Generating C-peptide Humanized Mice

by

Seana Nelson

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Institute of Medical Sciences
University of Toronto

© Copyright by Seana Nelson 2013
Generating Humanized C-peptide Mice

Seana Nelson

Master of Science

Institute of Medical Sciences
University of Toronto

2013

Abstract

Type 1 diabetes is primarily caused by the loss of insulin producing β-cells. Future therapies based on transplantation of in vitro generated β-cells hold great promise for providing a cure. The development of such treatments heavily relies on the use of mouse models. However, insulin expressed from transplanted, mouse derived β-cells cannot be differentiated from the host in allograft studies. ELISA can distinguish mouse and human C-peptide, a biomarker of insulin secretion. This project aims to create a humanized C-peptide mouse line that could be used to distinguish insulin production in allograft transplantation studies, or differentiate expression from the two mouse insulin genes. Humanized C-peptide insulin 1 chimeras have been generated and an insulin 2 founder has been identified. Following germline transmission, the lines will be crossed to obtain double homozygous humanized C-peptide mice. These animals will be an invaluable resource in allogeneic β-cell transplants and in monitoring insulin allele expression.
Acknowledgments

I would like to thank Dr. Andras Nagy for providing the opportunity to explore this exciting career option, for the extensive guidance and support. The environment that you have created in the lab much like a family, and it has made these past two years quite enjoyable.

Within the Nagy lab there are many people to be thankful for. To Kristina Nagy, thank you for starting this project and for your continued support throughout the Master’s project, including critical reading of this thesis. Peter Tonge for assisting with my Master’s project proposal and design, project guidance, and for acquiring $frt$-$F3$-$PGK/EM7$-$Kanamycin/Neomycin-frt$-$F3$ resistance cassette. To Claudio Monnetti for guidance, mentorship and technical support. Chen He for critical reading of my thesis, continued support with respect to gene targeting, screening, analyzing mice colonies during my leave of absence. Mira Puri for critical reading of the discussion of this thesis. Tracey Clatworthy: for keeping us organized and on task. Thank you to many members of the Nagy lab (2010-2013) for their support, helpful suggestions, and criticisms: Iacovos Michael, Samer Hussein, Balasz Varga, Maryam Faiz, Sabiha Hacibekiroglu, Qin Liang, Joe Son, Hoon-Ki Sung, Andrew Corso, Puzheng Zhang, Masha Mileikovsky, Marina Gertsenstein, Malgosia Kownacka.

My committee has been a tremendous resource and support throughout this process. They have introduced me to the diabetes community and individuals with whom I could contact for discussions and methodological/technical support. Thank you Dr. Drucker for discussions of my project and guidance in other life goals. Thank you Dr. Wheeler for allowing me to attend lab meetings and discuss my project with you and your group.

Thank you to the Rogers lab for their continued support throughout the last three years, including help with animal handling and $in vitro$ studies that laid the foundation for the “Future Studies”. To Dr. Rogers, thank you for the many helpful discussions and pieces of advice.
Acknowledgement of work completed by Dr. Chen He

Dr. Chen He used targeting vectors generated by the candidate to perform further gene targeting in mouse embryonic stem cells. He screened targeted clones by Southern blot using screening strategies that were published (Duvillié et al. 1997) (Ins1), that he generated (Ins1) or developed by the candidate (Ins2).
**Table of Contents**

Acknowledgments.............................................................................................................. iii

Table of Contents............................................................................................................... v

List of Tables........................................................................................................................ viii

List of Figures ......................................................................................................................... ix

List of Abbreviations ............................................................................................................. x

List of Appendices ................................................................................................................. xv

1) Literature Review: Introduction ......................................................................................... 1
   a) Diabetes is caused by insufficient insulin production and/or signaling ...................... 2
   b) Type 1 Diabetes ............................................................................................................... 3
   c) Stem cell therapy for T1D ............................................................................................. 4
   d) Insulin gene, expression and function ......................................................................... 5
      i) Evolutionary constraints of the primary and secondary structure of the insulin gene... 5
      ii) The C-peptide .......................................................................................................... 7
      iii) Regulation of insulin mRNA transcription ............................................................... 11
      iv) Post transcriptional modifications of insulin mRNA ................................................. 15
      v) Insulin biosynthesis .................................................................................................. 17
      vi) Insulin secretion ....................................................................................................... 18
      vii) Insulin function ...................................................................................................... 19
   e) Mouse Models to study T1D ......................................................................................... 21
      i) β-cell ablation models ............................................................................................... 21
      ii) Immune mediated models of diabetes ....................................................................... 23
   f) Differences between the mouse and human insulin structure .................................. 23
   g) Generating a targeted mouse mutation: the toolbox .................................................. 25
      i) Choice of genetic background for the generation of genetically engineered mouse models 25
      ii) Recombineering ...................................................................................................... 28
      iii) Target vector design .............................................................................................. 29
      iv) Chimeras ............................................................................................................... 30

2) Rationale, Hypothesis and Aims ....................................................................................... 31

3) Materials, Methods and Experimental Procedures ......................................................... 34
a) **Materials & Methods** ........................................................................................................... 35
   i) Transforming chemically competent cells ........................................................................... 35
   ii) Colony screening (plasmids) ............................................................................................... 35
   iii) Sequencing and analysis of results ....................................................................................... 36
   iv) Electroporating bacteria with pSC101-BAD-γβα Red/ET expression plasmid ...................... 36
   v) Recombineering ................................................................................................................... 38
   vi) Screening colonies (BACs and targeting vectors) ................................................................. 38
   vii) Targeting vector preparation ............................................................................................. 39
   viii) ESC thawing, culture, and freezing .................................................................................... 39
   ix) Electroporation of ESCs ........................................................................................................ 41
   x) ESC selection ....................................................................................................................... 41
   xi) ESC colony picking .............................................................................................................. 42
   xii) Passaging and freezing of ESC on 96-well plates ................................................................. 42
   xiii) Screening ESC clones by PCR .......................................................................................... 43
   xiv) Screening clones by Southern Blots .................................................................................... 44
b) **Experimental Procedures: Targeting of mouse Insulin 1 and 2 genes** .............................. 45
   i) Recombineering vector construction .................................................................................... 45
   ii) Targeting vector construction .............................................................................................. 47
   iii) Gene targeting in ESC .......................................................................................................... 49
   iv) Chimera and transgenic animal creation ............................................................................. 49

4) **Results** .................................................................................................................................. 52
   a) **Overview** ............................................................................................................................ 52
   b) **Generating mouse Insulin1 and 2 gene-targeting vectors** .............................................. 53
      i) Design of the humanized C-peptide recombinereing vectors ........................................... 53
      ii) Construction of the recombinereing vectors ................................................................... 53
      iii) Construction of gene-targeting vectors by recombinereing .......................................... 62
   c) **Gene targeting in ESC** ....................................................................................................... 71
   d) **Generating transgenic mice** ............................................................................................... 77

5) **Discussion** ............................................................................................................................. 80
   a) Tracing the expression of insulin with new mouse models by knocking in a humanized C-peptide into the mouse Insulin 1 and 2 genes ......................................................... 80
   b) C-peptide was chosen as a biomarker ................................................................................... 81
   c) A genetic approach to express human C-peptide as an endogenous protein in mice ...82
i) Design and generation of targeting vectors ................................................................. 83
d) Gene targeting efficiencies at the mouse Insulin 1 and Insulin 2 loci ......................... 90
e) Generating chimeras ................................................................................................. 93
f) Future directions to characterize humanized C-peptide Ins1 and Ins2 mouse models. 94
i) Breeding schematic to generate homozygous humanized C-peptide mice .............. 94
ii) Monitoring health and metabolic parameters ........................................................ 96
iii) Investigating relative insulin allele expression ...................................................... 97
iv) Investigating the functionality of in vitro modified insulin producing cells ........ 97
g) Limitations of the humanized C-peptide model .................................................... 98

6) Summary and Conclusions ..................................................................................... 100

References .................................................................................................................... 101

Appendices .................................................................................................................. 115

1 Southern Analysis Methods, Performed by Dr. Chen He ........................................ 115
List of Tables

TABLE 1. FBS-ES-DMEM (ES) MEDIA. ................................................................. 40
TABLE 2. PRIMERS USED IN ORDER OF APPEARANCE........................................ 50
TABLE 3. EFFICIENCY OF TARGETING VECTORS TVhCP1 AND TVhCP2. ..................... 76
TABLE 4. INSULIN 1 HUMAN C-PEPTIDE CHIMERAS GENERATED FROM CLONES 2E AND 10F. .................. 77
TABLE 5. INSULIN 2 HUMAN C-PEPTIDE CHIMERAS GENERATED FROM CLONES 7C AND 1A. ......................... 78
List of Figures

Figure 1. Human INS, mouse Ins2 and Ins1 genetic structures.................................................................9
Figure 2. COLBALT protein alignment of the insulin peptide sequence for human and mouse insulin1 and 2... 10
Figure 3. Schematic to generate a mouse line..........................................................................................26
Figure 4. Ins1 humanized C-peptide recombineering vector strategy......................................................55
Figure 5. Ins2 humanized C-peptide recombineering vector strategy......................................................56
Figure 6. Constructs used to generate recombineering vectors.................................................................57
Figure 7. Construct digest to insert the resistance cassette K/NR into pBMH_hCp1 or pBMH_hCp2.............58
Figure 8. Human C-peptide recombineering cassettes..............................................................................59
Figure 9 Human C-peptide recombineering cassettes were cloned correctly...........................................60
Figure 10. Linearized humanized C-peptide recombineering cassettes....................................................61
Figure 11. Insulin 1 humanized C-peptide recombineering strategy.......................................................64
Figure 12. The C-peptide of Bins1 is humanized......................................................................................65
Figure 13. Targeting vector retrieval strategy............................................................................................66
Figure 14. Ins1 humanized C-peptide targeting vector was retrieved with 6.4Kb 5’ and 5.4Kb homology arms. 67
Figure 15. Ins2 humanized C-peptide recombineering strategy...............................................................68
Figure 16. Intermediate vector BAIIns2....................................................................................................69
Figure 17. Insulin 2 humanized C-peptide targeting vector was retrieved with 6.4Kb 5’ and 5.4Kb 3’ homology arms.........................................................................................................................70
Figure 18. Gene targeting of the mouse Ins1 gene to humanize the C-peptide.........................................72
Figure 19. Targeting strategy to humanize the mouse Ins2.....................................................................73
Figure 20. Embedded PCR (K/NRSR3, Ins2GenoR1; K/NRSR2, Ins2GenoR2) distinguished two (1A and 9F) Ins2 targeted clones (5.7Kb) out of 104 G418 resistant clones.........................................................74
Figure 21. Mouse Ins2 human C-peptide ESC by Southern analysis reveals three correctly targeted clones.. 75
Figure 22. Ins2 human C-peptide F1 litter derived from clone 1A, screened by PCR and Southern.............79
Figure 23. Incomplete humanizing of the C-peptide with the hCp2 recombineering cassette.....................89
Figure 24. Primary PCR of humanized C-peptide Ins1 did not conclusively indicate any correctly targeted clones............................................................................................................................................92
Figure 25. Sample breeding schematic. A hCp1-K/N will be bred to FLP-O to excise the resistance cassette.. 95
## List of Abbreviations

<table>
<thead>
<tr>
<th>Description</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine triphosphate</td>
<td>ATP</td>
</tr>
<tr>
<td>Amino Acids</td>
<td>AA</td>
</tr>
<tr>
<td>Ampicillin recombineering vector with ins2 homology arms</td>
<td>Amp2</td>
</tr>
<tr>
<td>Ampicillin resistance</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Arabinose inducible</td>
<td>Arab&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
</tr>
<tr>
<td>ATP sensitive potassium channel</td>
<td>K&lt;sub&gt;ATP&lt;/sub&gt;</td>
</tr>
<tr>
<td>BAC containing mouse insulin 1</td>
<td>Bins1</td>
</tr>
<tr>
<td>BAC containing mouse insulin 2</td>
<td>Bins2</td>
</tr>
<tr>
<td>BAC with Amp&lt;sup&gt;R&lt;/sup&gt; cassette integrated into Ins1 or 2</td>
<td>BAIns1 or 2</td>
</tr>
<tr>
<td>BAC with hCp incorporated</td>
<td>BhC</td>
</tr>
<tr>
<td>BAC with human C-peptide allele</td>
<td>BhCp</td>
</tr>
<tr>
<td>BAC with pRET transfected</td>
<td>B+pRET</td>
</tr>
<tr>
<td>Bacterial Artificial Chromosome</td>
<td>BAC</td>
</tr>
<tr>
<td>C57BL/6N-Tg(CAG-Flpo)1Afst/Mmucd</td>
<td>FlpO</td>
</tr>
<tr>
<td>C57BL/6NTac-C2 ES cells</td>
<td>C2</td>
</tr>
<tr>
<td>cAMP response element</td>
<td>CRE</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>cAMP-dependent protein kinase A</td>
<td>PKA</td>
</tr>
<tr>
<td>Carboxypeptidase H</td>
<td>CPH</td>
</tr>
<tr>
<td>Coding DNA</td>
<td>cDNA</td>
</tr>
<tr>
<td>CRE binding proteins</td>
<td>CREB</td>
</tr>
<tr>
<td>CRE modulating proteins</td>
<td>CREM</td>
</tr>
<tr>
<td>Cyclic Adenosine Monophosphate</td>
<td>CAMP</td>
</tr>
<tr>
<td>Deoxyribonucleic acid</td>
<td>DNA</td>
</tr>
<tr>
<td>Deoxyribonucleotide</td>
<td>dNTP</td>
</tr>
<tr>
<td>Diphtheria Toxin-A</td>
<td>DTA</td>
</tr>
<tr>
<td>Double distilled H₂O</td>
<td>ddH₂O</td>
</tr>
<tr>
<td>Double stranded DNA</td>
<td>dsDNA</td>
</tr>
<tr>
<td>Electroporation Buffer</td>
<td>E.P.B</td>
</tr>
<tr>
<td>Embryonic Stem cells</td>
<td>ES cells</td>
</tr>
<tr>
<td>Endoplasmic Reticulum</td>
<td>ER</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>E. Coli</td>
</tr>
<tr>
<td>European Conditional Mouse Mutagenesis Program</td>
<td>EUCOMM</td>
</tr>
<tr>
<td>FBS-ES-DMEM media</td>
<td>ES media</td>
</tr>
<tr>
<td><em>frt-F3-PGK/EM7-Kanamycin/Neomycin-frt-F3 resistance cassette</em></td>
<td>K/N&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gap repair plasmid</td>
<td>pGRDT</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Gap repair plasmid with insulin 1 homology arms</td>
<td>pGRDTIns1</td>
</tr>
<tr>
<td>Gap repair plasmid with insulin 2 homology arms</td>
<td>pGRDTIns2</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>gDNA</td>
</tr>
<tr>
<td>Genticin</td>
<td>G418</td>
</tr>
<tr>
<td>Germline Transmitter</td>
<td>GLT</td>
</tr>
<tr>
<td>Glucagon-like peptide-1</td>
<td>GLP-1</td>
</tr>
<tr>
<td>Human C-peptide 1 and 2 recombineering vectors</td>
<td>hCp1 and hCp2</td>
</tr>
<tr>
<td>Human insulin gene</td>
<td>INS</td>
</tr>
<tr>
<td>Humanized C-peptide targeting vector for insulin 1</td>
<td>TVhCp1</td>
</tr>
<tr>
<td>Humanized C-peptide targeting vector for insulin 2</td>
<td>TVhCp2</td>
</tr>
<tr>
<td>Imprinting control region mice</td>
<td>ICR</td>
</tr>
<tr>
<td>Insulin like growth factor</td>
<td>IGF</td>
</tr>
<tr>
<td>International Knockout Mouse Consortium</td>
<td>IKMC</td>
</tr>
<tr>
<td>Kanamycin resistance</td>
<td>KanaR</td>
</tr>
<tr>
<td>Kilo base</td>
<td>Kb</td>
</tr>
<tr>
<td>Knockout Mouse Project</td>
<td>KOMP</td>
</tr>
<tr>
<td>Liquid Broth</td>
<td>LB</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
</tr>
<tr>
<td>Mesenchymal embryonic fibroblasts</td>
<td>MEFs</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Mouse embryonic stem cell</td>
<td>mESC</td>
</tr>
<tr>
<td>Mouse insulin gene</td>
<td>Ins</td>
</tr>
<tr>
<td>Neomycin resistance cassette</td>
<td>Neo&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Non-obese diabetic</td>
<td>NOD</td>
</tr>
<tr>
<td>North American Conditional Mouse Mutagenesis project</td>
<td>NorCOMM</td>
</tr>
<tr>
<td>Pancreatic duodenal homeobox-1</td>
<td>PDX-1</td>
</tr>
<tr>
<td>Passage#</td>
<td>p#</td>
</tr>
<tr>
<td>PGK-promoted Neomycin resistance</td>
<td>Neo&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Phosphate Buffer Solution</td>
<td>PBS</td>
</tr>
<tr>
<td>Polymerase Chain Reaction</td>
<td>PCR</td>
</tr>
<tr>
<td>polyoma enhancer/herpes simplex virus thymidine kinase promoter</td>
<td>MC1</td>
</tr>
<tr>
<td>polypyrimidine tract-binding proteins</td>
<td>PTB</td>
</tr>
<tr>
<td>Protein Convertase -1</td>
<td>PC-1</td>
</tr>
<tr>
<td>Protein Convertase -2</td>
<td>PC-2</td>
</tr>
<tr>
<td>pSC101-BAD-γβα Red/ET</td>
<td>pRET</td>
</tr>
<tr>
<td>Ribonucleic acid</td>
<td>RNA</td>
</tr>
<tr>
<td>Rough endoplasmic reticulum</td>
<td>RER</td>
</tr>
<tr>
<td>Single stranded DNA</td>
<td>ssDNA</td>
</tr>
<tr>
<td>Targeting vector</td>
<td>TV</td>
</tr>
<tr>
<td>Term</td>
<td>Short Form</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Tetracycline resistant</td>
<td>Tet(^R)</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>T1D</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>T2D</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>T2D</td>
</tr>
<tr>
<td>Untranslated region</td>
<td>UTR</td>
</tr>
<tr>
<td>Wild Type</td>
<td>WT</td>
</tr>
<tr>
<td>Zeocin resistance-p-chlorophenalanine</td>
<td>ZR/PN</td>
</tr>
</tbody>
</table>
List of Appendices

Southern Analysis Methods, Performed by Dr. Chen He

115
1) Literature Review: Introduction

The purpose of this Masters project is to generate a transgenic mouse line expressing an insulin biomarker enabling the precise measurement of insulin production from an allograft versus recipient origin. To do this, the mouse insulin C-peptide sequence will be modified to the human sequence. In order to achieve this, it was necessary to modify the mouse insulin gene. The literature review is subdivided into four sections: insulin gene regulation and function, differences between mouse and human insulin genes, models of diabetes and tools to generate a mouse line. The purpose of the first section is to familiarize the reader with insulin through a comprehensive review of insulin regulation and function. The review begins at level of insulin primary structure and continues to transcription, mRNA stability, translation, secretion, and biological activity. The structure, regulation and function of the human and mouse insulin genes and C-peptides will then be compared and summarized. In the third section, the current animal models of diabetes will be reviewed. The goal here is to provide an overview of the mechanisms of some models and their limitations. The final section will provide a review of the tools used to generate a mouse line.
a) Diabetes is caused by insufficient insulin production and/or signaling

The prevalence of diabetes is increasing rapidly worldwide. The international diabetes federation predicts that 552 million people will have diabetes by 2030. Within Ontario alone, there has been a 69% increase in the incidence of diabetes during 1995-2005 (Shapiro et al. 2006; Lipscombe & Hux 2007). Diabetes is hyperglycemia resulting from insufficient insulin to maintain glucose homeostasis. There are two general etiologies of diabetes: Type 1 and Type 2. Type 1 diabetes (T1D) is an autoimmune disease that destroys insulin-producing β-cells. T1D is largely a genetic disease (Todd 2010) and accounts for 5-10% of all diabetics (Daneman 2006). Type 2 diabetes (T2D) is more prevalent and accounts for the rest of all diabetics (Federation n.d.). T2D is caused by impaired insulin release or signaling and is influenced by genetics and lifestyle factors.

