



Genome-wide analysis reveals the crucial role of lncRNAs in regulating the expression of genes controlling pollen development

Neeta Lohani^{1,2} · Agnieszka A. Golicz^{1,3} · Annapurna D. Allu^{1,4} · Prem L. Bhalla¹ · Mohan B. Singh¹

Received: 2 October 2022 / Accepted: 25 November 2022 / Published online: 18 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Key message The genomic location and stage-specific expression pattern of many long non-coding RNAs reveal their critical role in regulating protein-coding genes crucial in pollen developmental progression and male germ line specification.

Abstract Long non-coding RNAs (lncRNAs) are transcripts longer than 200 bp with no apparent protein-coding potential. Multiple investigations have revealed high expression of lncRNAs in plant reproductive organs in a cell and tissue-specific manner. However, their potential role as essential regulators of molecular processes involved in sexual reproduction remains largely unexplored. We have used developing field mustard (*Brassica rapa*) pollen as a model system for investigating the potential role of lncRNAs in reproductive development. Reference-based transcriptome assembly performed to update the existing genome annotation identified novel expressed protein-coding genes and long non-coding RNAs (lncRNAs), including 4347 long intergenic non-coding RNAs (lincRNAs, 1058 expressed) and 2,045 lncRNAs overlapping protein-coding genes on the opposite strand (lncNATs, 780 expressed). The analysis of expression profiles reveals that lncRNAs are significant and stage-specific contributors to the gene expression profile of developing pollen. Gene co-expression networks accompanied by genome location analysis identified 38 cis-acting lincRNA, 31 cis-acting lncNAT, 7 trans-acting lincRNA and 14 trans-acting lncNAT to be substantially co-expressed with target protein-coding genes involved in biological processes regulating pollen development and male lineage specification. These findings provide a foundation for future research aiming at developing strategies to employ lncRNAs as regulatory tools for gene expression control during reproductive development.

Keywords Non-coding · lncRNAs · Pollen development · *Brassica rapa* · Transcriptome

Introduction

Pollen grains, the male gametophyte of flowering plants are produced in anthers, the male reproductive organs of flowers. The formation of mature viable pollen is the culmination of a highly specialised and strictly regulated developmental gene expression program (Borg et al. 2009; Haerizadeh et al 2006). Pollen/microspore mother cells (also known as meiocytes) undergo meiosis to form tetrads of haploid microspores, which then divide mitotically and differentiate, giving rise to the sperm cell-carrying mature pollen. The stages of pollen development are well defined with stage-specific markers making it an ideal system for studying plant developmental processes (Brownfield et al. 2009).

Recently, long non-coding RNAs (lncRNAs) have emerged as important, stage-specific regulators of developmental processes in animals and plants (Golicz et al. 2018a, b; Perry and Ulitsky 2016). No lncRNA conservation

Communicated by Kinya Toriyama.

✉ Mohan B. Singh
mohan@unimelb.edu.au

¹ Plant Molecular Biology and Biotechnology Laboratory, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Melbourne, VIC, Australia

² Present Address: School of Science, Western Sydney University, Richmond, Australia

³ Present Address: Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Gießen, Gießen, Germany

⁴ Present Address: Department of Biology, Indian Institute of Science Education and Research, Tirupati, India

between plants and animals has been reported. Still it is postulated that lncRNAs can be universal regulators of developmental processes and that their similar functions and mechanisms of action could be a result of convergent evolution (Golicz et al. 2018a, b). lncRNAs are RNA molecules with more than 200 base pairs in length, lack open reading frames more than 100 amino acids long, and have no protein-coding potential. The discretionary length limit defining lncRNAs distinguishes them from small non-coding RNAs, including microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), and small interfering RNAs (siRNAs). lncRNAs, which are primarily intergenic ncRNAs (lincRNAs), intronic ncRNAs (incRNAs), or natural antisense transcripts (NATs), often show polyadenylation and tend to have highly tissue-specific expression (Mattick and Rinn 2015). They act as decoys, molecular scaffolds, or target mimics of miRNAs and siRNA precursors to influence gene expression (Franco-Zorrilla et al. 2007; Wu et al. 2013). When acting as decoys, certain lncRNAs can bind with transcription factors, thereby precluding their interaction with DNA to promote the expression of target genes, while as molecular scaffolds, they can bind with DNA or protein-recruiting regulatory components to specific gene loci (Franco-Zorrilla et al. 2007; Wang and Chang 2011; Wu et al. 2013).

Several reports have highlighted the critical role of lncRNAs in plant biological processes such as stress response, development regulation and nutrient procurement by regulating modification of histones, transcription, alternative splicing, chromatin remodelling or target mimicry (Böhmendorfer and Wierzbicki 2015; Di et al. 2014; Li et al. 2016; Mattick and Rinn 2015; Yu et al. 2013; Yuan et al. 2016; Zhang et al. 2014). In Arabidopsis, during cold exposure, an antisense transcript—*COOLAIR* (*Cold Induced Long Antisense Intragenic RNA*)—and an intronic lncRNA—*COLDAIR* (*COLD ASSISTED INTRONIC NONCODING RNA*)—restrict the transcriptional activation of the floral repressor *FLOWERING LOCUS C* (*FLC*) via histone modification and thereby promote flowering (Csorba et al. 2014; Heo and Sung 2011; Rosa et al. 2016). Similarly, another cold-induced natural antisense lncRNA, *MAS* (*MAF4 antisense RNA*), is reported to direct the activation of *MADS AFFECTING FLOWERING4* (*MAF4*) via histone modification resulting in the suppression of early flowering in Arabidopsis (Zhao et al. 2018). In rice, the silencing of an antisense lncRNA—*LRK Antisense Intergenic RNA* (*LAIR*)—results in reduced plant growth along with reduced expression of *LEUCINE-RICH REPEAT SERINE/THREONINE-PROTEIN KINASE* (*LRK*) gene cluster (Wang et al. 2018). Lines overexpressing *LAIR*, on the other hand, show a significant increase in overall grain yield and increased expression of some members of the *LRK* gene cluster. It was reported that in rice, *LAIR* could variably activate the promoters of the *LRKs* gene by

binding to histone modification enriched in the *LRK1* gene area (Wang et al. 2018).

In plants, lncRNAs have also been linked to male reproductive development. In rice, under transcription of long-day conditions, long-day specific male-fertility-associated RNA, *LDMAR* is required for photoperiod-sensitive male sterility (PSMS) activation and proper pollen formation (Ding et al. 2012, Babaei et al. 2022). In young panicles of rice, overexpression of *LDMAR* impairs fertility under long-day conditions. In maize, high expression of lncRNA *Zm401* was observed in developing male gametophytes and mature pollen, and it was identified as the primary regulator of genes essential for pollen formation, such as *ZmC5*, *ZmMADS2*, and *MZm3-3* (Ma et al. 2008). Downregulation of *Zm401* leads to aberrant tapetum and microspore development, resulting in the production of sterile pollen. Furthermore, in Chinese cabbage (*Brassica campestris* L.), a novel pollen-specific lncRNA *BcMF11* was identified to regulate male reproductive development (Song et al. 2013, 2007). The silencing of *BcMF11* resulted in delayed tapetum degradation, abnormal microspore development and pollen abortion. These findings demonstrated that lncRNAs are essential for regulating pollen formation.

Here, we performed a genome-wide identification of lncRNAs during five stages (pollen mother cell, tetrad, microspore, bicellular pollen and mature pollen) of pollen development in field mustard (*Brassica rapa*) using strand-specific RNA sequencing (ssRNA60 Seq). lncRNAs exhibit stage-specific expression suggesting potential roles at well-defined developmental points. Next, we analysed the genomic location of lncRNAs and predicted *cis* and *trans*-acting lncRNAs and their potential target protein-coding genes. Differential expression and functional enrichment analysis highlighted the complex transcriptional reprogramming involved in the transition of diploid pollen/microspore mother cells into haploid trinucleate pollen. We further performed a weighted gene co-expression network analysis (WGCNA) coupled with gene expression correlation to identify lncRNA–mRNA pairs with a potential role in regulating pollen development progression. Collectively, our findings shed light on the roles of lncRNAs during pollen development and expand our knowledge of the molecular mechanisms underlying male reproductive development.

Results

Identification and characterisation of lncRNAs in *B. rapa* expressed during pollen development

Strand-specific RNA-Seq sequencing reads corresponding to five stages of pollen development (pollen mother cell—‘PMC’, tetrad—‘TET’, microspore—‘MIC’, binucleate

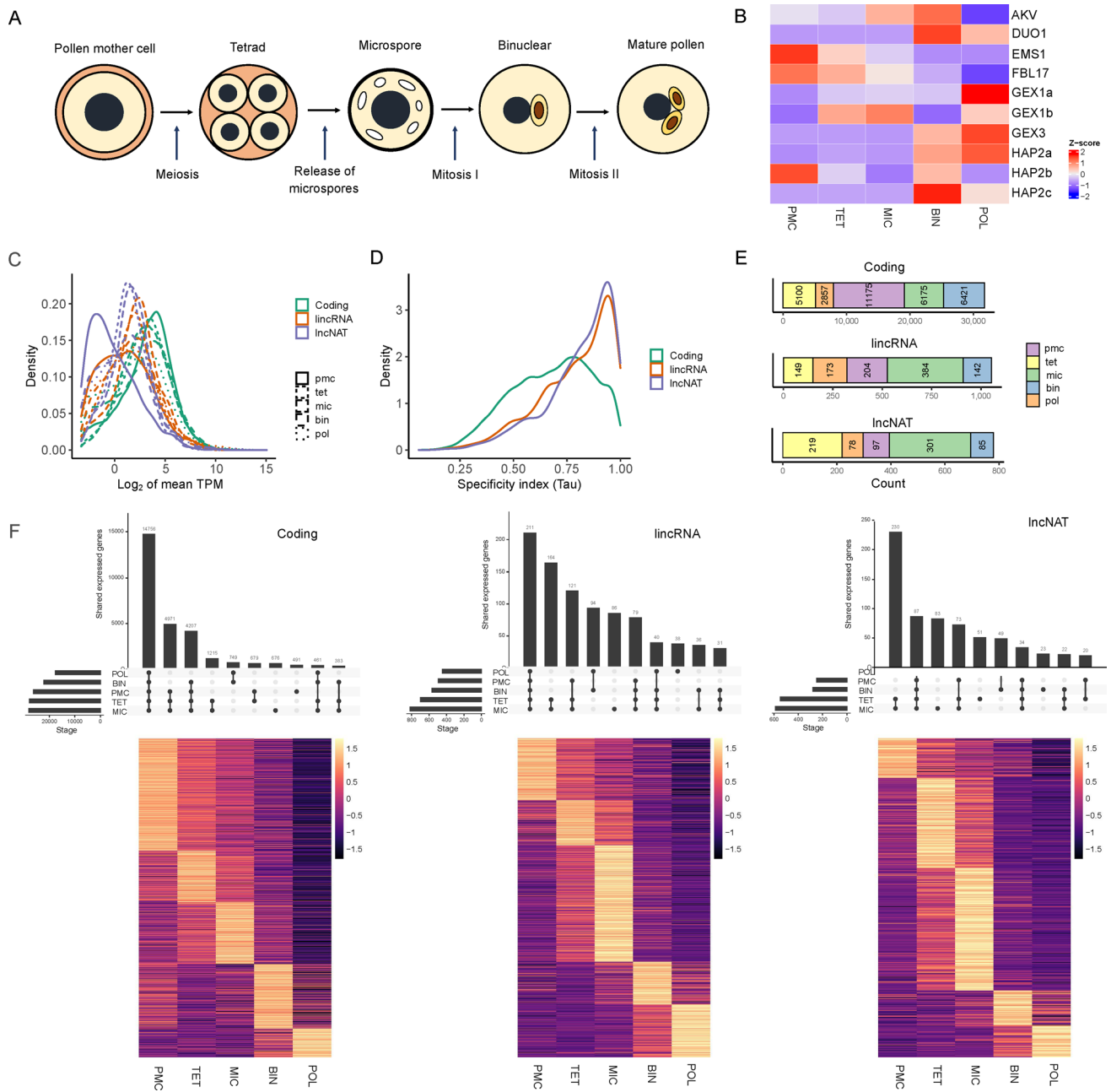


Fig. 1 Male gametophyte development and properties of the lncNATs and lincRNAs discovered. **A** The five stages of pollen development, **B** heat map of *B. rapa* homologs of known pollen development marker genes, **C** distribution of expression values for coding genes and lncRNAs, **D** expression specificity index for protein-coding genes and lncRNAs, **E** summary of the number of coding and lncRNA

