

Characterization of progenitors for mouse fetal Leydig cells

井上, 実紀

<https://hdl.handle.net/2324/4060011>

出版情報 : Kyushu University, 2019, 博士 (理学), 課程博士
バージョン :
権利関係 :

Characterization of progenitors for mouse fetal Leydig cells

Miki Inoue

Laboratory of Biology of Sex Differences
Division of Medical Molecular Cell Biology
Graduate School of Systems Life Sciences
Kyushu University

2020

List of abbreviation

α -MEM	α -modified Eagle's medium
Acta2	actin α 2
Actb	β -actin
Ad4BP/SF-1	Ad4 binding protein/steroidogenic factor 1
ALC	adult Leydig cell
Aldoa	aldolase A
Amh	anti-Müllerian hormone
Arx	aristaless related homeobox
CEL-Seq2	Cell expression by linear amplification sequencing 2
ChIP-seq	chromatin immunoprecipitation-sequence
Col15a1	collagen, type XV, α 1
COUP-TFII	chicken ovalbumin upstream promoter transcription factor II
Cyp11a1	cholesterol side chain cleavage P450
Cyp11b1	11 β -hydroxylase P450
Cyp17a1	17 α -hydroxylase/17,20-lyase P450
Cyp21a1	21-hydroxylase P450
DAPI	4'6'-diamidino-2-phenylindole
DAVID	Database for Annotation, Visualization and Integrated Discovery
Dcn	decorin
Ddx4	DEAD box polypeptide 4
DEG	differentially expressed gene
Dhh	desert hedgehog
Dlk1	delta-like 1 homolog
DMEM	Dulbecco's modified Eagle's medium

DMSO	dimethyl sulfoxide
E	embryonic day
ECAR	extracellular acidification rate
Ednra	endothelin receptor type A
EGFP	enhanced green fluorescent protein
Eno1	enolase 1
ERCC	External RNA Controls Consortium
Err α	estrogen-related receptor α
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
Fgf9	fibroblast growth factor 9
FLC	fetal Leydig cell
FLE	fetal Leydig enhancer
FPKM	fragments per kilobase of transcript per million mapped fragments
Gapdh	glyceraldehyde 3-phosphate dehydrogenase
Gli1	GLI-Kruppel family member GLI1
Gli2	GLI-Kruppel family member GLI2
Gli3	GLI-Kruppel family member GLI3
GO	Gene Ontology
Gpi1	phosphoglucose isomerase 1
Hes1	hes family bHLH transcription factor 1
Hey1	hairy/enhancer-of-split related with YRPW motif-like
Hk1	hexokinase 1
Hk2	hexokinase 2
Hsd17b3	17 β -hydroxysteroid dehydrogenase type 3
Hsd3b	3 β -hydroxysteroid dehydrogenase

Hsd3b1	3 β -hydroxysteroid dehydrogenase type 1
Hummr	hypoxia up-regulated mitochondrial movement regulators
Idh3g	isocitrate dehydrogenase 3 γ
Insl3	insulin-like 3
Jag1	jagged 1
Jund	jun D proto-oncogene
KD	knockdown
KEGG	Kyoto Encyclopedia of Genes and Genomes
Kras	kirsten rat sarcoma viral oncogene homolog
Lsp1	lymphocyte specific 1
M-MLV	Moloney Murine Leukemia Virus
Mash1	mammalian achaete-scute homolog 1
Mfap2	microfibrillar-associated protein 2
Mfap4	microfibrillar-associated protein 4
Mgarp	mitochondria localized glutamic acid rich protein
Mgp	matrix Gla protein
mRNA-seq	mRNA-sequencing
NADPH	nicotinamide adenine dinucleotide phosphate
Ndufs8	NADH: ubiquinone oxidoreductase core subunit S8
Nes	nestin
OCR	oxygen consumption rate
Osap	ovary specific acidic protein
OXPPOS	oxidative phosphorylation
PBS	phosphate-buffered saline
PCA	principal component analysis
Pdgf	platelet-derived growth factor

Pdgfra	Pdgf receptor α
Pdgfra	Pdgf receptor β
Pecam1	platelet/endothelial cell adhesion molecule 1
Pgam1	phosphoglycerate mutase 1
Pgk1	phosphoglycerate kinase 1
Pkm	pyruvate kinase, muscle
Pod1	E-box binding transcription factor
PPP	pentose phosphate pathway
PS	penicillin and streptomycin
Ptch1	patched 1
Ptch2	patched 2
qRT-PCR	quantitative RT-PCR
Rn18s	18S ribosomal RNA
S-EGFP	strongly EGFP-labeled cells
SAG	smoothened agonist
scRNA-seq	single-cell RNA-sequencing
Sdha	succinate dehydrogenase complex, subunit A
SEM	standard error of the means
siRNA	small interfering RNA
Smo	smoothened
Sox9	SRY-box 9
Srebf2	sterol regulatory element binding factor 2
Sry	sex-determining region of the Y chromosome
Star	steroidogenic acute regulatory protein
Suclg2	succinate-coenzyme A ligase, β subunit
t-SNE	t-distributed stochastic neighboring embedding

TCA cycle	tricarboxylic acid cycle
Tmsb10	thymosin β 10
Tmsb4x	thymosin β 4x
Tpi1	triosephosphate isomerase 1
UCSC	University of California Santa Cruz
UMI	unique molecular identifier
Uqcr10	ubiquinol-cytochrome c reductase, complex III subunit X
W-EGFP	weakly EGFP-labeled cells
Wnt4	wingless-related MMTV integration site 4

Table of Contents

ABSTRACT	12
INTRODUCTION	14
MATERIALS AND METHODS	
<i>Preparation of EGFP-labeled cells by FACS</i>	19
<i>Immunocytochemistry</i>	19
<i>In vitro testis reconstruction with cells prepared by FACS</i>	20
<i>Immunofluorescence analyses</i>	21
<i>mRNA-sequencing (mRNA-seq) and data processing</i>	22
<i>qRT-PCR analyses</i>	23
<i>Library preparation for single-cell RNA-sequencing (scRNA-seq) according to Cell expression by linear amplification sequencing 2 (CEL-Seq2)</i>	24
<i>Data analysis of single-cell transcriptomes</i>	25
<i>Measurement of extracellular acidification rate (ECAR) and oxygen consumption rate (OCR)</i>	26
<i>Small interfering RNA (siRNA) treatments</i>	26
<i>Luciferase reporter gene assay</i>	27
<i>Analyses of genomic sequences</i>	28
RESULTS	
<i>W-EGFP cells possibly containing FLC progenitors</i>	29
<i>Biological pathways activated in S-EGFP cells</i>	30
<i>Biological pathways activated in W-EGFP cells</i>	35
<i>Expression of genes related to growth factor signaling</i>	36
<i>scRNA-seq of interstitial cells in fetal testes</i>	37

<i>Possible candidate for FLC progenitors</i>	38
<i>Expression of steroidogenic and metabolic genes in the five clusters</i>	40
<i>Expression of marker genes in the five clusters</i>	41
<i>Unique expression of Tmsb10</i>	42
<i>Possible implication of Tmsb10 into FLC differentiation</i>	44
<i>Possible implication of Tmsb10 into DHH signaling</i>	45
<i>Activation of Ad4BP/SF-1 gene transcription by Gli</i>	45
<i>Implication of Tmsb10 into Gli1 gene expression</i>	47
<i>Regulation of primary cilia by Tmsb10</i>	48
DISCUSSION	
<i>Characterization of fetal Leydig progenitor cell population</i>	49
<i>FLC progenitors characterized by transiently expressed genes</i>	50
<i>Possible implication of Tmsb10 into DHH signaling to activate FLC differentiation</i>	52
<i>Activation of metabolism during differentiation of FLCs</i>	53
<i>Activation of Ad4BP/SF-1 expression by Gli</i>	56
REFERENCES	57
ACKNOWLEDGEMENTS	79
LIST OF FIGURES	
<i>Fig. 1, Structure and cellular components of a fetal testis</i>	80
<i>Fig. 2, S-EGFP and W-EGFP cells prepared from the fetal testes of FLE-EGFP mice</i>	81
<i>Fig. 3, Differentiation of S-EGFP to W-EGFP cells in</i>	

<i>reconstructed testes</i>	82
<i>Fig. 4, Comparison of gene expression between S-EGFP, W-EGFP, and Sertoli cells</i>	83
<i>Fig. 5, Steroidogenic and metabolic gene expression activated in S-EGFP cells</i>	84
<i>Fig. 6, Expression of genes involved in steroidogenesis in S-EGFP and W-EGFP cells</i>	85
<i>Fig. 7, Expression of genes involved in glycolysis and TCA cycle in S-EGFP and W-EGFP cells</i>	86
<i>Fig. 8, Expression of genes involved in OXPHOS in S-EGFP and W-EGFP cells</i>	88
<i>Fig. 9, Expression of genes involved in the mitochondrial pathway for fatty acid synthesis in S-EGFP and W-EGFP cells</i>	89
<i>Fig. 10, Energy metabolism in S-EGFP and W-EGFP cells</i>	90
<i>Fig. 11, Expression of genes involved in cholesterologenesis in S-EGFP and W-EGFP cells</i>	91
<i>Fig. 12, Quantitative analyses for cholesterologenic gene expression in S-EGFP and W-EGFP cells</i>	92
<i>Fig. 13, Expression of genes involved in PPP and NADPH production in S-EGFP and W-EGFP cells</i>	93
<i>Fig. 14, Quantitative analyses of genes activated in W-EGFP, Sertoli, and germ cells</i>	94
<i>Fig. 15, Expression of genes potentially implicated in the FLC differentiation</i>	95
<i>Fig. 16, Expression of genes involved in growth factor signaling</i>	97
<i>Fig. 17, Quality check of scRNA-seq datasets</i>	98

<i>Fig. 18, Characterization of S-EGFP and W-EGFP cells using scRNA-seq datasets</i>	99
<i>Fig. 19, Differential expression of steroidogenic and glycolytic genes in the five cell clusters</i>	100
<i>Fig. 20, Differential expression of TCA cycle genes in the five cell clusters</i>	101
<i>Fig. 21, Differential expression of OXPHOS genes in the five cell clusters</i>	102
<i>Fig. 22, Differential expression of genes involved in PPP and NADPH production in the five cell clusters</i>	105
<i>Fig. 23, Differential expression of genes involved in cholesterologenesis in the five cell clusters</i>	106
<i>Fig. 24, Activation of Tmsb10 in cluster C</i>	107
<i>Fig. 25, TMSB10-positive cells in the interstitial space</i>	108
<i>Fig. 26, Suppression of FLC differentiation by Tmsb10 KD</i>	109
<i>Fig. 27, Activation of FLC differentiation by SAG</i>	110
<i>Fig. 28, FLE-driven transcription activated by GLI1/GLI2</i>	111
<i>Fig. 29, Expression of genes required for hedgehog signaling affected by Tmsb10 KD in W-EGFP cells</i>	113
<i>Fig. 30, Expression of genes required for hedgehog signaling affected by Tmsb10 KD in NIH3T3 cells</i>	114
<i>Fig. 31, Transient activation of Gli1 in cluster C</i>	115
<i>Fig. 32, Primary cilia formation affected by Tmsb10 KD in W-EGFP cells</i>	116

LIST OF TABLES

<i>Table 1, Primary and secondary antibodies, and their dilution ratios used for immunofluorescence analyses</i>	117
<i>Table 2, Primers used for qRT-PCR analyses</i>	118

<i>Table 3, Primers used for CEL-Seq2 analyses</i>	120
<i>Table 4, List of top 20 genes with increased expression in S-EGFP cells</i>	121
<i>Table 5, List of top 20 genes with increased expression in W-EGFP cells</i>	122
<i>Table 6, List of top 20 terms obtained by GO and KEGG pathway analyses using genes with increased expression in S-EGFP cells</i>	123
<i>Table 7, List of top 20 terms obtained by GO and KEGG pathway analyses using genes with increased expression in W-EGFP cells</i>	124

ABSTRACT

Fetal Leydig cells (FLCs) appear in the interstitial space of fetal testes, and play pivotal roles for masculinization of fetuses through testosterone synthesis. Over the past two decades, the molecular mechanisms of FLC differentiation have been studied primarily with genetically modified animals. As the consequence, many transcription factors and growth factors have been identified to be implicated in FLC differentiation. However, to comprehend the mechanisms for FLC differentiation, the functional correlation among the factors above remains to be investigated.

Progenitor cells for FLCs, if identified, could be a powerful tool to study the mechanism for FLC differentiation. Enhanced green fluorescent protein (EGFP) transgenic mice previously established in our laboratory provided us with two cell populations, one of which was strongly EGFP-labeled FLCs (S-EGFP cells) and the other was weakly EGFP-labeled interstitial cells (W-EGFP cells). By *in vitro* testis reconstruction studies using these EGFP-labeled cell populations, I showed for the first time that the interstitial W-EGFP cells differentiated to S-EGFP (FLCs), indicating that the progenitor cells for FLCs are included in the W-EGFP cell population. Comparison of the transcriptomes of the purified FLCs (S-EGFP cells) and the W-EGFP cells including the progenitors revealed expectedly an elevated expression of the steroidogenic genes in the former cells. Interestingly, the genes involved in other metabolic pathways such as energy production (glycolysis, tricarboxylic acid cycle, and

oxidative phosphorylation) and syntheses for lipid and cholesterol were found to be activated in FLCs. Indeed, oxygen consumption and ATP production were elevated in FLCs.

Since W-EGFP cells seemed to consist of multiple cell populations including FLC progenitors, I performed single-cell transcriptome analyses. The transcriptome data revealed the presence of a unique fraction of cells, which were as if differentiating into FLCs. A few genes such as *Tmsb10* (*thymosin β 10*) were identified as a expressed gene in the cell population. As for metabolic genes, several glycolytic gene expressions were elevated in these particular cells. A nuclear receptor, Ad4BP/SF-1 (Ad4 binding protein/steroidogenic factor 1), has been known to be critical for FLC differentiation, and thus the expression is elevated in FLCs. Investigation of an FLC-specific enhancer of *Ad4BP/SF-1* gene raised a possibility that GLI transcription factors activate the gene expression at the downstream of hedgehog signaling. Functional investigation with *in vitro* testis reconstruction demonstrated that *Tmsb10* potentially regulates FLC differentiation possibly through modulating hedgehog signaling. Although additional studies are required to comprehend how *Tmsb10* modulates the hedgehog signaling, the present study advances our understanding of the molecular mechanisms underlying hedgehog dependent FLC differentiation.

INTRODUCTION

Mammalian gonadal sex is determined by the presence or absence of the *SRY* (*sex-determining region of the Y chromosome*) gene. Once SRY is expressed in sexually indifferent gonadal somatic cells, they fix their cell fate to develop into Sertoli cells (Albrecht and Eicher, 2001; Gubbay et al., 1990; Hawkins et al., 1992; Koopman et al., 1991; Lovell-Badge and Robertson, 1990; Sinclair et al., 1990). SRY triggers the expression of *SOX9* (*SRY-box 9*), and then SOX9-positive Sertoli cells immediately surround the primordial germ cells to form tubular structures called testicular cords (future seminiferous tubules) (Kidokoro et al., 2005; Sekido et al., 2004; Sekido and Lovell-Badge, 2008; Wilhelm et al., 2007). As the consequence, the basic structure consisting of two distinct compartments, testicular cords and interstitial space, is formed in the fetal testis (Brennan and Capel, 2004) (Fig. 1). Sertoli and germ cells are localized inside the testicular cords, while a variety of cells, such as Leydig cells, peritubular myoid cells, endothelial cells, macrophages, and uncharacterized interstitial cells, are localized at the interstitial space (Barsoum et al., 2013; DeFalco et al., 2014; DeFalco et al., 2011; Jameson et al., 2012; Martin, 2016; McClelland et al., 2015; Potter and DeFalco, 2017; Stevant et al., 2018; Svingen and Koopman, 2013).

In 1850, a German scientist, Franz Leydig, first described the cells residing in the interstitial space (Leydig, 1850). Thereafter, evidences supporting the notion that these cells were responsible for the synthesis of androgen were accumulated gradually,

resulting in recognition of their physiological roles of the interstitial cells. Honoring his finding of the physiologically important cells, the androgen producing cells localized in the interstitial space of the testis have been called “Leydig cells”.

The androgen produced by Leydig cells is essential for masculinization and the maintenance of male-specific characters. In mammals, fetal and adult Leydig cells (FLCs and ALCs, respectively) develop in prenatal and postnatal testes, respectively (Roosen-Runge and Anderson, 1959). FLCs appear in the interstitial space of the fetal testes shortly after the sex is determined at around embryonic day 12.5 (E12.5) in mice, and thereafter increase in the number during the fetal days (Griswold and Behringer, 2009; O'Shaughnessy et al., 2006). They produce androstenedione, which is converted to testosterone by HSD17B3 (17 β -hydroxysteroid dehydrogenase type 3) expressed in fetal Sertoli cells, indicating that two cell types, FLCs and Sertoli cells, are required for testosterone production in the fetal testis (O'Shaughnessy et al., 2000; Shima et al., 2013). Interestingly, *Hsd17b3* is expressed in postnatal ALCs whereas the expression disappears from the postnatal Sertoli cells. Therefore, ALCs can produce testosterone by themselves in the postnatal testes, and the male hormone induces the secondary sexual maturation for males.

Since FLCs rarely proliferate (Miyabayashi et al., 2013; Orth, 1982), it has been thought that their increase in number is primarily due to differentiation from progenitor cells, although they were not identified yet (Kerr and Knell, 1988; Shima and

Morohashi, 2017). Stem/progenitors for ALCs were found in the interstitial space of neonatal testes of rats (Ge et al., 2006). Likewise, in mice, NES (nestin)-positive interstitial cells in the postnatal testes have the potential to differentiate into multiple cell types including ALCs (Jiang et al., 2014; Ye et al., 2017). With respect to genes implicated in ALC differentiation, a gene disruption study showed that *COUP-TFII* (*chicken ovalbumin upstream promoter transcription factor II*, *Nr2f2*) is required for the differentiation probably through the expression in the progenitor cells (Kilcoyne et al., 2014; Qin et al., 2008). Taken together, the ALC stem/progenitors seem to be localized in the interstitial space of neonatal to adult testes. However, it remains to identify and characterize FLC progenitors in the fetal testes.

Gene disruption studies have identified multiple growth factors involved in FLC differentiation. DHH (desert hedgehog) (Barsoum et al., 2009; Yao et al., 2002) and PDGF (platelet-derived growth factors) signals (Brennan et al., 2003) have been shown to promote FLC differentiation. Interestingly, DHH receptors, *Ptch1* (*patched 1*), and PDGF receptor, *Pdgfra* (*Pdgf receptor α*), as well as the transcriptional mediators of the hedgehog signal transduction, the GLI-Krüppel family of transcription factors (*Gli1* and *Gli2*), are all expressed in the interstitial cells of the fetal testes (Barsoum and Yao, 2011; Barsoum et al., 2009; Brennan et al., 2003; Yao et al., 2002). Likewise, Notch signaling pathway was shown to contribute to FLC differentiation (Defalco et al., 2013; Tang et al., 2008). In addition to growth factors, many transcription factors have been

discovered from the studies of gene disrupted mice and human patients (Bielinska et al., 2007; Birk et al., 2000; Buaas et al., 2012; Cui et al., 2004; Hammes et al., 2001; Katoh-Fukui et al., 1998; Kitamura et al., 2002; Luo et al., 1994; Meeks et al., 2003; Miyabayashi et al., 2013; Miyado et al., 2012; Miyamoto et al., 1997; Ogata et al., 2012; Park et al., 2005; van den Driesche et al., 2012; Wen et al., 2016; Wen et al., 2014). Taken together, our understanding of the molecular mechanisms for gonad development has been advanced considerably. Unfortunately, however, how these signals are transduced in the FLC progenitor cells remains unclear.

Ad4BP/SF-1 (Ad4 binding protein/steroidogenic factor 1, NR5A1) is a nuclear receptor essential for establishing reproductive functions by regulating genes related to steroidogenesis (Honda et al., 1993; Lala et al., 1992; Morohashi and Omura, 1996; Parker and Schimmer, 1997), glycolysis (Baba et al., 2014), and cholesterologenesis (Baba et al., 2018). We previously identified a fetal Leydig enhancer (FLE) of *Ad4BP/SF-1* gene (Shima et al., 2012), and generated transgenic mice in which FLCs are labeled with an enhanced green fluorescent protein (EGFP) (Shima et al., 2013). Interestingly, fluorescence-activated cell sorting (FACS) analyses of the mouse fetal testes provided us with two distinct EGFP-positive cell populations, strongly EGFP-labeled (S-EGFP) and weakly EGFP-labeled (W-EGFP) cells. Quantitative RT-PCR (qRT-PCR) demonstrated that S-EGFP cells primarily consist of FLCs, and W-EGFP cells consist of interstitial cells.

Throughout my thesis study, I have attempted to comprehend the molecular mechanisms for FLC differentiation via characterizing W-EGFP cell population. An *in vitro* testis reconstruction study successfully demonstrated that W-EGFP cells contain FLC progenitors. The transcriptomes obtained from S-EGFP and W-EGFP cells suggests that differentiation of FLCs is accompanied by activation of multiple metabolic pathways. Since these studies revealed the presence of FLC progenitors in W-EGFP cells, I obtained single-cell transcriptomes of W-EGFP cells, and found *Tmsb10* (*thymosin β 10*) as a possible key gene regulating hedgehog signaling during FLC differentiation.

MATERIALS AND METHODS

Preparation of EGFP-labeled cells by FACS

Shima *et al.* identified a DNA fragment in *Ad4BP/SF-1* gene, which enhances gene expression specifically in FLCs. To label FLCs with EGFP, *FLE-EGFP* mice were generated using the DNA fragment (Fig. 2A) (Shima *et al.*, 2012). In addition to this mouse, I used *Sox9-EGFP* knock-in mice whose Sertoli cells were labeled with EGFP (Nel-Themaat *et al.*, 2009). Male mice from both lines were crossed with ICR females (Japan SLC, Shizuoka, Japan and Kyudo, Saga, Japan). EGFP-positive transgenic testes were harvested at E16.5 and were dispersed as described by Shima *et al.* (Shima *et al.*, 2013). The EGFP-positive cells were recovered using FACS (JASN; Bay bioscience, Kobe, Japan). In addition to these EGFP-positive cells, another cell population labeled weakly with EGFP was observed in the fetal testes. Thus, these two EGFP-positive cell populations were designated as S-EGFP and W-EGFP cells. All protocols for animal experiments were approved by the Animal Care and Use Committee of Kyushu University.

Immunocytochemistry

The sorted cells (2.0×10^4 cells/well) were cultured on collagen type I coated plate (96-well) (IWAKI, Tokyo, Japan) in α -modified Eagle's medium (α -MEM) (Nacalai Tesque, Kyoto, Japan) containing 10% fetal bovine serum (FBS) (Thermo Fisher

Scientific, Waltham, MA, USA), penicillin and streptomycin (PS) (Thermo Fisher Scientific) for 4 hrs at 37 °C with 5% CO₂. They were fixed with 4% paraformaldehyde (Nacalai Tesque) in phosphate-buffered saline (PBS) at room temperature for 15 min, and then washed with 0.1% Triton X-100 (WAKO, Tokyo, Japan) in PBS. After the cells were treated with 0.5% Triton X-100 in PBS at room temperature for 10 min, they were treated in 1.0% skim milk (WAKO) in PBS at room temperature for 20 min. The cells were treated with the primary antibodies for HSD3B (3 β -hydroxysteroid dehydrogenase) and EGFP at 4 °C overnight. After washing with 0.1% Triton X-100 in PBS, the cells were treated with the secondary antibodies at room temperature for 1 hrs. F-actin was stained with AlexaFluor-594 phalloidin (Thermo Fisher Scientific). Nuclei were counterstained with 4'6'-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich, St. Louis, MO, USA). The primary and secondary antibodies used in this study are listed in Table 1. Bright field and immunofluorescence images were obtained using a BZ-9000 fluorescence microscope (Keyence, Osaka, Japan).

In vitro testis reconstruction with cells prepared by FACS

Fetal testes were reconstructed as described by Yokonishi *et al.* (Yokonishi et al., 2013). Briefly, wild type testes at E16.5 were dispersed by incubating in 2 ml of 0.25% trypsin (Sigma-Aldrich) in PBS for 10 min at 37 °C. The reaction was arrested by adding 8 ml of Dulbecco's modified Eagle's medium (DMEM) (Nacalai Tesque)

containing 10% FBS (Thermo Fisher Scientific) and PS (Thermo Fisher Scientific). The resulting mixture was filtered through membranes with pore sizes of 70 μm and subsequently 40 μm (Greiner Bio-One, Kremsmünster, Austria). Approximately 8.0×10^5 testicular cells prepared from wild type were mixed with S-EGFP cells (1.0×10^4 cells) or W-EGFP cells (2.0 or 5.0×10^4 cells) isolated from E16.5 fetal testes of the *FLE-EGFP* transgenic mice. They were cultured on V-shaped agarose gel for 2 days and transferred onto bowl-shaped agarose gel, followed by culturing for 33 days at the longest in α -MEM (Nacalai Tesque) containing 10% Knockout Serum Replacement (Thermo Fisher Scientific) and PS at 37 °C with 5% CO₂. To investigate the effect of hedgehog signal on FLC differentiation, reconstructed tissues with W-EGFP cells were treated with smoothed agonist (SAG) (0.5 μM ; Adipogen Life Sciences, San Diego, CA, USA) or dimethyl sulfoxide (DMSO) (Nacalai Tesque). The reconstructed tissues were observed under a fluorescent stereomicroscope (SZX16/SZX2-ILLT; Olympus, Tokyo, Japan) and a BZ-X700 fluorescence microscope (Keyence).

Immunofluorescence analyses

Cryosections prepared from fetal testes or reconstructed tissues were subjected to immunostaining as described. (Miyabayashi et al., 2015; Shima et al., 2015; Yokoyama et al., 2019). Briefly, the fetal testes and reconstructed tissues were fixed with 4% paraformaldehyde at 4 °C overnight. The fixed samples were soaked in 20% sucrose

(Nacalai Tesque) in PBS at 4 °C overnight, and then embedded in OCT Compound (SAKURA, Tokyo, Japan). After washing with PBS, cryosections (10 µm) were boiled for 5 min in 10 mM sodium citrate (pH 6.0) to unmask antigen epitopes. For immunostaining of TMSB10 and PECAM1 (platelet/endothelial cell adhesion molecule 1), the sections were not boiled. They were treated with 2% skim milk in PBS at room temperature for 30 min, and then incubated with the primary antibodies at 4 °C overnight. After washing with PBS, the sections were treated with the secondary antibodies at room temperature for 1 hr. The primary and secondary antibodies used in this study are listed in Table 1. Can Get Signal Immunostain Solution A (TOYOBO, Osaka, Japan) was used to enhance the signals for HSD17B3. Nuclei were counterstained with DAPI (Sigma-Aldrich). Immunofluorescence images were captured using a BZ-9000 fluorescence microscope (Keyence), an LSM700 with Zeiss AxioObserver Z1 Apotome (Carl Zeiss, Jena, Germany), and an LSM 700 confocal laser scanning microscope (Carl Zeiss).

mRNA-sequencing (mRNA-seq) and data processing

Poly(A) RNAs prepared from S-EGFP and W-EGFP cells were subjected to library construction for mRNA-seq using TruSeq RNA Sample Prep Kit and TruSeq stranded mRNA Sample Prep Kit (Illumina, San Diego, CA, USA). After the quality of the library was validated using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa

Clara, CA, USA), the samples were subjected to sequencing (Genome Analyzer Iix and HiSeq 1000, Illumina). The sequence reads were aligned to the reference mouse genome sequence (UCSC mm10) using TopHat (version 2.1.0) with the default parameters (Trapnell et al., 2009). Cufflinks (version 2.2.1) was then used to assemble the transcripts and calculate fragments per kilobase of transcript per million mapped fragments (FPKM) with the default parameters (Trapnell et al., 2012). Genes of which FPKM values differed by more than 10 between S-EGFP and W-EGFP cells were extracted. After a pseudo-count of one was added to the FPKM values, genes of which FPKM values differed by more than a factor of two were extracted (Consortium., 2014). The extracted genes were subjected to Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (<http://david.abcc.ncifcrf.gov/>). mRNA-seq data have been deposited in DDBJ/EMBL/GeneBank under the accession code DRA004005 and DRA004075.

qRT-PCR analyses

cDNAs were synthesized from total RNAs (n = 3) using Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (Thermo Fisher Scientific) and random primers (Sigma-Aldrich). qRT-PCR was performed using a CFX96 real-time PCR system (Bio-Rad Laboratories, Hercules, CA, USA) and SYBR Select Master Mix

(Thermo Fisher Scientific). The expression level of *Actb* (β -actin) or *Rn18s* (*18S ribosomal RNA*) was used as internal control. Primers used for qRT-PCR are listed in Table 2. Data are presented as means +/- standard error of the means (SEM) were statistically analyzed using a two-tailed Student's *t*-test or a one-way ANOVA followed by Tukey's *post hoc* test.

