CHAPTER 2 MATERIALS AND METHODS

2.1. Tissue culture

2.1.1. Human pluripotent stem cell lines

H9 hESCs were purchased from WiCell (WA09). FSPS13.B hiPSCs were derived by a colleague from peripheral blood mononuclear cells (PBMCs) using Sendai virus as described previously (Yusa et al., 2011) by the Human induced pluripotent stem cell initiative (HIPSCI). The resulting cells are available for distribution through European Collection of Authenticated Cell Cultures (ECACC). Further information is available at http://www.hipsci.org/lines/#/lines/HPSI0813i-fpdm_2.

The generation of *GATA6* patient-derived hiPSCs was approved by the Great Ormond Street Hospital and Institute of Child Health Research Ethics Committee (ethics reference number: 08/H0713/82), and informed consent was obtained from all patients. Skin fibroblasts from *GATA6* patients were obtained with the help of Dr Ranna El Khairi. Skin punch biopsy samples were collected from volunteering patients and all hiPSC lines used were derived and validated by the Cambridge Biomedical Research Centre hiPSC Core Facility.

Reprogramming of the *GATA6* patient fibroblasts to derive *GATA6*-patient iPSCs was done with the help of the hiPSC core facility at the Anne McLaren Laboratory for Regenerative Medicine using sendai virus reprogramming as described previously (Yusa et al., 2011).

2.1.2. Growth conditions

hPSCs were routinely cultured in feeder-free conditions on vitronectin-coated (Stemcell Technologies #07180) sterile plates (Corning, Costar) with Essential 8 (E8) media (Life Technologies #A1517001). Subculture was performed every 4-6 days by aspirating the cell culture media, washing the cells once with D-PBS, incubating them in 0.5 mM EDTA (Invitrogen, #15575020) for 5 min at room temperature, aspirating the EDTA and adding appropriate volumes of fresh E8 media (Table 4) to the cells. hPSCs were further physically detached from the plates by gentle pipetting, collected in a 15 ml Falcon tube (BD Falcon) and allowed to settle for 10 min. The supernatant was aspirated and hPSCs were re-suspended in fresh E8 media, and then dissociated into smaller clumps by gentle pipetting. The cells were split and plated onto fresh vitronectin-coated plates (Table 5) containing appropriate volumes of media (Table 4). Cells were maintained by refreshing E8 media on a daily basis.

Prior to differentiation into any lineages, culture plates were pre-coated with 0.1% Gelatin (Millipore, #ES-006-B) for at least 20 min at room temperature. The gelatin solution was then aspirated and replaced with MEF media (Table 6 and Table 7). The MEF-coated plates were allowed to incubate at 37°C overnight, up to 1 week, and used when required.

Table 4. Volume of E8 used in splitting and maintenance of cells

Culture vessel	During splitting prior to physical detachment	Maintenance
10 cm dish	3 ml	7 ml
1 well of 6wp	1 ml per well	1.5 ml
1 well of 12wp	0.5 ml per well	1 ml
1 well of 24wp	0.3 ml per well	0.5 ml

Table 5. Splitting ratio of cells

Initial growth area	Final growth area
1 x 10 cm dish	4 x 10 cm dishes
1 well of 6wp	1 full 6wp, 12wp or 24wp

Table 6. MEF media formulation

Component	Volume
Advanced DMEM/F-12 (Gibco, #12634028)	450 ml
FBS (Biosera, #S04253S181S)	50 ml
L-Glutamine (Gibco, #25030024)	5 ml
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml
β-Mercaptoethanol (Sigma-Aldrich, #M6250)	3.5 µl

Table 7. Volume of MEF media used for coating

Culture vessel	Volume
10 cm dish	8 ml
1 well of 6wp	2 ml
1 well of 12wp	1.5 ml
1 well of 24wp	1 ml

2.1.3. Definitive endoderm (DE) differentiation

To plate hPSCs for differentiation, hPSCs were grown to about 80% confluency and then passaged as described in the previous section. The cells were broken into clumps smaller than routine passaging by doubling the number of times they pass through the pipette tip. They were then filtered through a 70 µm cell strainer (BD Falcon, #352350) to obtain a uniformly-sized cell suspension, re-suspended in fresh E8 media containing ROCK inhibitor Y-27632 (Sigma-Aldrich, #Y0503) and plated onto the appropriate MEF-coated culture plates. Prior to plating, the MEF media was aspirated and plates were washed once with D-PBS (Life Technologies, #14190094). Cells were plated at a density such that they were about 60 - 70% confluent at the start of DE differentiation. 24 hours after plating, the culture media was replaced with fresh E8 media without ROCK inhibitor Y-27632. DE differentiation commenced 48 hours after plating.

Using the lab's established DE differentiation protocol, the E8 media was aspirated and cells were washed once with D-PBS to remove any residual E8 media. Chemically defined medium-poly vinyl alcohol (CDM-PVA; Table 8) media of appropriate volume (Table 9) supplemented with 100 ng/ml Activin A (produced in-house), 20 ng/ml basic FGF (bFGF/FGF2; produced in-house), 10 ng/ml BMP4 (R&D Systems, #314-BP-010), 10 µM LY294002 (Promega, #V1201) and 3 µM CHIR99201 (Stratech Scientific, #CT99021) was added to the cells on Day 1 (Table 10). On Day 2, media from Day 1 was aspirated and CDM-PVA media containing 100 ng/ml Activin A, 20 ng/ml FGF2, 10 ng/ml BMP4, and 10 µM LY294002 was added to the cells. On Day 3, media from Day 2 was aspirated and RPMI-B27 media (Table 11) supplemented with 100 ng/ml Activin A and 20 ng/ml FGF2 was added to the cells. Cells were harvested on Day 4 for downstream analyses of the DE.

For DE differentiation using a commercial kit PSC Definitive Endoderm Induction Kit (Gibco, #A27654SA), cells were initially passaged and plated similarly to the lab's established protocol. Volumes of media used are indicated in Table 9. On Day 1, Medium A of the kit was added and on Day 2, Medium A was replaced with Medium B. Cells were harvested on Day 3 for downstream analyses of the DE.

Table 8. CDM-PVA formulation

Component	Volume
Ham's F-12+ GlutaMAX-1 (Gibco, #31765068)	250 ml
Iscove's Modified Dulbecco's Medium (Gibco, #21980065)	250 ml
Chemically Defined Lipid Concentrate (Gibco, #11905031)	5 ml
1-Thioglycerol (Sigma-Aldrich, #M6145)	20 µl
Transferrin (30 mg/ml; Roche #652202)	250 µl
Insulin (10 mg/ml; Roche #1376497)	350 µl
Poly vinly alcohol (Sigma-Aldrich, #P8136)	0.5 g
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml

Table 9. Volume of media used for DE differentiation

Culture vessel	Volume
10 cm dish	7 ml
1 well of 12wp	0.5 ml
1 well of 24wp	0.3 ml

Table 10. Lab DE differentiation protocol

Time point	Media	Supplements
Day -1 (plate)	E8	ROCK inhibitor
Day 0	E8	None
Day 1	CDM-PVA	100 ng/ml Activin A + 20 ng/ml FGF2 + 10 ng/ml BMP4 + 10 µM LY294002 + 3 µM CHIR99201
Day 2	CDM-PVA	100 ng/ml Activin A + 20 ng/ml FGF2 + 10 ng/ml BMP4 + 10 µM LY294002
Day 3	RPMI-B27	100 ng/ml Activin A + 20 ng/ml FGF2

Table 11. RPMI-B27 formulation

Component	Volume
RPMI 1640 + GlutaMAX (Gibco, #61870-10)	500 ml
B-27 supplement (Gibco, #17504-044)	10 ml
MEM Non-Essential Amino Acids Solution (Gibco, #11140-050)	5 ml
+/- Penicillin-Streptomycin (Gibco, #15140122)	5 ml

2.1.4. Pancreatic differentiation

Following on from DE differentiation, cells were further specified toward the pancreatic lineage by using a protocol that was previously established in the lab. On Day 4, Day 3 media was aspirated and replaced with Advanced DMEM/F-12 containing bovine serum albumin (ADV-BSA; Table 12) supplemented with 2 µM retinoic acid (RA; Sigma-Aldrich, #R2625), 50 ng/ml FGF10 (Autogen Bioclear, #ABC144), 150 ng/ml Noggin (R&D, #33-44NG/CF) and 10 µM SB-431542 (R&D, #1614/10) (Table 13) using appropriate volumes indicated in Table 14. On Days 5 and 6, media was refreshed with the same supplements as Day 4. Cells were harvested on Day 7 for downstream analyses of the primitive gut tube.

On Day 7, Day 6 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 50 ng/ml FGF10, 150 ng/ml Noggin and 1 mg/ml KAAD-cyclopamine (Toronto Research Chemicals Incorporated, #K171000). On Days 8 and 9, media was refreshed with the same supplements as Day 7. Cells were harvested on Day 10 for downstream analyses of the posterior foregut.

On Day 10, Day 9 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 50 ng/ml FGF10 and 0.25 µg/ml KAAD-cyclopamine (Toronto Research Chemicals Incorporated, #K171000). On Days 11 and 12, media was refreshed with the same supplements as Day 10. Cells were harvested on Day 13 for downstream analyses of the pancreatic progenitors.

