

1 **Full title**

2 **Environmental profiles of urban living relate to regional brain**
3 **volumes and symptom groups of mental illness through distinct**
4 **genetic pathways.**

5 **Author names and affiliations:** Jiayuan Xu^{1,#,*}, Nana Liu^{1,#}, Elli Polemiti^{2,3}, Liliana
6 Garcia Mondragon⁴, Jie Tang¹, Xiaoxuan Liu^{5,6}, Tristram Lett^{2,3}, Le Yu^{5,6}, Markus
7 Noethen⁷, Chunshui Yu^{1,8}, Andre Marquand⁹, Gunter Schumann^{2,3,*}

8
9 ¹ Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin
10 Medical University General Hospital, Tianjin 300052, P.R. China

11 ² Centre for Population Neuroscience and Stratified Medicine (PONS), Institute for
12 Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University,
13 Shanghai, P.R. China

14 ³ Centre for Population Neuroscience and Stratified Medicine (PONS), Charite Mental
15 Health, Dept. of Psychiatry and Psychotherapy, CCM, Charite Universitätsmedizin
16 Berlin, Germany

17 ⁴ International Max Planck Research School for Translational Psychiatry (IMPRS-TP),
18 Munich 80804, Germany

19 ⁵ Ministry of Education Key Laboratory for Earth System Modeling, Department of
20 Earth System Science, Tsinghua University, Beijing 100084, P.R. China

21 ⁶ Joint Center for Global Change Studies, Beijing 100875, P.R. China

22 ⁷ Institute of Human Genetics, University Hospital of Bonn, Venusberg-Campus 1, D-
23 53127 Bonn, Germany

24 ⁸ CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese
25 Academy of Sciences, Shanghai, 200031, P.R. China

26 ⁹ Predictive Clinical Neuroscience group at the Donders Institute, Radboud University
27 Medical Center, The Netherlands

28 [#] These authors contributed equally to this work.

29 ***Correspondence to:**

30 Professor Gunter Schumann, M.D., Ph.D.,
 31 Centre for Population Neuroscience and Stratified Medicine (PONS), Institute for
 32 Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University,
 33 Shanghai, P.R. China
 34 Email: gunter_schumann@fudan.edu.cn
 35 and
 36 Centre for Population Neuroscience and Stratified Medicine (PONS), Charite Mental
 37 Health, Dept. of Psychiatry and Clinical Neuroscience, CCM, Charite
 38 Universitätsmedizin Berlin, Germany
 39 E-mail: gunter.schumann@charite.de
 40
 41 Jiayuan Xu, M.D. Ph.D.
 42 Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin
 43 Medical University General Hospital, Tianjin 300052, P.R. China
 44 E-mail: xujiayuan@tmu.edu.cn
 45

46 Abstract

47 The majority of people worldwide live in cities, yet how urban living affects brain and
 48 mental illness is scarcely understood. Urban lives are exposed to a a wide array of
 49 environmental factors that may combine and interact to influence mental health. While
 50 individual factors of the urban environment have been investigated in isolation, to date
 51 no attempt has been made to model how the complex, real life exposure to living in the
 52 city relates to brain and mental illness, and how it is moderated by genetic factors. Using
 53 data of over 150,000 participants of the UK Biobank, we carried out sparse canonical
 54 correlation analyses (sCCA) to investigate the relation of urban living environment with
 55 symptoms of mental illness. We found three mental health symptom groups, consisting
 56 of affective, anxiety and emotional instability symptoms, respectively. These groups
 57 were correlated with distinct profiles of urban environments defined by risk factors
 58 related to social deprivation, air pollution and urban density, and protective factors
 59 involving green spaces and generous land use. The relations between environment and
 60 symptoms of mental illness were mediated by the volume of brain regions involved in
 61 reward processing, emotional processing and executive control, and moderated by
 62 genes regulating stress response, neurotransmission, neural development and
 63 differentiation, as well as epigenetic modifications. Together, these findings indicate
 64 distinct biological pathways by which different environmental profiles of urban living
 65 may influence mental illness. Our results also provide a quantitative measure of the
 66 contribution of each environmental factor to brain volume and symptom group. They

will aid in targeting and prioritizing important decisions for planning and public health interventions.

Introduction

More than 50% of the world population lives in an urban area, and it is estimated that by 2050, two-thirds of the world population will live in cities¹. This dramatic increase in urbanization means that living environments are going through drastic transformations: Lives in urban areas are led in increasingly higher density residential and commercial buildings¹, concomitant reduced access to green areas², increased exposures to potentially harmful substances³ and more stressful social conditions⁴. At the same time urbans residents potentially benefit from better infrastructure and more work opportunities than rural dwellers¹.

The impact of the urban living environment on mental health, is not well understood. Thus far there is evidence of urban-rural differences in the prevalence of psychiatric disorders⁵, and recent research has confirmed that individuals living in urban environments are at a higher risk of experiencing various mental health issues. The most prevalent of these issues are increased emotional problems, including symptoms of depression and anxiety⁵⁻⁸. In addition to the investigation of a general and complex environmental factor as urbanicity, there have been investigations of isolated environmental factors relevant to urban living, such as greenspaces^{9,10}, socio-economic status¹¹ and others¹². But these isolated factors have not been considered in the wider

environmental context that characterises a living environment. In order to develop targeted prevention and intervention programmes ranging from urban planning to individual psychosocial coping programmes, it is neither sufficient to regard urbanicity as one risk factor, nor is it sufficient to focus on single isolated environmental factors alone. The urban environment, as any other living environment, is composed of simultaneous interacting factors, which may form profiles or signatures that together can reduce or increase risk for mental illness^{13,14}.

The relation of symptoms of mental illness and brain structure with the exposure to environmental profiles of multiple factors composing a living environment are currently unknown, either in urban settings or otherwise. Furthermore, exposure to environmental adversity does not result in a uniform response, but individual differences have been well documented¹⁵. Genetic variations are known to be one important source of such individual differences¹⁶. In fact, activity of biological pathways, such as the stress response pathway or epigenetic modifications that are known mediators of the effect of stressful environmental stimuli on brain and mental illness have been shown to vary depending on particular genotypes¹⁷.

In this paper, we identify environmental profiles of urban living and relate them to groups of mental illness symptoms. We aim to understand what combinations of environmental profiles are most relevant for these emotional problems, and how within these combinations each single factor contributes to risk or resilience. We also identify regional brain areas that mediate the effect of different environmental profiles on

109 mental illness symptoms groups. To discover moderators that may underlie individual
110 differences in response to adverse environmental profiles, we investigate genetic
111 variations derived from genome-wide analyses of symptoms of mental illness and test
112 them for moderation of the regional brain volumes correlated to environmental
113 adversity. To this end, we carried out four analysis steps: (1) a two-way sparse
114 canonical correlation analysis (sCCA) to identify environmental profiles and their
115 related symptom groups of mental illness. (2) a genome-wide association study (GWAS)
116 of the symptom groups identified, followed by the computation of gene scores of
117 GWAS-significant genes. (3) a three-way multiple-sparse canonical correlation
118 analysis (msCCA) relating urban living environmental profiles and symptom groups
119 with regional brain volume. (4) a moderated mediation analysis to characterize the
120 mediation of environmental profiles and symptom groups by regional brain volume,
121 and its moderation by gene scores (Figure 1 and Supplementary Figure 1).

122 **Results**

123 **Overview of analyses**

124 Our analyses were carried out in participants of UK Biobank, a large and
125 predominantly urban population-based cohort. A subset of 156,375 participants that
126 were assessed for both, 21 mental symptoms measures as well as 128 environmental
127 variables linked to their home address. These environmental variables included air and
128 sound pollution, traffic, greenspace and coastal proximity, socioeconomic indices of
129 multiple deprivation (IMD), building class, destination accessibility, land use density,

130 terrain, normalized difference vegetation index (NDVI) and street network accessibility.

131 The participants from UK Biobank -Biobank Urban Morphometric Platform (UK-

132 BUMP), with complete urban living environmental data and mental health data

133 (n=156,375) were divided into datasets without neuroimaging data (n= 141,325, UKB-

134 nonNI) and with neuroimaging data (n=15,050, UKB-NI). At the time of our analyses,

135 brain neuroimaging was ascertained in 42,796 participants, of which 15,050 had

136 complete neuroimaging, mental health and environmental assessments (Supplementary

137 methods).

138 We apply a training and test data split design to investigate the relation of urban

139 living environment with symptoms of mental illness using sparse canonical correlation

140 analysis (sCCA), a multivariate analysis technique to determine multivariate

141 associations between two or more sets of variables. First, we characterise the

142 environmental profiles associated with groups of mental health symptoms using a

143 classical (binary) sCCA. Following a GWAS analysis, we infer a set of genes scores

144 associated with these symptoms groups. Next, using a multi-view sparse CCA

145 (msCCA), we identify the regional brain volumes jointly associated with environment

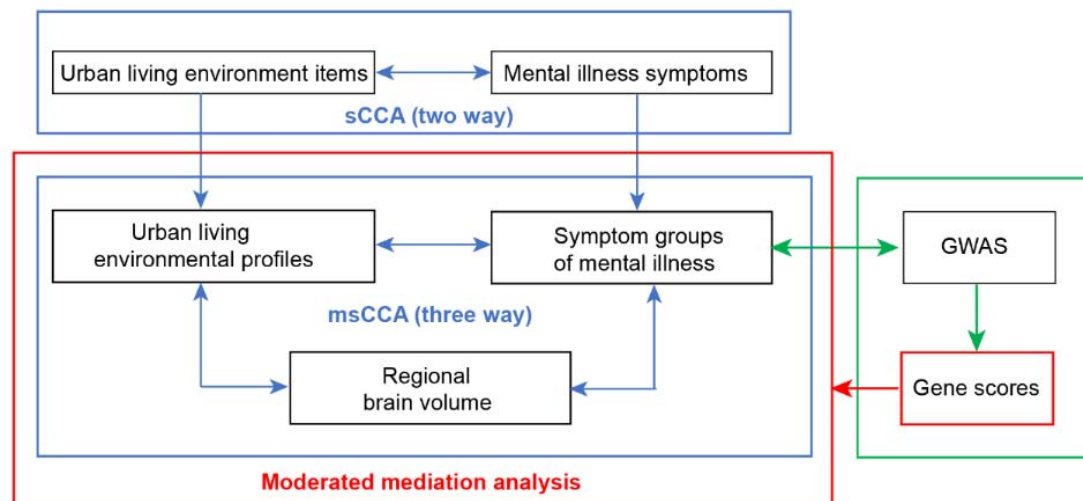
146 and mental health. We then describe the directionality of this relationship using a

147 moderated mediation analysis. Throughout the analysis we were careful to retain a strict

148 training-test separation to avoid bias. Demographic information on the specific

149 statistical analysis is shown in Table 1. Schematic summaries are shown in Figure 1

150 and Supplementary Figure 1.



