MS Thesis Proposal

Tim Vigers

Table of Contents

[Abstract 3](#_Toc37063366)

[Introduction 3](#_Toc37063367)

[Specific Aims 4](#_Toc37063368)

[Primary Aim 1: 4](#_Toc37063369)

[Secondary Aim 1a: 4](#_Toc37063370)

[Secondary Aim 1b: 4](#_Toc37063371)

[Methods 4](#_Toc37063372)

[Methylation 4](#_Toc37063373)

[Metabolomics 5](#_Toc37063374)

[Candidate selection 5](#_Toc37063375)

# Abstract

# Introduction

Type 1 diabetes (T1D) is an autoimmune disease characterized by the production of antibodies which target pancreatic -cells. This leads to both micro and macrovascular complications, and individuals with T1D are at an increased risk of experiencing a major cardiac event (e.g. stroke, angina, and myocardial infarction) and are less likely to have satisfactory outcomes after a coronary event1. The disease currently affects over 30 million people worldwide2, and is increasing by 3-4% per year on average3. However, the global burden of disease is difficult to estimate due to geographic variation in incidence1,3.

Genetic predisposition accounts for some of the etiology of T1D (sibling relative risk has been estimated at 154) and explains some geographic variation in incidence3. Human leukocyte antigen (HLA) genes were the first to be linked to T1D and account for much of the genetic predisposition to the disease, but genome-wide association studies (GWAS) have also identified more than 40 other loci associated with increased relative risk of T1D. However, it is still unclear how these multiple loci interact with one another and the environment to eventually produce a T1D diagnosis. And in addition to the complex genetics of T1D, low monozygotic (MZ) twin concordance (approximately 50%) and studies of migrant populations support the theory that non-genetic factors play an important role in T1D development2.

Epigenetic differences may be important contributors to non-genetic T1D etiology. Changes in methylation have been associated with other autoimmune conditions, and monozygotic twins can be epigenetically heterogenous despite sharing identical genetic code2. Rakyan et al. and Stefan et al. performed epigenome-wide association studies (EWAS) in discordant and concordant twin pairs and found that methylation profiles were more similar among participants with T1D than to unaffected twins. Epigenetics profiles were also combined with GWAS data and differentially expressed methylation sites were mapped to 6 well known T1D susceptibility genes, including two major histocompatibility complex (MHC) genes and several HLA loci2,5. These results suggest that methylation is vital in the progression of T1D.

Environmental factors including viruses, diet, and the metabolome have also been linked with T1D etiology1. Metabolites are small molecule products of metabolism, and are involved in many vital processes, including energy storage, cellular signaling and apoptosis, post-translational protein modification and transport, and maintenance of homeostasis in the cellular milieu6. Analysis of the metabolome can therefore quantify the integrated response to endogenous and exogenous disease factors or other physiological changes. Previous studies have found associations between T1D and differentially expressed phospholipids and sphingolipids, excretion of modified amino acids, and vitamin D (and related compounds on its metabolic pathway)1,7.

We have collected methylation and metabolomic data …

Integration of epigenetic and metabolomics data requires statistical methods capable of evaluating associations between many different variables in complex multilevel networks. Bayesian networks are becoming increasingly popular for analysis of large scale systems biology (“omics”) data, including protein-protein interactions and gene regulatory networks8. This approach can be used to generate intuitive graphical models, which represent probabilistic dependence between multiple variables8,9 and avoid many of the pitfalls of traditional mediation analyses10. We hope to derive a graphical model of the relationships between metabolites and methylation sites, in order to illustrate and make inferences on the epigenetic and environmental pathways that lead to T1D.

# Specific Aims

## **Primary Aim 1:**

Construct multi-Omics networks connecting T1D status, metabolite concentrations, and methylation levels at select CpG sites.

### Secondary Aim 1a:

Compare the triad and SmCCNet approaches to feature reduction.

### Secondary Aim 1b:

Compare BNA triad results from the “bnlearn” R package to likelihood-based causal mediation results from the “cit” R package.

## **Primary Aim 2:**

Use BNA on reduced feature set to learn the structure of methylation and metabolomic pathways related to T1D.

1. Use Bayesian network approaches to learn the structure of methylation and metabolomic pathways related to T1D. This final network structure will allow us to analyze the probabilistic dependence relationships between metabolites, methylation sites, and T1D.
2. Place the final network structure in biological context using known metabolomic and methylation pathways from previous research and public databases such as MethDB and the Human Metabolome Database (HMDB).

# Methods

## Methylation

Need details on probe selection, instruments, etc.

## Metabolomics

Ask about this too.

## Candidate selection

First, we will select methylation-metabolite pairs by using linear models to correlate all combinations of probe and metabolite. We will also use logistic regression to find probes and metabolites that are significantly associated with the T1D phenotype. Candidate pairs that are strongly associated (nominal p-value < 0.001) and contain either a probe or metabolite (or both) that is significantly associated with the T1D phenotype (nominal p-value < 0.05) will continue on to the next step of our Bayesian network analysis (BNA). We will not adjust these p-values for multiple comparisons because doing is complicated with many correlated variables, and there are no agreed upon best practices11. However, careful network development with strict thresholds in later steps should prevent false positives in the final results, as described in Rudra et al.

Next, we will perform BNA separately for each probe-metabolite-phenotype triplet. With no additional restrictions on edge number or direction (aside from the assumptions of Bayesian networks), there are 24 possible network structures for each triple. We will compare all possible network structures using a Bayesian Information Criterion (BIC) score, and the structure with the best score will continue on to the next selection step…

References

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *The Lancet*. 2014;383(9911):69-82. doi:10.1016/S0140-6736(13)60591-7

2. Rakyan VK, Beyan H, Down TA, et al. Identification of Type 1 Diabetes–Associated DNA Methylation Variable Positions That Precede Disease Diagnosis. *PLOS Genet*. 2011;7(9):e1002300. doi:10.1371/journal.pgen.1002300

3. Tuomilehto J. The Emerging Global Epidemic of Type 1 Diabetes. *Curr Diab Rep*. 2013;13(6):795-804. doi:10.1007/s11892-013-0433-5

4. Polychronakos C, Li Q. Understanding type 1 diabetes through genetics: advances and prospects. *Nat Rev Genet*. 2011;12(11):781-792. doi:10.1038/nrg3069

5. Stefan M, Zhang W, Concepcion E, Yi Z, Tomer Y. DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J Autoimmun*. 2014;50:33-37. doi:10.1016/j.jaut.2013.10.001

6. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol*. 2016;17(7):451-459. doi:10.1038/nrm.2016.25

7. Frohnert BI, Rewers MJ. Metabolomics in childhood diabetes. *Pediatr Diabetes*. 2016;17(1):3-14. doi:10.1111/pedi.12323

8. Shaddox E, Stingo FC, Peterson CB, et al. A Bayesian approach for learning gene networks underlying disease severity in COPD. *Stat Biosci*. 2018;10(1):59-85.

9. Sachs K, Perez O, Pe’er D, Lauffenburger DA, Nolan GP. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science*. 2005;308(5721):523-529. doi:10.1126/science.1105809

10. Rudra P, Shi WJ, Russell P, et al. Predictive modeling of miRNA-mediated predisposition to alcohol-related phenotypes in mouse. *BMC Genomics*. 2018;19(1):639. doi:10.1186/s12864-018-5004-3

11. Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *Am J Hum Genet*. 2007;81(6):1158-1168. doi:10.1086/522036