Diabetologia (2012) 55:694–706
DOI 10.1007/s00125-011-2379-y

ARTICLE

INS$^{GFP/w}$ human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells

S. J. Micallef · X. Li · J. V. Schiesser · C. E. Hirst · Q. C. Yu · S. M. Lim · M. C. Nostro · D. A. Elliott · F. Sarangi · L. C. Harrison · G. Keller · A. G. Elefanty · E. G. Stanley

Received: 27 July 2011 / Accepted: 20 October 2011 / Published online: 26 November 2011

Abstract
Aims/hypothesis We aimed to generate human embryonic stem cell (hESC) reporter lines that would facilitate the characterisation of insulin-producing (INS$^+$) cells derived in vitro.
Methods Homologous recombination was used to insert sequences encoding green fluorescent protein (GFP) into the *INS* locus, to create reporter cell lines enabling the prospective isolation of viable INS$^+$ cells.
Results Differentiation of *INS*$^{GFP/w}$ hESCs using published protocols demonstrated that all GFP$^+$ cells co-produced insulin, confirming the fidelity of the reporter gene. INS-GFP$^+$ cells often co-produced glucagon and somatostatin, confirming conclusions from previous studies that early hESC-derived insulin-producing cells were polyhormonal. *INS*$^{GFP/w}$ hESCs were used to develop a 96-well format spin embryoid body (EB) differentiation protocol that used the recombinant protein-based, fully defined medium, APEL. Like INS-GFP$^+$ cells generated with other methods, those derived using the spin EB protocol expressed a suite of pancreatic-related transcription factor genes including *ISL1*, *PAX6* and *NKX2.2*. However, in contrast with previous methods, the spin EB protocol yielded INS-GFP$^+$ cells that also co-expressed the beta cell transcription factor gene, *NKX6.1*, and comprised a substantial proportion of monohormonal INS$^+$ cells.
Conclusions/interpretation INS$^{GFP/w}$ hESCs are a valuable tool for investigating the nature of early INS$^+$ progenitors in beta cell ontogeny and will facilitate the development of novel protocols for generating INS$^+$ cells from differentiating hESCs.

Keywords Diabetes · Gene targeting · GFP · Human embryonic stem cells · Insulin

Abbreviations

BMP4	Bone morphogenetic protein 4
BrdU	Bromodeoxyuridine
EB	Embryoid body
GFP	Green fluorescent protein
FGF2	Fibroblast growth factor 2
hESC	Human embryonic stem cell
HGF	Hepatocyte growth factor
INS	Insulin
IPA	Insulin-positive aggregates
ISL	ISL LIM homeobox
KGF	Fibroblast growth factor 7
NKX2-2	NK2 homeobox 2
NKX6.1	NK6 homeobox 1
PAX6	Paired box 6

S. J. Micallef and X. Li contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-011-2379-y) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

S. J. Micallef · X. Li · J. V. Schiesser · C. E. Hirst · Q. C. Yu · S. M. Lim · D. A. Elliott · A. G. Elefanty · E. G. Stanley (✉)
Monash Immunology and Stem Cell Laboratories (MISCL), Level 3, Building 75, STRIP1, West Ring Road, Monash University, Clayton, Victoria 3800, Australia
e-mail: ed.stanley@monash.edu

M. C. Nostro · F. Sarangi · G. Keller
McEwen Centre for Regenerative Medicine, University Health Network, Toronto, ON, Canada

L. C. Harrison
The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

PDX1	Pancreatic and duodenal homeobox 1
RA	Retinoic acid
ROCK	Rho associated kinase

Introduction

Type 1 diabetes is an autoimmune disease characterised by destruction of beta cells in the pancreas, deficient insulin production, and persistent high levels of blood glucose. Treatment with exogenous insulin, although life-saving, does not restore physiological control of blood glucose, leaving people with type 1 diabetes at risk of long-term complications. Control can be improved by islet transplantation (reviewed by Speight et al. [1]), but this treatment option will always be limited by the scarcity of cadaveric donor tissue.

Beta cells derived from the differentiation of human embryonic stem cells (hESCs) in vitro potentially represent an inexhaustible source of insulin-producing cells for the treatment of type 1 diabetes. Several laboratories have demonstrated that hESC-derived endocrine cells can regulate blood glucose in a diabetic mouse model, providing proof of principle for future clinical application (for example, see studies by Kroon et al. [2] and Jiang et al. [3] and a review by van Hoof et al. [4]). However, while attempts to generate INS^+ cells from pluripotent stem cells have been encouraging, the biology of this process remains poorly understood. In this light, better tools and reagents to facilitate the understanding of beta cell development are required.

We describe the generation and characterisation of two independently derived hESC lines in which sequences encoding green fluorescent protein (GFP) have been targeted to the insulin locus ($INS^{GFP/w}$ hESCs). We demonstrate the utility of these lines by characterising the transcriptional signature of hESC-derived insulin-producing (INS^+) cells generated using established differentiation protocols. Analysis of these data in conjunction with immunofluorescence studies confirms that such cells display an immature phenotype, with the majority of INS^+ cells also producing glucagon. We used $INS^{GFP/w}$ hESCs to develop a novel 96-well format spin embryoid body (EB) differentiation protocol for the differentiation of hESCs to INS^+ pancreatic endoderm. This method is based on a protocol originally developed for the differentiation of hESCs to mesodermal populations [5] and uses a defined wholly recombinant protein-based medium (APEL) [6]. Characterisation of INS^+ cells generated with this platform reveals that, unlike INS^+ cells derived with previous methods, a substantial proportion also produce the beta cell-associated marker, NK6 homeobox 1 (NKX6.1), suggesting that the EB environment is conducive to ongoing differentiation. $INS^{GFP/w}$ hESCs are therefore a valuable tool for investigating and refining the generation of INS^+ cells from pluripotent stem cells in vitro.

Methods

Generation and identification of targeted $INS^{GFP/W}$ hESCs The INS-targeting vector comprised a 10.7 kb 5′ homology arm, GFP coding sequences, a loxP flanked phosphoglycerol kinase (PGK)-promoter-neomycin resistance cassette and 2.9 kb 3′ homology arm. The 5′ homology arm was derived from a bacterial artificial chromosome (RP11 889I17) encompassing the human insulin locus using ET cloning as described previously [7]. The 3′ homology arm was derived by PCR using the same bacterial artificial chromosome clone as a template. The vector was digested with the restriction enzyme PacI before electroporation into the hESC lines HES3 (http://www.escellinternational.co/) and MEL1 (Millipore, Billerica, MA, USA) as described previously [8]. Targeted hESC clones were identified by a PCR-based screening strategy using primer P1, a forward primer in the neomycin resistance gene, in conjunction with P2, a reverse primer located immediately 3′ of genomic sequences encompassed by the targeting vector. The fidelity of homologous recombination within the 5′ arm was confirmed by PCR using P3, a forward primer located immediately 5′ of genomic sequences included in the targeting vector, in conjunction with P4, a reverse primer in the GFP gene. By these criteria, a number of clones were identified in which the vector was correctly integrated into the *INS* locus in both HES3 and MEL1 lines. One HES3-derived and one MEL1-derived $INS^{GFPNeo/w}$ clone was expanded, and the neomycin resistance cassette removed as described previously [9]. Single-cell cloning was performed by single-cell deposition using a FACSaria FACS station as described previously [9]. Several colonies representing each primary clone were picked and screened for the loss of the neomycin resistance cassette by PCR. Southern blot analysis using a probe encompassing the coding sequences of enhanced green fluorescent protein (EGFP) (Invitrogen, Carlsbad, CA, USA) was performed on EcoRV-digested genomic DNA from each cell line (Fig. 1b). As this enzyme cuts only once within the vector, the presence of a single band indicated that each cell line contained a single integration of the targeting vector. The DNA fragments generated by PCR using the primers P1 and P2, P3 and P4 were cloned and sequenced to establish that the targeting vector had been correctly integrated into the *INS* locus.

hESC culture and differentiation hESCs were cultured and passaged as reported elsewhere [10]. The differentiation of

