Check for updates

RESEARCH ARTICLE

# Cell-specific cis-natural antisense transcripts (cis-NATs) in the sperm and the pollen vegetative cells of *Arabidopsis thaliana* [version 1; referees: 3 approved]

Peng Qin [iD] [1,2], Ann E. Loraine[3], Sheila McCormick[2]

[1]Rice Research Institute, Sichuan Agricultural University, Chengdu Wenjiang, Sichuan, 611130, China
[2]U.S. Department of Agriculture/Agricultural Research Service and Department of Plant and Microbial Biology, University of California, Berkeley, Albany, CA, 94710, USA
[3]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA

## Abstract

**Background:** cis-NATs **(**cis-natural antisense transcripts**)** are transcribed from opposite strands of adjacent genes and have been shown to regulate gene expression by generating small RNAs from the overlapping region. cis-NATs are important for plant development and resistance to pathogens and stress. Several genome-wide investigations identified a number of cis-NAT pairs, but these investigations predicted cis-NATS using expression data from bulk samples that included lots of cell types. Some cis-NAT pairs identified from those investigations might not be functional, because both transcripts of cis-NAT pairs need to be co-expressed in the same cell. Pollen only contains two cell types, two sperm and one vegetative cell, which makes cell-specific investigation of cis-NATs possible.
**Methods:** We investigated potential protein-coding cis-NATs in pollen and sperm using pollen RNA-seq data and TAIR10 gene models using the Integrated Genome Browser. We then used sperm microarray data and sRNAs in sperm and pollen to determine possibly functional cis-NATs in the sperm or vegetative cell, respectively.
**Results:** We identified 1471 potential protein-coding cis-NAT pairs, including 131 novel pairs that were not present in TAIR10 gene models. In pollen, 872 possibly functional pairs were identified. 72 and 56 pairs were potentially functional in sperm and vegetative cells, respectively. sRNAs were detected at 794 genes, belonging to 739 pairs.
**Conclusion:** These potential candidates in sperm and the vegetative cell are tools for understanding gene expression mechanisms in pollen.

## Keywords
Pollen, Sperm, Cis-Natural antisense transcript, Arabidopsis

**Open Peer Review**

**Referee Status:** ✓ ✓ ✓

| | Invited Referees | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| **version 1** published 22 Jan 2018 | ✓ report | ✓ report | ✓ report |

1 **Hailing Jin**, University of California, Riverside, USA

2 **Robert A. Martienssen**, Cold Spring Harbor Laboratory, USA
**Cristiane Alves**, Cold Spring Harbor Laboratory, USA

3 **Jules Deforges**, Paris Descartes University, France

**Discuss this article**

Comments (0)

**Corresponding author:** Sheila McCormick (sheilamc@berkeley.edu)

**Author roles: Qin P**: Formal Analysis, Investigation, Writing – Original Draft Preparation; **Loraine AE**: Investigation, Resources, Software; **McCormick S**: Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

## Introduction

Natural antisense transcripts (NATs) are endogenous transcripts that contain sequences complementary to each other. NATs have been shown to regulate gene expression by generating small RNAs from the overlapping region (Zhang *et al.*, 2013). NATs are classified into two subgroups according to the site of their biogenesis: trans-NATs and cis-NATs. Trans-NATs are transcribed from different genomic loci, whereas cis-NATs are transcribed from opposite strands of adjacent genes (Jin *et al.*, 2008). Based on the relative orientation and overlap degree of two transcripts, cis-NATs can be categorized into three types: head-to-head (5′ to 5′), tail-to-tail (3′ to 3′) and fully overlapping (Jin *et al.*, 2008). cis-NATs are widely present in plants, animals and fungi (Zhang *et al.*, 2013). In plants, cis-NATs are important for pathogen resistance (Katiyar-Agarwal *et al.*, 2006), stress tolerance (Borsani *et al.*, 2005), successful fertilization (Ron *et al.*, 2010), and phosphate homeostasis and plant fitness (Jabnoune *et al.*, 2013).

