

1 **Increasing Sample Diversity in Psychiatric Genetics – Introducing a new Cohort of**
2 **Patients with Schizophrenia and Controls from Vietnam – Results from a Pilot Study**

3

4 Nguyen VT^{a,b*}, Braun A^{c*}, Kraft J^{c*}, Ta TMT^{e,g}; Panagiotaropoulou GM^c, Nguyen VP^a,
5 Nguyen TH^{a,g}, Trubetsky V^c, Le CT^{a,b}, Le TTH^{a,b}, Pham XT^a, Heuser-Collier I^g, Lam
6 NH^f, Böge K^g, Hahne IM^g, Bajbouj M^g, Zierhut MM^{e,g}, Hahn E^{g*}, Ripke S^{c,d,*†}

7

8 **Affiliations**

9 *^aDepartment of Psychiatry, Hanoi Medical University, Hà Nội, Việt Nam;*

10 *^bNational Institute of Mental Health, Bach Mai Hospital, Hà Nội, Việt Nam;*

11 *^cCharité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy,
12 Campus Charité Mitte, Berlin, Germany;*

13 *^dStanley Center for Psychiatric Research, Broad Institute of MIT and Harvard,
14 Cambridge, MA, United States of America;*

15 *^eBerlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Academy, Clinician
16 Scientist Program;*

17 *^fHanoi Mental Hospital, Long Biên, Hà Nội, Việt Nam;*

18 *^gCharité – Universitätsmedizin Berlin, Department of Psychiatry, Campus Benjamin Franklin,
19 Berlin, Germany*

20

21 *** These authors contributed equally to the manuscript**

22

23 **† Corresponding author**

24 Stephan Ripke, Department of Psychiatry, Charité – Universitätsmedizin Berlin, Campus
25 Mitte, Charitéplatz 1, 10117 Berlin, Germany

26 Phone: +49 30 450 517 030

27 E-Mail: sripke@broadinstitute.org

28 **Increasing Sample Diversity in Psychiatric Genetics – Introducing a new Cohort of**
29 **Patients with Schizophrenia and Controls from Vietnam – Results from a Pilot Study**

30

31 **Abstract**

32 **Objectives:** Genome-Wide Association Studies (GWAS) of Schizophrenia (SCZ) have
33 provided new biological insights; however, most cohorts are of European ancestry. As a result,
34 derived polygenic risk scores (PRS) show decreased predictive power when applied to
35 populations of different ancestries. We aimed to assess the feasibility of a large-scale data
36 collection in Hanoi, Vietnam, contribute to international efforts to diversify ancestry in SCZ
37 genetic research and examine the transferability of SCZ-PRS to individuals of Vietnamese
38 Kinh ancestry.

39 **Methods:** In a pilot study, 368 individuals (including 190 SCZ cases) were recruited at the
40 Hanoi Medical University's associated psychiatric hospitals and outpatient facilities. Data
41 collection included sociodemographic data, baseline clinical data, clinical interviews assessing
42 symptom severity and genome-wide SNP genotyping. SCZ-PRS were generated using
43 different training data sets: i) European, ii) East-Asian and iii) trans-ancestry GWAS summary
44 statistics from the latest SCZ GWAS meta-analysis.

45 **Results:** SCZ-PRS significantly predicted case status in Vietnamese individuals using mixed-
46 ancestry (R^2 liability=4.9%, $p=6.83 \times 10^{-8}$), East-Asian (R^2 liability=4.5%, $p=2.73 \times 10^{-7}$) and
47 European (R^2 liability=3.8%, $p = 1.79 \times 10^{-6}$) discovery samples.

48 **Discussion:** Our results corroborate previous findings of reduced PRS predictive power
49 across populations, highlighting the importance of ancestral diversity in GWA studies.

50

51 **Keywords:** SCHIZOPHRENIA, GWAS, VIETNAM, EAST-ASIA, POLYGENIC PREDICTION

52 **1 Introduction**

53 Schizophrenia (SCZ) is a severe and often chronic psychiatric disorder that consistently
54 affects between 0,5-1% of the worldwide population (McGrath et al., 2008). It is a clinically
55 heterogeneous disorder characterised by perceptual distortions such as delusions and
56 hallucinations, disorganised thinking, negative symptoms including reduced affective
57 resonance, avolition, lack of motivation, marked cognitive deficits, as well as mood
58 disturbances (Galderisi et al., 2018). With a high disease burden (GBD 2016 Disease and
59 Injury Incidence and Prevalence Collaborators, 2017), loss of social functioning, and reduced
60 life-expectancy (Walker et al., 2015), SCZ has a substantial impact on public health-related
61 costs (Cloutier et al., 2016). To this date, SCZ remains challenging to treat, as only about
62 13.5% of patients meet the criteria for full recovery (Charlson et al., 2018), and advances in
63 treatment options have been hampered — at least in part — by our incomplete knowledge of
64 SCZ etiology and pathophysiological mechanisms.

