

**JAANA OIKKONEN**

## **Genetics and Genomics of Musical Abilities**



GENETICS, DEPARTMENT OF BIOSCIENCES  
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES  
DOCTORAL PROGRAMME IN INTEGRATIVE LIFE SCIENCE  
UNIVERSITY OF HELSINKI

# GENETICS AND GENOMICS OF MUSICAL ABILITIES

Jaana Oikkonen

Department of Biosciences,  
Faculty of Biological and Environmental Sciences  
University of Helsinki  
Finland

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in Infocenter Korona, Lecture hall 2, on 7<sup>th</sup> October, 2016 at 12 noon.

HELSINKI 2016

**Supervised by:** Dr. Päivi Onkamo  
Department of Biosciences, University of Helsinki, Helsinki, Finland  
*and*

Dr. Irma Järvelä  
Medicum, University of Helsinki, Helsinki, Finland

**Reviewed by:** Prof. Pekka Uimari  
Department of Agricultural Sciences, University of Helsinki, Helsinki,  
Finland

*and*

Prof. Fredrik Ullén  
Department of Neuroscience, Karolinska Institutet, Stockholm,  
Sweden

**Opponent:** Prof. Suzanne M. Leal  
Molecular and Human Genetics, Baylor College of Medicine, Houston,  
Texas, US

**Custos:** Prof. Liisa Holm  
Institute of Biotechnology, University of Helsinki, Helsinki, Finland

**Members of the advisory committee:**

Prof. Pekka Uimari  
Department of Agricultural Sciences, University of Helsinki, Helsinki,  
Finland

*and*

Dr. Rainer Lehtonen  
Institute of Biomedicine, University of Helsinki, Helsinki, Finland

ISBN 978-951-51-2463-0 (print)

ISBN 978-951-51-2464-7 (PDF)

ISSN 2342-3161 (print)

ISSN 2342-317X (online)

<http://ethesis.helsinki.fi/>

Cover art: Jaana Oikkonen

Cover layout: Anita Tienhaara

Press: Hansaprint

*To my extended family*



# Table of contents

LIST OF ORIGINAL PUBLICATIONS.....	6
ABBREVIATIONS .....	7
ABSTRACT .....	8
TIIVISTELMÄ .....	10
1 INTRODUCTION.....	12
2 REVIEW OF LITERATURE.....	13
2.1 BEHAVIOURAL GENETICS .....	13
2.1.1 Structure of the genome.....	13
2.1.2 Genetics of traits .....	14
2.1.3 Genome-wide gene mapping strategies for complex traits .....	15
2.1.3.1 Linkage.....	15
2.1.3.2 Association .....	17
2.1.3.3 Selection signature analysis methods.....	18
2.1.3.4 Multiple testing .....	19
2.1.3.5 Biological significance.....	20
2.2 MUSICAL ABILITIES .....	21
2.2.1 Biological origins of music .....	21
2.2.2 Evolutionary origins of music.....	21
2.2.3 Music-related phenotypes and measurements.....	22
2.2.4 Heritability of musical abilities.....	24
2.2.5 Physiological effects of music .....	26
2.2.6 Genetics of musical abilities .....	29
2.2.7 Music perception and production in animals .....	29
3 AIMS OF THE STUDY .....	32
4 MATERIALS AND METHODS.....	33
4.1 STUDY MATERIALS (I, II, III) .....	33
4.1.1 Sample.....	33
4.1.2 Genotypes .....	33
4.1.3 Phenotypes.....	33
4.2 METHODS (I, II, III).....	35
4.2.1 Heritabilities and correlations .....	35
4.2.2 Gene mapping (I, II) .....	35
4.2.2.1 Quality control.....	35
4.2.2.2 Linkage and association analysis (I).....	35
4.2.2.3 Linkage and linkage disequilibrium analysis (II).....	37
4.2.2.4 Enrichment analyses .....	37
4.2.2.5 Regional association plots .....	38
4.2.2.6 Prediction of putative regulatory sites (I) .....	38
4.2.3 Positive selection (III) .....	38
4.3 MATERIALS AND METHODS IN CONVERGENT EVIDENCE ANALYSIS (IV)..	39
4.3.1 Data .....	39
4.3.2 CFG method.....	41
4.3.3 Enrichment and network analyses.....	41

5	RESULTS.....	43
5.1	DESCRIPTIVE STATISTICS (I, II, III) .....	43
5.2	HERITABILITIES OF MUSICAL TRAITS (I, II) .....	45
5.3	GENE MAPPING OF MUSICAL APTITUDE (I).....	45
5.3.1	Linkage.....	46
5.3.2	Linkage disequilibrium .....	46
5.3.3	Enriched functions and pathways among the suggested genes .....	48
5.4	GENE MAPPING OF CREATIVE ACTIVITIES IN MUSIC (II) .....	48
5.4.1	Linkage.....	48
5.4.2	Joint linkage and linkage disequilibrium .....	48
5.4.3	Enriched functions and pathways among the suggested genes .....	48
5.5	REGIONS UNDER POSITIVE SELECTION IN MUSICAL APTITUDE (III) .....	49
5.5.1	haploPS.....	49
5.5.2	XP-EHH .....	50
5.5.3	$F_{ST}$ .....	51
5.5.4	Difference between the metrics.....	51
5.5.5	Enrichment analysis.....	52
5.6	CONVERGENT EVIDENCE OF MUSICAL ABILITIES (IV).....	52
6	DISCUSSION.....	56
6.1	GENOMIC REGIONS AFFECTING MUSICAL ABILITIES.....	56
6.2	LIMITATIONS IN THE GENE MAPPING STUDIES (I, II).....	56
6.3	DETECTED SELECTION SIGNATURE REGIONS.....	57
6.4	GENES SUGGESTED BY THE CFG METHOD.....	58
6.5	BIOLOGICAL FUNCTIONS RELATED TO MUSICAL TRAITS .....	59
6.5.1	Cognition.....	59
6.5.2	Auditory pathways and the inner ear .....	60
6.5.3	Speech, music and birdsong.....	60
6.6	PATHWAYS CHARACTERISTIC TO MUSICAL ABILITIES .....	61
7	CONCLUSIONS AND FUTURE PROSPECTS.....	63
8	ACKNOWLEDGEMENTS.....	65
9	REFERENCES.....	67

# List of original publications

This thesis is based on the following original publications, which are referred in the text by their Roman numerals:

- I. J. Oikkonen, Y. Huang, P. Onkamo, L. Ukkola-Vuoti, P. Raijas, K. Karma, V. J. Vieland and I. Järvelä. “A genome-wide linkage and association study of musical aptitude identifies genetic variants related to inner ear development and neurocognitive functions” 2015, *Molecular Psychiatry*, 20, 275-282.
- II. J. Oikkonen, T. Kuusi, P. Peltonen, P. Raijas, L. Ukkola-Vuoti, K. Karma, P. Onkamo, I. Järvelä. “Genome-wide analysis of self-reported creative activities in music” 2016, *PLOS One*, 11: e0148679.
- III. X. Liu, C. Kanduri, J. Oikkonen, K. Karma, P. Raijas, L. Ukkola-Vuoti, Y.-Y. Teo, I. Järvelä. ”Detecting signatures of positive selection associated with musical aptitude in the human genome” 2016, *Scientific Reports*, 6: 21198.
- IV. J. Oikkonen, P. Onkamo, I. Järvelä, C. Kanduri. “Convergent evidence for the molecular basis of musical traits” bioRxiv doi: <http://dx.doi.org/10.1101/061358> [Submitted manuscript]

## Author’s contributions

- I. Participated in designing the study. Performed data management, quality control procedures, literature search for candidate genes, haplotype association analysis, prediction of putative regulatory sites and enrichment analysis. Wrote the manuscript together with other authors.
- II. Participated in designing the study. Performed the linkage and linkage disequilibrium analyses, quality control procedures, literature search for candidate genes and enrichment analysis. Wrote the manuscript together with other authors.
- III. Performed data management and quality control procedures. Wrote the manuscript together with other authors.

Participated in designing the study. Collected the database and performed analyses. Wrote the manuscript together with other authors.

# Abbreviations

AMMA	Advanced Measures of Music Audiation
AMPAR	$\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
AP	Absolute pitch
AR	Androgen receptor
bp	Base pair
CFG	Convergent functional genomics
cM	Centimorgan
COMB	Combined musical aptitude scores (including KMT, SP and ST)
DTT	Distorted tunes test
EDU	Musical education
E-S	Elston-Stewart, algorithm for genetic linkage
FBAT	Family-based association test
FDR	False discovery rate
F <sub>ST</sub>	Fixation index, F-statistics
GWAS	Genome-wide association study
HGNC	HUGO Gene Nomenclature Committee
HRR	Haplotype-relative-risk
HVC	High vocal center
H <sup>2</sup>	Heritability
h <sup>2</sup>	Additive genetic heritability
IBD	Identical by descent
IEG	Immediate early gene
IPA	Ingenuity pathway analysis
KMT	Karma music test
LD	Linkage disequilibrium
LOD	Logarithm of odds
LTD	Long-term depression (neurophysiological process)
LTP	Long-term potentiation (neurophysiological process)
L-G	Lander-Green, algorithm for genetic linkage
MBEA	Montreal Battery of Evaluation of Amusia
N	Number of observations
NCM	Caudomedial nidopallium, forebrain auditory region
NCNA	Neither composing nor arranging
PPA	Pitch production accuracy
PPL	Posterior probability of linkage
PPLD	Posterior probability of LD
Pr	Probability
QTL	Quantitative trait locus
RA	Robust nucleus of the arcopallium
RNA	Ribonucleic acid
RR	Risk ratio
SMDT	Swedish Musical Discrimination Test
SNP	Single nucleotide polymorphism
SP	Seashore's test for pitch
ST	Seashore's test for time
TDT	Transmission disequilibrium test
VC	Variance-components

# Abstract

Music is an integral part of human culture. Most people have the capacity for music perception and production, but the degree of music competency varies between individuals. My aim is to study the genetic components affecting this variation. In this thesis, I studied abilities to identify pitch, tone duration and sound patterns, which can be considered as basic components of musicality, as well as realisation of music abilities, namely music practice, musicianship, composing, arranging and improvising. Musical ability is a diverse phenotype that includes both acquired and innate abilities. Earlier studies have already shown some genetic components affecting musical traits. However, the genetic basis of musical ability has remained mostly unknown.

Genetic predisposition of musical aptitude was first studied in family material (N=915) using three tests of musical aptitude: Karma's test for auditory structuring (KMT), and Seashore's tests for time (ST) and pitch (SP). Heritability estimates of the test results ranged from 21-68%. Genomic regions were assigned using linkage and linkage disequilibrium analyses implemented in KELVIN. The best association was obtained at 3q21.3 and the best linkage at 4p14. The best-associated region was near the *GATA2* gene, which is an important transcription factor and is related to hearing. The linked region at 4p14 is adjacent to the *PCDH7* gene, a neuronal receptor with also a suggested role in cochlear complexes. There were also other linked regions at chromosomes 4, 16, 18 and 22. The genes within these linked regions showed enrichment of inner ear development and schizophrenia related genes.

In the second part of this thesis, more complex features of musical abilities were studied: music-related creative activities. Again, the heritability estimates settled to approximately 30% for arranging and composing. Somewhat surprisingly, I found significant linkage at 18q21 for musically experienced individuals without creative activity in music (neither composing nor arranging). This region contains, for example, cadherin genes *CDH7* and *CDH19*. The creative activities in music were suggestively linked to 4q22.1 and 16p12.1-q12.1. Joint analysis of linkage disequilibrium and linkage of composing showed enrichment of the long-term depression pathway, which is a molecular pathway important in learning and memory.

The genetics of music test results were further studied with selection signature methods that can reveal regions under positive selection characteristic to the studied phenotype. Using three methods,  $F_{ST}$ , haploPS and XP-EHH, over 100 candidate genomic regions were detected to be under positive selection in individuals with high scores in the music tests but not in the individuals with low scores. However, there were only a few overlaps between the results from the three methods used. Enrichment analysis pointed to the development of the inner ear, corresponding to the enrichment results from the musical aptitude linkage analysis, even though there were no common regions between these two studies.

The fourth part of this thesis integrates our musical aptitude gene mapping results with other music-related studies. A convergent genomics approach was used to rank results from human musical trait studies and from music-related studies using other animals. Genes and biomarkers showing the most prominent association to musical abilities included *EGR1*, cortisol, *FOS* and *FOXP2*. The best 40 genes showed enrichment of cognition. Previous studies have shown that musical abilities share a common background with cognitive abilities, such as intelligence, and this study is the first to suggest the associated molecules.

In conclusion, with these studies, I suggest many new genes and pathways to be associated with musical traits. I also show that music has a genetic component that is partly overlapping with overall cognition and hearing. Some evidence of a common evolutionary background of birdsong, language and music was obtained. The genetic predisposition for music is affected by a large number of genes of which only a few have been identified. Here, a new approach using a convergent genomics method and integration of animal model results in the study of musical traits has been introduced that can also be used to evaluate genomics results in the future.

# Tiivistelmä

Musiikki on olennainen osa kulttuuriamme eikä tunneta yhtään ihmisen kulttuuria, jossa ei olisi musiikkia. Musiikin on todettu vaikuttavan ihmisiin sekä muihin eläimiin: musiikin kuuntelu ja tuottaminen muuttavat hormonituotantoa, rentouttavat, nopeuttavat paranemista sekä muuttavat aivoja jopa pysyvästi. Lähes kaikilla ihmisillä on kyky musiikin kuunteluun ja tuottamiseen, mutta yksilöiden välillä on eroja: kaikki eivät kasva musiikin ihmelapsiksi ja joillekin sävelen muistaminen on helpompaa kuin toisille. Väitöskirjatyöni aiheena on selvittää musikaalisuuden geneettistä taustaa, sitä yksilöiden välistä eroa, joka tekee meistä musiikillisesti erilaisia. Musikaalisuus sisältää paljon erilaisia ominaisuuksia, joista tässä työssäni keskityn kuullun musiikin erotteluun perustuviin Karman ja Seashoren testeihin, jotka mittaavat sävelkorkeuden, -pituuden ja rakenteiden erottelua, sekä musiikin tuottamiseen liittyviin ominaisuuksiin sisältäen erityisesti säveltämisen ja sovittamisen.

Ensimmäisessä osatyössä tutkimme assosiaatio- ja kytkentäanalyysin geneettisiä tekijöitä, jotka vaikuttavat musikaalisuustesteissä pärjäämiseen. Aineistona tutkimuksessa on suomalaisia sukuja (N=915). Heritabiliteetin, eli geneettisen osuuden yksilöiden välisestä eroista, arvioitiin olevan 21-68%. Paras assosiaatio löydettiin kromosomista 3q21.3 läheltä *GATA2*-geeniä, jonka tiedetään vaikuttavan mm. kuuloon sekä aivojen kehitykseen. Parhaiten kytkeytyi alue kromosomista 4p14, jonka lähellä assosioitui *PCDH7*-geeni. *PCDH7* on neuronireseptori, joka vaikuttaa myös kuuloon. Muita kytkeytyneitä alueita löydettiin kromosomeista 4, 16, 18 ja 22. Kytkeytyneiltä alueilta löytyneet geenit liittyivät oletettua enemmän sisäkorvan kehitykseen ja skitsofreniaan.

Musiikillisen luovuuden perimän haravoinnissa tutkimme säveltämiseen ja sovittamiseen liittyviä genomien alueita. Näiden ominaisuuksien heritabiliteetiksi arvioitiin noin 30%. Kromosomi 18q21 kytkeytyi musiikillisesti harrastuneisiin, mutta säveltämistä tai sovittamista harrastamattomiin, yksilöihin. Tämä alue pitää sisällään mm. kadheriinegeenit *CDH7* ja *CDH19*. Lisäksi löytyi viitteitä säveltämisen ja sovittamisen kytkeytymisestä kromosomeihin 4q22.1 ja 16p12.1-q12.1. Kytkeäpätasapainoa hyödyntävän analyysin osoittamilta alueilta löytyi oppimiseen ja muistiin liittyvään LTD-reaktoreittien vaikuttavia geenejä.

Kolmannessa osatyössä, musikaalisuuden evoluution tutkimuksessa, etsimme perimän alueita, joihin on kohdistunut positiivista valintaa ja jotka ovat ominaisia musikaalisuustesteissä pärjäävillä. Käytimme kolmea mittaria, haploPS, XP-EHH ja  $F_{ST}$ -laskentatapoja. Näillä menetelmillä löytyi perimän alueita, joissa sijaitsee sisäkorvan kehitykseen ja aivojen toimintaan vaikuttavia geenejä. Tutkimuksessa löytyi myös samoja geenejä, joiden tiedetään vaikuttavan puheen kehitykseen ja laululintujen laulamiseen. Eri menetelmien tulosten välillä oli kuitenkin vain vähän yhtäläisyyttä. Vaikka sisäkorvan kehitykseen liittyvät geenit olivat tässäkin osatyössä rikastuneet, ei positiivisesti valikoituneiden alueiden ja ensimmäisen osatyön kytkentäalueiden välillä ollut päällekkäisyyttä.

Viimeisessä osatyössä yhdistimme aiemmat osatyöt muihin musikaalisuuden geeni- ja molekyyli tutkimusten tuloksiin. Lisäksi hyödynsimme eläinmalleilla saatuja musiikkiin liittyviä tuloksia, kuten linnunlaulun geenitutkimuksia ja hiirien musiikkialtistuksen vaikutuksista tehtyjä molekulaarisia tutkimuksia. Konvergoivan genomikan menetelmällä arvoitimme näistä kaikista tutkimuksista saatuja tuloksia ja tunnistimme ne geneettiset tekijät, jotka näiden kaikkien tutkimusten valossa todennäköisimmin vaikuttavat musikaalisuuteen. Todennäköisimmiksi arvioitujen geenit ja hormonit sisälsivät *EGR1*-, *FOS*- ja *FOXP2*-geenit sekä kortisolin. Parhaiden 40 molekyylin joukossa oli rikastuneina erityisesti kognitiivisiin kykyihin liittyviä geenejä. Aiemman tutkimuksen valossa tiedetään, että kognitiivisilla kyvyillä ja musikaalisuudella on osittain yhteinen tausta, mutta yhteisiä geenejä ei ole tunnettu.

Musikaalisuus on monitekijäinen ominaisuus, johon vaikuttavat sekä synnynnäiset kyvyt että opitut taidot. Oletettavaa on, että musiikkiin liittyy satoja, jopa tuhansia geenejä. Näiden geenien tunteminen voi auttaa vastaamaan mm. kysymyksiin siitä, miksi musiikki on säilynyt ihmisen evoluutiossa, miten ja miksi musiikki meihin vaikuttaa, minkälaiset biologiset erot johtavat eroihin musikaalisuudessa tai onko puheen ja musiikin geneettinen tausta yhteinen. Tutkimalla normaalia vaihtelua ihmisten välillä – kuten musikaalisuutta – voimme selvittää mihin biologisiin mekanismeihin normaalit kognitiiviset toimintomme aivoissa perustuvat.



# 1 Introduction

Music is an integral part of human culture. Most people are interested in music and music has measurable biological effects in the body (VanderArk & Ely, 1993; Fancourt et al., 2014). There is genetic predisposition for music; even at birth, we can identify certain musical abstracts like pitch, tempo and consonance (Trehub, 2001). However, the genetic background for mediating the effects of music as well as for musical abilities has mainly been unknown.

Music itself is biologically interesting, as it has proved to affect brain plasticity (Hyde et al., 2009b), mood (Schulreich et al., 2014) and recovery from many neuropsychiatric and physiological issues (see for example Särkämö et al., 2016). Music is a strong factor that serves as a model for interaction between genes and the environment. The genetic factors behind musical abilities may help in understanding the molecular mechanisms of brain plasticity. Music-specific biological pathways could increase knowledge about music processing as well as the evolutionary roots of music.

The publication of the human genome in 2001 enhanced progress in the research of human behavioural genetics. The research started from behavioural disorders and expanded to also cover normal cognitive variation such as intelligence, memory and music. The genetic study of specific cognitive functions helps to identify genetic domains affecting overall cognition. For example, many genes known from psychiatric diseases also have an effect on normal cognitive functions (St Pourcain et al., 2013; Kavanagh et al., 2015). Thus, studying music may also complement the study of related phenotypes like speech and creativity.

Overall, musical ability includes a wide variety of skills both inherited and acquired. Our hypothesis is that genetic variants associated with musical abilities can have a pivotal role in musical practices and music evolution. In this study, we have assessed musical aptitude using the auditory structuring ability test (Karma Music Test, KMT) and Carl Seashore's subtests of pitch (SP) and time discrimination (ST). At the same time, I acknowledge that the tests cover only a limited portion of the abilities important in music. Thus, I have also studied composing and arranging, which show more complex features of musical abilities.