Several genome-wide investigations identified potential cis-NATs in plants, ranging from 1057 to 1710 pairs in *Arabidopsis* (Henz *et al.*, 2007; Jin *et al.*, 2008; Wang *et al.*, 2005; Zhan & Lukens, 2013), and 3819 pairs in rice (Lu *et al.*, 2012). However, all the expression data used in these cis-NAT investigations were from bulk samples such as seedlings, leaves or inflorescences, which include many cell types. For potential cis-NATs to be functionally relevant, the reverse and complementary transcripts must be co-expressed in the same cell. Some potential cis-NATs identified in those investigations might be expressed in different cells and, thus, the presence of overlapping transcripts in the same cell is not known. Moreover, the regions predicted to overlap in previous investigations were based on available gene model annotations, which might not fully represent potential overlaps, due to alternative splicing at different developmental stages or, to more extensive 5' or 3' UTRs than are annotated.

A pollen grain contains only two types of cell, one vegetative cell and two sperm cells, and thus can be used to investigate cell-specific cis-NAT pairs. Pollen RNA-seq data (Loraine *et al.*, 2013) provides accurate transcript lengths, helpful for precisely identifying overlapping regions of two adjacent genes. Sperm microarray data (Borges *et al.*, 2008) is helpful for defining whether two adjacent genes are expressed in the same cell. Moreover, a small RNA database for pollen and sperm (Slotkin *et al.*, 2009) can be used to determine if small RNAs were detected from any overlapping regions.

In this study, we investigated potential cis-NATs in *Arabidopsis* sperm and vegetative cells using pollen RNA-seq data, sperm microarray data and sRNA data in pollen and sperm. In total, we identified 1471 cis-NAT pairs, including 131 novel pairs, with 72 and 56 pairs being potentially functional cis-NATs in sperm and vegetative cells respectively. These cis-NATs are tools for understanding gene regulation mechanisms in sperm and vegetative cells.

## Methods

We investigated potential cis-NATs from protein-coding genes in pollen and sperm using pollen RNA-seq data (Borges *et al.*, 2008; Loraine, *et al.*, 2013) and TAIR10 gene models (TAIR10 gene annotation data available here) using the Integrated Genome Browser, available from http://www.bioviz.org (Freese *et al.*, 2016). We loaded TAIR10 gene models s and pollen RNA-seq into IGB, then manually scanned for cis-NATs in each chromosome, based on the following parameters: 1) the orientation of two adjacent genes in TAIR10 was reverse and complementary; 2) the length of transcripts mapping to the overlap of two adjacent genes was larger than 21nt, because the size of sRNA generated by cis-NATs is normally larger than 21 nt; and 3) both adjacent genes encoded proteins. The expression and sRNA data of each cis-NAT was merged with the cis-NAT data in Excel to generate sheet 1 of Table S1. Different categories of cis-NATs in sheets 2–8 of Table S1 were obtained based on the cis-NAT data in sheet 1 of Table S1.

## Results

We identified 1471 potential cis-NAT pairs, comprising 1373 pairs whose transcripts were complementary at their 3' ends, and 98 pairs whose transcripts were complementary at their 5' ends (Figure 1A, sheet 1 and 2 of Table S1). Among these 1471, in 37 pairs one transcript was completely internal to the other, 100 pairs comprised 50 sets that had three overlapping genes (sheet 1 of Table S1), and 131 pairs (8.9%) were not apparent using the TAIR 10 gene models, but were detected in the pollen RNA-seq data (Figure 1B and G, sheet 1 of Table S1).

