Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Luxi Liang Date

Polygenic score analysis of Bipolar Disorder

By

Luxi Liang
Master of Science in Public Health

Epidemiology

Jennifer Gladys Mulle
Committee Chair

Committee Member

Committee Member

Committee Member

Polygenic score analysis of Bipolar Disorder

By

Luxi Liang

B.S
Sun Yat-sen University
2014

Thesis Committee Chair: Jennifer Gladys Mulle, MHS, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology
2016

Abstract

Polygenic score analysis of Bipolar Disorder
By Luxi Liang

Bipolar Disorder (BD) is a life-long mental disorder with a high heritability. Although no single gene explains the disease, a great number of studies successfully identified the combined effect of small genetic factors contributing to BD. Polygenic risk score (PRS) is generated by counting and weighing these risk alleles among target individuals. In this study, we examine whether PRS generated from Bipolar I Disorder (BPI) is able to predict BPI status, age at onset of psychosis and Schizophrenia (SZ) status. We first used the entire BPI and controls set to identify associated SNPs at three inclusion thresholds ($P<0.1$, $P<0.05$, $P< 0.001$) and used these variants to calculate PRS in the exact same population. Afterwards, we divided the dataset evenly into a training set in which we obtained the genetic risk variants and a replication set in which we generated the PRSs. Last, we tested the PRS from the BPI in SZ patients. The results showed that the PRSs were significant higher among BPI cases compared to controls, regardless of the threshold in the original dataset ($P<0.0001$). The PRS from BD risk alleles were also higher among SZ cases compared to controls at threshold $P<0.05$ ($P=0.0003$) and threshold $P<0.001$ (0.0222). Nonetheless, the PRSs did not differ significantly between BPI cases and controls in the replication sample. Likewise, PRSs were unable to predict age at onset of psychosis in both dataset.

Keyword: Bipolar Disorder, Polygenic risk score, age at onset, Schizophrenia, predict

.

Polygenic score analysis of Bipolar Disorder

By

Luxi Liang

B.S
Sun Yat-sen University
2014

Thesis Committee Chair: Jennifer Gladys Mulle, MHS, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology
2016

Acknowledgements

I would like to thank my thesis advisor Dr. Jennifer G. Mulle of the Rollins School of Public Health at Emory University for her guidance and constant patience during my master program. She provided me enormous help to start the thesis program and steered me in the right direction. Her suggestion helped me refine my study greatly.

I would also like to thank Dr. Fernando S. Goes and other researchers who provided the materials used in my thesis. The study could not have been conducted without their input.

Finally, I would like to express my gratitude to my parents, my spouse Leiling Tao for their encouragement and support during my thesis project.

Author
Luxi Liang

Table of Contents

Background **1**

Risk Factors for BD **5**

Environmental and other factors 5

Genetic factors 5

Candidate Gene Associations 6

Overlap in genetics of Bipolar Disorder and Schizophrenia **9**

Polygenic risk score **10**

METHODS **12**

Study Questions and Hypothesis **12**

Material **12**

Study Sample 12

Source of genetic data 13

Method **14**

Data cleaning 14

GWAS and risk score calculation 14

Analysis of polygenic risk scores 15

Analysis of SZ samples 16

RESULTS **18**

DISCUSSION **23**

FUTURE DIRECTIONS **26**

References **27**

TABLES **36**

FIGURES AND FIGURE LEGENDS **37**

Background

Bipolar disorder (BD) is a life-long disorder that is characterized by recurrence of episodic, extreme changing mood from mania to depression[1]. According to the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V), BD and related disorders were placed as a bridge between Schizophrenia and depressive disorders with respect to symptoms, family history and genetic factors. BD can be further divided into two major subtypes, Bipolar I disorder (BPI) and Bipolar II disorder (BPII). A wide spectrum of related disorders exists (DSM-V).

The diagnosis of Bipolar I disorder (BPI) requires at least one manic episode during which the individual presents with perpetually and abnormally elevated mood, increased energy and activity represented by inflated self-esteem, less need for sleep, flight of ideas and excess involvement in either social or sexual activities (DSM-V). The manic episode cannot be explained by other psychiatric disorders or effect of substance and should last at least 1 week (DSM-V). A major depressive episode, in which an individual suffers from one or more clinically significant symptoms of depression— such as depressed mood, loss of interest, weight loss, insomnia and thoughts of death that impair social functions— is common, but not necessary for a BPI diagnosis. Hypomanic episode, a lesser form of manic episode is also common, but not required, for the diagnosis of BPI (DSM-V). The 12-month prevalence of BPI is 0.6% in the US with a male to female prevalence ratio approximately 1.1:1. It is significantly higher among

African Americans and Whites compared to Afro-Caribbeans. Mean age of onset is around 18 years old (DSM-V).

For diagnosis of Bipolar II Disorder (BPII), a history of at least one hypomanic episode or a current hypomanic episode, a milder form of manic episode that normally does not cause impairment to social functioning, is needed in combination with recurrence of periods of major depression which last at least 2 weeks. The 12 month prevalence of BPII is estimated to be 0.8% in the US, slightly higher than BPI. The average onset age is mid-20s (DSM-V). The major aspect differentiated BPI from BPII is the presence of true mania (DSM-V).

The population suffering from BD was estimated to be 29.5 million worldwide in 2004 [2]. The lifetime prevalence of type I is approximately 0.6% and type II 0.4%, according to an eleven population-based survey [3]. Though females are at an increased risk of depression, there was no evidence for a gender difference in BD prevalence [4]. Incidence rate (IR) of BD varies in different populations, ranging from 2.2/100,000 Person-Year in rural Irish populations [5] to 12.3/100,000 PY in Finland [6]. A cohort study conducted in a practitioner research database with 800,000 patient records in the Netherlands showed the overall IR of BD was 0.7/10,000 PY and IR of BPI and BPII are 0.43/10,000 PY and 0.19/10,000 PY, respectively. No statistically significant gender differences of hypomania and mania IR were detected in the study (Male IR: 0.7/10,000; Female IR: 0.68/10,000) [7]. The mean standardized mortality ratio (SMR), comparing BD patients to the general population, as obtained from Danish populations, was 1.7. In

this study a noticeable cause of death was suicide [8]. While prevalence and incidence rate appeared to be approximately equal in both genders, females are more likely to have rapid cycling BD than males [9].

Some studies have found that early-onset BD is associated with worse prognosis and response to treatment [10]. Age of onset (AAO) could serve as an alternative characterization of BD by grouping patients into early intermediate and late-onset by using 21 and 34 years old as cutoffs [10]. Evidence showed that age of onset matters in prognosis and treatment of BD in terms of clinical courses and biological implications. Early-onset BD patients often present worse outcomes compared with their late-onset counterparts, and these outcomes include higher suicide risk (Odd Ratio: 1.407), Axis I comorbidity (OR: 1.646), substance abuse (OR: 1.468) and rapid cycling course (OR: 2.082)[11]. A further study on age of onset indicated age might be a proxy for number of mood episodes and illness duration. Early onset patients suffer from more episodes of depressive or maniac mood, which reduced their response to psychosocial treatments for depression and collaborative care and significantly prolonged their recovery time [12].

The result of assessment of quality of life (QOL) indicates that BD patients performed far below the general population and worse than other mood disorder and anxiety disorder patients in emotional or psychosocial realms, mainly due to the depressive period [13]. Based on the Self-stigma Questionnaire (SSQ) results, BD patients also suffer from self-stigmatization, which leads to diminished self-esteem and self-efficacy [14]. The disease is also associated with increased morbidity and significant

impair to social functions, including occupation, education and ability to live independently [13]. BD patients have a 15-fold increased risk of suicide, compared to the general populations (DSM-V). In addition, a high BD recurrence risk of 70% was detected in a previous study, indicating the disease's chronicity [15].

The common treatments for BD include medication and psychotherapies, based on the type and severity of disease. Medication mainly consists of mood stabilizers, antidepressants and antipsychotics. Psychotherapies, including cognitive–behavioral therapy, family-focus therapy, interpersonal social rhythm therapy and functional remediation, are alternatives. Patient may also seek help from support groups [16]. The combinations of treatments above may exert different influence on treating depression and preventing possible relapses [17]. However, BD patients still have a significantly high relapse risk of 37% within the first year and 60% in the second year, even with treatment [16].

As one of the world's 10 most disabling conditions, BD contributed to 7.3 million years of healthy life lost through times spent in states of less than full health (YDL) in males, which was 2.5% of total global YDL and 7.1 million YDL in females, which was 2.3% of total YDL [2]. BD also inflicts a financial burden on public health systems, with cost per capita ranging from $16,500 to $35,000 for direct healthcare, mental healthcare and BD-related care [18]. Health services for BD patients cost 1.5 times higher than that of the general population matched with age and sex [19].

Risk Factors for BD

Environmental and other factors

Though BD has higher prevalence in high-income countries compared to low-income countries [20], the incidence rate of BD is higher in deprived areas compared to non-deprived areas. Early adulthood (age 15-24) is the peak age of onset [7]. Studies have found the onset and recurrence of BD could be triggered by stressful and negative events. Social support from family and friends is another environmental factor that impact patients' onset and recovery. BD individuals who have poor social support are more likely to relapse and need more time to recover [21]. Cognitive style, including person's self worth, self-schemata and attitude, contributes to BD, based on the evidence that BD patients exhibit a considerable degree of negative cognitive style. Parenting and maltreatment during childhood might also be associated with BD, though studies have inconsistent conclusions on these findings [21]. As for clinical risk factors, anxiety and behavioral disorders serve as precursors of BD in some studies[22].

Genetic factors

Besides the factors above, genetic factors greatly affect predisposition to major psychiatric disorders, including BD, according to findings from family and twin studies. The heritability of BD is estimated to be approximately 70% [23]. The lifetime risk of BD for a monozygotic co-twin of BD proband is 40-70% and for first-degree relative lifetime risk is 5-10%, suggesting that genetic factors affect BD susceptibility [1]. Segregation analysis, attempting to identify a mode of BD inheritance from family data, could not account for Mendelian transmission of BD [24]. However, a study conducted

among Old Order Amish families detected an autosomal dominant inheritance model that could explain the transmission of BPI in very closely related families and a polygenic effect contributing to BD among distantly related families [25]. Linkage studies showed certain chromosomes— 4p, 4q, 10p, 12q, 18q, 21q and Xq— were associated with BD in extended families. Additionally, some chromosomal abnormality, including missing and extra portion of DNA on chromosomes, or translocation were also associated with development of psychiatric disorders, such as BD and Schizophrenia (SZ) in a few families [26]. A recent study among Danish populations revealed a strong association between family history and occurrence of BD by detecting a 13.63 (95% CI: 11.81 – 15.71) fold increased risk of BD for a person who has a first-degree relative diagnosed with BD [27]. Consistent results found in another European population indicated that 76.9% of BD patients had at least one first-degree relative who had psychiatric disorders and 25.6% of BD patients have at least one grandparent who had one or more psychiatric disorders [28]. These studies supported the existence of the high inheritability of BD.