genes showing peak expression at a given stage, and **F** heat maps and Upset plots presenting overall expression patterns of coding genes lncNATs and lincRNAs across the five stages. *PMC* pollen/microspore mother cell, *TET* tetrads, *MIC* microspores to polarised microspores, *BIN* early to late binucleate pollen, *POL* trinucleate pollen

pollen—‘BIN’ and trinucleate pollen—‘POL’), were used to track changes in gene expression during male gametophyte development in *B. rapa* (Fig. 1A). Both poly(A) capture and ribosomal RNA (rRNA) depletion libraries were prepared. The reads were aligned to the *Brassica rapa* genome with a mapping rate for poly(A) capture libraries

between 78.18 and 90.85% (mean: 87.34%) and for the rRNA depletion libraries between 71.02 and 83.22% (mean: 76.74%; Table S1A). Because pollen development requires the participation of highly specialised tissues and cell types, some of the genes involved may not be found in the existing annotation. A reference-based (Zhang et al.

2022) transcriptome assembly was performed to update the existing genome annotation (using an in-house pipeline, Figure S1, (Golicz 2022)), identify novel expressed protein-coding genes and long non-coding RNAs (lncRNAs), including long intergenic non-coding RNAs (referred to as ‘lincRNAs’ hereafter) and lncRNAs overlapping protein-coding genes on the opposite strand (referred to as ‘lncNATs’ hereafter). In total, 49,577 protein-coding genes, 4347 lincRNAs and 2,045 lncNATs were identified. Comparison of the poly(A) capture and rRNA depletion libraries (TPM < 0.1 for all the poly(A) libraries and TPM > 0.1 in at least one rRNA depletion library) suggests that 1.3, 4.3 and 1.7% of loci produce non-polyadenylated transcripts for coding lincRNA and lncNAT genes, respectively. Principal component analysis (PCA) revealed high relatedness between the replicates of each sample (Figure S2A). Further, the Pearson correlation between the three biological replicates ranged from 0.911 to 0.989 (median: 0.968). The correlation between the coding genes, lincRNAs and lncNATs, was also significant across the five pollen developmental stages (Figure S2B). We have tested the concordance between expression observed in this dataset and the previously reported expression patterns of known male development markers in *Arabidopsis thaliana*. All the markers, other than AtMGH3 and AtGEX2, for which no confident orthologues were identified, had expected expression patterns (Fig. 1B and S2C).

Lastly, based on data mean–variance trend analysis, genes with low expression were filtered, and a CPM cutoff (> 1.0 CPM in at least three samples) was imposed, identifying 31,729 coding genes, 1,052 lincRNA and 780 lncNAT loci available for the analysis. A comparison of the protein-coding and lncRNA loci confirms that the latter have lower expression levels and more stage-specific expression (Fig. 1C and Fig. 1D), with different expression profiles of coding genes and lncRNAs. Among the samples used in this study, the highest number of protein-coding genes (35.22%) had peak expression in PMC and lincRNAs (36.50%), and lncNATs (38.59%) had peak expression in MIC (Fig. 1E, F). It is important to note that the peak expression stage has been defined as the stage with maximum gene expression measured by TPM (transcripts per million). Therefore, the peak expression stage is the stage where transcript abundance is the highest relative to the abundance of other transcripts at that stage.

The lncRNAs were shorter than coding genes with ~ 80 and 40% lncRNAs with one transcript and only one exon, respectively (Fig. 2A–C). Compared to lincRNAs, lncNATs had slightly higher proportion of lncNATs genes that had one transcript (lincRNAs: 78.61%, lncNATs: 82.18%) and multiple exon (lincRNAs: 52.96%, lncNATs: 55.28%). A/U content of the lincRNAs and lncNATs (particularly the lincRNAs) was also higher than the protein-coding sequences

(Fig. 2D). Among the lncRNAs with assigned chromosome locations (Fig. 1E), most expressed lncRNA loci (164 lincRNAs and 107 lncNATs) were mapped to chromosome A09, and the least was found to be present on chromosome A10 (48 lincRNAs and 65 lncNATs). The majority of the expressed mRNA loci were located on chromosome A03 (4618) and the least on chromosome A04 (2212) (Fig. 1E).

Conservation analysis of *B. rapa* lncRNAs

We investigated putative lncRNA conservation between *B. rapa* and three related *Brassicaceae* species namely *B. napus*, *B. oleracea* and *A. thaliana* by searching for collinear genomic sequences with similarity to annotated *B. rapa* lncRNA loci. The highest number of lncRNAs loci could be matched between *B. rapa* and *B. napus* A sub-genome followed by *B. oleracea* and *A. thaliana* (Fig. 2F, Table S2). The lower number of corresponding non-coding loci compared to protein coding genes, especially at higher evolutionary distance, is consistent with lineage specific nature of lncRNAs. It is important to note that comparisons are based on sequence similarity only, without evidence of expression.

Prediction of *cis*- and *trans*-acting lncRNAs

In the next step, the *cis* and *trans* interactions of the lncRNAs with the expressed protein-coding genes were predicted. The relative location of lncRNA to their neighbouring protein coding gene has been shown to be associated with the effect the lncRNA has on protein-coding gene expression (Rinn and Chang 2012). lncRNAs that act closer to the transcription site of neighbouring genes are identified as *cis*-acting lncRNAs (Figure S3A). In contrast, lncRNAs can regulate numerous genes throughout the genome by acting in a *trans* manner away from the transcription site (Figure S3A). The *cis*-acting lncRNAs are divided into several classes (Figure S3B) based on the direction (sense or antisense), type of interactions (intergenic or genic) and relative location (upstream or downstream) with respect to the interacting protein-coding gene (Kornienko et al. 2013). Figure 3A, B summarises the *cis* lincRNAs and lncNATs present on A01–A10 chromosomes, respectively. In this analysis, the lincRNAs are identified as intergenic, and their distribution between sense and antisense is roughly equal (Fig. 3A, Table S3). A slightly higher number of lincRNAs are located upstream (2540) of the protein-coding genes compared to the lincRNAs located downstream (1,947). lncNATs are identified as antisense and genic, the majority of which are located in exons of protein-coding genes (Fig. 3B, Table S4).

Further, these *cis*-acting lncRNA-protein-coding genes neighbouring pairs were filtered out to select the pairs in which both lncRNA and protein-coding genes were

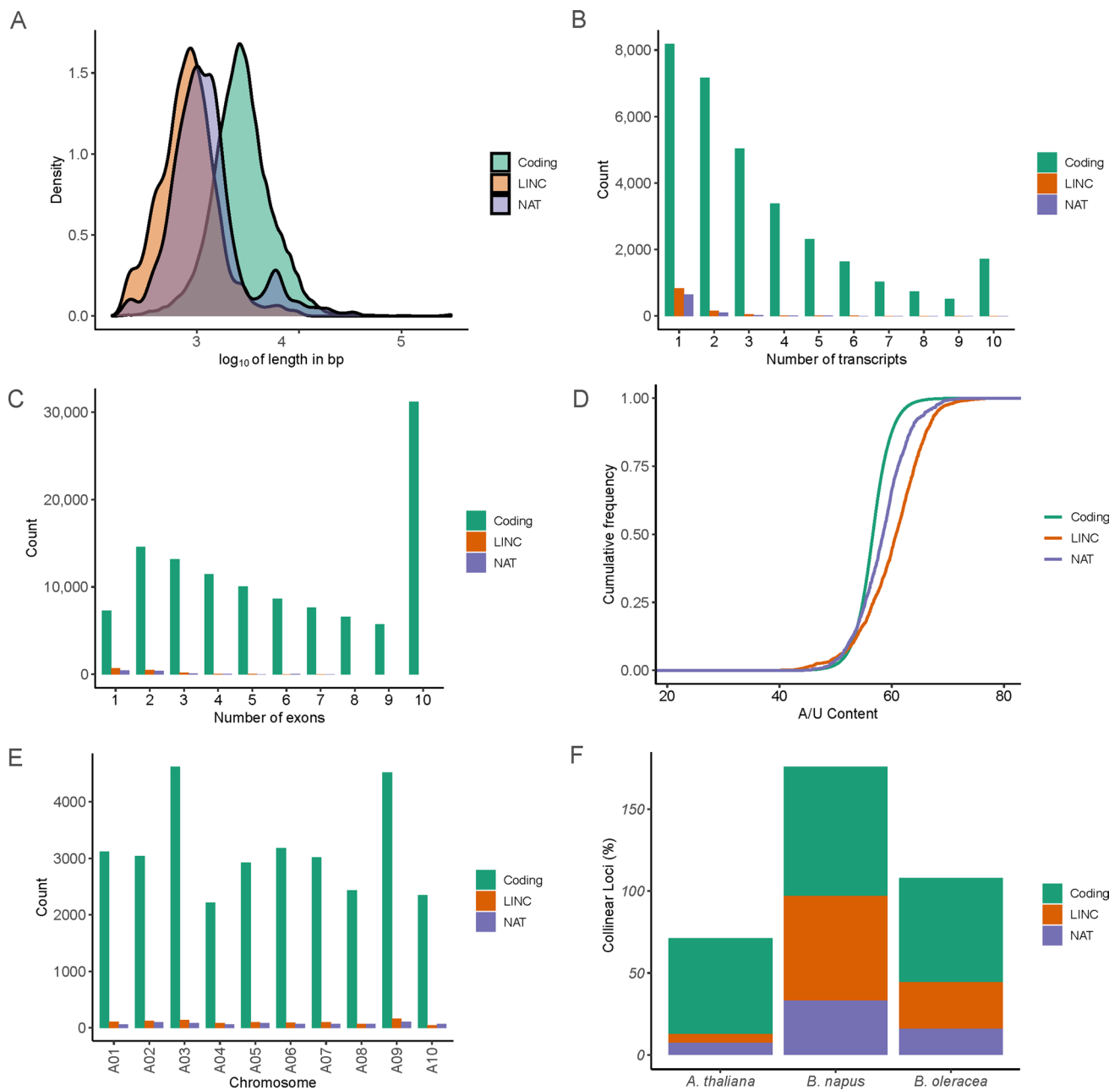


Fig. 2 **A** Distribution of transcript length of coding genes and lncRNAs, **B** number of transcripts per gene of coding genes and lncRNAs, **C** number of exons per transcript of coding genes and lncRNAs, **D** comparison of A/U content of coding transcripts and lncRNAs, **E**

chromosome distribution of coding genes and lncRNAs and **F** proportion of collinear loci of different types between *B. rapa* and three other *Brassicaceae* species

identified as expressed in the samples. The GO enrichment analysis of the protein-coding genes identified as partners of the *cis*-acting lincRNAs and lincNATs is provided in Fig. 3C, D, respectively. lincRNAs neighbouring proteins coding genes were associated with biological process categories such as “hormone-mediated signalling pathway”, “regulation of pollen tube development”, “cell communication”, “regulation of cell morphogenesis involved in differentiation”

and “transcription, DNA-templated”, (Fig. 3C, Table S5). The protein-coding genes neighbouring *cis*-acting lincNATs were involved in “carbohydrate utilisation”, “transmembrane transport”, “replication fork reversal”, “phosphorylation” and “stamen filament development” among other biological processes (Fig. 3D, Table S6).