151

152 **Figure 1: Characterization of the relation of urban living environment profile,**
153 **regional brain volume and mental illness symptoms, and its moderation by genes. In**
154 *141,325 participants without neuroimaging data (UKB-nonNI), we identified*
155 *environmental profiles correlated to symptom groups of emotional problems by*
156 *splitting our data in training and test datasets, and applying bootstrapping with*
157 *replacement and random resampling. Next, we carried out genome-wide association*
158 *analysis (GWAS) analyses of the symptom groups identified in 85,348 participants with*
159 *complete genomic, environment and mental health data from the UKB-nonNI dataset.*
160 *The dataset with neuroimaging data (n=15,050, UKB-NI) was used for independent*
161 *replication of the multivariate relation between urban living environment, genes and*
162 *mental health, as well as for additional neuroimaging analyses. We analysed relations*
163 *between the environmental profiles, regional brain volume and the emotional symptom*
164 *groups applying msCCA. Again, we used a split design with a training and a test dataset.*
165 *Using a moderated mediation analysis, we then investigated the mediation of the effect*
166 *of the environmental profiles on emotional symptom groups by regional brain volume*

167 *and its genetic moderation in 8,726 participants with complete genomic, environment,*
168 *brain volume and mental health data in UKB-NI.*

169

170 **Correlation of urban living environment with symptoms of mental illness.**

171 A total of 53 urban living environment categories including 128 items were
172 included in this study. In the 53 categories, there were 34 categories having one
173 independent item. In the remaining 19 categories, given that some items of each
174 category examined similar aspects of urban environment, ten fold cross-validation
175 confirmatory factor analysis (CFA) was performed to collapse the available information
176 into 19 latent urban living environmental categories using the R package *lavaan*
177 (<https://cran.r-project.org/web/packages/lavaan/>)¹⁸. Ten-fold cross-validation splits
178 were performed to ensure unbiased estimates of generalizability throughout the analytic
179 pipeline and to optimize the CFA models. To investigate the relation of urban living
180 environment with symptoms of mental illness, we used sCCA regression to link 53
181 independent urban living environment categories and 21 prospective symptoms of
182 mental illness (see Online Methods). To avoid overestimating the variance shared
183 between urban living environment and symptoms of mental illness, we used a split data
184 analysis design, which allows us to estimate effect sizes in an unbiased way. We carried
185 out model selection in a training dataset of 90% of the data (n = 127,134), and model
186 validation in the testing dataset of the remaining 10% (n = 14,132) in the 141,326
187 participants of the UKB-nonNI dataset. To enhance stability, we resampled the data

188 and retained only variables that contributed to the model in 90% of resamples (Methods
189 and Supplementary Figure 1)¹⁹.

190 Using msCCA regression, we found a significant relationship between urban
191 living environment profiles and a subset of five symptoms of mental illness in the
192 training dataset ($r=0.22$, $P_{\text{permutation}} < 0.001$), explaining 4.84% of the variance. The
193 model was also significant when applied to the test dataset ($r = 0.19$, $P_{\text{permutation}} < 0.001$,
194 $P_{\text{FDR}} < 0.001$) explaining 3.61% of the variance between environment profiles and
195 symptoms groups of mental illness (Figure 1). These symptoms of mental illness consist
196 of a group of five self-reported symptoms, namely frequency of unenthusiasm,
197 frequency of tiredness, loneliness, frequency of depressed mood and fed-up feelings
198 (Figure 2), which we summarised as the affective symptom group. This symptom group
199 is positively correlated with IMD score, air and noise pollution, measures of street
200 accessibility and neighbourhood walkability (street radial and centrality), traffic and
201 density of urban infrastructures (factories, retails, offices and community services),
202 while being negatively correlated with percentage of domestic garden, natural
203 environment and greenspace, and distance to urban facilities (community services,
204 factories, emergency, education, food stores, community and healthcare)
205 (Supplementary Table 5 and Figure 2). We assessed robustness in two ways, first we
206 used bootstrapping to resample the training data (with replacement) 1000 times, each
207 containing 10% to 150% of the training dataset in 10% increments. Then, we performed
208 a second validation by drawing 1000 resamples of the training dataset, each containing
209 90% of the dataset (Supplementary Table 6 and Figure 2). No specific gender effect

210 was detected (Supplementary Results and Supplementary Table 7). Our results indicate
211 that the affective symptom group is positively correlated with an environment profile
212 dominated by high degrees of social deprivation and air pollution, and to a lesser extent
213 traffic and short distance to infrastructural facilities. Other properties of urbanicity, in
214 particular various forms of green space and social infrastructure appear to be protective
215 in this model.

216 After determining the significance of the first canonical correlate, we removed the
217 effect of the first set of canonical vectors by projection deflation, as we have done
218 previously¹⁹. This approach is more appropriate for regularised CCA than standard
219 deflation methods more commonly used in the field²⁰ because it ensures that the
220 predictions for successive components are orthogonal (i.e. uncorrelated). This is also
221 important to keep in mind in order to correctly interpret our findings (e.g. the GWAS
222 analysis we report below) because the second canonical variables relate to covariance
223 between datasets that is orthogonal to the first set of canonical variables, as is also the
224 case in standard CCA. In other words, the second CCA component explains covariance
225 over and above what is explained by the first component. We then repeated the analysis
226 to investigate the presence of a second canonical correlation between the remaining
227 urban living environment and mental health variables. Here, we identified another
228 mental illness symptoms consisting of six self-report measures, including worrying too
229 long, anxious feelings, nervous feelings, nerves, tense and suffering from nerves
230 (Figure 2), summarised as the anxiety symptom group, which was significantly
231 associated with the environmental profile in the training dataset ($r = 0.15$,

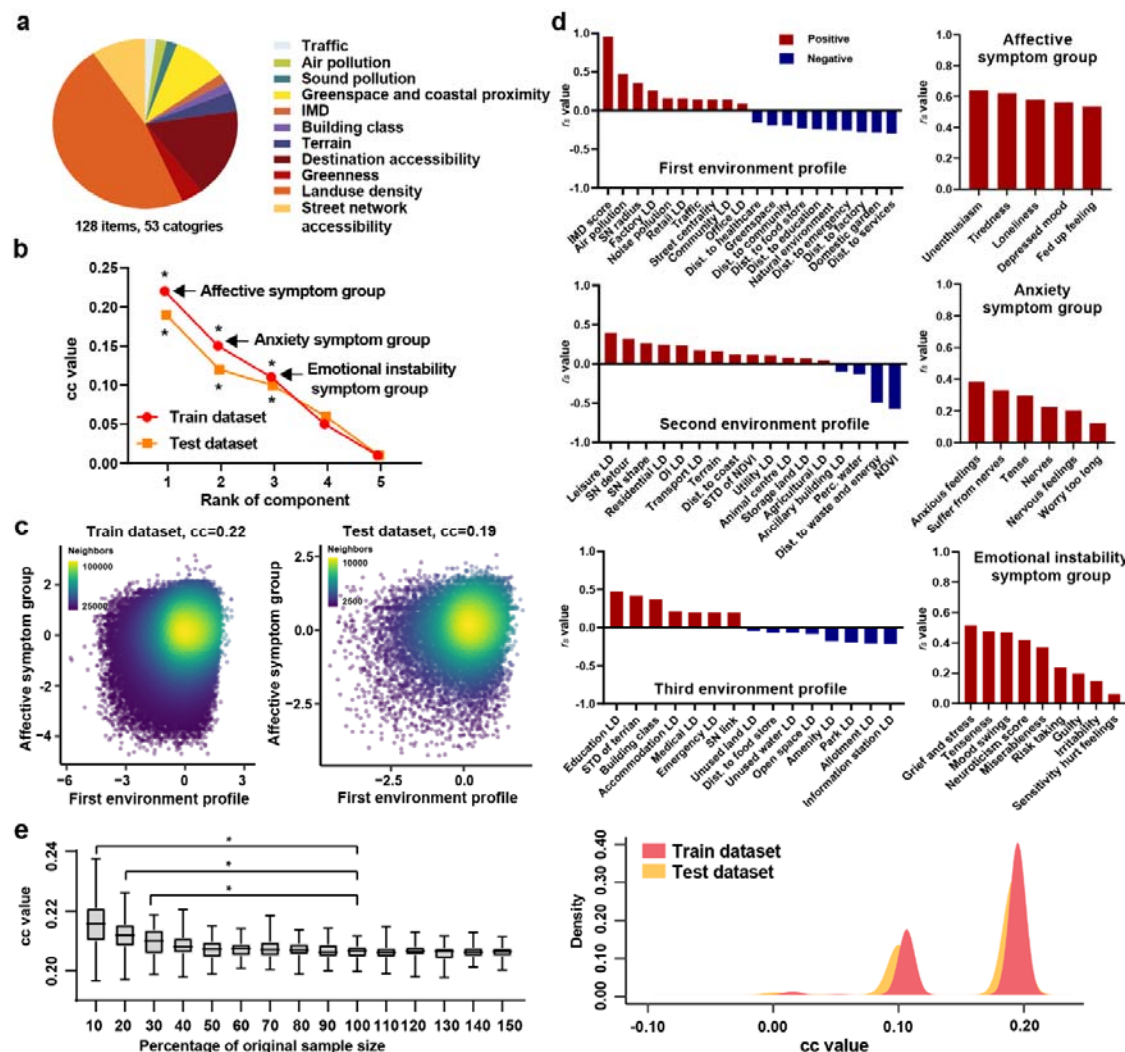
232 $P_{\text{permutation}} < 0.001$). The test sample correlation is also significant ($r = 0.12$,
 233 $P_{\text{permutation}} < 0.001$, $P_{\text{FDR}} < 0.001$), explaining 1.44% of the variance between
 234 environmental profile and anxiety symptom group. This symptom group was positively
 235 correlated with measures of urban morphology, including density of leisure places,
 236 street accessibility and neighbourhood walkability (street network detour and shape)
 237 (Supplementary methods and Supplementary Table 2-3), mean terrain, Euclidean
 238 distance to coast, variation of NDVI and density of mixed urban infrastructure
 239 (residential, transport, utility, animal centre, storage land and agriculture, et.al) while
 240 being negatively correlated with mean NDVI, distance to waste and energy as well as
 241 percentage of water (Supplementary Table 5 and Figure 2). The second environmental
 242 correlate captures a different urban profile, one that is dominated by green spaces and
 243 long distances to waste and energy facilities as well as presence of water, all of which
 244 are inversely correlated (protective) to symptoms of anxiety. The anxiety symptom
 245 group is also positively correlated with signs of denser urban build-up, such as density
 246 of streets and leisure places as well as urban regions with mixed residential, commercial
 247 and industrial use.

248 We removed the effects of the second canonical correlate and investigated the
 249 presence of a third canonical correlation between the remaining urban living
 250 environment and symptoms of mental illness. Here, the correlation coefficient between
 251 two canonical variables was 0.12 in the training dataset ($P_{\text{permutation}} < 0.001$) and 0.10 in
 252 the test dataset ($P_{\text{permutation}} < 0.001$, $P_{\text{FDR}} < 0.001$), explaining variance of 1.44% and
 253 1.00%, respectively. The canonical mental health correlate consists of nine symptoms

including feeling guilty, frequency of tenseness, miserableness, mood swings, neuroticism score, risk taking, irritability and sensitivity, hurt feelings, illness, injury, bereavement or stress in last 2 years (Figure 2), which we termed the emotional instability symptom group. This symptom group is positively correlated with density of education facilities, variation of terrain, building class (flats in highrisers, terraced houses etc.), measures of street accessibility and neighbourhood walkability (street network link characteristics), density of accommodation, medical and emergency facilities, while being negatively correlated with density of unused land, density of water, open space, amenity, park, allotment and information stations as well as distance to food store.

