DATA CLEANING, PRELIMINARY SUMMARY AND EVALUATION OF DIAGNOSTIC CRITERIA OF T-CELL DATA IN A JUVENILE ONSET DIABETES COHORT

by

Yihe Huang

B.S., Minzu University of China, China, 2010

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Yihe Huang

It was defended on

Aug. 9, 2012

and approved by

Thesis Advisor:
Vincent C. Arena, PhD
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee member:
Ingrid M Libman, MD, PhD
Assistant Professor
Department of Pediatrics
Children's Hospital of Pittsburgh
University of Pittsburgh Medical Center

Francis Pike, PhD
Assistant Professor
Critical Care Medicine
School of Medicine
University of Pittsburgh

DATA CLEANING, PRELIMINARY SUMMARY AND EVALUATION OF DIAGNOSTIC CRITERIA OF T-CELL DATA IN A JUVENILE ONSET DIABETES COHORT

Yihe Huang, M.S.

University of Pittsburgh, 2012

Type 1 diabetes mellitus (T1DM) is an autoimmune disease manifested by an autoimmune attack on pancreatic beta-islet cells. T1DM can occur at any age. However, it is most often diagnosed in children, adolescents, or young adults. My thesis is derived from a large longitudinal study of Juvenile Onset Diabetes (JOD) at Children's Hospital of Pittsburgh. The objectives are: 1) Data cleaning and preliminary summary of the cohort with respect to T-cell data. 2) Evaluating the T-cell data criteria used for the prediction of the diabetes.

An extensive data examination was made for accuracy and consistency. A preliminary summary of the stimulation index (SI) for the test analytes and the number of positive antigens was performed by demographic sub-groups, HLA-DQ serotype, and follow up time. Using the ROC analysis, an evaluation of diagnosis test performance based on two different criteria was performed.

The JOD dataset had few errors with an error rate under 0.5%. The accuracy and consistency of the data is good. New onsets and first degree relatives (FDRs) nonconverters had a relatively stable SI as well as positive antigen tests results. The SI level and positive test results are higher in new onsets when compared with FDRs. FDR-converters (those subsequently developing diabetes) prior to using insulin have SIs and number of positive antigens similar to FDR-nonconverters; and FDR-converters after starting insulin have results similar to new onsets. The recommended SI cutoff of 1.5 indicating positive response appears reasonable.

However, the cutoff still may be optimized for better prediction. Evidence suggests that a lower cutoff within 1.25 to 1.5 may be better and the number of positive antigens could move from $\geq$4 to greater than 5 or 6.

Public health significance: Development of a better understanding of the pattern of T-cell response in diabetes and non-diabetic children, and those progressing to diabetes, may give us tools to predict the early onset of disease. It is this point in time where therapeutic intervention could be focused to help stem the development of T1DM or to dramatically reduce its severity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

- I would like to thank to my advisor of this thesis, Dr. Vincent Arena for his valuable guidance and advice. He inspired me greatly to work on this thesis. His willingness to motivate me contributed tremendously to my thesis.
- This thesis would not have been possible without the support of many people. My deepest gratitude is also due to the members of the thesis committee; Dr. Ingrid Libman and Dr. Francis Pike without whose knowledge and assistance in this study would not have been successful.
- I would also like to acknowledge and thank Dr. Dorothy Becker, who is principal investigator of the Juvenile Onsets Diabetes (JOD) project. I am grateful to her for allowing me to use the data from her project.
- Finally, yet importantly, I would like to express my heartfelt thanks to my beloved parents for their blessings, my friends and classmates for their help and wishes for the successful completion of this thesis.

ACKNOWLEDGEMENTS

Research reported in this thesis was supported by the National Institute of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number R01DK024021. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

1.0 INTRODUCTION

A cohort design is a study in which patients who presently have a specified condition and/or receive a particular treatment are followed over time and compared with another group who are not affected by the condition under investigation. The cohort study is a form of long-term study often used in medicine, social science, actuarial science, and ecology. In a study of juvenile onset diabetes at Children's Hospital of Pittsburgh (CHP), we have observational data on a subset of 1,591 participants who have more than four years of follow up. The purpose of this thesis are two fold; 1) Perform data cleaning to detect and correct the potential mistakes in the original data, transforming the raw data into a form that allows statistical analysis, and the generating of the preliminary summary of the t-cell data for further analysis. 2) Perform ROC analysis for 10 test-antigens and to discuss the criteria for the prediction of type 1 diabetes mellitus.

1.1 OVERVIEW OF DIABETES MELLITUS

Diabetes mellitus is a group of metabolic diseases in which the body has trouble regulating its blood sugar (glucose) levels, either from the body not producing enough insulin, or as a result of cells not responding to the insulin that is produced. Either condition produces high blood sugar and the typical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger) [1]. Diabetes mellitus occurs throughout the world, but is more

common in the more developed countries. In 2010, an estimated 285 million people had diabetes [2]. There are 25.8 million people who have diabetes in the United States, of whom 7 million people remain undiagnosed. Another 57 million people are estimated to have prediabetes [3].

Based on the known causes and risk factors of the disease, there are three major types of diabetes mellitus: Type 1 Diabetes Mellitus (T1DM), Type 2 Diabetes Mellitus (T2DM), and Gestational Diabetes Mellitus (GDM). Briefly, T1DM is an immune disorder in which the body attacks and destroys insulin-producing beta cells in the pancreas; T2DM is a metabolic disorder that is characterized by high blood glucose in the context of insulin resistance and relative insulin deficiency. This is in contrast to diabetes mellitus type 1 in which there is an absolute insulin deficiency; GDM is a form of glucose intolerance during pregnancy without a previous diagnosis of diabetes.

1.2 TYPE 1 DIABETES MELLITUS

T1DM can occur at any age, but it is most often diagnosed in children, teens, or young adults, so T1DM was originally referred to as Juvenile-Onset Diabetes. T1DM causes an estimated 5–10% of all diabetes cases worldwide [4]. Incidence varies from eight to 17 per 100,000 in Northern Europe and the U.S., with a high of about 35 per 100,000 in Scandinavia, to a low of one per 100,000 in Japan and China [5].

The exact cause of T1DM is unknown. Most likely it is induced by one or more of the following: genetic susceptibility, a diabetogenic trigger and/or exposure to a driving antigen [6].

1.3 HLA AND HLA – DQ TYPE IN T1DM

1.3.1 Human leukocyte antigen (HLA)

Human leukocyte antigens (HLA) are proteins that are present on the surface of our bodies' cells. Depending on the HLA proteins displayed on the surface of our cells, the immune system recognizes our own cells as “self”, as opposed to “foreign”. In humans, HLA also is called Major Histocompatibility Complex (MHC). The HLA types are genetically determined; the corresponding genes to HLA proteins are located on chromosome 6. Different clusters of genes form three major classes of HLA based on their functions [7-8]:

Class I antigens (A, B & C) present peptides from inside the cell (including viral peptides if present). The proteins produced from these genes are present on the surface of almost all cells. MHC class I proteins display these peptides to the immune system. If the immune system recognizes the peptides as foreign (such as viral or bacterial peptides), it responds by triggering the infected cell to self-destruct.

Class II antigens (DP, DM, DOA, DOB, DQ, and DR) corresponding to MHC class II genes, which provide instructions for making proteins that are present almost exclusively on the surface of certain immune system cells. These proteins also display peptides to the immune system like MHC class I proteins.

Class III antigens corresponding to MHC class III genes have somewhat different functions; they are involved in inflammation and other immune system activities. The functions of some MHC genes are unknown. (See Figure 1)

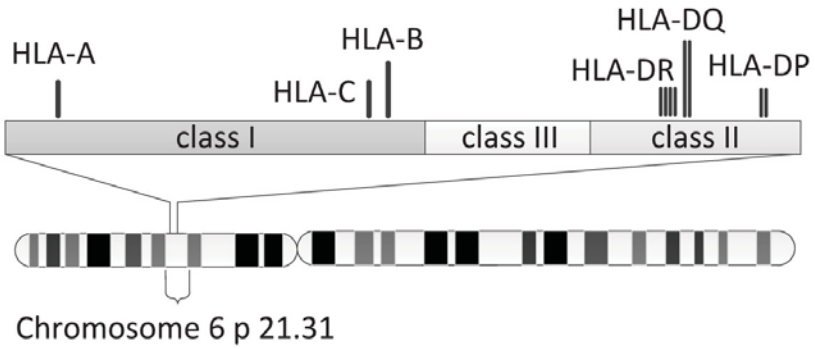


Figure 1. HLA region on Chromosome 6

(by Xie et al.[9])

1.3.2 HLA-DQ

HLA-DQ (DQ) is a cell surface type protein found on antigen presenting cells. It has an α chain and β chain which are coded by HLA-DQA1 and HLA_DQB2 genes, respectively. The two chains form a αβ heterodimer, where two chains vary greatly. The variance of subunit in α chain and β chain result in lots of different isoforms of DQ in human. Compared with α chain, the β chain has more forms of subunits and it helps us to discriminate the serotype of DQ. The current serotype of DQ include HLA-DQ2, -DQ3, -DQ4, -DQ5, -DQ6, -DQ7, -DQ8, -DQ9. DQ help the immune system to recognize and present foreign antigens on the cells' surface. It is believed that DQ also recognizes and presents self-antigens in order to develop tolerance for the immune system. However, the absent of tolerance to self-proteins incurs the autoimmune disease associated with DQ. For instance, DQ is involved in celiac disease [10] and T1DM.

Recent studies have demonstrated that some HLA-DQ genes are most strongly associated with T1DM susceptibility [11-14]. A combination of DQ2 and DQ8 genes would increase the risk of adult onset Type 1 Diabetes and ambiguous type I/II Diabetes [15][16].

1.3.3 Autoantibodies

Autoantibodies are antibodies (immune proteins) that falsely target and damage specific tissues or organs of the body. One or more autoantibodies may be produced by a person's immune system when it fails to distinguish between "self" and "foreign" proteins. That is to say, the immune system disorder has a certain association with autoantibodies when the immune system ceases to recognize the self-components. Specific autoantibodies are usually present in a percentage of people with a particular autoimmune disorder, so several autoantibodies tests can be performed to predict or diagnose the autoimmune disorder. One of the most common immunologic markers of individuals with autoimmune diabetes is the presence of autoantibodies against beta-cell autoantigens.

Diabetes-related (islet) autoantibody testing is primarily ordered to help distinguish between autoimmune type 1 diabetes and diabetes as a result of other causes. There are four common autoantibody tests used to distinguish between type 1 diabetes and diabetes resulting from other causes; ICA (Islet Cell Cytoplasmic Autoantibodies), recognizing islet cytoplasmic antigens, were detected many years ago in newly-diagnosed type 1 diabetic patients [17]. ICA can be present in 90% of type 1 diabetes patients at the time of diagnosis [18]. GADA (Glutamic Acid Decarboxylase Autoantibodies) has been found in approximately 50-80% of newly diagnosed type 1 diabetic patients [19-23]. IA-2A (Insulinoma-Associated-2 Autoantibodies) is found in about 55-80% of newly diagnosed type 1 diabetic patients [19,22,24]. IAA (Insulin Autoantibodies) is found in about 40-70% of newly diagnosed type 1 diabetic children.

ICA was previously widely used to study the clinical course and pathogenesis of type 1 diabetes. However, as the detection methods developed, ICA was replaced by GADA and IA-2A

in a large extent. Recently, autoantibodies against ZnT8 (ZnT8A) were discovered as an additional marker for type 1 diabetes [25].

2.0 MATERIALS AND METHODS

The juvenile onsets diabetes (JOD) cohort is a longitudinal cohort of new onsets with T1D and their first degree relatives (FDRs) initially without T1D. This study has been ongoing for over 30 years. New onsets are identified upon presentation with T1DM to Children's Hospital of Pittsburgh (CHP). The data for this thesis is derived from a subset of new onsets and their FDRs enrolled from January 2004 through June 2008, and who have results of T-cell data.

2.1 STUDY DATASET CREATION

Three datasets contained various information about the study participants. Information was extracted and merged into a single study dataset according the routine in Figure 2.

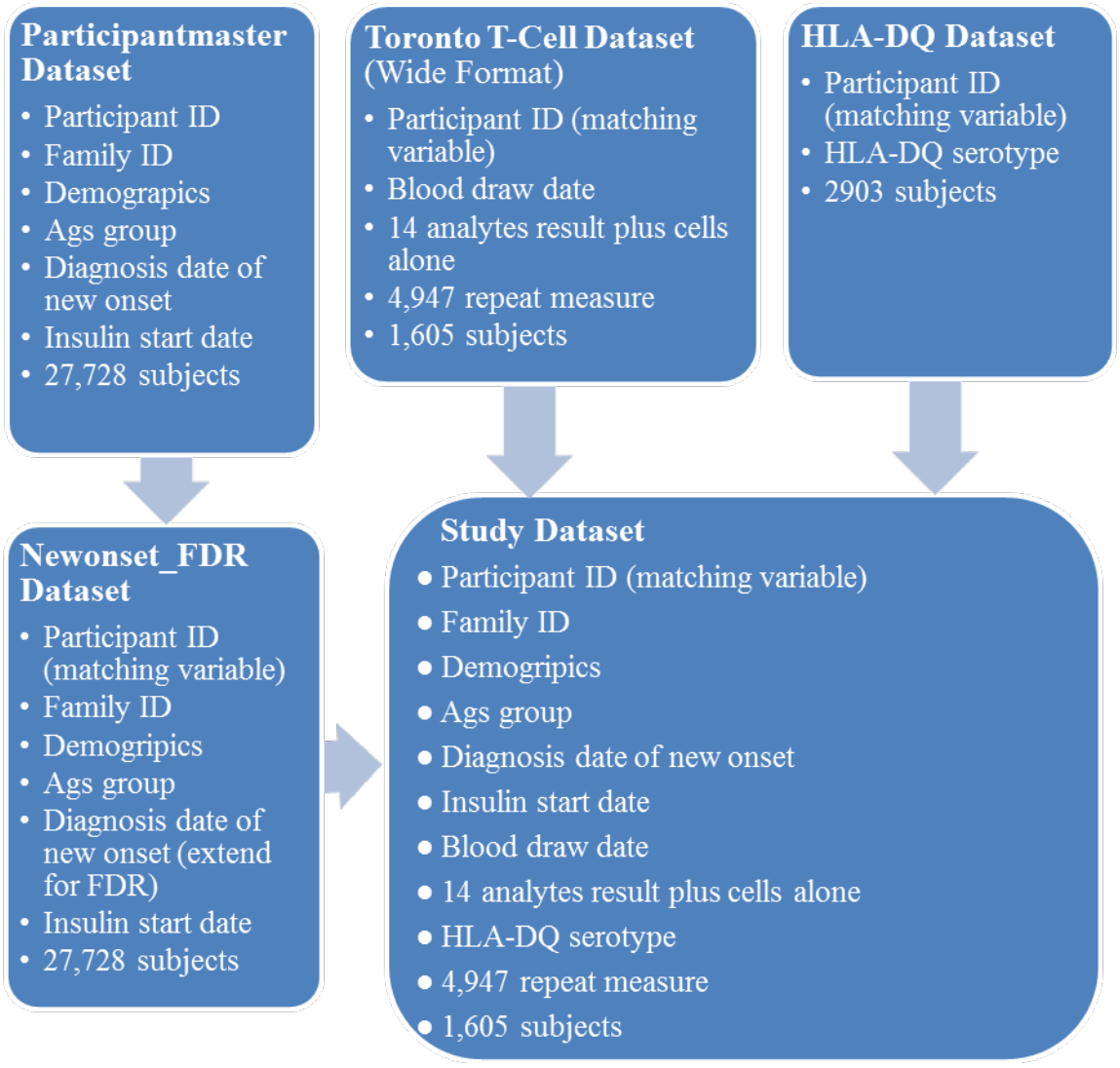


Figure 2. Study dataset creation

2.2 NEW VARIABLES CREATION

2.2.1 Important analytes and terms from original data

H.M. Dosch et al., introduced an assay system to make the detection of 14 antigens associated with T1DM [26]. Fourteen anlytes and cells alone were measured from the participants' blood draw samples (Table 1), including 10 test antigens, 2 positive controls and 2 negative controls. Specific terms and criteria displayed in Table 2.

Table 1. Description for cells alone and 14 analytes

Analytes	Contents	Useage
CNSA_MBP	Myelin Basic Protein (MBP), a myelin autoantigen commonly targeted in T1D & MS, (5µg/ml)	test antigen
CNSA_EX2	A major MBP autoimmune target epitope peptide (splice variant) expressed in developing myelin and remyelination only (5µg/ml)	test antigen
GLIA_GFAP	Glial Fibrillary Acidic Protein, A glial autoantigen often targeted in T1D & MS (0.5µg/ml)	test antigen
GLIA_S100	A glial autoantigen often targeted in T1D and MS (0.5µg/ml)	test antigen
ISLA_Gad	Glutamic Acid Decarboxylase, a T1D target autoantigen (5µg/ml)	test antigen
ISLA_Gad55	A GAD epitope peptide often targeted in T1D, also recognized by Nepom MHC class II tetramer (5µg/ml)	test antigen
ISLA_PI	Pro-Insulin (1µg/ml)	test antigen
ISLA_Tep69	An ICA69 autoantigen peptide often targeted in T1D, 2 way mimicry antigen crossreactive with ABBOS (5µg/ml)	test antigen
MIP_Abbos	ABBOS - a milk epitope peptide commonly targeted in T1D, 2-way mimicry antigen ICA69 (5µg/ml)	test antigen
MIP_BSA	Bovine Serum Albumin - cow milk protein (5µg/ml)	test antigen
PS_PHA	T-cell mitogen (Phytohemagglutinin) (1µg/ml)	positive control, proliferation competence
PS_TT:	Tetanus Toxin (0.1µg/ml)	positive control, post-vaccination response competence
OVA	Ovalbumin - dietary protein (5µg/ml)	negative control
CSA_Actin	Human Actin (5µg/ml)	negative control
Cells alone	Back ground	control

Table 2. Specific terms and criteria from original data for cells alone and 14 analytes

Term	Contents and Criteria
Well1	counts per minute (cpm)
Well2	counts per minute (cpm)
WellMean	mean cpm of Well1 and Well2
SI	Stimulation index (mean cpm experimental/ mean cpm cells alone)
Result	For test antigens, SI >= 1.5 is Positive, SI < 1.5 is Negative
Interpretation	For any given sample, with >= 4 positive responses to test antigens=Affected Less than 4 positive responses to test antigens = Unaffected

2.2.2 Data cleaning variables creation

The original dataset contained many hand calculations and as a result, potential mistakes could exist. In order to address this issue, it is important to recalculate and verify these values. Sequences of variables were generated to verify these calculations:

1. Based on formula WellMean= (Well1+Well2)/2, for each of the analytes tested, a "wellmean_chk" variable was generated to check these calculations.
2. Based on formula SI=WellMean/ (WellMean in cells alone), for each of analytes tested, a "SI_chk" variable was generated to check these calculations.

3. Based on the criterion "SI>=1.5 indicates a positive result for test antigen, SI<1.5 indicates a negative result for the test antigen"; a "Result_chk" variable was generated for each of the test antigen.
4. Based on the criterion that "for any given sample, with 4 or more positive responses to the test antigens indicates an affected sample, less than 4 positive responses to the test antigens indicates the sample unaffected"; an "Interpretation_chk" variable was generated for each blood draw test.

2.2.3 Preliminary summary and ROC analysis variables creation

Several variables were created to perform preliminary summaries and ROC analysis.

Age calculated at the participants' first blood draw.

Time window "win_ddx" was created based on the diagnosis date of T1DM for new-onsets. We stratified the time interval from date of diagnosis to each blood draw date into several month categories. For the first degree relatives (FDRs) of the new onsets would we used their corresponding new onsets' diagnosis date to generate these time windows. FDRs without new onsets records were coded as missing.

Time windows "win_bdd" was created to stratify the time interval from first blood draw date to subsequent follow-up blood draw dates and grouped them into several month categories.

Based on the blood draw date, a time sequencing variable "order" was created to indicate the progression of blood draw times in participants.

2.3 THE ROC METHOD

The accuracy of antigens test to discriminate T1DM cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis.

Usually we consider the results of an antigen test in two populations, one population with T1DM, the other population without T1DM; you will rarely observe a perfect separation between the two groups. Inevitably, the distribution of the test results will overlap, as depicted in the following table.

Table 3. Distribution of diagnosis test result

	T1DM Present	T1DM Absent	
Test +	a	c	a + c
Test -	b	d	b + d
	a + b	c + d	

Several indices can be derived from Table 3.

Sensitivity, also known as true positive rate (TPR), is the probability of a positive test result when the disease is present. Sensitivity = a / (a+b). Sensitivity relates to the test's ability to identify diseased individuals. If a test has high sensitivity, a positive result would more likely indicate the presence of disease.

Specificity, also known as true negative rate, is the probability of a negative test result when the disease is not present. Specificity = d / (c+d). Specificity relates to the ability of the test

to identify disease free individuals. If a test has high specificity, a negative result would more likely suggest the absence of disease.

Positive predictive value (PPV) is probability that the disease is present when the test is positive. PPV= a / (a+c). A high PPV would help a given test to conform the presence of disease when a positive result is yielded.

Negative predictive value (NPV) is probability that the disease is not present when the test is negative. NPV= d / (b+d). A high NPV would help a given test to conform the absence of disease when a negative result is yielded.

Note that PPV and NPV are both depended on prevalence [27], PPV and NPV should only be used if the ratio of the number of patients in the disease group and the number of patients in the healthy control group used to establish the PPV and NPV is equivalent to the prevalence of the diseases in the studied population. Otherwise, positive and negative likelihood ratios would be more accurate than PPV and NPV due to the likelihood ratios do not depend on prevalence.

LR+ is a ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease, LR+ = True positive rate / False positive rate = Sensitivity / (1-Specificity).

LR- is a ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease, LR- = False negative rate / True negative rate = (1-Sensitivity) / Specificity.

The greater the value of LR+, the more likely a positive test result is from a diseased individual. The greater the value of LR-, the more likely a negative test result is from a disease free individual.

The ROC curve has long been used to depict the tradeoff between true positive rates and false positive rate in signal detection theory [28]. Given different cutoff points, sensitivity is plotted with accompanying false positive rate (1- specificity) in a ROC curve (Figure 3).

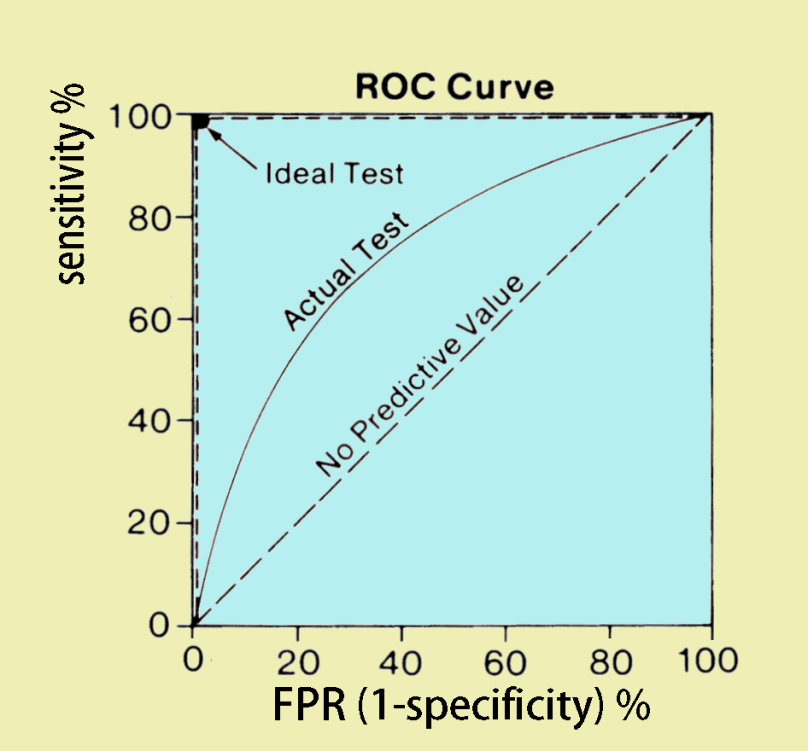


Figure 3. Example of a ROC curve

From Figure 3, we see that each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. Ideally, a test with perfect discrimination would have a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [29].

The area under the ROC curve (AUC) is one way to quantitate the goodness or accuracy of diagnosis test in discriminating between 2 states of health. A diagnosis test with no predictive value would have an AUC of 0.5 (also represented by the diagonal "chance" line above), while a diagnosis test with perfect discrimination would have an AUC of 1. The empirical AUC is

calculated via the "trapezoidal" rule, where a trapezoid is constructed from the lines drawn for each two consecutive points on the curve (Figure 4). The AUC would refer to the sum of areas of the trapezoids. Mathematically, the AUC may be defined as

$$AUC = \int_0^1 Sn(t)dt$$

where t is the false positive rate, Sn(t) is the corresponding sensitivity [30].

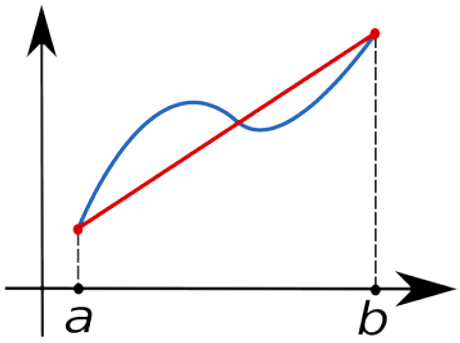


Figure 4. Illustration of trapezoidal rule

A general guide for determining the accuracy of a diagnostic test is the traditional academic point system [31]: (Table 4)

Table 4. Accuracy classification by AUC for a diagnostic test

Range of AUC	Classification
0.90-1	Excellent
0.80-0.90	Good
0.70-0.80	Fair
0.60-0.70	Poor
0.50-0.60	Fail

The AUC of a diagnosis is often compared to chance which has an AUC of 0.5. The statistical test involves estimating AUC_{test} - AUC_{chance} which is asymptotically normal. With Gaussian techniques, we can derive a p-value under null hypothesis that the difference equals to zero.

For the same study population, two diagnosis tests can also be compared using paired sample statistical techniques. The method exploits the mathematical equivalence of the AUC to the Mann-Whitney U-statistic [32]. According to this method, the comparison for ROCs of any two diagnosis test can be made by evaluating the difference of the AUCs which consider as asymptotically normal.

Statistical Analysis System (SAS) version 9.2 was used for all data cleaning, descriptive statistics, test statistics, and graphics.

3.0 ANALYSIS AND RESULTS

3.1 DATA CLEANING

3.1.1 Checking missing value

One of the first and most important steps in any data processing task is to verify that your data values are correct or, at the very least, conform to some set of rules. After the creation of the study dataset, it is important to look through the whole dataset to find missing values before we process with the recalculations.

In the study dataset, there are 31 observations whose well1 and well2 values for all 15 analytes are missing. (See Table 18) These missing values are the result of a bad sample, or hemolysis of the sample. One observation sample only missed 1 record in CSA_Actinwell1 or CSA_Actinwell2. The reason for this missing datum was unknown.

3.1.2 Recalculation for WellMean

There are 6 observations which contained miscalculated WELLMEAN. (See Table 19 in Appendix A)

3.1.3 Recalculation for SI

There are 29 observations which contained miscalculated SI (See Table 20 in Appendix A).

3.1.4 Checking analytes' result with corresponding SI

For the 10 test antigens: CNSA_EX2, CNSA_MBP, GLIA_GFAP, GLIA_S100, ISLA_Gad, ISLA_Gad55, ISLA_PI, ISLA_Tep69, MIP_Abbos, MIP_BSA, the result should be positive, negative, or missing.

If the dataset was correct, the SI value should match with its corresponding result. When the SI value of test antigen was >= 1.5, the result should be positive, or it was considered incorrect; When the SI value of test antigen < 1.5, the result should be negative, or it was considered incorrect. The result was considered as missing when the SI value was not reported.

There are 20 mismatched results in total; 6 falsely reported as negative, and 14 falsely reported as positive (See Table 21, Table 22 in Appendix A).

3.1.5 Checking analytes' interpretation

All analytes were interpreted as being affected, unaffected, or missing.

A check of the analytes' interpretation with corresponding number of positive test antigens was made.

If the number of positive test antigens >= 4, they were interpreted as affected, or falsely unaffected.

If the number of positive test antigens < 4, they were interpreted as unaffected, or falsely affected.

The interpretation was considered as missing when results were not reported.

There are 23 mismatched interpretations in total; 18 false-unaffected interpretations, and 5 false-affected (See Table 23, Table 24 in Appendix A).

