CHAPTER 3 RESULTS

3.1. An *in vitro* culture system to study the development of the pancreas

3.1.1. An efficient protocol to differentiate hPSCs into endocrine progenitors

To study human pancreas development in an *in vitro* system, a chemically defined and feeder-free pancreatic specification protocol that was previously established in the lab (refer to Chapters 2.1.3 and 2.1.4) was used to differentiate hPSCs into endocrine progenitors. To confirm the efficiency of the protocol in generating a near homogenous population of pancreatic progenitors from hPSCs, I performed comprehensive analyses throughout the differentiation protocol; cells were analysed at key stages of the pancreas development and assessed for the expression of key markers to ensure successful commitment to the pancreatic lineage (Figure 10).

I first performed pancreatic differentiation on H9, a well-characterised hESC line that was routinely used in the lab and has been extensively tested for successful pancreatic differentiation. Cells were assayed by qRT-PCR and immunocytochemistry (ICC) on days 3 (DE), 6 (primitive gut tube), 9 (posterior foregut), 12 (pancreatic endoderm), 15, 18 and 24 (endocrine progenitor) of pancreatic differentiation, while FACS was performed on days 3, 12, 15 and 24 of pancreatic differentiation.

On day 0 of differentiation, the high expression levels of pluripotency factors (OCT4, SOX2 and NANOG) as shown in ICC and the absence of early germ layer markers (EOMES and SOX17) (Figure 11) indicated that the H9 cells were indeed in a fully undifferentiated and pluripotent state.

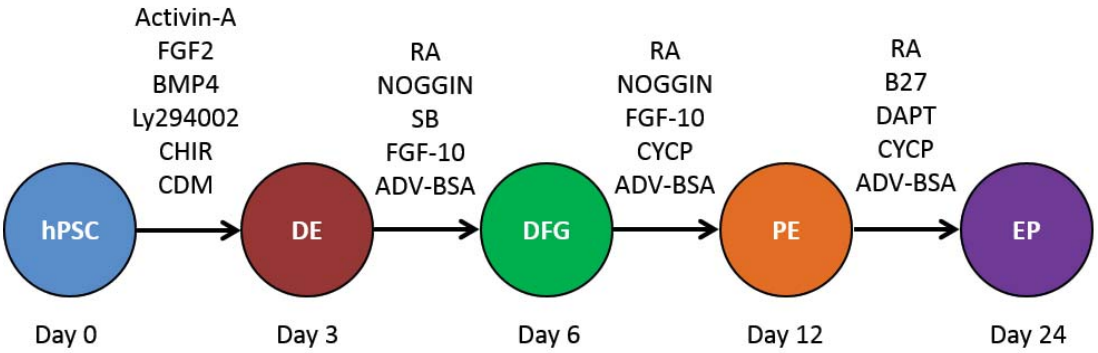


Figure 10. Schematic of the 24-day differentiation protocol. DE, definitive endoderm; DFG, dorsal foregut; PE, pancreatic endoderm; EP, endocrine progenitors. The culture medium and supplements indicated are BMP, bone morphogenetic protein 4; the PI3 kinase inhibitor Ly294002; the GSK3 inhibitor CHIR, CHIR99021; CDM, chemically defined medium; Adv-BSA, Advanced Dulbecco's Modified Eagle Medium/Ham's F-12 medium supplemented with BSA and L-glutamine; RA, retinoic acid; the ALK4/5/78 inhibitor SB-431542 (SB); FGF2, Fibroblast growth factor 2; FGF10, Fibroblast growth factor 10; the Hedgehog inhibitor, CYCP, Cyclopamine-KAAD; B27 supplement; the NOTCH inhibitor, DAPT.

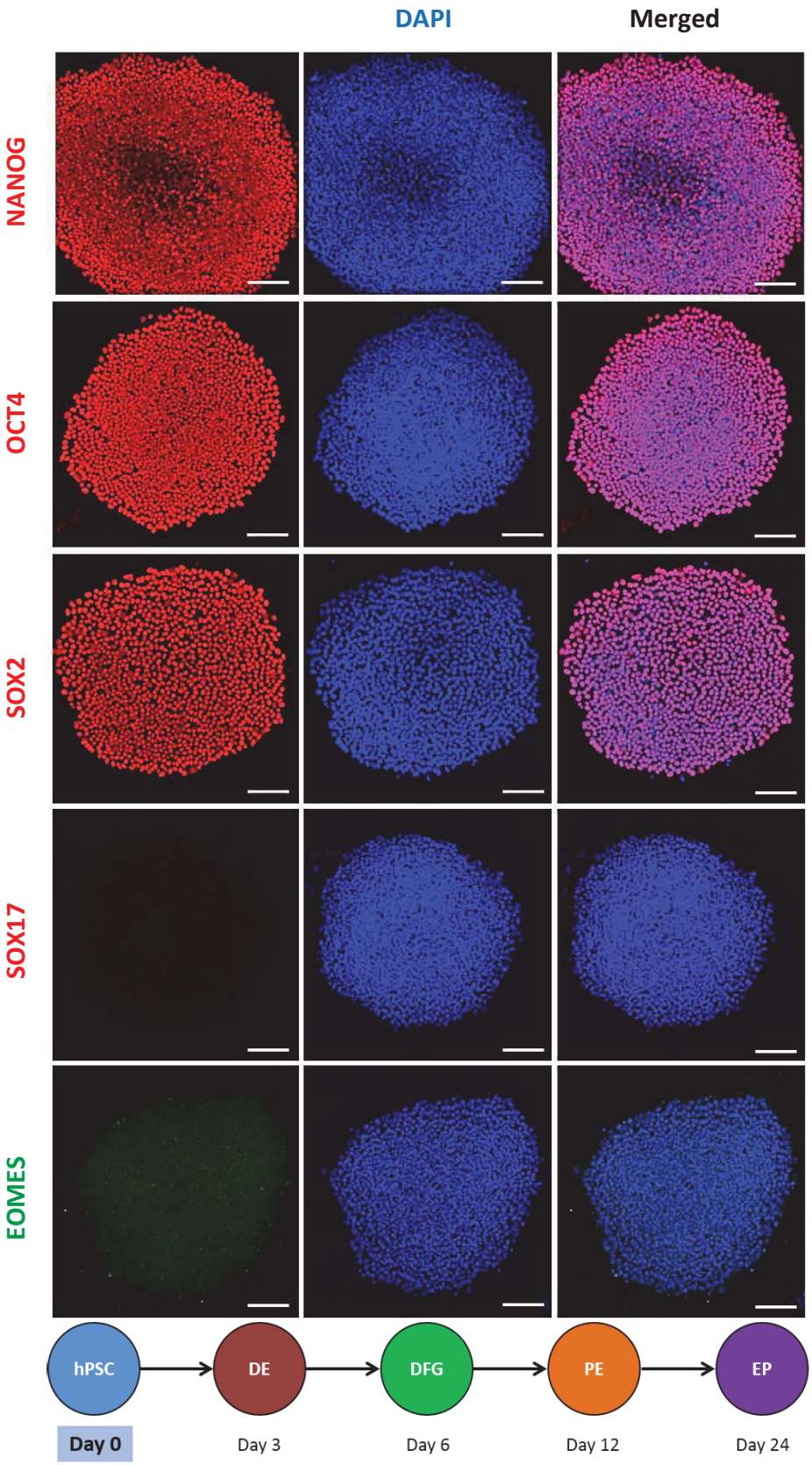


Figure 11. H9 cells are pluripotent and undifferentiated. Cells were grown in feeder-free culture conditions that maintain pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for pluripotency markers NANOG, OCT4 and SOX2 and early germ layer markers SOX17 and EOMES. Scale bar, 100 µm.

ICC analyses of DE differentiation after day 3 showed the up-regulation of endoderm marker SOX17 and early foregut marker FOXA2 accompanied by the down-regulation of pluripotency marker NANOG (Figure 12), indicating that the H9 cells successfully differentiated into the DE. FACS analyses revealed efficient DE formation with populations of 76.6% of SOX17+ cells and 67.7% of CXCR4+ cells (Figure 13). In addition, morphological analyses showed that the cells lost the compact colony morphology characteristic of stem cells and adapted a spread out "cobblestone-like" morphology indicating an epithelial-mesenchymal transition (EMT)-like phenomenon characteristic of endoderm cells (Figure 14).

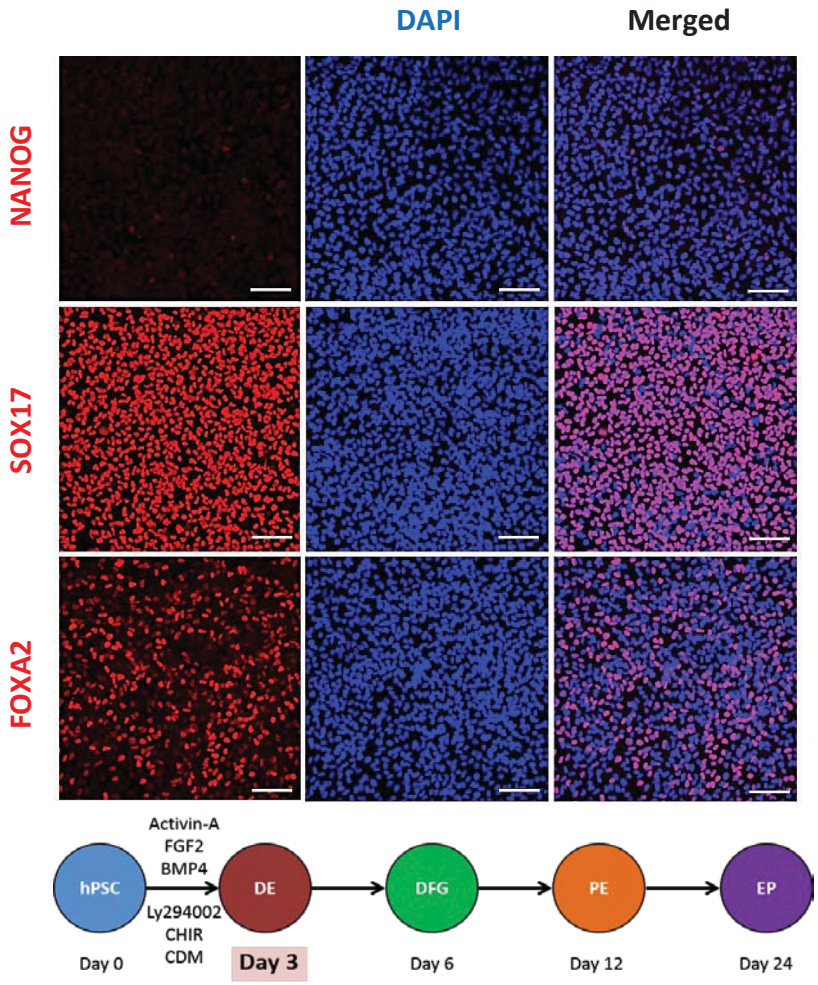


Figure 12. ICC of H9 cells differentiated into the DE. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG, as well as DE markers SOX17 and FOXA2. Scale bar, 100 µm.

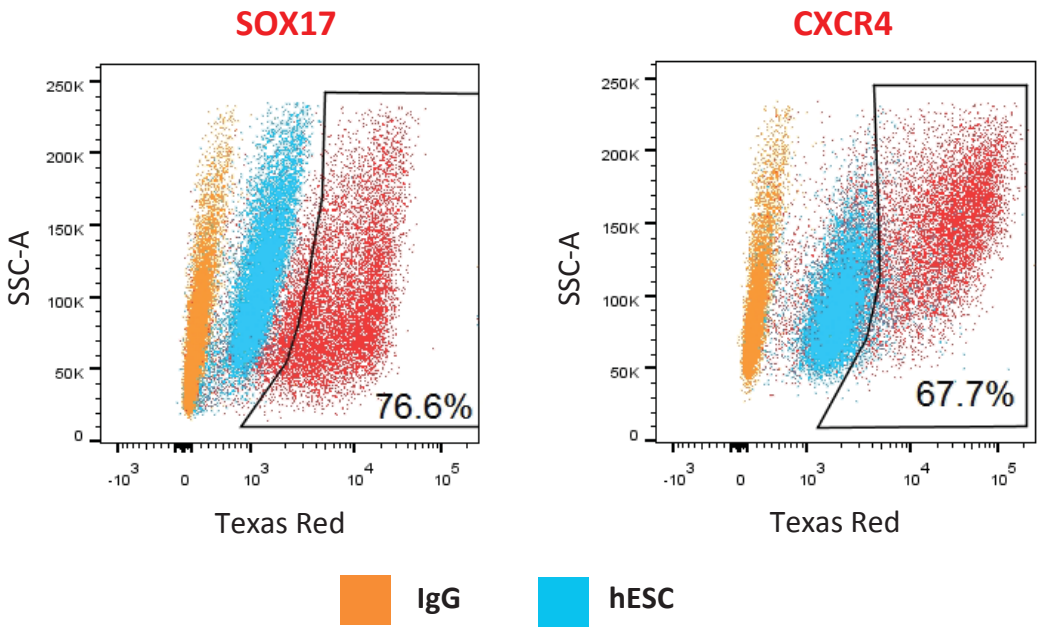


Figure 13. FACS of H9 cells differentiated into the DE. Cells were fixed on day 3 and were stained for DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

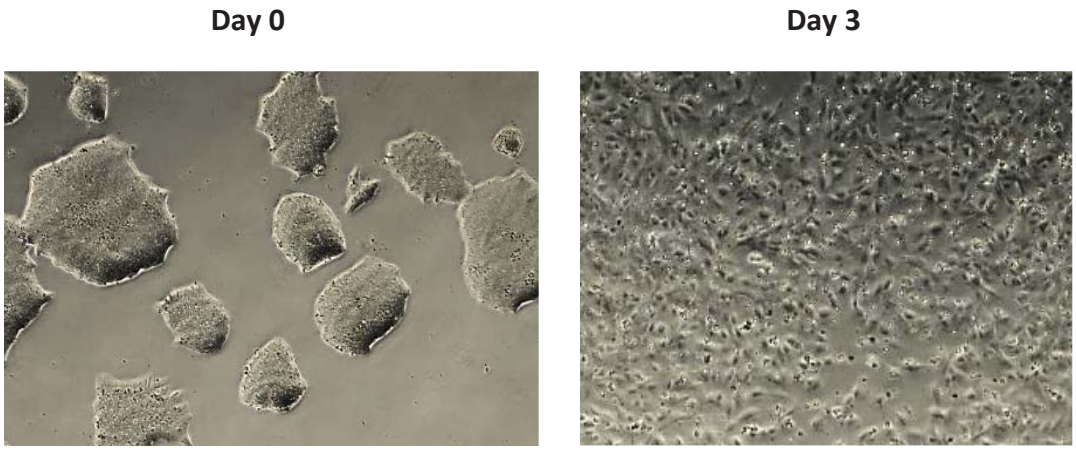


Figure 14. H9 exhibiting morphological changes upon differentiation into the DE. Bright field microscopy images showing the different morphologies between undifferentiated and DE cells.

Further differentiation toward the primitive gut tube on day 6 led to a near homogenous population of cells expressing key foregut markers such as FOXA2, HNF1B and HNF4A, as shown by ICC (Figure 15). Notably, the absence of HEX and CDX2 suggested a dorsal identity for these foregut cells (Figure 15). The cells continued to proliferate, resulting in a denser monolayer as the differentiation process progressed.

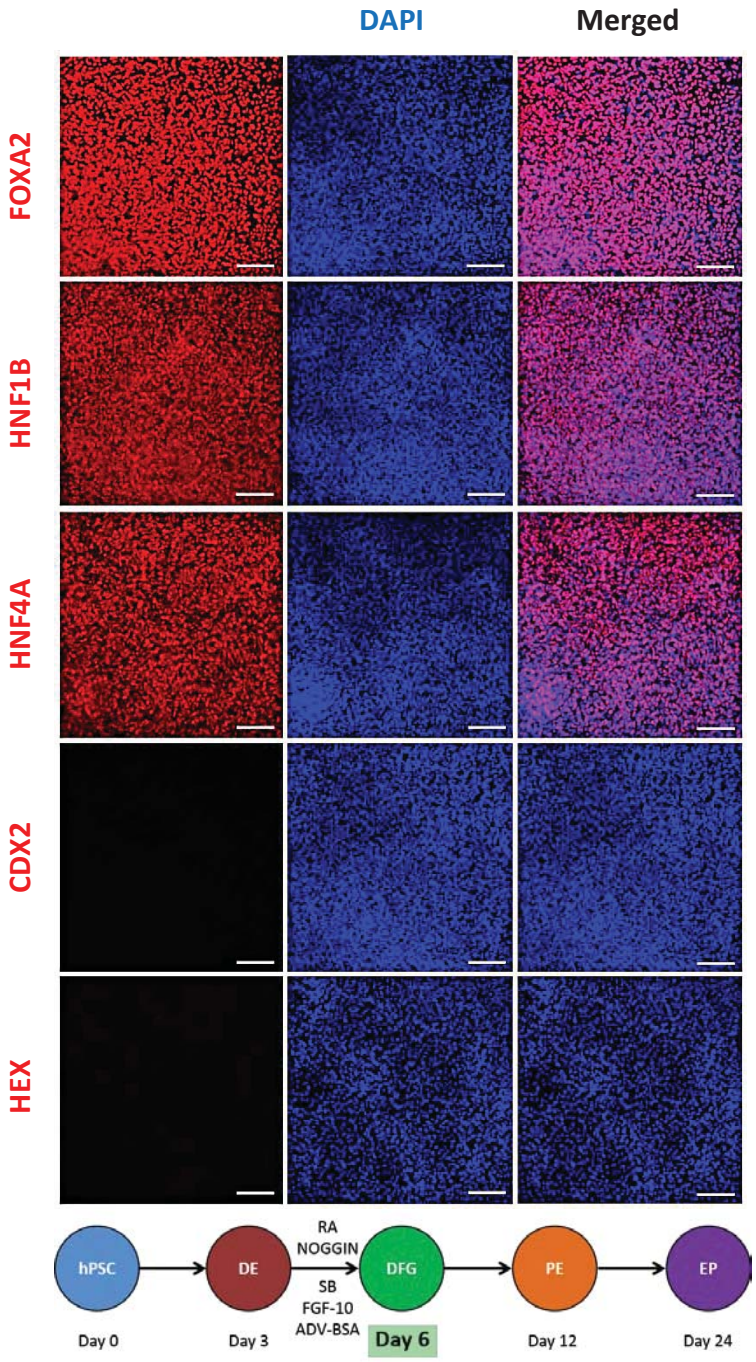


Figure 15. ICC of H9 cells differentiated into the primitive gut tube. Cells were grown in culture conditions that specified them toward the dorsal foregut and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for key markers FOXA2, HNF1B and HNF4A and negative markers CDX2 and HEX. Scale bar, 100 µm.

Three days later on day 9, the cells continued expressing key foregut markers FOXA2, HNF1B, SOX2 and HNF6 (Figure 16). The key pancreatic marker PDX1 was shown to be almost homogeneously expressed by the end of day 12, indicating acquisition of pancreatic fate (Figure 17), and FACS analyses confirmed this by showing a 75.9% population of PDX1+ cells (Figure 18).

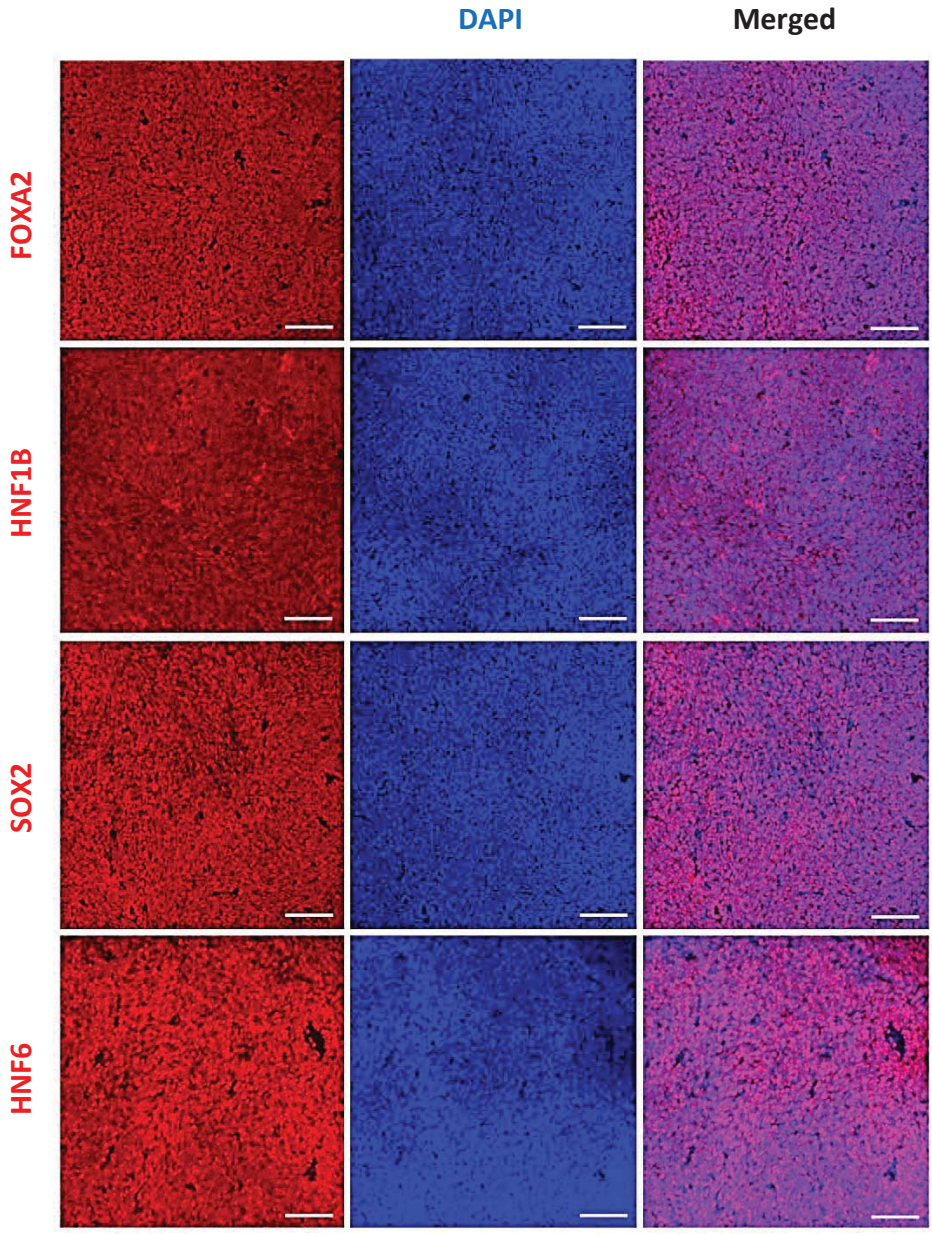


Figure 16. ICC of H9 cells differentiated into the posterior foregut. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for key markers FOXA2, HNF1B, SOX2 and HNF6. Scale bar, 100 µm.

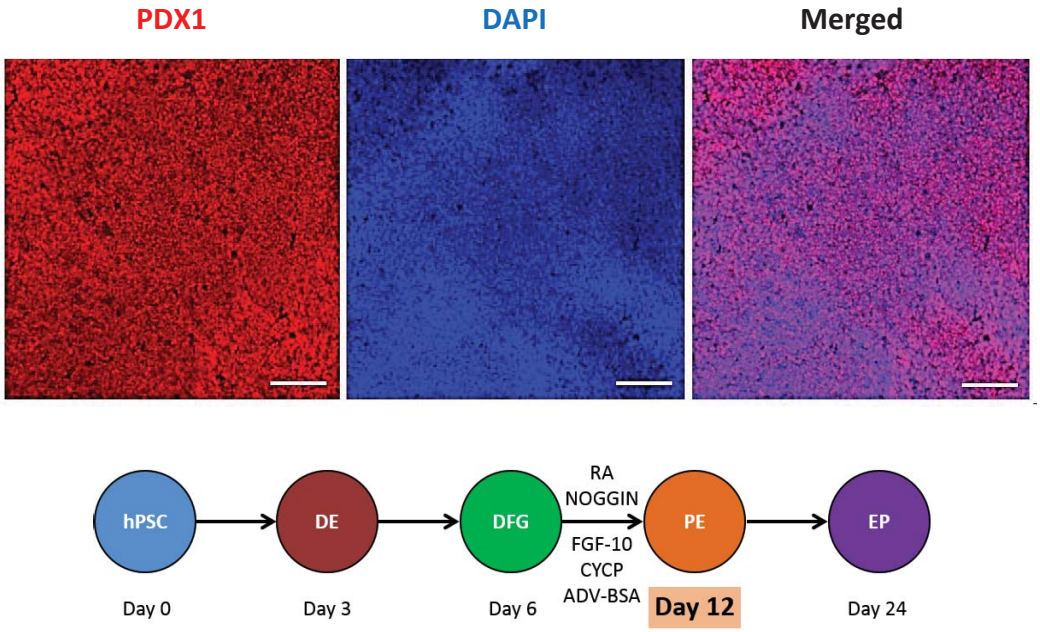


Figure 17. ICC of H9 cells differentiated into the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for the key pancreatic marker PDX1. Scale bar, 100 µm.

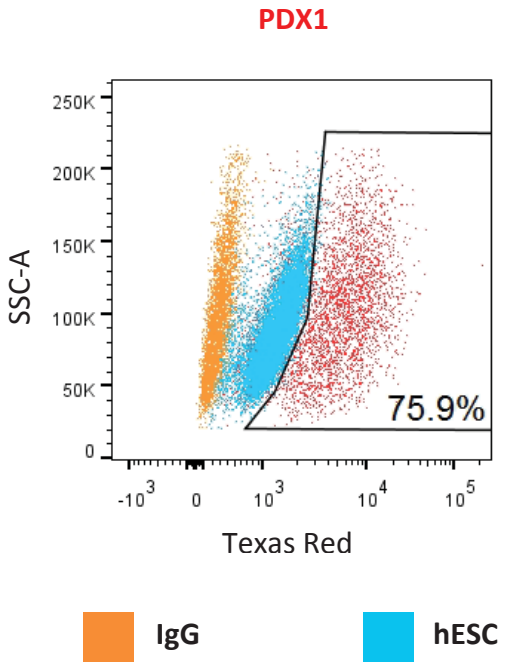


Figure 18. FACS of H9 cells differentiated into the pancreatic endoderm. Cells were fixed on day 12 and were stained for the key pancreatic marker, PDX1. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 15, NGN3 was shown to be expressed by ICC (Figure 19) and FACS, at a population of 12.1% (Figure 20). Endocrine progenitors on day 18 revealed an expression of the key pancreatic marker PDX1 and endocrine progenitor markers C-PEPTIDE, SST and GCG via ICC (Figure 21). From day 18 to day 24 as the cells matured into immature β-cells, ICC showed an increased number of cells expressing C-PEPTIDE, SST and GCG (Figure 22). FACS analyses revealed populations of 10.4% C-PEPTIDE+ cells, 1.76% SST+ cells, and 3.79% GCG+ cells (Figure 23) and 7.9% of mono-hormonal C-PEPTIDE+ cells on day 24 (Figure 24).

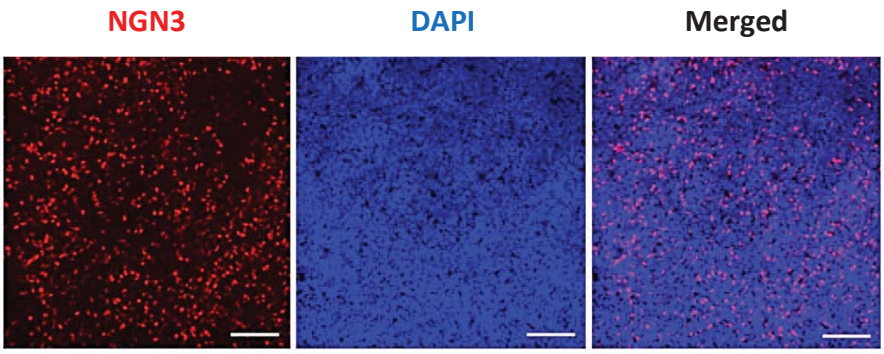


Figure 19. ICC of H9 cells on day 15. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 15 and were stained for NGN3. Scale bar, 100 µm.

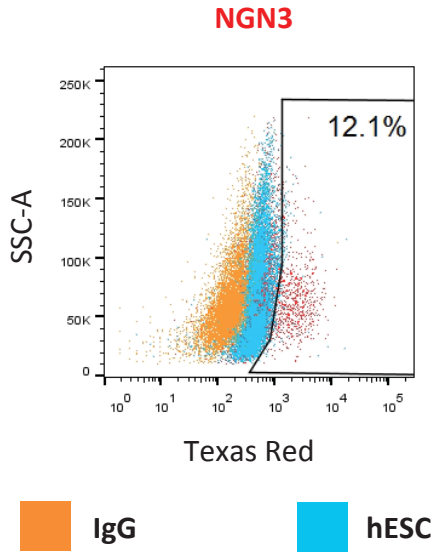


Figure 20. FACS of H9 cells on day 15. Cells were fixed on day 15 and were stained for NGN3. Data show results of one experiment that is representative of 3 independent experiments.

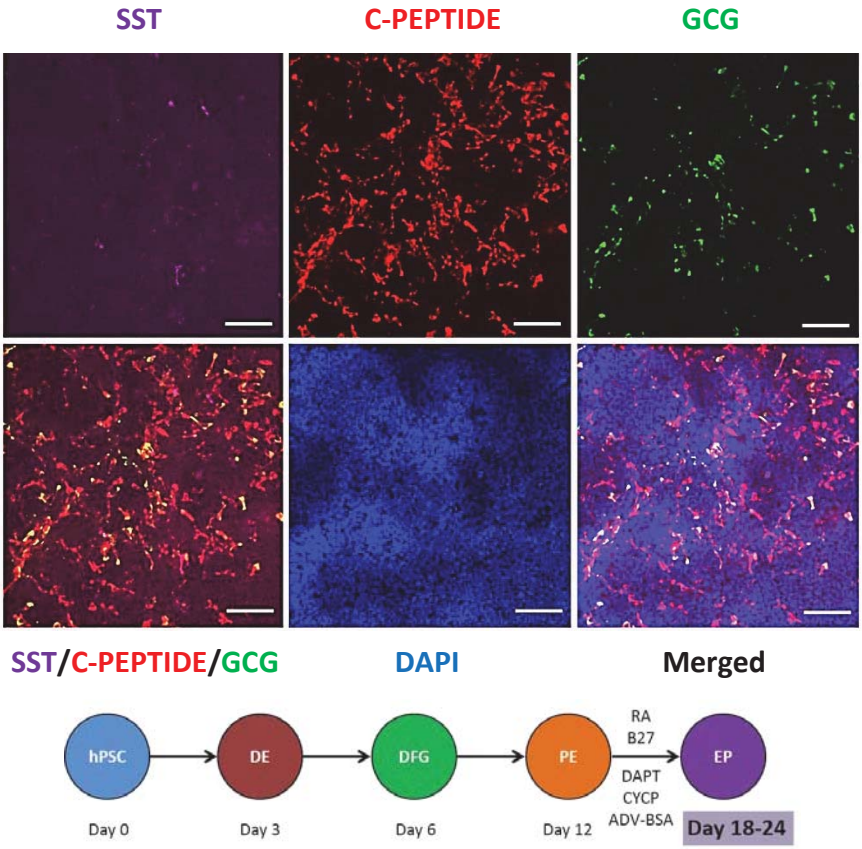


Figure 21. ICC of H9 cells on day 18. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 18 and were stained for key markers SST, C-PEPTIDE and GCG. Scale bar, 100 µm.

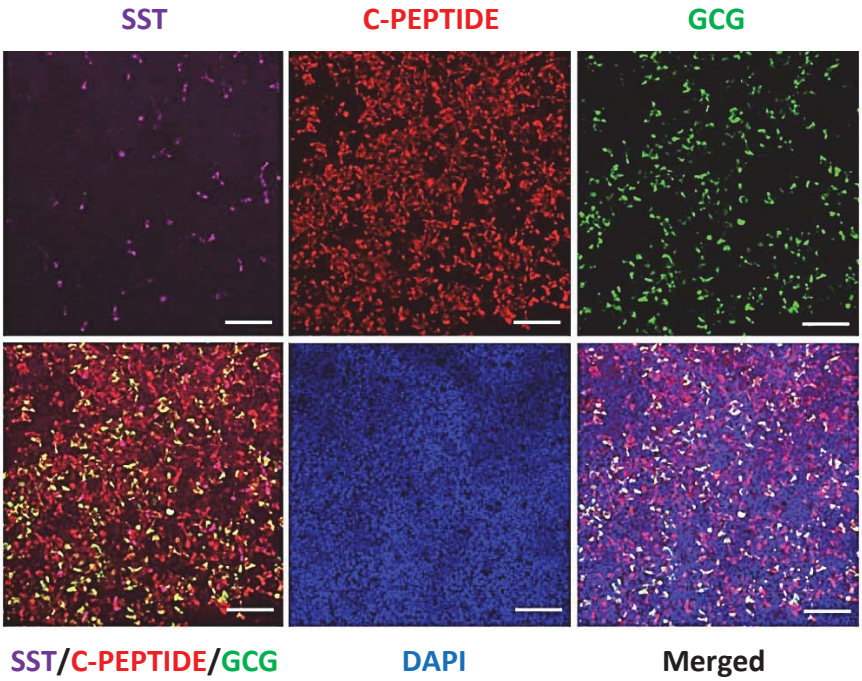


Figure 22. ICC of H9 cells differentiated into endocrine progenitors. Cells were grown in culture conditions that that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers SST, C-PEPTIDE and GCG. Scale bar, 100 µm.

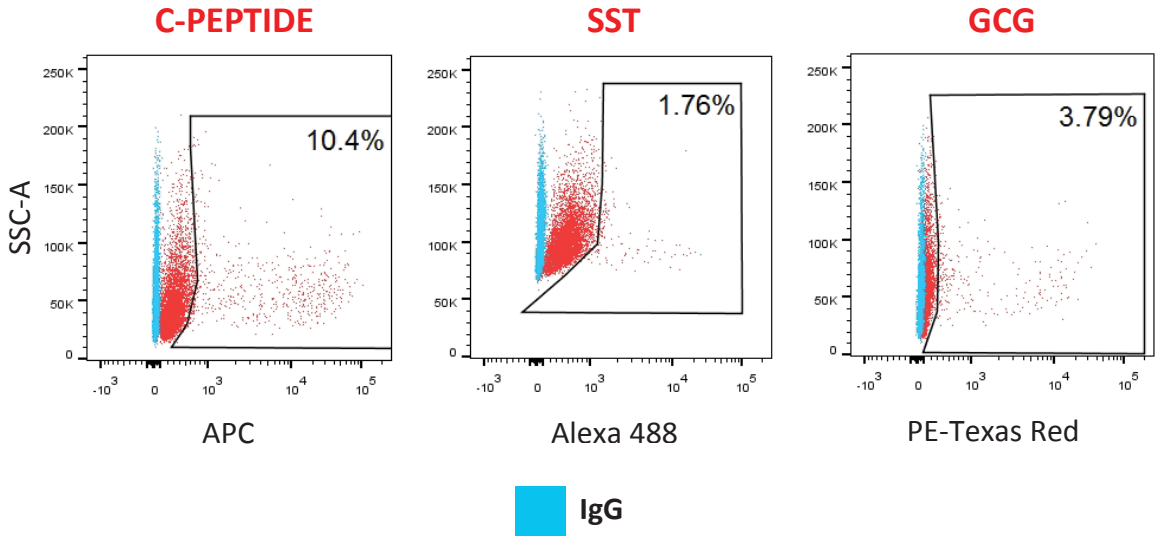


Figure 23. FACS of H9 cells differentiated into endocrine progenitors. Cells were fixed on day 24 and triple-stained for key markers, C-PEPTIDE, SST and GCG. Cell populations represent a combination of both poly- and mono-hormonal cells. Data show results of one experiment that is representative of 3 independent experiments.

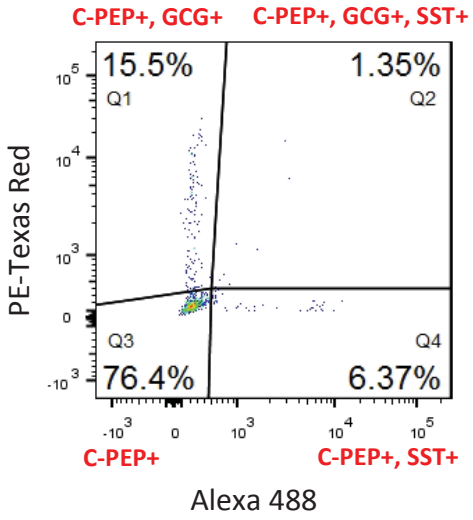


Figure 24. FACS of mono-hormonal H9 cells on day 24. Cell population consists of poly- or mono-hormonal cells which are C-PEPTIDE+ (C-PEPTIDE population of 10.4% from Figure 23). Q1 represents C-PEPTIDE+ and GCG+ poly-hormonal cells; Q2 represents C-PEPTIDE+, SST+ and GCG+ poly-hormonal cells; Q3 represents C-PEPTIDE+ mono-hormonal cells; Q4 represents C-PEPTIDE+ and SST+ poly-hormonal cells. Data show results of one experiment that is representative of 3 independent experiments.

Further qRT-PCR analyses not only validated the expression of key markers described earlier at each time point, but also provided an overview of the expression patterns of each marker over the course of pancreatic specification (Figure 25).

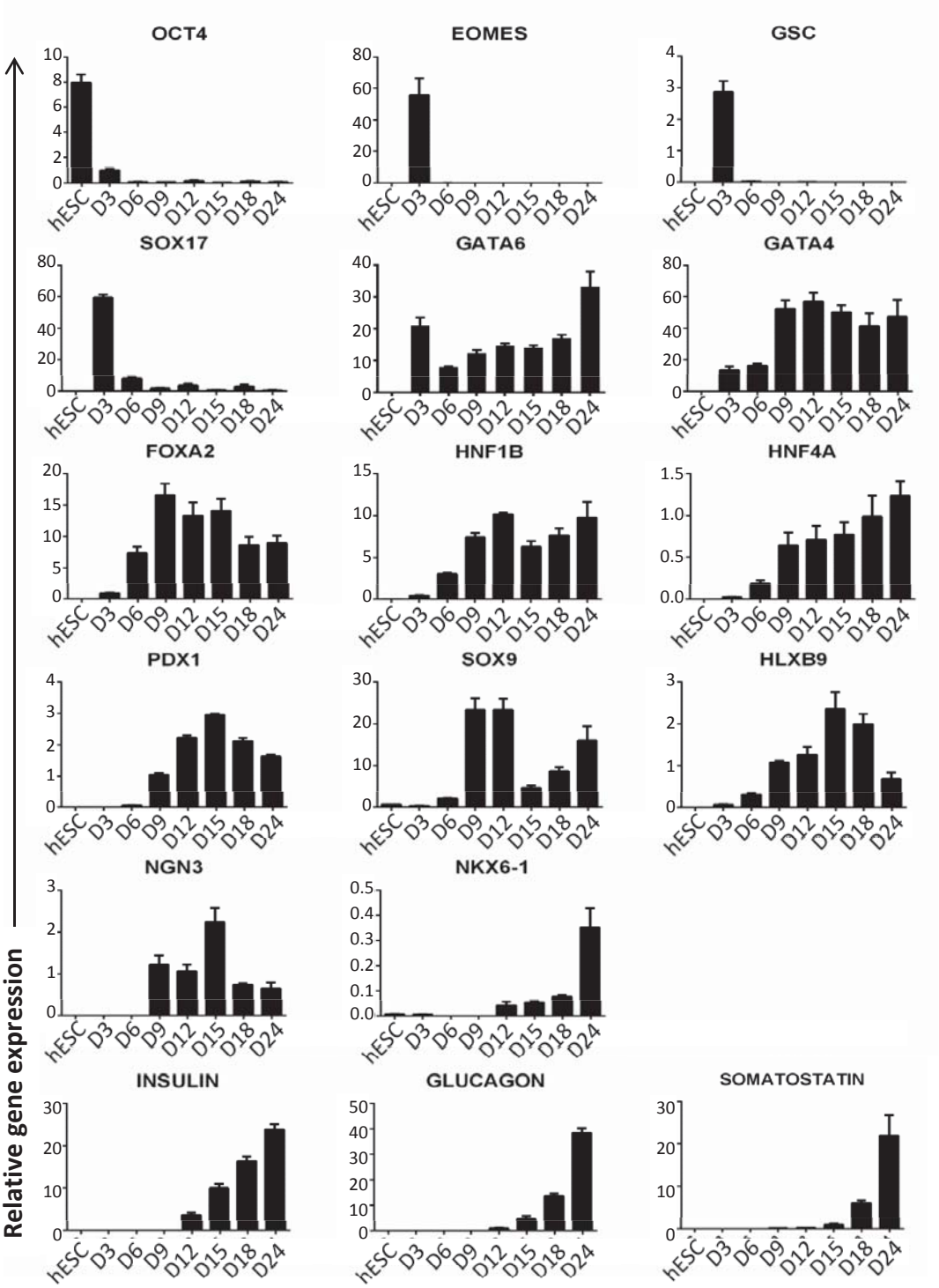


Figure 25. qRT-PCR analyses of H9 cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation. Values are relative to the housekeeping gene *PBGD*.

