Genomics Data 6 (2015) 12–15

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome-wide analysis of DNA methylation associated with HIV infection based on a pair of monozygotic twins

Yinfeng Zhang [a], Sai-Kam Li [a], Stephen Kwok-Wing Tsui [a,b,c,*]

[a] School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
[b] Hong Kong Bioinformatics Center, The Chinese University of Hong Kong, Hong Kong
[c] Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 25 June 2015
Accepted 7 July 2015
Available online 30 July 2015

Keywords:
HIV
DNA methylation
Monozygotic twins
MeDIP–microarray

ABSTRACT

Alteration of DNA methylation in mammalian cells could be elicited by many factors, including viral infections [1]. HIV has shown the ability to interact with host cellular factors to change the methylation status of some genes [2–4]. However, the change of the DNA methylation associated with HIV infection based on the whole genome has not been well illustrated. In this study, a unique pair of monozygotic twins was recruited: one of the twins was infected with HIV without further anti-retroviral therapy while the other one was healthy, which could be considered as a relatively ideal model for profiling the alterations of DNA methylation associated with HIV infection. Therefore, using methylated DNA immunoprecipitation–microarray method (MeDIP–microarray), we found the increased DNA methylation level in peripheral blood mononuclear cells from HIV infected twin compared to her normal sibling. Moreover, several distinguished differential methylation regions (DMRs) in HIV infected twin worth further study. The raw data has been deposited in Gene Expression Omnibus (GEO) datasets with reference number GSE68028.

Specifications	
Organism/cell line/tissue	Human peripheral blood mononuclear cells
Sex	Female
Sequencer or array type	NimbleGen Human DNA Methylation 2.1M Deluxe Promoter Array
Data format	Raw and processed
Experimental factors	One of the monozygotic twins was infected with HIV while the other one was not. In addition, the HIV + twin did not take any anti-retrovirus therapy.
Experimental features	DNA methylation from both twins was compared using the MeDIP–microarray to identify the alterations associated with HIV infection.
Consent	N/A
Sample source location	China

1. Deposited data

1.1. Direct link to deposited files

http://datalink.elsevier.com/midas/datalink/api/downloadfiles?items=16069-16070-16071.

1.2. Direct link to deposited genomic data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68028.

2. Experimental design, materials and methods

2.1. Peripheral blood mononuclear cell isolation and genomic DNA extraction

Peripheral blood mononuclear cells (PBMCs) were separated from the EDTA-blood collected from the pair of monozygotic twins. Both twins were 15 years old at the time of blood collection. In addition, it was around 7 years since the HIV infection occurred in HIV infected twin without receiving any anti-retrovirus treatments. The whole blood was mixed with the OptiPrep™ by repeated inversion (1/8 volume of the blood). Then, 0.5–1 ml RPMI1640 medium was added and

* Corresponding author at: School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong.
E-mail addresses: yinfengzh@gmail.com (Y. Zhang), kwtsui@cuhk.edu.hk (S.K.-W. Tsui).

http://dx.doi.org/10.1016/j.gdata.2015.07.024

Table 1
The differential methylated regions in HIV+ twin with peak value above 5.0.

