

# Single-cell transcriptome analysis of *Physcomitrella* leaf cells during reprogramming using microcapillary manipulation

Minoru Kubo<sup>1,\*</sup>, Tomoaki Nishiyama<sup>2</sup>, Yosuke Tamada<sup>3,4</sup>, Ryosuke Sano<sup>5</sup>, Masaki Ishikawa<sup>3,4</sup>, Takashi Murata<sup>3,4</sup>, Akihiro Imai<sup>6</sup>, Daniel Lang<sup>7</sup>, Taku Demura<sup>5</sup>, Ralf Reski<sup>7,8,\*</sup> and Mitsuyasu Hasebe<sup>3,4,\*</sup>

<sup>1</sup>Institute for Research Initiative, Nara Institute of Science and Technology, Ikoma 630-0192, Japan, <sup>2</sup>Advanced Science Research Center, Kanazawa University, Kanazawa 920-0934, Japan, <sup>3</sup>National Institute for Basic Biology, Okazaki 444-8585, Japan, <sup>4</sup>School of Life Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan, <sup>5</sup>Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma 630-0192, Japan, <sup>6</sup>Faculty of Life Sciences, Hiroshima Institute of Technology, Hiroshima 731-5193, Japan, <sup>7</sup>Plant Biotechnology, Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany and <sup>8</sup>Signaling Research Centres BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany

Received December 09, 2018; Revised March 01, 2019; Editorial Decision March 04, 2019; Accepted March 08, 2019

## ABSTRACT

Next-generation sequencing technologies have made it possible to carry out transcriptome analysis at the single-cell level. Single-cell RNA-sequencing (scRNA-seq) data provide insights into cellular dynamics, including intercellular heterogeneity as well as inter- and intra-cellular fluctuations in gene expression that cannot be studied using populations of cells. The utilization of scRNA-seq is, however, restricted to cell types that can be isolated from their original tissues, and it can be difficult to obtain precise positional information for these cells *in situ*. Here, we established single cell-digital gene expression (1cell-DGE), a method of scRNA-seq that uses micromanipulation to extract the contents of individual living cells in intact tissue while recording their positional information. With 1cell-DGE, we could detect differentially expressed genes (DEGs) during the reprogramming of leaf cells of the moss *Physcomitrella patens*, identifying 6382 DEGs between cells at 0 and 24 h after excision. Furthermore, we identified a subpopulation of reprogramming cells based on their pseudotimes, which were calculated using transcriptome profiles at 24 h. 1cell-DGE with microcapillary manipulation can be used to analyze the gene expression of individual cells

without detaching them from their tightly associated tissues, enabling us to retain positional information and investigate cell–cell interactions.

## INTRODUCTION

Whole plants can be regenerated from tissue samples such as branch cuttings or detached leaves via a callus (1), a mass of undifferentiated cells able to initiate shoot- and root-stem cells in the presence of the appropriate phytohormones (2). Several genes have been shown to function in regenerating cells (3,4); however, the elucidation of the transcriptome profiles involved in the regeneration process of each cell is a major challenge as it is not currently possible to separate and identify the limited numbers of stem cells that randomly emerge in a callus during regeneration.

A number of single-cell RNA-sequencing (scRNA-seq) methods utilizing next generation sequencing (NGS) have been developed to prepare cDNA libraries from isolated single cells containing trace amounts of RNA (5,6). Hundreds to thousands of single isolated cells derived from the human fetal cortex or mouse retinas have been simultaneously prepared into sequencing libraries using automated single-cell preparation systems such as Fluidigm C1 (7) and inDrop (8,9). By assessing the heterogeneity of expression profiles between individual cells in a population, including rare cell types, biological events such as different cell cycle stages and transcription bursts have been identified, revealing the trajectories of developmental cell states that were

\*To whom correspondence should be addressed. Tel: +81 743 72 6158; Fax: +81 743 72 5469; Email: ku-bo@bs.naist.jp  
Correspondence may also be addressed to Ralf Reski. Tel: +49 761 203 6969; Fax: +49 761 203 6967; Email: ralf.reski@biologie.uni-freiburg.de  
Correspondence may also be addressed to Mitsuyasu Hasebe. Tel: +81 564 55 7546; Fax: +81 564 55 7548; Email: mhasebe@nibb.ac.jp  
Present address: Daniel Lang, Plant Genome and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany.

not previously detectable in transcriptome analyses of samples containing multiple cells (10,11). In plants, single-cell transcriptome analyses of root cells in the flowering plant *Arabidopsis thaliana* (Arabidopsis) revealed a transition of cell identity during root regeneration (12–14). scRNA-seq has great potential for providing new biological insights into regeneration; however, using the methods described above, the positional information of the cells within their tissue is lost during the isolation process. Furthermore, it can be difficult to detach single cells from the tissues and organs of many plant species because their cell walls consisting of carbohydrate and proteoglycan polymers strongly adhere to each other.

The moss *Physcomitrella patens* (*Physcomitrella*) is a basal land plant with a simple body plan, including leaves formed of a single cell layer (15), which facilitates its observation and manipulation at the cellular level (16,17). When a *Physcomitrella* leaf is cut, some of the cells facing the cut change into chloronema apical stem cells without the addition of exogenous plant hormones, enabling the entire moss body to be regenerated (18). Several genes involved in this reprogramming have been characterized. Cyclin-dependent kinase A (PpCDKA) and cyclin D (PpCYCD;1) regulate the reentry into the cell cycle (18). The *WUSCHEL*-related *homeobox 13* (*PpWOX13*) genes are upregulated during reprogramming and required for the tip growth characteristic of the chloronema apical stem cells (19). The Cold-Shock Domain Protein 1 (PpCSP1) and PpCSP2, orthologous to the mammalian reprogramming factor Lin28A, also positively regulate reprogramming in *Physcomitrella* (20). Furthermore, a transcriptome analysis of whole excised leaves during reprogramming revealed that the expression levels of more than 3900 genes were altered within 24 hours after excision (21).

When *Physcomitrella* leaves are excised, only some of the leaf cells facing the cut are reprogrammed, while other cells neighboring the cut, as well as the intact cells that do not face the cut, are not reprogrammed (18). It is therefore difficult to distinguish between genes specifically expressed in the reprogramming cells and those expressed in non-reprogramming cells. Understanding the *in situ* regulation of reprogramming in an excised leaf is a challenge; when two neighboring leaf cells are isolated together, only one is reprogrammed, even though almost all cells isolated on their own can autonomously reprogram into protonema apical cells (22). This suggests the presence of cell–cell interactions between neighboring cells during reprogramming; however, the molecules and genes responsible for this mechanism have not been identified, partially because of the difficulty in isolating a single cell to investigate its transcriptome during the reprogramming process. When a pair of adjacent cells are isolated, both show features of the early phases of reprogramming, such as nuclear expansion and the expression of cell cycle-related genes; however, these become diminished in the non-reprogrammed cell (22). This suggests that the reprogrammed cells not only inhibit reprogramming in their neighbors, but that they actively revert their neighboring cells back to a leaf cell state. Although this is a good model for studying cell–cell interactions during reprogramming, it has meant that the mechanisms by which stem cells are determined and the factors involved in the in-

hibitory effect of the reprogrammed cells on their neighbors are poorly understood.

To explore the genes involved in cell–cell interactions of reprogramming in *Physcomitrella* leaves, we established a single cell transcriptome analysis method using microcapillary manipulation to physically extract the contents of individual living cells within a tissue and prepare a cDNA library of their trace amounts of RNA. We also introduced a unique molecular identifier (UMI) (23) to the cDNAs to reduce the amplification bias when using PCR.

## MATERIALS AND METHODS

### Plant materials and growth conditions

The wild-type moss *Physcomitrella patens* Gransden 2004 (24) and the transgenic *Physcomitrella* line GX8-NGG (25) were used for the total RNA extractions and the preparation of excised leaves, respectively. To propagate the gametophores, a small portion of GX8-NGG protonema was inoculated on BCDAT agar medium (26) and cultured in a growth chamber (MLR-352H; Panasonic, Tokyo, Japan) under 20–70  $\mu\text{mol}/\text{m}^2/\text{s}$  of continuous white light and 55% relative humidity at 23°C.

### Preparation of excised leaves

Gametophores were cultured for 21 days after inoculation on BCDAT medium, after which the distal half of the third leaf was cleanly cut with a razor blade, placed onto the BCDAT medium and covered with cellophane. The majority of the excised leaf, except for the living leaf cells facing the cut edge, was covered with additional layers of cellophane. Dishes containing the excised leaves were sealed with Parafilm and incubated under continuous white light at 23°C until the cell contents were extracted. For the sampling at 0 h, the cell contents were extracted within 35 min of leaf excision.

### Micromanipulation to extract cell contents

A set of oil hydraulic micromanipulators (MMO-220A and MMO-202ND; Narishige, Tokyo, Japan) and motor-driven manipulators MM-89 (Narishige) were equipped onto an inverted fluorescent microscope IX-70 (Olympus, Tokyo, Japan) with a fluorescence filter unit (U-MWIB3; excitation: 460–495 nm, emission: 510IF; dichromatic mirror: 505 nm; Olympus). To simultaneously observe the tip of the microcapillary and the GFP fluorescence in the nuclei of the leaf cells, the fluorescence microscopy was performed under dim bright-field illumination. The 1.0-mm capillary holder was connected to a microinjector (CellTram vario; Eppendorf, Hamburg, Germany) via a silicone tube filled with mineral oil (M-8410; Sigma-Aldrich, St. Louis, MO, USA), which was in turn attached to the MMO-220A micromanipulator. The parameters of the 1.0-mm glass capillaries were as follows: inner diameter: 20  $\mu\text{m}$ , pipette form: straight, beveled angle of tip: 40°, and pipette length: 55 mm (BioMedical Instruments, Zöllnitz, Germany). The bottom of the glass capillary contained a small amount of cell content extraction mix2 comprising 13% 10  $\times$  PCR buffer II (Thermo Fisher Scientific, Waltham, MA, USA), 7.8% 25

mM MgCl<sub>2</sub>, 6.5% 0.1 M DTT, 2.6% RNasin Plus RNase inhibitor (Promega, Madison, WI, USA), and 2.6% of a mix containing 2.5 mM of each dNTP (Takara Bio, Kusatsu, Japan). The capillary was attached to the capillary holder so that the beveled tip faced down without any air bubbles.

The attached microcapillary was gently filled with mineral oil under a microscope using the CellTram vario. After adjusting the tip position of the glass capillary to the center of the observation field, a dish containing an excised leaf was set on the microscope and the tip of the capillary was used to extract the nucleus and some surrounding cytoplasm from the target cell. The cell contents were immediately transferred into a 0.2-ml PCR tube containing 1.25  $\mu$ l RT oligos (0.05  $\mu$ M) and 2.35  $\mu$ l cell content extraction mix1, containing 0.45  $\mu$ l 10 $\times$  PCR buffer II (Thermo Fisher Scientific), 0.27  $\mu$ l MgCl<sub>2</sub> (25 mM), 0.225  $\mu$ l DTT (0.1 M), 0.09  $\mu$ l RNasin Plus RNase inhibitor (Promega), 0.09  $\mu$ l of a mix containing 2.5 mM of each dNTP (Takara Bio), and 0.1  $\mu$ l of 20 000-fold diluted ERCC RNA spike-in mix (Thermo Fisher Scientific). After a brief centrifugation, the samples were primed with an incubation on a thermal cycler at 70°C for 90 s, 35°C for 15 s, and cooled to 4°C. The tubes were kept on ice before reverse transcription. The RT oligos used in this work are listed in Supplementary Table S1. Extracted nuclear conditions were categorized into one of six sample quality classes: broken, damaged; broken, average quality; broken, good quality; broken, very good quality; intact, good quality; intact, very good quality.