To investigate the functional potential of the endocrine progenitors generated from the H9 hESCs, I performed C-peptide ELISA on cells that have undergone the 24-day differentiation protocol derived from the lab. C-peptide is often used as an alternative measurement for insulin at protein level, including ELISA, ICC and FACS. A major reason for this is because cells take up insulin present in the cell culture media, which may lead to false positive signals of insulin expression and contents in cells. The proteolytic cleavage of proinsulin prior to secretion produces the mature insulin molecule and the connecting peptide (C-peptide). As C-peptide is secreted in equimolar quantities to insulin, it serves as a substitute for the measurement of insulin. The 10% of H9-derived insulin-expressing cells seemed to elicit an inverse C-peptide releasing response upon glucose stimulation (Figure 26).

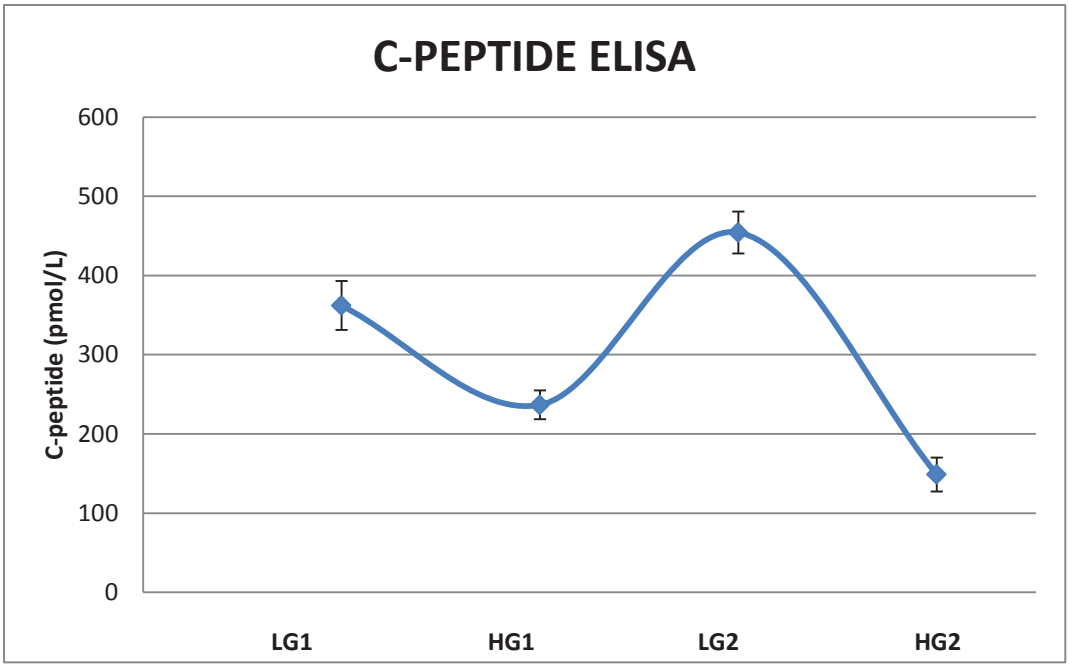


Figure 26. C-peptide secretion upon glucose stimulation on day 24. Cells were differentiated using the lab-derived protocol and assayed on day 24. Data are presented as the average of 3 biological replicates of one experiment and representative of three independent experiments. Error bars indicate standard deviation. LG1 is first incubation of low glucose; HG1 is first incubation of high glucose; LG2 is second incubation of low glucose; HG2 is second incubation of high glucose.

To enhance the reproducibility of the results, a second hPSC line was introduced into the project. This was a hIPSC line derived previously in the lab, named FSPS13.B. Pancreatic differentiation was performed on these cells and qRT-PCR and FACS analyses were done at various time points of the protocol. Expression patterns of the key genes analysed via qRT-PCR corroborated those from H9 cells (Figure 27). FACS analyses revealed 70.8% of SOX17+ cells and 83.7% of CXCR4+ cells on day 3 (Figure 28), and 86.8% of PDX1+ cells on day 12 (Figure 29), indicating similar differentiation efficiencies toward the DE and pancreatic endoderm lineages between H9 and FSPS13.B cells.

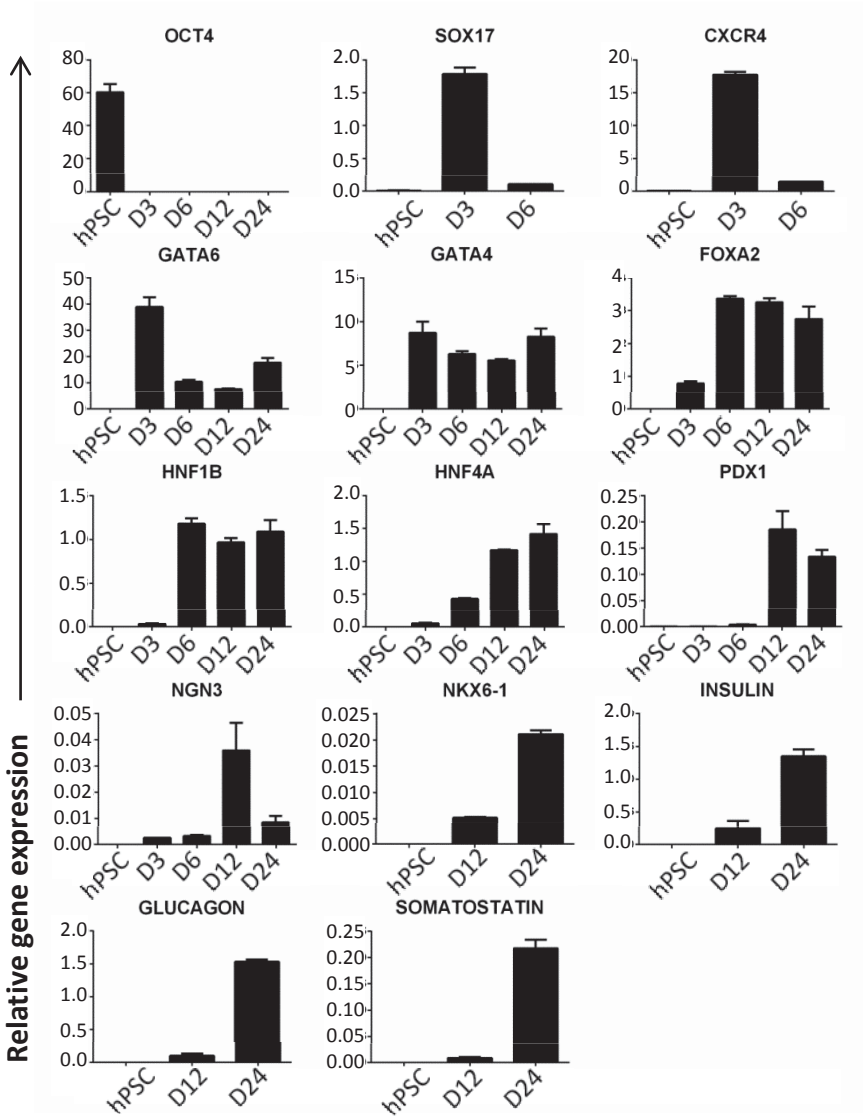


Figure 27. qRT-PCR analyses of FSPS13.B cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

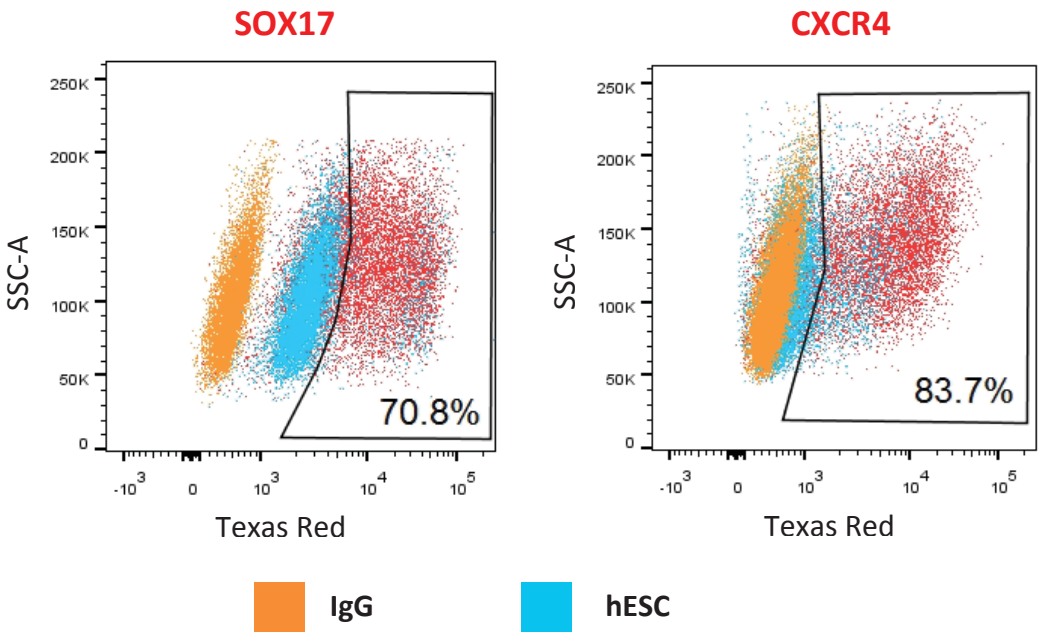


Figure 28. FACS of FSPS13.B cells differentiated into the DE. Cells were grown in culture conditions that specified them toward the DE lineage and then analysed by FACS. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

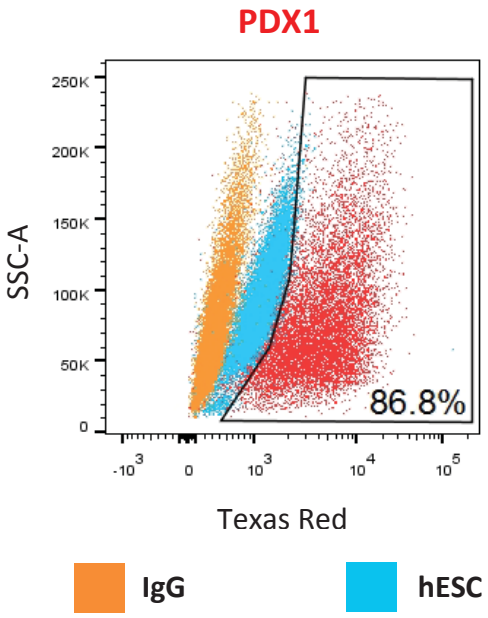


Figure 29. FACS of FPS13.B cells differentiated into the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and then analysed by FACS. Cells were fixed on day 12 and were stained for the pancreatic marker, PDX1. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 24, however, FSPS13.B displayed lower endocrine progenitor differentiation efficiency as compared with H9, with FACS analysis showing 5.75% of C-PEPTIDE+ cells (Figure 30). Due to the low number of C-PEPTIDE+ cells, I could not perform triple staining of C-PEPTIDE, SST and GCG to determine the percentage of mono-hormonal and poly-hormonal cells as was done on H9 cells (Figure 24)

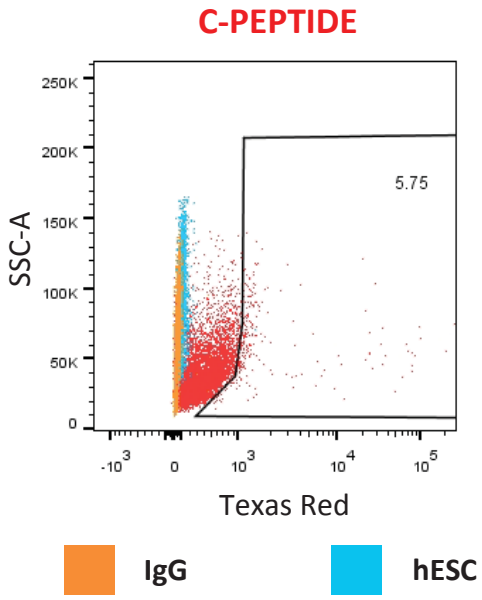


Figure 30. FACS of FPS13.B cells differentiated into the endocrine progenitors. Cells were grown in culture conditions that specified them toward the endocrine progenitor lineage and then analysed by FACS. Cells were fixed on day 24 and were stained for the endocrine marker, C-PEPTIDE. Data show results of one experiment that is representative of at least 3 independent experiments.

Together, except for the C-peptide ELISA, these results recapitulate pancreatic differentiation that was previously published (Cho et al., 2012), indicating to a large extent, the consistency of the pancreatic differentiation protocol in performing pancreatic differentiation on both hESC and hiPSC lines.

3.1.2. GATA6 is up-regulated upon definitive endoderm formation

To better understand of the role of GATA6 in the development of the human pancreas, I closely investigated the expression profile of GATA6 over the course of pancreatic specification. Using H9 cells, *in vitro* pancreatic differentiation was performed according to the protocol described in Chapters 2.1.3 and 2.1.4.

On day 0 of differentiation, GATA6 expression was demonstrated to be absent in undifferentiated and pluripotent cells, as shown by ICC (Figure 31) and qRT-PCR (Figure 33).

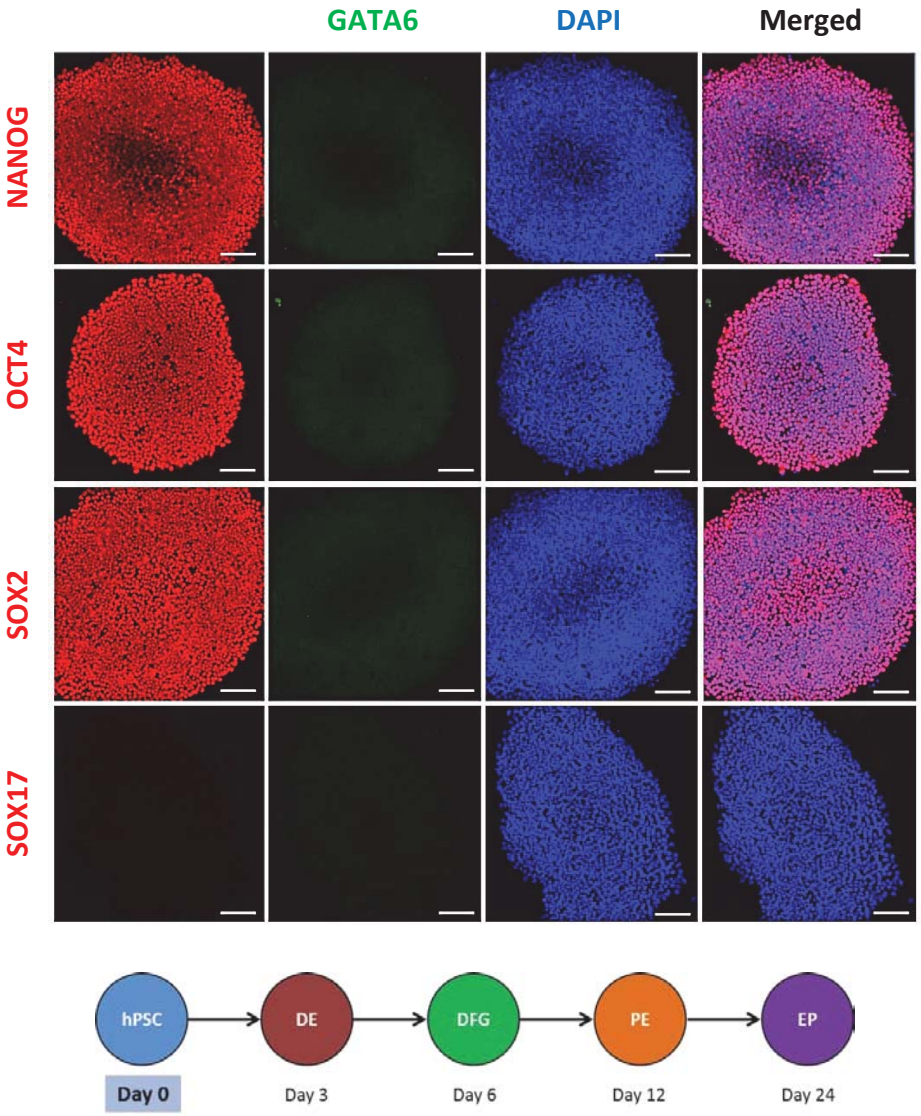


Figure 31. GATA6 expression is negligible in undifferentiated state. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for pluripotency markers NANOG, OCT4 and SOX2 and early DE marker SOX17. Scale bar, 100 µm.

On day 3, ICC showed that GATA6 is highly expressed in the DE and is co-localised with key DE markers (Figure 32). The rapid up-regulation of *GATA6* on day 3 was confirmed by qRT-PCR (Figure 33). FACS analysis showed that 97.8% of the cells on day 3 were GATA6+ (Figure 34).

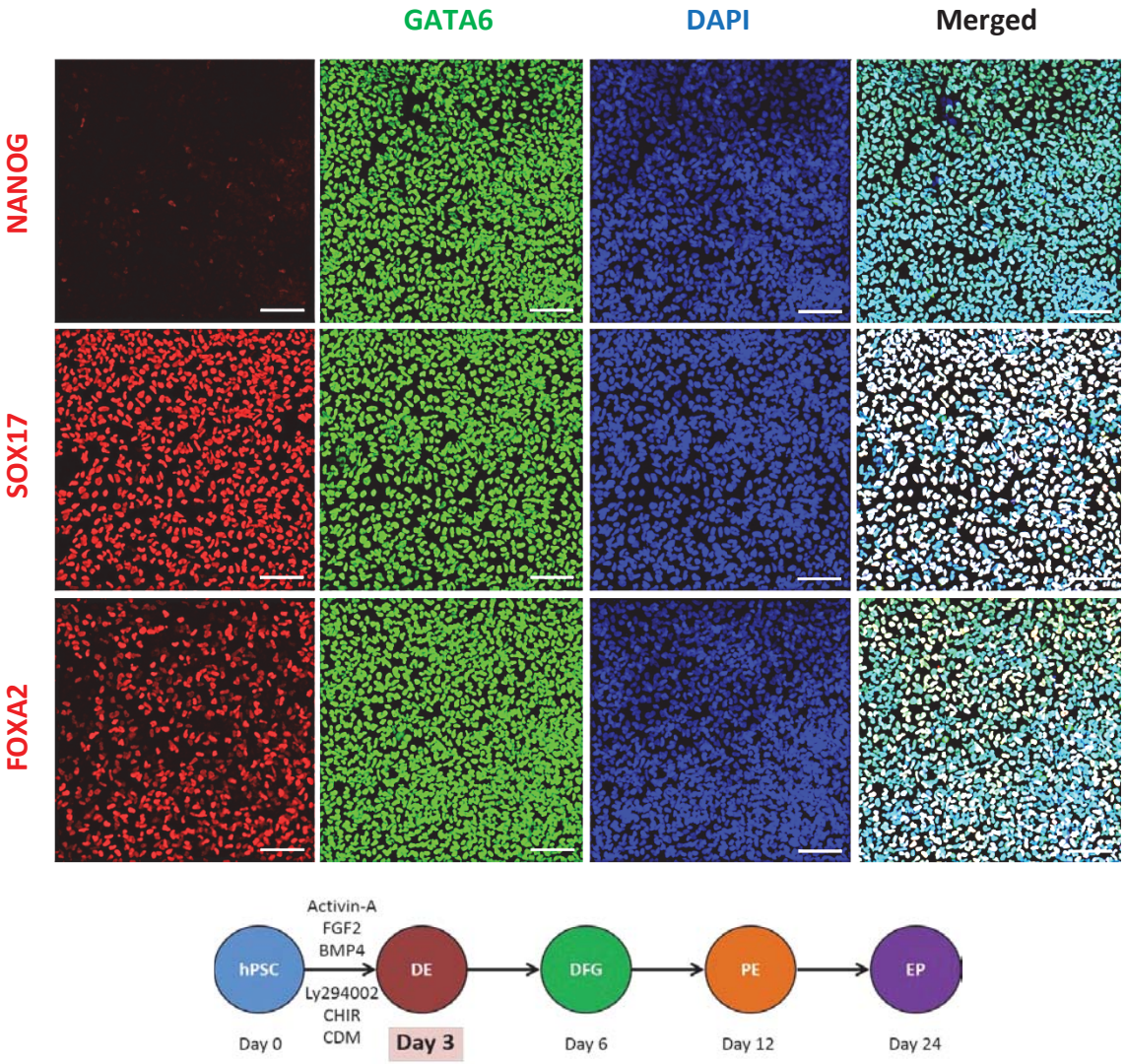


Figure 32. GATA6 is co-expressed with key DE markers. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG, as well as DE markers SOX17 and FOXA2. Scale bar, 100 µm.

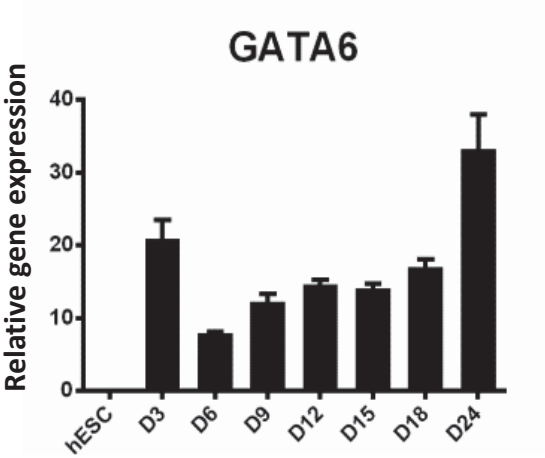


Figure 33. qRT-PCR analyses of *GATA6* expression levels in of H9 cells specified toward the pancreatic lineage. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation. Values are relative to the housekeeping gene *PBGD*.

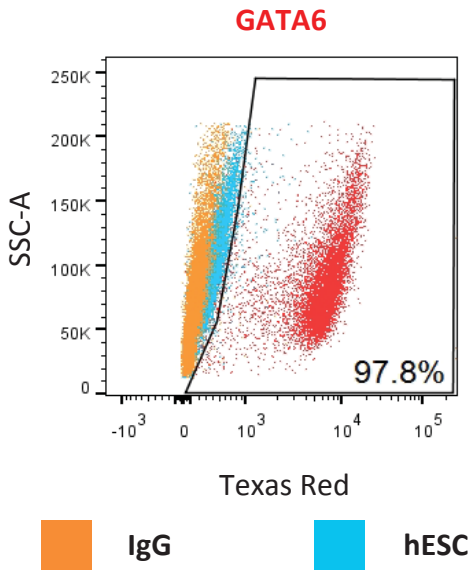


Figure 34. FACS of GATA6+ H9 cells at day 3. Cells were grown in culture conditions that specified them toward the DE lineage and then analysed by FACS. Cells were fixed on day 3 and were stained for GATA6. Data show results of one experiment that is representative of at least 3 independent experiments.

3.1.3. GATA6 is expressed throughout pancreatic development

On day 6, GATA6 continues to be expressed during pancreatic differentiation, although at lower levels compared with day 3 as indicated by the decreased staining intensity by ICC (Figure 35) and decreased relative expression level by qRT-PCR (Figure 33). On days 9 and 12, GATA6 remains expressed and co-localised with the key markers at each respective stage (Figure 36 and Figure 37). FACS analysis showed that 88.1% of cells on day 12 were GATA6+ (Figure 38).

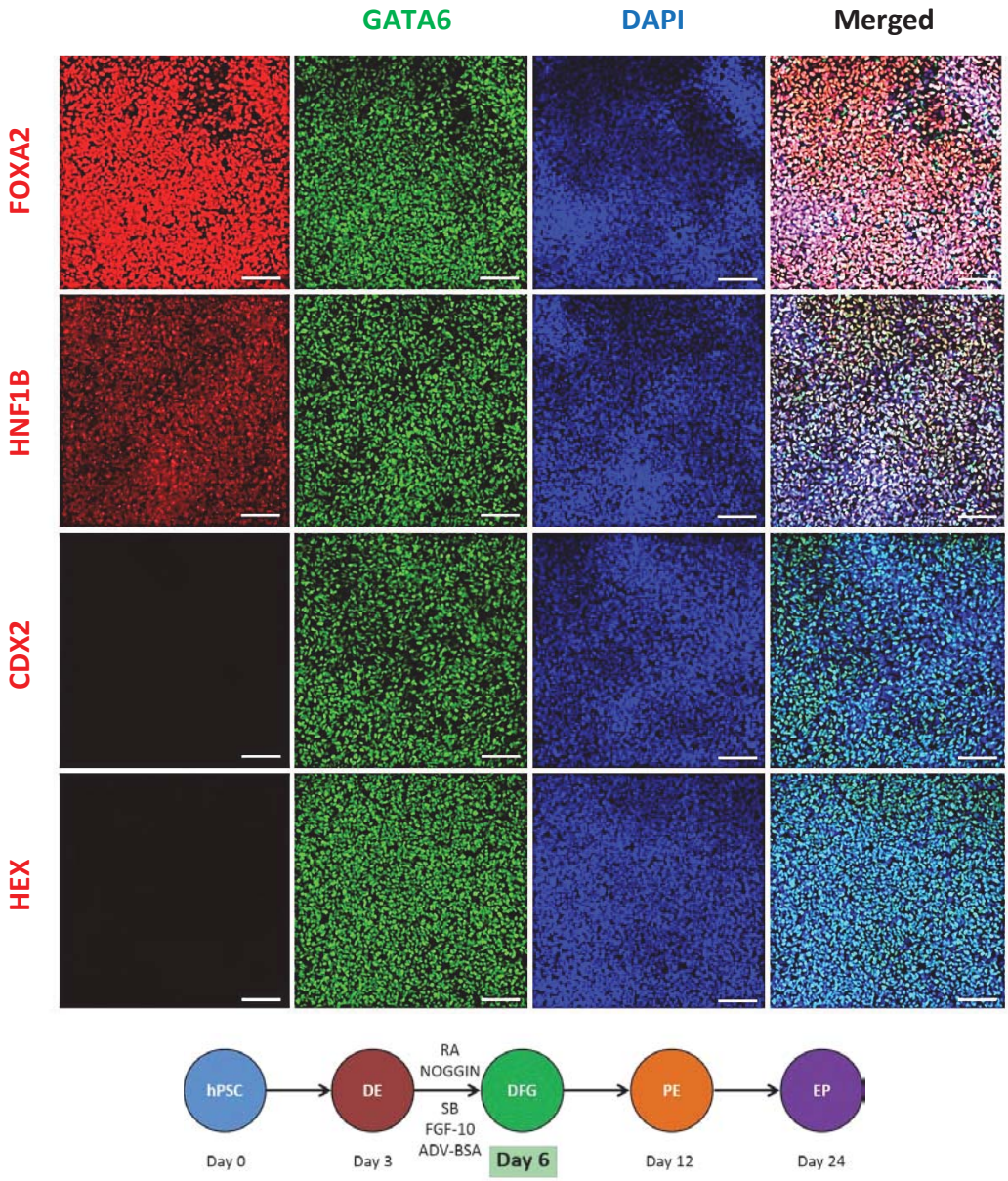


Figure 35. GATA6 is co-localised with key markers of the primitive gut tube. Cells were grown in culture conditions that specified them toward the dorsal foregut and analysed via immunofluorescence. Cells were fixed on day 6 and stained for key markers FOXA2 and HNF1B and negative markers CDX2 and HEX. Scale bar, 100 µm.

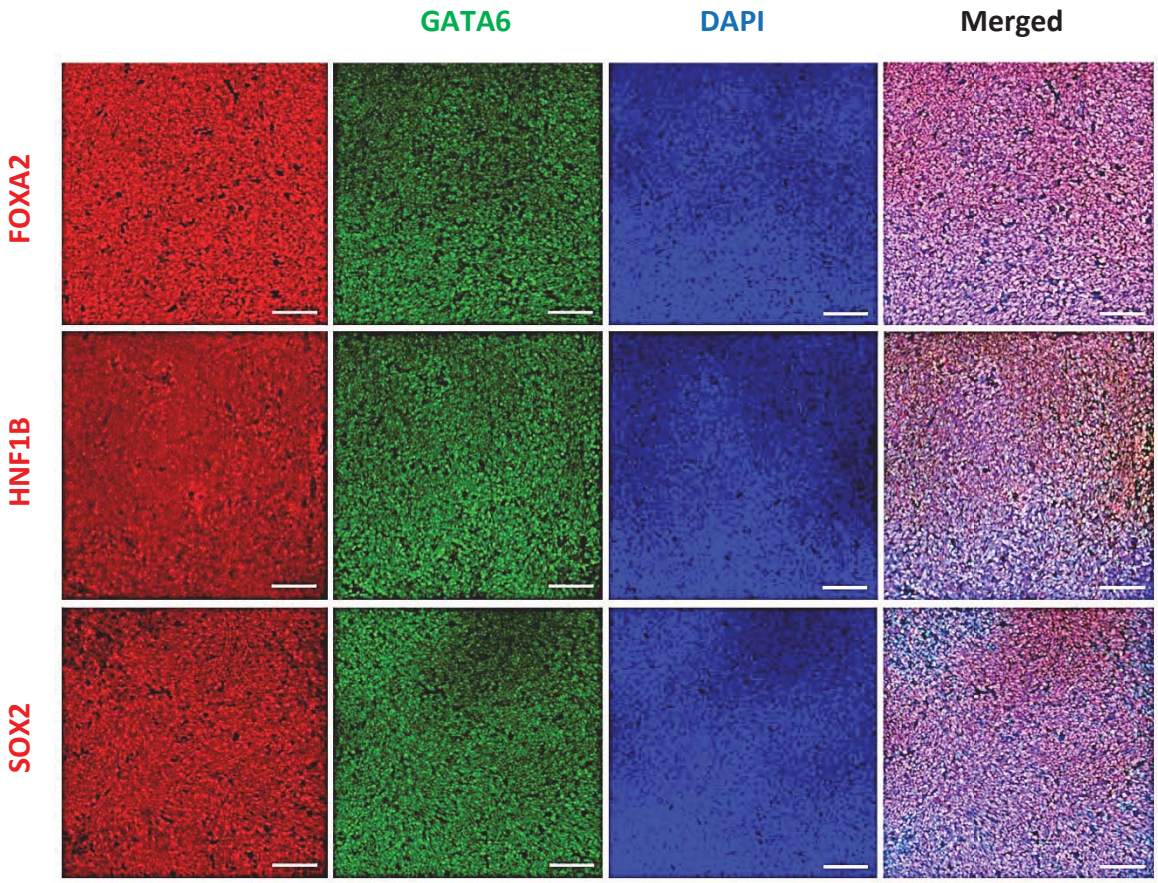


Figure 36. GATA6 is co-localised with key markers of the posterior foregut. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for key markers FOXA2, HNF1B and SOX2. Scale bar, 100 µm.

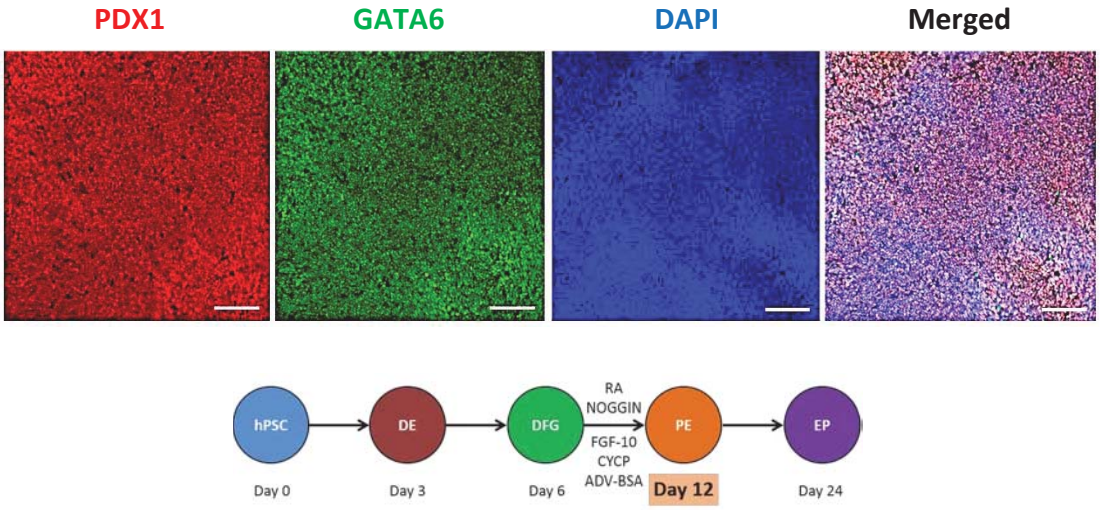


Figure 37. GATA6 is co-localised with key marker of the pancreatic endoderm. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for key pancreatic marker PDX1. Scale bar, 100 µm.

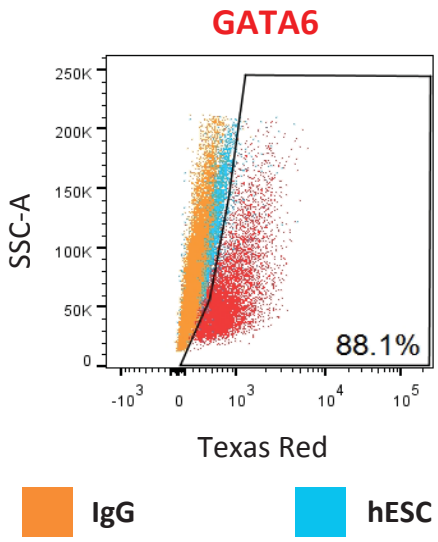


Figure 38. FACS of GATA6+ H9 cells at day 12. Cells were grown in culture conditions that specified them toward the pancreatic endoderm lineage and then analysed by FACS. Cells were fixed on day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least 3 independent experiments.

On day 24, GATA6 continues to be expressed at high levels as shown by ICC (Figure 39) and qRT-PCR (Figure 33). Together, these results suggest that GATA6 plays an important role during pancreatic specification.

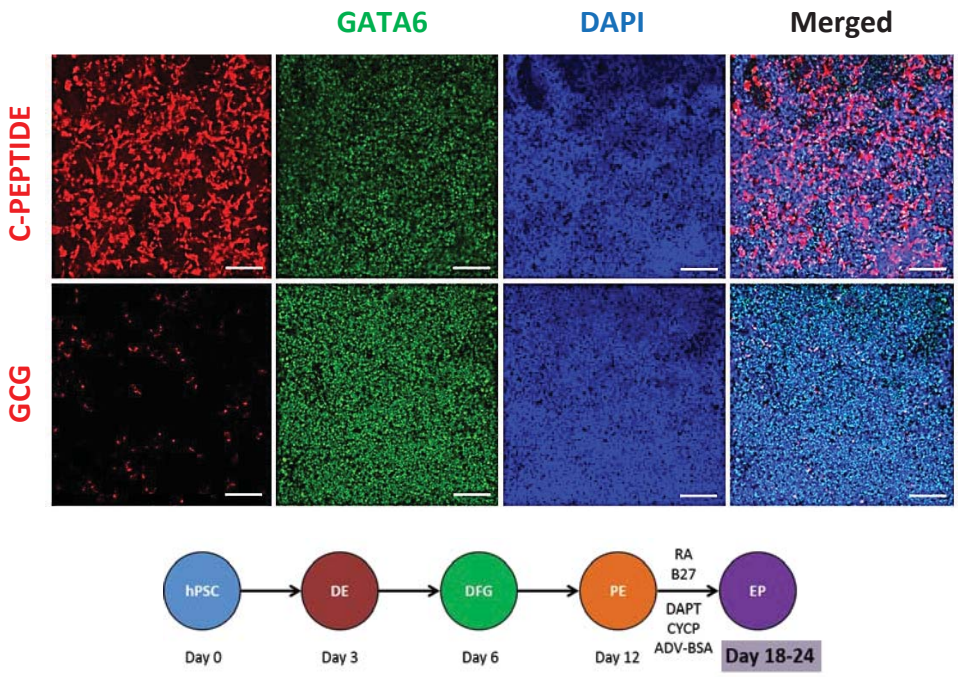


Figure 39. GATA6 is co-localised with key markers of endocrine progenitors. Cells were grown in culture conditions that allowed them to mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers C-PEPTIDE and GCG. Scale bar, 100 µm.

3.2. Successful derivation of *GATA6* mutant lines

Using an efficient endocrine progenitor differentiation protocol, I have determined the expression profile of the transcription factor *GATA6*, where its expression is first detected during the formation of the DE and continues to be expressed throughout human pancreas development. Next, I sought to generate heterozygous and homozygous *GATA6* mutant cell lines in order to further investigate the role of *GATA6* in the formation of the human pancreas. To generate *GATA6* mutants, I employed TALENs to induce mutations at two specific target sites that were selected based on their close proximity downstream of the first and second start codons of the *GATA6* gene to avoid generating partial protein products.

3.2.1. NHEJ pathway

To generate *GATA6* mutants via the NHEJ pathway, I employed TALENs to induce mutations at both target sites in the *GATA6* gene as described in Chapter 2.3.1 in both H9 and FSPS13.B cell hPSC lines. In the first instance, H9 cells were targeted at both *GATA6* TALEN1 and TALEN2 sites (Figure 40) and as described in Chapters 2.3.1 and 2.4.1. After TALEN targeting via electroporation, antibiotic selection and recovery, approximately 20-80 colonies derived from single cells were present in each culture dish. To reveal potential heterozygous or homozygous clones, restriction digest was performed to screen for mutations. Genomic DNA extracted from each colony was digested using restriction enzymes AfeI and PstI for TALEN1 and TALEN2 respectively. Colonies that were successfully cut by the TALENs and incorporated mutations upon DNA repair, and thus and remained undigested due to a disruption within the restriction enzyme site. Colonies that did not harbour any mutations have the restriction enzyme site intact, and therefore resulted in smaller and more numerous DNA fragments. This is apparent in Figure 41 where colonies numbered in red are colonies with a disrupted AfeI site showing a band resembling the undigested wild-type band, hence indicating the occurrence of a likely mutational event in an allele. These clones also showed the PCR fragments resulting from AfeI digestion on a wild-type allele, suggesting that they were likely heterozygous *GATA6* mutants.

