UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Investigating genetic links between grapheme–colour synaesthesia and neuropsychiatric traits

Tilot, A.K.; Vino, A.; Kucera, K.S.; Carmichael, D.A.; van den Heuvel, L.; den Hoed, J.; Sidoroff-Dorso, A.V.; Campbell, A.; Porteous, D.J.; St Pourcain, B.; van Leeuwen, T.M.; Ward, J.; Rouw, R.; Simner, J.; Fisher, S.E.

DOI
10.1098/rstb.2019.0026

Publication date
2019

Document Version
Final published version

Published in
Philosophical Transactions of the Royal Society B - Biological Sciences

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

Citation for published version (APA):
Tilot, A. K., Vino, A., Kucera, K. S., Carmichael, D. A., van den Heuvel, L., den Hoed, J., Sidoroff-Dorso, A. V., Campbell, A., Porteous, D. J., St Pourcain, B., van Leeuwen, T. M., Ward, J., Rouw, R., Simner, J., & Fisher, S. E. (2019). Investigating genetic links between grapheme–colour synaesthesia and neuropsychiatric traits. *Philosophical Transactions of the Royal Society B - Biological Sciences*, *374*(1787), Article 20190026. https://doi.org/10.1098/rstb.2019.0026

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date:12 Sept 2026

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Tilot AK *et al*. 2019 Investigating genetic links between grapheme–colour synaesthesia and neuropsychiatric traits. *Phil. Trans. R. Soc. B* **374**: 20190026.
http://dx.doi.org/10.1098/rstb.2019.0026

Accepted: 23 August 2019

One contribution of 16 to a discussion meeting issue 'Bridging senses: novel insights from synaesthesia'.

Subject Areas:
genetics, neuroscience, genomics, cognition

Keywords:
synaesthesia/synesthesia, polygenic scores, autism, schizophrenia, Generation Scotland

Author for correspondence:
Simon E. Fisher
e-mail: simon.fisher@mpi.nl

Investigating genetic links between grapheme–colour synaesthesia and neuropsychiatric traits

Amanda K. Tilot[1,2], Arianna Vino[1], Katerina S. Kucera[1], Duncan A. Carmichael[3], Loes van den Heuvel[1], Joery den Hoed[1], Anton V. Sidoroff-Dorso[4], Archie Campbell[5,6], David J. Porteous[5,7], Beate St Pourcain[1], Tessa M. van Leeuwen[8], Jamie Ward[9], Romke Rouw[10], Julia Simner[9] and Simon E. Fisher[1,8]

[1]Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands
[2]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA
[3]School of Applied Sciences, Edinburgh Napier University, Sighthill Court, Edinburgh EH11 4BN, UK
[4]Department of Psychology, Moscow Pedagogical State University, Moscow, 119435, Russia
[5]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
[6]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH16 4UX, UK
[7]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK
[8]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 HE Nijmegen, The Netherlands
[9]School of Psychology, University of Sussex, Brighton BN1 9RH, UK
[10]Department of Psychology, University of Amsterdam, 1018 WT Amsterdam, The Netherlands

AKT, 0000-0002-9107-8190; JdH, 0000-0001-6614-876X; TMvL, 0000-0001-7810-6348; JW, 0000-0001-7007-1902; RR, 0000-0001-9697-5360; JS, 0000-0003-4789-0138; SEF, 0000-0002-3132-1996

Synaesthesia is a neurological phenomenon affecting perception, where triggering stimuli (e.g. letters and numbers) elicit unusual secondary sensory experiences (e.g. colours). Family-based studies point to a role for genetic factors in the development of this trait. However, the contributions of common genomic variation to synaesthesia have not yet been investigated. Here, we present the SynGenes cohort, the largest genotyped collection of unrelated people with grapheme–colour synaesthesia ($n = 723$). Synaesthesia has been associated with a range of other neuropsychological traits, including enhanced memory and mental imagery, as well as greater sensory sensitivity. Motivated by the prior literature on putative trait overlaps, we investigated polygenic scores derived from published genome-wide scans of schizophrenia and autism spectrum disorder (ASD), comparing our SynGenes cohort to 2181 non-synaesthetic controls. We found a very slight association between schizophrenia polygenic scores and synaesthesia (Nagelkerke's $R^2 = 0.0047$, empirical $p = 0.0027$) and no significant association for scores related to ASD (Nagelkerke's $R^2 = 0.00092$, empirical $p = 0.54$) or body mass index ($R^2 = 0.00058$, empirical $p = 0.60$), included as a negative control. As sample sizes for studying common genomic variation continue to increase, genetic investigations of the kind reported here may yield novel insights into the shared biology between synaesthesia and other traits, to complement findings from neuropsychology and brain imaging.

This article is part of a discussion meeting issue 'Bridging senses: novel insights from synaesthesia'.

1. Introduction

Synaesthesia is a neurological phenomenon at the edge of natural variation in sensory perception. Individuals with synaesthesia (up to approx. 4% of the population) have unusual secondary sensory experiences when presented

© 2019 The Author(s) Published by the Royal Society. All rights reserved.

THE ROYAL SOCIETY PUBLISHING

with triggering stimuli, typically against a background of normal neural development. A diverse range of such experiences have been documented; for example, some people with synaesthesia visualize the days of the week or sequences of numbers having ordered locations in the space around them (sequence–space synaesthesia) while for others vivid taste perceptions are elicited by words (lexical–gustatory synaesthesia) [1–3]. Grapheme–colour synaesthesia is a relatively common form, with an estimated prevalence of 1.4% [1]. In this type of synaesthesia, which has been widely studied over the past 20 years, letters and numbers trigger the consistent perception of specific colours [4].

Evidence from a range of sources suggests that synaesthesia is influenced by genetic factors. The precise nature of these factors is poorly understood, but it is thought that several mechanisms are involved, ranging from rare DNA variants with large effects (more akin to monogenic inheritance) to combined actions of common polymorphisms, each with only a small effect on the trait. Investigations of families where multiple relatives are synaesthetic indicate that, even where rare variation could be responsible, there is substantial genetic heterogeneity, meaning that distinct genetic loci may be involved in different families [5–9]. Against this background, we previously performed a whole-exome sequencing study and found enrichment of rare variants in genes associated with axonogenesis in three families with sound–colour synaesthesia [9]. These results supported a long-standing hypothesis that synaesthesia may be caused in part by altered or hyperconnectivity between brain regions processing the inducing and concurrent sensory stimuli [10,11].

While studies of familial synaesthesia are beginning to show promise for pinpointing contributions of rare gene variants, a large proportion of synaesthetes are unaware of other relatives with similar experiences. Indeed, virtually nothing is known at present about the potential role of common genetic variation in synaesthesia. We sought to address this major gap in the literature. Since investigations of contributions of common polymorphisms require large sample sizes, in 2013, we initiated a new effort (SynGenes) to systematically recruit and genotype unrelated individuals with objectively verified grapheme–colour synaesthesia. Genome-wide association studies (GWAS) are often used as a tool for assessing the contributions of common genetic variants to a quantitative or binary trait. A typical GWAS involves determining the genotypes of hundreds of thousands of single nucleotide polymorphisms (SNPs) at different positions through a person's genome and then systematically testing each SNP to see if there is an association between the status of the allele and the trait of interest, in the cohort as a whole. Given the small effect sizes of individual common variants, and the multiple-testing burden involved, a GWAS becomes informative when the sample size of the cohort reaches several thousand individuals or cases. Two decades of human population genetics research has shown that, as well as being underpowered, a GWAS performed on smaller sample sizes can yield spurious results that fail to replicate in larger cohorts [12].

