



The Pancancer DNA Methylation Trackhub: A Window to The Cancer Genome Atlas Epigenomics Data

Izaskun Mallona, Alberto Sierco, and Miguel A. Peinado

Abstract

The Cancer Genome Atlas (TCGA) epigenome data includes the DNA methylation status of tumor and normal tissues of large cohorts for dozens of cancer types. Due to the moderately large data sizes, retrieving and analyzing them requires basic programming skills. Simple data browsing (e.g., candidate gene search) is hampered by the scarcity of easy-to-use data browsers addressed to the broad community of biomedical researchers. We propose a new visualization method depicting the overall DNA methylation status at each TCGA cohort while emphasizing its heterogeneity, thus facilitating the evaluation of the cohort variability and the normal versus tumor differences. Implemented as a trackhub integrated to the University of California Santa Cruz (UCSC) genome browser, it can be easily added to any genome-wide annotation layer.

To exemplify the trackhub usage we evaluate local DNA methylation boundaries, the aberrant DNA methylation of a CpG island located at the estrogen receptor 1 (ESR1) in breast and colon cancer, and the hypermethylation of the Homeobox HOXA gene cluster and the EN1 gene in multiple cancer types. The DNA methylation pancancer trackhub is freely available at http://maplab.cat/tega_450k_trackhub.

Key words DNA methylation, Pancancer, Data visualization, TCGA, The Cancer Genome Atlas

1 Introduction

Cancer research is being benefited by the unprecedented availability of multilayered molecular data gathered by international consortia. Namely, The Cancer Genome Atlas (TCGA) [1] offers a wide assortment of genomics and epigenomics data for more than 30 cancer types, thus facilitating hypothesis generation and testing. The genome-wide quality of the data and hence its big size might however set back browsing individual candidates. This gap is largely due to the lack of specific tools to extract and explore parts of the data; noteworthy, this is the case for experimentalists or clinicians, who are interested in candidate-based queries rather than in accessing whole-genome datasets at once [2]. Although there are many refined methods of data mining that generate valuable information

from these big datasets [3, 4], interpretation by biological and clinical domain experts is still a major way to retrieve basic and applicable knowledge from them.

(Epi)genomics data are still lacking tools to be effectively displayed, integrating both the user-friendliness and the big biodatasets mining. For instance, a common gateway to genomic data is genome browsers, which offer chromosome views with some annotated layers over it. The University of California Santa Cruz (UCSC) genome browser is widely used in (epi)genomic data visualization: it renders a genomic region offering an overview of selected datasets. Among the data included there are transcript and genes models, regulation, expression, epigenetics, disease association, variation, and evolutionary pairwise genomic alignments [5]. Such availability makes the tool very insightful for visual correlation. The main drawback is that data are usually displayed in a single window, limiting the maximum number of tracks. Thus data tracks are mostly useful when they summarize models or generalizations, rather than individual samples.

This chapter describes a pancancer set of DNA methylation tracks for the UCSC Genome Browser to visualize datasets with dozens to thousands of patients in a compact manner while depicting their inner variability and differences between tumor versus normal tissue (Fig. 1a). To exemplify the trackhub usability we display examples of local DNA methylation boundaries between CpG islands and their flanking sequences, the aberrant DNA methylation of the estrogen receptor 1 (ESR1) in breast and colon cancer and the hypermethylation of the Homeobox HOXA cluster and the EN1 gene in multiple cancer types.

2 Trackhub Generation

We fetched the open access tier of highly processed DNA methylation data (TCGA level 3) of over 450,000 genomic locations as measured by the Illumina's Infinium array [6] (Table 1; see Notes 1–3). As the methylation metadata refers to biospecimens stratified by cancer cohort and sample type, we retrieved those corresponding to solid primary tumors and matched normals (sample codes 01 and 11, respectively). Data was downloaded using TCGA-Assembler v1.0.3 [7].