Candidate Gene Associations

A case-only study in South Africa demonstrated BD patients who possess the G allele at the SNP rs6465084 (*GRM3*) have a 4 times greater risk of developing psychosis, and the interaction between rs701567 (*DAOA*) and rs1019385 (*GRIN2B*) influenced the number of hospitalizations for mania, supporting the previous finding that genetic disposition affects the course and severity of BD [29]. As an important feature of BD, recurrence is also found to be associated with genetic factors, demonstrated by an Italian case-only study in which rs4680 genotype (*COMT*) was detected to significantly

influence the recurrence of mania (P=0.0395) (Recurrence risk ratio: Val/Val 0.0709; Val/Met 0.0739; Met/Met 0.0446) [30]. Additionally, another study using the Italian population showed that rs10861688 was significantly related to the number of depressive events (P=0.048) with higher relapse in AA subjects (4.46 times) compared to AG (3.08 times) and GG (2.65 times) [31]. Although the possible biological mechanism of these finding are not yet known, further studies on the relationship between genetic factors and BD severity would be of great interest in terms of possible screening to early detect and target high risk BD populations.

In the 1980s, linkage studies on BD were conducted after a number of large pedigrees with high incidence of BD were found. No evidence supported the hypothesis of single major locus transmission (autosomal and X chromosome) of BD in either segregation studies or in linkage studies of 21 autosomal markers [24]. The first linkage studies involved Old Order Amish families in large multi-generational pedigrees. There was no evidence found of linkage between certain blood group antigen loci and serum protein with BD [32]. However, a further study conducted by the same group reported the Harvey-ras-1 (*HRAS*) and insulin (*INS*) loci on Chromosome 11 were linked to BD [33]. The latter study encouraged a series of linkage studies on the association between Chromosome 11 and BD.

Nevertheless, follow-up studies among Old Order Amish families found no significant linkage of BD to *HRAS* or *INS* in the pedigree [34], consistent with the finding of a diminished association between BD and *HRAS* and *INS* in re-analysis of an Old

Order Amish pedigree [35]. Later on, a full-genome scan study among the same population was unable to identify any significant LOD-score from 367 markers throughout the genome [36], consistent with the previously found non-significant markers on Chromosome 1 and 11 [37]. Significant linkage of BD to loci on Chromosome 18 and evidence of parent of origin effect was detected and reproduced in studies using other nuclear families [38, 39]. With the arising of linkage studies, a combined analysis on 11 linkage studies, which had a stronger statistical power, reported loci on chromosome 6 and 8 have genome-wide significance as well as loci on 9 and 20 met potential linkage pattern [40]. Nevertheless, linkage analysis did not reveal any consistent associations in BD.

Emerging genome-wide association studies (GWAS) have enabled substantial and replicable findings in genetic factors of BD [1]. The research of the genetics behind BD has evolved from the search of Mendelian disease genes to population-based studies detecting common risk variance to GWAS [1]. GWAS is a large-scale example of linkage disequilibrium (LD) mapping [41] in which markers in genomes or sets of DNA of individuals are scanned in order to detect association between diseases and genetic variations, most of which are single nucleotide polymorphisms (SNPs). GWAS is able to detect DNA sequence differences that influence genetic susceptibility using millions of catalogued human DNA sequences and rapid, accurate DNA assay technologies. Large-sample GWAS found associations between common SNPs and psychiatric disorders [41]. Copy number variants (CNVs), though they received little attention in BD studies before 2010, were believed to play an important role in susceptibility of diseases [42]. A study

analyzing rare CNV burden among a BD and Schizophrenia (SZ) population, using the genome-wide association data, generated by the Wellcome Trust Case Control Consortium (WTCCC), failed to observe any significant increased burden in CNVs among BD patients[42]. However, another genome-wide copy-number scan of 788 subjects, in which there were 185 cases of BD, identified 23 *de novo* CNVs significantly associated with BD ((p = 0.009, OR = 4.8 [1.4,16.0]) [43]. The combined GWAS conducted by Psychiatric Genomics Consortium (PGC), in which 34 SNPs were tested in 4,496 BD cases, and 42,422 independent controls, identified 18 of those SNPs to be statistically significant (P<0.05) [44]. Later on, the PGC did another BD GWAS study of 16,731 samples and 46,912 replication samples and found two significant loci, *CACNA1C* and *ODZ4* [45]. A final study combining GWAS analysis of SZ and BD yielded a strong estimate of association of SNPs in *CACNA1C* and *NEK4-ITIH1-ITIH3-ITIH4* [44]. The above studies indicated that the increased sample sizes would be critical to detect potential disease-related loci and yield statistically significant results. Thus, the advent of GWAS made the searching for genetic predisposition of BD promising.

Overlap in genetics of Bipolar Disorder and Schizophrenia

Although defined as two distinct mental disorders, BD and SZ have similar epidemiologic characteristic such as distribution of onset age, chronicity and even a heritability of approximately 70% [23]. BD and SZ also share considerable neurobiological features and genetic susceptibility [46]. Previous linkage studies identified various loci within 18p11, 13q32, 22q11, 10p14 and 8p22 that may be associated with both BD and SZ [47], though further meta-analysis on the same

population found no substantial overlap in the highest-ranking regions [48]. A 440 SNPs Screen of candidate genes among a Ashkenazi Jewish population identified six genes, namely *DPYSL2, DTNBD1, G30/G72, GRID1, GRM4* and *NOS1* that indicated overlap in BPI and SZ [49]. Considering that the findings above are subjected to small sample sizes or chance, a Sweden population-based study of 9,009, 202 individuals was conducted and the results showed relatives of BD probands were at elevated risk of SZ with a genetic correlation of 0.60, consistent with previous findings of shared genetic susceptibility of these two diseases [50]. Furthermore, a GWAS study in a European population with 3,322 cases and 3,587 controls revealed thousands of small effect alleles contributing to SZ also had an impact on BD [51]. Some loci identified in SZ GWAS were also associated with BD. All the evidence above suggested these two psychiatric disorders share substantial genetic overlap [52, 53]. Recently, a Schizophrenia GWAS of 36,989 cases and 113,075 controls identified 108 statistically significant conservatively defined loci associated with SZ, of which 75% include protein-coding genes [54].

Polygenic risk score

The polygenic risk score (PRS) is the weighted sum of the risk-related alleles [55]. Polygenicity, which means several small genetic risk factors contribute to the traits instead of a single gene, is useful in studies of psychiatric diseases [1]. The most useful application of PRS is to predict the level of risk for an individual by calculating the risk alleles and weighting them based on their association to diseases [56]. For those complex traits and diseases which cannot be explained by single genes, the ensemble of insignificant markers obtained in GWAS, which were not individually linked to the

disease of interest, could explain the diseases to some extent [55]. The application of PRS in SZ and BD was proved to be successful, confirming that a polygenic component involving a large number of alleles of small effect contributed to the risk of diseases [51].

The fact that SZ PRS has predictability for BD also implies the existence of overlapping risk factors [53]. In the study consisting of 130 childhood onset Schizophrenia patients and their 103 healthy siblings, COS patients have a higher PRS based on 108 risk loci than those in the sibling group (P=0.025), suggesting higher PRS was associated with earlier age at onset [57]. A plausible mechanism for high PRS in BD patients could be the increased activation in limbic regions associated with these alleles detected by fMRI [58]. The possibility of early detection and prediction for BD allows for disease intervention, helping to reduce the burden of BD.

METHODS

Study Questions and Hypothesis

This study will analyze whether PRSs generated from the BPI samples are able to predict BPI status, age at onset of psychosis and SZ status.

The specific aims of my thesis are given below:

1). To determine whether the polygenic risk score are increased in BPI cases compared to controls. We hypothesize that PRS will be higher in BPI cases compared to controls.

2). To determine whether a higher PRS in BPI cases is associated with earlier age at onset. We expect PRS to be higher in early onset cases compared to late onset cases.

3). To determine whether there are gender differences in the PRS for BPI cases. The hypothesis is females would have increased PRS compared to males.

4). To determine whether BPI cases with psychosis have higher PRS than BPI cases without psychosis or controls. The hypothesis is BPI cases with psychosis have elevated PRS compared to both cases without psychosis and controls.

5). To determine whether the polygenic risk score are increased in SZ cases compared to controls. We hypothesize that PRS will be higher in SZ cases compared to controls.

Material

Study Sample

The Ashkenazi Jewish [59] ethnic group is an isolated population, which possesses useful characteristics for genetic studies because of the population bottleneck it underwent [60]. Ascertainment of BD cases was limited to BPI because among BD

subtypes, BPI is more severe [20] and poses a burden to public health with a prevalence of approximately 1% worldwide [13]. In the study, SZ and BPI cases were AJ descent recruited through advertisement in Jewish newspapers and a research website managed by The Johns Hopkins University School of Medicine, or through Jewish community outreach. The subjects were interviewed by trained clinicians, using a revised edition of the Diagnostic Interview for Genetic Studies [61]. Their blood samples were taken. The diagnosis was made based on DSM-IV after two supervising clinicians reviewed the interview results and other medical records. The subjects also went through a family history interview to confirm their AJ ethnicity. The controls were psychiatric diseases-free and of AJ descent [62, 63].

The dataset contained 446 BPI cases and 508 controls, among which 433 were males (45.4%) and 521 (54.6%) were female. Gender ratio was close to 1 among cases but female was slightly over-represented among controls (Female vs. Male: 56.1% vs. 43.9%). There were 600 SZ subjects included in the dataset for further analysis, among which were 401 (66.8%) males and 199 (33.2%). Demographic [33] and clinical characteristics including age at onset among BPI cases were also collected (Table 1).

Source of genetic data

Genotypes were obtained using the Affymetrix Human Genomewide SNP Array 6.0 on DNA isolated from the blood samples of cases and controls, and were adjusted for batch effect. The SNPs demonstrating correlation with a specific batch were removed at

the threshold of P<0.0001 to correct the bias results from the potential effect of different batches used for genotyping cases and controls [62].