### Preparation of cDNA libraries for 1cell-DGE

For the reverse transcription, a 0.9  $\mu$ l RT mix containing 0.33  $\mu$ l SuperScriptIII reverse transcriptase (Thermo Fisher Scientific), 0.05  $\mu$ l RNasin plus RNase Inhibitor (Promega), and 0.07  $\mu$ l T4 gene 32 protein (New England Biolabs, Ipswich, MA, USA) was added to each primed RNA solution. After pipetting gently and centrifuging briefly, the tubes were incubated on a thermal cycler at 50°C for 30 min, 70°C for 10 min, then cooled to 4°C.

To digest the excess RT oligos, the samples were mixed with 0.8  $\mu$ l nuclease-free water (Qiagen, Hilden, Germany), 0.1  $\mu$ l 10 $\times$  exonuclease I buffer, and 0.1  $\mu$ l of 20 U/ $\mu$ l exonuclease I (New England Biolabs). After pipetting gently to mix and a brief centrifugation, the tubes were incubated on a thermal cycler using the following conditions: 4°C for 30 s, 37°C for 30 min, 80°C for 20 min with lid heating at 90°C, and cooled to 4°C. The tubes were then transferred onto ice for at least 1 min.

The poly(dA) mix for poly(dA) tailing with RNaseH was as follows: 4.44  $\mu$ l nuclease-free water (Qiagen), 0.6  $\mu$ l 10 $\times$  PCR buffer II (Thermo Fisher Scientific), 0.36  $\mu$ l MgCl<sub>2</sub> (25 mM), 0.18  $\mu$ l dATP (100 mM) (New England Biolabs), 0.3  $\mu$ l of 15 U/ $\mu$ l terminal deoxynucleotidyl transferase (Thermo Fisher Scientific), and 0.12  $\mu$ l of 5 U/ $\mu$ l RNaseH (New England Biolabs). A 6- $\mu$ l aliquot of this poly(dA) mix was added to each tube after the exonuclease I treatment. After pipetting to mix and a brief centrifugation, the samples were incubated on a thermal cycler using the following conditions: 4°C for 30 s, 37°C for 1.5 min, 70°C for 10 min with lid heating at 80°C, and cooled to 4°C.

For the second-strand synthesis, the following PCR mix1 was prepared: 50.68  $\mu$ l nuclease-free water (Qiagen), 15.2  $\mu$ l 5 $\times$  Q5 reaction buffer with MgCl<sub>2</sub> (New England Biolabs), 7.6  $\mu$ l of each dNTP (2.5 mM) (Takara Bio), 0.76  $\mu$ l NUP3 primer (100  $\mu$ M), and 1.76  $\mu$ l Q5 Hot Start High-Fidelity DNA polymerase (2 U/ $\mu$ l) (New England Biolabs). A 76- $\mu$ l volume of PCR mix1 was added into each tube after the poly(dA) tailing, pipetted to mix and briefly centrifuged, then the mixtures were divided into 21- $\mu$ l aliquots which were transferred into four new 0.2-ml PCR tubes. After centrifuging briefly, the tubes were incubated on a thermal cycler in the following conditions: 95°C for 3 min, 98°C for 20 s, 50°C for 2 min, 72°C for 10 min, then cooled to 4°C.

For the cDNA amplification, PCR mix2 was prepared, containing 12.73  $\mu$ l nuclease-free water (Qiagen), 3.8  $\mu$ l 5 $\times$  Q5 reaction buffer with MgCl<sub>2</sub> (New England Biolabs), 1.9  $\mu$ l of each dNTP (2.5 mM) (Takara Bio), 0.19  $\mu$ l BTEP7v2 primer (100  $\mu$ M), and 0.38  $\mu$ l Q5 Hot Start High-Fidelity DNA polymerase (2 U/ $\mu$ l) (New England Biolabs). A 19- $\mu$ l volume of PCR mix2 was added to each tube after the second-strand synthesis, pipetted to mix and briefly centrifuged. The tubes were incubated on a thermal cycler using the following conditions: an initial denaturation at 95°C for 3 min; followed by 22 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 6 min, which extended by 6 s at 72°C in each cycle; and stored at 4°C.

After the PCR amplification, the cDNA libraries were purified using a Purelink PCR purification kit with Binding Buffer High-Cutoff (Thermo Fisher Scientific), according to the manufacturer's instructions. To check availability of each sample, the quantity and quality of the cDNA libraries were measured using a Bioanalyzer 2100 with a High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA). Each cDNA library solution was placed in a 1.5-ml DNA Lo-bind tube (Eppendorf) and adjusted to a volume of 35  $\mu$ l with elution buffer (EB) containing 10 mM Tris-HCl (pH 8.0). To remove the byproducts in the cDNA libraries, a 0.55 $\times$  volume of SPRIselect beads (Beckman Coulter, Brea, CA, USA) were added to each cDNA library solution, which adhered the appropriately sized cDNAs. The tubes were placed on a Magna stand (Nippon Genetics, Tokyo, Japan) for 3 min, and the beads collected at the bottom of the tube. The supernatants were gently removed by aspiration, after which the beads were rinsed twice with 80% ethanol. After air-drying for 10 min, the beads were resuspended in 50  $\mu$ l EB then left to stand on the Magna stand for 3 min. The resulting supernatants were recovered into new 1.5-ml DNA Lo-bind tubes. Purification with the SPRIselect beads was carried out at least three times. The quantity and quality of the purified cDNA libraries were measured using a Bioanalyzer 2100 with a High Sensitivity DNA kit (Agilent Technologies), and the purified cDNA libraries were stored at -30°C until required. The oligo DNAs used in this work are listed in Supplementary Table S1.

### Bulk treatment for NGS library construction

In order to fragment the cDNAs to construct the NGS libraries, 2.5 nmol each of four or five purified cDNA libraries were combined, and the volume of the resulting solution was increased to 75  $\mu$ l with EB. The mixtures

were transferred into microTUBE AFA Fiber Pre-Snap-Cap tubes (Covaris, Woburn, MA, USA) and care was taken to prevent any air bubbles. cDNA shearing with a target peak of ~400 bp was carried out using an acoustic solubilizer, Covaris S2 (Covaris), under the following conditions; bath temperature: 4–8°C, degassing mode: continuous, power mode: frequency sweeping, duty cycle: 10%, intensity: 3, cycles/burst: 200, and time: 90 s. After this treatment, the fragmented cDNAs were transferred to new 1.5-ml DNA Lo-bind tubes (Eppendorf) and purified with a MinElute PCR purification kit (Qiagen). The quality of the fragmented cDNA was measured using a Bioanalyzer 2100 with a High Sensitivity DNA kit (Agilent Technologies).

To recover the fragmented cDNAs tagged with biotin, 20  $\mu$ l of streptavidin-linked beads, Dynabeads MyOne C1 (Thermo Fisher Scientific), were rinsed twice with 2  $\times$  BWT buffer containing 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, 2 M NaCl, and 0.02% Tween-20, then suspended in an equal volume of the fragmented cDNA solution. The solutions were left to stand for 10 min to bind the biotinylated cDNA fragments, then placed on a Magna stand for 30 s. The supernatants were discarded and the beads were rinsed three times with 1  $\times$  BWT buffer and resuspended in 25  $\mu$ l EBT buffer, which contained 10 mM Tris–HCl (pH 8.5) and 0.02% Tween-20.

For the end repair, a 25- $\mu$ l mixture containing 5  $\mu$ l 10  $\times$  NEBnext End Repair reaction buffer (New England Biolabs) and 2.5  $\mu$ l of NEBnext End Repair Enzyme Mix (New England Biolabs) was added to the mixture of beads and cDNA fragments. The solution was mixed by gently pipetting and centrifuging briefly, then incubated at room temperature with shaking at 400 rpm for 30 min. The tubes were stood on a Magna stand for 30 s and the supernatants were discarded. The beads were then rinsed twice with EBT buffer while on the Magna stand, after which the stand was removed to enable the beads to be resuspended in 21  $\mu$ l EBT buffer.

For the dA-tailing, a commercial reagent kit, NEBNext dA-tailing Module (New England Biolabs), was utilized. A 4- $\mu$ l aliquot of a mixture containing 2.5  $\mu$ l 10  $\times$  NEBnext dA-Tailing reaction buffer (New England Biolabs) and 1.5  $\mu$ l of Klenow Fragment (3'  $\rightarrow$  5' exo-) (New England Biolabs) was added to the solution of beads with end-repaired cDNA fragments. The mixtures were pipetted gently to mix and briefly centrifuged before being incubated at 37°C with shaking at 400 rpm for 30 min. To remove the reaction mix, the tubes were stood on a Magna stand for 30 s and the supernatants were discarded. The beads were rinsed twice with EBT buffer and resuspended in 25  $\mu$ l EBT buffer.

To ligate the adapters to the cDNA, a 25- $\mu$ l mixture containing 5  $\mu$ l 10  $\times$  T4 DNA ligase buffer (New England Biolabs), 1.5  $\mu$ l RP1 adaptor v2 (100  $\mu$ M), and 5  $\mu$ l of 400 U/ $\mu$ l T4 DNA ligase (New England Biolabs) was added to the solution containing the beads and dA-tailed cDNA fragments. The solutions were mixed by gently pipetting then centrifuged briefly, after which they were incubated at 20°C with shaking at 400 rpm for 20 min. A 5- $\mu$ l aliquot of 1 U/ $\mu$ l USER enzyme mix (New England Biolabs) was added to each tube, pipetted gently to mix, then incubated at 37°C with shaking at 400 rpm for 60 min. To remove the reaction mix, the tubes were stood on a Magna stand for 30 s and the

supernatants were discarded. The beads were rinsed twice with EBT buffer and resuspended in 25  $\mu$ l EBT buffer.

To fill the 5' overhang in the cDNA, a 5- $\mu$ l mixture containing 3  $\mu$ l 10 $\times$  NEB buffer 2 (New England Biolabs), 1  $\mu$ l of a mix containing 2.5 mM of each dNTP (Takara Bio), and 1  $\mu$ l of 10 U/ $\mu$ l DNA pol I (New England Biolabs) was added to the solution of adaptor-ligated cDNA fragments and beads. After pipetting gently to mix and centrifuging briefly, the tubes were incubated at 37°C with shaking at 400 rpm for 30 min. To remove the reaction mix, the tubes were stood on a Magna stand for 30 s and the supernatants were discarded. The beads were then rinsed twice with EBT buffer and resuspended in 25  $\mu$ l EBT buffer.