Most musical abilities are human-specific but some can be identified from other animals as well. For example, birdsong includes many musical features (Rothenberg et al., 2014) and the effects of music listening can be studied in animals. Similarly, basic components of musical abilities, like pitch memory, can be found from any animals able to hear.

This study was performed to identify genetic loci and variants associated with musical aptitude and creative activities in music. In addition, we assigned candidate regions showing positive selection with musical aptitude in the human genome. Finally, we combined genomic data from studies on humans and animals to identify common genes and pathways underlying music-related traits.

## 2 Review of literature

### 2.1 Behavioural genetics

#### 2.1.1 Structure of the genome

The human genome consists of 23 chromosomes; diploid organisms like humans inherit two copies of every chromosome, one from their mother and one from their father. Exceptions are the sex chromosomes (allosomes) and mitochondrial DNA, where mitochondrial DNA and one X-chromosome are always inherited from the mother. The father transmits his Y-chromosome to his sons and the X-chromosome to his daughters.

The genetic code of the human genome was published in 2001 (Cravchik et al., 2001; Lander et al., 2001). The human genome includes approximately 3.2 billion base pairs (bp) and 20,000 genes (Harrow et al., 2012). Until now, thousands of human genomes from different individuals have been published and a human reference genome has been compiled. The reference genome enables, for example, identification of novel mutations, genome-wide methods for gene mapping and understanding normal variation in the genome.

The DNA sequence is highly similar between individuals; 99.9% of our genomes are identical. The remaining differences between individuals cause, for example, one person to have dark hair and another to develop Huntington disease. The genetic variation is essentially caused by mutations. On average, every individual acquires 50-100 new mutations that they will pass to offspring (Lynch, 2010). These mutations are transmitted over generations where genetic recombination rearranges and random assortment shuffles these variants into gametes (Figure 1). As a result, it is very unlikely that any two individuals – with the exception of monozygotic twins – have exactly the same genome. Evolutionary processes change the prevalence of the genetic variants over time. In general, beneficial mutations will become more common and deleterious mutations will vanish. Neutral mutations drift stochastically in frequency over time. Different types of mutations exist: substitutions, deletions, duplications, inversions, insertions and translocations. Simple, single nucleotide variants, where an ancient substitution has drifted to a population frequency of at least 1% are called SNPs. Variants where a short DNA tract is repeated in variable times is called a microsatellite.

The 20,000 genes of the human genome are not constantly translated. Gene regulation dictates which genes are read and which silenced. Gene regulation makes it possible for different cells to develop in different directions. Some genes are silenced and some are enhanced in order for the cell to act appropriately. The cell's own actions, other cells and the overall environment affects the cell, mainly through gene regulation. Outside of genes, the genome includes numerous other DNA elements that affect gene regulation. The environment can change genome function, for example through epigenetics and DNA binding factors. Thus, the genome is not a straight-forward representation of genetic code but a constantly changing

cooperation of substances in the cell, the environment, the genetic code and other elements affecting the genome. One example of gene regulation is alternative splicing, where different proteins are produced from the same gene by choosing which exons are included at which time and in which tissue.

### 2.1.2 Genetics of traits

The first step to studying the genetics of any phenotype is to search for familial aggregation of the trait. If the phenotype has a genetic component affecting its outcome, relatives are more similar to each other than random individuals from the same population. Concerning heritable traits like height, siblings tend to resemble each other more than unrelated individuals. With heritable diseases, we can usually find families where there are multiple patients in one family.

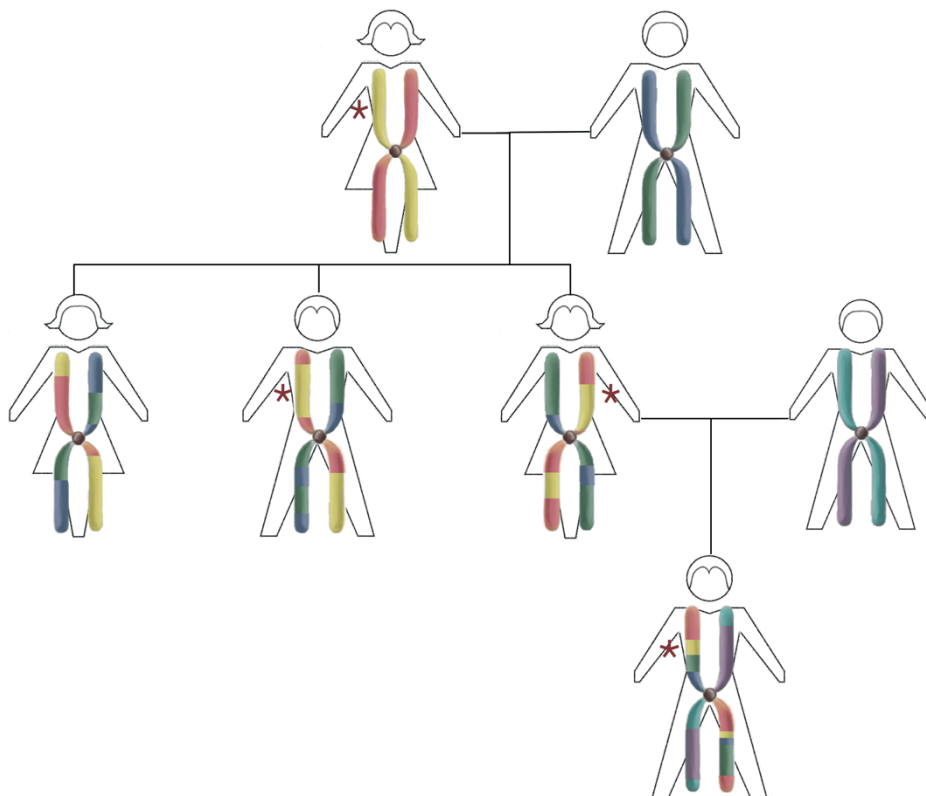
Some phenotypes like albinism are caused by a single gene. These traits are called Mendelian as they are inherited by the rules of Gregor Mendel (Mendel, 1866). Behavioural and other common traits are typically complex as they are affected by multiple sources of variation including many genetic components and environmental factors. The genetic predisposition of complex traits may include tens or even thousands of variants affecting the trait. The complex phenotypes have typically quantitative nature whereas single gene traits are typically discrete.

Resemblance between relatives can be used to estimate the proportion of the phenotypic variance that is due to genetic effects. For binary traits (i.e. dichotomous or case/control), relative risk (alias risk ratio, RR) measures the ratio of the probability for the disease between the normal population and the relatives of the affected individuals. For quantitative traits, heritability ( $H^2$ ) is used. Heritability gives the fraction of phenotypic variance explained by genetic variation. The more commonly used narrow-sense heritability ( $h^2$ ) considers only additive genetic variance. It is defined as:

$$h^2 = \frac{\text{additive genetic variance}}{\text{phenotypic variance}}.$$

In this thesis, heritability refers to this narrow sense heritability. One should also remember that the heritability estimate is always population specific as the genetic and environmental variation may differ between populations; thus,  $h^2$  estimated in one population does not even theoretically have to approximate the estimate from another population. The most reliable way of estimating the genetic influence on a trait is via twin studies where all variation between the monozygotic twins is caused by the environment. Naturally, it is not always possible to carry out a twin study.

For Mendelian traits, the mode of inheritance can be evaluated with segregation analysis, which is statistical detection of how phenotypic variation is passed to the next generation. This can be done with family material, estimating the phenotype ratios in offspring. The mode of inheritance tells whether the trait is inherited in dominant, recessive or additive fashion. However, with complex traits it may be impossible to know if the complex pattern is caused by incomplete penetrance, multiple loci or some other reason.



**Figure 1. Genetic recombination shuffles genetic variants from the mother and father into the offspring’s genomes.** Here, the inheritance of one fictitious diploid chromosome is shown in a pedigree. The original founder chromosomes in this family are shown in different colours. In general, siblings share on average half of the genetic variants. As an example, a variant linked to widow’s peak has been marked with asterisks.

Concerning music-related traits studied in this thesis, previous research refers to the traits as being complex rather than Mendelian (see also chapter 2.2.3 in this thesis). Thus, the following introduction to gene mapping methods focuses on methods applicable for complex traits and family data, because the data in this thesis consists mostly of families.

### 2.1.3 Genome-wide gene mapping strategies for complex traits

Gene mapping methods are used to locate genetic factors contributing to the variation of the studied phenotype. The locations of the genes are determined utilising simple genetic variants (SNPs or microsatellites) as landmarks. The relationship between the phenotypes and genetic variants can be studied in many ways. Two basic classes of gene mapping methods are linkage and association analyses.

#### 2.1.3.1 Linkage

In general, linkage analysis measures co-segregation of phenotype and genotypes. Figure 1 shows an example where widow’s peak is inherited together with a locus marked with an asterisk. In chromosomes, markers that are physically close to each

other are more probably inherited together. In linkage analysis, we essentially try to detect the most probable location for the underlying causative genetic locus (often called disease locus) through the observed markers. This unknown disease locus is inferred from the observed trait phenotypes. Family members sharing the same variants near the disease locus should resemble each other phenotypically. Linkage analysis gives the likelihood of a certain genomic marker locus to be linked to the disease locus in the studied families.

Most linkage methods are based on either the Elston-Stewart (E-S) (Elston & Stewart, 1971) or Lander-Green (L-G) (Lander & Green, 1987) algorithms (Ott & Hoh, 2000). The basic difference between the algorithms is that E-S factorises likelihood calculation by individuals and L-G by markers. As the calculation of the linkage is computationally expensive, E-S can handle large families but only a small number of markers whereas L-G can handle a large number of markers but only simple families. In many modern methods, these algorithms have either been extended to overcome some of this complication or some combination of these two have been used.

The linkage methods are further divided into parametric and non-parametric methods. In parametric linkage analysis, we need to specify the mode of inheritance including, for example, the genetic model (additive, recessive or dominant), the allele frequency in the population and the penetrance. Non-parametric methods do not require the mode of inheritance to be specified, however, parametric methods are often statistically more powerful than non-parametric methods (like in any statistical testing). As mentioned above, it may be hard to solve the mode of inheritance underlying a complex trait. Parametric methods are sensitive to model misspecification which can lead to spurious results and reduction of power (Almasy et al., 2015). Thus, non-parametric, or model-free, methods are preferred when the model is unknown.

Many non-parametric methods are based on allele sharing probabilities (see for example Kruglyak et al., 1996; Kong and Cox, 1997). The most common approach is to calculate the proportion of shared alleles that are inherited from the same ancestor. These are said to be identical by descent (IBD). The methods compare the IBD sharing alleles between affected relatives: expected versus observed proportions. When a marker is linked to the disease, the IBD between affected relatives will increase from its expected value.

Concerning quantitative traits, where simple dichotomous division to affected and unaffected individuals is not possible, variance-components (VC) methods are often used. VC methods, also called as linear mixed models, are based on IBD sharing (Amos, 1994; Almasy & Blangero, 1998). The phenotypically similar individuals are assumed to share alleles near the quantitative trait locus (QTL) more commonly than by coincidence. Within the VC method, the phenotypic covariance is explained by components of genetic and environmental effects. The VC method has been implemented in SOLAR software package (Almasy & Blangero, 1998).

Most common linkage methods are based on a traditional statistical theory called the frequentist theory, which focuses on frequencies seen in random samples. Another option, currently becoming more popular, is Bayesian inference, which is based on observing known data and measuring the probability that a certain reason for some event is true. The relationship between conditional probabilities can be written as:

$$\Pr(Y|X) = \frac{\Pr(X|Y) * \Pr(Y)}{\Pr(X)}$$

which describes how probable (Pr) it is that X happened because of Y compared to other possible reasons (Bayes, 1763). In linkage analysis, we can calculate the probability that a certain measured genotype (Y) would cause the observed trait variation (X). In frequentist statistics, the p-value gives the probability of getting a certain result from observed data when it is not true. In linkage analysis, the results are usually given on a logarithmic scale for the ratio of the probabilities for linkage and against linkage, which is called the LOD (logarithm of odds) score. In the Bayesian approach, results are given as probabilities that the reason, for example a certain genetic locus, caused the event, for example some phenotype, instead of other explanations. Thus, probabilities range from 0 to 1 where 1 stands for proof.

The Bayesian linkage methods include, for example, the posterior probability of linkage (PPL) framework (Vieland, 1998). This framework has been implemented in the KELVIN program that is based on the E-S algorithm (Vieland et al., 2011). Compared to the frequentist analysis where parameters are fixed, the parameters in the Bayesian framework are estimated from prior probabilities and the data. In addition, a multiple testing problem that challenges frequentist-based gene mapping, affects Bayesian-based analyses less because they compare the studied reason (Y) to all possible reasons in the likelihood calculation instead of comparing an alternative to the null hypothesis (for more information, see Scott and Berger, 2006).

### 2.1.3.2 Association

The other fundamental class of gene mapping methods include association analyses. Genetic association analysis studies the correlation between phenotype status and genotypes. The frequencies of genotypes (or alleles) are compared between cases and controls in a population sample. Association analyses are usually computationally less demanding than linkage analyses. The main focus of the association analysis is on single marker analyses and short haplotypes whereas linkage analysis usually focus on genomic regions.

Genome-wide association studies (GWAS) have been successful in the past decade showing more than 15,000 SNP-trait associations (Welter et al., 2014). The methodology for GWAS is widely discussed and protocols have been published (see for example Zondervan and Cardon, 2007; Visscher et al., 2012). GWAS have been published for qualitative and quantitative traits in different population samples.

Association analysis on families can increase power compared to the more typical case-control setting, as shown by simulations (Laird & Lange, 2006; Hiekkalinna et



al., 2012). However, the methodology for family settings is less extensive. Fortunately, the use of family-based data has gained new interest in recent years; new methods have been published for rare variant and sequence data analysis (Lee et al., 2014). For nuclear families, methods include the transmission disequilibrium test (TDT) that compares transmitted alleles to expected distribution of alleles, and the haplotype-relative-risk (HRR) method that uses the non-transmitted alleles as controls and transmitted alleles as cases (Ewens & Spielman, 1995). The family-based association test (FBAT) is built on the TDT and has been extended to be used with different kinds of traits (Horvath et al., 2001; Laird & Lange, 2006). These three types of methods are usually applied with small sized families.

Concerning large and mixed sized families, there are only a few LD (linkage disequilibrium) methods that can utilise the families as a whole and do not break the pedigrees into nuclear components (Hiekkalinna et al., 2012). Such methods include likelihood-based methods like Pseudomarker that can be used for binary traits (Hiekkalinna et al., 2011), and MENDEL that has many different analysis options (Lange et al., 2013). A partially likelihood-based method is also the posterior probability of LD (PPLD) method implemented in KELVIN that is most suitable for quantitative traits (Vieland et al., 2011). As all these listed methods are computationally intensive, faster association methods have been developed that correct for the relatedness instead of using it as a starting point. These methods include, for example, the GRAMMAR algorithm (Aulchenko, de Koning, et al., 2007) that has been implemented in GenABEL (Aulchenko, Ripke, et al., 2007). The methods that use full family information are usually most powerful, although HRR (which uses only nuclear families) has proven to maintain equivalent power in simulations (Hiekkalinna et al., 2012).

### 2.1.3.3 *Selection signature analysis methods*

Additional to linkage and association methods, genetic loci affecting a certain phenotype can be identified through signatures of selection; beneficial mutations or variants favourable in a new environment will become objects of positive selection. Positive selection changes the frequency of the variant and shapes the LD pattern around the variant, which can lead to excess of extremely long haplotypes or high frequency alleles. Thus, the selection pressure will leave signatures in genome that can be detected (Sabeti et al., 2007). Even weak selection can become notable after accumulating over sufficiently long period of time (Storz, 2005).

Regions with reduced variability may be a signal of positive selection. Thus, homozygosity tests can be performed to compare cases and controls to test if some haplotypes have achieved fixation in cases but remained polymorphic in the controls. An example of a method that is based on the homozygosity test is the Cross-Population Extended Haplotype Homozygosity (XP-EHH) method (Sabeti et al., 2007). Recent positive selection can lead to exceptionally long haplotypes; if the frequency of an allele increases fast in the population, there will not be enough time for recombination to fraction the haplotype around the allele (Sabeti et al., 2007). HaploPS searches for these kind of exceptionally long haplotypes (Liu et al.,

2013). Contrary to these haplotype methods, genetic differentiation between two populations can also be used to identify regions under selective pressure (Storz, 2005). The Fixation Index, F-statistics ( $F_{ST}$ ) parameter can be used to measure the genetic differentiation between populations from allele frequencies (Weir & Hill, 2002). When it is applied to cases and controls, it can be used to identify regions with maximum differentiation (Weir & Hill, 2002; Storz, 2005; Vitti et al., 2013).

However, there are no selection signature methods available for families. The use of related individuals in the existing methods would result in biased LD structure and allele frequency estimations.

#### 2.1.3.4 Multiple testing

Traditional frequentist statistical tests pose the fundamental problem of multiple testing, especially prominent in association studies where the number of tests performed easily becomes very large. As there can be differences between any two groups just by chance, differences in allele frequencies between cases and controls can emerge just by chance. In a genome-wide study, millions of genetic loci are used which means millions of statistical tests. If the chance of getting a false positive association is 1% in one test, there will be on average 10,000 such spurious associations when performing one million independent tests. Thus, the p-value thresholds for declaring a positive test result need to be corrected by the number of tests.

Rule-of-thumb p-value thresholds have been proposed for both linkage and association analyses. In GWAS, a p-value threshold of  $5 \times 10^{-8}$  has been commonly accepted for genome-wide significance (Li et al., 2012). This threshold should result in a false discovery rate (FDR) less than 5% when the whole genome is analysed. In linkage analyses, a LOD score of 3 has been commonly used as the threshold for a significant linkage, meaning 1000 times the likelihood of linkage compared to the likelihood of non-linkage. However, there are also other features affecting the probability that a detected linkage is true. For example, the width of the linked region may relate to significance (Ott & Hoh, 2000). Additional information, for example functional data or similar results from other studies, may also suggest a genuine result even though the significance thresholds were not met.

The multiple testing problem can also be corrected by empirical methods. There, the p-value distribution under the assumption that the null hypothesis is true is estimated from the data by simulations. In these simulations, data with the same properties as the real data is used to estimate the empirical p-value threshold for the chosen FDR. Whereas threshold methods are applied regardless of sample size and other factors, these are taken into consideration in the empirical correction methods. However, the genome-wide mapping methods can be computationally extensive. The simulations increase computations and, especially with families, can lead to infeasible computation times, at least with current computer power.

Some methods have been proposed to ease the multiple testing problem by making fewer statistical tests. The multipoint linkage, haplotype-based association and



selection signature methods are examples of these kinds of methods. Another approach is to use a preselection stage to reduce the number of markers used in the final single marker analyses (Wason & Dudbridge, 2012). With less markers in the actual analysis, the two-stage analysis reduces the multiple testing problem and computational burden. The approach is not limited to any particular association method but the two stages need to be independent (Dai et al., 2012). The two-stage analyses are not common, but methodology has been suggested for other computationally intensive analyses like interaction analyses and for sequence data (Boonstra et al., 2016).

#### 2.1.3.5 *Biological significance*

Gene mapping methods are used to uncover the genetic regions affecting the trait studied. However, it is crucial to note that the methods do not identify the genes but only genetic regions or variants. The majority of the variants associated with complex traits have been identified outside coding regions, which complicates the prediction of the biological function (Hindorff et al., 2009). With additional analyses or information, we can estimate whether a specific adjacent gene affects the trait outcome. For example, functional studies (like transgenic animal models) can be used to show whether a mutation has an effect on a specific gene that could explain the trait (see for example Freedman et al., 2011). Another method is to perform an enrichment analysis to search for functions or pathways related to multiple associated locations or genes adjacent to the regions. As genes interact with each other, many diseases are caused by mutations in genes working within certain pathways. For example, signaling in neurons requires a cascade of genes to work and mutation in any of those genes may impair the function. Similarly, there might be associations found in multiple genes within a pathway, which indicate the biological importance of that pathway for the studied trait.

Biologically relevant results should replicate between different studies and settings. In other words, findings that have been made in multiple studies are more probably biologically relevant and not spurious. These replicated signals can be identified through meta-analyses and other methods like convergent functional genomics (CFG). CFG is an algorithm for ranking results from multiple evidence layers (Bertsch et al., 2005). In the simplest form, it is counting votes for the genes. Normally, layer weights are considered. The layers can include evidence from multiple sources like gene mapping, gene and protein expression, inhibition and biomarker levels. The algorithm has been implemented in the GenRank package for Rstudio (Kanduri & Järvelä, 2016). The approach can be used for merely human data to get maximum phenotype specificity or to also integrate animal model data to achieve better sensitivity for the molecular evidence. It aims to identify the most probable candidate genes for the studied phenotype.

