Spatial fine-mapping for gene-by-environment effects identifies risk hot spots for schizophrenia

Chun Chieh Fan1,2, John J. McGrath3,4,5, Vivek Appadurai2,6, Alfonso Buil2,6, Michael J. Gandal7, Andrew J. Schork2,6, Preben Bo Mortensen3,6,8, Esben Agerbo3,6,8, Sandy A. Geschwind9, Daniel Geschwind7,10, Thomas Werge2,6,11,12, Wesley K. Thompson2,6,13 & Carsten Bøcker Pedersen3,6,8,14

Spatial mapping is a promising strategy to investigate the mechanisms underlying the incidence of psychosis. We analyzed a case-cohort study (n = 24,028), drawn from the 1.47 million Danish persons born between 1981 and 2005, using a novel framework for decomposing the geospatial risk for schizophrenia based on locale of upbringing and polygenic scores. Upbringing in a high environmental risk locale increases the risk for schizophrenia by 122%. Individuals living in a high gene-by-environmental risk locale have a 78% increased risk compared to those who have the same genetic liability but live in a low-risk locale. Effects of specific locales vary substantially within the most densely populated city of Denmark, with hazard ratios ranging from 0.26 to 9.26 for environment and from 0.20 to 5.95 for gene-by-environment. These findings indicate the critical synergism of gene and environment on the etiology of schizophrenia and demonstrate the potential of incorporating geolocation in genetic studies.

1 Center for Human Development, University of California, San Diego, CA 92093, USA. 2 Mental Health Center Sct. Hans, Capital Region of Denmark, Roskilde 4000, Denmark. 3 National Centre for Register-based Research, Aarhus University, Aarhus 8210, Denmark. 4 Queensland Brain Institute, University of Queensland, St. Lucia, QLD 4072, Australia. 5 Queensland Centre for Mental Health Research, Wacol, QLD 4076, Australia. 6 The Lundbeck Foundation Initiative for Integrative Psychiatric Research, IPSYCH, Aarhus and Copenhagen, Denmark. 7 Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA 90095, USA. 8 Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus 8210, Denmark. 9 Scientific Decision Consulting, Santa Monica, CA 90401, USA. 10 Department of Neurology, University of California, Los Angeles, CA 90095, USA. 11 Department of Clinical Sciences, University of Copenhagen, Copenhagen 2200, Denmark. 12 Institute of Biological Psychiatry, Mental Health Services of Copenhagen, Copenhagen 4000, Denmark. 13 Family Medicine and Public Health Division of Biostatistics, University of California, San Diego, CA 92093, USA. 14 Big data Centre for Environment and Health, Aarhus University, Aarhus 8210, Denmark. These authors contributed equally: Wesley K. Thompson, Carsten Bøcker Pedersen. Correspondence and requests for materials should be addressed to W.K.T. (email: wes.stat@gmail.com) or to C.B.P. (email: cbp@econ.au.dk)
or public mental health, it is critical to know which environmental factors can be modified to mitigate the risk of psychiatric disorders. However, identifying modifiable environmental factors has been a contentious issue, especially when the effects may depend on one’s genetic liability for illness. Take as an example one of the best-established environmental risks for schizophrenia, childhood upbringing in an urban area. Persons born and raised in urban areas have an approximately twofold increased risk of schizophrenia compared to those born and raised in rural areas. Researchers have examined potentially causal elements of urban upbringing, such as accessibility to health care, selective migration of individuals, air-pollution, infections, and socioeconomic inequality. Yet none of these factors have substantially explained the risk associated with urbanicity, nor are they highly correlated with instruments used in defining urbanicity, such as population density. The conditional relationships between genetic liabilities and putative environmental factors are even harder to detect despite several cohort studies suggesting an interaction between urban upbringing and family history of schizophrenia.