For the library enrichment, a PCR mix3 was prepared containing 50  $\mu$ l nuclease-free water (Qiagen), 20  $\mu$ l 5  $\times$  KAPAHiFi reaction buffer (KAPA Biosystems, MA, USA), 3  $\mu$ l a mix containing 10 mM of each dNTP (KAPA Biosystems), 3  $\mu$ l P5RP1 primer (100  $\mu$ M), 3  $\mu$ l EP7v2 primer (100  $\mu$ M), and 2  $\mu$ l of 1 U/ $\mu$ l KAPAHiFi Hot Start DNA polymerase (Roche, Basel, Switzerland). The 25- $\mu$ l mixture of 5'-end-filled cDNAs and the beads was transferred into a new 0.2-ml PCR tube and mixed with 75  $\mu$ l of PCR mix 3 and centrifuged briefly. The beads were resuspended by pipetting, after which the tubes were immediately set on a thermal cycler and a PCR was performed using the following conditions: an initial denaturation at 95°C for 2 min; followed by 10 cycles of 98°C for 20 s, 63°C for 30 s, and 72°C for 30 s; with a final extension at 72°C for 5 min, after which the samples were stored at 4°C. The enriched libraries were purified with a MinElute PCR purification kit (Qiagen) and eluted with 28  $\mu$ l EB, according to the manufacturer's instructions.

The NGS libraries were next subjected to size selection, where fragments measuring 300 bp to 800 bp were electrophoretically recovered from the enriched NGS libraries using BluePippin (Sage Science, Beverly, MA, USA) with a 1.5% dye-free agarose gel cassette and the internal standard R2, according to the manufacturer's instructions. Size-selected NGS libraries were purified with a MinElute PCR purification kit (Qiagen) and eluted in 28  $\mu$ l EB, according to the manufacturer's instruction. The quantity and quality of the NGS libraries were determined using a Bioanalyzer 2100 with a High Sensitivity DNA kit (Agilent Technologies). The oligo DNAs used in this work are listed in Supplementary Table S1.

### qPCR

Each 20- $\mu$ l qPCR mixture contained 2  $\mu$ l of the cDNA templates, LightCycler 480 SYBR Green I Master mix (Roche), and 0.5  $\mu$ M of each primer, which are listed in Supplementary Table S1. The qPCRs were performed using a LightCycler 480 (Roche) with the following conditions: 95°C for 8 min; followed by 35 to 50 cycles of 95°C for 10 s, 56°C for 20 s, and 72°C for 15 s. After the amplification cycles, the melting curves were checked to confirm the target validity using the following conditions: 95°C for 10 s, 65°C for 1 min, and heating to 97°C while determining the fluorescence intensity of SYBR Green I five times per 1°C increase. The transcript levels (copy numbers and  $C_p$  values) were calculated using standard curves for absolute quantification generated using

a dilution series (10, 1, 0.1, 0.01, 0.001, 0.0001 and 0.00001 pg/ $\mu$ l) of the following plasmids: *NGG*: pENTR::*NGG* (5.2 kb), *PpCYCD;1* (AJ428953): pJET::*PpCYCD;1* (4.5 kb), *PpEF1 $\alpha$*  (XM.001753007): pJET::*PpEF1 $\alpha$*  (4.9 kb) and *PpTUAI* (AB096718): pphb6e07 (4.9 kb). Using the molecular weight of the plasmids, the copy numbers of the transcripts were calculated as follows: weight in Daltons (g/mol) = (bp size of plasmids) (615[Da/bp]). Hence, (g/mol)/Avogadro's number = g/molecule = copy numbers. This calculation produced copy numbers equivalent to the double-stranded DNA. These experiments were carried out and evaluated using three sets of experimental replicates. Missing values were substituted for a value one-tenth of the minimum value of each transcript level.

## NGS

A 10- $\mu$ l aliquot of 10 nM sequence libraries consisting of either 32 cells facing the cut edge of the leaf at 0 h after its excision or 34 cells facing the cut at 24 h after leaf excision was denatured and loaded into a lane of the flow cell on a HiSeq1500 sequencer (Illumina, San Diego, CA, USA), following the manufacturer's instructions. The SBS condition in HighOutput v4 was 126 bp of Read1 and 20 bp of index read. The output data in the BCL files were converted to fastq files of read1 and the index read using bcl2fastq package v1.8.4 (Illumina), and the NGS data were deposited in DDBJ (accessions DRA006455 and DRA006456).

## Demultiplexing and tag counting of 1-cellDGE data

For the sequences with 18 bp of index reads, 8 bp were the multiplex index and 10 bp were the UMI. Therefore, 2 bp of 20 bp of index reads were trimmed from the 3' end of the reads in the fastq files, using the cutadapt package (27). To count the numbers of UMIs obtained from the 1-cell-DGE data, the fastq files of read1 and the trimmed index read were processed by the package UMI-SC (Nishiyama, 2016: <https://github.com/tomoakin/UMI-SC>). The 10-bp UMI sequences were inserted into the name of read1, then the sequences were demultiplexed using 6 bp or 8 bp within the trimmed index read for each sample. The read1 sequences were then trimmed and filtered with the trimomatic package (v0.36) (28) using the options 'ILLUMINA-CLIP:adapters.1cellDGE.fa:2:30:7 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:30', after which they were mapped onto the reference transcripts using bowtie (v1.1.2). Subsequently, one mapped read was randomly selected from the all mapped reads with the same UMI that are located on the same gene. Finally, the UMI counts for the genes and transcripts (isoforms) were estimated using RSEM (v1.3.0) (29) and exported a table of 1-cell-DGE data. The read1 data were mapped on *Physcomitrella* v3.3 gene models (30) (Phytozome12: <https://phytozome.jgi.doe.gov/pz/portal.html>).

## Statistical analyses of 1-cell DGE data

To check the quality of 1-cell-DGE data, the SinQC program (31) was run using the following settings: MAX FPR: 0.05,

TPM Cutoff: 1, Spearman's test *P*-value: <0.001, and Pearson's test *P*-value: <0.001. The UMI counts were normalized using a scaling normalization method with iDEGES implemented in the TCC package (32), using trimmed mean of M (TMM) values (33) and an exact test of edgeR (34) with the following settings: norm.method = 'tmm', test.method = 'edgeR', iteration = 3, FDR = 0.1, and floor-PDEG = 0.05.

To detect DEGs, the *q*.value was set to FDR < 0.01. GO term enrichment analysis was performed using cytoscape v3.4.0 with a BinGO plug-in (35), and the ontology of biological processes was assessed using GOSlim\_plants. The BinGO settings were as follows: statistical test: binomial test, multiple testing correction: Benjamini & Hochberg False Discovery Rate (FDR), and significance level: 0.01. Most of the statistical analyses were performed using R v3.3.3 with Rstudio v0.99.491. The plots were drawn using ggplot2 package v2.2.1. The ICA and pseudotime calculations were carried out using the monocle package v1.4.0 (36). Hoeffding's *D* tests of independence were performed using Hmisc package v4.0-3 (37).

## RESULTS

### Extraction of the contents of single cells in excised leaves

We employed microcapillary manipulation to isolate the contents of individual leaf cells in *Physcomitrella* while recording their positional information. Our initial attempts to generate cDNA from the extracted contents of entire cells were rarely successful, presumably because the central vacuole occupies ~ 90% of the plant cell volume (38) and accumulates RNases that degrade RNA molecules (39). Since the transcriptomes of isolated nuclei are reported to be similar to those of whole cells (40,41), we extracted nuclei labelled with a fusion protein (*NGG*) composed of a nuclear-localizing signal (42), sGFP (synthetic green fluorescent protein) (43) and GUS ( $\beta$ -glucuronidase) (44), along with some inevitable accompanying cytoplasm, monitoring the signals for sGFP instead of those of GUS to enable an immediate collection.

We excised the distal half of *Physcomitrella* leaves and, after 24 h, sucked the nucleus and surrounding cytoplasm from individual leaf cells facing the cut (Supplementary Figure S1, Supplementary Movie S1). We synthesized cDNA from the RNA in the cellular contents without any purification, and amplification using PCR, a quantitative PCR (qPCR) was used to determine the transcript levels of four genes: *NGG*, *CYCLIN D;1* (*PpCYCD;1*), *ELONGATION FACTOR 1 $\alpha$*  (*PpEF1 $\alpha$* ), and *TUBULIN  $\alpha$ 1* (*PpTUAI*) (Supplementary Figure S2). The transcript levels of *PpCYCD;1* (sample standard deviation: *s* = 7.9423) and *PpTUAI* (*s* = 7.9431), which are known to be upregulated during reprogramming (18,21), varied among the contents of the different cells, while those of the positive controls *NGG* (*s* = 2.8616) and *PpEF1 $\alpha$*  (*s* = 1.9492) were mostly stable in all contents, as expected. The variation in the *PpCYCD;1* and *PpTUAI* transcripts suggests that the isolated single cells include those in different stages of reprogramming as well as those not undergoing reprogramming, as previously observed (18). These results indicate that cDNAs

derived from single-cell contents extracted using a microcapillary can be used to detect gene expression in single leaf cells of *Physcomitrella*, as was previously shown for isolated cell contents of individual cells of tobacco, potato and cucumber (45).

### Preparation of cDNA libraries from traces of total RNA

To make a cDNA library of at least 40 femtomoles ( $2.4 \times 10^{10}$  molecules) for sequencing on Illumina next-generation sequencers, we amplified the single-cell cDNAs using PCR. Two methods of attaching an adaptor to the 3' ends of the first-strand cDNAs were evaluated: template switching (46) and polydeoxyadenines (dA) tailing (47–49). Template switching introduces several non-templated deoxycytosine residues to the 3' ends of the first-strand cDNA before synthesizing the complementary strands. Poly(dA) tailing extends the poly(dA) region at the 3' ends of the first-strand cDNAs using a terminal deoxynucleotidyl transferase, followed by the annealing of a specific primer with 20-nt oligo (dT) sequences to the poly(dA) tail. After second-strand synthesis using both types of cDNAs, we performed qPCR to detect cDNA quality and quantity using various RNA amounts for the libraries. Using the poly(dA) tailing libraries, the decreases in the crossing point PCR cycle ( $C_p$ ) values of *PpTUAI* cDNA were proportional to increases in the total RNA amounts from 10 pg to 10 ng (Supplementary Figure S3). By contrast, the changes in the  $C_p$  values when using the template switching libraries did not demonstrate this relationship. We therefore adopted poly(dA) tailing for our single-cell transcriptome analysis.