Theoretically, the methylation status of a given CpG is rather binary, being methylated or unmethylated; however, some cells might have an intermediate state, with the two epialleles present (imprinted). Strikingly, however, nonimprinted loci often follow a continuous bimodal distribution ranging from full unmethylation to complete methylation (beta values from 0 to 1) [6]. This is mainly due to the mixture of cell populations subjected to the measurement of the CpG methylation (except for imprinted

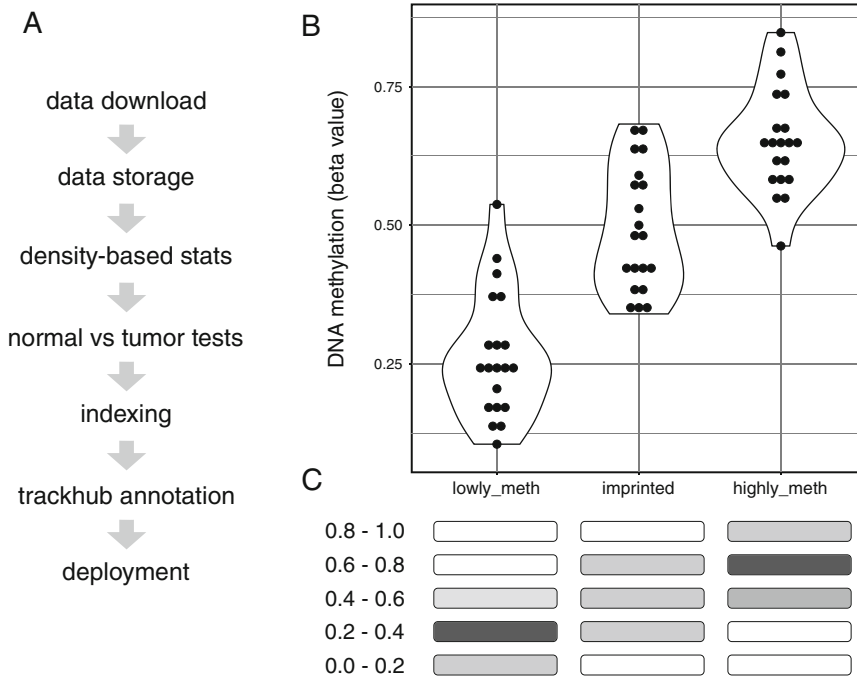


Fig. 1 Pancancer trackhub. (a) Data flow chart. (b) DNA methylation status for three loci (lowly methylated, imprinted and highly methylated) across a cohort (each point represents a sample) as depicted by violin plots. (c) Track-based representation of the same series summarizing cohort heterogeneity

sites). The DNA methylation status of numerous genomic regions, however, is conserved across samples belonging to the same tissue and physiological or pathological condition. Therefore, visualizing the DNA methylation barcode of a set of samples aid to interpret and/or infer their inner complexity. To do so we have developed a new visualization of the DNA methylation status in which the abundance of samples in a low, intermediate and high DNA methylation status is shown for each genomic position (Fig. 1b, c).

We represent a 5-row stacked track (namely, composite track) in which the methylation status is split and sorted into its layers from almost full methylation (beta value of 0.8–1) on top to minimal methylation (beta 0–0.2) on bottom, with three rows in between depicting intermediate methylation statuses (betas 0.8–0.6, 0.6–0.4 and 0.4–0.2). For each CpG, the color at each track reflects the proportion of samples which show that methylation status; such score ranges from 0 to 1000, with 1000 being the 100% of the data (Fig. 1c).

