Derivation of Poly-Methylomic Profile Scores for Schizophrenia

Watkeys, Oliver J. a, b
Cohen-Woods, Sarah c, d, e
Quidé, Yann a, b
Cairns, Murray J. f
Bronwyn Overs b
Fullerton, Janice M. b, g
Green, Melissa J. * a, b

a School of Psychiatry, University of New South Wales (UNSW Sydney), Sydney, NSW, Australia
b Neuroscience Research Australia, Sydney, NSW, Australia
c Discipline of Psychology, Flinders University, Adelaide, SA, Australia
d Flinders Centre for Innovation in Cancer, Adelaide, SA, Australia
e Centre for Neuroscience, Adelaide, SA, Australia
f School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW, Australia
g School of Medical Sciences, University of New South Wales (UNSW Sydney), Sydney, NSW, Australia

*Corresponding Author
Professor Melissa Green, UNSW Research Unit for Schizophrenia Epidemiology, Level 4, The O'Brien Centre, St Vincent's Hospital, 394-404 Victoria Street, Darlinghurst NSW 2010. Telephone: +61 02 8382 1584. Email: melissa.green@unsw.edu.au

Keywords: Epigenetics, psychosis, polygenic, bipolar, EWAS, methylation

Word count: 3,789
Number of Tables: 4
Number of Figures: 1
Supplementary Materials: 3 Tables
Poly-methylomic Profile Scores for schizophrenia

Abstract

Schizophrenia (SZ) and bipolar disorder (BD) share numerous clinical and biological features as well as environmental risk factors that may be associated with altered DNA methylation. In this study we sought to construct a Poly-Methylomic Profile Score (PMPS) for SZ, representing the degree of epigenome-wide methylation according to previously published findings; we then examined its association with SZ and BD in an independent sample. DNA methylation for 57 SZ, 59 BD cases and 55 healthy controls (HCs) was quantified using the Illumina 450K methylation beadchip. We constructed five PMPSs for different p-value thresholds using summary statistics reported in a large epigenome-wide schizophrenia case-control association study, weighted by individual CpG effect sizes. All SZ PMPSs were significantly elevated in SZ cases relative to HCs, with the score calculated at the most stringent threshold accounting for the greatest amount of variance in SZ (compared to other PMPSs derived at more inclusive p-value thresholds). However, none of the PMPSs were associated with BD, or a combined cohort of BD and SZ cases relative to HCs. Results demonstrating elevated PMPSs in SZ relative to BD did not survive correction for multiple testing. PMPSs were also not associated with positive or negative symptom severity. That this SZ-derived PMPS was elevated among SZ, but not BD participants, suggests that epigenome-wide methylation patterns associated with schizophrenia may represent distinct pathophysiology that is yet to be elucidated. Whether this PMPS may be associated with neuroanatomical or other biological endophenotypes relevant to SZ and/or BD remains to be determined.
Poly-methylomic Profile Scores for schizophrenia

1. Introduction

Owing to recent advances in genomic technologies and collaboration among global consortia, schizophrenia (SZ) and bipolar disorder (BD) have been shown to share a significant proportion of genetic liability (Hamshere et al., 2011; Ivleva et al., 2010; Purcell et al., 2009). For example, the degree to which multiple sites of genetic variation contribute to risk for SZ and BD can be summarised using ‘polygenic risk scores’, calculated as the sum of alleles associated with a particular trait, weighted by their respective effect sizes (Purcell et al., 2009). While the majority of these genetic variants would individually fail to reach even nominal levels of statistical significance, collectively they may account for a significant proportion of the variance in manifestation of SZ and BD (Purcell et al., 2009). However, there remains a significant amount of unexplained variance in shared genetic and environmental risk for these disorders (Lichtenstein et al., 2009), some of which may be accounted for by epigenetic processes associated with shared environmental risk factors such as childhood trauma, cannabis use, obstetric complications (Bourque et al., 2011; Buoli et al., 2016; Cannon et al., 2002; Kraan et al., 2016; Laursen et al., 2007).

Epigenetic processes, such as DNA methylation, refer to functionally significant alterations in the genome that do not alter the nucleotide sequence (Hannon et al., 2016; Montano et al., 2016; Pries et al., 2017). DNA methylation involves the addition of a methyl-group to the carbon-5 position of a cytosine residue (Razin and Cedar, 1991): this may be associated with increased or decreased gene expression depending on the genomic region in which it occurs (Jones, 2012; Watkeys et al., 2018). Both candidate- and genome-wide studies have revealed differential methylation of various sites in SZ (Hannon et al., 2016; Montano et al., 2016; Ruzicka et al., 2015; Yoshino et al., 2016), with some studies suggesting overlapping patterns of DNA methylation in BD (Dempster et al., 2011; Sugawara...
Poly-methylomic Profile Scores for schizophrenia et al., 2018), although there have been few replications at individual CpG sites (Pries et al., 2017; Teroganova et al., 2016).