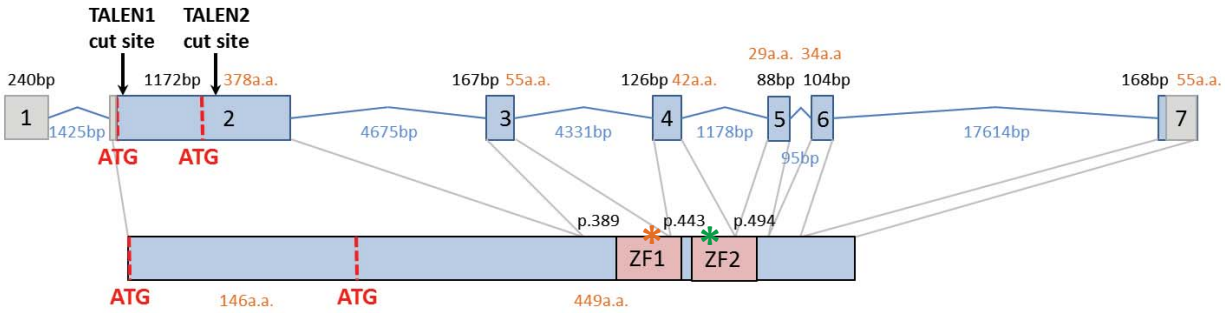


Figure 40. Schematic of TALEN1 and TALEN2 cut sites on the GATA6 locus. Top row represents *GATA6* cDNA showing the exons as boxes numbered accordingly and introns as lines. Grey regions of the exons represent non-coding regions while blue regions represent coding regions. The number of base pairs in each exon is shown in black text; the number of base pairs in each intron is shown in blue text; number of amino acids in each exon is shown in orange text. Bottom row illustrates the GATA6 protein and the grey lines connecting the cDNA to protein represents the corresponding area of the protein that each exon codes for. Protein translation commencing from the first ATG forms the long isoform of GATA6, which has a length of 595 amino acids. Protein translation commencing from the second ATG forms the short isoform of GATA6, which has a length of 449 amino acids. Green and orange asterisks indicate the locations of Patient A, *GATA6*$^{R465C/+}$ and Patient B, *GATA6*$^{c.1136\text{-}2A>G/+}$ mutations respectively.

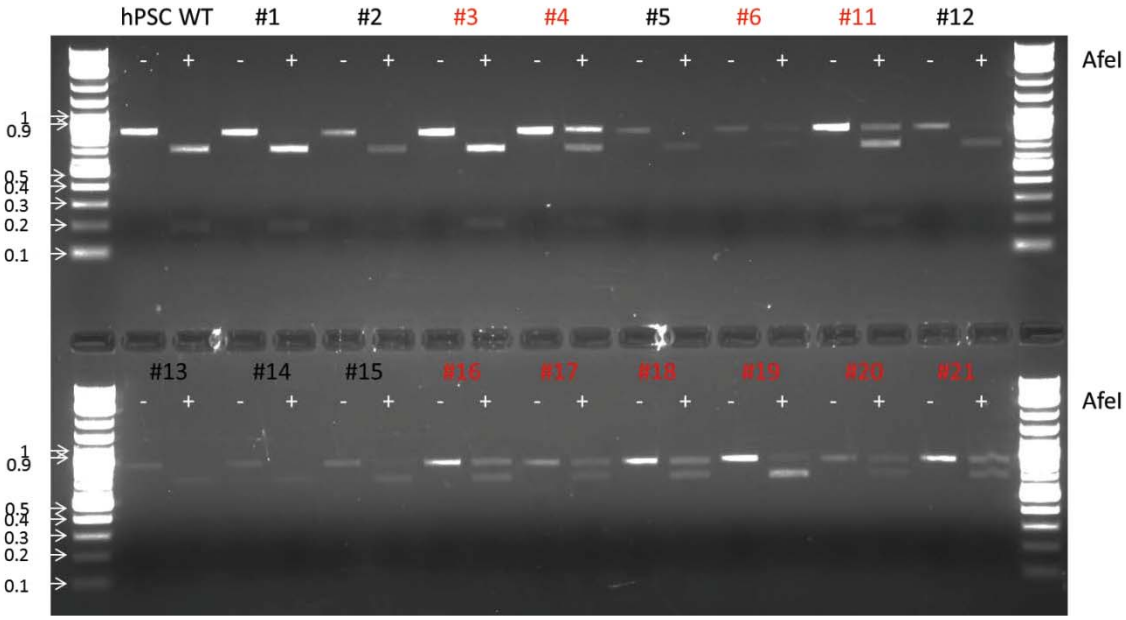


Figure 41. Representative DNA agarose gel picture of colonies screened via restriction enzyme digest. Genomic DNA was extracted from picked colonies and subjected to restriction digest using AfeI. Undigested and digested DNAs were run simultaneously. Colonies numbered red indicate the presence of heterozygous mutations while colonies numbered in black indicate wild-type.

24 clones were screened for each TALEN1 and TALEN2 targeting, with an observed cutting efficiency of approximately 80% at either targeting site, and a cutting efficiency of approximately 5% on both alleles at the TALEN2 target site. Unfortunately, no homozygous mutant for the TALEN1 cut site was generated in H9 cells. The colonies were then sequenced via Sanger sequencing to confirm their genotype. One heterozygous mutant derived from TALEN1 and TALEN2 target sites and one homozygous mutant derived from TALEN2 target site that harboured out of frame frameshift mutations resulting in a premature stop codon were selected for further western blotting validation (Figure 42). The TALEN-derived mutant lines were labelled according to their TALEN cut site (e.g. T1 for TALEN1 cut site or T2 for TALEN2 cut site) followed by the colony number.

To control for any potential off-target effects, I selected one colony that had been treated similarly to the mutants i.e. had undergone the TALEN targeting, antibiotic selection and re-growth process but harboured no observable mutations (Figure 42). All the TALEN-generated lines were karyotyped regularly to ensure chromosomal stability.

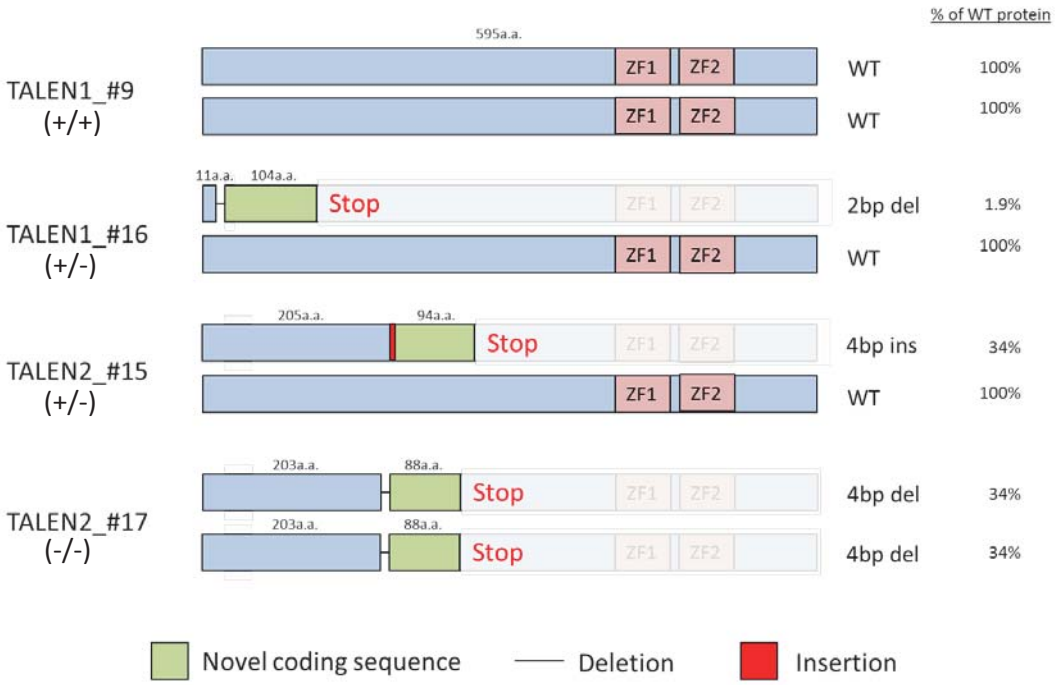


Figure 42. Schematic of selected H9 TALEN-derived *GATA6* cell lines. Four TALEN-derived cell lines were selected for use in downstream experiments. TALEN1_#9 is a TALEN-targeted but wild-type line; TALEN1_#16 contains a 2 base pair deletion frameshift mutation at the TALEN1 cut site; TALEN2_#15 contains a 4 base pair insertion frameshift mutation at the TALEN2 cut site; TALEN2_#17 contains an identical 4 base pair deletion on both *GATA6* alleles at the TALEN 2 cut site. All frameshift mutations were out of frame resulting in a premature stop codon.

Next, I performed western blotting to investigate GATA6 protein level in the TALEN-derived mutant lines. Cells were harvested on day 0 (D0) for negative control and day 3 (D3), as GATA6 is highly expressed on this day. Two antibodies, one recognising the N-terminus of the GATA6 protein and the other recognising the C-terminus of the GATA6 protein, were used. Expectedly, the N-terminal GATA6 antibody detected a partial protein product (PPP) for the mutants derived from the TALEN2 cut site, confirming the presence a truncated GATA6 protein caused by a premature stop codon in one *GATA6* allele (T2_#15) or both *GATA6* alleles (T2_#17) (Figure 43). The absence of GATA6 protein for T2_#17 mutant line using the C-terminal GATA6 antibody confirmed that this line was a homozygous mutant (Figure 44). The wild-type *GATA6* allele of heterozygous mutants T1_#16 and T2_#15 appeared similar to the wild-type lines H9 and T1_#9 (Figure 43 and Figure 44).

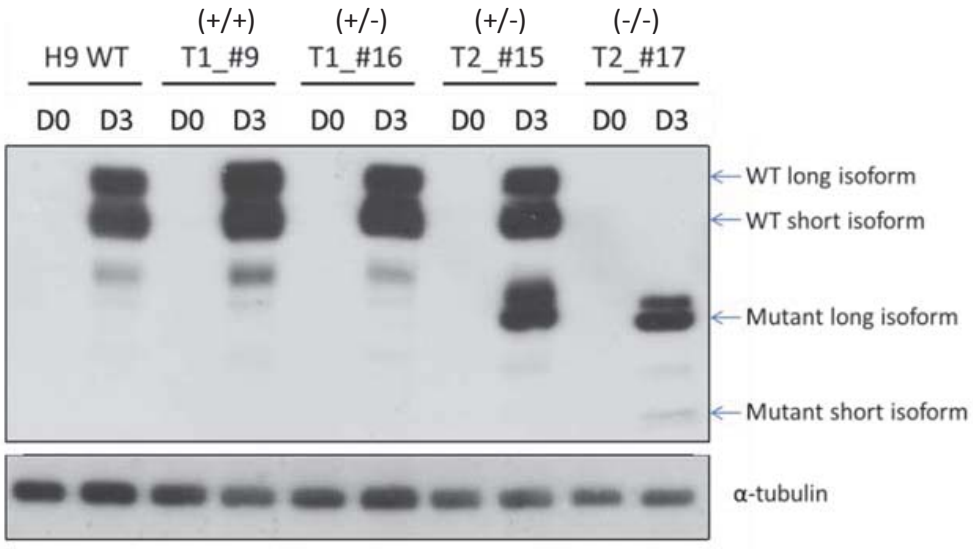


Figure 43. Western blot analysis of GATA6 protein levels in TALEN-derived H9 mutant lines using an N-terminal GATA6 antibody. Cells were harvested on day 0 (undifferentiated) and day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively; the partial protein products for T2_#15 and T2_#17 are approximately 30 kDa and 18 kDa for the long and short isoforms respectively.

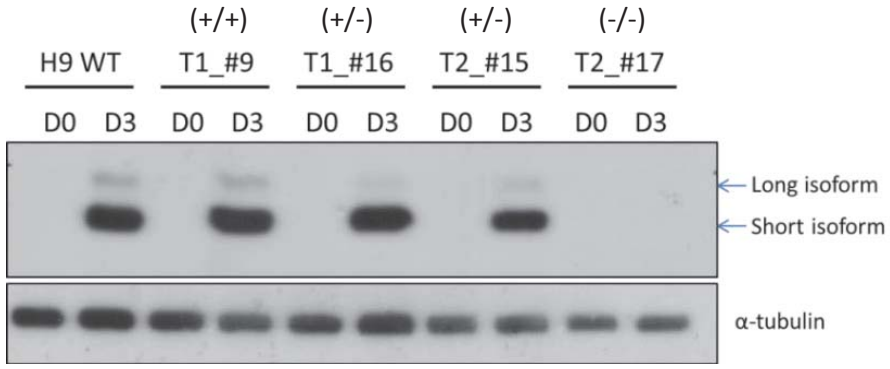


Figure 44. Western blot analysis of GATA6 protein levels in TALEN-derived H9 mutant lines using a C-terminal GATA6 antibody. Cells were harvested on day 0 (undifferentiated) and day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively. No GATA6 protein was present for the T2_#17 mutant.

Gene editing was next performed in FSPS13.B cells. I focused on using the TALEN1 cut site with the intention of eliminating the PPP and obtaining complete loss of function *GATA6* alleles. 78 clones were screened with an observed cutting efficiency of approximately 10% on one allele. I obtained an identical mutant for FSPS13.B, referred to as T1_#50, that contained the same 2 base pair deletion as T2_#16 in H9 cells (Figure 45). Similar to H9 cells, no homozygous mutant for the TALEN1 cut site was recovered. Thus, I subjected the T1_#50 mutant line to re-targeting at the TALEN1 site. 48 clones were screened with an observed cutting efficiency of approximately 2% of the remaining wild-type allele. A homozygous mutant, referred to as T1_#50_#42, containing the same 2 base pair deletion on both alleles was obtained (Figure 45).

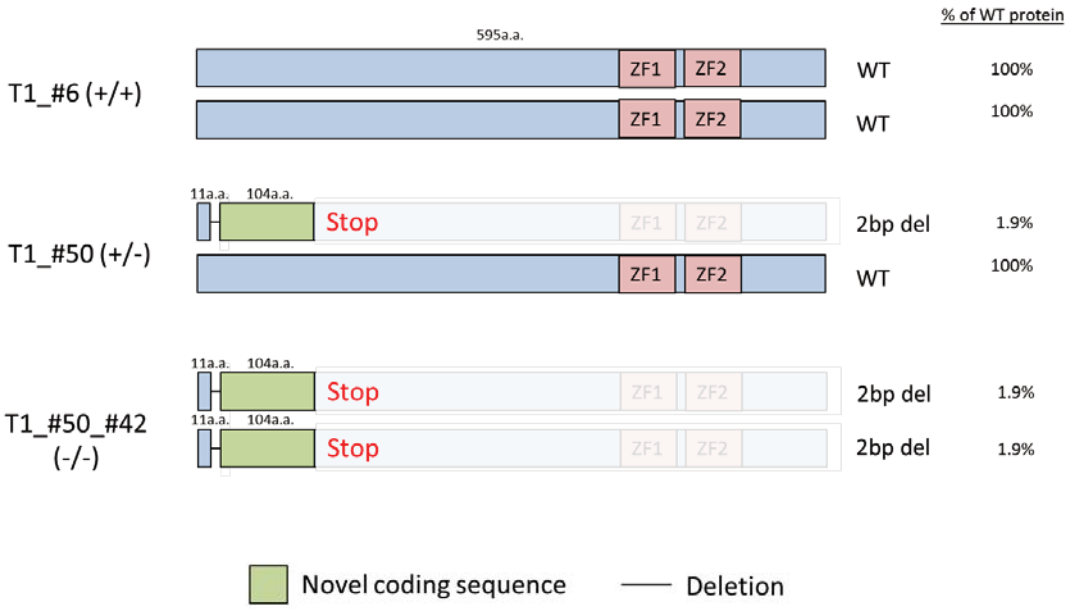


Figure 45. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN1 cut site. Three TALEN-derived cell lines were selected for use in downstream applications. T1_#6 is a TALEN-targeted but wild-type line; T1_#50 contains a 2 base pair deletion frameshift mutation; T1_#50_#42 contains an identical 2 base pair deletion frameshift mutation on both *GATA6* alleles. All frameshift mutations were out of frame resulting in a premature stop codon.

Surprisingly, upon performing western blot to verify the complete loss of GATA6 protein in the T1_#50_#42 mutant, the short protein isoform of GATA6 was observed using both N- and C-terminus GATA6 antibodies (Figure 46). The simplest explanation for this finding is ribosomal read through and translational initiation at the second *GATA6* ATG (Figure 47).

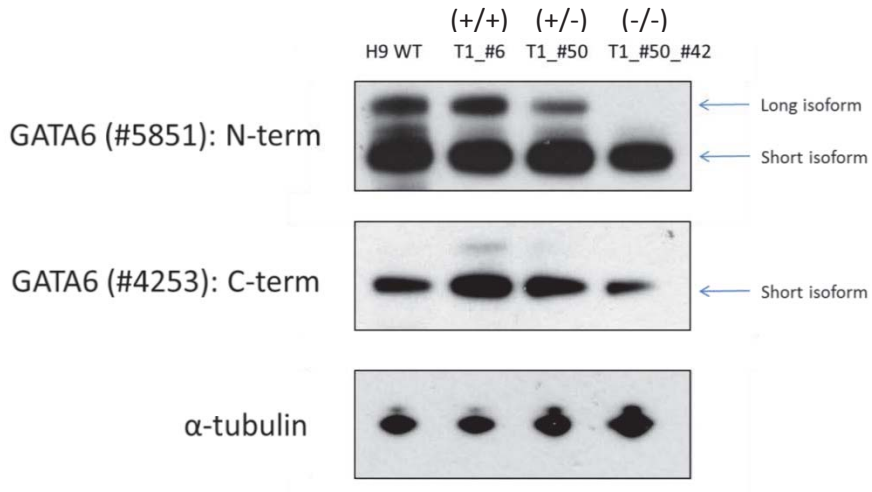


Figure 46. Western blot analysis of GATA6 protein expression in TALEN-derived FSPS13.B mutant lines using N- and C-terminal antibodies. Cells were harvested on day 3 (DE). Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively. The short isoform of GATA6 protein was present for the T1_#50_#42 mutant.

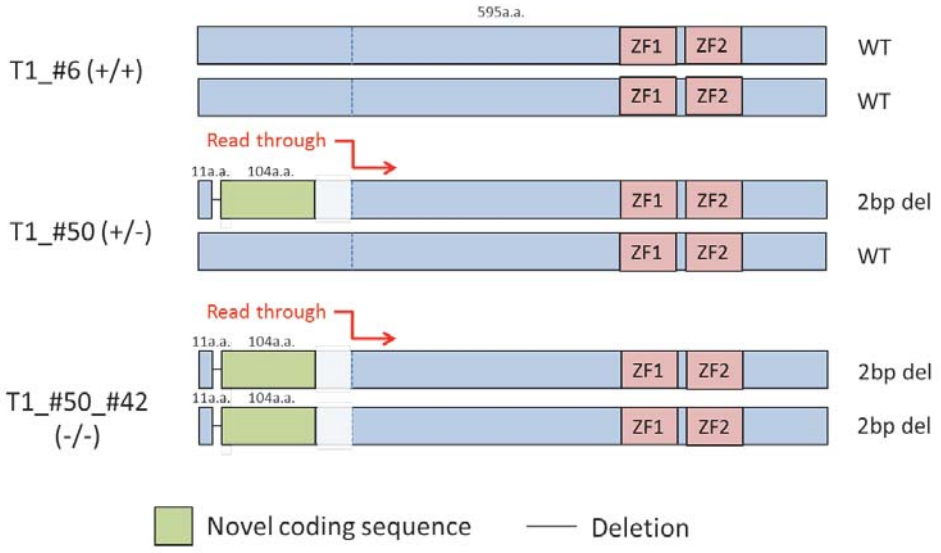


Figure 47. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN1 cut site with read-through. The ribosomal read through and translational initiation at the second *GATA6* ATG led to the translation of the short GATA6 isoform in the mutants. Blue dotted lines represent the second start codon.

Since it is now clear that the TALEN1 cut site is not suitable for generating *GATA6* loss of function mutant lines, I proceeded to target the FSPS13.B cell line at the TALEN2 cut site. 33 clones were screened with an observed cutting efficiency of 30% of one allele. I selected three colonies which after sequencing were found to harbour out of frame heterozygous frameshift mutations resulting in premature stop codons (Figure 48). Again, no homozygous mutants were obtained and so I proceeded to retarget mutant T2_#8 which harbours a 14 base pair out of frame frameshift mutation. 24 clones were screened with an observed cutting efficiency of 30% of the remaining wild-type allele. A homozygous mutant, referred to as T2_#8_#4, containing an 11 base pair out of frame frameshift mutation in the other *GATA6* allele was selected for use in downstream experiments (Figure 48).

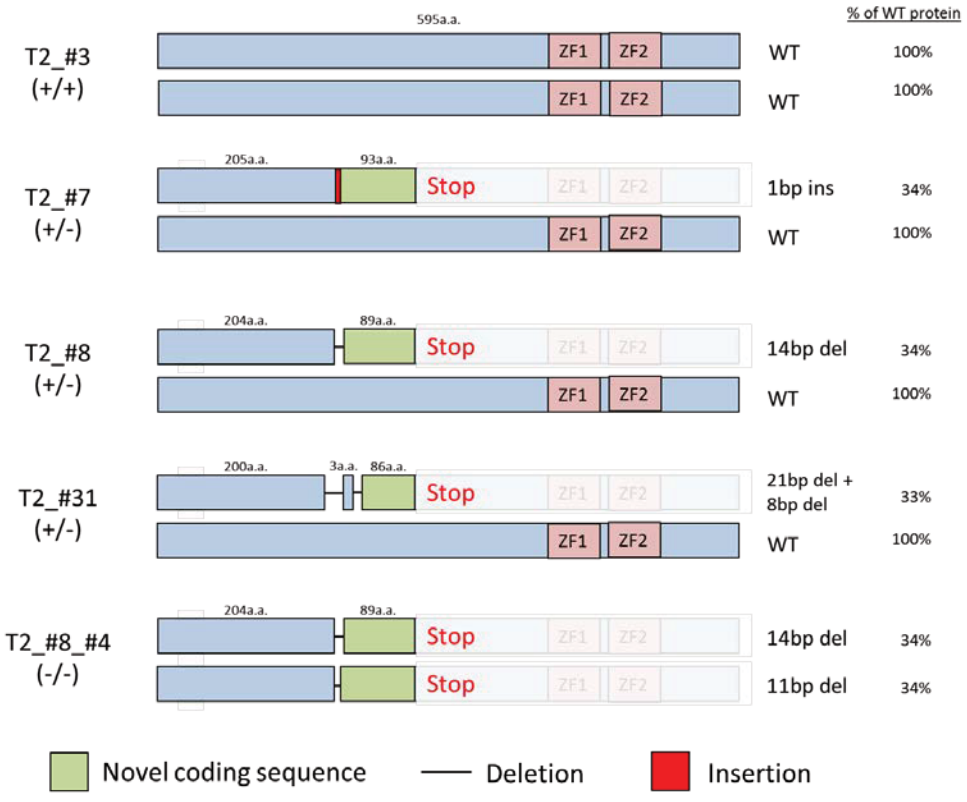


Figure 48. Schematic of selected FSPS13.B TALEN-derived *GATA6* cell lines at the TALEN2 cut site. Four TALEN-derived cell lines were selected for use in downstream applications. T2_#3 is a TALEN-targeted but wild-type line; T2_#7 contains a 1 base pair insertion frameshift mutation; T2_#8 contains a 14 base pair deletion frameshift mutation; T2_#31 contains a 21 and 8 base pair deletion frameshift mutation; T2_#8_#4 contains a 14 base pair deletion frameshift mutation in one *GATA6* allele and 11 in the other. All frameshift mutations were out of frame resulting in a premature stop codon.

GATA6 heterozygous and homozygous mutant H9 and FSPS13.B cell lines derived from TALEN targeting at the TALEN2 site that were selected for subsequent experiments are summarised in Table 51 with precise details of their genotypes. The nomenclature for the mutants that is used throughout the remainder of this thesis is listed in Table 51. All mutant lines were verified by western blotting and sequencing. They were also karyotyped and found to be normal and absent of any chromosomal abnormality.

Table 51. Summary of selected H9 and FSPS13.B mutants generated via NHEJ pathway

Cell line	TALEN cut site and colony number	Genotype	Nomenclature
H9	T1_#9	*GATA6* $^{+/+}$ (wild-type)	H9* (WT)
H9	T2_#15	*GATA6* $^{c.618\_619insTGCA/+}$	*GATA6* $^{4ins/+}$ (Het)
H9	T2_#17	*GATA6* $^{c.611\_614delACCT/c.611\_614delACCT}$	*GATA6* $^{\Delta4/\Delta4}$ (Hom)
FSPS13.B	T2_#3	*GATA6* $^{+/+}$ (wild-type)	FSPS13.B* (WT)
FSPS13.B	T2_#7	*GATA6* $^{c.615\_616insC/+}$	*GATA6* $^{1ins/+}$ (Het)
FSPS13.B	T2_#8	*GATA6* $^{c.del614\_627TGCAGGGGTCGGGC/+}$	*GATA6* $^{\Delta14/+}$ (Het)
FSPS13.B	T2_#31	*GATA6* $^{c.600\_621delinsTGGGCCAG/+}$	*GATA6* $^{\Delta21\_8ins/+}$ (Het)
FSPS13.B	T2_#8_#4	*GATA6* $^{c.del614\_627TGCAGGGGTCGGGC/}$ $^{c.del613\_623CTGCAGGGGTC}$	*GATA6* $^{\Delta14/\Delta11}$ (Hom)

3.2.2. HR pathway

Next, I generated *GATA6* knockout hPSC lines that are entirely devoid of a PPP. To accomplish this, I performed TALEN targeting at the TALEN1 site and introduced a donor/targeting vector (TG) that contained an emerald GFP (emGFP) reporter in frame with the endogenous *GATA6* ATG and a puromycin-resistance cassette (Figure 49). TALEN cutting and successful HR via the 5' and 3' homology arms results in either a heterozygous or homozygous loss-of-function *GATA6* mutation (Figure 49).

I targeted both H9 and FSPS13.B cells. Initial PCR screens on genomic DNA extracted from picked colonies using forward and reverse primers within exon 1 and intron 2 respectively as indicated by the red arrows in Figure 49 revealed the "knock-in" of the donor template from the targeting vector (Figure 50). For H9 cells, 12 clones were screened with an observed HR efficiency of approximately 20% on one allele and 10% on two alleles. For FSPS13.B cells, 24 clones were screened with an observed HR efficiency of 50% of one allele. Unfortunately, no homozygous mutants were obtained even after I attempted to re-target a heterozygous mutant and screened through 24 clones.

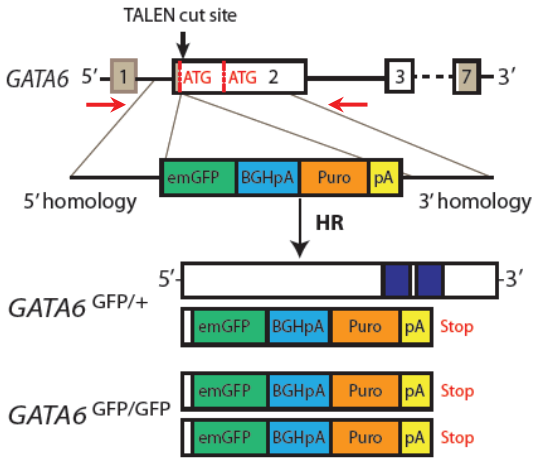


Figure 49. Schematic of generating heterozygous or homozygous loss-of-function *GATA6* mutations via HR. A "knock-in" vector that introduces an emerald GFP (emGFP) reporter in-frame with the first *GATA6* ATG and a puromycin resistance cassette into *GATA6* exon 2 is depicted. Successful homologous recombination resulted in both heterozygous (*GATA6* $^{GFP/+}$) and homozygous (*GATA6* $^{GFP/GFP}$) mutant cells. Red arrows indicate the positions of the primers used for PCR screening.

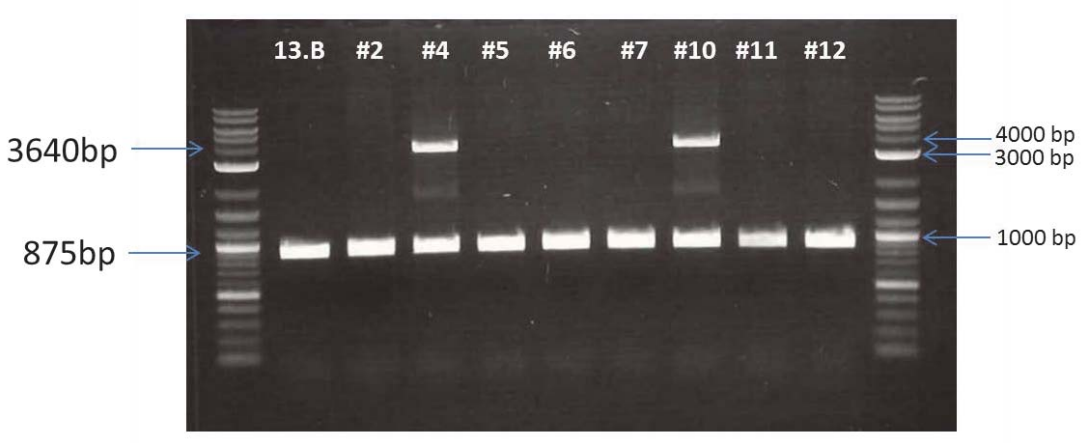


Figure 50. Representative DNA agarose gel picture of colonies screened via PCR to assess for successful HR. Genomic DNA was extracted from picked colonies and subjected to PCR using forward and reverse primers within exon 1 and intron 2 of the *GATA6* gene respectively. Successful HR of the donor template resulted in a 2765bp insertion as seen in colonies #4 and #10.

For H9 cells, I selected two heterozygous mutants and one homozygous mutant for use in downstream applications (Figure 51). For FSPS13.B cells, I selected two heterozygous mutants for use in downstream applications. To control for any potential off-target effects, I selected one colony that had been treated similarly to the mutants but harboured no observable HR or insertion of the donor template.

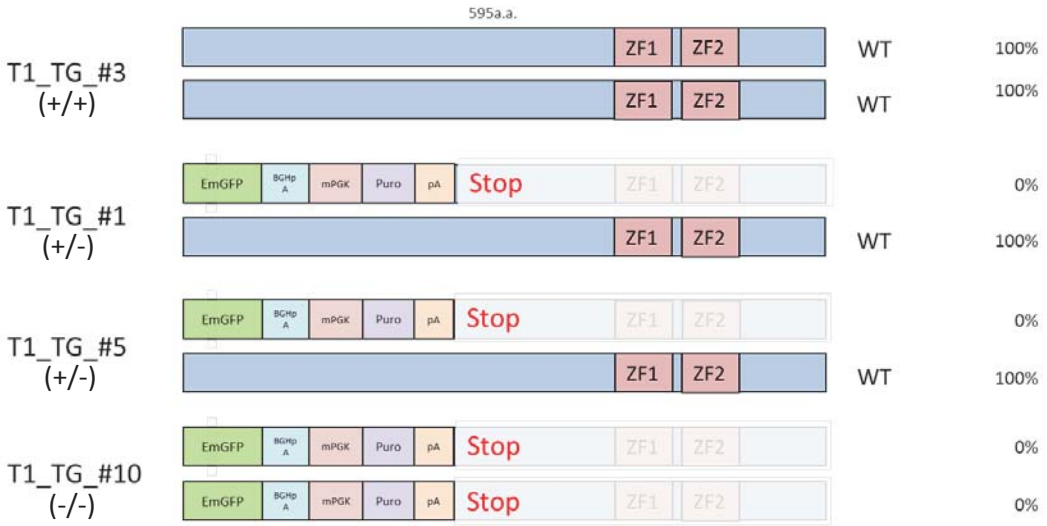


Figure 51. Schematic of selected H9 TALEN-derived *GATA6* cell lines via HR. Four TALEN-derived cell lines were selected for use in downstream applications. T1_TG_#3 is a TALEN-targeted but wild-type line; T1_TG_#1 and T1_TG_#5 are heterozygous *GATA6* knockouts; T1_TG_#10 is a homozygous *GATA6* knockout.

Western blotting using both N- and C-terminal antibodies confirmed the complete absence of GATA6 protein in *GATA6* $^{GFP/GFP}$ (Figure 52). All mutant lines were sequenced to ensure correct insertion of the donor template. This was done by sequencing the 3640bp PCR product containing the inserted fragment from the donor vector as shown in colonies #4 and #10 (Figure 50). They were also karyotyped and found to be normal and absent of any chromosomal abnormality. The mutants were further validated by the expression of emGFP (Figure 53).

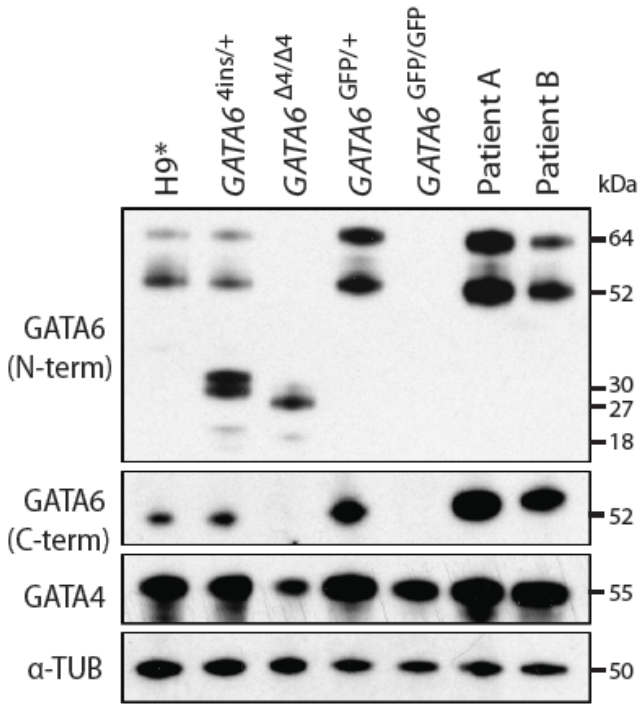


Figure 52. Western blot analysis of GATA6 and GATA4 protein levels in GATA6 mutant lines. H9*, *GATA6* $^{4ins/+}$, *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{GFP/+}$, *GATA6* $^{GFP/GFP}$ mutant cells as well as Patients A and B were differentiated to day 3 (DE) and harvested. Alpha-tubulin was used as a loading control. Long and short isoforms of wild-type GATA6 are 64 kDa and 52 kDa respectively; the partial protein products for GATA6$^{4ins/+}$ are 30 kDa and 18 kDa for the long and short isoforms respectively; the partial protein products for GATA6$^{\Delta4/\Delta4}$ are 27 kDa and 15 kDa for the long and short isoforms respectively. No GATA6 protein was present for the *GATA6* $^{GFP/GFP}$ mutant.

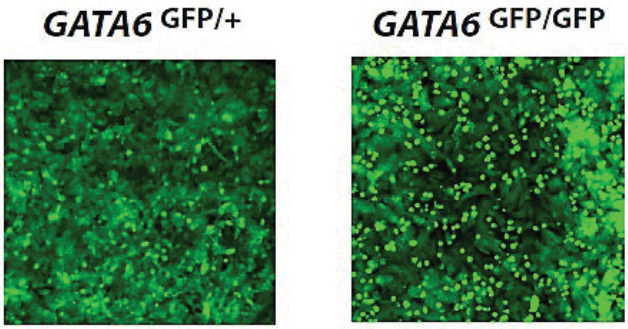


Figure 53. Immunofluorescence showing emGFP-expressing heterozygous *GATA6* $^{GFP/+}$ and homozygous *GATA6* $^{GFP/GFP}$ mutant cells on day 3. Cells were differentiated toward the DE and checked for the expression of emGFP.

GATA6 heterozygous and homozygous mutant H9 and FSPS13.B cell lines derived from TALEN targeting at TALEN1 site following HR repair pathway that were selected for use in downstream experiments are summarised in Table 52. The nomenclature for the mutants that is used for the subsequent sections of this report is also listed in Table 52.

Table 52. Summary of selected H9 and FSPS13.B mutants generated via HR pathway

Cell line	TALEN cut site and colony number	Nomenclature
H9	T1_TG_#3	H9* (WT)
H9	T1_TG_#1	*GATA6* $^{GFP/+}$ (Clone 1) (Het)
H9	T1_TG_#5	*GATA6* $^{GFP/+}$ (Clone 2) (Het)
H9	T1_TG_#10	*GATA6* $^{GFP/GFP}$ (Hom)
FSPS13.B	T1_TG_#2	FSPS13.B* (WT)
FSPS13.B	T1_TG_#4	*GATA6* $^{GFP/+}$ (Clone 1) (Het)
FSPS13.B	T1_TG_#10	*GATA6* $^{GFP/+}$ (Clone 2) (Het)

3.2.3. Reprogramming of GATA6 patient fibroblasts

In addition to the TALEN-derived GATA6 mutants, I also obtained fibroblasts from two *GATA6* patients in collaboration with Professor Andrew Hattersley.

Genotyping via Sanger sequencing revealed that Patient A (*GATA6* $^{R456C/+}$) has a missense mutation within the second zinc finger DNA-binding domain of the GATA6 protein, while Patient B (*GATA6* $^{c.1136\text{-}2A>G/+}$) contains a splice acceptor mutation in exon 3 (Figure 54). After reprogramming of the fibroblasts, three independent hiPSC clones were selected for each patient line, and these clones were also monitored for absence of the Sendai virus (Figure 55). All patient-derived cell lines were karyotypically normal and pluripotent (Figure 56).

Western blot analyses revealed the presence of both short and long isoforms of the GATA6 protein in both Patient A and B cell lines, suggesting the absence of a nonsense mediated decay (Figure 52).

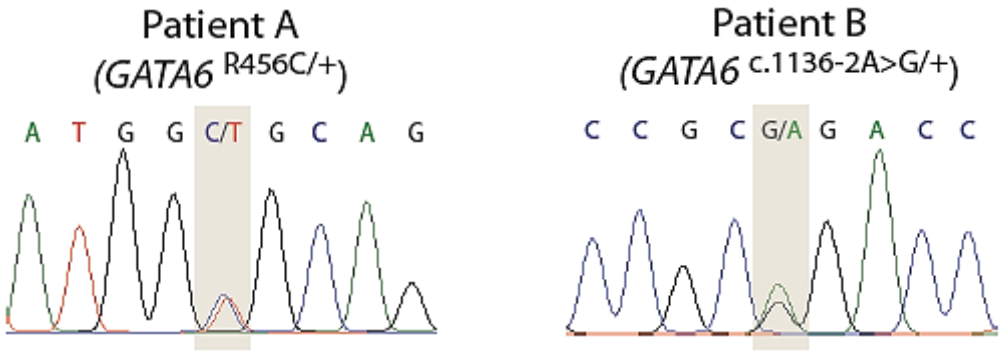


Figure 54. Genotype confirmation of Patients A and B by Sanger sequencing. Two *GATA6* patient-derived hiPSC lines—Patient A, *GATA6*$^{R465C/+}$ and Patient B, *GATA6*$^{c.1136\text{-}2A>G/+}$ were sequenced to confirm their heterozygous mutations.

