

ORIGINAL RESEARCH

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

A Next-Generation Sequencing Approach to Study the Transcriptomic Changes During the Differentiation of *Physarum* at the Single-Cell Level

Israel Barrantes¹, Jeremy Leipzig² and Wolfgang Marwan¹

¹Molecular Network Analysis Group, Magdeburg Centre for Systems Biology, and Lehrstuhl für Regulationsbiologie, Otto von Guericke University, Magdeburg, Germany. ²Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. Corresponding author email: walter.barrantes-bustinza@ovgu.de

Abstract: *Physarum polycephalum* is a unicellular eukaryote belonging to the amoebozoa group of organisms. The complex life cycle involves various cell types that differ in morphology, function, and biochemical composition. Sporulation, one step in the life cycle, is a stimulus-controlled differentiation response of macroscopic plasmodial cells that develop into fruiting bodies. Well-established Mendelian genetics and the occurrence of macroscopic cells with a naturally synchronous population of nuclei as source of homogeneous cell material for biochemical analyses make *Physarum* an attractive model organism for studying the regulatory control of cell differentiation. Here, we develop an approach using RNA-sequencing (RNA-seq), without needing to rely on a genome sequence as a reference, for studying the transcriptomic changes during stimulus-triggered commitment to sporulation in individual plasmodial cells. The approach is validated through the obtained expression patterns and annotations, and particularly the results from up- and downregulated genes, which correlate well with previous studies.

Keywords: cell differentiation, single-cell methods, RNA-seq, *Physarum polycephalum*

Gene Regulation and Systems Biology 2012:6 127–137

doi: [10.4137/GRSB.S10224](https://doi.org/10.4137/GRSB.S10224)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.

Introduction

Physarum polycephalum (“slime mold”) is a free-living amoebozoan protist. Due to its rich cell biology, *Physarum* is a model organism for cell motility¹ and cell differentiation.^{2–4} More recently in RNA editing,⁵ *Physarum* is also a model organism for DNA replication,⁶ optimization of cell morphology⁷ and behavioral intelligence.⁸ During its branched life cycle, *Physarum* differentiates into several cell types that occur in temporal order (Fig. 1). The multinucleate, macroscopic plasmodium, for example, may develop from microscopic, single nucleate amoebal cells or develop by germination of spherules.⁹ The number of nuclei contained in a plasmodium depends on its size. A plasmodium in a 9 cm petri dish, for example, contains approximately 10^7 nuclei that display natural synchrony in terms of cell cycle and differentiation status.^{10,11} During sporulation, the plasmodium differentiates into fruiting bodies containing spores which undergo a meiotic recombination to become

the precursor cells of amoebae that may develop into gametes later on. Sporulation occurs only in starved plasmodial cells and can be experimentally triggered, e.g. by exposure to far red light or to heat shock. Well-established Mendelian genetics and the occurrence of naturally synchronous macroscopic cells as source of homogeneous cell material for biochemical analyses have made *Physarum* an attractive model organism for studying the regulatory control of cell differentiation. Recent advances in sequencing technologies motivated detailed analyses of transcriptome dynamics during the life cycle of *Physarum*.^{12,13} These studies were performed using RNA isolated from cell populations.

In general, differentiation follows spatial and temporal changes in transcript abundance in a cell type-specific manner. Stochastic variations in gene expression presumably do impact cell-fate decisions,¹⁴ and therefore the time-resolved analysis of gene expression patterns in individual cells would provide valuable insight as compared to averaged data from measurements obtained from cell populations.^{14,15} Expression patterns of single-cells have been analyzed using deep RNA sequencing (RNA seq)¹⁶ to characterize the transcriptomes of individual embryonic mouse cells separated by technically complex procedures, and relying on the mouse genomic information for transcript assembly and mapping.^{17,18}

The *Physarum* genome has been sequenced but the data are still not assembled into a complete genome sequence (The Genome Institute, Washington University School of Medicine; http://genome.wustl.edu/genomes/view/physarum_polycephalum/). Therefore, here we evaluated the possibility of studying the global transcriptional changes during the differentiation of a single cell without relying on genomic information, and developed an approach to analyze the differential expression at several time points during the commitment of a plasmodial cell to sporulation.

Methods

Physarum macroplasmodia (apogamic strain WT31¹⁹) were cultured as previously described.^{13,20} Cells were grown and collected under two different conditions: (i) a plasmodium starved for 6 days (competent D1 and D2 individual cell samples); and (ii) a plasmodium starved for 6 days, exposed to far red light for 30 minutes, and returned to the dark for 6.5 hours

