BIOENGINEERED
2017, VOL. 8, NO. 2, 182–190
http://dx.doi.org/10.1080/21655979.2016.1218582

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS

DMAK: A curated pan-cancer DNA methylation annotation knowledgebase

Binhua Tang[a,b]

[a]Epigenetics and Function Group, School of Internet of Things, Hohai University, Jiangsu, China; [b]School of Public Health, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

Pan-cancer analysis can identify cell- and tissue-specific genomic loci and regions with underlying biological functions. Here we present an online curated DNA Methylation Annotation Knowledgebase, DMAK, which includes the pan-cancer analysis results for differentially-methylated loci and regions by the Reduced Representation Bisulfite Sequencing profiling technology. DMAK contains 3 modules of curated information and analysis results on 688,445 CpG sites across 19 cancer and embryonic stem cell lines from ENCODE, and further analysis of survival associations with clinical sources retrieved from TCGA. The knowledgebase covers all identified differentially-methylated CpG sites and regions of interest, further annotated genomic information, together with tumor suppressor genes information and calculated methylation level. DMAK provides meaningful clues for deriving functional association network and related clinical association results based on protein-coding genes, including tumor suppressor genes, identified from differentially methylated regions of interest. Thus DMAK constitutes a comprehensive reference source for the current epigenetic research and clinical study.

ARTICLE HISTORY
Received 19 July 2016
Revised 26 July 2016
Accepted 26 July 2016

KEYWORDS
differential analysis; DNA methylation; genomic annotation; pan-cancer

Introduction

Pan-cancer analysis can uncover cell- and tissue-specific genomic loci and regions with underlying biological functions of interest.[1-6] Meanwhile, it can provide meaningful insights by genome-wide interrogation and cross-cell genetic annotation.

Especially for the topics of public research consortiums, ENCODE (Encyclopedia Of DNA Elements), focusing on identifying all functional elements in the human genome sequence[7-9]; and TCGA (The Cancer Genome Atlas), providing comprehensive and multi-dimensional maps of the key genomic changes in 33 types of cancer.[1,5,10] Pan-cancer analysis on the consortium resources can unveil the molecular basis of cancer through genome-wide interrogation and deep learning.

While till now, due to data size and technique barrier, there is no comprehensive reference source for wet-lab experiment design and post-experiment validation purposes. Thus, this is an imperative for most biologists and biomedical researchers to improve their research output and efficiency.[11,12]

Here we present an online curated reference source for DNA methylation annotation and analysis purposes. The information knowledgebase provides multiple read-to-use analysis results and annotation information for pan-cancer interrogation and cross-validation usages.

For the first time, our work attempts to provide a rapid but thorough reference to the epigenetic research fields. Thus we deposit the curated information knowledgebase online for direct and interactive usage.

Structure and function of DMAK

In summary, DMAK contains 3 modules of curated information across 19 cell types retrieved from ENCODE Consortium portal.[13-16] The cell types analyzed as below include breast cancer (T-47D and MCF-7), cervical cancer (HeLa-S3), endometrial cancer (ECC-1), blood cancer (GM12878, GM12891, GM12892, HL-60 and K562), brain cancer (SK-N-MC, SK-N-SH, SK-N-SH_RA, PFSK-1 and U87), liver cancer (HepG2), colon cancer (HCT-116), pancreas cancer (PANC-1), lung cancer (A549), and human embryonic stem cell (H1-hESC).

CONTACT Binhua Tang bh.tang@outlook.com 200 N Jinling Rd., Hohai University, Jiangsu 213022, China.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/kbie.

