



UNIVERSITY OF
CAMBRIDGE



Somatic mutations in the pancreatic islets

Supervisor: Iñigo Martincorena



Pantelis Andreas Nicola

King's College

University of Cambridge

July 2018

This dissertation is submitted for the degree of
Master of Philosophy

Preface

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

It does not exceed the prescribed word limit prescribed by the Degree Committee for the Faculty of Biology. The word count is 19,956 exclusive of tables, footnotes, bibliography and appendices.

Abstract

The endocrine pancreas is composed of the islets of Langerhans. These micro-organs play a crucial role in glucose homeostasis by producing and regulating insulin and glucagon secretion. Dysfunction of these islets is part of the pathogenesis of diabetes mellitus. The global health burden of diabetes mellitus is growing, with an estimated one in eleven adults affected worldwide. A better understanding of the development and maintenance of the pancreatic islets could prove crucial to reversing this trend.

Whilst somatic mutations have been studied extensively in tumours, the exploration of normal tissue is still in its infancy. Here I present a novel workflow, using laser capture microdissection, whole-genome sequencing and innovative bioinformatics to accurately identify somatic mutations in healthy pancreatic islets. By sequencing 32 islets, all from a single individual, this work reveals islets to be polyclonal units formed by multiple embryonic founding lineages that often come to be dominated by one or two major lineages. The very low mutational burden observed here suggests these islets expand very early in embryogenesis and do not undergo large clonal expansions later in life. This is consistent with the islets being a slowly dividing tissue during adulthood, making it less likely that islet neogenesis or replenishment by a small number of progenitors, occurs in adults.

The 32 islets sequenced here all share a single common ancestor, one that also gives rise to the bladder urothelium. This is presumably the first cell that gave rise to all adult tissues. Spatially, nearby islets are more genetically similar than distant islets. This pattern demonstrates that different embryonic lineages contribute disproportionately to the islets in different areas of the pancreas, the implication being that this emerges during embryonic pancreatic development, or in adulthood through hypothetical islet fission events.

The translational potential of this line of work is substantial. Using this approach to understand the maintenance of islets in diabetes could yield a greater understanding of the pathogenesis, and whether somatic mutations could play a role in the disease. Applications to other normal tissues could similarly refine our knowledge of their development, maintenance and disease, with exciting prospects for clinical application.

Acknowledgements

This work could not have been possible without the kind generosity of the transplant donor, who so altruistically donated their tissues for research. The surgical team, led by Kourosh Saeb-Parsy at Addenbrooke's Hospital, Cambridge, collected the donor specimens and Yvette Hooks processed and sectioned them. The low-input sequencing pipeline was designed by Pete Ellis while the laser capture microdissection pipeline was developed by Luiza Moore, who also gave an expert histological opinion on all samples used in this project. Laura O'Neill was responsible for managing the sequencing submissions, which was undertaken by the Sanger Sequencing Core. The Cancer, Ageing and Somatic Mutation (CASM) IT department were central to the data analysis used here, as was Mathijs Sanders, who crafted the in-house variant filter used in this work.

CASM, and the Wellcome Sanger Institute, has been a stimulating environment full of friendly colleagues and I thank all those to whom I've spoken. Alex Cagan and Andrew Lawson have both been instrumental throughout this work and have shown themselves to not just be exceptional scientists, but natural-born mentors. Federico Abascal proved to be an encouraging teacher and I thank him for his help with the clustering analysis (section 4.10 and 8.4). Tim Coorens has been the backbone of this project and his computational input is omnipresent throughout this work, particularly in the design of the beta-binomial filter (section 3.10) and the production of the phylogenetic trees (section 3.15 and 4.13).

My thesis committee, made up of Peter Campbell and Kourosh Saeb-Parsy, provided critical insights for which I'm grateful. Additionally, I thank Peter Campbell for writing the clustering algorithm used here (section 3.15 and 8.1). My supervisor, Iñigo Martincorena, deserves the highest praise for taking me on with such passion and nurturing my development as a budding clinician-scientist. The lessons I've learnt from him, not just in research, will stay with me through the years.

Finally, I am humbled by the generous support that King's College, Cambridge and the Isaac Newton Trust have provided me with this year. To my family, friends and loved ones, you have all been unparalleled in your unconditional love and support and I dedicate this work to you all, with the hope that one day I can repay the sentiment.