Library preparation for single-cell RNA-sequencing (scRNA-seq) according to Cell expression by linear amplification sequencing 2 (CEL-Seq2)

Single S-EGFP or W-EGFP cells were plated by FACS (SH800, Sony, Tokyo, Japan) into individual wells of 384-plate (Piko PCR Plate, Thermo Fisher Scientific) pre-loaded with lysis buffer. The CEL-Seq2 protocol established by Hashimshony *et al.* was used for library preparation (Hashimshony *et al.*, 2016). Briefly, the RNA of each cell was reverse transcribed using CEL-Seq primers. The primers consisted of an anchored poly(T), a 6 bp unique molecular identifier (UMI), the 5' Illumina adapter (TruSeq small RNA kit; Illumina), a cell-specific 6 bp barcode, and a T7 promoter (Table 3). After second-strand reaction and cleanup of the cDNA, the double-stranded cDNAs were *in vitro* transcribed by T7 RNA polymerase. The synthesized RNAs were reverse transcribed using random primers with the 3' Illumina adapter. Finally, the libraries were amplified by PCR (11 cycles). The pair-ended libraries were sequenced by HiSeq 2500 (Illumina).

Data analysis of single-cell transcriptomes

A first quality check of the sequence data was done using FastQC (version 0.11.7) and thereafter the sequence data were analyzed according to the CEL-Seq2 pipeline (Hashimshony et al., 2016). The pipeline includes the following steps. First, the reads of CEL-Seq2 libraries were demultiplexed to each cell using the CEL-Seq barcode. To identify the transcript, the reads of each cell were mapped to the mouse reference genome (mm10) including EGFP sequence, mitochondrial genome sequences, and the External RNA Controls Consortium (ERCC) spike-in reference by the Bowtie 2 software (version 2.3.4.1) (Langmead and Salzberg, 2012). UMI sequence was identified in order to remove PCR duplicates. Finally, the mapped reads were counted using HTSeq (version 0.9.1) (Anders et al., 2015).

Low-quality samples were identified and removed by a previously published method (Lun et al., 2016). In brief, they were removed according to certain thresholds of low library sizes, few expressed genes (< 1000 genes), and high spike-in proportions. The library size is defined as the total sum of the read counts. The quality of libraries with small sizes is likely to be low probably due to inefficient preparation of RNA from cells, and thus the diverse transcript population has not been successfully captured. The numbers of the expressed genes in each cell are defined as the number of genes with non-zero counts. Another criterion of the quality is proportion of the reads mapped to

spike-in transcripts. The quantity of spike-in RNA added to each cell should be constant.

High spike-in RNA proportions may reflect low quality, possibly because the proportion should increase upon loss of endogenous RNA.

Cells at the G1 phase of the cell cycle were selected. Expression of the genes remained was normalized and denoised. The normalized data were subjected to hierarchical clustering on the significant principal components to divide scRNA-seq data into clusters. Differentially expressed genes (DEGs) of each cluster were identified in order to characterize the clusters. R packages including SingleCellExperiment (version 1.0.0), scater (version 1.6.3), scran (version 1.6.9), Seurat (version 3.0.2), and monocle (version 2.6.4) were used for analyses described above.

Measurement of extracellular acidification rate (ECAR) and oxygen consumption rate (OCR)

The ECAR and OCR of S- and W-EGFP cells was measured using an XF96 extracellular Flux Analyzer (Agilent Technologies), according to the manufacturer's instructions. After SAG treatment for 24 hrs, the W-EGFP cells were subjected to analysis.

Small interfering RNA (siRNA) treatments

W-EGFP and NIH3T3 cells were cultured on Advanced TC 24-well plates (Greiner Bio-One) in α -MEM and DMEM, respectively, supplemented with 10% FBS and PS. The cells were transfected with *siTmsb10* (Mm01_00030122: 5'-GCUUCGAUAAGGCAAGCUTT-3'; Sigma-Aldrich), *siTmsb4x* (Mm01_00163777: 5'-GCCGCCAAUAUGCACUGUA-3'; Sigma-Aldrich), and *siAd4BP/SF-1* (Mm Nr5a1_1635: 5'-CCUUUAUCUCCAUGUCGATT-3'; Sigma-Aldrich) using lipofectamine RNAiMAX reagent (Thermo Fisher Scientific) according to the manufacturer's instructions for 24 hrs. A control siRNA (Stealth RNAi Negative Control Medium GC Duplex; Thermo Fisher Scientific) was used as a negative control. After the treatment for 24 hrs, the W-EGFP cells were cultured in α -MEM and DMEM with SAG (0.5 μ M; Adipogen Life Sciences) or DMSO (Nacalai Tesque) to investigate the effect of hedgehog signal.

Luciferase reporter gene assay

The reporter genes FLE-Ad4BPpro-luc and Ad4BPpro-luc were constructed using the promoter less vector pGL3-basic (Promega, Madison, WI, USA). Ad4BPpro-luc carries a 0.9 kb upstream region of *Ad4BP/SF-1* gene, while FLE-Ad4BPpro-luc carries a 1.8 kb FLE fragment at the 5' upstream of the 0.9 kb promoter fragment (Shima et al., 2012). FLE-Ad4BPpro-luc, Ad4BPpro-luc, or pGL3 basic vector (10 ng) was transfected to NIH3T3 cells with increasing amounts of *Gli1* or *Gli2* expression

plasmids (20, 100, or 250 ng) using lipofectamine 2000 reagent according to the manufacture's instruction (Thermo Fisher Scientific). pCMV-SPORT- β gal (Thermo Fisher Scientific) was used as the internal control to evaluate transfection efficiency. The cells were harvested 48 hrs after the transfection to determine luciferase and β -galactosidase activities. The analyses were performed in biological triplicate. Data are presented as the means (SEM), and statistical differences between experimental groups were examined by the two-tailed Student's *t*-test.

Analyses of genomic sequences

Genomic sequence conservation of FLE locus among mammal species was examined using the University of California Santa Cruz (UCSC) genome browser (Kent et al., 2002). The genomic sequences were aligned using the multiple sequence alignment program, MAFFT (Katoh and Toh, 2008).

RESULTS

W-EGFP cells possibly containing FLC progenitors

FLE-EGFP transgenic mice were established using an EGFP reporter gene controlled by FLE and promoter region of *Ad4BP/SF-1* gene. Expectedly, FLCs in the fetal testes were labeled strongly with EGFP (S-EGFP cells) and thus the cells were obtained by FACS. In addition to S-EGFP cells, weakly EGFP-positive (W-EGFP) cells were found as a discrete cell population by FACS (Fig. 2B). qRT-PCR studies showed that S-EGFP and W-EGFP cells correspond to FLCs and interstitial cells other than FLCs, respectively (Shima et al., 2013).

Since FLC progenitors appear in the interstitial space of fetal testes (DeFalco et al., 2011; Jameson et al., 2012; McClelland et al., 2015; Stevant et al., 2018), it was hypothesized that W-EGFP cells contain the progenitors. To examine it, I performed an *in vitro* testis reconstruction study, in which whole cells recovered from the fetal testes of wild type (E16.5) were mixed and cultured with S-EGFP or W-EGFP cells. Although W-EGFP cells were recovered as a cell population weakly labeled by EGFP, the weak EGFP signal was not detectable by a fluorescent microscope (Fig. 2, C to F). Testicular cord-like structures were unclear yet after incubation for 4 days (Fig. 3, A, D, G, and J), and became obvious at day 9 (Fig. 3, B, E, H, and K). The EGFP signals of the S-EGFP cells were detectable soon after starting reconstruction (Fig. 3D), and the signals were maintained thereafter (Fig. 3, E and F). Although any EGFP signal was hardly

detectable in the testes reconstructed with the W-EGFP cells before day 9 (Fig. 3, J and K), the signals appeared at day 14 (Fig. 3L), and increased in number thereafter. These EGFP-positive cells were localized in the space outside the testicular cord-like structures (Fig. 3, E, F, and L).

The EGFP-positive cells were characterized by their immunohistochemical properties using antibodies for HSD3B expressed in Leydig cells (Shima et al., 2012), and for Ad4BP/SF-1 expressed at higher levels in Leydig and at lower levels in Sertoli cells (Hatano et al., 1994). HSD3B and Ad4BP/SF-1 were expressed in the EGFP-positive cells of the testes reconstructed with S-EGFP cells (Fig. 3, M and N). HSD3B single-positive cells in Fig. 3M and Ad4BP/SF-1 single-positive cells seen in Fig. 3N and 3Q were possibly derived from the wild type testes used for the reconstruction. Likewise, the EGFP-positive cells in the testes reconstructed with W-EGFP cells were positive for HSD3B and Ad4BP/SF-1 (Fig. 3, P and Q). These cells were expected to be FLCs but not ALCs. To confirm this, we examined the expression of HSD17B3, an ALC marker (O'Shaughnessy et al., 2000; Shima et al., 2013). As expected, HSD17B3 was not expressed in the EGFP-positive cells (Fig. 3, O and R).

Biological pathways activated in S-EGFP cells

To characterize the progenitor cells, I obtained transcriptome of S-EGFP and W-EGFP cells. Sequencing (Seq-I and Seq-II) was performed with poly(A) RNAs

prepared from two biologically independent samples each for S-EGFP and W-EGFP cell fraction. As a control, a single sample of Sertoli cells from fetal testes was used. Sequence reads qualitatively enough for bioinformatics analyses were obtained with all RNA samples above, for which mapping rates were approximately 98% (Fig. 4A). Sequence data from the biological duplicates for S-EGFP and W-EGFP cells showed high correlation coefficient (0.987 for S-EGFP cells and 0.991 for W-EGFP cells, Fig. 4B). Consistent with it, up-regulated genes in S-EGFP and W-EGFP cells revealed by Seq-I were quite similar to those revealed by Seq-II (Tables, 4 and 5). However, the sequence data indicated that the W-EGFP cell fraction used for Seq-I contained Sertoli cell contamination; a representative Sertoli cell marker gene, *Amh* (*anti-Müllerian hormone*), was ranked 9th in Seq-I study (Table 5). Therefore, I will describe mainly the results of Seq-II later in this thesis.

Comparison of the levels of expression between S-EGFP and W-EGFP cells revealed 1,406 and 1,048 genes whose expression was more than doubled in the former and the latter cells, respectively. Using the 1,406 genes enriched in S-EGFP cells, we evaluated the biological pathways activated in these cells by GO and KEGG pathway analyses (Table 6). As expected, pathways related to steroid metabolism were identified as an activated pathway in S-EGFP cells in both analyses. Indeed, steroidogenic genes such as *Star* (*steroidogenic acute regulatory protein*), *Cyp11a1* (*cholesterol side chain cleavage P450*), *Cyp17a1* (*17 α -hydroxylase/17,20-lyase P450*), and *Hsd3b1*

(*3 β -hydroxysteroid dehydrogenase type 1*) required for producing androstenedione were among the top 20 genes up-regulated in S-EGFP cells (Table 4). *Cyp21a1* (*21-hydroxylase P450*) and *Cyp11b1* (*11 β -hydroxylase P450*), essential for adrenocorticoid production, are usually expressed in the adrenal cortex but not Leydig cells. Unexpectedly, however, these genes were also listed as the genes enriched in S-EGFP cells, although their FPKM values were considerably lower than those required for androstenedione production. This ectopic expression may be a result of their development from common progenitor cells (Hatano et al., 1996). Indeed, FLCs were reported to have an ability to synthesize glucocorticoid in response to adrenocortical tropic hormone (O'Shaughnessy et al., 2003). The steroidogenic gene expression activated in S-EGFP cells was validated by qRT-PCR. *Star*, *Cyp11a1*, *Hsd3b1*, and *Cyp17a1* gene transcripts were more abundant in S-EGFP cells than in W-EGFP and Sertoli cells (Fig. 5, A to D, and Fig. 6). Consistent with previous reports (O'Shaughnessy et al., 2000; Shima et al., 2013), *Hsd17b3*, which mediates the conversion of androstenedione to testosterone, was expressed only in Sertoli cells in the fetal testes (Fig. 5E and Fig. 6). *Ad4BP/SF-1*, originally identified as a transcription factor to regulate steroidogenic genes, was expressed abundantly in S-EGFP cells (Fig. 5F and Fig. 6).

In addition to steroid metabolism, I found the pathways/terms related to mitochondrial structures and functions were activated in S-EGFP cells. As for the

functions, tricarboxylic acid cycle (TCA cycle), oxidative phosphorylation (OXPHOS), and fatty acid metabolism were listed with high *p*-values (Table 6, Figs 7, 8, and 9). Glycolysis/gluconeogenesis pathways were also listed as the pathways activated in S-EGFP cells. Indeed, sequence datasets demonstrated that the expression of glycolytic, TCA cycle, and OXPHOS genes was mostly increased in S-EGFP cells. The expression level of these genes was further validated by qRT-PCR. Expression of the glycolytic genes, *Aldoa* (*aldolase A*) and *Eno1* (*enolase 1*), was higher in S-EGFP than W-EGFP cells, whereas that of *Hk1* (*hexokinase 1*) was similar between the two cell types (Fig. 5, G to I, and Fig. 7). Similarly, TCA cycle genes, *Idh3g* (*isocitrate dehydrogenase 3 γ*), *Suclg2* (*succinate-coenzyme A ligase, β subunit*), and *Sdha* (*succinate dehydrogenase complex, subunit A*), were activated in S-EGFP cells (Fig. 5, J to L, and Fig. 7). *Ndufs8* (*NADH: ubiquinone oxidoreductase core subunit S8*) and *Uqcr10* (*ubiquinol-cytochrome c reductase, complex III subunit X*) are components of mitochondrial respiratory enzyme complexes I and III, respectively, and their products were enriched in S-EGFP cells (Fig. 5, M, N, and Fig. 8). Considering that glycolysis, TCA cycle, and OXPHOS are the central pathways for ATP production, these results suggest that energy production, as well as steroidogenesis, is activated in S-EGFP cells.

Therefore, I examined whether the elevated expression of the energy metabolizing genes leads to activation of energy metabolism in S-EGFP cells. As indicated in Fig. 10A, basal glycolytic activity of W-EGFP cells was lower than that of S-EGFP cells.

Likewise, mitochondrial basal respiration and ATP production were lower in W-EGFP than S-EGFP cells (Fig. 10, B and C). Taken together, these studies demonstrated that energy metabolism is upregulated in S-EGFP cells in terms of the gene expression and biological activity.

Moreover, Baba *et al.* recently indicated that the cholesterologenic genes are the targets of Ad4BP/SF-1 (Baba et al., 2018). Thus, I examined the sequence datasets of S-EGFP and W-EGFP cells and found that all cholesterologenic genes were up-regulated in S-EGFP cells (Fig. 11). The cholesterologenic gene expression elevated in S-EGFP cells was confirmed by qRT-PCR (Fig. 12). Interestingly, the expression of *Srebf2* (*sterol regulatory element binding factor 2*), a key regulator of the cholesterologenic genes, was slightly increased, by less than 2-fold (Fig. 11 and Fig. 12A), whereas the expressions of cholesterologenic genes were increased by 2- to 22-fold in S-EGFP cells (Fig. 11, and Fig. 12, B to U). Considering that Ad4BP/SF-1 itself was increased by approximately 17-fold (Fig. 5F and Fig. 6), the increased expression of the cholesterologenic genes in S-EGFP cells might be primarily due to Ad4BP/SF-1.

Ad4BP/SF-1 directly regulates expression of genes crucial for pentose phosphate pathway (PPP) and nicotinamide adenine dinucleotide phosphate (NADPH) synthesis (Baba et al., 2014; Li et al., 2017). Present mRNA-seq data indicated that the genes related to PPP and NADPH production were more abundant in S-EGFP than W-EGFP cells (Fig. 13). Together, it might be possible to assume that FLC differentiation is

accomplished by activation of energy production (glycolysis, TCA cycle, and OXPHOS), PPP, NADPH production, fatty acid synthesis, and cholesterologenesis in addition to steroidogenesis.

Biological pathways activated in W-EGFP cells

1,048 genes enriched in W-EGFP cells were subjected to pathway analyses. Many of the biological terms identified were related to the extracellular matrix, intracellular cytoskeleton, and cell surface structures (Table 7). Functional terms ‘focal adhesion’ was also identified. Consistent with the biological terms listed above, *Coll5a1* (*collagen, type XV, $\alpha 1$*), *Dcn* (*decorin*), *Mfap4* (*microfibrillar-associated protein 4*), and *Mgp* (*matrix Gla protein*) were found in the top 20 genes (Table 5). The increase of *Mgp* gene transcripts was validated by qRT-PCR (Fig. 14A). Terms related to active proliferation were also listed, which is consistent with the observation that interstitial cells other than FLCs proliferate actively (Miyabayashi et al., 2013; Orth, 1982). In contrast to S-EGFP cells, no genes related to metabolic pathways were enriched in W-EGFP cells.

Lsp1 (*lymphocyte specific 1*), *Ednra* (*endothelin receptor type A*), and *Acta2* (*actin $\alpha 2$*), which are markers for lymphocytes, vascular endothelial cells, and peritubular myoid cells, respectively, were also among the genes enriched in W-EGFP cells (Table 5). The increase of *Acta2* gene transcripts was validated by qRT-PCR (Fig. 14B).

Moreover, the Sertoli cell markers (*Amh*, *Sox9*, and *Dhh*) and germ cell marker (*Ddx4* (*DEAD box polypeptide 4*)) were significantly less in W-EGFP than in Sertoli and whole testicular cells (Fig. 14, C to F).

Expression of genes related to growth factor signaling

Heyl (*hairy/enhancer-of-split related with YRPW motif-like*) was one of the genes enriched in W-EGFP cells (Table 5). Because *Heyl* is a Notch downstream gene (Fischer and Gessler, 2007), and Notch signaling negatively regulates differentiation of FLCs (Defalco et al., 2013; Tang et al., 2008), I examined the expression of factors related to Notch signaling (Fig. 15A and Fig. 16). Among the four *Notch* genes, the sequence datasets showed that *Notch1*, *Notch2*, and *Notch3* were enriched in W-EGFP cells (Fig. 16). The qRT-PCR analysis confirmed that these three gene transcripts were more abundant in W-EGFP than S-EGFP cells (Fig. 15, B to D). The Notch ligands *Jag1* (*jagged 1*) and *Dlk1* (*delta-like 1 homolog*) were enriched in S-EGFP and W-EGFP cells, respectively (Fig. 15, E and F). Similar to *Heyl*, *Hes1* (*hes family bHLH transcription factor 1*), a transcription factor functioning downstream of the Notch signal (Fischer and Gessler, 2007; Tang et al., 2008), was enriched in W-EGFP cells (Fig. 15G).

Loss-of-function studies have established that the PDGF signal is crucial for FLC differentiation (Brennan et al., 2003). As expected, expression of the PDGF receptors

Pdgfra and *Pdgfrb* (*Pdgf receptor β*) was highest in W-EGFP cells (Fig. 15, H and I). Similarly, DHH secreted from Sertoli cells enhances FLC differentiation (Barsoum et al., 2009; Yao et al., 2002). Indeed, *Dhh* was enriched in Sertoli cells (Fig. 14E), and the receptors *Ptch1*, *Ptch2* (*patched 2*), and *Smo* (*smoothened*) were enriched in W-EGFP cells (Fig. 15, J to L). *Gli1*, *Gli2*, and *Gli3* regulate target gene transcription upon hedgehog stimulation. Expression of the three *Gli* genes was higher in W-EGFP than in S-EGFP cells, although the difference in *Gli3* expression was not significant (Fig. 15, M to O).

In addition to the genes above, I examined the expression of the transcription factors, *Pod1* (*E-box binding transcription factor, Tcf21*) (Cui et al., 2004) and *Arx* (*aristaless related homeobox*) (Kitamura et al., 2002; Miyabayashi et al., 2013), which were shown to be implicated in FLC differentiation. Both were enriched in W-EGFP cells (Fig. 15, P and Q). *Nr2f2* (*COUP-TFII*) (Kilcoyne et al., 2014; Qin et al., 2008; van den Driesche et al., 2012) and *Nes* (Jiang et al., 2014) are expressed in ALC stem/progenitors. These gene transcripts were also enriched in W-EGFP cells (Fig. 15, R and S). Interestingly, the distribution of ARX overlapped with that of COUP-TFII in the interstitial cells of the fetal testes (Fig. 15, T to V).

scRNA-seq of interstitial cells in fetal testes

My study with *in vitro* testis reconstruction demonstrated that W-EGFP cells contain progenitors for FLCs. To characterize the progenitor cells, I adopted a scRNA-seq analysis, CEL-Seq2 (Hashimshony et al., 2016). RNAs were prepared from W-EGFP cells of E16.5 mouse fetal testes. The cDNA libraries prepared from the cells were subjected to sequencing. S-EGFP cells were used as control. 0.1 to 0.2 million of mean reads were obtained (Fig. 17A). The median numbers of the expressed genes for S-EGFP and W-EGFP cells were 4,029 and 1,294, respectively. The quality of these transcriptome data was evaluated in terms of their library sizes and the number of the detected genes as described in Material and Methods (Fig. 17B). After the filtration of the data by the evaluation above, 80 datasets from the S-EGFP cells and 341 datasets from the W-EGFP cells were subjected to subsequent analyses.

Possible candidate for FLC progenitors

To assess how many cell groups are present in the interstitial space of the fetal testis, the datasets were subjected to hierarchical clustering on the significant principal components. Cell cycle genes were filtrated out because their expression changes merely depending on cell cycle but not differentiation state (Lun et al., 2016). Eventually, as indicated in the cluster dendrogram (Fig. 18A), W-EGFP cells were divided into three clusters (cluster A, B, and C) while S-EGFP cells were into two clusters (cluster D and E). The cluster dendrogram made me realized the following

points. Among the relative distances of the clusters, cluster E is the most distant from other clusters, while cluster A and B are not largely distant. Under milder grouping condition, these cells could be divided into three clusters, cluster A+B, cluster C+D, and cluster E. When considering that cluster C and D are originated from distinct populations, W-EGFP and S-EGFP cells, respectively, the presence of cluster C+D might suggest the presence of a certain cell population differentiating from W-EGFP to S-EGFP cells.

These cell clusters were visualized by principal component analysis (PCA) and t-distributed stochastic neighboring embedding (t-SNE) (Fig. 18, B and C). Consistent with the dendrogram presentation, both plots revealed that cluster E is segregated from other clusters, while cluster A and B distribute like a single-cell group. Cluster C is localized close to cluster A and B, while cluster D is localized between cluster C and E, supporting the notion above that the cells in cluster C and D are differentiating from those in cluster A+B to cluster E. The differentially expressed genes (DEGs) were further applied to monocle trajectory analysis, which predicts the developmental order of the five cell clusters. As indicated in Fig. 18D, it was predicted that the cells in cluster A and B differentiate to those in cluster C, and eventually to those in cluster E thought cluster D. This prediction is likely to be consistent with the results shown in Fig. 18B and 18C.

Expression of steroidogenic and metabolic genes in the five clusters

To characterize the cells in the five clusters, the expression of *Ad4BP/SF-1* and steroidogenic genes were plotted. In addition, our laboratory recently identified *Mgarp* (*mitochondria localized glutamic acid rich protein*), also called *Osap* (*ovary specific acidic protein*) or *Hummr* (*hypoxia up-regulated mitochondrial movement regulator*) as a target gene of Ad4BP/SF-1 (Baba et al., 2018). Considering that MGARP facilitates cholesterol transfer to mitochondrial outer membrane (Jinn, S., 2015), *Mgarp* seems to be an essential component for steroidogenesis.

As shown in Fig. 19A, the expression of *Ad4BP/SF-1* gene in cluster D and E is higher than that in cluster A, B, and C. As expected from the function of Ad4BP/SF-1, the steroidogenic genes, *Star*, *Cyp11a1*, *Hsd3b1*, *Cyp17a1*, and *Mgarp* were shown to be expressed similarly to that of Ad4BP/SF-1; the expressions of all these steroidogenic genes were lower in cluster A, B, and C, and highest in cluster E (Fig. 19, A to F). The expressions in cluster D are higher than A and B, but lower than E, suggesting that the cells in cluster D are not fully developed yet to FLCs.

As I mentioned, the expressions of the energy metabolizing genes are higher in S-EGFP than W-EGFP cells. Thus, the expressions of the genes implicated in the processes among the five cell clusters were examined. As expected, many glycolytic gene expressions were higher in cluster E, while lower in cluster A and B (Fig. 19, G to R). Interestingly, the expressions of *Aldoa*, *Gapdh* (*glyceraldehyde 3-phosphate*

dehydrogenase), *Eno1*, and *Pkm* (*pyruvate kinase, muscle*) were largely elevated in cluster C and D (Fig. 19, L, N, Q, and R). The expression of *Gpi1* (*phosphoglucose isomerase 1*), *Tpi1* (*triosephosphate isomerase 1*), *Pgk1* (*phosphoglycerate kinase 1*), and *Pgam1* (*phosphoglycerate mutase 1*) gene in cluster C may be in the process of fully activated (Fig. 19, I, M, O, and P). The expression of *Hk1* and *Hk2* (*hexokinase 2*) was exceptionally higher only in cluster D (Fig. 19, G and H). Concerning TCA cycle, OXPHOS, PPP, and NADPH production, many gene expressions are higher in cluster D and E, while lower in cluster A and B (Fig. 20, 21, and 22). The expressions of a few genes implicated in TCA cycle, OXPHOS, PPP, and NADPH production in cluster C were at intermediate levels between cluster A and B, and cluster D and E.

Similar to the energy metabolic genes, the expression of cholesterogenic genes is activated in S-EGFP cells (Fig. 11 and 12). Thus, the expression profiles of these genes were examined. All the gene expressions were elevated in cluster E (Fig. 23). The expression levels in cluster D was varied, including fully activated, less activated and not activated. In addition to steroidogenic genes, cholesterogenic gene expressions were not activated in cluster C.

Expression of marker genes in the five clusters

Genes of which expressions are altered, whichever increased or decreased, specifically in cluster C are considered to be critical for comprehend the mechanism for

FLC differentiation. Thus, a heatmap of gene expression was generated by paying attention to the genes above (Fig. 24A). The genes in group-I showed the highest expression in cluster E or cluster E and D. As expected, FLC marker genes such as *Insl3* (*insulin-like 3*), *Cyp11a1*, and *Hsd3b1* were included in this group-I. The genes in group-II showed higher expression in cluster C, D, and E. Interestingly, several large ribosomal subunit genes were included in this group. Finally, the genes in group-III show higher expression in cluster C than others. Considering the developmental trajectory shown in Fig. 18D, it might be possible to assume that the genes in group-III might act as key players for FLC differentiation.

Unique expression of Tmsb10

Among the genes in group-III, the expression of *Tmsb10* was examined in detail. *Tmsb10* belongs to a thymosin beta family of small proteins. Another member of the family, *Tmsb4x* (*thymosin β 4x*), is quite similar to *Tmsb10* in their amino acid sequences, both of which are composed of 44 amino acids (Goldstein, 2007). The function of TMSB4X has been known to sequester actin monomer to suppress actin polymerization (Safer et al., 1991; Safer et al., 1990). Likewise, TMSB10 was reported to have the same ability (Yu et al., 1993). In addition, another study uncovered the activity of TMSB10 to suppress RAS signaling through inhibiting RAS-RAF interaction (Lee et al., 2005). However, the function of TMSB10 still remains largely controversial.

Among the five cell clusters, *Tmsb10* showed the highest expression in cluster C, and then decreased in clusters D and E. The expression in cluster C seemed to be approximately 4-fold higher than that in cluster E (Fig. 24B). *Tmsb4x* showed an expression differently from *Tmsb10* (Fig. 24C).