## 2.2 Musical abilities

### 2.2.1 Biological origins of music

Music is a part of every human culture (Merriam, 1964; Savage et al., 2015). In rituals, music evokes strong emotions. Music can even be utilised in hospitals to help people recover faster (Särkämö et al., 2008). Music is known to affect us; in marketing, music is used to compel us to buy more (Alpert et al., 2005). Generally, music has many biological effects in our bodies, but not much is known about the specifics of these functions. Moreover, music does not only affect humans, but has measurable effect on some other animals as well (Chikahisa et al., 2006; Angelucci et al., 2007; Sanyal et al., 2013). The music-related capacities in animals are covered in chapter 2.2.7.

In many studies, music has been characterised as a product of general-purpose cognitive functions, but there is also evidence for music-specific neuronal pathways (Peretz & Zatorre, 2005). For example, speech and music can be differently impaired in brain injuries (Peretz & Zatorre, 2005; Peretz, 2006). This was also illustrated in the case of Finnish singer Kaisa Makkonen who lost the ability to speak after a brain infarction but was able to continue to sing (Nousiainen & Ranta, 2015). Otherwise, memory, learning and many other common cognitive functions are required in many musical tasks. Supposedly, most musical abilities are based on both music-specific and general-purpose cognitive functions.

There is genetic predisposition for music cognition. Even at birth we can identify certain musical abstracts like pitch, tempo and consonance (Perani et al., 2010). Most people have the capacity for music perception and production, but the degree of competency in music varies between individuals. In this thesis, I have studied differences between individual's abilities to identify pitch, note length and sound patterns, which can be considered as basic components of musicality (see for example Honing et al., 2015). Additionally, I have studied realisation of musical abilities, namely music practice, musicianship, composing, arranging and improvising. Musical ability itself comprises diverse phenotypes that includes both acquired and innate skills.

### 2.2.2 Evolutionary origins of music

It has been argued whether music is an adaptation as such or makes use of other existing skills (Honing et al., 2015). Probably, it has been a mixture of them both, some aspects of music are adaptations based on pre-existing capacities (Trainor, 2015). For example, auditory processing has supposedly evolved for other uses than music. That ability may have evolved into the pitch processing needed in music (Trainor, 2015). Similarly, a sense of rhythm is important in walking which may explain why it has evolved. However, it has been suggested that music is important for group cohesion and social bonds, which may have evoked selection pressure for music-related abilities even though the abilities would originally have been based on other, existing skills (Fukui & Toyoshima, 2014; Honing et al., 2015; Trainor, 2015).

Consequently, music and musicality may be, at least partially, cultural creations and may have evolved further from pre-existing capacities. But as Honing et al. (2015) stated, “*Despite the apparent impossibility of studying the evolution of complex mental processes such as cognition, we argue that a bottom-up approach involving the search for basic mechanisms that combine into a multicomponent trait like musicality can be fruitful. Such an approach has resulted in important insights in the domains of animal cognition and the evolution of language.*” Thus, the study of the genetics of music-related traits may unveil evolutionary information despite the argument on the evolutionary grounds of music.

### 2.2.3 Music-related phenotypes and measurements

Musical abilities comprise a wide variety of skills (see for example Levitin, 2012). For example, abilities to identify pitch, rhythm, tempo and consonance are prerequisites to understand music. These capacities establish the ground to understand more complex music properties like patterns, and to memorise and learn music. Music practice and producing also features other skills depending on the medium, such as hand and body coordination, singing, keeping rhythm, creativity, adaptability and aesthetic sense.

Pitch processing is an example of an ability that can readily be measured. Absolute pitch (AP) is one of the capacities relating to pitch processing. AP means the ability to identify musical notes without any reference tone. Similarly, amusia, also known as tone-deafness, is a musical disorder that appears as an inability to detect pitch. Amusics have impaired perception of music with no deficit in hearing. It has been estimated that 4 per cent of the population are amusics (Fry, 1948; Peretz et al., 2007) even though there has been debate about the prevalence (Henry & McAuley, 2010).

Karma’s auditory theory divides musical abilities into primary capacities and secondary music skills that are built upon them (Karma, 1994). According to the theory, the primary capacities reflect innate musical aptitude and the secondary skills are affected by environment. The primary capacities mostly include perceptual abilities. However, it is not definite which abilities belong to the primary or secondary capacities. Some skills, like musical creativity and understanding music, are clearly defined as secondary skills. Otherwise, musical memory can be considered as a partly primary and partly secondary skill; elementarily, musical memory includes any ability to deposit a heard auditory signal and to be able to compare it to following ones. As a more complex skill, it can be considered as representation and retention of complex music pieces (Burunat et al., 2014).

Many musical abilities can be measured with tests (Table 1). Standardised tests were developed, especially in the 20<sup>th</sup> century, but newer tests have recently been introduced to be used through the internet. The older tests include tests from Carl Seashore (Seashore et al., 1960), Herbert Wing (Wing, 1941) and Edwin Gordon (Gordon, 1979). These are test batteries covering many aspects of musicality but focus on perceptual skills. Seashore’s test battery covers tests for basic capacities including, for example, rhythm, duration, pitch and tonal memory. Wing’s test

battery also includes tests for more complex music properties like aesthetic understanding. Gordon's music test is a more complex test that focuses on giving a full picture of the musical abilities of the individual. It is based on Gordon's theory on the sense of hearing, named audiation (Gordon, 2007). Newer tests include KMT, which is a pattern recognition test (Karma, 2007) that has also been modified to be used without hearing (Karma, 1994). Some tests include music playing or singing like the Pitch Production Accuracy (PPA) test, a singing accuracy test that measures the ability to reproduce pitch (Park et al., 2012).

Amusia has been tested with the Distorted Tunes Test (DTT) and Montreal Battery of Evaluation of Amusia (MBEA) (Henry & McAuley, 2010). DTT includes popular melodies where individuals are required to recognise distorted tunes (Drayna et al., 2001). DTT illustrates pitch recognition ability where individuals identifying less than 18 items from the 26 are considered amusics. MBEA is a similar test but the melodies are new.

AP testing relies on identifying notes and, thus, only individuals who have received music training can be tested. One example of an AP test is the test by Baharloo et al. (2000). Some studies have used tests with no requirement for knowing note names, for example PitchMatch! (Ross et al., 2004; Gregersen et al., 2013). Due to differences in measuring, the prevalence for AP varies between studies. The prevalence for the strict definition has been estimated as 0.1% (Deutsch et al., 2009). Among music students, the prevalence of AP in individuals with European ancestry is estimated at 7%, whereas it was even five times higher in individuals speaking tone languages (Gregersen et al., 1999; Deutsch et al., 2009). This illustrates the effect of early environmental exposure that is important in music.

As with any tests, there are limitations in musical aptitude testing, even though the reliability of most of the tests has been reported to be relatively high (Shuter-Dyson & Gabriel, 1981). Many tests correlate with music practice. At least some of the tests

**Table 1. Music tests used in genetics studies.**

Music test	Capacities tested	Abbreviation	Reference
Advanced Measures of Music Audiation	Rhythm, pitch	AMMA	Gordon, 2007
Distorted tunes test	Amusia, pitch recognition	DTT	Drayna et al., 2001
Karma music test	Pattern recognition test	KMT	Karma, 2007
Montreal Battery of Evaluation of Amusia	Amusia	MBEA	Henry & McAuley, 2010
Pitch production accuracy test	Pitch, singing	PPA	Park et al., 2012
Swedish Musical Discrimination Test	Melody, rhythm and pitch	SMDT	Ullen et al., 2014
Seashore's test	Pitch, rhythm, duration, timbre, loudness and tonal memory	SP, ST	Seashore et al., 1960
Wing's test battery	Chord, pitch, memory, rhythm, harmony, intensity and phrasing	-	Wing, 1941

are not sensitive to practice, as was shown in a twin study using the Swedish Musical Discrimination Test (SMDT) (Mosing, Madison, et al., 2014). The music test results also had a positive correlation with intelligence (Mosing, Pedersen, et al., 2014). Additionally, many of the tests rely on concepts common in Western music, or are otherwise culture-bound like DTT. This has led some researchers to think that musicality is an overall culture-specific skill that cannot be measured. However, musical abilities comprise skills that have shown to be reliably measured. Moreover, the culture-dependency can be studied by comparing different populations.

One especially complex feature of musical abilities are emotions awakened by music, for example, the ability to recognize emotions conveyed by music. These have been studied based on experiences of “musical chills”, phenomenon where one can experience shivers and goose bumps during music listening (Nusbaum et al., 2014). Emotions have also been tested through hormone levels and heart rate (VanderArk & Ely, 1993). However, the measurability of emotions in the general population may not be reliable. For example, not everyone experiences musical chills despite the emotions music awakens in them (Nusbaum et al., 2014). Similarly, the reaction in hormone levels between music students and other students do vary; the effect on cortisol level was opposite between the groups (VanderArk & Ely, 1993).

Another complex feature related to musical ability is musical creativity. Improvising, composing and arranging are important in music. Improvising skills have been included in some test batteries like those of Gordon (1979), Wang (1985) and Webster (1983). In these tests, the participants produce sounds in various ways according to instructions. The produced sounds are rated to assess the measurement for improvising. However, this approach is difficult to implement in population studies. Thus, questionnaires have been used to study improvising, composing and arranging (see for example Sovansky et al., 2016).

Musical abilities include various skills that researchers have tried to measure for decades. There are still debates on what and how to measure. The question relates closely to what is music and what is not. These questions will most likely also remain debatable. From the biological point of view, music is specific auditory stimuli that is detected by the ear, auditory pathway and many regions in the brain. Genomics and bioinformatics methods can be used to explore the biology and evolution of music and sounds at the molecular level without any prior knowledge of the biological background of the trait (Lander, 2011). The biological view on the subject has been more practical; biological studies have concentrated on test reliability, applicability and direct measures of music production, as has also been done in this thesis.

#### 2.2.4 Heritability of musical abilities

It is generally agreed that music abilities are affected by both genetic and environmental factors. However, it has long been argued whether world-famous musicians and musical prodigies are born with the competencies or if their music perfection is produced by fruitful environment (see for example Howe et al., 1998,

and the related discussion). Many historical musicians, like Mozart, have been used as examples of innate competencies but these examples often lack information on all family members and childhood education (Shuter-Dyson & Gabriel, 1981; Levitin, 2006). For example, the environment and opportunities towards music in these families may have substantially differed from the normal population. Thus, twin and family studies have been used to estimate the proportion of genetic and environmental factors on many musical abilities (for example Stanton, 1922; Friend, 1939; Drayna et al., 2001).

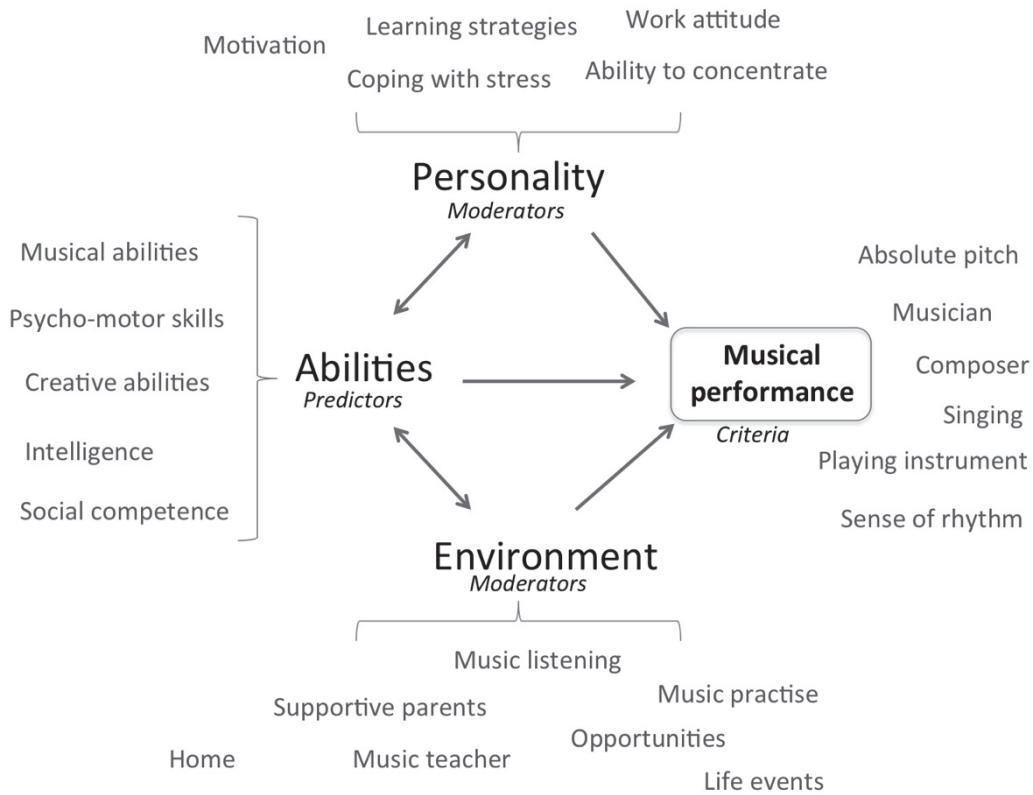
Estimates of heritability vary considerably between the phenotypes. Pitch recognition with British twins showed a heritability between 71% and 80% (Drayna et al., 2001). The phenotype definition was based on the DTT whose extreme values are used to define amusics. In another study, the PPA test showed heritability of 40% in Mongolian families (Park et al., 2012). The heritability of musical aptitude using three music tests in the Finnish population showed heritabilities from 21% to 57% (Pulli et al., 2008).

Sibling recurrence risk for AP has been estimated as 8-15 when controlling for early music training (Baharloo et al., 2000). Similarly, sibling recurrence risk for amusia in the United Kingdom was estimated as 10.8 with a prevalence of 4% (Peretz et al., 2007). Both of these two recurrence risk results show a strong role for genetic influence. A Swedish study using the SMDT showed heritability of 59% for melody, 50% for rhythm and 40% for pitch (Ullen et al., 2014). However, pitch recognition showed significant gender differences not included in the estimate. Separate heritabilities for the pitch test showed 30% for females and insignificant heritability for males.

It has been widely discussed whether practice of music affects musical aptitude or not (see for example Schellenberg, 2015). There is evidence for and against (Mosing, Madison, et al., 2014; Strait & Kraus, 2014; Saarikivi et al., 2016). As there are countless numbers of possible environmental factors affecting musical aptitude and related abilities, the interaction between the environment and genes is utterly complicated to study. The interaction is difficult to break down into separate factors that could be measured. The Munich Model of Giftedness presents the performance of, for example, music as an outcome from an interplay between abilities, personality and environment (Figure 2) (Heller et al., 2005). The model depicts personality as a moderator that makes it easier or harder to practice which affects the musical outcome. The specific impact of practice has been studied using twins to separate practice from other factors (Mosing, Madison, et al., 2014). In that study, it was shown that at least the perceptual abilities in music seem to be quite constant regardless of practice.

However, there are abilities like AP where childhood application of the ability, such as music practice, is of importance in acquiring or maintaining the skill (Wilson et al., 2012; Gervain et al., 2013). A similar sensitivity period has been reported in language development (White et al., 2013). It has also been suggested that practice during childhood may alter musical ability more than practice after this sensitivity





**Figure 2. Musical performance as an interplay between musical abilities, personality and environment.** The division into these four categories is based on the Munich Model of Giftedness (Heller et al., 2005). Music-related factors that are part of the interplay as well as some performance features are given. Similar models have been proposed, especially in expertise studies in music (see for example Ullen et al., 2015).

period (White et al., 2013; Baer et al., 2015). However, there is a lack of tests that would be reliable for pre-schoolers. Thus, it is difficult to study the effect of practice during the sensitivity period. Overall, early music exposure as an environmental factor suggests an interaction between genetic predisposition and environmental exposure.

In conclusion, a large part of the discussion as to whether music is an innate ability or a result from practice comes from the differences in definition of musical abilities, whether we discuss music performance or musical abilities. Abilities are less prone to environmental factors (like music practice) whereas music performance, like musicianship, depends greatly on environment and personality (Figure 2).

### 2.2.5 Physiological effects of music

Music listening and producing affect the body. Music listening has physiological effects on heart rate, blood pressure, respiration and mood (Mockel et al., 1994; Blood & Zatorre, 2001; Schulreich et al., 2014). Listening to pleasant music can induce dopamine release in brain regions of importance for reward (Salimpoor et al., 2011). Listening to music promotes recovery and helps patients with anxiety or

depression (Särkämö et al., 2008; van der Heijden et al., 2015). These effects have been utilised in ancient cultures by using music for the treatment of physical and neuropsychiatric conditions (Merriam, 1964). Nowadays, music has been successfully used in music therapy, both as receptive, especially in recovery, and as active music therapy for many neuropsychiatric disorders like dementia, depression and epilepsy (Särkämö et al., 2008; Dastgheib et al., 2014). The positive effects in recovery after surgery have been promoted in many recent meta-analyses (van der Heijden et al., 2015; Vetter et al., 2015). However, the biological mechanisms underlying these effects remain largely uncharacterised.

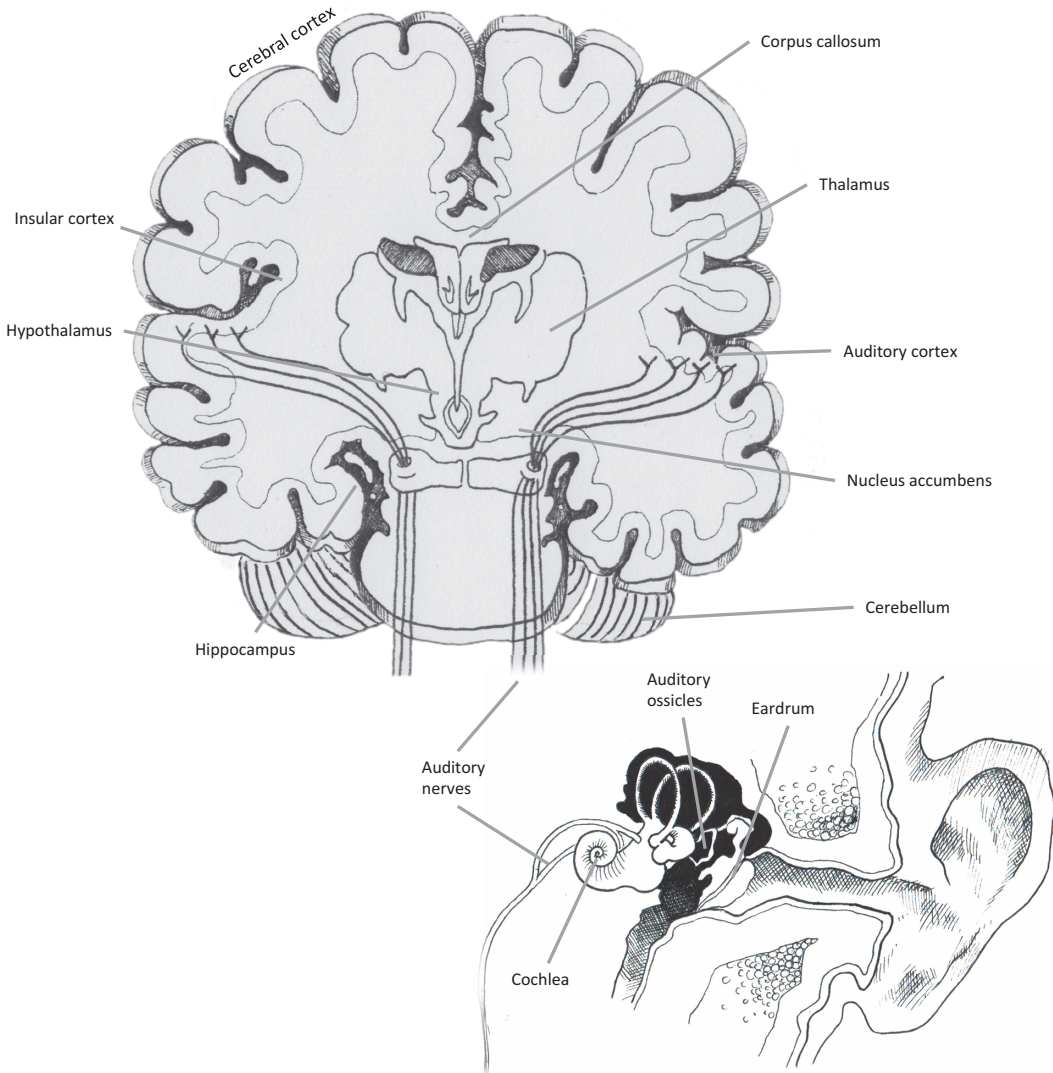
Music listening can reduce cortisol levels (Kreutz et al., 2004; Fukui & Toyoshima, 2013; Fancourt & Williamon, 2016). Some other hormones and neurotransmitters such as dopamine, noradrenalin and testosterone can also be affected by music listening (VanderArk & Ely, 1993; Fukui & Yamashita, 2003; Salimpoor et al., 2011). The direction of the effect on the hormone levels is dependent on the type of music and the listeners, for example, differences between experienced listeners and other individuals exist (VanderArk & Ely, 1993; McCraty et al., 1996; Gerra et al., 1998). Thus, the hormone level changes relate to emotional state, for example, the subjective feeling of pleasure and reward, induced by music (Chanda & Levitin, 2013).