3.2 DEMOGRAPHICS

The entire study population included 1,591 participants, 43.0% (685) of entire participants were males, and 57.0% (906) of entire participants were females. The complete population was made up of 95.8% (1,504) Whites, 2.3% (52) Blacks, 0.3% (4) Asians, and 0.6% (10) other or multiple races. There were 316 probands or new onsets, 1,275 first degree relatives (FDRs). FDRs were further subdivided into, FDR-converters and FDR-nonconverters. The FDR-converters are FDRs who converted/developed insulin dependent diabetes; the other, FDR-nonconverters are those FDRs who have not developed diabetes. The FDRs included 20 FDR-converters and 1,255 FDR-nonconverters. Table 5 shows the distribution of gender in New Onsets, FDR-converters and FDR-nonconverters.

Table 5. Gender Distribution for Entire Population

Gender	New Onsets	FDR-converters	FDR-nonconverters	Total
	% (Counts)	% (Counts)	% (Counts)	% (Counts)
Male	59.5 (188)	40.0 (8)	39.0 (489)	43.0(685)
Female	40.5 (128)	60.0 (12)	61.0 (766)	57.0(906)
Total	100 (316)	100 (20)	100 (1255)	100(1591)

Table 6 shows the race distribution in the study population, all FDR-converter consisted of white, there is no Asian in New onsets.

Table 6. Race Distribution for Entire Population

Race	New Onsets	FDR-converters	FDR-nonconverters	Total
	% (Counts)	% (Counts)	% (Counts)	% (Counts)
White	92.4 (292)	100 (20)	96.6 (1,192)	95.8(1504)
Blcak	5.7 (18)	0(0)	2.8 (34)	3.3(52)
Asian	0(0)	0(0)	0.3 (4)	0.3(4)
Other*	1.9 (6)	0(0)	0.3 (4)	0.6(10)
Total	100 (316)	100 (20)	100 (1234)**	100(1570)

* Including multiple races

** Frequency Missing = 21

Table 7 shows the age distribution for the whole study population. There are 2 New Onset patients and 4 FDR-converters with ages greater than 20 years of age at the time of their first blood draws.

Table 7. Age Distribution for Entire Population

Age	Mean	Std	Median	Min.	Max.
New Onsets	9.4	4.38	10	1	31
FDR-converter	14.3	11.47	10.5	2	41
FDR-nonconverter	28.6	13.95	34	0	61

Std = standard deviation

3.3 HLA-DQ SEROTYPE IN STUDY DATASET

We classified HLA-serotype into four categories, X, DQ2, DQ8, and DQ2/DQ8. X indicates the participants' HLA-serotype is neither DQ2 nor DQ8. DQ2, DQ8 indicates the people who only have either DQ2 or DQ8 serotype for their HLA. DQ2/DQ8 indicates the people have both DQ2 and DQ8 serotype. The greatest risk of T1DM is found among the DQ2/DQ8individuals, followed by the DQ8 individuals then the DQ2 individuals and finally the lowest risk is X serotype.

Table 8. HLA-DQ Distribution for Subpopulation

HLA-DQ Type	New Onsets	FDR-converters	FDR-nonconverters	Total
	% (Counts)	% (Counts)	% (Counts)	% (Counts)
X	17.5 (51)	18.7 (3)	32.6 (399)	29.6 (453)
DQ2	28.6 (83)	25.00 (4)	31.3 (383)	30.7 (470)
DQ8	35.2 (102)	43.8 (7)	30.5 (374)	31.6 (483)
DQ2/DQ8	18.7 (54)	12.50 (2)	5.6 (69)	8.1(125)
Total	100 (290)*	100 (16)**	100 (1225)***	100(1531)

* Frequency Missing = 26

** Frequency Missing = 4

*** Frequency Missing = 30

3.4 NUMBER OF POSITIVE ANTIGENS

The number of positive antigens is one of the most important criteria in predicting T1DM. The recommended cutpoint for predicting T1DM is 4; in constrast <4 should predict no T1DM.

3.4.1 Number of positive antigens in characteristics of subjects by subpopulation

The number of positive antigens was assessed among three subpopulations: New onsets, FDR-converters, and FDR-nonconverters. From Table 9, we found that new onsets group had a higher

number of positive antigens than FDR. The variance among FDR-converters is higher when compared with other two subpopulations. The FDR-converter group can be further divided into two subgroups by insulin start date. The FDR-converter who has a blood draw before starting insulin should have a similar number of positive antigens with the FDR-nonconverters. After using insulin, the FDR-converter would be expected to have similar number of positive antigens with the new onsets subjects. (Table 10)

Table 9. Number of positive antigens in demographics by subpopulation

Number of positive antigens	New Onsets					FDR-converter					FDR-nonconverter				
	Mean	**Std**	**Median**	**Q1**	**Q3**	**Mean**	**Std**	**Median**	**Q1**	**Q3**	**Mean**	**Std**	**Median**	**Q1**	**Q3**
Gender															
Male	8.6	2.8	10.0	9.0	10.0	2.8	4.3	0.0	0.0	8.0	0.9	2.5	0.0	0.0	0.0
Female	8.6	2.8	10.0	9.0	10.0	5.1	4.8	7.0	0.0	10.0	0.9	2.6	0.0	0.0	0.0
Race															
White	8.6	2.7	10.0	9.0	10.0	4.3	4.7	0.0	0.0	10.0	0.9	2.5	0.0	0.0	0.0
Black	8.3	3.3	10.0	9.0	10.0	.	.	.	.	.	1.5	3.2	0.0	0.0	0.0
Asian	.	.	.	.	.	.	.	.	.	.	0.0	0.0	0.0	0.0	0.0
Other	8.2	3.4	10.0	8.0	10.0	.	.	.	.	.	3.1	4.4	0.0	0.0	8.0
HLA-DQ type															
X	8.7	2.8	10.0	9.0	10.0	4.6	5.0	4.0	0.0	9.5	0.9	2.6	0.0	0.0	0.0
DQ2	8.5	2.9	10.0	8.0	10.0	1.0	3.0	0.0	0.0	0.0	0.8	2.4	0.0	0.0	0.0
DQ8	8.3	3.0	10.0	8.0	10.0	5.7	4.6	8.0	0.0	10.0	1.0	2.7	0.0	0.0	0.0
DQ2/DQ8	9.2	2.2	10.0	9.0	10.0	4.2	5.0	0.0	0.0	10.0	0.9	2.6	0.0	0.0	0.0
All	8.6	2.8	10.0	9.0	10.0	4.3	4.7	0.0	0.0	10.0	0.9	2.5	0.0	0.0	0.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 10. Number of positive antigens in demographics by subgroup of FDR-converter

	After insulin start					Before insulin start				
Number of positive antigens	**Mean**	**Std**	**Median**	**Q1**	**Q3**	**Mean**	**Std**	**Median**	**Q1**	**Q3**
Gender										
Male	6.8	4.3	9.0	3.0	10.0	1.5	3.6	0.0	0.0	0.0
Female	8.0	3.4	10.0	7.0	10.0	2.9	4.6	0.0	0.0	9.0
Race										
White	7.7	3.6	10.0	7.0	10.0	2.4	4.2	0.0	0.0	1.0
Black	.	.	.	.	.	.	.	.	.	.
Asian	.	.	.	.	.	.	.	.	.	.
Other	.	.	.	.	.	.	.	.	.	.
HLA-DQ type										
X	9.3	1.0	9.5	8.5	10.0	0.0	0.0	0.0	0.0	0.0
DQ2	3.3	5.8	0.0	0.0	10.0	0.1	0.4	0.0	0.0	0.0
DQ8	8.1	3.1	10.0	7.0	10.0	3.3	4.8	0.0	0.0	9.5
DQ2/DQ8	.	.	.	.	.	4.2	5.0	0.0	0.0	10.0
All	7.7	3.6	10.0	7.0	10.0	2.4	4.2	0.0	0.0	1.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Figure 5 and Figure 6 show the change in the number of positive antigens in three subpopulations over the follow up time. New onsets have the highest number of positive antigens and the FDR-nonconverters have the lowest number of positive antigens. The red line in Figure 5 and Figure 6, represents FDR-converters before using insulin and FDR-converters after using insulin, respectively. Figure 5 and Figure 6 also show that before the FDR-converters started using insulin, their number of positive antigens had a lower value; but the number of positive antigens increases shortly after using insulin.

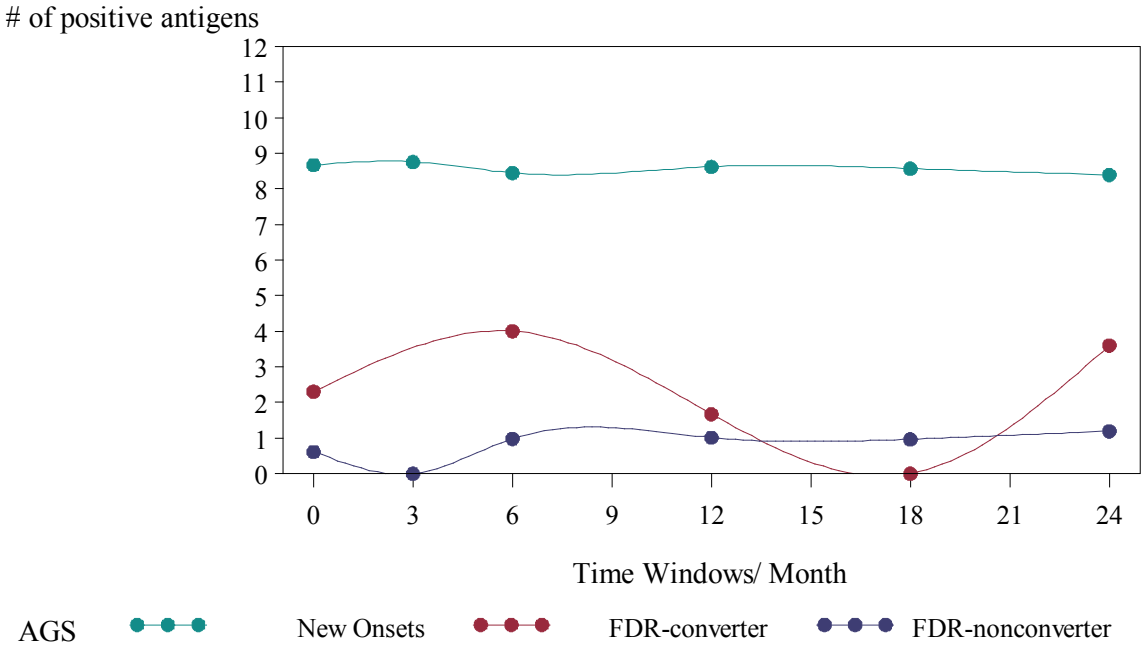


Figure 5. Comparison of number of positive antigens in New Onsets, FDR-converters before using insulin, and FDR-nonconverters over folloing time

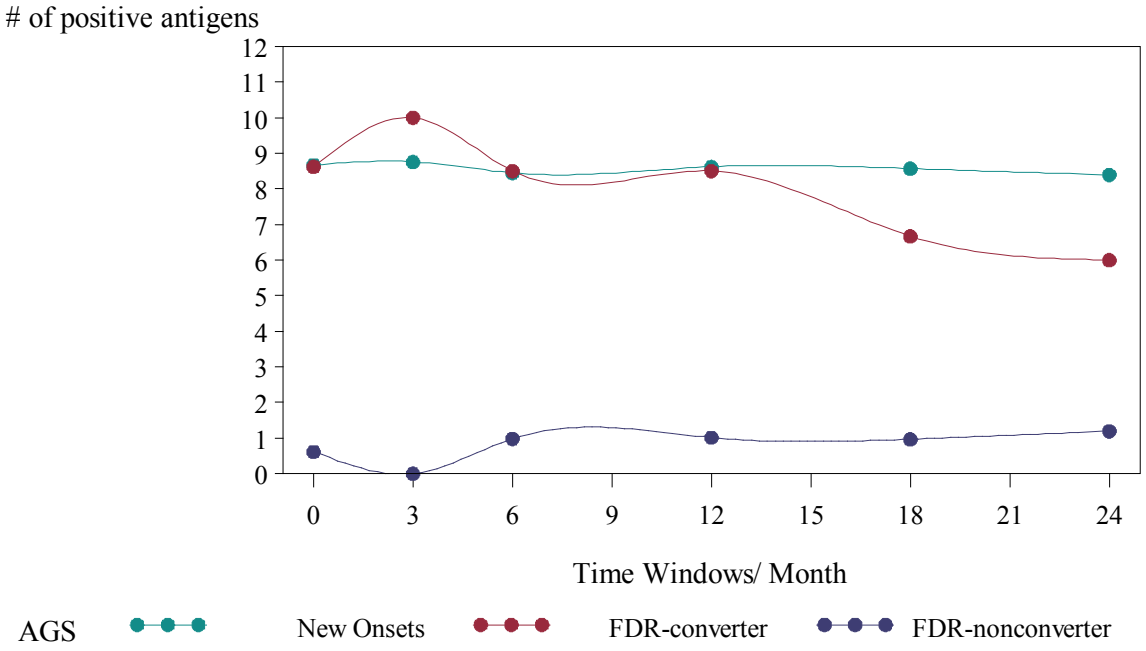


Figure 6. Comparison of number of positive antigens in New Onsets, FDR-converters after using insulin, and FDR-nonconverters over folloing time

3.4.2 T1DM affected status by subpopulations

According to the criterion, T1DM affected statuses were discriminated by four or more positive antigens, general affected status shown in Table 11. A p-value less than 0.0001 indicate that the affected status is significant among three subpopulations.

Table 11. T1DM status by subpopulations

Frequency Col Pct	New Onsets	FDR-converter	FDR-nonconverter	Total
Affected	286 91.4	9 45.0	79 6.3	374
Unaffected	27 8.6	11 55.0	1171 93.7	1209
Total	313	20	1250	1583

Frequency Missing = 8, P-value <0.0001

From Figure 7, it can be seen that the new onsets have a relatively higher percentage for affect rates than FDR-nonconverters who tend to have a lower affect rates.

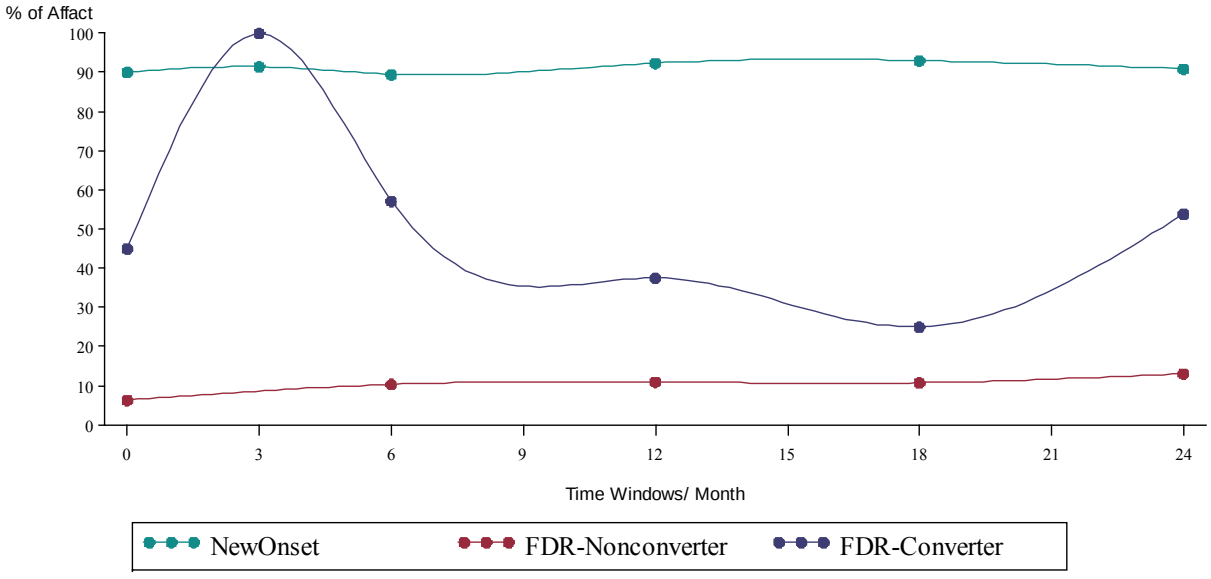


Figure 7. Percentage of T1DM affected in three subpopulations over time

3.5 TEN TEST ANTIGENS

The ten test antigens are also important predictors of T1DM. The prior recommended cutoff for SI was 1.5 for all test antigens. For test antigens with SI >=1.5, the test result would be considered as positive. For test antigens with SI<1.5, the test result would consider as negative.

A series of tables in Appendix B.1 shows the summary of SI of ten test antigens in characteristics of subjects by subpopulations. The new onsets have higher level of SI when compare with FDR.

The series of figures in Appendix B.1.2 (P. 79) shows the change of ten test antigens' SI over follow-up time by subpopulation. The SI in new onsets and FDR-nonconverters is relatively stable over time. New onsets and FDR-converters who used insulin had a higher SI level when compared with an FDR-converter who never used insulin, or an FDR-nonconverter.

3.6 EVALUATION OF DIAGNOSIS TEST

For the evaluation of the diagnostic test, we used the new onsets and the FDr-nonconverters. We included a total of 1,571 subjects for further evaluation. The blood test results were based on a subject's first blood draw, which would provide a baseline performance of the diagnosis test.

The ROC curve is a straightforward way to evaluate the diagnosis test. Two criteria were considered in this section to predict T1DM.

3.6.1 Criterion using test antigens

Ten test antigens were measured based on their SI in blood draw test. From prior recommendations, 1.5 would be the cutoff to discriminate between a positive or negative antigen. For an SI>=1.5, the test result would be positive, whilst an SI<1.5, the test result would be deemed as being negative. According to this criterion, we created Table 12. Generally, the ten test antigens gave us good test results with high sensitivity and specificity. Based on the cutoff of 1.5, we plotted the ROC curves for the ten test antigens. (Figure 8- Figure 17) The optimal cutpoint would be closer to the ideal point of (0, 1), which implies that the prediction of test antigens are good.

Table 12. Effect of SI on predict T1DM*

	2x2Table			*Sensitivity*	*Specificity*	*1-Specificity*	*PPV*	*NPV*
CNSA_EX2		A+	A-					
	Test +	251	55	80.2%	95.6%	4.4%	82.0%	95.1%
	Test -	62	1195					
CNSA_MBP		A+	A-					
	Test +	255	56	81.5%	95.5%	4.5%	82.0%	95.4%
	Test -	58	1194					
GLIA_GFAP		A+	A-					
	Test +	271	64	86.6%	94.9%	5.1%	80.9%	96.6%
	Test -	42	1186					
GLIA_S100		A+	A-					
	Test +	265	70	84.7%	94.4%	5.6%	79.1%	96.1%
	Test -	48	1180					
ISLA_Gad		A+	A-					
	Test +	279	79	89.1%	93.7%	6.3%	77.9%	97.2%
	Test -	34	1171					
ISLA_Gad55		A+	A-					
	Test +	278	74	88.8%	94.1%	5.9%	79.0%	97.1%
	Test -	35	1176					
ISLA_PI		A+	A-					
	Test +	284	102	90.7%	91.8%	8.2%	73.6%	97.5%
	Test -	29	1148					
ISLA_Tep69		A+	A-					
	Test +	281	90	89.8%	92.8%	7.2%	75.7%	97.3%
	Test -	32	1160					
MIP_Abbos		A+	A-					
	Test +	285	83	91.1%	93.4%	6.6%	77.4%	97.7%
	Test -	28	1167					
MIP_BSA		A+	A-					
	Test +	287	88	91.7%	93.0%	7.0%	76.5%	97.8%
	Test -	26	1162					

