Vβ EXPRESSION ANALYSIS IN TYPE 1 DIABETES POPULATION AND THEIR UNAFFECTED FIRST DEGREE RELATIVES: A PRELIMINARY ANALYSIS

By

Mei Han

B.S., Beijing Normal University, China, 1984

M.S., Renmin University of China, China, 2002

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Mei Han

It was defended on

August 8, 2008

and approved by

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

Committee Member:
Patrizia Luppi, MD
Department of Pediatrics
Children's Hospital of Pittsburgh
University of Pittsburgh Medical Center

Committee Member:
Ingrid M Libman, MD
Department of Pediatrics
Children's Hospital of Pittsburgh
University of Pittsburgh Medical Center

Vβ EXPRESSION ANALYSIS IN TYPE 1 DIABETES POPULATION AND THEIR UNAFFECTED FIRST DEGREE RELATIVES: A PRELIMINARY ANALYSIS

Mei Han, M.S.

University of Pittsburgh, 2008

Type 1 diabetes (T1D) is an autoimmune disease and is one of the most common diseases of children. It has a significant impact on public health costing millions of dollars in health care every year. Some studies have found that diabetes-associated T-cell receptor (TCR) bias can be detected in the peripheral circulation, and precede the onset of T1D by years. Therefore confirmation of this postulate would identify an acceleration time point that could be targeted by intervention strategies.

The overall purpose of the study is to investigate Vβ expressions in T1D. This thesis is a preliminary analysis to explore the basic relationships among the various data points and provides direction for the investigators. Results show that Vβ1 and Vβ13.1 expressions for new onsets were correlated with the number of days since diagnosis. Significantly higher values were observed within 7 days of diagnosis compared to those measured further away. In comparison to FDRs, new onsets with blood drawn within 7 days of diagnosis showed higher Vβ1 values. Male FDRs had higher Vβ1 expression than female FDRs. The Vβ7 expression among black new onsets was significantly higher than for whites. Black FDRs showed higher Vβ1 expression and lower Vβ13.1 expression than white FDRs. Our data suggests that Vβ expression does not correlate with antibody status (GAD, IA2, IAA, and ICA) or the number of positive antibodies. Moreover, new onsets and FDRs show no significant differences in Vβ expression among individuals with different non-ASP status, DQ2 and DQ8 alleles. Logistic regression models

that included all FDRs and new onsets with Vβ data available within 7 days of diagnosis suggest that Vβ1, Vβ13.1, DQ8 and age are associated with the onset of diabetes. Thus, Vβ expression is an independent predictor of T1D.

We recommend that when designing studies to assess Vβ expression and analyzing the resultant data, it is important to consider the timing of the blood draws and the number of days since diagnosis. Our evidence suggests that Vβ expression is higher at onset and then decreases over time and needs to be taken into account in the design stage and the analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like first and foremost to express my sincere gratitude to my thesis advisor, Dr. Vincent C. Arena, for all the help and guidance he has provided to me during the work on this thesis. I am so grateful to him for the opportunity to be involved in "The Etiology and Epidemiology of Type 1 Diabetes" study and to work with him and the research team. I appreciate all of the time and resources he extended to me along the way.

I would like to thank Dr. Patrizia Luppi and Dr. Ingrid M Libman for their efforts as well as their encouragement and suggestions. They not only assisted me with their excellent advice, but also elevated my interest in human genetics.

I would like to acknowledge Dr. Dorothy Becker, the project principal investigator, for allowing me to use her study data.

Finally, I would like to dedicate this thesis to my husband, my son, my parents as well as my brother and sister for their unwavering support and love. This would not have been possible without them.

1.0 INTRODUCTION

1.1 DIABETES OVERVIEW

Diabetes mellitus type 1 (T1DM) has been known as Insulin Dependent Diabetes Mellitus (IDDM). Because Type 1 diabetes is usually diagnosed in children and young adults, it was previously known as a disease of children and was called "childhood" or "juvenile" diabetes. T1DM is an autoimmune disease that results in the permanent destruction of insulin producing beta cells of the pancreas [1]. Insulin is a hormone that is needed to convert sugar (glucose), starches and other food into energy needed for daily life. Type 1 diabetes is fatal unless treated with exogenous insulin via injections to replace the missing hormone. In addition, frequent monitoring of blood glucose levels is essential.

The precise cause of diabetes and mechanisms triggering the beta-cell-specific attack remain unclear, despite the evidence that both genetics and environmental factors such as enterovirus (EV) and Coxsackie B (CVB) appear to play roles. There is currently no cure or no preventive measure that can be taken against Type 1 diabetes. Although some clinical trials aim to find interventions for preventing or slowing its development, so far none has proven successful.

According to an estimation of the American Diabetes Association [2], there are 20.8 million Americans (7 percent of the population) who have diabetes. Unfortunately, only 14.6

million (70%) have been actually diagnosed with diabetes, 6.2 million people, or nearly one-third, are unaware that they have the disease. Diagnosed diabetes patients account for 5.8 percent of the total U.S. population. About 176,500 people aged 20 years or younger have diabetes. This group represents 0.22% of all people in this age group. About one in every 400 to 600 children and adolescents has Type 1 diabetes. The peak incidence of T1DM is 11 years, but new cases frequently occur in adulthood. Type 1 diabetes onset is more common in Caucasians than in African Americans.

Diabetes is the fifth-deadliest disease in the United States and has a huge impact on public health. It is estimated that the total annual economic cost of diabetes was $174 billion in 2007. Medical expenditures attributed to diabetes totaled $116 billion, including $27 billion for direct diabetes care, $58 billion for diabetes-related chronic complications, and $31 billion in excess general medical costs. Based on death certificate data, diabetes contributed to 233,619 deaths in 2005. The toll of diabetes is actually much higher than officially reported since diabetes is generally under-reported on death certificates, particularly in the cases of older persons with multiples chronic conditions such as heart disease and hypertension. It is reported that 488 of the deaths associated with diabetes occurred among people age under 18 in 2007[2, 16].

1.2 ROLE OF TCR Vβ GENE EXPRESSIONS IN TYPE I DIABETES

It is generally believed that certain viral infections, especially those caused by enterovirus (EV) such as Coxsackie B (CVB) infections, are associated with damage of pancreas β-cell and with the clinical onset of diabetes, particularly amongst very young children [4, 5, 6]. CVB can

activate a pool of potentially autoreactive T cells present either in the peripheral circulation [3] or already accumulated in the periphery of the islets [7]. The repertoire of T cells activated by the virus seems to be restricted with expansion of certain variable portions of the β chain (Vβ) of the T cell receptor (TCR) [3]. It appears that this selective T cell activation is temporally associated with the occurrence of acute CVB infections in T1DM [3]. In particular, exposure to multiple CVB infections possibly characterizes the pre-diabetic period, thus reinforcing the hypothesis that viral infections can initiate or accelerate the process of β-cell destruction in humans [3].

The destruction of the insulin-producing beta-cells occurs over a prolonged period of time, eventually resulting in the delayed onset of the disease [1]. Detection of serum antibodies [8, 9, 10] and T-cell sensitization [11] to a vast array of islet-cell antigens during this period implies the presence of autoimmunity even at early stages in the course of the disease. The analysis of the T-cell receptor (TCR) repertoire of the circulating T cells is important because it might represent a direct means to ascertain the presence of T cells involved in the pathogenesis of Type I diabetes eventually infiltrating the pancreatic islets of Langherans [3].

A previous study reported that the majority of the islet infiltrating lymphocytes in the pancreas of two children who died at the onset of T1D expressed the TCR Vβ7 gene family [12]. A pilot study demonstrated that the preferential expression of the TCR Vβ 7 gene family can be also found in the peripheral blood of T1D patients at diagnosis [3] (Figure 1). There were no TCR biases among HLA-matched control subjects. Pre-diabetic FDRs showed expansion of the Vβ 7 gene family nearly four years prior to the onset of clinical diabetes. Other studies have also found evidence that diabetes-associated TCR bias can be detected in the peripheral circulation, and precede the onset of T1D by years.

It has been suggested that the presence of a TCR Vβ 7 expansion among the circulating T cells of T1D patients could be associated with the presence of enteroviral infection and of increased aggressiveness of the autoimmune insult. Confirmation of this postulate would identify an acceleration time point that could be targeted by intervention strategies and give impetus to strategies to prevent such infection.

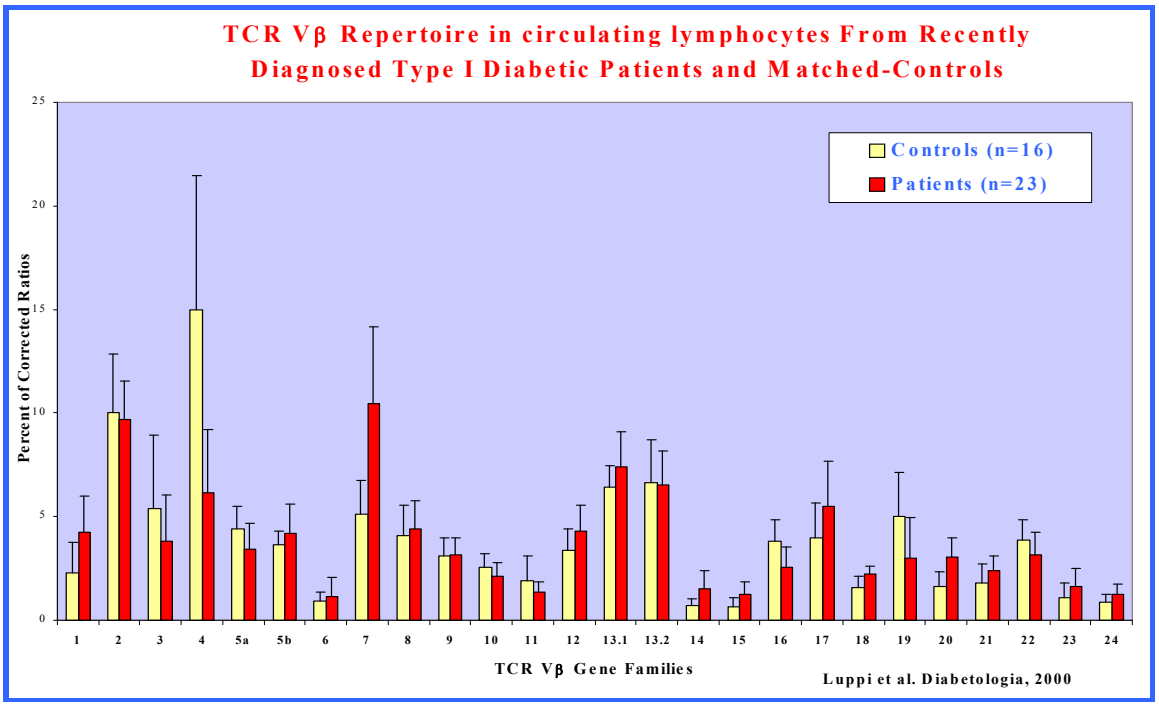

Figure 1: TCR Vβ repertoire from T1D patients and matched–control

Used with the permission from Luppi P et al. (2000) *Restricted TVR Vβ gene expression and enterovirus infection in Type I diabetes: a pilot study* Diabetologia 43: 1484-1497

2.0 STATEMENT OF THE PROBLEM

Overall Specific Aim: Investigation of TCR Vβ gene expression in Type 1 Diabetes To evaluate TCR Vβ expression, including preferential expression of TCR Vβ1, Vβ5.1, Vβ7, Vβ13.1, and Vβ20, in circulating T cells from a group of diagnosed T1DM patients and in a longitudinal prospective study of a group of high-risk first degree relatives, using the Children's Hospital of Pittsburgh T1D registry.

Hypothesis I: T-cell Vβ repertoire and viral infection Enteroviral infection or other triggering agents of T1D are associated with evidence of infection which induces a lymphocyte immune response characterized by a typical T-cell Vβ bias.

Hypothesis II: T-Cell precedes β-cell antigen-spreading in progression to T1D and is precipitated and accelerated by viral infections T-cell autoimmunity is precipitated by environmental triggers such as viral infections and precedes the appearance of autoantibodies in FDRs, which allows more sensitive and earlier detection and staging of pre-diabetes autoimmunity. T-cell and B-cell responses to increasing numbers of islet autoantigens (antigen spreading) is a marker of progressive pre-diabetes. These hypotheses will be tested by:

1) The evaluation of the ontogeny of skewing of T-cell Vβ immune responses by measuring frequencies of abnormal T-cell Vβ repertoires. Temporal relationships of these Vβ responses and serologic and genomic evidence of enteroviral infections in a prospectively followed cohort of FDR's with high-risk HLA genes will be examined.

2) The development and progression (antigen spreading) of both T-cell and B-cell autoimmune responses to islet antigens will be related to both T-cell Vβ skewing and these measures of enteroviral infection.

Aim of thesis project:

This study is still ongoing and the data has not been completely obtained. Thus, my thesis is a preliminary analysis of the available data and explores the basic relationships among the various data points and provides direction for the investigators. The results obtained will be conveyed to the researchers so that decisions and future direction of the study will be predicated on quantitative information.

3.0 MATERIALS AND METHODS

3.1 AVAILABLE POPULATIONS

This project is a part of research study on "The Etiology and Epidemiology of Type 1 Diabetes", which is conducted in the Children's Hospital of Pittsburgh. The population consists of 665 subjects enrolled from January 2004 to May 2008.

The probands, also known as new onsets, in the study were 212 children age 1 through 18 diagnosed with the onset of Type 1 diabetes. Peripheral blood mononuclear cells (PBMC) were collected at serial time points, at and after the onset of the disease. The first blood sample was to be drawn within 100 days of diagnosis in order to capture the autoantibody status at onset of the disease and was considered the baseline, as this status is believed to change over time, especially once exogenous insulin treatment has been started. Additional blood samples were collected at routine six month clinic visits.

The first degree relatives (FDRs) in the study were biologic parents or siblings of an insulin dependent diabetic child (proband) and recruited in the same time frame. Ages ranged from 2 to 54 years among subjects with no diabetes at the time of enrollment. Their peripheral blood mononuclear cells (PBMC) were collected approximately every 6 months. A total of 453 FDRs were available from the original Juvenile Onset Diabetes (JOD) and the current Antigen Spreading (AGS) studies.

3.2 DATA

Vβ data was provided by Dr. Patrizia Luppi from the Immunogenic Laboratory in the Division of Immunogenetics, Department of Pediatrics, University of Pittsburgh. Vβ expression was available for Vβ 1, Vβ 5.1, Vβ 7, Vβ 13.1, and Vβ 20. Demographic, antibody and genetic data was available from the master database of the etiology and epidemiology of Type 1 Diabetes study.

3.3 ANALYTIC STRATEGY AND STATISTICAL METHODS

Descriptive statistics were provided for the entire study population and for the subgroups of diabetic patients and their unaffected first degree relatives (FDRs). For categorical variables, frequency distributions were calculated and chi-square tests and linear trend tests were conducted to evaluate associations. For continuous variables, we estimated such statistics as mean, standard deviation, 95% confidence limits. Pearson correlations were computed to assess the associations among some continuous variables. To assess differences in continuous variables, we performed two-sample t-tests for comparison between two categories, the general linear model (GLM) for unbalanced analysis, and Duncan's multiple range tests for multiple groups. Plots such as histogram plots, box plots, and scatter plots were created to explore patterns in the data. Multiple logistic regression models were also built to investigate the association between the onset of diabetes and potential independent predictors.

Results from multiple blood draws were available for some individuals. For consistency we chose only one blood draw to use in our analyses. Among new onsets it was their baseline

blood draw. For FDR subjects, it was their first draw. Therefore, we verified that the blood draw date was within 100 days of diagnosis for new onsets. If this was not the case, the subject was excluded. For FDRs, they had to be unaffected for the purpose of these analyses. Therefore, the 7 FDRs with diabetes were excluded. Furthermore, if an individual had multiple blood draws, only the earliest was used.

Statistical Analysis System (SAS) version 9.0 was used for all descriptive statistics, test statistics, and graphic displays.

4.0 ANALYSIS AND RESULTS

4.1 DEMOGRAPHICS

The entire population included 665 participants, 212 probands or new onsets, 453 first degree relatives (FDRs). Table 1 and Table 2 show the distribution of gender and race for the two subpopulations: new onset and FDR. The total population was made up of 40% males and 60% females. Of the 212 new onsets, 62% were males and 38% were females. Of the 450 FDRs, 30% were males and 70% were females (Table 1). Ninety-one percent out of the new onsets and 97% out of the FDRs were white (Table 2).

According to the protocol, the study population should be the diabetic individuals whose first blood samples were collected within 100 days of diagnosis and the unaffected first degree relatives. We excluded 128 diabetic patients whose first blood samples were drawn beyond 100 days of diagnosis and 7 converters in FDRs from the entire population. Thus 84 patients and 446 FDRs were retained for the analysis. Table 3 and Table 4 show the distribution of gender and race, respectively. Table 5 shows race (only white and black) distribution by gender for our study population. For the 84 new onsets, 64% were males and 36% were females. 88% were white and 10% were black. Amongst the 74 white new onsets, 62% were males and 38% were females. Out of the 8 black new onsets, 87.5% were males and 12.5% were females. Of the 443 FDRs (excluding 3 missing), 30% were males and 70% were females. 97% were white and 2%

were black. In the 428 white FDRs, 30% were males and 70% were females. As for the 9 black FDRs, 22% were males and 78% were females.