Method

Data cleaning

Genomic data were managed using PLINK 1.07 [64]. Quality control steps were implemented, as described by [62]: Subject missing rate > 2%; Minor Allele frequency (MAF) < 1%; SNPs missing rates > 2% and HWE P-Value < 1×10^{-6} to exclude markers. Data cleaning in the BPI cases and controls resulted in removal of 1123 SNPs due to departure from the MAF threshold and 1 marker in control because of HWE. Among SZ cases, the data cleaning procedure removed 1,669 unqualified SNPs failing to meet MAF threshold and 11 SNPs that failed the missing SNPs rate. 363,104 SNPs remained in the BPI cases while 349,170 SNPs remained in the SZ cases for further analysis.

GWAS and risk score calculation

After data cleaning, association analysis comparing allele frequencies between 446 BPI cases and 508 controls was performed in PLINK 1.07 [64]. Manhattan Plots were created using the QQMAN package [65] in R. Odds ratios for all associations were transformed to natural log scale in R Studio [Version 0.99.447] [66]. To create polygenic risk scores [64] we used three p-value thresholds, in decreasing order of stringency: P-value < 0.1, P-value < 0.05 or P-value <0.001. This was done in two ways: First, we used the log transformed ORs obtained in association analysis in all 446 BPI cases and

508 controls to generate PRS in the same population by using SNP scoring function in PLINK 1.07 [64]. The PRS for a subject is the sum of the numbers of the reference alleles (0, 1 or 2) multiply by the weights we assigned to that SNP, in this case, the log transformed ORs [64], which could be a predictors of diseases of interest.

However, we note that this method is biased because we obtained the initial association results and the subsequent risk scores in the exact same population. To alleviate this bias, we then decided to split the original sample into a training sample and a replication sample. Previous study showed the greatest power in terms of testing association of score would be achieved when the original sample was approximately evenly split into two subsets [55]. Thus, we employed the survey function in SAS Software Version [9.4] to randomly sample half of the cases (223 BPI individuals) and controls (254 controls) from the entire sample as a training subset, in which we performed an association analysis to obtain ORs as described previously. The identical analysis procedures were done to obtain three score files based on different P-Value thresholds. We then applied the score files to the replication sample in PLINK 1.07 [64] using the scoring function. The score results as well as the demographic information of individuals were imported into SAS 9.4 for further analysis.

Analysis of polygenic risk scores

We conduct this in analysis in two ways. First, we analyzed the PRSs gained from the entire dataset. A descriptive analysis for variables including age at onset for Psychosis, sex and PRS based on different P-value thresholds was performed. Data were

visualized using box plots, histograms and scatter plots. Bivariate analysis between each variable (PRSs vs. Sex, PRSs vs. Age at onset, PRSs vs. Phenotype, Sex vs. Phenotype) was conducted. Association between age at onset and polygenic risk score was tested using the correlation function. To test the hypothesis that gender may affect PRSs, the association between sex and scores was tested with t-test function in in SAS software Version [9.4].

To assess the hypothesis that the PRSs are higher among BPI cases compared to controls, linear regression models were used with polygenic risk score as the outcome and phenotype, sex and their interaction terms as predictors ($PRS\ at\ P_i = \alpha + \beta_1 * Phenotype + \beta_2 * Sex + \delta_1 * (Phenotype * Sex) + \varepsilon, i = 1,2,3$). Logistic regression models were also used with phenotypes as outcome and sex, scores and their interactions terms are predictors. ($Logit\ (Phenotype) = \alpha + \beta_1 * PRS\ at\ P_i + \beta_2 * Sex + \delta_1 * (PRS * Sex) + \varepsilon, i = 1,2,3$) to test the whether PRSs could predict BPI. In order to further test the association between age at onset and scores, the linear regression models with scores as outcomes and age at onset as predictor were built among BD individuals who also had psychosis ($PRS\ at\ P_i = \alpha + \beta_1 * Age\ at\ onset + \varepsilon, i = 1,2,3$). Then, the identical analysis was performed with the newly acquired scores generated from the replication sample and their demographic information.

Analysis of SZ samples

To test the hypothesis that the PRSs generated using ORs from BPI sample are higher among SZ cases compared to controls, we also applied the score files to the SZ

individuals and generated PRS for them. Descriptive analysis and modeling were also done using the PRS generated from SZ individuals.

RESULTS

In the QC steps, all 446 BPI cases and 508 controls met the requirements. Demographic characteristics of the individuals included in the study are shown in Table 1. Three score files produced from the entire datasets in the association analysis included 37,069 SNPs (P_1 = P-value<0.1), 18,757 SNPs (P_2 = P-value<0.05) and 367 SNPs (P_3 = P-value<0.001), respectively (Figure 1). The score files produced from the training dataset contained 37,064 SNPs (P-value<0.1), 18,463 SNPs (P-value<0.05) and 337 SNPs (P-value<0.001) (Figure 2).

The subsequent scoring procedure generated the files containing scores, which were then combined into one SAS dataset with external demographics. The SAS dataset contained Family ID, Individual ID, Phenotype, Sex, Age at onset of Psychosis (Only among BD individuals who have psychosis, missing in all other individuals), PRS for that individual under P_1, PRS for that individual under P_2, PRS for that individual under P_3. Sex was recoded as 1 (female) and 0(male) comparing with the original PLINK output 2(female) and 1(male) in order to make the modeling more convenient. Phenotype was also recoded as 1 (BD individual) and 0(Control) comparing with the original PLINK output 2(BD individual) and 1(Control).

In the entire dataset:

Selecting risk alleles with P_1 as threshold, the average PRS was -0.004 (Standard Deviation: 0.0009) among controls and 0.0055 (Std Dev: 0.0009) among BPI individuals (Figure 3). Using P_2 as threshold to include risk alleles, the mean of PRS was -0.0055

(Std Dev: 0.0013) among controls and 0.0065 (Std Dev: 0.0013) among cases (Figure 4). With P_3 as threshold, the average PRS was -0.0174 (Std Dev: 0.0114) among controls and 0.0097 (Std Dev: 0.0113) among BPI patients (Figure 5). The PRSs increased as the threshold stringency decreased, regardless of phenotype. The PRSs were higher among BPI individuals compared to controls based on the highly statistically significant t-test results ($P<0.001$) regardless of the thresholds used. As the Pearson correlation statistics suggested a linear relationship between PRSs and phenotype, the linear regression models using three PRSs based on different thresholds as outcomes were built separately. After eliminating the insignificant predictors sex, the models indicated the BPI individuals had significant higher PRSs than the controls at threshold P_1 ($P<0.0001$, $R^2=0.9658$), threshold P_2 ($P<0.0001$, $R^2=0.9571$) and threshold P_3 ($P<0.0001$, $R^2= 0.5909$). The correlation test results revealed no gender difference of PRS regardless of thresholds ($P=0.1575$ when including risk alleles threshold was P_1; $P=0.1368$ when threshold was P_2; $P=0.2137$ when threshold was P_3).

In addition, in the logistic regression models we built using PRSs, sex and their interaction terms as indicators and phenotype as outcome, sex was not associated with phenotype at alleles inclusion threshold P_1 (OR=1.699, P=0.9116), P_2 (OR=1.949, P=0.879), P_3 (OR=0.846, P=0.5187). The interaction terms of sex and scores were found significant only at alleles inclusion threshold P_3 (P=0.0378) while no association were found at inclusion threshold P_1 (P=0.9845) and inclusion threshold P_2 (P=0.9386). When further testing the significance of the interaction terms of sex and PRS at threshold of P_3, likelihood ratio test yielded a none significant result (LR statistics= 4.649, $P_{(}x^2 \sim df_{1)}$

=0.97). Given the fact that sex and its interaction terms with PRSs were not associated with phenotype, they were eliminated from the models. After the elimination, PRSs were all highly significant with a P-value of 0.0029 at inclusion threshold of P_1, a P-value of 0.0026 at threshold of P_2 and a P-value less than 0.0001 at threshold of P_3 in models only contained PRSs as indicators. Among BPI patients who also have psychosis, the average age at onset of Psychosis was 23.72 (Std Dev: 7.96). The age at onset did not affect the PRSs at threshold of P_1 (P=0.9074, R^2= 0.0001), threshold of P_2 (P=0.7607, R^2 = 0.0004) or threshold P_3 (P=0.7627, R^2=0.0004) (Figure 9, 10 and 11).

In the replication dataset:

At threshold P_1 the mean PRS was 0.0014 (Std Dev: 0.0013) among controls and 0.0013 (Std Dev: 0.0013) among BPI individuals (Figure 6). Using P_2 as threshold, the average PRS in the replication dataset was 0.000985 (Std Dev: 0.002) among controls and 0.000991 (Std Dev: 0.0019) among cases (Figure 7). At P_3 threshold, the average PRS was 0.0065 (Std Dev: 0.0139) among controls and 0.0068 (Std Dev: 0.0138) among BPI patients (Figure 8). We did not detect significant differences of scores between BPI individuals and controls at threshold of P_1 (P=0.8229), threshold of P_2 (P=0.9702) and threshold of P_3 (P=0.7993), which was inconsistent with the finding in the entire dataset. The linear regression models predicting PRSs with phenotype suggested phenotype failed to be significant predictors at threshold P_1 (P=0.8229, R^2=0.0001), threshold P_2 (P=0.9702, R^2 =0) and threshold P_3 (P=0.7993, R^2 =0.0001), excluding the insignificant predictors sex and interaction terms between sex and phenotype. Likewise, sex has no correlation with PRSs at all thresholds (P_1: P=0.5325; P_2: P=0.6126; P_3: 0.7923). The

logistic regression models predicting phenotype with PRSs, sex and their interaction terms revealed none of the predictors were associated with phenotypes regardless of allele inclusion threshold. Meanwhile, the models contained only PRSs at different thresholds also yielded non-significant results (P_1: OR<0.001, P=0.8225; P_2: OR=6.083, P=0.9701; P_3: OR=5.429, P=0.7988). The average age at onset of psychosis for BPI individuals in the replication dataset was 24.13 (Std Dev: 8.52). No evidence supported the hypothesis that scores increases as the age at onset decreases (P_1: P=0.85, R^2=0.0003; P_2: P=0.2503, R^2=0.01; P_3: P=0.1281, R^2 =0.0175) (Figure 12, 13 and 14).

