

RESEARCH

Open Access



Identification of cell-type specific alternative transcripts in the multicellular alga *Volvox carteri*

Ravi Nicholas Balasubramanian^{1,2}, Minglu Gao² and James Umen^{2*}

Abstract

Background Cell type specialization is a hallmark of complex multicellular organisms and is usually established through implementation of cell-type-specific gene expression programs. The multicellular green alga *Volvox carteri* has just two cell types, germ and soma, that have previously been shown to have very different transcriptome compositions which match their specialized roles. Here we interrogated another potential mechanism for differentiation in *V. carteri*, cell type specific alternative transcript isoforms (CTSAI).

Methods We used pre-existing predictions of alternative transcripts and de novo transcript assembly with HISAT2 and Ballgown software to compile a list of loci with two or more transcript isoforms, identified a small subset that were candidates for CTSAI, and manually curated this subset of genes to remove false positives. We experimentally verified three candidates using semi-quantitative RT-PCR to assess relative isoform abundance in each cell type.

Results Of the 1978 loci with two or more predicted transcript isoforms 67 of these also showed cell type isoform expression biases. After curation 15 strong candidates for CTSAI were identified, three of which were experimentally verified, and their predicted gene product functions were evaluated in light of potential cell type specific roles. A comparison of genes with predicted alternative splicing from *Chlamydomonas reinhardtii*, a unicellular relative of *V. carteri*, identified little overlap between ortholog pairs with alternative splicing in both species. Finally, we interrogated cell type expression patterns of 126 *V. carteri* predicted RNA binding protein (RBP) encoding genes and found 40 that showed either somatic or germ cell expression bias. These RBPs are potential mediators of CTSAI in *V. carteri* and suggest possible pre-adaptation for cell type specific RNA processing and a potential path for generating CTSAI in the early ancestors of metazoans and plants.

Conclusions We predicted numerous instances of alternative transcript isoforms in *Volvox*, only a small subset of which showed cell type specific isoform expression bias. However, the validated examples of CTSAI supported existing hypotheses about cell type specialization in *V. carteri*, and also suggested new hypotheses about mechanisms of functional specialization for their gene products. Our data imply that CTSAI operates as a minor but important component of *V. carteri* cellular differentiation and could be used as a model for how alternative isoforms emerge and co-evolve with cell type specialization.

Keywords Gonidia, Soma, Transcript isoforms, Volvocine, *Chlamydomonas*, Convergent transcription, Evolution, Multicellularity

Background

Cell type specialization is a key feature of multicellular organisms that enables development of complex forms and functions, as exemplified by plants and animals with hundreds of cell types forming complex tissues and organs [1–3]. All multicellular eukaryotes evolved from

*Correspondence:

James Umen

JUmen@danforthcenter.org

¹ Yale University, New Haven, CT, USA

² Donald Danforth Plant Science Center, St. Louis, MO, USA



single-celled ancestors with limited capacity for cell-type specialization, and one of the major questions in evolutionary biology is in understanding the origins of cell type specialization in multicellular lineages [1, 3, 4]. In the case of animals and plants, their complexity and ancient divergence from unicellular ancestors makes it challenging to infer how cell type specialization may have arisen.

A major contributor to cell type specialization is differential gene expression. Differential expression is most commonly achieved through combinatorial activities of different transcription factors that control morphogenesis and cell differentiation programs for each cell type in a multicellular individual [5]. Modulation of gene expression through small RNAs (e.g. microRNAs) is another means of controlling gene expression that is widespread in plants and animals [6, 7]. Both above mechanisms modulate transcript abundance, but do not necessarily contribute directly to functional diversity of genes or gene products. Many multicellular eukaryotes employ a third mechanism for regulating gene expression that enables distinct cell-type-specific functions to be expressed from a single genetic locus using alternative transcript isoforms. Differential transcript isoform expression occurs via alternative splicing of pre-mRNAs and/or use of alternative transcription start/termination sites [8–12]. The different transcript isoforms may specify production of structurally and functionally different proteins from the same locus, most dramatically exemplified by complex alternative splicing of some neuronal transcripts in animal brains [13]. This mechanism can promote cell type differentiation when alternative isoforms encode proteins with distinct functional properties. Another less common but well-documented phenomenon for controlling gene expression involves convergent overlapping anti-sense transcripts produced from adjacent loci, where the expression of one transcript antagonizes production of its antisense partner [14–16]. It remains unclear to what extent each of these mechanisms may have been employed during the early transitions of life from single cells to multicellular individuals.

The volvocine algae are a related group of green algae within the Order Chlamydomonadales and are a model system for investigating the step-wise acquisition of multicellular organization and emergent behaviors [17, 18]. The unicellular volvocine species *Chlamydomonas reinhardtii* is an outgroup that is sister to different multicellular volvocine genera that are classified based on their cell number, organismal size and degree of cell type specialization [17, 19]. The estimated divergence time of volvocine algae from a unicellular ancestor is around 200–250 MYA, which is relatively recent compared to plants and animals that shared a common ancestor with

unicellular relatives >500 and >600–800 MYA respectively [17, 20, 21]. Multicellular *Volvox carteri* belongs to the most developmentally complex volvocine genus and has evolved some hallmark features of more complex multicellular organisms including germ-soma cell type differentiation and morphogenetic patterning. Each vegetative phase *V. carteri* individual is a spheroid with a diameter of up to 500 μm and contains ~2000 cells of just two types embedded in a clear structured extracellular matrix (ECM). On the periphery of each spheroid are ~2000 bi-ciliated terminally differentiated somatic cells that provide motility and are specialized for secretion of ECM [22]. Within each spheroid are around 16 large non-ciliated reproductive cells termed gonidia. Upon reaching maturity, each gonidium undergoes embryogenesis which involves a programmed series of twelve or thirteen symmetric and asymmetric cleavage divisions and morphogenesis to produce a juvenile spheroid with ~2000 small somatic precursor cells and 12–16 large gonidial precursor cells. Cell type differentiation is controlled in an unknown manner by cell size at the end of embryogenesis, with cells below 8 μm diameter adopting a somatic fate and cells above 8 μm diameter adopting a gonidial fate [23]. Juveniles remain within their mother spheroid for around a day as they continue to grow, and eventually hatch and mature to begin the two-day vegetative life cycle again.