Table 1: Gender distribution for the whole population

Gender	New Onsets		FDRs	
	Count	Percent (%)	Count	Percent (%)
Male	132	62.3	134	29.8
Female	80	37.7	316	70.2
Total	212	100.0	450	100.0

Note: Frequency missing = 3

Table 2: Race distribution for the whole population

Gender	New Onsets		FDRs	
	Count	Percent (%)	Count	Percent (%)
White	193	91.0	435	97.3
Black	15	7.1	9	2.0
Asian	0	0.0	1	0.2
Other*	4	1.9	2	0.5
Total	212	100.0	447	100.0

Note: Frequency missing = 6
* including multiple races

Table 3: Gender distribution for the study population

Gender	New Onsets		FDRs	
	Count	Percent (%)	Count	Percent (%)
Male	54	64.3	132	29.8
Female	30	35.7	311	70.2
Total	84	100.0	443	100.0

Note: Frequency missing = 3

Table 4: Race distribution for the study population

Gender	New Onsets		FDRs	
	Count	Percent (%)	Count	Percent (%)
White	74	88.1	428	97.3
Black	8	9.5	9	2.0
Asian	0	0.0	1	0.2
Other*	2	2.4	2	0.5
Total	84	100.0	440	100.0

Note: * including multiple races
Frequency missing = 6

Table 5: Race distribution by gender for the study population

Race*	Gender	New Onsets		FDRs	
		Count	Percent (%)	Count	Percent (%)
White	Male	46	62.2	128	29.9
	Female	28	37.8	300	70.1
	Total	74	100.0	428	100.0
Black	Male	7	87.5	2	22.2
	Female	1	12.5	7	77.8
	Total	8	100.0	9	100.0

Note: * Only white and black

4.2 VB EXPRESSIONS FOR NEW ONSETS AND FDRS

4.2.1 Correlations between Vβ expressions

The first blood samples were used in this analysis. Table 6 shows the correlations between all the Vβ expressions. From the p values we observed that for new onsets, Vβ1 was correlated with Vβ5.1, Vβ13.1, and Vβ20; and Vβ13.1 was correlated with Vβ20. For FDRs, Vβ7 was correlated with Vβ5.1, and Vβ13.1; and Vβ13.1 was correlated with Vβ20.

Table 6: Correlations between Vβ expressions

Coefficient p-value	Variable	Vβ 1	Vβ 5.1	Vβ 7	Vβ 13.1	Vβ 20
New Onset	**Vβ 1**	1.0000	0.2343 **0.0341**	0.0275 0.8063	0.3227 **0.0031**	0.2412 **0.0290**
	Vβ 5.1		1.0000	0.0333 0.7664	0.0779 0.4866	0.1934 0.0818
	Vβ 7			1.0000	-.0813 0.4678	0.0618 0.5810
	Vβ 13.1				1.0000	0.5395 **<.0001**
	Vβ 20					1.0000
FDR	**Vβ 1**	1.0000	0.0509 0.2933	-.0813 0.0931	-.0587 0.2266	0.0829 0.0876
	Vβ 5.1		1.0000	0.1929 **<.0001**	0.0427 0.3768	-.0679 0.1597
	Vβ 7			1.0000	-.0954 **0.0482**	-.0400 0.4086
	Vβ 13.1				1.0000	0.1280 **0.0079**
	Vβ 20					1.0000

4.2.2 Vβ expressions by days since diagnosis for new onsets

The new onsets in this study were the individuals with their first blood samples drawn within 100 days of diagnosis. We assessed if Vβ expressions were correlated with the number of days from the date of diagnosis to the time of the first blood draw. If some correlations existed, the whole time period (100 days) was to be divided into two time periods: within 7 days and 8-100 days. We would investigate the significance of Vβ expressions in these two different time periods.

Table 7 shows the correlations between Vβ expressions and the number of days between the date of diagnosis and the date of the blood draw. It appears that Vβ1 and Vβ13.1 were significantly correlated with the days, whereas Vβ20 was a borderline and also show somewhat correlation with the days. From Figure 2 decreased tendencies were observed in the Vβ1 expression and the Vβ13.1 expression, indicating the expression of Vβ1 and Vβ13.1 gene families tended to decrease as the time elapsed since diagnosis. A similar pattern was also noted in the Vβ20 gene family, even though the trend was not significant.

We further detected the changes in Vβ expression in the two time periods. Referencing to Table 8 and Figure 3, it appears that all the Vβ expressions except the Vβ7 expression were higher within 7 days of diagnosis relative to the values between 8 to 100 days of diagnosis. Significant changes were observed in both Vβ1 and Vβ13.1 gene families. For the Vβ1 and Vβ13.1 gene families, significantly higher values ($p = 0.0004$, $p < 0.0001$) were measured within 7 days of diagnosis in contrast to the values measured between 8 to 100 days of diagnosis. As for the Vβ20 expression, a borderline, a slightly higher value was noted within 7 days of diagnosis compared to the value measured between 8 to 100 days of diagnosis ($p = 0.0511$). The results demonstrated the time effect on Vβ expression.

Table 7: Correlations between Vβ expressions and days since diagnosis

Coefficient **p-value**	**Vβ 1**	**Vβ 5.1**	**Vβ 7**	**Vβ 13.1**	**Vβ 20**
No. of Days	-0.3751 **0.0005**	-0.1264 0.2579	0.0890 0.4268	-0.4011 **0.0002**	-0.1956 0.0782

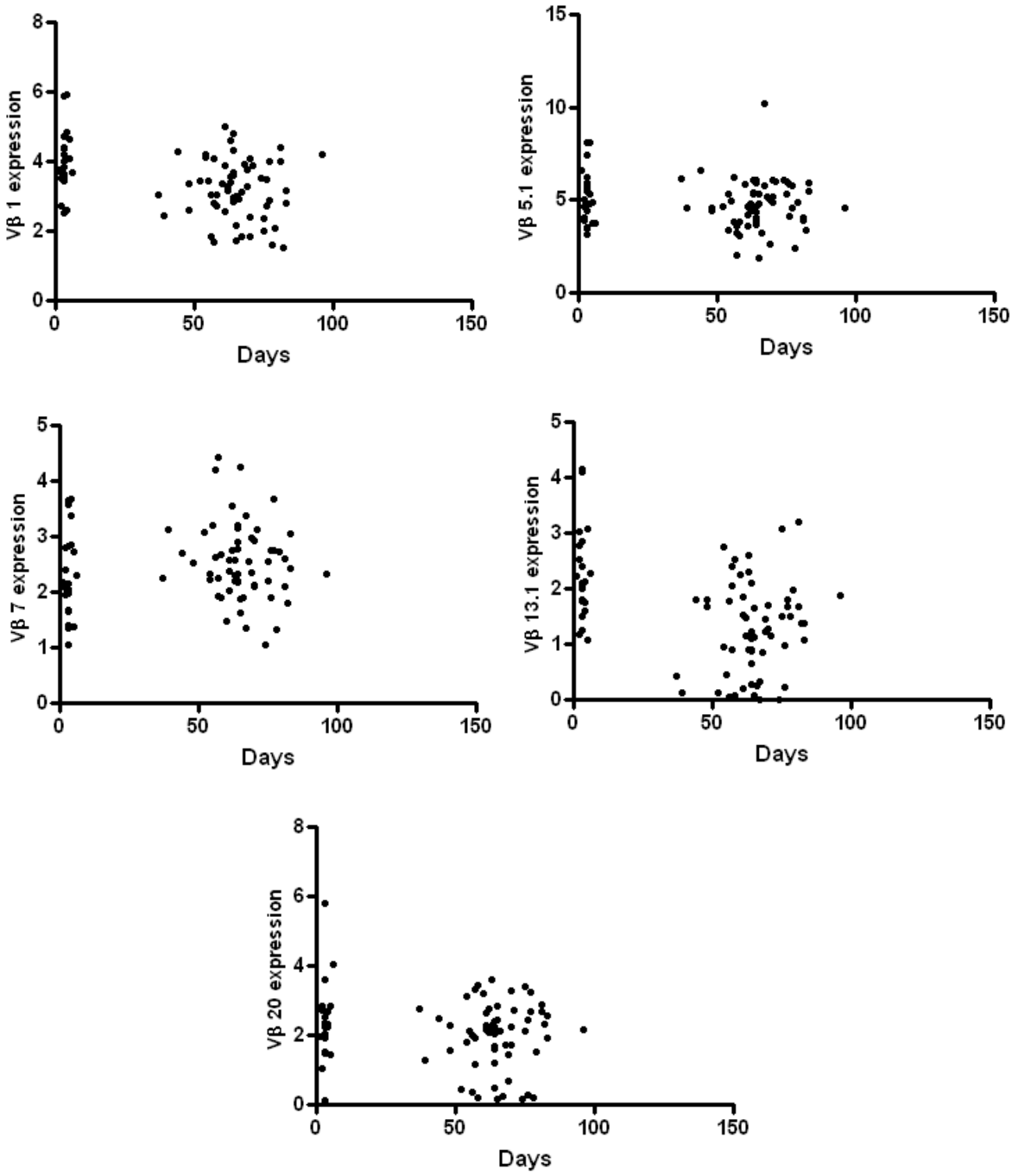

Figure 2: Scatter plots for Vβ expressions vs days since diagnosis

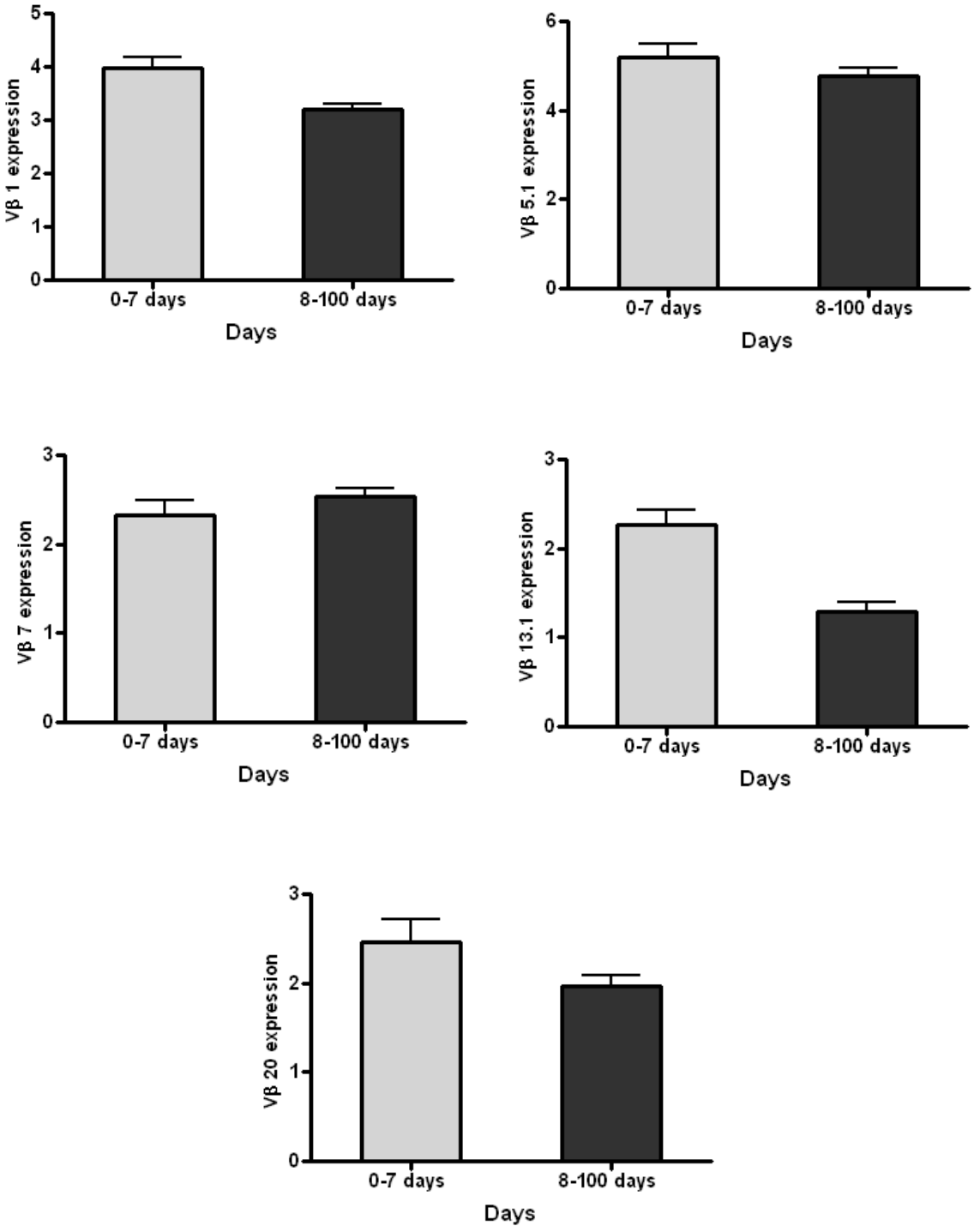

Figure 3: Vβ expressions by days since diagnosis

Table 8: Vβ expressions by days since diagnosis

Variable	No. of Days	N	Means	Std. Dev	Min	Max	p-value
Vβ 1	Within 7 days	22	3.99	0.88	2.51	5.93	**0.0004**
	8 - 100 days	60	3.19	0.86	1.52	5.01	
Vβ 5.1	Within 7 days	22	5.20	1.45	3.14	8.12	0.2160
	8 - 100 days	60	4.78	1.32	1.88	10.2	
Vβ 7	Within 7 days	22	2.32	0.78	1.06	3.67	0.2198
	8 - 100 days	60	2.54	0.68	1.06	4.42	
Vβ 13.1	Within 7 days	22	2.26	0.83	1.07	4.16	**<.0001**
	8 - 100 days	60	1.29	0.81	0.00	3.20	
Vβ 20	Within 7 days	22	2.46	1.16	0.12	5.82	0.0511
	8 - 100 days	60	1.96	0.97	0.16	3.60	

4.2.3 Vβ expressions by subpopulation

In this portion, we first assessed whether any differences existed in the expression of Vβ gene families between two subpopulations: FDRs and new onsets with their first blood samples drawn within 7 days of diagnosis. Table 9 shows that with the exception of Vβ1 expression every Vβ expression in new onsets had no significant differences from the value in FDRs. Based on the two-sample t-tests for Vβ expressions, the results indicate that for the Vβ1 gene family, a significantly lower value ($p = 0.0017$) was observed in FDRs (mean = 3.24%±1.09SD) compared to the new onsets whose first blood samples were collected within 7 days of diagnosis (mean = 3.99%±0.88SD).

Then we assessed the differences in the expression of Vβ gene families between two subpopulations: FDRs and new onsets with their first blood samples drawn within 100 days of diagnosis. Table 10 shows that with the exception of the Vβ13.1 expression every Vβ expression

in new onsets had no significant differences from the value in FDRs. For the Vβ13.1 gene family, a lower value was observed in new onsets (mean = 1.55%±0.92SD) relative to the FDRs (mean=2.05%±0.85SD), which was significantly different ($p < 0.0001$).

We found that the two comparisons had totally different results. The results of the second comparison show that the Vβ13.1 expression changed in a counter intuitive direction, that is, the Vβ13.1 expression was lower in new onsets than in FDRs. The most possible reason is the time effect on Vβ expression. Based on the analysis in 4.2.2 we knew that all the Vβ expressions except the Vβ7 expression in new onsets had negative correlations with the days. As time elapsed since diagnosis, the Vβ expressions decreased so that the Vβ1 expression in new onsets was not significantly higher than in FDRs, and the Vβ13.1 expression in new onsets decreased dramatically so that it was significantly lower than in FDRs. Therefore, it is recommended for the study that the baseline be restricted to the blood samples within 7 days of diagnosis.

It was shown that some subjects had multiple blood draws. For the new onsets whose Vβ expressions were available within 7 days of diagnosis, seven with 2 blood draws were identified. Table 11 indicates that with the exception of the Vβ7 expression, all the other Vβ expressions declined over time compared to the baseline. However, further analysis can not be conducted due to the small sample size ($n = 7$).

Table 9: Vβ expressions for new onsets and FDRs

Variable	Subjects	N*	Mean	Std Dev	Min	Max	p-value
Vβ 1	New Onset	22	3.99	0.88	2.51	5.93	**0.0017**
	FDR	428	3.24	1.09	0.84	10.3	
Vβ 5.1	New Onset	22	5.20	1.45	3.14	8.12	0.5656
	FDR	432	5.04	1.22	1.94	10.9	
Vβ 7	New Onset	22	2.32	0.78	1.06	3.67	0.4926
	FDR	432	2.49	1.11	0.76	12.5	
Vβ 13.1	New Onset	22	2.26	0.83	1.07	4.16	0.2671
	FDR	430	2.05	0.85	0.18	9.07	
Vβ 20	New Onset	22	2.46	1.16	0.12	5.82	0.1959
	FDR	430	2.17	1.05	0.02	7.65	

Note: * Results not available for every Vβ expression

Table 10: Vβ expressions for new onsets and FDRs

Variable	Subjects	N*	Mean	Std Dev	Min	Max	p-value
Vβ 1	New Onset	82	3.41	0.93	1.52	5.93	0.2021
	FDR	428	3.24	1.09	0.84	10.3	
Vβ 5.1	New Onset	82	4.89	1.36	1.88	10.2	0.3032
	FDR	432	5.04	1.22	1.94	10.9	
Vβ 7	New Onset	82	2.48	0.71	1.06	4.42	0.9531
	FDR	432	2.49	1.11	0.76	12.5	
Vβ 13.1	New Onset	82	1.55	0.92	0.00	4.16	**<.0001**
	FDR	430	2.05	0.85	0.18	9.07	
Vβ 20	New Onset	82	2.10	1.04	0.12	5.82	0.5820
	FDR	430	2.17	1.05	0.02	7.65	

Note: * Results not available for every Vβ expression

Table 11: Differences in Vβ expressions between two blood draws

Variable	N	Mean	Std Dev	Min	Max
Vβ 1	7	-0.9771	1.6029	-3.50	0.69
Vβ 5.1	7	-0.5771	1.5656	-3.07	1.70
Vβ 7	7	0.7729	1.3034	-0.80	2.22
Vβ 13.1	7	-0.2186	0.8921	-1.43	1.27
Vβ 20	7	-0.6729	0.6538	-1.76	-0.05

4.2.4 Vβ expressions by gender

The expression of Vβ gene families by gender are shown in Table 12. Based on the results of the two-sample t-tests, there were no significant differences in Vβ expressions between males and females in new onsets. With the exception of the Vβ1 expression, for FDRs, all the other Vβ expressions in males did not statistically differ from those in females. However, the Vβ1 expression in males (mean = 3.45%±1.22SD) was significantly higher (p = 0.0095) than the value in females (mean = 3.15%±1.02SD) (Figure 4).