The difficulty in isolating specific environmental risk elements underlying urbanicity effects on schizophrenia incidence exemplifies a serious methodological challenge. The process for discovering environmental risk factors typically relies on a hypothesis-driven “candidate environmental factor” approach. Researchers need to formulate a carefully constructed environmental hypothesis, measure it, and then determine if it associates with risk of the disease. Analyses is usually performed in a study of selected participants not necessarily representative of the entire population of interest. Similar to the candidate gene approach before the dawning of genome-wide association studies, the candidate environment approach suffers from the “spotlight effect”, ignoring the likely complexity of many environmental factors interacting with each other and with genetic liabilities to determine overall risk for illness. The environmental impact can even be a joint holistic effects from multiple environmental factors. Measurement of the specific environmental factor may also be imprecise, masking its relationship to the illness. For example, many instruments have been devised to characterize socioeconomic inequality, yet have not shown consistent effects on incidence of schizophrenia. Given the complexity of real-life socioeconomic forces, lack of association with schizophrenia could be caused by instrument measurement error or because the instrument does not capture the relevant socioeconomic factors.

An alternative to the candidate environment approach is to assess spatial patterns of disease risk without directly measuring environmental factors. As with John Snow isolating the environmental source of cholera outbreak via mapping the cases, identifying spatially localized disease “hot spots” can assist in the discovery of latent environmental factors. Advanced methods for disease mapping have been developed within the field of geostatistics, particularly in applying spatial random effect models to infer latent environmental variation in causal risk factors. As the urbanicity-related increase in risk for schizophrenia was first noted through spatial clustering of disease incidence, inferring risk hot spots to a finer resolution may provide insight into potential risk-modulating environmental elements before investing substantial resources in active measurement.

With this concept in mind, we develop a disease mapping strategy to address the need for discovering environmental factors without direct measurement. We use spatial random effects to map the geographic distribution of genetic liabilities (G), locale of upbringing (E), and their synergistic effects (GxE) on disease risk. By treating E and GxE as “latent random fields” on the map of Denmark, we avoid methodological issues inherent in the candidate environment approach. Although several studies have utilized random effect models to examine spatially localized risk for schizophrenia, our method differs by utilizing spatial fine-mapping and enabling the partition of risk into E and GxE components without the need for candidate environmental factors.

As a proof of concept, we examine geospatial variation in schizophrenia risk across Denmark. To do so, we apply this novel analytical approach to data from a population-based case-cohort study that includes subject genotyping and detailed residential information from birth up to age 7 years. We are thus able to assess locale of upbringing effects on schizophrenia risk with a resolution beyond conventionally defined levels of urbanicity, allowing us to assess variation in spatial risk, and to ask whether spatially localized environmental factors modulate genetic liability of risk for schizophrenia.

**Results**

**Spatial distribution of overall risk of schizophrenia.** We utilize the entire population cohort of iPYSCH, excluding cases, to derive locales. The resulting map contains 186 non-overlapping locales, with the number of cohort members ranging from 65 to 197 individuals in each locale (median = 105). Figure 1 displays the risk ratio (RR) from the Mantel-Haenszel analyses. With the exception of the southwestern portion of Denmark, the majority of rural regions have lower risk ratios while high-risk locales are concentrated in large cities (Fig. 1a). By plotting RR’s against the size of each locale, Fig. 1b demonstrates a general trend for spatial risks of schizophrenia, meaning locales with higher population density tend to have higher RR’s. Thus, the risk distribution recapitulates the known urbanicity effects. However, there is substantial variation in risk even controlling for locale size; for example, RR’s can range from protective to highly detrimental within densely populated areas (Fig. 1b).

The contribution of the E and GxE. Table 1 shows the estimations from multilevel models. Compared to rural regions, being born and living in densely populated urban area increases the risk of schizophrenia by (hazard ratio = 1.89, 95% CI: 1.53–2.33), which replicates previous studies on urbanicity effects. The inclusion of spatial random effects (E) reduces the urbanicity effect to hazard ratio = 1.64 with confidence interval encompassing 1. Model 3 with both E and GxE effects significantly contributes explanatory power to the variation in risk for schizophrenia (Log-likelihood ratio tests p < 2 × 10^{-10}), while the urbanicity effect is further reduced (hazard ratio = 1.46). Despite the concerns of residual confounds from interaction effects, Model 3 contains full pairwise interaction terms of fixed-effect covariates included in the model, i.e., PRS, genetic principal components, gender, and family history. Median hazard ratios for E and GxE components, defined as the median absolute difference in hazard ratios for all possible combinations of pairs of locales, are 2.22 and 1.78, respectively, representing a 122 and 78% expected change in risk if living in a high-risk locale.

