



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

## Data on RDM16 and STA1 regulate differential usage of exon/intron in RNA directed DNA Methylation pathway

Ravi Datta Sharma<sup>a,b,\*</sup>, Bert Bogaerts<sup>a</sup>, Neha Goyal<sup>c</sup><sup>a</sup> Centre for Microbial and Plant Genetics, KU Leuven, 3000, Belgium<sup>b</sup> Amity institute of Biotechnology and Amity institute of Integrative Sciences and Health, Amity University Haryana, NH-8, Panchgaon, Gurgaon, Haryana 122413, India<sup>c</sup> School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

## ARTICLE INFO

## Article history:

Received 24 January 2017

Received in revised form

16 March 2017

Accepted 31 March 2017

Available online 8 April 2017

## Keywords:

RNA directed DNA methylation pathway

Alternative splicing

RNA-Seq

Arabidopsis thaliana

## ABSTRACT

This article contains data on *RDM16* and *STA1* regulate differential usage of exon/intron in RNA directed DNA Methylation pathway (RdDM) (Sharma et al., 2016) [5]. This data include expression profiles of top 100 genes that has at least one exon or intron differentially expressed in three different contrast, i.e., WT (Wild type) vs *RDM16*, WT vs *STA1*, and *RDM16* vs *STA1*. Also we included the alignment of *MORC6* protein to the ATPase-C family members that have conserved three ATP binding sites and conserved Mg<sup>2+</sup>-binding sites in the spliced exon.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	Bioinformatics, Genomics
More specific subject area	Alternative splicing, Differential Expression
Type of data	Figures, Table, Alignment

DOI of original article: <http://dx.doi.org/10.1016/j.gene.2017.01.027>

\* Corresponding author at: Centre for Microbial and Plant Genetics, KU Leuven, 3000, Belgium.

E-mail address: [ravidattasharma@gmail.com](mailto:ravidattasharma@gmail.com) (R.D. Sharma).<http://dx.doi.org/10.1016/j.dib.2017.03.050>2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data was acquired	Gene Expression Omnibus (GEO) id: GSE44635, URL: <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44635">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44635</a> from the article [1]
Data format	Analyzed (Figs. 1–3, Alignment Figs. 4 and 5, Table 1)
Experimental factors	Secondary analysis of published data
Experimental features	Computational analysis
Data source location	–
Data accessibility	Accessible from this article

---

### Value of the data

- The data from article [5] shows the expression profiles of the genes that contain at least one alternative splicing event in different conditions. This information will be useful for other researchers to understand the regulation of gene expression by alternative splicing.
  - Alignment of *MORC6* protein to the ATPase-C family simplifies the mechanism by which splicing factor RDM16 regulate the *MORC6*.
  - This data provides the information of the genes that are affected in RdDM pathway by knockdown of RDM16 and STA1 splicing factors. This data will help other researcher to validate the findings of the exon/intron level analysis in RdDM pathway.
- 

