IDENTIFICATION, ISOLATION AND CHARACTERIZATION OF PROINSULIN PRODUCING THYMIC CELLS

MICHAEL O. PALUMBO

FACULTY OF MEDICINE, DEPARTMENT OF EXPERIMENTAL MEDICINE

MCGILL UNIVERSITY

AUGUST 2006

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENT OF THE DEGREE OF DOCTOR OF PHILOSOPHY

© Michael Palumbo

August 2006
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.
This thesis is dedicated to all patients suffering from an autoimmune disease, their families and to all those that support research efforts.

Disease brings suffering... time offers hope... knowledge provides a cure.
Abstract

The finding that more than 152 tissue-restricted antigens are expressed by thymic medullary epithelial cells is redefining the importance of thymic central tolerance induction in the prevention of autoimmune diseases. One of the tissue-restricted antigens in the thymus is proinsulin, and in both mice and humans, reduced thymic proinsulin levels have been shown to predispose to Type 1 diabetes. Using transgenic mice expressing a functional β-Galactosidase gene under the regulation of the Ins2 promoter we have determined that between 1-3% of all medullary thymic epithelial cells express proinsulin and that these cells are frequently part of the Hassall’s Corpuscles like structures in mice. Using a cross between the β-Galactosidase expressing mice and Immortomice (expressing SV40 large T Antigen under the regulation of the MHC I promoter), we have isolated and cultured two proinsulin and two non-proinsulin producing medullary epithelial cell lines. Microarray analysis and RT-PCR analysis of the cell lines revealed the over-expression of approximately 50 genes (>4 fold or more) in the proinsulin producing lineage, versus the non proinsulin producing lineage, and approximately half the over-expressed genes can be considered tissue-restricted antigens. We do not find any evidence for chromosomal clustering of the over-expressed genes nor do we report the expression of any other pancreatic β-cell antigens or specific pancreatic proinsulin regulatory proteins (Pdx-1, Glut-2 or GCK) within the proinsulin producing cell lines but we do detect their expression in whole thymus. Our results suggest that chromosomal clustering is not a phenomenon associated with thymic tissue-restricted antigen expression and that the mechanisms allowing for thymic tissue-restricted antigen
expression are not related to the expression mechanisms of such antigens in peripheral tissues.
Resume

The finding that more than 152 tissue-restricted antigens are expressed by thymic medullary epithelial cells is redefining the importance of thymic central tolerance induction in the prevention of autoimmune diseases. One of the tissue-restricted antigens in the thymus is proinsulin, and in both mice and humans, reduced thymic proinsulin levels have been shown to predispose to Type 1 diabetes. Using transgenic mice expressing a functional β-Galactosidase gene under the regulation of the Ins2 promoter we have determined that between 1-3% of all medullary thymic epithelial cells express proinsulin and that these cells are frequently part of the Hassall's Corpuscles like structures in mice. Using a cross between the β-Galactosidase expressing mice and Immortomice (expressing SV40 large T Antigen under the regulation of the MHC I promoter), we have isolated and cultured two proinsulin and two non-proinsulin producing medullary epithelial cell lines. Microarray analysis and RT-PCR analysis of the cell lines revealed the over-expression of approximately 50 genes (>4 fold or more) in the proinsulin producing lineage, versus the non proinsulin producing lineage, and approximately half the over-expressed genes can be considered tissue-restricted antigens. We do not find any evidence for chromosomal clustering of the over-expressed genes nor do we report the expression of any other pancreatic β-cell antigens or specific pancreatic proinsulin regulatory proteins (Pdx-1, Glut-2 or GCK) within the proinsulin producing cell lines but we do detect their expression in whole thymus. Our results suggest that chromosomal clustering is not a phenomenon associated with thymic tissue-restricted antigen expression and that the mechanisms allowing for thymic tissue-restricted antigen expression...
expression are not related to the expression mechanisms of such antigens in peripheral tissues.
Acknowledgements

First and foremost, I would like to thank Dr. Constantin Polychronakos for his supervision and guidance. Dr. Polychronakos is quite unique in that he is quite open-minded about student-life and he allows his students to enjoy and undertake other endeavors while simultaneously completing a rigorous research project. It is without a doubt that his support and understanding have allowed me to succeed not only in my PhD research but also at my medical studies. Certainly his guidance has and will likely continue to allow me to become a well-rounded clinician scientist.

I would also like to thank, Dr. Aziz Alami Chenoufi for forcing or at least trying to convince/force me to get to the lab early everyday, and for teaching me many of the techniques I initially required to complete an immunology related project, which previous to this I had no experience with at all.

Of course, I must obviously thank Rosemarie Grabs and Dina Levi for a number of reasons including: 1) putting up with my horrid jokes and 2) putting up with my immature behaviour...which for some reason only really seemed to come out when I would be around them. More importantly, though, is reason 3) their assistance which allowed me to complete a number of the chapters in this thesis – forever grateful for this ladies.

I must also thank the rest of my lab team (even those who are no longer there) and members of other labs that constantly provided me with assistance, including, Luc Marchand, Houria Ounissi-Benkalha, Marylène Rousseau, Xiaoyu Du, Huiqi Qu, Lu Yang, Yiota Papadopoulos, Cindy Wong, Adam Szymborski, François Bacot, Rosalie
Fréchette, Marie-Catherine Tessier, Gurvinder Kenth, Joy Osafo, Oriana Hoi-Yun and Wissam Shalish. Thanks for all your support ladies and gentlemen.

I would also like to thank Dr. Cindy Goodyer and the members of my committee, for although we did not meet, perhaps, as often as we should have, all the members, but especially Dr. C. Goodyer, were very helpful in making sure my thesis remained on the correct track. Other than Dr. C. Goodyer, the members of my committee included, Dr. Indra Gupta, Dr. Christine McCusker and Dr. Christian Sirard.

Of course, much of this thesis would not have been possible without the support of my family (yes, Mom, Dad, you should be seeing a degree up on the wall very soon), and perhaps most importantly my 3 year old niece Emilie. After spending most of the night (into the wee hours of the next morning) writing this thesis, her ability to continuously find that flute and start playing it at 7AM, surprisingly always seemed to put on smile on my face and gave me new found energy to keep writing.

I would also like to thank my incredible/brilliant/beloved girl-friend, Anna Khananian and her family for supporting me throughout this work. I must give special thanks to Anna because not many people would be capable of enduring my incessant conversation about the problems encountered with my research but Anna did so with unrelenting support.

Finally, I would like to thank a number of my closest friends including Johnny Zaccheo and his wife, Isabella along with Samer Hussein, for not forgetting about me while I disappeared into the depths of academic research for the past 6 years.

Last but not least, I would like to thank what is probably the most mysterious element known to mankind - call it whatever you may.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>III</td>
</tr>
<tr>
<td>RESUME</td>
<td>V</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>VII</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>IX</td>
</tr>
<tr>
<td>INDEX OF TABLES AND FIGURES</td>
<td>XII</td>
</tr>
<tr>
<td>1.0 CHAPTER ONE: Comprehensive Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>1.1 Diabetes Mellitus</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Type 1 Diabetes Mellitus</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Epidemiology of Type 1 Diabetes</td>
<td>5</td>
</tr>
<tr>
<td>1.2.2 Clinical Manifestations and Complications of Type 1 Diabetes</td>
<td>6</td>
</tr>
<tr>
<td>1.2.3 The Pancreatic β-Cell</td>
<td>8</td>
</tr>
<tr>
<td>1.3 The Immune System</td>
<td>11</td>
</tr>
<tr>
<td>1.3.1 Acquired Immunity</td>
<td>12</td>
</tr>
<tr>
<td>1.3.2 T and B Cell Activation</td>
<td>13</td>
</tr>
<tr>
<td>1.3.3 CD4+ T Cell and B Cell Activation</td>
<td>14</td>
</tr>
<tr>
<td>1.3.4 CD4+ and CD8+ T Cell Activation</td>
<td>14</td>
</tr>
<tr>
<td>1.3.5 Cell-Mediated vs Humoral Immunity - T_H1 vs T_H2 CD4+ T cells</td>
<td>17</td>
</tr>
<tr>
<td>1.3.6 Self-Tolerance</td>
<td>18</td>
</tr>
<tr>
<td>1.3.7 Central Tolerance</td>
<td>18</td>
</tr>
<tr>
<td>1.3.8 Peripheral Tolerance</td>
<td>20</td>
</tr>
<tr>
<td>1.3.9 Regulatory T Cells</td>
<td>22</td>
</tr>
<tr>
<td>1.4 Pathogenesis of Type 1 Diabetes Mellitus</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1 Antibodies and Autoantigens in Type 1 Diabetes</td>
<td>25</td>
</tr>
<tr>
<td>1.4.2 Insulitis and Cell Mediated Autoimmunity in Type 1 Diabetes</td>
<td>26</td>
</tr>
<tr>
<td>1.5 Susceptibility to Type 1 Diabetes</td>
<td>28</td>
</tr>
<tr>
<td>1.5.1 Environment and Type 1 Diabetes</td>
<td>29</td>
</tr>
<tr>
<td>1.5.2 Genetics of Type 1 Diabetes</td>
<td>29</td>
</tr>
<tr>
<td>1.5.3 Human Leukocyte Antigen Locus (HLA)/Major Histocompatibility Complex (MHC)</td>
<td>31</td>
</tr>
<tr>
<td>1.5.4 HLA Class I Alleles and Type 1 Diabetes</td>
<td>34</td>
</tr>
<tr>
<td>1.5.5 HLA Class II Alleles and Type 1 Diabetes</td>
<td>36</td>
</tr>
<tr>
<td>1.5.6 HLA Predisposing Alleles and Mechanism(s) of Action</td>
<td>39</td>
</tr>
<tr>
<td>1.5.7 IDDM2, The Insulin Variable Number of Tandem Repeat (VNTR)</td>
<td>41</td>
</tr>
<tr>
<td>Polymorphism and Type 1 Diabetes</td>
<td>41</td>
</tr>
<tr>
<td>1.5.8 PTPN22, CTLA-4 and Type 1 Diabetes</td>
<td>41</td>
</tr>
<tr>
<td>1.6 Thymic Insulin Expression, Insulin VNTR and Thymic Central Tolerance Induction</td>
<td>42</td>
</tr>
<tr>
<td>1.6.1 The Insulin Variable Number of Tandem Repeat (VNTR) Polymorphism</td>
<td>42</td>
</tr>
<tr>
<td>1.6.2 Thymic Insulin Expression</td>
<td>48</td>
</tr>
<tr>
<td>1.6.3 Thymic Central Tolerance Induction</td>
<td>50</td>
</tr>
</tbody>
</table>
## 2.0 CHAPTER TWO: Statement of Purpose and Objectives of the Research

### 3.0 CHAPTER THREE: Proinsulin Expression by Hassall’s Corpuscles in the mouse Thymus

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.01 Contribution of Authors</td>
<td>62</td>
</tr>
<tr>
<td>3.02 Objective</td>
<td>62</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>63</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>64</td>
</tr>
<tr>
<td>3.3 Methods</td>
<td>67</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>72</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>82</td>
</tr>
<tr>
<td>3.6 Acknowledgment</td>
<td>85</td>
</tr>
<tr>
<td>3.7 References</td>
<td>86</td>
</tr>
</tbody>
</table>

## 4.0 CHAPTER FOUR: Isolation and Characterization of Proinsulin Producing Medullary Thymic Epithelial Cell Clones

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.01 Contribution of Authors</td>
<td>91</td>
</tr>
<tr>
<td>4.02 Objective</td>
<td>91</td>
</tr>
<tr>
<td>4.1 Abstract</td>
<td>92</td>
</tr>
<tr>
<td>4.2 Introduction</td>
<td>93</td>
</tr>
<tr>
<td>4.3 Materials and Methods</td>
<td>96</td>
</tr>
<tr>
<td>4.4 Results</td>
<td>101</td>
</tr>
<tr>
<td>4.5 Discussion</td>
<td>112</td>
</tr>
<tr>
<td>4.6 Acknowledgments</td>
<td>115</td>
</tr>
<tr>
<td>4.7 References</td>
<td>116</td>
</tr>
</tbody>
</table>

## 5.0 CHAPTER FIVE: Microarray Analysis of Cultured Proinsulin Producing Medullary Thymic Epithelial Cells

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.01 Contribution of Authors</td>
<td>119</td>
</tr>
<tr>
<td>5.02 Objective</td>
<td>119</td>
</tr>
<tr>
<td>5.1 Abstract</td>
<td>120</td>
</tr>
<tr>
<td>5.2 Introduction</td>
<td>121</td>
</tr>
<tr>
<td>5.3 Materials and Methods</td>
<td>126</td>
</tr>
<tr>
<td>5.4 Results</td>
<td>128</td>
</tr>
<tr>
<td>5.5 Discussion</td>
<td>143</td>
</tr>
<tr>
<td>5.6 References</td>
<td>149</td>
</tr>
</tbody>
</table>
## 6.0 CHAPTER SIX: Class III INS VNTR alleles associated with silencing of thymic insulin predispose to type 1 diabetes

6.01 Contribution of Authors

6.02 Objective

6.1 Abstract

6.2 Introduction

6.3 Research Design and Methods

6.4 Results

6.5 Discussion

6.6 Acknowledgments

6.7 Abbreviations

6.8 References

## 7.0 CHAPTER SEVEN: DISCUSSION AND CONCLUSION

APPENDIX: ETHICS APPROVALS AND COPYRIGHTS

REFERENCES:
Index of Tables and Figures

Chapter 1.0 Tables and Figures

Index of Tables

Table 1: Etiological Classification of Diabetes Mellitus, adapted from [2]. Reproduced with permission from McGraw-Hill Companies. ................................................................. 3
Table 2: Susceptibility Loci for Type 1 Diabetes[87] ............................................................................... 30
Table 3: HLA Class II DR-DQ Linkage Patterns and IDDM Susceptibility in Whites [87]. ......................... 35
Table 4: Known tandem repeat elements of the INS-VNTR with the general consensus sequence in row a. ................................................................................................. 44

Index of Figures

Figure 1: Schematic diagram demonstrating CD4+ (A) and CD8+ (B) T cell binding to MHC II and MHC I expressing antigen presenting cells, respectively. C) Schematic diagram of the co-stimulatory molecules required for T cell activation. Adapted from [33, 34] ........................................................................................................ 15
Figure 2: A) Schematic diagram describing T cell development in the thymus. Positive T cell selection occurs in the thymic cortex whereas negative selection occurs largely in the thymic medulla. B) Only ~2% of thymocytes survive development as a result of positive and negative selection. Adapted from [32, 33] ........................................................................ 21
Figure 3: (Top Panel) Schematic diagram of the Human MHC/ Human Leukocyte Antigen (HLA) Locus and the homologous MHC in mice (referred to as H-2 or simply as I). (Bottom Panel) HLA alleles currently officially assigned by the WHO Nomenclature Committee for Factors of the HLA system. Adapted from [32] ........................................................................................................ 32
Figure 4: Schematic diagram portraying the Tyrosine Hydroxylase (TH) – Insulin (INS) - Insulin Like Growth Factor-2 (IGF-2) Region on human chromosome 11p15 also showing the 10 polymorphisms within the 4.1kB region. Open, closed and hatched boxes depict introns, exons and untranslated regions, respectively. Polymorphisms are designated with respect to the first base of the initiating ATG [91]. Reprinted by permission from Maxmillan Publishers LTD: Nature Genetics, 2005 ........................................... 44
Chapter 3.0 Index of Figures

Figure 1: Proinsulin/β-galactosidase expression in the thymus and pancreas. (A-B) Quantitative RT-PCR of proinsulin in the thymus of Ins1-KO and Ins2-KO mice normalised with cyclophilin. Lower band - competitor sequence, upper band - endogenous insulin, 1-8 are serial dilutions of competitor. (C) Semiquantitative RT-PCR of β-Gal in thymus and pancreas using serial dilutions of cDNA (1, 2, 3 = 1/1, 1/10, 1/100, normalised with cyclophilin). ................................................................. 73

Figure 2: FACS Sorting of Thymic Dendritic and Epithelial Cell Populations: Sort regions for DC without overnight incubation – with DEC205 (A), with CD11c and I-A^b (MHC II) (B). Epithelial sorts - G8.8^+ (C) (mainly TEC), and G8.8^+/CD45^-lo (D) (TEC). Dotted square - G8.8^+/CD45^- cells ........................................................................ 75

Figure 3: Proinsulin/β-galactosidase expression in thymic epithelial cells versus dendritic cells ................................................................................................................ 76

Figure 4: Histochemistry of β-Gal activity on pancreatic and thymic frozen sections. (A, B) pancreatic β-gal activity detected in β-cells of Ins2-KO mice but not in Ins1-KO (40x). (C, D) Thymic β-Gal activity detected in medullary epithelial cells (arrows) of HC of Ins2-KO mice but not in Ins-1KO mice (40x). (E, F) Zoom of HC cells positive for β-Gal activity ........................................................................................................................ 78

Figure 5: Immunohistochemistry of β-Gal activity and thymic stromal cells. (A) Colocalization of β-Gal activity and CD11c (DC) (40x). (B-E) Colocalization of β-Gal activity and G8.8 (TEC) (40x). No staining was observed in Ins-1KO sections (data not shown) ........................................................................................................ 80

Figure 6: β-galactosidase activity in sorted thymic G8.8^+ cells .................................................................................................................. 81

Chapter 4.0 Tables and Figures

Index of Tables

Table 1: Primers and PCR conditions ........................................................................................................ 100
Index of Figures

Figure 1: RT-PCR for proinsulin (detecting both 1 and 2) in cell colonies grown at 33°C with IFN-γ. All colonies shown here had above average β-galactosidase luminescence values (2-4 fold above background). N designates negative for proinsulin, P designates positive for proinsulin. 103

Figure 2: RT-PCR for proinsulin (detecting both 1 and 2) in cell colonies grown at 33°C without IFN-γ and placed for 3 days at 37°C without IFN-γ before RNA was extracted. N designates negative for proinsulin, P designates positive for proinsulin. All reactions done in triplicate for both negatives and one of the positives - the other positive was also tested and gave the same results (data not shown). 104

Figure 3: FACS analysis results for one of the proinsulin positive colonies. Identical results were obtained for the other proinsulin positive colony and the two proinsulin negative colonies. Top panel for each of the analysis is the negative or isotype control, bottom panel is with antibody or lectin (for UEA-1). UEA-1 specificity was assayed by incubating 10⁶ Jurkat cells with the same concentration of UEA-1 lectin used for the epithelial cells. No positive cells were observed (data not shown). No differences in signal were observed with cells incubated at 33°C or 37°C with or without IFN-γ except for MHC II which was up-regulated significantly in the presence of IFN-γ as shown. 106

Figure 4: Phase contrast microscopy of the proinsulin negative and positive mTECs demonstrating a cuboidal and "cobblestone" appearance characteristic of epithelial cells. 107

Figure 5: A) RT-PCR for CD80 in INS2KO/Immortal mouse pancreas, proinsulin negative mTEC colony, proinsulin positive mTEC colony and INS2KO/Immortal mouse thymus. B) RT-PCR for Ins2, Ins1, Pdx-1, Glut-2 and Gck in INS2KO/Immortal mouse pancreas and thymus. C) RT-PCR for Ins2, Ins1, Pdx-1, Glut-2 and Gck in the proinsulin negative and positive mTEC colonies. Pancreas and thymus RNA was obtained from a fourth 10 week old INS2KO/Immortal mouse mice. Cyclophilin was used to control for loading (data not shown). 109

Figure 6: The effect of preincubation with anti-lymphotoxin B antibody and trichostatin A (LTBR-Ab + TSA) on Aire (left panel) and insulin (right panel) by medullary thymic epithelial cell clones expressing or not expressing insulin, respectively Ins(+) mTEC and Ins(-) mTEC. The means and standard errors of seven (Aire) or four (insulin) separate preincubations are shown. Open bars represent untreated, shaded bars treated cells. 111
Chapter 5.0 Tables and Figures

Index of Tables

Table 1: Genes over-expressed by the PPC cell lines relative to the NPP cell lines. 47 genes are over-expressed, 23 of which can be considered tissue restricted (expressed in less than 5 tissues at either 10x the median or at a minimum 3x the median - labeled*). Gene symbols labeled (***) are Representative Public IDs and have yet be assigned a gene symbol. Gene symbols labeled (****) were represented by 2 or more probe sets. Fold differences are expressed in log2 format. 3 genes have been previously reported as over-expressed in mTECs vs cTECs, along with 9 genes that are very similar (ie. belonging to the same family and/or having a similar function) to others previously reported. ........ 135

Table 2: Genes over-expressed by the NPP cell lines relative to the PPC cell lines. 39 genes are over-expressed, 20 of which can be considered tissue restricted (expressed in less than 5 tissues at either 10x the median or at a minimum 3x the median - labeled*). Gene symbols labeled (***) are Representative Public IDs and have yet be assigned a gene symbol. Gene symbols labeled (****) were represented by 2 or more probe sets. Fold differences are expressed in log2 format. 8 genes have been previously reported as over-expressed in mTECs vs cTECs, along with 4 genes that are very similar (ie. belonging to the same family and/or having a similar function) to others previously reported. ........ 136

Table 3: Tissues represented by the tissue restricted antigens over-expressed by the PPC and NPP cell lines. .............................................................. 139

Table 4: Inter-gene distances for over-expressed genes in PPC cell lines ...................... 140

Table 5: Inter-gene distances for over-expressed genes in NPP cell lines ..................... 141

Table 6: Aire and Hand2/Tecl3 binding sites located within the 1000bp upstream of the Ins1 and Ins2 transcriptional start sites. Location refers to the distance from the translational start site. .................................................. 142

Index of Figures

Figure 1: Reproducibility of microarray data obtained between triplicate samples of the same cell line. A) Plot of microarray signals from PPC 2.2 versus PPC 2.3 and B) signals from NPP1.1 versus NPP1.2 .......................................................... 129

Figure 2: Degree of similarity of the microarray expression data obtained for each of the cell lines. A) Average signal from the triplicates of PPC 1 versus PPC 2 B) Average signal from the triplicates of NPP 1 versus NPP 2 C) Average signal from ............................................. 130

Figure 3: A) PPC1 – average signal from NPP1 and 2 versus PPC2 – average signal from NPP1 and 2 – note the linear correlation indicating the same genes are over-expressed. B) Same comparison for NPP1 and 2 and C) A random comparison ......................... 132

Figure 4: Number of over-expressed genes (4 fold or more) versus fold difference in log2 scale. 60 genes are over-expressed by the PPC cell lines relative to the NPP cell lines. 45 genes are over-expressed by the NPP cell lines relative to the PPC cell lines. Remaining comparisons have: AvePPC1NPP2-AvePPC2NPP2 (12 over-expressed genes), AvePPC2NPP1-AvePPC1NPP2 (8 over-expressed genes), AvePPC1NPP1-AvePPC2NPP2 (2 over-expressed genes) and PPC2NPP2-PPC1NPP1 (1 over-expressed gene) ........................................... 133
Chapter 6 Tables and Figures

Index of Tables

Table 1: Summary of transmissions from nondiabetic parents to a diabetic child. We examined 287 diabetic children and 227 pairs of parents (908 parental chromosomes). In the 60 HBDI families that had two children transmissions were calculated twice, once for each child. Therefore, the total number of transmissions was calculated twice, once for each child. Therefore, the total number of parental chromosomes. .................................................................................................... 168

Index of Figures

Figure 1: Amplification of all three classes of INS VNTR alleles by a single PCR reaction in heterozygous DNA samples. S1 and S2, DNA from the two thymus samples where insulin expression from the class III chromosome was silenced; E, thymus sample with enhanced thymic expression from the class III chromosome. Smaller alleles amplify more strongly. The length of the repetitive part was calculated by subtracting 668 bp of nonrepetitive flank sequence from that estimated for the PCR product. ......................... 162

Figure 2: Illustration of codominant segregation (a and b) and stable transmission (c and d) of INS VNTR alleles in families. a and b show undigested amplification products run under conditions that show both class I and class III alleles (m, Mother; f, father; c, child). The mother in family a is heterozygous for two different class III alleles (finer size discrimination between class III alleles is not shown here because it requires longer electrophoresis in which the class I allele is lost). c and d show the two major bands of the MspI digest used to discriminate between class III alleles of the same size, as in Fig. 3. DNA from the original S1 and S2 samples is run as standard. u, DNA sample from unrelated individual. ....................................................................................................... 163

Figure 3: Digestion of the PCR product with MspI gave two major polymorphic bands for most alleles. The two major bands in S1, predicted from the sequence, are 427 and 736 bp. S2 has an identical lower band, whereas the upper band appears to be shorter by one repeat unit. Known S1 and S2 samples (marked) were frequently interspersed among unknown samples to adjust for gel drift. Lanes containing unknown samples are indicated with a U. An arrow under the U signifies that the fingerprint is identical to either S1 (left) or S2 (right). An example of a III/III genotype is shown. Cl, PCR amplification from cloned DNA. .................................................................................................................... 165

Figure 4: Sequence of two class III alleles that have the same size and the same MspI fingerprint. INS expression from the class III chromosome in the thymus from which E1 was isolated was enhanced (as is the case with most class III alleles), whereas S1 was associated with monoallelic INS silencing. Differences are highlighted. The sequence is given in variants of the repeat unit, as defined by Bennett and Todd (29). The nucleotide sequence of each variant is given. Variants n–q (bold) have not been previously reported. All sequence was confirmed on at least two overlapping clones. ........................................... 167
1.0 CHAPTER ONE: Comprehensive Literature Review

1.1 Diabetes Mellitus

Glucose is the fundamental energy source of many cell types. Normal human physiology dictates that the circulating glucose level be vigilantly monitored by the pancreatic β-cell. In turn, the pancreatic β-cell secretes the corresponding level of insulin required, since glucose uptake by many cells is dependant on insulin’s interaction with its cell surface receptor. Lack of insulin or its receptor, or diminished insulin/insulin receptor affinity, results in the inability of cells to uptake glucose consequently leading to a wide variety of clinical manifestations one of which is the accumulation of systemic glucose referred to as hyperglycemia.

Diabetes Mellitus consists of a group of heterogeneous metabolic disorders that inevitably result in hyperglycemia [1, 2]. The American Diabetes Association classifies Diabetes Mellitus into four categories based on their widely differing etiologies (Table 1) [1, 2]. They also differ considerably in their degree of glucose homeostatic impairment and in their underlying causal factors. As an example, Maturity Onset Diabetes of the Young (MODY types 1 through 5) are exclusively the result of impaired pancreatic β-cell function due to an autosomal dominant mutation in one of the many β-cell transcription factors or in the glucokinase gene [3-6]. Such mutations result in either a primary decrease in insulin production and secretion or, in the case of
I. Type 1 diabetes (-cell destruction, usually leading to absolute insulin deficiency)
   A. Immune-mediated
   B. Idiopathic

II. Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly insulin secretory defect with insulin resistance)

III. Other specific types of diabetes
   A. Genetic defects of -cell function characterized by mutations in:
      1. Hepatocyte nuclear transcription factor (HNF) 4 (MODY 1)
      2. Glucokinase (MODY 2)
      3. HNF-1 (MODY 3)
      4. Insulin promoter factor (IPF) 1 (MODY 4)
      5. HNF-1 (MODY 5)
      6. NeuroD1 (MODY 6)
      7. Mitochondrial DNA
      8. Proinsulin or insulin conversion
   B. Genetic defects in insulin action
      1. Type A insulin resistance
      2. Leprechaunism
      3. Rabson-Mendenhall syndrome
      4. Lipodystrophy syndromes
   C. Diseases of the exocrine pancreas—pancreatitis, pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocystic pancreatitis
   D. Endocrinopathies—acromegaly, Cushing's syndrome, glucagonoma, pheochromocytoma, hyperthyroidism, somatostatinoma, aldosteronoma
   E. Drug- or chemical-induced—Vacor, pentamidine, nicotinic acid, glucocorticoids, thyroid hormone, diazoxide, -adrenergic agonists, thiazides, phenytoin, interferon, protease inhibitors, clozapine, beta blockers
   F. Infections—congenital rubella, cytomegalovirus, coxsackie
   G. Uncommon forms of immune-mediated diabetes—"stiff-man" syndrome, anti-insulin receptor antibodies

IV. Gestational diabetes mellitus (GDM)

---

**Table 1: Etiological Classification of Diabetes Mellitus, adapted from [2]. Reproduced with permission from McGraw-Hill Companies.**
glucokinase mutations, the inappropriate sensing of systemic glucose levels and thus the secondary diminished production and secretion of insulin [5].

In contrast to MODY, which accounts for a very small proportion of Diabetes Mellitus cases, Type 2 diabetes, previously referred to as Non-Insulin Dependant Diabetes Mellitus (NIDDM), accounts for the majority (~90%) of the Diabetes Mellitus cases and is the result of complex genetic interactions along with environmental factors and even life-style choices. Ultimately, it is characterized by variable degrees of insulin resistance, impaired insulin secretion and increased hepatic glucose production [1, 2]. Many studies strongly correlate Type 2 diabetes with obesity and it was for many years considered to be a disorder affecting older overweight adults. Interestingly, the obesity epidemic in North America seems to be resulting in a worrisome increase in the number of documented Type 2 diabetes cases not only in older adults but in obese children and young adults as well [1, 2, 7, 8].

With regards to severity of impaired glucose homeostasis, Type 1 diabetes, is certainly one of the most severe with almost all of the cases stemming from the autoimmune destruction of the pancreatic β-cells leading to the complete lack of insulin. Type 1 diabetes accounts for the majority of the remaining 10% of Diabetes Mellitus cases and will be the remaining focus of this thesis [1, 2, 9].

1.2 Type 1 Diabetes Mellitus

Type 1 diabetes is a condition in which pancreatic β-cell destruction usually leads to absolute insulin deficiency. Two forms have been identified – 1A and 1B. Type 1A results from the cell mediated autoimmune attack of the pancreatic β-cells whereas 1B is
much rarer, has no known cause and seems to occur more often in individuals of Asian or African descent with a very strong hereditary component [1, 2, 9]. The focus of this thesis is on Type 1A diabetes and is referred to simply as Type 1 diabetes in the remainder of the text.

1.2.1 Epidemiology of Type 1 Diabetes

*Insulin Dependant Diabetes Mellitus (IDDM)* and *Juvenile Diabetes* were previously coined terms used synonymously with Type 1 diabetes. The term *IDDM* is now considered erroneous and misleading, however, since many other Diabetes Mellitus types may require insulin, including patients suffering from Type 2 diabetes [1]. The term *Juvenile Diabetes* was often used due to the fact that the majority of new cases of Type 1 diabetes were thought to occur in childhood. Although true, more recent data suggest that while 50-60% of patients initially present younger than 16-18 years of age it is also now widely accepted that the disease actually occurs at a low incidence level throughout adulthood and can even be seen in the eight and ninth decade of life in some patients [1, 9]. In addition, using the age of onset of the disorder to distinguish the type of diabetes involved is becoming increasingly inappropriate due to the increasing incidence of Type 2 diabetes in children.

The epidemiology of Type 1 diabetes has been reported in many countries predominantly in children younger than 15-18 years of age. Four important observations from the epidemiological studies have been reported and include: i) incidence has been increasing at approximately 2-5% per year worldwide without a known cause [10-12]. The increase is much more dramatic in populations with a previously lower incidence of disease but even in countries with the highest rates, such as Finland, there has been no
leveling off since statistics started being recorded in the 1950s. ii) Geographical variation is quite significant with countries such as China reporting the lowest incidence (0.57 cases/100000 population younger than 18 years of age) in contrast to rates roughly 30 times higher in the United Kingdom and United States (18-20 per 100000 per year) to almost 100 fold higher, at approximately 48-49 per 100000 per year in Finland and Sardinia [13]. iii) There is a significant trend towards decreasing age at presentation most notably in children younger than 5 years of age. Swedish data indicate that from 1983 to 1998, the overall incidence of Type 1 diabetes did not go up in the 0-34 year age group, but median age at diagnosis fell [14]. Similar findings were reported from 1989 to 2000 in Belgium in those younger than 40 years of age [15]. iv) In Quebec, the incidence has remained stable at ~10/100,000 in the past two decades, with a tendency of a shift towards younger onset (Legault and Polychronakos, 2006). Migrating populations take on the incidence of rates of their new countries within a short time. As an example, incidence rates of Type 1 diabetes in Asian children in the United Kingdom are similar to those of white or other ethnic backgrounds in the same area which is in striking contrast to the very low rates reported in Asia [16]. This suggests that environmental factors are important, in interaction with genetic susceptibility.

1.2.2 Clinical Manifestations and Complications of Type 1 Diabetes

The rate of β-cell destruction in Type 1 diabetes can be quite variable with a much more rapid course seen mainly in infants and children and a more latent course observed typically in adults [17]. In both cases a prodrome of mildly disturbed glucose homeostasis can often be detected years before onset of overt symptoms which requires destruction of >80% of the β-cells [1, 2, 18]. Children and young adolescents tend to
more often present with ketoacidosis as the first manifestation of the disease as a result of
the rapid β-cell destruction [1, 9]. Ketoacidosis can be defined as dangerously high levels
of ketones in the blood – the result of fatty acid breakdown in an attempt to supply
energy to the cells when the ability to utilize glucose is severely restricted due, for
example, to the complete abrogation of pancreatic insulin secretion following the
autoimmune destruction of the pancreatic β-cells [2]. Other patients present with frank
hyperglycemia with or without minimal ketoacidosis and some patients present with only
mild to moderate fasting hyperglycemia that can exacerbate to severe hyperglycemia
and/or ketoacidosis in the presence of an underlying infection or other stress[1].
Toxicities of systemic ketone accumulation can result in severe acidosis, diabetic coma
and death. Adults, with the more latent disease course, tend to retain residual β-cell
function sufficient to prevent ketoacidosis for many years [1].

