LUDC-IRC Postdoctoral Program 2020 – Project 8

Bioinformatician for the study of α- and β-cell function in T2D – machine learning of human multi-omics data and functional follow-up

PI: Charlotte Ling
Co-PIs: Lena Eliasson and Tina Rönn

Purpose and aims

Overall aim: To identify and functionally validate novel, targetable disease mechanisms underlying Type 2 diabetes (T2D) with focus on the pancreatic islet α- and β-cells using multi-omics data and machine learning techniques.

Type 2 diabetes (T2D) is associated with deregulated hormonal secretion from the α- and β-cells within the pancreatic islets of Langerhans (1; 2). Although we today know much about the mechanisms behind defect β-cell function, there are still many open questions, and the regulation of α-cell glucagon secretion is largely unknown. In recent years we have discovered novel single entities affecting regulation of human islet gene expression essential for insulin secretion in the β-cells in development of T2D, e.g. differential DNA methylation (3; 4), differential expression of non-coding RNAs (5-7) and genetic variants (SNPs) (8-11). Although it is well established that there are strong interactions between genetics, epigenetics, gene expression and cell function (11; 12), most data analyses have been done using traditional statistical approaches and most omics data have been analysed separately. However, integration of all omics into one combined analysis allows for more accurate modelling which may generate new findings and give a more "true" picture of what is going on in the islets.

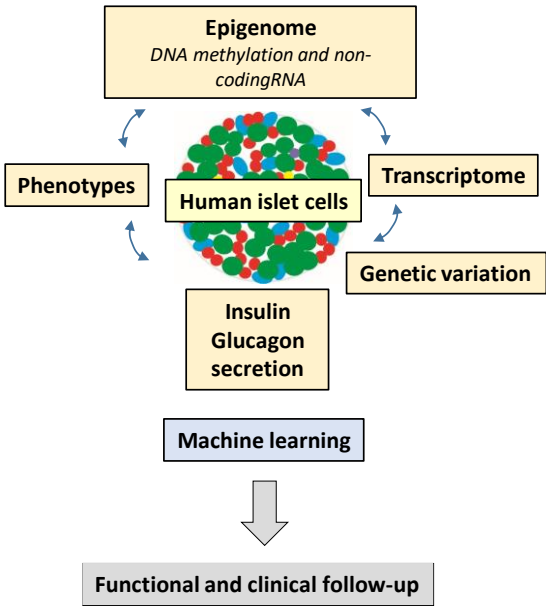

Fig. 1 Overview of the aim of this proposal

Here, we aim to 1) combine genome-wide DNA-methylation data, microRNA (miRNA) expression data, RNA-seq and SNP data from both whole islets and sorted human α- and β-cells using advanced bioinformatic analysis and machine learning approaches to obtain integrated gene networks controlling insulin and glucagon secretion, and 2) perform functional analyses of key nodes in these networks in both α- and β-cells (Fig. 1).

State-of-the-art/background

T2D is one of the largest global health emergencies predicting to reach >500 million people having the disease in 2030 (13). Poorly managed diabetes leads to devastating complications and early death. The disease is complex with a great diversity in the aetiology and clinical picture between patients, which complicates prevention as well as management of the disease. It is a multifactorial disease depending on both genetic and environmental factors such as diet and exercise (14). Better and more detailed characterization of islet biology together with measures of blood-based biomarkers in large cohorts (15; 16), may open new possibilities for investigation of hidden structures using mathematics for improved prediction of the disease – known as precision medicine (17). However, most of this work lies ahead of us.

We have long-time experience concerning studies on islet α- and β-cells (See e.g. ORCID: orcid.org/0000-0003-0587-7154, orcid.org/0000-0002-6467-5029 and orcid.org/0000-0001-7622-800X) with the aim to understand gene-regulation and function in the islets of healthy subjects and T2D patients. For example, recent work from us describes the role of epigenetic changes (3; 4; 12) and different types of non-coding RNAs (6; 7; 18-20) in transcriptional regulation of gene expression. Hence, our work describes how other factors than genetic variation can contribute to changes in gene expression. Moreover, previous work identified genetic variations associated with the risk of diabetes development, and demonstrated that most of the T2D risk SNPs associate with islet dysfunction (21). However, the identified risk SNPs only explain a modest part of the disease heritability and modest changes in gene expression in islet cells during diabetes development, why other factors must also contribute to the pathogenesis of T2D.

In this proposal we will combine multiple genome-wide omics data sets including expression of miRNAs, DNA methylation, RNA-sequencing and SNP data from sorted human α- and β-cells from both non-diabetic and T2D donors in a large network and use machine learning approaches to get an improved picture of α- and β-cell regulation in the disease. The same analysis will be performed using these omics also in our large cohort of human islets.

