Mapping the genetic and environmental aetiology of autistic traits in Sweden and the UK

Zoe E. Reed[1,2*], Henrik Larsson[4,5], Claire M.A. Haworth[1,2,3,6,7], Dheeraj Rai[2,3,8], Sebastian Lundström[9], Angelica Ronald[10], Abraham Reichenberg[11], Paul Lichtenstein[12], Oliver S.P. Davis[1,2,3,7]

* Corresponding author

1) MRC Integrative Epidemiology Unit, University of Bristol, UK

2) Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK

3) National Institute for Health Research Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK

4) School of Medical Sciences, Örebro University, Sweden

5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

6) School of Psychological Science, University of Bristol, UK

7) The Alan Turing Institute, London, UK

8) Avon and Wiltshire Partnership NHS Mental Health Trust

9) Gillberg Neuropsychiatry Centre, Centre for Ethics, Law and Mental Health, University of Gothenburg, Sweden

10) Department of Psychological Sciences, Birkbeck, University of London, UK

11) Department of Psychiatry and Seaver Autism Center for Research and Treatment, Ichan School of Medicine at Mount Sinai, New York, USA

12) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

Abstract

Autistic traits are influenced by both genetic and environmental factors, and are known to vary geographically in prevalence. But to what extent does their aetiology also vary from place to place? We applied a novel spatial approach to data from two large twin studies, the Child and Adolescent Twin Study in Sweden (CATSS) and the Twins Early Development Study (TEDS) in the UK, to explore how the influence of nature and nurture on autistic traits varies from place to place. We present maps of gene- and environment- by geography interactions that suggest, for example, higher heritability and lower non-shared environmental influence in more densely populated areas. We hope this systematic approach to aetiological interactions will inspire research to identify previously unknown environmental influences on the aetiology of autistic traits.

Background

Autism spectrum disorder (ASD) is a neurodevelopmental condition that manifests in childhood. ASD is generally characterised by persistent difficulties with social communication and repetitive behaviours. Reported prevalence of ASD varies, but in developed countries the prevalence is estimated to be between 1-1.5% [1–4] and this recorded prevalence has increased over the past few decades [1,3,4]. Factors such as diagnostic criteria, age, time of study and location of study may all contribute to this heterogeneity in prevalence estimates. ASD has a significant impact on child development, often including language difficulties and other co-occurring conditions which may persist into adulthood [5].

The aetiology of ASD reflects both genetic and environmental influences. Twin studies suggest that genetic differences between people explain around 80% of the population variance for ASD [6]. Most studies suggest that the remaining variance is explained by variation in the non-shared environment. That is, environmental influences that do not contribute to similarity within families. Similarly, a recent study of over 3.5 million twin and sibling pairs in Sweden found that 83% of the variance is explained by genetic differences and 17% by non-shared environmental influences [7]. Another study across 5 different countries (Sweden, Finland, Denmark, Western Australia and Israel) estimated heritability for ASD to be around 80%, using data from whole populations, although there was variation between countries [8].

ASD is known to vary in prevalence across geographical regions. For example, spatial analyses of ASD have revealed areas of increased prevalence in Salt Lake

County in Utah [9], in northern Taiwan [10] and in areas of California [11,12]. ASD also appears more common in those born in New England compared to those born in the south east of the United States (US) [13]. Similarly, a study of Greater Glasgow in Scotland identified variation in prevalence across the city [14]. Several studies have also suggested differences in prevalence between urban and rural areas, where living in or growing up in an urban environment is associated with greater risk of ASD compared to rural environments [10,15–18]. Possible reasons for geographical variation in prevalence of ASD across these areas include regional diagnostic bias, differences in the access to health services or diagnostic resources, different levels of parental awareness, air pollution exposure during pregnancy, green space in an area and local trends in socioeconomic status (SES).

If prevalence of autistic traits varies from place to place, is the same true of the aetiology? For example, does variation in the environment explain variation in autistic traits in some areas more than others? Similarly, does the environment in some areas draw out genetic differences between children in their propensity for developing autistic traits? We previously developed a spatial approach to twin model-fitting called spACE that showed genetic and environmental influences vary spatially within a country in response to geographically-distributed environments [19]. This approach has the potential to highlight gene-environment (G×E) and environment-environment (E×E) interactions for outcomes such as ASD traits. G×E and E×E represent variation in the aetiological genetic influences on a trait depending on environmental exposure. For example, genetic risk of a mental health disorder may be drawn out by a stressful

84 environment, genetic risk of asthma may be apparent only in polluted
85 environments, or genetic risk of hay fever may only reveal itself in pollen-rich
86 areas. The spACE approach allows us to investigate this, mapping geographical
87 patterns of nature and nurture without requiring the measurement of specific
88 genetic variants or specific environmental characteristics. This systematic
89 geographical approach may facilitate the discovery of novel specific genetic and
90 environmental influences.

91 Here we apply the spACE approach to data on autistic traits in Sweden and the
92 UK. Autistic traits and diagnostic categories of ASD show substantial aetiological
93 overlap [20,21], with genetic correlations from bivariate twin models of 0.52-0.89
94 and SNP based genetic correlations of 0.27-0.30. The heritability of ASD traits
95 does not change as a function of severity [22–24], and genetic links have been
96 identified between extreme and sub-threshold variation in ASD [22,24], so to
97 maximise power we have focussed on trait measures rather than diagnoses.

98 It seems likely that environments previously identified as important for the
99 development of autistic traits will also influence aetiology. For example, given
100 previous research on the social stress of urban compared to rural upbringing [25],
101 we hypothesise that urban-rural differences will be apparent in the aetiology of
102 autistic traits. However, more importantly, we hope that by systematically
103 mapping geographical differences in aetiology we will facilitate identification of
104 new environments and shed light on the mechanisms by which they act.

Methods

The Swedish Twin Registry and CATSS

The Swedish twin registry, established in the 1950s, currently includes over 194,000 twins [26]. Phenotypic information on the twins comes from a variety of sources such as medical registers and questionnaires and is regularly updated. Several sub-studies of the registry have been established, including the longitudinal Child and Adolescent Twin Study in Sweden (CATSS) [27]. CATSS was launched in 2004 to investigate childhood-onset neurodevelopmental problems such as ADHD and ASD in childhood and adolescence, for all twins turning 9 or 12 years since 2004. Parents of all Swedish twins aged 9 and 12 years old were asked to participate in a telephone interview to collect information on various health-related issues. By the time data on autistic traits were obtained in 2013, 8,610 parents had responded to this request, accounting for 17,220 twins. The CATSS-9/12 study obtained ethical approval from the Karolinska Institute Ethical Review Board: Dnr 03-672 and 2010/507-31/1, CATSS-9 – clinical 2010/1099-31/3 CATSS-15 Dnr: 2009/1599-32/5, CATSS-15/DOGSS Dnr: 03-672 and 2010/1356/31/1, and CATSS-18 Dnr: 2010/1410/31/1.