**Spatial distribution of the risk components of schizophrenia.** The geographical distribution of E and GxE are shown in Fig. 2. The E component mirrors the heightened risk in the southwestern part of the Denmark (Fig. 2a) and the southern portion of Copenhagen, the metropolitan area with highest population density (Fig. 2b). However, within the city boundary, hazard ratios vary strongly from protective to highly detrimental (hazard ratio: 0.26 to 9.26, Fig. 2c). The GxE component has a different spatial pattern compared to E (Fig. 2d). Within the metropolitan boundary, high-risk GxE locales are concentrated in the city center (Fig. 2e) and the modulating effect can range from a...
Table 1 Hazard ratio estimates from three nested Cox regression models of the iPSYCH case-cohort data

|                | Model 1                  | Model 2                  | Model 3*                  | p-value |
|----------------|--------------------------|--------------------------|---------------------------|---------|
|                | HR  | 95% CI        | HR  | 95% CI        | HR  | 95% CI        |         |
| Individual level |     |               |     |               |     |               |         |
| Gender (male)   | 1.05 | (0.99-1.11)  | 1.06 | (1.00-1.11)  | 1.08 | (1.01-1.13)  | 0.01    |
| Genetic PC 1    | 1.07 | (1.04-1.10)  | 1.08 | (1.05-1.12)  | 1.15 | (1.11-1.18)  | 2 × 10^-15 |
| Genetic PC 2    | 0.92 | (0.89-0.94)  | 0.97 | (0.95-1.01)  | 0.99 | (0.96-1.02)  | 0.59    |
| Genetic PC 3    | 0.92 | (0.89-0.95)  | 0.92 | (0.90-0.95)  | 0.93 | (0.90-0.95)  | 4 × 10^-7 |
| Family history  | 6.07 | (5.23-7.05)  | 4.61 | (3.93-5.04)  | 5.63 | (4.75-6.67)  | <2 × 10^-16 |
| PRSb            | 1.27 | (1.24-1.31)  | 1.26 | (1.23-1.29)  | 1.34 | (1.21-1.49)  | 2 × 10^-8 |
| Spatial level   |     |               |     |               |     |               |         |
| Population density (urban vs. rural)c | 1.89 | (1.53-2.33)  | 1.64 | (0.51-5.23)  | 1.46 | (0.49-4.38)  | 0.49    |
| E                      | 2.29 |               | 2.22 |               | 1.78 |               | <2 × 10^-16 |
| GxE                    |       |               |       |               |       |               | <2 × 10^-16 |

*Model 3 is a full interaction model, obtained by multiplying PRS with all other covariates. Since E and GxE are the effects of interest, no other interactions are shown here. p-values shown are for Model 3.

bPRS has been zero centered and standardized to unit variance. The PRS estimate measure the risk associated with a one unit increase in standard deviation of the standard normal distribution of the PRS for the entire population. Therefore, comparing to a person with first decile of the PRS, a person with highest decile of the PRS has a HR of 3.25, 3.23, and 3.43, in the Model 1, Model 2, and Model 3, respectively.

cUnit increase corresponds to going from 55 person/km² to 5220 person/km², equivalent to previous definition of rural to urban residence.

dThe hazard ratios for E and GxE are median hazard ratios (median of hazard ratio absolute difference overall possible pairs of regions). p-values are based on likelihood ratio test to the model without random effects.
decrease of risk of 80% to a sixfold increase (hazard ratios: 0.20 to 5.95, Fig. 2f).

Discussion
Our novel spatial mapping analysis strategy transforms the "candidate environment" approach for disease risk into a search for environmental hot spots, localizing where environmental factors appear to have a strong impact. The flexibility of this approach enables the estimation of the amount variance accounted for by E and GxE effects without direct measurement of environmental risk factors. Both simulations and empirical application demonstrate the utility of this strategy as an alternative to the candidate environment approach.