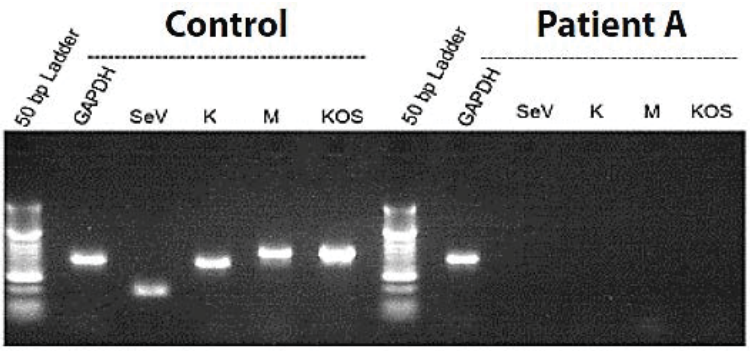


Figure 55. PCR showing loss of transgenes in Patient A mutant line, clone 1 compared with positive control. Picture is representative of 3 clones derived from each Patient A and Patient B mutant lines.

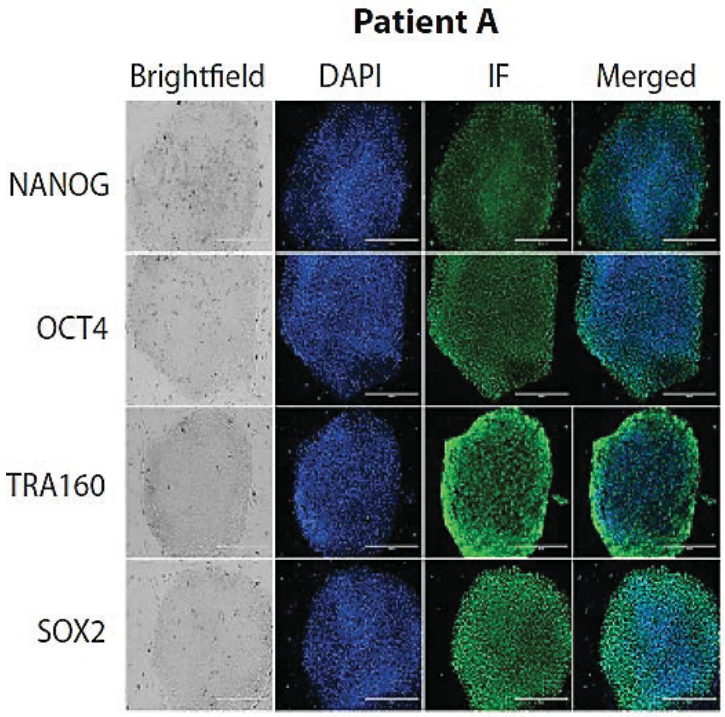


Figure 56. Immunofluorescence showing successful reprogramming of patient-derived Patient A (*GATA6* $^{R465C/+}$) mutant line via expression of pluripotency markers. Scale bars, 400 µm. Picture is representative of 3 clones derived from each Patient A and Patient B (*GATA6* $^{c.1136\text{-}2A>G/+}$) mutant lines.

3.2.4. Genome editing does not affect pluripotency

Next, I wanted to determine whether genome editing of the *GATA6* gene would affect pluripotency in the TALEN-derived *GATA6* mutants. ICC analyses on undifferentiated H9 *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells showed high expression levels of key pluripotency markers NANOG, OCT4 and SOX2 and absence of key DE marker SOX17 as well as GATA6, indicating that the cells are indeed in a pluripotent state (Figure 57 and Figure 58).

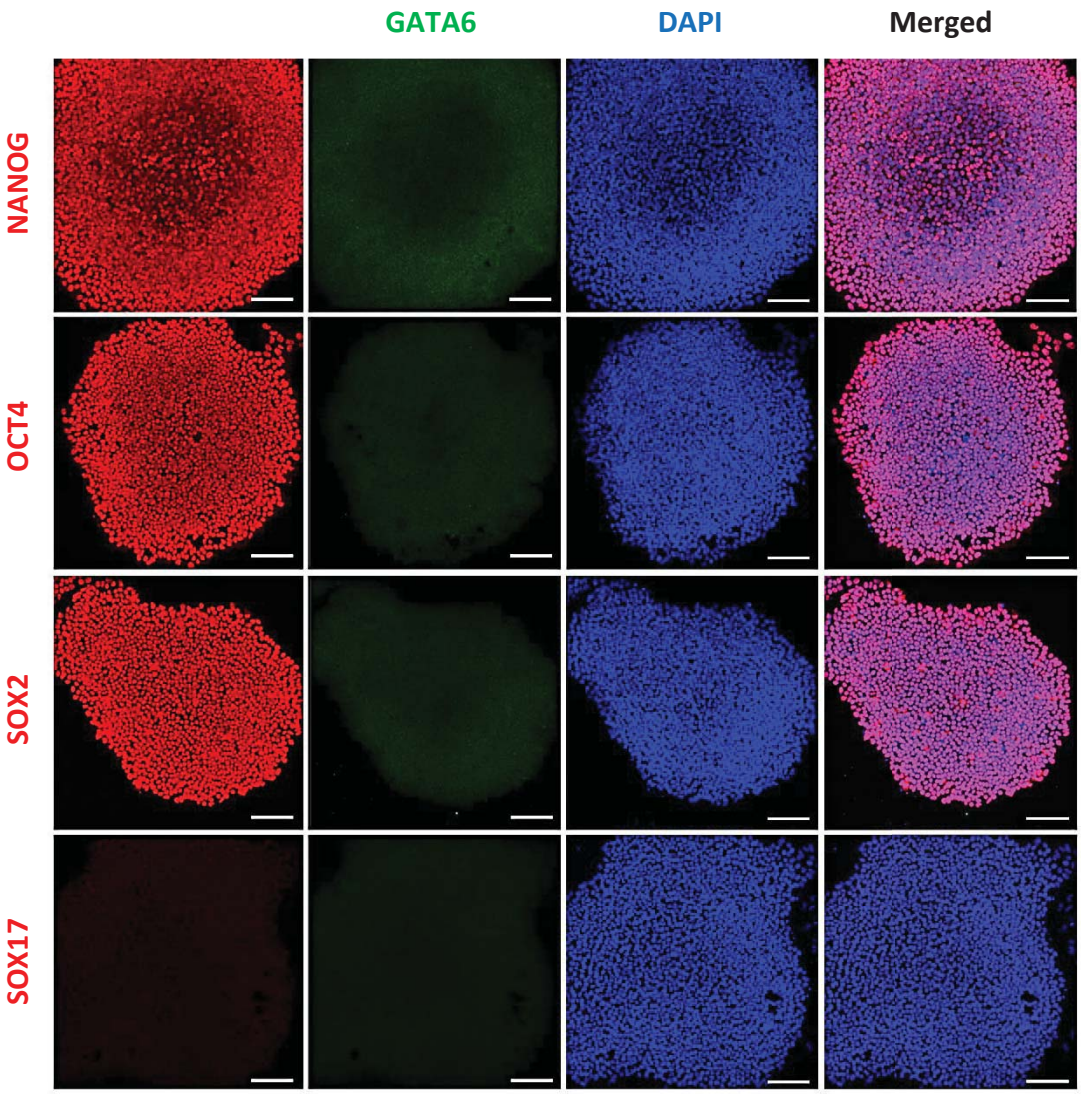


Figure 57. Pluripotency is maintained in *GATA6* $^{4ins/+}$ H9 cells. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for the pluripotency markers *NANOG*, *OCT4* and *SOX2* and early DE marker *SOX17*. Scale bar, 100 µm.

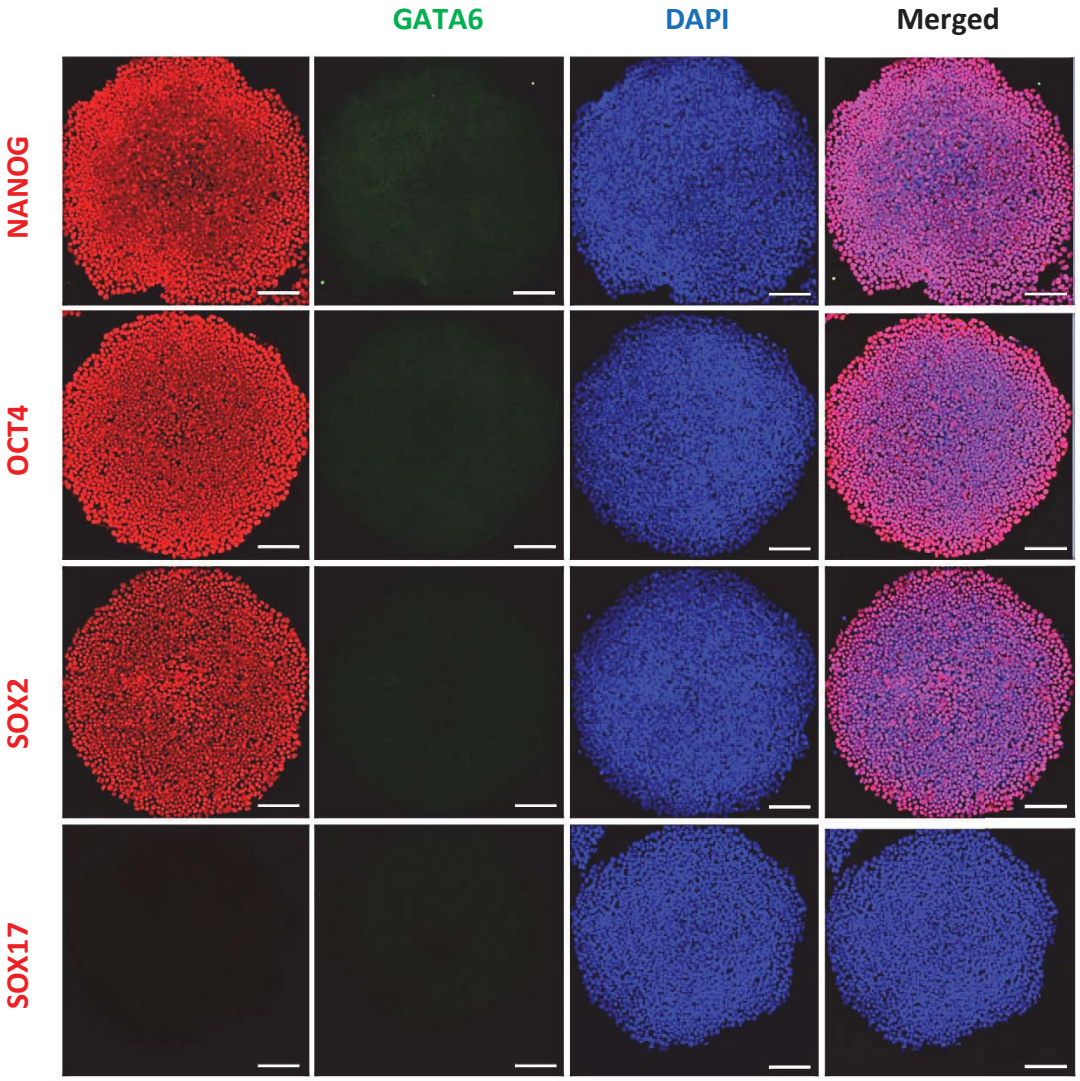


Figure 58. Pluripotency is maintained in *GATA6* $^{\Delta4/\Delta4}$ H9 cells. Cells were grown in culture conditions that maintained pluripotency and analysed via immunofluorescence. Cells were fixed on day 0 and were stained for the pluripotency markers *NANOG*, *OCT4* and *SOX2* and early DE marker *SOX17*. Scale bar, 100 µm.

Subsequent qRT-PCR analyses of the H9, FSPS13.B and patient *GATA6* mutant lines at D0 focusing on pluripotency genes such as *OCT4* and *SOX2* indicated that the mutant cell lines were similarly pluripotent , as described further in the next sections. Together, these results suggest genome editing on the *GATA6* gene does not affect pluripotency.

3.2.5. TALEN-derived wild-type cell lines resemble untargeted hPSCs

To assay for any potential off target effects, I differentiated TALEN-derived wild-type cells lines designated H9* and FSPS13.B* alongside their respective unmanipulated wild-type counterparts (H9 and FSPS13.B). qRT-PCR and FACS results show similar levels of key markers between the respective cell lines, indicating that DE and pancreatic differentiation were unaffected by the TALEN targeting, and further suggesting that there were no observable off target effects from the TALEN targeting (Figure 59, Figure 60 and Figure 61).

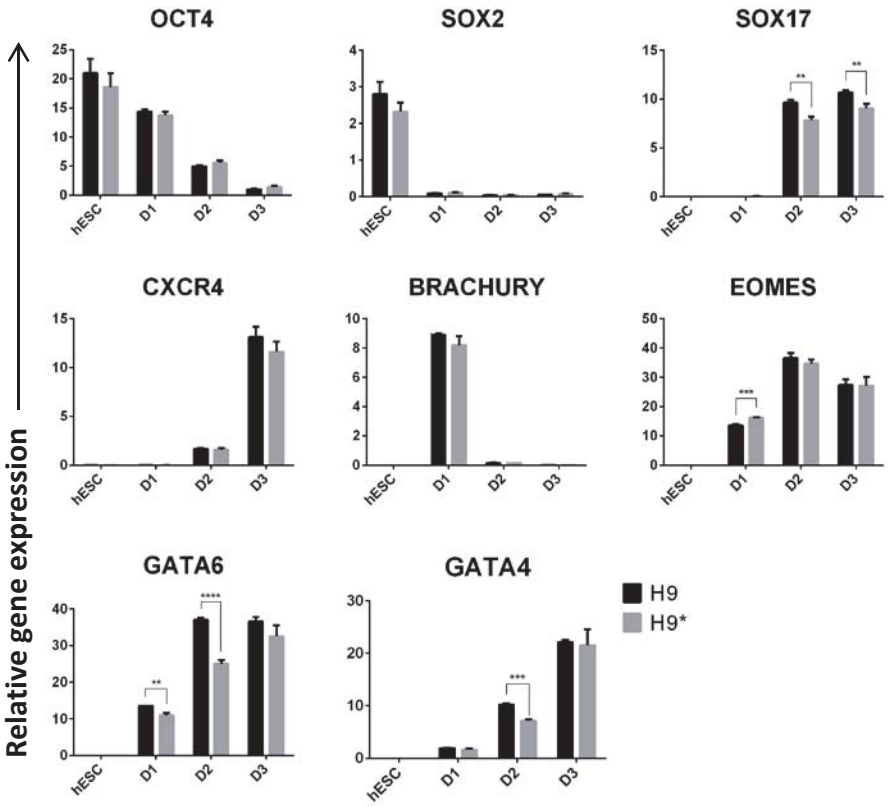


Figure 59. qRT-PCR analyses of H9 and H9* cells on days 1, 2 and 3. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

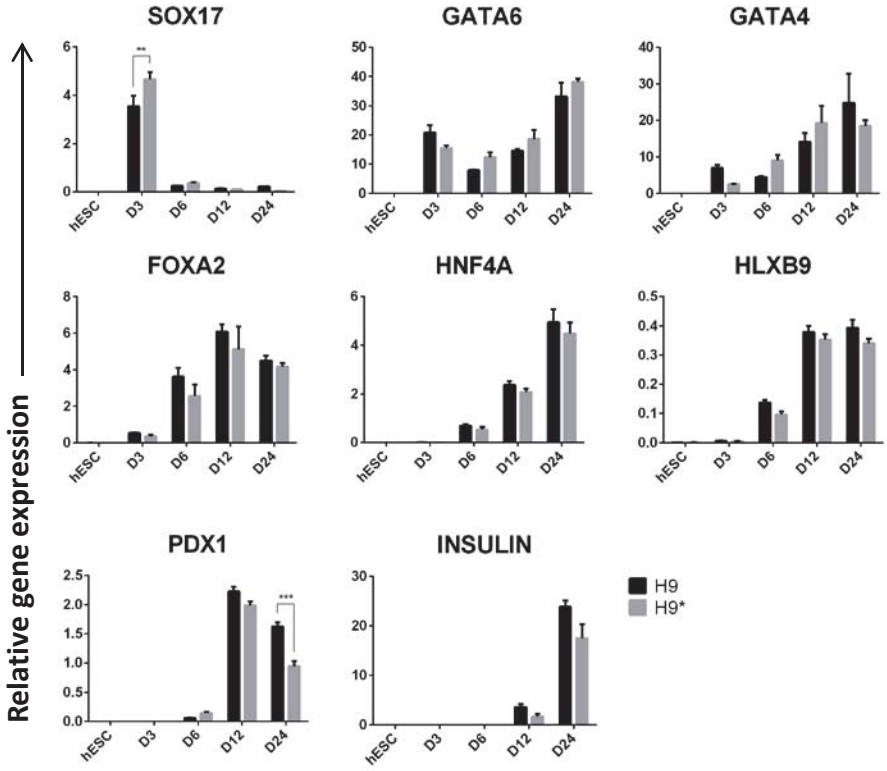


Figure 60. qRT-PCR analyses of H9 and H9* cells on days 3, 6, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

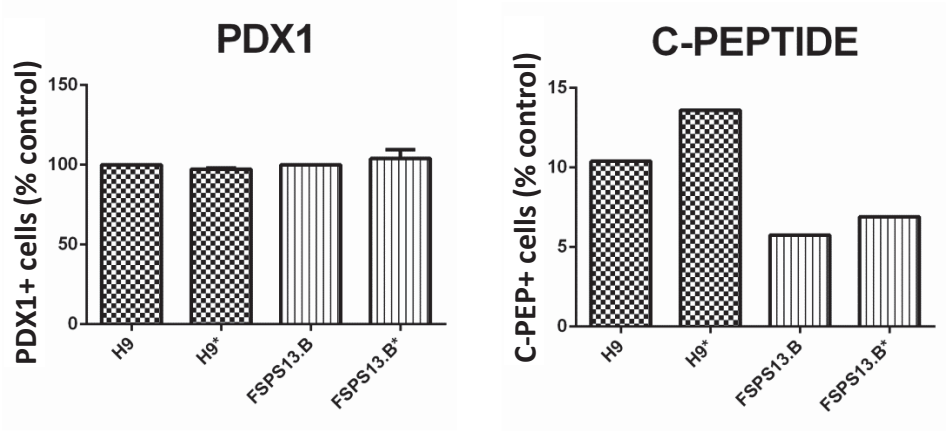


Figure 61. Summary of PDX1+ (day 12) and C-PEPTIDE+ (day 24) cells via FACS for H9, H9*, FSPS13.B and FSPS13.B. For PDX1, H9 and FSPS13.B cells are normalised to 100% and the relative PDX1+ cells to their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation. For C-PEPTIDE, absolute percentages of C-PEPTIDE+ cells in all cell lines are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of three independent experiments.

3.3. Endodermal formation is inconsistently impaired by heterozygous *GATA6* mutations

Next, to study the effects of heterozygous or homozygous loss of *GATA6* during human pancreatic development, I subjected the mutant lines to pancreatic differentiation in order to characterise their phenotypes and perform phenotypical comparisons between the various genotypes with an initial focus at the DE stage.

3.3.1. Biallelic loss of N-terminal of GATA6 protein impairs DE formation

To determine if loss of the N-terminal GATA6 protein upstream of the second start codon affects the formation of the DE, I differentiated the FSPS13.B lines derived from TALEN1 targeting toward the DE and performed FACS of key DE marker CXCR4 at day 3 (refer to Figure 47). Results from FACS suggested that in FSPS13.B cells, the loss of GATA6 protein between the first and second start codon on one allele does not affect DE differentiation (84% of CXCR4+ cells for T1_#50), but the biallelic loss of GATA6 protein between the first and second start codon impairs DE specification by approximately 30% (57% of CXCR4+ cells for T1_#50_#42) (Figure 62). This is in concordance to an early study where it was reported that the N-terminal 146 amino acids of GATA-6 contains transactivational activity, and its deletion reduced luciferase expression by 50% (Brewer et al., 1999).

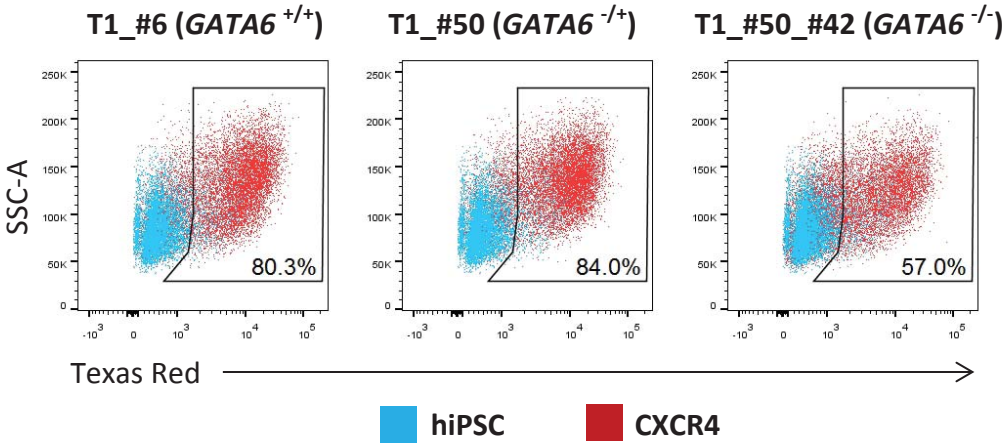


Figure 62. FACS of CXCR4+ cells for FSPS13.B TALEN1-targeted mutant cells on day 3. Cells were grown in culture conditions that specified them toward the DE lineage and they were analysed via FACS. Cells were fixed on day 3 and were stained for CXCR4. Gates are set according to undifferentiated FSPS13.B cells. Data show results of one experiment that is representative of at least 3 independent experiments.

3.3.2. Homozygous *GATA6* mutants fail to form the DE

It was reported that mouse embryos lacking both copies of *Gata6* die during gastrulation (Morrisey et al., 1998, Koutsourakis et al., 1999). This early embryonic lethality is believed to be a consequence of extraembryonic endoderm dysfunction and arrest development at E6.5-7, as this deficiency could be overcome by providing *Gata6*-null embryos with a wild-type extraembryonic endoderm with the use of tetraploid embryo complementation (Zhao et al., 2005).

To determine if this early endoderm dysfunction is recapitulated in humans using our *in vitro* model system, I differentiated TALEN-derived homozygous mutants in both H9 (*GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$) and FSPS13.B (*GATA6* $^{\Delta14/\Delta11}$) cell lines. At day 3, the loss of pluripotency marker NANOG in all cell lines (Figure 63) and the high expression of key DE marker SOX17 in H9* cells (Figure 64) indicated the successful differentiation of cells toward the DE lineage. H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ homozygous mutants, however, displayed negligible expression of SOX17 as shown by ICC (Figure 64). FACS analyses showed a 90-100% decrease of SOX17+ and CXCR4+ cells (Figure 65 and Figure 66). FACS analyses of FSPS13.B-derived *GATA6* $^{GFP/GFP}$ homozygous mutant also showed a similar result (Figure 67). The loss of DE markers in these homozygous mutants was confirmed by qRT-PCR analyses (Figure 68, Figure 69 and Figure 70). That CXCR4+ cells were similarly abolished in H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutants suggest that the PPP in *GATA6* $^{\Delta4/\Delta4}$ mutant cells does not have a functional role during differentiation (Figure 71). Expectedly, the homozygous mutants failed to develop into the subsequent key stages toward pancreatic development such as the dorsal foregut (Figure 72, Figure 73 and Figure 74).

Taken together, these results are in concordance with previously published data in mice where *Gata6* $^{-/-}$ embryos displayed early embryonic lethality, which is believed to be a consequence of extraembryonic endoderm dysfunction. My results show that absence of *GATA6* in *GATA6* $^{-/-}$ hPSC mutants generated from TALEN genome editing abolished the formation of the DE.

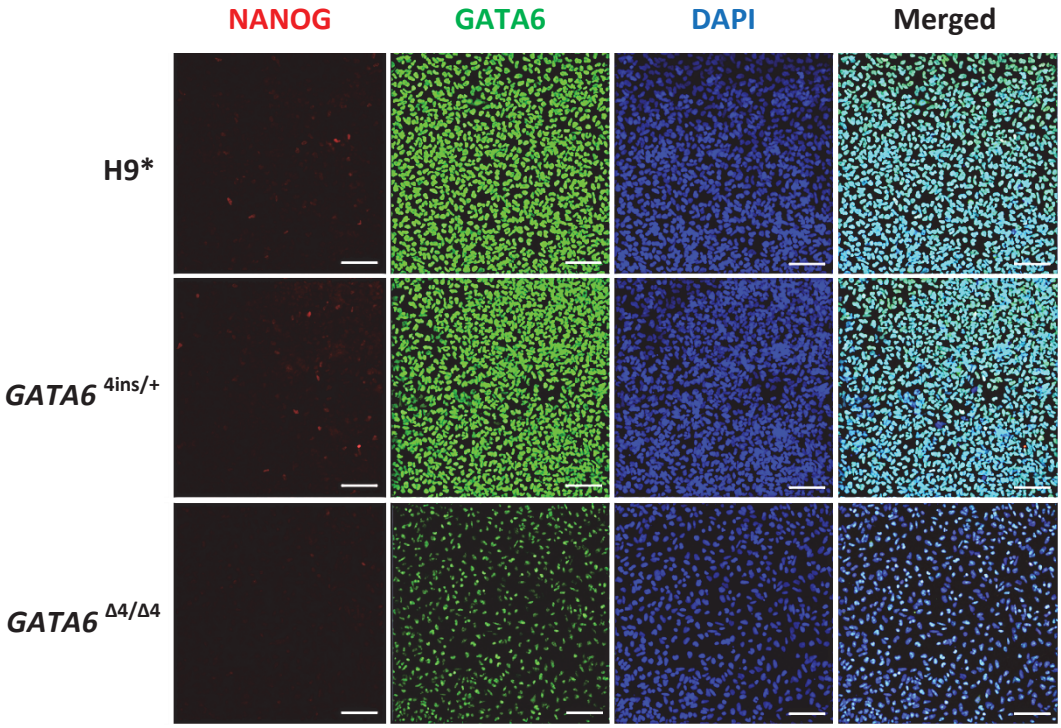


Figure 63. *GATA6* $^{4ins/+}$ and *GATA6* $^{Δ4/Δ4}$ mutants had the capacity to differentiate. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the pluripotency marker NANOG. Scale bar, 100 µm.

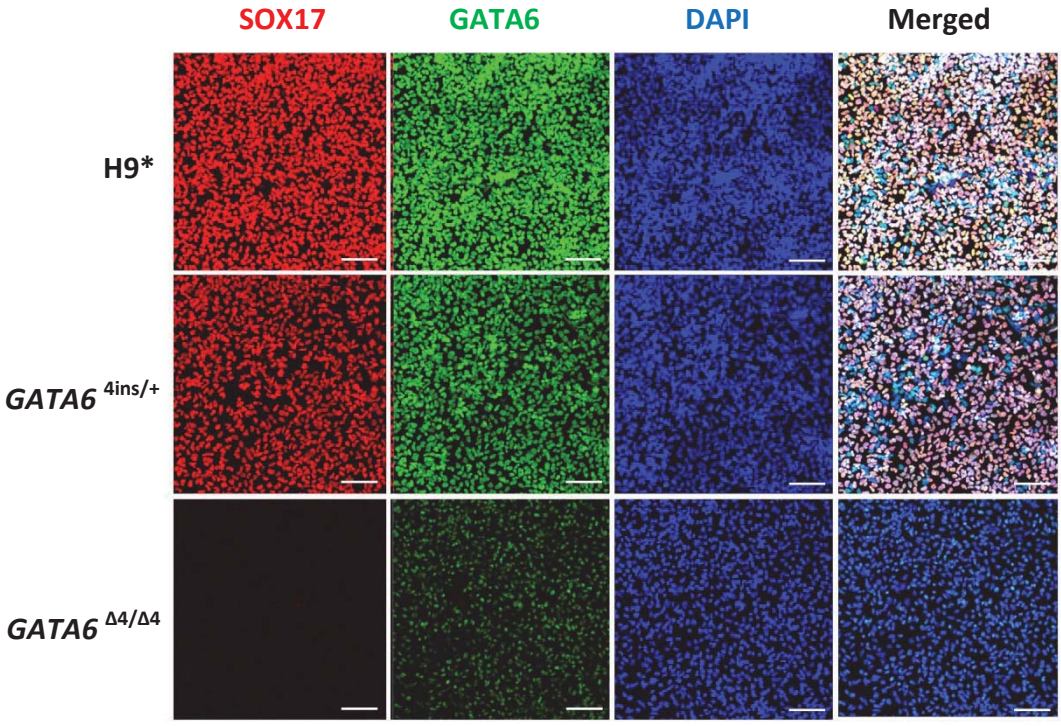


Figure 64. SOX17 expression is abolished *GATA6* $^{Δ4/Δ4}$ cells. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the DE marker SOX17. Scale bar, 100 µm.

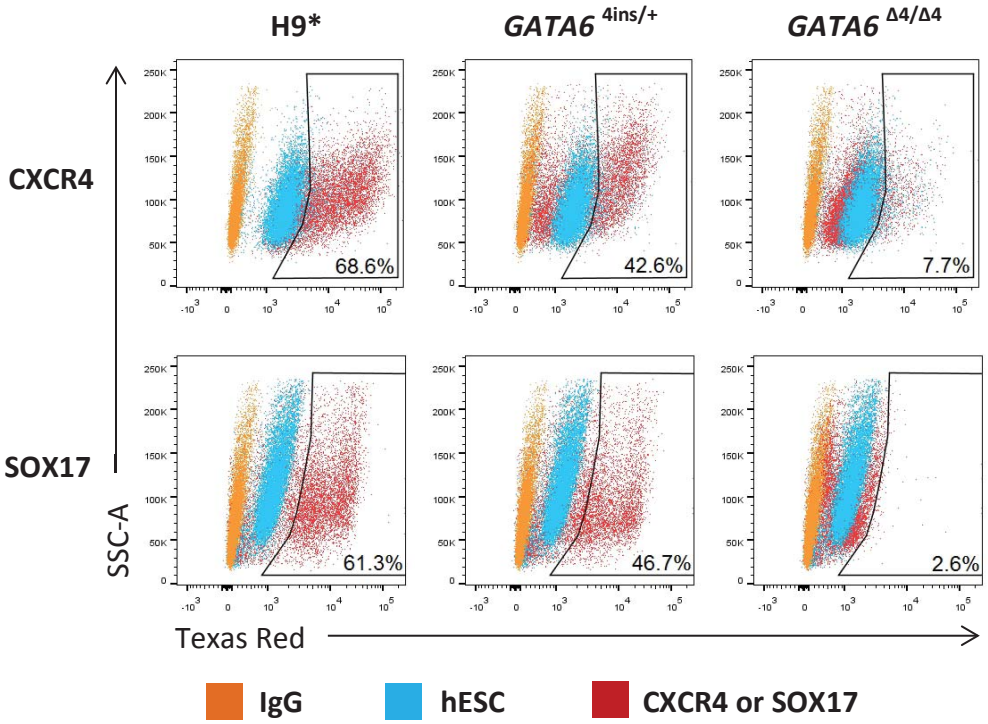


Figure 65. The number of CXCR4+ and SOX17+ cells are decreased in *GATA6* $^{4ins/+}$ cells and are almost completely absent in *GATA6* $^{\Delta4/\Delta4}$ H9 cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

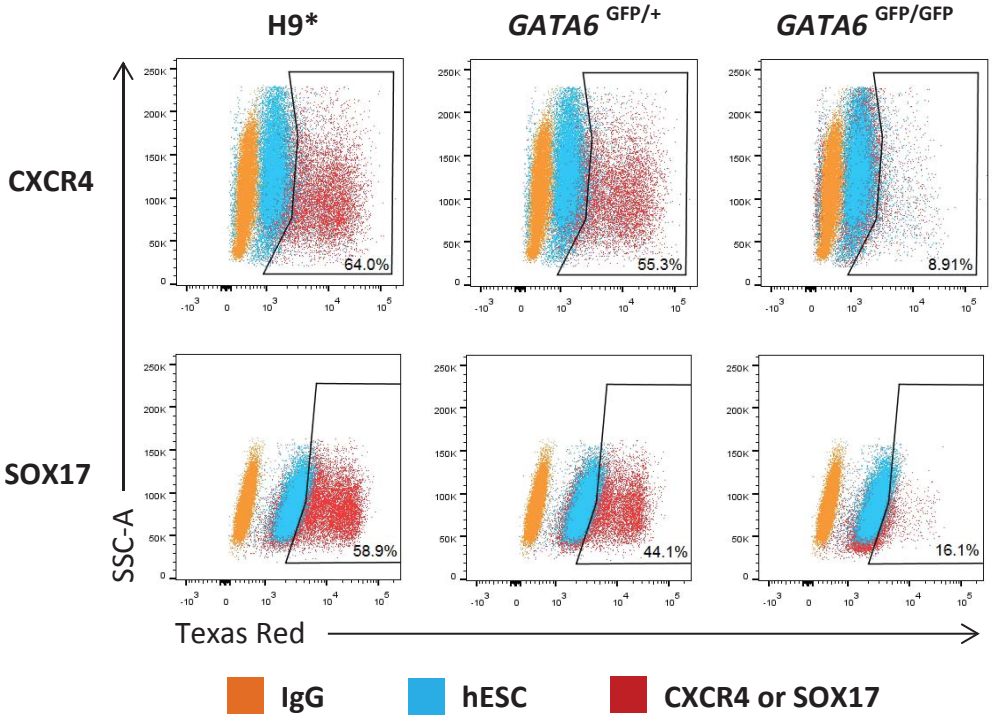


Figure 66. The number of CXCR4+ and SOX17+ cells are decreased in *GATA6* $^{GFP/+}$ cells and are almost completely absent in *GATA6* $^{GFP/GFP}$ H9 cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

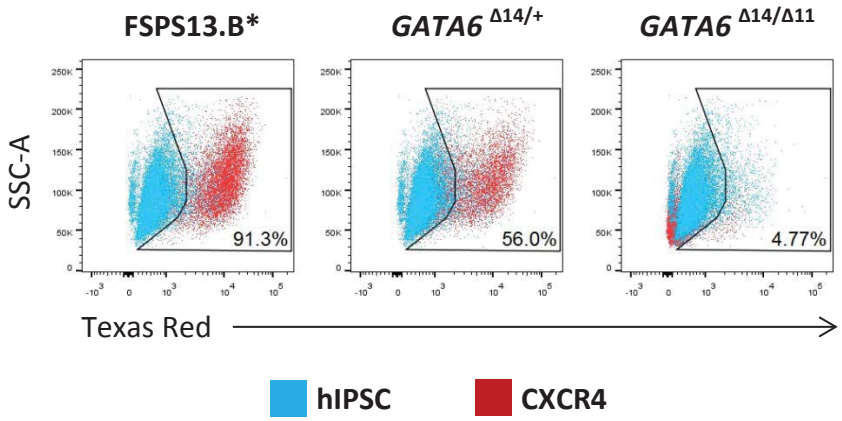


Figure 67. The number of CXCR4+ is decreased in *GATA6* $^{\Delta 14/+}$ cells and are almost completely absent in *GATA6* $^{\Delta 14/\Delta 11}$ FSPS13.B cells. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

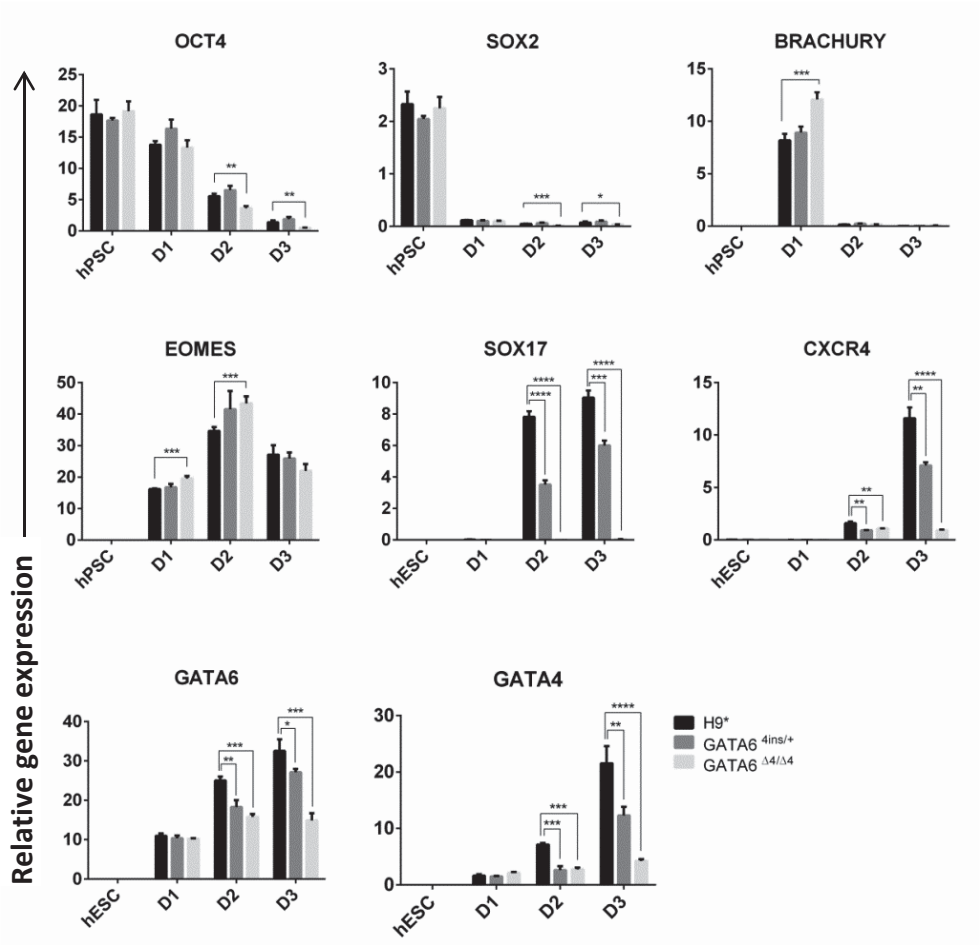


Figure 68. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta 4/\Delta 4}$ cells on days 1, 2 and 3. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

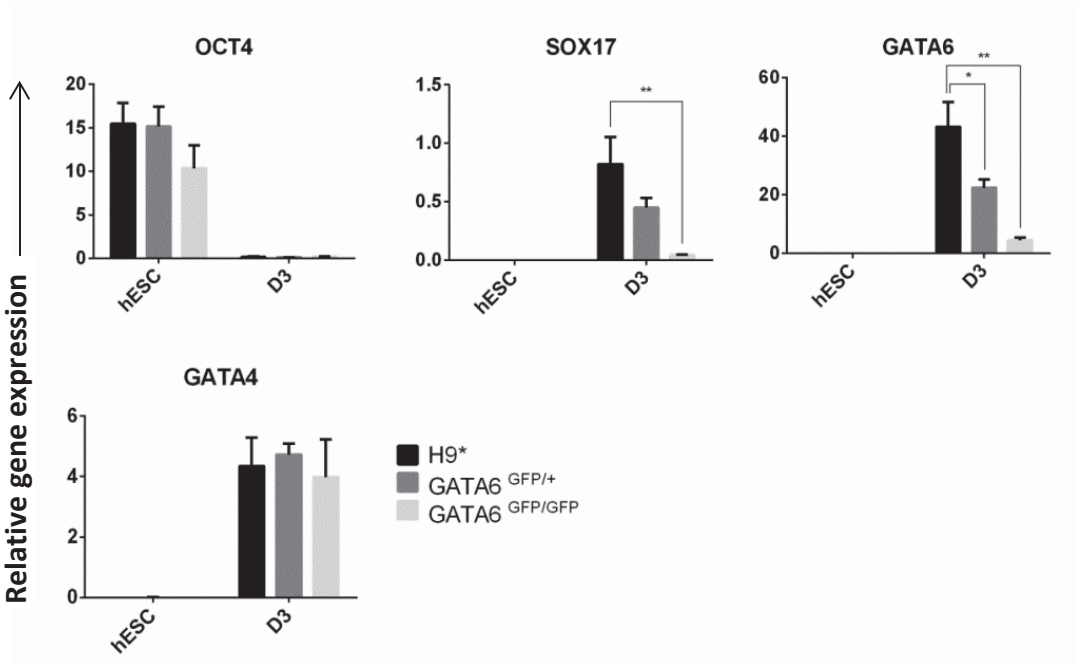


Figure 69. qRT-PCR analyses of H9*, *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

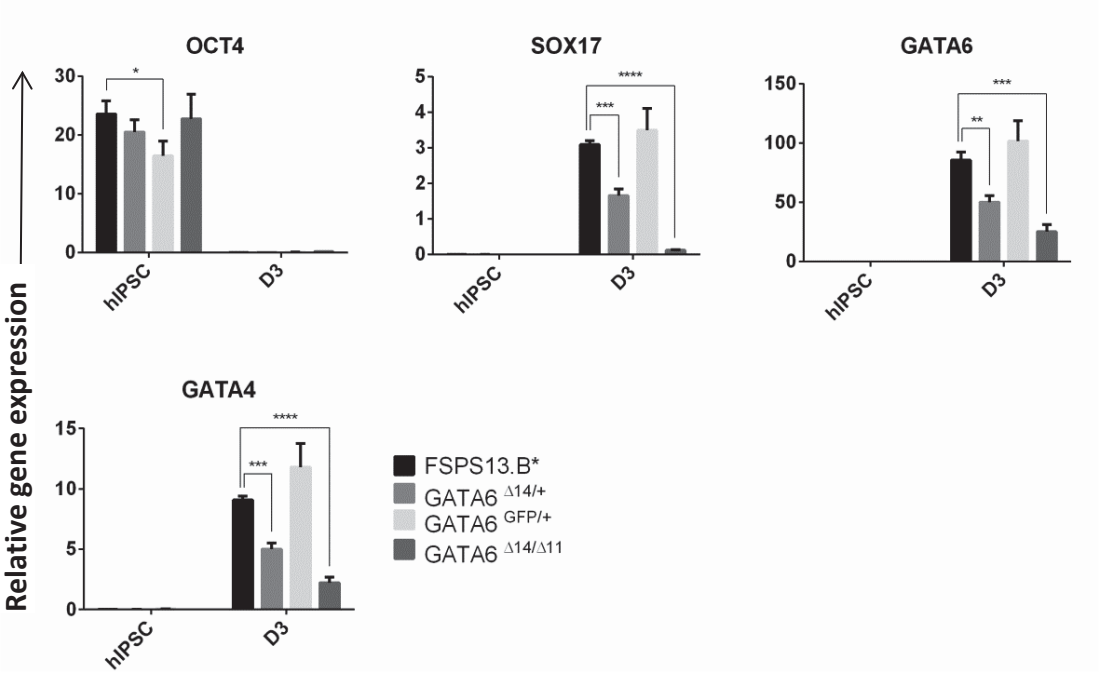


Figure 70. qRT-PCR analyses of FSPS13.B*, *GATA6* $^{\Delta 14/+}$, *GATA6* $^{GFP/+}$ and *GATA6* $^{\Delta 14/\Delta 14}$ cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

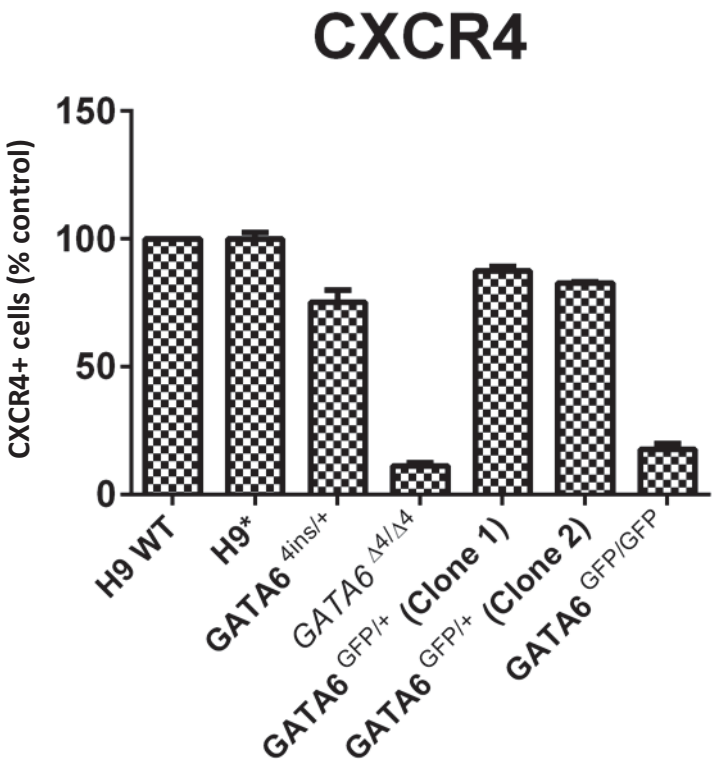


Figure 71. Summary of CXCR4+ cells via FACS for all H9-derived mutant cells on day 3. Wild-type H9 cells are normalised to 100% and the relative CXCR4+ cells of H9* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

3.3.3. *GATA6* heterozygous mutants exhibit endodermal defects using lab-derived protocol

After determining the phenotypic consequence of homozygous *GATA6* mutations, I here shift my focus to heterozygous *GATA6* hPSC lines.