The most common symptoms of marked hyperglycemia, often designated as the
classic signs of diabetes, include polyuria, polydipsia and polyphagia. Weight loss,
lethargy, abdominal cramps and blurred vision may also occur. Impairment of growth and
susceptibility to certain infections can also accompany chronic hyperglycemia [1, 2].

The long-term complications of diabetes include retinopathy with potential loss of
vision; nephropathy leading to renal failure; peripheral neuropathy with risk of foot
ulcers, amputation; and autonomic neuropathy causing gastrointestinal, genitourinary, and
cardiovascular symptoms and sexual dysfunction[1, 9]. These are referred to as
microvascular complications. Furthermore, patients with diabetes have an increased
incidence of atherosclerotic cardiovascular, peripheral vascular, and cerebrovascular
disease referred to as macrovascular complications [1, 9]. Hypertension, abnormalities of
lipoprotein metabolism, and periodontal disease are often found in people with diabetes. Glycation of tissue proteins and other macromolecules and excess production of polyol compounds from glucose are among the mechanisms thought to produce tissue damage due to chronic hyperglycemia [1, 2, 9].

If we include patients with Type 2 diabetes, diabetes is the most common cause of renal failure, in the developed world [9, 19], and the most common cause of acquired blindness in the Western World [20]. When we consider that patients suffering from Type 1 diabetes have a more severe course and earlier onset of disease in comparison to patients with Type 2 diabetes, patients with Type 1 diabetes are certainly at significant increased risk for the many diabetes associated complications. A recently released report from the US Centers for Disease Control estimated that a 10 year old child developing Type 1 diabetes in the year 2000 would lose, on average, approximately 20 years of their life, compared with their non-diabetic peers [21]. One should also not forget that the emotional and social impact of diabetes and the demands of therapy may also cause significant psychosocial dysfunction in patients and their families.

Of importance to note is that recent results from the Diabetes Control and Complications Trial (DCCT) and followed in the Epidemiology of Diabetes Interventions and Complications Study (DCCT/EDIC) suggests that very tight glycemic control can significantly reduce the incidence of microvascular and potentially macrovascular complications [9].

1.2.3 The Pancreatic β-Cell

The pancreas has both exocrine and endocrine functions. The exocrine functions involve the production of pancreatic juices and enzymes required for the digestion and
absorption of food. The exocrine cells make up the bulk of the pancreas and are grouped into acini with ducts enabling the secretion of the digestive juices. Dispersed amongst all the acini are the endocrine portions of the pancreas that are organized into structures called the Islets of Langerhans. In a normal human adult there are approximately $10^6$ islets but this accounts for only 1-2% of the entire mass of the pancreas. Each islet contains approximately 1000 endocrine cells which includes insulin producing β-cells (65-80% of the islet cells), glucagon producing α-cells (15-20% of the islets), somatostatin producing σ-cells (3-10% of the islet cells), and pancreatic polypeptide producing PP cells (1% of the islet cells).

Secretion of insulin from the β-cell can be initiated in minutes, can vary as much as 50-100 fold and can be sustained for several hours before changes in insulin gene transcription are required. This is because in the normal state, the pancreatic β-cell has many insulin filled granules ready for secretion upon demand. In addition to this, approximately 15% of the total pancreatic β-cell mRNA consists of insulin which can be translated and processed to active insulin (see below) within 30 minutes of a high glucose load. These two mechanisms combined can provide enough insulin for high glucose loads of approximately 3-4 hours in duration. For longer time frames, transcription of insulin must take place [2, 22].

Glucose is the key regulator of insulin secretion although fats, amino acids, ketones, gastrointestinal peptides and neurotransmitters also influence secretion. With respect to glucose the mechanism is as follows: Glucose stimulation of insulin secretion begins with the uptake of glucose into the pancreatic β-cell via the GLUT-2 glucose transporter. Following this, glucose is phosphorylated by glucokinase which is the rate
limiting step in the regulation of insulin secretion by glucose. Further metabolism of
glucose generates ATP which inhibits ATP sensitive K⁺ channels which results in
depolarization of the cell. This depolarization opens voltage dependant Ca²⁺ channels
stimulating insulin secretion [23-25].

In humans, insulin is transcribed from a single insulin gene located on chromosome
11. Although many transcription factors are important in its production, one absolutely
required factor is Pdx-1. In fact, Pdx-1 is not only required for insulin transcription, it is
also absolutely required for pancreas development [24, 26]. Another important note
about insulin production in the pancreas is that it is not dependant on insulin gene copy
numbers as it is subject to metabolic feedback control and transcription from even a
single gene can be up-regulated to maintain homeostasis [2, 22, 27].

Transcription and translation of the insulin gene initially results in an 86 amino
acid precursor polypeptide denoted preproinsulin. Subsequent proteolytic processing,
once in the endoplasmic reticulum, results in the removal of the aminoterminal signal
peptide giving rise to proinsulin. Proinsulin can at this point be divided into three regions
— an amino terminal A region of 21 amino acids, a carboxyl terminal B region of 30
amino acids and an internal sequence, referred to as the C region of length 31 amino
acids. Cleavage by the endoproteases proconvertase 1 and 2 (PC 1 and 2) gives rise to
each of the individual fragments (A, B and C) and the A and B fragments assemble via
disulfide linkages to form insulin. The exoprotease CPH is another enzyme that aids
PC1 and PC2 in the cleavage of proinsulin [2, 22].

Interestingly, mice, rats and three species of fish have two insulin genes [28, 29]. In
mice, the two genes are located on two different chromosomes, more specifically, gene I
is on chromosome 19 whereas gene II is on chromosome 7. Although gene I lacks an intron with respect to gene II (in both mice and rats), the two sequences are highly homologous and expressed in approximately equal amounts in pancreas. Gene II is believed to be the original gene and gene I is hypothesized to have occurred as a non-viral functional retroposon [28, 29]. As in humans, Pdx-1 expression is absolutely required for normal pancreatic development and insulin expression in mice and very likely in rats as well [24, 26].

1.3 The Immune System

Given that type 1 diabetes is due to autoimmune destruction of the β-cells, an overview of the immune system is highly relevant to the object of my thesis.

The immune system is a highly organized arrangement of cells and molecules with specific roles aimed at defending the organism against invasion from foreign substances including foreign organisms. In humans, there exist two fundamentally different types of immune response – innate and acquired. The innate response, which is usually involved in the initial phases of the host response to an invasion, involves a variety of bone marrow derived cells, more specifically from hematopoietic stems cells, and includes basophils, eosinophils, neutrophils, mast cells, phagocytes macrophages and natural killer cells (NK cells). The system is present in all individuals, at all times, does not increase with repeated exposure to the pathogen and does not discriminate between pathogens. The acquired immune response usually follows the innate and leads to clonal selection of antigen specific lymphocytes subsequently leading to long term protection from the specific invasive agent (immunological memory) [30-32].
1.3.1 Acquired Immunity

Cells acting within acquired immunity involve B lymphocytes, CD4+ and CD8+ T lymphocytes, as well as antigen presenting cells (dendritic cells, macrophages and B lymphocytes) (APCs). B lymphocytes express immunoglobulin B cell receptors (BCRs) on their cell surface whereas T cells express T cell receptors (TCRs) on their surface. BCRs on the B lymphocytes are capable of binding soluble antigens (eg. peptides, polysaccharides) as well as antigens presented to them via direct contact with the invading agent (eg. glycoproteins on the surface of a bacterium). On the other hand, TCRs from CD4+ T lymphocytes are solely capable of binding peptides that are presented to them bound to major histocompatibility class II molecules (MHC) and CD8+ cells only bind peptides presented to their TCRs via MHC class I molecules [30-32].

The MHC locus is one of great diversity and polymorphism in humans and is the main reason for transplanted tissue/organ rejections [30-32]. MHC class I molecules are expressed by all cell types except for red blood cells, and in general, present small peptides of all proteins produced within a cell on that respective cell’s surface [30-32]. MHC class II molecules, which are expressed mainly by APCs and in some instances by epithelial cells, also present peptides on the cell’s surface but these are usually peptides of proteins that have been phagocytosed by the cell [30-32]. If a T cell binds strongly to an MHC/antigen complex, this can potentially result in clonal expansion of the cell and an immune response against all cells bearing that MHC/antigen complex.

Both BCRs and TCRs have a constant region and a variable region, as well as three hypervariable regions within the variable region [30-32]. Although the variable
regions of the BCR or TCR are identical on any single lymphocyte, they differ from one lymphocyte to the next, and this allows for recognition of up to $\sim 10^{11}$ different types of antigens - when considering the complete lymphocyte repertoire [30-32]. This diversity of the TCRs and immunoglobulins is accomplished via an extraordinary feat using approximately 400 genes. It involves somatic DNA rearrangements such as cutting, splicing and modification of the genes encoding the variable sites of the BCR or TCR in a unique way within each lymphocyte [30-32]. As will be discussed in sections 1.3.6 – 1.3.9, a variety of mechanisms exist to ensure that T and B cells do not recognize molecules belonging to the organism proper (self-antigens).

1.3.2 T and B Cell Activation

B and T lymphocytes circulate throughout the lymphatic system and the blood as "naïve cells" – that is to say they have never encountered a specific antigen for their BCR or TCR, and thus have never been activated. Their activation, to stimulate the production of effector B and T lymphocytes, often, but not in all cases, requires binding of their receptor (TCR or BCR) to the antigen (MHC/antigen complex in the case of TCRs) as well as co-stimulatory signals [30-32]. These co-stimulatory signals are released when "danger" is sensed by the organism’s immune system. What exactly distinguishes an antigen as "dangerous" is still not known and is the subject of much research. It is important to note that a variety of lymphocyte activating pathways exist and this paper will not attempt in any way to cover all the pathways, but a few major ones will be briefly described.
1.3.3 CD4+ T Cell and B Cell Activation

Upon BCR binding of an antigen by a B lymphocyte, either a soluble antigen or one present directly on the surface of a bacteria or viral shell, the B cells are capable of internalizing the bound molecule. These internalized substances are subsequently presented on the B cell surface via MHC class II molecules. Simultaneously, if “danger” is sensed, there is induction of CD40 receptor and B7 ligands (B7-1, B7-2 often referred to as CD80 and CD86 respectively), all of which are expressed on the B cell surface. Antigen specific CD4+ T cells bind the antigen via the TCR/MHC II-antigen interaction, which induces expression of the CD40 ligand and CD28 receptor (receptor for B7 ligands) on the CD4+ T-cell surface. The receptor ligand interactions of B7 and CD28 cause activation and proliferation of CD4+ cells to either inflammatory T cells (TH1) or T helper cells (TH2). The CD40 receptor/ligand interaction leads to the activation of the B lymphocyte, and results in its differentiation to B effector cells (plasma cells that produce antibody) and cells which migrate to the lymphoid tissues to induce further response via germinal centers [30-32].

1.3.4 CD4+ and CD8+ T Cell Activation

CD4+ T cells can also be activated by macrophages and or dendritic cells (APCs), which present peptides from phagocytosed bacteria, viruses or other engulfed substances on their cell surface, via MHC class II molecules (Figure 1A). In some cases these antigens may also be presented via MHC class I molecules, which activates CD8+ T cells (see below and Figure 1B) [30-32]. Macrophages are found in many sites throughout the body and ingest dead and or foreign substances in the surrounding milieu. If considered “dangerous”, they present these antigenic substances to CD4+ T-cells leading to their
Figure 1: Schematic diagram demonstrating CD4+ (A) and CD8+ (B) T cell binding to MHC II and MHC I expressing antigen presenting cells, respectively. C) Schematic diagram of the co-stimulatory molecules required for T cell activation. Adapted from [33, 34].
subsequent activation to either a T\textsubscript{H1} or T\textsubscript{H2} phenotype. In the absence of "danger" signals, such antigen presentation may result in activation-induced apoptosis, anergy, or the generation of regulatory T-cells (peripheral tolerance). Dendritic cells are also found in many sites throughout the body and in the immature state are capable of engulfing cells and molecules. Once a substance that is considered "dangerous" is engulfed, the dendritic cell is activated, and it matures and begins migration towards the lymphoid tissues, where it begins the development of an immune response to the antigen via CD4+ T-cell activation once again via presentation of the antigen through MHC class II molecules. As mentioned previously, what results in a substance being considered "dangerous" is still not known but in both cases, whether it be dendritic cells, macrophages or B-lymphocytes, the engulfed substance must lead to the stable expression of the B7 ligands on its surface in order to activate the CD4+ T lymphocytes [30-32]. It should be noted that an immune response generated by dendritic cells is often more potent than one caused by macrophages or B-cells.

CD8+ T cells develop into cytotoxic T-cells (CTL) upon activation by APCs (dendritic cells and macrophages). CTL cells lead to an extremely potent immune response as they are capable of killing cells directly by either inducing cell apoptosis via direct expression of surface ligands for apoptosis-inducing receptors, such as FasL and membrane-bound TNF-\textalpha, or they can induce apoptosis secondary to perforin release which facilitates the passage of protease granzymes that lead to apoptosis [30-32]. Activation of CD8+ T cells usually proceeds through presentation of the antigens on the cell surface via MHC class I molecules [30-32]. Once again, B7 ligand (CD80 and/or 86)
expression on the dendritic cell surface is upregulated when “danger is sensed” and is necessary for CD8+ T-cell activation and proliferation. (Figure 1C).

1.3.5 Cell-Mediated vs Humoral Immunity – T\textsubscript{H}1 vs T\textsubscript{H}2 CD4+ T cells

The above descriptions of B and T cell activation are greatly simplified. The reality is that there is significant cross-talk between all the cells involved which gears the immune response towards either Cell-Mediated Immunity or Humoral Immunity. The cross-talk between cells is predominantly mediated by cytokines and often the induced immune response is actually a combination of both cell mediated and humoral immunity but one usually tends to dominate depending on which cytokines are present in the milieu. Cell-mediated immunity usually occurs when CD4+ T cells differentiate to T\textsubscript{H}1 inflammatory T cells resulting in the production of strongly opsonizing antibodies and the activation of macrophages. Activated macrophages can destroy intracellular pathogens but can also damage cells in the surrounding milieu by the production of cytotoxic substances such as oxygen radicals, hydroxyl radicals, nitric oxide and proteases. Activation of CD8+ T cells to cytotoxic T cells very often accompanies a T\textsubscript{H}1 response resulting in direct cell death. On the other hand, differentiation of CD4+ T cells to T\textsubscript{H}2 helper T cells usually results in a predominantly humoral immune response with the production of weakly opsonizing or neutralizing antibodies with minimal cytotoxic events. What determines whether a T\textsubscript{H}1 or T\textsubscript{H}2 response will occur is as of yet not completely understood but as mentioned previously cross-talk between cells using cytokines is very important. Cytokines associated with cell mediated immunity include Interleukin-1 (IL-1), IL-2, IL-3, IL-6, IL-12, IFN-\gamma and Tumor Necrosis Factor-\alpha and \beta.
Cytokines associated with a humoral response include IL-4, IL-5 and IL-10, TGF-β and IL-β [30-32].

1.3.6 Self-Tolerance

With the large repertoire of TCRs and BCRs described earlier, one may begin to wonder how an immune response does not build up against vital substances produced by our body. This would of course be deleterious to the host organism since its own cells producing required substances would be eradicated. Mechanisms which exist in the human body for developing tolerance to self-antigens include central and peripheral tolerance along with Regulatory T cells.

1.3.7 Central Tolerance

B lymphocytes begin their life cycle in the bone marrow derived from hematopoietic stem cells. Within the marrow they develop from Pro-B cells (not expressing any antibody) to mature differentiated B cells (each one expressing a single immunoglobulin specificity). During the progression to mature lymphocytes, the still immature B cells, right before the final maturation phase, have the capacity to bind antigens within the bone marrow that are either cell bound or soluble within the matrix. If the antigen is present at a high enough concentration, all B cells specific for that antigen and that bind with high enough affinity will die upon antigen binding [30-32]. B cells that bind directly to cell surface antigens die via apoptosis. Those that bind to soluble antigens become anergic, that is they are inactivated, and die via apoptosis later on [30-32].
T lymphocytes also develop from hematopoietic stem cells in the marrow, but their progenitors migrate to the thymus where they differentiate and learn to distinguish between self and non-self by positive and negative selection. Once in the thymus, the precursor T cells are referred to as thymocytes and during their initial phase of maturation they begin expressing both CD4 and CD8 on their cell surface. With the development of their TCRs, positive selection, a process largely limited to the thymic cortex, occurs which involves the selecting of T cells having TCRs that can bind “loosely” to the host's MHC (I or II) molecules expressed on cortical epithelial cells (cTECs). T cells that do not bind at all immediately undergo apoptosis as these T cells are of no use for the organism in an immune response as TCR recognition of an MHC/antigen complex is unlikely to occur if the TCRs cannot recognize the MHC. T cells that bind very tightly also undergo apoptosis as they would be deleterious to the host's cells - this latter process occurs much more extensively in the medulla (described next and section 1.6) [30-32]. Note that while in the cortex, T cells with TCRs capable of binding MHC I and II molecules are directed to become uniquely CD8+ and CD4+ T cells, respectively.

Loosely binding T cells subsequently migrate to the medulla where APCs, including medullary epithelial cells and dendritic cells present self-antigens to the remaining T cells via MHC I and MHC II molecules. Any T cells that are capable of binding tightly to these self-antigens/MHC complexes are also killed. This completes the negative selection for autoreactive T cells and those cells that have survived develop into mature T lymphocytes and are released into the periphery as naïve T cells [30-32]. Just like in the bone marrow, the expression of the self-antigen in the thymus must be high enough to obtain the most stringent of T cell selection [30-32]. Approximately 5% of T
cells survive positive selection and roughly 2% of all T cells entering the thymus survive both positive and negative selection [30-32] (Figure 2).

It should be noted that it was once believed that the thymus was important for T cell development solely in the early stages of life, but recent evidence suggests that despite the partial involution of the thymus that occurs at puberty, T cells continue to develop in the thymus throughout life [30, 31].

1.3.8 Peripheral Tolerance

Although central tolerance as described above leads to the selection of lymphocytes that do not react to self, central tolerance very likely does not eliminate all autoreactive T and B cells and it is conceivable that some autoreactive T cells do escape into the periphery for a variety of reasons (ie. some self-antigens may not be expressed in the bone marrow or thymus, or their expression level may be too low for efficient negative selection of autoreactive T cells) yet we do not develop a reaction towards them [30, 31]. This important process is termed peripheral tolerance. Although a B cell or T cell may bind a self-antigen somewhere in the body, to proliferate to effector T and B cells requires co-stimulatory signals – a “danger” must be sensed. If these signals are not present the lymphocytes bound to the antigen undergo anergy and eventually die. This process of lymphocytes binding self-antigens but not receiving any co-stimulatory signals, thus leading to their inactivation and death, is believed to be the major mechanism of peripheral tolerance induction [30, 31].
Figure 2: A) Schematic diagram describing T cell development in the thymus. Positive T cell selection occurs in the thymic cortex whereas negative selection occurs largely in the thymic medulla. B) Only ~2% of thymocytes survive development as a result of positive and negative selection. Adapted from [32, 33]
It should be noted that immunological ignorance is also believed to be important in peripheral tolerance which involves the sequestering of self-antigens from the immune system. In some cases, breaking down of barriers required for sequestering self-antigens results in an autoimmune disease (ie. uveitis following ocular lens trauma).

1.3.9 Regulatory T Cells

A sub-population of CD4+ T cells that are also CD25+ (IL-2 receptor α-chain) have been described in both humans and mice and they have been shown to have strong immunosuppressive effects both \textit{in vitro} and \textit{in vivo} (reviewed in [35]). Although referred to as suppressor T cells, the term regulatory T cell is often preferred because these cells are hypothesized as being capable of both enhancing or suppressing (ie. modulating) an immune response. It should be noted, however, that at present, the CD4+/CD25+ population has only been shown to suppress the immune response. The idea of regulatory T cells was originally described in the 1970s but many investigators lost interest in the subject when the cells described could not be cloned and eventually the field demised by the 1980s. Sakaguchi and associates, rekindled interest in the concept, in the mid-1990s, when they identified a population of CD4+/CD25+ cells and demonstrated that this population was crucial for the control of autoreactive T cells \textit{in vivo} [36, 37]. Although since then much work has been done to attempt to comprehend the mechanism by which these cells function, to date there are many more questions than answers.

Recent work done in mice has shown that CD4+/CD25+ cells develop in the thymus as a sub-population of the developing CD4+ cells [38, 39]. TCR antigen interaction has been shown to be important with at least one group reporting high-affinity
interaction with the antigen favoring the formation of the CD4+/CD25+ population [38, 39]. These same mouse models have been used to demonstrate that antigen presentation by either medullary or cortical epithelial cells but not dendritic cells or macrophages is sufficient for CD4+/CD25+ T cell formation. It should be noted, however, that this notion was recently challenged by Watanabe et al. [40] as their recently published work suggests dendritic cells play an important role in regulatory T cell formation.

Interestingly, it has also been shown that regulatory or suppressor T cells can be induced in the periphery through exposure to antigen in the presence of suppressive cytokines. This has been accomplished both in vitro and in vivo but the similarities/differences between these peripherally induced regulatory T cells versus those produced in the thymus is as of yet not known [35].

It is important to note that recent work has identified a vital transcription factor for the differentiation of regulatory T cells in the thymus and their maintenance in the periphery. Foxp3 is an X-chromosome encoded forkhead transcription factor that is indispensable in the differentiation and formation of the regulatory T-cells. Lack of expression of the transcription factor leads to an early onset, highly aggressive and fatal autoimmune disease referred to as immune dysregulation, polyendocrinopathy enteropathy, X-linked syndrome (IPEX) which in some patients results in an autoimmune destruction of the pancreatic β-cells [41].

1.4 Pathogenesis of Type 1 Diabetes Mellitus

The autoimmune reaction, in Type 1 diabetes, is very specific for β-cells as no other cells within the islets are destroyed (reviewed in [42]). The β-cell autoantigen responsible for inducing the autoimmune attack is as of yet unknown but there are a
number of suspected candidates that will be discussed in the subsequent section. A couple of points important to mention before discussing the pathogenesis of Type 1 diabetes include: i) the concordance of Type 1 diabetes development in identical twins is approximately 40% implying that there is both a genetic and environmental component in the etiology[43]. A number of hypotheses have been put forward to explain the environmental component but none have been conclusively proven. One of the more popular notions implicates a viral infection prior to onset of disease that results in an autoimmune response via either molecular mimicry (with β-cell antigens) or through direct viral induced β-cell damage releasing and exposing β-cell autoantigens in a “dangerous” environment (see section 1.3.2) resulting in an autoimmune reaction [9, 42].

ii) Studying the pathogenesis of Type 1 diabetes is difficult in humans therefore much of the detailed pathophysiological work has been performed in two animal models of Type 1 diabetes which include the Non-Obese Diabetic mouse (NOD) and the Diabetes Prone Bio Breeding rat (DP-BB). With respect to the NOD mouse, overt diabetes can be detected in 5% of animals as early as 14 weeks of age and 77% of the female mice and 28% of the male mice become afflicted by 27 weeks of age [44]. Within the DP-BB rat line, the incidence of spontaneous autoimmune diabetes is approximately 77% (females) and 87% (males) by 18 weeks of age [45]. As in humans, insulitis (see below) is observed in both the mouse and rat models prior to the autoimmune destruction of the β-cells [42, 45].

iii) Although unavoidable, the usage of animals to study the pathogenesis of Type 1 diabetes also, to some extent, limits our understanding of the complexity of the disease [45, 46]. As will be discussed in subsequent sections, Type 1 diabetes is a multifactorial polygenic disease with many genes involved in predisposition and pathogenesis.
with each gene carrying its own burden and having its own mechanism whereby it predisposes to disease [45-47]. Different combinations of these predisposing genes can all potentially result in the development of Type 1 diabetes via unique mechanisms. The NOD mice and BB rats, each of which is inbred, share an identical genotype and as such really correspond to the development of Type 1 diabetes with regards to two specific genotypes. In a sense, this can be likened to studying the onset of diabetes in only two patients (each of the animal models corresponding to one patient). In addition, it should also be pointed out that although only a proportion of these animals develop disease, an environmental etiology causative of disease is difficult to identify because these animals generally live in identical controlled conditions making the comparison with regards to environmental causation between human and animal disease more obscure.

1.4.1 Antibodies and Autoantigens in Type 1 Diabetes

Years prior to the onset of overt diabetes, antibodies against a variety of β-cell antigens can often be found in the peripheral blood of Type 1 diabetes patients. The same antibodies are also found prior to overt diabetes development in the animal models and include antibodies against insulin [48], islet-specific glucose 6-phosphatase catalytic subunit related protein (IGRP)[49], glutamic acid decarboxylase (GAD65 in humans, GAD65 and 67 in mice) [50], tyrosine phosphatase (IA-2 and IA-2β) [51, 52], carboxypeptidase H [53], islet cell antigen 69 (ICA69) [54], GM gangliosides [55], a 38kd autoantigen [56] and SOX13 [57]. Interestingly, the antibodies are not believed to have a pathogenic role in Type 1 diabetes development as disease cannot be induced in other animals following transfer of the antibodies [2]. However, the risk of developing diabetes is strongly related to the number of autoantibody markers a patient has, that is,
the presence of 2 or more autoantibodies gives a higher probability of developing the
disease than the presence of a single autoantibody alone [58]. Of the antibodies found,
anti-insulin, anti-IGRP, anti-GAD and the anti-tyrosine phosphatase have been the most
extensively studied. Each of them can be found in approximately 50-80% of diabetes
patients and all three of the targeted antigens (ie. insulin, IGRP, GAD and tyrosine
phosphatase) are equally suspected as being the causal autoantigen but definitive proof
implicating one specifically is lacking [42, 59, 60]. It should be noted that T cells
reactive against each of these antigens can also be found during overt disease in both
humans and mice [42, 61-63]. The current notion is that the antibodies and T cells
develop against all of the antigens as a result of epitope spreading but that one is actually
the causal antigen. That being said, evidence implicating insulin as the causal
autoantigen includes the fact that anti-insulin antibodies are usually the first antibodies to
be detected and they are usually the first indication of an ongoing autoimmune process in
both patients and animal models [60, 64]. In addition, besides IGRP, insulin is really the
only β-cell specific antigen implicated and recent work has demonstrated tolerance
induction towards insulin and not IGRP, in the NOD mouse, prevents diabetes [65, 66].

1.4.2 Insulitis and Cell Mediated Autoimmunity in Type 1 Diabetes

In humans, histological analysis of the pancreas from patients with recent onset
Type 1 diabetes reveals an infiltration of the islets of Langerhans by mononuclear cells
which have been identified as T and B lymphocytes, monocytes/macrophages, and
natural killer cells (NK) [67-69]. In animals, the initial cells to infiltrate the pancreas are
macrophages and dendritic cells followed not too long after by B-lymphocytes [70-73].
The migration of these APCs into the pancreas is hypothesized to occur due to some
event resulting in β-cell turn over and or damage (viral or otherwise). Uptake and presentation of β-cell autoantigens either within the islets or in the pancreatic lymph nodes, by these APCs to naïve CD4+ and CD8+ T cells is then believed to result in T cell infiltration of the islets. This infiltration occurs both in animals and humans and is referred to as insulitis. Interestingly there is no evidence of β-cell destruction during insulitis [74]. In fact, some animal models have been shown to develop insulitis but never progress to autoimmune destruction of the β-cells [74]. At some point following insulitis, the very specific destruction of β-cells begins and this can progress rapidly or more gradually until overt diabetes symptoms develop (>80% of the β-cell mass must be destroyed).

Animal studies have revealed some important findings regarding disease progression: i) T cells are absolutely required for disease as athymic mice do not develop insulitis or diabetes. Additionally, the transfer of CD4+ T and CD8+ T cells but not antibodies and or B-lymphocytes, from affected mice into unaffected mice transfer disease. ii) The autoimmune destruction is cell mediated with cytotoxic T cell involvement that requires activation with CD4+ Th1 T cells. Transfer of CD4+ T cell clones alone from affected mice into non-affected mice tends to result in insulitis without disease progression whereas transfer of CD8+ T cell clones alone can usually transfer disease. When transfers of CD8+ T cell clones alone do not result in disease, disease can be induced with CD4+ T cell transfers that result in the recruitment of the CD8+ T cells into the pancreas. Further evidence implicating a Th1 response includes the findings that administration of Th2 associated cytokines (IL-4, IL-10, TGF-β) and inhibition of Th1 cytokines (soluble IL-1 receptor, anti-IFN-γ, anti-TNF-α) can prevent disease with a
direct reduction in cytotoxic T cell production and activity. Additionally, administration of Th1 cytokines can exacerbate disease progression. iii) CD8+ T cells are strongly implicated in β-cell destruction by inducing apoptosis via FasL, membrane bound TNF-α and or perforin. Systemically Fas deficient NOD mice do not develop insulitis or diabetes. Diabetes development is also inhibited in TNF receptor-1 deficient NOD mice and NOD mice lacking perforin expression. iv) Macrophages are important in disease initiation and progression. The incidence of diabetes in NOD mice and BB rats is significantly reduced in macrophage depleted animals. In such animals, Th1 associated cytokines are significantly reduced whereas expression of Th2 cytokines is increased. Once activated in a Th1 environment, macrophages have also been shown to have a role in β-cell destruction by the production of oxygen radicals, NO species, hydroxyl radicals and proteases. Studies have shown that the β-cells are more susceptible to these cytotoxic substances than are the other cells within the islets.

1.5 Susceptibility to Type 1 Diabetes

As stated previously, the etiology of Type 1 diabetes has been shown to be dependent on both genetic and environmental factors. The probability of monozygotic twins (identical twins) both inheriting the disease has been reported to be anywhere between 30-70% indicating that there is a strong environmental component[43, 75]. Dizygotic twins (non-identical), from a parent with type 1 diabetes, have approximately the same probability of developing the disease as a regular offspring (~5%) [76]. The likelihood of developing diabetes without one of the parents affected is a ten fold less [75, 76]. Note that if a single gene locus was involved simple Mendelian genetics would be observed and if the environment was not involved identical twins would be concordant
100% of the time. Such figures do indicate that offspring with an affected first-degree relative have an increased risk of developing the disease. However it is important to note that 85% of newly diagnosed cases are actually sporadic, in that they do not have a first degree affected relative[47].

1.5.1 Environment and Type 1 Diabetes

Suspected environmental triggers in Type 1 diabetes are many and include viruses (eg. enteroviruses, coxsackie, congenital rubella)[77, 78], environmental toxins (eg. nitrosamines) [79] or foods (eg. early exposure to cow's milk proteins, cereals, or gluten) [80-84]. Of these factors, a close relation, albeit a very modest one, has been identified only with congenital rubella [77, 78]. Recently, a study also disproved a possible correlation between type 1 diabetes and childhood vaccinations[85, 86]. Due to the lack of a strong association with any of these factors the search for a definitive environmental trigger continues.