## 1. Data

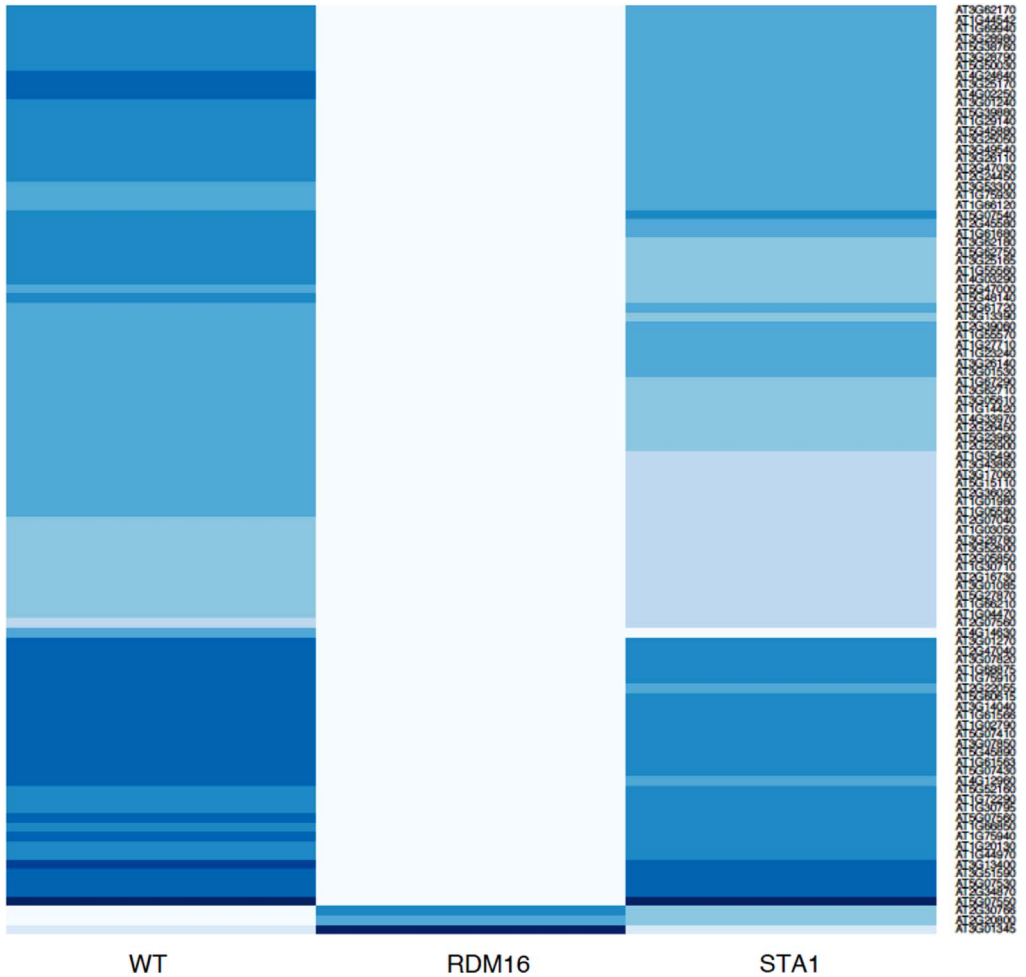
Figs. 1–3 depict expression profile of top 100 genes that has at least one exon or intron differentially expressed in WT vs *RDM16*, WT vs *STA1*, and *RDM16* vs *STA1* respectively. The color key is given with Fig. 3.

Fig. 4: Figure shows the alignment of *MORC6* protein to the ATPase-C family members that have conserved three ATP binding sites at 8, 11 and 14th position of the alignment. There are few more ATP binding sites at 55–65, 104–107, 123–125, 166–169 but may not be contributing in the ATP binding since co-factor binding site is only available in the protein sequence that is coded by exon4 in *MORC6* (region highlighted in yellow).

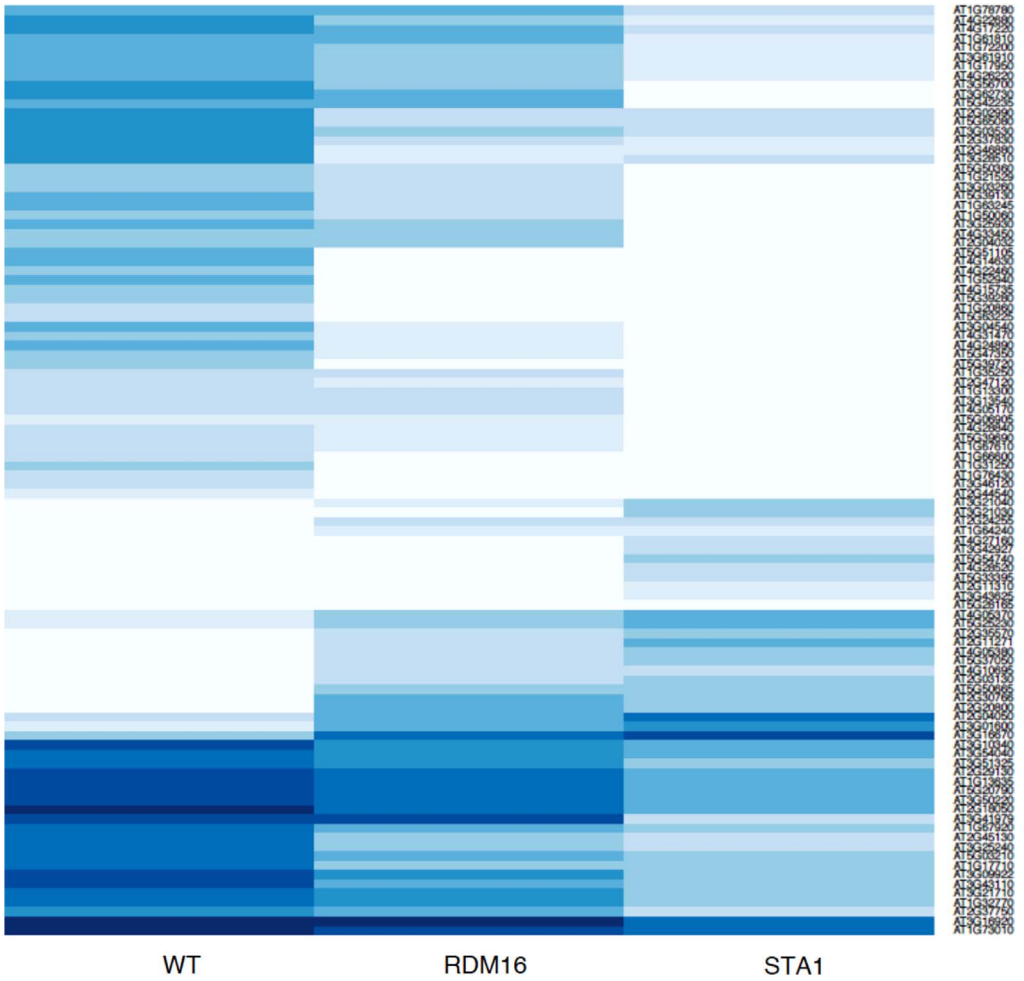
Fig. 5: Figure shows the alignment of *MORC6* protein to the ATPase-C family members that have conserved Mg<sup>2+</sup> binding site at 11th position of the alignment. Highlighted (yellow color) query sequence shows the protein sequence that is coded by exon4 in *MORC6*. ASP (D) and ASN (N) are essential amino acid for Mg<sup>2+</sup> binding but do not contribute in it [7].