Insulin is required for cells to utilize glucose and amino acids normally. Lack of glucose in human and rodent cells leads to death from acute acidosis from fatty acid oxidation. In response to insulin depletion, glucagon secretion from the liver is not inhibited so that glycogen is continuously catabolized into glucose. Circulating glucose levels increase, eventually leading to osmotic diuresis. Lipid stores are used to create energy in liver through lipolysis in the liver. Fatty acid oxidation leads to the generation of ketone bodies (acetoacetate, β-hydroxybutyrate, acetone) whose production drops blood pH. This cascade of events causes the patient to become acidotic, leading to hyperventilation, diuresis, polydypsia, and risk of cerebral edema leading to death. Delivery of exogenous insulin or amplification of insulin secretion or signaling can mitigate the direct effects of diabetes. However, even if insulin levels are carefully monitored and controlled, long-term serious secondary nerve and blood vessel damage is often a devastating reality.
Most of the pancreas is designated to produce exocrine enzymes aiding in the digestion of nutrients within the intestines. However, 1% is referred to as islets of Langerhans. These islets contribute to maintaining metabolic homeostasis. Insulin producing β-cells represent approximately 80% of the islet while the remaining 20% secrete glucagon, somatostatin, pancreatic polypeptide, and neuropeptides. Alpha cells secrete glucagon that results in the releases of glucose from glucagon stores in the liver. Delta cells secrete somatostatin that inhibits secretion from the neurocrine and endocrine system. PP cells secrete pancreatic polypeptide that regulates both the endocrine and exocrine functions of the pancreas as well as hepatic glycogen stores and gut secretions.

Insulin secretion is regulated at the level of transcription, translation and secretion. Once secreted, it functions to enhance glucose uptake and anabolic pathways while inhibiting catabolic pathways in insulin responsive tissues such as muscle and adipose

b) Type 1 Diabetes

T1D is an autoimmune disease resulting in the destruction of β-cell. It is primarily a genetic disease. However, it has become clear that factors beyond genetics play a role in the development of T1D(Gittes 2009; Maslowski & Mackay 2011; Oliver-Krasinski & Stoffers 2008; Filippi & Herrath 2008; Redondo et al. 1999). Individuals with T1D have a profound lack of β-cells, while other cells within the islets are generally preserved. In early diabetes, insulinitis occurs within the islets, this is characterized by enlarged nuclei, degranulated β-cells and chronic inflammatory infiltrates. In the early stages of diabetes, β-cells with high levels of insulin have more leukocyte infiltration(Dor et al. 2004; Foulis et al. 1986; Salpeter et al. 2010; Porat et al. 2011). This led to the hypothesis that insulin itself is the auto-antigen causing diabetes. When insulin is expressed in the thymus, the site of T-cell maturation, immune cells are sensitized to insulin and autoimmune reactions are less likely to occur. This seems to confer protection from diabetes(Kawaguchi et al. 2002; Jolicoeur et al. 1994; Deutsch et al. 2001;
Pugliese et al. 1997; Johansson et al. 2007; Yechoor et al. 2009). In the absence of such protection, destruction of β-cells proceeds, but is not clinically apparent until approximately 80% of the cells are destroyed. Once destroyed, the β-cell population seldom recovers -although there have been rare reports in young individuals(Van Hoof et al. 2009; Gepts & In 't Veld 1984; Cohen & Melton 2011; Cheng et al. 2013) (as cited by(Kroon et al. 2008; Epstein & Atkinson 1994; Nostro et al. 2011)).

c) Stem cell therapy for T1D

T1D patients regulate their glycaemia by the administration of exogenous insulin. However, even with stringent glycemic control, individuals with diabetes have a higher probability of developing cardiovascular disease and retinal diseases(Hentze et al. 2009; Nathan et al. 2005; Reichard et al. 1993). Insulin pumps provide an improvement to injections(Cheng et al. 2012; Pickup 2012) but do not prevent secondary morbidities. Islet cell transplantation has been investigated as a means of regulating diabetes(Sneddon et al. 2013; Shapiro et al. 2000). However, there are numerous limitations to this therapy including the lack of sufficient numbers of donors and prevalent graft loss at follow-up(Harper et al. 1981; Shapiro et al. 2006). A recent summary of islet transplantation trials (2007-2010) suggests that these issues are still current, though overall outcomes have improved(Barton et al. 2012).

Stem cell therapy offers a solution to these problems, as it would provide a renewable source of islet cells or insulin producing progenitor cells. Renewable, functional β-cells derived from stem cells would be superior to exogenous insulin because they would regulate glucose homeostasis in a much more sensitive fashion. Scientists have attempted to harness our understanding of β-cell differentiation in vivo (reviewed here(Davies et al. 1994; Gittes 2009; Oliver-Krasinski & Stoffers 2008)) to direct cell fate for therapeutic use. This includes trying to understand the regenerative potential of β-cells in vivo(Duvillié et al. 1998; Dor et al. 2004; Salpeter et al. 2010; Porat et al. 2011), endogenous β-cell precursor populations(Steiner et al. 1985;
Kawaguchi et al. 2002; Melloul et al. 2002; Deutsch et al. 2001; Weiss 2013; Johansson et al. 2007; Yeohoor et al. 2009), and differentiating cells in vitro(reviewed here (Van Hoof et al. 2009; Cohen & Melton 2011; Cheng et al. 2013)). While the first two methods, in vivo trans-differentiation and β–cell precursors, highlight the plasticity of cells they do not hold promise of providing sufficient cell numbers for therapeutic use. The latter, in vitro differentiation generates the most excitement for the regenerative medicine community. Early attempts to produce β-cells directly from ESCs failed to produce fully mature, functional and safe cells. These cells displayed an immature, polyhormonal phenotype(Pugliese et al. 1997; Kroon et al. 2008; Nostro et al. 2011). Furthermore, they were contaminated with undifferentiated ESC, making them dangerous for forming teratomas (Steiner et al. 1985; Hentze et al. 2009).

Within the last two years, there has been a push to differentiate β-cells from an endocrine precursor population. Purely endodermal populations can be propagated in culture and can then be differentiated into β-cells. β-cells derived from these endodermal progenitor cells respond to glucose in culture(Weiss 2013; Cheng et al. 2012; M. Liu, Hodish, Rhodes & Arvan 2007b) or in vivo(Shoelson et al. 1992; Sneddon et al. 2013). This method limits the possibility of propagating and transplanting undifferentiated, teratoma-forming contaminants(Cheng et al. 2012), and appears to improve the generation of mono-hormonal, glucose responsive β-cells. The future of stem cell research will continue to improve differentiation efficiencies. These in vitro differentiated cells could then be used as cell therapies for T1D.

d) Insulin gene, expression and function

i) Evolutionary constraints of the primary and secondary structure of the insulin gene

Insulin is conserved across species, including mice and humans, though there are notable differences. Humans have one insulin gene (INS) that lies on chromosome 11(X. Zhu et al. 2002; Harper et al. 1981), while mice and rats have two non-allelic insulin genes Ins1 and Ins2.
In mice, *Ins1* is found on chromosome 19 (Steiner 2004; Davies et al. 1994) while *Ins2* is resides on chromosome 7 (Steiner 2004; Duvillié et al. 1998).

The structural characteristics, regulation and processing of insulin have been described in detail (reviewed here (Chen et al. 2002; Steiner et al. 1985; Melloul et al. 2002; Weiss 2013)). Insulin is translated as preproinsulin consisting of a 24 amino acid (AA) signal peptide, followed by proinsulin: N-term-B-chain-Arginine(R)-R-C-peptide-Lysine(K)-R-A-chain-C-term. The signal peptide is removed to create proinsulin comprised of B and A chains separated by the C-peptide. The C-peptide facilitates the protein’s tertiary structure by folding so that the 21 AA A chain can link to the 30 AA B chain by disulphide bonds. The mature 6-kilo Dalton insulin protein is generated following cleavage of the C-peptide. There are several proteins that have similar structural characteristics to insulin including insulin like growth factors (IGF) and relaxin (Steiner et al. 1985). IGF-I and II consist of conserved disulphide bridges and a linking sequence like C-peptide that is not normally removed. Relaxin has similar primary and secondary structures to insulin.

The human *INS* and mouse *Ins2* genes are homologous (Figure 1A, B). They consist of three exons separated by two divergent introns. The introns vary in length, but not in location. Intron 1 is found within the 5’ untranslated region (UTR), while intron 2 separates residues 6 and 7 of the C-peptide chain. This division is so highly evolutionary conserved that it is even found in hagfish (Powell 1988; Steiner et al. 1985). The location of the human insulin on chromosome 11 is relatively homologous to that of Ins2 on chromosome 7 in mice. The upstream region consists of highly repetitive VNTR elements that contribute to diabetes penetrance (M. Liu et al. 2003; Pugliese et al. 1997). The flanking region of *INS* and *Ins2* are typical of many proteins: a eukaryotic promoter located approximately -22bp of the transcriptional start site consists of TATAAG and a polyadenylation sequence, AATAA is found +20bp downstream from the sequence.
Not surprisingly, the structure of *Ins1* differs from *Ins2* and the human *INS* (Figure 1) as the two sequences diverged approximately 20-35 million years ago (M. Liu, Hodish, Rhodes & Arvan 2007b; Steiner et al. 1985; Micallef et al. 2007). *Ins1* results from the retrotransposition of an incompletely processed (mRNA) *Ins2* transcript. It consists of two exons separated by a single intron, where the second exon encodes the entire preproinsulin protein.

The preproinsulin gene is under evolutionary constraint, with more selection being placed on the B and A chains, while the flanking regions of the insulin gene have evolved in a selectively neutral fashion. The sequence necessary for segregation of the nascent peptide chain across the rough endoplasmic reticulum (RER) is also well conserved. Between mouse and human, there are ten AA mutations within the B-chain and one within the A-chain (Figure 2) (Steiner et al. 1985). The B and A chains are evolutionary conserved, as these are the biologically active chains. The conservation is found within sequences relating to packaging of insulin, molecular self-association, hexamerization and those relating to insulin receptor binding affinities. In humans, mutations can affect metabolism by causing protein misfolding (Polonsky et al. 1986; Weiss 2013; M. Liu, Hodish, Rhodes & Arvan 2007b) or changes in insulin receptor binding affinities (Ohtomo et al. 1996; Shoelson et al. 1992).

ii) The C-peptide

The C-peptide is under less selective pressure than the B and A chains as it is ultimately removed from insulin during processing. There are nine mutations between the mouse and human C-peptides and five mutations between the two mouse sequences (Figure 2). There are some reports that the C-peptide functions as a molecular spacer, allowing for proper folding of insulin, so the B and A chains can align and are processed correctly (Rigler et al. 1999; X. Zhu et al. 2002). To function as such, there are some constraints within the sequence. In general, the C-peptide is approximately 30 AA and is comprised of mainly nonpolar and glycine residues (Rigler et al. 1999; Steiner 2004). The glycine residues allow for high levels of
Acidic residues play two roles within the peptide; they maintain neutrality and assist in proper folding (Wahren et al. 2000; Steiner 2004). Mutations within the first five acidic AA can greatly inhibit protein refolding properties (Wahren et al. 2007; Chen et al. 2002). In contrast, studies have also shown that the C-peptide sequence can be grossly mutated without affecting insulin folding or processing. Deletion of the C-peptide including dibasic peptides for cleavage (Karlsson et al. 1987; Powell 1988) or doubling its length (M. Liu et al. 2003) does not seem to influence protein folding in vitro. Additionally, insertion of a fluorescent tag (Inagaki et al. 1992; M. Liu, Hodish, Rhodes & Arvan 2007b; Micallef et al. 2007) does not alter insulin processing or secretion. The sequence flanking the C-peptide sequence codes for four basic amino acids that provides signals for insulin processing enzymes discussed further below.

The C-peptide is secreted with insulin from the β-cell and is commonly used as a biomarker of insulin production. It has a half-life of 30 minutes, which is much longer than the 5 minutes of insulin. Due to its predictable storage and degradation, the C-peptide can be used to extrapolate an individual’s insulin secretion patterns (Naya et al. 1995; Polonsky et al. 1986).

The relatively long half-life of this peptide has led to the idea that it is biologically inert. However, there is evidence that the C-peptide binds to renal tubule cells (Cordle et al. 1991; Ohtomo et al. 1996) through a pentapetide sequence EGSLQ (Lu et al. 1997; Rigler et al. 1999). When binding, it may act through a G-protein coupled receptor to improve the functionality of endothelial cells and neural tissues (Melloul et al. 2002; Rigler et al. 1999) and prevent structural abnormalities often found in diabetic patients (Petersen et al. 1994; Wahren et al. 2000; Marshak et al. 1996; German & J. Wang 1994). Though there is substantial literature regarding this peptide’s physiological effects, there are limitations to these arguments. The pentapetide sequence is not conserved across species, which suggests it is not necessary for metabolic homeostasis. If the C-peptide were to play a role in normal physiology, then individuals with T2D with high levels of the circulating peptide would have reduced incidence of neurovascular and renal comorbidities. On the contrary, individuals with T2D and high C-peptide levels still develop endothelial and neural complications, but by different mechanisms than T1D (Ohlsson et al. 1993; Wahren et al. 2007; Melloul et al. 1993).
Figure 1. Human *INS*, mouse *Ins2* and *Ins1* genetic structures. A, B Human insulin and mouse insulin 2 are homologs, C insulin 1 is the result of retrotransposition of the mouse insulin 2 gene. [Boxes represent exons, filled boxes represent coding sequence, empty boxes represent untranslated regions, and lines represent introns.]
Figure 2. COLBALT protein alignment of the insulin peptide sequence for human and mouse insulin1 and 2. The insulin sequence is well conserved between mouse and human sequences. Dissimilarities lie within the B-chain and C-peptide. The C-peptide of the mouse and human genes can be distinguished by monoclonal antibodies.
iii) Regulation of insulin mRNA transcription

Insulin gene expression is regulated in part by transcription from the insulin locus through promoter activation of the sequence upstream of the insulin gene. Transcription is positively regulated from cis acting factors binding to E, A and GAGA- boxes, RIPE3b sequences, cAMP response elements (CRE), C2 elements, and variable number tandem repeats (VNTR) regions. E, A, CRE, C2 and VNTR elements will be discussed here. Silencers negatively regulate insulin expression by reducing promoter activity. Nutrients and hormones affect transcription levels by activating proteins that bind to response elements within the promoters. Promoter activation leads to changes in the chromatin status of the insulin gene so that it can be readily transcribed.
(1) Promoter elements

E boxes are common to rat Ins1 and human sequences (Inagaki et al. 1992; Karlsson et al. 1987) and bind basic helix-loop-helix family of transcription factors. E boxes (consensus sequence CANNTG) come in two variations, E1 and E2. These boxes bind transcription factors in a tissue specific manner. The transcription factor NeuroD binds to E boxes (Philippe & Missotten 1990; Naya et al. 1995) in cooperation with other tissue specific elements to activate the insulin enhancer. Several proteins such as c-jun (Crowe & Tsai 1989; Inagaki et al. 1992), BETA3 (Melloul et al. 2002; Naya et al. 1995), Id (Grill & Cerasi 1974; Cordle et al. 1991), c/EBPß (Drucker et al. 1987; Lu et al. 1997) inhibit transcriptional activation, perhaps by limiting expression of a coactivator (Roach 1991; Melloul et al. 2002).

A boxes are AT rich sequences that bind the homeodomain proteins. These sequences are grouped into five groups (A1-A5) where A2 sequences have the consensus sequence GGAAAT and A1,3,4 and 5 have the consensus sequence TAAT. A3 is considered to have the most dramatic effect on insulin transcription (Chrivia et al. 1993; Petersen et al. 1994; Marshak et al. 1996; German & J. Wang 1994) which binds pancreatic duodenal homeobox– (Pdx-1) (Foulkes et al. 1991; Ohlsson et al. 1993; Melloul et al. 1993). Pdx-1 is involved in early commitment of the primitive gut to a pancreatic fate. This homeodomain protein also activates insulin gene expression as well as other metabolism related proteins including glut2, cokinase, IAPP, and somatostatin.

cAMP response elements bind basic region leucine zipper family members of transcription factors CRE binding proteins (CREB). The human INS gene (Gepts & In ’t Veld 1984; Inagaki et al. 1992; Inada et al. 1999) and rat Ins1 (Inada et al. 1998; Philippe & Missotten 1990) and Ins2 (Read et al. 1997; Crowe & Tsai 1989) genes contain CRE elements. Generally, the consensus sequence of CRE is TGACGTCA, but rat and human have different sequences which may lead to a variation in their response to cAMP (Fujitani et al. 1999; Melloul et al. 2002).
cAMP increases in response to glucose (Oliver-Krasinski & Stoffers 2008; Grill & Cerasi 1974), glucagon and glucagon-like peptide-1 (GLP-1) (Smith et al. 1999; Drucker et al. 1987). In response to elevated cAMP, PKA phosphorylates CREB (reviewed here (Kennedy et al. 1995; Roach 1991)) allowing it to interact with transcriptional coactivators (Moriyama et al. 2003; Chrivia et al. 1993) and basal transcriptional units. CRE modulator proteins (CREM) are isoforms of CREB produced through alternative splicing (Bell et al. 1982; Foulkes et al. 1991). CREM can either be activators or repressors. CREM activates insulin production more so than CREB (Pugliese et al. 1997; Gepts & In ’t Veld 1984; Inada et al. 1999). Repressors compete with activators to repress insulin secretion (Epstein & Atkinson 1994; Inada et al. 1998; Pugliese et al. 1997).

C2 are found upstream of the human (Dubois-Lafforgue et al. 2002; Read et al. 1997; Thébault-Baumont et al. 2003) and rat insulin genes. These elements have similar structures in rodents and humans (German et al. 1990; Fujitani et al. 1999). They bind transactivator PAX6 (Shushan & Cerasi 1999; Oliver-Krasinski & Stoffers 2008; Peshavaria 2000) and repressor PAX4 (Macfarlane et al. 1999; Smith et al. 1999). PAX4 represses PAX6 activity during embryogenesis but is turned off after birth. PAX6 expression remains activated throughout adulthood.

The VNTR region is located upstream of INS and Ins2 in the mouse. VNTRs consist of a consensus sequence ACAGGGGT(G/C)(T/C)GGGG with a variable number of repeats. Class I consists of approximately 40 repeats and predisposes individuals to type I diabetes (German et al. 1990; Kennedy et al. 1995; Melloul et al. 1993). In a mouse model of T1D, expression of Ins1 under the regulation of Class I VNTR promoter accelerates disease onset (Lu et al. 1997; Moriyama et al. 2003; Curry et al. 1968). In humans, Class III VNTRs (140-200 repeats (Lu et al. 1997; Bell et al. 1982) (as cited by (Ritz-Laser 1999; Pugliese et al. 1997; Gremlich et al. 1997; Briaud et al. 2001; Jacqueminet et al. 2000)) have been reported to confer expression of INS to the thymus (Nathan et al. 2005; Epstein & Atkinson 1994; Koranyi et al. 1992; Pugliese et al. 1997; Reichard et al. 1993) and thus reduced susceptibility to T1D. Class III VNTRs have been reported in the mouse Ins2 promoter, leading to protective affects against hyperglycemia in
a model of T1D (Leibowitz et al. 2003; Dubois-Lafforgue et al. 2002; Thébault-Baumont et al. 2003). Putting this together, the two mouse Ins genes confer different susceptibilities to metabolic disease. This could lead to inaccurate translation of results from mouse models of human T1D.

(2) Nutrient and hormonal effects on insulin gene expression

Insulin transcription is activated in response to metabolic cues and mediated through the transcription factors mentioned above. As would be expected, insulin is expressed in the presence of high glucose. However, prolonged glucose and elevated free-fatty acid levels can have deleterious effects on insulin transcription. Hormones such as insulin itself, GLP-1, leptin and growth hormones also affect insulin transcription.

Insulin is secreted in response to elevated glucose, concomitantly insulin pre-mRNA transcription increases as well. Glucose regulates insulin gene transcription through mediator proteins such as cAMP (X. Wang et al. 2001; German et al. 1990; Skoglund et al. 2000), PKA, and PI3K, among others. Up-regulation of insulin pre-mRNA, an un-spliced transcript, increases within 60 minutes of exposure to high glucose, while mature mRNA levels increase within 48 hours (Evans-Molina et al. 2007). This change in insulin transcripts is largely regulated by Pdx-1 transcription factor (Itoh et al. 1978; Shusan & Cerasi 1999; Peshavaria 2000). Elevation of glucose results in a conformational change of Pdx-1 that allows it to be transported to the nucleus from the cytoplasm (Dubois-Lafforgue et al. 2002; Macfarlane et al. 1999; Brunstedt & Chan 1982; Itoh & Okamoto 1980). Once in the nucleus, Pdx-1 binds it’s A3 sequence to activate transcription. In the absence of an A3 binding domain, glucose responsiveness is inhibited (M. Welsh et al. 1985; German et al. 1990; Melloul et al. 1993).
In contrast, chronically elevated glucose levels are cytotoxic and result in reduced insulin transcription. Elevated glucose leads to decreased pdx-1 mRNA and concomitant decreased in PDX-1 activation of A3 box. Furthermore, there is a decreased affinity of other transcription factors for their promoter element(Tillmar 2001; Lu et al. 1997; Curry et al. 1968; Marshak et al. 1999) and increased expression of some repressors of transcription(Wicksteed 2001; Lu et al. 1997).

Similar to chronic glucose elevation, increased non-esterified free-fatty acids (FFA) can have paradoxical affects on insulin transcription. In normo-physiological levels, FFA enter ß-cell metabolism and provide energy for the cell. However, as is the case for glucose, prolonged exposure to FFA inhibits insulin transcription(Tillmar 2001; Ritz-Laser 1999; Gremlich et al. 1997; Briaud et al. 2001; Jacqueminet et al. 2000).

