two-sided Wilcoxon rank sum test.
13 To evaluate driver gene-PC associations, the test was performed for every driver gene and each
14 of the top five methylation PCs; samples were ranked based on their coordinates on the PC, and
15 the mutated cohort was compared with the non-mutated cohort. To evaluate site-specific
16 associations, the test was performed for every possible driver gene-probe pair. Here, samples
17 were ranked based on their β values at the probe, and the mutated cohort was compared with the
18 non-mutated cohort. Finally, we performed the same association test for every driver gene and
19 the HyperZ and HypoZ indices, to identify driver genes potentially associated with genome-wide
20 CGI hypermethylation and open sea hypomethylation. In each case, q-values were computed by
21 correcting for all tests performed for a given cancer type [54]. Associations were considered
22 significant at $q < 0.05$. For associations between every gene-probe pair, the empirical false
23 discovery rate was also estimated by permuting the mutation status for every driver gene. The

1 results showed that the empirical false discovery rate was controlled (< 0.05) at the theoretical
2 cutoff ($q < 0.05$) for each cancer type (S5 Fig).

3

4 **Correlations between methylation PCs, mutated driver genes, and mitotic indices**

5 A DNAm-based mitotic index called epiTOC (epigenetic Timer Of Cancer) was proposed by
6 Yang et al. to estimate tumor cell proliferation rates. Performance was validated by correlating
7 with a mRNA expression-based mitotic index [7]. epiTOC is the average methylation level
8 across 385 probes whereas the expression-based index is the average expression level across 9
9 genes: *CDKN3*, *ILF2*, *KDEL2*, *RFC4*, *TOP2A*, *MCM3*, *KPNA2*, *CKS2*, and *CDC2*. For each
10 index, we computed its Spearman correlation with each of the top 5 PCs for each cancer type and
11 called statistical significance at $q < 0.05$. To remove epiTOC-correlated probes, we removed any
12 probes correlated with epiTOC ($q < 0.05$; Pearson correlation) and recomputed methylation PC-
13 driver gene associations for each cancer type. Finally, we used Wilcoxon rank sum test to test the
14 associations between mutated driver genes and each of the two indices and called statistical
15 significance at $q < 0.05$ for each cancer type.

16

17 **Differential expression analysis of thyroid carcinoma molecular subtypes**

18 First, we visually identified two THCA molecular subtypes based on driver gene mutations and
19 DNA methylation patterns (Fig 4A): (1) *BRAF*-mutated tumors (the *BRAF* group) and (2) *NRAS*-
20 and *HRAS*-mutated tumors (the *NRAS-HRAS* group). Next, we assembled four differentially
21 expressed gene sets (up- or downregulated in one group but identical or regulated in the opposite
22 direction in another group, compared with controls). The search was restricted to genes whose
23 aberrant expression levels coincided with hyper- or hypomethylated probes associated with

1 *BRAF*, *HRAS*, and *NRAS* mutations (see section below). Using a hypergeometric model, genes in
2 each of the four sets were tested for enrichment by 1,330 canonical pathways collected by
3 MSigDB [55]. The highly transcribed, differential genes were defined by expression levels that
4 are greater than 10 (in \log_2 RSEM) and at least double those of controls.

5 6 **Association between driver gene-associated aberrant methylation and aberrant gene** 7 **expression in thyroid carcinomas**

8 We sought genes whose aberrant expression was correlated with aberrant methylation in the
9 presence of *BRAF*, *HRAS*, or *NRAS* mutations. First, we identified probes that fell within 1,500
10 bp of the transcription start site (TSS) or within gene bodies, and whose β values were
11 significantly correlated with expression levels of the corresponding gene. Because a methylated
12 CpG may up- or downregulate gene expression in a context-dependent manner [50, 56, 57], we
13 computed the significance of the Spearman correlation between each individual probe's
14 methylation and the expression level of its corresponding gene. We considered gene-probe pairs
15 significant at $q < 0.05$. Second, we analyzed the association between β values and *BRAF*, *HRAS*,
16 and *NRAS* mutation status for all probes using the two-sided Wilcoxon rank sum test. For each
17 driver gene, aberrantly methylated probes ($q < 0.05$) were classified as hyper- or hypomethylated
18 relative to control samples. Third, we integrated the results from the first and second steps to
19 identify aberrantly methylated probes whose methylation levels were significantly correlated
20 with the expression levels of their corresponding genes for each group of *BRAF*-, *HRAS*-, and
21 *NRAS*- mutated samples. Finally, for each group of mutated samples, significantly up- or
22 downregulated genes were identified from the aberrant methylation-matched genes identified in
23 the third step using the two-sided Wilcoxon rank sum test relative to the expression levels of

1 control samples ($q < 0.05$). For example, when we looked for genes that were upregulated in
2 *BRAF*-mutated samples but exhibited no change or were downregulated in *HRAS*- and *NRAS*-
3 mutated samples, we restricted the search to genes that were hyper- (or hypo-) methylated in
4 *BRAF*-mutated samples but exhibited no change or were hypo- (or hyper-) methylated in *HRAS*-
5 and *NRAS*-mutated samples in the third step. Thus, for each driver gene, we obtained a list of
6 aberrantly methylated probes associated with genes that were aberrantly expressed.

7

8 **Acknowledgements**

9 We thank Kristin Harper for editorial assistance.

10

1 References

- 2 1. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al.
3 Increased methylation variation in epigenetic domains across cancer types. *Nat Genet.*
4 2011;43(8):768-75. doi:10.1038/ng.865
- 5 2. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large
6 hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors.
7 *Genome Med.* 2014;6(8):61. doi:10.1186/s13073-014-0061-y
- 8 3. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric
9 adenocarcinoma. *Nature.* 2014;513(7517):202-9. doi:10.1038/nature13480
- 10 4. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA
11 methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol.*
12 2015;9(3):555-68. doi:10.1016/j.molonc.2014.10.012
- 13 5. Chen Y, Breeze CE, Zhen S, Beck S, Teschendorff AE. Tissue-independent and tissue-
14 specific patterns of DNA methylation alteration in cancer. *Epigenetics Chromatin.* 2016;9:10.
15 doi:10.1186/s13072-016-0058-4
- 16 6. Nejman D, Straussman R, Steinfeld I, Ruvolo M, Roberts D, Yakhini Z, et al. Molecular
17 rules governing de novo methylation in cancer. *Cancer Res.* 2014;74(5):1475-83.
18 doi:10.1158/0008-5472.CAN-13-3042
- 19 7. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, et al. Correlation of an
20 epigenetic mitotic clock with cancer risk. *Genome Biol.* 2016;17(1):205. doi:10.1186/s13059-016-
21 1064-3
- 22 8. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours.
23 *Nature.* 2012;490(7418):61-70. doi:10.1038/nature11412
- 24 9. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med.*
25 2004;10(8):789-99. doi:10.1038/nm1087
- 26 10. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer
27 genome landscapes. *Science.* 2013;339(6127):1546-58. doi:10.1126/science.1235122
- 28 11. Kawano H, Saeki H, Kitao H, Tsuda Y, Otsu H, Ando K, et al. Chromosomal instability
29 associated with global DNA hypomethylation is associated with the initiation and progression of
30 esophageal squamous cell carcinoma. *Ann Surg Oncol.* 2014;21 Suppl 4:S696-702.
31 doi:10.1245/s10434-014-3818-z
- 32 12. Mudbhary R, Hoshida Y, Chernyavskaya Y, Jacob V, Villanueva A, Fiel MI, et al. UHRF1
33 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell.*
34 2014;25(2):196-209. doi:10.1016/j.ccr.2014.01.003
- 35 13. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al.
36 Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.
37 *Cancer Cell.* 2010;17(5):510-22. doi:10.1016/j.ccr.2010.03.017
- 38 14. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, et al. IDH1 mutation is
39 sufficient to establish the glioma hypermethylator phenotype. *Nature.* 2012;483(7390):479-83.
40 doi:10.1038/nature10866
- 41 15. Fang M, Ou J, Hutchinson L, Green MR. The BRAF oncoprotein functions through the
42 transcriptional repressor MAFK to mediate the CpG Island Methylator phenotype. *Mol Cell.*
43 2014;55(6):904-15. doi:10.1016/j.molcel.2014.08.010