A recently published epigenome-wide association study (EWAS) (Hannon et al., 2016) of blood-derived DNA methylation patterns reported 14 differentially methylated CpG sites that were robustly associated with SZ in both a large discovery sample (comprising 353 SZ cases and 322 healthy controls) and a replication sample (comprising 414 SZ cases and 433 healthy controls), at a stringent significance level \( (p < 1 \times 10^{-7}) \). A larger number of CpGs \( (n=213) \) were associated with SZ in both the discovery and replication case-control samples at a more liberal significance threshold \( (p < 1 \times 10^{-5}) \). Hannon et al.’s gene ontology analysis suggested that genes containing CpG sites associated with SZ were implicated in inflammatory processes and brain development. Importantly, Hannon et al. (Hannon et al., 2016) reported effect sizes and \( p \)-values for all CpG sites differentially methylated in SZ for all samples (with replication \( p \)-values ranging from \( 4.42 \times 10^{-18} \) to 0.991), and included covariates (age, sex, technical batch effects, as well as methylation-derived estimates of cell composition and smoking status). A comparably sized study published in the same year reported only CpG sites significant at an FDR corrected threshold of \( p<0.2 \) in their replication sample (number CpGs = 172) (Montano et al., 2016).

In this study, we adapted methodology used to calculate PRS from GWAS summary statistics, to estimate a series of poly-methylomic profile scores (PMPSs) for SZ. The calculation of these scores was based on discovery effect sizes of individual differentially methylated CpG sites at various \( p \)-value thresholds as reported by Hannon et al. (2016). We then established the extent to which these PMPSs were associated with an independent sample of cases with SZ, BD, and the combined SZ-BD group (each compared to healthy controls; HC), and directly compared PMPSs between SZ and BD. We also explored
Poly-methylomic Profile Scores for schizophrenia

associations with positive and negative symptoms of psychosis for the PMPS that was most strongly associated with SZ.

2. Experimental/Materials and methods.

Study procedures were approved by the UNSW Human Research Ethics committees (HC12384), St. Vincent’s Hospital (HREC/10/SVH/9), and the South East Sydney and Illawarra Area Health Service (HREC 09/081).

2.1. Participants

Peripheral blood samples were obtained from a total of 191 study participants (64 SZ, 64 BD, 60 HC). Following the application of exclusion criteria and QC processes (see below) the final sample (N=171) comprised 57 people who met ICD-10 (World Health Organisation, 2004) criteria for SZ or schizoaffective disorder (referred to as SZ), 59 individuals with BD-I disorder (BD), and 55 HCs. Clinical participants were recruited from outpatient services of the South Eastern Sydney-Illawarra Area Health Service (SESIAHS), the Australian Schizophrenia Research Bank (Loughland et al., 2010), and the Sydney BD Clinic (Mitchell et al., 2009). Healthy controls were recruited from the local community. Study exclusion criteria included current neurological disorder, current substance abuse or dependence, and/or electroconvulsive treatment in the past six months. Healthy controls were also required to have no lifetime history of a DSM-IV Axis-I diagnosis (American Psychiatric Association, 1994) confirmed using the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998), and no history of psychotic disorders among first-degree biological relatives.

2.2. Cognitive and symptom measures

Estimates of premorbid and current intelligence quotients (IQ) were calculated using the Wechsler Test of Adult Reading (Wechsler, 2001), and the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), respectively. The Scales for the Assessment of Positive
Symptoms (SAPS) and Negative Symptoms (SANS) were used to assess symptom severity (Andreasen, 1984a, b).

2.3. Methylation quantification and quality control

Methylation quantification was undertaken on DNA samples from 191 participants at the Institute for Molecular Bioscience, University of Queensland (UQ), according to standard manufacturer’s instructions using the Illumina 450K BeadChip. Samples were randomised on the plate to avoid bias due to chip or position on chip. Quantification was completed using the Qubit fluorimeter, and bisulfite conversions were completed using the EZ-96 DNA Methylation kit according to the manufacturers protocol (Zymo Research, Orange, CA, USA), and assessed on the Epoch Microplate Spectrophotometer (Biotek, Winooski, VT). All chips were imaged and analysed for DNA methylation status on the Illumina iScan system using the GenomeStudio methylation module software (Illumina, San Diego, CA, USA). This produced a methylation score as a continuous $\beta$-value representing the ratio of methylated (M) probe intensity relative to total intensity of methylated and unmethylated (U) probes defined as $\beta=M/(M+U)$. Raw and total probe intensities were extracted.

Of the 191 samples, 94 had DNA derived from whole blood, 94 firstly underwent ficoll density gradient separation with DNA extracted from the cells in the ficoll/PBS/plasma layer, and 3 were derived from the isolated lymphoblast cell population after ficoll separation; each with different cellular compositions in each tissue source. Ten samples were excluded from methylation QC analyses owing to their lymphoblast origin (3 samples) or having undergone 450k hybridisation in a different batch (7 samples); 181 DNA samples (whole blood or ficoll-treated) were included in methylation QC procedures.

The R package ‘meffil’ (Min et al., 2017) was used to extract signal intensities and perform background correction and normalisation of methylation beta values for individual CpG sites. Our QC procedure involved excluding individual subjects for whom: (i) more than
10% of probes had less than three beads (0 samples), (ii) the predicted median methylation signal was in excess of three standard deviations from the regression line generated from methylated vs. unmethylated signal intensities (1 sample), (iii) a gender mismatch (0 samples) or XY outlier status was detected (2 samples), (iv) methylation signal intensities deviated excessively from mean values for control probes (4 samples), (v) more than 10% of probes had a detection $p$-value $> 0.01$ (0 samples), or (vi) a genotype mismatch with genotypes derived from Illumina psych-chip was detected (1 sample). Individual probes were excluded if (i) $>10\%$ of samples showed detection $p$-values $> 0.01$ for that probe ($n = 137$), (ii) $>10\%$ of samples demonstrated less than 3 beads at a given probe ($n = 141$), or (iii) the probe was previously demonstrated to be cross-reactive ($n = 29\,233$) (Chen et al., 2013).

