The role of transcription factor *GATA6* in the development of the human pancreas

Crystal Ying Chia

Homerton College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

July 2017

Acknowledgements

"Follow your dreams", my parents always reminded me.

Growing up, I had never imagined I would do a PhD, much less in Cambridge. Being enrolled at the University of Cambridge to read a doctoral degree is a dream come true for me. The path to getting to where I am today is attributed to a mixture of hard work and luck in meeting the right people who not only opened doors and made this once in a lifetime opportunity possible, but also believed in me and supported me throughout this journey. There are many people for whom I have heartfelt gratitude, but sadly, it is only possible to particularly thank some here.

To Professor Ludovic Vallier, my Cambridge supervisor, who has been instrumental in driving the progress of my PhD project from start to finish. Thank you for accepting me as your student and for providing me with this valuable opportunity. You have been a great mentor who has taught me so much. Thank you for all the guidance, understanding and support you have shown me over the years, and for moulding my development not only in stem cell biology, but as a scientist. You have provided me with such an all-rounded PhD experience, from teaching me important set of skills for my scientific career, to exploring Europe and USA on the annual lab retreats and conferences.

To Dr Norris Ray Dunn, my Singapore supervisor, who has moved mountains for me. Thank you for accepting me as your student and for providing me with this valuable opportunity. I am truly lucky and blessed to have you as a mentor, and you have been nothing but supportive, encouraging and caring. Thank you for imparting valuable skills to me not only in research, but also in writing and presentations. Thank you for making my PhD experience such an amazing and wholesome one and for teaching me so much that is not only important for my scientific career, but also in life.

To my family, thank you for all the encouragement, support, care and concern. To my parents, for giving me the freedom and opportunities to pursue all my desires in life. To my brother and sister-in-law, for always being there no matter the distance.

To my friends and colleagues in Cambridge, Li Meng, Kasia, Sapna, Hiroko- you have been there for me since day one. Thank you for being my pillars of support through the good and the bad times, and for making my life in Cambridge so wonderful.

To my friends in Singapore and Malaysia who have not forgotten me despite being away for two years, Atee, Jieying, Daphne, Xuefang and Anita. Thank you for your friendship over the many, many years and for always pushing me to excel in the things I do.

To CJ, thanks for all the motivation, lessons of patience and for being there for me during the final push of my PhD.

To A*STAR, for providing me with the opportunity to pursue a PhD degree by funding my course of study.

To members of the Vallier and NRD lab who have created a fun and lively place to work in. Thank you for all the help, support and friendship.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

Crystal Ying Chia

July 2017

Statement of Length

This dissertation does not exceed the word limit of 60,000 words excluding figures, photographs, tables, appendices and bibliography for the Biology Degree Committee.

List of Publications and Presentations

Work from this thesis contributed to the following publications and presentations:

Academic publication:

Chia, C.Y., Madrigal, P., Denil, S.L.I.J., El-Khairi, R., Chhatriwala, M., Hattersley, A.T., N. Dunn, N.R., and Vallier, L.. GATA6 cooperates with EOMES/SMAD2/3 to deploy the gene regulatory network governing human definitive endoderm and pancreas formation. To be submitted (2017).

Poster presentations:

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Keystone Symposia Conference: Emerging Concepts and Targets in Islet Biology held from April 6 – 11, 2014, Colorado, USA.

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Company of Biologist Workshops: From Stem Cells to Human Development held from September 21 – 24, 2014, Wotton House, Surrey, UK.

Chia, C.Y., Chhatriwala, M., Dunn, N.R., and Vallier, L.. (2014). The role of transcription factor GATA6 in human pancreas and liver. Poster presented at the Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute: SCI PhD Day held on July 22, 2014, Cambridge, UK.

Summary

While there has been an opulence of data and studies surrounding the study of the developing pancreas in mammals and other vertebrates, the focus has largely been in mice. The paucity of research in the development of the human pancreas has led to diminished knowledge in the area, compared to other species. Recent discoveries provide growing evidence for discrepancies between mouse and human pancreatic development and diseases and highlight the fact that developmental studies of the pancreas in humans are imperative. The need to develop therapies for diabetes, a growing and one of the leading health problems worldwide, further compels more exploration in this area to deepen our understanding in the different aspects of diabetes in humans and its underlying causes.

Research involving modelling human diseases *in vitro* enables the investigation of the cellular and molecular mechanisms underlying these diseases as well as the development of therapies for treating them. The availability of hPSCs brings with it the advantage of overcoming the limitations of animal models for certain disorders such as pancreatic agenesis, the focus of my project. The use of site-specific nucleases such as TALENs for such a purpose represents a paradigm shift in disease modelling, where TALENs are capable of directly correcting disease-causing mutations, therefore permanently eliminating the symptoms with precise genome modifications. Alternatively, TALENs can also be used to inactivate specific genes by inducing site-specific mutations.

