

RESEARCH ARTICLE

Open Access



# Whole transcriptome RNA-Seq analysis reveals extensive cell type-specific compartmentalization in *Volvox carteri*

Benjamin Klein<sup>1</sup>, Daniel Wibberg<sup>2</sup> and Armin Hallmann<sup>1\*</sup>

## Abstract

**Background:** One of evolution's most important achievements is the development and radiation of multicellular organisms with different types of cells. Complex multicellularity has evolved several times in eukaryotes; yet, in most lineages, an investigation of its molecular background is considerably challenging since the transition occurred too far in the past and, in addition, these lineages evolved a large number of cell types. However, for volvocine green algae, such as *Volvox carteri*, multicellularity is a relatively recent innovation. Furthermore, *V. carteri* shows a complete division of labor between only two cell types – small, flagellated somatic cells and large, immotile reproductive cells. Thus, *V. carteri* provides a unique opportunity to study multicellularity and cellular differentiation at the molecular level.

**Results:** This study provides a whole transcriptome RNA-Seq analysis of separated cell types of the multicellular green alga *V. carteri* f. *nagariensis* to reveal cell type-specific components and functions. To this end, 246 million quality filtered reads were mapped to the genome and valid expression data were obtained for 93% of the 14,247 gene loci. In the subsequent search for protein domains with assigned molecular function, we identified 9435 previously classified domains in 44% of all gene loci. Furthermore, in 43% of all gene loci we identified 15,254 domains that are involved in biological processes. All identified domains were investigated regarding cell type-specific expression. Moreover, we provide further insight into the expression pattern of previously described gene families (e.g., pherophorin, extracellular matrix metalloprotease, and *VARL* families). Our results demonstrate an extensive compartmentalization of the transcriptome between cell types: More than half of all genes show a clear difference in expression between somatic and reproductive cells.

**Conclusions:** This study constitutes the first transcriptome-wide RNA-Seq analysis of separated cell types of *V. carteri* focusing on gene expression. The high degree of differential expression indicates a strong differentiation of cell types despite the fact that *V. carteri* diverged relatively recently from its unicellular relatives. Our expression dataset and the bioinformatic analyses provide the opportunity to further investigate and understand the mechanisms of cell type-specific expression and its transcriptional regulation.

**Keywords:** Cellular differentiation, Cell types, Gene expression, Green algae, RNA sequencing, Transcript level, Whole transcriptome sequencing, Volvocales, Volvocine algae, *Volvox carteri*

\* Correspondence: armin.hallmann@gmx.de

<sup>1</sup>Department of Cellular and Developmental Biology of Plants, University of Bielefeld, Universitätsstr. 25, 33615 Bielefeld, Germany

Full list of author information is available at the end of the article



## Background

The development and radiation of clonally developing multicellular organisms with different types of cells is one of evolution's most important achievements [1–5]. Among the eukaryotes, simple multicellularity has evolved at least 25 times from unicellular ancestors, making such a development step less rare than might have been expected [1–3, 6–16]. Complex multicellularity with cell-cell adhesion, intercellular communication, and cellular differentiation has evolved ten times in eukaryotes – once in Animalia, three times in Fungi (chytrids, ascomycetes, and basidiomycetes), and six times in the three major photosynthetic clades [5], namely Phaeophyta (brown algae), Rhodophyta (red algae), and Viridiplantae (green algae and land plants). Evolution of cellular differentiation is a milestone through which two or more cell types with clear-cut identities arise from one embryonic cell accompanied by the loss of reproductive capacity in somatic cells. *Prima facie*, it is hard to understand how the waiving of reproductive capacity of many cells of an organism can be beneficial for the whole organism and, therefore, different theories about the evolution of cellular differentiation have emerged [1, 3, 4, 8, 9, 17–24].

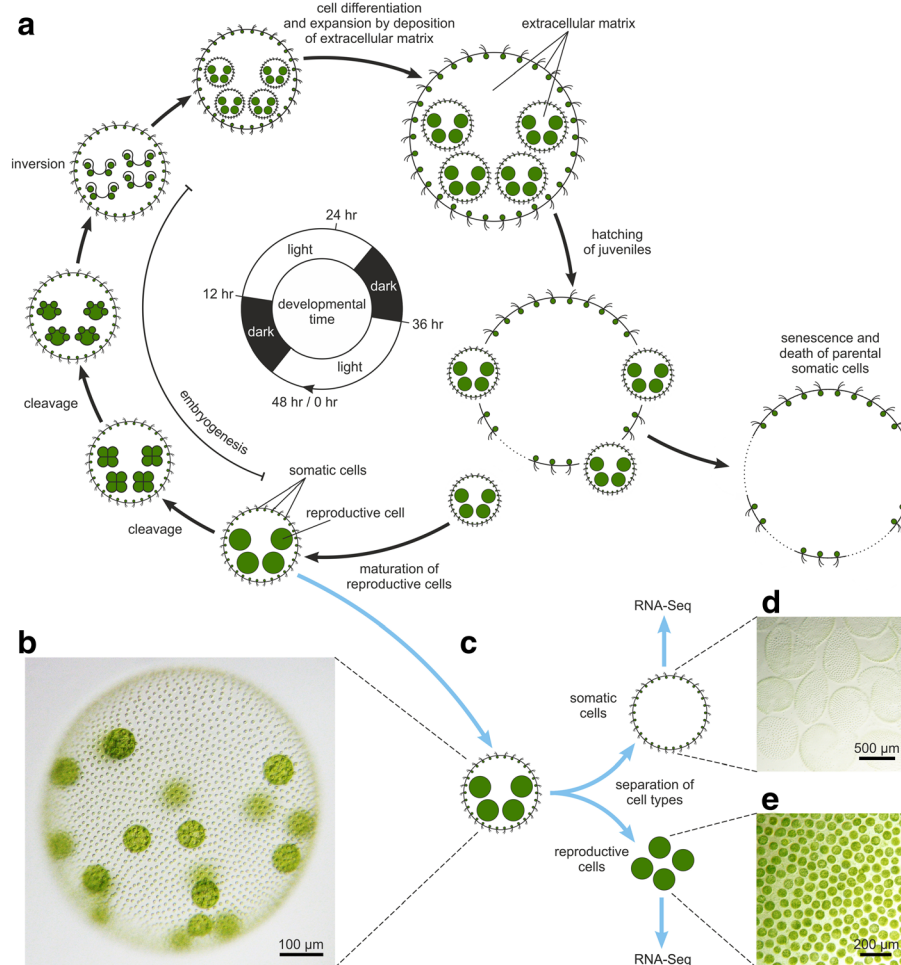
In most lineages, the investigation of aspects of multicellularity and cellular differentiation at the molecular level are challenging since the transitions occurred too long ago and organisms have evolved numerous different cell types [25]. In contrast, multicellular members of the volvocine green algae group, such as *Volvox carteri*, diverged relatively recently from their unicellular relatives [23, 25, 26], thus representing a unique opportunity to study multicellularity and cellular differentiation at the molecular level. Furthermore, *V. carteri* exhibits a complete division of labor between mortal somatic cells and immortal germ cells. Given the above and further unique properties, *V. carteri* remains one of the simplest multicellular model organisms in developmental biology [8, 27–35].

*V. carteri* is a spherically organized, mobile, obligate photoautotrophic alga of 0.5 to 2 mm in diameter, with a distinct male-female sexual dimorphism [8, 35]. In nature, it lives in freshwater ponds, puddles, and ditches, where it reproduces asexually as long as the conditions are favorable. An asexual cycle begins when each mature reproductive cell of an adult spheroid initiates a rapid series of cleavage divisions, some of which are asymmetric and produce large reproductive initials and small somatic initials (Fig. 1a). After completion of cleavage and cellular differentiation, the embryo needs to turn itself right-side out in a morphogenetic process called inversion. Following inversion, both the adult spheroid and the juvenile spheroids within it increase in size by depositing large quantities of extracellular matrix (ECM)

(Fig. 1a). Finally, the juveniles hatch out of the parenteral spheroid and the asexual cycle starts again. However, when the habitat of an asexually reproducing *Volvox* population begins to dry out, e.g., in the heat of late summer, the algae switch to sexual reproduction and produce dormant zygotes with hard cell walls that survive the drought (Additional file 1: Figure S1). As soon as favorable conditions return, the zygotes undergo meiosis, germinate, and develop into asexually reproducing males or females. In the asexual mode of reproduction, both male and female algae contain approximately 2000 small, terminally differentiated, biflagellate somatic cells embedded in the surface of a transparent sphere of glycoprotein-rich ECM. Furthermore, approximately 16 large reproductive germ cells (called gonidia) are positioned slightly below the surface of the spheroid (Fig. 1b). Each cell has a single, large cup-shaped chloroplast to conduct photosynthesis [8]. The somatic cells are specialized for motility and phototaxis, incapable of dividing, and programmed to die when only a few days old, whereas reproductive cells are immotile, specialized for growth and reproduction, and potentially immortal [8, 27–35].

Based on molecular studies, a minimal model for the genetic program of cellular differentiation into somatic and germline cells in *V. carteri* has been established [8, 27, 30, 32, 33, 35–42] (Fig. 2). The model includes four master regulatory genes, namely *glsA*, *hsp70A*, *lag*, and *regA*. After several symmetric cell divisions, *glsA* and *hsp70A* genes act to shift cell-division planes in one half of the embryo, resulting in the asymmetric divisions that set apart large-small sister-cell pairs. After cleavage divisions, cell specialization results from cell size-specific expression of the regulatory genes *lag* and *regA*, which are supposed to code for transcriptional repressors. The *lag* gene acts only in the large cells to repress the development of somatic characteristics, while the *regA* gene acts only in the small cells to repress reproductive development. After activation of either a somatic or germline program, small cells develop into biflagellate somatic cells and large cells develop into non-motile germline cells.

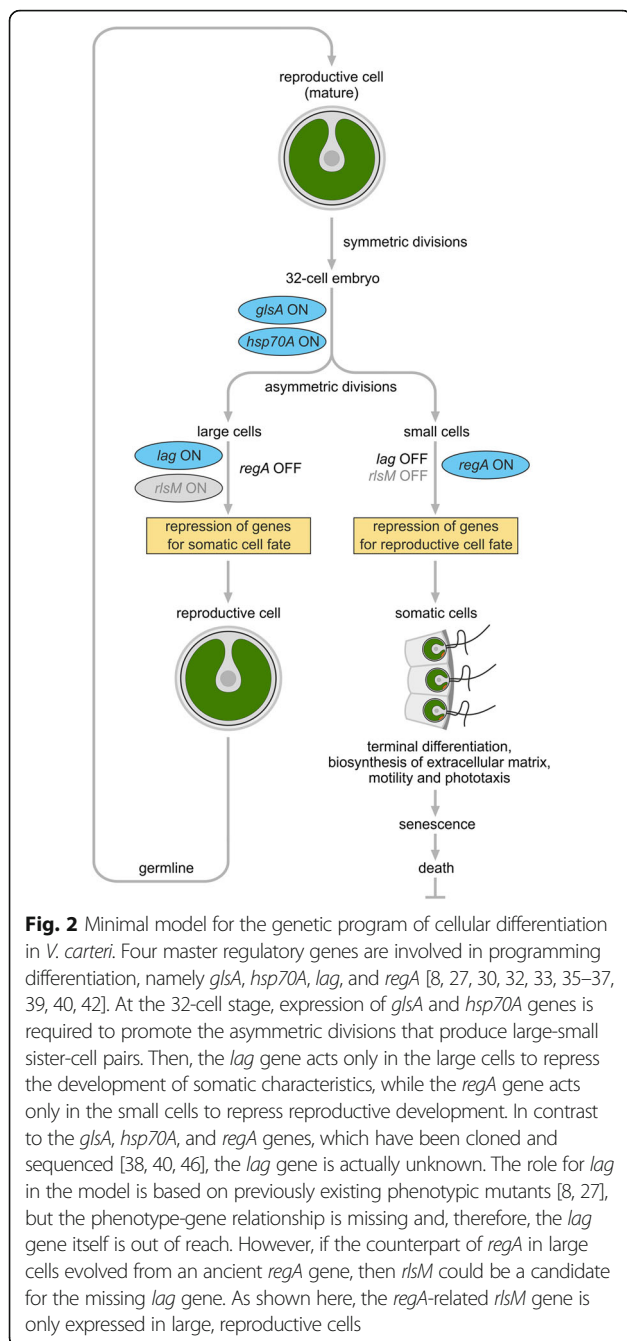
Although this minimal model is very helpful, it is only an interim outcome towards complete understanding of cellular differentiation in *V. carteri*. It remains unclear which other components are involved and how the identified master regulatory genes fit into a larger regulatory network that governs cell type-specific gene expression levels. Over 30 years ago, David and Marilyn Kirk [43] recognized that it is also necessary to identify the genes or proteins that are expressed differentially in the two cell types in order to better understand cellular differentiation. At that time, they showed that somatic and reproductive cells of *V. carteri* display substantially



**Fig. 1** Asexual development of *Volvox carteri*, wild-type phenotype and separation of cell types. **a** Asexual development of *V. carteri* [8, 35]. *Volvox* algae exist as distinct males and females. However, during asexual development the males look just like the females (for sexual development see Additional file 1: Figure S1). During embryogenesis, mature asexual reproductive cells (gonidia) undergo a rapid series of 11–12 cleavage divisions, some of which are asymmetric. The fully cleaved embryo contains all of the cells of both types that will be present in an adult but it is inside out with respect to the adult configuration. This awkward condition is quickly corrected by a gastrulation-like inversion process [144]. Then, both the adult spheroid and the juvenile spheroids within it expand by the deposition of the extracellular matrix (ECM). The juveniles eventually hatch from their parent spheroid and the somatic cells of the parent undergo senescence and die, while the reproductive cells of the juvenile spheroids mature. Under standard conditions [117, 133, 134], the asexual life-cycle takes 48 h. For clarity, each parent spheroid in this figure contains only 4 of the ~16 reproductive cells, embryos, or descendant spheroids. **b** Wild-type phenotype of an asexual female of *V. carteri* containing approximately 2000 small, terminally differentiated, biflagellate somatic cells at the surface and approximately 16 large reproductive cells in the interior. The reproductive cells are at the developmental stage just before the beginning of embryogenesis. More than 95% of the volume of such a spheroid consists of a complex but transparent ECM. **c** Mechanical separation of the cell types of three biological replicates was performed at the developmental stage just before the onset of cell cleavage of reproductive cells. The separated cell types were then used for the RNA-Seq analysis. **d** Isolated somatic cell sheets. **e** Isolated reproductive cells

different patterns of both newly synthesized and accumulated proteins [43]. However, it was not possible to obtain amino acid sequences of these proteins, so their identity remained unknown. The first cell type-specific expressed mRNAs of *V. carteri* were identified by northern blots using radiolabeled restriction-digested DNA as probes [44]. However, the investigators identified only approximately 30 different mRNA species and they did not obtain the sequence of these mRNAs. Without a

sequence, the molecular functions of these mRNAs remained unresolved. A few years later, 18 mRNAs with cell type-specific expression in reproductive cells were sequenced and functionally classified [45]. Remarkably, these mRNAs turned out to be expressed both in reproductive cells and *regA*<sup>-</sup> mutant somatic cells, but not in *regA*<sup>+</sup> wild-type somatic cells. Moreover, many of these mRNAs encoded chloroplast proteins. These findings contributed to the current model for somatic cell



differentiation (Fig. 2) involving repression of genes for reproductive development, whereby several of these genes are required for chloroplast biogenesis [45]. The *regA* gene and its gene product, which acts as key regulator in small cells (later somatic cells) to suppress reproductive development, have been identified by analyzing mutants and by Mendelian analysis [46] (Fig. 2). In a similar way, another key regulator, the *lag* protein, which acts in large cells (later reproductive cells) to repress somatic development, has been characterized [8, 47, 48] (Fig. 2).