65 As with all complex traits and diseases, schizophrenia exhibits a diffuse pattern of
66 underlying genetic variation in which many common genetic variants contribute to disease
67 risk. In recent years large-scale genome-wide association studies (GWAS) have provided
68 robust evidence for many susceptibility loci within the genome (Schizophrenia Working Group
69 of the Psychiatric Genomics Consortium, 2014); the most recent SCZ GWAS meta-analysis
70 identified 276 distinct genomic regions (Schizophrenia Working Group of the Psychiatric
71 Genomics Consortium et al., 2020). While genetic studies can provide insights into the
72 biological underpinnings of SCZ, they also offer potential biomarkers that could inform
73 treatment strategies, patient management, prognosis, and possibly early intervention. A
74 growing number of studies have directed their attention towards the utility of polygenic risk
75 scores (PRS) in clinical psychiatry. PRS are typically constructed using genome-wide
76 summary statistics and represent an individual's propensity towards certain traits or disorders
77 attributable to common genetic variation (International Schizophrenia Consortium et al., 2009).
78 To establish the clinical relevance of genetic findings, it becomes increasingly important to
79 invest in large, well-phenotyped SCZ cohorts; especially as patients affected by severe

80 psychotic disorders are often underrepresented in population-based research initiatives (J.
81 Martin et al., 2016)

82 Although increases in sample size boost statistical power, and consequently the
83 discovery of novel risk-conferring variants (Levinson et al., 2014) and predictive capacity of
84 PRS, cohorts of non-European ancestries — especially those from low- and middle-income
85 countries — are still sparse in GWA studies. Multiple works have demonstrated that the
86 prediction accuracy of PRS depends on the ancestral composition of discovery cohorts and
87 their genetic divergence from the target sample (Bigdeli et al., 2020; Lam et al., 2019). As a
88 result, polygenic scores derived from Eurocentric GWAS suffer from an attenuated predictive
89 power when transferred to non-European populations (Duncan et al., 2019; A. R. Martin et al.,
90 2019). Consequently, PRS are less informative for non-European populations; hence, patients
91 from these populations will benefit less from polygenic scores and their potential applications.
92 Although there is emerging evidence for a shared genetic architecture of schizophrenia across
93 diverse populations (Lam et al., 2019), the incomplete transferability of GWAS findings is
94 driven by three fundamental factors: 1) differences in allele frequencies across populations,
95 as GWAS are powered only for common variants of small or modest effect-size; 2) differences
96 in linkage disequilibrium patterns that affect how well causal variants are tagged by genotype
97 arrays and imputation reference panels; 3) population genetic factors such as genetic drift,
98 polygenic selection, and demographic history (Duncan et al., 2019; A. R. Martin et al., 2017,
99 2019). To improve GWAS findings' generalizability, encouraging ancestral diversity in study
100 samples has taken high priority in the field of psychiatric genetics (Mills & Rahal, 2020).

101 Despite a surge of East-Asian samples in the latest schizophrenia GWAS
102 (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020), a
103 representative Vietnamese cohort is still lacking. To close this gap, we conducted a case-
104 control pilot study in Hanoi, Vietnam, in which sociodemographic-, genetic-, and clinical data
105 from a total of 378 participants were collected. Here we provide a first report on data
106 collection's feasibility while providing insights into clinical data and current treatment of
107 patients with schizophrenia recruited in Hanoi, Vietnam. Moreover, we aim to facilitate genetic

108 discoveries by contributing to global research endeavors to diversify genetic samples and
109 clinical data in schizophrenia research. Ultimately, this pilot study sets the foundation for the
110 recruitment of a sizeable Vietnamese well characterized cohort that includes state-of-the-art
111 phenotyping procedures in the field of psychiatric genetics.

112

113 **2 Methods**

114 ***2.1 Local Recruitment and Data Collection***

115 For this pilot study, 190 patients with schizophrenia and 188 control subjects, all of Vietnamese
116 Kinh ancestry, were recruited between August and December 2017. All participants were at
117 least 18 years old and provided full written informed consent upon study enrollment. Before
118 data collection, study procedures were reviewed and approved by the institutional ethical
119 review board of the Hanoi Medical University. All in-patients were recruited at teaching
120 hospitals of the Department of Psychiatry at Hanoi Medical University.

121 Additionally, outpatients were recruited from associated community medical centers around
122 the Hanoi metropolitan area. In outpatient facilities, data collection was supervised by senior
123 psychiatrists of the staff at the Department of Psychiatry, Hanoi Medical University. Diagnosis
124 of schizophrenia according to ICD-10 criteria (F20.X) was ascertained by clinicians with at
125 least three years of psychiatric experience and confirmed by a senior attending psychiatrist.
126 Current symptom severity was assessed with the Vietnamese Translation of the Positive and
127 Negative Syndrome Scale (Kay et al., 1987), which encompasses the dimensions of positive,
128 negative, and general psychopathology. Each patient was interviewed and rated
129 independently by two clinicians who received prior training in conducting the semi-structured
130 interview. If no consensus was reached, ratings were discussed with a senior psychiatrist.
131 Medical and demographic parameters, including medication, duration of illness, and age at
132 onset, were recorded during routine examinations and/or extracted from the patients' medical
133 chart. Control subjects were recruited from the staff at the Hanoi Medical University across
134 different occupational groups. To obtain genetic information, saliva samples were collected

135 from each subject using OraGene-510 DNA-Self-Collection Kits (Genotek, Ottawa, Ontario,
136 Canada) according to the manufacturer's instructions.

137

138 **2.2 Genetic Data**

139 *2.2.1 Genome-wide Genotyping*

140 Genetic material was extracted from saliva following established standard protocols. Genome-
141 wide genotyping of DNA samples was performed on Infinium Global Screening array-23
142 (GSA) MD BeadChip (Illumina, San Diego, CA) at ERASMUS Medical Center, Human
143 Genotyping Facility (HuGe-F), Rotterdam, Netherlands. SNP-level data was analyzed using a
144 central pipeline 'RICOPILI' (Lam et al., 2020) on the LISA computer cluster.