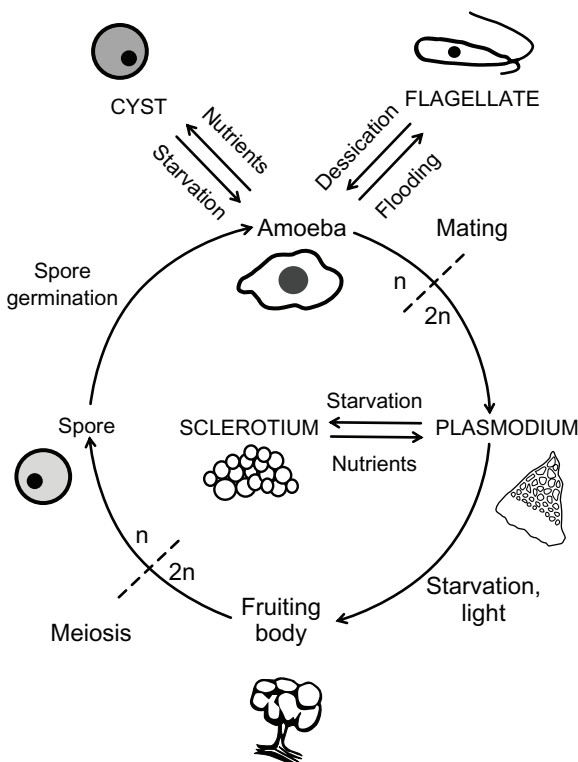


Figure 1. The life cycle of *Physarum polycephalum*.^{3,9,35}

Notes: Spores, released from mature fruiting bodies, germinate into mononuclear amoebae (n), which propagate by mitosis. At high population density, amoebae are able to mate, to form a zygote ($2n$). This diploid cell later develops into a multinuclear plasmodium ($2n$), through multiple nuclear divisions. Following starvation, the plasmodium can be induced to sporulation by visible light. Later, the plasmodial mass develops into individual fruiting bodies, which will subsequently yield haploid spores (n).



(L1 and L2 photoinduced cells; Table 1). During this time period the cell becomes irreversibly committed to sporulation.²¹ Samples were frozen with liquid nitrogen and Poly(A)+ RNA was isolated from the total RNA samples (by two rounds of oligo(dT) affinity chromatography), and fragmented with ultrasound (4 pulses of 30 sec at 4 °C). Subsequently, the RNA fragments were poly(A)-tailed using poly(A) polymerase, followed by treatment with tobacco acid pyrophosphatase (TAP). Then a RNA adapter was ligated to the 5'-monophosphate of the RNA. First-strand complementary DNA (cDNA) synthesis was performed using an oligo(dT)-adapter primer and the Maloney murine leukemia virus (M-MLV) reverse transcriptase. The resulting cDNAs were polymerase chain reaction (PCR)-amplified for 14–15 cycles to about 20–30 ng/μL, including distinctive 4-bp 5'-barcodes for each sample (Table 1), and using a high fidelity DNA polymerase. The PCR products were purified with the Agencourt AMPure XP kit (Beckman Coulter Genomics), and pooled in equivalent amounts. cDNAs in the range of 200 to 400 bp were fractionated from agarose gels and sequenced using the Illumina HiSeq 2000 system. Poly(A)+ RNA isolation, cDNA synthesis and cDNA library preparation as described here were carried out by vertis Biotechnologie (Freising-Weihenstephan, Germany). The 100-bp sequencing outputs were then trimmed for quality

(Phred score > 33), and later assembled de novo, using velvet²² and oases (k-mers: 31, 41, 51). Later, CAP3²³ was employed to reduce redundancy in the assembly. The annotation of this assembly was carried out first through BLAST²⁴ searches (*e*-value 1E-3) against the SwissProt²⁵ protein database. A search for *Physarum* noncoding RNAs was not included due to the lack of complete gene models and a finished reference genome in this species. Afterwards, domains and protein signature patterns were associated from matches to the InterPro database, and Gene Ontology (GO) based annotations were assigned using Blast2GO,²⁶ from annotations pertaining to orthologs (annotation *e*-value cutoff < 1E-6). Gene names and descriptions were filtered using the Blast Description Annotation tool from Blast2GO.²⁶ Significant differences in GO annotations between sets of up- and downregulated genes from each cDNA library were evaluated using Fisher exact tests, as implemented in Blast2GO.

To assess the differential expression between the several single cells, the sequencing output was split using the barcode information for each sample. Then the decoded outputs were mapped to the novel assembly with Bowtie.²⁷ Samtools²⁸ and Tablet²⁹ were later used to obtain mapped read counts. For expression comparisons, we obtained the following for each transcript: (i) the number of mapped reads, and (ii) the normalized expression value, as

Table 1. Summary of the single-cell transcriptome sequencing.

| Culture | Competent plasmodium 1 | Competent plasmodium 2 | Induced plasmodium 1 | Induced plasmodium 2 |
|---|------------------------|------------------------|----------------------|----------------------|
| Replicate ID | D1 | D2 | L1 | L2 |
| Starvation time (days) | 6.0 | 6.0 | 6.0 | 6.0 |
| Light exposure (hrs) | – | – | 0.5 | 0.5 |
| Collection time after exposure (hrs) | – | – | 6.5 | 6.5 |
| RNA samples | | | | |
| Concentration (ng/μL) | 274.0 | 250.2 | 472.8 | 666.4 |
| Amount (ug) | 120.6 | 81.3 | 93.6 | 129.3 |
| cDNA amplification cycles | 14 | 15 | 14 | 15 |
| Sequencing | | | | |
| Barcode | CAGATC | ACTTGA | GATCAG | TAGCTT |
| Total bases (bp) | 1,994,171,100 | 1,949,941,700 | 1,827,921,300 | 1,934,609,000 |
| Total reads | 19,941,711 | 19,499,417 | 18,279,213 | 19,346,090 |
| Reads used for assembly | 19,930,198 | 19,489,244 | 18,269,297 | 19,334,649 |
| %GC of reads used for assembly | 34% | 34% | 31% | 32% |
| Total mapped reads | 2,128,193 | 2,076,582 | 1,567,249 | 1,937,222 |
| Assembled transcripts with mapped reads | 9,365 | 9,222 | 9,087 | 9,226 |

Notes: A summary of the single-cell RNA libraries is shown, corresponding to the competent (libraries D1 and D2), and photoinduced individual cell replicates (L1 and L2).

measured in reads per kilobase per million mapped reads (RPKM).³⁰ To identify differentially expressed transcripts between starved and photoinduced cells, the non-normalized mapped read count data was analyzed using the R-based package DESeq.³¹ Transcript abundances for each gene were estimated as a weighted mean of mapped read counts from each replicate, normalized to the library size. *P*-values (adjusted for false discovery rate) were generated for each gene in pairwise comparisons between different conditions (competent and induced cells). We used the per-condition method and fit-only sharing mode. A summary of experiments and bioinformatic analyses is depicted in Figure 2.