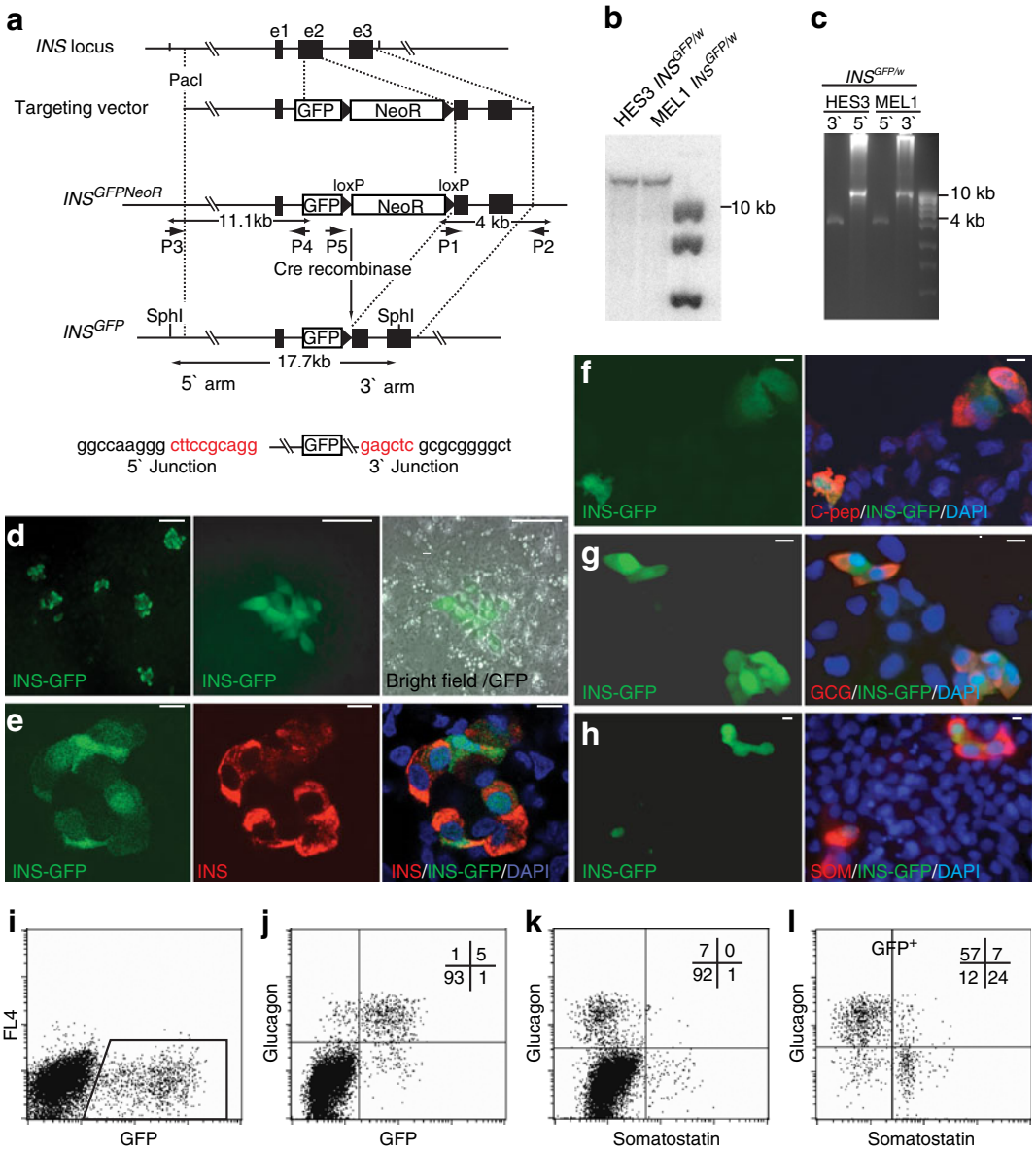

Fig. 1 Generation and characterisation of $INS^{GFP/w}$ hESCs. **a** Vector used to target the *INS* locus in hESCs. The G418 antibiotic resistance (NeoR) cassette, flanked by loxP sites (black triangles), was removed with Cre recombinase. PCR primers (P1, P2) and (P3, P4) were used to identify targeted clones. The sequences of PCR fragments shown below represent the junctions between the extremities of the targeting vector (red text) and the *INS* locus (black text). **b** Southern blot analysis of SphI-digested genomic DNA isolated from single-cell cloned MEL1 and HES3 $INS^{GFP/w}$ hESCs showed a single *GFP* insert present in both lines. **c** PCR analysis with primer pairs (P3, P4) and (P5, P2) generated DNA fragments of ~10 and 4 kb, respectively, the size predicted for correct integration of the targeting vector into the *INS* locus. **d** Images showing clusters of GFP^+ cells in differentiated cultures of $INS^{GFP/w}$ hESCs. Scale bar, 100 μm. **e**–**h** Immunofluorescence analysis of flat cultures showing that INS-GFP^+ cells produce insulin, C-peptide (C-pep), glucagon (GCG) and somatostatin (SOM). Scale bar, 10 μm. **i** Flow cytometry analysis showing that INS-GFP^+ cells can be clearly distinguished from the GFP^- population. **j**–**l** Intracellular flow cytometry (ICF) analysis confirming that the majority of INS-GFP^+ cells co-produce glucagon and/or somatostatin. FACS plot shows the production of glucagon and somatostatin in cells gated on INS-GFP^+ expression. Note: GFP intensity is decreased by ~1 log in processing samples for ICF

hESCs into INS^+ cells was performed using several different protocols. Adherent, flat culture differentiations based on the work of D'Amour et al. [11] and Kroon et al. [2] (referred to as 'flat cultures'). Spin EB differentiations (referred to as 'spin EBs') [5], were set up in APEL medium [6]. Differentiation of spin EBs under pancreatic-specific conditions was as follows. EBs were formed by the forced aggregation of 2,000 (HES3) or 3,500 (MEL1) hESCs in APEL (the protein-free hybridoma medium component was omitted from this formulation) containing 10 ng/ml bone morphogenetic protein 4 (BMP4) and 150–200 ng/ml activin A (batch dependent) in low-attachment

96-well plates. After 3 days, medium was replaced with APEL containing 200–400 ng/ml noggin (batch dependent). At day 6, medium was replaced with APEL containing 1×10^{-5} mol/l retinoic acid (RA). At day 9, the medium was changed to APEL without polyvinyl alcohol (AEL) containing 1×10^{-5} mol/l RA, 100 μmol/l glucagon-like peptide 1 (GLP1), 1×B27 and 10 mmol/l nicotinamide. At day 15 of differentiation, EBs were transferred to gelatinised, adherent 96-well plates, and insulin production was induced in AEL containing 10 mmol/l nicotinamide and 50 ng/ml IGF-I. With this system, most EBs contained INS-GFP$^+$ cells by day 30 of differentiation. In addition, INS-GFP$^+$ cells were also differentiated according to a protocol developed by Nostro and colleagues [12], referred to as the 'Nostro protocol'. Recombinant human activin A, fibroblast growth factor 10 (FGF10), fibroblast growth factor 7 (KGF), IGF-I and hepatocyte growth factor (HGF) were purchased from R&D Systems (Minneapolis, MN, USA). Basic FGF (FGF2) was purchased from Peprotech (Rocky Hill, NJ, USA). Wingless-type MMTV integration site family, member 3A (WNT3A) and noggin were purchased from R&D Systems or provided by the Australia Stem Cell Centre (Melbourne, VIC, Australia). KAAD-cyclopamine was purchased from Toronto Research Chemicals (North York, ON, Canada); all-*trans* RA, nicotinamide, SB431542 and GLP1 were purchased from Sigma-Aldrich (St Louis, MO, USA).