Applying this strategy to nationwide, population-based longitudinal data enriched with genetic information, we recapitulate the well-known urban-rural gradient in schizophrenia risk based on the residential information alone. Furthermore, we show that locale of upbringing significantly contributes to the risk for schizophrenia even after controlling for population density. Both E and GxE spatial effects demonstrate substantial variation within city boundaries and account for a higher proportion of schizophrenia risk than simple urban-rural contrasts. In terms of schizophrenia risk, results indicate that the locale an individual was born and raised in is more important than urban-rural differences per se, even within the confines of a single city. Our patterns of E and GxE across Denmark can be regarded as reference distribution. The partitioned risk contour serves as an

Fig. 2 Risk distribution of E and GxE. The estimated E component is shown in the upper panel (a–c) while the estimated GxE component is shown in the lower panel (d–f). All colors were centered on national average while scaled according to risk deciles. a Hazard ratios distribution of E component in Denmark. b Hazard ratios distribution of E component in the metropolitan area, Copenhagen. c Histograms of E risk distribution within the metropolitan area. d Hazard ratios distribution of GxE component in Denmark. e Hazard ratios distribution of GxE component in the metropolitan area, Copenhagen. f Histograms of GxE risk distribution within the metropolitan area.
initial guide to find the true risk element. Further comparisons with putative environmental factors can reveal the underlying elements that are highly relevant for the etiology of schizophrenia.

As a proof of concept study, our current analysis is not without limitations. First, the average age of the iPSYCH case-cohort is younger than the expected incidence peak of schizophrenia. Although the age range of our cohort is 8–32 years, encompassing the incidence peak of schizophrenia, some cohort members are still at risk for schizophrenia. Right-censoring among cohort members reduces the power of statistical analyses. However, by analyzing the case-cohort with age-adjusted RR’s and survival analyses with inverse probability of sampling weights, we obtain unbiased estimates of incidence proportions. Second, our case-cohort is relatively young, while existing GWAS of schizophrenia tend to recruit more chronic patients in middle age39. Thus, the PRS we used may be biased toward older patients, reducing the predictive power of the already weak biological instrument. Third, the diagnostic uncertainty of very early-onset schizophrenia (onset age lesser than 13-years-old) can impact observed associations. However, a recent validation study of schizophrenia diagnoses using the Danish registry has shown good reliability in both early-onset (age 13 years to 18 years) and very early-onset (age <13 years) schizophrenia, with diagnostic concordance greater than 82 percent30. Another concern with the relatively young age of the iPSYCH sample is the inclusion of cohort members younger than 10-years-old who have very low-risk of being diagnosed as schizophrenia. These subjects are handled in the Cox proportional hazards model by treating their potential future diagnoses as right-censored outcomes, and hence have little impact on the model outputs. To verify this, we performed a sensitivity analysis on Model 3. We removed anyone younger than age 10 at study end and re-ran Model 3. As expected, the results are almost identical, with the E component on-average increasing risk by 127 percent (originally 122 percent) and GxE component on-average increasing the risk by 77 percent (originally 72 percent). Fourth, as shown in our simulations, the size of the GxE effect depends upon the predictive accuracy of the G effect. Because the PRS is a weak instrument of G, the true size of the GxE effect is probably several times larger than our current estimate, as suggested by our simulations. Fifth, we did not examine the impact of migration on locale effects. Since we cannot differentiate GxE from the gene by environment correlation introduced by migration, we restricted our analyses to individuals who have Danish parents and defined the locales as the place of birth. Although by this we intended to reduce the influence of migration, migration itself can be an important contributor for spatially-embedded risk8, as many migrants tend to live in clusters, especially in urban areas. A recent study on community samples across several countries showed that individuals with higher genetic risks of schizophrenia tend to migrate to urban area40. However, the spatial patterns we observe are unlikely due to the confounding effects of within generational drift43 since locale of upbringing was assessed before age 7, at which age no one had yet been diagnosed with schizophrenia. Inter-generational drift might still cause spatial aggregation of individuals with high genetic liabilities. A Swedish family-based study suggested urbanicity effects on schizophrenia can be explained by familial aggregation of risk13. Nevertheless, familial risk might not be the result of genetic liability but shared environmental risks within families. Danish registry studies using a cohort independent of our sample showed no evident urban aggregation of polygenic risk42, and the polygenic risk scores associated with incidence of schizophrenia independent of family history31. Therefore, there is little evidence to suggest that the identified spatial patterns is driven by inter-generational drift of families with high genetic liability for schizophrenia. Finally, we did not investigate a variety of possible socioeconomic factors in our current analyses. The potential importance such factors mandates in-depth examination in the future research; however, obtaining, validating, and analyzing socioeconomic variables as potential candidate environmental factors in the iPSYCH sample needs to be handled carefully and is beyond the scope of current paper.

