The cytoskeletal linker desmoplakin mediates reprogramming and regeneration

Jeongmin Ha1,2†, Beomseok Kim3†, Byungkuk Min4, Juhyeon Nam1,2, Jae-Geun Lee2,5, Minhyung Lee1, Byoung-Ha Yoon6, Yoon Ha Choi3, Ilkyun Im7, Jung Sun Park4, Hyosun Choi8, Areum Baek1, Sang Mi Cho9, Mi-Ok Lee1,2, Ki-Hoan Nam9, Ji Young Mun10, Mirang Kim2,6, Mi Young Son1,2, Yong-Kook Kang2,4, Jeong-Soo Lee5,11\*, Jong Kyoung Kim3\*, Janghwan Kim1,2\*

1Stem Cell Convergence Research Center, Korea Research Institute Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea.

2Department of Functional Genomics, KRIBB school of Bioscience, Korea University of Science & Technology (UST), Daejeon 34113, Republic of Korea.

3Department of New Biology, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Republic of Korea.

4Development and Differentiation Research Center, KRIBB, Daejeon 34141, Republic of Korea.

5Disease Target Structure Research Center, KRIBB, Daejeon 34141, Republic of Korea.

6 Personalized Genomic Medicine Research Center, KRIBB, Daejeon 34141, Republic of Korea.

7Department of Predictive Toxicology, Korea Institute of Toxicology, Daejeon 34114, Republic of Korea.

8BK21 Plus Program, Department of Senior Healthcare, Graduate School, Eulji University, Daejeon, 34824, Republic of Korea.

9Laboratory Animal Resource Center, KRIBB, Cheongju 28116, Republic of Korea.

10Neural Circuit Research Group, Korea Brain Research Institute, Daegu 41062, Republic of Korea.

11Dementia DTC R&D Convergence Program, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea.

**†**These authors contributed equally to this work.

**\*Corresponding author. E-mail: janghwan.kim@kribb.re.kr (J.K.); jkkim@dgist.ac.kr (J.K.K.); jeongsoo@kribb.re.kr (J.-S.L.)**

**SUMMARY:**

Amphibians and fishes show considerable regeneration potential via dedifferentiation of somatic cells into blastema cells1. In terms of dedifferentiation, *in vitro* cellular reprogramming has been suspected to share common process with *in vivo* tissue regeneration2,3, although the details are elusive. Here we identify the cytoskeletal linker protein desmoplakin (Dsp) as a common factor mediating both reprogramming and regeneration. Single cell RNA sequencing revealed that *Dsp* was enriched in distinct intermediate cells during *in vitro* reprogramming. Knockdown of *Dsp* impeded *in vitro* reprogramming into induced pluripotent stem cells and induced neural stem cells as well as *in vivo* regeneration of zebrafish fins. Notably, reduced *Dsp* expression impaired the formation of the intermediate cells and inhibited the expression of blastema-specific genes. These findings suggest that there was a Dsp-mediated cellular link between mammalian *in vitro* reprogramming and *in vivo* regeneration.

**Body Text:**

Ectopic expression of the reprogramming factors *Oct3/4*, *Sox2*, *Klf4*,and *c-Myc* (OSKM) promotes the dedifferentiation of somatic cells into induced pluripotent stem cells (iPSCs)4. iPSC reprogramming (iPSCR) involves removal of somatic cell signatures and acquisition of pluripotency5, however, some developmental genes, particularly epidermis-related genes, are temporally up-regulated6-8 and distinct cell populations, different from iPSCs, occur transiently9-11, indicating that iPSCR is not a simple reversion of normal developmental differentiation. Amphibians and fishes can regenerate entire missing body parts via dedifferentiation of wounded cells, which form a blastema, a mass of undifferentiated pluripotent stem cells that retain a memory of their tissue origin, and re-differentiate into the cell types of the damaged tissues or organs12. Upregulation of OSKM or pluripotency factors is observed in developing blastema cells and these factors are required for regeneration2,3,13. In mammalian cells, temporary exposure of OSKM can convert cells directly into lineage-specific cells14,15. This pluripotency factor-mediated direct reprogramming (PDR) is thought to involve intermediate cells, which are distinct from iPSCs, and not yet fully characterized.

**Common path of reprogramming**

To address the mechanism of OSKM-mediated reprogramming and potentially expand our understanding of *in vivo* regeneration, we decided to characterize intermediate cells. To investigate cell status induced by temporary activation of OSKM *in vitro*, we conducted comparative analysis of iPSCR and PDR, using secondary mouse embryonic fibroblasts (2° MEFs) that express *OSKM* in response to doxycycline (dox). We reprogrammed the 2° MEFs to iPSCs (± LIF treatment until day 5), induced neural stem cells (iNSCs), and dopaminergic neuronal progenitors (iDPs)16,17 (Fig. 1a). LIF did not seem to affect the early phase of iPSCR and the four reprogramming regimes generated appropriate target cells that expressed target cell-specific markers (Extended Data Fig. 1a–c). For comparative transcriptomic analysis of these four reprogramming processes, RNA samples were collected every two days (Extended Data Fig. 2a). The global correlation, determined by principal component analysis (PCA), revealed that all samples were very similar until day 6 (cluster I) (Fig. 1b). Samples after day 6 fell into two distinct groups, iPSCR (cluster II) and PDR (cluster III). Pearson’s correlation heat map analysis also revealed three independent clusters (Extended Data Fig. 2b), suggesting that iPSCR and PDR went through a common path until day 6.

**Distinct signature of intermediate phase**

As somatic cell signatures were lost and the pluripotency-related genes not yet activated (Extended Data Fig. 1b), we considered day 6 as the “intermediate phase” in our 2° MEF reprogramming system. To identify the genes contributed to the establishment of the intermediate phase, we analyzed differentially expressed genes (DEGs) that changed by more than three-fold between day 0–6 (Extended Data Fig. 2c, d, Supplementary Table 1). Clustering the genes into four groups by k-means clustering revealed that expression of commonly down-regulated genes decreased rapidly and remained low until the end of reprogramming. Expression of commonly up-regulated genes was differentially clustered as reprogramming proceeded (Extended Data Fig. 2e, f). We classified the commonly up-regulated genes in early and intermediate phases (between day 0–6) into four groups based on their expression patterns after day 6 (Fig. 1c, d, Extended Data Fig. 2g, h, Supplementary Table 2). The Group IV covered more than 70% of the genes analyzed and represented the genes showed intermediate phase-specific expression pattern. These results were also observed in another reprogrammable cell system (4F2A MEFs) (Extended Data Fig. 3a–d, Supplementary Table 3). Therefore, most genes activated by OSKM were involved mainly in the intermediate phase.

To further characterize Group IV genes showing an intermediate phase-specific expression pattern in both iPSCR and PDR, we performed network-based analyses of Gene Ontology (GO) terms using ClueGO on Group IV genes commonly expressed both in the 2° MEF and 4F2A MEF systems. The 244 genes were rarely expressed in the starting fibroblasts and end-point cells, i.e., iPSCs, iNSCs, and iDPs, but were highly expressed on day 6 (Extended Data Fig. 4a–c, Supplementary Table 4). The top-ranked group included GO terms such as keratinocyte differentiation and epidermis development, and the next highest ranked group included terms related to the desmosome, an intercellular junction in the epidermis18 (Extended Data Fig. 4d, Supplementary Table 5). We used gene set enrichment analysis (GSEA) to characterize the global expression patterns of intermediate phase-specific genes, comparing day 6 samples to starting fibroblasts and end-point cells in both the 2° MEF and 4F2A MEF systems. Ranking the top 15 GO terms by normalized enrichment score (NES) revealed terms associated with the epidermis (Extended Data Fig. 4e, Supplementary Table 6), consistent with the ClueGO results (Extended Data Fig. 4d). Interestingly, the highest ranked gene set was “desmosome” (Fig. 1e). Thus, epidermis-related signatures, in particular desmosome-related signatures, are common characteristics of intermediate phase-specific genes in both iPSCR and PDR.

**Transient formation of desmosomes**

The epidermis is a type of epithelial tissue. Since mesenchymal-to-epithelial transition (MET) is an essential process in the early phase of reprogramming19, we compared the expression of desmosome-related genes with epithelial and mesenchymal genes to investigate whether the transient upregulation of desmosome-related genes is part of the MET process. Among the epithelial genes, non-desmosomal component genes (*Cdh1*, *Ep-cam, and Cldn6*) showed increased expression in all samples on day 6 and in iPSCs, whereas mesenchymal genes (*Snail*, *Slug*, *Zeb1*, *Twist1*, *Twist2*, *Cdh2*, and *Fn*) showed decreased expression in all samples. However, desmosomal component genes (*Dsp* and *Pkp1*) and cytokeratins (*Krt8* and *Krt19*) were increased in all samples on day 6, but reduced in iPSCs, suggesting that desmosome-related epithelial genes are regulated differently from the epithelial genes involved in MET, such as *Cdh1* (Extended Data Fig. 5a). These results reveal that desmosomal component genes are controlled differently from the other epithelial genes.

Desmosomes are composed of desmosomal cadherins, i.e., desmocollins (Dsc) and desmogleins (Dsg), the armadillo proteins, i.e., plakoglobins and plakophilins, and the cytoskeletal linker protein, desmoplakin (Dsp). To confirm the formation of desmosomes in the intermediate phase, we further examined the expression of desmosomal component during reprogramming. All desmosomal components were up-regulated during reprogramming, and that desmosomal proteins were expressed specifically on day 6 of iPSCR and iNSCR, except desmoglein2 (*Dsg2*) which was mainly expressed in iPSCs20,21 (Extended Data Fig. 5b–d). Temporal expression of desmosomal components also occurred during reprogramming of 4F2A MEFs and 2° B-cells into iPSCs (Extended Data Fig. 5e, f). We assessed the intermediate phase-specific formation of desmosomes by transmission electron microscopy (TEM) and mature desmosomes are only found in cells on day 6 (Fig. 1f, Extended Data Fig. 5g, h). These results suggest that the formation of mature desmosomes is an intermediate phase-specific feature of OSKM-mediated reprogramming.

**The role of Dsp in reprogramming**

To investigate whether the formation of mature desmosome is necessary to proceed OSKM-mediated reprogramming, we suppressed Dsp expression using IPTG-inducible short hairpin RNAs against *Dsp* (sh*Dsp*) during iPSCR and iNSC reprogramming (iNSCR) (Extended Data Fig. 6a)*.* Dsp is essential for the formation of mature desmosomes22, linking intermediate filament networks to the desmosomal plaque23,24. *Dsp* knockdown (KD) reduced the efficiency of iPSC generation, with the number of colonies positive for alkaline phosphatase (AP) and Nanog, markers for pluripotency, being significantly decreased (Fig. 2a, b). *Dsp* KD also significantly reduced iNSC generation (Fig. 2c). When we over-expressed *DSP* during iPSCR and iNSCR of human cells, the reprogramming efficiencies were significantly increased (Extended Data Fig. 6b–d). Thus, Dsp appears to be an essential factor controlling the efficiency of OSKM-mediated reprogramming.