The prediction of *trans* regulation of protein-coding genes by lincRNAs depends on the formation of complementary

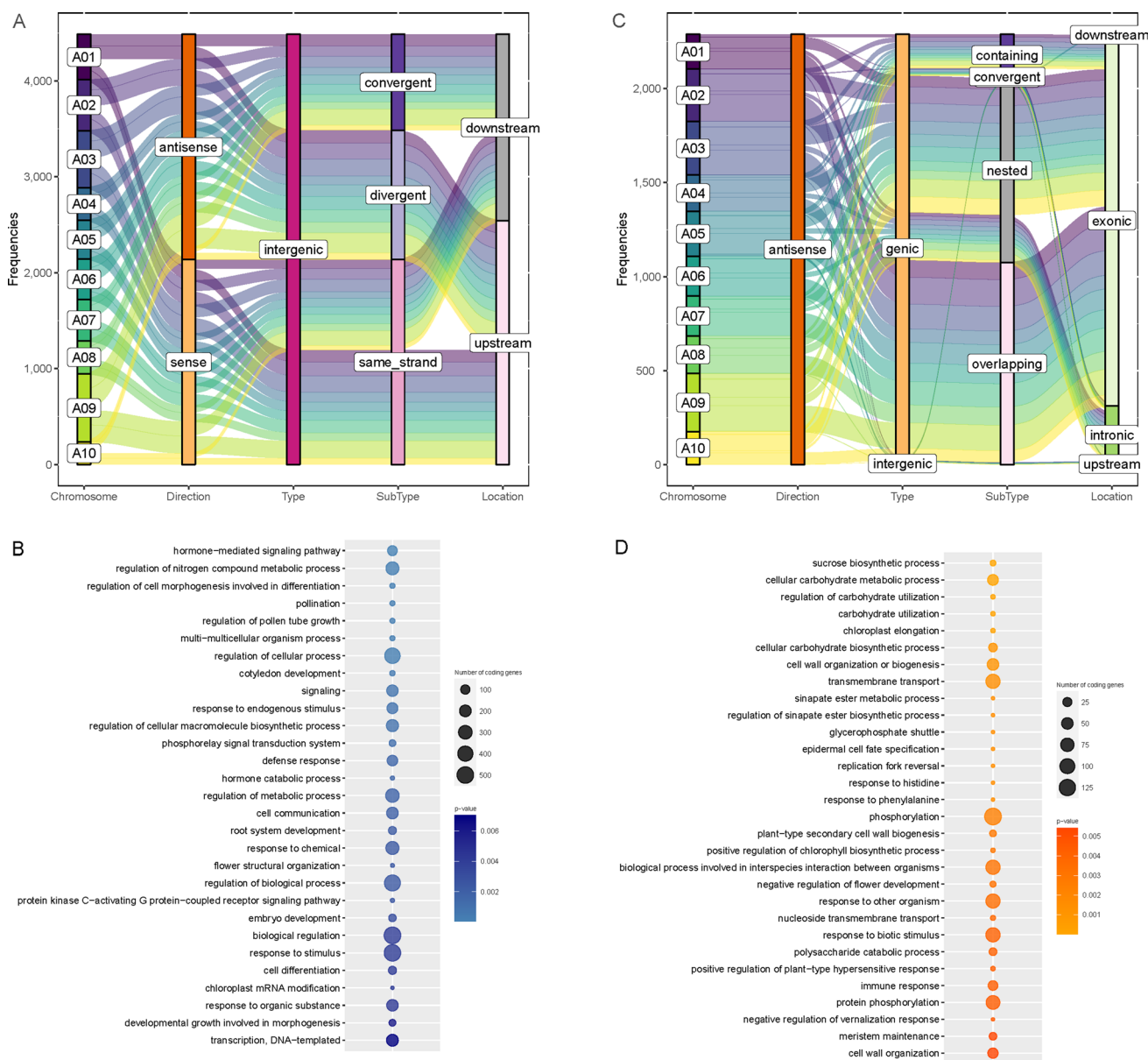


Fig. 3 **A** lincRNA *cis* interactions classification per chromosome in *B. rapa*, **B** top significant non-redundant GO terms associated with expressed protein-coding genes identified as partners of *cis*-acting lincRNAs, **C** lincNAT *cis* interactions classification per chromosome

in *B. rapa* and **D** top significant non-redundant GO terms associated with expressed protein-coding genes identified as partners of *cis*-acting lincNATs

hybrids and the associated interaction energy between the lincRNA and the associated protein-coding genes. Interactions in the scaffold were discarded since the scaffold is unplaced, and one cannot determine the *bona fide* of the *trans* interactions. Initially, the maximum threshold of interaction energy was set at -20 J to retain significant interactions, and 103,545 interactions were identified for lincRNA transcripts. For lincNAT transcripts, 82,606 total *trans* interactions were identified. However, a number of significant *trans* interactions in the order of hundreds of thousands are unlikely. The

distribution of the energy of interactions (Fig. 4A) shows that most of these interactions have low energy (below -100 J). Thus, setting a more stringent arbitrary threshold of -100 J (red vertical line in Fig. 4A) brings down the number of *trans* interactions to 1418 for lincRNAs and 1061 for lincNATs. The 1418 identified *trans* interactions involved 548 lincRNAs (Table S7), out of which $\sim 43\%$ significantly interacted with only 1 protein-coding gene, whereas 9 lincRNAs interacted with ≥ 10 protein-coding genes. *LINC_BRAPST00049411* interacted with the maximum number of protein-coding

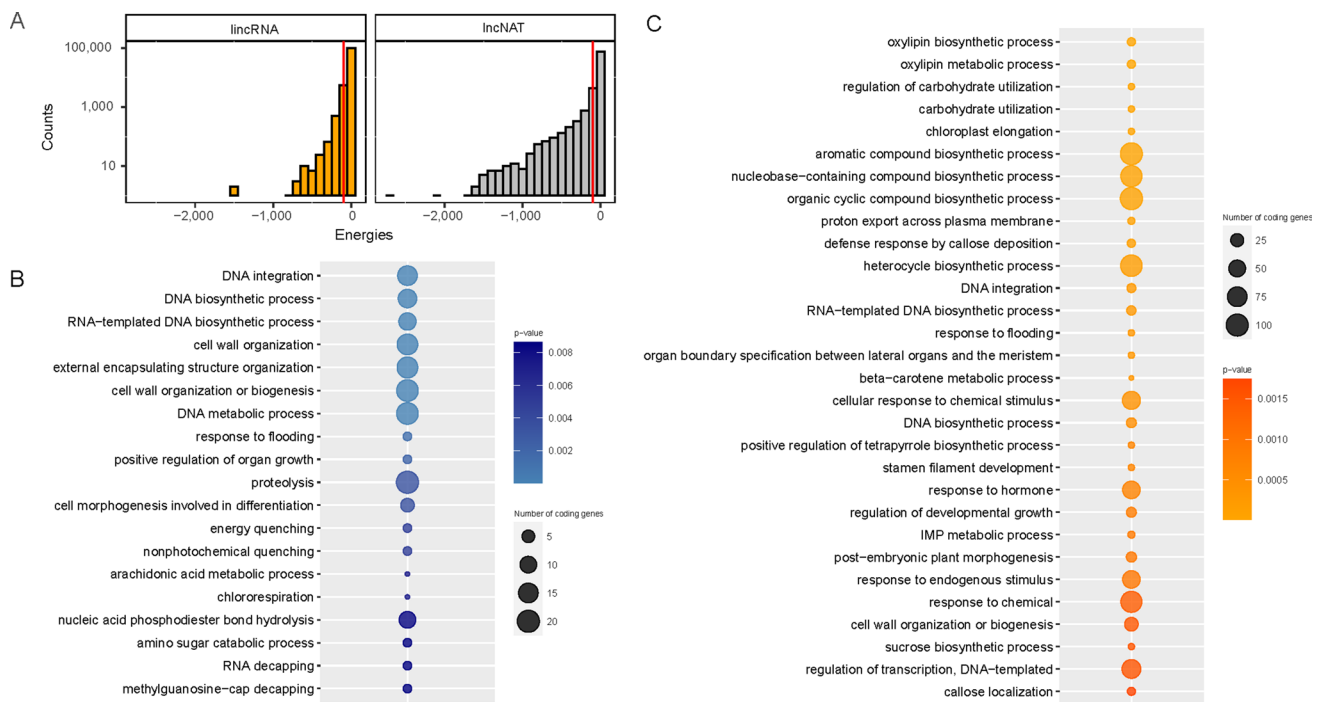


Fig. 4 **A** Comparison of *trans* interactions free energy distribution for lincRNAs (LINC) and lincNATs (NAT), **B** top significant non-redundant GO terms associated with expressed protein-coding genes identi-

fied as partners of *trans*-acting lincRNAs and **C** top significant non-redundant GO terms associated with expressed protein-coding genes identified as partners of *trans*-acting lincNATs

genes (39) in a *trans* manner. In contrast, 1061 *trans* interactions involved 575 lincNATs (Table S8), with only 65% lincNATs interacting with 1 protein-coding gene, and 9 lincNATs interacted with ≥ 10 protein-coding genes. Among the lincNATs, *NAT_BRAPST00007879* interacted with 28 protein-coding genes. Further, these lincRNA-protein coding genes *trans* interacting pairs were filtered out to select the pairs in which both lincRNA and protein-coding genes were identified as expressed in the samples.

Functional enrichment of protein-coding genes identified as potentially regulated by lincRNAs in a *trans* manner revealed their association with “DNA integration”, “cell wall organisation”, “proteolysis”, “cell morphogenesis involved in differentiation” and “regulation of cell growth” among other biological process categories (Fig. 4B). Furthermore, *trans*-acting lincNATs potentially regulated protein-coding genes involved in biological processes such as “oxylipin biosynthetic process”, “carbohydrate utilisation”, “DNA integration”, “stamen filament development”, and “response to hormone” (Fig. 4C).

lincRNA as potential miRNAs targets and precursors

microRNAs (miRNAs) play an important role in regulating gene expression by influencing mRNA degradation and translational repression (Bartel 2004). lincRNAs, like

mRNA, can be miRNA targets and operate as miRNA decoys, suppressing the interaction between miRNAs and their target genes (Franco-Zorrilla et al. 2007). Out of the 1052 lincRNAs, only 22 were predicted as potential targets of 18 miRNAs, and 21 out of 780 lincNATs were predicted to be targeted by 36 miRNAs (Table S9). Majority of the identified *B. rapa* lincRNAs targeted by miRNAs were potentially regulated by cleavage, and very few lincRNA were inhibited at the translational level. The low number of lincRNAs detected as miRNA targets in this analysis is probably due to the lack of male reproductive tissue-specific miRNAs available in published miRNAs.

Some lincRNAs are also considered small RNA (miRNA and siRNA) precursors (Amor et al. 2009; Arikrit et al. 2013; Ma et al. 2014; Wei et al. 2022). Plant small RNAs provide a crucial regulatory role in gene expression and genome integrity by silencing transposons during plant reproduction (Liu et al. 2020; Pokhrel et al. 2021). The comparison of the lincRNA sequences to the miRbase collection found only 0.95% of lincRNA and 0.90% of lincNATs as potential small RNA precursors (have 100% similarity to known mature miRNAs). Furthermore, to identify high confidence targets of the miRNAs for which lincRNAs served as precursors, psRNATarget with a stringent expectation cutoff of 0 was employed (Table S9). Three lincRNA–miRNA–mRNA modules were identified (Fig. 5A).