145

146 *2.2.2 Quality Control*

147 According to the Psychiatric Genomics Consortium (PGC) standards, quality control was
148 performed on the newly recruited cohort. The default quality control parameters for retaining
149 subjects and SNPs were: SNP missingness < 0.05 (before sample removal); subject
150 missingness < 0.02 ; autosomal heterozygosity deviation ($| F_{\text{het}} | < 0.2$); SNP missingness $<$
151 0.02 (after sample removal); difference in SNP missingness between cases and controls $<$
152 0.02 ; and SNP Hardy-Weinberg equilibrium ($P > 10^{-6}$ in controls or $P > 10^{-10}$ in cases). Seven
153 individuals who did not meet QC criteria were excluded resulting in a data set of 370 individuals
154 (184 cases and 186 controls) and 367,920 SNPs. Post-QC Manhattan and QQ plots did not
155 show unexpected genome-wide significant association or general inflation.

156

157 *2.2.3 Genotype Imputation*

158 Genotype imputation was performed using the pre-phasing/imputation stepwise approach
159 implemented in EAGLE/MINIMAC3 (with a variable chunk size of 132 genomic chunks and
160 default parameters). The imputation reference set consisted of 54,330 phased haplotypes with
161 36,678,882 variants from the publicly available HRC reference release 1.1 (McCarthy et al.,
162 2016).

163 2.2.4 Population Stratification and Relatedness Testing

164 After imputation, we identified SNPs with very high imputation quality (INFO >0.8) and low
165 missingness (<1%) for further quality control. After linkage disequilibrium pruning ($r^2 > 0.02$)
166 and frequency filtering (MAF > 0.05), there were 116,203 autosomal SNPs in the data set.
167 This SNP set was used for robust relatedness testing and population structure analysis using
168 PLINK (Purcell et al., 2007). Thirteen pairs of subjects with identity-by-descent PI-HAT value
169 of > 0.2 were identified. One member of each pair was removed at random after preferentially
170 retaining cases over controls. Principal component estimation was done with the same
171 collection of autosomal SNPs, and fifteen ancestry outliers were removed from the data set
172 (see supplementary figures S2-S3). Supplementary figure S3 (A) shows that Vietnamese
173 individuals cluster within other Asian populations from the 1000 Genomes reference
174 population (Auton et al., 2015). Figure S3 (B) illustrates that the Vietnamese Kinh population
175 is distinctly stratified from China and Japan when plotted only against Asian populations. We
176 tested the first 20 principal components for phenotype association (using logistic regression
177 with study indicator variables included as covariates). Seven principal components, namely
178 PC1, PC2, PC3, PC4, PC5, PC8, and PC14, were significantly associated ($P < 0.05$) with SCZ
179 status, and these were added to the GWAS model as covariates in order to control for
180 population stratification. The final analysis consisted of 343 individuals (174 cases and 169
181 controls) and 8,332,179 SNPs. We tested all SNPs for association under an additive logistic
182 regression model using PLINK and the derived principal components as covariates.

183

184 2.2.5 Polygenic Risk Scoring

185 We obtained summary statistics from the latest PGC schizophrenia GWAS (Schizophrenia
186 Working Group of the Psychiatric Genomics Consortium et al., 2020) to calculate PRS for
187 individuals in our target data set. Polygenic scoring was performed using (a) East-Asian (EAS)
188 specific meta-analysis (14,004 cases and 16,757 controls), (b) European (EUR) specific meta-
189 analysis (53,386 cases and 77,258 controls) and (c) trans-ancestry (MIX) meta-analysis
190 (67,390 cases and 94,015 controls) as training statistics. Genome-wide summary statistics

191 were LD pruned and “clumped” by discarding variants within 500 kb of, and in $r^2 \geq 0.1$ with
192 another (more significant) marker, resulting in 61 805 (EAS), 90,713, (EUR) 87,671 variants
193 (MIX), respectively. Clumping for the East-Asian ancestry was conducted with the LD
194 information of the East-Asian subset of the HRC reference panel. The EUR and MIX training
195 data were clumped using LD information from the entire HRC reference panel. This set of LD-
196 independent SNPs was used to construct a PRS at a range of P-value thresholds (5×10^{-8} ,
197 1×10^{-6} , 1×10^{-4} , 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0). For a given sample, we take each LD-
198 independent SNP and multiply the marginal SNP effect from the logistic regression by the
199 number of risk alleles. The resulting products were summed over all variants for each
200 individual so that a single score reflecting the overall polygenic burden for schizophrenia is
201 obtained.

202

203 *2.2.6 Consistency of Directions of Effects*

204 In an additional final step, we assessed the robustness of genetic associations in the
205 Vietnamese pilot sample by performing a replication analysis using the PGC trans-ancestry
206 discovery sample as a discovery data set. We used a binomial sign test to establish that SNP
207 associations replicated *en masse* in an independent PGC sample at various p-value
208 thresholds by showing that the number of concordant effect directions for index SNPs was
209 significantly greater than the null hypothesis of randomly oriented effects. The Vietnam
210 summary statistics were then combined with those from the PGC primary GWAS dataset using
211 an inverse variance-weighted fixed effects model and overlapping LD-clumps were merged
212 into distinct genomic regions.

213

214 **2.3 Polygenic Association Testing**

215 Individual-level PRS at each p-value threshold (P_d) were tested for association with SCZ
216 status (Schizophrenia vs. No Schizophrenia) in logistic regression models. To adjust for fine-
217 grained, population-stratification principal components (PCs) 1, 2, 3, 4, 5, 8, and 14 (selected
218 for being significantly associated with SCZ status) were included as covariates in each model.