Early work on cell type differentiation in *Volvox carteri* (Volvox) focused on gene expression programs that are different between the two cell types [24] and light-dependent translational regulation [25]. The Volvox genome sequence [26] facilitated application of more complete transcriptome profiling in the two cell types. Matt & Umen [22] showed a high degree of transcriptome asymmetry between the two vegetative Volvox cell-types suggesting that, like in many other developmental systems, differential transcript abundance is probably the most important component of cell type differentiation. While transcriptional regulation is expected to be a major factor in cell type differentiation in Volvox, other regulatory mechanisms may also be involved including microRNAs, some of which show preferential cell type expression [27, 28]. *C. reinhardtii* (Chlamydomonas) also possesses miRNAs and siRNAs that resemble those in multicellular eukaryotes, implying that such complex RNA-silencing mechanisms evolved before multicellularity [29]. However, there exists little conservation between the Chlamydomonas and the Volvox miRNAs [27–29] suggesting rapid evolution of miRNA loci in the volvocine algae.

An additional potential mechanism for cell-type differentiation that has not been previously examined systematically for Volvox or other multicellular volvocine

species is cell-type specific alternative transcript isoforms (CTSAI) that can be generated through alternative transcription initiation/termination sites or through alternative pre-mRNA processing [30, 31]. Here we made use of prior annotations and previously published cell-type-specific transcriptome data to identify candidate loci with transcripts exhibiting CTSAI. Using two independent prediction methods for alternative isoforms, we found less than 20% of expressed genes had more than one isoform. Among these, the number of genes showing cell type bias in expression of one or both isoforms was less than 100 total, including a high proportion of false positives. Nonetheless, some predicted cases of cell type specific transcript isoforms could be verified and provided potential insights into *V. carteri* cell-type specification. A possible mechanism for controlling cell type specific RNA processing in *Volvox* could come from RNA binding proteins (RBPs) which can mediate alternative splicing in plants or animals. We found several dozen predicted RBP genes in *Volvox* that are expressed in a cell type specific manner, and which could be mediators of CTSAI or CTSAI.

Results

Identifying *Volvox* genes with multiple transcripts

Raw data from a previously published transcriptome study of purified *Volvox* gonidial and somatic cells was re-analyzed for cell-type-specific alternative transcript isoforms (CTSAI) using the Ballgown software package with additional filtering and curation as described below. We first verified that each set of replicates for somatic and gonidial samples showed correlated expression values and that differential cell type expression was comparable to that in the prior study of Matt and Umen, 2018 (Supplemental Figures S1 and S2). CTSAI can only occur if two or more transcript isoforms are produced from the same gene. Out of the 11,095 *Volvox* genes that had average expression across all samples >1 FPKM, 9860

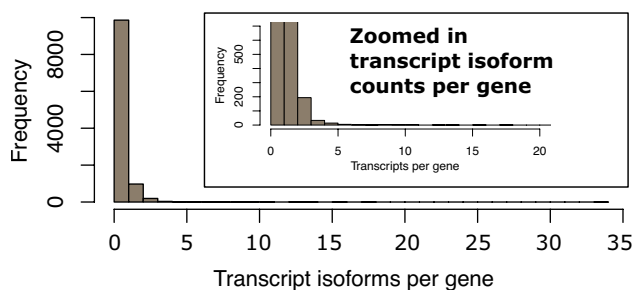


Fig. 1 Histogram of transcript counts per gene. 9860 genes have one isoform, 974 have two isoforms, while the remaining 261 genes have three or more isoforms. The largest number of isoforms per gene is 34. The inset in the top right corner is the histogram zoomed up at lower numbers of isoforms per gene

had only one transcript isoform and were not analyzed further (Fig. 1). 1235 remaining genes had multiple predicted transcripts and were candidates for CTSAI. 974 of these genes (78%) had just two predicted isoforms and were analyzed further while those with three or more isoforms were not.

To quantify cell type expression bias of transcript isoforms we first computed the \log_2 fold change as \log_2 (expression level in soma)/(expression level in gonidia). Large ratios ($>>1$) indicated bias towards soma, while small ratios ($<<1$) indicated bias towards gonidia. CTSAI are present when the computed cell type ratios of each transcript isoform significantly differ from each other.

We focused on genes whose transcript isoforms showed opposite cell type biases. We therefore considered cases where one transcript isoform was somatic-biased (s/g ratio >2) while the other isoform was gonidial-biased (s/g ratio $<1/2$) with p -values for differential expression <0.1 . Additionally, we required the overall cell type bias, defined as [isoform 1 (s/g)/ isoform 2 (s/g)], to be greater than 8. 60 Genes passed these two

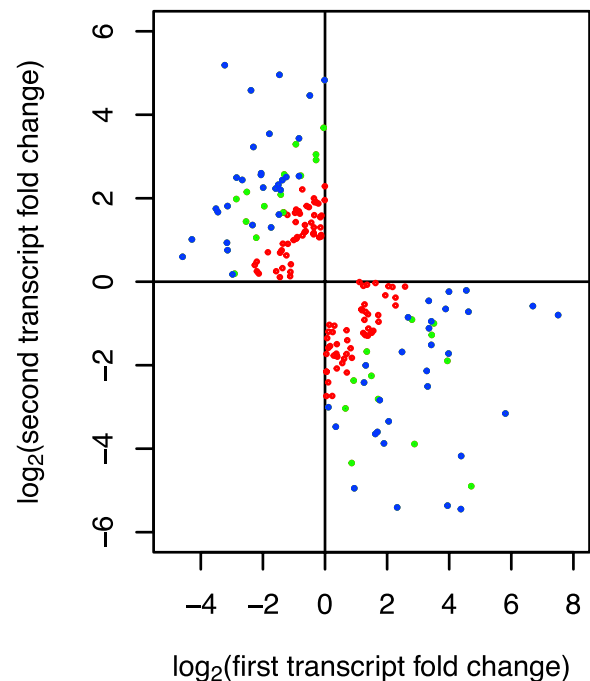


Fig. 2 Sequentially applied filters for identifying CTSAI genes. The first cutoff criterion is that \log_2 ER is greater or less than one in one of the cell types and is expressed oppositely in the other cell type (187 candidates, red points). The second cutoff criterion is that the total cell type bias (ER soma/ER gonidia) is greater than 8 (86 candidates, green points). The third criterion selects genes where the p -values for isoform expression ratio in each cell type are less than 0.1 (aggregate p value <0.01 ; 60 genes, blue points). Plots show \log_2 expression ratios (ER = expression in soma over gonidia) for second vs. first isoform in genes with two isoforms and opposite cell-type bias with color coding as described above

criteria and were assigned as candidates for CTSAI with an aggregate p -value < 0.01 since the individual p -values can be multiplied (Fig. 2). Results using alternative filter cutoffs with higher or lower stringencies are in Supplemental Figures S3 and S4. Expression data for all 60 genes are in Supplemental Table S1.