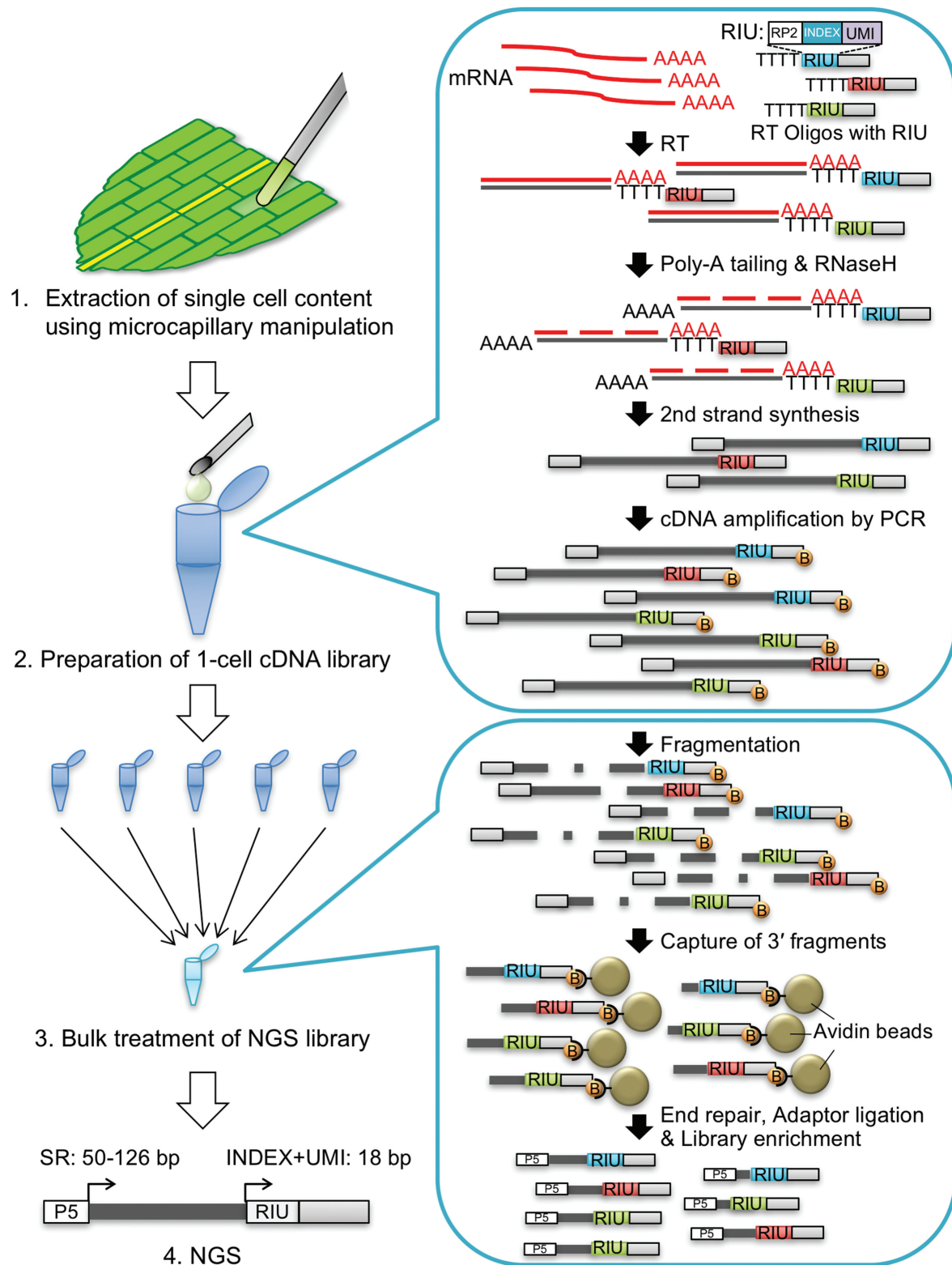
Another problem is the amplification bias in cDNAs when using PCR. These biases are caused by differences in amplification efficiencies, which depend on the length, nucleotide contents, and sequences of the DNA fragments, as well as stochastic fluctuations (50). To reduce the template-dependent biases, we adopted the unique molecular identifier (UMI) method, in which random barcode sequences are introduced into the first-strand cDNA at the time of reverse transcription (23). When the sequence reads are mapped to the reference genome (24,30), reads mapped to the same gene locus with the same UMI were considered to have originated from the same cDNA, and were therefore unified as one UMI count. When reads mapped at different gene loci, the reads were counted as independent reads even if they had the same UMI. To test the UMI method, we sequenced the cDNA libraries derived from 5  $\mu$ g and 20 pg of total RNA extracted from protonema cells from *Physcomitrella* (Supplementary Figure S4, Supplementary Table S2). In the two 5- $\mu$ g library replicates, 780 202 and 972 450 reads were mapped onto the *Physcomitrella* v3.3 gene model, which were unified to 680 993 and 832 593 UMI counts, respectively. As a result, 6.95% and 8.34% of reads were unified and removed, respectively. In the two 20-pg library replicates, 606 512 and 660 397 reads were mapped, and were unified to 82 229 and 101 533 UMI counts, respectively. As a result, 84.21% and 82.17% of the reads, respectively, were unified and removed. We found that the read counts of each gene in the 5- $\mu$ g libraries were strongly correlated between the duplicated samples ( $R^2 = 0.9891$ ), even if the UMIs were not unified ( $R^2 = 0.9743$ ). On the other hand,

the read counts in the 20-pg libraries tended to vary if the UMIs were not unified ( $R^2 = 0.8610$ ); however, the UMI counts of each gene were strongly correlated between duplicated samples, to a similar level as those of the 5- $\mu$ g libraries ( $R^2 = 0.9677$ ). We found similar tendencies for replicates of the External RNA Controls Consortium (ERCC) RNA spike-in mix, which were added to the total RNA samples as an external control (Supplementary Figure S4C, D) (51). Furthermore, the UMI counts reflected the nominal concentrations of the ERCC RNA spike-in mix, although the correlation was somewhat lower in the 20-pg libraries than the 5- $\mu$ g libraries (Supplementary Figure S4E–H).

### Preparation of cDNA libraries from the contents of single cells and NGS

To apply the methods described above to individual *Physcomitrella* leaf cells, we made cDNA libraries from single cells facing the cut of an excised leaf after 0 and 24 h (Figure 1). The cell contents, including the nuclear region marked by the NGG, were extracted using microcapillary manipulation from 32 cells at 0 and 34 cells at 24 h after the leaves were excised. The content of each cell was transferred to a PCR tube and cDNA was synthesized using reverse transcription with an exonuclease I treatment, poly(dA) tailing, second-strand synthesis, and cDNA amplification. Subsequently, the cDNA quality and quantity were measured, and the samples that showed a peak of cDNAs between 500 bp and 5000 bp in length were purified by removing the byproducts detected between 100 bp and 400 bp. To prepare the Illumina sequencing libraries, four or five samples with different multiplex sequences were mixed equally in a single batch, and were subsequently subjected to fragmentation, end-repair, dA-tailing, adaptor-ligation, library enrichment and library purification.

After quantification and qualification, these batches of NGS libraries were equally mixed and sequenced on an Illumina HiSeq sequencer set to generate 126-bp single-end reads and 18-bp index reads, including 8 bp of multiplex index and 10 bp of UMI. A total of 384 993 923 reads were obtained and cleaned by trimming and filtering to remove inaccurate sequences and ribosomal sequences, respectively. As a result, 98.1% (377 495 213) of the reads were sorted to their respective samples using the multiplex index sequences, generating 2.8–8.5 million reads per sample for 66 samples (Supplementary Figure S5, Supplementary Table S3). Mapping these reads to *Physcomitrella* gene models, means of 5 566 262 and 4 966 189 mapped reads were obtained for the samples taken at 0 and 24 h, respectively, which equated to mapping rates of 90.9% and 92.6%, respectively. These mapped reads were unified to remove duplicated reads located at the same gene locus and with the same UMI sequences, because the UMI sequences were introduced during the reverse transcription step and later amplified using PCR. Only sequences with different UMI sequences at the same gene locus were therefore considered as distinct cDNAs in the quantification of the original numbers of transcripts. For the cells sampled at 0 h and 24 h, the mean UMI counts were 102 145 and 91 851, respectively, and the UMI-unified rates, indicating the rate of duplicated reads (same locus with the same UMI), were 98.3%



**Figure 1.** Schematic representation of the workflow of 1cell-DGE. 1cell-DGE involves four major steps; 1, extraction of single-cell contents by microcapillary manipulation; 2, preparation of 1-cell cDNA libraries; 3, bulk treatment of next-generation sequencing (NGS) libraries; and 4, NGS. The conditions for the NGS using illumina sequencers were as follows: single read (SR): 80 bp; INDEX read: 18 bp, which includes 8 bp of multiplex indexes and 10 bp of UMI. In the figure, RIU indicates the oligonucleotides, including the RP2 sequences for INDEX read priming, 8 bp of multiplex indexes, and 8 bp or 10 bp of random oligonucleotide sequences for the UMI. B indicates the biotin modification of the 5' end of the DNA oligonucleotides for cDNA amplification, which capture cDNA fragments with avidin-conjugated magnetic beads.

and 98.3%, respectively (Supplementary Figure S6, Supplementary Table S3). At 0 h and 24 h, the mean numbers of transcribed genes per sample were 5277 and 7297, respectively.

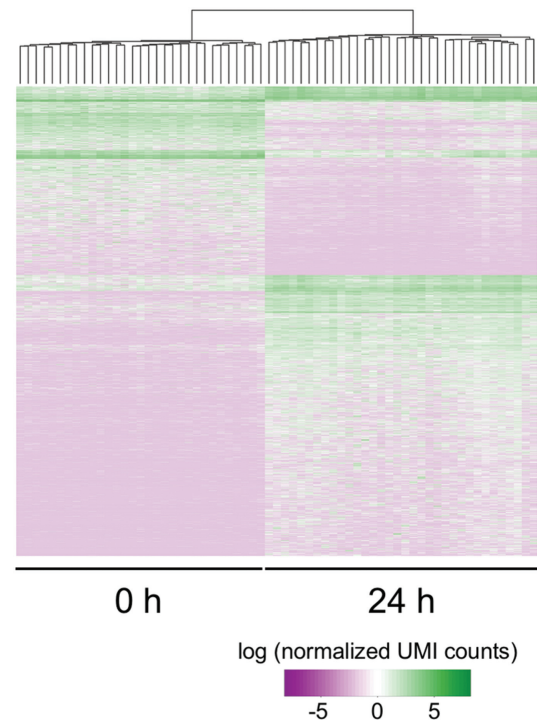
Generally, validating transcriptome data involves comparing the transcript levels of internal control genes with a similar expression level among all samples; however, it was difficult to choose an appropriate gene because the expression of many genes, including those which are generally accepted as housekeeping genes (e.g. *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* and  $\alpha$ -*TUBULIN*), were found to fluctuate substantially at the single-cell level (Supplementary Figure S2). We therefore checked the quality of the 1cell-DGE data using a statistically analyzed population of single-cell transcriptome data in the SinQC package (31), judging the outliers based on the statistics of mapping rates, the number of detected genes, and read complexity. All but one sample (31 cells at 0 h and 34 cells at 24 h after leaf excision) passed this evaluation with a max false positive rate (FPR) of 0.05 and using the following settings: TPM Cutoff: 1, Spearman's test  $P$ -value:  $< 0.001$ , Pearson's test  $P$ -value:  $< 0.001$  (Supplementary Table S3). These data were therefore used for further analyses.

To estimate how many reads are adequate for single-cell profiling using 1cell-DGE, we calculated the number of detected genes and UMI-unified rates within a limited number of reads randomly extracted from 1cell-DGE data at 0 and 24 h (Supplementary Figure S7). We did not detect any significant differences in the tendencies of these statistics between the 0-h and 24-h samples or among the selected index sequences. Although the numbers of detected genes increased as the number of sampled reads increased, the rate of change slowed as more sampled reads were considered. The UMI-unified rates also increased as the number of sampled reads increased, although they appeared to have close to an asymptotic relationship. At 2 million and 5 million reads, UMI-unified rates of 98.0% and 98.4%, respectively, were calculated for the samples taken at 0 h. For the samples taken after 24 h, UMI-unified rates of 97.1% and 98.0%, respectively, were calculated.