*A+ =New Onsets, A- =FDR-nonconverter

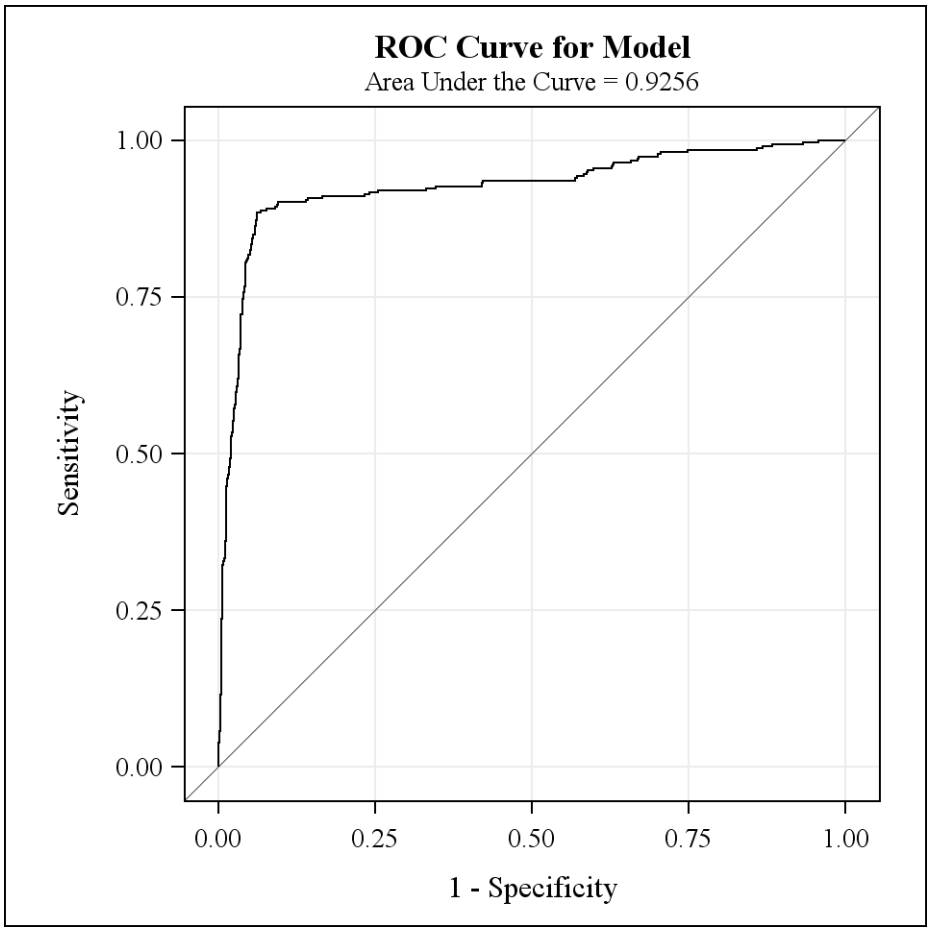


Figure 8. ROC curve for CNSA_EX2

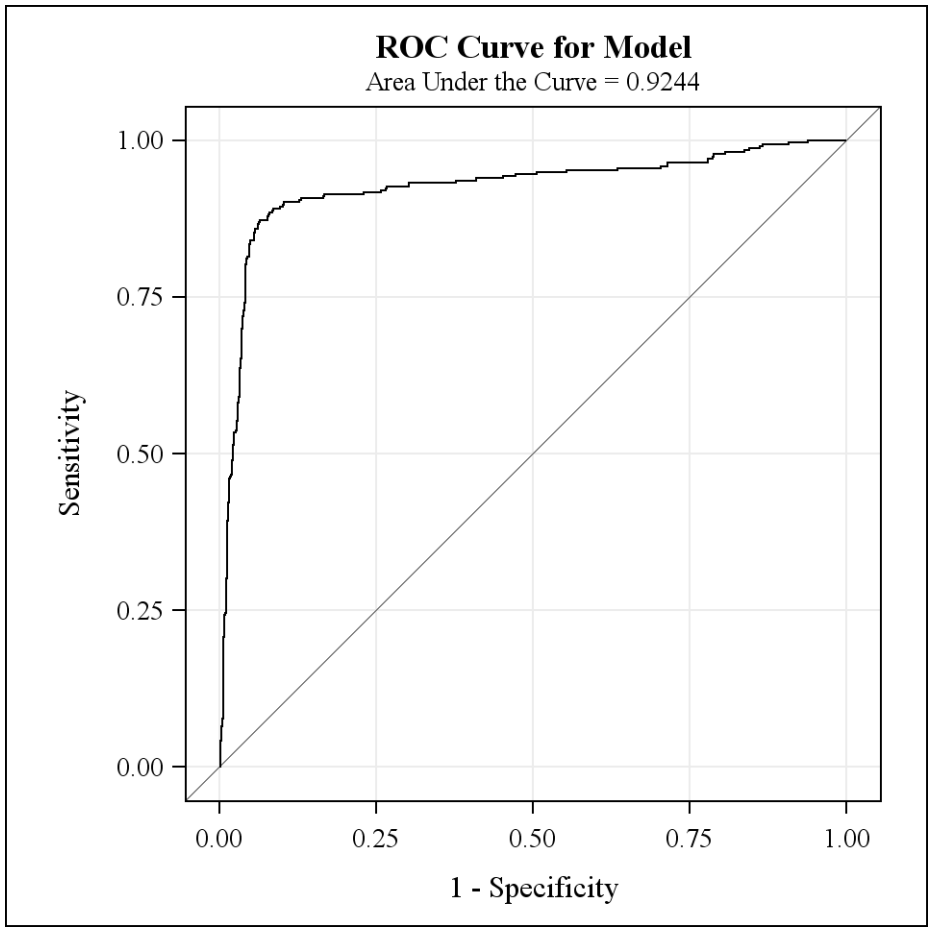


Figure 9. ROC curve for CNSA_MBP

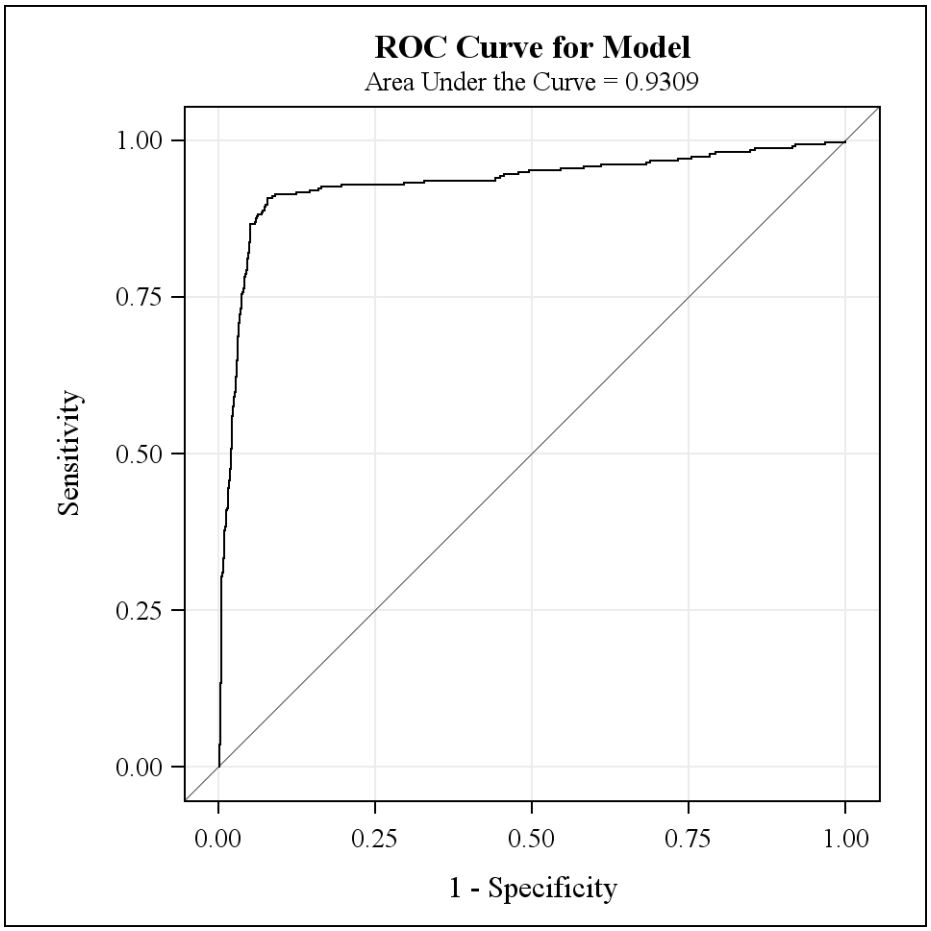


Figure 10. ROC curve for GLIA_GFAP

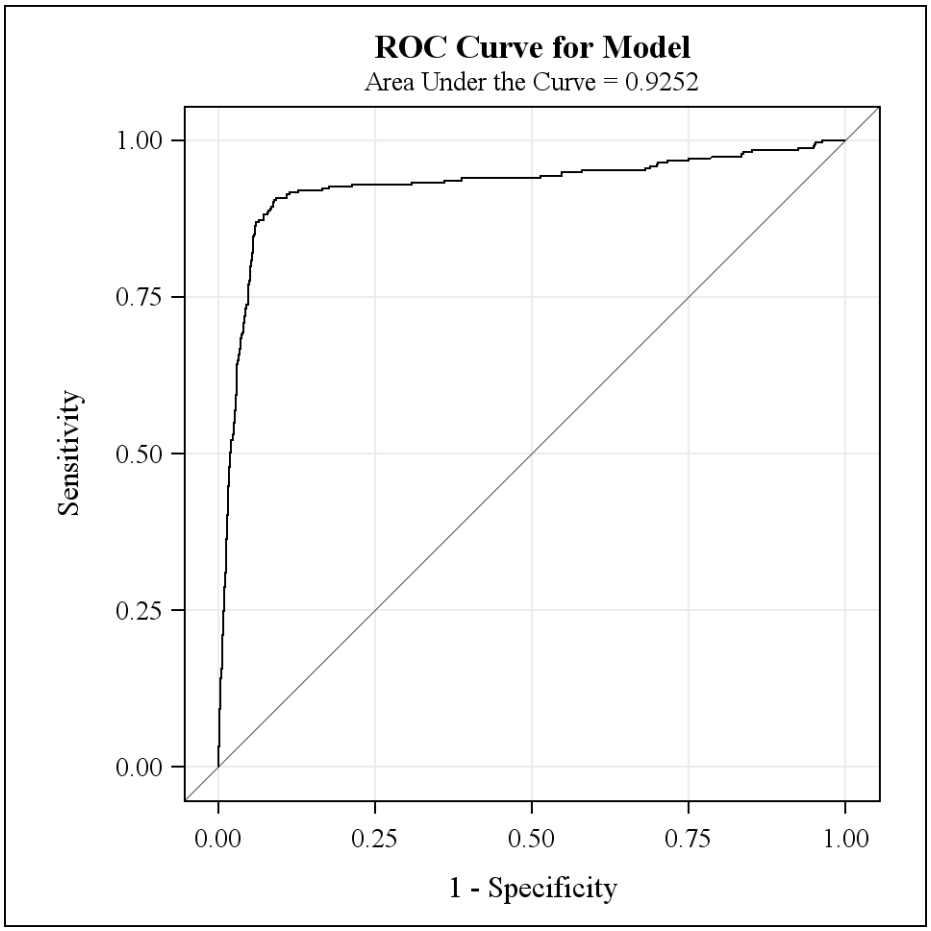


Figure 11. ROC curve for GLIA_S100

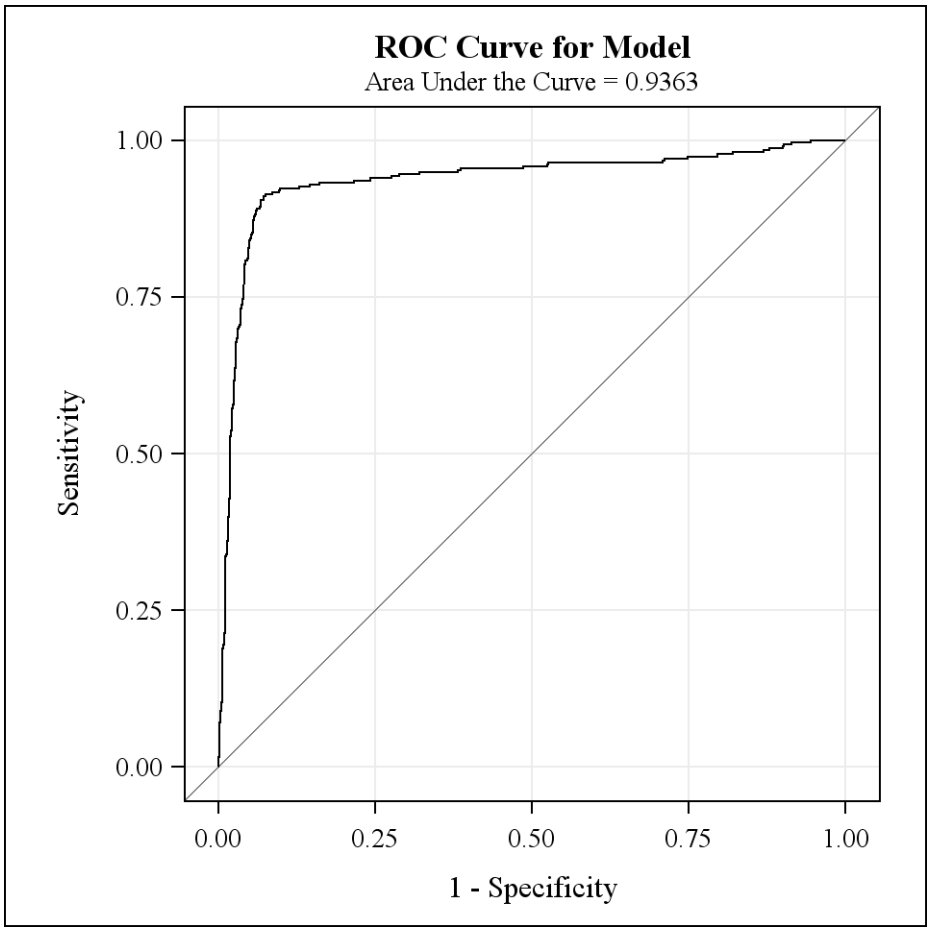


Figure 12. ROC curve for ISLA_Gad

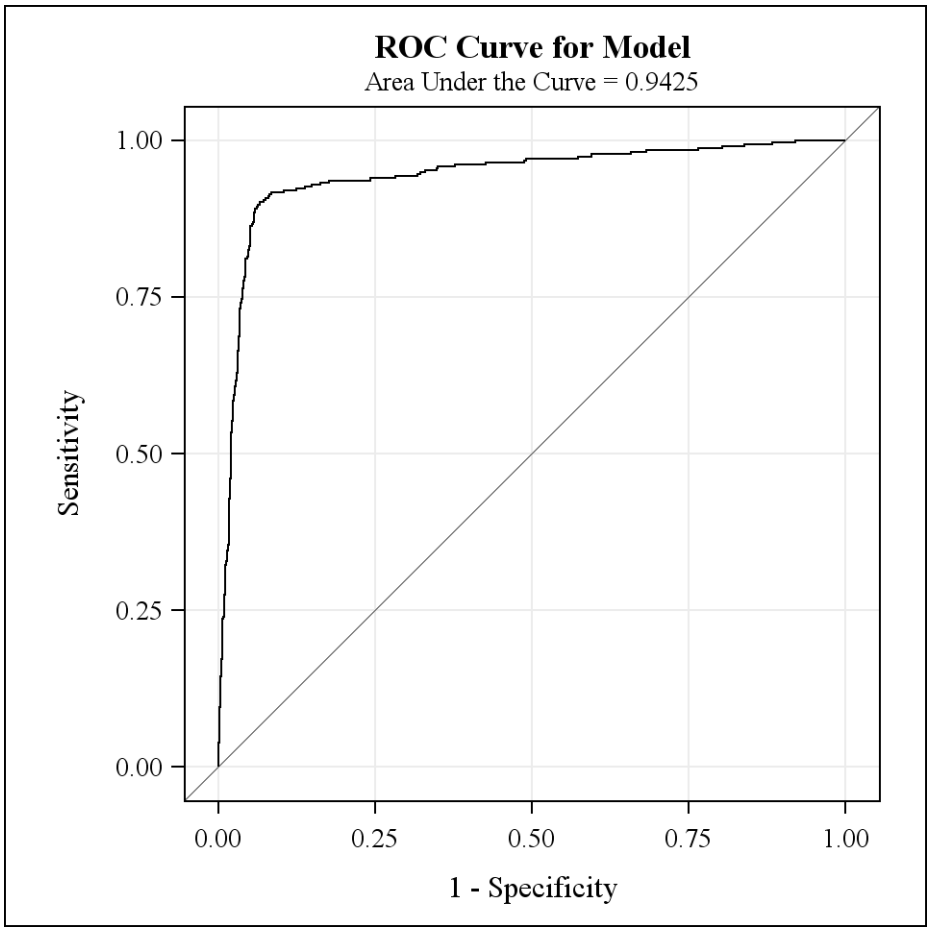


Figure 13. ROC curve for ISLA_Gad55

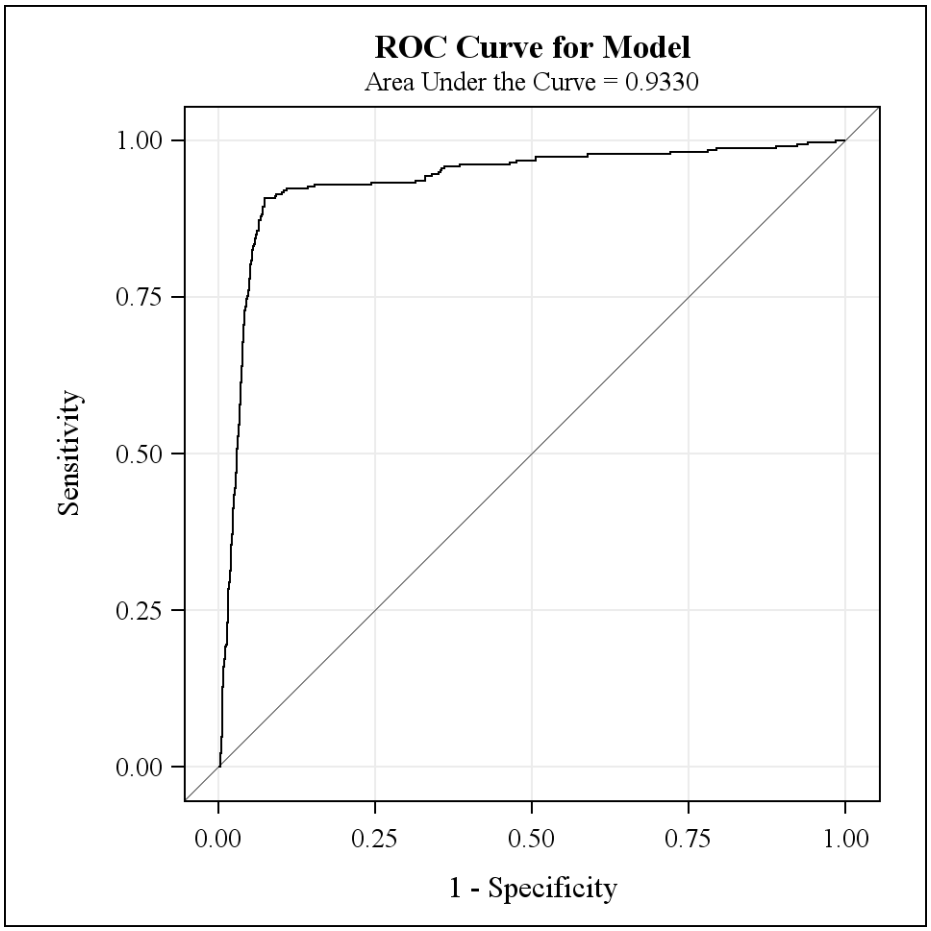


Figure 14. ROC curve for ISLA_PI

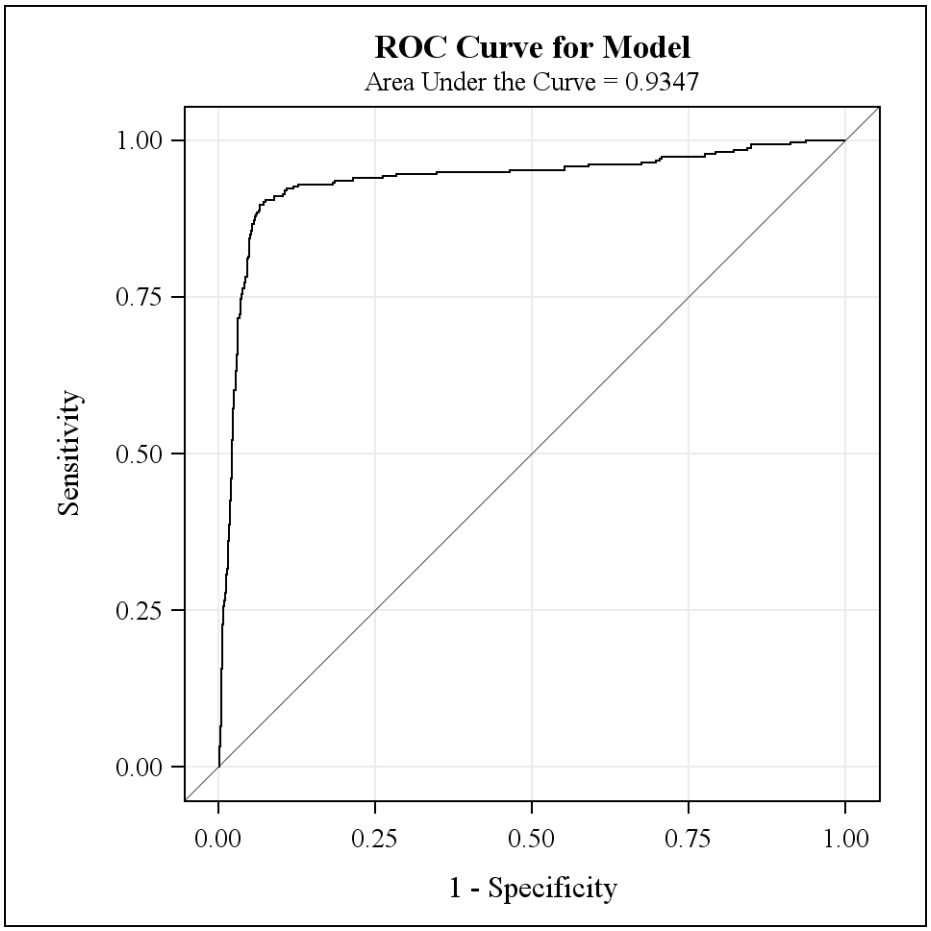


Figure 15. ROC curve for ISLA_Tep69

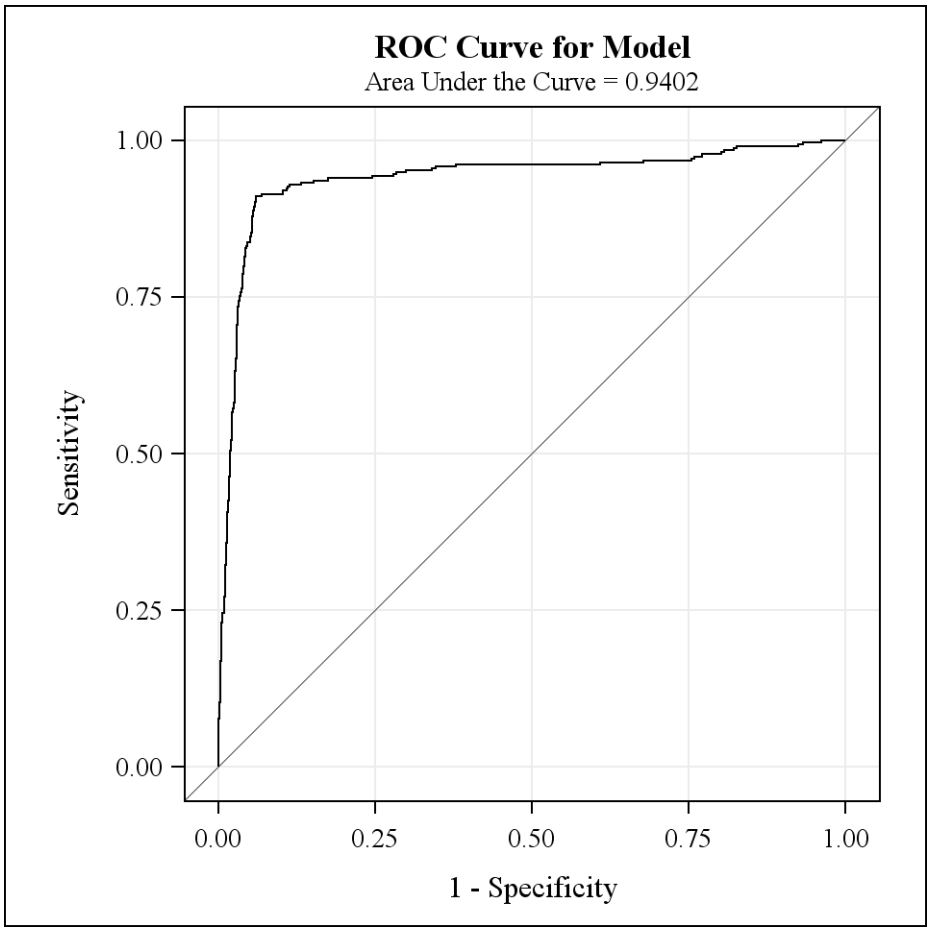


Figure 16. ROC curve for MIP_Abbos

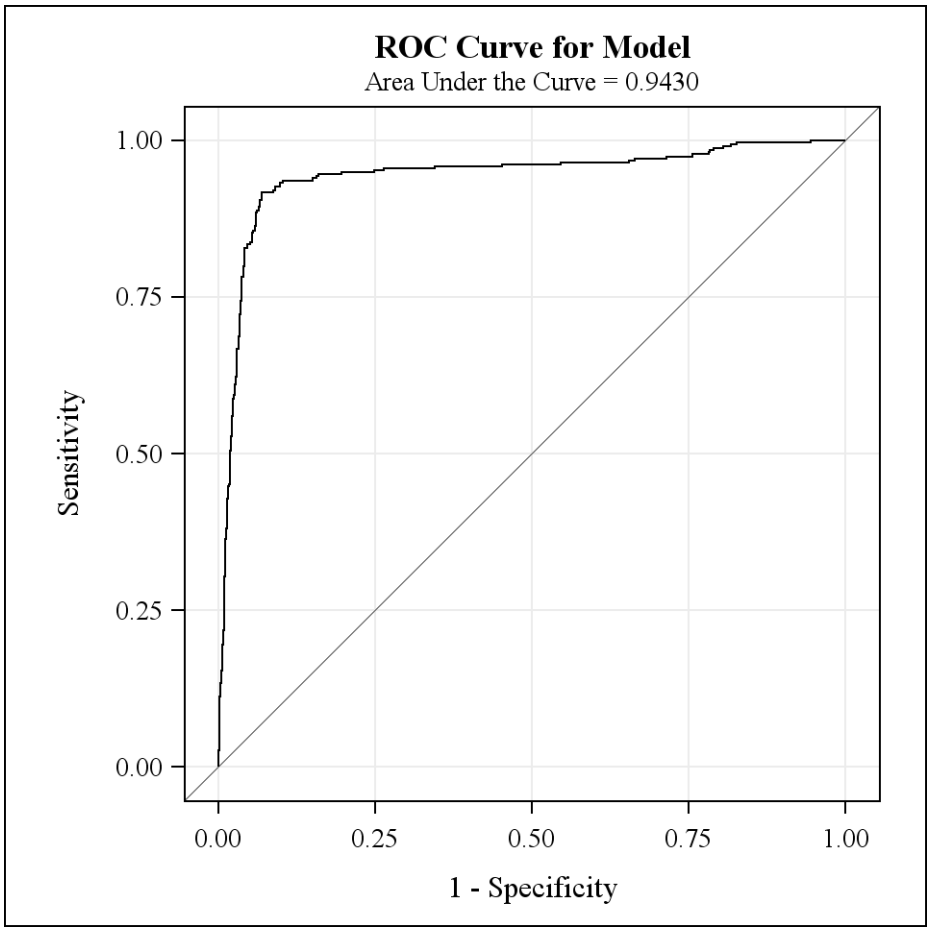


Figure 17. ROC curve for MIP_BSA

Table 13, Table 14 and Figure 18 indicate that 1.5 is a good cutoff because of the higher value of AUC and ranges from 0.9244 to 0.930 for these test antigens.

Table 13. AUC for 10 test antigens and chance*

ROC Association Statistics							
ROC Model	Mann-Whitney				Somers' D (Gini)	Gamma	Tau-a
	Area	Standard Error	95% Wald Confidence Limits				
CNSA_EX2	0.9256	0.0104	0.9052	0.9460	0.8513	0.8513	0.2728
CNSA_MBP	0.9244	0.0106	0.9036	0.9452	0.8488	0.8488	0.2720
GLIA_GFAP	0.9309	0.0103	0.9107	0.9512	0.8618	0.8618	0.2762
GLIA_S100	0.9252	0.0108	0.9040	0.9465	0.8505	0.8505	0.2726
ISLA_Gad	0.9363	0.00974	0.9172	0.9554	0.8726	0.8726	0.2797
ISLA_Gad55	0.9425	0.00846	0.9260	0.9591	0.8851	0.8851	0.2837
ISLA_PI	0.9330	0.00914	0.9151	0.9509	0.8661	0.8661	0.2776
ISLA_Tep69	0.9347	0.00976	0.9155	0.9538	0.8693	0.8693	0.2786
MIP_Abbos	0.9402	0.00943	0.9218	0.9587	0.8805	0.8805	0.2822
MIP_BSA	0.9430	0.00892	0.9255	0.9605	0.8859	0.8859	0.2840
Chance	0.5000	0	0.5000	0.5000	0	.	0

*Chance represents the diagonal line without predictive value

Table 14. AUC comparison between 10 test antigens and chance*

ROC Contrast Estimation and Testing Results by Row						
Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
CNSA_EX2 - Chance	0.4256	0.0104	0.4052	0.4460	1669.8120	<.0001
CNSA_MBP - Chance	0.4244	0.0106	0.4036	0.4452	1598.0204	<.0001
GLIA_GFAP - Chance	0.4309	0.0103	0.4107	0.4512	1740.7075	<.0001
GLIA_S100 - Chance	0.4252	0.0108	0.4040	0.4465	1537.2854	<.0001
ISLA_Gad - Chance	0.4363	0.00974	0.4172	0.4554	2007.7545	<.0001
ISLA_Gad55 - Chance	0.4425	0.00846	0.4260	0.4591	2738.5590	<.0001
ISLA_PI - Chance	0.4330	0.00914	0.4151	0.4509	2243.2769	<.0001
ISLA_Tep69 - Chance	0.4347	0.00976	0.4155	0.4538	1982.7429	<.0001
MIP_Abbos - Chance	0.4402	0.00943	0.4218	0.4587	2180.8534	<.0001
MIP_BSA - Chance	0.4430	0.00892	0.4255	0.4605	2465.5871	<.0001

*Chance represents the diagonal line without predictive value

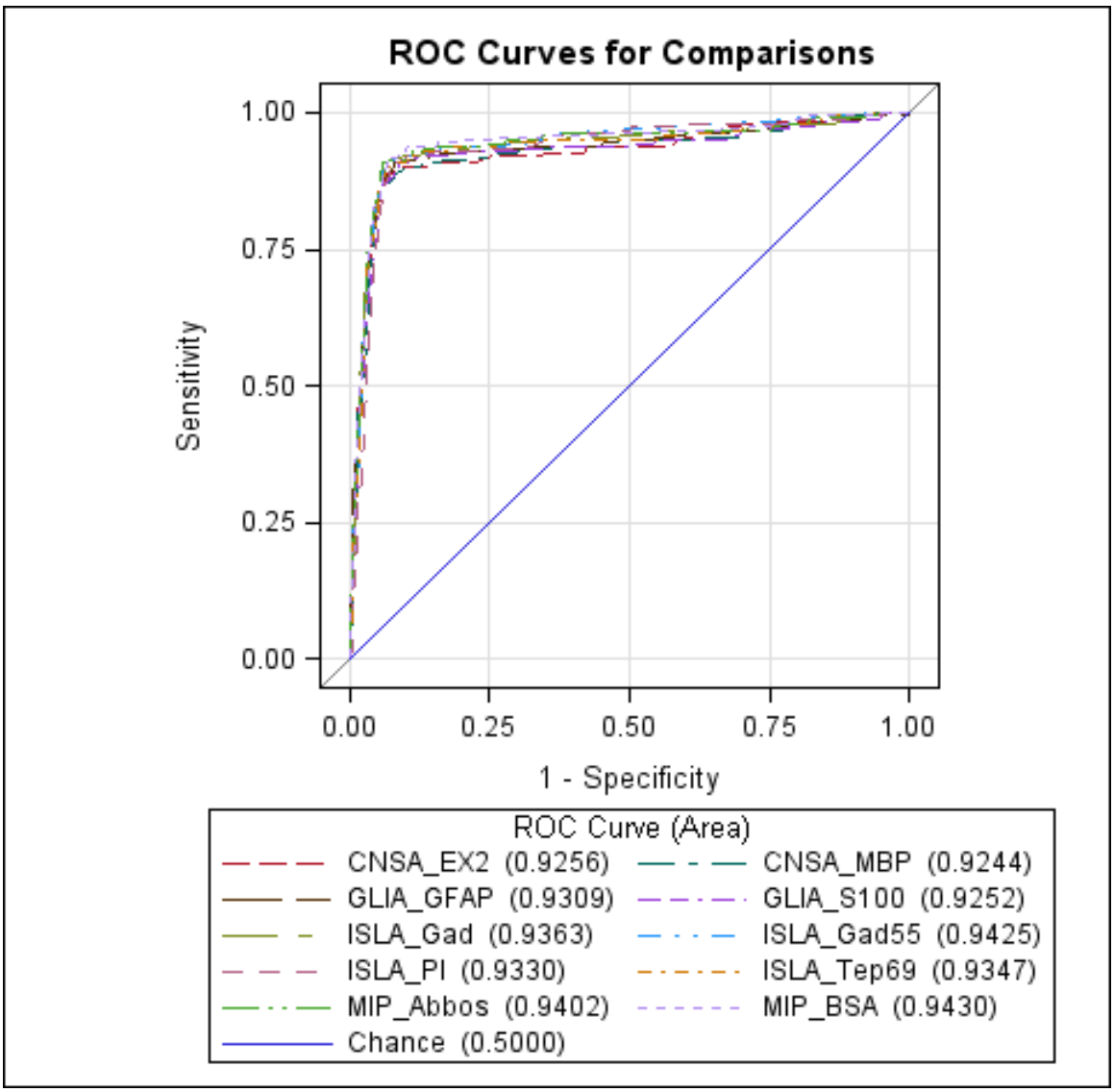


Figure 18. ROC curve for 10 test antigens and chance line

3.6.2 Criterion using number of test antigens

The numbers of positive results from the ten test antigens were counted for predicting the T1DM. The recommendation cutoff of four was adopted. The prediction at this cutoff seems good according to Figure 19.

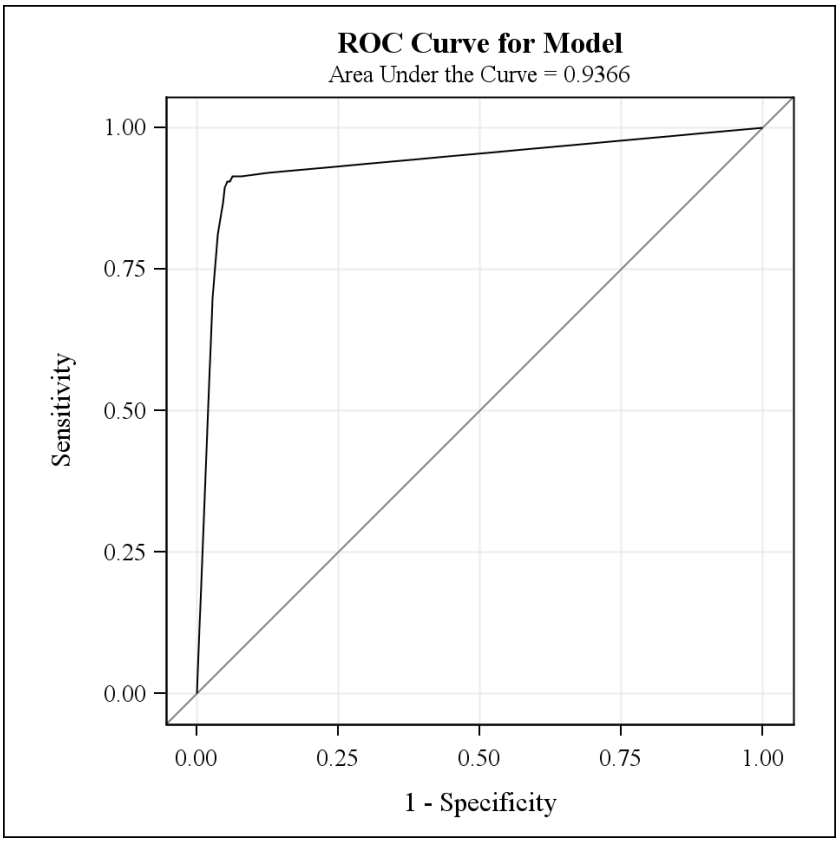


Figure 19. ROC curve for number of positive antigens under the cutoff= 4

3.6.3 Alternative cutoff for diagnosis test

Usually, when the cutoff is decided, a Student's T-test can be used to measure the effect of the diagnosis test. From Table 15 and Table 16, it is clear that the mean of SI among new onsets is significantly higher than that of FDR-nonconverters. However, we want to know how well the cutoff of 1.5 performed. It is still interesting to determine whether 1.5 is too strict or too loose for prediction. One method of analysis is the sensitivity and specificity at several cut points.