After application of these QC procedures, 456,041 CpG sites remained for analysis in the 171 participants. However, in the current study we only included the 1051 independent CpG sites demonstrating a replication $p$-value $< 0.5$ in Hannon et al.’s (Hannon et al., 2016) first replication cohort.

Blood cell count estimates were derived from methylation data for Bcells, CD4T, CD8T, granulocytes, monocytes, and natural killer cells using meffil’s gse35069 profile references. Shinymethyl (Fortin et al., 2014) was used to conduct principle components analysis with the final QCed data. We found two components significantly associated with ethnicity ($p < 0.05$), which were included as covariates in all subsequent analysis.

2.4. Poly-Methylomic Profile Score (PMPS)

PMPSs were calculated utilising data from multiple CpG sites previously reported as demonstrating an association with SZ (Hannon et al., 2016), weighted on the basis of discovery sample effect sizes. Sites were included in the PMPSs if they replicated across the two cohorts tested by Hannon et al. (2016) at five different $p$-value thresholds ($1 \times 10^{-7}$, $1 \times 10^{-5}$, 0.001, 0.05, 0.5). The use of multiple $p$-value thresholds is standard in the construction of
Poly-methylomic Profile Scores for schizophrenia

polygenic risk scores, in order to identify the threshold that enables optimal prediction (Purcell et al., 2009). The five $p$-value thresholds included 14, 210, 359, 684, and 1051 CpG sites in each PMPS, respectively. The first two scores were calculated based on CpG sites which replicated at their respective $p$-value threshold across both the discovery and first replication cohort. In contrast, the final three PMPSs were calculated based on sites that evidenced at least a discovery threshold ($p < 1 \times 10^{-5}$) association with schizophrenia and replicated at each of the three thresholds listed above, respectively (Hannon et al., 2016). Each PMPS score was calculated using previously described methods (Shah et al., 2015) as the sum of weighted beta values associated with SZ, or:

$$PMPS_i = \sum B_{ij} x \Delta j$$

Where PMPS$_i$ is the polymethylomic profile score for the $i^{th}$ participant, B$_{ij}$ is the beta value for the $i^{th}$ participant at the $j^{th}$ CpG site, and $\Delta j$ is the effect size for the $j^{th}$ CpG site.

To ensure that each PMPS was not inflated by correlated methylation among CpGs, the independence of CpG sites meeting thresholds for inclusion in any PMPS was assessed via correlation analyses. In the event that two or more CpG sites within 500 base-pairs (bp) of each other demonstrated a Pearson’s correlation coefficient > 0.1, a single site demonstrating the strongest association with SZ (as indicated by replication $p$-values from Hannon et al. (2016) was retained for PMPS calculation, whilst the other correlated sites (n=10) were excluded (see Supplementary Table 1) (Huynh et al., 2013; Ong and Holbrook, 2013; Shah et al., 2015).

As smoking behaviour influences methylation (Zeilinger et al., 2013), we used the method described above to estimate a methylation smoking score, using smoking-related methylation effect sizes for CpG sites demonstrating a significant replicated association with smoking status ($p$-value $< 5 \times 10^{-5}$) reported previously (Zeilinger et al., 2013). We validated this score using smoking data collected for our clinical participants; the smoking score was
Poly-methylomic Profile Scores for schizophrenia

significantly associated with both lifetime smoking \( (B = 0.880, OR = 2.410, p = 6.49 \times 10^{-5}) \)
and Fagerstrom test for nicotine dependency (Heatherton et al., 1991) \( (B = 0.395, OR = 1.484, p = 0.004) \).

2.5. Statistical Analyses

An initial exploratory analysis investigated whether patterns of methylation for individual CpG sites reported by Hannon et al. (2016) were differentially methylated in the current SZ sample. Beta values for each CpG site were used as dependent variables and clinical status (SZ/HC) as the main explanatory variable. For this and all subsequent analyses, covariates of age, sex, tissue source (ficol or whole blood), estimated methylation-derived cell counts, two methylation-derived ethnicity principle components, and an estimated methylation smoking score were included as fixed factors, and 450k slide ID as a random factor. Individual CpG sites were only considered to be significantly associated with SZ if they evidenced a False Discovery Rate (FDR) corrected \( q \)-value < 0.05.

A series of logistic regression analyses were then performed to assess if the PMPSs were associated with: (i) SZ (relative to HC); (ii) BD (relative to HC); (iii) the combined BD and SZ group (relative to HC); and (iv) SZ (relative to BD). The PMPSs were converted to z-scores before analysis so that regression coefficients and effect sizes would be comparable between models. FDR correction was applied across the five PMPSs assessed within each group comparison. In addition, two linear regression analyses were used to assess if the optimal PMPS was associated with SAPS and SANS global subscale scores respectively.

To investigate any potential confounding effects of medication, we ran three linear regression models (one model each for BD, SZ, and the combined SZ-BD cohort), with the PMPS that explained the most variance in SZ as the dependent variable, and imipramine (antidepressant) equivalency dosage, chlorpromazine (antipsychotic) equivalency dosage, and
mood stabiliser usage included as predictors in addition to covariates used in all other analyses.