Results and Discussion

Four datasets consisting a total of 77.07 million 100-base reads from single cell *Physarum* plasmodia were obtained from the Illumina sequencing (77.02 million reads with Phred score > 33; 7.12 Gb). This RNA-seq output was deposited in the European Nucleotide Archive³² (study accession ERP001220). Replicate data distributions were 1.85 and 1.82 Gb corresponding to the starved plasmodium (cDNA library replicates D1 and D2), and 1.67 and 1.78 Gb for the cells collected 6 hours after photoinduction (libraries L1 and L2; Table 1). Therefore, assuming a 10% of protein-encoding genes, we obtained a 237.32x coverage for the 300 Mb

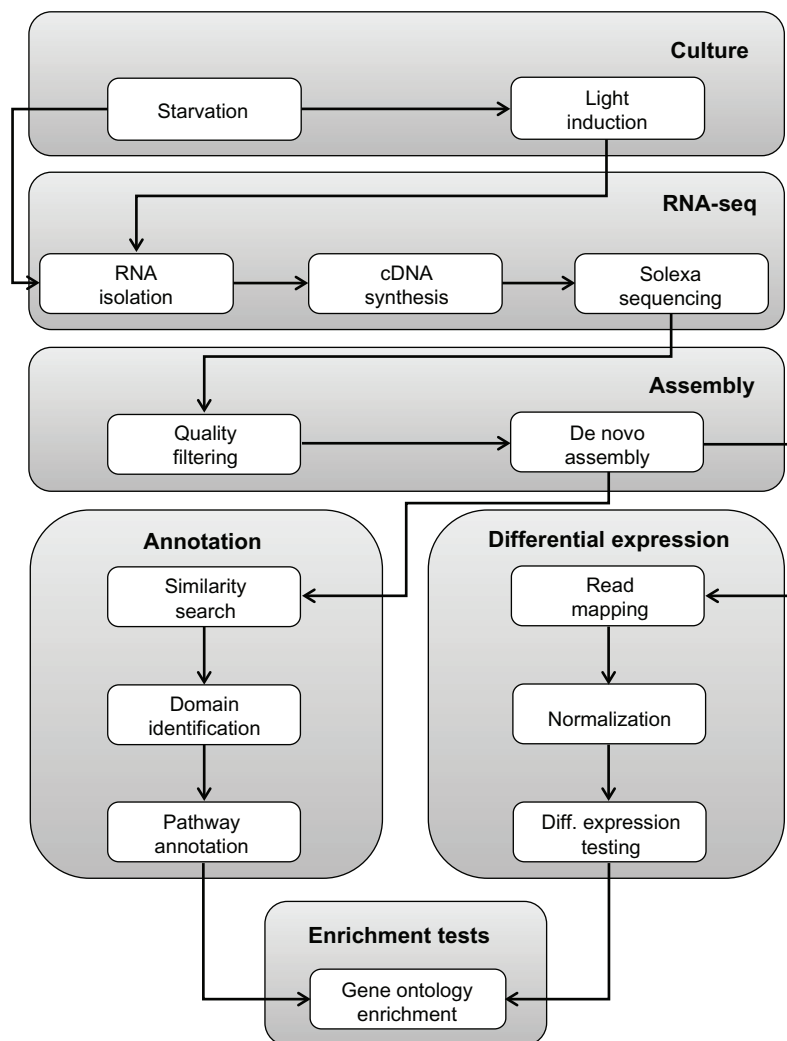


Figure 2. Overview of the experimental design.

Notes: A summary of experiments and computational analyses is shown. RNA samples were taken from competent and light-induced plasmodia (culture). These RNAs were employed as templates for cDNA synthesis, which were later sequenced using the Illumina HiSeq 2000 platform (RNA-seq). Reads were assembled (assembly), and the obtained contigs were annotated at every bioinformatic level (annotation). Then, to evaluate differentially expressed transcripts, reads were mapped to the assembly and normalized (differential expression). Finally, gene ontology annotations were tested for enrichment between up- and downregulated transcripts (enrichment tests).

genome of *Physarum*.¹³ Later, from the de novo assembly of the trimmed sequencing output, we obtained 909,505 sequences, with a N50 contig size of 371 bp. Large cDNAs from this assembly (>500 bp) were then clustered into 16,822 contigs (N50 length: 778 bp) with CAP3, to create a comprehensive set of representative transcripts. 10,278 of these contigs included transcript abundance data in the form of mapped reads in at least one cell sample, which was used as a measure of gene expression. Transcriptionally active genes were then defined as contigs with at least one mapped read present in all four samples or differentiation stages; in this regard, we observed 8,149 transcripts with mapped reads in all libraries.