The expression of TMSB10 was examined with immunofluorescence in E16.5 fetal testes of *FLE-EGFP* mice where FLCs are labeled by EGFP. Sertoli and germ cells as localized in the testicular cords indicated by the broken lines in Fig. 25C, G and K were all negative for TMSB10. Likewise, S-EGFP FLCs did not express TMSB10 as indicated by the open arrowheads in Fig. 25A to D. Strong signals for TMSB10 were detected in certain cells in the interstitial space as indicated by the arrows in Fig. 25A to D. Unfortunately, however, the signals were colocalized with the signals for PECAM1, an endothelial cell marker. With careful observation, I noticed that cells showing weak TMSB10 signals are present in the interstitial space as indicated by the closed arrowheads in Fig. 25E to L. Many of these cells were weakly stained with EGFP, and these double positive cells were frequently observed around the testicular cords. Because I wanted to exclude the possibility that endothelial cells showing a strong TMSB10 signal were contaminated into W-EGFP cell population, EGFP-negative cells were examined if they contain PECAM1 positive endothelial cells. qRT-PCR revealed that the *Pecam1*-positive endothelial cells were collected in the EGFP-negative but not

W-EGFP cell fraction (Fig. 25, M to O), indicating that TMSB10-positive cells other than endothelial cells are present in the interstitial space.

Possible implication of *Tmsb10* into FLC differentiation

Because of the unique expression, I examined whether *Tmsb10* is required for FLC differentiation using knockdown (KD) technique with siRNA. First, KD efficiencies were examined with W-EGFP cells. All siRNAs for *Tmsb10*, *Tmsb4x*, and *Ad4BP/SF-1* led to clear reduction of their corresponding gene expression (Fig. 26, A to C). The expression of *Tmsb10* was likely affected even by the treatment with the siRNAs for *Tmsb4x* probably due to their similarity of the nucleotide sequences. Although I tried KD with another siRNA for *Tmsb4x*, such cross inhibition was not dissolved.

After W-EGFP cells were treated with the siRNAs, they were applied to testis reconstruction assays. As indicated in Fig. 26G, EGFP-positive (S-EGFP) cells were detected in the reconstructed testes using the W-EGFP cells treated with control siRNA (*siControl*). The treatment with the siRNA for *Tmsb10* (*siTmsb10*) resulted in reduction of S-EGFP cells in number (Fig. 26D), whereas the treatment with *Tmsb4x* siRNA (*siTmsb4x*) unlikely gave any effect (Fig. 26E). Quantitative examination indicated that the differentiation efficiency from W-EGFP to S-EGFP cells was decreased to 36.7% by the treated with *siTmsb10*, whereas not affected by *siTmsb4x* as compared to *siControl*

(Fig. 26H). *Ad4BP/SF-1* siRNA (*siAd4BP/SF-1*) suppressed the appearance of S-EGFP cells more strongly than *Tmsb10* siRNA (Fig. 25, F and H).

Possible implication of Tmsb10 into DHH signaling

It was demonstrated with genes disrupted mice that DHH secreted from Sertoli cells activates FLC differentiation (Barsoum and Yao, 2011; Brennan et al., 2003; Yao et al., 2002). Thus, I examined whether the action of DHH can be recapitulated in testis reconstruction assays. Expectedly, the treatment with a SMO agonist (SAG) increased the number of S-EGFP cells in the reconstructed testis by approximately 7-fold (Fig. 27, A to G). Concomitantly, the latency of the differentiation was shortened. Under this condition, it was examined whether *Tmsb10* affects the SAG activated differentiation. The number of S-EGFP cells increased by the SAG treatment was largely decreased by *Tmsb10* KD, while *siTmsb4x* did not give any significant effect (Fig. 27H). The treatment with *siAd4BP/SF-1* led to larger decrease. These results strongly suggested that *Tmsb10* regulates FLC differentiation activated by DHH signal.

Activation of Ad4BP/SF-1 gene transcription by Gli

FLC-specific disruption of *Ad4BP/SF-1* gene resulted in disappearance of FLCs (Shima et al., 2018). This study clearly indicated that *Ad4BP/SF-1* is indispensable for FLC differentiation, and at the same time suggested that *Ad4BP/SF-1* gene transcription

is enhanced during FLC differentiation from progenitors in W-EGFP cells. Thus, I assumed that the DHH signal might lead ultimately to activation of *Ad4BP/SF-1* gene transcription. As for the transcriptional regulation by hedgehog signal, it has been established that glioma associated Zn-finger transcription factors (GLI1 and GLI2) activate target gene expression (Hui and Angers, 2011; Jiang and Hui, 2008). Therefore, I examined whether GLI activates *Ad4BP/SF-1* gene transcription in NIH3T3 cells, which have been widely utilized to monitor hedgehog signaling (Pietrobono et al., 2018; Rohatgi et al., 2007; Tukachinsky et al., 2010). Luciferase reporter genes were constructed using the basal promoter (Ad4BPpro-luc), or FLE together with the basal promoter (FLE-Ad4BPpro-luc) of the mouse *Ad4BP/SF-1* gene. These reporter genes were co-transfected with the expression plasmid for *Gli1* or *Gli2* into NIH3T3 cells. Luciferase activity driven by FLE-Ad4BPpro-luc was increased by the plasmids. However, such increase was detected neither with Ad4BPpro-luc nor control luc plasmid (pGL3-basic) (Fig. 28, A and B). Taken together, these results strongly suggested that hedgehog signal activates *Ad4BP/SF-1* gene transcription through GLI1 and GLI2 binding to target sequence in FLE. To localize the GLI binding sites, FLE was divided into three fragments (FLE1, FLE2, and FLE3 in Fig. 28C), and they were subjected to luciferase reporter gene assays. FLE2 and FLE3 reporter genes were not activated by ectopically expressed GLI2 (Fig. 28D). In contrast, the FLE1 reporter gene was activated by GLI2 in a dose-dependent manner, suggesting the presence of

functional GLI binding sites in FLE1 but not in FLE2 and FLE3. Indeed, potential binding sites for GLI were identified in FLE1, and they are conserved among animal species (Fig. 28E).

Implication of Tmsb10 into Gli1 gene expression

To gain insights into the mechanisms how *Tmsb10* regulates hedgehog signaling, it was examined with W-EGFP cells whether *Tmsb10* KD affects the expression of the components required for hedgehog signaling. First, KD efficiencies of *siTmsb10* and *siTmsb4x* were evaluated. The amount of *Tmsb10* mRNA was largely decreased by the *siTmsb10* treatment both in the absence or presence of SAG. Unexpectedly again, the *siTmsb4x* treatment decreased slightly *Tmsb10* mRNA (Fig. 29, A and B). The amount of *Tmsb4x* mRNA was decreased by the *siTmsb4x* treatment, and any such effect was not observed in the amount of *Tmsb10*. The expression of *Tmsb10* and *Tmsb4x* were unaffected by SAG treatment (*siControl*).

Under this experimental condition, I examined the expression of hedgehog signaling components in W-EGFP cells. SAG treatment induced *Ptch1*, *Ptch2*, *Gli1*, and *Gli2* expressions (*siCnt* in Fig. 29, C, D, F, and G). Interestingly, the expression of these genes in the presence of SAG was significantly decreased by *Tmsb10* KD, although such decrease was not observed by *Tmsb4x* KD. The expression of *Smo* was affected neither by SAG treatment nor the KD (Fig. 29E). Similar results were obtained

with NIH3T3 cells (Fig. 30). Finally, I examined the expressions of the genes above in the five cell clusters identified by the single-cell transcriptome study. As shown in Fig. 31, the expression profile of *Tmsb10* is similar to that of *Gli1*.

Regulation of primary cilia by Tmsb10

It is unclear how TMSB10 regulates hedgehog signaling. Primary cilia play an essential role in transduction of hedgehog signaling (Ezratty et al., 2011; Grisanti et al., 2016; Pala et al., 2017; Schneider et al., 2010; Schneider et al., 2005; Stasiulewicz et al., 2015). They are ubiquitous microtubule-based organelles in most vertebrate cells, including testicular interstitial cells (Wainwright et al., 2014). Thus, I examined whether primary cilia formation is controlled by *Tmsb10* using immunostaining. There seems to be no difference between *Tmsb4x* knockdown and control (Fig. 32, B and C), whereas knockdown of *Tmsb10* gene results in significant decreases in ciliated cells (Fig. 32A). These results indicated that TMSB10 regulates hedgehog signaling through primary cilia.

DISCUSSION

Characterization of fetal Leydig progenitor cell population

We previously identified a fetal Leydig enhancer (FLE) of *Ad4BP/SF-1* (*Ad4 binding protein/steroidogenic factor 1*) gene, and established a transgenic mouse line in which fetal Leydig cells (FLCs) are labeled strongly with an enhanced green fluorescent protein (EGFP) (Shima et al., 2012). Interestingly, we noticed the presence of weakly labeled cells (W-EGFP cells) in the transgenic fetal testes. Marker gene expression indicated that W-EGFP cells correspond to interstitial cells except for S-EGFP FLCs (Shima et al., 2013). FLCs increase in number rapidly during the fetal days (Griswold and Behringer, 2009; O'Shaughnessy et al., 2006), although FLCs rarely proliferate (Miyabayashi et al., 2013; Orth, 1982). Therefore, the increase has been thought to be due to differentiation from progenitor cells possibly present in the interstitial space of the fetal testis (Kerr et al., 1985; Shima and Morohashi, 2017). However, it was not clarified yet whether FLC progenitors are present in the interstitial cells. In this study, I demonstrated for the first time that not all but a certain cell population in the testicular interstitial space potentially differentiate into FLCs by an *in vitro* testis reconstruction study. Although in general the efficiency of differentiation *in vitro* is possibly lower than that *in vivo*, only 20% or less of the W-EGFP cells used the *in vitro* study was found to be differentiated into FLCs. It remains to be clarified whether all interstitial cells or particular cells have the ability to differentiate into FLCs.

FLC progenitors characterized by transiently expressed genes

Implication of multiple growth factors into FLC differentiation has been studied by gene disruption. Notch signaling, for example, suppresses (Defalco et al., 2013; Tang et al., 2008) while PDGF (platelet-derived growth factor) signals activated FLC differentiation (Brennan et al., 2003). Moreover, DHH (desert hedgehog) secreted by Sertoli cells activates FLC differentiation by binding to the DHH receptors PTCH1 (patched1) and/or PTCH2 (patched2) (Barsoum and Yao, 2011; Barsoum et al., 2009; Yao et al., 2002). Although these knockout mouse studies have indicated that many growth factors regulate FLC differentiation, it has been unclear yet how the cells activated by the signals undergo differentiation into FLCs. To uncover this process, it seemed to be important to identify and characterize FLC progenitors present in W-EGFP cells.

Therefore, I obtained single-cell transcriptomes of W-EGFP cells together with S-EGFP cells. Eventually, I found a cell population characterized by the unique gene expression: *Jund* (*jun D proto-oncogene*), *Mfap2* (*microfibrillar-associated protein 2*), and *Tmsb10* (*thymosin β 10*) were found to be expressed in the particular cells more than other populations of cells. Interestingly, the analyses of the single-cell datasets predicted that the unique cells are differentiating into FLCs. If it is the case, it reasoned to assume that the genes above are expressing transiently in the progenitors.

Transient expression of genes has been reported to be critical for various cell fate determination. Mammalian sex-determining gene, *Sry* (*sex-determining region of the Y chromosome*), is transiently expressed during a sex-determining period to fix the cell fate into male supporting Sertoli cells (Albrecht and Eicher, 2001; Gubbay et al., 1990; Hawkins et al., 1992; Koopman et al., 1991; Lovell-Badge and Robertson, 1990; Sinclair et al., 1990). The transient expression is critical to enhance male-enriched expression of FGF9 (fibroblast growth factor 9) and suppress female-enriched expression of WNT4 (wingless-related MMTV integration site 4) (Colvin et al., 2001; Kim et al., 2006; Schmahl et al., 2004; Vainio et al., 1999). In fact, artificial delay of the *Sry* expression could not induce Sertoli cell differentiation (Hiramatsu et al., 2009).

Transient expression of *Mash1* (*mammalian achaete-scute homolog 1*) is necessary for differentiation to ventral, sympathetic and olfactory neurons from progenitors (Guillemot et al., 1993; Hirsch et al., 1998; Lo et al., 1991; Uchida et al., 2007).

Continued expression of *Mash1* resulted in development of neuroblastoma (Axelson, 2004; Ichimiya et al., 2001) possibly due to accelerated proliferation (Isogai et al., 2011). Based on these studies, it has been well accepted that transiently expressed genes act frequently as switches to guide cells to proper differentiation, and thus the expressions are strictly controlled to prevent any harmful effects. Likewise, the genes transiently expressed in certain interstitial of fetal testes cells may be involved in the

establishment of the FLC progenitors and/or in regulating differentiation of the progenitors into FLCs.

Possible implication of *Tmsb10* into DHH signaling to activate FLC differentiation

Among the transiently expressed genes, I paid attention to *Tmsb10* encoding a small molecular weight protein. In fact, *Tmsb10* knockdown (KD) inhibits differentiation of W-EGFP cells into FLCs. Although the functions of *Tmsb10* are still controversial, previous studies have shown that TMSB10 suppresses actin polymerization through binding to actin monomer (Yu et al., 1993). In addition, TMSB10 was reported to interfere with RAS signaling by blocking RAS-RAF interaction (Lee et al., 2005).

It has been well established that PTCH receptor upon binding to hedgehog releases SMO (smoothened) to activate the transcription factor GLI (Briscoe and Therond, 2013; Kong et al., 2019; Lee et al., 2016), and thereafter the activated GLI controls cell differentiation through regulating target genes. Being consistent with it, a SMO agonist (SAG) treatment induced the expression of GLIs in W-EGFP cells. Interestingly, the up-regulated *Gli* expression was suppressed by *Tmsb10* KD, suggesting that *Tmsb10* is implicated somehow in hedgehog signals.

Primary cilium is a structure of non-motile projection from cell surface, and composed mainly of microtubules (Singla and Reiter, 2006). Many studies have

demonstrated that primary cilia are essential for hedgehog signal transduction (Bangs and Anderson, 2017; Corbit et al., 2005; Haycraft et al., 2005; Huangfu et al., 2003; Kong et al., 2019; Rohatgi et al., 2007). In addition to hedgehog signal, it has been shown that components of other signals such as PDGF and Notch accumulate in primary cilia, suggesting that the structure act as a place for the signals to communicate mutually (Ezratty et al., 2011; Grisanti et al., 2016; Pala et al., 2017; Schneider et al., 2010; Schneider et al., 2005; Stasiulewicz et al., 2015). Considering the possible interaction between the primary cilium and *Tmsb10*, it is interesting to notice that the cilium formation is induced by down-regulation of actin polymerization (Drummond et al., 2018) and suppressed by activated KRAS (kirsten rat sarcoma viral oncogene homolog) (Lauth et al., 2010; Seeley et al., 2009). These results may suggest that *Tmsb10* promotes elongation of the primary cilium by interfering with actin polymerization and RAS signals, and consequently promotes hedgehog signal transduction in particular cells of fetal testes. Based on this hypothesis, further investigation concerning the effect of *Tmsb10* on primary cilium formation could lead to understanding of the mechanism of FLC differentiation by hedgehog signal.

Activation of metabolism during differentiation of FLCs

Steroidogenic genes are regulated directly by Ad4BP/SF-1 (Honda et al., 1993; Lala et al., 1992; Morohashi and Omura, 1996; Parker and Schimmer, 1997). The expression

of Ad4BP/SF-1 is largely increased in S-EGFP FLCs, whereas the expression in W-EGFP cells is very low. Therefore, the higher steroidogenic activity in S-EGFP FLCs can be explained by the increased expression of Ad4BP/SF-1. In addition to steroidogenesis, our recent studies with chromatin immunoprecipitation-sequence (ChIP-seq) demonstrated that not all but many genes involved in glycolytic (Baba et al., 2014) and cholesterologenic pathways (Baba et al., 2018), and key genes for nicotinamide adenine dinucleotide phosphate (NADPH) production (Li et al., 2017) are activated directly by Ad4BP/SF-1 in mouse Y-1 adrenocortical cells and adult Leydig cells. The transcriptome data obtained by my present study showed that the genes for glycolysis, pentose phosphate pathway, NADPH production, and cholesterologenesis are expressed at higher levels in S-EGFP than W-EGFP cells. Although ChIP-seq for Ad4BP/SF-1 has not performed yet with FLCs, my study together with the data obtained by our group strongly suggested that these genes are activated during the FLC differentiation directly by Ad4BP/SF-1.

In addition to these metabolic pathways, transcriptomes obtained by my study reveals that the genes involved in tricarboxylic acid cycle (TCA cycle) and oxidative phosphorylation (OXPHOS) are activated in S-EGFP cells. However, it has been unclear whether Ad4BP/SF-1 regulates directly these metabolic genes. Concerning the possible implication of Ad4BP/SF-1 into these metabolic pathways, it is interesting to notice the ChIP-seq data with another nuclear receptor, ERR α (estrogen-related

receptor α). $ERR\alpha$ belongs to a nuclear receptor subfamily same with Ad4BP/SF-1. Previously published studies with ChIP-seq demonstrated that $ERR\alpha$ regulates the genes for glycolysis, TCA cycle, OXPHOS, and lipid metabolism in the mouse liver through direct binding (Charest-Marcotte et al., 2010; Chaveroux et al., 2013). As the consensus binding sequence of $ERR\alpha$ is shared by Ad4BP/SF-1 (Morohashi et al., 1992; Sladek et al., 1997), Ad4BP/SF-1 may regulate the genes involved in TCA cycle, OXPHOS, and lipid metabolism under certain physiological conditions.

The present scRNA-seq demonstrated that the expression of some, if not all, glycolytic genes start to increase in the putative progenitors for FLCs. Although such increase was not obvious in TCA cycle and OXPHOS genes, these results suggested a possibility that up-regulation of glycolysis might be prerequisite for or simultaneous with FLC differentiation. More interestingly, such expression of Ad4BP/SF-1 in the putative progenitor cells did not occur. Thus, the increased expression of the glycolytic genes in the putative progenitors might be due to yet-unknown factor other than Ad4BP/SF-1. Growth factors are also capable of regulating metabolisms through noncanonical pathways. Notch signaling affects glycolytic activity through a phosphatidylinositol 3-kinase/AKT serine/threonine kinase pathway or in a p53-dependent manner (Landor et al., 2011). Hedgehog signaling activates glycolysis through enhanced expression of glycolytic genes (Di Magno et al., 2014). and/or through enhanced phosphorylation of MAPK pathway (Ge et al., 2015). Onset of

glycolytic gene activation earlier than that of Ad4BP/SF-1 in the putative progenitor cell population may be directly activated by growth factors through the noncanonical pathways.

Activation of Ad4BP/SF-1 expression by Gli

Our laboratory previously identified an FLE in the upstream region of *Ad4BP/SF-1* gene (Shima et al., 2012). Investigation of the enhancer function identified two *cis*-acting sequences, Ad4 site and E-box. The presence of the functional Ad4 site strongly suggested that Ad4BP/SF-1 forms an autoregulatory loop to maintain its own gene expression. Similar autoregulatory loop was found in the fetal adrenal-specific enhancer of *Ad4BP/SF-1* gene to maintain the expression in the tissue (Zubair et al., 2006; Zubair et al., 2009). In addition to the maintenance step, it is reasonable to assume that a step for initial activation occurs. In this study, I found functional GLI binding sites in FLE. Indeed, reporter gene analyses showed that GLI1/GLI2 could activate the expression of *Ad4BP/SF-1* via FLE, suggesting that DHH signaling may drive the first step of *Ad4BP/SF-1* gene transcription in addition to activation of glycolytic genes.

REFERENCES

Albrecht, K.H., and Eicher, E.M. (2001). Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor. *Developmental biology* 240, 92-107.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* 31, 166-169.

Baba, T., Otake, H., Inoue, M., Sato, T., Ishihara, Y., Moon, J.-Y., Tsuchiya, M., Miyabayashi, K., Ogawa, H., Shima, Y., *et al.* (2018). Ad4BP/SF-1 regulates cholesterol synthesis to boost the production of steroids. *Communications Biology* 1, 18.

Baba, T., Otake, H., Sato, T., Miyabayashi, K., Shishido, Y., Wang, C.Y., Shima, Y., Kimura, H., Yagi, M., Ishihara, Y., *et al.* (2014). Glycolytic genes are targets of the nuclear receptor Ad4BP/SF-1. *Nature communications* 5, 3634.

Bangs, F., and Anderson, K.V. (2017). Primary Cilia and Mammalian Hedgehog Signaling. *Cold Spring Harbor perspectives in biology* 9.

Barsoum, I., and Yao, H.H. (2011). Redundant and differential roles of transcription factors Gli1 and Gli2 in the development of mouse fetal Leydig cells. *Biology of reproduction* 84, 894-899.

Barsoum, I.B., Bingham, N.C., Parker, K.L., Jorgensen, J.S., and Yao, H.H.C. (2009). Activation of the Hedgehog pathway in the mouse fetal ovary leads to ectopic

appearance of fetal Leydig cells and female pseudohermaphroditism. *Developmental biology* 329, 96-103.

Barsoum, I.B., Kaur, J., Ge, R.S., Cooke, P.S., and Yao, H.H. (2013). Dynamic changes in fetal Leydig cell populations influence adult Leydig cell populations in mice. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 27, 2657-2666.

Bielinska, M., Seehra, A., Toppari, J., Heikinheimo, M., and Wilson, D.B. (2007). GATA-4 is required for sex steroidogenic cell development in the fetal mouse. *Developmental dynamics : an official publication of the American Association of Anatomists* 236, 203-213.

Birk, O.S., Casiano, D.E., Wassif, C.A., Cogliati, T., Zhao, L., Zhao, Y., Grinberg, A., Huang, S., Kreidberg, J.A., Parker, K.L., *et al.* (2000). The LIM homeobox gene *Lhx9* is essential for mouse gonad formation. *Nature* 403, 909-913.

Brennan, J., and Capel, B. (2004). One tissue, two fates: molecular genetic events that underlie testis versus ovary development. *Nature reviews Genetics* 5, 509-521.

Brennan, J., Tilmann, C., and Capel, B. (2003). *Pdgfr*-alpha mediates testis cord organization and fetal Leydig cell development in the XY gonad. *Genes & development* 17, 800-810.

Briscoe, J., and Therond, P.P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature reviews Molecular cell biology* 14, 416-429.

Buaas, F.W., Gardiner, J.R., Clayton, S., Val, P., and Swain, A. (2012). In vivo evidence for the crucial role of SF1 in steroid-producing cells of the testis, ovary and adrenal gland. *Development* *139*, 4561-4570.

Charest-Marcotte, A., Dufour, C.R., Wilson, B.J., Tremblay, A.M., Eichner, L.J., Arlow, D.H., Mootha, V.K., and Giguere, V. (2010). The homeobox protein Prox1 is a negative modulator of ERR α /PGC-1 α bioenergetic functions. *Genes & development* *24*, 537-542.

Chaveroux, C., Eichner, L.J., Dufour, C.R., Shatnawi, A., Khoutorsky, A., Bourque, G., Sonenberg, N., and Giguere, V. (2013). Molecular and genetic crosstalks between mTOR and ERR α are key determinants of rapamycin-induced nonalcoholic fatty liver. *Cell metabolism* *17*, 586-598.

Colvin, J.S., Green, R.P., Schmahl, J., Capel, B., and Ornitz, D.M. (2001). Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell* *104*, 875-889.

Consortium., S.M.-I. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature biotechnology* *32*, 903-914.

Corbit, K.C., Aanstad, P., Singla, V., Norman, A.R., Stainier, D.Y., and Reiter, J.F. (2005). Vertebrate Smoothed functions at the primary cilium. *Nature* *437*, 1018-1021.

Cui, S., Ross, A., Stallings, N., Parker, K.L., Capel, B., and Quaggin, S.E. (2004).
Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice.
Development *131*, 4095-4105.

DeFalco, T., Bhattacharya, I., Williams, A.V., Sams, D.M., and Capel, B. (2014).
Yolk-sac-derived macrophages regulate fetal testis vascularization and morphogenesis.
Proceedings of the National Academy of Sciences of the United States of America *111*,
E2384-2393.

Defalco, T., Saraswathula, A., Briot, A., Iruela-Arispe, M.L., and Capel, B. (2013).
Testosterone levels influence mouse fetal Leydig cell progenitors through notch
signaling. *Biology of reproduction* *88*, 91.

DeFalco, T., Takahashi, S., and Capel, B. (2011). Two distinct origins for Leydig cell
progenitors in the fetal testis. *Developmental biology* *352*, 14-26.

Di Magno, L., Manzi, D., D'Amico, D., Coni, S., Macone, A., Infante, P., Di
Marcotullio, L., De Smaele, E., Ferretti, E., Screpanti, I., *et al.* (2014). Druggable
glycolytic requirement for Hedgehog-dependent neuronal and medulloblastoma growth.
Cell cycle (Georgetown, Tex) *13*, 3404-3413.

Drummond, M.L., Li, M., Tarapore, E., Nguyen, T.T.L., Barouni, B.J., Cruz, S., Tan,
K.C., Oro, A.E., and Atwood, S.X. (2018). Actin polymerization controls cilia-mediated
signaling. *The Journal of cell biology* *217*, 3255-3266.

Ezratty, E.J., Stokes, N., Chai, S., Shah, A.S., Williams, S.E., and Fuchs, E. (2011). A role for the primary cilium in Notch signaling and epidermal differentiation during skin development. *Cell* *145*, 1129-1141.

Fischer, A., and Gessler, M. (2007). Delta-Notch--and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. *Nucleic acids research* *35*, 4583-4596.

Ge, R.S., Dong, Q., Sottas, C.M., Papadopoulos, V., Zirkin, B.R., and Hardy, M.P. (2006). In search of rat stem Leydig cells: identification, isolation, and lineage-specific development. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 2719-2724.

Ge, X., Lyu, P., Gu, Y., Li, L., Li, J., Wang, Y., Zhang, L., Fu, C., and Cao, Z. (2015). Sonic hedgehog stimulates glycolysis and proliferation of breast cancer cells: Modulation of PFKFB3 activation. *Biochemical and biophysical research communications* *464*, 862-868.

Goldstein, A.L. (2007). History of the discovery of the thymosins. *Annals of the New York Academy of Sciences* *1112*, 1-13.

Grisanti, L., Revenkova, E., Gordon, R.E., and Iomini, C. (2016). Primary cilia maintain corneal epithelial homeostasis by regulation of the Notch signaling pathway. *Development* *143*, 2160-2171.

Griswold, S.L., and Behringer, R.R. (2009). Fetal Leydig cell origin and development.

Sexual development : genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation 3, 1-15.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346, 245-250.

Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C., and Schedl, A. (2001). Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell* 106, 319-329.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., *et al.* (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome biology* 17, 77.

Hatano, O., Takayama, K., Imai, T., Waterman, M.R., Takakusu, A., Omura, T., and Morohashi, K. (1994). Sex-dependent expression of a transcription factor, Ad4BP, regulating steroidogenic P-450 genes in the gonads during prenatal and postnatal rat development. *Development* 120, 2787-2797.

Hawkins, J.R., Taylor, A., Berta, P., Levilliers, J., Van der Auwera, B., and Goodfellow, P.N. (1992). Mutational analysis of SRY: nonsense and missense mutations in XY sex reversal. *Human genetics* 88, 471-474.

Haycraft, C.J., Banizs, B., Aydin-Son, Y., Zhang, Q., Michaud, E.J., and Yoder, B.K. (2005). Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS genetics* 1, e53.

Hiramatsu, R., Matoba, S., Kanai-Azuma, M., Tsunekawa, N., Katoh-Fukui, Y., Kurohmaru, M., Morohashi, K., Wilhelm, D., Koopman, P., and Kanai, Y. (2009). A critical time window of Sry action in gonadal sex determination in mice. *Development* 136, 129-138.

Honda, S., Morohashi, K., Nomura, M., Takeya, H., Kitajima, M., and Omura, T. (1993). Ad4BP regulating steroidogenic P-450 gene is a member of steroid hormone receptor superfamily. *The Journal of biological chemistry* 268, 7494-7502.

Huangfu, D., Liu, A., Rakeman, A.S., Murcia, N.S., Niswander, L., and Anderson, K.V. (2003). Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature* 426, 83-87.

Hui, C.C., and Angers, S. (2011). Gli proteins in development and disease. *Annual review of cell and developmental biology* 27, 513-537.

Jameson, S.A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D.M., Mork, L., Munger, S.C., and Capel, B. (2012). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS genetics* 8, e1002575.