219 Nagelkerke's R^2 , as a measure of "variance explained" for logistic regression, was determined
220 by comparing the full model (with PRS and PCs) against the null model containing only PCs.
221 We report Nagelkerke's R^2 on the liability scale (Lee et al., 2012), which corrects for a higher
222 case-control-ratio in the target dataset relative to the general population's prevalence
223 estimated 0.7% (McGrath et al., 2008). Using LD-clumped genome-wide summary statistics
224 without any p-value filter, individuals were grouped into five quintiles based on their individual
225 polygenic burden (SCZ-PRS). Odds ratios for schizophrenia diagnosis were obtained by
226 comparing individuals with low polygenic risk (bottom 20%, or bottom quintile) to individuals in
227 strata with increased polygenic loadings for SCZ (second to the fifth quintile). Each step was
228 repeated for three different training data sets: East-Asian GWAS (EAS), European GWAS
229 (EUR), and mixed-ancestry GWAS (PGC3 SCZ-GWAS).

230

231 **3 Results**

232 **3.1 Sample Characteristics**

233 A total of 378 participants were included in the study, of which 190 are cases diagnosed with
234 schizophrenia and 188 are control subjects. Sociodemographic parameters of all case and
235 control subjects are shown in Table 1. All clinical parameters of the case sample are presented
236 in Table 2.

237

238 **3.2 Polygenic Risk Profiling**

239 Next, we assessed how much variance in schizophrenia risk could be attributed to underlying
240 genetic variation. After quality control (see **Supplementary** material), 343 individuals
241 remained for polygenic risk score analysis. Schizophrenia polygenic risk scores (SCZ-PRS)
242 were generated with GWAS summary statistics from East-Asian (EAS), European (EUR), and
243 mixed-ancestry populations as training sets. In line with previous findings, individual-level
244 SCZ-PRS were significantly associated with schizophrenia case status (**Fig 1**,
245 **Supplementary Table S1**). However, the predictive accuracy of SCZ-PRS among
246 Vietnamese participants varied with the ancestral composition of the training data sets. The

247 strongest effect sizes were observed when PRS were derived from trans-ancestry summary
248 statistics, which explained ~4.9 % of the variance in schizophrenia liability ($p = 6.83 \times 10^{-8}$ at
249 $P_d < 1$). Likewise, PRS constructed on East-Asian GWAS summary statistics (PGC SCZ wave
250 3) accounted for ~4.5 % of the variability in schizophrenia risk ($p = 2.73 \times 10^{-7}$ at $P_d < 1$). PRS
251 trained on European ancestry GWAS results also predicted SCZ status in the Vietnamese
252 target sample with an explained variance of only ~3.9 % ($p = 1.79 \times 10^{-6}$ at $P_d < 0.01$). As
253 expected, best overall predictions were observed at more inclusive p-value thresholds that
254 capture a larger proportion of the polygenic signal by including more independent common
255 variants below the genome-wide significance level.

256 A similar pattern can be observed when inspecting odds ratios for the association
257 between SCZ-PRS and schizophrenia diagnosis (**Fig 1c, Supplementary Table S3**). Overall,
258 individuals who carry a higher polygenic burden are at higher odds for having a schizophrenia
259 diagnosis, ORs range from 1.97 (95% CI, 0.96 - 4.12) at the second PRS-quintile up to 5.74
260 (95% CI, 2.70 - 12.69) at the fifth PRS-quintile, using the first quintile as the reference group
261 and mixed-ancestry training populations for polygenic scoring. When considering scores
262 obtained from east-Asian ancestry GWAS, odds ratios for SCZ are slightly lower, ranging from
263 1.50 (95% CI, 0.72 - 3.14) at the second PRS-quintile up to 4.47 (95% CI, 2.13 – 9.66) at the
264 top PRS-quintile. By comparison, observed ORs are markedly lower for EUR-derived SCZ-
265 PRS; the OR at the fifth PRS quintile is 3.39 (95% CI, 1.63 – 7.26) relative to the first PRS
266 quintile.

267 Overall, reported predictive values of PRS and odds ratios at the top PRS-quintile in
268 our Vietnamese sample are consistent with those of other East-Asian cohorts reported within
269 the latest SCZ GWAS meta-analysis (Schizophrenia Working Group of the Psychiatric
270 Genomics Consortium, 2020).

271

272 **3.3 Replication Analysis**

273 Complementing PRS association testing, we explored if SNP associations at varying p-value
274 thresholds (P_d) in the mixed-ancestry SCZ3-GWAS replicate in the present Vietnamese-

275 ancestry sample (replication data set). We observed that the fraction of SNPs with directions
276 of effect congruent to the replication data set was consistently greater than 50%. Specifically,
277 of the 182 index SNPs that attained genome-wide significance in the SCZ3-GWAS meta-
278 analysis and that were also available in the Vietnamese cohort, 110 SNPs (60%) showed the
279 same direction of effect, which was significantly greater than expected by chance (binomial p-
280 value = 0.003). Results of the replication analysis and binomial sign test across different p-
281 value thresholds are reported in the supplementary material (**Supplementary Table S3**).
282 When meta-analyzed with SCZ3-GWAS, 249 loci reach genome-wide significance, in contrast
283 to 244 loci identified by SCZ3-GWAS alone, indicating that by adding this or a potentially larger
284 Vietnamese cohort could help to discover novel SCZ-associated loci in future GWAS meta-
285 analysis.