1.5.2 Genetics of Type 1 Diabetes

To date, linkage analysis has tentatively identified more than 18 loci that have been associated with Type 1 diabetes, named by using the abbreviation IDDM and a number that usually corresponds to the chronological order in which they were reported (e.g. IDDM1 – IDDM18) (Table 2) [87]. Some of the loci have yet to be named and for most, a responsible gene has yet to be mapped. Most are now believed to be statistical artifacts[88]. The candidate-gene approach, however, implicated two specific genes and genome wide scans have also consistently linked them to type 1 diabetes. IDDM1, is the first of these two loci described and it maps to the Human Leukocyte Antigen (HLA)
<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Candidate Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDDM1</td>
<td>6p21.3</td>
<td>HLA DR/DQ</td>
</tr>
<tr>
<td>IDDM2</td>
<td>11p15.5</td>
<td>INSULIN VNTR</td>
</tr>
<tr>
<td>PTPN22</td>
<td>1p13</td>
<td>PTPN22(LYP)</td>
</tr>
<tr>
<td>SUMO4</td>
<td>6q25 (IDDM5)</td>
<td>SUMO4</td>
</tr>
<tr>
<td>IDDM3</td>
<td>15q26</td>
<td></td>
</tr>
<tr>
<td>IDDM4</td>
<td>11q13.3</td>
<td>MDU1, ZFM1, RT6, ICE, LRP5, FADD, CD3</td>
</tr>
<tr>
<td>IDDM5</td>
<td>6q25</td>
<td>SUMO4, MnSOD</td>
</tr>
<tr>
<td>IDDM6</td>
<td>18q12-q21</td>
<td>JK (Kidd), ZNF236</td>
</tr>
<tr>
<td>IDDM7</td>
<td>2q31-33</td>
<td>NEUROD</td>
</tr>
<tr>
<td>IDDM8</td>
<td>6q25-27</td>
<td></td>
</tr>
<tr>
<td>IDDM9</td>
<td>3q21-25</td>
<td></td>
</tr>
<tr>
<td>IDDM10</td>
<td>10p11-q11</td>
<td></td>
</tr>
<tr>
<td>IDDM11</td>
<td>14q24.3-q31</td>
<td>ENSA, SEL-1L</td>
</tr>
<tr>
<td>IDDM12</td>
<td>2q33</td>
<td>CTLA-4</td>
</tr>
<tr>
<td>IDDM13</td>
<td>2q34</td>
<td>IGFBP2, IGFBP5, NEUROD, HOXD8</td>
</tr>
<tr>
<td>IDDM15</td>
<td>6q21</td>
<td></td>
</tr>
<tr>
<td>IDDM16</td>
<td>14q32</td>
<td>IGH</td>
</tr>
<tr>
<td>IDDM17</td>
<td>10q25</td>
<td></td>
</tr>
<tr>
<td>IDDM18</td>
<td>5q31.1-33.1</td>
<td>IL12B</td>
</tr>
<tr>
<td></td>
<td>1q42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16p12-q11.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16q22-q24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17q25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19q11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3p13-p14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9q33-q34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12q14-q12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19p13.3-p.13.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Susceptibility Loci for Type 1 Diabetes[87]
Locus/ The Major Histocompatibility Complex (MHC) located on the short arm of chromosome 6 [89]. The second, \textit{IDDM2}, maps to the insulin gene VNTR/variable number of tandem repeat region, a polymorphic region located just upstream of the insulin gene on chromosome 11p15 [90-92]. Other genes associated with Type 1 diabetes include \textit{PTPN22} and \textit{CTLA-4 (IDDM12)} but their effect, as is probably the case for many of the other suspected genes, is less than \textit{IDDM1} and 2 as only modest linkage has been demonstrated and linkage of these genes is often not confirmed in all genome scans [88] Although true that the effect of these remaining genes is probably less than \textit{IDDM1 and 2}, it is important to note that Type 1 diabetes predisposition is very likely a function of all or many of these loci contributing a variable percentage of risk – ie. Type 1 diabetes is a multifactorial polygenic disease. This notion was originally proposed by Eisenbarth in the mid 1980s as he stated that everyone is born with some degree of susceptibility to Type 1 diabetes – an individual could be at high, modest or low risk depending on the pattern of inheritance of the susceptibility genes [93]. As such, the overall risk becomes a combination of the inherited risk and environmental factors.

1.5.3 Human Leukocyte Antigen Locus (HLA)/Major Histocompatibility Complex (MHC)

In this section, a brief review of the HLA locus and nomenclature is given. The HLA class I locus contains three major genes, A, B and C (Figure 3). They are homologous genes of the same family and extremely polymorphic with approximately 303 A alleles known, 559 B alleles and 150 C alleles. The polymorphic product of each of these genes is an \(\alpha\)-chain which complexes to a non-polymorphic glycoprotein known as \(\beta2\) – microglobulin. Class I A, B and C are expressed on nearly all nucleated cells and
Figure 3: (Top Panel) Schematic diagram of the Human MHC/ Human Leukocyte Antigen (HLA) Locus and the homologous MHC in mice (referred to as H-2 or simply as I). (Bottom Panel) HLA alleles currently officially assigned by the WHO Nomenclature Committee for Factors of the HLA system. Adapted from [32].
are essential for presenting antigenic peptides to CD8+ Cytotoxic T lymphocytes during a cellular immune response (see sections 1.3.3 and 1.3.4). Another function of these class I alleles includes positive selection of host/MHC compatible CD8+ T lymphocytes as well as deletion of autoreactive CD8+ T cells via self-antigen presentation during T cell maturation in the thymus (see sections 1.3.5-1.3.8). The nomenclature for the class I alleles is quite simple; the 8th polymorphic allele of the class IB type is labeled HLA-B8. The HLA class II locus also contains three major genes known as DR, DQ and DP. These molecules consist of two chains an α and a β chain, all of which are polymorphic except for the α chain of the gene DR (Figure 3). The DR gene does, however, have two different β-chain loci available to pair with the DRα chain. These HLA class II molecules are expressed by dendritic cells, B-lymphocytes, macrophages, activated T-lymphocytes (in humans) and most epithelial cells can also be induced to express them as well. They are important for initiating CD4+ T cell immune responses, and they, just like the class I alleles are also important for the positive selection of host/MHC compatible CD4+ T-lymphocytes as well as the negative deletion of autoreactive CD4+ T cells (see sections 1.3.3-1.3.8). Their nomenclature is slightly more complicated since polymorphisms exist for all the chains except DRα. For example, the DR molecule DRB1*0401 refers to the β-chain termed number 1 for the DR gene, and the 0401 refers to the specific polymorphic allele of that β-chain. Note that since the α-chain of DR is constant, it is not necessary to include it in the nomenclature and also note that the 04 refers to the serological specificity. For a specific DQ molecule, the following nomenclature would be used. DQA1*0301-DQB1*0302. Observe that in this situation there are two specific polymorphic chains that must be accounted for and hence the
additional specificity of the β-chain. Taken together, pairs of alleles at the two chains correspond to DQ serotypes. This specific DQ haplotype corresponds to the DQ8 serotype.

Nomenclature regarding the MHC genes also exists for haplotypes of DQ and DR (among others) alleles that are in very high linkage disequilibrium with each other. Unfortunately, the names of such haplotypes remain quite ambiguous, but the relevant ones for this paper are listed in table 3. All material in this section can be further reviewed in references [32, 47].

1.5.4 HLA Class I Alleles and Type 1 Diabetes

The association of MHC alleles or Human Leukocyte Antigen (HLA) alleles with type 1 diabetes was recognized more than 20 years ago by Nerup and colleagues and it is now widely believed that the HLA locus accounts for up to 50% of the susceptibility to Type 1 diabetes [94]. Eisenbarth and colleagues demonstrated the association between the HLA-B8 allele and Type 1 diabetes. Subsequently two other alleles of the class I B type, HLA – B18 and B15 were also associated and affected sib-pair studies provided evidence for linkage of all three of these class I B alleles with Type 1 diabetes [95-97]. Several other reports have not only suggested that HLA class I genes confer susceptibility to diabetes but that they also affect particular aspects of the disease such as age onset and the rate of β-cell destruction [98-103]. Further studies, however, went on to demonstrate a much stronger association existed between the HLA class II alleles and Type 1 diabetes and much work since then has focused on the genes of the class II locus [95-97].
<table>
<thead>
<tr>
<th>HLA-DR</th>
<th>DQA1</th>
<th>DQB1</th>
<th>DRB1</th>
<th>Susceptibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR2</td>
<td>102</td>
<td>602</td>
<td>1501</td>
<td>Protective</td>
</tr>
<tr>
<td>DR2</td>
<td>102</td>
<td>0502(AZH)</td>
<td>1601</td>
<td>Predisposing</td>
</tr>
<tr>
<td>DR2</td>
<td>103</td>
<td>601</td>
<td>1502</td>
<td>Neutral</td>
</tr>
<tr>
<td>DR3</td>
<td>501</td>
<td>201</td>
<td>301</td>
<td>High Risk</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>302</td>
<td>401</td>
<td>High Risk</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>302</td>
<td>402</td>
<td>Predisposing</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>302</td>
<td>403</td>
<td>Neutral</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>302</td>
<td>404</td>
<td>Predisposing</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>302</td>
<td>405</td>
<td>High Risk</td>
</tr>
<tr>
<td>DR4</td>
<td>301</td>
<td>301</td>
<td>401</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

Table 3: HLA Class II DR-DQ Linkage Patterns and IDDM Susceptibility in Whites [87].
1.5.5 HLA Class II Alleles and Type 1 Diabetes

Two strongly predisposing haplotypes of the class II alleles are DR4 and DR3 (table 3). The DR4 haplotype consists of the DQA1*0301-DQB1*0302 (DQ8) alleles along with either DRB1*0401 or DRB1*0405. DR3, on the other hand, consists of DRB1*0301 with DQA1*0501-DQB1*0201 (also referred to as DQ2). Ninety percent of individuals with type 1 diabetes have at least one of these high-risk haplotypes compared to 20% of the general population [47]. To further substantiate this, a study in a U.S. population indicated that 35% of patients were compound heterozygotes for DR3/DR4 vs. 2.4% of the general population [47]. In addition, studies have shown that if a non-affected sibling of a patient with diabetes carries both high risk haplotypes (DR3/DR4), the risk of diabetes is approximately 25 - 40% vs. the reported 5% [47].

For some time, distinguishing which of the two genes, DQ or DR, played a greater role in Type 1 diabetes susceptibility was very difficult. The two genes are very close together on the chromosome and in tight linkage disequilibrium (the result of infrequent recombinations between the two genes). As DNA-based sequencing techniques and typing technology evolved, the HLA-DQ locus was found to be more strongly associated with diabetes susceptibility than the DR locus [87]. As one might expect the suspected DQ alleles present within the DR3 and DR4 haplotypes DQA1*0301, DQB1*0302, DQA1*0501 and DQB1*0201 were believed to be the greatest contributors to this predisposition [104-106]. To decipher between which of these alleles were more causative it was noted that patients carrying the DR4 haplotype contain the DQB1*0302 allele very frequently and is found less in neutral haplotypes whereas DQA1*0301 is more commonly found in the general population. This along
with other studies led to the notion that the DQB1*0302 allele was a more favored candidate for increased diabetes susceptibility [87]. The studying of DQB1*0201 as an independent predisposing locus to diabetes has been difficult since it is in very high linkage disequilibrium with the DR3 haplotype, but it is also located on the DR7 haplotype, namely in linkage with the DRB1*0701 allele and in this context is not predisposing [107, 108]. As such, it has been hypothesized that the increased risk of the DR3 haplotype is related the DQA1*0501 allele, but no studies have yet demonstrated its independent association in diabetes either [87].

Establishing a mechanism of action of these predisposing alleles has been quite difficult since many theories exist but proving or disproving them is difficult. The more predisposing DQB1*0302 allele differs from the DQB1*0301 allele at position 57, with the *0302 lacking an aspartic acid residue at that position [109]. (Note that it has been substituted for alanine at that position). Interestingly, the NOD mouse has the mouse homologue of the DQB1*0302 gene as its DQ counterpart – in mice it is referred to as the I-A^b7 molecule. NOD mice are homozygous for this I-A^b7 molecule and it also lacks the aspartic acid residue at position 57 [109]. It should be noted, though, that the NOD mouse has its mouse homologue “DR” allele, I-E, completely inactive, as there is a deletion in its first exon of the α gene [110]. The exact role of this aspartic acid residue at position 57 is unknown, but it appears to be critical for peptide binding and recognition [111]. Other DQ-β alleles known to lack aspartic acid at position 57 are the DQB1*0201 allele, but as was mentioned earlier its role in diabetes susceptibility is of yet questionable [106, 112]. Furthermore, the importance of aspartic acid has recently been questioned after a study in Japanese patients indicated that a significant percentage of
patients carried the aspartic acid variation [113]. Two other residues have been associated with diabetes risk and they may also play a critical role in peptide binding. In the DQ-β molecule, it has been hypothesized that a combined variation at position 70 along with the aspartic acid variation at position 57 may result in a greater predisposition [87]. A variation in the DQ-α chain, more specifically at position 52, has also been associated with diabetes risk [107]. The difficulty with pinpointing a specific amino acid variation as the main causative factor, especially in a disease such as diabetes, is that as stated previously, diabetes is multifactorial polygenic disease and as such, what may be a predisposing genetic polymorphism in one race/culture of humans, may not be such a predisposing factor in other races due to differences in genetic background.

Interestingly, HLA alleles also confer protection from diabetes. One such allele that is very well known, and deemed “extremely protective” is DQB1*0602 [47]. It is very commonly present on the protective DR2 haplotype shown in table 1. Indeed approximately 20 percent of Americans and Europeans are at least heterozygous for the DQA1*0102-DQB1*0602 (DR2) haplotype, while less than one percent of children with type 1 diabetes carry these alleles [47]. In fact, it has also been shown that less than 1% of all diabetic individuals in Asian, African American and Mexican American populations carry this haplotype [87]. Three lines of evidence pinpoint the DQB1*0602 allele in the haplotype to play the protective role. First, the DQB1*0602 allele is the only MHC class II allele that is present on the DR2 protective haplotype while all the other alleles (DQA1*0102, DQA1*0103, DQB1*0601 and DQB1*0502 along with all the DRB1 alleles found in the DR2 protective haplotype) are also found in neutral or moderately predisposing haplotypes [47, 87]. A second interesting point of evidence
results from the finding that a few type 1 diabetic patients that have been shown to carry the protective DR2 haplotype have mutated DQB1*0602 alleles [47, 87]. The third line of evidence, is a study which demonstrated that the DQB1*0602 appears to be dominant over diabetes high risk HLA haplotypes and seems to prevent the onset of diabetes [47, 87]. Be that as it may, recently some diabetic subjects have been found that carry the DQB1*0602 allele, but the number of patients is very small.

The DRB1 alleles, although believed to be of less impact in diabetes susceptibility in comparison to the DQ alleles have been studied for sometime as well. The DRB1*0405 and 0401 alleles have been reported as predisposing. DRB1*0402 and DRB1*0401 as neutral and *0403, *0406 and *0407 appear to be protective [87]. Some studies have also reported that TAP1 and TAP2 as well as the DPB1 gene, all genes found within the class II MHC region, may play a role but more studies are required to substantiate this [114-116]. In addition, polymorphisms located within the class III MHC region, within and around genes such as HSP70 (heat shock protein 70) and TNF-α (Tumor Necrosis Factor-α) have also been associated with type 1 diabetes [117]. TNF-α is a strong candidate as it influences the inflammatory response substantially, and some recent studies have shown that it is associated with diabetes independently of MHC class II alleles since previous reports indicated that association may simply be a result of the short chromosomal distance between the class II genes and TNF [100, 102].

1.5.6 HLA Predisposing Alleles and Mechanism(s) of Action

As mentioned previously, elucidating the mechanism(s) of how specific MHC alleles result in diabetes susceptibility is a very difficult task. Protective HLA alleles have been hypothesized to have a higher affinity for one or several peptides in
comparison to predisposing alleles. Such a hypothesis was suggested by the observation that the I-A\textsuperscript{b7} molecule (lacking the aspartic acid residue) in the NOD mouse, which, as stated earlier, is the mouse homologue of the human DQB1*0302 molecule, is a very poor peptide binder of the suspected diabetogenic antigens [118]. Studies have in fact confirmed that the more protective DQA*0102 and DQB1*0602 have differing peptide affinities and peptide specificities from the predisposing DQA1*0301 and DQB1*0302 alleles. Peptide antigens in this study also consisted of some of the primary candidate autoantigens suspected in diabetes including insulin and GAD [119, 120]. Similarly, affinity differences have been reported in predisposing vs. protective DR molecules with those more protective having a stronger affinity for the peptide [121]. How exactly this increased affinity and binding specificity may prevent diabetes is unclear. One possible mechanism is that the deletion of autoreactive T cells in the thymus is faulty due to the poor presentation of antigen to the T cells in the thymus as a result of the weaker DQ/DR-peptide interaction. Poor presentation may not only affect T cell deletion via presentation of peptide within the thymus but T cell activity may also be regulated by presentation of peptides via APCs in the peripheral tissues. Although such theories are plausible, it is important to point out that there is as of yet no evidence definitively linking peptide affinities and Type 1 diabetes susceptibility. That being said, there is some good evidence suggesting that diabetogenic MHC alleles represent a loss of function as it has shown that transgenic mice over-expressing the very same diabetogenic alleles are protected from Type 1 diabetes [122].

Recently, an article in Nature Immunology compared the structures of the DQ8 and DQ2 molecules along with the I-A\textsuperscript{b7} molecule, from mice, using X-ray
crystallography in the presence of insulin, more specifically residues 9-23 from the B chain [123]. Looking back at section 3.2, DQ8 and DQ2 are the DQ haplotypes most commonly found on the highly predisposing DR3/DR4 haplotypes. According to the authors, all three molecules have a similar three-dimensional structure of the peptide binding pockets and the authors suggest that the similar structure of the pockets results in a specific conformation of antigen presentation, which results in significant diabetes susceptibility. More studies will have to be conducted to confirm this, which should of course include crystallization, and X-ray crystallography studies of neutral and protective class II MHC molecules.

1.5.7 IDDM2, The Insulin Variable Number of Tander Repeat (VNTR) Polymorphism and Type 1 Diabetes

"IDDM2 maps to a 4.1 Kb region containing the INS gene and its flanking regions which contain several polymorphisms in linkage disequilibrium[124, 125]. The bulk of the Type 1 diabetes susceptibility at this locus has been attributed to the VNTR 365bp upstream of the transcriptional start site of insulin [90, 91, 126-129]. The VNTR’s association with and role in Type 1 diabetes will be discussed in depth in section 1.6.

1.5.8 PTPN22, CTLA-4 and Type 1 Diabetes

"PTPN22, located on chromosome 1p13 encodes the LYP protein, a lymphoid tyrosine phosphatase. The polymorphism associated with Type 1 diabetes appears to be a missense mutation that changes an arginine at position 620 to a tryptophane and thereby abrogates the ability of the LYP molecule to bind to the signaling molecule CSK [95, 96]. The LYP-CSK complex normally downregulates T cell receptor signaling, the lack of
this downregulation results in enhanced T cell receptor signaling and T-cell hyperactivity. Consistent with a general effect on immune function is the finding that the tryptophane encoded allele is associated with a series of autoimmune disorders including rheumatoid arthritis [96], lupus erythematosus [97] and Type 1 diabetes [130].

The CTLA-4 gene is a strong candidate for autoimmune diseases since it encodes a molecule that functions as a key negative regulator of T cell activation. Susceptibility to various autoimmune diseases, including Grave’s disease, Addison’s disease and Type 1 diabetes was mapped to a polymorphism in the non-coding 6.1 kb 3’ end, more specifically a G/A polymorphism at +6230 (with respect to the transcription start site of CTLA-4) with the A polymorphism, possibly [131], associated with higher messenger RNA levels, of the soluble form of CTLA-4, along with higher numbers of CD4+CD25+ regulatory T cells and decreased susceptibility to autoimmune diseases [132-135]. The functional mechanism resulting in these benefits of the A polymorphism is as of yet unknown. It should be noted that ctla-4 is also a candidate susceptibility gene in the NOD mouse (mouse iddm5) but just as in humans, a mechanism has yet to be determined. Lastly, it should also be noted that some studies suggest that CTLA-4 plays a bigger role in Grave’s disease susceptibility than in Type 1 diabetes susceptibility both in humans and the NOD mouse [136].

1.6 Thymic Insulin Expression, Insulin VNTR and Thymic Central Tolerance Induction

1.6.1 The Insulin Variable Number of Tandem Repeat (VNTR) Polymorphism

IDDM2 was the second susceptibility locus identified and was mapped to a 4.1 kb region on chromosome 11p15 which contained, within its span, the gene for insulin
(INS), a Variable Number of Tandem Repeat (VNTR) polymorphic sequence, located 365bp upstream to the transcriptional start site of insulin (or 596bp upstream of the translational start site), and 10 other polymorphisms all in strong linkage disequilibrium [90, 91, 126-129] (Figure 4). Subsequent association and linkage analysis studies narrowed down the diabetes susceptibility to the VNTR [90, 91, 126-129]. Because of its close proximity to insulin it is usually referred to as the INS-VNTR. Cloning and sequencing of the INS-VNTR alleles revealed that they were tandem repeats of 14-15 bp polymorphic units, variants of the general consensus sequence ACAGGGGTGTGGGG [137-139]. At present, there are 18 repeat elements known [140] (see table 4 and Chapter 6). As more INS-VNTR regions were sequenced they were segregated into three classes depending on the number of repeat elements with the class I alleles consisting of 26-63 repeats, class III alleles classified as greater than or equal to 140 repeats and the rare class II alleles, which were of a length in between class I and class III alleles [141]. Further studies went on to show that the class I alleles, although the more frequent alleles in the general population, were predisposing to Type 1 diabetes whereas the majority of class III alleles were protective [90, 91, 127, 128, 142]. The general consensus from these studies was that class I homozygotes were at greatest risk for diabetes, followed by class I/III heterozygotes followed by class III/III homozygotes and that on the whole the IDDM2 locus accounted for approximately 10% of the Type 1 diabetes genetic susceptibility. It should be noted that although many studies demonstrated the class I alleles were predisposing and that the class III alleles protective, there were occasional class III alleles identified that seemed to result in increased susceptibility and
Figure 4: Schematic diagram portraying the Tyrosine Hydroxylase (TH) – Insulin (INS) - Insulin Like Growth Factor-2 (IGF-2) Region on human chromosome 11p15 also showing the 10 polymorphisms within the 4.1kB region. Open, closed and hatched boxes depict introns, exons and untranslated regions, respectively. Polymorphisms are designated with respect to the first base of the initiating ATG [91]. Reprinted by permission from Maxmillan Publishers LTD: Nature Genetics, 2005.

Table 4: Known tandem repeat elements of the INS-VNTR with the general consensus sequence in row a.
rare class I alleles that seemed to be protective - such alleles were, however, very infrequent [90, 91, 129]. It should also be noted that susceptibility based on class II alleles was not assessed because of the rarity of such alleles in Europeans.

With the susceptibility of Type 1 diabetes now linked to the class of the VNTR, a causal mechanism resulting in susceptibility was sought. A number of hypothesis were put forward on how the VNTR, an un-transcribed DNA sequence, could potentially regulate expression patterns of genes in close proximity. Of course, the closest and most obvious gene implicated was insulin, but two other genes in close proximity were exon 13 of tyrosine hydroxylase and exon 1 of insulin like growth factor-2 (IGF-2) - both of which were within 50kB of the VNTR. With this in mind, a number of groups began to demonstrate that the VNTRs acted in cis to modulate insulin transcription. More specifically, such studies demonstrated that insulin mRNA transcription, in both adult and fetal pancreatic tissue, was slightly increased from INS genes in cis with a class I allele in contrast to INS genes in cis with class III alleles [91, 142]. The difference was estimated as approximately 20-30%. Albeit an interesting finding, it was also quickly pointed out that such differences were largely insignificant and there was no clear explanation as to how such a difference could possibly predispose to Type 1 diabetes. Interestingly, the detection of pancreatic specific genes in mouse thymus, more specifically proinsulin, led two groups, Vafiadis et al. and Pugliese et al., to study the same phenomenon in human thymus [141, 143]. Using human fetal thymus samples, the two groups independently reported finding proinsulin mRNA in all thymi analysed and in addition they demonstrated that proinsulin mRNA levels varied in cis with the class of the VNTR. Both groups demonstrated a 2-3 fold increase in proinsulin mRNA when the
**INS** gene was in cis with a class III VNTR as opposed to a class I VNTR. They suggested that the protective effect of the class III alleles was linked to this higher level of thymic proinsulin expression which possibly played a role in thymic central tolerance induction. More specifically, they hypothesized that the more proinsulin was expressed by the tolerizing cells the more likely proinsulin specific T cells would be exposed to it and undergo negative selection. Of particular interest was that both groups also came across rare thymic samples in which proinsulin mRNA from an **INS** gene in cis with a class III VNTR could not be detected. These samples were class I/III heterozygotes (as were all samples in the study) and proinsulin mRNA was clearly detectable from the **INS** gene in cis with the class I VNTR. Sequencing of the insulin promoter and a portion of the 5' region of the **INS** gene itself, in cis with the class III allele, did not reveal any mutations that could explain the findings and the two groups hypothesized that such rare class III VNTRs were directly insulin silencing and were likely the Type 1 diabetes predisposing class III VNTRs detected in previous studies. Both groups suggested that heterozygotes for these alleles (ie. class III-S, s for silencing/class I alleles) have a higher degree of susceptibility than class I homozygotes since thymic proinsulin mRNA levels would be even less in such individuals. Indeed, in subsequent studies, in Vafiadis et al. we went on to demonstrate that the class III-S VNTRs were more frequently transmitted to children suffering from diabetes than to non-affected children [140] (see Chapter 6). We also sequenced a class III-S VNTR in the hopes of finding a sequence variation that could explain the monoallelic silencing of **INS** but the differences in sequence between the class III-S VNTR and an insulin “enhancing” class III VNTR were very subtle with only minor variations. It must be acknowledged that attempting to describe how the
The hypothetical mechanism relating thymic insulin levels to Type 1 diabetes susceptibility is a mechanism that is plausible in humans because of the \textit{INS-VNTR}. Mice and rats do not have an \textit{INS-VNTR} but a similar mechanism could also be possible in the NOD mouse as one study reported that thymic proinsulin mRNA levels were reduced in the autoimmune diabetes prone NOD mice versus wild type mice—an approximately 2 fold difference [148]. Such a differential expression, if confirmed, must be determined in \textit{trans} because neither of the two mouse insulin genes maps to a NOD locus. A follow-up to this concept was conducted by my supervisor’s lab using \textit{Ins2} and \textit{Ins1} knockout mice that were engineered to contain between 1 and 4 functional \textit{Ins} gene copies [27]. Removal of three of the \textit{Ins} gene copies (leaving only one functional \textit{Ins1} copy) did not alter pancreatic proinsulin expression levels at the mRNA or protein level, as expected from the known metabolic feedback regulation of pancreatic Ins transcription (section 1.2.3). The same genotype resulted in a marked decrease in thymic proinsulin
mRNA and protein (approximately 20 fold decrease in mRNA and 10 fold decrease in protein), suggesting that the thymus is not subject the same metabolic feedback regulation. The group also demonstrated that in mice with both INS2 copies knocked out (resulting in approximately 20% of the normal thymic proinsulin levels) the specific decrease in thymic, and not peripheral or systemic insulin levels resulted in detectable peripheral proinsulin specific T cell reactivity. Against a non-diabetes susceptible background, none of the thymic proinsulin deficient mice developed autoimmune diabetes. Interestingly, however, the Ins2 knockout against the NOD background accelerated the onset of and exacerbated the incidence of autoimmune diabetes [44]. As mentioned previously, in the pure NOD line overt diabetes can be detected in 5% of animals by 14 weeks of age and 77% of the female mice and 28% of the male mice develop autoimmune diabetes by 27 weeks of age. In contrast 100% of females and 100% of the males developed autoimmune diabetes by 15 and 21 weeks of age, respectively, in the Ins2 deficient NOD line. Such studies clearly indicated a role for thymic proinsulin expression/levels in the susceptibility to autoimmune diabetes in animals and strongly corroborated a similar role for the INS-VNTR in humans.

1.6.2 Thymic Insulin Expression

Up until approximately a decade ago, the importance of thymic central tolerance induction, as described in section 1.3.7, was believed to be limited to ubiquitously expressed antigens. That is to say, peripheral tissue specific self-antigens such as insulin or myelin basic protein were not believed to be expressed in the thymus and tolerance induction towards such self-antigens was believed to be induced by other mechanisms -
ie. peripheral tolerance. Serendipitous discoveries by both Hanahan [149] and Heath et al. [150] challenged this notion when they managed to induce tolerance towards the SV40 large T-antigen and the K^b antigen, respectively, when these antigens were expressed in transgenic mice under the regulation of the rat insulin 2 promoter. Using the SV40 transgenic mice, Jolicoeur et al. [151] went on to demonstrate that SV40 large T-antigen mRNA was detectable in the thymus of these mice and that the mRNA of a number of pancreatic specific genes such as, proinsulin, glucagon, pancreatic polypeptide, somatostatin, trypsin, elastase and GAD67 was detectable in the thymus of non-transgenic mice. Antonia et al. [152] further went on to demonstrate that double transgenic mice expressing the SV40 large T-antigen, under the regulation of the elastase promoter, which were also transgenic for SV40 large T-antigen specific TCRs underwent significant deletion and anergy of their T antigen specific T cells. Using mouse chimeras, this group identified the necessary tolerizing element as thymic epithelium. Smith et al. [153] took the studies one step further and demonstrated that non-transgenic mice could be tolerized against the SV40 large T antigen by transplantation of transgenic thymus into non-transgenic recipients. However, in contrast to Antonia et al., Smith et al., using thymic tissue digestion, FACS sorting and RT-PCR studies of various thymic cell isolations, came to the conclusion that the thymic cells responsible for the expression of the peripherally restricted antigens were low density MHCII positive cells hypothesized as being either dendritic cells, macrophages or thymic epithelial cells (in decreasing order of probability).
1.6.3 Thymic Central Tolerance Induction

By the beginning of the second millennium, it was not only shown that proinsulin mRNA and protein could be detected in fetal and childhood thymus but so could mRNA for GAD65/67, ICA69 and I-A/2 [143]. It should be noted that although proinsulin protein is detectable, there is currently no existing evidence for the conversion of proinsulin to insulin in the thymus. In 2001, Pugliese et al. [154] released the results of a study on human thymus, spleen and peripheral lymph nodes (tissues were from fetuses, children and adults) demonstrating that mRNA for proinsulin, GAD65/67, ICA69 and I-A/2 were detectable in all tissues. They also showed that a small number of cells in each of these tissues stained positively for proinsulin, GAD65/67, ICA69 and I-A/2, some cells staining positively for more than one antigen. Double staining with markers such as CD83, CD11c, CD40, CD14 and MHC II revealed that these cells were APCs, more specifically dendritic cells and macrophages. They were never able to co-localize any of the pancreatic antigens with cytokeratin – a marker of epithelial cells. In addition, co-staining experiments with the antibody for autoimmune regulator protein (AIRE) were also negative. At the time, Aire, was a recently discovered transcription factor expressed by thymic medullary epithelial cells and was known to play an important role in tolerance induction [155]. Indeed, humans with mutations in the AIRE gene were found to develop a multiorgan autoimmune disease known as Autoimmune Polyendocrine Syndrome type 1 (APS-1) [156, 157].

The results described by Pugliese et al. were unfortunately not reproduced by many groups including our group that specifically attempted to detect proinsulin mRNA in peripheral blood monocytes without success (unpublished data). That being said, less
than one year later, Derbinski *et al.* [158] published dramatically different observations from those obtained by Pugliese *et al.* while studying mouse thymus. Essentially using a combination of thymus tissue digestion, density gradient separations and FACS sorting, RT-PCR and antibody staining, Derbinski *et al.* reported that pancreatic specific antigens, including proinsulin, were expressed by thymic cells with the phenotype G8.8+/CD45-lo/CDRI-, a phenotype characteristic of medullary thymic epithelial cells (mTECs). In addition, they reported detecting an additional 20 non-pancreatic peripheral tissue specific antigens that were exclusively expressed by mTECs and only a small number (~6) of these tissue specific antigens were concurrently expressed by cortical epithelial cells (cTECs), dendritic cells or macrophages. Proinsulin was solely expressed by mTECs. They also determined, using antibody staining, that only a small portion of the mTECs expressed the antigens — ie. proinsulin was produced by approximately 1-3% of all mTECs. They could not determine if individual mTECs produced more than one antigen and they confirmed our findings that proinsulin mRNA (and that of GAD and I-A/2) could not be detected in peripheral lymph node cells.

In an attempt to determine which of the findings, either Pugliese *et al.* or Derbinski *et al.* were correct our group set out to identify which cells in the thymus specifically produced proinsulin using transgenic mice containing a functional *LacZ* gene knocked into the *Ins2* promoter [159, 160](see Chapter 3). Using the thymus digestion and FACS isolation methods described by Derbinski *et al.* and Crowley *et al.* [161] to isolate TEC and thymic dendritic cells, we found that although a very faint proinsulin band could be obtained in dendritic cell isolates, by RT-PCR, the signal was consistently and significantly greater in TEC isolates. Additionally, we also showed by X-gal
immunocytochemistry and immunohistochemical staining for TEC specific markers and dendritic cell specific markers, that X-gal activity was clearly detectable in mTECs but not in dendritic cells. Of interest, as well, was that the X-gal staining was often localized in medullary epithelial cells often located within the Hassall’s Corpuscles. Lastly, we demonstrated that a population of β-Galactosidase producing mTECs could be identified by FACS using an enzymatic substrate for β-Galactosidase, 5-dodecanoylaminofluorescein di-D-galactopyranoside (C12FDG), that emits fluorescein green fluorescence upon cleavage – such experiments indicated that between 1-3% of all mTECs produced proinsulin. These published findings quite convincingly implicated mTECs as the proinsulin producing cells of the thymus and suggested that these were also the cells likely responsible for expression of other peripheral tissue specific antigens.

It should be noted that although the work by Derbinski et al. and our group refutes the findings of Pugliese et al., some recent albeit disputed findings have suggested that although mTECs can produce the self-antigens, the processing and presentation of these antigens by mTECs may be limited to CD8+ T-cells and that efficient presentation to CD4+ T-cells requires self-antigen uptake and presentation by dendritic cells [162-164]. In conjunction with this, some recent studies also showed that the presentation of antigen to thymocytes via dendritic cells in the immediate proximity of Hassall's Corpuscles appears to be involved in the generation of regulatory T cells [40]. Such studies may offer an explanation for the discrepancy between Derbinsky et al. and our results with that of Pugliese et al., at least at the protein level.