In 2006, approximately 40 genes with quite different functions were characterized by quantitative real-time RT-PCR with respect to cell type-specific expression [49]. Even if the number of investigated genes is low, it is the largest analysis on mRNA expression of separated cell types in *Volvox* so far. Beyond that, only an additional 12 genes of *Volvox* have been analyzed in the same way [50].

Although large-scale transcriptome analyses have already been performed in *V. carteri*, they did not deal with cell type-specific mRNAs but had their own different objectives. Large-scale transcriptome analyses using expressed sequence tags were utilized to develop and confirm gene models [16] and to explore alternative splicing in *Volvox* [50]. However, these large-scale analyses could not provide any information about cell type-specific expression because the mRNA came from whole organisms. Even large-scale transcriptome analyses using RNA sequencing data and small RNA sequencing data have been generated in *Volvox* but only Argonaute 3-associated microRNAs have been analyzed for cell type-specific expression [51].

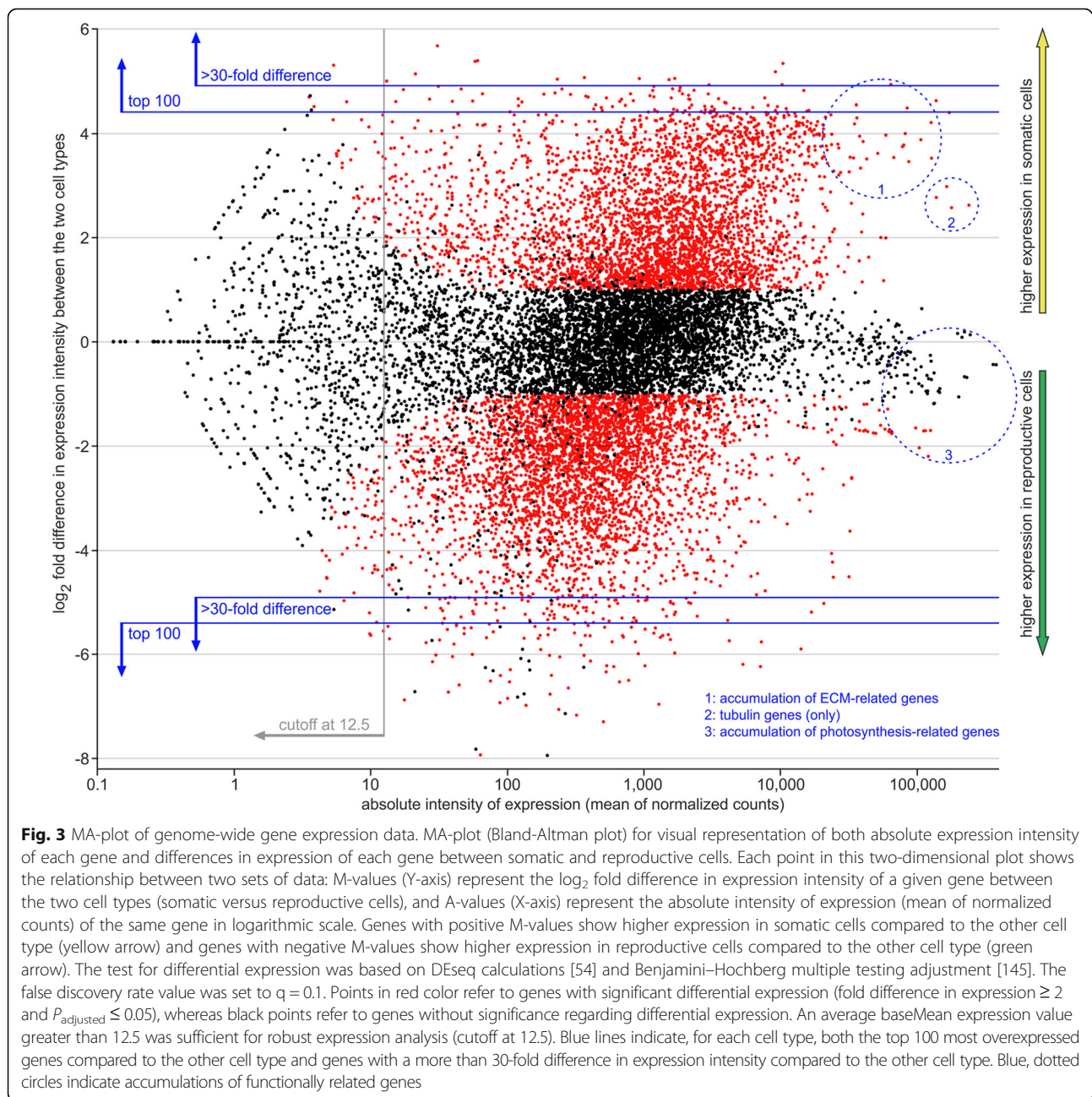
Here, we show a whole transcriptome RNA-Seq analysis of separated cell types of the multicellular alga *V. carteri* f. *nagariensis* to reveal cell type-specific mRNAs and their functions. We provide valid expression data for 93% of the 14,247 gene loci in *V. carteri*. Furthermore, all expressed genes were searched for known protein domain encoding sequences and we present which identified domains show cell type-specific expression. Since the scientific literature contains information on or at least a brief mention of approximately 400 *Volvox* genes, we look at the expression of those genes in more detail. In this connection, we also provide further insight into the expression pattern of previously described gene families, such as pherophorin, ECM metalloprotease, and *VARL* (volvocine algal *regA* like) families. Overall, our results demonstrate an extensive compartmentalization of the transcriptome between cell types, since more than half of all genes show a clear difference in expression between somatic and reproductive cells.

## Results

### RNA isolation and high throughput sequencing

The objective of our study was the generation of global gene expression profiles of somatic cells and reproductive cells of *V. carteri* separately from each other. Mechanical separation of the cell types of three biological replicates was performed at the developmental stage just before the onset of cell cleavage of reproductive cells (Fig. 1a, c; procedure see Methods); only this stage allows for separation of somatic and reproductive cells (Fig. 1d, e).





However, the determination of gene structures is limited to genes with good coverage by mapped reads and, therefore, the expression needs to exceed a certain expression threshold corresponding to a baseMean value of 450 as computed by the R package DESeq [54–56].

To gain an overview of how well our mapped RNA-Seq data, which produce the expression profiles, match the exons of the predicted *Volvox* genes of annotation v2.1 on the Phytozome V12 platform, we used the software suite BEDTools *intersect* [57] for a genome-wide examination of overlaps. The analysis showed that 87% of our mapped data had an overlap

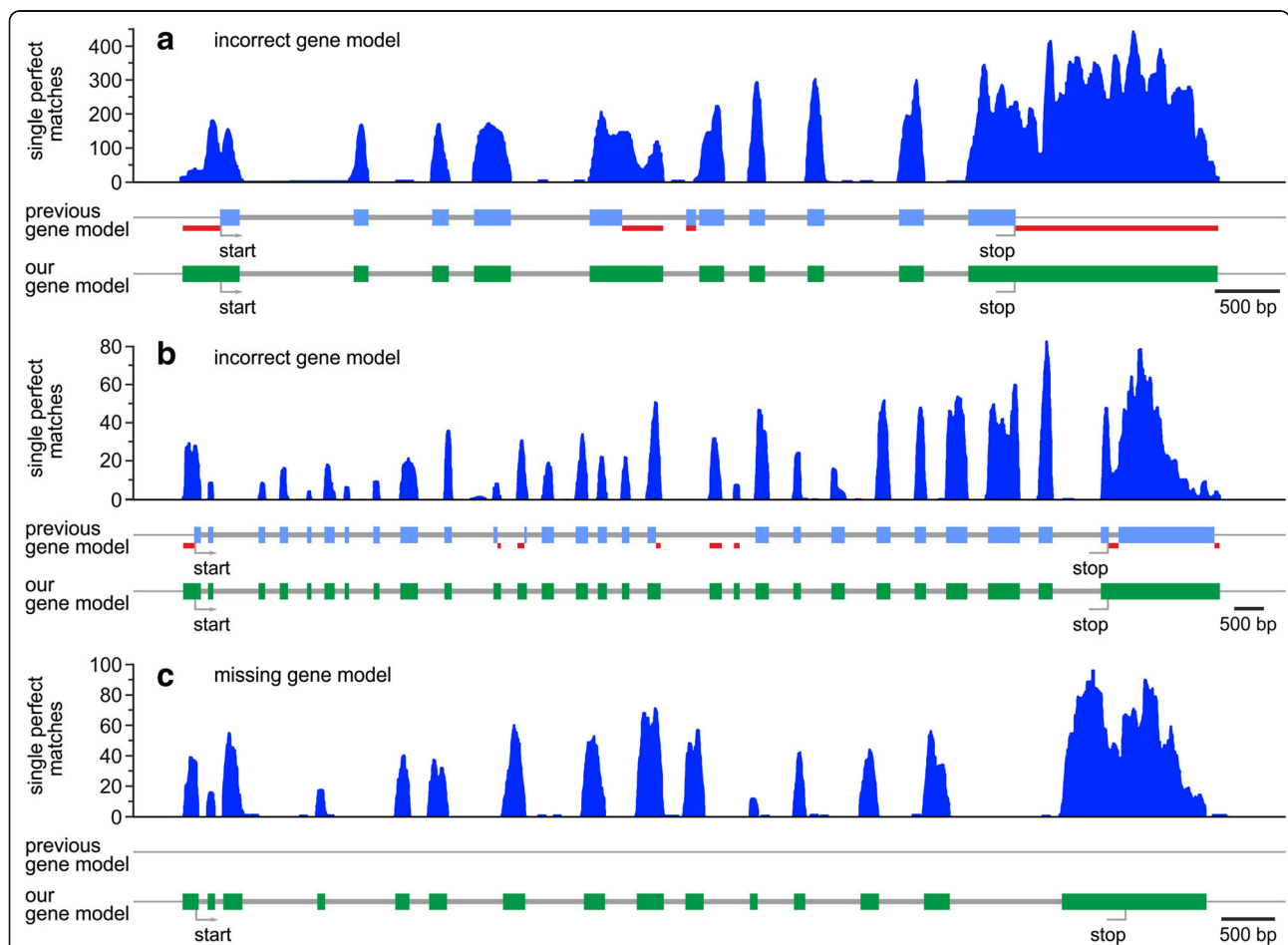
with a predicted exon. However, 13% mapped outside predicted exons, but most of these mappings were localized to the UTRs (e.g., due to a 3'-UTR that is actually longer than predicted).

For a more detailed picture of the nature of these discrepancies, we individually and very closely checked all generated expression profiles found on the first 1 million base pairs of the randomly chosen scaffold 9 (Additional file 2: Table S1) (see Methods). This section of the genome covers approximately 100 gene loci. Moreover, we performed the same analysis for 100 randomly chosen gene loci from the complete list of

previously known gene loci (Additional file 3: Table S2). Both examinations on a random basis revealed that approximately one fifth (21% and 17%, respectively) of the sufficiently expressed genes showed discrepancies within the coding sequence. Examples of previous incorrect gene predictions that affect the coding sequence and, thus, have strong impact on the deduced amino acid sequences, are shown in Fig. 4a, b. In addition, more than two-fifths (45% and 42%, respectively) of the critically inspected gene loci showed discrepancies within the UTRs, which in reality are frequently longer than predicted. We also identified genes with

clear expression profiles that were not predicted at the corresponding genomic position according to gene annotation v2.1 of the *V. carteri* genome available on the *Volvox* pages of the Phytozome V12 platform (Fig. 4c).

We cannot extrapolate the results of these sample analyses to the whole genome, but we can say that, apparently, the prediction algorithms used in the previous computer-based genome-wide analysis did not produce perfect results. Our RNA-Seq data can serve as a reliable basis for manual verification of gene predictions.