286

287 **4 Discussion**

288 With this pilot study, the feasibility of a large-scale data collection of patients diagnosed with
289 schizophrenia in Vietnam has been shown for the first time. We were able to implement a
290 successful recruitment scheme at the Department of Psychiatry of the Hanoi Medical
291 University, with the National Institute of Mental Health staff at Bach Mai Hospital and the Hanoi
292 Psychiatric Hospital participating in data collection. This study established an efficient pipeline
293 for data generation and subsequent analyses that will be upscaled with future collaborative
294 endeavors. With these central foundations set in place, we aim to continue and intensify our
295 data collection efforts and establish more in-depth phenotyping going forward.

296 There are a few practical and clinical considerations concerning the recruitment of
297 patients with schizophrenia in Vietnam. First, patients were primarily recruited through in-
298 patient facilities. For genetic research, that may prove to be an advantage since the clinical
299 population in Vietnam comprises mostly cases with severe illness courses of schizophrenia,
300 long duration of illness, multiple prior hospitalizations, and a high symptom load, especially for
301 negative and residual symptoms (Maramis et al., 2011). This pattern can be considered
302 representative of inpatients often treated for long durations for chronic symptoms of

303 schizophrenia in Vietnam, since individuals with a less severe course of illness typically won't
304 seek or receive hospital treatment due to stigma or a lack of illness recognition for less severe
305 illness courses (Martensen et al., 2020; Nguyen et al., 2019; Ta et al., 2016, 2018). This
306 pattern is reflected by the fact that even today, more than 50% of all patients treated in
307 psychiatric hospitals in Vietnam have a primary psychiatric diagnosis of schizophrenia
308 (Maramis et al., 2011; Martensen et al., 2018, 2020). In our pilot sample, most of the patients
309 were hospitalised several times due to schizophrenia. Nearly all patients, including the
310 outpatients, still exhibited chronic, residual, or negative symptoms when assessed by the
311 PANSS at the time of recruitment. To increase clinical heterogeneity among patients, we will
312 integrate more outpatient facilities in the Hanoi metropolitan area as additional study sites in
313 future studies.

314 In the PRS and replication analysis, we corroborated some key findings of previous
315 publications. First, we observed that PRS derived from trans-ancestry GWAS perform better
316 than EAS-derived PRS, and both predict schizophrenia more accurately than EUR-derived
317 PRS. Similarly, leave-one-out PRS analysis in the latest SCZ GWAS meta-analysis
318 demonstrated that PRS in Asian cohorts explained consistently more variance in
319 schizophrenia risk when mixed-ancestry summary statistics were used for polygenic scoring
320 compared to Asian training populations (Schizophrenia Working Group of the Psychiatric
321 Genomics Consortium et al., 2020). Consistent with this finding, Lam et al. (2019)
322 demonstrated that EUR-derived PRS accounted for less variance in schizophrenia liability
323 than EAS-derived PRS. Second, we demonstrated that risk-conferring SNPs in the present
324 Vietnamese samples largely showed same-direction effects compared to SCZ PGC3
325 summary statistics. The inclusion of our sample, especially after scaling-up the sample size,
326 will probably uncover novel susceptibility loci, especially since an earlier meta-analysis of
327 East-Asian (Lam et al., 2019) and Latino and admixed African samples (Bigdeli et al., 2020)
328 strongly support this notion. Due to practical limitations, we focused on Kinh ancestry
329 individuals, representing 87% of the Vietnamese population. Given the genetic diversity
330 among Vietnamese populations (Macholdt et al., 2020), our genotyping initiative might be

331 extended to other ethnic groups living in Vietnam and bordering geographical areas in South-
332 East-Asia. Thus, the addition of currently understudied populations such as Vietnamese
333 ethnicities will likely help refine our understanding of the genetic architecture of schizophrenia.

334 Ancestral diversity among study participants in discovery GWAS has a substantial
335 impact on prediction accuracy of polygenic scores and their applicability across populations.
336 Further inclusion of East- and South-East-Asian cohorts will likely improve the polygenic risk
337 prediction in Vietnamese individuals and vice versa, more than by a comparable addition of
338 European ancestry samples. This is increasingly important as PRS are beginning to find their
339 way into clinical practice outside of psychiatry. Beyond disease risk prediction, the clinical
340 utility of PRS may range from genotype-informed screening methods and stratification to
341 tailored intervention strategies (Torkamani et al., 2018). To translate genome-wide data into
342 clinically actionable information, in-depth clinical and valid phenotypic information must be
343 collected along with genetic data. While these pilot data are too small to perform meaningful
344 secondary analyses, e.g., on symptom profiles, such cohorts will play a pivotal role in making
345 potential PRS-based applications available for a wider population.

346

347 **5 Conclusion and Outlook**

348 This pilot research project's main objective was to establish key research and clinical
349 infrastructures for a larger cross-sectional Vietnamese cohort of patients with schizophrenia
350 and healthy controls. In the next phase, we will extend our data collection efforts by introducing
351 additional study sites in Vietnam, both in northern and southern provinces, as well as a
352 translated and validated comprehensive phenotyping battery. This includes trained
353 assessment, that is harmonised with our current research initiative in Berlin, Germany. This
354 assessment is comprised of an in-depth neuropsychological evaluation and a selection of
355 questionnaires that assess environmental risk factors, a broad range of health-related data,
356 as well as outcome measures such as quality of life and social functioning. With this
357 prospective cohort, we intend to build a valuable resource for trans-ancestry GWAS and
358 secondary analyses on cross-cultural gene-environment interactions to elucidate potential

359 etiological trajectories associated with an increased liability to Schizophrenia. In doing so, we
360 aim to contribute to a worldwide endeavor to enhance well-powered and ancestrally diverse
361 GWAS samples and explore trans-ancestry and culture specific gene-environment
362 interactions in psychiatry.