Table 12: Vβ expressions by gender

Subject	Variable	Gender	N*	Mean	Std Dev	Min	Max	p-value
New Onset	**Vβ 1**	Male	54	3.52	0.90	1.86	5.93	0.1241
		Female	28	3.19	0.95	1.52	4.80	
	Vβ 5.1	Male	54	5.06	1.41	1.88	10.2	0.1286
		Female	28	4.57	1.22	2.03	6.26	
	Vβ 7	Male	54	2.57	0.68	1.36	4.25	0.1353
		Female	28	2.32	0.75	1.06	4.42	
	Vβ 13.1	Male	54	1.59	0.93	0.06	4.16	0.6055
		Female	28	1.48	0.91	0.00	3.20	
	Vβ 20	Male	54	2.15	1.00	0.16	5.82	0.5117
		Female	28	1.99	1.11	0.12	3.61	

Table 12 continued

Subject	Variable	Gender	N*	Mean	Std Dev	Min	Max	p-value
FDR	**Vβ 1**	Male	129	3.45	1.22	1.02	10.3	**0.0095**
		Female	299	3.15	1.02	0.84	6.30	
	Vβ 5.1	Male	129	5.07	1.31	1.98	10.9	0.7413
		Female	303	5.03	1.18	1.94	8.56	
	Vβ 7	Male	129	2.57	1.47	0.88	12.5	0.3174
		Female	303	2.45	0.91	0.76	10.4	
	Vβ 13.1	Male	129	2.03	0.75	0.18	4.84	0.7139
		Female	301	2.06	0.90	0.24	9.07	
	Vβ 20	Male	129	2.08	0.98	0.08	7.65	0.2884
		Female	301	2.20	1.08	0.02	7.22	

Note: * Results not available for every Vβ expression

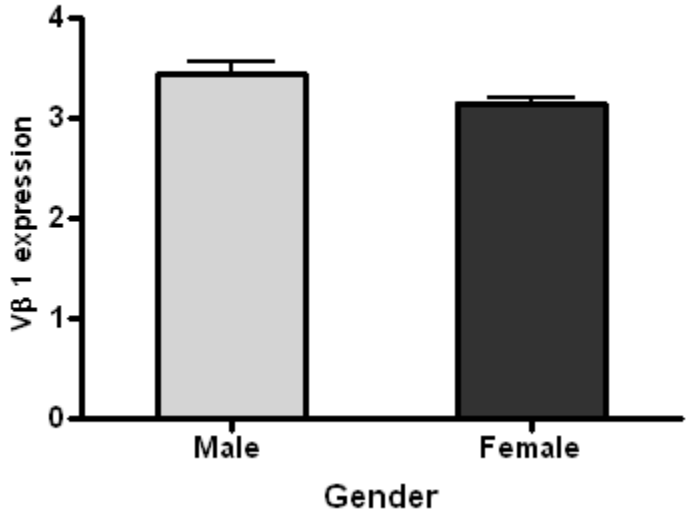

Figure 4: Vβ 1 expression by gender for FDRs

4.2.5 Vβ expressions by race

Table 4 indicates that most subjects in our study population were white, followed by black, with few were Asians and other races. Therefore we only considered white and black. Table 13 shows Vβ expressions by race and the p values which were based on the two-sample t-tests for the comparison of the Vβ expressions between white and black in new onsets and FDRs. It appears that the Vβ7 expression for the black new onsets was 3.14%±0.62SD, which was significantly higher than the value for the white new onsets (mean = 2.42%±0.69SD) with p value of 0.0057. For the Vβ1 gene family, the black FDRs (mean = 4.07%±1.44SD) showed significantly higher value (p= 0.0216) than the white FDRs (mean = 3.22%±1.08SD). For the Vβ13.1 expression, the black FDRs (mean = 1.27%±0.77SD) was significantly lower (p = 0.0050) than the white FDRs (mean = 2.08%±0.85SD).

Table 13: Vβ expressions by race

Subject	Variable	Race	N*	Mean	Std Dev	Min	Max	p value
New Onset	**Vβ 1**	White	72	3.36	0.92	1.52	5.87	0.1939
		Black	8	3.81	1.05	2.61	5.93	
	Vβ 5.1	White	72	4.78	1.31	1.88	10.2	0.1197
		Black	8	5.56	1.64	3.85	8.12	
	Vβ 7	White	72	2.42	0.69	1.06	4.42	**0.0057**
		Black	8	3.14	0.62	2.54	4.20	
	Vβ 13.1	White	72	1.51	0.88	0.00	4.16	0.5560
		Black	8	1.72	1.29	0.06	4.11	
	Vβ 20	White	72	2.02	0.96	0.12	4.03	0.1596
		Black	8	2.57	1.57	0.36	5.82	
FDR	**Vβ 1**	White	413	3.22	1.08	0.85	10.3	**0.0216**
		Black	9	4.07	1.44	2.05	7.31	
	Vβ 5.1	White	417	5.02	1.22	1.94	10.9	0.1080
		Black	9	5.68	0.80	4.85	7.45	
	Vβ 7	White	417	2.49	1.12	0.76	12.5	0.5144
		Black	9	2.73	0.88	1.39	4.48	
	Vβ 13.1	White	415	2.08	0.85	0.18	9.07	**0.0050**
		Black	9	1.27	0.77	0.59	3.05	
	Vβ 20	White	415	2.16	1.04	0.02	7.65	0.0742
		Black	9	2.79	1.18	0.63	4.25	

Note: * Results not available for every Vβ expression

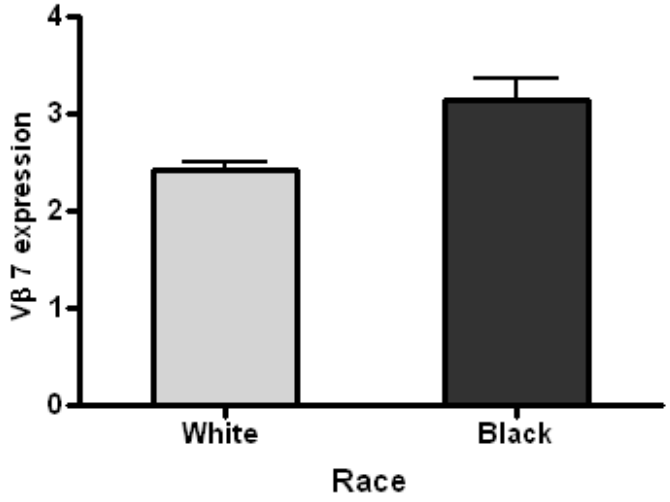

Figure 5: Vβ 7 expression by race for new onsets

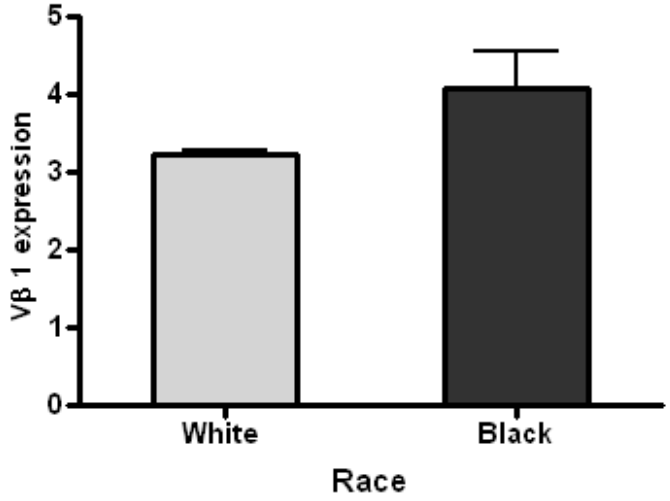

Figure 6: Vβ 1 expression by race for FDRs

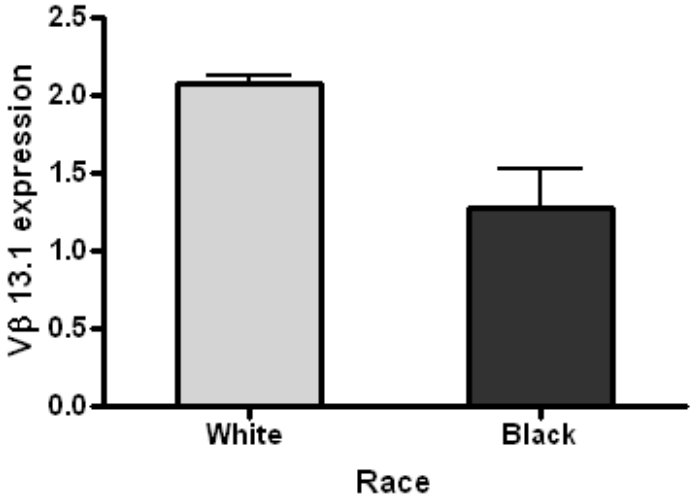

Figure 7: Vβ 13.1 expression by race for FDRs

4.2.6 Vβ expressions by positive antoantibodies

The presence of autoantibodies in the pancreatic islet cells is often discovered in diabetic patients. We detected four such antibodies: glutamic acid decarboxylase (GAD), insulinoma-associated protein 2 (IA-2), insulin autoantibodies (IAA), and islet cell antibodies (ICA). The previous studies have demonstrated that up to 90% of newly diagnosed diabetics have at least one of these antibodies in contrast to just 1% of people in the general population [15]. In our study we tested each new onset for the antibodies. Then we investigated if the expression of TCR Vβ gene families was associated with the antibodies. The Vβ expressions by positive antibodies GAD, IA-2, IAA, and ICA in new onsets are shown in Table 14 to Table 17, respectively. Table 18 shows Vβ expression by the number of positive antibodies. The comparison of each Vβ expression by positive antibodies was conducted by using the GLM

analysis and all the p values were greater than 0.05. It appears that for new onsets the expression of TCR Vβ gene families was not associated with positive autoantibodies GAD, IA-2, IAA, ICA, and the number of antibodies.

Table 14: Vβ expressions by antibody GAD for new onsets

Variable	N/P	N	Mean	Std Dev	Min	Max	p value
Vβ 1	N	34	3.23	0.93	1.52	5.01	0.0517
	P	42	3.65	0.89	1.73	5.93	
Vβ 5.1	N	34	4.70	1.25	2.03	6.60	0.4614
	P	42	4.92	1.29	1.88	8.12	
Vβ 7	N	34	2.50	0.63	1.33	4.42	0.9528
	P	42	2.49	0.78	1.06	4.25	
Vβ 13.1	N	34	1.66	0.89	0.08	4.16	0.5422
	P	42	1.53	0.95	0.00	4.11	
Vβ 20	N	34	2.29	0.98	0.20	4.03	0.2078
	P	42	1.98	1.10	0.12	5.82	

Note: N = Negative P = Positive

Table 15: Vβ expressions by antibody IA-2 for new onsets

Variable	N/P	N	Mean	Std Dev	Min	Max	p value
Vβ 1	N	30	3.35	0.78	1.85	5.01	0.5891
	P	47	3.47	1.03	1.52	5.93	
Vβ 5.1	N	30	5.00	1.03	3.09	6.60	0.6675
	P	47	4.85	1.59	1.88	10.2	
Vβ 7	N	30	2.43	0.51	1.06	3.37	0.5179
	P	47	2.54	0.83	1.06	4.42	
Vβ 13.1	N	30	1.39	0.88	0.00	3.08	0.2771
	P	47	1.63	0.98	0.00	4.16	
Vβ 20	N	30	2.00	0.89	0.16	3.60	0.7176
	P	47	2.09	1.12	0.12	5.82	

Note: N = Negative P = Positive

Table 16: Vβ expressions by antibody IAA for new onsets

Variable	N/P	N	Mean	Std Dev	Min	Max	p value
Vβ 1	N	15	3.83	0.81	2.51	5.93	0.8989
	P	13	3.88	1.08	1.73	5.87	
Vβ 5.1	N	15	4.95	1.38	3.14	8.12	0.2768
	P	13	5.51	1.29	3.48	8.08	
Vβ 7	N	15	2.25	0.74	1.06	3.67	0.3491
	P	13	2.53	0.78	1.40	3.68	
Vβ 13.1	N	15	2.09	0.87	0.90	4.16	0.7927
	P	13	2.00	0.93	0.42	4.11	
Vβ 20	N	15	2.37	0.91	0.12	3.61	0.7832
	P	13	2.48	1.16	1.04	5.82	

Note: N = Negative P = Positive

Table 17: Vβ expressions by antibody ICA for new onsets

Variable	N/P	N	Mean	Std Dev	Min	Max	p value
Vβ 1	N	27	3.34	0.85	1.85	5.93	0.4514
	P	49	3.51	0.98	1.52	5.97	
Vβ 5.1	N	27	5.00	1.23	3.09	8.12	0.5759
	P	49	4.81	1.50	1.88	10.2	
Vβ 7	N	27	2.34	0.57	1.06	3.67	0.1354
	P	49	2.59	0.77	1.33	4.42	
Vβ 13.1	N	27	1.46	0.96	0.00	3.08	0.6683
	P	49	1.55	0.93	0.00	4.16	
Vβ 20	N	27	2.13	0.99	0.16	3.61	0.9962
	P	49	2.13	1.09	0.16	5.82	

Note: N = Negative P = Positive

Table 18: Vβ expressions by number of positive antibodies

Variable	Number of positive antibodies	N	Mean	Std Dev	Min	Max	p value
Vβ 1	**0**	15	2.93	0.70	1.85	4.30	0.0816
	1	14	3.65	0.52	2.80	5.01	
	2	22	3.42	1.13	1.52	5.93	
	3	23	3.48	0.78	2.16	4.83	
	4	6	4.07	1.38	1.73	5.87	
Vβ 5.1	**0**	15	4.91	1.21	3.09	6.60	0.5824
	1	14	4.94	0.90	3.26	6.11	
	2	22	4.92	1.81	2.03	10.2	
	3	23	4.63	1.11	1.88	6.14	
	4	6	5.70	1.85	3.48	8.08	
Vβ 7	**0**	15	2.40	0.39	1.90	3.15	0.3131
	1	14	2.30	0.62	1.06	3.21	
	2	22	2.40	0.85	1.06	4.42	
	3	23	2.75	0.69	1.39	4.25	
	4	6	2.35	0.99	1.40	3.65	
Vβ 13.1	**0**	15	1.34	0.95	0.08	3.08	0.0862
	1	14	1.42	0.88	0.00	2.86	
	2	22	1.85	0.81	0.33	4.16	
	3	23	1.29	0.93	0.00	3.20	
	4	6	2.20	1.05	1.14	4.11	
Vβ 20	**0**	15	2.18	1.05	0.20	3.61	0.7748
	1	14	2.10	1.05	0.16	4.03	
	2	22	2.05	0.92	0.12	3.59	
	3	23	1.97	1.04	0.16	3.30	
	4	6	2.60	1.65	1.04	5.82	

4.2.7 Vβ expressions and BMI / BMIZ

Body mass index (BMI) and z-score for BMI-for-age (BMIZ) are important factors contributing to Type 1 diabetes. BMI is calculated by the formula: BMI = weight (kg) / [height (m)]^2 . Table 19 shows the correlations between Vβ expressions and BMI at baseline or BMI at 3 months, and BMIZ at baseline or BMIZ at 3 months for new onsets. From the correlation coefficients and the corresponding p values we concluded that the expression of TCR Vβ gene families had no correlations with BMI at baseline, BMI at 3 months, BMIZ at baseline, and BMIZ at 3 months.

Table 19: Correlations between Vβ expressions and BMI / BMIZ

Coefficient p-value	N	Vβ 1	Vβ 5.1	Vβ 7	Vβ 13.1	Vβ 20
BMI	80	-.0684 0.5466	-.0129 0.9098	-.0661 0.5633	-.0501 0.6593	-.0841 0.4582
BMI:3 months	75	0.0151 0.8977	0.0254 0.8287	-.0800 0.4952	0.0155 0.8951	-.0814 0.4874
BMIZ	80	-0.1196 0.2906	-0.0334 0.7688	-0.1148 0.3105	-0.0147 0.8968	0.0284 0.8028
BMIZ:3 months	75	0.1510 0.1959	-0.0184 0.8753	-0.0115 0.9217	0.1238 0.2898	0.0757 0.5188

4.2.8 New onsets vs. FDRs with ASP haplotype

The FDRs carrying ASP or 0 non-ASP haplotype are individuals with the least risk of diabetes among FDRs. In the 446 FDRs, 40 were the individuals with ASP haplotype. We compared them with the 23 new onsets with their first blood samples collected within 7 days of the diagnosis to assess the differences in Vβ expressions between these two groups.

Two-sample t-test was performed for each Vβ gene family. We found that the Vβ1 expression in the FDRs carrying ASP haplotype (mean = 3.47%±0.92SD) was significant lower (p = 0.0348) than that in new onsets (mean = 3.99%±0.88SD), which is similar to the result in the comparison between the new onsets and all the FDRs. As for other Vβ gene families, there were no significant differences between new onsets and the FDRs with ASP alleles (Table 20).

Table 20: Two-sample t-test for Vβ expressions (new onsets vs ASP FDRs)

Variables	Mean New Onset	FDR	Mehtods	DF	t-value	p-value
Vβ 1	3.99	3.47	Pooled	59	-2.16	**0.0348**
Vβ 5.1	5.20	4.90	Pooled	59	-0.90	0.3703
Vβ 7	2.32	2.50	Pooled	59	0.83	0.4116
Vβ 13.1	2.26	1.96	Pooled	59	-1.61	0.1117
Vβ 20	2.46	2.23	Pooled	59	-0.86	0.3942

4.3 VB EXPRESSIONS AND HLA GENES

The fact that the first degree relatives of T1D probands appear to have a higher risk of developing diabetes indicates that there might be a genetic susceptibility to diabetes. The Human Leukocyte Antigen (HLA) genes are identified as diabetes susceptibility genes and play the largest role in a genetically heterogenous disease, comprising approximately 50% of the genetic risk [13, 14]. We assessed whether an expression of certain TCR Vβ gene families associates with HLA genes.

