Rare variants in axonogenesis genes connect three families with sound–color synesthesia

Amanda K. Tilot*, Katerina S. Kucera*, Arianna Vino*, Julian E. Asherb, Simon Baron-Cohenb,1, and Simon E. Fishera,c,1,2

*Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; bAutism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge CB2 8AH, United Kingdom; and cDonders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 HE Nijmegen, The Netherlands

Edited by Edward M. Hubbard, University of Wisconsin–Madison, Madison, WI, and accepted by Editorial Board Member Randolph Blake January 23, 2018 (received for review September 5, 2017)

Synesthesia is a rare nonpathological phenomenon where stimulation of one sense automatically provokes a secondary perception in another. Hypothesized to result from differences in cortical wiring during development, synesthetes show atypical structural and functional neural connectivity, but the underlying molecular mechanisms are unknown. The trait also appears to be more common among people with autism spectrum disorder and savant abilities. Previous linkage studies searching for shared loci of large effect size across multiple families have had limited success. To address the critical lack of candidate genes, we applied whole-exome sequencing to three families with sound–color (auditory–visual) synesthesia affecting multiple relatives across three or more generations. We identified rare genetic variants that fully cosegregate with synesthesia in each family, uncovering 37 genes of interest. Consistent with reports indicating genetic heterogeneity, no variants were shared across families. Gene ontology analyses highlighted six genes—COL4A1, ITGA2, MYO10, ROBO3, SLC9A6, and SLIT2—associated with axonogenesis and expressed during early childhood when synesthetic associations are formed. These results are consistent with neuroimaging-based hypotheses about the role of hyperconnectivity in the etiology of synesthesia and offer a potential entry point into the neurobiology that organizes our sensory experiences.

Significance

Our physical senses are separated not only into distinct experiences but also into specialized regions within the cerebral cortex. Synesthesia is a neurological phenomenon that causes unusual links between sensory experiences, and its molecular basis is completely unknown. We demonstrate that three families who experience color when listening to sounds are connected by rare genetic variants affecting genes that contribute to axonogenesis, a process essential for neuronal connections within and across brain regions. Multiple genes with similar activity patterns during neural development fall within parts of the genome previously linked to the condition. Our results connect synesthetes’ altered structural and functional connectivity to genes that support the development of those connections.

Author contributions: A.K.T., K.S.K., S.B.-C., and S.E.F. designed research; A.K.T., K.S.K., A.V., and J.E.A. performed research; A.K.T., K.S.K., A.V., and S.E.F. analyzed data; A.K.T. and S.E.F. wrote the paper; and J.E.A. and S.B.-C. collected synesthesia families.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.M.H. is a guest editor invited by the Editorial Board. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The datasets generated during the current study are available upon request from The Language Archive (TLA: https://corpus1.mpi.nl/davis/10), a public data archive hosted by the Max Planck Institute for Psycholinguistics. The data are stored under the node IDs MP11756324 and MP11813562# and are accessible at https://hdl.handle.net/1839/00-0000-0000-001A-8756-484view.

5B.C. and S.E.F. contributed equally to this work.

6To whom correspondence should be addressed. Email: simon.fisher@mpi.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1715492115/-/DCSupplemental.
candidate genes from these regions failed to detect rare variants segregating with the phenotype (19). Our understanding of the developmental processes that define the boundaries between our senses is impeded by our lack of knowledge about the genetic underpinnings.

We performed a genome-wide sequence-level analysis of synesthesia to begin addressing the major neurobiological hypotheses in this field from a genetic perspective. We focused on three unrelated families, ascertained on the basis of at least five members affected with sound–color synesthesia across three or more generations, validated through an established test battery (22). Using whole-exome sequencing (WES), we identified rare coding variants that perfectly segregate with the phenomenon within each family. Variants preferentially fell within genes tied to the processes of axonogenesis and cell migration, forming a common theme across families, one that aligns with contemporary theories about the neurodevelopmental origins of synesthesia. These results suggest that molecular approaches can help increase understanding of the neurobiology of our sensory experiences, beyond pathology.