We independently replicated these correlations in the UKB-NI dataset (n=15,050) by applying the same sCCA split data (90%/10%) with a resampling threshold of 90%. The replication analysis yielded three statistically significant canonical correlates in the training and test datasets, which were identical to those of the primary analyses. In the first affective symptoms group, the canonical correlation was 0.20 in the training dataset (n=13,545; $P_{\text{permutation}} < 0.001$) and 0.16 in test dataset (n=1,505; $P_{\text{permutation}} < 0.001$, $P_{\text{FDR}} < 0.001$), in the second anxiety symptoms group, the correlation coefficient was 0.10 in the training dataset ($P_{\text{permutation}} < 0.001$) and 0.08 in the test dataset ($P_{\text{permutation}} < 0.001$, $P_{\text{FDR}} < 0.001$). In the third emotional instability symptoms group the correlation coefficient between two canonical variables was 0.06 in the training dataset ($P_{\text{permutation}} < 0.001$) and 0.04 in the test dataset ($P_{\text{permutation}} < 0.0039$, $P_{\text{FDR}} = 0.0273$) (Supplementary Table 8).

276



277

278 **Figure 2. Multivariate relation between urban living environment profile and**
 279 **symptoms of mental illness.** a. A total of 128 items corresponding to 53 categories of
 280 urban living environment are included; b. The sCCA-regression model linking urban
 281 living environment to symptoms of mental illness identified three significant correlates
 282 in train datasets (red dot), including affective symptoms group ($r=0.22$,
 283 $P_{\text{permutation}} < 0.001$), anxiety symptoms group ($r=0.15$, $P_{\text{permutation}} < 0.001$) and emotional

instability symptoms group ($r=0.12$, $P_{\text{permutation}} < 0.001$). These results were still significant in test datasets of affective ($r=0.19$, $P_{\text{permutation}} < 0.001$), anxiety ($r=0.12$, $P_{\text{permutation}} < 0.001$) and emotional instability symptoms group ($r=0.10$, $P_{\text{permutation}} < 0.001$). c. An example of correlation map between urban environment profile and affective symptoms group in train dataset ($r=0.22$, $P_{\text{permutation}} < 0.001$) and test dataset ($r=0.19$, $P_{\text{permutation}} < 0.001$). d. In the first (top), second (medium) and third (bottom) correlates, urban living environment profiles contributing to this relationship were shown on the left, symptoms group of mental illness contributing to this relationship were shown on the right. The bars demonstrated the structure coefficient value (r_s) of each variable in sCCA-regression analysis. e. Robustness assessment. Left: We used bootstrapping to resample the training data (with replacement) 1000 times, each containing 10% to 150% of the training dataset in 10% increments. Stability in correlation coefficient after about 30% of the sample size were observed; Right: To estimate of the stability of the findings across subsamples, we resampled the same proportion 90% of original sample size as train dataset for 1000 times, reran the sCCA algorithm and calculated the correlation between the resulting feature in the remaining 10% test dataset. CC value, canonical correlation coefficient; IMD, Index of Multiple Deprivation; NDVI, Normalized difference vegetation index; r_s , structure coefficient value; STD, standard deviation; UE, urban living environment.

304 **GWAS analyses of the mental health symptoms groups that related to urban** 305 **environment profile**

306 We next performed a GWAS analysis of the canonical variates of the affective,
307 anxiety and emotional instability symptom groups in 85,348 participants with complete
308 genetic, urban environment and mental health data in UKB-nonNI datasets (Table 1).
309 Gene-set enrichment analysis was then performed to explore biological mechanism
310 underlying the genes associated with the symptoms groups using TopGene²¹. To reduce
311 dimensionality, we generated scores for each of the genes where significant SNPs were
312 localised (see Supplementary methods). The score of each gene is calculated as the sum
313 of the count of risk alleles multiplied by the corresponding beta value from GWAS
314 across the index SNPs of each clump after adjusting for linkage disequilibrium (see
315 Supplementary methods). These gene scores were then analysed for moderation of the
316 relation of environment, regional brain volume and symptoms group of mental illness
317 (see below).

318 For the affective symptom group, we found 2,983 significant associations with
319 SNPs at Bonferroni $P < 0.05$ (uncorrected $P < 0.05/139,187,27/3 = 1.20 \times 10^{-9}$,
320 Supplementary Table 9), located in 43 genes. The lambda genomic control inflation
321 factor was 1.0195 and the intercept value in linkage disequilibrium score regression
322 (LDSC) was 1.011, indicating that population stratification was minimal in this study
323 and did not cause significant inflation to the test statistics. By far the strongest
324 association with the affective symptom group was found in a genomic region of

chromosome 17q21.3 spanning from position 43487217 to 44862162 (hg19), the site of a human supergene candidate that encodes several genes previously implicated in mental illness (Figure 3)²². The lead SNP was rs62062288, located in intron 6 of microtubule associated protein tau (*MAPT*) gene of chromosome 17 ($P=6.09 \times 10^{-15}$), a gene that encodes Tau protein in neurons and has been shown to be involved in affective symptoms²³, alcohol disorders²⁴ as well as risk taking behaviour²⁵. In the same region of chromosome 17q21.3, we also found strong association of the affective symptom group with Corticotrophin Release Hormone Receptor 1 (*CRHR1*, a critical regulatory gene for neuroendocrinological and behavioural stress response²⁶ as well as alcohol use disorders^{27,28}). The remaining top associated genes were also encoded in this region. They include *ARL17B* (ADP ribosylation factor-like GTPase17B, potentially regulating G-protein signalling²⁹), *KANSL1* (KAT8 regulatory NSL complex subunit1, part of a histone acetyltransferase complex, an epigenetic regulator³⁰), and *MAPT-AS1* (*MAPT* antisense RNA 1, involved in progression of tau pathology in neurodegeneration³¹), as well as a pseudogene *RP11-259G18* (Figure 3 and Supplementary Table 9). Additional associations were found on genomic region of chromosome 18q21.2, at the *DCC* gene locus (DCC Netrin 1 Receptor, regulates synaptic plasticity³² and neuronal migration³³) and *TCF4* gene locus (transcription factor 4, involved in neural differentiation³⁴), chromosome 14q24.1 (*DCAF5* and *EXD2* gene locus) and chromosome 3q22 (*STAG1* and a pseudogene *RP11-102M11.2*). Bioinformatics analyses of the 43 genes using TopGene revealed enrichment of neuronal differentiation, regulation of axogenesis and, importantly, CRH-binding and activity (Supplementary Table 12). We then calculated

43 genes scores. The relationship between urban living environment profile and affective symptom group have shown different vulnerabilities depending on the genetic background. For example, we found that participants with lower *CRHR1* gene scores demonstrate smaller correlation of urban environment profile with the affective symptom group compared to those with higher *CRHR1* genes scores (Figure 3).

We found significant associations of the anxiety symptom group with 29 SNPs covering 11 genes at Bonferroni $P < 0.05$ (Supplementary Table 10 and Supplementary Figure 2). The drop in genome-wide significant hits compared to the affective symptom group GWAS is likely caused by the decreased covariance of the second CCA correlate, the anxiety symptom group, following deflation of the correlates of the first CCA correlate, the affective symptom group, as described above. The lead SNP associated with the anxiety symptom group was rs77641763 located in the intron 15 of Exonuclease 3'-5' Domain Containing 3 (*EXD3*) gene of chromosome 9 ($P = 4.08 \times 10^{-10}$). *EXD3*, involved in nucleic acid binding and widely expressed in the brain, has been associated with a group of anxiety, phobic and dissociative disorders³⁵. The other top significant genes include *NOLCI* (Nucleolar And Coiled-Body Phosphoprotein 1), a chaperone for shuttling between the nucleolus and cytoplasm^{36,37}, *ELOVL3* (Elongation of very long chain fatty acids protein 3, provides precursors for synthesis of sphingolipids and ceramides³⁸), *LBX1-AS1* (Ladybird homeobox 1- antisense RNA 1, involved in neuronal determination processes³⁹), *PITX3* (Paired Like Homeodomain 3, regulates differentiation and maintenance of midbrain dopamine neurons during development⁴⁰), *GBF1* (Golgi Brefeldin A Resistant Guanine Nucleotide Exchange

Factor 1, involved in axonal neuropathy⁴¹), *CNNM2* (Cyclin And CBS Domain Divalent Metal Cation Transport Mediator 2, linked to neurodevelopmental impairments⁴²) and *NT5C2* (cytosolic 5' -nucleotidase II, regulating AMPK signalling and protein translation during early neurodevelopment⁴³) (Supplementary Table 10). The 11 genes associated with the anxiety symptom group were enriched for small nucleolar ribonucleoprotein (snoRNP) complex binding involved in serotonin receptor regulation (Supplementary Table 12). And participants with lower *EXD3* gene scores demonstrate smaller correlation of urban environment profile with the anxiety symptom group compared to those with higher *EXD3* genes scores (Figure 3).

In the emotional instability symptom group, there were 6 significant genetic associations at Bonferroni $P < 0.05$ (Supplementary Table 11 and Supplementary Figure 2). The lead SNP was rs77786116 located in Intraflagellar Transport 74 (*IFT74*) gene of chromosome 9 ($P = 4.16 \times 10^{-10}$). *ITF74* is a critical factor in neuronal migration, a cause for Bardet-Biedl syndrome and Joubert syndrome^{44,45} and is associated with paranoid schizophrenia⁴⁶. The other significant genes include *TMPO* (thymopoietin, involves in neuron proliferation), *LDHC* (lactate dehydrogenase C, involved in anaerobic and aerobic glycolysis), *SLC9A7PI* (solute carrier family 9 member 7 pseudogene 1). Together, they are enriched for cerebellar granular layer development process (Supplementary Table 12).

We independently replicated the SNPs that in the discovery GWAS were significantly associated with symptom groups of mental illness (UKB-nonNI dataset)

in a dataset of 8,726 participants of UKB-NI. The significance threshold was Bonferroni $P < 0.05$ (uncorrected $P < 0.05/3018$, the numbers of all significant SNPs from discovery GWAS analysis). Of these 3018 significant SNPs from GWAS analysis of mental illness symptom groups, we replicated in the independent UKB-NI dataset 2034 SNPs associated with the affective symptom group, 18 SNPs associated with the anxiety symptom group and 3 SNPs associated with the emotional instability symptom group (Supplementary Table 13-15). We then calculated the corresponding genes scores as before and validated the associations between genes scores and symptoms group of mental illness in the UKB-NI dataset. Of 43 genes scores associated with affective symptom group in the discovery analysis, we replicated 36 genes in the replication analysis, the top ten associated genes including *MAPT*, *CRHR1*, *DCC*, *DNDIP1*, *KANSL1*, *PLEKHM1*, *RP11-363J20.1:EXD2*, *STAG1*, *C5orf17* and *DCAF5*. Of 11 genes scores in the anxiety symptom group, we replicated 6 genes, including *EXD3*, *LBX1-AS1*, *NT5C2*, *GBF1*, *NOLC1* and *CNNM2*. Of 6 genes scores associated with emotional instability symptom, we replicated 3 genes including *LDHC*, *IFT74* and *TMPO* (Supplementary Table 15).