Apart from visualizing the methylation profiles, we generated direct comparisons between normal and tumor for cohorts in which both datasets are present. To statistically assess such differences, we performed independent nonparametric Wilcoxon rank sum tests on each CpG in which the beta value showed an absolute difference of at

Table 1
DNA methylation dataset sizes

Description	Code	Normal tissue	Tumor tissue
Adrenocortical carcinoma	ACC	0	80
Bladder urothelial carcinoma	BLCA	21	358
Brain lower grade glioma	LGG	0	511
Breast invasive carcinoma	BRCA	98	743
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	3	256
Colon adenocarcinoma	COAD	38	302
Esophageal carcinoma	ESCA	16	185
Glioblastoma multiforme	GBM	2	129
Head and neck squamous cell carcinoma	HNSC	50	528
Kidney chromophobe	KICH	0	66
Kidney renal clear cell carcinoma	KIRC	160	324
Kidney renal papillary cell carcinoma	KIRP	45	226
Liver hepatocellular carcinoma	LIHC	50	256
Lung adenocarcinoma	LUAD	32	463
Lung squamous cell carcinoma	LUSC	43	361
Lymphoid neoplasm diffuse large B + AC0-cell lymphoma	DLBC	0	48
Mesothelioma	MESO	0	37
Ovarian serous cystadenocarcinoma	OV	0	10
Pancreatic adenocarcinoma	PAAD	10	146
Pheochromocytoma and paraganglioma	PCPG	3	179
Prostate adenocarcinoma	PRAD	49	340
Rectum adenocarcinoma	READ	7	98
Sarcoma	SARC	4	242
Skin cutaneous melanoma	SKCM	2	92
Stomach adenocarcinoma	STAD	2	339
Thyroid carcinoma	THCA	56	507
Uterine carcinosarcoma	UCS	0	57
Uterine corpus endometrial carcinoma	UCEC	46	438
Uveal melanoma	UVM	0	80

Tumors depict the number of primary tumors samples scrutinized. Normal tissue refers to adjacent to the primary tumor; note that some cohorts lack the adjacent normal counterpart and therefore will lack the statistics track

least 0.2 [6]. Tests did not assume normal-tumor pairing and compared the whole distribution of normal against that of tumor tissues (two-tailed, null hypothesis: there are no differences between normal and tumor samples). Significance p -value cutoff was set to 0.001 ($p < 0.001$ there are detectable differences between normal and tumor samples). To ease visualization we depicted no differences ($p \geq 0.001$) in gray, tumor hypermethylations in red, and tumor hypomethylations in blue (*see* **Notes 4** and **5**).

3 Querying and Cohort Selection

UCSC browser can be launched by entering http://maplab.cat/tcga_450k_trackhub from any Web browser. This will render a default genomic location, which can be readily changed to another region of interest by typing a position (i.e., chr6:152,107,634–152,148,574), a gene symbol (e.g., GLDC) or Illumina Infinium probe name (e.g., cg16029534) (*see* **Notes 3–9**). After pressing the “go” button the browser will point to the new location (zooming in and out might be necessary).

The setup panel named “Pan-cancer DNA methylation profiling” located below the track window allows to select the cohort to focus on. Due to the abundance of data, most cohorts are hidden by default, and manual selection of a target cohort is required. To start exploring the Bladder Urothelial Carcinoma (BLCA) dataset just click the second drop down menu (named “blca”) and select “show.” Then, click to the “refresh” button at the upper right corner of the setup panel (*see* **Note 4**). This will render a composite track describing the DNA methylation distribution across tumor samples, another equivalent for the matching normal; and a simple track for the normal vs. tumor statistical comparison.

4 Data Visualization and Interpretation

Both the normal and the tumor composite tracksets are composed of a stack of five tracks depicting the abundance of samples belonging to five strata of DNA methylation, from nearly fully methylated (top track, named “a_blca_1.0”, with beta values between 0.8 and 1) to almost totally unmethylated (bottom track, named “e_blca_0.2”, with beta values between 0 and 0.2), with three tracks of intermediate methylation between them. The shade of gray at each track depicts the abundance of samples at that level of DNA methylation: black represent that all samples belong to the DNA methylation interval; and white, none (*Fig. 1*).

Differences in DNA methylation between tumor and normal samples are depicted for those cohorts where both datasets exist

(Table 1). The statistics track simply depicts in grey the scrutinized CpGs; and tumor hypomethylation in blue and tumor hypermethylation in red.