- Jiang, J., and Hui, C.C. (2008). Hedgehog signaling in development and cancer. *Developmental cell* 15, 801-812.
- Jiang, M.H., Cai, B., Tuo, Y., Wang, J., Zang, Z.J., Tu, X., Gao, Y., Su, Z., Li, W., Li, G., *et al.* (2014). Characterization of Nestin-positive stem Leydig cells as a potential source for the treatment of testicular Leydig cell dysfunction. *Cell research* 24, 1466-1485.
- Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics* 9, 286-298.
- Katoh-Fukui, Y., Tsuchiya, R., Shiroishi, T., Nakahara, Y., Hashimoto, N., Noguchi, K., and Higashinakagawa, T. (1998). Male-to-female sex reversal in M33 mutant mice. *Nature* 393, 688-692.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome research* 12, 996-1006.
- Kerr, J.B., Donachie, K., and Rommerts, F.F. (1985). Selective destruction and regeneration of rat Leydig cells in vivo. A new method for the study of seminiferous tubular-interstitial tissue interaction. *Cell and tissue research* 242, 145-156.
- Kerr, J.B., and Knell, C.M. (1988). The fate of fetal Leydig cells during the development of the fetal and postnatal rat testis. *Development* 103, 535-544.

Kidokoro, T., Matoba, S., Hiramatsu, R., Fujisawa, M., Kanai-Azuma, M., Taya, C., Kurohmaru, M., Kawakami, H., Hayashi, Y., Kanai, Y., *et al.* (2005). Influence on spatiotemporal patterns of a male-specific Sox9 activation by ectopic Sry expression during early phases of testis differentiation in mice. *Developmental biology* 278, 511-525.

Kilcoyne, K.R., Smith, L.B., Atanassova, N., Macpherson, S., McKinnell, C., van den Driesche, S., Jobling, M.S., Chambers, T.J., De Gendt, K., Verhoeven, G., *et al.* (2014). Fetal programming of adult Leydig cell function by androgenic effects on stem/progenitor cells. *Proceedings of the National Academy of Sciences of the United States of America* 111, E1924-1932.

Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.C., Poulat, F., Behringer, R.R., Lovell-Badge, R., and Capel, B. (2006). Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. *PLoS biology* 4, e187.

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., *et al.* (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nature genetics* 32, 359-369.

Kong, J.H., Siebold, C., and Rohatgi, R. (2019). Biochemical mechanisms of vertebrate hedgehog signaling. *Development* 146.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991).

Male development of chromosomally female mice transgenic for Sry. *Nature* *351*, 117-121.

Lala, D.S., Rice, D.A., and Parker, K.L. (1992). Steroidogenic factor I, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor I.

Molecular endocrinology *6*, 1249-1258.

Landor, S.K., Mutvei, A.P., Mamaeva, V., Jin, S., Busk, M., Borra, R., Gronroos, T.J., Kronqvist, P., Lendahl, U., and Sahlgren, C.M. (2011). Hypo- and hyperactivated Notch signaling induce a glycolytic switch through distinct mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* *108*, 18814-18819.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* *9*, 357-359.

Lauth, M., Bergstrom, A., Shimokawa, T., Tostar, U., Jin, Q., Fendrich, V., Guerra, C., Barbacid, M., and Toftgard, R. (2010). DYRK1B-dependent autocrine-to-paracrine shift of Hedgehog signaling by mutant RAS. *Nature structural & molecular biology* *17*, 718-725.

Lee, R.T., Zhao, Z., and Ingham, P.W. (2016). Hedgehog signalling. *Development* *143*, 367-372.

Lee, S.H., Son, M.J., Oh, S.H., Rho, S.B., Park, K., Kim, Y.J., Park, M.S., and Lee, J.H. (2005). Thymosin β (10) inhibits angiogenesis and tumor growth by interfering with Ras function. *Cancer research* 65, 137-148.

Leydig, F. (1850). *Zur anatomie der mannlichen geschlechtsorgane und anldrusen der saugethiere*, Vol 2 (Z Wiss Zool).

Li, B., Baba, T., Miyabayashi, K., Sato, T., Shima, Y., Ichinose, T., Miura, D., Ohkawa, Y., Suyama, M., and Morohashi, K.I. (2017). Role of Ad4-binding protein/steroidogenic factor 1 in regulating NADPH production in adrenocortical Y-1 cells. *Endocrine journal* 64, 315-324.

Lovell-Badge, R., and Robertson, E. (1990). XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. *Development* 109, 635-646.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* 5, 2122.

Luo, X., Ikeda, Y., and Parker, K.L. (1994). A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* 77, 481-490.

Martin, L.J. (2016). Cell interactions and genetic regulation that contribute to testicular Leydig cell development and differentiation. *Molecular reproduction and development* 83, 470-487.

McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., and Bowles, J. (2015). Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells. *Biology of reproduction* 92, 145.

Meeks, J.J., Crawford, S.E., Russell, T.A., Morohashi, K.-i., Weiss, J., and Jameson, J.L. (2003). Dax1 regulates testis cord organization during gonadal differentiation. *Development* 130, 1029-1036.

Miyabayashi, K., Katoh-Fukui, Y., Ogawa, H., Baba, T., Shima, Y., Sugiyama, N., Kitamura, K., and Morohashi, K. (2013). Aristaless related homeobox gene, Arx, is implicated in mouse fetal Leydig cell differentiation possibly through expressing in the progenitor cells. *PLoS One* 8, e68050.

Miyabayashi, K., Tokunaga, K., Otake, H., Baba, T., Shima, Y., and Morohashi, K.-i. (2015). Heterogeneity of Ovarian Theca and Interstitial Gland Cells in Mice. *PLoS ONE* 10, e0128352.

Miyado, M., Nakamura, M., Miyado, K., Morohashi, K.-i., Sano, S., Nagata, E., Fukami, M., and Ogata, T. (2012). Maml1 Deficiency Significantly Reduces mRNA Expression Levels of Multiple Genes Expressed in Mouse Fetal Leydig Cells but Permits Normal Genital and Reproductive Development. *Endocrinology* 153, 6033-6040.

Miyamoto, N., Yoshida, M., Kuratani, S., Matsuo, I., and Aizawa, S. (1997). Defects of urogenital development in mice lacking *Emx2*. *Development* *124*, 1653-1664.

Morohashi, K., Honda, S., Inomata, Y., Handa, H., and Omura, T. (1992). A common trans-acting factor, Ad4-binding protein, to the promoters of steroidogenic P-450s. *The Journal of biological chemistry* *267*, 17913-17919.

Morohashi, K.I., and Omura, T. (1996). Ad4BP/SF-1, a transcription factor essential for the transcription of steroidogenic cytochrome P450 genes and for the establishment of the reproductive function. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* *10*, 1569-1577.

Nel-Themaat, L., Vadakkan, T.J., Wang, Y., Dickinson, M.E., Akiyama, H., and Behringer, R.R. (2009). Morphometric analysis of testis cord formation in *Sox9*-EGFP mice. *Developmental dynamics : an official publication of the American Association of Anatomists* *238*, 1100-1110.

O'Shaughnessy, P.J., Baker, P.J., Heikkila, M., Vainio, S., and McMahon, A.P. (2000). Localization of 17 β -hydroxysteroid dehydrogenase/17-ketosteroid reductase isoform expression in the developing mouse testis--androstenedione is the major androgen secreted by fetal/neonatal leydig cells. *Endocrinology* *141*, 2631-2637.

O'Shaughnessy, P.J., Baker, P.J., and Johnston, H. (2006). The foetal Leydig cell--differentiation, function and regulation. *International journal of andrology* *29*, 90-95; discussion 105-108.

Ogata, T., Sano, S., Nagata, E., Kato, F., and Fukami, M. (2012). MAMLD1 and 46,XY disorders of sex development. *Seminars in reproductive medicine* 30, 410-416.

Orth, J.M. (1982). Proliferation of Sertoli cells in fetal and postnatal rats: a quantitative autoradiographic study. *The Anatomical record* 203, 485-492.

Pala, R., Alomari, N., and Nauli, S.M. (2017). Primary Cilium-Dependent Signaling Mechanisms. *International journal of molecular sciences* 18.

Park, S.Y., Meeks, J.J., Raverot, G., Pfaff, L.E., Weiss, J., Hammer, G.D., and Jameson, J.L. (2005). Nuclear receptors Sf1 and Dax1 function cooperatively to mediate somatic cell differentiation during testis development. *Development* 132, 2415-2423.

Parker, K.L., and Schimmer, B.P. (1997). Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocrine reviews* 18, 361-377.

Pietrobono, S., Santini, R., Gagliardi, S., Dapporto, F., Colecchia, D., Chiariello, M., Leone, C., Valoti, M., Manetti, F., Petricci, E., *et al.* (2018). Targeted inhibition of Hedgehog-GLI signaling by novel acylguanidine derivatives inhibits melanoma cell growth by inducing replication stress and mitotic catastrophe. *Cell death & disease* 9, 142.

Potter, S.J., and DeFalco, T. (2017). Role of the testis interstitial compartment in spermatogonial stem cell function. *Reproduction (Cambridge, England)* 153, R151-r162.

Qin, J., Tsai, M.J., and Tsai, S.Y. (2008). Essential roles of COUP-TFII in Leydig cell differentiation and male fertility. *PLoS One* 3, e3285.

Rohatgi, R., Milenkovic, L., and Scott, M.P. (2007). Patched1 regulates hedgehog signaling at the primary cilium. *Science (New York, NY)* 317, 372-376.

Roosen-Runge, E.C., and Anderson, D. (1959). The development of the interstitial cells in the testis of the albino rat. *Acta anatomica* 37, 125-137.

Safer, D., Elzinga, M., and Nachmias, V.T. (1991). Thymosin beta 4 and Fx, an actin-sequestering peptide, are indistinguishable. *The Journal of biological chemistry* 266, 4029-4032.

Safer, D., Golla, R., and Nachmias, V.T. (1990). Isolation of a 5-kilodalton actin-sequestering peptide from human blood platelets. *Proceedings of the National Academy of Sciences of the United States of America* 87, 2536-2540.

Schmahl, J., Kim, Y., Colvin, J.S., Ornitz, D.M., and Capel, B. (2004). Fgf9 induces proliferation and nuclear localization of FGFR2 in Sertoli precursors during male sex determination. *Development* 131, 3627-3636.

Schneider, L., Cammer, M., Lehman, J., Nielsen, S.K., Guerra, C.F., Veland, I.R., Stock, C., Hoffmann, E.K., Yoder, B.K., Schwab, A., *et al.* (2010). Directional cell migration and chemotaxis in wound healing response to PDGF-AA are coordinated by the primary cilium in fibroblasts. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 25, 279-292.

Schneider, L., Clement, C.A., Teilmann, S.C., Pazour, G.J., Hoffmann, E.K., Satir, P., and Christensen, S.T. (2005). PDGFR α signaling is regulated through the primary cilium in fibroblasts. *Current biology* : CB 15, 1861-1866.

Seeley, E.S., Carriere, C., Goetze, T., Longnecker, D.S., and Korc, M. (2009).

Pancreatic cancer and precursor pancreatic intraepithelial neoplasia lesions are devoid of primary cilia. *Cancer research* 69, 422-430.

Sekido, R., Bar, I., Narvaez, V., Penny, G., and Lovell-Badge, R. (2004). SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. *Developmental biology* 274, 271-279.

Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 453, 930-934.

Shima, Y., Matsuzaki, S., Miyabayashi, K., Otake, H., Baba, T., Kato, S., Huhtaniemi, I., and Morohashi, K. (2015). Fetal Leydig Cells Persist as an Androgen-Independent Subpopulation in the Postnatal Testis. *Molecular endocrinology* 29, 1581-1593.

Shima, Y., Miyabayashi, K., Baba, T., Otake, H., Katsura, Y., Oka, S., Zubair, M., and Morohashi, K. (2012). Identification of an enhancer in the Ad4BP/SF-1 gene specific for fetal Leydig cells. *Endocrinology* 153, 417-425.

Shima, Y., Miyabayashi, K., Haraguchi, S., Arakawa, T., Otake, H., Baba, T.,

Matsuzaki, S., Shishido, Y., Akiyama, H., Tachibana, T., *et al.* (2013). Contribution of

Leydig and Sertoli cells to testosterone production in mouse fetal testes. *Molecular endocrinology* 27, 63-73.

Shima, Y., Miyabayashi, K., Sato, T., Suyama, M., Ohkawa, Y., Doi, M., Okamura, H., and Suzuki, K. (2018). Fetal Leydig cells dedifferentiate and serve as adult Leydig stem cells. *Development* 145.

Shima, Y., and Morohashi, K.I. (2017). Leydig progenitor cells in fetal testis. *Molecular and cellular endocrinology* 445, 55-64.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346, 240-244.

Singla, V., and Reiter, J.F. (2006). The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science (New York, NY)* 313, 629-633.

Sladek, R., Bader, J.A., and Giguere, V. (1997). The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. *Molecular and cellular biology* 17, 5400-5409.

Stasiulewicz, M., Gray, S.D., Mastromina, I., Silva, J.C., Bjorklund, M., Seymour, P.A., Booth, D., Thompson, C., Green, R.J., Hall, E.A., *et al.* (2015). A conserved role for

Notch signaling in priming the cellular response to Shh through ciliary localisation of the key Shh transducer Smo. *Development* *142*, 2291-2303.

Stevant, I., Neirijnck, Y., Borel, C., Escoffier, J., Smith, L.B., Antonarakis, S.E., Dermitzakis, E.T., and Nef, S. (2018). Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell reports* *22*, 1589-1599.

Svingen, T., and Koopman, P. (2013). Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes & development* *27*, 2409-2426.

Tang, H., Brennan, J., Karl, J., Hamada, Y., Raetzman, L., and Capel, B. (2008). Notch signaling maintains Leydig progenitor cells in the mouse testis. *Development* *135*, 3745-3753.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* *25*, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* *7*, 562-578.

Tukachinsky, H., Lopez, L.V., and Salic, A. (2010). A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. *The Journal of cell biology* *191*, 415-428.

Vainio, S., Heikkila, M., Kispert, A., Chin, N., and McMahon, A.P. (1999). Female development in mammals is regulated by Wnt-4 signalling. *Nature* *397*, 405-409.

van den Driesche, S., Walker, M., McKinnell, C., Scott, H.M., Eddie, S.L., Mitchell, R.T., Seckl, J.R., Drake, A.J., Smith, L.B., Anderson, R.A., *et al.* (2012). Proposed role for COUP-TFII in regulating fetal Leydig cell steroidogenesis, perturbation of which leads to masculinization disorders in rodents. *PLoS One* *7*, e37064.

Wainwright, E.N., Svingen, T., Ng, E.T., Wicking, C., and Koopman, P. (2014). Primary cilia function regulates the length of the embryonic trunk axis and urogenital field in mice. *Developmental biology* *395*, 342-354.

Wen, Q., Wang, Y., Tang, J., Cheng, C.Y., and Liu, Y.X. (2016). Sertoli Cell Wt1 Regulates Peritubular Myoid Cell and Fetal Leydig Cell Differentiation during Fetal Testis Development. *PLoS One* *11*, e0167920.

Wen, Q., Zheng, Q.S., Li, X.X., Hu, Z.Y., Gao, F., Cheng, C.Y., and Liu, Y.X. (2014). Wt1 dictates the fate of fetal and adult Leydig cells during development in the mouse testis. *American journal of physiology Endocrinology and metabolism* *307*, E1131-1143.

Wilhelm, D., Hiramatsu, R., Mizusaki, H., Widjaja, L., Combes, A.N., Kanai, Y., and Koopman, P. (2007). SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development. *The Journal of biological chemistry* 282, 10553-10560.

Yao, H.H., Whoriskey, W., and Capel, B. (2002). Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes & development* 16, 1433-1440.

Ye, L., Li, X., Li, L., Chen, H., and Ge, R.S. (2017). Insights into the Development of the Adult Leydig Cell Lineage from Stem Leydig Cells. *Frontiers in physiology* 8, 430.

Yokonishi, T., Sato, T., Katagiri, K., Komeya, M., Kubota, Y., and Ogawa, T. (2013). In Vitro Reconstruction of Mouse Seminiferous Tubules Supporting Germ Cell Differentiation. *Biology of reproduction*.

Yokoyama, C., Chigi, Y., Baba, T., Ohshitanai, A., Harada, Y., Takahashi, F., and Morohashi, K.I. (2019). Three populations of adult Leydig cells in mouse testes revealed by a novel mouse HSD3B1-specific rat monoclonal antibody. *Biochemical and biophysical research communications* 511, 916-920.

Yu, F.X., Lin, S.C., Morrison-Bogorad, M., Atkinson, M.A., and Yin, H.L. (1993). Thymosin beta 10 and thymosin beta 4 are both actin monomer sequestering proteins. *The Journal of biological chemistry* 268, 502-509.

Zubair, M., Ishihara, S., Oka, S., Okumura, K., and Morohashi, K. (2006). Two-step regulation of Ad4BP/SF-1 gene transcription during fetal adrenal development: initiation by a Hox-Pbx1-Prep1 complex and maintenance via autoregulation by Ad4BP/SF-1. *Molecular and cellular biology* 26, 4111-4121.

Zubair, M., Oka, S., Parker, K.L., and Morohashi, K. (2009). Transgenic expression of Ad4BP/SF-1 in fetal adrenal progenitor cells leads to ectopic adrenal formation. *Molecular endocrinology* 23, 1657-1667.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Prof. Ken-ichirou Morohashi for the exact guidance and constant support throughout my study. I wish to express my gratitude to Dr. Takashi Baba for helpful discussion and technical advice. I am also grateful to Dr. Kanako Miyabayashi and Dr. Yuichi Shima (Department of Anatomy, Kawasaki Medical School, Kurashiki, Okayama, Japan) for their help on transgenic mouse studies, helpful discussion, and technical advice. I thank all the current and past members of Morohashi laboratory.

I would like to thank Dr. Tetsuhiro Yokonishi and Dr. Takehiko Ogawa (Institute of Molecular Medicine and Life Science, Yokohama City University, Yokohama, Japan) for instructing *in vitro* testis reconstruction. I also would like to thank Dr. Akihito Harada and Dr. Yasuyuki Ohkawa (Division of Transcriptomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka Japan) for their technical supports for single-cell and deep sequencing studies. I am grateful to Dr. Tetsuya Sato, Dr. Daisuke Saito, and Dr. Mikita Suyama (Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University, Fukuoka Japan) for their help for computational analyses of the sequence datasets. I thank Dr. Haruhiko Akiyama (Department of Orthopaedics, Gifu University Graduate School of Medicine, Gifu, Japan) for his kind gift of *Sox9-EGFP* mice. I appreciate the technical assistance from

The Research Support Center, Research Center for Human Disease Modeling, Kyushu
University Graduate School of Medical Sciences.

FIGURES

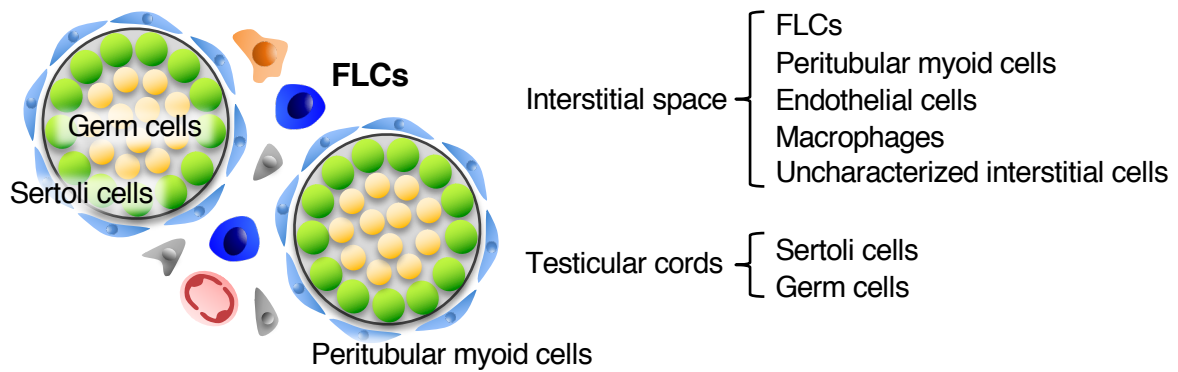


Fig. 1, Structure and cellular components of a fetal testis

Fetal testis is thematically illustrated. The fetal testis is divided into two distinct compartments: testicular cords (called seminiferous tubules after birth) and interstitial space. Sertoli (indicated with green) and germ cells (indicated with light yellow) are localized inside the testicular cords, while a variety of cell types, such as FLCs (indicated with blue), peritubular myoid cells (indicated with light blue), endothelial cells (indicated with red), macrophages (indicated with orange), and uncharacterized interstitial cells (indicated with gray), are localized in the interstitial space.

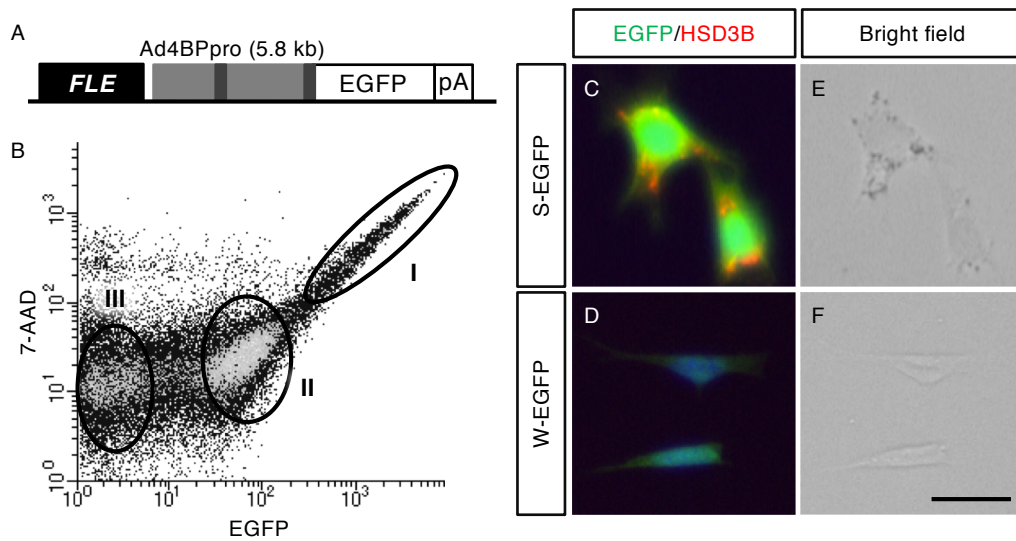


Fig. 2, S-EGFP and W-EGFP cells prepared from the fetal testes of *FLE-EGFP* mice

The construct of *FLE-EGFP* transgenic mouse is shown (A). The cells prepared from *FLE-EGFP* mouse testes (E16.5) were subjected to FACS after staining with 7-amino-actinomycin D (B). Fractions I, II and III contain S-EGFP, W-EGFP, and EGFP-negative cells, respectively. S-EGFP and W-EGFP cells were cultured for 4 hrs (C-F), and subjected to immunostaining for EGFP (green) and HSD3B (red). Nuclei were stained with DAPI (blue). Overlaid images are shown (B, C). Bright field images for S-EGFP (C) and W-EGFP cells (D) are shown. Scale bar = 25 μm.

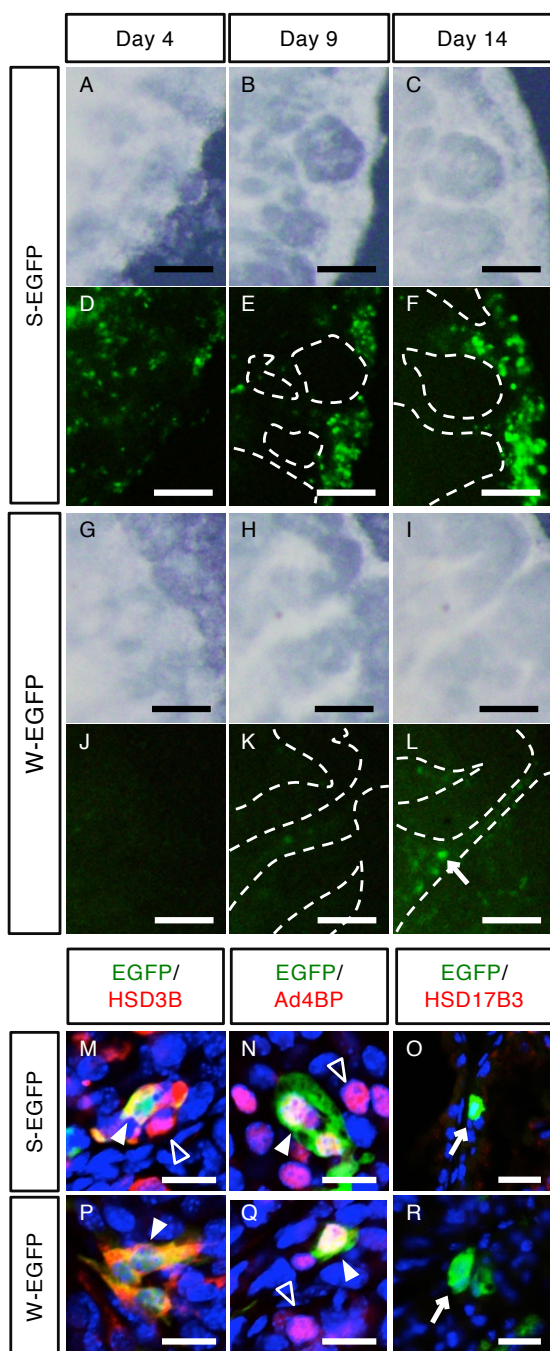


Fig. 3, Differentiation of S-EGFP to W-EGFP cells in reconstructed testes

Whole cells prepared from wild type testes (E16.5) were mixed with S-EGFP (A–F) and W-EGFP cells (G–L), and thereafter cultured for 4 (A, D, G, J), 9 (B, E, H, K), and 14 days (C, F, I, L). These reconstructed testes were observed under a microscope (A–C, G–I) or a fluorescence microscope (D–F, J–L). The arrow in L indicates EGFP-positive cells emerged after the reconstructed testes with W-EGFP cells were incubation for 14 days. Broken lines in E, F, K, and L indicate testicular cord-like structures. Scale bar = 50 μm . The reconstructed tissues with S-EGFP (M–O) and W-EGFP cells (P–R) at day 35 were subjected to double immunostaining with the antibodies to EGFP (green) and HSD3B (red) (M, P), EGFP (green) and Ad4BP/SF-1 (red) (N, Q), and EGFP (green) and HSD17B3 (red) (O, R). Nuclei were stained with DAPI (blue). Overlaid images are shown. Closed and open arrowheads in M and P indicate double-positive cells for EGFP and HSD3B, and a single-positive cell for HSD3B, respectively. Closed and open arrowheads in N and Q indicate double-positive cells for EGFP and Ad4BP/SF-1, and single-positive cells for Ad4BP/SF-1, respectively. Arrows in O and R indicate single positive cells for EGFP. Scale bar = 10 μm (M, N, P, and Q), 25 μm (O and R).

A	Seq-I		Seq-II		Sertoli cell
	S-EGFP	W-EGFP	S-EGFP	W-EGFP	
Total reads	41,976,499	43,462,406	17,451,881	26,652,369	42,123,295
Mapping rate (%)	98.4	98.4	98.6	99.2	98.3

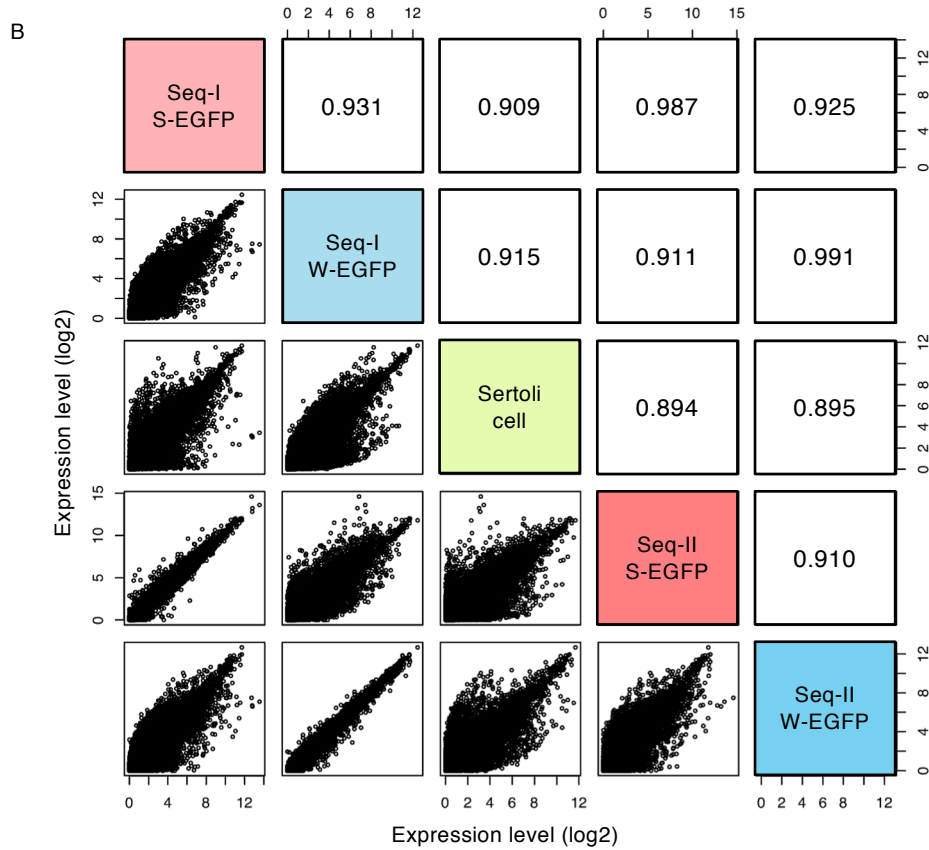


Fig. 4, Comparison of gene expression between S-EGFP, W-EGFP, and Sertoli cells

Sequence datasets were obtained from S-EGFP, W-EGFP, and Sertoli cells prepared from fetal testes (E16.5) by FACS. The Sequencing was performed with biological duplicates of S-EGFP and W-EGFP cells (Seq-I and Seq-II). More than 15 million reads were obtained from every sample, and 98–99% of the reads were mapped to the reference genome (A). Gene expression was compared between the samples. Scatter plots (log₂ scale) are shown below the diagonal and the correlation coefficients above the diagonal (B).