For ASD traits, 16,677 participants had data available (including 8,307 complete pairs and 63 incomplete pairs of twins). Interviews were carried out when the twins were around the age of 9 or 12 years and 51% were male.

CATSS measures of autistic traits

The Autism-Tics, ADHD and other Comorbidities (A-TAC) inventory, based on the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria, was used in the

telephone interview with parents to collect information on a range of neurodevelopmental problems. This inventory has previously been validated in both clinically diagnosed children and the general population [28–32]. The inventory includes 17 items that assess ASD symptoms, where respondents can answer 'yes/1', 'yes, to some extent/0.5', and 'no/0'. Following the standard approach, we created a symptom score for each individual by summing these item scores. Further details can be found in previous publications [32].

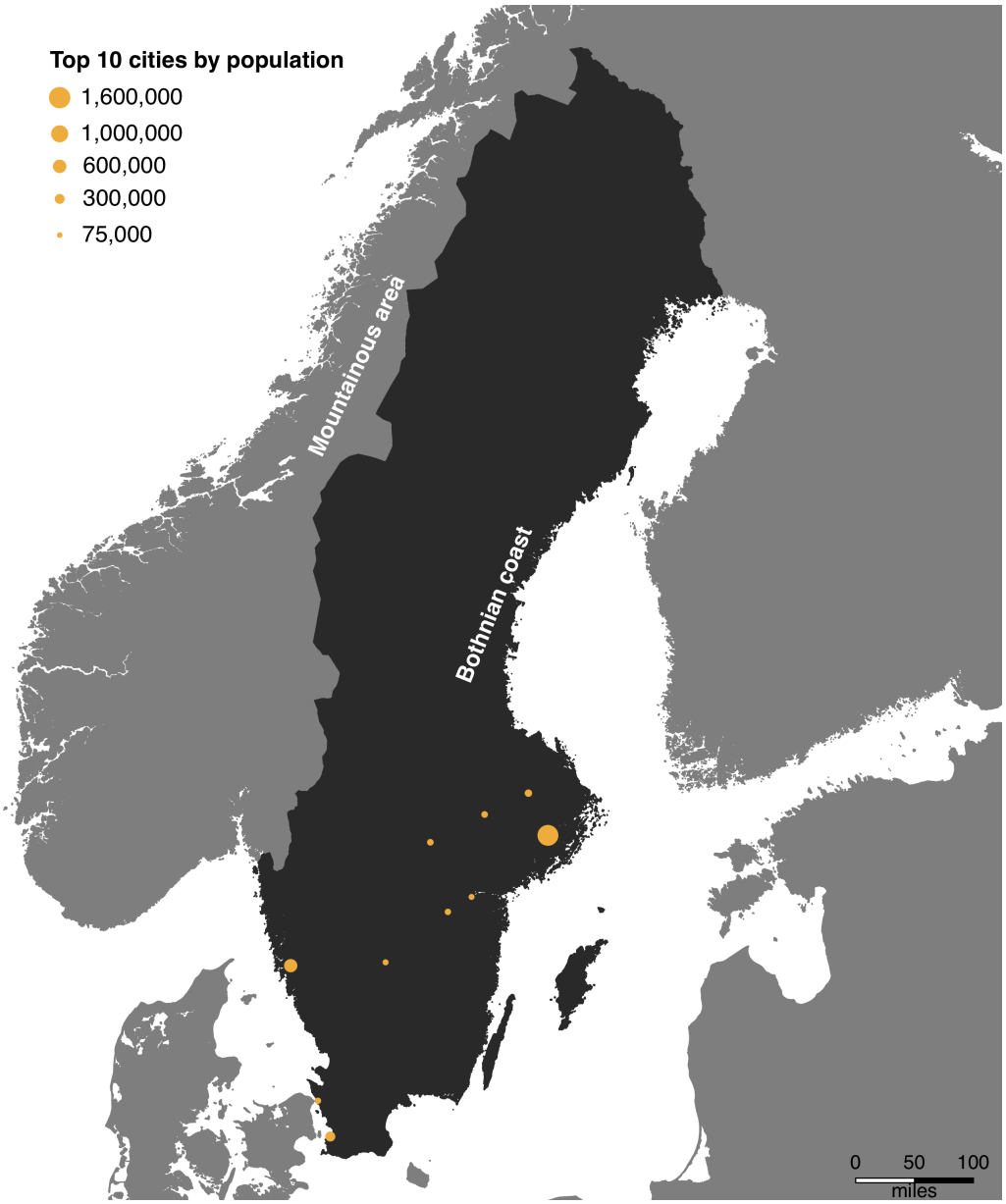
In our sample the median score was 0.00 (interquartile range [IQR]=1.00) for autistic symptoms, where, in previous validation studies a low and high cut-off of 4.5 and 8.5 for ASD have been established for broad screening and for use as a clinical proxy, respectively. This indicates that most people, in this general population sample, score well below these cut-offs. As expected, the distribution of this symptom score was zero-skewed, as shown in the histogram in **supplementary figure 1**.

CATSS location data

To conduct the spACE analysis, we assigned a geographical location to each family. In CATSS we matched each twin pair to a Small Areas for Market Statistics (SAMS) location, for the most recent location data we had available up to 2009, using data from Statistics Sweden (http://www.scb.se/en/) and assigned coordinates based on the centroid of the SAMS location. There are approximately 9,200 SAMS in Sweden, subdivisions of 290 municipalities. The average population within each SAMS is 1,000 people and therefore the area covered by each SAMS varies by population density.

To provide context for the results for Sweden, it is useful to understand a little about its geography. **Figure 1** shows a map of Sweden and some general indicators of the country's geography; the **supplementary materials** contain a detailed description. In summary, Sweden is split into a more rural north and central area, known as the lowlands and the more populated areas in a belt from Gothenburg in the west to Stockholm in the east and the very south near Malmö. Much of Sweden is covered by forest and lakes. The capital, Stockholm, is in the east with a mix of tourist-centred and residential areas and a number of islands. Gothenburg, the second largest city with a port, also has varied areas like Stockholm and an archipelago. Further inland is rural Värmland. South-west Sweden is a coastal and lowland area and is the third most populated area in Sweden. The main city in this area is industrialised and multicultural Malmö. South-east Sweden is heavily forested with some large lakes and a number of large towns. Sweden's two largest islands are also found here, Öland and Gotland, which are popular summer destinations, due to their warmer climate. They both have fairly rural landscapes with small towns and villages. The Bothnian coast is the most populated area in the north, with some large towns along the coastline. Central Sweden is a sparsely populated, rural, lowland area covered in forests, with numerous lakes and mountains along the Norwegian border. Further north is Swedish Lapland, a very remote area with a mountainous, rural landscape.

