

# The genetics of colored sequence synesthesia: Evidence of linkage to chromosome 16q and genetic heterogeneity for the condition

Stephanie S. Nelson<sup>1</sup>, Nili Avidan<sup>2</sup>, Anand K. Sarma<sup>1</sup>, Rejnal Tushe<sup>1</sup>, Dianna M. Milewicz<sup>2</sup>, Kwanghyuk Lee<sup>3</sup>, Molly Bray<sup>4</sup>, Suzanne M. Leal<sup>3</sup>, David M. Eagleman<sup>1,5\*</sup>

<sup>1</sup>Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA; <sup>2</sup>Division of Medical Genetics, Department of Internal Medicine, University of Texas Medical School at Houston, Houston, TX, USA; <sup>3</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; <sup>4</sup>Departments of Epidemiology and Genetics, University of Alabama at Birmingham, Birmingham, AL, <sup>5</sup>Department of Psychiatry, Baylor College of Medicine, Houston, TX, USA

\*Correspondence: David Eagleman, PhD, Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA. Email: eagleman@bcm.edu

## Abstract

Synesthesia is a perceptual condition in which normal sensory stimulation can trigger anomalous sensory experiences. For example, synesthetes may experience colors in response to sounds, tastes in response to words, or smells in response to touch. We here focus on colored sequence synesthesia, in which color experiences are triggered by learned ordinal sequences such as letters, numbers, weekdays and months. Although synesthesia has been noted in the scientific literature for over a century, it is understood only at the level of the phenomenology, and not at the molecular and neural levels. We have performed a linkage analysis to identify the first genetic loci responsible for the increased neural crosstalk underlying colored sequence synesthesia. Our analysis has identified a 23 MB region on chromosome 16 as a putative locus for the trait. Our data provide the first step in understanding neural crosstalk from its molecular basis to its behavioral consequences, opening a new inroad into the understanding of the multisensory brain.

**Keywords:** synesthesia; linkage analysis; chromosome 16

## Introduction

In synesthesia, ordinary stimuli elicit anomalous perceptual experiences. In one form, known as grapheme-color synesthesia, viewing a letter or digit triggers the visual experience of a specific color; in closely related forms, days of the week or months of

the year trigger color experiences. We will refer to these forms collectively as colored sequence synesthesia (CSS), since they involve the triggering of color experiences by learned sequences (Figure 1). The particular color pairings for each stimulus are

idiosyncratic for each synesthete. Grapheme-color is one of the most common varieties, estimated in random sample studies at >1% of the population (Simner et al 2006), and we show below that weekday- and month-synesthesias are highly correlated.

Synesthetic experiences differ from imagery in their consistency and automaticity. These associations are often involuntary and constant over time, much like the memory of a friend's face upon hearing her name. This is thought to occur by an increasing degree of crosstalk between brain areas, such that activity in one area kindles activity in another.

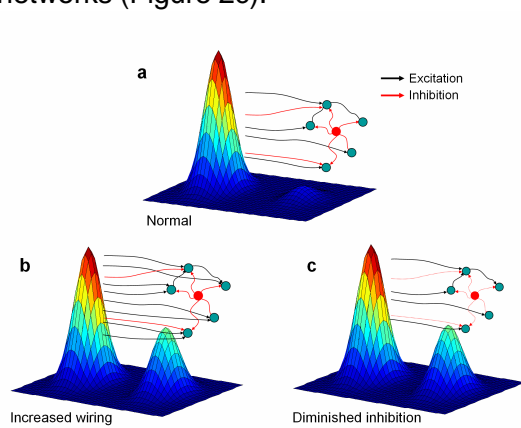
0	A	N	Sunday	January
1	B	O	Monday	February
2	C	P	Tuesday	March
3	D	Q	Wednesday	April
4	E	R	Thursday	May
5	F	S	Friday	June
6	G	T	Saturday	July
7	H	U		August
8	I	V		September
9	J	W		October
	K	X		November
	L	Y		December
	M	Z		

**Figure 1.** Data from a representative subject tested at synesthete.org, showing the color experiences triggered by numbers, letters, weekdays and months.