Hormones that regulate metabolism, such as insulin, GLP-1, growth hormone and leptin can also affect insulin transcription. It is debated whether the presence of insulin provides a negative feedback to inhibit its own expression(Wagner & García-Blanco 2001; Nathan et al. 2005; Koranyi et al. 1992; Reichard et al. 1993), or if it has no affect at all(Y. K. Kim et al. 2000; Leibowitz et al. 2003). GLP-1 is an incretin hormone that is secreted from L-cells in the intestine. It increases insulin transcription within ß-cells through cAMP and PKA activation dependent pathways(Moreira et al. 1998; X. Wang et al. 2001; Skoglund et al. 2000). While the growth hormone somatropin also increases ß-cell proliferation and stimulates gene transcription, leptin, in line with its function in metabolism, inhibits insulin transcription.

iv) Post transcriptional modifications of insulin mRNA

Insulin biosynthesis is considered to be the rate-limiting step in insulin secretion(Dreyfuss & V. N. Kim 2002; Itoh et al. 1978), so ß-cells rely on large stores of insulin mRNA for translation.
Pre-mRNA is processed to form mature insulin mRNA. Insulin mRNA stability is increased by elevated glucose levels. This increased stability is mediated by binding of polypyrimidine tract-binding proteins (PTB) to the 3’untranslated region of insulin mRNA.

Insulin mRNA stability rapidly increases in the presences of glucose. Short term (0-3 hour) exposure of islets to glucose leads to increased insulin levels without generation of newly synthesized mRNA(Webb et al. 2000; Dubois-Lafforgue et al. 2002; Fred & N. Welsh 2009; Brunstedt & Chan 1982; Itoh & Okamoto 1980; M. Welsh et al. 1985). In the presence of low glucose, the half life of the mRNA is 29 hours, while in high glucose its stability increases to 77 hours(Knoch et al. 2004; M. Welsh et al. 1985). Considering the tight metabolic regulation of insulin transcription cited previously, this increase in mRNA stability leads to an increase in total mRNA transcripts over time in the presence of high glucose(Xie et al. 2003; Tillmar 2001).

Increase in insulin mRNA stability is likely mediated by the 3’UTR and binding of PTB. Wicksteed et al. identified a sequence (UUGAA) within the 3’UTR, found after the translational stop site and before the polyadenylation sequence, that is conserved among species(Permutt 1974; Wicksteed 2001; Itoh et al. 1978; Tillmar 2001; Jahr et al. 1980). This sequence was thought to increase mRNA stability by serving as binding sites for PTB(Gilligan et al. 1996; Tillmar 2001). PTB assist in post-transcriptional modification through RNA splicing(Wicksteed et al. 2007; Wagner & Garcia-Blanco 2001; Wicksteed et al. 2010), cap-independent translation(Goodge & Hutton 2000; Y. K. Kim et al. 2000), polyadenylation(Huang & Arvan 1995; Moreira et al. 1998), RNA localization and mRNA stabilization(Orci et al. 1987; Dreyfuss & V. N. Kim 2002). PTB expression increases in the presence of high glucose(Bennett & Eisenberg 1994; Webb et al. 2000; Fred & N. Welsh 2009). In addition to binding to insulin, PTB also binds other insulin vesicle associated proteins(Davidson et al. 1988; Knoch et al. 2004) that are involved in insulin post-translational processing. cAMP dependent PKA increase in β-cells leads to PTB phosphorylation of Ser-16, resulting in nuclear export of PTB bound mRNA(Alarcón et al. 1993; Xie et al. 2003). Following nuclear export of the mRNA sequence it can be translated.
v) Insulin biosynthesis

Insulin secretion is partially regulated at the level of translation. Mediators of glucose responsiveness with respect to translation have not been elucidated. Translation occurs in the ER, and the protein is then processed throughout its transport to the secretory vesicles where it is primed for exocytosis.

In response to glucose, preproinsulin translation increases rapidly. There is a fifty fold increase in insulin biosynthesis in response to elevated glucose(Rhodes et al. 1992; Permutt 1974; Itoh et al. 1978; Jahr et al. 1980), as well as other metabolites. To initiate this increase in translation, there is a large increase in ribosome recruitment, initiation and elongation factors. However, these events do not appear to be regulated by glucose. Translation initiation factors such as eEF2, eIF2alpha, eIF4E and phosphoprotein are phosphorylated to initiate translation. This phosphorylation is not mediated by elevation in glucose(Curry et al. 1968; Gilligan et al. 1996; Wiser et al. 1999). Instead, the insulin mRNA itself may regulate translation. A stem loop structure within the 5’UTR stimulates translation of insulin(Curry et al. 1968; Wicksteed et al. 2007; Wicksteed et al. 2010). This stem loop structure is also observed in the 5’UTR of insulin processing proteins that are upregulated in the presence of glucose(Xu et al. 1999; Goodge & Hutton 2000).

Insulin is translated like other exocytosed proteins. The signal peptide mediates co-translational translocation to the rough ER. Translation then proceeds through the B-chain, C-peptide and A chains. Once inside the ER, the signal peptide is rapidly removed, and the disulphide bonds form between the A and B chains. The proinsulin moves through the ER to the Golgi apparatus where it is sorted and packaged.
The Golgi provides a $\text{Zn}^{2+}$, $\text{Ca}^{2+}$ rich environment that facilitates packaging of the proinsulin molecules. The proinsulin has one aromatic-like side, and one side that is complex. This complex side interacts with $\text{Zn}^{2+}$ where the carboxylates of the amino acid side chains likely form a metal ion-binding site within their core. Six proinsulin molecules aggregate to form hexamers around $\text{Zn}^{2+}$ ions (Rorsman & Renström 2003; Huang & Arvan 1995). The A and B chains of insulin are on the interior of the hexamer, while the C-peptide chain is on the surface. The soluble proinsulin hexamers bud off from the Golgi to form vesicles.

Proinsulin processing occurs within the vesicles (Eliasson et al. 1997; Orci et al. 1987) by the membrane bound enzymes Protein Convertase (PC)-1, -2 and Carboxypeptidase H (CPH) (Barg et al. 2002; Bennett & Eisenberg 1994; Barg et al. 2001). PC1 and PC2 are Ca$^{2+}$ dependent endoproteases (Eddlestone et al. 1985; Davidson et al. 1988) that are regulated by elevated glucose (Leto & Saltiel 2012; Alarcón et al. 1993). The cleavage of proinsulin occurs by cleaving the A/C then the B/C junctions through a des$^{31,32}$–proinsulin intermediate (Taniguchi et al. 2006; Rhodes et al. 1992). PC1 cleaves the B/C terminus of the proprotein at the dibasic R-R site. PC2 then cleaves at the C/A junction at the K-R residues. CPH, a Zn$^{2+}$ dependent exoprotease, then removes the dibasic residues RR and KR resulting in insulin. With the release of the C-peptide, the insulin hexamers become insoluble and aggregate within the vesicles.

vi) Insulin secretion

Insulin secretion from the β-cell is controlled by glucose and other metabolites through changes in membrane potential. Interaction between membrane bound channels and the vesicles allow ready release in response to stimulus. An adenosine triphosphate (ATP) sensitive potassium ($\text{K}_{\text{ATP}}$) channel maintains a constant efflux of potassium resulting in a negative membrane potential. Glucose enters the cell through GLUT2 transporter and is metabolized. The resulting increase in cellular ATP closes the $\text{K}_{\text{ATP}}$ channel, depolarizing the cell. Calcium enters the cell
through L-type Ca\(^{2+}\) channels(A. J. F. King 2012; Curry et al. 1968; Lenzen 2007; Wiser et al. 1999) and the influx of Ca\(^{2+}\) triggers exocytosis of the insulin containing vesicles.

The release of insulin occurs in a biphasic manner(Deeds et al. 2011; Curry et al. 1968; Makhlouf et al. 2003), with a rapid initial phase and a prolonged secondary phase. Normal individuals have a biphasic release of insulin secretion, while those with T2D have a blunted first phase. Pools of secretory granules must undergo priming so that they can be released without any additional modifications. This vesicle priming includes a number of processes including the formation of SNARE complexes(Smukler et al. 2011; Xu et al. 1999), and occurs in an ATP, Ca\(^{2+}\), temperature and time dependent fashion. In response to an increase in intracellular Ca\(^{2+}\) levels, there is a rapid transient release of primed vesicles(Baeyens et al. 2005; Rorsman & Renström 2003). When the primed granules are released, ATP energy and time is required to prime newly synthesized vesicles(A. J. F. King 2012; Eliasson et al. 1997). These vesicles are released over a longer period of time at a lower intensity. Priming may also be dependent on acidification of the granules. Inhibition of V-type H\(^{+}\)ATPases or CIC3 channels inhibits vesicle priming and secondary insulin secretion(Shapiro et al. 2000; Barg et al. 2002; Lenzen 2007; Barg et al. 2001). Compounds that increase cellular cAMP levels increase granule mobilization(H. R. Kim et al. 1994; Eddlestone et al. 1985). Individuals with T2D exhibit a loss of the first phase of insulin secretion because of a decreased ability to produce ATP, reducing the ability to form vesicles(Muoio & Newgard 2008). Amplification of insulin secretion can increase the size, priming and mobilization of granules(Lebovitz 2011). Insulin circulates systemically and activates insulin receptors in the tissues of muscle and adipose tissue.

vii)Insulin function

The function of insulin is to increase cellular uptake of glucose in tissues. Tissues require glucose to create energy in the form of ATP. Cellular intake of glucose is mediated by GLUT transporters -primarily GLUT4 within adipose tissues and skeletal muscle. Failure of GLUT4
action contributes to the development of insulin resistance and T2D (Rees & Alcolado 2005; Leto & Saltiel 2012).

Tissues have different glucose requirements. For example, the brain requires glucose constantly. Hypoglycemia can lead to loss of consciousness or death. However, prolonged hyperglycemia can lead to blindness, renal failure, cardiac failure, peripheral vascular disease and neuropathy. To maintain normoglycemia, the liver supplies glucose by glycogen catabolism during periods of fasting. After eating, increased blood sugar leads to secretion of insulin as discussed above.

Insulin signaling regulates GLUT4 transport to the cell membrane (Leiter 1982; Taniguchi et al. 2006). Once mobilized to the cell membrane, glucose can enter the cell. Uptake of glucose is the rate-limiting step in glucose metabolism. Furthermore, insulin signaling inhibits glucagon secretion, decreases hepatic glucose production and lowers free fatty acid concentrations.
e) Mouse Models to study T1D

i) ß-cell ablation models

ß-cell ablation models mimic T1D by elevating blood glucose. Destruction of ß-cell by chemical induction is the most commonly used approach (Grossman et al. 2010; A. J. F. King 2012; Lenzen 2007) as it is a cost effective and relatively quick method. However, spontaneous mutants and transgenic models are also available. ß-cell ablation models are appropriate for studies that aim to decrease blood glucose in a ß-cell independent manner, for example by delivering exogenous insulin. This is an ideal model to investigate cell therapies (Ashcroft & Rorsman 2012; Deeds et al. 2011; Ron 2002; Makhlof et al. 2003; Støy et al. 2007). Cell therapies include in vitro differentiated ß-cells (Sneddon et al. 2013; Cheng et al. 2013; Kroon et al. 2008), insulin producing progenitor cells (Hong et al. 2007; Smukler et al. 2011). Insulin producing cells are often transplanted into the kidney capsule or fat pad (Rees & Alcolado 2005). Animals that are cured need to have the cells excised to rule out regeneration of endogenous ß-cells (Mathews et al. 2002; Baeyens et al. 2005).

There are two chemically induced models; alloxan and streptozoticin (STZ). These are relatively unstable molecules (Mathews et al. 2002; A. J. F. King 2012; Pearson et al. 2008) that could lead to technical errors as they need to be used quickly. These chemicals are glucose analogues that enter through GLUT2 transporters. They compete with glucose for uptake, so function best under fasting conditions. GLUT2 is expressed primarily on ß-cells, but is also expressed on tubular and hepatocyte cells of the kidney and liver. These off target effects can lead to misinterpretation of results. These chemical means of inducing ß-cell death differs from that of naturally occurring diabetes where ß-cells are destroyed by immune reactions – here they die by necrosis.
Alloxan is primarily used to study the effects of reactive oxygen species (ROS) on β-cell toxicity (Barber 2005; Shapiro et al. 2000; Gurley 2005; Lenzen 2007). Alloxan is taken up by GLUT2 transporters and reacts with glucokinase. Inhibition of glucokinase decreases glucose oxidation, reducing the cellular levels of ATP and inhibiting ATP-dependent insulin processing and secretion mechanisms. Alloxan leads to cell death by entering a redox cycling reaction that generates significant ROS leading to DNA fragmentation and cell death. Cell death may occur through other mechanisms as well, including altering calcium homeostasis (Pozzilli et al. 1993; H. R. Kim et al. 1994; Rees & Alcolado 2005).

STZ is another glucose analogue as well as a nitrosourea analogue and is the most popular form of β-cell ablation (Hanafusa et al. 1994; Rees & Alcolado 2005). It has low affinity for GLUT2 but accumulates within β-cells. Its toxicity is largely dependent on the alkylating activity of the molecule, which results in methylation of DNA. DNA repair mechanisms become over-activated, leading to depletion of cellular ATP. A single high dose leads to rapid β-cell ablation. Multiple low doses over a prolonged period of time leads to decreased β-cell volume and decreased insulin secretion (Wicker et al. 2005; Leiter 1982). STZ models have been shown to be unstable and β-cells can regenerate over time with tight glycemic control (Todd & Wicker 2001; Grossman et al. 2010). This suggests that the more effective a potential therapy is at maintaining glycemic control, the less effective the STZ model becomes.

The AKITA mouse is a model in which a spontaneous mutation in the Ins2 gene leads to protein misfolding, ER stress and β-cell death. This mimics a specific type of progressive neonatal onset diabetes (M. King et al. 2008; Ashcroft & Rorsman 2012; Ron 2002; Støy et al. 2007), or insulin deficient non obese type 2 diabetes (M. King et al. 2008; Hong et al. 2007; A. J. F. King 2012). Male animals develop diabetes at 3-4 weeks of age, where males experience higher disease penetrance. This model is an alternative to the commonly used STZ model (Hay & Docherty 2006; Mathews et al. 2002; Bucchini et al. 1986). Animals exhibit fewer unrelated systemic side effects caused by chemical induction (Warram & Krolewski 1984; Mathews et al. 2002; Moore et al. 2001; Pearson et al. 2008). This renders this model more suitable for investigation of some

ii) Immune mediated models of diabetes

The most commonly used model of insulin dependent diabetes is the non-obese diabetic (NOD) mouse (Deltour et al. 1993; Pozzilli et al. 1993; Rees & Alcolado 2005). This model is used to study the onset of autoimmune mediated diabetes and therapies that may prevent onset. Mice develop diabetes at 3-4 weeks of age, with CD4+ CD8+ T-lymphocyte infiltrates in islets. By 10-14 weeks of age, autoimmune disease leads to the death of over 90% of β-cells. Females are more likely to develop the disease than males (Leroux et al. 2001; Hanafusa et al. 1994). There are numerous similarities between the mouse and human systems of disease (Mehran et al. 2012; Wicker et al. 2005), including MHCII (Bell et al. 1982; Todd & Wicker 2001) and the functional elements of the promoter of human INS and mouse Ins2. Humanizing of the mouse immune system has significantly improved this model (Shapiro et al. 2006; M. King et al. 2008; Fan et al. 2009). However, many treatments that have been successful in NOD mice have not translated to the clinic (Babaya et al. 2006; M. King et al. 2008; Dubois-Lafforgue et al. 2002; A. J. F. King 2012; Thébault-Baumont et al. 2003; Moriyama et al. 2003).

d) Differences between the mouse and human insulin structure

Human and mouse metabolism are similar as evidenced by the numerous translational studies that have been spear headed in rodents. However, with numerous difficulties in translation of therapies from mice to humans, many argue the differences between the two species are too drastic to draw appropriate conclusions. One culprit to overcome is purely genetic: humans have a single allele that is regulated in a similar manner to Ins2 in mice, while Ins1 seems to be regulated differently.
Human INS and mouse Ins2 have the same expression distribution, while Ins1 shows a different pattern. The promoter elements of the mouse Ins2 and INS genes are not identical, but are functionally conserved (Thomas & Capecchi 1987; Hay & Docherty 2006; Bucchini et al. 1986). In both species, it is expressed within the β-cell as well as in the thymus, brain and yolk sac, while Ins1 is not. During development, the INS (Copeland et al. 2001; Warram & Krolewski 1984; Moore et al. 2001) and Ins2 genes (Gertsenstein et al. 2010; Duvillié et al. 1998; Kuroda et al. 2009) are subject to paternal imprinting in the yolk sac, while there is little evidence for imprinting of Ins1. Throughout development and in adulthood, Ins2 is expressed at approximately two times the level of Ins1 (Ware et al. 2003; Deltour et al. 1993; Seong et al. 2004). Although they are both expressed at different ratios in wild type animals, one Ins gene can functionally compensate for knockout (KO) of the other (Gertsenstein et al. 2010; Leroux et al. 2001). Ins2 KO mice develop β-cell hyperplasia, with islets increasing to approximately three-fold wild type mass. Though Ins1 KO mice also develop significantly larger islets, they do not expand to the same extent of Ins2 KO. Their ability to compensate could have intriguing consequences for the study of the development of type IV hypersensitivity (autoimmunity) in diabetes. A recent report used the difference in mouse insulin gene expression to dissect the role of hyperinsulinemia in obesity development (Chinwalla et al. 2002; Mehran et al. 2012).

Insulin is a major autoantigen in diabetes both in humans and mouse models. As described previously, CD4+ CD8+ T-cells mediate destruction of insulin containing β-cells. T-cells mature in the thymus. When introduced to insulin in during maturation, they are less likely to initiate an immune response to it. Therefore, it has been proposed that expression of insulin in the thymus leads to reduce incidence of β-cell targeted autoimmune reactions (International Mouse Knockout Consortium et al. 2007; Bell et al. 1982). Animals with a thymus specific deletion of insulin have an elevated risk of diabetes (Zhang et al. 1998; Shapiro et al. 2006; P. Liu 2003; Fan et al. 2009). NOD Ins1+/−,Ins2−/− animals develop rapid onset of diabetes (Osoegawa et al. 2000; Babaya et al. 2006; Dubois-Lafforgue et al. 2002; Thébault-Baumont et al. 2003; Moriyama et al. 2003), while those with Ins1+/−,Ins2+/+ were protected
from diabetes. This further highlights the need to distinguish \textit{Ins1} and \textit{Ins2} expression in mouse models of diabetes in order to improve the clinical translation of therapeutics.

\textbf{g) Generating a targeted mouse mutation: the toolbox}

Mouse models are valuable tools for investigating human physiology and therapeutics. They are generated through deletion, insertion or mutagenesis of the genome most commonly either by gene targeting utilizing mouse ESC (mESC), random chemical mutagenesis, or pronuclear injection of transgenes. Precise modifications generally utilize homologous recombination in mESCs. This is done by designing targeting vectors \textit{in silico} and producing them either by classical molecular cloning techniques (Sharan et al. 2009; Thomas & Capecchi 1987) or recombineering (Riele et al. 1992; Copeland et al. 2001). These targeting vectors can then be introduced into mESC by transfection. Clones with the correct genetic modifications can then be used to generate chimeras (Figure 3). Germ-line transmission of the altered genome is the foundation of the new mouse line that can be characterized and used as a tool for a given research model.
Figure 3. Schematic to generation of a mouse line. Designing mutations in silico, and generating a mutated vector by gene synthesis are used to generate mouse lines. Recombineering can be used to finish the vector by insertion of the mutated construct into a bacterial artificial chromosome (BAC) and then retrieving homology arms. Targeting in ESC will generate mutated clones, after which a chimera can be generated. [HA50bp = 50bp homology arm, hcp = human C-peptide, RC = Resistance cassette, HA = homology arm]
i) Choice of genetic background for the generation of genetically engineered mouse models

The C57BL/6 inbred strain has historically been the most widely used genetic background to maintain transgenic mouse lines. This led it to be the first strain to be fully sequenced. By using the C57BL/6 background both for the in silico design, targeting and screening, results from consecutive experiments can reliably be compared. However, gene targeting is traditionally done with ESCs derived from the 129 inbred genetic background (Deng & Capecchi 1992; Gertsenstein et al. 2010) as C57BL/6 is by far less permissive both for derivation of ESCs as well as germline competence of the resulting chimeras (Deng & Capecchi 1992; Ware et al. 2003; Seong et al. 2004).

Recently, the Nagy lab succeeded to derive C57BL/6Ntac mESC (McCarrick et al. 1993; Gertsenstein et al. 2010) (hereafter referred to as C2) and demonstrated their capacity to efficiently contribute to the germline of chimeric animals.