Phytozome predicted alternative transcripts

To supplement our analysis using Ballgown software which relied on de novo transcript assembly, we repeated the same analysis above using alternative transcripts predicted from the version 2.1 genome assembly of *V. carteri* hosted on Phytozome [26, 32, 33]. Out of the 9750 genes that had average expression across all samples > 1 FPKM, 9297 only had one isoform and were removed from the analysis. Of the remaining 453 genes, which had multiple isoforms, 415 had only two isoforms, and were analyzed further.

We proceeded to use the same filtering criteria as were used for Ballgown-predicted transcripts – opposite expression ratios between cell types for the two isoforms, at least one expression ratio in the isoform pair being greater than 2 or less than 0.5, an overall cell-type-bias of the two isoforms greater than 8, and p -values for differential expression of each individual isoform less than 0.1. This filtering resulted in identification of 20 candidate genes from Phytozome for CTSAI, 13 of which were predicted by Ballgown and 7 of which were unique.

Manual curation

Each of the 67 combined total CTSAI candidate loci predicted by Ballgown and Phytozome were visualized on the Phytozome browser [32, 33] along with transcriptome coverage in each cell type replicate [22] to assess the results of automated alternative transcript predictions and the evidence for presence of CTSAI (Supplemental Table S2). We further filtered candidates following some additional criteria: a) overall read mapping of both isoforms was sufficiently dense to support transcript structures [RD—read density]; b) reads mapping to differential portions of each transcript supported a significant difference in isoform abundance by cell type and not just overall transcript abundance between cell types [REL – relative expression level] (examples in Supplemental Figure S5). In doing so we identified 52 false positives that were excluded for one or more reasons as described above while retaining 15 genes as strong candidates for having CTSAI – 12 were in the Ballgown analysis and 4 were in the Phytozome analysis, with one gene in common between the two sets (Supplemental Table S3).

Of the 12 genes that were identified as strong candidates for CTSAI in the Ballgown analysis, 10 were not in the set of candidates for CTSAI from the Phytozome

analysis because Phytozome did not previously identify alternate isoforms for them or predicted that those genes have more than two isoforms (exon coordinates of these twelve genes are in Supplemental Document S1).

Convergent transcripts

Among our 67 candidate CTSAI loci, there were 4 clear cases where regions of overlapping convergent transcription from distinct adjacent loci were mis-identified by Ballgown as alternative isoforms (letter code in Supplemental Table S1 = CT). The Phytozome IDs that were assigned to these convergent genes by Ballgown were Vocar20012504m.g (version 2.0), Vocar.0004s0224, Vocar.0042s0098, and Vocar.0006s0317 (version 2.1). Although the available transcriptome reads used by Ballgown did not contain strand information, during manual curation the direction of transcription could be inferred for each transcript from a combination of Phytozome gene model predictions, annotations, ESTs, and/or polarity of consensus splice site sequences in each model. Notably, in most instances of convergent transcription that were originally mis-identified as CTSAI, the cell type specificity was correctly estimated, meaning that a different member of the convergent pair dominated in each of the two cell types. These convergently transcribed overlapping gene pairs might, therefore, have an antagonistic regulatory relationship that contributes to their cell type expression specificity (see DISCUSSION). Examples of cell type specific convergent transcript pairs are shown in Supplemental Figures S6a and b.

Experimental validation of alternative splicing

We selected three genes showing CTSAI for further validation: Vocar.0001s1758, Vocar.0011s0285, and Vocar.0018s0107 (Figs. 3). Vocar.0001s1758 encodes a predicted UDP-glucose pyrophosphorylase and is required for UDP-glucose synthesis, a critical substrate for protein glycosylation. Vocar.0011s0285 has a universal stress protein (USP) domain (PFAM 00582) and its *Chlamydomonas* ortholog, Cre03.g211521, is a cilia-associated protein, FAP165 [34]. Vocar.0018s0107 is a predicted flavin-dependent oxido-reductase with three flavodoxin binding domains (PFAM 00258) followed by a FAD domain (PFAM 00667). To confirm the bioinformatic results, we PCR amplified *Volvox* cDNA produced from one of the replicate RNA samples [22] with primers that were isoform specific. We created 3 primers for each tested gene: a shared primer which would bind to both transcript isoforms, and a specific primer for each isoform that could pair with the shared primer. We used isoform specific primer sets alone or combined in single PCR reactions from each cell type, and assessed relative

band intensities. For all three genes we found a qualitative match to the cell type splicing patterns predicted using transcriptome data (Fig. 4, Supplemental Figure S7).

The gonidial form of Vocar.0001s1758 has 13 exons, while the somatic form has one extra exon at the 3' end which is joined to an internal 5' splice site within exon 13 of the gonidial isoform (Fig. 3). We compared transcript structures and the protein sequences encoded by Vocar.0001s1758 isoforms to that of the orthologous gene in *Chlamydomonas reinhardtii* (*Chlamydomonas*), Cre04.g229700, which has a single isoform that can be used as a potential proxy for the ancestral transcript structure and protein sequence (Supplemental Figure S8). Cre04.g229700 has an overall similar structure to that of the *Volvox* transcripts, but with two extra introns that are not present in either *Volvox* isoform. At its 3' end the *Chlamydomonas* transcript matches the structure of

the gonidial isoform in *Volvox* which is most likely the ancestral gene structure, with the somatic form being derived. Both predicted transcript isoforms retained the coding region for the catalytic domain and differ in their C-terminal domains. UDP-glucose pyrophosphorylases show highest activity as monomers, but can also oligomerize to a form that is less active [35]. The solved crystal structure and additional conformational modeling of *Arabidopsis thaliana* UDP-glucose pyrophosphorylase paralog At3g03250 revealed a potential explanation for the impact of dimerization on activity: in dimeric form the C-terminal portion of each subunit partly occludes the active site in the N-terminal domain [36]. Thus, altering the C-terminal region in the two *Volvox* isoforms has the potential to affect dimerization efficiency and/or activity of the dimeric form, either of which could cause somatic and gonidial isoforms of Vocar.0001s1758 to have different activities and/or allosteric regulatory responses.