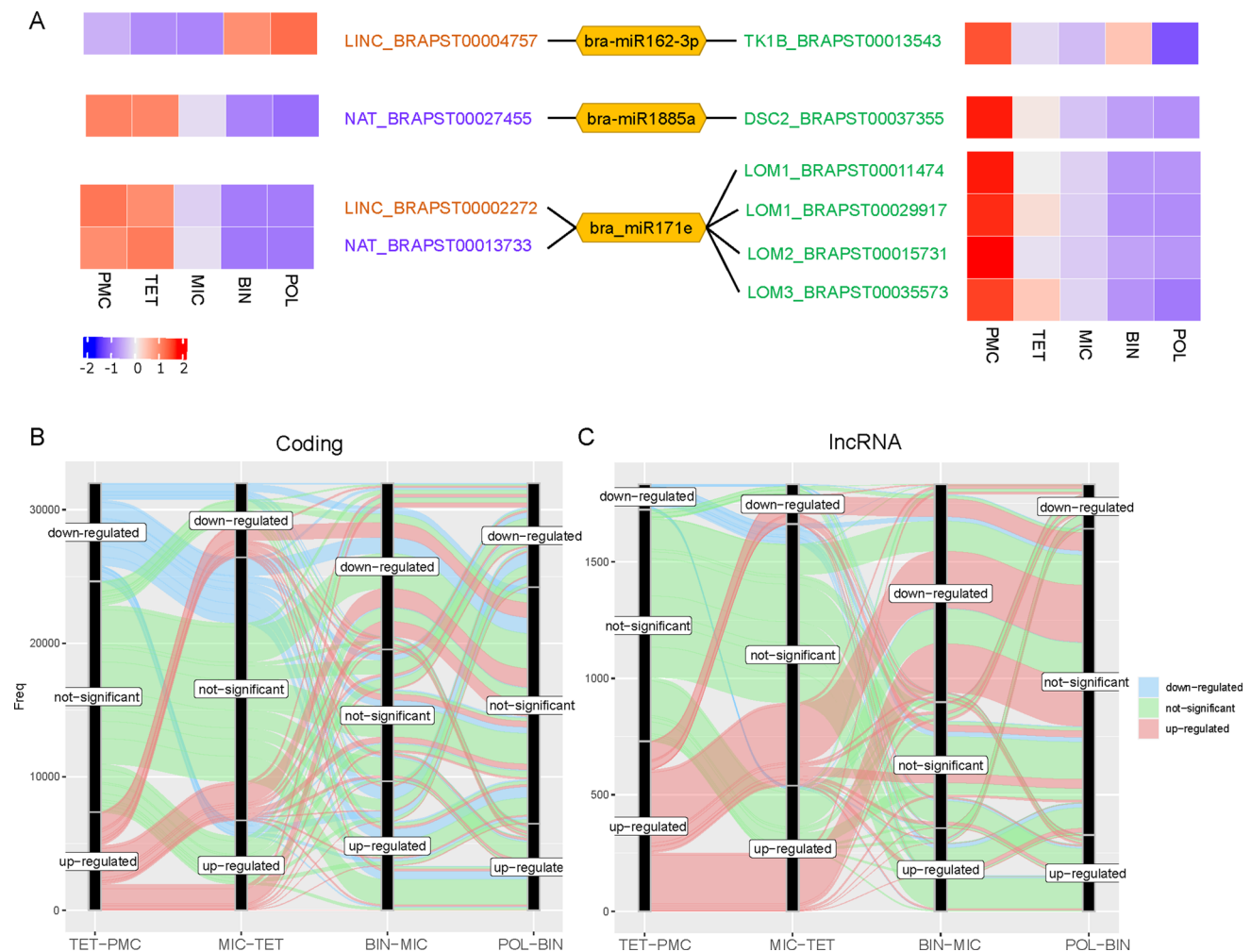


Fig. 5 **A** Three identified lncRNA-miRNA-mRNA modules, where the lncRNA acts as a precursor of the miRNA, and the mRNA or the protein-coding gene is the direct target of miRNA. Heat maps represent the expression profiles of lncRNA and protein-coding genes dur-

ing pollen development, **B** differential regulation of protein-coding genes during pollen development as depicted by an alluvial plot and **C** differential regulation of lncRNAs during pollen development as depicted by an alluvial plot

For one of the modules, the expression profile of the lncRNA was antagonistic to the protein-coding gene expression profile. *LINC_BRAPST00004757* potentially acts as a precursor of *bra-miR162-3p*, which then targets the expression of *BRAPST00013543*. Functional annotation identified *BRAPST00013543* as a gene encoding thymidine kinase that salvages DNA precursors. The pyrimidine salvage pathway is crucial for genome replication and maintaining of its integrity. *BRAPST00013543* showed the highest expression in the PMC stage, and its expression gradually decreased, whereas *LINC_BRAPST00004757* expression increased as pollen development progressed (Fig. 5A). Thus, it can be postulated that *LINC_BRAPST00004757* regulates the expression of *BRAPST00013543* during male gametophyte development in *B. rapa*.

Differential transcriptional reprogramming during pollen development

lncRNAs identified in the datasets used in this study had lower expression levels than protein-coding genes. lncRNAs with low abundance might get filtered out while performing differential expression analysis (Assefa et al. 2018). *limma* R package employed in this study to perform differential expression analysis runs a moderated *t* test after an empirical Bayes correction (Ritchie et al. 2015), a generic and suitable for the differential expression of processed lncRNA expression data. In the RNA-Seq libraries, 49,577 protein-coding genes, 4347 lincRNAs and 2045 lncNATs were identified. A CPM cutoff of 1 in at least 3 samples was used to identify 31,729 coding genes, 1,052 lincRNAs and 780 lncNATs

expressed during pollen development. To investigate the regulation of protein-coding genes and lincRNA during pollen development, we performed differential expression analysis ($\log_2\text{fc}$ cutoff = 0.585, adjusted p value cutoff < 0.01) across four contrasts (TET-PMC, MIC-TET, BIN-MIC, and POL-BIN) by comparing each pollen developmental stage with the previous one (Table S10, Figure S4A). In total, 92.58, 89.73 and 93.21% of the expressed protein-coding genes, lincRNAs and lincNATs, respectively, were differentially regulated across the four contrasts. When the uninucleate microspore transitions into a binucleate microspore, a significantly higher percentage of genes and lincRNAs were differentially regulated with a higher proportion of downregulated genes (Fig. 5B, C, Figure S4A). These observations align with the reported findings that during male germline development, a decreasing trend in transcriptome size and complexity throughout microsporogenesis and microgametogenesis is observed in flowering plants (Singh et al. 2008; Wei et al. 2010). Only 19 and 84 genes were commonly upregulated or downregulated among the protein-coding genes across all developmental stage contrasts (Figure S4B). Interestingly, no common lincRNAs were identified to be differentially regulated across all four contrasts, further highlighting their stage-specific regulation (Figure S4B).

Gene ontology (GO) analysis of the differentially expressed protein-coding genes was performed to unravel their role during different pollen developmental stages (Figure S5). When pollen/microspore mother cell transitions into tetrads, protein-coding genes associated with biological process categories such as “mRNA processing”, “transcription by RNA polymerase II”, “proteolysis”, “gene expression” and “histone modification” were upregulated. Further, as the pollen development progresses, biological process categories including “ribosome biogenesis”, “ncRNA processing”, “translation” and “gene expression” were upregulated. As highlighted earlier, gene expression significantly downregulates as the uninucleate microspore transitions into binucleate pollen, contributing to the decreasing complexity of transcription in binucleate pollen. This was further supported by the downregulation of biological process categories involved in “regulation of transcription, DNA-templated” and “regulation of gene expression”. In contrast, the upregulated genes were associated with “cellular localization” and “protein transport” among others. During the final trinucleate pollen stage, the genes involved in transcription and protein synthesis were generally expressed at lower levels, as indicated by the downregulation of GO terms “ribosome biogenesis”, “translation” and “mRNA processing”. Furthermore, protein-coding genes involved in processes including “localization”, “regulation of pollen tube growth”, “ion transmembrane transport” and “nucleotide-sugar metabolic process” were upregulated in trinucleate pollen. The functional

annotation of differentially expressed genes highlighted the stage-specific differential regulation of an array of biological processes during the progression of male gametophyte development.

lincRNA–mRNA co-expression analysis

To predict the regulatory roles of lincRNA during pollen development, the co-expression networks between the protein-coding genes and lincRNAs (expressed genes, > 1 CPM in at least three samples) were identified using the WGCNA tool. The tool identified 24 modules in the dataset (Figure S6). For further analysis, the top three modules associated with the five pollen developmental stages were identified (Table S11). A different number of lincRNAs were present in the selected modules. Based on the *cis* and *trans* regulation of protein-coding genes by lincRNA and lincNATs, we next investigated the hub genes in the selected modules and identified lincRNA–protein-coding genes interactions and grouped them as *cis*-lincRNA–protein-coding gene, *cis*-lincNAT–protein-coding gene, *trans*-lincRNA–protein-coding gene, and *trans*-lincNAT–protein-coding gene co-expressed pairs. We also performed correlation analysis to supplement the WGCNA analysis and further filtered out lincRNA–coding gene pairs with a Pearson correlation coefficient of less than 0.8 or more than -0.8. Additionally, only those pairs were selected in which the protein-coding gene was differentially expressed and had > 65% similarity with its homolog in *A. thaliana*. In total, 54 *cis*-lincRNA–protein-coding genes, 58 *cis*-lincNAT–protein-coding gene, 8 *trans*-lincRNA–protein-coding genes and 18 *trans*-lincNAT–protein-coding gene interacting co-expressed pairs were identified (Table S12).