167 **Figure 1.** Map of Sweden with top 10 cities by population



168

169 *Sweden is shown in dark grey with the surrounding countries in a lighter shade of grey. The top 10 most*

170 *populated cities are shown on the map with orange circles, the area of which reflects the population of the*

171 *city.*

The Twins Early Development Study

The Twins Early Development Study (TEDS) contacted parents of twins born in England and Wales between January 1994 and December 1996 [33]. 16,810 pairs of twins were initially recruited, and currently there are over 10,000 twin pairs still enrolled in TEDS. The participants are demographically representative of the UK population of a similar age, with the majority identifying themselves as white British and with English as their first language. TEDS has collected wide-ranging data on cognitive and behavioural development, using approaches that include questionnaire booklets, telephone testing and web-based tests. The twins, their parents and teachers have all participated in data collection. Ethical approval for TEDS research is provided by the Institute of Psychiatry, Psychology and Neuroscience Ethics Committee, King's College London.

Full phenotypic data for autistic traits were available for 11,594 TEDS participants (including 5,796 complete pairs and 62 incomplete pairs of twins). For these twins the mean age was 11.30 (SD=0.72) and 48% were male.

TEDS measures of autistic traits

Parents in TEDS completed the Childhood Autism Spectrum Test (CAST) when the twins were age 12 years. The CAST consists of 30 items, scored 1 for yes or 0 for no [34]. For participants included in our analyses, the median score for ASD symptoms was 4.0 (IQR=4.84). The CAST score considered indicative of ASD is 15.

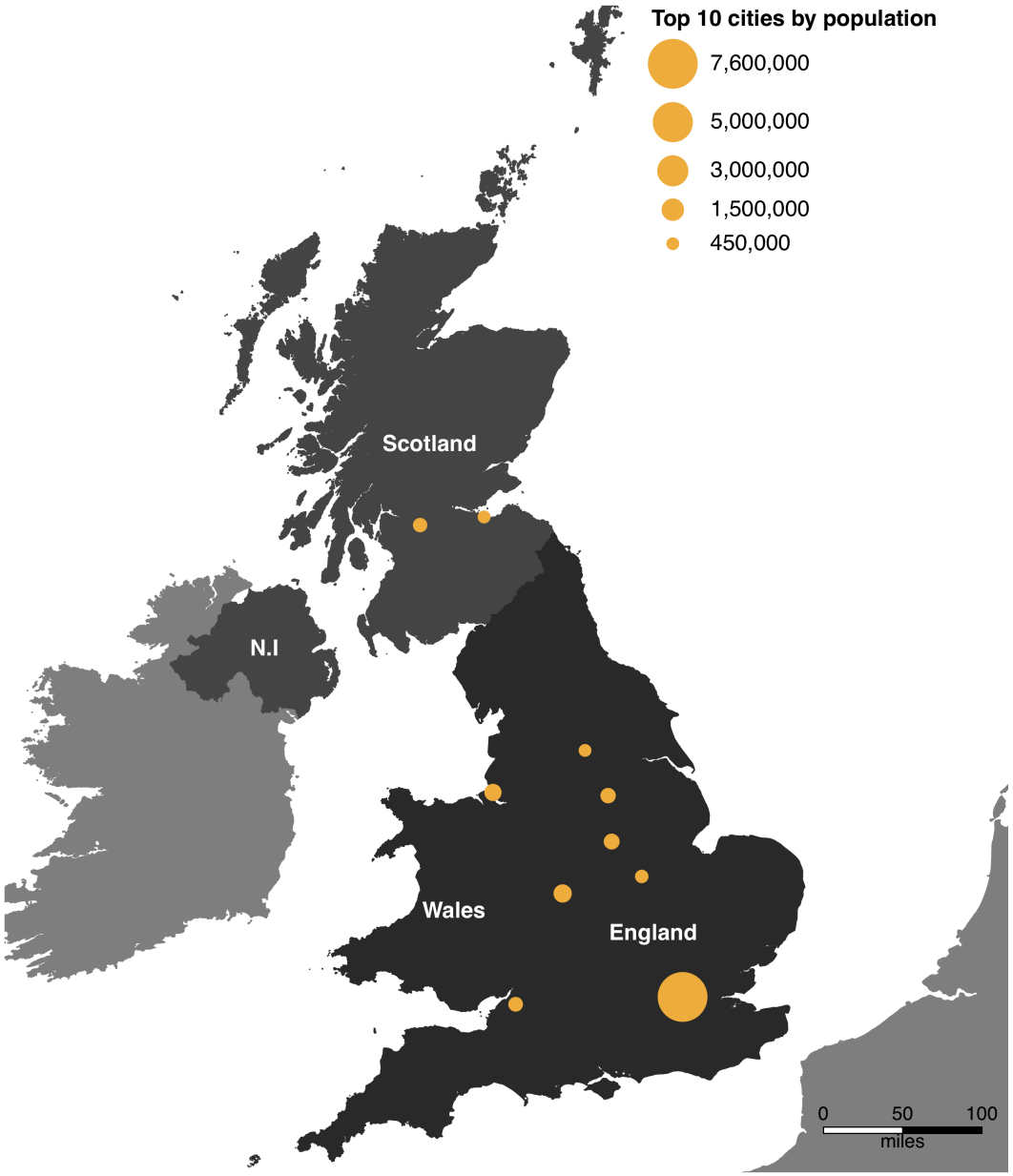
TEDS location data

We assigned each twin pair geographical coordinates based on the centroids of their postcodes at age 12. There are over 1.5 million postcode units in the UK, covering, on average, 15 properties. As with SAMS, the area covered by each postcode varies depending on population density.

To provide context for the results for the UK, **Figure 2** displays a map and some general indicators of the country's geography; the **supplementary materials** include a detailed description. The UK is split into England, Wales, Scotland and Northern Ireland (although we do not describe Northern Ireland here because it was not included in the TEDS recruitment area and few participants have moved there since recruitment). Generally, the south of the UK has a milder climate compared to the north and has more low-lying land. The UK is a mix of some very urban, previously (or still) industrial areas and more rural traditional countryside areas. London, a diverse, multicultural city in the south-east, is the capital, with its own distinct boroughs. The south-east of England has many commuter areas and is surrounded by coastline. This area is historically rich and has a mix of industrial and countryside areas and a number of seaside towns. The south of the UK is fairly rural, has many historical sites and also many coastal towns as well as the New Forest. More inland are the areas of Oxfordshire, with the city of Oxford and the picturesque, rural Cotswolds. In the west are the areas of cosmopolitan Bristol, spa-city Bath and rural Somerset, with the wooded Mendips, the Quantock and Exmoor National Park, on the Bristol channel. South-west England consists of pre-industrial Devon and Cornwall, popular Summer destinations with many seaside and fishing towns and plenty of farmland, and Dartmoor National Park. East Anglia is an area of flatland, wetlands and coastal areas.

