Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1136/jmedgenet-2012-101209</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:10578921">http://nrs.harvard.edu/urn-3:HUL.InstRepos:10578921</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
ORIGINAL ARTICLE

Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population

Hansoo Park,1,2 Seungbok Lee,1,3 Hyun-Jin Kim,1,3 Young Seok Ju,1,4 Jong-Yeon Shin,1,5 Dongwan Hong,1,6 Marcin von Grothuss,2 Dong-Sung Lee,1,3 Changho Park,7 Jennifer Hayeon Kim,1 Boram Kim,1 Yun Joo Yoo,9 Sung-Il Cho,9 Joohon Sung,9 Charles Lee,2 Jong-II Kim,1,3,5,7 Jeong-Sun Seo1,3,4,5,7

ABSTRACT

Background Musical abilities such as recognising music and singing performance serve as means for communication and are instruments in sexual selection. Specific regions of the brain have been found to be activated by musical stimuli, but these have rarely been extended to the discovery of genes and molecules associated with musical ability.

Methods A total of 1008 individuals from 73 families were enrolled and a pitch-production accuracy test was applied to determine musical ability. To identify genetic loci and variants that contribute to musical ability, we conducted family-based linkage and association analyses, and incorporated the results with data from exome sequencing and array comparative genomic hybridisation analyses.

Results We found significant evidence of linkage at 4q23 with the nearest marker D4S2986 (LOD=3.1), whose supporting interval overlaps a previous study in Finnish families, and identified an intergenic single nucleotide polymorphism (SNP) (rs1251078, p=8.4×10−13) near UGT8, a gene highly expressed in the central nervous system and known to act in brain organisation. In addition, a non-synonymous SNP in UGT8 was revealed to be highly associated with musical ability (rs1418254, p=8.0×10−13), and a 6.2 kb copy number loss near UGT8 showed a plausible association with musical ability (p=2.9×10−4).

Conclusions This study provides new insight into the genetics of musical ability, exemplifying a methodology to assign functional significance to synonymous and non-coding alleles by integrating multiple experimental methods.

INTRODUCTION

Song as a communication signal and as an instrument in sexual selection has been recognised since it was first proposed by Darwin.1–5 Musical ability is a non-verbal and complex cognitive skill, and appears to have a latent biological basis in that infants can differentiate frequencies and ‘carry a tune’ without receiving extensive formal musical training.

Researchers have described certain aspects of how the architecture of the brain affects facets of musical ability. Perception and vocal production of singing seem to be based on the auditory and motor domains of the brain.4,5 Studies of impaired language skills with spared musical abilities and impaired musical abilities with normal language skills have revealed a dissociation between these two skill sets,6 leading to the proposal of a distinct mental module associated with separate neural substrates and a set of neurally isolatable processing components. A minority of humans exhibit extreme musical abilities in the form of either absolute pitch (the ability to accurately label tones with specific musical notes) or amusia (the inability to accurately identify and mimic tones).7,8

Recent studies have identified genetic components of musical ability. For example, absolute pitch has a significant familial basis and is predominant in females.9 A twin study has shown substantial heritability for musical ability10 and linkage studies have found loci for musical aptitude and absolute pitch.11,12 Some polymorphisms of specific genes in association with musical ability have begun to be reported, including variants of AVPR1A and SLC6A4.13,14

As part of the GENDISCAN study (GENe DIScovery for Complex traits in large isolated families of Asians of the Northeast), which was designed to investigate genetic influences on complex traits in extended Asian families of rural Mongolia, we investigated the processing of pitch using 1008 subjects from 73 families. It was expected that several points of the GENDISCAN study would increase the power of genetic loci discovery in normal complex traits, considering (1) the study population has little ethnic admixture, (2) consists of large extended families, and (3) represents a community-based population unbiased by health status.15