One hypothesis of synesthetic crosstalk stipulates that excess connections in newborn brains are insufficiently pruned, leading to increased cross-wiring in adulthood. A variant of this idea is that there is increased outgrowth of neurons in a synesthete's brain. What both

of these ideas (insufficient pruning and increased arborization) have in common is the idea that a synesthetic brain harbors an abnormal profile of synaptic connections (Figure 2b).

A second hypothesis implicates unusual inhibition as the cause of synesthesia. The idea is that excitation is counterbalanced by inhibition in normal brains, while in synesthetic brains the excitation can overcome the weakened inhibition. In this framework, the same rich connectivity is present in all brains; the only difference between normals and synesthetes is the functioning of the inhibitory networks (Figure 2c).



**Figure 2.** Two models of synesthesia. **a.** Under normal conditions, excitation and inhibition are balanced such that activity in one brain area does not lead to activity in another. Black lines represent excitatory input, red lines represent inhibitory input, and the red neuron represents local inhibition by interneurons. **b.** An increase of excitatory wiring causes activity in one area to kindle activity in the other. **c.** When inhibition is decreased, either by diminished input or decreased receptor

These two hypotheses are indistinguishable by behavioral measures, electroencephalography and neuroimaging –

all of which demonstrate functional crossactivation but cannot discriminate between hypotheses of the underlying molecular basis.

In 1883, Sir Francis Galton observed that synesthesia runs in families and suggested that the condition is heritable . Cytowic (1989) examined the inheritance patterns of eight self-reported families and suggested that synesthesia is transmitted as a dominant trait . Ward & Simner (2005), also relying on self-report, looked at inheritance patterns in 72 families of varying size and conjectured that the trait may be inherited on the X–chromosome based on their observation that the trait never seemed to pass from father to son . Asher et al (2009) reported two cases of male-to-male transmission, which is problematic for the X–linked hypothesis because a single authentic case invalidates it . However, X-linkage remains a possibility if there is incomplete penetrance. Incomplete penetrance seems likely, as evidenced by the finding of identical twins wherein one twin is synesthetic while the other is not .

One shortcoming of previous genetic studies such as Cytowic (1989), Ward & Simner (2005) is the failure to objectively verify the phenotype. Instead, participants were deemed synesthetic based solely on self-report. Furthermore, what counted as a ‘pedigree’ was often as small as a single parent and child. Therefore, at present, the mode of inheritance is unknown.

In an attempt to identify the first gene for colored-sequence synesthesia, we have performed a large-scale family linkage analysis in synesthetic family pedigrees. We have limited the present study to individuals

with CSS and we have verified the consistency of these color-sequence associations with a host of test batteries.

## Methods

### *Family characterization*

To phenotype synesthetes, we leveraged the fact that synesthetes show greater consistency than non-synesthetes – that is, when asked to match stimuli to the colors they trigger, and then re-tested some time later, synesthetes are largely consistent in their matches, whereas control subjects have a distinguishably worse performance . For example, when subjects are asked to describe the color experienced on hearing items from a list of words, phrases, and letters, synesthetes show 92.3% consistency over a year later, while controls cluster around 32% consistency .

For the purposes of rigorous phenotyping, we have developed the Synesthesia Battery ([www.synesthete.org](http://www.synesthete.org)), an online battery of tests that sensitively discriminates synesthetes from non-synesthetes . During the battery, participants are serially presented with a full set of graphemes (A-Z, 0-9), three times each in a random order. For each grapheme, they are asked to choose the best associated color from an online palette of 16.7 million colors. The color choices are then analyzed for consistency within the session (e.g., did the participant choose a similar shade of green each time she saw the letter J?). Controls display low consistency when associating colors to graphemes over 108 trials, while synesthetes perform with high accuracy. Participants were retested after six months to ensure consistency.

Once we established synesthesia in each proband, we investigated the presence of synesthesia in other members of the family. All participants in this study completed the Synesthesia Battery to ensure proper classification. Each member of a pedigree was classified in one of three ways: synesthetic, unknown, or unaffected.