The tracks depicting the DNA methylation distribution are designed to be visualized as dense tracks by default, so they arrange all the DNA methylation data inline. This display mode can be changed in order to expand the methylation data to multiple rows. Taking the BLCA normal dataset as an example, the display mode can be updated by right-clicking to the gray vertical bar on their left at the browsing panel (“configure Methylation in blca normal track set”). The settings page present a drop-down named “display mode” which permits bulk updates (i.e., expanding or collapsing many datasets at once). If desired, each track can be fine tuned (e.g., hidden) independently. Clicking to “submit” after selecting the proper visualization method will save the changes and reload the browser window with the updated display mode (*see* **Notes 4** and **5**).

4.1 Pancancer Tumor vs. Normal Evaluation

The stacked DNA methylation density tracks allow to visually correlate hypermethylation and hypomethylation. Independently, the statistics track simplifies the visualization as it summarizes in a single line the significant changes. To perform a pancancer analysis evaluating the statistical differences for a given region across cohorts (i.e., in bulk), the configure button (below the browsing window) launches a form for the “Pan-cancer DNA methylation profiling” in which all the statistics tracks display options must be set to “dense” (they appear as “hidden” by default). Then, clicking the “submit” button at the top will save the changes and draw an updated browsing window in which shared and divergent patterns can be easily detected (as in Figs. 2, 3, 4, and 5).

4.2 Candidate CpGs Numeric Data Acquisition

The pancancer trackhub renders the DNA methylation distribution with a grayscale color palette; and, for cohorts with both normal and tumor data, with an extra track testing hypermethylation and hypomethylations. The user, however, can extract the numeric data of both types of representation (for instance, the p -value of the statistical test).

To do so, if starting from the statistical test track, a two-step procedure must be followed. First, clicking to any CpG from a default densely stacked track will force the browser to render a “full” display version of the track, representing each individual CpG separately. Then, if clicking at a given CpG, a new Web page depicting the mean DNA methylation difference between normal and tumor will be generated. This page includes extra information, such as the Wilcoxon (Mann–Whitney) test p -value and statistic (two tailed, unpaired) and details about the CpG, such as the cytoband or genomic coordinate. The DNA sequence of the region (which can be extended up and downstream according to the user