To overcome the difficulties of identifying genetic variations underlying common complex diseases, an approach that allows for recruitment of homogeneous and isolated populations was proposed. However, only a few studies have incorporated this approach due to difficulties in sample recruitments. The inner Mongolian steppes are still inhabited by small populations; geographically isolated populations are commonly found in rural provinces of Mongolia. We recruited Mongolian individuals from an isolated population with large extended pedigrees. These individuals possess a
homogeneous genetic background and close genetic affinity to populations of the northern part of East Asia. Previous binary familiarity tests have mostly been used to indicate whether or not each song part sounds similar to assess musical ability. By shifting the pitch of melody one semitone higher or lower, participants were asked to classify two melodies as the same or different. In this study, we created a test to analyse subjects’ acoustic outputs followed by hearing specific tones using cochlear implants (CI). There are advantages to this approach, which include the possibility to study musical ability as a whole and the better availability of subjects. We determined the pitch discrimination limen with a simulated CI coding strategy and employed the complementary nature of linkage- and association-based methods for musical ability. The functional importance of results was screened through the incorporation of data from exome sequencing and array-based comparative genomic hybridisation (aCGH). This combined approach provides a method by which to discover additional novel genetic loci underlying complex traits.

**METHODS**

**Study subjects and phenotype measurement**

In 2006, a total of 2008 volunteers were recruited in Dashbalbar, Dornod Province, Mongolia for the GENDISCAN project, which was designed to discover the genetic backgrounds of several complex traits (figure 1). For this project, we selected an isolated population composed of large extended families. This population is highly appropriate for gene mapping research due to its genetic homogeneity, decreased environmental heterogeneity, and restricted geographical distribution. Extended multi-generation families comprising a small number of founders are known to increase the genetic power. Traits included in this project are summarised in online supplementary table S1.

In this study, we chose 1008 individuals who are derived from 73 extended families and have precise pedigree structures. Table 1 lists descriptive characteristics of the study population. The average age of the participants is 31.0 years and 51.6% are women. The family structure in this population is very complicated, with multiple generations and many family pairs such as 1794 parent–offspring pairs, 734 full-siblings, 395 half-siblings, and 888 avuncular pairs. The average family size and standard deviation are 19.6 and 11.3, respectively. Peripheral blood sample was collected for each study subject, and DNA was extracted according to standard protocols. The extracted DNA was stored in solution at −20°C.

To examine the musical ability of subjects, we used a pitch-production accuracy (PPA) test based on the difference limen of a pitch paradigm in a psychophysical experiment with a simulated CI coding strategy. PPA is given by \((100−10×(|ν_1−ν_s|/ν_s×100))\), subtracting 10 points for each 1% error, where \(ν_s\) is the standard auditory frequency emitted by a pitch-producing device and \(ν_i\) is the vocal pitch frequency produced by the individual, who hear a specific tone through a headset and recite the sound. A harmonic tone complex with a sound pressure level of 70 dB intensity and sex-dependent fundamental frequency was used as a stimulus (see online supplementary table S2).

The participants with PPA values higher than 60 were categorised as individuals with good musical ability because they were consistently and accurately able to produce tones differing by less than a semitone from one another; the number of subjects with a PPA score over 60 was 357 (35.4%). However, for further analyses, participants with borderline PPA values between 50 and 70 were excluded to eliminate ambiguous PPA values; the number of subjects with PPA score over 70 was 268 (31.1%).
We used the Kosambi mapping function (derived from the Markov chain Monte Carlo method) to convert map distances into recombination fractions. For the multipoint linkage analyses, the Sequential Oligogenic Linkage Analysis Routines package was used.\textsuperscript{35} We performed 10,000 permutation tests using the lodadj option to obtain the empirical p value. In addition, we estimated the adjusted narrow-sense heritability (h\textsuperscript{2}) (i.e., the proportion of phenotype variance attributable to additive genetic variance). In all analyses, we used age and sex as covariates.