On Day 13, Day 12 media was aspirated and replaced with ADV-BSA supplemented with 2 µM RA, 1% vol/vol B27 supplement, 1 µM DAPT (Sigma-Aldrich, #D5942) and 0.1 mM 6-Bnz-cAMP sodium salt (BNZ; Sigma-Aldrich, #B4560). Media was left unchanged on Days 14 and 15. Cells were harvested on Day 16 for downstream analyses of the endocrine progenitors.

On Days 16, 18 and 22, cells were then allowed to mature into immature β-cells by aspirating media from Days 15, 17 and 21 respectively and adding ADV-BSA supplemented with 2 µM RA, 1% B27 supplement and 0.25 µg/ml KAAD-cyclopamine. Cells were harvested on Days 19 and 25 for downstream analyses of immature endocrine cells and β-cells respectively.

Table 12. ADV-BSA formulation

Component	Volume
Advanced DMEM/F-12 (Gibco, #21634010)	500 ml
Bovine Serum Albumin (Europa Bioproducts, Lot #1260)	2.5 g
L-Glutamine (Gibco, #25030024)	5 ml
Penicillin-Streptomycin (Gibco, #15140122)	5 ml

Table 13. Pancreatic differentiation protocol

Time point	Media	Supplements
Day 4	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 5	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 6	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 10 µM SB-431542
Day 7	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 8	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 9	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 150 ng/ml Noggin + 0.25 µg/ml KAAD-cyclopamine
Day 10	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Day 11	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Day 12	ADV-BSA	2 µM RA + 50 ng/ml FGF10 + 0.25 µg/ml KAAD-cyclopamine
Days 13-15	ADV-BSA	2 µM RA + 1% B27 + 1 µM DAPT + 0.1 mM BNZ
Days 16-18	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine
Days 18-21	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine
Days 22-24	ADV-BSA	2 µM RA + 1% B27 + 0.25 µg/ml KAAD-cyclopamine

Table 14. Volume of media used for pancreatic differentiation

Culture vessel	Volume
10 cm dish	7 ml each day from Days 4 to 9, 14 ml on Days 13, 16, 18 and 22
1 well of 12wp	0.5 ml each day from Days 4 to 9, 1.5 ml on Days 13, 16, 18 and 22
1 well of 24wp	0.3 ml each day from Days 4 to 9, 1 ml on Days 13, 16, 18 and 22

2.1.5. Pancreatic differentiation using STEMdiff pancreatic progenitor kit

Alternatively, cells were also differentiated into pancreatic progenitors using the STEMdiff pancreatic progenitor kit (Stem Cell Technologies, #05120). They were cultured in E8 media as described in section 2.1.1 and plated for differentiation as described in section 2.1.3. Cells were differentiated according to the manufacturer's protocol and were harvested on Days 4 and 14 for downstream FACS analyses of the DE and pancreatic progenitors respectively, and Days 4, 7, 10 and 14 for downstream qRT-PCR analyses of the DE, primitive gut tube, posterior foregut and pancreatic progenitors respectively.

2.1.6. Glucose response assay and Enzyme linked immunosorbent assay (ELISA) for C-peptide

hPSCs differentiated into the endocrine progenitors after Day 24 are cultured in differentiation medium without insulin for 24 hr prior to ELISA. Cells were washed thrice in D-PBS and pre-incubated in DMEM supplemented with 2.5 mM glucose (Invitrogen) for 60 min at 37°C. To perform glucose-induced insulin secretion, the buffer is replaced with 22.5 mM glucose (high glucose) alternatively with 5.5 mM glucose (low glucose) twice at 30 min intervals. The supernatant is collected after each round of incubation and stored at -80°C for determination of C-peptide release. For control, cells were incubated in DMEM supplemented with 2.5 mM glucose.

C-peptide ELISA was measured using the Mercodia Ultrasensitive C-peptide ELSIA kit (Mercodia) following the manufacturer's instructions. Absorbance was read at 450 nm using Tecan Infinite 200 Pro plate reader.

2.1.7. Hepatic differentiation

Following on from DE differentiation, cells were further specified toward the hepatic lineage by using a protocol that was previously established in the lab. From Day 4 to Day 9, cells were cultured in 0.5 ml per well of a 12 well plate of RPMI-B27 (Table 11) and supplemented with 50 ng/ml Activin A (Table 15). Media was refreshed daily.

From Days 10 to 26, cells were cultured in 1 ml per well of a 12 well plate of Hepatozyme (Table 16) and supplemented with 10ng/ml Oncostatin M (OSM; R&D, #295-OM) and 50ng/ml Hepatocyte growth factor (HGF; Peprotech, #100-39). Media was refreshed every other day.

Table 15. Hepatic differentiation protocol

Time point	Media	Supplements	Volume
Day 4	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 5	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 6	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 7	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 8	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 9	RPMI-B27	50 ng/ml Activin A	0.5 ml
Day 10-11	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Day 12-13	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Day 14-15	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 16-17	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 18-19	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 20-21	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 22-23	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml
Days 24-26	Hepatozyme	10 ng/ml OSM + 50 ng/ml HGF	1 ml

Table 16. Hepatozyme formulation

Component	Volume
Hepatozyme (Gibco, #17705-021)	500 ml
MEM Non-Essential Amino Acids Solution (Gibco, #11140-050)	10.6 ml
Chemically Defined Lipid Concentrate (Gibco, #11905031)	10.6 ml
L-Glutamine (Gibco, #25030024)	5.3 ml
Transferrin (30 mg/ml; Roche #652202)	530 µl
Insulin (10 mg/ml; Roche #1376497)	742 µl
+/- Penicillin-Streptomycin (Gibco, #15140122)	5.3 ml

2.2. Cloning of plasmid DNA constructs

2.2.1. Transformation of plasmids into *Escherichia coli* cells

XL-10 Gold Ultracompetent cells (Agilent Technologies, #200314) were thawed on ice. The competent cells were gently mixed by gentle tapping of the tube then 50 µl of competent cells were aliquoted into tubes containing DNA plasmids or ligation reactions while keeping all tubes on ice. The tubes were gently tapped to mix. The cells were incubated on ice for 20 min, then heat-shocked for 45 s in a 42°C water bath. They were then placed back on ice for a further 2 min. 500 µl of room temperature S.O.C. Medium (Invitrogen) was then added to the tubes and they were incubated shaking at 225 rpm for 1 hr at 37°C. 200-500 µl of the reaction mix was then spread on LB plates containing the appropriate antibiotics and incubated overnight at 37°C.

2.2.2. Small scale DNA plasmid purification and colony selection

Plasmid DNA constructs were prepared using QIAprep Spin Miniprep kit (Qiagen, #27106). They were analysed by restriction digest to confirm presence of the insert in the correct orientation.

2.2.3. Genotyping via Sanger sequencing

Plasmid DNA constructs were sent to AITBiotech, Singapore and GATC Biotech, UK for genotype verification.

2.3. Constructs

For DNA plasmid constructs used in cell culture, they were prepared using EndoFree Plasmid Maxi Kit (Qiagen, #12362). For all other applications, DNA plasmid constructs were prepared using QIAprep Spin Miniprep kit (Qiagen, #27106).

2.3.1. Transcription activator-like effector nuclease (TALEN) vectors

Suitable TALEN target sites on the *GATA6* gene were generated using an online TALEN targeter software tool (Cermak et al., 2011, Doyle et al., 2012). Two TALEN targets within exon 2 of the *GATA6* gene were selected based on higher numbers of RVDs HD and NH and the presence of a restriction site in the spacer region; one targeting a site that is 6 base pairs downstream of the first start codon while the other targeting a site that is 149 base pairs downstream of the second start codon. They are termed TALEN 1 and TALEN 2 respectively. The sequences of the two selected TALEN target pairs are listed in Table 17.

Table 17. Two selected TALEN target sites for *GATA6*

TALEN pair	Name	Sequence (5' → 3')
TALEN 1 (T1)	Left arm	GACTGACGGCGGCTGGT
	Right arm	CCGCACCCGCGGCCCCG
TALEN 2 (T2)	Left arm	GCTGCCCGGCCTACCGT
	Right arm	GGCTGGCCCACTGCCC

TALEN vectors were assembled using the Joung Lab REAL Assembly TALEN kit (Addgene, #1000000017). Modifications of the pTAL scaffold into the second generation GoldyTALEN scaffold and NN RVDs into NH were performed by Dr Norihiro Tsuneyoshi. Plasmids obtained from this kit used for the TALEN assemblies and their corresponding RVDs are listed in Table 18. The selected target sequences were entered into a ZiFiT targeter software via the REAL assembly method prior to TALEN assembly (Sander et al., 2007, Sander et al., 2010). Assembly of the TALE repeat arrays were performed in three polymerase chain reaction (PCR) steps (Table 19 for T1 and Table 20 for T2) with the primers listed in Table 21. Primers used in the final PCR step included the Kozak sequence. PCR was performed using PrimeSTAR Max DNA polymerase (Takara, # R045B) according to the manufacturer's protocol.