## 2. Experimental design, materials and methods

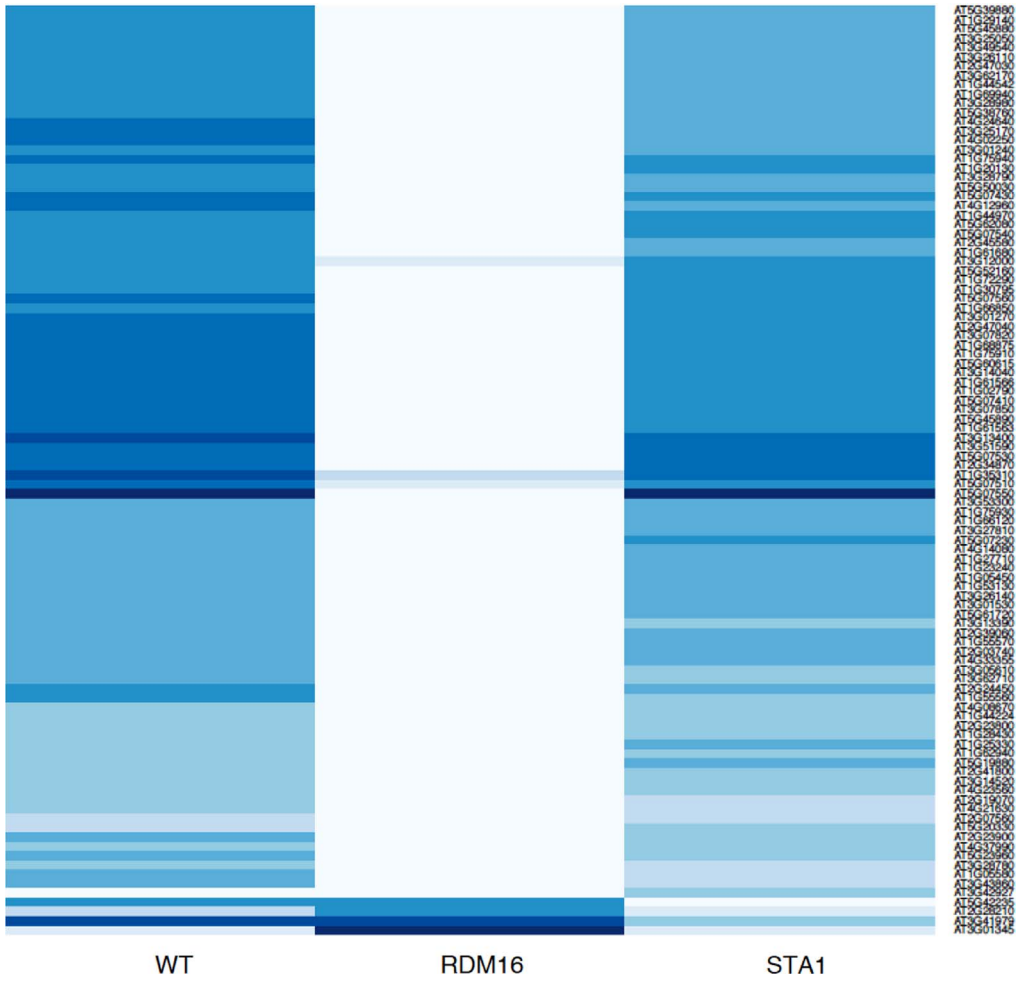
The experiment contains RNA-Seq samples in three conditions; WT (wild type), mutant *RDM16* and *STA1*. The raw data were downloaded from Gene Expression Omnibus (GEO) with accession number GSE44635. The alignment of the reads were done using TopHat2 pipeline [2] (Table 1) and the reads were counted via featurecount function in Rsubread package [4]. We used edgeR in order to find the differentially expressed exons and introns [6]. Figs. 1–3 were prepared using in-built functions in R. The alignment of the *MORC6* protein to ATPase-C family members was done using ClustalX software [3]