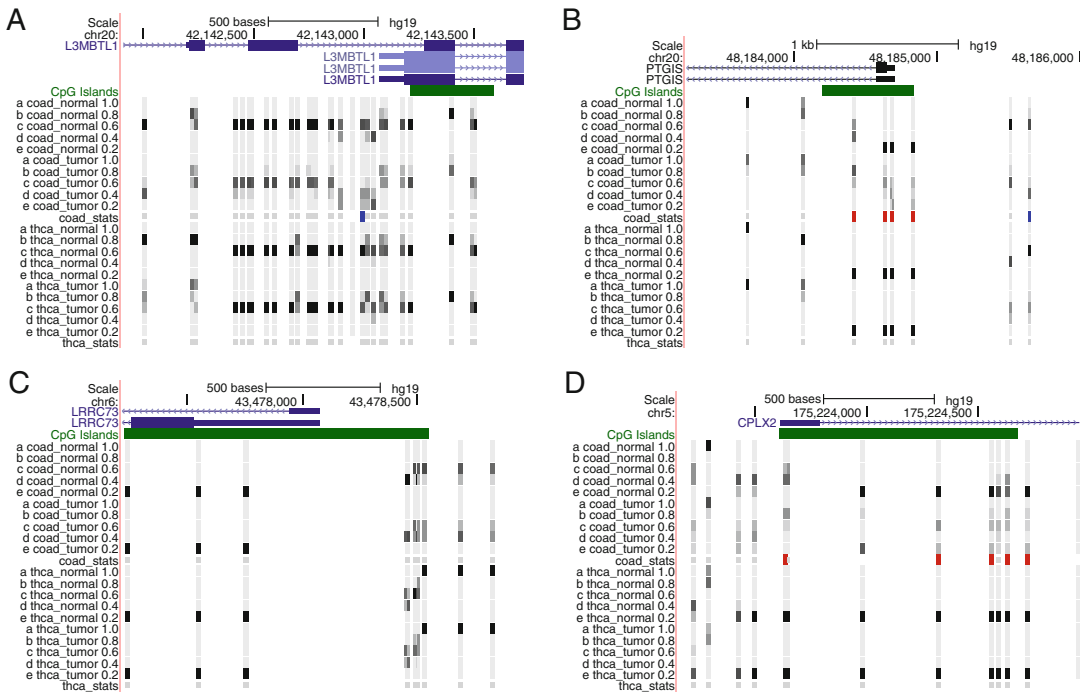


Fig. 2 Four DNA methylation landscapes at the local level. (a) Paternally imprinted region in the L3MBTL1 gene, in which beta values are at the 0.4–0.6 range (middle track). (b–d) Different types of DNA methylation boundaries in CpG islands and their flanking regions in thyroid and colon tissues. Track descriptions have been omitted

request) can be readily downloaded by clicking “view DNA for this feature.”

The same two-step procedure can be applied to the DNA methylation profile composite track. In this case, a full description of the CpG including its score, represented as the percentile of samples at the CpG range multiplied by a factor of 10, will be rendered. That is, if 50% of the samples are located at that given DNA methylation range the score of the CpG will be 500.

4.3 Data Sharing

A raster image of the browsing window can be exported at any time by right-clicking at any point of the visualization window and selecting “view image”. To save a high-resolution, vector image the user must click to the “View” menu at the blue navigation bar at the top of the screen (“PDF/PS” link).

As opposed to taking just a static image of the browsing window, the sessions tool allows to save a snapshot of the location and a full set of annotation tracks, including the overall aspect (display settings, track order, and zoom level, among others). This snapshot can be launched and browsed interactively. To do so, the sessions tool <https://genome.ucsc.edu/cgi-bin/hgSession> allows to download a file with the session details (“Save Settings” section).

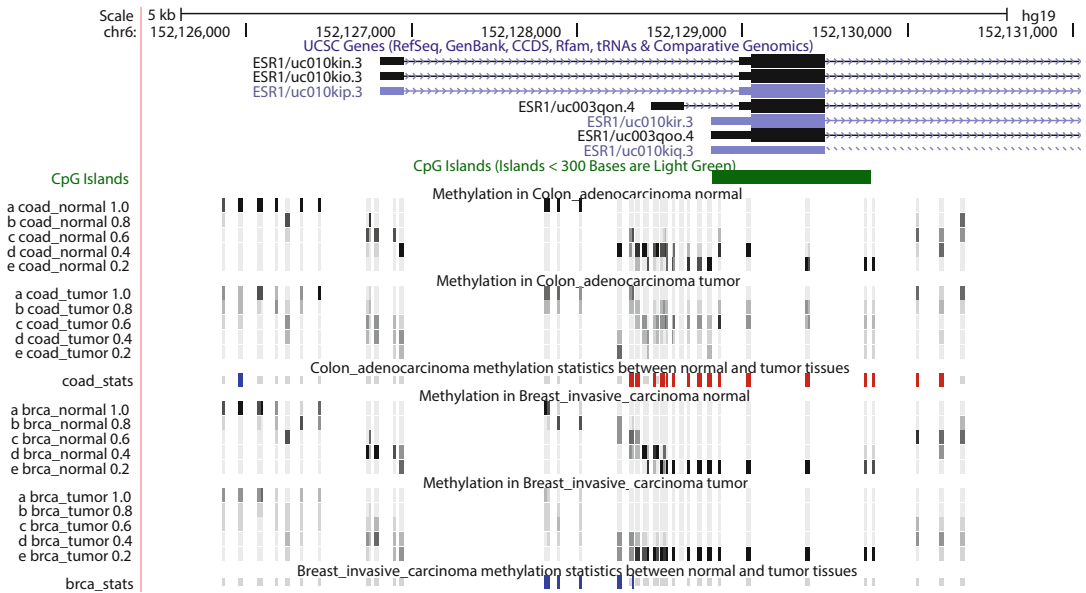


Fig. 3 DNA methylation landscape of the ESR1 gene illustrating hypermethylation in colon tumor (red ticks in coad_stats) and hypomethylation of the 5' region upstream of the CpG island in breast cancer (blue ticks at brca_stats)