#### *Inclusion criteria*

Although synesthesia has been reported in many forms (e.g., tasting shapes, seeing sounds, etc), for our analysis we included only verified colored sequence synesthesias, which here include the triggering of color perception by letters, numbers, weekdays, or months. This decision was motivated by our analysis of 713 verified colored sequence synesthetes, which indicated that having a color experience triggered by one type of sequence (e.g., letters) gives ~79% likelihood of having synesthesia for the other types of sequences (weekdays, months or numbers), but no more likely than a random person in the population to possess another type of synesthesia (Eagleman and Cheng, in preparation). In contrast, other forms of synesthesia show a much lower and typically non-significant correlation, suggesting the possibility that colored sequences share a common genetic mechanism, while other forms of synesthesia may have different bases (including possibly non-genetic sources). Therefore, we only pursue the genetic etiology of colored sequence synesthesia here.

#### *Genotyping*

After obtaining informed consent, DNA from 48 individuals in five families of predominantly

Caucasian descent was obtained. Each pedigree contained at least four verified CSS synesthetes, as determined by the Synesthesia Battery. The saliva samples were collected using OraGene sample kits (OraGene, DNA Genotek). The samples were purified according to Oragene DNA purification protocol and quantified using PicoGreen (Invitrogen, Carlsbad, CA, USA). The concentration of all samples was normalized to 50 ng/ $\mu$ l. A single-nucleotide-polymorphism (SNP) array (Linkage-12 BeadChip, Illumina, Inc., San Diego, CA) containing 6,090 SNPs was used to genotype 24 affected and 24 unaffected individuals from five large families. The SNP markers were distributed approximately evenly throughout the genome based on February 2009 human reference sequence NCBI Build 37. The assay was performed using 4  $\mu$ L of DNA from each individual. Fluorescent image analysis was performed using BeadStudio Version 3 genotyping software which provides automatic cluster separation, genotype calling, and quality control.

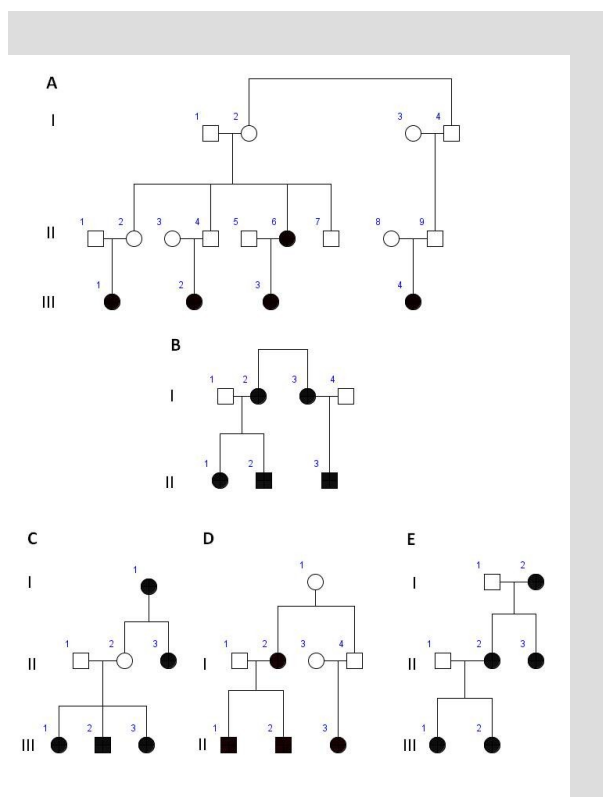
#### *Linkage analysis*

PEDCHECK was used to identify Mendelian inconsistencies, and the MERLIN program was used to detect double recombination events over short genetic distances that were probably due to genotyping errors. Parametric and Non-parametric linkage analysis (allele sharing method) was carried out using the ALLEGRO and MERLIN programs. A disease allele frequency of 0.001 was used for the analysis. Allele frequencies for SNP marker loci were obtained from the Hapmap data

using the CEU (Caucasian) population. For the parametric linkage analysis, an autosomal dominant mode of inheritance with reduced penetrance (0.8) and no phenocopies was used. Additionally, an affected-only parametric linkage analysis was performed also using an autosomal dominant mode of inheritance. For the parametric linkage analyses, HLOD scores were calculated allowing for linkage admixture. For the multi-point linkage analysis, genetic map distances were then obtained from the Rutgers combined linkage-physical map of the human genome using interpolation. It has been demonstrated that densely distributed SNP marker loci can increase the type-1-error rate due to Linkage Disequilibrium (LD) between marker loci. MERLIN can control for intermarker LD when present. Haplotypes were reconstructed using ALLEGRO.