It is still possible to investigate the genetic architecture of a trait when sample sizes are too small for a full-scale GWAS, through more recently developed methods that use aggregated genotype information from multiple loci across the genome to examine how two traits are related at a genetic level. Specifically, alleles associated with increased risk for a binary trait (or associated with increases in a quantitative trait, like height) can be aggregated into a polygenic risk score (PGS) to assess genetic association with other phenotypes [13,14]. A PGS weights the number of independent 'increaser' alleles that an individual carries at different SNPs in their genome with the respective effect sizes, as measured in an independent GWAS with sufficient power. This aggregation of information across the multiple markers yields a score that reflects the individual's net balance of alleles that increase or decrease risk for the binary trait, or that increase/decrease scores on a quantitative trait (as in the height example). A recent study on the application of schizophrenia PGS across populations found significant variation between groups with differences in ancestry, leading to best practice recommendations that a PGS derived from the GWAS of one population (most commonly Europeans) should only be applied to participants from that same population [14,15]. Despite this limitation, the method is increasingly being used to link quantitative measurements of brain-related traits (e.g. executive functioning, neuroticism) in the normal population to genetic risk for disorders [16]. This provides a feasible way to study the genetic relationships between one trait for which there are large-scale GWAS data available and another trait that is measured in a cohort of more limited size. For the former, this is usually a trait that has been the focus of a meta-analytic effort across multiple cohorts in a consortium, such as the Psychiatric Genomics Consortium (PGC) [17].

With these limitations in mind, we sought to understand whether neuropsychological traits that have been previously linked to synaesthesia at a phenotypic level have any deeper, genetic relationships. Beyond the synaesthetic experiences themselves, there is considerable literature on other ways that synaesthetes differ from non-synaesthetes. These include cognitive traits like improved memory performance in synaesthetes, and perceptual traits like increased sensory sensitivity and mental imagery, and as well as greater positive schizotypy as measured through self-reported unusual perceptual experiences [18,19]. There is also an increased prevalence of synaesthesia among individuals with autism spectrum disorder (ASD) [20–22].

When selecting traits for inclusion in the current study, we were limited to traits with relevant and sufficiently powered GWAS. While schizophrenia and ASD are complex at both the phenotypic and genotypic levels, they each include an element of unusual perceptual differences that may show overlapping genetic architecture with synaesthesia in neurotypical individuals. Such a shared genetic component could be observable as synaesthetes having higher PGSs for either trait. Although schizotypy has not been the subject of a GWAS, one study found that individuals with a higher PGS for schizophrenia (derived from GWAS sample sizes that are greater than 30 000) also had higher levels of positive and negative schizotypy [23]. GWAS efforts in ASD are just beginning to bear fruit, with a recent meta-analysis of 18 381 cases identifying five genome-wide significant loci [24].

The current study presents the first set of results from the SynGenes cohort, the largest collection of genotyped individuals with validated grapheme–colour synaesthesia, involving genetic analyses of over 700 unrelated people. We used PGS analyses to assess whether an individual's aggregate genetic risk for schizophrenia or ASD influences the likelihood that they experience synaesthesia. To confirm that our case–control design was free from confounding differences in genetic ancestry, we also looked for differences in polygenic scores for body mass index (BMI), using this heritable trait as a negative

control. We found that PGSs for schizophrenia are likely to explain a very small amount (less than 1%) of the variance in synaesthesia status, an effect similar in size to the relationship between schizophrenia PGS and creativity, reported in a prior study [25]. We did not observe any relationships between PGSs for ASD or BMI and synaesthesia. Our results suggest that PGSs for schizophrenia do not have a meaningful impact on whether a person experiences synaesthesia, while at the same time identifying a small piece of shared biology, perhaps related to unusual perceptual experiences, that may tie this intriguing sensory phenomenon into the fabric of modern psychiatric genetics.

2. Methods

(a) Participant recruitment

Synaesthete participants were recruited at the Max Planck Institute for Psycholinguistics (MPI) through multiple routes. These included social media (e.g. Twitter, posts in synaesthesia-related Facebook groups, a Reddit 'Ask Me Anything' event in August 2016), flyers posted at Radboud University, advertising to the Russian synaesthesia community database, and recontacting consenting individuals who had previously participated in synaesthesia research through the Groot Nationaal Onderzoek (in English, 'Large National Study', http://gno.mpi.nl/tests) [26], the University of Amsterdam, or the University of Sussex. Those recruited to the genetics study directly at the University of Edinburgh were contacted through similar methods, including online advertisements, student forms and mailing lists, and recontacting previous participants.

Participants initially recruited by the MPI were asked to read a participant information sheet and provide their informed consent prior to completing initial surveys and synaesthesia testing. During the genetic phase of the study, participants were again provided with the information sheet as well as informed consent forms. Participants aged between 12 and 18 years were additionally asked for parent or guardian consent, and the consent forms used for participants aged between 4 and 12 years asked the parent or guardian for consent on behalf of the child. Ethical approval for the study was granted by the Ethische Commissie Gedragswetenschappelijk Onderzoek (Ethics Committee, Faculty of Social Sciences) at Radboud University (application number ECG2013-2504-105). Participants from the above sources are collectively referred to as the 'SynGenes' cohort.

Both synaesthetes and controls were recruited from the Scottish Family Health Study, a cohort of over 20 000 individuals with genetic and health data collected as part of Generation Scotland [27]. The advertisement for the study included a potential reward of £100 via a prize draw. All participants provided written informed consent, and ethical approval for the Generation Scotland project was granted by the Tayside Committee on Medical Research Ethics, on behalf of the National Health Service (reference number: 05/S1401/89). The recruitment and data collection process for Generation Scotland has been described in detail elsewhere [27–29].

Additional non-synaesthetic and population controls were included from the Nijmegen 'Brain Imaging Genetics' cohort (BIG), a Dutch population-based sample of healthy volunteers that has been described in detail in several previous studies [30–32]. The BIG study was approved by the regional ethics committee, and all participants provided written informed consent.

(b) Questionnaires and synaesthesia consistency testing

Participants directly recruited by the Max Planck Institute were able to join the study through three routes: *The Synesthesia Battery* (hosted by Baylor College of Medicine until 2017, later by the University of Sussex); a web-based survey hosted by the MPI (available at www.mpi.nl/synaesthesia); or the *SynQuiz* app (available on iOS and Android) [33,34]. Each included a survey of synaesthesia types, including an open field for other forms not listed, basic demographics (age, gender) and questions about whether the participant experienced any potentially relevant neurological conditions that might mimic the symptoms of synaesthesia (e.g. headaches, migraines and epilepsy). There was also a final open box for participants to share anything else that they felt was relevant to their experience with synaesthesia. Participants who reported that their synaesthetic experiences resulted from a non-developmental trigger (e.g. psychoactive drug use or epilepsy) were excluded.

Generation Scotland participants who consented to being recontacted were invited to join the current study by email, including a link to synaesthesia diagnostic tests hosted by the University of Sussex. Demographic and relevant health information was available through previous surveys conducted by Generation Scotland.

BIG participants were asked a screening question about whether or not they thought that they experienced synaesthesia, as part of a larger survey. Those who replied that they did not experience synaesthesia were eligible for the non-synaesthetic control group. If the synaesthesia questionnaire information was not available, participants were considered 'synaesthesia status unknown' controls (table 2).

In the MPI cohort and Generation Scotland participants who replied positively to a screening question about synaesthesia, grapheme–colour synaesthesia was assessed through a commonly used and validated online diagnostic test (known as the 'test of consistency'; see below) [33,34]. Letters (A–Z) and numbers (0–9) are randomized and presented to the participant along with a colour palette. Participants must choose the colour they feel most closely matches their synaesthetic association, and each grapheme is presented three times during the test. The test is scored by calculating the distance in colour space between the three colours chosen, with smaller distances reflecting more similar colour choices. For example, similar shades of red would generate a low score, while a combination of pale pink, bright red and deep purple would produce a higher score. Lower scores thus indicate greater consistency in colour selection, a known diagnostic feature of synaesthesia [35]. If a non-native English-speaker completed the grapheme–colour consistency test through *The Synesthesia Battery*, they had the option of using a non-Roman alphabet (e.g. Cyrillic or Hebrew).

Participants whose consistency test scores were collected by the MPI were considered synaesthetic if they scored below 1.5 on the consistency test, and if at least 15 graphemes were linked to colour sensations (thus only having synaesthetic experiences for numbers 0–9 would be insufficient). Participants for whom synaesthesia status was confirmed as part of prior studies at the University of Sussex and the University of Edinburgh qualified for this genetics study with a grapheme–colour consistency score of 1.43 or less (following [36]). Participants from the Groot Nationaal Onderzoek were invited to the genetics portion of the study if they scored below 1.35 on the consistency test. University of Amsterdam participants needed to pass the grapheme–colour test within *The Synesthesia Battery* with a score of 1.0 or lower.