4.3.1 Vβ expressions by non-ASP haplotype

Multiple studies demonstrated a strong correlation between having a non-charged amino acid such as alanine, valine, or serine at codon 57 (non-ASP) in diabetic patients instead of having aspartic acid (ASP) as was most often seen in non-diabetic patients. Therefore, the HLA results were stratified by non-ASP haplotype. For the non-ASP hyplotype, each of the participants was evaluated for the presence of zero (ASP/ASP), one (Non-ASP/ASP), and two non-ASP alleles (Non-ASP/ Non-ASP). For those with one non-ASP plus DQB1*0602 was recorded separately as Non-ASP/0602, which shows a protective effect against the development of T1D even in the presence of autoantibodies. The order of risk from lowest to highest is ASP/ASP, non-ASP/0602, non-ASP/ASP, and non-ASP/non-ASP. 32 of the FDRs were unable to determine their status and need a further analysis of the alleles.

As seen in Table 21 below, out of the 77 patients, only 1% of the patients were in the ASP/ASP group, 99% had at least one non-ASP HLA haplotype, including 3% with non-ASP/0602 which is a protective haplotype, 25% with one non-ASP allele and 71% with 2 non-ASP alleles. The higher incidence in the non-ASP group is to be expected as it is the presence of non-ASP alleles that is associated with the risk of developing T1D. Looking at the presence of non-ASP in the 407 FDRs, 90% of them had at least one non-ASP HLA haplotype with a break down of 45% with one non-ASP allele, 34% with 2 non-ASP alleles, and 11% with the non-ASP/0602 allele. 10% of FDRs had 0 non-ASP HLA haplotype. A Chi-Square test and a Cochran-Armitage trend test were performed. It appears that differences existed among non-ASP haplotype categories ($p < 0.0001$), which showed a linear tendency ($p = 0.0027$).

Figure 8 shows Vβ expressions by non-ASP haplotype for the study population. According to the data in Table 22, lower Vβ7 expressions were measured in T1D patients

carrying the ASP/ASP and non-ASP/0602 haplotype with means 1.91% and 1.88% in contrast to values of 2.58% and 2.50% in those with the non-ASP/ASP and non-ASP/non-ASP alleles. The lowest Vβ7 expression was observed in the FDRs carrying the non-ASP/0602 haplotype (mean = 2.05%±0.76SD) in contrast to values ranging from 2.50% to 2.57% in other haplotype groups. The results of the Duncan's multiple range test indicated FDRs carrying the non-ASP/0602 haplotype had significant lower Vβ7 expression than other FDRs. This evidence supports the fact that DQB1*0602 is a protective haplotype. For the TCR Vβ13.1 gene family, the lowest expression was detected in the ASP/ASP group of new onset with mean 0.25% compared to values ranging from 1.50% to 1.93% in the other haplotypes. The highest Vβ13.1 expression for FDRs was found in the Non-ASP/Non-ASP group (mean = 2.18%±1.07SD) relative to values ranging from 1.96% to 2.01% in other haplotype groups. Using the GLM analysis, however, we concluded that both new onsets and FDRs showed no significant differences in each Vβ expression among the individuals with different non-ASP haplotypes.

Table 21: Non-ASP haplotype distribution for new onsets and FDRs

Frequency Percent RowPct Col Pct	ASP/ASP	Non-ASP /0602	Non-ASP /ASP	Non-ASP /Non-ASP	Total
New Onset	1	2	19	55	77
	0.21	0.41	3.93	11.36	15.91
	1.30	2.60	24.68	71.43	
	2.44	4.35	9.41	28.21	
FDR	40	44	183	140	407
	8.26	9.09	37.81	28.93	84.09
	9.83	10.81	44.96	34.40	
	97.56	95.65	90.59	71.79	
Total	41	46	202	195	484
	8.47	9.50	41.74	40.29	100.00

Table 22: Vβ expressions by non-ASP haplotype

Subject	Variable	ASP Haplotype	N*	Mean	Std Dev	Min	Max
New Onset	Vβ 1	ASP/ASP	1	2.92	.	2.92	2.92
		Non-ASP/0602	2	3.97	0.47	3.64	4.30
		Non-ASP/ASP	19	3.20	0.89	1.59	4.80
		Non-ASP/Non-ASP	53	3.44	0.98	1.52	5.93
	Vβ 5.1	ASP/ASP	1	3.27	.	3.27	3.27
		Non-ASP/0602	2	6.24	0.50	5.89	6.60
		Non-ASP/ASP	19	4.10	1.06	2.03	6.26
		Non-ASP/Non-ASP	53	5.12	1.37	1.88	10.2
	Vβ 7	ASP/ASP	1	1.91	.	1.91	1.91
		Non-ASP/0602	2	1.88	1.17	1.06	2.71
		Non-ASP/ASP	19	2.58	0.83	1.33	4.42
		Non-ASP/Non-ASP	53	2.50	0.68	1.06	4.25
	Vβ 13.1	ASP/ASP	1	0.25	.	0.25	0.25
		Non-ASP/0602	2	1.93	0.19	1.80	2.07
		Non-ASP/ASP	19	1.64	0.78	0.00	3.03
		Non-ASP/Non-ASP	53	1.50	0.94	0.00	4.11

Table 22 continued

Subject	Variable	ASP Haplotype	N*	Mean	Std Dev	Min	Max
New Onset	Vβ 20	ASP/ASP	1	2.12	.	2.12	2.12
		Non-ASP/0602	2	1.31	1.68	0.12	2.49
		Non-ASP/ASP	19	2.00	1.10	0.22	3.61
		Non-ASP/Non-ASP	53	2.10	1.03	0.16	5.82
FDR	Vβ 1	ASP/ASP	39	3.47	0.92	1.70	5.17
		Non-ASP/0602	43	3.35	1.23	0.90	6.13
		Non-ASP/ASP	179	3.27	1.14	1.02	10.3
		Non-ASP/Non-ASP	135	3.14	1.04	0.84	6.95
	Vβ 5.1	ASP/ASP	39	4.90	1.11	2.70	8.04
		Non-ASP/0602	43	5.20	1.29	2.54	8.14
		Non-ASP/ASP	180	4.94	1.15	1.94	8.61
		Non-ASP/Non-ASP	137	5.08	1.27	1.98	10.9
	Vβ 7	ASP/ASP	39	2.50	0.81	1.16	5.45
		Non-ASP/0602	43	2.05	0.76	1.05	6.23
		Non-ASP/ASP	180	2.57	1.05	0.76	11.3
		Non-ASP/Non-ASP	137	2.56	1.39	0.88	12.5
	Vβ 13.1	ASP/ASP	39	1.96	0.61	0.85	3.31
		Non-ASP/0602	43	2.01	0.64	0.52	3.73
		Non-ASP/ASP	178	1.97	0.76	0.18	4.84
		Non-ASP/Non-ASP	137	2.18	1.07	0.26	9.07
	Vβ 20	ASP/ASP	39	2.23	0.95	0.20	4.09
		Non-ASP/0602	43	2.15	1.08	0.02	5.16
		Non-ASP/ASP	178	2.07	0.99	0.06	7.22
		Non-ASP/Non-ASP	137	2.29	1.17	0.04	7.65

Note: * Results not available for every Vβ expression

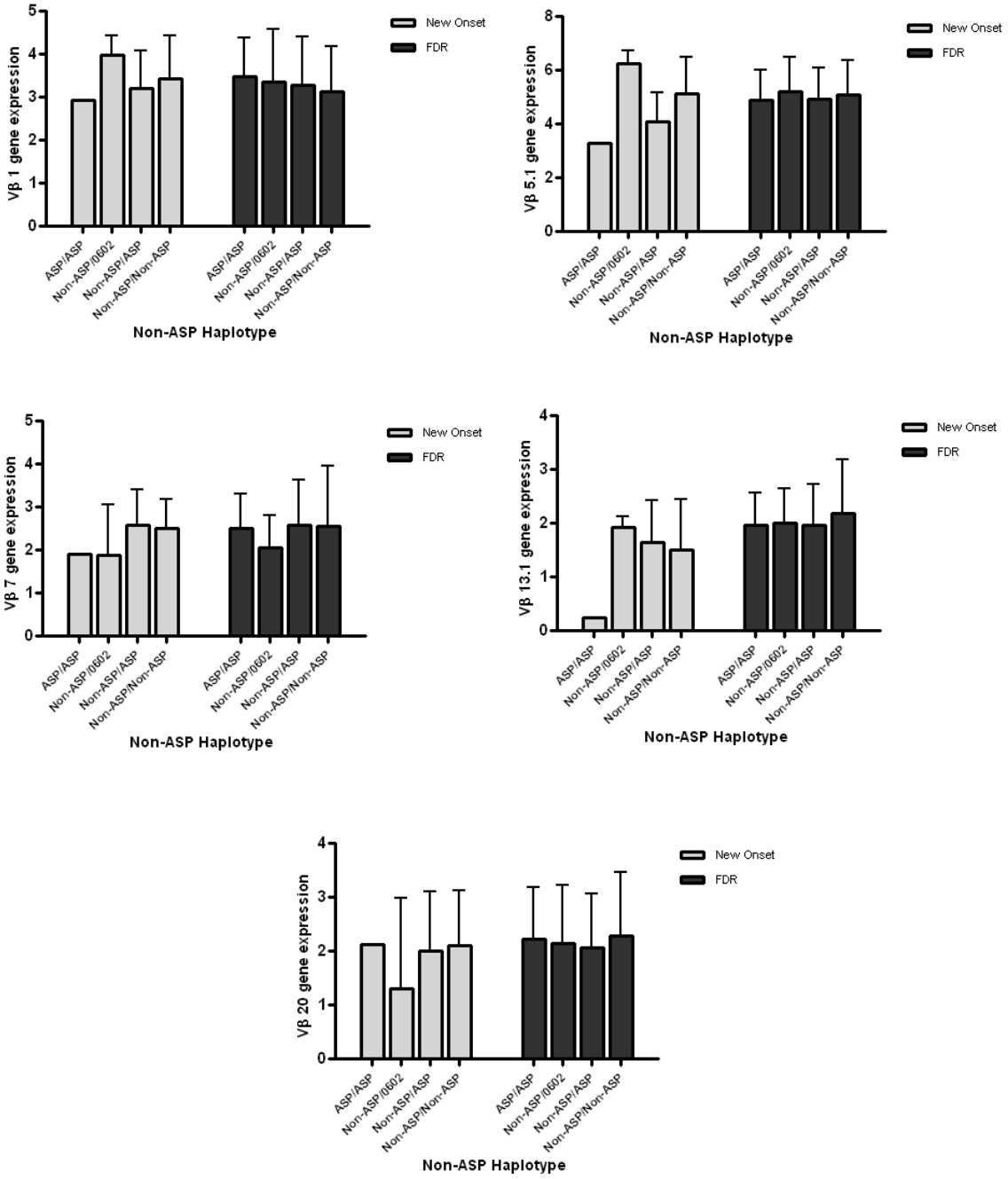

Figure 8: Vβ expressions by non-ASP haplotype

4.3.2 Vβ expressions by DQ2/DQ8 allele

The DQ2 and DQ8 alleles have been shown to trigger an increased genetic risk in developing T1D and are indicators for diabetes risk. The DQ results were stratified by the presence of the DQ2 and DQ8 alleles. Each of the subjects was assessed for the presence of the DQ2 (DQA1*0501-DQB1*0201) and DQ8 (DQA1*0301-DQB1*0302) alleles. Recorded results indicated if there was one (DQ2/X) or two (DQ2/DQ2) copies of the DQ2 haplotype, one (DQ8/X) or two (DQ8/DQ8) copies of the DQ8 haplotype. The order of risk from highest to lowest is DQ2/DQ8 (~1:15), DQ8/DQ8 (~1:60-200), DQ8/X (~1:60-200), DQ2/DQ2 (~1:60-200), and DQ2/X (~1:300) (where "X" indicates the presence of any other alpha and beta allele with the exception of DQB1*0602) [15].

4.3.2.1 Vβ expressions by DQ2 allele

Individuals were classified by the presence of 0 (X/X), 1 (DQ2/X), or 2 (DQ2/ DQ2) copies of the DQ2 alleles (where "X" indicates the presence of any other alpha and beta allele with the exception of DQB1*0602).

In regard to the DQ2 alleles, Table 23 shows that 55% of the new onsets and 61% of the FDRs possessed no DQ2 alleles, whereas 45% of the new onsets and 39% of the FDRs had at least 1 DQ2 allele.

Based on the GLM analysis, there were no significant differences in each Vβ expression for new onsets and FDRs. Table 24 and Figure 9 show that there were increasing tendencies in Vβ5.1, Vβ7, and Vβ20 expressions for new onsets with the increase in the number of DQ2 alleles. Similar patterns were also noted in the Vβ1 and Vβ5.1 expressions for FDRs.

Table 23: Number of DQ2 alleles distribution for new onsets and FDRs

Frequency Percent RowPct Col Pct	Number of DQ2 alleles 0	 1	 2	 Total
New onset	42 8.16 54.55 13.50	30 5.83 38.96 15.96	5 0.97 6.49 31.25	77 14.95
FDR	269 52.23 61.42 86.50	158 30.68 36.07 84.04	11 2.14 2.51 68.75	438 85.05
Total	311 60.39	188 36.50	16 3.11	515 100.00

Note: Frequency Missing = 15

Table 24: Vβ expressions by DQ2 allele

Subject	Variable	Number of DQ2 alleles	N*	Mean	Std Dev	Min	Max
New Onset	Vβ 1	0	40	3.31	0.89	1.59	4.83
		1	30	3.53	1.06	1.52	5.93
		2	5	3.19	0.54	2.44	3.89
	Vβ 5.1	0	40	4.59	1.45	1.88	10.2
		1	30	5.16	1.30	3.41	8.12
		2	5	5.36	0.62	4.60	6.00
	Vβ 7	0	40	2.46	0.76	1.06	4.42
		1	30	2.52	0.72	1.36	4.20
		2	5	2.62	0.63	1.63	3.14
	Vβ 13.1	0	40	1.48	0.92	0.00	3.20
		1	30	1.67	0.90	0.06	4.11
		2	5	1.06	0.58	0.12	1.65
	Vβ 20	0	40	1.89	1.14	0.12	4.03
		1	30	2.25	0.96	0.36	5.82
		2	5	2.25	0.65	1.28	2.86
FDR	Vβ 1	0	261	3.22	1.00	0.84	6.30
		1	152	3.27	1.22	1.03	10.3
		2	11	3.73	1.24	1.58	6.10
	Vβ 5.1	0	263	4.86	1.13	1.98	8.14
		1	154	5.31	1.28	1.94	10.9
		2	11	5.40	1.36	2.77	7.57
	Vβ 7	0	263	2.42	0.75	0.76	6.23
		1	154	2.58	1.43	0.88	12.5
		2	11	2.04	0.52	1.36	2.82
	Vβ 13.1	0	261	2.10	0.81	0.18	7.55
		1	154	1.99	0.91	0.24	9.07
		2	11	1.99	1.07	0.68	4.89
	Vβ 20	0	261	2.18	1.07	0.02	7.65
		1	154	2.19	1.04	0.10	6.94
		2	11	1.69	0.76	0.04	2.80

Note: * Results not available for every Vβ expression

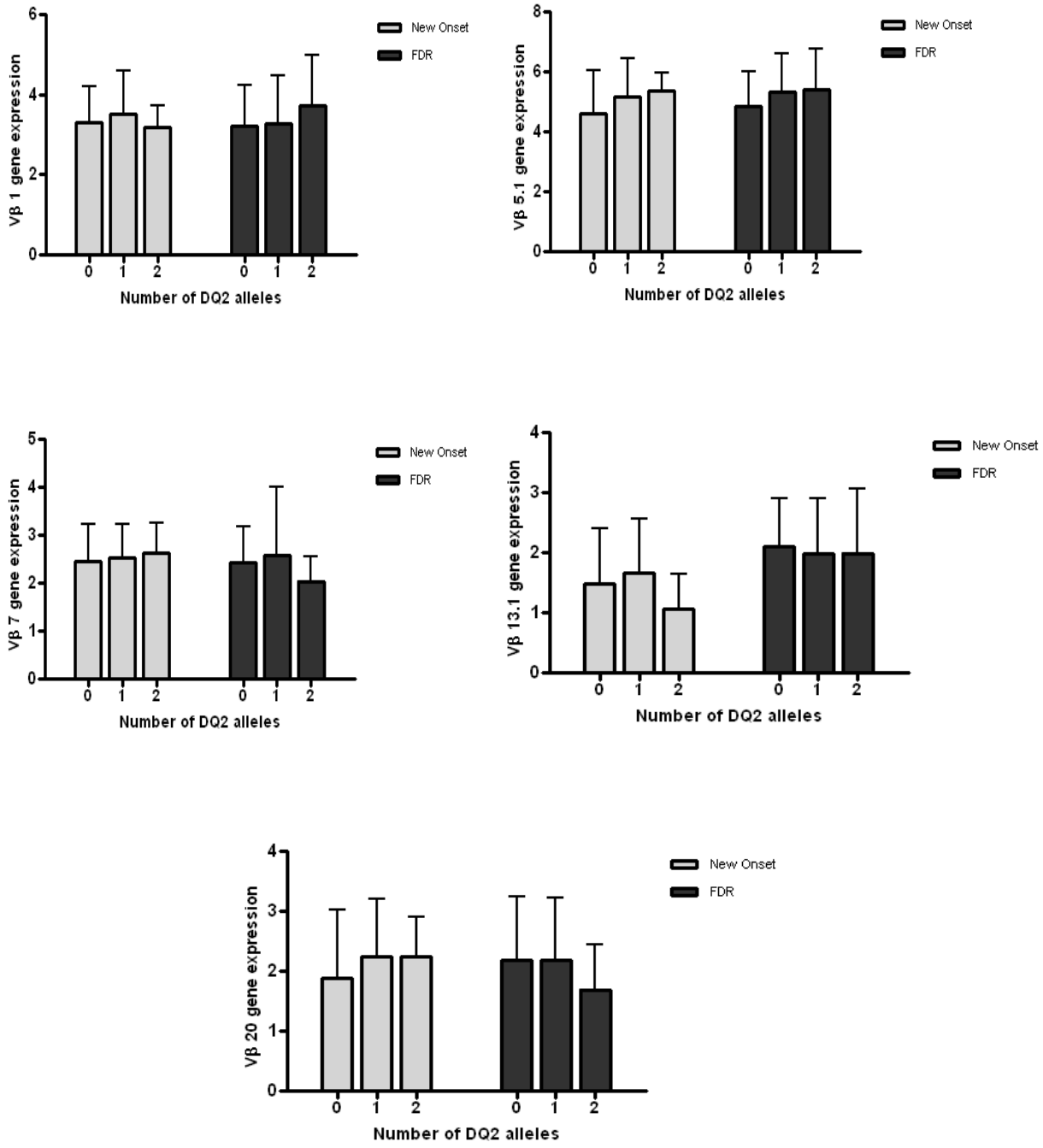

Figure 9: Vβ expressions by DQ2 allele

4.3.2.2 Vβ expressions by DQ8 allele

Individuals were also classified as having 0 (X/X), 1 (DQ8/X), or 2 (DQ8/ DQ8) copies of the DQ8 alleles (where "X" indicates the presence of any other alpha and beta allele with the exception of DQB1*0602).