Given that Dsp was exclusively expressed on day 6 of reprogramming, and that iPSCs and iNSCs did not express Dsp (Extended Data Fig. 5d), we hypothesized that a distinct cell population requiring Dsp expression should appear around day 6, at the time intermediate cells (ICs) emerge. Indeed, on day 6 we observed cell aggregates that were distinct from iPSC or iNSC colonies (Extended Data Fig. 1a), and the number of the cell aggregates decreased significantly upon *Dsp* KD (Fig. 2d, e).

**Two types of intermediate cells**

To further characterize the cell aggregates that appeared only in the intermediate phase, we stained the cell aggregates with antibodies to desmosomal components, Dsp and Dsc3 (Fig. 2f, g). We found two different cell populations: those co-expressed Dsp and Dsc3, named “desmosomal-component expressing intermediate cells (dICs)”, and those that did not express Dsp and Dsc3, which had a granular morphology, named “granular intermediate cells (gICs)”. These two types of ICs often observed together in one cell aggregate. Moreover, since gICs were not formed after Dsp KD even though they did not express Dsp (Fig. 2d–g), we hypothesized that gICs might originate from dICs during reprogramming. Retrospective tracing from the established colonies by iPSCR and iNSCR supported the possibility of the same hierarchical relationship between dICs and gICs (Extended Data Fig. 6e, f).

To investigate reprogramming process, in particular the intermediate phase, in depth, we established 33,966 single cell libraries from iPSCR- and iNSCR-derived cells, as well as the starting fibroblasts, iPSCs, and iNSCs, and then performed single cell RNA sequencing (scRNAseq) (Extended Data Fig. 7a, b). As seen in the microarray analysis (Fig. 1b), the reprogramming trajectories begin to bifurcate from day 6 onwards (Extended Data Fig. 7c, d). Previously, there are concerns of the existence of iPSC-like pluripotent intermediates in our PDR approach25,26. However, we were unable to detect any single cells showing iPSC-like expression of pluripotency genes (Extended Data Fig. 8a, b). While we observed high expression of *Oct4* in part of cells, 11 cells co-expressed only with *Dppa5a* even from day 8(Extended Data Fig. 8a, Supplementary Table 7). Thus, ICs temporarily appearing in OSKM-mediated reprogramming are a unique cell population that is distinct from iPSCs.

To further understand the characteristics of IC populations, we focused on cells that existed between days 5–7. We observed very similar cellular distributions until day 6 in both iPSCR and iNSCR (Fig. 3a). A previously published scRNAseq analysis of iPSCR has suggested that *Shisa8* is an early marker of the successful MET27. In our scRNAseq analysis, the *Shisa8*-positive population formed two sub-populations, distinguished by *Dsp* expression (Fig. 3b). Further analysis of gene set expression revealed the rare expression of pluripotency markers in all cells and the elevated expression of stromal cell signatures (e.g., SASP and MEF identity) in *Shisa8*-negative cells (Fig. 3c). These results suggest that the *Shisa8*+/*Dsp*+ ICs on successfully reprogrammed path on UMAP are dICs.

Given the gICs were highly proliferative (Movie S1), the distinctive cell cycle signature of *Shisa8*+/*Dsp*- cells looked correlated to gICs (Fig. 3b, c). To confirm this idea, we further characterized gICs. We found that the gICs were easily detached from the culture dish and formed additional iPSC-like colonies (Movie S2). This feature reminded us the ‘satellite iPSCs’ that appear during reprograming with the 2° MEF system28. When we isolated the floating gICs from the culture on day 6, re-attached to the culture dish, and continued reprogramming, the gICs were converted into iPSCs (Fig. 3d). Although they had potential to be iPSCs, the gICs rarely expressed pluripotency markers, i.e., *Nanog, Rex1,* and *Esrrb* (Fig. 3e, Extended Data Fig. 8c), suggesting that gICs are unique cells distinct from iPSCs. In isolated gICs, *Dsp* expression remained low, but *Shisa8* and *Cdk1* were highly expressed (Fig. 3e), indicating that the *Shisa8*+/*Dsp*- population are gICs (Fig. 3b).To further investigate the relationship between the dICs and gICs, we used the pseudotime algorithm Slingshot29 (Fig. 3f). The pseudotime analysis revealed that the dIC population (Cluster 0; *Shisa8*+/*Dsp*+) was divided into two branches by *Shisa8* expression. In the *Shisa8*-positive branch, dICs converted into *Shisa8*+/*Dsp*- populations (clusters 1 and 6), suggesting that gICs are derived from dICs and these two IC populations are in the *Shisa*8+ successful reprogramming path.

**Dsp-Akt signaling axis in reprogramming**

Because Dsp is a junctional molecule, we wondered how Dsp modulates reprogramming processes. Dsp expression is known to affect Akt22,30. Indeed, *Dsp* KD resulted in a significant reduction of phosphorylated Akt in both iPSCR and iNSCR (Extended Data Fig. 9a). Given that *Dsp* KD reduced the number of IC aggregates (Fig. 2d, e) and Akt signaling activity (Extended Data Fig. 9a), we hypothesized that Dsp promotes the formation of IC aggregates via Akt signaling. To test this idea, we used the small molecule MK2206 to inhibit Akt signaling31 during iPSCR and iNSCR. MK2206 reduced phosphorylated Akt levels, substantially diminished the number of IC aggregates (Extended Data Fig. 9b–d), and significantly decreased the efficiencies of iPSCR and iNSCR (Extended Data Fig. 9e, f). Given MK2206 did not affect the self-renewal of iPSCs and iNSCs (Extended Data Fig. 9g–j), we conclude that Dsp-Akt signaling is required for the generation of IC populations.

**The role of Dsp in regeneration**

During *in vivo* regeneration in axolotls and fishes, reprogramming factors are transiently up-regulated2,3,13. The blastema cells, which mediate *in vivo* regeneration, are multipotent, producing various types of three germ layer cells in the damaged tissues3 and even regenerate missing body parts. Thus, we hypothesized that the IC populations, requiring temporary expression of the reprogramming factors and producing all three germ layer cells in PDR15,16,32,33, are comparable to blastema cells. To confirm the relations between ICs and blastema, we sought to confirm that Dsp would play an essential role in *in vivo* regeneration, as in *in vitro* reprogramming. To assess whether dsp expression affects zebrafish larvae fin fold regeneration, we injected the *dspa-* and *dspb-*morpholino (*dsp* mix MO) into one-cell eggs (Extended Data Fig. 10a). *dsp* morphant embryos showed the reduction of fin fold regrowth (~26%) within 24 hours (Fig. 4a). An adult fin regrowth model also revealed that *dsp* KD in the amputated tips of dorsal fins impeded dorsal regrowth within 24 hours, to 68% of the regeneration seen in untreated ventral fin tips (Fig. 4b). This reduction was similar to that seen in transfection of the *pou5f3*/*oct4* morpholino (Fig. 4b), as previously reported3. Whole-mount *in situ* hybridization analysis revealed a substantial decrease in the blastema marker gene (*junbb*) in amputated *dsp* morphant embryos (Fig. 4c, Extended Data Fig. 10b, Supplementary Table 8). Moreover, other blastema marker genes (*msx1b, fgf20a, her6,* and *rpl13a*) also decreased in amputated *dsp* morphant embryos (Extended Data Fig. 10c). Altogether, these results suggest that Dsp contributes to blastema-mediated *in vivo* regeneration as it does in IC-mediated *in vitro* reprogramming.

*In vivo* regeneration observed in amphibians and fishes is still mysterious and it has long been questioned why mammalian might lost this regeneration potential. Here, we intensely analyzed the intermediate phase of *in vitro* induced reprogramming and found two distinct IC populations, dICs and gICs showing differential expression of Dsp, are formed, and *Dsp* knockdown critically affected the efficiency of iPSCR and iNSCR. Although we could not confirm these IC populations are mammalian version of blastema cells, we found that dsp also plays an essential role during zebrafish regeneration through the modulation of blastema induction. Thus, our results imply that OSKM-mediated reprogramming, in particular PDR, could be an evolutionary relic of *in vivo* regeneration and might provide a revolutionary clue for mammalian artificial blastema cells.

**Methods**

**Reprogramming of MEF to iPSCs, iNSCs, and iDPs**

Secondary mouse embryonic fibroblasts (2° MEFs) and 4F2A MEFs were prepared as previously described16,34,35. For iPSC reprogramming, 2° MEFs were thawed (p2) and plated (p3) on Geltrex (Thermo Fisher Scientific, Waltham, MA, USA)-coated culture dishes at 2×104 cells/cm2 in MEF derivation culture medium (MEF medium: Dulbecco’s modified Eagle’s medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific), 1X Glutamax (Thermo Fisher Scientific), 1X MEM-NEAA (Thermo Fisher Scientific), and 1% Penicillin/Streptomycin (Thermo Fisher Scientific)). Doxycycline (Sigma-Aldrich, St. Louis, MO, USA) induction was initiated the next day and continued the designated day. The cells were cultured in MEF medium for an additional day and the medium was changed to reprogramming initiation medium (RepM-Ini: KnockOut DMEM (Thermo Fisher Scientific) supplemented with 10% KnockOut Serum Replacement (Thermo Fisher Scientific), 5% fetal bovine serum, 1X Glutamax, 1X MEM-NEAA, 1% 1X Penicillin/Streptomycin, and 0.055 mM β-mercaptoethanol (Thermo Fisher Scientific)) or an pluripotent stem cell reprogramming medium (RepM-PSC: same as RepM-Ini containing 1,000 U/mL mLIF; Millipore, Billerica, MA, USA). For iNSC and iDP reprogramming, the medium was changed to neural stem cell reprogramming medium (RepM-NSC: advanced DMEM/F12 (Thermo Fisher Scientific) and neurobasal (Thermo Fisher Scientific) were mixed at 1:1 and supplemented with 0.05% BSA, 1X N2 (Thermo Fisher Scientific), 1X B27 (Thermo Fisher Scientific), 1X Glutamax, 0.11 mM β-mercaptoethanol with 20 ng/mL FGF2 (Peprotech, Rocky Hill, NJ, USA), 2 ng/mL FGF4 (Peprotech), and 20 ng/ mL EGF (Peprotech)) and dopaminergic progenitor reprogramming medium (RepM-DP: advanced DMEM/F12 and neurobasal were mixed by 1:1 and supplemented with 0.05% BSA, 1X N2, 1X B27, 1X Glutamax, 0.11 mM β-mercaptoethanol with 200 ng/mL SHH (Peprotech), and 100 ng/mL FGF8b (Peprotech)) after day five16,17.