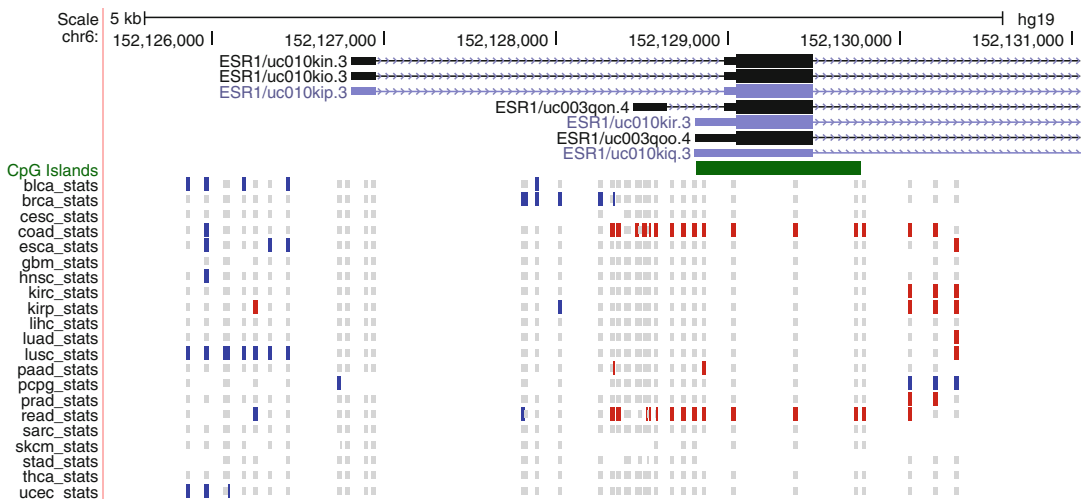


Fig. 4 Pancancer DNA methylation landscape of differential methylation in the ESR1 gene illustrating hypermethylation and hypomethylation of the 5' region upstream of the CpG island in different tumor types. Track descriptions have been omitted

The upload of such a file will result in genome browser configured exactly as it was when the session was saved. Users can avoid the need to download a configuration file by registering to the UCSC



Fig. 5 (a) Pancancer visualization of the local aggregation of significantly hypermethylated CpGs (red ticks) at the HOXA cluster. Despite of the fact that differential methylation between normal and tumor samples are plotted independently, long-range effects can be easily detected. Track descriptions have been omitted. **(b)** Pancancer visualization of the local aggregation of significantly hypermethylated CpGs (red ticks) at the CpG island cluster neighboring the EN1 gene. Track descriptions have been omitted

(free, requires a contact email, username and password). After registering, the UCSC browser permits to store and browse sessions for private use, as well to share them with other investigators.

5 Usage Examples

The pancancer DNA methylation trackhub depicts DNA methylation variability for discrete CpG loci. Independently, it also provides a statistically assessed straightforward tumor vs. normal comparison. Despite of representing genomic positions, the trackhub can be used at different scales, from the fine-grained nucleotide level to regions of hundreds of genes. To help the local use case, the trackhub adaptively resizes the width of each CpG to effectively display highly packed areas; and, when zooming out, it increases the

thickness of the bars depicting the statistically significant changes between normal and tumor samples to emphasize them.

To exemplify the trackhub usage versatility we briefly describe three use cases, from the closest to the nucleotide level to gene clusters.