Data index	Chromosome	Start	End	Peak value	Feature track	Feature strand	Name	CpG island
3115	chr19	52,390,044	52,391,710	5.9178	Transcription start site	−	ZNF577	Yes
4258	chr12	133,308,491	133,310,370	5.9045	Primary Transcript	−	ANKLE2	Yes
3334	chr22	18,120,437	18,121,372	5.7215	Transcription start site	+	BCL2L13	Yes
3956	chr7	774,083	775,359	5.6891	Primary Transcript	+	HEATR2	Yes
3695	chr2	3,290,959	3,291,604	5.6532	Primary Transcript	−	TSSC1	Yes
4035	chr8	99,958,705	99,959,730	5.5693	Primary Transcript	+	OSR2	Yes
429	chr2	88,752,443	88,753,744	5.5688	Transcription start site	−	FOXI3	Yes
526	chr2	2E+08	2E+08	5.5561	Transcription start site	−	SATB2	Yes
2582	chr15	68,567,915	68,569,664	5.5512	Transcription start site	+	FEM1B	Yes
5746	chr19	19,531,445	19,532,608	5.5192	Primary Transcript	+	GATAD2A	No
397	chr2	55,645,362	55,646,491	5.4771	Transcription start site	−	CCDC88A	Yes
5571	chr15	23,383,697	23,384,784	5.4762	Intergenic region		chr15:23,375,259–23,386,259	No
3608	chr1	1,087,609	1,089,246	5.4729	Tiled Region		chr1:1,087,484–1,118,109	No
2558	chr15	52,822,062	52,823,038	5.4564	Transcription start site	−	MYO5A	Yes
4243	chr12	1.09E+08	1.09E+08	5.4532	Primary Transcript	−	CORO1C	Yes
2198	chr12	46,767,058	46,768,326	5.4271	Transcription start site	−	SLC38A2	Yes
1004	chr5	1.31E+08	1.31E+08	5.4156	Transcription start site	−	FNIP1	Yes
4007	chr8	989,268	990,711	5.3297	Intergenic region		chr8:983,687–986,272	Yes
4592	chrX	1,561,952	1,562,785	5.3226	Primary Transcript	−	ASMTL	Yes
3335	chr22	18,588,750	18,589,791	5.3223	Transcription start site	+	TUBA8	Yes
3505	chrX	84,362,643	84,363,990	5.3138	Transcription start site	−	SATL1	No
3702	chr2	42,718,764	42,720,113	5.305	Primary Transcript	−	KCNG3	Yes
2196	chr12	44,149,645	44,150,881	5.2788	Transcription start site	+	IRAK4	Yes
3436	chrX	37,026,929	37,028,906	5.2767	Transcription start site	+	FAM47C	Yes
4844	chr2	99,286,979	99,287,918	5.2546	Primary Transcript	−	MGAT4A	No
1007	chr5	1.32E+08	1.32E+08	5.2447	Transcription start site	+	ANKRD43	Yes
2616	chr15	99,788,997	99,789,953	5.2296	Transcription start site	+	LRRC28	Yes
4394	chr16	89,646,203	89,647,802	5.2046	Primary Transcript	+	CPNE7	Yes
5239	chr8	1.02E+08	1.02E+08	5.1728	Intergenic region		chr8:102,373,120–102,384,120	No
4395	chr16	89,900,511	89,901,696	5.1493	Primary Transcript	+	SPIRE2	Yes
415	chr2	71,291,030	71,292,085	5.1481	Transcription start site	+	NAGK	Yes
1566	chr8	1.19E+08	1.19E+08	5.1422	Transcription start site	−	EXT1	Yes
1032	chr5	1.43E+08	1.43E+08	5.1276	Transcription start site	−	NR3C1	No
1772	chr10	14,876,960	14,877,948	5.1176	Transcription start site	−	CDNF	Yes
1637	chr9	37,032,945	37,034,183	5.0996	Transcription start site	−	PAX5	Yes
1450	chr7	1.59E+08	1.59E+08	5.0942	Transcription start site	−	VIPR2	Yes
2218	chr12	53,490,029	53,491,060	5.0796	Transcription start site	+	IGFBP6	Yes
2371	chr13	1.01E+08	1.01E+08	5.0787	Transcription start site	−	TMTC4	Yes
2837	chr17	43,595,250	43,597,099	5.0624	Transcription start site	−	LOC652203	No
2677	chr16	53,738,716	53,739,977	5.0471	Transcription start site	+	FTO	Yes
2970	chr18	76,825,666	76,827,858	5.0264	Transcription start site	+	ATP9B	Yes
4726	chr1	84,757,044	84,758,504	5.0263	Intergenic region		chr1:84,756,048–84,771,333	No
4218	chr12	6,399,331	6,400,806	5.0153	Intergenic region		chr12:6,399,262–6,400,771	Yes
3931	chr6	1.37E+08	1.37E+08	5.0098	Intergenic region		chr6:137,235,401–137,246,401	Yes
3920	chr6	1.01E+08	1.01E+08	5.0078	Primary Transcript	−	SIM1	Yes

centrifuged at 900 g for 30 min at 20 °C. Later, the middle layer comprised of the PBMC was collected and washed by phosphate buffered saline twice. Total genomic DNA was extracted from the PBMCs using QIAamp DNA blood mini Kit (QIAGEN) according to the manufacturer's instructions. Genomic DNA quality and quantity were determined by using a NanoDrop 2000c (Thermo Scientific).

2.2. MeDIP–microarray analysis

The sonication was employed for genomic DNA fragmentation ranging from 250 to 500 bp in length. Then the Methylated-DNA IP Kit (Zymo Research) was used for the immunoprecipitation of cytosine methylated DNA fragments from both twins. Afterwards, DNA labeling and hybridization were performed according to NimbleGen's standard protocol. The immunoprecipitated CpG-methylated DNA from HIV positive subject (test) and from HIV negative subject (input control) was labeled with fluorescent dyes Cy5 and Cy3 using NimbleGen Dual-Color DNA Labeling Kit (Roche–NimbleGen). The combined test and input control DNA samples were suspended in hybridization buffer (Roche–NimbleGen) and co-hybridized onto NimbleGen Human DNA Methylation 2.1M Deluxe Promoter Array for 20 h at 42 °C, following washing with the Wash Kit (Roche–NimbleGen).