We further searched for an association between protein-coding genes and lincRNAs expressed during male gametophyte development based on functional annotation of genes. We collected genes annotated with the GO biological process terms associated with male gametophyte and pollen development. We also collected genes predicted to be transcription factors or homologous to pollen-specific genes in *A. thaliana*. In total, we found 38 *cis*-lincRNA–protein-coding genes, 31 *cis*-lincNAT–protein-coding genes, 7 *trans*-lincRNA–protein-coding genes and 14 *trans*-lincNAT–protein-coding gene pairs of interest (Figs. 6 and 7). Several key pollen developmental regulators were found among the genes identified, including genes involved in “regulation of cell cycle”, “microtubule-based movement”, “pollen development”, “pollen tube growth”, “cell wall organisation” and “transmembrane transport” along with genes showing pollen-specific expression (Figs. 6 and 7). Analysis of the function of protein-coding genes in the identified lincRNA–protein-coding gene pairs revealed genes involved in transcription regulation, such as transcription factors

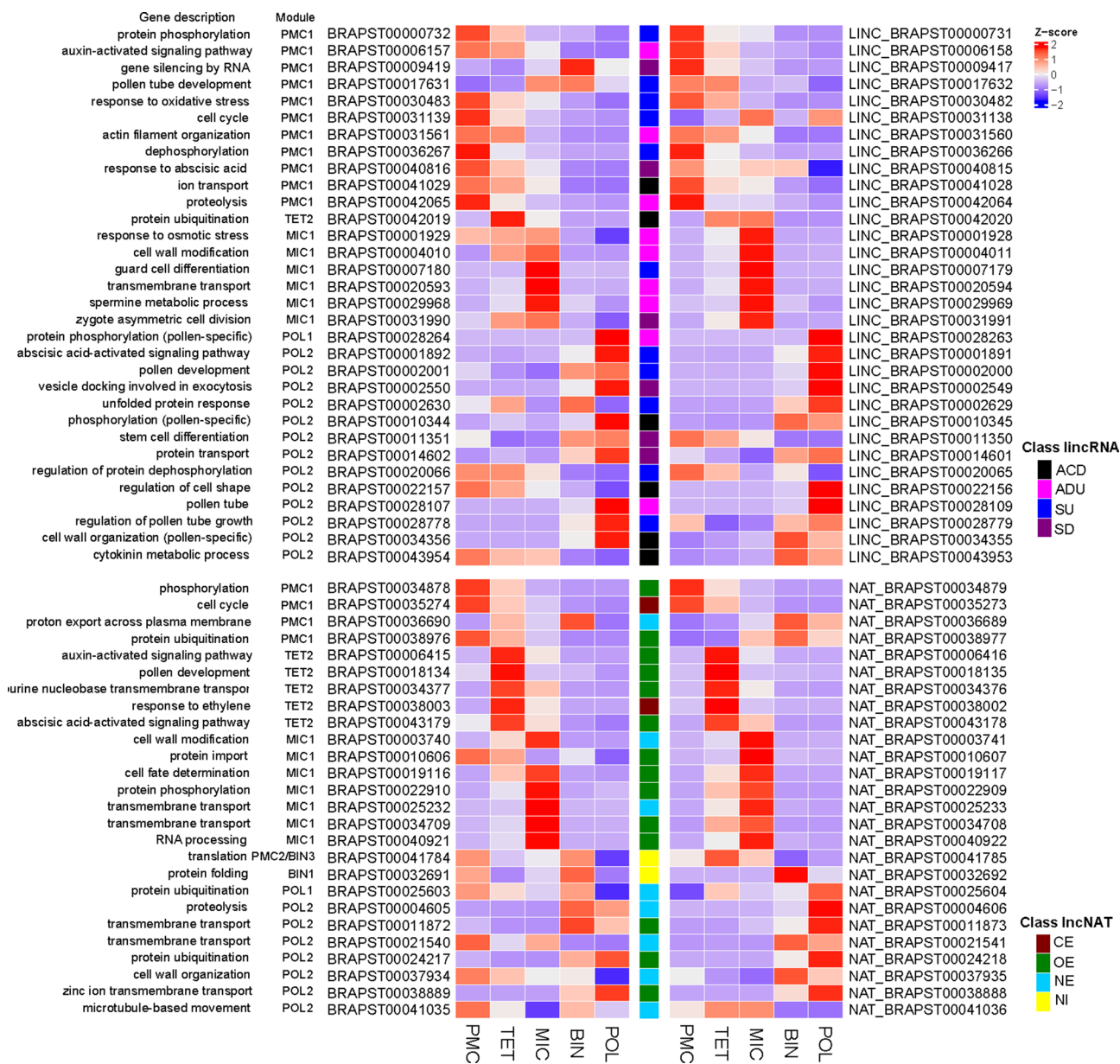


Fig. 6 Heat map representing the *cis*-lincRNA–protein-coding gene and *cis*-lincNAT–protein-coding gene pairs in the top modules of pollen development stages as identified by the co-expression analysis. The GO term description of the protein-coding gene highlights the involvement of the co-expressed pairs in biological processes critical to pollen development. The class of *cis* interaction between the

lincRNAs, lincNATs and their respective co-expressed protein-coding gene is also illustrated. *cis* classification of lincRNAs *ACD* antisense convergent downstream, *ADU* antisense divergent upstream, *SU* same strand upstream, *SD* same strand downstream. *cis* classification of lincNATs *CE* contained exonic, *OE* overlapping exonic, *NE* nested exonic *NI* nested intronic

belonging to WRKY, bHLH and NAC TF families (Fig. 7). The proximity of lincRNAs and our previous results suggesting a possible regulatory relationship between lincRNAs, and their co-expressed protein-coding gene partners suggest that the expression of these developmental regulators and transcription factors could be affected by lincRNAs and that they present targets for future investigation.

Discussion

Long non-coding RNAs (lincRNAs) play diverse roles in regulating biological processes (Golicz et al. 2018a, b; Perry and Ulitsky 2016; Waseem et al. 2020; Yu et al. 2019). Recent research has shown an increasing number of lincRNAs with significant tissue-specific expression

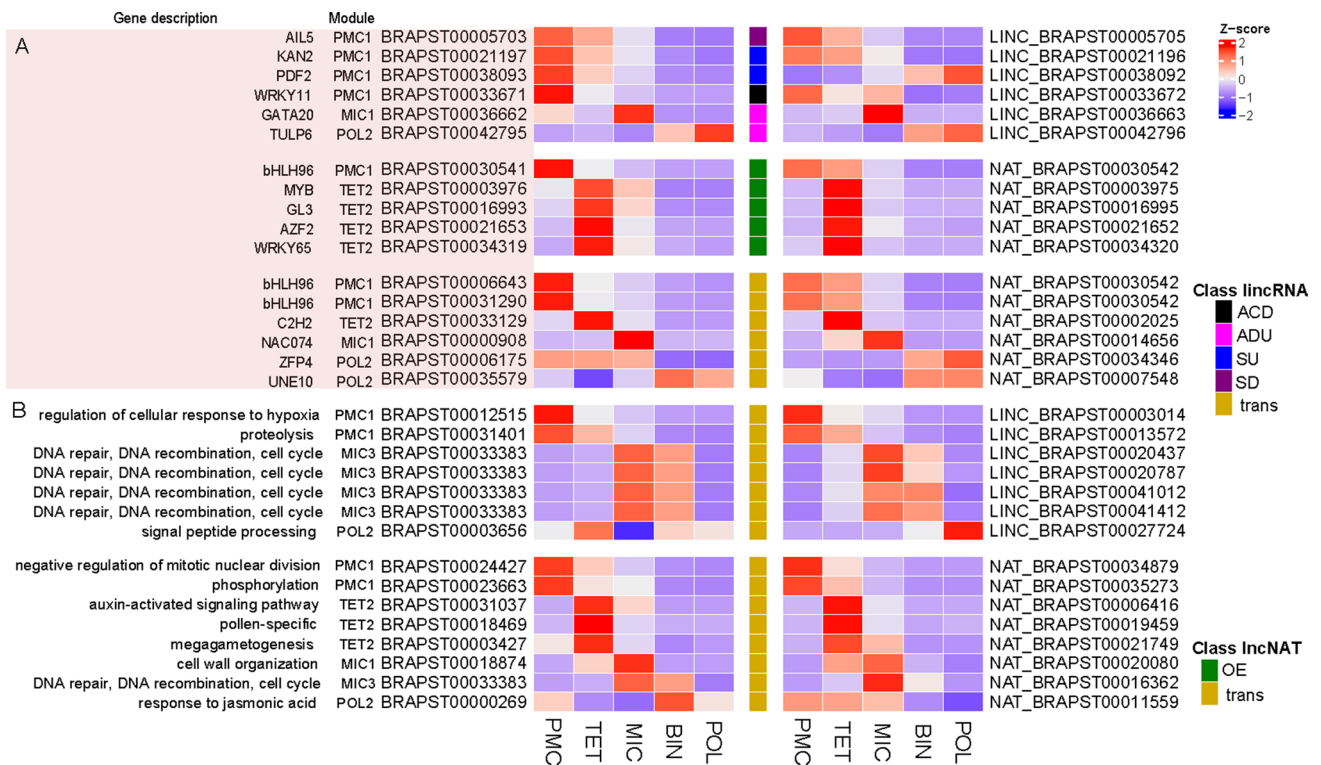


Fig. 7 **A** Heat map representing the *cis*- and *trans*-lincRNA–protein-coding gene, lincNAT–protein-coding gene pairs in the top modules of pollen development stages as identified by the co-expression analysis. The protein-coding genes in these pairs were identified as transcription factors (light red box), **B** heat map representing the *cis*-lincRNA–protein-coding gene and *cis*-lincNAT–protein-coding gene pairs in the top modules of pollen development stages as identified by the co-expression analysis. The GO term description of the protein-

coding gene highlights the involvement of the co-expressed pairs in biological processes critical to pollen development. The class of *cis* interaction between the lincRNAs, lincNATs and their respective co-expressed protein-coding gene is also illustrated. *cis* classification of lincRNAs- *ACD* antisense convergent downstream, *ADU* antisense divergent upstream, *SU* same strand upstream, *SD* same strand downstream. *cis* classification of lincNATs- *CE* contained exonic, *OE* overlapping exonic, *NE* nested exonic, *NI* nested intronic

patterns, implying that they play regulatory roles in developmental processes (Gawronski and Kim 2017; Wang et al. 2015; Ward et al. 2015). Plant reproductive development is an essential feature of crop breeding, and identification of lincRNAs that influence reproductive development is becoming increasingly important. The progression of pollen development from diploid pollen mother cells to haploid microspores to highly specialised haploid trinucleate pollen provides the opportunity for dissecting molecular program controlling male lineage development and identification of transcriptional network in each stage associated with cell identity (Okada et al. 2005, 2007; Russell et al. 2012), cell cycle, cell fate determination (Haerizadeh et al. 2006; Singh and Bhalla 2007; Sharma et al. 2011) as well as male gametophyte specification processes. Here, we identified sets of lincRNAs and protein-coding genes expressed during pollen development (Fig. 1A) in field mustard (*B. rapa*). As an important vegetable crop, *B. rapa* is an attractive option for use as a reference species in *Brassica* genome investigations (Zhang et al. 2022). We predicted regulatory roles of lincRNAs based on genome location and co-expression

analysis. Our results help resolve the distinctive identity of pollen developmental stages and provide a rich data source for further describing the mechanisms underlying male lineage development.

We identified 6,392 lincRNAs (4347 lincRNAs and 2,045 lincNATs) during pollen development in *B. rapa* (Fig. 1E). A study in *B. rapa* conducted a time series of RNA-seq experiments at five developmental stages during pollen development and three different time points after pollination and identified 12,051 putative lincRNAs (Huang et al. 2018). This difference in results can be attributed using different cell types for sequencing. Huang et al. (Huang et al. 2018) used whole buds representing the five pollen developmental stages for sequencing, whereas we isolated pollen mother cells, tetrads, microspores, binuclear pollen and trinucleate pollen for sequencing. Another reason can be the use of a different variety of *B. rapa* and the employment of different bioinformatic tools for identification and discovering of lincRNAs.

The lincRNAs (lincNATs and lincRNAs) identified in *B. rapa* during pollen development (Fig. 2A–C) were shorter

and had fewer isoforms and exons per transcript than protein-coding genes; these properties were consistent with previous reports of genome-wide lncRNA discovery (Golicz et al. 2018a, b; Liu et al. 2012; Wang et al. 2015; Zhang et al. 2014). About half of the identified lncRNAs reported in Arabidopsis and rice have one transcript and include only a single exon (Liu et al. 2012; Zhang et al. 2014). A higher A/U content is another feature of lncRNAs observed in *B. rapa* lncRNAs (Fig. 1D). The A/U composition may reflect the underlying sequence. However, it is attractive to speculate that it may be a feature of lncRNAs that facilitates recognition by RNA-binding proteins. The flexible nature of lncRNAs in interacting with other transcripts is potentially indicated by a high A/U content (Smith and Mattick, 2017), as transcripts rich in A/U content are less stable (Barreau et al. 2005).

lncRNAs are reported to be low in abundance and show tissue-specific expression in plants and animals (Golicz et al. 2018a, b). In this study also, *B. rapa* lncRNAs showed lower expression in comparison to protein-coding genes (Fig. 1B). The expression patterns of lncRNAs (Fig. 1C–F) likely reflect their stage-specific roles during pollen development and are consistent with previous observations of high lncRNA transcription in buds containing the microspore stage (Huang et al. 2018). lncRNAs showed peak expression at the microspore stage, whereas a higher fraction of protein-coding genes had peak expression at pollen/microspore mother cells (Fig. 1E). Few lncRNAs were also expressed exclusively at a single pollen developmental stage (Fig. 1F). Furthermore, differential expression analysis revealed no common *B. rapa* lncRNA to be differentially regulated between all pollen developmental stage contrasts. Different expression windows for lncRNAs suggest they are part of a coordinated expression program rather than a result of non-specific pervasive transcription. These results indicate that lncRNA expression profiles may precisely determine the male lineage specifications. Further investigation of these stage-specific lncRNAs may lead to a better understanding of the molecular control of pollen development.