Since the completion of the mouse genome project (Hanson & Sedivy 1995; Chinwalla et al. 2002) a large effort has been initiated to characterize the function of each of the genes discovered. The international knockout mouse consortium (IKMC) (Lamotte et al. 2004; International Mouse Knockout Consortium et al. 2007) has undertaken the task of knocking out as many genes as possible in the C57BL/6 mouse model. With this resource, we have access not only to many KO cell line to study metabolism, but also tools to generate further allelic modification of interest.
ii) Recombineering

*Escherichia coli (E.Coli)* based recombineering is *in vivo* genetic engineering mediated by λ prophage derived proteins. Proteins *gam*, *bet* and *exo* enable homologous recombination between short (40-60bp) regions of homology (Mansour et al. 1988; Zhang et al. 1998; P. Liu 2003) which in turn allows DNA to be manipulated without the need for restriction sites or complex cloning procedures. Due to the small homology region required for engineering, short oligonucleotides conferring homology to a region of interest can be designed to flank a DNA element of interest. Nearly any recombineering vector could then be produced readily by polymerase chain reaction (PCR). The three λ prophage proteins interact to mediate recombination. While the *gam* gene inhibits RecBCD single stranded DNA (ssDNA) exonuclease activity that allows linear DNA to be transfected, *bet* is a ssDNA binding protein that guides the transfected DNA to regions of homology found in the lagging strand during DNA synthesis. The *exo* gene is a 5'→3' DNA exonuclease, which together with *bet* facilitates homologous recombination. The DNA to be modified can be a low copy plasmid, such as a bacterial artificial chromosome (BAC). Within this project, BACs are large (150-300 kilobasepairs (kb)) C57BL/6 mouse genome fragments cloned into pBACe3.6 chloramphenicol resistant vectors (McCarrick et al. 1993; Osoegawa et al. 2000). These BACs are useful for generating transgenic animals because the gene of interest and all its cis regulatory sequences are found within a single BAC (Smukler et al. 2011; Sharan et al. 2009; Ledermann 2000; Auerbach et al. 2000; Schoonjans et al. 2003; Poueymirou et al. 2007). After modifying the gene of interest within the BAC, the tailored element and large targeting-vector homology arms can be retrieved by another round of recombineering using gap-repair plasmids. Gap repair plasmids are linearized plasmids with homology regions at the 5’ and 3’ end of the vector. When recombineering is induced, homologous recombination occurs between the plasmid and gap repair plasmid, where the gap is fixed by incorporation of homology arms. Taken together, the recombineering system provides the means to generate gene-targeting vectors using only homologous recombination.
iii) Target vector design

Homology arms are stretches of DNA that flank the 5’ and 3’ ends of the target gene of interest and facilitate homologous recombination between vector and genome. Long regions of isogenic DNA are preferable for these arms. Isogenic DNA is DNA of the same background, in this case C57BL/6 genomic DNA. In a study targeting the Rb locus in 129 mESC, using BALB/c genomic DNA to build the construct greatly reduced the targeting efficiency (Nagy et al. 2010; Riele et al. 1992). Deng and Capecchi (Steiner et al. 1985; Deng & Capecchi 1992; Pugliese et al. 1997; Moore et al. 2001) demonstrated that homologous recombination increases as the length of homology arms are increased. The maximum efficiency of targeting was reached with 14kb arms, with efficiency being reduced with arms that were longer or shorter (Leroux et al. 2001; Deng & Capecchi 1992; Deltour et al. 1993; Hay & Docherty 2006).

To improve gene-targeting efficiencies, cells that have undergone homologous recombination are selected for. The most efficient means of selecting for targeting events is positive-negative selection (Paronen et al. 2003; McCarrick et al. 1993; M. King et al. 2008; Rees & Alcolado 2005). Positive selection is common to all cloning techniques and is used in targeting. Here, the modified gene of interest confers an antibiotic resistance. By treating the cells with the appropriate antibiotic, only the cells that contain the resistance gene will survive. In a reciprocal fashion, negative selection confers cellular toxicity. Depending on the negative selection used, treatment with a drug such as in Gancyclovir (Nathan et al. 2005; Hanson & Sedivy 1995; Reichard et al. 1993), or the mere presence of the cassette, as in diphtheria toxin A-chain (DTA) (Patterson et al. 2009; Lamotte et al. 2004), will cause cell death. In general gene targeting, positive selection is used to select for homologous recombination events, while the negative selection will induce apoptosis in random integration events. Such dual selection strategies have been shown to increase the rate of homologous recombination events 2000 fold (Grossman et al. 2010; Mansour et al. 1988). This frequency further improves when DTA is used (Steiner 2004; McCarrick et al. 1993; B. Zhu et al. 2007; M. Welsh et al. 1985).
iv) Chimeras

Following successful gene targeting, a new mouse line can be generated via germline transmitting chimeric animals. Chimeras can be produced in many ways (Smukler et al. 2011; Ledermann 2000; Auerbach et al. 2000; Schoonjans et al. 2003; Poueymirou et al. 2007), but most efficiently by aggregation of ESC with diploid or tetraploid mouse embryos (Nagy et al. 2010). In diploid aggregation, ESC are aggregated with morula or 8 cell stage embryos and implanted into the uterus of pseudo-pregnant females.
2) Rationale, Hypothesis and Aims

Animal models are a valuable resource for biomedical research. Genetically modified and disease induced mouse strains have been widely used to mimic common features of human metabolic syndromes. However, there are differences between mouse and human metabolism that may limit translation of pre-clinical findings. Drug induced models of T1D have been used extensively to study this disease. However, there are severe limitations to this approach.

Humans have a single insulin gene, while mice have two insulin genes that are expressed from different genetic loci. In comparison to the human, these loci are located in homologous (Ins2) (Steiner et al. 1985; Pugliese et al. 1997; Moore et al. 2001) and non-homologous (Ins1) (Steiner et al. 1985; Leroux et al. 2001; Guzman et al. 1995; Deltour et al. 1993; Melloul et al. 2002; Hay & Docherty 2006; Weiss 2013) genomic settings. This raises concerns with respect to the current mouse models (Paronen et al. 2003; M. King et al. 2008; Rees & Alcolado 2005). Co-expression of insulin from two different alleles may not reproduce human physiology, rendering mouse models of diabetes less representative of the human condition. At this time, there is no simple means of monitoring insulin expression from each mouse Ins gene.

T1D is characterized by the complete loss of β-cells. Individuals with this disease require exogenous sources of insulin for the duration of their lifetime. Even with stringent glucose control, individuals with T1D will still develop severe comorbidities (Nathan et al. 2005; Reichard et al. 1993). The possibility of creating a renewable source of β-cells in vitro that could be used therapeutically would improve the lives of millions (Patterson et al. 2009). It is therefore not surprising that there is a considerable interest in β-cell differentiation from stem cells and expansion from progenitor populations.
The most valuable assay of assessing these cell sources is by transplanting them into diabetic mice and monitoring their ability to regulate blood fluctuating blood glucose levels. The currently most widely used STZ model is not ideal as β-cells may regenerate over time (Wei 2013; Grossman et al. 2010; Osoegawa et al. 2000; M. Liu, Hodish, Rhodes & Arvan 2007b). For experimental purposes, when β-cells or their progenitors are derived from mice, transplanting them back into other mice can make interpretations of results difficult because the graft and the host insulin production are indistinguishable.

These issues can be addressed by generating a mouse model that will tag the mouse insulin production.

C-peptide is a biomarker of insulin secretion both in humans and mouse models. The sequences of the human and mouse C-peptide share several important characteristics (Steiner et al. 1985; Steiner 2004; B. Zhu et al. 2007; M. Welsh et al. 1985) for the proper folding and processing of proinsulin molecules. However, they vary sufficiently between the two species to make them readily distinguishable by monoclonal antibodies and ELISA.

I hypothesize that replacing the mouse Ins1 and Ins2 C-peptides with the human C-peptide will allow the activity of two insulin alleles to be distinguished in vivo. This would be a valuable tool to distinguish Ins1 production from Ins2 within a single mouse. Furthermore, this tool could be used to differentiate insulin production of a mouse β-cell graft from that of its’ host. To exchange the mouse C-peptide for that of the human, I propose to generate a ‘humanized C-peptide’ mouse line by gene targeting of C57BL/6 mESC followed by germ line transmission. To create this mouse model, I have divided the project into four specific aims that apply to the mouse insulin alleles, Ins1 and Ins2,

1) Generation of humanized C-peptide recombineering cassettes.
2) Generation of humanized C-peptide targeting vectors.
3) Humanizing the mouse C-peptide in mESC.
4) Generation and characterization of a humanized C-peptide mouse line.

The scope of this thesis comprises Steps 1 to 3, thus providing the key biological resources for the development of a novel mouse model for pre-clinical studies of diabetes.

1) My first aim was to generate a vector with the human C-peptide sequence within the mouse \( \text{Ins1} \) and \( \text{Ins2} \) genes. I inserted a positive selection tag within the sequence of the human C-peptide flanked by the mouse proinsulin gene sequences to create recombineering vectors for the mouse \( \text{Ins1} \) and \( \text{Ins2} \) genes. This enabled me to seamlessly ‘humanize’ the mouse C-peptide sequence without interfering with its processing or function.

2) My second aim was to create vectors to target the human C-peptide into the mouse (aim 3). In this aim, I could use the recombineering vectors from aim 1 to modify the mouse proinsulin gene sequences in BACs. The humanized C-peptide proinsulin sequence and flanking genomic regions were retrieved from the BAC to create gene-targeting vectors for the mouse \( \text{Ins} \) genes.

3) My third aim was to replace the mouse C-peptide with the human sequence in mESC. To achieve this, the targeting vector from aim 2 was applied to mESC. The genome of the cell and the targeting vector underwent homologous recombination, inserting the human C-peptide within the mouse proinsulin gene. By targeting the mouse genome with the human C-peptide targeting vector, I created the cell resource that would be the foundation for this mouse line tool.
3) Materials, Methods and Experimental Procedures

Materials and methods provide detailed experimental protocols used for each technique, while experimental procedures describe the technique applied to generating the \textit{Ins1} and \textit{Ins2} humanized C-peptide mouse line. All primers and oligos are listed in order of appearance at the end of the section (Table 2).
a) Materials & Methods

i) Transforming chemically competent cells

Chemically competent DH5α (Invitrogen) *Escherichia coli* (E. Coli) cells were used to expand plasmids. Manufacturer guidelines were followed. A 50µL aliquot of cells was thawed on ice. Plasmid (1-10ng) or ligation product (1uL of 5 fold diluted ligation reaction) was gently mixed with the cells. Cells were incubated on ice for 30 minutes, heat shocked for 30 seconds at 42°C using a heat block and returned to ice for 2 minutes. 1mL of liquid broth (LB) media was added and the cells were agitated at 225rpm for one hour. Cells were spun down at 4000rpm for 4 minutes. The supernatant was removed and the cell pellet was re-suspended in 50µL of LB. The 50uL was streaked on selective agar plates and incubated at 37°C overnight.

ii) Colony screening (plasmids)

Colonies resulting from transformations were isolated by miniprep. 2mL of selective LB media was aliquoted to 15mL bacterial culture tubes. In general 4-15, colonies were picked using a 200uL pipette tip and touched to the aliquoted LB in the tube, the tip was discarded. Media was placed on a shaker at 225rpm and 37°C overnight (16-18 hours). Miniprep procedures were followed to isolate plasmid DNA (QIAprep Spin Miniprep Kit QIAGen, Valencia, CA, USA).

Isolated plasmids were screened by restriction digest: 5µL DNA, 0.5uL of enzyme (All restriction enzymes were purchased from NEB, Ipswich, MA, USA), 1.5uL of appropriate 10×NEB buffer, 0.15µL of 100× BSA and 7.85uL ddH₂O. Digests were placed at the appropriate temperature for the restriction enzyme for 2hours. Digests resolved on a 1% agarose
gel made in 1×TAE buffer at 100V for 30 minutes. Pictures of the gel were taken using a UV camera. Digestion products were compared to the predicted products from downloaded or manipulated sequences using ApE plasmid editor (http://biologylabs.utah.edu/jorgensen/wayned/ape/)(http://biologylabs.utah.edu/jorgensen/wayned/ape/).

iii) Sequencing and analysis of results

Sequencing was performed at The Centre for Applied Genomics, Sequencing facility (DNA Sequencing Facility, The Centre for Applied Genomics, TMDT-MaRS Building 101 College St. Rm 14-201, East Tower Toronto ON M5G 1L7, http://tcag-sequencing.ccb.sickkids.ca/Finch/). Isolated plasmid or PCR DNA [Plasmids - 200-300ng, PCR product (>4kb) - 150-200ng, PCR products (2-4kb) - 100-150ng, PCR products (1-2kb) - 50-100ng] was resuspended in 7.2uL ddH₂O and 0.7µL of 5µM of a primer. Results were analyzed with nucleotide NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) by aligning the sequencing result with the relevant annotated sequence.

iv) Electroporating bacteria with pSC101-BAD-γβα Red/ET expression plasmid

DH10B E.Coli(Polonsky et al. 1986; Herrath & Nepom 2009; Osoegawa et al. 2000; A. J. F. King 2012; Shoelson et al. 1992; Grossman et al. 2010) cell strain containing BACs were used for recombineering purposes, a process that relies on pSC101-BAD-γβα Red/ET (pRET) (GeneBridges, Heidelberg Germany) expression. pRET is a Tetracycline resistant (Tet²), Arabinose inducible (Arab¹) plasmid that carries genes required for homologous recombination (λ-phage encoded: γ- inhibits E.Coli RecBDC exonuclease activity, Redα-5→3’ exonuclease, Redβ- single stranded DNA
annealing protein) under the control of the pBAD arabinose inducible promoter (Buchini et al. 1986; Steiner et al. 1985; Guzman et al. 1995; Melloul et al. 2002; Weiss 2013). To electroporate DH10B cells with pRET, the procedure recommended by GeneBridges Quick and Easy BAC Modification Kit was followed. BAC colonies were picked and grown in 3mL of selective LB shaking at 225rpm and 32°C overnight (16-18 hours). Clones were sub-cultured at 25µl/mL media from the overnight culture and cultured shaking at 225rpm for 2-3 hours at 32°C. Electroporation cuvettes and ddH₂O were placed in the fridge during this time. Cultures were spun down at 11,000rpm for 30 seconds at 4°C. The supernatant was aspirated and the cells were washed gently with 1mL of the cold ddH₂O. The centrifugation and aspiration were repeated and the cell pellet was re-suspended in 50µL of cold ddH₂O. 1-2ng of pRET was added to the cell mixture. The cell mixture was transferred to a chilled electroporation cuvette. The cells were electroporated at 1.8kV, 25µF and 400Ohms (BioRad GenePulserI/II system, BioRad, ON, CA). The electroporated cells were resuspended in LB and allowed to recover shaking at 225rpm for 1 hour at 32°C, plated on Chloramphenicol and Tetracycline selective agar and incubated at 32°C for 16-18 hours. This selects for cells containing Chloramphenicol resistant BACs and Tetracycline resistant pRET (B+pRET).
v) Recombineering

Recombineering occurred in B+pRET. To perform recombineering, cells were induced, electroporated with a linearized recombineering vector, then recombineering could take place. B+pRET colonies were picked and grown in 3mL of Chloramphenicol and Tetracycline selective LB shaking at 225rpm and 32°C overnight (16-18 hours). Clones were subcultured at 25μl/mL media from the overnight culture and cultured shaking at 225rpm for 2-3 hours at 32°C. Electroporation cuvettes and ddH₂O were placed in the refrigerator during this time. At 2 hours the OD₆₀₀ of the culture was measured, when the culture reached an OD₆₀₀ of approximately 0.3, 25μL/mL of 10% L-Arabinose was added. Cultures were moved to 37°C for 45 minutes or 1 hour shaking at 225rpm. Cultures were spun down at 11,000rpm for 30 seconds at 4°C. The supernatant was aspirated and the cells were washed gently with 1mL of the cold ddH₂O. The centrifugation and aspiration steps were repeated and the cell pellet was resuspended in 50μL of cold ddH₂O. 1ug of linearized vector was added to the cell mixture. The cell mixture was transferred to a chilled electroporation cuvette. The cells were electroporated at 1.8kV, 25μF and 400Ohms. The electroporated cells were resuspended in LB and allowed to recover shaking at 225rpm for 1 hour at 37°C, plated on selective agar and incubated at 32°C for 24-36 hours.

vi) Screening colonies (BACs and targeting vectors)

Colonies resulting from recombineering were isolated by miniprep using a QIAprep Spin Miniprep Kit buffers followed by ethanol precipitation. Colonies were picked and grown in selective LB shaking at 225rpm and 32°C overnight (16-18 hours). Cells were centrifuged at 4000rpm for 10 minutes at 4°C. Pellets were re-suspended in 250μL of P1 with RNaseA and transferred to an eppendorf tube, 250μL of P2 was added and mixed by gently inverting, 250μL of P3 was added and mixed by inverting. The mixture was centrifuged for 10 minutes at 13,000rpm. The supernatant was transferred to a new eppendorf tube and 3M sodium acetate at
pH4 was added to 10% of the volume. Two volumes of ethanol (Fisher Scientific, Pittsburgh, PA, USA) were added to precipitate the DNA. The mixture was placed in the freezer for 30 minutes then centrifuged at 13000rpm for 30 minutes. The DNA pellet was washed in 1mL of 70% ethanol and centrifuged for 10 minutes at 13,000rpm. The 70% ethanol was removed and the pellet was air dried for 5 minutes. The DNA pellet was re-suspended in 10µL of ddH2O. This was used for screening by restriction digest or PCR.

To screen by PCR, the following reaction mixture was used: 2.5µL of 10xBuffer (with Mg2+), 0.5µL of 10mM deoxyribonucleotide (dNTP), 0.5µL of 10mM Forward primer, 0.5µL of 10mM Reverse primer, 0.2µL of Taq, 18.8µL of ddH2O, and 2µL of BAC DNA. The appropriate PCR program was used depending on the primer pair, and products were resolved on a 1% agarose gel made in 1XTAE buffer. PCR products were compared to the predicted size using ApE plasmid editor. Appropriate band sizes were gel purified (QIAquick Gel Extraction Kit, QIAGen, Valencia, CA, USA) using the manufacture guidelines and sent for sequencing.

vii) Targeting vector preparation

Targeting vectors (TV) that were retrieved through recombineering procedures were linearized and used for gene targeting. 100µg of TV was digested with 200U of AsisI for 16-18hours. 5µL of linearized vector was digested with XhoI, this product and 0.5µL of the linearized TV were resolved on an agarose gel as previously to confirm linearization. The linearized TV was purified by ethanol precipitation. The DNA pellet was then re-suspended in 50µL of ddH2O.

viii) ESC thawing, culture, and freezing
C2 cells were maintained in FBS-ES-DMEM (ES) media. ES media was prepared as described in Table 1.

**Table 1. FBS-ES-DMEM (ES) media.** This media was used for culturing C2 cells, and as a base for selective media following electroporation and freezing media.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Final Concentration</th>
<th>Company, Catalogue#</th>
</tr>
</thead>
<tbody>
<tr>
<td>High glucose DMEM</td>
<td>83%</td>
<td>Invitrogen, #11960-044</td>
</tr>
<tr>
<td>FBS</td>
<td>15%</td>
<td>Invitrogen, #12483-020</td>
</tr>
<tr>
<td>Gluta MAX</td>
<td>2mM</td>
<td>Invitrogen, #35050-061</td>
</tr>
<tr>
<td>Penicillin-Streptomycin</td>
<td>50µg/mL</td>
<td>Invitrogen, #15140-148</td>
</tr>
<tr>
<td>Non essential amino acids</td>
<td>0.1mM</td>
<td>Invitrogen, #11140-050</td>
</tr>
<tr>
<td>Sodium pyruvate</td>
<td>1mM</td>
<td>Invitrogen, #11360-070</td>
</tr>
<tr>
<td>Beta-mercaptoethanol</td>
<td>0.11mM</td>
<td>Invitrogen, #21985-023</td>
</tr>
<tr>
<td>Leukemia inhibiting factor</td>
<td>500 µg/mL</td>
<td>Millipore, #ESG1107</td>
</tr>
</tbody>
</table>

C2 passage 11 (p11) cells were thawed into ES media onto a 35mm plate coated with mitotically inactivated mouse embryonic fibroblasts (MEFs). Cells were always grown on MEFs in ES media and incubated at 37°C with 5% CO₂. Media was changed daily. MEFs were always plated onto the desired surface 24 hours before use. Cells were passed when they reached 70-80% confluency, first to a 60mm dish, then to a 100mm dish. Cells were always passaged 1:3 in this manner: the media was aspirated, the cells were washed with phosphate buffer solution (PBS) using the same volume as ES media previously used, trypsin was added at 1/10th the volume of ES media used, plates were moved to 37°C for 2-5 minutes, and cells were resuspended in half
the volume of ES media. Cells were centrifuged at 1200rpm for 3 minutes, the supernatant was removed and the cells were resuspended in new media and plated to the new surface.