- 1 16. Serra RW, Fang M, Park SM, Hutchinson L, Green MR. A KRAS-directed transcriptional
2 silencing pathway that mediates the CpG island methylator phenotype. *Elife*. 2014;3:e02313.
3 doi:10.7554/eLife.02313
- 4 17. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al.
5 Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*.
6 2013;499(7457):214-8. doi:10.1038/nature12213
- 7 18. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat*
8 *Rev Genet*. 2012;13(7):484-92. doi:10.1038/nrg3230
- 9 19. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human
10 colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific
11 CpG island shores. *Nat Genet*. 2009;41(2):178-86. doi:10.1038/ng.298
- 12 20. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, et al. Regions of
13 focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with
14 nuclear lamina-associated domains. *Nat Genet*. 2012;44(1):40-6. doi:10.1038/ng.969
- 15 21. Yang Z, Jones A, Widschwendter M, Teschendorff AE. An integrative pan-cancer-wide
16 analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer.
17 *Genome Biol*. 2015;16:140. doi:10.1186/s13059-015-0699-9
- 18 22. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG
19 island methylator phenotype underlies sporadic microsatellite instability and is tightly associated
20 with BRAF mutation in colorectal cancer. *Nat Genet*. 2006;38(7):787-93. doi:10.1038/ng1834
- 21 23. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck
22 squamous cell carcinomas. *Nature*. 2015;517(7536):576-82. doi:10.1038/nature14129
- 23 24. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, et al. Oncometabolite 2-
24 hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.
25 *Cancer Cell*. 2011;19(1):17-30. doi:10.1016/j.ccr.2010.12.014
- 26 25. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and
27 IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair
28 hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553-67. doi:10.1016/j.ccr.2010.11.015
- 29 26. Yu DH, Ware C, Waterland RA, Zhang J, Chen MH, Gadkari M, et al. Developmentally
30 programmed 3' CpG island methylation confers tissue- and cell-type-specific transcriptional
31 activation. *Mol Cell Biol*. 2013;33(9):1845-58. doi:10.1128/MCB.01124-12
- 32 27. Dong Y, Loessner D, Irving-Rodgers H, Obermair A, Nicklin JL, Clements JA. Metastasis
33 of ovarian cancer is mediated by kallikrein related peptidases. *Clin Exp Metastasis*.
34 2014;31(1):135-47. doi:10.1007/s10585-013-9615-4
- 35 28. Ip YC, Cheung ST, Lee YT, Ho JC, Fan ST. Inhibition of hepatocellular carcinoma
36 invasion by suppression of claudin-10 in HLE cells. *Mol Cancer Ther*. 2007;6(11):2858-67.
37 doi:10.1158/1535-7163.MCT-07-0453
- 38 29. Zhang W, Hou T, Niu C, Song L, Zhang Y. B3GNT3 Expression Is a Novel Marker
39 Correlated with Pelvic Lymph Node Metastasis and Poor Clinical Outcome in Early-Stage
40 Cervical Cancer. *PLoS One*. 2015;10(12):e0144360. doi:10.1371/journal.pone.0144360
- 41 30. Tarnowski M, Schneider G, Amann G, Clark G, Houghton P, Barr FG, et al. RasGRF1
42 regulates proliferation and metastatic behavior of human alveolar rhabdomyosarcomas. *Int J Oncol*.
43 2012;41(3):995-1004. doi:10.3892/ijo.2012.1536
- 44 31. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate
45 breast cancer metastasis to the brain. *Nature*. 2009;459(7249):1005-9. doi:10.1038/nature08021

- 1 32. Lin H, Huang JF, Qiu JR, Zhang HL, Tang XJ, Li H, et al. Significantly upregulated
2 TACSTD2 and Cyclin D1 correlate with poor prognosis of invasive ductal breast cancer. *Exp Mol*
3 *Pathol.* 2013;94(1):73-8. doi:10.1016/j.yexmp.2012.08.004
- 4 33. Zang M, Zhang B, Zhang Y, Li J, Su L, Zhu Z, et al. CEACAM6 promotes gastric cancer
5 invasion and metastasis by inducing epithelial-mesenchymal transition via PI3K/AKT signaling
6 pathway. *PLoS One.* 2014;9(11):e112908. doi:10.1371/journal.pone.0112908
- 7 34. Garcia-Murillas I, Sharpe R, Pearson A, Campbell J, Natrajan R, Ashworth A, et al. An
8 siRNA screen identifies the GNAS locus as a driver in 20q amplified breast cancer. *Oncogene.*
9 2014;33(19):2478-86. doi:10.1038/onc.2013.202
- 10 35. Liu Z, Chen X, Wang Y, Peng H, Wang Y, Jing Y, et al. PDK4 protein promotes
11 tumorigenesis through activation of cAMP-response element-binding protein (CREB)-Ras
12 homolog enriched in brain (RHEB)-mTORC1 signaling cascade. *J Biol Chem.*
13 2014;289(43):29739-49. doi:10.1074/jbc.M114.584821
- 14 36. Oh ET, Park HJ. Implications of NQO1 in cancer therapy. *BMB Rep.* 2015;48(11):609-17.
- 15 37. Ameziane-El-Hassani R, Schlumberger M, Dupuy C. NADPH oxidases: new actors in
16 thyroid cancer? *Nat Rev Endocrinol.* 2016. doi:10.1038/nrendo.2016.64
- 17 38. Tiedemann RL, Hlady RA, Hanavan PD, Lake DF, Tibes R, Lee JH, et al. Dynamic
18 reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. *Oncotarget.*
19 2016;7(2):1927-46. doi:10.18632/oncotarget.6481
- 20 39. Kolbe DL, DeLoia JA, Porter-Gill P, Strange M, Petrykowska HM, Guirguis A, et al.
21 Differential analysis of ovarian and endometrial cancers identifies a methylator phenotype. *PLoS*
22 *One.* 2012;7(3):e32941. doi:10.1371/journal.pone.0032941
- 23 40. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors
24 promoted by DNA hypomethylation. *Science.* 2003;300(5618):455. doi:10.1126/science.1083557
- 25 41. Nasr AF, Nutini M, Palombo B, Guerra E, Alberti S. Mutations of TP53 induce loss of
26 DNA methylation and amplification of the TROP1 gene. *Oncogene.* 2003;22(11):1668-77.
27 doi:10.1038/sj.onc.1206248
- 28 42. Struhl K. Is DNA methylation of tumour suppressor genes epigenetic? *Elife.*
29 2013;3:e02475. doi:10.7554/eLife.02475
- 30 43. O'Hagan HM, Wang W, Sen S, Destefano Shields C, Lee SS, Zhang YW, et al. Oxidative
31 damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members
32 to promoter CpG Islands. *Cancer Cell.* 2011;20(5):606-19. doi:10.1016/j.ccr.2011.09.012
- 33 44. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, et al. Tumour
34 hypoxia causes DNA hypermethylation by reducing TET activity. *Nature.* 2016;537(7618):63-8.
35 doi:10.1038/nature19081
- 36 45. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu
37 Y, et al. Integrated genomic characterization of endometrial carcinoma. *Nature.*
38 2013;497(7447):67-73. doi:10.1038/nature12113
- 39 46. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid
40 carcinoma. *Cell.* 2014;159(3):676-90. doi:10.1016/j.cell.2014.09.050
- 41 47. Fecteau RE, Lutterbaugh J, Markowitz SD, Willis J, Guda K. GNAS mutations identify a
42 set of right-sided, RAS mutant, villous colon cancers. *PLoS One.* 2014;9(1):e87966.
43 doi:10.1371/journal.pone.0087966
- 44 48. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun.*
45 2015;6:8971. doi:10.1038/ncomms9971

- 1 49. Fahrner JA, Eguchi S, Herman JG, Baylin SB. Dependence of histone modifications and
2 gene expression on DNA hypermethylation in cancer. *Cancer Res.* 2002;62(24):7213-8.
- 3 50. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation
4 can alter gene expression and is a therapeutic target in cancer. *Cancer Cell.* 2014;26(4):577-90.
5 doi:10.1016/j.ccr.2014.07.028
- 6 51. Luo J, Li YN, Wang F, Zhang WM, Geng X. S-adenosylmethionine inhibits the growth of
7 cancer cells by reversing the hypomethylation status of c-myc and H-ras in human gastric cancer
8 and colon cancer. *Int J Biol Sci.* 2010;6(7):784-95.
- 9 52. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al.
10 A beta-mixture quantile normalization method for correcting probe design bias in Illumina
11 Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29(2):189-96.
12 doi:10.1093/bioinformatics/bts680
- 13 53. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al.
14 Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium
15 HumanMethylation450 microarray. *Epigenetics.* 2013;8(2):203-9. doi:10.4161/epi.23470
- 16 54. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad
17 Sci U S A.* 2003;100(16):9440-5. doi:10.1073/pnas.1530509100
- 18 55. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene
19 set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression
20 profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50. doi:10.1073/pnas.0506580102
- 21 56. Spruijt CG, Vermeulen M. DNA methylation: old dog, new tricks? *Nat Struct Mol Biol.*
22 2014;21(11):949-54. doi:10.1038/nsmb.2910
- 23 57. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer
24 Cell.* 2012;22(1):9-20. doi:10.1016/j.ccr.2012.06.008
25
26