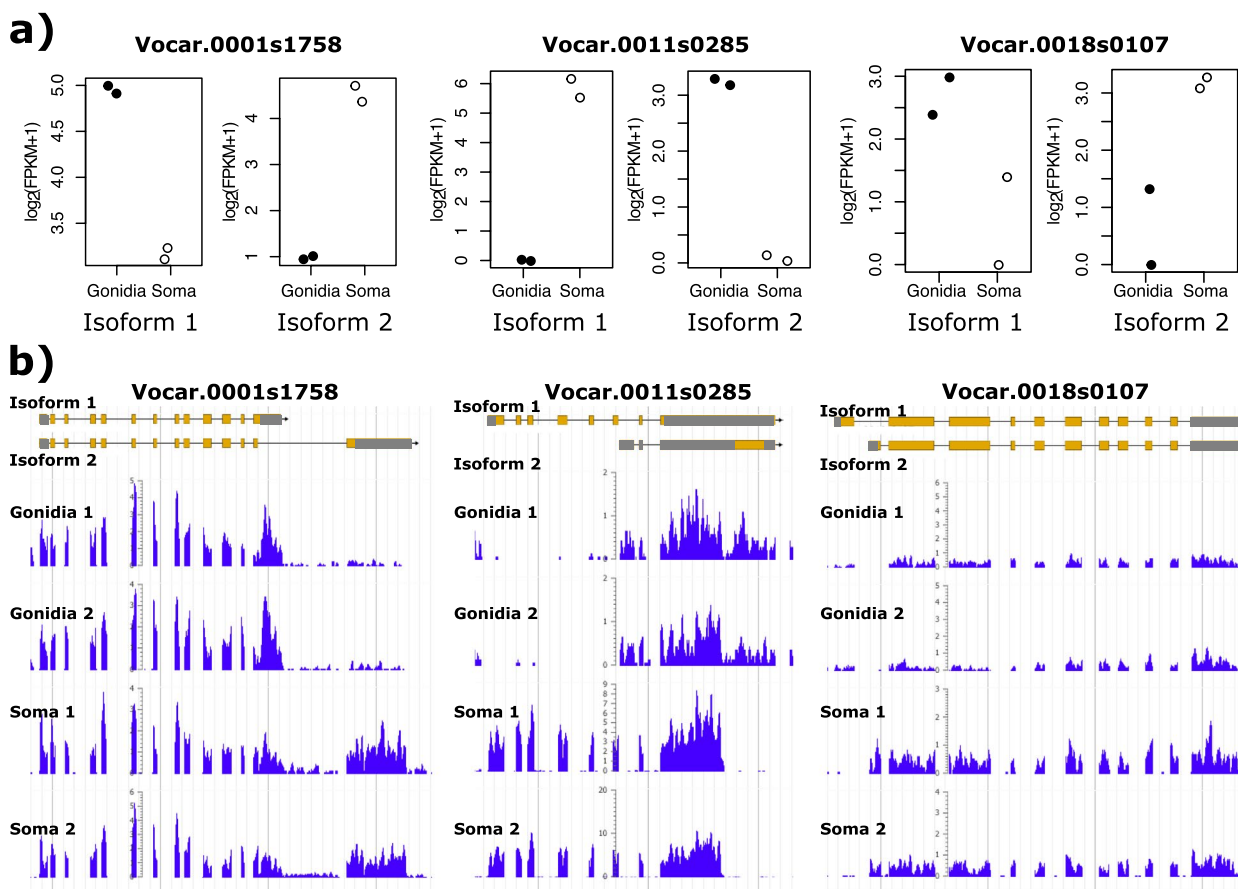


Fig. 3 Three of the top candidates for CTSAl identified in *V. carteri*. **a)** CTSAl expression estimates by cell type for Vocar.0001s1758, Vocar.0011s0285, and Vocar.0018s0107 genes. Dark circles and light circles show replicate measurements in gonidia and soma respectively. **b)** Transcriptome sequence coverage supporting CTSAl. Left, middle and right panels are browser screen shots from Phytozome of Vocar.0001s1758, Vocar.0011s0285, and Vocar.0018s0107 respectively with the two isoforms diagrammed on top and duplicate somatic and gonidial tracks showing read density across the locus