For knockdown of *Dsp*, IPTG (1 mM; Sigma-Aldrich) was treated for six days from the day after plating.

gICs were isolated from the supernatant at day 6 of iPSCR and iNSCR. The harvested supernatant centrifuged at 300 ×g for 3 min, washed with Dulbecco’s phosphate-buffered saline (DPBS; WELGENE, Daegu, Korea), and re-attached on geltrex-coated culture plates at 2×104 cells/cm2 in RepM-PSC with doxycycline.

For MEF derivation, husbandry, animal care, and blastocyst injection were performed in accordance with guidelines from the Korea Research Institute of Bioscience and Biotechnology (KRIBB) and approved by KRIBB-IACUC (approval number: KRIBB-AEC-20044).

**RNA preparation, cDNA synthesis, and quantitative PCR**

Total RNA was extracted from the samples using the RNeasy Plus mini kit with QIAshredder (Qiagen, Hilden, Germany) and cDNA was synthesized from 1 μg of total RNA using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). Quantitative PCR was performed using a 1/50 concentration of the obtained cDNA with iQ SYBR Green Supermix (Bio-Rad) on an Applied Biosystems 7500 Fast Real-Time PCR instrument system (Thermo Fisher Scientific). The cycle threshold (Ct) value for each target gene was determined using a software provided by the manufacturer. The expression data were normalized to Ct value of *Rpl7*32,36. The primer sequences used in this study are listed in Supplementary Table 9.

**Alkaline phosphatase staining**

Samples were washed once with DPBS and fixed with 10% formalin solution (Sigma-Aldrich) for 30 sec. Alkaline phosphatase (AP) staining was performed with the Leukocyte Alkaline Phosphatase kit (Sigma-Aldrich) according to the manufacturer’s instruction. Briefly, fixed samples were washed once and then incubated with an AP substrate solution for 20 min in the dark.

**Immunofluorescence staining**

Cells were grown in 24- or 4-well tissue culture plates and washed once with DPBS. Samples were fixed with 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA) and 0.15% picric acid (Sigma-Aldrich) in DPBS for 15 min. After then blocked and permeabilized with 3% bovine serum albumin (BSA; Thermo Fisher Scientific) and 0.3% Triton X-100 (Sigma-Aldrich) in DPBS for 1 hour at room temperature. All primary antibodies were diluted in 1% BSA and incubated overnight at 4 °C. After iterative washing with 0.1% BSA in DPBS, the samples were incubated with Alexa-594- or Alexa-488-conjugated secondary antibodies (Thermo Fisher Scientific) for 1 hour at room temperature. All fluorescent images were acquired using an Axio Vert.A1 microscope (Carl Zeiss, Oberkochen, Germany) and an Evos FL auto 2 imaging system (Thermo Fisher Scientific). The antibodies used in this study are listed in Supplementary Table 10.

**Microarray analysis**

Total RNA was extracted using the RNeasy Plus mini kit with QIAshredder (Qiagen) and global gene expression profiles were analyzed by an Agilent Mouse Whole Genome 4 × 44K arrays (V2) chip (one-color platform; Agilent Technologies, Santa Clara, CA, USA). Briefly, RNA quality of all samples was checked by the Agilent 2100 Bioanalyzer System, followed by amplification, labeling, and hybridization steps. All experiments were performed according to the manufacturer’s protocols.

The microarray data were processed using GeneSpring software (Agilent Technologies) and normalized using global scale normalization. To analyze functionally grouped GO terms, we used the Cytoscape software platform (v3.5.1; <http://www.cytoscape.org/what_is_> cytoscape.tml). A functionally grouped GO term network was visualized by ClueGO plug-in (v2.3.4; <http://apps.cytoscape.org/apps/cluego>) and CluePedia (v1.3). ClueGO analysis of ranked the 244 genes into five annotation groups (group p-value < 0.002)37. The enrichment of genes in biological process, cellular component, and molecular function was analyzed by GSEA software (v3.0)38. Parameters for GSEA were set as 1,000 permutations of gene sets, a classic enrichment statistic and signal-to noise separation metric. Starting cells contained 2° MEFs and 4F2A MEFs; Day 6 samples contained P6, P’6, N6, and D6 in 2° MEF system and P6 and N6 in 4F2A MEF system; Endpoint cells contained P12, P’12, N12, and D12 in 2° MEF system and iPSC and iNSC in 4F2A MEF system.

**Western blot**

Whole cell extracts were prepared using RIPA buffer (Sigma-Aldrich) containing 1 mM PMSF (Sigma-Aldrich) and a cocktail of protease inhibitors (Roche, Basel, Switzerland) and then centrifuged at 7,000 ×g for 5 min at 4 °C. The protein concentrations were determined using a Pierce BCA protein assay kit (Thermo Fisher Scientific). An equal amount of total protein was separated on MP TGX Precast Gels (Bio-Rad) and transferred to polyvinylidene fluoride membranes (Bio-Rad). The membranes were blocked in Tris-buffered saline (LPS Solution, Daejeon, Korea) containing 0.05% Tween-20 (TBST; Sigma-Aldrich) with 5% non-fat milk (BD Biosciences, Franklin Lakes, NJ, USA) or 1.5% BSA for 1 hour at room temperature and then incubated with specific primary antibodies overnight at 4 °C. After washing with TBST six times for 30 min, the samples were incubated for 1 hour at room temperature with anti-rabbit or anti-mouse horseradish peroxidase-conjugated secondary antibodies (Cell Signaling Technology, Danvers, MA, USA; 1:5000). The blots were developed using ECL Select Western Blotting Detection Reagent (GE Healthcare, Little Chalfont, UK). The antibodies used in this study are listed in Supplementary Table 10.

**Transmission electron microscopy imaging**

To analyze desmosome formation, reprogramming was performed on geltrex-coated coverslips (Electron Microscopy Sciences, Hatfield, PA, USA). Cells on the coverslip were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate solution (pH 7.0) for 1 hour, followed by 2% osmium tetroxide for 2 hours at 4 °C. The cells were dehydrated with a graded acetone series and embedded into Spurr medium (Electron Microscopy Services). The samples were sectioned (60 nm) with an ultra-microtome (RMC MTXL; Boeckeler Instruments, Tuscon, AZ, USA), and double-stained with 2% uranyl acetate for 20 min and lead citrate for 10 min. The sections were then viewed under a H-7600 (HITACHI, Tokyo, Japan) transmission electron microscope at 80 kV39.

**Lentivirus production and transduction**

The pLKO-GFP-IPTG-3xLacO construct containing shRNA against mouse *Dsp* (GCCTACAAGAAAGGTCTCATT) was designed by Sigma-Aldrich. The viruses were produced in HEK293T cells. Briefly, the cells at 60‒80% confluence (16 hours after plating) were co-transfected with lentiviral plasmid along with vesicular stomatitis virus coat protein plasmid (pMD2.G; Addgene, Watertown, MA, USA) and packaging plasmid (psPAX2; Addgene) using TransIT-2020 transfection reagent (Mirus Bio, Madison, WI, USA). Transfected cells were washed with DPBS (8-16 hours after transfection) and grown for additional 48 h with MEF medium. Viral supernatants were collected and used to infect 2° MEFs with 6 μg/mL polybrene (Sigma-Aldrich).

**Fluorescence-activated cell sorting (FACS)**

The infected 2° MEFs were isolated using FACS. Cells were dissociated in 0.05% Trypsin-EDTA for 5 min, washed with DPBS, and re-suspended with FACS buffer (0.5% BSA in DPBS) for isolation using a FACSaria cell sorter (BD Biosciences).

**Human iPSC and iNSC reprogramming using SeV**

Human iPSC and iNSC reprogramming were performed as previously described40, with slight modifications. Briefly, human fibroblasts, CRL-2097 (American Type Culture Collection, Manassas, VA, USA), were plated into 24-well plates at a density of 30,000 cells/cm2, and then transduced with SeV mixtures (CytoTune™-iPS 2.0 Sendai reprogramming kit; Thermo Fisher Scientific), according to the manufacturer’s instructions. On the next day, the SeV mixtures were washed with DPBS. For human iPSC reprogramming, the cells were incubated in fibroblast medium consisting of MEM (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum, 1X Sodium pyruvate (Thermo Fisher Scientific), and 1X MEM-NEAA for an additional two days. At 3 days post-transduction (dpt), the medium was replaced with iPSC reprogramming medium consisting of mTeSR-1 (STEMCELL Technologies, Vancouver, Canada) supplemented with 3.0 μM CHIR99021 (Tocris Bioscience, Bristol, UK), 0.5 μM A83-01 (Tocris), and 0.2 mM NaB (Sigma-Aldrich).

For human iNSC reprogramming, the SeV mixtures were washed with DPBS and replaced with human neural reprogramming medium supplemented with 3.0 μM CHIR99021, 0.5 μM A83-01, and 10 ng/mL hLIF (Peprotech). The reprogramming medium was replaced every other day.

For *DSP* overexpression, human fibroblasts were transfected with 1 µg of *DSP* plasmid (Addgene, Cat.32227) and empty plasmid (pEGFP-N1; Takara Bio, Shiga, Japan) at 4 dpt.

For reprogramming purposes, the use of human fibroblasts was exempted from IRB review by the Public Institutional Review Board Designated by Ministry of Health and Welfare (P01-201802-31-001).

**Single-cell RNA sequencing (scRNAseq)**

A total of 33,966 cells passed our stringent quality control criteria, with an average of 3,962 genes detected per cell. scRNAseq libraries were generated using the Chromium Single Cell 3’ Reagent Kit v2 (PN-120267, 10X Genomics), Chromium Single Cell A Chip Kit (PN-120236, 10X Genomics) and Chromium Single Cell i7 Multiplex Kit (PN-120262, 10X Genomics). The cell viability was estimated to be over 90% by trypan blue staining under the microscope. Cells were diluted to 2×105 ~ 2×106 cells/mL with 0.04% BSA in DPBS and loaded on the Chromium microfluidic platform, aiming to capture 3,000 cells per each channel. Subsequent library preparation was performed according to the manufacturer’s instructions. Libraries were sequenced on an Illumina HiSeq 4000 platform (2×100bp).