5.1 Fine-Grained CpG Evaluation

Despite of the preponderant bimodal distribution of DNA methylation beta values, with CpGs being either methylated or unmethylated, DNA methylation landscapes are zonally heterogeneous in both mean values (e.g., the methylation status) and variances (e.g., the heterogeneity of methylation levels within the cohort). Therefore, different signatures can be detected by visual inspection of candidate regions, which might be interpreted in terms of regulatory potential. For instance, most samples display intermediate methylation values (0.4–0.6 beta values range, central density track) at the L3MBTL1 locus, a Polycomb family member located in a region of chromosome 20 and imprinted in human, indicating hemimethylation and low variability (Fig. 2a).

On the other hand, extreme methylation statuses are predominant. Namely, CpG islands are often fully unmethylated whereas other compartments, such as repetitive elements, are methylated, in normal tissue. Nonetheless, transitions between unmethylated CpG islands and their methylated flanked sequences can produce both sharp and sloping boundaries. For instance, in the downstream region of the PTGIS gene associated CpG island thyroid tissues display a blunt boundary with most samples in either the 0–0.2 and 0.8–1 beta values range, while colon samples show a gradual DNA methylation profile (Fig. 2b). Another interesting example is the upstream region of the LRRC73 gene CpG island that is fully methylated in thyroid samples but shows partial methylation in colon tissues, in which the boundary is not so precisely set, indicating variability at the cohort level (Fig. 2c). Finally, the CPLX2 gene promoter CpG island also displays a tight profile in thyroid tissues but a broad variability in colon with a hypermethylation trend in tumor samples (Fig. 2d). As a whole, the differential DNA methylation profiles pinpoint potential diversity in the genomic regulation of the examined genes.

5.2 Pancancer Close-Up Comparison

The aberrant methylation of the estrogen receptor 1 (ESR1) gene drives hormone insensitivity in breast cancer and has been reported for other malignancies [8, 9]. According to the UCSC gene annotation (“knownGene” table), ESR1 encodes for 27 isoforms, including uc003qon.4 (NM_001122740; transcript variant 2 translated to P03372), whose transcription start site lies near a CpG island which is extensively scrutinized by the Infinium 450k chip.

In breast cancer and matched normal samples (BRCA cohort) the CpG island is almost completely unmethylated, with all the

samples concentrated at the 0–0.2 beta values range (the density track at the bottom of each trackset; Fig. 3). The upstream flanking region, however, points to a significant tumor hypomethylation, as the profile of proximal CpGs to the island displace the samples located at almost full methylation in normal tissues (0.8–1 range, density track at the top of the trackset) to an almost continuous representation at intermediate methylation statuses (i.e., all the tracks from the trackset with a uniform shade of gray). This tumor hypomethylation reached statistical significance according to the Mann–Whitney test (blue ticks at the stats track; Fig. 3).

In colon cancer, the CpG island and its flanks are reshaped in a different manner (COAD cohort; Figs. 3 and 4). In this case, the CpG island gets significantly hypermethylated (red ticks at the stats track), while non detectable changes occur at its 5', where most breast tumor did hypomethylated. Noteworthy, different DNA methylation changes at the ESRI CpG island and its flanking region can be found across the different cohorts (Fig. 4), pointing to the locus propensity to dysregulation and its putative functional relevance. On the other hand, colon and rectum carcinomas show a highly consistent island hypermethylation pattern, in line with their clinicopathological similarities.

5.3 *Pancancer Regional Landscapes*

Discrete genomic loci repression by means of de novo methylation of CpG islands (epigenetic silencing) is a hallmark of human cancer. Remarkably, local CpG island hypermethylation can spread to megabase-sized blocks, altering the epigenetic landscape of large genomic regions whose gene expression gets disrupted (long-range epigenetic gene silencing [10]). Among the long range coordinated events in multiple cancer types, hypermethylations of the Homeobox clusters at chromosomes 2, 7, 12, and 17 are widespread [11].

Even though the trackhub is built upon the methylation status of discrete CpGs, the statistics tracks effectively display long range phenomena when zooming out. Namely, the cluster HOXA depicts a consistent tumor vs normal hypermethylation in an almost 100 KB range for most of cancer types, with the higher abundance of differentially methylated positions for lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) to the lower for prostate adenocarcinoma (PRAD) and thyroid carcinoma (THCA) (Fig. 5a, red ticks indicate tumor hypermethylation).