### Expression profiles of individual cells at 0 h and 24 h after leaf excision

To detect differentially expressed genes (DEGs) in the 1cell-DGE data taken at 0 h and 24 h after leaf excision, we carried out a statistical analysis after normalization using the iterative differentially expressed gene exclusion strategy (iDEGES) method (32). A total of 6382 genes were identified as DEGs, of which 2382 and 4000 genes were expressed at higher levels in the samples taken at 0 h (0 h-high) and 24 h (24 h-high), respectively, when calculated using the criterion of a false discovery rate (FDR)  $< 0.01$ . Using these gene expression profiles, we performed a hierarchical clustering and found that profiles for 0 and 24 h were clearly categorized into separate populations, indicating characteristic transcript profiles (Figure 2).

We performed a gene ontology (GO) term enrichment analysis for the 1978 of the total 2382 DEGs at 0 h and 3648 of the total 4000 DEGs at 24 h that were putatively homologous to annotated Arabidopsis genes (Figure 3) (35).

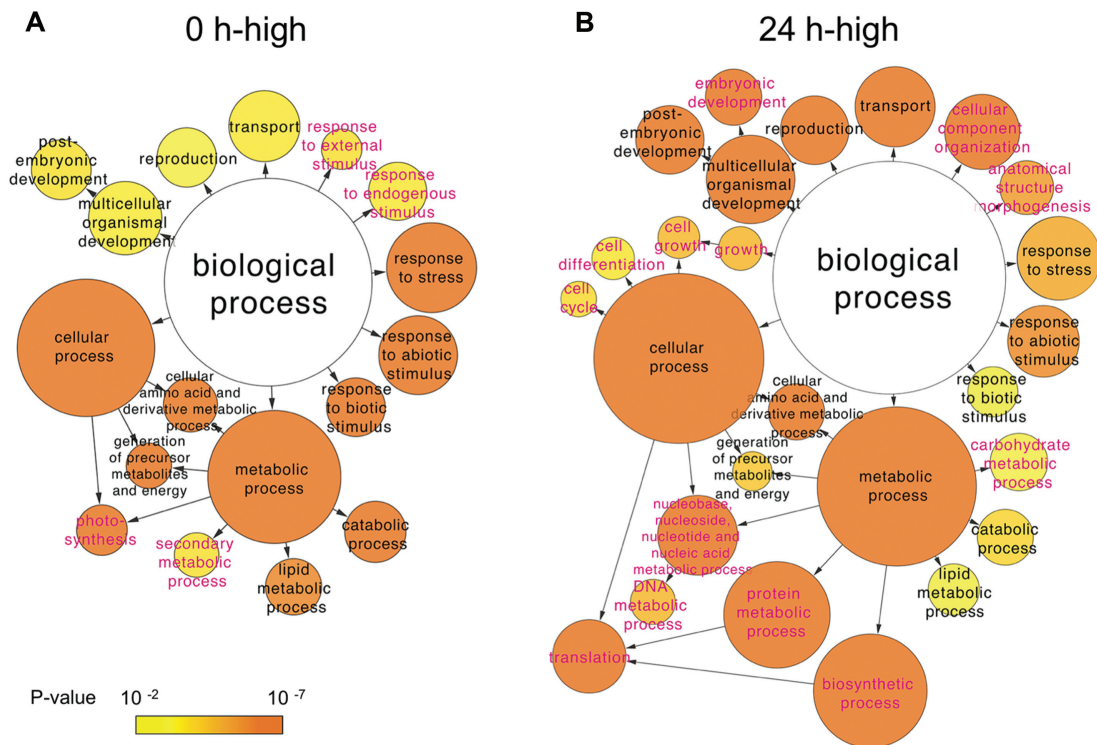


**Figure 2.** Hierarchical clustering of differentially expressed genes (DEGs) in individual cells at 0 and 24 h after leaf excision. Using 1cell-DGE data taken from 31 and 34 cells at 0 and 24 h after leaf excision, respectively, 6,382 genes were identified as DEGs (FDR  $< 0.01$ ) after normalization with iDEGES using the TCC package (32). Of these, 2382 and 4000 were more highly expressed in the cells at 0 and 24 h, respectively. Hierarchical clustering was performed using hclust in the stats package. The colored bars indicate the normalized UMI counts as the expression levels of the DEGs on a log scale.

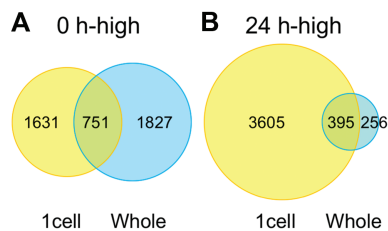
Using GOSlim\_plants to categorize the genes, we revealed an enrichment of genes involved in the responses to stress and abiotic and biotic stimuli, the generation of precursor metabolites and energy, metabolic processes involving cellular amino acids and their derivatives, lipid metabolic processes, catabolic processes, post-embryonic development, reproduction, and cellular transport in both the 0 h-high and 24 h-high DEGs. In addition, the GO terms of photosynthesis, secondary metabolic process, and response to endogenous/external stimulus were enriched at 0 h, whereas those of cell growth, cell cycle, cell differentiation, embryonic development, DNA and protein metabolic process, biosynthetic process, translation, carbohydrate metabolic process, anatomical structure morphogenesis, and cellular component organization were enriched at 24 h (Supplementary Tables S4–S14).

In previous studies, transcriptome analyses of whole excised leaves during reprogramming were performed using 5'DGE (19,21). We therefore compared the DEGs identified using 1cell-DGE with those reported using the 5'DGE method for whole excised leaves (Figure 4). After remapping the 5'DGE data onto the Physcomitrella v3.3 gene models (30) and counting the read tags for each gene locus, 2578 and 651 DEGs with a FDR  $< 0.01$  were detected as 0 h-high and 24 h-high genes, respectively. A total of 751 of the 0 h-high DEGs were commonly identified in both the 1cell-





**Figure 3.** GO term enrichment analysis of differentially expressed genes (DEGs) in individual cells at 0 and 24 h after leaf excision. A total of 1978 of the 2,382 DEGs more highly expressed at 0 h and 3648 of the 4000 DEGs more highly expressed at 24 h after leaf excision could be annotated based on their homology to genes in *Arabidopsis thaliana*. Using the associated *Arabidopsis* gene identifiers, a GO term enrichment analysis was performed using cytoscape v.3.4.0 with the BinGO plug-in, and the ontology of biological processes was assessed using GOSlim\_plants. The terms in magenta text indicate sub-categories are only represented in the DEGs more highly expressed at either 0 or 24 h after leaf excision. The circles are colored based on the statistical significance of their enrichment.



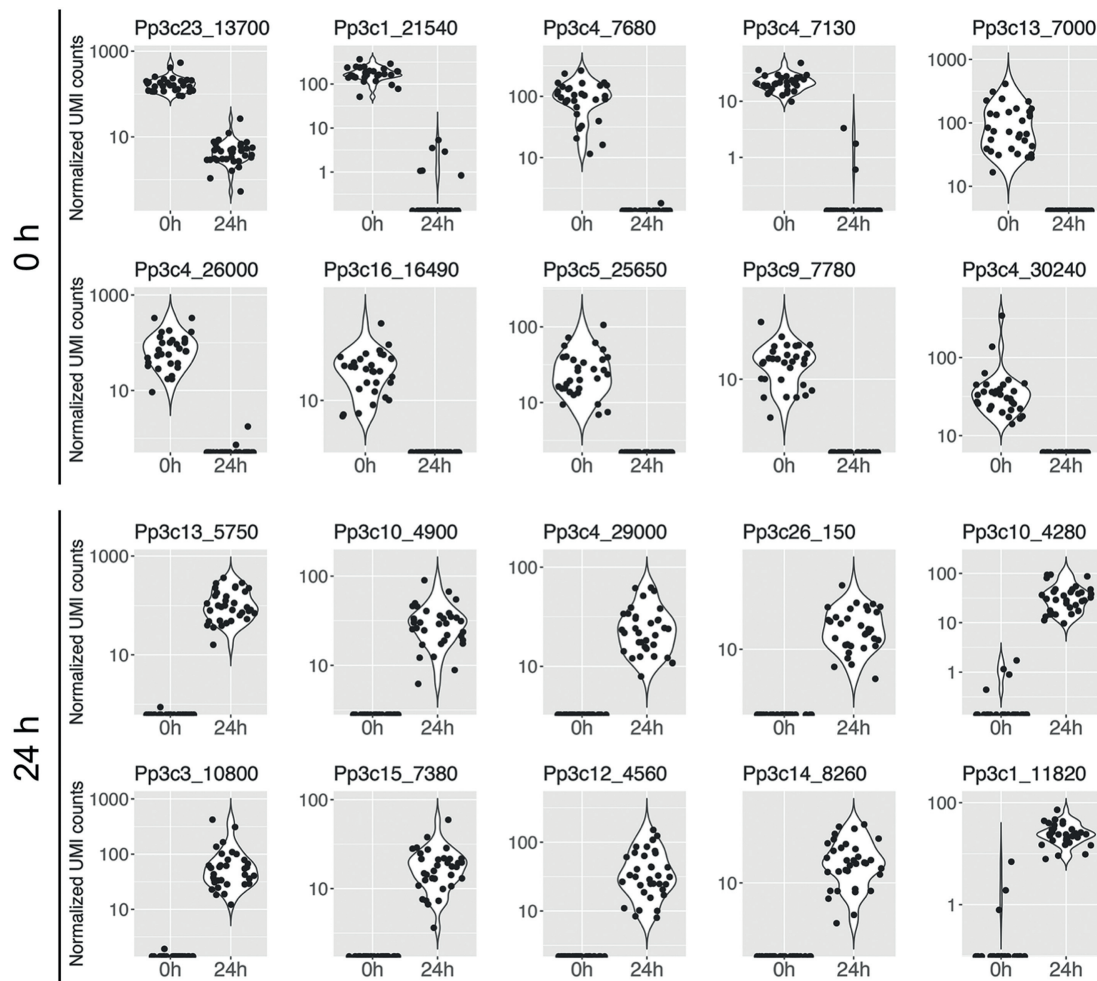
**Figure 4.** Comparison of differentially expressed genes (DEGs) between 1cell-DGE and whole leaf-5'DGE. Venn diagrams show the number of DEGs identified using 1cell-DGE and 5'DGE in single cells of excised leaves or the whole leaf, respectively, which were more highly expressed at 0 or 24 h after leaf excision. The 5'DGE data from (19,21) were remapped onto *Physcomitrella* v3.3 gene models (30) and normalized using the iDEGES method with the TCC package (32).

DGE and 5'DGE analyses, while 395 of the 24 h-high DEGs were commonly identified between datasets.