1 **List of figures**

2 **Fig 1. Driver gene mutations are significantly associated with DNA methylation in various**
3 **cancers.** (A) Examples of mutations in 15 driver genes display an uneven distribution along the
4 first principal component (PC1) of DNAm, biased toward either the positive extreme (+) or the
5 negative extreme (-). Tumor samples are ordered vertically by their coordinates on PC1, from
6 small (-, *bottom*) to large (+, *top*). A black line indicates the presence of the mutated driver gene
7 in a sample, whereas a white line indicates its absence. Note that a sample at an extreme (+/-) of a
8 PC does not necessarily correspond to high or low methylation. (B) Example of a cancer with
9 driver gene mutations unevenly distributed on PC1, resulting in distinct methylation patterns:
10 *ARID1A/PIK3CA*-mutated stomach adenocarcinomas (STADs) display a methylation pattern
11 distinct from *TP53*-mutated STADs. Shown is a heat map of methylation levels for the top 1,000
12 most heavily weighted probes in PC1. Each column represents a sample ordered by its PC1
13 coordinate, from small (-, *left*) to large (+, *right*). Each row represents a probe site. Five column
14 sidebars are shown at the top. The bottom three indicate mutation statuses for *TP53*, *PIK3CA*, and
15 *ARID1A*. *TP53*-mutated STADs display lower methylation levels at the selected CpG sites
16 compared with the majority of *ARID1A/PIK3CA*-mutated STADs. The top two indicate cell
17 proliferation rates estimated by the DNAm-based mitotic index (epiTOC) and the mRNA
18 expression-based mitotic index [7]. Each index is normalized to the range between 0 (white) and
19 1 (black). Samples ordered by PC1 are significantly correlated with epiTOC ($q=0$) but not the
20 mRNA expression-based index ($q=0.22$) [see results in (C)]. (C) In 15 of 18 cancer types examined,
21 mutated driver genes were associated with one or more of the top five methylation PCs, shown as
22 rows. The three driver genes most significantly associated with each PC are reported. Driver genes
23 associated with a negative extreme of the PC are blue, whereas associations with the positive

1 extreme are written in red. Background colors indicate correlations ($q < 0.05$; Spearman correlation)
2 with two mitotic indices: DNAm-based index (epiTOC) (light green), or mRNA-based index (light
3 yellow), or both (orange). See Table 1 for cancer type abbreviation.

4
5 **Fig 2. Driver gene-methylation associations and CpG subsets.** (A) The total number of probes

6 associated with any driver gene is shown for each cancer type (*top of each column*). Each point
7 represents the fraction of corresponding probes associated with a driver gene (y-axis). Names are
8 shown for the top three driver genes if they account for more than 10% of total probes (dotted line).

9 (B) Driver genes with the most probe associations in each cancer type (gene names in panel). The
10 bar plots show the proportion of associated probes in each of the three CpG subsets (CGIs; SSs;
11 or open sea), stratified by the direction of association (+/-). Dashed lines indicate divisions
12 expected if associations were proportionally distributed. No probes were associated with driver
13 genes in BLCA, LUSC, and READ. See Table 1 for cancer type abbreviations.

14
15 **Fig 3. Proportions of positive and negative associations with methylation for 17 recurrently**

16 **mutated driver genes.** (A) Barplots show the proportion of methylation probes for each driver
17 gene (*labels at bottom*) and cancer type (*labels at top*) displaying positive and negative
18 associations. Positive associations are plotted above the horizontal line, negative associations are

19 below the horizontal line. Associations are further stratified by CpG subset: CGI (CpG islands),
20 SS (shores and shelves of CGIs; 4kb regions flanking CGIs), and open sea (regions outside CGIs
21 and SSs). Driver genes can be classified into 3 groups based on directionality of their predominant
22 associations (*negative, positive, inconsistent*). Genes in figure were associated with more than

23 1,000 probes, in at least two cancer types. (B) Plotted as in (A), using: (1) positively associated

1 and hypermethylated probes and (2) negatively associated and hypomethylated probes. *Hyper- or
2 hypomethylated probes were not identified for GBM, STAD, SKCM, and TGCT due to a lack of
3 control samples. See Table 1 for cancer type abbreviations.

4
5 **Fig 4. Driver gene-associated methylation patterns can be used to identify molecular**
6 **subtypes.** Heat maps for (A) thyroid carcinoma (THCA) and (B) uterine corpus endometrial
7 carcinoma (UCEC) depict hierarchical clustering of methylation values of the union set of the 500
8 probes most significantly associated with each of the three dominant driver genes in each cancer
9 type. Each column represents a sample, and each row represents a probe. Mutation status shown
10 in upper panel. Left panel sidebars indicate CpG subset and average methylation level across
11 control samples. The arrow in (B) indicates methylation probes displaying more hypomethylation
12 in samples where *PTEN* and *CTNNB1* mutations co-occurred than samples with *PTEN* mutations
13 alone.

14
15 **Fig 5. Differential expression of JAK and STAT family genes correlated with differential**
16 **methylation in thyroid cancer.** (A) Shown are box plots for gene expression levels (*top plots*)
17 and methylation levels (*bottom plots*) for *STAT1*, *STAT3*, *STAT4*, *STAT5A*, *JAK1*, and *JAK3*,
18 grouped by *BRAF*-mutated (*red*), *HRAS*-mutated (*blue*), *NRAS*-mutated (*green*), and control (*grey*)
19 samples. Gene names and Spearman rho (with p-value) between gene expression and methylation
20 among tumor samples are shown on top. Probe names (where methylation levels were measured)
21 are shown at bottom. (B) Shown are snapshots from the UCSC genome browser for *STAT1* and
22 *JAK3*, with CpG islands (CGIs) indicated below (*green arrow*: CGIs; *red arrow*: probes flanking
23 the CGIs). Methylation levels of the indicated regions are shown in panel (C). (C) Box plots show

1 methylation levels (y-axis) at probes in *STAT1* and *JAK3* for *BRAF*-mutated, *HRAS*-mutated,
2 *NRAS*-mutated, and control samples. The shown probes fall in the south shores and shelves [or
3 Ss, indicated by red arrows in (B)] of the *STAT1* promoter CGI and the 3' CGI of *JAK3* [indicated
4 by green arrows in (B)].
5

1 List of tables

2 Table 1. Number of tumor samples, control samples, and driver genes across 18 cancer types

Cancer type	# tumor samples	# control samples	# driver genes ^a
Bladder urothelial carcinoma (BLCA)	130	21	48
Breast invasive carcinoma (BRCA)	652	102	51
Colon adenocarcinoma (COAD)	219	38	650
Glioblastoma multiforme (GBM)	144	2	16
Head and neck squamous cell carcinoma (HNSC)	306	60	34
Kidney renal clear cell carcinoma (KIRC)	245	160	14
Kidney renal papillary cell carcinoma (KIRP)	153	45	10
Liver hepatocellular carcinoma (LIHC)	202	50	11
Lung adenocarcinoma (LUAD)	407	32	84
Lung squamous cell carcinoma (LUSC)	74	43	10
Pancreatic adenocarcinoma (PAAD)	90	10	31
Prostate adenocarcinoma (PRAD)	261	50	22
Rectum adenocarcinoma (READ)	80	7	7
Skin cutaneous melanoma (SKCM)	372	3	56
Stomach adenocarcinoma (STAD)	244	2	366
Testicular germ cell tumor (TGCT)	142	0	25
Thyroid carcinoma (THCA)	446	56	5
Uterine corpus endometrial carcinoma (UCEC)	135	46	62

3 ^a Defined as MutSigCV-reported driver genes mutated in at least five tumor samples from our data
4 set for a given cancer type (see MutSigCV in [17]).

5

6

7

8

9

10

11

12

13

14

15

1 **Table 2. Cell proliferation rate and associated mutated driver genes**

Cancer type	High proliferation (epiTOC)	Low proliferation (epiTOC)	High proliferation (mRNA)	Low proliferation (mRNA)
BLCA			RB1, TP53	
BRCA	CDH1, FOXA1		BCORL1, RB1, TP53	CDH1, MAP3K1, PIK3CA
COAD				
GBM				
HNSC	CASP8, CTCF, EPHA2, HRAS	TP53		
KIRC	BAP1, PBRM1, SETD2		BAP1, PTEN, SETD2, TP53	PBRM1, VHL
KIRP	NF2, SETD2		SETD2	
LIHC			RB1, TP53	
LUAD	KRAS	KEAP1	COL5A2, GLDC, KEAP1, MYH7, RB1, SMARCA4, STRA8, TP53	KRAS
LUSC			TP53	SLC28A1
PAAD	CDKN2A, KRAS			
PRAD	SPOP		TP53	
READ				
SKCM			ALPK2, IL5RA, NBP1, PTEN, TP53, XIRP2	
STAD	189 driver genes (not including TP53)		52 driver genes (including TP53)	CDH1, RHOA
TGCT	MLLT3, MUC6			KIT
THCA			BRAF	NRAS
UCEC	ARID1A, ARID5B, BCOR, CUX1, KRAS, PIK3R1, PTEN, SIN3A, ZFH3	TP53	NA ^b	NA ^b

2 ^a Association is computed by Wilcoxon rank sum test at $q < 0.05$ for each cancer.

3 ^b Too few ($n=5$) with both gene expression and mutation data for UCEC.