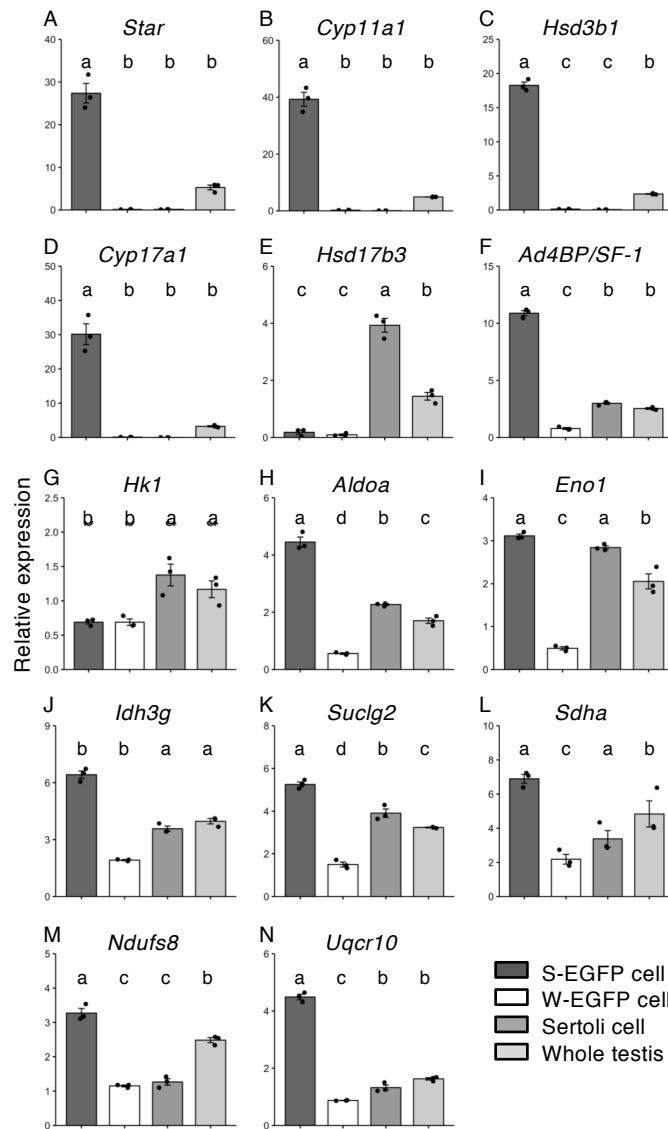
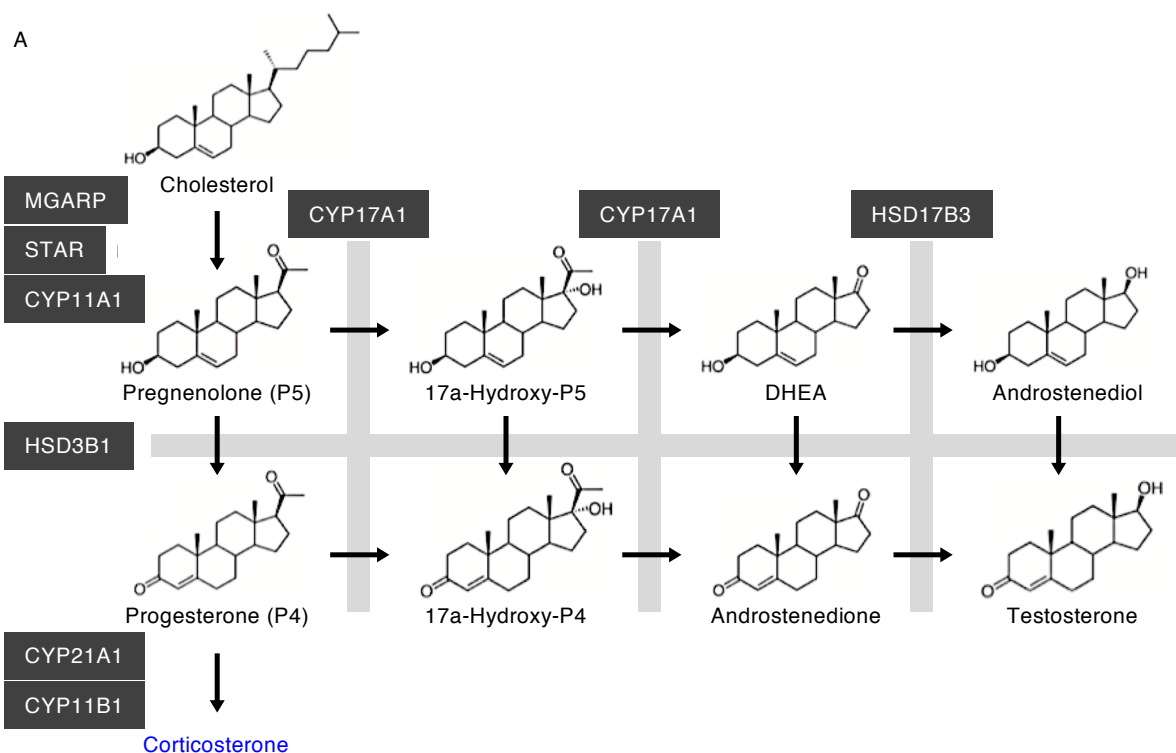


Fig. 5, Steroidogenic and metabolic gene expression activated in S-EGFP cells

Total RNAs (n = 3) were prepared from S-EGFP (dark gray bars), W-EGFP (open bars), Sertoli cells (gray bars), and whole testes (light gray bars) at E16.5. Expressions of the steroidogenic genes (*Star* (A), *Cyp11a1* (B), *Hsd3b1* (C), *Cyp17a1* (D), and *Hsd17b3* (E)), *Ad4BP/SF-1* (F), glycolytic genes (*Hk1* (G), *Aldoa* (H), and *Eno1* (I)), TCA cycle genes (*Idh3g* (J), *Suclg2* (K), and *Sdha* (L)), and OXPHOS genes (*Ndufs8* (M) and *Uqcr10* (N)) were quantified by qRT-PCR. The primers used for PCR are listed in Table 2. The data were standardized using *Actb*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test.



B Steroidogenesis

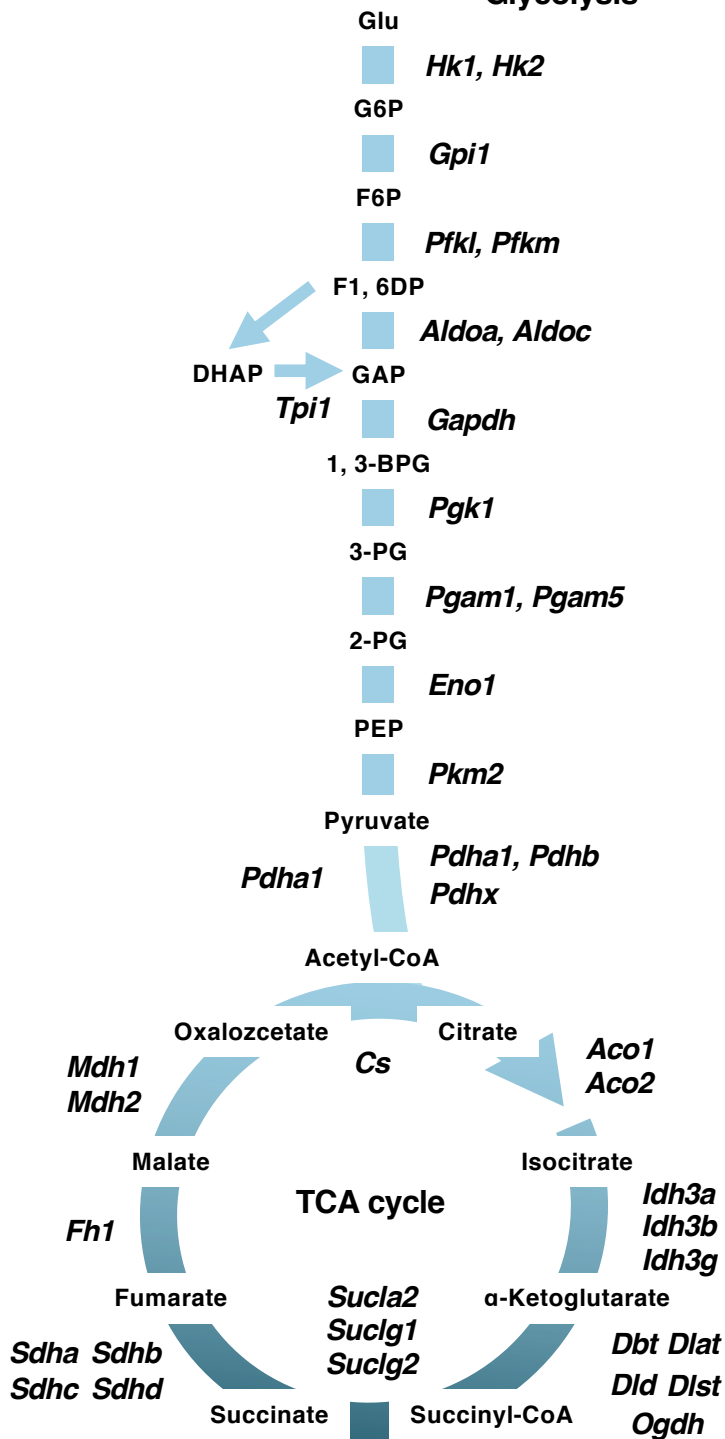
Gene	FPKM		S/W
	S-EGFP	W-EGFP	
<i>Ad4BP/SF-1</i>	850.19	49.93	16.70
<i>Mgrap</i>	134.15	2.85	35.08
<i>Star</i>	1249.96	10.18	111.89
<i>Cyp11a1</i>	9598.19	95.84	99.13
<i>Hsd3b1</i>	3506.79	40.92	83.69
<i>Cyp17a1</i>	12483.90	135.62	91.38
<i>Hsd17b3</i>	2.37	6.89	0.43
<i>Cyp11b1</i>	60.08	0.04	58.98
<i>Cyp21a1</i>	306.43	0.36	225.96

S/W 3 1 0.3

Fig. 6, Expression of genes involved in steroidogenesis in S-EGFP and W-EGFP cells

Intermediate metabolites and genes for steroidogenesis from cholesterol are shown (A). FPKM values of *Ad4BP/SF-1* and the steroidogenic genes in S-EGFP and W-EGFP cells are shown (B). The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while that less than one is highlighted with blue. Deeper colors indicate larger differences.

Glycolysis



Glycolysis

Gene	FPKM		S/W
	S-EGFP	W-EGFP	
<i>Hk1</i>	13.34	21.56	0.64
<i>Hk2</i>	10.03	13.92	0.74
<i>Gpi1</i>	171.91	50.25	3.37
<i>Pfk1</i>	67.67	21.23	3.09
<i>Pfkm</i>	37.29	27.14	1.36
<i>Aldoa</i>	1657.73	224.93	7.34
<i>Aldoc</i>	97.61	0.74	56.55
<i>Tpi1</i>	417.89	218.64	1.91
<i>Gapdh</i>	816.52	235.25	3.46
<i>Pgk1</i>	225.32	103.59	2.16
<i>Pgam1</i>	595.85	156.22	3.80
<i>Pgam5</i>	30.36	20.05	1.49
<i>Eno1</i>	322.53	64.44	4.94
<i>Pkm</i>	335.31	197.34	1.70

TCA cycle

Gene	FPKM		S/W
	S-EGFP	W-EGFP	
<i>Pdha1</i>	25.25	53.74	0.48
<i>Pdhb</i>	131.30	54.44	2.39
<i>Pdhx</i>	50.59	10.23	4.60
<i>Cs</i>	128.59	71.85	1.78
<i>Aco1</i>	67.37	31.84	2.08
<i>Aco2</i>	145.03	71.49	2.01
<i>Idh3a</i>	32.34	22.08	1.44
<i>Idh3b</i>	120.09	69.73	1.71
<i>Idh3g</i>	193.32	86.15	2.23
<i>Dbt</i>	13.21	6.94	1.79
<i>Dlat</i>	20.97	9.17	2.16
<i>Dlst</i>	88.95	47.53	1.85
<i>Dld</i>	98.12	39.36	2.46
<i>Ogdh</i>	47.87	29.00	1.63
<i>Sucla2</i>	55.89	27.08	2.03
<i>Suclg1</i>	149.64	47.98	3.08
<i>Suclg2</i>	111.72	35.89	3.06
<i>Sdha</i>	222.11	84.02	2.62
<i>Sdhb</i>	191.48	73.94	2.57
<i>Sdhc</i>	158.10	116.05	1.36
<i>Sdh</i>	135.56	64.51	2.08
<i>Fh1</i>	127.50	67.17	1.88
<i>Mdh1</i>	227.98	83.16	2.72
<i>Mdh2</i>	525.54	170.41	3.07

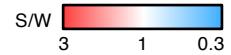


Fig. 7, Expression of genes involved in glycolysis and TCA cycle in S-EGFP and W-EGFP cells

Intermediate metabolites and genes for glycolysis (upper left) and TCA cycle (lower left) are shown. The expression levels of glycolytic and TCA cycle genes in S-EGFP and W-EGFP cells are shown with FPKM values at the right. Isozymes of which FPKM values were less than 10.0 in both cell types are excluded from the list. The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while those less than one are highlighted with blue. Deeper colors indicate larger differences.

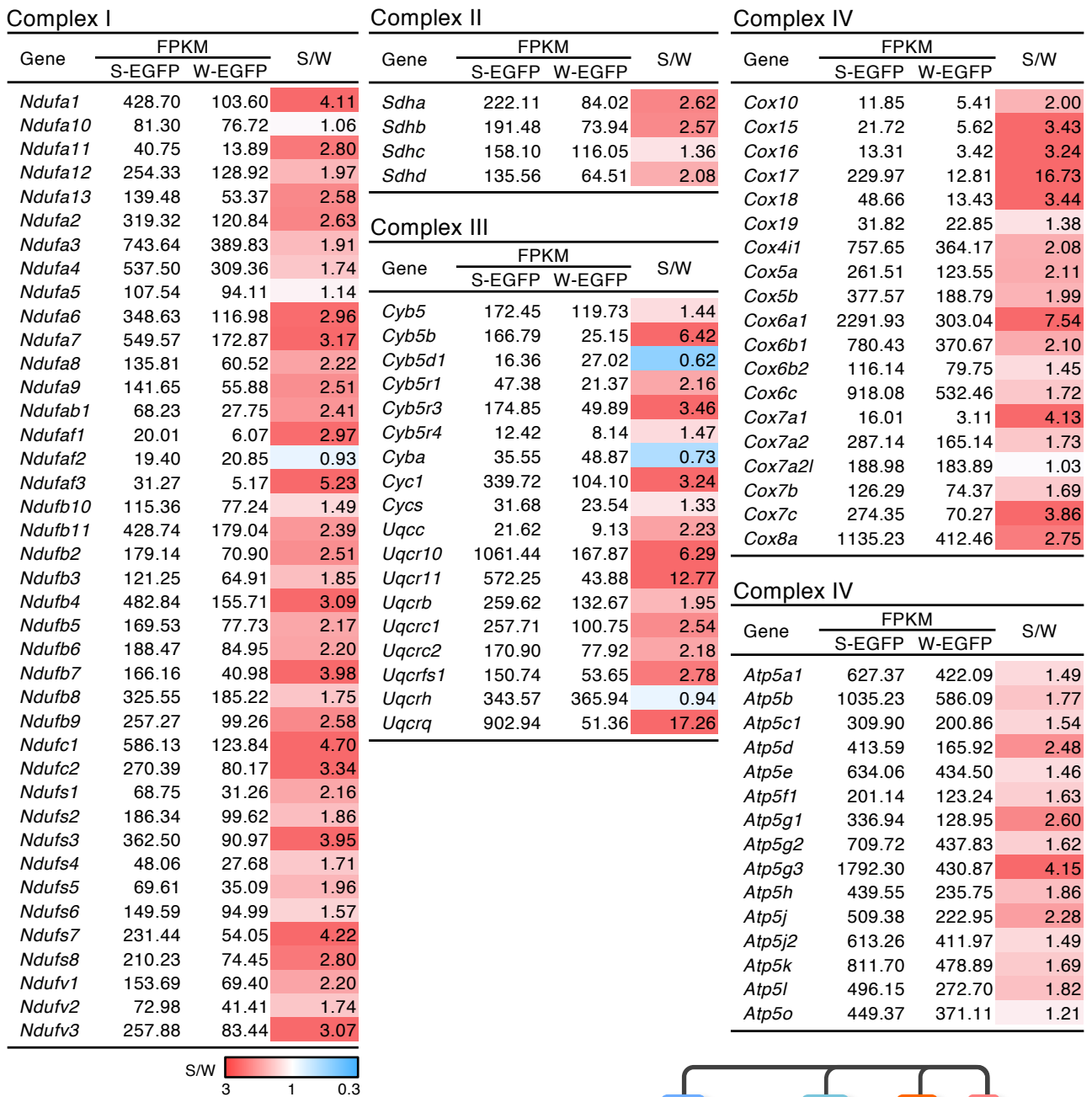


Fig. 8, Expression of genes involved in OXPHOS in S-EGFP and W-EGFP cells

FPKM values of OXPHOS genes (complex I to V schematically illustrated at the lower right) in S-EGFP and W-EGFP cells are shown. Isozymes of which FPKM values were less than 10.0 in both cell types are excluded from the list. The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while those less than one are highlighted with blue. Deeper colors indicate larger differences.

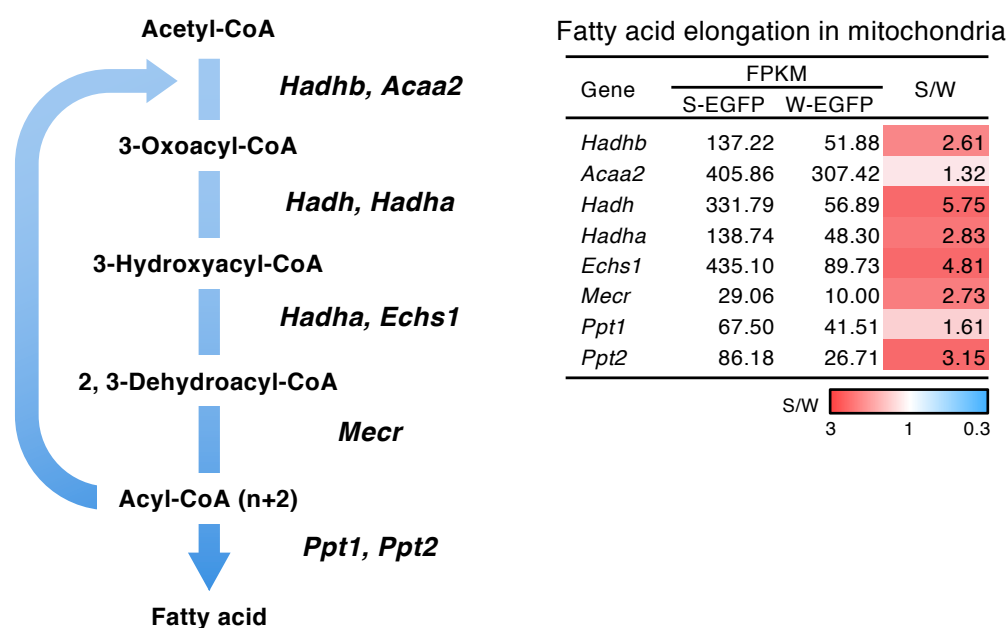


Fig. 9, Expression of genes involved in the mitochondrial pathway for fatty acid synthesis in S-EGFP and W-EGFP cells

Intermediate metabolites and the genes for mitochondrial fatty acid synthesis are shown at the left. FPKM values of genes involved in the mitochondrial pathway for fatty acid elongation in S-EGFP and W-EGFP cells are shown at the right. The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while those less than one are highlighted with blue. Deeper colors indicate larger differences.

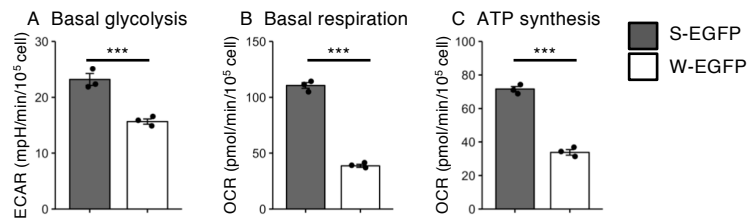


Fig. 10, Energy metabolism in S-EGFP and W-EGFP cells

Basal glycolysis (ECAR in A), basal respiration (OCR in B), and ATP synthesis (OCR in C) were determined in S-EGFP (gray bars) and W-EGFP cells (open bars). The ECAR and OCR were normalized by cell number and are presented as means (SEM). $n = 3$. $***p < 0.001$. Two-tailed Student's t -test.

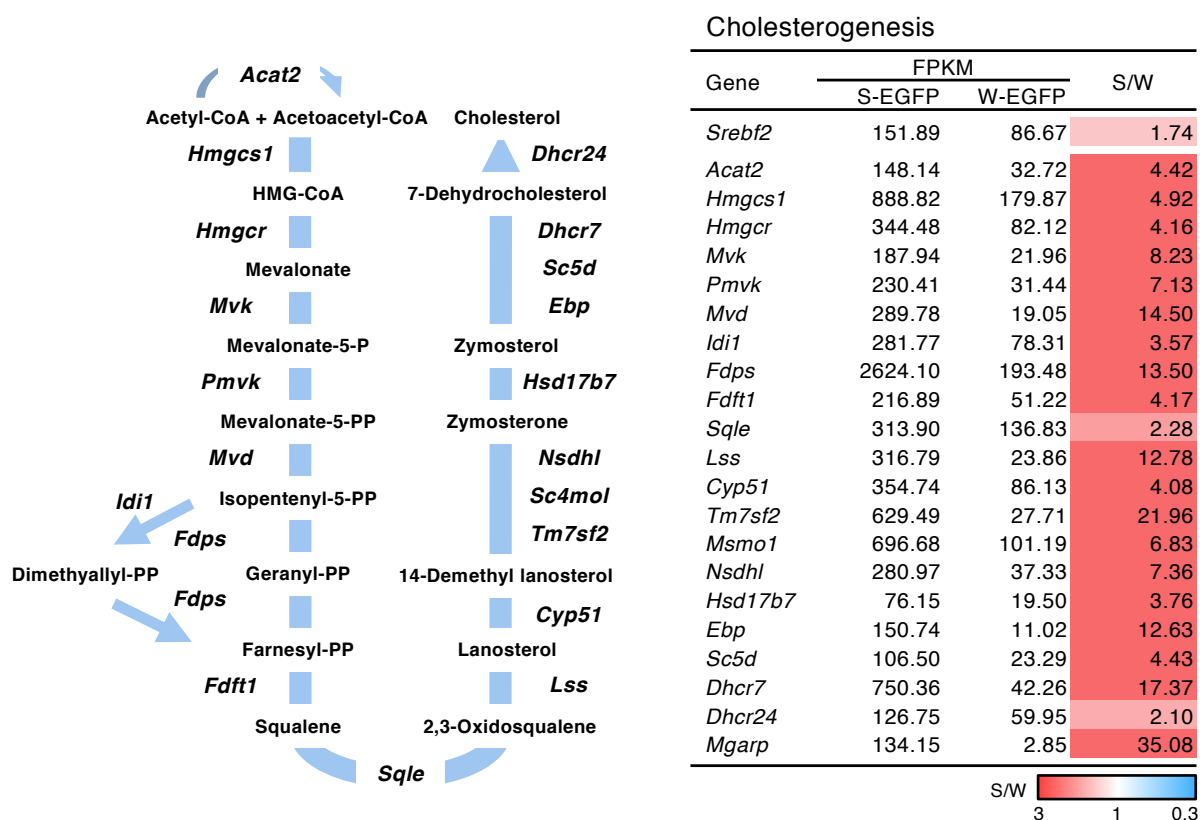


Fig. 11, Expression of genes involved in cholesterogenesis in S-EGFP and W-EGFP cells

Intermediate metabolites and genes for cholesterogenesis are shown at the left. FPKM values of *Srebf2* and cholesterogenic genes in S-EGFP and W-EGFP cells are shown at the right. The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while those less than one are highlighted with blue. Deeper colors indicate larger differences.

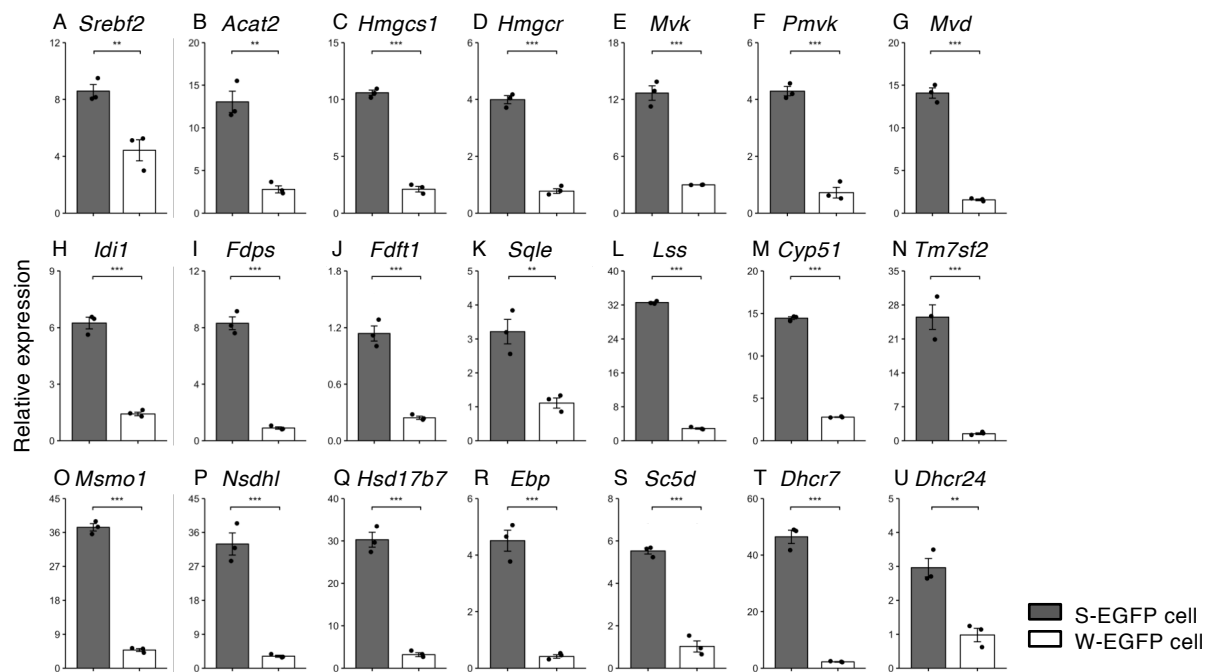


Fig. 12, Quantitative analyses for cholesterologenic gene expression in S-EGFP and W-EGFP cells

Total RNAs (n = 3) were prepared from S-EGFP (dark gray bars) and W-EGFP (open bars) of E16.5 testes. Expressions of *Srebf2* (A) and cholesterologenic genes (*Acat2* (B), *Hmgcs1* (C), *Hmgcr* (D), *Mvk* (E), *Pmvk* (F), *Mvd* (G), *Idi1* (H), *Fdps* (I), *Fdft1* (J), *Sqle* (K), *Lss* (L), *Cyp51* (M), *Tm7sf2* (N), *Msmo1* (O), *Nsdhl* (P), *Hsd17b7* (Q), *Ebp* (R), *Sc5d* (S), *Dhcr7* (T), *Dhcr24* (U)) were determined quantitatively by qRT-PCR. The primers used for PCR are listed in Table 2. The data were standardized using *Actb*, and are presented as means (SEM). ** $p < 0.01$. *** $p < 0.001$. Two-tailed Student's *t*-test.