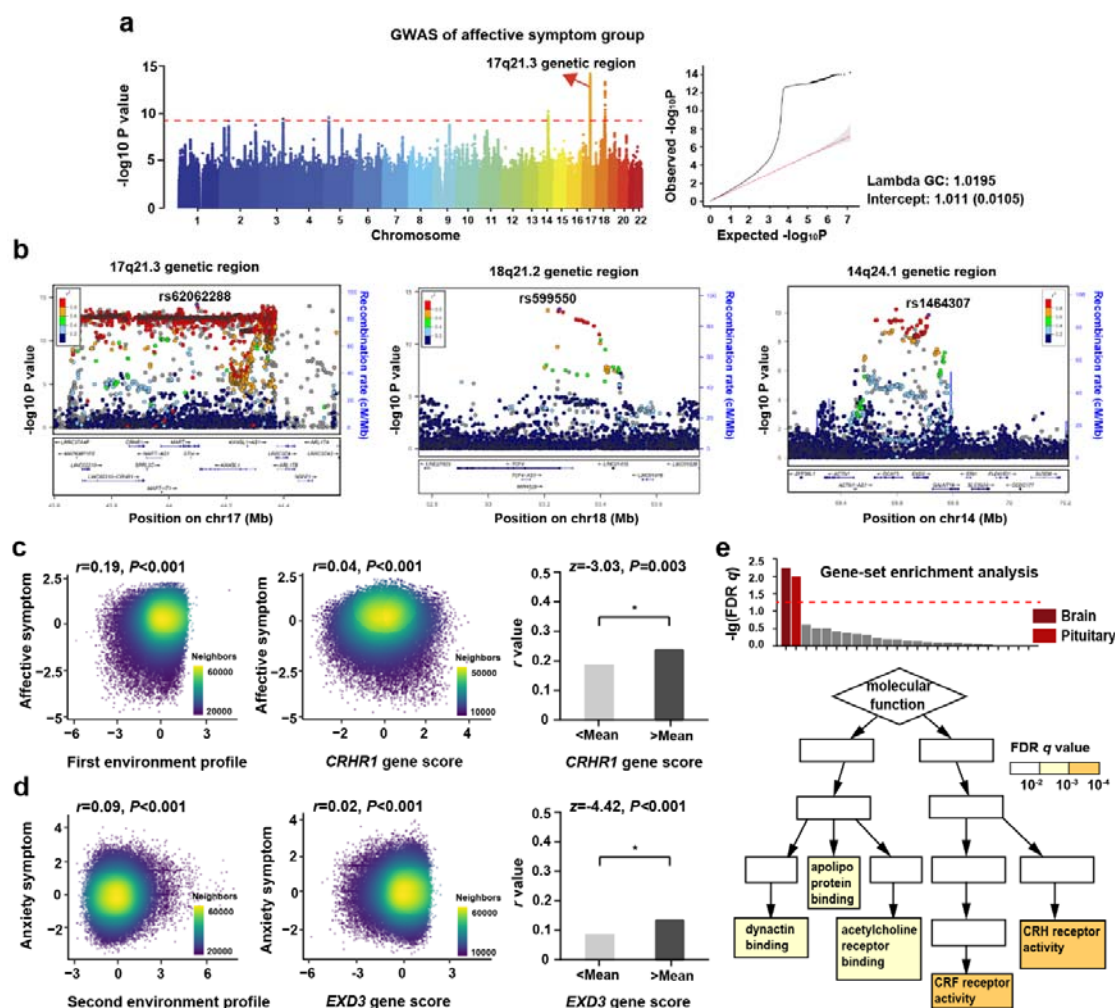


Figure 3. Associations of genomics, urban living environment profiles and mental health symptoms. *a. Left. Manhattan plot of GWAS analysis. GWAS of the affective symptom group identifies 2,983 significantly associated SNPs. The strongest association with the affective symptom group was found in a genomic region of chromosome 17q21.3 spanning from position 43487217 to 44862162 (hg19) containing a potential supergene that encodes several genes previously implicated in mental illness. An intron variant rs62062288 located in MAPT gene of chromosome 17 ($P=6.09 \times 10^{-15}$) was the lead SNP; Right. Quantile-quantile plot of GWAS analysis.*

415 The lambda genomic control inflation factor of 1.0195 and intercept value in LDSC of
 416 1.011, which indicate that population stratification was minimal in this study to cause
 417 significant inflation to the test statistics. b. Locus zoom plots of significant genomic
 418 region of 17q21.3 (left), 18q21.2 (medium) and 14q24.1 (right) in GWAS analysis of
 419 affective symptom group. The purple dots demonstrate the lead SNPs of each genomic
 420 region. c-d. Two examples of the affective symptom group influenced by both genomics
 421 and urban living environment profile. c. Left. The first urban living environment profile
 422 showed significant positive correlation with the affective symptom group; Medium.
 423 Participants with higher CRHR1 gene scores show more severe affective symptoms
 424 compared to those with lower CRHR1 gene scores; Right: Participants with lower
 425 CRHR1 gene scores demonstrate smaller correlation of urban environment profile with
 426 the affective symptom group compared to those with higher CRHR1 genes scores. d.
 427 Left. The second urban living environment profile showed significant positive
 428 correlation with the anxiety symptom group; Medium. Participants with higher EXD3
 429 gene scores show more severe anxiety symptoms compared to those with lower EXD3
 430 gene scores; Right: Participants with lower EXD3 gene scores demonstrate smaller
 431 correlation of urban environment profile with the anxiety symptom group compared to
 432 those with higher EXD3 genes scores. e. Gene-set enrichment analysis of affective
 433 symptom group-associated 43 genes. Top: Tissue-specific expression analysis showed
 434 affective symptom group associated 43 genes were significantly overexpressed in
 435 brain and pituitary tissues. Bottom: In the 43 genes associated with the symptom group,
 436 we found over-representation in the molecular function of CRH receptor activity (FDR

437 $q_c = 4.37 \times 10^{-4}$) and CRF receptor activity ($FDR\ q_c = 4.37 \times 10^{-4}$); CRH,
 438 corticotropin-releasing hormone; CRF, corticotropin-releasing factor; GWAS,
 439 genome-wide association analysis; lambda GC, lambda genomic control inflation
 440 factor.

441

442 **Brain volume changes underlying the associations between urban living** 443 **environment and symptom groups of mental illness**

444 To further investigate the neurobiological mechanisms underlying the associations
 445 of urban living environment with mental symptoms groups, we carried out msCCA on
 446 the urban living environment profiles, brain volume measures and symptoms group of
 447 mental illness. This analysis was conducted in an independent dataset of 15,050 UKB
 448 participants, split in a training dataset (90%) and a test dataset (10%). We found 13
 449 regional brain volumes significantly associated with the first urban environmental
 450 profile and the affective symptom group ($r=0.052$, $P_{\text{permutaiton}}=0.01$ in training dataset;
 451 $r=0.046$, $P_{\text{permutaiton}}=0.01$ in test dataset). These brain volumes include the left amygdala
 452 and right ventral striatum, right frontal pole, right occipital fusiform gyrus, as well as
 453 bilateral cerebellar lobules VIIa and VIIb, right posterior cerebellum crus I and II
 454 (Figure 4 and Supplementary Table 17). The first urban environment profile was
 455 negatively correlated with brain volume in these areas and positively correlated with
 456 affective symptoms group. We also found 11 regional brain volumes significantly
 457 associated with the second urban environmental profile and the anxiety symptom group

($r=0.045$, $P_{\text{permutation}}=0.006$ in training dataset; $r=0.045$, $P_{\text{permutation}}=0.007$ in test dataset) (Supplementary Table 17). These brain volumes include the inferior frontal regions and the right amygdala, as well as the bilateral cerebellar lobules VIIIa and VIIb, posterior cerebellum crus I, right cerebellar lobule V and left lobule VI (Figure 4 and Supplementary Table 17). Finally, there were 13 regional brain volumes, including bilateral frontal pole, amygdala, precentral gyrus, insular and left lateral occipital cortex, associated with the third urban environment profile and the emotional instability symptom group ($r=0.057$, $P_{\text{permutation}}<0.001$ in the training dataset; $r=0.060$, $P_{\text{permutation}}<0.001$ in test dataset) (Figure 4 and Supplementary Table 17).

467

468 **Moderated mediation analysis between urban living environment profile, brain** 469 **volume and mental illness symptoms groups modulated by genomics**

470 To test whether the relation of urban living environment with symptoms group of
471 mental illness is mediated by brain volume and moderated by genetics, we
472 independently performed moderated mediation analysis in each replicated gene score
473 (moderating variable), three urban living environment profiles (independent variable),
474 three brain volume correlates (mediated variable) and three symptoms groups of mental
475 illness (dependent variable) in 8,726 participants with complete data. Explained
476 mediation effect (EME) were reported for moderated mediation analyses. Thus, a total
477 of 21 moderated mediations analysis were tested (13 gene scores of the affective
478 symptom group, 6 gene scores of the anxiety symptom group and 2 gene scores of the

emotional instability symptom group). Of the 13 replicated gene scores in the affective symptom group, we found that the *CRHR1* gene score (EME=5.01%), *MAPT* gene score (EME=3.62%), *TCF4* gene score (EME=1.62%) and *DCC* gene score (EME=1.96%) moderate the mediation pathway from environment profile to brain correlate to affective symptom group. For example, participants with higher *CRHR1* genetic risk living in areas of more urban environment exposure had lower brain volume and demonstrated more severe affective symptoms (Figure 4). We found moderation of the mediation pathway of the anxiety symptom group by the *EXD3* gene score (EME=3.65%) and of the emotional instability symptom group by the *IFT74* gene score (EME=2.86%) (Figure 4).

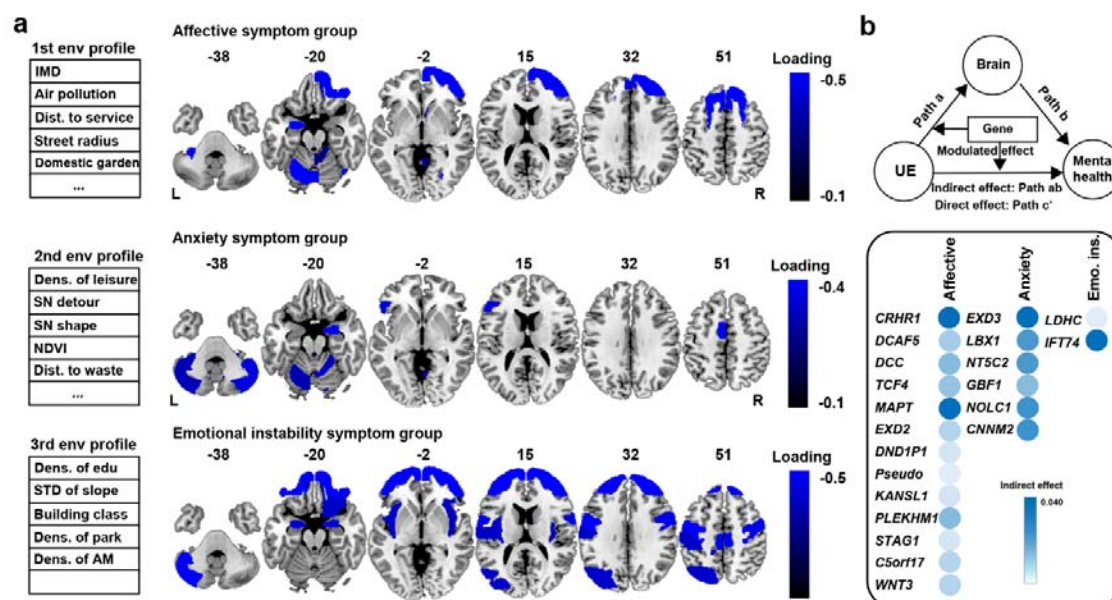


Figure 4. Three-way msCCA regression between urban living environmental profile, regional brain volume and mental illness symptom group, and moderated mediation

