bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 # The DNA methylation landscape of five pediatric-tumor
2 types

3 Alyssa C. Parker[1], Badi I. Quinteros[1], Stephen R. Piccolo[1,*]

4 1 - Department of Biology, Brigham Young University, Provo, UT, USA

5 * - Please address correspondence to S.R.P. at stephen_piccolo@byu.edu.

6 # Abstract

7 Fewer DNA mutations have been identified in pediatric tumors than adult tumors, suggesting that
8 alternative tumorigenic mechanisms, including aberrant DNA methylation, may play a prominent role in
9 pediatric tumors. Methylation is an epigenetic process of regulating gene expression in which methyl
10 groups are attached to DNA molecules, often in promoter regions. In Wilms tumors and acute myeloid
11 leukemias, increased levels of epigenetic silencing have been associated with worse patient outcomes.
12 However, to date, researchers have studied methylation primarily in adult tumors and for specific genes
13 but not on a pan-pediatric cancer scale. We addressed these gaps first by aggregating methylation data
14 from 309 noncancerous samples and establishing baseline expectations for each gene. Even though these
15 samples represent diverse tissue types and population ancestral groups, methylation levels were highly
16 consistent for most genes. Second, we compared tumor methylation levels against these baseline values
17 for five pediatric cancer types—Wilms tumors, clear cell sarcomas of the kidney, rhabdoid tumors,
18 neuroblastomas, and osteosarcomas. Hypermethylation was more common than hypomethylation—as
19 many as 11.8% of genes were hypermethylated in a given tumor, compared to a maximum of 4.8% for
20 hypomethylated genes. For each cancer type, genes with the highest variance exhibited consistently

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

21 divergent methylation patterns for distinct patient subsets. We evaluated whether genomic and
22 epigenomic abnormalities contribute to pediatric tumorigenesis in a mutually exclusive manner but did
23 not find evidence of this phenomenon. Furthermore, even though oncogenes are commonly upregulated in
24 tumors, and tumor-suppressor genes are commonly downregulated in tumors, we did not find statistical
25 evidence that methylation drives such patterns on a broad scale in pediatric tumors.

26 Introduction

27 Pediatric tumors are the leading cause of disease-related death for children in developed countries[1], and
28 those who survive pediatric cancer often experience adverse health challenges later in life[2]. Many
29 mutations and structural variants have been associated with adult forms of cancer[3]; however, significantly
30 fewer genomic abnormalities have been identified in pediatric tumors[4]. The mutation rate in pediatric
31 tumors is 14 times less than the mutation rate in adult tumors, implying that different mechanisms may be
32 involved in pediatric-cancer development than in adult cancers[5]. Of the mutations that have been
33 identified in pediatric tumors, many are associated with epigenetic regulation[1]. In many pediatric tumors,
34 molecular profiling does not identify genomic abnormalities but does show abnormal DNA methylation
35 patterns[6,7], suggesting the DNA methylation may play a critical role in tumorigenesis in these cases.

36 Gene-expression levels in cancer cells often vary from those in normal cells[8]. Such abnormalities alter
37 cellular environments and manipulate cellular processes, leading to increased survival, rapid proliferation,
38 and metastasis[9]. Oncogenes are one type of genes that contribute to the development of these abnormal
39 features. These genes are often expressed at higher levels in cancer cells than in normal cells[10]. Another
40 type of genes known as tumor suppressor genes counteracts cellular changes that lead to cancer. These
41 genes are often expressed at lower levels in tumor cells than in normal cells[11], potentially leading to rapid
42 cellular proliferation. Methylation of the promoter region is often negatively correlated with gene
43 expression levels, suggesting that DNA methylation plays a role in regulating gene expression[12,13].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

44 However, relatively little is known about global methylation patterns in pediatric tumors, the interplay
45 between methylation events and mutations in pediatric tumors, or how these observations may differ
46 between known cancer genes (oncogenes and tumor suppressors) and other genes. Prior studies have
47 focused on cancer cell lines, a single tumor type at a time, or adult cancers[14–16].

48 Computational models have been developed to identify differentially expressed genes across large sets of
49 methylation data[15,17]. Applications of these methods have found several genes with highly variant
50 methylation in adult tumors[17]. Many of the genes that exhibited highly variant methylation in tumors were
51 not previously associated with cancer, and genomic aberrations in these genes were not characteristic of
52 tumors. A small-scale study of Wilms tumors (a common pediatric cancer) showed similar results[17]. It has
53 been shown that several distinct types of cancer, including endometrioid adenocarcinomas and
54 glioblastomas, share many differentially methylated regions, suggesting that these epigenetic markers
55 may be a universal feature of cancer[18].

56 While these findings hint that aberrant methylation patterns may also be characteristic of pediatric cancer,
57 most pediatric cancers have not been analyzed for DNA methylation patterns. One study investigated
58 genomic, transcriptomic, and epigenomic patterns in acute myeloid leukemia and identified dozens of
59 hypermethylated genes and age-specific patterns[19]. Additionally, structural variations were found to be
60 more common than single nucleotide polymorphisms. A separate analysis of acute lymphoblastic
61 leukemia identified a number of genes on chromosome 3, including PPP2R3A, THRB and FBLN2, that
62 were frequently hypermethylated[16].

63 Little work has been done to specifically investigate aberrant methylation of oncogenes and tumor
64 suppressor genes in cancer. One study about endometrial cancer identified seven oncogenes that were
65 hypomethylated and upregulated and twelve tumor suppressor genes that were hypermethylated and
66 downregulated[14], suggesting that changes in DNA methylation impact gene expression and may target
67 oncogenes and tumor suppressor genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

68 We address these gaps by analyzing methylation data for five types of pediatric cancers: Wilms tumors,
69 clear cell sarcomas of the kidney, rhabdoid tumors, neuroblastomas, and osteosarcomas. We compare
70 these cancer types against each other. Furthermore, as a baseline reference, we compare the tumor data
71 against methylation levels for fetuses and children representing normal conditions for diverse cell types
72 and human populations. Because methylation events tend to be gene specific[20], we evaluate gene-level
73 patterns. But we also consider global methylation patterns. For many of the tumors, we identify genomic
74 alterations—single-nucleotide variants, small indels, and structural variants—in the tumors and evaluate
75 whether these somatic mutations exhibited gene-specific mutual exclusivity with hypo- or
76 hypermethylation events. In addition, we evaluate the consistency of these findings for oncogenes and
77 tumor suppressor genes.

78 Results

79 Our goal was to evaluate gene-level DNA methylation patterns for pediatric-tumor cells and normal cells.
80 We used publicly available data to characterize five types of pediatric cancers as well as baseline
81 methylation levels for normal cells. First, we evaluated the consistency of methylation levels representing
82 non-cancerous states in fetuses and children. Second, we compared tumor methylation levels against the
83 normal values and identified genes and tumor samples that exhibited patterns of hypomethylation or
84 hypermethylation. Next, under the assumption that tumors with aberrant methylation would be subject to
85 evolutionary constraints that are redundant with those resulting from somatic mutations, we evaluated
86 whether these two event types were mutually exclusive in a given tumor. Finally, we examined these
87 patterns within known oncogenes and tumor-suppressor genes.

88 Consistency of methylation levels in normal samples and in pediatric tumors

89 To aid in understanding how methylation levels change in cancer cells, we first characterized baseline
90 methylation levels for individual genes. We obtained Illumina Infinium 450K data for four normal

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

91 datasets. We used data from diverse cell types and human populations with the goal of identifying
92 methylation patterns that broadly represent baseline methylation states for healthy children. The datasets
93 were from chorionic villi, kidney, spinal cord, brain, muscle, nasal epithelial, and blood cells and included
94 data for a total of 309 patients of North American (n = 94), African American (n = 36), and Australian
95 ancestry (n = 179).

96 Because DNA methylation contributes to regulating gene expression, we expected most genes to exhibit a
97 consistent baseline methylation range. We anticipated that many genes (such as tumor suppressor genes)
98 would have consistently low methylation levels because those genes must be consistently expressed to
99 properly regulate cellular division, growth, and other proliferation activities. We anticipated that other
100 genes (such as oncogenes) would have consistently high levels of methylation because proper cellular
101 functioning requires that these genes remain relatively inactive. We expected to see some variance across
102 samples because we included data from multiple cell types and because methylation levels change as cells
103 respond to internal and external cues. We also anticipated that some genes would deviate from these
104 patterns, perhaps in part because they are regulated by mechanisms other than DNA methylation. Based
105 on a preliminary inspection of the data, we identified thresholds for categorizing genes based on the
106 magnitude and variance of methylation levels. We considered genes with a median value less than 0.2
107 across all samples in a given dataset to be methylated at "low" levels, genes with a median greater than
108 0.6 to be methylated at "high" methylation levels and the remaining genes as having "medium"
109 methylation levels. We categorized genes with a coefficient of variation (CV) less than 0.5 in a given
110 dataset as having "low" variance and the remaining genes as having "high" variance. In the largest normal
111 dataset (GSE89278), most genes exhibited low (58.9%) or medium methylation levels (25.6%).
112 Relatively few genes exhibited high methylation (15.5%), meaning that under normal conditions, most
113 genes appear not to be subject to strong expression constraints as a result of DNA methylation. Nearly all
114 genes (95.1%) exhibited low variance. The other normal datasets reflected similar patterns.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

115 By combining these two ways of categorizing the genes (Additional Data File S1), we found that the most
116 common combination across all datasets was low methylation / low variance (56.9-57.2% per dataset).
117 For GSE89278, the following numbers of genes fell into each category: low methylation / low variance:
118 12,874; low methylation / high variance: 534; medium methylation / low variance: 3,898; medium
119 methylation / high variance: 14; high methylation / low variance: 5,246; high methylation / high variance:
120 0. These results were similar for the other normal datasets. To explore the consistency of these patterns
121 across the normal datasets, we classified each gene into the following consistency levels: consistent (same
122 category in all four datasets), semiconsistent (same category in three datasets), and inconsistent (same
123 category in two or fewer datasets). Of the 22,253 genes that we profiled, 20,089 were consistent, 1,512
124 were semiconsistent, and 652 were inconsistent (Figure 1). Thus even though the normal datasets differed
125 based on cell type and population ancestral groups, gene-level methylation levels were largely consistent.