For further association analysis, 53 extended families composed of 630 family members were genotyped using an Illumina Human610-Quad BeadChip kit by Macrogen (Macrogen Inc, Seoul, Korea). We evaluated the Mendelian inconsistencies in single nucleotide polymorphism (SNP) data using PEDCHECK.\textsuperscript{34} Non-Mendelian genotype errors were detected using Merlin.\textsuperscript{36} SNP quality control assessment was done using the deCODE map) to convert map distances into recombination fractions. For the multipoint linkage analyses, the Sequential Oligogenic Linkage Analysis Routines package was used.\textsuperscript{35} We performed 10,000 permutation tests using the lodadj option to obtain the empirical p value. In addition, we estimated the adjusted narrow-sense heritability (h\textsuperscript{2}) (i.e., the proportion of phenotype variance attributable to additive genetic variance). In all analyses, we used age and sex as covariates.

For further association analysis, 53 extended families composed of 630 family members were genotyped using an Illumina Human610-Quad BeadChip kit by Macrogen (Macrogen Inc, Seoul, Korea). We evaluated the Mendelian inconsistencies in single nucleotide polymorphism (SNP) data using PEDCHECK.\textsuperscript{34} Non-Mendelian genotype errors were detected using Merlin.\textsuperscript{36} SNP quality control assessment was based on SNP call rate, marker error rate, and minor allele frequency (MAF); minimum per-SNP call rate of 99%, less than 1% marker error rate, and higher than 5% MAF. In addition, we also removed genotypes with Hardy-Weinberg equilibrium p values <1×10\textsuperscript{-6}. We focused on the putative linkage region in chromosome 4 for this analysis (1-LOD Unit Support Interval: 99–118 cM). A total of 5424 SNPs that met quality control criteria were included in the putative linkage region, and the FBAT tool in HelixTree software (V6.4, GoldenHelix) was used for family-based association test (FBAT), which can control population stratification or population admixture.\textsuperscript{15, 39}

The null hypothesis was ‘linkage and no association (sandwich variance)’,\textsuperscript{40} which can be useful for expanded pedigrees by calculating a robust variance. We used the generalised estimating equation for the FBAT test statistic, and hypothesised an additive model. The association result was adjusted by covariates of age and sex.

Genome-wide linkage scan and family-based association study under linkage region

We genotyped 862 samples from 70 families with deCODE 1039 microsatellite marker platform throughout the autosomes for genome-wide linkage analysis. We checked family relationships through PREST\textsuperscript{33} using an average identity-by-descent (IBD)-based method. PEDCHECK was used to examine Mendelian inconsistencies in genotype data.\textsuperscript{34} and non-Mendelian genotype errors were detected with SimWalk.\textsuperscript{35} After fixing the genotype errors, multipoint identity-by-descent-matrices were calculated at each 1 cM distance, and converted using the Markov chain–Monte Carlo method by LOKI.\textsuperscript{36} We used the Kosambi mapping function (derived from the deCODE map) to convert map distances into recombination fractions. For the multipoint linkage analyses, the Sequential Oligogenic Linkage Analysis Routines package was used.\textsuperscript{35} We performed 10,000 permutation tests using the lodadj option to obtain the empirical p value. In addition, we estimated the adjusted narrow-sense heritability (h\textsuperscript{2}) (i.e., the proportion of phenotype variance attributable to additive genetic variance). In all analyses, we used age and sex as covariates.

For further association analysis, 53 extended families composed of 630 family members were genotyped using an Illumina Human610-Quad BeadChip kit by Macrogen (Macrogen Inc, Seoul, Korea). We evaluated the Mendelian inconsistencies in single nucleotide polymorphism (SNP) data using PEDCHECK.\textsuperscript{34} Non-Mendelian genotype errors were detected using Merlin.\textsuperscript{36} SNP quality control assessment was based on SNP call rate, marker error rate, and minor allele frequency (MAF); minimum per-SNP call rate of 99%, less than 1% marker error rate, and higher than 5% MAF. In addition, we also removed genotypes with Hardy-Weinberg equilibrium p values <1×10\textsuperscript{-6}. We focused on the putative linkage region in chromosome 4 for this analysis (1-LOD Unit Support Interval: 99–118 cM). A total of 5424 SNPs that met quality control criteria were included in the putative linkage region, and the FBAT tool in HelixTree software (V6.4, GoldenHelix) was used for family-based association test (FBAT), which can control population stratification or population admixture.\textsuperscript{15, 39}

The null hypothesis was ‘linkage and no association (sandwich variance)’,\textsuperscript{40} which can be useful for expanded pedigrees by calculating a robust variance. We used the generalised estimating equation for the FBAT test statistic, and hypothesised an additive model. The association result was adjusted by covariates of age and sex.