We then checked the expression levels of the top 10 DEGs detected by the statistical significance of their differences at 0 and 24 h using a q-value (Figure 5). Pp3c23\_13700 (unknown), Pp3c1\_21540 (aluminium induced protein-like), Pp3c4\_7680 (membrane protein, putative), Pp3c4\_7130 (unknown), Pp3c13\_7000 (glyoxal oxidase-related protein-like), Pp3c4\_26000 (chaperone DnaJ-domain superfamily protein-like), Pp3c16\_16490 (unknown), Pp3c5\_25650 (unknown), Pp3c9\_7780 (calcium-dependent lipid-binding

family protein-like), and Pp3c4\_30240 (TOXICOS EN LEVADURA 2-like) were selected as the top 10 DEGs more highly expressed at 0 h. Pp3c13\_5750 (lactoyl-glutathione lyase / glyoxalase I family protein-like), Pp3c10\_4900 (unknown), Pp3c4\_29000 (unknown), Pp3c26\_150 (unknown), Pp3c10\_4280 (bHLH protein), Pp3c3\_10800 (adenosine kinase 2-like), Pp3c15\_7380 (dihydrodipicolinate reductase-like), Pp3c12\_4560 (expansin A9-like), Pp3c14\_8260 (succinyl-CoA ligase, alpha subunit-like), and Pp3c1\_11820 (unknown) were selected as the top 10 DEGs more highly expressed at 24 h.

In order to take full advantage of single-cell transcriptome data, it is possible to calculate the pseudotime, a hypothetical time scale estimating the transition between cell states during development and differentiation based on similar gene expression profiles (11,36). First, an independent component analysis (ICA) was carried out to reduce the dimensions of the gene expression profiles (Figure 6). Like the hierarchical clustering, nodes indicating the individual samples in the ICA were clearly separated between cells sampled at 0 or 24 h after excision (Figure 6A). Furthermore, we could not find any relationship between the cell profiles for the extracted nuclear condition, leaf excision date, cDNA amount, or byproduct contamination (Supplementary Figure S8), further confirming the correlation between the ICA result and the other criteria. When each point in the ICA plot was colored according to its pseudotime, almost



**Figure 5.** Expression profiles of 20 genes differentially expressed in single cells at 0 h and 24 h after leaf excision. The expression profiles of the genes which ranked in the top 10 DEGs at 0 h (upper two rows) and 24 h (bottom two rows) are shown as violin and jitter plots. The top 10 DEGs were selected by ranking the DEGs based on their  $m$ -values,  $P$ -values and  $q$ -values calculated with the TCC package (32).

all points for both the 0-h and 24-h samples were found to be arranged in order of their pseudotime (Figure 6B).

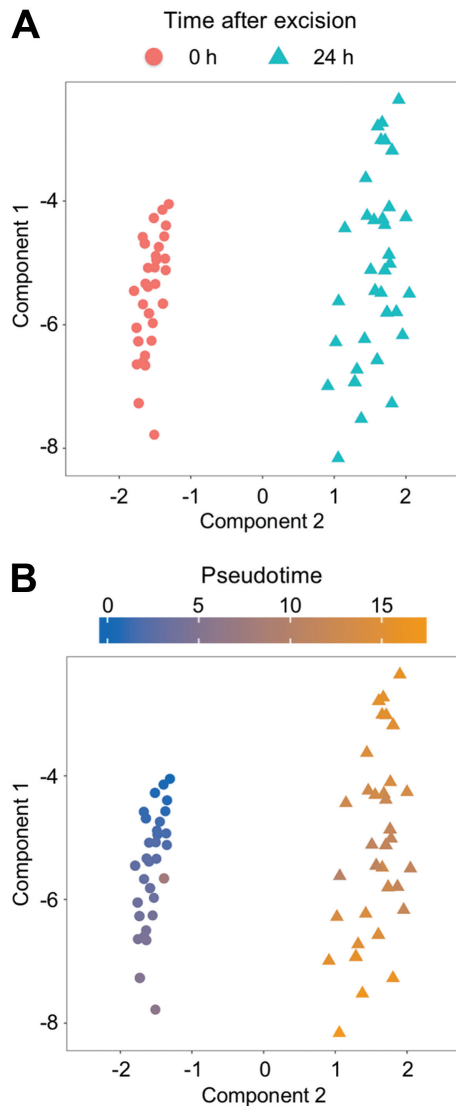
When the expression profiles of the *Physcomitrella* reprogramming-related genes *PpCSP1*, *PpCSP2* and *PpCYCD;1* (18,20) were plotted against pseudotime (Figure 7), they were generally found to be expressed at low levels in the early phase of pseudotime, with the exception of several cells with high *PpCSP1* expression. Further along the pseudotime scale, *PpCSP1* was the most highly expressed in cells at 24 h after the leaf excision. In contrast, *PpCYCD;1* expression varied substantially among cells at 24 h after the leaf excision, which is likely attributable to the heterogeneity in the reprogramming ability of the cells at the cut edge (22).

We also compared the correlation measures between the pseudotime and NGS statistics. Using Hoeffding's  $D$  test of independence for nonparametric and non-monotonic relationships (37), we identified a low correlation between the mapped read counts and the pseudotime ( $D = 0.014$ ,  $P = 0.0493$ ), but found a moderate correlation between the number of detected genes and the pseudotime ( $D = 0.285$ ,  $P = 10^{-8}$ ) (Figure 8).

## DISCUSSION

### Single cell transcriptome analysis using microcapillary manipulation

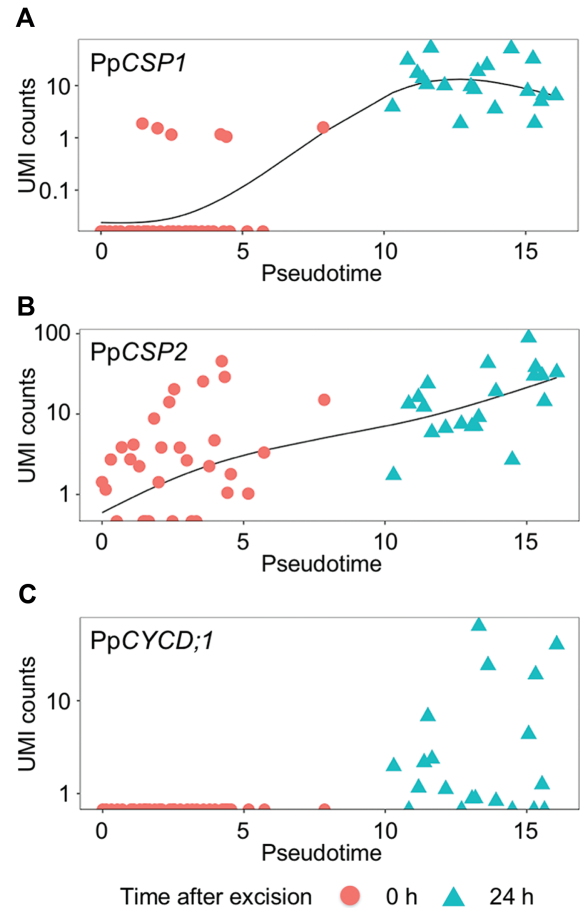
Since the first report of scRNA-seq (47), many methods of scRNA-seq have been developed and improved (5,6,52), but two major concerns have arisen: how can a single cell or its contents be isolated, and how can cDNA be efficiently and accurately prepared from trace amounts of RNA. The former challenge has been partly overcome using cell sorting with fluorescence-activated cell sorting (FACS) and microfluidics (7–9); however, applying these systems require the separation of individual cells from their tissues or cultures. When the cells are separated using these techniques, the resulting samples no longer have accurate positional information. In addition, plant cells are often tightly attached to each other by their cell walls, making it difficult to mechanically or enzymatically detach them from each other while keeping their cellular contents intact. Microcapillary manipulation or laser microdissection can be used to extract the contents of single cells from tissues without detaching them, overcoming these challenges and enabling the tran-



**Figure 6.** Independent component analyses of the 1cell-DGE data. Normalized 1cell-DGE data were calculated to reduce the dimensions of the expression profiles using an independent component analysis, and were plotted with the monocle package (36). Each dot indicates an independent cell sample categorized by the time it was sampled after excision (A) or the pseudotime (B).

scriptomic analysis of individual cells while retaining positional information.

Reverse transcription polymerase chain reaction (RT-PCR) has previously been successfully performed with cDNA derived from the contents of single cells extracted from plant leaf cells using glass microcapillaries (45); therefore, we attempted to prepare cDNA libraries from the contents of individual leaf cells from the moss *Physcomitrella* (Figure 1). Marking the nuclei with GFP (Supplementary Figure S1, Supplementary Movie S1) enabled us to reproducibly recover the cell contents including nuclear region using microcapillary manipulation and prepare cDNA that could be used for qPCR (Supplementary Figure S2). Previous transcriptome analyses of isolated nuclei have demonstrated that they have similar expression profiles to those

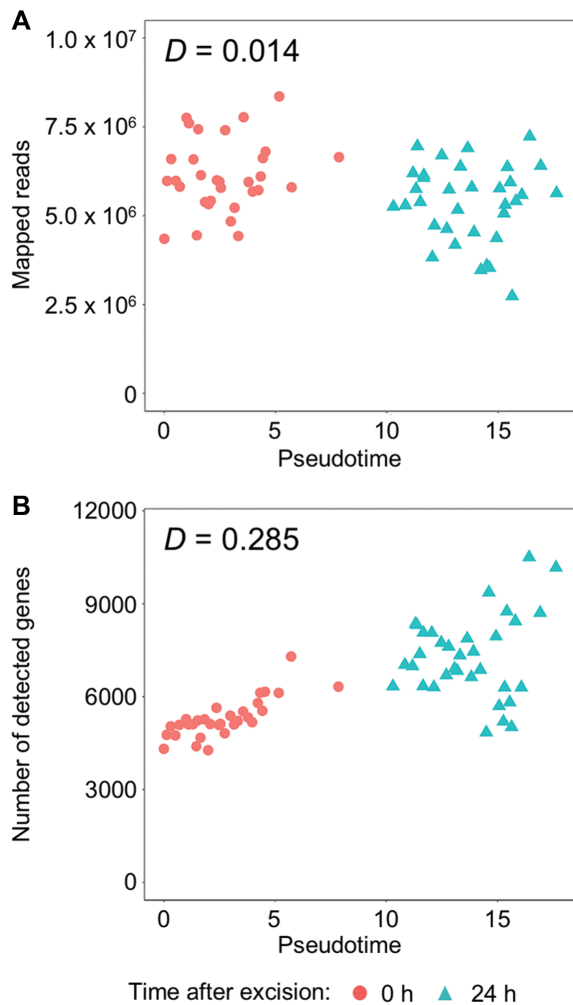


**Figure 7.** Expression profiles of reprogramming-related genes ordering by pseudotime. PpCSP1 and PpCSP2, genes encoding cold-shock protein 1 and 2 (20); PpCYCD;1: gene encoding cyclin D;1 (18). These were calculated and plotted using the monocle package (36). Each dot indicates an independent cell depicted as the time it was sampled after the excision.

of whole cells (40,41), indicating that this is an appropriate technique to use for the preparation of cDNA from individual cells.

Another issue was how to prepare cDNA from a small amount of RNA without excessive amplification bias which depends on the length, nucleotide contents, and sequences of the cDNAs (50). Using conventional methods, NGS libraries are prepared for RNA-seq by purifying the mRNAs and fragmenting them before reverse transcription. In contrast, for scRNA-seq, the mRNAs are not purified, and are instead directly reverse-transcribed to cDNA from the crude cell contents. Generally, template switching or poly(dA) tailing is utilized to attach the adaptor oligonucleotides at the 3' end of the cDNA after reverse transcription. In this study, we employed the latter technique for the 1cell-DGE based on our results (Supplementary Figure S3). While we found template switching to be less effective than poly(dA) tailing, this could be improved by the use of a short template-switching oligo and locked nucleic acid (LNA)-linked nucleotides (53–55), and might be suitable for use with 1cell-DGE following such improvements.

