T1D & T2D Methylation Analysis

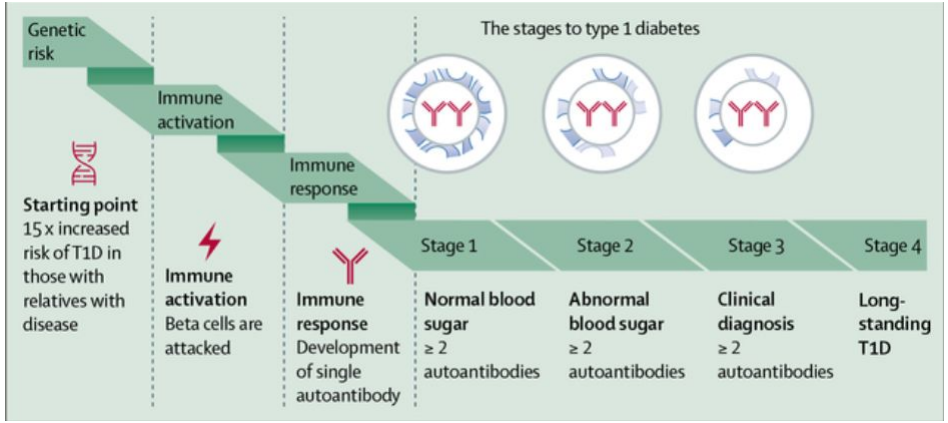
Jen Zeiger & Emily Baranowski

T1D Introduction

-Autoimmune disease caused by genetic & **non-genetic factors**

-Leads to destruction of insulin-secreting pancreatic islet beta cells by the immune system

-Characterized by insulin deficiency, due to beta cell loss



DiMeglio, Linda A et al. “Type 1 diabetes.” Lancet (London, England) vol. 391,10138 (2018): 2449-2462.

Identification of Type 1 Diabetes–Associated DNA Methylation Variable Positions That Precede Disease Diagnosis

Hypothesis - Epigenetic variation may contribute to the non-genetic factors of T1D, such variation may affect immune effector cell function/disease susceptibility

Previous studies have found that monozygotic (MZ) twins can be epigenetically discordant

- ➔ MZ Twins - 50% change of a T1D affected co-twin developing the disease as well, even though they are genetically identical

Method: DNA methylation profiling with Illumina 27K

Genome-wide DNA methylation analysis of CD14+ monocytes from MZ twins discordant for T1D

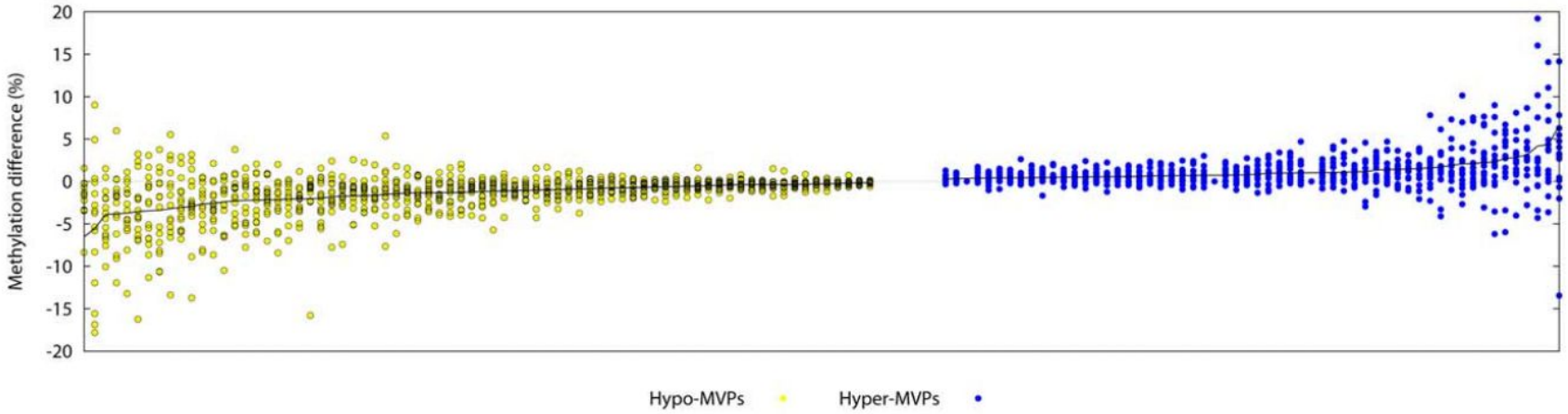
➔ CD14+ monocytes give rise to macrophages associated with islet cell destruction/insulin deficiency

Illumina 27k allows for DNA methylation measurements at 27,458 different CpG sites across the genome

➔ Examined methylation differences between T1D vs Unaffected MZ twins to identify T1D methylation variable positions (MVPs)

Results

T1D–specific methylation variable positions (T1D–MVPs) in the T1D–affected co-twins were identified



R Analysis

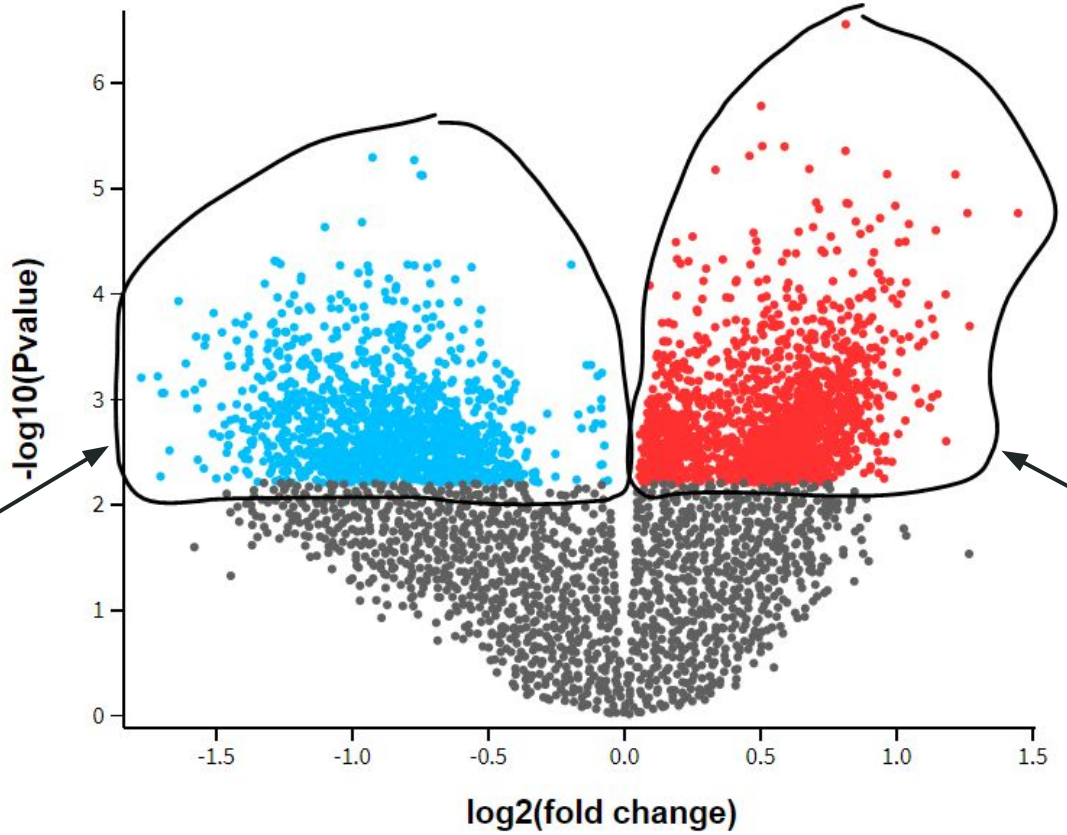
Step 1	Step 2	Step 3	Step 4 (ongoing)
Grouped samples into: 1. T1D/Unaffected Discordant Twins CD4 tissue vs Normal MZ Twins CD4 tissue 2. T1D CD4 tissue vs T1D CD14+ monocytes	Ran R script through GEO2R (available via NCBI) 	Modified R script to reflect chosen data / graphs	Attempting to run Quality Control on chosen data, but having issues with this :'(

T1D/Unaffected Discordant Twins CD4 tissue vs Normal MZ Twins CD4 tissue

GEO Dataset contained CD4 tissue samples from MZ twins discordant for T1D as well as normal MZ twins

- ➔ Article results mainly discussed CD14+ monocytes despite data for CD4 tissue being present
- ➔ Comparison made to determine if differences exist between different twin types

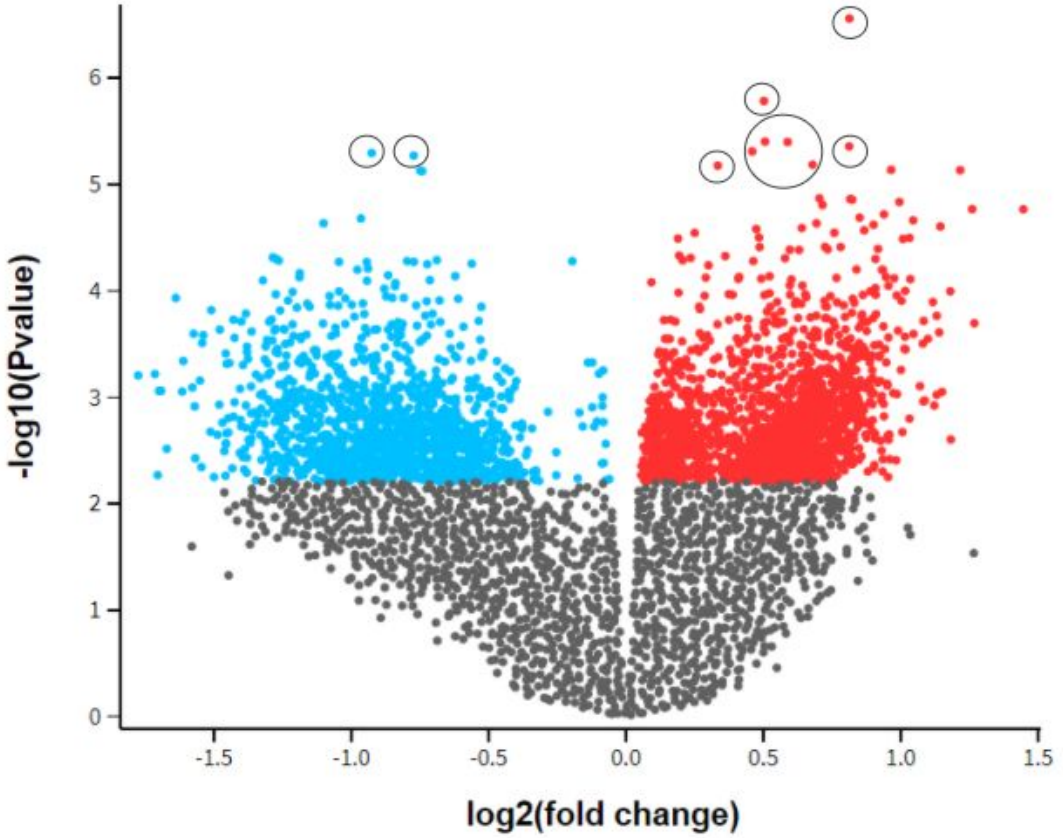
Volcano Plot



Hypomethylated CpG sites

Hypermethylated CpG sites

Top 10 Differentially Expressed Genes



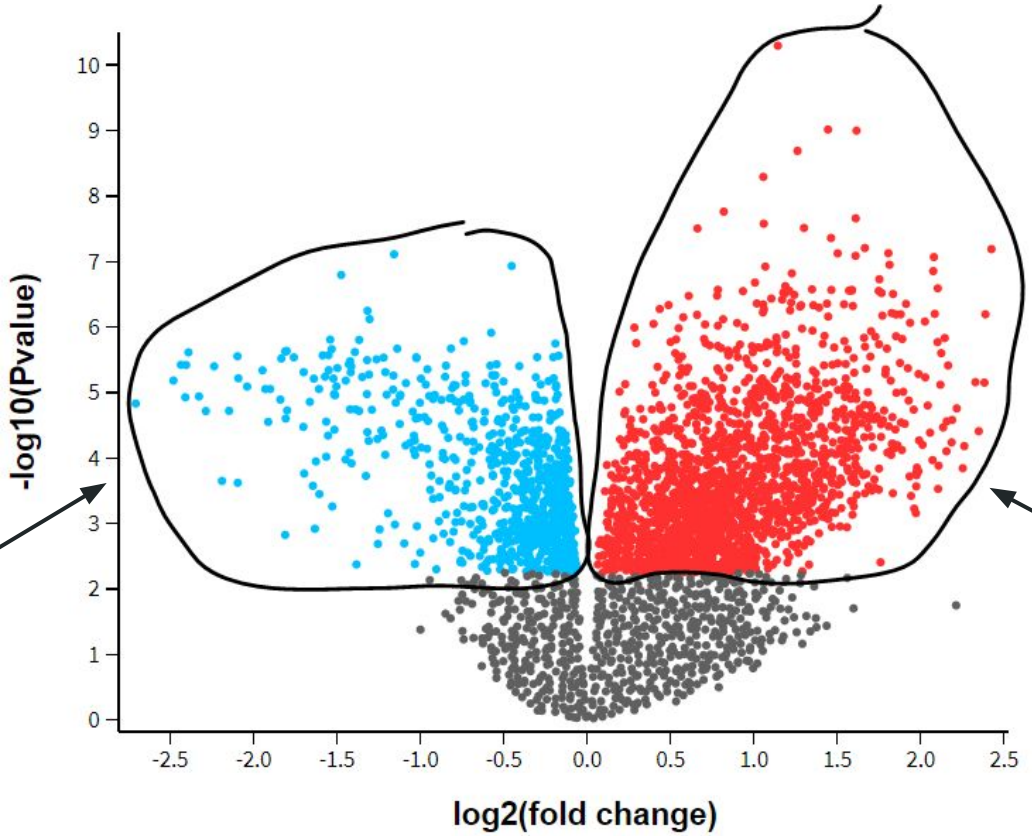
ID	Symbol	CPG_ISLAND
cg11304734	PRAF1	TRUE
cg08209133	SLC10A4	TRUE
cg17189494	MAD2L2	TRUE
cg00015770	GPR103	TRUE
cg03963198	IRX4	TRUE
cg23616741	PGAM2	FALSE
cg16909364	THAP2	TRUE
cg12620499	DSU	TRUE
cg03923934	OPCML	TRUE
cg24178740	FEV	TRUE

T1D CD4 tissue vs T1D CD14+ monocytes

GEO Dataset contained T1D CD4 tissue samples and CD14+ monocyte samples from the same twin with T1D

- ➔ Article did not mention any differences between CD4 and CD14+ T1D samples despite data being present for both tissue types
- ➔ Comparison made to determine if differences exist between different tissue types within patients with T1D

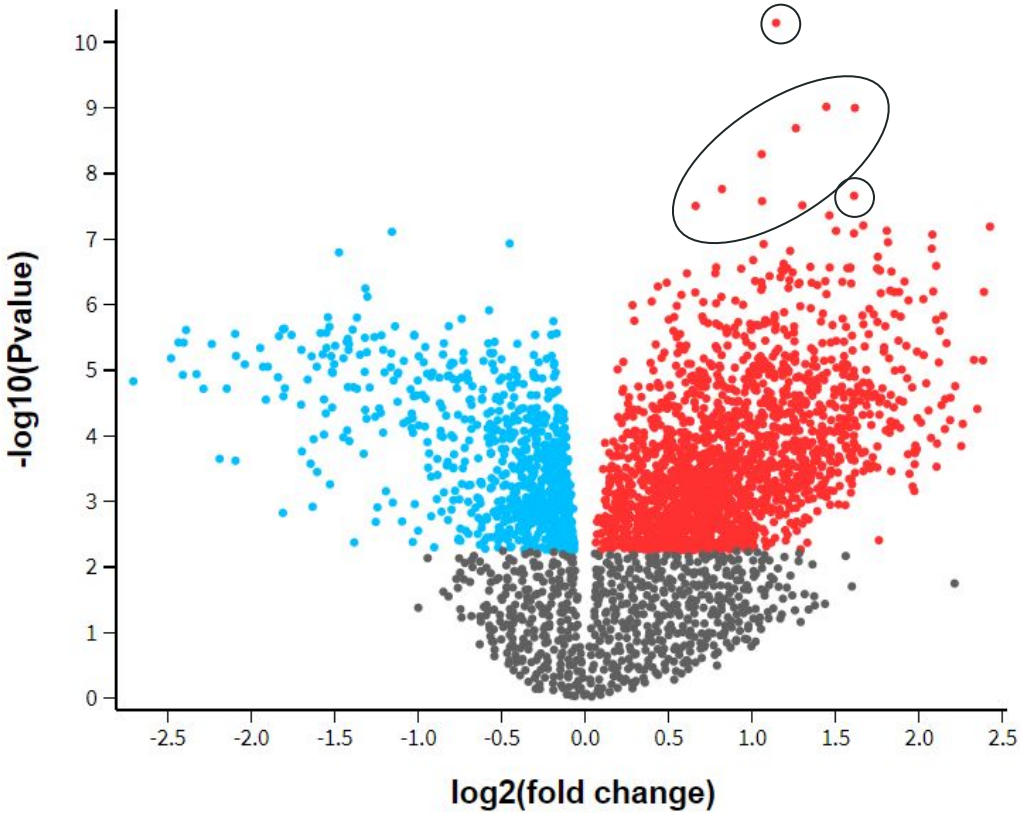
Volcano Plot



Hypomethylated CpG sites

Hypermethylated CpG sites

Top 10 Differentially Expressed Genes



ID	Symbol	CPG_ISLAND
cg06800962	TM4SF1	FALSE
cg08116137	NAT8	FALSE
cg00563926	TGFBR3	TRUE
cg10451565	GPR77	FALSE
cg21491308	TEX101	FALSE
cg06750167	TESC	FALSE
cg08510456	BSN	TRUE
cg17980508	IFI44L	FALSE
cg21792432	POT1	FALSE
cg12113819	THRAP2	TRUE

T2D Introduction

-Diabetes caused by muscle / fat / liver cells becoming resistant to insulin

-Cells do not take in enough sugar; pancreas does not produce enough insulin to regulate blood sugar levels

-As with T1D, there are both genetic and non-genetic factors that lead to disease

Methylated DNA Immunoprecipitation (MeDIP) using a custom type 2 diabetes and related phenotypes array

Integrated Genetic and Epigenetic Analysis Identifies Haplotype-Specific Methylation in the *FTO* Type 2 Diabetes and Obesity Susceptibility Locus

Hypothesis = Genotype - epigenotype interactions may underlie the aetiopathogenesis of T2D.

Method: NimbleGen human 385k tiling array & MeDIP

NimbleGen tiling array is used to probe intensively for sequences that are known to exist in a contiguous region; 385k array can tile up to that many 50 to 75-mer probes.

MeDIP, or methylation DNA immunoprecipitation, is performed to quantify DNA methylation.

Results

Table 1

Average Methylation in Association SNP LD blocks by Genotype.

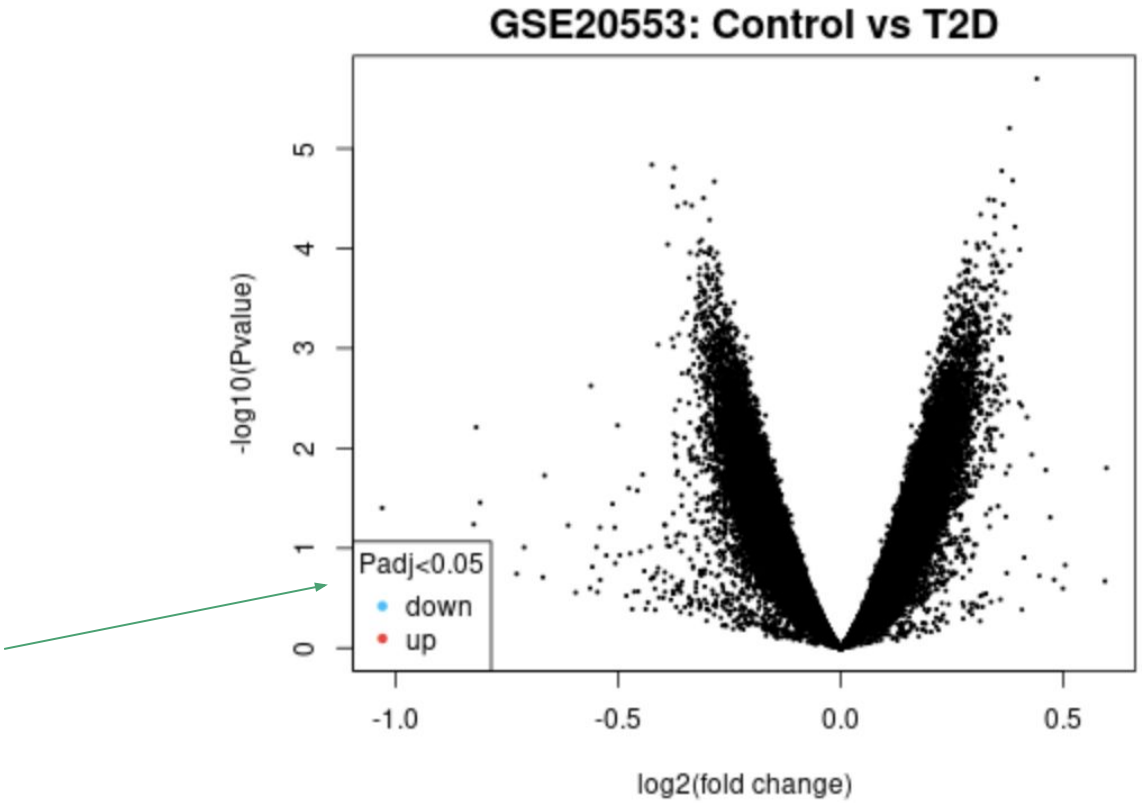
					Average Methylation			
Chr	LD Block		Genotyped SNP	Gene/Locus	11	12	22	*p*-value
1	120236149	120398430	rs2934381	*NOTCH2*	0.496	0.495	0.502	0.187
2	43529937	43617946	rs7578597	*THADA*	0.492	0.512	0.504	0.564
3	12298413	12372392	rs1801282	*PPARG*	0.512	0.509	0.511	0.640
3	64673853	64705161	rs4607103	*ADAMTS9*	0.477	0.472	0.481	0.760
3	186971576	187031377	rs4402960	*IGF2BP2*	0.502	0.493	0.502	0.016
4	6317902	6363877	rs10010131	*WFS1*	0.581	0.604	0.590	0.982
7	28147081	28175361	rs864745	*JAZF1*	0.501	0.492	0.494	0.317
8	118252732	118254914	rs13266634	*SLC30A8*	0.333	0.350	0.303	0.865
9	22122209	22126489	rs10811661	*CDKN2A/CDKN2B*	0.543	0.512	0.512	0.389
10	12367941	12368040	rs12779790	*CDC123/CAMK1D*	0.611	0.590	0.671	0.913
10	94426831	94467199	rs1111875	*HHEX/IDE*	0.480	0.483	0.483	0.369
11	17350649	17365206	rs5219	*KCNJ11*	0.501	0.502	0.498	0.540
12	69942990	69949369	rs7961581	*TSPAN8*	0.457	0.461	0.470	0.289
16	52357008	52402988	rs8050136	*FTO*	0.497	0.510	0.531	$9.397'10^{-4}$
17	33170413	33182480	rs757210	*HNF1B*	0.423	0.430	0.427	0.382

Bell et. al. (2010). Integrated Genetic and Epigenetic Analysis Identifies Haplotype-Specific Methylation in the *FTO* Type 2 Diabetes and Obesity Susceptibility Locus. *PLOS One*, *5*(11), e14040. doi: 10.1371/journal.pone.0014040