Live cell imaging and immunofluorescence Live cell imaging of spin EBs in a 96-well plate format was performed with a Leica TCS NT inverted microscope, and images were processed with ImageJ software. For immunofluorescence analysis of flat cultures, differentiated cells were fixed for 15 min in 4% (wt/vol.) paraformaldehyde in PBS, permeablised in 0.2% (vol./vol.) Triton X-100 at room temperature for 10 min, and blocked for 60 min in 10% (vol./vol.) goat serum. Primary antibodies were incubated overnight at 4°C, and secondary antibodies were incubated for 1 h at 24°C. The following antibodies were used: rabbit anti-pancreatic and duodenal homeobox 1 (PDX1) (kindly provided by C. Wright, Vanderbilt University, Nashville, TN, USA); mouse anti-NK2 homeobox 2 (NKX2-2) (Developmental Studies Hybridoma Bank (DSHB; Iowa City, IA, USA; clone 74.5A5); mouse anti-NKX6.1 (DSHB); mouse anti-ISL LIM homeobox (ISL)1/2 (DSHB clone 39.4D5); mouse anti-paired box 6 (PAX6) (DSHB); guinea pig anti-insulin (Dako, Glostrup, Denmark; clone A0564); rabbit anti-C-peptide (Millipore; clone 4020; note that this antibody detects both C-peptide and proinsulin); rabbit anti-glucagon (Dako; clone A0565); anti-glucagon (Sigma; clone K79bB10); rat anti-somatostatin (Millipore; clone MAB354). Secondary antibodies used were Alexa-488- and Alexa-568-conjugated goat antibodies against mouse, rat, rabbit and goat (Invitrogen) and a tetramethyl rhodamine iso-thiocyanate (TRITC)-conjugated antibody against guinea pig (Sigma).

For wholemount immunofluorescence of spin EBs [13], differentiated EBs were removed from 96-well plates and fixed for 90 min on ice in 4% (wt/vol.) paraformaldehyde in PBS and permeablised in 1% (vol./vol.) Triton X-100 at room temperature for 90 min. EBs were blocked for 90 min in 10% (vol./vol.) goat serum. Incubation of primary and secondary antibodies was as described above. All washes were for 15 min in PBS/10% FCS.

Flow cytometric analysis For flow cytometric analysis and sorting of live cells, hESCs differentiated in flat cultures or as spin EBs were dissociated with TrypLE-select (Invitrogen) to give a single-cell suspension and purified as described previously [7]. High-throughput flow cytometric analysis of cells in 96-well plates was performed with an LSR II multi-laser benchtop flow cytometer (BD Biosciences, Franklin Lakes, NJ, USA), and FACS plots processed using Gatelogic software (www.inivai.com/GatelogicHome.html).

Reculture experiments By using flow cytometry we purified day 20 INS-GFP$^+$ cells generated with the flat culture protocol and then added between 2×10^3 and 5×10^3 INS-GFP$^+$ cells to each well of a low-attachment 96-well tray in APEL medium containing 10 μmol/l rho-associated, coiled-coil containing protein kinase 1 (ROCK) inhibitor Y27632 [7, 14]. FGF10, HGF, FGF2, BMP4, KGF and noggin (10–100 ng/ml) were added singly or in combination at the time of aggregation or after 24 h. Medium containing the ROCK inhibitor was replaced with APEL containing combinations of the above growth factors after aggregates had formed (usually 24–72 h). Half of the medium was changed every 3–4 days over a 3-week period. When used, 20 μl Matrigel (diluted 2:1 in APEL medium) was added directly to reaggregated INS-GFP$^+$ cells. After Matrigel polymerisation, 100 μl APEL medium containing combinations of the above growth factors was added to each well. Medium was refreshed periodically as described above. Intracellular flow cytometric analysis was performed as described by Nostro et al. [12]. Bromodeoxyuridine (BrdU) incorporation measured by flow cytometry was performed according to the manufacturer's (BD Biosciences) instructions.

Gene expression analysis RNA preparation, Illumina microarray analysis and real-time quantitative PCR was performed essentially as described previously [7]. Briefly, total RNA for each sample was amplified, labelled and hybridised to human WG-6v2, human HT12v3 or HT12v4 BeadChips according to Illumina standard protocols (Illumina, San Diego, CA, USA) at the Australian Genome Research Facility. Initial data analysis was performed using GenomeStudio version 2010.3 (Illumina),

using average normalisation across all the samples. Alternatively, data were analysed using R/BioConductor using algorithms within the lumi package [15] (function: bgAdjust.affy and quantile normalisation [16]). Subsequent data analysis was performed using MultiExperiment Viewer [17, 18]. Hierarchical clustering was performed using Pearson correlation with average linkage clustering. Differentially expressed genes were subjected to functional clustering analysis using the DAVID public database (Database for Annotation, Visualization and Integrated Discovery) [19, 20].

Results

We targeted a GFP reporter gene to the *INS* locus of HES3 and MEL1 hESCs, creating a reagent to enable detailed study of the potential and characteristics of INS$^+$ cells (Fig. 1a). Undifferentiated *INS*$^{GFP/w}$ hESCs had a normal karyotype (46XX for the HES3-derived line and 46XY for the MEL1-derived line), produced stem cell markers and generated teratomas containing derivatives of the three primary germ layers upon transplantation into immunodeficient mice (electronic supplementary material [ESM] Fig. 1). Southern blotting analysis indicated that both lines contained a single GFP insertion, while PCR and DNA sequencing confirmed that the targeting vector had been integrated by homologous recombination (Fig. 1a–c). Differentiation of *INS*$^{GFP/w}$ hESCs in flat culture (ESM Fig. 2) revealed that INS-GFP$^+$ cells appeared as small clusters, which co-stained with insulin and C-peptide (Fig. 1d–f), confirming the fidelity of the reporter gene. Immunofluorescence experiments also demonstrated that INS-GFP$^+$ cells co-produced somatostatin and glucagon (Fig. 1g, h), confirming previous reports [11] that early hESC-derived INS$^+$ cells are polyhormonal. This conclusion was supported by flow cytometry analysis, which showed that ~80% of INS-GFP$^+$ cells co-produced glucagon, and ~20% produced somatostatin (Fig. 1i–l). INS-GFP$^+$ cells producing neither hormone constituted a minor proportion of the population; however, it is possible that these cells produced other hormones that were not assayed (e.g. pancreatic polypeptide [PPY], ghrelin [GHRL]).

Two distinct but not necessarily mutually exclusive scenarios may explain the prevalence of polyhormonal INS$^+$ cells generated by many current in vitro differentiation protocols. First, it has been postulated that the first wave of INS$^+$ cells emerging during human embryonic development represent a primary wave of endocrine cells, the contribution of which to the mature endocrine organ is still undetermined [21]. Alternatively, the cells produced in current differentiation systems may have the potential for further maturation and proliferation, but this requires appropriate culture conditions. We attempted to address these possibilities by asking if INS-GFP$^+$ cells derived from flat culture differentiation could mature further in vitro. To perform these experiments, purified INS-GFP$^+$ cells were reaggregated in APEL medium containing 10 μmol/l ROCK inhibitor Y27632 [7] (Fig. 2a). This reaggregation step substantially improved the viability of cells after sorting. After 24 h, INS-GFP$^+$ cells formed tight E-cadherin$^+$ clusters (insulin-positive aggregates [IPAs]) that displayed the same spectrum of GFP intensities as present in the original sorted population and continued to co-produce glucagon (Fig. 2b, c).

The growth and differentiation potential of IPAs was subsequently examined in vitro. We first tested the ability of previously reported pancreatic growth factors to either sustain INS-GFP production over a 3-week period or promote expansion of the GFP$^+$ population. Extended cultures of IPAs in APEL medium alone revealed that INS-GFP production waned rapidly after IPA formation (data not shown). We also observed that factors previously reported to have a role in expansion of the pancreatic primordium, such as HGF and FGF10 [22–24], promoted the slow growth of the population overall, particularly in the presence of Matrigel (Fig. 2d and data not shown). However, none of the factor combinations tested (see Methods) sustained GFP production for more than 3 weeks nor promoted expansion of the INS-GFP$^+$ pool. Instead, GFP$^-$ cells (either derived from INS$^+$ cells or representing contaminants in the original sorted population) eventually became the predominant cell type within the IPAs (Fig. 2d), raising the possibility that INS-GFP$^+$ cells generated in flat cultures were postmitotic. To address this, we performed BrdU incorporation analysis of cells differentiated using the flat culture protocol. This analysis indeed showed that, at later differentiation stages, few INS-GFP$^+$ cells and/or their immediate precursors incorporated BrdU (Fig. 2e), consistent with the notion that these cells were essentially non-proliferative.