Acknowledgements

We thank the Schizophrenia Working Group of the Psychiatric Genomics Consortium for granting permission to use European, East-Asian and trans-ancestry GWAS summary statistics (wave 3) for this study. The Stanley Center for Psychiatric Research financially supported the collection of this sample. The statistical analysis was carried out on the Dutch LISA computer cluster powered by SURFsara. We further thank the participants and additional staff of Hanoi Psychiatric Hospital, Hanoi Medical University, and the National Institute of Mental Health, Vietnam, for taking part in this study.

Disclosure of interest

None of the authors involved in this manuscript report conflicts of interest.

Table 1. Sociodemographic characteristics of the health control and patient sample

		Vietnamese sample (N=378)	
		Patients (N=190)	Controls (N=188)
Age [years]	Mean (SD)	44.44 (10.62)	44.21 (10.43)
Sex (%)			
	Male	147 (77.37)	134 (71.28)
	Female	43 (22.63)	54 (28.72)
Weight [kg]	Mean (SD)	57.92 (9.81)	62.46 (8.97)
Height [cm]	Mean (SD)	162.45 (7.35)	164.34 (7.14)
BMI [kg/m²]	Mean (SD)	21.90 (3.33)	23.06 (2.40)
Marital Status (%)			
	Single	97 (51.32)	9 (4.86)
	Married	58 (30.69)	172 (92.97)
	Living separately	10 (5.29)	1 (0.54)
	Divorced	21 (11.11)	0 (0.00)
	Widowed	3 (1.59)	3 (1.62)
School Education (%)			
	Illiterate	2 (1.06)	0 (0.00)
	Primary School	29 (15.4)	1 (0.53)
	Secondary School	59 (31.38)	21 (11.23)
	High School	100 (53.19)	165 (88.24)
Higher Education (%)			
	Apprentice	12 (27.91)	6 (4.32)
	Under-College	9 (20.93)	40 (28.78)
	College / University	22 (51.65)	67 (48.22)
	Post-University	0 (0.00)	26 (18.71)

Table 2. Clinical parameters

		N (%)	Mean (SD)
Clinical information			
Duration of illness [years]			18.72 (10.09)
Average age of onset of positive symptoms			25.55 (7.52)
No. of hospitalisations ¹			11.51 (15.50)
Positive Family History ²	at least 1 family member affected	54 (28.42)	
Recruitment site	Inpatients	96 (50.53)	
	Outpatients	94 (49.47)	
Treatment			
Antipsychotic Medication ³	Olanzapine	59 (31.05)	
	Risperidone	14 (7.37)	
	Clozapine	12 (6.32)	
	Haloperidol	7 (3.68)	
	Chlorpromazine	11 (5.79)	
	Levomepromazine	68 (35.79)	
	Other antipsychotics	17 (9.45)	
Symptom severity			
Symptoms within the last year	Complete remission	1 (0.53)	
	Currently symptomatic	188 (98.95)	
PANSS	PANSS-Pos		14.80 (6.50)
	PANSS-Neg		24.42 (8.81)
	PANSS-Gen		34.01 (12.08)
	PANSS-Total		73.12 (22.52)
Domains of psychopathology	Delusions	75 (39.47)	
	Hallucinations	80 (42.11)	
	Disorganisation	77 (40.53)	
	Catatonia	21 (11.05)	
	Negative symptoms	189 (99.47)	
	Hostile behavior	58 (30.53)	
	Marked affective symptoms	21 (11.05)	
	Other psychopathological symptoms	5 (2.63)	