Table 15. Mean of SI between New onsets and FDR-converters

	New Onsets(N=316)	*FDR-nonconverter(n=1,255)*	*p-value**
CNSA_EX2si	1.83	1.10	<.0001
CNSA_MBPsi	1.79	1.11	<.0001
GLIA_GFAPsi	1.91	1.13	<.0001
GLIA_S100si	1.91	1.12	<.0001
ISLA_Gadsi	2.03	1.13	<.0001
ISLA_Gad55si	2.04	1.13	<.0001
ISLA_PIsi	2.10	1.17	<.0001
ISLA_Tep69si	2.12	1.14	<.0001
MIP_Abbossi	2.14	1.14	<.0001
MIP_BSAsi	2.18	1.14	<.0001

*The mean were compared using Student's t-test for unequal variance.

Table 16. Mean of SI between New onsets and FDR-converters

	New Onsets(N=313)	*FDR-nonconverter(n=1,250)*	*p-value**
Number of positive antigens	8.7	0.6	<.0001

Using 4 as the cutoff criterion for the number of positive antigens, we built a pair comparison table (Table 17) with corresponding ROC curves (Figure 20). It was hypothesized that 5 or 6 would be a better cutoff to predict the T1DM due to a better trade-off between sensitivity and specificity.

Table 17. Statistical output of number of positive antigens for new onsets

Obs	CUTOFF	COUNT	PERCENT	PCT_ROW	TP_RATE	FP_RATE
1	0	155	9.86633	34.9887	91.1392	12.3506
2	1	99	6.30172	25.7143	90.5063	7.8884
3	2	88	5.60153	23.5294	90.5063	7.0120
4	3	79	5.02864	21.6438	90.5063	6.2948
5	4	74	4.71038	20.7283	89.5570	5.8964
6	5	67	4.26480	19.1429	89.5570	5.3386
7	6	61	3.88288	17.8886	88.6076	4.8606
8	7	57	3.62826	17.3780	85.7595	4.5418
9	8	47	2.99173	15.6146	80.3797	3.7450
10	9	34	2.16423	13.4387	69.3038	2.7092

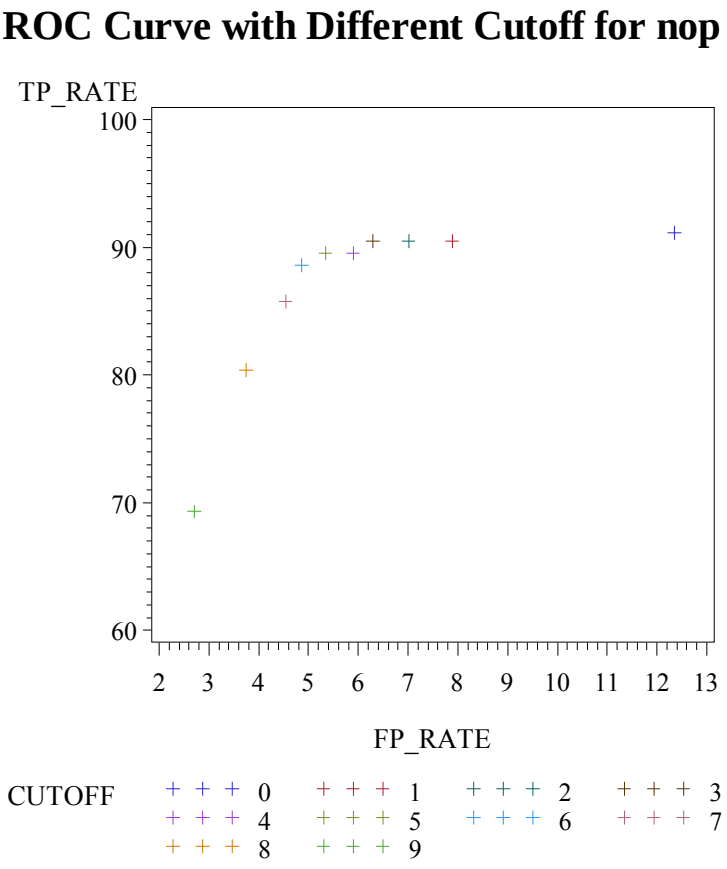


Figure 20. Optimal points in ROC curve under different cutoff of number of positive antigens

Using 1.5 as the cutoff criterion for the 10 test antigens, Appendix B.1.3 (Page 90) indicates the optimal cutoff would be a little lower than 1.5, and ranges from 1.25-1.50.

4.0 CONCLUSION

The data cleaning process is one of the most important steps in the data analysis, and it helps us to avoid false conclusions and misdirected investigations using incorrect or inconsistent data. Based on previous data, a number of hand calculations were involved. Thus, data cleaning is necessary in this setting. Despite the existence of a few mistakes, the accuracy of the hand calculations was quite good, and the error rate was less than 1%. However, a high quality data analysis has zero tolerance for mistakes and it reminds us of the importance of data checking and cleaning prior to analysis.

My thesis is a first look at the SI values and number of positive antigens in new onsets, FDR-converters, and FDR-nonconverters from this longitudinal study of juvenile onset diabetics and their family members. Preliminary summary was performed to generate information in the various subpopulations. The value of SI is stable in new onsets and FDR-nonconverters. The SI in new onsets has higher values than that of FDRs. The SI values in FDR-converters are higher than that of FDR-nonconveters. When we consider two subgroups among the FDR-converters based on the start of insulin, the SI values in FDR-converters before the start of insulin have similar values (<1.5) to that of FDR-nonconveters; the SI values in FDR-converters after insulin start have similar values (>=1.5) with that found in new onsets.

The number of positive antigens in three subpopulations also has consistency similar to the SI values. The number of positive antigens is stable in new onsets and FDR-nonconverters,

where new onsets have higher values and FDR-nonconverters have lower values. After considering the subgroups in the FDR-converters by their insulin start date, the number of positive antigens in FDR-converters before the start of insulin have similar values (<4) compared with FDR-nonconverters; and, the number of positive antigens in FDR-converters after the start of insulin have similar values (>=4) found in the new onsets.

The evaluation of a diagnostic test is an exploratory part of this thesis. ROC curves are frequently used to evaluate diagnostic tests to differentiate "healthy" individuals from "diseased" individuals or predict the existence of disease. We applied ROC analysis curves to evaluate the two criteria of T1DM prediction. The overall performance of two criteria is excellent. This indicates that the prior cutpoint recommendations for these diagnostic tests work well. However, it is necessary to explore the optimal cutoff for these two criteria for our cohort.

We found evidence to suggest that:

- For the SI, 1.5 is a good cutoff with high AUC. However, after careful examination of the previous cutoff comparison for the 10 test antigens, a slightly lower cutoff value may provide a more optimal tradeoff for the diagnostic test, i.e. using an SI cutoff of (1.25-1.5).
- For the number of positive antigens, 4 is a good cutoff for discrimination of T1DM. However, there is suggestion that a slightly higher value may provide a more optimal tradeoff for the diagnostic test, i.e. 5 or 6.

Before changes are implemented further analytical work needs to be performed. Most importantly, this should include the identification and inclusion of a greater number of FDR-converters and their respective antigen test results.

In conclusion, the development of a better understanding of the pattern of T-cell response among diabetes and non-diabetic children, and those progressing to diabetes may give us tools to predict the early onset of disease. It is this point in time where therapeutic intervention could be focused to help stem the development of T1DM or to dramatically reduces its severity. Focusing on prevention and/or attenuating progression is the public health approach to the treatment of disease.

APPENDIX A

APPENDIX FOR DATA CLEANING

Table 18. Both Well1 and Well2 Missing Observations List

PID	BloodDrawDate	PID	BloodDrawDate
116	2-May-07	48830103	21-Aug-06
204	28-Aug-07	49020101	12-Feb-07
226	23-Oct-07	49180101	8-Aug-07
233	11-Jun-07	49900103	12-Feb-07
27240101	13-Feb-06	49900104	12-Feb-07
29200104	31-Mar-05	70820103	30-Nov-05
32170101	9-May-07	70870101	4-Apr-07
36420102	31-Mar-08	71110103	31-Dec-05
40080102	12-Feb-07	71540101	12-Feb-07
40690101	4-Feb-06	71630103	25-Jul-07
43320103	30-Dec-05	71640103	18-Sep-06
43320106	30-Dec-05	71700104	12-Feb-07
45750102	12-Feb-07	71720103	24-Mar-08
45920102	26-Sep-07	71890101	14-Mar-07
47330104	4-Feb-04	71960103	10-Jul-06
48830103	18-May-06		

Table 19. Miscalculated WELLMEAN Observations List

PID	BloodDrawDate
38700101	26-Sep-05
40840101	25-Mar-08
47800101	26-Sep-05
48510106	27-Jul-05
48920101	10-Sep-05
70600105	29-Jul-05

Table 20. Miscalculated SI Observations List

PID	BloodDrawDate	PID	BloodDrawDate
186	23-Jan-07	48790103	25-Mar-06
254	24-Jul-07	48920101	10-Sep-05
256	24-Jul-07	48950102	27-Oct-07
258	24-Jul-07	49090103	9-Oct-06
35270104	1-Apr-06	49110104	13-Feb-06
38700101	26-Sep-05	49300101	13-Feb-06
39730204	9-Jul-07	49400105	28-Apr-08
40660102	27-Oct-07	49410101	13-Feb-06
43330102	21-Oct-06	49650103	10-Jul-06
46390101	25-Jul-07	49910103	15-Feb-06
47800101	26-Sep-05	70600105	29-Jul-05
47980101	22-Jan-07	71030101	24-Jul-07
47980104	25-Jul-07	71390101	15-Feb-06
48510106	27-Jul-05	71390103	15-Feb-06
48660101	6-Jun-07		

Table 21. False-negative results: SI>1.5 with "Negative" result

PID	FID	BloodDrawDate	Analyte
48790103	4879	25-Mar-06	CNSA_MBP
254	10186	24-Jul-07	GLIA_GFAP
48920101	4892	10-Sep-05	ISLA_PI
48920101	4892	10-Sep-05	ISLA_Tep69
48920101	4892	10-Sep-05	MIP_Abbos
48920101	4892	10-Sep-05	MIP_BSA

Table 22. False-positive results: Si<1.5 with "Positive" result

PID	FID	BloodDrawDate	Analyte
49400105	4940	28-Apr-08	CNSA_EX2
49400105	4940	28-Apr-08	CNSA_MBP
49400105	4940	28-Apr-08	GLIA_GFAP
49400105	4940	28-Apr-08	GLIA_S100
49400105	4940	28-Apr-08	ISLA_Gad
49400105	4940	28-Apr-08	ISLA_Gad55
42280101	4228	7-Nov-07	ISLA_PI
49400105	4940	28-Apr-08	ISLA_PI
39730204	3973	9-Jul-07	ISLA_Tep69
49400105	4940	28-Apr-08	ISLA_Tep69
71030101	7103	24-Jul-07	ISLA_Tep69
39730204	3973	9-Jul-07	MIP_Abbos
49400105	4940	28-Apr-08	MIP_Abbos
49400105	4940	28-Apr-08	MIP_BSA

Table 23. False-unaffected interpretations: number of positive test antigens >4, but interpreted as "Unaffected".

PID	FID	BloodDrawDate
32960101	3296	21-Apr-08
36569899	3656	5-Dec-05
40410104	4041	20-Jun-05
40660102	4066	10-Oct-05
41700101	4170	14-Nov-05
44520102	4452	16-Aug-07
44580102	4458	2-Jul-07
46570101	4657	5-Dec-05
47280101	4728	18-Apr-05
47930101	4793	16-Aug-07
48460103	4846	5-Nov-05
48890102	4889	28-Mar-07
48920101	4892	10-Sep-05
49010103	4901	14-Jun-06
49420203	4942	10-Jan-07
71230101	7123	20-Jan-07

Table 24. False-affected interpretations: number of positive test antigens <4, but interpreted as "Affected".

PID	FID	BloodDrawDate
46420101	4642	25-Mar-08
48910103	4891	21-Sep-06
49400105	4940	28-Apr-08
71770204	7177	5-Aug-06
71830103	7183	30-Sep-06

APPENDIX B

APPENDIX FOR PRELIMINARY SUMMARY AND DIAGNOSTIC TEST EVALUATION

B.1 SI OF TEN TEST ANTIGENS

B.1.1 SI of ten test antigens in demographics by subpopulation

Following 20 tables shows the SI in ten test antigens stratified by the demographics and subpopulation.

Table 25. SI of CNSA_EX2 in demographics before insulin start

CNSA_EX2si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.8	0.4	1.8	1.5	2.0	0.8	4.0	1.1	0.2	1.0	1.0	1.1	0.9	1.6	1.1	0.2	1.1	1.0	1.1	0.6	2.8
Female	1.8	0.7	1.8	1.5	2.0	0.9	12.4	1.3	0.4	1.1	1.0	1.7	1.0	2.2	1.1	0.2	1.1	1.0	1.1	0.7	2.8
Race																					
White	1.8	0.6	1.8	1.5	2.0	0.8	12.4	1.2	0.4	1.1	1.0	1.2	0.9	2.2	1.1	0.2	1.1	1.0	1.1	0.6	2.8
Black	1.8	0.5	1.7	1.5	2.0	1.0	3.3	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.9	1.8
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	1.8	0.6	1.6	1.4	2.1	1.1	3.1	.	.	.	.	.	.	.	1.3	0.4	1.1	1.1	1.5	0.9	2.2
HLA-DQ type																					
X	1.8	0.4	1.7	1.6	2.0	0.9	3.0	1.0	0.1	1.0	1.0	1.1	1.0	1.2	1.1	0.2	1.1	1.0	1.1	0.7	2.8
DQ2	1.8	0.8	1.8	1.5	2.0	0.8	12.4	1.0	0.1	1.1	1.0	1.1	0.9	1.1	1.1	0.2	1.1	1.0	1.1	0.7	2.5
DQ8	1.8	0.5	1.7	1.5	2.0	0.8	4.0	1.3	0.4	1.1	1.1	1.5	1.0	2.2	1.1	0.2	1.1	1.0	1.1	0.6	2.5
DQ2/DQ8	1.9	0.4	1.8	1.6	2.0	0.9	3.5	1.3	0.5	1.1	1.0	1.7	0.9	2.1	1.1	0.2	1.1	1.0	1.1	0.8	2.2
All	1.8	0.6	1.8	1.5	2.0	0.8	12.4	1.2	0.4	1.1	1.0	1.2	0.9	2.2	1.1	0.2	1.1	1.0	1.1	0.6	2.8

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 26. SI of CNSA_EX2 in demographics after insulin start

CNSA_EX2si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.8	0.4	1.8	1.5	2.0	0.8	4.0	1.6	0.5	1.6	1.1	1.8	1.1	2.5	1.1	0.2	1.1	1.0	1.1	0.6	2.8
Female	1.8	0.7	1.8	1.5	2.0	0.9	12.4	1.8	0.5	1.7	1.5	2.1	1.0	2.9	1.1	0.2	1.1	1.0	1.1	0.7	2.8
Race																					
White	1.8	0.6	1.8	1.5	2.0	0.8	12.4	1.7	0.5	1.6	1.5	2.0	1.0	2.9	1.1	0.2	1.1	1.0	1.1	0.6	2.8
Black	1.8	0.5	1.7	1.5	2.0	1.0	3.3	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.9	1.8
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	1.8	0.6	1.6	1.4	2.1	1.1	3.1	.	.	.	.	.	.	.	1.3	0.4	1.1	1.1	1.5	0.9	2.2
HLA-DQ type																					
X	1.8	0.4	1.7	1.6	2.0	0.9	3.0	1.9	0.3	1.7	1.7	2.0	1.6	2.3	1.1	0.2	1.1	1.0	1.1	0.7	2.8
DQ2	1.8	0.8	1.8	1.5	2.0	0.8	12.4	1.3	0.4	1.1	1.0	1.8	1.0	1.8	1.1	0.2	1.1	1.0	1.1	0.7	2.5
DQ8	1.8	0.5	1.7	1.5	2.0	0.8	4.0	1.8	0.5	1.6	1.5	2.1	1.1	2.9	1.1	0.2	1.1	1.0	1.1	0.6	2.5
DQ2/DQ8	1.9	0.4	1.8	1.6	2.0	0.9	3.5	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.1	0.8	2.2
All	1.8	0.6	1.8	1.5	2.0	0.8	12.4	1.7	0.5	1.6	1.5	2.0	1.0	2.9	1.1	0.2	1.1	1.0	1.1	0.6	2.8

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 27. SI of CNSA_MBP in demographics before insulin start

CNSA_MBPsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.8	0.4	1.7	1.5	2.0	0.8	3.9	1.1	0.2	1.1	1.0	1.1	0.9	1.8	1.1	0.2	1.1	1.0	1.1	0.7	3.2
Female	1.8	0.5	1.8	1.5	2.0	0.9	3.3	1.3	0.4	1.1	1.0	1.7	1.0	2.3	1.1	0.3	1.1	1.0	1.1	0.5	9.0
Race																					
White	1.8	0.4	1.8	1.5	2.0	0.8	3.9	1.3	0.4	1.1	1.0	1.3	0.9	2.3	1.1	0.3	1.1	1.0	1.1	0.5	9.0
Black	1.8	0.5	1.7	1.5	2.0	0.9	3.3	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.7	2.2
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	1.0	1.2
Other	1.8	0.7	1.7	1.5	2.2	1.0	3.2	.	.	.	.	.	.	.	1.3	0.4	1.1	1.0	1.4	1.0	2.3
HLA-DQ type																					
X	1.8	0.4	1.8	1.5	2.0	0.9	3.1	1.1	0.1	1.0	1.0	1.1	1.0	1.2	1.1	0.3	1.1	1.0	1.1	0.5	9.0
DQ2	1.8	0.5	1.7	1.5	2.0	0.8	3.6	1.0	0.1	1.1	1.0	1.1	0.9	1.2	1.1	0.2	1.1	1.0	1.1	0.6	2.8
DQ8	1.8	0.5	1.7	1.5	2.0	0.9	3.9	1.3	0.4	1.1	1.0	1.6	1.0	2.3	1.1	0.3	1.1	1.0	1.1	0.6	6.0
DQ2/DQ8	1.9	0.4	1.8	1.6	2.1	0.9	3.2	1.4	0.5	1.3	1.1	1.8	1.0	2.1	1.1	0.3	1.1	1.0	1.1	0.8	3.2
All	1.8	0.4	1.8	1.5	2.0	0.8	3.9	1.3	0.4	1.1	1.0	1.3	0.9	2.3	1.1	0.3	1.1	1.0	1.1	0.5	9.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 28. SI of CNSA_MBP in demographics after insulin start

CNSA_MBPsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.8	0.4	1.7	1.5	2.0	0.8	3.9	1.6	0.6	1.6	1.2	1.8	1.0	2.6	1.1	0.2	1.1	1.0	1.1	0.7	3.2
Female	1.8	0.5	1.8	1.5	2.0	0.9	3.3	1.7	0.5	1.6	1.5	1.9	1.1	2.8	1.1	0.3	1.1	1.0	1.1	0.5	9.0
Race																					
White	1.8	0.4	1.8	1.5	2.0	0.8	3.9	1.7	0.5	1.6	1.5	1.9	1.0	2.8	1.1	0.3	1.1	1.0	1.1	0.5	9.0
Black	1.8	0.5	1.7	1.5	2.0	0.9	3.3	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.7	2.2
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	1.0	1.2
Other	1.8	0.7	1.7	1.5	2.2	1.0	3.2	.	.	.	.	.	.	.	1.3	0.4	1.1	1.0	1.4	1.0	2.3
HLA-DQ type																					
X	1.8	0.4	1.8	1.5	2.0	0.9	3.1	1.8	0.3	1.7	1.6	2.1	1.5	2.3	1.1	0.3	1.1	1.0	1.1	0.5	9.0
DQ2	1.8	0.5	1.7	1.5	2.0	0.8	3.6	1.3	0.4	1.1	1.0	1.7	1.0	1.7	1.1	0.2	1.1	1.0	1.1	0.6	2.8
DQ8	1.8	0.5	1.7	1.5	2.0	0.9	3.9	1.8	0.5	1.6	1.5	1.9	1.1	2.8	1.1	0.3	1.1	1.0	1.1	0.6	6.0
DQ2/DQ8	1.9	0.4	1.8	1.6	2.1	0.9	3.2	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.1	0.8	3.2
All	1.8	0.4	1.8	1.5	2.0	0.8	3.9	1.7	0.5	1.6	1.5	1.9	1.0	2.8	1.1	0.3	1.1	1.0	1.1	0.5	9.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 29. SI of GLIA_GFAP in demographics before insulin start

GLIA_GFAPsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.9	0.5	1.9	1.6	2.2	0.9	3.9	1.2	0.5	1.0	0.9	1.2	0.9	2.6	1.1	0.3	1.1	1.0	1.2	0.5	3.0
Female	1.9	0.5	1.8	1.6	2.1	0.5	3.6	1.4	0.5	1.1	1.0	1.7	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.6	23.1
Race																					
White	1.9	0.5	1.9	1.6	2.2	0.5	3.9	1.3	0.5	1.1	1.0	1.3	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.5	23.1
Black	1.9	0.5	1.8	1.6	2.1	1.0	3.1	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.1	0.8	1.9
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.0	0.6	1.8	1.6	2.6	1.1	3.0	.	.	.	.	.	.	.	1.3	0.3	1.2	1.0	1.6	1.0	1.9
HLA-DQ type																					
X	1.9	0.4	1.9	1.7	2.1	0.9	3.4	1.1	0.1	1.1	1.1	1.2	1.0	1.3	1.1	0.3	1.1	1.0	1.2	0.7	3.5
DQ2	1.9	0.5	1.9	1.6	2.2	0.9	3.6	1.1	0.2	1.0	1.0	1.1	0.9	1.4	1.1	0.3	1.1	1.0	1.1	0.6	3.0
DQ8	1.9	0.5	1.8	1.6	2.2	0.9	3.9	1.4	0.6	1.1	1.0	1.7	0.9	2.6	1.2	0.7	1.1	1.0	1.2	0.5	23.1
DQ2/DQ8	2.0	0.5	1.9	1.7	2.2	0.5	3.6	1.5	0.7	1.1	1.1	1.9	0.9	2.7	1.1	0.2	1.1	1.0	1.2	0.7	2.3
All	1.9	0.5	1.9	1.6	2.2	0.5	3.9	1.3	0.5	1.1	1.0	1.3	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.5	23.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 30. SI of GLIA_GFAP in demographics after insulin start

GLIA_GFAPsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.9	0.5	1.9	1.6	2.2	0.9	3.9	1.7	0.7	1.6	1.1	2.5	1.0	2.6	1.1	0.3	1.1	1.0	1.2	0.5	3.0
Female	1.9	0.5	1.8	1.6	2.1	0.5	3.6	1.9	0.4	1.9	1.6	2.2	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.6	23.1
Race																					
White	1.9	0.5	1.9	1.6	2.2	0.5	3.9	1.8	0.5	1.9	1.4	2.2	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.5	23.1
Black	1.9	0.5	1.8	1.6	2.1	1.0	3.1	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.1	0.8	1.9
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.0	0.6	1.8	1.6	2.6	1.1	3.0	.	.	.	.	.	.	.	1.3	0.3	1.2	1.0	1.6	1.0	1.9
HLA-DQ type																					
X	1.9	0.4	1.9	1.7	2.1	0.9	3.4	1.7	0.4	1.8	1.4	1.9	1.0	1.9	1.1	0.3	1.1	1.0	1.2	0.7	3.5
DQ2	1.9	0.5	1.9	1.6	2.2	0.9	3.6	1.5	0.5	1.4	1.1	2.0	1.1	2.0	1.1	0.3	1.1	1.0	1.1	0.6	3.0
DQ8	1.9	0.5	1.8	1.6	2.2	0.9	3.9	1.9	0.5	2.1	1.6	2.3	0.9	2.7	1.2	0.7	1.1	1.0	1.2	0.5	23.1
DQ2/DQ8	2.0	0.5	1.9	1.7	2.2	0.5	3.6	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.7	2.3
All	1.9	0.5	1.9	1.6	2.2	0.5	3.9	1.8	0.5	1.9	1.4	2.2	0.9	2.7	1.1	0.5	1.1	1.0	1.2	0.5	23.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 31. SI of GLIA_S100 in demographics before insulin start

GLIA_S100si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.2	0.4	1.1	1.0	1.1	0.8	2.3	1.1	0.3	1.1	1.0	1.2	0.5	2.9
Female	1.9	0.5	1.9	1.6	2.2	0.9	3.8	1.4	0.5	1.1	1.0	1.7	0.9	2.6	1.1	0.3	1.1	1.0	1.2	0.4	5.9
Race																					
White	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.3	0.5	1.1	1.0	1.3	0.8	2.6	1.1	0.3	1.1	1.0	1.2	0.4	5.9
Black	1.9	0.5	1.8	1.6	2.2	1.0	3.1	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	2.0	0.6	1.8	1.5	2.6	1.0	3.1	.	.	.	.	.	.	.	1.3	0.3	1.1	1.0	1.5	0.9	1.9
HLA-DQ type																					
X	1.9	0.4	1.9	1.7	2.2	0.9	3.0	1.1	0.1	1.0	1.0	1.2	1.0	1.2	1.1	0.3	1.1	1.0	1.2	0.5	3.5
DQ2	1.9	0.5	1.9	1.6	2.2	0.7	3.7	1.1	0.1	1.1	1.0	1.1	0.8	1.3	1.1	0.3	1.1	1.0	1.1	0.6	5.9
DQ8	1.9	0.5	1.8	1.6	2.2	0.9	4.1	1.4	0.6	1.1	1.0	2.0	0.9	2.4	1.1	0.3	1.1	1.0	1.2	0.4	2.9
DQ2/DQ8	2.0	0.5	1.9	1.7	2.2	1.0	3.8	1.5	0.6	1.1	1.1	2.0	1.0	2.6	1.1	0.2	1.1	1.0	1.2	0.8	2.4
All	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.3	0.5	1.1	1.0	1.3	0.8	2.6	1.1	0.3	1.1	1.0	1.2	0.4	5.9

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 32. SI of GLIA_S100 in demographics after insulin start

GLIA_S100si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.9	0.7	1.8	1.2	2.3	1.1	3.1	1.1	0.3	1.1	1.0	1.2	0.5	2.9
Female	1.9	0.5	1.9	1.6	2.2	0.9	3.8	1.8	0.5	1.7	1.4	2.2	1.0	2.6	1.1	0.3	1.1	1.0	1.2	0.4	5.9
Race																					
White	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.8	0.5	1.7	1.4	2.2	1.0	3.1	1.1	0.3	1.1	1.0	1.2	0.4	5.9
Black	1.9	0.5	1.8	1.6	2.2	1.0	3.1	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	2.0	0.6	1.8	1.5	2.6	1.0	3.1	.	.	.	.	.	.	.	1.3	0.3	1.1	1.0	1.5	0.9	1.9
HLA-DQ type																					
X	1.9	0.4	1.9	1.7	2.2	0.9	3.0	1.6	0.2	1.6	1.5	1.8	1.4	1.9	1.1	0.3	1.1	1.0	1.2	0.5	3.5
DQ2	1.9	0.5	1.9	1.6	2.2	0.7	3.7	1.5	0.4	1.3	1.1	1.9	1.1	1.9	1.1	0.3	1.1	1.0	1.1	0.6	5.9
DQ8	1.9	0.5	1.8	1.6	2.2	0.9	4.1	1.9	0.6	2.0	1.6	2.3	1.0	3.1	1.1	0.3	1.1	1.0	1.2	0.4	2.9
DQ2/DQ8	2.0	0.5	1.9	1.7	2.2	1.0	3.8	.	.	.	.	.	.	.	1.1	0.2	1.1	1.0	1.2	0.8	2.4
All	1.9	0.5	1.9	1.6	2.2	0.7	4.1	1.8	0.5	1.7	1.4	2.2	1.0	3.1	1.1	0.3	1.1	1.0	1.2	0.4	5.9

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 33. SI of ISLA_Gad in demographics before insulin start