4

1 **Table 3. Associations between mutated driver genes and HyperZ and HypoZ indices or site-**
2 **specific methylation alterations.**

Cancer	# driver genes associated with HyperZ and/or HypoZ	# driver genes associated with any probe
Bladder urothelial carcinoma (BLCA)	1	0
Breast invasive carcinoma (BRCA)	2	50
Colon adenocarcinoma (COAD)	67	189
Glioblastoma multiforme (GBM)	3	16
Head and neck squamous cell carcinoma (HNSC)	2	34
Kidney renal clear cell carcinoma (KIRC)	4	14
Kidney renal papillary cell carcinoma (KIRP)	2	10
Liver hepatocellular carcinoma (LIHC)	2	11
Lung adenocarcinoma (LUAD)	4	80
Lung squamous cell carcinoma (LUSC)	1	0
Pancreatic adenocarcinoma (PAAD)	0	7
Prostate adenocarcinoma (PRAD)	2	12
Rectum adenocarcinoma (READ)	0	0
Skin cutaneous melanoma (SKCM)	0	43
Stomach adenocarcinoma (STAD)	30	357
Testicular germ cell tumor (TGCT)	NA ^a	25
Thyroid carcinoma (THCA)	4	5
Uterine corpus endometrial carcinoma (UCEC)	4	53

3 ^a Controls were not available to calculate HyperZ/HypoZ indices

4

1 **Supporting information**

2 **S1 Fig. Driver gene-methylation associations across 18 cancer types (*rows*), stratified by CpG**
3 **subsets (*columns*).** The numbers (1–5) indicate the top five principal components (PCs) for each
4 probe set, whereas the color shows the significance of the strongest association between each
5 methylation PC and any driver gene. The probe sets are methylome (all probes), CpG island (CGI)
6 probes, shore and shelf (SS) probes, and open sea probes, further stratified by hyper- and
7 hypomethylation status. For GBM, STAD, SKCM, and TGCT, there were not enough control
8 samples to compute associations for hyper-/hypomethylated probes (shown in dark grey).

9
10 **S2 Fig. Driver gene mutations are significantly associated with epiTOC-uncorrelated DNA**
11 **methylation variation in various cancers.** The DNAm-based mitotic index called epiTOC
12 (epigenetic Timer Of Cancer) was used to approximate cell proliferation rate in cancer [7]. In
13 of 18 cancer types examined, driver gene mutations were associated with one or more of the top
14 five epiTOC-uncorrelated methylation principal components (PCs). Shown is a grid depicting the
15 three driver genes most significantly associated with each PC. A gene name in blue indicates that
16 mutations in a given gene were significantly associated with the negative extreme of the PC,
17 whereas red indicates association with the positive extreme of the PC. For each PC, the light green
18 background color indicates its correlation with epiTOC ($q < 0.05$; Spearman correlation). See Table
19 1 for cancer type abbreviation.

20
21 **S3 Fig. Distribution of driver gene-associated methylation probes throughout the genome in**