126 The gene with the smallest variance across the normal datasets was BTG3 (variance = 2.09E-5). In each
127 of the normal datasets, BTG3 exhibited low methylation, suggesting that a biological constraint
128 suppresses methylation of this gene under normal circumstances. The process that normally keeps BTG3
129 methylation low may be disrupted in tumors. Indeed, hypermethylation of the promoter region of BTG3
130 has previously been associated with several types of cancer, including breast[21], prostate[22], and renal[23]. The
131 gene with the largest variance across the normal datasets was DOCK11. Information about this gene is
132 sparse in the literature. Its role in cancer[24] and hypercholesterolemia[25] has been discussed. Its high
133 variance in normal cells suggests that consistent expression of DOCK11 is inessential to normal cellular
134 function and/or that processes other than promoter methylation regulate its expression.

135 **Tumor methylation relative to normal levels**

136 Under the hypothesis that tumorigenesis alters local and global methylation patterns, we evaluated the
137 extent to which methylation levels and variances differed for a given gene between normal and tumor
138 conditions. For each gene, we compared the most common median / variance category across the four
139 normal datasets against the most common category across the five cancer datasets. Typically genes stayed

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

140 under the same category. For example, of the 12,368 genes that were categorized as low methylation / low
141 variance for the normal datasets, only 549 (4.4%) changed categories: either to low methylation / high
142 variance (n = 462) or to medium methylation / low variance (n = 87) (Table 1). The 227 genes in the
143 medium methylation / high variance category were most likely to change categories, with 137 genes
144 (60.0%) changing to low methylation / high variance and 19 genes (8.4%) changing to medium
145 methylation / low variance. For the 512 genes that changed median methylation levels, approximately half
146 (n = 259, 50.6%) increased (from low to medium or medium to high). For the 562 genes that changed
147 variance categories, 519 (92.3%) increased from low to high variance. These increased variances suggest
148 that the factors that normally keep methylation levels stable under normal conditions often become
149 dysregulated in tumors. Altered expression of these genes may play a role in tumor development or may
150 lead to downstream effects that affect tumor development. For example, if methylation levels of an
151 oncogene are decreased, higher expression of the gene may result, leading to increased cellular growth,
152 proliferation, or survival[9]. On the other hand, increased methylation levels of a tumor suppressor gene
153 could cause lower gene expression and prevent it from performing regulatory functions[13].

154 To identify genes that may influence pediatric tumorigenesis, we compared tumor methylation levels for
155 each gene against the respective normal values on a per-cancer basis. We used a two-sided, Mann-
156 Whitney U test and adjusted for multiple tests using the Benjamini-Hochberg False Discovery Rate
157 (FDR)[26]. We considered genes with an FDR < 0.05 and an absolute methylation change > 0.02 to be
158 statistically significant. Out of the 22,253 genes we analyzed, 37 exhibited differential methylation for at
159 least one cancer type. Of these genes, 19 had decreased methylation in tumors, and 18 had increased
160 methylation, including 15 genes that were differentially methylated for Wilms tumors, 0 for clear cell
161 sarcomas, 6 for rhabdoid tumors, 17 for neuroblastomas, and 4 for osteosarcomas; 4 genes were
162 statistically significant for 2 or more cancer types (Table S1; Figure 2). We note that the number of
163 significant genes was smaller for cancer types with relatively small sample sizes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

164 We performed a pathway analysis for the significant genes from each cancer type using reactome.org[27,28].
165 For Wilms tumors, rhabdoid tumors, and osteosarcomas, pathways associated with ubiquitination,
166 transcriptional regulation, and growth signaling were significant; for neuroblastomas, growth-signaling,
167 integrin-signaling, and blood-biosynthesis pathways were among the most significantAdditional Data
168 Files S2-S5. These functions have plausible connections to tumor biology because they are central to
169 cellular function in general,[29,30] but it is difficult to make precise inferences due to the relatively small
170 numbers of significant genes.

171 To characterize differences in methylation across the five cancer types, independent of normal
172 methylation levels, we performed a one-way ANOVA test for each gene and adjusted the p-values using
173 the FDR correction. Methylation levels of most genes were consistent across the 5 cancer types, including
174 for genes that we had identified as being differentially methylated in tumors compared to normal cells.
175 Methylation levels for 19 genes differed significantly across the cancer types (FDR < 0.05; Table S2;
176 Figure 3). But in most cases, the mean differences were small. For example, the largest absolute
177 difference in mean methylation between tumor types for any of the significant genes was an increase of
178 0.24 in CCSK compared to RT (KCNQ1OT1 gene). Of the 19 genes, 10 fell into the medium methylation
179 / low variance category, a higher proportion (0.526) than the overall proportion of genes in this category
180 (0.175).

181 Next we focused on the 20 genes for which methylation levels varied most across all tumor types and
182 plotted the methylation values, relative to normal levels, for each patient (Figure 4). Some tumors
183 exhibited relatively high methylation levels for nearly all of these genes, whereas other tumors exhibited
184 relatively low methylation levels for the same genes. This was especially true for Wilms tumors, rhabdoid
185 tumors, and osteosarcomas. For example, the average methylation difference (relative to normals) for
186 these 20 genes was 0.13 for 58 (44.3%) of the Wilms tumors but -0.11 for the remaining tumors. For the
187 osteosarcomas and rhabdoid tumors, we observed a similar pattern in which subsets of 49 (57.0%) and 35
188 (51.5%) tumors, respectively, showed markedly higher methylation levels than the remaining tumors.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

189 These patterns of consistently divergent methylation might be useful for identifying tumor subtypes in a
190 precision-medicine context[31] and may be driven by factors such as the presence of somatic mutations and
191 structural alterations that affect global methylation in characteristic ways.

192 To investigate global methylation patterns, we calculated gene-level z-scores for each tumor using the
193 normal data as a reference. In cases where a tumor's methylation value was more than three standard
194 deviations higher than the mean normal value for a particular gene, we classified that gene as being
195 hypermethylated in that tumor. In cases where a tumor's methylation value was more than three standard
196 deviations lower than the mean normal value for a particular gene, we classified that gene as being
197 hypomethylated in that tumor. Then we calculated the proportion of hypomethylated and hypermethylated
198 genes for each patient. We categorized tumors for which greater than 1% of genes were hypermethylated
199 as being "frequently hypermethylated" and tumors for which greater than 1% of genes were
200 hypomethylated as being "frequently hypomethylated." Frequent hypermethylation was more common
201 (17.5% of tumors) than frequent hypomethylation (3.4%)(Figure 5). The maximum percentage of
202 hypermethylated genes for any particular patient was 11.8%, compared to a maximum of 4.8% for
203 hypomethylated genes.

204 Cancer mutation data analysis

205 Although they occur less frequently in pediatric tumors than in adult tumors, somatic mutations often
206 contribute to pediatric tumorigenesis[4,5]. To evaluate the frequency of and interplay between somatic
207 mutations and methylation events, we examined pediatric tumors for which both data types were
208 available. Mutation data were available for somatic single nucleotide variants (SNVs),
209 insertions/deletions (indels), and RNA fusions. We used the RNA fusion data as indicators of structural
210 DNA variants. Copy-number data were available for only a small number of tumors, so we did not
211 include this data type in the analysis. Methylation, SNV, indel, and RNA fusion data were available for
212 Wilms tumors (n = 41), neuroblastomas (n = 65), and osteosarcomas (n = 66) but not for the other tumor
213 types. For a given tumor, we considered genes with at least one SNV, indel, or RNA fusion event to be

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

214 "mutated." Across all tumor types, aberrant methylation—via either hypomethylation or
215 hypermethylation of a given gene—occurred less frequently (in 1.1% of tumors) than mutations (2.8%);
216 the largest disparity occurred for neuroblastomas (Table S3). Wilms tumors and osteosarcomas were
217 aberrantly methylated nearly twice as often as neuroblastomas (1.4% vs. 0.8%; Table S3). In contrast,
218 Wilms tumors and osteosarcomas were mutated less than half as often as neuroblastomas (1.7-1.8%
219 vs. 4.6%).