Despite these caveats, we demonstrate that locale effects and modulating effects of locale on genetic risk account for a substantial proportion of urbanicity effects in Denmark. Living in a locale with a high E component increases the risk for schizophrenia by as much as 122 percent, independent of genetic liability and family history. Meanwhile, living in a locale with a high GxE component can increase risk due to genetic liability for schizophrenia by as much as 78 percent. Because our results demonstrate risk variation with finer resolution and stronger effects than urban-rural demarcation, there must be specific factors underlying previously observed urban effects. However, identification of factors explicating urban risk has been unsuccessful to date42,47. Given the uncertainty involved, invalid constructs or measurement error could be contributors to lower power to detect risk associations with specific environmental factors. Our spatial mapping strategy is an alternative approach, since finding high-risk locales does not depend on correct specification of a purported environmental risk factor.

In the nineteenth century, epidemiology pioneer John Snow mapped high-density regions of cholera cases onto London streets and thus identified the water source as the key infectious medium. By demonstrating that the locale of upbringing significantly contributes to risk and modulates genetic susceptibility to schizophrenia, we hope this is the first step in isolating the source of spatial risk variation, facilitating the design of future public health interventions for severe mental disorders.

Methods
Our spatial mapping approach follows three steps: (1) defining neighboring locales to characterize the latent environment field, (2) estimating random effects associated with each locale, and (3) mapping the spatial distribution based on the realized effects on locales. These three steps are calibrated to ensure a good balance between fine spatial resolution and adequate statistical power. Furthermore, the modeling strategy partitions observed effects on risk for schizophrenia into different components: locale of upbringing (E), genetics (G), and the synergistic effects of spatial locale and genetics (GxE).

Defining locales for risk mapping
We exploit the duality between Delaunay triangulation and Voronoi tessellation32, ensuring each defined locale has a sufficient number of study subjects to be well-powered while achieving a fine spatial resolution (Supplemental Information). The Voronoi tessellation partitions the whole map into smaller units based on individuals’ coordinates on the map, making sure every point in a given unit area is closer to its centroid than any other. Their neighborhood relationships are defined simultaneously because the centroids are connected by the dual of Voronoi tessellation, i.e., Delaunay triangulation. After defining neighborhood relationships, individuals are grouped with their closest neighbors, making the locale growing in size, until the number of individuals in the defined locale reaches a pre-defined range (Supplementary Fig. 1 and Supplemental Information). The algorithm thus achieves a balance between spatial resolution and a sufficient number of subjects in each locale by adaptively merging neighboring locales with too few individuals into larger locales. The primary advantage from this approach is to localize the regions as much as possible while retaining high statistical power to estimate locale (E) and gene x locale (GxE) spatial random effects. This also prevents potential bias introduced by estimating spatial risks via a smoothing kernel, as exemplified by one twin study that used an isotropic smoothing kernel to estimate the spatial distribution of the risk in mental illness, inadvertently biasing all outcomes, regardless of diagnosis, toward densely populated areas34.

Estimating the effects associate with the locale
Mixed effects models provide the necessary tools to estimate the latent environmental and gene x environmental effects. Fixed effects in the model control for potential confounding factors, whereas random locale effects approximate the latent field across all spatial locations. Once the random effect variance is estimated and determined to be significant, spatial mapping is achieved through computing the posterior means of the random effects for each locale, defined by the best linear unbiased predictors33.
To ensure the validity of this approach, we performed 1000 Monte Carlo simulations to determine how well we can estimate E and GxE via the spatial mixed effects model. Given a sample size of 30,000 individuals with disease prevalence of one percent and heritability of 70 percent (similar to the profile of schizophrenia), we obtained an unbiased estimation of spatial effects (E), while GxE effects are conservatively bounded by the predictive power of the genetic instrument (Supplementary Fig. 2 and Supplemental Information). As variance explained of the genetic liabilities increases for the genetic instrument, the amount of GxE effects explained is also increased.