Following the identification of mTECs as the promiscuous antigen presenting cells of the thymus, much work was published that contributed details regarding the
phenotypic characteristics of the antigen presenting cells and the importance of their expression. With the engineering of Aire deficient mice, Anderson et al. [165], demonstrated that these mice developed very similar pathologies as humans with the APS-1 disorder – including the development of autoimmune diabetes. Using microarray analysis (of a limited set of genes), and real-time PCR, Anderson’s group demonstrated that Aire deficiency resulted in decreased thymic expression of over 100 genes that included a significant subset (>30) of peripheral tissue specific antigens expressed in the thymus including proinsulin. Other affected peripheral tissue specific antigens, such as salivary protein 1 and retinal S antigen could be directly implicated in the autoimmune pathologies observed in these mice – lymphocytic salivary gland infiltration and uveitis, respectively. Using bone marrow chimeras, the group went on to demonstrate that the specific defect in these animals was localized in the thymic epithelial cells in accord with the previous findings that the bulk of Aire expression is localized in mTECs [155, 158]. The authors thus concluded that AIRE’s role was to upregulate expression of a number of tissue specific antigens to maximize tolerance induction. Interestingly, subsequent studies by the group (reviewed in [166]) revealed that regulatory T cell function and numbers were normal in these mice suggesting Aire deficiency directly impacted autoreactive T cell deletion. The most recent study from the group has additionally revealed that tolerance induction in Aire deficient animals is not only lost towards antigens whose levels are decreased following Aire mutation but also towards antigens whose levels were unaffected – the preliminary work from the study suggests AIRE not only induces expression of various tissue specific antigens but also plays a role in their physical presentation.
It should be noted that even though the bulk of Aire expression has been localized to mTECs it is known that not all mTECs express Aire [167, 168]. Indeed, it is also known that mTECs are quite heterogeneous in their expression of a number of surface molecules including MHC II, CD80 (existing as low expressing and high expressing mTECs) and the binding sites for the lectin UEA-1 (note that the Ulex europaeus agglutinin I (UEA-1) binds glycoproteins and glycolipids containing α-linked fucose residues) [169]. With this in mind and using murine thymus tissues, real-time PCR and microarray analysis, Derbinsky et al., [170] attempted to thoroughly characterize the importance of AIRE and the self-antigen expressing characteristics of the CD80^{lo} and CD80^{hi} subsets of mTECs (the CD80^{hi} subset concomitantly expresses Aire which usually co-localizes with MHC II+, UEA-1+ cells). As an initial analysis they found that mTECs over-express, in comparison to cTECs (their most similar cell type for comparison), over 545 genes, approximately 152 (28%) of which could be classified as tissue specific (i.e. tissue specific expression was arbitrarily designed as expression of a unique self-antigen in less than 5 tissues). Self-antigens from virtually every tissue in the body were detected. Segregation of the pooled cells into CD80^{hi} and CD80^{lo} subsets revealed that approximately 500 genes were over-expressed in the CD80^{hi} subset, when compared to the CD80^{lo} subset, and that 300 genes were over-expressed in the CD80^{lo} subset when compared to the CD80^{hi} subset. The CD80^{hi} subset also over-expressed a larger number of tissue specific antigens relative to the CD80^{lo} subset (33% versus 14%). The same analysis was conducted in Aire deficient mice and CD80^{hi} subsets showed a reduction in the total number of over-expressed genes by 25% (125 genes) and the relative percentage of tissue specific antigens decreased to 21%. As Aire is not expressed
in CD80\textsuperscript{lo} cells, there was virtually no change in the expression characteristics of the CD80\textsuperscript{lo} subset. These results, along with previous studies co-localizing CD80\textsuperscript{hi} expression with Aire expression, UEA-1 binding and MHC II positivity led the group to hypothesize that this subset of cells was a more mature/differentiated subset capable of widening their scope of peripheral tissue specific antigen expression. Interestingly, mTECs of the Hassall’s Corpuscles are hypothesized to be the most mature/differentiated form of mTECs [171, 172].

The mechanism allowing mTECs to over-express numerous peripheral tissue specific antigens is still a mystery. Recently, a few studies, using a combination of microarray analysis and computational/statistical analysis on subsets of mTECs, have suggested that chromosomal clustering is a characteristic of over-expressed mTEC genes [170, 173, 174]. These studies, done on both murine and human tissues, have identified clusters usually between 3-5 genes in length on virtually every chromosome – even clusters as long as 16 genes in length have been identified. Indeed, chromosomal clustering of expressed tissue specific genes is known to take place in peripheral tissues and is believed to be important for the efficient transcription of genes [175, 176]. These groups have thus suggested that an elaborate mechanism resulting in the expression of numerous clusters of genes is occurring in mTECs without any need for the tissue specific transcription factors normally required by the peripheral tissues. They hypothesize events such as chromosomal re-modelling as potentially important processes that could allow for this expression. As discussed above, the transcription factor \textit{AIRE} plays an important role in the expression of a subset of self-antigens. Indeed, \textit{Aire} deficiency has been shown to result in the reduction in the number of clusters and the
number of genes expressed per cluster but it does not completely abrogate the expression of all clusters suggesting that other as of yet unidentified mechanisms are necessary for this phenomenon to occur [174].
2.0 CHAPTER TWO: Statement of Purpose and Objectives of the Research

Three major objectives and one minor objective were proposed for the doctoral research project which are described in detail below. As a whole, the objectives were aimed at improving our current understanding of the role of thymic central tolerance induction, more specifically the role of thymic proinsulin expression, in insulin self-tolerance and Type 1 diabetes risk. Such knowledge would not only help in building strategies to potentially identify individuals predisposed to Type 1 diabetes but would also likely allow for the formulation of therapeutic interventions that could potentially prevent the disease. Although Type 1 diabetes was the major disease focus, it is important to note that the study designs necessary to accomplish the objectives also generated and provided much information about central tolerance induction towards other autoimmune diseases.

The first major objective (Chapter 3) was to identify the proinsulin producing cells of the thymus. As described in the comprehensive literature review, section 1.6 – 1.6.3, there was considerable opposing data in the literature stipulating which cells of the thymus produce proinsulin [149, 150]. At the onset of the project, the most recent work attempting to answer this question was by Pugliese et al.[174] implicating dendritic cells and Derbinsky et al.[155] implicating thymic medullary epithelial cells (mTECs). Our approach consisted of using Ins2KO mice engineered to have Ins2 knocked out and replaced with a LacZ gene capable of producing functional β-Galactosidase enzymes [157]. Using a combination of thymic tissue digestion, FACS sorting, RT-PCR, immunocytochemistry and immunohistochemistry we confirmed that the proinsulin
producing cells of the thymus were medullary epithelial cells often located in Hassall's Corpuscles and that proinsulin producing cells consisted of 1-3% of all mTECs.

It should be noted that this first objective was initially attempted by creating Rat Insulin Promoter-2 – Enhanced Green Fluorescent Protein (RIP2-EGFP) transgenic mice since RIP2 was shown to effectively express self-antigens in the murine thymus [146-148] and usage of EGFP, via FACS and microscopy, as opposed to the usage of an immunocytochemical based assay, was expected to greatly facilitate identification and subsequent isolation of the proinsulin positive cells. Unfortunately, two attempts were made at creating transgenic mice with functional RIP2-EGFP plasmids (as determined by in vitro expression studies) but RT-PCR and microscopy analysis, for EGFP expression, revealed none of the ~ 20 transgenic founders expressed the EGFP. It was later determined, by a personal communication from Dr. Christian Sirard, that EGFP expression, but not Green Fluorescent Protein (GFP), can be very difficult to stably acquire when engineering transgenic mice likely due to some form of silencing mechanism. Overall, this work encompassed two years of the thesis.

The second objective (Chapter 4) was to isolate and culture individual colonies of the proinsulin producing mTECs. The purpose of this objective was to develop an in vitro model to unequivocally study the small 1-3% of all mTECs that produce proinsulin. Points of interest, upon establishment of the lines, would subsequently include studying the transcriptional and translational regulation of proinsulin along with the potential to study the regulation and processing of the translated proinsulin protein. In addition, transfection of the cells with Insulin-VNTR constructs could facilitate the study of the VNTR’s mechanism of action. Such cell lines could also be used, as part of a
collaborative effort, to study T cell/mTEC interaction in vitro and potentially in vivo as well. As described in Chapter 5, using the F1 generation of Ins2KO mice crossed with mice expressing a temperature-sensitive mutant of the SV40 large T-antigen, under the regulation of the MHC class I promoter (Immortomice) [175], we established two proinsulin producing mTECs cell lines and, for comparison, two proinsulin-negative mTEC lines.

The third objective (Chapter 5) involved characterization of the cultured mTEC cell lines by microarray analysis. Such an analysis was designed to determine which peripheral tissue specific self-antigens were being produced by the isolated mTEC cell lines (from the 2nd objective) as presently it is unknown if individual mTECs produce one, some or all of the tissue specific antigens known to be expressed in the thymus. Such an analysis would also allow us to confirm recent reports describing chromosomal clustering of peripheral tissue specific genes expressed in the thymus [167, 170, 171]. In addition, microarray analysis would allow us to begin elucidating a mechanism for proinsulin (and other peripheral tissue specific self-antigens) expression by identifying required transcription factors and other potentially required regulatory molecules.

The fourth minor objective (Chapter 6) involved the sequencing of two class III Variable Number of Tandem Repeat (VNTR) alleles of ~2.1kB in length. More specifically, sequencing of a class III-s (s for thymic insulin silencing) and an insulin "enhancing" class III VNTR was undertaken following the determination that the class III-s alleles are found in cis with insulin genes that are monoallelically and inexplicably silenced at the level of the thymus [138, 140]. Such alleles have been shown to be predisposing to Type 1 diabetes [137]. It was hypothesized that a variation in sequence
could explain the silencing characteristics of the class III-s allele. Sequencing of the alleles was accomplished by creating linearized overlapping segments of each of the alleles using an exonuclease enzyme. Only minor sequence variations were found between the two alleles, and with the current limited knowledge of the functionality of the VNTRs a simple explanation could not be provided for the silencing phenomenon.
"Imagination is more important than knowledge..."

Albert Einstein
3.0 CHAPTER THREE: Proinsulin Expression by Hassall's Corpuscles in the mouse Thymus

Aziz Alami Chentoufi*, Michael Palumbo* and Constantin Polychronakos
*Both authors contributed equally to the work


3.01 Contribution of Authors

Aziz Alami Chentoufi: Experimental design and strategy for thymus extraction, specific cell isolations (dendritic cells and epithelial cells) and β-Galactosidase immunohistochemistry.

Michael Palumbo: All experimental work including cell isolations, RT-PCRs and immunohistochemical staining. Experimental design and strategy for RT-PCRs, G8.8 and cd11c double staining with β-Galactosidase immunohistochemistry, fluorescence assisted cell sorting (FACS) of double stained G8.8+ and β-Galactosidase positive cells.

Dr. Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

3.02 Objective

To identify the proinsulin producing cells of the thymus using transgenic mice expressing β-Galactosidase under the regulation of the Ins2 promoter and β-Galactosidase activity assays.
3.1 Abstract

The thymus expresses proinsulin, among many other tissue-specific antigens, and the inheritance of genetically determined low thymic proinsulin expression has been associated with impaired proinsulin-specific autoreactive T cell tolerance and type 1 diabetes susceptibility. The cellular and molecular biology of proinsulin expression in the thymus remains unknown and contradictory reports exist regarding the identity of proinsulin-producing cells. Using knock-in mice expressing β-galactosidase under the control of an endogenous insulin promoter, we found that thymic proinsulin and β-galactosidase transcripts were detectable at high levels in purified thymic epithelial cells. Immunohistochemical analysis of β-galactosidase activity showed that most proinsulin expression can be accounted for by rare medullary epithelial cells of the Hassall’s Corpuscles. Moreover, flow cytometry analyses of β-galactosidase positive cells showed that only 1-3% of all epithelial cells express proinsulin and this technique will now provide us with a methodology for isolating the proinsulin producing cells in mouse thymus.
3.2 Introduction

Type 1 diabetes (T1D) results from autoimmune destruction of the insulin-producing pancreatic β-cells by autoreactive T-lymphocytes (1). The inherited immunological defects responsible for this disease are polygenic in nature. Besides the major histocompatibility complex (HLA) on chr. 6p21 (2-4) one of the few genetic loci that have been confirmed and functionally studied maps to the insulin gene regulatory region, containing a variable number of tandem repeats (VNTR) polymorphism. VNTR alleles have little effect on pancreatic insulin expression but in the thymus the predisposing class I alleles correlate with low and the protective class III alleles with high levels of insulin expression (5,6). Low expression of insulin in the thymus of Non-Obese Diabetic (NOD) mice has been reported (7) and may play a role in diabetes susceptibility. More importantly, we have recently shown that mice with thymus-restricted insulin deficiency showed a strong peripheral proinsulin-specific T cell reactivity when compared to control animals (8). This corroborates our hypothesis (5) that genetically determined low thymic insulin expression levels predispose to human type 1 diabetes through less efficient insulin-specific autoreactive T cell selection. In addition, it has recently been shown that NOD mice with drastically reduced insulin expression in the thymus present an accelerated type 1 diabetes development (9).

The thymus is a central lymphoid organ responsible for the generation of the T cell repertoire, during which those T cells that recognize self-antigens must be eliminated during T cell development in the thymus by negative selection through clonal deletion
The importance of negative selection for tolerance to tissue-specific autoantigens such as insulin had, until recently, been in doubt. Such antigens were traditionally thought to be unavailable for presentation in the thymus and tolerance to them was believed to occur by extrathymic peripheral mechanisms such as anergy, ignorance or deletion (14-16). Growing evidence suggests that this may not be the case and an increasing number of proteins previously considered to be strictly tissue-specific can be detected in the thymus. These include hormones, transcription factors and secreted proteins. Many of the ectopically expressed antigens (proinsulin, thyroglobulin, myelin basic protein and retinal S-antigen) are associated with organ-specific autoimmune disease (type 1 diabetes, thyroiditis, multiple sclerosis and uveitis, respectively) (17-19). However, the expression of these self-antigens is limited to a very small subset of thymic cells and studying how these specific cells result in the deletion of a wide variety of autoreactive T cells has not been straightforward.

The thymus is anatomically divided into subcapsular, cortical, and medullary compartments. Thymic stroma contains a variety of professional antigen-presenting cells (APCs) including bone marrow-derived dendritic cells (DC), macrophages and B-cells and endoderm-derived cortical and medullary epithelial cells (cTECs and mTECs, respectively) (20). A striking morphological feature of the medulla is the presence of Hassall’s Corpuscles (HC), which consist of concentric whorls of stratified keratinizing epithelium, and share antigenic properties with ectodermic epithelium (21). Well developed in human thymus, they are small and sparse in murine thymus. The function of Hassall’s Corpuscles remains unknown, although recent studies, in human thymus, have
demonstrated that they are the only regions in the medulla to contain apoptotic T cells (22). In addition to this, the medullary epithelial cells making up the Corpuscles have been shown to express a leukocyte and hemotopoietic attractant, stromal derived factor-1 (SDF-1), and CD30-L – a membrane associated glycoprotein related to tumor necrosis factor (TNF) that is involved in T-cell signaling (23,24). Moreover, it has also been shown that CD30 deficient mice contain elevated numbers of thymocytes and show a gross defect in negative but not positive selection (25). Together these results suggest that the Hassall’s Corpuscles are central to negative T cell selection.

In mice, conflicting reports exist regarding the identity of proinsulin producing cells in the thymus (19, 26). To provide a definite answer, we used knock-in mice expressing β-galactosidase (β-Gal) under the control of endogenous Ins2 promoter. We show that most of insulin-promoter directed transcription can be accounted for by epithelial cells in Hassall’s Corpuscles which can be isolated by flow cytometry using a fluorescent β-galactosidase substrate.
3.3 Methods

**Animals:** Ins1-KO and Ins2-KO mice were a gift of Dr. J. Jami (Institut Cochin, Paris, France) and were generated as described by Duvillié et al. (27). Ins2-KO mice have their Ins2 gene replaced by a functional LacZ gene that is under the regulation of the endogenous Ins2 gene promoter, while Ins1 was disrupted by introducing a neomycin cassette, against the same genetic background. All mice were bred at our animal facility under conditions specified by the Canadian Council of Animal Care.

**RNA preparation and cDNA synthesis:** Pancreatic and thymic RNA from 2 to 4 week-old mice was isolated by Trizol (Gibco, Rockville, MD) and treated with DNAse (Gibco). RNA was isolated from enriched thymic cell populations with RNeasy Mini Kit (Qiagen Inc. Mississauga, Ont.) with typical yields of RNA between 50-400ng. An equal amount of each RNA (2.5 μg of tissue-extracted and 50-100ng of single cells or more as with the case of DC) was reverse-transcribed using random primers and Superscript II Reverse-Transcriptase (Gibco). Parallel samples in which Reverse Transcriptase was omitted, were always included to confirm the absence of DNA contamination.

**Quantitative and semi-quantitative RT-PCR:** Insulin expression in the thymus and pancreas extracted-tissues was analyzed by quantitative RT-PCR. An equal amount of each cDNA sample (2μl from thymic and 2μl from 1/300 diluted pancreatic cDNA) was then added to serial dilutions of the insulin competitor (56, 28, 14, 7, 3.5, 1.7 or 0 amol/tube) or with the cyclophilin competitor (2.5, 1, 0.5, 0.25, 0.1 or 0 x10^9 copies number/tube) for loading control. The insulin competitor is an internally deleted cloned
insulin sequence amplified with the same primers but 45bp shorter, constructed as described by Forster (28). For cyclophilin analysis, we used the PCR Kit for cyclophilin (Quantum mRNA; Ambion, Austin, TX) as described by the manufacturer. The PCR products were resolved using PAGE and bands were quantified using Geldoc Software (Bio-Rad, Hercules, CA). The insulin concentrations were normalized for the amount of starting RNA using cyclophilin measurement as described (8). β-galactosidase expression in pancreas and thymus was analyzed by semi-quantitative RT-PCR. An equal amount of cDNA (2μl of diluted 1/1, 1/10, 1/100 cDNA samples from tissues and undiluted amount for cell preparations) was amplified by PCR. The thermal cycler conditions for β-galactosidase and insulin amplification were: 2’30 and 3’ minutes at 92-95°C followed by 40 cycles of 92-94 °C for 30-60 sec, 53-55 °C for 30-60 sec and 74-72 °C for 60-90 sec respectively. The following primer pairs were used (forward and reverse primers respectively): Insulin specific primers: 5’-GGCTTTCTTCTACACACCCCA-3’ and 5’-TACCAGCTGGAGAACTACTG-3’ and bacterial-specific β-galactosidase primers: 5’-TGGCGGACCGCTAT-CAGGAC-3’ and 5’-TGTATCATCTGGTCGCTGG -3’.

**Immunohistochemistry:** Organs from mice were harvested, embedded in Tissue-Tek (Sakura, Finetech, Torrance, CA) (OCT compound), snap-frozen in isopentane precooled in dry ice and sectioned in 10 μm thick sections.

For β-galactosidase activity, the sections were rinsed in 0.1 M phosphate buffer (PBS), fixed with 0.2% glutaraldehyde, 5mM EGTA and 2mM MgCl2 in PBS for 10min at room temperature, washed 3 times (5 min, 10 min and 15 min) in PBS then, stained with color solution (1mg/ml of X-gal, 5 mM of K4Fe (CN)6, 5 mM of K3Fe (CN)6 and 2mM MgCl2

68
in PBS) for 24-48h at 37°C. The sections were rinsed in PBS then counter-stained with Nuclear Fast Red (DAKO, Mississauga, Ont.) as described by the manufacturer.

For Immunohistochemistry, the sections stained with X-gal were rinsed, endogenous peroxidase activity was quenched for 5 min in 0.3% H₂O₂, saturated with normal goat or rabbit sera and then incubated with anti-CD11c (N418, Serotec, Raleigh, NC) and the anti-gp40 antibody (murine homologue of human EpCAM) G8.8 (DSHB, University of Iowa, IA) Abs for 30min. After three successive washings in PBS (3 x 5 min) the sections were incubated with affinity purified biotin-conjugated secondary antibodies goat anti-hamster IgG Ab (Vector Laboratories, CA) and rabbit anti-rat IgG (Vector Laboratories) respectively. Negative control sections were incubated only with secondary Abs. Slides were then incubated with Vectastatin Elite ABC reagent for 30 min (Vector Laboratories) using the horseradish peroxidase substrate diaminobenzidine tetrahydrochloride (DAB, Vector).

**Isolation of thymic cells:** Thymic DC were isolated as described by Crowley et al (29). In the first digestion, individual thymic lobes were perfused with 100 U/ml collagenase D (Sigma, Oakville, Ont.) and then incubated 15min at 37°C with 400U/ml collagenase D. The glands were then disrupted and all released cells were centrifuged at 1400rpm for 10min. The cell pellet was resuspended in bovine serum albumin (BSA 35%, ρ = 1.080. Sigma), overlaid with 2ml RPMI medium and centrifuged for 20min at 7500g. In an alternative protocol, thymi were minced and digested with two 15 minute incubations at 37°C with collagenase IV (Sigma), pelleted and resuspended in bovine serum albumin. The low-density floating fraction was collected, washed, and incubated for 90 min in a
Petri dish. The non-adherent cells were removed by pipetting. The monolayer was incubated for an additional hour to remove additional small round lymphocytes. Then the monolayer was cultured overnight, at which time most DC became nonadherent and were harvested. For further purification and enrichment for DC, the cells were sorted in a Vantage FACS fluorescent-activated cell sorter (Becton Dickinson and Co., Franklin Lakes, New Jersey, USA) for a fluorescence-activated cell sorter using anti-CD11cAbs.

To examine freshly isolated DC without the phenotypic changes that might occur during overnight culture, the low-density floating fraction was incubated in 25 mM EDTA/PBS at 37°C for 5 minutes and DC were then directly FACS sorted with CD11c alone, DEC205 (Serotech, NLDC-145) alone, or CD11c and I-A^b (MHC II) (Pharmingen, Germany) with essentially identical results.

Thymic epithelial cells were isolated according to Klein et al (30). Briefly, 4-5 thymi of 2-4 week old mice were digested for 30-40 min with Collagenase/Dispase (Sigma) (0.2 mg/ml each) and DNase (Roche, Mannheim, Germany) (25 μg/ml) in PBS-1%FCS. EDTA was added to a final concentration of 10 mM for the last 5 minutes. Cells were then separated on a discontinuous Percoll (Amersham Biosciences, Uppsala, Sweden) density gradient. The TEC-enriched fraction (the two upper phases) were harvested, washed and stained with G8.8 followed by phycoerythrine-conjugated rabbit anti-rat IgG Ab (Serotech). After 30 min at 4°C, cells were stained with anti-CD45-FITC (Pharmingen), to distinguish epithelial cells from G8.8-positive marrow-derived cells. Thymic epithelial cells were sorted according to the phenotype of CD45^-/lo and G8.8^+.

Blocking with the anti-FcR mAb 2.4G2 (Pharmingen) preceded the stainings for DC
while blocking with 1% horse serum was done for the TEC isolations. The average cell yield per thymus was 30,000 TEC.

**β-galactosidase activity assay/FACS:** Approximately $10^6$ cells from the Percoll gradient isolation (epithelial cells), or from the dendritic cells isolation (low density fraction) were labeled with G8.8 or CD11c Abs and resuspended in 1 ml RPMI (10% FCS) with 10µl of endogenous β-Gal-like activity inhibitor from the C12FDG kit (ImaGene Green Kit, Molecular Probes, Eugene, Oregon), for half hour at 37°C. Following this, 0.5 ml of pre-warmed C12FDG (15µM) substrate was added to the cells and incubated for 20 min at 4°C, room temperature or at 37°C. Results from the incubations at all three temperatures were very similar, but the background was significantly lower at 4°C and the results shown (Figure 6) were obtained at this temperature. This method was used to compare the percentage of cells positive for both G8.8 (or CD11c) and C12FDG between Ins2-KO mice (β-galactosidase expressed under the Ins2 promoter) and Ins1-KO mice (no β-gal transgene). The difference in double positive cells between the two genotypes was taken to represent cells truly expressing transgenic β-galactosidase. The number, as a percentage of all G8.8+ cells was much higher (10-30 fold higher) in Ins2-KO mice than in Ins1-KO in all four experiments whether the background, in the absence of C12FDG, was subtracted from each or not. In CD11c+ cells, no significant difference from background was seen.
3.4 Results

Proinsulin/β-Gal expression in the thymus and pancreas

In mice, two unlinked, nonallelic, \textit{Ins1} and \textit{Ins2} genes encode insulin. \textit{Ins1} is expressed at 2-3 times lower than \textit{Ins2} in the pancreas and thymus (8, 31, 32). Figure 1A, shows proinsulin expression in the thymus of 2-4 week old mice. Proinsulin is expressed two to three times less in the Ins2-KO mice when compared to the Ins1-KO mice confirming previous results (8). For both mice the concentration of proinsulin was normalized with cyclophilin (Figure 1B). As expected, β-Gal expression was detected in thymus and pancreas of Ins2-KO but not Ins1-KO mice (Figure 1C). These results show that proinsulin is abundantly expressed in the thymus of Ins1-KO mice and bioactive β-galactosidase in the thymus of Ins2-KO mice. These mice were utilized for the analyses of proinsulin/β-Gal expression in fractionated thymic stromal cells.

Proinsulin/β-Gal expression in thymic stromal cells

To examine proinsulin/β-Gal gene expression in thymic stromal cells, we enriched distinct cell populations to high purity by a combination of density fractionation and cell sorting. Dendritic cells were isolated from thymi of two to four week old mice using the standard protocol (29) that includes enrichment by adherence to culture dishes after overnight incubation prior to cell sorting with anti-CD11c Ab. Alternatively, to avoid possible phenotypic changes during the overnight incubation, possibly comparable to DC maturation \textit{in vivo} (33), the overnight incubation step was omitted, and the carefully removed low density cell fraction from the BSA gradient was sorted immediately using
Figure 1: Proinsulin/β-galactosidase expression in the thymus and pancreas. (A-B) Quantitative RT-PCR of proinsulin in the thymus of Ins1-KO and Ins2-KO mice normalised with cyclophilin. Lower band - competitor sequence, upper band - endogenous insulin, 1-8 are serial dilutions of competitor. (C) Semiquantitative RT-PCR of β-Gal in thymus and pancreas using serial dilutions of cDNA (1, 2, 3 = 1/1, 1/10, 1/100, normalised with cyclophilin).
an anti-CD11c or anti-DEC205 Ab (Figure 2A). Both techniques yielded CD11c positive cells within the expected range of 30,000 to 70,000 DC per mouse thymus (29). Without overnight incubation, sorting by CD11c and I-A\textsuperscript{b} (MHC II) Abs yielded similar results (Figure 2B). Virtually the entire population of CD11c\textsuperscript{+} cells were also positive for I-A\textsuperscript{b}.

Epithelial cells were isolated as described by Klein et al. (30) and enriched by cell sorting using G8.8 and CD45 Abs. The G8.8\textsuperscript{+} population consists of 40-60\% thymic epithelial cells, with the remainder of cells being mostly thymocytes (30, 34). Marrow derived cells were eliminated by selecting for G8.8\textsuperscript{+}/CD45\textsuperscript{−ho} providing exclusively medullary and cortical epithelial cells (19, 30) (Figure 2 C, D). Such isolations yielded very similar results to Derbinski et al. (19), as we consistently isolated approximately 30,000 epithelial cells/thymus. G8.8\textsuperscript{+}/CD45\textsuperscript{−} and G8.8\textsuperscript{+}/CD45\textsuperscript{hi} cell fractions were also isolated (Figure 2D).

By RT-PCR, strong proinsulin and β-Gal bands were detected in G8.8\textsuperscript{+}/CD45\textsuperscript{−ho} TECs from Insl1-KO and Insl2-KO mice, respectively, while only a very faint band was detected in the CD11c\textsuperscript{+}, DEC205\textsuperscript{+} or CD11c\textsuperscript{+}/I-A\textsuperscript{b+} DC from the same mice, isolated without overnight incubation (Figure 3B, lanes 1-5, Figure 3C 1-4). Even with a 2-3 fold greater number of DC, as compared to TECs, and with a substantially larger amount of RNA used for RT-PCR, controlled for by cyclophilin, only very weak proinsulin and β-Gal bands could be detected within the DC extracts (Figure 3B, 1,2 and 4, Figure 3C lanes 1 and 2). Reducing the cDNA quantity in the DC only reduced the intensity of the
Figure 2: FACS Sorting of Thymic Dendritic and Epithelial Cell Populations: Sort regions for DC without overnight incubation — with DEC205 (A), with CD11c and I-A^b (MHC II) (B). Epithelial sorts - G8.8^+ (C) (mainly TEC), and G8.8^+/CD45^- (D) (TEC). Dotted square - G8.8^+/CD45^- cells.
Figure 3: Proinsulin/β-galactosidase expression in thymic epithelial cells versus dendritic cells.

A) Proinsulin expression in Ins1-KO mice: 1. CD11c+ (DC) (O/N = with overnight incubation) (n = 4), 2. G8.8+ cells (mainly TEC) (n = 4) and 3. G8.8+/CD45lo cells (n = 8) (TEC). Cyclophilin - loading control. 4. Pancreatic proinsulin.
β-Gal expression in Ins2-KO mice: 5. CD11c+ (DC) (O/N) (n = 4) 6. G8.8+ cells (mainly TEC) (n = 4) and 7. G8.8+/CD45lo cells (n = 8) (TEC). 8. Pancreatic β-Galactosidase.

B) Proinsulin RT-PCR from FACS sorted 1. CD11c+ cells (DC, Ins1-KO mice), 2. half as much cDNA as 1 used for the PCR, 3. G8.8+/CD45lo cells. Corresponding cyclophilin levels for 2 and 3 are shown. β-Gal RT-PCR from Ins2-KO mice FACS sorted 4. CD11c+ (DC) and 5. G8.8+/CD45lo cells with the corresponding cyclophilin levels.

C) Proinsulin RT-PCR in CD11c+ and DEC205+ sorts with the corresponding cyclophilin levels (1 and 2). Proinsulin RT-PCR in G8.8+/CD45lo and G8.8+/CD45+ sorts (3 and 4, isolated as per Fig. 2D) along with the corresponding cyclophilin levels.

D) RT-PCR for proinsulin (1) and β-Gal (2) in G8.8+/CD45hi extractions along with the corresponding cyclophilin levels. (n = 2).
proinsulin and β-Gal bands ruling out saturation of cDNA as a possible inhibitor of the PCR (Figure 3B, lanes 1 and 2). Similarly, and as expected, the intensity of proinsulin band increased with epithelial enrichment of the G8.8+ fraction by double sorting (G8.8+/CD45lo, Figure 3A, lanes 2 and 3). β-Gal was also detectable in G8.8+ cells and G8.8+/CD45lo cells from Ins2-KO-mice (Figure 3, lanes 6 and 7, note that loading was not normalized for this result). No proinsulin/β-Gal expression was detected in G8.8+/CD45hi cells (Figure 3D) and from DC with overnight incubation (Figure 3A lanes 1 and 5). Proinsulin was also strongly detected in G8.8+/CD45+ epithelial cells even when poor RNA yields did not allow for the detection of cyclophilin (Figure 3C, lane 4). All these data demonstrate that in the thymus proinsulin mRNA transcripts are predominantly detected in thymic epithelial cells.

Proinsulin/β-Gal is produced by rare thymic epithelial cells of the Hassall’s Corpuscles

To be more specific about which population of cells are responsible for insulin gene transcription in the thymus, we localized β-Gal activity in histological sections. Figure 4 shows histochemistry results of the β-Gal activity on pancreatic and thymic frozen sections of Ins1-KO and Ins2-KO mice. These results show that β-Gal activity is specifically detected in pancreatic islets of Ins2-KO mice but not in Ins1-KO mice (Figure 4 A, B). In the thymus, β-Gal activity is detected in Ins2-KO mice, but not in Ins1-KO, and was localized in the medullary epithelial cells forming the concentric structure that resembles Hassall’s Corpuscles. Approximately 1 in every 10 HC were positive for β-Gal activity, some of which contained as many as 4 X-gal stained cells.
Figure 4: Histochemistry of β-Gal activity on pancreatic and thymic frozen sections. (A, B) Pancreatic β-gal activity detected in β-cells of Ins2-KO mice but not in Ins1-KO (40x). (C, D) Thymic β-Gal activity detected in medullary epithelial cells (arrows) of HC of Ins2-KO mice but not in Ins-1KO mice (40x). (E,F) Zoom of HC cells positive for β-Gal activity.
(Figure 4C-F). Because our sections represent only a slice of a three-dimensional HC, the true proportion must be higher and it is, indeed, possible that every HC contains these cells. Immunohistochemistry colocalized all of the β-Gal positive cells with G8.8 (Figure 5B-E) but never with CD11c (DC) (Figure 5A). These data demonstrate that proinsulin-producing cells in the mouse thymus are medullary epithelial cells of the Hassall's Corpuscles.