220 Aberrant methylation levels and damaging mutations can have similar downstream effects in tumors[32].
221 After one of these alteration types has occurred in a given gene, it may be less likely for cells with a
222 second alteration in the same gene to gain an additional selective advantage. Thus, having both a mutation
223 and aberrant methylation in one gene may be less likely than expected by random chance. This mutual-
224 exclusivity hypothesis has been examined extensively for pairs of genes in which somatic mutations
225 might occur across diverse types of cancers[33–36]. It has also been investigated for DNA methylation
226 events, though to a lesser extent[37]. Little is known about mutual exclusivity between methylation events
227 and somatic mutations in pediatric tumors.

228 Treating each combination of tumor and gene as an independent observation, we examined whether DNA
229 methylation events are mutually exclusive with somatic mutations. Mutations and aberrant methylation
230 co-occurred rarely (0.03%) in the same gene and the same tumor (Table S3). Because both of these event
231 types are rare, we reduced the data to the 962 genes for which at least 5 mutation events and at least 5
232 aberrant methylation events had occurred across the cancer types. We used a permutation test to evaluate
233 whether mutations and aberrant methylation in the same gene and tumor co-occurred less frequently than
234 would be expected by random chance. Across all tumors and the 962 genes, we observed 318 co-
235 occurrences, whereas the average number of co-occurrences in the permuted data was 341.6. However,
236 this difference was not statistically significant ($p = 0.092$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

237 **Oncogenes and tumor suppressor genes**

238 Because oncogenes are typically expressed at relatively high levels and tumor suppressor genes are
239 typically expressed at relatively low levels, we expected that oncogenes would have higher methylation
240 levels than tumor suppressor genes. We identified 80 "tier 1" oncogenes and 141 "tier 1" tumor
241 suppressor genes in the Cancer Genome Census[38] and calculated the mean methylation value for each
242 gene in the normal datasets and used a two-sample t-test to evaluate whether these mean values differed
243 between the oncogenes and tumor suppressor genes. Under the hypothesis that oncogenes would be
244 methylated at higher levels than tumor-suppressor genes, we used a one-sided test. The mean of the
245 means for oncogenes was 0.02 higher than for tumor-suppressor genes; however, this difference was not
246 statistically significant ($p = 0.109$). Many tumor-suppressor genes were highly methylated, and many
247 oncogenes were methylated at low levels (Figure S1).

248 Next we evaluated the extent to which methylation levels differed for oncogenes and tumor-suppressor
249 genes between tumor and normal conditions. First, we filtered the results from the two-sided Mann-
250 Whitney U tests described above to include only oncogenes and tumor suppressor genes. Using $FDR <$
251 0.05 and an absolute mean difference > 0.02 as constraints, only 1 tumor suppressor gene was statistically
252 significant (CTCF in Wilms tumors). No oncogenes reached statistical significance. We relaxed the
253 threshold to $FDR < 0.2$ and removed the mean difference constraint. In this case, CTCF was significant
254 for all cancer types except CCSKFigure S2. DNM2 (a tumor suppressor gene) was significant for OS, and
255 GNA11 (an oncogene) was significant for NBLTable S4. Mean methylation values for none of the
256 oncogenes or tumor suppressor genes differed across the tumor types ($FDR < 0.2$; Table S2).

257 Across the tumor types, mutation rates for oncogenes and tumor suppressor genes (0.031-0.089) were
258 approximately twice the mutation rates for other genes (0.017-0.046) (Table 2), which aligns with prior
259 evidence that mutations in these genes are associated with tumorigenesis. However, aberrant methylation
260 rates for oncogenes and tumor suppressor genes (0.007-0.014) were approximately equal to methylation
261 rates for other genes (0.008-0.014). In Wilms tumors and osteosarcomas, mutation rates for oncogenes

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

262 and tumor suppressor genes were approximately three times higher than aberrant methylation rates, while
263 mutation and methylation rates were approximately equal for other genes. In neuroblastomas, mutation
264 rates for oncogenes and tumor suppressor genes were approximately twelve times higher than aberrant
265 methylation rates but only six times higher for other genes.

266 We performed a modified version of the permutation-based, mutual-exclusivity analysis in which we
267 searched for co-occurrences of either 1) a mutation in an oncogene and aberrant hypomethylation of the
268 same gene or 2) a mutation in a tumor suppressor gene and hypermethylation of the same gene. Because
269 few genes met these criteria, we performed this analysis with the oncogenes (n = 4) and tumor-suppressor
270 genes (n = 29) that were mutated in at least 2 tumors and aberrantly methylated in at least 2 tumors. These
271 event combinations were not mutually exclusive for oncogenes (p = 0.90) nor for tumor suppressors (p =
272 0.54).

273 Discussion

274 Using publicly available data, we examined methylation patterns for five childhood cancers. We
275 summarized baseline methylation levels for healthy children, identified deviations from these baseline
276 levels in tumors, and evaluated mutual exclusivity of methylation events and somatic mutations.
277 Subsequently, we evaluated these patterns for oncogenes and tumor-suppressor genes specifically. In the
278 309 healthy samples we studied, DNA methylation levels were highly consistent, suggesting that
279 biological processes maintain this consistency in diverse tissue types and ancestral groups. In the 531
280 tumors we studied, hypermethylation or hypomethylation of promoter regions was a common feature,
281 providing additional evidence that the tumor epigenome contributes to pediatric tumorigenesis.
282 Hypermethylation was more common than hypomethylation. Hypermethylation of promoter regions has
283 been associated with decreased gene expression[12]—tumors with frequent hypermethylation may result in
284 broad silencing of proteins necessary for normal cellular function. Hypermethylation affecting multiple

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

285 genes in the same tumor has been identified in adult thyroid neoplasms[39] and colorectal tumors[40] and has
286 been associated with multidrug resistance in cell cultures[41]. However, global hypomethylation has
287 garnered more attention than multigene hypermethylation[42]. Yet in the pediatric-tumor samples that we
288 examined, widespread hypomethylation in a given tumor was uncommon. We also observed tumor
289 subsets that exhibited divergent methylation—methylation levels for specific genes were consistently
290 either high or low in many tumors of a given cancer type. While methylation dysregulation may be a
291 common feature of pediatric tumors in general, specific mechanisms may drive these divergent changes in
292 each tumor subset, and there may be overlap in these mechanisms across cancer types. Identifying such
293 mechanisms may be useful for developing targeted treatments and informing patients.

294 Our mutual-exclusivity analysis did not provide evidence that somatic mutations and aberrant methylation
295 are mutually exclusive with each other, whether for oncogenes and tumor-suppressor genes specifically or
296 across all genes. However, we had access to a limited number of tumors (n = 172) for which both
297 mutation data and methylation data were available; a larger-scale analysis is warranted. Furthermore, we
298 evaluated single-nucleotide variants, indels, and structural variants because data were available for these
299 mutation types. But large-scale amplifications and deletions occur regularly in tumors[43], so including
300 copy-number data in this type of analysis would also be useful in future research.

301 In normal cells, methylation levels did not differ significantly between known cancer genes (oncogenes
302 and tumor suppressors) and other genes. In tumors, few oncogenes and tumor suppressors had high rates
303 of aberrant methylation, suggesting that while aberrant methylation occurs in pediatric tumors,
304 dysregulation likely does not occur disproportionately in these known gene categories. Global
305 hypermethylation and/or hypomethylation patterns may be more important indicators of pediatric-tumor
306 biology than local events. However, CTCF may be one exception. CTCF promoter methylation was
307 significantly increased relative to normal conditions in every cancer type that we examined, except clear
308 cell sarcoma of the kidney, which had a small sample size. CTCF codes for a zinc finger nuclease that is
309 highly conserved in eukaryotes[44]. As a regulatory protein involved in DNA methylation[45], improper

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

310 function of this gene could lead to epigenetic abnormalities. Mutations in CTCF have previously been
311 associated with several cancer types, including breast cancers, prostate cancers, and Wilms tumors[46,47].
312 Our findings across multiple pediatric cancer types suggest that CTCF hypermethylation—and
313 consequent inhibition of gene expression— may influence pediatric tumorigenesis broadly and cause
314 downstream effects.

315 In performing these analyses, we needed to choose arbitrary thresholds at times. For example, we
316 specified three standard deviations above or below the mean in the normal data as a conservative
317 threshold to indicate aberrant methylation. Using this threshold, we found that tumors were mutated more
318 frequently than they were aberrantly methylated. However, if we had relaxed this threshold to two
319 standard deviations, the average aberrant methylation rate would have been 0.045 rather than 0.012 (the
320 average mutation rate was 0.027). Thus, our conclusions would have been moderately different. The use
321 of arbitrary thresholds in research is common and cannot be completely avoided, especially when
322 discipline-specific precedents have yet to be specified by the research community. In this scenario, we
323 have introduced a new way of identifying aberrant methylation events. Future quantitative, experimental,
324 and translational work will be necessary to improve our ability to determine when a particular gene in a
325 given biological sample is methylated to an extent that alters that gene's behavior and may in turn have
326 clinical relevance.