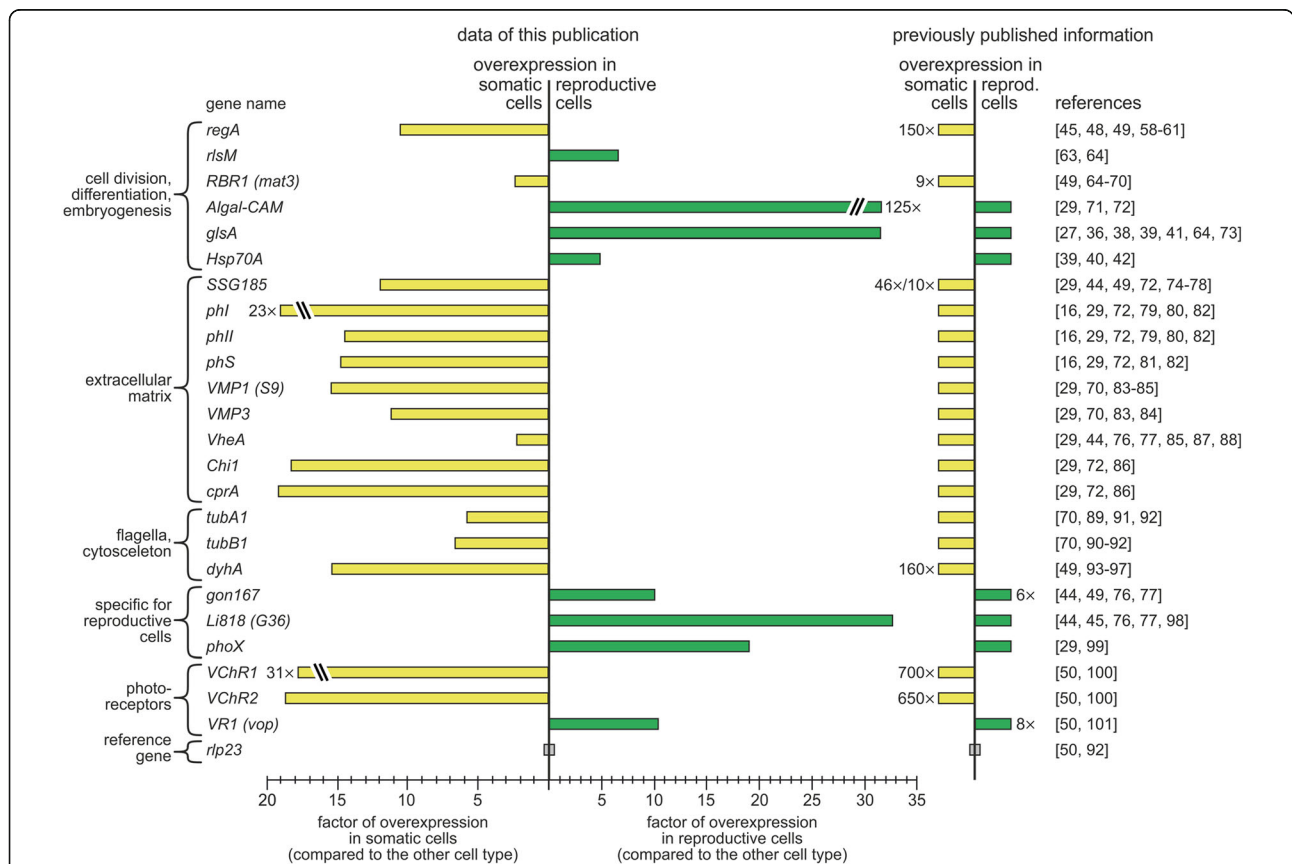
**Fig. 4** Examples of expression profiles of genes with previously absent or incorrect gene models. The previous gene models originate from gene annotation v2.1 of the *Volvox carteri* genome available on the *Volvox* pages of the Phytozome V12 platform. **a** Example of a gene with an incorrectly predicted gene model: the gene model of gene Vocar.0007s0316 is not in accordance with the expression profiles at exon 5 (prediction too short), exon 6 (non-existent in reality), and the 5'- and 3'-UTRs (no prediction). **b** Example of another gene with an incorrectly predicted gene model: the gene model of gene Vocar.0001s0415 is not in accordance with the expression profiles at exons 11, 12, and 17 (prediction too short). There are also inconsistencies on the 5'-UTR (no prediction) and on the 3'-UTR (prediction somewhat too short). Moreover, there are in fact two additional exons between exons 17 and 18 (no prediction), and there is definitely no intron within the 3'-UTR. **c** Example of a gene with previously absent gene model: the gene is located on scaffold 4 at position 1,393,429 to 1,402,937. It codes for a protein with 837 amino acid residues. In all panels (a–c), the previous exon-intron predictions for these genes according to annotation v2.1 of the *V. carteri* genome are given directly below the expression profiles; exons are shown as blue bars and introns as thick gray lines. Thick red lines indicate differences between the previous prediction according to annotation v2.1 and our prediction. Our own exon-intron predictions for these genes are shown at the bottom; exons are illustrated as green bars and introns as thick gray lines. Our predictions are supported by the obtained expression profiles (single perfect matches) shown in dark blue

**Gene expression of previously investigated *Volvox* genes**

To provide new information about the expression of previously investigated *Volvox* genes and to validate our RNA-Seq dataset, we composed a list of all available *Volvox* genes that were at least briefly mentioned in the literature. All 376 *Volvox* genes of this list were investigated regarding cell type-specific gene expression and the results were compared to previous expression data if available (Additional file 4: Table S3). For clarity reasons, only the best-researched genes were collated and arranged by gene function in Fig. 5. The comparison illustrates that our results were qualitatively well fitted to previous results. It should be noted that the comparison needs to be performed qualitatively, because previous expression data were either obtained qualitatively anyway or were quantified with a different experimental approach and/or developmental stage. It is noticeable that especially genes that previously had a very high fold difference in expression,

showed a more moderate fold difference in expression in our analysis.

Particularly, we checked several key genes in cell division, differentiation, and embryogenesis. The genes *regA* [42, 45, 46, 49, 58–64], *RBR1 (mat3)* [49, 64–70], *Algal-CAM* [29, 71, 72], *glsA* [27, 36, 38, 39, 41, 64, 73], and *Hsp70A* [39, 40, 42] of this group showed the expected cell type-specific expression (Additional file 4: Table S3, Fig. 5). The *regA* gene is one of the best-studied genes in *Volvox*, coding for a transcriptional repressor [42, 45, 46, 49, 58–64]. It is important to mention that the expression analysis of the counterpart of *regA*, the *lag* gene (Fig. 2), is missing because it actually has never been sequenced or assigned to a specific gene locus, although it is frequently referenced [8, 27, 30, 32, 33, 35–37, 42]. The *regA* gene is a member of the VARL (volvocine algal *regA* like) gene family due to a common VARL



**Fig. 5** Cell type-specific gene expression of previously investigated *Volvox* genes. The figure shows expression data and references of the best-researched *Volvox* genes, which are arranged by gene function. These data were extracted from Additional file 4: Table S3, which contains a much more extensive table with expression data, further information, and references of 376 *Volvox* genes that are at least briefly mentioned in the literature. The length of the expression bar illustrates the fold higher expression of a given gene within the given cell type with respect to the other cell type. Yellow: higher expression in somatic cells; green: higher expression in reproductive cells. The previously published information about expression of these genes is presented in a qualitative, 'digital' format because they were either obtained qualitatively anyway or were quantified with a different experimental approach and/or developmental stage, which make comparisons rather difficult. Nonetheless, the numerical value of the previously determined fold difference in expression is indicated, if available



domain that includes a DNA-binding SAND domain [63]. The *VARL* gene family includes 14 members in *V. carteri*, which have not been previously subject to an expression analysis [63, 64]. Here, we show that 10 *VARL*-genes including *regA* are overexpressed in somatic cells, three *VARL*-genes show no clear cell type-specific expression, and a single gene, *rlsM*, is overexpressed in reproductive cells (Additional file 4: Table S3, Fig. 5). Due to its overexpression in reproductive cells, the *rlsM* gene is of particular interest because it could correspond to the missing *lag* gene. Like *regA*, *lag* was suggested to be a transcriptional repressor with a DNA-binding domain in the opposite cell type and both may have descended from the same ancestral gene [8, 27, 30, 32, 33, 35–37, 42]; *rlsM* fits these conditions.

Several genes of ECM (glyco)proteins have been previously investigated, including the pherophorins (e.g., *SSG185*, *phI*, *phII*, *phS*) [16, 29, 44, 49, 72, 74–82], ECM metalloproteases (e.g., *VMP1*, *VMP3*) [29, 70, 83–85], and other ECM enzymes (e.g., *VheA*, *Chi1*, *cprA*) [29, 44, 72, 76, 77, 85–88]. Due to previous experimental results and since ECM biosynthesis has been previously attributed to somatic cells only [29, 72, 79–82], all these genes were expected to be overexpressed in somatic cells [29], as indeed shown herein (Additional file 4: Table S3, Fig. 5). Regarding the pherophorins, it should be noted that we not only investigated the expression pattern of previously characterized pherophorins but also that of many others. Among the latter, we surprisingly identified pherophorins that are clearly overexpressed in reproductive cells (see below).

Tubulins (e.g., *tubA1*, *tubB1*) [70, 89–92] and dyneins (e.g., *dyhA*) [49, 93–97] are important components of flagella. Because only somatic cells have flagella, the genes of these proteins are expected to be expressed predominantly in somatic cells, as shown herein (Additional file 4: Table S3, Fig. 5). Likewise, we confirmed that genes that code for proteins known to be specific for reproductive cells (e.g., *gon167*, *Li818*, *phoX*) [29, 44, 45, 49, 76, 77, 98, 99] show overexpression in reproductive cells.

Photoreceptors (e.g., *VChR1*, *VChR2*) [50, 100] are known to be expressed predominantly in somatic cells, except for two weakly expressed photoreceptors that showed no cell type-specific expression and one, *VRI* [50, 101], which is overexpressed in reproductive cells. Our expression analysis is in accordance with these expectations, except for a lack of confirmation of the previous extremely high fold difference in expression of *VChR1* and *VChR2* (Additional file 4: Table S3, Fig. 5).

The gene *rlp23* [50, 92], which codes for a structural component of the ribosome, was suggested as a reference gene in expression analysis of different cell types (e.g., with real time qRT-PCR). Here, we show that *rlp23* is uniformly expressed in both cell types (Additional file

4: Table S3, Fig. 5), which confirms its suitability as a reference gene when target genes are examined for cell type-specific expression.

#### Identification of the most highly expressed genes

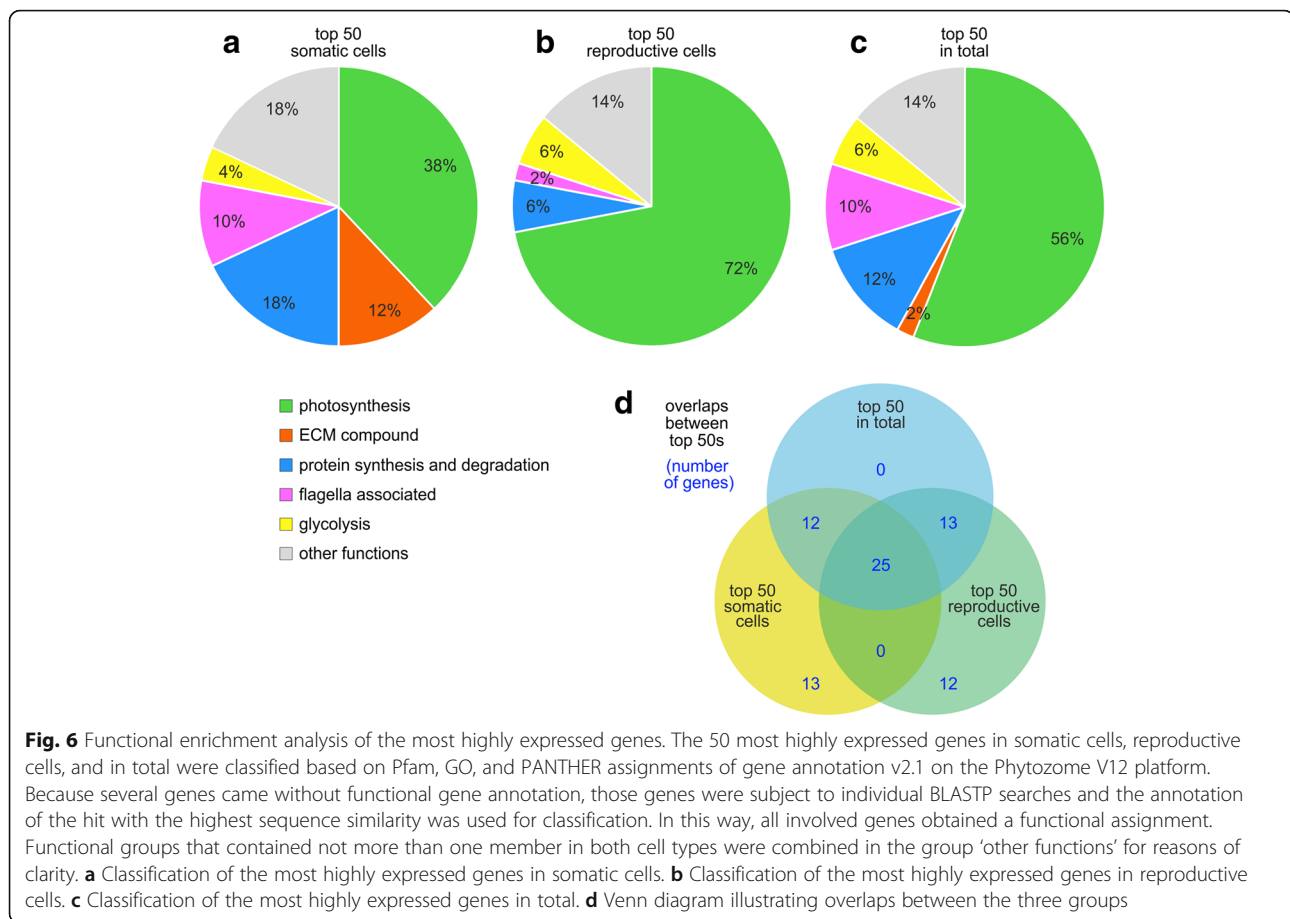
To identify the most highly expressed genes in *V. carteri*, the RNA-Seq raw data were normalized both to the total read count per sample and to the transcript length [54]. The 50 most highly expressed genes in somatic cells, reproductive cells, and in total were subjected to a functional classification based on the Pfam [102–104], PANTHER [105–107], and GO [108, 109] assignments of gene annotation v2.1 of the *V. carteri* genome on the Phytozome V12 platform [53]. Genes without functional gene annotation were subject to individual BLASTP searches [110–112] and the annotation of the hit with the highest sequence similarity was used for classification. A functional enrichment analysis of the most highly expressed genes in each cell type and in total is shown in Fig. 6 and Additional file 5: Table S4.