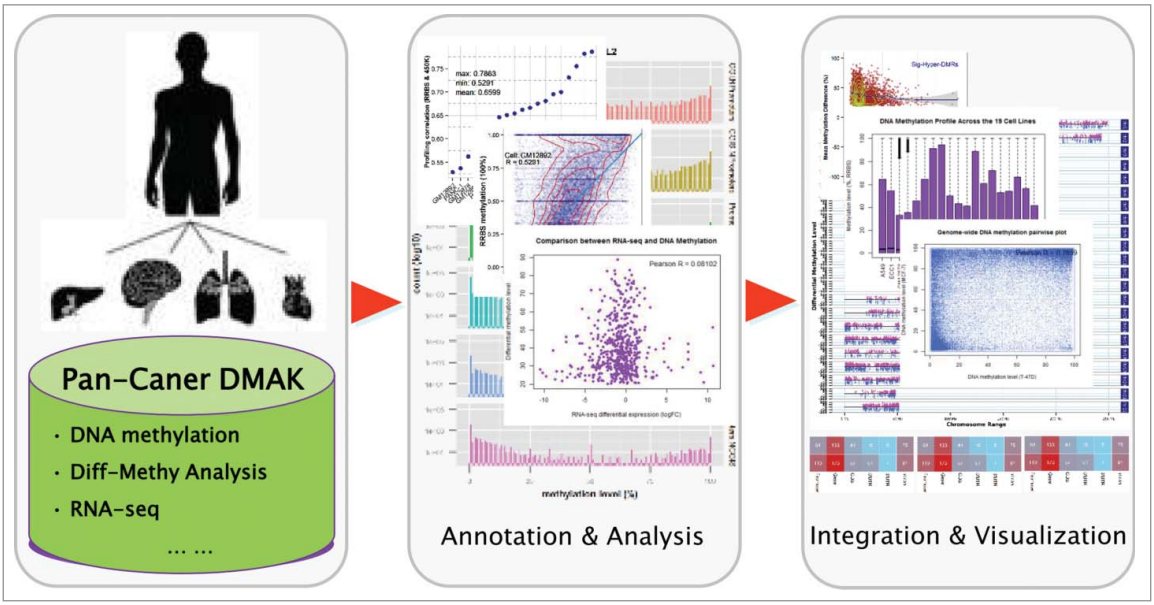

Figure 1. Schematic illustration for DMAK structure and function. The left panel covers data preprocess for pan-cancer cell lines (namely, cell line curation and data format process); the middle panel includes annotation and integrative analysis on the curated ENCODE data, namely DNA methylation CpGs annotation, identification of differentially-methylated CpGs and regions; the right panel covers function integration and visualization, which provides clues for further multi-scale validation.

As depicted in Fig. 1, the first module of DMAK is the curation of raw data sources from ENCODE, including DNA methylation, RNA-seq, Tumor Suppressor Gene (TSG) and corresponding genetic annotation and analysis; within the work, we emphasize on cross-cell DNA methylation profiling information for detecting differentially-methylated loci and regions with the 3 benchmark cell lines, lung cancer A549, breast cancer MCF-7 and T-47D.

The second module mainly focuses on genomic annotation and cross-cell function analysis on the curated DNA methylation data in RRBS format[17,18], we have implemented function annotation for methylated CpG sites, identified differentially-methylated regions (DMR), and classified the hyper- and hypo-methylated regions (or differential DMR candidates).[19] The detailed analysis procedure and results are given in the following section.

The third module includes the function integration and visualization for the annotated results, which includes the functional association network for tumor suppressor genes identified from the hyper- or hypo-DMRs derived from the above analysis, Gene Ontology and corresponding clinical outcome analysis.

We curated information and constructed the comprehensive knowledgebase using NGS data sources (namely RRBS, ChIP-seq and RNA-seq) mainly from ENCODE and TCGA, and the clinical survival resources retrieved from TCGA, together with other commonly-used tools, and the self-compiled programs.

Annotation and analysis procedure in DMAK

This section discusses the functions and analysis procedure in DMAK. As depicted in Fig. 2, the panel provides the 4 categories of annotation information, namely, the site methylation levels for all 19 cell lines, individual DMR-Genes (hyper- and hypo-methylated cases for each cell line), specific DMR-Genes (cell-specific hyper- and hypo-methylated cases) and common DMR-Genes (common hyper- and hypo-methylated cases) for the 3 selected benchmark cell lines (A549, MCF-7 and T-47D), respectively. Right after those radio options are the corresponding drop-down selection; together with Segment to View, Rows to View and Download options, which constitute an integrative operation panel for DMAK (Fig. 2).

Thus, corresponding to the panel, the annotation and analysis mainly covers the following sections,

Statistical information detected for sequencing read coverage

We performed statistical calculation for the sequencing reads coverage counts (Cs and Ts) for the 688,445 CpG sites across all 19 cell lines listed above. For consistence, all DNA methylation data from ENCODE are based on the RRBS platform. The illustrative output is given in Table 1,