In order to further characterise INS-GFP$^+$ cells generated using the flat culture and Nostro protocols, cells were isolated by flow cytometry and processed for gene expression microarray analysis (Fig. 3a). Data were compared with those of the INS-GFP$^-$ fraction from the same cultures, as well as from fetal pancreas, adult pancreas and adult islets. Differentially expressed genes were identified by performing sequential pair-wise comparisons between the GFP$^+$ and GFP$^-$ fractions from four independent sorting experiments using two different cell lines, performed in two different laboratories. A total of 202 probes (186 genes) were identified that were upregulated more than twofold across all four comparisons (see ESM Fig. 3 for scatter plots for individual pair-wise comparisons). For display purposes, a selected subset of these 186 genes was grouped into functional categories based on their

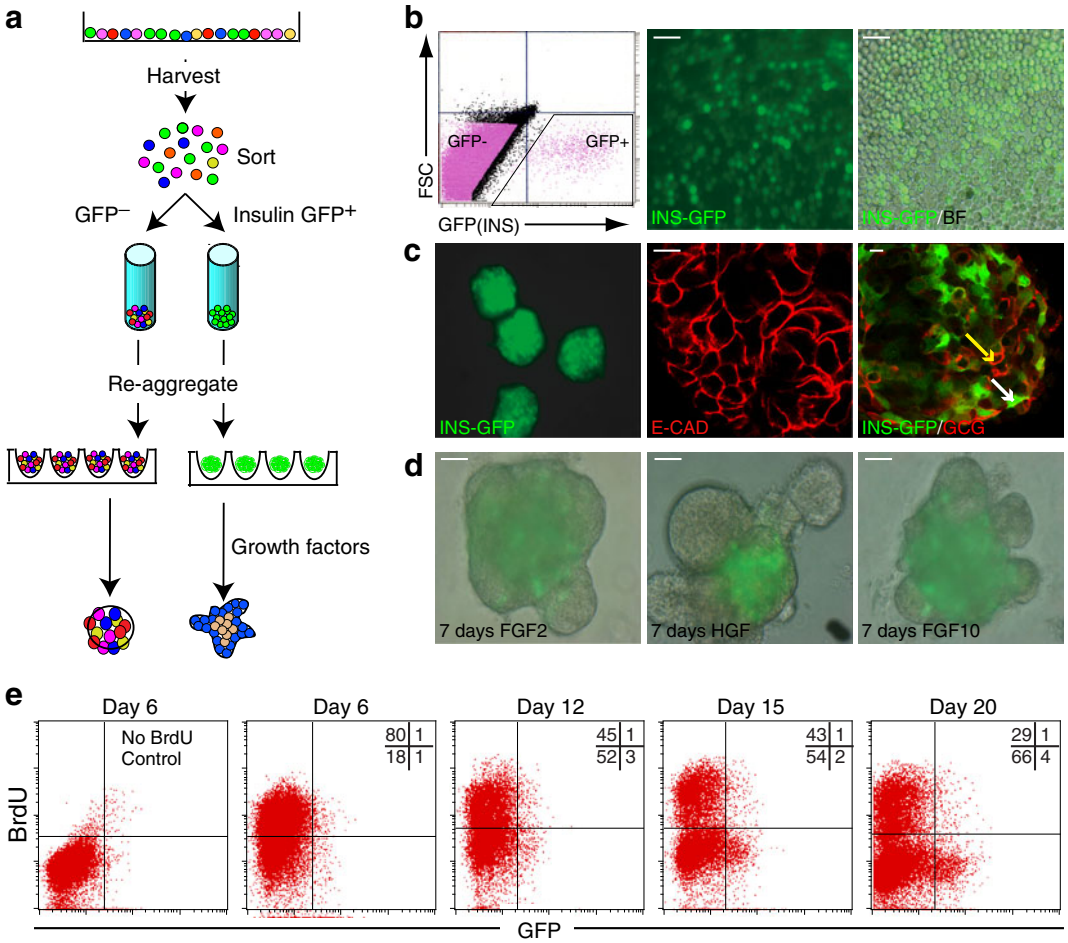

Fig. 2 INS-GFP$^+$ cells can be viably isolated and recultured. **a** Schematic depicting the experimental outline of sorting and reculture experiments. **b**, **c** Cells sorted on the basis of INS-GFP expression display a wide spectrum of GFP fluorescence intensities (**c**; scale bar, 50 μm). After reaggregation, INS-GFP$^+$ cells form tight E-cadherin$^+$ (E-CAD) insulin producing aggregates (IPAs) that continue to express glucagon (GCG). Scale bar, 10 μm. **d** Bright field (BF)–GFP overlay images of IPAs cultured in APEL medium supplemented with 100 ng/ml of the indicated growth factors for 7 days. Note the significant growth of the GFP$^-$ population and diminishing proportion of INS-GFP$^+$ cells present within these aggregates. Scale bar, 50 μm. **e** Flow cytometric analysis of differentiating cultures of *INS*$^{GFP/w}$ hESCs labelled with BrdU for 24 h at the indicated time points. Note that GFP expression is not observed before day 15. At day 20, the majority of INS-GFP$^+$ cells had not incorporated BrdU. The percentage of cells within each quadrant is indicated. FSC, forward scatter

gene ontology (secretory, transcription factors, diabetes associated) (Fig. 3a).

Comparison of GFP$^+$ and GFP$^-$ fractions indicated that GFP$^+$ cells had upregulated a suite of genes that confirmed the commitment of this population to endocrine differentiation (Fig. 3b). Genes that were substantially upregulated in the GFP$^+$ fraction included those for hormones traditionally associated with pancreatic endocrine cells (*GCG*, *INS*, *SST*, *PPY*), a suite of known pancreatic transcription factor genes (*NKX2.2* [also known as *NKX2-2*], *ARX*, *NEUROD1*, *MAFB*) as well as a number of genes associated with type 1 diabetes (*HLA*, *GAD*, *PTPRN*). This analysis also revealed that, within this restricted set, INS-GFP$^+$ samples were most similar to islets, consistent with their endocrine nature.

Gene profiling data were compared with the results of immunofluorescence experiments, confirming that INS-GFP$^+$ cells produced transcription factors such as PAX6, NKX2.2 and ISL1 (Fig. 4). We also investigated the expression of *PDX1* and *NKX6.1*, two genes that were absent from the composite list derived from comparison of all four sorting experiments as well as being absent from pair-wise comparison of individual experiments. In the case of *PDX1*, quantitative PCR studies (ESM Fig. 4) suggested that its absence from the GFP$^+$ fraction appeared to partly reflect low sensitivity of the Illumina probe set to detect this particular gene. However, although GFP$^+$ cells appeared to produce elevated levels of PDX1 (Fig. 4), it is also possible that the relatively high frequency of *PDX1*$^+$ cells present in the INS$^-$ fraction overall may have contributed to the poor enrichment of *PDX1* transcripts observed in the former population (see ESM Fig. 5). In contrast with PDX1, co-production of NKX6.1 with INS-GFP was not observed in