3. Results

Demographic information for included participants is summarised in Table 1.

3.1. Differential methylation analysis

Of 1051 CpG sites included for analysis in the present study, only 79 were associated with SZ in our sample at a nominal \( p \)-value threshold (\( p < 0.05 \)), with none surviving FDR correction for multiple testing (FDR < 0.05; see Supplementary Tables 2 and 3 for full results). Of the 14 CpG sites most robustly associated with SZ in both Hannon et al.’s (2016) discovery and replication cohorts (\( p < 1 \times 10^{-7} \)), only two sites (cg13803727 and cg27541604) demonstrated nominally significant associations with SZ in our cohort (\( p < 0.05 \)) (see Supplementary Table 2). There was 100% consistency at these 14 sites between the direction of hypo-/hyper-methylation reported in Hannon et al.’s (Hannon et al., 2016) discovery cohort and our own sample. The consistency in the direction of methylation between Hannon et al. (2016) discovery cohort and our own (81.42%, 74.93%, 64.91%, and 58.23%) tended to decrease at more liberal \( p \)-value thresholds (1 \( \times 10^{-5} \), 0.001, 0.05, and 0.5, respectively).

3.2. PMPS model comparisons in association with SZ

All five PMPSs were significantly associated with SZ relative to HC both before and after FDR correction (see Table 2). All models explained a similar amount of the variance in SZ (58-60%), however models incorporating PMPSs based on more liberal \( p \)-value inclusion thresholds tended to explain less of the variance (see Table 2). The PMPS that explained the greatest amount of variance in SZ diagnosis incorporated the 14 CpG sites which were those most robustly associated with SZ across both discovery and replication samples from Hannon et al. (2016) (Nagelkerke Pseudo \( R^2 = 0.595 \)). Within this model, SZ cases showed
Poly-methylomic Profile Scores for schizophrenia

heightened PMPS relative to controls ($B = 1.263$, $OR = 3.538$, 95% CI=1.227-10.197, $p = 0.018$; see Figure 1).

### 3.3. SZ-PMPS in association with BD and the cross-disorder sample

In contrast to the SZ-HC PMPS models (described above), models comparing BD and the combined “psychosis spectrum” group (SZ + BD = 116 cases) to HCs, explained considerably less of the variance (39-41% and 31-33%, respectively). Furthermore, none of the SZ PMPSs within these models were significantly associated with a diagnosis of BD (relative to HCs), nor were they significantly associated with the combined “psychosis spectrum” group relative to HCs either before or after FDR correction. Conversely, a considerable degree of the variance in BD versus SZ status was accounted for by the PMPS models (51-53%). While there was no significant difference between BD and SZ participants on the PMPS calculated at the most stringent $p$-value threshold ($p < 1 \times 10^{-7}$) ($B = 0.852$, $OR = 2.345$, 95% CI = 0.970-5.666, $p = 0.055$) all other PMPS scores were significantly elevated in SZ relative to BD. However, none of these associations withstood FDR correction (see Table 2).

### 3.4. Associations between optimal PMPS and positive and negative symptoms of psychosis

There were no relationships between the optimal PMPS and any subscales of the SAPS and SANS, respectively (see Table 3).

### 3.5. Potential effects of medication on PMPS scores

Irrespective of clinical group (SZ, BD, or cross-diagnostic “psychosis spectrum” group), all linear regression analyses indicated that medication usage was not significantly associated with the optimal PMPS (all $p > 0.05$; see Table 4).

### 4. Discussion

Here we demonstrate the construction of a Poly-Methylomic Profile Score for schizophrenia based on differentially methylated CpG sites reported in a recent EWAS in
schizophrenia (2016); we demonstrate significantly higher scores in an independent schizophrenia sample, relative to healthy controls, at both stringent and increasingly liberal thresholds of significance. Although we found little evidence for significant differences in methylation at those individual CpG sites in this small schizophrenia sample, likely due to limited power, there was 100% consistency in the direction of DNA hyper-/hypo-methylation in our schizophrenia sample among the 14 CpG sites previously reported as most strongly associated with schizophrenia (Hannon et al., 2016). Interestingly, schizophrenia PMPSs were not associated with BD or the combined psychosis group compared to healthy controls, suggesting that the SZ-derived PMPS may reflect distinct pathophysiology in SZ. Furthermore, whilst the PMPS differences between SZ and BD were not significant after FDR correction, these associations were only just above the threshold of statistical significance. Whilst it is possible that this indicates shared methylation patterns between SZ and BD, it is also possible that our study was underpowered to detect true differences in the methylation profiles between SZ and BD.