Across plant species, lncRNAs are reported to show low conservation and are mostly species specific (Ke et al. 2019; Simopoulos et al. 2019). Because lncRNAs are highly evolved, there is less sequence conservation across plant and animal taxa, resulting in fewer phylogenetic connections (Simopoulos et al. 2019). In the present study, we found that the proportion of collinear non-coding loci between *B. rapa* and three other Brassicaceae species non-coding decreased with increasing evolutionary distance between plant species consistent with the lineage specific nature of lncRNAs. In our study, a relatively high percentage (64%) of *B. rapa* lncRNA loci showed similarity to collinear genomic loci in the *B. napus* A sub-genome. Interestingly *B. napus* is amphidiploid species

originating from the hybridisation of *B. rapa* and *B. oleracea* and contains complete diploid chromosome sets of both parental genomes. Since the A genome in *B. napus* represents the diploid *B. rapa* genome, the high level of conservation of *B. rapa* lncRNAs in the A genome of amphidiploid *B. napus* species is expected. According to Liu et al. (2012), 2% of all putative lncRNAs identified in *A. thaliana* are conserved across the plant kingdom. Comparing maize (*Zea mays*) lncRNAs to *A. thaliana* lncRNAs yielded a similar conservation level (L. Li et al. 2014a, b). With the increasing availability of whole plant genomes, it will be feasible to address concerns of non-coding sequence conservation in a phylogenomic setting using genetic collinearity in addition to sequence similarity. The expanding lncRNA database and comparative genomics may further the understanding of the functional conservation of lncRNAs, and the underlying mechanism across different tissues and plant species.

Recent investigations have revealed the various functions mediated by the action of lncRNAs in plants (Nejat and Mantri 2018). Different interactions between lncRNAs and protein-coding genes point towards the diverse mode of action of lncRNAs (Rinn and Chang 2012). lncRNAs, like mRNA, can be miRNA targets and operate as miRNA decoys, suppressing the interaction between miRNAs and their target genes (Franco-Zorrilla et al. 2007). Out of the 1052 lincRNAs, only 22 were predicted as potential targets of 18 miRNAs, and 21 out of 780 lncNATs were predicted to be targeted by 36 miRNAs (Table S9). Some lncRNAs are also considered small RNA (miRNA and siRNA) precursors (Amor et al. 2009; Arikiti et al. 2013; Ma et al. 2014; Wei et al. 2022). We found that only 10 lincRNA and 8 lncNATs are potential small RNA precursors (100% similarity to known mature miRNAs). The small fraction of lncRNAs identified as targets of miRNA and the lack of similarity between lncRNAs and mature miRNAs suggest that the majority of lncRNAs are unlikely to be miRNA decoys or act as small RNA precursors and have other, independent modes of regulation. However, it is notable that the publicly available databases may lack some miRNAs specific to pollen development; therefore, the roles of lncRNAs as miRNA decoys or miRNA precursors of yet unknown miRNAs cannot be excluded.

lncRNAs can regulate gene expression in a *cis* or *trans* manner (Fatica and Bozzoni 2014). Different classes of lncRNAs may play distinct roles in regulating protein-coding gene expression changes in relative abundance. lncRNAs act closer to the transcription site of the neighbouring genes while acting in a *cis* manner (Guil and Esteller 2012). Contrary to this, they can regulate multiple genes throughout the genome by functioning away from the transcription site while acting in a *trans* manner (Fatica and Bozzoni 2014). The present study predicted *cis* and *trans* interactions between all the annotated lncRNAs and protein-coding genes (Figs. 3 and 4, Table S3–S8). The *cis* interactions

detected for lincRNAs and lncNATs highlighted that *cis*-acting lincRNAs are completely identified as intergenic. The distribution between sense/antisense and upstream/downstream description is approximately similar. *Cis*-acting lncNATs are mostly identified as antisense and genic, and most of them are located in exons of protein-coding genes. A higher number of significant *trans* interactions for lncNATs as compared to lincRNA were identified in this study. It is highly likely that the lncNATs overlapping with coding genes in the *B. rapa* genome potentially emerged from an ancestral genome triplication event (Cai et al. 2017; Cheng et al. 2014; Zhang et al. 2022). A substantial proportion of *B. rapa* genes would be present in groups of paralogs having sequence similarity scattered in different positions on the genome (Mun et al. 2009). Therefore, the chance of finding a high similarity between a lncNAT and a distant protein-coding gene is higher than for lincRNA, which is instead located outside protein-coding genes.

In plants, few studies have reported lncRNAs linked to male reproductive development, for instance, *LDMAR* in rice, *Zm401* in maize and *BcMF11* in *B. campestris* (Ding et al. 2012; Ma et al. 2008; Song et al. 2013, 2007). Huang et al. (2018) also reported lncRNA–mRNA pairs, including 10 genes involved in pollen and pollen development functions. Weighted gene co-expression network and correlation analysis revealed 90 pairs of *cis*- or *trans*-acting lncRNAs and protein-coding genes, including several genes involved in transcriptional regulation, regulation of cell cycle, microtubule-based movement, pollen development, pollen tube growth, cell wall organisation and transmembrane transport (Fig. 6, 7 and Table S12). Interestingly, a higher number of *cis*-acting lncRNAs containing pairs were identified, and most of them were positively correlated (Table S12). The next obvious step is determining whether this link between *cis*-acting lncRNAs and protein-coding genes is causal. The intimate relationship between lncRNA/protein-coding gene pairs could indicate the presence of sophisticated gene expression regulatory mechanisms during pollen development. Several studies have reported *cis*-acting lncRNAs regulating the activation of neighbouring genes' expression (Csorba et al. 2014; Krishnan and Mishra 2014; Rosa et al. 2016; Statello et al. 2021; Vance and Ponting 2014; Yap et al. 2010). In our study, the lncRNAs that are strongly co-expressed with neighbouring protein-coding genes are, thus, promising candidates for further research. Even though *cis*-acting lncRNA and neighbouring protein-coding genes are typically positively correlated, we also found few lncRNA that potentially negatively regulated the expression of neighbouring genes. Several *trans*-acting linked lncRNAs were also discovered during pollen development. These results indicate that an intimate relationship between lncRNAs and protein-coding transcripts may be mediated by various molecular interactions.

Material and methods

Plant material, sample collection and sequencing

Brassica rapa accession no. ATC 92,270 Y.S (AND)-168 was used in this study. The plants were grown in a growth cabinet under the following conditions (21/18 °C day/night, a photoperiod of 16/8 h light/dark, 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity and 60% humidity). The different stages of pollen development were harvested as per previously published protocol (Babaei et al. 2021; Golicz et al. 2021). Briefly, the following five groups of buds were identified: <0.5 mm buds (pollen/microspore mother cells, PMC), 0.8–1 mm buds (tetrads, TET), 1–2.5 mm (microspores to polarised microspores, MIC), 3–4.5 mm (early to late binucleate pollen, BIN) and 5–6 mm (trinucleate pollen, POL). Anthers were carefully dissected from the buds of the last two groups and crushed in the B5 medium in a 1.5 mL tube. The whole buds were crushed in B5 medium for the first three groups. The crushed suspension was then filtered through a 44 μm nylon mesh into 15 mL tubes. The filtrate was centrifuged at 150 g for 3 min at 4 °C. The supernatant was discarded, and the pellet was washed using 0.5 \times B5 medium and was again centrifuged at 150 g for 3 min at 4 °C. The supernatant was removed, and the pellet was immediately frozen in liquid nitrogen and stored at –80 °C. An aliquot from each isolation was analysed to check the developmental stage.

The buds collected from different plants were used as one biological replicate, and three independent biological replicates were prepared for each sample. The total RNA was isolated using the mirVana™ miRNA Isolation Kit (Thermo-Fisher; Part Numbers AM1560, AM1561, Carlsbad, CA, USA) according to the manufacturer's instructions. The isolated RNA samples were treated with TURBO™ DNase (Ambion, Carlsbad, CA, USA) to remove DNA contamination. The libraries were prepared using Illumina TruSeq stranded mRNA kit with poly (A) selection (Golicz et al. 2021). Additionally, five libraries were prepared using rRNA depletion (Babaei et al. 2021). The sequencing was performed at the Australian Genome Research Facility (AGRF), Melbourne.

B. rapa genome re-annotation and lncRNA discovery

The reads were mapped to the *B. rapa* 'Chiifu' genome assembly v3.5 (Zhang et al. 2022) using STARv2.7.9a (Dobin and Gingeras 2015) two-pass mapping as described by (Veeneman et al. 2016). Default settings were used for generating the first-pass genome index. The filters and parameters for the two-pass mapping are provided in Table S1B. The transcripts for each library were assembled separately using Stringtie v2.1.4

(Perteza et al. 2015; Varabyou et al. 2021) using existing the *B. rapa* ‘Chiifu’ genome assembly v3.5 genome annotation as a guide. The individual assemblies were merged using Stringtie v2.1.4, again using the *B. rapa* ‘Chiifu’ genome assembly v3.5 genome annotation as a guide.

The coding potential of all genes was evaluated by (1) Coding potential calculator2, CPC2-beta, (Kang et al. 2017) (2) PLEK tool (A. Li et al. 2014a, b) (3) DIAMOND blastx v0.9.30 comparison against the RefSeq database (obtained on: 23.05.2021) (Buchfink et al. 2015) and (4) DIAMOND blastp comparison of the extracted longest open reading frames (ORFs; TransDecoder v5.5.0) against RefSeq database. The genes for which none of the transcripts was identified as coding were designated as non-coding. Non-coding loci with at least one transcript ≥ 200 bp in length and no matched in the Rfam v14.7 database were identified as lncRNAs. The overlap between lncRNAs and protein-coding loci was identified by bedtools v2.30.0 (Quinlan 2014). lncRNAs that did not overlap any coding loci were designated lincRNAs, while lncRNA that overlapped at least one protein-coding locus was designated as lncNATs. A schematic representation of the pipeline for *B. rapa* genome re-annotation, lncRNA discovery and alternative splicing analysis is provided in Figure S1.

cis and trans regulation of protein-coding genes by lncRNAs

Classification of lncRNAs acting in *cis* is performed with the FEELnc_classification script from the software FEELnc (Wucher et al. 2017). This software applies a 100 kb sliding window and classifies lncRNAs based on its relationship with the closest mRNA in the window (Figure S3B). A custom python script was used for generating statistics per chromosome of the annotated lncRNAs. The results are visualised in a Sankey plot with a custom R script using the packages ggplot2 and ggalluvial (Wickham 2016). For simplicity, only the best-predicted interactions for lncRNAs localised in chromosomes are visualised, and the lncRNAs located on scaffold chromosomes were removed.

Possible RNA–RNA interactions in *trans* between *B. rapa* lncRNAs and mRNAs were performed using the RIBLAST software (Fukunaga and Hamada 2017). RIBLAST was chosen for this analysis as it is over 64 times faster than other similar software for achieving a similar level of precision.