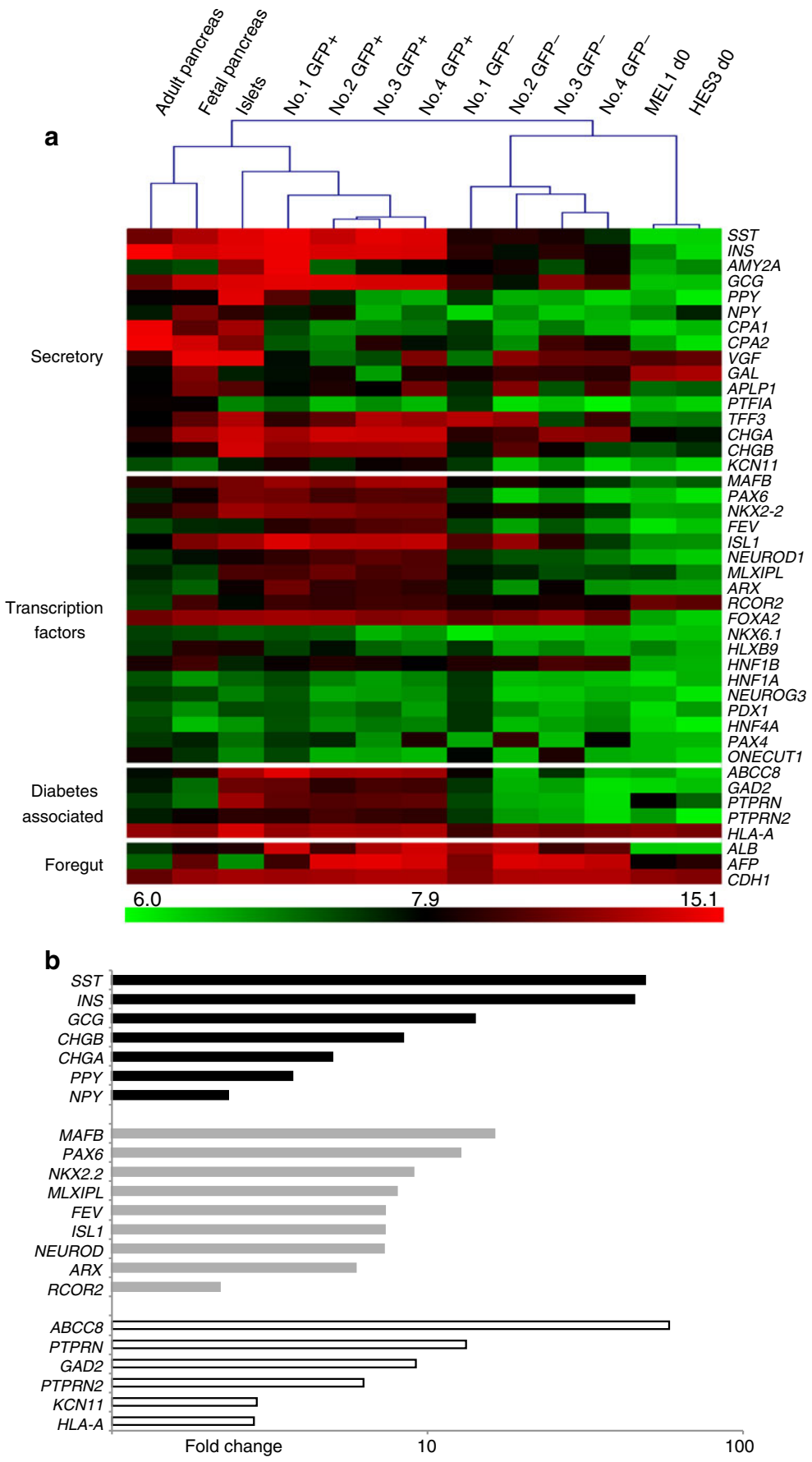

Fig. 3 Gene expression profiling of INS-GFP$^+$ cells using Illumina microarrays. **a** Two-colour heat map representation of selected data derived from Illumina gene chip analysis of GFP$^+$ and GFP$^-$ cell fractions from four independent differentiation experiments on two independent *INS*$^{GFP/w}$ hESC lines. Genes included were upregulated more than twofold in all pair-wise comparisons of each of the independently derived samples. Heat map represents $\log_2$ transformed data that have been average normalised. Hierarchical clustering of the samples was performed by Pearson correlation with average linkage. **b** Bar graph showing selected genes, the expression of which was upregulated more than twofold in the INS-GFP$^+$ cells versus the GFP$^-$ cells, calculated as an average of all four differentiation experiments. Secretory, black bars; transcription factor, grey bars; diabetes associated, white bars

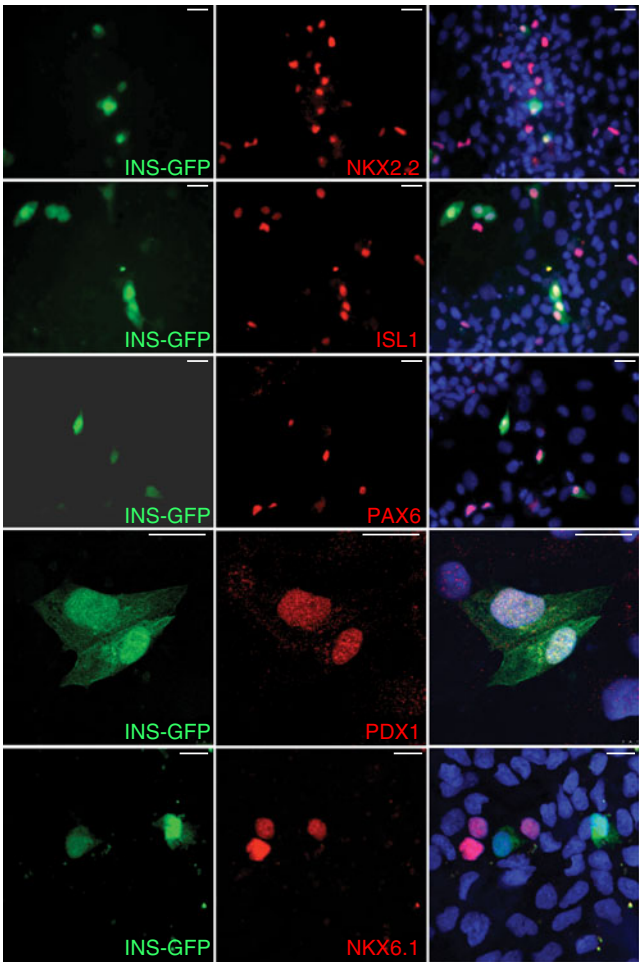

Fig. 4 Gene expression analysis reveals that INS-GFP$^+$ cells express a suite of pancreas-associated transcription factors. Immunofluorescence studies show that INS-GFP$^+$ cells generated with the flat culture protocol express *PAX6*, *NKX2.2*, *ISL1* (scale bar, 20 μm) and *PDX1* (scale bar, 10 μm) but not *NKX6.1* (scale bar, 20 μm). Note that all transcription factors are also produced by GFP$^-$ cells present in these cultures

flat cultures, and transcripts were not detected in samples generated in either the flat culture or the Nostro protocols. This last observation is consistent with other reports and the notion that these hESC-derived INS$^+$ cells represent an immature precursor population that is not yet fully committed to beta cell differentiation [11].

Analysis of INS-GFP$^+$ cells confirmed that both protocols generated cells at a stage of pancreas development that precedes the onset of overt beta cell differentiation. Therefore, we reasoned that the further testing of variables affecting the course of in vitro pancreatic differentiation would be required before more mature phenotypes could be generated. We used *INS*$^{GFP/w}$ hESCs to develop a novel 96-well format spin EB protocol [5] for generating INS-GFP$^+$ cells (Fig. 5a). With this method, we observed that INS-GFP$^+$ cells emerged with slower kinetics compared with the flat culture or the Nostro protocols, with the onset of GFP expression at differentiation day 20 (compared with day 15 for flat cultures). Late-stage spin EBs displayed a spectrum of morphologies with respect to the localisation of INS-GFP$^+$ cells (Fig. 5b). In general, these cells appeared in clusters either within the main body of the EB or as isolated spheres surrounding the GFP$^-$ EB core. Visual inspection suggested that some EBs contained a substantial fraction of GFP$^+$ cells, an observation confirmed by FACS analysis that showed that single EBs contained up to 37% INS-GFP$^+$ cells (Fig. 5c). However, without specific preselection of GFP$^+$ EBs, the overall frequency of GFP$^+$ cells generated with this protocol was ~2–5%.

Immunofluorescence and flow cytometry analysis of spin EB-derived cells indicated that, unlike cells generated using the flat culture protocol, a substantial fraction (40%) of INS-GFP$^+$ cells produced neither glucagon nor somatostatin (Fig. 6a–e). Immunofluorescence analysis showed that, like cells generated in flat cultures, spin EB-derived INS-GFP$^+$ cells produced PAX6, ISL1, NKX2.2 and PDX1. In addition, similar to that observed with flat cultures, PDX1$^+$ nuclei were also observed in a substantial proportion of GFP$^-$ cells within spin EBs that contained INS-GFP$^+$ cells (Fig. 6f). However, unlike cells derived with either the flat culture or the Nostro protocols, many INS-GFP$^+$ cells within spin EBs co-produced NKX6.1 (Fig. 6c) (see ESM Fig. 6 for single-colour images). In a survey of EBs immunostained for NKX6.1, we observed that 34% (122/363) of cells were GFP$^+$ and 16% (58/363) of cells were NKX6.1$^+$. In these EBs, INS-GFP$^+$/NKX6.1$^+$ cells comprised 11% (39/363) of the total cell population or 32% (39/122) of the INS-GFP$^+$ cell population (data not shown). The presence of an INS-GFP$^+$/NKX6.1$^+$ population using the spin EB method suggests that the differentiation conditions are conducive to ongoing differentiation of pancreatic endoderm.