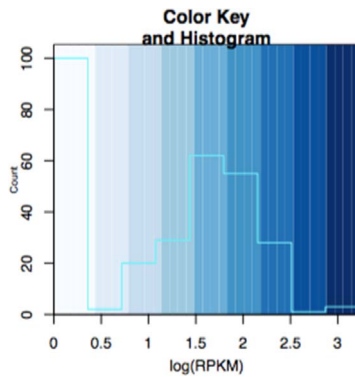
**Fig. 1.** This figure depicts expression profile of top 100 genes that has at least one exon or intron differentially expressed in WT vs RDM16. Color key used in expression profiles of genes in different contrasts is given with Fig. 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** Depicts expression profile of top 100 genes that has at least one exon or intron differentially expressed in WT vs STA1. Color key used in expression profiles of genes in different contrasts is given with Fig. 3.



**Fig. 3.** Depicts expression profile of top 100 genes that has at least one exon or intron differentially expressed in RDM16 vs STA1. The color key is given with figure.



Color key used in expression profiles of genes in different contrasts (for fig. 1, 2 and 3).

```

          10      20      30      40      50      60      70      80
...*...|...*...|...*...|...*...|...*...|...*...|...*...|...*...|
Feature 1
1A4H      37 KEIFLRELISNASDALdkirykslsdpkqletepdLFIRI--TPKpe---qkVLRDSGIGMtkaelinnlgtiaksgt 111
query    129 afgVAELLDNAVDEIqgn-----gaTFVIVdkTTNprd-gatALLIQDDGGGMdpqamrhcmgfgf--- 188
gi 7470847 532 LVELLTKLLDNAIKFTpt-----nGRISI--AVDrpnsqLEVITDTGRGIepnrletvdrfy--- 589
gi 12644105 126 LYKIFDEIIVNADNKVrdp-----nmntLKVTL--DPEa----nVISIYNNGKGIPieihdkekiyipelif 187
gi 6016274 25 KEIFLRELISNASDALdklkyealvdgtykqlhceARIDI--APEed---aqLRVVRDTGIGMnaediranlgtiarsgt 99
gi 17865492 25 KEIFLRELISNASDAIdkiiyralsddsitfnkddyFIKV--TANke---drTLVSDTIGIGMtkeelesnlgtiaksgs 99
gi 1708337 24 KEIFLRELISNASDAIdklklflsltnekfnkialePKIEI--SFdd----kSLLIKDNGIGMddegdltnhlgviaksgt 96
gi 17865496 26 HEIFLREIVSNAVDATqklkltltsv-gefkgetgdLRVTV--SVdev---arTITVSDRGVGMteevekyinqiafssa 99
gi 2495364 29 KDAFLRELISNASDALdklriealrnklevdtsdLHIEI--DADka---arTLVVRDNGIGMmareevvdligtlaksgt 103
gi 1708314 33 KEIFLRELISNASDALdkirfesltdkskldaqpeLFIRL--VPDkt---nkTLSIIDSGVGMakadlvnnlgtiarsgt 107
          90      100     110     120     130     140     150     160
...*...|...*...|...*...|...*...|...*...|...*...|...*...|...*...|
Feature 1
1A4H      112 k----afmealsagadvsmigqfGVGFYSLFLVAD---RVQVISKsnndeqyiwesnagg-----sftvtldevn 174
query    189 -----sdkksdsaigryGNGFKTSTMrl---gaDVIVFSRhsknqtltsqsigllsytyltrtghdrivvpildy 254
gi 7470847 590 -----geegalrrsrggtGIGLAICRQIVsgwggEIWAASDgk----- 627
gi 12644105 188 gn--lltssnyddnqkvvtggrnYGAKLCNIPSt---EFVETAdkermkkykqtwydnm-----srksepvitsl 254
gi 6016274 100 ka--flstltrdqkqdsnligqfGVGFYSAFMVAS---KVEVITKkaentvwkwtsegqayltldevdaaafpvegv 173
gi 17865492 100 l----afkteneskdgghdiigqfGVGFYSAFMVAd---KVTVTKalgeesgyqwestg-----adgytilpi 160
gi 1708337 97 ke--finnlkqdekkasligqfGVGFYSAFIVSe---KVEVTSKkalesdayiwssdg-----ktgyeieka 159
gi 17865496 100 e----eflekykddkaaiighfGLGFYSAFMVSe---RVDVITRsfredatavkwscdg-----speytlepa 160
gi 2495364 104 aelraqreaknaaaseeligqfGIGFYSSFMVAd---KVQLLTRkagesaatrwessg-----egtytiesv 168
gi 1708314 108 k----efmealqagadvsmigqfGVGFYSAYLVAE---KVIVTTKhndeqyiwesqagg-----sftvtrdvvg 170
          170
...*...|...
Feature 1
1A4H      175 erigrGTILRLFL 187
query    255 efnasagefktlq 267
gi 7470847 628 ---nhGTQPHFTV 637
gi 12644105 255 kkpdeYKITFKP 267
gi 6016274 174 aegsaGTCVVLHL 186
gi 17865492 161 akesvGTEIRLKL 173
gi 1708337 160 kkesGTEIKLYL 172
gi 17865496 161 dkadrGTDIVMHI 173
gi 2495364 169 edapqGTSVTLHL 181
-----