In SZ cases:

Applying the score files obtained from the BPI training dataset to 600 SZ cases, the average PRSs were 0.00152 (Std Dev: 0.0014) at P_1, 0.00154 (Std Dev: 0.00137) at P_2 and 0.009 (Std Dev: 0.00781) at P_3, respectively. The PRSs were significantly higher among SZ cases than controls at threshold P_2 (P= 0.0003) and P_3 (P=0.0222) at α=5 level. Sex did not influence the PRSs at all thresholds (P_1: P=0.1639; P_2: P=0.1082; P_3: P=0.9968). Removing insignificant predictors sex and interaction terms from linear models ($PRS\ at\ P_i = \alpha + \beta_1 * Phenotype + \beta_2 * Sex + \delta_1 * (Phenotype * Sex) + \varepsilon, i = 1,2,3$), the final model ($PRS\ at\ P_i = \alpha + \beta_1 * Phenotype + \varepsilon, i = 1,2,3$) significantly explained 1.51% (P=0.0003) variance due to phenotype at threshold P_2 and 0.61% variance at threshold P_3, although phenotype had no effect on PRS at P_1 (P=0.1422). The logistic models containing no interaction term ($Logit\ (Phenotype) = \alpha + \beta_1 * PRS\ at\ P_i + \beta_2 * Sex + \varepsilon, i = 1,2,3$) demonstrated both PRSs and Sex were

significant in terms of predicting SZ cases at threshold P_2 (Sex: OR=0.4045, P<0.0001; PRS: P=0.0009) and P_3 (Sex: OR=0.3955, P<0.0001; PRS: P=0.0203).

DISCUSSION

In this study, we initially conducted a GWAS in 446 BPI cases and 508 controls of AJ descent and calculated the PRSs, with descending allele inclusion thresholds: $P<0.1$, $P<0.05$ and $P<0.001$, in the same population. The results supported the hypothesis that PRSs were higher among BPI cases comparing to controls regardless of thresholds ($P<0.0001$). Nonetheless, we did not find any evidence of gender difference of PRSs. In addition, PRSs was not affected by age at onset among BPI cases with psychosis at any thresholds.

The limitation of this strategy was that the score profiles were applied to the original sample in which they were generated from. The score profiles were likely to be correlated with the phenotype in the original sample [64]. To address this issue, we then created two independent samples from the original dataset. The PRSs of the replication sample (BPI cases=223, Controls=254), using score profiles generated from the training sample (BPI cases=223, Controls=254), found no association between phenotype and PRSs at all threshold (P_1: $P=0.8229$; P_2: $P=0.9702$; $P=0.7993$). No gender differences of PRSs were detected. Likewise, the results failed to support the hypothesis that PRSs were higher among earlier onset psychosis among BPI cases at all thresholds (P_1: $P=0.85$; P_2: $P=0.2503$; P_3: $P=0.1281$). The results above demonstrated that the PRSs did not predict the BPI status and age at onset of psychosis once we corrected the bias in this sample.

We observed an increased polygenic risk burden among BPI cases in the entire sample, which we believe was biased. However, the non-significant results yielded from

the corrected method might be correlated with a small sample sizes. Despite the AJ sample we used might have higher number of common genetic variant because of homogeneity, the sample size might be inadequate to produce genome-wide significant results [62]. Previous simulation study showed that a reliable prediction of disease risk using genetic markers primarily depended on a large sample sizes being genotyped (>1,000 cases and 1,000 controls) and some even larger samples to perform replication studies in [56]. In studies where a polygenic inheritance was detected, the training samples and the replication samples were normally large. The International Schizophrenia Consortium used 3,322 SZ cases and 4,687 controls in GWAS and found polygenic components in SZ [67]. The polygenic markers used by Goes et el derived from the Psychiatric GWAS Consortium 2 schizophrenia study[62], which was a field-wide mega-analyses program. Thus, we might be able to achieve more accurate estimates of PRSs and a stronger statistical power if we use a larger sample.

When we applied the BPI score profiles to obtain PRSs in 600 SZ cases of AJ descent, we identified significantly higher PRSs among SZ cases compared to controls at the more stringent thresholds: $P<0.05$ ($P= 0.0003$) and $P<0.001$ ($P=0.0222$). PRSs along with sex significantly predicted SZ phenotype at threshold $P<0.05$(Sex: OR=0.4045, $P<0.0001$; PRS: $P=0.0009$) and $P<0.001$ (Sex: OR=0.3955, $P<0.0001$; PRS: $P=0.0203$).

The finding that the polygenic components to the risk of BPI also contributed to the risk of SZ is similar to the finding of International Schizophrenia Consortium's study, in which they identified the shared polygenic variance between BD and SZ [67]. The

evidence of overlap polygenic variance of BPI and SZ in our study was also consistent with the finding of a high genetic correlation due to sharing SNPs in BD and SZ (0.68±STD DEV 0.04) in a prior GWAS [68]. Additionally, a recent study demonstrated that PRSs derived from SZ was able to predict BD case-control status [69]. Our results not only consolidated the findings in these studies, but also confirmed the hypothesis that the PRSs generated using ORs from BPI sample are able to predict SZ status. The increasing evidence of shared genetic variance between BD and SZ encourages further research of the common traits and shared underlying causes of these two diseases. These findings might also shred light on developing treatments and therapies that are beneficial to patients suffer from these two diseases.

Although we failed to observe any association between age at onset of psychosis among BPI patients and their PRSs, the lack of data might explain the null results. Another explanation could be that the association between PRSs and psychosis among BPI individuals was not as strong as we expected, as the PRSs among BPI individual failed to predict psychosis in a prior study[70]. However, our results do suggest a decreasing stringent threshold could increase the R^2 of age at onset of psychosis explained by PRSs.

The limitation of our study was primarily represented by the limited sample sizes. The number of available training sample was small because we hoped to achieve a relatively high statistical power by splitting the entire sample evenly. Meanwhile, the limited number of BPI cases with psychosis along with the scarce age at onset data

contributed to the weak statistical power of analysis on the association between age at onset and PRSs. Additionally, we did not control for any ancestry based covariates in the models, which might lead to bias.

The strength of the study is that we used a homogenous ethnic group, avoiding false-positive results caused by allele frequency differences among different population substructure [71].

Future Direction

As we speculate that the null result of the association between BPI status and PRSs detected from our study was partially caused by a small sample size, further analysis can be done using public available markers, such as the risk score training set from the PGC, or similar set of markers obtained from larger samples, to generate score files. We may find a more profound correlation between BPI status and PRSs with the said score files.

Furthermore, the common variant shared by BD and SZ identified in our study need replication study. To assess whether a shared genetic components between BD and SZ exists in other populations, more studies are required to be done.

Meanwhile, the shared genetic components between BD and SZ encourage further analysis on the effect of BD risk alleles on other psychiatric disorders, such as unipolar disorder.

References

1. Craddock, N. and P. Sklar, *Genetics of bipolar disorder.* The Lancet, 2013. **381**(9878): p. 1654-1662.
2. *THE GLOBAL BURDEN OF DISEASE 2004 UPDATE Introduction.* Global Burden of Disease: 2004 Update, 2008.
3. Merikangas, K.R., et al., *Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative.* Arch Gen Psychiatry, 2011. **68**(3): p. 241-51.
4. Gold, J.H., *Gender differences in psychiatric illness and treatments: a critical review.* J Nerv Ment Dis, 1998. **186**(12): p. 769-75.
5. Scully, P.J., et al., *First-episode schizophrenia, bipolar disorder and other psychoses in a rural Irish catchment area: incidence and gender in the Cavan-Monaghan study at 5 years.* Br J Psychiatry Suppl, 2002. **43**: p. s3-9.
6. Rasanen, P., J. Tiihonen, and H. Hakko, *The incidence and onset-age of hospitalized bipolar affective disorder in Finland.* J Affect Disord, 1998. **48**(1): p. 63-8.
7. Kroon, J.S., et al., *Incidence rates and risk factors of bipolar disorder in the general population: a population-based cohort study.* Bipolar Disord, 2013. **15**(3): p. 306-13.
8. Medici, C.R., et al., *Mortality and secular trend in the incidence of bipolar disorder.* J Affect Disord, 2015. **183**: p. 39-44.

9. Calabrese, J.R. and M.J. Woyshville, *A medication algorithm for treatment of bipolar rapid cycling?* J Clin Psychiatry, 1995. **56 Suppl 3**: p. 11-8.

10. Geoffroy, P.A., et al., *Reconsideration of bipolar disorder as a developmental disorder: importance of the time of onset.* J Physiol Paris, 2013. **107**(4): p. 278-85.

11. Carter, T.D.C., et al., *Early age at onset as a risk factor for poor outcome of bipolar disorder.* Journal of Psychiatric Research, 2003. **37**(4): p. 297-303.

12. Peters, A., et al., *Age at onset, course of illness and response to psychotherapy in bipolar disorder: results from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD).* Psychological Medicine, 2014. **44**(16): p. 3455-3467.

13. Michalak, E.E.M., Greg; Young, Allan H; Lam, Raymond W; , *Burden of Bipolar Depression*

Impact of Disorder and Medications on Quality of Life. CNS Drugs, 2008. **22**(5): p. 19.

14. Karidi, M.V., et al., *Bipolar disorder and self-stigma: A comparison with schizophrenia.* J Affect Disord, 2015. **184**: p. 209-15.

15. Morgan, V.A., P.B. Mitchell, and A.V. Jablensky, *The epidemiology of bipolar disorder: sociodemographic, disability and service utilization data from the Australian National Study of Low Prevalence (Psychotic) Disorders.* Bipolar Disord, 2005. **7**(4): p. 326-37.

16. Geddes, J.R. and D.J. Miklowitz, *Treatment of bipolar disorder.* Lancet, 2013. **381**(9878): p. 1672-82.

17. Sachs, G.S., et al., *Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD).* Biological Psychiatry, 2003. **53**(11): p. 1028-1042.

18. Kleine-Budde, K., et al., *Cost of illness for bipolar disorder: a systematic review of the economic burden.* Bipolar Disord, 2014. **16**(4): p. 337-53.

19. Kupfer, D.J., *The increasing medical burden in bipolar disorder.* JAMA, 2005. **293**(20): p. 2528-30.

20. Association, A.P., *DIAGNOSTIC AND STATISTICAL*

MANUAL OF

MENTAL DISORDERS

FIFTH EDITION. 2013.

21. Alloy, L.B., et al., *The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors.* Clin Psychol Rev, 2005. **25**(8): p. 1043-75.