Significance and scientific novelty

The current exponential elevation of T2D cases, with increased morbidity and mortality, will lead to large individual sufferings and a huge burden on health care systems worldwide. Thus, it is important to identify the cause of the disease and ways it may be ameliorated through prediction and treatment regimes. The pancreatic islet hormones are central in glucose homeostasis and defects in insulin and glucagon secretion are central in the development of T2D. With the current project we focus on human islet α- and β-cells and aim to significantly contribute to better prediction and treatment of the disease, with emphasis on gene interaction networks controlling the complex islet cell function. The project is very timely since it relies on several recent advances in both molecular biology and computational biology, as well as large scale human genetics/transcriptomics. Within this project we will use high quality human islet cell data and novel bioinformatic tools combined with machine learning to build cell-type specific SNP – DNA methylation – miRNA – mRNA – hormone secretion networks followed by functional analyses in both cell lines, primary cells and animal models. Our aim is to use machine learning to perform multi-omics analyses and thereby understand, with cell type specificity, which gene regulatory networks that leads to the development of T2D. This effort will generate large-scale cell type-specific gene-interactomic maps of the pancreatic islet α- and β-cells that are likely to be suitable for drug targets. Our findings will have great importance for both patients and the society.

Preliminary and previous results

Novel interaction networks: We have established a pipeline to analyse multi-omics data in an integrated way using machine learning. This was based on a human case control cohort/data set containing four different omics. The human cohort was randomly split into a train and a test set. Then, feature pre-selection from the different omics was performed to reduce dimensionality and balance the data set with regards to number of human samples. Finally, we used machine learning for the multi-omics integration and validated the selected targets on our test set. Our preliminary data show that the prediction accuracy for the multi-omics data generated is very high (AUC 0.94±0.09). We plan to use a similar approach in the present proposal and further develop our existing machine learning pipeline for integrated multi-omics analysis. We have also performed gene network analysis of miRNA expression and islet gene expression from RNA sequencing data from the same donors as a high-throughput miRNA target validation approach. We have constructed scale-free miRNA – mRNA network and our preliminary data confirm that each miRNA has several targets. Among the most abundant miRNA we found the well-studied miR-375 (22; 23).

DNA-methylation data in human islets: We have generated genome-wide DNA methylation data from our large human islet cohort and we are currently analysing these data. The preliminary results look promising and these epigenetic data will be used in the present proposal.

miRNAs and associated novel target proteins have impact on β-cell function: We have novel data on sets of single miRNAs such as miR-130a, miR-130b and miR-152 in human islets. These miRNAs showed increased expression in islets from T2D donors ((6); Fig. 2). Detailed target analysis could validate pyruvate dehydrogenase E1 alpha (PDHA1) as target. Overexpression of the individual miRNAs and/or knockdown of

PDHA1 had effects on β-cell function and resulted in decreased glucose-induced elevation of ATP and reduced insulin secretion. We have promising preliminary data of other miRNAs including expression data from human islets and sorted α- and β-cells and functional follow up indicating effects on glucagon- and insulin secretion, respectively. One miRNA that is overexpressed in islets from T2D donors ($p<0.05$) showed reduced glucose-stimulated insulin secretion when overexpressed in an insulin-secreting cell line ($p<0.05$). Moreover, using a specific antagomir to lower the expression of this miRNA in human islet resulted in improved insulin secretion.