327 A better understanding of DNA methylation's role in pediatric tumors could shed light on mechanisms of
328 tumorigenesis and eventually lead to insights about patient care and treatments. DNA methylation
329 inhibitors have proven effective in some adult cancers, especially hematologic malignancies[20] and may
330 prove beneficial in pediatric cases. However, these therapies primarily target hypermethylation in a broad
331 sense and thus may not be suitable for targeting specific genes[48,49]. Furthermore, as we have shown,
332 hypermethylation events may be more common than hypomethylation events in pediatric tumors. Little is
333 understood about how to reverse hypomethylation in vivo; however, global-hypomethylation patterns

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

334 may be useful as biomarkers for therapies for targeting genes that have been activated as a result of
335 hypomethylation.

336 Methods

337 Normal methylation datasets

338 We downloaded datasets containing DNA methylation levels for cohorts of normal patients as a way to
339 establish a baseline against which tumor methylation could be compared for pediatric samples. We
340 selected four datasets from Gene Expression Omnibus that used Illumina Infinium
341 HumanMethylation450K arrays to profile normal cells in healthy fetuses and children aged 18 or under
342 for which raw (.idat) files were available. Illumina Infinium HumanMethylation450K is a microarray
343 platform that detects DNA methylation at over 450,000 locations in the human genome. Beta values from
344 these arrays indicate the ratio of methylated signals to unmethylated signals[50]. Values closer to one
345 indicate relatively high methylation levels; values closer to zero indicate relatively low methylation
346 levels. Below we describe each dataset included in the study and provide accession identifiers from Gene
347 Expression Omnibus.

348 GSE69502 originated from a study of Canadian second trimester fetuses[51]. Tissue was collected from
349 chorionic villi, kidney, spinal cord, brain, and muscle. The original study analyzed DNA methylation
350 differences in fetuses with spina bifida and anencephaly compared to normal fetuses. We included only
351 data from the 65 normal fetuses.

352 GSE65163 originated from a study of nasal epithelial cells from African American children aged 10-12[52].
353 The original study analyzed DNA methylation differences in children with or without asthma. We
354 included only data from the 36 children without asthma.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

355 GSE109446 originated from a study of children aged 5-18 living in Cincinnati, Ohio (USA)[53]. Similar to
356 GSE65163, nasal epithelial cells were used, and the goal of the original study was to understand
357 methylation differences in children with or without asthma. We included only data from the 29 children
358 without asthma.

359 GSE89278 originated from a study of Australian infant blood spot samples taken at birth[54]. The goal of
360 the original study was to understand methylation differences in children with and without
361 docosahexaenoic acid deficiency. We included only control data from 179 infants.

362 We chose these samples because they represented a variety of cell types and ancestral groups, which we
363 hoped would make our findings more broadly generalizable than if our samples had been from a single,
364 homogeneous population. Another factor in our decision was the age of the individuals. We searched for
365 datasets representing relatively young patients to limit the possible confounding factor of methylation
366 changes accumulating throughout life. In this process, we considered one additional normal dataset
367 (GSE72556)[55]; however, we found that the beta levels from this dataset were consistently different from
368 the other four normal datasets. This dataset originated from saliva samples, which often result in
369 systematically different methylation levels than other types of samples[56–58], perhaps due to external
370 contaminants or different sample-collection procedures. As a result, we excluded this dataset from our
371 analysis.

372 **Normal data processing**

373 We processed the data using the minfi package (version 1.34.0) from R (version 4.0.2) and the
374 Bioconductor suite[59,60]. We followed a standard workflow to process the methylation array files. The steps
375 in this workflow included preprocessing, ratio conversion, and beta value calculations. We then
376 summarized the beta values at the gene level. We mapped probes to genes based on an annotation file
377 created by Price et al. (see http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16304)[61]. Because
378 we were interested in methylation changes in the promoter regions of genes, we included only probes

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

379 within 300 base pairs of transcription start sites. We then calculated gene-level beta values for each
380 patient as the mean beta value across all remaining probes in a given gene.

381 **Methylation data acquisition and processing**

382 We obtained methylation data representing five tumor types from the Therapeutically Applicable
383 Research to Generate Effective Treatments (TARGET) Data Matrix (retrieved August 24, 2020 from
384 Website, https://ocg.cancer.gov/programs/target/data-matrix). We acquired data for 131 patients with
385 Wilms tumor, 11 patients with clear cell sarcoma of the kidney, 68 patients with rhabdoid tumor, 235
386 patients with neuroblastoma, and 86 patients with osteosarcoma. We downloaded the data from TARGET
387 using the rvest package (version 0.3.6)[62]. The data for Wilms tumors, clear cell sarcomas of the kidney,
388 neuroblastomas, and osteosarcomas were generated using the Illumina Infinium HumanMethylation450K
389 platform, and the data for rhabdoid tumors were generated using the Illumina Infinium MethylationEPIC
390 platform. The HumanMethylation450K platform produces beta values for 22,579 genes, and EPIC
391 produces values for 22,411 genes; we limited our analysis to the 22,253 genes that overlapped between
392 the two platforms.

393 We followed the same workflow that we used to process the normal data. We calculated probe-level beta
394 values using minfi and summarized values at the gene-level. We used principal component analysis to
395 assess high-level patterns across the datasets and visualized the results using a scatter plot of the first two
396 principal components. Methylation samples from each dataset generally clustered tightly with other
397 samples from the same dataset, demonstrating that batch effects were presentFigure S3. To reduce the
398 impact of these effects, we applied batch correction to a combined dataset that included all of the normal
399 datasets. We used the dataset identifier as the batch variable. In addition, we logit transformed the data
400 before performing the batch correction and inverse logit transformed the data after performing the batch
401 correction to ensure the beta values stayed between 0 and 1. We performed these transformations using
402 functions from the gtools package (version 3.8.2)[63]. Because the logit function cannot handle exact values
403 of 0 or 1, we removed genes in which the beta values were exactly 0 or 1. We also removed genes in

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

404 which beta values were NA. Filtering the data in this way removed 155 (0.7%) genes (out of 22,411 total
405 genes). To perform the batch correction, we used the ComBat function from the sva package (version
406 3.36.0)[64]. After completing the batch correction, we performed another principal component analysis and
407 visualized the results. The points no longer clustered by datasetFigure S3. We used these newly corrected
408 values for all remaining analyses.

409 **Somatic-mutation data acquisition and processing**

410 We obtained somatic-mutation data for TARGET patients via the Genomic Data Commons portal[65–67].
411 We first selected the Repository section. Under the Cases tab, we specified the TARGET program. We
412 also specified Wilms tumor, neuroblastoma, and osteosarcoma as the tumor types of interest (data were
413 unavailable for the other two tumor types). Under the Files tab, we selected "simple nucleotide variation"
414 and "annotated somatic mutation" as the data category and type, and we indicated that we would use
415 variants that had been called using Mutect2[68]. The data were stored in VCF format (version 4.2)[69].

416 We wrote custom code to parse the mutation data for each patient. We included only mutations that had
417 either 1) "HIGH" impact according to the variant annotations or 2) had "MODERATE" impact and were
418 considered to be "deleterious" by SIFT[70] or either "damaging" or "probably_damaging" according to
419 Polyphen-2[71].We considered using the specified minor-allele-frequency (MAF) values for filtering, but
420 those values were unavailable for a significant portion of the variants, so we focused on functional
421 annotations.

422 **RNA-fusion data acquisition and processing**

423 We downloaded data for RNA fusions via the Genomic Data Commons portal. Under the Cases tab, we
424 applied the same filters that we used for obtaining the somatic-mutation data. Under the Files tab, we
425 selected "structural variation" and "Transcript Fusion" as the data category and type. We specified
426 "STAR-Fusion"[72] as the workflow type and "bedpe" as the data format. We wrote custom code to parse
427 the RNA fusion data for each patient. The bedpe files described RNA transcripts from each patient that

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

428 had genetic information originating from two separate genes. We considered an RNA fusion to have a
429 functional impact on both genes affected by the fusion.

430 Methylation and mutation data integration

431 Different naming conventions were used to identify neuroblastoma and osteosarcoma patients in the
432 methylation data files versus the somatic-mutation and RNA fusion data files. We used a sample sheet
433 provided by the Genomic Data Commons (gdc_sample_sheet.2021-03-11.tsv) and the sdrf metadata files
434 for neuroblastoma and osteosarcoma to map the patient identifiers between these sources (see
435 https://target-
436 data.nci.nih.gov/Public/OS/methylation_array/METADATA/TARGET_OS_MethylationArray_20161103
437 .sdrf.txt and https://target-
438 data.nci.nih.gov/Public/NBL/methylation_array/METADATA/TARGET_NBL_MethylationArray_20160
439 812.sdrf.txt). A few files did not map properly across naming conventions. One fusion data sample was
440 mapped to two patients, so we excluded this sample. There were also two pairs of single nucleotide
441 mutation samples that were mapped to single patients, so we excluded these samples.