22. Mortensen, P.B., *Clinical risk factors for bipolar disorders: a systematic review of prospective studies.* Arch Gen Psychiatry 2003. **60**: p. 7.

23. Berrettini, W., *Evidence for shared susceptibility in bipolar disorder and schizophrenia.* Am J Med Genet C Semin Med Genet, 2003. **123C**(1): p. 59-64.

24. Goldin, L.R., et al., *Segregation and linkage analyses in families of patients with bipolar, unipolar, and schizoaffective mood disorders.* Am J Hum Genet, 1983. **35**(2): p. 274-87.

25. Pauls, D.L., et al., *Complex segregation analyses of old order Amish families ascertained through bipolar I individuals.* Am J Med Genet, 1995. **60**(4): p. 290-7.

26. Blackwood, D.H., P.M. Visscher, and W.J. Muir, *Genetic studies of bipolar affective disorder in large families.* Br J Psychiatry, 2001. **178**(Suppl 41): p. S134-6.

27. Mortensen, P.B.P., C.B; , *Individual and Familial Risk Factors For Bipolar Affective Disorders in Denmark.* Arch Gen Psychiatry, 2003. **60**: p. 7.

28. Thesing, C.S., et al., *Childhood abuse, family history and stressors in older patients with bipolar disorder in relation to age at onset.* J Affect Disord, 2015. **184**: p. 249-55.

29. Dalvie, S., et al., *Psychosis and relapse in bipolar disorder are related to GRM3, DAOA, and GRIN2B genotype.* Afr J Psychiatry (Johannesbg), 2010. **13**(4): p. 297-301.

30. Benedetti, F., et al., *Recurrence of bipolar mania is associated with catechol-O-methyltransferase Val(108/158)Met polymorphism.* J Affect Disord, 2011. **132**(1-2): p. 293-6.

31. Drago, A., et al., *CRY1 Variations Impacts on the Depressive Relapse Rate in a Sample of Bipolar Patients.* Psychiatry Investig, 2015. **12**(1): p. 118-24.

32. Egeland, J.A. and A.M. Hostetter, *Amish Study .1. Affective-Disorders among the Amish, 1976-1980.* American Journal of Psychiatry, 1983. **140**(1): p. 56-61.

33. Egeland, J.A., et al., *Bipolar Affective-Disorders Linked to DNA Markers on Chromosome-11.* Nature, 1987. **325**(6107): p. 783-787.

34. Law, A., et al., *Genetic-Linkage Analysis of Bipolar Affective-Disorder in an Old Order Amish Pedigree.* Human Genetics, 1992. **88**(5): p. 562-568.

35. Kelsoe, J.R., et al., *Re-Evaluation of the Linkage Relationship between Chromosome-11p Loci and the Gene for Bipolar Affective-Disorder in the Old Order Amish.* Nature, 1989. **342**(6247): p. 238-243.

36. LaBuda, M.C., et al., *A follow-up report of a genome search for affective disorder predisposition loci in the old order Amish.* American Journal of Human Genetics, 1996. **59**(6): p. 1343-1362.

37. Gerhard, D.S., et al., *Initial Report of a Genome Search for the Affective-Disorder Predisposition Gene in the Old-Order Amish Pedigrees - Chromosome-1 and Chromosome-11.* American Journal of Medical Genetics, 1994. **54**(4): p. 398-404.

38. Stine, O.C., et al., *Evidence for Linkage of Bipolar Disorder to Chromosome-18 with a Parent-of-Origin Effect.* American Journal of Human Genetics, 1995. **57**(4): p. 130-130.

39. Mulle, J.G., et al., *Dense SNP association study for bipolar I disorder on chromosome 18p11 suggests two loci with excess paternal transmission.* Mol Psychiatry, 2007. **12**(4): p. 367-75.

40. McQueen, M.B., et al., *Combined analysis from eleven linkage-studies of bipolar disorder provides strong evidence for susceptibility loci on chromosomes 6q and 8q.* American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 2005. **138b**(1): p. 3-4.

41. Psychiatric, G.C.C.C., et al., *Genomewide association studies: history, rationale, and prospects for psychiatric disorders.* Am J Psychiatry, 2009. **166**(5): p. 540-56.

42. Grozeva, D., et al., *Rare Copy Number Variants A Point of Rarity in Genetic Risk for Bipolar Disorder and Schizophrenia.* Archives of General Psychiatry, 2010. **67**(4): p. 318-327.

43. Malhotra, D., et al., *High frequencies of de novo CNVs in bipolar disorder and schizophrenia.* Neuron, 2011. **72**(6): p. 951-63.

44. Consortium, P.G. *Results* 2015; Available from: http://www.med.unc.edu/pgc/results - BP.

45. Group, P.G.C.B.D.W., *Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4.* Nature Genetics, 2012. **44**(9): p. 1072-1072.

46. Dacquino, C., P. De Rossi, and G. Spalletta, *Schizophrenia and bipolar disorder: The road from similarities and clinical heterogeneity to neurobiological types.* Clin Chim Acta, 2015.

47. Berrettini, W.H., *Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia.* Biol Psychiatry, 2000. **47**(3): p. 245-51.

48. Segurado, R., et al., *Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder.* Am J Hum Genet, 2003. **73**(1): p. 49-62.

49. Fallin, M.D., et al., *Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios.* Am J Hum Genet, 2005. **77**(6): p. 918-36.

50. Lichtenstein, P.Y., Benjamin H; Bjork, Camilla; Pawitan, Yudi; Cannon, Tyrone D; Sullivan, Patrick F; Hultman, Christina M;, *Common genetic determinants of schizophrenia and bipolar*

disorder in Swedish families: a population-based study. The Lancet, 2009. **373**: p. 6.

51. International Schizophrenia, C., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-52.

52. Consortium, C.-D.G.o.t.P.G., *Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.* The Lancet, 2013. **381**(9875): p. 1371-1379.

53. Neale, B.M. and P. Sklar, *Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation.* Curr Opin Neurobiol, 2015. **30**: p. 131-8.

54. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.

55. Dudbridge, F., *Power and predictive accuracy of polygenic risk scores.* PLoS Genet, 2013. **9**(3): p. e1003348.

56. Wray, N.R., M.E. Goddard, and P.M. Visscher, *Prediction of individual genetic risk to disease from genome-wide association studies.* Genome Research, 2007. **17**(10): p. 1520-1528.

57. Ahn, K., et al., *Common polygenic variation and risk for childhood-onset schizophrenia.* Mol Psychiatry, 2014.

58. Whalley, H.C., et al., *The influence of polygenic risk for bipolar disorder on neural activation assessed using fMRI.* Transl Psychiatry, 2012. **2**: p. e130.

59. Sajatovic, M.J., Janis H; Safavi, Roknedin; , *Personal and societal construction of illness among individuals with rapid-cycling biploar disorder: A life-trajectory perspective.* The American Journal of Geriatric Psychiatry, 2008. **16**: p. 9.

60. Ostrer, H. and K. Skorecki, *The population genetics of the Jewish people.* Hum Genet, 2013. **132**(2): p. 119-27.

61. Nurnberger, J.I., Jr., et al., *Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative.* Arch Gen Psychiatry, 1994. **51**(11): p. 849-59; discussion 863-4.

62. Goes, F.S., et al., *Genome-wide association study of schizophrenia in Ashkenazi Jews.* Am J Med Genet B Neuropsychiatr Genet, 2015.

63. Fallin, M.D., et al., *Bipolar I disorder and schizophrenia: A 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios.* American Journal of Human Genetics, 2005. **77**(6): p. 918-936.

64. Purcell, S., *PLINK: a toolset for whole-genome association and population-based*

linkage analysis. American Journal of Human Genetics, 81. 2007.

65. Turner, S.D., *qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots*. 2014.

66. Team, R., *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA* 2015.

67. Purcell, S.M., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-752.

68. Lee, S.H., et al., *Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.* Nature Genetics, 2013. **45**(9): p. 984-+.

69. Schulze, T.G., et al., *Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder.* World J Biol Psychiatry, 2014. **15**(3): p. 200-8.

70. Tesli, M., et al., *Polygenic risk score and the psychosis continuum model.* Acta Psychiatr Scand, 2014. **130**(4): p. 311-7.

71. Tian, C., P.K. Gregersen, and M.F. Seldin, *Accounting for ancestry: population substructure and genome-wide association studies.* Hum Mol Genet, 2008. **17**(R2): p. R143-50.

TABLES

Table 1. Demographic characteristics of Bipolar I Disorder Subjects, Schizophrenia subjects and controls

	BPI Subjects (n=446)	SZ Subjects (n=600)	Controls (n=508)
Female, n	236 (52.9%)	199 (33.2%)	285 (56.1%)
Age at onset of psychosis, median (range)	22 (4-53)	N/A	N/A

FIGURES AND FIGURE LEGENDS

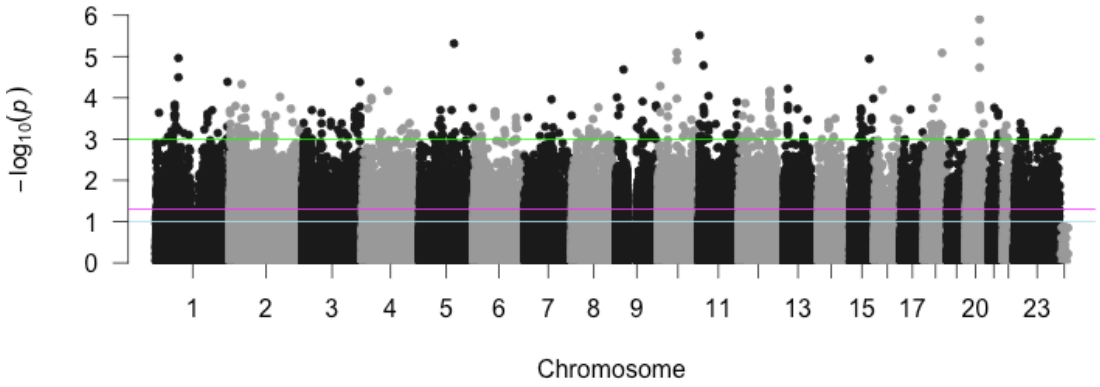


Figure 1. Manhattan plot showing three thresholds of P values of association (P<0.1, P<0.05, P<0.001)