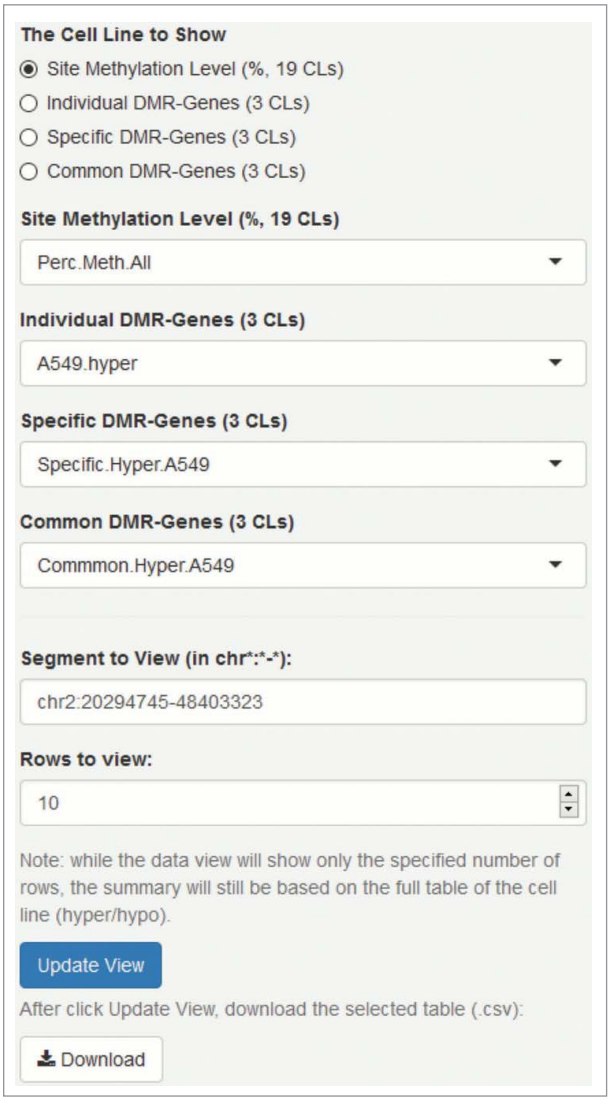

Figure 2. Schematic panel of annotation and analysis procedure in DMAK. It provides the 4 categories of annotation information, together with genomic Segment to View, Rows to View and Download options.

Statistical analysis and annotation for the identified genes from DMRs

We identified genes overlapping with all DMRs (hyper-DMRs and hypo-DMRs) with reference to each cell type (A549, MCF-7 and T-47D), respectively; then we further annotated those gene candidates with other information (symbol, log2 fold change of RNA-seq expression profile, TSG, genomic location and methylation level), thus it provides a thorough overview for those DMRs, depicted in below Table 2.

Summary panel of genome-wide DNA methylation for the 19 cell lines

This panel gives the statistical summary for the genome-wide methylation level for the 19 cell lines, which provides a general guide for comparing DNA methylation status across multiple cells, pairwise comparison or cross-cell analysis in Table 3.

Function integration and visualization

This section discusses function integration and visualization for the analysis results, including pairwise DNA methylation, differential DNA methylation and corresponding differential RNA expression comparison between cell lines, and identified genes of interest that are overlapped with the hyper- and hypo-DMRs, respectively (Fig. 3).

Furthermore we attempt to detect whether there exist any functional association between those identified genes from hyper-DMRs and hypo-DMRs, from protein level we can determine whether or not there is any potential functional link among those identified protein-coding genes,[20] which can further explain the differential expression between those genes qualitatively, especially for the genes belonging to tumor suppressor genes (TSG).[18,21,22]

Thus we annotated the genes identified from DMRs with TSG information, filtered out those from unknown sources, and constructed the TSG functional association networks for hyper-DMR and hypo-DMR, respectively.

Due to space limitation, Fig. 4 depicts the 20-TSG functional association structures for hyper-DMRs. For validating the high fidelity of the analysis results, those 20 TSGs are randomly selected from the TSG list for each case.

Table 1. Schematic illustration of statistical information for calculated methylation level (in percentage) from RRBS profiling technology.

Table View | Plot View | Summary | About

	Chr	Start	End	A549	ECC1	GM12878	GM12891	GM12892	H1hESC	HCT116	HeLaS3	HepG2	HL60	K562	MCF7	PANC1	PFSK1	SKNMC	SKNSH	SKNSHRA	T47DE2	U87
366990	chr2	20306626	20306627	74.47	69.44	85.53	90.91	85.59	69.23	88.66	95.56	87.69	85.71	76.88	86.96	84.72	86.44	87.34	85.14	86.02	66.67	85.37
366991	chr2	20306633	20306634	72.34	66.67	82.89	81.82	71.17	42.31	77.32	77.78	73.85	71.43	57.80	82.61	79.17	76.27	86.08	81.08	80.65	50.57	73.17
366992	chr2	20306637	20306638	72.34	69.44	75.00	63.64	76.58	53.85	75.26	77.78	70.77	71.43	60.12	82.61	75.00	71.19	84.81	79.73	73.12	48.28	75.61
366993	chr2	20306639	20306640	70.21	80.56	60.53	45.45	67.57	42.31	69.07	80.00	69.23	57.14	60.12	80.43	80.56	74.58	84.81	75.68	70.97	43.68	75.61
366994	chr2	20306656	20306657	8.51	5.56	6.58	18.18	0.90	3.85	3.09	8.89	10.77	7.14	5.20	8.70	13.89	8.47	10.13	13.51	7.53	0.00	0.00
366995	chr2	20306660	20306661	29.79	25.00	30.26	0.00	21.62	11.54	36.08	31.11	27.69	14.29	16.76	39.13	41.67	32.20	26.58	29.73	27.96	19.54	21.95

Table 2. Schematic illustration for the identified gene information (SYMBOL and ENTREZ ID), log2 fold change, methylation percentage, tumor suppressor gene category (TRUE/FALSE), loci (Promoter, CDS, Gene, 5′UTR, 3′UTR and Intron) and related methylation level (HYPER/HYPO) from DMRs with reference to T-47D cell type.