Screening functional significance of candidates using exome sequencing and aCGH data integration

To assign a functional significance to candidates, we used exome sequencing data of 40 founders and 180K aCGH results of 30 founders, both of which were included in this study and previously genotyped in our group. The experimental summary of each is described in data supplement (see online supplementary tables S3–S5, supplementary methods). Among SNPs and short insertions/deletions (indels) called from exomes, we selected coding sequence SNPs and indels, and canonical splice-site variants as candidates, along with the copy number variants (CNVs) called from the aCGH experiment. Focusing on variants in the putative linkage region, we further narrowed our candidates by linkage disequilibrium (LD) estimation with the top 10 SNPs of our association study. Haplovew software (V3.2) was used for this LD estimation.

Among the candidates showing a significant level of LD, we selected one SNP and one CNV to be genotyped in our study population and compared their p values with the association results. For the SNP selected, three-dimensional (3D) modelling was conducted to predict its functional impact on the corresponding protein (see online supplementary methods).

RESULTS

Family-based linkage and association study

The heritability explained by the additive genetic portion of musical ability was estimated as 40% (p<0.0001, 95% CI 20.4% to 59.6%), and linkage regions with LOD>1.0 were found for musical ability from the genome-wide linkage scan (see online supplementary table S6). The maximum LOD score was 5.1 at chromosome 4q23 with the nearest marker D4S2986 (figure 2A), and the putative linkage region encompassing a maximum 1-LOD unit supports an interval range from 99 cM to 118 cM (figure 2B). In the next phase, we conducted FBAT to identify candidate variants within the putative linkage interval. Table 2 shows the top 10 SNPs that were significantly associated with musical ability, and all of these have reached the strict genome-wide significance of p<1×10\textsuperscript{-8}. The strongest association (p=8.4×10\textsuperscript{-17}) was found for rs12510781, an intronic SNP near UGT8 (MIM 601291), and the others were located near ALPK1 (MIM 607347) and ELOVL6 (MIM 603610), and the others were located near ALPK1 (MIM 607347) and ELOVL6 (MIM 611546).

Table 1 Descriptive characteristics of study participants

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of samples</td>
<td>1008</td>
</tr>
<tr>
<td>No. of females (%)</td>
<td>520 (51.6)</td>
</tr>
<tr>
<td>Mean (SD) age (in years)</td>
<td>31.0 (15.5)</td>
</tr>
<tr>
<td>No. of sample with PPA score (%)</td>
<td></td>
</tr>
<tr>
<td>≥70</td>
<td>268 (26.5)</td>
</tr>
<tr>
<td>≥60</td>
<td>357 (35.4)</td>
</tr>
<tr>
<td>&lt;60</td>
<td>651 (64.6)</td>
</tr>
<tr>
<td>&lt;50</td>
<td>594 (58.9)</td>
</tr>
<tr>
<td>Family information</td>
<td></td>
</tr>
<tr>
<td>No. of families</td>
<td>73</td>
</tr>
<tr>
<td>Mean size (SD) of family members</td>
<td>19.61 (11.3)</td>
</tr>
<tr>
<td>No. of pairs</td>
<td></td>
</tr>
<tr>
<td>Parent–offspring</td>
<td>1794</td>
</tr>
<tr>
<td>Full-sibling</td>
<td>734</td>
</tr>
<tr>
<td>Sister–sister</td>
<td>198</td>
</tr>
<tr>
<td>Brother–brother</td>
<td>167</td>
</tr>
<tr>
<td>Sister–brother</td>
<td>399</td>
</tr>
<tr>
<td>Half-sibling</td>
<td>395</td>
</tr>
<tr>
<td>Grandparent–grandchild</td>
<td>1202</td>
</tr>
<tr>
<td>Avuncular pairs</td>
<td>888</td>
</tr>
<tr>
<td>First cousins</td>
<td>598</td>
</tr>
</tbody>
</table>

PPA, pitch-production accuracy.