22 **KIRC.** Chromosomes 1 to 22 are plotted on a circle, with each chromosome plotted proportional
23 to chromosome length and labeled in the outermost track. The 14 inner tracks correspond to all 14

1 driver genes in KIRC; gene names and the number of associated probes for each are shown. For
2 each driver gene, associated probes are plotted as line segments in the corresponding track at the
3 appropriate chromosome location. The chromosome length scale is labeled for chromosome 1 (a
4 major interval indicates 90 Mb).

5
6 **S4 Fig. Probes that are negatively associated with *TP53* are shared across cancer types.** A bar
7 plot shows the number of *TP53*-negatively associated probes (y-axis in log 10 scale; the number
8 also indicated at the top of each bar) reported in at least 1-9 cancer types (x-axis).

9
10 **S5 Fig. Empirical false discovery rates (FDRs) are controlled by theoretical cutoffs at $q=0.05$**
11 **for site-specific associations.** The site-specific association was tested between every driver gene
12 and every probe. Significant associations were called at the theoretical FDR (q-value) less than
13 0.05 for each cancer type. The empirical FDR (y-axis) was estimated for the theoretical cutoff
14 ($q=0.05$) using the permutation test for each cancer type (*column*). Here, all points are below the
15 line (empirical FDR=0.05), indicating that empirical FDRs are controlled by the theoretical cutoffs.

16
17 **S1 Table.** Driver genes associated with the HyperZ and HypoZ indices across 18 cancer types,
18 stratified by direction of association

19
20 **S2 Table.** Counts of driver gene-methylation probe associations for 18 cancer types, stratified by
21 CpG subsets: methylome (all probes), CpG island probes, shores and shelves probes, and open sea
22 probes, as well as by hypo- and hypermethylated status (+: positive associations, -: negative
23 associations)

1
2 **S3 Table.** Sorted median expression levels of genes differentially expressed (relative to control
3 samples) in the two thyroid carcinoma molecular subtypes identified in Figure 4A: *BRAF*-mutated
4 versus *NRAS*-/*HRAS*-mutated; each Excel tab separates genes by the direction of expression
5 dysregulation (*up* or *down*) and the group of mutated samples showing the expression
6 dysregulation (*BRAF* or *HRAS/NRAS*)

7
8 **S4 Table.** Pathways enriched with genes dysregulated in the two thyroid carcinoma molecular
9 subtypes identified in Figure 4A: *BRAF*-mutated versus *HRAS*-/*NRAS*-mutated, stratified by the
10 direction of expression dysregulation (*up* or *down*) and the mutant group (*BRAF* or *HRAS/NRAS*)

11

Figure 1

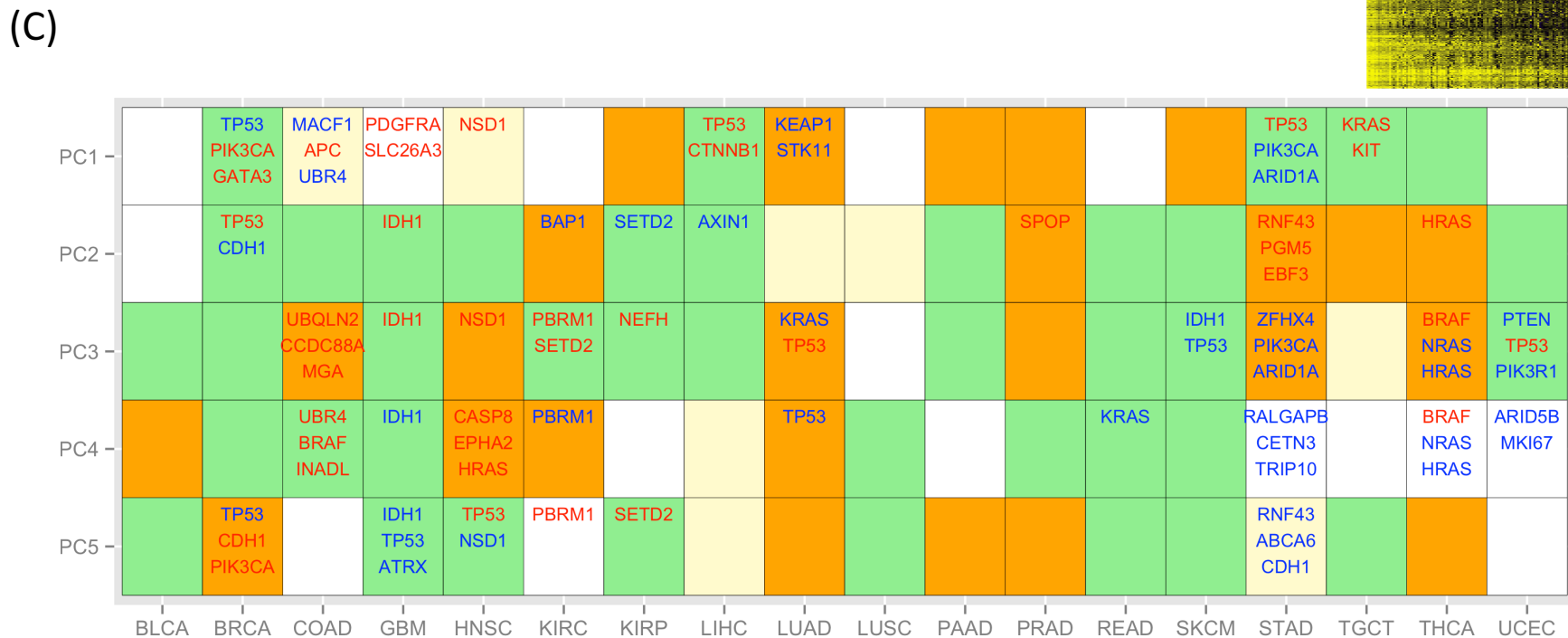
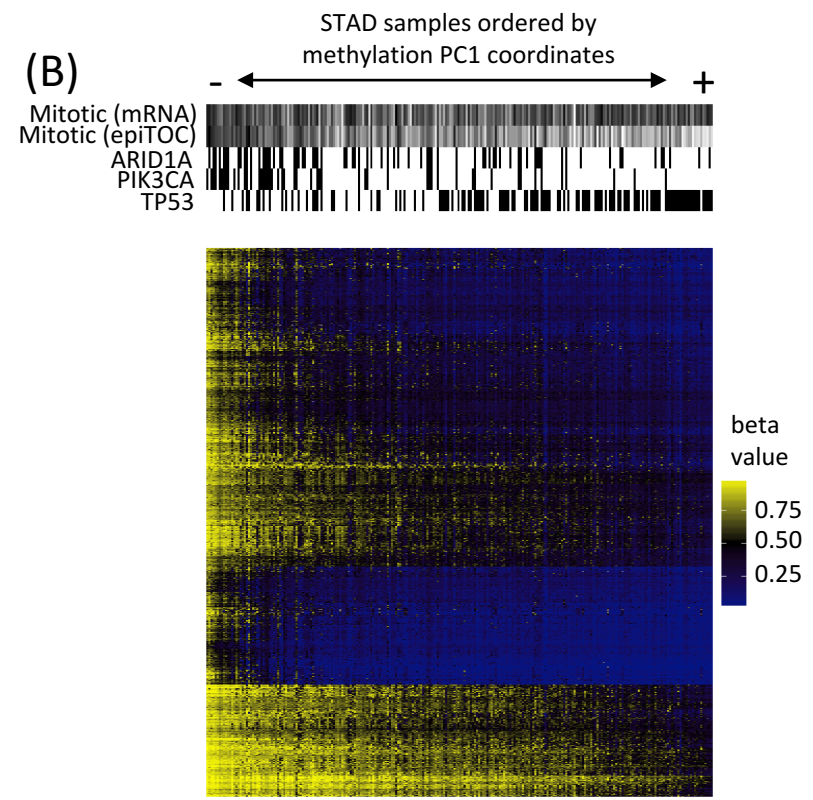
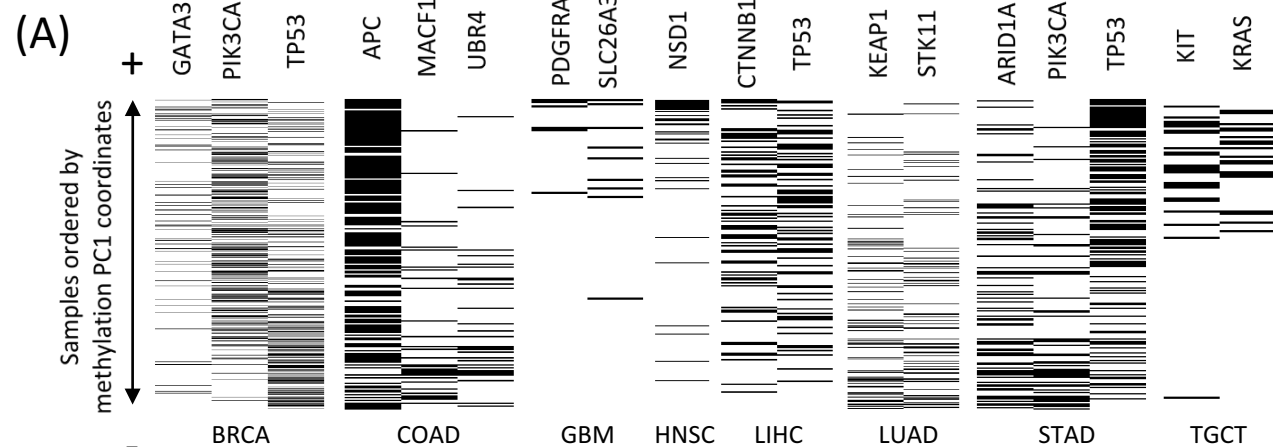
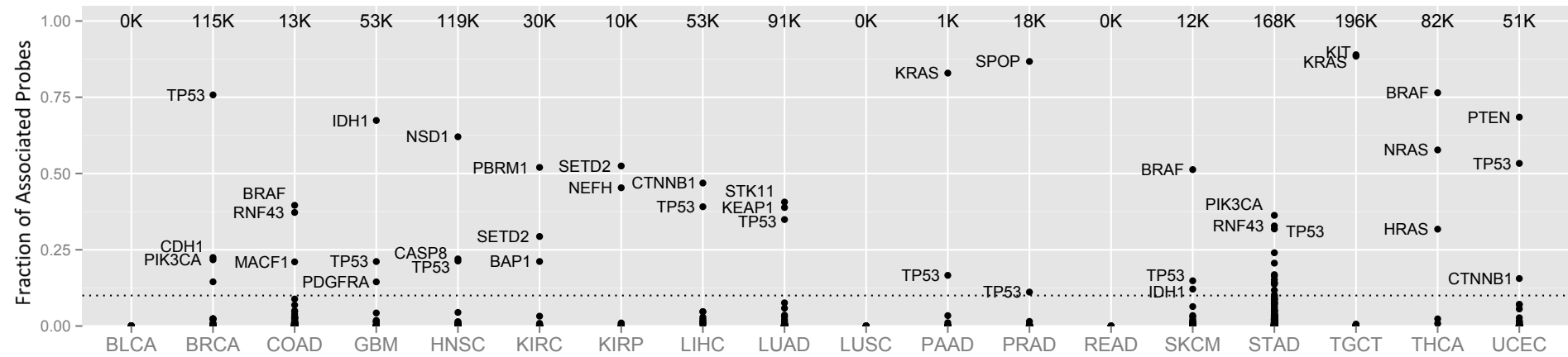