**Flow Cytometry Analysis of β-Gal/Proinsulin-Producing Rare Thymic Epithelial Cells**

We further localized β-Gal activity in purified populations of thymic stromal cells by flow cytometry using fluorescent β-Gal substrate (C12FDG) and cell-type specific antibodies. The results consistently revealed a population of between 1-3% G8.8+/β-Gal positive cells within the purified G8.8+ thymic epithelial cell extract (TEC) from Ins2-KO mice (Figure 6B). We could not detect a similar β-Gal positive population in G8.8+ cells of control Ins1-KO mice (Figure 6A). In a separate experiment we incubated sorted G8.8+ cells with C12FDG, under similar conditions as above, and observed them by fluorescence microscopy. The results showed a similar proportion of β-Gal positive cells in the Ins2-KO preparation, consistent with the FACS data (Figure 6C). The same experiments were done for sorted CD11c+ DC and no population of β-gal positive DC was observed in INS2KO mice that was above the background levels obtained with DC from Ins1-KO mice (data not shown).
Figure 5: Immunohistochemistry of β-Gal activity and thymic stromal cells. (A) Colocalization of β-Gal activity and CD11c (DC) (40x). (B-E) Colocalization of β-Gal activity and G8.8 (TEC) (40x). No staining was observed in Ins-1KO sections (data not shown).
Figure 6: β-galactosidase activity in sorted thymic G8.8+ cells.

(A,B) Percoll purified TECs were incubated with C12FDG substrate (see methods). A percentage of 1-3% more G8.8+/β-Gal positive epithelial cells was consistently observed in the Ins2-KO sorts when compared to the Ins1-KO sorts. Percentages shown are normalized for the percentage of G8.8+ cells in each extraction. No similar increase over the Ins1-KO control was seen in BSA purified CD11c+ cells (DC) C) G8.8+ TEC cells were incubated with C12FDG substrate and viewed by fluorescence microscopy (10x). An increased number of β-Gal positive cells in the Ins-2KO sample as compared to the Ins1-KO sample was observed (data not shown) confirming the FACS data.
3.5 Discussion

Our results show that in mouse thymus, proinsulin expression is limited to a very rare fraction of medullary epithelial cells forming the Hassall’s Corpuscles. Previous studies attempting to identify which cells are responsible for this proinsulin expression, by RT-PCR, gave contradictory results with expression being detected in DC (26) or TECs (19) but not both. In order to resolve this discrepancy and to best characterize the identity of these cells, we used β-gal as a marker of transcriptional activity of the endogenous Ins2 promoter. Semiquantitative RT-PCR, shows that proinsulin/β-Gal is very strongly and easily detected in TECs but only very faintly detected in DC that are sorted without overnight incubation. Throsby et al. (26) also detected a very faint band in DC extracts and they required using a radiolabelled probe to see it adequately. This band may be the result of some epithelial cell contamination in the DC extracts, or it might represent expression at much lower levels or much fewer cells among the DC.

Immunohistolocalization experiments revealed that these thymic cells were confined to the thymic medulla and were part of the Hassall’s Corpuscles. The expression of insulin and other ectopic or promiscous tissue-specific self-antigens by medullary epithelial cells has been extensively documented by Derbinski et al. (19). The authors did not, however, identify the exact location of the medullary epithelial cells responsible for the production of insulin or other antigens and their spatial relationship to the rest of the stroma. Our immunohistochemistry of β-Gal activity showed that proinsulin-production is a unique property of rare medullary thymic epithelial cells of the Hassall’s Corpuscles.
It has been reported that self-antigen expressing cells in the thymus, including proinsulin producing cells, were frequently found in a cluster of two to four positive cells, when analyzed ex vivo (35). This may reflect in vivo clustering, as suggested by our observation of clusters of more than one β-gal positive cells. Rodewald et al. (36) have demonstrated that the thymus medulla consists of epithelial cell islets, each derived from a single progenitor. It is not known how this clonality relates to the epigenetic differences that distinguish the rare cells involved in the thymic expression of tissue-restricted antigens from the rest of the TECs. It is also not known whether such expression is the property of a large number of cells, each specializing in a particular antigen, or of a small number of cells that have developed the promiscuous transcriptional machinery to express many diversely regulated antigens. Our ability to isolate the insulin-producing cells by scaling up the flow cytometry depicted in Fig. 6B will enable us, in studies currently under way, to provide an answer to these questions.

Thymocyte negative selection occurs mainly in the medulla (37, 38) and the HC cells have been associated with apoptotic thymocytes (22). DCs and macrophages have been frequently observed around HC (23, 38) and the DC were shown to be responsible for phagocytosis of apoptotic thymocytes (23). In human thymus, Pugliese et al. found that insulin can be detected in a number of thymic and secondary lymphoid DC by double immunostaining (39). Our findings are compatible with the presence of a small number of proinsulin-expressing DC but it is also quite conceivable that these DC do not express insulin at the transcriptional level but merely take it up for presentation. In order to tolerate both CD4+ and CD8+ T cells, presentation of the self-antigen by MHC class I
and II molecules is necessary. It is known that mTECs can present endogenously produced self-antigens by both classes while DCs do not seem to present endogenously produced antigen on class II (40).

If proinsulin expression in thymus medulla is important in negative selection, how can contact of all or most thymocytes with the few, scattered insulin-producing cells be assured? This may be the result of chemoattractant properties (23) that drive each thymocyte to scan many HC. Alternatively, or in parallel, other APC, particularly dendritic cells, may also contribute by taking up secreted antigen for presentation.

We have previously shown that thymic insulin expression levels modulate insulin-specific autoreactive T cell tolerance even in non-diabetes prone mice (8). Against the NOD background, the Ins2 knockout mice (with extremely low insulin expression in the thymus but normal levels in the pancreas) develop diabetes with higher incidence and earlier than the wild type (20, confirmed by our own preliminary observations and those of the group of Dr. G. Eisenbarth, Moriyama et al, personal communication-submitted).

We believe that the isolation of proinsulin-producing cells, using the techniques we have developed here, will allow detailed study of the cellular and molecular biology of proinsulin production, central tolerance mechanisms and the possibility of designing novel forms of gene or cellular therapy.
3.6 Acknowledgment

This work was supported by a grant from the Canadian Diabetes Association in honor of the late Emily C. Haynes. The authors thank Dr. Jacques Jami for his generous gift of the Knockout mice and Rosemarie Grabs for all her technical assistance.
3.7 References


4.0 CHAPTER FOUR: Isolation and Characterization of Proinsulin Producing Medullary Thymic Epithelial Cell Clones
Michael Palumbo, Dina Levi, Aziz Alami Chentoufi and Constantin Polychronakos

Diabetes, Vol. 55, Sept. 2006

4.01 Contribution of Authors

Michael Palumbo: Experimental design and strategy for proinsulin producing medullary thymic epithelial cell(s) isolation (including animal work), fluorescence assisted sorting and analysis (FACS), microscopy and RT-PCR for all transcripts except CD80.

Dina Levi: Real-Time PCRs for AIRE and Insulin, RT-PCR for CD80.

Aziz Alami Chentoufi: Technical advice.

Dr. Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

4.02 Objective

Following our studies (see Chapter 3) confirming that the proinsulin producing cells of the thymus are medullary epithelial cells, our next objective was to isolate and culture the proinsulin producing cells using a cross of Immortomice (which contain an SV40 large T antigen under the major histocompatibility class I promoter) and Ins2KO mice (which contain a functional β-Galactosidase gene under the regulation of the Ins2 promoter). Immortomice have been previously used to culture primary epithelial cells [1-3] and enzymatic activity assays will allow for the isolation of the β-Galactosidase/proinsulin producing cells.
4.1 Abstract

Proinsulin, like many tissue-specific antigens, is expressed by rare (1-3%) cells of the thymus medullary stroma, presumably for the purpose of self-tolerance; levels of this expression are associated with Type 1 diabetes susceptibility in humans and in the NOD mouse. To further understand the mechanism of central tolerance induction by these rare cells we have isolated and cultured two proinsulin producing epithelial cell clones from murine thymus. These cells have a typical epithelial morphology and, by flow cytometry, a surface phenotype representative of mature thymic medullary epithelial cells (G8.8+/UEA-1+/DEC205+/CD45+/MHC II+). By RT-PCR they express predominantly \textit{Ins2} as opposed to \textit{Ins1}, as does whole thymus. Expression of the transcription factor \textit{Aire}, implicated in enhancing promiscuous thymic expression of tissue-specific antigens, fell to very low levels after a few passages but increased twenty-fold upon exposure to an agonistic anti-lymphotoxin B antibody, concurrent with 2.5-fold enhanced insulin expression. RNA of \textit{Pdx-1}, \textit{Glut-2} and \textit{Gck}, was detectable by RT-PCR in whole thymus but not in the clones, suggesting thymic proinsulin expression is \textit{Pdx-1} independent and that \textit{Pdx-1}, \textit{Glut-2} and \textit{Gck} are likely expressed in the thymus as antigens, not as regulatory molecules.
4.2 Introduction

The contribution of central tolerance to the prevention of autoimmune disease has acquired increased importance with the finding that many tissue specific self-antigens are expressed in the thymic medulla [4, 5]. Proinsulin is one of these self-antigens, and it is also one of the major self-antigens involved in the T-cell mediated destruction of the pancreatic \( \beta \)-cells that causes type 1 diabetes (T1D) [6, 7]. Other antigens include Glutamic Acid Decarboxylase (GAD65 and 67) and the tyrosine phosphatase I-A/2. However proinsulin is the only T1D autoantigen that is exclusively expressed by the \( \beta \)-cell, and the only one that maps to a confirmed genetic susceptibility locus ([7] and references therein).

The importance of thymic proinsulin expression first became obvious from observations of tolerance induction to \( K^b \) and the SV40 large T-antigen when expressed in the thymus under the insulin 2 promoter, tolerance that was transferable by thymus transplantation [8-10]. In addition, reducing proinsulin levels in thymus, while maintaining it normal in pancreas, increased the number of proinsulin specific auto-reactive T-cells in mice [11, 12] an effect also transferable by thymic transplantation [13]. Against the NOD background, this genetic manipulation results in a significant acceleration of autoimmune diabetes in females and drastically increased frequency in males [14]. These findings corroborate genetic evidence in humans, where a Variable Number of Tandem Repeat (VNTR) polymorphism just 5' to the human proinsulin gene specifically modulates thymic proinsulin levels, and alleles resulting in reduced thymic expression predispose to Type I Diabetes [7, 15].
The majority of thymic self-antigens are expressed by medullary thymic epithelial cells (mTECs) in both mouse and humans [4, 5]. We and others have shown that proinsulin RNA can be detected very strongly in mTECs as opposed to other medullary thymic cell populations and also that this expression is limited to 1-3% of all thymic mTECs [4, 16]. In addition, we localized most of this proinsulin expression to what are believed to be the most mature of all thymic mTECs – those that make up the Hassall Corpuscles-like structures in mice [16]. Although this has yet to be confirmed using another experimental strategy, there is work demonstrating that self-antigens are produced in the Hassall’s Corpuscles of human thymus [17, 18] which was also recently shown to be a site of production of CD4+/CD25+ regulatory T-cells [17-19]. Thymic dendritic cells (DC), as well as peripheral DCs have also been reported to be positive for proinsulin/insulin peptides [20], a finding that has not been reproduced in attempts by others [4, 5, 16, 20]. Interestingly, recent work demonstrated that although mTECs can produce self-antigens, their processing and presentation of the antigen is limited to CD8+ T-cells and that efficient presentation to CD4+ T-cells requires self-antigen uptake and presentation by DCs [21] which, in the Hassall’s corpuscles, appears to be involved in the generation of T-regulatory cells [19, 21].

The rarity of proinsulin producing mTECs has precluded detailed study aimed at answering important questions. Do such cells produce all the thymic self-antigens, only proinsulin, or a defined subset of self-antigens in addition to proinsulin? How are proinsulin and the other expressed antigens processed and what is the role of this
processing? Indeed, if a subset of other self-antigens is expressed by these cells, what is the transcriptional mechanism involved and is it related to the reported chromosomal clustering of the genes encoding those antigens? [5, 22] These findings have prompted the proposal that more than one lineage of mTECs exist, each of which produces a subset of all self-antigens localized in close proximity along the same chromosome. Additional questions include; what is the transcriptional and translational regulation of proinsulin (and the other self-antigens produced by these cells) and how do these cells develop the capacity to express a single, many or all of the tissue specific antigens? Recently, the Autoimmune Regulator Element (Aire) has been shown to play a very important role in the expression of proinsulin along with many other but not all thymic expressed tissue specific antigens [22, 23]. Mutations in mouse Aire abolish or drastically reduce ectopic thymic expression of tissue-specific antigens and give a phenotype similar to mutations in the human orthologue which results in autoimmune polyendocrinopathy type 1 (APECED syndrome), in which diabetes is a common feature [23].

Towards answering these questions, we developed a strategy to isolate and culture the proinsulin producing cells from mouse thymus. Here we report the generation and basic characterization of two insulin-positive clonal mTEC lines.
4.3 Materials and Methods

Animals: INS2KO (C57Bl6/129) mice were a gift from Dr. J. Jami (Institut Cochin, Paris, France) and were generated as described by Duvillié et al. [24], by replacing the 
Ins2 gene with a functional LacZ copy, regulated by the endogenous Ins2 promoter. The 
expression profile of both β-galactosidase and proinsulin (both Ins1 and Ins2) in the 
thymus and pancreas of these mice has been previously described [11, 16]. These were 
crossed with mice homozygous for the Immortomouse transgene, a temperature sensitive 
SV40 large T-antigen, which induces proliferation of thymic epithelial cells at 33°C [1-3]. At 37°C, these cells revert to a more differentiated phenotype [1-3]. Mice carrying 
this transgene against the C57Bl6 background were purchased from Charles Rivers (Saint 
Constant, Quebec, Canada) and have been previously described by Jat et al. [1]. All mice 
were bred at our animal facility under conditions specified by the Canadian Council of 
Animal Care. Homozygous Ins2KO and Immortomouse animals were crossed to obtain 
F1 heterozygous Ins2KO/Immortomouse (INS2KO/Imm).

Isolation of Thymic Epithelial Cells: We modified the previously described protocol 
[4, 16] as follows: thymi from three 10 week old INS2KO/Imm were individually 
extracted, finely minced, resuspended in RPMI (no phenol red, Cat# 11835-030, Gibco, 
Rockville, MD, USA) and stirred at room temperature for 10 minutes to release 
thymocytes. Following this, individual thymi were digested with Collagenase/Dispase 
(Sigma, Oakville, Ontario, Canada, 125μg/ml and DNase I (Roche, Mannheim, Germany, 
125μg/ml) for two rounds of approximately 45 minutes each at 37°C. Cells were then 
washed with PBS and resuspended in DMEM (US Biological, Swampscott, MA, USA,
Cat-M3861-20, with D-Valine - to inhibit fibroblast growth, 100μg/ml IFN-γ (Roche), 10% FBS, 1x Sodium Pyruvate (Gibco), 1x Non-essential amino acids (Gibco), 50μM β-mercaptoethanol and 50μg/ml Gentamicin antibiotic (Gibco) and placed at 33°C. Media were changed every 3-4 days the first 1-2 weeks followed by passaging every 3-4 days at 3-4 weeks.

**Enrichment by β-galactosidase Activity Assay (C12FDG):** After approximately 4 weeks in culture, two of the three individual lines of extracted cells were treated with the fluorescent β-galactosidase substrate C12FDG (ImaGene Green Kit, Molecular Probes, Eugene, Oregon, USA) and green fluorescent cells were FACS sorted (Vantage FACS, Becton Dickinson, Franklin Lakes, New Jersey, USA) and re-cultured. Briefly, cells were immersed, while still in culture adhered to 10cm plates, in RPMI media (standard RPMI with phenol red) with 10% FBS and 100μl of endogenous β-gal like activity inhibitor from the C12FDG kit was added for 30 min at 37°C. Following this, pre-warmed C12FDG substrate in RPMI media without FBS, was added to a final concentration of 15μmol/l and the cells were incubated at 37°C until green fluorescent cells could be observed by microscopy using the Enhanced Green Fluorescent Protein filter (Leica FZFLIII, AG, Wetzlar, Germany). Once green fluorescent cells were observed (~1 hour), the reaction media was removed, cells were trypsinized washed, and incubated with PBS (10mM EDTA, 1% BSA, fraction V, Sigma) that included a stop reagent provided with the C12FDG kit. Green fluorescent cells (approximately 20-40% of the cell mix) were then FACS-sorted and placed in culture at 33°C with 100μg/ml IFN-γ (required to maximize expression of the SV40-Tag transgene).
Limiting Dilutions and β-Galactosidase Luminescence Assays: After sorted cells were cultured for 1-2 weeks sorted and non-sorted cells were cultured in flat-bottom 96 well plates at 33°C with IFN-γ, diluted to 1 cell / well (0.1ml). Colonies with clearly epithelial morphology were identified, trypsinized and grown in 24 well plates. It should be noted that although we continuously cultured our cells in D-valine containing media, a number of cells with fibroblast morphology were observed in those limiting dilutions. The selected colonies were than analyzed for β-galactosidase luminescence using the Galacto-Star Kit (Tropix Inc, PE Biosystems, Bedford, Massachusetts, USA). Cells from the 24 well plate were trypsinized and an equivalent number of cells (approximately 2 x 10^5) from each plate were lysed in 200μl freshly made lysis buffer (0.5% NP40, 0.1 M Tris HCl pH=7.9, 0.01M DTT). Galacton-Star substrate was added (1/50) to 100μl of reaction buffer (100μM Sodium Phosphate, NaH2PO4, 1μM MgCl2, and 5% Sapphire Enhancer, Tropix Inc., PE Biosystems), in a non-translucent 96 well assay plate (Corning, NY, USA) followed by the addition of 20μl-30μl of cell extract to the wells. All reactions were done in triplicate. Luminescence readings were read at 30 and 60 minutes in an EG&G Berthold Microplate Luminometer, LB96V (Bad Wildbad, Germany). Four colonies with reproducibly higher than average luminescence, were selected from one of the enriched cultures and six from the other, for a total of ten. As negative controls, we selected 5 colonies with background luminescence values, taken from the non-enriched plate. The difference in luminescence values between the positive colonies and the average colonies was 2-4 fold.
RNA Preparation, cDNA Synthesis and RT-PCR: RNA was extracted using the RNeasy Mini Kit (Qiagen, Mississauga, Ontario, Canada) and treated with Ambion DNase I (Austin, Texas, USA). Approximately 2μg of RNA was reverse transcribed using Random Primers (Gibco) and Superscript II Reverse Transcriptase (Gibco) and parallel samples in which Reverse Transcriptase was omitted were always included to confirm the absence of DNA contamination. Primer sequences, conditions, reagents and amount of DNase I treated RNA used for each of the PCRs are shown in Table 1. RT-PCR for insulin and cyclophilin was performed on RNA extracted from cells cultured at either 33°C or 37°C and with or without IFN-γ.

FACS Analysis: Antibodies used to characterize the surface-marker phenotype of our cells were: Rat IgG2a G8.8 (Anti-gp40, mouse homologue of human EPCAM, DSHB, University of Iowa, Iowa City, IA, USA) with Goat Anti-Rat IgG RPE (Serotech, Raleigh, NC, USA), Rat Anti-Mouse DEC205 FITC (Sero tec), Mouse Anti-Mouse I-A<sup>b</sup> FITC or RPE (Phar mingen, Mississauga, Ontario, Canada), Mouse Anti-Mouse CD45.2 FITC (Pharmingen). Cells were washed with PBS and then incubated with PBS, 10mM EDTA for approximately 3 minutes. Cells were then resuspended (10<sup>6</sup> cells/ml) in PBS, 5mM EDTA, 1% FCS (fraction V) and approximately 10<sup>6</sup> cells were incubated with the appropriate antibody. Generally, a 1/200 dilution was used for G8.8, whereas 1/10 to 1/20 dilutions were used for all other antibodies. Isotype controls were used at the same concentrations and included RPE conjugated Rat IgG2a (Pharmingen) and FITC conjugated Mouse IgG2a (Pharmingen). Fc blocking was done using Rat Anti-Mouse CD16/CD32 (Pharmingen) for all reactions except when Goat Anti-Rat IgG was used as
<table>
<thead>
<tr>
<th>Gene</th>
<th>Sense Primer</th>
<th>Antisense Primer</th>
<th>Conditions</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ins1/2</td>
<td>GGCTTCTTCTACACCCCA</td>
<td>CAGTAGGYTCTCCACGCTGGTA</td>
<td>1**</td>
<td>181bp</td>
</tr>
<tr>
<td>Ins2</td>
<td>CCAAGCTATAATCCAGAGCCA</td>
<td>GCTGAAAAAAAAGCCACGCT</td>
<td>2*</td>
<td>197bp</td>
</tr>
<tr>
<td>Pde-1</td>
<td>TGGCTGGGATCACCTGGAGCA</td>
<td>GGTTCGCCTGTGTAAAGACACC</td>
<td>2</td>
<td>275bp</td>
</tr>
<tr>
<td>Gsat-1</td>
<td>GACCCAGGAGCCCGTTCTCA</td>
<td>GTCGAAGACCCAGACACCA</td>
<td>2</td>
<td>150bp</td>
</tr>
<tr>
<td>Gck</td>
<td>TGGATGACAGACAGCAGAGG</td>
<td>ACTCTGTAGCCTTCTGGGAGT</td>
<td>3*</td>
<td>208bp</td>
</tr>
<tr>
<td>Aire</td>
<td>ACACGTGGGACCCACTTTCTG</td>
<td>AGGGATTCAGACCCATGGCGCAG</td>
<td>2</td>
<td>295bp</td>
</tr>
<tr>
<td>Cd80</td>
<td>GAAACCCATCTGAGACACTA</td>
<td>GACAACGATGACGACGACTG</td>
<td>4**</td>
<td>338bp</td>
</tr>
<tr>
<td>Cyclophilin</td>
<td>see reference 10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conditions:**

1 94°C 2min30, 94°C 1min, 60°C 1min, 72°C 1min50 - 32x
2 95°C 10min - 1x, 97°C 30sec, 64°C 30sec, 72°C 30sec - 5x, 95°C 30sec, 64°C 30sec, 72°C 30sec - 45x, 72°C 6min - 1x
3 95°C 10min - 1x, 97°C 30sec, 62°C 30sec, 72°C 30sec - 5x, 95°C 30sec, 62°C 30sec, 72°C 30sec - 30x, 72°C 6min - 1x
4 95°C 10min - 1x, 97°C 30sec, 61.5°C 30sec, 72°C 30sec - 5x, 95°C 30sec, 61.5°C 30sec, 72°C 30sec - 30x, 72°C 6min

*60ng of DNASE1 treated RNA used for PCR as opposed to 200ng of treated RNA for all other PCRs.
*Also indicates 15µl as opposed to 25µl final reaction volume used.
Reagent Concentrations in final reaction mix: 2mM MgCl2, 1X PCR Buffer, 0.2mM dNTPs, 0.1U AmpliTaq Gold, 1µ of each primer.
MgCl2, 10x PCR Buffer and AmpliTaq Gold are Applied Biosystems (Foster City, CA, USA), Products Cat# N808-0249
**0.1U RedTaq (Sigma) was used for Ins1/2, regular AmpliTaq was used for CD80.
All primers obtained from Alpha DNA (Montreal, Quebec, Canada).

Table 1: Primers and PCR conditions
a secondary antibody – horse serum was added to these samples instead. Similarly, 10^6 cells were incubated with a 1/10 dilution of Fluorescein Ulex Europaeus Agglutinin I (UEA-1, an mTEC marker, Vector Laboratories, Burlingame, CA, USA). In some instances, cells were fixed with 1% formalin in PBS, 5mM EDTA, 1%FCS (fraction V) if FACS analysis could not be done immediately.

**Treatment to reactivate Aire expression:** The mTEC clones were incubated with 100nM of Trichostatin A (A.G. Scientific) for 24 hours. After the incubation rat anti-mouse lymphotoxin beta receptor antibody (Serotec – IgG2a azide-free antibody) was added to the cells at a concentration of 1μg/ml for 6 hours.

Real-time PCR was performed on 200 ng of cDNA, generated using random primers, both for *Aire* and insulin gene expression and normalized to 18s rRNA gene expression, using TaqMan Gene Expression Assay Kits for all genes (Applied Biosystems).

### 4.4 Results

**Isolation and Culturing of Thymic Proinsulin Producing Epithelial Cells:** As described in the methods, two of the three plates of thymic cells cultured from INS2KO/Imm mice were subjected to an enrichment step, using FACS with C12FDG. In our previous paper [16], using a shorter incubation time, we were able to identify and isolate to quite high purity the β-Gal/proinsulin producing cells (1-3% of the entire epithelial cell population). To maximize yield, at the expense of specificity, we extended the incubation time to 1 hour, at which point 20-40% of the cells in culture were
fluorescent green. This represents a 2-4 fold enrichment of proinsulin producing cells, in preparation for limiting dilution.

Over 1000 limiting dilutions were prepared from both enriched and non-enriched mTEC cultures. Once growing in 96 well plates, ~25 epithelial colonies were chosen from each of the enriched cultures and ~10 were chosen from the non-enriched cultures. A total of 10 colonies from the enriched cells had β-gal luminescence values above the average signal for all colonies combined (2-4 fold above the average). All 10 of these colonies along with 5 colonies with average luminescence values from the non-enriched plate were analyzed for proinsulin by RT-PCR. Two were found to be positive for proinsulin (PCR primers designed to detect both Ins1 and Ins2) – 1 positive colony was obtained from each of two enriched cultures indicating that they were separate colonies and not clonal (Figure 1). Figure 2 shows RT-PCR proinsulin results of two negative colonies and one of the positive colonies, after 45 passages, in triplicate with cyclophilin as a loading control. Similar results were obtained whether the cells were cultured in the presence of IFN-γ at 37°C for 3 days prior to RNA isolation, or in the complete absence of IFN-γ at 37°C. Assuming a 2 fold enrichment in proinsulin producing cells by C12FDG-FACS, the percentage of proinsulin producing cells in our cultures was 2/50 ~ 4% and thus corresponds reasonably well with what was determined to be the actual percentage in vivo (~2%) [4, 16]. Of note is that the levels of proinsulin detected by RT-PCR in the cultured cells seems to be lower than what can be detected in the whole thymus when equal RNA amounts are used (data not shown). This is related to Aire
Figure 1: RT-PCR for proinsulin (detecting both 1 and 2) in cell colonies grown at 33°C with IFN-γ. All colonies shown here had above average β-galactosidase luminescence values (2-4 fold above background). N designates negative for proinsulin, P designates positive for proinsulin.
Figure 2: RT-PCR for proinsulin (detecting both 1 and 2) in cell colonies grown at 33°C without IFN-γ and placed for 3 days at 37°C without IFN-γ before RNA was extracted. N designates negative for proinsulin, P designates positive for proinsulin. All reactions done in triplicate for both negatives and one of the positives - the other positive was also tested and gave the same results (data not shown).
expression (see below) and is probably attributable to loss of some endogenous stimulator. Identification of such factor(s) will be the subject of future research.

**FACS Analysis:** Typical mTEC surface phenotype by FACS can be defined as G8.8* [25, 26], CD45~hi [4], DEC205~ [27, 28], UEA-1~ [29, 30] and MHC II~ [3, 30], with the UEA-1 positive phenotype believed to be restricted to more mature mTECs [31]. No changes in surface phenotype were observed with changes in temperature (33°C vs. 37°C) but I-A^b (MHC II) levels were markedly upregulated in the presence of IFN-γ as has been previously shown [3] (Figure 3). The phenotype of two non-proinsulin producing colonies and both proinsulin producing colonies were found to conform to all these criteria (Figure 3). Of note is that all colonies initially analyzed were G8.8^+ but at approximately 30 or so passages, the G8.8 positive phenotype was lost – this was assumed to be the result of cell culturing and the no longer requiring of adhesion molecules. Also to note is that previous studies have shown that MHC II expression is undetectable in the absence of IFN-γ [2, 3] but FACS analysis with our colonies cultured without IFN-γ for over 5 weeks time still demonstrated a population of MHC II positive cells (Figure 3). Figure 4 is a phase contrast image of the proinsulin positive and negative colonies demonstrating a cuboidal, cobblestone epithelial morphology.
Figure 3: FACS analysis results for one of the proinsulin positive colonies. Identical results were obtained for the other proinsulin positive colony and the two proinsulin negative colonies. Top panel for each of the analysis is the negative or isotype control, bottom panel is with antibody or lectin (for UEA-1). UEA-1 specificity was assayed by incubating $10^6$ Jurkat cells with the same concentration of UEA-1 lectin used for the epithelial cells. No positive cells were observed (data not shown). No differences in signal were observed with cells incubated at 33°C or 37°C with or without IFN-γ except for MHC II which was up-regulated significantly in the presence of IFN-γ as shown.
Figure 4: Phase contrast microscopy of the proinsulin negative and positive mTECs demonstrating a cuboidal and "cobblestone" appearance characteristic of epithelial cells.
RT-PCR analysis for expression of *Ins1*, *Ins2* and Promiscuous Expression Markers:

Unlike the murine pancreas, the murine thymus expresses very low levels of *Ins1* compared to *Ins2* [11, 32]. To determine whether this was also the case in our proinsulin positive mTECs we assayed, by RT-PCR, for each individually. As shown in Figure 5B and 5C, *Ins2* is easily detectable in pancreas, thymus and our proinsulin positive mTECs. *Ins1* in contrast, is easily detectable in pancreas but only gives a very faint band in thymus or the proinsulin positive mTECs.

In addition to being positive for UEA-1 [31], more mature mTECs, are also characterized by the expression of CD80 and *Aire* both of which are expressed only by a subset of the UEA-1⁺ mTECs [23, 31, 33]. CD80 was detectable by RT-PCR in both the positive and negative colonies (Figure 5A) but *Aire* was not. *Aire* was detected by RT-PCR when the cells were first placed in culture but expression was lost quickly with passaging, a phenomenon that has been previously observed [34]. *Aire* and CD80 have been associated with expression of tissue restricted antigens in the murine thymus, including insulin [23, 35]. However, although *Aire* knockout reduces proinsulin expression significantly, some proinsulin is still detectable in *Aire* deficient CD80⁺ mTECs [35]. This explains how we could detect insulin in the absence of detectable *Aire* expression.
Figure 5: A) RT-PCR for CD80 in INS2KO/Immortomouse pancreas, proinsulin negative mTEC colony, proinsulin positive mTEC colony and INS2KO/Immortomouse thymus. B) RT-PCR for Ins2, Ins1, Pdx-1, Glut-2 and Gck in INS2KO/Immortomouse pancreas and thymus. C) RT-PCR for Ins2, Ins1, Pdx-1, Glut-2 and Gck in the proinsulin negative and positive mTEC colonies. Pancreas and thymus RNA was obtained from a fourth 10 week old INS2KO/Immortomouse mice. Cyclophilin was used to control for loading (data not shown).
Induction of Aire Expression and Ins2 Enhancement with anti-Lymphotoxin B Antibody.

It has been previously reported that treatment with an agonistic antibody against the lymphotoxin B receptor (LTBR-Ab) in the presence of trichostatin A (TSA) restores Aire expression by mTECs in culture [34].

Indeed, after LTBR-Ab+TSA treatment, as described in Methods, Aire mRNA, measured by real-time RT-PCR, increased by twenty-fold in the Ins2-expressing mTECs. This was accompanied by a 2.5-fold increase in Ins2 mRNA (Figure 6, average of four separate experiments, p=0.01).

Other β-cell Specific Genes

In the search of other important proinsulin regulatory molecules, we also examined regulators known to be important in pancreatic β-cells. These were Pdx-1, a β-cell specific transcription factor crucial for pancreas development and insulin expression, Glut-2 (glucose transporter) and Glucokinase (Gck) - 2 proteins essential in the sensing of systemic glucose levels and the corresponding β-cell insulin response [36-38].

Interestingly, Pdx-1, Glut-2 and Gck transcripts were all detectable in whole thymus but were not detectable in either the proinsulin or non-proinsulin producing mTECs (Figure 5B and C). This suggests that these molecules are expressed in the thymus as antigens rather than active insulin regulators.
Figure 6: The effect of preincubation with anti-lymphotoxin B antibody and trichostatin A (LTBR-Ab + TSA) on Aire (left panel) and insulin (right panel) by medullary thymic epithelial cell clones expressing or not expressing insulin, respectively Ins(+) mTEC and Ins(-) mTEC. The means and standard errors of seven (Aire) or four (insulin) separate preincubations are shown. Open bars represent untreated, shaded bars treated cells.

**p≤0.01, otherwise NS.
4.5 Discussion

To our knowledge, this is the first time proinsulin producing mTEC cell lines have been established. Several important pieces of evidence suggest that our proinsulin producing mTEC cells are representative of their in vivo counterparts in the murine thymus. In addition to having an appearance characteristic of epithelial cells, they also have a surface phenotype characteristic of mTECs with their UEA-1 positive phenotype suggesting that they are mature mTECs. This mature mTEC phenotype is associated with the expression of Aire and CD80, two proteins that characterize cells that express a number of tissue-restricted antigens, including insulin [35]. mTECs from Aire deficient mice show a greatly reduced level of proinsulin expression, consistent with the lower Ins2 expression in our mTECs compared to fresh whole thymus (which also explains the low β-gal expression we encountered during the cell cloning). As this reduction of Ins2 expression was partly reversed by inducing Aire expression, we can conclude that loss of Aire in culture is, at least partly, to blame. The reason for such loss is not clear but absence of stimulation by soluble factors and/or cell-to-cell interactions in the thymus environment is a very plausible explanation. Lymphotoxin, the endogenous LTBR ligand, is a prime candidate in this respect.