Results

We used WES to identify genetic variants within coding parts of the genome for three multigenerational families with sound–color synesthesia (Fig. 1). For each family, we obtained exome sequences from four or five verified synesthetes across three generations plus at least one nonsynesthetic family member. All cases in this study had sound–color synesthesia as confirmed using the established Test of Genuineness (SI Materials and Methods), although it is possible that some also experience additional, more common types of synesthesia (e.g., colored weekdays). The families were identified from the Cambridge Synaesthesia Research Group database and were originally included in a 2009 study using microsatellite markers to look for evidence of genetic linkage across 43 families of varying sizes with forms of auditory–visual synesthesia (20). The families in the present study were chosen as the most feasible for studying with WES based on size, structure, and availability of suitable DNA samples.

Using nonparametric linkage analysis and assuming genetic heterogeneity, the authors of the 2009 study found three regions with suggestive linkage (5q33, 6p12, and 12p12) (23). Chromosome 2q24 showed significant linkage after gene-drop simulations (heterogeneity log-odds of the odds score, HLOD 3.02, P value 0.047) (20). An independent investigation of five other families with sequence–color synesthesia (color associations for sequences such as numbers or months) reported suggestive evidence for a locus on chromosome 16q, but this was supported by only two of the five families (19). These studies suggest that synesthesia involves considerable genetic heterogeneity, with different genetic factors contributing in different families.

Following the Genome Analysis Toolkit’s best practices guidelines for calling DNA variants, we identified 11,597 variants across our three sound–color families after removing low-quality variants and those with low sequencing depth (family 2: 8,195; family 11: 9,202; and family 16: 8,074) (24). To elevate potentially causative variants, we applied filtration criteria based on our limited knowledge of synesthesia’s genetic architecture and the prevalence of the sound–color variety (familial or sporadic). A 2006 study established that up to 4.4% of the UK population may experience at least one form of synesthesia, but the prevalence of sound–color synesthesia is not well studied (7, 23, 25). Estimates range from 1 in 500 unselected individuals in the United Kingdom (from the same prevalence study) to 41% of self-referred Dutch and German synesthetes (7, 25). Given the uncertainty in these estimates, we chose a relatively inclusive maximum minor allele frequency (MAF) of 0.01 for highlighting variants of potential interest. In total, there were 3,864 variants across the families that were rare enough to be considered further (family 2: 1,812; family 11: 2,727; family 16: 1,862; note that, prior to the further filtering described below, these included some partially overlapping variants across families).

Based on pedigree structures of the three families, we next retained variants that followed dominant inheritance with full penetrance (Fig. L4). Given the findings of prior genetic screens, we hypothesized that while familial sound–color synesthesia may be due to single variants within a particular family, such variants were unlikely to be shared across different families (20). To test this, we looked in our retained variants for rare, heterozygous mutations that perfectly segregated with synesthesia within each respective family. This yielded a total of 37 variants across 37 genes (Table 1). No variants were shared across all synesthetes in all families, consistent with the hypothesis of genetic heterogeneity. A missense variant in RGS21 was detected in all synesthetes from family 16 but was found in only one synesthete from family 2. Further supporting genetic heterogeneity in synesthesia, no single gene contained a perfectly segregating variant in all three families. None of the 37 highlighted variants fell within the suggestive linkage peaks reported in prior studies (19, 20).

To further define shared mechanisms across the three families, we assessed whether there were enriched gene ontology terms within our set of putative candidate genes from WES. We excluded inferred electronic annotations, as these are considered less informative than manually curated entries. The ontology analysis identified significant enrichment for a narrow set of terms primarily related to neural development (Table 2). Notably, six genes were associated with axonogenesis (GO:0007409) and cell migration (GO:0016477), biological categories with clear relevance to the hyperconnectivity account of synesthesia. These genes, SLIT2, MYO10, ROBO3, ITGA2, COL4A1, and SLC9A6 (marked by an asterisk in Table 1), span the three families and may point to particular processes that can be investigated at higher levels (e.g., hyperconnectivity).

Having identified variants in a core set of six candidate genes with potential functional relevance for neurobiological accounts of synesthesia etiology, we went on to study their expression in neural tissues. We used two complementary
datasets, the Genotype-Tissue Expression project (GTEx) based on RNA-sequencing (RNAseq) and the microarray-based Allen Human Brain Atlas (ABA), to determine expression patterns of the candidates (26, 27). GTEx includes data from >100 postmortem human brains with RNA sampled from 13 sites, including the frontal cortex (26). Each of the six genes had detectable expression in these samples, with SLC9A6 and ITGA2 having the highest and lowest expression, respectively (Table S1).

As protein levels are not always well correlated with RNA expression (28), we combined these findings with immunohistochemistry data from the Human Protein Atlas, which includes manually quantified protein expression from human cerebral cortical tissue of three adult donor brains (age range, 37–70 y) (29). The Atlas included results for each of the relevant proteins, except ROBO3. All five remaining proteins were observed in neuronal cells, albeit to varying degrees (Table S1 shows ranges

### Table 1. Rare variants segregating with sound–color synesthesia within each family

| Family | Position | Ref/Alt | Gene | ExAC MAF | Impact | CADD | Eigen score | PolyPhen | SIFT |
|--------|----------|---------|------|----------|--------|------|-------------|----------|------|
| 11     | 1:1231208| C/T     | ACAP3| 0.002    | Missense| 18.65| −0.59       | Benign   | Deleterious |
| 2      | 1:146696486| C/G   | FMOS | 0.005    | Splice donor| 27.50| 1.18        | NA       | NA |
| 2      | 1:151665566| T/C   | SNN27| 0.005    | Regulatory| 1.22 | 0.42        | NA       | NA |
| 2      | 1:154574699| T/G   | ADAR | 0.005    | Missense | 24.80| 0.14        | Possibly damaging | Deleterious |
| 2      | 1:156899248| G/A   | LRRC71| 0.007   | Intron | 5.29 | 0.29        | NA       | NA |
| 2      | 1:162760613| C/G   | HSD17B7| 0.005  | Missense | 29.00| 0.80        | Probably damaging | Deleterious |
| 2      | 1:162825253| AGAGCA | C1orf110| 0.005 | Regulatory | 14.18| NA        | NA       | NA |

### Table 2. Gene ontology terms enriched in the combined set of synesthesia-associated variants

| Gene ontology term | Corrected P value | Terms Overlap | Intersecting genes |
|--------------------|-------------------|---------------|-------------------|
| BP: axonogenesis    | 0.0459            | 605           | 6 SLITZ, MYO10, ROBO3, ITGA2, COL4A1, SLC9A6 |
| BP: substrate-dependent cell migration | 0.0252 | 18 | 2 SLITZ, ITGA2 |
| BP: cell morphogenesis involved in differentiation | 0.0345 | 799 | 7 SLITZ, MYO10, ROBO3, ITGA2, FGA, COL4A1, SLC9A6 |
| MF: proteoglycan binding | 0.0022 | 17 | 2 CECR1, SLITZ |
| MF: hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines | 0.0452 | 24 | 2 CECR1, SLITZ |
| MF: deaminase activity | 0.0490 | 25 | 2 CECR1, SLITZ |

ExAC, Exome Aggregation Consortium; PolyPhen, polymorphism phenotyping; Ref/Alt, reference/alternative; SIFT, sorting intolerant from tolerant.

*Genes associated with axonogenesis (GO:0007409) and cell migration (GO:0016477), biological categories with relevance to the hyperconnectivity account of synesthesia.

Tilot et al.

PNAS Latest Articles | 3 of 6
when multiple antibodies produced different results). Beyond neurons, multiple candidates showed staining in neuropil and endothelial cells, while only SLIT2 was observed at high levels in glia. These results support the GTEx RNAseq data, indicating that the six genes are active in adult brain tissue.

The GTEx and Human Protein Atlas resources lack data from several cortical regions with relevance for synesthesia studies, and the tissues were primarily collected from middle-aged adults. Thus, we next sought to determine if the highlighted genes are active in the auditory, visual, and parietal cortices and to examine their expression patterns during development. Reports vary on the neuroanatomical regions and activity patterns that may mediate synesthetic experiences; some support a role for the visual cortex, and others emphasize sensory integration in the parietal lobe (11, 14, 30). We used microarray data from six postmortem brains, sampled at ~500 locations, with the data mapped to structural MRI scans (ABA), to visualize gene expression in a more fine-grained fashion (27). Despite the differing methodologies, expression values from the GTEx cortical samples were well correlated with frontal cortical data from the ABA (Pearson’s, \( r = 0.95, P = 0.004 \)). In the ABA, we found that each of the six genes was widely expressed across the brain (Fig. 2A), including the auditory, visual, and parietal cortices.

Longitudinal studies show that synesthetic associations often form during primary school years, and axonogenesis primarily occurs from early fetal development through the first years of life, prompting us to examine expression of our candidate genes across human brain development (15). To do this, we used data from the Allen Institute’s BrainSpan project, which includes human brain tissue from age 8 wk post conception to 40 y, sampling an average of 13 regions (range, 1–17) from one to three brains per time point, measured using RNAseq (31). Each of the six genes of interest has detectable expression in auditory, visual, and parietal cortices during fetal development and early childhood (Fig. 2B). Finally, to determine whether the six candidates were preferentially expressed in any specific neural cell types, we used an RNAseq dataset of six cell types isolated from the mouse cortex with a series of antibody-based purification steps to ensure high cell-type specificity (32). The candidate genes were broadly expressed across the neural and glial transcriptomes, with relatively lower expression levels in microglia (Fig. 2C).

Previous synesthesia genetics studies not only faced limitations of small family sizes but also were restricted by the use of lower-precision methods than those adopted here. We thus sought to determine whether the results of the current study might be relevant for those earlier reports. Coexpression patterns can reveal genes involved in similar neurodevelopmental processes, and so we tested whether genes coexpressed with the six candidates fell within the putative linkage peaks of prior studies, while acknowledging that most evidence from previous reports was at the suggestive level (33). The peaks reported by Asher et al. (20) were unlikely to be a result of linkage within the three families used in the present study, as we found no segregating variants within those regions (Table 1). Thus, the peaks from 5q33, 6p12, 12p12, 2q24 (sound-color synesthesia) and 16q12.2–23.1 (sequence-color synesthesia) represent signals from other families. Using the six putative candidates as seed genes, we found 109 genes coexpressed at \( r > 0.7 \) during neural development (31). The complete set of coexpressed genes contained six located within a previously reported synesthesia linkage peak. However, a formal test of enrichment did not achieve statistical significance (Fisher’s exact test, odds ratio = 2.069, \( P = 0.08 \)). Mapping known protein–protein interactions from Reactome within this extended set yielded a network that was relatively sparse but included \( ITGA2 \) and \( COL4A1 \) as hubs within a subnetwork (Fig. S3).

![Fig. 2. Six synesthesia candidate genes are widely expressed across neural development. (A) Expression data from the ABA, mapped to a unified anatomic framework (27). (B) Gene expression in human parietal, primary auditory, and primary visual cortices from 12 wk post conception to age 40 y, from the BrainSpan atlas (one to three independent measurements per gene, region, and time point); the y axis uses a log2 scale to visualize change over time for each gene. (C) Neural gene expression in specific cell types isolated from adult mice. Boxes represent the 25th to 75th percentiles; whiskers extend to 1.5 times the interquartile range. Data used in C from the Barres laboratory are available at https://web.stanford.edu/group/barres_lab-brain_rnaseq.html. FPKM, fragments per kilobase of transcript per million mapped reads.](https://www.pnas.org/content/1715492115/FB)
Finally, recent findings suggest that people with autism spectrum conditions (henceforth, “autism”) and savant abilities are more likely to experience synesthesia, while synesthetes as a group show atypical sensory sensitivities similar to those seen in autism (34, 35). We thus hypothesized that genetic variants uncovered in synesthetic individuals without autism may show links to genetic pathways implicated in autism. The presence of comorbid neurological conditions (including autism) was an exclusion criterion when recruiting the families for this work (20). Three of the 37 variants that tracked with synesthesia status in our families are within genes found in the Simons Foundation Autism Database catalog of autism-associated genes: FGA (family 16), HYDIN (family 2), and SLC9A6 (family 11) (36). However, the specific genetic variants seen in the synesthesia families (Table 1) are not shared with the autism cases currently contained in Autism Database. Also, the three genes did not represent a significant enrichment (Fisher’s exact test, odds ratio = 2.01, P value = 0.20); thus the presence of these variants is not unusual, given the large number of genes associated with autism. Four of the 109 coexpressed genes were previously associated with autism (Fig. S3), although, again, this did not represent significant enrichment (Fisher’s exact test, odds ratio = 2.084, P = 0.13).

Discussion

Synesthesia represents the outer edges of natural variation in sensory perception. In this study, we identified rare genetic variants perfectly cosegregating with the trait in three unrelated multigenerational families with at least five people affected with sound–color synesthesia. A core set of six genes was related to axonogenesis, the process by which immature neurons send out their primary process to connect with other brain regions by following secreted guidance cues. Among other cortical sites, these genes were expressed throughout development in auditory and visual cortex, consistent with a potential role in developmental processes spanning both brain regions. Using a network approach, we uncovered six additional genes with similar neural expression patterns that fall within suggestive linkage regions originally identified in screens of larger numbers of families, which could be further investigated in follow-up studies (20).

Potentially supporting emerging data that synesthesia shares some phenotypic and mechanistic features with autism, the orthologs of genes implicated in Alzheimer’s disease coexpressed with genes in the developing and adult brain fits with the wider synesthesia phenotype, which includes enhanced memory performance as well as altered sensory sensitivity (34, 46).

Since studies of the genetic basis for synesthesia are in early stages, there were limitations to our approach and dataset. Only one prevalence study has asked about sound–color synesthesia, confirming its presence in 1 of 500 UK college students surveyed (7). This consent with a survey of Dutch and German synesthetes with a much higher percentage reported sound–color associations (25). We erred on the side of inclusivity when determining the MAF threshold to use here. Despite this, we note that most of the synesthesia-specific variants we report are rare, falling well below this cutoff. We further focused on variants that perfectly tracked with the phenotype in each family. The families in this study were the largest multiplex families for whom we could obtain sufficient DNA for WES and validation, but they are still relatively small for linkage-based approaches, and we were unable to narrow the segregating variants to a single site per family. Further family-based studies of sound–color synesthesia and other forms will help clarify the genetic architecture and the potential role of altered neuronal morphology.
Over 130 y after the first reports of familial synesthesia, these results provide a molecular starting point for studies addressing the origins of healthy variation in sensory integration. While cellular and animal models will likely prove illustrative, behavioral experiments to test cross-modal perception in animals may elude the field for some time. It remains to be seen if other forms of synesthesia will involve alterations in axon growth and guidance. This study focused on relatively rare genetic variation occurring in families with multiple generations of synesthetes; assessing potential roles of common variation represents an intriguing question for the future.

Materials and Methods

Three families with sound–color (auditory–visual) synesthesia, without a history of drug use or any neurological, ophthalmological, or psychiatric disorder, were identified from the Cambridge Synaesthesia Research Group database. Ethical approval was granted by the Human Biology Research Ethics Committee (Ref: 2011.06). Informed consent was obtained from all participants. Variants in exome-sequencing data were called using the Genome Analysis Toolkit best practices pipeline (24). Gene ontology analyses were performed with gprofilerR, and human gene expression data were downloaded from GTEx, the Allen Human Brain Atlas, and the BrainSpan database, while mouse RNAseq data were accessed from Zhang et al. (26, 31, 32). The datasets generated during the current study are available upon request from The Language Archive (TLA: https://corpus1.mpi.nl/ds/asv/70), a public data archive hosted by the Max Planck Institute for Psycholinguistics. An extended description of materials and methods appears in Supporting Information.

ACKNOWLEDGMENTS. We thank the families from the Cambridge Synaesthesia Research Group Database for their contributions to this study; Allen Chan and Laura Murphy for assistance with resampling the families; Han Brunner, Joris Veltman, and Alex Hoischen for assistance with exome sequencing; and Anthony Monaco for his contribution to earlier phases of the genetic analysis for these families. Funding was provided by European Commission Grant H2020-MSCA-IF-2015 #704933 (to A.K.T.) and the Hertie Foundation through the Eric Kandel Young Neuroscientists Prize (to S.E.F.). K.S.K., A.V., A.K.T., and S.E.F. were supported by the Max Planck Society. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.