Table View | Plot View | Summary | About

	Chr	Strand	Start	End	SYMBOL	ENTREZID	logFC	Methy	TSGi	Loci	MethyLevel
117	chr2	-	38294745	38303323	CYP1B1	1545	0.40	37.78	FALSE	PROMOTER	HYPER
118	chr2	-	38294745	38303323	CYP1B1	1545	0.40	37.78	FALSE	5-UTR	HYPER
119	chr2	-	38294745	38303323	CYP1B1	1545	0.40	37.78	FALSE	INTRON	HYPER
120	chr2	-	38294745	38303323	CYP1B1	1545	0.40	23.81	FALSE	CDS	HYPER
200	chr2	-	31133330	31361592	GALNT14	79623	-0.44	26.02	FALSE	INTRON	HYPER
400	chr2	+	42275160	42285668	PKDCC	91461	-0.26	28.16	FALSE	CDS	HYPER
437	chr2	+	37571752	37600465	QPCT	25797	-0.79	37.46	FALSE	5-UTR	HYPER

And interestingly, we find most of those TSGs are functionally associated to form clusters. In Fig. 4, only 4 out of 20 TSGs are dissociated from the TSG cluster. Those structures further confirm TSGs are highly physically connected and functional associated in DMRs for the T-47D breast cancer case.

Next, we attempt to identify the clinical association with those gene candidates from DMRs, here for demonstration purpose, we resort to lung carcinoma study (A549 cell) in Fig. 3.

In lung carcinoma, it is recently reported that a long non-coding RNA, UCA1 (Urothelial cancer associated

Table 3. Summary panel of statistical information calculated from the RRBS profiling data across the 19 ENCODE cell lines.

Table View | Plot View | Summary | About

A549	ECC1	GM12878	GM12891	GM12892
Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.00	Min. : 0.000
1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.00	1st Qu.: 0.000
Median : 3.226	Median : 3.922	Median : 2.778	Median : 0.00	Median : 2.439
Mean : 27.786	Mean : 26.675	Mean : 22.275	Mean : 23.21	Mean : 24.740
3rd Qu.: 64.706	3rd Qu.: 54.545	3rd Qu.: 33.333	3rd Qu.: 35.48	3rd Qu.: 45.714
Max. :100.000	Max. :100.000	Max. :100.000	Max. :100.00	Max. :100.000
H1hESC	**HCT116**	**HeLaS3**	**HepG2**	**HL60**
Min. : 0.000	Min. : 0.000	Min. : 0.00	Min. : 0.000	Min. : 0.000
1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.00	1st Qu.: 0.000	1st Qu.: 0.000
Median : 2.041	Median : 4.167	Median : 22.22	Median : 3.636	Median : 1.562
Mean : 25.865	Mean : 37.934	Mean : 43.22	Mean : 26.394	Mean : 25.182
3rd Qu.: 64.706	3rd Qu.: 91.176	3rd Qu.: 94.59	3rd Qu.: 50.000	3rd Qu.: 43.243
Max. :100.000	Max. :100.000	Max. :100.00	Max. :100.000	Max. :100.000
K562	**MCF7**	**PANC1**	**PFSK1**	**SKNMC**
Min. : 0.000	Min. : 0.00	Min. : 0.000	Min. : 0.000	Min. : 0.000
1st Qu.: 0.000	1st Qu.: 0.00	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
Median : 4.444	Median : 5.00	Median : 2.564	Median : 5.085	Median : 3.846
Mean : 24.288	Mean : 35.38	Mean : 26.726	Mean : 30.558	Mean : 26.570
3rd Qu.: 41.379	3rd Qu.: 89.06	3rd Qu.: 60.833	3rd Qu.: 72.131	3rd Qu.: 52.778
Max. :100.000	Max. :100.00	Max. :100.000	Max. :100.000	Max. :100.000
SKNSH	**SKNSHRA**	**T47DE2**	**U87**	
Min. : 0.000	Min. : 0.000	Min. : 0.00	Min. : 0.000	
1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.00	1st Qu.: 0.000	
Median : 3.226	Median : 1.449	Median : 3.03	Median : 4.348	
Mean : 26.363	Mean : 26.806	Mean : 26.37	Mean : 24.694	
3rd Qu.: 54.167	3rd Qu.: 66.667	3rd Qu.: 56.41	3rd Qu.: 41.667	
Max. :100.000	Max. :100.000	Max. :100.00	Max. :100.000	