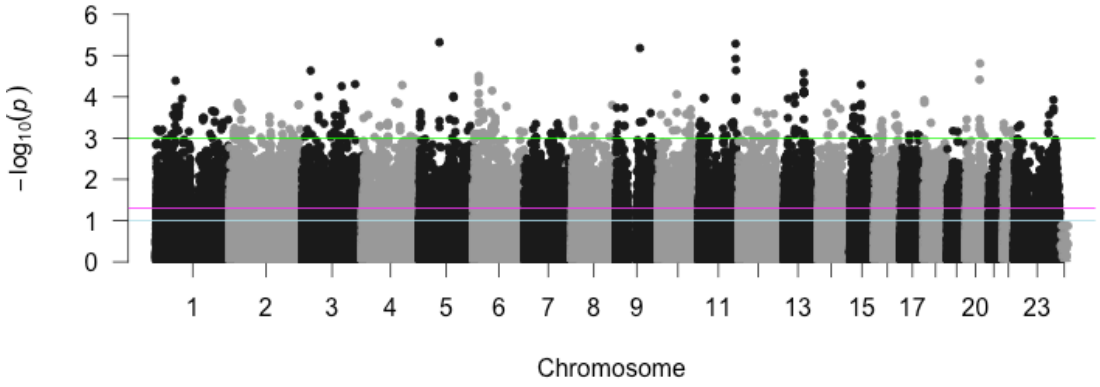


Figure 2. Manhattan plot showing three thresholds of P values of association (P<0.1, P<0.05, P<0.001) in the replication set

Figure 3. Polygenic Risk Score Distribution at Threshold P<0.1

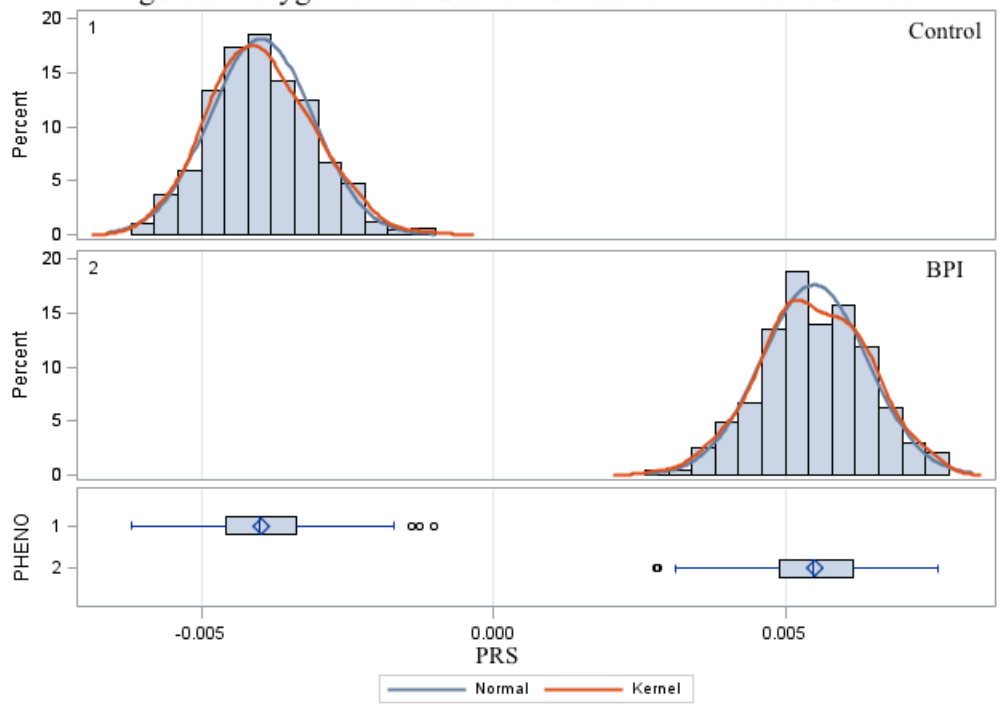


Figure 4. Polygenic Risk Score Distribution at Threshold P<0.05

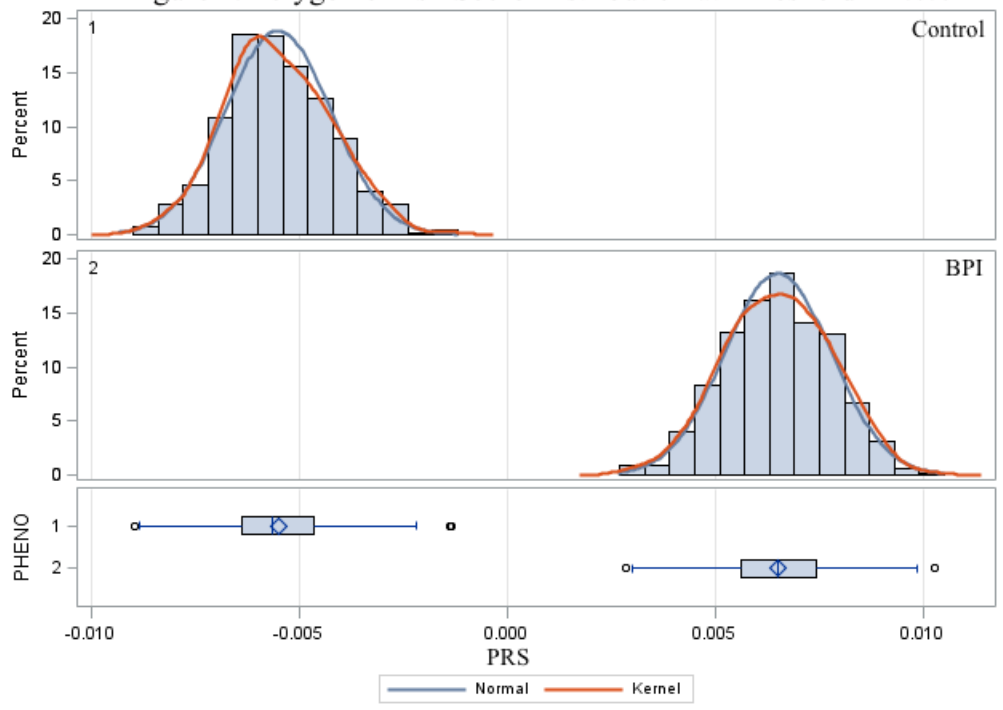


Figure 5. Polygenic Risk Score Distribution at Threshold P<0.001

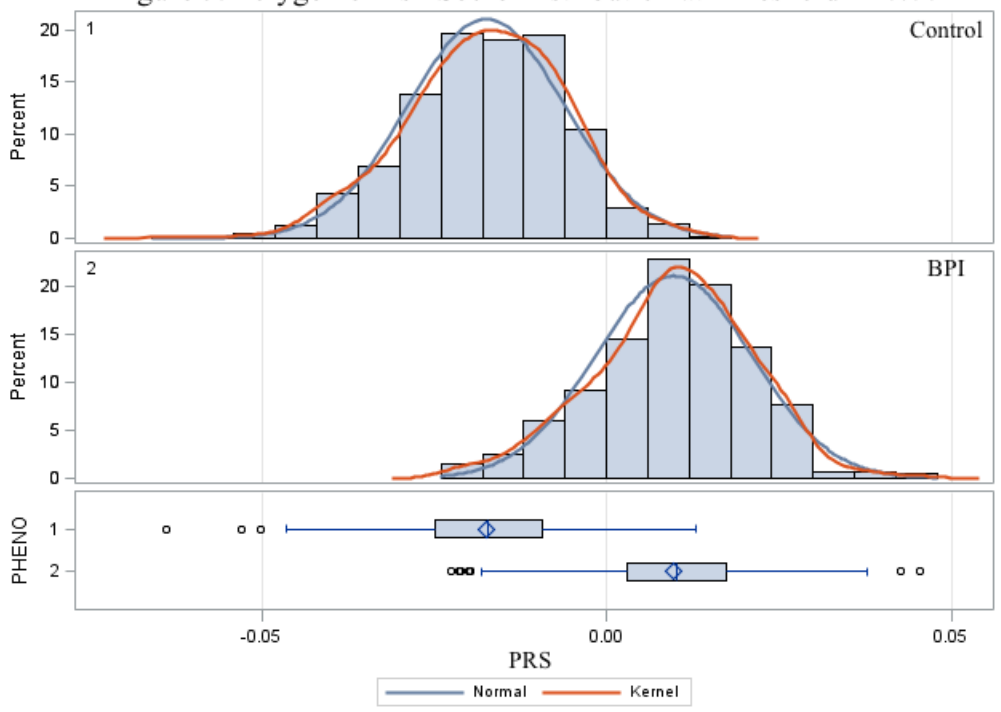


Figure 6. Polygenic Risk Score Distribution at Threshold P<0.1 in Replication Set

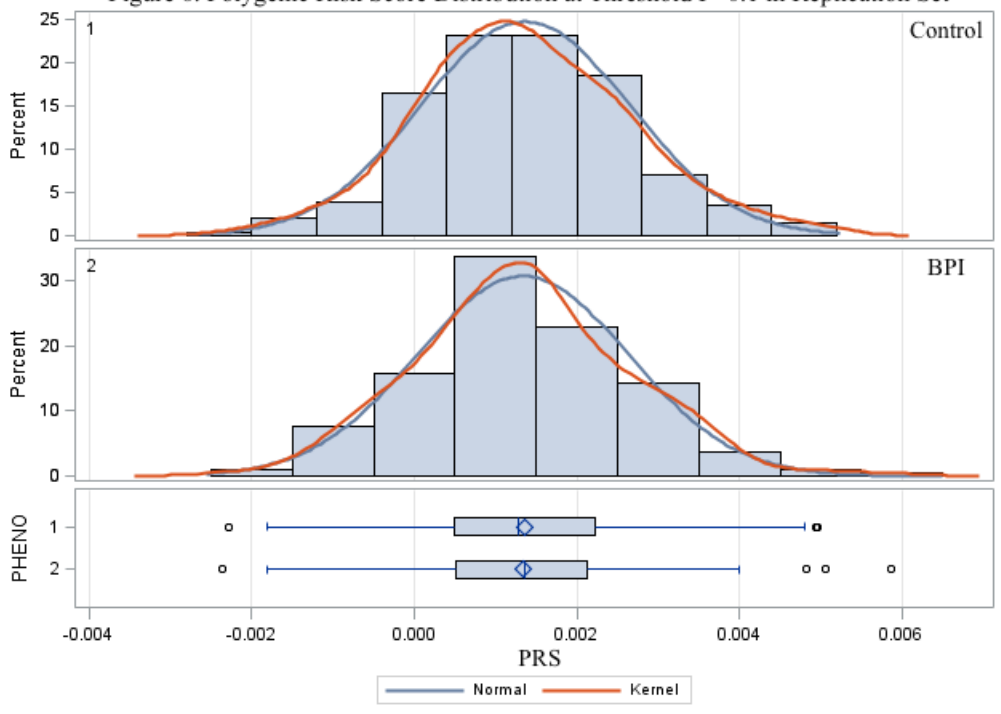


Figure 7. Polygenic Risk Score Distribution at Threshold P<0.05 in Replication Set