Within Generation Scotland, any individuals who reported themselves as synaesthetic but failed the consistency test (scoring above 1.5) were excluded from both the synaesthesia cases and control group.

The number of types of synaesthesia a participant experienced was calculated for all participants where we had access to a complete survey of synaesthesia types. Following the categories described by Novich *et al.* [37], synaesthesia forms were grouped into five clusters [37]. Self-report of one or more forms of synaesthesia within a category added that category to an individual's score, for a maximum score of five (experiencing at least one form of synaesthesia from each category).

(c) DNA sampling

Participants who met the criteria for grapheme–colour synaesthesia (see above for site-specific thresholds) were invited to join the genetics portion of the study. At the MPI, mailing addresses were requested via email. Those who replied with their address were sent an Oragene DNA OG-500 saliva collection kit, along with the participant information sheet, an ancestry survey and informed consent documents. At the University of Edinburgh, saliva samples were collected onsite using the same Oragene kits. All saliva kits were processed at the MPI according to the manufacturer's instructions. Kits that yielded an insufficient amount of DNA for genotyping were excluded from further analysis.

BIG participants contributed saliva samples using Oragene kits as part of initial recruitment [30]. In the Generation Scotland project, DNA was isolated from blood as previously described [27].

(d) Genotyping and pre-imputation quality control

A total of 768 DNA samples processed at the MPI were genotyped in two batches using the Illumina Human OmniExpressExome genotyping array. Initial quality control steps were performed using Illumina GenomeStudio software (v.2.0), following the protocol by Guo *et al*. [38], including clustering, removing all samples with genotyping rates less than 98% and any SNPs that were missing in more than 5% of samples (two samples were removed at this step) [38]. The resulting genotypes were exported in PLINK bfile format, and the remaining pre-imputation quality control steps were performed using PLINK v.1.9 and 2.0. At the SNP-level, the data were further filtered by removing SNPs that were out of Hardy–Weinberg equilibrium ($pHWE < 1 \times 10^{-5}$).

At the subject-level, samples were checked for sex-mismatches (a potential indicator of a sample processing error; none were found). The process for recruiting synaesthetes did not exclude family members from participating as long as they passed the consistency test, and so a small number of synaesthetic parents and siblings were genotyped. As close relatives would artificially skew the case–control comparisons owing to their genetic similarity, one individual from each pair of relatives was removed. This was done by computing the proportion of identity-by-descent (pi-hat) for every pair of participants, and removing one individual from each pair where the pi-hat value was greater than 0.185 [39]. This threshold identifies relatives between second and third degree, and a total of 19 samples were removed at this step. Finally, as the current study is focused on participants with European ancestry, EIGENSTRAT v.6.1.4 was used (following the protocol outlined in [38]) to identify and remove non-European individuals via principal component analysis (PCA), leading to the removal of 23 samples.

The two genotyping batches were combined prior to imputation, and the PLINK .bim files were checked against the Human Reference Consortium (HRC) reference SNP list using scripts provided by William Rayner (https://www.well.ox.ac.uk/~wrayner/tools/, v.4.2.9) [40].

DNA samples from the BIG cohort were genotyped in three batches, using the Affymetrix 6.0, Illumina OmniExpress and Psychiatric Genetics Consortium PsychChip arrays [32]. Pre-imputation quality control was performed in PLINK as above, and the batches were processed separately. Generation Scotland genotype data were previously cleaned for missingness, and close relatives (12 396 out of 20 032 original subjects) were removed using PLINK 1.9 as above. Both the BIG and Generation Scotland genotype data were checked for alignment with the HRC reference list prior to imputation.

(e) Imputation and final quality control

To ensure the maximum number of overlapping SNPs between cohorts, each set of genotyping data was imputed using the Michigan Imputation Server using the HRC r1.1 (2016) (see https://imputationserver.sph.umich.edu/start.html#!pages/hrc-r1.1) imputation panel and SHAPEIT v2.r790 [41] for phasing [42]. Poorly imputed SNPs ($R^2 < 0.8$) were removed and post-imputation quality control filters (SNP or sample missingness greater than 10% and/or $pHWE < 1 \times 10^{-6}$) were applied to each cohort separately. The cohorts were merged, removing any SNPs that were multi-allelic or missing in one or more cohorts, resulting in a final set of well-imputed SNPs that were present in all cohorts. The combined dataset was given a final check for closely related individuals to guard against the possibility that a participant from the SynGenes cohort had a close relative who participated in Generation Scotland or BIG—none were found.

Table 1. Participants with European ancestry in SynGenes, top 10 countries of origin.

country	number of participants
UK	199
The Netherlands	179
USA	117
Germany	38
Canada	29
Australia	23
Russia	19
Switzerland	12
Belgium	9
Italy	8
other	90

Table 2. Participant demographics.

cohort	*N*	syn[a]/non-syn/unknown	*n* female (%)	*n* Dutch (%)
SynGenes	723	723/0/0	621 (86)	191 (26)
Generation Scotland	1612	4/399/1209	1399 (87)	0 (0)
BIG	573	0/214/359	480 (84)	573 (100)
total	2908	727/613/1568	2500 (86)	764 (26)

[a]Synaesthetic.

(f) Control matching

The combined imputed genotype data (excluding regions with high linkage disequilibrium, e.g. the major histocompatibility complex region) were pruned for linkage disequilibrium and 20 principal components (ancestry PCs) were calculated using PLINK.

In order to choose non-synaesthete control samples that were as closely matched to the SynGenes cohort as possible, we first removed clear outlying samples from that cohort based on the first two ancestry PCs. One additional synaesthete sample was inadvertently dropped from the analysis at this point, bringing the final total to 723 SynGenes samples. Control samples from BIG and Generation Scotland were prioritized by initially including all known non-synaesthetes (replied 'no' to screening questions), followed by balancing the male/female, Dutch/non-Dutch ratios with 'population controls' for whom synaesthesia status was unknown (table 2). Within each subgroup (e.g. non-Dutch females), controls were randomly sampled so that the total number of controls was three times the number of cases

(oversampling for extra rigour). The ancestry PCs were recalculated using the final set of synaesthesia cases and controls, for use as covariates in the polygenic score calculations.

(g) Polygenic score calculation

GWAS summary statistics for ASD and schizophrenia were downloaded from the PGC (https://www.med.unc.edu/pgc/results-and-downloads). The ASD data come from the combined iPSYCH-PGC GWAS originally conducted in 2017 (18 381 ASD cases) and the schizophrenia dataset is from their 2018 combined study of schizophrenia and bipolar disorder (33 426 schizophrenia cases) [24,43]. Summary statistics for BMI are based on the UK Biobank sample of 361 194 men and women, and were downloaded from Benjamin Neale's laboratory (http://www.nealelab.is/uk-biobank). The GWAS effect measurements (odds ratios, betas) were aligned to the risk (schizophrenia, ASD) or trait (BMI) increasing alleles prior to calculating polygenic scores.

Polygenic scores for ASD, schizophrenia and BMI were calculated using the PRSice software package [13]. Following best practices for polygenic scoring, we included sex and 20 ancestry principal components as covariates [14]. Grapheme–colour synaesthesia was treated as a binary target phenotype with a prevalence of 0.014 [1]. The PRSice calculations excluded the major histocompatibility complex region and included running 10 000 permutations of the best-fitting model in order to generate an empirical *p*-value for the association between the GWAS trait and synaesthesia that is controlled for Type 1 error.

3. Results

(a) Generating a case–control sample for genetic studies of grapheme–colour synaesthesia

As synaesthetic experiences had not been surveyed in any large cohorts with existing genotype data, we first needed to recruit and collect DNA samples from a large number of participants with this condition. Synaesthesia takes a range of different forms, and it is not yet known whether similar genetic underpinnings are shared across them. Thus, to minimize heterogeneity for these first studies, we ascertained our participants based on one particular form of synaesthesia: grapheme–colour synaesthesia. Our choice was based on several pragmatic factors: (i) the availability of well-validated consistency tests that are reliable and can be completed online; (ii) its relatively high population prevalence (1.4%); and (iii) the fact that this form of synaesthesia is one of the more extensively studied in the prior literature, albeit not yet at the genetic level. All participants completed standard diagnostic (consistency) tests and met the recruitment site-specific threshold for grapheme–colour synaesthesia (scores below 1.5–1.0, see §2).