442 Inferring hypermethylation and hypomethylation states

443 To determine hypermethylation and hypomethylation status, we scaled the beta values for each gene to
444 have a zero mean and a standard deviation of one across all normal samples. Then for each tumor sample,
445 we compared the beta value for a given gene against the standardized values from the normal samples. If
446 a tumor's beta value was more than three standard deviations above the mean of the normal samples, we
447 classified that tumor as being hypermethylated for that gene. If a tumor's beta value was more than three
448 standard deviations below the mean of the normal samples, we classified that tumor as being
449 hypomethylated for that gene.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

450 **Mutual exclusivity**

451 We evaluated whether aberrant methylation—either hyper- or hypomethylation—was mutually exclusive
452 with somatic-mutation events for a given gene. We calculated the total number of times a gene was both
453 mutated and aberrantly methylated in the same tumor. Next we permuted methylation status for all tumors
454 and kept mutation status constant. We repeated the permutation process 10,000 times to create an
455 empirical null distribution to use as a baseline. For each permutation, we calculated the number of times a
456 gene was both mutated and aberrantly methylated. We then compared the number of times that mutations
457 co-occurred with aberrant methylation in the non-permuted data relative to the permuted data and
458 calculated a p-value based on these numbers.

459 **Oncogenes vs tumor suppressor genes**

460 We identified genes known to be oncogenes or tumor suppressor genes, based on information from The
461 Cancer Gene Census[38]. We included only tier 1 genes classified as "oncogene" or "TSG" and ignored any
462 for which it was ambiguous or unknown. To be classified as a tier 1 gene, there must be scientific
463 evidence that a gene plays a role in cancer development and that mutations in the gene modify the activity
464 of the associated protein. Several genes were listed in the Cancer Genome Census as both oncogenes and
465 tumor suppressor genes; because of this ambiguity, we excluded these from our analysis.

466 In evaluating our hypothesis that oncogenes generally have decreased methylation levels in tumor
467 samples relative to normal samples and that tumor suppressor genes generally have increased methylation
468 levels in tumor samples relative to normal samples, we applied a two-sided, Mann-Whitney U test to each
469 oncogene and tumor suppressor gene to identify genes that did (or did not) align with these expectations.

470 **Abbreviations**

Abbreviation	Term
WT	Wilms tumor

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CCSK	Clear cell sarcoma of the kidney
RT	Rhabdoid tumor
NBL	Neuroblastoma
OS	Osteosarcoma
TSG	Tumor suppressor gene

471

472

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

473

Tables

474 **Table 1: Summary of changes in methylation level/variance categories between normal and cancer**
475 **datasets.** We assigned each gene to a category that indicated whether it was methylated at low, medium,
476 or high levels and whether it had low or high variance across samples in a given dataset. The table shows
477 the total number of genes in each category for the normal datasets and the number (and percentage) of
478 genes that changed from one category to another in the tumor datasets.

Normal category	Tumor category	Total # genes	# genes changed	% genes changed
high methylation / low variance	medium methylation / low variance	3403	41	1.2
low methylation / high variance	low methylation / low variance	604	20	3.3
low methylation / high variance	medium methylation / high variance	604	9	1.5
low methylation / high variance	medium methylation / low variance	604	4	0.66
low methylation / low variance	low methylation / high variance	12368	462	3.7
low methylation / low variance	medium methylation / low variance	12368	87	0.7
medium methylation / high variance	low methylation / high variance	227	137	60
medium methylation / high variance	medium methylation / low variance	227	19	8.4
medium methylation / low variance	high methylation / low variance	5651	159	2.8
medium methylation / low variance	low methylation / high variance	5651	23	0.41
medium methylation / low variance	low methylation / low variance	5651	52	0.92
medium methylation / low variance	medium methylation / high variance	5651	34	0.6

479

480

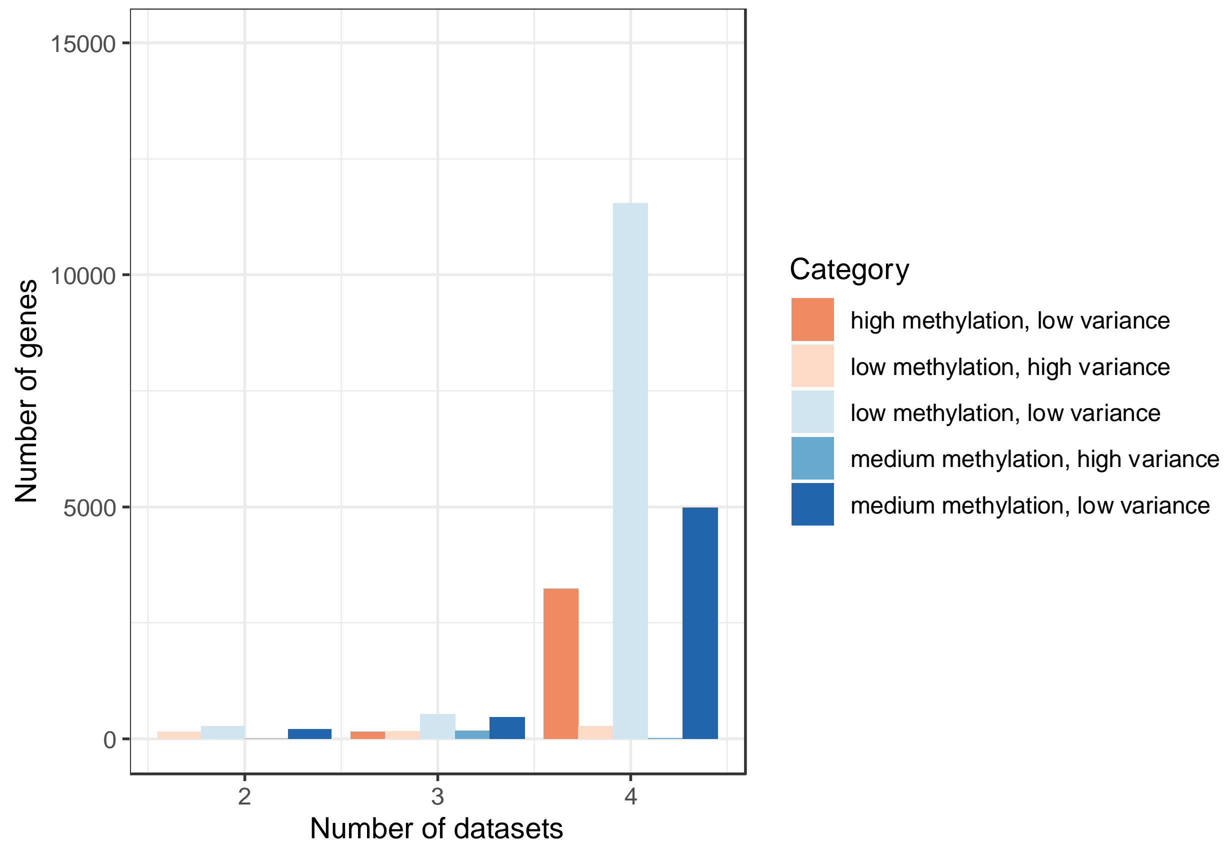
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

481 **Table 2: Aberrant methylation and mutation rates in oncogenes, tumor suppressor genes, and all**
482 **other genes.** For three pediatric tumor types, we identified aberrant methylation events (either
483 hypomethylation or hypermethylation) that had occurred in a given tumor and gene. For the same
484 tumor/gene combinations, we identified somatic single-nucleotide variants, indels, and structural variants
485 that had occurred. These numbers indicate overall rates of aberrant methylation or mutation across all
486 tumors of a given type. Methylation rates and mutation rates were typically similar across all three gene
487 categories, but mutation rates for oncogenes and tumor suppressor genes were always higher than for
488 other genes.

Tumor type	Aberration type	Oncogenes	TSG	Other
Wilms tumor	Mutation	0.04	0.034	0.017
Wilms tumor	Methylation	0.012	0.012	0.014
Neuroblastoma	Mutation	0.083	0.089	0.046
Neuroblastoma	Methylation	0.007	0.007	0.008
Osteosarcoma	Mutation	0.031	0.034	0.018
Osteosarcoma	Methylation	0.014	0.011	0.014