Highlights

1. The whole genomes of 32 pancreatic islets were sequenced from a single human donor.
2. An unmatched analysis proved efficient in the removal of germline variants and artefacts, as well as accurately calling somatic mutations, including those occurring in early embryonic development.
3. The observed somatic mutation burden in the pancreatic islets is low and is driven by intrinsic mutational processes.
4. Almost all somatic mutations identified appeared to have no functional impact.
5. Islets are polyclonal units.
6. Pancreatic islets and bladder urothelium share the same most recent common ancestor.
7. Multiple embryonic founders establish each pancreatic islet but islets develop major and minor lineages.
8. Pancreatic islets do not appear to be maintained by a rapidly-dividing stem cell population, but whether there are multiple stem cell populations, or self-duplicating islet cells, needs further study.
9. The spatial distribution of islets and their embryonic lineages reveals their founding cells are non-randomly distributed.
10. There are potential applications of this work to the fields of tissue development, maintenance and disease. The possible role of somatic mutations in diabetes mellitus is a target of future research.

Contents

1. Introduction

1.1	Somatic mutations are acquired throughout life	9
1.2	The somatic mutational burden in normal tissues is comparable to some tumours	10
1.3	The mutational signatures in normal tissues and cancer give insight into mutational processes	11
1.4	The clonality of a tissue sample can be estimated using the somatic mutations present	12
1.5	Phylogenetic tree reconstruction of early embryogenesis has been demonstrated in clonal organoids	14
1.6	Phylogenetic reconstruction with subclonal tissues requires a framework to identify cell populations	16
1.7	The pancreatic islets perform endocrine functions	18
1.8	Primitive islets develop early in foetal development and numbers peak in the post-natal period	20
1.9	There are physiological, and pathological, causes for β -cell proliferation	21
1.10	The maintenance of the pancreatic islets, through adulthood, is unclear	22
1.11	Summary	24

2. Aims

2.1	Develop a robust workflow for analysing somatic mutations in normal tissue	25
2.2	Characterise the landscape of somatic mutations in the normal pancreatic islets	25
2.3	Elucidate the early phylogeny of the pancreatic islets	25
2.4	Gain insight into the maintenance of the pancreatic islets through life	27

3. Methods

3.1	Pancreatic specimens were obtained and prepared for dissection	28
3.2	Slides were stained with haematoxylin and eosin	28

3.3	Slides were imaged using the Leica LMD7 Microscope (Leica Microsystems GmbH, Wetzlar, Germany)	29
3.4	Laser capture microdissection was used to excise pancreatic islets	30
3.5	Excised tissue underwent protein digestion prior to whole-genome sequencing	31
3.6	Whole-genome sequencing of the pancreatic islets	32
3.7	A somatic variant caller was used to generate the matched and unmatched calls	33
3.8	Copy number analysis was performed to assess for losses and gains	34
3.9	A mean VAF filter was applied to remove germline variants	34
3.10	The beta-binomial distribution identified over-dispersed somatic variants	35
3.11	A depth filter ensured sufficient read numbers supported variants	35
3.12	Estimation of the observable mutational burden per cell	35
3.13	Mutational signature analysis identifies the distinct mutational processes active in a sample	36
3.14	Assessment of clonality using variant allele frequencies	36
3.15	A phylogenetic tree of pancreatic islet development can be reconstructed using data clustering algorithms	37
4.	Results	
4.1	Whole-genome sequencing of 32 pancreatic islets from the same individual	38
4.2	Successful identification of somatic mutations in individual islets	40
4.3	Use of the beta-binomial distribution to identify variable sites	43
4.4	The unmatched analysis provided comparable results to the matched analysis	46
4.5	Early embryonic mutations are identified by unmatched variant calling	49
4.6	Almost all mutations identified had no apparent functional impact	50
4.7	The observed mutational burden in the pancreatic islets is low	54
4.8	Intrinsic mutational processes dominate the pancreatic islets	56
4.9	The pancreatic islets are not clonal units	59

4.10	Phylogenetic reconstruction of the early embryonic lineage tree	60
4.11	The early ancestry of the islets appears to be fully explained	67
4.12	The early embryonic lineages of the bladder appear incomplete	69
4.13	Islets are composed of different embryonic lineages and are often dominated by one or two major lineages	72
4.14	Integrating the spatial location of the islets with their lineages reveals a non-random distribution across the pancreatic tissue	74
5. Discussion		
5.1	LCM with an unmatched analysis may prove to be a reproducible workflow in other normal tissues	78
5.2	Novel insights into somatic mutations of the pancreatic islets	78
5.3	The founding model of the pancreatic islets is still only partially understood	79
5.4	Limitations of the current methodology	82
6. Future directions		
6.1	Single-cell derived splenocyte colonies may help enrich the phylogenetic tree	84
6.2	Immunohistochemistry could play a role in explaining the lineage proportions	84
6.3	Targeted genotyping of pancreatic tissues may reveal more detailed insights into development and maintenance	85
7. References		87
8. Appendices		
8.1	Variant clustering with the n-dimensional hierarchical Dirichlet process	97
8.2	Copy number analysis revealed no detectable gains or losses across samples	98
8.3	Additional support for the mutation calling filters	99
8.4	Application of the pigeonhole principle to each of the 42 samples	101