Genomics Data 4 (2015) 143–145

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Genome-wide gene expression profiling to investigate molecular phenotypes of Arabidopsis mutants deprived in distinct histone methyltransferases and demethylases

Wei Zhao, Sarfraz Shafiq [1], Alexandre Berr, Wen-Hui Shen *

Institut de Biologie Moléculaire des Plantes, UPR2357 CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France

ARTICLE INFO

Article history:
Received 1 April 2015
Accepted 6 April 2015
Available online 14 April 2015

Keywords:
Chromatin
Histone methylation
Transcription
Arabidopsis

ABSTRACT

Histone lysine (K) methylation is a type of epigenetic modification involved in regulation of DNA-based processes, including transcription, replication and repair. It can either activate or repress transcription depending on the histone K residue on which methylation occurs and on chromatin context of additional other modifications. In both animals and plants, methylation on one histone K residue can be deposited by several different histone methyltransferases and *vice versa* removed by different histone demethylases. It is of great interest to know which histone enzyme regulates which genes in the genome. Here we describe in details the contents and quality controls for the gene expression data of Arabidopsis mutants deprived in distinct histone methyltransferases (SDG26, SDG25, ATX1, CLF) and histone demethylases (LDL1, LDL2), in association with the study recently published by Berr and colleagues in The Plant Journal (Berr et al., 2015). The microarray dataset has been deposited in Gene Expression Omnibus with accession number GSE55167.

Specifications	
Organism/cell line/tissue	*Arabidopsis thaliana*, 16-day-old seedlings.
Sex	NA
Sequencer or array type	Agilent's Whole Arabidopsis Gene Expression Microarray (4 × 44K).
Data format	Raw data: TAR of TXT files, normalized data: TXT and XLS of SOFT.
Experimental factors	Mutant *vs.* wild-type plants.
Experimental features	Seed germination and plant growth for all mutants and wild-type were performed under the same conditions; plant age used in analysis corresponds to 16 days counted after seed sawing.
Consent	NA
Sample source location	Strasbourg, France.

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55167.

Experimental design, materials and methods

Plant materials and growth conditions

All Arabidopsis strains used in this work are in the Columbia (Col) ecotype background. Description of mutant strains can be found from previous publications: *sdg26-1* [1], *sdg25-1* [2], *atx1-2* [3], *clf-29* [4], *ldl1-2 ldl2* [5,6], and the combined mutants *sdg26 sdg25*, *sdg26 atx1*, *sdg26 clf* as well as *sdg26 ldl1 ldl2* [7]. Seeds of wild-type (*Col0*) and the abovementioned mutant strains were produced from plants grown on soil in glasshouse. For seed surface sterilization, about 100 seeds were disposed in an open 2 ml Eppendorf tube and the tube was placed in a desiccator under a fume hood. A beaker containing 20 ml of bleach (FLOREAL Haagen GmbH, http://www.eau-de-javel.info/) was placed close to the seed tube, 5 ml of concentrated HCl (37%) was added to the bleach, and the lid of the desiccator was immediately closed to keep the produced chlorine gas in the desiccator compartment. After 3 to 6 h, sterilization was complete and the seeds could be used for *in vitro* plant culture.

Surface-sterilized seeds were plated on agar-solidified Murashige and Skoog (MS) medium M0255 (Duschefa, http://www.duchefa-biochemie.com/) supplemented with 0.9% sucrose in Petri Dishes. After stratification in the dark at 4 °C for 48 h that helps to synchronize seed germination time, the Petri Dishes were moved to the growth chamber and incubated for plant growth under Medium Day (MD; 12 h light and 12 h dark) photoperiods at 22 °C.