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

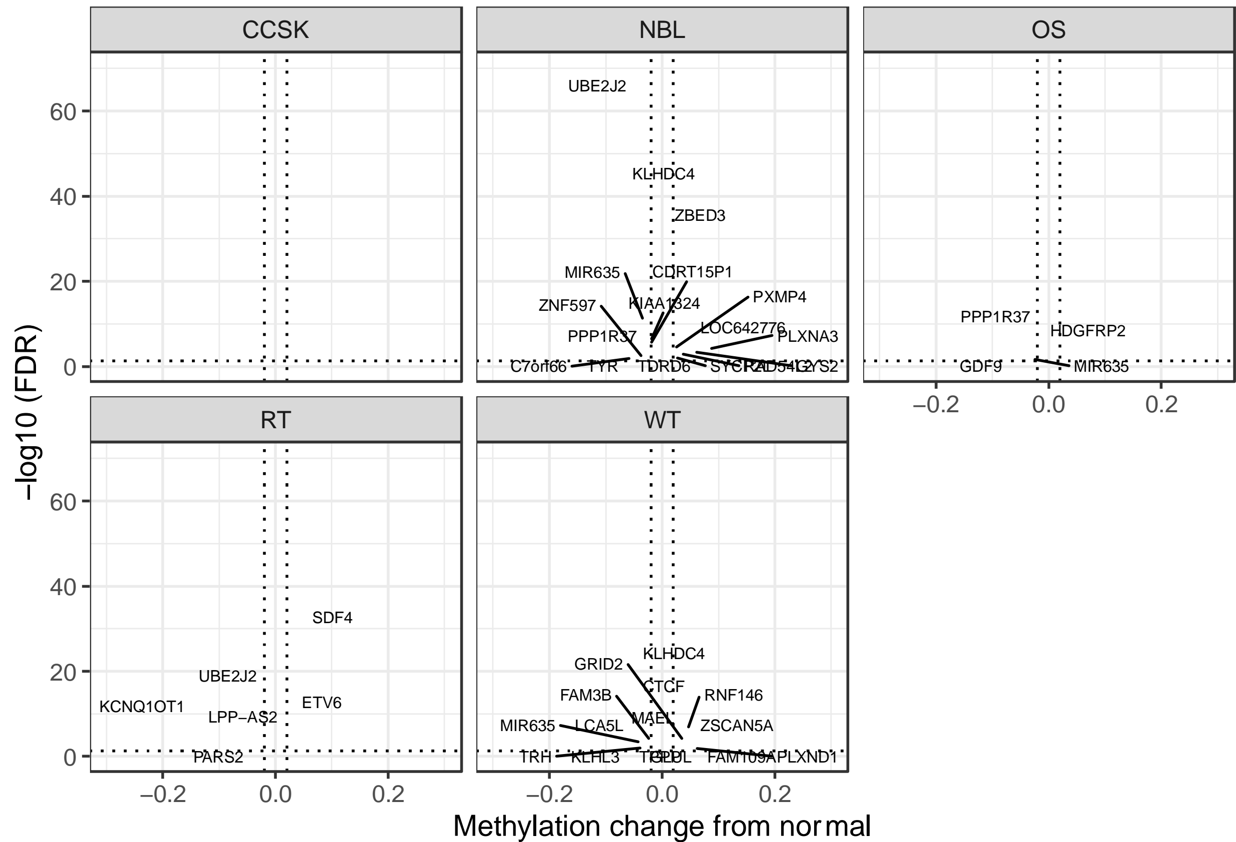
489 # Figures



490

491 **Figure 1: Consistency of DNA methylation levels and variances in normal cells.** We assigned each
492 gene to a category that indicated whether it was methylated at low, medium, or high levels and whether it
493 had low or high variance across samples in a given dataset. For each gene, we counted the maximum
494 number of datasets for which the level / variance category was consistent. For most genes, the category
495 was consistent across all four datasets.

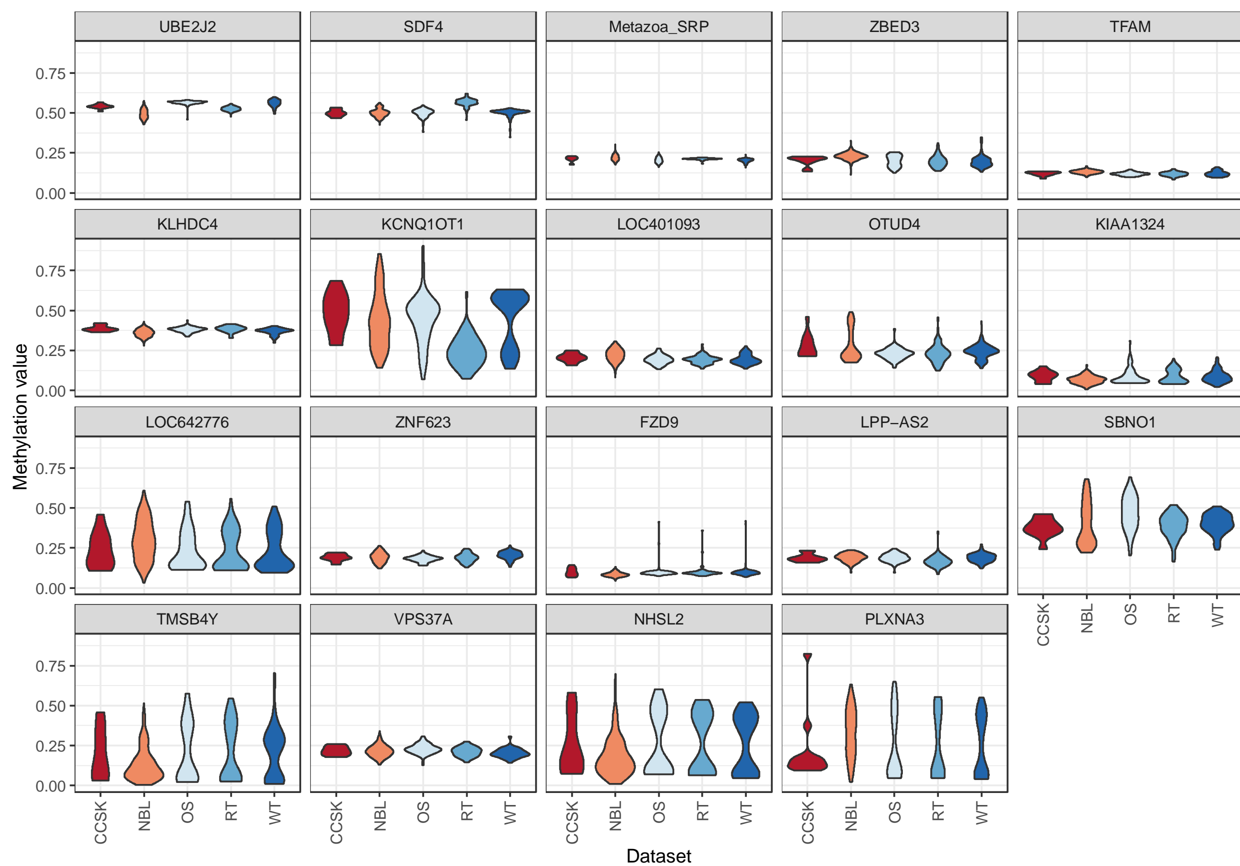
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



496

497 **Figure 2: Volcano plots showing differentially methylated genes for each tumor type.** For each tumor
498 type, we compared methylation levels at the gene level between tumors and the normal samples. Genes
499 showing significantly different methylation levels between tumor and normal conditions are highlighted.

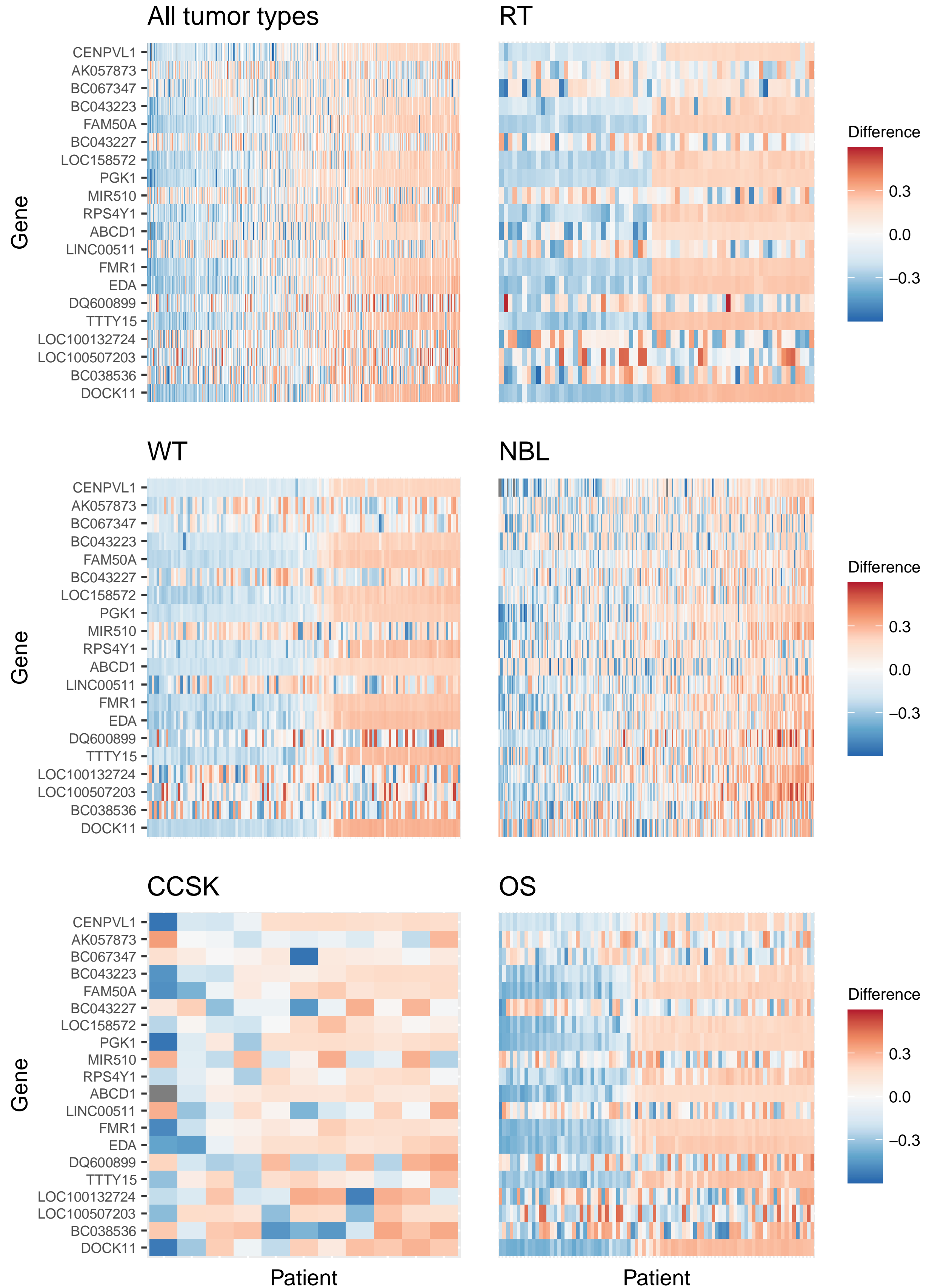
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