To complement this newly developed 'SynGenes' cohort, we also surveyed two existing cohorts about synaesthetic experiences. Generation Scotland participants who indicated that they experienced grapheme–colour synaesthesia were given the same consistency test as those in the SynGenes cohort. Although the total number of verified synaesthetes was low ($n = 4$), the survey also generated a pool of participants who almost certainly do not experience synaesthesia. As 25% of the final SynGenes cohort identified as Dutch (table 1), we also included control participants taken from the Nijmegen 'Brain Imaging Genetics' (BIG) cohort (see §2).

Polygenic scores are known to be sensitive to subtle population stratification [15] and we took several steps to match the synaesthesia cases and controls on ancestry. We used PCA to represent subjects' genotype data as 20 components that capture increasingly small fractions of the genetic differences among the subjects in the analysis. This technique can clearly separate individuals from different population backgrounds (figure 1*c*) and is frequently used as both a quality control step to remove obvious outliers and as a source of covariates to control for potentially confounding ancestry differences in GWAS and other population genetics methods.

Twenty-three subjects were removed from the SynGenes cohort owing to non-European ancestry based on PCA (see §2). We then used a further round of PCA with the combined set of SynGenes, BIG and Generation Scotland to remove significant outliers along the first two components (22 subjects removed, figure 1). Following genotype quality control (see §2) and this further PCA step, the final set of synaesthetes (SynGenes plus four synaesthetes from Generation Scotland) numbered 727 (table 2).

The control samples were selected by first including all remaining subjects who were known to be non-synaesthetic, followed by randomly sampling from the Generation Scotland and BIG subjects whose synaesthesia status was unknown, ensuring that the female : male and Dutch : non-Dutch ratios matched the synaesthesia cases (table 2).

(b) Assessing polygenic scores across synaesthetes and controls

In order to begin placing synaesthesia in the context of other brain-related traits with well-studied genetic bases, we chose to assess differences in polygenic scores between synaesthetes and controls. We focused the PGS analysis on two neuropsychiatric conditions previously linked to synaesthesia at the phenotypic level—schizophrenia and ASD. Levels of positive and disorganized schizotypy were higher in individuals who experienced synaesthetic associations with colour in a 2012 study by Banissy *et al.* [19]. The positive schizotypy finding was later replicated and found to be unrelated to synaesthetes' elevated capacity for visual imagery [44]. Multiple studies have demonstrated relationships between aspects of ASD and synaesthesia, including increased sensory sensitivity in synaesthetes and an increased prevalence of synaesthesia among people with ASD who also have savant skills [20–22].

Briefly, the analysis involved using logistic regression to predict synaesthesia status based on the PGS, with sex and 20 ancestry principal components included as covariates. For each trait (schizophrenia, ASD and BMI), we took the prior published GWAS data summary statistics and used that information to construct a PGS. Consistent with standard procedures, we varied the *p*-value threshold that we used for determining which SNPs from the GWAS should be included in the PGS, and compared its performance in the model. A stricter threshold means that the PGS is constructed with fewer SNPs and accounts for a smaller degree of variance in the trait tested in the GWAS. Lowering the threshold, to be more permissive in how many SNPs are included, has the potential to include more SNPs that contribute to the variance in the GWAS trait (figure 2*c*). The best-fitting PGS was the one that made the greatest improvement to the model's power to accurately predict synaesthesia status.

Using GWAS summary statistics from the Psychiatric Genetics Consortium's (PGC) most recent published study of schizophrenia and bipolar disorder (33 426 schizophrenia cases, mainly of European ancestry), we calculated

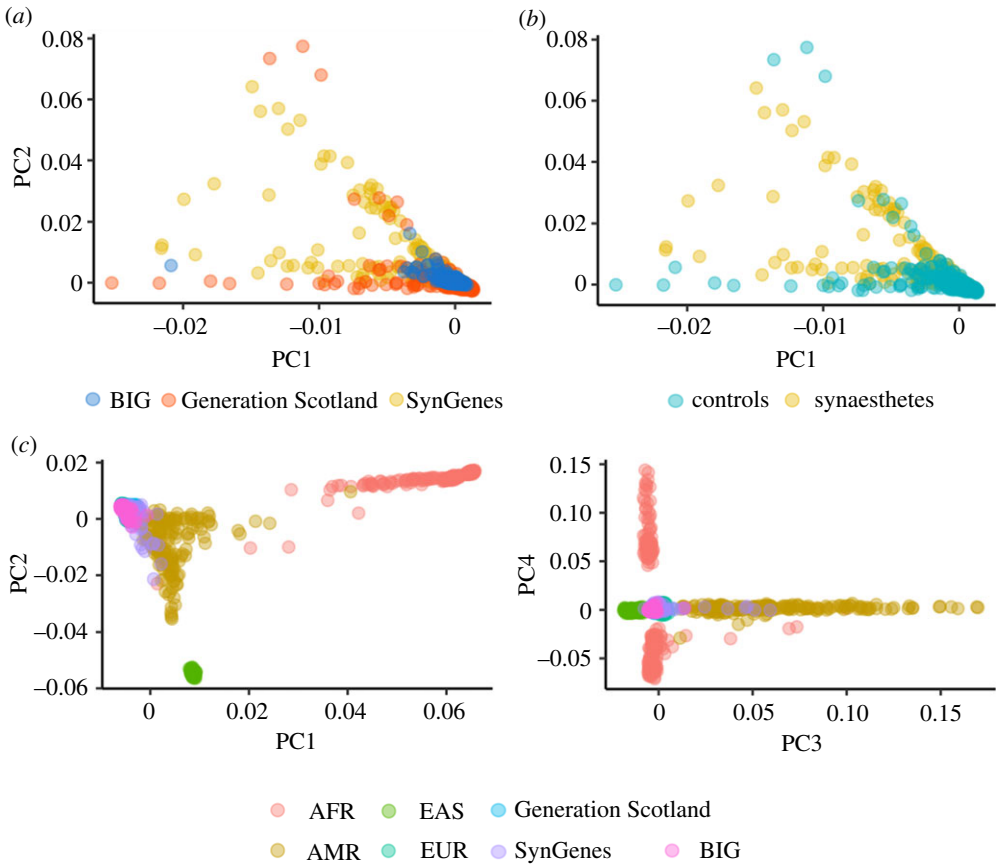

Figure 1. Final principal components (PCs) reflecting genetic variation owing to ancestry differences after matching synaesthetes and controls on self-reported ancestry. (*a*) The first two PCs for genetic ancestry, with colours indicating contributing cohorts (BIG = controls only; Generation Scotland = mainly controls; SynGenes = cases only). (*b*) The same PCs split by case/control status. (*c*) PCs 1–4 from a separate PCA that includes samples from the four major populations studied in the 1000 Genomes Project, to illustrate how the study cohorts compare to non-European populations (AFR, Africa; EAS, East Asian; AMR, Admixed American; EUR, European).

schizophrenia PGSs using multiple *p*-value thresholds [43]. The best-fitting model, assessing evidence for an association between schizophrenia PGS and synaesthesia, was identified for alleles selected at a *p*-value threshold of 0.0003, and included 1925 SNPs from the base schizophrenia GWAS (figure 2*c*, blue line). At this threshold, the PGS accounted for 0.47% of the variance in synaesthesia status (Nagelkerke's $R^2 = 0.0047$, adjusted for ascertainment), an association that remained statistically significant after 10 000 permutations (unadjusted $p = 0.00015$, empirical $p = 0.0027$) (figure 2*a*).