Table 25 indicates that 62% of the FDRs had no DQ8 alleles and 38% possessed at least one DQ8 allele. In the breakdown, almost 36% were DQ8 heterozygotes and 2% were DQ8 homozygotes. Of the 77 new onsets, individuals with 0, 1, and 2 DQ8 haplotypes were 52%, 43%, and 5%, respectively.

Referencing to Table 26 and Figure 10, for the Vβ13.1 gene family, both new onsets and FDRs carrying zero DQ8 allele showed the lowest value. As the number of the DQ8 alleles increased, the Vβ13.1 expression for both new onsets and FDRs increased. A similar tendency can be observed in the Vβ1 gene family for new onsets. However, the results based on the GLM analysis indicate that there were no significant differences in each Vβ expression by the number of DQ8 alleles for both new onsets and FDRs.

Table 25: Number of DQ8 alleles distribution for new onsets and FDRs

Frequency Percent RowPct Col Pct	Number of DQ8 alleles 0	 1	 2	 Total
New onset	40	33	4	77
	7.77	6.41	0.78	14.95
	51.95	42.86	5.19	
	12.90	17.37	26.67	
FDR	270	157	11	438
	52.43	30.49	2.14	85.05
	61.64	35.84	2.51	
	87.10	82.63	73.33	
Total	310	190	15	515
	60.19	36.89	2.91	100.00

Note: Frequency Missing = 15

Table 26: Vβ expressions by DQ8 allele

Subject	Variable	Number of DQ8 alleles	N*	Mean	Std Dev	Min	Max
New Onset	Vβ 1	0	39	3.34	0.91	1.52	5.87
		1	32	3.42	1.04	1.67	5.93
		2	4	3.67	0.33	3.29	4.09
	Vβ 5.1	0	39	4.76	1.23	2.42	8.08
		1	32	5.04	1.59	1.88	10.2
		2	4	4.61	0.96	3.26	5.48
	Vβ 7	0	39	2.58	0.66	1.33	4.20
		1	32	2.40	0.81	1.06	4.42
		2	4	2.34	0.64	1.67	2.98
	Vβ 13.1	0	39	1.50	0.95	0.00	4.11
		1	32	1.54	0.88	0.00	3.09
		2	4	1.70	0.53	1.18	2.40
	Vβ 20	0	39	2.15	1.05	0.20	5.82
		1	32	1.89	1.07	0.12	4.03
		2	4	2.48	0.79	1.45	3.31
FDR	Vβ 1	0	262	3.35	1.13	1.02	10.3
		1	151	3.09	1.03	0.84	6.95
		2	11	2.92	0.79	1.23	3.80
	Vβ 5.1	0	264	5.15	1.20	1.94	10.9
		1	153	4.85	1.21	1.98	8.14
		2	11	4.73	1.16	3.24	6.26
	Vβ 7	0	264	2.48	1.13	1.10	12.5
		1	153	2.45	0.91	0.76	6.63
		2	11	2.55	0.65	1.42	3.70
	Vβ 13.1	0	262	1.97	0.71	0.18	4.89
		1	153	2.17	1.04	0.26	9.07
		2	11	2.35	0.92	0.82	4.14
	Vβ 20	0	262	2.16	0.98	0.04	5.16
		1	153	2.19	1.19	0.02	7.65
		2	11	2.12	0.83	0.22	3.28

Note: * Results not available for every Vβ expression

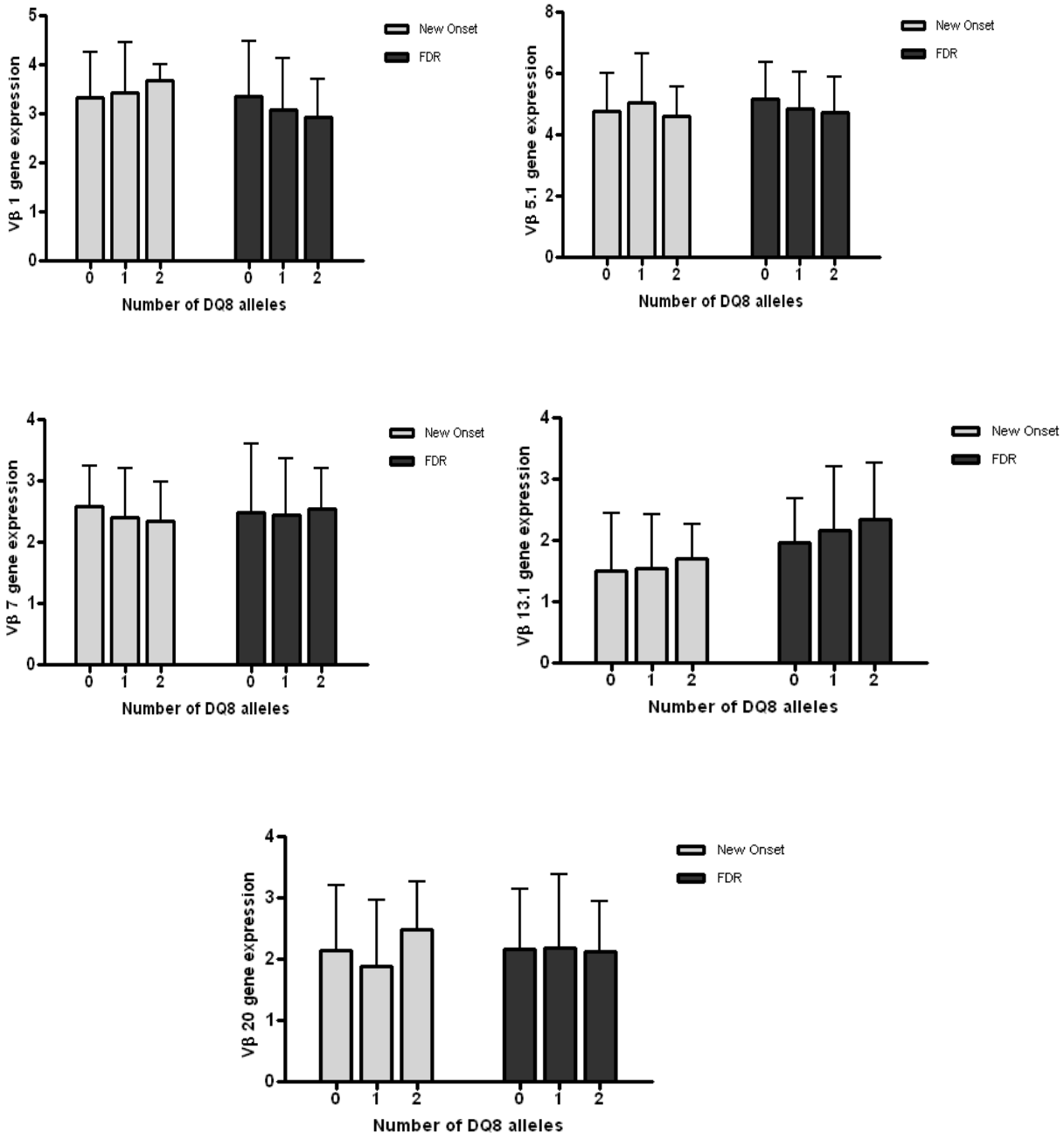

Figure 10: Vβ expressions by DQ8 allele

4.4 AGE EFFECT ON VB EXPRESSIONS

Table 27 shows that the average age for new onsets was 9.4, ranging from 1 to 18, and the average age for FDRs was 33.7, ranging from 2 to 54. In new onsets, all Vβ expressions were not correlated with age (Table 28 and Figure 11). However, the evidence that the Vβ5.1 expression was correlated with age was observed in FDRs (Table 28 and Figure 12).

We further investigated the expression of Vβ gene families among 5-year age groups. Table 33 and Figure 13 in Appendix A show Vβ expressions by age group for new onsets. Table 34 and Figure 14 in Appendix B show Vβ expressions by age group for FDRs. The GLM analyses were conducted to test the significance of the age groups that may contribute to any differences in each Vβ gene family. The results demonstrate that each Vβ gene family appeared to be not significantly different among four age groups in new onsets. But in FDRs the Vβ5.1 expression in age group 50 to 55 years (mean = 6.83%±0.68SD) was significantly higher than other age groups ($p = 0.0015$).

Table 27: Mean ages for new onsets and FDRs

Subjects	N	Mean	Std Dev	Min	Max
New Onset	84	9.4	4.16	1.0	18.0
FDR	443	33.7	12.72	2.0	54.0

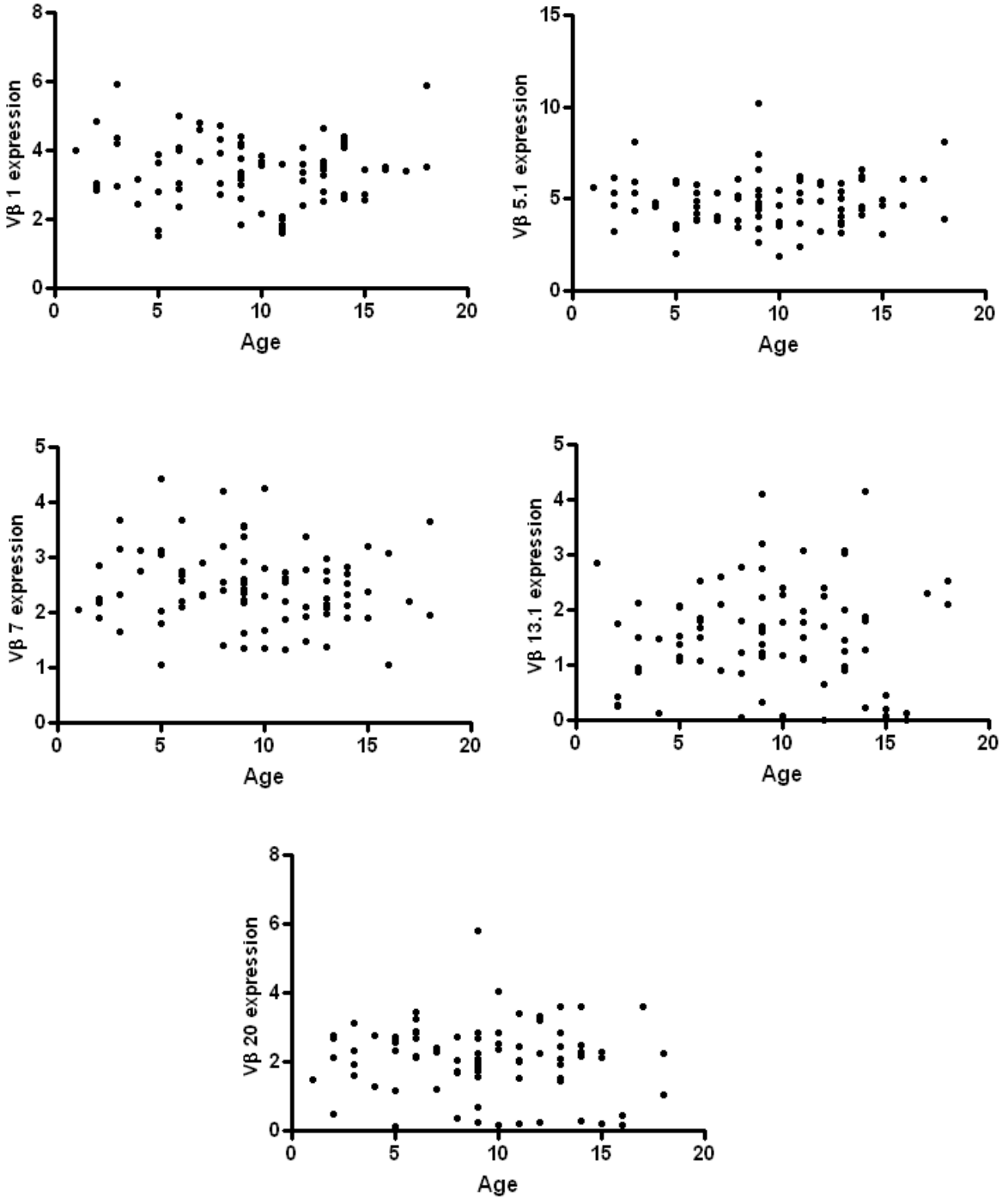

Figure 11: Scatter plots for Vβ expressions vs age for new onsets

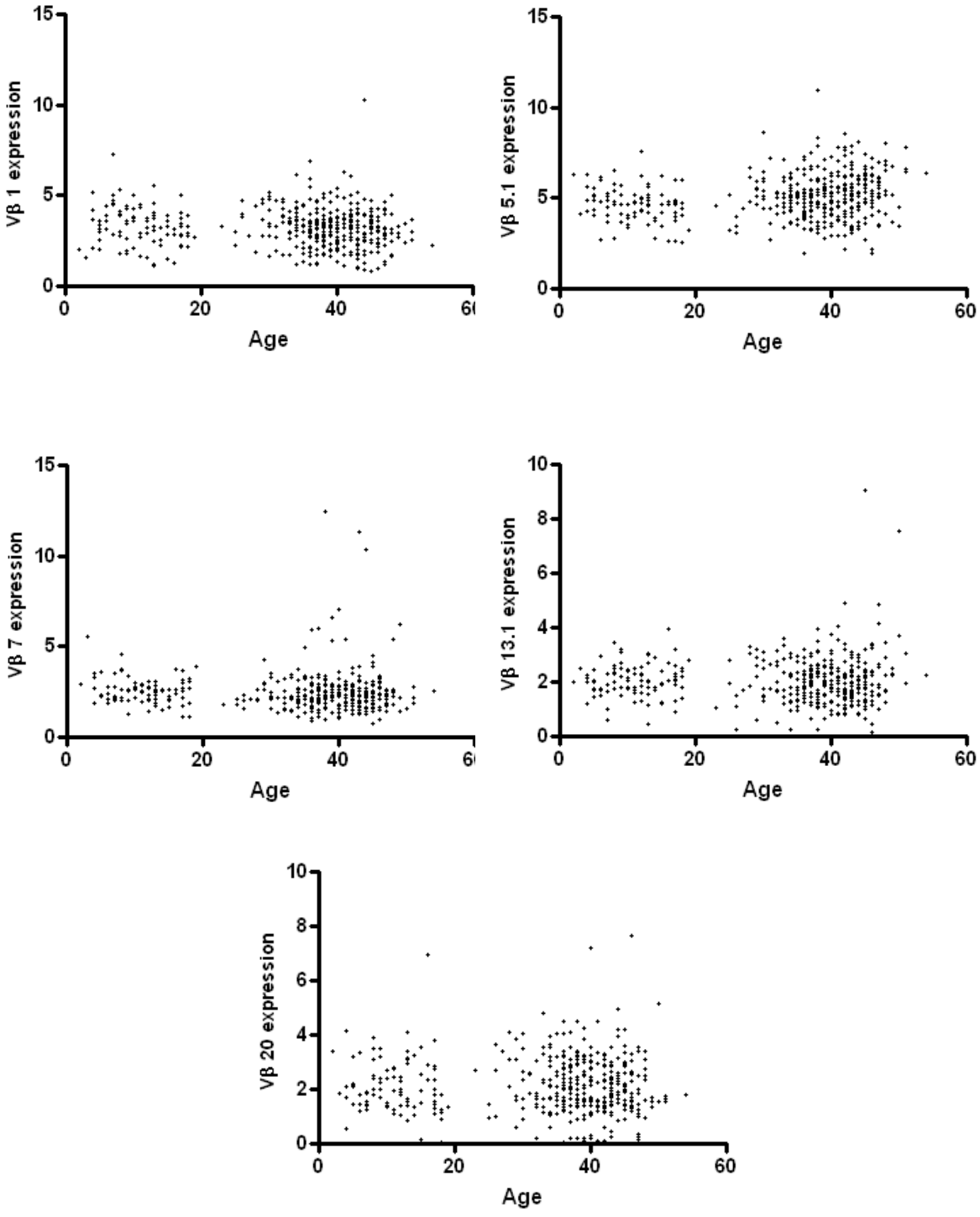

Figure 12: Scatter plots for Vβ expressions vs age for FDRs

Table 28: Correlations between Vβ expressions and age

Coefficient p-value	Vβ 1	Vβ 5.1	Vβ 7	Vβ 13.1	Vβ 20
New Onset	-.0404 0.7185	0.0222 0.8428	-.1076 0.3359	0.0364 0.7457	-.0702 0.5307
FDR	-.0808 0.0949	0.1848 **0.0001**	-.0329 0.4959	0.0305 0.5288	-.0131 0.7862

4.5 LOGISTIC REGRESSION

In the final portion of our analysis, logistic regression modeling was conducted to identify independent factors that are associated with Type 1 diabetes. We included all FDRs and only new onsets with a blood draw within seven days of diagnosis. The factors considered included Vβ1, Vβ5.1, Vβ7, Vβ13.1, Vβ20, age at the blood draw, the number of the DQ2 and DQ8 alleles, race and gender. HLA haplotype was not included in the models as only one new onset was in the ASP/ASP group, the remainder being Non ASP. Four logistic regression models were built for the purpose of assessing the factors associated with T1D. Model 1 was fitted with all of the 10 independent variables and included DQ2 and DQ8 as categorical variables; Model 2 included all of these variables and built the regression using forward selection. Model 3 was fitted using all 10 explanatory variables treating DQ2 and DQ8 as continuous; Model 4 included all of these variable and built the regression using forward selection.