In somatic cells, 19 out of the 50 most highly expressed genes (38%) were shown to encode photosynthesis-related proteins (Fig. 6a and Additional file 5: Table S4); 9 (18%) genes to encode proteins involved in protein synthesis and degradation, 6 (12%) ECM compounds, 5 (10%) flagella-related proteins, and 2 (4%) proteins of the glycolysis pathway. The remaining 9 (18%) genes were scattered across quite different functional groups.

In reproductive cells, 36 out of the 50 most highly expressed genes (72%) were shown to encode photosynthesis-related proteins, but none were shown to encode an ECM compound (Fig. 6b and Additional file 5: Table S4); 3 (6%) genes were shown to encode proteins involved in protein synthesis and degradation, 1 (2%) a flagella-related protein, and 3 (6%) proteins of the glycolysis pathway. The remaining 7 (14%) genes were scattered across quite different functional groups.

We also determined the most highly expressed genes in total, i.e., without taking the cell type into account (Fig. 6c and Additional file 5: Table S4). Also in this approach, genes encoding photosynthesis-related proteins formed the largest group (56%), followed by genes that encode proteins involved in protein synthesis and degradation (12%) and genes encoding flagella-related proteins (10%).

Overall, genes encoding photosynthesis-related proteins dominated all three groups of the top 50 most highly expressed genes. In somatic cells, several highly expressed genes that encode ECM compounds stood out, whereas no genes encoding ECM compounds were among the top 50 of reproductive cells. Genes encoding flagella-related proteins and proteins involved in protein synthesis and



degradation were also more represented among the top 50 genes in somatic cells.

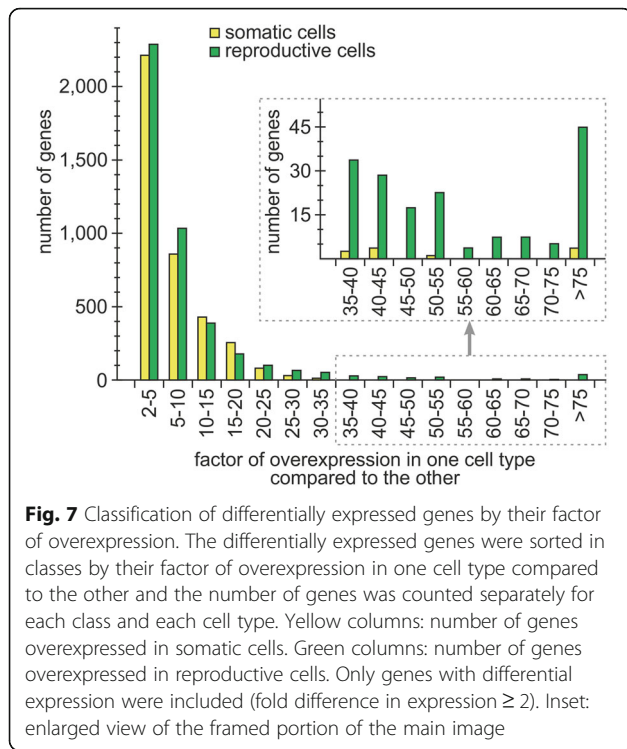
It is noteworthy that 25 genes belong to all the three groups of the top 50 most highly expressed genes. An overview of all intersections of the three groups is shown in Fig. 6d and all involved genes are mentioned by name in Additional file 5: Table S4.

#### Detection of differentially expressed genes between somatic and reproductive cells

The MA-plot in Fig. 3 provides an overview of the genome-wide gene expression differences between somatic and reproductive cells and highlights the differentially expressed genes by using red dots. It is remarkable that more than half of all predicted genes (54%) showed a clear difference in expression between somatic cells and reproductive cells. The total number of 7691 genes with significant cell type-specific overexpression (and a baseMean expression value of at least 12.5) was split more or less evenly between 3728 genes overexpressed in somatic cells and 3963 genes overexpressed in reproductive cells (ratio 48.5:51.5). A further 129 genes showed cell type-specific overexpression (36 somatic, 93

reproductive) but did not reach the baseMean expression limit of 12.5.

It is notable that the point cloud above the horizontal zero line did not have a mirror-image relationship with the point cloud below the zero line (Fig. 3). Obviously, there were more genes with a high factor of overexpression compared to the other cell type in reproductive cells than in somatic cells. To illustrate this effect more clearly, we identified all genes with more than 30-fold difference in expression compared to the other cell type (and a baseMean expression value of at least 12.5) (Fig. 3). Overall, 193 genes fulfilled these requirements, whereby 175 genes showed such a high factor of overexpression in reproductive cells whereas only 18 did so in somatic cells (ratio 90.7:9.3). Moreover, we identified the 100 most overexpressed genes of each cell type (Fig. 3) and determined the factor of overexpression of each of these genes. On average, the 100 most overexpressed genes in somatic cells showed 29-fold overexpression compared to the other cell type. However, the average value of the 100 most overexpressed genes in reproductive cells was much higher, showing 85-fold overexpression compared to the other cell type. To find the underlying cause, all genes with cell type-specific



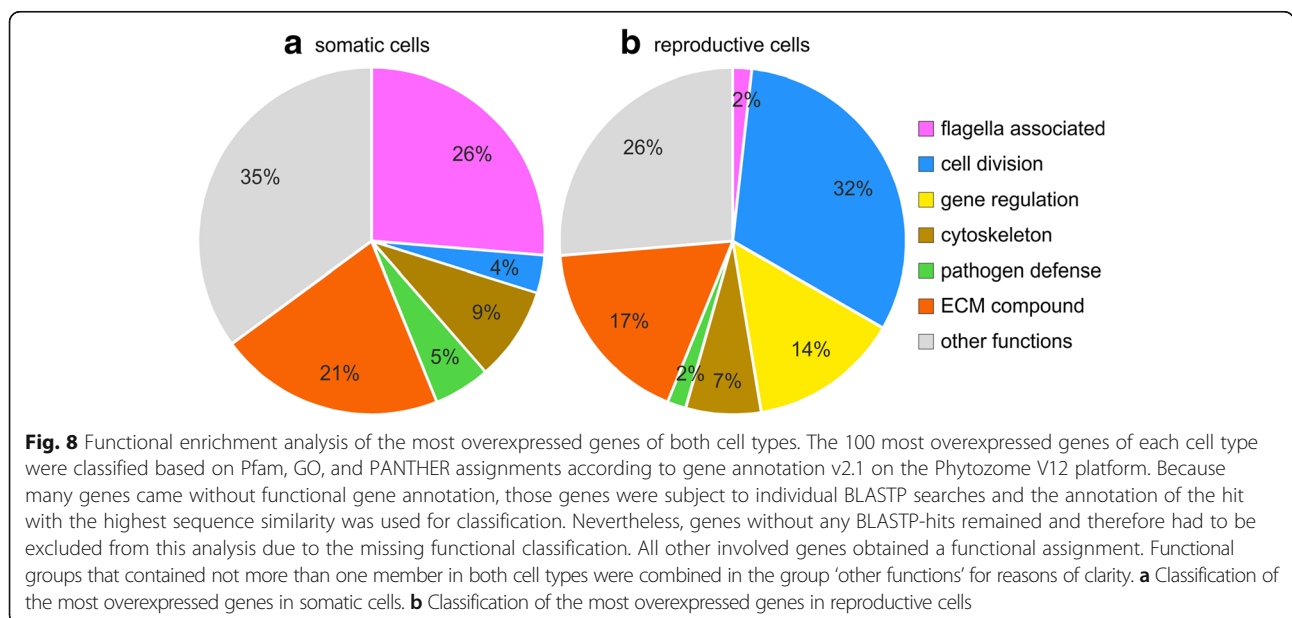
overexpression were sorted in classes by their factor of overexpression in one cell type compared to the other and the number of genes was counted separately for each class and each cell type (Fig. 7). From a more general perspective, the number of genes per class decreased with increasing factors of overexpression, which was to be expected. However, the number of genes with a high factor of overexpression was different between cell types.

For the classes containing 20- to 25-fold overexpression and higher, a similar result was always obtained, wherein a greater number of genes had higher factors of overexpression in reproductive cells than in somatic cells. Actually, there were hardly any genes that exhibited a 35-fold or higher expression in somatic cells compared to reproductive cells (Fig. 7). In the classes containing 15- to 20-fold overexpression or less, there was a more balanced situation between the number of contributing genes from each cell type (Fig. 7).

**Functional enrichment analysis of the most differentially expressed genes**

To get an idea of the function of the most differentially expressed genes in *V. carteri*, the 100 most overexpressed genes of each cell type were subject to a functional classification based on the Pfam [102–104], PANTHER [105–107], and GO [108, 109] assignments according to gene annotation v2.1 of the *V. carteri* genome on the Phytozome V12 platform. Because several genes came without functional gene annotation, those genes were subject to individual BLASTP searches [110–112] and the annotation of the hit with the highest sequence similarity was used for classification. Nevertheless, genes without any BLASTP-hits also remained and therefore had to be excluded from this analysis due to the missing possibility of functional classification. The results of the functional enrichment analysis of the most overexpressed genes of each cell type are shown in Fig. 8 and Additional file 6: Table S5.

The biggest functional group within the most overexpressed genes in somatic cells (compared to reproductive cells) were genes coding for flagella associated

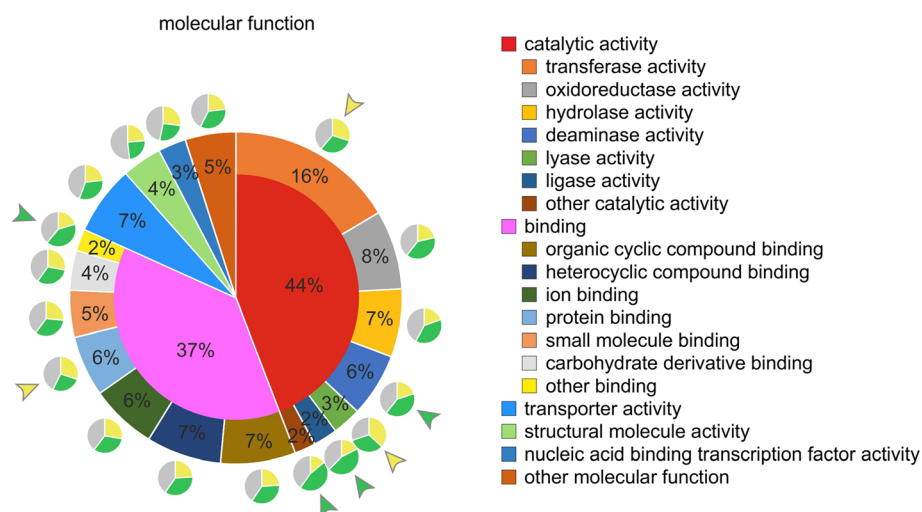


proteins (26%), followed by genes coding for ECM compounds (21%) and genes coding for components of the cytoskeleton (9%) (Fig. 8a). The biggest functional group within the most overexpressed genes in reproductive cells (compared to somatic cells) were genes related to cell division (32%) followed by genes coding for ECM compounds (17%), genes involved in gene regulation (14%), and genes coding for components of the cytoskeleton (7%) (Fig. 8b).

Remarkably, two major groups within the most overexpressed genes in reproductive cells, namely cell division and gene regulation, were only poorly (4%) or not at all represented in somatic cells (Fig. 8). In somatic cells, genes coding for flagella-associated proteins stood out, yet these were poorly (2%) represented within the most overexpressed genes in reproductive cells. It was expected that genes coding for ECM compounds would belong to the most overexpressed genes in somatic cells because it was assumed that somatic cells are solely or at least largely responsible for the biosynthesis of the extensive ECM [29, 72, 113]. However, genes coding for ECM compounds also represented a large proportion (21%) within the most overexpressed genes in reproductive cells (Fig. 8).

### Classification of all protein domains and screening for cell type-specific expression

Initially, all *Volvox* genes were screened for assigned Pfam, GO, and PANTHER identifiers regarding their molecular function and the identifiers were assigned to higher level GO-terms. A total of 6216 genes had at least one protein domain with an assigned molecular function and the total number of identified protein domains with assigned molecular function was 9435. The identified protein domains were sorted into groups and subgroups using QuickGO [114, 115]. The percentage share of each group and subgroup within the total number of protein domains with assigned molecular functions was determined (Fig. 9). Each group or subgroup was analyzed for the proportion of protein domains with overexpression (fold difference in expression  $\geq 2$ ) in somatic cells or reproductive cells, or without distinct differences in expression between the two cell types (Fig. 9). As a reference, we determined the cell type-specific proportions for all domains in 'molecular function' as a whole, which was 25.0% overexpressed in somatic cells, 34.2% overexpressed in reproductive cells, and 40.8% without overexpression. Groups and subgroups that differed clearly from this distribution of the total quantity (deviation  $\geq 20\%$ ) are highlighted in Fig. 9. For example, protein domains with transferase activity, lyase activity, and protein



**Fig. 9** Protein domain classification by molecular function followed by analysis for cell type-specific distribution of expression. All genes were screened for assigned Pfam, GO, and PANTHER identifiers regarding their molecular function and the identifiers were assigned to higher level GO-terms. The identified protein domains were sorted into groups and subgroups using QuickGO. The percentage share of each group and subgroup within the total number of protein domains with assigned functions was determined. Small groups or subgroups with a percentage share of less than 2% were combined for reasons of clarity (see 'other...'). The very large groups 'catalytic activity' (44%) and 'binding' (37%) are shown with their subgroups. Each group or subgroup was analyzed for the proportion of protein domains with overexpression (fold difference in expression  $\geq 2$ ) in somatic cells or reproductive cells, or without distinct differences in expression between the two cell types. Small pie charts show the results. Yellow color: share of protein domains that are overexpressed in somatic cells; green color: share of protein domains that are overexpressed in reproductive cells; gray color: share of protein domains without cell type-specific overexpression. Groups and subgroups that differ clearly from the distribution of the total quantity (deviation  $\geq 20\%$ ) are highlighted by colored arrowheads. Yellow arrowhead: larger proportion of domains with overexpression in somatic cells. Green arrowhead: larger proportion of domains with overexpression in reproductive cells

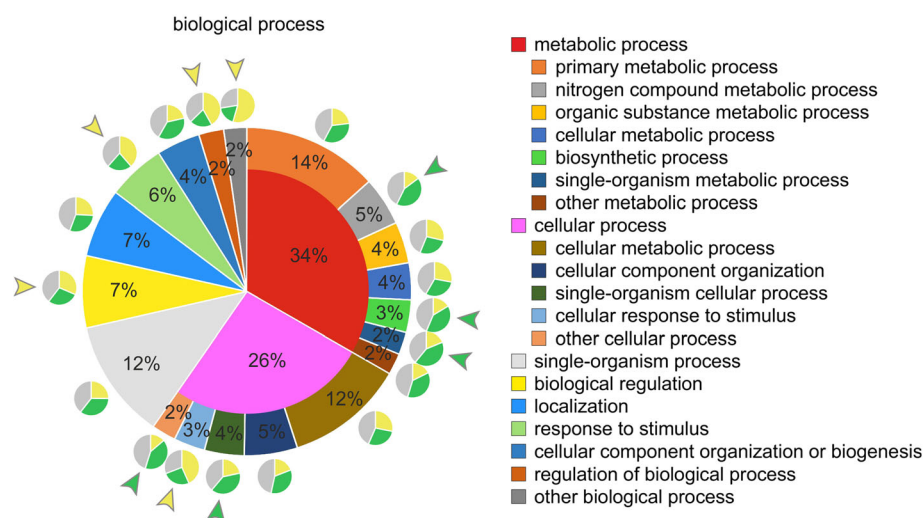
binding showed a larger proportion of domains with overexpression in somatic cells (Fig. 9). Whereas protein domains with deaminase activity and ligase activity showed a larger proportion of domains with overexpression in reproductive cells (Fig. 9).