Figure 2

(A)



(B)

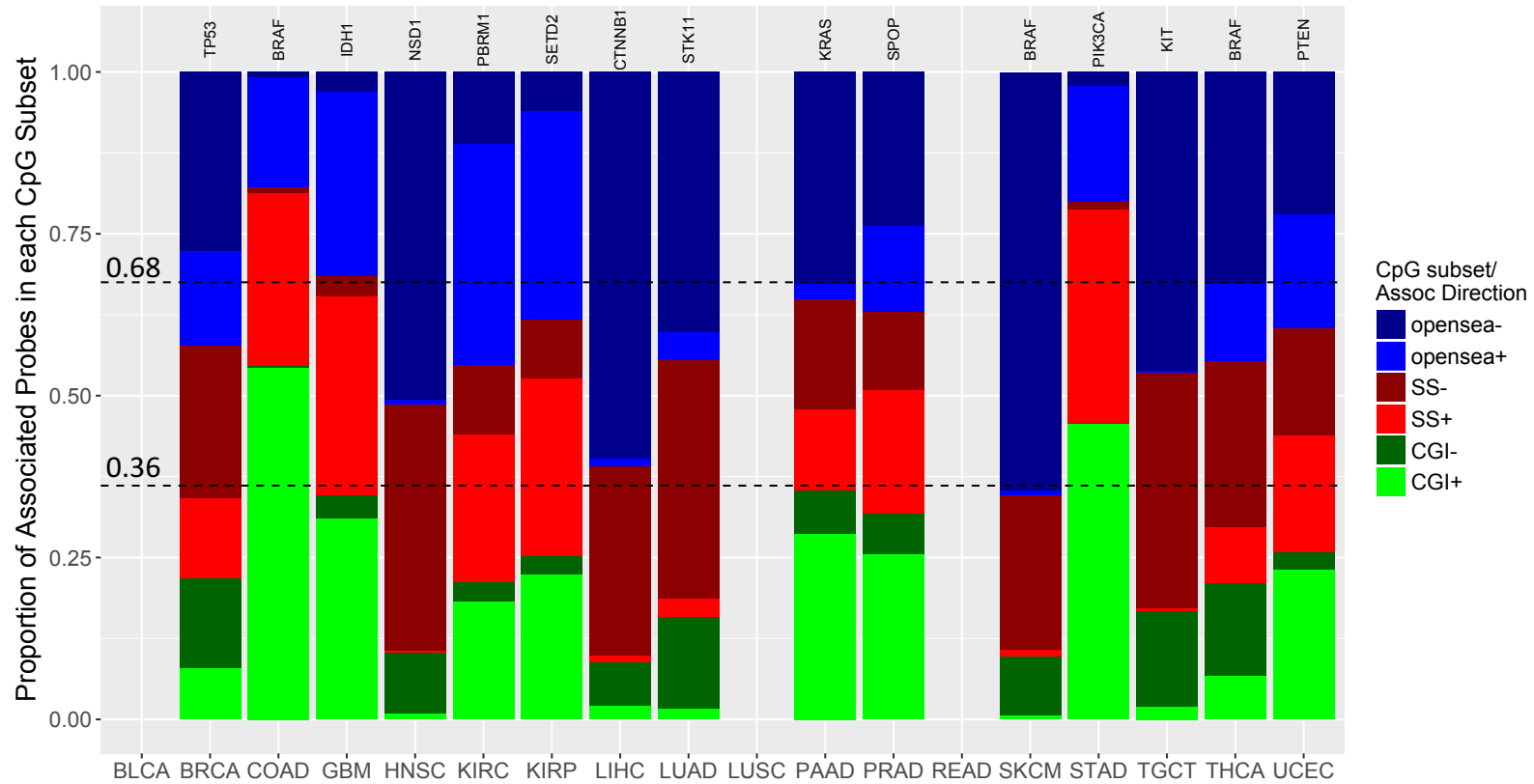


Figure 3

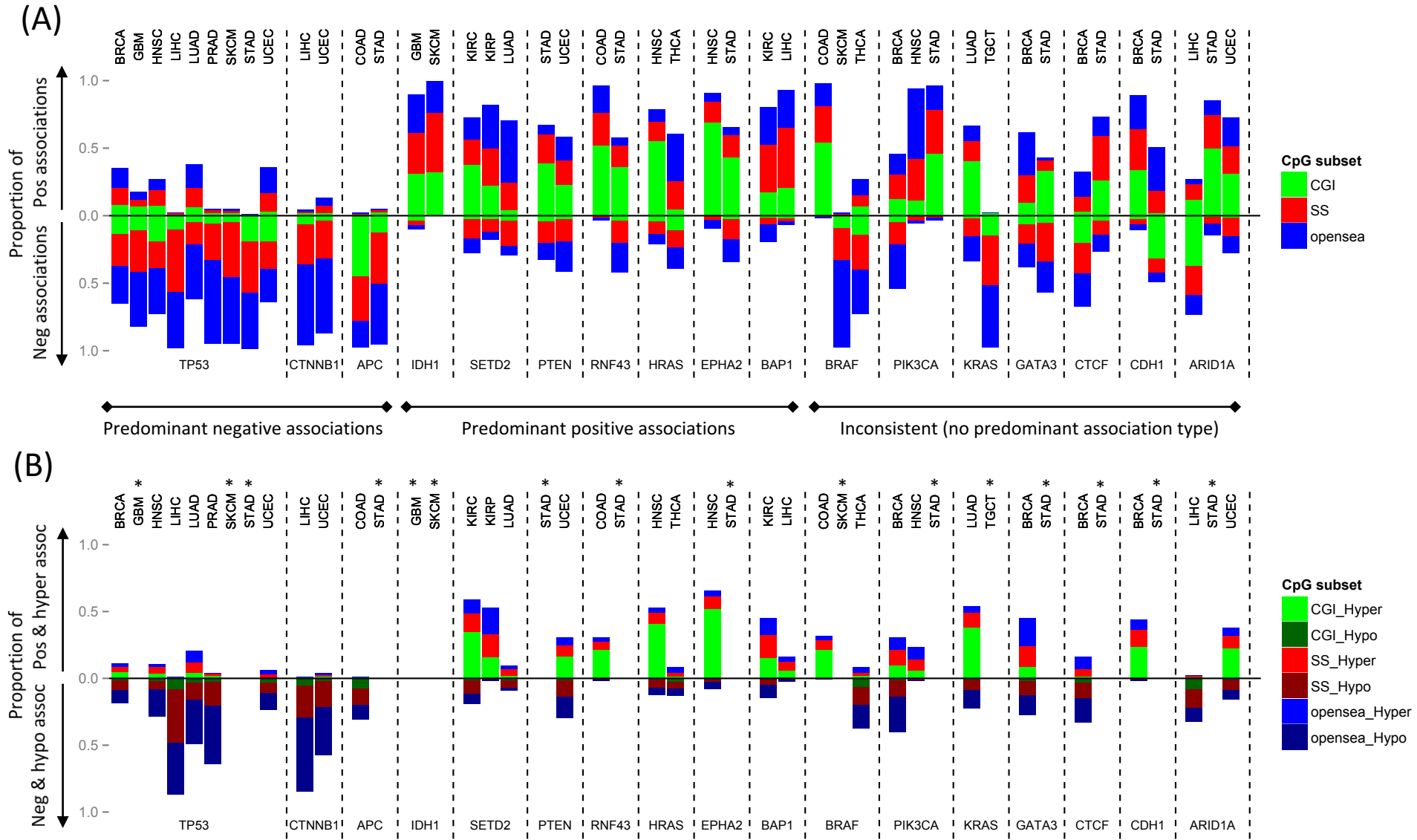
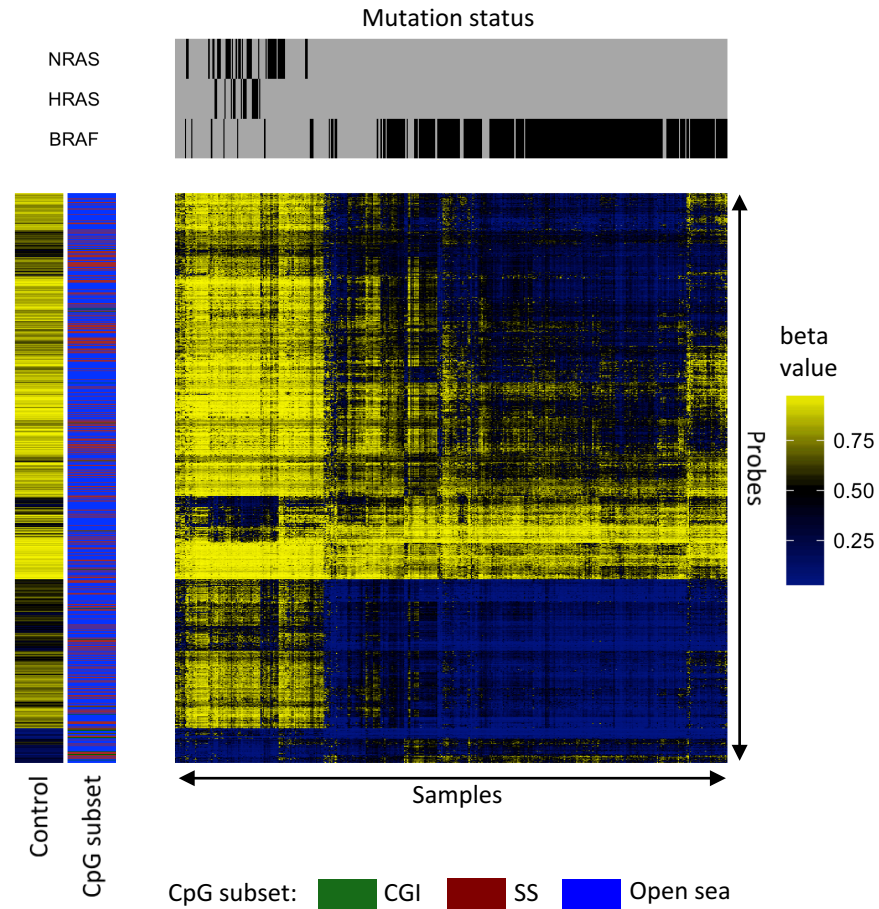


Figure 4

(A)

THCA



(B)