Table 18. Plasmids used and their corresponding RVDs

Plasmid	RVD
TAL006	NI
TAL007	HD
TAL009	NN to NH
TAL010	NG
TAL011	NI
TAL012	HD
TAL014	NN to NH
TAL015	NG
TAL016	NI
TAL017	HD
TAL019	NN to NH
TAL020	NG
TAL021	NI
TAL022	HD
TAL024	NN to NH
TAL025	NG

Table 19. TALEN 1 repeat array assembly via three step PCR method

PCR step	TALEN name	Primer name	RVD number	Plasmid
1	T1 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL009
			5	TAL009
			9	TAL009
			13	TAL007
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2	TAL011
			6	TAL011
			10	TAL012
			14	TAL015
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL017
			7	TAL017
			11	TAL019
			15	TAL019
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL025
			8	TAL024
			12	TAL024
			16	TAL024
	T1 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL007
			5	TAL006
			9	TAL009
			13	TAL007

		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2 6 10 14	TAL012 TAL012 TAL012 TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3 7 11 15	TAL019 TAL017 TAL019 TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4 8 12 16	TAL022 TAL022 TAL024 TAL022
2	T1 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL011
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL017, TAL025
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL009, TAL011
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL019, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL007, TAL015
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL019, TAL024
	T1 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL019, TAL022
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL006, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL022
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL019, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL017, TAL022

3	T1 left arm	TALEN-RVDs 1-4 Fwd TALEN-RVDs 1-4 Rev	1 2 3 4	TAL009, TAL011, TAL017, TAL025
		TALEN-RVDs 5-8 Fwd TALEN-RVDs 5-8 Rev	5 6 7 8	TAL009, TAL011, TAL017, TAL024
		TALEN-RVDs 9-12 Fwd TALEN-RVDs 9-12 Rev	9 10 11 12	TAL009, TAL012, TAL019, TAL024
		TALEN-RVDs 13-16 Fwd TALEN-RVDs 13-16 Rev	13 14 15 16	TAL007, TAL015, TAL019, TAL024
	T1 right arm	TALEN-RVDs 1-4 Fwd TALEN-RVDs 1-4 Rev	1 2 3 4	TAL007, TAL012, TAL019, TAL022
		TALEN-RVDs 5-8 Fwd TALEN-RVDs 5-8 Rev	5 6 7 8	TAL006, TAL012, TAL017, TAL022
		TALEN-RVDs 9-12 Fwd TALEN-RVDs 9-12 Rev	9 10 11 12	TAL009, TAL012, TAL019, TAL024
		TALEN-RVDs 13-16 Fwd TALEN-RVDs 13-16 Rev	13 14 15 16	TAL007, TAL012, TAL017, TAL022

Table 20. TALEN 2 repeat array assembly via three step PCR method

PCR step	TALEN name	Primer name	RVD number	Plasmid
1	T2 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1 5 9 13	TAL009 TAL007 TAL009 TAL006
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2 6 10 14	TAL012 TAL012 TAL012 TAL012

		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL020
			7	TAL017
			11	TAL017
			15	TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL024
			8	TAL024
			12	TAL025
			16	TAL024
	T2 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 1 Rev	1	TAL009
			5	TAL009
			9	TAL007
			13	TAL009
		TALEN-RVDs 2 Fwd TALEN-RVDs 2 Rev	2	TAL014
			6	TAL014
			10	TAL011
			14	TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 3 Rev	3	TAL017
			7	TAL017
			11	TAL017
			15	TAL017
		TALEN-RVDs 4 Fwd TALEN-RVDs 4 Rev	4	TAL025
			8	TAL022
			12	TAL025
2	T2 left arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL020, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	5 6	TAL007, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	7 8	TAL017, TAL024
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	9 10	TAL009, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	11 12	TAL017, TAL025
		TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	13 14	TAL006, TAL012
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	15 16	TAL017, TAL024
	T2 right arm	TALEN-RVDs 1 Fwd TALEN-RVDs 2 Rev	1 2	TAL009, TAL014
		TALEN-RVDs 3 Fwd TALEN-RVDs 4 Rev	3 4	TAL017, TAL025

		TALEN-RVDs 1 Fwd	5	TAL009,
		TALEN-RVDs 2 Rev	6	TAL014
		TALEN-RVDs 3 Fwd	7	TAL017,
		TALEN-RVDs 4 Rev	8	TAL022
		TALEN-RVDs 1 Fwd	9	TAL007,
		TALEN-RVDs 2 Rev	10	TAL011
		TALEN-RVDs 3 Fwd	11	TAL017,
		TALEN-RVDs 4 Rev	12	TAL025
		TALEN-RVDs 1 Fwd	13	TAL009,
		TALEN-RVDs 2 Rev	14	TAL012
		TALEN-RVDs 3 Fwd	15	TAL017
		TALEN-RVDs 4 Rev		
3	T2 left arm	TALEN-RVDs 1-4 Fwd	1	TAL009,
		TALEN-RVDs 1-4 Rev	2	TAL012,
			3	TAL020,
			4	TAL024
		TALEN-RVDs 5-8 Fwd	5	TAL007,
		TALEN-RVDs 5-8 Rev	6	TAL012,
			7	TAL017,
			8	TAL024
		TALEN-RVDs 9-12 Fwd	9	TAL009,
		TALEN-RVDs 9-12 Rev	10	TAL012,
			11	TAL017,
			12	TAL025
		TALEN-RVDs 13-16 Fwd	13	TAL006,
		TALEN-RVDs 13-16 Rev	14	TAL012,
			15	TAL017,
			16	TAL024
	T2 right arm	TALEN-RVDs 1-4 Fwd	1	TAL009,
		TALEN-RVDs 1-4 Rev	2	TAL014,
			3	TAL017,
			4	TAL025
		TALEN-RVDs 5-8 Fwd	5	TAL009,
		TALEN-RVDs 5-8 Rev	6	TAL014,
			7	TAL017,
			8	TAL022
		TALEN-RVDs 9-12 Fwd	9	TAL007,
		TALEN-RVDs 9-12 Rev	10	TAL011,
			11	TAL017,
			12	TAL025
		TALEN-RVDs 13-16 Fwd	13	TAL009,
		TALEN-RVDs 13-16 Rev	14	TAL012,
			15	TAL017

Table 21. Primers used to assemble TALEN repeat arrays

Primer name	Primer sequence (5' → 3')
TALEN-RVDs 1 Fwd	CTGACCCCAGACCAGGTAGTCGCA
TALEN-RVDs 1 Rev	CACGACTTGATCCGGTGTAAGGCCGTGGTCTTGACAAAGG
TALEN-RVDs 2 Fwd	CCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTG
TALEN-RVDs 2 Rev	TACAACTTGATCGGGAGTCAGCCCGTGGTCTTGACAGAGA
TALEN-RVDs 3 Fwd	TCTCTGTCAAGACCACGGGCTGACTCCCGATCAAGTTGTA
TALEN-RVDs 3 Rev	GACCACTTGGTCAGGCGTCAAACCGTGATCTTGACACAAC
TALEN-RVDs 4 Fwd	GTTGTGTCAAGATCACGGTTTGACGCCTGACCAAGTGGTC
TALEN-RVDs 4 Rev	TCCATGATCCTGGCACAGTACAGG
TALEN-RVDs 1-4 Fwd	tcagGGTCTCAGAACCTGACCCCAGACCAGGTAGTC
TALEN-RVDs 1-4 Rev	tcagGGTCTCTAGTCCATGATCCTGGCACAGT
TALEN-RVDs 5-8 Fwd	tcagGGTCTCAGACTGACCCCAGACCAGGTAGTC
TALEN-RVDs 5-8 Rev	tcagGGTCTCTGTCAGTCCATGATCCTGGCACAGT
TALEN-RVDs 9-12 Fwd	tcagGGTCTCATGACCCCAGACCAGGTAGTC
TALEN-RVDs 9-12 Rev	tcagGGTCTCTCAGTCCATGATCCTGGCACAGT
TALEN-RVDs 13-16 Fwd	tcagGGTCTCAACTGACCCCAGACCAGGTAGTC
TALEN-RVDs 13-16 Rev	tcagGGTCTCTTCAGTCCATGATCCTGGCACAGT

Plasmids derived from PCR steps 1 and 2 were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, #28706) and the final TALE repeat array plasmids derived from PCR step 3 were purified using QIAquick PCR Purification Kit (Qiagen, #28106). The TALE repeat arrays were then digested with restriction enzyme BsaI-HF (New England Biolabs, #R3535S) for 2 hours at 37°C then purified by gel extraction, while the respective vector backbones for each TALE repeat array (Table 22) were digested with restriction enzyme BsmBI (New England Biolabs, #R0580S) for 3 hours at 55°C then purified by gel extraction. The TALE repeat array were next ligated with their respective vector backbones by using T4 DNA ligase (New England Biolabs, #M0202S) for 2 hours at room temperature. The success of the TALEN assembly was verified by Sanger sequencing (Table 23).