Cells were frozen in 10% DMSO, 40% FBS, 50% ES media. A 2× freezing media (20% DMSO, 80% FBS) was prepared in advance and chilled at 4°C. When cells reached 70-80% confluency, cells were washed, trypsinized and centrifuged as above. The 2× freezing media was diluted with ES media and kept on ice. Following centrifugation, cells were resuspended in 1mL 1× freezing media per vial, aliquoted to freezing vials and placed on ice. Cells were quickly moved to -80°C freezer and transferred to liquid nitrogen the following day.

ix) Electroporation of ESCs

C2 p13 ESCs at 70-80% confluency growing on a 100mm dish were electroporated for gene targeting. On the morning of electroporation, media was changed and two electroporation cuvettes were placed in the fridge to chill at 4°C. A minimum of three hours later, the cells were washed, trypsinized and centrifuged. Following centrifugation, the supernatant was aspirated and the cell pellet was resuspended in 5mL of chilled electroporation buffer (E.P.B) (Millipore, Billerica, MA, USA). The cells were centrifuged again, and the E.P.B was aspirated. The cell pellet was re-suspended in 1.6mL of E.P.B, 0.8mL was aliquoted to each of the two chilled cuvettes. The linearized vector was added to each cuvette and resuspended. Cells were electroporated at 220V, 500µF, and ∞ Ohms. Cuvettes were placed on ice for 5 minutes then placed at ambient temperature and media was added to the cuvettes. Five minutes later the cells were resuspended, the contents of each cuvette were divided between two 100mm surfaces coated with half density feeders and ES media.

x) ESC selection
Following electroporation, ESCs were grown under Genticin (G418) selection to select for clones that had undergone homologous recombination with the targeting vector. Cells were put under selection 24-48 hours following electroporation, depending on their recovery. Well-recovered cells were placed under selection 24-30 hours later. Sparse or poorly recovered cells were placed under selection 36-48 hours after electroporation. Cells were cultured in ES media containing 150μg/mL of G418 until picking.

x) ESC colony picking

Colonies began to appear day 5 following electroporation, and continued to appear until day 8. As colonies appeared, they were picked and passed to 96-well plates. MEFs were aliquoted to 96-well plates, and MEF media was replaced with ES media. 30μL of trypsin was aliquoted to U-bottom 96-well plates. To pick, media was aspirated from the 100mm dish and replaced with 5mL of PBS. The colonies were picked under a phase-contrast microscope and transferred one colony per well to the trypsin containing 96-well U-bottom plate. Every 24 colonies, the plate was moved to the 37°C incubator for 2-3 minutes. The wells were resuspended in 100μL of non-selective ES media. Media was replaced with selective media 12-18 hours after picking.

xii) Passaging and freezing of ESC on 96-well plates

When the majority of the wells reached 80% confluency, the cells were passaged in triplicate MEF containing 96well plate format. MEF media was replaced with 50μL of ES media. To passage cells, media was aspirated and cells were washed with 100μL of PBS. PBS was aspirated and 30μL of Trypsin was aliquoted to each well. The plate(s) were moved to 37°C for 2-3 minutes. Cells were resuspended in 150μL of selective ES media. The 180μl was divided equally amongst the triplicate plates.
As the cells reached 70-80% confluency, one of three plates was passaged again as above. The remaining two plates were washed and trypsinized as above and the cells were resuspended in 100uL of 1× freezing media. Plates to be frozen were kept on ice, wrapped in an insulator and placed at -80°C.

**xiii) Screening ESC clones by PCR**

The final three plates that were passaged above were used to extract DNA for genetic screening to identify correctly targeted clones. Cells were allowed to reach 100% confluency. At this point, the media was aspirated and cells were washed three times with 100uL of PBS. PBS was aspirated and 50uL of lysis buffer (10mM Tris-HCl, pH 7.5; 10mM EDTA; 10 mM NaCL; 0.5% sarcosyle and 1mg/ml Proteinase K) was added to each well. Plates were incubated at 55°C overnight in a humid atmosphere. A NaCl/Ethanol mixture was prepared (150µL of 5M NaCl per 10mL 100% ethanol) and placed in the freezer overnight. The following morning 100µL of the cold NaCl/Ethanol mixture was added to each well. Plates were left undisturbed for 3 hours. Plates were inverted to remove the liquid, and wells were rinsed three times with 70% ethanol. Two plates were stored at -20°C in 70% ethanol.

The plate(s) that were to be used to screen by PCR were inverted and allowed to air dry for 5-15 minutes. 30-50µL of ddH20 was added to each well. PCR proceeded using a ¼ (12.5µL) reaction according to the manufacturers guidelines (KOD Xtreme HotStart Taq or Phusion Hot Start II High-Fidelity DNA Polymerase). 8µL/well of PCR product was resolved on a 1% agarose gel made in 1×TAE buffer at 100V for 3 minutes. 10µL of ddH2O was added to the remaining PCR and a second PCR was performed with primers pairs internal to the primary PCR. The PCR products were resolved as previously.
xiv) Screening clones by Southern Blots

Potential positive clones were screened by Southern Blot analysis. A 24-well plated was prepared in advance by coating it with MEFs, and the MEF media replaced with ES media prior to use. The 96-well plate containing the potential positive clone/s were thawed by placing it in a 37°C, 5%CO₂ incubator. When cells began to thaw, 80µL of media was added to the wells of interest, and the plate was returned to the incubator. Once completely thawed, the contents of each well were moved to a well of the 24-well plate. Media was replaced with selective media the following day. The cells were passed to a 6 well plate. When cells reached 70-80% confluency, two vials were frozen, and one volume was passed onto a new 6 well plate.

To extract DNA for Southern, these cells were allowed to reach 100% confluency, cells were washed, trypsinized and centrifuged as previously. Following centrifugation, cells were resuspended in PBS and re-centrifuged. The PBS was aspirated and the cell pellet was transferred to an eppendorf tube. 500µL of cell lysis buffer was added to the cell pellet and the cells were placed at 55°C overnight. 500µL of Pheno-Chlorormisoamyl was added to the digest. The digests were then centrifuged at 11000rpm for 10 minutes at 4°C. The liquid phase was transferred to a new eppendorf tube, and ethanol precipitation proceeded. The pellet was resuspended in 100µL TE buffer. The concentration was measured with a Nanodrop spectrometer (Fisher Scientific, Nepean ON). This DNA was then used for Southern blot analysis.

Southern probes were amplified by PCR (PrimeSTAR, Takara Bio, Otsu, Japan), resolved on a 1% agarose gel made in 1×TAE buffer for 30 minutes at 100V. The band of interest was excised and purified by gel extraction. 300ng of probe DNA was labeled with DIG-High prime (DIG-High Prime DNA Labeling and Detection Starter Kit II, Roche Applied Science, Laval, QC, CA) overnight following the manufacturer guidelines.
To prepare for a Southern, 15µg of DNA was digested in a 40µL reaction with 40 units of enzyme. The digest was resolved on a 0.75% agarose gel made in 1XTAE buffer for 18-20 hours at 20V. The gel was shaken in 0.8% HCl for 15 minutes, denatured in denaturation buffer for 30 minutes, neutralization buffer for 15 minutes, and 20× SSC for 10 minutes. A standard nylon (Amersham Hybond-N+ membrane, Cat# RPN303B, GE Health) membrane transfer was set up, and DNA was transferred for 18-24 hours. The membrane was UV cross-linked using Stratagene UV Stratalinker 1800, washed with ddH2O and air-dried. Once dried, DIG-Southern labeling and detection proceeded according to the manufacturer guidelines, with hybridization and high stringency washing temperatures specific to each probe.

b) Experimental Procedures: Targeting of mouse *Insulin 1* and 2 genes

i) Recombineering vector construction

All components of the humanized C-peptide recombineering vectors were designed *in silico* and generated by gene synthesis (Biomatik) and cloning.

(1) The design of recombineering vectors:

Using NCBI Ensembl
(http://useast.ensembl.org/Mus_musculus/Info/Indexhttp://useast.ensembl.org/Mus_musculus/Info/Index) the *Mus Musculus* (strain C57BL/6J) *Ins1*, *Ins2* and *INS* genomic and the annotated coding DNA (cDNA) sequences were retrieved. The sequences of genomic DNA (gDNA) were manually annotated in ApE plasmid editor using Ensembl annotated cDNA sequences. After
annotating the insulin cDNA, the mouse C-peptide sequences were replaced with the \textit{INS} C-peptide sequence within the plasmid editor to create human C-peptide 1 and 2 genes.

Gene synthesis (Biomatik) generated distinct mouse \textit{Ins1} and \textit{Ins2} humanized C-peptide recombineering vectors cloned into pBMH vectors, pBMH\_hCp1 (Figure 6A) and pBMH\_hCp2 (Figure 6B.) respectively. These plasmids were transformed into DH5\(\alpha\) chemically competent E. Coli, expanded and sequenced by T7 forward and M13 Reverse standard primers.

A resistance cassette with selection markers was cloned into the synthesized construct using Fse1 and Not1 restriction sites. An _frt-F3-PGK/EM7-Kanamycin/Neomycin-frt-F3_ resistance cassette (K/N\(^R\)) was generously provided by Andrew Smith (Edinburgh University). This resistance cassette provided EM7-promoted Kanamycin resistance (Kana\(^R\)) in bacteria during the recombineering process, and PGK-promoted Neomycin resistance (Neo\(^R\)) in embryonic stem cells. The synthesized constructs (pBMH\_hCp1 and pBMH\_hCp2) and K/N\(^R\) were digested with Fse1 and Not1. To confirm that the constructs were linearized, 2\(\mu\)L of the digestion was resolved on a 1\% agarose and a 1.8kb band containing the K/N\(^R\) cassette was gel extracted and purified. The linearized pBMH\_hCp1 and pBMH\_hCp2 were each ligated with the K/N\(^R\) digestion product. Following overnight ligation, the products were transformed to DH5\(\alpha\) E.Coli and streaked on Ampicillin and Kanamycin selective agar plates to select for colonies with Ampicillin resistant pBMH and Kana\(^R\) K/N\(^R\). Positive colonies were sequenced with T7 and M13-Reverse primers that confirmed the insertion of the resistance cassette. These plasmids, K/NR\_pBMH\_hCp1 and K/NR\_pBHM\_hCp2, were then used to isolate the recombineering vectors.

To generate the recombineering vectors, K/NR\_pBMH\_hCp1 and K/NR\_pBHM\_hCp2 were amplified in DH5\(\alpha\) \textit{Escherichia coli} (E. Coli), digested with Asc1 and purified by gel purification following the manufacture’s instruction. The Asc1 linearized recombineering vector was further amplified by PrimeSTAR HS DNA Polymerase PCR using oligonucleotides...
(Ins2OF, InsOR), gel extracted and purified. The K/NR_pBMH_hCp1 and K/NR_pBHM_hCp2
Asc1 were each Asc1 linearized, PCR amplified, DNA were then referred to as hCp1 and hCp2
recombineering vectors, respectively. These were used to generate gene-targeting vectors.

ii) Targeting vector construction

Humanized C-peptide targeting vectors were generated by homologous recombination with
BACs and the humanized C-peptide recombineering vectors.

The RP23-BAC library consisting of C57/BL6 gDNA fragments were cloned into
Chloramphenicol resistant pBACe3.6 vectors(Chen et al. 2002; Frengen et al. 1999; X. Zhu et
al. 2002; Bucchini et al. 1986) and grown in DH10B E.Coli bacteria(Weiss 2013; Osoegawa et
al. 2000; M. Liu, Hodish, Rhodes & Arvan 2007b). \textit{Ins1} and \textit{Ins2} RP23- BACs (B\textit{Ins1}: RP23-
232L8, 233M2, 478C24; B\textit{Ins2}: RP23-50N22, 51J21, 473M23) were retrieved (SP6, T7-
Reverse), and internally sequenced at \textit{Ins1} (Ins1Fwd, Ins1Rvs) or \textit{Ins2} (Ins2Fwd, Ins2Rvs)
locations. [Of note: Ins2 BACs RP23-17N3, 479C1, 299I6, 209O22 and 92L23 did not proceed
through recombineering and were discarded.]

\textit{Insulin 1 humanized C-peptide BAC}. Bins1+pRET were electroporated with hCp1.
Recombineering was allowed to take place and bacteria were plated on selective Tetracycline,
Chloramphenicol and Kanamycin plates. This selects for Chloramphenicol resistant BACs that
incorporated the Kanamycin resistant hCp1 and maintained pRET (BhCp1+pRET). The
BhCp1+pRET were screened by PCR and confirmed by sequencing. \textit{Ins1} PCR primers
amplified across the insulin-gDNA-resistance cassette junction (Ins1Fwd, K/N^{R}SR). Positive
colonies were sequenced by Ins1Fwd, Ins1Rvs, K/N^{R}SR and K/N^{R}EF.
**Insulin 1 humanized C-peptide targeting vector retrieval.** The target vector was retrieved by recombineering-mediated gap-repair. A gap-repair plasmid, pGRDT, was linearized and amplified with oligonucleotides (HAto5’Ins1, HAto3’Ins1) with Prime Star High Fidelity polymerase. The vector is now called pGRDTIns1. BhCp1+pRET were induced and electroporated with pGRDTIns1. Recombineering-mediated retrieval was allowed to take place, and bacteria were streaked on Ampicillin and Kanamycin selective plates. This selected for colonies in which Kanamycin resistant hCp1 targeting vector (TVhCp1) was retrieved with Ampicillin resistant pGRDTIns1. Candidate colonies were screened by EcoR1 and EcoRV digests. Sequencing with T7, M213 Rev, Ins1Fwd, Ins1Rvs, K/N^R SR and K/N^R EF were used to confirm that the homology arms were retrieved correctly and the humanized C-peptide element remained intact.

**Insulin 2 humanized C-peptide targeting vector.** An Ampicillin resistance (Amp^R) cassette flanked by the 50bp Ins2 recombineering homology arms was generated (hcp2_AmpF, hcp2_AmpR) Bins2+pRET were electroporated with the ampicillin recombineering cassette (Ins2Amp). Recombineering was allowed to take place and bacteria were streaked on Chloramphenicol, Ampicillin and Tetracycline selective plates. This selected for colonies with Chloramphenicol resistant BACs that had undergone homologous recombination with the Ampicillin cassette and maintained pRET (BAIns2+pRET). Colonies were screened by PCR (Ins2Fwd, Ins2Rvs) BAIns2+pRET were electroporated with the hCp2. Recombineering was allowed to take place and bacteria were grown in Chloramphenicol, Kanamycin and Tetracycline selective LB liquid media. This selected for colonies with Chloramphenicol resistant BACs that had undergone homologous recombination with hCp2 and maintained pRET (BhCp2+pRET). The retrieval vector pGRDT was linearized and amplified with PrimeSTAR HS PCR (TaKaRa Bio) (HAto5’Ins2, HAto3’Ins2). The vector was then referred to as pGRDTIns2. BhCp2+pRET were subsequently induced and electroporated with pGRDTIns2. Recombineering was allowed to take place and bacteria were plated on Ampicillin and Kanamycin selective plates. This selected for colonies in which Kanamycin resistant hCp2 targeting vector (TVhCp2) was retrieved with Ampicillin resistant pGRDTIns2. Possible candidates were screened by Sph1 and EcoR1 digests, and successfully recombineered and gap-repaired vectors were sequenced with T7, M213 Rev, Ins2Fwd, Ins2Rvs, K/N^R SR and K/N^R EF.
to confirm that the homology arms were retrieved correctly and the humanized C-peptide element was intact.

iii) Gene targeting in ESC

Gene targeting, selection and screening occurred as previously described. Resistant clones were picked and screened by embedded PCR, where the primary PCR product was diluted 4-fold and used as template for the secondary PCR (Ins2: Primer set 1: K/NRSR3, Ins2GenoR1; Primer set 2: K/NRSR2, Ins2GenoR2). Positive clones were screened using Southern analysis by Dr. Chen He in our laboratory (methods found in Appendices).

An external genomic probe confirmed Ins2 positive cells lines. The genomic probe was amplified from RP23-50N22 BAC (Probe 3F, Probe 3R) with PrimeSTAR HS DNA Polymerase 15ug of DNA extracted from the positive PCR screened clones were digested with 20 units of Sbf1 overnight and run at 20V on a 0.75% agarose gel overnight. DIG-High Prime DNA Labeling and Detection Starter Kit I was used to label and detect the probe.

iv) Chimera and transgenic animal creation

The Toronto Centre for Phenogenomics (Toronto, ON, CA) generated the chimeras. Positive cell lines were used in a diploid aggregation with outbred albino ICR embryos at the morula stage. Chimeras are bred to ICR females to determine germline transmitters (GLT). GLT will produce pups with black eyes, as they will have the C57BL/6 phenotype from the targeted C2 ESC. To determine GLT breeding pairs consist of a male chimera being bred to two ICR females or two female chimeras being bred male to a single ICR male.
### Table 2. Primers used in order of appearance.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 forward</td>
<td>gtaatacgactcactataggg</td>
</tr>
<tr>
<td>M13 Reverse</td>
<td>caggaacagctatgaccatg</td>
</tr>
<tr>
<td>Ins2OF</td>
<td>accttggtgtgagggggagctggtcttctcactacacccatgctccgcccgt</td>
</tr>
<tr>
<td>Ins2OR</td>
<td>tgttagagggagcagcatgctggtcagcactgatctacaatgcccacgcctt</td>
</tr>
<tr>
<td>SP6</td>
<td>atttaggtgacactatag</td>
</tr>
<tr>
<td>T7-Reverse</td>
<td>taatacgactcactataggggga</td>
</tr>
<tr>
<td>Ins1Fwd</td>
<td>cctgcctatctttagctagg</td>
</tr>
<tr>
<td>Ins1Rvs</td>
<td>tccatttccccctgcctagg</td>
</tr>
<tr>
<td>Ins2Fwd</td>
<td>gggctatgttggtgtaggcag</td>
</tr>
<tr>
<td>Ins2Rvs</td>
<td>ctcacccagccctatctcctcag</td>
</tr>
<tr>
<td>HAt5’Ins1</td>
<td>caatgtataaatctctagttttttcactttaaatcattgagcattttggctgtgatggtggttttc</td>
</tr>
<tr>
<td>HAt3’Ins1</td>
<td>gataagactcaagatcataatatgttctgctttagcacttaaactattgttaaaccagggccagttg</td>
</tr>
<tr>
<td>hcp2_AmpF</td>
<td>tgttagagggagcagcatgctggtcagcactgatctacaatgcccacgcctttgataaatggtttcatagcaggt</td>
</tr>
<tr>
<td>hcp2_AmpR</td>
<td>cttctacccggtggttggtggggagctggtctctacacccatgtctcaaggggatttgggtcatgaga</td>
</tr>
<tr>
<td>HAt5’Ins2</td>
<td>ccagctgggtctctctctctctgcaatcctcaagaccatggtgacctgctgtgaggtgttttc</td>
</tr>
<tr>
<td>HAt3’Ins2</td>
<td>taagccctttctcatgttctctttgcattcataatgcctctcaaaagttttgtaaaccagggccagttg</td>
</tr>
<tr>
<td>Ins2Fwd</td>
<td>gggctattgttggtgtaggcag</td>
</tr>
<tr>
<td></td>
<td>Sequence</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Ins2Rvs</td>
<td>ctcaacccagctatcttcag</td>
</tr>
<tr>
<td>K/NRSR3</td>
<td>cagaggccacttgtgtagcg</td>
</tr>
<tr>
<td>Ins2GenoR1</td>
<td>cagtcgggtgttcttacccact</td>
</tr>
<tr>
<td>K/NRSR2</td>
<td>catgctccagactgccttgg</td>
</tr>
<tr>
<td>Ins2GenoR2</td>
<td>ctcctctttctctcact</td>
</tr>
<tr>
<td>Probe 3F</td>
<td>agaacccaggagaacctgt</td>
</tr>
<tr>
<td>Probe 3R</td>
<td>gacataaaggaccttaggaaa</td>
</tr>
</tbody>
</table>
4) Results

a) Overview

The goal of this project was to establish two mouse models that carry the human C-peptide in the mouse $Ins1$ and $Ins2$ alleles respectively. To target these two genes, two recombineering vectors were first generated, in which the human C-peptide genomic sequence was flanked by the appropriate mouse insulin gene sequences at the knock-in site. In order to facilitate selection of targeted clones, a cassette with antibiotic resistance genes was included in the vector. Through a standard recombineering and gap retrieval procedure, the final gene-targeting vectors were obtained. Subsequent gene targeting followed the standard protocol of NorCOMM (North American Conditional Mouse Mutagenesis project). The resulting ESC clones that showed Geneticin resistance were further screened by PCR and Southern blots. The ESC clones that bore the correct recombination were used to generate chimeras, which are currently being tested for germ line transmission or have successfully transmitted to germline.
b) Generating mouse *Insulin1* and 2 gene-targeting vectors

i) Design of the humanized C-peptide recombineering vectors.