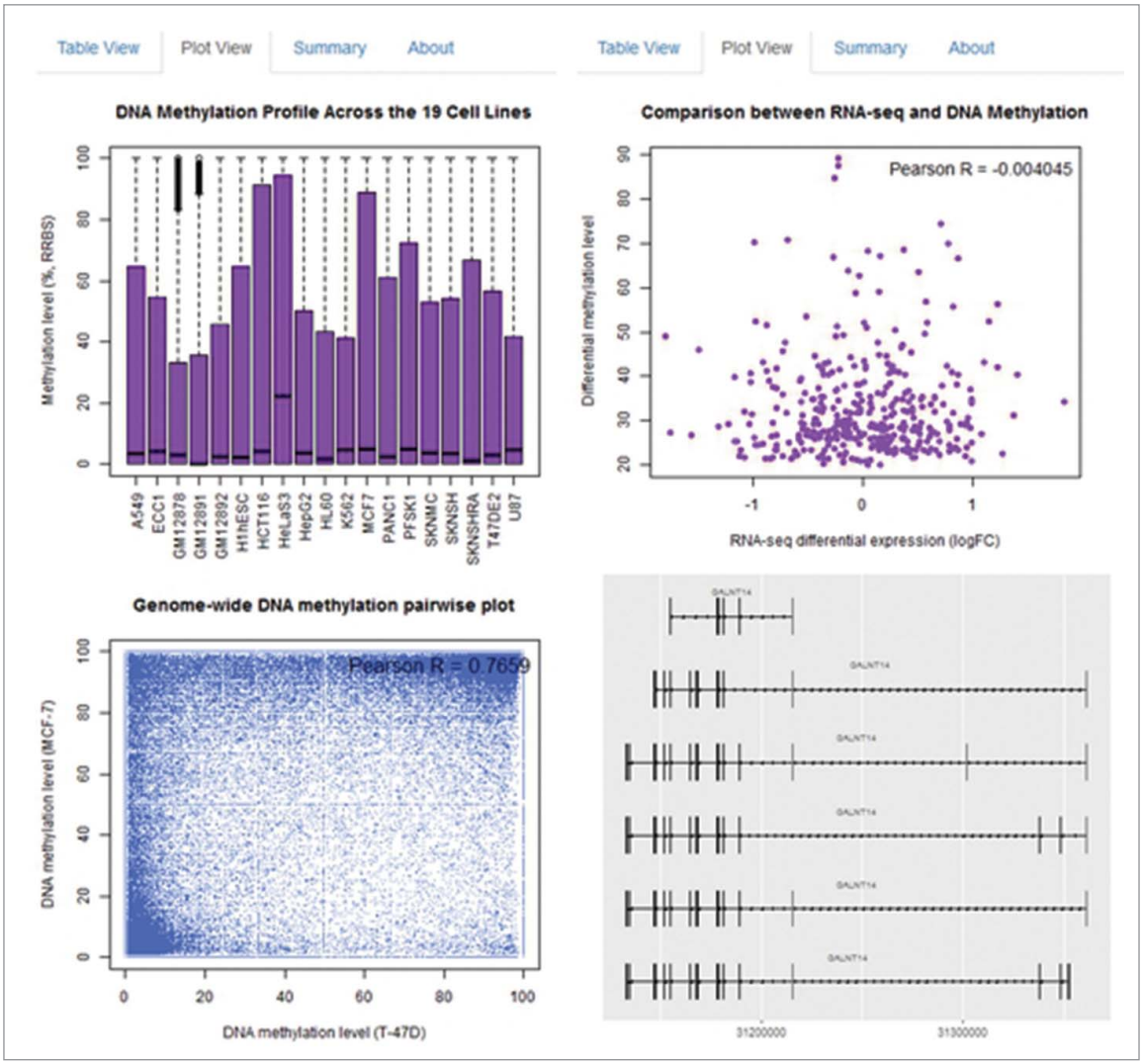

Figure 3. Schematic illustration of function integration and visualization for the annotated results for the ENCODE cell lines. Top left panel depicts the DNA methylation profile statistics for the 19 cell lines; bottom left panel for the genome-wide DNA methylation pairwise comparison between MCF-7 and T-47D cells, the methylation correlation coefficient is 0.7659 although both are from breast cancers; top right panel gives pairwise comparison between differential DNA methylation and RNA expression status for A549, with weak anti-correlation value, -0.004045, for the hyper-DMR case, bottom right depicts the gene information for GALNT3 at 2q24.1.

1) can up-regulate a potent oncogene ERBB4 (Erb-B2 receptor tyrosine kinase 4) by binding a microRNA, miR-193-3p, during transcriptional regulation.[23]

We use ERBB4 (2q33.3-q34) and UCA1 (19p13.12) as the study case for lung cancer, together with Kaplan-Meier probability analysis on the RNA-seq data and clinical survival information (lung adenocarcinoma, LUAD) retrieved from TCGA.[1,5]

Thus, based on the total clinical trial enrolment of 3,568 patients (LUAD), we calculate the clinical association anchored with the 2 candidate genes; as illustrated in Fig. 5, the results validate the 2 genes as the promising biomarkers or potential therapy targets in lung carcinoma.