One criterion for functional cis-NATs is that the two adjacent genes of a cis-NAT pair are expressed. To identify potentially functional cis-NATs in pollen, we analyzed the expression level of the 1471 gene pairs, defining genes with reads per million (RPM) ≥1 as expressed. There were 599 pairs for which the RPM of two adjacent genes was lower than 1 (sheet 3 of Table S1), suggesting that those 599 pairs might not produce relevant cis-NATs in pollen. Most sperm-specific genes, such as *GEX1*, *GEX2* (Rutley & Twell, 2015), and *Kokopelli* (Ron *et al.*, 2010), are detectable in pollen RNA-seq data, but sperm-specific genes with lower expression levels, such as *ARI14*, might not be detected, as the proportion of RNA from the vegetative cell is much larger than that from the sperm cells. So potential cis-NATs with one expressed gene in pollen RNA-seq data might still be functional in pollen. Based on this, there were 872 possibly functional cis-NATs pairs, for which the RPM of either or both adjacent genes was ≥1 (Figure 1C, sheet 4 of Table S1), of which 62 pairs did not overlap in TAIR10 gene models. Note that we did not detect a cis-NAT pair between *Kokopelli* and *ARI14* (Ron *et al.*, 2010), because the TAIR 10 gene model does not show them overlapping, and *ARI14* expression is low in wild type. Thus it is possible that other cis-NATs might similarly not be included in sheet 4 of Table S1 (see below).

Another criterion for functional cis-NATs is that both adjacent genes are expressed in the same cell. In order to accurately
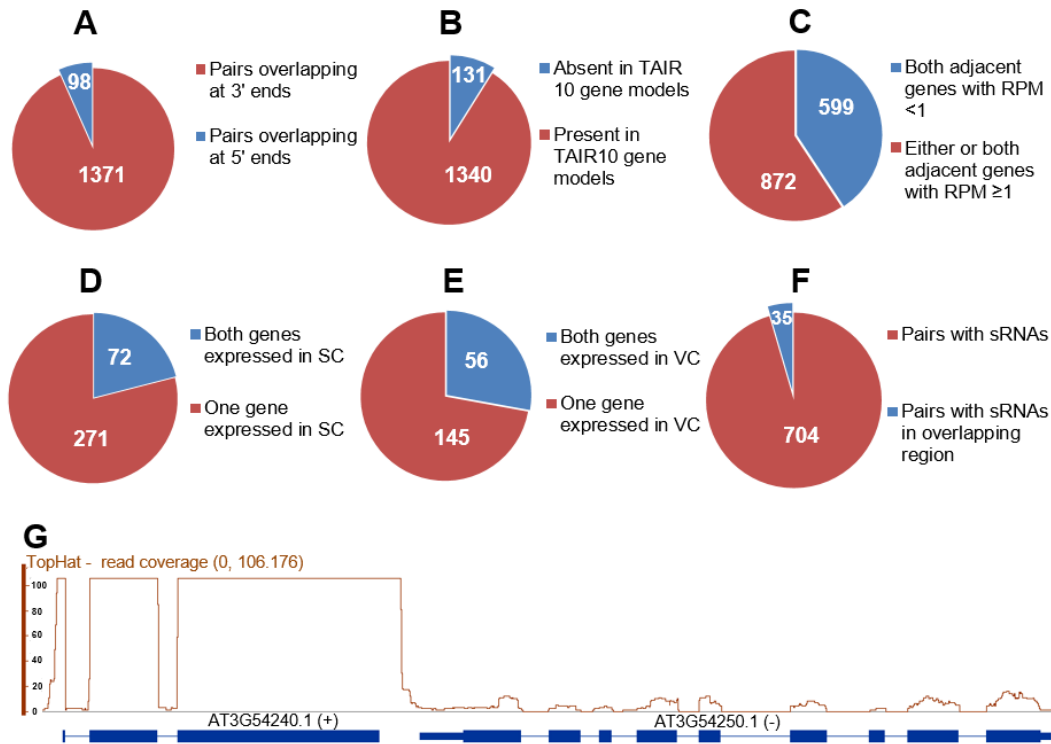
**Figure 1. Overview of potential cis-NATs in sperm and the pollen vegetative cell.** (**A**) Overview of potential cis-NATs with different overlapping types. (**B**) Overview of potential cis-NATs present or non-present in TAIR 10 gene models. (**C**) Overview of potential cis-NATs based on reads per million (RPM) from pollen RNA-seq data. (**D**) Overview of potential cis-NATs possibly functional in sperm cells (SC). (**E**) Overview of potential cis-NATs possibly functional in the vegetative cell (VC). (**F**) Overview of potential cis-NATs with sRNAs in vegetative and sperm cells.(**G**) Potential cis-NATs not present in TAIR10 gene models, but detected in pollen RNA-seq data.