**Empirical study on the risk of schizophrenia.** We demonstrate the feasibility of our spatial mapping approach by characterizing E and GxE effects of schizophrenia in the Danish population. To map the synergistic effects of locale of upbringing and schizophrenia, we used the Danish Civil Registration System. To focus on the early life experience, we excluded all psychiatric contacts until 31 December 2013 were obtained from the Danish Psychiatric Central Research Register, using diagnostic classification of Diseases, 10th revision. Genetic PCs were derived based on principal component analysis with a set of 15,000 single nucleotide polymorphisms (SNPs) from the Psychiatric Genomics Consortium (PGC) Schizophrenia GWAS. The PRS is the sum of the products of effect sizes of SNPs estimated from this independent GWAS and the dosage of those SNPs from the iPSYCH case-cohort. The included SNPs were pruned to ensure independence, while no significance threshold was set to filter SNPs. Parameters for calculating PRS include: clumping ($r^2 = 0.1$, distance = 250 kb), and pruning ($r^2 = 0.8$, window = 2 kb, increment = 2 kb). Nonetheless, PRS is inherently a weak genetic instrument, so our estimate on GxE is as a conservative lower bound of interaction effects (Supplementary Fig. 2).

**Code availability.** The code used for simulations, empirical analysis, and visualization can be found at https://chancheeфан.shinyapps.io/iPSYCH_geo_tess_S2/. The interactive version of the disease mapping is shown on the web portal while all the relevant codes can be downloaded on it. All analyses are implemented in R. R packages employed include spatstat and coxme. The geographical visualization is done with ggmap, which extracts geographical information from Google Maps. An interactive version of the risk map is generated using leaflet and shiny.

**Data availability**

Data for generating figures are provided as Supplementary Information. All relevant data is available upon request.

Received: 14 May 2018 Accepted: 16 November 2018
Published online: 13 December 2018