Materials and methods

DNA methylation arrays 450K

The HumanMethylation 450K Beadchip assay is a CpG-specific array technology and allows for the high-resolution, genome-wide DNA methylation profiling with over 450,000 CpGs covering 99% of all RefSeq genes.[24-26]

Reduced representation bisulfite sequencing (RRBS)

Reduced representation bisulfite sequencing, or RRBS, is a large-scale random approach for analyzing and comparing genomic methylation patterns. BglII restriction fragments of 500–600 bp sized

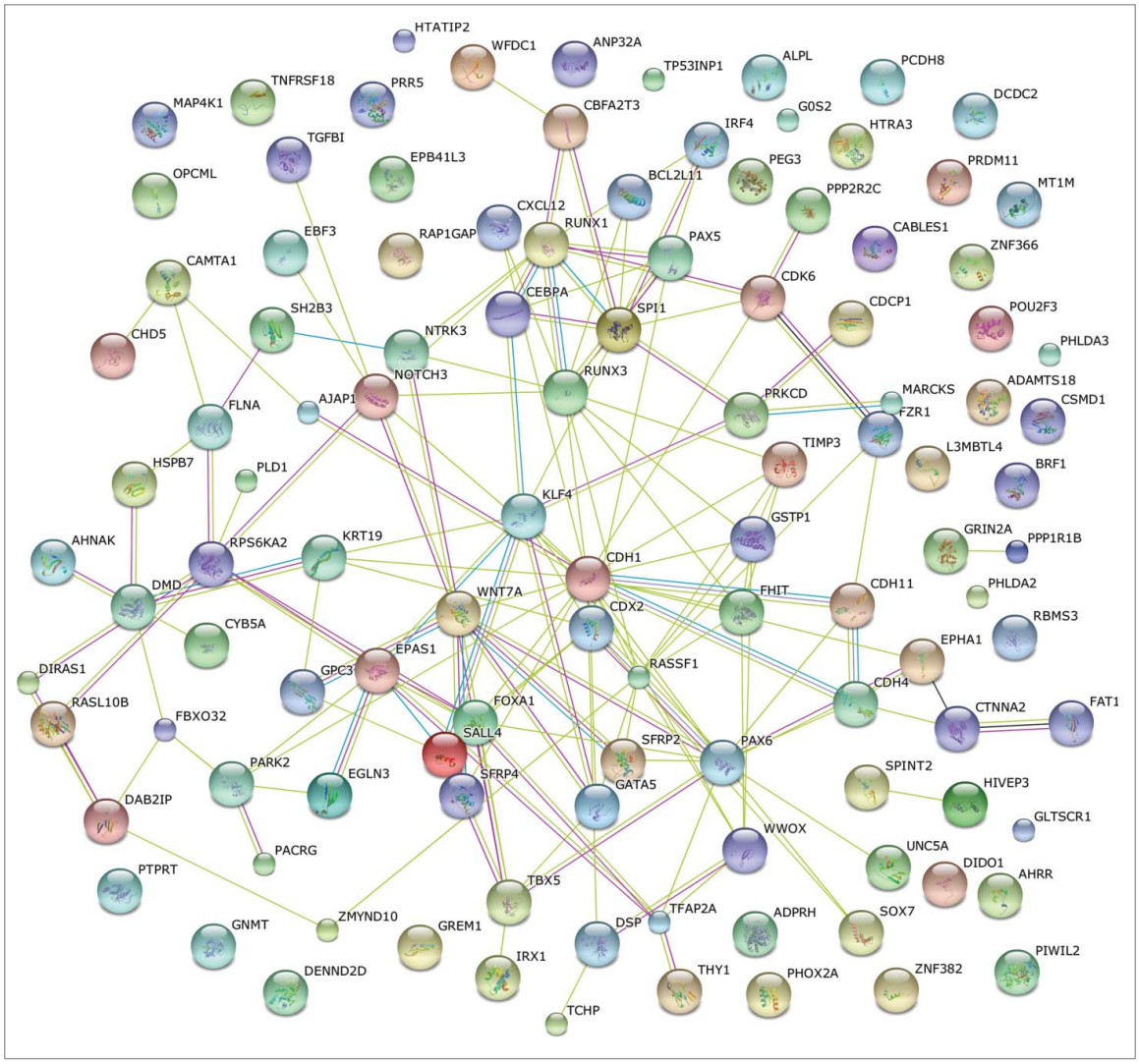

Figure 4. Illustrative diagram for functional protein association network inferred for the tumor suppressor genes (TSG), identified from hyper-DMRs. The nodes represent the protein-coding genes (including TSGs) detected from DMRs, and links represent the association evidences.

selected, together with adapters assembled, were further treated with bisulfite, PCR amplification and clone, and finally sequenced to target methylated CpG sites. From the converted and unconverted read counts at each CpG, the sample coverage and methylation level (in percentage) can be acquired.[11,27,28]

Annotation for the significant differentially-methylated CpG sites (SDMC)

Here we select one cell line (A549, MCF-7 and T-47D) as the reference cell type, and the annotation results are further filtered based on the lifted methylation difference threshold (at least 25% methylation difference for the paired groups). And the SDMC list contains 106,252 DMCs,[29,30] together the related statistical p-value and adjusted q-value are also provided.