Table 22. TALE repeat arrays and their corresponding vector backbone

TALEN name	Vector backbone
T1 left arm	JDS78
T1 right arm	JDS74
T2 left arm	JDS78
T2 right arm	JDS71

Table 23. Primers used for sequencing TALEN assembly

Primer name	Primer sequence (5' → 3')
TALEN-006seq Fwd	TCGCAATCGCGTCGAACATTG
TALEN-006seq Rev	GCTTGCTTTCCCCCAATGTTC
TALEN-007seq Fwd	TCGCAATCGCGTCACATGAC
TALEN-009seq Fwd	TCGCAATCGCGTCAAATCAT
TALEN-009seq Rev	CTTGCTTTCCCCCATGATTT
TALEN-010seq Fwd	TCGGCAATCGCGTCAAACGGA
TALEN-010seq Rev	CTTGCTTTCCCCCTCCGTTT
TALEN-011seq Fwd	GTGGCCATTGCAAGCAACATC
TALEN-011seq Rev	GAGCCTGTTTGCCACCGATG
TALEN-012seq Fwd	TGGCCATTGCATCCCACGAC
TALEN-012seq Rev	CTGTTTGCCACCGTCGTGG
TALEN-020seq Rev	AATGCTTGTTTCCCTCCACCG
TALEN-022seq Rev	CTGCTTACCGCCATCATGG
TALEN-024NHseq Rev	AGCGCCTGCTTACCGCCATG
TALEN-025seq Fwd	TCGCCATCGCCTCGAATGGC
TALEN-025seq Rev	CTGCTTACCGCCGCCATTC

Next, the assembled TALEN arms were cloned into vectors containing a CAG promoter and an antibiotic resistance gene (Table 24) that were already available in the lab. To do this, the TALEN arms were PCR amplified using PrimeSTAR Max DNA polymerase using primers listed in Table 25 and PCR conditions listed in Table 26. They were then purified using QIAquick PCR Purification Kit, digested by NheI restriction enzyme (New England Biolabs, #R0131S), blunted using Quick Blunting Kit (New England Biolabs, #E1201S) for 5 min at 72°C, purified using QIAquick PCR Purification Kit once more, digested by XhoI restriction enzyme (New England Biolabs, #R0146S), then gel purified by gel extraction. For the vectors, they were first digested by EcoRI-HF restriction enzyme (New England Biolabs, # R3101S), blunted as described earlier, purified using QIAquick PCR Purification Kit, digested by XhoI restriction enzyme, then gel purified by gel extraction. The TALEN arms were next ligated with their respective vector containing the appropriate antibiotic resistance gene by using T4 DNA ligase for 2 hours at room temperature. The final TALEN constructs were then sequenced to confirm that the TALEN arms were cloned in the correct orientation (Table 27).

Table 24. Antibiotic resistance gene specific to each TALEN arm

TALEN construct name	Antibiotic resistance gene
T1 left arm	Puromycin
T1 left arm	Blasticidin
T1 right arm	Zeocin
T2 left arm	Puromycin
T2 right arm	Zeocin

Table 25. Primers used to amplify TALEN arms

Primer name	Primer sequence (5' → 3')
T7 fwd	AATACGACTCACTATAG
TALEN-pCAG-IRES Rev	AACTTTTAAACCGGTCTCGAGCTGA

Table 26. Parameters for PCR cycling reaction to amplify TALEN arms

Step	Temperature	Time	Number of cycles
Denaturation Annealing Extension	98°C 55°C 72°C	10 s 15 s 5 s/1kb	35
Soak	4°C	Indefinite	1

Table 27. Primers used for sequencing TALEN constructs

Primer name	Primer sequence (5' → 3')
pre-KpnI TALEN-Core-Fwd	GTTGTAAAACGACGGCCAGTG
025seq-Fwd	TCGCCATCGCCTCGAATGGC
post-BamHI TALEN-Core-Rev	TACGCCAAGCTTGCATGCAGG
Fok1-Fwd	GTGAACTGGAGGAGAAGAAATCTG
pCAG-IP +1760 Rev	GGGCGGAATTTACGTAGCGG
007seq-Fwd	TCGCAATCGCGTCACATGAC
006seq-Fwd	TCGCAATCGCGTCGAACATTG
011seq-Rev	GAGCCTGTTTGCCACCGATG

A donor plasmid aimed at terminating transcription of *GATA6* prematurely by inserting a 'donor template' through HR was also constructed. Within the donor plasmid is a cassette which contains 5' and 3' homology arms each 1kb in length recognising the flanking regions of the TALEN 1 target site, an EmGFP gene, a puromycin antibiotic resistant gene and a polyA tail. PCR amplification of genomic DNA from H9 hESC line and a vector already available in the lab containing the EmGFP gene, puromycin gene and polyA tail was performed to obtain PCR products of the 5' and 3' homology arms and the other components of the donor plasmid respectively (Table 28). PCR was performed using PrimeSTAR Max DNA polymerase and conditions listed in Table 26. The PCR products were then ligated and the final construct was sequenced to confirm that the donor plasmid was cloned successfully (Table 29).

Table 28. Primers used to construct the donor plasmid

Primer name	Primer sequence (5' → 3')
5' Arm-KpnI-GATA6 Fwd	tcagGGTACCTTTGGGGTCGCCTCGGCTCTGG
5' Arm-GATA6 Rev	CTTGCTCACCATGGTGGCCACGGTCCGGCGCCGCTCCAA
5' Arm-GATA6-emGFP Fwd	CGCCGGACCGTGGCCACCATGGTGAGCAAGGGCGAGGAGC
3' Arm-XbaI-TALEN1 Fwd	tcagTCTAGAAAGCGCTTCGGGGCCGCGGGTG
3' Arm-SacI-TALEN1 Rev	tcagGAGCTCTGGCGCCCCCACGTAGGGCGAG

Table 29. Primers used for sequencing donor plasmid

Primer name	Primer sequence (5' → 3')
EmGFP3'-Fwd	TCACATGGTCCTGCTGGAGTTC
BGHpA-mid-Rev	TTAGGAAAGGACAGTGGGAGTG
EmGFP5'-Rev	CGCTGAACTTGTGGCCGTTTAC
EmGFP-mid-Rev	GACCTTGTGGCTGTTGTAGTTG
mPGKpA-Fwd	AAGAAGGGTGAGAACAGAGTACC
M13-Rev (-24)	GGAAACAGCTATGACCATG
M13-Fwd (-20)	GTAAAACGACGGCCAGT
pCAGGS pre-SA Fwd	CTGCTAACCATGTTCATGCCTTC

2.3.2. Introducing constructs into hPSCs

Electroporation using the Amaxa Nucleofector Technology (Lonza, Human Stem Cell Nucleofector Kit 1) was used to introduce the TALEN constructs and donor plasmid into the hPSCs. Cells were grown in 6 well plates to a confluency of 70-80%, then washed once with D-PBS and treated with 0.5 ml of StemPro Accutase Cell Dissociation Reagent (Gibco, #A1110501) per well for 5 min at 37°C. 1 ml of 5% fetal bovine serum (FBS; Gibco, #10082147) diluted in D-PBS was added per well and the cells were detached from the wells by gentle pipetting. Cells were collected in a 15 ml tube and an aliquot was taken for cell counting. Cell counting was performed using an automated cell counter (Biorad, TC20). 8×10^5 cells were used per electroporation and the appropriate volume that contained 8×10^5 cells was taken and divided into individual 15 ml tubes. The tubes were centrifuged at 1,200 rpm to pellet the cells. Cells were electroporated with a pre-determined amount of DNA for each construct. Electroporation was performed according to the manufacturer's protocol using Nucleofector programme B-016. The electroporated cells were next plated as single cells onto 10 cm dishes containing 10 ml of E8 with ROCK inhibitor.

2.4. Generation of *GATA6* mutant lines

2.4.1. Non-homologous end joining (NHEJ) pathway

TALENs were introduced into H9 and FSPS13.B hPSC lines via electroporation as described in section 2.3.2. For electroporation, 2.5 µg of DNA for each of the corresponding TALEN arm was used (Table 30). 24 hr after electroporation, simultaneous antibiotic selection of 1 µg/ml puromycin (Sigma-Aldrich, #P8833) and 2.5 µg/ml zeocin (Gibco, #R250-01) was done for 24 hr, after which the cells were allowed to recover and form colonies. Colonies formed approximately 7 days after the antibiotic selection, and they were individually picked using a pipette tip, re-plated and expanded. During picking of each colony, half of one colony was re-plated whereas the other half was pipetted into 8-tube PCR strips for subsequent screening.

Table 30. Electroporation of TALENs into hPSCs

TALEN target site	TALEN construct name	Antibiotic resistance gene	Amount of DNA electroporated
1	T1 left arm	Puromycin	2.5 µg
	T1 right arm	Zeocin	2.5 µg
2	T2 left arm	Puromycin	2.5 µg
	T2 right arm	Zeocin	2.5 µg

Screening of the colonies was performed by first extracting the genomic DNA of each colony using a direct PCR approach. 150-200 µl of D-PBS was added to each tube on the PCR strips. The strips were next centrifuged for 10 s to pellet the cells. The supernatant was removed carefully and 10 µl of Proteinase K reaction mix was added to each tube containing cells from one colony (Table 31 and Table 32). The strips were then loaded onto a PCR machine and subjected to conditions listed in Table 33 for genomic extraction. PCR of the genomic region flanking the TALEN 1 and 2 target sites was performed using PrimeSTAR GXL DNA polymerase (Takara, #R050A) and the reagents were set up as shown in Table 34. PCR cycling conditions are shown in Table 35 using primers listed in Table 36.