The trace amount of first-strand cDNA generated from the RNA of single cells necessitates their amplification be-



**Figure 8.** Correlation between pseudotime and NGS statistics. The pseudotimes of the 1cell-DGE data were plotted against (A) the mapped read counts and (B) the numbers of detected genes. Each dot indicates an independent cell categorized by the time it was sampled after the leaf excision.  $D$  indicates the statistical values calculated using Hoeffding's  $D$  test of independence in the Hmisc package (37).

fore they can be sequenced. To overcome an amplification bias, we introduced sequences of 6 or more random nucleotides, UMIs, to the cDNAs to enable their later discrimination (Figure 1) (23). We designed 103 species of oligo (dT) nucleotides comprising 10 nt of UMI and 8 nt of multiplex index, which enabled us to identify the original index even if one substitution error occurred on the index sequence (Supplementary Table S1) (56). Using these RT oligos, we can mix samples with different multiplex indexes after the synthesis of the first-strand cDNA and subsequently prepare the NGS libraries as bulk samples. Moreover, as sequencing generated single-end reads of 50–126 bp with 18 bp of index reads, this approach is expected to reduce sequencing costs and more efficiently generate analyzable reads than conventional scRNA-seq with pair-end reads. After NGS, the original numbers of first-strand cDNAs can be estimated by unifying the reads derived from the same molecule, which are defined as the reads mapped to the same gene locus that possess the same UMI.

To test this, we performed pilot sequencing using total RNA purified from *Physcomitrella protonema* tissues. The 5- $\mu$ g samples showed relatively low unified rates of 6.95% and 8.34%, which indicated that the starting mRNA materials were sufficient for the preparation of cDNAs with different UMIs (Supplementary Table S2). By contrast, the 20-pg samples showed higher unified rates of 84.21% and 82.17%, which means that the cDNAs were prepared from trace amounts of mRNA and amplified using PCR, suggesting that the single-cell samples contained many duplicated reads after the preparation of the NGS libraries. In 5- $\mu$ g samples of cDNA, which had not been amplified using PCR, we found a similar determination coefficient ( $R^2$ ) between the read counts and the UMI counts (Supplementary Figure S4A). By contrast, in the 20-pg samples, the  $R^2$  value of the UMI counts was higher than that of the read counts (Supplementary Figure S4B). The quantification values of the ERCC RNA spike-in mix added to the pilot sequencing samples (Supplementary Figure S4C, D) also showed high  $R^2$  values for both the read counts and the UMI counts between replicates. We confirmed that 1cell-DGE using UMIs enabled the highly reproducible quantification of cDNA from trace samples of RNA. In addition, we found a high correlation between the concentrations and UMI counts of the ERCC RNA spike-in mix in both the 5- $\mu$ g and 20-pg samples (Supplementary Figure S4E–H). The lower  $R^2$  values in the 20-pg samples may originate from the low coverage of the pilot sequencing (Supplementary Table S2), and the drop-out reads known to be a feature of scRNA-seq (57,58). Our results therefore demonstrate that we can quantify the original transcript abundance with high reproducibility and sufficient accuracy using UMIs.

Next, we analyzed the transcriptomes of single cells extracted from the cells of an excised leaf after 0 and 24 h. We extracted the cell contents from 32 and 34 cells at 0 h and 24 h, respectively, which were used for the preparation of NGS libraries with 1cell-DGE. A total of 2.8 million and 8.5 million reads were obtained, representing mapping rates of 89.9% and 91.5% for the 0 and 24 h samples, respectively (Supplementary Figure S5, Supplementary Table S3). These numbers indicate that our 1cell-DGE method can be used to efficiently construct NGS libraries. Although 98.3% of the mapped reads were removed when the UMI-unifying was performed (Supplementary Figure S6, Supplementary Table S3), our simulation of the relationship between read counts and the UMI-unified rate indicated that these read counts are more than sufficient to analyze the transcriptome profiles of the cells used. Our results suggested that a UMI-unified rate of one to two million read counts per sample were sufficient to enable the estimation of the expression profiles to a similar level as that of five million counts per sample, which is similar to estimations reported in previous studies (6) (Supplementary Figure S7). Our 1cell-DGE approach therefore generated adequate reads for single-cell transcriptome analyses.

### Gene expression profiles of individual leaf cells during reprogramming

We executed the SinQC program (31) to check the library quality of the single-cell samples, as it was not known which

mRNAs would be similarly abundant in the individual cells. Based on the distribution of the NGS statistics, all but one of the 66 samples met the quality criteria (Supplementary Table S3). While SinQC is suitable for the quality control of single-cell transcriptome data without an internal control (31), alternative methods may be more appropriate for the quality control of samples from cell populations including rare cell types or with a smaller number of samples (59).

We identified 2382 and 4000 DEGs that were more highly expressed in cells facing the cut edge of a leaf at 0 and 24 h after excision, respectively (Figures 2 and 4). Similar numbers of 0 h-high DEGs were identified in transcriptome profiles generated using whole excised leaves (19,21), although only 751 genes were found to overlap when using the 5'DGE and 1cell-DGE methods. This may indicate that whole gametophores, comprising a variety of cell types in addition to the leaves, were sampled at 0 h in the study using 5'DGE (19,21). On the other hand, a six-fold more DEGs were found to be more highly expressed after 24 h using 1cell-DGE compared to 5'DGE. In addition, 11 GO terms, cell growth, cell cycle, cell differentiation, embryonic development, DNA and protein metabolic process, biosynthetic process, translation, carbohydrate metabolic process, anatomical structure morphogenesis, and cellular component organization were statistically enriched in the 24 h-high data of 1cell-DGE, which corresponded to 41–774 DEGs that were not previously detected by 5'DGE (Supplementary Tables S4–S14). The newly identified genes categorized in the GO analysis may enhance our understanding of the molecular mechanisms underpinning plant reprogramming. These results are concordant with the fact that the whole excised leaves at 24 h after excision would have contained more heterogeneous cells, such as reprogramming and non-reprogramming cells, than those at 0 h. The 1cell-DGE approach was highly sensitive to differences in the expression of cell-state specific genes because only the cells facing the cut were analyzed.

The GO term enrichment analysis revealed that the DEGs were enriched in biological process terms related to specific cell states, with photosynthesis genes being more highly expressed at 0 h, while genes involved in the cell cycle, cell differentiation, translation, and DNA metabolic processes were upregulated at 24 h after excision (Figure 3). The expression of *PpCYCD;1*, a partner of *PpCDKA*, which coordinates cell cycle progression and the acquisition of the protonema cell characteristics involved in reprogramming (18), was not detected at 0 h; however, it was detected in many cells at 24 h (Figure 7). Furthermore, *PpCSP1* and *PpCSP2*, which were identified as the common reprogramming factors among plants and animals (20), were more highly expressed at 24 h after the leaf excision than at 0 h (Figure 7). Our results are consistent with previous works related to reprogramming in *Physcomitrella*, in which low levels of *PpCSP1* promoter activity were detected in the cells of intact leaves, but drastically upregulated in cells facing the cut edge of a leaf (20). On the other hand, we also detected several cells at 0 h with high levels of *PpCSP1* expression and some at 24 h with low levels of *PpCYCD;1* (Figure 7). These variations most likely reflect the heterogeneity of the cells at the cutting edge, where some cells are reprogrammed into stem cells but others are not (22). By

contrast, the top 10 DEGs detected using 1cell-DGE exhibited no or low levels of expression at 0 h and high levels of expression at 24 h after the leaf excision (Figure 5). These genes may be suitable for use as new cell state markers to discriminate between resting and reprogramming leaf cells in future research.

In addition to these conventional analyses of transcriptomes, pseudotime is an attractive concept for use with scRNA-seq, because the trajectory of the cell states can be predicted even if not all of the various states of the cell profile have been sampled in the analysis (11,36). Using only the profiles of individual leaf cells at 0 and 24 h after the leaf excision, the transcriptome profiles were found to be ordered according to pseudotime (Figure 6B). This suggests that the gene expression profiles at 0 h and 24 h fluctuated and might indicate the pattern of reprogramming in cells facing the cut. Furthermore, we did not detect any significant correlation between the time of sampling and the pseudotime at 0 h (Spearman correlation coefficient;  $\rho = 0.0387$ , Supplementary Table S3, Supplementary Figure S9), suggesting that the gene expression profiles of cells sampled 5 to 35 min after leaf excision were very similar. We found that pseudotime was correlated to the numbers of detected genes (Figure 8), suggesting that thousands of genes are transiently expressed during reprogramming or that the number of expressed genes increases during the reprogramming of leaf cells into stem cells. Furthermore, at the late phases of pseudotime, the transcriptomes of the cells sampled after 24 h appeared to be separated into two subpopulations with higher and lower numbers of detected genes (Figure 8). This may be the result of the spontaneous arrest or lateral inhibition of the reprogramming of some cells (22). To clarify this in future research, individual cells separated from other cells and cells facing the cut edge of the leaf should be analyzed at different time points.

We established 1cell-DGE with microcapillary manipulation as a new scRNA-seq technique, successfully using it to profile the transcriptomes of single cells with high reproducibility and accuracy. Although 1cell-DGE is not a method with as high of a throughput as automated single-cell preparation solutions such as Fluidigm C1 (7) and in-Drop (8,9), it can be used to analyze the contents of single cells from living tissues and organs without the preparation of isolated cells and the associated loss of positional information. This will not only widen the scope of single-cell transcriptome analyses using various types of cells, but also contribute to novel insights into cell–cell interactions in the complicated higher-order structures of multicellular organisms. In the future, microcapillary manipulation may be performed by an automated robotics system (60). Using this system, it will be possible to simultaneously recover thousands of cell contents from living tissues and organs.

## DATA AVAILABILITY

The NGS data have been deposited in DDBJ under accession numbers DRA006455 and DRA006456.

The package UMLSC is an open source available in the GitHub repository (<https://github.com/tomoakin/UMLSC>).

The wild-type moss *Physcomitrella patens* Gransden 2004 (24) and the transgenic *Physcomitrella* line GX8-NGG (25), as well as the plasmids pENTR::NGG, *Pp-CYCD;1* (AJ428953), *PpEFL $\alpha$*  (XM.001753007), and *Pp-TUAI* (AB096718); pphb6e07 are available from the corresponding author Mitsuyasu Hasebe (mhasebe@nibb.ac.jp) at the National Institute for Basic Biology (NIBB), Japan.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful to Dr Olaf Faustmann at Eppendorf for help with the micromanipulation. We also thank Drs Shuji Shigenobu and Katsushi Yamaguchi at the National Institute for Basic Biology (NIBB) and Drs Asao Fujiyama and Atsushi Toyoda at the National Institute of Genetics for their help with the next-generation sequencing. We are grateful to Dr Yukiko Kabeya at NIBB and Ritsuko Okamoto at the Nara Institute of Science and Technology for their technical assistance. The computational analyses were partially performed on the NIG supercomputer at the ROIS National Institute of Genetics.