In H9-derived *GATA6* $^{4ins/+}$ heterozygous mutant cells, SOX17 is expressed on day 3, indicating that DE formation is not abolished (Figure 64). FACS analyses on day 3 for key DE markers CXCR4 and SOX17 for *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant cells shows a moderate decrease in the number of CXCR4+ and SOX17+ cells of approximately 20% (Figure 65 and Figure 66). The relative number of CXCR4+ cells from FACS analyses comparing the *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ clones 1 and 2 mutants to H9* cells (normalised to 100%) indicate that the population of CXCR4+ cells for heterozygous mutants are consistently approximately 80% (Figure 71). The phenotypic similarity between *GATA6* $^{4ins/+}$ and both clones 1 and 2 of *GATA6* $^{GFP/+}$ also suggests that the one copy of the PPP present in the *GATA6* $^{4ins/+}$ mutant cells is non-functional during differentiation. ICC analyses showing expression of FOXA2 on days 3 and 6 (Figure 72), the absence of CDX2 (Figure 75) and HEX (Figure 76) on day 6 and expression of HNF1B on day 6 and *SOX2* on day 9 (Figure 77) indicate that *GATA6* $^{4ins/+}$ mutant cells are capable of differentiating into the dorsal foregut and foregut lineages.

Using the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines, I next asked whether *GATA6* haploinsufficiency impacts the very early differentiation from mesendoderm (corresponding to days 1 and 2) to DE (day 3). qRT-PCR analyses revealed the expression of primitive streak (*BRACHYURY*) and mesendoderm (*EOMESODERMIN* (*EOMES)*) markers were relatively unchanged across the control H9* and mutant lines, suggesting that early mesendoderm formation was not affected by either single or biallelic loss of *GATA6* (Figure 68). In contrast, consistent with FACS analyses, key DE markers *SOX17* and *CXCR4* were downregulated beginning on day 2 by approximately 30% in *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ cells (Figure 68 and Figure 69).

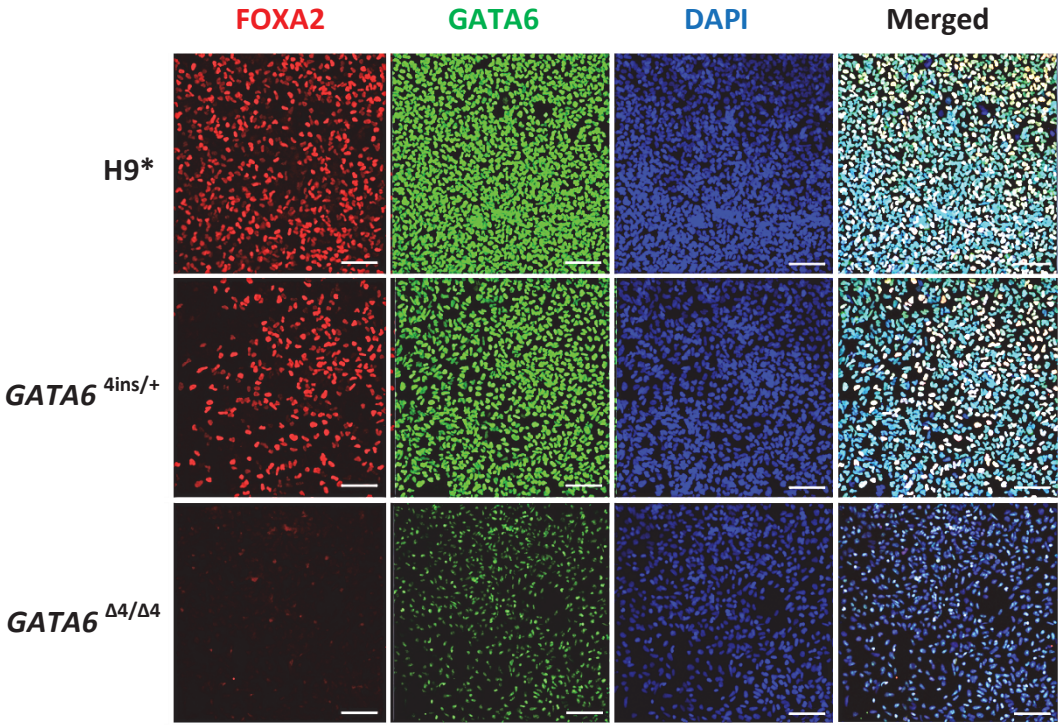


Figure 72. FOXA2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells and decreased in *GATA6* $^{4ins/+}$ cells on day 3. Cells were grown in culture conditions that specified them toward the DE lineage and analysed via immunofluorescence. Cells were fixed on day 3 and were stained for the foregut marker FOXA2. Scale bar, 100 µm.

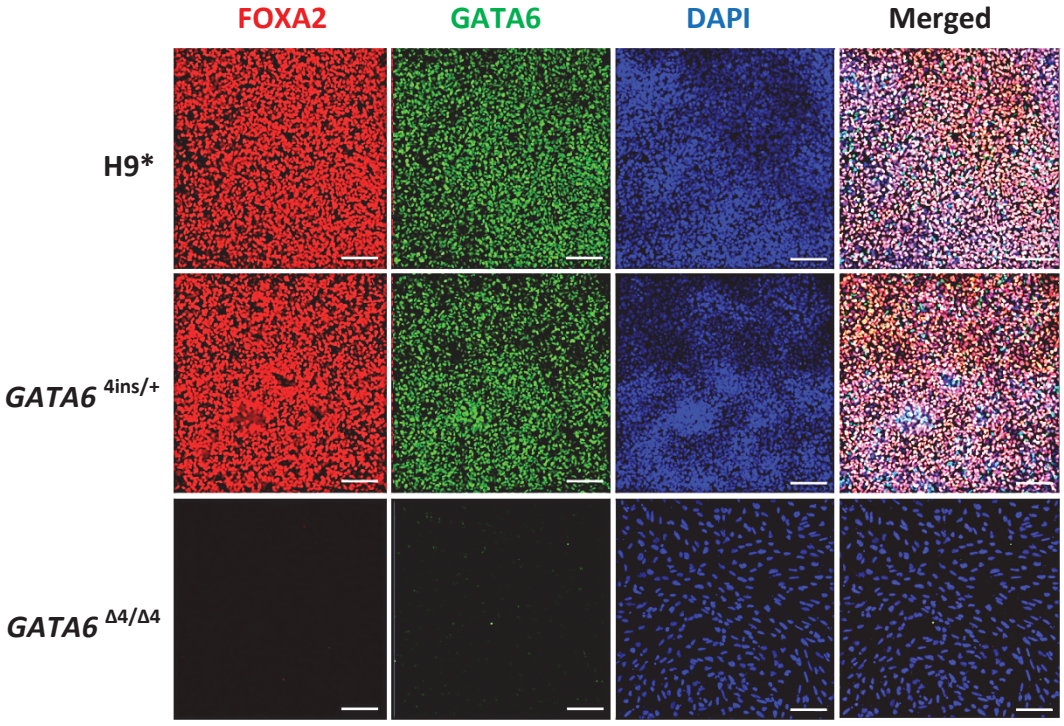


Figure 73. FOXA2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the foregut marker FOXA2. Scale bar, 100 µm.

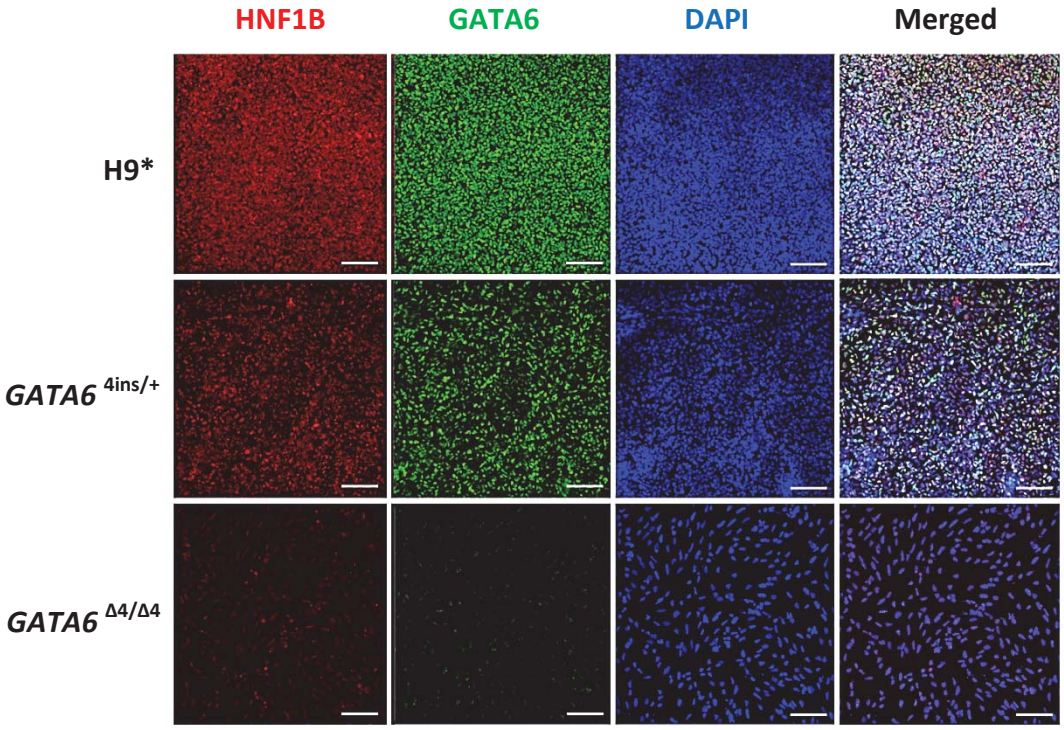


Figure 74. HNF1B expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the foregut marker HNF1B. Scale bar, 100 µm.

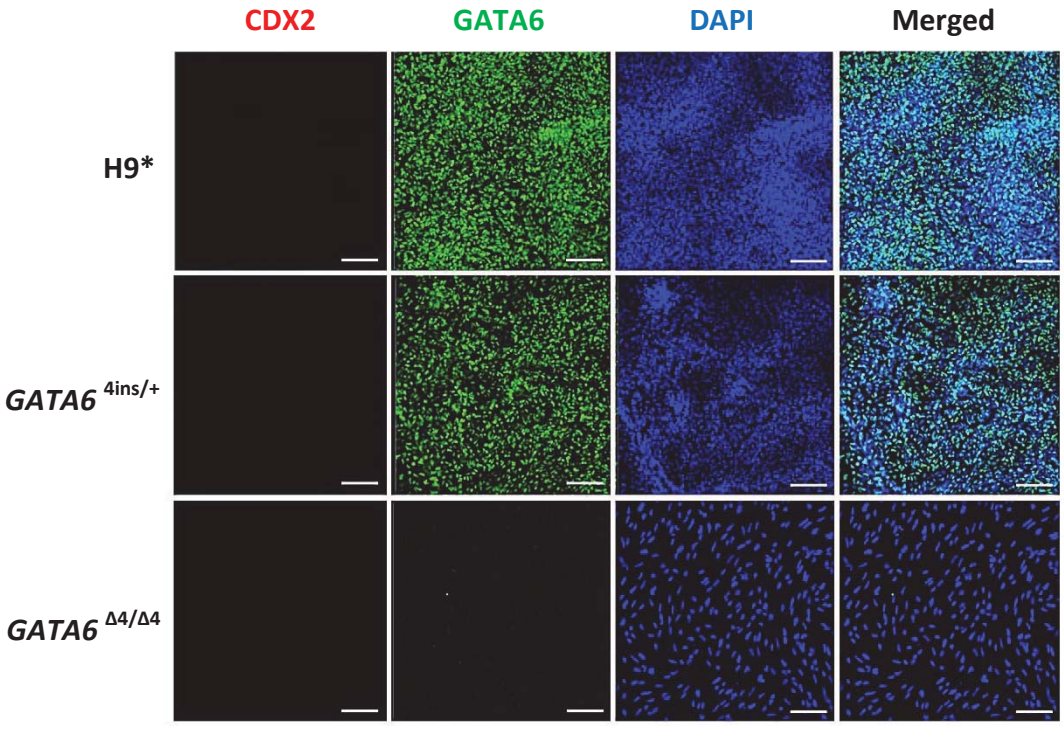


Figure 75. CDX2 remains unexpressed in all cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the hindgut marker CDX2. Scale bar, 100 µm.

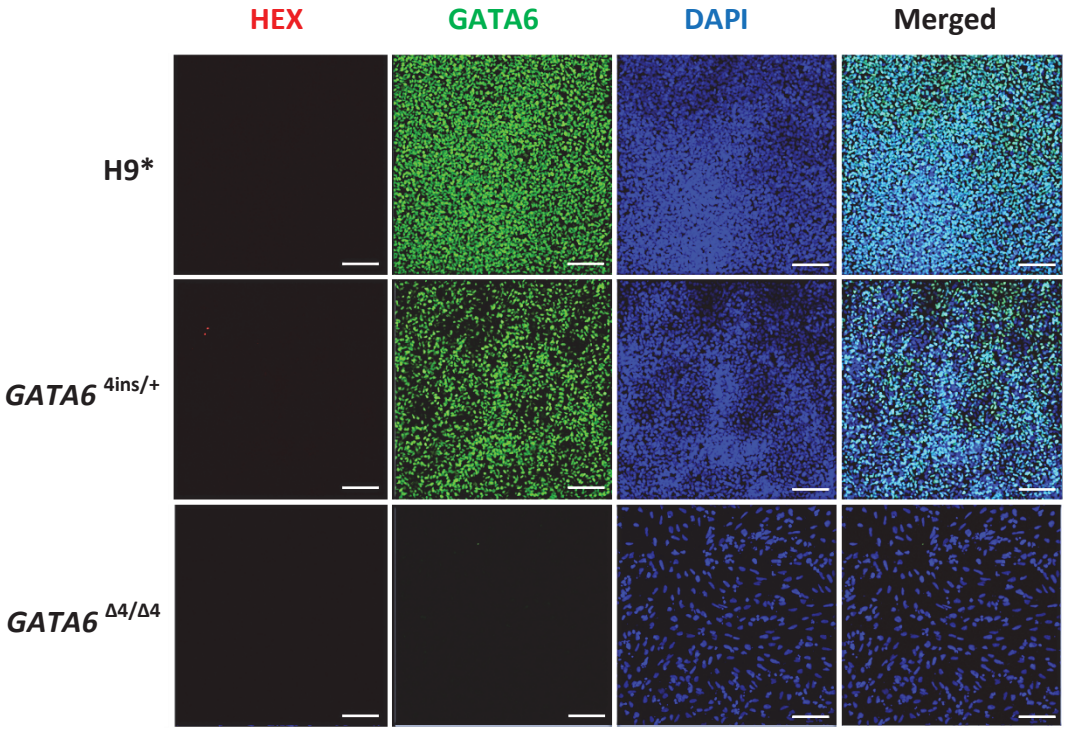


Figure 76. HEX remains unexpressed in all cells on day 6. Cells were grown in culture conditions that specified them toward the dorsal foregut lineage and analysed via immunofluorescence. Cells were fixed on day 6 and were stained for the ventral foregut marker HEX. Scale bar, 100 µm.

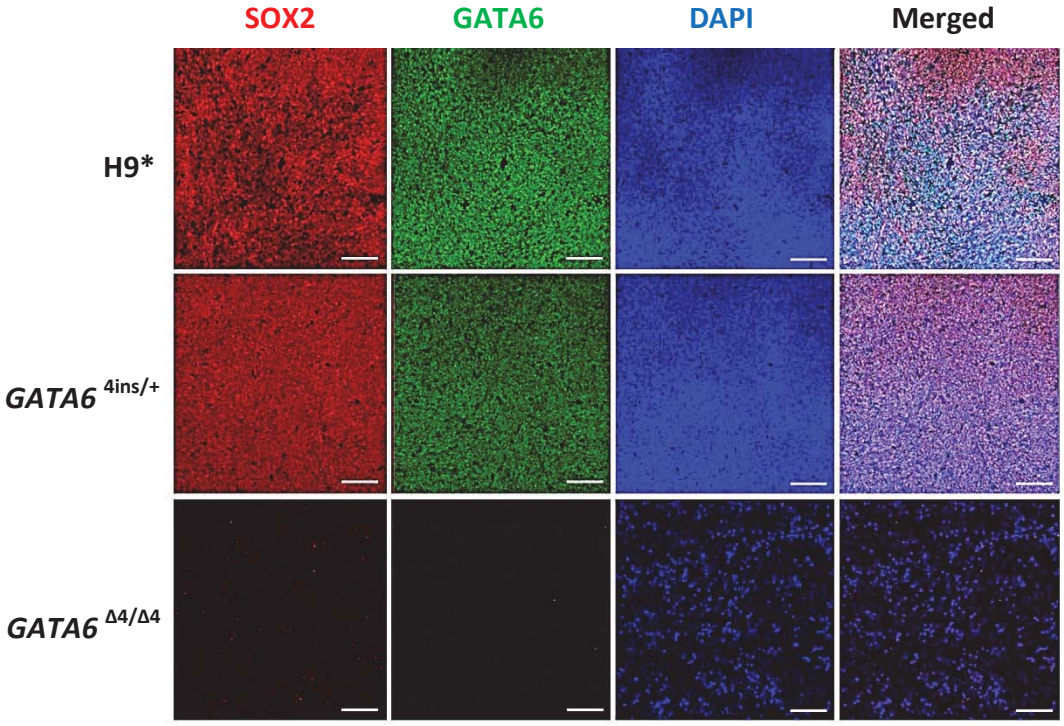


Figure 77. SOX2 expression is abolished in *GATA6* $^{\Delta4/\Delta4}$ cells on day 9. Cells were grown in culture conditions that specified them toward the foregut lineage and analysed via immunofluorescence. Cells were fixed on day 9 and were stained for the foregut marker SOX2. Scale bar, 100 µm.

As *GATA6* and *GATA4* have been shown to interact with each other (Charron et al., 1999), I looked at the mRNA levels of *GATA4* to investigate whether its regulation is affected by mutations in the *GATA6* gene. Notably, on day 3 at the transcriptional level, I observed that *GATA4* expression was decreased in H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 68), suggesting that the GATA6 PPPs generated from these mutant alleles interact with *GATA4* to negatively impact its expression, leading to its down-regulation. This was in contrast to the *GATA4* levels observed in *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells, where *GATA4* levels remained relatively unchanged at day 3 (Figure 69). Nevertheless, a similar DE phenotype was seen with H9-derived *GATA6* $^{GFP/+}$ and *GATA6* $^{GFP/GFP}$ cells based on *SOX17* levels (Figure 69), suggesting that even though the PPPs seemed to decrease *GATA4* levels, the defect on DE development is independent of this decrease.

Interestingly, in FSPS13.B-derived heterozygous mutants, differing DE phenotypes were observed as compared to H9-derived heterozygous mutants. FACS analyses of CXCR4 on day 3 in *GATA6* $^{1ins/+}$, *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta21\_8ins/+}$ mutant cells showed a 40-60% decrease in CXCR4+ cells (Figure 67 and Figure 79), indicating a stronger defect in these cells compared to the respective H9 mutant (*GATA6* $^{4ins/+}$). qRT-PCR was consistent with this observation, showing an approximately 50% decrease of *SOX17* expression levels in *GATA6* $^{\Delta14/+}$ mutant cells (Figure 70). At the transcriptional level on day 3, similar to H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells, I also observed that *GATA4* expression was decreased in *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta14/\Delta11}$ cells (Figure 70).

In contrast, FACS analyses of the key DE markers CXCR4 and SOX17 on day 3 in FSPS13.B-derived heterozygous mutants *GATA6* $^{GFP/+}$ clones 1 and 2 both displayed no defects in DE formation (Figure 78 and Figure 79). This data was recapitulated by qRT-PCR analyses where expression levels of *SOX17*, *GATA6* and *GATA4* had no significant changes (Figure 70).

Consistent with H9-derived homozygous *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells, FSPS13.B-derived homozygous *GATA6* $^{\Delta14/\Delta11}$ mutant cells failed to form the DE as shown by FACS analyses of CXCR4 at day 3 (Figure 67 and Figure 79). qRT-PCR analyses also showed almost complete ablation of *SOX17*, and strong decrease of *GATA6* and *GATA4* (Figure 70).

Patient A (clones 1-3) and Patient B (clones 1-3) displayed a similar DE phenotype to FSPS13.B-derived heterozygous *GATA6* $^{1ins/+}$, *GATA6* $^{\Delta14/+}$ and *GATA6* $^{\Delta21\_8ins/+}$ mutants cells, where FACS analyses of the key DE marker CXCR4 at day 3 showed a 50-60% decrease in CXCR4+ cells (Figure 80, Figure 81, Figure 82 and Figure 83). qRT-PCR analyses also showed a strong down-regulation of DE markers *SOX17* and *CXCR4* in both patient lines (Figure 84). *GATA4* levels, and to a lesser extent *GATA6*, in both patient lines were also down-regulated (Figure 84).

Taken together, these finding suggest that complete loss of *GATA6* in TALEN-edited hPSCs significantly perturbs the gene regulatory network (GRN) required for DE specification, resulting in the failure of DE formation. Heterozygous loss of *GATA6* seems to impair DE formation with a varying penetrance of phenotype from 20% in H9-derived heterozygous mutants, to 40-60% in FSPS13.B-derived heterozygous mutants and Patients A and B, and to no impairment in FSPS13.B-derived *GATA6* $^{GFP/+}$ mutants (summarised in Figure 85).

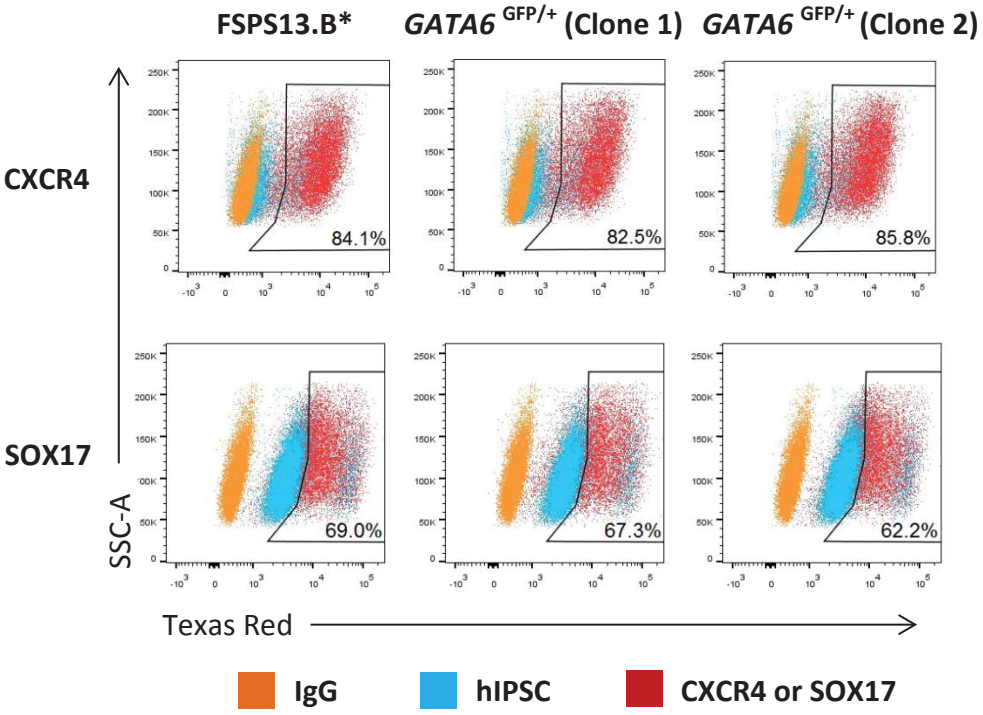


Figure 78. The number of CXCR4+ and SOX17+ cells is not decreased in both clones 1 and 2 of *GATA6* $^{GFP/+}$ FSPS13.B cells. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least 2 independent experiments.

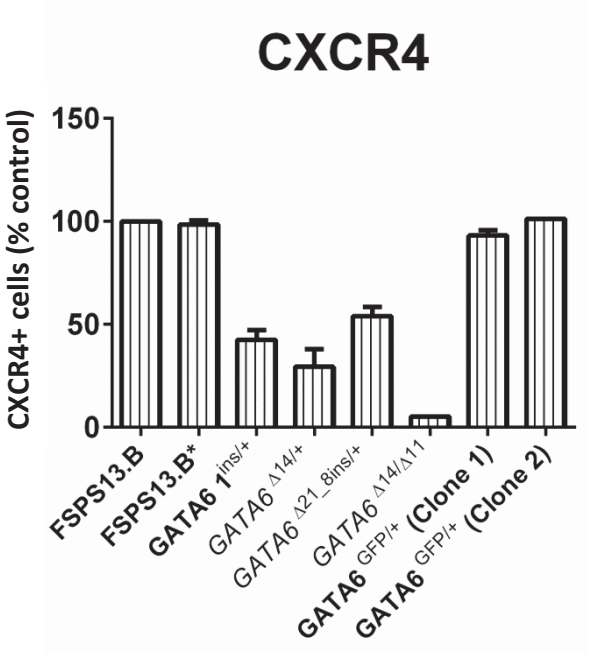


Figure 79. Summary of CXCR4 levels via FACS for all FPSP13.B-derived mutant cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of FSPS13.B* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

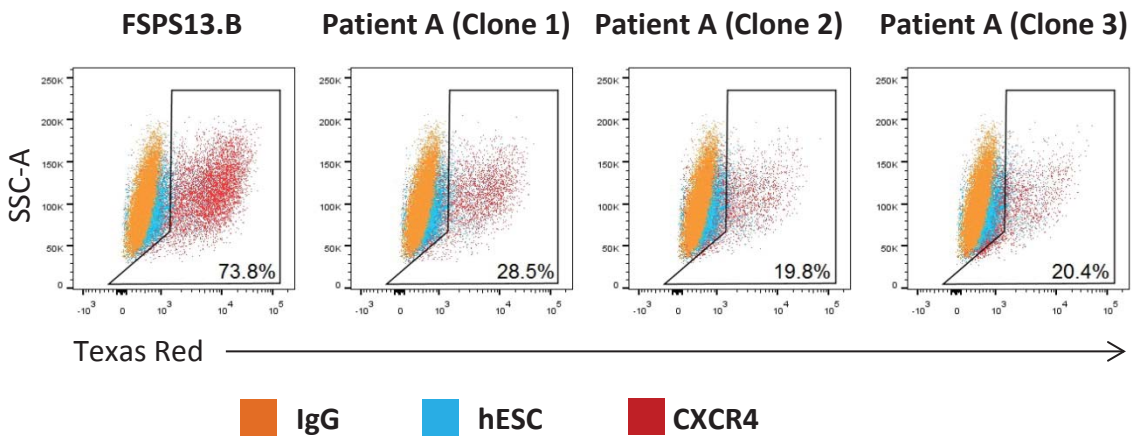


Figure 80. The number of CXCR4+ cells is decreased in Patient A. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment that is representative of at least 3 independent experiments.

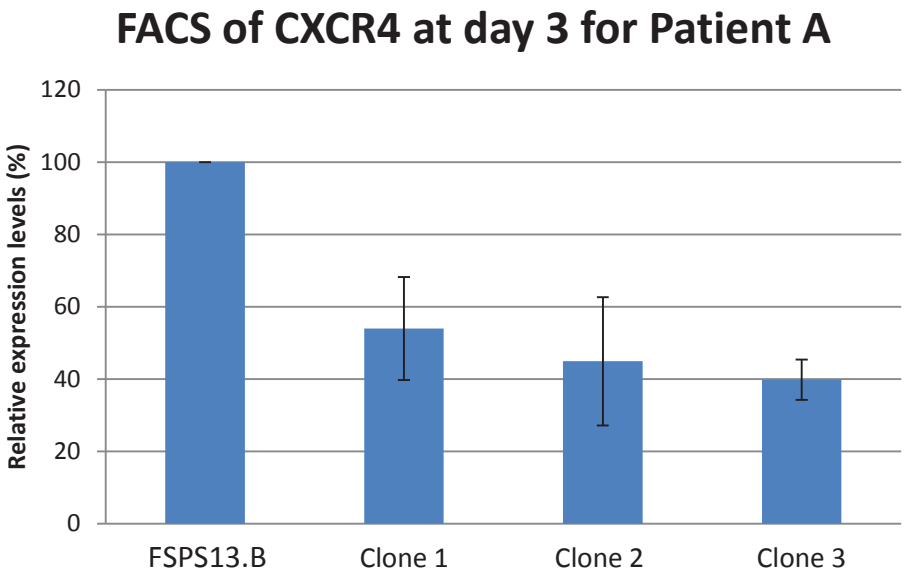


Figure 81. Summary of CXCR4 levels via FACS for Patient A cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of clones 1, 2 and 3 of Patient A are shown. Data show results of two experiments and error bars indicate standard deviation.

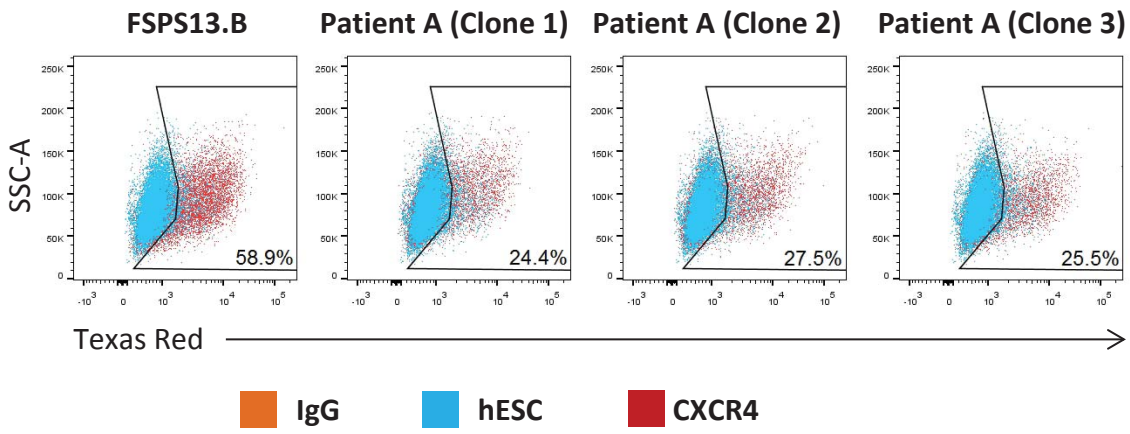


Figure 82. The number of CXCR4+ cells is decreased in Patient B. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Data show results of one experiment.

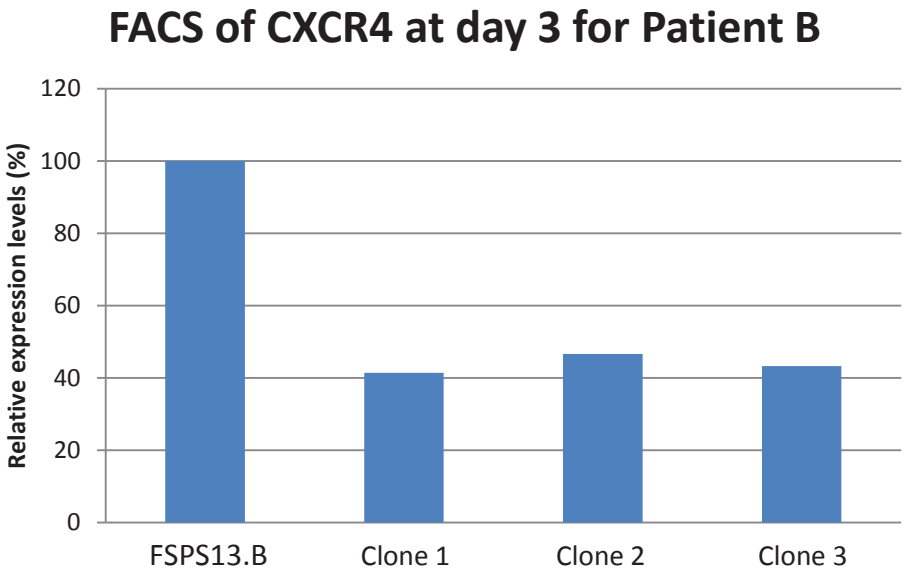


Figure 83. Summary of CXCR4 levels via FACS for Patient B cells on day 3. Wild-type FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of clones 1, 2 and 3 of Patient B are shown. Data show results of one experiment.

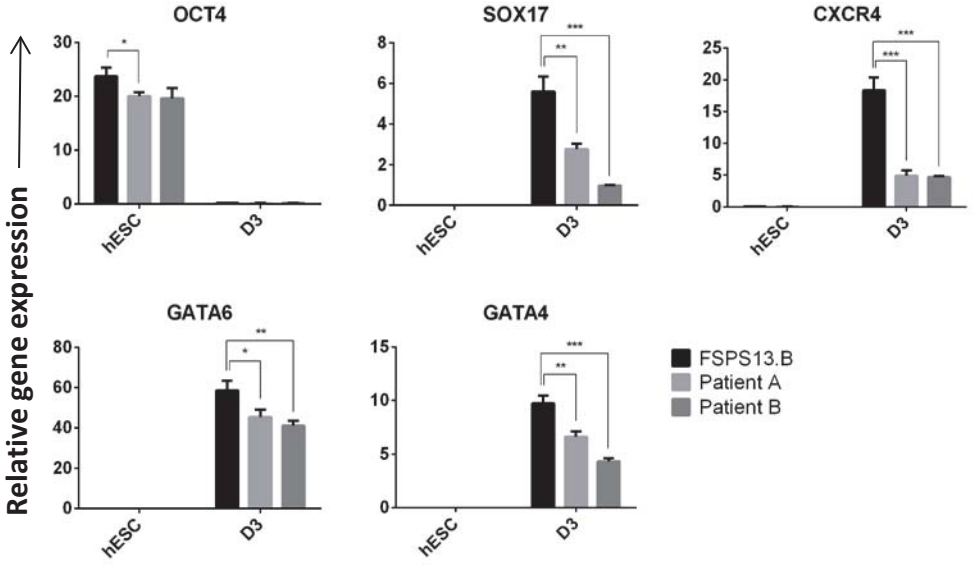


Figure 84. qRT-PCR analyses of Patient A and B cells on day 3. RNA was extracted on day 3 and the expression patterns of key endoderm markers were determined. Data are triplicate samples of one experiment and representative of two independent experiments (Patient A) and one experiment (Patient B). Error bars indicate standard deviation.