2.3. Microarray data analysis

The raw image files were obtained by MS 200 Microarray Scanner and MS 200 Data Collection Software. Then, the DEVA version 1.2.1 software (Data Extraction Visualization Analysis software, Roche,

Table 2
The differential methylated regions in HIV− twin with peak value above 5.0.

Data index	Chromosome	Start	End	Peak value	Feature track	Feature strand	Name	CpG island
408	chr4	3,310,903	3,312,010	5.4753	Transcription start site	+	RGS12	Yes
2640	chrX	85,510,778	85,511,909	5.2014	Primary Transcript	+	DACH2	No
1047	chr11	1.05E+08	1.05E+08	5.1255	Transcription start site	−	CASP5	No
852	chr8	1.44E+08	1.44E+08	5.0597	Transcription start site	+	GPIHBP1	No
2119	chr3	1.13E+08	1.13E+08	5.0144	Intergenic region		chr3: 112923374–112938044	No
382	chr3	1.53E+08	1.53E+08	5.0026	Transcription start site	+	C3orf79	No

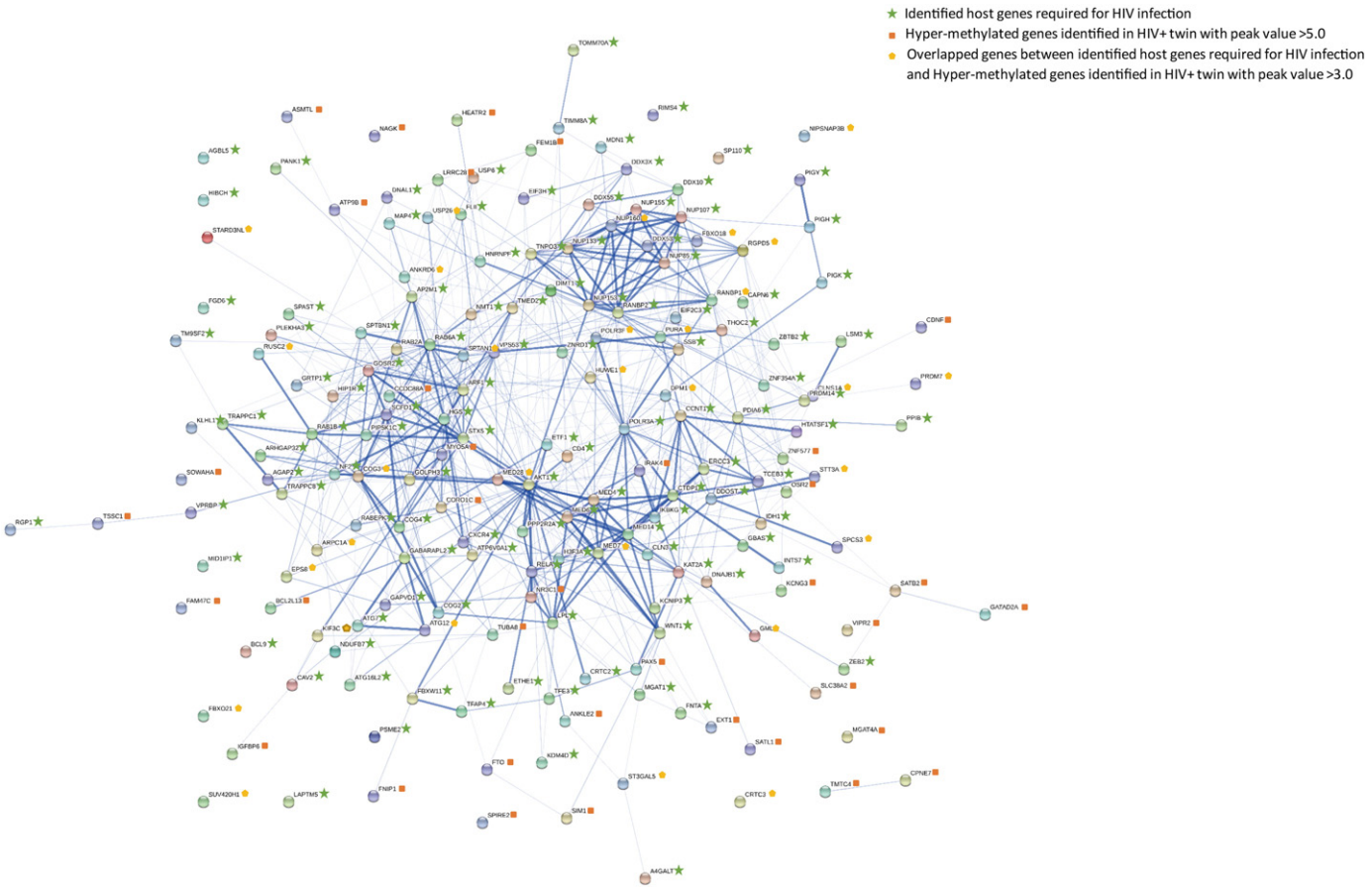

Fig. 1. Potential interactions between hyper-methylated genes in HIV + twin and known genes required for HIV infection by STRING [6]. The green pentagram indicated that the genes were know genes required for HIV infection reported by others [5]; the orange rectangle indicated that the genes were hyper-methylated in HIV + twin with peak value > 5.0; the yellow pentagon indicated that the genes were known to be required in HIV infection, but could also be hyper-methylated in HIV + twin with peak value > 3.0.