The west Midlands are a mix of lowlands and hilly areas with the industrial city of Birmingham (England's second largest city) and the Peak district. The east midlands has a number of large urbanised cities and is an old coal mining area, but rural areas can still be found. North-west England has the large cities of Manchester and Liverpool and the seaside resort of Blackpool, but also the unspoilt, mountainous Isle of Man. The scenic Lake district is also found in the north as well as the varied area of Yorkshire, with urbanised, coastal and rural areas. Wales is split into the more populated and coastal south, hilly, rural Mid-Wales and mountainous north Wales. Scotland is split into the Highlands in the north and the Lowlands in the south. Southern Scotland is home to the main cities of cosmopolitan and medieval Edinburgh and urban Glasgow and this area has coastal towns, forests and agricultural land. Central Scotland is more varied with large lochs and forests in the west, rural and industrialised areas and fishing villages in the east and peaks in the north. Argyll in the west is a remote area, transitioning between lowland and highland and with numerous islands. North east Scotland has a number of industrial cities and port towns, although further north becomes mountainous. The Highlands are a very remote but unspoilt area, with forests, lochs, mountains and rugged coastline. Scotland has a number of island clusters, the Inner and Outer Hebrides in the west and the Orkney and Shetland islands in the north.

233 **Figure 2.** Map of the UK with top 10 cities by population



234

235 *The recruitment area (England and Wales) for the Twins Early Development Study (TEDS) is shown in dark*
236 *grey, with the rest of the UK (Scotland and Northern Ireland [N.I]) in a lighter shade of grey. Other countries*
237 *are shown in the lightest shade of grey. The top 10 most populated cities are shown on the map with orange*
238 *circles, the area of which reflects the population of the city.*

Statistical analyses

ACE models and maps in CATSS and TEDS

In twin analysis, within-pair similarity of monozygotic (MZ) and dizygotic (DZ) twins is compared to estimate parameters for additive genetic (A), shared environmental (C) and non-shared environmental (E) influences on a trait. In this context, the shared environment refers to influences other than DNA similarity that make children growing up in the same family more similar to each other, whilst the non-shared environment refers to influences that do not contribute to similarity within families. Although tempting, it is not possible to assign specific environments to one or the other environmental component, because most environments themselves show both shared and non-shared (and often genetic) influences. We can estimate the contribution of genetic and environmental influences because of the different ways these influences are shared in MZ and DZ twin pairs. For MZ twins, who share 100% of their segregating alleles, A influences correlate 1, whereas for DZ twins they correlate 0.5 because DZ twins share, on average, 50% of their segregating alleles. For both MZ and DZ twins growing up in the same family the shared environmental correlation is 1. In contrast, the non-shared environment is uncorrelated and contributes to differences between twins [35]. In this study, we applied a version of the spACE analysis method [19] to explore how A, C and E for ASD traits vary geographically. To do this, we fit full information maximum likelihood structural equation models to twin data in R (version 3.3.1) using the OpenMx package (version 2.9.4), calculating A, C and E at many different target locations across an area. The contribution of each twin pair to a model is weighted by a function of the inverse Euclidean distance of the twin pair from the target location. In this study we built on our previous work by applying the weights within the structural

equation modelling framework, rather than by calculating weighted correlation matrices and using those as input (although for normally distributed measures the results are the same with either approach). In twin analysis it is possible to model non-additive genetic effects (D) instead of shared environmental effects (C), and D influences are sometimes found with ASD. However, the D component is highly correlated with the A component, which means confidence intervals are wide and the tendency of variance to swap between these two components makes it difficult to compare results across locations. Because of this, we have fitted ACE models, although in this case, A should be considered broad-sense heritability, including both additive and non-additive genetic influences.

For target locations in Sweden we used the centroid of each unique SAMS that included at least one twin pair. Because UK postcodes give more precise locations than Swedish SAMS, we instead selected UK target locations representative of local population density to preserve participant anonymity. All twin pairs contributed to the results at each location, but contributions were weighted according to the distance of each twin pair from the target location:

$$w_i(x) = \frac{1}{d(x, x_i)^p}$$

where x represents the target location, x_i represents the location of a twin pair, d is the Euclidean distance between x and x_i, and p is the power parameter that controls the rate of drop-off of a twin pair's influence over distance (0.5 for these analyses). We included sex as a covariate in all the models (accounting for on average 2.59% of the variance), and age in the TEDS data (where it accounted for on average 0.34% of the variance).

Further detail on the spACE approach can be found in the original article [19]. We plotted maps to visualise the results (**figures 3 and 4**). In the maps each target location is coloured according to the value of the estimate at that location compared to the full range of values across the map. Low values appear blue and high values appear red, with increasing brightness of the colour representing increasing distance from the mean. To avoid outliers having a large effect on the distribution of colours in the maps, we assigned the highest 4% of values to the brightest red and the lowest 4% of values to the brightest blue before assigning colour values to equal ranges between the two. The histograms show the distribution of results and the corresponding colours.

We estimated 95% confidence intervals for A, C and E at each target location and using the CATSS data we performed sensitivity analyses for how A, C and E estimates vary based on the historical residential location used for the twin pairs. To do this we repeated analyses based on participants' locations at different ages and we combined the resulting maps into a video (**supplementary video 1**). Changes across time may allow identification of critical developmental periods when the geographical environment is particularly influential; for example, if clear patterns are seen when participant locations for the analysis are based on their location at a specific age.

Sex limitation models

While some previous studies have identified no aetiological sex differences for ASD, others have. For example, one study using the Missouri twin study and a continuous measure of autistic traits, found no sex differences [36], and neither did previous work in TEDS[24], but modest sex differences were found in previous work with the Swedish Twin Study[37]. To maximise power, in the main text we report results that equate the

aetiological influences for males and females. But for the Swedish data, where we replicated quantitative sex differences in aetiology, we conducted further separate analyses for males and females, and we include these in supplementary materials (**supplementary figures 4 and 5**).

Data availability

The data used in this study are available to researchers directly from CATSS and TEDS. Procedures for accessing the data are described at https://ki.se/en/meb/the-child-and-adolescent-twin-study-in-sweden-catss (CATSS) and https://www.teds.ac.uk/researchers/teds-data-access-policy (TEDS).