493 **analysis of gene scores.** *a. Left. Urban living environment variables contributing to the*
494 *three way correlates are shown on the left. Right. Regional brain volume associated with*
495 *urban living environment profile and affective (top), anxiety (medium) and emotional*
496 *instability (down) symptom groups, respectively, in both training and test datasets.*
497 *Negative loading was shown in blue. b. Top. A schematic diagram of moderated*
498 *mediation analysis between genomics, urban living environment profile, brain and*
499 *mental illness symptom group. Bottom. Columns show an indirect effect in the*
500 *mediation analysis between environmental profile, brain volume and mental illness*
501 *symptom groups, which were moderated by each gene score. We found that the CRHR1*
502 *gene score (EME=5.01%), MAPT gene score (EME=3.62%), TCF4 gene score*
503 *(EME=1.62%) and DCC gene score (EME=1.96%) moderate the mediation pathway*
504 *from environmental profile to brain correlate to the affective symptom group. The*
505 *EXD3 gene moderates the mediation pathway from environmental profile to brain*
506 *correlate to the anxiety symptom group (EME=3.65%). IFT74 gene score moderates*
507 *the mediation pathway from environment profile to brain correlate to emotional*
508 *instability symptom group (EME=2.86%). AM, allotment. EME, explained mediation*
509 *effect; IMD, Index of Multiple Deprivation; NDVI, normalized difference vegetation*
510 *index; SN, street network measures; STD, standard deviation.*

511

512 **Discussion**

513 We describe how the urban living environment affects brain and mental illness
514 by identifying environmental profiles derived from a comprehensive set of measures of
515 deprivation, pollution, land use, and infrastructure that are correlated with distinct
516 groups of affective, anxiety and emotional instability symptoms, mediated by
517 reductions in regional brain volume and moderated by genes involved in pertinent
518 biological pathways.

519 Our approach and findings are novel, as they characterise specific multimodal
520 environmental profiles in an integrated way, while enabling a qualitative and
521 quantitative assessment of each factor of the profile. This is an advance beyond the
522 isolated assessment of individual environmental factors, as has previously been the
523 norm^{9,47}. As a consequence, we are able to explain a greater degree of variance of
524 integrated environmental factors (3.61%) than comparable studies measuring
525 individual environmental factors, such as night-time light (2.56%), built-up areas
526 (1.21%) and NDVI (1.00%)⁸, alone. Furthermore, the integrated approach enables us
527 to assess the role of each individual environmental factor in a context that is relevant
528 for mental illness. In another innovation, we describe how the effect of environmental
529 profiles on symptoms groups of mental illness is mediated by regional brain volume
530 and moderated by genetic factors.

531 By providing evidence for the brain-related correlates of environmental
532 adversity and their consequences for mental health, we are broadening the evidence
533 required for responsible urban and development planning. Furthermore, we are

534 enabling the development of neurobehavioural interventions that may convey adaptive
535 coping skills for environmental adversity by targeting specific brain mechanisms, for
536 example through neurofeedback-guided virtual reality sessions^{48,49}. The identification
537 of genetic moderators points towards biological pathways that might underlie the
538 observed relation between environment, brain and behaviour. It may also enable the
539 selection of individuals that are more sensitive to environmental adversity, and thus
540 might be more likely to benefit from targeted interventions.

541 Specifically, our analyses reveal the first environmental profile dominated by
542 high degrees of deprivation and air pollution, and to a lesser extent traffic, short distance
543 to infrastructural facilities and lack of green space. This environmental profile evokes
544 the image of a poor, dense inner city neighbourhood. It is correlated with increased
545 affective symptoms, in particular high levels of unenthusiasm, tiredness, loneliness, as
546 well as depressed mood and feelings of being fed-up. The correlation is mediated by
547 volume reductions in brain regions linked to emotional processing, such as the left
548 amygdala⁵⁰, regions linked to reward processing and exploration, such as the right
549 ventral striatum⁵¹ and the right frontal pole⁵², as well as other brain areas implicated in
550 affective processing, such as several cerebellar regions⁵³ and the right occipital fusiform
551 gyrus⁵⁴. While these findings point towards plausible mediators of a stressful
552 environment on affective symptoms, they also suggest the presence of different
553 underlying neurobehavioural mechanisms, a possibility that is supported by the
554 findings of our genetic analysis. Here we identified different moderators that may
555 influence distinct brain mechanisms underlying the affective symptoms identified: First

and foremost, *CRHRI*, which is also expressed in the amygdala⁵⁵, is a critical regulator of both the hypothalamic as well as the behavioural extrahypothalamic stress response. In addition, the environmental effect is moderated by genes regulating brain structure, including *MAPT*, involved in neurodegeneration⁵⁶, *TCF4*, inducing neural differentiation⁵⁷ and *DCC*, an adhesion molecule that guides axon growth⁵⁸. Apart from these moderators, we identified several genes that were associated with the affective component and were involved in relevant neural mechanisms, including G-protein signalling (*ARL17B*)⁵⁹, epigenetic regulation (*KANSL1*)⁶⁰ and others. These moderating genes are located in two genomic loci of chromosome 17q21.3 and 18q21.2. Notably, the chromosome 17q21.3 genomic locus is the site of a human supergene candidate, a cluster of tightly linked functional genetic elements spanning approximately 900 kb that control balanced phenotypes and are inherited as a unit⁶¹. Haplotypes of this cluster have been associated with brain morphology and different cognitive and behavioural traits, including depressive behaviour, neuroticism and risk taking behaviour^{62,63}. The 18q21.2 region at the transcription factor 4 gene *TCF4* and netrin 1 receptor gene *DCC* featured the most pleiotropic association with eight psychiatric disorders⁶⁴, and has been previously associated with both depression and neuroticism^{65,66}. The product of *DCC* plays a critical role in guiding axonal growth during neurodevelopment and serves as a master regulator of midline crossing and white matter projections⁶⁷.

The second environmental profile captures a different urban profile, one that is dominated by green spaces and long distances to waste and energy facilities as well as presence of water, all of which are inversely correlated (protective) to symptoms of

578 anxiety. These symptoms are ‘worry too long, anxious feelings, nervous feelings,
579 nerves, tense and suffer from nerves’. The anxiety symptoms are also positively
580 correlated with signs of denser urban build-up, such as density of streets and leisure
581 places as well as urban regions with mixed residential, commercial and industrial use.
582 Together, these correlations point to an important role of green spaces and a less dense
583 and more generous land use in protecting against symptoms of anxiety, thus extending
584 previous findings relating urban green spaces to mental health⁸. The relation between
585 environmental profiles and anxiety symptoms are mediated by volume reductions in the
586 inferior frontal regions, the right amygdala and cerebellar regions. They are moderated
587 by variations in the *EXD3* gene, involved in nucleic acid binding and widely expressed
588 in the brain. *EXD3* has previously been associated with anxiety, phobic and dissociative
589 disorders³⁵. While *EXD3* has a low tissue specificity, its expression in the brain is
590 highest in frontal cortical brain areas. Similar to the second environmental profile, the
591 third environmental profile shows positive correlations of measures of density of land
592 use and urban infrastructure with a group of symptoms best summarised as emotional
593 instability, which include feeling guilty, frequency of tenseness, miserableness, mood
594 swings, neuroticism score, risk taking, irritability and sensitivity, hurt feelings, illness,
595 injury, bereavement or stress in last 2 years. These symptoms were mediated by frontal
596 pole, amygdala, precentral gyrus, insular cortex and cerebellum, and moderated by
597 *ITF74*, a critical factor in neuronal migration, a cause for Bardet-Biedl syndrome and
598 Joubert syndrome^{44,45}, which has been found associated with paranoid schizophrenia⁴⁶.

599 While our findings differentiate distinct groups of environmentally-related
600 symptoms of mental illness, the overarching role of social deprivation⁶⁸ and urban
601 density⁶⁹ as a risk factor is evident. Conversely, we found green spaces^{9,10} and generous
602 land use⁷⁰ to be protective factors against affective and anxiety symptoms. These
603 environmental factors may engage brain mechanisms related to behavioural inhibition
604 (frontal cortex), reward processing (ventral striatum) and emotional processing
605 (amygdala) that may be moderated by biological pathways involved in stress response,
606 neuronal plasticity, epigenetic, transcriptional and neurotransmitter regulation.
607 Interestingly, the posterior cerebellar lobule VIIb and VIIIa, posterior cerebellar crus I
608 are among the brain mediators of environmental effects on symptoms of mental illness,
609 supporting a role for the cerebellum in coordinating and regulating not only voluntary
610 movements, but also cognitive and affective perceptions of the environment^{8,71}.

611 Our data do not enable the characterisation of individual biological pathways
612 that mediate defined environmental adversity. To carry out such mechanistic
613 investigation and to identify biomarkers for risk and resilience, a more deeply
614 phenotyped dataset with neurobehavioural characterisation is required. Our findings
615 generate hypotheses, that may be tested in well characterised samples of a much smaller
616 size. Another important question raised by this study is the generalizability of its results
617 beyond industrialised high income countries to low and medium income countries,
618 where socioeconomic conditions between rural and urban dwellers may be different
619 compared to Europe. Further studies validating our results in a global mental health
620 context are required.

621 By providing first evidence for comprehensive environmental profiles that
622 affect distinct groups of symptoms of mental illness, our results lay the ground work
623 for an exact characterisation of biological mechanisms underlying complex, real-life
624 environmental adversity. The quantification of the relative contribution of each
625 environmental factor to brain and symptoms of mental illness and their interplay in a
626 living environment are novelties of this study that will aid in targeting and prioritizing
627 important decisions for planning and public health interventions.

628

629 **Material and methods**

630 **Participants**

631 **UK Biobank Project**

632 UK Biobank (UKBB) is a population-based cohort including 502,616 participants
633 living in United Kingdom, where over 40,000 participants have obtained neuroimaging
634 scans (at the time of our analyses datasets of 42,796 participants were available).
635 Individuals were invited to the study if they were registered with the National Health
636 Service and if they lived within a 35 km radius of one of the 22 assessment centres
637 located across the UK at the time of recruitment, which took place between 2007 and
638 2010. Baseline assessments included genomics, physical and social exposures,
639 sociodemographic, lifestyle, occupational, psychosocial and environmental measures.
640 Informed consent was obtained from all UKBB participants. Ethical procedures are
641 controlled by a dedicated Ethics and Guidance Council

(<http://www.ukbiobank.ac.uk/ethics>) that has developed with UKBB an Ethics and Governance Framework (<http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>), with IRB approval also obtained from the North West Multi-center Research Ethics Committee. The data collected at baseline was used in this study. Demographic information of each statistical analysis was shown in Table 1.