500

501 **Figure 3: Methylation levels for genes that differed significantly across the tumor types.** One-way
502 ANOVA tests applied to tumor methylation levels identified 19 genes for which the means differed
503 significantly across the tumor types. These violin plots show the range and density of the methylation
504 values for these genes across the tumor types.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



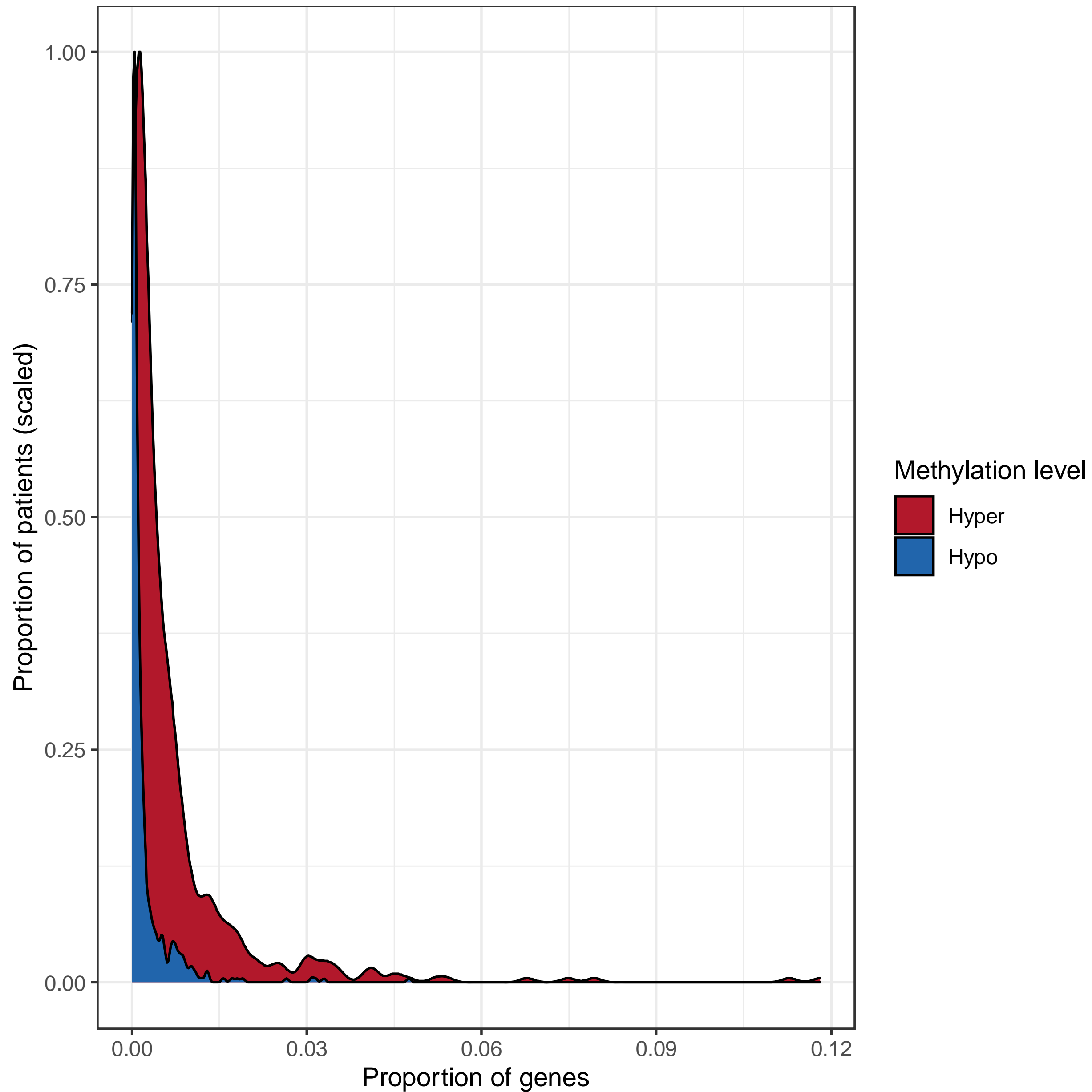
505

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

506 **Figure 4: Gene-level DNA methylation changes for high-variance genes.** Rows in these heatmaps
507 indicate methylation levels, relative to the normal data, for the 20 genes with the largest variance across
508 the tumor types. Columns represent individual tumors. Approximately half of all tumors exhibited
509 consistently lower methylation levels than the remaining tumors.

510

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



511

512 **Figure 5: Distributions of the proportion of hypermethylated or hypomethylated genes in a given**
513 **tumor.** Using the normal data as a reference, we identified genes that were hypermethylated or
514 hypomethylated in a given tumor. All five tumor types are represented. A relatively large number of
515 hypermethylated genes in a given tumor was more common than a relatively large number of
516 hypomethylated genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

We thank the research participants who donated samples for molecular profiling and the researchers who consented to have their data shared publicly. We thank the College of Life Sciences at Brigham Young University for providing funding to ACP through a College Undergraduate Research Award. The results published here are in whole or part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative, phs000218, managed by the NCI. The data used for this analysis are available from the National Cancer Institute Genomic Data Commons (https://gdc.cancer.gov). Information about TARGET can be found at http://ocg.cancer.gov/programs/target.

Ethics approval and consent to participate

Brigham Young University's Institutional Review Board approved this study under exemption status. This study uses data collected from public repositories only. We played no part in recruiting patients or obtaining consent.

Competing Interests

The authors declare that they have no competing interests.

Code and data availability

All of the code used to perform the analysis is publicly available on Open Science Framework so that others can verify and build upon our work (https://osf.io/79yfb/).

Author Contributions

The contributions listed below correspond to the CRediT Taxonomy[73].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

537 ACP: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software,
538 Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

539 BIQ: Conceptualization, Formal Analysis, Software, Visualization, Writing – Original Draft Preparation,
540 Writing – Review & Editing

541 SRP: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources,
542 Supervision, Writing – Review & Editing

543 References

544 1. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for
545 childhood cancer. *Nat. Med.* **25**, 367–376 (2019).
546 2. Oeffinger, K. C., Eshelman, D. A., Tomlinson, G. E., Buchanan, G. R. & Foster, B. M. Grading
547 of late effects in young adult survivors of childhood cancer followed in an ambulatory adult setting.
548 *Cancer* **88**, 1687–1695 (2000).
549 3. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724
550 (2009).
551 4. Yiu, T. T. & Li, W. Pediatric cancer epigenome and the influence of folate. *Epigenomics* **7**, 961–
552 973 (2015).
553 5. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**,
554 321–327 (2018).
555 6. Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy.
556 *Nature* **506**, 445–450 (2014).
557 7. Bayliss, J. *et al.* Lowered H3K27me3 and DNA hypomethylation define poorly prognostic
558 pediatric posterior fossa ependymomas. *Science Translational Medicine* **8**, 366ra161–366ra161 (2016).
559 8. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–259 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

560 9. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674
561 (2011).
562 10. Croce, C. M. Oncogenes and Cancer. *N. Engl. J. Med.* **358**, 502–511 (2008).
563 11. Macleod, K. Tumor suppressor genes. *Current Opinion in Genetics & Development* **10**, 81–93
564 (2000).
565 12. Spainhour, J. C., Lim, H. S., Yi, S. V. & Qiu, P. Correlation Patterns Between DNA Methylation
566 and Gene Expression in The Cancer Genome Atlas. *Cancer Inform* **18**, 1176935119828776 (2019).
567 13. Esteller, M. Epigenetic gene silencing in cancer: The DNA hypermethylome. *Hum. Mol. Genet.*
568 **16**, R50–R59 (2007).
569 14. Liu, J. *et al.* Identification of aberrantly methylated differentially expressed genes and associated
570 pathways in endometrial cancer using integrated bioinformatic analysis. *Cancer Med.* **9**, 3522–3536
571 (2020).
572 15. Shi, M., Tsui, S. K.-W., Wu, H. & Wei, Y. Pan-cancer analysis of differential DNA methylation
573 patterns. *BMC Medical Genomics* **13**, 154 (2020).
574 16. Dunwell, T. L. *et al.* Epigenetic analysis of childhood acute lymphoblastic leukemia. *Epigenetics*
575 **4**, 185–193 (2009).
576 17. Saghafinia, S., Mina, M., Riggi, N., Hanahan, D. & Ciriello, G. Pan-Cancer Landscape of
577 Aberrant DNA Methylation across Human Tumors. *Cell Reports* **25**, 1066–1080.e8 (2018).
578 18. Karlow, J. A., Miao, B., Xing, X., Wang, T. & Zhang, B. Common DNA methylation dynamics
579 in endometriod adenocarcinoma and glioblastoma suggest universal epigenomic alterations in
580 tumorigenesis. *Commun Biol* **4**, 1–16 (2021).
581 19. Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent
582 structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103–112 (2018).
583 20. Issa, J.-P. J. & Kantarjian, H. M. Targeting DNA Methylation. *Clin Cancer Res* **15**, 3938–3946
584 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