Code availability

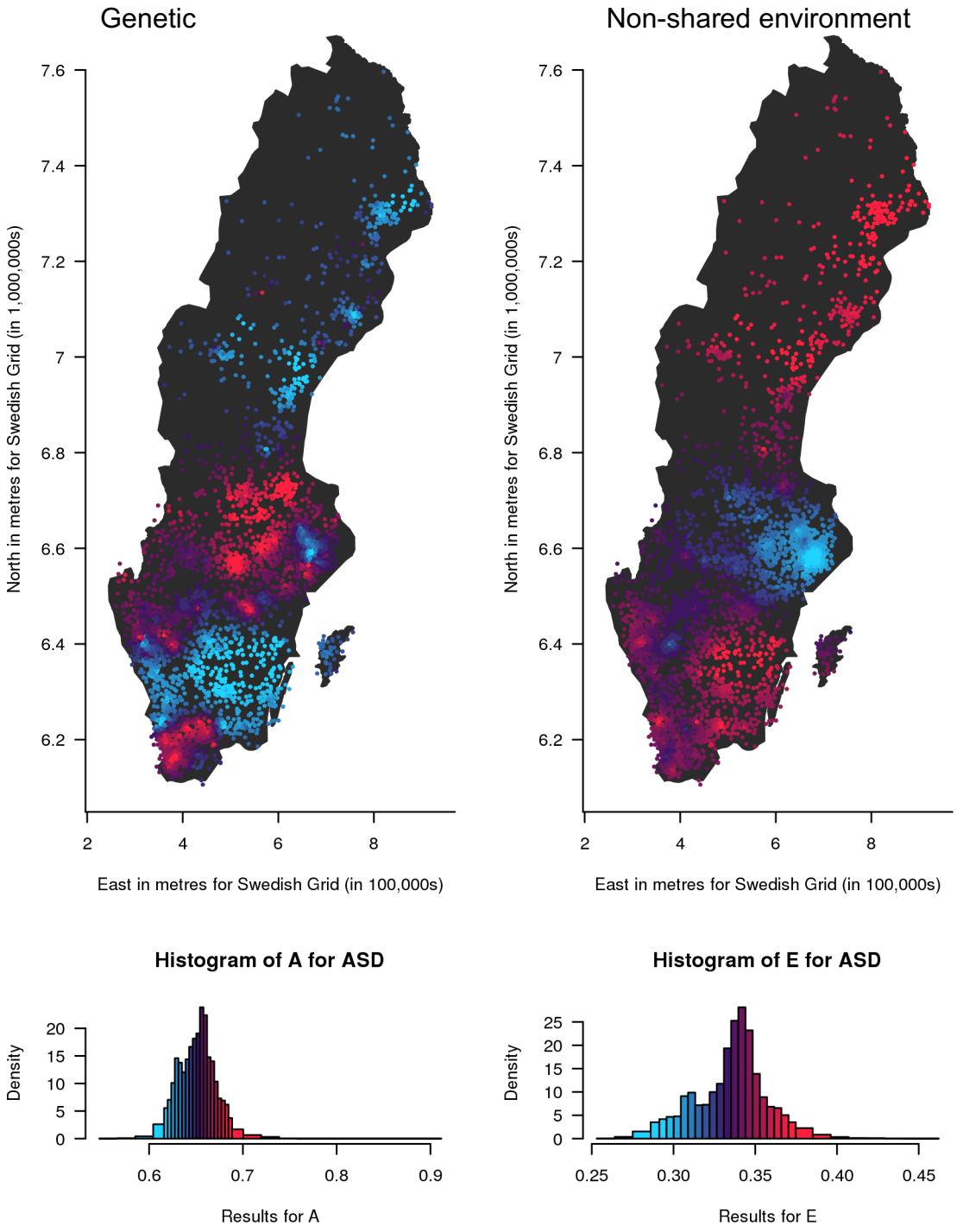
Code that implements the spACE model described here is available in the scripts directory at https://github.com/DynamicGenetics/spACEjs/.

Results

Mapping the aetiology of autistic traits in Sweden

We plotted the results from each of the 4,199 locations on a map (**Figure 3**, an interactive version is available at https://dynamicgenetics.github.io/spACEjs/), where red points represent locations where results fall above the population mean, and blue points represent locations where they fall below. The brighter the points, the further they are from the population mean. This is shown in the histograms below the maps, where the colours of the bars match the points in the map above. Because we modelled raw variance after standardising data to mean 0 and SD 1 at the population level (i.e. we did not standardize the A, C and E estimates at each location to add up to one) genetic and environmental influences are not reciprocal at each location, so it is possible for a location to show both strong genetic and environmental influences. The consequence of this is that each map stands alone: differences in genetic influence really do imply differences in the genetic component, and not just a reflection of differences in a reciprocal environmental component. Maps with A and E constrained to add up to one in each location (i.e. proportional) are shown in **supplementary figure 2** and show similar results to those for the raw variance. For comparison, we have also plotted results of the weighted means of scaled autistic trait scores at the same locations in **supplementary figure 3** and we observe geographical variation for mean autistic trait scores, reflecting the expected variation in the prevalence of ASD.

Figure 3. Mapping genetic (A) and non-shared environmental (E) influences on autistic traits in Sweden suggests that genetic variation is more influential in more densely-populated areas



343

344 *Geographical variation in genetic (A) and non-shared (E) influences on childhood ASD traits in Sweden*
345 *(results are overlaid on an outline of the SAMS areas). The contributions of A and E range from low (blue) to*
346 *high (red). The histograms below show the distribution of the estimates, coloured in the same way as the*
347 *points on the map. The estimates are not standardised and are therefore not constrained to add up to one.*
348 *Shared environment (C) estimates were approximately zero across the whole map, so they are not shown*
349 *here. An interactive version of this map is available at https://dynamicgenetics.github.io/spACEjs/.*

The results suggest that the amount of variation in autistic traits explained by genetic influences (A) is generally greater in urban areas and lesser in the sparsely populated north and more rural southern belt. The non-shared environment (E) frequently shows the opposite pattern, with the variation explained generally less in and around the capital and more in Southern and Northern rural areas. However, we also observe greater contribution of E in the areas around the cities of Gothenburg and Malmö. Variation in A and E can also be seen within local areas, such as around Stockholm, where there are both low and high values for A, suggesting genetic influences are moderated by other factors beyond urbanicity. The histograms for the raw variance indicate that the variance explained by genetic influences ranges from 0.55 to 0.91, with most values around the mean of 0.65 (SD=0.02). The variance explained by E ranges from 0.25 to 0.46, again with most values around the mean of 0.34 (SD=0.02). Variation in autistic traits explained by C was approximately zero over the whole of Sweden. Confidence intervals for estimates at each location are provided in **supplementary table 1. Supplementary video 1** shows that the overall patterns for variation in A and E remained similar irrespective of the historical location used for each twin pair.