Utilisation of exome sequencing and aCGH data to assign functional significance to candidate variants

Among the candidates from the exome data (347 SNPs and seven indels in the putative linkage region), we narrowed down to four SNPs that were in strong LD with the top 10 SNPs identified via FBAT (r²>0.6, online supplementary table S7). We found that a non-synonymous SNP (nsSNP) in UGT8 (rs4148254) showed perfect LD with rs12510781, the most significant SNP from FBAT (r²=1.0), and this SNP was genotyped in 611 FBAT samples for the association analysis. As a result, the LD between rs4148254 and rs12510781 was re-estimated (r²=0.93), and the rs4148254 SNP was found to have the most significant association with musical ability in this study (p=8.0×10⁻¹⁷). The effect estimate of this SNP in founder samples was also higher than that of rs12510781 (OR=3.4, 95% CI 1.2 to 9.9 vs OR=3.0, 95% CI 1.1 to 8.2, online supplementary tables S8,S9). The 3D modelling of UGT8 protein showed that Pro226, which is changed to leucine by the SNP, Figure 2

Figure 2  Summary of genome-wide linkage and association results for musical ability. (A) Genome-wide linkage results for musical ability. (B) The linkage peak on chromosome 4 and association plot under the linkage support region. The linkage support interval is indicated by a green line (99–118 cm). The red dot is the top single nucleotide polymorphism (SNP) by family-based association test. The SNPs 2–10 are labelled with green dots. (C) Regional plot of association results for SNPs from analysis (−log₁₀ p) for UGT8 (±300 kb position from top SNP). The SNPs close to rs12510781, the most significant SNP (blue diamond), are colour-coded to reflect their linkage disequilibrium with this SNP (r²<0.2; white, 0.2≤r²<0.4; yellow, 0.4≤r²<0.8; orange, r²≥0.8; red).

Table 2  Top 10 SNPs significantly associated with musical ability by FBAT under the putative linkage region of chromosome 4

<table>
<thead>
<tr>
<th>SNP</th>
<th>*Position</th>
<th>Alleles</th>
<th>Effect</th>
<th>Other</th>
<th>Frequency of effect allele</th>
<th>p Value (FBAT)</th>
<th>tNearest gene(s)</th>
<th>Location (distance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12510781</td>
<td>115 860 030</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>0.12</td>
<td>8.4×10⁻¹⁷</td>
<td>UGT8</td>
<td>Intergenic (42.3 kb)</td>
</tr>
<tr>
<td>rs9307160</td>
<td>96 586 977</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>0.10</td>
<td>3.0×10⁻¹³</td>
<td>UNC5C</td>
<td>Intronic (°)</td>
</tr>
<tr>
<td>rs17628408</td>
<td>113 574 860</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>0.91</td>
<td>7.1×10⁻¹¹</td>
<td>ALPK1</td>
<td>Intronic (°)</td>
</tr>
<tr>
<td>rs2074385</td>
<td>113 598 098</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>0.91</td>
<td>7.1×10⁻¹¹</td>
<td>ALPK1</td>
<td>Intronic (14.8 kb)</td>
</tr>
<tr>
<td>rs4148255</td>
<td>115 764 226</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>0.88</td>
<td>2.7×10⁻¹⁰</td>
<td>UGT8</td>
<td>Synonymous (°)</td>
</tr>
<tr>
<td>rs11097397</td>
<td>95 087 875</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>0.28</td>
<td>4.8×10⁻¹⁰</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rs10024217</td>
<td>115 677 564</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>0.28</td>
<td>6.1×10⁻⁹</td>
<td>UGT8</td>
<td>Intergenic (61.4 kb)</td>
</tr>
<tr>
<td>rs1903364</td>
<td>115 681 713</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>0.28</td>
<td>6.1×10⁻⁹</td>
<td>UGT8</td>
<td>Intergenic (57.3 kb)</td>
</tr>
<tr>
<td>rs12504058</td>
<td>115 718 566</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>0.28</td>
<td>6.1×10⁻⁹</td>
<td>UGT8</td>
<td>Intergenic (20.4 kb)</td>
</tr>
<tr>
<td>rs6845765</td>
<td>111 177 613</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>0.86</td>
<td>8.2×10⁻³</td>
<td>ELOVL6</td>
<td>Intergenic (12.0 kb)</td>
</tr>
</tbody>
</table>