In order to seamlessly knock-in the human C-peptide sequence, a recombineering protocol was adopted to construct gene-targeting vectors to the mouse *Ins1* and *Ins2* genes. To allow insertion of the human C-peptide precisely into mouse insulin genes, the recombineering vector was constructed with a structure as shown in Figure 4 and Figure 5. The vector contains the human C-peptide DNA in the center, a selection cassette and flanking (50bp) homology regions. To build the vector, reference mouse and human genomic DNA as well as cDNA were obtained from NCBI Ensembl (http://useast.ensembl.org/index.html).

ii) Construction of the recombineering vectors.

For the recombineering vector of mouse *Ins1*, homology arms were created within the 1B chain flanking the 5’ of the C-peptide and 9 bp following the TAA stop codon within the 3’UTR. Unique Fse1 and Not1 restriction sites separated by 5 arbitrary base pairs were inserted 5 bp following the stop codon. Unique Asc1 sites were added to the 5’ and 3’ of this vector (Figure 4). This construct was generated by gene synthesis and was cloned into a pBMH vector. The synthesized product and vector is now called pBMH_hCp1 (Figure 6A).

For the recombineering vector of mouse *Ins2*, homology arms were created to include the entire cDNA-coding region. 96 bp flanking the 5’ and 69 bp flanking the 3’ of the C-peptide were included in the final gene synthesis product. The intron of *Ins2* separates the human C-peptide at residues 6 and 7 between exon 2 and 3 (Exon 2: EVEDPQ, Exon 3: VAQLEGGLGAGDLQTLALEVAQQ). This organization is conserved between mouse and
human genes (Steiner et al. 1985). Unique Fse1 and Not1 restriction sites separated by 5 arbitrary base pairs were inserted into the intron 15 bp 3’ of the second exon and 473 bp 5’ of the 3rd exon. Unique Asc1 sites were added to the 5’ and 3’ of this vector (Figure 5). This construct was generated by gene synthesis and was cloned into a pBMH vector. The synthesized product and vector was now referred to as pBMH_hCp2 (Figure 6B).

To incorporate K/NR into both pBMH_hCp1 and pBMH_hCp2, the cassette was cloned between the Fse1 and Not1 restriction sites (Figure 7). This generated K/NR_pBMH_hCp1 and K/NR_pBHM_hCp2 (Figure 8). The resulting K/NR_pBMH_hCp1 was screened with EcoRV and HindIII (expected band size: 2.8kb, 1.8kb and 0.2kb bands) and K/NR_pBHM_hCp2 was screened by EcoRI and HindIII restriction digestion (expected band size: 2.8kb, 1.7kb and 0.6kb bands) (Figure 9). Positive colonies were sequenced to confirm the insertion of the resistance cassette. K/NR_pBMH_hCp1 and K/NR_pBHM_hCp2 were then linearized to generate recombineering vectors.

To finalize the recombineering vectors, the K/NR_pBMH_hCp1 and K/NR_pBHM_hCp2 vectors were digested with Asc1 (Figure 10A). From K/NR_pBMH_hCp1 Asc1 digest, a 2.1kb band was extracted. From K/NR_pBHM_hCp2 Asc1 digest, a 2.5kb band was extracted (Figure 10B). The Asc1 K/NR_pBHM_hCp2 linearized recombineering vector was amplified by PCR to improve the frequency of recombineering.
**Figure 4.** *Ins1* humanized C-peptide recombineering vector strategy. The *Ins1* human C-peptide recombineering vector created such that the human C-peptide sequence was flanked by 50bp of the 1B and the 1A chain. Two unique restriction sites FseI and NotI were inserted within the 3‘UTR, which were used to insert a Kanamycin/Neomycin resistance cassette (K/N\(^R\)). 50bp of the 3‘UTR followed the resistance cassette. The two 50 bp regions were used as homology arms for recombineering.
Figure 5. *Ins2* humanized C-peptide recombineering vector strategy. The mouse 2B and 2A chains flanked the human C-peptide. Unique FseI and NotI restriction sites were inserted into the intron sequence, which were used to insert K/N\textsuperscript{R}. The 2B and 2A chains were used as recombineering homology arms.
Figure 6. Constructs used to generate recombineering vectors. A, B Gene synthesized humanized C-peptide constructs that will be cloned with C the K/N\textsuperscript{R} resistance cassette. Fse1 and Not1 restriction sites used for cloning are marked in black on the vectors.
Figure 7. Construct digests to insert the resistance cassette K/NR into pBMH_hCp1 or pBMH_hCp2. FseI and NotI restriction digests were used to clone the 1.8kb K/N\textsuperscript{R} (green box) into pBMH_hCp1 or pBMH_hCp2 (red box). Expected sizes can be seen on vector maps found in Figure 6.
Figure 8. Human C-peptide recombineering cassettes. Vector maps show the elements of the human C-peptide recombineering cassettes and restriction digest sites of A K/NR_pBMH_hCp1 (with AscI, EcoRV and HindIII sites shown in black) and B K/NR_pBHM_hCp2 (with AscI, EcoRV and HindIII sites shown in black).
Figure 9 Human C-peptide recombineering cassettes were cloned correctly. Screening digests of K/NR_pBMH_hCp1 with EcoRV and HindIII (2.8kb, 1.8kb and 0.2kb) and K/NR_pBHM_hCp2 with EcoR1 and HindIII (2.8kb, 1.7kb and 0.6kb) demonstrate generation of human C-peptide recombineering cassettes with the pBMH cloning vector.
Figure 10. Linearized humanized C-peptide recombineering cassettes. A pBMH_hCp1_K/NR and pBMH_hCp2_K/NR were digested with AscI B to generate linear recombineering vectors hCp1 (*2.1kb) and hCp2 (**2.5kb). Vector maps can be seen in Figure 8.
iii) Construction of gene-targeting vectors by recombineering.

Humanized C-peptide targeting vectors were generated by phage protein mediated homologous recombination between BACs and the humanized C-peptide recombineering vectors described above.

To obtain the *Ins1* humanized C-peptide targeting vector, homologous recombination took place between Bins1 and hCp1 (Figure 11). Mouse Bins1+pRET was electroporated with hCp1. Colonies resulting from recombineering were screened by PCR amplification of the junction between *Ins1* 5′UTR and the resistance cassette junction. Positive clones generated a 471bp amplicon. PCR amplifying the *Ins1* 5′UTR to human C-peptide junction, resulting in a 280bp amplicon was also used to screen colonies (Figure 12). Sequencing confirmed that the human C-peptide recombined with the BAC sequence at the correct location and generated the humanized C-peptide *Ins1* BACs (BhCp1). I then retrieved the TVhCp1 by recombineing BhCp1 with a gap repair plasmid (Figure 13) pGRDT that was linearized with EcoRV. The linearized plasmid was amplified with oligonucleotides that added 50 bp to the 5′ and 3′ of the plasmid, designated pGRDITins1. These additional 50bp arms are homologous to the *Ins1* genomic DNA at the 3′ and 5′ ends, respectively, of the final targeting vector. BhCp1+pRET was electroporated with pGRDITins1. Recombination was induced and TVhCp1 targeting vector was retrieved (Figure 14A). Restriction digest was used to screen candidate colonies. Digestion with EcoRI was expected to result in 5.1kb, 3.9kb, 3.2kb, 2.0kb, 1.4kb, 0.9kb, 0.2kb bands, and with EcoRV in 4.7kb, 3.8kb, 3.5kb, 2.8kb, 1.6kb, 1.1kb, 0.18kb, and 0.16kb bands (Figure 14B). Sequencing of the homology arms and the humanized C-peptide element confirmed that TVhCp1 was retrieved with 6.4kb 5′ and 5.4kb 3′ homology arms.

To obtain the *Ins2* humanized C-peptide targeting vector, homologous recombination between Bins2 and hCp2 was induced. The intron dividing the mouse *Ins2* C-peptide was initially replaced with an Ampicillin selection (Figure 15). Due to a large homology region between the
intron located in mouse *Ins2* C-peptide sequence and the sequence located in the middle of hCp2, unwanted recombination events often occurred. To increase the recombineering efficiency, an Amp<sup>R</sup> cassette flanked by the 50bp *Ins2* recombineering homology arms was generated to replace the mouse *Ins2* C-peptide and intron (Figure 16A). BIns2+pRET was electroporated with the Ampicillin recombineering cassette and resulting colonies screened by PCR showed that this Ampicillin cassette replaced the *Ins2* intron (Figure 16B) and reduced unwanted recombination with the recombineering vector. Following this step, recombineering with hCp2 and the targeting vector retrieval took place in liquid culture so that screening occurred concomitantly when the final vector was retrieved. The linearized pGRDT retrieval plasmid was amplified with oligonucleotides that added 50 bp to the 5’ and 3’ of the plasmid, designated pGRDTIns2. These additional 50bp arms are homologous to the *Ins2* genomic DNA at the 3’ and 5’ ends, respectively, of the final targeting vector. Following recombineering with hCp2 and gap repair with pGRDTIns2, TVhCp2 was retrieved (Figure 17A). Restriction digest was used to screen candidate clones. Sph1 digests resulted in 10.8kb, 4.2kb, 1.6kb, 1.1kb and 0.45kb bands, and EcoR1 digests resulted in 8.1kb, 5.0kb, 2.6kb, 2.2kb and 0.2kb bands (Figure 17B). Sequencing of the homology arms and the humanized C-peptide element confirmed TVhCp2 was retrieved correctly with 7.8kb 5’ and 4.5kb 3’ homology arms.
Figure 11. Insulin 1 humanized C-peptide recombineering strategy. Human C-peptide and K/N<sup>R</sup> flanked by 50bp of 1B chain and 50 bp of the 3’UTR comprised the recombineering vector used to modify the wild type mouse insulin sequence.
**Figure 12. The C-peptide of Bins1 is humanized.** BhCp1 by PCR A Ins1FB to HCPR to give a 280bp band, B Ins1FB to M214SR to give a 418bp band. [- = Control mouse BAC DNA]
Figure 13. **Targeting vector retrieval strategy.** Example of vector retrieval strategy with pGRDTIns1. Blue boxes represent regions of homology that were used for recombineering-based vector retrieval.
Figure 14. *Ins1* humanized C-peptide targeting vector was retrieved with 6.4kb 5’ and 5.4kb homology arms. A TVhCp1 retrieved from BhCp1 into pGRDTIns1 and B screened by EcoR1 (5.1kb, 3.9kb, 3.2kb, 2.0kb, 1.4kb, 0.9kb, 0.2kb) and EcoRV (4.7kb, 3.8kb, 3.5kb, 2.8kb, 1.6kb, 1.1kb, 0.18 0.16 kb) digests demonstrating that the TVhCp1 was retrieved correctly.
Figure 15. *Ins2* humanized C-peptide recombineering strategy. The recombineering homology arms were used to insert an Ampicillin resistance cassette to remove the *Ins2* intron and C-peptide. The recombineering vector hCp2 was then used to humanize the C-peptide sequence.
Figure 16. **Intermediate vector BAIns2.** A Amplified Ins2Amp recombineering cassette used to replace the intron and C-peptide of *Ins2* within BIns2. B An example of a BAIns2 clone screened by PCR (Ins2FB-Ins2RB) demonstrating the Ampicillin cassette integrated into Bins2 (1.54kb).
Figure 17. Insulin 2 humanized C-peptide targeting vector was retrieved with 6.4kb 5' and 5.4kb 3' homology arms. A TVhCp2 retrieved from BhCp2 into pGRDTIns2 and B screened by EcoR1 (8.1kb, 5.0kb, 2.6kb, 2.2kb, 0.2kb) and Sphl (10.8kb, 4.2kb, 1.6kb, 1.1kb, 0.45kb) digests.
c) Gene targeting in ESC

Each of the *Ins1* and *Ins2* humanized C-peptide targeting vectors were linearized and electroporated into C2 ES cells and exposed to geneticin selection (150 µg/ml). Successfully recombined cells bearing the Neo<sup>R</sup> cassette were expected to survive. Resistant clones were picked and screened by a PCR strategy specific to the vector. The positive clones were screened by Southern blots using probes internal and external to the targeted locus.

*Ins1* targeting (Figure 18A) was repeated six times. Due to inaccurate screening, no correctly targeted clones were identified out of 305 screened. Chen He completed the sixth targeting experiment, and screening by Southern Blot where four clones (1F, 2E, 5F, 10F) were targeted correctly out of 80 screened according to Southern analysis. The targeting efficiency of TVhCp1 was 1% (Table 3).

*Ins2* targeting (Figure 19) was completed two times. Two clones (1A, 9F) out of a total of 104 genticin resistant colonies were PCR-confirmed positive (Figure 20). Chen He screened an additional 43 colonies, two (7B, 7C) (Figure 21) were targeted correctly according to Southern screen. The targeting efficiency of TVhCp2 was 2% (Table 3).
Figure 18. Gene targeting of the mouse *Ins1* gene to humanize the C-peptide. A Gene targeting and Southern analysis schematic. B Screening candidate clones by neo probe and C genomic probe revealed 4 correctly targeted clones (1F, 2E, 5F, 10F). *Southern Blots (B, C)* courtesy of Dr. Chen He.
Figure 19. Targeting strategy to humanize the mouse Ins2. The humanized C-peptide 2 targeting vector (TVhCp2) with 6.4kb 5’ and 5.4kb 3’ homology arms was used for targeted insertion of the human C-peptide sequence and a neo resistance cassette (previously K/NR). PCR screening from the NeoR to outside the 3’ homology arm (5.7kb band) was used for the first two targeting rounds. Southern Blot was performed using an external probe and Nsi (N) and EcoRV (E) (WT: 15kb; Mutant: 11.5kb) digest.
Figure 20. Embedded PCR (K/NRSR3, Ins2GenoR1; K/NRSR2, Ins2GenoR2) distinguished two (1A and 9F) *Ins2* targeted clones (5.7kb) out of 104 G418 resistant clones. [+= BhCp2, - = C57BL/6]
Figure 21. Mouse *Ins2* human C-peptide ESC by Southern analysis reveals three correctly targeted clones. Clone 1A (by SMLN), 7B, 7C (by Dr. Chen He) reveal both WT: 15kb and HCP2: 11.5kb bands. *Figure courtesy of Dr. Chen He.*
Table 3. Efficiency of targeting vectors TVhCp1 and TVhCp2.

<table>
<thead>
<tr>
<th>Vector</th>
<th>G418 resistant</th>
<th>Screened</th>
<th>Positive by PCR screening*</th>
<th>Positive by internal probe Southern**</th>
<th>Positive by genomic probe Southern**</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVhCp1</td>
<td>385</td>
<td>385</td>
<td>9 (all negative by Southern)</td>
<td>4</td>
<td>4</td>
<td>1%</td>
</tr>
<tr>
<td>TVhCp2</td>
<td>147</td>
<td>147</td>
<td>2</td>
<td>3</td>
<td></td>
<td>2%</td>
</tr>
</tbody>
</table>

*Performed by SMLN

**Performed by Dr. Chen He
d) Generating transgenic mice

Humanized C-peptide mouse *Ins1* clones 2E and 10F were used in a diploid aggregation with outbred ICR embryos at the morula stage to generate chimeras. 15 chimeras were weaned, 13 were male (Table 4). 8 male mice with >50% chimerism were crossed with ICR females to determine germ line transmitters. Unfortunately, no germ-line transmitters have been identified at this time.

**Table 4. Insulin 1 human C-peptide chimeras generated from clones 2E and 10F.** *Table modified, courtesy of Toronto Centre for Phenogenomics.*

<table>
<thead>
<tr>
<th>Clone ID</th>
<th># Embryos aggregated</th>
<th># Pups born</th>
<th># Chimeras weaned</th>
<th>Percent chimerism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>HCP1 2E</td>
<td>112</td>
<td>11</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>HCP1 10F</td>
<td>105</td>
<td>22</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

Humanized C-peptide mouse *Ins2* clones 1A and 7C were used in diploid aggregation with outbred ICR embryos at the morula stage to generate chimeras. 23 chimeras were weaned and mice with >50% chimerism were bred to determine a germ line transmitter (Table 5). Four highly chimeric animals transmitted to germline, including two mice derived from each of clone 1A (100%, 70% chimera) and clone 7C (90%, 70% chimera). Representative PCR screening and Southern analysis is shown in Figure 22.
Table 5. Insulin 2 human C-peptide chimeras generated from clones 7C and 1A. Table modified, courtesy of Toronto Centre for Phenogenomics.

<table>
<thead>
<tr>
<th>Clone ID</th>
<th># Embryos aggregated</th>
<th># Pups born</th>
<th># Chimeras weaned</th>
<th>Percent chimerism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>HCP2 1A</td>
<td>105</td>
<td>22</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>HCP2 7C</td>
<td>127</td>
<td>22</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 22. *Ins2* human C-peptide F1 litter derived from clone 1A, screened by PCR and Southern. Four F1 generation mice bore the HCP2 allele. Here one such litter is shown, screened by **A** PCR screening of K/NR (368bp) and **B** PCR of KN/R -HCP2 junction (349bp). **C** Southern analysis shows four F1 agouti mice (3,6,11,12) carry the HCP2 allele (WT: 15kb; mutant: 11.5kb). *Figure, modified, courtesy of Dr. Chen He.*
5) Discussion

a) Tracing the expression of insulin with new mouse models by knocking in a humanized C-peptide into the mouse Insulin 1 and 2 genes.

Murine models are often used to understand the pathophysiology of diabetes, as well as to develop possible therapies. However, numerous caveats exist in these systems including, the heterogeneity of the human disease compared to that of the mouse, correlating scale and dosage of therapeutics from mouse to human (Ido 1997; Herrath & Nepom 2009; A. J. F. King 2012), and the fact that the mouse has two functional Ins genes that could contribute to differences in mouse and human physiology. The mouse Ins genes evolved separately and are under differing regulatory controls on different chromosomes, compared to humans who have a single insulin gene. This would imply that the mouse responds to deregulated glycaemia in a fundamentally different way to humans. If we are unable to distinguish the expression of these two insulin genes in a mouse model, pathophysiology and treatments discovered in mice may not be translatable to humans. Furthermore, methods used to induce diabetes in mice are time consuming and difficult, such that full onset of diabetes may not occur. In this case, a mouse may continue to produce insulin or may recover following STZ induction (Ohtomo et al. 1996; Shoelson et al. 1992; Grossman et al. 2010). This limits the current mouse models as a recipient of insulin-producing cells in cell grafting experiments. To solve these issues, I initiated development of a new animal model that allows tracking of the two insulin alleles, as well as allows differentiation of graft and host insulin production for models using allograft transplantation.
b) C-peptide was chosen as a biomarker

In the current study, the two mouse C-peptides were replaced with the human C-peptide to make a mouse-human hybrid proinsulin protein. The C-peptide was chosen as a marker based on its specific biochemical properties. The C-peptide is a ~30 amino acid chain within proinsulin that connects the insulin A and B chains. During proteolytic processing to produce the mature insulin, the C-peptide is cleaved and released in equimolar ratio to insulin. Due to its longer half-life in circulation compared to mature insulin, the C-peptide is commonly used as an indicator of insulin production (Bucchini et al. 1986; Polonsky et al. 1986; Selden et al. 1986). Furthermore, because human and mouse C-peptides have evolved independently, the amino acid sequences vary sufficiently between the two species to allow them to be distinguished using species-specific antibodies. Although human and mouse C-peptide amino acid sequence are not identical, the tertiary structure of the human sequence would be expected to retain functionality in mouse proinsulin. Both human and mouse C-peptides share conserved acidic residues at B-chain/C-peptide junction (human: RR/EAED; mouse: RR/EVED) and conserved basic residues at the C-peptide/A-chain junction (human: LQ/KR; mouse Ins1 C-peptide: RQ/KR, mouse Ins2 C-peptide: QQ/KR) (Steiner et al. 1985). The mouse conversion endoproteases PC1/3 and PC2 can recognize and cleave either human or mouse cleavage sites (Smukler et al. 2011; Bucchini et al. 1986; Zhou et al. 2008). Thus, it is unlikely that processing of human C-peptide by those conversion endoproteases could be interrupted in the hybrid mouse/human proinsulin. Moreover, the N-terminal acidic residues of C-peptide (human: EAED; mouse: EVED) that play a role during proinsulin folding (Nagy 2003; Chen et al. 2002) are maintained in human and mouse peptides. Introducing the human C-peptide sequence did not change the biochemical characteristic of this peptide so it is not expected to interfere with the folding of mouse proinsulin. This is further supported by the observation that transgenic mice carrying human insulin secrete human C-peptide (Sharan et al. 2009; X. Zhu et al. 2002; Bucchini et al. 1986). Therefore, the human C-peptide is expected to be interchangeable with that of the mouse and retain its functionality in proinsulin processing and usefulness as a biomarker.
Studies have suggested the human C-peptide consensus sequence EGSLQ confers some biological function (Yang et al. 2009; Chinwalla et al. 2002; Rigler et al. 1999; Buchholz et al. 1998; Schlake & Bode 1994) including improving vascular and neural health (Wicksteed et al. 2007; Ido 1997) or regulating renal filtration (Tillmar 2001; Ohtomo et al. 1996). Constitutively expressing human C-peptide in mice, whose sequence differs ($Ins1$: EVARQ, $Ins2$: EVAQQ), may influence these physiological functions. In previous human insulin knock-in studies, neurovascular and renal functions were not directly studied (Wicksteed 2001; Bucchini et al. 1986; Selden et al. 1986), so it may be interesting to assess this here.