Sex limitation models for ASD traits

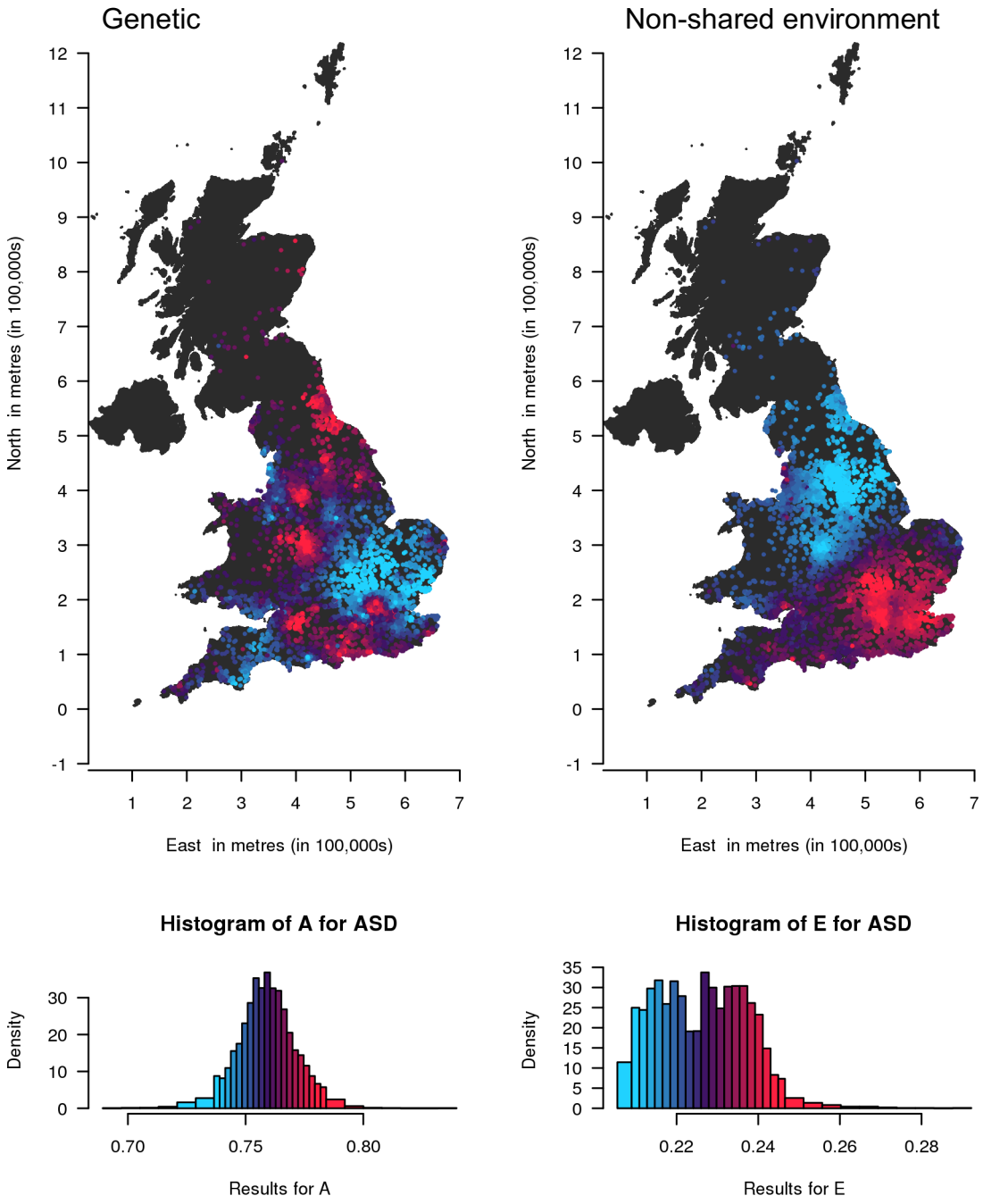
Population-level sex limitation model results for autistic traits are shown in **supplementary table 2**. We used nested sub-models to test for sex differences in aetiology. The common effects model is the most parsimonious model that adequately fits the data, indicating quantitative, but not qualitative, sex differences. However, because this is a large and well-powered sample, the parameter estimates for males and females that are "significantly" different at an alpha of 0.05 are actually within 1% of each other. To maximise power, we have presented the maps for males and females

combined in the main text, but maps for males and females separately are shown in **supplementary figures 4 and 5**.

Mapping the aetiology of autistic traits in the UK

Figure 4 maps genetic and environmental influences on autistic traits in the UK at 6,758 locations chosen to represent sample density across the UK (an interactive version of this map is available at https://dynamicgenetics.github.io/spACEjs/). Again, this is a map of the raw variance, so A, C and E are not constrained to add up to one. However, maps with A, C and E constrained to add to one at each location are shown in **supplementary figure 6**, with very similar results.

Figure 4. Similar to findings in Sweden, genetic influences for autistic traits in the UK appear more influential in more densely populated areas, although patterns of non-shared environmental influences follow a north-south divide



386

387 *Geographical variation in genetic (A) and non-shared (E) influences on childhood autistic traits in the UK.*
388 *The contributions of A and E range from low (blue) to high (red). The histograms below show the*
389 *distribution of the estimates, coloured in the same way as the points on the map. The A and E estimates are*
390 *not standardised and are therefore not constrained to add up to one at each location. An interactive version*
391 *of this map is available at https://dynamicgenetics.github.io/spACEjs/.*

The raw results for A are consistent with those from the CATSS sample in Sweden where we observed higher heritability in more densely populated areas. The mean of A is slightly higher in the UK than in Sweden: 0.76 (SD=0.01) compared to 0.65 (SD=0.02). For non-shared environment (E) the patterns are less similar across countries, as are the mean values 0.23 (SD=0.01) in the UK, compared to 0.34 (SD=0.02) in Sweden. London, the capital city, and the surrounding south-east of the UK show greater influence of E compared to the north and some regions in the mid-west of England and Wales. In contrast, Sweden's capital, Stockholm, and the surrounding areas show lower estimates of E. Again, C is approximately zero for autistic traits across all regions. As before, local variation in A and E is apparent within large cities such as London. As the histograms show, A is fairly normally distributed between 0.69 and 0.84 across regions. E ranges more narrowly from 0.21 to 0.29 in a bimodal distribution with a positive skew. Confidence intervals for estimates at each location are provided in **supplementary table 3.**

Discussion

In this study we looked at how genetic and environmental influences on symptoms of ASD vary geographically in Sweden and the UK. Our results are consistent with previous population-level estimates of genetic and environmental influences, and demonstrate geographical variation in genetic and non-shared environmental influences on autistic traits in Sweden and the UK.

These geographical differences in genetic and environmental influences on autistic traits are indicative of gene-environment and environment-environment interactions where the interacting environmental variable varies by location. Where we find areas of increased genetic or environmental influences for autistic traits this means that the environment in these areas draws out genetic or environmental influence, in the same way that the presence of airborne pollen would reveal individual differences in genetic risk for hay fever. By studying this in a systematic way, rather than relying on a specific measured environment, we can use our results to develop novel hypotheses about currently unknown environmental influences.

Our findings complement previous research that has focused on geographical prevalence differences in ASD [9–14,38]. Similarly, alongside aetiological differences, we observe geographical variation in mean autistic trait scores. These mean differences may be linked to aetiological differences. For example, areas of greater prevalence could represent regions where the environment triggers genetic predisposition to ASD traits. This provides a basis for future research into specific geographically distributed environments that draw out or mitigate genetic or environmental risk, which could in turn be useful for population health measures seeking to reduce the impact of ASD.

From our results we can hypothesise about what these factors could be. For example, we find that there is generally higher heritability in more densely populated areas of Sweden, such as Gothenburg, Malmö, and Stockholm, and in a band running from Gothenburg to Stockholm where the majority of the population live. We find lower heritability in the southern highlands and northern regions, which are less populated. This may suggest that urban environments draw out genetic differences in predisposition to autistic traits between people of the same ancestral background. These geographically distributed environments might include psychosocial factors such as the stress of urban living or income inequality, or aspects of the physical environment such as air pollution. This explanation fits with neuroscience literature that suggests that living in an urban environment is associated with specific neural correlates in response to stress, which may influence the onset of related mental health disorders [25]. The literature on prevalence suggests that other potentially important factors may include geographical differences in access to healthcare, diagnostic bias and parental awareness, socio-economic status, neighbourhood deprivation, infrastructure of the area, or access to green space. However, factors such as rater effects or access to healthcare are less likely to play a role in this aetiological variation as we have used data from structured interviews in population representative samples, and environmental influences on prevalence are not necessarily the same as environmental influences on aetiology. For non-shared environmental influences urban-rural differences are confined to areas in and around Stockholm, the Swedish capital. Therefore, it may be that there are environments related specifically to living in or around the capital that result in decreased non-shared environmental influences compared to other areas in Sweden.