648

649 **Data collection**

650 **Urban living environment data**

651 Indices of multiple deprivation (IMD), traffic, residential pollution, greenspace
652 and coastal proximity as well as urban morphometric measures were used to measure
653 the urban living social and physical environment around participants available in the
654 category ‘local environment’ in UK Biobank (data-field 113). A total of 53 categories
655 including 128 urban living environment items were included in the urban living
656 environment data. The detailed items and categories used are shown in Supplementary
657 Table 1 and 2. To exclude items with extremely skewed data distribution, we used the
658 function *nearZeroVar* from the *caret* R package⁷² and no items excluded. In the 128
659 items, we calculated the median absolute deviation (MAD) and removed values larger
660 than 4 MAD in each environment item. For further analyses, we used 216,341
661 participants with complete 128 environmental items.

662

663 ***Indices of Multiple Deprivation***

664 Indices of Multiple Deprivation (IMD) scores were used to classify the relative
665 deprivation (a measure of poverty) in British local councils which published by UK
666 government ([https://www.gov.uk/government/collections/english-indices-of-](https://www.gov.uk/government/collections/english-indices-of-deprivation)
667 deprivation). IMD scores were calculated separately in England (EIMD), Scotland
668 (SIMD) and Wales (WIMD), because multiple different components of deprivation are
669 weighted with different strengths and compiled into a single score of deprivation. The
670 EIMD score is composed of seven domain indices, including income deprivation
671 (Income subdomain, Income Deprivation Affecting Children Index and Older People
672 Index), employment deprivation, health deprivation and disability, education, skills and
673 training deprivation (children and young people subdomain and adults' skills
674 subdomain), barriers to housing and services (wider and geographical barriers
675 subdomain), living environment deprivation (indoors and outdoors subdomain) and
676 crime. The SIMD score is composed of seven domain indices, including crime (only
677 from 2006), current income, education, skills and training, employment, geographic
678 access, health and housing. The WIMD score is composed of eight domain indices for
679 income, employment, health, education, access to services, community safety, physical
680 environment and housing.

681 ***Traffic***

682 Traffic consists of seven items: (1) Close to major road; (2) Inverse distance to the
683 nearest major road; (3) Inverse distance to the nearest road; (4) Sum of road length of

684 major roads within 100m; (5) Total traffic load on major roads; (6) Traffic intensity on
685 the nearest major road; (7) Traffic intensity on the nearest road.

686 ***Residential pollution***

687 Residential air pollution consists of five items: (1) Nitrogen dioxide air pollution
688 from 2005 to 2010; (2) Nitrogen oxides air pollution in 2010; (3) Particulate matter 10
689 um air pollution in 2007 and 2010; (4) Particulate matter 2.5 um air pollution in 2010;
690 (5) Particulate matter 2.5-10um air pollution in 2010. Residential sound pollution
691 consists of five items: (1) Average 16-hour sound level of noise pollution; (2) Average
692 24-hour sound level of noise pollution; (3) Average daytime sound level of noise
693 pollution; (4) Average evening sound level of noise pollution; (5) Average nighttime
694 sound level of noise pollution.

695 ***Greenspace and coastal proximity***

696 Greenspace and coastal proximity category contains environmental indicators
697 relating to greenspace exposure and distance to the coast that was attributed to
698 participants based on 300m home location buffers, including five items: (1) Natural
699 environment percentage estimate compared to the 'built environment'; (2) Greenspace
700 percentage estimates; (3) Domestic garden percentage estimates; (4) Domestic water
701 percentage estimates; (5) Distance from home location to coast.

702 ***Urban Morphometric Platform measures***

703 UK Biobank Urban Morphometric Platform (UKBump) was an individual-level
704 built environment database of health-specific urban exposures within residential street
705 catchments of UK Biobank participants' geocoded home address⁷³. Spatial and network
706 modelling were performed upon multiple UK-wide dataset, including AddressBase
707 Premium data of Ordnance Survey GB, remotely sensing data, digital terrain
708 topographical models and other datasets based on the anonymized UK Biobank
709 participants' geocoded home address⁷³.

710 A total of six metrics including 104 urban environment measures from UKBump
711 were used. The six metrics used in this study include: (1) Building class (n=3); (2)
712 Destination accessibility (n=33); (3) Greenness (n=2); (4) Land use density (n=46); (5)
713 Street network accessibility based on 400m home location buffers (n=18)
714 (Supplementary Table 3); (6) Terrain (slope) (n=2).

715

716 **Genomics data**

717 We used the imputed genomic data (Version 3) made available by UK Biobank
718 with 487,411 individuals⁷⁴, which was imputed from the Haplotype Reference
719 Consortium (HRC) reference panel⁷⁵ and a merged UK10K and 1000 Genomes phase
720 3 reference panels⁷⁶. In participants-level quality control, we applied exclusion filters
721 for participants as follows: (1) participants with a mismatch in reported sex and
722 chromosome X imputed sex or with putative sex chromosome aneuploidy; (2)
723 participants with genetic kinship to other participants; (3) excess heterozygosity or
724 missing rates; (4) non-Caucasian participants; (5) without calculated genetic principal

725 components. In SNPs-level quality control, we applied exclusion filters for SNPs as
726 follows: (1) minor allele frequency (MAF) < 0.001; (2) imputation info quality score >
727 0.3. A total of 275,988 participants and 139,187,27 SNPs were finally used in the
728 further analysis.

729

730 **Mental health measures**

731 There are 44 mental health items in the category ‘Mental health’ in the UK
732 Biobank that cover symptoms of affective and anxiety disorders, as well as personality
733 (category id:100060). These items were obtained from a standardised mental health
734 questionnaire that participants answered at the time of recruitment. Of this
735 questionnaire 21 items were excluded because the missing rate was larger than 50%
736 from 502,616 participants of UK Biobank. To exclude mental health items with
737 extremely skewed data distribution, we used the function *nearZeroVar* from the *caret*
738 R package⁷² and excluded 2 items. Finally, a total of 21 mental health items with
739 complete data in 365,201 participants were included in the further analysis. A full list
740 of the 21 mental health measures is shown in Supplementary Table 4.

741

742 **Neuroimaging data**

743 In this study, neuroimaging data were acquired from one 3.0-Tesla MRI scanner
744 from Siemens® Skyra running VD13A SP4 with a standard 32-channel radiofrequency
745 receive head coil at UK Biobank imaging center in Cheadle Manchester. The standard

parameters of a 3D MPAGE sequence are shown in
https://biobank.ctsuo.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf.

A total of 154 regional image-derived phenotypes (IDPs) were investigated: 111 cortical and subcortical gray matter volume (GMV) segmentations from the FAST segmentations (data-field 1101), 28 cerebellum GMV segmentations from the FAST segmentations (data-field 1101), and 15 subcortical volumes from the FIRST (data-field 1102). The detailed information are shown in
https://biobank.ctsuo.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf. A total of 42,796 participants' neuroimaging data were used for the present study.

Confounding variables

Age, gender and assessment centers were adjusted as confounding covariates in the further analysis. The 21 mental health variables were firstly corrected for confounding variables and normalized as well. For neuroimaging-related analyses, total intracranial volume (TIV) was additionally corrected.

Statistical analysis

Train and test sample split design

Participants from UK Biobank with complete urban living environmental data and mental health data (n=156,375) were divided into datasets without neuroimaging data

(UKB non-NI dataset, n=141,325) and with neuroimaging data (UKB NI dataset, n=15,050). Participants without neuroimaging data (n=141,325) were divided into train and test dataset to ensure validity of the results: 90% of the participants were used as a training dataset (n=127,193) and 10% of the participants (n=14,132) were used as a test dataset for model validation. The dataset with neuroimaging data (n=15,050) was used for independent replication of the relation between urban living environment, genes and mental health, as well as for additional neuroimaging analyses. (Table 1).

773

774 **Table 1. Demographics of participants used in specific statistical analysis.**

Analysis	Required data	Sample size (n)	Age (years)	Gender (Male/Female)
CFA and sCCA	E, MH	141,325	59.22(8.22)	65,729/75,596
GWAS	G, MH, E	85,348	59.37(7.98)	40,966/44,382
msCCA	E, B, MH	15,050	57.75(7.50)	7,305/7,745
Modulated mediation	G, E, B, MH	8,726	58.03(7.42)	4,294/4,432

775 *B, brain volume data; CFA, confirmatory factor analysis; E, urban living environment data; G, genomic*
776 *data; GWAS, genome-wide analysis; MH, mental health data; msCCA, multiple sparse canonical*
777 *correlation analysis; sCCA, sparse canonical correlation analysis.*

778

779 **Urban living environmental variables construction**

780 In the CFA models, ten-fold cross-validation was performed to ensure unbiased
781 estimates of generalizability throughout the analytic pipeline and to optimize the CFA
782 models. For each fold, 90% of participants were used to build the CFA model, and the

783 optimized CFA model was used to calculate latent variables for the remaining 10% of
 784 participants in each environment subcategory. We used two criteria to optimize the
 785 CFA model by selecting appropriate environment measures. The first criterion was the
 786 goodness of fit of the CFA model assessed by Tucker-Lewis index (TLI), Comparative
 787 Fit Index (CFI), chi square, root mean square error of approximation (RMSEA) and
 788 standard root mean square residual (SRMR). The criteria for excellent model fit were
 789 TLI>0.95, CFI>0.95, RMSEA<0.06 and SRMR<0.08⁵⁷⁻⁵⁹. The second criterion was the
 790 inclusion of environment measures that best reflect different aspects of urban
 791 environment. For example, in residential noise pollution variables, we initially
 792 constructed a CFA model by including all five noise pollution measures in the training
 793 dataset. Based on factor loadings of the five noise pollution measures, we removed the
 794 ‘Average night-time sound level of noise pollution’ item with the smallest factor
 795 loading and repeated the CFA modelling. These steps were iterated until the resulting
 796 CFA model satisfied our criteria for excellent model fit in the training dataset. The
 797 factor loadings of the optimized CFA model were used to calculate the latent residential
 798 noise pollution measure in the test dataset. This process was applied into 10 folds to
 799 predict all out-of sample 19 environmental variables.

800

801 **Multivariate relation of urban living environment with mental health**

802 To investigate the multivariate relation between urban living environment and
 803 mental health, we conducted multivariate analyses using sparse canonical correlation

804 analysis (sCCA) implemented in MATLAB (version R2018a) based on our previous
805 work¹⁹. In detail, the analysis design was carried out as follows:

806 (1) The full dataset was randomly split into training and test datasets. The training
807 dataset was made up of 90% of the data whilst the testing set was made up of the
808 remaining 10%.

809 (2) The training dataset was then randomly split into 100 resamples. Each resample
810 was made up of $nt/2$ participant scans, where nt is the total number of participants in
811 the training dataset.

812 (3) The first stage of the msCCA-regression algorithm was then applied to each
813 resample, with a sparsity constraint of 0.5 in each view of the data¹⁹.

814 (4) The resulting weights for each environment and mental health variable were
815 recorded for each resample. The environment and mental health variables with non-
816 zero loading above 90% across the resamples were selected and retained as stable
817 variables in subsequent analyses.

818 (5) We then re-applied the sCCA algorithm to the data, without sparsity constraints,
819 on the stable urban living environment and mental health measures in the training
820 dataset. The canonical correlation value between urban living environment and mental
821 health measures were recorded.