Statistical analysis for the differentially-methylated regions

We identified 16,277 DMR candidates from all the DMCs, with the adjusted q-value $\leq$0.01, CpG base methylation difference cutoff, 25, and DMR mean methylation difference cutoff, 20. Within those candidates, 8,936 entries present hyper-methylated and 7,341 with hypo-methylated status. With the lifted thresholds, namely adjusted q-value $\leq$ 0.001, differentially-methylated CpG base count $\geq$5, we further detected 7,537 significant DMRs (Sig-DMRs), where 3,512 entries are significantly hypermethylated-DMRs

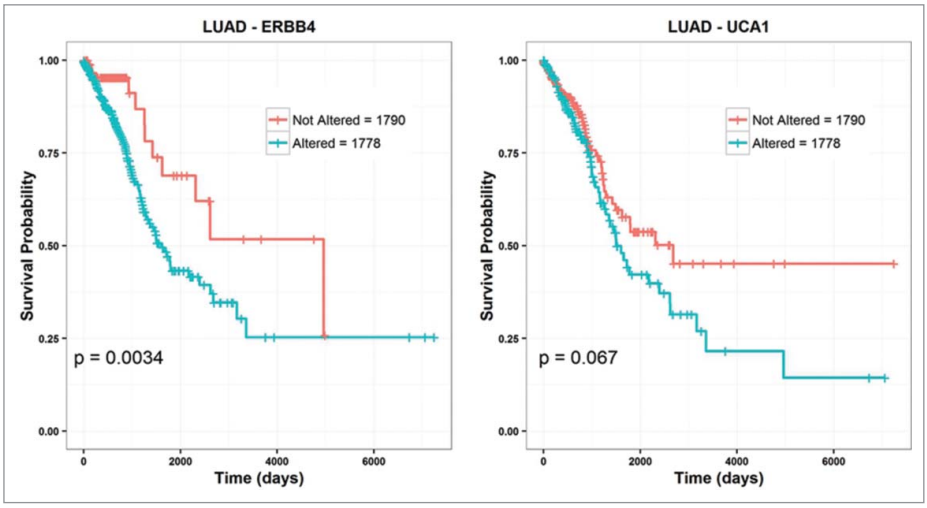

Figure 5. Kaplan-Meier survival probability plots for ERBB4 and UCA1 in LUAD clinical data, with the log-rank p-values given on the bottom right. The total clinical trial enrollment includes 3,568 patients (LUAD), with corresponding RNA-seq data retrieved from TCGA.

(Sig-Hyper-DMRs), and 4,025 significantly hypomethylated-DMRs (Sig-Hypo-DMRs).

Tools used in the curation and analysis

Bowtie2[31] was used to align sequencing reads; SAMtools[32] and BAMtools[33] were used to process the aligned sequencing reads; methylKit[30] was used to analyze part of RRBS data, and DEseq[34] was used to analyze RNA-seq data.

Conclusion

DMAK provides a comprehensive annotation and analysis knowledgebase for pan-cancer study. It contains 3 modules of curated reference results for ready-to-use information sharing and rapid reanalysis.

The first module of the knowledgebase is about raw data preprocess, and we retrieved DNA methylation data from ENCODE and clinical resources from TCGA. The second is for annotation and function analysis; in this study case, we focused on DNA methylation in breast cancer cell, T-47D, annotated and identified the differentially-methylated sites and regions, and further identified the underlying tumor suppressor genes within the regions. The third is for function integration and visualization procedures. We further constructed the functional association network for the identified tumor suppressor genes, and further performed the clinical association study with the DMR genes of interest, which can provide statistically significant evidences for the hyper-methylated and hypo-methylated processes in the transcriptional regulation context.

Our work provides a versatile and comprehensive platform for the corresponding biomedical research, especially for the genome-wide study, to interrogate and validate their hypothesis in an efficient and uniform way.

In coming days, further annotation and analysis results concerning pan-cancer analysis will be updated into the knowledgebase, thus it constitutes an interactive and efficient approach for biologists to carry out their research with knowledgebase.

Availability

DMAK is deployed at gladex.shinyapps.io/DMAK/ and dma2.hhuc.edu.cn/DMAK/.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Funding

This work was supported by the Natural Science Foundation of Jiangsu, China (Nos: BE2016655 and BK20161196), Fundamental Research Funds for China Central Universities (No. 2016B08914) and Changzhou Science & Technology Program (No. CE20155050). This work made use of the resources supported by the NSFC-Guangdong Mutual Funds for Super Computing Program (2nd Phase), and the Open Cloud Consortium (OCC)-sponsored project resource, which supported in part by grants from Gordon and Betty Moore Foundation and the National Science Foundation (USA) and major contributions from OCC members.