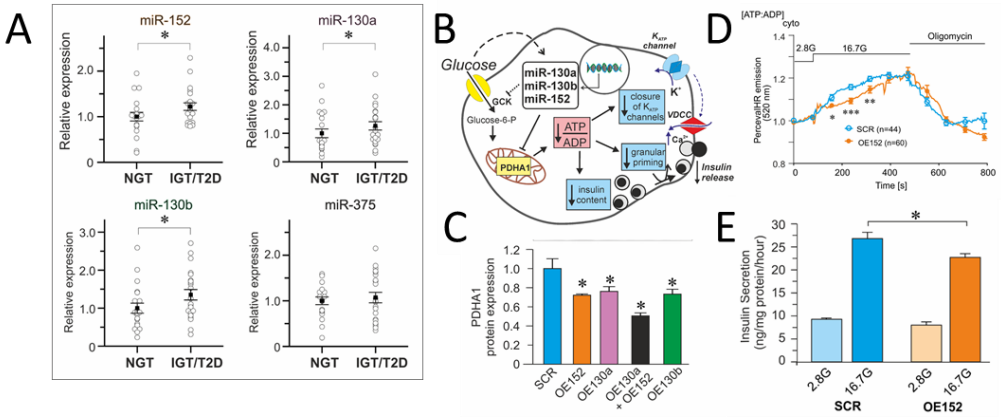

Fig. 2. Data from our recently published paper Ofori et al Scientific Reports 2017 **A.** Expression of miR-152, miR-130a, miR-130b and miR-375 in human islets from NGT and IGT/T2D donors. Notice upregulation of all except miR-375. **B** Summary of results. **C** Protein expression of the validated target PDHA1 after overexpression of miR-152, miR-130a and miR-130b in INS1-832/13 cells. **D** Perceval measurements of dynamic ATP:ADP ratio in control (SCR) and miR-152 overexpressing (OE152) cells. **E**. Glucose-stimulated insulin secretion in miR-152 overexpressing cells, compared

Research plan

AIM 1. Investigate gene-interactomic networks in α- and β-cells using human islet cell data and machine learning.

Rationale: It is possible that integrating multiple layers of information (multi-omics data) can give a more accurate understanding of T2D compared to separately analysing the individual omics.

Methods: Here, we plan to use a machine learning approach for performing an integrative analysis of DNA methylation, miRNA, gene expression, SNPs and phenotypic information from ~200 human pancreatic islets of T2D and non-diabetic donors as well as from sorted human α- and β-cells from a subset of the islet cohort. Human islets will be provided through the Human Tissue Lab at LUDC and high profile flow-cytometry has been utilized for sorting of the human islet cells. We have generated a large part of the genome-wide data needed for this human cohort and are in the process of generating miRNA data through sequencing. For machine learning and integration of multi-omics data, we plan to use linear supervised models since our cohort is relatively small with regards to numbers of features tested and the clear phenotype of interest (T2D cases vs. non-diabetic islet donors). The human cohort will be randomly split into a train (75% of samples) and test set (25% of samples), repeated 100 times. We will perform a feature pre-selection step from the different omics to reduce dimensionality and select the features that best separate the two groups. Next, we will include data from all omics into the integrative model. After 100 iteration we will be able to build confidence intervals of the predictive model and visualize the relations between the markers. This will be done both for data from human islets and for the data from the sorted human α- and β-cells.

Risk assessment: We believe that the risk of this proposal is relatively small since the human islet cohort is already available, we have established the sorting of human α- and β-cells with high purity fractions (>98% purity) and most individual omics data exist. Also, the machine learning pipeline has been established, in collaboration with Dr Nikolay Oskolkov at WABI, and is ready to be used.

AIM2. Functional follow up of key nodes and pathways in these networks.

Rationale: Do genes identified through machine learning in human islets and sorted human α- and β-cells affect insulin and/or glucagon secretion? And if so, through which mechanisms?

Methods: Functional follow up of key genes identified in AIM 1 will be performed by knock-down and/or over-expression experiments as well as by epigenetic editing using dead CRISPR/Cas9. We have access to human islets, α- and β-cell lines as well as diabetic rodent animal models where these experiments can be performed. Apart from hormone secretion measurements, cell and animal models will also be studied using

live confocal microscopy, seahorse and patch-clamp. In addition, we have access to a large number of transmisson electron microscopic images from the same human islets.

Risk assessment: We believe the risk of this proposal is relatively small since we have large experience from similar experiments and all functional methods are available within our groups. It is possible that some identified genes in AIM 1 do not affect α- and β-cell function. However, since we aim to follow up several genes, some are likely to play a role in insulin and/or glucagon secretion and thereby the pathogenesis of T2D.

Timeline

AIM 1 (year 2020-2021) and AIM 2 (year 2021-2022)

Implementation

Prof Charlotte Ling and Dr Tina Rönn are experts in epigenetics, they have developed a pipeline for machine learning to integrate multi-omics data. Prof Lena Eliasson is an expert in α-and β-cell physiology and the role of miRNAs in islet cell function. We have all tools needed to perform the experiments and are looking for a post-doc that can assist in the bioinformatic analysis of the data.

Project organization

Charlotte Ling, Lena Eliasson and Tina Rönn are the PI:s of this project. The recruited post-doc will take a large part in the bioinformatics/machine learning analyses with assistance from members of CL and LE groups. Functional analysis will be performed in collaboration with members of CL and LE groups (Karl Bacos – siRNA and overexpression in α- and β-cells; Sabrina Ruhrmann – epigenetic editing; Jones Ofori - sorting of islet cells; Anna Wendt - TEM and patch clamp; Jonathan Esguerra – molecular biology and confocal microscopy).