Apart from the classification by molecular function, all *Volvox* genes were also screened for assigned identifiers regarding their participation in biological processes and the identifiers were again assigned to higher level GO-terms. A total of 6089 genes had at least one protein domain with an assigned biological process and the total number of identified protein domains with assigned biological process was 15,254. The identified protein domains were sorted into groups and subgroups and analyzed as described above. As a reference, we determined the cell type-specific proportions for all domains in 'biological process' as a whole, which was 26.4% overexpressed in somatic cells, 32.4% overexpressed in reproductive cells, and 41.2% without overexpression. Groups and subgroups that differed clearly from this distribution of the total quantity (deviation  $\geq 20\%$ ) are highlighted in Fig. 10. For example, protein domains involved in responses to stimuli and regulation of biological processes showed a larger proportion with overexpression in somatic cells (Fig. 10). Whereas protein domains involved in nitrogen compound metabolic

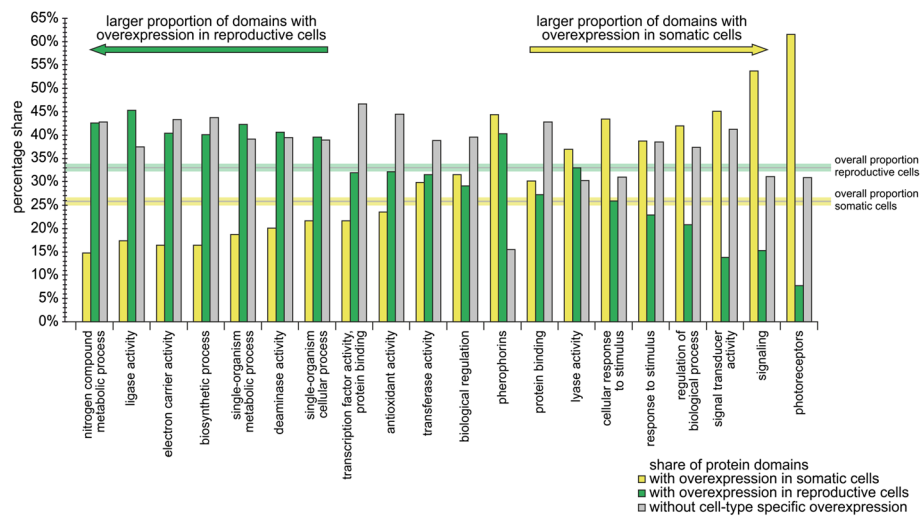
processes, biosynthetic processes, single-organism metabolic processes, and single-organism cellular processes showed a larger proportion with overexpression in reproductive cells (Fig. 10).

Overall, 24,689 protein domains were assigned to the 14,247 genes, which corresponds to an average of approximately 1.7 domains per gene. However, the real number is likely to be higher than 1.7 because new and undescribed domains, as well as domains with significant evolutionary changes, remain undetected.

To more clearly show the differences in the composition of groups or subgroups of protein domains regarding cell type-specific overexpression, we picked out 20 groups or subgroups and arranged them according to their composition. In Fig. 11, these groups or subgroups are sorted by the percentage ratio of protein domains with overexpression in somatic cells to protein domains with overexpression in reproductive cells. The clearest cell type-specific differences can be found in groups or subgroups that contain protein domains with very specific functions. Noticeable is the predominance of photoreceptor domains in somatic cells and the over-weighting of domains with ligase activity in reproductive cells (Fig. 11). Particularly remarkable is also the absence of a clear cell type-specific imbalance in a group that contains domains of well-known ECM proteins, the



**Fig. 10** Protein domain classification by biological process followed by analysis for cell type-specific distribution of expression. All genes were screened for assigned Pfam, GO, and PANTHER identifiers with regard to their participation in biological processes and the identifiers were assigned to higher level GO-terms. The identified protein domains were sorted into groups and subgroups using QuickGO. The percentage share of each group and subgroup within the total number of protein domains with assigned functions was determined. Small groups or subgroups with a percentage share of less than 2% were combined for reasons of clarity (see 'other...'). The very large groups 'metabolic processes' (34%) and 'cellular process' (26%) are shown with their subgroups. Each group or subgroup was analyzed for the proportion of protein domains with overexpression (fold difference in expression  $\geq 2$ ) in somatic cells or reproductive cells, or without distinct differences in expression between the two cell types. Small pie charts show the results. Yellow color: share of protein domains that are overexpressed in somatic cells; green color: share of protein domains that are overexpressed in reproductive cells; gray color: share of protein domains without cell type-specific overexpression. Groups and subgroups that differ clearly from the distribution of the total quantity (deviation  $\geq 20\%$ ) are highlighted by colored arrowheads. Yellow arrowhead: larger proportion of domains with overexpression in somatic cells. Green arrowhead: larger proportion of domains with overexpression in reproductive cells



**Fig. 11** Proportions of cell type-specifically expressed domains in selected groups or subgroups of protein domains. As described in the legends of Figs. 9 and 10, protein domains were sorted into groups and subgroups and each group or subgroup was analyzed for the proportion of protein domains with overexpression (fold difference in expression  $\geq 2$ ) in somatic cells or reproductive cells, or without distinct differences in expression between the two cell types. Here, shares of cell type-specific expressed domains in selected groups or subgroups of protein domains are shown in greater detail. The groups are sorted from left to right by the percentage ratio of protein domains with overexpression in somatic cells to protein domains with overexpression in reproductive cells. For comparison, we also determined the proportions for all domains as a whole, which is 25.8% overexpressed in somatic cells (yellow horizontal line), 33.1% overexpressed in reproductive cells (green horizontal line), and 41.1% without overexpression. The listed groups/subgroups come from the classifications by both molecular function (Fig. 9) and biological process (Fig. 10). Some of the subgroups shown here are subgroups of the groups/subgroups shown in Figs. 9 and 10

pherophorins (Fig. 11). Previously, pherophorins have only been associated with ECM biosynthesis in somatic cells [29, 72, 79–82]. The issue is discussed below.

## Discussion

The focus of this study was a whole transcriptome RNA-Seq analysis of mechanically separated cell types of *V. carteri* f. *nagariensis* (Fig. 1) to disclose cell type-specific components and functions. After quality filtration, 246 million reads were mapped to the genome and 13,204 genes showed adequate coverage for quantitative analysis of expression. This study thus provides valid expression data for 93% of the total 14,247 *V. carteri* gene loci (Fig. 3). Moreover, our RNA-Seq data can serve as a reliable basis for manual verification of gene predictions (Fig. 4).

There is no doubt that mRNA expression and its regulation are essential for key developmental events such as cellular differentiation. However, an observed expression level of a particular gene can be cause or effect of a cellular condition or phenotype. Expression comparisons alone cannot distinguish between these possibilities. In such a situation, the identification and detailed characterization of key genes can help to unravel cause-and-effect networks. A detailed characterization of selected genes nevertheless involves expression analyses. In this respect, the analysis of cell type-specific gene expression of many previously investigated *Volvox* genes enabled us to provide new information about these genes (Additional file 4:

Table S3, Fig. 5). The most thoroughly investigated genes with regard to cellular differentiation are *glsA*, *hsp70A*, *lag*, and *regA* [8, 27, 30, 32, 33, 35–42, 45, 46, 49, 58–64, 73] (Fig. 2). Both *glsA* and *hsp70A* are known to be expressed maximally in asymmetrically dividing embryos to shift cell-division planes but there is already a significant overexpression in reproductive cells before the onset of cell cleavage [27, 36, 38–42, 64, 73]. Our analysis clearly confirms this overexpression in reproductive cells prior to the onset of cell cleavage (Fig. 5). Based on the model for the genetic program of cellular differentiation (Fig. 2), cell specialization results from cell size-specific expression of the regulatory genes *lag* and *regA*. Therefore, these genes were suggested as key components added to the genome during evolution to make possible the conversion of the ancestral, sequential form of cyto-differentiation into the dichotomous form that characterizes *Volvox* [36]. The *regA* gene is expressed only in small cells to repress reproductive development and, thus, to produce somatic cells [42, 45, 46, 49, 58–64]. Accordingly, we observed a strong overexpression of *regA* in somatic cells (Fig. 5). The RegA protein was classified as a transcriptional repressor belonging to the VARL family [63, 64], which has 14 members in *V. carteri*. Our analysis provides the first expression data of the complete family. One of these VARL genes, *rlsM*, is particularly interesting because it is overexpressed in reproductive cells and, therefore, could correspond to the previously undiscovered *lag* gene [8, 27,

30, 32, 33, 35–37, 42]. The Lag protein has been presented and frequently referenced as a transcriptional repressor that acts in reproductive cells to prevent somatic development [8, 27, 30, 32, 33, 35–37, 42]; thus, Lag represents the counterpart of RegA. Previous work on Lag mainly deals with a characteristic mutant phenotype [8, 27, 37, 116] in which presumptive reproductive cells temporarily develop into larger-than-normal somatic cells with long flagella and large eyespots [8], similar to the RegA phenotype, in which presumptive somatic cells develop into reproductive cells [46]. Thus, the *rlsM* gene now appears as the most obvious candidate to be the undiscovered key gene that was previously named *lag*.

Two other key players in cell division and embryogenesis are *RBR1* (*mat3*) and *Algal-CAM*. *RBR1* is known to be involved in cell-size control of somatic cells [65]. Here, we confirm the overexpression of this gene in somatic cells (Fig. 5), even if the maximum expression of this gene is not expected prior to the onset of cell cleavage but later in embryogenesis [65]. Similarly, we confirm the overexpression of *Algal-CAM* in reproductive cells (Fig. 5). *Algal-CAM* is a cell adhesion molecule required in early embryogenesis [71].

Our actual results regarding cell type-specific expression of previously investigated genes coding for ECM proteins, flagella components, reproductive-cell specific proteins and photoreceptors are all in accordance with the expected results (Additional file 4: Table S3, Fig. 5). However, with regard to a large family of ECM proteins, the pterophorins, we also investigated new members that showed a cell type-specific expression behavior that was opposite to that of previously investigated pterophorins. This issue is further discussed below with regards to the most overexpressed genes.

The results regarding the most highly expressed genes are quite plausible. Since the obligate photoautotrophic *V. carteri* alga uses sunlight alone as the primary energy source, all required energy is provided by converting light energy into chemical energy by photosynthesis, which is conducted in the big chloroplast of each cell. As a consequence, it is reasonable that photosynthesis-related genes dominate the group of the most highly expressed genes in both cell types (Fig. 6a–c). A reproductive cell in the stage just prior to the onset of cell cleavage (Fig. 1a) requires sufficient energy for cell growth and for the forthcoming cleavage divisions. It should also be noted that, in *V. carteri*, reproductive cells grow far larger than somatic cells because a new juvenile is produced in a series of very rapid, synchronous cleavage divisions [8, 117]. The reproductive cell is focused not only on the synthesis of proteins that are required for photosynthesis but also of those for energy conversion by glycolysis, as was apparent in the group of the most highly expressed genes in this cell type (Fig. 6b).

In somatic cells, photosynthesis and glycolysis-related genes were also among the most highly expressed genes (Fig. 6a). Even if somatic cells do not grow and divide, they have other energy-intense functions that are not relevant in reproductive cells, requiring sufficient energy to continuously build and restructure the surrounding ECM and to produce and operate the flagella. Fittingly, we also identified ECM- and flagella-related genes among the most highly expressed genes (Fig. 6a).

The examination of differential gene expression revealed that more than half of all genes showed a clear difference in expression between somatic and reproductive cells. More specifically, 7691 of 14,203 genes (54%) exhibited a fold difference in expression of 2 or more and a significance value of less than 0.05, and are therefore considered to be differentially expressed; this large proportion demonstrates an extensive compartmentalization of the transcriptome between cell types. Even though we analyzed the developmental stage prior to the onset of divisions, the group of the most overexpressed genes in reproductive cells (compared to somatic cells) contained large shares of genes related to cell division and gene regulation, respectively (Fig. 8b). Shortly before the beginning of cell divisions, reproductive cells appear to have begun with the synthesis of cleavage-related compounds to allow for subsequent rapid divisions. Initiation of mitosis, for example, requires severe regulation through a network of regulators [118, 119]. Moreover, expression of the corresponding genes obviously needs to be highly cell type specific (Fig. 8b). The fact that there are many more genes with greater than 30-fold overexpression in reproductive cells compared to somatic cells, rather than vice versa (Figs. 3 and 7), might indicate that highly cell type-specific gene expression is more important for reproductive cells than for somatic cells, particularly with regards to pivotal processes with high requirements of regulation such as mitosis. However, expression of such genes does not need to be exceptionally strong because they do not appear within the most highly expressed genes of reproductive cells (Fig. 6b). The group of the most overexpressed genes in somatic cells contains a considerable number of genes that are flagella associated (Fig. 8a), which is hardly surprising since their expression is required for the biogenesis, turnover, and operation of flagella and only somatic cells develop flagella [120]. Nevertheless, some flagellar proteins are also involved in other processes (e.g., transport), which are relevant for both cell types.