Using these tools, I found that *GATA6* is required for the formation of the definitive endoderm and pancreas in humans; hPSCs harbouring homozygous *GATA6* mutations fail to form the definitive endoderm, and consequently the pancreas, whereas hPSCs harbouring heterozygous *GATA6* mutations exhibited impairment in definitive endoderm development, although it remains unclear if this is a protocol-dependent defect. At the pancreatic stage, heterozygous *GATA6* mutations consistently compromised pancreas formation regardless of protocol used. I also found that *GATA6* transcriptionally activates the development of the definitive

endoderm and pancreatic endoderm, and possibly represses the development of mesoderm. Furthermore, I also established that *GATA6* directly interacts with key definitive endoderm markers *CXCR4* and *SOX17*, and pancreatic marker *PDX1*.

Taken together, the work herein demonstrates the successful use of hPSCs coupled with the TALEN genome editing technology as a unique *in vitro* system for disease modelling. These findings also establish two developmental windows, the DE and pancreatic progenitor stages, where *GATA6* haploinsufficiency can result in the impairment of pancreatic development leading to pancreatic hypoplasia observed in human *GATA6* heterozygous patients. Lastly, my work also provides the molecular mechanism by which *GATA6* regulates pancreatic development.

Overall, this study provided new insights in the role of *GATA6* during development of the human pancreas. These results will be important in developing new methods of differentiation for hPSCs and understanding the interconnection between early organogenesis and late onset of diabetes.

Table of Contents

List of Abbreviations

ABCC8	ATP-binding cassette, sub-family C, member 8
ADV-BSA	Advanced DMEM/F-12 containing bovine serum albumin
AIP	Anterior intestinal portal
bFGF	Basic fibroblast growth factor
BNZ	6-Bnz-cAMP sodium salt
DE	Definitive endoderm
CDM	Chemically defined media
CDM-PVA	Chemically defined medium-poly vinyl alcohol
CEL	Carboxyl ester lipase
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regulatory interspaced short palindromic repeats
CS	Carnegie stages
CXCR4	Chemokine (C-X-C motif) receptor 4
DE	Definitive endoderm
Dpc	Days post-conception
DSB	Double-stranded break
E8	Essential 8
EIF2AK3	Eukaryotic translation initiation factor 2 alpha kinase 3
ELISA	Enzyme linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
EP	Endocrine progenitors
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FLASH	Fast Ligation-based Automatable Solid-phase High-throughput

FOX	Forkhead box
GATA	GATA binding protein
GCK	Glucokinase
GLIS3	GLIS family zinc finger 3
GO	Gene ontology
GSC	Goosecoid
GSIS	Glucose-stimulated insulin secretion
GRN	Gene regulatory network
hESC	Human embryonic stem cell
HGF	Hepatocyte growth factor
hiPSC	Human induced pluripotent stem cell
HIPSCI	Human induced pluripotent stem cell initiative
hPSC	Human pluripotent stem cell
HMG	High mobility group
HNF	Hepatocyte nuclear factor
HR	Homologous recombination
HRP	Horseradish peroxidase
ICA	Iterative Capped Assembly
ICC	Immunocytochemistry
ICM	Inner cell mass
INS	Insulin
iPSC	Induced pluripotent stem cells
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11
LIF	Leukaemia inhibitory factor
MEF	Mouse embryonic fibroblasts
MODY	Maturity onset diabetes of the young

NDM	Neonatal diabetes mellitus
NEUROD1	Neuronal differentiation 1
NGN	Neurogenin
NHEJ	Non-homologous end-joining
NKX	Nirenberg and Kim homeobox factor
ORF	Open reading frame
OSM	Oncostatin M
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PDX1	Pancreatic and duodenal homeobox factor 1
PE	Pancreatic endoderm
PFA	Paraformaldehyde
PMSF	Phenylmethylsulfonylfluoride
PNDM	Permanent neonatal diabetes mellitus
PPP	Partial protein product
PSC	Pluripotent stem cell
PTF1A	Pancreas transcription factor 1A
qRT-PCR	Quantitative real time polymerase chain reaction
RA	Retinoic acid
RFX6	Regulatory factor X6
RVD	Repeat variable di-residues
SCT	Stem Cell Technologies
SHH	Sonic hedgehog
SOX	SRY (sex determining region Y)-box
T1D	Type 1 diabetes
T2D	Type 2 diabetes

TALEN	Transcription activator-like effector nuclease
TF	Transcription factor
TGF-β	Transforming growth factor-β
TNDM	Transient neonatal diabetes mellitus
TSS	Transcription start site
Wp	Well plate
Wpc	Weeks post-conception
ZFN	Zinc finger nuclease

List of Figures

List of Tables