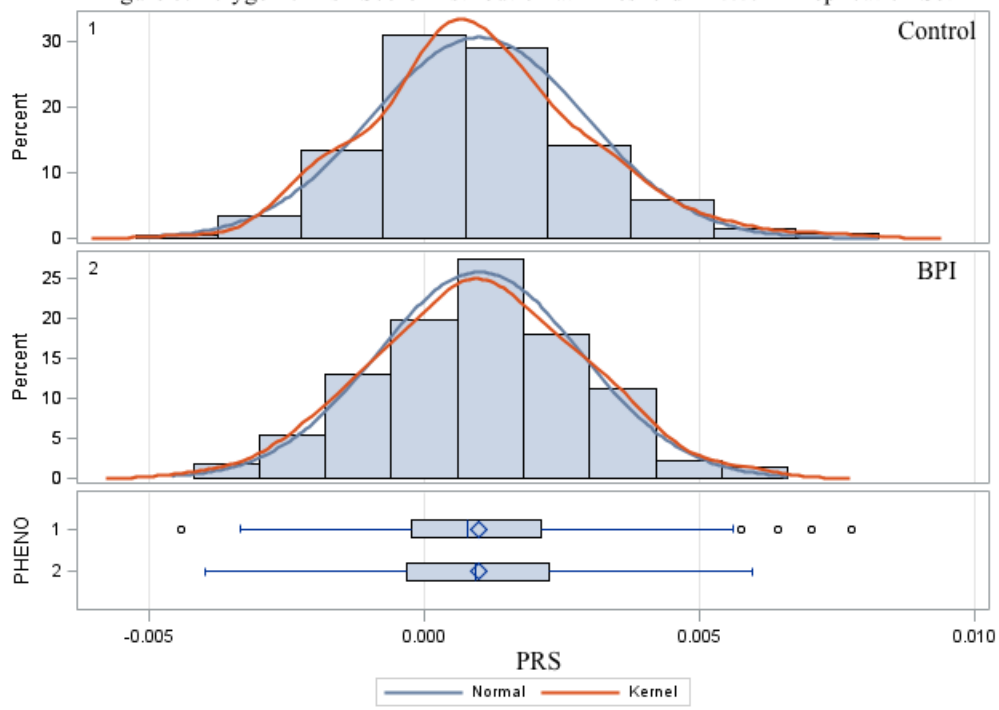
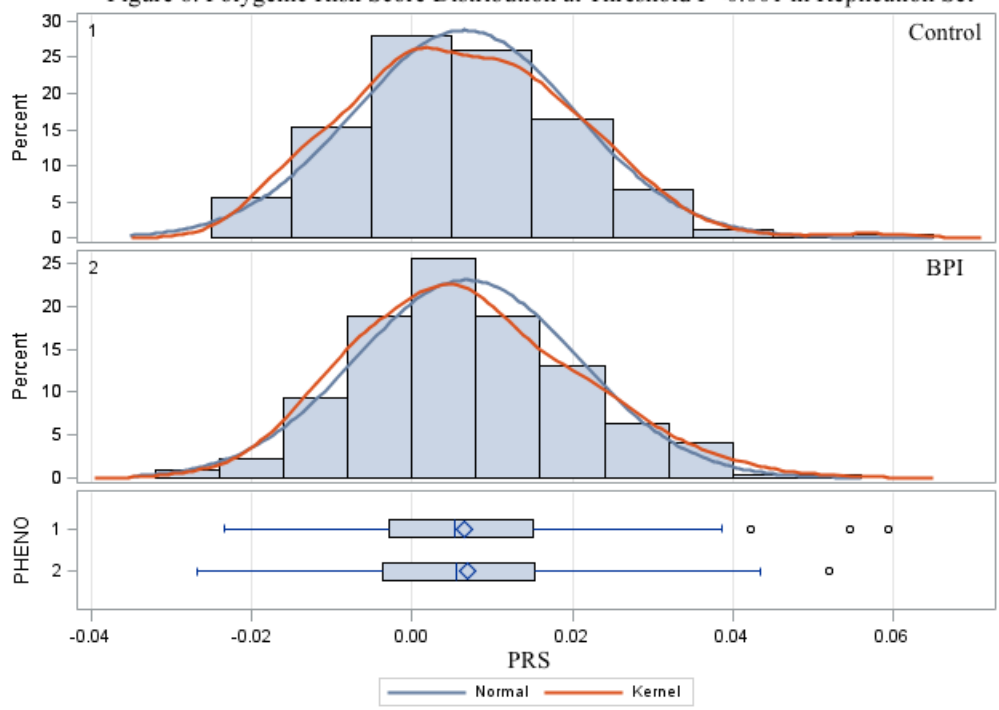


Figure 8. Polygenic Risk Score Distribution at Threshold P<0.001 in Replication Set