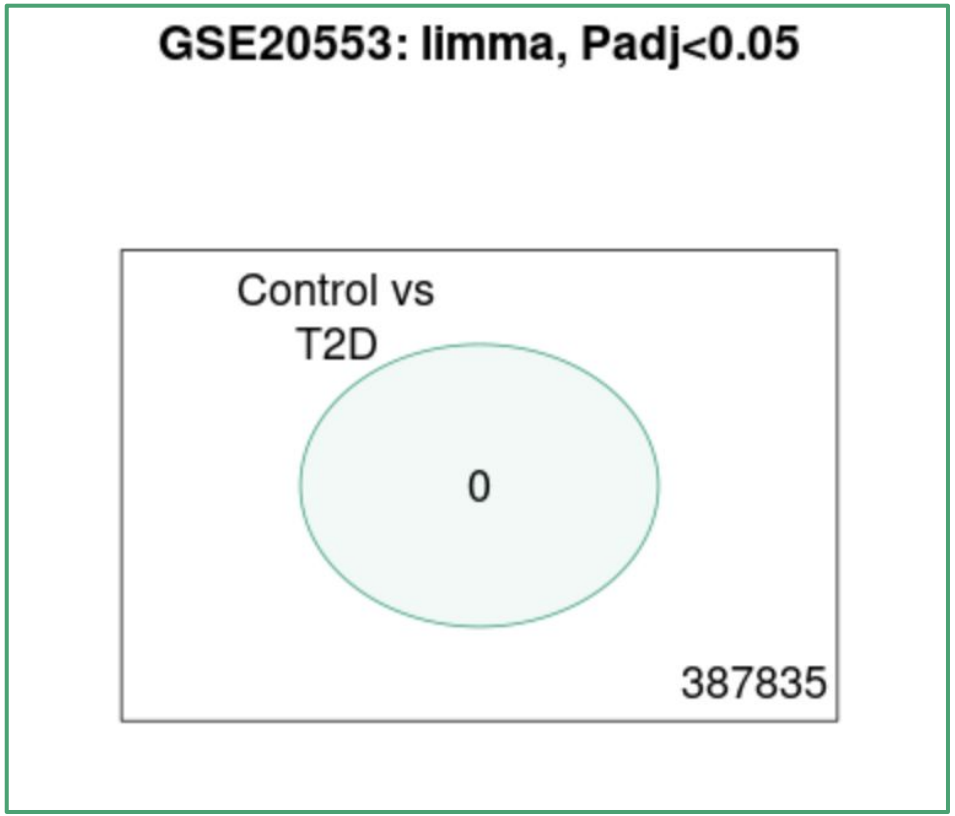
R Analysis / Workflow

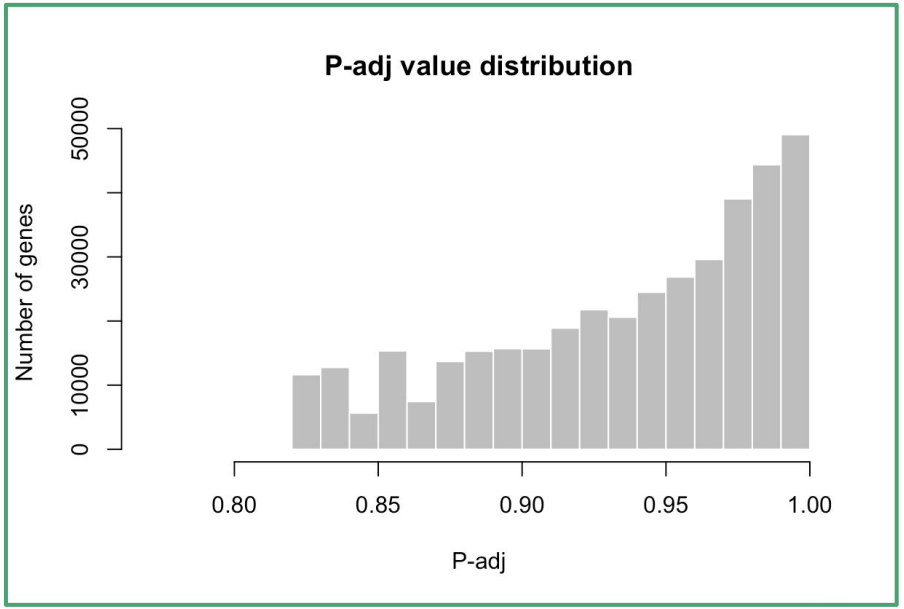
Step 1	Step 2	Step 3	Step 4 / Current
Grouped samples into control or T2D → 60 samples total = 30 control, 30 case	Ran R script through GEO2R (available via NCBI)	Modified R script to reflect chosen data / graphs	Attempted to run “qcreport” on data; currently having issues with this :(

Volcano Plot



Histogram of adjusted P-values and Limma Plot





Next Step

Quality control in R-

```
# read in the sample sheet for the experiment

targets <- read.metharray.sheet(dataDirectory, pattern="SampleSheet.csv")
```

Issue - dataDirectory - ‘base not found’ error message pops up

This step needs to be done before the quality control

Final Remarks

-Comparison between the T1D discordant twins and normal MZ twins showed that 9 of the 10 top differentially expressed genes had CpG sites present. Looking at CpG methylation differences between T1D and control subjects was part of the initial proposal.

-Comparison between T2D and control (from that same study) did not show any significant difference in differentially expressed genes

-Unfortunately, data between T1D and T2D is too different to compare the two with each other

-Will troubleshoot QC issue and hopefully resolve! :)