* Corresponding author.
E-mail address: wen-hui.shen@ibmp-cnrs.unistra.fr (W.-H. Shen).
[1] Current address: Center for Plant Biology, School of Life Sciences, Tsinghua University, and Tsinghua–Peking Center for Life Sciences, Beijing 100084, China.

http://dx.doi.org/10.1016/j.gdata.2015.04.006

Sample collection and RNA preparation

Sample collection and RNA preparation were performed using plants 16 days after seed plating (16-day-old seedlings). For each sample set, about 30 to 40 plants were collected from three replicating Petri Dishes. For each plant genotype/strain, three independent sets were pooled. Thus, a total of 33 samples were analyzed in this work. Total RNA was isolated from plant samples using the Nucleospin RNA Plant kit (Macherey-Nagel, http://www.mn-net.com) followed by RNeasy mini kit (Qiagen, https://www.qiagen.com/) clean-up according to manufacturer's instructions. The quality of the RNA obtained from each sample was assessed based on the RNA profile generated by the Agilent Bioanalyzer 2100 (Agilent technologies, Santa Clara, CA). Our entire samples' RNAs are in high quality in line with the requirement for microarray analysis (Table 1).

Microarray data

Microarray analyses were performed using Agilent's Whole Arabidopsis Gene Expression Microarray (G2519F, V4, 4 × 44K) *via* custom service of the Shanghai Huaguan Biochip Co. (http://www.bio-equip.cn/). Total RNA was amplified and Cy3-labeled using Low Input Quick Amp Labeling Kit, one-color by following the manufacturer's instructions (Agilent technologies, http://www.genomics.agilent.com/). Labeled cRNA was purified using RNeasy mini kit (Qiagen), slide hybridization at 65 °C for 17 h and image scanning (Scan resolution = 5μm, PMT 100%, 10%, 16 bits) were performed by following standard Agilent's protocols. Data extraction was performed using Agilent Feature Extraction software. Feature Extraction was used for background subtraction and normalization between the set of arrays was done by Quantile algorithm, a method that gives the same distribution of probe intensities to each array in the set [8]. Microarray and sample annotation data were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE55167.

Quality control

In addition to good quality of RNAs (Table 1), our microarray hybridization and detection of expressed genes were also at high quality (Table 2). The percent coefficient of variation (%CV) values for our entire microarray sets were largely inferior to 15%, a value proposed as validation threshold by Agilent. A lower median %CV value indicates better reproducibility of signal across the microarray than a higher value. Moreover, independent RT-PCR analyses confirmed microarray data on several selected genes (see below).

Basic and focused analysis

From previous studies we know that full-length transcripts of mutated genes are missing in the corresponding mutant strains used in this work. Therefore, it is predicted that our microarray data should confirm this information. Indeed, mutated genes are among the down-regulated genes in the respective mutants (GSE55167; [7]), *e.g. SDG26* is downregulated in *sdg26*, *sdg26 sdg25*, *sdg26 atx1*, *sdg26 clf* and *sdg26 ldl1 ldl2* but not in the other mutants, *SDG25* is downregulated in *sdg25* and *sdg26 sdg25* but not in the other mutants, *etc*. This confirmatory information on multiple mutated genes in various single, double and triple gene mutants in a same microarray analysis provides a good validation of our data.

Our focused analysis on flowering time regulatory genes revealed that *FLOWERING LOCUS C* (*FLC*) is upregulated or downregulated (−)

Table 1
RNA quality assessment of samples studied in this work.

Experiment	Arabidopsis	Sample	A260/A280(a)	RIN(b)	28S/18S(b)
Repeat-1	*Col*	1	1.97	7.1	1.3
	sdg25	2	2.04	7.0	1.3
	sdg26	3	2.06	7.0	1.3
	sdg25 sdg26	4	1.95	7.2	1.6
	atx1	5	2.06	7.0	1.3
	sdg26 atx1	6	2.03	7.0	1.4
	clf	7	1.86	7.1	1.3
	sdg26 clf	8	2.05	7.0	1.6
	ldl1 ldl2	9	1.92	7.4	1.7
	sdg25 ldl1 ldl2	10	2.04	7.0	1.3
	sdg26 ldl1 ldl2	11	1.91	7.2	1.4
Repeat-2	*Col*	12	1.92	7.4	1.5
	sdg25	13	1.88	7.0	1.3
	sdg26	14	2.02	7.0	1.5
	sdg25 sdg26	15	1.95	7.0	1.4
	atx1	16	2.01	7.0	1.3
	sdg26 atx1	17	2.03	7.0	1.3
	clf	18	1.85	7.0	1.3
	sdg26 clf	19	2.02	7.0	1.4
	ldl1 ldl2	20	1.91	7.4	1.6
	sdg25 ldl1 ldl2	21	1.96	7.6	1.4
	sdg26 ldl1 ldl2	22	2.02	7.3	1.4
Repeat-3	*Col*	23	1.98	7.2	1.9
	sdg25	24	1.93	7.2	1.8
	sdg26	25	1.94	7.0	1.6
	sdg25 sdg26	26	2.01	7.0	1.4
	atx1	27	1.95	7.3	1.5
	sdg26 atx1	28	1.97	7.1	1.4
	clf	29	1.95	7.1	1.4
	sdg26 clf	30	1.93	7.0	1.6
	ldl1 ldl2	31	1.93	7.4	1.5
	sdg25 ldl1 ldl2	32	1.95	7.1	1.7
	sdg26 ldl1 ldl2	33	2.01	7.4	1.7