Several brain regions have been shown to be involved in music. The auditory cortex is involved in the processing of any kind of auditory signal, including music (Skouras et al., 2014). However, musical abilities like pitch processing also include more complex cognitive processes outside of auditory cortex (Rogenmoser et al., 2015). Even passive music listening activates cerebral cortex regions including the anterior insula in the insular cortex, the limbic lobe including the hippocampus and prefrontal anterior cingulate, basal ganglia regions including the nucleus accumbens, and hypothalamus (Figure 3) (Brown et al., 2004; Koelsch & Skouras, 2014). Many of these regions have functions in emotions, for example, through the dopaminergic system. In musical ability, the cerebellum has been shown to be important in the working memory for rhythm and timing performance (Jerde et al., 2011; Baer et al., 2015). Improvisation and composition of music have been linked to the medial prefrontal cortex, premotor areas and the auditory cortex (for a review, see Dietrich and Kanso, 2010).

Music practice has also been found to affect brains (Koelsch, 2011). As mentioned above, music practice may enhance brain plasticity, for example, instrumental music practice has been connected to enlarged motor and auditory regions (Hyde et al., 2009b). Musicians and non-musicians have differential reactions for music in the brain (Koelsch, 2011; Burunat et al., 2015). Musicians have also been shown to have enlarged corpus callosum and increased functional symmetry between left and right hemispheres (Burunat et al., 2015). However, many studies lack controlling for the selection of individuals who start music education (Schellenberg, 2015). Thus, some differences between the musician and non-musician brains may have existed prior to music practice.



Recently, the effect of music listening on human gene expression profiles has been studied where especially the *SNCA* gene was found to be upregulated (Kanduri, Raijas, et al., 2015). The activity of genes involved in dopamine secretion and transport, synaptic function, learning and memory were enhanced. Similarly, a music practice expression study involving professional musicians showed differential expression of genes related to dopaminergic neurotransmission, motor function, neuronal plasticity, learning and memory (Kanduri, Kuusi, et al., 2015). For example, *FOS*, *DUSP1* and *SNCA* were differentially expressed. Another study showed increased *DRD4* expression among musicians and autistic individuals in lymphocytes (Emanuele et al., 2010).



**Figure 3. Brain and inner ear structures that are important in music perception.**

## 2.2.6 Genetics of musical abilities

There has been increasing interest in the genetics of musical abilities. When knowledge about the physical and mental effects of music increased, studies about the molecular mechanisms of music have become more acceptable. Similarly, increasing numbers of studies about specific cognitive abilities like educational attainment and mathematical aptitude have been published lately (Davis et al., 2015; Okbay et al., 2016).

The first studies about the genetics of musical abilities were twin studies where it was shown that musical abilities are partially inherited (Shuter, 1966; Coon & Carey, 1989; Drayna et al., 2001). Following DNA studies were candidate gene studies that particularly focused on *AVPR1A* and *SLC6A4* with various music-related phenotypes (Bachner-Melman et al., 2005; Granot et al., 2007; Ukkola et al., 2009; Ukkola-Vuoti et al., 2011; Morley et al., 2012; Fukui & Toyoshima, 2013). The *AVPR1A* gene is a receptor for arginine vasopressin, a hormone which influences, for example, social interaction and mood (Granot et al., 2013). *AVPR1A* may especially affect musical memory (Granot et al., 2013). The *SLC6A4* gene encodes a serotonin transporter in neurons. It has been related to choir participation (Morley et al., 2012) and dancing (Bachner-Melman et al., 2005). Another hormone receptor, androgen receptor (AR) has been associated with the AMMA music test results (Fukui & Toyoshima, 2013).

A few genome-wide studies of musical abilities have been published. The first genome-wide linkage study of musical aptitude was published from our pilot material in 2008 (Pulli et al., 2008). There, significant evidence for linkage was found at 4q22 and suggestive evidence at 8q13-21 for three music tests: Seashore's tests for pitch and time, and Karma's test for auditory structuring. In a study by Park et al. (2012), a region at 4q23 was linked for the PPA test in the Mongolian population. Within the linkage region, variants affecting the *UGT8* gene were associated. An AP study showed linkage at chromosome 8q24.21 in the European population and found suggestive evidence for 7q22.3 (replicated in East Asian samples), 8q21.11 and 9p21.3 (Theusch et al., 2009). Another study focusing on AP and synaesthesia found evidence at 6q14.1-6q16.1 and 2q24.1 for the combined phenotype and at chromosome 2q22.1 for AP (Gregersen et al., 2013). Contrary to previous findings, the hormone receptor genes that were suggested in the candidate gene studies were not replicated in these genome-wide studies.

## 2.2.7 Music perception and production in animals

As music is mostly considered as an art form and culture-dependent factor, animals are not considered to produce music in the same sense as humans do. However, there are animals that sing, like birds and whales (Honing et al., 2015). Also, many animals from goldfishes to primates and birds are capable of listening to and some are even able to discriminate music pieces (Watanabe & Sato, 1999; Chase, 2001; Otsuka et al., 2009; Shinozuka et al., 2013). Basic capacities for pitch and rhythm are available in many animals even though they are hard to measure (Weisman et al.,

2010; Patel, 2014). Thus, many music-related operations exist in most animals even though they do not produce music per se.

There are few animal models that have already been used to study music-related phenotypes. For example, songbirds like zebra finches (*Taeniopygia guttata*) have been widely used to study auditory perception skills and vocal learning (White, 2010). However, it is uncertain whether these phenotypes relate more to speech than music (Scharff & White, 2004; Earp & Maney, 2012; Simonyan et al., 2012; Patel, 2014; Rothenberg et al., 2014; Honing et al., 2015; Rohrmeier et al., 2015). Some music-related features, like repertoire size and motif duration, have shown very low heritability in a captive population of 808 cross-fostered zebra finches (Forstmeier et al., 2009). There has also been criticism that these kinds of measures, like repertoire size, are not relevant to song in the social context (Rothenberg et al., 2014). On the other hand, some features like the sensitivity period for AP and possibly music practice in humans (Zatorre, 2003; Bailey & Penhune, 2013; Gervain et al., 2013), vocal learning in mice (Yang et al., 2012) and song learning in zebra finches, seems to be similar between the species (Clayton, 2013). There is a large variability between bird species as to what extent they learn new songs after the sensitivity period; zebra finches only learn new songs during their sensitivity period whereas canaries learn songs throughout their lives (Bolhuis et al., 2010). In addition, vocal learning has been shown to share similar features between humans and songbirds (Jarvis, 2004). It is difficult to identify the heritable phenotypes in animal models that are parallel to musical abilities in humans. These parallel phenotypes are needed to be able to use animals to study the genetics of musical abilities.

Studies on animals have already indicated that the hypothesis regarding music listening enhancing brain plasticity seems to hold in neuronal level (Polley et al., 2006; Shetake et al., 2012). Overall, brain effects due to music that cannot be studied in humans have been studied in these animal models. For example, it has been shown that music exposure alters neurotrophin production in the hypothalamus in mice (Angelucci et al., 2007). Similarly, gene expression changes in the brain during the sensitivity period have been studied (see for example Olson et al., 2015).

The zebra finch brain studies have concentrated on song control regions of the brain like area X at the basal ganglia, high vocal center (HVC) and the robust nucleus of the arcopallium (RA). Short- and long-term electrophysiological and molecular responses for singing and song listening in these regions have been published (Wada et al., 2006; Tremere et al., 2009; Hilliard et al., 2012; Velho et al., 2012; Yoder et al., 2015). The singing and song listening evoked gene expression changes like that of *EGR1* have also been proven in other songbirds like canaries (Mello et al., 1992), black-capped chickadees (Avey et al., 2008) and budgerigars (Eda-Fujiwara et al., 2003). *EGR1* has been shown to express as a response to song in the bird's caudomedial nidopallium (NCM) which is similar to the human auditory cortex.

Naturally, there are differences between humans and other animals. Despite the differences, many molecular and structural parallels have been shown in auditory perception and hearing. Many auditory related brain structures were shown to

express genes similarly even between songbirds and humans (Pfenning et al., 2014). Similar structures have been identified even between katydid and mammal inner ears (Montealegre-Z et al., 2012). Moreover, convergent sequence evolution was also seen in hearing related genes between echo-locating bats and dolphins (Parker et al., 2013). Genes under differential expression in music producing and listening also included genes affecting song learning and singing in songbirds (Kanduri, Kuusi, et al., 2015; Kanduri, Raijas, et al., 2015). These data indicate evolutionary conservation of molecular mechanisms in auditory perception.

### 3 Aims of the study

The aims of the study were

1. To identify genetic loci for musical aptitude in Finnish families
2. To identify genetic loci for creative activities in music
3. To identify candidate regions showing positive selection with musical aptitude in the human genome
4. To integrate published data about musical abilities in order to recognize plausible candidate genes and pathways from the identified regions

## 4 Materials and methods

### 4.1 Study materials (I, II, III)

#### 4.1.1 Sample

The sample included 915 individuals from the Finnish population. The participants were recruited via advertisements in magazines, mailing lists, webpages and by contacting the relatives of the participants. In the early stage of the study, families with several music professionals were specifically sought for. Subsequently, any families were invited. Moreover, all family members regardless of their music education level were invited to the study in all phases. Of the final sample, 79 individuals were unrelated and 836 participants were from 99 families. The family sizes ranged from 2 to 50, consisting of participants from 1 to 4 generations. In publications I, II and III, samples consisted of families that were available at the time of the analysis and included phenotyped individuals (Table 2).

#### 4.1.2 Genotypes

Each participant older than 12 years was asked to provide a blood sample for the DNA study. An informed consent was obtained from all participants or their parents. The Ethical Committee of Helsinki University Central Hospital approved the studies. DNA was extracted from blood using the phenol-chloroform method and was obtained from 799 participants. The samples were genotyped with Illumina Human OmniExpress 12 1.0V SNP chip (Illumina Inc., San Diego, CA, USA) at the Wellcome Trust Center for Human Genetics, Oxford University. GenomeStudio was used for genotype calling and genotyping quality control. Twelve samples were reanalysed as they failed in the first run due to chip error. Two samples failed in the genotyping for other reasons. The remainder had call rates above 99% (minimum 99.18%, mean 99.74%). The OmniExpress chip is reliable for allele frequencies above 5%.

#### 4.1.3 Phenotypes

Phenotypes were assessed through musical aptitude tests and questionnaires. The musical aptitude scores were based on the KMT (Karma, 2007) and two tests from Seashore (Seashore et al., 1960). The tests were chosen to show different aspects of

**Table 2. Number of participants in publications I, II and III.** The total material consists of 915 participants.

	Sample size	Music test scores	Questionnaire answers	Valid genotypes	Families	Unrelated individuals
Total	915	867	805	780	99	79
Publication I	767	699	<i>not used</i>	632	76	0
Publication II	577	<i>not used</i>	568	572	79	103
Publication III	148*	148	<i>not used</i>	148	0	148

\* Chosen among 283 unrelated individuals

perceptual abilities in music. The KMT measures auditory structuring abilities that rely on recognition of melodic contour, grouping, relational pitch processing and gestalt principles (how music is perceived mentally). Participants have to detect repeated sound patterns in musical items and compare the pattern with a test sequence. These patterns are played with varying instruments. The test consists of 40 items and in each item participants choose whether the pattern matches the test item or not. Examples are available from [www.hi.helsinki.fi/music/english/samples.htm](http://www.hi.helsinki.fi/music/english/samples.htm).

From Carl Seashore's test panel, we used two tests in our study: subtest for pitch (SP) and time discrimination (ST). These tests measure the ability to compare pitch (in SP) or duration (in ST) of two sequentially presented tones. Both of these tests consists of 50 pairs of tunes.

The tests were performed in groups through loud speakers. The testing lasted approximately one hour. The objective of each test was first explained and a few example items were given before the start of the actual test.

In addition to the separate tests, we created a combined test score by assigning each test to a range of 25-50 and summing them together (COMB scores). The three separate music tests and COMB scores were used in musical aptitude gene mapping (I).

For selection signature analyses (III), the COMB scores were used to partition the subjects into cases and controls. Because there are no selection signature methods suitable for family material, only unrelated individuals were used. From 283 unrelated subjects, the cases and controls were chosen with linear regression. Linear regression was fitted between the COMB scores and age. The individuals with residual less than -10 were assigned to controls (N=74) and those with residual higher than 11 as cases (N=74).

A questionnaire was designed to explore the musical background of the individuals and what kind of music-related activities they had undertaken (Peltonen, 2013). One part of the questionnaire investigated creative activities in music. These included composing, arranging and improvising; the phenotypes used in gene mapping of creative activities (II). Each term was explained to the participants prior to answering. The participants were asked if they did or had done these activities. For comparison, we also asked if the participants considered themselves as creative in any non-musical field (visual, scientific, technical, physical or verbal). Additional questions about purpose, frequency and the way of performing the activity were used to ensure the reported activities (Peltonen, 2013).

The composing, arranging and improvising phenotypes were further processed to improve their reliability. For example, creative inactivity in music was restricted only to the individuals who had at least two years of music education or practice. This restriction was made to exclude individuals who simply might have lacked opportunity to perform these tasks or who did not have a medium to share or present their music. To separately analyse the individuals who were inactive considering musical creativity despite being musically educated, we constructed



another variable: neither composed nor arranged (NCNA). This phenomenon is well known to exist in musicians (Sovansky et al., 2016). Almost half of the musically educated individuals neither composed nor arranged.

## 4.2 Methods (I, II, III)

### 4.2.1 Heritabilities and correlations

The heritabilities were estimated with Solar (Almasy & Blangero, 1998) for quantitative traits and with DMU, a package for analysing multivariate mixed models (Madsen & Jensen, 2013), for qualitative traits. In DMU, the logit link model was used to connect the binary information into underlying quantitative trait. The phenotypic correlations were estimated with SPSS, version 22 (IBM), as non-parametric correlations. Regression models were used for music scores to correct for the effects of age and/or gender.

### 4.2.2 Gene mapping (I, II)

#### 4.2.2.1 *Quality control*

Quality control for the genotype data was performed with PLINK 1.07 (Purcell et al., 2007). The relatedness was checked through IBD calculation and compared with reported relationships. Population structure was further checked with principal components analysis using unrelated individuals with EIGENSOFT (Patterson et al., 2006). Three subjects were removed due to incorrect relatedness. Mendelian error rates were estimated with PedCheck (O'Connell & Weeks, 1998) as <0.1% per individual and incorrect genotypes were marked as missing. Markers were checked and removed for missingness (>5%), Mendelian inconsistencies, low minor allele frequencies (<5%) and Hardy-Weinberg equilibrium ( $p$ -value <0.001). Uninformative family members were excluded in both of the studies to ease the computational burden.

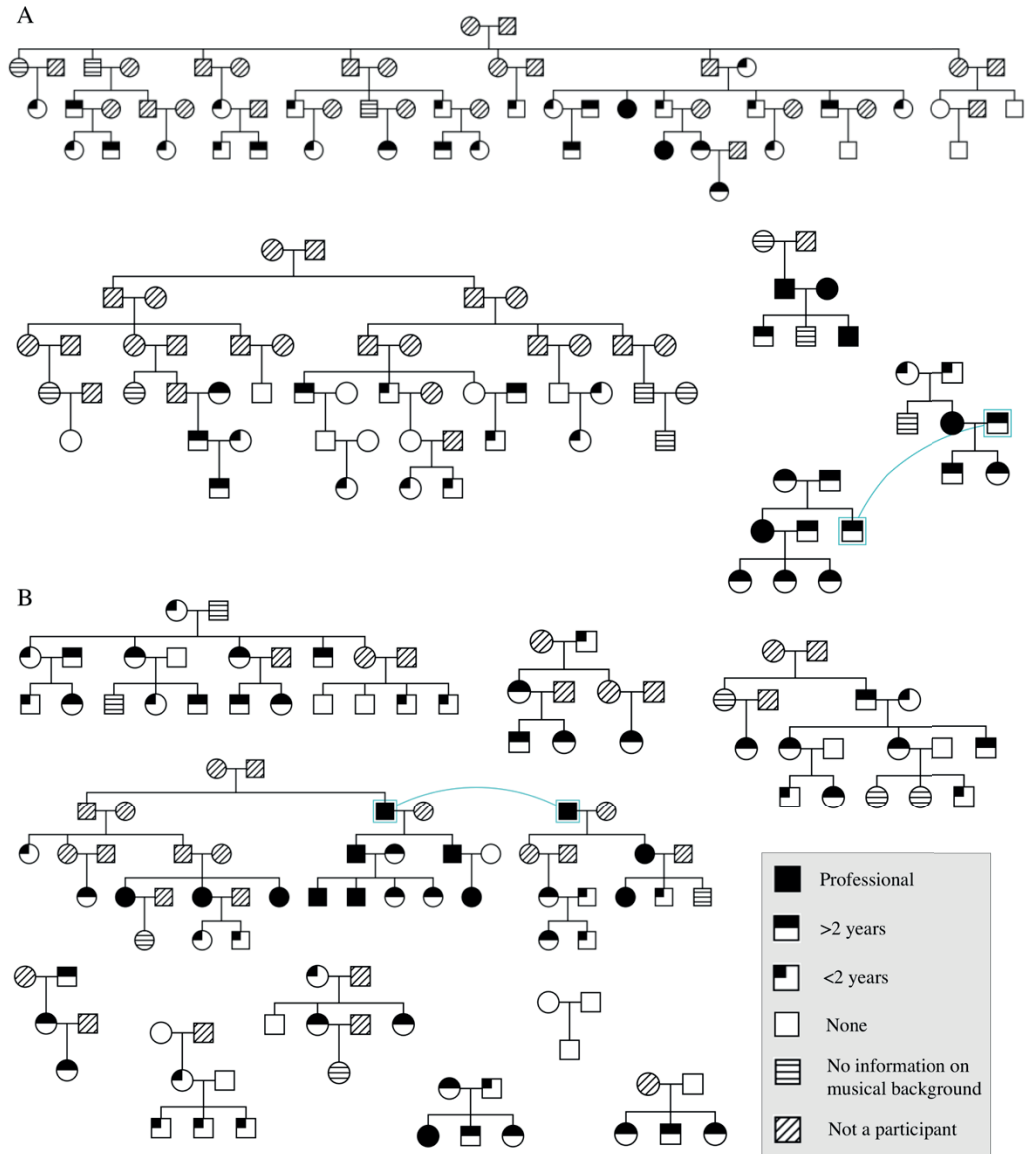
For linkage analyses, the marker map was thinned to ensure the independence of the signals between nearby markers. Minor allele frequency above 25% was required to maximise the information available for the selected markers. The SNPs in high LD with other SNPs were removed with the variant inflation factor method in PLINK using a variant inflation factor limit of 1.25 that corresponds to an LD of 0.2 as measured by  $R^2$ . Additionally, the remaining SNPs were pruned for a map distance of 0.2cM (Rutgers map v.217).

#### 4.2.2.2 *Linkage and association analysis (I)*

Our study was complicated by variably-sized families with plenty of missing data, as not all family members participated in the study (Figure 4). For example, information about the founders was missing in most of the families. The linkage and association methods had to, accordingly, be carefully selected to suit this type of data. We used the Bayesian method for quantitative linkage and association which is implemented in package KELVIN (Vieland et al., 2011). KELVIN can handle



extended families and large numbers of markers simultaneously. Thus, it was suitable for our data. It performs PPL and PPLD analyses. The data was analysed as two batches to take the ascertainment differences between the first families and the subsequently collected samples into account; the first batch that was collected based on having professional musicians included 14 families, the rest of the data were collected with no prior information about musical experience and included 62 families (Figure 4).



**Figure 4. Examples of the pedigrees in the data.** The first collected families (A) were collected based on having professional musicians whereas the rest of the families were collected with no prior information on musical experience (B). Musical experience was evaluated from the questionnaire answers and classified into four classes: no experience, less than two years of experience, more than two years of experience and professional musician or other music professional. The families included plenty of missing data as can be seen from the striped nodes.