*Positions are based on Build 36 from NCBI.
†Nearest gene, within ±100 kb of the SNP.
FBAT, family-based association test; SNP, single nucleotide polymorphism.
might be part of the loop exposed outside of the predicted 3D structure, and the loop with the Pro226 residue contains sequence motifs including TRFH domain docking and USP7-binding motifs (see online supplementary figure S1).

At the level of CNVs, only one copy number (CN) loss was found to have moderate LD with rs4148255, the fifth most significant SNP in FBAT ($r^2=0.48$; online supplementary table S10). This CN loss (Chr4: 115 727 257–115 753 452) is located 5.6 kb upstream of the UGT8 gene. We genotyped it in 618 FBAT samples and the frequencies of heterozygous and homozygous CN losses were shown to be 45.15% and 10.03% in our study subjects (allele frequency = 32.61%). This CNV was negatively associated with musical ability ($p=2.9 \times 10^{-6}$) and, interestingly, a diploid status at this position was shown to potentiate the positive effect of rs4148254 in founders (see online supplementary table S11). In addition, we identified a significant interaction effect between this CNV and rs4148254 using a logistic regression model ($p=0.01$).

**DISCUSSION**

In this study, we explored the genetic determinants of musical ability by combining several methodologies, namely family-based linkage and association studies supported by exome sequencing and aCGH data analyses. This study was conducted as a part of the GENDISCAN project, which was designed to discover the genetic backgrounds of complex traits in Mongolia.

Musical ability is a well-known complex trait determined by multiple environmental and genetic factors. As this trait consists of several factors including perception, cognition, learning, and emotions, a variety of genes have an effect on one’s musical ability, both independently and interactively. To discover genetic backgrounds of these complex traits, studies should be designed from the first to increase the power to detect genetic loci. In this regard, our study has some strong points as described in the Introduction and Methods, which include little ethnic admixture and large extended families. In addition, we excluded samples with borderline phenotypes from all the analyses to derive more accurate results.

Our results support the view that musical ability is heritable and have shown significant evidence of linkage for musical ability in large families. Previously, a linkage study for musical aptitude was performed with samples in a small number of Finnish multigenerational families, composed of predominantly white subjects. That study found an association of the chromosomal region 4q22 with musical aptitude in the Finnish study population, which overlaps with our linkage interval on chromosome 4q. Despite several differences in methodology, we believe that overlapping results for musical ability in different ethnic populations enhance the reliability of this linkage region on chromosome 4q.

We also discovered common variants strongly associated with musical ability, suggesting a biological mechanism for this finding. Including the most significant, five SNPs among the top 10 were shown to lie near or within UGT8. In addition, there was no LD structure between rs12510781 and rs4148255. These two unrelated variants on one gene, associated with the same phenotype, increase the possibility of UGT8 being one of the true susceptibility genes for musical ability.