Gene expression quantification, data pre-processing and differential expression analysis

Transcript expression was quantified using Kallisto v0.45.1 (Bray et al. 2016), and the downstream analysis was performed using the limma R package. Tximport R package version 1.10.0 (lengthScaledTPM method) was used to generate read counts and transcript per million reads (TPMs)

(Soneson et al. 2015). Transcripts and genes with low expression were filtered based on data mean–variance trend analysis. A gene was considered expressed if any transcripts had CPM ≥ 1 in at least 3 of the 15 samples. Normalisation of the gene and transcript read counts to \log_2 -CPM was performed by the TMM method (Bullard et al. 2010). Principal component analysis (PCA) was undertaken to determine the relatedness between the biological replicates (Figure S2A). Batch effects were corrected using the RUVSeq R package (Risso 2015). Four contrast groups were defined to perform the differential expression analysis: TET vs PMC, MIC vs TET, BIN vs MIC and POL vs BIN. For a gene to be considered significantly differentially expressed in a contrast group, a cutoff adjusted *p* value < 0.01 and \log_2 fold change ≥ 0.5 was used. *p* values of multiple testing were adjusted to correct the false discovery rate (FDR) using the BH method (Benjamini and Yekutieli 2001).

Functional annotation and enrichment analysis

Homologous *B. rapa* coding genes, compared to the Arabidopsis proteome, were identified using the BlastP program with an *e* value $\leq 1e-05$. PANNZER2 (Törönen et al. 2018) was employed for the functional annotation of the protein-coding genes. PANNZER provides a GO term and a functional description for the query protein sequences. 24,179 out of 31,729 expressed protein coding genes were assigned a GO term. The GO enrichment analysis was performed using the GOstats R package, and a hypergeometric test (*p* value < 0.01) was performed to identify overrepresented GO terms (Falcon and Gentleman 2007). The annotated expressed genes (24,179) were used as background for GO enrichment analysis. Further, ReViGO was used to retain the significant non-redundant GO terms (Supek et al. 2011).

lncRNA conservation analysis

Conservation of lncRNA loci was investigated by searching for sequence similarity between *B. rapa* lncRNA loci and the genome assemblies of *B. napus* (A genome) (<https://academic.oup.com/gigascience/article/9/12/giaa137/6034787>), *B. oleracea* (https://phytozome-next.jgi.doe.gov/info/Boleraceacapitata_v1_0) and *A. thaliana* (https://phytozome-next.jgi.doe.gov/info/Athaliana_TAIR10) using LiftOff (Shumate and Salzberg 2021). Collinearity between loci with sequence similarity was confirmed with MCSanX (Wang et al., 2012).

Analysis of lncRNA as miRNA targets or precursors

The identification of *B. rapa* miRNA targets were done using psRNATarget v2 (2017 update) (Dai et al., 2018). Using the default parameters, the alignment of identified *B. rapa*

lncRNAs and protein-coding genes expressed during pollen development was performed against the *B. rapa* miRNAs available on the psRNATarget database. A strict expectation threshold of ≤ 3 was used to filter potential targets.

Mature *B. rapa* sequences were downloaded from miRbase release 22.1 (Kozomara et al. 2019). To identify whether lncRNAs acted as potential precursors or targets of small RNAs (miRNA and siRNA), the lncRNAs were compared to the miRNA sequences using BLASTN v2.7 (-task blastn -outfmt '6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send eval evalue bitscore qcovs qlen slen'). The matches were filtered to retain only matches with 100% sequence identity and match length equal to miRNA. The matches were filtered to retain only matches with 100% sequence identity and match length equal to miRNA length.

lncRNA–mRNA co-expression analysis by WGCNA

Weighted gene co-expression network analysis using the WGCNA R package v1.63 (Langfelder and Horvath 2008) was performed to identify coding genes and lncRNA co-expressed networks. As per the WGCNA guidelines, normalised gene counts were used. The construction of a co-expression network with an approximate scale-free topology of above 0.80 was achieved with a soft power of 10. A minimum module size of 30 is chosen. Moreover, gene modules that exhibit highly similar expression patterns are merged. After evaluating the module dendrogram, an arbitrary threshold of 0.20 (corresponding to 0.80 correlation) was set as limit to merge modules with a similar expression. The top three best correlating modules for each developmental stage were selected, and then the transcripts with the best significance in the selected correlated modules were identified, and functional enrichment analysis was performed. lncRNA–mRNA co-expressed pairs were identified among the hub genes of the selected modules and the potential regulatory relationship (cis or trans) was characterised between the pair. Further, the Pearson correlation coefficient between the lncRNA–protein-coding gene co-expressed pairs was calculated using the R package.

Conclusion

In summary, we investigated the expression of long non-coding RNAs during pollen development and identified 1,832 lncRNA (1052 lincRNA, 780 lincNAT) and 31,729 protein-coding genes as expressed during pollen development in *B. rapa*. The lncRNAs have defined stage-specific expression patterns. The lncRNAs were subdivided into classes based on their genomic location and orientation relative to protein-coding gene neighbours. lncRNA belonging to

different classes have distinct properties suggesting possible differences in function and/or mode of action. The analysis of expression patterns of lncRNA–protein-coding genes points to the involvement of lncRNAs in the modulation of the protein-coding gene associated with several biological processes regulating pollen development. Overall, genome-wide identification, characterisation and functional analysis enabled the identification of lncRNAs candidates and their functional associations with protein-coding genes, potentially revealing regulatory and molecular mechanisms underlying male reproductive development in *B. rapa*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00299-022-02960-0>.

Acknowledgements This research was supported by Spartan HPC (Lev Lafayette GS, Linh Vu, Bernard Meade October 27, 2016 (Meade et al. 2017)) at the University of Melbourne, Australia.

Author contributions Samples collection and RNA isolation, ADA; data curation and analysis, N.L. and AAG; initial draft preparation, N.L. and AAG; review and editing, MBS and PLB; supervision, MBS and PLB. All authors have read and agreed to the published version of the manuscript.

Funding The research was supported by ARC Discovery grant DP0988972 and McKenzie Fellowship.

Data availability All data generated in this study are available in the article and its Supplementary Materials. The Poly-A strand-specific RNA-Sequencing data were deposited in NCBI SRA under PRJNA529957 and rRNA depletion RNA-Seq data were deposited in NCBI SRA under PRJNA763698.

Declarations

Conflict of interest The authors declare no conflict of interest.

Informed consent Not applicable.

Institutional review board statement Not applicable.

References

- Amor BB, Wirth S, Merchan F, d' Aubenton Laporte P, Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM (2009) Novel long non-protein coding RNAs involved in arabidopsis differentiation and stress responses. *Genome Res* 19(1):57–69. <https://doi.org/10.1101/gr.080275.108>
- Arikat S, Zhai J, Meyers BC (2013) Biogenesis and function of rice small RNAs from non-coding RNA precursors. *Curr Opin Plant Biol* 16(2):170–179. <https://doi.org/10.1016/j.pbi.2013.01.006>
- Assefa AT, De Paepe K, Everaert C, Mestdagh P, Thas O, Vandensompele J (2018) Differential gene expression analysis tools exhibit substandard performance for long non-coding RNA-sequencing data. *Genome Biol* 19(1):1–16. <https://doi.org/10.1186/s13059-018-1466-5>
- Babaei S, Singh MB, Bhalla PL (2021) Circular RNAs repertoire and expression profile during Brassica rapa pollen development. *Int J Mol Sci* 22(19):10297. <https://doi.org/10.3390/ijms221910297>

- Babaei S, Singh MB, Bhalla PL (2022) Role of non-coding RNAs in rice reproductive development. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2022.1040366>
- Barreau C, Paillard L, Osborne HB (2005) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 33(22):7138–7150. <https://doi.org/10.1093/nar/gki1012>
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281–297. [https://doi.org/10.1016/s0092-8674\(04\)00045-5](https://doi.org/10.1016/s0092-8674(04)00045-5)
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. <https://doi.org/10.1214/aos/1013699998>
- Böhmendorfer G, Wierzbicki AT (2015) Control of chromatin structure by long noncoding RNA. *Trends Cell Biol* 25(10):623–632. <https://doi.org/10.1016/j.tcb.2015.07.002>
- Borg M, Brownfield L, Twell D (2009) Male gametophyte development: a molecular perspective. *J Exp Bot* 60(5):1465–1478. <https://doi.org/10.1093/jxb/ern355>
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34(5):525–527. <https://doi.org/10.1038/nbt0816-888d>
- Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D (2009) A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet* 5(3):e1000430. <https://doi.org/10.1371/journal.pgen.1000430>
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59–60. <https://doi.org/10.1038/nmeth.3176>
- Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform* 11(1):1–13. <https://doi.org/10.1186/1471-2105-11-94>
- Cai C, Wang X, Liu B, Wu J, Liang J, Cui Y, Cheng F, Wang X (2017) Brassica Rapa genome 2.0: a reference upgrade through sequence re-assembly and gene re-annotation. *Mol Plant* 10(4):649–651. <https://doi.org/10.1016/j.molp.2016.11.008>
- Cheng F, Wu J, Wang X (2014) Genome triplication drove the diversification of Brassica plants. *Horticul Res*. <https://doi.org/10.1038/hortres.2014.24>
- Csorba T, Questa JI, Sun Q, Dean C (2014) Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci* 111(45):16160–16165. <https://doi.org/10.1073/pnas.1419030111>
- Dai X, Zhuang Z, Zhao PX (2018) PsRNATarget: a plant small RNA target analysis server (2017 release) *Nucleic Acids Res* 46(W1):W49–W54. <https://doi.org/10.1093/nar/gky316>
- Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhan T, Qi Y, Gerstein MB, Guo Y (2014) Characterization of stress-responsive linc RNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J* 80(5):848–861. <https://doi.org/10.1111/tbj.12679>
- Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci* 109(7):2654–2659. <https://doi.org/10.1073/pnas.1121374109>
- Dobin A, Gingeras TR (2015) Mapping RNA-seq reads with STAR. *Curr Protoc Bioinform* 51(1):11.14.11–11.14.19. <https://doi.org/10.1002/0471250953.bi1114s1>
- Falcon S, Gentleman R (2007) Using GOSTats to test gene lists for GO term association. *Bioinformatics* 23(2):257–258. <https://doi.org/10.1093/bioinformatics/btl567>
- Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15(1):7–21. <https://doi.org/10.1038/nrg3606>
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 39(8):1033–1037. <https://doi.org/10.1038/ng2079>
- Fukunaga T, Hamada M (2017) RIBlast: an ultrafast RNA–RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* 33(17):2666–2674. <https://doi.org/10.1093/bioinformatics/btx287>
- Gawronski KA, Kim J (2017) Single cell transcriptomics of non-coding RNAs and their cell-specificity. *Wiley Interdiscipl Rev* 8(6):e1433. <https://doi.org/10.1002/wrna.1433>
- Golicz AA (2022) Long intergenic noncoding RNA (lincRNA) discovery from non-strand-specific RNA-seq data. *Plant bioinformatics*. Humana Press, New York, pp 465–482. https://doi.org/10.1007/978-1-0716-2067-0_24
- Golicz AA, Bhalla PL, Singh MB (2018a) lincRNAs in plant and animal sexual reproduction. *Trends Plant Sci* 23(3):195–205. <https://doi.org/10.1016/j.tplants.2017.12.009>
- Golicz AA, Singh MB, Bhalla PL (2018b) The long intergenic non-coding RNA (LincRNA) landscape of the soybean genome. *Plant Physiol* 176(3):2133–2147. <https://doi.org/10.1104/pp.17.01657>
- Golicz AA, Allu AD, Li W, Lohani N, Singh MB, Bhalla PL (2021) A dynamic intron retention program regulates the expression of several hundred genes during pollen meiosis. *Plant Reprod* 34(3):225–242. <https://doi.org/10.1007/s00497-021-00411-6>
- Guil S, Esteller M (2012) Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol* 19(11):1068. <https://doi.org/10.1038/nsmb.2428>
- Haerizadeh F, Singh MB, Bhalla PL (2006) Transcriptional repression distinguishes somatic from germ cell lineages in a plant. *Science* 313(5786):496–499
- Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331(6013):76–79. <https://doi.org/10.1126/science.1197349>
- Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J (2018) Systematic identification of long non-coding RNA s during pollen development and fertilization in Brassica Rapa. *Plant J* 96(1):203–222. <https://doi.org/10.1111/tbj.14016>
- Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 45(W1):W12–W16. <https://doi.org/10.1093/nar/gkx428>
- Ke L, Zhou Z, Xu XW, Wang X, Liu Y, Xu Y, Huang Y, Wang S, Deng X, Chen LL (2019) Evolutionary dynamics of linc RNA transcription in nine citrus species. *Plant J* 98(5):912–927. <https://doi.org/10.1111/tbj.14279>
- Kornienko AE, Guenzl PM, Barlow DP, Pauler FM (2013) Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* 11(1):1–14. <https://doi.org/10.1186/1741-7007-11-59>
- Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47(D1):D155–D162. <https://doi.org/10.1093/nar/gky1141>
- Krishnan J, Mishra RK (2014) Emerging trends of long non-coding RNA s in gene activation. *FEBS J* 281(1):34–45. <https://doi.org/10.1111/febs.12578>
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform* 9(1):1–13. <https://doi.org/10.1186/1471-2105-9-559>
- Li A, Zhang J, Zhou Z (2014a) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinform* 15(1):1–10. <https://doi.org/10.1186/1471-2105-15-311>
- Li L, Eichten SR, Shimizu R, Petsch K, Yeh C-T, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE (2014b) Genome-wide