PPP and NADPH production

Gene	FPKM		S/W
	S-EGFP	W-EGFP	
<i>G6pdx</i>	67.85	19.45	3.37
<i>Pgd</i>	210.60	33.29	6.17
<i>Me1</i>	738.95	91.43	8.01
<i>Me2</i>	7.01	19.99	0.38
<i>ldh1</i>	419.51	56.76	7.28
<i>ldh2</i>	302.38	223.01	1.35
<i>Mthfd1</i>	17.05	18.47	0.93
<i>Mthfd2</i>	68.48	16.55	3.96
<i>Tkt</i>	182.45	70.62	2.56
<i>Taldo1</i>	188.83	72.39	2.59
<i>Rpia</i>	41.19	14.73	2.68

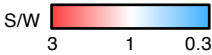


Fig. 13, Expression of genes involved in PPP and NADPH production in S-EGFP and W-EGFP cells

FPKM values of genes involved in PPP and NADPH production in S-EGFP and W-EGFP cells are shown. The values in S/W columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with red, while those less than one are highlighted with blue. Deeper colors indicate larger differences.

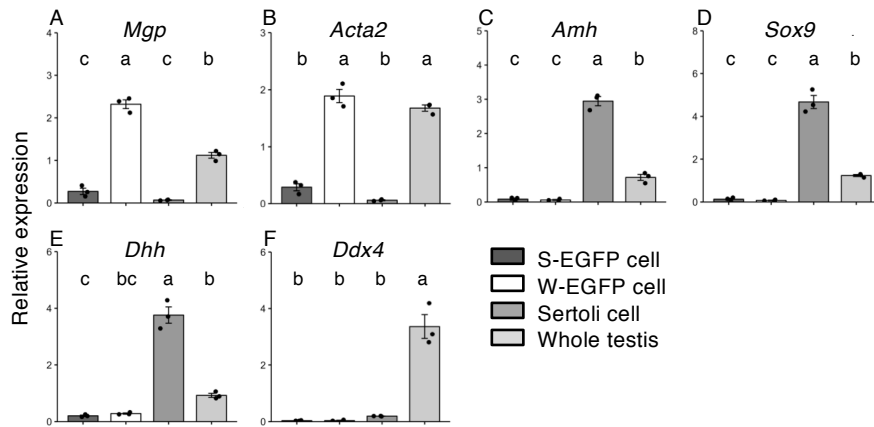


Fig. 14, Quantitative analyses of genes activated in W-EGFP, Sertoli, and germ cells
 Total RNAs (n = 3) were prepared from S-EGFP (dark gray bars), W-EGFP (open bars), Sertoli cells (gray bars), and whole testes (light gray bars) at E16.5 testes. Expressions of *Mgp* (A), *Acta2* (B), Sertoli cell marker genes (*Amh* (C), *Sox9* (D)), and germ cell marker gene (*Dhh* (E)), and germ cell marker gene (*Ddx4* (F)) were quantified using qRT-PCR. The primers used for the PCR are listed in Table 2. The data were standardized using *Actb*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

A

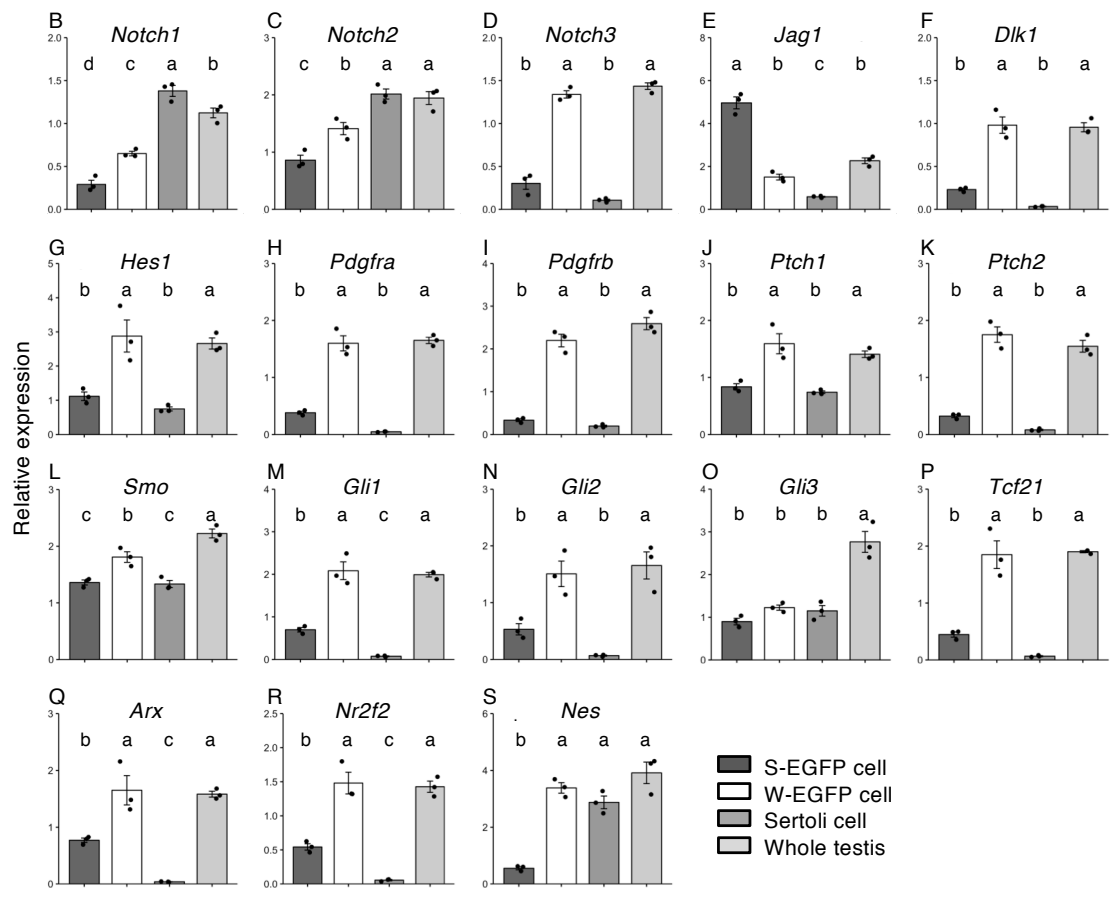
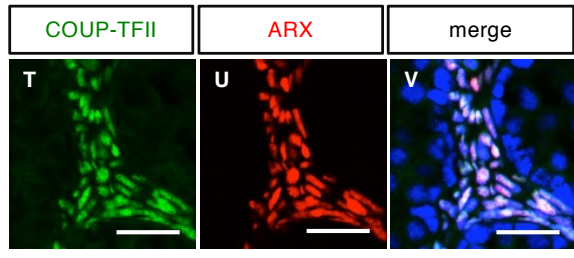
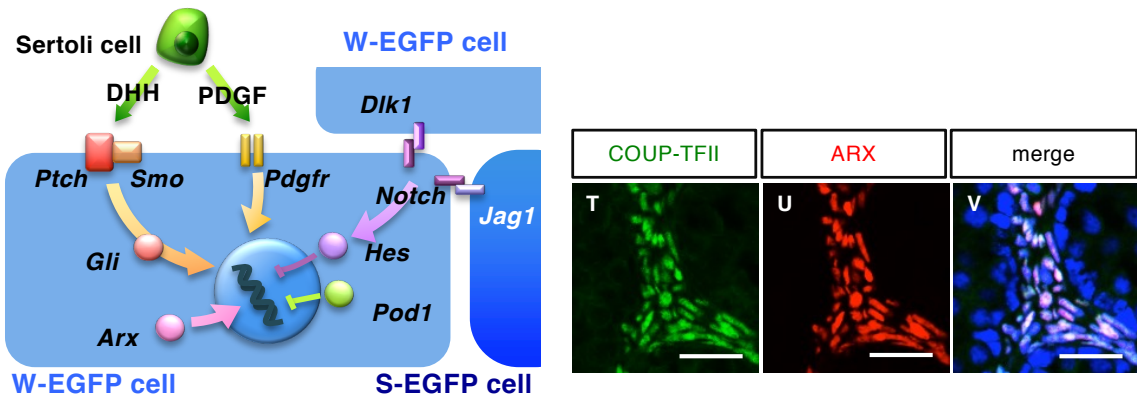


Fig. 15, Expression of genes potentially implicated in the FLC differentiation

Functional relationships of genes potentially implicated in the differentiation of FLCs is illustrated (A). Total RNAs (n = 3) were prepared from S-EGFP (dark gray bars), W-EGFP (open bars), Sertoli cells (gray bars), and whole testes (light gray bars) at E16.5. The expression of genes related to Notch signaling (*Notch1* (B), *Notch2* (C), *Notch3* (D), *Jag1* (E), *Dlk1* (F), and *Hes1* (G)), receptors of PDGF signaling (*Pdgfra* (H) and *Pdgfrb* (I)), and those related to hedgehog signaling (*Ptch1* (J), *Ptch2* (K), *Smo* (L), *Gli1* (M), *Gli2* (N), and *Gli3* (O)) was examined by qRT-PCR. The expression of transcription factors, *Tcf21* (P), *Arx* (Q), and *Nr2f2* (R), potentially implicated in FLC differentiation, and *Nes* (S), a marker for stem/progenitor cells of Leydig cells, were also examined. The primers used for qPCR are listed in Table 2. The data were standardized using *Actb*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test. Distribution of ARX and COUP-TFII in the fetal testes at E16.5 was determined using immunofluorescence. COUP-TFII (T), ARX (U), and their overlaid images (V) are shown. DAPI was used for nuclear staining. Scale bar = 50 μm .

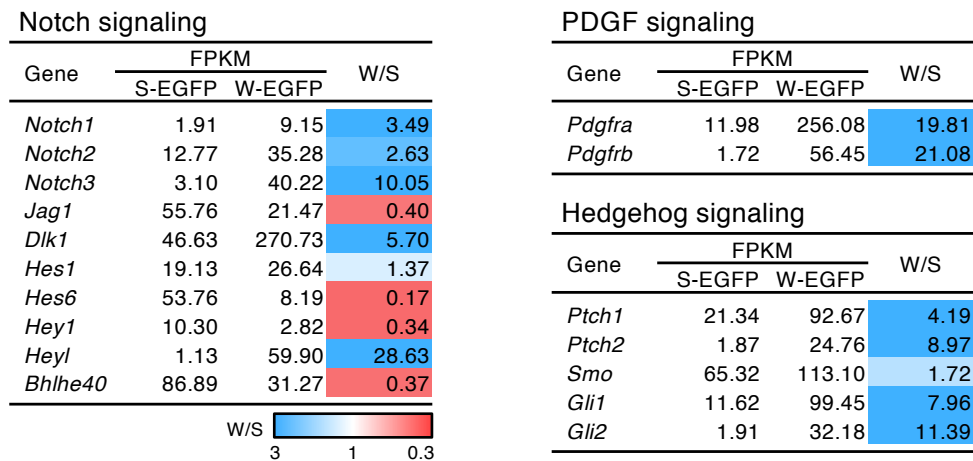


Fig. 16, Expression of genes involved in growth factor signaling

FPKM values of genes related to Notch, PDGF, and hedgehog signaling in S-EGFP and W-EGFP cells are shown. Isozymes of which FPKM values were less than 10.0 in both cell types are excluded from the list. The values in W/S columns are the ratios of the FPKM values of the genes indicated at the left between S-EGFP and W-EGFP cells. Ratios larger than one are highlighted with blue, while those less than one are highlighted with red. Deeper colors indicate larger differences.

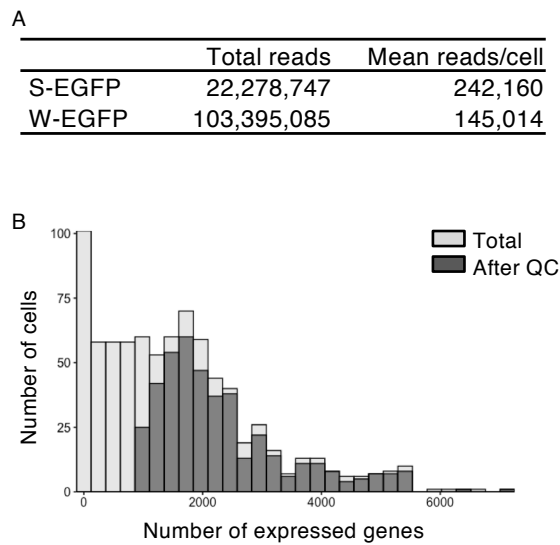


Fig. 17, Quality check of scRNA-seq datasets

Datasets of scRNA-seq were obtained from S-EGFP and W-EGFP cells at E16.5. Mean reads obtained from a single-cell of S-EGFP and W-EGFP cells were calculated (A). Cellular distribution is shown as a function of the number of genes expressed (B). Cell number is set on the vertical axis, while number of expressed genes is set on the horizontal axis. The light gray bars represent the distribution of whole cells used for single-cell sequence (Total), and the dark gray bars represent the distribution of the cells whose sequence datasets were evaluated to maintain a certain level of quality (After QC). The sequence quality was assessed as described in 'Materials and Methods' by a quality control analysis.

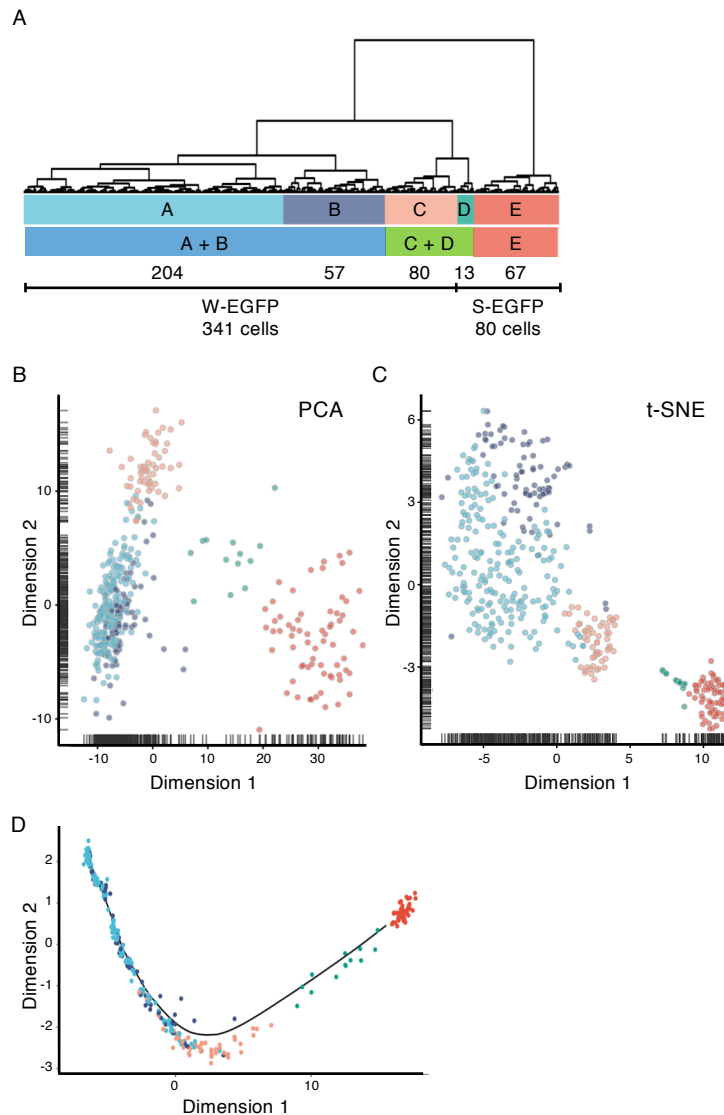


Fig. 18, Characterization of S-EGFP and W-EGFP cells using scRNA-seq datasets

80 datasets of S-EGFP cells and 341 datasets of W-EGFP cells were subjected to hierarchical clustering. Eventually, these cells were divided into five cell clusters: Cluster A containing 204 W-EGFP cells (light blue), cluster B containing 80 W-EGFP cells (blue), cluster C containing 57 W-EGFP cells (orange), cluster D containing 13 S-EGFP cells (green), and cluster E containing 67 S-EGFP cells (red) (A). Under milder grouping condition, these cells could be divided into three clusters, cluster A+B, cluster C+D, and cluster E. S-EGFP and W-EGFP cells were subjected to PCA (B), t-SNE (C), or Monocle pseudo-time trajectory analysis (D). Cellular distribution revealed by these analyses is shown by dots with colors labeling the five clusters in (A).

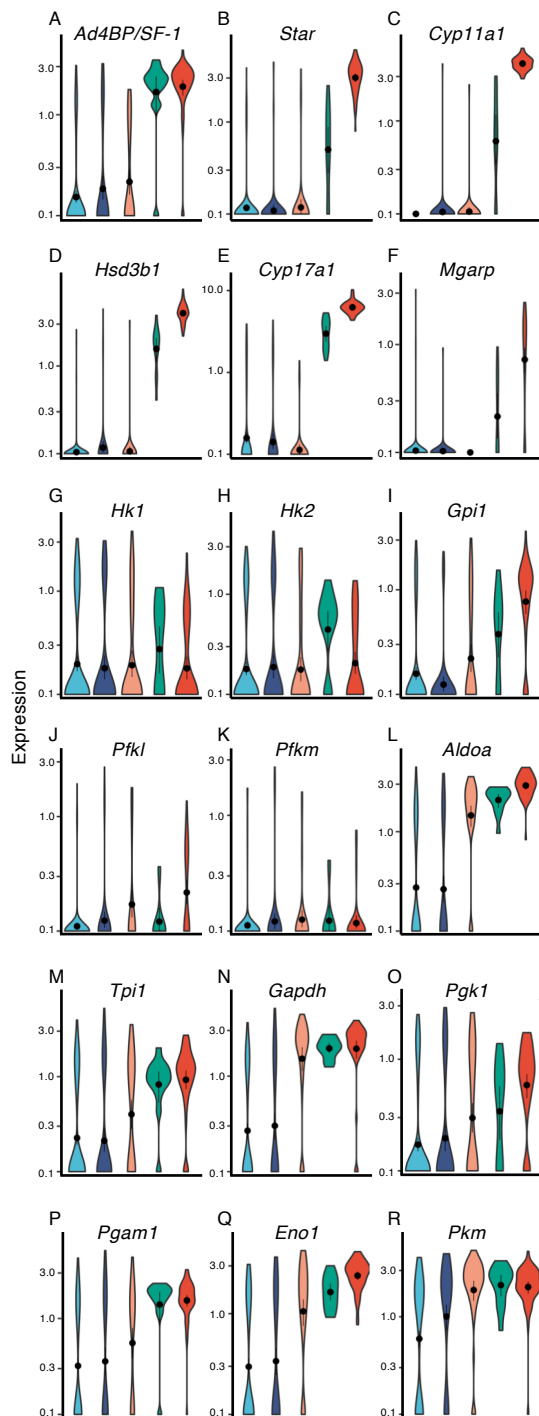


Fig. 19, Differential expression of steroidogenic and glycolytic genes in the five cell clusters

Expressions of *Ad4BP/SF-1* (A), steroidogenic genes (*Star* (B), *Cyp11a1* (C), *Hsd3b1* (D), and *Cyp17a1* (E)), and *Mgarp* (F) in cluster A (light blue), cluster B (blue), cluster C (orange), cluster D (green), and cluster E (red) are indicated by violin plot. Expressions of glycolytic genes (*Hk1* (G), *Hk2* (H), *Gpi1* (I), *Pfk1* (J), *Pfk1* (K), *Aldoa* (L), *Tpi1* (M), *Gapdh* (N), *Pgc1* (O), *Pgam1* (P), *Eno1* (Q), and *Pkm* (R)) in the five cell clusters are indicated by violin plot.

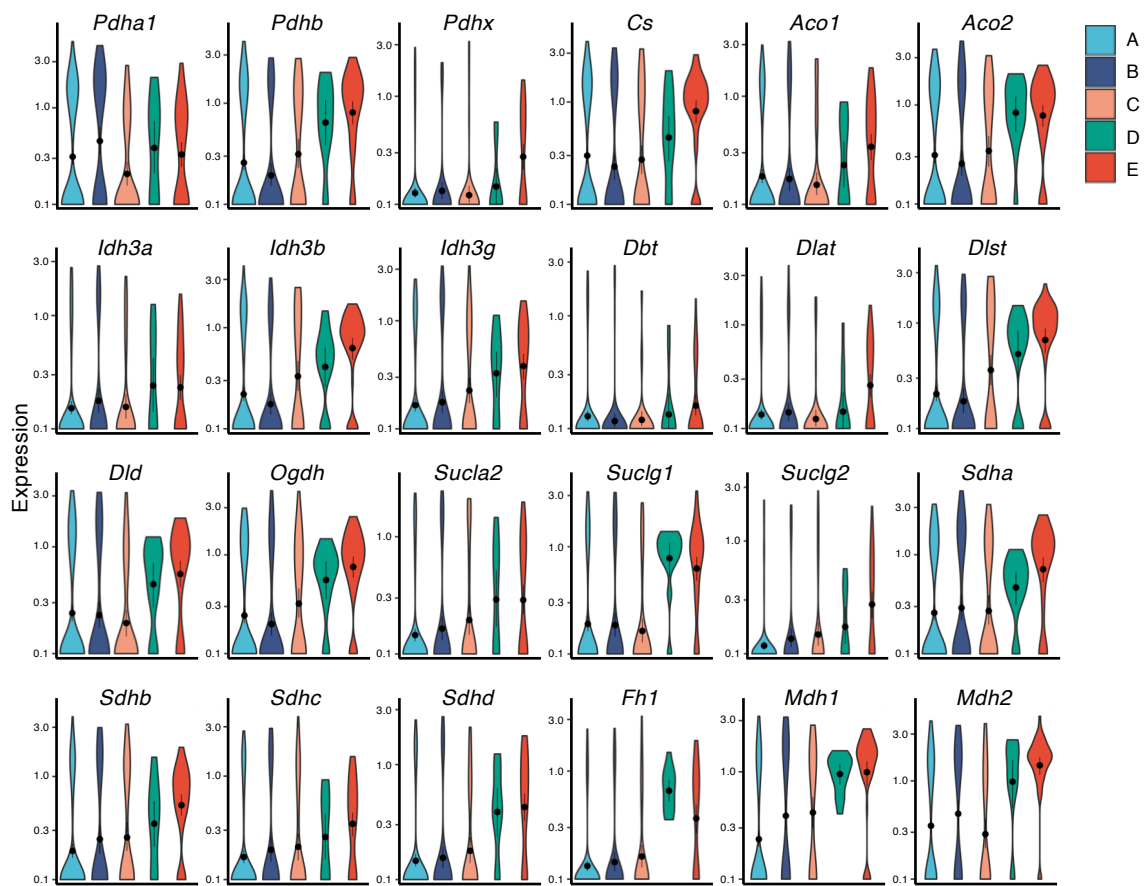


Fig. 20, Differential expression of TCA cycle genes in the five cell clusters

Expressions of TCA cycle genes in clusters A to E are shown by violin plot. The isozymes which did not detect by scRNA-seq are excluded from the list.

A Complex I

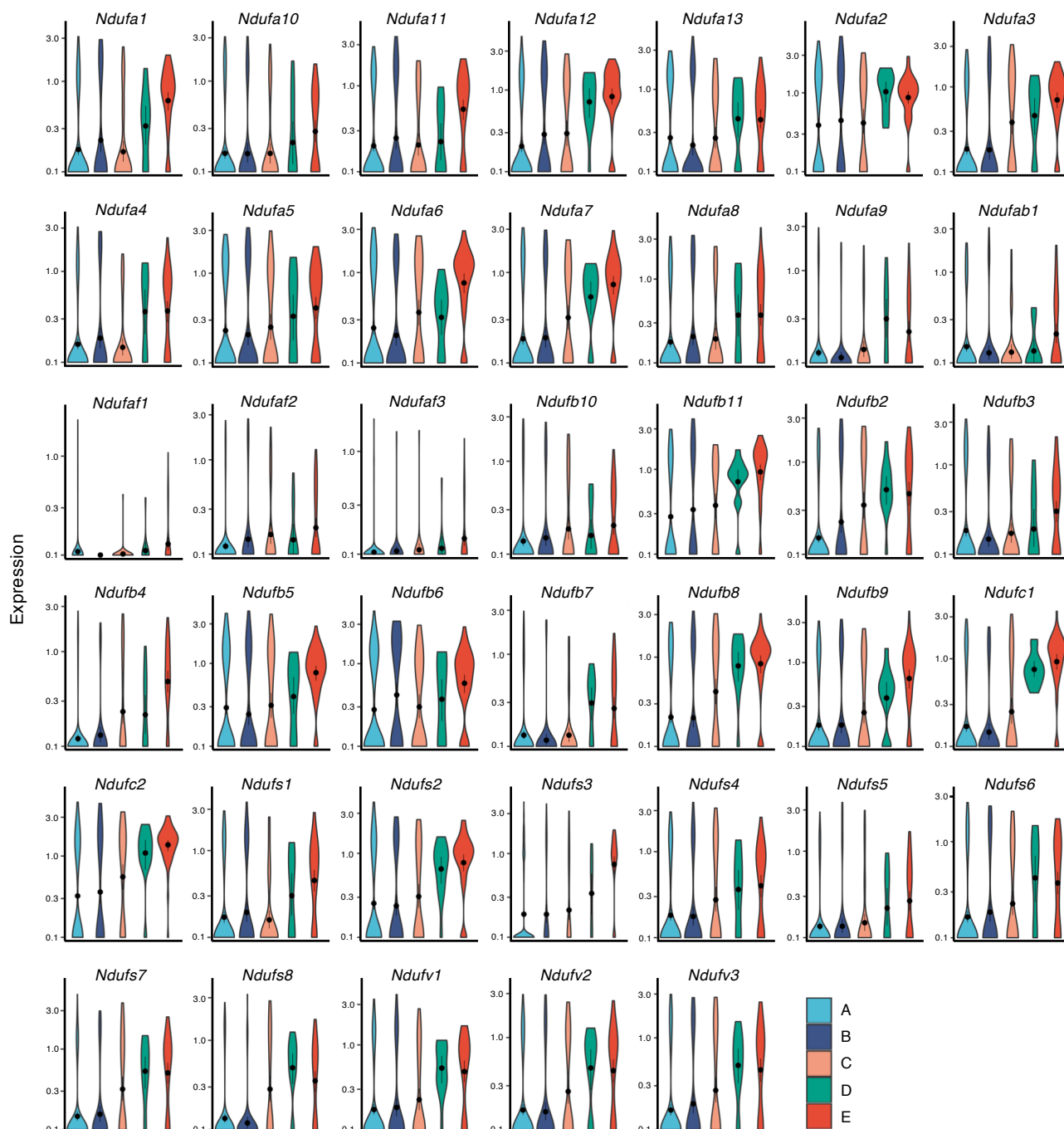
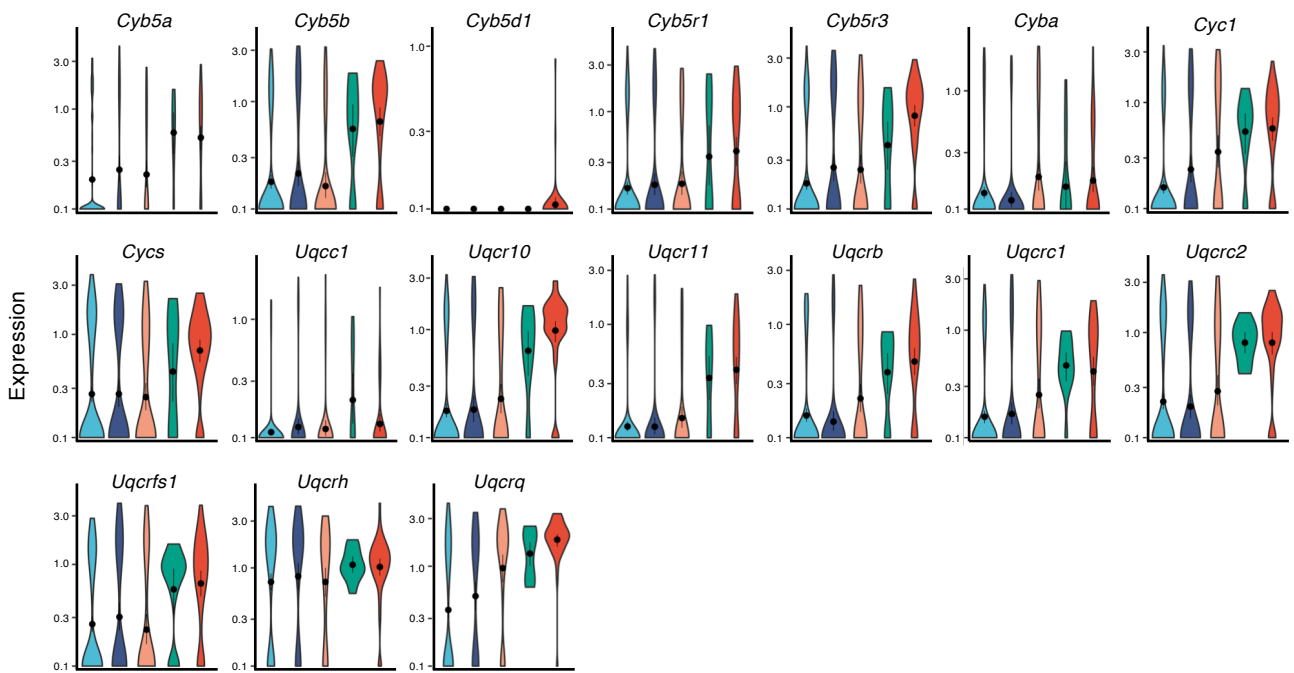


Fig. 21, Differential expression of OXPHOS genes in the five cell clusters

Expressions of OXPHOS genes (complex I (A), III (B), IV (C), and V (D)) in cluster A to E are shown by violin plot. The isozyms which did not detect by scRNA-seq are excluded from the list.