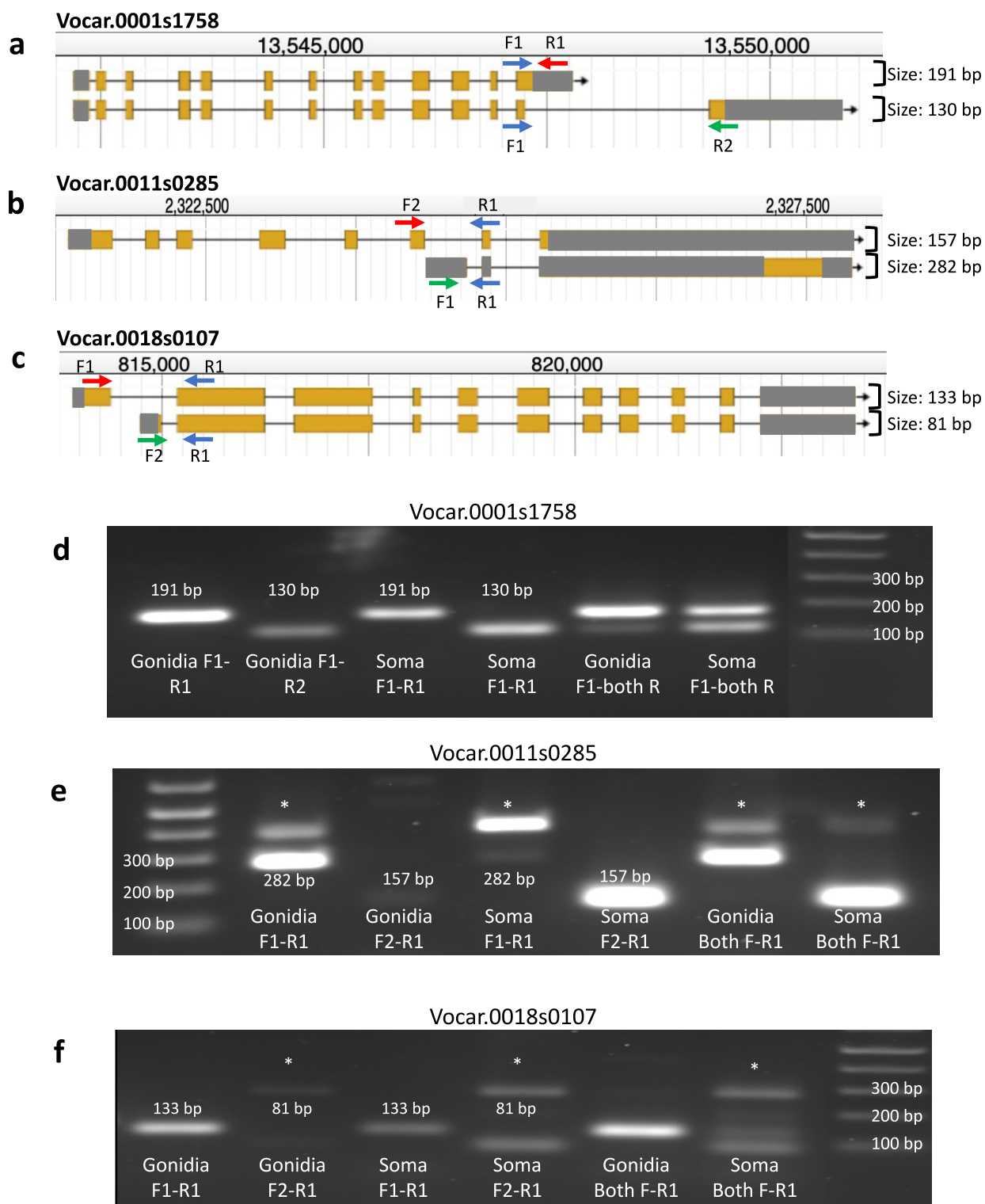


Fig. 4 a-c) Graphic displaying primer binding sites and amplicon lengths for RT-PCR primer sets used for Vocar.0001s1758, Vocar.0011s0285 and Vocar.0018s0107, respectively. Arrows show primer direction with the common primer in blue and the isoform specific primers in red and green. Predicted product lengths are also indicated. d-f) Gel electrophoresis images from amplification reactions using primers in panels a-c. Lanes are marked with primer combinations and cell types. The end lanes have DNA size markers. Asterisks indicate non-specific amplification products

While there is no direct evidence of differential activity of UDP-glucose pyrophosphorylase in *V. carteri* cell types, it is striking that nearly every enzyme in the starch-sugar nucleotide synthesis pathway in *V. carteri* has a transcript which is subject to cell type regulation with high expression in gonidia and low expression in soma—except for Vocar.0001s1758 which shows no difference in overall transcript abundance between cell types [22]. Our data showing that this gene produces cell type specific alternative isoforms of UDP-glucose pyrophosphorylase provides a rationale for why cell type expression level control is not observed for this member of the pathway, as regulation of activity is achieved post-transcriptionally by cell type specific alternative splicing.

The predicted somatic isoform of Vocar.0011s0285 encodes a protein that is orthologous to the *Chlamydomonas* gene Cre03.g211521 (Supplemental Figure S8), while the gonidial isoform uses an internal transcription start site and excludes the first six exons found in the somatic isoform (Fig. 4b). Instead, the gonidial transcript begins with a unique alternative 5' exon followed by a short penultimate exon and final exon, both of which are shared with the somatic isoform. The Cre03.g211521 predicted protein was found in the ciliary proteome and designated FAP165, but its function has not been characterized [34]. Its only domain (also shared by the somatic isoform of Vocar.0011s0285) is PFAM 00582/IPR006016 that is designated a universal stress protein domain [37], though its specific function in stress survival is unclear. Interestingly, several other chlorophyte algae, mostly within the *Chlamydomonadales*, have homologs of Cre03.g211521 with similarity extending outside of the conserved N-terminal PFAM 00582/IPR006016 domain. The closest matches of this protein outside of related green algae are in prokaryotes, and only in the conserved PFAM 00582/IPR006016 domain. It is thus possible that homologs of Cre03.g211521/Vocar.0011s0285 originated as a horizontal gene transfer event from prokaryotes into one or more subgroups of chlorophycean green algae. How the function of this protein domain was incorporated into cilia is an intriguing question. Our finding that only the somatic isoform of Vocar.0011s0285 can produce a FAP165-related protein supports the idea that this protein functions in the cilium of volvocine algae since most cilia-related proteins are not expressed at all in *V. carteri* gonidial cells [22]. Indeed, the gonidial isoform of Vocar.0011s0285 appears to be a non-coding transcript produced from an internal initiation site whose only significant open reading frame would encode a short peptide that is unrelated to the somatic protein isoform. Whether the gonidial transcript has a function is unknown, as is the reason why this transcript came under cell type regulation by alternative transcript isoforms and not through

cell type expression as is the case for most of the cilia-related genes in *Volvox* [22].

The two isoforms of Vocar.0018s0107 have different transcription start sites and first exons but share all remaining exons in common and encode predicted proteins that differ in their N-termini (Fig. 4c, Supplemental Figure S8). The two predicted *Volvox* proteins are equally similar to the *Chlamydomonas* ortholog encoded at locus Cre16.g683550 which does not align with either *Volvox* protein in the N-terminal region. Interestingly, the predicted localizations for the two *Volvox* isoforms by the Predalgo algorithm are different, with the somatic isoform predicted to be imported to the chloroplast and the gonidial isoform predicted to be excreted [38]. The *Chlamydomonas* ortholog is also predicted to enter the secretory pathway, and this may be the ancestral localization for the protein. Vocar.0018s0107 is a predicted oxidoreductase with three flavodoxin domains (IPR008254) followed by a ferredoxin reductase-type FAD binding domain (IPR017927). These flavodoxin domains often serve as redox reaction activators [39]. Interestingly this domain architecture is shared with a family of predicted proteins in many other green algae and in the phytophthora clade of downy mildew plant pathogens, but to our knowledge has not been functionally characterized in any species. Its alternative splicing in *Volvox* may influence its function via differential subcellular localization in somatic versus gonidial cells, but its specific biochemical function(s) and substrates remain to be determined.