In the group of the most overexpressed genes in somatic cells, genes coding for ECM compounds were also prominent (Fig. 6a), which is as expected considering that somatic cells are known to secrete large amounts of ECM during ontogenesis, causing cells to move apart from neighboring cells and a rapid organism size growth

[29, 72]. Once the process is complete, the volume of the ECM constitutes approximately 99% of the sphere. The complex ECM environment is considered an essential feature required for the evolutionary transition to multicellular volvocine algae [33, 35]. Previously, several ECM compounds were shown to be synthesized by somatic cells [43, 74, 79–81, 121–124]. It was previously assumed that reproductive cells, which lie below the somatic cell sheet, play only a minor role in ECM biosynthesis, if at all. However, in our analysis, it was surprising to find that, in the group of the most overexpressed genes in reproductive cells, the share of genes coding for ECM compounds was almost as large as that of such genes in somatic cells (Fig. 8a, b). Obviously, the involvement of reproductive cells in ECM biosynthesis is more important than previously thought. The reproductive cell is enclosed by a robust but thin ECM structure, termed the '(glycoprotein) vesicle' (although it contains no membrane). This vesicle protects the reproductive cell and grows during embryogenesis until completion of the inversion [125]. The exact composition of the vesicle and the mechanism of its synthesis remain unclear, yet it is likely that the vesicle is synthesized by the reproductive cells. The fact that both cell types have a significant share of genes coding for ECM compounds within the group of the most overexpressed genes indicates that, even though both cell types contribute significantly to the components and functions of the ECM, these contributions are clearly distinct.

Among the ECM compounds of *Volvox* is a large family of well-known ECM proteins, the pherophorins. Like ECM biosynthesis in general, the synthesis of pherophorins was previously attributed to somatic cells alone [29, 72, 79–82]. The expression of some pherophorin genes was shown to be induced by both a sex-inducer, which triggers sexual development, and by mechanical wounding [82, 86, 126]; however, in our analysis, these genes were also mainly expressed by somatic cells, as expected. Nevertheless, we identified nine pherophorins within the top 100 most highly overexpressed genes in reproductive cells. Furthermore, we found pherophorins among the most overexpressed genes in both somatic cells and reproductive cells to approximately the same extent (Fig. 11). More specifically, we identified 97 pherophorin domains in total, with 44% of these being overexpressed in somatic cells and 40% in reproductive cells (the remaining 16% showed none or unclear cell type specificity). The clear overexpression of two-fifths of all pherophorins in reproductive cells was another surprising result. Concerning ECM biosynthesis, it appears therefore likely that reproductive cells account for more than the synthesis of the glycoprotein vesicle. Future studies need to assess the role of this large number of pherophorins expressed by reproductive cells.

In our search for protein domains with assigned molecular function we identified 9435 domains in 6216 genes, corresponding to 44% of all gene loci. Furthermore, we detected 15,254 domains involved in biological processes, accounting for 43% of all gene loci. Then, we sorted the protein domains into groups and subgroups separately for 'molecular function' and 'biological process', determined the percentage share of each group and subgroup, and analyzed each group or subgroup for cell type-specific overexpression (Figs. 9 and 10). However, it should be noted that the automated classification was not sufficient to obtain a precise and definite assignment of the domain functions of a particular gene product but was a helpful resource to obtain representative insights into gene expression patterns of some functional groups. An example is the group of genes containing gene products with electron carrier activity, which includes several components of the photosynthetic electron transport system (Fig. 11). The expression analysis demonstrated that 40% of these genes were overexpressed in reproductive cells, whereas only 16% were overexpressed in somatic cells, which correlates with the expectation that reproductive cells have a higher photosynthetic performance. Another interesting group with a larger share of domains overexpressed in reproductive cells compared to somatic cells was 'antioxidant activity' (Fig. 11). This group includes, among others, genes coding for peroxiredoxins, thioredoxins, glutathione reductases, and superoxide dismutases. The corresponding proteins are involved in redox signaling and serve as scavengers for reactive oxygen (ROS) and nitrogen species (RNS) [127, 128]. Given the higher photosynthetic performance, and subsequent production of ROS and RNS, of reproductive cells, ROS and RNS scavenging seems to be more important for reproductive cells than for somatic cells. Some peroxiredoxins have been shown to be directly involved in the protection of the photosynthetic apparatus [127], which is more abundant in reproductive cells. Furthermore, preventing ROS or RNS damage to DNA and the resulting mutations is more important in reproductive cells because they represent the germ line.

Conversely, there were groups with a larger share of domains overexpressed in somatic cells compared to reproductive cells. The biggest difference was seen in the group of photoreceptors (Fig. 11), which seems to reflect the fact that only somatic cells have an eyespot apparatus with a basic visual system [50, 100, 129, 130]. Further, there was also a large share of signal-transducer-activity domains and signaling domains overexpressed in somatic cells (Fig. 11). The reason for this could be that somatic cells have a broader range of cellular tasks compared to



reproductive cells [8] and, therefore, may require a more complex signaling network.

A closer assessment was carried out for the group of transcription factors since control of gene transcription is a requirement for cell type-specific expression. More than half of all genes with transcription factor domains were shown to be differentially expressed with a fold difference in expression of 2 or more, whereby 32% of these domains were overexpressed in reproductive cells and 22% in somatic cells (Fig. 11). These large differences in the expression of transcription factor genes between the cell types reflects the strong differentiation of cell types despite *V. carteri* having only relatively recently diverged from its unicellular relatives [23, 25, 26].

## Conclusion

This study presents the first transcriptome-wide RNA-Seq analysis of separated cell types of *V. carteri* focusing on gene expression. Our expression dataset and the bioinformatic analyses provide a solid basis for further investigations of cell type-specific expression and its transcriptional regulation. Particularly, the expression analysis of potential key regulatory genes, like *rlsM* and other members of the *VARL* family, could promote the clarification of molecular mechanisms involved in the regulation of cellular differentiation. In addition, the large number of identified ECM-related genes with overexpression in reproductive cells was unexpected and should be subject to further studies.

## Methods

### Strain and culture conditions

The previously described female wild-type strain Eve10 of *V. carteri* f. *nagariensis* was used in this study [131]. Eve10 is a descendant of strains HK10 (female) and 69-1b (male) [117, 132], which originate from Japan. Cultures were grown synchronously under vegetative conditions in *Volvox* standard medium [117, 133]. Growth was performed at 28 °C in a cycle of 8 h dark/16 h cool fluorescent white light [134], at an average of ~100  $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$  photosynthetically active radiation in glass tubes or Fernbach flasks. The glass tubes had caps that allowed for gas exchange and the Fernbach flasks were aerated with approximately 50  $\text{cm}^3$  sterile air/min. To obtain three biological replicates for further processing (see below), three Fernbach flasks were inoculated with four Eve10 spheroids each and grown separately to a density of approximately 15 spheroids/mL each.

### Separation of cell types

The three separately grown and harvested cultures were kept separate. For each of these biological

replicates the two cell types, i.e., somatic cells and reproductive cells, were separated mechanically from each other as previously described [49]. Briefly, *Volvox* spheroids were harvested from Fernbach flasks shortly before the onset of cell cleavage of reproductive cells (Fig. 1a, c). The spheroids were broken in a Dounce homogenizer and cell size-based separation of cell types was achieved by successive filtration on screens of different mesh sizes and a centrifugation step. The separated somatic cell sheets and reproductive cells were quickly checked microscopically for possible contamination with the other cell type (Fig. 1d, e). Immediately after the separation of cell types, cells were flash-frozen in liquid nitrogen and stored at -80 °C for subsequent RNA extraction.

### RNA isolation

Total RNA was extracted separately from both cell types of each of the three biological replicates using the phenol-based TRI Reagent (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's instructions and as previously described [100]. The extracted RNA was dissolved in RNase-free water and stored at -70 °C. The RNA concentration was adjusted to 100  $\text{ng}/\mu\text{L}$  and quality was determined as described previously [92]. In addition, the quality of the cell type-specific RNA was re-checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), which performs microfluidic electrophoretic separation on microfabricated chips [135]. All of our six RNA samples passed both quality controls. The OD 260/280 was 1.8–2.2 and the RNA integrity number value was  $\geq 8$ .

### Whole transcriptome sequencing

The six samples with separately isolated total RNA of both cell types of each of the three biological replicates were prepared for RNA-Seq analysis. For each sample, 1  $\mu\text{g}$  of total RNA was subjected to poly(A)<sup>+</sup> RNA selection and mRNA fragmentation. This was followed by random primer cDNA synthesis and adapter ligation. Massively parallel sequencing was performed on an Illumina HiSeq2500 system with a read length of 50 bp. Purification of poly(A)<sup>+</sup> RNA and sequencing of the six samples was done by GATC Biotech (Konstanz, Germany). With sequencing by synthesis technology, each base in a read was assigned a quality score (Q-score) by a Phred-like algorithm [136, 137]. Across all reads, 92% of the bases had quality scores above Q30, which indicates that the probability of an incorrect base call was less than 1 in 1000. The sequenced reads were quality filtered by applying the FASTX Toolkit [138], which is a collection of command line tools for short-read FASTA/FASTQ file preprocessing. Quality-based trimming was performed with the GenePattern

Trimmomatic module using the following settings: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:36 [139]. All RNA-Seq data are available at the EMBL-EBI ArrayExpress repository under accession number E-MTAB-5691 [140–142].

### Mapping, data analysis, and bioinformatics

Quality filtered and trimmed sequence reads of all replicates of both cell types were mapped to the *V. carteri* f. *nagariensis* genome assembly v2 [16] using TopHat2 [143], with the following settings: tophat2 -r 300 -o \$Folder --library-type fr-unstranded -p 8 \$reference \$forward1\_R1.fastq,\$forward2\_R1.fastq \$reverse1\_R2.fastq,\$reverse2\_R2.fastq. Considering sequencing errors and the read length of 50 bp, up to two mismatches were allowed in mapping. Expression analysis and visualization was performed by using the short-read mapping analysis platform ReadXplorer 2.2.3 [52].

The ReadXplorer platform also includes the R package DESeq [54–56], which was used to normalize the count data and calculate mean values (baseMean, baseMeanA, baseMeanB), fold differences in expression, and *P* values (raw and adjusted) of a test for differential gene expression based on generalized linear models using negative binomial distribution errors. The baseMean is the mean normalized expression level of a given transcript, averaged over all replicates from all conditions; baseMeanA is the mean normalized expression level of a given transcript, averaged over the three biological replicates coming from reproductive cells (i.e., all condition A replicates); and baseMeanB is the mean normalized expression level of a given transcript, averaged over the three biological replicates coming from somatic cells (i.e., all condition B replicates) [54–56]. For quality control purposes, we also determined a minimum expression threshold. Therefore, we individually checked 100 expression profiles of genes that show an extremely low expression level, reflected by an extremely low baseMean value. We came to the assessment that a baseMean expression value of at least 12.5 is sufficient for robust expression analysis. As a consequence, we cannot make reliable statements regarding the expression of 999 predicted genes (7%) with a baseMean expression value below 12.5.

### Manual examination of gene structures on a random basis

The determination of gene structures is limited to genes with good coverage by mapped reads and therefore the expression needs to exceed a certain expression threshold. Thus, we individually checked 100 expression profiles of genes with low expression. We came to the assessment that genes with a baseMean expression value of at least 450 show sufficient coverage by mapped reads to estimate the correctness of

gene structures. An example of a typical expression profile of such a gene with an expression value of 450 is shown in Additional file 7: Figure S2A. Based on this empirical threshold baseMean expression value of 450, we calculated that our RNA-Seq data provide a source for the quality control of exon-intron gene structures of at least 8893 out of the 14,247 predicted genes, which corresponds to 62% of all predicted genes. For comparison, we also checked expression profiles of genes with average and high expression individually. The average of all baseMean expression values in our dataset of 14,203 genes with mapped reads was 2822. An example of a typical expression profile of a gene with average expression is shown in Additional file 7: Figure S2B. Gene structures can be examined easily. The same applies, of course, to genes with high expression (Additional file 7: Figure S2C).

For quality control purposes, we examined the intron-exon structures of a series of genes on a random basis by using the threshold of 450. To this effect, we manually and systematically checked all expression profiles mapped to the first 1 million base pairs of the randomly chosen scaffold 9. The expression profiles allowed the identification of 103 genes in this section of the genome, 102 of which are known genes and one gene is new (Additional file 2: Table S1). Among the 103 genes are 66 genes with a baseMean expression value of at least 450, which enables estimation of correctness of the corresponding gene structures. The expression profiles revealed that 14 of these 66 genes show discrepancies within the coding sequence as compared to the *Volvox* gene predictions of gene annotation v2.1 on the Phytozome V12 platform, which represents 21% of the sufficiently expressed genes in this section of the genome (Additional file 2: Table S1). Moreover, 35 of the 66 sufficiently expressed genes show discrepancies within the UTRs (45%).

In addition to this part of scaffold 9, we performed the same analysis for 100 randomly selected gene loci from the complete list of all gene loci. For this purpose, we used the random number generator of random.org to randomly pick 100 gene loci out of the total list of 14,247 previously known gene loci (Additional file 3: Table S2). Among these 100 genes were 60 sufficiently expressed genes. The expression profiles revealed that 10 of these 60 sufficiently expressed genes showed discrepancies within the coding sequence as compared to the *Volvox* gene predictions according to gene annotation v2.1 on the Phytozome V12 platform, which represents 17% of the sufficiently expressed genes (Additional file 3: Table S2). Moreover, 25 of the 60 sufficiently expressed genes showed discrepancies within the UTRs (42%).