ISLA_Gadsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.3	0.5	1.1	1.0	1.2	0.9	2.5	1.2	0.3	1.1	1.0	1.2	0.5	6.3
Female	2.0	0.5	2.0	1.7	2.3	0.9	4.0	1.4	0.5	1.1	1.0	1.8	0.9	2.7	1.1	0.3	1.1	1.0	1.2	0.5	5.7
Race																					
White	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.3	0.5	1.1	1.0	1.5	0.9	2.7	1.1	0.3	1.1	1.0	1.2	0.5	6.3
Black	2.0	0.5	2.0	1.7	2.3	1.0	3.0	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	2.2
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.0	0.1	1.0	0.9	1.1	0.9	1.1
Other	2.1	0.7	2.0	1.6	2.9	0.9	3.5	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	0.9	1.9
HLA-DQ type																					
X	2.0	0.4	2.0	1.7	2.2	0.9	3.3	1.1	0.2	1.1	0.9	1.2	0.9	1.2	1.1	0.3	1.1	1.0	1.2	0.5	3.2
DQ2	2.0	0.5	1.9	1.7	2.3	0.8	3.7	1.1	0.2	1.1	1.0	1.2	0.9	1.5	1.1	0.3	1.1	1.0	1.1	0.6	3.8
DQ8	2.0	0.6	2.0	1.6	2.3	0.8	4.7	1.4	0.6	1.1	1.0	2.0	1.0	2.5	1.2	0.4	1.1	1.0	1.2	0.5	6.3
DQ2/DQ8	2.1	0.5	2.1	1.8	2.4	0.9	3.8	1.6	0.6	1.5	1.1	2.0	0.9	2.7	1.1	0.3	1.1	1.0	1.2	0.7	2.8
All	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.3	0.5	1.1	1.0	1.5	0.9	2.7	1.2	0.3	1.1	1.0	1.2	0.5	6.3

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 34. SI of ISLA_Gad in demographics after insulin start

ISLA_Gadsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.9	0.8	1.8	1.4	2.5	1.1	3.0	1.2	0.3	1.1	1.0	1.2	0.5	6.3
Female	2.0	0.5	2.0	1.7	2.3	0.9	4.0	1.9	0.6	2.0	1.5	2.4	1.1	3.0	1.1	0.3	1.1	1.0	1.2	0.5	5.7
Race																					
White	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.9	0.6	2.0	1.4	2.4	1.1	3.0	1.1	0.3	1.1	1.0	1.2	0.5	6.3
Black	2.0	0.5	2.0	1.7	2.3	1.0	3.0	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	2.2
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.0	0.1	1.0	0.9	1.1	0.9	1.1
Other	2.1	0.7	2.0	1.6	2.9	0.9	3.5	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	0.9	1.9
HLA-DQ type																					
X	2.0	0.4	2.0	1.7	2.2	0.9	3.3	1.7	0.3	1.6	1.5	1.8	1.4	2.0	1.1	0.3	1.1	1.0	1.2	0.5	3.2
DQ2	2.0	0.5	1.9	1.7	2.3	0.8	3.7	1.4	0.6	1.1	1.1	2.1	1.1	2.1	1.1	0.3	1.1	1.0	1.1	0.6	3.8
DQ8	2.0	0.6	2.0	1.6	2.3	0.8	4.7	2.1	0.6	2.2	1.5	2.5	1.1	3.0	1.2	0.4	1.1	1.0	1.2	0.5	6.3
DQ2/DQ8	2.1	0.5	2.1	1.8	2.4	0.9	3.8	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.2	0.7	2.8
All	2.0	0.5	2.0	1.7	2.3	0.8	4.7	1.9	0.6	2.0	1.4	2.4	1.1	3.0	1.2	0.3	1.1	1.0	1.2	0.5	6.3

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 35. SI of ISLA_Gad55 in demographics before insulin start

ISLA_Gad55si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.0	0.5	2.0	1.7	2.3	0.8	3.8	1.2	0.4	1.1	1.0	1.2	0.9	2.3	1.2	0.3	1.1	1.0	1.2	0.6	3.1
Female	2.0	0.5	2.0	1.7	2.3	0.9	4.2	1.4	0.5	1.1	1.1	1.6	0.9	2.8	1.2	0.3	1.1	1.0	1.2	0.5	3.6
Race																					
White	2.0	0.5	2.0	1.7	2.3	0.8	4.2	1.3	0.5	1.1	1.0	1.5	0.9	2.8	1.2	0.3	1.1	1.0	1.2	0.5	3.6
Black	2.0	0.5	1.9	1.7	2.2	1.0	3.5	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.4
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.0	1.1	1.0	1.1
Other	2.1	0.6	2.0	1.7	2.6	1.1	3.1	.	.	.	.	.	.	.	1.3	0.4	1.1	1.1	1.7	1.0	1.9
HLA-DQ type																					
X	2.0	0.5	2.0	1.7	2.3	1.0	3.6	1.1	0.2	1.1	1.0	1.2	0.9	1.3	1.2	0.3	1.1	1.0	1.2	0.6	3.5
DQ2	2.0	0.5	2.0	1.7	2.3	0.9	4.2	1.1	0.1	1.0	1.0	1.2	1.0	1.2	1.1	0.3	1.1	1.0	1.2	0.5	3.6
DQ8	2.0	0.6	2.0	1.7	2.3	0.8	4.2	1.4	0.6	1.1	1.1	2.0	0.9	2.4	1.2	0.3	1.1	1.0	1.2	0.5	3.1
DQ2/DQ8	2.1	0.5	2.1	1.8	2.4	0.9	3.9	1.5	0.6	1.5	1.1	1.7	1.0	2.8	1.1	0.3	1.1	1.0	1.1	0.8	2.5
All	2.0	0.5	2.0	1.7	2.3	0.8	4.2	1.3	0.5	1.1	1.0	1.5	0.9	2.8	1.2	0.3	1.1	1.0	1.2	0.5	3.6

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 36. SI of ISLA_Gad55 in demographics after insulin start

ISLA_Gad55si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.0	0.5	2.0	1.7	2.3	0.8	3.8	1.9	0.6	1.8	1.6	2.3	1.1	2.6	1.2	0.3	1.1	1.0	1.2	0.6	3.1
Female	2.0	0.5	2.0	1.7	2.3	0.9	4.2	1.9	0.6	2.0	1.6	2.3	0.9	3.2	1.2	0.3	1.1	1.0	1.2	0.5	3.6
Race																					
White	2.0	0.5	2.0	1.7	2.3	0.8	4.2	1.9	0.6	2.0	1.6	2.3	0.9	3.2	1.2	0.3	1.1	1.0	1.2	0.5	3.6
Black	2.0	0.5	1.9	1.7	2.2	1.0	3.5	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.4
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.0	1.1	1.0	1.1
Other	2.1	0.6	2.0	1.7	2.6	1.1	3.1	.	.	.	.	.	.	.	1.3	0.4	1.1	1.1	1.7	1.0	1.9
HLA-DQ type																					
X	2.0	0.5	2.0	1.7	2.3	1.0	3.6	1.8	0.3	1.8	1.6	2.1	1.6	2.1	1.2	0.3	1.1	1.0	1.2	0.6	3.5
DQ2	2.0	0.5	2.0	1.7	2.3	0.9	4.2	1.3	0.6	1.1	0.9	2.0	0.9	2.0	1.1	0.3	1.1	1.0	1.2	0.5	3.6
DQ8	2.0	0.6	2.0	1.7	2.3	0.8	4.2	2.1	0.5	2.1	1.6	2.4	1.1	3.2	1.2	0.3	1.1	1.0	1.2	0.5	3.1
DQ2/DQ8	2.1	0.5	2.1	1.8	2.4	0.9	3.9	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.1	0.8	2.5
All	2.0	0.5	2.0	1.7	2.3	0.8	4.2	1.9	0.6	2.0	1.6	2.3	0.9	3.2	1.2	0.3	1.1	1.0	1.2	0.5	3.6

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 37. SI of ISLA_PI in demographics before insulin start

ISLA_PIsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.6	2.1	1.8	2.5	0.8	5.9	1.2	0.3	1.1	1.0	1.2	1.0	2.1	1.2	0.4	1.1	1.0	1.2	0.5	7.1
Female	2.2	0.6	2.1	1.8	2.5	0.9	4.3	1.4	0.6	1.1	1.0	1.8	0.9	2.8	1.2	0.4	1.1	1.0	1.2	0.4	8.0
Race																					
White	2.2	0.6	2.1	1.8	2.5	0.8	5.9	1.3	0.5	1.1	1.0	1.3	0.9	2.8	1.2	0.4	1.1	1.0	1.2	0.4	8.0
Black	2.1	0.6	2.1	1.7	2.4	1.0	3.7	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.4	1.2	2.0	1.6	2.9	1.1	5.6	.	.	.	.	.	.	.	1.5	0.7	1.1	1.0	1.9	0.9	3.1
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.9	3.9	1.0	0.2	1.0	0.9	1.2	0.9	1.3	1.2	0.4	1.1	1.0	1.2	0.5	7.1
DQ2	2.2	0.7	2.1	1.7	2.5	0.8	5.9	1.0	0.1	1.0	1.0	1.1	1.0	1.1	1.2	0.4	1.1	1.0	1.2	0.4	8.0
DQ8	2.2	0.6	2.1	1.7	2.5	0.9	4.3	1.4	0.5	1.1	1.0	1.8	1.0	2.3	1.2	0.3	1.1	1.0	1.2	0.5	3.7
DQ2/DQ8	2.3	0.7	2.2	1.9	2.6	1.0	5.6	1.6	0.7	1.2	1.1	2.1	1.0	2.8	1.2	0.3	1.1	1.0	1.2	0.7	3.0
All	2.2	0.6	2.1	1.8	2.5	0.8	5.9	1.3	0.5	1.1	1.0	1.3	0.9	2.8	1.2	0.4	1.1	1.0	1.2	0.4	8.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 38. SI of ISLA_PI in demographics after insulin start

ISLA_PIsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.6	2.1	1.8	2.5	0.8	5.9	1.8	0.6	1.7	1.5	2.1	1.1	2.7	1.2	0.4	1.1	1.0	1.2	0.5	7.1
Female	2.2	0.6	2.1	1.8	2.5	0.9	4.3	2.1	0.6	2.2	1.5	2.6	1.0	3.0	1.2	0.4	1.1	1.0	1.2	0.4	8.0
Race																					
White	2.2	0.6	2.1	1.8	2.5	0.8	5.9	2.0	0.6	1.9	1.5	2.4	1.0	3.0	1.2	0.4	1.1	1.0	1.2	0.4	8.0
Black	2.1	0.6	2.1	1.7	2.4	1.0	3.7	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.4	1.2	2.0	1.6	2.9	1.1	5.6	.	.	.	.	.	.	.	1.5	0.7	1.1	1.0	1.9	0.9	3.1
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.9	3.9	2.2	0.6	2.0	1.8	2.6	1.7	3.0	1.2	0.4	1.1	1.0	1.2	0.5	7.1
DQ2	2.2	0.7	2.1	1.7	2.5	0.8	5.9	1.4	0.6	1.1	1.1	2.1	1.1	2.1	1.2	0.4	1.1	1.0	1.2	0.4	8.0
DQ8	2.2	0.6	2.1	1.7	2.5	0.9	4.3	2.0	0.6	2.0	1.5	2.6	1.0	2.9	1.2	0.3	1.1	1.0	1.2	0.5	3.7
DQ2/DQ8	2.3	0.7	2.2	1.9	2.6	1.0	5.6	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	3.0
All	2.2	0.6	2.1	1.8	2.5	0.8	5.9	2.0	0.6	1.9	1.5	2.4	1.0	3.0	1.2	0.4	1.1	1.0	1.2	0.4	8.0

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 39. SI of ISLA_Tep69 in demographics before insulin start

ISLA_Tep69si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.1	0.5	2.1	1.8	2.4	0.8	4.2	1.2	0.6	1.0	0.9	1.3	0.7	2.7	1.2	0.3	1.1	1.0	1.2	0.7	6.7
Female	2.1	0.7	2.1	1.8	2.4	0.8	11.9	1.4	0.6	1.1	1.0	1.8	0.9	2.6	1.2	0.3	1.1	1.0	1.2	0.3	6.9
Race																					
White	2.1	0.6	2.1	1.8	2.4	0.8	11.9	1.3	0.6	1.1	1.0	1.4	0.7	2.7	1.2	0.3	1.1	1.0	1.2	0.3	6.9
Black	2.0	0.6	2.1	1.7	2.3	1.0	3.9	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.9	2.0
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.2	0.6	2.1	1.9	2.6	0.9	3.1	.	.	.	.	.	.	.	1.4	0.5	1.2	1.1	1.9	0.9	2.4
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.8	3.6	1.1	0.2	1.0	1.0	1.2	1.0	1.3	1.2	0.3	1.1	1.0	1.2	0.3	4.0
DQ2	2.1	0.5	2.1	1.8	2.4	0.8	4.3	1.0	0.2	1.0	0.9	1.2	0.7	1.4	1.2	0.4	1.1	1.0	1.2	0.5	6.9
DQ8	2.1	0.6	2.0	1.7	2.4	0.8	4.2	1.4	0.7	1.1	1.0	2.1	0.8	2.7	1.2	0.3	1.1	1.0	1.2	0.5	3.2
DQ2/DQ8	2.2	0.9	2.1	1.9	2.5	1.0	11.9	1.6	0.7	1.2	1.0	2.1	0.9	2.6	1.2	0.4	1.1	1.0	1.2	0.8	6.7
All	2.1	0.6	2.1	1.8	2.4	0.8	11.9	1.3	0.6	1.1	1.0	1.4	0.7	2.7	1.2	0.3	1.1	1.0	1.2	0.3	6.9

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 40. SI of ISLA_Tep69 in demographics after insulin start

ISLA_Tep69si	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.1	0.5	2.1	1.8	2.4	0.8	4.2	2.0	0.7	1.9	1.4	2.7	1.1	3.0	1.2	0.3	1.1	1.0	1.2	0.7	6.7
Female	2.1	0.7	2.1	1.8	2.4	0.8	11.9	2.0	0.6	2.1	1.6	2.4	1.0	2.9	1.2	0.3	1.1	1.0	1.2	0.3	6.9
Race																					
White	2.1	0.6	2.1	1.8	2.4	0.8	11.9	2.0	0.6	2.1	1.6	2.4	1.0	3.0	1.2	0.3	1.1	1.0	1.2	0.3	6.9
Black	2.0	0.6	2.1	1.7	2.3	1.0	3.9	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.9	2.0
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.0	1.1	1.1	1.1	1.1	1.1
Other	2.2	0.6	2.1	1.9	2.6	0.9	3.1	.	.	.	.	.	.	.	1.4	0.5	1.2	1.1	1.9	0.9	2.4
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.8	3.6	2.0	0.3	2.0	1.7	2.3	1.7	2.3	1.2	0.3	1.1	1.0	1.2	0.3	4.0
DQ2	2.1	0.5	2.1	1.8	2.4	0.8	4.3	1.4	0.6	1.1	1.0	2.1	1.0	2.1	1.2	0.4	1.1	1.0	1.2	0.5	6.9
DQ8	2.1	0.6	2.0	1.7	2.4	0.8	4.2	2.1	0.6	2.2	1.6	2.7	1.0	3.0	1.2	0.3	1.1	1.0	1.2	0.5	3.2
DQ2/DQ8	2.2	0.9	2.1	1.9	2.5	1.0	11.9	.	.	.	.	.	.	.	1.2	0.4	1.1	1.0	1.2	0.8	6.7
All	2.1	0.6	2.1	1.8	2.4	0.8	11.9	2.0	0.6	2.1	1.6	2.4	1.0	3.0	1.2	0.3	1.1	1.0	1.2	0.3	6.9

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 41. SI of MIP_Abbos in demographics before insulin start

MIP_Abbossi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.9	2.1	1.8	2.5	0.8	18.5	1.3	0.5	1.1	1.0	1.2	0.8	2.6	1.2	0.4	1.1	1.0	1.2	0.7	7.1
Female	2.2	0.6	2.1	1.8	2.5	0.9	4.2	1.4	0.6	1.1	1.0	1.9	0.9	2.9	1.2	0.3	1.1	1.0	1.2	0.5	4.1
Race																					
White	2.2	0.8	2.1	1.8	2.5	0.8	18.5	1.4	0.6	1.1	1.0	1.5	0.8	2.9	1.2	0.3	1.1	1.0	1.2	0.5	7.1
Black	2.1	0.6	2.1	1.8	2.4	1.0	3.8	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.0	0.0	1.0	1.0	1.1	1.0	1.1
Other	2.3	0.7	2.2	2.0	3.0	1.0	3.4	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	0.9	2.2
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	1.0	3.8	1.1	0.1	1.1	1.0	1.2	0.9	1.2	1.2	0.4	1.1	1.0	1.2	0.6	7.1
DQ2	2.1	0.6	2.1	1.8	2.5	0.9	4.4	1.1	0.2	1.1	1.0	1.2	0.9	1.5	1.2	0.3	1.1	1.0	1.2	0.5	4.1
DQ8	2.2	1.0	2.1	1.7	2.5	0.8	18.5	1.4	0.6	1.1	1.0	2.1	0.8	2.6	1.2	0.3	1.1	1.0	1.2	0.6	3.2
DQ2/DQ8	2.3	0.6	2.2	1.9	2.6	0.9	4.4	1.6	0.7	1.2	1.0	1.9	0.9	2.9	1.2	0.3	1.1	1.0	1.2	0.7	2.8
All	2.2	0.8	2.1	1.8	2.5	0.8	18.5	1.4	0.6	1.1	1.0	1.5	0.8	2.9	1.2	0.3	1.1	1.0	1.2	0.5	7.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 42. SI of MIP_Abbos in demographics after insulin start

MIP_Abbossi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.9	2.1	1.8	2.5	0.8	18.5	2.0	0.8	1.9	1.5	2.6	1.1	3.1	1.2	0.4	1.1	1.0	1.2	0.7	7.1
Female	2.2	0.6	2.1	1.8	2.5	0.9	4.2	2.1	0.6	2.3	1.6	2.6	1.1	3.0	1.2	0.3	1.1	1.0	1.2	0.5	4.1
Race																					
White	2.2	0.8	2.1	1.8	2.5	0.8	18.5	2.1	0.6	2.3	1.6	2.6	1.1	3.1	1.2	0.3	1.1	1.0	1.2	0.5	7.1
Black	2.1	0.6	2.1	1.8	2.4	1.0	3.8	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.8	2.1
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.0	0.0	1.0	1.0	1.1	1.0	1.1
Other	2.3	0.7	2.2	2.0	3.0	1.0	3.4	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	0.9	2.2
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	1.0	3.8	2.0	0.4	2.1	1.7	2.4	1.5	2.4	1.2	0.4	1.1	1.0	1.2	0.6	7.1
DQ2	2.1	0.6	2.1	1.8	2.5	0.9	4.4	1.5	0.7	1.2	1.1	2.3	1.1	2.3	1.2	0.3	1.1	1.0	1.2	0.5	4.1
DQ8	2.2	1.0	2.1	1.7	2.5	0.8	18.5	2.2	0.6	2.3	1.6	2.7	1.1	3.1	1.2	0.3	1.1	1.0	1.2	0.6	3.2
DQ2/DQ8	2.3	0.6	2.2	1.9	2.6	0.9	4.4	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.2	0.7	2.8
All	2.2	0.8	2.1	1.8	2.5	0.8	18.5	2.1	0.6	2.3	1.6	2.6	1.1	3.1	1.2	0.3	1.1	1.0	1.2	0.5	7.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 43. SI of MIP_BSA in demographics before insulin start

MIP_BSAsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.6	2.1	1.8	2.5	0.8	4.5	1.3	0.5	1.1	1.0	1.3	0.9	2.5	1.2	0.4	1.1	1.0	1.2	0.6	7.1
Female	2.2	0.6	2.1	1.8	2.5	1.0	4.1	1.4	0.6	1.1	1.0	1.8	0.9	2.7	1.2	0.3	1.1	1.0	1.2	0.7	4.7
Race																					
White	2.2	0.6	2.1	1.8	2.5	0.8	4.5	1.4	0.6	1.1	1.0	1.5	0.9	2.7	1.2	0.3	1.1	1.0	1.2	0.6	7.1
Black	2.1	0.5	2.1	1.8	2.3	1.0	3.9	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.3	0.7	2.3
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	2.3	0.7	2.3	2.0	2.8	1.0	3.2	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	1.0	2.0
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.9	4.1	1.1	0.2	1.0	1.0	1.2	0.9	1.3	1.2	0.3	1.1	1.0	1.2	0.7	3.6
DQ2	2.2	0.6	2.1	1.8	2.4	0.9	4.1	1.1	0.2	1.0	1.0	1.2	0.9	1.5	1.2	0.4	1.1	1.0	1.2	0.7	7.1
DQ8	2.1	0.6	2.1	1.8	2.5	0.8	4.5	1.5	0.7	1.1	1.0	2.2	1.0	2.7	1.2	0.3	1.1	1.0	1.2	0.6	4.7
DQ2/DQ8	2.3	0.6	2.2	1.9	2.7	0.8	4.5	1.5	0.6	1.2	1.0	1.9	0.9	2.7	1.1	0.3	1.1	1.0	1.1	0.8	2.9
All	2.2	0.6	2.1	1.8	2.5	0.8	4.5	1.4	0.6	1.1	1.0	1.5	0.9	2.7	1.2	0.3	1.1	1.0	1.2	0.6	7.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

Table 44. SI of MIP_BSA in demographics after insulin start

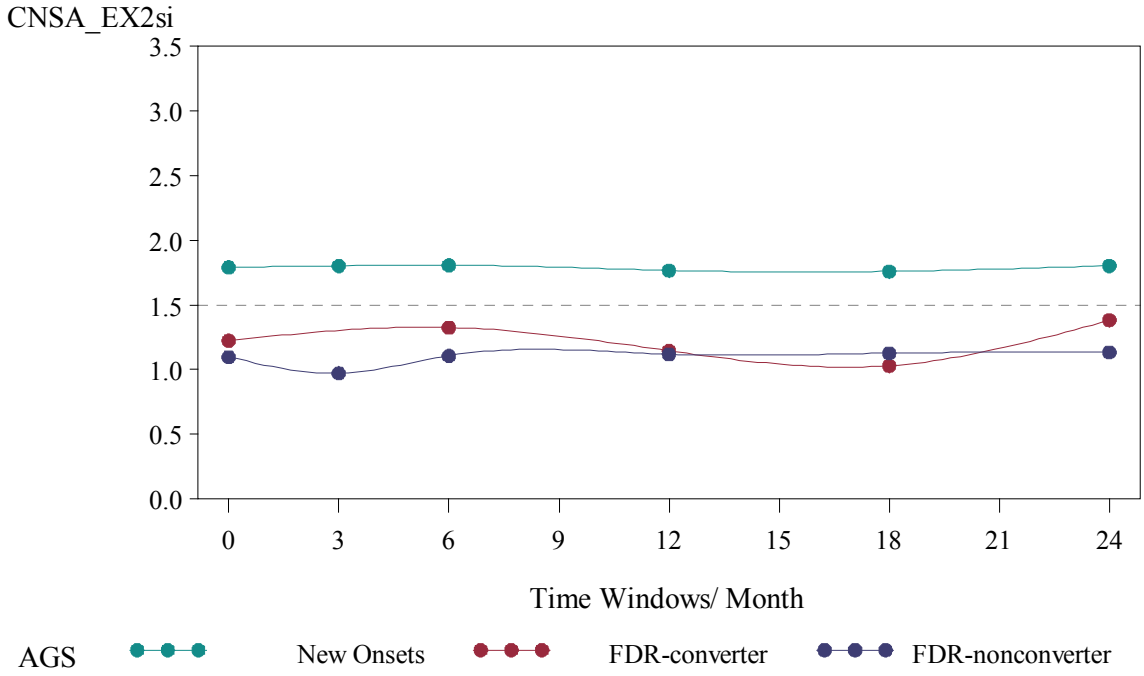
MIP_BSAsi	New Onsets							FDR-converter							FDR-nonconverter						
	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max	Mean	Std	Median	Q1	Q3	Min	Max
Gender																					
Male	2.2	0.6	2.1	1.8	2.5	0.8	4.5	2.0	0.7	2.0	1.4	2.5	1.1	2.9	1.2	0.4	1.1	1.0	1.2	0.6	7.1
Female	2.2	0.6	2.1	1.8	2.5	1.0	4.1	2.1	0.6	2.2	1.6	2.6	1.1	2.8	1.2	0.3	1.1	1.0	1.2	0.7	4.7
Race																					
White	2.2	0.6	2.1	1.8	2.5	0.8	4.5	2.1	0.6	2.2	1.6	2.5	1.1	2.9	1.2	0.3	1.1	1.0	1.2	0.6	7.1
Black	2.1	0.5	2.1	1.8	2.3	1.0	3.9	.	.	.	.	.	.	.	1.2	0.3	1.1	1.0	1.3	0.7	2.3
Asian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1.1	0.1	1.1	1.0	1.1	1.0	1.1
Other	2.3	0.7	2.3	2.0	2.8	1.0	3.2	.	.	.	.	.	.	.	1.4	0.4	1.2	1.1	1.8	1.0	2.0
HLA-DQ type																					
X	2.1	0.5	2.1	1.8	2.4	0.9	4.1	2.1	0.3	2.1	1.8	2.3	1.7	2.3	1.2	0.3	1.1	1.0	1.2	0.7	3.6
DQ2	2.2	0.6	2.1	1.8	2.4	0.9	4.1	1.6	0.7	1.2	1.1	2.3	1.1	2.3	1.2	0.4	1.1	1.0	1.2	0.7	7.1
DQ8	2.1	0.6	2.1	1.8	2.5	0.8	4.5	2.2	0.6	2.3	1.6	2.7	1.1	2.9	1.2	0.3	1.1	1.0	1.2	0.6	4.7
DQ2/DQ8	2.3	0.6	2.2	1.9	2.7	0.8	4.5	.	.	.	.	.	.	.	1.1	0.3	1.1	1.0	1.1	0.8	2.9
All	2.2	0.6	2.1	1.8	2.5	0.8	4.5	2.1	0.6	2.2	1.6	2.5	1.1	2.9	1.2	0.3	1.1	1.0	1.2	0.6	7.1

Std = standard deviation; Q1 = lower quartile; Q3 = uppper quartile

B.1.2 SI of ten test antigens over time by subpopulations

Based on the insulin start date, two subgroup of FDR-converters stratified, SI of ten test antigens plotted into two parts for one test antigens, which means FDR-converter with using insulin and FDR-converter without using insulin compared to other subpopulations separately.