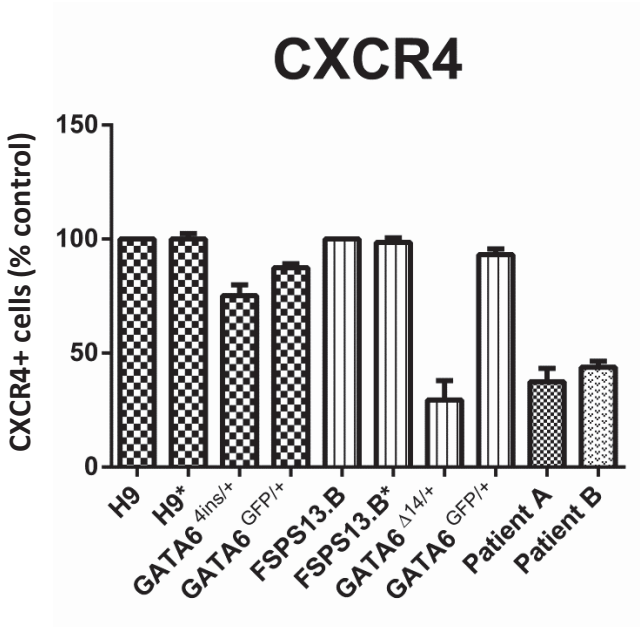


Figure 85. Summary of CXCR4 levels via FACS for H9 and FSPS13.B selected heterozygous mutant cells, and Patients A and B on day 3. H9 and FSPS13.B cells are normalised to 100% and the relative CXCR4 expression levels of their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation.

3.3.4. *GATA6* heterozygous mutants display similar endodermal defects using a commercial kit from STEMCELL Technologies

I further validated these results using the STEMdiff Definitive Endoderm kit from Stem Cell Technologies (SCT) using H9*, H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines. A similar endoderm defect was observed in the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells using this protocol as compared to the lab protocol. FACS analyses at day 3 showed that SOX17+ and CXCR4+ cells were decreased by approximately 30-50% in *GATA6* $^{4ins/+}$ cells and 90% in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 86). This was consistently seen in qRT-PCR analyses (Figure 87). Expression patterns of *OCT4*, *GATA6* and *GATA4* on days 3 and 6 also resembled that of the lab's in-house protocol (Figure 87). These results using a commercially available kit align well with those presented earlier in this chapter and give me confidence in my conclusion that GATA6 haploinsufficiency indeed impacts early formation of the DE lineage.

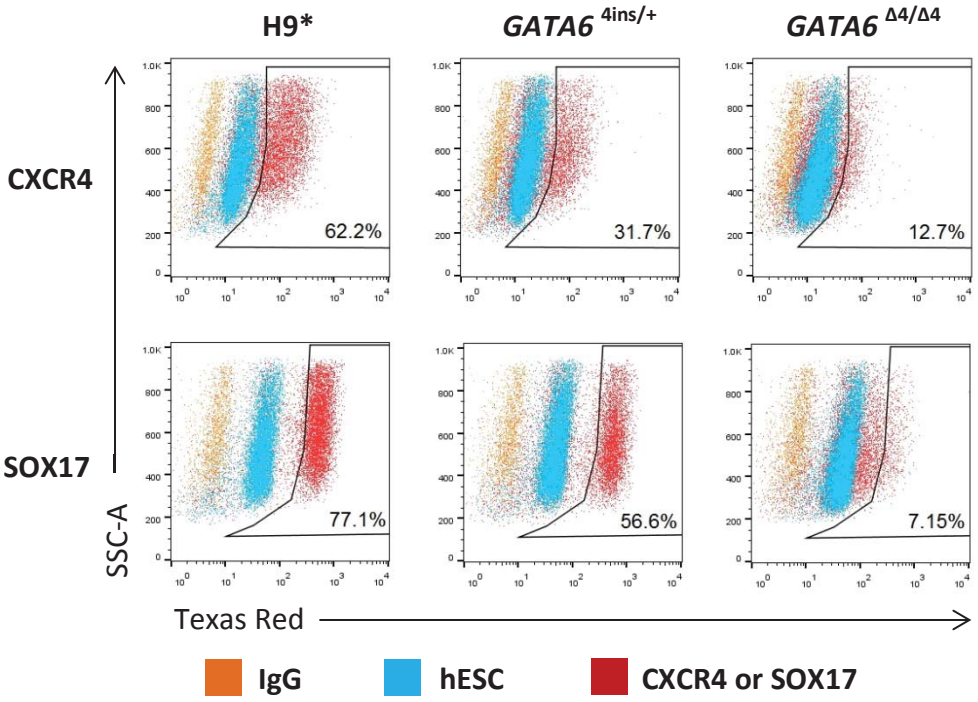


Figure 86. The number of CXCR4+ and SOX17+ cells is decreased in *GATA6* $^{4ins/+}$ cells and are almost completely absent in *GATA6* $^{\Delta4/\Delta4}$ H9 cells differentiated via STEMCELL Technologies kit. Cells were fixed on day 3 and were stained for the DE markers SOX17 and CXCR4. Data show results of one experiment that is representative of at least two independent experiments.

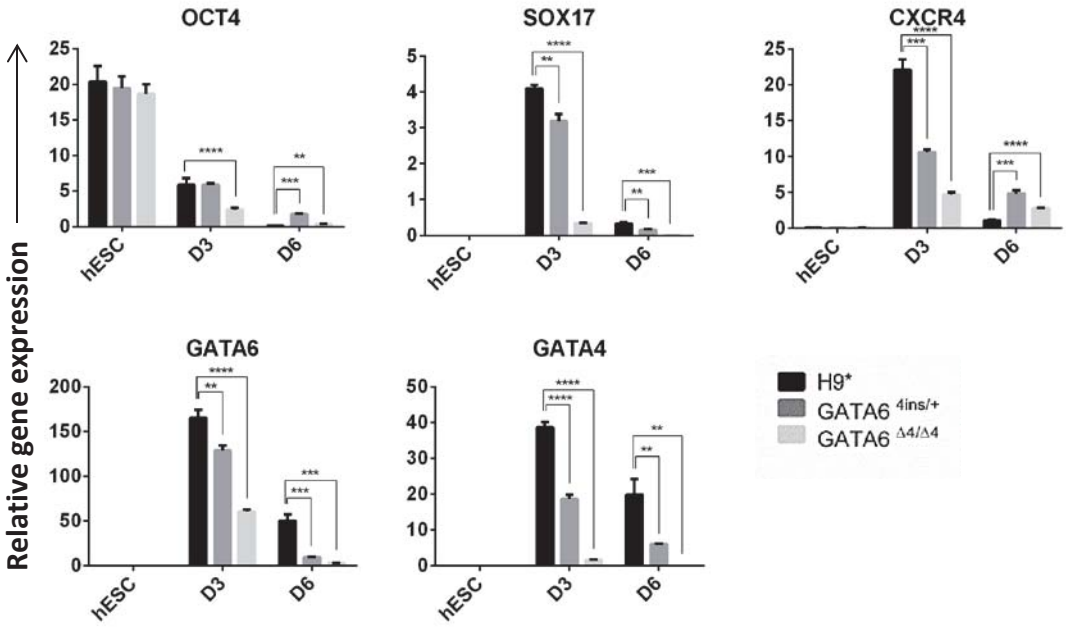


Figure 87. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3 and 6 differentiated via STEMCELL Technologies kit . RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

3.3.5. *GATA6* heterozygous mutants did not exhibit endodermal defects using PSC Definitive Endoderm Induction Kit from Life Technologies

In addition to the commercial kit from SCT, I also performed DE differentiation using PSC Definitive Endoderm Induction Kit from Life Technologies. H9 and FPSP13.B cells were differentiated alongside the TALEN-derived and Patient B lines as a control for DE differentiation efficiency compared to the lab and the STEMdiff Definitive Endoderm kit from SCT. From FACS analyses, CXCR4+ cells were consistently over 80% in both H9 and FSPS13.B cells, suggesting a slightly more efficient DE differentiation than the lab and SCT protocols (Figure 88).

TALEN-derived FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$, *GATA6* $^{\Delta 14/\Delta 11}$ cell lines and Patient B (clone 1) were differentiated using the PSC DE induction kit. Interestingly, FACS analyses revealed no endodermal defect in the *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ and Patient B heterozygous mutant cell lines (Figure 88). qRT-PCR analyses recapitulate results from FACS, except for Patient B (Figure 89). However, qRT-PCR was only performed once, whereas FACS was performed twice. So, it is possible that there could have been an error with the qRT-PCR result of Patient B. FACS and qRT-PCR results for *GATA6* $^{\Delta 14/\Delta 11}$ cell line was consistent with both lab and SCT protocols, with an approximately 90% decrease of CXCR4+ cells (Figure 88 and Figure 89).

I attempted to pursue this inconsistency between protocols further by requesting to know the components of the PSC DE Kit, but Life Technologies was unwilling to share the formulation of their kit. Hence, I was unable to compare the components between the protocols and determine which growth factor(s) and/or pathway inhibitors could have attributed to this discrepancy in the results. Taken together, the results from the PSC DE kit suggest that the impaired DE phenotype seen in the heterozygous *GATA6* mutant lines could be attributed to a protocol-dependent defect.

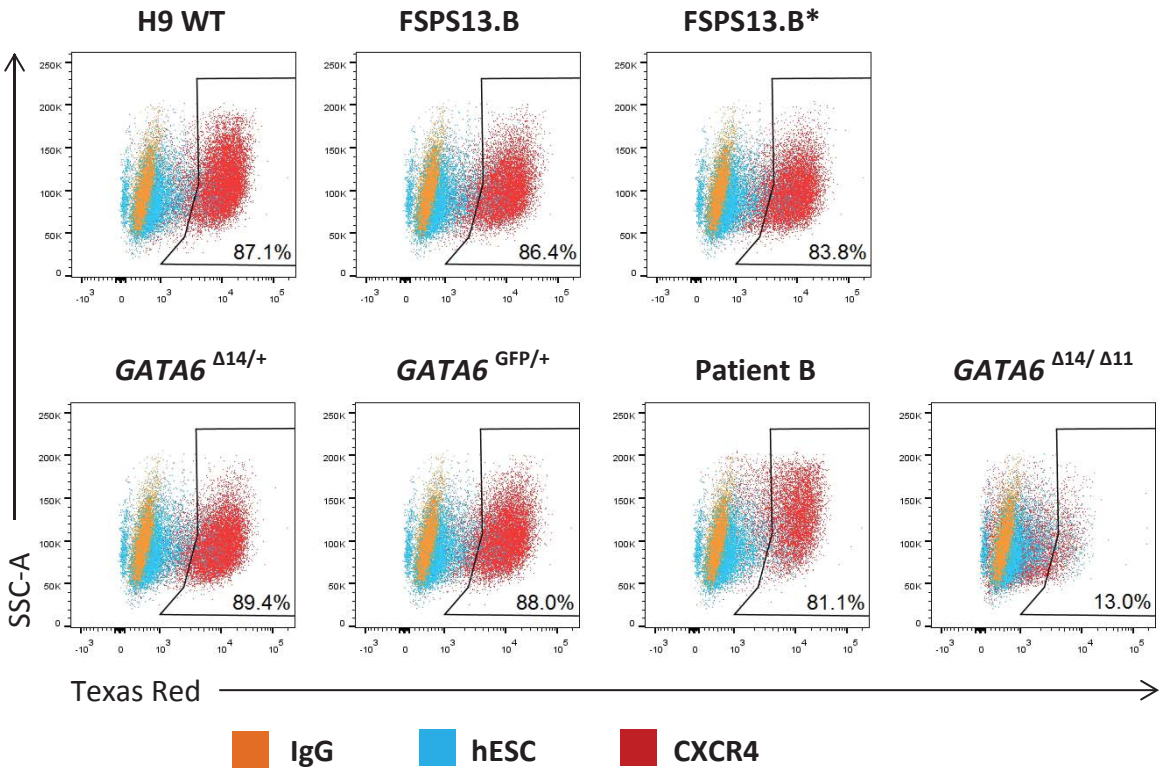


Figure 88. The number of CXCR4+ cells is not decreased in *GATA6* heterozygous FSPS13.B mutant cells and Patient B differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. Cells were fixed on day 3 and were stained for the DE marker CXCR4. Gates were set according to hESC. Data show results of one experiment that is representative of two independent experiments.

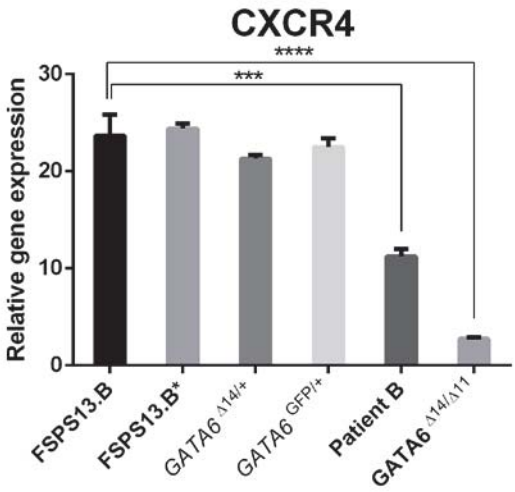


Figure 89. qRT-PCR analyses of FSPS13.B, FPSP13.B*, FSPS13.B-derived mutant cells and Patient B on day 3 differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. RNA was extracted on day 3 and the expression of *CXCR4* was determined. Data are triplicate samples of one experiment, error bars indicate standard deviation.

3.4. *GATA6* is required for differentiation into the pancreatic lineage

To further study the effects of heterozygous and homozygous loss of *GATA6* during human pancreatic development, I continued the differentiation process with the mutant lines in order to characterise their later phenotypes during acquisition of pancreatic identity (pancreatic endoderm; PE on day 12) and allocation to the endocrine lineage (endocrine progenitors; EP on day 24).

3.4.1. Homozygous *GATA6* mutants fail to enter the pancreatic lineage

To characterise the phenotype of homozygous loss of *GATA6* using our *in vitro* model system, I differentiated H9 TALEN-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells toward the EP stage. Expectedly, ICC analyses of *GATA6* $^{\Delta4/\Delta4}$ cells showed the absence of key markers PDX1 at day 12 (Figure 90), NGN3 at day 15 (Figure 91) and C-PEPTIDE, SST and GCG at day 24 (Figure 92), indicating that the *GATA6* $^{\Delta4/\Delta4}$ mutant entirely failed to enter the pancreatic lineage.

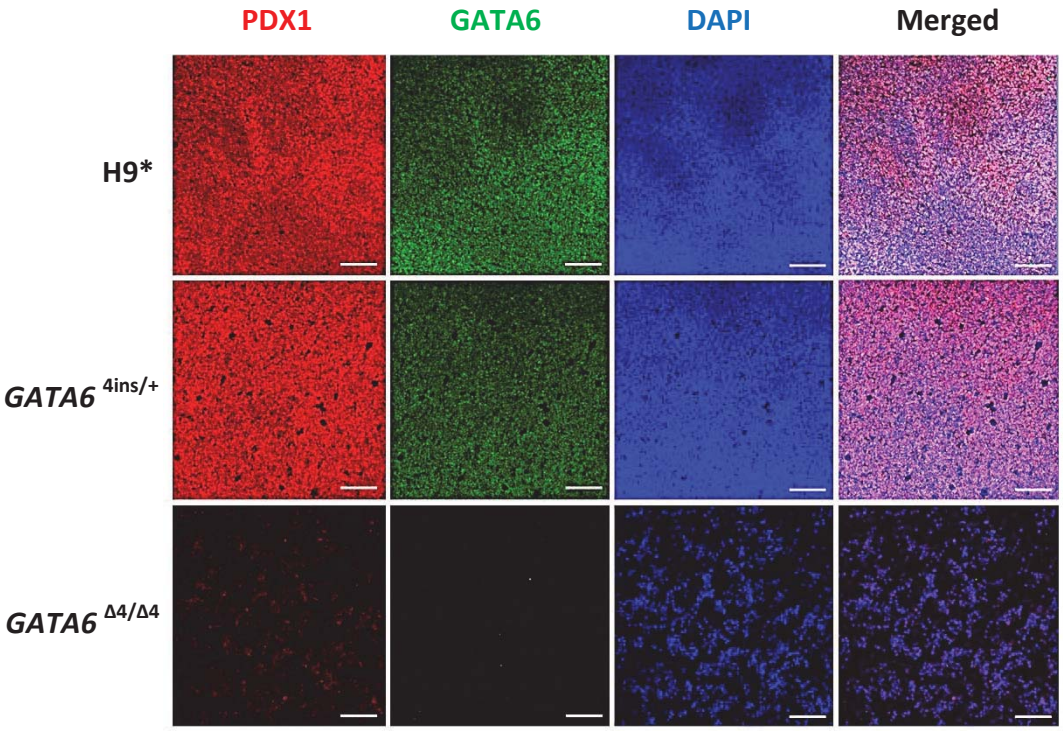


Figure 90. PDX1 is not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 12 of differentiation. Cells were grown in culture conditions that specified them toward the pancreatic progenitor lineage and analysed via immunofluorescence. Cells were fixed on day 12 and were stained for the pancreatic marker PDX1. Scale bar, 100 µm.

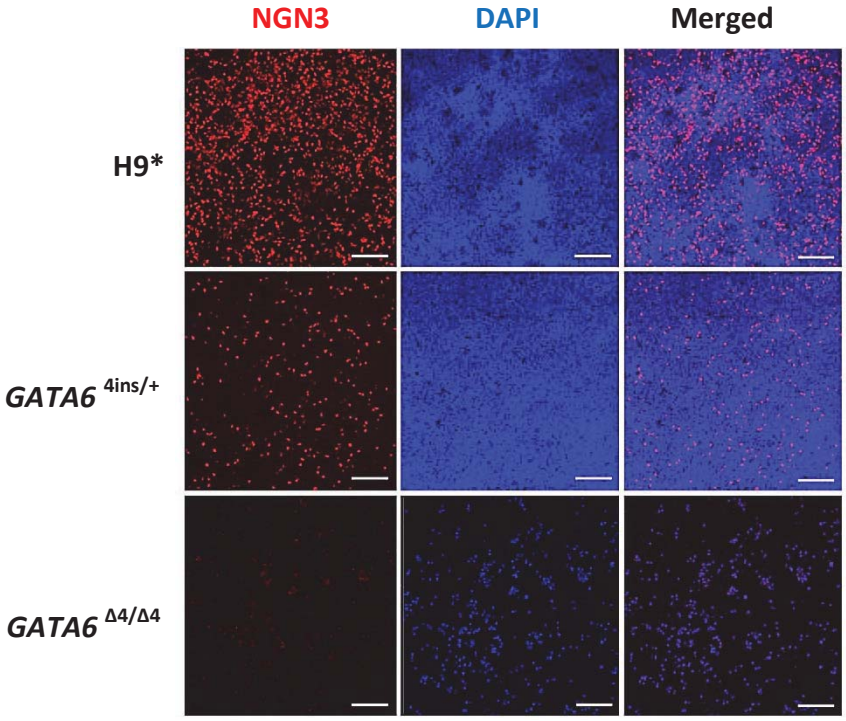


Figure 91. NGN3 is not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 15 of differentiation. Cells were grown in culture conditions that specified them toward the endocrine lineage and analysed via immunofluorescence. Cells were fixed on day 15 and were stained for the key endocrine marker NGN3. Scale bar, 100 µm.

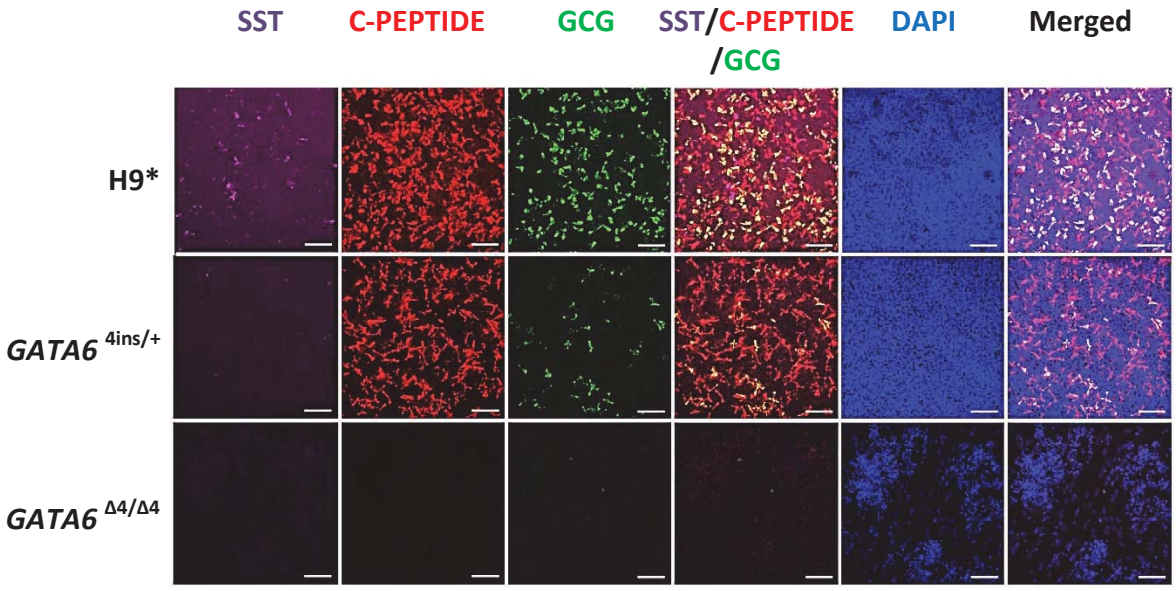


Figure 92. C-PEPTIDE, SST and GCG expression are not activated in *GATA6* $^{\Delta4/\Delta4}$ cells by day 24 of differentiation. Cells were grown in culture conditions that mature into endocrine progenitors and analysed via immunofluorescence. Cells were fixed on day 24 and were stained for key markers C-PEPTIDE, SST and GCG. Scale bar, 100 µm.

FACS analyses of *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells on day 12 showed a 90-100% loss of PDX1+ cells compared to H9 wild-type and H9* cells (Figure 93 and Figure 94). On day 24, C-PEPTIDE levels in *GATA6* $^{GFP/GFP}$ cells were almost negligible (Figure 95 and Figure 96). In some experiments, *GATA6* $^{\Delta4/\Delta4}$ cells did not survive up to day 24 and died around day 18 (Figure 95).

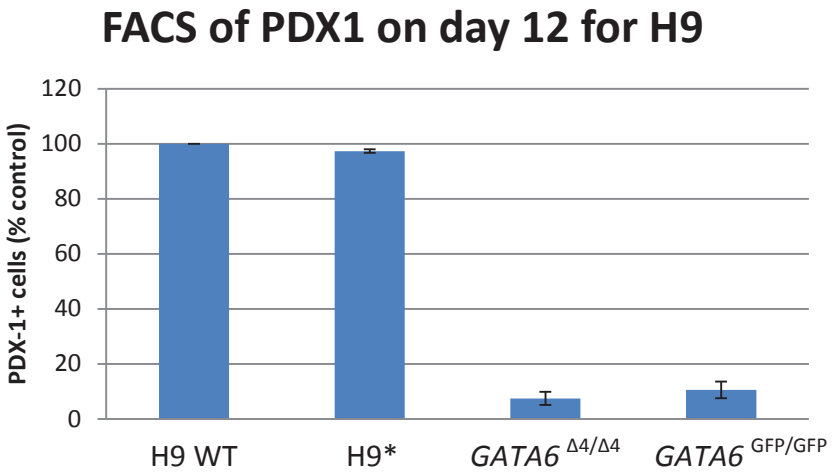


Figure 93. Summary of PDX1 levels via FACS for H9 homozygous mutant cells on day 12. Wild-type H9 cells are normalised to 100% and the relative PDX1 levels of H9*, H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells are shown. Data show results of two independent experiments and error bars indicate standard deviation.

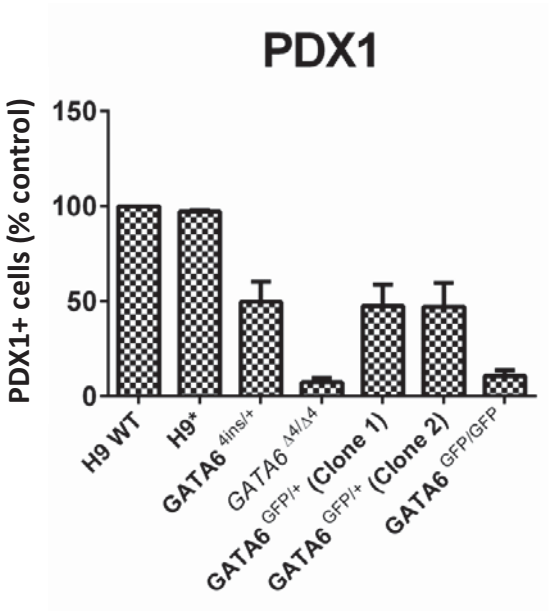


Figure 94. Summary of PDX1 levels via FACS for all H9-derived mutants cells on day 12. Wild-type H9 cells are normalised to 100% and the relative PDX1 levels of H9* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

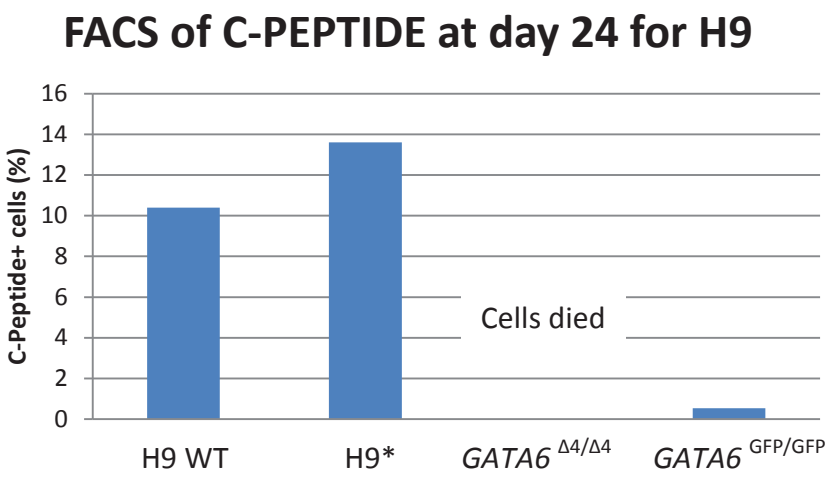


Figure 95. Summary of C-PEPTIDE levels via FACS for H9 homozygous mutant cells on day 24. Absolute percentage of C-PEPTIDE-positive cells in H9 wild-type, H9* and H9-derived *GATA6* $^{\Delta4/\Delta4}$ and *GATA6* $^{GFP/GFP}$ mutant cells are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of two independent experiments.

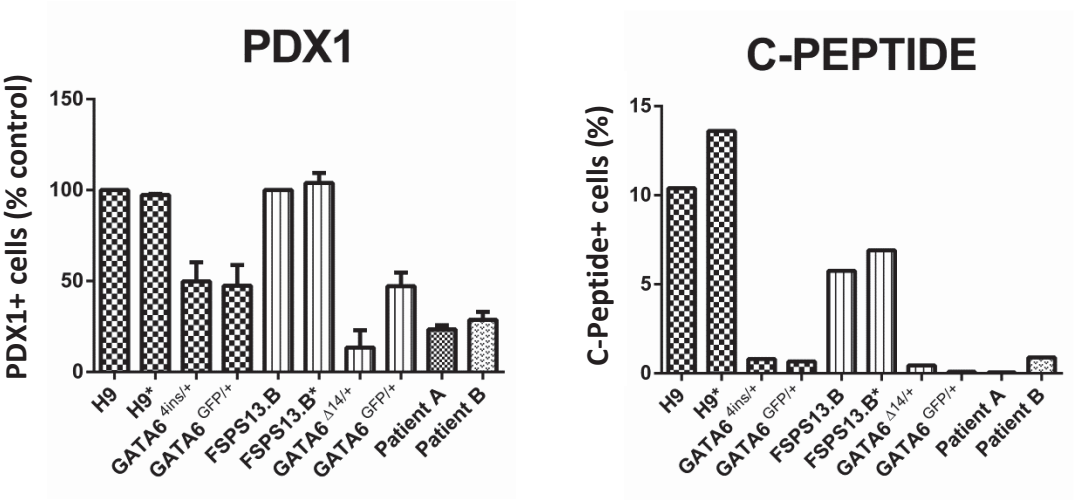


Figure 96. Summary of PDX1 (day 12) and C-PEPTIDE (day 24) expression via FACS for H9 and FSPS13.B selected heterozygous mutant cells, and Patients A and B. For PDX1, H9 and FSPS13.B cells are normalised to 100% and the relative PDX1 levels of their respective mutant cell lines are shown. Data show results of three independent experiments and error bars indicate standard deviation. For C-PEPTIDE, absolute percentages of C-PEPTIDE-positive cells in all cell lines are shown. Each bar represents one biological sample, and the graph was taken from one experiment, which is representative of three independent experiments.

3.4.2. Heterozygous *GATA6* mutants elicit a pancreatic defect in all protocols

Next, to characterise the phenotype of heterozygous loss of *GATA6* using our *in vitro* model system, I first differentiated H9 TALEN-derived *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ mutant cells via the lab-derived protocol toward the EP stage. FACS analyses of key PE marker PDX1 at day 12 for *GATA6* $^{4ins/+}$ and clones 1 and 2 of *GATA6* $^{GFP/+}$ mutant cells indicate a decrease in PDX1+ cells by about 50% in both lines (Figure 94 and Figure 96), even though *GATA6* $^{GFP/+}$ seemed to have lower levels of *PDX1* (Figure 97). In contrast, ICC analyses of *GATA6* $^{4ins/+}$ cells did not show an observable difference in PDX1 expression levels compared to H9* cells (Figure 90). This observation could be due to the fact that the cells analysed by ICC are grown very densely, resulting in overlapping growth that may mask PDX1-negative cells.

On day 24, C-PEPTIDE levels by FACS were almost negligible in both *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant lines (Figure 96). Similar to *PDX1*, *GATA6* $^{GFP/+}$ have lower levels of *Insulin* mRNA compared to *GATA6* $^{4ins/+}$ (Figure 97). qRT-PCR analyses of key genes showed a similar down-regulation of *HNF4A* and *GATA4* in both *GATA6* $^{4ins/+}$ and *GATA6* $^{GFP/+}$ mutant lines at day 24 (Figure 97). *HLXB9*, on the other hand, was strongly decreased in *GATA6* $^{4ins/+}$ mutant cells but not in *GATA6* $^{GFP/+}$ mutant cells at day 24 (Figure 97).

In addition, I performed C-peptide ELISA on H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Similar to the H9 wild-type cells that was described earlier in Chapter 3.1.1, H9* elicited an inverse C-peptide release response upon glucose stimulation (Figure 98). The levels of C-Peptide stimulation and down-regulation in H9*is very similar to H9 wild-type, suggesting the lack of off target effects from the TALEN targeting, if any. In contrast, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant lines did not elicit any glucose-stimulated C-peptide release, and the basal C-Peptide levels these lines are much lower than H9 wild-type and H9* (Figure 98).

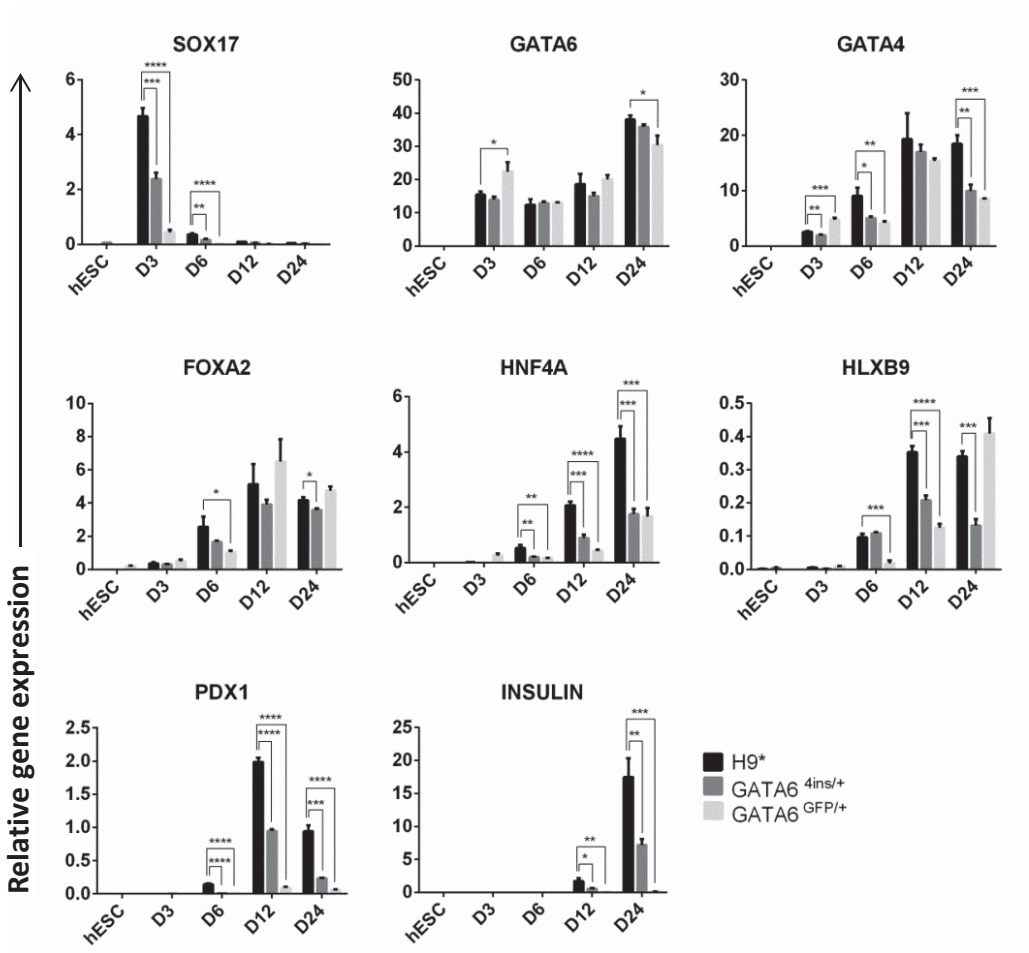


Figure 97. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3, 6, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

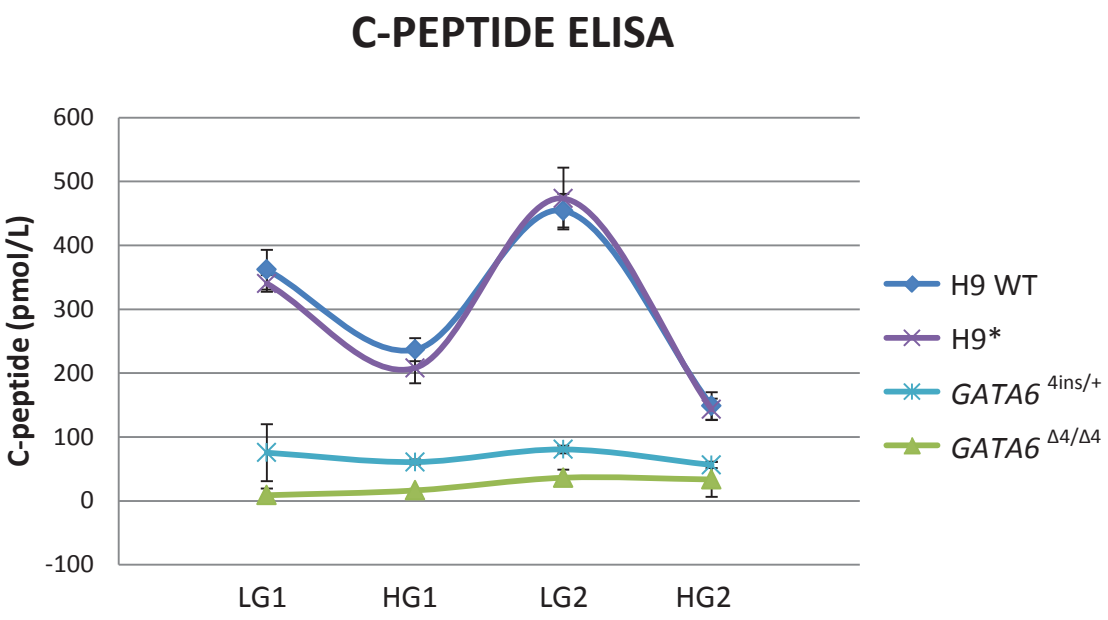


Figure 98. C-peptide secretion of TALEN-derived H9 mutants upon glucose stimulation on day 24. Cells were differentiated using the lab-derived protocol and assayed on day 24. Data are presented as the average of 3 biological replicates of one experiment and representative of two independent experiments. Error bars indicate standard deviation. LG1 is first incubation of low glucose; HG1 is first incubation of high glucose; LG2 is second incubation of low glucose; HG2 is second incubation of high glucose.

In FSPS13.B TALEN-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells, *PDX1* levels were also strongly down-regulated on days 12 and 24 by qRT-PCR (Figure 99). FACS analyses showed an approximately 80-90% decrease of PDX1+ cells in *GATA6* $^{\Delta 14/+}$ mutant cells and 50% decrease of PDX1+ cells in *GATA6* $^{GFP/+}$ mutant cells on day 12 (Figure 96). PDX1 FACS analyses for all the other heterozygous mutant lines are shown in Figure 100. qRT-PCR analyses of key genes showed down-regulation of *HNF4A* and *Insulin* on days 12 and 24, but no change for *HLXB9* (Figure 99). *GATA4* expression was down-regulated in *GATA6* $^{\Delta 14/+}$ mutant cells at all time points but remained relatively unchanged in *GATA6* $^{GFP/+}$ mutant cells (Figure 99). On day 24, C-PEPTIDE+ cells in both *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant lines numbered near zero (Figure 96).