585 21. Yu, J. *et al.* Methylation-mediated downregulation of the B-cell translocation gene 3 (BTG3) in
586 breast cancer cells. *Gene Expr* **14**, 173–182 (2008).
587 22. Majid, S. *et al.* Genistein reverses hypermethylation and induces active histone modifications in
588 tumor suppressor gene B-Cell translocation gene 3 in prostate cancer. *Cancer* **116**, 66–76 (2010).
589 23. Majid, S. *et al.* BTG3 tumor suppressor gene promoter demethylation, histone modification and
590 cell cycle arrest by genistein in renal cancer. *Carcinogenesis* **30**, 662–670 (2009).
591 24. Almstrup, K. *et al.* Improved gene expression signature of testicular carcinoma in situ. *Int J*
592 *Androl* **30**, 292–302; discussion 303 (2007).
593 25. Reeskamp, L. F. *et al.* Differential DNA methylation in familial hypercholesterolemia.
594 *EBioMedicine* **61**, 103079 (2020).
595 26. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful
596 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
597 27. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498–D503
598 (2020).
599 28. Fabregat, A. *et al.* Reactome pathway analysis: A high-performance in-memory approach. *BMC*
600 *Bioinformatics* **18**, (2017).
601 29. Futreal, P. A. *et al.* A CENSUS OF HUMAN CANCER GENES. *Nat Rev Cancer* **4**, 177–183
602 (2004).
603 30. Goode, E. L., Ulrich, C. M. & Potter, J. D. Polymorphisms in DNA Repair Genes and
604 Associations with Cancer Risk. *Cancer Epidemiol Biomarkers Prev* **11**, 1513–1530 (2002).
605 31. Forrest, S. J., Geoerger, B. & Janeway, K. A. Precision medicine in pediatric oncology. *Curr*
606 *Opin Pediatr* **30**, 17–24 (2018).
607 32. Bodily, W. R. *et al.* Effects of germline and somatic events in candidate BRCA-like genes on
608 breast-tumor signatures. *PLoS One* **15**, (2020).
609 33. Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies
610 oncogenic network modules. *Genome Res* **22**, 398–406 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

611 34. Babur, Ö. *et al.* Systematic identification of cancer driving signaling pathways based on mutual
612 exclusivity of genomic alterations. *Genome Biol* **16**, 45 (2015).
613 35. Kang, S. *et al.* Mutual exclusiveness between PIK3CA and KRAS mutations in endometrial
614 carcinoma. *Int J Gynecol Cancer* **18**, 1339–1343 (2008 Nov-Dec).
615 36. Szczurek, E. & Beerenwinkel, N. Modeling mutual exclusivity of cancer mutations. *PLoS*
616 *Comput Biol* **10**, e1003503 (2014).
617 37. Ding, W., Feng, G., Hu, Y., Chen, G. & Shi, T. Co-occurrence and Mutual Exclusivity Analysis
618 of DNA Methylation Reveals Distinct Subtypes in Multiple Cancers. *Front Cell Dev Biol* **8**, 20 (2020).
619 38. Tate, J. G. *et al.* COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*
620 **47**, D941–D947 (2019).
621 39. Keelawat, S. *et al.* Detection of global hypermethylation in well-differentiated thyroid neoplasms
622 by immunohistochemical (5-methylcytidine) analysis. *J Endocrinol Invest* **38**, 725–732 (2015).
623 40. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *PNAS* **96**, 8681–8686
624 (1999).
625 41. Segura-Pacheco, B. *et al.* Global DNA hypermethylation-associated cancer chemotherapy
626 resistance and its reversion with the demethylating agent hydralazine. *Journal of Translational Medicine*
627 **4**, 32 (2006).
628 42. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143–153
629 (2004).
630 43. Shlien, A. & Malkin, D. Copy number variations and cancer. *Genome Medicine* **1**, 62 (2009).
631 44. Ohlsson, R., Renkawitz, R. & Lobanenkov, V. CTCF is a uniquely versatile transcription
632 regulator linked to epigenetics and disease. *Trends in Genetics* **17**, 520–527 (2001).
633 45. Wiehle, L. *et al.* DNA (de)methylation in embryonic stem cells controls CTCF-dependent
634 chromatin boundaries. *Genome Res* **29**, 750–761 (2019).
635 46. Kemp, C. J. *et al.* CTCF Haploinsufficiency Destabilizes DNA Methylation and Predisposes to
636 Cancer. *Cell Reports* **7**, 1020–1029 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

637 47. Oh, S., Oh, C. & Yoo, K. H. Functional roles of CTCF in breast cancer. *BMB Rep* **50**, 445–453
638 (2017).
639 48. Roberti, A., Valdes, A. F., Torrecillas, R., Fraga, M. F. & Fernandez, A. F. Epigenetics in cancer
640 therapy and nanomedicine. *Clinical Epigenetics* **11**, 81 (2019).
641 49. Cheishvili, D., Boureau, L. & Szyf, M. DNA demethylation and invasive cancer: Implications for
642 therapeutics. *Br J Pharmacol* **172**, 2705–2715 (2015).
643 50. Dedeurwaerder, S. *et al.* A comprehensive overview of Infinium HumanMethylation450 data
644 processing. *Brief Bioinform* **15**, 929–941 (2014).
645 51. Price, E. M. *et al.* Profiling placental and fetal DNA methylation in human neural tube defects.
646 *Epigenetics Chromatin* **9**, 6 (2016).
647 52. Yang, I. V. *et al.* The nasal methylome and childhood atopic asthma. *J Allergy Clin Immunol* **139**,
648 1478–1488 (2017).
649 53. Zhang, X. *et al.* Nasal DNA methylation is associated with childhood asthma. *Epigenomics* **10**,
650 629–641 (2018).
651 54. van Dijk, S. J. *et al.* Effect of prenatal DHA supplementation on the infant epigenome: Results
652 from a randomized controlled trial. *Clin Epigenetics* **8**, 114 (2016).
653 55. Oelsner, K. T., Guo, Y., To, S. B.-C., Non, A. L. & Barkin, S. L. Maternal BMI as a predictor of
654 methylation of obesity-related genes in saliva samples from preschool-age Hispanic children at-risk for
655 obesity. *BMC Genomics* **18**, 57 (2017).
656 56. Armstrong, D. A., Lesseur, C., Conradt, E., Lester, B. M. & Marsit, C. J. Global and gene-
657 specific DNA methylation across multiple tissues in early infancy: Implications for children's health
658 research. *FASEB J.* **28**, 2088–2097 (2014).
659 57. Godderis, L. *et al.* Global Methylation and Hydroxymethylation in DNA from Blood and Saliva
660 in Healthy Volunteers. *BioMed Res. Int.* **2015**, e845041 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

661 58. Smith, A. K. *et al.* DNA extracted from saliva for methylation studies of psychiatric traits:
662 Evidence tissue specificity and relatedness to brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168**,
663 36–44 (2015).
664 59. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of
665 Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
666 60. Amezquita, R. A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**,
667 137–145 (2020).
668 61. Price, M. E. *et al.* Additional annotation enhances potential for biologically-relevant analysis of
669 the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
670 62. Wickham, H. *Rvest: Easily harvest (scrape) web pages*. (2020).
671 63. Warnes, G. R., Bolker, B. & Lumley, T. *Gtools: Various r programming tools*. (2020).
672 64. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for
673 removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**,
674 882–883 (2012).
675 65. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**,
676 1109–1112 (2016).
677 66. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279–284
678 (2013).
679 67. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic
680 landscape of Wilms tumor. *Nat Genet* **49**, 1487–1494 (2017).
681 68. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019)
682 doi:10.1101/861054.
683 69. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
684 70. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for
685 genomes. *Nat. Protoc.* **11**, 1–9 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.02.454814; this version posted August 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

686 71. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense
687 Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1–7.20.41 (2013).
688 72. Haas, B. J. *et al.* Accuracy assessment of fusion transcript detection via read-mapping and de
689 novo fusion transcript assembly-based methods. *Genome Biology* **20**, 213 (2019).
690 73. Brand, A., Allen, L., Altman, M., Hlava, M. & Scott, J. Beyond authorship: Attribution,
691 contribution, collaboration, and credit. *Learn. Publ.* **28**, 151–155 (2015).