## Results

Using the phenotyping and sample collection methods described above, we recruited five families with four or more individuals who have verified colored sequence synesthesia (Figure 3). We began our analysis with a three-generation family with multiple colored sequence synesthetes (Family A).



**Figure 3** Five families with colored sequence synesthesia. Rectangles designate males, circles designate females, and black designates a verified color sequence synesthete.

The synesthesia phenotype appears in each generation but contains several carriers without phenotype, such as A.II.2, A.II.4, and A.II.6. This observation is consistent with an autosomal dominant mode of inheritance with reduced penetrance. An HLOD score of 2.2 ( $p = 0.36$ ) was observed in the 16q12.2-23.1 region. Pedigrees A and B produce multipoint LOD scores of 2.6 and 0.9, respectively to this same region. The affected-only analyses yielded similar results.

Synesthetes in pedigrees A and B share the same haplotypes (Figure 4). The reconstructed haplotype of family A was used to define the 23 MB critical interval. This haplotype was not found in the other three

families nor did these families display a haplotype which segregated with the synesthesia phenotype in the 16q12.2-23.1 region, suggesting that families C, D and E are not linked to this region.

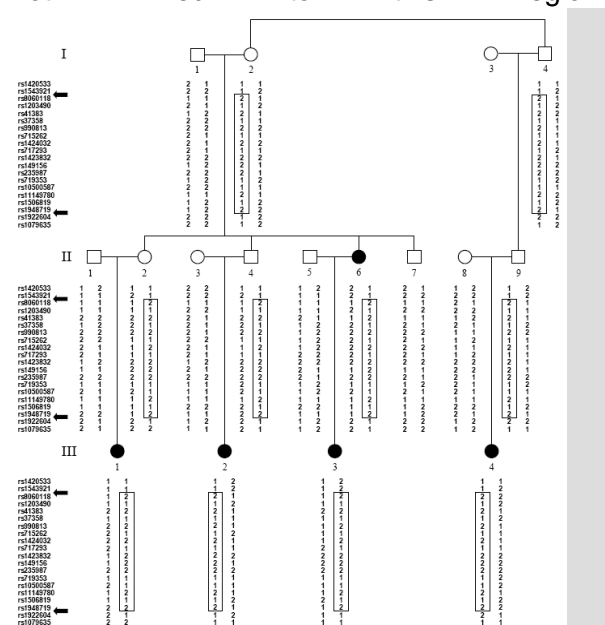


Figure 4. Boxes represent haplotypes in family A within the 23MB region of chromosome 16 which are common to all synesthetes. '1' represents the wild type and '2' represents the mutation that is identified by the SNP label in the list to the left. Incomplete penetrance explains the presence of the haplotype in unaffected individuals. Black symbols

### Investigation of candidate genes

The genomic region between markers rs1543921 and rs1922604 contains over 100 known genes that express in the brain based on UCSC (<http://genome.ucsc.edu>) and Ensembl (<http://www.ensembl.org>) genomic browsers. We selected six brain-specific genes whose known function and expression could potentially fit the hypothesis of synesthesia, and we sequenced the exon regions of these genes using intron-based,

exon-specific primers in two synesthetes from pedigree A and two synesthetes from pedigree B. The genes were: (1) GABARAPL2, a GABA(A) receptor-associated protein promotes GABA<sub>A</sub> receptor clustering and has been demonstrated to be an important receptor molecule in intercellular communication. (2) *NDRG4*, a member of the NDRG gene family that has been related to brain development. (3) PLPP, a gene involved in myelination and ion transport. (4) KATNB1 (WD repeat containing), which produces a microtubule-severing enzyme *katanin p80*. (5) Cytokine-induced apoptosis inhibitor 1 (CIAPIN1), a newly identified anti-apoptotic molecule that is ubiquitously expressed in normal human fetal and adult brains. (6) Guanine nucleotide-binding regulatory protein Go-alpha (GNAO1), a protein with significantly decreased expression in individuals with schizophrenia compared to unaffected family controls. The sequencing results were compared to the NCBI SNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP>), which contains 17 million human single nucleotide polymorphisms. None of the known polymorphisms segregated with all four of the affected individuals, and no novel polymorphisms were discovered.