(a) A ratio of A260/A280 > 1.8 indicates little protein contamination.
(b) RNA integration number (RIN) ≥ 7.0 and 28S/18S ≥ 0.7 indicate good quality for microarray assay.

Table 2
Quality assessment of microarray hybridization reproducibility and detection rate for samples studied in this work.

Experiment	Arabidopsis	Sample	%CV(*)	Detection rate	GEO file
Repeat-1	*Col*	1	7.40	70.47	GSM1330673
	sdg25	2	4.52	70.58	GSM1330674
	sdg26	3	3.93	69.44	GSM1330675
	sdg25 sdg26	4	7.17	65.57	GSM1330676
	atx1	5	6.07	72.13	GSM1330677
	sdg26 atx1	6	5.07	75.74	GSM1330678
	clf	7	4.50	76.60	GSM1330679
	sdg26 clf	8	5.61	72.02	GSM1330680
	ldl1 ldl2	9	10.69	72.51	GSM1330681
	sdg25 ldl1 ldl2	10	5.97	73.48	GSM1330682
	sdg26 ldl1 ldl2	11	6.64	71.34	GSM1330683
Repeat-2	*Col*	12	5.52	70.18	GSM1330684
	sdg25	13	3.22	75.06	GSM1330685
	sdg26	14	3.80	78.41	GSM1330686
	sdg25 sdg26	15	9.75	70.20	GSM1330687
	atx1	16	2.83	76.88	GSM1330688
	sdg26 atx1	17	3.09	76.88	GSM1330689
	clf	18	4.47	76.22	GSM1330690
	sdg26 clf	19	8.57	73.12	GSM1330691
	ldl1 ldl2	20	3.17	75.81	GSM1330692
	sdg25 ldl1 ldl2	21	2.33	76.61	GSM1330693
	sdg26 ldl1 ldl2	22	9.97	72.81	GSM1330694
Repeat-3	*Col*	23	6.84	70.20	GSM1330695
	sdg25	24	9.19	69.06	GSM1330696
	sdg26	25	3.81	74.97	GSM1330697
	sdg25 sdg26	26	3.39	75.79	GSM1330698
	atx1	27	3.96	74.84	GSM1330699
	sdg26 atx1	28	4.50	77.53	GSM1330700
	clf	29	3.86	72.12	GSM1330701
	sdg26 clf	30	3.90	76.70	GSM1330702
	ldl1 ldl2	31	3.54	76.34	GSM1330703
	sdg25 ldl1 ldl2	32	4.26	75.76	GSM1330704
	sdg26 ldl1 ldl2	33	3.70	76.52	GSM1330705

(*) %CV indicates for percent coefficient of variation.