Through humanizing the mouse C-peptide, it will be possible to distinguish the expression of two different mouse insulin proteins. This tool is expected to allow tracing of the two insulin alleles in vivo by comparing relative mouse and human C-peptide when only one allele is humanized. Additionally mouse-to-mouse islet, β-cell, or β-cell precursor (Steiner et al. 1985; Smukler et al. 2011; Zhou et al. 2008) transplantation can be improved because graft and host insulin production could be distinguished when either the donor or recipient is homozygous for the humanized C-peptide at both insulin alleles.

c) A genetic approach to express human C-peptide as an endogenous protein in mice

In order to generate the humanized C-peptide mouse lines, standard methods of generating a genetically altered mouse line were undertaken (Clancy 2008; Nagy 2003). A recombineering vector was generated with the modified element of interest and a resistance cassette, and was used to modify a BAC. A final targeting vector was retrieved from the modified BAC, consisting of gene-targeting homology arms, the modified DNA element and selection cassettes. When targeted in ESC, the vector facilitates homologous recombination with the ESC genome, resulting in insertion of the DNA element of interest. ESC clones that were successfully targeted were used for chimera generation. Chimeras can then be bred to identify animals in which the
modified ES cells have contributed to the germ cells, and therefore to propagate the modification in its offspring.

To this end, I successfully designed and targeted the mouse $Ins1$ and $Ins2$ alleles with a humanized C-peptide vector. In the recombineering vectors, human C-peptide seamlessly replaces the mouse $Ins1$ and $Ins2$ C-peptide respectively. Following insertion of a resistance cassette into the 3’UTR ($Ins1$) or the intron ($Ins2$) (Figure 9), the recombineering vector was used to modify BACs by homologous recombination ($Ins1$: Figure 12; $Ins2$: Figure 16). The humanized C-peptide gene-targeting vectors were retrieved directly from their respective modified BACs ($Ins1$: Figure 14; $Ins2$: Figure 17) and used to target the mouse genome in C2 ESC ($Ins1$: Figure 18; $Ins2$: Figure 21). Knock-in $Ins1$ and $Ins2$ human C-peptide chimeras have been generated ($Ins1$: Table 4; $Ins2$: Table 5). $Ins2$ human C-peptide has transmitted to the germline (Figure 22).

i) Design and generation of targeting vectors

The recombineering method was used in the current project to generate gene-targeting vectors, as it is an efficient and precise process. Traditional methods of generating gene-targeting vectors rely on the presence of unique restriction sites and in vitro manipulation of large fragments of DNA. Recombineering on the other hand relies on induced homologous recombination between a BAC and a recombineering vector that targets a gene of interest.

Recombineering is in vivo genetic engineering. It relies on the $\lambda$ bacteriophage derived proteins $gam$, $bet$ and $exo$ to facilitate homologous recombination between short (40-60bp) regions of homology. Due to the small homology region required for engineering, as short as 50bp flanking homology arms can be used to modify a DNA element of interest. The DNA to be modified is often a low copy plasmid, such as a BAC. BACs are particularly useful for generating
transgenic animals in that endogenous regulatory elements of a target gene can be found and retrieved within a single plasmid (McCarrick et al. 1993; Sharan et al. 2009; Skarnes et al. 2011). After modifying the gene of interest (e.g. human C-peptide) and selectable drug marker, the modified element and large targeting-vector homology arms can be retrieved by another round of homologous recombination using gap-repair plasmids. Taken together, the recombineering system provides an ideal method to generate gene-targeting vectors, by facilitating seamless engineering of the gene of interest and acquisition of ample targeting-region homology arms.

In the current study, the BACs contain C57BL/6 genomic DNA and our C2 ES cells are also from C57BL/6 mice. The C57BL/6 genome has been sequenced (Ohtomo et al. 1996; Chinwalla et al. 2002; Collier 1975), allowing for the BACs, ES cells and bioinformatics tools to share the same genomic information. Therefore, designing and evaluating outcomes following gene manipulation and targeting is readily possible.

The basic components of recombineering vectors were designed in silico and generated by gene synthesis. The most important component is the humanized C-peptide. Humanizing the C-peptide involved a straightforward exchange of the encoding mouse C-peptide sequence for the human variant.

Another essential component is a resistance cassette to facilitate clone selection. It was important to place K/N\(^R\) where it would have the least affect on insulin activity stability. The resistance cassette K/N\(^R\) consists of Kana\(^R\) and Neo\(^R\) flanked by frt-F3 sites. The frt-F3 sites are 34 bp sequences that are recognized by Flp recombinases (McCarrick et al. 1993; Buchholz et al. 1998; Schlake & Bode 1994). When two frt-F3 sites are in the same orientation in cis, expression of a Flp recombinase leads to deletion of the intervening sequence and one frt-F3 site. The remaining frt-F3 site may potentially interfere with splicing (Riele et al. 1992; Yang et al. 2009; Deng & Capecchi 1992) or endogenous regulatory elements at the DNA, mRNA or protein level. To minimize this risk, the K/N\(^R\) was inserted into the 3’UTR of the mouse Ins1
(Figure 4) and into the intron of the mouse Ins2 genes respectively (Figure 5). In general, the 5’UTR is involved in protein synthesis by facilitating ribosomal entry. With respect to the mouse proinsulin, the 5’UTR may also regulate glucose dependent translation (Bird & Bradshaw 1997; Wicksteed et al. 2007; Seong et al. 2004). The 3’UTR generally plays a role in mRNA stability and processing (Duvillié et al. 1997; Tillmar 2001; Micallef et al. 2007; Ben Yehudah et al. 2005; Austin et al. 2004; Lamotte et al. 2004; Fan et al. 2009). For proinsulin mRNA translation, the 3’UTR acts synergistically with the 5’UTR to regulate insulin protein production in a β-cell specific manner (Duvillié et al. 1997; Wicksteed 2001), and may have a role in suppressing insulin biogenesis. The 12 bp consensus sequence of this potential regulatory 3’UTR occurs following the insulin stop codon and before the polyadenylation sequence. To limit effects of the remaining frt-F3 sequence, the K/N<sup>R</sup> was placed after TAA Ins1 stop codon, but before this potential regulatory element. The K/N<sup>R</sup> was inserted into Ins2 intron between exon 2 and exon 3. The intron sequences in mammals are not conserved and vary considerably in length. However, the locations of the introns of human Ins and mouse Ins2 are conserved (Ben Yehudah et al. 2005; Steiner et al. 1985; Micallef et al. 2007). At the terminal ends of intron 2 are the characteristic GU-AG splice donor-acceptor sequence, with the pyrimidine tract at the 3’ of the intron and an adenosine branch-point residue in the interior (Shirohzu 2004; Clancy 2008). The K/N<sup>R</sup> was inserted into the intron of Ins2 following the GU splice donor sequence, and prior to any within the intron adenosine residue. In light of the fact that the Ins2 intron 2 is widely variable among species, it was predicted that an frt-F3 site in this location would not impact insulin mRNA stability. In summary, the humanized C-peptide recombineering cassettes were designed by incorporating the positive selection cassette into regions that are least likely to impact insulin biogenesis.

The third component in the recombineering vector that delivered human C-peptide and the resistance cassette to the insulin gene in the BAC is the short recombineering homology arms. They were designed to flank the above two elements - the human C-peptide and K/N<sup>R</sup>. The recombineering homology arms of the humanized Ins1 C-peptide cassette included 50bp of the mouse 1B chain 3’ of the C-peptide and 50 bp of the 3’UTR 5’ of the insertion site of the resistance cassette (Figure 4). To generate the Ins2 recombineering cassette, the entire 2B and 2A chains were incorporated. This consisted of more than the necessary 50 bp required for
recombineering (Figure 5). Due to the intron exceeding the necessary length for recombineering, I anticipated challenges humanizing both regions of the C-peptide located in the two exons. The longer homology arms were designed to increase the efficiency of recombineering. Ultimately, the longer homology arms were not used and instead the strategy outlined in Figure 15 was undertaken to humanize the mouse Ins2 sequence.

An additional selection step was necessary for successful recombineering with Ins2 (Figure 15) as the intron of Ins2 caused unwanted recombination to occur (Figure 23). This culprit was sidestepped by removing the intron and C-peptide sequence. When generating knockout mouse strains or inserting point mutations into a sequence by recombineering, positive-negative selection cassettes are often used (Dolgin 2011; McCarrick et al. 1993; Skarnes et al. 2011). This concept involves inserting a cassette by recombineering that contains a positive selection marker, meaning it confers an antibiotic resistance and it is selected for after insertion. The cassette is then replaced by the modified DNA element of interest. The dual selection cassette is then used as negative selection, where it becomes toxic when challenged by the appropriate drug. A Zeocin resistance-p-chlorophenalanine (ZR/PN) cassette flanked by 50bp of the 2B chain and 2A chain respectively was used initially to try to remove the intron. Following recombineering, zeocin was used to select for insertion of the cassette. The hCp2 cassette was then used, and p-chlorophenalanine was used as counter selection to ensure that hCp2 replaced ZR/PN. Unfortunately, this cassette consistently resulted in a high number of false positives, despite titration of the selectable markers and trouble-shooting. In an attempt to resolve this culprit, a similar concept was applied using a different cassette that took advantage of positive selection. A robust ampicillin resistance cassette was amplified using the hCp2 recombineering homology arms. This Amp2 cassette was used in recombineering to replace the mouse C-peptide and the intron. β-lactamase, the protein that confers resistance to Ampicillin is secreted by gram-negative bacteria such as E. Coli. Therefore, selection could not take place in liquid culture. When recombineering products were plated on Ampicillin selective agar, Amp2 had recombined in all of the screened colonies. These BAIIns2 BACs were then used in high-throughput liquid culture recombineering with hCp2 and gap repair.
Following the generation of the humanized C-peptide BACs, the homology arms for the gene targeting constructs of the \textit{Ins1} and \textit{Ins2} alleles were retrieved by gap-repair. Gap repair is a concept from cloning in \textit{Saccharomyces cerevisiae}. A linear dsDNA with ends that are homologous to the DNA to be retrieved is transfected into the cell. Recombineering facilitates homologous recombination between these regions of homology, the gap in the linearized vector is “repaired” and the DNA of interest is inserted into the plasmid. The gap-repair plasmid with the targeting homology arms and modified DNA element can then be linearized and used as a gene-targeting vector with the gap repair elements located in a small 4kb region at the 5’ or 3’ of the targeting vector. The gap repair plasmid used in this procedure was pGRDT, a vector conferring ampicillin resistance and containing a DTA cassette. The Amp\textsuperscript{R} was used during the gap repair recombineering process in E. Coli. The DTA toxin is expressed from the constitutively active polyoma enhancer/herpes simplex virus thymidine kinase (MC1) promoter is useful in gene targeting. Diphtheria is an A-B toxin derived from \textit{Corynebacterium diphtheria}, where the A chain is the catalytic domain that inhibits protein synthesis(Seong et al. 2004; Ohtomo et al. 1996; Fedorov et al. 1997; Collier 1975). When used in gene targeting in combination with a positive selection marker such as Neo\textsuperscript{R}, the combination provides positive-negative selection which greatly increases targeting efficiencies(Seong et al. 2004; McCarrick et al. 1993; Gertsenstein et al. 2010). The positive selection selects for clones that have integrated the modified DNA of interest. Though the long homology arms provide stretches of DNA for homologous recombination, there is no guarantee that targeted homologous recombination took place. If homologous recombination does take place, DTA is not integrated into the host genome. In the case of non-homologous integration however, the negative selection cassette will kill the cell.

Homology arms have a large impact on targeting efficiencies. It is generally accepted that longer stretches of isogenic DNA are superior in homologous recombination when compared to shorter homology regions(Leroux et al. 2001; Riele et al. 1992; Deng & Capecchi 1992), however, this is not the case for all targeting vectors or targeting loci(M. Liu, Hodish, Rhodes & Arvan 2007b; Bird & Bradshaw 1997; Seong et al. 2004). The insulin loci have been successfully targeted previously(Rigler et al. 1999; Duvillié et al. 1997; Micallef et al. 2007; Ben Yehudah et al. 2005; Austin et al. 2004; Lamotte et al. 2004; Fan et al. 2009). \textit{Ins1}
knockout targeting vector consisted of 2.5kb 5’ and 6.5kb 3’, while Ins2 knockout targeting vector consisted of 2.7kb 5’ and 5.5kb 3’ homology arms (Rigler et al. 1999; Duvillié et al. 1997; Ohtomo et al. 1996; Ido 1997). More recently, the international knockout mouse consortium generated targeted knock out in the C57BL/6N background with 5.6kb 5’ and 4kb 3’ Ins1 (KOMP ID: EM:01219), and 4kb 6.2kb 5’ and 3.8kb 3’ Ins2 homology arms (EUCOMM ID: 30841). Other targeting vectors are consistent with these homology arms (Seong et al. 2004; Ben Yehudah et al. 2005; Gertsenstein et al. 2010; Micallef et al. 2007), one arm greater than or equal to 5kb and the other being less than 5kb. To generate the humanized C-peptide targeting vectors, I used the KOMP and EUCOMM genotyping primers as homology regions for targeting vector retrieval. In appreciating the repetitive nature of the insulin loci, and the common difficulty of targeting in C57BL/6 cell lines, I slightly increased the length of the published homology arms.
**Figure 23. Incomplete humanizing of the C-peptide with the hCp2 recombineering cassette.** Homology within the intron of *Ins2* and the recombineering cassette led to incomplete recombination. The C-peptide of the third exon was not the human sequence so alternative methods were used (Figure 15).
d) Gene targeting efficiencies at the mouse *Insulin 1* and *Insulin 2* loci

Gene targeting of the insulin gene with TVhCp1 and TVhCp2 resulted in low targeting efficiencies (Table 3). 385 *Ins1* and 153 *Ins2* clones were G418 resistant indicating that their genome contained the Neo<sup>R</sup> cassette of the humanized C-peptide targeting vectors. Southern screening using an internal probe indicated 4 putative clones, with 4 correctly targeted according to Southern analysis with an external genomic probe (Data: Chen He, Figure 18). Southern analysis by external genomic probe of *Ins2* humanized C-peptide clones indicated that 3 had undergone correct homologous recombination (Data: Chen He, Figure 21).

A low targeting efficiency of the humanized C-peptide targeting constructs could be due to factors including the targeting vector itself and the target loci. As mentioned previously, the targeting vectors were designed with longer homology arm than had been previously used for this locus. Normally, targeting vectors consist of a short (<5kb) homology arm to allow for PCR screening and a longer arm, with the two arms totaling approximately 10kb. Here, short arms consisted of approximately 5kb and the long arm >6kb. This caused significant difficulties of screening G418 resistant colonies by PCR (Figure 24), leading to the first 308 apparently negative *Ins1* clones. Southern screening in a 96-well format was undertaken with the subsequent 80 clones, 4 of which were correctly targeted.

The regions surrounding *Ins1* and 2 are highly repetitive in sequence and structure. In fact, as the sequences move further from the insulin exons, the sequence becomes more repetitive. The sequence between *Th* and *Ins2* on chromosome 7 area consists of a 210kb made up entirely of tandem repeats and retro-elements (Leroux et al. 2001; Shirohzu 2004). These are considered to have played a role in *Ins2* retrotransposition. Due to the highly repetitive nature of this
sequence, it is possible that adding extra length to the homology arms hindered targeting performance by targeting more repetitive sequences.
Figure 24. Primary PCR of humanized C-peptide *Ins1* did not conclusively indicate any correctly targeted clones. Green boxes indicate clones that were used in the secondary PCR while the black box was used as an internal negative control in the secondary PCR.
e) Generating chimeras

The humanized C-peptide mouse lines were generated in the C57BL/6 background. C57BL/6 mice are relatively fecund, generating large numbers of progeny at quick intervals. As an inbred strain, they are used for genetic studies to dissociate mutations from outbred phenotypes. With IKMC near completion (M. Liu, Hodish, Rhodes & Arvan 2007a; Dolgin 2011), the availability of thousands of knockouts is further reason for choosing this line to generate a new mouse tool.

There are three common routes for the generation of chimeric mice; 1) injection of ES cells into the blastocoel cavity of blastocyst stage embryos, 2) ES cell injection under the zona pellucida of eight-cell stage embryos, and 3) aggregation of ES cells with eight-cell stage embryos, the zona of which have been removed. Of these alternatives, blastocyst injection is the most widely used technique. However, it not only requires critical conditions to be met such as qualified technical skills and inbred donor embryos but also the chimeric embryos need to be transferred immediately. Aggregation is less technically demanding but results in a similar efficiency of germline transmitting chimeric mice (Nagy et al. 2010). Therefore, we chose to use this method to generate chimeras from the two clones of each Ins1 and Ins2 humanized C-peptide. In 129 derived chimeras, a high bias towards the male sex and high chimerism is a strong indication that germline transmission would be achieved (Rigler et al. 1999; Seong et al. 2004; Fedorov et al. 1997). However, in C57BL/6 derived chimeras, the sex distribution and level of chimerism is much less indicative of the chance of germline transmission from these animals (Rigler et al. 1999; Seong et al. 2004; Ohtomo et al. 1996; Gertsenstein et al. 2010; Ido 1997). With respect to Ins2 humanized C-peptide chimeras, highly chimeric animals (70%, 70%, 90%, 100%) transmitted to the germline. It is still to be determined if Ins1 humanized C-peptide chimeras will transmit to the germline.
f) Future directions to characterize humanized C-peptide *Ins1* and *Ins2* mouse models.

The humanized C-peptide mouse line must be fully characterized in order for it to be useful as a research tool. First, it will be bred to homozygosity and/or to the reciprocal knockout animal. To characterize the mouse, I propose to monitor its overall health, then subject the line to metabolic challenge. Following characterization of the metabolic phenotype of healthy and high-fat diet challenged mice, the humanized C-peptide can be a tool to dissect the stoichiometry of the *Ins* allele expression and secretion pattern, and as a tool to monitor β-cell function.

i) Breeding schematic to generate homozygous humanized C-peptide mice

The knock-in K/N<sup>R</sup> may affect insulin mRNA stability. Therefore, germline-transmitting chimeras will be bred to C57BL/6N-Tg(CAG-Flpo)1Afst/Mmucd (FlpO) mice to excise the resistance cassette (Figure 25A). For the purpose of routine genotyping, the mice can be genotyped by PCR amplification of the insulin - human C-peptide junction or remaining frt-F3 site. Humanized C-peptide mice will then be crossed to wild type C57BL/6 to transmit the modified allele, and to siblings bearing human C-peptide to generate homozygous humanized C-peptide mice. A homozygous human C-peptide line will be crossed to the reciprocal homozygous humanized C-peptide line (Figure 25C). As only one humanized C-peptide transmitted at a time, the line could be crossed to the reciprocal knock-out animal (Figure 25B), because the *Ins1* and *Ins2* can compensate for each other(Jeon et al. 2012; Leroux et al. 2001; Sneddon et al. 2013).
Figure 25. Sample breeding schematic. A hCp1-K/N^R will be bred to Flp-O to excise the resistance cassette. Following excision of the resistance cassette, siblings can be mated to generate homozygosity at one allele (e.g. Ins1). Homozygous human C-peptide (hCp) mice can be crossed to B the reciprocal Ins knock out.
ii) Monitoring health and metabolic parameters

Following generation of a humanized C-peptide line that is homozygous at both alleles or one allele and a knockout at the reciprocal location, the health of the mice will be closely monitored. This monitoring includes weight, longevity, reproductive patterns as well as non-fasting and fasting blood glucose and insulin levels. Standard ELISA to the human C-peptide can quantify the levels of human C-peptide from the serum, and immunohistochemistry can identify the expression of the human C-peptide in sectioned pancreas or isolated islets. Islet preproinsulin, proinsulin and insulin mRNA and protein levels will be measured. Animals will be challenged by glucose and insulin tolerance tests. As suggested previously, the humanized C-peptide is not expected to cause any change in insulin processing or secretion compared to wild type animals. In the case where glucose homeostasis has been altered, it would be important to dissect out whether or not insulin expression or protein synthesis has been changed in the knocked-in mice. As it is likely to be a cell autonomous effect, this change can be investigated directly by examining the existence of insulin mRNA, protein levels, and secretion in β-cells. Noticeably, misfolding of insulin can lead to endoplasmic reticulum stress and cytotoxicity (M. Liu, Hodish, Rhodes & Arvan 2007b). However, the misfolding events cited here are due to mutations found within the A or B chains of insulin and unlikely to occur in the humanized C-peptide model.