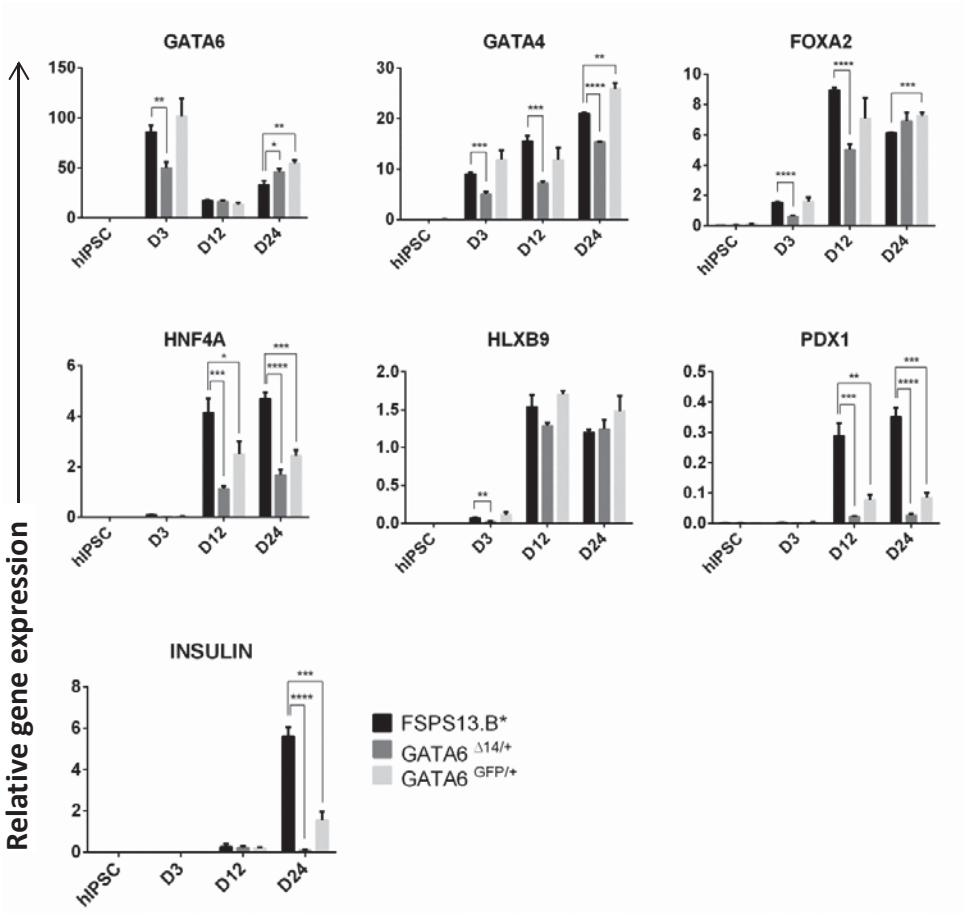


Figure 99. qRT-PCR analyses of FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells on days 3, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

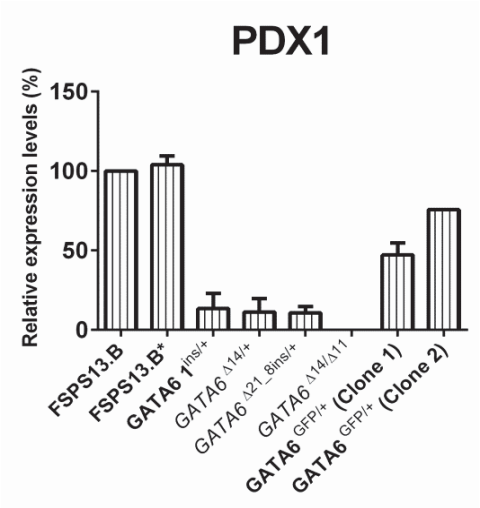


Figure 100. Summary of *CXCR4* levels via FACS for all FPSP13.B-derived mutant cells on day 12. Wild-type FSPS13.B cells are normalised to 100% and the relative *CXCR4* levels of FSPS13.B* or mutant cells are shown. Data show results of two experiments and error bars indicate standard deviation.

In Patients A and B, *PDX1* levels were also strongly down-regulated on days 12 and 24, as shown by qRT-PCR (Figure 101). FACS analyses showed an approximately 70-80% decrease of PDX1+ cells in both Patients A and B lines on day 12 (Figure 96). qRT-PCR analyses of key genes showed down-regulation of *GATA4*, *FOXA2*, *HNF4A*, *HLXB9* and *Insulin* on days 12 and 24 (Figure 101). On day 24, FACS of C-PEPTIDE+ cells were near zero in both Patient A and B lines (Figure 96).

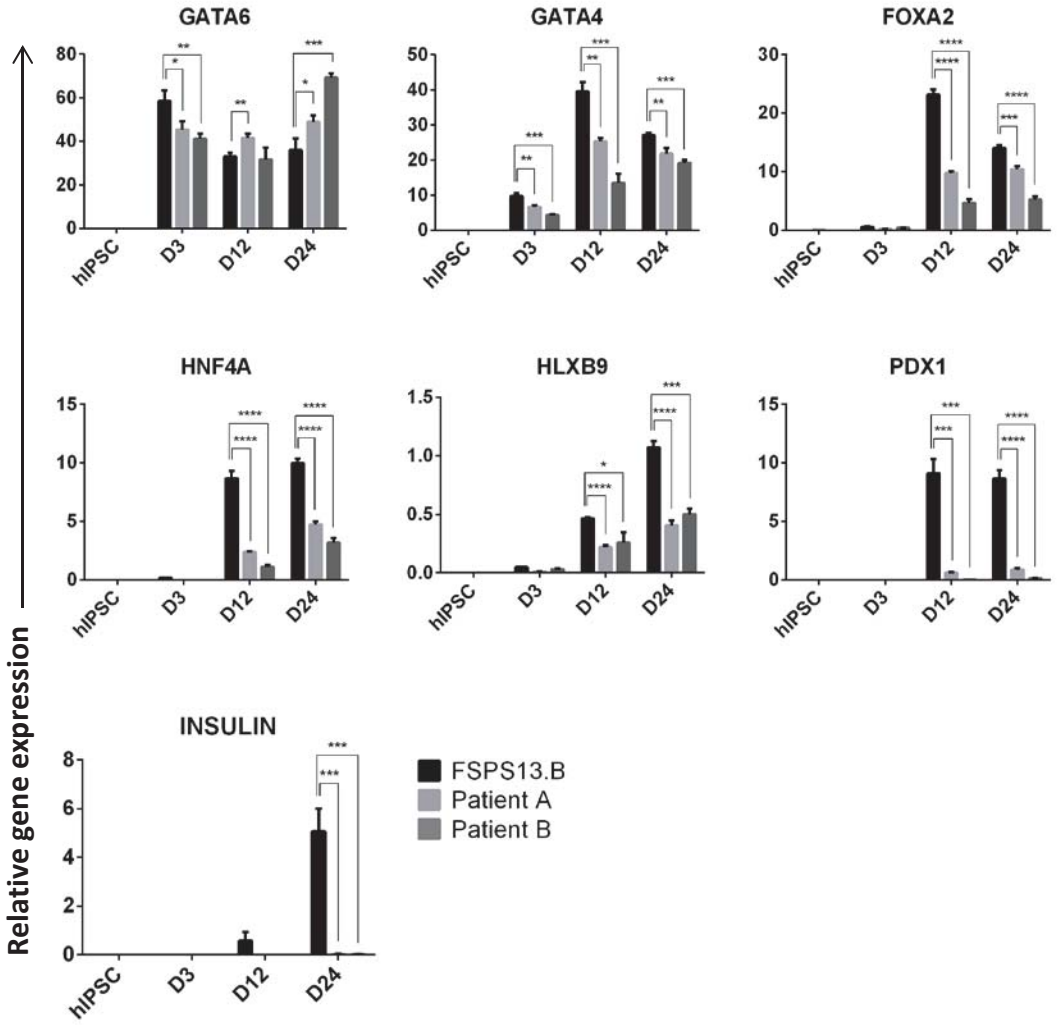


Figure 101. qRT-PCR analyses of FSPS13.B, Patient A and Patient B cells on days 3, 12 and 24. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

Using the STEMdiff pancreatic progenitor kit from SCT, I differentiated H9*, H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ lines toward the PE lineage. A similar pancreatic defect was observed in the H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells using this protocol as compared to the lab protocol. FACS analyses on day 12 showed that PDX1+ cells were decreased by approximately 50% in *GATA6* $^{4ins/+}$ cells and absent in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 102). This was consistently seen in qRT-PCR analyses where key markers such as *GATA4*, *HNF4A*, *PDX1* and *NKX6-1* were strongly down-regulated in *GATA6* $^{4ins/+}$ mutant cells and absent in *GATA6* $^{\Delta4/\Delta4}$ cells (Figure 103). These results complement the results derived from the lab protocol.

Using the PSC Definitive Endoderm Induction Kit from Life Technologies, I differentiated FSPS13.B* and FSPS13.B-derived mutants *GATA6* $^{\Delta14/+}$, *GATA6* $^{GFP/+}$ and *GATA6* $^{\Delta14/\Delta11}$ mutant cell lines and Patient B toward the PE lineage. Interestingly, despite the heterozygous mutants showing no impairment of DE formation (described in Chapter 3.3.5), a very similar pancreatic defect was observed in the heterozygous mutants using this protocol as compared to the lab protocol. FACS analyses on day 12 showed that PDX1+ cells were decreased by approximately 80-90% in *GATA6* $^{\Delta14/+}$cells and Patient B cells, 60% in *GATA6* $^{GFP/+}$ cells and is almost completely absent in *GATA6* $^{\Delta14/\Delta11}$ cells (Figure 104). qRT-PCR analysis focusing on *PDX1* on day 12 confirms the FACS data (Figure 105).

Taken together, these finding confirm that complete loss of *GATA6* in TALEN-edited hPSCs results in failure of pancreatic development, which is most likely explained by the disruption of the GRN required for earlier DE specification. Heterozygous loss of *GATA6* seems to impair pancreatic formation (day 12) by 50-90%, which is a smaller variation in phenotype penetrance across all heterozygous mutants as compared to DE. By day 24, all heterozygous mutants fail to differentiate into endocrine cells at similar rates, thereby reducing the variation in phenotype penetrance further. The similarity of phenotypes between the lab-derived, SCT and Life Technologies protocols suggest that the pancreatic phenotypes observed in the heterozygous mutants are not a consequence of the differentiation protocol used, and are true effects of heterozygous loss of *GATA6*.

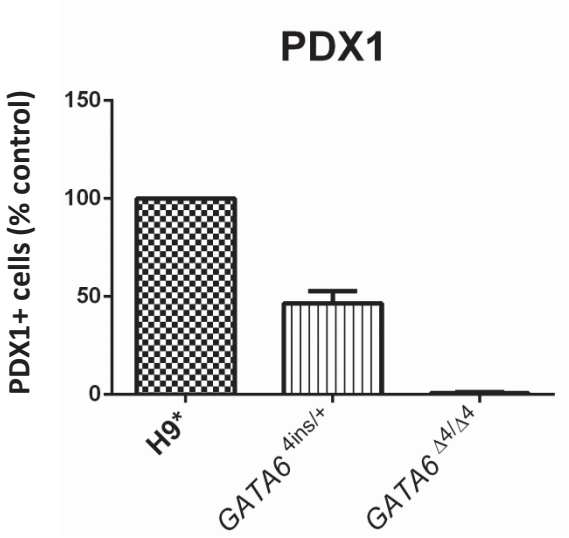


Figure 102. The number of PDX1+ cells is decreased in *GATA6* $^{4ins/+}$ cells and is almost completely zero in *GATA6* $^{\Delta4/\Delta4}$ H9 cells differentiated via STEMCELL Technologies kit. Cells were fixed on day 12 and were stained for PDX1. Data show results of one experiment that is representative of at least two independent experiments.

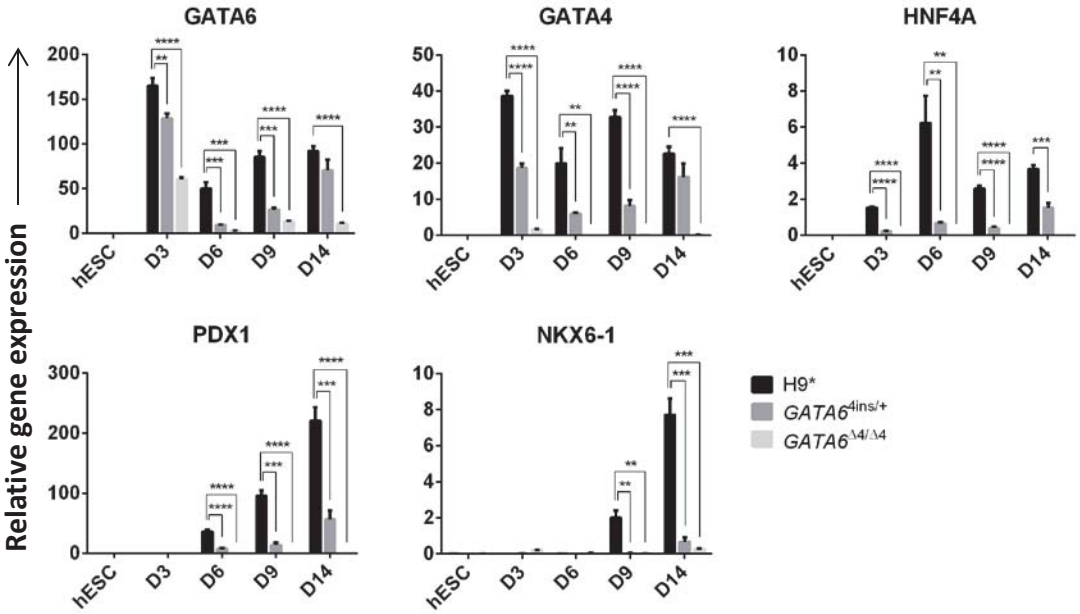


Figure 103. qRT-PCR analyses of H9*, *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ cells on days 3, 6, 9 and 14 differentiated via STEMCELL Technologies kit. RNA was extracted at specific stages and the expression patterns of key markers were determined. Data are triplicate samples of one experiment and representative of three independent experiments. Error bars indicate standard deviation.

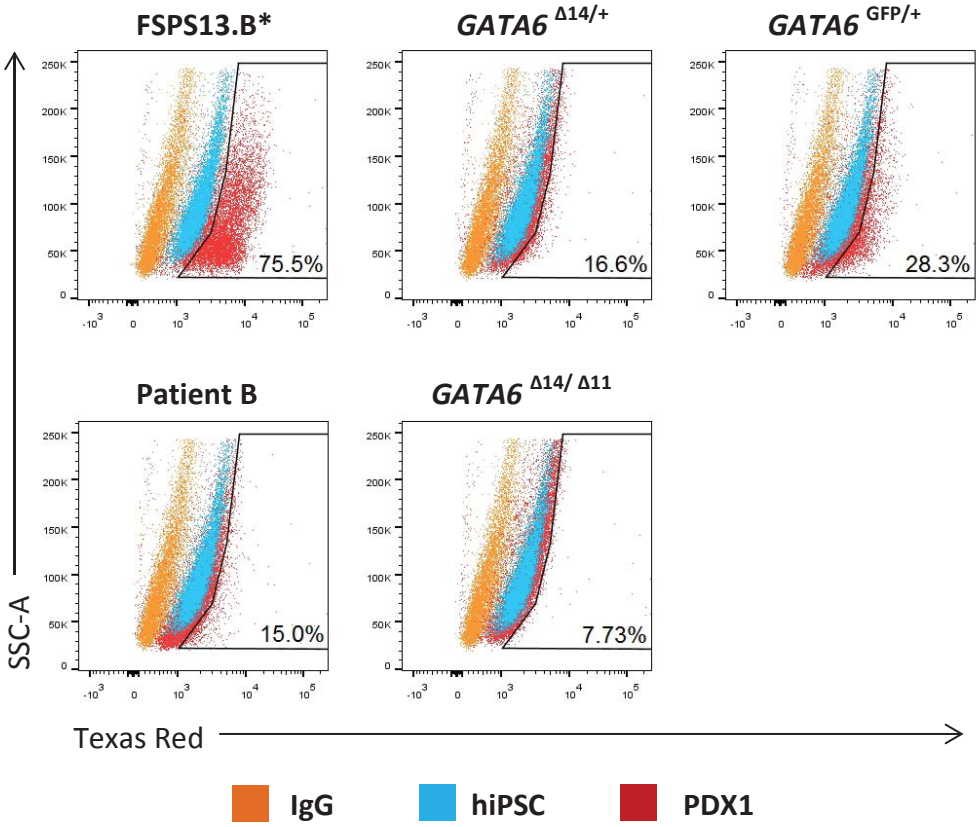


Figure 104. The number of PDX1+ cells is decreased in *GATA6* heterozygous FSPS13.B mutant cells and Patient B differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. Cells were fixed on day 12 and were stained for PDX1. Data show results of one experiment that is representative of at least two independent experiments.

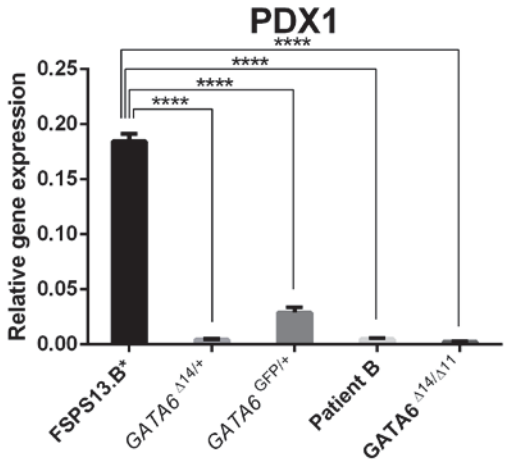


Figure 105. qRT-PCR analyses of FPSP13.B*, FSPS13.B-derived mutant cells and Patient B on day 12 differentiated via PSC Definitive Endoderm Induction Kit from Life Technologies. RNA was extracted on day 12 and the expression of *PDX1* was determined. Data are triplicate samples of one experiment, error bars indicate standard deviation.

3.5. *GATA6* is a key regulator of DE and pancreatic specification

Next, to investigate if *GATA6* is a master regulator of the DE and pancreatic transcriptional networks, I conducted RNA-sequencing (RNA-seq) on H9*, H9-derived *GATA6* $^{4ins/+}$, *GATA6* $^{\Delta4/\Delta4}$ and Patient A (clones 1-3) mutant cells at the DE and pancreatic stages of the lab differentiation protocol. Specifically, I performed RNA-seq for biologically triplicate samples on days 3 and 12 for *GATA6* $^{4ins/+}$ mutant cells, days 3 and 12 for Patient A cells, days 2 and 3 for *GATA6* $^{\Delta4/\Delta4}$ mutant cells and days 2, 3, and 12 for H9* cells. In addition, to identify the direct interacting partners of GATA6, I also perform chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for biologically duplicate samples on days 3 and 12 for H9* and H9-derived *GATA6* $^{4ins/+}$ mutant cells, and on day 12 for FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta14/+}$, *GATA6* $^{GFP/+}$ mutant cells. Pre-processing and downstream analyses of the RNA- and ChIP-seq data were performed with the help of Dr. Pedro Madrigal. These experiments will elucidate the molecular mechanism of *GATA6* and reveal the functional consequences and global transcriptional profiles of *GATA6*.

3.5.1. Loss of GATA6 perturbs the DE transcriptional network and promotes mesoderm formation

Using RNA-seq data, I first performed differential gene expression analyses between *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells at day 2 to investigate the effect of biallelic loss of *GATA6* on the GRN prior to DE formation. I found 4,679 differentially expressed genes between H9* and *GATA6* $^{\Delta4/\Delta4}$ cells (Table S1) (adjusted p-value 0.01; fold-change ≥ 2). Of these, 3,649 were protein coding genes; 2,239 down-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells and 1,410 up-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells. Interestingly, gene ontology (GO) analysis via DAVID revealed the enrichment of many mesodermal developmental pathways, which were found up-regulated in *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 106). Conversely, endodermal developmental pathways were down-regulated in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. In a heat map showing the top 10 most variable genes, *SOX17* was amongst the most highly down-regulated genes, suggesting that the expression of *SOX17* is regulated by *GATA6* (Figure 107).

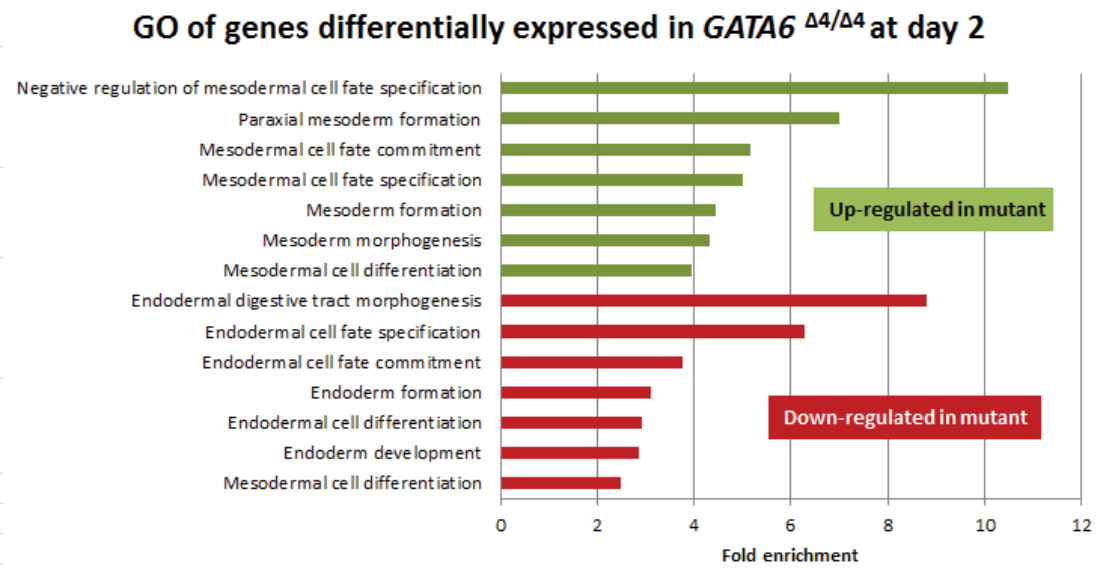


Figure 106. Enriched GO of *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells on day 2 from RNA-seq. GO shows fold enrichment of many mesodermal and endodermal developmental pathways up- and down-regulated respectively in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Data show results of triplicate samples in one experiment representative of three independent experiments.

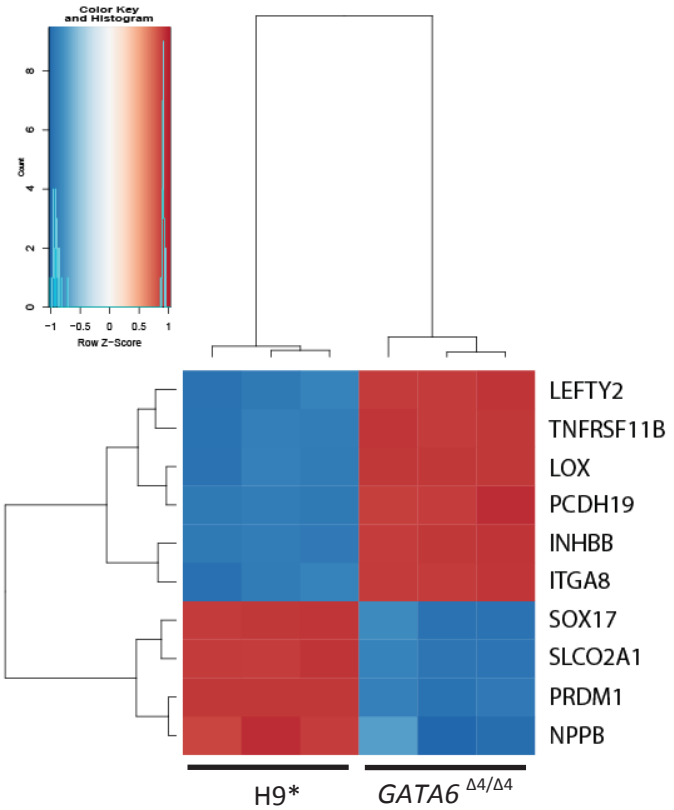


Figure 107. Key DE marker, *SOX17*, is one of the most highly down-regulated genes in *GATA6* $^{\Delta4/\Delta4}$ mutant cells. Heat map illustrating differential gene expression of the top 10 most variable genes in *GATA6* $^{\Delta4/\Delta4}$ mutant cells against H9* cells.

From RNA-seq results on day 3, I found 7,472 differentially expressed genes between H9* and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Table S2), 2,898 genes between H9* and *GATA6* $^{4ins/+}$ (Table S3), and 6,977 between clones 1-3 of Patient A mutant cells (Table S4). Of the 7,472 genes, 5,393 were protein coding genes; 2,702 down-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells and 2,691 up-regulated in *GATA6* $^{\Delta4/\Delta4}$ cells. Of the 2,898 genes, 1,460 were protein coding genes; 729 down-regulated in *GATA6* $^{4ins/+}$ cells and 731 up-regulated in *GATA6* $^{4ins/+}$ cells. Of the 6,977 genes, 3,981 were protein coding genes; 2,154 down-regulated in Patient A cells and 1,827 up-regulated in Patient A cells.

In a heat map derived from RNA-seq results focusing on genes involved in ectoderm, endoderm and mesoderm development, I observed that consistent with my day 2 RNA-seq data, *GATA6* $^{\Delta4/\Delta4}$ mutant cells displayed a decreased endodermal signature, with a concomitant increase in the expression of mesodermal genes (Figure 108). A similar trend was observed when comparing between H9*, *GATA6*$^{4ins/+}$ and clones 1 to 3 of Patient A (Figure 109).

Next, in order to subsequently identify the direct interacting partners of GATA6, I performed ChIP on H9* cells and H9-derived *GATA6* $^{4ins/+}$ cells which were differentiated to the DE stage via the lab protocol to identify those genes bound directly by GATA6. I validated the ChIP via qPCR using primers recognising a region on the human HNF4α P2 promoter as a GATA6 positive control (primer sequences are listed in Chapter 2.10.5). The primer sequences were kindly provided to us by our collaborator Dr. Santi Rodriguez. Indeed, GATA4 and GATA6 have been shown to bind to HNF4α (Sumi et al., 2007). Results from qPCR show a 4-fold increase in binding in H9* cells and only a 1-fold increase in binding in *GATA6* $^{4ins/+}$ mutant cells compared to IgG (Figure 110). To ensure that this was not a result of a decreased number of cells, I performed FACS on GATA6 at the DE stage and found that both H9* and *GATA6* $^{4ins/+}$ cells similarly express approximately 95% GATA6-positive cells (Figure 111), indicating that the decreased DNA binding observed in *GATA6* $^{4ins/+}$ mutant cells is indeed a consequence of heterozygous loss of *GATA6*.

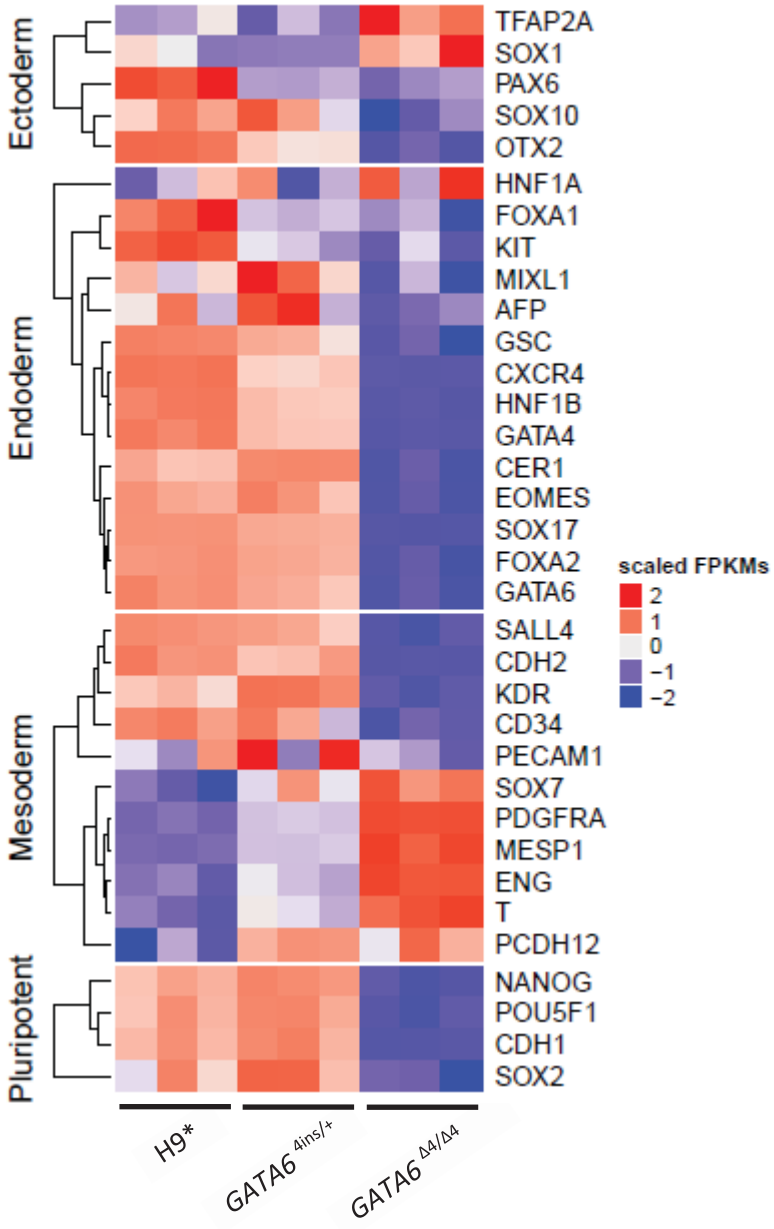


Figure 108. Heat map illustrating differential gene expression of key germ layer markers via RNA-seq between H9*cells and H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells at the DE stage (day 3). *n* = 3 biological replicates for each cell line.

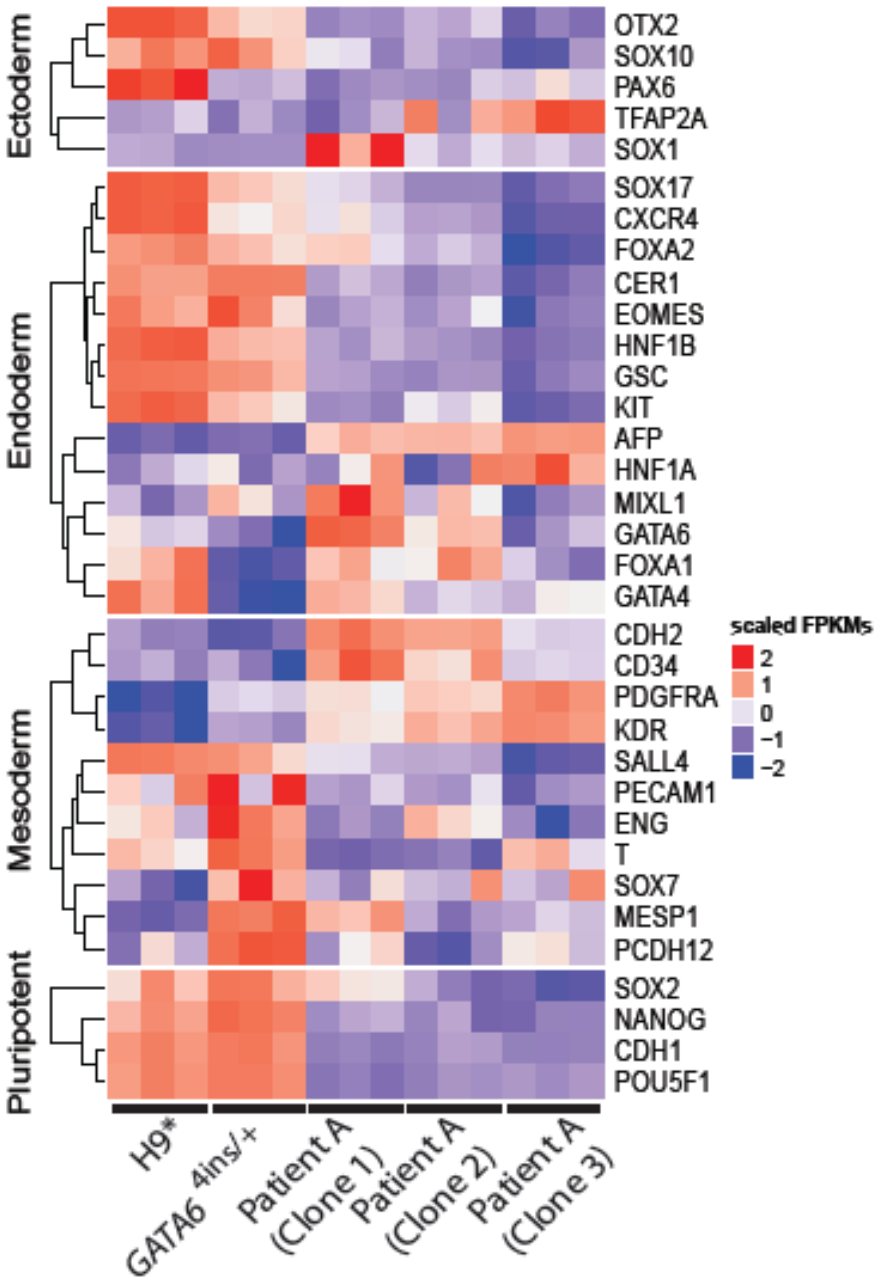


Figure 109. Heat map illustrating differential gene expression of key germ layer markers via RNA-seq between H9*cells and H9-derived *GATA6* $^{4ins/+}$ and clones 1-3 of Patient A mutant cells at the DE stage (day 3). *n* = 3 biological replicates for each cell line.

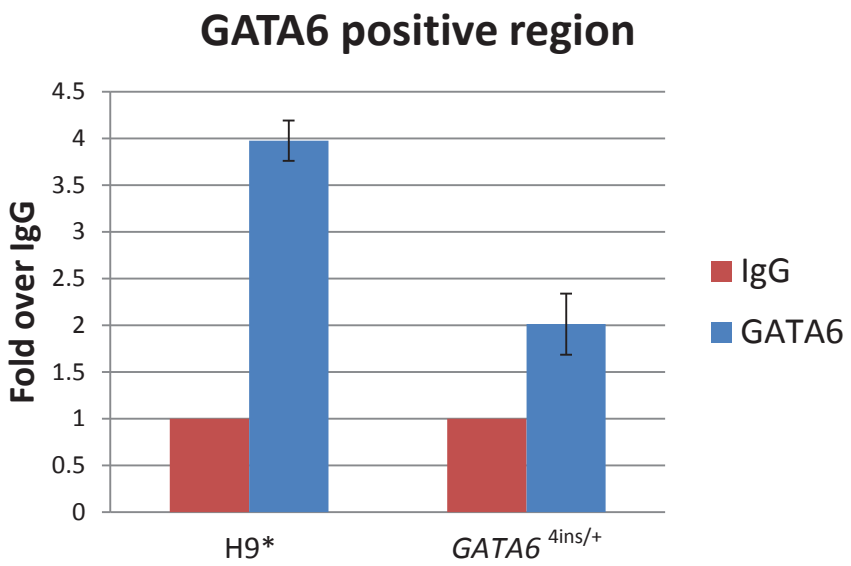


Figure 110. ChIP validation via qPCR using primers specific to a GATA6 positive binding region on day 3. Values of input samples were subtracted from the values of IgG or GATA6 samples and the graph shows fold over IgG (normalised to 1) of H9* and H9-derived *GATA6* $^{4ins/+}$ cells.

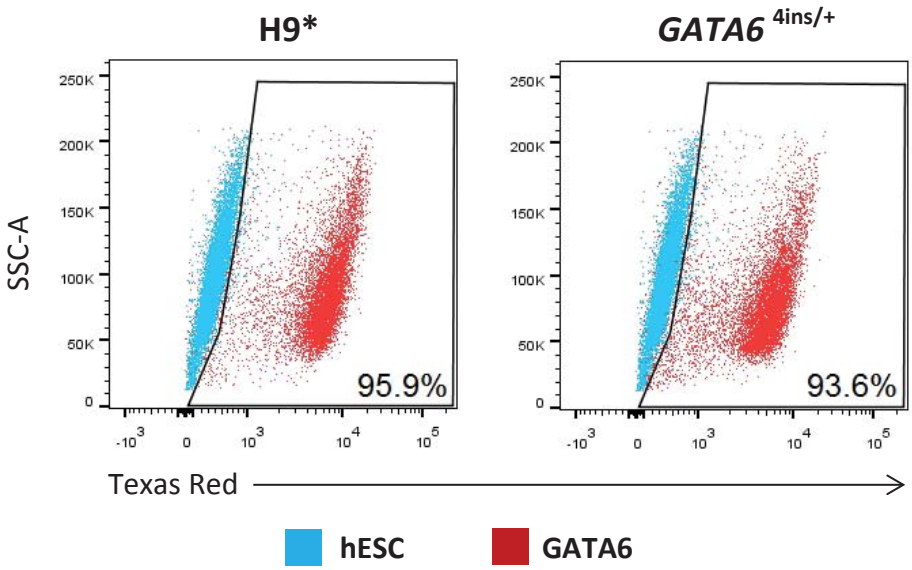


Figure 111. The number of GATA6+ cells is similar in H9* and H9-derived *GATA6* $^{4ins/+}$ cells differentiated via the lab protocol on day 3. Cells were fixed on day 3 and were stained for *GATA6*. Data show results of one experiment that is representative of at least two independent experiments.

Now that I had validated the ChIP, I proceeded to perform ChIP-seq on these samples with the aim of revealing the direct binding targets of GATA6. From the ChIP-seq data at the DE stage, 12,098 peaks were called (IDR ≤ 0.05; median width=417 bp) for H9* cells, of which 10,669 were associated to genes, 4,790 of them protein coding (Table S5). For *GATA6* $^{4ins/+}$ mutant cells, 2,220 peaks were called, of which 1,137 of them were protein coding (Table S6).

Comparing *GATA6* binding at the *GATA4* locus between H9* and *GATA6*$^{4ins/+}$ cells, I saw an enrichment of *GATA6* binding in H9* cells, suggesting that *GATA6* directly regulates *GATA4* during DE development (Figure 112). RNA-seq data showing *GATA4* expression levels in *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells (Figure 112) are also consistent with qRT-PCR data (Figure 68), where *GATA4* is down-regulated in *GATA6* $^{4ins/+}$ cells by approximately 50% and almost completely abolished in *GATA6* $^{\Delta4/\Delta4}$ mutant cells, thus further validating the RNA-seq results.

To relate *GATA6* binding to global gene expression dynamics and investigate genes that were not only direct interacting partners of GATA6, but also differentially expressed, I compared the ChIP-seq analysis to the RNA-seq data set to identify the subset of *GATA6*-bound genes that were up- or down-regulated. In *GATA6* $^{\Delta4/\Delta4}$ mutant cells, I found that 1,120 protein coding genes were *GATA6*-bound and down-regulated including DE genes such as *CXCR4* and *SOX17*, and posterior foregut markers such as *HNF1B* and *HNF4A*, while 745 genes were *GATA6*-bound and up-regulated (Figure 113). 337 and 607 genes were *GATA6*-bound and down-regulated in *GATA6* $^{4ins/+}$ and Patient A cells respectively, while 254 and 616 genes were *GATA6*-bound and up-regulated in *GATA6* $^{4ins/+}$ and Patient A cells respectively (Figure 113). Overlapping the RNA-seq differentially expressed gene sets of *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A identified 143 genes commonly down-regulated and 104 genes commonly up-regulated in these three separate data sets (Figure 114). Key endoderm markers *CXCR4*, *SOX17*, *GATA4* were among the 143 genes commonly down-regulated, suggesting that *GATA6* is a direct regulator of DE development.