SI of CNSA_EX2si overtime before insulin start in FDR-converter compare with other subpopulations



SI of CNSA_EX2si overtime after insulin start in FDR-converter compare with other subpopulations

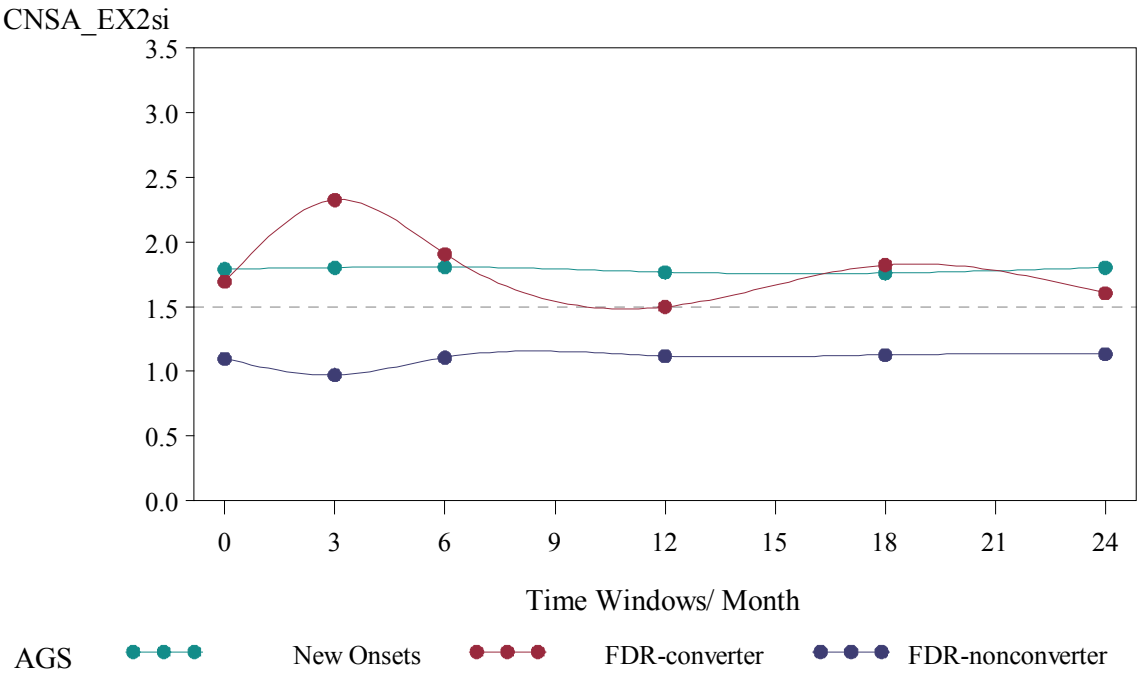
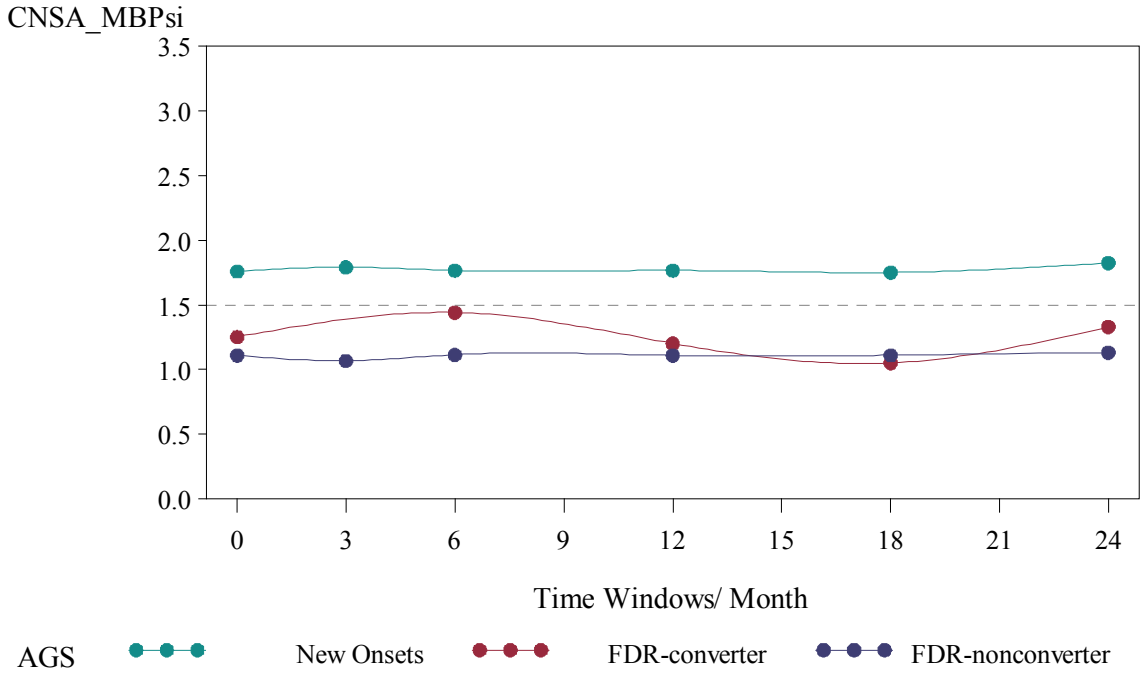


Figure 21. Chang in SI over time for CNSA_EX2

SI of CNSA_MBPsi overtime before insulin start in FDR-converter compare with other subpopulations



SI of CNSA_MBPsi overtime after insulin start in FDR-converter compare with other subpopulations

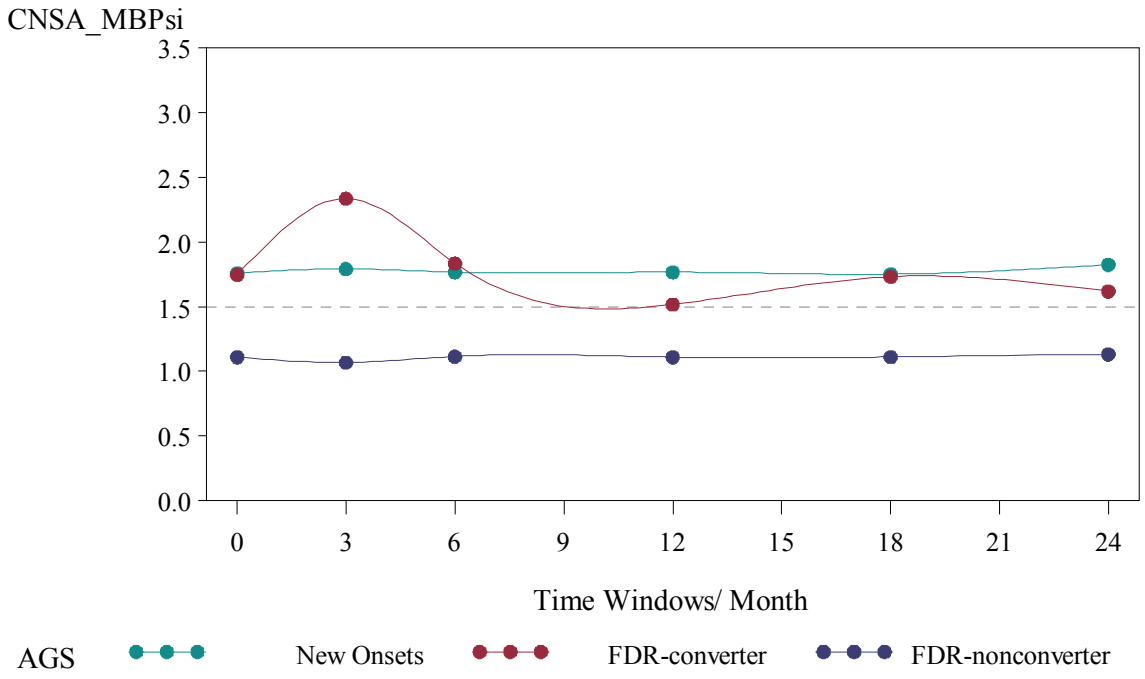
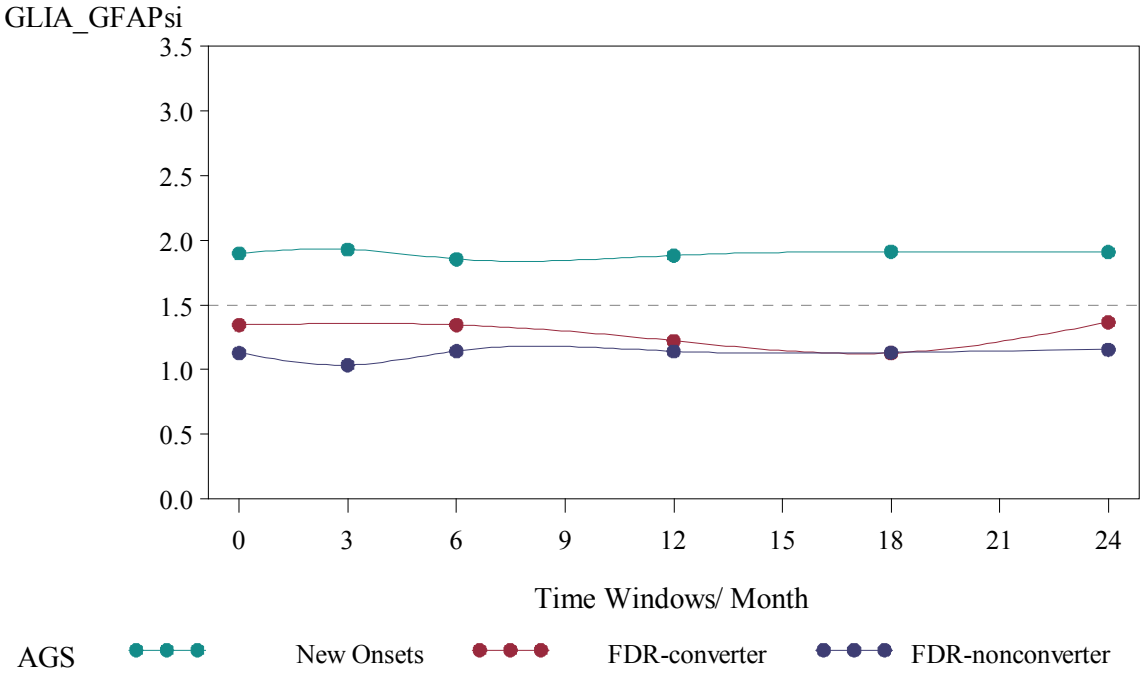


Figure 22. Chang in SI over time for CNSA_MBP

SI of GLIA_GFAPsi overtime before insulin start in FDR-converter compare with other subpopulations



SI of GLIA_GFAPsi overtime after insulin start in FDR-converter compare with other subpopulations

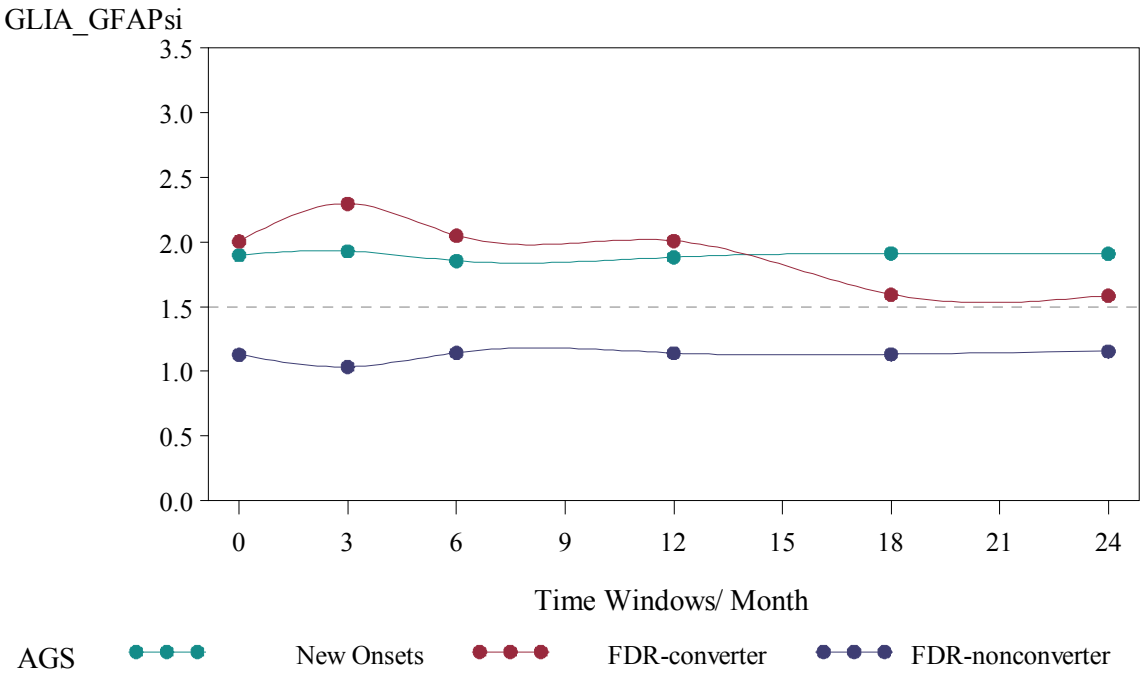
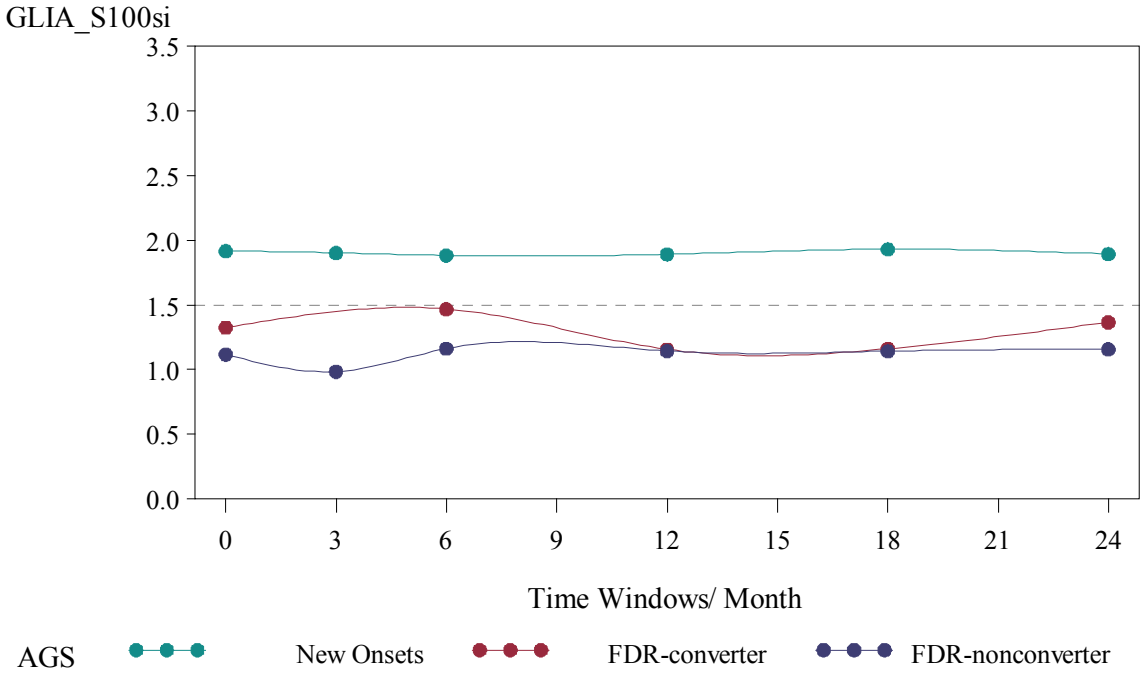


Figure 23. Chang in SI over time for GLIA_GFAP

SI of GLIA_S100si overtime before insulin start in FDR-converter compare with other subpopulations



SI of GLIA_S100si overtime after insulin start in FDR-converter compare with other subpopulations

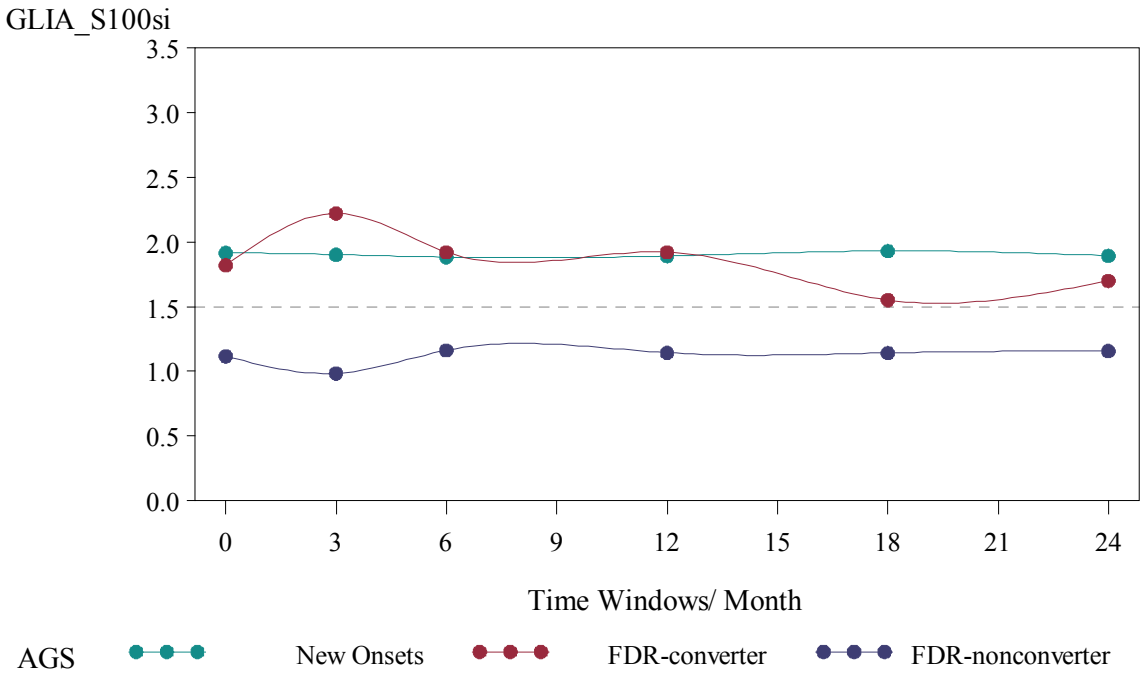
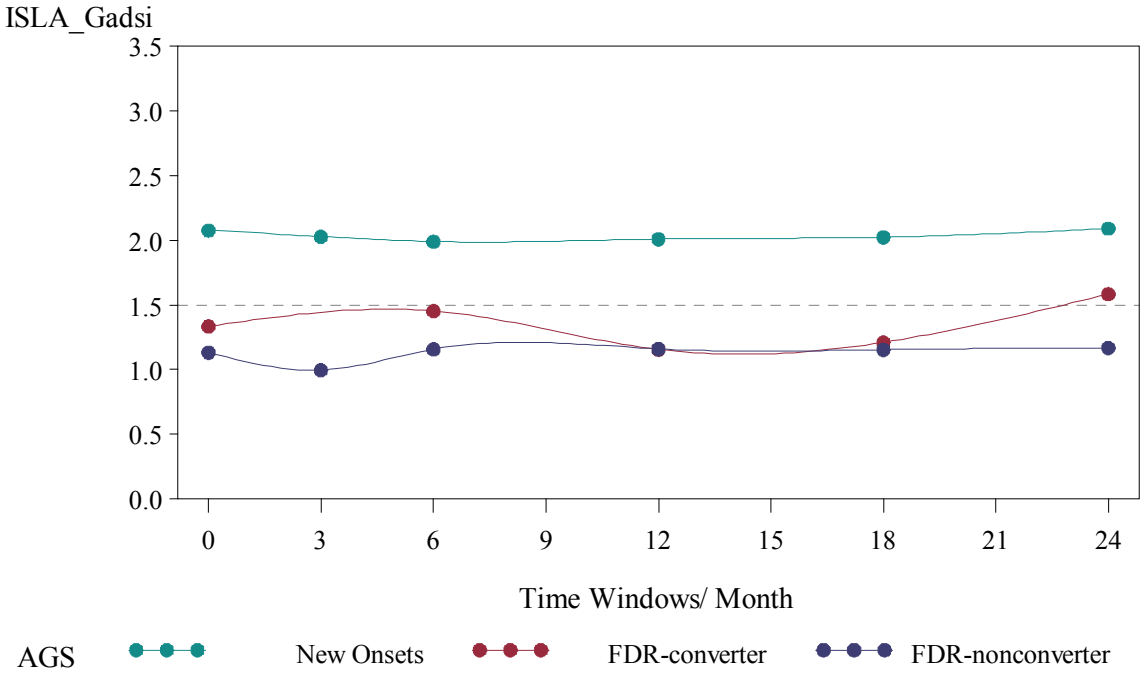


Figure 24. Chang in SI over time for GLIA_S100

SI of ISLA_Gadsi overtime before insulin start in FDR-converter compare with other subpopulations



SI of ISLA_Gadsi overtime after insulin start in FDR-converter compare with other subpopulations

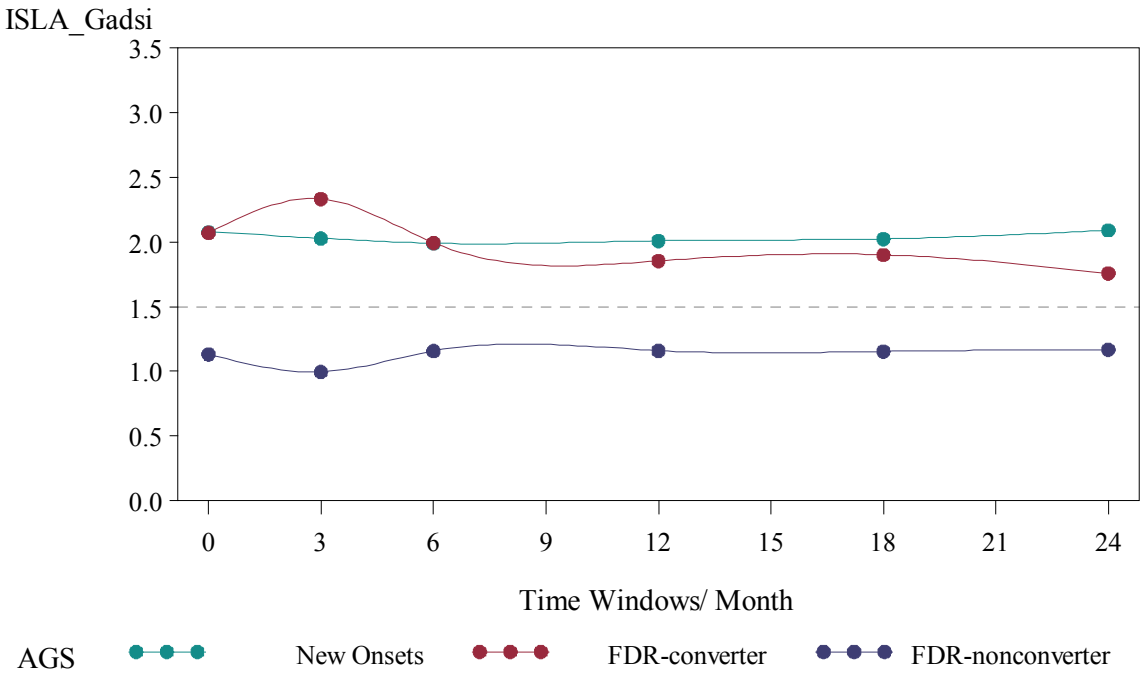
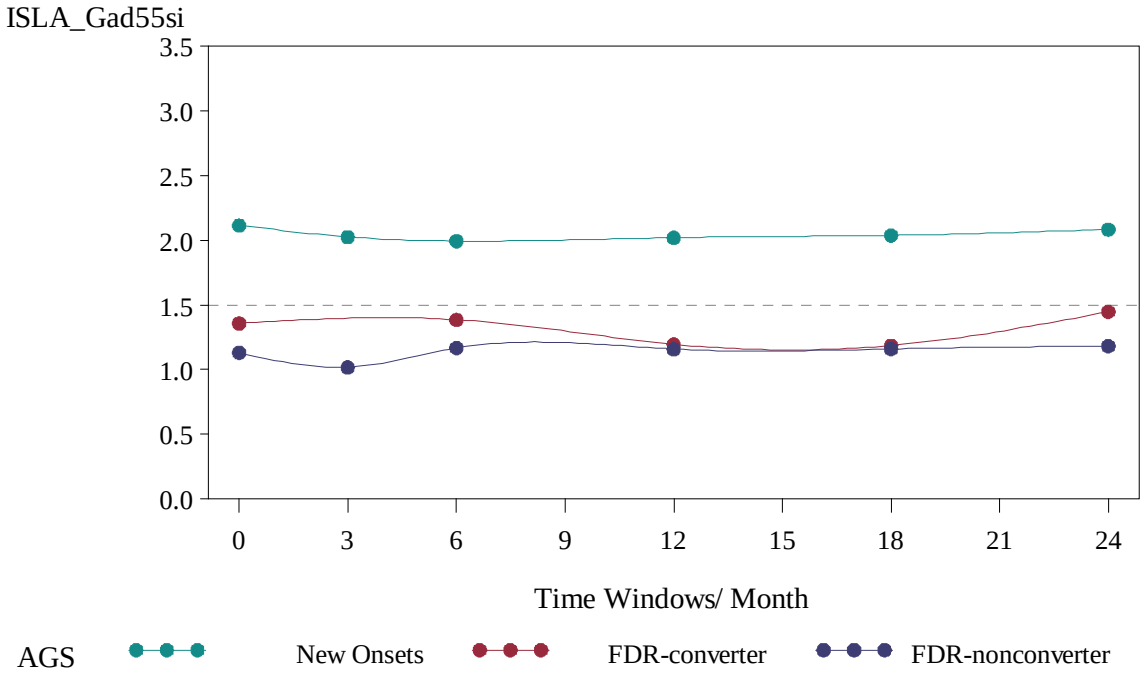


Figure 25. Chang in SI over time for ISLA_Gad

SI of ISLA_Gad55si overtime before insulin start in FDR-converter compare with other subpopulations



SI of ISLA_Gad55si overtime after insulin start in FDR-converter compare with other subpopulations

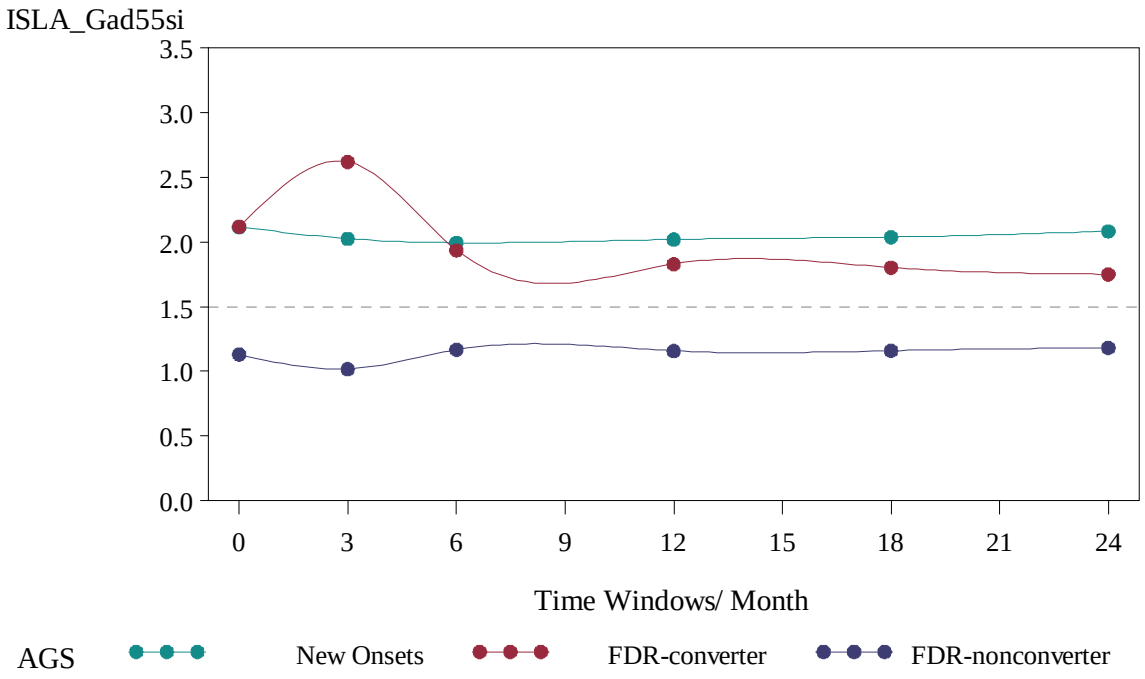
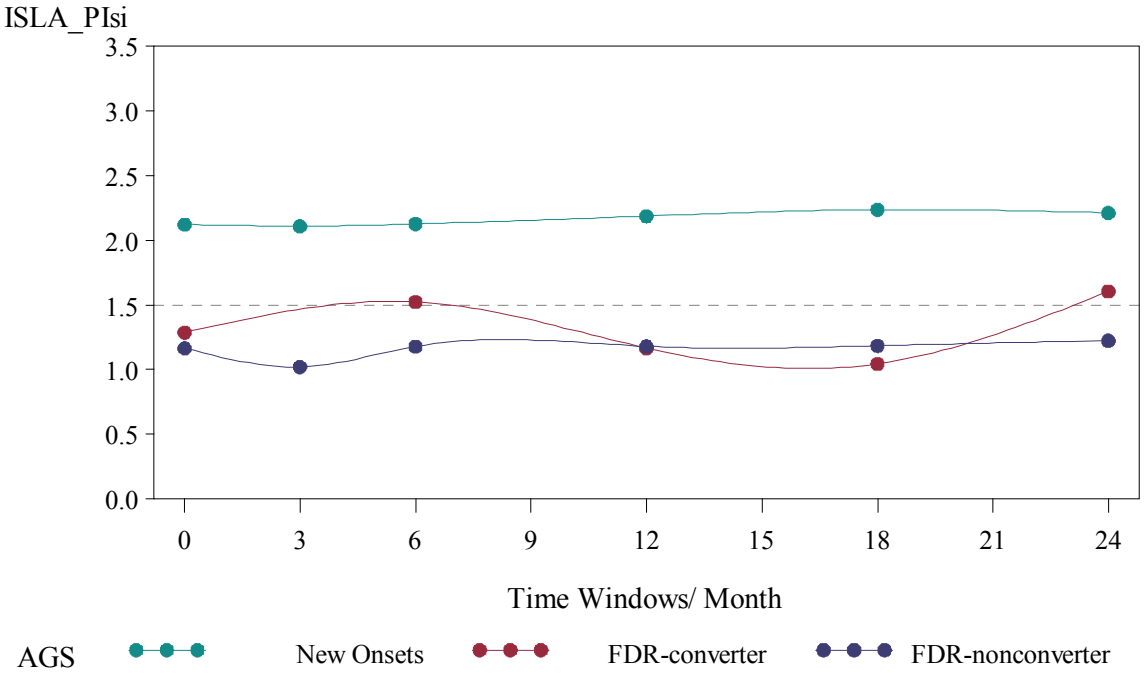