In order to assess the reproducibility of this single-cell approach, we normalized the mapped read counts for each transcriptionally active transcript as reads per

kilobase per million mapped reads (RPKM).³⁰ We then compared the RPKM-normalized data for competent plasmodia (D1 and D2 cells) and photoinduced cells (L1 and L2) that were separately processed (Fig. 3). Accordingly, cells from related developmental stages exhibit very similar transcriptomic profiles (competent cells: $r = 0.99$; light-induced plasmodia: $r = 0.98$). On the other hand, lower correlations were found between competent and photoinduced cells ($r = 0.96$ between D2 and L1; $r = 0.97$ in all other cases; Fig. 3). However, further studies involving comparisons between several cell types and cell stages are required to establish whether these lower correlations account for the variations between individual cells.

Afterwards, we estimated the differentially expressed transcripts using DESeq³¹ from the raw mapped reads. For these analyses, only contigs with a combined count of 300 mapped reads among all the

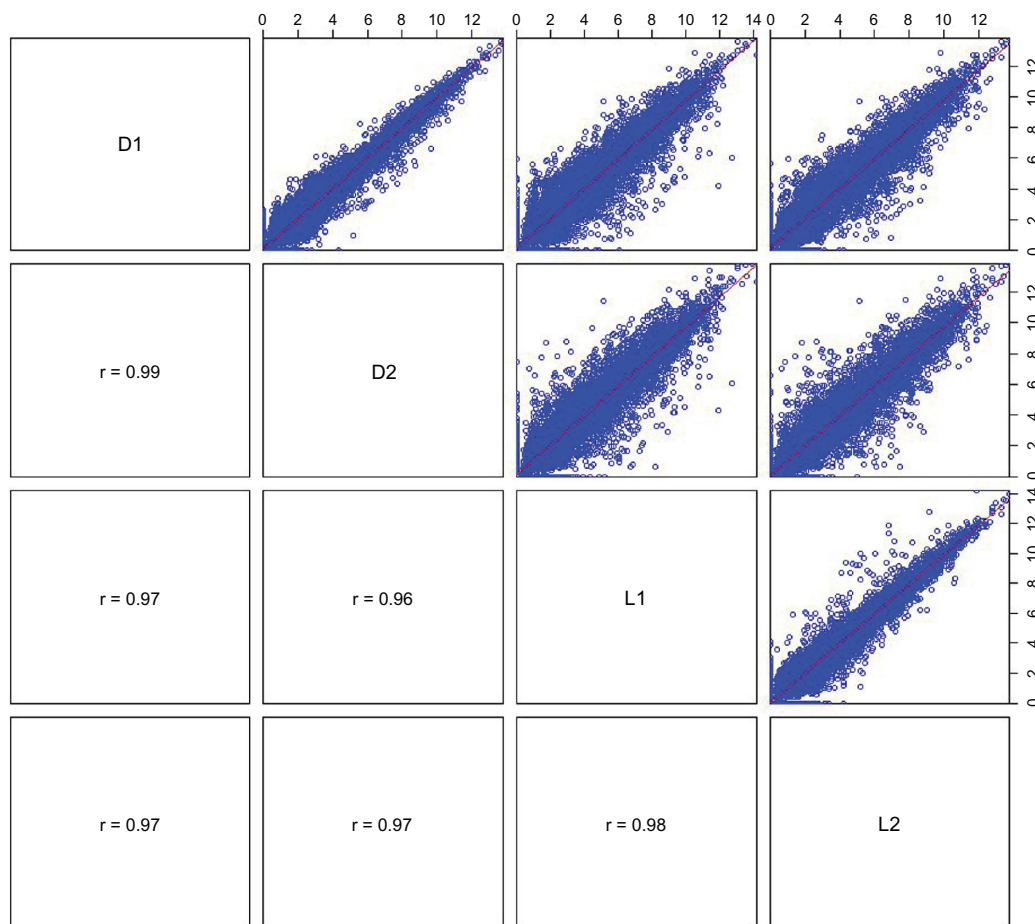


Figure 3. Assessment of the reproducibility of the approach.

Notes: Correlation plots of the RPKM-normalized reads³⁰ for competent (D1 and D2) and light-induced (L1 and L2) plasmodia were employed to assess the reproducibility of the single-cell RNA-seq in *Physarum*. Reads mapped to the novel transcriptomic assembly were used for plotting. Values of correlation coefficients (r) are shown in the corresponding boxes and the red lines indicate no fold changes in expression. Labels of both x - and y -axis are the Log_2 of the RPKM-normalized reads.

samples were considered; i.e., 3,164 contigs were then selected that fitted these criteria. This mapped read count threshold was selected to reduce noise caused by spurious contigs and alignments. Upon normalization, the distribution of mapped reads reflected the presence of differentially expressed transcripts, and genes with other kinds of regulation, with a slightly greater set of genes with higher expression in light-induced cells (Fig. 4). Specifically, we identified 556 upregulated transcripts (P -value < 0.05 ; 504 of these with a false discovery rate (FDR) < 0.1) and 531 downregulated transcripts (475 with FDR < 0.1) between the photoinduced and competent cell libraries for transcriptionally active contigs with mapped reads (Fig. 4).