**scRNAseq data analysis**

Raw FASTQ files were processed with Cell Ranger (v.2.1.0) using default arguments. Reads were aligned to the mouse reference genome (GRCm38) with the Ensembl GRCm38.92 annotation. A gene-by-cell unique molecular identifier (UMI) count matrix for each condition was generated with “expect-cells=3,000”, and aggregated into a single count matrix. Empty droplets were identified and filtered out using the emptyDrops function of the DropletUtils (v0.99) R package41 with FDR ≤ 0.05. By visually inspecting outliers in the principal component analysis (PCA) plot on the quality control metrics using the scater (v.1.7.18) R package42 as described previously43, low-quality cells with less than 1,000 UMIs, with less than 102.5 detected genes (103 detected genes for iNSCs), and with greater than 10% of UMIs mapped to mitochondrial genes were removed. The raw count matrix was normalized by cell-specific size factors estimated by the scran (v1.8.4) R package44, and then log2-transformed with a pseudo-count of 145.

Using the normalized count matrix, scRNAseq profiles of whole cells were scaled and visualized on a UMAP using RunUMAP function in Seurat (v3.1.5) R package46. Gene signature scores were computed by ‘score\_gene\_sets’ function implemented in WOT27 using pre-defined lists of marker genes also employed in Schiebinger et al, and cells were color masked by the signature scores using home-brew R codes to examine the overall dispersion of cell identities throughout the reprogramming processes. To characterize dICs and gICs and perform the trajectory analysis on the selected cells, expression profiles of cells in day 6 were extracted. UMAP plots were generated, and cells were marked with corresponding signature scores for marker genes (*Dsp* and *Shisa8*) in day 6. Next, cells were clustered using FindClusters function in Seurat, and potential trajectory of reprogramming paths were predicted using slingshot (v3.11), a R package for cell lineage inference29. Finally, trajectories of branching lineages were calculated by setting cluster 0, which mainly contains the cells on day 5, as the starting point and plotted on UMAP plots generated earlier.

**Maintaining zebrafish**

Zebrafish (Danio rerio) AB (wild type) strain was maintained at 28.5 °C under a standard condition. Fish were fed daily with a combination of dry food and brine shrimp and maintained under a light schedule of 14 hours light and 10 hours dark. Zebrafish husbandry and animal care were performed in accordance with guidelines from the KRIBB and approved by KRIBB-IACUC (approval number: KRIBB-AEC-17073, KRIBB-AEC-20056).

**Adult caudal fin regeneration**

Adult caudal fin regeneration experiments were performed with 6-9 months old adults as previously described47. Adult zebrafish were adapted at 33 °C day before experiments. Fish were anesthetized in tricaine prior to their caudal fin amputation experiments, and adult fin was amputated at the 7 bony segments distal to the fin girdle. Following the surgery, the fish were returned to a 33 °C tank. The following morpholinos (Gene Tools Inc., Philomath, OR, USA) were used in this study.

*pou5f3* MO : 5`-CGCTCTCTCCGTCATCTTTCCGCTA

*dspa* MO : 5`-AAACTAAAACCGAGGCTGACCTTCT

*dspb* MO : 5`-CTGACTGTGTTTCAGACTGACCTGT

Each morpholino contained a 3-fluorescein tag and was resuspended in water. *dspa* and *dspb* morpholino were mixed by 1:1 ratio to knockdown both desmoplakin genes simultaneously. At 48 hours post amputation, 1 mM morpholinos was injected into the regenerating tissue on the dorsal side of each zebrafish tail fin using PV380 Pneumatic picopump (World Precision Instruments, Sarasota, FL, USA). Each morpholino injection was targeted to the regenerative tissue just distal to each bony ray and approximately 70 nL of morpholino solution was injected per bony ray.

Electroporation of both the dorsal and ventral (to control for a non-specific electroporation effects) sides of the fin using NEPA21 Electroporator (Nepa Gene Co., Ldt., Chiba, Japan) was performed immediately following the injections. Electroporation parameter used three consecutive 50 msec pulses at 15 V with a 50 msec pause between pulses using CUY647 15 mm diameter platinum pate electrode (Protech International Inc., Cornelius, NC, USA). Electroporation was preformed two times using same parameters to increase electroporation efficiency. Fish were then returned to the same 33 °C tank. At 72 hours post-amputation, each fin was photographed using an Olympus SZX16 microscope equipped with TUCSEN Dhyana 400DC digial camera. The area of both the dorsal (D) and ventral (V) fin re-growth was calculated using Image J software (National Institutes of Health, Bethesda, MD, USA). The percentage of regeneration was calculated by ((D3dpa-D2dpa)/(V3dpa-V2dpa))×100 and the statistical significance of the morpholino on regeneration was analyzed using student’s *t*-test.

**Larvae fin fold regeneration**

Zebrafish embryos of AB strain, were maintained in E3 egg water (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, 0.33 mM MgSO4) in a petri dish at 28.5 °C. 0.4 mM control or *dspa* and *dspb* mix morpholinos at approximately 1 nL were injected into one-cell eggs. Zebrafish larvae at 2 dpf were anesthetized with tricaine in egg water and their fin fold was amputated using a surgical razor blade. The fin fold was carefully amputated at same sites just posterior to notochords. For quantification of fin fold regeneration, the lengths from the notochord end to the posterior tip of fin fold at 0 hours post amputation and 24 hours post amputation were measured using image J software and was analyzed using student’s *t*-test.

Zebrafish embryos at each stage were harvested with TRI reagent solution (Thermo Fisher Scientific), followed by purifying total RNA with Direct-zol RNA miniprep kit (Zymo Research, Irvine, CA, USA) and synthesizing cDNA with SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific). The synthesized cDNA was amplified by semi-quantitative PCR using the specific primers that are listed in Supplementary Table 9.

**Whole mount in situ hybridization (WISH)**

WISH in zebrafish embryos was performed as previously reported48. To make *in situ* probes, the DNA templates for zebrafish blastema marker gene, *junbb*, was amplified from cDNA of WT zebrafish using PCR.

Forward probe: 5`-TGGGTTACGGTCACAACGAC

Reverse probe: 5`-CAGTGTCCGTTCTCTTCCGT

PCR products were agarose gel separated and purified, then cloned into pCR™Blunt II-TOPO® vector. Dig-labelled anti-sense probes were in vitro transcribed by SP6 or T7 RNA polymerase kit (Roche) and purified with NucAway spin columns (Thermo Fisher Scientific). Caudal fin of adult and larvae zebrafish for WISH were prepared by fixing with 4% paraformaldehyde in 1X PBS solution, dehydrating using methanol, stored at -20 °C over 30 min, and serially rehydrated with 1X PBST solution. The rehydrated embryos were treated with proteinase K in 1X PBS and post-fixed with 4% paraformaldehyde. The antisense probes were hybridized with the fixed embryos at each developmental stage in hybridizing solution (5 mg/mL torula yeast RNA type VI, 50 ug/mL heparin, 50% formamide, 5X SSC, 0.1% Tween-20, 1 M citric acid used to adjust pH 6.0) at 70 °C overnight. The probes were washed serially using 2X SSCT-F (2X SSCT, 50% formamide, 0.1% Tween20), 2X SSCT (2X SSCT, 0.1% Tween20), 0.2X SSCT (0.2X SSCT, 0.1% Tween20) at 70 °C and 1X PBST at RT. The embryos were blocked with the blocking solution (5% horse serum, 1X PBST) at RT, and the alkaline phosphatase conjugated anti-digoxigenin antibody (Roche) was added into the blocking solution at 4 °C overnight. To detect the expression signal of transcript, NBT/BCIP solution (Roche) was used as alkaline phosphatase (AP) substrate. The expression pattern of transcripts was observed by using Olympus SZX16 microscope and imaged with TUCSEN Dhyana 400DC49. The quantification of WISH was performed according to AP activity in the fin fold. We pixelated raw data using Photoshop (Adobe Systems Inc., San Jose, CA, USA) for the calculation of gene expression, and we count the number of AP-positive (blue) pixels through Image J software. See Supplementary Table 8 for detailed calculation.

**Statistical analysis**

Statistical significance was calculated with the Student’s *t*-test using Microsft Excel (Microsoft Office, WA, USA). For comparisons of two group with equal variance as determined by the *F*-test, and all statistical analyses were two-tailed. Statistically significant differences are indicated as \**P* < 0.05, \*\**P* < 0.01 or \*\*\**P* < 0.001. No statistical methods were used to predetermine sample size. Experimental triplicates were performed for all Reverse-transcription qPCRor semi-PCR analyses.

**Reporting summary**

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability   
Microarray and scRNAseq data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) database under accession numbers: GSE156888 (microarray) and PRJNA555736 (scRNAseq), respectively.

Code availability  
All data were analysed with standard programs and packages as detailed. Scripts can be found at https://github.com/ References and Notes

1 Tsonis, P. A. Regeneration in vertebrates. *Dev Biol* **221**, 273-284, doi:10.1006/dbio.2000.9667 (2000).

2 Maki, N. *et al.* Expression of stem cell pluripotency factors during regeneration in newts. *Dev Dyn* **238**, 1613-1616, doi:10.1002/dvdy.21959 (2009).

3 Christen, B., Robles, V., Raya, M., Paramonov, I. & Izpisua Belmonte, J. C. Regeneration and reprogramming compared. *BMC Biol* **8**, 5, doi:10.1186/1741-7007-8-5 (2010).

4 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).

5 Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet* **14**, 427-439, doi:10.1038/nrg3473 (2013).

6 Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617-1632, doi:10.1016/j.cell.2012.11.039 (2012).

7 O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* **499**, 88-91, doi:10.1038/nature12243 (2013).

8 Kuno, A., Nishimura, K. & Takahashi, S. Time-course transcriptome analysis of human cellular reprogramming from multiple cell types reveals the drastic change occurs between the mid phase and the late phase. *BMC Genomics* **19**, 9, doi:10.1186/s12864-017-4389-8 (2018).

9 Lujan, E. *et al.* Early reprogramming regulators identified by prospective isolation and mass cytometry. *Nature* **521**, 352-356, doi:10.1038/nature14274 (2015).

10 Schwarz, B. A. *et al.* Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. *Cell Stem Cell* **23**, 289-305 e285, doi:10.1016/j.stem.2018.06.013 (2018).

11 Kagawa, H. *et al.* OVOL1 Influences the Determination and Expansion of iPSC Reprogramming Intermediates. *Stem Cell Reports* **12**, 319-332, doi:10.1016/j.stemcr.2018.12.008 (2019).

12 Jopling, C., Boue, S. & Izpisua Belmonte, J. C. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat Rev Mol Cell Biol* **12**, 79-89, doi:10.1038/nrm3043 (2011).