Microarray analysis confirmed that INS-GFP$^+$ cells generated by the spin EB method had upregulated a similar cohort of pancreatic genes to those derived using the flat culture differentiation protocol (Fig. 7a). However, given the presence of INS-GFP$^+$/NKX6.1$^+$ cells in spin EBs, we used microarray analysis to explicitly compare the gene expression profiles of INS-GFP$^+$ cells generated using all three protocols. Specifically, we were interested in genes that might provide information relating to differentiation status and/or cellular identity. This comparison showed that INS-GFP$^+$ cells derived from flat cultures expressed elevated levels of a number of genes that are associated with non-pancreatic derivatives of foregut endoderm (Fig. 7b). In particular, transcripts encoding apolipoproteins (liver), claudin 18 (CLDN18) (lung) and pepsinogen (stomach) were more abundant in INS-GFP$^+$ cells generated using previously described methods. Conversely, INS-GFP$^+$

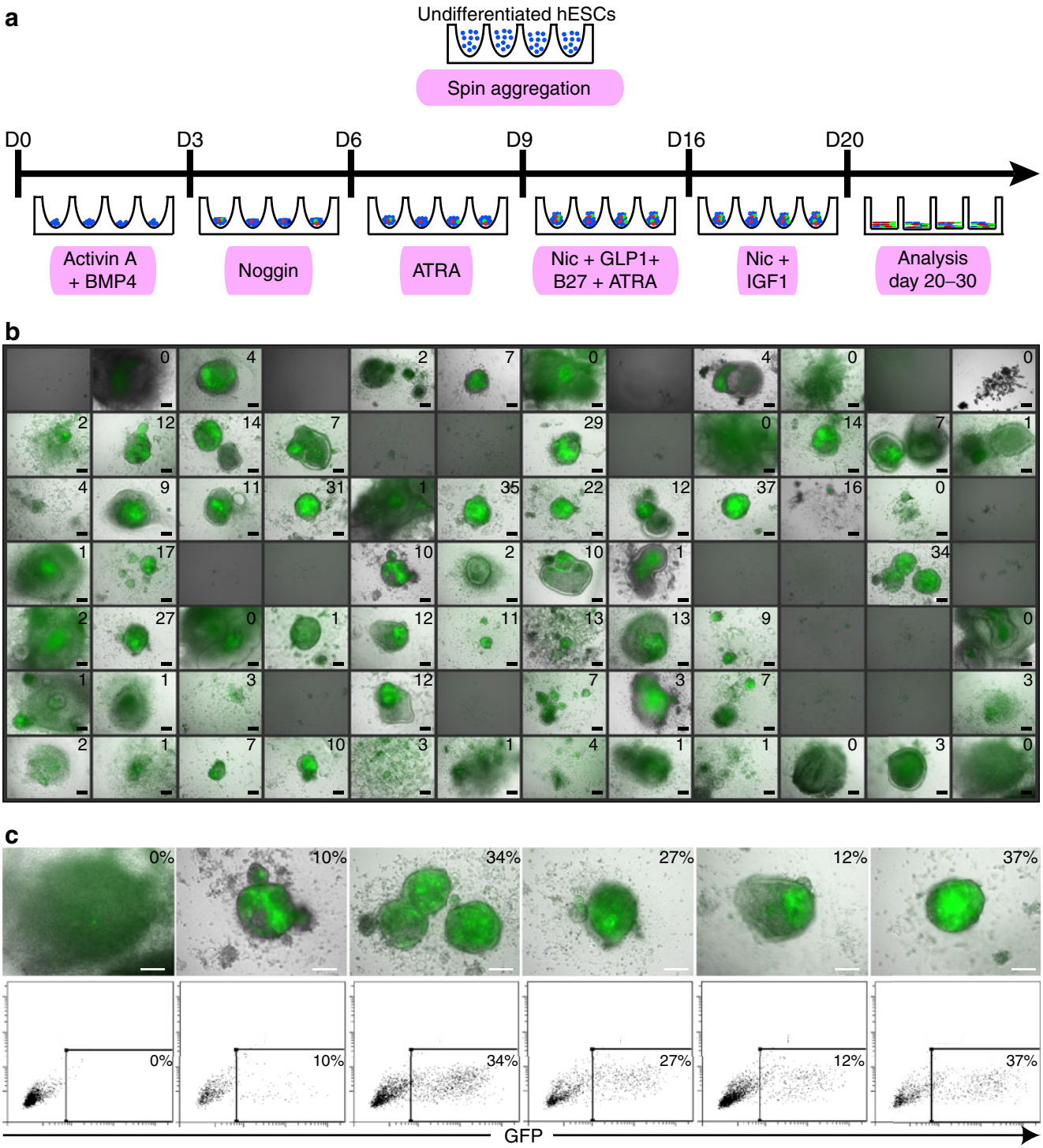

Fig. 5 Generation and analysis of INS-GFP$^+$ cells using the spin EB differentiation platform. **a** Schematic representation of the spin EB protocol, which uses the wholly recombinant protein-based medium, APEL. Concentration ranges given for specific growth factors reflect variation in the specific activity of individual batches. ATRA, all-*trans* retinoic acid; Nic, nicotinamide. **b** Composite image showing the variation in the size and INS-GFP$^+$ cell content of individual EBs. Greyed boxes indicate wells that contained EBs lost during processing (medium changes). The numbers in the upper right hand corners indicate the percentage of GFP$^+$ cells in each EB determined by flow cytometric analysis. **c** Selected images from **a** showing the morphology of INS-GFP$^+$ EBs and the primary flow cytometry data associated with each sample. Scale bar, 100 μm