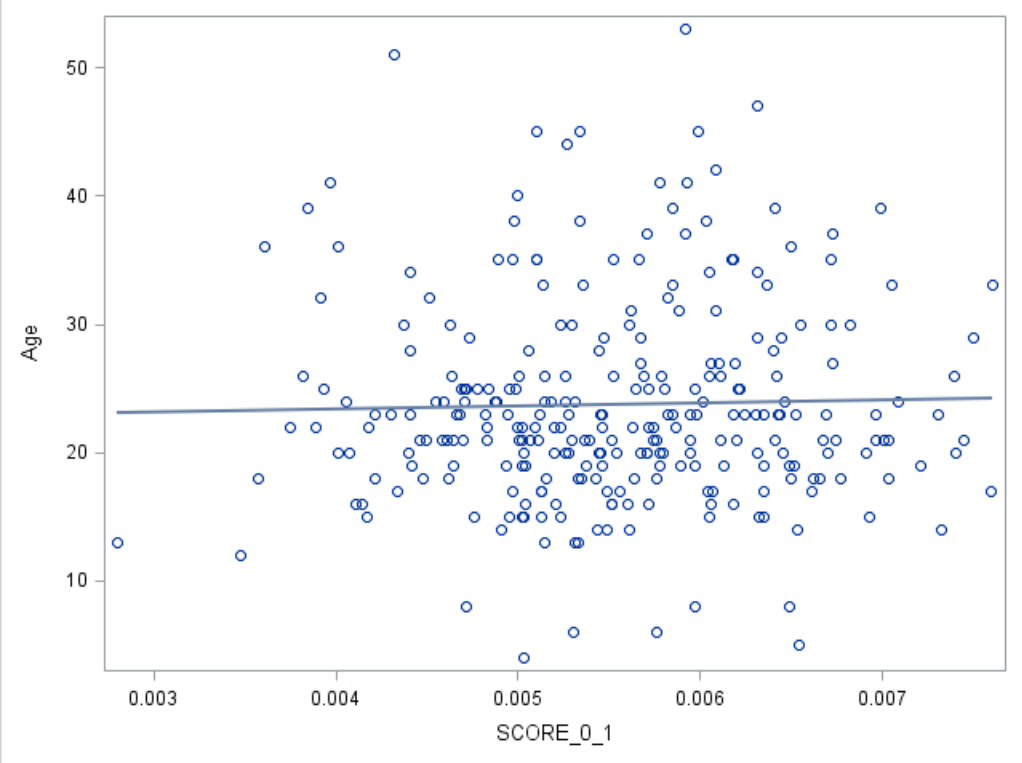


Figure 9. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.1

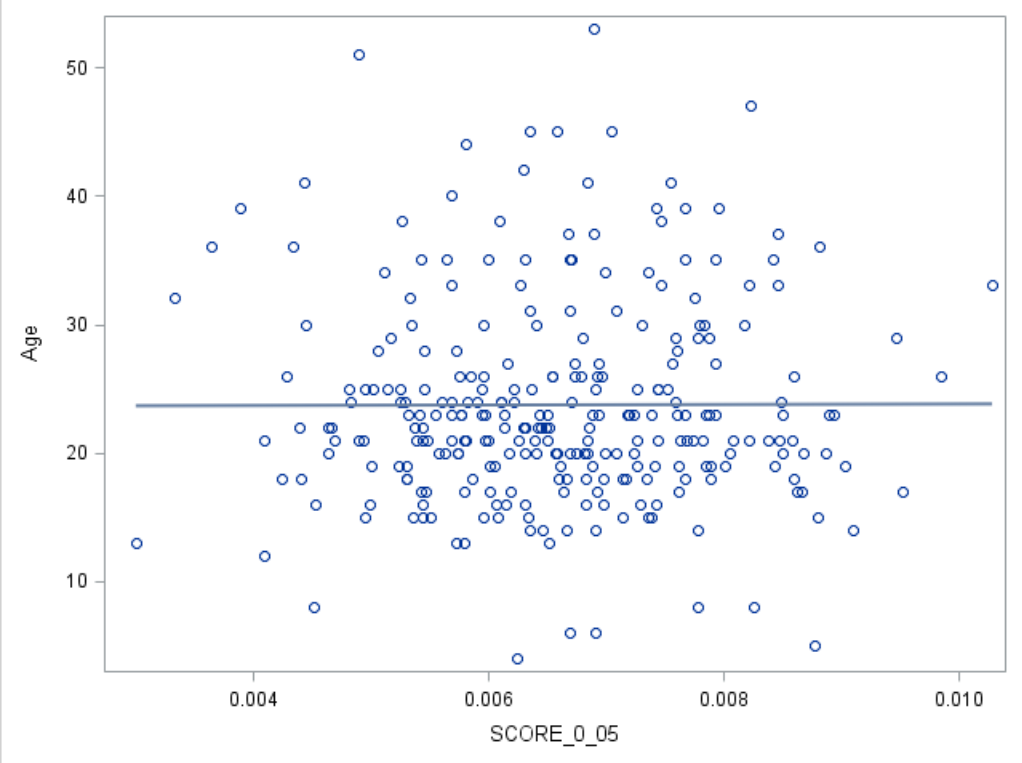


Figure 10. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.05

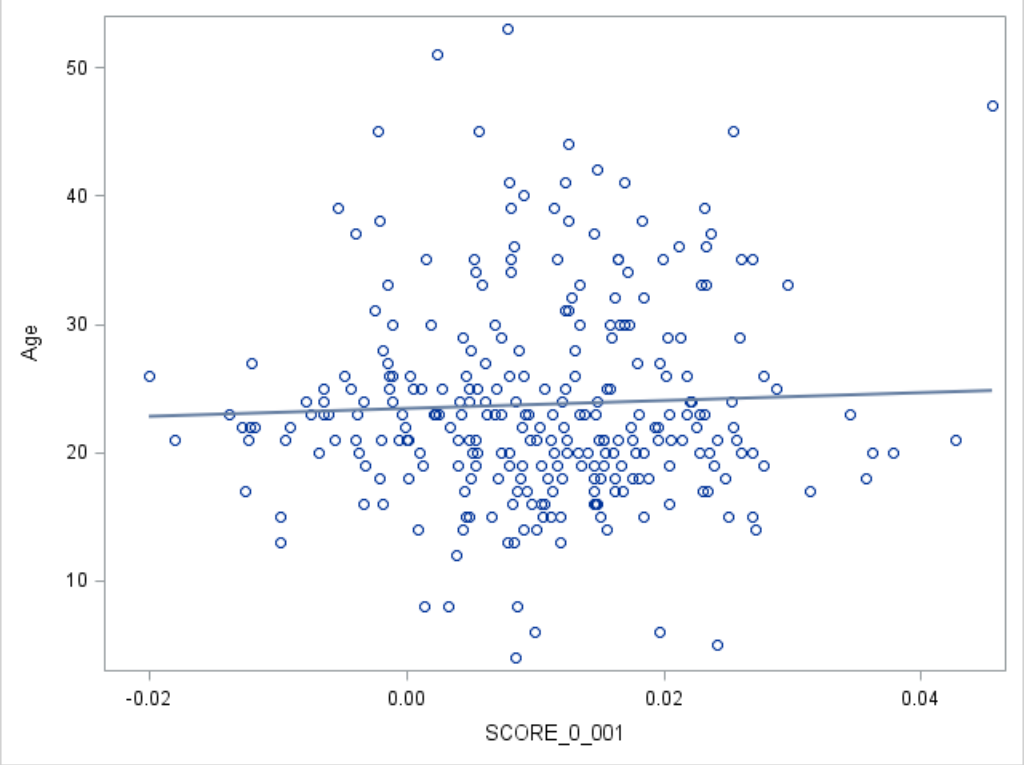


Figure 11. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.001

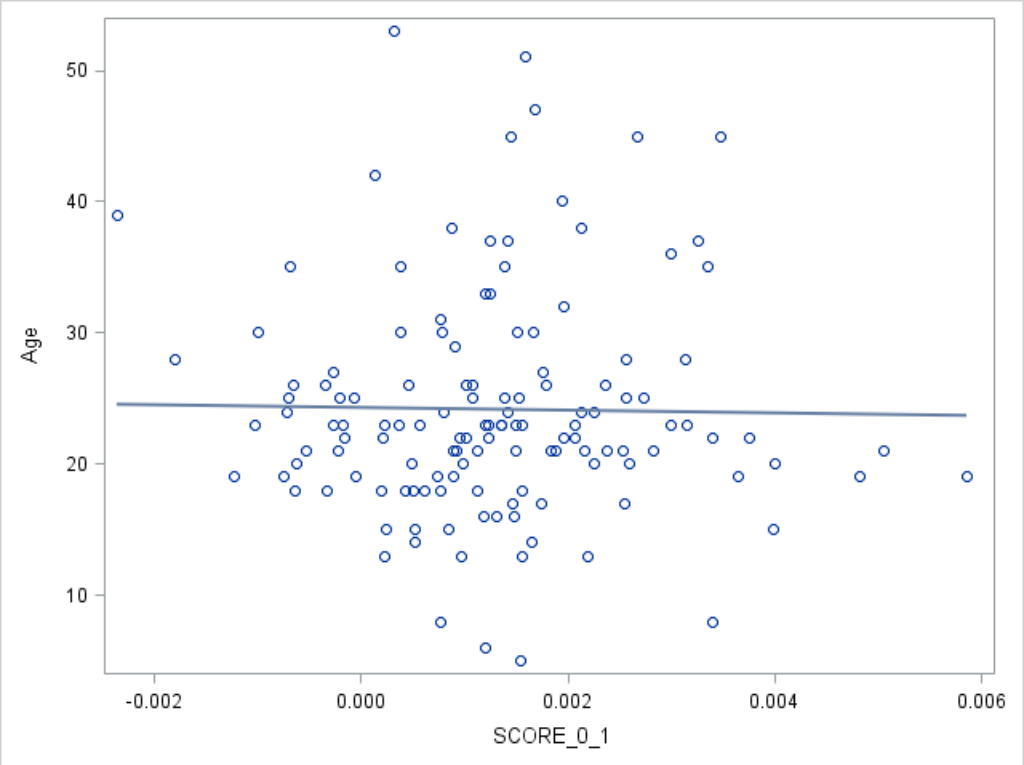


Figure 12. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.1 in the replication set

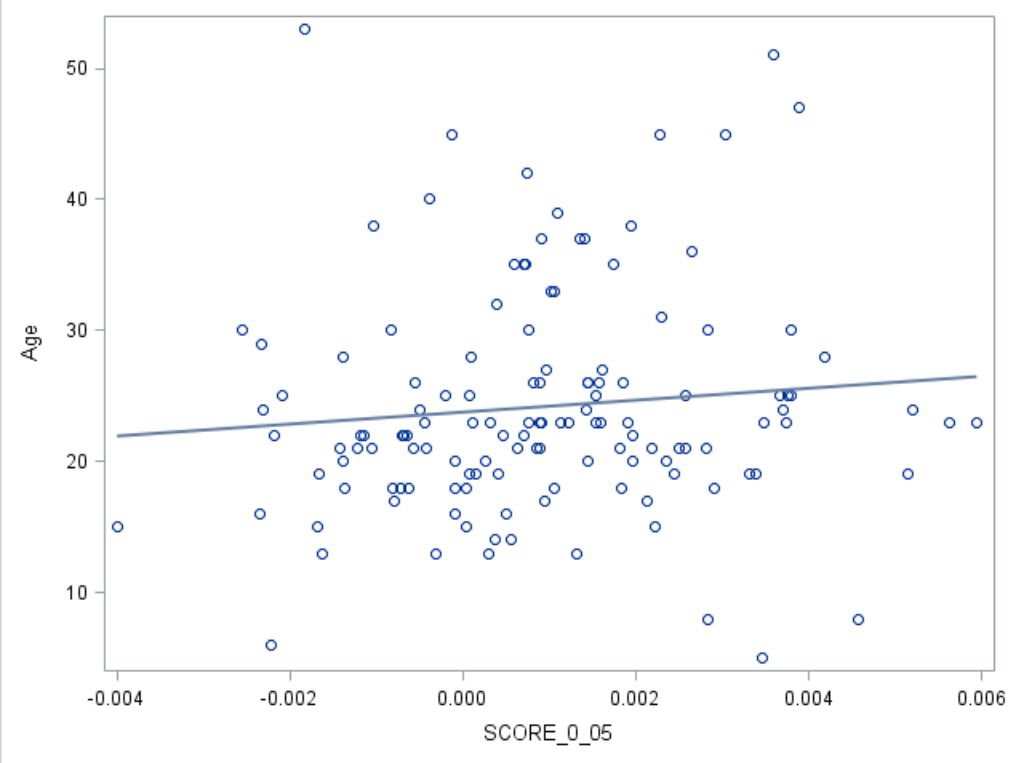


Figure 13. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.05 in the replication set

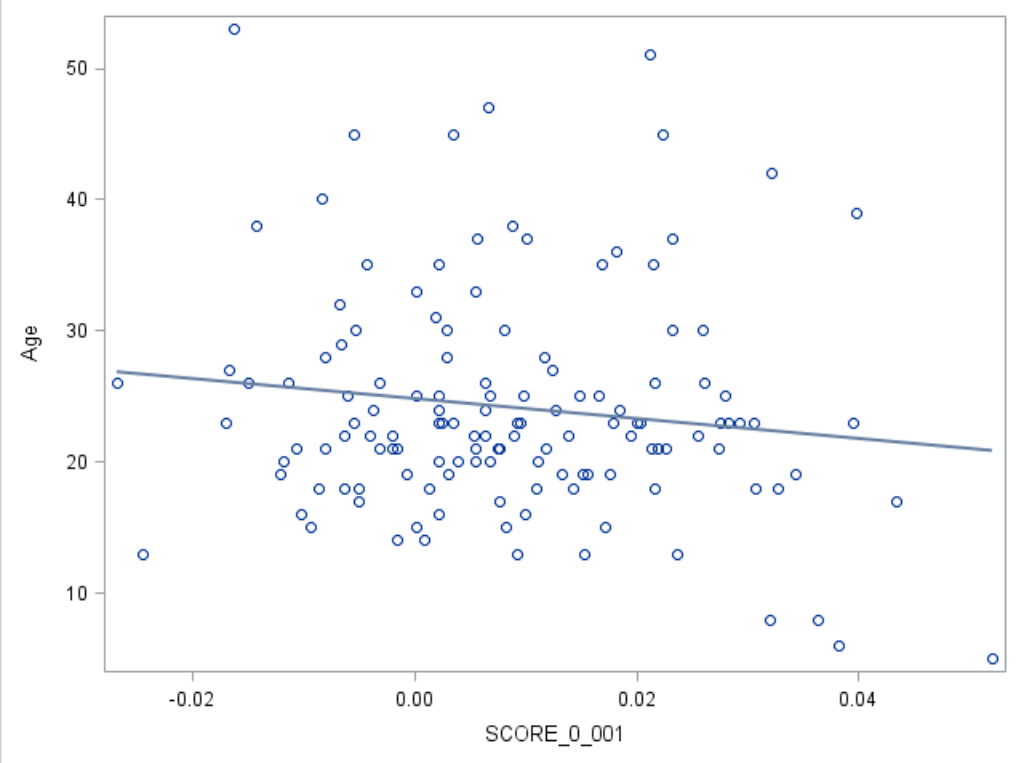


Figure 14. The association between age at onset among BPI individuals with psychosis and polygenic risk score at threshold P<0.001 in the replication set