The capacity to demonstrate associations between the constructed poly-methylomic scores and schizophrenia is consistent with the idea that methylation changes in schizophrenia span multiple CpG sites of small effect throughout the genome, and that collectively, the summed effects across multiple CpG sites should explain a larger proportion of the variance in the manifestation of schizophrenia than individual CpG sites alone. Notably, the amount of variance explained by the PMPSs decreased as increasingly liberal p-value thresholds were adopted for score calculation. This stands in stark contrast to genetic polygenic risk scores, which tend to explain more of the variance when more liberal p-value thresholds are utilised (Purcell et al., 2009). This difference may reflect the smaller EWAS cohorts compared to those used to generate polygenic risk scores for schizophrenia (Purcell et al., 2009). However, additional consideration should also be given to the increased extent to
which DNA methylation is susceptible to confounding factors. That is, whilst genetic sequence variants are inherited directly from parents and remain stable across the lifespan (except in the case of de novo events), epigenetic processes are susceptible to environmental modification and change over time (Dempster et al., 2013). These factors make it inappropriate to characterise a trait-associated methylation score as indexing risk for the trait in question. Rather, any such score might be instead regarded as summarising the methylation profile associated with a given trait. Indeed, such scores have already been calculated for phenotypes such as body-mass index (BMI) and environmental exposures such as smoking status, education, and alcohol consumption (Elliott et al., 2014; Hamilton et al., 2018; McCartney et al., 2018). Our study reflects the first attempt to derive such scores from methylome-wide CpG data for established schizophrenia. Application of this profile to first-episode psychosis samples, and the further derivation of PMPSs from EWASs conducted in first-episode psychosis and other medication naive high-risk samples, will be useful to understand its potential biological relevance to disease risk and the emerging pathophysiology of schizophrenia.

Interestingly, none of the SZ-derived PMPS scores were associated with BD or the cross-disorder psychosis group relative to HCs, suggesting that the methylation patterns summarised by these scores may reflect distinct schizophrenia pathophysiology. This was somewhat unexpected given increasing evidence of biological overlap between schizophrenia and BD (Dempster et al., 2011; Ivleva et al., 2010; Laursen et al., 2007; Purcell et al., 2009; Sugawara et al., 2018). Furthermore, whilst PMPSs were not significantly elevated in SZ compared to BD after correcting for multiple testing, this could reflect lack of power. Larger discovery EWASs may be required to identify CpGs that are differentially methylated in both conditions. It would also be interesting to determine if a PMPS calculated using discovery
data from a BD cohort is associated with both BD and SZ, however large discovery EWAS datasets in bipolar disorder are currently lacking.

Several limitations should be considered when interpreting these results. First, the reference profile used for estimating cellular composition was based on a small sample of men (Reinius et al., 2012). Whilst this method is widely used, questions remain regarding the validity of this measure as a means of estimating cellular composition in the broader population. Other potential sources of variation in DNA methylation include participants’ smoking behaviours (Zeilinger et al., 2013) and body mass index (BMI) (McCartney et al., 2018). We partially addressed this by estimating a methylation-based smoking score for use as a covariate (Elliott et al., 2014), since smoking status can be accurately estimated using methylation profile scores (Elliott et al., 2014). However, a methylation-derived score of BMI was not estimated, as there is substantially less reliability where BMI estimation is concerned (McCartney et al., 2018), and phenotypic BMI data was not available for our sample.

Regarding potential medication effects, we saw no significant associations between PMPS for schizophrenia and antipsychotic and antidepressant dosages, or mood stabiliser use, but caution that further research in medication-naïve cohorts is warranted. It is also important to note that the tissue specificity inherent to DNA methylation patterns (Davies et al., 2012; Murphy et al., 2005; van den Oord et al., 2016; Walton et al., 2016; Xin et al., 2010) which are highly influenced by cellular composition (Jaffe and Irizarry, 2014), limits the generalisability of methylation patterns obtained from blood. Replicating the methods described here using discovery/replication cohorts derived from brain tissue will be important to understand the significance of methylation patterns over the course of illness.

In conclusion, we have developed a novel method for calculating a poly-methylomic profile of schizophrenia based on published EWAS data; all profile scores derived at various p-value thresholds were associated with SZ, but not with BD, or the cross-disorder sample of
Poly-methylomic Profile Scores for schizophrenia

SZ and BD cases, relative to healthy participants. The PMPS has potential utility for estimating how epigenetic patterns across multiple differentially methylated sites might be related to intermediate (endo)phenotypes of schizophrenia and related disorders, such as cognitive functioning and brain structural and functional abnormalities. Future studies in larger independent samples should be mindful of potential demographic and experimental confounders of DNA methylation, such as tissue source of DNA, cellular composition, smoking status, BMI, age and medication exposure. Finally, efforts should be made to assess the extent to which blood-based PMPSs for schizophrenia index methylation patterns in the brain, and the extent to which these markers can shed light on the pathophysiology of psychosis and related disorders.
Poly-methylomic Profile Scores for schizophrenia

References


Andreasen, N.C., 1984a. Scale for the Assessment of Negative Symptoms (SANS). University of Iowa, Iowa City, IA.

Andreasen, N.C., 1984b. Scale for the Assessment of Positive Symptoms (SAPS). University of Iowa, Iowa City, IA.


Poly-methylomic Profile Scores for schizophrenia


Poly-methylomic Profile Scores for schizophrenia


Ruzicka, W.B., Subburaju, S., Benes, F.M., 2015. Circuit- and Diagnosis-Specific DNA Methylation Changes at gamma-Aminobutyric Acid-Related Genes in Postmortem Human Hippocampus in Schizophrenia and Bipolar Disorder. JAMA psychiatry 72(6), 541-551.


Teroganova, N., Girshkin, L., Suter, C.M., Green, M.J., 2016. DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. BMC Genetics 17, 27.