References

1. Ashcroft FM, Rorsman P: Diabetes mellitus and the beta cell: the last ten years. Cell 2012;148:1160-1171
2. D'Alessio D: The role of dysregulated glucagon secretion in type 2 diabetes. Diabetes, obesity & metabolism 2011;13 Suppl 1:126-132
3. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Ronn T, Bacos K, Ling C: Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS genetics 2014;10:e1004160
4. Volkov P, Bacos K, Ofori JK, Esguerra JL, Eliasson L, Ronn T, Ling C: Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. Diabetes 2017;66:1074-1085
5. Eliasson L, Esguerra JLS: MicroRNA Networks in Pancreatic Islet Cells: Normal Function and Type 2 Diabetes. Diabetes 2020;69:804-812
6. Ofori JK, Salunkhe VA, Bagge A, Vishnu N, Nagao M, Mulder H, Wollheim CB, Eliasson L, Esguerra JL: Elevated miR-130a/miR130b/miR-152 expression reduces intracellular ATP levels in the pancreatic beta cell. Sci Rep 2017;7:44986
7. Salunkhe VA, Ofori JK, Gandasi NR, Salo SA, Hansson S, Andersson ME, Wendt A, Barg S, Esguerra JLS, Eliasson L: MiR-335 overexpression impairs insulin secretion through defective priming of insulin vesicles. Physiol Rep 2017;5
8. Andersson SA, Olsson AH, Esguerra JL, Heimann E, Ladenvall C, Edlund A, Salehi A, Taneera J, Degerman E, Groop L, Ling C, Eliasson L: Reduced insulin secretion correlates with decreased expression of exocytotic genes in pancreatic islets from patients with type 2 diabetes. Mol Cell Endocrinol 2012;364:36-45
9. Dayeh TA, Olsson AH, Volkov P, Almgren P, Ronn T, Ling C: Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia 2013;56:1036-1046
10. Koeck T, Olsson AH, Nitert MD, Sharoyko VV, Ladenvall C, Kotova O, Reiling E, Ronn T, Parikh H, Taneera J, Eriksson JG, Metodiev MD, Larsson NG, Balhuizen A, Luthman H, Stancakova A, Kuusisto J, Laakso M, Poulsen P, Vaag A, Groop L, Lyssenko V, Mulder H, Ling C: A common variant in TFB1M is associated with reduced insulin secretion and increased future risk of type 2 diabetes. Cell Metab 2011;13:80-91
11. Olsson AH, Volkov P, Bacos K, Dayeh T, Hall E, Nilsson EA, Ladenvall C, Ronn T, Ling C: Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. PLoS genetics 2014;10:e1004735
12. Hall E, Volkov P, Dayeh T, Esguerra JL, Salo S, Eliasson L, Ronn T, Bacos K, Ling C: Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. Genome Biol 2014;15:522

13. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B: IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018;138:271-281
14. Ling C, Groop L: Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 2009;58:2718-2725
15. Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, Gjesing AP, Eiberg H, Tuomi T, Almgren P, Groop L, Eliasson L, Vaag A, Dayeh T, Ling C: Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. Nat Commun 2016;7:11089
16. Gallo W, Esguerra JLS, Eliasson L, Melander O: miR-483-5p associates with obesity and insulin resistance and independently associates with new onset diabetes mellitus and cardiovascular disease. PLoS One 2018;13:e0206974
17. Fitipaldi H, McCarthy MI, Florez JC, Franks PW: A Global Overview of Precision Medicine in Type 2 Diabetes. Diabetes 2018;67:1911-1922
18. Henaoui IS, Jacovetti C, Guerra Mollet I, Guay C, Sobel J, Eliasson L, Regazzi R: PIWI-interacting RNAs as novel regulators of pancreatic beta cell function. Diabetologia 2017;60:1977-1986
19. Mollet IG, Malm HA, Wendt A, Orho-Melander M, Eliasson L: CAMTA1-Calcium-Calmodulin Transcriptional Activator 1, a new player in the regulation of microRNAs and insulin secretion. Diabetologia 2015;58:S88-S88
20. Motterle A, Gattesco S, Peyot ML, Esguerra JLS, Gomez-Ruiz A, Laybutt DR, Gilon P, Burdet F, Ibberson M, Eliasson L, Prentki M, Regazzi R: Identification of islet-enriched long non-coding RNAs contributing to beta-cell failure in type 2 diabetes. Mol Metab 2017;6:1407-1418
21. Groop L, Pociot F: Genetics of diabetes--are we missing the genes or the disease? Mol Cell Endocrinol 2014;382:726-739
22. Eliasson L: The small RNA miR-375 - a pancreatic islet abundant miRNA with multiple roles in endocrine beta cell function. Mol Cell Endocrinol 2017;456:95-101
23. Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, Stoffel M: A pancreatic islet-specific microRNA regulates insulin secretion. Nature 2004;432:226-230