- discovery and characterization of maize long non-coding RNAs. *Genome Biol* 15(2):1–15. <https://doi.org/10.1186/gb-2014-15-2-r40>
- Li S, Yamada M, Han X, Ohler U, Benfey PN (2016) High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. *Dev Cell* 39(4):508–522. <https://doi.org/10.1016/j.devcel.2016.10.012>
- Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell* 24(11):4333–4345. <https://doi.org/10.1105/tpc.112.102855>
- Liu Y, Teng C, Xia R, Meyers BC (2020) PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell* 32(10):3059–3080. <https://doi.org/10.1105/tpc.20.00335>
- Ma J, Yan B, Qu Y, Qin F, Yang Y, Hao X, Yu J, Zhao Q, Zhu D, Ao G (2008) Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J Cell Biochem* 105(1):136–146. <https://doi.org/10.1002/jcb.21807>
- Ma X, Shao C, Jin Y, Wang H, Meng Y (2014) Long non-coding RNAs. *RNA Biol* 11(4):373–390. <https://doi.org/10.4161/rna.28725>
- Mattick JS, Rinn JL (2015) Discovery and annotation of long non-coding RNAs. *Nat Struct Mol Biol* 22(1):5–7. <https://doi.org/10.1038/nsmb.2942>
- Meade B, Lafayette L, Sauter G, Tosello D (2017) Spartan HPC-cloud hybrid: delivering performance and flexibility. University of Melbourne, Melbourne, pp 10–49
- Mun J-H, Kwon S-J, Yang T-J, Seol Y-J, Jin M, Kim J-A, Lim M-H, Kim JS, Baek S, Choi B-S (2009) Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol* 10(10):1–18. <https://doi.org/10.1186/gb-2009-10-10-r111>
- Nejat N, Mantri N (2018) Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. *Crit Rev Biotechnol* 38(1):93–105. <https://doi.org/10.1080/07388551.2017.1312270>
- Okada T, Singh MB, Bhalla PL (2005) Expressed sequence tag analysis of *Lilium longiflorum* generative cell. *Plant Cell Physiol* 47(6):698–705
- Okada T, Singh MB, Bhalla PL (2007) Transcriptome profiling of *Lilium longiflorum* generative cells by cDNA microarray. *Plant Cell Rep* 26(7):1045–1052
- Perry RB-T, Ulitsky I (2016) The functions of long noncoding RNAs in development and stem cells. *Development* 143(21):3882–3894. <https://doi.org/10.1242/dev.140962>
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33(3):290–295. <https://doi.org/10.1038/nbt.3122>
- Pokhrel S, Huang K, Bélanger S, Zhan J, Caplan JL, Kramer EM, Meyers BC (2021) Pre-meiotic 21-nucleotide reproductive phasiRNAs emerged in seed plants and diversified in flowering plants. *Nat Commun* 12(1):1–12. <https://doi.org/10.1038/s41467-021-25128-y>
- Quinlan AR (2014) BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinform* 47(1):11.12.11–11.12.34. <https://doi.org/10.1002/0471250953.bi1112s47>
- Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81:145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>
- Risso D (2015) RUVSeq: remove unwanted variation from RNA-seq data. *Bioconductor* <https://bioconductor.org/packages/release/bioc/html/RUVSeq.html>
- Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 43(7):e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rosa S, Duncan S, Dean C (2016) Mutually exclusive sense–anti-sense transcription at FLC facilitates environmentally induced gene repression. *Nat Commun* 7(1):1–7. <https://doi.org/10.1038/ncomms13031>
- Russell SD, Gou X, Wong CE, Wang X, Wei X, Bhalla PL, Singh MB (2012) Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytol* 195(3):560–573
- Sharma N, Russell SC, Bhalla PL, Singh MB (2011) Putative cis-regulatory elements in genes highly expressed in rice sperm cells. *BMC Res Notes* 4:319
- Shumate A, Salzberg SL (2021) Liftoff: accurate mapping of gene annotations. *Bioinformatics* 37(12):1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>
- Simopoulos CM, Weretilnyk EA, Golding GB (2019) Molecular traits of long non-protein coding RNAs from diverse plant species show little evidence of phylogenetic relationships. *G3 Genes Genomes Genet* 9(8):2511–2520. <https://doi.org/10.1534/g3.119.400201>
- Singh MB, Bhalla PL (2007) Control of male germ cell development in flowering plants. *BioEssays* 29(11):1124–1132
- Singh MB, Bhalla PL, Russell SD (2008) Molecular repertoire of flowering plant male germ cells. *Sex Plant Reprod* 21(1):27–36
- Smith MA, Mattick JS (2017) Structural and functional annotation of long noncoding RNAs. *Bioinformatics*. Humana Press, New York, pp 65–85. https://doi.org/10.1007/978-1-4939-6613-4_4
- Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. <https://doi.org/10.12688/f1000research.7563.2>
- Song J-H, Cao J-S, Yu X-L, Xiang X (2007) BcMF11, a putative pollen-specific non-coding RNA from *Brassica campestris* ssp. chinensis. *J Plant Physiol* 164(8):1097–1100. <https://doi.org/10.1016/j.jplph.2006.10.002>
- Song J-H, Cao J-S, Wang C-G (2013) BcMF11, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility. *Plant Cell Rep* 32(1):21–30. <https://doi.org/10.1007/s00299-012-1337-6>
- Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22(2):96–118. <https://doi.org/10.1038/s41580-020-00315-9>
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6(7):e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Törönen P, Medlar A, Hol L (2018) PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res* 46(W1):W84–W88. <https://doi.org/10.1093/nar/gky350>
- Vance KW, Ponting CP (2014) Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* 30(8):348–355. <https://doi.org/10.1016/j.tig.2014.06.001>
- Varabyou A, Salzberg SL, Pertea M (2021) Effects of transcriptional noise on estimates of gene and transcript expression in RNA sequencing experiments. *Genome Res* 31(2):301–308. <https://doi.org/10.1101/gr.266213.120>
- Veeneman BA, Shukla S, Dhanasekaran SM, Chinnaiyan AM, Nesvizhskii AI (2016) Two-pass alignment improves novel splice junction quantification. *Bioinformatics* 32(1):43–49. <https://doi.org/10.1093/bioinformatics/btv642>
- Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. *Mol Cell* 43(6):904–914. <https://doi.org/10.1016/j.molcel.2011.08.018>

- Wang M, Yuan D, Tu L, Gao W, He Y, Hu H, Wang P, Liu N, Lindsey K, Zhang X (2015) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol* 207(4):1181–1197. <https://doi.org/10.1111/nph.13429>
- Wang Y, Luo X, Sun F, Hu J, Zha X, Su W, Yang J (2018) Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat Commun* 9(1):1–9. <https://doi.org/10.1038/s41467-018-05829-7>
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40(7):e49–e49. <https://doi.org/10.1093/nar/gkr1293>
- Ward M, McEwan C, Mills JD, Janitz M (2015) Conservation and tissue-specific transcription patterns of long noncoding RNAs. *J Human Transcript* 1(1):2–9. <https://doi.org/10.3109/23324015.2015.1077591>
- Waseem M, Liu Y, Xia R (2020) Long non-coding RNAs, the dark matter: an emerging regulatory component in plants. *Int J Mol Sci* 22(1):86. <https://doi.org/10.3390/ijms22010086>
- Wei LQ, Xu WY, Deng ZY, Su Z, Xue Y, Wang T (2010) Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genomics* 11(1):1–20. <https://doi.org/10.1186/1471-2164-11-338>
- Wei L, Zhang R, Zhang M, Xia G, Liu S (2022) Functional analysis of long noncoding RNAs involved in alkaline stress responses in wheat. *J Exp Bot* 73(16):5698–5714. <https://doi.org/10.1093/jxb/erac211>
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, Cham. <https://doi.org/10.1007/978-0-387-98141-3>
- Wu H-J, Wang Z-M, Wang M, Wang X-J (2013) Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol* 161(4):1875–1884. <https://doi.org/10.1104/pp.113.215962>
- Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H (2017) FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* 45(8):e57–e57. <https://doi.org/10.1093/nar/gkw1306>
- Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 38(5):662–674. <https://doi.org/10.1016/j.molcel.2010.03.021>
- Yu X, Yang J, Li X, Liu X, Sun C, Wu F, He Y (2013) Global analysis of cis-natural antisense transcripts and their heat-responsive natural siRNAs in *Brassica rapa*. *BMC Plant Biol* 13(1):1–13. <https://doi.org/10.1186/1471-2229-13-208>
- Yu Y, Zhang Y, Chen X, Chen Y (2019) Plant noncoding RNAs: hidden players in development and stress responses. *Annu Rev Cell Dev Biol* 35:407. <https://doi.org/10.1146/annurev-cellbio-100818-125218>
- Yuan J, Zhang Y, Dong J, Sun Y, Lim BL, Liu D, Lu ZJ (2016) Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC Genomics* 17(1):1–16. <https://doi.org/10.1186/s12864-016-2929-2>
- Zhang Y-C, Liao J-Y, Li Z-Y, Yu Y, Zhang J-P, Li Q-F, Qu L-H, Shu W-S, Chen Y-Q (2014) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 15(12):1–16. <https://doi.org/10.1186/s13059-014-0512-1>
- Zhang Z, Guo J, Cai X, Li Y, Xi X, Lin R, Liang J, Wang X, Wu J (2022) Improved reference genome annotation of *Brassica Rapa* by Pacific Biosciences RNA sequencing. *Front Plant Sci* 13:841618. <https://doi.org/10.3389/fpls.2022.841618>
- Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y (2018) Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat Commun* 9(1):1–12. <https://doi.org/10.1038/s41467-018-07500-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.