We see similar patterns for genetic influences on in the UK, with higher heritability estimates in city areas such as central and south London, Birmingham, Bristol, Manchester, Newcastle. Estimates are generally lower in East Anglia, the south west, Wales and other less densely populated areas. Again, as in Sweden, non-shared environmental influence shows a more complex pattern in the UK.

Whilst we see similarities in patterns of aetiology between Sweden and the UK for autistic traits, there are also substantial differences. There are several possible reasons for this. For example, it could be due to differences in the measurement of ASD symptoms in the cohorts, or it could be due to environmental differences between the two countries, for example differences in the level of awareness of ASD and therefore possible differential reporting in ASD symptoms, or differences in the physical or social environments, which may vary between countries in the same way as they do within each country. It will be important to investigate this in other countries to explore these international similarities and differences further.

When interpreting these results there are a few important points to consider. First, in some areas the effective sample size is lower than others, for example in densely populated areas the proximity of some twin pairs relative to others can weight their influence relatively highly. However, across all areas we have taken care to maintain effective sample sizes in the thousands for both identical and fraternal twin pairs, so estimates remain reasonably precise. Second, due to how the weighting of participants' contributions to the analyses works, i.e. participants contribute more to analysis the closer they are to the target location, this results in smoothing over the estimates for A, C and E. The amount by which results over the area are smoothed depends on the tuning parameter used in the weighting. There is a trade-off when selecting the tuning

parameter between smoothing over noise and detecting real variation, or between accurately estimating variance components and accurately localising them. Here, we have chosen the tuning parameter to result in some smoothing towards the population mean, but this may mean that some larger localised variation remains undetected. In interpreting the maps, it is important to take into account both the pattern of results shown on the map, and the range of estimates shown by the histogram, while bearing in mind that the effect sizes are smoothed towards the population mean. Third, in common with the previous literature, we find that an ADE model is often a slightly better fit to the data, but here we have fitted ACE models and generally presented results for A and E alone because the high correlation between A and D brings noise to spatial analysis due to switching between the two across locations [39]. Instead, we interpret A here as a broad genetic component, without the usual connotation of additivity. Fourth, as with any statistical analysis, it is important to consider the assumptions of the model. For twin modelling, these include random mating within the population, that MZ and DZ twins share their environments to the same extent (at least where those environments are not genetically influenced), and that twins are representative of the general population for the traits studied [35]. These assumptions have generally been found to be reasonable [40], although there is some evidence to suggest that there is assortative mating for ASD, for example a study in Sweden that found phenotypic correlations of 0.48 for ASD [41]. This would have the effect of inflating the shared environmental influences, which we find to be approximately zero across locations. For our geographical analyses we do not assume that there is no gene-environment interaction or correlation, because we are explicitly modelling them as our main point of interest.

Our systematic analysis shows geographical variation in genetic and non-shared environmental influences for symptoms of ASD in both Sweden and the UK. These results will inform further studies of measured geographically distributed environments, beyond those already identified as influencing prevalence in the literature. For example, by correlating the spatial distribution of these environments with the spatial distribution of the aetiological estimates or by using formal continuous moderator models. Identifying these environments and understanding how they draw out or mask genetic predisposition may lead to population health and social policy innovation to support people with ASD.

509 **References**

510 1. Lundström, S., Reichenberg, A., Anckarsäter, H., Lichtenstein, P. & Gillberg, C.
511 Autism phenotype versus registered diagnosis in Swedish children: Prevalence
512 trends over 10 years in general population samples. *BMJ* **350**, h1961 (2015).

513 2. Brugha, T. S. *et al.* Epidemiology of autism spectrum disorders in adults in the
514 community in England. *Arch. Gen. Psychiatry* **68**, 459–466 (2011).

515 3. Idring, S. *et al.* Changes in Prevalence of Autism Spectrum Disorders in 2001–
516 2011: Findings from the Stockholm Youth Cohort. *J. Autism Dev. Disord.* **45**, 1766–
517 1773 (2015).

518 4. Lyall, K. *et al.* The Changing Epidemiology of Autism Spectrum Disorders. *Annu.*
519 *Rev. Public Health* **38**, 81–102 (2017).

520 5. Seltzer, M. M., Shattuck, P., Abbeduto, L. & Greenberg, J. S. Trajectory of
521 development in adolescents and adults with autism. *Ment. Retard. Dev. Disabil.*
522 *Res. Rev.* **10**, 234–247 (2004).

523 6. Tick, B., Bolton, P., Happé, F., Rutter, M. & Rijsdijk, F. Heritability of autism
524 spectrum disorders: A meta-analysis of twin studies. *J. Child Psychol. Psychiatry*
525 *Allied Discip.* **57**, 585–595 (2016).

526 7. Sandin, S. *et al.* The Heritability of Autism Spectrum Disorder. *JAMA* **318**, 1182–
527 1184 (2017).

528 8. Bai, D. *et al.* Association of Genetic and Environmental Factors With Autism in a 5-

Country Cohort. *JAMA Psychiatry* 1–9 (2019). doi:10.1001/jamapsychiatry.2019.1411

9. Bakian, A. V., Bilder, D. A., Coon, H. & McMahon, W. M. Spatial Relative Risk Patterns of Autism Spectrum Disorders in Utah. *J. Autism Dev. Disord.* **45**, 988–1000 (2015).

10. Chen, C. Y., Liu, C. Y., Su, W. C., Huang, S. L. & Lin, K. M. Urbanicity-related variation in help-seeking and services utilization among preschool-age children with autism in Taiwan. *J. Autism Dev. Disord.* **38**, 489–497 (2008).

11. Mazumdar, S., King, M., Liu, K. Y., Zerubavel, N. & Bearman, P. The spatial structure of autism in California, 1993-2001. *Heal. Place* **16**, 539–546 (2010).

12. Mazumdar, S., Winter, A., Liu, K. Y. & Bearman, P. Spatial clusters of autism births and diagnoses point to contextual drivers ofincreased prevalence. *Soc. Sci. Med.* **95**, 87–96 (2013).

13. Hoffman, K. *et al.* Geographic Patterns of Autism Spectrum Disorder Among Children of Participants in Nurses ' Health Study II. *Am. J. Epidemiol.* **186**, 834–842 (2017).