Another interesting example illustrating the usefulness of this tool refers to the hypermethylation of the EN1 CpG island, that has been postulated as a biomarker with diagnostic and prognostic applications in colorectal cancer [12]. A quick view of the EN1 CpG island cluster differential DNA methylation panorama in other tumor types (Fig. 5b) may aid the selection of the most appropriate loci to be screened in each specific tumor type.

6 Notes

1. The trackhub corresponds to data from human assembly GRCh37 (hg19).
2. TCGA produces DNA methylation readouts with a standard pipeline which includes quality checking and normalization, making final results comparable across datasets at their highest processing level, which corresponds to level 3. Data processing includes the removal of ambiguous readouts or data matching to regions with low mappability, such as most of the genomic repetitive elements.
3. The Infinium chip evaluates the DNA methylation status of individual cytosines (Cs) in CpG contexts, which are scattered along the genome, though particularly abundant in CpG islands [6]. Given the difficulty of visualizing 2-bp long features immersed in the roughly 3000 million basepairs-long human genome, we expanded each CpG 10 bp up and downstream, thus increasing its effective size. When CpG lied close together, we adaptively resized the intervals to accommodate them (*see Note 8*).
4. Unsorted or messed DNA methylation tracks (i.e., after manual drag and dropping, or showing different cohorts interleaved) can be recovered to the default sorting by pressing the “Default track order” link located at the “View” button on the blue bar on top of the page; or by clicking to the “default order” below the browsing window.
5. To simplify the display of pancancer analysis and increase its readability, track descriptions (the text over each trackset) and guides (the fade blue vertical lines) can be disabled by pressing the “Configure” button below the browsing window and then unchecking both “Display description above each track” and “Show light blue vertical guidelines.”
6. An useful track summarizing all the human CpG islands is available at the UCSC’s “Regulation trackset.” This depicts all CpG Islands over 300 bp in green and the rest in light green.
7. The UCSC Genome Browser described here corresponds to version 341; button names/locations can vary between different releases.
8. The procedure to build the trackhub can be generalized to any collection of samples evaluated by the Illumina’s Infinium array. The source code to generate the trackhub (starting from a populated postgres database) can be accessed at https://bitbucket.org/imallona/pancancer_dnameth_trackhub.

9. The pancancer trackhub can be disabled by clicking “disconnect” at the top-right corner of its configuration box, just below the browsing window.

Acknowledgments

We thank Iñaki Martínez de Ilarduya for his excellent technical support. The trackhub published here is based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. This work was supported by the Spanish Ministry of Economy and Competitiveness [SAF2011/23638 and SAF2015-64521-R to M.A.P.]. CERCA Program/Generalitat de Catalunya.

References

1. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M, Yao L, Kasprzyk A (2011) International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford)* 2011: bar026
2. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N (2013) Visualizing multidimensional cancer genomics data. *Genome Med* 5(1):9
3. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5):401–404
4. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J (2013) Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. *Sci Rep* 3:2652
5. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, Heitner S, Harte RA, Haussler M, Gurusvadoo L, Fujita PA, Eisenhart C, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res* 44 (D1):D717–D725
6. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98 (4):288–295
7. Zhu Y, Qiu P, Ji Y (2014) TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat Methods* 11(6):599–600
8. Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8(4):286–298
9. Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. *Cancer Res* 63 (22):7641–7645
10. Frigola J, Song J, Storzaker C, Hinshelwood RA, Peinado MA, Clark S (2006) Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat Genet* 38(5):540–549
11. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, Kernstine KH, Riggs AD, Pfeifer GP (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A* 104 (13):5527–5532
12. Mayor R, Casadome L, Azuara D, Moreno V, Clark SJ, Capella G, Peinado MA (2009) Long-range epigenetic silencing at 2q14.2 affects most human colorectal cancers and may have application as a non-invasive biomarker of disease. *Br J Cancer* 100 (10):1534–1539