Poly-methylomic Profile Scores for schizophrenia


Poly-methylomic Profile Scores for schizophrenia

Figure/Table Legends

**Figure 1**
Title: Figure 1. Group comparisons on raw PMPS score

Legend: Boxplots showing group differences on the PMPS calculated at the most stringent $p$-value threshold ($p 1 \times 10^{-7}$) between: A) schizophrenia cases and healthy controls, B) bipolar disorder cases and healthy controls, C) combined clinical cases (SZ+BD) and healthy controls, and D) bipolar disorder and schizophrenia.

**Table 1**
Title: Table 1. Sample demographics

Legend: Demographics statistics for each group (schizophrenia, bipolar disorder, and healthy controls)

**Table 2**
Title: Table 2. PMPSs and clinical groups

Legend: Results of logistic regression analyses comparing various groups on each PMPS.

**Table 3**
Title: Table 3. PMPSs and SAPS/SANS subscale scores

Legend: Results for regression models examining association between PMPS ($p < 1 \times 10^{-7}$) and SANS/SAPS global subscales.

**Table 4**
Title: Table 4. PMPSs and Medication

Legend: Results from linear regression analyses investigating relationship between PMPS ($1 \times 10^{-7}$) for SZ and medication use in cases.
Table 1. Sample demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>HC</th>
<th>BD</th>
<th>SZ</th>
<th>Group difference statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (% Male)</td>
<td>55 (49.09)</td>
<td>59 (28.81)</td>
<td>57 (61.40)</td>
<td>$X^2 = 12.661, p = 0.002$</td>
</tr>
<tr>
<td>Mean Age in years (SD)</td>
<td>36.32 (11.78)</td>
<td>37.66 (12.56)</td>
<td>43.02 (11.12)</td>
<td>$F(2,168) = 5.061, p = 0.007, \eta^2_p = 0.057$</td>
</tr>
<tr>
<td>Mean Education in years (SD)</td>
<td>16.84 (2.46)</td>
<td>15.68 (2.79)</td>
<td>14.23 (2.49)</td>
<td>$F(2,168) = 14.289, p = 1.858 \times 10^{-3}, \eta^2_p = 0.145$</td>
</tr>
<tr>
<td>Mean WTAR (SD)</td>
<td>110.62 (10.30)</td>
<td>105.02 (14.78)</td>
<td>100.21 (14.43)</td>
<td>$F(2,167) = 8.495, p = 3.066 \times 10^{-3}, \eta^2_p = 0.092$</td>
</tr>
<tr>
<td>Mean WASI (SD)</td>
<td>117.74 (12)</td>
<td>114.82 (11.39)</td>
<td>107.7 (14.077)</td>
<td>$F(2,124) = 7.121, p = 0.001, \eta^2_p = 0.103$</td>
</tr>
<tr>
<td>Mean SAPS (SD)</td>
<td>NA</td>
<td>6.75 (8.19)</td>
<td>17.75 (14.61)</td>
<td>$F(1,111) = 24.469, p = 2.706 \times 10^{-3}, \eta^2_p = 0.181$</td>
</tr>
<tr>
<td>Mean SANS (SD)</td>
<td>NA</td>
<td>15.3 (15.83)</td>
<td>29.84 (17.26)</td>
<td>$F(1,110) = 21.605, p = 9.361 \times 10^{-3}, \eta^2_p = 0.164$</td>
</tr>
<tr>
<td>Mean Age of illness onset</td>
<td>NA</td>
<td>23.42 (9.70)</td>
<td>25.04 (8.51)</td>
<td>$F(1,105) = 0.831, p = 0.364, \eta^2_p = 0.008$</td>
</tr>
<tr>
<td>Mood stabiliser usage (%; BD vs. SZ)</td>
<td>NA</td>
<td>45 (76.27)</td>
<td>12 (21.05)</td>
<td>$X^2 = 33.195, p = 8.34 \times 10^{-4}$</td>
</tr>
<tr>
<td>Mean Imiprimine-equivalent dosage (SD)</td>
<td>NA</td>
<td>26.06 (60.05)</td>
<td>63.69 (121.99)</td>
<td>$F(1,114) = 4.491, p = 0.036, \eta^2_p = 0.038$</td>
</tr>
<tr>
<td>Mean Chlorpromazine-equivalent dosage (SD)</td>
<td>NA</td>
<td>194 (571.09)</td>
<td>476.9 (886.43)</td>
<td>$F(1,114) = 4.204, p = 0.043, \eta^2_p = 0.036$</td>
</tr>
</tbody>
</table>

HC= healthy controls; BD= bipolar disorder; SZ= schizophrenia; WTAR= Wechsler Test of Adult Reading; WASI= Wechsler Abbreviated Scale of Intelligence; SAPS= Scale for Assessment of Positive Symptoms; SANS= Scale for Assessment of Negative Symptoms; SD= standard deviation
Table 2. Associations between each of the five PMPS scores and clinical groups