As a histone deacetylase inhibitor, TSA might turn genes on in a non-specific fashion. We do not believe this to be an explanation for the effect we saw, for a number of reasons. First, in the absence of LTBR-Ab, no effect was seen. Second, the specificity of our Ins2-positive vs. Ins2-negative mTECs was maintained after LTBR-Ab+TSA treatment, which caused a negligible (in absolute terms) increase in insulin mRNA in the
negative clone (Figure 6). Third, the values shown in Fig. 6 have been normalized for total RNA, using 18S RNA determined by real-time RT-PCR.

The fact that our proinsulin positive cells express Ins2 at a much higher level than Ins1 is also very encouraging evidence that we indeed isolated the true proinsulin positive mTECs. Ins1 is expressed at very low levels; indeed it is sometimes undetectable in thymus, whereas, in pancreas, its expression level is of the same order of magnitude as that of Ins2. These findings suggest that the mechanism of proinsulin expression in these cell lines has remained the same as what takes place in thymic mTECs.

Of much interest are the results with regards to Pdx-1, Glut-2 and Gck expression. The expression of these three proteins in the β-cell is crucial for β-cell formation and proinsulin production (Pdx-1), and for proinsulin level modulation with regards to peripheral glucose levels (Glut-2 and Gck). Thus, we confirm a previously reported finding that thymic proinsulin expression is Pdx-1 independent [32]. However, our finding of Pdx-1, Glut-2 and Gck in whole thymus but not in our proinsulin producing mTECs also suggests that proinsulin expression in the epithelial cultures and thymus is not Pdx-1 dependent and that proinsulin expression is unresponsive to systemic glucose levels, consistent with our previous finding of compensatory Ins1 and Ins2 upregulation in the pancreas but not in the thymus in heterozygous knockouts [11]. These results also suggest that Pdx-1, Glut-2 and Gck may be expressed in thymus as self-antigens rather than functional regulatory molecules. A previous study [31] failed to detect Pdx-1 in the thymus, a finding that is probably only a question of sensitivity, as that study used
multiplex PCR for detection of several pancreatic transcription factors simultaneously and PCR was limited to 27 cycles.

A considerable amount of work remains to be done to truly characterize these proinsulin producing mTEC cell lines. Microarray expression profiling will reveal other tissue restricted antigens co-expressed with insulin along, perhaps, with important regulatory molecules and transcription factors. Such profiling with or without prior induction of Aire will allow us to directly study the importance of Aire expression on tissue restricted antigens and any corresponding transcriptions factors. It will be also interesting to see how our mTEC cultures react to a variety of environmental stimuli in order to further understand their regulation in the thymus. Does the expression profile of the cells change in the presence of cytokines and other regulatory molecules that are likely present in the thymic medulla? These cells will also be used to see whether proinsulin is secreted as intact peptide or directly processed into antigenic epitopes.

Thus, the cell lines we generated can be used to address a number of questions that are very important in understanding the role of the thymus in self-tolerance to tissue-specific antigens in general and to diabetes pathogenesis in relation to insulin autoreactivity in particular.
4.6 Acknowledgments

This work was supported by a grant from the Canadian Diabetes Association in honor of the late Emily C. Haynes and the Juvenile Diabetes Foundation. M.P. was supported by studentships from the Montreal Children’s Hospital Research Institute, Fonds de la Recherche en Santé Québec, McGill University Medicine, and the Canadian Institutes of Health Research. The authors thank Dr. Jacques Jami for his generous gift of the knockout mice and the support of all the students and staff at the Montreal Children’s Hospital Research Institute.
4.7 References


5.0 CHAPTER FIVE: Microarray Analysis of Cultured Proinsulin Producing Medullary Thymic Epithelial Cells
Michael Palumbo, Dina Levi and Constantin Polychronakos

Manuscript in preparation

5.01 Contribution of Authors

Michael Palumbo: Experimental design and strategy. Cell cultures, RNA extraction, RT-PCRs and microarray data analysis.

Dina Levi: Verification of proinsulin expression by RT-PCR, assisted with microarray data analysis.

Dr. Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

5.02 Objective

Having now isolated two proinsulin producing and two non-proinsulin producing thymic medullary epithelial cell lines (chapter 4), our next goal was to characterize the expression patterns of the proinsulin producing cell lines versus the non-producing cell lines.
5.1 Abstract

We have previously isolated two proinsulin producing and two non-proinsulin producing medullary thymic epithelial cell lines. Although the cell lines lost the expression of *Aire* once cultured, by real-time PCR we demonstrated that the proinsulin producing cell lines express proinsulin 8-10 fold over the non-proinsulin producing cell lines and we further demonstrated that all the isolated cell lines exhibit a medullary epithelial cell phenotype by FACS analysis (G8.8+ , UEA-1+, CD45-, DEC205- and MHCII+) and a cuboidal/cobblestone/epithelial morphology by microscopy. Previous groups comparing mixed medullary vs. cortical thymus epithelial cells reported the over-expression of over 500 genes, including over 150 tissue restricted antigens in the medullary epithelial cells. Using statistical analysis these groups reported that the over-expressed genes are organized into chromosomal clusters containing anywhere between 3-16 genes within 200kB regions and as such they suggest chromosome re-modeling may play an important role in tissue restricted antigen expression. Microarray analysis of the proinsulin producing and the non-proinsulin producing cell lines revealed over-expression of approximately 50 genes, in the proinsulin producing lines, with approximately half of these genes corresponding to tissue restricted antigens and with virtually every tissue of the body represented. Mapping of the over-expressed genes did not reveal any significant clustering of genes suggesting that the clustering phenomenon does not occur in the more narrow/specialized lineages of medullary epithelial cells and that clustering may be an artifact associated with using large subsets of medullary epithelial cells for analysis.
5.2 Introduction

As the site of T cell development, the thymus plays a critical role in positively selecting for T cells whose TCR (T cell receptor) is capable of recognizing antigen presented by host MHC (Major Histocompatibility Complex) while simultaneously negatively selecting for T cells that can recognize self-antigens. Positive selection occurs in the thymic cortex [4] whereas negative selection occurs quite extensively in the thymic medulla [5]. The effectiveness of negative T cell selection, which is also referred to as thymic central tolerance induction, has been shown to be dependant on the expression level of the self-antigen(s) in question within the thymic medulla [6-11]. Additionally, although specific details are lacking, it is generally accepted that the mechanism of thymic central tolerance induction requires the binding of the T cell’s TCR to self-antigen, via antigen presenting cells (APS), invoking either apoptotic deletion of autoreactive T cells [5, 12] or the formation of regulatory/suppressor T cells [13].

The efficiency of thymic central tolerance induction is not believed to be perfect but rather it is hypothesized to function in conjunction with peripheral tolerance to prevent autoimmune disease. Be that as it may, the discovery, in the last decade, that many (>152) peripheral tissue-restricted antigens (TRAs) are expressed in the thymic medulla is redefining the weight thymic central tolerance induction is believed to have in preventing autoimmune disease [14, 15]. In addition, although it was previously believed that the thymus was important only early on in host development, largely because the thymus involutes with age, more recent studies have suggested that despite this involution, T cells continue to develop in the thymus throughout life [16, 17].

121
The cells expressing the bulk of the TRAs detected in the thymic medulla are medullary thymic epithelial cells (mTECs). Indeed, a number of studies using purified preparations of mTECs, cortical thymic epithelial cells (cTECs), dendritic cells (DC) and macrophages, in conjunction with microarray analysis, real-time PCR and RT-PCR, have demonstrated that only a minor number of the TRAs are expressed by thymic DC, cTECs and macrophages while the vast preponderance are exclusively expressed by mTECs [14, 15, 18, 19]. In addition, it should be noted that many TRAs expressed by DC or cTECs are also expressed by mTECs [19]. Although such results implicate mTECs as the main cells responsible for tolerance induction, as stated previously, efficient central tolerance induction is believed to require T cell TCR interaction with APCs. Interestingly, recent albeit controversial results, have suggested that although mTECs can efficiently present self-antigens to CD8+ T cells the same is not true for CD4+ T cells – tolerance induction in the CD4+ subset seems to require uptake and presentation of self-antigens by DC [12, 20]. In addition, a very recent study implicated the uptake and presentation of self-antigens by DC, located in the vicinity of Hassall’s Corpuscles, as being necessary for the formation of regulatory T cells [13]. Thus, although expression of antigens is highly restricted to mTECs, direct presentation of the antigens to specific subsets of T cells could be mediated by other APCs in the medulla.

At present, there is very little known about the mechanism(s) allowing mTECs to express and present such a vast array of TRAs. The transcription factor, Autoimmune Regulator (AIRE), has been shown to be important for the expression of approximately 12% of the TRAs known to be expressed in the thymus, but a defined mechanism for AIRE induction is still lacking [14, 21]. A number of microarray studies comparing the
expression profiles of mTECs vs cTECs, isolated from both human and mouse thymus, have demonstrated that a significant number of the genes expressed by mTECs but not by cTECs were located within chromosomal clusters – clusters within windows of 35kB to 200kB were noted as being significant [14, 15]. Clustering of over-expressed genes was also found in microarray analysis of aire/- deficient mTECs although fewer clusters and fewer genes per cluster were identified [14, 22]. Such findings suggest that chromosomal alterations/re-modeling likely play an important role in allowing for and regulating TRA expression. Although possibly true, it is also important to note that such studies have important limitations in that analysis were conducted on total thymus mTECs that are likely composed of many different mTEC lineages, each specializing in different antigens. It is known that proinsulin is expressed only by 1-3% of all mTECs [18, 19], the remaining mTECs likely specializing in different sets of antigens. Microarray studies on such mixed subsets do not clearly define which antigens are expressed by unique mTEC lineages (eg. the proinsulin producing lineage) and what mechanisms allow for TRA expression in these lineages.

Two theories of mTEC development have been proposed to explain TRA expression – the “terminal differentiation model” [14] and the “mosaic model”[23]. The terminal differentiation model holds that mTEC maturation/differentiation results in the expression of an assortment of tissue restricted antigens. This model denotes CD80^hi/UEA-1+/ Aire expressing mTECs as more mature/differentiated mTECs capable of expressing a greater number of TRAs - it is further proposed that such cells do not share similar characteristics to the peripheral cells expressing the antigens. This is in contrast to the mosaic model that supposes that the medulla is a “patch quilt” of mTECs
emulating the expression patterns of peripheral cell lineages (ie. lung epithelial cells, pancreatic β-cells, etc...) with the corresponding mechanism of TRA expression/regulation (occurring in the peripheral cells) maintained. Evidence in support of this model was recently published describing the coordinated expression of numerous respiratory/lung antigens by an organized subset of thymic epithelial cells[24]. This model is not supported, however, by our findings demonstrating that thymic proinsulin expression is differentially regulated from pancreatic proinsulin[6].

Thymic proinsulin expression and its importance in the predisposition to Type 1 diabetes is of significant interest to our group. Indeed, in both mice and humans, a number of studies have directly demonstrated that thymic proinsulin levels can modify Type 1 diabetes susceptibility [6-11, 25]. In addition, findings from our lab have confirmed previous findings that proinsulin is expressed by 1-3% of all mTECs [18, 19] - more specifically our group showed that proinsulin positive mTECs were often located within Hassall’s Corpuscles-like structures in mice [18] which have been implicated in numerous tolerance induction mechanisms including the aforementioned formation of regulatory T cells [13, 26-29].

We have previously reported (Chapter 4) the isolation of two proinsulin producing (PPC) and two non-proinsulin producing (NPP) mTEC cell lines from INS2KO/Immortomice. By real-time PCR we demonstrated that the PPC cell lines over-expressed proinsulin approximately 8-10 fold over the NPP lines. Interestingly, proinsulin has been shown to be an *Aire* sensitive TRA but CD80<sup>hi</sup> mTECs still produce proinsulin albeit at greatly reduced levels in the absence of *Aire* [14, 21]. We in fact demonstrated that although the PPC lines lost the expression of *Aire*, following a very
short time in culture, they were positive for CD80, by RT-PCR, and they also stained positively for UEA-1, a marker for the hypothesized more mature/differentiated mTEC. In addition, all cell lines were found to have typical epithelial cuboidal/cobblestone morphology and typical mTEC (G8.8⁺, CD45⁻, MHC II⁺, DEC205⁻) surface-marker phenotype [18, 19, 30-32]. Additionally, stimulation of Aire expression resulted in significant up-regulation (2.5 fold) of proinsulin expression in the PPC lines but not the NPP lines.

We describe here the microarray analysis of the two PPC lines versus the NPP lines in the absence of Aire expression. We report the over-expression, of 4 fold or more, of 47 genes in the PPC cell lines versus the NPP lines and that approximately 50% of these genes can be considered TRAs. We do not report chromosomal clustering.
5.3 Materials and Methods

Isolation of Proinsulin Producing Medullary Thymic Epithelial Cells

Cells were isolated and cultured as described in Chapter 4 and passaged at 33°C without IFN-γ.

RNA Preparation, cDNA Synthesis, RT-PCR and Microarray Analysis

Cells from each of the cell lines, PPC1, PPC2, NPP1 and NPP2 were grown in triplicate in 10 cm tissue culture dishes to 80-90% confluency, at 33°C in DMEM with D-Valine (US Biological, Swampscott, MA, USA, Cat-M3861-20, 10% FBS, 1x Sodium Pyruvate (Gibco, Rockville, MD, USA), 1x Non-essential amino acids (Gibco), 50μM β-mercaptoethanol and 50μg/ml Gentamicin antibiotic (Gibco)) media without IFN-γ. They were subsequently transferred into a 37°C incubator for 48hrs before extracting RNA. RNA was extracted using the RNeasy Mini Kit (Qiagen, Mississauga, Ontario, Canada) and approximately 2μg of the extracted RNA was treated with Ambion DNase I (Austin, Texas, USA) and reverse transcribed using Random Primers (Gibco) and Superscript II Reverse Transcriptase (Gibco). Parallel samples in which Reverse Transcriptase was omitted were always included to confirm the absence of DNA contamination. PCRs for proinsulin, (detecting both INS1 and 2, confirmed the proinsulin producing nature of the PPC cell lines (data not shown, see Chapter 4). Primer sequences, conditions, reagents and amount of DNase I treated RNA used for each of the PCRs have been previously described (Chapter 4). 10μg of RNA from each of the triplicate samples was sent to the Montreal Genome Center for Microarray analysis on the Affymetrix (Santa Clara, CA 95051, USA) Mouse Genome 430 2.0 Arrays.
(containing 45102 probe sets representing 39,000 transcripts). Prior to microarray analysis, RNA quality was verified using electrophoretic detection of the 18S and 28S peaks in the Agilent 2100 Bioanalyzer.

**Microarray Data Analysis**

Microarray data was provided to us in both MicroArray Suite MAS5.0 (Affymetrix) and Robust Multiple-array Averaging (RMA) formats [33]. Preliminary analysis, indicated that usage of either data, MAS5.0 or RMA, resulted in similar expression patterns. The entire analysis was thus completed using RMA formatted data. Detailed manipulation of data is described in the results.

**Identification of Transcription Factor Binding Sites in the Insl and Ins2 Promoters**

One thousand base-pairs upstream from the transcriptional start site of Murine *Ins1* and *Ins2* genes known to contain the important regions for pancreatic *Ins* expression [34] (Ensembl ID for *Ins1* ENSMUSG0000035804, and *Ins2* ENSMUSG0000000215) were analysed for the presence of transcription factor binding sites using MatInspector [35].
5.4 Results

As an initial analysis we compared the reproducibility of the microarray data obtained from the three replicate samples of a single cell line. A plot of the data from PPC2.2 versus PPC2.3 is shown in figure 1A demonstrating a very linear symmetrical plot indicating very similar results were obtained for each of these two samples. A similar plot was obtained for the comparison of NPP1.1 versus NPP1.2, as is shown in figure 1B, and identical plots were also obtained for all the remaining possible replicate comparisons (data now shown).

We thus proceeded to determine the degree of similarity of the microarray expression profiles across the cell lines (ie. PPC1 versus PPC2). Average signals from the triplicate replicates of PPC1 were plotted against the average signals of the triplicate replicates of PPC2 yielding the plot in figure 2A. Figure 2B is the same comparison but using the average signal of the two NPP cell lines and figure 1C is a plot of the average signal from the NPP1 cell line versus the PPC1 cell line. As was expected, based on their isolation as proinsulin producing mTECs and thus very likely cell lines from the same narrow proinsulin producing mTEC lineage, the plot of PPC1 versus PPC2 had fewer variations in expression pattern, when compared to the NPP1 versus PPC1 cell lines – these latter two cell lines should only be related through their broad mTEC phenotype. Unexpectedly, the plot of NPP1 versus NPP2 also showed less variation than the NPP1 versus PPC1 even though we did not anticipate any similarities between the two NPP cell lines.
Figure 1: Reproducibility of microarray data obtained between triplicate samples of the same cell line. A) Plot of microarray signals from PPC 2.2 versus PPC 2.3 and B) signals from NPP1.1 versus NPP1.2
Figure 2: Degree of similarity of the microarray expression data obtained for each of the cell lines. A) Average signal from the triplicates of PPC 1 versus PPC 2 B) Average signal from the triplicates of NPP 1 versus NPP 2 C) Average signal from the triplicates of NPP 1 versus PPC 1
Our next point of interest was to determine if both of the PPC cell lines overexpressed the same genes (i.e. were positive for over-expression of the same probe sets) relative to the NPP cell lines. We considered over-expression to mean a difference of 4 fold or more. To determine this, we plotted the average signal from the NPP (cell lines 1 and 2 combined) cell lines subtracted from the signal obtained from either the PPC1 or PPC2 cell line (figure 3A). Note that with such a plot, and as is evident in figure 3A, a linear correlation between the two cell lines is obtained if they share a given number of over-expressed genes. A linear correlation was also obtained when graphing this plot for the NPP cell lines (figure 3B: the average signal of both PPC cell lines was subtracted from the signal obtained from either NPP1 or 2), confirming that indeed the NPP cell lines did over-express a given number of the same genes. This is in contrast to a plot of random combinations (figure 3C, the average of PPC1 and NPP2 subtracted from NPP1 versus the average of PPC2 and NPP1 subtracted from NPP2) which yielded a dispersed scatter-plot with no obvious linear correlation indicating the PPC cell lines and NPP cell lines did not share a pattern of over-expressed genes. To appreciate this data further, we plotted graphs of the number of over-expressed genes (>4 fold difference or more) in the PPC versus the NPP lines by subtracting the average signal of the NPP cells lines from the average signal of the PPC cell lines (figure 4). We did the same for a number of combinations including subtracting the average signal obtained from the PPC cell lines from the average signal of the NPP cell lines to determine which genes are over-expressed in the NPP cell lines relative to the PPC cell lines. As shown in figure 4, the only comparisons yielding a significant number of the same over-expressed genes were
Figure 3: A) PPC1 – average signal from NPP1 and 2 versus PPC2 – average signal from NPP1 and 2 – note the linear correlation indicating the same genes are overexpressed. B) Same comparison for NPP1 and 2 and C) A random comparison
Figure 4: Number of over-expressed genes (4 fold or more) versus fold difference in log₂ scale. 60 genes are over-expressed by the PPC cell lines relative to the NPP cell lines. 45 genes are over-expressed by the NPP cell lines relative to the PPC cell lines. Remaining comparisons have: AvePPC1NPP2-AvePPC2NPP2 (12 over-expressed genes), AvePPC2NPP1-AvePPC1NPP2 (8 over-expressed genes), AvePPC1NPP1-AvePPC2NPP2 (2 over-expressed genes) and PPC2NPP2-PPC1NPP1 (1 over-expressed gene).
the PPC-NPP and NPP-PPC plots with 60 and 45 genes (probe-sets) over-expressed greater than four fold, respectively. The remaining combinations had 18, 9, 2 and 1 gene(s)/probe sets over-expressed.

From the above data, we identified each of the genes and noted that some genes were represented more than once via different probe sets. These duplicate representations of the genes were removed retaining only the probe set that noted the greatest fold difference for the gene. Tables 1 and 2 list the over-expressed genes in the PPC versus the NPP cell lines and the NPP versus the PPC cell lines, respectively. Note that following the removal of the duplicate genes there remained 47 over-expressed genes in the PPC cell lines and 39 in the NPP cell lines. Included in the tables 1 and 2 are the gene symbol, gene title/proposed function, the fold difference (in log₂ scale) and, the tissue over-expression profile. Gene name, gene title/supposed function were obtained from the Affymetrix database for the Mouse 430 2.0 microarray. The tissue over-expression profile was obtained from SymAtlas [36] and Unigene [37]. We also included, in the tables, whether or not the same genes (or similar genes, ie. belonging to the same family and/or having a similar function) had been previously reported as over-expressed in mTECs (versus cTECs) in mouse thymus [14]. The notation (***) with regards to gene titles, refers to genes that were represented by 2 or more probe sets. With regards to tissue expression profiles, we reported the expression of the gene as restricted or specific (ie. TRA) if it was expressed in less than 5 tissues at at a minimum 3 fold higher (those at the minimum of 3 fold higher are denoted with a *) than the median expression of the gene in all tissues. Using this definition, 23 and 20 genes were
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Title</th>
<th>Fold</th>
<th>Tissue expressed: 10x median (*3x median)</th>
<th>Previously Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>1110033J19Rik</td>
<td>RIKEN cDNA 1110033J19 gene (ribosomal RNA binding protein)</td>
<td>6.10</td>
<td>placenta, testis, umbilical cord, gall bladder*</td>
<td></td>
</tr>
<tr>
<td>Rpp56</td>
<td>retinoblastoma GTPase regulator interacting protein 1</td>
<td>4.89</td>
<td>embryo, retina*</td>
<td></td>
</tr>
<tr>
<td>AA14J917</td>
<td>expressed sequence AA14J917 (amino acid metabolism)</td>
<td>4.49</td>
<td>*</td>
<td>Galm</td>
</tr>
<tr>
<td>Galm</td>
<td>glycine amidotransferase (L-arginine-glycine amidotransferase)</td>
<td>4.26</td>
<td>*</td>
<td>Galm</td>
</tr>
<tr>
<td>Fgl1**</td>
<td>fibroblast growth factor 5</td>
<td>4.24</td>
<td>embryo</td>
<td></td>
</tr>
<tr>
<td>111001J22Rik</td>
<td>RIKEN cDNA 111001J22 gene (DNA Binding Protein)</td>
<td>4.17</td>
<td>heart</td>
<td></td>
</tr>
<tr>
<td>Pir</td>
<td>pirin</td>
<td>4.16</td>
<td>* (Unigene)</td>
<td></td>
</tr>
<tr>
<td>Opp4***</td>
<td>dippeptidylpeptidase 4</td>
<td>3.06</td>
<td>small intestine</td>
<td></td>
</tr>
<tr>
<td>Mine</td>
<td>Membrane metallo endopeptidase, mRNA</td>
<td>3.02</td>
<td>prostate and small intestine (very high) also fat tissue and kidney</td>
<td></td>
</tr>
<tr>
<td>IG5cl***</td>
<td>immunoglobulin family member 21A (intracellular based movement)</td>
<td>2.97</td>
<td>trigeminal, dorsal root ganglia</td>
<td>KGB</td>
</tr>
<tr>
<td>Ang1***</td>
<td>angiopegin, ribonuclease A family, member 1</td>
<td>3.34</td>
<td>liver</td>
<td>Ang4</td>
</tr>
<tr>
<td>Odc1y***</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked</td>
<td>3.41</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Meta2***</td>
<td>membrane-spanning 4-domains, subfamily A, member (C020)</td>
<td>3.39</td>
<td>&gt;4 but 6 cells very high</td>
<td></td>
</tr>
<tr>
<td>Cd44***</td>
<td>CD44 antigen (hematopoietic progenitor cell antigen)</td>
<td>3.30</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Pou2f1</td>
<td>POU domain, class 3, transcription factor 1</td>
<td>3.28</td>
<td>epidermis, embryo, dorsal striatum, hippocampus*</td>
<td></td>
</tr>
<tr>
<td>Zfp533</td>
<td>zinc finger protein 533</td>
<td>3.13</td>
<td>CNS, ovary, retina</td>
<td>Zfp532</td>
</tr>
<tr>
<td>K01-19</td>
<td>keratin complex 1, acidic, gene 19</td>
<td>3.00</td>
<td>&gt;4</td>
<td>K01-19</td>
</tr>
<tr>
<td>LOC346190</td>
<td>similar to CDNA sequence BC061212</td>
<td>2.87</td>
<td>No available data</td>
<td></td>
</tr>
<tr>
<td>BBD244752**</td>
<td>PREDICTED; similar to OTTHUMP00000065831</td>
<td>2.73</td>
<td>cerebellum, testes, kidney, embryonic (Unigene)</td>
<td></td>
</tr>
<tr>
<td>BB114368**</td>
<td>---</td>
<td>2.71</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Hend2***</td>
<td>heart and neural crest derivatives expressed transcript</td>
<td>2.70</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Gabra1</td>
<td>gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 1</td>
<td>2.69</td>
<td>CNS, retina</td>
<td></td>
</tr>
<tr>
<td>1810011O10Rik</td>
<td>chromosome 8 open reading frame 4</td>
<td>2.62</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>3732412O22Rik***</td>
<td>RIKEN cDNA 3732412O22 gene (myosin light chain)</td>
<td>2.61</td>
<td>heart, fertilized egg, oocyte*</td>
<td></td>
</tr>
<tr>
<td>Ath663860</td>
<td>EED A356396</td>
<td>2.57</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Eif4sk</td>
<td>eukaryotic translation initiation factor 2, subunit 3</td>
<td>2.55</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Zic1</td>
<td>zinc finger protein of the cerebellum 1</td>
<td>2.55</td>
<td>CNS, embryo</td>
<td></td>
</tr>
<tr>
<td>Frmomo***</td>
<td>FERM domain containing 4</td>
<td>2.53</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Ahhr</td>
<td>aryl-hydrocarbon receptor repressor</td>
<td>2.50</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>C03004500D6Rik</td>
<td>RIKEN cDNA C03004500D6 gene</td>
<td>2.46</td>
<td>&gt;4 (Unigene)</td>
<td></td>
</tr>
<tr>
<td>Adcy7</td>
<td>adenylate cyclase 7</td>
<td>2.42</td>
<td>trachea, lymph node, T cells, B cells</td>
<td></td>
</tr>
<tr>
<td>Casp1</td>
<td>caspase 1</td>
<td>2.41</td>
<td>&gt;4, small and large intestine very high</td>
<td>Casp1</td>
</tr>
<tr>
<td>Pbx2</td>
<td>paired-like homeodomain transcription factor 2</td>
<td>2.26</td>
<td>umbilical cord, tongue, pituitary</td>
<td></td>
</tr>
<tr>
<td>Jund1d</td>
<td>junD, AT rich interactive domain 1D (Rbp2 like)</td>
<td>2.34</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Elav2**</td>
<td>ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2</td>
<td>2.29</td>
<td>CNS, fertilized egg, oocyte</td>
<td></td>
</tr>
<tr>
<td>9230117T10Rk</td>
<td>RIKEN cDNA 9230117T10 gene</td>
<td>2.24</td>
<td>CNS, lung, lymph nodes, trachea**</td>
<td></td>
</tr>
<tr>
<td>Tcf21</td>
<td>transcription factor 21</td>
<td>2.21</td>
<td>lung, uterus, ovary</td>
<td>Tcf3</td>
</tr>
<tr>
<td>Cystai***</td>
<td>cysteinyl leukotriene receptor 1</td>
<td>2.20</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Adh7</td>
<td>alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide</td>
<td>2.19</td>
<td>&gt;4</td>
<td>Adh4</td>
</tr>
<tr>
<td>Neg5</td>
<td>neuronal growth regulator 1</td>
<td>2.13</td>
<td>CNS, pituitary, retina**</td>
<td></td>
</tr>
<tr>
<td>BC048916</td>
<td>CDNA sequence BC048916</td>
<td>2.10</td>
<td>&gt;4 (Unigene)</td>
<td></td>
</tr>
<tr>
<td>All427515</td>
<td>expressed sequence AI427515</td>
<td>2.10</td>
<td>&gt;4 (Unigene)</td>
<td></td>
</tr>
<tr>
<td>9030811N11Rk</td>
<td>RIKEN cDNA 9030811N11 gene</td>
<td>2.07</td>
<td>&gt;4</td>
<td>Akr1c12, Akr1c21</td>
</tr>
<tr>
<td>Ino1</td>
<td>inositol phosphorelated homodimer (Drosophila)</td>
<td>2.03</td>
<td>&gt;4 but mammary gland very high</td>
<td></td>
</tr>
<tr>
<td>Gsta4</td>
<td>glutathione S-transferase, alpha 4</td>
<td>2.01</td>
<td>stomach, trachea, bladder, retina</td>
<td>Gsta2</td>
</tr>
<tr>
<td>Mmp3</td>
<td>matrix metalloproteinase 3</td>
<td>2.01</td>
<td>lung, retina, epididymis, adipose tissue</td>
<td>Mrmp7</td>
</tr>
<tr>
<td>Id4</td>
<td>inhibitor of DNA binding 4</td>
<td>2.00</td>
<td>CNS, umbilical cord, bladder</td>
<td>Id4</td>
</tr>
</tbody>
</table>

Table 1: Genes over-expressed by the PPC cell lines relative to the NPP cell lines. 47 genes are over-expressed, 23 of which can be considered tissue restricted (expressed in less than 5 tissues at either 10x the median or at a minimum 3x the median labeled*). Gene symbols labeled (**) are Representative Public IDs and have yet to be assigned a gene symbol. Gene symbols labeled (***) were represented by 2 or more probe sets. Fold differences are expressed in log2 format. 3 genes have been previously reported as over-expressed in mTECs vs cTECs, along with 9 genes that are very similar (ie. belonging to the same family and/or having a similar function) to others previously reported.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Title</th>
<th>Fold Δ</th>
<th>Tissue expressed in 10x median (*3x median)</th>
<th>Previously Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xlat**</td>
<td>inactive X specific transcripts</td>
<td>7.41</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>CD91</td>
<td>cysteine dioxygenase 1, cytosolic</td>
<td>6.06</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Sat**</td>
<td>bone marrow stromal cell antigen 1</td>
<td>5.08</td>
<td>small intestine</td>
<td></td>
</tr>
<tr>
<td>Scara3</td>
<td>scavenger receptor class A, member 3</td>
<td>4.28</td>
<td>digits, trachea, bone</td>
<td></td>
</tr>
<tr>
<td>Gpx7</td>
<td>glutathione peroxidase 7</td>
<td>4.13</td>
<td>ovary, umbilical cord</td>
<td></td>
</tr>
<tr>
<td>Mdk</td>
<td>midkine embryonic expression, neural/adrenal development</td>
<td>4.08</td>
<td>embryo, uterus, ovary, umbilical cord</td>
<td></td>
</tr>
<tr>
<td>Cd200</td>
<td>Cd200 antigen, protein binding</td>
<td>4.08</td>
<td>proteoglycan, hypothalamus, lung, adipose tissue</td>
<td></td>
</tr>
<tr>
<td>Plac8</td>
<td>placenta-specific 8</td>
<td>3.80</td>
<td>&gt;4</td>
<td>Plac8, Myl4,7</td>
</tr>
<tr>
<td>Myb</td>
<td>myosin, light polypeptide 9, regulatory</td>
<td>3.79</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>BC021831**</td>
<td>Muc mucosae, clone IMAGE:3963021</td>
<td>3.78</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Tbid</td>
<td>thrombomodulin, placental and lung development function (intestinal)</td>
<td>3.76</td>
<td>lung, digits</td>
<td></td>
</tr>
<tr>
<td>Pipm</td>
<td>protein tyrosine phosphatase, receptor type, M</td>
<td>3.28</td>
<td>olfactory, heart, lung</td>
<td></td>
</tr>
<tr>
<td>Rex3</td>
<td>reduced expression 3</td>
<td>3.06</td>
<td>embryo, ovary, pituitary</td>
<td></td>
</tr>
<tr>
<td>Myo7b</td>
<td>myosin Vila</td>
<td>2.77</td>
<td>adrenal, testes, umbilical cord*</td>
<td>Myo5b, Myl4,7</td>
</tr>
<tr>
<td>Csr2</td>
<td>carboxyl reductase 2</td>
<td>2.65</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>Ctgrb5 III Mhp</td>
<td>C1q and tumor necrosis factor related protein 5</td>
<td>2.62</td>
<td>mammary gland, prostate, tongue, digits</td>
<td></td>
</tr>
<tr>
<td>Cish</td>
<td>calpepsin H</td>
<td>2.61</td>
<td>&gt;4</td>
<td>Cish</td>
</tr>
<tr>
<td>Dkk2</td>
<td>Dickkopf homolog 2 (Xenopus laevis), lens development, Wnt pathway</td>
<td>2.61</td>
<td>&gt;4</td>
<td>Dkk1</td>
</tr>
<tr>
<td>Trmb2**</td>
<td>tropomyosin T1, cardiac, muscle and heart development</td>
<td>2.39</td>
<td>heart</td>
<td></td>
</tr>
<tr>
<td>Hsp92</td>
<td>HIV-1 tat interactive protein 2, homolog (human), apoptosis</td>
<td>2.37</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>2410129F14Rik***</td>
<td>RIKEN cDNA 2410129F14 gene, GTPase activity</td>
<td>2.36</td>
<td>thymus, ovary, mammary gland, embryo*</td>
<td></td>
</tr>
<tr>
<td>Cd100</td>
<td>CD100 antigen, lactin, sugar binding</td>
<td>2.32</td>
<td>liver, ovary, adipose tissue</td>
<td></td>
</tr>
<tr>
<td>BB745211**</td>
<td>—</td>
<td>2.29</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>0610006010Rik</td>
<td>RIKEN cDNA 0610006010 gene</td>
<td>2.26</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>BE370819**</td>
<td>Transcribed locus</td>
<td>2.23</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Nid1</td>
<td>nidogen 1, glycoprotein, necessary for lung/cardiac development</td>
<td>2.22</td>
<td>adipose tissue, placenta, umbilical cord, blastocyst</td>
<td></td>
</tr>
<tr>
<td>0530406810Rik</td>
<td>Bone marrow stromal cell antigen 1</td>
<td>2.19</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>MGC239678</td>
<td>3-ketoacyl-CoA thiolase B</td>
<td>2.15</td>
<td>liver, bladder, kidney</td>
<td></td>
</tr>
<tr>
<td>Icam1</td>
<td>intercellular adhesion molecule</td>
<td>2.15</td>
<td>lung, lymph node, trachea, immune cells</td>
<td></td>
</tr>
<tr>
<td>Ankrd1***</td>
<td>ankyrin repeat domain 1 (cardiac muscle), cardiac progenitor cells</td>
<td>2.14</td>
<td>heart, umbilical cord</td>
<td>Ankrd1</td>
</tr>
<tr>
<td>BB770665**</td>
<td>Transcribed locus</td>
<td>2.14</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Has2</td>
<td>hyaluronan synthase 2, several embryonic functions</td>
<td>2.14</td>
<td>umbilical cord</td>
<td>Has2</td>
</tr>
<tr>
<td>B3gal3</td>
<td>UDP-GalbetaGalCNA beta 1,3-galactosyltransferase, polyphenol 3</td>
<td>2.13</td>
<td>CNS, (ovary, ovary, umbilical cord, placenta) pituitary*</td>
<td></td>
</tr>
<tr>
<td>Serping1</td>
<td>serine (or cysteine) peptidase inhibitor, clade G, member 1</td>
<td>2.08</td>
<td>more than 5</td>
<td>Serping1</td>
</tr>
<tr>
<td>Agrp1</td>
<td>angiotensin receptor 1</td>
<td>2.07</td>
<td>adrenal gland, prostate, adipose tissue</td>
<td></td>
</tr>
<tr>
<td>Tcf5</td>
<td>transcription factor-like 5 (basic helix-loop-helix), spermatogenesis</td>
<td>2.05</td>
<td>testes</td>
<td></td>
</tr>
<tr>
<td>BG071670**</td>
<td>Transcribed locus</td>
<td>2.04</td>
<td>no data available</td>
<td></td>
</tr>
<tr>
<td>Mmne1</td>
<td>multimerin 1</td>
<td>2.01</td>
<td>&gt;4</td>
<td></td>
</tr>
<tr>
<td>BB097841**</td>
<td>Transcribed locus</td>
<td>2.01</td>
<td>no data available</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Genes over-expressed by the NPP cell lines relative to the PPC cell lines. 39 genes are over-expressed, 20 of which can be considered tissue restricted (expressed in less than 5 tissues at either 10x the median or at a minimum 3x the median - labeled*). Gene symbols labeled (**) are Representative Public IDs and have yet be assigned a gene symbol. Gene symbols labeled (****) were represented by 2 or more probe sets. Fold differences are expressed in log2 format. 8 genes have been previously reported as over-expressed in mTECs vs cTECs, along with 4 genes that are very similar (ie. belonging to the same family and/or having a similar function) to others previously reported.
determined as being tissue restricted in the PPC and NPP cell lines, respectively. Three
of the genes over-expressed in the PPC cell lines had been previously reported and 9
other genes were similar (i.e. belonging to the same family and/or having a similar
function) to others reported whereas in the NPP cell lines, 8 had been previously reported
and 4 others were similar to others reported. Of the previously reported or similar genes,
8 and 6 have a tissue restricted expression profile in the PPC and NPP cell lines,
respectively. It is important to note that proinsulin was not one of the genes over-
expressed in the PPC cell lines and that \textit{XIST} was one of the genes over-expressed in the
NPP cell lines.