identify potentially functional cis-NATs in pollen, we investigated the gene expression level of 872 pairs in sperm cells. The microarray signal calls "Present (P) or Absent (A)" (Borges *et al.,* 2008) were used to categorize genes as expressed or not expressed, respectively. We identified 72 pairs for which both adjacent genes were expressed in sperm, supporting the likelihood that these 72 cis-NATs pairs exist in sperm (sheet 5 of Table S1). There were an additional 271 pairs for which only one of the adjacent genes was expressed in sperm (Figure 1D and sheet 6 of Table S1), some of which might pertain to the *Kokopelli/ARI14* example. To test if these pairs might function in the vegetative cell, we defined the gene as expressed in the vegetative cell if either 1): both pollen and sperm signals were called "P", and the ratio of the pollen to sperm signal was > 3; or 2): the pollen signal was "P" and sperm signal was "A". This exercise yielded 56 pairs for which both adjacent genes were expressed in the vegetative cell (sheet 7 of Table S1), and 145 pairs for which only one was expressed in the vegetative cell (Figure 1E and sheet 8 of Table S1). Another hallmark of func-

tional cis-NATs is that there are small RNAs generated from the overlapping region. We therefore investigated the sRNAs of pollen and sperm (Slotkin *et al.,* 2009) for these potentially functional cis-NAT pairs. sRNAs were detected at 794 genes, belonging to 739 pairs (Figure 1F and sheet 1 of Table S1). Of these, 35 cis-NATs pairs had sRNAs from the overlapping region.

## Discussion

cis-NATs are widely present in plants, and play an important role in regulating gene expression. However, in plants precise identification of cis-NATs at a cell-specific level to support whether cis-NATs might be functional is lacking. One possible reason was that it is difficult to get specific cell types for RNA-seq. As pollen grains contain only two types of cell, they are an excellent model to investigate cell-specific cis-NATs.

The vegetative cell forms a pollen tube that transports two sperm cells into the ovule for fertilization. Successful fertiliza-

tion needs proper gene regulation in both the vegetative and sperm cells (Ron *et al.,* 2010). The precise identification of cis-NATs in pollen is a tool for understanding the molecular mechanism of pollen tube growth and fertilization. The 131 novel potential cis-NATs in pollen, and the 72 and 56 potentially functional cis-NATs in sperm and the vegetative cell, respectively, provide candidates toward further uncovering the regulatory mechanisms of gene expression in the sperm and vegetative cells.

## Conclusion
We identified 1471 potential cis-NAT pairs, including 131 pairs only detected in the pollen RNA-seq data. There were 872 pairs expressed in the same cell and thus possibly functional in pollen, while 72 and 56 pairs were potentially functional in sperm and vegetative cell, respectively.

## Data availability
*Arabidopsis* pollen RNA-seq alignments data were loaded into Integrated Genome Browser from the IGB Quickload site

Unprocessed sequence data are available from the Sequence Read Archive under accession SRP022162 (Loraine *et al.,* 2013).

The pollen and sperm microarray data was from (Borges *et al.,* 2008) (available at https://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-9-87)

sRNAs in *Arabidopsis* pollen and sperm were downloaded from https://mpss.danforthcenter.org/dbs/index.php?SITE=at_sRNA&lib_type=sRNA&lib_id=487 and https://mpss.danforth-center.org/dbs/index.php?SITE=at_sRNA&lib_type=sRNA&lib_id=489, respectively (Slotkin *et al.,* 2009).

## Supplemental material
**Table S1.** Investigation of cis-NATs in pollen and sperm cells of *Arabidopsis*.

Click here to access the data.

Sheet 1) List of all potential cis-NATs identified using TAIR 10 gene models and pollen RNA-seq data; Sheet 2) List of potential cis-NAT pairs with three overlapping genes; Sheet 3) List of potential cis-NATs pairs for which the RPM of both adjacent genes is < 1; Sheet 4) List of potential cis-NATs pairs for which the RPM of either or both adjacent genes is ≥ 1; Sheet 5) List of potential cis-NAT pairs for which both adjacent genes are expressed in sperm cells; Sheet 6) List of potential cis-NAT pairs for which only one gene is expressed in sperm cells; Sheet 7) List of potential cis-NAT pairs for which both adjacent genes are expressed in the vegetative cell; Sheet 8) List of potential cis-NAT pairs for which only one gene is expressed in the vegetative cell.