Figure 26. Chang in SI over time for ISLA_Gad55

SI of ISLA_PIsi overtime before insulin start in FDR-converter compare with other subpopulations



SI of ISLA_PIsi overtime after insulin start in FDR-converter compare with other subpopulations

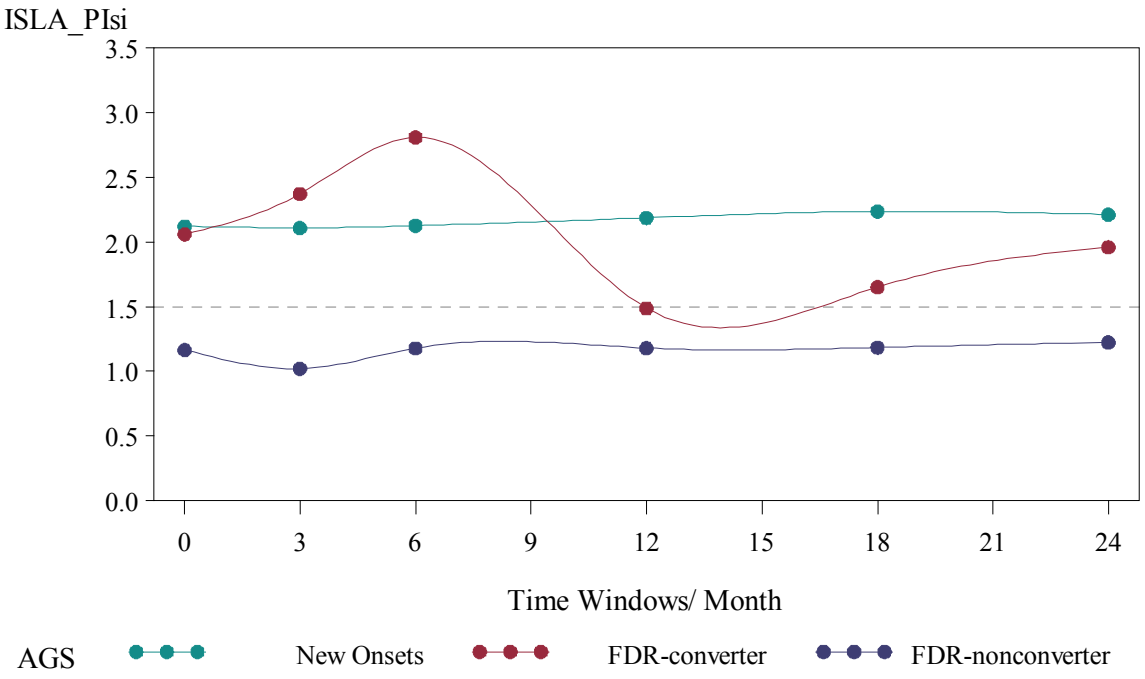
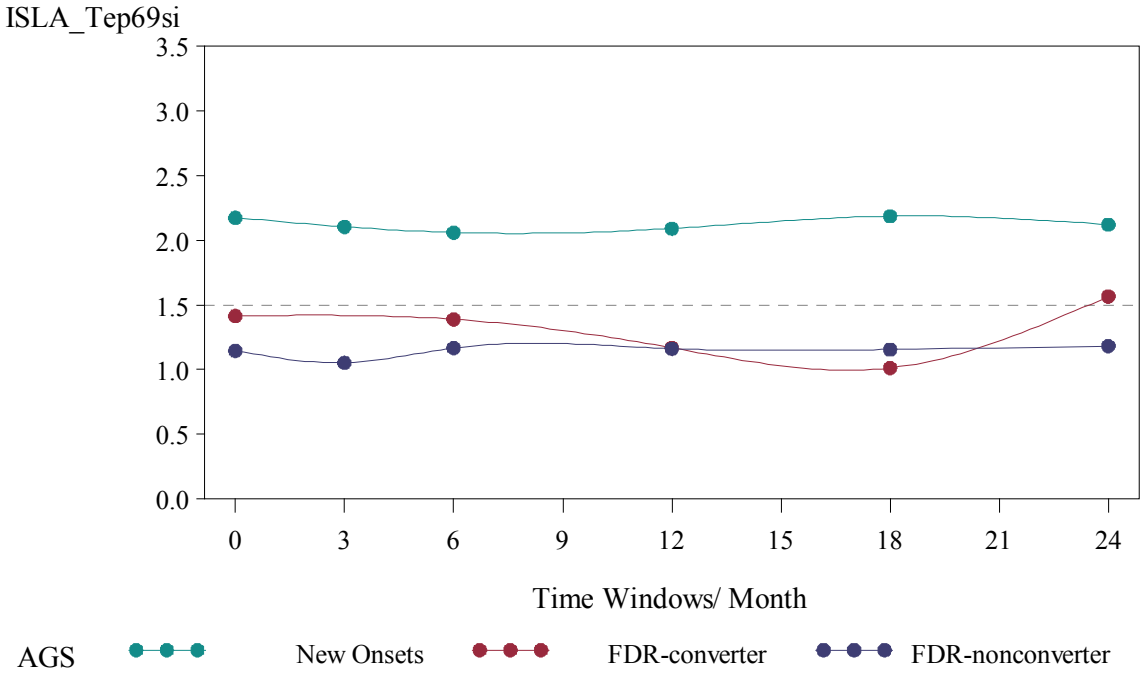


Figure 27. Chang in SI over time for ISLA_PI

SI of ISLA_Tep69si overtime before insulin start in FDR-converter compare with other subpopulations



SI of ISLA_Tep69si overtime after insulin start in FDR-converter compare with other subpopulations

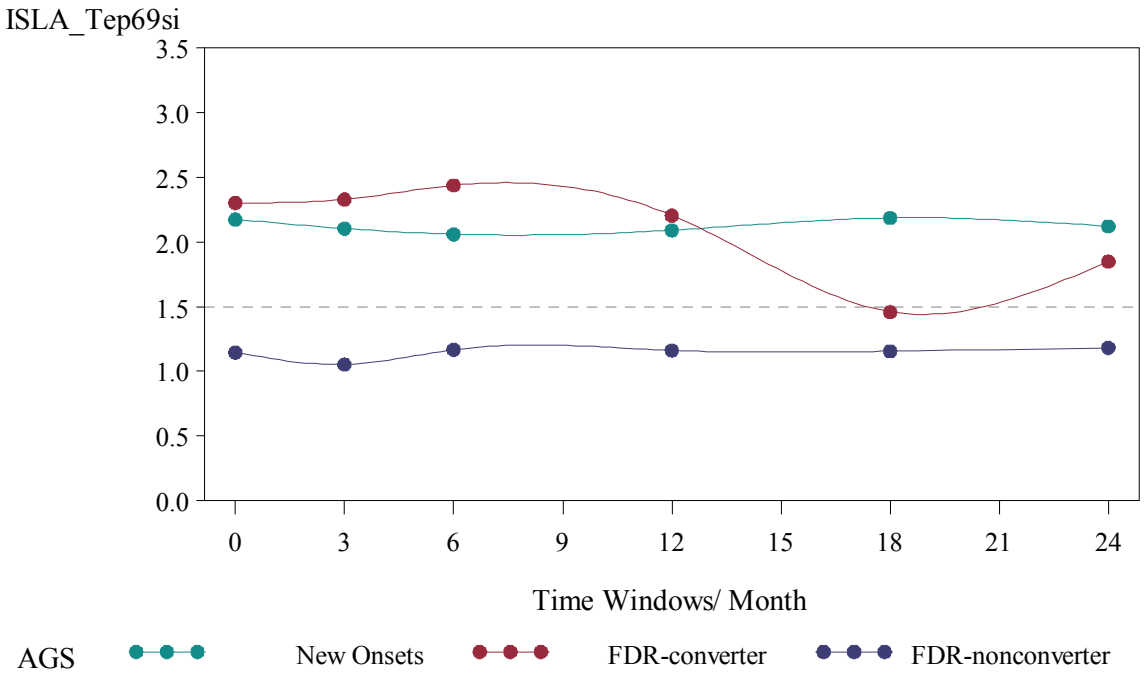
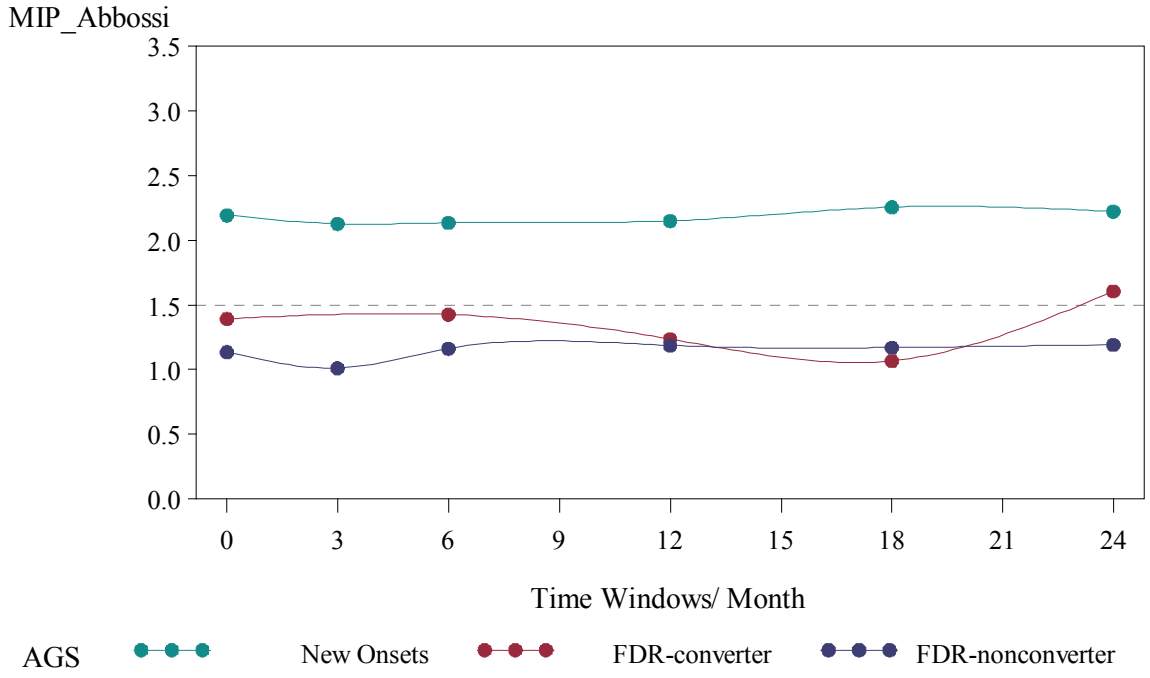


Figure 28. Chang in SI over time for ISLA_Tep69

SI of MIP_Abbossi overtime before insulin start in FDR-converter compare with other subpopulations



SI of MIP_Abbossi overtime after insulin start in FDR-converter compare with other subpopulations

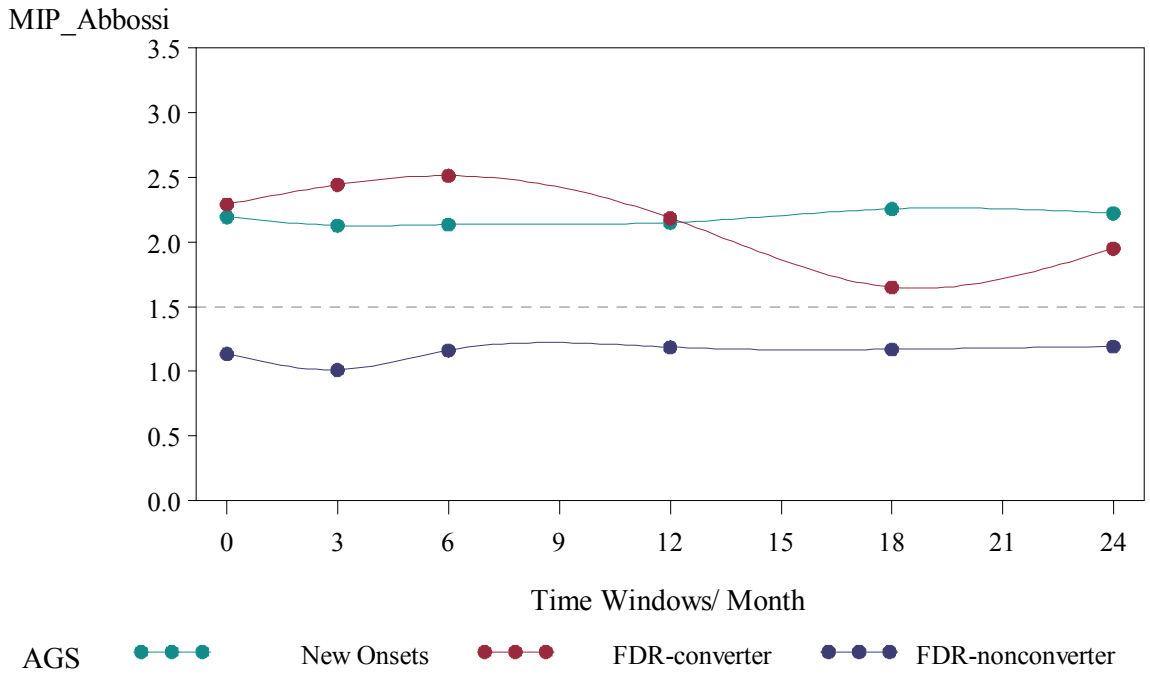
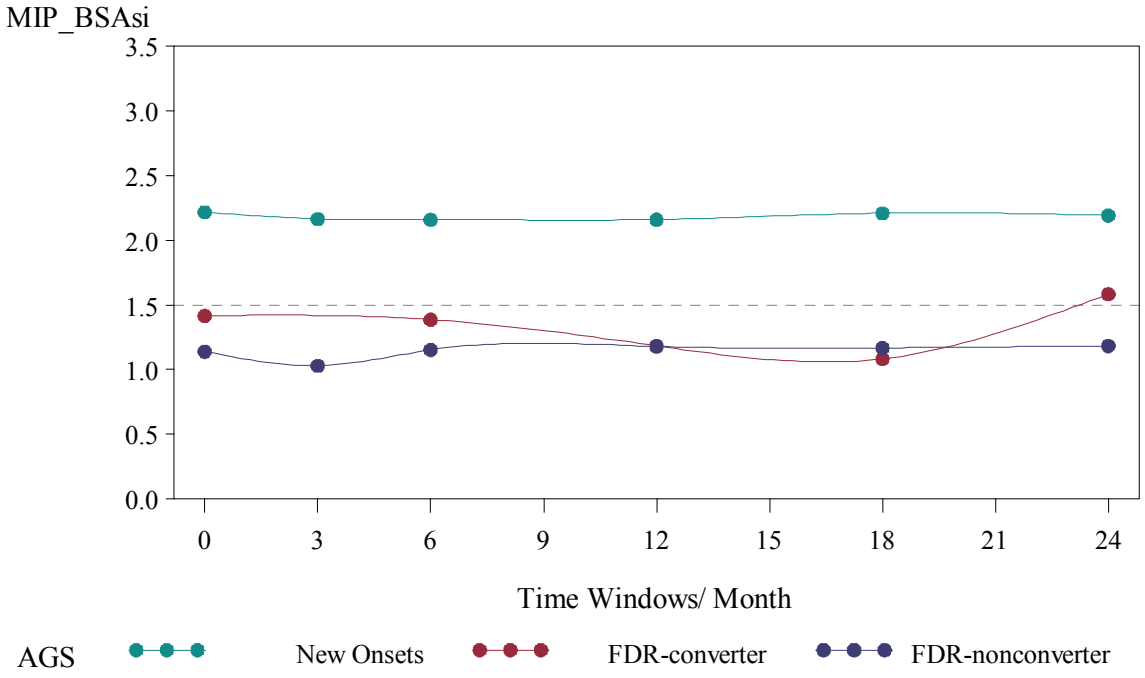


Figure 29. Chang in SI over time for MIP_Abbos

SI of MIP_BSAsi overtime before insulin start in FDR-converter compare with other subpopulations



SI of MIP_BSAsi overtime after insulin start in FDR-converter compare with other subpopulations