Gene mapping of musical aptitude was performed for three phenotypes: SP, KMT and COMB scores. The PPL analyses included 10,742 SNPs and the PPLD analyses 664,177 SNPs. Both analyses included 767 participants.

The PPLD results were further pruned to exclude signals that were based only on one SNP because they are most probable false positive signals. Only clusters that had at least two SNPs near each other ( $R^2 \geq 0.5$ ) with  $PPLD \geq 0.2$  were considered as true positive signals.

Haplotype analysis was used to further study the region near the best-associated markers. Haplotyping and haplotype association analysis was carried out with PLINK 1.07 (Purcell et al., 2007). The haplotype analysis in PLINK only handles nuclear family structures and, thus, splits the used large families, which may inflate the results.

#### 4.2.2.3 *Linkage and linkage disequilibrium analysis (II)*

Linkage analyses for the binary traits of creative activities in music were performed with MERLIN (Abecasis et al., 2002) as non-parametric multipoint analyses (Kong & Cox, 1997). Linkage analysis was performed for arranging, composing, NCNA and non-music-related creativity. Contrary to the quantitative method used in the musical aptitude linkage analyses (I), this method seeks for co-segregation of the affected individuals and the genomic region. Both ends of the variation, being creative and not being creative, were considered interesting. Therefore, the NCNA was analysed in addition to composing and arranging.

Uninformative families and family members as well as unrelated individuals were removed prior to the linkage analyses. The largest families were split, because MERLIN cannot analyse very complex families. The linkage analyses included 10,826 SNPs and 467 individuals from 76 families.

The association analyses were performed as two-stage analyses (see chapter 2.1.3.4). In the first stage, we used PLINK 1.07 with no correction for relatedness and a loose p-value threshold of 0.01. As such, the test should be independent of the next stage where family information and linkage was incorporated. In some studies, even looser p-values, especially 0.05, have been used. However, we chose to use 0.01 to improve the power (Murcray et al., 2009). We also incorporated all SNPs within the suggestively linked regions. At the second stage, Pseudomarker (Gertz et al., 2014) was used to perform full single marker analysis for the mixed data. This analysis was performed as joint linkage and LD analysis. From the 590,979 markers in the first stage, 14,135, 13,069, 11,875 and 10,749 markers were included in the second stage for arranging, composing, NCNA and other creativity, respectively. For this analysis, family data and case-control data were combined. The combined data included 572 genotyped individuals, 105 individuals more than in the linkage analysis.

#### 4.2.2.4 *Enrichment analyses*

The linked and associated regions were further studied with enrichment analyses. Enrichment of functional classes and pathways among genes near suggestively

associated or linked markers were analysed with Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, [www.ingenuity.com](http://www.ingenuity.com)). The analysis compares the number of suggested genes belonging to the pathway (or class) to the total number of genes in the pathway (or class).

In musical aptitude gene mapping (I), all genes within 2cM regions (or at least 2Mb) around linkage results above 0.2 PPL were included, which resulted in 286 genes. The functional classes were restricted to those that included at least five suggested genes. For the PPLD results, there were only few regions where PPLD was above 0.2. Thus, not enough genes were included in the PPLD regions to perform informative enrichment analysis.

In creative activity gene mapping (II), the enrichment analysis was carried out for the regions with joint linkage and LD results with p-value below  $1 \times 10^{-4}$ . All genes 1Mb upstream and downstream were included. All four phenotypes were analysed separately. These thresholds resulted in 668, 672, 583 and 637 genes in the analyses of arranging, composing, NCNA and other creativity, respectively. The functional classes were restricted to those with at least 3 suggestively associated genes.

#### 4.2.2.5 Regional association plots

For the best-associated regions (I and II), regional association plots were produced. The regional association plots were made with R. An R script from P. I. de Bakker ([www.broadinstitute.org/diabetes/scandinavians/figures.html](http://www.broadinstitute.org/diabetes/scandinavians/figures.html)) was modified to make the plots. Background LD information was acquired from HapMap (International HapMap, 2003) and LD information between the analysed markers were estimated from the data.

#### 4.2.2.6 Prediction of putative regulatory sites (I)

The best-associated region was studied to see if the associated site might have some regulatory effect on the nearby genes. First, known regulatory sites were gathered from the literature to define the regulatory region of *GATA2*. Other nearby genes did not have known functions or published regulatory sites. Secondly, DNase I hypersensitivity sites from ENCODE data (Encode Project Consortium, 2012) and their conservation (Placental Chain/Net track, hg19) were gathered from UCSC (Meyer et al., 2013) to predict putative regulatory sites. Regulatory elements are known to show increased DNase I hypersensitivity (Heintzman et al., 2007) and enhancers are often conserved in evolution (Pennacchio et al., 2006). Thus, these two signals were used to predict putative regulatory sites near *GATA2*.

#### 4.2.3 Positive selection (III)

Regions showing signatures of positive selection for abilities that contribute to musical aptitude were searched with three metrics: haploPS (Liu et al., 2013), XP-EHH (Sabeti et al., 2007) and  $F_{ST}$  (Weir & Hill, 2002). Of these, haploPS and XP-EHH are haplotype-based methods that search for long haplotypes which show signatures of positive selection and are shared among cases but not with controls.

XP-EHH searches signatures from homozygosity that extend over long regions. XP-EHH scores were calculated around every marker for a  $\pm 1\text{Mb}$  region as a logarithm of the ratio of homozygosity in cases compared to controls. Based on these raw scores, the top 0.1% of the markers were selected to identify regions with multiple significant markers less than 200kb apart from each other.

HaploPS searches for long haplotypes separately within cases and within controls. The haplotypes were constructed around seed markers. Significance of the selection regions was calculated from the length (in cM) and the number of SNPs in the haplotype. The selection signatures related to musical aptitude were identified by comparing the regions observed using cases to the ones observed using controls. HaploPS was run for haplotypes with frequencies between 0.05 and 0.95. The differential selection signals were identified as having a p-value below 0.05 in cases and above 0.10 in controls.

$F_{ST}$  is a parameter to describe the genetic structure of populations (Weir & Hill, 2002), and is calculated from SNP allele frequencies between cases and controls. Applied to populations of cases versus controls, it can be used to search regions with maximum differentiation.  $F_{ST}$  values were calculated for every SNP and 200kb windows were used to compute window statistic from the top three SNPs within the region. Only windows with at least 5 SNPs were included. The top 1% of the window statistics distribution was used as a threshold for significant finding (Akey, 2009).

Haplotype phases that were used in the haploPS and XP-EHH analyses were estimated with Beagle 3.3.2 (Browning & Browning, 2007) for 286 samples. The 148 cases and controls (see section 4.1.3) were used for the selection signature analyses. Annotations were based on the hg19 genome annotation.

Gene ontology analysis was performed with Genetrail2 (<http://genetrail2.bioinf.uni-sb.de>). The resulting regions from all three methods were analysed together. The enrichment analysis was performed with a hypergeometric distribution test with a false discovery rate of 0.005. Genes belonging to the same gene families may cluster into the same genomic region, which may lead to inflated significance of the related ontology class. To avoid this problem, we excluded ontology terms that had related genes only at one candidate region.

## 4.3 Materials and methods in convergent evidence analysis (IV)

### 4.3.1 Data

For the integrative analysis of published data, I collected music-related scientific articles which had gene-related results. All studies regardless of the studied animal species were included. The articles were searched through Google Scholar (<https://scholar.google.fi>), Pubmed ([www.ncbi.nlm.nih.gov/pubmed/](http://www.ncbi.nlm.nih.gov/pubmed/)), Web of Science (<http://apps.webofknowledge.com>) and from the references of already included studies and related reviews. We searched music-related articles that

included genotyping, sequencing, expression, knockdown or other molecular evidence (such as hormonal levels). The searched phenotypes were restricted to: music (listening, exposure, memory, aptitude, abilities, cognition, creativity, improvisation, composition or making), singing, playing instruments; vocalization, vocal/auditory phenotypes (learning, memory, plasticity, processing, perception, stimulation, imitation, improvisation or behaviour), absolute pitch, amusia, song (listening, exposure, memory, producing, complexity or variability) or musicians. Furthermore, articles with phenotypes restricted to speech (or speech-related characteristics) or considering any sounds were excluded; for example, not all studies considering singing were chosen. Additionally, we searched for studies about vocal/song control system but they were only included if suitable phenotype was used.

Based on the titles and abstracts, I chose a list of 326 studies that were further examined to find those with relevant results. Studies were excluded if the phenotype was not relevant in music, if there was similar work from the same group already included, or if there were no significant molecular results reported. Animal studies where phenotype was related to music perception, listening or practice were included. Results were extracted from 101 articles using publicly available data or by contacting the authors (Table 3).

The human gene information was gathered for the extracted animal model genes and the human genomic regions using biomaRt (Durinck et al., 2009) from Bioconductor in RStudio. Hormones and biomarkers that have no encoding genes but are synthesised from other substances were included as such. HGNC (HUGO Gene Nomenclature Committee) gene symbols were used when available (Gray et al., 2015). Genomic locations of the association and linkage studies were translated into hg38 reference genome (Ensembl, GRCh38.p5) through reported markers to assign adjacent genes to them. Genes within  $\pm 500\text{kb}$  region around the associated markers were included. All genes within linkage regions with  $\text{LOD} \geq 3$  were included. When the borders of the linkage region were only reported as cM (without reference map), the regions around the reported peak markers were translated into bp information using the formula  $1\text{cM} = 1\text{Mb}$ . From our own gene mapping of musical aptitude (I), I included all regions showing probability score above 0.2 and 0.3 from association and linkage, respectively.

The human homologous genes for the genes and molecules identified in bird and other species studies were gathered primarily from biomaRt, and for the unavailable genes data was gathered using the Ensembl ([www.ensembl.org](http://www.ensembl.org)), UniProt

**Table 3. Studies used in the CFG analysis of musical traits.** Some studies included evidence from multiple molecular levels (DNA, RNA, protein or other).

Species	# studies	DNA, GW	DNA, candidate	RNA, GW	RNA, candidate	Other levels	Brain tissue or imaging used
Human	33	5	3	3	1	20	1
Songbirds	55	0	0	8	16	32	All
Other animals	13	0	0	2	2	11	All

GW, genome-wide

([www.uniprot.org](http://www.uniprot.org)), miRBase (<http://mirbase.org>, version 21), EggNOG (<http://eggnogdb.embl.de>), OrthoDB (<http://cegg.unige.ch>), Bird Base (<http://birdbase.arizona.edu>) and BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) databases. Most of the genes that were successfully found from the databases were matched to at least one human homolog. However, some non-human microarray studies included large proportion of probes that were not linked to any known gene.

#### 4.3.2 CFG method

Chakravarthi Kanduri implemented the CFG method “GenRank” in RStudio and it is available from Bioconductor (<http://bioconductor.org>) and Github (<https://github.com/chakri9/GenRank>). The identified genes are ranked based on the input scoring of each layer and the number of layers the gene has been identified on. Multiple hits per gene within one layer are ignored. All genes within a layer are given the same input score.

Each study was treated as a layer. Exceptions were the studies where similar methods and the same samples were used. For example, when upregulated and downregulated genes were reported in separate studies, the results were combined. Contrary, linkage and association related results from the same article were considered as two separate layers. Additionally, a study by Drnevich et al. (2012) reanalysed results from multiple original studies of which the song-related results were included as separate layers instead of the original studies. As a result, there were 105 layers in the analysis.

All studies were given different scores based on the sample size, homology translation and phenotype. Sample sizes were scored from 0.8 to 1 within two study categories: DNA and expression level studies. The expression level includes RNA, protein and biomarker studies where large scale studies are often ethically infeasible. Contrary, the golden standard for DNA level studies like GWAS require thousands of individuals. The study with the lowest sample size within the category was assigned 0.8 and the highest was assigned 1. The homology translation may include errors or some genes may not be found from animal models or their function may be different and, thus, the animal model studies were given a score of 0.8 and human studies 1. Studies considering musical abilities and settings including music were given a score of 1 and phenotypes only related to music were given 0.8. These three scorings were combined by multiplication, which resulted in the final scores ranging from 0.512 to 1. It was not possible to fully identify all factors affecting reliability or specificity for music. Thus, the scoring was kept relatively simple.

#### 4.3.3 Enrichment and network analyses

The ranked results were further annotated with enrichment and network analyses with IPA, as in the gene mapping studies (I, II). Contrary to other common enrichment methods, genes and biomarkers could be used as inputs in the same analysis. The top 40 molecules were used to evaluate the biological functions,

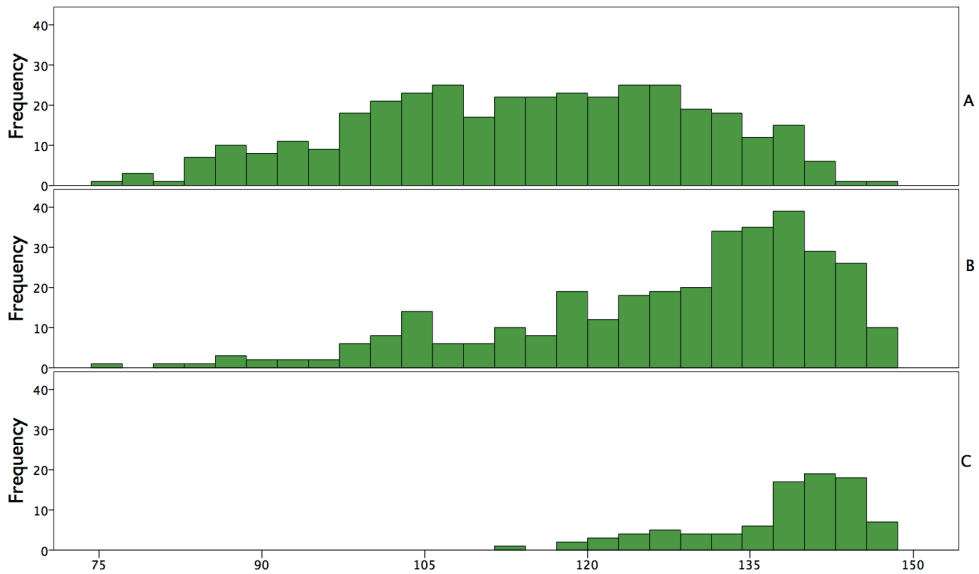
pathways, upregulators and interaction networks related to musical traits. There is no generally accepted way to define the number of molecules for the following analyses; the chosen number was considered small enough not to include many false positives and large enough to perform the analyses. In IPA, interaction networks are evaluated by growing network around most interconnected molecules of the input molecules, here the top 40 molecules, and adding minimum number of connectors to show interactions to other top molecules. The interconnectivity is measured by the number of triangles (when three genes interact with each other, they form a triangle) the gene belongs to. Biological function of each predicted interaction network was estimated from all the genes within the network, including also the added connectors.



## 5 Results

### 5.1 Descriptive statistics (I, II, III)

Music-related phenotypes of 915 Finnish individuals were studied. The phenotypes included musical aptitude scores from the three tests and the questionnaire (Tables 4 and 5). Among the individuals with no or only a little (<2 years) formal education in music, the musical aptitude scores showed Gaussian distribution (Figure 5a). The scores of the musically educated skewed towards lower scores (Figure 5c).



**Figure 5 Histograms of COMB.** Histograms are shown in three sub groups of the data: A) less than 2 years of music education; B) more than 2 years of music education and C) musicians and other music professionals.

The proportion of musically-educated individuals is supposedly higher in our sample than in the general population (12% musicians or other professionals in music, see also Table 5). The main cause for this is self-selection bias as individuals interested in music were more likely to attend the study. The family setting can reduce the bias as relatives were requested to attend the study regardless of their interest in music. However, the large number of musically-educated individuals enabled us to study phenotypes not evident in the musically non-educated individuals, such as musical creativity.

**Table 4. Characteristics of the quantitative traits (N=915)**

	Min	Max	Mean	Mean among females	Mean among males	N
Age	7	94	40.6	39.9	41.1	893
KMT	20	40	31.7	31.5	31.8	867
SP	25	50	41.9	41.8	42.1	865
ST	25	49	40.3	39.6	41.2	866
COMB	75	148	121.8	120.8	123.2	865

**Table 5. Characteristics of the qualitative traits**

	<i>Females</i>	<i>Males</i>
Gender	536 (58.6%)	379 (41.4%)
	<i>Yes</i>	<i>No</i>
Arranging*	120 (42.3%)	164 (57.7%)
Composing*	103 (36.0%)	183 (64.0%)
Improvising*	178 (62.9%)	105 (37.1%)
Other creativity	259 (61.5%)	162 (38.5%)
NCNA	159 (51.6%)	149 (48.4%)
	<i>&gt;2 years</i>	<i>&lt;2 years</i>
Musical education	428 (53.6%)	371 (46.4%)

\* Of the musically experienced individuals

Most of the music-related traits correlated with each other, as expected, because they all relate to underlying musicality. The music scores from different tests correlated with each other significantly (Table 6). ST was found to be the least correlated with the other tests. It has also been shown to have the lowest reliability of the three tests (Cronbach's alpha 0.78, whereas 0.88 and 0.91 for KMT and SP, respectively) (Pulli et al., 2008). However, the estimates of correlation were not corrected for the relatedness of the test subjects, and are thus only approximate. Prior studies with population samples have, however, shown similar correlations (Shuter-Dyson & Gabriel, 1981).

The music scores also showed correlation with most variables illustrating music education. An exception was the ST score; ST did not correlate with music education, nor starting age for music practice, unlike SP and KMT. It must be noted that the correlation between the test results and music education do not inform about causality; based on the correlation we cannot say whether individuals will become better because of music practice, or do they choose to start playing music because they are talented towards music. However, there have been more comprehensive studies where an early starting age for music practice actually seems to have a causal effect on music abilities (Hyde et al., 2009a; Baer et al., 2015). Also, the nominal music-related variables show connection with the music scores. For example, individuals active in composing, arranging and/or improvising activities show significantly higher music test scores than the non-active participants.

There are differences between age groups in the music test scores as well as other music-related traits, as is also shown in earlier studies (Stanton, 1922; Friend, 1939). Children achieve lower scores in the music tests than individuals on average (Pulli et al., 2008). This can be due to a weaker ability to concentrate in the testing situation,

**Table 6 Correlations between music test scores and some other variables.** The correlation coefficients for the Spearman correlation are shown for the significant bivariate correlations.

	<b>SP</b>	<b>ST</b>	<b>Starting age</b>	<b>Age</b>
KMT	0.62	0.48	0.22	0.25
SP		0.41	0.23	0.24
ST			*	0.19
Starting age				0.36

\* Not significant at 0.01 level

or to understand the concepts in the test. Also, older individuals achieve lower scores and improvise, arrange and compose music less frequently than other age groups. An early study by Stanton (1922) suggested that the reason was that younger individuals are more adaptable to the testing situation. Another explanation is differences in hearing. Moreover, the childhood environment between the oldest participants and the younger ones has been very different. The opportunity to hear professionally played music has greatly increased over the last decades. Younger individuals are also encouraged to be creative, in music and in other fields.

Males achieved higher scores from ST (p-value 0.001, corrected for pedigree information) and were more active in composing and arranging. However, there is no difference between musical education level, age or own assessment of creativity in non-musical fields between males and females. The gender differences in musical abilities have also been found in previous studies. For example, AP is more common in females (Profita & Bidder, 1988) and males did perform better in an off-beat test in a twin study by Seesjärvi et al. (2016). Moreover, there are differences between genders in music processing in the brain as measured by EEG (electroencephalography) (Koelsch et al., 2003). It has been proposed that there are differences in the preferences and on certain components but not in the overall musicality (O'Neill, 1997; Hallam et al., 2008).

## 5.2 Heritabilities of musical traits (I, II)

The heritabilities varied from 11% up to 68% (Table 7). The phenotypes with heritability below 25% were excluded from the gene mapping studies. Generally, the estimated heritabilities were similar to the ones reported in previous music-related studies.