Irrespective of the way DQ2 and DQ8 were treated, categorical or continuous, Vβ7, DQ8, and age at blood draw were predictors (Tables 29 and Table 31). The estimates and odds ratios show that the lower Vβ7 expression indicates higher risk of the disease. In addition, with the

increase in the number of DQ8 alleles, the risk of developing T1D increased significantly, whereas the risk of developing T1D decreased significantly as the age at blood draw increased.

When we fitted the models 2 and 4 with forward selection, we found that Vβ1, Vβ13.1, DQ8, and age at the blood draw were important factors associated with T1D. From estimates and odds ratios in Table 30 and Table 32, we see that higher expression of Vβ1 and Vβ13.1 gene families, as well as the number of DQ8 alleles is associated with increased risk of the disease. In contrast an inverse relationship with age at blood draw was noted with T1D.

Hosmer and Lemeshow goodness-of-fit tests all show the model fits well (all p values >0.9000). Based on the AIC criteria, the lowest value was observed in model 4 (Table 32) and suggests that this model is the "best" of the four models.

Table 29: Parameter estimates, test statistics, and odds ratio for model 1

Parameter	Estimate	Standard Error	p-value of Wald Statistics	Odds Ratio	AIC
Intercept	-4.314	2.545	0.090	0.013	97.872
Vβ 1	0.575	0.316	**0.069**	1.778	
Vβ 5.1	0.506	0.340	0.137	1.659	
Vβ 7	-1.162	0.584	**0.046**	0.313	
Vβ 13.1	0.537	0.389	0.168	1.710	
Vβ 20	0.311	0.309	0.315	1.364	
DQ2-1	0.629	0.729	0.388	1.876	
DQ2-2	-12.00	558.1	0.983	0.000	
DQ8-1	1.175	0.743	0.114	3.239	
DQ8-2	2.966	1.260	**0.019**	19.41	
Race(black)	-0.099	1.689	0.953	0.906	
Gender(female)	-0.483	0.680	0.478	0.617	
Age at blood draw	-0.183	0.047	**<0.001**	0.833	

Table 30: Parameter estimates, test statistics, and odds ratio for model 2

Parameter	Estimate	Standard Error	p-value of Wald Statistics	Odds Ratio	AIC
Intercept	-5.171	1.550	**0.001**	0.006	91.719
Vβ 1	0.760	0.278	**0.006**	2.139	
Vβ 13.1	0.769	0.358	**0.032**	2.157	
DQ8-1	1.087	0.642	**0.090**	2.965	
DQ8-2	2.649	1.074	**0.014**	14.15	
Age at blood draw	-0.171	0.043	**<0.001**	0.843	

Table 31: Parameter estimates, test statistics, and odds ratio for model 3

Parameter	Estimate	Standard Error	p-value of Wald Statistics	Odds Ratio	AIC
Intercept	-4.638	2.555	0.070	0.010	95.605
Vβ 1	0.564	0.320	**0.078**	1.758	
Vβ 5.1	0.510	0.337	0.130	1.665	
Vβ 7	-1.080	0.571	**0.059**	0.340	
Vβ 13.1	0.511	0.375	0.174	1.666	
Vβ 20	0.348	0.303	0.251	1.416	
DQ2	0.229	0.635	0.718	1.258	
DQ8	1.371	0.581	**0.018**	3.939	
Race(black)	0.203	1.659	0.902	1.225	
Gender(female)	-0.457	0.683	0.503	0.633	
Age at blood draw	-0.173	0.044	**<0.001**	0.841	

Table 32: Parameter estimates, test statistics, and odds ratio for model 4

Parameter	Estimate	Standard Error	p-value of Wald Statistics	Odds Ratio	AIC
Intercept	-5.194	1.547	**0.001**	0.006	**89.849**
Vβ 1	0.754	0.277	**0.007**	2.126	
Vβ 13.1	0.751	0.353	**0.033**	2.120	
DQ8	1.238	0.494	**0.012**	3.449	
Age at blood draw	-0.169	0.042	**<0.001**	0.845	

5.0 CONCLUSIONS

As this is a preliminary analysis of the partially available data, we presented only general summaries of the data. All of our work was exploratory and not hypothesis testing.

For new onsets, correlations were indentified between the number of days from diagnosis and Vβ expression. Vβ expression in blood samples collected within one week of diagnosis tended to be higher than those collected further away. For Vβ1 and Vβ13.1 this difference was significant and for Vβ20 approached significance. We had multiple blood draws for a few individuals and were able to see a general decrease in Vβ expression at later times. Thus further suggesting a time relationship for Vβ expression. Therefore when characterizing Vβ expression at the time of diagnosis, it is important to obtain samples within a few days of onset. Sample collected at later times may not reflect the activity at baseline. In analyzing Vβ expression, adjustment needs to be made for time since diagnosis.

To investigate the differences in Vβ expressions between new onsets and FDRs, we first assessed the differences in the expression of Vβ gene families between FDRs and new onsets whose first blood sample was drawn within 7 days of diagnosis. The results show that for the Vβ1 gene family, a significantly lower value was observed among FDRs compared with the new onsets whose first blood sample was drawn within 7 days of diagnosis ($p = 0.0017$). When we expanded the number of days since diagnosis in new onsets to 100 days a significantly lower value was observed in the Vβ13.1 gene family for new onsets compared with FDRs. These

findings are in contrast with one another and the results of the second comparison showing a lower Vβ13.1 expression in new onsets is counter intuitive. The most possible reason is the time effect on Vβ expression. At onset of diabetes Vβ expression is increased. However, as time since diagnosis elapses, Vβ expression decreases so that the Vβ1 expression in new onsets was not significantly higher than in FDRs, and the Vβ13.1 expression in new onsets decreased. Therefore, it is highly recommended that the study of baseline Vβ expression be restricted to using blood samples within 7 days of diagnosis.

The forgoing summarizes Vβ expression by various subject characteristics. We did not adjust for time since diagnosis as the choice of samples analyzed appear to be random with respect to the number of days since diagnosis. Among FDRs, there is no corresponding metric, and thus no adjustment was possible. For Vβ expressions by gender, the Vβ1 expression in male FDRs was significantly higher than that in females ($p = 0.0095$). As for Vβ expressions by race, the Vβ7 expression in the black new onsets was significantly higher than that in the white new onsets ($p = 0.0057$). For the Vβ 1 family, black FDRs showed significantly higher value than white FDRs ($p = 0.0216$). The Vβ 13.1 expression for black FDRs was significantly lower than that for white FDRs ($p = 0.0005$). Vβ expressions were not associated with the presence of positive antibodies GAD, IA-2, IAA, ICA, or the number of positive autoantibodies. Findings from the GLM analysis suggest no association of Vβ expressions with non-ASP status, the DQ2 alleles, and the DQ8 alleles for either the new onsets or FDRs. Assessment of age with Vβ gene expressions did not reveal any real association.

The logistic regression analysis was conducted for the purpose of assessing the factors associated with T1D. All FDRs and those new onsets with Vβ results available within 7 days of diagnosis were analyzed. The potential factors considered were Vβ1, Vβ5.1, Vβ7, Vβ13.1,

Vβ20, age at the blood draw, the number of the DQ2 or DQ8 alleles, race and gender. The HLA haplotype was not involved in the analyses due to only one new onset in the ASP/ASP group. Four logistic regression models were built. In different models, we treated DQ2 and DQ8 as categorical and continuous. The models showed similar conclusions that the Vβ1 and Vβ13.1 gene families, DQ8 alleles, and lower age were independently associated with the onset of T1D.

In summary, there appears to be some evidence that Vβ expression is higher at the diagnosis of T1D in new onsets compared to a control group of FDRs. Furthermore, this elevated level of Vβ expression is transient and begins to diminish over time. These levels may eventually be similar to or possibly lower than levels found in FDRs.

APPENDIX A

VB EXPRESSION BY 5-YEAR AGE GROUP FOR NEW ONSETS

Table 33: Vβ expressions by age group for new onsets

Variable	Age Group	N	Mean	Std Dev	Min	Max
Vβ 1	**<6 yrs**	17	3.42	1.11	1.52	5.93
	6 - 10 yrs	32	3.55	0.83	1.85	5.01
	11 - 15 yrs	28	3.14	0.86	1.59	4.63
	16 - 20 yrs	5	3.95	1.07	3.40	5.87
Vβ5.1	**<6 yrs**	17	5.01	1.42	2.03	8.12
	6 - 10 yrs	32	4.77	1.48	1.88	10.2
	11 - 15 yrs	28	4.80	1.13	2.42	6.60
	16 - 20 yrs	5	5.77	1.60	3.93	8.08
Vβ 7	**<6 yrs**	17	2.55	0.83	1.06	4.42
	6 - 10 yrs	32	2.60	0.74	1.36	4.25
	11 - 15 yrs	28	2.32	0.51	1.33	3.37
	16 - 20 yrs	5	2.39	1.01	1.06	3.65
Vβ13.1	**<6 yrs**	17	1.29	0.76	0.12	2.86
	6 - 10 yrs	32	1.72	0.87	0.06	4.11
	11 - 15 yrs	28	1.54	1.01	0.00	4.16
	16 - 20 yrs	5	1.42	1.24	0.00	2.54
Vβ20	**<6 yrs**	17	2.01	0.86	0.12	3.13
	6 - 10 yrs	32	2.23	1.11	0.16	5.82
	11 - 15 yrs	28	2.11	1.00	0.20	3.61
	16 - 20 yrs	5	1.50	1.43	0.16	3.60

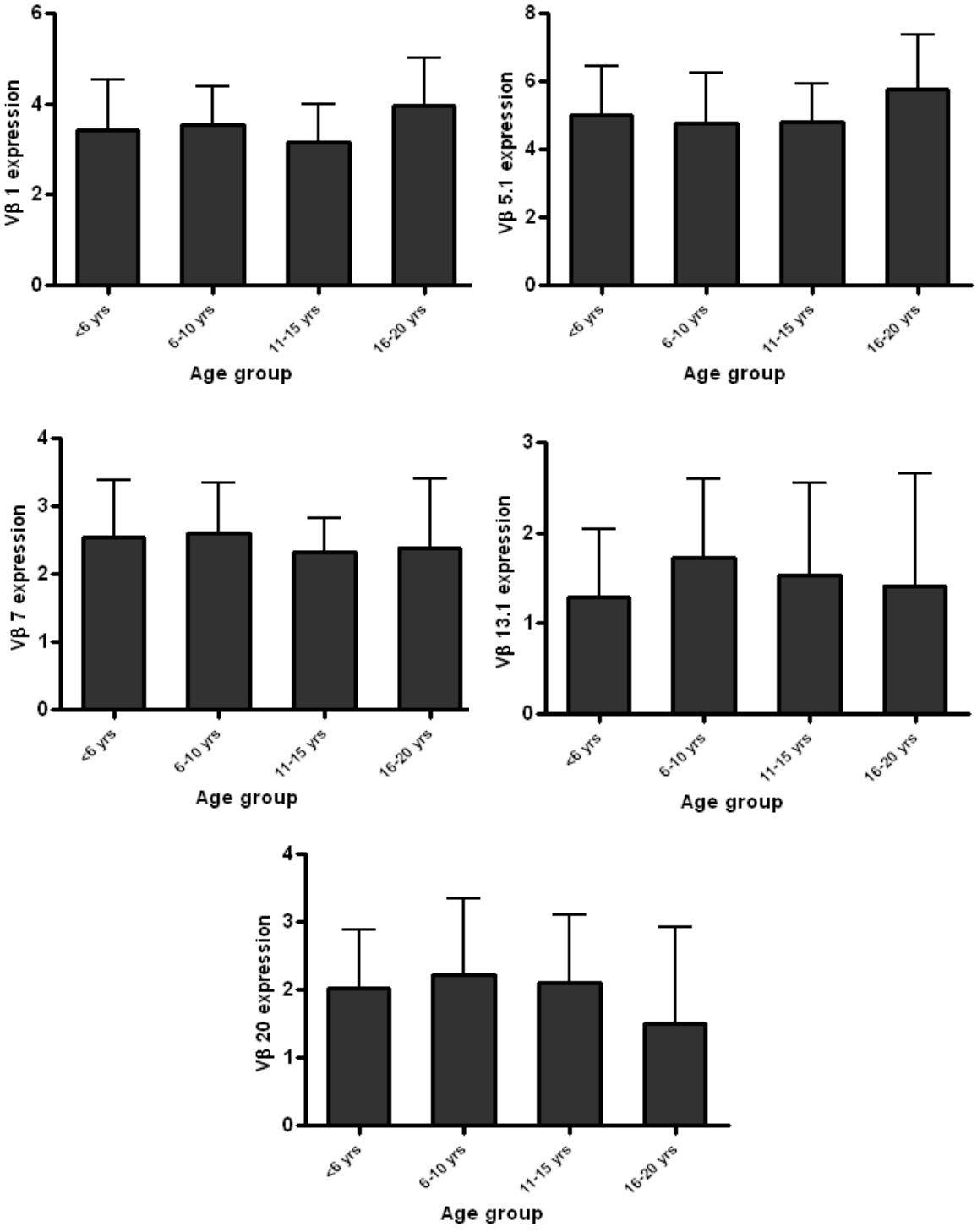

Figure 13: Vβ expressions by 5-year age group for new onsets

APPENDIX B

VB EXPRESSIONS BY 5-YEAR AGE GROUP FOR FDR'S

Table 34: Vβ expressions by age group for FDRs

Variable	Age Group	N	Mean	Std Dev	Min	Max
Vβ 1	**<6 yrs**	12	3.05	1.04	1.61	5.20
	6 - 10 yrs	30	3.80	1.17	1.81	7.31
	11 - 15 yrs	31	3.14	1.02	1.13	5.53
	16 - 20 yrs	19	3.09	0.83	1.32	5.01
	21 - 25 yrs	3	2.85	0.54	2.25	3.30
	26 - 30 yrs	20	3.76	1.04	1.66	5.17
	31 - 35 yrs	48	3.31	1.05	1.23	6.13
	36 - 40 yrs	112	3.22	1.02	1.18	6.95
	41 - 45 yrs	109	3.20	1.26	0.84	10.3
	46 - 50 yrs	41	3.00	0.93	1.02	5.03
	51 - 55 yrs	4	2.89	0.64	2.26	3.72
Vβ 5.1	**<6 yrs**	12	5.07	0.78	4.10	6.32
	6 - 10 yrs	30	4.66	0.99	2.69	6.51
	11 - 15 yrs	31	4.76	0.92	2.82	7.57
	16 - 20 yrs	19	4.24	1.00	2.54	6.03
	21 - 25 yrs	3	4.35	0.98	3.27	5.19
	26 - 30 yrs	20	5.32	1.23	3.12	8.61
	31 - 35 yrs	48	4.99	0.95	2.74	7.23
	36 - 40 yrs	113	5.04	1.29	1.94	10.9
	41 - 45 yrs	112	5.20	1.30	2.16	8.56
	46 - 50 yrs	41	5.33	1.32	1.98	8.04
	51 - 55 yrs	4	6.83	0.68	6.37	7.83
Vβ 7	**<6 yrs**	12	2.91	1.00	1.87	5.53
	6 - 10 yrs	30	2.63	0.70	1.32	4.58
	11 - 15 yrs	31	2.37	0.44	1.46	3.09

Table 34 continued

Variable	Age Group	N	Mean	Std Dev	Min	Max
Vβ 7	**16 - 20 yrs**	19	2.69	0.80	1.11	3.92
	21 - 25 yrs	3	1.91	0.17	1.79	2.11
	26 - 30 yrs	20	2.52	0.66	1.56	4.29
	31 - 35 yrs	48	2.32	0.75	1.16	4.94
	36 - 40 yrs	113	2.53	1.37	0.88	12.5
	41 - 45 yrs	112	2.47	1.34	0.76	11.3
	46 - 50 yrs	41	2.44	0.95	0.96	6.23
	51 - 55 yrs	4	2.36	0.41	1.88	2.80
V β 13.1	**<6 yrs**	12	1.97	0.38	1.23	2.53
	6 - 10 yrs	30	2.13	0.63	0.62	3.47
	11 - 15 yrs	31	1.94	0.59	0.46	3.07
	16 - 20 yrs	19	2.25	0.68	0.90	3.95
	21 - 25 yrs	3	1.94	0.89	1.04	2.81
	26 - 30 yrs	20	2.03	0.88	0.27	3.31
	31 - 35 yrs	47	1.94	0.80	0.24	3.60
	36 - 40 yrs	113	2.04	0.66	0.26	3.96
	41 - 45 yrs	111	1.98	1.00	0.66	9.07
	46 - 50 yrs	41	2.38	1.29	0.18	7.55
	51 - 55 yrs	4	2.31	0.53	1.94	3.08
V β 20	**<6 yrs**	12	2.27	0.94	0.58	4.15
	6 - 10 yrs	30	2.13	0.86	0.02	3.92
	11 - 15 yrs	31	2.14	0.94	0.16	4.09
	16 - 20 yrs	19	2.09	1.46	0.04	6.94
	21 - 25 yrs	3	1.70	0.88	0.98	2.69
	26 - 30 yrs	20	2.50	1.10	0.63	4.09
	31 - 35 yrs	47	2.23	0.96	0.20	4.81
	36 - 40 yrs	113	2.20	1.09	0.06	7.22
	41 - 45 yrs	111	2.11	0.96	0.04	4.95
	46 - 50 yrs	41	2.13	1.33	0.17	7.65
	51 - 55 yrs	4	1.70	0.10	1.58	1.79