References

[1] The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013; 45(10):1113-20; PMID:24071849; http://dx.doi.org/10.1038/ng.2764

[2] Kristensen VN, Lingjærde OC, Russnes HG, Vollan HK, Frigessi A, Børresen-Dale AL. Principles and methods of integrative genomic analyses in cancer. Nat Rev Cancer 2014; 14(5):299-313; PMID:24759209; http://dx.doi.org/10.1038/nrc3721

[3] Witte T, Plass C, Gerhauser C. Pan-cancer patterns of DNA methylation. Genome Medicine 2014; 6(8):1-18; PMID:24433494; http://dx.doi.org/10.1186/s13073-014-0066-6

[4] Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet 2015; 47(2):106-114; PMID:25501392; http://dx.doi.org/10.1038/ng.3168

[5] The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012; 489(7417):519-525; PMID:22960745; http://dx.doi.org/10.1038/nature11404

[6] The Cancer Genome Atlas Research Network. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490(7418):61-70; PMID:23000897; http://dx.doi.org/10.1038/nature11412

[7] The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004; 306(5696):636-640; PMID:15499007; http://dx.doi.org/10.1126/science.1105136

[8] The Encode Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol 2011; 9(4):e1001046; PMID:21526222; http://dx.doi.org/10.1371/journal.pbio.1001046

[9] The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007; 447(7146):799-816; PMID:17571346; http://dx.doi.org/10.1038/nature05874

[10] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490(7418):61-70; PMID:23000897; http://dx.doi.org/10.1038/nature11412

[11] Roadmap Epigenomics Consortium. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015; 518(7539):317-330; PMID:25693563; http://dx.doi.org/10.1038/nature14248

[12] Bock C, Lengauer T. Computational epigenetics. Bioinformatics 2008; 24(1):1-10; PMID:18024971; http://dx.doi.org/10.1093/bioinformatics/btm546

[13] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57-74; PMID:22955616; http://dx.doi.org/10.1038/nature11247

[14] Pennisi E. ENCODE project writes eulogy for junk DNA. Science 2012; 337(6099):1159-1161; PMID:22955811; http://dx.doi.org/10.1126/science.337.6099.1159

[15] de Souza N. Genomics: The ENCODE project. Nat Meth 2012; 9(11):1046-1046; http://dx.doi.org/10.1038/nmeth.2238

[16] Tang B, Wang X. Inferring genome-wide interplay landscape between DNA methylation and transcriptional regulation P. J. Pharm. Sci. 2015; 28(1):349-352

[17] Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 2013; 500(7463):477-481; PMID:23925113; http://dx.doi.org/10.1038/nature12433

[18] Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, Farnham PJ. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biology 2014; 15(9):469; PMID:25239471; http://dx.doi.org/10.1186/s13059-014-0469-0

[19] Kemp CJ, Moore JM, Moser R, Bernard B, Teater M, Smith LE, Rabaia NA, Gurley KE, Guinney J, Busch SE, et al.. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep 2014; 7(4):1020-1029; PMID:24794443; http://dx.doi.org/10.1016/j.celrep.2014.04.004

[20] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucl Acids Res 2015; 43(D1):D447-D452; PMID:25352553; http://dx.doi.org/10.1093/nar/gku1003

[21] Bedi U, Mishra VK, Wasilewski D, Scheel C, Johnsen SA. Epigenetic plasticity: A central regulator of epithelial-to-mesenchymal transition in cancer. Oncotarget 2014; 5(8):2016-2029; PMID:24840099; http://dx.doi.org/10.18632/oncotarget.1875

[22] Zhao M, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. Nucleic Acids Research 2013; 41(D1):D970-D976; PMID:23066107; http://dx.doi.org/10.1093/nar/gks937

[23] Nie W, Ge HJ, Yang XQ, Sun X, Huang H, Tao X, Chen WS, Li B. LncRNA-UCA1 exerts oncogenic functions in non-small cell lung cancer by targeting miR-193a-3p. Cancer Letters 2016; 371(1):99-106; PMID:26655272; http://dx.doi.org/10.1016/j.canlet.2015.11.024

[24] Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium humanmethylation450 BeadChips. Genome Biology 2012; 13(6):R44; PMID:22703947; http://dx.doi.org/10.1186/gb-2012-13-6-r44

[25] Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. Genomics 2011; 98(4):288-295; PMID:21839163; http://dx.doi.org/10.1016/j.ygeno.2011.07.007

[26] Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462 (7271):315-322; PMID:19829295; http://dx.doi.org/10.1038/nature08514

[27] Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Research 2005; 33(18):5868-5877; PMID:16224102; http://dx.doi.org/10.1093/nar/gki901

[28] Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J, et al. The DNA methylation landscape of human early embryos. Nature 2014; 511(7511):606-610; PMID:25079557; http://dx.doi.org/10.1038/nature13544

[29] Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, Milne TA, Huang Y, Biswas D, Hess JL, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet 2012; 8(6): e1002781; PMID:22737091; http://dx.doi.org/10.1371/journal.pgen.1002781

[30] Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology 2012; 13 (10):R87; PMID:23034086; http://dx.doi.org/10.1186/gb-2012-13-10-r87

[31] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth 2012; 9(4):357-359; http://dx.doi.org/10.1038/nmeth.1923

[32] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25(16):2078-2079; PMID:19505943; http://dx.doi.org/10.1093/bioinformatics/btp352

[33] Barnett D, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011; 27 (12):1691-1692; PMID:21493652; http://dx.doi.org/10.1093/bioinformatics/btr174

[34] Anders S, Huber W. Differential expression analysis for sequence count data. Gen Biol 2010; 11(10):R106; http://dx.doi.org/10.1186/gb-2010-11-10-r106