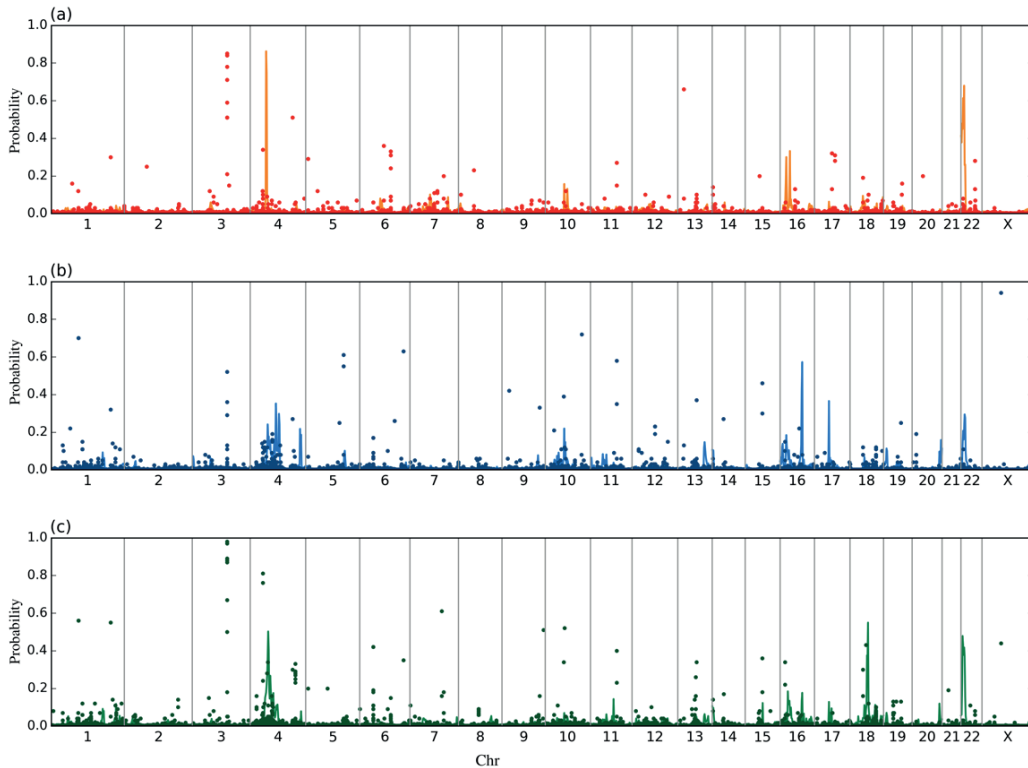
## 5.3 Gene mapping of musical aptitude (I)

We performed PPL and PPLD analyses for music test scores (KMT, SP and COMB) in extended families. The results are given as posterior probabilities ranging from 0 to 1 where 1 stands for proof. Values below 0.02 denote evidence against linkage or LD.

**Table 7. Heritability estimates (I and II)**

<i>Trait</i>	<b>Yes</b>		<b>No</b>		<b>Total</b>	<b>Heritability</b>
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>b</i> <sup>2</sup>
Composing	103	36.0	183	64.0	286	0.33
Arranging	120	42.3	164	57.7	284	0.33
NCNA			149	48.4	308	0.29
Improvising	178	62.9	105	37.1	283	0.12
Other creativity	259	61.5	162	38.5	421	0.34
KMT*					757	0.46
SP*					757	0.68
ST*					757	0.21
COMB*					757	0.60

\* continuous traits



**Figure 6. Linkage and LD results for musical aptitude.** The linkage and LD analyses were performed for SP (a), KMT (b) and COMB (c). The lines denote linkage results and dots association results. Note, all association results are shown here regardless of probability or signal cluster information (see chapter 4.2.2.2).

### 5.3.1 Linkage

We found multiple linked regions throughout the genome. Region at 4p14 was the best-linked to SP (PPL 0.86). Moreover, the region that spanned from 4p15 to 4q26 was found to be linked to all phenotypes: KMT, SP and COMB (Figure 6).

Region at 22q11.1-21 was also linked to all three phenotypes (best PPL 0.68). The region has also been linked to 22q11 deletion syndrome, also known as DiGeorge syndrome (MIM#188400) that causes learning disabilities and has hearing deficits as one of the minor symptoms (Bassett et al., 2005; Umlauf, 2008). The combined scores were best linked to 18q12.3-21.1 (PPL 0.55) that includes the *LOXHD1* gene whose mutations lead to auditory defects (Grillet et al., 2009). Other linked regions included 16q21-22.1 (PPL 0.57), 17q11.2 (PPL 0.37) and 10q21.1 (PPL 0.22) for KMT, and 16p13.2 (PPL 0.30) and 16p12.3 (PPL 0.33) for SP.

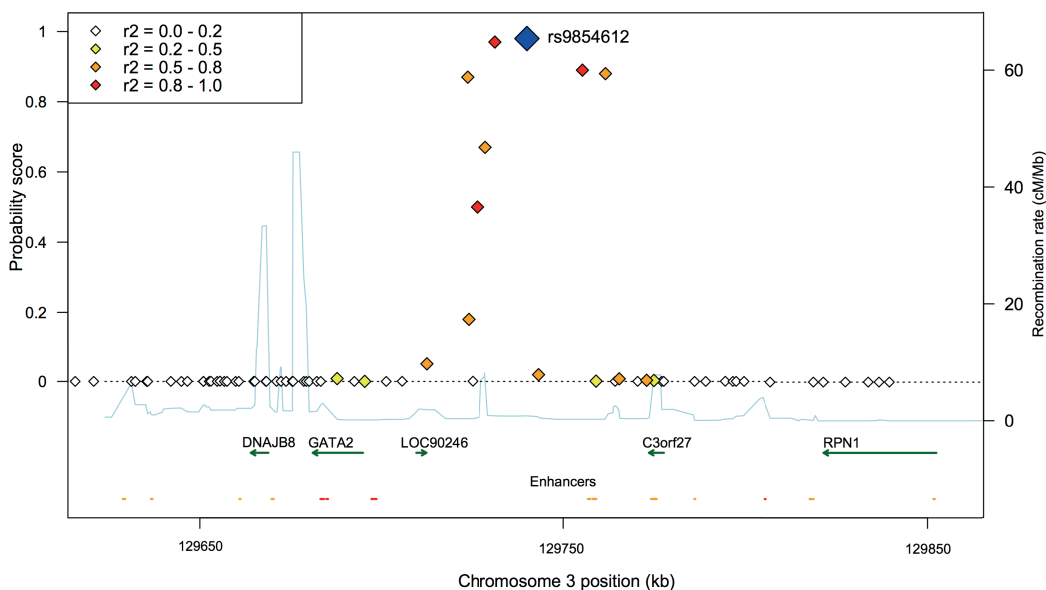
### 5.3.2 Linkage disequilibrium

The most prominent association for musical aptitude was at 3q21.3 (PPLD 0.98; Figure 6). The region is located near *GATA2*, between known enhancer binding sites affecting *GATA2* expression (Figure 7) (Zhou et al., 2000; Kobayashi-Osaki et al., 2005; Grass et al., 2006; Khandekar et al., 2007; Wozniak et al., 2008; Nozawa et

al., 2009). Based on the haplotype analysis, the best haplotype to spans over the entire gene. In the region, there are also genes *LOC90246* and *C3orf27*. Further analysis showed a putative enhancer site near the best-associated SNP rs9854612. The *GATA2* gene is a general transcription factor involved in development of several organs, including the inner ear and the inferior colliculus, both important in auditory processing (Haugas et al., 2010; Lahti et al., 2013).

The highest PPLD probabilities were located outside of the linked regions. For example, the associated regions at chromosome 3 and 6 were not spotted in linkage analyses. Similarly, at the best-linked region at 4p14, there was no evident PPLD within the linked region. However, there were PPLD results for SP and COMB near the region at 4p14 adjacent to the *PCDH7* gene which has been shown to affect the development of the chicken cochlea (Lin et al., 2012) and mouse amygdaloid system (Hertel et al., 2012). Also, the 10q21.1 region that was linked with KMT includes rs2660180 that was associated with COMB (PPLD 0.39). The SNP is within *PCDH15* whose mutations cause deafness (Ahmed et al., 2001).

The linked region for COMB at the q arm of chromosome 4 showed PPLD for rs2412501. The SNP was reported to be located near the *PDGFRA* gene using the hg19 genome annotation. However, using the newer hg38 annotation and RefSeq genes, the SNP is closer to genes *LNK1* and *CHIC2*. *CHIC2* is a membrane structure gene and *LNK1* encodes for a membrane-bound protein with a possible function in tumorigenesis.



**Figure 7. The most prominent association was found at chromosome 3 near the *GATA2* gene.** The region contains known enhancers for *GATA2* in both sides of the associated markers (red enhancers at the bottom of the figure). The linkage disequilibrium between the best associated SNP, rs9854612 (blue diamond), and the other SNPs are color-coded according to their pairwise R-squared values (see figure legend). The region was further analysed to look for putative enhancers (see chapter 4.2.2.6) that are shown here as orange lines at the bottom of the figure.

### 5.3.3 Enriched functions and pathways among the suggested genes

The genes within linked regions were further studied to find biological functions enriched in the results. The enriched functions included schizophrenia (p-value  $1.0 \times 10^{-7}$ , 27 genes within linked regions) as well as other psychiatric disorders, and inner ear development (p-value  $4.7 \times 10^{-4}$ , 7 genes within linked regions). Additionally, we identified genes with hearing and inner ear related functions with PPLD, such as *GATA2*, *PCDH7* and *PCDH15*. The protocadherins, such as *PCDH15* and *PCDH7*, also relate to serotonergic systems and psychiatric disorders. Thus, they may also have other functions important for music in addition to hearing. Unfortunately, there were not enough genes to perform informative enrichment analysis for PPLD results.

## 5.4 Gene mapping of creative activities in music (II)

We performed gene-mapping analyses for self-reported composing, arranging and non-music-related creativity. We also compiled a new phenotype, NCNA, that included music-experienced individuals who neither composed nor arranged music (see chapter 4.1.3). These four phenotypes were studied with linkage and LD analyses.

### 5.4.1 Linkage

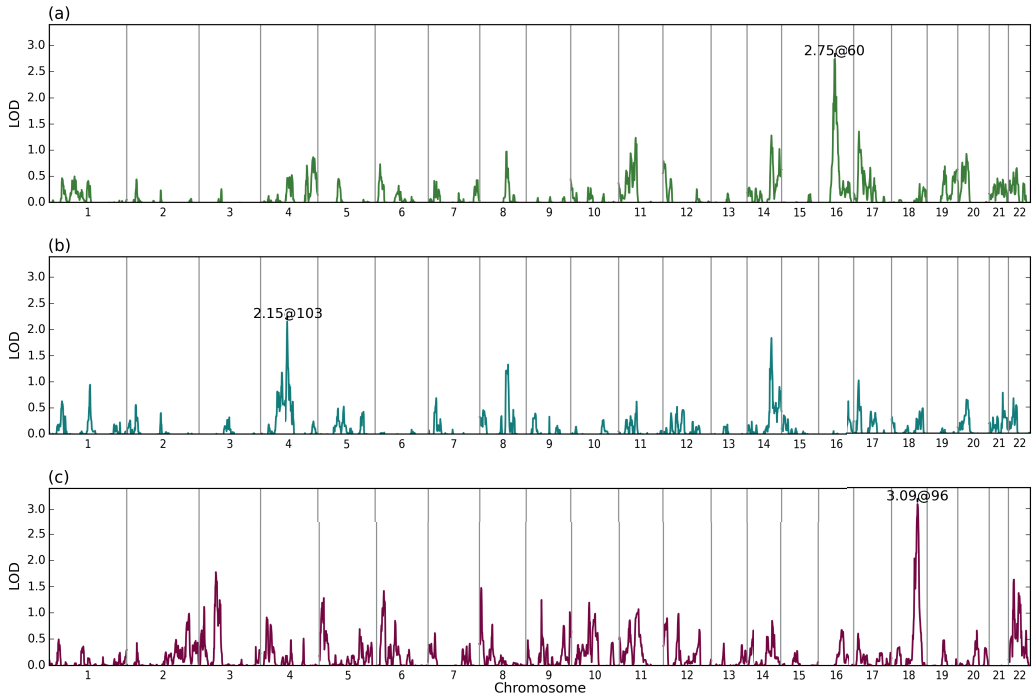
We found significant linkage at 18q21 for NCNA (LOD 3.09, Figure 8). The region includes, for example, cadherin genes (*CDH7*, *CDH19* and *CDH20*), which affect neuropsychiatric disorders (Redies et al., 2012). Composing was suggestively linked at 4q22.1 (LOD 2.15) and arranging at 16p12.1-q12.1 (LOD 2.75). All of these three linked regions are adjacent to the regions identified previously in the musical aptitude gene mapping (I). The other creativity phenotype showed suggestive linkage at Xp11.23 (LOD 2.50) and no linkage at the music-related regions.

### 5.4.2 Joint linkage and linkage disequilibrium

Single marker analyses were performed as joint linkage and LD analyses. The smallest p-value was obtained at 10q22.1 for 'other creativity' (p-value  $1.5 \times 10^{-7}$ ). The best result for the music-related traits was found at 2p24.3 for arranging (p-value  $1.7 \times 10^{-7}$ ) near the *MYCN* gene that causes, for example, Feingold syndrome including intellectual disability and deafness (van Bokhoven et al., 2005). Only a few of the signals were found near the linked regions.

### 5.4.3 Enriched functions and pathways among the suggested genes

Enrichment analysis of the genes pinpointed by the joint linkage and LD were performed separately for each of the phenotypes. Genes belonging to the long-term depression (LTD) pathway (p-value  $9.0 \times 10^{-8}$ ) were enriched for composing. The pathway reduces synaptic efficacy after a stimulus for subsequent stimuli and is considered as a cellular model for synaptic plasticity and memory (Collingridge et al., 2010). Of the genes associated with the LTD pathway, 17 were suggestively



**Figure 8. Linkage results for creative activities in music: arranging (a), composing (b) and NCNA (c).** The three best signals are annotated with relative LOD score and peak region as cM.

associated with composing. Concerning arranging, the enrichment analysis showed X-linked autism and X-linked hereditary disease ( $p$ -values  $2.1 \times 10^{-6}$  and  $4.0 \times 10^{-6}$ , respectively).

Additional interest in the LTD pathway is raised by a single-point linkage that was previously obtained in the intron of the *GSG1L* gene (rs9933639, LOD 4.22). *GSG1L* is part of a form of the  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) complex. Interestingly, the AMPAR complex is a part of the LTD pathway.

## 5.5 Regions under positive selection in musical aptitude (III)

In Study III, we studied signatures of positive selection for abilities that contribute to musical aptitude using three different metrics: haploPS, XP-EHH and  $F_{ST}$ . The phenotype was based on the COMB scores. The results for each of these methods are shown in Figure 9.

### 5.5.1 haploPS

With haploPS, we identified twelve regions that showed signature of positive selection only in cases (Figure 9b). Additionally, there were 34 identified regions evident in both cases and controls. The best two case-specific signals were found at 12q15-21.2 (haploPS  $p$ -value 0.0053 and 0.0077). The third best region was found at 2q12 ( $p$ -value 0.0115). These three signals included haplotypes spanning over 3.5cM regions but had relatively low frequency (0.05). The other nine identified

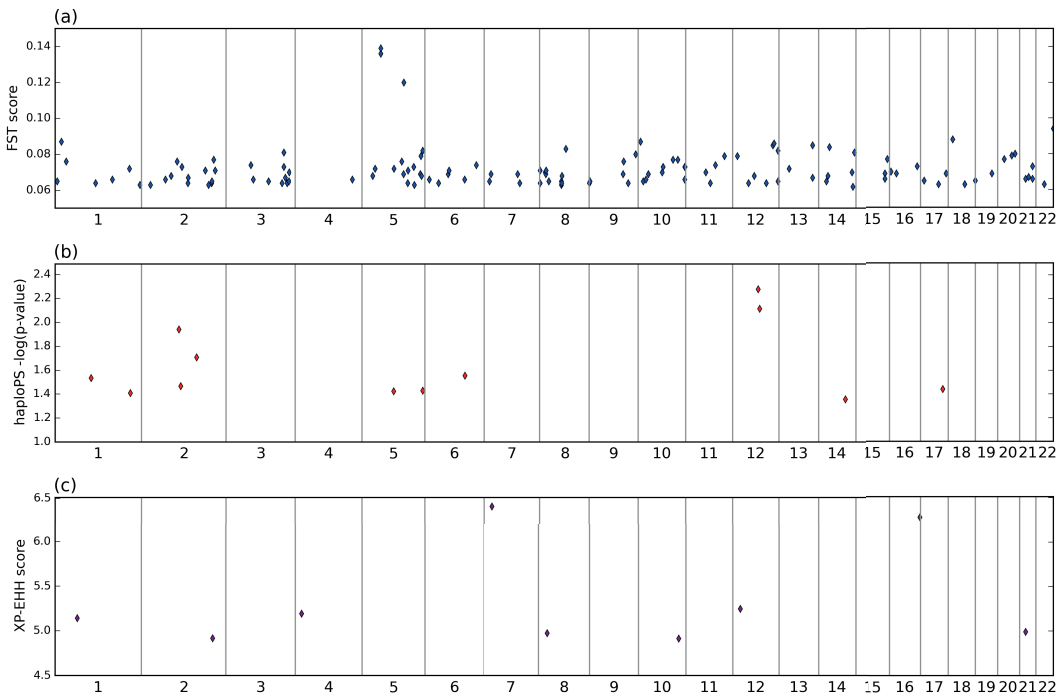


regions were shorter and more variant. For example, the region at 2q24.1 spanned only 0.2cM but the frequency was as high as 0.9 ( $p$ -value 0.0197).

Because most of the regions were very long there were many genes per region. The regions contained multiple brain-related genes. For example, the chromosome 12 regions include *SYT1*, that works in synapses (Craxton, 2004), and *KCNC2* that has been associated with neurodevelopmental disease (Rajakulendran et al., 2013). Some hearing-related genes were also seen in the case-specific regions. For example, *ADGRV1* (previously *GPR98*) is located at the 5q14.3 region and *USH2A* at the 1q32.3-41 region, which are both related to development of hearing-related cells (McGee et al., 2006; Liu et al., 2007). Moreover, the *ADGRV1* gene has been shown to suppress during song listening and also shows evidence for positive selection in zebra finches (Warren et al., 2010).

## 5.5.2 XP-EHH

Only 9 regions showing signatures of selection were identified with the XP-EHH method (Figure 9c). The regions were also shorter than the ones identified with haploPS. One of the top signals was found at 7p15.3 near *RAPGEF5*, a RAS protein activator that has been related to memory functions (Ostroveanu et al., 2010). Most of the identified regions contained genes with unknown function or no genes at all.



**Figure 9. Positively selected genomic regions related to musical aptitude.** The three metrics  $F_{ST}$  (a), haploPS (b) and XP-EHH (c) found signals mostly at different regions. Additionally, the methodological differences resulted in variability in length of the identified regions. Here, the lengths of the regions are not on scale. Note, all three metrics use different scoring and thus, the scales (y-axis) of the results of these three methods vary. The haploPS  $p$ -values are transformed as  $-\log(p\text{-value})$ .

### 5.5.3 $F_{ST}$

In total, 128 regions were detected within the top 1% of  $F_{ST}$  variation (Figure 9a). The best region was found at 5q11.2 covering two adjacent windows ( $F_{ST}$  score 0.139). The region includes multiple genes, for example, *GZMA*, *ESM1* and *MCIDAS*. The *GZMA* gene is downregulated during music listening in humans (Kanduri, Raijas, et al., 2015) and *ESM1* homolog *Esml* is upregulated during music exposure in mice (Meng et al., 2009).

At 12p13.1 ( $F_{ST}$  score 0.079), there is *GRIN2B* whose homolog has been shown to be affected by music exposure in rats (Xu et al., 2009). In zebra finches, *GRIN2B* expression is affected by song isolation during the sensitivity period (Singh et al., 2000) as well as singing (Hilliard et al., 2012; Whitney et al., 2014). *GRIN2B* encodes a subunit of the N-methyl-D-aspartate (NMDA) receptor that is an excitatory neurotransmitter receptor in brain. In humans, *GRIN2B* variants have been associated with memory (de Quervain & Papassotiropoulos, 2006) and other cognition related phenotypes like dyslexia (Mascheretti et al., 2015). However, these regions with a  $F_{ST}$  score below 0.1 are less reliable.

Additionally, there was another region at 5q23.1 showing an  $F_{ST}$  score over 0.10. Within this region, there are no genes but there is a suggestive GWAS result (rs13166268) for verbal declarative memory (Debette et al., 2015). Another interesting region was at 21q22.12 where *DOPEY2* has been found to be upregulated during music performance (Kanduri, Kuusi, et al., 2015).

### 5.5.4 Difference between the metrics

None of the identified regions were identified with all three methods. HaploPS and  $F_{ST}$  recognised a common region at 6q21 which includes *MARCKS*, *HDAC2* and *LINC01268*. Homologs of *MARCKS* and *HDAC2* have been shown to be regulated during singing in zebra finches (Hilliard et al., 2012; Whitney et al., 2014).