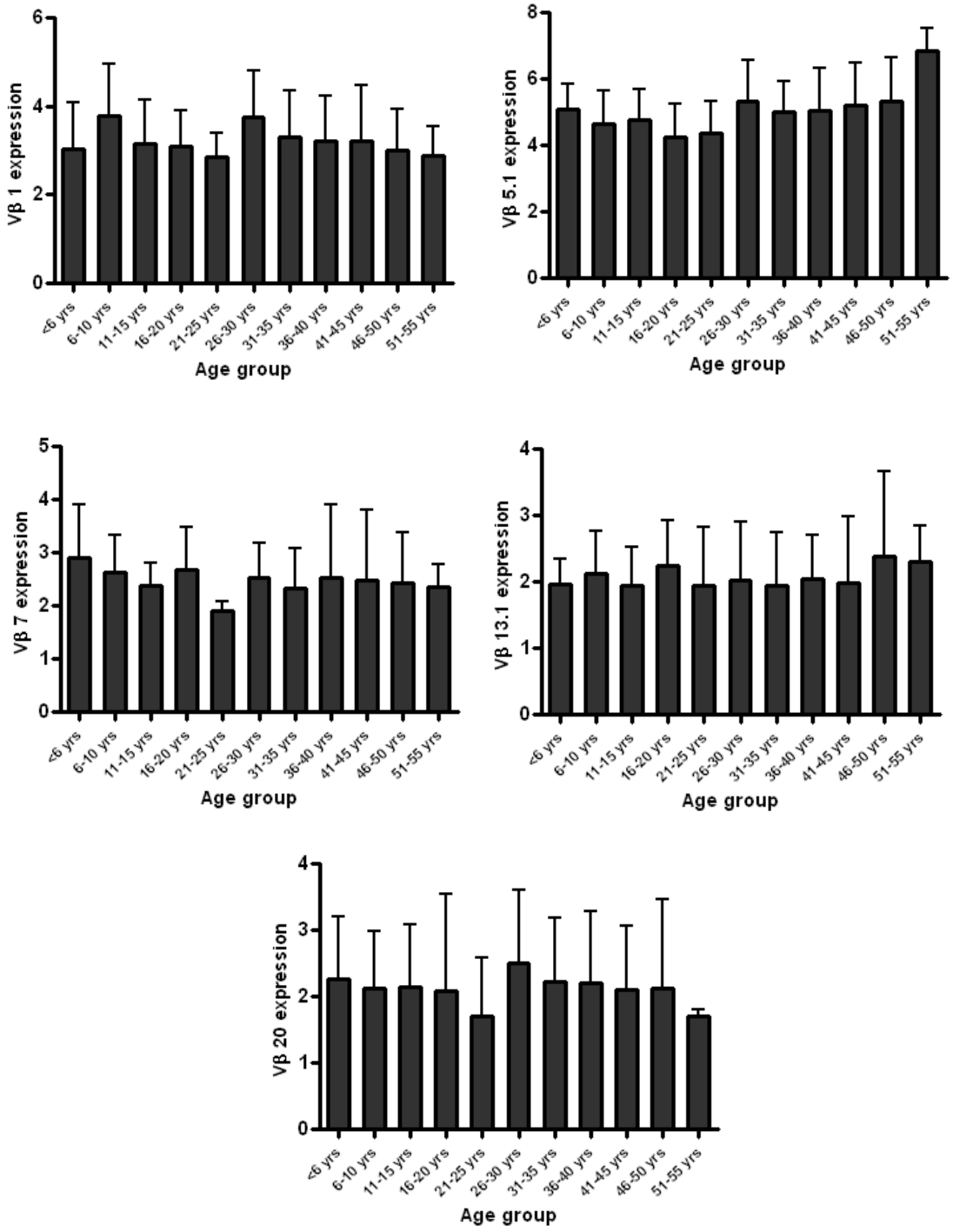

Figure 14: Vβ expression by 5-year age group for FDRs

APPENDIX C

SAS PROGRAM USED FOR THE ANALYSIS

```
libname   library  "F:\thesis" ;
filename  format   "F:\thesis\jod_formats_08_05_01.sas" ;
filename  format_m   "F:\thesis\jod_formats_mei.sas" ;

run ;

*  Use and create SAS format from existing program;

%include format ;
%include format_m ;
run ;


title1     "VB Study" ;


*  Create permanent SAS data set ;

data library.vb_info  (drop = i) ;

     set library.vb_info_abi_08_07_23 ;

*    Assigning missing values to the data ;
     missing  m d n r s b ;

     array tmpvar_n(*)   _NUMERIC_  ;
     do i=1 to dim(tmpvar_n) ;
        if tmpvar_n(i)=-999 then tmpvar_n(i)="m" ;
        if tmpvar_n(i)=-888 then tmpvar_n(i)="s" ;
        if tmpvar_n(i)=-777 then tmpvar_n(i)="b" ;
     end ;

     array tmpvar_c(*)   _character_ ;
     do i=1 to dim(tmpvar_c) ;
        if compress(tmpvar_c(i))="-999" then tmpvar_c(i)="m" ;
        if compress(tmpvar_c(i))="-888" then tmpvar_c(i)="d" ;
     end ;
```

```
      attrib  pid  label = "Participant ID" ;

      *  Extracted date part from variable ;
      blooddrawdate = datepart(blooddrawdate) ;
      attrib  blooddrawdate  format = date9.  label = "Blood draw date" ;
      attrib  countofpid  label = "Number of samples" ;
      attrib  fid  label = "Family ID" ;
      attrib  ags  format = lags.  label = "AGS study" ;
      attrib  agsdisp  format = lagsdisp.  label = "AGS study" ;
      attrib  gender  format = lgender.  label = "Gender" ;
      attrib  race  format = lrace.  label = "Race" ;

      *  Extracted date part from variable ;
      dob = datepart(dob) ;
      attrib  dob  format = date9.  label = "Birthdate" ;

      *  Extracted date part from variable ;
      ddx = datepart(ddx) ;
      attrib  ddx  format = date9.  label = "Date of diagnosis" ;

      attrib  age_dx  label = "Age at diagnosis" ;
      attrib  age_blooddraw  label = "Age at blood draw" ;
      attrib  nonasp_status  format = lasp.  label = "ASP status" ;
      attrib  calc_nonasp  format = lcalc_nonasp.  label = "ASP
            haplotype" ;
      attrib  calc_dq2  label = "Number of DQ2 alleles" ;
      attrib  calc_dq8  label = "Number of DQ8 alleles" ;
      attrib  vb_1      label = "VB 1 percent expression" ;
      attrib  vb_5_1    label = "VB 5.1 percent expression" ;
      attrib  vb_7      label = "VB 7 percent expression" ;
      attrib  vb_13_1   label = "VB 13.1 percent expression" ;
      attrib  vb_20     label = "VB 20 percent expression" ;

run ;



/* Create data set only for the first blood draw*/

proc sort  data =  library.vb_info ;
    by pid blooddrawdate;
run ;

data library.vb_subjec_newonset library.vb_subjec_fdr ;
      set library.vb_info ;
      by pid blooddrawdate;
      pid_first = first.pid;
      pid_last = last.pid;
      if first.pid;

      n_day = blooddrawdate - ddx;
      age = floor((blooddrawdate - dob)/365.25);

      if  0  <= age <6   then agegp_5 = 1;
      if  6 <= age <11 then agegp_5 = 2;
      if 11 <= age <16 then agegp_5 = 3;
      if 16 <= age <21 then agegp_5 = 4;
```

```
      if 21 <= age <26 then agegp_5 = 5;
      if 26 <= age <31 then agegp_5 = 6;
      if 31 <= age <36 then agegp_5 = 7;
      if 36 <= age <41 then agegp_5 = 8;
      if 41 <= age <46 then agegp_5 = 9;
      if 46 <= age <51 then agegp_5 = 10;
      if 51 <= age <56 then agegp_5 = 11;
      if 56 <= age <61 then agegp_5 = 12;

      if ags in(1 2 5) then output library.vb_subjec_newonset;
      else output library.vb_subjec_fdr;
run ;

/* Create data set only for the probands with the first blood draw within 100
days of diagnosis */

data library.vb_subjec_newonset;
     set library.vb_subjec_newonset;
     if -7 < n_day <=100 and age  < 20 ;
run;

/* Create data set only for the probands with the first blood draw within 7
days of diagnosis */

data library.vb_subjec_newonset_7;
     set library.vb_subjec_newonset;
     if -7 < n_day <= 7 and age  < 20 ;
run;

/*create data set with unaffected FDRs*/

data library.vb_subjec_fdr;
     set library.vb_subjec_fdr;
     if  n_day = . ;
run;

/* Create data set for the probands with the first blood draw within 100 days
of diagnosis and unaffected FDRs*/

data library.vb_subject;
     merge library.vb_subjec_newonset library.vb_subjec_fdr;
     by pid;
     if ags in(1 2 5) then newonset = 1;
     else if ags in(3 4)then newonset = 0;
run;

/* Create data set for the probands with the first blood draw within 7 days
of diagnosis and unaffected FDRs*/

data library.vb_subject_7;
     merge library.vb_subjec_newonset_7 library.vb_subjec_fdr;
     by pid;
     if ags in(1 2 5) then newonset = 1;
     else if ags in(3 4)then newonset = 0;
run;
```

```
/*create the data set containing variable, n_samples.*/

proc freq data = library.vb_info ;
     tables pid / out = vb_freq (drop=percent rename=(count=n_samples));
run;

data library.vb_freq;
     set vb_freq;
     label n_samples = 'Number of samples';
run;

title3    "Participant Summary" ;

proc freq data = library.vb_freq;
     tables n_samples;
run;


/* Basic information for subjects and samples*/
proc freq  data =  library.vb_subject ;
     tables  gender  race  ags ;
     tables  n_day nonasp_status  calc_nonasp  calc_dq2  calc_dq8 ;
run ;

proc freq  data =  library.vb_subject ;
     tables  ags * (gender  race ) ;
     format  ags lags_grouped. ;
run;

proc freq  data =  library.vb_subject ;
     tables  ags ;
     format  ags lags_grouped. ;
run ;

proc freq  data =  library.vb_subject ;
     tables  ags  * (nonasp_status  calc_nonasp  calc_dq2 calc_dq8)
           /chisq trend ;
     format  ags lags_grouped. ;
run ;

proc freq  data =  library.vb_subject ;
     where calc_nonasp in(0 1 2 3);
     tables  ags  *  calc_nonasp /fisher trend expected;
     format  ags lags_grouped. ;
run ;

proc means data = library.vb_subject  n mean std median p5 p25 p75 p95 min
     max range   fw = 4 maxdec = 2 ;
     var  age_dx ;
run;

proc means data = library.vb_subject  n mean std median p5 p25 p75 p95 min
      max range   fw = 4 maxdec = 2 ;
      class ags;
      var  n_day age;
      format ags lags_grouped.;
run;
```

```
title3    "Sample summary" ;
proc means data = library.vb_info  n mean std median p5 p25 p75 p95 min max
     range   fw = 4 maxdec = 2 ;
     var  age_blooddraw  vb_1 -- vb_20 ;
run ;


/* ************************************************************************/
/*                              New Onset vs FDR                          */
/************************************************************************/

/*Analysis A*/

/* New Onset VS FDR: mean VB expression based on the 1st blood sample*/

/* 1. calculate mean VB expression for New Onset and FDR*/

proc means data = library.vb_subject n mean std median min max range
     nonobs fw = 4 maxdec = 2 ;
     class ags;
     types () ags;
     format ags lags_grouped.;
     var vb_1 -- vb_20;
     title 'Means of VB Expression by AGS Study';
run;

/*2. calculate mean VB expression by days since diagnosis for New Onset */

proc means data = library.vb_subjec_newonset n mean std min max nonobs fw = 4
maxdec = 2 ;
     class n_day;
     format n_day ldaygp.;
     var vb_1 -- vb_20;
     title 'Means of VB Expression by days since diagnosis';
run;

/*a. Scatter plot For new onsets*/

proc gplot data = library.vb_subject;
  where ags in(1 2 5);
  plot (vb_1 vb_5_1 vb_7 vb_13_1 vb_20) * n_day;
  title 'New Onsets: Scatter Plot For VB vs. Days';
run;

quit;

/*b. t-test for VB expression in new onsets*/

proc ttest data = library.vb_subject;
     where ags in(1 2 5) ;
     class n_day;
     var vb_1 vb_13_1 vb_20;
     format n_day ldaygp.;
run;


/*C. correlation between VB's and the days */
```

```
proc corr  data = library.vb_subjec_newonset ;
     var  vb_1 -- vb_20 ;
     with n_day ;
run ;

/* 3. Graphs for VB expression: New onset vs FDR*/

proc sort data = library.vb_subject out = vb_subject;
     by ags;
run;

data vb_subject2;
     set vb_subject;
     if ags in(1 2 5) then ags_group = "New onset";
     else if ags in(3 4) then ags_group = "FDR";
     else ags_group = "Unknown";
     label ags_group = " Subject";
run;


/*For VB1*/

title 'New Onset vs FDR for VB1 Expression';

proc boxplot data = vb_subject2;
     plot vb_1 * ags_group;
run;

proc gchart data = library.vb_subject;
     vbar ags  / sumvar = vb_1 discrete
           type = mean inside = mean
           errorbar = top
           width = 18 space = 4;
     format ags lags_grouped.;
run;

*comparison the difference in VB1 between two groups;

proc ttest data = library.vb_subject_7;
     class newonset;
     var vb_1 --vb_20;
run;

/*4. correlation between VB's */

proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     title 'New Onset: Correlation between VB families ';
run;

proc corr data = library.vb_subject;
     where ags in(3 4);
     var vb_1 -- vb_20;
     title 'FDRs: Correlation between VB families ';
run;
```

```
/*5.Correlation between age and VB's*/

proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     with age;
     title 'New Onset: Correlation between age and VB families ';
run;

proc corr data = library.vb_subject;
     where ags in(3 4);
     var vb_1 -- vb_20;
     with age;
     title 'FDRs: Correlation between age and VB families ';
run;

/*6.Correlation between BMI and VB's*/

proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     with bmi;
     title 'New Onset: Correlation between BMI and VB families ';
run;

/*7.Correlation between BMI_3months and VB's*/

proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     with bmi_3months;
     title 'New Onset: Correlation between BMI_3months and VB families ';
run;

/*8.Correlation between BMIZ and VB's*/

proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     with bmiz;
     title 'New Onset: Correlation between BMIZ and VB families ';
run;

/*9.Correlation between BMIZ_3months and VB's*/
proc corr data = library.vb_subject;
     where ags in(1 2 5);
     var vb_1 -- vb_20;
     with bmiz_3months;
     title 'New Onset: Correlation between BMIZ_3months and VB families ';
run;

/*10.VB expression by gender*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     class ags gender;
     var vb_1 --vb_20;
     title "VB Expression by gender";
```

```
     format ags lags_grouped.;
run;


/*11.VB expression by race*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where race in(1 2);
     class ags race;
     var vb_1 --vb_20;
     title "VB Expression by race";
     format ags lags_grouped.;
run;

/*12.VB expression by Positive Antibody: GAD*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where ags in(1 2 5);
     class pos_antibodies_gad;
     var vb_1 --vb_20;
     title "New onset: VB Expression by Positive Antibody: GAD";
run;

/*13.VB expression by Positive Antibody: IA2*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where ags in(1 2 5) ;
     class pos_antibodies_ia2;
     var vb_1 --vb_20;
     title "New onset: VB Expression by Positive Antibody: IA2";
run;

/*14.VB expression by Positive Antibody: IAA*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where ags in(1 2 5) ;
     class pos_antibodies_iaa;
     var vb_1 --vb_20;
     title "New onset: VB Expression by Positive Antibody: IAA";
run;

/*15.VB expression by Positive Antibody: ICA*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where ags in(1 2 5) ;
     class pos_antibodies_ica;
     var vb_1 --vb_20;
     title "New onset: VB Expression by Positive Antibody: ICA";
run;

/*16.VB expression by any Positive Antibody*/

proc means data = library.vb_subject nonobs n mean std min max range ;
     where ags in(1 2 5) ;
     class pos_antibodies;
     var vb_1 --vb_20;
     title "New onset: VB Expression by Any Positive Antibody: ICA";
```

```
run;

/*17.Comparison*/

/*a.VB by race*/

title 'VB expression by race';

/*for new onset*/
proc ttest data = library.vb_subject;
     where race in(1 2) and ags in(1 2 5) ;
     class race;
     var vb_1 —- vb_20;
run;

/*for FDR*/
proc ttest data = library.vb_subject;
     where race in(1 2)and ags in(3 4);
     class race;
     var vb_1 -- vb_20;
run;


/*b.VB by gender*/

title 'VB expression by gender';

/*for FDR*/
proc ttest data = library.vb_subject;
     where ags in(3 4);
     class gender;
     var vb_1 --vb_20;
run;

/*c.VB expression by antibody GAD*/

title 'VB expression by antibody GAD';

/*for New Onset*/
proc ttest data = library.vb_subject;
     where ags in(1 2 5);
     class pos_antibodies_gad;
     var vb_1 -- vb_20;
run;

/*d.VB expression by antibody IA2*/

title 'VB expression by antibody IA2';

/*for New Onset*/
proc ttest data = library.vb_subject;
     where ags in(1 2 5);
     class pos_antibodies_ia2;
     var vb_1 -- vb_20;
run;
```

```
/*e.VB expression by antibody IAA*/

title 'VB expression by antibody IAA';

/*for New Onset*/
proc ttest data = library.vb_subject;
     where ags in(1 2 5);
     class pos_antibodies_iaa;
     var vb_1 -- vb_20;
run;


/*f.VB expression by antibody ICA*/
title 'VB expression by antibody ICA';

/*for New Onset*/
proc ttest data = library.vb_subject;
     where ags in(1 2 5);
     class pos_antibodies_ica;
     var vb_1 -- vb_20;
run;

/*g.VB expression by the number of posotive antibodies*/

title 'VB expression by Number of posotive antibodies';

/*for New Onset*/

proc glm data = library.vb_subject;
      where ags in(1 2 5);
      class pos_antibodies;
      model vb_20 = pos_antibodies;
      means pos_antibodies;
run;



/*Analysis B*/

/* HLA and VB expressions (using the first sample)*/

* 1. VB expression by ASP status;
proc means data = library.vb_subject n mean std median min max fw = 4 maxdec
      = 2;
     class ags nonasp_status;
     var vb_1 -- vb_20;
     format ags lags_grouped.;
     title 'Mean VB Expression by ASP Status';
run;


* 2. VB expressions by non-ASP Haplotype;
proc means data = library.vb_subject n mean std median min max fw = 4 maxdec
       = 2;
     where calc_nonasp in(0 1 2 3);
     class ags calc_nonasp;
     var vb_1 -- vb_20;
```

```
     format ags lags_grouped.;
     title 'Mean VB Expressions by non-ASP Haplotype';
run;

proc gchart data = library.vb_subject;
     where calc_nonasp in(0 1 2 3);
     vbar calc_nonasp / sumvar = vb_1 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 1 Expression by ASP Haplotype';
run;

proc gchart data = library.vb_subject;
     where calc_nonasp in(0 1 2 3);
     vbar calc_nonasp / sumvar = vb_5_1 discrete
            type = mean mean nozero
            group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 5.1 Expression by ASP Haplotype';
run;

proc gchart data = library.vb_subject;
     where calc_nonasp in(0 1 2 3);
     vbar calc_nonasp / sumvar = vb_7 discrete
            type = mean mean nozero
            group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB7 Expression by ASP Haplotype';
run;

proc gchart data = library.vb_subject;
     where calc_nonasp in(0 1 2 3);