¹ Number of lifetime in-patient admissions due to schizophrenia

² Family history of psychiatric illnesses

³ All patients received at least one antipsychotic medication

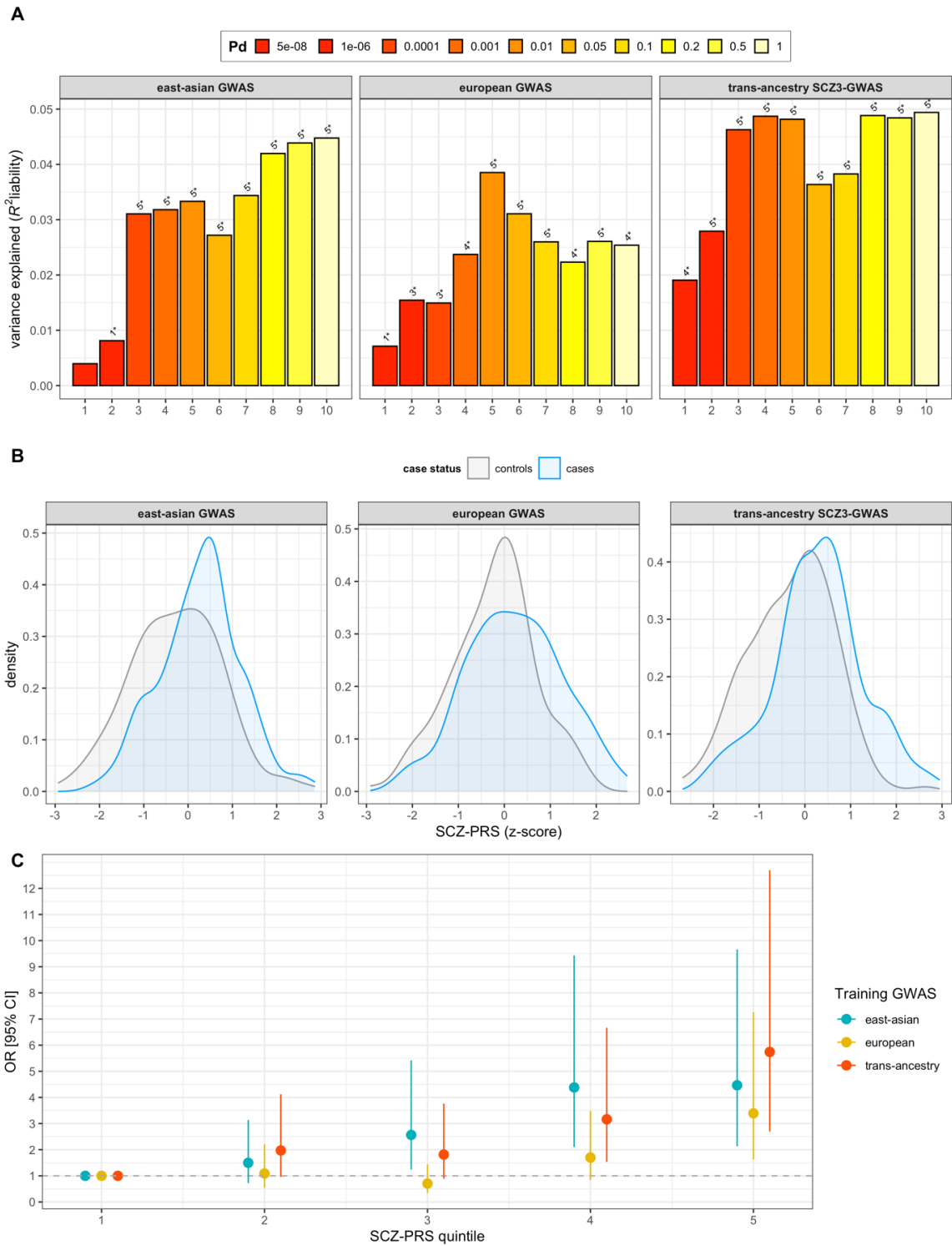


Fig 1. Associations of SCZ-PRS with schizophrenia in a cohort of 169 controls and 174 cases of Vietnamese Kinh ancestry. **A)** The variance explained (R^2 liability) by SCZ-PRS at different p-value thresholds (Pd; x-axis) is shown on the y-axis. P-values are displayed on top of each bar: 1* < 0.05 , 2* < 0.01 , 3* < 0.005 , 4* < 0.001 , 5* $< 1 \times 10^{-4}$. SCZ-PRS were constructed on different GWAS discovery samples: East-Asian GWAS (left panel), European-ancestry GWAS (center panel), and trans-ancestry SCZ3-GWAS (right panel). **B)** Distribution of z-transformed polygenic risk scores among cases and controls at Pd < 1 . **C)** odds ratios for schizophrenia per SCZ-PRS (Pd < 1) quintile relative to the first quintile. Error bars represent 95% Confidence Intervals.

References

- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., ... National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Bigdeli, T. B., Genovese, G., Georgakopoulos, P., Meyers, J. L., Peterson, R. E., Iyegbe, C. O., Medeiros, H., Valderrama, J., Achtyes, E. D., Kotov, R., Stahl, E. A., Abbott, C., Azevedo, M. H., Belliveau, R. A., Bevilacqua, E., Bromet, E. J., Byerley, W., Carvalho, C. B., Chapman, S. B., ... Genomic Psychiatry Cohort (GPC) Consortium. (2020). Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Molecular Psychiatry*, *25*(10), 2455–2467. <https://doi.org/10.1038/s41380-019-0517-y>
- Charlson, F. J., Ferrari, A. J., Santomauro, D. F., Diminic, S., Stockings, E., Scott, J. G., McGrath, J. J., & Whiteford, H. A. (2018). Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophrenia Bulletin*, *44*(6), 1195–1203. <https://doi.org/10.1093/schbul/sby058>
- Cloutier, M., Aigbogun, M. S., Guerin, A., Nitulescu, R., Ramanakumar, A. V., Kamat, S. A., DeLucia, M., Duffy, R., Legacy, S. N., Henderson, C., Francois, C., & Wu, E. (2016). The Economic Burden of Schizophrenia in the United States in 2013. *The Journal of Clinical Psychiatry*, *77*(6), 764–771. PubMed. <https://doi.org/10.4088/jcp.15m10278>
- Duncan, L., Shen, H., Gelaye, B., Meijssen, J., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1), 3328. <https://doi.org/10.1038/s41467-019-11112-0>
- GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the