Volvox and Chlamydomonas orthologs with alternative splicing or transcript isoforms

Little is known about the origin and persistence of alternative splicing or transcript isoforms in volvocine algae. The last common ancestor of *Chlamydomonas reinhardtii* and *Volvox carteri* existed around 200 MYA [40], but they still share many 1:1 orthologous gene pairs [26]. However, it is unknown if alternative transcript isoforms were conserved between the two species. To estimate the degree of conservation of alternative splicing patterns in homologous genes of *V. carteri* and *C. reinhardtii*, we compared whether presence of alternative isoforms was a feature for each of 10,447 predicted 1:1 orthologs between the two species identified based on mutual best hit similarity searching [22]. The merged Ballgown and Phytozome data set of genes with two isoforms (1978 genes total without duplicates) had 1520 in the 1:1 candidate ortholog list. We then analyzed data from a recent study in *Chlamydomonas* that identified 2182 alternatively spliced genes [41], and detected 1441 that were also found in the 1:1 candidate ortholog dataset. We found 273 gene pairs where both members showed alternative splicing. We repeated the same analysis using

Phytozome v5.5 predictions of alternative transcripts, which contained 1531 genes, 790 of which were in the 1:1 homolog dataset, and 161 of these had multiple isoforms in both species. However, among the set of 15 validated CTSAI candidate genes, only one had alternative isoforms in *Chlamydomonas*. This suggests that new cases of CTSAI in *Volvox* mostly evolved from genes that previously did not have cell type specific alternative splicing patterns or ones where it was lost in the lineage leading to *C. reinhardtii*. This result of low conservation of CTSAI is consistent with high variability of alternative splicing between species also found in land plants and vertebrates [42–44]. However, we cannot rule that some *C. reinhardtii* genes whose *V. carteri* counterparts show CTSAI may have undetected alternative splicing. Overall, there appears to be minimal conservation of alternative splicing in volvocine algae, suggesting that this feature of gene expression is labile, a property that could facilitate the evolution of cell-type specific alternative transcript functions.

Cell type specific RNA binding proteins are potential mediators of alternative splicing

The identification of cell type alternative splicing in *Volvox* implies that there is some cell type specificity to its RNA processing machinery. One broad class of proteins that can influence alternative RNA processing are those with RNA binding domains [45, 46]. We therefore identified 126 predicted RNA binding domain proteins in *Volvox* based on their automated domain annotations, including RNA recognition motifs (PFAM 00076), KH domains (PFAM 00013), G patch domains (PFAM 01585), RNA helicase domains (PFAM 00270/1) and several others (Supplemental Tables S4 and S5). Proteins with these domains whose expression is cell type specific are candidate mediators of cell type specific alternative splicing. We queried the transcripts for these 126 predicted RNA binding proteins for their cell type expression patterns as determined in our earlier study [22] and found that most were expressed at similar levels in both cell types (< twofold cell type expression ratio) or showed modest cell type bias (2–fivefold cell type expression ratio). However, 40 predicted RNA binding protein coding genes had transcripts that were gonidial specific and 4 that were somatic specific (> fivefold cell type expression ratio) (Supplemental Tables S4 and S5). These 44 proteins are candidate mediators of CTSAS. Many are related to RNA metabolism, but among them are the somatic biased gene encoding predicted U2 auxiliary factor large subunit gene (U2AF 68) which is required for pre-mRNA splice site recognition, and gonidial biased gene transcription termination factor 64 (CTSF 64) required for 3' end formation.

Discussion

Based on bioinformatics predictions alone, the frequency of genes with two alternative isoforms in *Volvox* was relatively high, with around 12–13% of all loci predicted to have two or more alternative isoforms. 19.7% of expressed genes in *Chlamydomonas reinhardtii* and about 42% of intron containing genes in *Arabidopsis thaliana* are alternatively spliced [41, 47], so the overall frequency of alternative transcript isoforms in *V. carteri* appears to be lower; but the differences in these estimates could also be due to the amount of data available and the methods used to identify alternative transcripts. Moreover, the sexual cycle in *Volvox* involves modified embryogenesis and differentiation of additional cell types including eggs, sperm and somatic sexual cells, and these might also partly rely on alternative splicing of genes only expressed in the sexual phase of the life cycle that was not queried here [26]. We note also that there was evidence of extensive alternative isoforms in the *MAT3/RBR* gene residing in the sex determining region (SDR) of *Volvox* [48] and it is likely that this region of the genome has additional diversity of alternative isoforms, but is not fully represented in the Phytozome *Volvox* genome assembly [32, 33].

Only 67 loci in *Volvox* that had two predicted alternative isoforms also showed cell type bias for expression of those isoforms. Moreover, of the 67 loci with high confidence and strong cell type isoform expression bias, most were false positives. In some cases, the transcripts that were predicted by Ballgown were not supported by expression data, or the expression data was too sparse to adequately support the existence of an alternative transcript. A few false positives were due to adjacent genes being so close together that Ballgown fused them into erroneous transcript predictions. However, many false positives were also found among the Phytozome predictions, and these errors became apparent when visually inspecting transcriptome read mapping data (Supplemental Figure S5c). For future work it would be helpful to use Iso-seq type methods [49] to capture full length transcripts and/or use sequencing methods that preserve transcript strand information to help reduce false positives.

During curation of automated prediction of CTSAS/CTSAI we also encountered cases of cell-type specific convergent transcription in which expression for a different member of the convergent pair dominates in each cell type. The overlapping 3'UTRs in these examples may be a mechanism for gene regulation that functions by antagonizing production, translation or mRNA stability of a convergent partner [15, 50]. A focused search for more examples of convergent transcripts with opposite cell type expression patterns could be used to determine

how widespread is this phenomenon for possibly generating or reinforcing cell-type-specific transcript levels in *Volvox*.

RNA binding proteins are potential mediators of CTSAS. Here we identified 126 predicted RNA binding proteins from *Volvox*, and a subset of 44 that showed cell type specific expression bias. These 44 may be potentiators for the evolution of CTSAS which could occur through direct binding to pre-mRNA sequences that influence splice site or transcription start/termination site usage [45, 46]. Moreover, there exist other possible means of regulating transcript processing such as riboswitches interacting with metabolites or small molecules [51], through small RNA (sRNA) or microRNA (miRNA) pathways [52], or by modulation of transcription start/termination sites [31]. The prevalence and relative importance of these different mechanisms for differentiation of *Volvox* is not known, and they are worthy of further investigation.