822 (6) We then permuted the training data, and repeated steps (2)-(5). This was done
823 for 10,000 different permutations of the training data labelling. In each case, we

824 recorded the canonical correlation value between urban living environment and mental
825 health measures. In this way, we built up a permutation distribution to assess the
826 significance of the relationship between urban living environment and mental health
827 measures in the experimental labelling, within the training dataset.

828 (7) We then applied the trained model to the test dataset to produce canonical
829 correlates of urban living environment and mental health measures. We recorded
830 associations for training and testing dataset.

831 (8) We then randomly permuted the data rows in the test dataset and recalculated
832 correlation values between urban living environment and mental health canonical
833 correlates. We recorded associations between urban living environment and mental
834 health correlates, for each of 10,000 permutations of the experimental labelling. False
835 Discovery Rate (FDR) correction was used to control for multiple testings and a
836 $P_{FDR} < 0.05$ was considered statistically significant.

837 (9) After determining the significance of the first canonical correlate, we remove
838 the effect of the first set of canonical vectors by projection deflation as we did before¹⁹,
839 and repeat the sCCA analysis between the remaining urban living environment and
840 mental health variables. These steps were iterated until the resulting canonical correlate
841 were not significant any more.

842 (10) To ensure reliability and reproducibility, we undertook further analyses
843 including: a. sample composition using bootstrapping with replacement and random
844 resampling; b. the effect of sex (Supplemental Methods).

845 Finally, we calculated the canonical correlations coefficient between the urban
846 living environment canonical variables (which we refer to as ‘urban environmental
847 profile’) and mental health canonical variables (referred to as ‘mental illness symptom
848 group’), as well as structure coefficient value (r_s) of individual variables

849

850 **GWAS analyses of the mental health symptoms groups that related to urban** 851 **environment profile**

852 For the significant mental illness symptom group that correlated with urban living
853 environment profile from sCCA results, we conducted a genome-wide association
854 analysis (GWAS) of the corresponding mental illness symptom group in 85,348
855 participants with complete genomic, urban environment and mental health data (Table
856 1). Using BGENIE v1.2 (<https://jmarchini.org/bgenie/>), we fit an additive model of
857 association at each variant, using expected genotype count (dosage) from the imputed
858 genetic data. The covariates included age, gender, assessment centre, processing batch
859 and the top 10 ancestry principal components. Bonferroni $P_c < 0.05$ (uncorrected
860 $P < 0.05/139,187,27 \times$ numbers of significant mental illness symptom group from sCCA
861 results) was considered as a statistically significant threshold. We employed the FUMA
862 online platform (<https://fuma.ctglab.nl/>) to fine mapping to genes and perform
863 annotation of significant SNPs of each GWAS analysis. All SNPs with genome-wide
864 significance were mapped to genes based on physical distance. Each SNP was mapped
865 to the nearest gene in the human reference assembly (GRCh37/hg19).

866 **Gene-set enrichment analysis**

867 To better understand the biological function of fine-mapping genes associated with
868 mental health components, these genes were functionally annotated using ToppGene
869 portal (<https://toppgene.cchmc.org/>) to identify significant enrichments for gene
870 ontology (GO). The Benjamini and Hochberg method for false discovery rate (FDR-
871 BH correction) ($q_c < 0.05$) was applied to correct for multiple comparisons. The default
872 full reference gene list of each category in ToppGene were used as background gene
873 set.

875 **Replication analysis**

876 To replicate the multivariate relation between urban living environment and
877 mental health, we applied the sCCA analysis in an independent dataset of 15,050
878 participants with complete environmental, mental health and neuroimaging data from
879 UKB-NI dataset. Again, we used a training dataset (n=13,545, 90%) and a test dataset
880 (n=1,505, 10%), a re-sampling method to ensure variable stability (with a threshold of
881 90% for non-zero weights from re-sampled data to consider as stable variables) and
882 permutation tests to assess the significance of the results (10,000 times) as we used in
883 the discovery sCCA analysis. Next, we independently replicated the significant SNPs
884 associated with symptom groups of mental illness surviving from the discovery GWAS
885 analysis (UKB-nonNI dataset) in an independent 8,726 participants of UKB-NI dataset
886 at Bonferroni $P_c < 0.05$ (uncorrected $P < 0.05$ / the numbers of all significant SNPs of

887 GWAS from discovery analysis). Then we calculated the corresponding genes scores
888 as the same way we did in the discovery analysis. And finally we independently
889 validated the associations between genes scores and symptoms group of mental illness
890 in the UKB-NI dataset.

891

892 **Brain volume changes underlying the associations between urban living** 893 **environment and mental health**

894 To investigate neurobiological mechanisms underlying the associations between
895 urban living environment and mental health, we carried out a multiple sparse canonical
896 correlation analysis (msCCA) between the environmental component, brain volume
897 measures and the mental health component. This analysis was conducted in an
898 independent sample of 15,050 participants with environmental, mental health and
899 neuroimaging data from UK Biobank. Again, we used a training dataset (n=13,545,
900 90%) and a test dataset (n=1,505, 10%), a re-sampling method to ensure variable
901 stability (with a threshold of 85% for non-zero weights from re-sampled data to
902 consider as stable variables) and permutation tests to assess the significance of the
903 results (10,000 times). An in-house Matlab script based on our previous work¹⁹ was
904 employed for this analysis. Finally, the brain volume canonical variables (referred to as
905 brain component), the canonical correlations coefficient between environment, brain
906 and mental health canonical variables, structural loadings and weights of corresponding
907 brain volume variables were calculated.

908 **Moderated mediation analysis between urban living enviroment, brain volume** 909 **and mental health correlates modulated by genomics**

910 The above analysis was carried out separately to identify the associations of mental
911 health symptoms with urban living environment, genetic variation and brain volume,
912 leaving the complex associations of urban living, genomics, brain and mental health
913 unexplored. To formally test whether urban living environment-mental health
914 relationship can be mediated by brain volume and modulated by genetics, we carried
915 out a modulated mediation analysis in 8,726 participants. Moderated mediation analysis
916 is an extension of mediation analysis, which is a valuable technique for assessing
917 whether an indirect effect is conditional on a moderating variable. The basis of
918 moderation and mediation effect were integrated into a combined model of moderated
919 mediation within a linear regression framework. Finally, genomics was defined as
920 modulated variables, the urban living environment correlate was defined as an
921 independent variable, the brain volume correlates a mediator variable and the mental
922 health correlates a dependent variable.

923 In moderated mediation analyses, all indirect effects are estimated in one multiple
924 regression analysis with independent variable and all mediators as predictor variables.
925 This means that the indirect effect of one mediator was estimated when the other
926 mediators are taken into account. We used bootstrapping to assess the significance of
927 the mediation effect. After 5,000 bias-corrected bootstrapping, we estimated the
928 distribution of the indirect effect and calculate its 95% confidence intervals (CI). If zero

929 does not fall between the resulting 95% confidence interval of the bootstrapping method,
 930 we confirmed the existence of a significant mediation effect ($P < 0.05$). It should be
 931 emphasized that in the multiple mediation analysis of this study, mediators and
 932 dependent variables were measured contemporaneously, thus not allowing
 933 establishment of any causal directionality. Explained mediation effect (EME) and 95%
 934 CI were reported for moderated mediation analyses. Confounding factors were
 935 controlled in the moderated mediation model. The moderated mediation analysis with
 936 a nonparametric bootstrap method was conducted using the *mediation* R package⁷⁷.

937

938 References

- 939 1. Heilig, G.K. World urbanization prospects: the 2011 revision. *United Nations,*
 940 *Department of Economic and Social Affairs (DESA), Population Division,*
 941 *Population Estimates and Projections Section, New York* **14**(2012).
 942
- 943 2. Callaghan, A. *et al.* The impact of green spaces on mental health in urban
 944 settings: a scoping review. *J Ment Health* **30**, 179-193 (2021).
 945
- 946 3. Morgan, N. & Mall, S. Pathways between urbanization and harmful substance
 947 use. *Curr Opin Psychiatry* **32**, 218-223 (2019).
- 948 4. Lederbogen, F., Haddad, L. & Meyer-Lindenberg, A. Urban social stress-risk
 949 factor for mental disorders. The case of schizophrenia. *Environ Pollut* **183**, 2-6
 950 (2013).
 951
- 952 5. Peen, J. *et al.* The current status of urban-rural differences in psychiatric
 953 disorders. *Acta Psychiatr Scand* **121**, 84-93 (2010).
 954
- 955 6. Gruebner, O. *et al.* Cities and mental health. *Dtsch Arztebl Int* **114**, 121 (2017).
 956
- 957 7. Lederbogen, F. *et al.* City living and urban upbringing affect neural social stress
 958 processing in humans. *Nature* **474**, 498-501 (2011).
 959

- 960 8. Xu, J. *et al.* Global urbanicity is associated with brain and behaviour in young
961 people. *Nat Hum Behav* **6**, 279-293 (2021).
962
- 963 9. Crouse, D.L. *et al.* Residential greenness and indicators of stress and mental
964 well-being in a Canadian national-level survey. *Environ Res* **192**, 110267
965 (2021).
966
- 967 10. Tost, H. *et al.* Neural correlates of individual differences in affective benefit of
968 real-life urban green space exposure. *Nat Neurosci* **22**, 1389-1393 (2019).
969
- 970 11. Kumar, S., Kroon, J., Laloo, R. A systematic review of the impact of parental
971 socio-economic status and home environment characteristics on children's oral
972 health related quality of life. *Health Qual Life Outcomes* **12**, 1-15 (2014).
973
- 974 12. Abrahamyan Empson, L. *et al.* Urbanicity: The need for new avenues to explore
975 the link between urban living and psychosis. *Early Interv Psychiatry* **14**, 398-
976 409 (2020).
977
- 978 13. Gruebner, O. *et al.* Cities and Mental Health. *Dtsch Arztebl Int* **114**, 121-127
979 (2017).
980
- 981 14. Okkels, N. *et al.* Urban mental health: challenges and perspectives. *Curr Opin*
982 *Psychiatry* **31**, 258-264 (2018).
983
- 984 15. Pluess, M. Individual differences in environmental sensitivity. *Child Dev.*
985 *Perspect* **9**, 138-143 (2015).
986
- 987 16. Uher, R. & Zwickler, A. Etiology in psychiatry: embracing the reality of poly-
988 gene-environmental causation of mental illness. *World Psychiatry* **16**, 121-129
989 (2017).
990
- 991 17. Tost, H., Champagne, F.A. & Meyer-Lindenberg, A.. Environmental influence
992 in the brain, human welfare and mental health. *Nat Neurosci* **18**, 1421-1431
993 (2015).
994
- 995 18. Rosseel, Y. Lavaan: An R package for structural equation modeling and more.
996 Version 0.5–12 (BETA). *J Stat Softw* **48**, 1-36 (2012).
997
- 998 19. Ing, A. *et al.* Identification of neurobehavioural symptom groups based on
999 shared brain mechanisms. *Nat Hum Behav* **3**, 1306-1318 (2019).
1000
- 1001 20. Monteiro, J.M. *et al.* A multiple hold-out framework for Sparse Partial Least
1002 Squares. *J Neurosci Methods* **271**, 182-94 (2016).
1003