As selecting a *p*-value threshold based on its ability to predict synaesthesia status generates an R^2 value that is necessarily overfitted, we also assessed the relationship between synaesthesia status and the schizophrenia PGS using weighted information from all available SNPs across the genome. A previous meta-analysis of schizophrenia GWAS by the PGC used this unconstrained approach to determine how much of the variance in schizophrenia diagnosis can be explained by cumulative genetic risk for schizophrenia [43]. Other studies testing associations of schizophrenia PGSs with different traits have tested a range of *p*-value thresholds and reported the one that best predicted the target trait (as above), leading to adoption of a wide variety of *p*-value thresholds across the literature [45]. We found that the unthresholded schizophrenia PGS could account for 0.27% of the variance in synaesthesia status in our cohort (Nagelkerke's $R^2 = 0.0027$, adjusted for ascertainment), a statistically significant association (unadjusted $p = 0.0039$, empirical $p = 0.0041$).

To better understand how a schizophrenia PGS is associated with the odds of experiencing grapheme–colour synaesthesia, we split the combined set of cases and controls into 20 quantiles based on the best-fitting PGS. We observed a trend for increasing odds of having synaesthesia with increasing schizophrenia (SCZ) PGS load, although the 95% confidence intervals around the odds-ratios were wide and included 1.0 for both the lowest and highest quantiles of schizophrenia PGS (figure 2*b*).

In an exploratory follow-up analysis, we investigated potential relationships between additional aspects of synaesthetic experiences and schizophrenia PGS. A full survey of synaesthetic forms was available for 292 individuals. The number of forms of synaesthesia each person reported was scored as in Novich *et al.* [37], where related forms are combined and the maximum possible score is five [37]. The total number of forms of synaesthesia that a person experienced was not significantly correlated with their schizophrenia PGS (Spearman's $Rho = 0.012$, $p = 0.84$, average number of types = 2.55, s.d. = 1.07). The most common forms (following the Novich descriptions) were, in order, coloured sequence synaesthesia ($n = 292$), sequence–space synaesthesia ($n = 145$), coloured sensation ($n = 138$), coloured music ($n = 123$) and non-visual sequalae synaesthesia ($n = 47$). A total of 217 participants also completed the Projector-Associator survey as part of the *Synesthesia Battery* [46]. The survey asks whether synaesthetic experiences are perceived 'in the mind's eye' (associators) or externally (projectors) [47,48]. This facet of synaesthetic experience was also unrelated to the schizophrenia PGS (two-sample *t*-test, $t = 0.55$, d.f. = 58, $p = 0.59$).

We used summary statistics from the PGC-iPSYCH autism GWAS (18 381 ASD cases, European ancestry) to

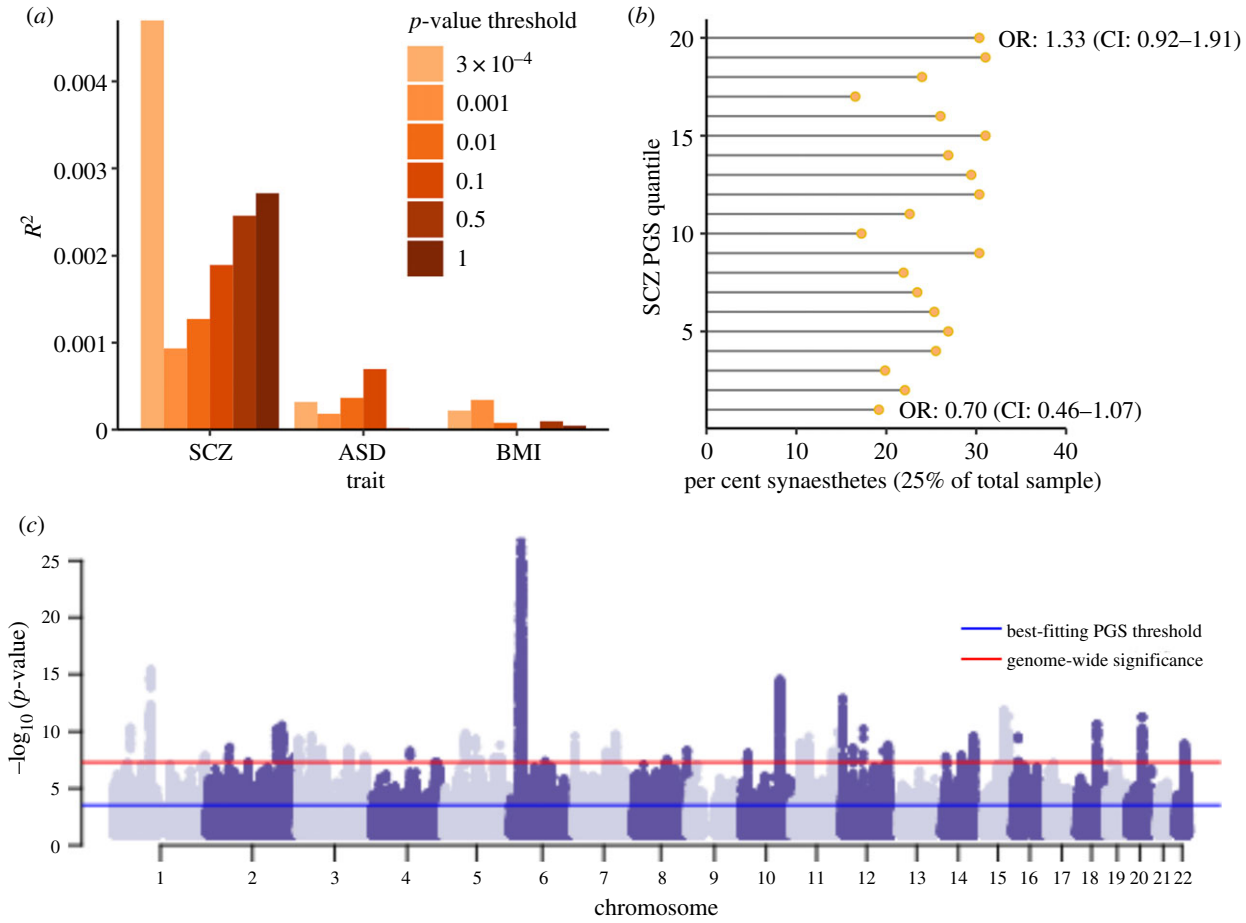

Figure 2. Relationship between schizophrenia polygenic scores and grapheme–colour synaesthesia. (*a*) Comparison of PGS model fit across several *p*-value thresholds for SCZ, ASD and BMI. (*b*) Schizophrenia PGS for the entire sample broken into 20 quantiles, with dots showing the proportion of synaesthetes within each quantile. Odds ratios for synaesthesia for individuals in the highest (20) and lowest (1) quantiles of schizophrenia risk are annotated. (*c*) Manhattan plot of the 2018 PGC schizophrenia GWAS results [43] with points indicating the negative log of the association *p*-value for each SNP. The best-fitting PGS *p*-value threshold (blue dotted line) and the standard threshold for a significant association between schizophrenia and a single SNP (red dashed line) are annotated. (Online version in colour.)

construct PGSs for both synaesthetes and controls [24]. Using the same covariates and synaesthesia prevalence as above, we calculated autism PGSs across the same range of *p*-value thresholds as before. We found that the best-fitting model used a *p*-value threshold of 0.0001, including 285 SNPs from the autism GWAS. In this case, even this best-fitting model was unable to predict synaesthesia status based on autism PGS (Nagelkerke's $R^2 = 0.00092$, $p = 0.091$, empirical $p = 0.54$), indicating that polygenic risk for ASD, at least from the currently available GWAS data, cannot distinguish synaesthetes from non-synaesthetes (figure 2*a*).

In order to determine if the subtle effects we saw for schizophrenia PGSs were owing to non-specific genetic differences between the synaesthetes and non-synaesthetes (e.g. owing to remaining differences in ancestry), we calculated PGSs for BMI using summary statistics available from the UK Biobank. BMI is an ideal negative control as we do not *a priori* expect genetic risk for high BMI to have any relationship with synaesthesia status, and the GWAS for this unrelated trait involved a very large sample of European ancestry ($n = 336\,107$). The PGS for BMI performed the worst of the three traits examined. The best-fitting model, based on a *p*-value threshold of 0.0027, including 30 454 SNPs from the BMI GWAS, failed to predict synaesthesia ($R^2 = 0.00058$, $p = 0.18$, empirical $p = 0.60$) (figure 2*a*). These results suggest that the cases and controls are well-matched, as uncorrected population stratification between the samples would be expected to introduce differences in non-psychiatric traits that also have a genetic basis.