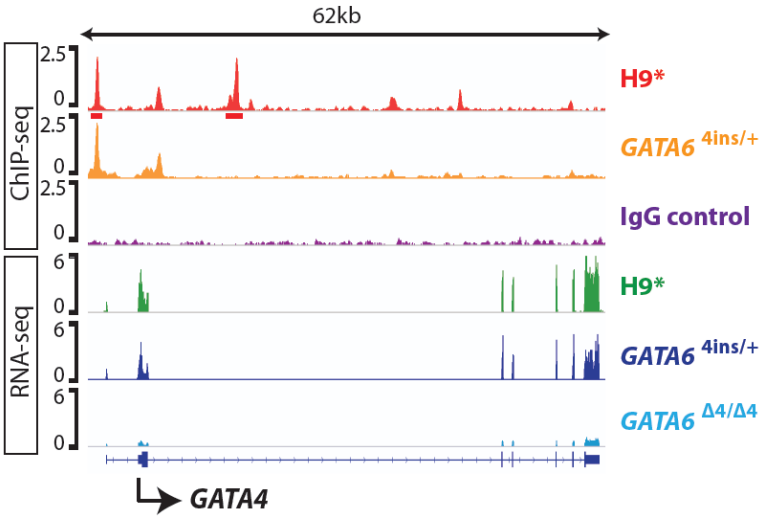
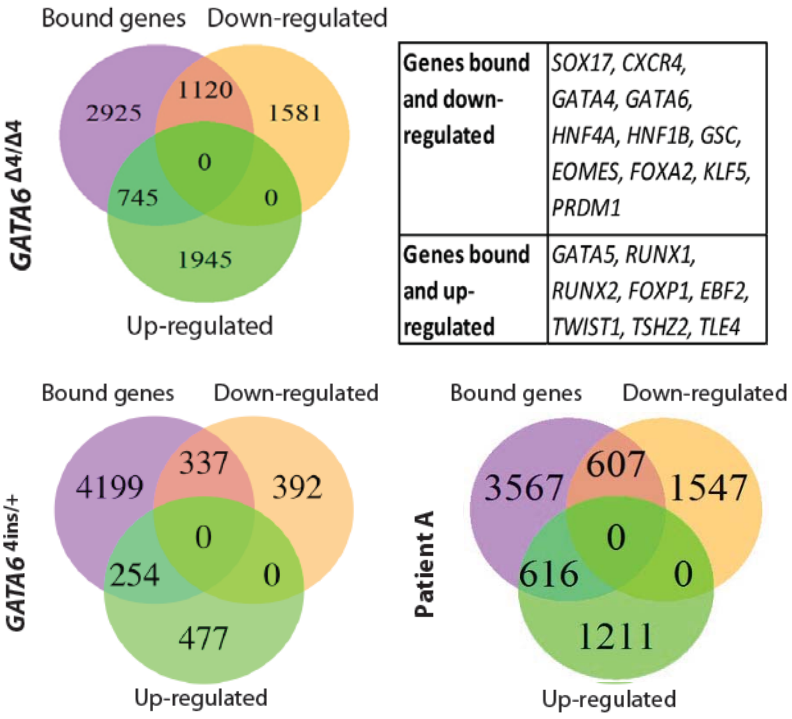


Figure 112. *GATA6* binding is enriched near the *GATA4* gene and *GATA4* is decreased in H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells. ChIP-seq binding profiles of H9* and GATA6 $^{4ins/+}$ and RNA-seq representation of *GATA4* locus showing gene expression in H9*cells and H9-derived *GATA6* $^{4ins/+}$ and *GATA6* $^{\Delta4/\Delta4}$ mutant cells at the DE stage (day 3). The input control profile (IgG control) is included for comparison. ChIP-seq binding profile is derived from merging two biological replicates.



Genes bound and down-regulated	*SOX17, CXCR4, GATA4, GATA6, HNF4A, HNF1B, GSC, EOMES, FOXA2, KLF5, PRDM1*
Genes bound and up-regulated	*GATA5, RUNX1, RUNX2, FOXP1, EBF2, TWIST1, TSHZ2, TLE4*

Figure 113. Venn diagrams indicating the overlap of GATA6-bound genes from ChIP-seq and differentially expressed genes from RNA-seq at the DE stage (day 3). Diagram shows GATA6-bound down- or up-regulated genes of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells compared to H9* cells. Key bound genes up-or down-regulated for *GATA6* $^{\Delta4/\Delta4}$ are indicated in the table.

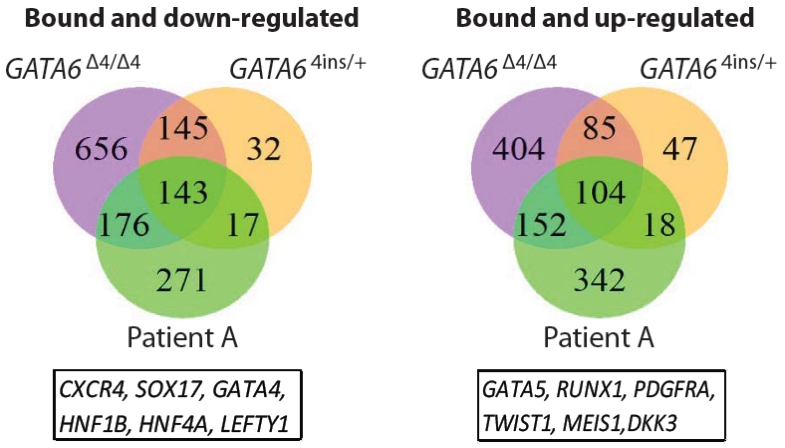
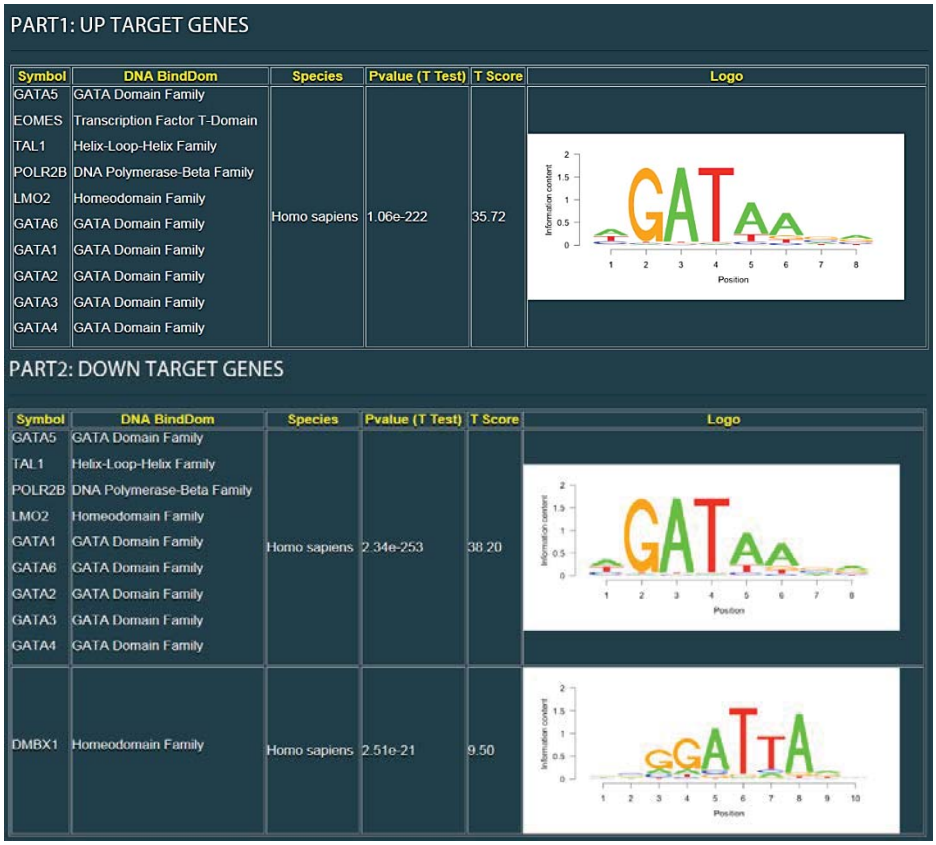


Figure 114. Venn diagrams indicating the triple overlap of *GATA6*-bound genes at the DE stage (day 3). Down- or up-regulated genes of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells are represented. Key bound genes up-or down-regulated common in all three mutants are indicated in the respective boxes.

To infer genes that are directly targeted and regulated by GATA6 and with the help of Dr. Denil, I used a software package binding and expression target analysis (BETA) to integrate the ChIP-seq dataset of H9* with differential gene expression data of *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A from RNA-seq (Wang et al., 2013). Motif analyses generated by BETA showed the GATA motif as highly enriched in both up- and down-regulated target genes, thus further validating the ChIP (Figure 115). Results from activating/repressive function prediction did not show an activating or repressive function of GATA6 in *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cell lines (Figure 116). Next, I performed gene ontology (GO) analyses on the direct targets prediction of up- and down-regulated gene lists generated by BETA data set using the DAVID tool (Huang da et al., 2009b, Huang da et al., 2009a). I found that endoderm development is consistently up-regulated in H9* cells compared to *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells, while mesoderm development is consistently up-regulated in all three mutant cell lines compared to H9* cells (Figure 117). Together, these results show a direct molecular function of *GATA6* in driving the development of the endoderm. Loss of one or two alleles of *GATA6* promotes mesodermal formation.



PART1: UP TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	1.06e-222	35.72	
EOMES	Transcription Factor T-Domain				
TAL1	Helix-Loop-Helix Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA6	GATA Domain Family				
GATA1	GATA Domain Family				
GATA2	GATA Domain Family				
GATA3	GATA Domain Family				
GATA4	GATA Domain Family				

PART2: DOWN TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	2.34e-253	38.20	
TAL1	Helix-Loop-Helix Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA1	GATA Domain Family				
GATA6	GATA Domain Family				
GATA2	GATA Domain Family				
GATA3	GATA Domain Family				
GATA4	GATA Domain Family				
DMBX1	Homeodomain Family	Homo sapiens	2.51e-21	9.50	

Figure 115. Screenshot of binding motif analysis on UP and DOWN target regions of GATA6 ChIP-seq on day 3 derived from BETA analysis. Similar motifs are grouped together, and the motif logo of the most significant factor in the group is provided in the last column. The motif symbol, DNA-binding domain and species are shown in the first three columns; the *t* score and the *P* value from the *t* test are shown in the middle two columns.

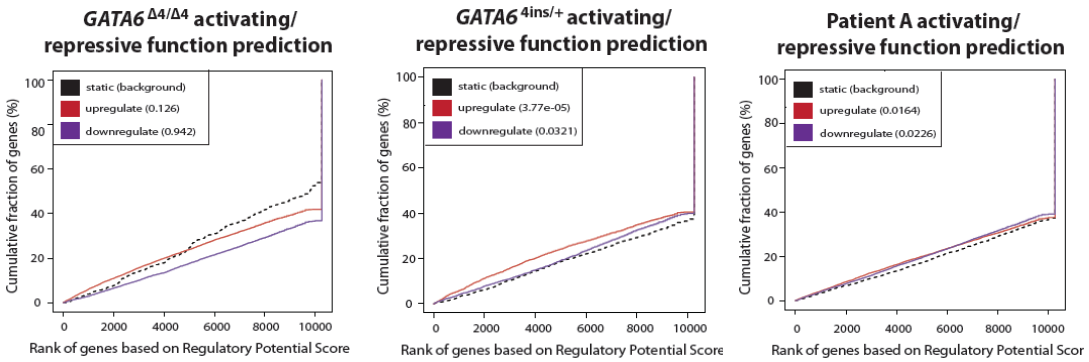


Figure 116. BETA output of activating/repressive function prediction of H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 3. The red and the purple lines represent the up-regulated and down-regulated genes, respectively. The dashed black line indicates the non-differentially expressed genes as background. Genes are cumulated by the rank on the basis of the regulatory potential score from high to low. *P*-values that represent the significance of the UP or DOWN group distributions are compared with the NON group by the Kolmogorov-Smirnov test.

Gene Ontology of genes bound and differenitally expressed in *GATA6* $^{\Delta4/\Delta4}$

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	8.74E-04	GDF3, COL4A2, HNF1B, FGF8, NODAL, EOMES, SMAD2, VTN, MMP15, HMGA2, HSBP1, COL5A2, MIXL1, DUSP5, HHEX, DUSP1, GATA6
Pancreas development	0.03667	INSM1, HNF1B, FOXA2, WFS1, SOX4, SMAD2, BAD, ISL1, TCF7L2, HHEX, BAK1, ACVR2B, MEIS2, GATA6, EIF2AK3, BMP5, BMP6
Up-regulated in *GATA6* $^{\Delta4/\Delta4}$		
Skeletal system development	2.66E-06	THRA, NDST1, HEXB, EDN1, TGFB3, BCAN, GJA5, TGFB2, MBTD1, INSIG2, PAX7, TRIM45, CHST11, RAB23, PHOSPHO1, HHIP, ANO6, ALX1, COL10A1
Mesoderm formation	0.001386	FGFR2, SIX2, SMAD3, ITGA3, WLS, PAX2, SNAI1, WNT3, DKK1, HAND1, SFRP2, ITGA8, EPB41L5, FOXC1, TLX2, ACVR1

Gene Ontology of genes bound and differentially expressed in *GATA6* $^{4ins/+}$

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	6.78E-04	GDF3, COL4A2, NOG, HNF1B, NODAL, SMAD2, MMP15, HMGA2, HSBP1, DUSP5, HHEX, DUSP1, GATA6, ITGA7, COL11A1
Up-regulated in *GATA6* $^{4ins/+}$		
Mesoderm formation	3.31E-04	FGFR2, SIX2, ITGA3, WLS, SMAD1, ITGB1, SNAI1, WNT3, DKK1, HAND1, SFRP2, ITGA8, FOXC1, TLX2

Gene Ontology of genes bound and differentially expressed in Patient A

Category	P-value	Gene symbol
Up-regulated in WT		
Endoderm development	0.026782	GDF3, NANOG, HNF1B, ONECUT1, NODAL, EOMES, MMP15, KIF16B, HSBP1, ZFP36L1, HHEX, LHX1, ITGA7, COL6A1
Up-regulated in Patient A		
Mesoderm development	0.00119	FGFR2, NOG, FGF8, TP63, ITGB3, PAX2, WNT3, OSR1, HAND1, PPP2CA, YAP1, PALB2, TLX2, TBX3, SMAD4, SIX2, SMAD3, ITGA2, ITGA3, SMAD1, WLS, POGLUT1, ACVR2B, CTDNEP1, DKK1, ITGA8, EPB41L5, PRKAR1A, ACVR1

Figure 117. Enriched gene ontology derived from BETA analysis showing developmental pathways. Tables are derived from direct target genes differentially expressed between H9* (WT), H9-derived *GATA6* $^{\Delta4/\Delta4}$, *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 3. *P*-values of the developmental pathways are indicated along with the respective genes.

Based on results from integrating the RNA-seq and ChIP-seq datasets, GATA6 seems to be a master regulator of DE development. This prompted me to investigate whether GATA6 interacts with other know DE regulators such as EOMES and SMAD2/3. With the help of Dr. Denil, I looked at the physical overlap of potential transcription factor binding sites from my *GATA6* ChIP-seq dataset on day 3 with previously published EOMES (Teo et al., 2011) and SMAD2/3 ChIP data (Brown et al., 2011) (Figure 118). The data were reprocessed as described in Chapter 2.11.2. This resulted in 16,303, 20,089 and 2,613 sites respectively. The overlap between *GATA6* and *EOMES* was 8,126/20,089 (40.5%); the overlap between *GATA6* and *SMAD2/3* was 950/2,613 (36.4%). Overlap of the three data sets is found in 858 locations. Of those sites upstream from a gene, the mean distance to the transcription start site was 831bp (median 0). Results from this experiment suggests that GATA6 co-operates with EOMES and SMAD2/3 to deploy the gene regulatory network governing human DE development.

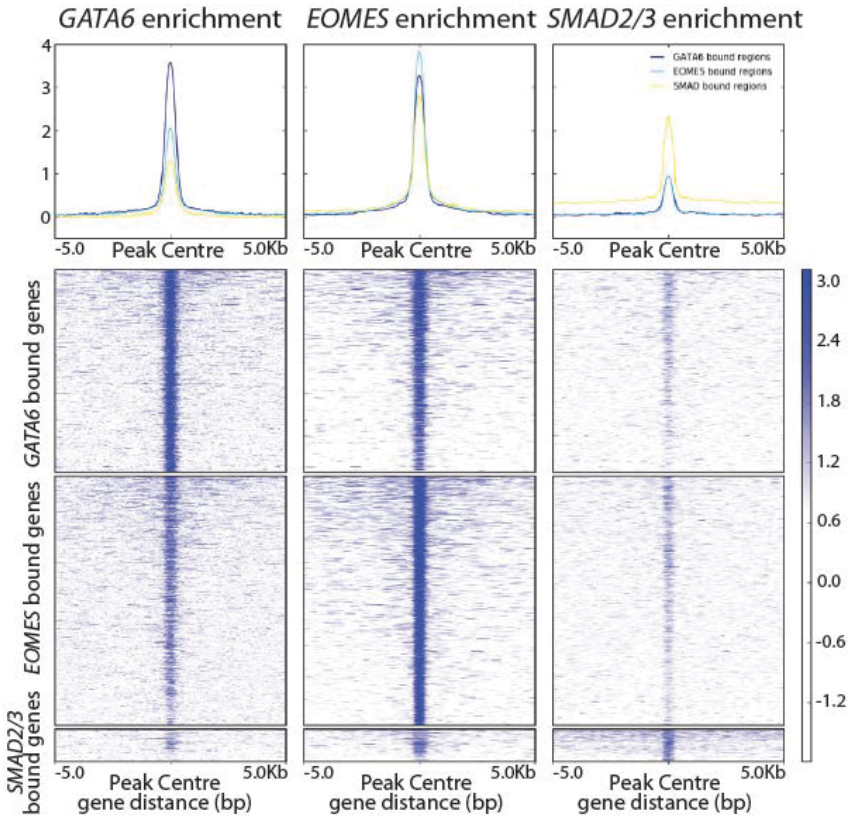


Figure 118. Density heat maps of *GATA6*-binding peak intensity at DE. Figure shows direct overlaps with known endodermal regulators including *SMAD2/3* and *EOMES* within a 10-kb window centered at the transcription start site (TSS).

3.5.2. GATA6 haploinsufficiency perturbs the pancreatic transcriptional network

Next, to investigate whether *GATA6* also plays a master regulator role at the pancreatic stage, I repeated the RNA-seq at the PE stage (day 12) for H9* cells, H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells. H9* RNA-seq data closely recapitulated a previously profiled transcriptome using the same protocol (Spearman **ρ** =0.77 for in vitro, **ρ** =0.59 for in vivo, P < 2.2e-16), which was used for the analyses of *in vitro* MPCs described in Cebola et al. (Cebola et al., 2015).

I found 1,423 differentially expressed genes between H9* and *GATA6* $^{4ins/+}$ (Table S7), and 6,093 between clones 1-3 of Patient A mutant cells (Table S8). Of the 1,423 genes, 1,230 were protein coding genes; 899 down-regulated in *GATA6* $^{4ins/+}$ cells and 331 up-regulated in *GATA6* $^{4ins/+}$ cells. Of the 6,093 genes, 4,148 were protein coding genes; 2,424 down-regulated in Patient A cells and 1,724 up-regulated in Patient A cells. In a heat map derived from RNA-seq results at the PE stage focusing on genes involved in pancreatic development, I observed that consistent with qRT-PCR (Figure 97 and Figure 101) and FACS results (Figure 96) discussed earlier, *GATA6* $^{4ins/+}$ and Patient A mutant cells displayed decreased pancreatic formation (Figure 119).

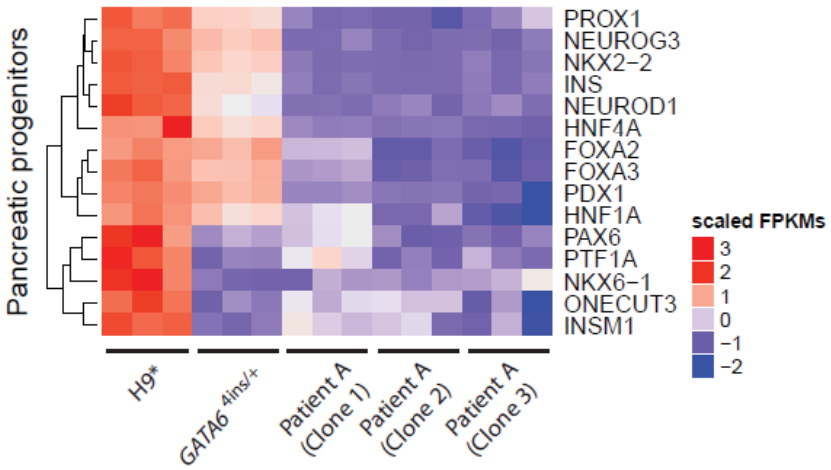


Figure 119. Heat map illustrating differential gene expression of key pancreatic markers via RNA-seq between H9*cells, H9-derived *GATA6* $^{4ins/+}$ and clones 1-3 of Patient A mutant cells at the PE stage. *n* = 3 biological replicates for each cell line.

Next, I performed ChIP on H9* cells and H9-derived *GATA6* $^{4ins/+}$ cells which were differentiated to the PE stage via the lab protocol to identify those genes bound directly to and interacting with GATA6 at the pancreatic stage. I validated the ChIP via qPCR using primers that would amplify a region on the human HNF4α P2 promoter, which acts as a GATA6 positive control. In addition, I also used a primer pair that amplifies a positive PDX1 binding location, again kindly provided to us by our collaborator Dr. Santi Rodriguez. Results from qPCR show an approximately 10-fold higher binding at the GATA6 positive region in H9* cells compared to *GATA6* $^{4ins/+}$ mutant cells and an approximately 7-fold higher binding at the PDX1 positive region in H9* cells compared to *GATA6* $^{4ins/+}$ mutant cells (Figure 120). To ensure that this was not a result of a decreased number of cells, I performed FACS on GATA6 at the PE stage and found that both H9* and *GATA6* $^{4ins/+}$ cells had similar numbers of GATA6+ cells of approximately 75% GATA6+ cells (Figure 121), indicating that the decreased DNA binding observed in *GATA6* $^{4ins/+}$ mutant cells is indeed a consequence of heterozygous loss of *GATA6*.

Unfortunately, when I proceeded to perform ChIP-seq on H9 WT cells on day 12, the ChIP-seq data retrieved was of low quality, which resulted in only 171 peaks called. This was in contrast to GATA6 ChIP-seq data published by Cebola et al., where using the same parameters, 2,060 peaks were called (Cebola et al., 2015). The ChIP experiment was repeated two more times and ChIP-seq was performed on samples derived from these two independent experiments, but the data quality remained low.

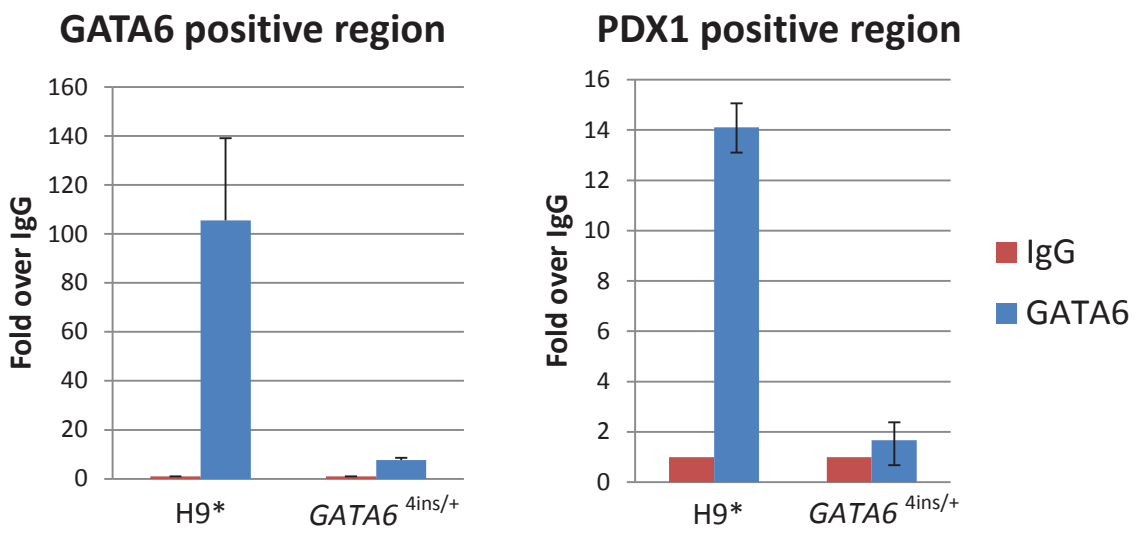


Figure 120. ChIP validation via qPCR using primers specific to a GATA6 and PDX1 positive binding region on day 12 for H9* and H9-derived *GATA6* $^{4ins/+}$ mutant cells. Values of input samples were subtracted from the values of IgG or *GATA6* samples and the graph shows fold over IgG (normalised to 1) of H9* and H9-derived *GATA6* $^{4ins/+}$ cells.

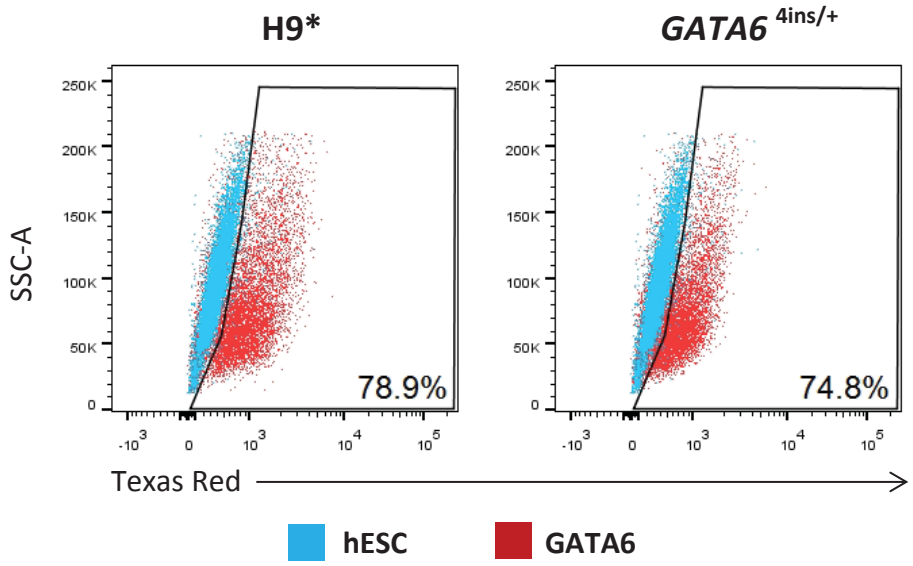


Figure 121. Number of GATA6+ cells is similar in H9* and H9-derived *GATA6* $^{4ins/+}$ cells differentiated via the lab protocol at day 12. Cells were fixed at day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least two independent experiments.

As a result, I was prompted to use FSPS13.B cell line and its respective heterozygous mutant cell lines for ChIP and ChIP-seq experiments. I hence performed ChIP on FSPS13.B* cells and FSPS13.B -derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells which were differentiated to the PE stage via the lab protocol. Validation of the ChIP via qRT-PCR showed an approximately 7-fold and 4-fold higher binding at the GATA6 and PDX1 positive binding regions respectively in FSPS13.B* cells compared to *GATA6* $^{\Delta 14/+}$ mutant cells (Figure 122). On the other hand, an approximately 2-fold higher binding at the GATA6 and PDX1 positive binding regions in FSPS13.B* cells compared to *GATA6* $^{GFP/+}$ mutant cells was observed (Figure 122). From FACS data of GATA6 in FSPS13.B*, *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells, FSPS13.B* and *GATA6* $^{GFP/+}$ mutant cells had similar levels of GATA6+ cells at the PE stage, whereas *GATA6* $^{\Delta 14/+}$ mutant cells had an approximately 30% decrease of GATA6+ cells (Figure 123). Although a decrease was seen in the number of GATA6-expressing cells in *GATA6* $^{\Delta 14/+}$ mutant cells, the decrease in binding was more so for ChIP, suggesting that the decrease in binding could be a combined consequence of fewer cells and the loss of one allele of GATA6.

Sequencing was next performed on these ChIP samples. From the ChIP-seq data at the PE stage, 2,306 peaks were called for FSPS13.B* cells, of which 1,096 of them protein coding (Table S9). For *GATA6* $^{\Delta 14/+}$ mutant cells, 543 peaks were called, of which 234 of them were protein coding (Table S10). For *GATA6* $^{GFP/+}$ mutant cells, 2,376 peaks were called, of which 1,157 of them were protein coding (Table S11). Notably, PDX1 was found to bind GATA6 in both FSPS13.B* and *GATA6* $^{GFP/+}$ mutant cells, but not in *GATA6* $^{\Delta 14/+}$ mutant cells.

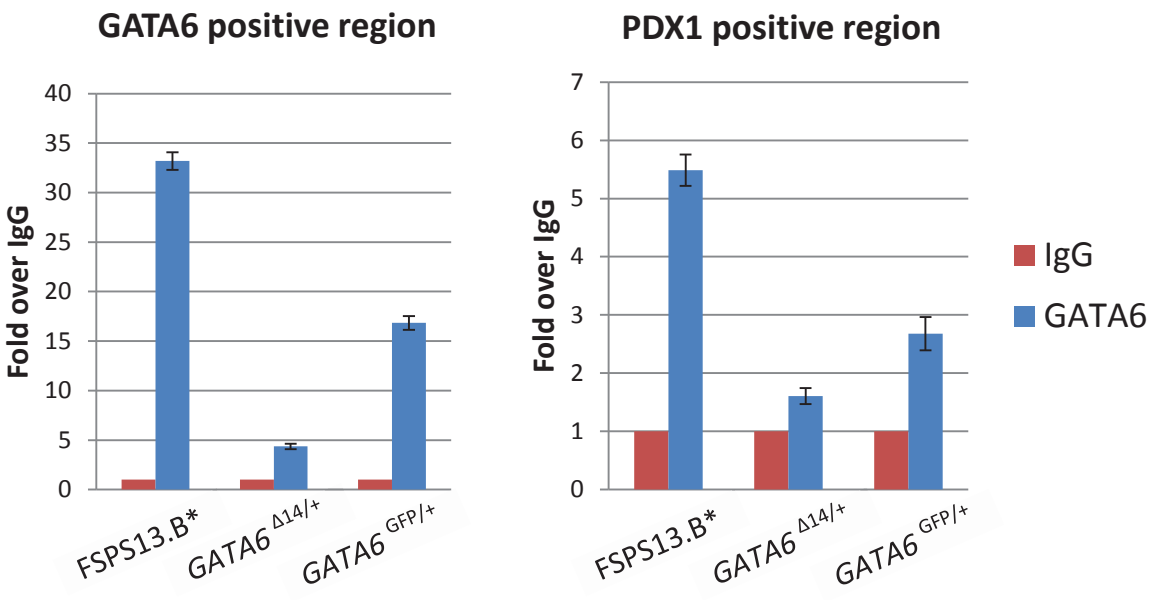


Figure 122. ChIP validation via qPCR using primers specific to a GATA6 and PDX1 positive binding region on day 12 for FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ mutant cells. Values of input samples were subtracted from the values of IgG or GATA6 samples and the graph shows fold over IgG (normalised to 1) of FSPS13.B* and FSPS13.B-derived *GATA6* $^{\Delta 14/+}$ and *GATA6* $^{GFP/+}$ cells.

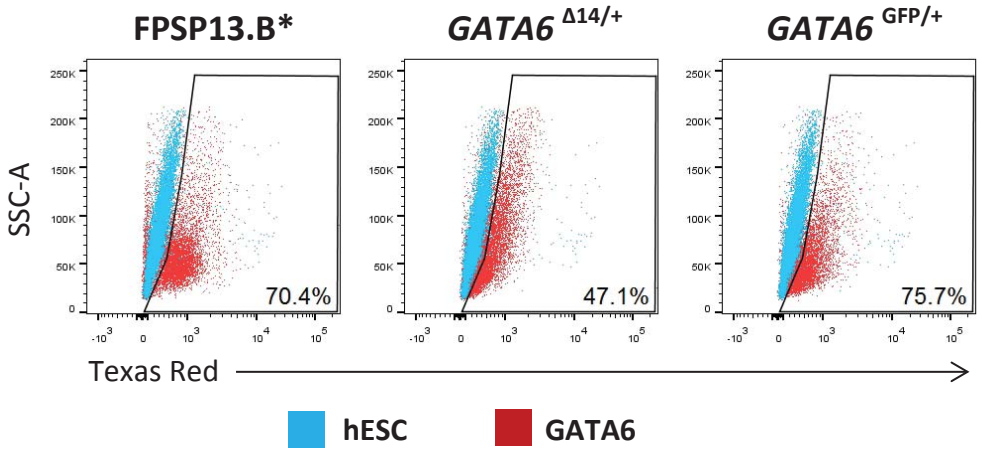
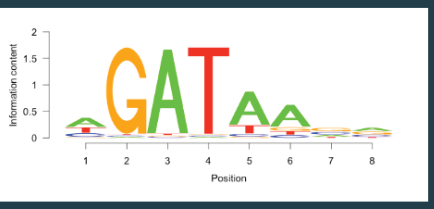


Figure 123. The number of GATA6+ cells is similar in FSPS13.B* and FSPS13.B-derived *GATA6* $^{GFP/+}$ cells but down-regulated in *GATA6* $^{\Delta 14/+}$ cells differentiated via the lab protocol on day 12. Cells were fixed on day 12 and were stained for GATA6. Data show results of one experiment that is representative of at least two independent experiments.

Next, Dr. Denil helped to repeat the BETA analyses to integrate the ChIP-seq dataset of FSPS13.B* with differential gene expression data of H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. Motif analyses generated by BETA showed the GATA motif as highly enriched in both up- and down-regulated target genes, thus further validating the ChIP (Figure 124). Results from activating/repressive function prediction did not show an activating or repressive function of GATA6 in *GATA6* $^{\Delta 14/+}$ and Patient A mutant cell lines at the PE stage (Figure 125).

Next, I performed gene ontology (GO) using DAVID on the up- and down-regulated gene lists generated by BETA and found that pancreas-related terms such as regulation of insulin secretion, endocrine and pancreas development are up-regulated in FSPS13.B* cells compared to *GATA6* $^{\Delta 14/+}$ and Patient A mutant cells (Figure 126). Interestingly, skeletal and nervous system developments, which are of mesodermal and ectodermal origins respectively, were observed to be up-regulated in both *GATA6* $^{\Delta 14/+}$ and Patient A mutant cells (Figure 126). Together, these results show a direct molecular function of *GATA6* in driving the specification of the human pancreas.

PART1: UP TARGET GENES

Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA4	GATA Domain Family	Homo sapiens	1.66e-45	16.77	
TAL1	Helix-Loop-Helix Family				
GATA6	GATA Domain Family				
POLR2B	DNA Polymerase-Beta Family				
LMO2	Homeodomain Family				
GATA3	GATA Domain Family				
GATA2	GATA Domain Family				
GATA1	GATA Domain Family				
GATA5	GATA Domain Family				
DPRX	Homeodomain Family	Homo sapiens	5.36e-04	3.31	

PART2: DOWN TARGET GENES

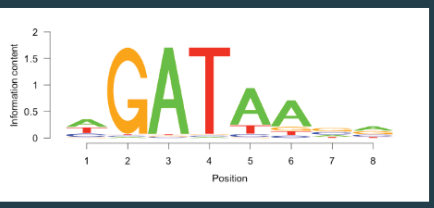
Symbol	DNA BindDom	Species	Pvalue (T Test)	T Score	Logo
GATA5	GATA Domain Family	Homo sapiens	2.02e-17	9.26	
DPRX	Homeodomain Family				
LMO2	Homeodomain Family				
POLR2B	DNA Polymerase-Beta Family				
GATA3	GATA Domain Family				
GATA6	GATA Domain Family				
GATA2	GATA Domain Family				
GATA1	GATA Domain Family				
GATA4	GATA Domain Family				
RHOXF1	Homeodomain Family	Homo sapiens	1.58e-04	3.67	

Figure 124. Screenshot of binding motif analysis on UP and DOWN target regions of GATA6 ChIP-seq on day 12 derived from BETA analysis. Similar motifs are grouped together, and the motif logo of the most significant factor in the group is provided in the last column. The motif symbol, DNA-binding domain and species are shown in the first three columns; the *t* score and the *P* value from the *t* test are shown in the middle two columns.

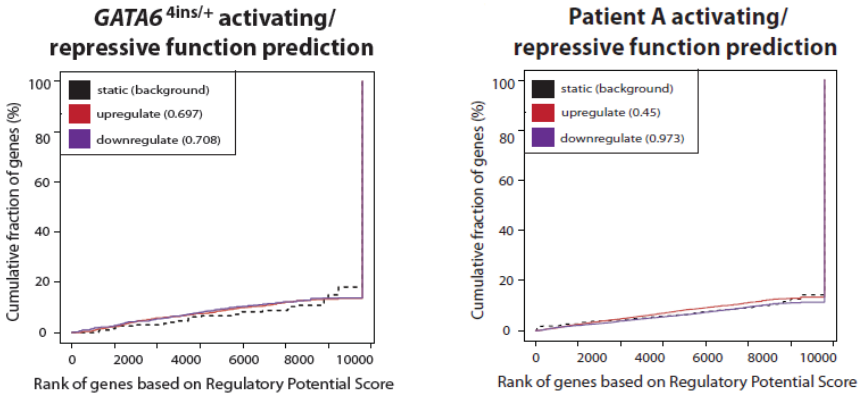


Figure 125. BETA output of activating/repressive function prediction of H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. The red and the purple lines represent the up-regulated and down-regulated genes, respectively. The dashed black line indicates the non-differentially expressed genes as background. Genes are cumulated by the rank on the basis of the regulatory potential score from high to low. *P*-values that represent the significance of the UP or DOWN group distributions are compared with the NON group by the Kolmogorov-Smirnov test.

Gene Ontology of genes bound and differenitally expressed in *GATA6* $^{4ins/+}$

Category	P-value	Gene symbol
Up-regulated in WT		
Regulation of insulin secretion	0.05181	STX1A, HNF4A, BRSK2, PDX1, C2CD2L
Regulation of heart contraction	5.70E-04	TNNT2, MYL4, SCN1B, ATP2B4, ADM, MYL3, NPPA, KCNK3, CACNA1B
Up-regulated in *GATA6* $^{4ins/+}$		
Skeletal system development	1.43E-09	LUM, TBX1, GLI2, FOXP1, HOXB4, HOXB1, DHRS3, OSR2, HOXB7, HOXB8, HOXB5, HOXC4, HOXB6, HOXC5, HOXB9, RARB, IGFBP4
Nervous system development	2.04E-06	LUM, GLI2, EPHB2, OVOL2, DYNLL2, HLX, PAFAH1B3, RANBP1, H2AFX, RARB, NR2F1, GBA, CENPF, ARID1A, TBX1, SHANK1, SHANK2, FOXP1, ATF5, HOXB1, DHFR, HOXB8, CLIC5, MNX1, EFNA5, KDM4A, LRP2, CALM1

Gene Ontology of genes bound and differenitally expressed in Patient A

Category	P-value	Gene symbol
Up-regulated in WT		
Endocrine system development	0.001505	HES1, GATA2, HNF1B, FGF8, APOA1, FOXA2, SALL1, MNX1, PAX6, DLL1, TBX1, PDX1, INSR
Endocrine pancreas development	0.003831	HES1, HNF1B, FOXA2, MNX1, PAX6, DLL1, PDX1
Up-regulated in Patient A		
Skeletal system development	4.45E-05	RBP4, HOXA13, LUM, PTH1R, SLC38A10, GLI2, FUZ, HOXC6, HOXC8, HOXC9, OSR2, CD44, HOXA5, HOXC4, HOXA6, RARB, GHR, HAPLN1
Nervous system development	3.24E-06	STIL, CCDC85C, STAR, PPP2R5D, HOXC8, PACSIN1, SMARCD3, GATA3, SMARCD1, S1PR5, DYNC2H1, H2AFX, ROBO2, RARB, MATN2, SCRT2

Figure 126. Enriched GO derived from BETA analysis showing developmental pathways. Tables are derived from direct target genes differentially expressed between H9* (WT), H9-derived *GATA6* $^{4ins/+}$ and Patient A mutant cells on day 12. *P*-values of the developmental pathways are indicated along with the respective genes.