**References**

1. Keller, M. C. Genexenvironment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. Biol. Psychiatry 75, 18–24 (2014).
2. McAllister, K. et al. Current challenges and new opportunities for geneenvironment interaction studies of complex diseases. Am. J. Epidemiol. 186, 753–761 (2017).
3. Rappaport, S. M. & Smith, M. T. Environment and disease risks. Science 330, 369 (2010).
4. Pedersen, C. B. & Mortensen, P. B. Family history, place and season of birth as risk factors for schizophrenia in Denmark: a replication and reanalysis. Br. J. Psychiatry 179, 46–52 (2001).
5. Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A. & Lewis, C. M. Meta-analysis of the association of urbanicity with schizophrenia. Schizophr. Bull. 30, 1118–1123 (2012).
6. Pedersen, C. B. No evidence of time trends in the urban–rural differences in schizophrenia risk among five million people born in Denmark from 1910 to 1986. Psychol. Med. 36, 211–219 (2006).
7. Pedersen, C. B. Persons with schizophrenia migrate towards urban areas due to the development of their disorder or its prodromata. *Schizophr. Res.* 168, 284–208 (2015).
8. Colodro-Conde, L. et al. Association between population density and genetic risk for schizophrenia. *JAMA Psychiatry* 75, 901–910 (2018).
9. Pedersen, C. B. & Mortensen, P. B. Urbanization and traffic related exposures as risk factors for schizophrenia. *BMC Psychiatry* 6, 2 (2006).
10. Brown, A. S. & Derkits, E. J. Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. *Am. J. Psychiatry* 167, 261–280 (2009).
11. Werner, S., Malaspina, D. & Rabinowitz, J. Socioeconomic status at birth is associated with risk of schizophrenia: Population-based multilevel study. *Schizophr. Bull.* 33, 1373 (2007).
12. Kirkbride, J. B., Jones, P. B., Ullrich, S. & Coid, J. W. Social deprivation, inequality, and the neighborhood-level incidence of psychotic syndromes in East London. *Schizophr. Bull.* 40, 169–180 (2014).
13. Sariaslan, A. et al. Does population density and neighborhood deprivation predict schizophrenia? A nationwide swedish family-based study of 2.4 million individuals. *Schizophr. Bull.* 41, 494–502 (2015).
14. Pedersen, C. B. & Mortensen, P. B. Are the cause(s) responsible for urban-rural differences in schizophrenia risk rooted in families or in individuals? *Am. J. Epidemiol.* 163, 971–978 (2006).
15. Zammit, S. et al. Individuals, schools, and neighborhood: A multilevel longitudinal study of variation in incidence of psychotic disorders. *Arch. Gen. Psychiatry* 67, 914–922 (2010).
16. van Os, J., Harris, M., Bak, M., Bijl, R. V. & Vollger, W. Do urbanicity and familial liability coparticipate in causing psychosis? *Am. J. Psychiatry* 160, 477–482 (2003).
17. van Os, J., Pedersen, C. B. & Mortensen, P. B. Confirmation of synergy between urbanicity and familial liability in the causation of psychosis. *Am. J. Psychiatry* 161, 2312–2314 (2004).
18. Kirkbride, J. B., van Os, J. Schizophrenia and urbanicity: a major environmental influence—conditional on genetic risk. *Schizophr. Bull.* 31, 795–799 (2005).
19. Grecch, A., van Os, J. & Investigators, G. Evidence that the urban environment moderates the level of familial clustering of positive psychotic symptoms. *Schizophr. Bull.* 43, 325–331 (2017).
20. Paksaian, D. et al. The role of genetic liability in the association of urbanicity at birth and during upbringing with schizophrenia in Denmark. *Psychol. Med.* 48, 1–10 (2017).
21. Visscher, P. M. et al. 10 Years of GWAS discovery: Biology, function, and translation. *Am. J. Human. Genet.* 101, 5–22 (2017).
22. Snow, J. *On the Mode of Communication of Cholera.* (John Churchill, London, 1855).
23. Kelsall, J. & Wakefield, J. Modeling spatial variation in disease risk: a geostatistical approach. *J. Am. Stat. Assoc.* 97, 692–701 (2002).
24. Faris, R. E. L. & Dunham, H. W. Mental Disorders in Urban Areas: an Ecological Study of Schizophrenia and Other Psychoses. (Univ. Chicago Press, Oxford, England, 1939).
25. Kirkbride, J. B. et al. Neighbourhood variation in the incidence of psychotic disorders in Southeast London. *Social. Psychiatry Psychiatr. Epidemiol.* 42, 438–445 (2007).
26. Torrey, E. F., Mortensen, P. B., Pedersen, C. B., Wohlfahrt, J. & Melbye, M. Risk factors and confounders in the geographical clustering of schizophrenia. *Schizophr. Res.* 49, 295–299 (2001).
27. Davis, O. S. P., Haworth, C. M. A., Lewis, C. M. & Plomin, R. Visual analysis of geocoded twin data puts nature and nurture on the map. *Mol. Psychiatry* 17, 867 (2012).
28. Austin, P. C., Wagner, P. & Merlo, J. The median hazard ratio: a useful measure of variance and general contextual effects in multilevel survival analysis. *Stat. Med.* 36, 928–938 (2017).
29. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. *Mol. Psychiatry* 21, 969–974 (2016).
30. Veral, D. L. et al. Validation study of the early onset schizophrenia diagnosis in the Danish Psychiatric Central Research Register. *Eur. Child & Adolesc. Psychiatry* 27, 965–975 (2018).
31. Lu, Y. et al. Genetic risk scores and family history as predictors of schizophrenia in Nordic registers. *Psychol. Med.* 48, 1201–2008 (2018).
32. Barr, C. D. & Schoenberg, F. P. On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika* 97, 977–984 (2010).
33. Hilker, R. et al. Heritability of schizophrenia and schizophrenia spectrum based disorders in Southeast London. *Mol. Psychiatry* 83, 492–498 (2016).
34. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for case-cohort designs. *Clin. Epidemiol.* 52, 1165–1172 (1999).
35. O’Connell, J. et al. Haplotype estimation for biobank-scale data sets. *Nat. Genet.* 48, 817–820 (2016).
36. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904 (2006).
37. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of biobank-scale genotype datasets. *Bioinformatics* 33, 2776–2778 (2016).