     vbar calc_nonasp / sumvar = vb_13_1 discrete
            type = mean mean nozero
            group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 13.1 Expression by ASP Haplotype';
run;

proc gchart data = library.vb_subject;
     where calc_nonasp in(0 1 2 3);
     vbar calc_nonasp / sumvar = vb_20 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Means of VB 20 Expression by ASP Haplotype';
run;


* 3. VB Expressions by Number of DQ2 Alleles;
proc means data = library.vb_subject n mean std median min max fw = 4
      maxdec = 2;
     class ags calc_dq2;
     var vb_1 -- vb_20;
     format ags lags_grouped.;
     title 'Mean VB Expressions by Number of DQ2 Alleles';
```

```
run;

proc gchart data = library.vb_subject;
     vbar calc_dq2 / sumvar = vb_1 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 1 Expression by Number of DQ2 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq2 / sumvar = vb_5_1 discrete
                 type = mean mean nozero
                 group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 5.1 Expression by Number of DQ2 Alleles';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq2 / sumvar = vb_7 discrete
            type = mean mean nozero
            group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 7 Expression by Number of DQ2 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq2 / sumvar = vb_13_1 discrete
                 type = mean mean nozero
                 group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 13.1 Expression by Number of DQ2 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq2 / sumvar = vb_20 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
      format ags lags_grouped.;
      title 'Mean VB 20 Expression by Number of DQ2 Alleles ';
run;

* 4. VB Expressions by Number of DQ8 Alleles;
proc means data = library.vb_subject n mean std median min max fw = 4
      maxdec = 2 ;
     class ags calc_dq8;
     var vb_1 -- vb_20;
     format ags lags_grouped.;
     title 'Mean VB Expression by Number of DQ8 Alleles';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq8 / sumvar = vb_1 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 1 Expression by Number of DQ8 Alleles ';
```

```
run;

proc gchart data = library.vb_subject;
     vbar calc_dq8 / sumvar = vb_5_1 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 5.1 Expression by Number of DQ8 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq8 / sumvar = vb_7 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 7 Expression by Number of DQ8 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq8 / sumvar = vb_13_1 discrete
           type = mean mean nozero
           group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB 13.1 Expression by Number of DQ8 Alleles ';
run;

proc gchart data = library.vb_subject;
     vbar calc_dq8 / sumvar = vb_20 discrete
                 type = mean mean nozero
                 group = ags gspace = 8 width = 4;
     format ags lags_grouped.;
     title 'Mean VB20 Expression by Number of DQ8 Alleles';
run;


/*Comparison*/

/*1.VB expression by non-asp alleles*/

title 'VB expression by non-asp alleles';

/*for New Onset*/
proc glm data = library.vb_subject;
     where ags in(1 2 5) and calc_nonasp in(0 1 2 3);
     class calc_nonasp;
     model vb_5_1 = calc_nonasp;
     means calc_nonasp /duncan;
run;

/*for FDR*/
proc glm data = library.vb_subject;
     where ags in(3 4) and calc_nonasp in(0 1 2 3);
     class calc_nonasp;
     model vb_20 = calc_nonasp;
     means calc_nonasp /duncan;
run;
```

```
/*2.VB expression by number of DQ2 alleles*/

title 'VB expression by number of DQ2 alleles';

/*for New Onset*/
proc glm data = library.vb_subject;
     where ags in(1 2 5);
     class calc_dq2;
     model vb_7 = calc_dq2;
     means calc_dq2 /duncan;
run;

/*for FDR*/
proc glm data = library.vb_subject;
     where ags in(3 4) ;
     class calc_dq2;
     model vb_13_1 = calc_dq2;
     means calc_dq2 /duncan;
run;


/*3.VB expression by number of DQ8 alleles*/

title 'VB expression by number of DQ8 alleles';

/*for New Onset*/
proc glm data = library.vb_subject;
     where ags in(1 2 5);
     class calc_dq8;
     model vb_20 = calc_dq8;
     means calc_dq8 /duncan;
run;

/*for FDR*/
proc glm data = library.vb_subject;
     where ags in(3 4) ;
     class calc_dq8;
     model vb_20 = calc_dq8;
     means calc_dq8 /duncan;
run;



/*Analysis C*/

/* Compare VB between new onset and FDRs with only ASP haplotype*/

/*step 1. create the data set with new onset and FDRs with ASP haplotype */

data library.vb_newonset_fdrasp;
     set library.vb_subject;
     if nonasp_status NE 0 and newonset = 0 then delete;
run;

proc freq data = library.vb_newonset_fdrasp;;
     tables  newonset;
     format newonset lnewonset_group.;
```

```
run ;

/*step 2. Compare VB between new onset and FDRs with ASP Alleles*/

/* Two-sample t-test */
proc ttest data = library.vb_newonset_fdrasp;
     class newonset;
     title "Two-sample t-test for VB1";
     var vb_1 -- vb_20;
run;



/*Analysis E */

/* VB expressions by Age for the subjects */

/*1. Correlations between VB families and age*/

proc corr data = library.vb_subject;
     where ags in (1 2 5);
     var vb_1 -- vb_20;
     with age;
     title ' Correlation for New Onsets between VB family and age';
run;

proc corr data = library.vb_subject;
     where ags in (3 4);
     var vb_1 -- vb_20;
     with age;
     title ' Correlation for FDRs between VB family and age';
run;

/* 2. VB expressions by age groups for new onsets*/

/* a. by 5 years age group*/
proc means data = library.vb_subjec_newonset nonobs n mean std  min max fw =
      4 maxdec = 2 ;
     class age;
     format age lage5yr_t1d.;
     var vb_1 --vb_20;
     title 'Mean VB by Age for New Onset';
run;


/*b. Comparison: VB expression by 5-year age groups */

title 'New Onset: VB expression by 5-year age groups ';

proc glm data = library.vb_subjec_newonset;
     class agegp_5;
     model vb_7 = agegp_5;
     means agegp_5 /tukey;
run;

/*c. Scatter plot For new onsets*/
proc gplot data = library.vb_subject;
```

```
  where ags in(1 2 5);
  plot (vb_1 vb_5_1 vb_7 vb_13_1 vb_20) * age;
  title 'New Onsets: Scatter Plot For VB vs. Age';
run;

/* 3. VB expressions by 5-year age group for FDRs*/

/*a. VB expressions by 5 years age group */

proc means data = library.vb_subjec_fdr nonobs n mean std min max fw = 4
     maxdec = 2 ;
     class age;
     format age lage5yr_fdr.;
     var vb_1 -- vb_20;
     title 'Mean VB by Age for FDR';
run;

/*b. Comparison: VB expression by 5-year age groups */

title 'FDR: VB expression by 5-year age groups ';

proc glm data = library.vb_subjec_fdr;
class agegp_5;
model vb_5_1 = agegp_5;
means agegp_5 /duncan;
run;


/*c. Scatter plot For FDRs*/
proc gplot data = library.vb_subject;
  where ags in(3 4);
  plot (vb_1 vb_5_1 vb_7 vb_13_1 vb_20) * age;
  title 'FDRs: Scatter Plot For VB vs. Age';
run;



/*Analysis E */

/*Logistic regression */

title 'Logistic regression';

/*model 1*/

proc logistic data = vb_subject  descending ;
    where race in(1 2) and n_day <= 7 ;
    class black(ref='0') female(ref='0') calc_dq2(ref='0') calc_dq8(ref='0')
          / param=ref  ;
    model newonset = vb_1 vb_5_1 vb_7 vb_13_1 vb_20
                     calc_dq2 calc_dq8  black female age_blooddraw /expb
                     selection = none lackfit risklimits  stb ;
run;

/*model 2*/

proc logistic data = vb_subject  descending ;
```

```
    where race in(1 2) and n_day <= 7 ;
    class  black(ref='0') female(ref='0') calc_dq2(ref='0') calc_dq8(ref='0')
           / param=ref  ;
    model newonset = vb_1 vb_5_1 vb_7 vb_13_1 vb_20
                     calc_dq2 calc_dq8  black female  age_blooddraw /expb
                     selection = forward lackfit risklimits  stb ;
run;

/*model 3*/

proc logistic data = vb_subject  descending ;
    where race in(1 2) and n_day <= 7 ;
    class  black (ref='0') female (ref='0') / param=ref  ;
    model newonset = vb_1 vb_5_1 vb_7 vb_13_1 vb_20
                     calc_dq2 calc_dq8  black female  age_blooddraw /expb
                     selection = none lackfit risklimits  stb ;
run;

/*model 4*/

proc logistic data = vb_subject  descending ;
    where race in(1 2) and n_day <= 7 ;
    class  black (ref='0') female (ref='0') / param=ref  ;
    model newonset = vb_1 vb_5_1 vb_7 vb_13_1 vb_20
                     calc_dq2 calc_dq8  black female  age_blooddraw /expb
                     selection = forward lackfit risklimits  stb ;
run;


/*Analysis F*/

/*calculate the difference in VB expression between baseline and 1st test and
the No. of days between baseline and 1st blood draw*/

/*Step 1: create the data set containing the subjects with more than two
blood draws*/

data vb_info_moretests;
     set library.vb_info ;
     by pid blooddrawdate;
     pid_first = first.pid;
     pid_last = last.pid;
     n_day = blooddrawdate - ddx;
     if pid_first = 1 and pid_last = 1 then delete;
run;

proc sort data = vb_info_moretests;
     by pid;
run;

proc rank data = vb_info_moretests out = blooddraw;
     by pid;
     var blooddrawdate;
     ranks blooddraw_no;
run;
```

```
/*Step 2: get the data set containing new onsets with their first blood
samples within 7 days of diagnosis*/

data  vb_1st ;
      set blooddraw ( keep = pid fid blooddrawdate ags vb_1 --vb_20
      blooddraw_no n_day);
      if blooddraw_no = 1 and -7 < n_day <=7 and ags in(1 2 5);
      rename blooddrawdate = blooddraw_1st
               vb_1   = vb_1_1st
               vb_5_1 = vb_5_1_1st
               vb_7   = vb_7_1st
               vb_13_1= vb_13_1_1st
               vb_20  = vb_20_1st;
run;

/*Step 3: get the data set containing the second blood samples for new
onsets*/

data  vb_2nd ;
      set blooddraw ( keep = pid fid blooddrawdate ags vb_1 --vb_20
          blooddraw_no );
      if blooddraw_no = 2 and ags in(1 2 5);
      rename blooddrawdate = blooddraw_2nd
               vb_1   = vb_1_2nd
               vb_5_1 = vb_5_1_2nd
               vb_7   = vb_7_2nd
               vb_13_1= vb_13_1_2nd
               vb_20  = vb_20_2nd;
run;


/*Step 4: calculate the difference in VB expression between 1st and 2nd
sample and the No. of days between 1st and 2nd blood draw*/

data vb_diff;
     merge vb_1st (drop = blooddraw_no) vb_2nd (drop = blooddraw_no);
      by pid;
      days = blooddraw_2nd - blooddraw_1st;
      vb_1_diff   = vb_1_2nd - vb_1_1st;
      vb_5_1_diff = vb_5_1_2nd - vb_5_1_1st;
      vb_7_diff   = vb_7_2nd - vb_7_1st;
      vb_13_1_diff= vb_13_1_2nd - vb_13_1_1st;
      vb_20_diff  = vb_20_2nd - vb_20_1st;

      if vb_1_diff = . and vb_5_1_diff = . and vb_7_diff = .
         and vb_13_1_diff = . and vb_20_diff = . then delete;

      label days          = '# of days between baseline and 2nd test'
            vb_1_diff     = 'VB1 diff. between baseline and 2nd test'
            vb_5_1_diff   = 'VB5.1 diff. between baseline and 2nd test'
            vb_7_diff     = 'VB7 diff. between baseline and 2nd test'
            vb_13_1_diff  = 'VB13.1 diff. between baseline and 2nd test'
            vb_20_diff    = 'VB20 diff. between baseline and 2nd test'
            blooddraw_1st = '1st Blood Draw Date'
            blooddraw_2nd = '2nd Blood Draw Date';

run;
```

```
proc print data = vb_diff;
      var pid n_day days vb_1_diff--vb_20_diff;
run;

/*1. Scatter plot for VB expression vs days*/

proc gplot data = vb_diff;
     where ags in(1 2 5);
     plot (vb_1_diff vb_5_1_diff vb_7_diff vb_13_1_diff vb_20_diff)* days;
     title 'New Onsets: Scatter Plot For VB_diff vs. Days';
run;

quit;

/*3. correlations between Diff. in VB families and days*/

proc corr data = vb_diff;
     where ags in(1 2 5);
     var vb_1_diff -- vb_20_diff;
     with days;
     title 'New Onsets: Correlation between VB families and Days';
run;

/* 4. Mean VB Expressions Difference between Two Tests */

proc means data = vb_diff nonobs n mean std median min max clm fw = 4
           maxdec = 2;
     var vb_1_diff -- vb_20_diff;
     title 'Mean VB Expression Difference between Two Tests';
run;
```

BIBLIOGRAPHY

1. Eisenbarth GS (1986) *Type I diabetes mellitus: a chronic autoimmune disease.* N Engl J Med 314: 1360-1368

2. American Diabetes Association. *All About Diabetes*. [cited 2008 June]; Available from: http://www.diabetes.org/about-diabetes.jsp.

3. Luppi P et al. (2000) *Restricted TVR Vβ gene expression and enterovirus infection in Type I diabetes: a pilot study* Diabetologia 43: 1484-1497

4. Clements GB, Galbraith DN, Taylor KW (1995) *Coxsackie B virus infection and onset of childhood diabetes.* Lancet 346: 221-223

5. Nairn C, Galbraith DN, Taylor KW, Clements GB (1999) *Enterovirus variants in the serum of children at the onset of Type I diabetes mellitus.* Diabet Med 16: 509-513

6. Juhela S, Hyoty H, Uibo R, Meriste SH, Uibo O, Lonnrot M, Halminen M, Simell O, Ilonen J (1999*) Comparison of enterovirus-specific cellular immunity in two populations of young children vaccinated with inactivated or live poliovirus vaccines.* Clin Exp Immunol 117: 100-105

7. Serreze DV, Ottendorfer EW, Ellis TM, Gauntt CJ, Atkinson MA (2000) *Acceleration of Type I diabetes by a coxsackievirus infection requires a preexisting critical mass of autoreactive T-cells in pancreatic islets* 49:708-711

8. Bottazzo GF, Florin-Christensen A, Doniach D (1974*) Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies.* Lancet ii: 1279-1282

9. Lipton RB, Kocova M, LaPorte RE et al. (1992) *Autoimmunity and genetics contribute to the risk of insulin-dependent diabetes mellitus in families: islet cell antibodies and HLA DQ heterodimers*. Am J Epidemiol 136: 503-512

10. Verge CF, Gianani R, Kawasaki E et al. (1996) *Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies.* Diabetes 45: 926-933

11. Brooks-Worrell BM, Starkebaum GA, Greenbaum C, Palmer JP (1996) *Peripheral blood mononuclear cells of insulin-dependent diabetic patients respond to multiple islet cell proteins*. J Immunol 157: 5668-5674

12. Conrad B, Weidmann E, Trucco G, Rudert WA, Behboo R, Ricordi C, Rodriquez-Rilo H, Finegold D, Trucco M. (1994) *Evidence for superantigen involvement in IDDM aetiology*. Nature 371:351-355.

13. Pietropaolo, M. and D. Le Roith (2001) *Pathogenesis of diabetes: our current understanding.* Clin Cornerstone **4**(2): p. 1-16.

14. Redondo M.J., Babu S., Zeidler A., Orban T., Yu L, Greenbaum C., Palmer J.P., Cuthbertson D., Eisenbarth G. S, Krischer J. P, Schatz D., (2006) *Specific human leukocyte antigen DQ influence on expression of antiislet autoantibodies and progression to type 1 diabetes*. J Clin Endocrinol Metab **91**(5): p. 1705-13.

15. Brandy Marie Smolnik *Analysis of the HLA-DQ Alleles in the Type 1 Diabetes Population and Their Unaffected Relatives* Available from: http://etd.library.pitt.edu/ETD/available/etd-04132007-163327/

16. American Diabetes Association. *Economic costs of diabetes in the U.S. in 2007.* [cited 2008 August]; Available from: http://care.diabetesjournals.org/misc/econcosts.pdf.