Table 3 denotes the tissues represented, by the tissue restricted antigens, and their
number of representations in the PPC and NPP cell lines. Interestingly, in both the PPC
or NPP cell lines, a wide array of different tissues are represented with some degree of
overlap between them and with numerous tissues being represented more than once.

Tables 4 and 5 list the chromosomal locations of all the over-expressed genes and
their start and stop sites obtained from MapIt along with their calculated inter-gene
distances for the PPC and NPP cell lines, respectively. All inter-gene distances are
greater than 200kB except for one pair of genes on chromosome 6 (gene symbol \textit{Frmd4b}
and \textit{Bc049816}) and a second pair on chromosome Y (gene symbol \textit{Jarid1d} and \textit{Eif2s3y})—
both of which are found in the PPC cell lines.

Analysis of 1kB upstream from the transcriptional start sites of \textit{Ins1} and \textit{Ins2},
using MatInspector, did not identify any obvious binding sites for transcription factors
over-expressed in the PPC. Two binding sites were identified, however, for a
heterodimer of the transcription factors \textit{Hand2} [38] (heart and neural crest derivatives
expressed transcript) and Tcf3 (transcription factor E2a) [39]—two basic helix-loop-helix transcription factors that must heterodimerize together to bind DNA as they cannot do so alone [40]. As is indicated in table 1, Hand2 is a transcription factor over-expressed in our PPC but expression of Tcf3 was not found. Interestingly, MatInspector analysis of the Ins upstream regions indicated that Ins1 only has a single Aire binding site as opposed to Ins2 that has two such binding sites (table 6).
Table 3: Tissues represented by the tissue restricted antigens over-expressed by the PPC and NPP cell lines.

<table>
<thead>
<tr>
<th>Proinsulin Positive Cell Lines</th>
<th>Non Proinsulin Producing Cell Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tissues</strong></td>
<td><strong># of Representations</strong></td>
</tr>
<tr>
<td>CNS</td>
<td>10</td>
</tr>
<tr>
<td>Retina</td>
<td>6</td>
</tr>
<tr>
<td>Embryo</td>
<td>5</td>
</tr>
<tr>
<td>Umbilical Cord</td>
<td>3</td>
</tr>
<tr>
<td>Trachea</td>
<td>3</td>
</tr>
<tr>
<td>Lung</td>
<td>3</td>
</tr>
<tr>
<td>Testis</td>
<td>2</td>
</tr>
<tr>
<td>Small Intestine</td>
<td>2</td>
</tr>
<tr>
<td>Pituitary</td>
<td>2</td>
</tr>
<tr>
<td>Ovary</td>
<td>2</td>
</tr>
<tr>
<td>Oocyte</td>
<td>2</td>
</tr>
<tr>
<td>Lymph Nodes</td>
<td>2</td>
</tr>
<tr>
<td>Kidney</td>
<td>2</td>
</tr>
<tr>
<td>Heart</td>
<td>2</td>
</tr>
<tr>
<td>Fertilized Egg</td>
<td>2</td>
</tr>
<tr>
<td>Fatty Tissue</td>
<td>2</td>
</tr>
<tr>
<td>Epidermis</td>
<td>2</td>
</tr>
<tr>
<td>Bladder</td>
<td>2</td>
</tr>
<tr>
<td>Uterus</td>
<td>1</td>
</tr>
<tr>
<td>Tongue</td>
<td>1</td>
</tr>
<tr>
<td>Stomach</td>
<td>1</td>
</tr>
<tr>
<td>Prostate</td>
<td>1</td>
</tr>
<tr>
<td>Placenta</td>
<td>1</td>
</tr>
<tr>
<td>Liver</td>
<td>1</td>
</tr>
<tr>
<td>Immune Cells</td>
<td>1</td>
</tr>
<tr>
<td>Gall Bladder</td>
<td>1</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Chromosome</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>C030045D06Rik</td>
<td>1</td>
</tr>
<tr>
<td>BB024475**</td>
<td>1</td>
</tr>
<tr>
<td>Cd34</td>
<td>1</td>
</tr>
<tr>
<td>Dpp4</td>
<td>2</td>
</tr>
<tr>
<td>Zfp533</td>
<td>2</td>
</tr>
<tr>
<td>Gatm</td>
<td>2</td>
</tr>
<tr>
<td>Mme</td>
<td>3</td>
</tr>
<tr>
<td>Pltx2</td>
<td>3</td>
</tr>
<tr>
<td>Adh7</td>
<td>3</td>
</tr>
<tr>
<td>Negr1</td>
<td>3</td>
</tr>
<tr>
<td>Elavl2</td>
<td>4</td>
</tr>
<tr>
<td>Pou3f1</td>
<td>4</td>
</tr>
<tr>
<td>3732412D22Rik</td>
<td>5</td>
</tr>
<tr>
<td>Fgt5</td>
<td>5</td>
</tr>
<tr>
<td>LOC546190</td>
<td>5</td>
</tr>
<tr>
<td>Frm4b</td>
<td>6</td>
</tr>
<tr>
<td>BC049816</td>
<td>6</td>
</tr>
<tr>
<td>At256396</td>
<td>6</td>
</tr>
<tr>
<td>1110033J19Rik</td>
<td>6</td>
</tr>
<tr>
<td>18100110I10Rik</td>
<td>8</td>
</tr>
<tr>
<td>Hand2***</td>
<td>8</td>
</tr>
<tr>
<td>Adcy7</td>
<td>8</td>
</tr>
<tr>
<td>At427515</td>
<td>8</td>
</tr>
<tr>
<td>Casp1</td>
<td>9</td>
</tr>
<tr>
<td>Mmp3</td>
<td>9</td>
</tr>
<tr>
<td>Gsta4</td>
<td>9</td>
</tr>
<tr>
<td>BB114398**</td>
<td>9</td>
</tr>
<tr>
<td>Zic1</td>
<td>9</td>
</tr>
<tr>
<td>Tcf21</td>
<td>10</td>
</tr>
<tr>
<td>Gabra1</td>
<td>11</td>
</tr>
<tr>
<td>Kr11-19</td>
<td>11</td>
</tr>
<tr>
<td>9030611N15Rik</td>
<td>13</td>
</tr>
<tr>
<td>Id4</td>
<td>13</td>
</tr>
<tr>
<td>Irx2</td>
<td>13</td>
</tr>
<tr>
<td>Ahhr</td>
<td>13</td>
</tr>
<tr>
<td>Ang1</td>
<td>14</td>
</tr>
<tr>
<td>Rprgr1p1</td>
<td>14</td>
</tr>
<tr>
<td>Kif21a</td>
<td>15</td>
</tr>
<tr>
<td>AA415817</td>
<td>16</td>
</tr>
<tr>
<td>Ms4a2</td>
<td>19</td>
</tr>
<tr>
<td>9230117N10Rik</td>
<td>19</td>
</tr>
<tr>
<td>1110018J23Rik</td>
<td>19</td>
</tr>
<tr>
<td>Cysltr1</td>
<td>X</td>
</tr>
<tr>
<td>Pir</td>
<td>X</td>
</tr>
<tr>
<td>Jarid1d</td>
<td>Y</td>
</tr>
<tr>
<td>Eif2e3y</td>
<td>Y</td>
</tr>
<tr>
<td>Ddx3y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 4: Inter-gene distances for over-expressed genes in PPC cell lines.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Chromosome</th>
<th>Start (bp)</th>
<th>Stop (bp)</th>
<th>Inter-Gene Distance (Kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tnnt2</td>
<td>1</td>
<td>137652816</td>
<td>137668671</td>
<td>24481</td>
</tr>
<tr>
<td>Cd302</td>
<td>2</td>
<td>60052832</td>
<td>60085257</td>
<td>24481</td>
</tr>
<tr>
<td>Serpina1</td>
<td>2</td>
<td>84566224</td>
<td>84576243</td>
<td>7154</td>
</tr>
<tr>
<td>Mdk</td>
<td>2</td>
<td>91730660</td>
<td>91733121</td>
<td>56363</td>
</tr>
<tr>
<td>Thbd</td>
<td>2</td>
<td>148095912</td>
<td>148099544</td>
<td>56363</td>
</tr>
<tr>
<td>Myl9</td>
<td>2</td>
<td>156466480</td>
<td>156473099</td>
<td>24078</td>
</tr>
<tr>
<td>Tcf15</td>
<td>2</td>
<td>180551368</td>
<td>180571890</td>
<td>24078</td>
</tr>
<tr>
<td>B3gal3</td>
<td>3</td>
<td>69662115</td>
<td>69686790</td>
<td>53164</td>
</tr>
<tr>
<td>BB397841**</td>
<td>3</td>
<td>122860900</td>
<td>122861603</td>
<td>53164</td>
</tr>
<tr>
<td>Dkk2</td>
<td>3</td>
<td>132022807</td>
<td>132117610</td>
<td>9171</td>
</tr>
<tr>
<td>BE370618**</td>
<td>3</td>
<td>N/AP</td>
<td>N/AP</td>
<td>N/AP</td>
</tr>
<tr>
<td>Gpx7</td>
<td>4</td>
<td>107898343</td>
<td>107904646</td>
<td>48285</td>
</tr>
<tr>
<td>Bst1</td>
<td>5</td>
<td>44107156</td>
<td>44131559</td>
<td>46662</td>
</tr>
<tr>
<td>Plac8</td>
<td>5</td>
<td>100794035</td>
<td>100812907</td>
<td>46662</td>
</tr>
<tr>
<td>Mmrn1</td>
<td>6</td>
<td>60873938</td>
<td>60918956</td>
<td>46662</td>
</tr>
<tr>
<td>Hlatip2</td>
<td>7</td>
<td>49627174</td>
<td>49642016</td>
<td>46662</td>
</tr>
<tr>
<td>Myo7a</td>
<td>7</td>
<td>97926593</td>
<td>97992926</td>
<td>46662</td>
</tr>
<tr>
<td>BG071670**</td>
<td>8</td>
<td>119454200</td>
<td>119455860</td>
<td>46662</td>
</tr>
<tr>
<td>Icam1</td>
<td>9</td>
<td>20766362</td>
<td>20779199</td>
<td>23074</td>
</tr>
<tr>
<td>C1qtnf5 /// Mfrp</td>
<td>9</td>
<td>43852837</td>
<td>4390182</td>
<td>23074</td>
</tr>
<tr>
<td>Cth</td>
<td>9</td>
<td>89852402</td>
<td>89873682</td>
<td>5992</td>
</tr>
<tr>
<td>MGC29978</td>
<td>9</td>
<td>118996741</td>
<td>119005791</td>
<td>29123</td>
</tr>
<tr>
<td>5830408B19Rik</td>
<td>10</td>
<td>9312780</td>
<td>9313970</td>
<td>29123</td>
</tr>
<tr>
<td>Cbr2</td>
<td>11</td>
<td>120545599</td>
<td>120548085</td>
<td>29123</td>
</tr>
<tr>
<td>2410129E14Rik</td>
<td>12</td>
<td>11908230</td>
<td>11909670</td>
<td>29123</td>
</tr>
<tr>
<td>BB745211**</td>
<td>12</td>
<td>77368500</td>
<td>77369720</td>
<td>29123</td>
</tr>
<tr>
<td>Nid1</td>
<td>13</td>
<td>13229817</td>
<td>13304390</td>
<td>17040</td>
</tr>
<tr>
<td>Agtr1</td>
<td>13</td>
<td>30343922</td>
<td>30390332</td>
<td>93579</td>
</tr>
<tr>
<td>BB021831**</td>
<td>13</td>
<td>123968950</td>
<td>123972990</td>
<td>93579</td>
</tr>
<tr>
<td>Scar3</td>
<td>14</td>
<td>64873505</td>
<td>64907854</td>
<td>93579</td>
</tr>
<tr>
<td>Has2</td>
<td>15</td>
<td>56495712</td>
<td>56524587</td>
<td>925670</td>
</tr>
<tr>
<td>BB776065**</td>
<td>15</td>
<td>982194360</td>
<td>982195430</td>
<td>925670</td>
</tr>
<tr>
<td>Cd200</td>
<td>16</td>
<td>45301440</td>
<td>45328223</td>
<td>21244</td>
</tr>
<tr>
<td>Ptprn</td>
<td>17</td>
<td>66572488</td>
<td>67259332</td>
<td>21244</td>
</tr>
<tr>
<td>Cdo1</td>
<td>18</td>
<td>46838536</td>
<td>46853673</td>
<td>21244</td>
</tr>
<tr>
<td>0810006010Rik</td>
<td>19</td>
<td>8838041</td>
<td>8839520</td>
<td>27338</td>
</tr>
<tr>
<td>Ankrd1</td>
<td>19</td>
<td>36177109</td>
<td>36184988</td>
<td>27338</td>
</tr>
<tr>
<td>Rex3</td>
<td>X</td>
<td>57764683</td>
<td>57765387</td>
<td>41898</td>
</tr>
<tr>
<td>Xist</td>
<td>X</td>
<td>99663092</td>
<td>99685952</td>
<td>41898</td>
</tr>
</tbody>
</table>

Table 5: Inter-gene distances for over-expressed genes in NPP cell lines.
<table>
<thead>
<tr>
<th></th>
<th>Ins1</th>
<th>Ins2</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Present</td>
<td>Aire</td>
<td>-750</td>
<td></td>
</tr>
<tr>
<td>Aire</td>
<td>Aire</td>
<td>-250</td>
<td></td>
</tr>
<tr>
<td>Hand2/Tcf3</td>
<td>Hand2/Tcf3</td>
<td>-240</td>
<td></td>
</tr>
<tr>
<td>Hand2/Tcf3</td>
<td>Hand2/Tcf3</td>
<td>-110</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: *Aire* and *Hand2/Tcf3* binding sites located within the 1000bp upstream of the *Ins1* and *Ins2* transcriptional start sites. Location refers to the distance from the translational start site.
5.5 Discussion

The finding that more than 152 peripheral tissue restricted antigens are expressed in mTECs [41] and that TRA expression levels can modulate susceptibility to autoimmune pathologies [6-9, 11, 25] is reshaping the scope and thus importance of thymic central tolerance in preventing the onset of autoimmune diseases. What allows for such a broad expression profile of self-antigens in mTECs is presently not known but the transcription factor Aire has been shown to result in the up-regulation of expression of approximately 12% of the TRAs known to be expressed in the thymus. Statistical analysis of microarray studies, has provided indirect evidence for the expression of clustered genes suggesting a mechanism involving chromosomal alteration [14, 15, 21, 22]. It is important to note, however, that the latter studies were performed on mixed mTECs, invariably composed of many lineages, of which the 1-3% of mTECs that produce proinsulin is but one. Thus, the importance of chromosomal clustering in these more specific individual lineages can only be assumed.

The objective of our group is to understand the mechanism(s) by which reduced thymic proinsulin levels confer susceptibility to Type 1 diabetes. Therefore, understanding the transcriptional regulation and subsequent translational and protein processing of proinsulin, as well as the functional mechanism behind the mTEC cells, is of utmost importance. To this end, we performed microarray analysis on the cell lines to determine: 1) if other self-antigens, besides proinsulin, are produced by the PPC, 2) if chromosomal clustering of over-expressed genes occurs in these cell lines and 3) identify transcription factors expressed by the PPC which can potentially result in their TRA expression pattern.
As described in the results section, initial analysis of the microarray data demonstrated that the two PPC cell lines were more similar in expression pattern than the PPC and NPP cell lines (figure 2). This was of course expected as the PPC cell lines should be from the same narrow mTEC lineage based on their isolation as proinsulin producing cell lines. Admittedly, when looking at the graphs (figure 2A versus C), the greater expression variation in C is very subtle but this was also expected since the cell lines are first and foremost mTECs and very likely express the same genes except for a small number that are characteristic of that narrow mTEC lineage. Interestingly, and unexpectedly, the comparison of the NPP cell lines also showed a more similar expression profile when compared to the profile of the PPC versus the NPP cell lines. Under normal circumstances, and with the assumption that many narrow/specialized mTEC lineages exist, we expected that random selection of two mTEC cell lines would result in the isolation of two different narrow mTEC lineages. We realized, however, that the isolation of the two similar NPP cell lines was likely a consequence of our methodology during their isolation (chapter 4). As opposed to the isolation of the PPC cell lines which were isolated independently from two 10cm culture dishes, the NPP cell lines were isolated from the same culture dish and it is possible that we selected two cell lines from the same clone following clonal growth/expansion. As such, the two NPP cell lines are likely clonal and not necessarily two primary NPP mTEC lineages that coincidentally happened to be of the same lineage.

The subsequent analysis described in figures 3 and 4 further support the similar expression profile of the two PPC cell lines and the NPP cell lines demonstrating that the PPC cell lines over-express 47 of the same genes and that the NPP cell lines over-express
39 of the same genes at >4 fold or more (when results from >1 probe sets for the same gene are removed from the set, figure 4). The significance of these findings can be appreciated when one considers the random comparisons (figure 4), with only one such comparison yielding some minor degree of similarity with 10 of the same genes being over-expressed with the majority of these genes on the lower end of the fold scale (Again, note that 10 over-expressed genes is obtained when duplicate genes are removed). The other comparisons over-expressed 6, 2 and 1 of the same gene(s).

Approximately 50% of the genes over-expressed in the PPC and NPP cell lines could be denoted as TRAs based on their expression in less than 5 tissues (tables 1 and 2). As shown in table 3, a wide variety of tissues were represented in each of the cell lines. Interestingly, more than a single self-antigen was over-expressed in both the NPP and PPC cell lines for numerous tissues including lung tissue (as has been reported in vivo [24]), but proinsulin was the only representative of the pancreas (see Chapter 4 for additional corroborating data). Importantly, proinsulin was not detected by microarray in the PPC cell lines but this is very likely due to the absence of *Aire* expression and the poor sensitivity of the proinsulin probe sets. In support of this, Derbinski *et al.* [14] were also not able to detect proinsulin when performing microarray analysis on *aire-/-* mTECs and their own observations suggested that the sensitivity of microarray technology is approximately 50%. It should be noted that RT-PCR for proinsulin was performed on all the RNA samples prior to being sent for microarray analysis simply to confirm the proinsulin producing nature of the PPC cell lines (data not shown). When comparing our results to those obtained by Derbinski *et al.*, approximately 11 of the genes over-expressed in either the PPC or NPP cell lines were reported as being over-expressed in
mTECs versus cTECs (tables 1 and 2). In addition, another 12 of the over-expressed genes were very similar (ie. belonging to the same family and/or having a similar function) to those already reported (tables 1 and 2). Considering Derbinski et al. reported the over-expression of approximately 500 genes (in mTECs versus cTECs), finding 11 exact matches out of 86 genes over-expressed in our cell lines, out of a pool of 30000 or so genes, is unlikely by chance alone, especially when one considers that that the microarrays used by Derbinski et al. had approximately 1/3 the number of transcripts of the Mouse 430 2.0 arrays. A further interesting point is that Derbinski et al., report the overexpression of approximately 150 TRAs (in mTECs versus cTECs) but estimate, based on the number of transcripts and sensitivity of the microarrays used (MGU74Av2) that >1000 TRAs are expressed in mTECs. We report approximately 20 or so TRAs expressed in the PPC and NPP cell lines and if we extrapolate these findings to be representative of the expression characteristics of other mTEC lineages approximately 50 different lineages would be required to express ~1000 TRAs, thus suggesting that 1-3% of all mTECs express proinsulin.

In contrast to a number of groups that have studied mTEC expression patterns in mixed mTECs, we do not report chromosomal clustering of the over-expressed genes in either the PPC or NPP cell lines, suggesting that chromosomal clustering is not a phenomenon correlated with the expression profiles of the narrower more specialized mTEC lineages. Of course, our present analysis is limited due to lack of Aire expression in the cell lines which would increase the number of over-expressed genes per cluster but is not likely to result in the over-expression of clustered genes as clustering has been reported even in the absence of Aire. Subsequent studies required to offer a potential
explanation for the expression profiles of the PPC and NPP lines will necessitate in-depth analysis of the promoter regions of these over-expressed genes. We performed a preliminary analysis of the *Ins1* and *Ins2* promoter regions which did not reveal any obvious binding sites for transcription factors over-expressed in our PPC cell lines. We did find, however, two binding sites for the heterodimer consisting of the transcription factors *Hand2* [38] (heart and neural crest derivatives expressed transcript) and *Tcf3* [39] (transcription factor E2a) – two basic helix-loop-helix transcription factors that must heterodimerize together to bind DNA [40]. Their collaborative effort has been shown to be crucial in heart, CNS and limb development and their expression has been documented in a number of tissues including pancreas, thymus and reproductive tissues [36, 42, 43]. Interestingly, within both the *Ins1* and *Ins2* promoter, one of the *Hand2/Tcf3* binding sites is directly adjacent to the *Aire* binding site suggesting a possible interaction. Based on the microarray results, *Tcf3* is not over-expressed in the PPC lines but *Hand2* is, yet both have been shown to be expressed in thymus tissue and more specifically Derbinksi *et al.*, reported *Tcf3* as over-expressed in the mTECs population. One possible explanation for not detecting *Tcf3* in the PPC lines is that it may be expressed in a multitude of mTEC lineages, including the NPP cell lines. Of further interest to note is that the PPC cell lines over-express a transcription factor belonging to the same family as *Tcf3* – *Tcf21* (table 1) which may be capable of binding with *Hand2* as *Tcf3* does. It is also important to point out that since transcription factors tend to be expressed at low levels the actual transcription factors allowing for TRA expression may not have been detected by microarray analysis. Further studies on the promoters of the over-expressed
genes and the expression of transcription factors both below and above the limit of microarray detection will be required to substantiate such suggestions.

The work presented in this paper has a number of limitations that require mention. The first is that the cell lines being used are SV40 transformed mTEC cell lines that underwent 20 or so passages before microarray analysis was performed. Just as we lost G8.8 and Aire expression in the few months following isolation and culturing we may have lost the expression of a number of important factors that play a role in mTEC TRA expression and function. That being said, the fact that proinsulin is still expressed suggests that other than Aire loss such changes are hopefully minimal. A second limitation of the work is that differential expression due to sex may be playing a role in the over-expression of a number of genes. As mentioned in the results, Xist is expressed by the NPP cell lines but not the PPC cell lines. Thus the PPC cell lines are male while the NPP cell lines are female. In the tissue culture setting, sex differences due to the in vivo hormonal environment should be absent. However, hormonal imprinting and differences due to Y chromosome genes or genes escaping X inactivation could potentially play a role in modifying expression characteristics of cultured cells. We do not believe this to be the case, however, since we do not seem to find over-expression of predominantly sex hormone associated genes. We are at this time, however, performing microarray analysis on cultured female and male mTEC cell lines to determine if a sex-linked expression pattern phenomenon is present.
5.6 References


6.0 CHAPTER SIX: Class III INS VNTR alleles associated with silencing of thymic insulin predispose to type 1 diabetes.

Petros Vafiadis, Houria Ounissi-Benkalha, Michael Palumbo, Rosemarie Grabs, Marylène Rousseau, Cynthia G. Goodyer, Constantin Polychronakos

The Journal of Clinical Endocrinology & Metabolism Vol. 86, No. 8 3705-3710, 2001
Copyright 2001, The Endocrine Society.

6.01 Contribution of Authors

Petros Vafiadis: Experimental design and strategy, Variable Number of Tandem Repeat (VNTR) cloning.

Houria Ounissi-Benkalha: VNTR cloning and VNTR genotyping, strategy development for VNTR sequencing.

Michael Palumbo: VNTR sequencing, C++ software development for annotating VNTR sequences.

Rosemarie Grabs: VNTR cloning and VNTR genotyping.

Marylène Rousseau: VNTR cloning and VNTR genotyping.

Cynthia G. Goodyer: Hypothesis development, manuscript preparation.

Dr. Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

6.02 Objective

Having identified two class III VNTR (18,19) alleles that were found to be in cis with silenced thymic INS genes, our group set-out to determine if such class III VNTRs predisposed to Type 1 diabetes. In addition, sequencing of one of these “silencing” VNTR alleles and one of the more commonly found INS enhancing VNTRs alleles was carried out.
6.1 Abstract

Type 1 diabetes results from autoimmune destruction of the insulin-producing pancreatic β-cells. The insulin gene (INS) is also expressed in human thymus, an ectopic expression site likely involved in immune tolerance. The IDDM2 diabetes susceptibility locus maps to a minisatellite composed of a variable number of tandem repeats (VNTR) situated 0.5 kb upstream of INS. Chromosomes carrying the protective long INS VNTR alleles (class III) produce higher levels of thymic INS mRNA than those with the predisposing, short class I alleles. However, complete silencing of thymic INS transcripts from the class III chromosome was found in a small proportion of heterozygous human thymus samples. We hypothesized that the specific class III alleles found on these chromosomes silence rather than enhance thymic insulin expression. To test the prediction that these alleles are predisposing, we developed a DNA fingerprinting method for detecting two putative "silencing" alleles found in two thymus samples (S1, S2). In a set of 287 diabetic children and their parents we found 13 alleles matching the fingerprint of the S1 or S2 alleles. Of 18 possible transmissions, 12 of the S1-S2 alleles were transmitted to the diabetic offspring, a frequency of 0.67, significantly higher than the 0.38 seen in the remaining 142 class III alleles; p=0.025. This confirms our prediction and represents an additional level of correlation between thymic insulin and diabetes susceptibility, which supports a thymic enhancer effect of the INS VNTR as the mechanism of IDDM2 and refines the contribution of IDDM2 genotyping to diabetes risk assessment.
6.2 Introduction

Insulin has an important place among the antigens involved in the autoimmune process that results in type 1 diabetes. Of the known autoantigens, insulin has the highest expression specificity for pancreatic β-cells, and it is the only antigen whose gene has been mapped to a genetic susceptibility locus. Anti-insulin antibodies are most often the first to appear in prediabetic individuals (1), and, in rodent models, autoimmune diabetes can be prevented by peripheral tolerance induction to insulin (2,3).

Central T-cell tolerance is determined in the thymus, where cells with autoreactive TCR rearrangements are deleted. The discovery that the thymus expresses small amounts of insulin (and other proteins with tissue-restricted expression) has generated interest in the role of central mechanisms in immune tolerance to these autoantigens (reviewed in Ref. 4). Direct evidence for this comes from studies in which thymus grafts can transfer tolerance to allo-(5) or xeno-(6) antigens transgenically expressed under the insulin promoter to non-transgenic syngeneic recipients. Indirect evidence for the importance of this mechanism in the pathogenesis of type 1 diabetes comes from studies of the IDDM2 locus of type 1 diabetes susceptibility.

IDDM2 maps 0.5 kb upstream of the insulin gene (INS), to a repeat polymorphism which consists of a variable number of tandem repeats (VNTR) of the consensus sequence ACAGGGGTGTGGGG (7,8). The number of repeats ranges from 30 to >150 in Caucasian chromosomes and this, combined with slight sequence variations in the repeat unit (Figure 4) results in a complex, hypervariable allele system. About 80% of Caucasoid alleles are in the range of 30-44 repeats (class I) and virtually all of the rest are longer than 110 repeats (class III). Intermediate lengths (class II) are rare.
The presence of at least one class III allele is associated with a 3-5 fold reduction in the risk of type 1 diabetes compared to the common I/I homozygote genotype (9-15). This dominant protective effect of class III alleles has been precisely located to the INS VNTR itself (9,12,15) which does not alter protein sequence, and therefore must exercise its biological effect through transcriptional regulation in cis. In adult (9) and fetal (16) human pancreas class III alleles are associated with 15-20% lower INS mRNA. By contrast in the thymus, where insulin is also expressed (17-19) class III is associated with substantially higher INS mRNA levels than class I, a gain of function much more likely to explain their dominant protective effect. Thus, in 17 of the 22 thymus samples heterozygous for class I/III INS VNTR alleles originally studied by us (18) and Pugliese (19), insulin mRNA was more than 2-fold higher from the class III-containing chromosome than from the class I. This suggested a mechanism for the protective effect of the class III alleles involving more efficient deletion of insulin-autoreactive T-cells, a process known to be dose-dependent (20-22). In the remaining 5 samples in the two studies, mRNA from the INS copy linked with the class III allele was completely absent.

The cause of this monoallelic silencing is at present not known but it may be due to genomic imprinting, the property of certain genes to be expressed only from the copy inherited from a parent of a given sex. Human INS and mouse Ins2 are located in syntenic domains that contain several genes that are imprinted in both species. Ins2 expresses both copies in pancreas, but has exclusive paternal expression in mouse yolk sac (23). We hypothesized that INS may be silenced in thymus by parental imprinting, but only on chromosomes carrying specific alleles within class III (allele-restricted imprinting). Alternatively, some class III alleles could act as thymus-specific silencing elements.
regardless of parental origin. Either version of this hypothesis predicts that those specific alleles predispose to diabetes through decreased thymic insulin expression, in contrast to class III as a whole which is protective.

Direct testing of the hypothesis that the silencing is due to the VNTR allele in cis is a major aim of our laboratory but is not simple. We are accumulating more thymus samples but the alleles are infrequent, and it will require a large number to reach statistical proof. In vitro testing with reporter constructs is hampered by the fact that the effect is thymus-specific, and insulin-expressing cells in this organ are rare, incompletely characterized (4), and no cell lines exist. While the effort for direct proof is underway, we decided to address the hypothesis through one testable major prediction it makes: if these specific alleles silence insulin in the thymus, they must be predisposing to diabetes.