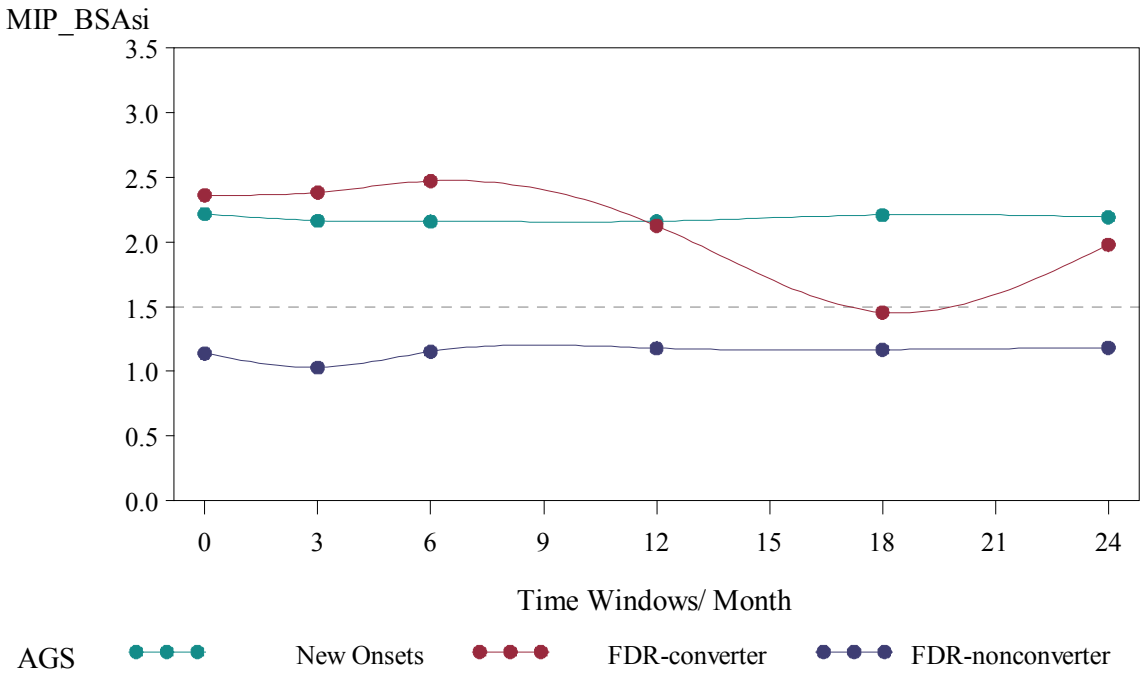


Figure 30. Chang in SI over time for MIP_BSA

B.1.3 Optimal cutoff for ten test antigens

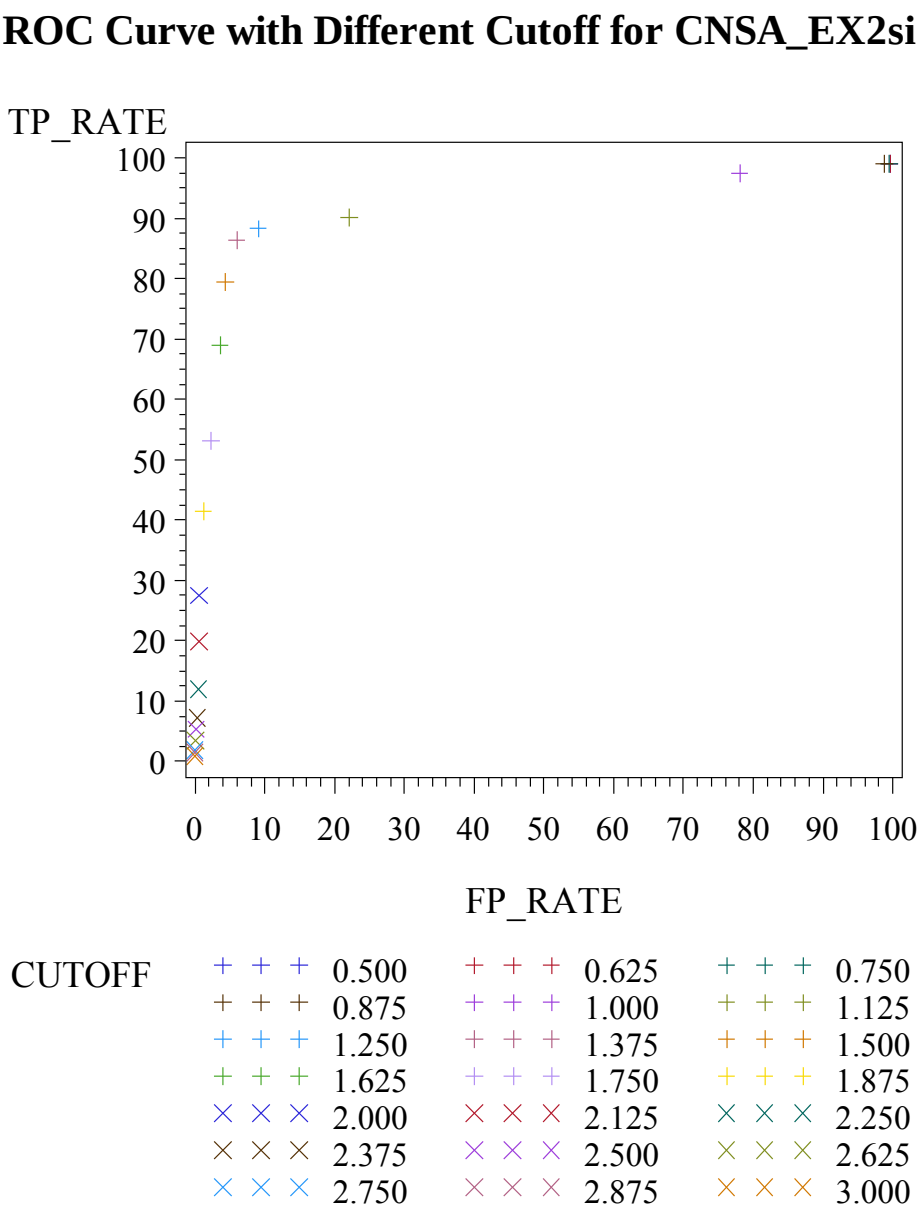


Figure 31. ROC curve at different cutoffs for CNSA_EX2

ROC Curve with Different Cutoff for CNSA_MBPsi

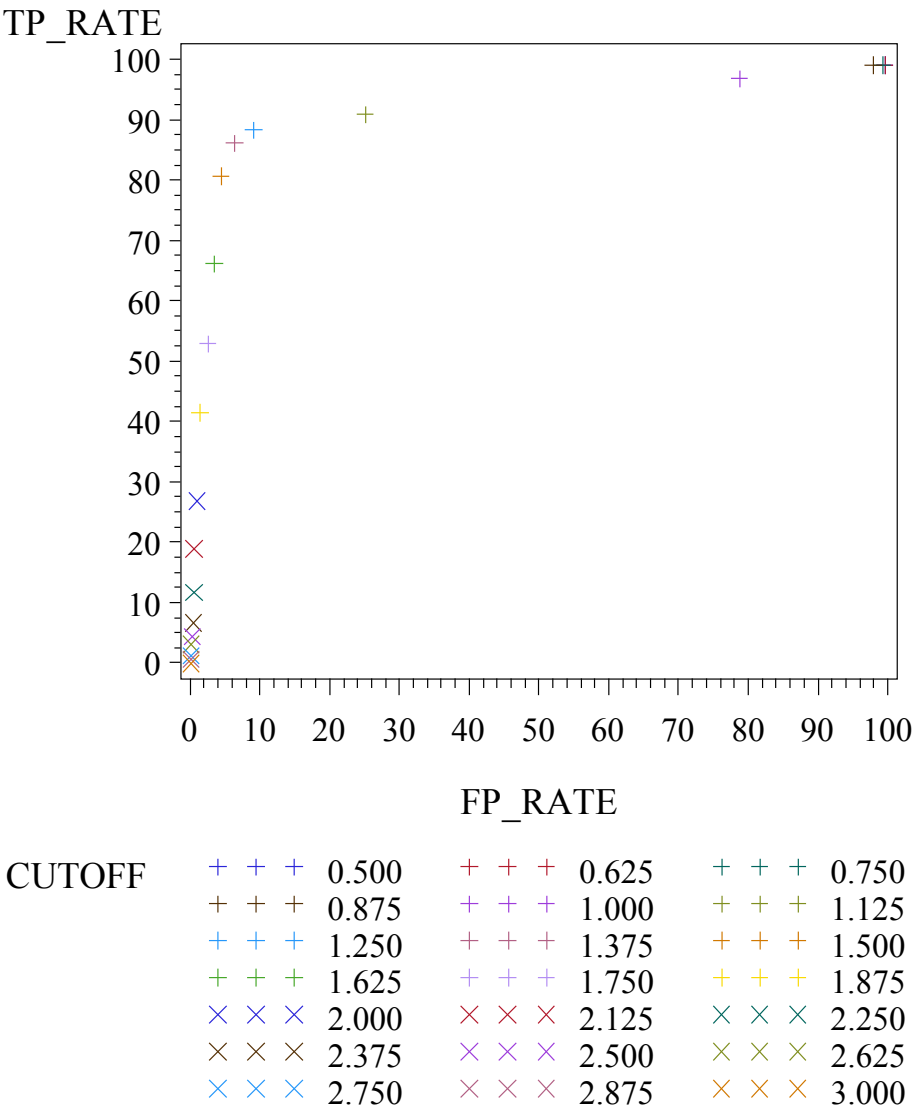


Figure 32. ROC curve at different cutoffs for CNSA_MBP

ROC Curve with Different Cutoff for GLIA_GFAPsi

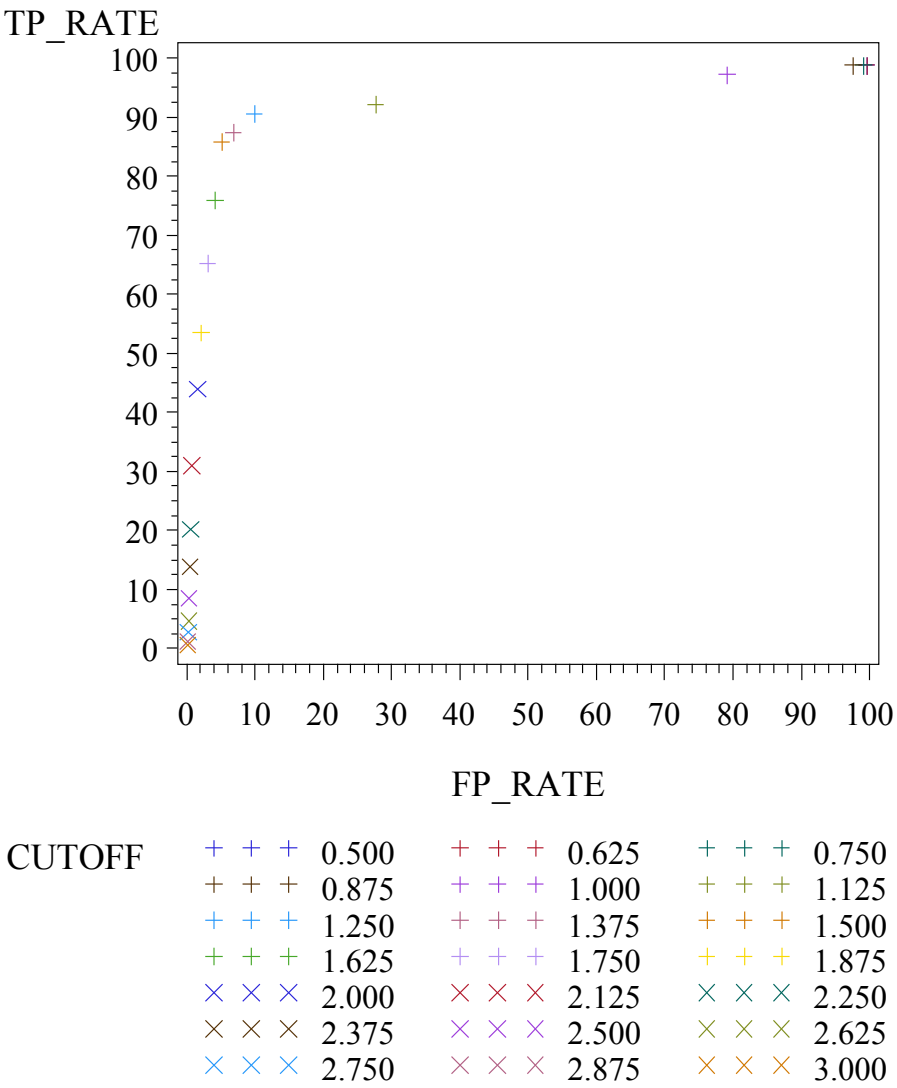


Figure 33. ROC curve at different cutoffs for GLIA_GFAP

ROC Curve with Different Cutoff for GLIA_S100si

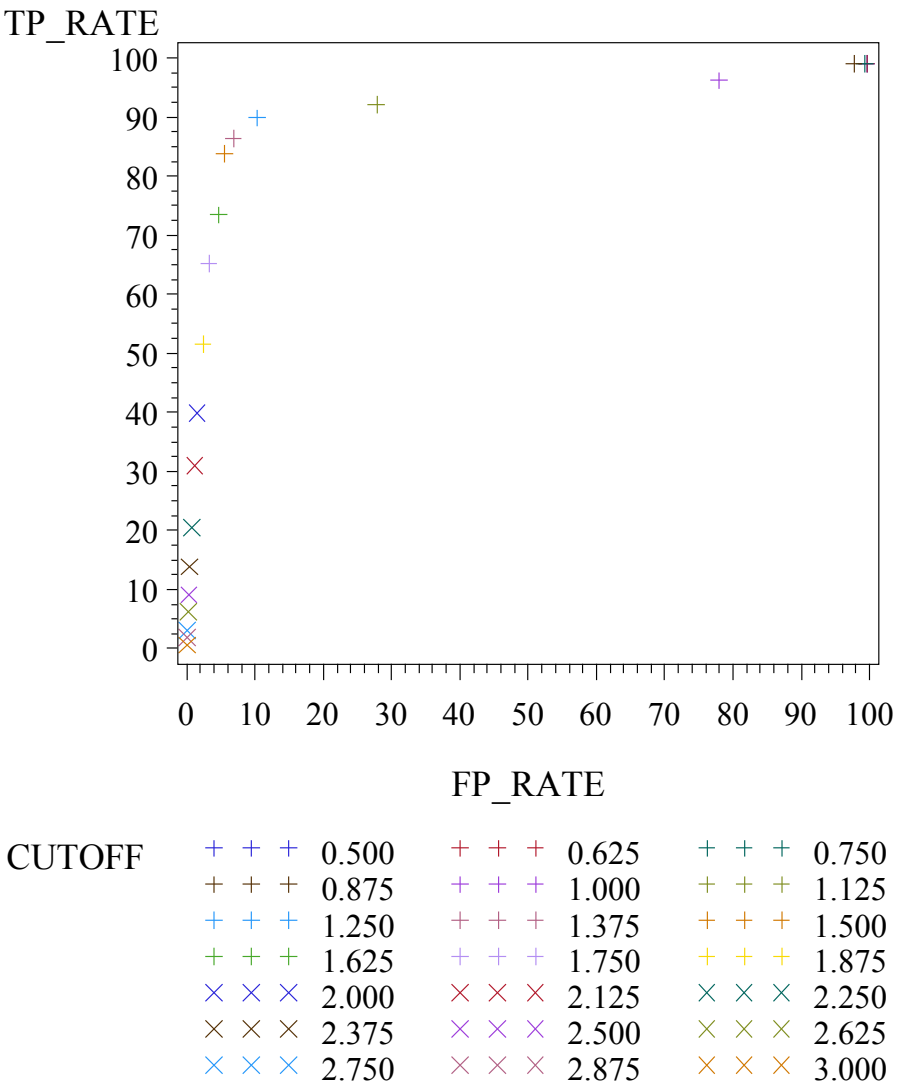


Figure 34. ROC curve at different cutoffs for GLIA_S100

ROC Curve with Different Cutoff for ISLA_Gadsi

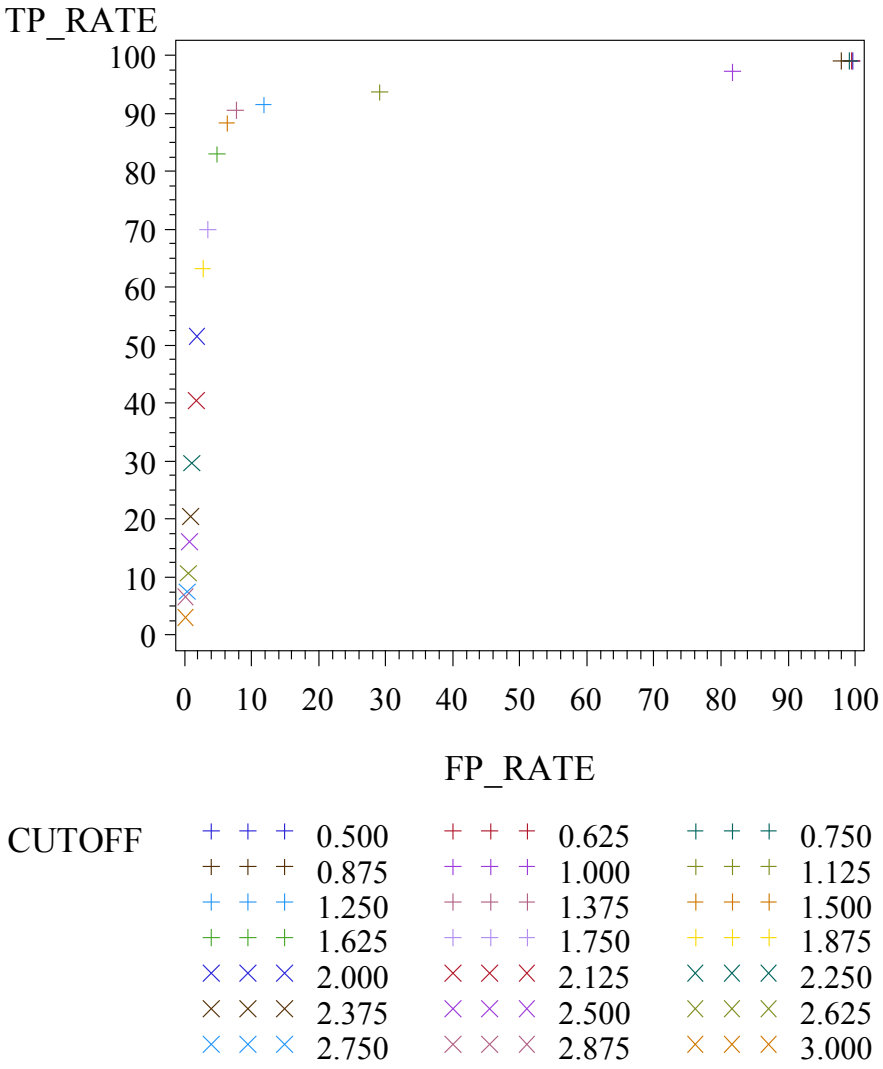


Figure 35. ROC curve at different cutoffs for ISLA_Gad

ROC Curve with Different Cutoff for ISLA_Gad55si

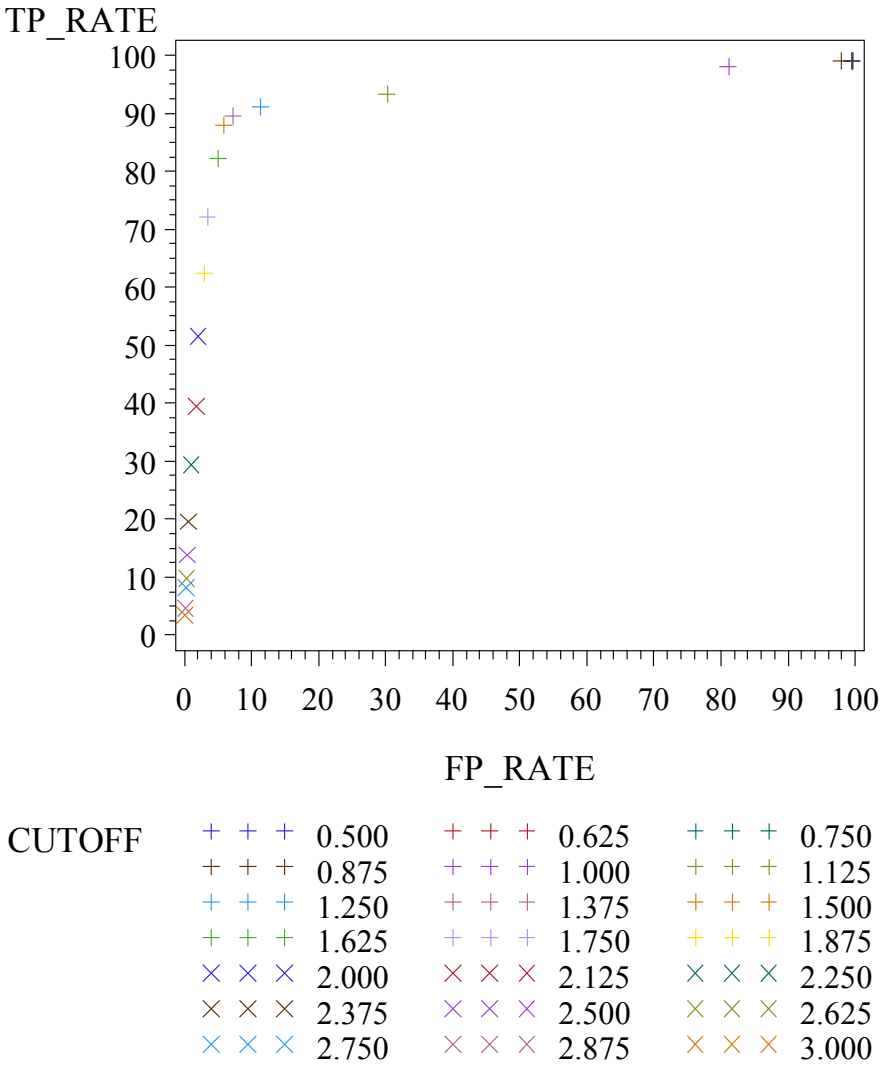


Figure 36. ROC curve at different cutoffs for ISLA_Gad55

ROC Curve with Different Cutoff for ISLA_PIsi

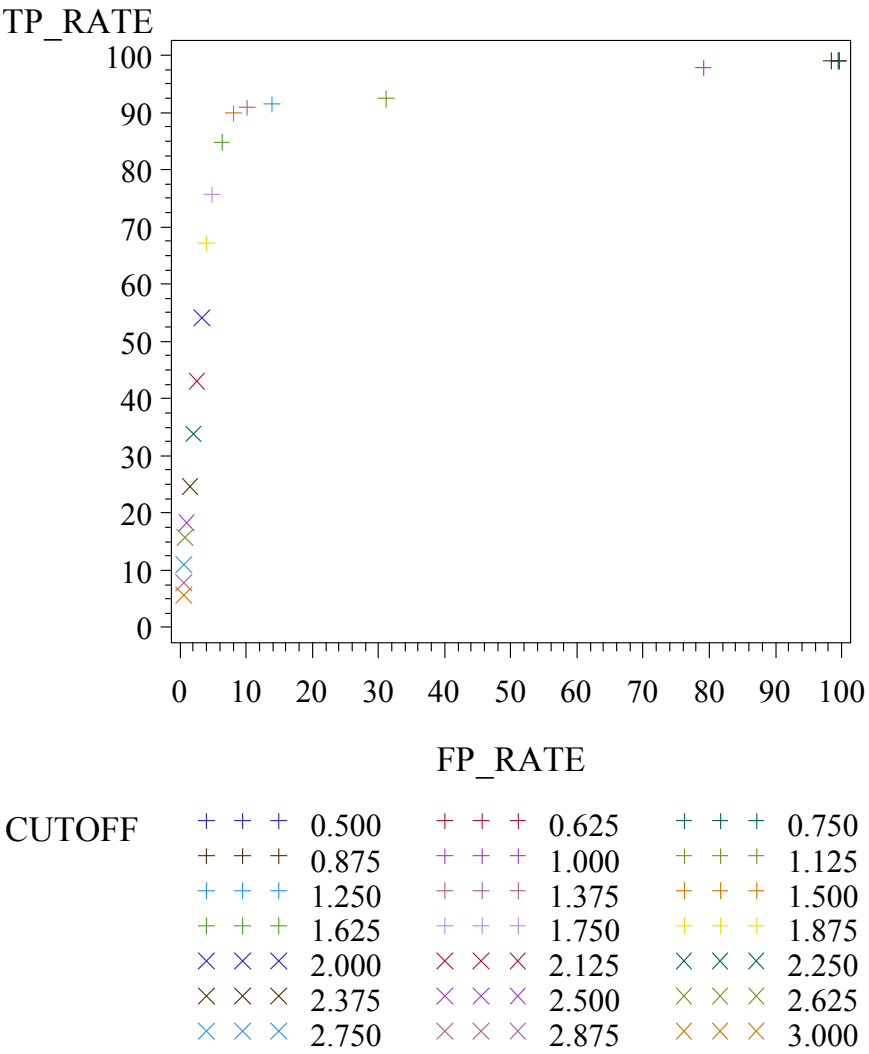


Figure 37. ROC curve at different cutoffs for ISLA_PI

ROC Curve with Different Cutoff for ISLA_Tep69si

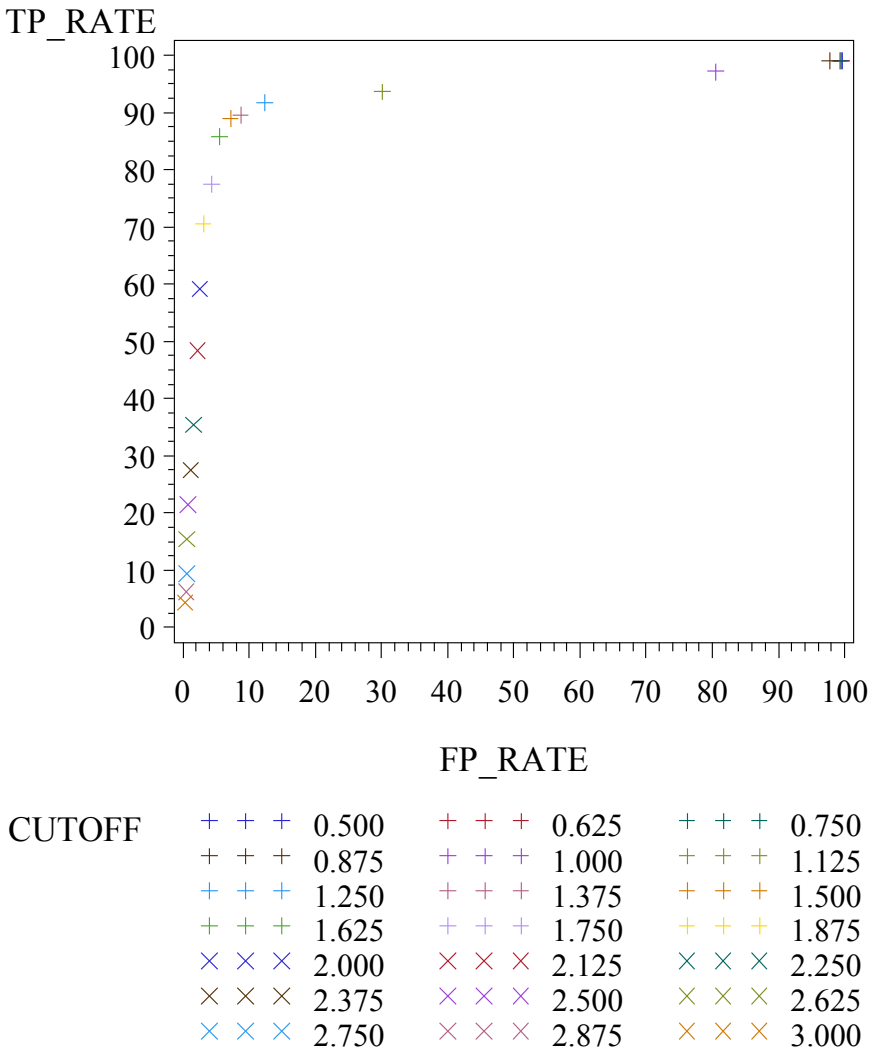


Figure 38. ROC curve at different cutoffs for ISLA_Tep69

ROC Curve with Different Cutoff for MIP_Abbossi

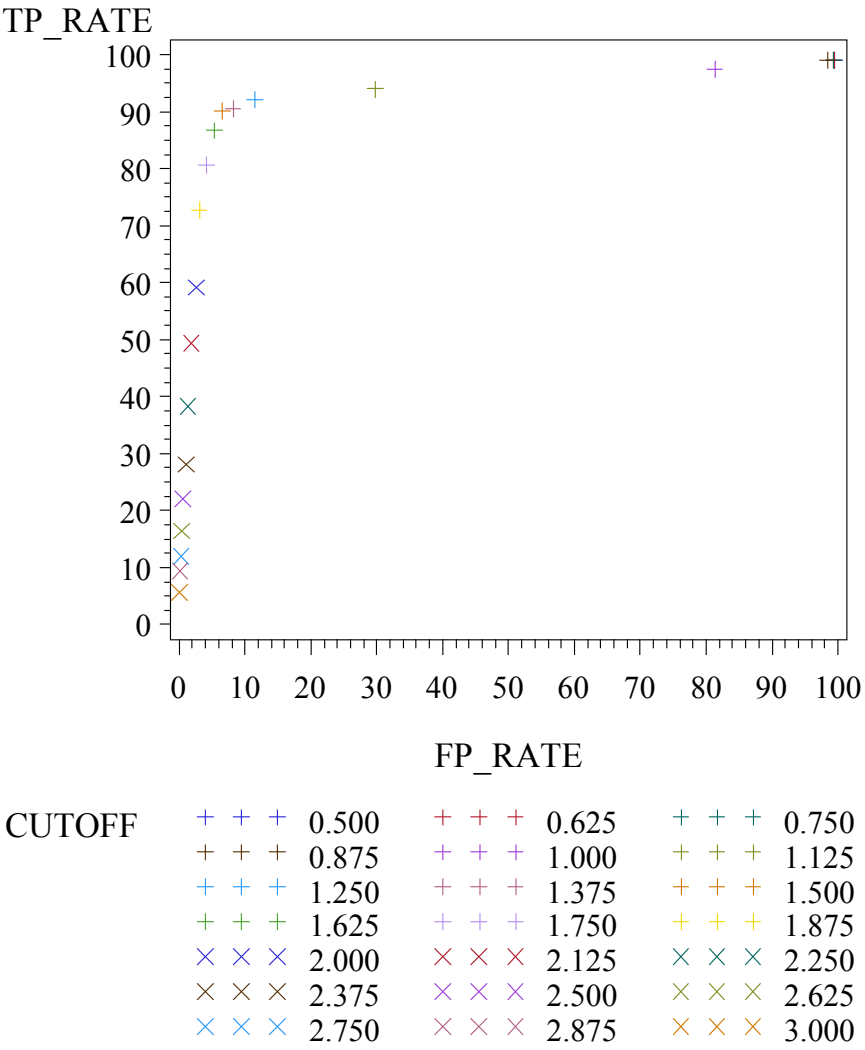


Figure 39. ROC curve at different cutoffs for MIP_Abbos

ROC Curve with Different Cutoff for MIP_BSAsi

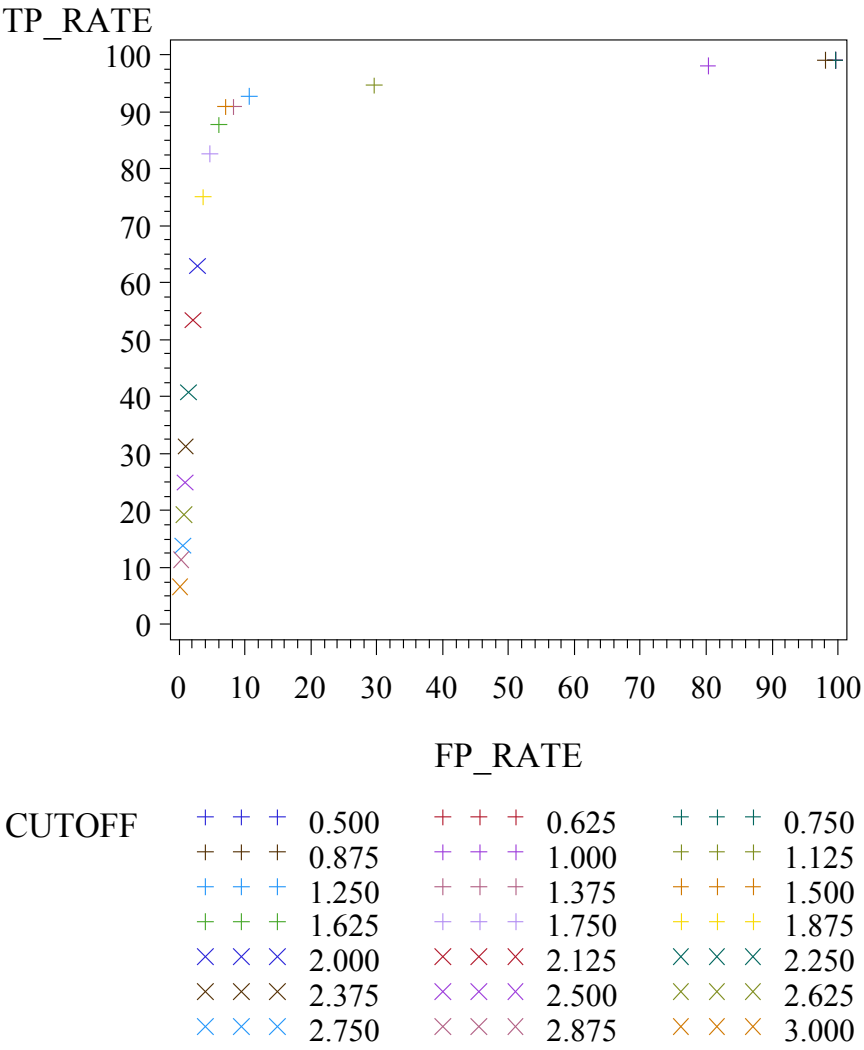


Figure 40. ROC curve at different cutoffs for MIP_BSA

BIBLIOGRAPHY

[1] Cooke, D. W., and L. Plotnick. Type 1 diabetes mellitus in pediatrics. Pediatrics in Review. 2008 Nov;29(11):374-84; quiz 385. Review.

[2] Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010 Jan;87(1):4-14.

[3] Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.

[4] Daneman D. Type 1 diabetes. Lancet. 2006 Mar 11;367(9513):847-58. Review.

[5] Soltesz G, Patterson CC, Dahlquist G; EURODIAB Study Group. Worldwide childhood type 1 diabetes incidence--what can we learn from epidemiology? Pediatr Diabetes. 2007 Oct;8 Suppl_6:6-14.

[6] Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Akerblom HK. Environmental triggers and determinants of type 1 diabetes. Diabetes. 2005 Dec;54 Suppl_2:S125-36. Review.

[7] Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens. 2004 Dec;64(6):631-49. Review.

[8] Janeway CA Jr, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. The major histocompatibility complex and its functions. Retrieved Aug. 8th, 2012 from: http://www.ncbi.nlm.nih.gov/books/NBK27156/

[9] Xie M, Li J, Jiang T. Accurate HLA type inference using a weighted similarity graph. BMC Bioinformatics. 2010 Dec 14;11 Suppl 11:S10.

[10] Louka AS, Nilsson S, Olsson M, Talseth B, Lie BA, Ek J, Gudjónsdóttir AH, Ascher H, Sollid LM. HLA in coeliac disease families: a novel test of risk modification by the 'other' haplotype when at least one DQA1*05-DQB1*02 haplotype is carried. Tissue Antigens. 2002 Aug;60(2):147-54.

[11] Owerbach D, Lernmark A, Platz P, Ryder LP, Rask L, Peterson PA, Ludvigsson J. HLA-D region beta-chain DNA endonuclease fragments differ between HLA-DR identical healthy and insulin-dependent diabetic individuals. Nature. 1983 Jun 30;303(5920):815-7.

[12] Cohen-Haguenauer O, Robbins E, Massart C, Busson M, Deschamps I, Hors J, Lalouel JM, Dausset J, Cohen D. A systematic study of HLA class II-beta DNA restriction fragments in insulin-dependent diabetes mellitus. Proc Natl Acad Sci U S A. 1985 May;82(10):3335-9.

[13] Festenstein H, Awad J, Hitman GA, Cutbush S, Groves AV, Cassell P, Ollier W, Sachs JA. New HLA DNA polymorphisms associated with autoimmune diseases. Nature. 1986 Jul 3-9;322(6074):64-7. PubMed PMID: 3014346.

[14] Nepom BS, Palmer J, Kim SJ, Hansen JA, Holbeck SL, Nepom GT. Specific genomic markers for the HLA-DQ subregion discriminate between DR4+ insulin-dependent diabetes mellitus and DR4+ seropositive juvenile rheumatoid arthritis. J Exp Med. 1986 Jul 1;164(1):345-50.

[15] Horton V, Stratton I, Bottazzo GF, Shattock M, Mackay I, Zimmet P, Manley S, Holman R, Turner R. Genetic heterogeneity of autoimmune diabetes: age of presentation in adults is influenced by HLA DRB1 and DQB1 genotypes (UKPDS 43). UK Prospective Diabetes Study (UKPDS) Group. Diabetologia. 1999 May;42(5):608-16.

[16] Bakhtadze E, Borg H, Stenström G, Fernlund P, Arnqvist HJ, Ekbom-Schnell A, Bolinder J, Eriksson JW, Gudbjörnsdottir S, Nyström L, Groop LC, Sundkvist G. HLA-DQB1 genotypes, islet antibodies and beta cell function in the classification of recent-onset diabetes among young adults in the nationwide Diabetes Incidence Study in Sweden. Diabetologia. 2006 Aug;49(8):1785-94.

[17] Bottazzo GF, Florin-Christensen A, Doniach D. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. Lancet. 1974 Nov 30;2(7892):1279-83.

[18] Notkins AL, Lernmark A. Autoimmune type 1 diabetes: resolved and unresolved issues. J Clin Invest. 2001 Nov;108(9):1247-52.

[19] Bonifacio E, Genovese S, Braghi S, Bazzigaluppi E, Lampasona V, Bingley PJ, Rogge L, Pastore MR, Bognetti E, Bottazzo GF, et al. Islet autoantibody markers in IDDM: risk assessment strategies yielding high sensitivity. Diabetologia. 1995 Jul;38(7):816-22.

[20] Sabbah E, Savola K, Kulmala P, Veijola R, Vähäsalo P, Karjalainen J, Akerblom HK, Knip M. Diabetes-associated autoantibodies in relation to clinical characteristics and natural course in children with newly diagnosed type 1 diabetes. The Childhood Diabetes In Finland Study Group. J Clin Endocrinol Metab. 1999 May;84(5):1534-9.

[21] Strebelow M, Schlosser M, Ziegler B, Rjasanowski I, Ziegler M. Karlsburg Type I diabetes risk study of a general population: frequencies and interactions of the four major Type I

diabetes-associated autoantibodies studied in 9419 schoolchildren. Diabetologia. 1999 Jun;42(6):661-70.

[22] Winter WE, Harris N, Schatz D. Type 1 diabetes islet autoantibody markers. Diabetes Technol Ther. 2002;4(6):817-39. Review.

[23] Holmberg H, Vaarala O, Sadauskaite-Kuehne V, Ilonen J, Padaiga Z, Ludvigsson J. Higher prevalence of autoantibodies to insulin and GAD65 in Swedish compared to Lithuanian children with type 1 diabetes. Diabetes Res Clin Pract. 2006 Jun;72(3):308-14.

[24] Williams AJ, Norcross AJ, Dix RJ, Gillespie KM, Gale EA, Bingley PJ. The prevalence of insulin autoantibodies at the onset of Type 1 diabetes is higher in males than females during adolescence. Diabetologia. 2003 Oct;46(10):1354-6.

[25] Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, Rewers M, Eisenbarth GS, Jensen J, Davidson HW, Hutton JC. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci USA. 2007 Oct 23;104(43):17040-5.

[26] Dosch H, Cheung RK, Karges W, Pietropaolo M, Becker DJ. Persistent T cell anergy in human type 1 diabetes. J Immunol. 1999 Dec 15;163(12):6933-40.

[27] Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994 Jul 9;309(6947):102.

[28] Swets, J.A., Dawes, R.M., Monahan, J. Better decisions through science. Scientific American. 2000; 283, 82–87.

[29] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993 Apr;39(4):561-77. Review.

[30] Pepe, Margaret S. The Statistical Evaluation of Medical Tests for Classification and Prediction. New York, NY: Oxford University Press 2003.

[31] Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988; 240: 1285-93.

[32] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988 Sep;44(3):837-45.