14. Campbell, M., Reynolds, L., Cunningham, J., Minnis, H. & Gillberg, C. Autism in Glasgow: Cumulative incidence and the effects of referral age, deprivation and geographical location. *Child. Care. Health Dev.* **39**, 688–694 (2011).

15. Lauritsen, M. B., Pedersen, C. B. & Mortensen, P. B. Effects of familial risk factors

549 and place of birth on the risk of autism: A nationwide register-based study. *J. Child*
550 *Psychol. Psychiatry* **46**, 963–971 (2005).

551 16. Lauritsen, M. B. *et al.* Urbanicity and autism spectrum disorders. *J. Autism Dev.*
552 *Disord.* **44**, 394–404 (2014).

553 17. Wu, J. & Jackson, L. Inverse relationship between urban green space and
554 childhood autism in California elementary school districts. *Environ. Int.* **107**, 140–
555 146 (2017).

556 18. Vassos, E., Agerbo, E., Mors, O. & Bøcker Pedersen, C. Urban-rural differences in
557 incidence rates of psychiatric disorders in Denmark. *Br. J. Psychiatry* **208**, 435–
558 440 (2016).

559 19. Davis, O. S. P., Haworth, C. M. A., Lewis, C. M. & Plomin, R. Visual analysis of
560 geocoded twin data puts nature and nurture on the map. *Mol. Psychiatry* **17**, 867–
561 874 (2012).

562 20. Colvert, E. *et al.* Heritability of autism spectrum disorder in a UK population-
563 based twin sample. *JAMA Psychiatry* **72**, 415–423 (2015).

564 21. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and
565 neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555
566 (2016).

567 22. Robinson, E. B. *et al.* Evidence that autistic traits show the same etiology in the
568 general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch.*

Gen. Psychiatry **68**, 1113–1121 (2011).

23. Lundström, S. *et al.* Autism spectrum disorders and autisticlike traits: Similar etiology in the extreme end and the normal variation. *Arch. Gen. Psychiatry* **69**, 46–52 (2012).

24. Ronald, A. *et al.* Genetic heterogeneity between the three components of the autism spectrum: A twin study. *J. Am. Acad. Child Adolesc. Psychiatry* **45**, 691–699 (2006).

25. Lederbogen, F. *et al.* City living and urban upbringing affect neural social stress processing in humans. *Nature* **474**, 498–501 (2011).

26. Magnusson, P. K. E. *et al.* The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments. *Twin Res. Hum. Genet.* **16**, 317–329 (2012).

27. Anckarsäter, H. *et al.* The Child and Adolescent Twin Study in Sweden (CATSS). *Twin Res. Hum. Genet.* **14**, 495–508 (2011).

28. Hansson, S. L. *et al.* Psychiatric telephone interview with parents for screening of childhood autism - Tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): Preliminary reliability and validity. *Br. J. Psychiatry* **187**, 262–267 (2005).

29. Anckarsäter, H. *et al.* Child neurodevelopmental and behavioural problems are intercorrelated and dimensionally distributed in the general population. *The Open Psychiatry Journal* **2**, 5–11 (2008).

589 30. Mårland, C. *et al.* The Autism-Tics, ADHD and other Comorbidities inventory (A-
590 TAC): Previous and predictive validity. *BMC Psychiatry* **17**, 403 (2017).

591 31. Larson, T. *et al.* Predictive properties of the A-TAC inventory when screening for
592 childhood-onset neurodevelopmental problems in a population-based sample.
593 *BMC Psychiatry* **13**, 233 (2013).

594 32. Larson, T. *et al.* The Autism-Tics, AD/HD and other Comorbidities inventory (A-
595 TAC): further validation of a telephone interview for epidemiological research.
596 *BMC Psychiatry* **10**, 1 (2010).

597 33. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study
598 (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioural
599 Development From Childhood to Young Adulthood. *Twin Res. Hum. Genet.* **16**,
600 117–125 (2013).

601 34. Scott, F. J., Baron-Cohen, S., Bolton, P. & Brayne, C. The CAST (Childhood Asperger
602 Syndrome Test): preliminary development of a UK screen for mainstream
603 primary-school-age children. *Autism* **6**, 9–31 (2002).

604 35. Rijsdijk, F. V & Sham, P. C. Analytic approaches to twin data using structural
605 equation models. *Brief. Bioinform.* **3**, 119–133 (2002).

606 36. Constantino, J. N. & Todd, R. D. Autistic Traits in the General Population. *Arch. Gen.*
607 *Psychiatry* **60**, 524 (2003).

608 37. Ronald, A., Larsson, H., Anckarsäter, H. & Lichtenstein, P. A twin study of autism

609 symptoms in Sweden. *Mol. Psychiatry* **16**, 1039–1047 (2011).

610 38. Hoffman, K., Kalkbrenner, A. E., Vieira, V. M. & Daniels, J. L. The spatial distribution
611 of known predictors of autism spectrum disorders impacts geographic variability
612 in prevalence in central North Carolina. *Environ. Heal.* **11**, 80 (2012).

613 39. Rietveld, M. J. H., Posthuma, D., Dolan, C. V. & Boomsma, D. I. ADHD: Sibling
614 interaction or dominance: An evaluation of statistical power. *Behav. Genet.* **33**,
615 247–255 (2003).

616 40. Evans, D. M. & Martin, N. G. The validity of twin studies. *Genescreen* **1**, 77 (2000).

617 41. Nordsletten, A. E. *et al.* Patterns of nonrandom mating within and across 11 major
618 psychiatric disorders. *JAMA Psychiatry* **73**, 354–361 (2016).

619

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the TEDS and CATSS and their research teams. This work was supported in part by the UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (Grant ref: MC_UU_12013/1). ZR is supported by a Wellcome Trust PhD studentship (Grant ref: 109104/Z/15/Z). OSPD and CMAH are funded by the Alan Turing Institute under the EPSRC grant EP/N510129/1. This study was also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol (BRC-1215-2011). The CATSS study is supported by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), the Swedish Council for Health, Working Life and Welfare, the Söderström-Königska Foundation, and the Swedish Research Council (Medicine and SIMSAM). TEDS is funded by Medical Research Council grant MR/M021475/1 to Robert Plomin. AbR is supported by the Beatrice and Samuel L Seaver Foundation.

Author contributions

Conceptualization: ZER, OSPD; Methodology: ZER and OSPD; Software: ZER and OSPD; Formal Analysis: ZER; Resources: OSPD, CMAH, ARo and PL; Data Curation: ZER; Writing—Original Draft: ZER; Writing—Review and Editing: OSPD, HL, CMAH, DR, SL, AnR, AbR and PL; Visualization: ZER and OSPD; Supervision: OSPD; Project Administration: OSPD; Funding Acquisition: OSPD.

641 **Competing interests**

642 Angelica Ronald has received a consultancy fee for writing for the National Childbirth
643 Trust and receives an annual honorarium as joint editor of the Journal of Child
644 Psychology and Psychiatry. H. Larsson has served as a speaker for Evolan Pharma and
645 Shire and has received research grants from Shire; all outside the submitted work.

646