to more than 2 folds in several of the mutant strains analyzed in microarray (GSE55167; Table 3). We validated the observed differential *FLC* expression in mutants by quantitative RT-PCR analysis. For qRT-PCR analyses, first-strand cDNA was synthesized from 2 μg of total RNA pretreated with 2 units of DNase I using the Impro-II Reverse Transcriptase system with oligo$(dT)_{20}$ primer by following the manufacturer's instruction (Promega, http://www.promega.com). The synthesized cDNA was analyzed by quantitative PCR in a 384-well optical plate on a BioRad i-cycler apparatus using 5 μl of PCR master mix (Roche) containing 480 SYBER® Green I fluorescent reporter with 2.5 μM forward and reverse *FLC*-specific primers. PCR reaction was performed by a step of 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C, 30 s at 60 °C and 15 s at 72 °C. Melting curves of PCR reactions were checked to insure the quality of PCR reaction and to avoid any DNA contamination. The threshold cycle value (CT) was set so that the fluorescent signal was above the baseline noise but as low as possible in the exponential amplification phase. The relative expression level of *FLC* was calculated using the Lightcycler 480 software and normalized using *ACT2* and *GAPDH* as internal reference genes. Each sample was analyzed in triplicate, and mean ± SD was shown as result (Table 3). Our qRT-PCR data show an overall agreement with microarray results obtained on *FLC* expression changes in different mutants.

Table 3
Validation of microarray data by quantitative RT-PCR analysis on *FLC* expression in different mutants as compared to wild-type (set as 1).

Mutant strain	Fold change in microarray	Relative level in RT-PCR
sdg25	−5.5	0.15 ± 0.02
sdg26	2.4	3.24 ± 0.31
sdg25 sdg26	<2.0	1.07 ± 0.15
atx1	−6.0	0.24 ± 0.04
sdg26 atx1	<2.0	1.08 ± 0.16
clf	3.8	3.80 ± 0.20
sdg26 clf	6.7	7.52 ± 0.11
ldl1 ldl2	<2.0	1.92 ± 0.23
sdg25 ldl1 ldl2	<2.0	1.17 ± 0.03
sdg26 ldl1 ldl2	5.1	4.57 ± 0.62

Discussion

We described here a unique dataset of microarray analyses on multiple different mutants deprived of one, two or three histone-methyltransferase(s)/histone-demethylase(s). This dataset has been recently used in a study focused on plant flowering time regulation [7]. We anticipate that the dataset may also be useful for comparative study of other mutants as well as for investigation of roles of histone methylations in processes beyond plant flowering time control.

Acknowledgments

We thank Dr. Yu Yu (Fudan University) for assistant in microarray analysis. This work was supported by the French Agence Nationale de la Recherche (ANR-12-BSV2-0013-02) and the European Commission (FP7-PEOPLE-2013-ITN, grant no 607880). S.S. received a PhD fellowship from the Pakistan government.

References

[1] L. Xu, Z. Zhao, A. Dong, L. Soubigou-Taconnat, J.-P. Renou, A. Steinmetz, W.-H. Shen, Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Mol. Cell. Biol. 28 (2008) 1348–1360.
[2] A. Berr, L. Xu, J. Gao, V. Cognat, A. Steinmetz, A. Dong, W.-H. Shen, *SET DOMAIN GROUP25* encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering. Plant Physiol. 151 (2009) 1476–1485.
[3] S. Pien, D. Fleury, J.S. Mylne, P. Crevillen, D. Inzé, Z. Avramova, C. Dean, U. Grossniklaus, *ARABIDOPSIS TRITHORAX1* dynamically regulates *FLOWERING LOCUS C* activation via histone 3 lysine 4 trimethylation. Plant Cell 20 (2008) 580–588.
[4] L. Xu, W.H. Shen, Polycomb silencing of *KNOX* genes confines shoot stem cell niches in Arabidopsis. Curr. Biol. 18 (2008) 1966–1971.
[5] D. Jing, W. Yang, Y. He, R.M. Amasino, Arabidopsis relatives of the human lysine-specific demethylase1 repress the expression of *FWA* and *FLOWERING LOCUS C* and thus promote the floral transition. Plant Cell 19 (2007) 2975–2987.
[6] S. Shafiq, A. Berr, W.-H. Shen, Combinatorial functions of diverse histone methylations in *Arabidopsis thaliana* flowering time regulation. New Phytol. 201 (2014) 312–322.
[7] A. Berr, S. Shafiq, V. Hamzavi-Pinon, A. Dong, W.-H. Shen, The trxG family histone-methyltransferase SET DOMAIN GROUP 26 promotes flowering via a distinctive genetic pathway. Plant J. 81 (2015) 316–328.
[8] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19 (2003) 185–193.