Subsequently, to assign functions to the novel sequences, annotations were associated to the transcriptome assembly. In this way we obtained 92,641 Gene Ontology (GO) terms, corresponding to 5,722 SwissProt orthologs, where 64,730 GO terms belong to 4,222 sequences with mapped reads. cDNAs were linked biological processes ($n = 1,135$), molecular functions ($n = 1,558$), and cellular components ($n = 576$). From the transcriptionally active genes with mapped reads, 231 annotated transcripts were

upregulated, and 264 transcripts were downregulated (Supplementary Data). A comparison of GO annotations between sets of up- and downregulated transcripts revealed two terms exclusively featured in upregulated genes ('symplast,' GO:0055044; and 'auxiliary transport protein,' GO:0015457). Both annotations are related to the extracellular transport via pores. Conversely, six GO terms were identified only in downregulated transcripts ('synapse,' GO:0045202; 'synapse part,' GO:0044456; 'antioxidant activity,' GO:0016209; 'translation regulator activity,' GO:0045182; 'immune system process,' GO:0002376; and 'viral reproduction,' GO:0016032; Fig. 5). These groups of GO annotations are associated with the regulation of translation. Next, we tested for the enrichment of GO terms in up- and downregulated contigs against the full list of annotated transcripts, using the Fisher's exact test as implemented in Blast2GO.²⁶ In this manner, significant overrepresentation was found only in upregulated transcripts (FDR = 0.037; P -value = 0.046), with all GO terms belonging to the molecular function category of ontologies: metal ion binding (GO:0046872), calcium ion binding (GO:0005509), ion binding (GO:0043167), and cation binding (GO:0043169). All these belong to the same hierarchy of ontologies, so all these can be summarized with the lower and more specific category, i.e., the 'calcium ion binding' GO term. Both analyses of GO annotations correlate well with previous studies that point to the upregulation of genes associated with the ion transport in the light-induced plasmodium.^{13,20}

Later, to evaluate the genes with similar regulation, we clustered the fully annotated transcripts for the highest statistically significant up- and downregulation levels, as compared to the starved plasmodium cell libraries (Tables 2 and 3). In spite of the apparent diversity of the annotations, we inferred potential functions based on ortholog identities and gene ontology assignments. In this way, we identified upregulated transcripts encoding endopeptidases (PHYSA), phospholipases (PLDG) and stress response proteins (BPM1, NAH1), as well as genes related to biosynthetic processes (COAD, IOD1, PYR1), development (STX3), chromatin remodeling (YA27), and signaling (ARF1, CYH4, SAR1, LTBP2), that are highly expressed 6.5 hours after photoinduction (Table 2).

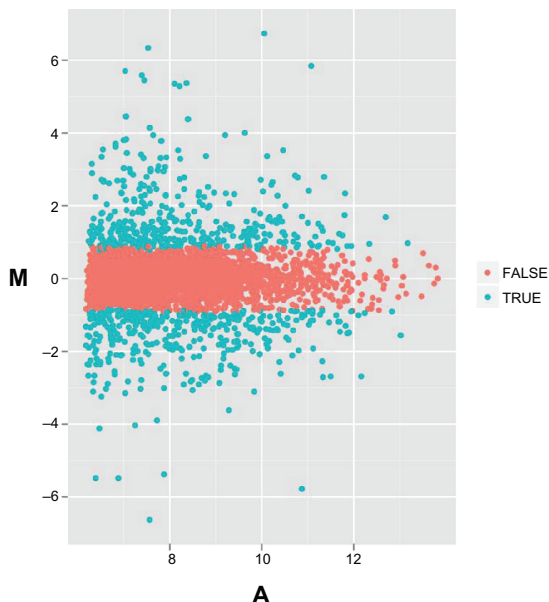


Figure 4. Fold change and significance.

Notes: Log_2 fold-changes (M) of normalized mapped reads are plotted on the y-axis, and Log_2 -normalized means (A) are plotted on the x-axis. Differentially expressed transcripts (turquoise points) were identified between photoinduced and starved single cells of *Physarum*, using DEseq³¹ (version 1.6.1, FDR < 0.05). Transcripts with high fold change may not be significantly differentially expressed simply due to high variance.