- 1004 21. Chen, J. *et al.* ToppGene Suite for gene list enrichment analysis and candidate
1005 gene prioritization. *Nucleic Acids Res* **37**, W305-11 (2009).
1006
- 1007 22. Ward, J. *et al.* The genomic basis of mood instability: identification of 46 loci
1008 in 363,705 UK Biobank participants, genetic correlation with psychiatric
1009 disorders, and association with gene expression and function. *Mol Psychiatry*
1010 **25**, 3091-3099 (2020).
1011
- 1012 23. Nagel, M. *et al.* Item-level analyses reveal genetic heterogeneity in neuroticism.
1013 *Nat Commun* **9**, 905 (2018).
1014
- 1015 24. Sanchez-Roige, S. *et al.* Genome-wide association study meta-analysis of the
1016 Alcohol Use Disorders Identification Test (AUDIT) in two population-based
1017 cohorts. *Am J Psychiatry* **176**, 107-118 (2019).
1018
- 1019 25. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance
1020 and risky behaviors in over 1 million individuals identify hundreds of loci and
1021 shared genetic influences. *Nat Genet* **51**, 245-257 (2019).
1022
- 1023 26. Normann, C. & Buttenschøn, H.N. Gene-environment interactions between
1024 HPA-axis genes and stressful life events in depression: a systematic review.
1025 *Acta Neuropsychiatr* **31**, 186-192 (2019).
1026
- 1027 27. Blomeyer, D. *et al.* Interaction between CRHR1 gene and stressful life events
1028 predicts adolescent heavy alcohol use. *Biol Psychiatry* **63**, 146-51 (2008).
1029
- 1030 28. Treutlein, J. *et al.* Genetic association of the human corticotropin releasing
1031 hormone receptor 1 (CRHR1) with binge drinking and alcohol intake patterns
1032 in two independent samples. *Mol Psychiatry* **11**, 594-602 (2006).
1033
- 1034 29. Falola, M.I. *et al.* Genomic copy number variants: evidence for association with
1035 antibody response to anthrax vaccine adsorbed. *PLoS One* **8**, e64813 (2013).
1036
- 1037 30. León, L.E. *et al.* Partial microduplication in the histone acetyltransferase
1038 complex member KANSL1 is associated with congenital heart defects in 22q11.
1039 2 microdeletion syndrome patients. *Sci Rep* **7**, 1-8 (2017).
1040
- 1041 31. Simone, R. *et al.* MIR-NATs repress MAPT translation and aid proteostasis in
1042 neurodegeneration. *Nature* **594**, 117-123 (2021).
1043
- 1044 32. Horn, Katherine E. *et al.* DCC Expression by Neurons Regulates Synaptic
1045 Plasticity in the Adult Brain. *Cell Reports* **3**, 173-185 (2013).
1046

- 1047 33. Junge, H.J. *et al.* Netrin1/DCC signaling promotes neuronal migration in the
1048 dorsal spinal cord. *Neural Development* **11**, 19 (2016).
1049
- 1050 34. Forrest, M.P. *et al.* The emerging roles of TCF4 in disease and development.
1051 *Trends Mol Med* **20**, 322-331 (2014).
1052
- 1053 35. Stein, M.B. *et al.* Genomic Characterization of Posttraumatic Stress Disorder in
1054 a Large US Military Veteran Sample. *bioRxiv*, 764001 (2019).
1055
- 1056 36. Meier, U.T. & Blobel, G. Nopp 140 shuttles on tracks between nucleolus and
1057 cytoplasm. *Cell* **70**, 127-138 (1992).
1058
- 1059 37. Meier, U.T. & Blobel, G. A nuclear localization signal binding protein in the
1060 nucleolus. *J Cell Biol* **111**, 2235-2245 (1990).
1061
- 1062 38. Guillou, H. *et al.* The key roles of elongases and desaturases in mammalian fatty
1063 acid metabolism: Insights from transgenic mice. *Prog Lipid Res* **49**, 186-199
1064 (2010).
1065
- 1066 39. Luo, M. *et al.* The Susceptibility and Potential Functions of the LBX1 Gene in
1067 Adolescent Idiopathic Scoliosis. *Front Genet* 1802 (2021).
1068
- 1069 40. Chung, S. *et al.* The homeodomain transcription factor Pitx3 facilitates
1070 differentiation of mouse embryonic stem cells into AHD2-expressing
1071 dopaminergic neurons. *Mol Cell Neurosci* **28**, 241-252 (2005).
1072
- 1073 41. Mendoza-Ferreira, N. *et al.* De Novo and inherited variants in GBF1 are
1074 associated with axonal neuropathy caused by golgi fragmentation. *Am J Hum*
1075 *Genet* **107**, 763-777 (2020).
1076
- 1077 42. Arjona, F. *et al.* CNNM2 Mutations Cause Impaired Brain Development and
1078 Seizures in patients with hypomagnesemia. *PLoS Genet* **10**, e1004267 (2014).
1079
- 1080 43. Duarte, R.R. *et al.* The psychiatric risk gene NT5C2 regulates adenosine
1081 monophosphate-activated protein kinase signaling and protein translation in
1082 human neural progenitor cells. *Biol Psychiatry* **86**, 120-130 (2019).
1083
- 1084 44. Lindstrand, A. *et al.* Copy-number variation contributes to the mutational load
1085 of Bardet-Biedl syndrome. *Am J Hum Genet* **99**, 318-336 (2016).
1086
- 1087 45. Kleinendorst, L. *et al.* Second case of Bardet-Biedl syndrome caused by
1088 biallelic variants in IFT74. *Eur J Hum Genet* **28**, 943-946 (2020).
1089

- 1090 46. Thakurela, S. *et al.* Mapping gene regulatory circuitry of Pax6 during
1091 neurogenesis. *Cell Discov* **2**, 15045 (2016).
1092
- 1093 47. Guxens, M. *et al.* Air Pollution Exposure During Fetal Life, Brain Morphology,
1094 and Cognitive Function in School-Age Children. *Biol Psychiatry* **84**, 295-303
1095 (2018).
1096
- 1097 48. Subramanian, L. *et al.* Functional magnetic resonance imaging neurofeedback-
1098 guided motor imagery training and motor training for Parkinson's disease:
1099 randomized trial. *Front Behav Neurosci* **10**, 111 (2016).
1100
- 1101 49. Tarrant, J., Viczko, J. & Cope, H. Virtual reality for anxiety reduction
1102 demonstrated by quantitative EEG: a pilot study. *Front Psychol* 1280 (2018).
1103
- 1104 50. Janak, P.H. & Tye, K.M. From circuits to behaviour in the amygdala. *Nature*
1105 **517**, 284-92 (2015).
1106
- 1107 51. Burton, A.C., Nakamura, K. & Roesch, M.R. From ventral-medial to dorsal-
1108 lateral striatum: neural correlates of reward-guided decision-making. *Neurobiol*
1109 *Learn Mem* **117**, 51-9 (2015).
1110
- 1111 52. Mansouri, F.A. *et al.* Behavioral consequences of selective damage to frontal
1112 pole and posterior cingulate cortices. *Proc Natl Acad Sci U S A* **112**, E3940-
1113 E3949 (2015).
1114
- 1115 53. De Smet, H.J. *et al.* The cerebellum: its role in language and related cognitive
1116 and affective functions. *Brain Lang* **127**, 334-42 (2013).
1117
- 1118 54. Gong, J. *et al.* Common and distinct patterns of intrinsic brain activity
1119 alterations in major depression and bipolar disorder: voxel-based meta-analysis.
1120 *Transl Psychiatry* **10**, 353 (2020).
1121
- 1122 55. Savarese, A. & Lasek, A.W. Regulation of anxiety-like behavior and Crhr1
1123 expression in the basolateral amygdala by LMO3. *Psychoneuroendocrinology*
1124 **92**, 13-20 (2018).
1125
- 1126 56. Strang, K.H., Golde, T.E. & Giasson, B.I. MAPT mutations, tauopathy, and
1127 mechanisms of neurodegeneration. *Lab Invest* **99**, 912-928 (2019).
1128
- 1129 57. Shariq, M. *et al.* Adult neural stem cells have latent inflammatory potential that
1130 is kept suppressed by Tcf4 to facilitate adult neurogenesis. *Sci Adv* **7**(2021).
1131
- 1132 58. Duman-Scheel, M. Netrin and DCC: axon guidance regulators at the
1133 intersection of nervous system development and cancer. *Curr Drug Targets* **10**,
602-10 (2009).

- 1134
- 1135 59. Vargová, R. *et al.* A eukaryote-wide perspective on the diversity and evolution
- 1136 of the ARF GTPase protein family. *Genome Biol Evol* **13**, evab157 (2021).
- 1137
- 1138 60. Meunier, S. *et al.* An epigenetic regulator emerges as microtubule minus-end
- 1139 binding and stabilizing factor in mitosis. *Nat Commun* **6**, 1-10 (2015).
- 1140
- 1141 61. Campoy, E. *et al.* Genomic architecture and functional effects of potential
- 1142 human inversion supergenes. *Philos Trans R Soc Lond B Biol Sci* **377**,
- 1143 20210209 (2022).
- 1144
- 1145 62. Adams, H.H. *et al.* Novel genetic loci underlying human intracranial volume
- 1146 identified through genome-wide association. *Nat Neurosci* **19**, 1569-1582
- 1147 (2016).
- 1148
- 1149 63. Grasby, K.L. *et al.* The genetic architecture of the human cerebral cortex.
- 1150 *Science* **367**, eaay6690 (2020).
- 1151
- 1152 64. Lee, P.H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic
- 1153 Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11
- 1154 (2019).
- 1155
- 1156 65. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary
- 1157 statistics using MTAG. *Nat Genet* **50**, 229-237 (2018).
- 1158
- 1159 66. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants
- 1160 and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681
- 1161 (2018).
- 1162
- 1163 67. Bendriem, R.M. & Ross, M.E. Wiring the human brain: A user's handbook.
- 1164 *Neuron* **95**, 482-485 (2017).
- 1165 68. Orben, A. *et al.* A. The effects of social deprivation on adolescent development
- 1166 and mental health. *Lancet Child Adolesc Health* **4**, 634-640 (2020).
- 1167
- 1168 69. Jackson, L.E. The relationship of urban design to human health and condition.
- 1169 *Landsc Urban Plan* **64**, 191-200 (2003).
- 1170
- 1171 70. Dannenberg, A.L. *et al.* The impact of community design and land-use choices
- 1172 on public health: a scientific research agenda. *Am J Public Health* **93**, 1500-
- 1173 1508 (2003).
- 1174
- 1175 71. Stoodley, C.J. & Schmahmann, J.D. Evidence for topographic organization in
- 1176 the cerebellum of motor control versus cognitive and affective processing.
- 1177 *Cortex* **46**, 831-844 (2010).

1178
1179 72. Kuhn, M. *et al.* Package ‘caret’. *The R Journal* **223**, 7 (2020).
1180
1181 73. Sarkar, C., Webster, C. & Gallacher, J. UK Biobank Urban Morphometric
1182 Platform (UKBUMP)—a nationwide resource for evidence-based healthy city
1183 planning and public health interventions. *Ann. GIS* **21**, 135-148 (2015).
1184
1185 74. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic
1186 data. *Nature* **562**, 203-209 (2018).
1187
1188 75. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype
1189 imputation. *Nat Genet* **48**, 1279-83 (2016).
1190
1191 76. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using
1192 the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
1193
1194 77. Tingley, D. *et al.* Mediation: R package for causal mediation analysis. *J Stat*
1195 *Softw* (2014).

1196

1197

1198

1199