The number of identified regions was considerably higher with  $F_{ST}$  than with the other two methods.  $F_{ST}$  only considers the top three SNPs from every 200kb region whereas all SNPs are considered in the other two methods. The other two methods are based on haplotypes, contrary to individual SNPs in  $F_{ST}$ . Also, the resulting regions predicted with  $F_{ST}$  and XP-EHH are approximately 200kb long, whereas the top region from haploPS was almost 5Mb long. Similarly, the three methods consider the phenotype differently. HaploPS and XP-EHH search for case-specific selection signals whereas  $F_{ST}$  studies the frequency differences between cases and controls, which resembles more the association setting. Even for the haplotype-based methods, the sharing of detected positive selection is low as demonstrated by Liu et al. (2013). XH-EHH and haploPS are powered at detecting selection signals of different derived allele frequency ranges.  $F_{ST}$  and haplotype-based methods search for very different kind of signatures from the genome, which results in differences in the identified regions.

### 5.5.5 Enrichment analysis

The enrichment analysis of all genes falling within the regions identified by any of the three metrics referred to, once again, inner ear development (p-value 0.002, 12 genes within the resulting regions) and to cellular component assembly involved in morphogenesis of different organs (p-value 0.02, 14 genes within the resulting regions).

## 5.6 Convergent evidence of musical abilities (IV)

The studies related to music included evidence for 7883 different genes. Most of the genes have been nominated by a single study (5889 genes), and many within linked regions (1753 genes). The CFG method ranked *EGR1*, cortisol, *FOS* and *FOXP2* as most likely music related molecules (Table 8). *EGR1* (also called *ZENK*) had been pinpointed in 29 studies investigating various songbirds (for example Mello et al., 1992; Jarvis et al., 1998; Warren et al., 2010; Hilliard et al., 2012; Whitney et al., 2014; Mori and Wada, 2015) and in one study using frogs (Mangiamele & Burmeister, 2008). *EGR1* had been shown to be upregulated during song listening in frogs and during song listening and singing in song nuclei of zebra finch and other songbirds. It is also related to song memory at the NCM in the forebrain auditory region in zebra finches (Dong & Clayton, 2008). *EGR1* is an immediate early gene (IEG), which is a term used for genes that respond rapidly to cellular stimuli. *EGR1* functions as a transcriptional regulator and affects learning (Davis et al., 2010) and immune system by interleukin receptor signaling (Decker et al., 1998).

*FOS* is also an IEG. It has a role in signal transduction and cell differentiation. It has been linked to music-related abilities in 18 studies; including expression analysis of music performance where it was upregulated (Kanduri, Kuusi, et al., 2015). Similarly, *FOS* was found to be upregulated by music in the auditory cortex in mice (Arnauld et al., 1996; Meng et al., 2009). In zebra finches, the expression is highest after 30 minutes of song listening (Velho et al., 2005). *FOXP2* is a transcription factor that is commonly known to affect language development in humans (Fisher & Scharff, 2009). Concerning music, it was found by 12 zebra finch studies and two studies considering vocalisation in mice. In zebra finches, it is downregulated during singing, but adequate levels are needed in song imitation (Murugan et al., 2013; Heston & White, 2015).

Cortisol was linked to music only in human studies (total of 14 studies). Cortisol is a hormone affecting stress response by inhibiting production of many interleukins and other factors. It also affects long-term memory (de Quervain et al., 2000).

The functional enrichment analysis of the top 40 ranked genes showed cognition, memory, learning and excitation of neurons (p-value  $<1 \cdot 10^{-11}$ ). There were 16 cognition related genes among the results, of which 8 related to excitation of neurons and 10 to long-term potentiation (LTP). Additionally, network analysis showed four non-overlapping networks of which a network relating to behaviour (11 ranked genes) was the most significant (see Table 8 for full list of related genes). The non-overlapping networks may tell about separate biological pathways affecting music-

related traits. The enrichment analysis of pathways pointed out CDK5 signaling pathway (Table 8) that affects brain development and cognition (Shah & Lahiri, 2014).

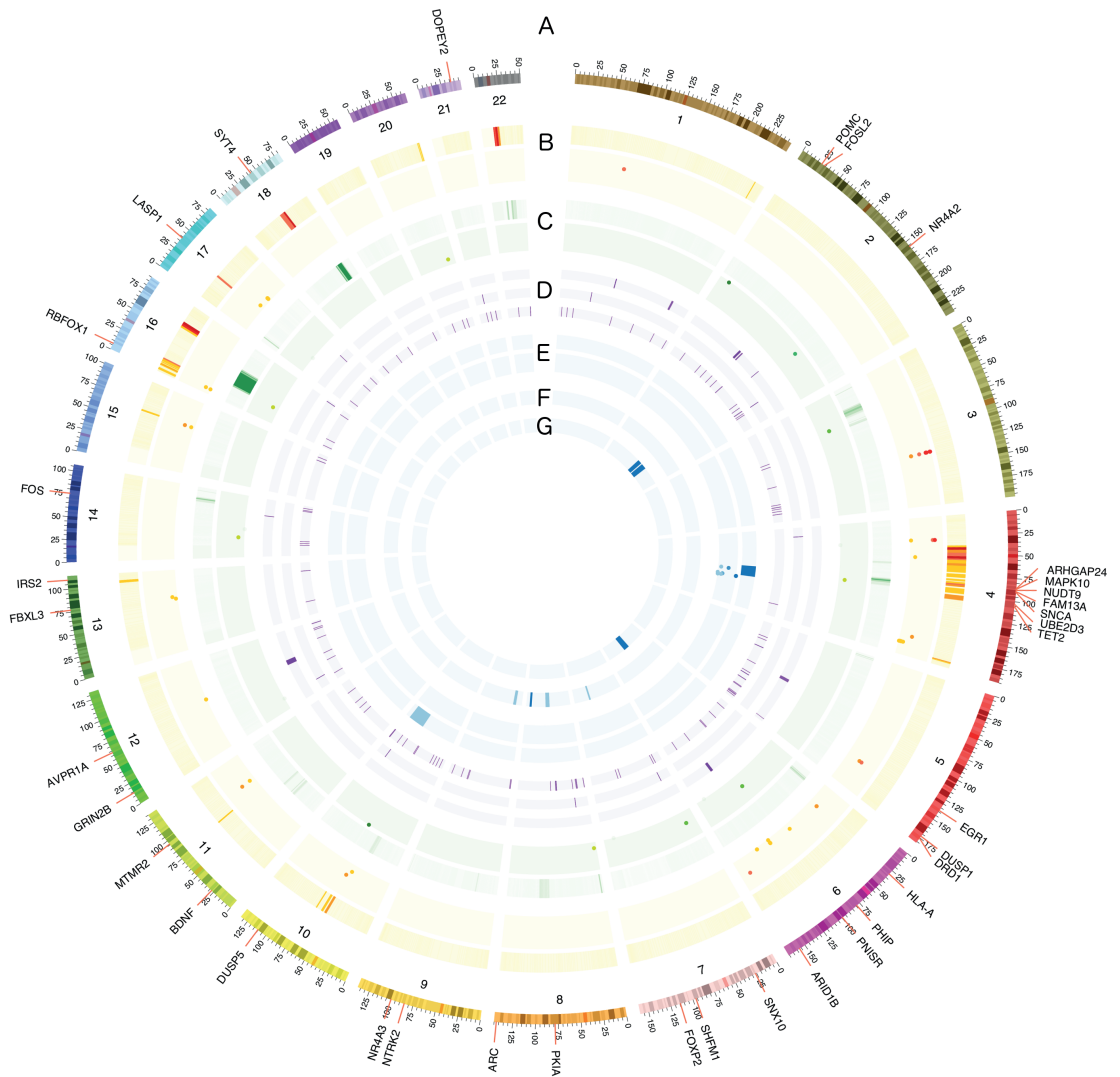
**Table 8. Top 40 ranked genes in CFG analysis of music-related abilities.**

Gene information		Result	Indicated by	Indicated by human studies	Network **	Indicated in this thesis
<i>Gene symbol</i>	<i>Genomic location in humans</i>	<i>CFG Score</i>	<i># studies</i>	<i># studies</i>	<i>Part of</i>	
<i>EGR1*</i>	5q23-q31	0.233	30	-	A	-
Cortisol	-	0.166	13	14	B	-
<i>FOS</i>	14q24.3	0.148	18	1	D	-
<i>FOXP2</i>	7q31	0.109	14	-	-	-
<i>ARC</i>	8q24.3	0.071	9	-	A	-
Dopamine*	-	0.062	7	1	A	-
<i>BDNF</i>	11p14.1	0.073	8	-	A	-
<i>GRIN2B</i>	12p13.1	0.052	6	1	A	III
<i>SYT4</i>	18q12.3	0.052	6	1	D	I linkage
<i>PHIP</i>	6q14	0.050	5	2	B	I association
<i>MAPK10*</i>	4q22-q23	0.050	5	2	D	I linkage
<i>DRD1*</i>	5q34-q35	0.049	6	-	A	-
<i>SNCA</i>	4q21.3-q22	0.048	4	3	D	-
<i>NR4A3</i>	9q22	0.046	6	-	A	-
<i>IRS2</i>	13q34	0.046	6	-	B	-
<i>ARHGAP24</i>	4q22.1	0.046	4	3	-	I linkage
Noradrenalin	-	0.046	4	3	A	-
<i>MTMR2</i>	11q22	0.045	5	1	B	I association
<i>NR4A2</i>	2q22-q23	0.045	5	1	A	-
<i>DUSP1</i>	5q35.1	0.045	5	1	B	-
<i>DUSP5</i>	10q25	0.044	5	1	B	-
<i>PKLA</i>	8q21.13	0.043	5	1	B	-
<i>PNISR</i>	6q16.3	0.043	5	1	C	-
Estradiol	-	0.043	5	1	-	-
<i>TET2</i>	4q24	0.042	4	2	-	I linkage
<i>UBE2D3</i>	4q24	0.042	4	2	C	I linkage
<i>FAM13A</i>	4q22.1	0.042	4	2	C	I linkage
<i>NUDT9</i>	4q22.1	0.042	4	2	B	I linkage
<i>DOPEY2</i>	21q22.2	0.041	4	2	C	III
<i>NTRK2*</i>	9q22.1	0.041	5	-	A	-
<i>AVPR1A</i>	12q14.2	0.039	3	3	-	I association
<i>FOSL2</i>	2p23.3	0.039	5	-	B	-
<i>HLA-A</i>	6p21.3	0.038	3	2	D	I association
<i>FBXL3</i>	13q22	0.038	4	1	C	I association
<i>LASP1</i>	17q11-q21.3	0.038	4	1	B	I association
<i>RBFOX1</i>	16p13.3	0.038	4	1	C	I association
<i>SNX10</i>	7p15.2	0.038	4	1	C	-
<i>ARID1B</i>	6q25.3	0.037	4	1	C	I association
<i>POMC</i>	2p23	0.037	3	3	A	-
<i>SHFM1</i>	7q21.3	0.037	4	1	C	-

\* Part of CDK5 signaling pathway

\*\* A, behaviour-related network; B, C, D interaction networks with unknown function

The regions found in the gene mapping studies contained some of the highly ranked molecules (Figure 10). For example, *MAPK10*, *ARHGAP24* and *SNCA* are located at 4q22 near the linked regions. However, *SNCA* was not included in the layer of the genome-wide scan of musical aptitude (I) because the strict *SNCA* region had a probability below 0.3. *SNCA* encodes alpha synuclein that affects, for example, presynaptic signaling and is associated to early-onset Parkinson disease (Nuytemans et al., 2010). It is upregulated during music listening and performance (Kanduri, Kuusi, et al., 2015; Kanduri, Raijas, et al., 2015). The *SNCA* knockout mouse exhibits more vocalisations than wild-type mice (Kurz et al., 2010). *MAPK10* has only been detected in linkage analyses in humans. *ARHGAP24* has been detected in a music listening expression study (Qu et al., 2013) additional to the linkage regions. Musical aptitude association signals included the top-ranked *PHIP* that affects insulin signaling. Interestingly, a nonsense mutation in *PHIP* has been found from a girl with intellectual disability including speech delay (de Ligt et al., 2012).



**Figure 10. Top genes from the CFB analysis in context of the gene mapping results in this thesis.** The top-ranked genes are shown on the outermost circle of the figure (A). The inner circles show the best linkage results as bars and association results as dots for musical aptitude (B) and creativity (C). Selection signature regions are denoted with lilac colour (D). The blue bars and dots denote other published linkage and association regions: Park et al. 2012 (E), Theusch et al. 2009 (F) and Gregersen et al. 2013 (G). The darker the colour, the more significant or probable the result.

## 6 Discussion

### 6.1 Genomic regions affecting musical abilities

The gene mapping results of musical aptitude (I) as well as composing (II) showed evidence for linkage at chromosome 4. The linked region for musical aptitude was large, ranging from 4p15 to 4q26 and covers almost one third of the chromosome. The best linkage was found at 4p14, adjacent to the association signal near *PCDH7*. The region contains four other candidate genes (*PDGFRA*, *KCTD8*, *CHRNA9* and *PHOX2B*) that affect inner ear development and are expressed in the amygdala or hippocampus. Composing was linked to 4q22.1, the previously identified region of the KMT (I) and singing-based pitch accuracy test. The latter was associated to a coding variant of the *UGT8* gene (Park et al., 2012). In our data, we did not find associations nearby the *UGT8* gene, which may result from phenotypic or population differences – the allele associating in Mongolian families is rare in Europeans.

Unfortunately, the 4q region did not show prominent associations that would have helped to identify possible affective genes. The linkage peak of the KMT and COMB was of bimodal shape that may indicate multiple signals from the region. Whereas linkage analysis can indicate multiple separate signals with relatively weak individual effects, association analysis with our relatively modest sample size could fail in the same setting. Interestingly, the 4q region contains the *SNCA*, *MAPK10* and *ARGHAP24* genes, each of which were suggested in the CFG analysis. *MAPK10* is activated by extracellular stimuli and has been found to be upregulated in Alzheimer patients (Gourmaud et al., 2015). Its homolog has been differentially expressed during singing and song listening in zebra finches (Drnevich et al., 2012; Hilliard et al., 2012; Whitney et al., 2014). *ARHGAP24* is a GTPase regulator that has been found to be upregulated during music listening (Qu et al., 2013).

The shared linked regions between gene mapping analyses of musical aptitude and music-related creative activities also included regions at 18q21 and at chromosome 16 (Figure 10). The 18q21 region is located near the *CDH7* and *CDH19* genes that were pointed out by two-point linkage of NCNA. These type II cadherin genes encode calcium-dependent cell-cell adhesion molecules and *CDH7* has been associated with bipolar and depressive disorder (Redies et al., 2012; Li et al., 2014). Another interesting gene at chromosome 18 is *SYT4* because it was also highly ranked by CFG. *SYT4* affects noradrenalin flux that can be of interest in music (Wang et al., 2001). At chromosome 16, arranging was linked to a region located between regions linked to the KMT and SP. The SP-linked region at 16p includes association with *RBFOX1* (rs6500963) that was also highly ranked by CFG.

### 6.2 Limitations in the gene mapping studies (I, II)

There are several limitations in the performed gene mapping studies. The sample sizes were limited, especially for the creative activities traits (II). There were also



multiple phenotypes analysed which were not corrected in the chosen significance thresholds. Additionally, the splitting of the largest pedigrees may have even reduced the power (Cummings et al., 2013). Thus, the linkage and LD analyses of creative activities had insufficient power to detect other than major signals (see also chapter 2.1.3.2). On the other hand, the involvement of large families in the linkage and LD analyses yield higher power than the same sample size in a case-control setting (Hiekkalinna et al., 2012). Additionally, the power depends on the mode of inheritance of the underlying trait, which was unknown.

Overall, the linkage and association results indicate certain genomic regions instead of genes. It is not straightforward to infer the genes that might underlie the detected signals. Enrichment analysis of pathways and functions of genes within or nearby those assigned regions was performed in this thesis. However, this may involve complications. The causative variant can be further away from the detected locus than what was assumed by the selected window sizes. However, larger windows would have included even more genes that would have brought more noise to the analysis and the gene-rich regions would have had even more weight. Additionally, there is a lot of missing information about gene functions, and the functional classes are biased towards the commonly studied subjects. Thus, the CFG analysis was performed to more comprehensively rank the genes located near the loci detected by gene mapping methods.

### 6.3 Detected selection signature regions

The selection signature methods showed over one hundred regions in the genome characteristic to musical aptitude. These regions included over 500 genes, for example, *GRIN2B* and *DOPEY2* that ranked high in the CFG analysis. The regions detected by any of the three selection signature methods included inner ear development related genes. The same function was also found in gene mapping of musical aptitude using the same phenotype, even though the genomic regions were different.

The use of selection signature methods in the context of complex traits can be problematic. The methods may not perform optimally in the case of weak effects resulting from polygenic selection, even under high selection pressure (Pritchard et al., 2010; Kemper et al., 2014). The  $F_{ST}$  signals especially showed a dubious pattern. The  $F_{ST}$  is one of the most widely used statistics to study differentiation between different populations. Significant signals should show as outliers from the detected  $F_{ST}$  score variation. However, the 128 detected signals (99<sup>th</sup> percentile of the variation) did not rise high above the overall variation and, thus, can include spurious signals. Therefore, the signals detected with haploPS and XP-EHH may be more reliable, though they are also not designed for detecting polygenic selection. In fact, most of the inner ear related genes belonged to the haploPS regions. Overall, the methodology for detecting selection signatures affecting complex traits is still in developing stage (Qian et al., 2013).

Interestingly, the chromosome 4 linkage region did not include any selection signatures. There are multiple reasons that could explain the situation. First, only positively selected regions were searched for. If the region with the QTL is under balancing or negative selection, these methods cannot detect it. Second, the methods look for recent selection, whereas an ancient variant, where the haplotype has already been divided into shorter fragments over time, could not be detected. These favoured alleles, where time has already caused diversity in the surrounding region, can be detected with linkage (Vitti et al., 2013). The  $F_{ST}$  method is parallel to association, which was also not detected at the linked region. In fact,  $F_{ST}$  hits were more frequent in all other chromosomes compared to chromosome 4.

## 6.4 Genes suggested by the CFG method

The best-ranked molecules from CFG method were *EGR1*, *FOXP2*, *FOS* and cortisol (see Table 8 for a longer list). *EGR1*, *FOXP2* and *FOS* were also the top presented genes by songbird studies. *FOS* has been additionally shown to be upregulated in humans, mice and songbirds in perception or practice of music or singing (for example, Arnault et al., 1996; Kimpo and Doupe, 1997; Drnevich et al., 2012; Hilliard et al., 2012; Kanduri, Kuusi, et al., 2015). Cortisol has only been indicated in human studies. The top genes also included *PHIP*, *MAPK10*, *SNCA* and *ARHGAP24* that have been associated with musical abilities in more than one human study (see chapter 6.1).

Many of the top genes which were suggested only in animal models have a function in humans that can be important in music perception. For example, *BDNF* encodes a neurotrophin affecting a range of neuronal responses, like neuronal cell-to-cell signaling and LTP, which is an important molecular pathway affecting memory and learning. These functions can be important in music as were suggested by enrichment analyses (see chapter 6.5.1). A large difference between the included human and animal model studies was the source of the tissue. Brain tissue is mostly unavailable in humans; only in one human study brain tissue was used (Salimpoor et al., 2011) whereas all songbird and other animal model studies used brain as the main source tissue. Peripheral blood, that has been used as a proxy for unavailable tissues in many human studies, has been estimated to share most of the transcriptome with many of the brain tissues (Liew et al., 2006; Sullivan et al., 2006; Tylee et al., 2013). However, some genes like *EGR1* have shown considerable tissue-specificity in the brain of the zebra finch. When many responses occur in the brain and may not be directly shown in the blood, the information gap between, for example music stimuli and the expression changes in blood, need to be filled by using information from animal models.

The main concern about the CFG analysis regards the animal models that have been used. First, there are no commonly accepted animal models for music practice and perception. We gave a reduced weight for the animal model studies in the analysis to correct for uncertainty. However, as the number of the animal studies was relatively large, they do affect the results. Ultimately, it was not possible to leave all

uncertain phenotypes out, due to the small number of studies overall. Second, there are limitations when interpreting genetic effects identified in animal models back to humans. Even though the expressions have proven to be mostly correlated between the songbird and human auditory regions (Pfenning et al., 2014), there can be differences in the function of the genes. Moreover, not all genes exist in the animal models. For example, songbirds do not have any homolog for the *SYN1* gene (Warren et al., 2010).

Some concern arises with the fast evolution of cognition related genes. Because cognitive capacities in humans have recently evolved (Fisher & Ridley, 2013), it may have also driven the music-related genetic pathways in humans far from other animals. This may complicate the combined analysis of animal models and human studies. The genes shown with the CFG analysis supposedly have a conserved function between the species. Thus, the analysis may lack in showing any fast evolved genes. Hence, the lack of high scores in the CFG analysis for *PCDH7*, which was strongly proposed in the gene mapping of musical aptitude (I), or *UGT8* that has been proposed elsewhere (Park et al., 2012), may not indicate their non-existent impact on musical abilities. They may have evolved faster in humans, there can be human-specific features, or these genes may have different functions in the songbirds that would result in low scores in the analysis.