## FUNDING

Japan Society for the Promotion of Science [Young Researcher Overseas Visits Program for Vitalizing Brain Circulation to M.K., A.I., M.H.]; Ministry of Education, Culture, Sports, Science and Technology (MEXT) KAKENHI [JP26113514 to M.K.]; Nara Institute of Science and Technology [Humanophilic Innovation Project to M.K., T.D., Big Data Project to M.K.]; Excellence Initiative of the German Federal and States Governments [EXC294 to R.R.]. Funding for open access charge: Nara Institute of Science and Technology.

*Conflict of interest statement.* None declared.

## REFERENCES

- Steward,F., Mapes,M. and Smith,J. (1958) Growth and organized development of cultured cells. I. Growth and division in freely suspended cells. *Am. J. Bot.*, **45**, 693–703.
- Valvekens,D., Van Montagu,M. and Van Lijsebettens,M. (1988) Agrobacterium tumefaciens-mediated transformation of Arabidopsis thaliana root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 5536–5540.
- Ikeuchi,M., Ogawa,Y., Iwase,A. and Sugimoto,K. (2016) Plant regeneration: cellular origins and molecular mechanisms. *Development*, **143**, 1442–1451.
- Sang,Y., Cheng,Z. and Zhang,X. (2018) Plant stem cells and de novo organogenesis. *New Phytol.*, **218**, 1334–1339.
- Liu,S. and Trapnell,C. (2016) Single-cell transcriptome sequencing: recent advances and remaining challenges. *FI000Res*, **5**, 182.
- Ziegenhain,C., Vieth,B., Parekh,S., Reinius,B., Guillaumet-Adkins,A., Smets,M., Leonhardt,H., Heyn,H., Hellmann,I. and Enard,W. (2017) Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell*, **65**, 631–643.
- Pollen,A., Nowakowski,T., Shuga,J., Wang,X., Leyrat,A., Lui,J., Li,N., Szpankowski,L., Fowler,B., Chen,P. *et al.* (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.*, **32**, 1053–1058.
- Klein,A., Mazutis,L., Akartuna,I., Tallapragada,N., Veres,A., Li,V., Peshkin,L., Weitz,D. and Kirschner,M. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
- Macosko,E., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A., Kamitaki,N., Martersteck,E. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
- Stegle,O., Teichmann,S. and Marioni,J. (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.*, **16**, 133–145.
- Trapnell,C. (2015) Defining cell types and states with single-cell genomics. *Genome Res.*, **25**, 1491–1498.
- Brennecke,P., Anders,S., Kim,J., Kotodziejczyk,A., Zhang,X., Proserpio,V., Baying,B., Benes,V., Teichmann,S., Marioni,J. *et al.* (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods*, **10**, 1093–1095.
- Efroni,I., Ip,P., Nawy,T., Mello,A. and Birnbaum,K. (2015) Quantification of cell identity from single-cell gene expression profiles. *Genome Biol.*, **16**, 9.
- Efroni,I., Mello,A., Nawy,T., Ip,P., Rahni,R., DelRose,N., Powers,A., Satija,R. and Birnbaum,K. (2016) Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell*, **165**, 1721–1733.
- Cove,D., Bezanilla,M., Harries,P. and Quatrano,R. (2006) Mosses as model systems for the study of metabolism and development. *Annu. Rev. Plant Biol.*, **57**, 497–520.
- Cove,D. and Knight,C. (1993) The moss *Physcomitrella patens*, a model system with potential for the study of plant reproduction. *Plant Cell*, **5**, 1483–1488.
- Kofuji,R. and Hasebe,M. (2014) Eight types of stem cells in the life cycle of the moss *Physcomitrella patens*. *Curr. Opin. Plant Biol.*, **17**, 13–21.
- Ishikawa,M., Murata,T., Sato,Y., Nishiyama,T., Hiwatashi,Y., Imai,A., Kimura,M., Sugimoto,N., Akita,A., Oguri,Y. *et al.* (2011) *Physcomitrella* cyclin-dependent kinase A links cell cycle reactivation to other cellular changes during reprogramming of leaf cells. *Plant Cell*, **23**, 2924–2938.
- Sakakibara,K., Reisewitz,P., Aoyama,T., Friedrich,T., Ando,S., Sato,Y., Tamada,Y., Nishiyama,T., Hiwatashi,Y., Kurata,T. *et al.* (2014) WOX13-like genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*. *Development*, **141**, 1660–1670.
- Li,C., Sako,Y., Imai,A., Nishiyama,T., Thompson,K., Kubo,M., Hiwatashi,Y., Kabeya,Y., Karlson,D., Wu,S. *et al.* (2017) A Lin28 homologue reprograms differentiated cells to stem cells in the moss *Physcomitrella patens*. *Nat. Commun.*, **8**, 14242.
- Nishiyama,T., Miyawaki,K., Ohshima,M., Thompson,K., Nagashima,A., Hasebe,M. and Kurata,T. (2012) Digital gene expression profiling by 5'-end sequencing of cDNAs during reprogramming in the moss *Physcomitrella patens*. *PLoS One*, **7**, e36471.
- Sato,Y., Sugimoto,N., Hirai,T., Imai,A., Kubo,M., Hiwatashi,Y., Nishiyama,T. and Hasebe,M. (2017) Cells reprogramming to stem cells inhibit the reprogramming of adjacent cells in the moss *Physcomitrella patens*. *Sci. Rep.*, **7**, 1909.
- Kivioja,T., Vähäräutio,A., Karlsson,K., Bonke,M., Enge,M., Linnarsson,S. and Taipale,J. (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods*, **9**, 72–74.
- Rensing,S., Lang,D., Zimmer,A., Terry,A., Salamov,A., Shapiro,H., Nishiyama,T., Perroud,P., Lindquist,E., Kamisugi,Y. *et al.* (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*, **319**, 64–69.
- Kubo,M., Imai,A., Nishiyama,T., Ishikawa,M., Sato,Y., Kurata,T., Hiwatashi,Y., Reski,R. and Hasebe,M. (2013) System for stable  $\beta$ -estradiol-inducible gene expression in the moss *Physcomitrella patens*. *PLoS One*, **8**, e77356.
- Nishiyama,T., Fujita,T., Shin-I,T., Seki,M., Nishide,H., Uchiyama,I., Kamiya,A., Carninci,P., Hayashizaki,Y., Shinozaki,K. *et al.* (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8007–8012.
- Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.*, **17**, 10–12.

28. Bolger, A., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
29. Li, B. and Dewey, C. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
30. Lang, D., Ullrich, K., Murat, F., Fuchs, J., Jenkins, J., Haas, F., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R. *et al.* (2018) The *P. patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.*, **93**, 515–533.
31. Jiang, P., Thomson, J. and Stewart, R. (2016) Quality control of single-cell RNA-seq by SinQC. *Bioinformatics*, **32**, 2514–2516.
32. Sun, J., Nishiyama, T., Shimizu, K. and Kadota, K. (2013) TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*, **14**, 219.
33. Robinson, M. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.
34. Robinson, M., McCarthy, D. and Smyth, G. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
35. Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.
36. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N., Livak, K., Mikkelsen, T. and Rinn, J. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, **32**, 381–386.
37. de Siqueira Santos, S., Takahashi, D., Nakata, A. and Fujita, A. (2014) A comparative study of statistical methods used to identify dependencies between gene expression signals. *Brief. Bioinform.*, **15**, 906–918.
38. Owens, T. and Poole, R. (1979) Regulation of cytoplasmic and vacuolar volumes by plant cells in suspension culture. *Plant Physiol.*, **64**, 900–904.
39. Hillwig, M., Contento, A., Meyer, A., Ebany, D., Bassham, D. and Macintosh, G. (2011) RNS2, a conserved member of the RNase T2 family, is necessary for ribosomal RNA decay in plants. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1093–1098.
40. Barthelson, R., Lambert, G., Vanier, C., Lynch, R. and Galbraith, D. (2007) Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics*, **8**, 340.
41. Slane, D., Kong, J., Schmid, M., Jürgens, G. and Bayer, M. (2015) Profiling of embryonic nuclear vs. cellular RNA in *Arabidopsis thaliana*. *Genom Data*, **4**, 96–98.
42. Kalderon, D., Roberts, B., Richardson, W. and Smith, A. (1984) A short amino acid sequence able to specify nuclear location. *Cell*, **39**, 499–509.
43. Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. (1996) Engineered GFP as a vital reporter in plants. *Curr. Biol.*, **6**, 325–330.
44. Jefferson, R., Burgess, S. and Hirsh, D. (1986) Beta-glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 8447–8451.
45. Brandt, S., Kehr, J., Walz, C., Imlau, A., Willmitzer, L. and Fisahn, J. (1999) Technical advance: a rapid method for detection of plant gene transcripts from single epidermal, mesophyll and companion cells of intact leaves. *Plant J.*, **20**, 245–250.
46. Ramsköld, D., Luo, S., Wang, Y., Li, R., Deng, Q., Faridani, O., Daniels, G., Khrebtkova, I., Loring, J., Laurent, L. *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.*, **30**, 777–782.
47. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B., Siddiqui, A. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, **6**, 377–382.
48. Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V., Lao, K. and Surani, M. (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.*, **5**, 516–535.
49. Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K., Imai, T. and Ueda, H. (2013) Quartz-Seq: a highly reproducible and sensitive single-cell RNA-Seq reveals non-genetic gene expression heterogeneity. *Genome Biol.*, **14**, R31.
50. Aird, D., Ross, M., Chen, W., Danielsson, M., Fennell, T., Russ, C., Jaffe, D., Nusbaum, C. and Gnirke, A. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.*, **12**, R18.
51. External RNA Controls Consortium (2005) Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics*, **6**, 150.
52. Wang, Y. and Navin, N. (2015) Advances and applications of single-cell sequencing technologies. *Mol. Cell*, **58**, 598–609.
53. Arguel, M., LeBrigand, K., Paquet, A., García, Ruiz, Zaragoza, S., Barbry, L. and Waldmann, R. (2017) A cost effective 5' selective single cell transcriptome profiling approach with improved UMI design. *Nucleic Acids Res.*, **45**, e48.
54. Harbers, M., Kato, S., de Hoon, M., Hayashizaki, Y., Carninci, P. and Plessy, C. (2013) Comparison of RNA- or LNA-hybrid oligonucleotides in template-switching reactions for high-speed sequencing library preparation. *BMC Genomics*, **14**, 665.
55. Picelli, S., Björklund, A., Faridani, O., Sagasser, S., Winberg, G. and Sandberg, R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
56. Faircloth, B.C. and Glenn, T.C. (2012) Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE*, **7**, e42543.
57. Efroni, I. and Birnbaum, K. (2016) The potential of single-cell profiling in plants. *Genome Biol.*, **17**, 65.
58. Lun, A.T., Bach, K. and Marioni, J.C. (2016) Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.*, **17**, 75.
59. Bacher, R. and Kendziorski, C. (2016) Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.*, **17**, 63.
60. Yachie, N., Natsume, T., Takahashi, K., Katayama, T., Sakurada, T., Kanda, G., Takagi, E., Hirose, T., Katsura, T., Moriya, T. *et al.* (2017) Robotic crowd biology with Maholo LabDroids. *Nat. Biotechnol.*, **35**, 310–312.