13 Shyh-Chang, N. *et al.* Lin28 enhances tissue repair by reprogramming cellular metabolism. *Cell* **155**, 778-792, doi:10.1016/j.cell.2013.09.059 (2013).

14 Kim, J., Ambasudhan, R. & Ding, S. Direct lineage reprogramming to neural cells. *Curr Opin Neurobiol* **22**, 778-784, doi:10.1016/j.conb.2012.05.001 (2012).

15 Zhu, S., Wang, H. & Ding, S. Reprogramming fibroblasts toward cardiomyocytes, neural stem cells and hepatocytes by cell activation and signaling-directed lineage conversion. *Nat Protoc* **10**, 959-973, doi:10.1038/nprot.2015.059 (2015).

16 Kim, J. *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc Natl Acad Sci U S A* **108**, 7838-7843, doi:10.1073/pnas.1103113108 (2011).

17 Kim, H. S., Kim, J., Jo, Y., Jeon, D. & Cho, Y. S. Direct lineage reprogramming of mouse fibroblasts to functional midbrain dopaminergic neuronal progenitors. *Stem Cell Res* **12**, 60-68, doi:10.1016/j.scr.2013.09.007 (2014).

18 Johnson, J. L., Najor, N. A. & Green, K. J. Desmosomes: regulators of cellular signaling and adhesion in epidermal health and disease. *Cold Spring Harb Perspect Med* **4**, a015297, doi:10.1101/cshperspect.a015297 (2014).

19 Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51-63, doi:10.1016/j.stem.2010.04.014 (2010).

20 Eshkind, L. *et al.* Loss of desmoglein 2 suggests essential functions for early embryonic development and proliferation of embryonal stem cells. *Eur J Cell Biol* **81**, 592-598, doi:10.1078/0171-9335-00278 (2002).

21 Park, J. *et al.* DSG2 Is a Functional Cell Surface Marker for Identification and Isolation of Human Pluripotent Stem Cells. *Stem Cell Reports* **11**, 115-127, doi:10.1016/j.stemcr.2018.05.009 (2018).

22 Gallicano, G. I. *et al.* Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. *J Cell Biol* **143**, 2009-2022, doi:10.1083/jcb.143.7.2009 (1998).

23 Getsios, S., Huen, A. C. & Green, K. J. Working out the strength and flexibility of desmosomes. *Nat Rev Mol Cell Biol* **5**, 271-281, doi:10.1038/nrm1356 (2004).

24 Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. Desmoplakin is essential in epidermal sheet formation. *Nat Cell Biol* **3**, 1076-1085, doi:10.1038/ncb1201-1076 (2001).

25 Bar-Nur, O. *et al.* Lineage conversion induced by pluripotency factors involves transient passage through an iPSC stage. *Nat Biotechnol* **33**, 761-768, doi:10.1038/nbt.3247 (2015).

26 Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat Biotechnol* **33**, 769-774, doi:10.1038/nbt.3270 (2015).

27 Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943 e922, doi:10.1016/j.cell.2019.01.006 (2019).

28 Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol* **28**, 521-526, doi:10.1038/nbt.1632 (2010).

29 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).

30 Wan, H., South, A. P. & Hart, I. R. Increased keratinocyte proliferation initiated through downregulation of desmoplakin by RNA interference. *Exp Cell Res* **313**, 2336-2344, doi:10.1016/j.yexcr.2007.01.010 (2007).

31 Nishimura, K. *et al.* A Role for KLF4 in Promoting the Metabolic Shift via TCL1 during Induced Pluripotent Stem Cell Generation. *Stem Cell Reports* **8**, 787-801, doi:10.1016/j.stemcr.2017.01.026 (2017).

32 Efe, J. A. *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat Cell Biol* **13**, 215-222, doi:10.1038/ncb2164 (2011).

33 Margariti, A. *et al.* Direct reprogramming of fibroblasts into endothelial cells capable of angiogenesis and reendothelialization in tissue-engineered vessels. *Proc Natl Acad Sci U S A* **109**, 13793-13798, doi:10.1073/pnas.1205526109 (2012).

34 Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* **26**, 916-924, doi:10.1038/nbt1483 (2008).

35 Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat Methods* **7**, 56-59, doi:10.1038/nmeth.1410 (2010).

36 Ha, J. *et al.* Simplified Brain Organoids for Rapid and Robust Modeling of Brain Disease. *Front Cell Dev Biol* **8**, 594090, doi:10.3389/fcell.2020.594090 (2020).

37 Son, M. Y. *et al.* Distinctive genomic signature of neural and intestinal organoids from familial Parkinson's disease patient-derived induced pluripotent stem cells. *Neuropathol Appl Neurobiol* **43**, 584-603, doi:10.1111/nan.12396 (2017).

38 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

39 Yamindago, A. *et al.* Acute toxic effects of zinc oxide nanoparticles on Hydra magnipapillata. *Aquat Toxicol* **205**, 130-139, doi:10.1016/j.aquatox.2018.10.008 (2018).

40 Lee, M. *et al.* Efficient exogenous DNA-free reprogramming with suicide gene vectors. *Exp Mol Med* **51**, 82, doi:10.1038/s12276-019-0282-7 (2019).

41 Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, 63, doi:10.1186/s13059-019-1662-y (2019).

42 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186, doi:10.1093/bioinformatics/btw777 (2017).

43 Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* **17**, 29, doi:10.1186/s13059-016-0888-1 (2016).

44 Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* **17**, 75, doi:10.1186/s13059-016-0947-7 (2016).

45 Lee, M. *et al.* Single-cell RNA sequencing identifies shared differentiation paths of mouse thymic innate T cells. *Nat Commun* **11**, 4367, doi:10.1038/s41467-020-18155-8 (2020).

46 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

47 Thummel, R. *et al.* Inhibition of zebrafish fin regeneration using in vivo electroporation of morpholinos against fgfr1 and msxb. *Dev Dyn* **235**, 336-346, doi:10.1002/dvdy.20630 (2006).

48 Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* **3**, 59-69, doi:10.1038/nprot.2007.514 (2008).

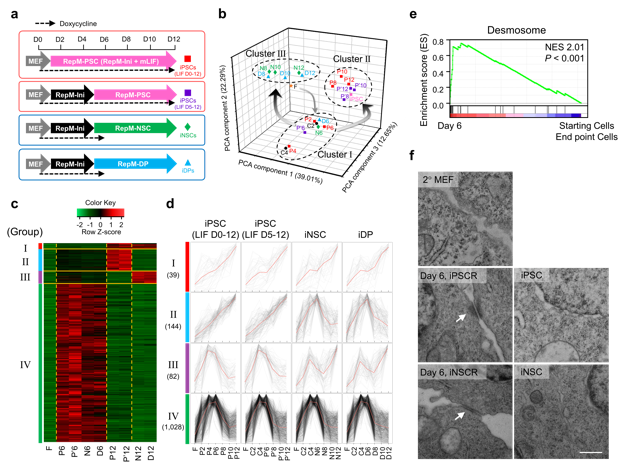
49 Cho, H. J. *et al.* Vascular defects of DYRK1A knockouts are ameliorated by modulating calcium signaling in zebrafish. *Dis Model Mech* **12**, doi:10.1242/dmm.037044 (2019).

**Acknowledgements** This work was supported by grants from the National Research Foundation of Korea (2019R1A2C2087606 and 2018R1A5A1025511) to J. K. and J. K. K., by Samsung Research Funding Center of Samsung Electronic (SRFC-MA1601-06) to J. K., by National Research Council of Science & Technology (NST) of Ministry of Science and ICT of Korea (CRC-15-04-KIST) to J.-S. L., KBRI basic research program through Korea Brain Research Institute funded by Ministry of Science and ICT (20-BR-01-09) to J. Y. M. and KRIBB research initiative program funded by the Ministry of Science and ICT. The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

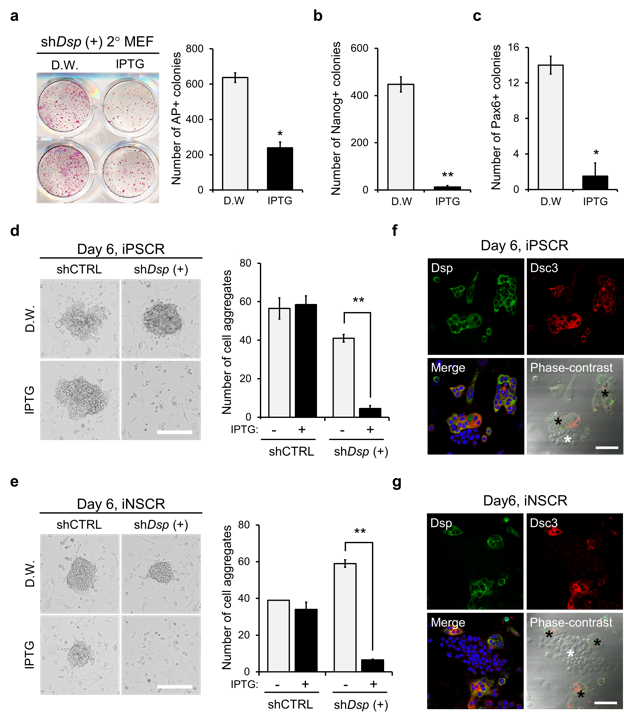
**Author contributions** J. H. designed and performed the experiments, analyzed the data, and wrote the manuscript; B. K., B. M., and Y. H. C. performed bioinformatics analyses with single cell RNA sequencing (scRNAseq) data; B.-H. Y. performed preliminary bioinformatics analyses with microarray data; I.I. prepared samples for 4F2A MEF microarray and scRNAseq; J. K. K. and M. K. supervised the bioinformatics analyses; J. N. and J. –G. L. performed *in vivo* regeneration in zebrafish with advice from J. –S. L; M. L. performed human iPSC and iNSC reprogramming; H. C. performed TEM analysis with advice from J.-Y. M; J.-S. P. and S.-M. C. performed blastocyst injection for generation of 2°MEF with advice from Y.-K. K. and K.-H. N.; A. B. helped with genomic analyses; M.-O. L. and M.-Y. S. provided helpful comments; J. K. supervised the overall project, analyzed the data, and wrote the manuscript.