```

**Fig. 4.** Alignment of *MORC6* protein to the ATPase-C family members that have conserved three ATP binding sites at 8, 11 and 14th position of the alignment. There are few more ATP binding sites at 55–65, 104–107, 123–125, 166–169 but may not be contributing in the ATP binding since co factor binding site is only available in the protein sequence that is coded by exon 4 in *MORC6* (region highlighted in yellow).



**Fig. 5.** Alignment of *MORC6* protein to the ATPase-C family members that have conserved Mg<sup>2+</sup> binding site at 11<sup>th</sup> position of the alignment. Highlighted (yellow color) query sequence shows the protein sequence that is coded by exon4 in *MORC6*. ASP (D) and ASN (N) are essential amino acid for Mg<sup>2+</sup> binding but do not contribute in it (Jorgensen et al. [7]).

**Table 1**  
Summary of the TopHat2 alignment. (Values are in millions).

Sample	Pairs	Aligned pairs (%)	Multiple alignments (%)	Discordant alignments (%)	Concordant pairs (%)
WT	25.74	24.10 (93.6%)	1.51 (6.2%)	0.02 (0.1%)	24.08 (93.5%)
RDM-16	26.62	24.93 (93.7%)	1.69 (6.7%)	0.02 (0.1%)	24.91 (93.6%)
STA1	26.56	24.90 (93.8%)	2.15 (8.5%)	0.02 (0.1%)	24.88 (93.7%)

**Acknowledgement**

We would like to thank Prof. Vera van Noort, Computational System Biology Group, CMPG, KU Leuven, Belgium, for support and valuable suggestions.

**Transparency document. Supplementary material**

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.03.050>.

## References

- [1] C.F. Huang, D. Miki, K. Tang, H.R. Zhou, Z. Zheng, W. Chen, Z.Y. Ma, L. Yang, H. Zhang, R. Liu, X.J. He, J.K. Zhu, A Pre-mRNA-splicing factor is required for RNA-directed DNA methylation in Arabidopsis, *PLoS Genet.* 9 (9) (2013) e1003779.
- [2] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.* 14 (4) (2013) R36.
- [3] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics*, 23(21), 2007, pp. 2947–2948.
- [4] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* 30 (7) (2014) 923–930. <http://www.ncbi.nlm.nih.gov/pubmed/24227677>.
- [5] R.D. Sharma, B. Bogerts, N. Goyal, RDM16 and STA1 regulate differential usage of exon/intron in RNA directed DNA methylation pathway, *Gene* 609 (2017) 62–67. <http://dx.doi.org/10.1016/j.gene.2017.01.027>.
- [6] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (1) (2010) 139–140.
- [7] P.L. Jorgensen, P.A. Pedersen, Structure–function relationships of Na<sup>+</sup>, K<sup>+</sup>, ATP, or Mg<sup>2+</sup> binding and energy transduction in Na,K-ATPase, *Biochim. Biophys. Acta - Bioenergy* 1505 (1) (2001) 57–74. [http://dx.doi.org/10.1016/S0005-2728\(00\)00277-2](http://dx.doi.org/10.1016/S0005-2728(00)00277-2).