4. Discussion

In this study, we present a new cohort of over 700 unrelated individuals with verified grapheme–colour synaesthesia and genotype data, and provide the first analyses of genetic relationships between synaesthesia and other brain-related traits, based on common polymorphisms.

We found that there was a significant relationship between aggregate genetic risk for schizophrenia and synaesthesia. However, the amount of variance explained was extremely small and synaesthetes were indistinguishable from non-synaesthetes at either the highest or lowest quantiles of schizophrenia PGS. We did not see significant effects with PGSs for risk of autism, another brain-related trait that has been hypothesized to have connections to synaesthesia based on prior work. Nor did we see any genetic relationship with higher BMI, used here as a negative control. In investigation of a subset of the sample with available data, we did not find a relationship between schizophrenia PGSs and how many types of synaesthesia participants reported. Nor did we see an association with associator–projector status, i.e. whether participants experience the secondary percept in the mind's eye or externally.

Synaesthesia is a perceptual phenomenon that can occur in otherwise neurotypical individuals, although being neurotypical is not a diagnostic criterion. It has been argued that experiencing synaesthesia should be considered an all-or-nothing designation rather than the tail end of a distribution that extends into cross-modal correspondences in the general population [49]. Other studies describing the psychological and/or neuropsychiatric profiles of synaesthetes using tools designed to measure levels of disorder-related traits in healthy populations (e.g. the Autism-Spectrum Quotient or the Oxford-Liverpool Inventory of Feelings and Experiences, O-LIFE) have shown that synaesthetes are only subtly different from non-synaesthetes on these other dimensions [19,20,44,50]. Informed by these results, we framed synaesthesia in this study as a binary trait that may have modest quantitative relationships with aspects of other neurological traits at the genetic level.

We examined the potential for such a relationship between synaesthesia and schizophrenia based on reports that synaesthetes experience higher levels of positive schizotypy compared to controls [19,44]. Specifically, in two separate studies, synaesthetes scored higher on the Unusual Experiences subscale of the O-LIFE, which measures positive schizotypy through questions on perceptual aberrations, hallucinations and magical thinking. In prior work connecting schizotypy in the general population to schizophrenia, levels of positive and negative schizotypy were positively correlated with schizophrenia PGS in healthy controls as well as healthy relatives of individuals with schizophrenia [23]. Our analyses of schizophrenia PGSs in the present study demonstrate an extremely subtle, but measurable, relationship between genetic risk for schizophrenia and grapheme–colour synaesthesia (figure 2*a*,*b*). This result should be considered from two angles: its relevance as a predictor of grapheme–colour synaesthesia, and what it reveals about the genetic basis of a perception-related trait-like synaesthesia. Although not robust enough to distinguish synaesthetes from non-synaesthetes even at the extreme ends of the distribution (figure 2*b*), we speculate that research into the genetic overlap between these traits may reveal firmer connections as the sample of synaesthetes increases.

Just as for other neuropsychiatric traits, the role of common genetic variation in schizophrenia is highly complex, with GWAS efforts identifying a large number of genome-wide significant risk alleles and many more loci that still contribute to its heritability [43,51] (figure 2*c*). Schizophrenia shows positive genetic correlations with bipolar disorder and major depressive disorder (and weakly with ASD) such that increased risk for one condition increases the risk for the others owing to shared genetic factors. Using the PGS approach instead of genetic correlation methods, schizophrenia PGSs can explain around 6% of the variation in bipolar disorder and major depressive disorder [45]. We wish to emphasize that not all of the shared biology points toward disorder—schizophrenia risk is also genetically correlated with increased educational attainment, a trait strongly linked to higher cognitive performance [52–54]. A recent systematic review of the schizophrenia PGS literature notes that these scores have much lower association with cognitive traits, explaining a maximum of 0.7% of the variance [45]. Previous work also showed a genetic link between schizophrenia and creativity, with schizophrenia PGSs explaining 0.24% of the variance in whether someone was a member of a professional society related to a creative pursuit (e.g. visual arts, writing and acting) [25]. Interestingly, alongside the differences in disorder-related traits, synaesthetes show enhanced performance on a variety of learning and memory tasks, and are more likely to be involved in creative pursuits [55–57]. It is important to emphasize that common genetic variation (i.e. polymorphisms that are observed at appreciable frequency in the population) only captures a portion of what makes these traits heritable, and that this limits the scope of genetic correlations between traits that are derived from SNP data. A long-term goal of such research is to understand how these traits form an interconnected web of shared biology, and the current study offers just a first impression of how synaesthesia may fit into this framework.

Proposed relationships between synaesthesia and ASD are nuanced, but the clearest suggested link between them is likely altered sensory sensitivity [20,58]. While the prevalence of synaesthesia among people with ASD varies by study, it is consistently elevated and likely highest for individuals with savant abilities [21,22,59]. Unfortunately, there are no published investigations of genomic associations with sensory sensitivity, and sample sizes for GWAS studies of ASD lag far behind those of schizophrenia and other neuropsychiatric traits, with only a handful of genome-wide significant loci identified so far [24]. By contrast, rare mutations in hundreds of different genes have been linked to ASD, and such rare genetic variation is thought to be a major contributor to the condition [60]. It is possible that common genetic variation plays a smaller role in ASD than schizophrenia, and that even with a more powerful GWAS the number of significant loci would remain low [61]. This would limit the utility of PGS-based methods for linking ASD to traits like synaesthesia and elevate study designs that instead compare the downstream biological consequences of rare genetic variation across traits. In a previous study, we identified rare genetic variants in three families with sound–colour synaesthesia, but found little overlap with known genes that have been implicated in ASD risk [9]. Pathway-based approaches focusing on genes rather than specific variants within genes may prove more effective at deciphering the biology underlying the shared sensory sensitivity and increased prevalence of synaesthetic experiences among people with ASD [20–22].

The limitations facing the current study stem from sample size, both for the synaesthesia cohort and the available GWAS data used in the PGS calculations. As the largest ASD GWAS includes roughly half as many cases as the most recent schizophrenia meta-analysis, we expected a less information-rich dataset with which to construct the PGSs. We calculated PGSs for a range of p-value thresholds to identify the best-fitting model; however, the relative lack of strong effect sizes in the ASD GWAS led to a model based on only 285 SNPs compared to 1925 in the schizophrenia PGS. Applicability across populations is a general limitation of PGS—current best practices warn against applying a PGS built using genetic data from Europeans to non-European populations. We worked to reduce the impact of population stratification on our results by matching the synaesthetes and controls on self-reported ancestry and including 20 principal components as covariates in the PGS calculations (tables 1 and 2 and figure 1). As the BMI GWAS in UK Biobank is 10-times larger than the schizophrenia sample, the lack of difference in BMI PGSs between synaesthesia cases and controls in the present study suggests that our groups were generally well-matched. Finally, we were limited here to a PGS approach because we are underpowered to more directly detect genetic correlations between synaesthesia and other

traits, as we are currently unable to accurately measure how much of the heritability of synaesthesia is accounted for by common genetic variation. In studying SNP-based genetic correlations between two traits the heritability accounted for by common variation (called 'SNP-heritability', to distinguish it from heritability measures based on twin-studies) must be known for both traits. We expect that future studies, with increased sample sizes for both synaesthesia and ASD cohorts, will overcome these hurdles and offer a better understanding of how the traits are connected.

In sum, this study introduces a new cohort of consistency test-verified synaesthetes, with accompanying genotype information across the genome, and offers the first genetic threads linking this unusual phenomenon to better understood brain-related traits. Large-scale GWAS efforts, like those available for educational attainment and schizophrenia, have already uncovered a network of interrelated traits with varying degrees of genetic overlap. While easy-to-survey measures like years of education or neuropsychiatric diagnoses are assessed for ever larger and more powerful genotyped cohorts, traits related to perception lag behind. There is still much to learn about how synaesthesia, as a model of natural variation in sensory perception, is connected to this larger network. Such efforts may reveal how genetic variation ripples through interconnected neurological traits, not only boosting or suppressing risk for disorder, but fine-tuning the neural substrates on which perception is built.