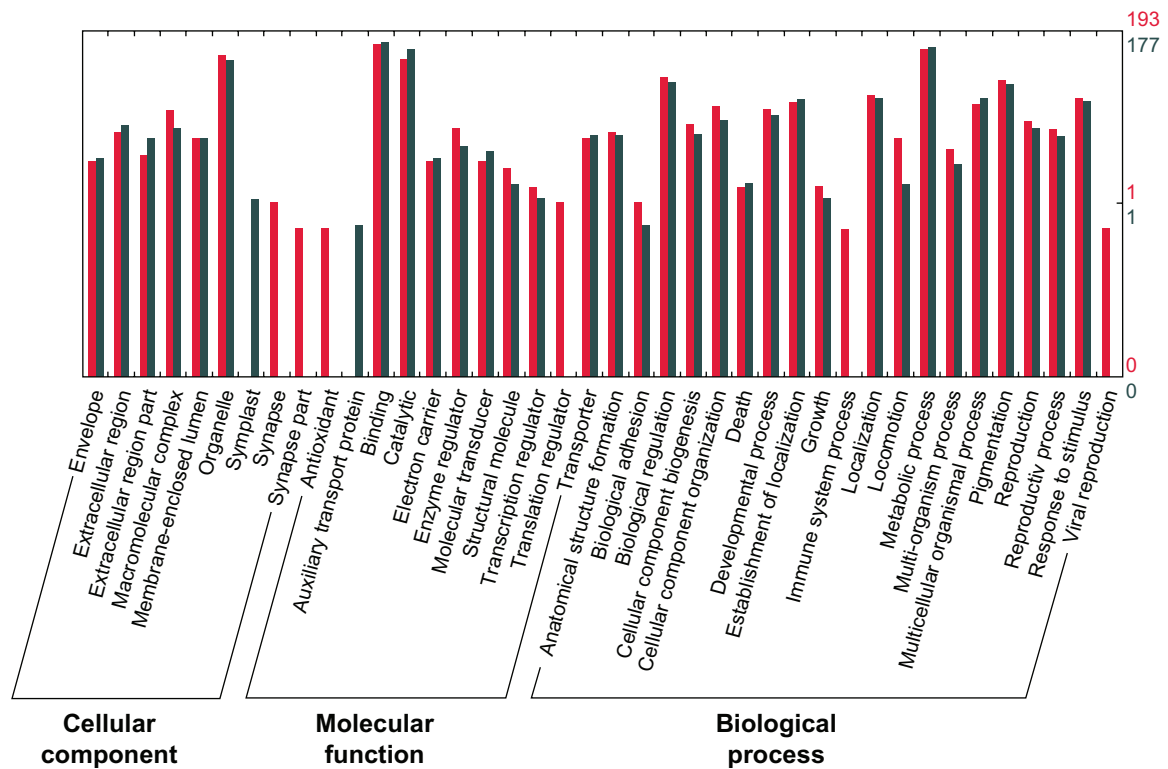


Figure 5. Gene Ontology (GO) classification of differentially expressed transcripts.

Notes: A comparison of the three main GO categories (biological process, molecular function, and cellular component) between the different expression groups using WEGO,³⁴ is shown. Up- and downregulated transcripts are indicated with dark green and red colors, respectively. The y-axis represents the number of transcripts for each GO category, plotted on a logarithmic scale.

Table 2. Top 20 annotated transcripts upregulated after photoinduction.

| Contig ID | UniProt | Annotation | D | L | Fold change | P-value |
|--------------|---------|---|-----|-------|-------------|----------|
| s432k3t27235 | O00909 | ADP-ribosylation factor 1, ARF1 | 43 | 3,475 | 106.3591 | 3.75E-50 |
| s431k3t6841 | Q8MZS4 | Physarolisin, PHYSA | 10 | 655 | 81.1591 | 6.79E-39 |
| s431k4t817 | P0CR31 | Small COPII coat GTPase SAR1, SAR1 | 161 | 7,079 | 57.2024 | 6.92E-47 |
| s431k4t520 | Q86AV9 | Phospholipase D, PLDG | 15 | 607 | 48.1411 | 1.39E-34 |
| s431k4t26745 | Q8MZS4 | Physarolisin, PHYSA | 32 | 1,043 | 39.0623 | 5.44E-37 |
| s422k4t53788 | Q8L765 | BTB/POZ and MATH domain-1, BPM1 | 67 | 1,057 | 20.8130 | 6.62E-28 |
| s422k4t53789 | Q8L765 | BTB/POZ and MATH domain-1, BPM1 | 44 | 589 | 17.6746 | 8.25E-24 |
| s431k3t89 | P49894 | Iodothyronine deiodinase, IOD1 | 201 | 2,474 | 16.1363 | 6.09E-24 |
| s431k3t4056 | Q20797 | Syntaxin-3, STX3 | 157 | 1,885 | 15.3925 | 2.41E-23 |
| s431k3t8494 | Q8L765 | BTB/POZ and MATH domain-1, BPM1 | 53 | 619 | 15.3912 | 1.31E-22 |
| s432k3t59570 | Q28019 | Latent-transforming growth factor beta-binding 2, LTBP2 | 38 | 459 | 14.1221 | 2.17E-22 |
| s432k4t17 | P20054 | PYR1-3 CAD homolog, PYR1 | 37 | 441 | 13.9181 | 3.32E-22 |
| s422k4t53790 | Q8L765 | BTB/POZ and MATH domain-1, BPM1 | 66 | 692 | 13.8263 | 4.95E-22 |
| s432k3t79109 | Q5UPW1 | F-box and FNIP repeat-protein, YR286 | 37 | 382 | 12.2992 | 1.83E-20 |
| s422k4t53791 | Q8L765 | BTB/POZ and MATH domain-1, BPM1 | 101 | 884 | 11.4944 | 2.26E-20 |
| s432k4t1366 | Q09698 | Uncharacterized C2F7.07c, YA27 | 493 | 4,849 | 11.4741 | 2.67E-22 |
| s431k3t50779 | Q99271 | Na ⁺ /H ⁺ antiporter, NAH1 | 48 | 410 | 10.9637 | 4.53E-18 |
| s431k4t11834 | Q80YW0 | Cytohesin-4, CYH4 | 425 | 3,365 | 10.3168 | 8.72E-19 |
| ctg9928 | P08955 | CAD protein, PYR1 | 90 | 737 | 9.9023 | 4.73E-19 |
| s431k4t20190 | C1DIB2 | Phosphopantetheine adenylyltransferase, COAD | 68 | 519 | 9.8979 | 4.5E-17 |

Notes: A list of transcripts with unambiguous annotations, significant differential expression ($P < 0.05$), and with the highest levels of upregulation between the competent and light-induced libraries, is shown. Sums of mapped reads (D: starved, L: photoinduced) and fold changes are indicated for each transcript on a given condition. Blast2GO²⁶ automatic annotations were used, and manual corrections of annotations were included in some cases.