Table 31. 50X detergent mix formulation

Component	Volume
Tween-20	10 µl
Igepal CA-630	10 µl
dH_2O	180 µl

Table 32. Proteinase K reaction mix formulation

Component	Volume
50x detergents mix	0.2 µl
5x PrimeSTAR GXL PCR Buffer	2 µl
20 mg/mL Proteinase K	0.2 µl
dH_2O	7.6 µl

Table 33. Parameters for PCR cycling conditions for genomic DNA extraction

Temperature	Time
50°C	30 min
95°C	5 min
4°C	Indefinitely

Table 34. Volume of reagents in genomic PCR reaction

Component	Volume
dH_2O	13.4 µl
5x PrimeSTAR GXL PCR Buffer	4 µl
2.5 mM dNTP	1.6 µl
100 µM Primer Fwd	0.05 µl
100 µM Primer Rev	0.05 µl
Genomic DNA	0.5 µl
PrimeSTAR GXL DNA polymerase	0.4 µl

Table 35. Parameters for PCR to amplify TALEN 1 and 2 genomic regions

TALEN target site	Step	Temperature	Time	Number of cycles
1	Denaturation Annealing Extension	98°C 55°C 72°C	10 s 5 s 5 s	30
	Soak	4°C	∞	1
2	Denaturation Annealing Extension	98°C 62°C 72°C	10 s 5 s 5 s	30
	Soak	4°C	∞	1

Table 36. Primers used to amplify TALEN 1 and 2 genomic regions

TALEN target site		Primer sequence (5' → 3')
1	F	CTTTGAGAAGTCAGATCCCATTTGA
	R	CGCCTCCGCTGCCGTATGGAGGGCT
2	F	CGCCAGCAAGCTGCTGTGGTCCAGC
	R	TCCGCGCACCCGGACGAGAAAGTCC

The PCR products were next treated with ExoSAP-IT (Affymetrix, #78250) then digested with restriction enzymes AfeI (New England Biolabs, #R0652S) and PstI (New England Biolabs, #R0140S) for TALEN 1 and TALEN2 PCR products respectively for 1 hr at 37°C. The digested PCR products were analysed by agarose gel electrophoresis as a first pass screen for any mutations. PCR products of the colonies that showed a potential for the occurrence of mutations were next sent for sequencing as a second pass screen and confirmation for mutations using forward primers listed in Table 36.

2.4.2. Homologous recombination (HR) pathway

TALENs were introduced into H9 and FSPS13.B hPSC lines via electroporation as described in section 2.3.2. For electroporation, 2 μg of DNA for each TALEN arm targeting the TALEN 1 site and 1 μg of DNA for the donor plasmid was used (Table 37). 24 hr after electroporation, simultaneous antibiotic selection of 3.5 μg/ml blasticidin and 1 μg/ml puromycin was done for 24 hr. After which, simultaneous antibiotic selection of 3.5 μg/ml blasticidin and 2.5 μg/ml zeocin was done for the next 24 hours. Colonies formed approximately 7 days after the antibiotic selection, and they were individually picked using a pipette tip, re-plated and expanded. Genomic DNA was performed the same way as described in section 2.4.1. Successful HR was determined by PCR using PrimeSTAR GXL DNA polymerase (Takara, #R050A) with reaction mix listed in Table 38. The PCR cycling conditions are shown in Table 39 using forward and reverse primers for the TALEN 1 target site listed in Table 36. Colonies that showed positive HR via PCR were confirmed by sequencing.

Table 37. Electroporation of TALENs with donor plasmid into hPSCs

TALEN target site	TALEN construct name	Antibiotic resistance gene	Amount of DNA electroporated
1	T1 left arm	Blasticidin	2 μg
	T1 right arm	Zeocin	2 μg
	Donor plasmid	Puromycin	1 μg

Table 38. Volume of reagents in PCR reaction to check for HR

Component	Volume
dH_2O	13.4 μl
5x PrimeSTAR GXL PCR Buffer	4 μl
2.5 mM dNTP	1.6 μl
100 μM Primer Fwd	0.05 μl
100 μM Primer Rev	0.05 μl
Genomic DNA	50 ng
PrimeSTAR GXL DNA polymerase	0.4 μl

Table 39. Parameters for PCR cycling reaction to check for HR

TALEN target site	Step	Temperature	Time	Number of cycles
1	Initial denaturation	94°C	3 min	1
	Denaturation Annealing Extension	98°C 62°C 68°C	10 s 15 s 1 min	30
	Final extension	68°C	2 min	1
	Soak	4°C	∞	1

2.5. Western blot

2.5.1. Cell lysate preparation and normalisation

hPSCs or their differentiated progenitors were washed once in D-PBS and incubated in 0.5 ml of Accutase per well of a 6 well plate for 5 min at 37°C. The Accutase was neutralised by adding 1 ml of 5% FBS diluted in D-PBS per well and the cells were dissociated by gentle pipetting. The cells were washed twice with D-PBS and pelleted by centrifuging at 1,200 rpm. The pelleted cells were re-suspended in 50-200 µl of Lysis Buffer (Table 40) containing freshly added inhibitors cOmplete Protease Inhibitor Cocktail (Roche, #11697498001), Sodium Fluoride (NaF; New England Biolabs, #P0759), Sodium Vanadate (Na_3VO_4; New England Biolabs, #P0758) according to Table 41. One tablet of the cOmplete Protease Inhibitor Cocktail was dissolved in 1 ml of Lysis Buffer to make up a 25x stock solution. The cell lysates were kept on ice for at least 15 min, vortexed at maximum speed for 15 s then centrifuged for 30 min at 15,000 g at 4°C. The supernatants were collected and protein concentrations were determined by Bradford assay (Protein Assay Dye Reagent Concentrate, Bio-Rad) according to the manufacturer's protocol. The protein concentrations of the cell lysates were normalised to 10 µg of protein for probing with *GATA6* and *GATA4* and 1 µg for probing with alpha-tubulin. The normalised cell lysates were heat denatured at 98°C in the presence of Laemmli Sample Buffer (Bio-Rad) and β-mercaptoethanol for 5 min, then subjected to SDS-PAGE electrophoresis.

Table 40. Lysis Buffer formulation

Component	Working concentration
1M Tris-Cl pH 7.5 (Cambridge Bioscience, #600201)	50mM
5M NaCl (Ambion, #AM9759)	150mM
Triton X-100 (Sigma-Aldrich, #T8787)	1%
Glycerol (Sigma-Aldrich, #49781)	10%
Deocycholate (Sigma-Aldrich, #D6750)	0.1%
β-glycerophosphate (Sigma-Aldrich, #G5422)	25mM

Table 41. Lysis Buffer with protease inhibitors formulation

Component	Volume
25x Complete solution	40 µl
Lysis Buffer	930 µl
50 mM NaF	20 µl
100 mM Na_3VO_4	10 µl

2.5.2. SDS-page, blotting and blocking

Electrophoresis of NuPAGE Novex 4-12% Bis-Tris Protein Gels was performed using the XCell SureLock Mini-Cell (Invitrogen). 1x SDS Running Buffer was prepared by adding 50 ml of 20x NuPAGE MOPS SDS Running Buffer to 950 ml deionised water. Precision Plus protein kaleidoscope standards (Bio-Rad, #161-0375) protein ladder was used. Electrophoresis was carried out at a constant volt of 120 V until the ladders were completely separated. The separated proteins were next transferred from the gel onto Immun-Blot PVDF membrane (Bio-Rad, #162-0177) using Mini Trans-Blot Cell (Bio-Rad) at 25 V overnight at 4°C.

2.5.3. Antibody incubation and detection

The membranes were incubated shaking in 5% Blotting-Grade Blocker (Bio-Rad, #170-6404) diluted in 0.1% Triton X-100 in D-PBS (PBST) for 1 hr at room temperature. The blocking solution was removed and primary antibodies (Table 42) diluted in PBST were added to the membranes and incubated shaking for 2 hr at room temperature. Membranes were then subjected to three 10 min washes in PBST and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies in 5% Blotting-Grade Blocker in PBST for 1 hr at room temperature while shaking. Unbound antibodies were removed by three 10 min washes while shaking in PBST. Proteins were detected via chemiluminescence using SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific, #PI34095) and finally developed using Amersham Hyperfilm ECL (GE Healthcare).

Table 42. Antibodies used in western blot experiments

Primary antibody	**Dilution ratio**	**Duration**	**Expected molecular weight**
Rabbit anti-human GATA6 (N-terminus; Cell Signaling, #5851)	1:2000	2 hr	Long isoform: 64 kDa Short isoform: 52 kDa
Rabbit anti-human GATA6 (C-terminus; Cell Signalling, #4253)	1:2000	2 hr	Short isoform: 52 kDa
Rabbit anti-human GATA4 (Cell Signalling, #36966)	1:2000	2 hr	55 kDa
Mouse anti-alpha-Tubulin (Sigma-Aldrich, #T6199)	1:5000	1 hr	50 kDa
Secondary antibody	**Dilution ratio**	**Duration**	
Anti-Rabbit IgG-Peroxidase antibody produced in goat (Sigma-Aldrich, #A6154)	1:10,000	1 hr	-
Anti-Mouse IgG-Peroxidase antibody produced in goat (Sigma-Aldrich, #A5278)	1:10,000	1 hr	-

2.6. Immunocytochemistry (ICC)

2.6.1. Fixation and blocking

Cells in 12 well plates were fixed by aspirating the culture media then immediately adding 500 µl of 4% paraformaldehyde (PFA; VWR, #43368.9M) solution diluted in D-PBS per well and incubating for 20 min at 4°C. They were then washed thrice in D-PBS. Cells were next incubated in 500 µl of PBST (0.1% Triton X-100 in D-PBS) containing 10% donkey serum (AbD Serotec, #C06SB) per well for 20 min at room temperature for blocking.