To pursue this, we first precisely defined the class III alleles associated with INS silencing in two of our thymus samples by size and restriction fingerprinting. We then examined the transmission frequency of class III alleles indistinguishable from these “silencing” alleles (which we call S1 and S2) from heterozygous parents to diabetic offspring. We predicted that S-type alleles would be transmitted to diabetic offspring more often than the rest of class III.
6.3 Research Design and Methods

Sample sources and preparation

DNA was extracted from blood samples collected from 167 patients who either attended the Montreal Children's Hospital diabetes clinic, or participated in the Minneapolis branch of a multicenter study of the natural history of diabetic nephropathy in type 1 diabetes. Samples were obtained with signed, informed consent and approval by the Institutional Review Board of the respective institutions. In addition, we examined 120 DNA samples from 60 diabetic sibling pairs from the Human Biological Data Interchange. DNA from both parents was available in all cases. All patients developed insulin-dependent diabetes under the age of 19, except for two HBDI patients who were 29 and 34 at onset but were included because the other sibling in the pair had a young onset. All patients were of mixed European background.

Human fetal thymus tissues were obtained at the time of pregnancy termination with written consent from the mother, approved by the Institutional Review board of the Maisonneuve-Rosemont Hospital. The tissue was flash-frozen and pulverized under liquid nitrogen to extract DNA and RNA. INS mRNA from class I vs. III chromosomes was quantitatively differentiated by RT-PCR using a transcribed PstI polymorphism in tight linkage disequilibrium with the VNTR, as we previously described (16,18).

PCR protocol for all classes of INS VNTR alleles

The PCR reaction for amplification of all classes of INS VNTR alleles contained approximately 100-200 ng of DNA, 0.2 mM of each dNTP, 1 μCi of P32-dCTP, 1 mM MgCl2, Ammonium PCR reaction buffer, 0.4 U ID-Zyme thermostable polymerase and
100 ng of sense and antisense primer. ID-Zyme contains a proprietary mixture of high-fidelity polymerases. We used the primers described by Bennett et al. (9), as well as an additional pair of internal primers that could amplify both genomic and cloned DNA:

Sense 5'-GGCATCTTGGCATCCGGGACTG-3'
Anti-sense 5'-GCAGGGCGGGCTCTTTGCGCTG-3'

The PCR was carried out for 25 cycles of: 94°C/30 sec denaturing, 62°C/30 sec annealing and 70°C/3 min 30 sec with a 4 sec extension per cycle extension step. Products, internally labeled with $^{32}$P-dCTP, were visualized by 8% polyacrylamide gel electrophoresis and autoradiography. To distinguish small differences in size, PCR product from the S1 and S2 alleles were loaded in every 2 to 4 lanes, interspersed throughout the unknown samples (Fig. 3). All class III alleles analyzed were either larger than both S1 and S2, smaller than S1 and S2 or equal in size to either S1 or S2. Thus, the S1 allele is only slightly larger than the S2 allele, and there are no allele sizes between them that can be resolved by our method. Furthermore, each class III allele was loaded into two separate wells of the polyacrylamide gel for comparison to S1 and S2 separately. Those that were equal in size to S1 were always larger than the S2 allele, and those that were equal in size to S2 were always smaller than S1 allele. This indicates that the same alleles behave in a consistent manner between separate loadings in PAGE and that our method can consistently identify small differences in migration distance.
Cloning and sequencing of INS VNTR alleles

S1 and several non-S alleles were directionally cloned by double digestion of the PCR product with NcoI and PstI. Cloned and genomic DNA template gave identical results and were used interchangeably for PCR amplification. The identity and integrity of the cloned S1 allele was repeatedly demonstrated by showing electrophoretic co-migration of PCR amplification products (n>15). Similarly, both cloned and genomic DNA demonstrated the same RFLP band pattern after MspI digestion (MspI recognizes an uncommon variant of the repeat unit, ref. 8,12), suggesting the same arrangement of internal repeat sequences.

The highly repetitive nature of the VNTR precludes the use of restriction subcloning or internal primers for sequencing. Therefore, the S1 allele and a class III allele associated with enhanced thymic insulin expression (E1), were sequenced by generating a series of overlapping unidirectional deletions from a NcoI-PstI fragment subcloned into the pGEM-T® vector using the Exo-Size® deletion kit, (New England BioLabs). Exonuclease III digestion of the construct linearized by double digestion with SphI and NcoI, both on the 5' end of the insert, shortens the insert but not the vector, which is protected by the 3' overhang left by SphI.

DNA fingerprinting of class III alleles

 Alleles of the same size could be further distinguished using two highly polymorphic MspI fragments (700-800 bp and 400-500 bp) which, in class I/III individuals, are always located above the largest class I allele digestion fragment. This was determined by comparing the MspI digestion fragments in a number of class I/III individuals obtained from PCR product where only the class I allele is amplified (using
the previously described PCR conditions) (9,18) and product from our PCR protocol which amplifies both class I and III alleles.
6.4 Results

PCR amplification of all INS VNTR classes

Our protocol was successful in amplifying all INS-VNTR classes (Fig. 1), an important technical advance in the study of the IDDM2 susceptibility locus. Due to the high GC-rich content and repetitive nature of this sequence, resulting in a highly stable intramolecularly folded structure (24-26), there have been, to date, no published reports of successful PCR amplification of the long class III alleles. Co-dominant segregation of class III alleles within families confirmed the high fidelity of the method and the stability of these alleles within families. It is illustrated in Fig. 2.

Characterization of “silencing” class III INS VNTR alleles

First, in DNA from the two thymus samples that showed monoallelic expression, we sequenced the non-repetitive part of the INS promoter/enhancer region from just downstream of the VNTR to the first INS exon. No unknown sequence variant was found that could account for the silencing of one allele in this region. However, the possibility remains that S1 or S2 may be merely markers for a polymorphic silencing element outside the promoter region we sequenced. We then proceeded to define the VNTR in these two chromosomes where INS was apparently silenced.

Class III alleles ranged in size from 1.8 kb to 2.5 kb. The S1 and S2 silencing alleles were in the middle of the range, at around 150 repeats (2.1 kb),
Figure 1: Amplification of all three classes of INS VNTR alleles by a single PCR reaction in heterozygous DNA samples. S1 and S2, DNA from the two thymus samples where insulin expression from the class III chromosome was silenced; E, thymus sample with enhanced thymic expression from the class III chromosome. Smaller alleles amplify more strongly. The length of the repetitive part was calculated by subtracting 668 bp of nonrepetitive flank sequence from that estimated for the PCR product.
Figure 2: Illustration of codominant segregation (a and b) and stable transmission (c and d) of INS VNTR alleles in families. a and b show undigested amplification products run under conditions that show both class I and class III alleles (m, Mother; f, father; c, child). The mother in family a is heterozygous for two different class III alleles (finer size discrimination between class III alleles is not shown here because it requires longer electrophoresis in which the class I allele is lost). c and d show the two major bands of the MspI digest used to discriminate between class III alleles of the same size, as in Fig. 3. DNA from the original S1 and S2 samples is run as standard. u, DNA sample from unrelated individual.
and probably differing from each other by a single repeat. S1 was distinguishably longer
than S2, without observable alleles between the two.

Class III alleles identical in size but different in terms of composition in variants
of the repeat unit were further distinguished by restriction fingerprinting (Figure 3). The
PCR product was digested with MspI, whose recognition site (5'-CCGG-3') is present
only in certain of the less common variants. Alleles of identical size often had fingerprints
that could be distinguished using the two bands that could always be seen above the class
I allele bands. Most MspI digests of class III alleles had two major bands clustered around
700-800 and 400-500 bp, plus a number of smaller or less common bands. The exception
were class III alleles that were part of the unusual haplotype termed VPH by Bennett et al.
(9), which were richer in MspI sites and gave no band over 350 bp, confirming their
derivation from a different ancestral chromosome. These are the same alleles recently
termed class III.b by Stead and Jeffreys (27), whose rich content in MspI-containing
repeats explains the absence of visible long bands in the restriction digest.

Transmission analysis of “silencing” alleles in type 1 diabetes

Using both size and MspI fingerprint, we determined that 7 of the 908 parental
chromosomes in families with at least one type 1 diabetes-affected child carried alleles
indistinguishable from S1 and 6 from S2. Among 31 thymus samples showing enhanced
expression from the class III chromosome (ref 18 and
Figure 3: Digestion of the PCR product with MspI gave two major polymorphic bands for most alleles. The two major bands in S1, predicted from the sequence, are 427 and 736 bp. S2 has an identical lower band, whereas the upper band appears to be shorter by one repeat unit. Known S1 and S2 samples (marked) were frequently interspersed among unknown samples to adjust for gel drift. Lanes containing unknown samples are indicated with a U. An arrow under the U signifies that the fingerprint is identical to either S1 (left) or S2 (right). An example of a III/III genotype is shown. Cl, PCR amplification from cloned DNA.
an additional 21 informative fetal thymus samples tested subsequently), 30 were distinct from S1-S2 by size alone, while only one allele associated with enhancing (termed E1) had identical size and fingerprint to S1. The two alleles were sequenced and found to differ in four out of 150 repeats (Figure 4), none of which affected the MspI restriction pattern. In any event, alleles identical to S1 or S2 by both size and fingerprint but which are not associated with INS silencing are too infrequent (1/33) to interfere with a statistical testing of the hypothesis that S1 and S2 predispose to diabetes.

We proceeded to test our hypothesis by genotyping the diabetic children of all parents who had a class III allele. There were no S/S homozygous genotypes, therefore all parents who had a “silencing” (S) allele were heterozygous for one S and one “non-silencing” INS VNTR allele (either class I or non-S class III). Alleles present in the child were counted as transmitted, and those absent as non-transmitted. In families with more than one diabetic child, each opportunity for transmission was counted separately in this way.

The results of the transmission analysis are summarized in Table I. Non-S class III alleles were transmitted at a frequency significantly less than 0.5, as expected from the known dominant protective effect of class III as a whole. As our hypothesis had predicted, S alleles behaved as predisposing rather than protective: they were much more frequently transmitted than all other class III alleles (p=0.025 by the Fischer exact test). Since in most cases the transmitting parent's other allele was a class I, S-type alleles appear to behave as more
Sequence comparison of E1 and S1 alleles

S1  cniabaabababaacaaaaacaaacaaaacaacaacaaaaakabacaabakabaacaa
E1  cniabaabababaacaaaaacaaaacaacaacaaaaakabacaabakabaacaa

S1  abakababaacaacaaakaakabaabaabaacabaeacbaacaaacha
E1  abakababaacaacaaakaakabaabaabaacabaeacbaacaaacha

S1  aaacaaaaeaacaacaadaaadaaacaacghcccb
E1  aaacaaaaeaacaacaadaaadaaacaacghcccb

| a | ACAGGGGTGTGGGG | j | ACAGGGGTGTGAGG |
| b | ACAGGGGTCTGGGG | k | ACAGGGGTCCGGGGG |
| c | ACAGGGGTCCGGGG | l | ACAGGGGTCCGGGT |
| d | ACAGGGGTCCGGGG | m | ACAGGGGTCCGG |
| e | ATAGGGGTCTGGGG | n | ACAGGGGTCCGGGG |
| f | ATAGGGGTCTGGGG | o | ACAGGGGTCTGGGG |
| g | ATAGGGGTCTGGGG | p | ATAGGGGTCTGGGG |
| h | ACAGGGGTCTGGGG | q | ACAGGGGTCTGGGG |
| i | ACAGGGGTCTGGGG | r | ACAGGGGTCTGGGG |

Figure 4: Sequence of two class III alleles that have the same size and the same MspI fingerprint. INS expression from the class III chromosome in the thymus from which E1 was isolated was enhanced (as is the case with most class III alleles), whereas S1 was associated with monoallelic INS silencing. Differences are highlighted. The sequence is given in variants of the repeat unit, as defined by Bennett and Todd (29). The nucleotide sequence of each variant is given. Variants n–q (bold) have not been previously reported. All sequence was confirmed on at least two overlapping clones.
Table 1: Summary of transmissions from nondiabetic parents to a diabetic child. We examined 287 diabetic children and 227 pairs of parents (908 parental chromosomes). In the 60 HBDI families that had two children transmissions were calculated twice, once for each child. Therefore, the total number of transmissions was higher than the number of parental chromosomes.

<table>
<thead>
<tr>
<th>Transmissions involving:</th>
<th>Total Transmissions</th>
<th>S1 or S2</th>
<th>All other class III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransmitted</td>
<td></td>
<td>6</td>
<td>99</td>
</tr>
<tr>
<td>Transmitted</td>
<td></td>
<td>12</td>
<td>62</td>
</tr>
<tr>
<td>Total</td>
<td>1148</td>
<td>18</td>
<td>161</td>
</tr>
</tbody>
</table>

\[ p=0.025 \]
predisposing than even class I alleles, as our hypothesis would predict based on thymic insulin expression levels. Due to the small sample size, we were not able to evaluate over-transmission of S-type alleles in terms of parental origin, and the question of imprinting vs. parent-of-origin independent silencing remains open.

6.5 Discussion

Our previous findings have led to the hypothesis that the expression of insulin (or antigenic epitopes thereof) in the thymus is important in insulin-specific T-cell tolerance, and that the effect of the IDDM2 locus is due to allele-dependence of the levels of such expression: higher expression would better induce immune tolerance to INS-encoded antigens, resulting in a decreased probability of an autoimmune response against the pancreatic β-cells. The dose-responsiveness of thymic T-cell precursor fate is supported by studies in vitro that have demonstrated positive selection of thymocytes at low concentrations of a selecting peptide in fetal thymus organ cultures, while increasing concentrations of the selecting peptide (20,21) or peptide-MHC complexes (22) resulted in decreased positive selection and induction of negative selection.

A direct corollary of this hypothesis is that the rare individuals who silence INS on the class III-bearing chromosome in the thymus would be predisposed to diabetes. If this silencing depends on the cis-effect of specific alleles within class III, either through the direct transcriptional effect of S1 or S2, or because these alleles are markers for a sequence variant outside the regions sequenced, those alleles will be predisposing. Alternatively, silencing could simply require the presence of any class III allele, the difference between silencing and enhancing being determined by genetic influences in
trans or environmental factors. Our data clearly support the model of a specific genetic effect in cis by confirming its one key prediction. This finding justifies the major investment in effort needed to directly prove the effect of these alleles in thymic insulin-producing cells.

Our characterization of two of these INS silencing-associated class III alleles by PCR amplification and restriction fingerprinting was an important technical advance that allows distinguishing a large number of different alleles within class III. Recently, the diversity of INS-VNTR alleles was demonstrated with MVR-PCR by Stead and Jeffreys (27). This method defines the position of the six most common repeat variants within a given allele, and has demonstrated the presence of 89 distinct patterns in 171 class-III bearing chromosomes. In light of these data it is unlikely that all alleles identified by our restriction fingerprinting as either S1 or S2 represent the exact same DNA sequence. It seems more likely that these patterns are markers for some common structural feature that confers the functional property of INS silencing in the thymus. To explain the observed phenomenon, this structure-function relationship must be subtle, as illustrated by a comparison of the two alleles we sequenced, Eland 81 and S1. In the thymus samples from which they were obtained, these alleles are associated with, respectively, enhanced or silenced INS expression, yet they only differ in 4 out of the 150 repeat units.

A meaningful discussion of how these subtle differences could affect function would require understanding of the mechanism by which individual INS VNTR alleles could exercise transcriptional effects on INS. Such understanding is currently sketchy. The INS VNTR has a complex secondary structure (24-26) that may interact with
transcriptional complexes and whose precise nature may be influenced by relatively subtle changes in size and the exact repeat composition.

Whatever the mechanism of polymorphic INS silencing is, it may involve parental imprinting. Ins2 is imprinted in mouse yolk sack (23) but in both mouse and human pancreas, INS has demonstrated biallelic expression (9,13,16,23). Polymorphic parental imprinting of INS as a cause of monoallelic silencing in the thymus could not be evaluated as parental DNA was not available in any of the five monoallelic samples. If imprinting is the mechanism involved in silencing of thymic INS expression, then it likely requires the presence of specific class III alleles in cis (allele-restricted imprinting). Therefore, silencing-type alleles may only be predisposing when transmitted from a parent of a specific sex. This is a focus of on-going work in our laboratory.

Functional heterogeneity among alleles within class III has three important implications: a) The additional concordance between diabetes susceptibility and thymic INS expression lends further support to the latter as the biological mechanism involved in IDDM2, b) it supports the INS VNTR as the disease locus in IDDM2, a concept not yet proven beyond all doubt (28), and c) it enhances the contribution of IDDM2 to the predictive power of DNA testing for diabetes risk.
6.6 Acknowledgments

Supported by the Juvenile Diabetes Foundation International and the Quebec Diabetes Association (QDA). P.V. was supported by a Doctoral Research Award from the Medical Research Council of Canada, and by the QDA. Many thanks to Tracey Hollingdrake, Danielle Drouin, Tami Dalzel, M. Mauer, and Jaqueline Dufresne for patient recruitment and sample collection. We also thank George Eisenbarth and Sunanda Babu for making the HBDI samples available to us.
6.7 Abbreviations

INS  human insulin gene
Ins2  mouse insulin gene
VNTR variable number of tandem repeats
RFLP restriction fragment length polymorphism
PCR  polymerase chain reaction
S1/S2 silencing-type allele 1 or 2
E1   enhancing-type allele 1
VPH  very protective haplotype
6.8 References


7.0 CHAPTER SEVEN: Discussion and Conclusion

The work described herein provides novel and important information regarding the expression of proinsulin and other TRAs within the thymus. At the outset of the work, there existed a major controversy as to which cells within the thymus produced the TRAs, including proinsulin. A great deal of immunohistochemical work, using antibodies for proinsulin, GAD, and I-A/2 implicated dendritic cells [177] whereas purification of thymic cell types followed by RT-PCR analysis for proinsulin and other TRAs implicated medullary epithelial cells [158]. In chapter 3, we describe the unambiguous identification of the proinsulin producing cells as thymic medullary epithelial cells using transgenic mice expressing a functional β-Galactosidase gene under the regulation of the Ins2 promoter. We demonstrate that anywhere between 1-3% of all mTECs produce proinsulin and we also go on to localize the proinsulin producing cells within the Hassall’s Corpuscle-like structures in mice suggesting a possible function for these conserved thymic formations.

It should be noted that our results reported in chapter 3 have for the most part been reported/reproduced by other groups as well, but the localization of the TRA expressing mTECs within Hassall’s Corpuscles had not been previously demonstrated. Two groups did reportedly localize the expression of peripheral antigens in mTECs of Hassall’s Corpuscles of human thymus but these molecules cannot be exclusively classified as TRAs [171, 172]. Additional, indirect evidence for the involvement of Hassall’s Corpuscles comes from studies demonstrating marked T-cell apoptosis in and around the Corpuscles [178] and findings demonstrating that dendritic cells in the
immediate vicinity of Hassall’s Corpuscles take antigen up from the surrounding milieu and present the antigen to CD4+ cells which seems to be important for Regulatory T cell formation [40] – note that the close proximity of the dendritic cells to Hassall’s Corpuscles is suggestive of the production of antigen within the surrounding milieu by the mTECs of the Hassall’s Corpuscles. Other groups, however, have demonstrated, via antibody labeling [179] or in situ hybridization [180], that the majority of the TRA producing mTECs are located within the cortico-medullary junction and not specifically within Hassall’s Corpuscles. Such differences in observations are difficult to reconcile, but it should be pointed out that all the methods used thus far (i.e. immunohistochemistry with X-gal staining, antibody labeling or in situ hybridization) have important limitations. That is to say, all of these methodologies, even when the work is done quite rigorously, can often result in considerable background activity especially when assaying for such a small number of positive cells and when long incubation times are necessary due to the low level expression of Proinsulin/β-Galactosidase. In fact, at the outset of our studies, it was initially desired not to use antibody labeling, in situ hybridization or transgenic mice requiring the use of an enzymatic assay to identify the cells because of the above described limitations. Instead, our group attempted at engineering transgenic mice expressing Enhanced Green Fluorescent Protein (EGFP) under the regulation of the Rat Insulin II Promoter and it was deemed possible to use fluorescence microscopy to identify the proinsulin producing cells and their location within the thymic medulla. Unfortunately, after numerous attempts at engineering the mice, using in vitro tested constructs, expression of EGFP, by RT-PCR, in any of 11 transgenic founder mice was not detected. As mentioned previously, a possible explanation for this was provided in a
personal communication from Dr. Christian Sirard as he had been made aware by others that EGFP (but not GFP) is often silenced in a variety of cell types for an as of yet unknown reason. As a result, we decided to proceed with the experiments using the LacZ expressing mice.

The work described in Chapter 4, the isolation of two proinsulin producing thymic medullary epithelial and two non-proinsulin producing cell lines (PPC and NPP, respectively), is novel work that to our knowledge has not been reported by any other groups to date. In addition to allowing us to perform microarray analysis on lineage specific mTECs (chapter 5), as opposed to the presently more commonly performed microarray analysis on mixed populations of mTECs, isolation of these cell lines will provide our group with the ability to study the regulation of proinsulin transcription and translation. In addition to this, inducing expression of the transcription factor Autoimmune Regulator (Aire) in these cell lines will allow us to study the specific mechanism by which Aire augments proinsulin expression as well as other TRAs. Finally, such cell lines can be used to study VNTR function - more specifically the VNTR associated allele-specific mechanisms allowing for enhanced (class III VNTRs), reduced (class I VNTRs) or completely abrogated (class III proinsulin silencing type VNTRs) proinsulin expression (chapter 6).

Microarray analysis of the two proinsulin producing cell lines versus the two non-proinsulin producing cell lines (chapter 5), in the absence of Aire (as Aire expression was lost in culture), demonstrated the over-expression (>4 fold or more) of approximately 50 genes in the PPC cell lines (both lines) and approximately half of these genes can be described as TRAs. Coincidently, the NPP cell lines were likely clonal in origin and as
such shared a very similar expression profile with each other. The NPP cell lines overexpressed, in comparison with the PPC cell lines, approximately 40 genes again half of which can be described as TRAs. Interestingly, the TRAs expressed by either the NPP or PPC cell lines were representative of virtually every tissue in the body (including embryonic and reproductive) with some tissues represented more than once in both cell lines. Chromosomal localization of the over-expressed genes in either the PPC or NPP cell lines did not reveal chromosomal clustering (ie. two genes localized within 200kB of each other), a phenomenon described when microarray analysis of mixed mTECs subsets was undertaken [170, 174]. It is presently hypothesized that chromosomal clustering is the result of chromosomal alterations that allows for promiscuous TRA expression.

A number of important findings were obtained from the results of the microarray analysis. Previous to this work, it was unknown how many self-antigens are expressed by an individual mTEC. Indeed, much work has shown that a number of antigens, including proinsulin, are expressed only by 1-3% of all mTECs but it was not established if the same mTECs were expressing all these antigens. The work described demonstrates that approximately 20-25 TRAs (in the absence of Aire), are likely expressed in each of the G8.8+/UEA-1+ lineages and as Aire expression seems to affect approximately 12% of total TRAs known to be expressed in the thymus [170], we could estimate that each of these mTEC lineages expresses approximately 25-30 TRAs in vivo. Interestingly, seeing as that it is estimated that greater than 1000 TRAs are expressed in the thymic medulla [170], this could require anywhere between 30-50 mTEC lineages which closely agrees with the findings that 1-3% of all mTECs are proinsulin producing mTECs. As such, our results suggest that each mTEC lineage expresses approximately 25-30 TRAs.
A second important finding from the microarray analysis is that antigens expressed within individual mTEC lineages are unrelated by tissue type. Only one pancreatic TRA (proinsulin) was found in the PPC cell lines while crucial pancreatic proinsulin regulatory molecules, such as Pdx-1, Glut-2 and GCK [23-25] were absent in this lineage by both microarray and RT-PCR although they are present in whole thymus—possibly expressed as TRAs in other mTEC lineages. It is noteworthy to point out that because of the limited sensitivity of the microarrays, future work should include RT-PCR for other pancreatic specific TRAs that are strongly suspected in the development of Type 1 Diabetes. Importantly, the lack of expression of Pdx, Glut-2 and GCK suggest that the mechanisms directing the expression and regulation of TRAs, more specifically proinsulin, in mTECs is not similar to those operating in peripheral cells, and that the mTECs, as opposed to what the “mosaic model or patch quilt model” theory [170, 181] proposes, are not simply emulating the expression patterns of peripheral cell types.

A final important observation from the microarray analysis are that the over-expressed genes, be it TRAs or other genes, do not seem to be localized in close proximity to each other as the chromosomal clustering hypothesis suggests, questioning the importance of the clustering phenomenon in allowing for TRA expression. With regards to TRA expression, it should be mentioned that a number of transcription factors are differentially expressed in the PPC vs. the NPP cell lines but whether or not they are expressed as TRAs or are expressed for functional reasons (ie. directing TRA expression) has yet to be determined. With regards to proinsulin expression, we performed a preliminary analysis of the Ins1 and Ins2 promoter regions which revealed two binding sites for the heterodimer transcription factors Hand2 (heart and neural crest derivatives
expressed transcript) and Tcf3 (transcription factor E2a) – two basic helix-loop-helix transcription factors that must heterodimerize together to bind DNA [182, 183]. Based on the microarray results, only Hand2 is over-expressed in the PPC cell lines but Tcf3 has been previously shown to be expressed in mTECs[170]. One possible explanation for not detecting Tcf3 in the PPC lines is that it may be expressed in a wide range of mTEC lineages, including the NPP cell lines or its expression may be dependant on Aire. Of further interest to note is that the PPC cell lines over-express a transcription factor belonging to the same family as Tcf3 – Tcf21 which may be capable of binding with Hand2 as Tcf3 does. It is also important to point out that transcription factors tend to be expressed at low levels and that a number of these factors important for TRA expression may very well be below the limit of detection for microarray studies. Further studies on the promoters of the over-expressed genes and the expression of transcription factors expressed in the PPC and NPP cell lines will be necessary to answer these questions.

Of course, the work described in chapters 4 and 5 has important limitations which require mentioning. As is the case when cells are cultured in vitro, cell phenotype and expression patterns can potentially change. Indeed, culturing of our mTEC cell lines resulted in the loss of G8.8+ and Aire expression, both of which were detectable upon initial isolation of the cell lines. Although we can potentially compensate for Aire loss by inducing Aire expression and re-performing the microarrays in the presence of Aire (which we are doing at this time), there may be a number of other factors lost that we cannot account for that can potentially prevent investigators from solving the puzzle.

Another potential limitation of the microarray work, as mentioned in chapter 5, is that the proinsulin positive cell lines were determined as being derived from male mice
and the two negative cell lines from female mice. *In vivo*, biological differences between sexes are mostly determined by different gonadal steroids (androgens vs. estrogens), which are likely irrelevant in lines that have been cultured in the same media for many cell generations. However, some gonadal steroid effects persist long after hormone exposure (hormonal imprinting) and sex differences may also be determined by the (very few) genes that escape X-inactivation. To compensate for such possible confounders, we are currently performing microarray analysis on cultured mTEC cell types from female and male mice and we believe there will be only minor associations with sex type.

As a final point, it is worthwhile to mention that should any of this work be redone in the future, in particular, determining the location of the proinsulin producing cells in the thymic medulla and microarray analysis of the proinsulin producing mTEC lineage, usage of Timer mice [184] (transgenic mice that have proinsulin II tagged with a live-cell fluorescent reporter protein termed Timer) along with the novel technology of performing microarray analysis with 100 ng as opposed to 10 μg of RNA would greatly facilitate the work while at the same reduce the limitations of the results. The Timer mice have been shown to express the fluorescent protein in the pancreatic β-cells. Confirmation of thymic expression and identification of the fluorescent cells within the thymic medulla would allow for fluorescence automated cell sorting of primary fluorescent cells (proinsulin producing cells) immediately extracted from murine thymus.
Appendix: Ethics Approvals and Copyrights
This is a License Agreement between Michael O Palumbo ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

License Number 1662301123570
License date Mar 05, 2007
Licensed content publisher Nature Publishing Group
Licensed content publication Nature Genetics
Licensed content title Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus
Volume number 9
Issue number 3
Pages 284-292
Year of publication 1995
Portion used Figures
Number of figures 1
Requestor type Student
Type of Use Thesis / Dissertation
Total $0.00
Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.
Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).

NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:

Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

For AOP papers, the credit line should read:

Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.JOURNAL.ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of up to a whole article do not require NPG approval. The translation should be credited as follows:

Translabeled by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

https://s100.copyright.com/App/PrintableLicenseFrame.jsp?licenseID=2007030_117307408... 3/5/2007
Additional Comments:

I will reprint the entire article in my dissertation.

The Endocrine Society grants permission to reproduce the entire selected article stated above in your dissertation contingent upon the following conditions: 1) That you give proper credit to the author(s) and to include in your citation, the title of journal, title of article, volume, issue number, date, and page numbers. 2) That you include the statement Copyright 2001, The Endocrine Society. Please understand that permission is granted for one-time use only. Permission must be requested separately for future editions, revisions, translations, derivative works, and promotional pieces.

Name: Karen Epps
Title: Journal Publications Coordinator
Date: January 2, 2007
PERMISSION LICENSE: PRINT REPUBLICATION

Request ID/Invoice Number: MIC46025

Date: January 04, 2007

To: Michael Palumbo
    McGill University
    298 Justin
    Laval QC H7P 6G7
    CANADA
    "Licensee"

McGraw-Hill Material

Author: Kasper
Title: HARRISON’S PRINCIPLES OF INTERNAL MEDICINE, 16 e © 2005
Description of material: Table 323-1 on page 2153

Fee: $WAIVED

Licensee Work:

Author: Michael Palumbo
Title: Identification, Isolation and Characterization of Preinsulin Producing Thymic Cells [thesis]
Publisher: McGill University
Publication Date: June 2006
Distribution Territory: USA
Languages: English

Permission for the use described above is granted under the following conditions:

1. The permission fee of $WAIVED must be received by The McGraw-Hill Companies on or before 1/4/2007, and MUST BE ACCOMPANIED BY A SIGNED COPY OF THIS AGREEMENT. A check should be made payable to The McGraw-Hill Companies, and sent to The Permissions Department, The McGraw-Hill Companies, Two Penn Plaza, NY, NY 10121-2298. Please include the invoice number indicated at the top of this form on your check.

2. No adaptations, deletions, or changes will be made in the material without the prior written consent of The McGraw-Hill Companies.

www.mheducation.com
blood samples from immature animals will not be necessary.

From strain 2 and strain 3 mice, pancreas, salivary glands and the thyroid will be extracted for histological analysis and thymus, spleen and bone marrow will also be extracted for isolation of T-cells and insulin-producing epithelial cells. These extractions will be carried out on 8 mice at the ages indicated, in triplicate. All mice will be CO2 euthanised prior to dissection (protocol UACC#13).

Strain 4 animals are prone to brain tumours and thymus tumours. Our research does not involved the study of these tumours, so we will keep only healthy animals. There is no test procedure to detect subclinical tumours. The first symptoms consist of decreased growth and decreased grooming. Animals will be inspected daily and immediately euthanized at the first appearance of symptoms suggesting tumour.

7. Endpoints

a) For B and C level of invasiveness.

The procedures are the same as the original protocol: YES ☒ NO ☐

IF NO, supply new endpoints that are different from the original protocol:

b) For D level of invasiveness.

Include here ALL endpoints, including the ones described in the original protocol as well as new and changed endpoints in CAPS:

Development of diabetes
Euthanize to remove tissues

8. Hazards: (check here if none are used: ☒)

a) Are the hazards different from original protocol? (infectious, radioactive, toxic, carcinogen, tumours)

YES ☒ NO ☐ if yes, supply details (material, risks, precautions):

b) Have the cell lines been tested for human and animal pathogens? YES: ☒ NO: ☐ None used: ☐

9. Description of Animals to be used in the coming year (only):

<table>
<thead>
<tr>
<th>Species</th>
<th>Sp/str 1</th>
<th>Sp/str 2</th>
<th>Sp/str 3</th>
<th>Sp/str 4</th>
<th>Sp/str 5</th>
<th>Sp/str 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier/Source:</td>
<td>mus/C57Bl/6</td>
<td>mus/NOD</td>
<td>mus/NOD</td>
<td>mus/C57Bl6</td>
<td>mus/C57Bl6/129</td>
<td>mus/129</td>
</tr>
<tr>
<td>Strain:</td>
<td>Charles River</td>
<td>Jackson labs</td>
<td>Bred locally or from Jackson Labs</td>
<td>Charles River</td>
<td>Bred locally obtained from Jackson Labs</td>
<td>generated at the RVH Alan Peterson AUP #4558</td>
</tr>
<tr>
<td>Sex</td>
<td>M+F</td>
<td>M+F</td>
<td>M+F</td>
<td>M+F</td>
<td>M+F</td>
<td>M+F</td>
</tr>
<tr>
<td>Ins1 knockout and Ins2 knockout</td>
<td>Immortomouse (heterozygous/homozygous)</td>
<td>Ins2 knockout and Ins1 knockout mice</td>
<td>human insulins/VNTR transgenics</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References:


