Supplementary Information

Read records & provide summary statistics

Perform predictive analytics from clinical data

Integrate gene expression profiles from GEO

Provide proof-of-concept for combined analytics
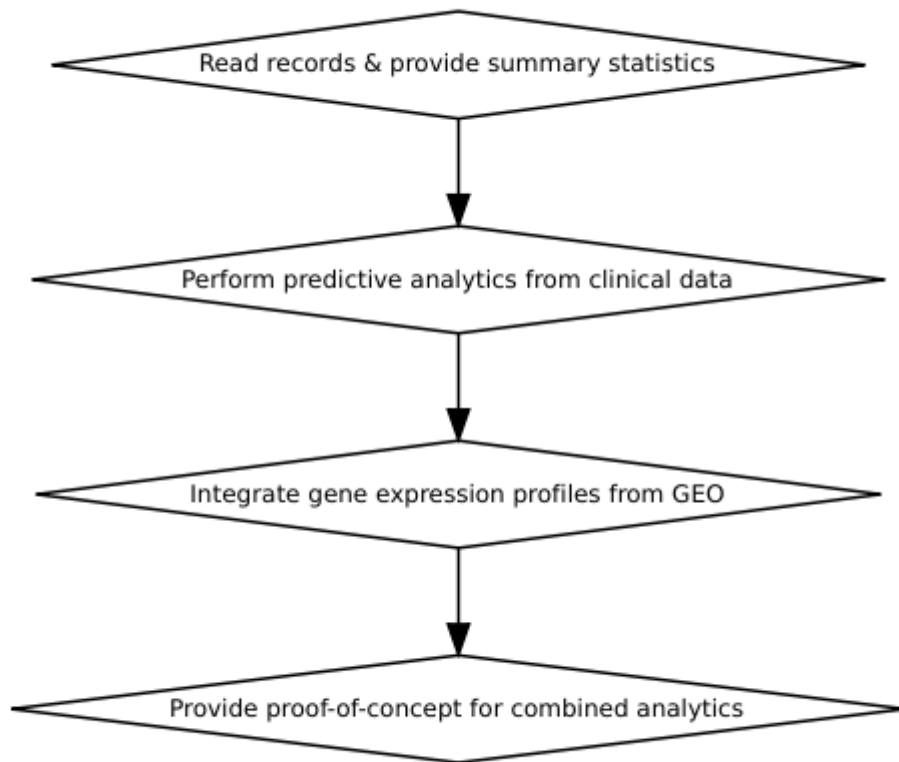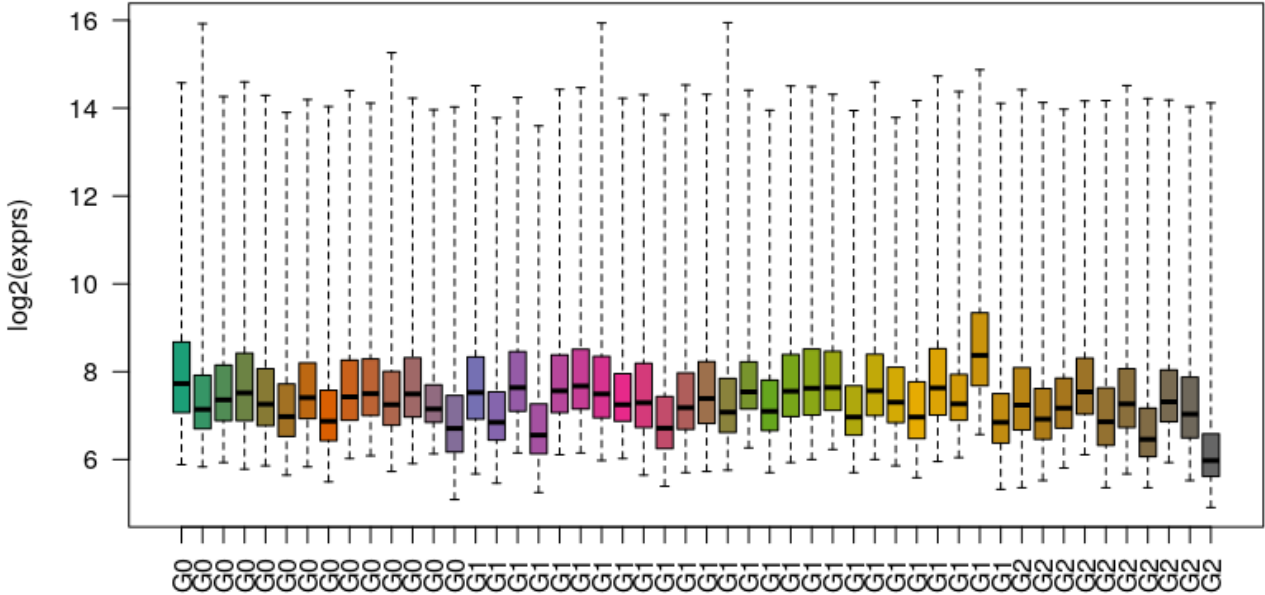
Clinical Data

1. Disorder Frequency

Frequency of each metabolics disorder

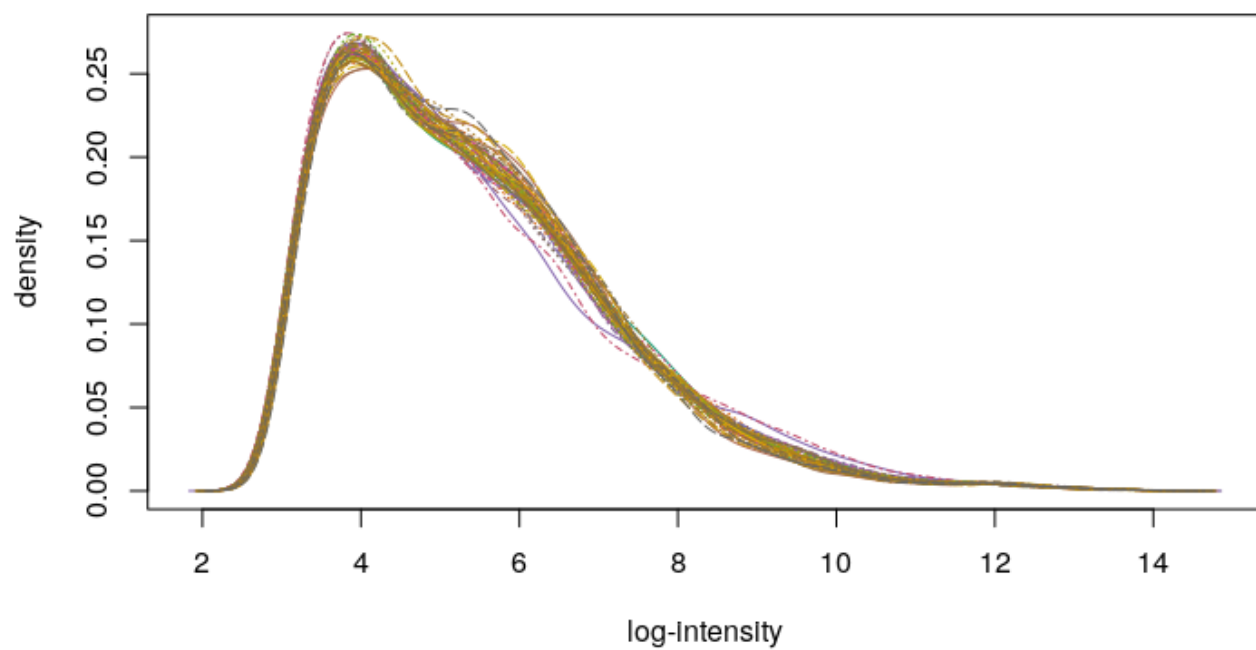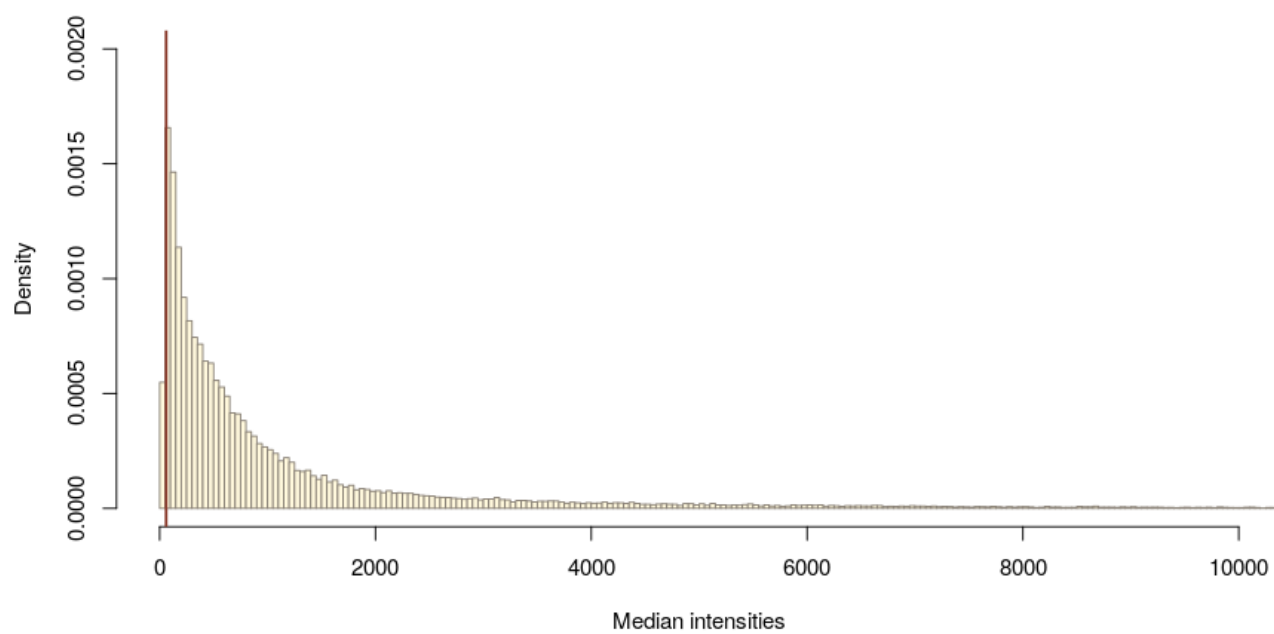| Disorder	Count
Stroke	54
Proteinuria due to type 2 diabetes mellitus (disorder)	3
Proliferative diabetic retinopathy due to type II diabetes mellitus (disorder)	1
Prediabetes	293
Nonproliferative diabetic retinopathy due to type 2 diabetes mellitus (disorder)	10
Neuropathy due to type 2 diabetes mellitus (disorder)	23
Myocardial infarction	19
Microalbuminuria due to type 2 diabetes mellitus (disorder)	5
Metabolic syndrome X (disorder)	48
Macular edema and retinopathy due to type 2 diabetes mellitus (disorder)	2
Hypertriglyceridemia (disorder)	49
Hypertension	270
Hyperglycemia (disorder)	41
History of myocardial infarction (situation)	19
History of cardiac arrest (situation)	26
Diabetic retinopathy associated with type II diabetes mellitus (disorder)	19
Diabetic renal disease (disorder)	19
Diabetes	54
Coronary Heart Disease	58
Chronic kidney disease stage 3 (disorder)	3
Chronic kidney disease stage 2 (disorder)	2
Chronic kidney disease stage 1 (disorder)	18
Cardiac Arrest	26
Body mass index 40+ - severely obese (finding)	104
Body mass index 30+ - obesity (finding)	446
Atrial Fibrillation	35
Anemia (disorder)	295

GSE25462

Raw Data

Norm Data
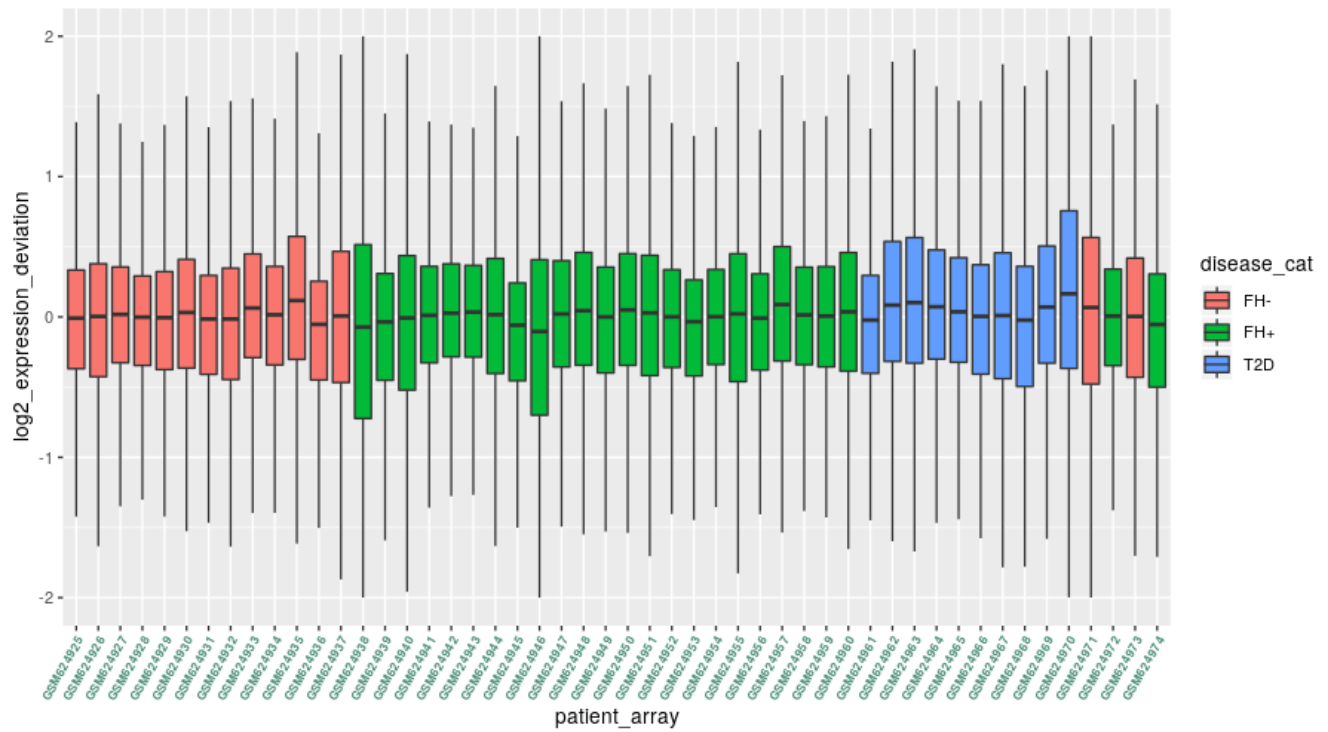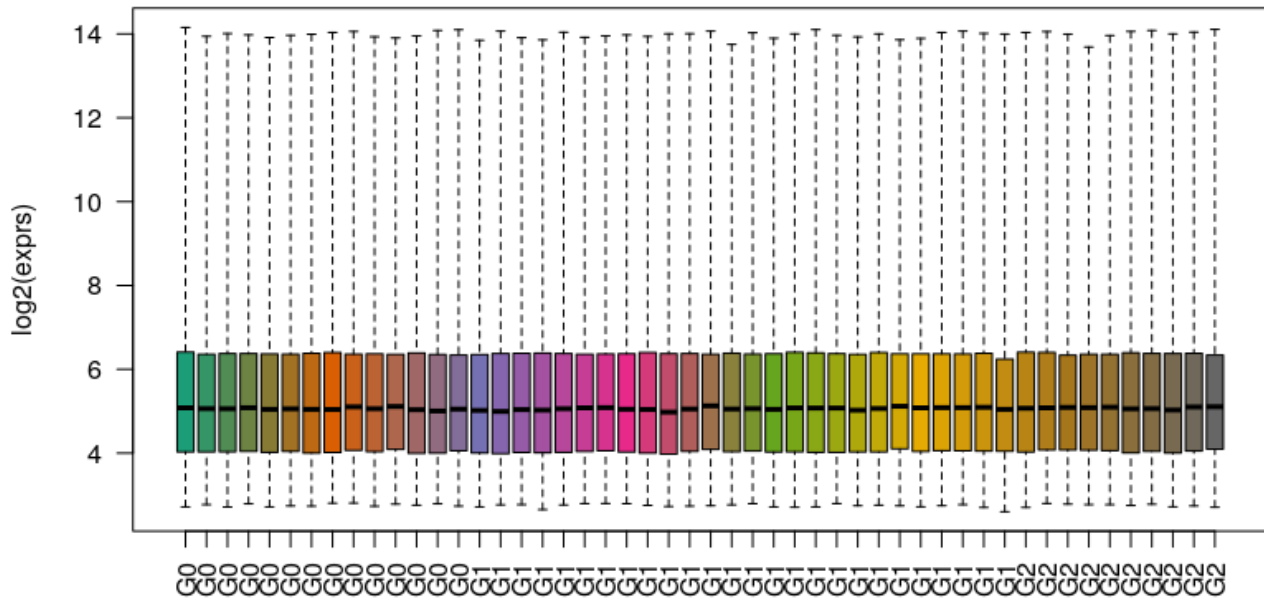
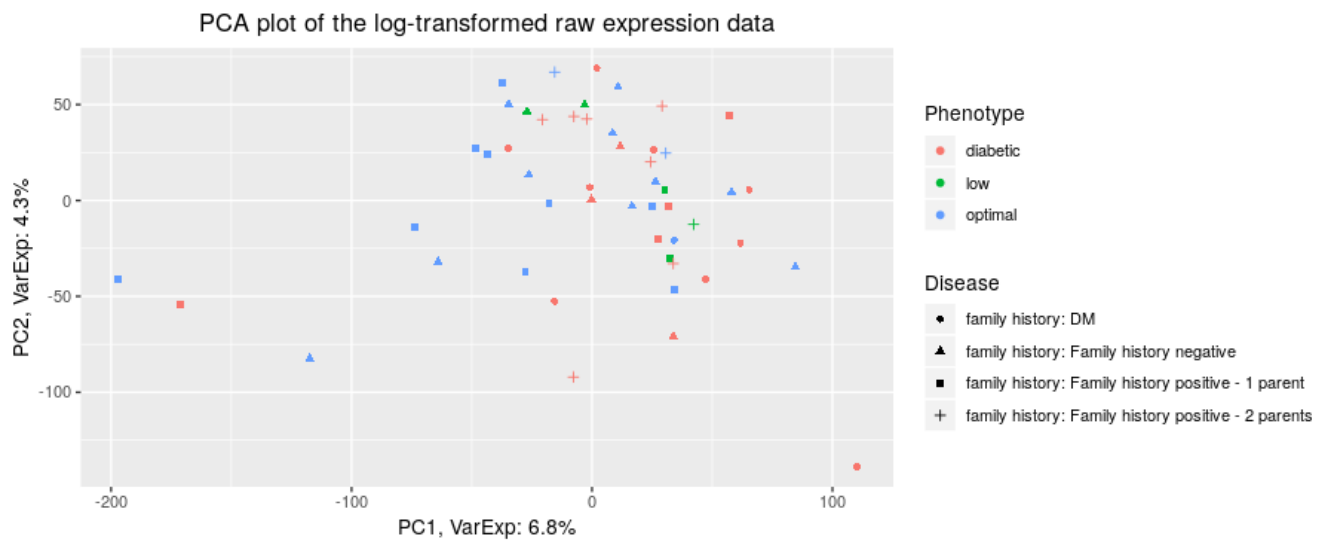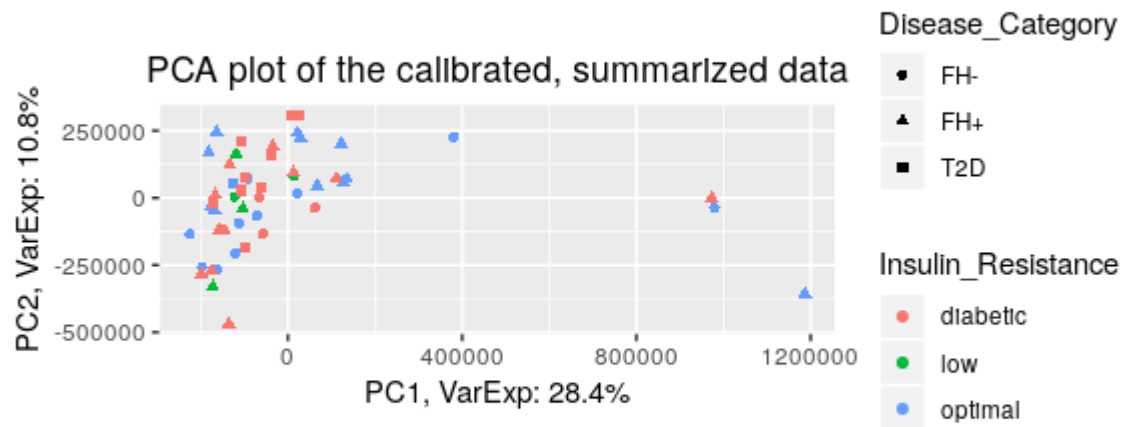**Normalised Expression Density**

Histogram of the median intensities

PCA plot of the calibrated, summarized data

Disease_Category
● FH-
▲ FH+
■ T2D

Insulin_Resistance
● diabetic
● low
● optimal

PC2, VarExp: 10.8%

PC1, VarExp: 28.4%

PCA plot of the log-transformed raw expression data

Phenotype
● diabetic
● low
● optimal

Disease
● family history: DM
▲ family history: Family history negative
■ family history: Family history positive - 1 parent
+ family history: Family history positive - 2 parents
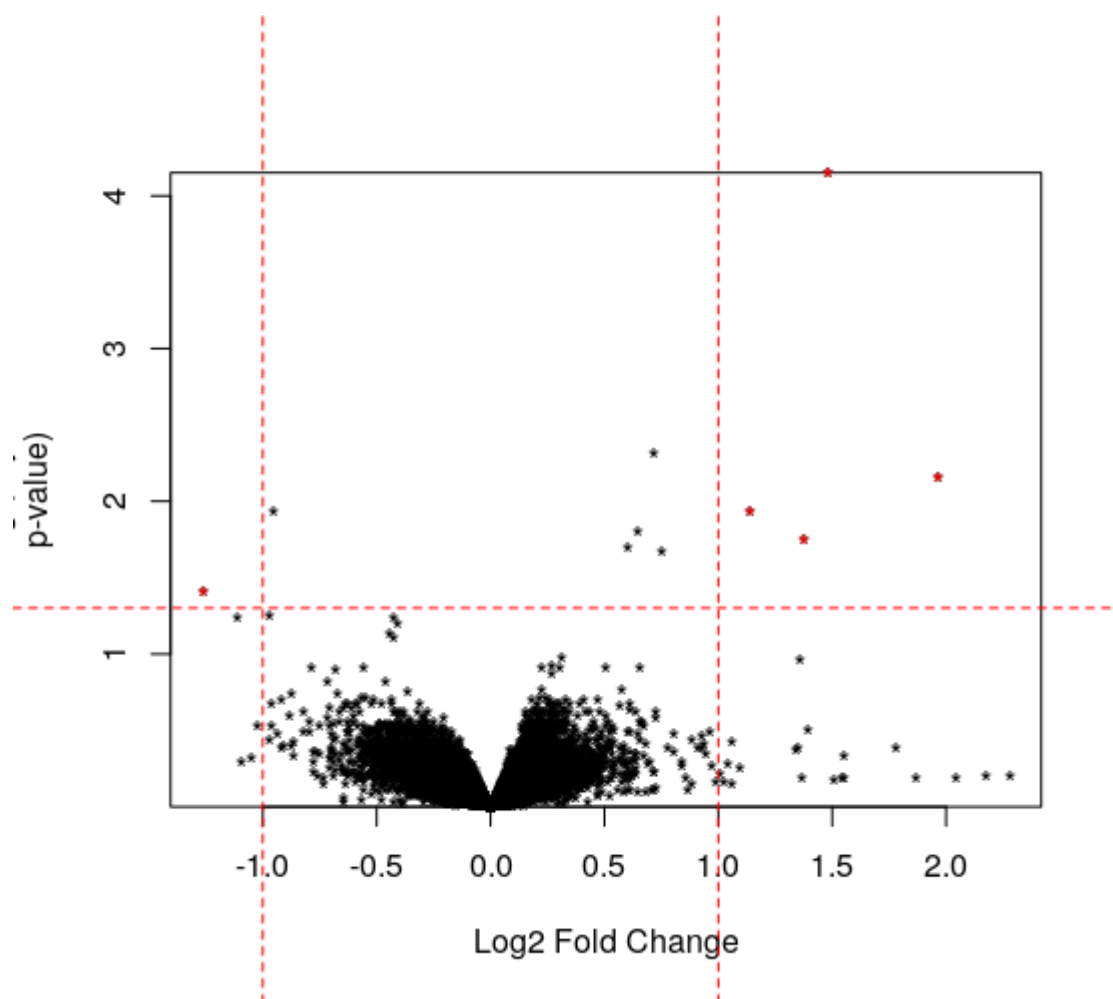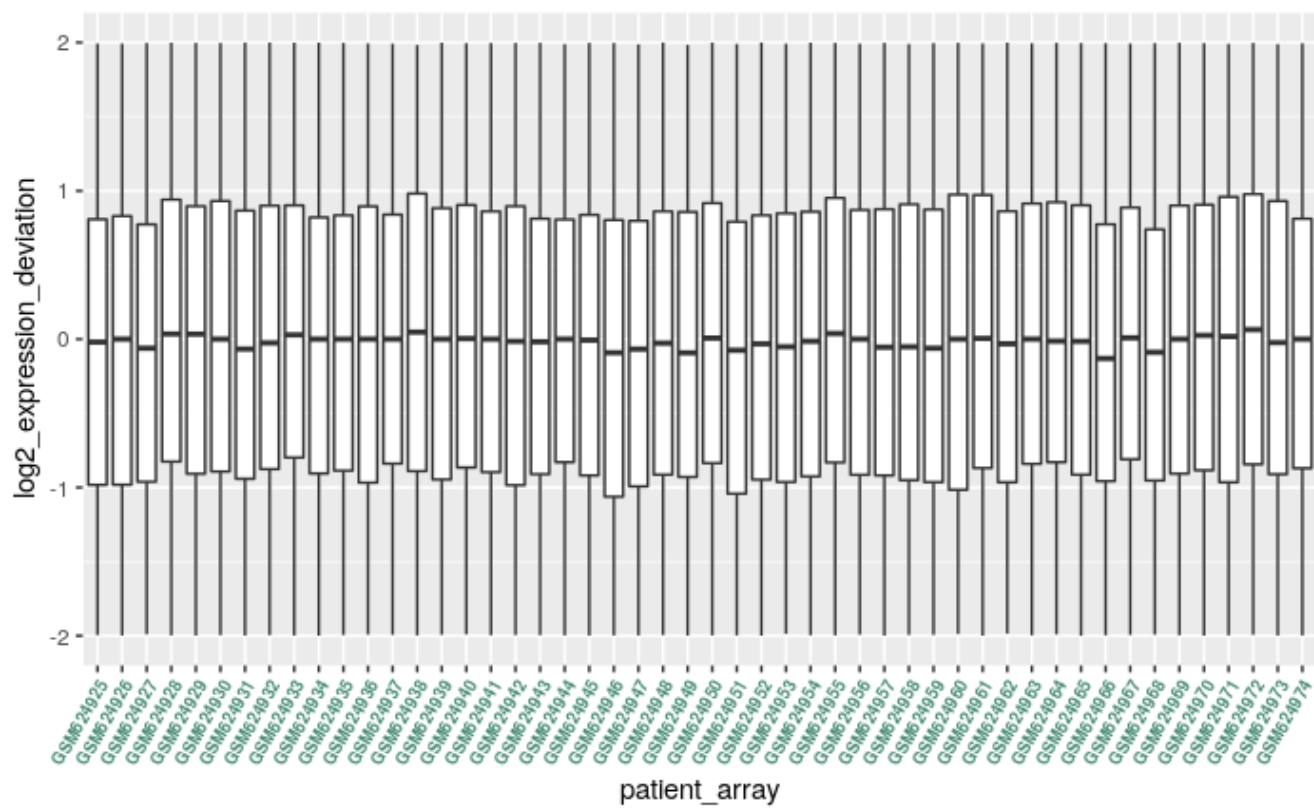
PC2, VarExp: 4.3%

PC1, VarExp: 6.8%

Clustering heatmap for the calibrated samples

large distance

small distance

GSM624938
GSM624946
GSM624971
GSM624955
GSM624970
GSM624940
GSM624973
GSM624951
GSM624966
GSM624939
GSM624974
GSM624929
GSM624964
GSM624953
GSM624961
GSM624945
GSM624949
GSM624962
GSM624963
GSM624925
GSM624972
GSM624944
GSM624947
GSM624959
GSM624935
GSM624948
GSM624969
GSM624937
GSM624967
GSM624968
GSM624926
GSM624960
GSM624927
GSM624941
GSM624942
GSM624943
GSM624928
GSM624958
GSM624956
GSM624965
GSM624933
GSM624954
GSM624950
GSM624957
GSM624936
GSM624931
GSM624934
GSM624930
GSM624932
GSM624952

1. 3D PCA

GSE115313

PCA plot of the log-transformed raw expression data

PC1, VarExp: 50.1%

PC2, VarExp: 5.4%

Phenotype
- Colon cancer Tumor
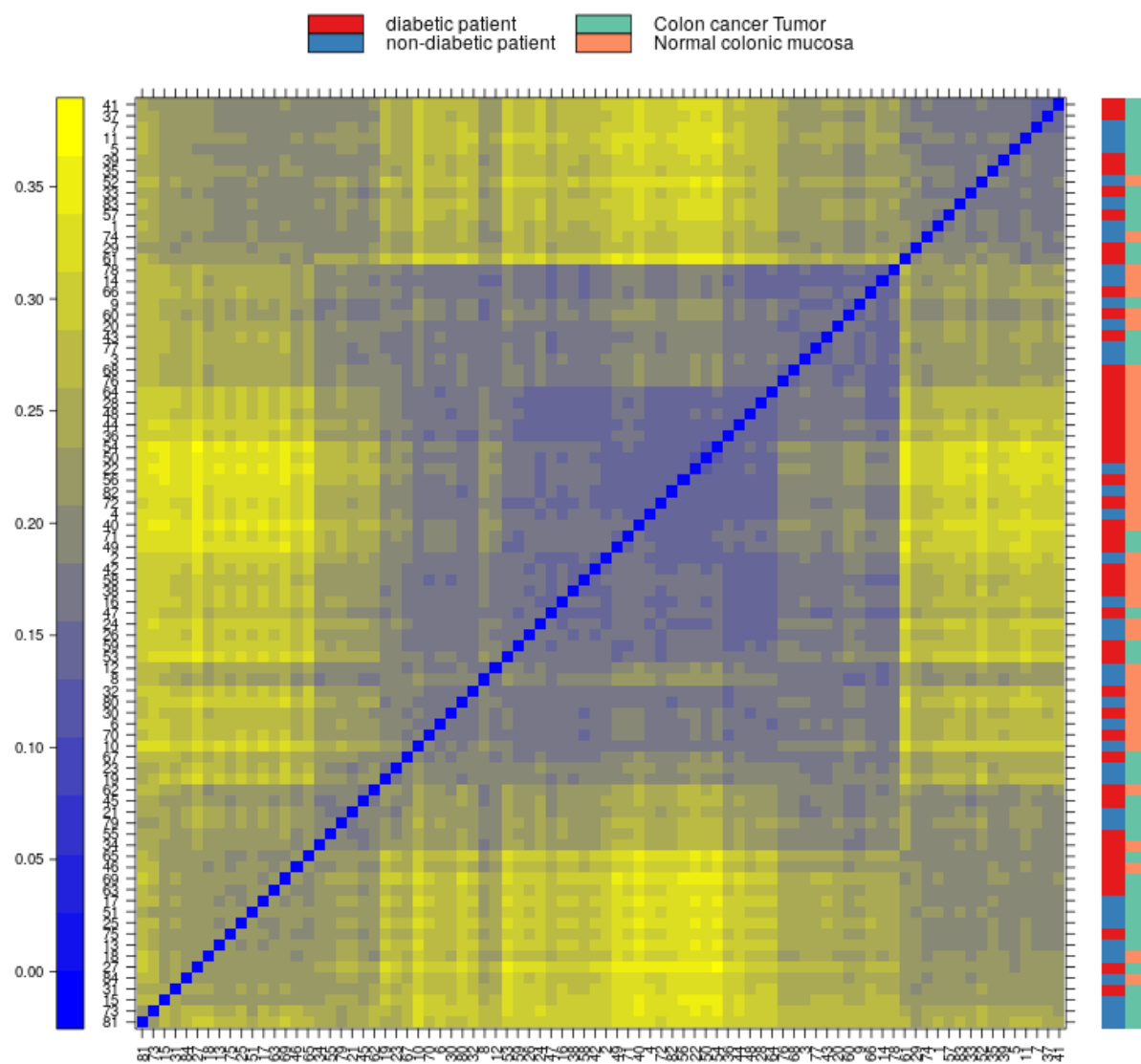- Normal colonic mucosa

Disease
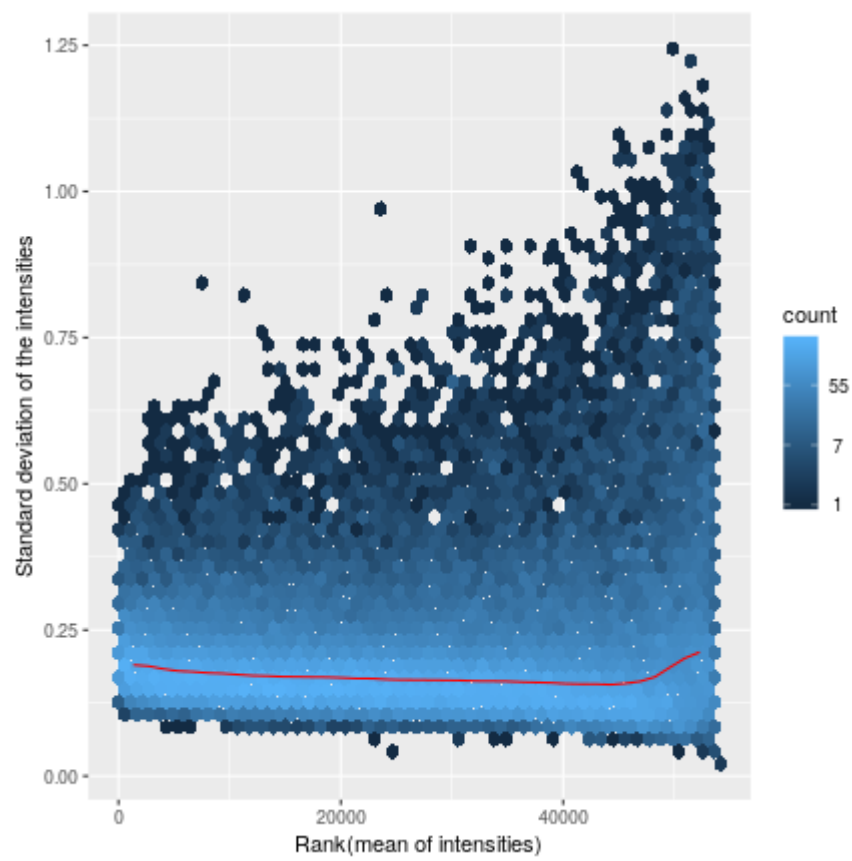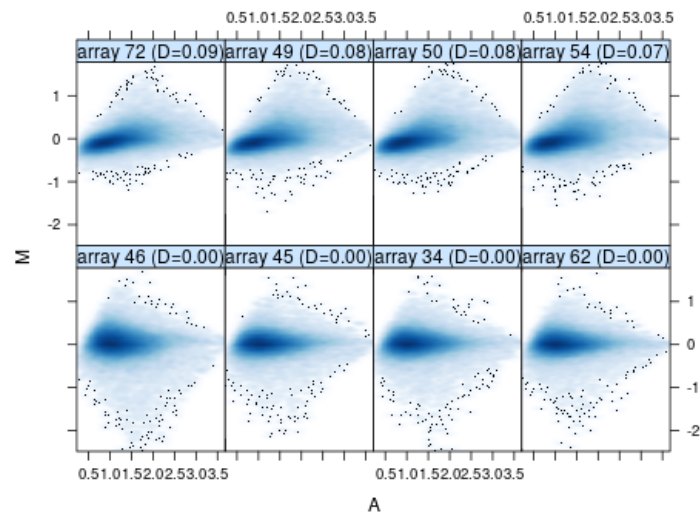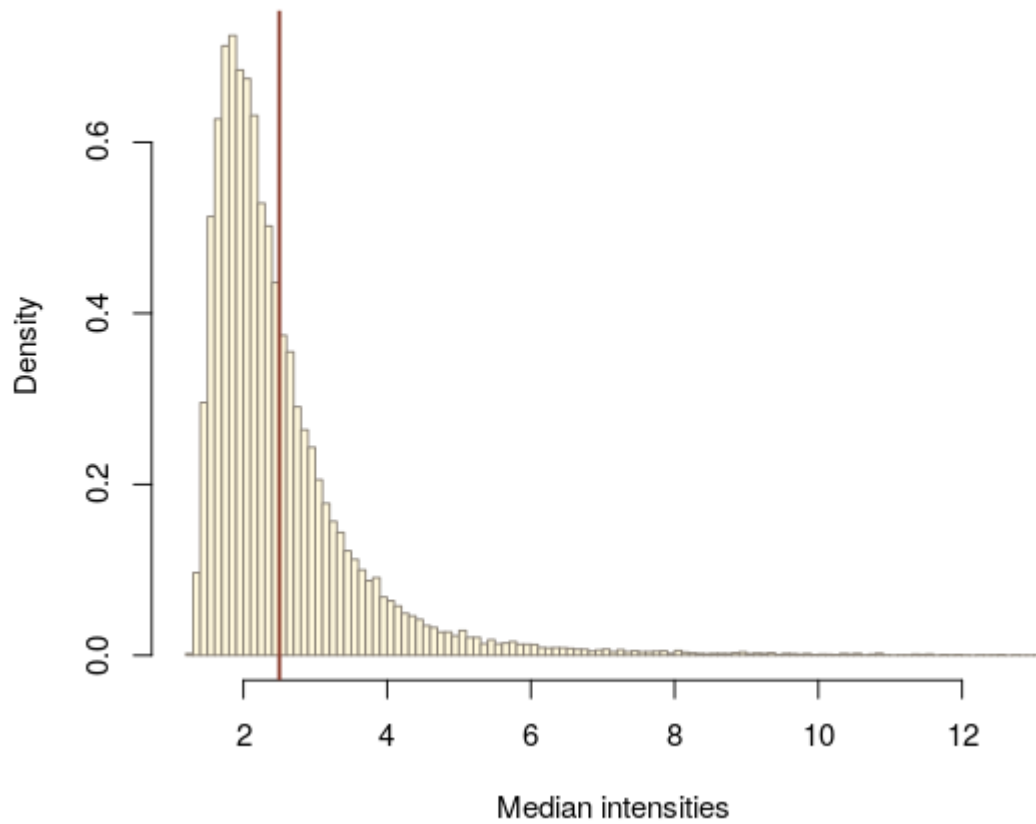- × diabetic patient
- ■ non-diabetic patient

Histogram of the median intensities

From other sources

Synthea