Conclusions

Our study successfully identified and characterized cell type specific transcript isoform usage in the multicellular alga *Volvox carteri*. Although many candidates initially identified using automated approaches were false positives, we nonetheless identified fifteen cases that passed manual inspection, and three of these were validated experimentally. These cases immediately suggested functional hypotheses about cell type specific functions of alternative transcripts. We also identified RNA binding proteins with cell-type expression biases that may serve as mediators for CTSAS. Additionally, our manual curation of CTSAs candidates revealed multiple cases of cell-type specific convergent transcription which may be an underappreciated means of enforcing mutually exclusive cell type specificity for each gene of the transcript pair. In summary, cell type specific alternative transcript isoforms seem to represent only a minor component of cell differentiation in *Volvox* where biased cell type expression was found previously for thousands of genes and appears to be the predominant means of cell differentiation in this simple multicellular species. Overall, our data imply that CTSAs operates as a component of *V. carteri* cellular differentiation, and this simple multicellular organism could be a model for how alternative isoforms emerge and co-evolve with cell type specialization.

Methods

HISAT2 RNA-seq protocol

We used HISAT2 RNA-seq software with the default settings laid out in the original protocol [53] (Supplemental Figure S9) to analyze previously published *Volvox carteri* transcriptome data. First, RNA-seq reads obtained from

Matt & Umen, 2018 [22] were passed through HISAT2 and aligned to a reference annotation of the *Volvox* genome available from Phytozome [32, 33]. We made annotation file following the methods of Matt and Umen, 2018 [22] by combining all 14,247 *V. carteri* version 2.1 gene models currently available on Phytozome [26, 32] with 1761 nonoverlapping (<20% sequence overlap) version 2.0 Phytozome gene models [26, 32] to generate a final set of 16,008 unique protein-coding gene models. Aligned reads were merged together into transcripts whose abundances were estimated using the StringTie software package. We then used StringTie to merge the transcripts across samples to get a full transcriptome for analysis, allowing us to restore transcripts with missing fragments in individual samples to their full length. Transcript abundances were then re-estimated by StringTie and normalized in Fragments per Kilobase per Million Reads (FPKM). Identification numbers were assigned to each transcript and gene, and when possible, gene names from the reference annotation in Phytozome [26, 32, 33] were associated with each locus. Previously unidentified transcripts were indexed by the Ballgown-generated ID number of the associated gene. We applied an expression filter of 1 FPKM to each transcript across all samples to remove low abundance transcripts prior to our analysis. Finally, we used the Ballgown software package [54] to identify candidate transcript isoforms with significant differences in expression between somatic and gonidial transcriptomes.

RT-PCR

RT-PCR experiments were performed using the methods and RNA samples of Matt and Umen (2018). cDNA was prepared from two somatic and two gonidial samples described in Matt and Umen (2018) with 5 µg total RNA following the manufacturer's protocol for ThermoScript (Invitrogen, Carlsbad, CA) using a 10:1 mixture of oligo dT and random hexamer for priming and the following cDNA synthesis reaction temperatures: 25 °C 10', 42 °C 10', 50 °C 20', 55 °C 20', 60 °C 20', 85 °C 5'. After cDNA synthesis reactions were treated with RNaseH and diluted 1:10 with 10 mM Tris pH 8.0, 1 mM EDTA (TE) and stored at -20 °C. PCR amplification reactions (20 µL total per reaction) contained 1X Taq polymerase buffer, 0.2 mM dNTPs, 2% DMSO, 1 µL purified recombinant Taq polymerase, 0.5 mM each primer (Supplemental Table S6) and 1 µL cDNA template. Samples were amplified in a peqSTAR thermocycler as follows 1) 94 °C for 2 min; 2) 35 cycles of 94 °C for 25 s, 55 °C for 30 s, and 72 °C for 30 s; 3) 72 °C for 10 min. After thermocycling, samples were separated on a 2% agarose gel containing ethidium bromide and imaged using a UV transilluminator and CCD camera.

Abbreviations

CTSAI	Cell-type specific alternative isoforms
CTSAS	Cell-type specific alternative splicing
ECM	Extracellular matrix
RT-PCR	Reverse transcription polymerase chain reaction
UDP	Uridine diphosphate

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09558-0>.

Additional file 1: Figure S1. Verification gene and transcript expression within replicates. **Figure S2.** Verification of differences in gene and transcript expression between cell types. **Figure S3.** Alternate restriction criteria: all genes with two transcripts. **Figure S4.** Alternate restriction criteria: two oppositely expressed transcripts with large biases. **Figure S5.** Hand Curation Examples. **Figure S6.** Convergent Transcription Examples. **Figure S7.** Full-length gel-electrophoresis images for CTSAI verification experiments. **Figure S8.** Three-way alignments with *Chlamydomonas* homologs. **Figure S9.** HISAT2-Ballgown protocol schematic. **Document S1.** Exon coordinates for the manually curated HISAT2 candidates.

Additional file 2: Table S1. HISAT2 candidate genes. **Table S2.** HISAT2 and Phytozome candidate genes. **Table S3.** Final set of candidate genes. **Table S4.** RNA binding proteins of *V. carteri* with *A. thaliana* homologs. **Table S5.** Full functional annotations of the RNA binding proteins of *V. carteri*. **Table S6.** PCR primers for experimental verification experiments.

Acknowledgements

We acknowledge the kind mentorship of Gavriel Matt in training R.N.B. with lab methods and experimental protocols.

Author's contributions

RNB performed all bioinformatic analyses with supervision from JU. RNB and JU manually curated candidate genes with CTSAI. MG and RNB designed and performed RT-PCR validation experiments. RNB and JU wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by National Science Foundation Grant IOS 1755430 to JU.

Availability of data and materials

The datasets produced, used, and analyzed during the study are available from the corresponding author upon request. Transcriptome data are available at the NCBI Gene Expression Omnibus repository under accession number GSE104835.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 March 2023 Accepted: 6 August 2023