<table>
<thead>
<tr>
<th>Group comparison</th>
<th>PMPs p-value threshold</th>
<th># CpGs</th>
<th>B</th>
<th>OR</th>
<th>CI 95%</th>
<th>p</th>
<th>FDR q value</th>
<th>Pseudo R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1e-07 (discovery and replication)</td>
<td>14</td>
<td>1.263</td>
<td>3.538</td>
<td>1.227-10.197</td>
<td>0.018</td>
<td>0.029</td>
<td>0.5958</td>
</tr>
<tr>
<td></td>
<td>1e-05 (discovery and replication)</td>
<td>210</td>
<td>1.127</td>
<td>3.086</td>
<td>1.166-8.166</td>
<td>0.021</td>
<td>0.029</td>
<td>0.5934</td>
</tr>
<tr>
<td></td>
<td>0.001 (replication)</td>
<td>359</td>
<td>0.975</td>
<td>2.651</td>
<td>1.091-6.441</td>
<td>0.029</td>
<td>0.029</td>
<td>0.5886</td>
</tr>
<tr>
<td></td>
<td>0.05 (replication)</td>
<td>684</td>
<td>0.923</td>
<td>2.516</td>
<td>1.126-5.622</td>
<td>0.022</td>
<td>0.029</td>
<td>0.5919</td>
</tr>
<tr>
<td></td>
<td>0.5 (replication)</td>
<td>1051</td>
<td>0.803</td>
<td>2.233</td>
<td>1.080-4.617</td>
<td>0.028</td>
<td>0.029</td>
<td>0.5883</td>
</tr>
<tr>
<td>SZ vs. HC</td>
<td>1e-07 (discovery and replication)</td>
<td>14</td>
<td>-0.077</td>
<td>0.926</td>
<td>0.423-2.024</td>
<td>0.845</td>
<td>0.979</td>
<td>0.3998</td>
</tr>
<tr>
<td></td>
<td>1e-05 (discovery and replication)</td>
<td>210</td>
<td>-0.010</td>
<td>0.990</td>
<td>0.472-2.078</td>
<td>0.979</td>
<td>0.979</td>
<td>0.3995</td>
</tr>
<tr>
<td></td>
<td>0.001 (replication)</td>
<td>359</td>
<td>-0.130</td>
<td>0.878</td>
<td>0.424-1.817</td>
<td>0.722</td>
<td>0.979</td>
<td>0.4006</td>
</tr>
<tr>
<td></td>
<td>0.05 (replication)</td>
<td>684</td>
<td>-0.273</td>
<td>0.761</td>
<td>0.379-1.530</td>
<td>0.437</td>
<td>0.979</td>
<td>0.4045</td>
</tr>
<tr>
<td></td>
<td>0.5 (replication)</td>
<td>1051</td>
<td>-0.281</td>
<td>0.755</td>
<td>0.381-1.495</td>
<td>0.414</td>
<td>0.979</td>
<td>0.4051</td>
</tr>
<tr>
<td>BD vs. HC</td>
<td>1e-07 (discovery and replication)</td>
<td>14</td>
<td>0.472</td>
<td>1.603</td>
<td>0.875-2.935</td>
<td>0.123</td>
<td>0.238</td>
<td>0.3246</td>
</tr>
<tr>
<td></td>
<td>1e-05 (discovery and replication)</td>
<td>210</td>
<td>0.465</td>
<td>1.592</td>
<td>0.911-2.783</td>
<td>0.100</td>
<td>0.238</td>
<td>0.3269</td>
</tr>
<tr>
<td></td>
<td>0.001 (replication)</td>
<td>359</td>
<td>0.381</td>
<td>1.464</td>
<td>0.860-2.494</td>
<td>0.157</td>
<td>0.238</td>
<td>0.3223</td>
</tr>
<tr>
<td></td>
<td>0.05 (replication)</td>
<td>684</td>
<td>0.304</td>
<td>1.355</td>
<td>0.827-2.222</td>
<td>0.224</td>
<td>0.238</td>
<td>0.3188</td>
</tr>
<tr>
<td></td>
<td>0.5 (replication)</td>
<td>1051</td>
<td>0.284</td>
<td>1.328</td>
<td>0.826-2.137</td>
<td>0.238</td>
<td>0.238</td>
<td>0.3183</td>
</tr>
<tr>
<td>Case vs. Control</td>
<td>1e-07 (discovery and replication)</td>
<td>14</td>
<td>0.852</td>
<td>2.345</td>
<td>0.970-5.666</td>
<td>0.055</td>
<td>0.055</td>
<td>0.5179</td>
</tr>
<tr>
<td></td>
<td>1e-05 (discovery and replication)</td>
<td>210</td>
<td>0.834</td>
<td>2.303</td>
<td>1.035-5.121</td>
<td>0.038</td>
<td>0.055</td>
<td>0.5232</td>
</tr>
<tr>
<td></td>
<td>0.001 (replication)</td>
<td>359</td>
<td>0.766</td>
<td>2.151</td>
<td>1.021-4.532</td>
<td>0.041</td>
<td>0.055</td>
<td>0.5219</td>
</tr>
<tr>
<td></td>
<td>0.05 (replication)</td>
<td>684</td>
<td>0.789</td>
<td>2.202</td>
<td>1.094-4.433</td>
<td>0.025</td>
<td>0.055</td>
<td>0.5292</td>
</tr>
<tr>
<td></td>
<td>0.5 (replication)</td>
<td>1051</td>
<td>0.655</td>
<td>1.925</td>
<td>0.998-3.714</td>
<td>0.048</td>
<td>0.055</td>
<td>0.5198</td>
</tr>
<tr>
<td>SZ vs. BD</td>
<td>1e-07 (discovery and replication)</td>
<td>14</td>
<td>1.263</td>
<td>3.538</td>
<td>1.227-10.197</td>
<td>0.018</td>
<td>0.029</td>
<td>0.5958</td>
</tr>
<tr>
<td></td>
<td>1e-05 (discovery and replication)</td>
<td>210</td>
<td>1.127</td>
<td>3.086</td>
<td>1.166-8.166</td>
<td>0.021</td>
<td>0.029</td>
<td>0.5934</td>
</tr>
<tr>
<td></td>
<td>0.001 (replication)</td>
<td>359</td>
<td>0.975</td>
<td>2.651</td>
<td>1.091-6.441</td>
<td>0.029</td>
<td>0.029</td>
<td>0.5886</td>
</tr>
<tr>
<td></td>
<td>0.05 (replication)</td>
<td>684</td>
<td>0.923</td>
<td>2.516</td>
<td>1.126-5.622</td>
<td>0.022</td>
<td>0.029</td>
<td>0.5919</td>
</tr>
<tr>
<td></td>
<td>0.5 (replication)</td>
<td>1051</td>
<td>0.803</td>
<td>2.233</td>
<td>1.080-4.617</td>
<td>0.028</td>
<td>0.029</td>
<td>0.5883</td>
</tr>
</tbody>
</table>
Figure 1. Group comparisons on raw PMPS score