Ethics. Ethical approval for the study was granted by the Ethische Commissie Gedragswetenschappelijk Onderzoek (Ethics Committee, Faculty of Social Sciences) at Radboud University (application number ECG2013-2504-105). All participants provided written informed consent, and ethical approval for the Generation Scotland project was granted by the Tayside Committee on Medical Research Ethics, on behalf of the National Health Service (reference number: 05/S1401/89). The BIG study was approved by the regional ethics committee, and all participants provided written informed consent.

Data accessibility. GWAS summary statistics used to construct the PGS are available from https://www.med.unc.edu/pgc/results-and-downloads (ASD and schizophrenia) and https://docs.google.com/spreadsheets/d/1kvPoupSzsSFBNSztMzl04xMoSC3Kcx3CrjVf4yBmESU/edit?ts=5b5f17db#gid=227859291 (BMI). Anonymized data (PGS scores, synaesthesia status and covariates) are available from The Language Archive, a public data archive hosted by the Max Planck Institute for Psycholinguistics. The data are accessible at https://hdl.handle.net/1839/00-0000-0000-001A-8756-4. Full access to the Generation Scotland dataset is available on application to access@generationscotland.org.

Authors' contributions. A.K.T. conceived of the study and participated in its design, recruited participants, collected DNA samples, participated in DNA isolation, analysed the genetic and phenotypic data, and drafted the manuscript. A.V.S.-D. coordinated and participated in sample preparation and DNA isolation. K.S.K. participated in the design of the study, recruited participants, collected DNA samples and participated in DNA isolation. D.A.C. and L.v.d.H. recruited participants and collected DNA samples. A.V.S.-D., T.M.v.L. and R.R. recruited participants. A.C. and D.J.P. organized the participation of the Generation Scotland cohort. B.S. participated in the design of the study and the statistical genetics analyses. J.S. participated in the design of the study and recruited participants. S.E.F. conceived of the study, oversaw the work and drafted the manuscript. All authors critically revised the manuscript, gave approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. The study was supported by the Max Planck Society. A.K.T. was supported by European Commission Grant H2020-MSCA-IF-2015 grant no. 704393 (SynGenes). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council (grant no. HR03006). Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award 'STratifying Resilience and Depression Longitudinally' (STRADL) Reference 104036/Z/14/Z). This work made use of the BIG database, first established in Nijmegen, The Netherlands, in 2007. This resource is now part of Cognomics (www.cognomics.nl), a joint initiative by researchers of the Donders Centre for Cognitive Neuroimaging, the Human Genetics and Cognitive Neuroscience departments of the Radboud University Medical Center and the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. The Board of the Cognomics Initiative consists of Barbara Franke, Simon Fisher, Peter Hagoort, Han Brunner, Jan Buitelaar, Hans van Bokhoven and David Norris. The Cognomics Initiative has received support from the participating departments and centres and from external grants, i.e. the Biobanking and Biomolecular Resources Research Infrastructure (Netherlands) (BBMRI-NL), the Hersenstichting Nederland, the Netherlands Organisation for Scientific Research (NWO) and the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreements no. 602450 (IMAGEMEND). In addition, the work was supported by a grant for the ENIGMA Consortium (grant no. U54 EB020403) from the BD2 K Initiative of a cross-NIH partnership. We wish to thank all persons who kindly participated in the BIG research.

Acknowledgements. We thank all of the participants for their time and support of synaesthesia genetics research. We also thank Peter Withers and Kees van der Veer for excellent technical support, Generation Scotland for granting permission to use their resources and recontact participants, and the PGC for making the summary statistics from their GWAS publicly available.

References

1. Simner J, Mulvenna C, Sagiv N, Tsakanikos E, Witherby S, Fraser C, Scott K, Ward J. 2006 Synaesthesia: the prevalence of atypical cross-modal experiences. *Perception* **35**, 1024–1033. (doi:10.1068/p5469)
2. Simner J, Carmichael DA. 2015 Is synaesthesia a dominantly female trait? *Cogn. Neurosci.* **6**, 68–76. (doi:10.1080/17588928.2015.1019441)
3. Price MC, Mentzoni RA. 2008 Where is January? The month-SNARC effect in sequence-form synaesthetes. *Cortex* **44**, 890–907. (doi:10.1016/j.cortex.2006.10.003)
4. Baron-Cohen S, Burt L, Smith-Laittan F, Harrison J, Bolton P. 1996 Synaesthesia: prevalence and familiality. *Perception* **25**, 1073–1079. (doi:10.1068/p251073)
5. Bosley HG, Eagleman DM. 2015 Synesthesia in twins: incomplete concordance in monozygotes suggests extragenic factors. *Behav. Brain Res.* **286**, 93–96. (doi:10.1016/j.bbr.2015.02.024)
6. Gregersen PK, Kowalsky E, Lee A, Baron-Cohen S, Fisher SE, Asher JE, Ballard D, Freudenberg J, Li W. 2013 Absolute pitch exhibits phenotypic and genetic overlap with synesthesia. *Hum. Mol. Genet.* **22**, 2097–2104. (doi:10.1093/hmg/ddt059)
7. Asher JE, Lamb JA, Brocklebank D, Cazier J-B, Maestrini E, Addis L, Sen M, Baron-Cohen S, Monaco AP. 2009 A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. *Am. J. Hum. Genet.* **84**, 279–285. (doi:10.1016/j.ajhg.2009.01.012)
8. Tomson SN, Avidan N, Lee K, Sarma AK, Tushe R, Milewicz DM, Bray M, Leal SM, Eagleman DM. 2011 The genetics of colored sequence synesthesia: suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behav. Brain Res.* **223**, 48–52. (doi:10.1016/j.bbr.2011.03.071)
9. Tilot AK, Kucera KS, Vino A, Asher JE, Baron-Cohen S, Fisher SE. 2018 Rare variants in axonogenesis

genes connect three families with sound–color synesthesia. *Proc. Natl Acad. Sci. USA* **115**, 3168–3173. (doi:10.1073/pnas.1715492115)