### Discussion

Our family linkage analysis demonstrates suggestive linkage to the 16q12.2-23.1 region for colored sequence synesthesia, with a combined NPL score of 3.65 for two of the five families. These same two families also share a haplotype within our region, suggesting that these families express the same variant. Three of the five families do not support

linkage to this region, suggesting that there is genetic heterogeneity for the CSS phenotype.

Using DNA from four family members from families A and B, we sequenced six candidate genes: *GABARAPL2*, *NDRG4*, *PLPP*, *KATNB1*, *CIAPIN1*, and *GNAO1*. No mutations were found in the affected individuals, suggesting that these six genes do not contribute to the synesthetic phenotype. However, we have only thus far examined the exons and flanking intronic regions of these genes, and it remains possible that non-coding introns in this 23 MB region affect genetic expression or regulation. In addition, our region contains over 100 genes expressed in the brain, only 6 of which we have sequenced. In order to draw more definite conclusions, our next steps are to narrow the 23 MB region by collecting more families, and then sequence the region.

The male-female ratio of synesthetes has been previously examined in several studies, but with widely variable outcomes. Our sample of five families includes 78 individuals, only 46 of which were genetically informative and included in our genetic analysis. Twenty four of the 46 participants were synesthetes, of which nineteen were female and only five were male. This synesthetic male-female ratio of 1:3.8 is interesting in light of the fact that there is no gender imbalance within the families (ratio 1:1.42 for the 46 genetically-informative individuals in Fig 3; ratio 1:1.15 for the 84 individuals in the broader family trees). Note that in the past, male-female ratios have been estimated at 1:5.5 (n=26 subjects; , 1:7.7 (n=6 families; , 1:3.7 (n=84 families , and close to 1:1 . Larger sample sizes in the future will allow us to determine if the gender

bias we find is due to undersampling or a biological factor. In addition, higher numbers will allow us to determine if different pedigrees are consistent with an autosomal dominant mode of inheritance.

Our results identify a collection of candidate genes, positioning us to identify the first color sequencing synesthesia gene in the near future. We are currently collecting additional CSS families to provide additional evidence of linkage of CSS to the 16q12.2-23.1 region and to further refine the genetic interval. The fact that three families are not linked to the same locus may be due to genetic heterogeneity (non-allelic heterogeneity) for this phenotype, indicating that other genes may be involved in developing the colored sequence synesthesia phenotype.

If we succeed in identifying the responsible genes in the future, we will then be able to explore who in the non-synesthetic population might express it. That is, if the gene were expressed in non-sensory areas (e.g., leading to increased crosstalk between regions of the prefrontal lobe), such an individual would not be synesthetic by our definitions and perceptual tests. Instead, areas involved in reasoning, planning, decision-making, and others might have increased crosstalk. What would such a phenotype look like? We do not yet know, although we speculate that the gene could be an attractive candidate for the ability of highly creative people to derive connections between seemingly unrelated concepts . To summarize, if we successfully find the gene that correlates with synesthesia, our next task will be to find out who carries the gene in the general population and to search for unusual cognitive or behavioral traits.

Many examples illustrate the heavy crosstalk in the normal brain, including cross-sensory illusions , anatomical tracing , and sensory substitution experiments . The difference between the synesthetic and non-synesthetic brain therefore appears to be not whether there is crosstalk, but rather how much there is.

More generally, this research may shed light on the way in which abnormal neural crosstalk arises. Abnormal crosstalk, whether from defects in pruning connections or imbalances in inhibitory networks, is something which is poorly understood but thought to underlie aspects of both autism and schizophrenia .

### **Literature Cited**

Using synesthesia as a model for abnormal crosstalk, we hope to uncover the functional and molecular mechanisms of neural crosstalk; a discovery that could potentially improve our ability to characterize and treat developmental disorders such as schizophrenia and autism.

### **Acknowledgments**

We thank Deepak Sagaram and Robert LiKamWa for programming contributions. This work was supported in part by a grant from the Mind Science Foundation (DME).