**Table 3.** Top 20 annotated transcripts downregulated after photoinduction.

| Contig ID | UniProt | Annotation | D | L | Fold change | P-value |
|--------------|---------|---------------------------------------|--------|-------|-------------|----------|
| s424k4t102 | Q53G44 | Interferon-induced 44-like, IF44L | 792 | 7 | 0.0101 | 6.44E-36 |
| s422k3t11297 | Q6WP50 | Botrydial synthesis protein 2, BOT2 | 7,869 | 124 | 0.0183 | 1.80E-39 |
| s424k4t12560 | O74467 | SET domain-containing 5, SET5 | 616 | 30 | 0.0607 | 1.51E-23 |
| s424k3t14992 | O74467 | SET domain-containing 5, SET5 | 852 | 46 | 0.0672 | 2.04E-21 |
| s424k3t34437 | P29473 | Nitric oxide synthase, NOS | 504 | 48 | 0.1139 | 8.76E-17 |
| s424k3t13022 | Q60106 | Xanthomonalisin, XANP | 1,507 | 174 | 0.1337 | 5.27E-13 |
| s424k4t29841 | Q27597 | NADPH-cytochrome P450 reductase, NCPR | 380 | 45 | 0.1352 | 2.49E-14 |
| ctg353 | O61063 | Tectonin-1, TCT1 | 2,055 | 243 | 0.1365 | 3.61E-13 |
| ctg4361 | P08799 | Myosin-2 heavy chain, MYS2 | 311 | 39 | 0.1383 | 7.02E-13 |
| s424k3t6526 | Q9EST2 | Histamine N-methyltransferase, HMNT | 1,322 | 165 | 0.1411 | 2.54E-12 |
| ctg5105 | O01840 | Peptide transporter 3, PEPT3 | 1,646 | 199 | 0.1427 | 1.28E-12 |
| s424k4t30424 | Q27597 | NADPH-cytochrome P450 reductase, NCPR | 1,344 | 166 | 0.1449 | 2.12E-12 |
| s424k4t31287 | Q60106 | Xanthomonalisin, XANP | 1,584 | 208 | 0.1504 | 7.97E-12 |
| s424k4t4957 | Q8T8C0 | Nitric oxide synthase, NOS | 1,535 | 195 | 0.1513 | 4.38E-12 |
| s424k4t2011 | Q6B9X6 | Alpha-protein kinase, VWKA | 9,604 | 1,216 | 0.1532 | 1.56E-13 |
| s424k4t319 | Q03380 | Comitin, COMA | 17,238 | 2,208 | 0.1552 | 9.65E-13 |
| s422k4t3112 | Q03380 | Comitin, COMA | 1,664 | 217 | 0.1570 | 6.19E-12 |
| s422k4t11955 | P08799 | Myosin-2 heavy chain, MYS2 | 311 | 44 | 0.1577 | 8.38E-12 |
| s422k3t4555 | Q03380 | Comitin, COMA | 1,223 | 163 | 0.1634 | 8.55E-12 |
| s422k3t4907 | Q6B9X6 | Alpha-protein kinase, VWKA | 5,034 | 683 | 0.1644 | 8.63E-13 |

Notes: Transcripts with unambiguous annotations, significant differential expression (P -value < 0.05), with mapped reads in all libraries, and that possess the highest levels of downregulation between the competent and light-induced samples, are listed. Annotations, UniProt accessions, mapped reads, and probability values follow the same convention as in Table 2.

On the other hand, a different group of genes is downregulated upon light exposure. In this case, we found transcripts associated with actin-binding (MYS2, COMA), FMN-binding (NOS, NCPR), signaling (VWKA), sugar- (TCT1) and cation-binding proteins (XANP, BOT2), a transporter (PEP3) and transferases (SET5, HMNT), that were downregulated 6.5 hours after light induction (Table 3). These measurements of transcriptional regulation at different time points correlate well with previous results in cell pools, where actin-binding and signaling proteins were identified as core members of the regulatory network during sporulation.¹³

We also noticed the expression of multiple transcript isoforms in the same cell at the same time point, both in upregulated transcripts (PYR1, BPM1, PHYSA; Table 2), and in downregulated transcripts (COMA, VWKA, NOS, NCPR, XANP, MYS2, SET5; Table 3). This phenomenon has also been observed in previous single-cell studies, and has been attributed to the complexity of transcript variants.³³ Whether these genes encode isoforms controlling stage-specific signaling pathways remains to be studied in detail.

To date, two studies have reported the RNA-seq analysis of transcriptomes in eukaryotic organisms, using

single embryonic cells as models.^{17,18} In these works, both the assembly and mapping procedures were achieved using the mouse genome as a reference. Previously, we succeeded in measuring the expression on a set of 35 differentially regulated genes in populations of single-cells, using the GeXP RT-PCR platform.²⁰ Here, using the power of RNA-seq to obtain whole transcriptomes without relying on previous genomic information, we were able to characterize a large set of expressed genes in different samples during the sporulation of *Physarum*, an organism without a known genomic sequence. Furthermore, in order to obtain single cells, all former studies on single-cell multiplex gene expression analysis required complex separation methods, such as pipetting cells manually, or using laser microdissection or fluorescence-activated cell sorting.¹⁵ In this study, we used the plasmodium, a natural macroscopic multinucleate single-cell stage from *Physarum*, whose culture and handling is straightforward, and for which there are several well established methods for genetic manipulation.³

Conclusions

By combining the power of next-generation sequencing technologies with the simplicity for obtaining



single cells from *Physarum*, we developed an approach to characterize the whole transcriptome through the differentiation of this lower eukaryote at the single-cell level. The observed regulation patterns correlate well with previous studies on differential gene expression during the commitment to sporulation in the slime mold, particularly with respect to proteins involved in signaling and actin binding. We expect that improvements in single-cell transcriptomics, such as discrimination in sense and antisense transcripts, the ability to sequence a more diverse range of nucleic acid species, and other future developments, will help us to display a more precise picture of the regulatory network controlling differentiation in this organism.

Acknowledgements

Authors thank Jan-Thierry Wegener (Otto von Guericke University) for his help with computational issues.