Boxplots showing group differences on the PMPS calculated at the most stringent $p$-value threshold ($p < 1 \times 10^{-7}$) between: A) schizophrenia cases and healthy controls, B) bipolar disorder cases and healthy controls, C) combined clinical cases (SZ+BD) and healthy controls, and D) bipolar disorder and schizophrenia.
Table 3. PMPSs and SAPS/SANS subscale scores

<table>
<thead>
<tr>
<th>SANS/SAPS subscale score</th>
<th>B</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS Affective Flattening</td>
<td>0.117</td>
<td>1.124</td>
<td>0.933-1.354</td>
<td>0.216</td>
</tr>
<tr>
<td>SANS Alogia</td>
<td>-0.047</td>
<td>0.954</td>
<td>0.797-1.142</td>
<td>0.607</td>
</tr>
<tr>
<td>SANS Avolition</td>
<td>0.056</td>
<td>1.057</td>
<td>0.900-1.242</td>
<td>0.491</td>
</tr>
<tr>
<td>SANS Asociality</td>
<td>0.029</td>
<td>1.030</td>
<td>0.880-1.204</td>
<td>0.712</td>
</tr>
<tr>
<td>SANS Attention</td>
<td>-0.083</td>
<td>0.920</td>
<td>0.784-1.079</td>
<td>0.302</td>
</tr>
<tr>
<td>SAPS Hallucinations</td>
<td>0.033</td>
<td>1.034</td>
<td>0.874-1.222</td>
<td>0.696</td>
</tr>
<tr>
<td>SAPS Delusions</td>
<td>0.114</td>
<td>1.121</td>
<td>0.918-1.369</td>
<td>0.258</td>
</tr>
<tr>
<td>SAPS Bizarre behaviour</td>
<td>0.004</td>
<td>1.004</td>
<td>0.800-1.259</td>
<td>0.975</td>
</tr>
<tr>
<td>SAPS Thought disorder</td>
<td>-0.055</td>
<td>0.946</td>
<td>0.827-1.082</td>
<td>0.414</td>
</tr>
</tbody>
</table>
Table 4. Results from linear regression analyses investigating relationship between PMPS (1 x 10^7) for SZ and medication use in cases.

<table>
<thead>
<tr>
<th>Cohort and Medication</th>
<th>B</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SZ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imipramine Equivalent Dosage</td>
<td>-0.001</td>
<td>0.999</td>
<td>0.996-1.002</td>
<td>0.378</td>
</tr>
<tr>
<td>Chlorpromazine Equivalent Dosage</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000-1.001</td>
<td>0.115</td>
</tr>
<tr>
<td>Mood Stabiliser Usage</td>
<td>0.217</td>
<td>1.242</td>
<td>0.482-3.198</td>
<td>0.641</td>
</tr>
<tr>
<td><strong>BD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imipramine Equivalent Dosage</td>
<td>0.003</td>
<td>1.003</td>
<td>0.996-1.009</td>
<td>0.419</td>
</tr>
<tr>
<td>Chlorpromazine Equivalent Dosage</td>
<td>0.000</td>
<td>1.001</td>
<td>0.999-1.001</td>
<td>0.807</td>
</tr>
<tr>
<td>Mood Stabiliser Usage</td>
<td>-0.055</td>
<td>0.946</td>
<td>0.485-1.848</td>
<td>0.867</td>
</tr>
<tr>
<td><strong>Combined psychosis group (SZ+BD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imipramine Equivalent Dosage</td>
<td>0.001</td>
<td>1.001</td>
<td>0.999-1.002</td>
<td>0.460</td>
</tr>
<tr>
<td>Chlorpromazine Equivalent Dosage</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000-1.000</td>
<td>0.332</td>
</tr>
<tr>
<td>Mood Stabiliser Usage</td>
<td>-0.202</td>
<td>0.817</td>
<td>0.586-1.138</td>
<td>0.228</td>
</tr>
</tbody>
</table>

Note: The optimal PMPS score for SZ was based on the top 14 CpG sites associated with SZ as reported by Hannon et al.5