2.6.2. Antibody incubation and detection

Cells in 12 well plates were then incubated overnight at 4°C with 300 µl of primary antibodies (Table 43) diluted in PBST containing 1% donkey serum. Cells were then washed thrice with PBST to remove unbound primary antibodies and thereafter incubated with 300 µl of fluorescence-dye-conjugated secondary antibodies (Table 43) diluted in PBST containing 1% donkey serum in for 1 hr at room temperature. Unbound antibodies were removed by three 5 min washes in D-PBS. 4’,6-Diamidino-2-phenylindole dihydrochloride (DAPI; Sigma-Aldrich, #D-8417) at a dilution of 1:1000 was added to the first wash.

Table 43. Antibodies used in immunocytochemistry experiments

Primary antibody	**Dilution ratio**	**Duration**
Goat anti-human Nanog (R&D, #AF1997)	1:100	Overnight
Goat anti-human Sox2 (R&D, #AF2018)	1:100	Overnight
Goat anti-human Oct4 (Santa Cruz, #sc-8628)	1:100	Overnight
Rabbit anti-human Eomes (Abcam, #Ab23345)	1:100	Overnight
Rabbit anti-human GATA6 (Cell Signaling, #5851)	1:200	Overnight
Mouse anti-human GATA4 (Santa Cruz, #sc25310)	1:100	Overnight
Goat anti-human Sox17 (R&D, #AF1924)	1:200	Overnight
Goat anti-human FoxA2 (R&D, #AF2400)	1:100	Overnight
Mouse anti-human Hex (Abcam, #Ab117864)	1:100	Overnight
Mouse anti-human CDX2 (CDX-88; Abcam, #Ab86949)	1:100	Overnight
Goat anti-human HNF1B C-20 (Santa Cruz, #sc7411)	1:100	Overnight
Rabbit anti-human HNF4A H-171 (Santa Cruz, #sc8987)	1:100	Overnight
Rabbit anti-human HNF6 H-100 (Santa Cruz, #sc13050)	1:100	Overnight
Goat anti-human PDX1 (R&D, #AF2419)	1:100	Overnight
Rabbit anti-human Sox9 H-90 (Santa Cruz, #sc20095)	1:100	Overnight
Sheep anti-human NGN3 (R&D, #AF3444)	1:100	Overnight
Mouse anti-human C-Peptide (Acris Antibodies, #BM270S)	1:100	Overnight
Goat anti-human Glucagon G-17 (Santa Cruz, #sc7780)	1:100	Overnight
Rabbit anti-human Somatostatin (Daka, #A0566)	1:200	Overnight
Secondary antibody	**Dilution ratio**	**Duration**
Alexa Fluor 568 Donkey Anti-Goat IgG (H+L) (Invitrogen, #A11057)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Mouse IgG (H+L) (Invitrogen, #A10037)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Rabbit IgG (H+L) (Invitrogen, #A10042)	1:1000	1 hr
Alexa Fluor 568 Donkey Anti-Sheep IgG (H+L) (Invitrogen, #A21099)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Goat IgG (H+L) (Invitrogen, #A11055)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A21202)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A21206)	1:1000	1 hr
Alexa Fluor 488 Donkey anti-Sheep IgG (H+L) (Invitrogen, #A11015)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Goat IgG (H+L) (Invitrogen, #A21447)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A31571)	1:1000	1 hr
Alexa Fluor 647 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A31573)	1:1000	1 hr

2.7. Fluorescence activated cell sorting (FACS) analysis

2.7.1. Cell preparation

Cells in 12 well plates were washed twice in D-PBS and incubated in 0.3 ml of Accutase per well for 5 min at 37°C. The Accutase was neutralised by adding 0.6 ml of 5% FBS diluted in D-PBS and the cells were dissociated by gentle pipetting. Cells were re-suspended in D-PBS at approximately 0.1-1 x 10^6 cells/ml and washed twice with D-PBS. They were then pelleted and fixed by re-suspending in 500 µl of 4% PFA solution diluted in D-PBS per well and incubating at for 20 min at 4°C, then washed twice in D-PBS.

2.7.2. Antibody incubation and detection

Next, for all primary antibodies except CXCR4, cells were permeabilised in 500 µl of D-PBS containing 1% Saponin (Sigma-Aldrich, #47036-50G-F) for 30 min at room temperature. Cells were then incubated for 2 hr at room temperature with primary antibody (Table 44) diluted in 100 µl of Staining Solution (1% Saponin and 5% FBS in D-PBS). After which, they were washed three times with 1 ml of Staining Solution per wash and incubated with secondary antibodies (Table 44) diluted in 100 µl of Staining Solution for 30 min at room temperature. Unbound antibody was then removed by three washes in 1 ml of Staining Solution per wash and cells were re-suspended in 200 µl of 2% FBS diluted in D-PBS prior to analysis.

For CXCR4 staining, cells were fixed in 4% PFA and washed as described above. Thereafter, primary antibody (Table 44) diluted in 100 µl of 5% FBS in D-PBS was added to the cells and incubated for 1 hr at room temperature. Unbound antibody was then removed by three washes of 1ml 2% FBS in D-PBS per wash. Cells were then re-suspended in 200 µl of 2% FBS in PBS prior to analysis.

Analyses were performed using a BD LRSFortessa cell analyser (BD Biosciences). All flow cytometry experiments were gated using unstained cells. On all flow cytometry plots, the undifferentiated population is shown in blue. All gates shown on scatterplots were set according to the undifferentiated population control.

Table 44. Antibodies used in FACS experiments

Primary antibody	**Dilution ratio**	**Duration**
Goat anti-human Sox17 (R&D, #AF1924)	1:20	2 hr
Rabbit anti-human GATA6 (Cell Signaling, #5851)	1:20	2 hr
Goat anti-human PDX1 (R&D, #AF2419)	1:20	2 hr
Sheep anti-human NGN3 (R&D, #AF3444)	1:20	2 hr
Mouse anti-human C-Peptide (Acris Antibodies, #BM270S)	1:100	2 hr
Goat anti-human Glucagon G-17 (Santa Cruz, #sc7780)	1:20	2 hr
Rabbit anti-human Somatostatin (Daka, #A0566)	1:200	2 hr
Secondary antibody	**Dilution ratio**	**Duration**
Alexa Fluor 568 Donkey Anti-Goat IgG (H+L) (Invitrogen, #A11057)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Mouse IgG (H+L) (Invitrogen, #A10037)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Rabbit IgG (H+L) (Invitrogen, #A10042)	1:1000	30 min
Alexa Fluor 568 Donkey Anti-Sheep IgG (H+L) (Invitrogen, #A21099)	1:1000	30 min
Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) (Invitrogen, #A31571)	1:1000	30 min
Alexa Fluor 488 Donkey anti-Rabbit IgG (H+L) (Invitrogen, #A21206)	1:1000	30 min
Conjugated primary and secondary antibody	**Dilution ratio**	**Duration**
Anti-Human CD184 (CXCR4) PE (eBioscience, #12-9999-41)	1:50	1 hr

2.8. Quantitative RT-PCR (qRT-PCR)

2.8.1. Total RNA isolation

Cells were grown in 12 well plates for total RNA isolation. 3 wells were harvested per sample to obtain technical triplicates. RNeasy Mini Kit together with the Qiacube (Qiagen) was used for total RNA extraction. Cell culture media was aspirated and the cells were washed once with D-PBS. The D-PBS was completely aspirated and cells were lysed directly in the 12 well plates by adding 350 µl of Buffer RLT. Cell lysates were transferred to 2 ml tubes and were either frozen at -80°C or used immediately with the Qiacube for RNA extraction. Each sample was treated with RNase-Free DNase (Qiagen) to avoid DNA contamination. RNA was eluted in a volume of 30 µl. RNA was either frozen at -80°C or immediately taken to the next step of first strand cDNA synthesis. If RNA samples were frozen, they were thawed on ice to prevent degradation.

2.8.2. First strand cDNA synthesis

500 ng of RNA samples were made up to a total volume of 11.875 µl with nuclease free water. The following components were then added to a nuclease-free 96 well plate (Table 45).

Table 45. Reagents to denature RNA and primer

Component	Volume
500 ng of total RNA	11.875 µl
Random primer (Promega, #C1181)	0.5 µl
dNTP (Promega, #U1511)	1 µl

The plate was centrifuged briefly to ensure that the reagents were at the bottom of the tube. The plate was incubated in a PCR machine for 5 min at 65°C then quickly chilled on ice. The plate was again centrifuged briefly and the reagents listed in Table 46 were prepared as a master mix then added to the samples. The final volume of each sample was 20 µl.

Table 46. Reagents for reverse transcription of RNA

Component	Volume
5x First-strand buffer	4 µl
0.1 M DTT	2 µl
RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, #10777019)	0.5 µl
SuperScript® II Reverse Transcriptase (Invitrogen, #18064014)	0.125 µl

The plate was again centrifuged briefly to ensure that all reagents were at the bottom of the tube. The plate was then incubated in a PCR machine programmed at 10 min at 25°C for the primer annealing step, 50 min at 42°C for the extension step, and finally 15 min at 70°C for the inactivation of the enzyme. The resulting cDNA was diluted to a final volume of 600 µl with nuclease-free water prior to use for qRT-PCR.