B Complex III



C Complex IV

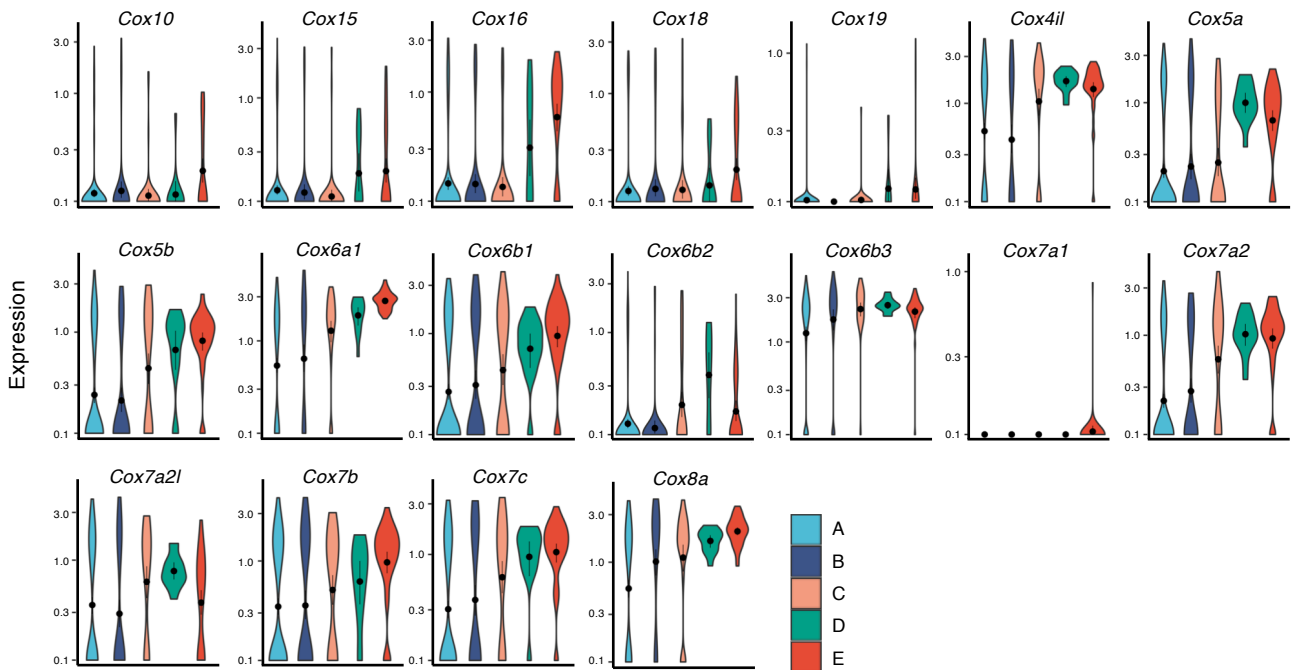


Fig. 21, Differential expression of OXPHOS genes in the five cell clusters (Continued)

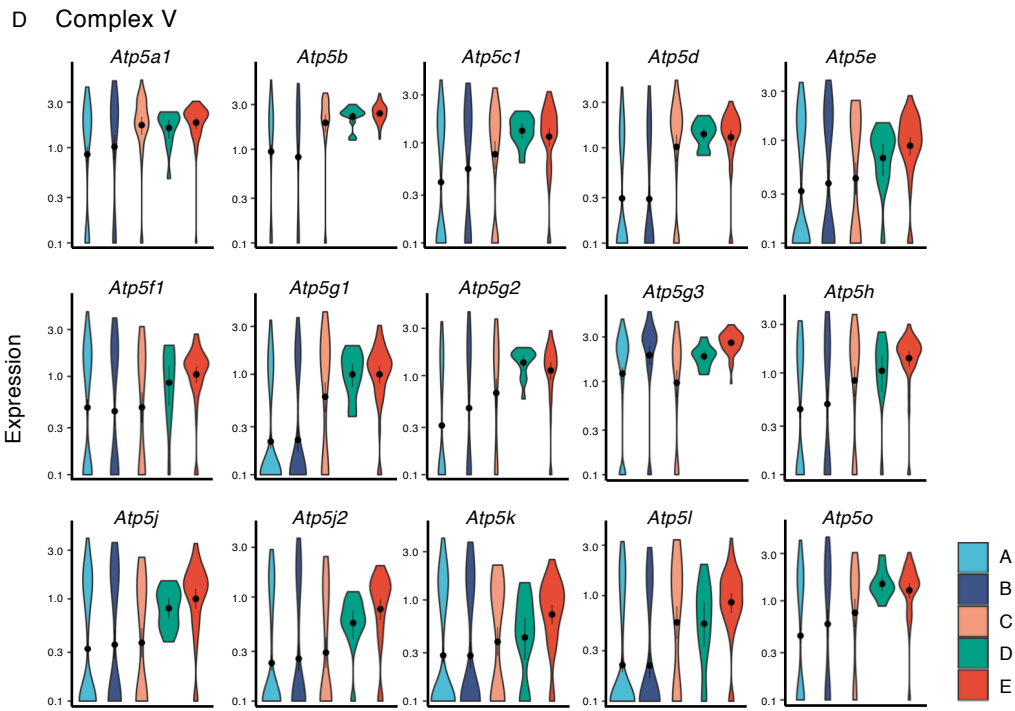


Fig. 21, Differential expression of OXPHOS genes in the five cell clusters (Continued)

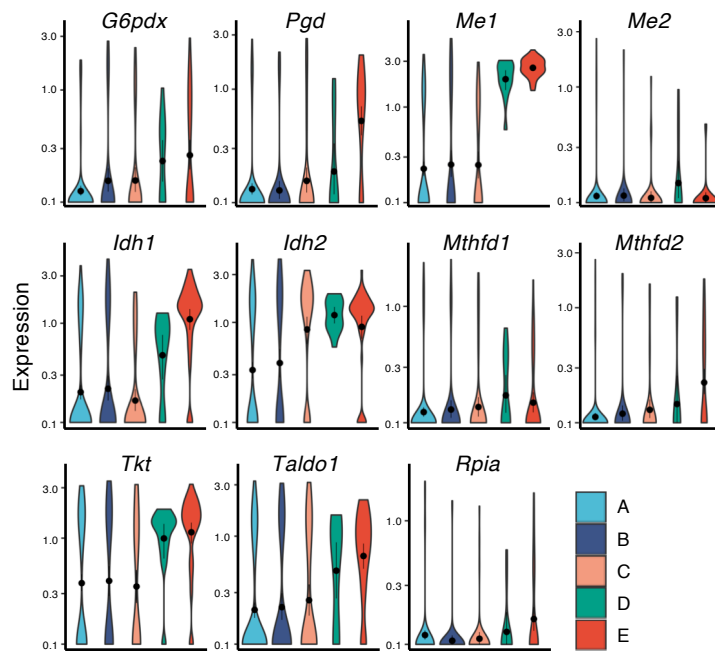


Fig. 22, Differential expression of genes involved in PPP and NADPH production in the five cell clusters

Expressions of genes involved in PPP and NADPH production in cluster A to E are shown by violin plot. The isozymes which did not detect by scRNA-seq are excluded from the list.

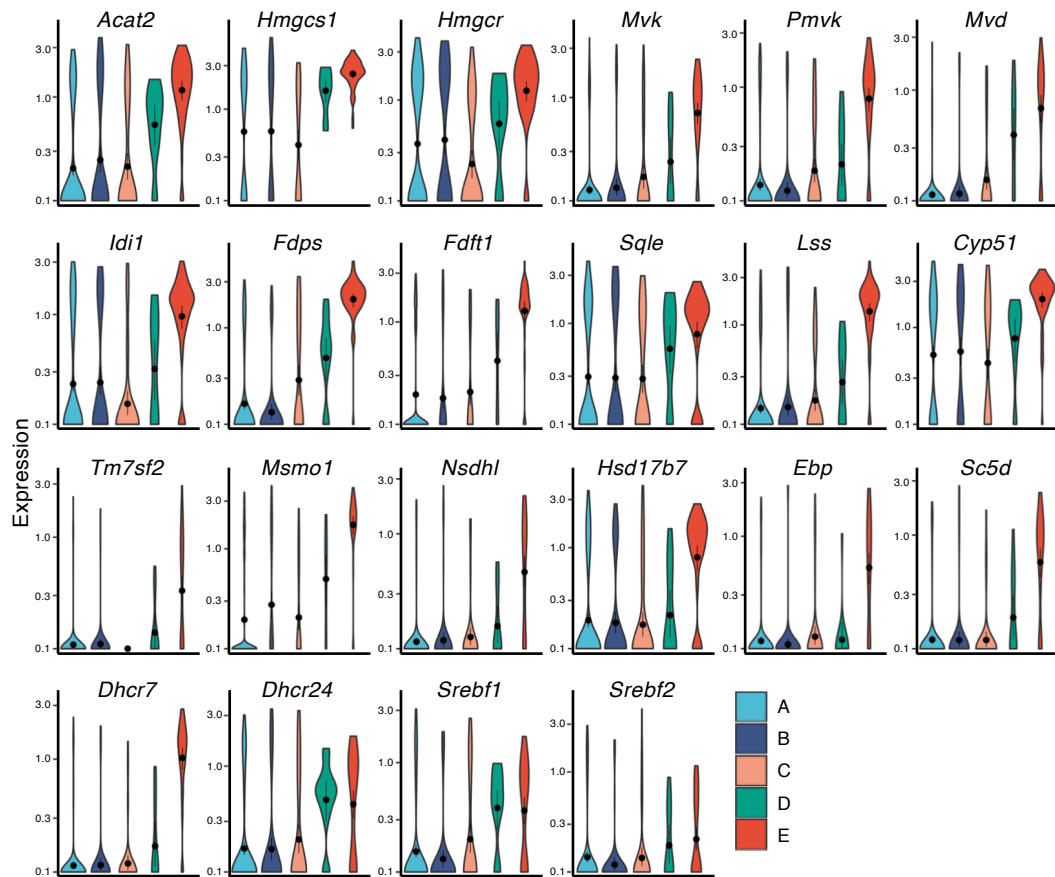


Fig. 23, Differential expression of genes involved in cholesterologenesis in the five cell clusters

Expressions of cholesterologenic genes in cluster A to E are shown by violin plot. *Srebf1* and *Srebf2* whose products are implicated in regulation of the genes for lipid metabolism and cholesterologenesis, respectively, are included. The isozymes which did not detect by scRNA-seq are excluded from the list.

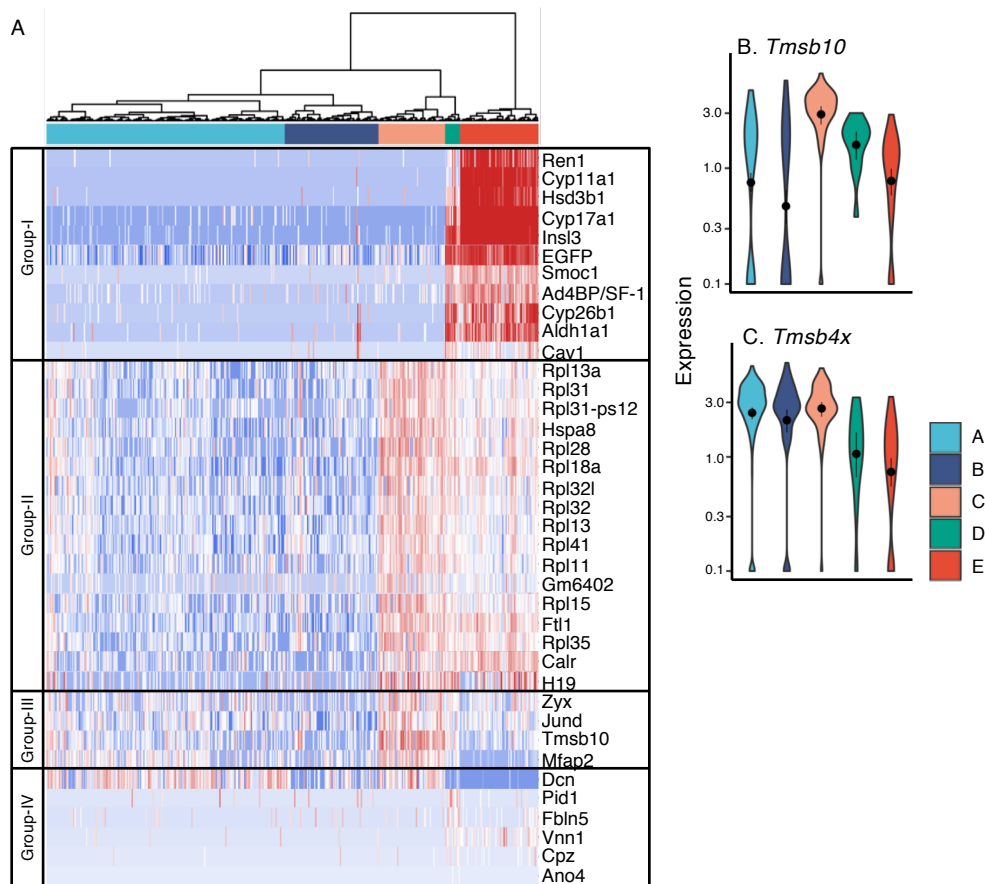


Fig. 24, Activation of *Tmsb10* in cluster C

Heatmap of gene expression was generated with paying attention to the genes whose expression is altered in cluster C (A). Representative differentially expressed genes (DEGs) shown at the right were divided into 4 groups. Genes in group-I show higher expression in cluster D and E. Genes in group-II show higher expression in Cluster C, D, and E. Genes in group-III show higher expression in cluster C. Expression of *Tmsb10* (B) and *Tmsb4x* (C) are shown by violin plots of cluster A (light blue), cluster B (blue), cluster C (orange), cluster D (green), and cluster E (red).

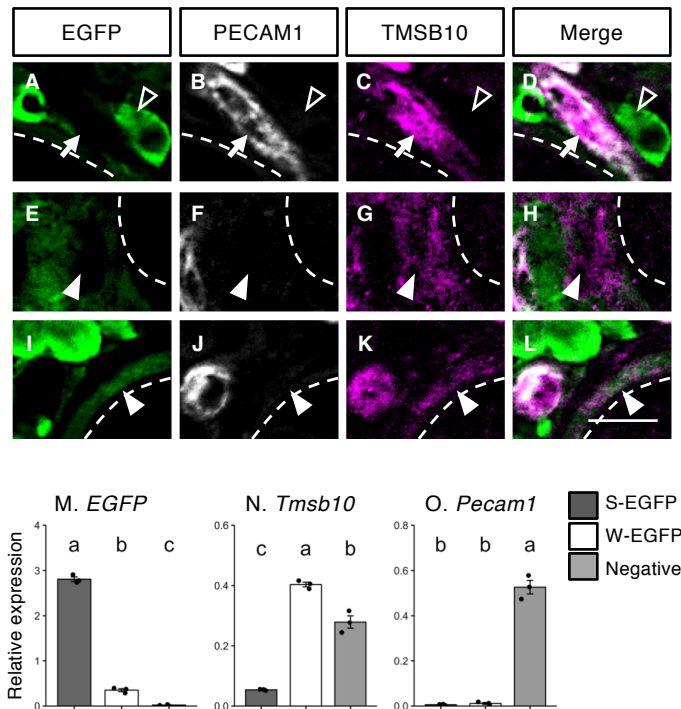


Fig. 25, TMSB10-positive cells in the interstitial space

Distribution of EGFP (green), PECAM1 (white), and TMSB10 (magenta) in the fetal testes at E16.5 was examined by immunofluorescence. EGFP (A, E, I), PECAM1 (B, F, J), TMSB10 (C, G, K), and the merge image (D, H, L) are shown. Arrow indicates TMSB10 and PECAM1-positive cells (A–D). Closed and open arrowhead indicates weakly TMSB10-positive cells (E–L) and EGFP-positive FLCs (A–D), respectively. Broken lines indicate testicular cords (A–L). Scale bar = 25 μ m. Total RNAs (n = 3) were prepared from S-EGFP (dark gray bars), W-EGFP (open bars), and EGFP-negative cells (light gray bars) of E16.5 testes. Expression of *EGFP* (M), *Tmsb10* (N), and *Pecam1* (O) was quantified by qRT-PCR. The primers used for PCR are listed in Table 2. The data were standardized using *Rn18s*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

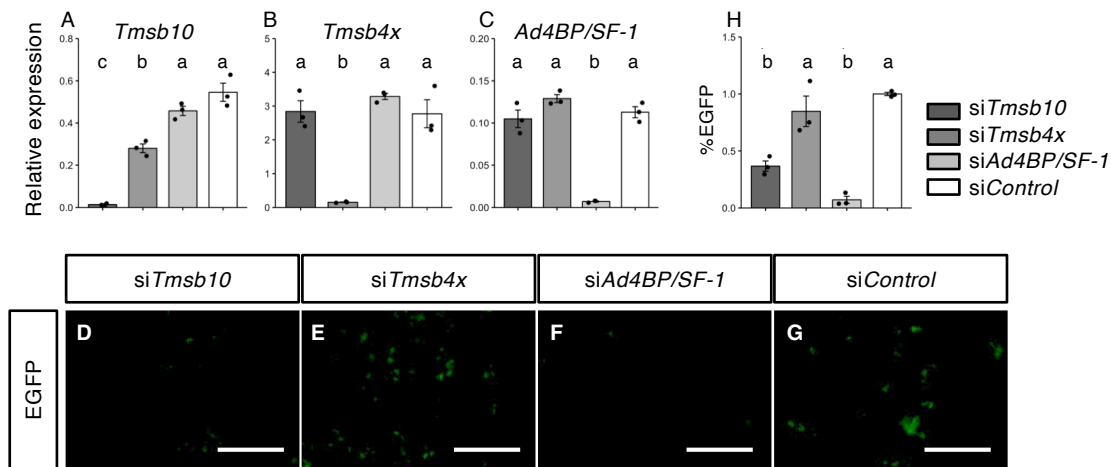


Fig. 26, Suppression of FLC differentiation by *Tmsb10* KD

W-EGFP cells prepared from E16.5 testes were cultured on plates and then treated with siRNA for *Tmsb10* (si*Tmsb10*, dark gray bars), *Tmsb4x* (si*Tmsb4x*, gray bars), *Ad4BP/SF-1* (si*Ad4BP/SF-1*, light gray bars), and control siRNA (si*Control*, open bars). Total RNAs (n = 3) were prepared from the cells and subjected to qRT-PCR for *Tmsb10* (A), *Tmsb4x* (B), and *Ad4BP/SF-1* (C). The primers used for PCR are listed in Table 2. The data were standardized using *Rn18s* (ribosomal 18S RNA) and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test. Whole cells prepared from the wild type testes at E16.5 were mixed with E16.5 W-EGFP cells pretreated for 24 hrs with siRNA for *Tmsb10* (si*Tmsb10*) (D), *Tmsb4x* (si*Tmsb4x*) (E), *Ad4BP/SF-1* (si*Ad4BP/SF-1*) (F), and control siRNA (si*Control*) (G), and thereafter cultured for 35 days. These reconstructed testes were observed under a fluorescence microscope. Scale bar = 250 μm. The EGFP positive cells in the reconstructed testes (n = 3 for each siRNA treatment) were counted. The ratios of EGFP positive cells per W-EGFP cells used for the reconstruction are shown (H). Average values for the si*Control*-treated cells were normalized to 1.0. The data are presented as means (SEM). Difference among the reconstructed tissues was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

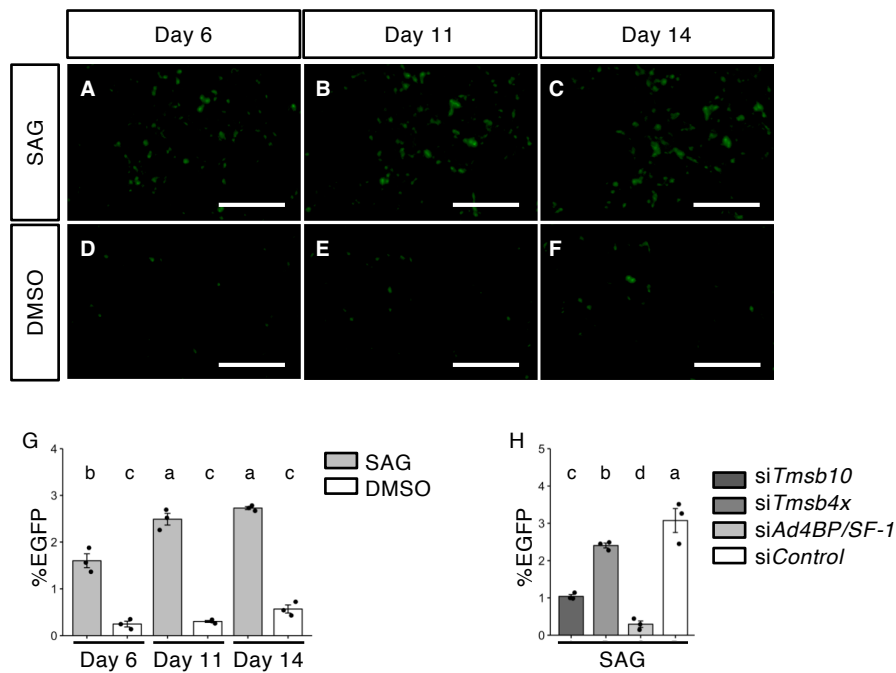


Fig. 27, Activation of FLC differentiation by SAG

Reconstructed testes with E16.5 whole testis and W-EGFP cells were cultured in the presence (A, B, and C) or absence (D, E, and F) of SAG for 6 (A and D), 11 (B and E), and 14 days (C and F). These reconstructed testes were observed under a fluorescence microscope. Scale bar = 250 μ m. The EGFP-positive cells were counted at day 6, 11, and 14 of incubation in the presence (gray bars) or absence (open bars) of SAG (n = 3 for each term of incubation). The ratios of the EGFP positive cells per W-EGFP cells used for the reconstruction are shown (G). Whole fetal testicular cells (E16.5) were mixed with E16.5 W-EGFP cells pretreated for 24 hrs with siRNA for *Tmsb10* (si*Tmsb10*), *Tmsb4x* (si*Tmsb4x*), *Ad4BP/SF-1* (si*Ad4BP/SF-1*), and control siRNA (si*Control*). These reconstructed testes were cultured in the presence of SAG for 21 days and were observed under a fluorescence microscope. The EGFP-positive cells in the reconstructed testes (n = 3 for each siRNA treatment) were counted, and the ratios of the EGFP positive cells per W-EGFP cells used for the reconstruction are shown (H). The data are presented as means (SEM). Difference among the reconstructed tissues was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

Fig. 28, FLE-driven transcription activated by GLI1/GLI2

Control reporter gene (pGL3-basic, open bars), Ad4BPpro-luc (Ad4BPpro, light gray bars), and FLE-Ad4BPpro-luc (FLE, dark gray bars) were transfected into NIH3T3 cells together with expression vectors for Gli1 (A) and Gli2 (B) (0, 20, 100, and 250 ng). Average relative luciferase activity (RLU) of control reporter gene with 0 ng of these expression plasmids was normalized to 1. Luciferase reporter gene constructs (FLE-Ad4BPpro-luc, FLE1-Ad4BPpro-luc, FLE2-Ad4BPpro-luc, and FLE3-Ad4BPpro-luc) are shown (C). FLE-Ad4BPpro-luc (FLE, dark gray bars), FLE1-Ad4BPpro-luc (FLE1, gray bars), FLE2-Ad4BPpro-luc (FLE2, light gray bars), and FLE3-Ad4BPpro-luc (FLE3, open bars) were transfected into NIH3T3 cells together with expression vectors for Gli2 (0, 20, and 100 ng) (D). RLU of FLE-Ad4BPpro-luc with 0 ng of these expression plasmids was normalized to 1. The data are presented as means (SEM). Alignment of FLE1 nucleotide sequences is shown (E). Nucleotide sequences were compared among mouse, rat, human, cow, and dog by alignment using MAFFT. Asterisks indicate nucleotides conserved in all animal species. Gray, filled boxes indicate short blocks of highly conserved regions that consist of more than five consecutively conserved nucleotides.

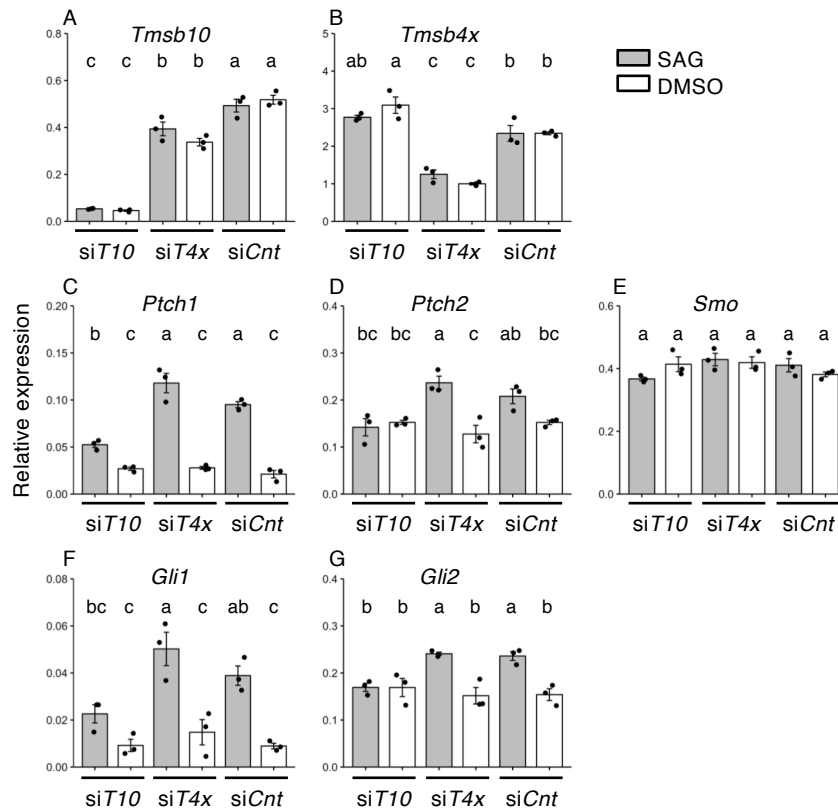


Fig. 29, Expression of genes required for hedgehog signaling affected by *Tmsb10* KD in W-EGFP cells

Total RNAs (n = 3) were prepared from E16.5 W-EGFP cells treated with siRNA for *Tmsb10* (siT10), *Tmsb4x* (siT4x), and control (siCnt). After siRNA treatment for 24 hrs, the cells were cultured in the presence (gray bars) or absence (open bars) of SAG for 24 hrs. The expression of *Tmsb10* (A), *Tmsb4x* (B), *Ptch1*(C), *Ptch2* (D), *Smo* (E), *Gli1* (F), and *Gli2* (G) was examined by qRT-PCR. The primers used for PCR are listed in Table 2. The data are standardized using *Rn18s*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

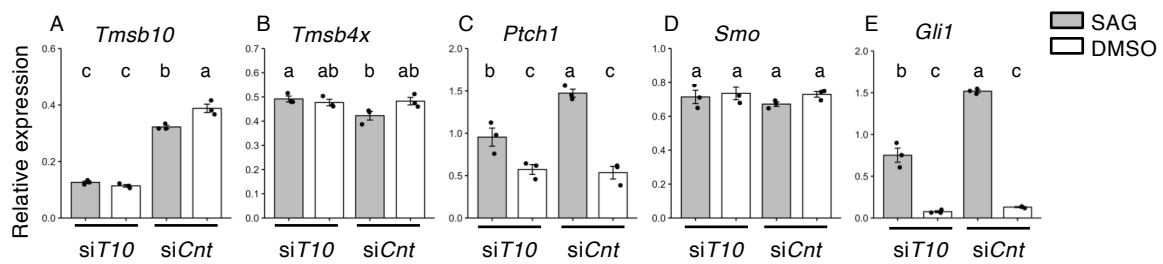


Fig. 30, Expression of genes required for hedgehog signaling affected by *Tmsb10* KD in NIH3T3 cells

Total RNAs (n = 3) were prepared from NIH3T3 cells treated with siRNA for *Tmsb10* (siT10) and control (siCnt). After siRNA treatment for 24 hrs, the cells were cultured in the presence (gray bars) or absence (open bars) of SAG for 24 hrs. Expression of *Tmsb10* (A), *Tmsb4x* (B), *Ptch1*(C), *Smo* (D), and *Gli1*(E) was quantified by qRT-PCR. The data are standardized using *Rn18s*, and are presented as means (SEM). Difference in the gene expression among the cells was assessed by one-way ANOVA followed by Tukey's *post hoc* test.