UCEC

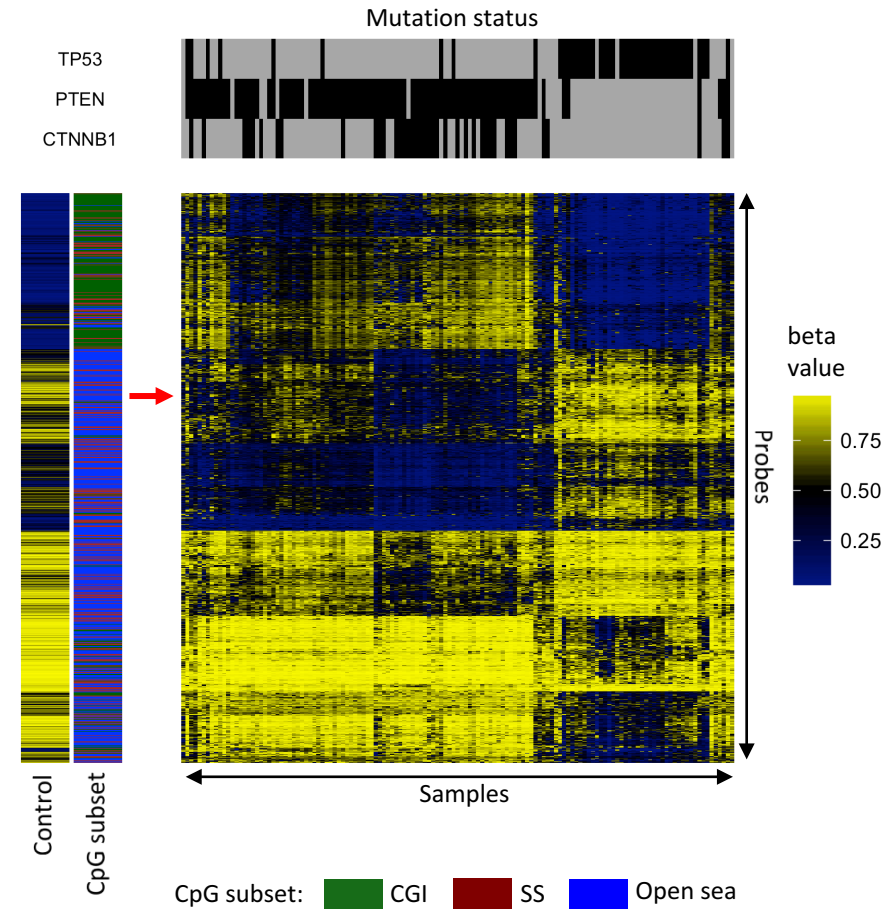


Figure 5

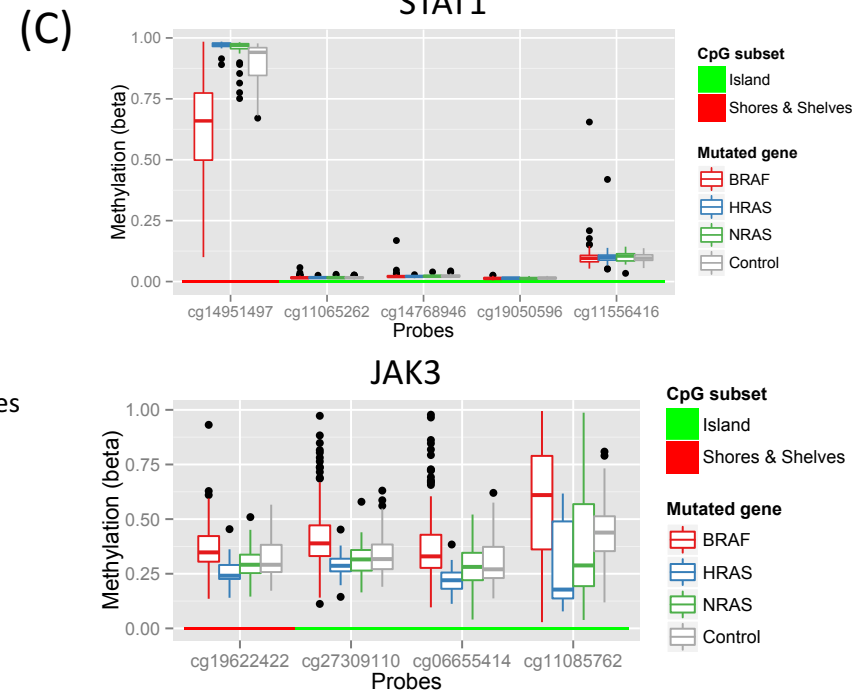
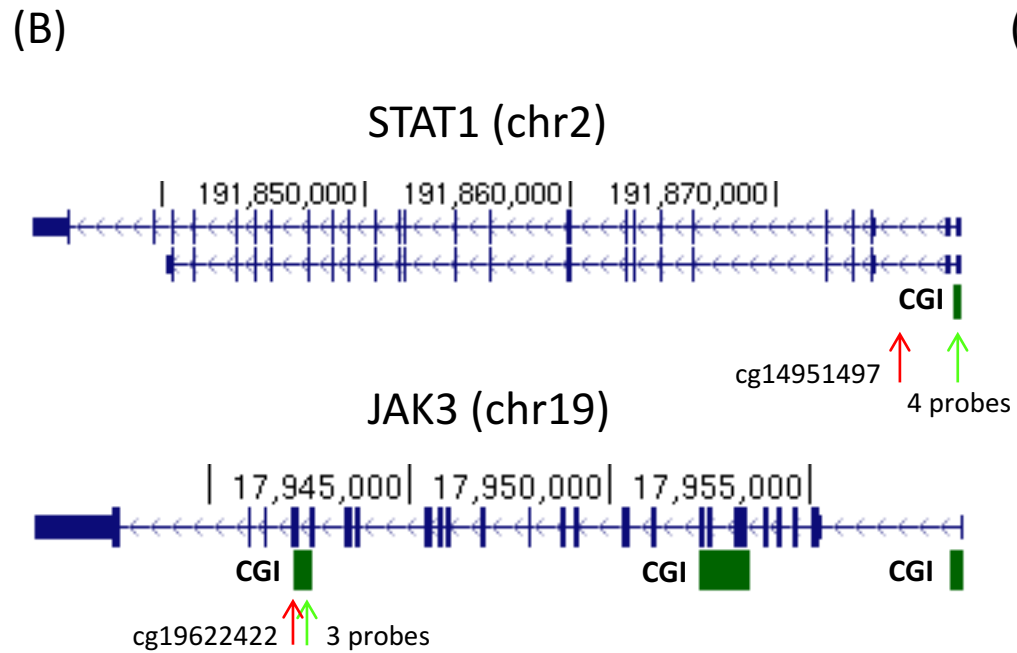
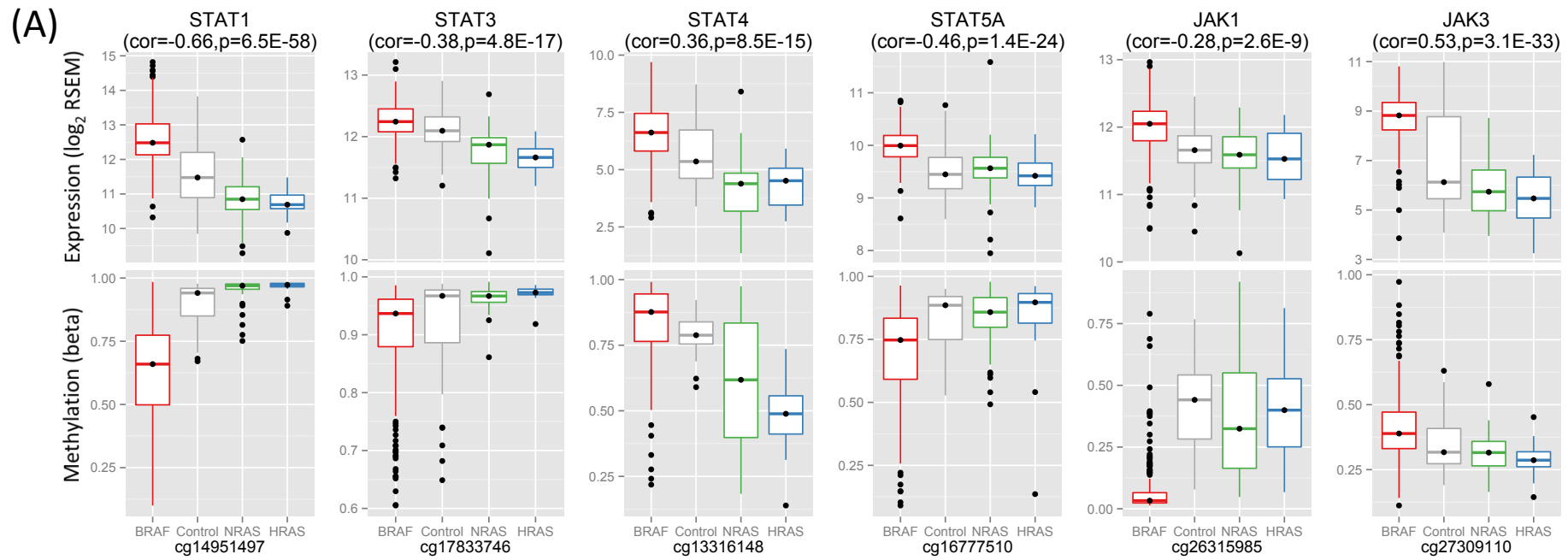