2.8.3. qRT-PCR

qRT-PCR master mix was prepared using Sensi Mix Sybr Low Rox Kit (Bioline, #QT625-20). The reaction was prepared according to Table 47.

Table 47. Reagents for qRT-PCR

Component	Volume
cDNA template	5 µl
Forward primer (5 µM)	0.6 µl
Reverse primer (5 µM)	0.6 µl
Sensi Mix (2x)	7.5 µl
Nuclease free water	1.3 µl

qRT-PCR reactions were performed using Mx3005P system (Stratagene) with cycling conditions as listed in Table 48. Samples were run in technical triplicates and normalised to *PBGD*. Gene-specific primers are listed in Table 49. Data represents the mean of one experiment which is representative of three independent experiments and error bars indicate standard deviation of triplicates.

Table 48. Parameters for qRT-PCR cycling conditions

Stage	Cycles	Temperature	Time
1	1	95°C	10 min
2	40	95°C	30 s
		60°C	30 s
		72°C	30 s
3	1	95°C	1 min
		55°C	30 s (↑ 1°C/30 s)
		95°C	30 s

Table 49. Primers used in qRT-PCR

Gene		**Primer sequence (5' → 3')**
OCT4	F	AGTGAGAGGCAACCTGGAGA
	R	ACACTCGGACCACATCCTTC
NANOG	F	CATGAGTGTGGATCCAGCTTG
	R	CCTGAATAAGCAGATCCATGG
SOX2	F	TGGACAGTTACGCGCACAT
	R	CGAGTAGGACATGCTGTAGGT
GSC	F	GAGGAGAAAGTGGAGGTCTGGTT
	R	CTCTGATGAGGACCGCTTCTG
BRACHURY	F	TGCTTCCCTGAGACCCAGTT
	R	GATCACTTCTTTCCTTTGCATCAAG
EOMESODERMIN	F	ATCATTACGAAACAGGGCAGGC
	R	CGGGGTTGGTATTTGTGTAAGG
GATA4	F	TCCCTCTTCCCTCCTCAAAT
	R	TCAGCGTGTAAAGGCATCTG
GATA6	F	TGTGCAATGCTTGTGGACTC
	R	AGTTGGAGTCATGGGAATGG
SOX17	F	CGCACGGAATTTGAACAGTA
	R	GGATCAGGGACCTGTCACAC
CXCR4	F	CACCGCATCTGGAGAACCA
	R	GCCCATTTCCTCGGTGTAGTT
FOXA2	F	GGGAGCGGTGAAGATGGA
	R	TCATGTTGCTCACGGAGGAGTA
GCG	F	AAGCATTTACTTTGTGGCTGGATT
	R	TGATCTGGATTTCTCCTCTGTGTCT
HLXB9	F	CACCGCGGGCATGATC
	R	ACTTCCCCAGGAGGTTCGA
HNF1B	F	TCACAGATACCAGCAGCATCAGT
	R	GGGCATCACCAGGCTTGTA
HNF4A	F	CATGGCCAAGATTGACAACCT
	R	TTCCCATATGTTCCTGCATCAG

HNF6	F	GTGTTGCCTCTATCCTTCCCAT
	R	CGCTCCGCTTAGCAGCAT
INSULIN	F	GAAGCGTGGCATTGTGGAAC
	R	GCTGCGTCTAGTTGCAGTAGT
NGN3	F	GCTCATCGCTCTCTATTCTTTTGC
	R	GGTTGAGGCGTCATCCTTTCT
NKX6.1	F	GGCCTGTACCCCTCATCAAG
	R	TCCGGAAAAAGTGGGTCTCG
PDX1	F	GATTGGCGTTGTTTGTGGCT
	R	GCCGGCTTCTCTAAACAGGT
SST	F	CCCCAGACTCCGTCAGTTTC
	R	TCCGTCTGGTTGGGTTCAG
PBGD	F	GGAGCCATGTCTGGTAACGG
	R	CCACGCGAATCACTCTCATCT
SOX9	-	Quantitect primers (QT00001498)

2.9. RNA-sequencing

2.9.1. Illumina sequencing

Library preparation and deep sequencing were performed at the Wellcome Trust Sanger Institute (Hinxton, UK). RNA-sequencing was run on Illumina Hiseq v3 with read length 75bp and paired-ends, and a library fragment size of 100-1000bp using a multi-plex strategy. Samples were run in biological triplicates.

2.9.2. RNA enrichment analysis

RNA-sequencing analyses were performed with partial help from Dr Pedro Madrigal. Tophat v2 (Kim et al., 2013), provided by Ensembl release 76 annotations, was used to align short reads from each sample to the reference human genome assembly (GRCh38/hg20). Feature counts was used to summarize paired-end reads and count fragments with parameters '-p -T 8 -t exon -g gene_id' (Liao et al., 2014). DESeq2 Bioconductor package was used to search for significant differences between samples, requiring at least a 2 fold expression change and adjusted p-value less than 0.01 (Love et al., 2014). R package edgeR function 'rpkm' was used with default parameters to normalize count gene expression (Robinson et al., 2010). Raw bedGraphs were normalized per million mapped reads in the library per library size in all ChIP-seq and RNA-seq samples. Genome browser panels were generated using IGV (Thorvaldsdóttir et al., 2013).

2.9.3. Functional annotations

Gene Ontology (GO) analyses were performed using Amigo2 separately for up- and down- regulated differentially expressed genes (Carbon et al., 2009). Spearman's correlation values were calculated for FPKM values for gene expressed at more than 5 FPKM in at least one sample.

2.10. Chromatin Immunoprecipitation (ChIP)

2.10.1. Cross-linking of protein and DNA

Co-binding of DNA to DNA-binding proteins was determined by ChIP against *GATA6* (Cell Signaling, #5851) on approximately 1 x 10^7 cells per antibody or control sample. hPSCs grown on 10 cm dishes were differentiated and harvested either at the endoderm (D3) or pancreatic progenitor (D12) stage. 1 x 10 cm dish of cells harvested at D3 was used for one antibody immunoprecipitation (IP). 1 x 10 cm dish of cells harvested at D12 was used for four antibody IPs. Cells were cross-linked by adding 312.5 µl of 16% formaldehyde (ThermoFisher UK, #11586711) to 5 ml of media to make a final concentration of 1%. The cells were incubated rocking for 10 min at room temperature to allow cross-linking of protein-DNA complexes.

The reaction was quenched by adding 312.5 µl of 2 M glycine (Millipore, #357002) to make a final concentration of 0.125 M and incubated for 5 min with rocking. Thereafter, the media was aspirated and cells were washed twice with 5 ml of ice-cold PBS. The cells were detached by scraping into 3 ml of ice-cold PBS containing freshly-added protease inhibitors (10 µl/ml of 5 mg/ml phenylmethylsulfonylfluoride (PMSF; Sigma-Aldrich, #93482), 10 µl/ml of 1 M Sodium Butyrate (Sigma-Aldrich, #303410) and 1 µl/ml of 1 mg/ml Leupeptin (Roche, #11017101001)) and pooled into 15 ml tubes, each tube containing approximately 2 x 10^7 cells, or 2 IPs. The cells were then centrifuged for 5 min at 1,200 rpm at 4°C to pellet.

For all subsequent steps, the samples were kept on ice. For all subsequent buffers used, the aforementioned protease inhibitors were added freshly to the buffers before use. The pelleted cells were subsequently re-suspended in 2 ml of ice-cold Cell Lysis Buffer (10 mM Tris-Cl pH 8.0, 10 mM NaCl and 0.2% NP-40) per 15 ml tube and incubated on ice for 10 min. The cells were then centrifuged for 5 min at 1,800 rpm at 4°C. The supernatant was discarded and the pellet was gently re-suspended in 1.25 ml of ice-cold Nuclear Lysis Buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA and 1% SDS) per 15 ml tube and incubated on ice for 10 min. 0.75 ml of ice-

cold IP Dilution Buffer (20 mM Tris-Cl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.01% SDS, 1% Triton X-100) was then added per 15 ml tube.

The chromatin was then transferred into 15 ml Diagenode sonication tubes containing sonication beads (Diagenode, #C01020031) that were pre-washed twice with 10 ml D-PBS each time and once with 10 ml IP Dilution Buffer. Next, chromatin was sonicated in Diagenode Biorupter Pico for 10 cycles of 30s on/45s off. The sonicated chromatin were transferred to 1.5 ml tubes and centrifuged at 14,000 rpm for 10 min at 4°C to pellet debris. The pellet was discarded, and two 15 ml tubes, or 4 IPs, worth of sonicated chromatin was pooled into a fresh 15 ml tube. 3.5 ml of IP Dilution Buffer was added and mixed gently. The cross-linked DNA was pre-cleared by incubating with rotation 10 µg of rabbit IgG (Sigma-Aldrich, #I5006) for 1 hr at 4°C, followed by incubating with rotation 100 µl of Protein G agarose beads (50% v/v; Roche, #11243233001) pre-washed twice with D-PBS for 1 hr at 4°C. The samples were then centrifuged for 3 min at 3,000 rpm at 4°C and the supernatant was transferred to a fresh 15 ml tube. An aliquot of 300 µl for Input sample was transferred to a fresh 1.5 ml Eppendorf tube and stored at 4°C.