In summary, both the sample analysis of a genome fragment and the analysis of 100 randomly selected gene loci gave approximately the same percentage of discrepancies between the *Volvox* gene predictions according to gene annotation v2.1 on the Phytozome V12 platform and our expression profiles.

### Annotation

We obtained standard gene information of *V. carteri* f. *nagariensis* from the Phytozome V12 platform [53], including Pfam [102–104], GO [108, 109], and PANTHER annotation data [105–107]. The *V. carteri* gene information of Phytozome V12 came originally from the US Department of Energy's Joint Genome Institute annotation v2.1 of *V. carteri* genome assembly v2 [16]. Genes were screened for the corresponding identifiers, the identifiers were assigned to higher level GO-terms, and genes were sorted into groups of molecular function and biological process using QuickGO [114, 115]. In gene subsets of particular importance (e.g., 50 most abundant transcripts per cell type), we re-checked the contained gene products without annotation by using the BLASTP algorithm to screen all non-redundant protein sequences in the National Center for Biotechnology Information protein databases for sequence similarities [110–112]. If a gene product without previous annotation was found to be similar to sequences with assigned function, and if the e-value was less than  $1 \times 10^{-10}$ , the same function was assigned to this gene product.

### Additional files

**Additional file 1: Figure S1.** Complete life cycle of *V. carteri*. (PDF 5372 kb)

**Additional file 2: Table S1.** Examples for discrepancies between gene predictions according to *V. carteri* genome annotation v2.1 on the Phytozome V12 platform and our expression profiles: examination of all expression profiles mapped to the first 1 million base pairs of the randomly chosen scaffold 9. (PDF 30 kb)

**Additional file 3: Table S2.** Examples for discrepancies between gene predictions according to *V. carteri* genome annotation v2.1 on the Phytozome V12 platform and our expression profiles: examination of 100 randomly chosen gene loci using the random number generator of random.org. (PDF 32 kb)

**Additional file 4: Table S3.** Cell type-specific gene expression, further information and references of 376 *Volvox* genes that show at least a brief mention in the literature. (PDF 116 kb)

**Additional file 5: Table S4.** Functional enrichment analysis of the most highly expressed genes in somatic cells, reproductive cells, and in total. (PDF 42 kb)

**Additional file 6: Table S5.** Functional enrichment analysis of the most overexpressed genes of both cell types. (PDF 54 kb)

**Additional file 7: Figure S2.** Examples of typical expression profiles of genes with low, average, and high expression. (PDF 236 kb)

### Abbreviations

BLASTP: basic local alignment search tool for protein-protein searches; cDNA: complementary deoxyribonucleic acid; ECM: extracellular matrix; GO: gene ontology (database); MA-plot: plot of log-intensity ratios (M-values)

versus log-intensity averages (A-values); mRNA: messenger ribonucleic acid; PANTHER: protein analysis through evolutionary relationships (database); Pfam: protein families (database); RNA-Seq: ribonucleic acid sequencing; RNS: reactive nitrogen species; ROS: reactive oxygen species

### Acknowledgements

We thank Dr Arash Kianianmomeni for RNA isolation and the availability of raw sequencing data. We are also grateful to Yannick Kerkhoff for his assistance with data analysis. Finally, we thank Kordula Puls for technical assistance and Eva Laura von der Heyde for constructive feedback. The bioinformatics support of the BMBF-funded project 'Bielefeld-Gießen Center for Microbial Bioinformatics – BiGi (Grant number 031A533)' within the German Network for Bioinformatics Infrastructure (de.NBI) is gratefully acknowledged.

### Funding

This work was supported by Bielefeld University. We also acknowledge support of the publication fee by Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University.

### Availability of data and materials

All RNA-Seq data supporting the conclusions of this article are available at the EMBL-EBI ArrayExpress repository under accession number E-MTAB-5691 [140–142]. Further data supporting the conclusions of this article are included in the article and its additional files.

### Authors' contributions

BK conducted the analytical part and wrote the first version of the manuscript. DW performed the bioinformatics approaches. AH (corresponding author) conceived and coordinated the study, critically evaluated the data, and finalized the manuscript. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

The use of the green alga *Volvox carteri* for research purposes does not require ethical approval.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

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

### Author details

<sup>1</sup>Department of Cellular and Developmental Biology of Plants, University of Bielefeld, Universitätsstr. 25, 33615 Bielefeld, Germany. <sup>2</sup>Center for Biotechnology (CeBiTec), University of Bielefeld, Bielefeld, Germany.

Received: 7 September 2017 Accepted: 30 October 2017

Published online: 28 November 2017

### References

- Bonner JT. The origins of multicellularity. *Integr Biol.* 1998;1:27–36.
- Maynard Smith J, Szathmáry E. *The Major Transitions in Evolution.* Oxford: Freeman; 1995.
- Grosberg RK, Strathmann R. The evolution of multicellularity: a minor major transition? *Annu Rev Ecol Evol Syst.* 2007;38:621–54.
- Knoll AH. The multiple origins of complex multicellularity. *Annu Rev Earth Planet Sci.* 2011;39:217–39.
- Niklas KJ, Newman SA. The origins of multicellular organisms. *Evol Dev.* 2013;15(1):41–52.
- Buss LW. *The Evolution of Individuality.* Princeton: Princeton University Press; 1987.
- Cavalier-Smith T. Cell diversification in heterotrophic flagellates. In: Patterson DJ, Larsen J, editors. *The biology of free-living heterotrophic flagellates.* Oxford: Clarendon Press; 1991. p. 113–31.
- Kirk DL. *Volvox: molecular-genetic origins of multicellularity and cellular differentiation.* 1st ed. Cambridge: Cambridge University Press; 1998.

9. Bonner JT. *First Signals: The Evolution of Multicellular Development*. Princeton: Princeton University Press; 2001.
10. Carroll SB. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*. 2001;409(6823):1102–9.
11. Kaiser D. Building a multicellular organism. *Annu Rev Genet*. 2001;35:103–23.
12. Medina M, Collins AG, Taylor JW, Valentine JW, Lipps JH, Amaral-Zettler L. Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int J Astrobiol*. 2003;2:203–11.
13. Baldauf SL. The deep roots of eukaryotes. *Science*. 2003;300(5626):1703–6.
14. King N. The unicellular ancestry of animal development. *Dev Cell*. 2004;7(3):313–25.
15. Rokas A. The molecular origins of multicellular transitions. *Curr Opin Genet Dev*. 2008;18(6):472–8.
16. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, et al. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*. 2010;329(5988):223–6.
17. Bonner JT. On the origin of differentiation. *J Biosci*. 2003;28(4):523–8.
18. Hallmann A. Experienced developers of multicellularity - the Volvocales. *Bioforum Europe*. 2003;6:326–8.
19. Hoenigsberg HF, Tijaro MH, Sanabria C. From unicellularity to multicellularity - molecular speculations about early animal evolution. *Genet Mol Res*. 2008;7(1):50–9.
20. Rokas A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet*. 2008;42:235–51.
21. Sachs JL. Resolving the first steps to multicellularity. *Trends Ecol Evol*. 2008;23(5):245–8.
22. Willensdorfer M. On the evolution of differentiated multicellularity. *Evolution*. 2009;63(2):306–23.
23. Herron MD. Many from one: lessons from the volvocine algae on the evolution of multicellularity. *Commun Integr Biol*. 2009;2(4):368–71.
24. Miller SM. *Volvox*, *Chlamydomonas*, and the evolution of multicellularity. *Nat Educ*. 2010;3(9):65.
25. Herron MD, Michod RE. Evolution of complexity in the volvocine algae: transitions in individuality through Darwin's eye. *Evolution*. 2008;62(2):436–51.
26. Rausch H, Larsen N, Schmitt R. Phylogenetic relationships of the green alga *Volvox carteri* deduced from small-subunit ribosomal RNA comparisons. *J Mol Evol*. 1989;29(3):255–65.
27. Kirk DL. Germ-soma differentiation in *Volvox*. *Dev Biol*. 2001;238(2):213–23.
28. Nozaki H. Origin and evolution of the genera *Pleodorina* and *Volvox* (Volvocales). *Biologia (Bratisl)*. 2003;58(4):425–31.
29. Hallmann A. Extracellular matrix and sex-inducing pheromone in *Volvox*. *Int Rev Cytol*. 2003;227:131–82.
30. Schmitt R. Differentiation of germinal and somatic cells in *Volvox carteri*. *Curr Opin Microbiol*. 2003;6(6):608–13.
31. Kirk DL. Seeking the ultimate and proximate causes of *Volvox* multicellularity and cellular differentiation. *Integr Comp Biol*. 2003;43(2):247–53.
32. Kirk MM, Kirk DL. Exploring germ-soma differentiation in *Volvox*. *J Biosci*. 2004;29(2):143–52.
33. Kirk DL. A twelve-step program for evolving multicellularity and a division of labor. *BioEssays*. 2005;27(3):299–310.
34. Michod RE, Viossat Y, Solari CA, Hurand M, Nedelcu AM. Life-history evolution and the origin of multicellularity. *J Theor Biol*. 2006;239(2):257–72.
35. Hallmann A. Evolution of reproductive development in the volvocine algae. *Sex Plant Reprod*. 2011;24:97–112.
36. Kirk DL. The genetic program for germ-soma differentiation in *Volvox*. *Annu Rev Genet*. 1997;31:359–80.
37. Kirk DL. Genetic control of reproductive cell differentiation in *Volvox*. In: Wiessner W, Robinson DG, Starr RC, editors. *Experimental Phycology 1: Cell Walls and Surfaces, Reproduction and Photosynthesis*. Berlin: Springer-Verlag; 1990. p. 81–94.
38. Miller SM, Kirk DL. *glsA*, a *Volvox* gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. *Development*. 1999;126(4):649–58.
39. Cheng Q, Pappas V, Hallmann A, Miller SM. Hsp70A and GlS A interact as partner chaperones to regulate asymmetric division in *Volvox*. *Dev Biol*. 2005;286(2):537–48.
40. Cheng Q, Hallmann A, Edwards L, Miller SM. Characterization of a heat-shock-inducible *hsp70* gene of the green alga *Volvox carteri*. *Gene*. 2006;371(1):112–20.
41. Pappas V, Miller SM. Functional analysis of the *Volvox carteri* asymmetric division protein GlS A. *Mech Dev*. 2009;126(10):842–51.
42. Matt G, Umen J. *Volvox*: A simple algal model for embryogenesis, morphogenesis and cellular differentiation. *Dev Biol*. 2016;419(1):99–113.
43. Kirk DL, Kirk MM. Protein synthetic patterns during the asexual life cycle of *Volvox carteri*. *Dev Biol*. 1983;96(2):493–506.
44. Tam LW, Kirk DL. Identification of cell-type-specific genes of *Volvox carteri* and characterization of their expression during the asexual life cycle. *Dev Biol*. 1991;145(1):51–66.
45. Meissner M, Stark K, Cresnar B, Kirk DL, Schmitt R. *Volvox* germline-specific genes that are putative targets of RegA repression encode chloroplast proteins. *Curr Genet*. 1999;36(6):363–70.
46. Kirk MM, Stark K, Miller SM, Müller W, Taillon BE, Gruber H, Schmitt R, Kirk DL. *regA*, a *Volvox* gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. *Development*. 1999;126(4):639–47.
47. Kirk DL. The ontogeny and phylogeny of cellular differentiation in *Volvox*. *Trends Genet*. 1988;4(2):32–6.
48. Kirk DL, Kaufman MR, Keeling RM, Stamer KA. Genetic and cytological control of the asymmetric divisions that pattern the *Volvox* embryo. *Dev Suppl*. 1991;1:67–82.
49. Nematollahi G, Kianianmomeni A, Hallmann A. Quantitative analysis of cell-type specific gene expression in the green alga *Volvox carteri*. *BMC Genomics*. 2006;7:321.
50. Kianianmomeni A, Hallmann A. Transcriptional analysis of *Volvox* photoreceptors suggests the existence of different cell-type specific light-signaling pathways. *Curr Genet*. 2015;61(1):3–18.
51. Dueck A, Evers M, Henz SR, Unger K, Eichner N, Merkl R, Berezikov E, Engelmann JC, Weigel D, Wenzl S, et al. Revealing insights into evolution and origins *Volvox carteri* reveal insights into evolution and origins of small RNA systems in plants. *BMC Genomics*. 2016;17:853.
52. Hilker R, Stadermann KB, Doppmeier D, Kalinowski J, Stoye J, Straube J, Winnebal J, Goesmann A. ReadXplorer - visualization and analysis of mapped sequences. *Bioinformatics*. 2014;30(16):2247–54.
53. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(Database issue):D1178–1186.
54. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
55. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc*. 2013;8(9):1765–86.
56. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2016. <http://www.R-project.org/>. Accessed 13 Nov 2017.
57. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
58. Harper JF, Huson KS, Kirk DL. Use of repetitive sequences to identify DNA polymorphisms linked to *regA*, a developmentally important locus in *Volvox*. *Genes Dev*. 1987;1(6):573–84.
59. Kirk DL, Baran GJ, Harper JF, Huskey RJ, Huson KS, Zagris N. Stage-specific hypermutability of the *regA* locus of *Volvox*, a gene regulating the germ-soma dichotomy. *Cell*. 1987;48(1):11–24.
60. Stark K, Kirk DL, Schmitt R. Two enhancers and one silencer located in the introns of *regA* control somatic cell differentiation in *Volvox carteri*. *Genes Dev*. 2001;15(11):1449–60.
61. Babinger K, Hallmann A, Schmitt R. Translational control of *regA*, a key gene controlling cell differentiation in *Volvox carteri*. *Development*. 2006;133(20):4045–51.
62. Duncan L, Nishii I, Howard A, Kirk D, Miller SM. Orthologs and paralogs of *regA*, a master cell-type regulatory gene in *Volvox carteri*. *Curr Genet*. 2006;50(1):61–72.
63. Duncan L, Nishii I, Harryman A, Buckley S, Howard A, Friedman NR, Miller SM. The *VARL* gene family and the evolutionary origins of the master cell-type regulatory gene, *regA*, in *Volvox carteri*. *J Mol Evol*. 2007;65(1):1–11.
64. Olson BJ, Nedelcu AM. Co-option during the evolution of multicellular and developmental complexity in the volvocine green algae. *Curr Opin Genet Dev*. 2016;39:107–15.
65. Kianianmomeni A, Nematollahi G, Hallmann A. A gender-specific retinoblastoma-related protein in *Volvox carteri* implies a role for the retinoblastoma protein family in sexual development. *Plant Cell*. 2008;20(9):2399–419.
66. van den Heuvel S, Dyson NJ. Conserved functions of the pRB and E2F families. *Nat Rev Mol Cell Biol*. 2008;9(9):713–24.