## References

Borges F, Gomes G, Gardner R, *et al.*: **Comparative transcriptomics of Arabidopsis sperm cells.** *Plant Physiol.* 2008; **148**(2): 1168–1181.
PubMed Abstract | Publisher Full Text | Free Full Text

Borsani O, Zhu J, Verslues PE, *et al.*: **Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*.** *Cell.* 2005; **123**(7): 1279–1291.
PubMed Abstract | Publisher Full Text | Free Full Text

Freese NH, Norris DC, Loraine AE: **Integrated genome browser: visual analytics platform for genomics.** *Bioinformatics.* 2016; **32**(14): 2089–2095.
PubMed Abstract | Publisher Full Text | Free Full Text

Henz SR, Cumbie JS, Kasschau KD, *et al.*: **Distinct expression patterns of natural antisense transcripts in Arabidopsis.** *Plant Physiol.* 2007; **144**(3): 1247–1255.
PubMed Abstract | Publisher Full Text | Free Full Text

Jabnoune M, Secco D, Lecampion C, *et al.*: **A rice *cis*-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness.** *Plant Cell.* 2013; **25**(10): 4166–4182.
PubMed Abstract | Publisher Full Text | Free Full Text

Jin H, Vacic V, Girke T, *et al.*: **Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis.** *BMC Mol Biol.* 2008; **9**: 6.
PubMed Abstract | Publisher Full Text | Free Full Text

Katiyar-Agarwal S, Morgan R, Dahlbeck D, *et al.*: **A pathogen-inducible endogenous siRNA in plant immunity.** *Proc Natl Acad Sci U S A.* 2006; **103**(47): 18002–18007.
PubMed Abstract | Publisher Full Text | Free Full Text

Loraine AE, McCormick S, Estrada A, *et al.*: **RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing.** *Plant Physiol.* 2013; **162**(2): 1092–1109.
PubMed Abstract | Publisher Full Text | Free Full Text

Lu T, Zhu C, Lu G, *et al.*: **Strand-specific RNA-seq reveals widespread occurrence of novel *cis*-natural antisense transcripts in rice.** *BMC Genomics.* 2012; **13**: 721.
PubMed Abstract | Publisher Full Text | Free Full Text

Ron M, Alandete-Saez M, Eshed Williams L, *et al.*: **Proper regulation of a sperm-specific *cis*-nat-siRNA is essential for double fertilization in *Arabidopsis*.** *Genes Dev.* 2010; **24**(10): 1010–1021.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rutley N, Twell D: **A decade of pollen transcriptomics.** *Plant Reprod.* 2015; **28**(2): 73–89.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Slotkin RK, Vaughn M, Borges F, *et al.*: **Epigenetic reprogramming and small RNA silencing of transposable elements in pollen.** *Cell.* 2009; **136**(3): 461–472.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Wang XJ, Gaasterland T, Chua NH: **Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*.** *Genome Biol.* 2005; **6**(4): R30.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Zhan S, Lukens L: **Protein-coding cis-natural antisense transcripts have high and broad expression in Arabidopsis.** *Plant Physiol.* 2013; **161**(4): 2171–2180.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Zhang X, Lii Y, Wu Z, *et al.*: **Mechanisms of small RNA generation from *cis*-NATs in response to environmental and developmental cues.** *Mol Plant.* 2013; **6**(3): 704–715.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

# Open Peer Review

## Current Referee Status: ✓ ✓ ✓

**Version 1**

Referee Report 16 May 2018

✓ **Jules Deforges**

CNRS (French National Center for Scientific Research), UMR 8015, Laboratory of Crystallography and NMR Biology, Paris Descartes University, Paris, France

In this study, the authors have addressed an important question often neglected regarding the regulation of gene expression by cis-Natural Antisense Transcripts. Cis-NATs usually identified by sequencing of tissues containing many different cell types are assumed to potentially regulate their cognate coding genes co-expressed in the same tissue. However, direct evidence that both transcripts of the pairs are actually expressed in the same cells is often missing.