2.10.2. Immunoprecipitation of protein-DNA complex

The supernatant was split equally into four 15 ml tubes, each representing one sample i.e. 1 x 10^7 cells worth of material per tube. 10 µg of *GATA6* antibody or rabbit IgG control was added per tube, and samples were incubated rotating overnight at 4°C. Antibody-bound chromatin was then collected using 60 µl of Protein G agarose beads (50% v/v) pre-washed twice with D-PBS for each tube by incubating with rotation for 1 hr at 4°C. Thereafter, the tubes were centrifuged for 3 min at 3,000 rpm at 4°C. The supernatant was discarded and the pellet containing the protein-DNA complexes bound onto the protein G agarose beads were kept.

2.10.3. DNA extraction

500 µl of IP Wash Buffer 1 (20 mM Tris-Cl pH 8.0, 2 mM EDTA, 50 mM NaCl, 0.1% SDS and 1% Triton X-100) was added to each tube and the tubes were vortexed. The samples were transferred to fresh 1.5 ml Eppendorf tubes and were then centrifuged at 14,000 rpm for 1 min at 4°C. The samples were washed with IP Wash Buffer 1 once more by removing the supernatant, adding 500 µl of IP Wash Buffer 1, vortexing, and centrifuging at 14,000 rpm for 1 min at 4°C. 500 µl of IP Wash Buffer 2 (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% NP-40 and 1% Sodium deoxycholic acid) was then added to each tube after removing the supernatant and the tubes were vortexed and centrifuged at 14,000 rpm for 1 min at 4°C. The wash was repeated once more. Thereafter, the samples were washed twice with 500 µl of TE Buffer (10mM Tris-Cl pH 8.0, 1mM EDTA) according to the procedures above. After the supernatant was aspirated, the chromatin was eluted from the Protein G beads by washing twice with 150 µl of Elution Buffer (100 mM $NaHCO_3$ and 1% SDS), vortexing and centrifuging at 14,000 rpm for 1 min at 4°C. The supernatants were collected and pooled in fresh 1.5 ml Eppendorf tubes.

ChIP and Input DNA cross-links were reversed and RNA degraded by adding 1 µl of 1 mg/ml RNase A and 18 µl of 5M NaCl and incubating at 67°C in a heat block with shaking at 1,300 rpm overnight. Protein was degraded by adding 3 µl of 20 mg/ml Proteinase K and incubating for 3 hrs at 45°C in a heat block with shaking at 1,300 rpm. DNA was extracted using 300 µl of phenol/chloroform wash with vortexing then centrifuged at 14,000 rpm for 5 min at room temperature. The aqueous layer containing pulled down genomic DNA was transferred to fresh 1.5 ml Eppendorf tubes. 30 µl of 3M NaAc pH 5.2 (Ambion, #AM9740), 30 µg glycoblue (Ambion, #AM9516) and 750 µl of 100% ethanol were added to the samples, which were then vortexed. The samples were next incubated for at least 30 min at -80°C to precipitate the DNA. Precipitated DNA was pelleted by centrifuging at 14,000 rpm for 30 min at 4°C. The DNA pellet was then washed with ice-cold 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C. The ethanol was removed and the pellet

air dried. 70 µl of deionised water was added to Input samples whereas 30 µl of deionised water was added to ChIP samples.

2.10.4. Bioanalyser

Chromatin fragments after sonication were determined by a Bioanalyser (Agilent 2100 Bioanalyzer) to ensure that fragmented chromatins are within the range of 150-300 base pairs in size.

A 10 µl aliquot was taken directly from freshly sonicated samples and transferred to a clean 1.5 ml Eppendorf tube. The tubes were centrifuged at 14,000 rpm for 10 min at 4°C to remove any insoluable material. The supernatant was transferred to a fresh 1.5 ml Eppendorf tube. Thereafter, fast reversal of the cross-links was performed by adding 74 µl of nuclease free water, 4 µl of 5M NaCl, 8 µl of 20 mg/ml Proteinase K and 4 µl of 10 mg/ml RNase A to the aliquots. The aliquots were incubated at 65°C for 2 hr in a heat block with shaking at 1,300 rpm. DNA was purified using QIAquick PCR Purification Kit (Qiagen) and DNA was eluted in 20 µl of elution buffer. DNA was diluted 1:10 in nuclease free water and 1 µl was used for analysis using High Sensitivity DNA Kit (Agilent, #5067-4626) according to the manufacturer's protocol.

2.10.5. qPCR detection

Pulldown enrichment was validated by qPCR using KAPA SYBR FAST Master Mix (2X) ROX Low qPCR Kit (KAPA Biosystems). Gene-specific primers kindly provided by our collaborator, Jorge Ferrer's group, are listed in Table 50. Results were expressed as normalised values against a negative control region and fold change compared between antibody pulldown and IgG control.

Table 50. Primers used in ChIP qPCR

Primer name		**Primer sequence (5' → 3')**
hGATA6 positive region	F	CATGGAGACAGCAACAGTCC
hGATA6 positive region	R	ACCGCCCGGTTATCTTATTG
hPDX1 positive region	F	TTTCTCGCTGCCCTTTACTC
hPDX1 positive region	R	GTGCTGTGGCTCAACTCTGA
NR0B2 positive for hGATA6 and hPDX1	F	GCTGCCCCTTATCAGATGAC
NR0B2 positive for hGATA6 and hPDX1	R	CTGGCTTAGCAAAAGCCCTA
Amy2A negative control	F	TGCTGCCAGAACCTAAGAAAA
Amy2A negative control	R	TTGAGGGCAAACTGTTTATTCA
Nanog negative control	F	AAAGCTTGCCTTGCTTTGAA
Nanog negative control	R	AGTCTCCGTGTGAGGCATCT

2.11. ChIP-sequencing

2.11.1. Illumina sequencing

Library preparation and deep sequencing were performed at the Wellcome Trust Sanger Institute (Hinxton, UK). RNA-sequencing was run on Illumina Hiseq v4 with read length 75bp and paired-ends, and a library fragment size of 100-1000bp using a multi-plex strategy. Samples were run in biological duplicates.

2.11.2. Bioinformatics analyses

ChIP-sequencing analyses were performed with partial help from Dr. Pedro Madrigal and Dr. Denil Simon Lieven Imanuel Johannes. Short-insert paired-end reads were aligned to the reference human genome assembly (GRCh38/hg18) using the Burrows-Wheeler Aligner (BWA) 0.5.10 (Li and Durbin, 2009) with -q 15 and default for the rest of parameters. Reproducibility between replicates was first assessed using the Pearson Correlation Coefficient (PCC) for the two biological replicates, using the genome-wide normalized read (extended to 300 bp) count distribution on a single nucleotide resolution. For this, the UCSC tool bigwigCorrelate provided in http://hgdownload.cse.ucsc.edu/admin/exe/. PCC is equal to 0.949326 was used.

Peak calling was performed using MACS version 2.0.10 (Zhang et al., 2008), allowing a p-value cut-off of p-value ≤ 1e-3, and default for the rest of parameters. Relaxed thresholds are suggested in order to enable the correct computation of IDR values (Landt et al., 2012). Following the recommendations for the analysis of self-consistency and reproducibility between replicates, the negative control samples (IgG and input DNA) were combined into one single control; code for IDR analysis was downloaded from https://sites.google.com/site/anshulkundaje/projects/idr (Li et al., 2011). This is also beneficial as control samples with substantially higher number of reads are recommended for peak calling (Bailey et al., 2013).

To estimate the Irreproducible Discovery Rate (IDR) between replicates, top 35k peaks for each biological replicate were submitted for IDR analysis. For IDR computation using MACS results, we used p-values rather than q-values as suggested in https://sites.google.com/site/anshulkundaje/projects/idr (Li et al., 2011). The number of peaks found passing a threshold of IDR ≤ 5% (12,107) was selected as a conservative estimated number of candidate transcription factor binding sites. Autosomal and sex chromosomes were also excluded.

Co-localization plots of the transcription factors *GATA6*, *EOMES* and *SMAD2/3* ChIP-seq, was generated with deepTools (Ramirez et al., 2014). The input data was obtained by combining my ChIP data of H9 cells at day 3 (*GATA6*) with previously published *EOMES* (Teo et al., 2011) and *SMAD2/3* ChIP data (Brown et al., 2011). To make the results more comparable, the 3 data sets were re-mapped with STAR v2.5.1a (Dobin et al., 2013) (BWA failed on short single end SMAD reads) and processed with MACS version 2.0.10 and IDR as described earlier. The resulting peak files (bed format) were used as input for deepTools. The mapped read files (bam format) were pre-processed with deepTools' "bamCompare" function (bin size = 50, assumed genome size = 2451960000 bp, ignoring chromosomes X and Y for normalization and extending single end reads by 250bp).

2.12. Statistical analyses

Unpaired two-tailed Student's t-test were used to assess statistical significance. Statistical analysis was done by GraphPad Prism software (Version 6.0 for Windows. GraphPad Software, San Diego, CA). p value < 0.05 was considered as statistically significant, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. Data are presented as mean ± s.d. as indicated in the figure legends. All sample numbers listed indicate the number of biological replicates employed in each experiment. For experiments showing data of one experiment that is representative of three independent experiments, this was done due to variations between different differentiation experiments; combining the experiments resulted in inaccurately large standard deviations that were not representative of the results.