**Competing interests** The authors declare no competing interests.

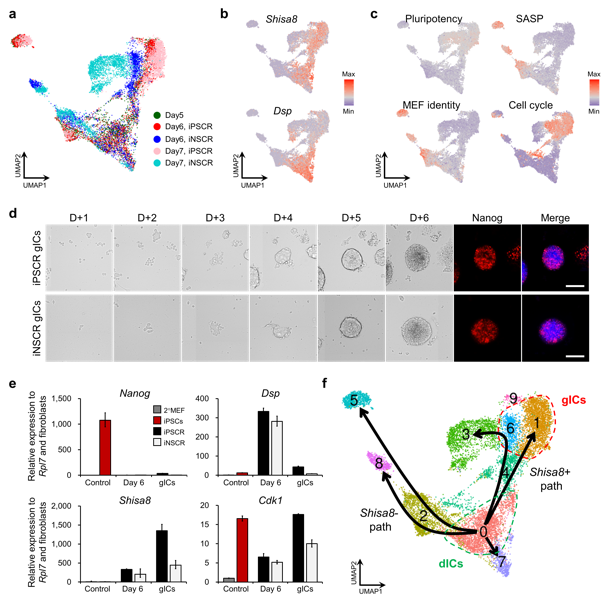
**Correspondence and requests for materials** should be addressed to J. K. K. or J. –S. L. or J. K.

****

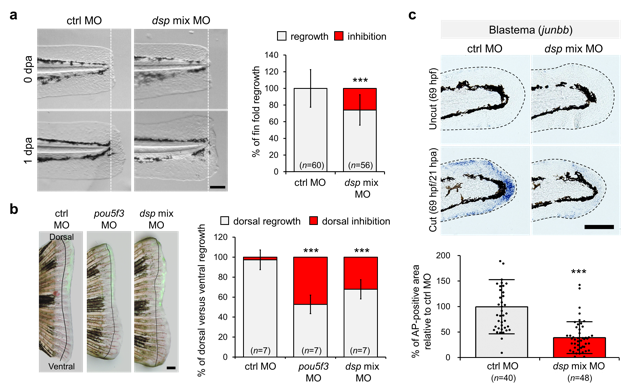
**Fig. 1 | Desmosomes are formed only in the intermediate phase of reprogramming. a**, Reprogramming schematic from 2° MEFs to iPSCs, iNSCs, or iDPs. D, day; MEF, medium for mouse embryonic fibroblast (MEF); RepM-Ini, reprogramming initiation medium; RepM-PSC, reprogramming medium for pluripotent stem cell (PSC); RepM-NSC, reprogramming medium for neural stem cell (NSC); RepM-DP, reprogramming medium for dopaminergic progenitor (DP). **b**, Principal component analyses (PCA) of the microarray datasets. Each reprogramming sample is colored by reprogramming regime. The numbers indicate the days after reprogramming when RNA was collected. F, 2° MEFs; C, common intermediates without LIF; P, iPSC reprogramming (LIF D0-12); P’, iPSC reprogramming (LIF D5-12); N, iNSC reprogramming; D, iDP reprogramming. **c**, Heat map showing gene expression of four groups in 2° MEFs, at day 6 samples, and day 12 samples. **d**, The individual expression patterns of four groups are superimposed. The number of genes in each group is presented. All values were normalized to the average expression of each gene and visualized by a line plot. Red line represents mean of gene expression. **e**, Gene set enrichment analysis showing that “Desmosome” represents the highest normalized enrichment score (NES). NESs and *p*-values are shown. **f**,Representative transmission electron microscopy images of cellular junction regions in 2° MEFs, day 6 samples of iPSCR and iNSCR, iPSCs and iNSCs. White arrows indicate desmosome. Scale bar represents 0.5 µm. Data representative of experimental duplicates.



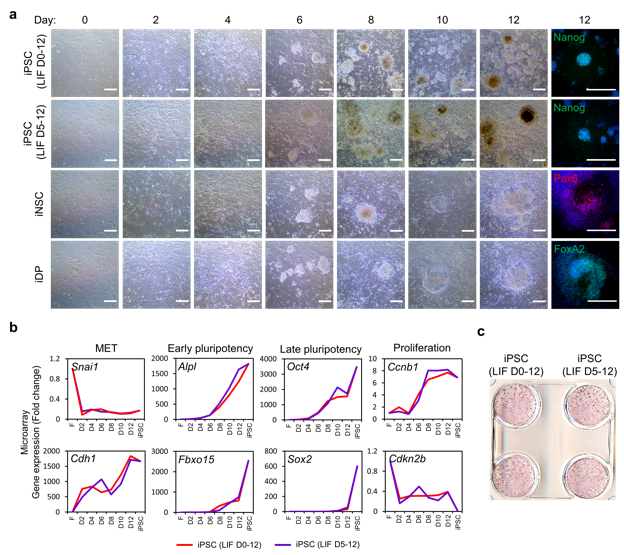
**Fig. 2 | Loss of *Dsp* impeded OSKM-mediated reprogramming by controlling intermediate cell formation. a**,**b**, Whole-well imaging of alkaline phosphatase (AP) staining and quantitative analysis of the AP+ colonies (**a**) and Nanog+ colonies (**b**) at day 12 of iPSCR with D.W. or IPTG. D.W., distilled water (**a**, *n* = 2 wells, *P* = 0.0112; **b**, *n* = 2 wells, *P* = 0.0056). **c**, Quantitative analysis of Pax6+ colonies at day 12 of iNSCR with D.W. or IPTG (*n* = 2 wells; P = 0.0202). **d**,**e**,Representative phase-contrast images and quantitative analysis of cell aggregates on day 6 of iPSCR (**d**) and iNSCR (**e**) with D.W. or IPTG. Scale bars represent 200 µm (**d**, *n* = 2 wells; shCTRL, *P* = 0.80; shDSP+, *P* = 0.0047) (**e**, *n* = 2 wells; shCTRL, *P* = 0.34; shDSP+, *P* = 0.0015). **f**,**g**, Immunostaining and phase-contrast images with desmosomal components (Dsp and Dsc3) on day 6 of iPSCR (**f**) and iNSCR (**g**). The black asterisk indicates positive cells and white asterisk indicates negative cells of the indicated desmosomal components in one colony. Scale bars represent 50 µm. Error bars represent ± s.d. *P* values < 0.05 were considered significant (\* *P* < 0.05; \*\* *P* < 0.01), two-tailed, paired Student’s *t* test was used. Data representative of 2 independent experiments.

****

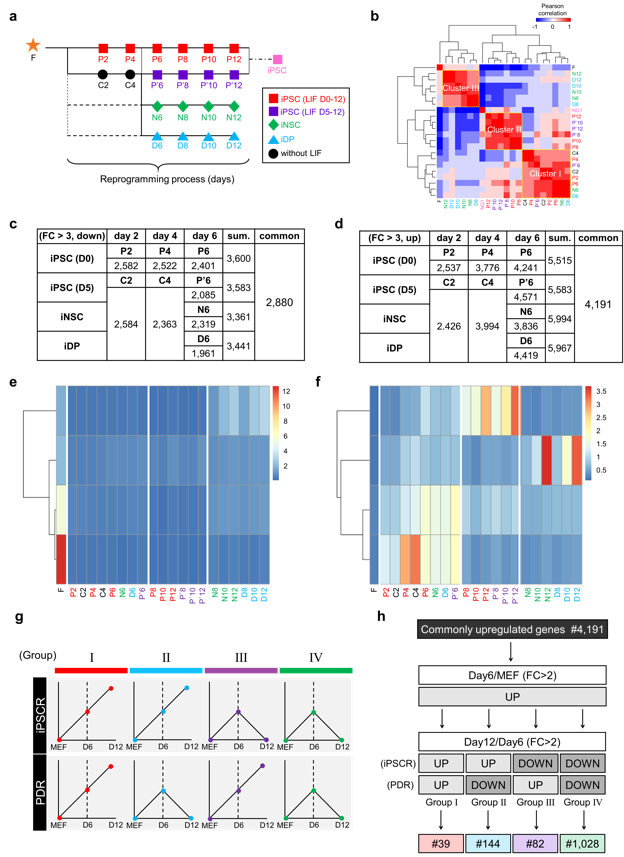
**Fig. 3 | gICs are derived from dICs. a**,UMAP plots of scRNAseq data showing cells between day 5–7 of iPSCR and iNSCR. **b**,**c**, Signature score distribution for indicated genes (**b**) and identity gene sets (**c**) defining dICs and gICs on the UMAP. Scale bars represent the minimum (min) and maximum (max) signature scores during the whole reprogramming process. **d**,Phase-contrast and immunostaining images of the same region of interest after re-attachment of gICs from day 6 of iPSCR or iNSCR (D0). Immunostaining was performed for Nanog on day 6 after re-attachment. (D+6) Scale bars represent 100 µm. **e**,Reverse-transcription qPCR for indicated genes. 2° MEFs were used as a negative-control (grey) and iPSCs were used as a positive-control for the pluripotency marker (red). Day 6 samples included whole cells of day 6 and gICs were isolated on day 6. **f**,Predicted reprogramming trajectories using Slingshot showing the dICs branching to *Shisa8*-positive and -negative paths. Cells are colored by cluster numbers generated using Seurat3. Error bars represent ± s.e. Data representative of experimental triplicates.

****

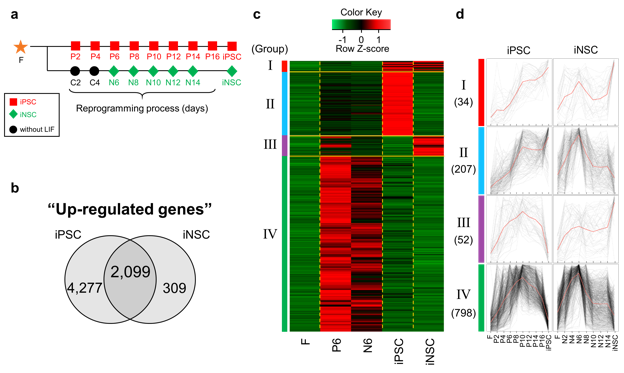
**Fig. 4 | Loss of *dsp* impedes *in vivo* regeneration by controlling blastema cell induction. a**,Phase-contrast images of larvae fin folds after amputation. Fin fold amputation was done at 2 dpf and photographed 24 h later. The graph shows the percentage of 1 dpa versus 0 dpa fin fold outgrowth (in grey) and average inhibition of 1 dpa versus 0 dpa fin fold (in red). Scale bar represents 100 µm (*n* = 60, 56 fishes, *P* < 0.0001). dpa, days post-amputation; ctrl, control; MO, morpholino; *dsp* mix MO, *dspa* MO *+ dspb* MO. **b**,Phase-contrast and fluorescence merged images of adult fins after injection and electroporation with ctrl MO, *pou5f3* MO (positive-control), and *dsp* mix MO. Injection was done into the dorsal half of a 2 dpa blastema and photographed 24 h later. The graph shows the percentage of dorsal versus ventral fin outgrowth (in grey) and average inhibition of dorsal versus ventral fin (in red) of fins. The line on each image denotes the fins at 2 dpa. Scale bar represents 500 µm (*n* = 7 fishes, *P* < 0.0001). **c**, Whole-mount in situ hybridization analysis of larvae fin folds with control MO or *dsp* mix MO. Representative phase-contrast images and quantitative analysis show the expression of blastema marker gene (*junbb*) in regenerating fin folds. Uncut controls were equivalently stained. Dotted lines indicate the boundary of the fin folds. Scale bars represent 100 µm. (*n* = 40, 48 fishes, *P* < 0.0001). hdf, hours post-fertilization; hpa, hours post-amputation. Error bars represent ± s.d. *P* values < 0.05 were considered significant (\* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001), two-tailed, paired Student’s *t* test was used. Data representative of 2 independent experiments. See method for exact calculation.