10. Rouw R, Scholte SH. 2007 Increased structural connectivity in grapheme–color synesthesia. *Nat. Neurosci.* **10**, 792–797. (doi:10.1038/nn1906)
11. Newell FN, Mitchell KJ. 2016 Multisensory integration and cross-modal learning in synaesthesia: a unifying model. *Neuropsychologia* **88**, 140–150. (doi:10.1016/j.neuropsychologia.2015.07.026)
12. Kraft P, Zeggini E, Ioannidis JP. 2009 Replication in genome-wide association studies. *Stat. Sci* **24**, 561–573. (doi:10.1214/09-sts290)
13. Euesden J, Lewis CM, O'Reilly PF. 2015 PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468. (doi:10.1093/bioinformatics/btu848)
14. Choi S, Mak T, O'Reilly P. 2018 A guide to performing Polygenic Risk Score analyses. *bioRxiv*, 416545. (doi:10.1101/416545)
15. Curtis D. 2018 Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatr. Genet.* **28**, 85–89. (doi:10.1097/YPG.0000000000000206)
16. Schork AJ, Brown TT, Hagler DJ, Thompson WK, Chen C-H, Dale AM, Jernigan TL, Akshoomoff N. 2017 Polygenic risk for psychiatric disorders correlates with executive function in typical development. *Genes Brain. Behav.* **18**, e12480. (doi:10.1111/gbb.12480)
17. Foo JC *et al.* 2019 Evidence for increased genetic risk load for major depression in patients assigned to electroconvulsive therapy. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 35–45. (doi:10.1002/ajmg.b.32700)
18. Rothen N, Meier B, Ward J. 2012 Enhanced memory ability: insights from synaesthesia. *Neurosci. Biobehav. Rev.* **36**, 1952–1963. (doi:10.1016/j.neubiorev.2012.05.004)
19. Banissy MJ, Cassell JE, Fitzpatrick S, Ward J, Walsh VX, Muggleton NG. 2012 Increased positive and disorganised schizotypy in synaesthetes who experience colour from letters and tones. *Cortex* **48**, 1085–1087. (doi:10.1016/j.cortex.2011.06.009)
20. Ward J, Hoadley C, Hughes JE, Smith P, Allison C, Baron-Cohen S, Simner J. 2017 Atypical sensory sensitivity as a shared feature between synaesthesia and autism. *Sci. Rep.* **7**, 41155. (doi:10.1038/srep41155)
21. Hughes J, Simner J, Baron-Cohen S, Treffert DA, Ward J. 2017 Is synaesthesia more prevalent in autism spectrum conditions? Only where there is prodigious talent. *Multisensory Res.* **30**, 391–408. (doi:10.1163/22134808-00002558)
22. Baron-Cohen S, Johnson D, Asher J, Wheelwright S, Fisher SE, Gregersen PK, Allison C. 2013 Is synaesthesia more common in autism? *Mol. Autism* **4**, 1–6. (doi:10.1186/2040-2392-4-40)
23. van Os J, van der Steen Y, Islam MA, Gülöksüz S, Rutten B, Simons C, Investigators G. 2017 Evidence that polygenic risk for psychotic disorder is expressed in the domain of neurodevelopment, emotion regulation and attribution of salience. *Psychol. Med.* **47**, 1–17. (doi:10.1017/S0033291717000915)
24. Grove J *et al.* 2019 Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444. (doi:10.1038/s41588-019-0344-8)
25. Power RA *et al.* 2015 Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat. Neurosci.* **18**, 953–955. (doi:10.1038/nn.4040)
26. Cuskley C, Dingemanse M, Kirby S, van Leeuwen TM. 2019 Cross-modal associations and synesthesia: categorical perception and structure in vowel–color mappings in a large online sample. *Behav. Res. Methods* **51**, 1651–1675. (doi:10.3758/s13428-019-01203-7)
27. Smith BH *et al.* 2013 Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700. (doi:10.1093/ije/dys084)
28. Smith BH *et al.* 2006 Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* **7**, 74. (doi:10.1186/1471-2350-7-74)
29. Kerr SM *et al.* 2013 Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med. Genet.* **14**, 38. (doi:10.1186/1471-2350-14-38)
30. Guadalupe T *et al.* 2014 Measurement and genetics of human subcortical and hippocampal asymmetries in large datasets. *Hum. Brain Mapp.* **35**, 3277–3289. (doi:10.1002/hbm.22401)
31. Hibar DP *et al.* 2017 Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624. (doi:10.1038/ncomms13624)
32. Gunz P *et al.* 2019 Neandertal introgression sheds light on modern human endocranial globularity. *Curr. Biol.* **29**, 120–127. (doi:10.1016/j.cub.2018.10.065)
33. Eagleman DM, Kagan AD, Nelson SS, Sagaram D, Sarma AK. 2007 A standardized test battery for the study of synesthesia. *J. Neurosci. Methods* **159**, 139–145. (doi:10.1016/j.jneumeth.2006.07.012)
34. Carmichael DA, Down MP, Shillcock RC, Eagleman DM, Simner J. 2015 Validating a standardised test battery for synesthesia: does the synesthesia battery reliably detect synesthesia? *Conscious. Cogn.* **33**, 375–385. (doi:10.1016/j.concog.2015.02.001)
35. Simner J, Hubbard E, Johnson D, Allison C, Baron-Cohen S. 2013 The prevalence of synesthesia: the Consistency Revolution. In *Oxford Handbook of Synesthesia* (eds J Simner, E Hubbard), pp. 3–23. Oxford, UK: Oxford University Press. (doi:10.1093/oxfordhb/9780199603329.001.0001)
36. Rothen N, Seth AK, Witzel C, Ward J. 2013 Diagnosing synaesthesia with online colour pickers: maximising sensitivity and specificity. *J. Neurosci. Methods* **215**, 156–160. (doi:10.1016/j.jneumeth.2013.02.009)
37. Novich S, Cheng S, Eagleman DM. 2011 Is synaesthesia one condition or many? A large-scale analysis reveals subgroups. *J. Neuropsychol.* **5**, 353–371. (doi:10.1111/j.1748-6653.2011.02015.x)
38. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, Samuels DC, Shyr Y, Long J. 2014 Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **9**, 2643–2662. (doi:10.1038/nprot.2014.174)
39. Coleman JR, Euesden J, Patel H, Folarin AA, Newhouse S, Breen G. 2015 Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings Funct. Genomics* **15**, 298–304. (doi:10.1093/bfgp/elv037)
40. Haplotype Consortium *et al.* 2016 A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283. (doi:10.1038/ng.3643)
41. O'Connell J *et al.* 2014 A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* **10**, e1004234. (doi:10.1371/journal.pgen.1004234)
42. Das S *et al.* 2016 Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287. (doi:10.1038/ng.3656)
43. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2018 Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715. (doi:10.1016/j.cell.2018.05.046)
44. McErlean AB, Banissy MJ. 2016 Examining the relationship between schizotypy and self-reported visual imagery vividness in grapheme-color synaesthesia. *Front. Psychol.* **7**, 131. (doi:10.3389/fpsyg.2016.00131)
45. Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. 2018 The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: systematic review. *Schizophr. Res.* **197**, 2–8. (doi:10.1016/j.schres.2017.10.037)
46. Carmichael DA, Down MP, Shillcock RC, Eagleman DM, Simner J. 2015 Validating a standardised test battery for synesthesia: Does the Synesthesia Battery reliably detect synesthesia? *Conscious. Cogn.* **33**, 375–385. (doi:10.1016/j.concog.2015.02.001)
47. Cohen MX, Weidacker K, Tankink J, Scholte H, Rouw R. 2015 Grapheme-color synesthesia subtypes: stable individual differences reflected in posterior alpha-band oscillations. *Cogn. Neurosci.* **6**, 56–67. (doi:10.1080/17588928.2015.1017450)
48. Anderson HP, Ward J. 2015 Principle component analyses of questionnaires measuring individual differences in synaesthetic phenomenology. *Conscious. Cogn.* **33**, 316–324. (doi:10.1016/j.concog.2015.01.013)
49. Deroy O, Spence C. 2013 Why we are not all synesthetes (not even weakly so). *Psychon. Bull. Rev.* **20**, 643–664. (doi:10.3758/s13423-013-0387-2)
50. Simmonds-Moore CA, Alvarado CS, Zingrone NL. 2018 A survey exploring synesthetic experiences: exceptional experiences, schizotypy, and psychological well-being. *Psychol. Conscious. Theory Res. Pract.* **6**, 99–121. (doi:10.1037/cns0000165)

51. Ripke S *et al.* 2014 Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427. (doi:10.1038/nature13595)
52. Li Z *et al.* 2017 Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583. (doi:10.1038/ng.3973)
53. Okbay A *et al.* 2016 Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542. (doi:10.1038/nature17671)
54. Lee JJ *et al.* 2018 Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121. (doi:10.1038/s41588-018-0147-3)
55. Bankieris KR, Aslin RN. 2016 Explicit associative learning and memory in synesthetes and nonsynesthetes. *i-Perception* **7**, 2041669516658488. (doi:10.1177/2041669516658488)
56. Lunke K, Meier B. 2018 New insights into mechanisms of enhanced synaesthetic memory: benefits are synaesthesia-type-specific. *PLoS ONE* **13**, e0203055. (doi:10.1371/journal.pone.0203055)
57. Lunke K, Meier B. 2018 Creativity and involvement in art in different types of synaesthesia. *Brit. J. Psychol.* (doi:10.1111/bjop.12363)
58. Ward J, Brown P, Sherwood J, Simner J. 2018 An autistic-like profile of attention and perception in synaesthesia. *Cortex* **107**, 121–130. (doi:10.1016/j.cortex.2017.10.008)
59. Neufeld J, Roy M, Zapf A, Sinke C, Emrich HM, Prox-Vagedes V, Dillo W, Zedler M. 2013 Is synesthesia more common in patients with Asperger syndrome? *Front. Hum. Neurosci.* **7**, 847. (doi:10.3389/fnhum.2013.00847)
60. Wiśniowiecka-Kowalnik B, Nowakowska B. 2019 Genetics and epigenetics of autism spectrum disorder—current evidence in the field. *J. Appl. Genet.* **60**, 37–47. (doi:10.1007/s13353-018-00480-w)
61. Sullivan PF, Geschwind DH. 2019 Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* **177**, 162–183. (doi:10.1016/j.cell.2019.01.015)