NimbleGen) was used for further analysis. Then, the result of microarray was analyzed by DEVA version 1.2.1 software (Roche, NimbleGen) using default parameters. In detail, 100929_HG19_Deluxe_Prom_Meth_HX1.ncd, 100929_HG19_Deluxe_Prom_Meth_HX1.ndf, 100929_HG19_Deluxe_Prom_Meth_HX1.pos and 100929_HG19_Deluxe_Prom_Meth_HX1.gff files were imported as the design files; the hg19 genome build was selected for organisms. The proper annotation files were selected for the identification of the features of the probes. After analyzing, the results of log2 ratios, P-score derived from Kolmogorov–Smirnov (KS) test were processed and obtained. The final results were presented as peak value based on P-score. The larger peak value from the designated region indicated that its differential methylation level was higher. The threshold for defining the differential methylation regions (DMRs) was set to 3.0. In Tables 1 and 2, the resulting list of the most differentially methylated regions (peak value > 5.0) from the HIV + twin and HIV − twin was shown. In addition, the hyper-methylated genes identified in HIV + twin were combined with the reported host genes required for HIV infection [5] to predict the potential protein–protein interactions by STRING [6]. The potential or known relationships among the genes were shown in Fig. 1.

3. Discussion

Nowadays, viruses are known to be able to change the DNA methylation pattern in host [1], such as HIV [2-4]. In addition, high throughput microarray method has been widely used to investigate the DNA methylation status across the whole genome. We showed the differential DNA methylation pattern in a rare pair of identical twins, while the difference was potentially associated with HIV infection. The further analysis and interpretation of the results were included in Zhang et al. (2015) [7]. Although the dataset was valuable in identifying the host genes which may play important roles in the course of AIDS, it would be noted that the meticulous experimental validation should be performed before any conclusion could be drawn.

Acknowledgments

This project was partly funded by the Scheme B funding on the project “Establishment of the Centre for Microbial Genomics and Proteomics”, the Scheme D funding on the project “Enhancing the Capability of the Hong Kong Bioinformatics Centre” of the Focused Investment Scheme and two grants from Human Medical Research Fund of the Food and Health Bureau, Hong Kong Special Administrative Region (reference nos.: 12110582 and 14130302).

References

[1] R. Ferrari, A. Berk, S. Kurdistani, Viral manipulation of the host epigenome for oncogenic transformation. Nat. Rev. Genet. 10 (2009) 290–294, http://dx.doi.org/10.1038/nrg2539.
[2] B. Youngblood, N.O. Reich, The early expressed HIV-1 genes regulate DNMT1 expression. Epigenetics 3 (2008) 149–156, http://dx.doi.org/10.4161/epi.3.3.6372.
[3] M. Pion, D. Jaramillo-Ruiz, A. Martinez, M.A. Munoz-Fernandez, R. Correa-Rocha, HIV infection of human regulatory T cells downregulates Foxp3 expression by increasing DNMT3b levels and DNA methylation in the FOXP3 gene. AIDS 27 (2013) 2019–2029, http://dx.doi.org/10.1097/QAD.0b013e32836253fd.

[4] J.Y. Fang, J.A. Mikovits, R. Bagni, C.L. Petrow-Sadowski, F.W. Ruscetti, Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases de novo methylation. J. Virol. 75 (2001) 9753–9761, http://dx.doi.org/10.1128/JVI.75.20.9753-9761.2001.
[5] A. Brass, D. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. Xavier, J. Lieberman, S. Elledge, Identification of host proteins required for HIV infection through a functional genomic screen. Science 319 (2008) 921–926, http://dx.doi.org/10.1126/science.1152725.
[6] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. Tsafou, M. Kuhn, P. Bork, L. Jensen, C. Mering, STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43 (2015) D447–D452, http://dx.doi.org/10.1093/nar/gku1003.
[7] Y. Zhang, S.K. Li, K.Y. Yang, M. Liu, N. Lee, X. Tang, H. Wang, L. Liu, Z. Chen, C. Zhang, J. Wang, S.K. Tsui, Whole genome methylation array reveals the down-regulation of IGFBP6 and SATB2 by HIV-1. Sci. Rep. 5 (2015) 10806, http://dx.doi.org/10.1038/srep10806.