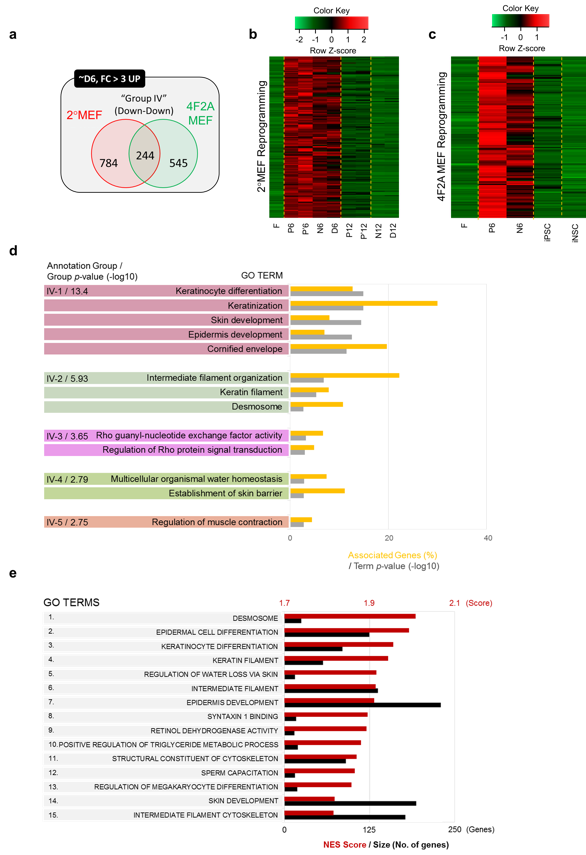
Extended Data Fig. 1 | Analyses of reprogramming procedures in 2° MEFs. a, Phase-contrast and immunostaining images of cells in four reprogramming regimes on day 12 with markers for iPSC (Nanog) or iNSC (Pax6) or iDP (FoxA2). Scale bar, 200 µm. b, Expression analysis of indicated genes during iPSC reprogramming (iPSCR) and in established iPSCs. Red lines depict iPSCR with LIF treatment D0-12; purple lines depict iPSCR with LIF treatment D5-12. c, Alkaline phosphatase staining of iPSCR samples on day 12 of iPSCR with LIF treatment D0-12 or D5-12 (*n* = 2 wells). Data representative of 2 independent experiments.



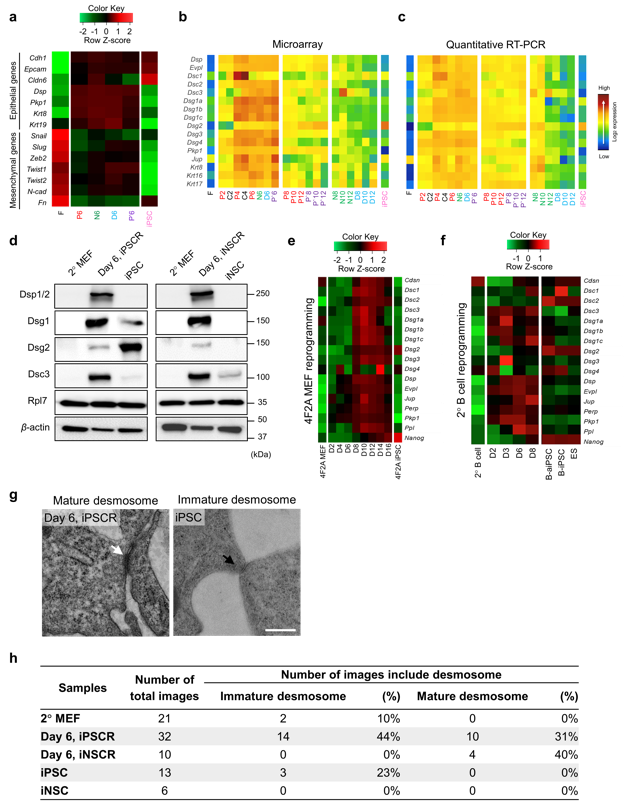
Extended Data Fig. 2 | Differentially expressed genes during OSKM-mediated reprogramming in 2° MEFs. a, Experimental design for microarray shows four reprogramming regimes and sample collection for transcriptome analysis in 2° MEFs. b, Pearson correlation matrix with hierarchical clustering. c,d, Tables show the number of genes whose expression was down-regulated (c) or up-regulated (d) more than 3-fold compared to fibroblasts. e,f, K-means (k = 4) clustering of commonly down-regulated genes (e) and commonly up-regulated genes (f) at indicated reprogramming time points. g, Schematic representing grouping strategies by gene expression pattern in starting 2° MEFs, day 6, and day 12. h, Flow chart for selection of DEGs. Genes were classified into four groups by expression patterns.



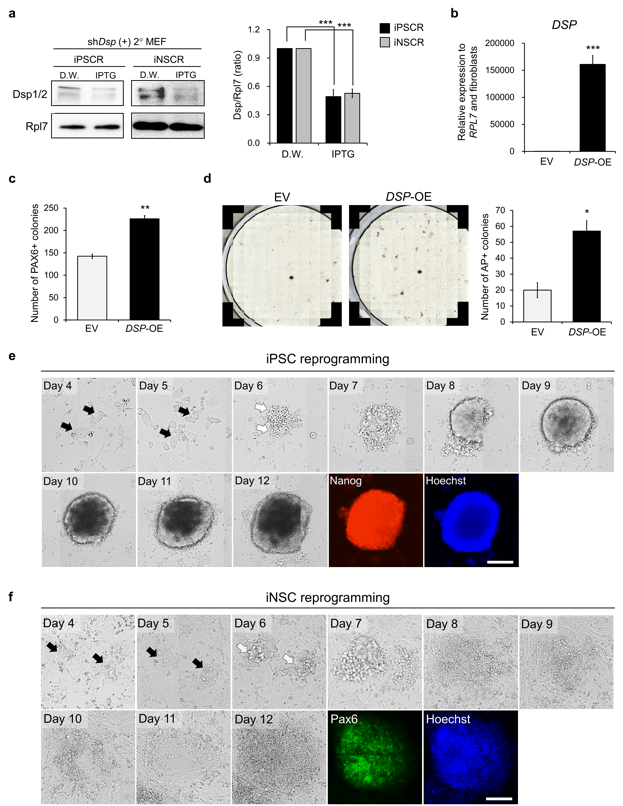
**Extended Data Fig. 3 | Transcriptome analysis during OSKM-mediated reprogramming of 2° MEFs and 4F2A MEFs.** **a**, Experimental design for microarray shows four reprogramming regimes and sample collection for transcriptome analysis in4F2A MEFs. The iPSC and iNSC colonies appeared at day 16 and 14, respectively. **b**, Venn diagram represents the number of genes whose expression was up-regulated more than 3-fold up to day 6 compared with 4F2A MEFs. **c**, Heat map showing gene expression of four groups in 4F2A MEFs, at day 6, iPSCs, and iNSCs. **d**,The individual expression patterns of Groups I–IV are superimposed. In iPSCR, genes representing intermediate phase-specific expression show maximum expression on day 12, as the reprogramming kinetics of the 4F2A MEF system was slower than the 2° MEF system. The number of genes included in each group was presented. All values were normalized to the average expression of each gene and displayed using line plots.



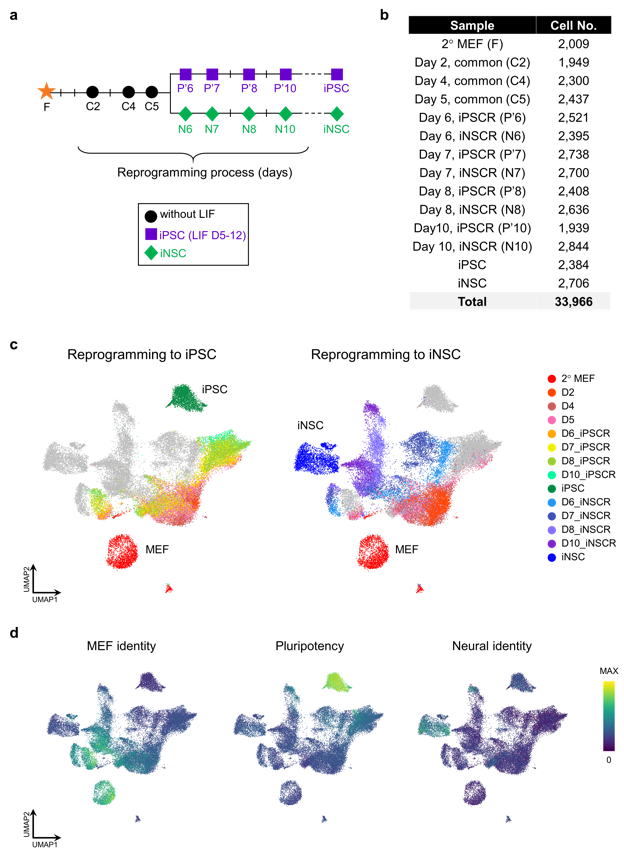
Extended Data Fig. 4 | Gene ontology analysis of common genes in group IV between 2° MEFs and 4F2A MEFs. a, Venn diagram represents the number of Group IV genes in each cell system. b,c, Heat maps showing the expression of the 244 common genes on 2° MEF (b) or 4F2A MEF (c) reprogramming. d, The five most significantly enriched groups were characterized using ClueGO analysis. The left panel shows representative GO terms and the rank of each group together with group *p*-value. The bars in the right panel represent the percentage of genes associated with a specific GO term (yellow) and the term *p*-value (grey). e, Top 15 GO terms ranked by GSEA. GSEA was performed using transcriptomes of ICs (day 6 samples in 2° MEF and 4F2A MEF reprogramming) versus transcriptomes of initial cells (2° MEFs and 4F2A MEFs) and final products (P12, P’12, N12, D12, 4F2A iPSCs, and 4F2A iNSCs). The bars represent the normalized enrichment scores (NES; dark red) and number of genes included in each GO term (black).