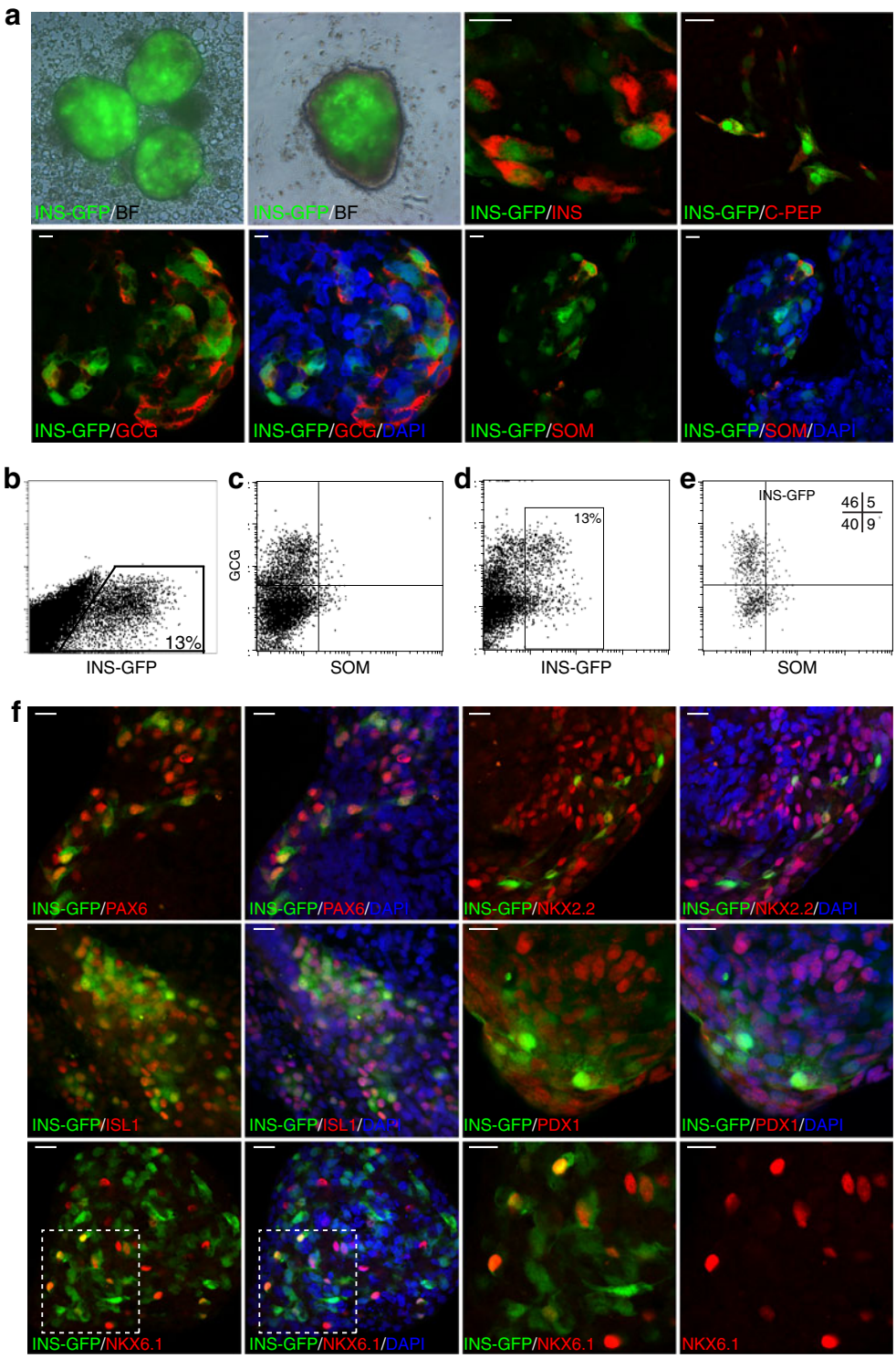

Fig. 6 Hormone and transcription factor production by cells within INS-GFP$^+$ spin EBs. **a** Individual spin EBs containing INS-GFP$^+$ cells were collected, immunolabelled with antibodies directed against endocrine hormones as indicated and imaged by confocal microscopy. Scale bar, 10 μm. **b–e** Flow cytometric analysis of HES3 INS-GFP$^+$ cells representing selected EBs for co-production of glucagon (GCG) and somatostatin (SOM). Note that GFP intensity is lost during fixation and permeabilisation (compare **b** with **d**). The percentage of cells within specific regions or quadrants is indicated. **e** The expression of glucagon and somatostatin in the INS-GFP$^+$ population. **f** Whole-mount immunofluorescence of INS-GFP$^+$ EBs generated with the spin EB platform showing production of PAX6, NKX2.2, ISL1, PDX1 and NKX6.1. Scale bar, 20 μm. Note that all transcription factors are also produced by GFP$^-$ cells present in these cultures

cells derived from spin EB cultures were enriched for transcripts encoding a number of *HOX* genes, some of which have been previously associated with axial patterning of the developing gut tube [25]. Whether these differences are indicative of differences in relative developmental maturity or reflect some underlying difference in the specification process remains to be determined.

Discussion

INS$^{GFP/w}$ hESCs reported on here represent a novel reagent for the study of beta cell differentiation in vitro. By directly tagging the *INS* locus, cells can be viably isolated for further studies or followed in real time, allowing their growth and response to culture manipulations to be directly

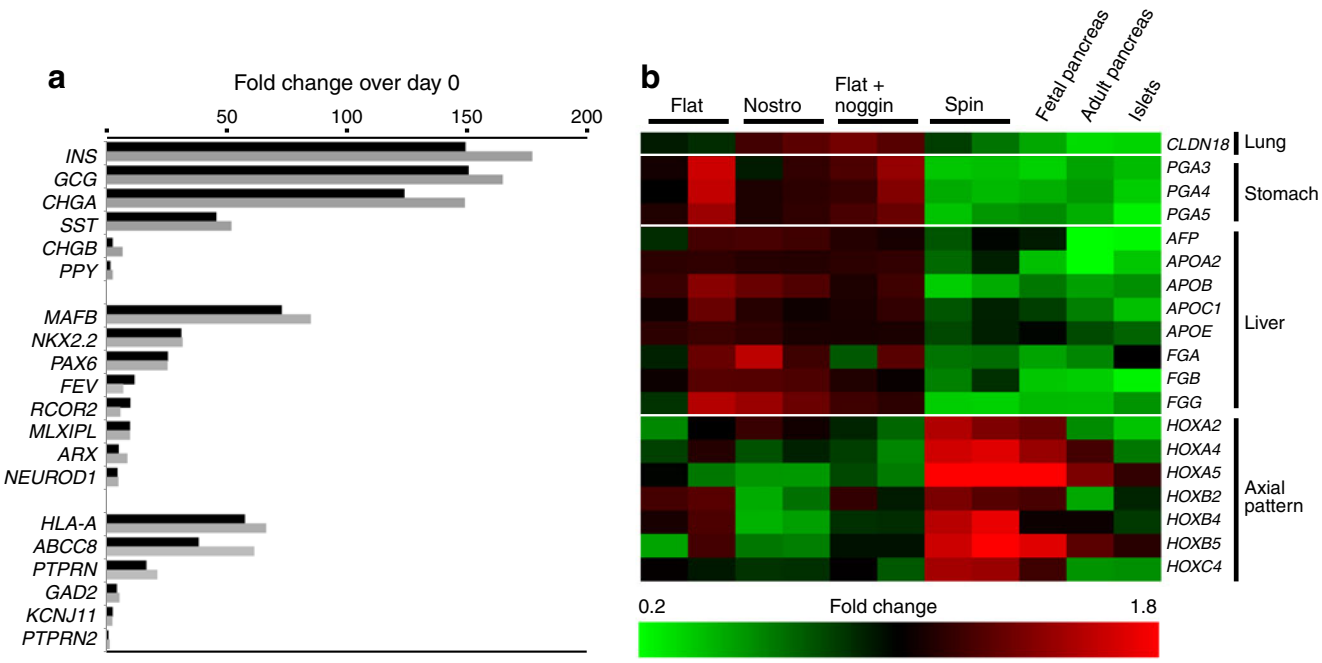

Fig. 7 Gene expression profiling of spin EB-derived INS-GFP$^+$ cells using Illumina microarrays. **a** Bar graph showing selected pancreas-associated genes, the expression of which was upregulated in INS-GFP$^+$ cells derived from MEL1- and HES3-based *INS*$^{GFP/w}$ hESCs differentiated using the spin EB (grey) and flat culture + noggin (black) protocols. Fold change was calculated from the average signal obtained from two independent spin EB- and flat culture-derived samples divided by the average signal obtained for those genes in the undifferentiated MEL1 and HES3 *INS*$^{GFP/w}$ hESCs. **b** Two-colour heat map representation of selected data derived from Illumina gene chip analysis of INS-GFP$^+$ cell fractions generated using the protocols indicated. Only genes that were upregulated more than twofold in pair-wise comparisons from two independently derived samples from the spin EB and flat culture differentiation protocols are shown. Expression levels were background adjusted, quantile normalised and $\log_2$ transformed using R/Bioconductor. The heat map represents expression of each gene relative to its average expression level across all eight INS-GFP$^+$ samples

monitored. We have used these cells to generate a set of gene profiling data that will serve as a baseline for future studies on hESC-derived endocrine cells. These data, in conjunction with immunofluorescence studies, reveal that INS-GFP$^+$ cells generated with two distinct but related protocols [11, 12] have hallmarks of immature endocrine cells. This conclusion was drawn from the observation that most INS-GFP$^+$ cells in late-stage (day 20–25) cultures produced other endocrine hormones, most commonly glucagon. Although cells producing multiple hormones are present in the developing human pancreas, their relative abundance as a fraction of the hormone-positive population is minor [26]. Several theories have been proposed to account for polyhormonal cells in cultures of differentiating hESCs. The preponderance of INS$^+$ cells that express other hormones may indicate a bona fide differentiation-intermediate population, further development of which is arrested because culture conditions are not appropriate. Alternatively, the appearance of this cell type may signify that current culture conditions drive the generation of an in vitro artefact that lacks the capacity for further differentiation along the beta cell lineage. BrdU labelling experiments suggest that the polyhormonal cells generated under these conditions are postmitotic, a conclusion consistent with other studies suggesting that the major source of new islets during development is not pre-existing hormone-producing cells [27–30].