67. Hallmann A. Key elements of the retinoblastoma tumor suppressor pathway in *Volvox carteri*. *Commun Integr Biol.* 2009;2(5):396–9.
68. Ferris P, Olson BJ, De Hoff PL, Douglass S, Casero D, Prochnik S, Geng S, Rai R, Grimwood J, Schmutz J, et al. Evolution of an expanded sex-determining locus in *Volvox*. *Science.* 2010;328(5976):351–4.
69. Hiraide R, Kawai-Toyooka H, Hamaji T, Matsuzaki R, Kawafune K, Abe J, Sekimoto H, Umen J, Nozaki H. The evolution of male-female sexual dimorphism predates the gender-based divergence of the mating locus gene *MAT3/RB*. *Mol Biol Evol.* 2013;30(5):1038–40.
70. Hanschen ER, Marriage TN, Ferris PJ, Hamaji T, Toyoda A, Fujiyama A, Neme R, Noguchi H, Minakuchi Y, Suzuki M, et al. The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat Commun.* 2016;7:11370.
71. Huber O, Sumper M. Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. *EMBO J.* 1994;13(18):4212–22.
72. Sumper M, Hallmann A. Biochemistry of the extracellular matrix of *Volvox*. *Int Rev Cytol.* 1998;180:51–85.
73. Cheng Q, Fowler R, Tam LW, Edwards L, Miller SM. The role of GlsA in the evolution of asymmetric cell division in the green alga *Volvox carteri*. *Dev Genes Evol.* 2003;213(7):328–35.
74. Ertl H, Mengele R, Wenzl S, Engel J, Sumper M. The extracellular matrix of *Volvox carteri*: molecular structure of the cellular compartment. *J Cell Biol.* 1989;109(6 Pt 2):3493–501.
75. Holst O, Christoffel V, Fründ R, Moll H, Sumper M. A phosphodiester bridge between two arabinose residues as a structural element of an extracellular glycoprotein of *Volvox carteri*. *Eur J Biochem.* 1989;181(2):345–50.
76. Tam LW, Kirk DL. The program for cellular differentiation in *Volvox carteri* as revealed by molecular analysis of development in a gonidialess/somatic regenerator mutant. *Development.* 1991;112(2):571–80.
77. Tam LW, Stamer KA, Kirk DL. Early and late gene expression programs in developing somatic cells of *Volvox carteri*. *Dev Biol.* 1991;145(1):67–76.
78. Mengele R, Sumper M. Gulose as a constituent of a glycoprotein. *FEBS Lett.* 1992;298(1):14–6.
79. Sumper M, Berg E, Wenzl S, Godl K. How a sex pheromone might act at a concentration below  $10^{-16}$  M. *EMBO J.* 1993;12(3):831–6.
80. Godl K, Hallmann A, Rappel A, Sumper M. Pherophorins: a family of extracellular matrix glycoproteins from *Volvox* structurally related to the sex-inducing pheromone. *Planta.* 1995;196(4):781–7.
81. Godl K, Hallmann A, Wenzl S, Sumper M. Differential targeting of closely related ECM glycoproteins: the pherophorin family from *Volvox*. *EMBO J.* 1997;16(1):25–34.
82. Hallmann A. The pherophorins: common, versatile building blocks in the evolution of extracellular matrix architecture in Volvocales. *Plant J.* 2006;45(2):292–307.
83. Hallmann A, Amon P, Godl K, Heitzer M, Sumper M. Transcriptional activation by the sexual pheromone and wounding: a new gene family from *Volvox* encoding modular proteins with (hydroxy)proline-rich and metalloproteinase homology domains. *Plant J.* 2001;26(6):583–93.
84. Heitzer M, Hallmann A. An extracellular matrix-localized metalloproteinase with an exceptional QEXXH metal binding site prefers copper for catalytic activity. *J Biol Chem.* 2002;277(31):28280–6.
85. Shimizu T, Inoue T, Shiraishi H. Cloning and characterization of novel extensin-like cDNAs that are expressed during late somatic cell phase in the green alga *Volvox carteri*. *Gene.* 2002;284(1–2):179–87.
86. Amon P, Haas E, Sumper M. The sex-inducing pheromone and wounding trigger the same set of genes in the multicellular green alga *Volvox*. *Plant Cell.* 1998;10(5):781–9.
87. Fukada K, Inoue T, Shiraishi H. A posttranslationally regulated protease, VheA, is involved in the liberation of juveniles from parental spheroids in *Volvox carteri*. *Plant Cell.* 2006;18(10):2554–66.
88. Nishimura M, Nagashio R, Sato Y, Hasegawa T. Late Somatic Gene 2 disrupts parental spheroids cooperatively with *Volvox* hatching enzyme A in *Volvox*. *Planta.* 2017;245(1):183–92.
89. Mages W, Salbaum JM, Harper JF, Schmitt R. Organization and structure of *Volvox* alpha-tubulin genes. *Mol Gen Genet.* 1988;213(2–3):449–58.
90. Harper JF, Mages W. Organization and structure of *Volvox* beta-tubulin genes. *Mol Gen Genet.* 1988;213(2–3):315–24.
91. Mages W, Cresnar B, Harper JF, Brüderlein M, Schmitt R. *Volvox carteri* alpha 2- and beta 2-tubulin-encoding genes: regulatory signals and transcription. *Gene.* 1995;160(1):47–54.
92. Kianianmomeni A, Hallmann A. Validation of reference genes for quantitative gene expression studies in *Volvox carteri* using real-time RT-PCR. *Mol Biol Rep.* 2013;40(12):6691–9.
93. Mitchell DR, Brown KS. Sequence analysis of the *Chlamydomonas* alpha and beta dynein heavy chain genes. *J Cell Sci.* 1994;107(Pt 3):635–44.
94. Porter ME, Knott JA, Myster SH, Farlow SJ. The dynein gene family in *Chlamydomonas reinhardtii*. *Genetics.* 1996;144(2):569–85.
95. Mitchell DR, Brown KS. Sequence analysis of the *Chlamydomonas reinhardtii* flagellar alpha dynein gene. *Cell Motil Cytoskeleton.* 1997;37(2):120–6.
96. Myster SH, Knott JA, O'Toole E, Porter ME. The *Chlamydomonas Dhc1* gene encodes a dynein heavy chain subunit required for assembly of the 11 inner arm complex. *Mol Biol Cell.* 1997;8(4):607–20.
97. Yagi T, Minoura I, Fujiwara A, Saito R, Yasunaga T, Hirono M, Kamiya R. An axonemal dynein particularly important for flagellar movement at high viscosity. Implications from a new *Chlamydomonas* mutant deficient in the dynein heavy chain gene *DHC9*. *J Biol Chem.* 2005;280(50):41412–20.
98. Savard F, Richard C, Guertin M. The *Chlamydomonas reinhardtii* *L1818* gene represents a distant relative of the *cab1/III* genes that is regulated during the cell cycle and in response to illumination. *Plant Mol Biol.* 1996;32(3):461–73.
99. Hallmann A. Enzymes in the extracellular matrix of *Volvox*: an inducible, calcium-dependent phosphatase with a modular composition. *J Biol Chem.* 1999;274(3):1691–7.
100. Kianianmomeni A, Stehfest K, Nematollahi G, Hegemann P, Hallmann A. Channelrhodopsins of *Volvox carteri* are photochromic proteins that are specifically expressed in somatic cells under control of light, temperature, and the sex inducer. *Plant Physiol.* 2009;151(1):347–66.
101. Ebnat E, Fischer M, Deininger W, Hegemann P. Volvoxrhodopsin, a light-regulated sensory photoreceptor of the spheroidal green alga *Volvox carteri*. *Plant Cell.* 1999;11(8):1473–84.
102. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40(Database issue):D290–301.
103. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222–230.
104. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44(D1):D279–285.
105. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* 2003;31(1):334–41.
106. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013;41(Database issue):D377–386.
107. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 2016;44(D1):D336–342.
108. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–9.
109. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004;32(Database issue):D258–261.
110. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
111. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server issue):W5–9.
112. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezuk Y, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 2013;41(Web Server issue):W29–33.
113. Hallmann A, Kirk DL. The developmentally regulated ECM glycoprotein ISG plays an essential role in organizing the ECM and orienting the cells of *Volvox*. *J Cell Sci.* 2000;113(Pt24):4605–17.
114. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for gene ontology searching. *Bioinformatics.* 2009;25(22):3045–6.

115. Huntley RP, Binns D, Dimmer E, Barrell D, O'Donovan C, Apweiler R. QuickGO: a user tutorial for the web-based gene ontology browser. Database. 2009;2009:bap010.
116. Kirk DL, Kirk MM, Stamer KA, Larson A. The genetic basis for the evolution of multicellularity and cellular differentiation in the volvocine green algae. In: Dudley EC, editor. *The Unity of Evolutionary Biology*. Portland: Dioscorides Press; 1990. p. 568–81.
117. Starr RC. Structure, reproduction and differentiation in *Volvox carteri f. nagariensis* lyengar, strains HK 9 & 10. Arch Protistenkd. 1969;111:204–22.
118. Feilotter H, Lingner C, Rowley R, Young PG. Regulation of the G2-mitosis transition. Biochem Cell Biol. 1992;70(10-11):954–71.
119. Cross FR, Umen JG. The *Chlamydomonas* cell cycle. Plant J. 2015;82(3):370–92.
120. Marshall WF, Rosenbaum JL. Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. J Cell Biol. 2001;155(3):405–14.
121. Wenzl S, Sumper M. Sulfation of a cell surface glycoprotein correlates with the developmental program during embryogenesis of *Volvox carteri*. Proc Natl Acad Sci U S A. 1981;78(6):3716–20.
122. Wenzl S, Sumper M. The occurrence of different sulphated cell surface glycoproteins correlates with defined developmental events in *Volvox*. FEBS Lett. 1982;143(2):311–5.
123. Ender F, Godl K, Wenzl S, Sumper M. Evidence for autocatalytic cross-linking of hydroxyproline-rich glycoproteins during extracellular matrix assembly in *Volvox*. Plant Cell. 2002;14(5):1147–60.
124. Wenzl S, Thym D, Sumper M. Development-dependent modification of the extracellular matrix by a sulphated glycoprotein in *Volvox carteri*. EMBO J. 1984;3:739–44.
125. Ueki N, Nishii I. Controlled enlargement of the glycoprotein vesicle surrounding a *Volvox* embryo requires the InvB nucleotide-sugar transporter and is required for normal morphogenesis. Plant Cell. 2009;21(4):1166–81.
126. Ender F, Hallmann A, Amon P, Sumper M. Response to the sexual pheromone and wounding in the green alga *Volvox*: induction of an extracellular glycoprotein consisting almost exclusively of hydroxyproline. J Biol Chem. 1999;274(49):35023–8.
127. Dayer R, Fischer BB, Eggen RI, Lemaire SD. The peroxiredoxin and glutathione peroxidase families in *Chlamydomonas reinhardtii*. Genetics. 2008;179(1):41–57.
128. Dietz KJ. Peroxiredoxins in plants and cyanobacteria. Antioxid Redox Signal. 2011;15(4):1129–59.
129. Hegemann P. Vision in microalgae. Planta. 1997;203(3):265–74.
130. Kianianmomeni A, Hallmann A. Algal photoreceptors: in vivo functions and potential applications. Planta. 2014;239(1):1–26.
131. Adams CR, Stamer KA, Miller JK, McNally JG, Kirk MM, Kirk DL. Patterns of organellar and nuclear inheritance among progeny of two geographically isolated strains of *Volvox carteri*. Curr Genet. 1990;18(2):141–53.
132. Starr RC. Control of differentiation in *Volvox*. Dev Biol Suppl. 1970;4:59–100.
133. Provasoli L, Pintner IJ. Artificial media for fresh-water algae: problems and suggestions. In: Tryon CA, Hartman RT, editors. *The Ecology of Algae, a symposium held at the Pymatuning Laboratory of Field Biology on June 18 and 19, 1959*. 1st ed. Pittsburgh: The Pymatuning Symposia in Ecology, Special Publication No. 2., University of Pittsburgh; 1959. p. 84–96.
134. Starr RC, Jaenicke L. Purification and characterization of the hormone initiating sexual morphogenesis in *Volvox carteri f. nagariensis* lyengar. Proc Natl Acad Sci U S A. 1974;71(4):1050–4.
135. Mueller O, Hahnenberger K, Dittmann M, Yee H, Dubrow R, Nagle R, Rlsley D. A microfluidic system for high-speed reproducible DNA sizing and quantitation. Electrophoresis. 2000;21(1):128–34.
136. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II Error probabilities. Genome Res. 1998;8(3):186–94.
137. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I Accuracy assessment. Genome Res. 1998;8(3):175–85.
138. Pearson WR, Wood T, Zhang Z, Miller W. Comparison of DNA sequences with protein sequences. Genomics. 1997;46(1):24–36.
139. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
140. Annotare. European Bioinformatics Institute. 2017. <https://www.ebi.ac.uk/fg/annotare>. Accessed 13 Nov 2017.
141. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi M, Burdett T, et al. ArrayExpress update - simplifying data submissions. Nucleic Acids Res. 2015;43(Database issue):D1113–1116.
142. Arrayexpress. European Bioinformatics Institute. 2017. <https://www.ebi.ac.uk/arrayexpress/>. Accessed 13 Nov 2017.
143. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.
144. Kirk DL, Viamontes GI, Green KJ, Bryant Jr JL. Integrated morphogenetic behavior of cell sheets: *Volvox* as a model. In: Subtelny S, Green PB, editors. *Developmental Order: Its Origin and Regulation*. New York: Alan R. Liss; 1982. p. 247–74.
145. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (Methodological). 1995;57(1):289–300.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