Figure S1

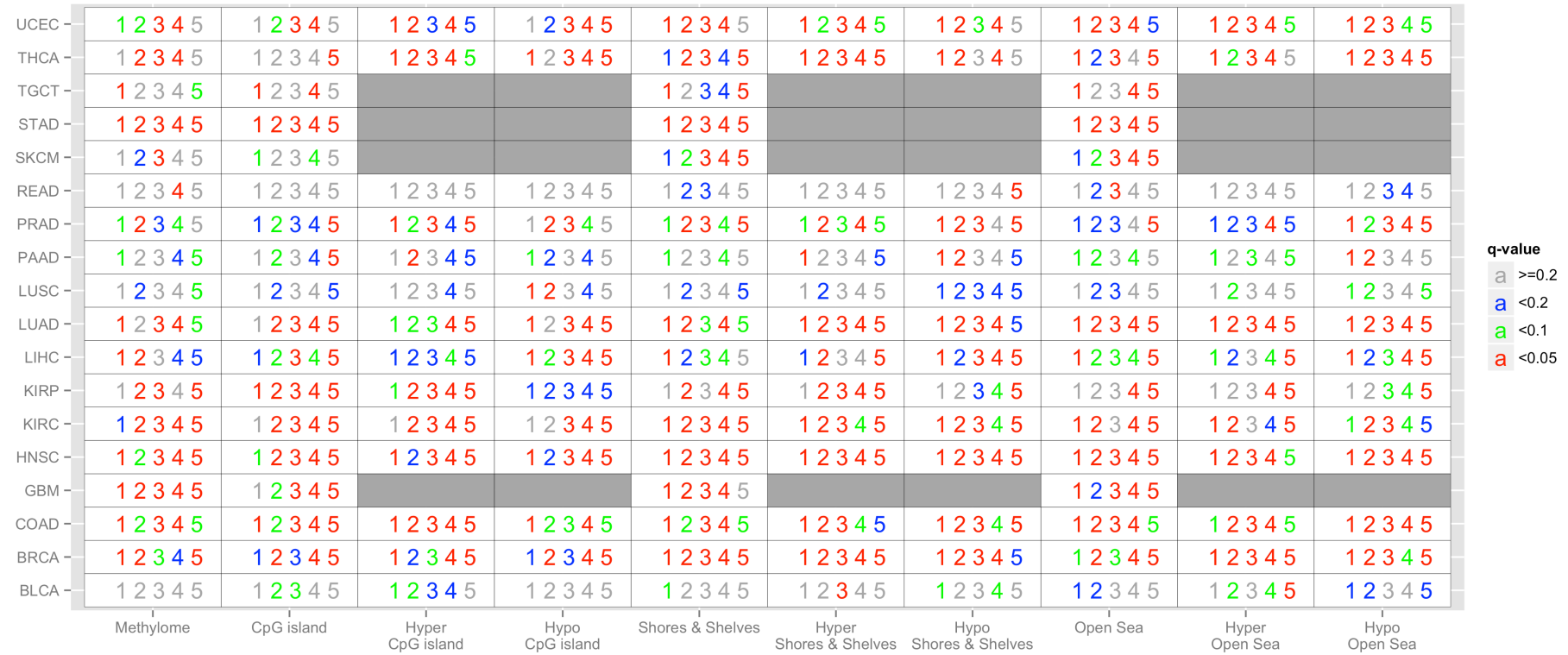


Figure S2

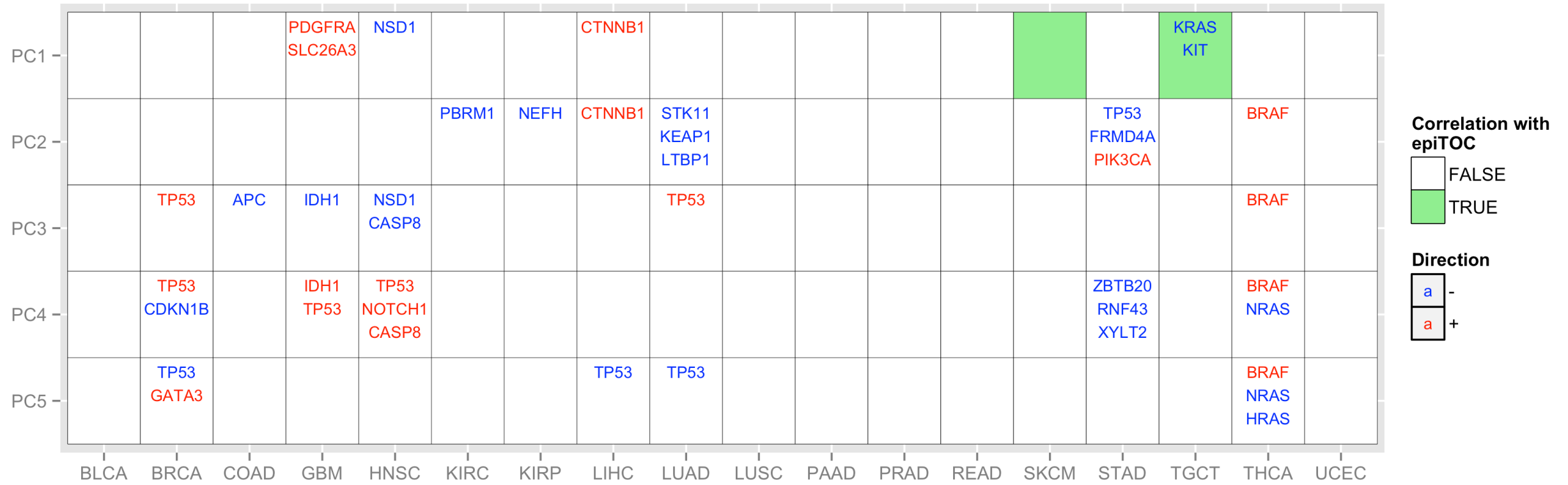


Figure S3

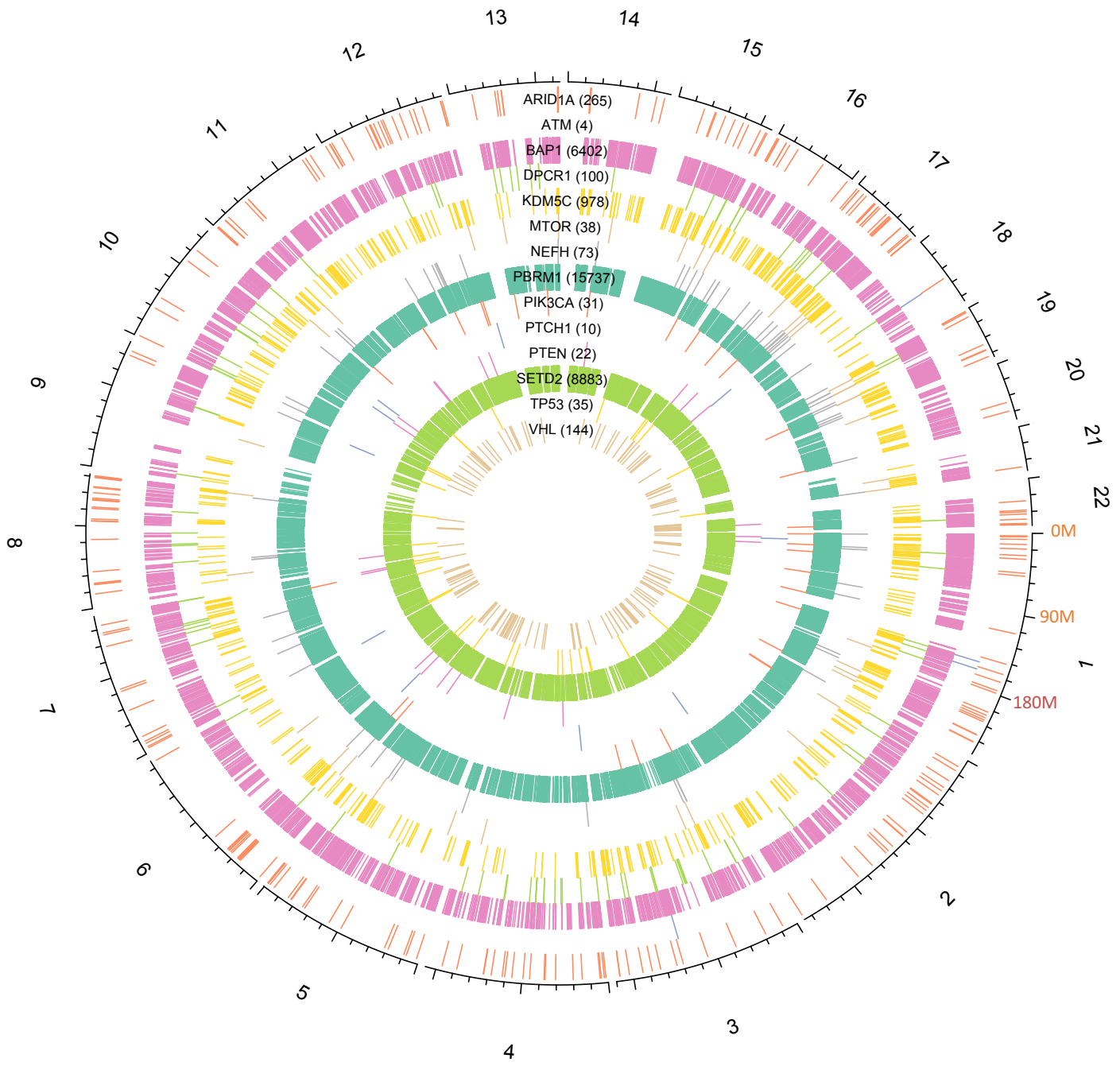


Figure S4

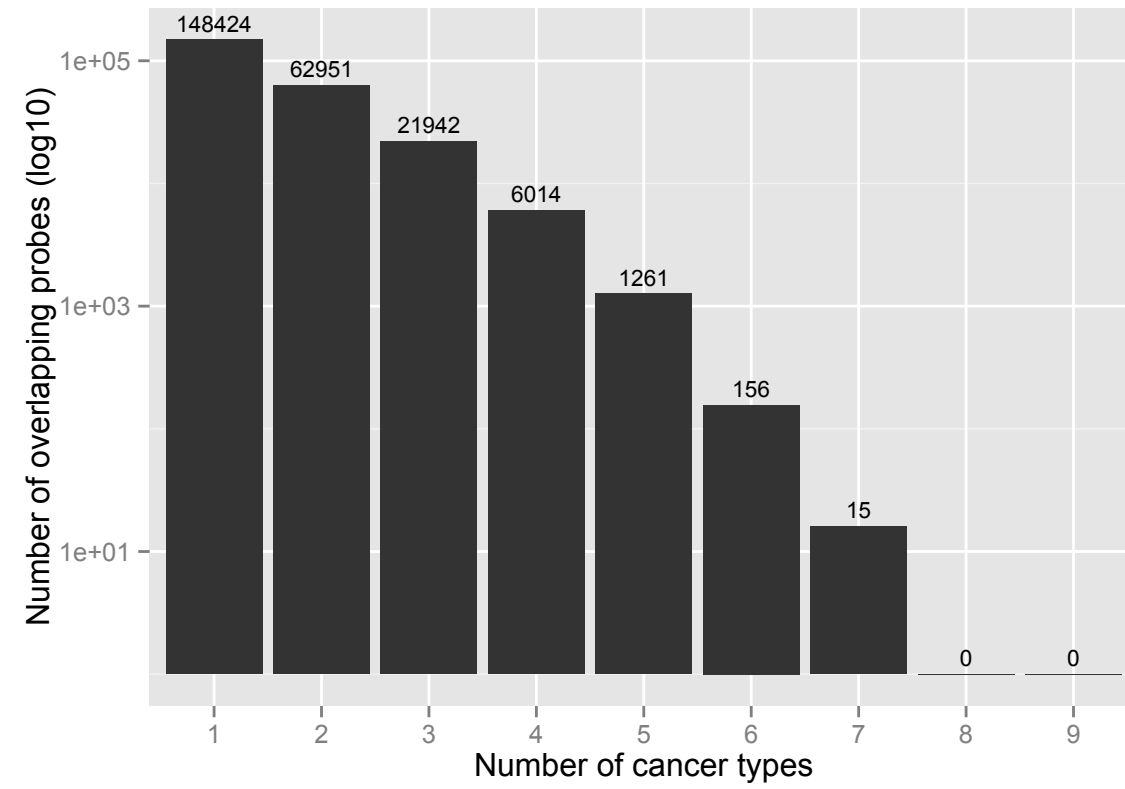


Figure S5