Taken together, the above observations emphasise that further work will be required before mature beta cells can be readily generated from hESCs. Therefore, methodologies that lend themselves to testing large numbers of variables will assist efforts to refine or reconstruct hESC to beta cell differentiation protocols. In this context, we used *INS*$^{GFP/w}$ hESCs to develop a 96-well format spin EB differentiation protocol that used the recombinant protein-based medium, APEL [6]. This platform has a number of advantages that will facilitate further exploration of pathways governing pancreatic differentiation of hESCs. First, the 96-well format is compatible with high-throughput methodologies that enable the simultaneous assessment of large numbers of variables. Second, because APEL contains only recombinant proteins (albumin, transferrin and insulin), inconsistencies arising from batch to batch variation intrinsic to media components such as BSA are minimised. Using the spin EB platform, we observed that INS-GFP$^+$ cells appeared in the context of a variety of morphologically distinct structures, some of which appeared to derive and/or separate from the main mass of the EB and resembled

islet-like clusters described by others [3, 31]. Interestingly, islet-like clusters in both reports contained a substantial fraction of INS$^+$ cells that did not co-produce either glucagon or somatostatin, mirroring our findings with spin EB-derived INS-GFP$^+$ cells. This phenotype would be consistent with the idea that spin EB-derived INS-GFP$^+$ cells represented a more mature stage of development, a conclusion supported by the substantial number of INS-GFP$^+$ cells in spin EBs producing the later-stage beta cell differentiation marker, NKX6.1 [32].

Nevertheless, with the notable exception of *INS*, *GCG* and *SST*, our microarray data indicated that INS-GFP$^+$ cells generated using all of the protocols (flat, Nostro or spin EB) expressed relatively low levels of genes recently reported by Dorrell and colleagues [33] to be associated with mature alpha or beta cells (ESM Fig. 7). This observation reinforces the notion that establishing culture conditions that promote appropriate maturation represents a significant hurdle for the generation of functional beta cells in vitro.

Our microarray data also suggested that spin EB-derived INS-GFP$^+$ cells expressed lower levels of transcripts associated with non-pancreatic endodermal cell types and higher levels of *HOX* genes—genes with a known role in setting axial position within the developing embryo. In this light, it is tempting to speculate that the spin EB environment may provide differentiative cues that more precisely specify positional identity and/or more closely resemble those in the embryo. While further studies are necessary to determine if INS-GFP$^+$ cells can be further differentiated, the *INS*$^{GFP/w}$ hESCs and 96-well spin EB format protocol described here represent new tools for optimising generation of beta cells in vitro for the treatment of type 1 diabetes.

Acknowledgements We thank R. Mayberry, K. Koutsis and A. Bruce for the provision of hESCs, S. Hawes for the provision of human fetal pancreas RNA, FlowCore for flow cytometric sorting, and Monash Micro Imaging for confocal microscopy and time-lapse video services. All work involving experimentation with human ES cell lines was performed under approval of the Monash University Human Ethics Committee (2002-225MC). This work was supported by the Australian Stem Cell Centre (ASCC), The National Health and Medical Research Council (NHMRC) of Australia and the Juvenile Diabetes Research Foundation. M. C. Nostro was supported by a postdoctoral fellowship from the McEwen Centre for Regenerative Medicine. L. C. Harrison is a Senior Principal Research Fellow, and A. G. Elefanty and E. G. Stanley are Senior Research Fellows of the NHMRC.

Contribution statement SJM contributed to the conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing and final approval of the manuscript. XL, JVS, CEH, QCY, SML, MCN, DAE and SF contributed to the collection and assembly of data, data analysis and interpretation and critical revision, and gave final approval of the manuscript. LCH and GK contributed to provision of study materials, data analysis and interpretation and critical revision, and gave final approval of manuscript. AGE and EGS contributed to conception and design, data analysis and interpretation, manuscript writing and financial support, and gave final approval of manuscript.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

References

1. Speight J, Reaney MD, Woodcock AJ, Smith RM, Shaw JA (2010) Patient-reported outcomes following islet cell or pancreas transplantation (alone or after kidney) in type 1 diabetes: a systematic review. Diabet Med 27:812–822
2. Kroon E, Martinson LA, Kadoya K et al (2008) Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat Biotechnol 26:443–452
3. Jiang W, Shi Y, Zhao D et al (2007) In vitro derivation of functional insulin-producing cells from human embryonic stem cells. Cell Res 17:333–344
4. Van Hoof D, D'Amour KA, German MS (2009) Derivation of insulin-producing cells from human embryonic stem cells. Stem Cell Res 3:73–87
5. Ng ES, Davis RP, Azzola L, Stanley EG, Elefanty AG (2005) Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. Blood 106:1601–1603
6. Ng ES, Davis R, Stanley EG, Elefanty AG (2008) A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat Protoc 3:768–776
7. Goulburn AL, Alden D, Davis RP et al (2011) A targeted NKX2.1 human embryonic stem cell reporter line enables identification of human basal forebrain derivatives. Stem Cells 29:462–473
8. Costa M, Dottori M, Sourris K et al (2007) A method for genetic modification of human embryonic stem cells using electroporation. Nat Protoc 2:792–796
9. Davis RP, Costa M, Grandela C et al (2008) A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nat Protoc 3:1550–1558
10. Costa M, Dottori M, Ng E et al (2005) The hESC line Envy expresses high levels of GFP in all differentiated progeny. Nat Methods 2:259–260
11. D'Amour KA, Bang AG, Eliazer S et al (2006) Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat Biotechnol 24:1392–1401
12. Nostro MC, Sarangi F, Ogawa S et al (2011) Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. Development 138:861–871
13. Micallef SJ, Li X, Janes ME et al (2007) Endocrine cells develop within pancreatic bud-like structures derived from mouse ES cells differentiated in response to BMP4 and retinoic acid. Stem Cell Res 1:25–36. doi:10.1634/stemcells.2006-0713

14. Watanabe K, Ueno M, Kamiya D et al (2007) A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat Biotechnol 25:681–686
15. Du P, Kibbe WA, Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics 24:1547–1548
16. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193
17. Saeed AI, Bhagabati NK, Braisted JC et al (2006) TM4 microarray software suite. Methods Enzymol 411:134–193
18. Saeed AI, Sharov V, White J et al (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34:374–378
19. Dennis G Jr, Sherman BT, Hosack DA et al (2003) DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 4:P3
20. da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
21. Pang K, Mukonoweshuro C, Wong GG (1994) Beta cells arise from glucose transporter type 2 (Glut2)-expressing epithelial cells of the developing rat pancreas. Proc Natl Acad Sci USA 91:9559–9563
22. Miralles F, Czernichow P, Ozaki K, Itoh N, Scharfmann R (1999) Signaling through fibroblast growth factor receptor 2b plays a key role in the development of the exocrine pancreas. Proc Natl Acad Sci USA 96:6267–6272
23. Otonkoski T, Cirulli V, Beattie M et al (1996) A role for hepatocyte growth factor/scatter factor in fetal mesenchyme-induced pancreatic beta-cell growth. Endocrinology 137:3131–3139
24. Bhushan A, Itoh N, Kato S et al (2001) Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. Development 128:5109–5117
25. Kawazoe Y, Sekimoto T, Araki M, Takagi K, Araki K, Yamamura K (2002) Region-specific gastrointestinal Hox code during murine embryonal gut development. Dev Growth Differ 44:77–84
26. Jeon J, Correa-Medina M, Ricordi C, Edlund H, Diez JA (2009) Endocrine cell clustering during human pancreas development. J Histochem Cytochem 57:811–824
27. Castaing M, Peault B, Basmaciogullari A, Casal I, Czernichow P, Scharfmann R (2001) Blood glucose normalization upon transplantation of human embryonic pancreas into beta-cell-deficient SCID mice. Diabetologia 44:2066–2076
28. Bouwens L, Lu WG, de Krijger R (1997) Proliferation and differentiation in the human fetal endocrine pancreas. Diabetologia 40:398–404
29. Piper K, Brickwood S, Turnpenny LW et al (2004) Beta cell differentiation during early human pancreas development. J Endocrinol 181:11–23
30. Castaing M, Duvillie B, Quemeneur E, Basmaciogullari A, Scharfmann R (2005) Ex vivo analysis of acinar and endocrine cell development in the human embryonic pancreas. Dev Dyn 234:339–345
31. Jiang J, Au M, Lu K et al (2007) Generation of insulin-producing islet-like clusters from human embryonic stem cells. Stem Cells 25:1940–1953
32. Lyttle BM, Li J, Krishnamurthy M et al (2008) Transcription factor expression in the developing human fetal endocrine pancreas. Diabetologia 51:1169–1180
33. Dorrell C, Schug J, Lin CF et al (2011) Transcriptomes of the major human pancreatic cell types. Diabetologia 54:2832–2844