To identify more detailed causal variants, we integrated additional technologies such as exome sequencing and aCGH, resulting in the discovery of another rsSNP in UGT8 and a CN loss located 5.6 kb upstream of this gene. The SNP rs4148254, which changes amino acid 226 of the UGT8 protein from proline to leucine, was not included in the platform we used, and has shown a lower $p$ value than rs12510781 in our study population (see online supplementary figure S1A,B). Because the BLOSUM score predicts this to be damaging, the SNP might affect the function of the UGT8 protein. Moreover, this proline amino acid seems to be conserved among vertebrates (see online supplementary table S12). The three other SNPs (rs55508602, rs2074381, and rs3828559), which were in high LD ($r^2>0.6$) with the top 10 SNPs, were predicted to be benign by PolyPhen-2 and the BLOSUM scores were ‘2’, ‘1’, and ‘1’, respectively (see online supplementary table S7). In case of the CN loss, even though it was not more significant than the associated SNP allele, the synergetic effect of this variant with rs4148254 was suggested in the founder analysis.

The protein encoded by UGT8 is UDP glycosyltransferase 8, which is highly expressed in brain (see online supplementary figure S2). It is the first enzyme involved in complex lipid biosynthesis in the myelinating oligodendrocyte and clearance of long-chain ceramides (IcCer). IcCer clearance in neurons is mediated by glucosylceramide synthase (GCS) and studies have shown that decreased GCS leads to abnormally high IcCer. A significant early downregulation in glial GCS expression was associated with an increase in UGT8 mRNA in Alzheimer’s disease, and some patients with Alzheimer’s disease have been observed to preserve musical ability long after losing all other cognitive functions.

Although this study primarily focused on UGT8, there are other genes such as UNCS3, ALPK1, and ELOVL6 equally worth our attention. The protein encoded by UNCS3 plays a role in the chemoreceptive effect of netrin-1 in axon guidance. This gene was previously suggested as a susceptibility gene for musical ability in the Finnish linkage study. Regarding the other two, one study has shown that mice homozygous for disrupted copies of Alpk1 exhibited coordination defects, and ELOVL6 was once reported as one of the susceptibility loci for attention-deficit/hyperactivity disorder in a genome-wide association study. Several previous findings, as listed above, have supported the neural involvement of those candidate genes; however, more evidence should be given to associate them with musical ability.

Music is a complex cognitive skill in the neuronal network affected by several potential covariates. We first considered language ability as a potential covariate besides age and sex. However, we found no language skill defects in our study subjects, and previous studies have reported that it is possible for language skills to be impaired while musical abilities are spared (aphasia without amusia); likewise, musical abilities can be impaired while language skills are spared (amusia without aphasia). In addition, more factors including special musical training, education status, and education duration might be considered as potential covariates, since it has been reported that the skill of absolute pitch could be developed at a very young age by special musical training. However, our participants lived in an isolated area with a homogeneous culture, and most of them were educated in the same public school without any additional musical training. In this study, therefore, we did not take those factors into account for analyses.

In summary, we have demonstrated for the first time that common genetic variants in UGT8 are associated with musical ability, exemplifying a methodology to assign functional significance to the results of various association studies, which in many cases yield synonymous or non-coding alleles.
Genome-wide studies

Acknowledgements The authors appreciate the help of all study participants and collaborators. We thank Omer Gokcumen and Raju Govindarajan at Harvard Medical School for their personal comments regarding this manuscript.

Contributors J-SS planned and managed the project. J-IK, HP, JY, SY, SL, and JS recruited and measured phenotypes of the Mongolian samples. HP, H-JK, YJF, and J-IK analysed linkage data and family-based association studies. SL, J-YS, DH, and JK executed exome sequencing and analysed sequence data. HP, HI, D-SL, CP, JHK, and BK executed and analysed aCGH experiments. MvG performed the 3D modelling and motif analysis. CL supervised research at Brigham and Women’s Hospital/Harvard Medical School. J-SS, HP, H-JK, SL, and J-IK wrote the manuscript and CL edited the manuscript.

Funding This work was supported by the Korean Ministry of Education, Science and Technology (Grant No. 2003-2001558) and the US National Institutes of Health (Grant No. HG004221).

Competing interests None.

Ethics approval This study was approved by the Institutional Review Board of the Seoul National University Hospital (approval number, H-0307-105-002).

Data sharing statement The whole data of exome and aCGH experiments, some of which were used for this study, have not been published yet. We can provide the part of data related to this project upon request.

REFERENCES


752