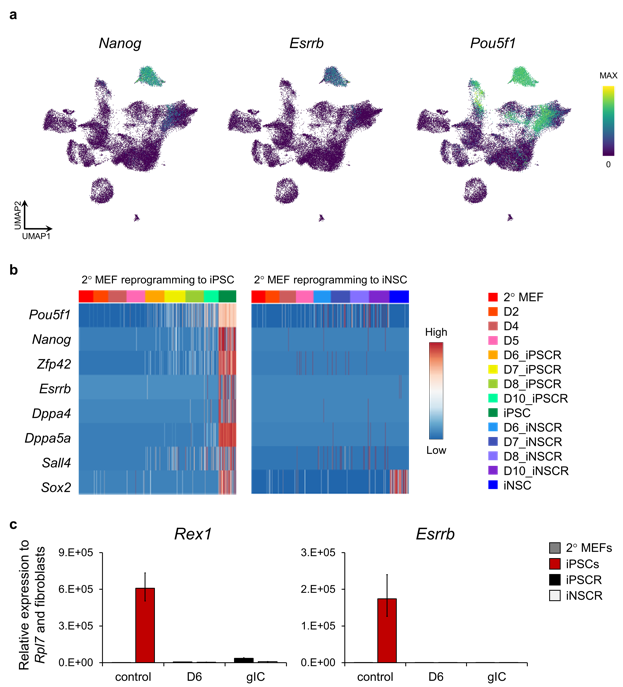
Extended Data Fig. 5 | Expression of desmosomal components in OSKM-mediated reprogramming. a, Heat map showing the expression of indicated genes including epithelial and mesenchymal markers. b,c, Microarray analysis (b) and quantitative RT-PCR (c) for desmosomal components during 2° MEF reprogramming. Heat map values were normalized after being log transformed. d, Western blot analysis for desmosomal components during 2° MEF reprogramming. Rpl7 and β-actin were used as loading controls. e,f, Heat map showing the expression of desmosomal components in microarrays during 4F2A MEF (e) and 2° B-cell reprogramming (f). To analyze 2° B-cell reprogramming, publicly available microarray data were used (GSE52397). g, Representative transmission electron microscopy images of cellular junction regions at day 6 of iPSCR and iPSC. The white arrow indicates a mature desmosome and the black arrow indicates an immature desmosome. Scale bar represents 0.5 µm. h, Table shows the number of mature or immature desmosome in the indicated samples. Data are representative of 2 independent experiments (d, g) or experimental triplicates (c).



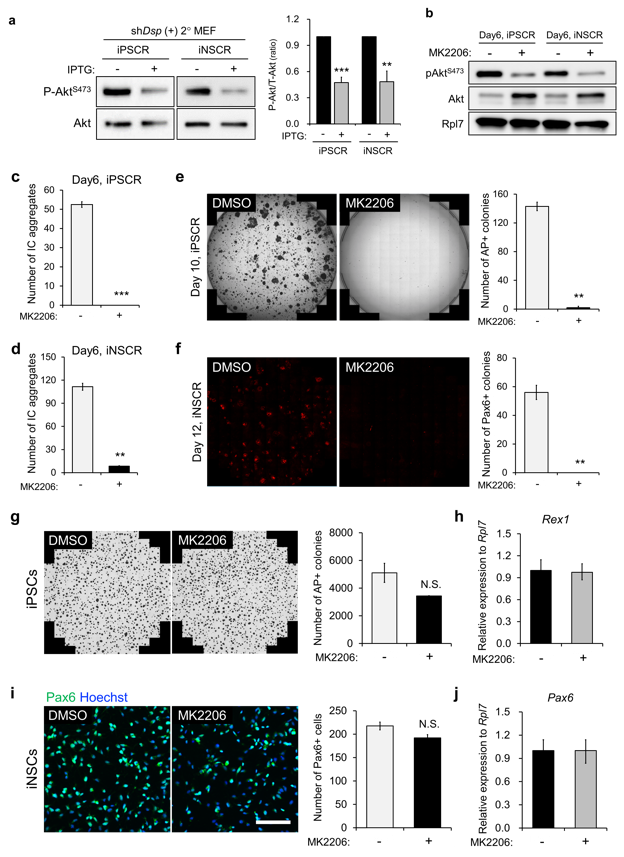
**Extended Data Fig. 6 | Altered reprogramming efficiency by controlling Dsp expression in both mouse and human cells. a**, Western blot for Dsp and Rpl7 during iPSCR and PDR of sh*Dsp*-GFP positive sorted 2° MEFs. The bar graph indicates the quantitative expression of Dsp. Samples were harvested at day 6 in both iPSCR and iNSC. Four independent experiments were performed and Rpl7 was used as the loading control (*n* = 2 wells, *P* < 0.0001). **b**, Quantitative RT-PCR analysis of *DSP* expression following *DSP* overexpression (*DSP*-OE), normalized to expression of *RPL7,* in human fibroblasts. EV, empty vector (EV) (*n* = 2 wells, *P* < 0.0001**)**. **c**, Quantitative analysis of PAX6+ colonies at day 16 of iNSCR upon *DSP* overexpression (*n* = 2 wells, *P* = 0.0053**)**. **d**, Representative AP staining images and quantitative analysis of the AP+ colonies at day 16 in human iPSCR upon *DSP* overexpression (*n* = 2 wells, *P* = 0.0102). **e**,**f**, Phase-contrast images and fluorescence images of the same region of interest during iPSCR (e) and iNSCR (f). Black arrows indicate dIC-like cells and white arrows indicate gIC-like cells. Scale bars represent 100 µm. Error bars represent ± s.e.m. (**a**) or ± s.e. (**b**) or ± s.d. (**c, d**). *P* values < 0.05 were considered significant (\* *P* < 0.05; \*\*\* *P* < 0.001), two-tailed, paired Student’s *t* test was used. Data representative of 3 independent experiments (**a, c, d**) or experimental triplicates (**b**).



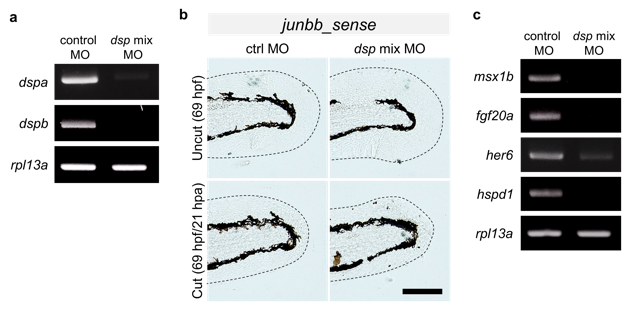
Extended Data Fig. 7 | Single cell RNA sequencing of OSKM-mediate reprogramming. a, Schematic of two reprogramming regimes and sample collection for scRNAseq. b, The number of single cell libraries at each reprogramming day. c,d, UMAP plots of cells during iPSC or iNSC reprogramming colored by collection time points (c) or gene set signature scores (d).



Extended Data Fig. 8 | The expression of pluripotency markers during OSKM-mediate reprogramming. a, UMAP plots of cells during iPSC or iNSC reprogramming colored indicated gene signature scores. b, Expression of the pluripotency signature in cells from iPSCR and iNSCR. Reprogramming day is indicated by colors in upper bar. c, Reverse-transcription qPCR analysis with pluripotency markers. 2° MEF was used as a negative-control (grey) and iPSC was used as a positive control for pluripotency markers (red). Day 6 samples included whole cells of day 6 and gICs were isolated on day 6. Error bars represent s.e. Data are representative of experimental triplicates



Extended Data Fig. 9 | Akt signaling inhibition during OSKM-mediate reprogramming. a, Representative western blot images and quantitative graph for phospho-Akt and total Akt at day 6 of iPSCR and iNSCR with or without *Dsp* knockdown. Values normalized to total protein on the same blot (*n* = 2 wells; iPSCR, *P* = 0.0002; iNSCR, *P* = 0.0055). b, Representative western blot for phospho-Akt and total Akt at day 6 of iPSCR and iNSCR with or without MK2206 (Akt inhibitor). Rpl7 was used as a loading control (*n* = 2 wells). c,d, Quantitative analysis of IC aggregates at day 6 of iPSCR (c) and iNSCR (d) with or without MK2206 (c, *n* = 2 wells, *P* = 0.0009; d, *n* = 2 wells, *P =* 0.0019). e,f, Whole-well imaging and quantitative analysis of AP+ (e) and Pax6+ (f) colonies at day 12 of iPSCR and iNSCR, respectively, with or without MK2206 (e, *n* = 2 wells, *P* = 0.0020; f, *n* = 2 wells, *P =* 0.0078). g, Whole-well imaging and quantitative analysis of AP+ colonies in iPSCs with or without MK2206. h, Reverse-transcription qPCR analysis of pluripotency-associated marker (*Rex1*) in iPSCs with or without MK2206 (*n* = 2 wells, *P* = 0.13). i, Representative fluorescence images and quantitative analysis of the Pax6+ cells in iNSCs with or without MK2206. Scale bar represents 100 µm (*n* = 2 wells, *P =* 0.15). j, Reverse-transcription qPCR analysis of a neural stem cell marker (*Pax6*) in iNSCs with or without MK2206. Error bars represent s.e.m. (a) or s.d. (c–g, i) or s.e. (h, j). *P* values < 0.05 were considered significant (\*\* *P* < 0.01; \*\*\* *P* < 0.001), two-tailed, paired Student’s *t* test was used. Data representative of 3 independent experiments (c–g, i) or experimental triplicates (h, j).



Extended Data Fig. 10 | The expression of blastema marker genes in *dsp* morphant embryos. a, semi-PCR shows altered *dsp* (*dspa* and *dspb*) expression induced by *dsp* mix MO (*n* = 18 fishes). b,Whole-mount in situ hybridization analysis of the larvae fin fold, control MO or *dsp* mix MO. All samples were equivalently stained with the sense probe of *junbb*. Dotted lines indicate the boundary of the fin fold. Scale bar represents 100 µm (*n* = 8 fishes). c, Semi-PCR shows altered expression of blastema marker genes(*msx1b, fgf20a, her6,* and *hspd1*) induced by *dsp* mix MO(*n* = 18 fishes). RNA samples were harvested at 2 dpa and *rpl13a* was used as an internal control. Data representative of 2 independent experiments (b) or experimental triplicates (a, c).