Published online: 30 October 2023

References

- Rueffler C, Hermisson J, Wagner GP. Evolution of functional specialization and division of labor. *Proc Natl Acad Sci U S A*. 2012;109:E326–35.
- Kirk DL. Developmental and cell biology series: *Volvox*: A search for the molecular and genetic origins of multicellularity and cellular differentiation series number 33. Cambridge, England: Cambridge University Press; 2011.
- Buss LW. Evolution, development, and the units of selection. *Proc Natl Acad Sci U S A*. 1983;80:1387–91.
- Buss LW. The evolution of individuality. Princeton, NJ: Princeton University Press; 2016.
- Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. *Nat Rev Genet*. 2016;17:744–57.
- Cui J, You C, Chen X. The evolution of microRNAs in plants. *Curr Opin Plant Biol*. 2017;35:61–7.
- Zhang B, Wang Q, Pan X. MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol*. 2007;210:279–89.
- Brett D, Pospisil H, Valcárcel J, Reich J, Bork P. Alternative splicing and genome complexity. *Nat Genet*. 2002;30:29–30.
- Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*. 2002;418:236–43.
- Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*. 2017;18:437–51.
- Szakonyi D, Duque P. Alternative splicing as a regulator of early plant development. *Front Plant Sci*. 2018;9:1174.
- Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet*. 2022;23:697–710.
- Su C-H, Dhananjaya D, Tarn W-Y. Alternative splicing in neurogenesis and brain development. *Front Mol Biosci*. 2018;5:12.
- Wright BW, Molloy MP, Jaschke PR. Overlapping genes in natural and engineered genomes. *Nat Rev Genet*. 2021;23:1–15.
- Gullerova M, Proudfoot NJ. Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. *Nat Struct Mol Biol*. 2012;19:1193–201.
- Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. *Nat Rev Genet*. 2013;14:880–93.
- Umen JG. *Volvox* and volvocine green algae. *Evodevo*. 2020;11:13.
- Balasubramanian RN, McCourt RM. *Volvox barberi* (Chlorophyceae) actively forms two-dimensional flocks in culture. *J Phycol*. 2021;57:967–74.
- Herron MD. Origins of multicellular complexity: *Volvox* and the volvocine algae. *Mol Ecol*. 2016;25:1213–23.
- Umen JG. Green algae and the origins of multicellularity in the plant kingdom. *Cold Spring Harb Perspect Biol*. 2014;6:a016170.
- Leys SP, Kahn AS. Oxygen and the energetic requirements of the first multicellular animals. *Integr Comp Biol*. 2018;58:666–76.
- Matt GY, Umen JG. Cell-type transcriptomes of the multicellular green alga *Volvox carteri* yield insights into the evolutionary origins of germ and somatic differentiation programs. *G3 (Bethesda)*. 2018;8:531–50.
- Kirk MM, Ransick A, McRae SE, Kirk DL. The relationship between cell size and cell fate in *Volvox carteri*. *J Cell Biol*. 1993;123:191–208.
- Tam LW, Stamer KA, Kirk DL. Early and late gene expression programs in developing somatic cells of *Volvox carteri*. *Dev Biol*. 1991;145:67–76.
- Kirk DL, Kirk MM. Protein synthetic patterns during the asexual life cycle of *Volvox carteri*. *Dev Biol*. 1983;96:493–506.
- Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, et al. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*. 2010;329:223–6.
- Dueck A, Evers M, Henz SR, Unger K, Eichner N, Merkl R, et al. Gene silencing pathways found in the green alga *Volvox carteri* reveal insights into evolution and origins of small RNA systems in plants. *BMC Genomics*. 2016;17:853.
- Li J, Wu Y, Qi Y. MicroRNAs in a multicellular green alga *Volvox carteri*. *Sci China Life Sci*. 2014;57:36–45.
- Molnár A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature*. 2007;447:1126–9.
- Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet*. 2010;11:75–87.
- de Klerk E, t Hoen PAC. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet*. 2015;31:128–39.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(Database):D1178–86.

33. Phytozome v13. <http://phytozome.jgi.doe.gov/pz/portal.html>. Accessed 14 Sep 2022.
34. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol.* 2005;170:103–13.
35. Martz F, Wilczynska M, Kleczkowski LA. Oligomerization status, with the monomer as active species, defines catalytic efficiency of UDP-glucose pyrophosphorylase. *Biochem J.* 2002;367(Pt 1):295–300.
36. McCoy JG, Bitto E, Bingman CA, Wesenberg GE, Bannen RM, Kondrashov DA, et al. Structure and dynamics of UDP-glucose pyrophosphorylase from *Arabidopsis thaliana* with bound UDP-glucose and UTP. *J Mol Biol.* 2007;366:830–41.
37. Freestone P, Nyström T, Trinei M, Norris V. The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol.* 1997;274:318–24.
38. Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugière S, et al. Pre-dAlgo: a new subcellular localization prediction tool dedicated to green algae. *Mol Biol Evol.* 2012;29:3625–39.
39. Hall DA, Vander Kooi CW, Stasik CN, Stevens SY, Zuiderweg ER, Matthews RG. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. *Proc Natl Acad Sci U S A.* 2001;98:9521–6.
40. Herron MD, Hackett JD, Aylward FO, Michod RE. Triassic origin and early radiation of multicellular volvocine algae. *Proc Natl Acad Sci U S A.* 2009;106:3254–8.
41. Pandey M, Stormo GD, Dutcher SK. Alternative splicing during the *Chlamydomonas reinhardtii* cell cycle. G3 (Bethesda). 2020;10:3797–810.
42. Ling Z, Brockmüller T, Baldwin IT, Xu S. Evolution of alternative splicing in eudicots. *Front Plant Sci.* 2019;10:707.
43. Verta J-P, Jacobs A. The role of alternative splicing in adaptation and evolution. *Trends Ecol Evol.* 2022;37:299–308.
44. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science.* 2012;338:1587–93.
45. Bandziulis RJ, Swanson MS, Dreyfuss G. RNA-binding proteins as developmental regulators. *Genes Dev.* 1989;3:431–7.
46. Fu X-D, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet.* 2014;15:689–701.
47. Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, et al. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 2010;20:45–58.
48. Ferris P, Olson BJSC, Hoff PLD, Douglass S, Casero D, Prochnik S, et al. Evolution of an expanded sex-determining locus in *Volvox*. *Science.* 2010;328:351–4.
49. Gonzalez-Garay ML. Introduction to isoform sequencing using pacific Biosciences technology (Iso-seq). In: *translational bioinformatics*. Dordrecht: Springer; 2016. p. 141–60.
50. Wang L, Jiang N, Wang L, Fang O, Leach LJ, Hu X, et al. 3' Untranslated regions mediate transcriptional interference between convergent genes both locally and ectopically in *Saccharomyces cerevisiae*. *PLoS Genet.* 2014;10: e1004021.
51. Winkler WC, Breaker RR. Genetic control by metabolite-binding riboswitches. *ChemBioChem.* 2003;4:1024–32.
52. Vaucheret H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* 2006;20:759–71.
53. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat Protoc.* 2016;11:1650–67.
54. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* 2015;33:243–6.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