- Global Burden of Disease Study 2016. *Lancet (London, England)*, 390(10100), 1211–1259. PubMed. [https://doi.org/10.1016/S0140-6736\(17\)32154-2](https://doi.org/10.1016/S0140-6736(17)32154-2)
- International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748–752. PubMed. <https://doi.org/10.1038/nature08185>
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276. <https://doi.org/10.1093/schbul/13.2.261>
- Lam, M., Awasthi, S., Watson, H. J., Goldstein, J., Panagiotaropoulou, G., Trubetsky, V., Karlsson, R., Frei, O., Fan, C.-C., De Witte, W., Mota, N. R., Mullins, N., Brügger, K., Lee, S. H., Wray, N. R., Skarabis, N., Huang, H., Neale, B., Daly, M. J., ... Ripke, S. (2020). RICOPILI: Rapid Imputation for COnsortias PipeLine. *Bioinformatics*, 36(3), 930–933. <https://doi.org/10.1093/bioinformatics/btz633>
- Lam, M., Chen, C.-Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S.-W., Kubo, M., Kusumawardhani, A. A. A. A., Liu, C.-M., Ma, H., ... Huang, H. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*, 51(12), 1670–1678. PubMed. <https://doi.org/10.1038/s41588-019-0512-x>
- Lee, S. H., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2012). A Better Coefficient of Determination for Genetic Profile Analysis. *Genetic Epidemiology*, 36(3), 214–224. <https://doi.org/10.1002/gepi.21614>
- Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R., & Sullivan, P. F. (2014). Genetic studies of major depressive disorder: Why are there no genome-wide association study findings and what can we do about it? *Biological Psychiatry*, 76(7), 510–512. PubMed. <https://doi.org/10.1016/j.biopsych.2014.07.029>

- Macholdt, E., Arias, L., Duong, N. T., Ton, N. D., Van Phong, N., Schröder, R., Pakendorf, B., Van Hai, N., & Stoneking, M. (2020). The paternal and maternal genetic history of Vietnamese populations. *European Journal of Human Genetics*, 28(5), 636–645. <https://doi.org/10.1038/s41431-019-0557-4>
- Maramis, A., Van Tuan, N., & Minas, H. (2011). Mental health in southeast Asia. *The Lancet*, 377(9767), 700–702. [https://doi.org/10.1016/S0140-6736\(10\)62181-2](https://doi.org/10.1016/S0140-6736(10)62181-2)
- Martensen, L. K., Hahn, E., Cao, T. D., Schomerus, G., Nguyen, M. H., Böge, K., Nguyen, T. D., Mungee, A., Dettling, M., Angermeyer, M. C., & Ta, T. M. T. (2018). Impact of perceived course of illness on the desire for social distance towards people with symptoms of schizophrenia in Hanoi, Vietnam. *Psychiatry Research*, 268, 206–210. <https://doi.org/10.1016/j.psychres.2018.05.046>
- Martensen, L. K., Hahn, E., Duc, C. T., Schomerus, G., Böge, K., Dettling, M., Angermeyer, M. C., Nguyene, V. T., & Ta, T. M. T. (2020). Impact and differences of illness course perception on the desire for social distance towards people with symptoms of depression or schizophrenia in Hanoi, Vietnam. *Asian Journal of Psychiatry*, 50, 101973. <https://doi.org/10.1016/j.ajp.2020.101973>
- Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., & Kenny, E. E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics*, 100(4), 635–649. PubMed. <https://doi.org/10.1016/j.ajhg.2017.03.004>
- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), 584–591. <https://doi.org/10.1038/s41588-019-0379-x>
- Martin, J., Tilling, K., Hubbard, L., Stergiakouli, E., Thapar, A., Davey Smith, G., O'Donovan, M. C., & Zammit, S. (2016). Association of Genetic Risk for Schizophrenia With Nonparticipation Over Time in a Population-Based Cohort Study. *American Journal of Epidemiology*, 183(12), 1149–1158. PubMed. <https://doi.org/10.1093/aje/kww009>

- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., ... Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. PubMed. <https://doi.org/10.1038/ng.3643>
- McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews*, *30*(1), 67–76. <https://doi.org/10.1093/epirev/mxn001>
- Mills, M. C., & Rahal, C. (2020). The GWAS Diversity Monitor tracks diversity by disease in real time. *Nature Genetics*, *52*(3), 242–243. <https://doi.org/10.1038/s41588-020-0580-y>
- Nguyen, T., Tran, T., Tran, H., Tran, T., & Fisher, J. (2019). The burden of clinically significant symptoms of common and severe mental disorders among adults in Vietnam: A population-based cross-sectional survey. *BMC Public Health*, *19*(1), 1173. <https://doi.org/10.1186/s12889-019-7513-7>
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. <https://doi.org/10.1038/nature13595>
- Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Walters, J. T., & O'Donovan, M. C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*, 2020.09.12.20192922. <https://doi.org/10.1101/2020.09.12.20192922>
- Ta, T. M. T., Böge, K., Cao, T. D., Schomerus, G., Nguyen, T. D., Dettling, M., Mungee, A., Martensen, L. K., Diefenbacher, A., Angermeyer, M. C., & Hahn, E. (2018). Public attitudes towards psychiatrists in the metropolitan area of Hanoi, Vietnam. *Asian Journal of Psychiatry*, *32*, 44–49. PubMed. <https://doi.org/10.1016/j.ajp.2017.11.031>

- Ta, T. M. T., Zieger, A., Schomerus, G., Cao, T. D., Dettling, M., Do, X. T., Mungsee, A., Diefenbacher, A., Angermeyer, M. C., & Hahn, E. (2016). Influence of urbanity on perception of mental illness stigma: A population based study in urban and rural Hanoi, Vietnam. *International Journal of Social Psychiatry*, 62(8), 685–695.
<https://doi.org/10.1177/0020764016670430>
- Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9), 581–590.
<https://doi.org/10.1038/s41576-018-0018-x>
- Walker, E. R., McGee, R. E., & Druss, B. G. (2015). Mortality in Mental Disorders and Global Disease Burden Implications: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, 72(4), 334–341. <https://doi.org/10.1001/jamapsychiatry.2014.2502>