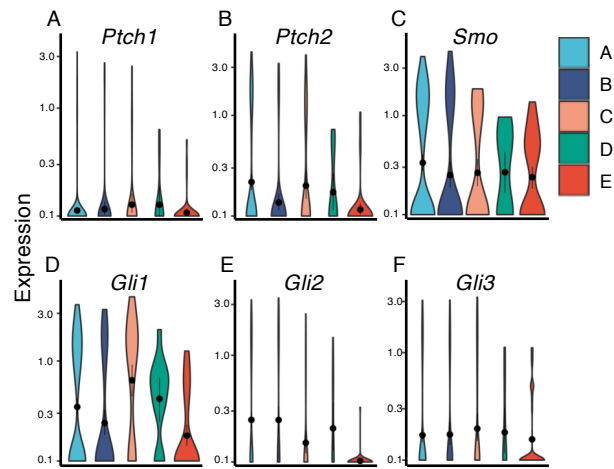


Fig. 31, Transient activation of *Gli1* in cluster C

Expressions of genes related to hedgehog signaling (*Ptch1* (A), *Ptch2* (B), *Smo* (C), *Gli1* (D), *Gli2* (E), and *Gli3* (F)) in cluster A (light blue), cluster B (blue), cluster C (orange), cluster D (green), and cluster E (red) were indicated by violin plots.

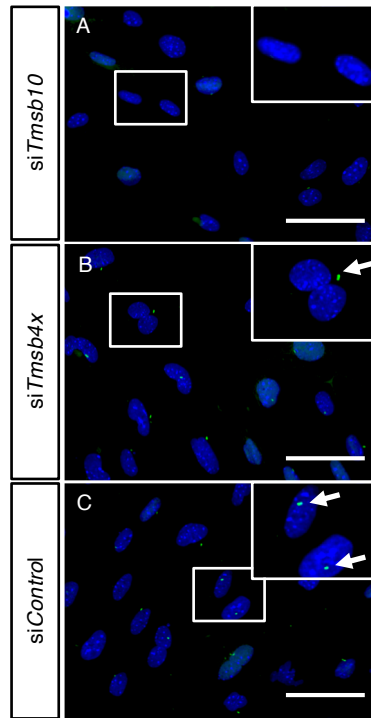


Fig. 32, Primary cilia formation affected by *Tmsb10* KD in W-EGFP cells

Distribution of ARL13B (green) in the E16.5 W-EGFP cells treated with siRNA for *Tmsb10* (si*Tmsb10*, A), *Tmsb4x* (si*Tmsb4x*, B), and control (si*Control*, C) was examined by immunofluorescence. Nuclei were stained with DAPI (blue). Enclosed areas are enlarged at the top right in each panel. Arrow indicates primary cilia (B, C). Scale bar = 10 μm .

Table 1. Primary and secondary antibodies, and their dilution ratios used for immunofluorescence analyses

Target protein	Source	Reference	Dilution
EGFP	Chicken polyclonal	Abcam; ab13970	1:1000
HSD3B	Rabbit polyclonal	(Fatchiyah et al., 2006)	1:2000
Ad4BP/SF-1	Rabbit polyclonal	(Morohashi et al., 1993)	1:2000
HSD17B3	Rat monoclonal	(Shima et al., 2013)	1:50
COUP-TFII	Mouse monoclonal	Perseus Proteomics, Inc; PP-H7147-00 A2	1:1000
ARX	Rabbit polyclonal	(Kitamura et al., 2002)	1:500
TMSB10	Rabbit polyclonal	Abcam; ab198738	1:100
PECAM1	Rat monoclonal	BD Biosciences; 557355	1:1000
ARL13B	Mouse monoclonal	Antibodies Incorporated; 75-287	1:200
Chicken IgY	Goat anti-chicken IgY H&L (ALEXA Fluor® 488)	Abcam; ab150169	1:500
Rabbit IgG	ALEXA Fluor® 555 goat anti-rabbit IgG (H+L)	Life Technologies; A-21428	1:500
Rat IgG	ALEXA Fluor® 555 goat anti-rat IgG (H+L)	Life Technologies; A-21434	1:500
Mouse IgG	ALEXA Fluor® 488 goat anti-mouse IgG (H+L)	Life Technologies; A-11001	1:500
Rabbit IgG	Goat anti-rabbit IgG H&L (ALEXA Fluor® 647)	Abcam; ab150083	1:500

Table 2. Primers used for qRT-PCR analyses

Gene	Forward (5' → 3')	Reverse (5' → 3')
<i>Actb</i>	AGGGTGTGATGGTGGGAATGG	TGGCTGGGGTGTGAAGGTCT
<i>Star</i>	TACATCCAGCAGGGAGAGGTG	CAGCGCACGCTCACGAAGTCT
<i>Cyp11a1</i>	CGAATCGTCCTAAACCAAGAG	CACTGATGACCCCTGAGAAAT
<i>Hsd3b1</i>	CAAGTGTGCCAGCCTTCATCT	TTCATGATTCTGTTCTCTCGTG
<i>Cyp17a1</i>	GGGTGACCCCAAGGTGGTGGTCTTTCT	TGGGATCCGGGACGTTAGATTCCGG
<i>Hsd17b3</i>	ATGGAGTCAAGGAGGAAAGGC	GGCTGTAAAGAGGCCAGGG
<i>Ad4BP/SF-1</i>	AAGCCACTCTGTAGGACCAAGC	TGTAAATCTGACGCGAAAGCAG
<i>Hkl</i>	ATGCATTGCCCAAGCGTGAACGTA	TCCGGCAGTGTCTGCTGGTGATAA
<i>Aldoa</i>	GCCGCAGCCAGTGAATCTCTCTTC	TTCACAGACAACACCCGCACACGAG
<i>Eno1</i>	TAGGCATCCACACCTGACCACCAG	GGGCTCCAGACACTAGCTGGGAAG
<i>Idh3g</i>	ATTAATGGACGGGCTGTGGAGGCT	TATTCTGGGTCCTGTCCCCCAAAGC
<i>Suclg2</i>	TATCCTCAAGAGCAGCGGACTCCC	CTTCATGGGAGCCAGGGAGAGCAT
<i>Sdha</i>	TGGTCCATGGGCATGTCTCTGAGG	CATGCATGGGGAAGGCTCAGCTTG
<i>Ndufs8</i>	TGGCGGCAACGTACAAGTAT	GTAGTTGATGGTGGCAGGCT
<i>Uqcr10</i>	TACTCCTTGCTGTTCCGCAG	CCACAGTTCCCCTCGTTGA
<i>Sreb2</i>	TCCCAGAGCTGTGAGGATGAAGGC	TGGGCTTGACCGTGGATTTACCA
<i>Acat2</i>	GTGTCTGCGGCAATAGCTAAAGAA	CAGCCAGATGCTCCCAGAGGATG
<i>Hmgcs1</i>	AATGACCACAGTTTGGATGAAGGA	AGGGAGTCTTGGCACTTTCTTAGC
<i>Hmgcr</i>	AGCCTTGGCAGCAGGACATCTTGT	TCTTGGTGACGTTCCCTGAAGAT
<i>Mvk</i>	CAAGTAACGGCAGCACACGGACTG	TGGCTTGCTCTAGACCTGGCTTCA
<i>Pmvk</i>	AGTAGTGGCCTCGGAGCAGAGTCG	AAAGTCCCAAAGTTGTCCAGACC
<i>Mvd</i>	GGGTCCAGTACATCATTGCCACTC	GCAGTCCATCCTGGCCTAGCAGAT
<i>Idi1</i>	CTTGAAAGCCGAGTTGGGAATAC	CCATCAGATTGGGCCTTGTAGTAA
<i>Fdps</i>	TTCAGTGTCTGCTACGAGCCTCTC	CTTTCACCCGAGCCACTTTTTCTG
<i>Fdft1</i>	AACATGCCTGCCGTCAAAGCTATC	GAGATGACCTGCTTGGTTTTGCTT
<i>Sqle</i>	AAACTTGGTGGAGAGTGTGTGACC	CAACGGAAAAGAAGTGTGCAATCA
<i>Lss</i>	GGGATCAGATGTCTGCTAGGGAAG	GTAGCTGATGGCACAGGACTTGTT
<i>Cyp51</i>	CTTACAGGATAACCCAGCATCAGG	TAGGCAAAATTTTCTCCAACACAA
<i>Tm7sf2</i>	CTCCAAAAGTATGGCCGTGCCTGG	AGCAGTGGCTTCCTCTTGCTTTCA
<i>Msmo1</i>	ACCATACGTTTGCTGGAAACCATC	AGCGCCCGTATAAAAAGGAACCAA
<i>Nsdhl</i>	GACACATCTTAGCCGCTGAGCAC	CAGAAAGGGATTGGTTCATCGTT
<i>Hsd17b7</i>	CATGGCACACGTGGCTTTCCATCT	TGTGTGCACGCCTGCAAAAATGTG
<i>Ebp</i>	CTTCCGCTTTGTCCTACAGCTTG	TGGAGTCCTTCGTGTAGCTCTGTC
<i>Sc5d</i>	TACTGGATTCATAGGGGCCTGCAC	GGTGAAAAGCATGACTTGCAAACG
<i>Dhcr7</i>	GGTGGTACCTAGGCTGGGGAGATTG	GGAGAGCTGCACAGGGTGGTACA
<i>Dhcr24</i>	ATGAGGCAGCTGGAGAAGTTTGT	ATCTCCAGAATTCCTCGCGGTTT

Table 2. Primers used for qRT-PCR analyses (Continued)

Gene	Forward (5' → 3')	Reverse (5' → 3')
<i>Mgp</i>	AAGCCCAAAGAGAGTCCAGG	AAGTAGCGGTTGTAGGCAGC
<i>Acta2</i>	G TTCCTTGCCTATCAGAATGG	AAAAACACACAAAAACCCAACC
<i>Amh</i>	GAACCTCTGCCCTACTCGGG	AAGTCCACGGTTAGCACCAAA
<i>Sox9</i>	CACAAGAAAGACCACCCCGA	GATTGCCCAGAGTGCTCGC
<i>Dhh</i>	AGCGCTTCCGGGACCTCGTA	CCCGCTCTTTGCAACGCTCT
<i>Ddx4</i>	GCACACGTTGAATACAGCGGGGAT	TGGGAGGAAGAACAGAAGAACAGG
<i>Notch1</i>	ACAGTGCAACCCCCTGTATG	AGTTGTTCCGTAGCTGGTCCG
<i>Notch2</i>	TGCTGTGCATGCCAGGTTT	ATGTCGATCTGGCACACTGG
<i>Notch3</i>	TGATCGGCTCTGTGGTGATG	ATCCCGAAGTGGGTATGGGA
<i>Jag1</i>	CCTCCTGTCGGGATTTGGTT	AGTGACCCCCATTCAAGCAG
<i>Dlk1</i>	CCTGGCTGTGTCAATGGAGT	TGGCAGTCCTTCCAGAGAA
<i>Hes1</i>	CAACACGACACCGGACAAAC	GGAATGCCGGGAGCTATCTT
<i>Pdgfra</i>	CAAACCCTGAGACCACAATGG	TGATGCCACATAGCCTTCAT
<i>Pdgfrb</i>	AGGACAACCGTACCTTGGGTGACT	CAGTTCTGACACGTACCGGGTCTC
<i>Ptch1</i>	CCGACCCAGATTGCCCTGCC	CAGGGCGTGAGCGCTGACAAG
<i>Ptch2</i>	CTGGCCTCATAGTGCTGGTC	GGGGATGGCACTCAGTTTGA
<i>Smo</i>	GAGCCTTTGCGCTACAACG	CGGAGGCCGGACCAGA
<i>Gli1</i>	ACACAAGTGCACGTTTGAAGG	TCTCATTGGAGTGGGTCCGA
<i>Gli2</i>	CACTAGCCCCTTGAATATGAG	GAGAGGATGCCCCCTAATCTTT
<i>Gli3</i>	GGGAAGAACTTTGCATGTAG	ATAAGGAACCTCTGGGAGAAGG
<i>Tcf21</i>	CAGTCAACCTGACGTGGCCCTT	GCGGTTGGCGGTCACCACTT
<i>Arx</i>	TGCCACTGCCAATGCTACTCCAAC	ATCCGGAGGAAGAGAAGGTTGGGC
<i>Nr2f2</i>	TGCGGAGGAACCTGAGCTAC	CTGTACAGCTTCCCGTCTCAT
<i>Nes</i>	GGGAGAGTCGCTTAGAGGTG	ACAGCCAGCTGGAACCTTTC
<i>EGFP</i>	TATATCATGGCCGACAAGCA	TGTTCTGCTGGTAGTGGTCCG
<i>Tmsb10</i>	AAGCCGGACATGGGGGAAAT	GTTCAATGGTCTCTTTGGTCCG
<i>Pecam1</i>	ACCAGTCCCCGAAGCAGCACT	GTGGAGCAGCTGGCCTGGAC
<i>Tmsb4x</i>	TCCTCTGCCTTCAAAGAAACAAT	AGAAGGCAATGCTCGTGGAA

Table 3. Primers used for CEL-Seq2 analyses

	Primer (5' → 3')
1s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNAGACTCTTTTTTTTTTTTTTTTTTTTTTIV
3s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNAGCTCATTTTTTTTTTTTTTTTTTTTTTIV
8s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCACTAGTTTTTTTTTTTTTTTTTTTTTIV
9s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCAGATCTTTTTTTTTTTTTTTTTTTTTTIV
10s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTCACAGTTTTTTTTTTTTTTTTTTTTTIV
18s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTCGACATTTTTTTTTTTTTTTTTTTTTTIV
25s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNGTGCATTTTTTTTTTTTTTTTTTTTTIV
29s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNACCATGTTTTTTTTTTTTTTTTTTTTTIV
32s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNACGTACTTTTTTTTTTTTTTTTTTTTTTIV
35s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCTAGACTTTTTTTTTTTTTTTTTTTTTTIV
39s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCTCAGATTTTTTTTTTTTTTTTTTTTTTIV
44s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTGCAACTTTTTTTTTTTTTTTTTTTTTTIV
48s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTGTGATTTTTTTTTTTTTTTTTTTTTTIV
57s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNAGTGAGTTTTTTTTTTTTTTTTTTTTTIV
61s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCAACCATTTTTTTTTTTTTTTTTTTTTTIV
71s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTCTGTCTTTTTTTTTTTTTTTTTTTTTTIV
72s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNGTCTCTTTTTTTTTTTTTTTTTTTTTTIV
75s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNGTGAAGTTTTTTTTTTTTTTTTTTTTTIV
77s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNACAGGATTTTTTTTTTTTTTTTTTTTTTIV
83s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNCTTCTGTTTTTTTTTTTTTTTTTTTTTIV
88s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNTGGTGTTTTTTTTTTTTTTTTTTTTTTIV
91s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNGAAGTGTTTTTTTTTTTTTTTTTTTTTIV
96s	GCCGGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNNNNGAGTGATTTTTTTTTTTTTTTTTTTTTTIV

Table 4. List of top 20 genes with increased expression in S-EGFP cells

Seq-I				Seq-II					
Rank	Gene	FPKM		S/W	Rank	Gene	FPKM		S/W
		S-EGFP	W-EGFP				S-EGFP	W-EGFP	
1	<i>Clps</i>	112.34	0.09	104.37	1	<i>Cyp21a1</i>	306.43	0.36	225.96
2	<i>Star</i>	1735.35	16.61	98.57	2	<i>Adh1</i>	2909.46	15.81	173.12
3	<i>Adh1</i>	2122.64	21.20	95.67	3	<i>Insl3,Jak3</i>	24966.60	179.80	138.10
4	<i>Npy</i>	206.39	1.60	79.62	4	<i>Clps</i>	170.56	0.32	130.40
5	<i>Cyp11a1</i>	7197.24	102.69	69.42	5	<i>Star</i>	1249.96	10.18	111.89
6	<i>Cyp17a1</i>	12032.00	172.62	69.30	6	<i>Inha</i>	2417.19	23.19	99.97
7	<i>Cyp11b1</i>	105.09	0.69	62.78	7	<i>Cyp11a1</i>	9598.19	95.84	99.13
8	<i>Hsd3b1</i>	2766.07	43.25	62.54	8	<i>Cyp17a1</i>	12483.90	135.62	91.38
9	<i>Cyp21a1</i>	138.73	1.31	60.56	9	<i>Npy</i>	291.71	2.31	88.56
10	<i>Hao2</i>	200.07	2.36	59.93	10	<i>Hsd3b1</i>	3506.79	40.92	83.69
11	<i>Insl3,Jak3</i>	6767.43	114.56	58.57	11	<i>Lefty2</i>	89.14	0.18	76.46
12	<i>Lefty2</i>	70.52	0.52	47.17	12	<i>Crhr1</i>	226.52	2.21	70.92
13	<i>Atp1a3</i>	86.82	0.97	44.53	13	<i>Ren1,Ren2</i>	7173.08	102.68	69.20
14	<i>Fam129a</i>	260.52	5.20	42.21	14	<i>Hao2</i>	199.21	1.96	67.70
15	<i>Cst8</i>	412.86	9.25	40.39	15	<i>Gpha2</i>	75.11	0.15	66.44
16	<i>Lhcgr</i>	230.90	4.84	39.72	16	<i>Fabp3</i>	421.34	5.95	60.77
17	<i>Ren1,Ren2</i>	7085.40	181.86	38.75	17	<i>Susd3</i>	390.67	5.59	59.40
18	<i>Crhr1</i>	126.91	2.42	37.42	18	<i>Cyp11b1</i>	60.08	0.04	58.98
19	<i>Pgbd5</i>	220.76	5.71	33.05	19	<i>Agt</i>	452.86	6.85	57.78
20	<i>Gramd1b</i>	402.97	11.59	32.08	20	<i>Atp1a3</i>	78.56	0.40	56.95
37	<i>Gpha2</i>	28.09	0.27	22.84	22	<i>Pgbd5</i>	311.41	4.67	55.14
41	<i>Agt</i>	196.41	8.54	20.69	25	<i>Cst8</i>	226.50	3.74	47.98
46	<i>Susd3</i>	222.59	10.20	19.96	28	<i>Lhcgr</i>	229.13	4.04	45.68
54	<i>Inha</i>	1376.13	82.26	16.54	39	<i>Fam129a</i>	130.11	3.02	32.58
76	<i>Fabp3</i>	285.20	20.43	13.35	73	<i>Gramd1b</i>	286.19	12.67	21.00

Top 20 genes showing high S/W values revealed by Seq-I study (left) and Seq-II studies (right) are listed. The genes below the double line in the left table are the genes listed within the top 20 by the Seq-II study but not within the top 20 by the Seq-I study, while the genes below the double line in the right table are the genes listed within the top 20 by the Seq-I study but not within the top 20 by the Seq-II study.

Table 5. List of top 20 genes with increased expression in W-EGFP cells

Seq-I					Seq-II				
Rank	Gene	FPKM		W/S	Rank	Gene	FPKM		W/S
		S-EGFP	W-EGFP				S-EGFP	W-EGFP	
1	<i>Col15a1</i>	9.15	210.61	20.85	1	<i>Lsp1</i>	6.69	324.56	42.32
2	<i>Colla1</i>	48.87	1037.46	20.82	2	<i>Col15a1</i>	3.32	174.86	40.72
3	<i>Mgp</i>	15.48	337.19	20.52	3	<i>Dcn</i>	16.21	661.75	38.51
4	<i>Dcn</i>	22.24	467.74	20.17	4	<i>Lox</i>	1.34	73.66	31.97
5	<i>Pdgfrb</i>	3.04	69.27	17.40	5	<i>Ednra</i>	6.64	237.33	31.20
6	<i>Col6a3</i>	2.74	61.77	16.76	6	<i>Ngfr</i>	1.16	66.21	31.07
7	<i>Lsp1</i>	15.62	276.00	16.66	7	<i>Acta2</i>	15.54	501.75	30.40
8	<i>Acta2</i>	32.28	543.21	16.35	8	<i>Cdh11</i>	5.19	177.61	28.86
9	<i>Amh</i>	7.11	130.50	16.21	9	<i>Heyl</i>	1.13	59.90	28.63
10	<i>Ngfr</i>	4.27	83.83	16.10	10	<i>Mmp2</i>	10.78	328.21	27.94
11	<i>Lum</i>	1.36	36.88	16.03	11	<i>Mfap4</i>	37.31	1054.64	27.55
12	<i>Igfbp5</i>	16.87	278.13	15.62	12	<i>Igfbp5</i>	6.20	189.06	26.39
13	<i>Heyl</i>	2.99	60.53	15.41	13	<i>Mgp</i>	16.99	472.76	26.33
14	<i>Ednra</i>	12.31	202.13	15.26	14	<i>Fn1</i>	14.24	389.36	25.62
15	<i>Rgs5</i>	0.49	20.90	14.69	15	<i>Penk</i>	1.06	48.35	23.91
16	<i>Islr</i>	17.74	271.54	14.54	16	<i>Has2</i>	0.27	29.02	23.56
17	<i>Mfap4</i>	47.51	699.67	14.44	17	<i>Cd34</i>	1.70	61.82	23.31
18	<i>Pdlim3</i>	33.04	487.78	14.36	18	<i>Adcy7</i>	2.92	89.72	23.16
19	<i>Mmp2</i>	18.61	280.02	14.33	19	<i>Rarres2</i>	5.40	146.38	23.03
20	<i>Tagln</i>	8.29	129.48	14.05	20	<i>Akap12</i>	2.05	67.63	22.53
22	<i>Lox</i>	2.64	49.38	13.85	22	<i>Pdlim3</i>	19.90	449.46	21.56
35	<i>Fn1</i>	33.99	446.60	12.79	23	<i>Col6a3</i>	1.07	43.01	21.22
42	<i>Akap12</i>	4.26	63.27	12.21	25	<i>Pdgfrb</i>	1.72	56.45	21.08
47	<i>Rarres2</i>	7.77	104.04	11.98	26	<i>Lum</i>	1.61	53.92	21.05
54	<i>Cd34</i>	2.81	42.09	11.32	30	<i>Islr</i>	12.27	270.50	20.46
94	<i>Cdh11</i>	16.26	165.25	9.63	42	<i>Colla1</i>	33.42	596.52	17.36
105	<i>Adcy7</i>	6.32	67.65	9.37	73	<i>Tagln</i>	6.49	107.62	14.51
124	<i>Penk</i>	3.19	35.02	8.60		<i>Amh</i>	17.39	20.39	1.16
153	<i>Has2</i>	1.18	16.52	8.04		<i>Rgs5</i>	0.25	8.05	7.27

Top 20 genes showing high W/S values revealed by Seq-I study (left) and Seq-II studies (right) are listed. The genes below the double line in the left table are the genes listed within the top 20 by the Seq-II study but not within the top 20 by the Seq-I study, while the genes below the double line in the right table are the genes listed within the top 20 by the Seq-I study but not within the top 20 by the Seq-II study. Although *Amh* is ranked 9th in the Seq-I study, the gene is not listed as the gene whose W/S ratio is more than 2-fold. Since this calculation was done with the gene whose expression is more than 10.0 at least either in S-EGFP or W-EGFP, *Rgs5* gene is not listed.

Table 6. List of top 20 terms obtained by GO and KEGG pathway analyses using genes with increased expression in S-EGFP cells

GO analysis	P-value	KEGG pathway analysis	P-value
mitochondrion	2.35E-121	oxidative phosphorylation	6.74E-25
mitochondrial part	1.45E-72	valine, leucine and isoleucine degradation	7.81E-18
oxidation reduction	4.50E-54	Parkinson's disease	8.84E-17
mitochondrial envelope	1.42E-43	Alzheimer's disease	1.40E-15
mitochondrial inner membrane	6.47E-40	propanoate metabolism	4.83E-15
mitochondrial membrane	1.38E-39	Huntington's disease	2.59E-12
organelle inner membrane	3.13E-39	pyruvate metabolism	1.75E-10
mitochondrial matrix	2.93E-38	steroid biosynthesis	6.20E-10
mitochondrial lumen	2.93E-38	citrate cycle	6.61E-10
generation of precursor metabolites and energy	1.07E-35	fatty acid metabolism	1.26E-09
organelle envelope	1.63E-34	glycolysis / gluconeogenesis	2.27E-08
envelope	2.47E-34	glutathione metabolism	1.37E-07
organelle membrane	1.10E-29	pentose phosphate pathway	4.11E-06
electron transport chain	1.10E-25	beta-Alanine metabolism	5.21E-06
cofactor metabolic process	1.74E-22	terpenoid backbone biosynthesis	5.46E-06
steroid biosynthetic process	2.53E-22	butanoate metabolism	6.45E-06
cofactor binding	2.60E-21	glyoxylate and dicarboxylate metabolism	1.96E-04
lipid biosynthetic process	7.45E-21	fatty acid elongation in mitochondria	2.11E-04
sterol biosynthetic process	8.00E-20	galactose metabolism	2.80E-04
sterol metabolic process	8.54E-20	biosynthesis of unsaturated fatty acids	2.80E-04

Table 7. List of top 20 terms obtained by GO and KEGG pathway analyses using genes with increased expression in W-EGFP cells

GO analysis	P-Value	KEGG pathway analysis	P-Value
extracellular matrix	5.52E-24	focal adhesion	1.17E-17
proteinaceous extracellular matrix	1.33E-23	ECM-receptor interaction	2.71E-12
cytoskeleton organization	1.77E-18	cell cycle	2.76E-09
cytoskeletal protein binding	3.41E-17	DNA replication	2.98E-09
cell division	1.56E-16	pathways in cancer	2.09E-08
cell cycle	3.86E-16	dilated cardiomyopathy	3.99E-08
vasculature development	4.53E-16	regulation of actin cytoskeleton	2.27E-06
extracellular region part	9.64E-16	hypertrophic cardiomyopathy	4.79E-06
extracellular matrix part	1.85E-15	axon guidance	1.79E-05
cytoskeleton	5.69E-15	arrhythmogenic right ventricular cardiomyopathy	2.03E-05
actin cytoskeleton organization	8.64E-15	gap junction	2.79E-05
blood vessel development	1.43E-14	small cell lung cancer	3.40E-04
actin filament-based process	1.60E-14	TGF-beta signaling pathway	4.35E-04
cell cycle process	9.22E-14	vascular smooth muscle contraction	5.19E-04
M phase of mitotic cell cycle	4.83E-13	cell adhesion molecules	5.49E-04
non-membrane-bounded organelle	7.99E-13	progesterone-mediated oocyte maturation	1.14E-03
intracellular non-membrane-bounded organelle	7.99E-13	pancreatic cancer	2.71E-03
mitotic cell cycle	1.08E-12	leukocyte transendothelial migration	3.56E-03
nuclear division	1.10E-12	mismatch repair	7.51E-03
mitosis	1.10E-12	hedgehog signaling pathway	1.20E-02