By taking advantage of RNAseq, micro-array as well as sRNA data from pollen, a tissue containing only 2 cell types, the authors identified 72 cis-NAT pairs for which both transcripts were expressed in Sperm cells and 56 in vegetative cells. In addition, they identified sRNA falling in the overlapping region of 35 cis-NAT pairs out of 739 in pollen, providing insight about the functional role of these cis-NATs.

Qin et al. provide useful data about cis-NAT expression in sperm and vegetative cells in pollen. They also provide insight about their potential function in sRNA production, contributing to a better understanding of the mechanisms of gene regulation by cis-NATs.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 10 May 2018

**doi:**10.5256/f1000research.14444.r33361

Robert A. Martienssen [1], Cristiane Alves [2]

[1] Department of Plant Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA
[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

The have authors addressed an interesting question regarding gene regulation in vegetative and sperm cells in pollen by cis-natural antisense transcripts (cis-NATs). Qin et al. explain how cis-NATs are generated and their importance for plant development, pathogen resistance, stress response, homeostasis and fertilization. The main question for this work regards the discrepancy that may be occurring in previously published data, as most analysis were performed on bulked samples containing different types of cells leading to possible misassignment of cis-NAT pairing. cis-NATs should be co-expressed in the same cell, as the authors described, so the analysis should be done at the least on the same cell type. To address this issue, the authors used RNA-seq and microarray publicly available data from vegetative and sperm cells from *Arabidopsis thaliana* pollen in order to investigate cell-specific cis-NATs pairs, using specific parameters avoiding possible mispairing. The authors identified possibly functional cis-NATs in pollen, 56 pairs in the vegetative cell and 72 in sperm cells, also, the authors screened for small RNAs generated from the pairing regions, finding that 35 out of 739 pollen cis-NATs from the overlapping region, suggesting a regulatory function. It would be interesting to summarize in the text if any of these were in sperm vs VN.

Function analysis is yet to be addressed, as the authors pointed out, to confirm the biological role of cis-NATs in plant development, stress response, pathogen resistance, homeostasis and fertilization. But the authors choice to survey pollen grains where cell specificity is known could be key to understanding molecular mechanisms regarding fertilization and heredity.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Referee Expertise:* Epigenetics and small RNA in plants

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 21 March 2018

**Hailing Jin**
Department of Plant Pathology and Microbiology, Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, USA

This manuscript by Qin et al. addresses an important question in the gene regulation field: How cis-natural antisense transcripts (cis-NATs) are co-regulated within the same cell. Although many genome-wide studies on cis-NATs have been reported in different species, most expression analysis was based on data from bulk samples that include many different cell types. The conclusions from those studies may not be entirely accurate because it is not clear if the two transcripts of each cis-NAT pair indeed express within the same cell. The authors of this study used a unique system – pollen for the analysis because pollen only has two cell types, sperm and vegetative cells, which allows easy identification of NATs that expressed in the same cell type. They identified 872 potential functional cis-NAT pairs in pollen. Microarray expression analysis identified 72 and 56 pairs were potentially functional in sperm and vegetative cells, respectively, as both adjacent genes were expressed. In addition, 271 pairs and 145 pairs for which only one was expressed in sperm or vegetative cell, respectively. It is possible that one of the transcripts within a pair that is expressed may suppress the expression of the other transcript in the same cell. Furthermore, 739 pairs of cis-NATs generated sRNAs and 35 cis-NATs pairs had sRNAs from the overlapping region, suggesting that sRNAs play a role in the co-regulation of NAT genes.

Co-regulation analysis of cis-NATs within the same cell is an important but often overlooked criteria, this study provided an excellent example to take the localization of gene expression into consideration, which not only identified a set of pollen-specific functional cis-NATs and sRNAs for future study, but also help us understand the regulatory mechanism of cis-NATs. Thus, I support this manuscript to be indexed without delay.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research