Funding Sources

This research was partially supported by the German Federal Ministry of Education and Research (FORSYS—MaCS, Magdeburg Centre for Systems Biology). IB is an International Max Planck Research School—Otto von Guericke University fellow.

Competing Interests

WM's institution has received grants and study support from the German Federal Ministry of Education and Research. Other authors disclose no competing interests.

Author Contributions

Conceived and designed the research: WM, IB. Analyzed the data: IB, JL. Wrote the first draft of the manuscript: IB. Contributed to the writing of the manuscript: WM, JL. Agree with manuscript results and conclusions: WM, JL. Jointly developed the structure and arguments for the paper: WM, IB. All authors reviewed and approved of the final manuscript.

Disclosures

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and

confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. The external blind peer reviewers report no conflicts of interest.

References

1. Nachmias VT, Huxley HE. Electron microscope observations on actomyosin and actin preparations from *Physarum polycephalum*, and on their interaction with heavy meromyosin subfragment I from muscle myosin. *J Mol Biol.* 1970;50(1):83–90.
2. Sauer HW, Babcock KL, Rusch HP. Sporulation in *Physarum polycephalum*: a model system for studies on differentiation. *Exp Cell Res.* 1969;57(2):319–27.
3. Marwan W. Detecting functional interactions in a gene and signaling network by time-resolved somatic complementation analysis. *Bioessays.* 2003;25(10):950–60.
4. Marwan W, Sujatha A, Starostzik C. Reconstructing the regulatory network controlling commitment and sporulation in *Physarum polycephalum* based on hierarchical Petri Net modelling and simulation. *J Theor Biol.* 2005;236(4):349–65.
5. Bundschuh R, Altmüller J, Becker C, Nürnberg P, Gott JM. Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. *Nucleic Acids Res.* 2011;39(14):6044–55.
6. Bénard M, Maric C, Pierron G. Low rate of replication fork progression lengthens the replication timing of a locus containing an early firing origin. *Nucleic Acids Res.* 2007;35(17):5763–74.
7. Tero A, Takagi S, Saigusa T, et al. Rules for biologically inspired adaptive network design. *Science.* 2010;327(5964):439–42.
8. Saigusa T, Tero A, Nakagaki T, Kuramoto Y. Amoebae anticipate periodic events. *Phys Rev Lett.* 2008;100(1):018101.
9. Burland TG, Solnica-Krezel L, Bailey J, Cunningham DB, Dove WF. Patterns of inheritance, development and the mitotic cycle in the protist *Physarum polycephalum*. *Adv Microb Physiol.* 1993;35:1–69.
10. Beach D, Piper M, Shall S. Isolation of newly-initiated DNA from the early S phase of the synchronous eukaryote, *Physarum polycephalum*. *Exp Cell Res.* 1980;129(1):211–21.
11. Guttus E, Guttus S. Chapter 3. Mitotic Synchrony in the Plasmodia of *Physarum polycephalum* and Mitotic Synchronization by Coalescence of Microplasmodia. *Methods Cell Physiology Volume 1.* 1964;New York: Academic Press.
12. Glöckner G, Golderer G, Werner-Felmayer G, Meyer S, Marwan W. A first glimpse at the transcriptome of *Physarum polycephalum*. *BMC Genomics.* 2008;9:6.
13. Barrantes I, Glöckner G, Meyer S, Marwan W. Transcriptomic changes arising during light-induced sporulation in *Physarum polycephalum*. *BMC Genomics.* 2010;11:115.
14. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* 2010;28(6):281–90.
15. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. *Nat Methods.* 2011;8(Suppl 3):S6–11.
16. Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 2008;320(5881):1344–9.
17. Tang F, Barbacioru C, Bao S, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell.* 2010;6(5):468–78.



18. Islam S, Kjällquist U, Moliner A, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 2011;21(7):1160–7.
19. Starostzik C, Marwan W. Kinetic analysis of a signal-transduction pathway by time-resolved somatic complementation of mutants. *J Exp Biol.* 1998;201(Pt 13):1991–9.
20. Hoffmann XK, Tesmer J, Souquet M, Marwan W. Futile attempts to differentiate provide molecular evidence for individual differences within a population of cells during cellular reprogramming. *FEMS Microbiol Lett.* 2012;329(1):78–86.
21. Starostzik C, Marwan W. Functional mapping of the branched signal transduction pathway that controls sporulation in *Physarum polycephalum*. *Photochem Photobiol.* 1995;62(5):930–3.
22. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18(5):821–9.
23. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res.* 1999;9(9):868–77.
24. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
25. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 2003;31(1):365–70.
26. Götz S, García-Gómez JM, Terol J, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36(10):3420–35.
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
28. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
29. Milne I, Bayer M, Cardle L, et al. Tablet—next generation sequence assembly visualization. *Bioinformatics.* 2010;26(3):401–2.
30. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621–8.
31. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.
32. Leinonen R, Akhtar R, Birney E, et al. The European Nucleotide Archive. *Nucleic Acids Res.* 2011;39(Database issue):D28–31.
33. Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods.* 2009;6(5):377–82.
34. Ye J, Fang L, Zheng H, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 2006;34(Web Server issue):W293–7.
35. Carlile MJ, Watkinson SC, Gooday GW. *The Fungi*, 2nd. Ed. 2001; London: Academic Press.



Supplementary Data

- Transcript Abundance Data and Annotation (“Supplementary.xls”). Annotations and mapped reads for all the assembled cDNAs (Excel spreadsheet).