The C-peptide has been observed to have biological activity with a cognate receptor (Rigler et al. 1999). The human C-peptide may not bind to this receptor in mice and this could have consequences on neural, endothelial, or renal function (Rigler et al. 1999; Ohtomo et al. 1996; Ido 1997). In the absence of gross abnormalities in the kidney, brain and cardiovascular system, specific physiological functions of these organs (glomerular filtration rate, vascular tone or neural excitability) could be tested. Alternatively, the humanized C-peptide could be crossed to models with genetic susceptibility to renal or neurovascular diseases.
iii) Investigating relative insulin allele expression

To study the relative expression of insulin from the *Ins1* and *Ins2* genes, mice will need to be homozygous for one humanized C-peptide while the reciprocal gene remains wild type. To draw proper conclusions, it would be important that both alleles are humanized and compared to the reciprocal wild type allele. Differences in allele expression and secretion may be accounted for by variations in mouse and human C-peptide biodegradation. In parallel, mRNA and protein C-peptide levels would be measured to verify the expected differences in insulin transcription and translation. Immunohistochemistry and fluorescently activated cell sorting could further support this information. Fluorescently activated cell sorting would be able to discern the distribution of human:mouse/Ins1:Ins2 levels in β-cells based on commercially available kits. Each β-cell would be expected to contain both C-peptides, so double positive cells would be observed. However, in response to metabolic challenges, the distribution of double positives may change, for example from high human C-peptide, low mouse C-peptide expression. This would indicate a shift in allele translation that may or may not be reciprocated in secreted insulin. Secreted insulin would be measured by circulating C-peptide levels that will be measured by ELISA. After establishing the expression and secretion patterns in wild type adult animals, the same parameters could be monitored through development or under metabolic stress. Understanding the stoichiometry of insulin expression may be important for understanding the translatability of the mouse model to human disease.

iv) Investigating the functionality of *in vitro* modified insulin producing cells

This mouse would be an excellent recipient for mouse derived β-cells or insulin producing progenitor cells that have been expanded or manipulated *in vitro*. Here, the host human C-peptide mouse would be treated with STZ to induce diabetes, and then *in vitro* differentiated cells (Jeon et al. 2012; Sneddon et al. 2013) could be transplanted. The ability of the transplanted
cells to regulate glucose could be monitored by measuring mouse C-peptide levels against human levels.

This concept could be applied further to discern the cell autonomous response of β-cells in response to differing environments. For example, the IKMC is nearing its completion, which has made thousands of knockout mouse strains readily available. If a metabolic phenotype is observed in a knockout and it is suspected that the environment is contributing to this observation, islets could be isolated and transplanted into healthy or diabetic humanized C-peptide mice. If the environment is indeed contributing to this phenotype, then following transplantation minimal phenotype and relatively equal mouse: human C-peptide levels should be observed in the recipients. However, if this is a cell autonomous defect in insulin function, then aberrant mouse C-peptide will be observed in the recipient. The reciprocal would also hold true: if an environment is contributing to a metabolic phenotype, the transplantation of humanized C-peptide islet should have limited effect on the phenotype. However, if the endogenous β-cells are dysfunctional, then transplantation of wild type humanized C-peptide islets should rescue the phenotype. Any rescue could be confirmed by monitoring the human to mouse C-peptide levels. This may be a valuable tool in studies investigation insulin resistance.

**g) Limitations of the humanized C-peptide model**

The humanized C-peptide is an excellent tool that will provide researchers a means of addressing questions in ways that were not previously possible. This being said, there are a few limitations inherent in the mouse line and the concept of its utility. The C-peptide must be fully homozygous to be able to discern insulin production between the two genes or those expressed from a cell graft versus host.
A major drive in creating this mouse line was for its use in transplantation studies. The human C-peptide is distinguishable from the mouse C-peptide. Therefore, it is possible that following transplantation among isogenic C57BL/6 mouse lines, the human or mouse C-peptide could cause an immune reaction, but this is not expected as discussed previously.
6) Summary and Conclusions

The purpose of this project was to generate a tool that would allow differentiation of the expression of two insulin genes in the mouse. Discriminating insulin expression from *Ins1* and *Ins2* will be important in the translation potential of mouse metabolic studies. Additionally, it will be useful in mouse allograft transplantation studies. Such mouse lines would also be appreciated in variety of research involved in islet development, β-cell physiology, and β-cell progenitor differentiation.

To develop this tool, I have undertaken the humanization of the C-peptide in the C57BL/6 mouse strain. I have created two gene-targeting vectors to target the mouse *Ins1* and *Ins2* genes and have generated chimeras carrying the targeted mouse *Ins1* and *Ins2* gene, where the human C-peptide tagged *Ins2* allele has contributed to the germline. In order to generate such mouse lines and subsequently to dissect their biological characteristics, five aims in this project were initially proposed and the first four have been achieved. In summary, the necessary materials to establish an *Ins1* and *Ins2* humanized C-peptide mouse line has been created. This line can provide foundation for future metabolic studies in mice, making it a valuable resource for biomedical research.
References


Crowe, D.T. & Tsai, M.J., 1989. Mutagenesis of the rat insulin II 5'-flanking region defines sequences important for expression in HIT cells. *Molecular and cellular biology*.


Grill, V. & Cerasi, E., 1974. Stimulation by d-Glucose of Cyclic Adenosine 3" : 5-
"Monophosphate Accumulation and Insulin Release in Isolated Pancreatic Islets of the
Rat. *Journal of Biological Chemistry*.

Grossman, E.J. et al., 2010. Glycemic control promotes pancreatic beta-cell regeneration in


Guzman, L.M. et al., 1995. Tight regulation, modulation, and high-level expression by vectors
containing the arabinose PBAD promoter.


Hanson, K.D. & Sedivy, J.M., 1995. Analysis of biological selections for high-efficiency gene

distal end of the short arm of chromosome 11. *Proceedings of the National Academy of
Sciences of the United States of America*, 78(7), pp.4458–4460.


Hentze, H. et al., 2009. Teratoma formation by human embryonic stem cells: Evaluation of


Hong, E.G. et al., 2007. Nonobese, insulin-deficient Ins2Akita mice develop type 2 diabetes
phenotypes including insulin resistance and cardiac remodeling. *AJP: Endocrinology
and Metabolism*, 293(6), pp.E1687–E1696.

Structural maturation probed by disulfide accessibility. *The Journal of biological
chemistry*, 270(35), pp.20417–20423.


Inada, A. et al., 1999. The cyclic AMP response element modulator family regulates the insulin
gene transcription by interacting with transcription factor IID. *The Journal of biological
chemistry*, 274(30), pp.21095–21103.

Inada, A. et al., 1998. Transcriptional repressors are increased in pancreatic islets of type 2
diabetic rats. *Biochemical and biophysical research communications*, 253(3), pp.712–
718.

Inagaki, N. et al., 1992. c-Jun represses the human insulin promoter activity that depends on
multiple cAMP response elements. *Proceedings of the ....

International Mouse Knockout Consortium et al., 2007. A mouse for all reasons. *Cell*, 128(1),


Shapiro, A.M. et al., 2000. *Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen*.

Shapiro, A.M.J. et al., 2006. *International trial of the Edmonton protocol for islet transplantation*.


Xie, J. et al., 2003. Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. Proceedings of the ....


Zhu, B. et al., 2007. In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations,
Appendices

1 Southern Analysis Methods, Performed by Dr. Chen He

1. Digestion of genomic DNA: 10µg of genomic DNA from each sample (ES cells or mouse tails) were digested with 50 units of restriction enzyme(s) at 37°C for 12-16 hours.

2. The digested DNA samples were resolved in a 0.7% Agarose gel (1XSYBR® Safe DNA Gel Stain mixed in the gel) with 1XTAE buffer, by 20V for 10-12 hours.

3. A digital image of the separated DNA was taken on a UV trans illuminator by the end of the electrophoresis.

4. The gel was soaked in 1 liter of 1XDepurination solution with gentle shaking for 15min or until the Bromophenol blue front turned yellow. The gel was rinsed with distilled water.

5. The gel was soaked in 1 liter of 1XDenaturation solution with gentle shaking for 30min and the Bromophenol blue front turned blue. The gel was rinsed with distilled water.

6. The gel was soaked in 1 liter of 1XDenaturation solution with gentle shaking for 30min. The gel was rinsed with distilled water.

7. The gel was soaked in 1 liter of 20XSSC solution with gentle shaking for 15min.

8. Setting up the transfer device: Amersham Hybond-N+ membrane was cut 0.5 cm larger than the gel dimensions. The membrane was pre-wetted with water followed by 20XSSC. A tray was filled with 1 liter of 20XSSC and a solid stand with a flat surface big enough to hold the gel was positioned over the tray, 1-2cm above the surface of the 20XSSC. A double layer of filter paper, the width of which matched the size of the gel and the length of which could reach the 20XSSC buffer, was soaked with 20XSSC and lay on the flat surface of the stand. Both ends of the filter paper emerged into the 20XSSC in the tray, forming a bridge. The side of the gel with the loading wells faced the filter paper bridge and lay down on top the filter paper. Air bubbles were driven out. The Hybond-N+ membrane was then laid on the smooth side of the gel tightly.
Air bubbles were cleaned out. 6 layers of filter paper were cut matching the dimensions of the gel. They were soaked with 20XSSC and laid layer by layer on the top of the Hybond-N+ membrane. Air bubbles were driven out between layers. Multiple layers of paper towels (up to 10cm) were laid on the top of the filter paper. A 250g weight was applied on the top of the pile of paper towels. The transfer device was wrapped with plastic wraps and the transfer carried on for 12-16 hours.

9. By the end of the transfer, the gel shrank to a thin layer. The paper towels and the filter paper were removed. The loading wells of the gel were marked on the Hybond-N+ membrane with a pencil. The first well was marked with an arrow sign.

10. A piece of filter paper of the same size of the Hybond-N+ membrane was wetted with 20XSSC. The Hybond-N+ membrane lay on the filter paper with the DNA side facing up. The membrane was subjected to UV crosslinking with a UV Stratalinker using 120,000 microjoules.

11. Following crosslinking, the membrane was washed in water for 1 min and air-dried for one hour.

12. The dried membrane could be stored in plastic hybridization bags or directly subjected to hybridization.

To perform hybridization with DIG labeled probes

1. 300ng of the probe DNA (a PCR product or a restriction enzyme digested, recovered fragment) was labeled with DIG-High Prime DNA Labeling and Detection Starter Kit II for 20 hours.

2. ExpressHyb™ Hybridization Solution was preheated at 65°C overnight. The temperature of the hybridization oven was set at 60°C. The dried membrane was soaked with 2XSSC and rolled into a hybridization tube. The side of membrane with DNA faced the lumen of the tube. The 2XSSC was removed, leaving no air bubbles between the membrane and the wall of the tube. The preheated hybridization solution (0.133ml/cm2 of membrane) was added into the tube. The speed of the rolling was set at 1.75. The prehybridization carried on for 2 hours.
3. To perform hybridization, the concentration of the DIG labeled probe was 25ng/ml in the hybridization buffer. Before use, the designated amount of probe was diluted with 50µl of DNase free water and denatured at 95°C for 5-10min and quickly chilled on ice for 2min. At 60°C, the probe was mixed with the hybridization buffer, the same amount of prehybridization buffer. The prehybridization buffer in the tube was replaced with the buffer with the probe. The hybridization carried on overnight.

4. By the end of hybridization, the membrane was washed in tube twice with the DIG-probe wash buffer at room temperature and twice in a tray on a fast shaking shaker for 30min each time.

5. The DIG-probe wash buffer at 50°C was preheated to 52°C. The membrane was then washed at 50°C for 20min, twice with fast shaking.

6. The membrane was then soaked in Maleic Acid buffer and equilibrated for 5min.

7. The 1XBlocking buffer was made from the 10X stock in the DIG-High Prime DNA Labeling and Detection Starter Kit. The blocking took 30-60min.

8. AP conjugated anti-DIG antibody was diluted in the blocking buffer 1:10,000. The membrane was blotted for 30min at room temperature.

9. The membrane was washed in Maleic Acid wash buffer for 15min, twice.

10. The membrane was equilibrated in pH9.5 Detection buffer for 2min.

11. The NBT/BCIP substrate solution was applied evenly over the DNA side of the membrane at room temperature. The membrane was incubated with the substrate for 5min. Extra solution was removed and membrane was sealed in a plastic hybridization bag without air bubbles. The DNA side of the membrane was exposed to a Amersham Hyperfilm ECL film for 20min. The film was developed with an X-ray developer.

To perform hybridization with P32 labeled probes.
1. The membrane was rolled in the hybridization tube, 0.15ml of Salmon sperm DNA (10mg/ml) was denatured at 95°C and chilled on ice for 2min. 15ml of the Church buffer for radioisotope labeled hybridization was pre-heated at 42°C. The denatured Salmon sperm DNA was added into the Church buffer. The pre-heated Church buffer replaced 2XSSC in the hybridization tube. The prehybridization carried on at 42°C for 2 hours.

2. 100ng of probe DNA in 45µl of water was denatured at 95°C for 10min and chilled on ice for 2min.

3. In the radioisotope room, 5µl of dCTP [α-32P] was mixed with the probe and the labeling cocktail from the Rediprime II DNA Labeling kit. The labeling reaction carried on at 37°C for one hour.

4. The labeled probe was separated from free dCTP [α-32P] by a ProbeQuant G-50 spin-column. The column was centrifuged at 2,000rpm to elute out the preservatives. The solution from the labeling reaction was applied to the top of the column and the column was centrifuged for 2min at 2,000rpm. The flow-through (probe) was collected and the free dCTP [α-32P] was trapped in the column. 1µl of the probe was mixed with 10mls of scintillation buffer in a scintillation vial. The radioactivity was counted by a Beckman Coulter LS 6500 scintillation counter.

5. The specific activity was calculated according to the kit instructions

6. The labeled probe was denatured at 95°C for 5min and chilled on ice for 2min. The probe was then added into the hybridization tube directly. The hybridization carried on overnight.

7. By the end of hybridization, the membrane was first washed in tube twice with P32-probe wash buffer at room temperature. The waste was collected in a liquid waste bottle. The membrane was then washed for 15min, three times in a tray on a shaker. The waste was collected in the liquid waste bottle.

8. The membrane was further washed at 55°C for 10min, at least three times. A Geiger counter was used frequently to check the radioactivity on the membrane (the probed area versus the area without DNA or the area supposed not to be probed).
9. The membrane was semi-dried and wrapped in plastic wrap. The membrane was exposed to an Amersham Molecular Dynamics Phosphor Scan screen for 24 hours. The screen was scanned with a BioRad BMI Molecular Imager.

UVP Visi-Blue Trans illuminator and BioDoc-It Imaging System
Stratagene UV Stratalinker 1800.
Beckman Coulter LS 6500 scintillation counter
ALLPRO 100 Plus DRIVE GEAR X-RAY PROCESSOR
BioRad BMI Molecular Imager
Amersham Molecular Dynamics Phosphor Scan screen
GE Health Amersham Hybond-N+
GE Health Rediprime II DNA Labeling System
Invitrogen SYBR® Safe DNA Gel Stain
Perkin Elmer dCTP, [α-32P]- 3000Ci/mmol 10mCi/ml EasyTide, 250μCi, 25μl.
DIG-High Prime DNA Labeling and Detection Starter Kit II (including: DIG-High Prime, 5x concentrated
DIG-labeled Control DNA (5 μg/ml), pBR328, DNA Dilution Buffer, Anti-digoxigenin-AP Conjugate, NBT/BCIP Stock Solution, concentrated Blocking Solution, 10x concentrated and DIG Easy Hyb Granules)
Clontech ExpressHyb™ Hybridization Solution
GE Health Amersham Hyperfilm ECL
GE Health ProbeQuant™ G-50 Micro Columns
1XDepurination solution (1 liter)
250 mM HCl, 20.75 ml in water

1XDenaturation solution (1 liter)
1.5 M NaCl, 87.7 g
0.5 M NaOH, 20 g

1XNeutralization solution (1 liter)
1.5 M NaCl, 87.7 ml
0.5 M Tris-HCl, pH 7.5, 78.8 ml

20x SSC (1 liter)
NaCl, 175.3 g
Na3C6H5O7·2H2O, 88.2 g

DIG-probe wash buffer at 50°C (1 liter)
20X SSC, 5 ml
20%SDS, 5 ml
DIG-probe wash buffer at room temperature (1 liter)

20X SSC, 100ml

20%SDS, 5ml

Maleic Acid buffer (0.1M, pH7.5) (1 liter)

Maleic Acid, 11.61g

5M NaCl, 30ml

Using 10M NaOH to adjust pH.

Maleic Acid wash buffer

0.3% Tween 20 in Maleic Acid buffer

Maleic Acid blocking buffer

Blocking Solution (10x concentrated) diluted in Maleic Acid buffer to 1X.

Detection buffer (1 liter)

1M Tris buffer pH9.5, 100ml

5M NaCl, 20ml

Church buffer for radioisotope labeled hybridization (250ml)

10% BSA, 25ml
Sodium phosphate buffer pH7.2, 125ml

0.5M EDTA pH8.0, 0.5ml

20%SDS, 87.5ml

Prehybridization and hybridization buffer for radioisotope labeled hybridization (15ml)

Church buffer, 10.5ml

Formamide (Fisher Bioreagents), 4.5ml

Denatured Salmon sperm DNA solution (Invitrogen), 0.15ml

P32-probe wash buffer at room temperature (1 liter)

20XSSC, 100ml

20%SDS, 25ml

P32-probe wash buffer at 55°C (1 liter)

20XSSC, 10ml

20%SDS, 25ml

Neomycin resistance gene hybridization probe for Southern

\textit{Ins2}, human C-peptide knocked-in

SSpI digestion, fragment size 9kb, detection of the insertion from the 5' region of mouse genomic DNA
DraIII digestion, fragment size 6.6kb, detection of the insertion from the 3’ region of mouse genomic DNA

*Ins1*, human C-peptide knocked-in

PmeI+Agel-HF digestion, fragment size 8.5kb, detection of the insertion from the 3’ region of mouse genomic DNA

ScaI+HindIII digestion, fragment size 9.6kb, detection of the insertion from the 5’ region of mouse genomic DNA

*Ins1* (HCP1) Gene-specific Southern probe

```
ATGAAAGAAATGGAAAAATTTGAAAAAGCAATGATCTAAGTCATAAGTGGCAAT
GAGTTTGACACAAAGAAAATGTTATTTTCCTATAATTATAATCATTTCAACAGCAC
CTGAAAAACGGTCAAACAAGGAAGATTGGACCAGGATGATTTTTGTTTTCAAGTTAAC
ACCGTACAATGAAAGGAATAATGTATTTTCTAATTGAAATGTAATGAAACAACA
AGAGAAAAATGGCTAATCATACTATATTGATAGTCCAATGCTGCTCCGATTCT
AAATACTTTAGGTGCAGCACAATTACTGCGCATGCAAATCACAAGCGATTCT
GTTGAGGTGTTAAAGACCTCAGAATTGTTGAGAAGCTAAACCTTCAGAGCCCTAA
CTCGCAAGAAAATCATAGTGGAGCCAGTTTCAGGTACAGTGACAGCATAATTCTAT
CCACCTATTTTAATAAGGGTCACAACCTTTCTTGTAGCCATCACCAGCA
```

PvuII digestion, WT size: 7.9kb; mutant size: 8.7kb

PCR primers for the *Ins13P* probe:

*Ins13P*probeFP (59.7)
5'- GCA ATG ATC TAA GTC ATA AAG TTG GCA ATG AGT TTG -3'

*Ins13PprobeRP* (63.8)

5'- TGC TGG GTG ATG GGC TAA AGG AAA GGT -3'

*Ins2* (HCP2) gene-specific 5' Southern probe

```
AAGAATTAACCTTGGAAGGAGTTATGCCCTACTGATAATCAGCAGAAGTTTTT
AGTCCCCAGAAAGACATCCAGAGAAGCTTTAAGAACCATAAAAGAACATGGGAAAGA
AATTAGGCTTCATTGAGTCCACGAGGTTTTTAGGCTCAATACCTGAATGCTTAAGCT
CCAGTGAGAACATGTAGAAGATTTTAGGATATGGAAGCTGGGAGATGCTTAAGA
CTAAGTGATACTTTTGGAGAAGCTTTGGGATCCCAAGGATATCTCTAAGATGGCCCA
CTAGGCCAAGTGAGAAAGTGCAAGATGTGTTGAGCCACAGTGAGAACATGGGAAA
AGCCTTCAGACCCAGTGAAGTTTCCAAGGGCTTTGGAATGGCTGTGGGAACCTGG
GAAAAGTTTTTAGGCAAGAATTTCTGGGTAAATGTCCTAGGTCTCCAGTGAGAAT
CTGGGCAAGGCTTTAAACAGACTGAGTACATGGGAAAGGCTTTAGAGGCCAGGAGACCTGT
```

Nsi+EcoRV digestion: WT size: 15kb, Mutant: 11.5kb

PCR primers for the *Ins2* (HCP2) gene-specific 5’ Southern probe:

*Ins2ProbeLane5FP*

5'- AAG AAT TAA AAA CCT TGG AAG GAG TTA TGC CCT AC -3'

*Ins2ProbeLane5RP*

5'- ACA GGT TCT CCC TGG GTT CTA AGG -3'