It has been shown that correlation on one molecular level may not predict evidence on another level (Bauernfeind et al., 2015). For example, some gene expression changes, detected at the RNA level, do not necessarily predict the protein level changes. In the CFG method, all evidence levels from DNA to proteins were considered. As such, the method ranks those genes highest that have evidence in multiple levels. Because brain tissue is mostly unavailable in humans, the direct brain evidence from the protein and RNA levels comes solely from animal studies. To cover these molecular levels, the use of animal models is needed even though there are concerns.

## 6.5 Biological functions related to musical traits

### 6.5.1 Cognition

The regions identified in the gene mapping studies showed enrichment of psychiatric diseases and memory-related pathways. As indicated in previous studies, part of the genetic predisposition for musical abilities is shared with cognitive abilities like intelligence (Mosing, Pedersen, et al., 2014; Mosing et al., 2016). Also, memory and other cognition-related traits are important in musical abilities. Thus, it was no surprise that the CFG-ranked genes showed enrichment of overall cognition related genes.

Moreover, enrichment of psychiatric diseases can also suggest a more direct link between the studied phenotypes and psychiatric diseases, as has been suggested between creativity and psychiatric diseases (Kyaga et al., 2013; Power et al., 2015). For example, dopamine and serotonin related genes have been associated with both psychiatric diseases and artistic creativity (Carson, 2011). In the musical aptitude

study, these psychiatric disease-related genes included *PCDH7* and multiple GABA A receptors. Composing and arranging can be considered as creative tasks and thus part of creativity. Additionally, music practice and listening have both been shown to induce brain plasticity (Herholz & Zatorre, 2012). Similarly, it has been suggested that some creativity related alleles may cause cognitive flexibility that may lead to psychiatric disorders in adverse environments (Keri, 2009; Hall et al., 2015; Power et al., 2015). Moreover, music has been used as therapy in many psychiatric conditions. These observations are in line with the suggested link between music-related phenotypes and psychiatric diseases.

Overall, many of the psychiatric diseases include impairment of cognitive abilities of which a part are also musical abilities (Peretz, 2006; Millan et al., 2012). Thus, it may be impossible to separate any of these cognition-related genetic signals for one particular phenotype. Some pathways may be more important, for example, in musical abilities than in other cognitive abilities. Also, the same variants under a different environment may develop into different outcomes.

### 6.5.2 Auditory pathways and the inner ear

The enrichment analyses of the genes identified in the publications I and III also showed hearing related pathways. This fits the assumptions that musicality comprises auditory capacities. Sounding music is essentially sound stimuli transmitted from the cochlea through the auditory pathway into the auditory cortex. Part of the pitch processing is performed in the inferior colliculus, which makes it an area of interest, in addition to the inner ear. For example, *GATA2* affects the development of both, the inferior colliculus and the cochlea. The ability to detect these stimuli is necessary in music perception. Especially because we used perceptual tests, it is most likely that hearing related genetic variants affect the phenotype. Expectations were also met in the creative activities gene mapping where there was no enrichment of hearing related genes.

The CFG analysis showed much less inner ear or hearing related genes than the gene mapping studies (I, II, III). Most of the studies used in CFG were zebra finch studies using the song nuclei of the brain. Thus, it is not surprising that the hearing related genes are not commonly found.

The most significant hearing-related function in the enrichment analyses of gene mapping studies was inner ear development (I and III). There were seven genes, for example, *PHOX2B* and *CHRNA9*, related to this function in the musical aptitude linked region and 12 genes in positively selected regions including, for example, *LGR5*, *ADGRV1* and *USH2A*.

### 6.5.3 Speech, music and birdsong

There were many speech-related genes like *FOXP2* and *VLDLR* among the top results in CFG (IV) and all three gene mapping studies (I, II, III). To date, most researchers have thought that birdsong is a model for social communication more so than for singing or other music-related concepts in humans (Scharff & White,

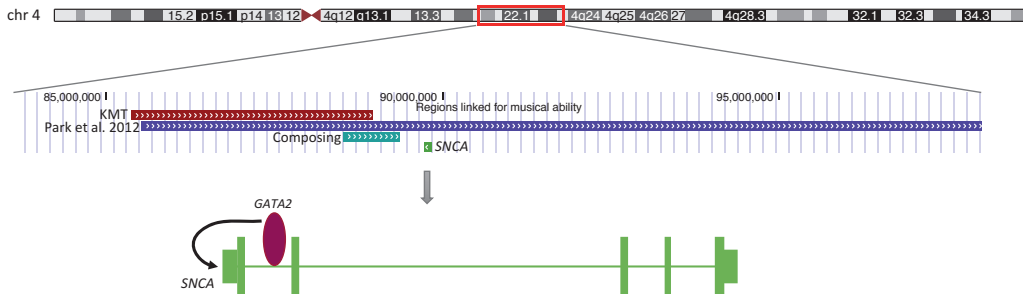
2004; Simonyan et al., 2012). In CFG, the proportion of studies using unclearly music-related phenotypes remained large. Some genes like *EGR1* may in fact relate more to complex social structures and social communication instead of plain auditory stimuli (Avey et al., 2005; Schubloom & Woolley, 2016). However, birdsong may relate partially to both speech and music (Hauser & McDermott, 2003; Rothenberg et al., 2014). In human studies, it has been shown that musically-trained individuals perform better in speech discrimination (Schellenberg, 2015). Also, *FOXP2*-related genetic mutation has been found in a family with language and rhythm impairment (Alcock et al., 2000; Lai et al., 2001). Similarly, the *GRIN2B* gene that was found in the selection signature region (III) and ranked high in the CFG analysis relates to both producing and listening to music and has been shown to affect dyslexia (Mascheretti et al., 2015).

Another common aspect of speech and music is hearing and perception. The processing of music compared to speech is different at the cortical level (Zatorre & Baum, 2012). In spite of this, musicians perform better in many speech-related tasks, like understanding speech in noise (Slater et al., 2015). Studies on songbirds have shown that conspecific song affects the birds more than other auditory stimuli, although other auditory stimuli also have some effect (Velho et al., 2005). Thus, the genes and pathways may be common but the strength and brain region of the expressions may be different. There are examples where speech has been impaired by brain injury when singing remained intact (Peretz, 2006). Thus, there is at least structural differences in the brain between music practice and speech. It is not known whether the speech related genes, like *FOXP2*, are differentially expressed in different brain regions during singing and speaking. Overall, similar features and genes are needed in music and in speech but their relationship in evolution and the brain structures associated with these abilities remain unknown.

## 6.6 Pathways characteristic to musical abilities

The *SNCA* gene, upregulated by listening and performing music (Kanduri, Kuusi, et al., 2015; Kanduri, Raijas, et al., 2015) and affecting the dopaminergic pathway, is located on chromosome 4q22. The *GATA2* gene that was best-associated gene in musical aptitude gene mapping regulates expression of *SNCA* (Figure 11) (Scherzer et al., 2008). *GATA2* is abundantly expressed in dopaminergic neurons (Willett & Greene, 2011) whereas dopaminergic systems are important in music perception in humans (Salimpoor et al., 2011; Zatorre & Salimpoor, 2013) and in vocal learning in songbirds (Simonyan et al., 2012; Hoffmann et al., 2016). Thus, the *GATA2-SNCA* connection seems plausible to affect musical ability.

The top-ranked genes from the CFG analysis showed enrichment of excitation of neurons and LTP. Similarly, cerebellar LTD was enriched in genes in composing-related genomic regions. These long-term memory pathways are important in learning and memory, both important in musical abilities and overall cognitive skills (Collingridge et al., 2010). These pathways enable strengthening or weakening of



**Figure 11. *GATA2* regulates *SNCA*.** *GATA2* was associated in musical aptitude gene mapping (1). *SNCA* which was highly ranked in CFG analysis binds to intron-1 of *SNCA* to regulate its expression. Figure updated from Kanduri, Rajjas, et al. (2015) with permission.

neuronal activation in response to stimuli. The excitatory response of neurons is a part of synaptic transmission whose basic functions are excitation or inhibition.

Network analysis of the 40 top-ranked genes showed four non-overlapping networks (see Table 8). The most prominent network of these four was related to behaviour. These networks comprised most of the top-ranked genes. An interesting part of the networks were the connectors (hubs) that were added to connect the top-ranked genes. For example, adrenergic receptors were not highly ranked but they were predicted to affect other highly ranked genes within the behaviour-related network. Similarly, growth hormone was included in the post-transcriptional modification related network, but it has only been suggested in one music listening study (Gerra et al., 1998). The hearing-related network includes three miRNAs of interest: miR-30c-5p, miR-6967-5p and miR-519a-3p that are the major hubs in the network. Not many studies have even studied any miRNAs and future studies will show if they actually have function in music perception.

The networks and pathways discussed here may work as suggestions on important pathways that can mediate the effects of the candidate genes in music. However, their true importance can only be studied by functional analyses where the genes and proteins also relate to specific brain, inner ear or other tissue structures. In these analyses, all interaction information from any cell lines were used. Thus, the real networks in the tissues relevant to music may work differently.



## 7 Conclusions and future prospects

The study represents pioneering research on biological processes underlying musical abilities like perceptual skills and creativity in music. The findings provide novel molecular evidence that genes related to inner ear development, dopaminergic systems, cognition and memory are potentially interconnected with musical abilities in humans. The biological pathways identified in this project may not only explain the molecular mechanisms underlying the musical abilities in humans, but also serve the purpose of comparisons between other cognitive traits both in normal (for example language and intelligence) and diseased conditions (for example schizophrenia and dementia). The findings can be utilised widely, including gene-environment interaction studies of music and studies on the effect of music as therapy.

In this thesis, I suggest that singing in songbirds can be used as a model for musical traits. Some prior studies have suggested this approach, but songbirds have not been commonly utilised in music studies previously. The gathered evidence points to common genes between human musical traits and song related phenotypes in birds. Developing on these animal models, I suggest cognition-related networks and pathways to be important in musical traits.

Musical abilities relate to auditory processing; fine-tuned hearing is important for musicians. Hence, it was not surprising that the results obtained in this work indicate some common genetic pathways between musical abilities and hearing. The results also give further support for a shared evolutionary background of auditory processing between vocalising birds and humans.

Music-specific pathways and networks have previously been suggested in brain studies. The evidence in this work mostly suggests that genes related to general cognition and hearing affect music practice and perception. No music-specific molecular pathways or genes were identified. The CFG method was applied to prioritise sensitivity for the molecules whereas specificity for music was weakened. Overall, more studies would be needed to verify the findings suggested by the CFG method, and to improve specificity. Animal models designed for studying music perception and practice could bring more information on the evolution of music and the specific molecular background for music, especially in the brain and inner ear. Whether the music-specificity observed in brain studies results from the common genetic pathways functioning on music-specific brain regions, or from music-specific molecular pathways, remains unsolved.

As musicians have always noted, there are differences between musicians. Some are good at improvising while others are better at learning new pieces. The findings here provide evidence of genetic predisposition affecting the differences between musically experienced individuals: the drive to arrange or compose. The relatively large role of inherited component in this kind of complex behaviour was not expected. Even though there are genetic factors affecting the musical abilities, individual's future expertise cannot be predicted from the genetic information.



However, the genetic studies about musical abilities may help individuals to cope with their innate differences; for example, knowledge about genetic background of amusia can change the way amusics are considered in schools.

Future studies should further assess the importance of the findings. Next generation sequencing could be used to identify possible causative variation in the suggested regions, for example, at chromosome 4q sequencing could reveal genomic variants explaining the detected linkage. At this thesis, some aspects of the genetics of music are perceived but a great deal requires further clarification.

## 8 Acknowledgements

In 2009 I was looking for a place to do my bachelor's thesis. After Leila Kauppinen's kind suggestion, I ended up at Päivi Onkamo's office and started to work in the music project that is run by Irma Järvelä. It did not take long until someone mentioned doctoral thesis. So I stayed and now, seven years later, I have this thesis in my hand.

I want to thank everyone involved in this process. First of all, I want to thank my supervisors Päivi Onkamo and Irma Järvelä. Thank you Päivi for being an inspiring teacher and for all the optimism and realism that has helped during this journey. You have always asked difficult, sometimes impossible, questions and helped me develop. Thank you Irma for all the support during this project and during difficult times of my life, especially thanks for all the help in the writing processes throughout the project. Your enormous knowledge about genes and passion towards research has always amazed me.

I would like to thank Pekka Uimari and Fredrik Ullén for the valuable suggestions and comments on this thesis. I am grateful for Suzanne Leal for accepting to act as an opponent.

I want to thank my colleagues Liisa Ukkola-Vuoti and Chakravarthi Kanduri. Liisa guided my first years in the MUSGEN project and made me feel welcome. I am grateful to Chakri especially for driving the convergent genomics article. My gratitude goes also to everyone else involved in the music project: music experts Pirre Raijas, Kai Karma and Tuire Kuusi who all gave valuable comments whenever I misused the music terms or concepts, statistics master Petri Peltonen who gathered new data and project organizer Minna Ahvenainen. Special thanks to Minna for language revision.

I am also grateful to other people who have worked with me in Biomedicum, in the Department of Medical Genetics. Especially thanks for good company to office mates Katri Kantojärvi, Anju Phillips, Fatma Doagu, Preethy Nair, Paula Salo, Emmi Joensuu, Heidi Marjonen and Annalaura Jokiniemi. I also want to thank people in Viikki: former and current office mates Kerttu Majander, Sanni Översti, Elina Salmela, Martin Heger, AP Junno and all the people at the old seventh floor. Thanks for the funny (maybe loud) discussions, bad jokes, shared complaints, wine and whiskey, travel companion, good advice and listening. Thanks also for scientific help and discussions. You all made my days during these years! Additionally, thanks to Osmo Hakosalo for all the help with the gene mapping methodology and Daniel Borshagovski for the *GATA2* information.

I want to express my gratitude to the Academy of Finland, the Finnish Brain Foundation, the Finnish Concordia Fund and the University of Helsinki for funding my thesis.

I also would like to thank my family for support during these years: dear sisters, brothers and father and all the people in our extended family who share ups and

downs, worries and laughs with me. Huge thanks for patient Otto and kind (and funny) Meira, you are my everything. Thanks to Sami and other people at the pottery for shifting my thoughts from the thesis and all the laughs. And sorry for all friends who I have been ignoring when concentrating on this thesis.

Finally, I sincerely want to thank all the participants of the MUSGEN study.

*Helsinki, September 2016*

A handwritten signature in black ink, appearing to be 'Jana', with a large loop at the beginning and a horizontal line extending to the right.

## 9 References

- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*, *30*, 97-101. doi:10.1038/ng786
- Ahmed, Z. M., Riazuddin, S., Bernstein, S. L., Ahmed, Z., Khan, S., Griffith, A. J., . . . Wilcox, E. R. (2001). Mutations of the protocadherin gene PCDH15 cause Usher syndrome type 1F. *Am J Hum Genet*, *69*, 25-34. doi:10.1086/321277
- Akey, J. M. (2009). Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res*, *19*, 711-722. doi:10.1101/gr.086652.108
- Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain Lang*, *75*, 34-46. doi:10.1006/brln.2000.2323
- Almasy, L., & Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*, *62*, 1198-1211. doi:10.1086/301844
- Almasy, L., Kos, M., & Blangero, J. (2015). Linkage Mapping: Localizing the Genes That Shape Human Variation. In R. Duggirala, L. Almasy, S. Williams-Blangero, S. F. D. Paul, & C. Kole (Eds.), *Genome Mapping and Genomics in Human and Non-Human Primates* (Vol. 5, pp. 33-52): Springer Berlin Heidelberg.
- Alpert, M. I., Alpert, J. I., & Maltz, E. N. (2005). Purchase occasion influence on the role of music in advertising. *J Bus Res*, *58*, 369-376. doi:10.1016/S1048-2963(03)00101-2
- Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet*, *54*, 535-543.
- Angelucci, F., Ricci, E., Padua, L., Sabino, A., & Tonali, P. A. (2007). Music exposure differentially alters the levels of brain-derived neurotrophic factor and nerve growth factor in the mouse hypothalamus. *Neurosci Lett*, *429*, 152-155. doi:10.1016/j.neulet.2007.10.005
- Arnauld, E., Jeantet, Y., Arsaut, J., & Demotes-Mainard, J. (1996). Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Brain Res Mol Brain Res*, *41*, 27-35.
- Aulchenko, Y. S., de Koning, D. J., & Haley, C. (2007). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, *177*, 577-585. doi:10.1534/genetics.107.075614
- Aulchenko, Y. S., Ripke, S., Isaacs, A., & van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, *23*, 1294-1296. doi:10.1093/bioinformatics/btm108
- Avey, M. T., Kanyo, R. A., Irwin, E. L., & Sturdy, C. B. (2008). Differential effects of vocalization type, singer and listener on ZENK immediate early gene response in black-capped chickadees (*Poecile atricapillus*). *Behav Brain Res*, *188*, 201-208. doi:10.1016/j.bbr.2007.10.034
- Avey, M. T., Phillmore, L. S., & MacDougall-Shackleton, S. A. (2005). Immediate early gene expression following exposure to acoustic and visual components of courtship in zebra finches. *Behav Brain Res*, *165*, 247-253. doi:10.1016/j.bbr.2005.07.002
- Bachner-Melman, R., Dina, C., Zohar, A. H., Constantini, N., Lerer, E., Hoch, S., . . . Ebstein, R. P. (2005). AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS Genet*, *1*, e42. doi:10.1371/journal.pgen.0010042
- Baer, L. H., Park, M. T., Bailey, J. A., Chakravarty, M. M., Li, K. Z., & Penhune, V. B. (2015). Regional cerebellar volumes are related to early musical training and finger tapping performance. *Neuroimage*, *109*, 130-139. doi:10.1016/j.neuroimage.2014.12.076
- Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *Am J Hum Genet*, *67*, 755-758. doi:10.1086/303057
- Bailey, J. A., & Penhune, V. B. (2013). The relationship between the age of onset of musical training and rhythm synchronization performance: validation of sensitive period effects. *Front Neurosci*, *7*, 227. doi:10.3389/fnins.2013.00227

- Bassett, A. S., Chow, E. W., Husted, J., Weksberg, R., Caluseriu, O., Webb, G. D., & Gatzoulis, M. A. (2005). Clinical features of 78 adults with 22q11 Deletion Syndrome. *Am J Med Genet A*, *138*, 307-313. doi:10.1002/ajmg.a.30984
- Bauernfeind, A. L., Soderblom, E. J., Turner, M. E., Moseley, M. A., Ely, J. J., Hof, P. R., . . . Babbitt, C. C. (2015). Evolutionary Divergence of Gene and Protein Expression in the Brains of Humans and Chimpanzees. *Genome Biol Evol*, *7*, 2276-2288. doi:10.1093/gbe/evv132
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. 1763. *MD Comput*, *8*, 157-171.
- Bertsch, B., Ogden, C. A., Sidhu, K., Le-Niculescu, H., Kuczynski, R., & Niculescu, A. B. (2005). Convergent functional genomics: a Bayesian candidate gene identification approach for complex disorders. *Methods*, *37*, 274-279. doi:10.1016/j.ymeth.2005.03.012
- Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *P Natl Acad Sci U S A*, *98*, 11818-11823. doi:10.1073/pnas.191355898
- Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci*, *11*, 747-759. doi:10.1038/nrn2931
- Boonstra, P. S., Mukherjee, B., Gruber, S. B., Ahn, J., Schmit, S. L., & Chatterjee, N. (2016). Tests for Gene-Environment Interactions and Joint Effects With Exposure Misclassification. *Am J Epidemiol*, *183*, 237-247. doi:10.1093/aje/kwv198
- Brown, S., Martinez, M. J., & Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport*, *15*, 2033-2037.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*, *81*, 1084-1097. doi:10.1086/521987
- Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex*, *57*, 254-269. doi:10.1016/j.cortex.2014.04.012
- Burunat, I., Brattico, E., Puolivali, T., Ristaniemi, T., Sams, M., & Toiviainen, P. (2015). Action in Perception: Prominent Visuo-Motor Functional Symmetry in Musicians during Music Listening. *PLoS one*, *10*, e0138238. doi:10.1371/journal.pone.0138238
- Carson, S. H. (2011). Creativity and psychopathology: a shared vulnerability model. *Can J Psychiatry*, *56*, 144-153.
- Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn Sci*, *17*, 179-193. doi:10.1016/j.tics.2013.02.007
- Chase, A. R. (2001). Music discriminations by carp (*Cyprinus carpio*). *Anim Learn Behav*, *29*, 336-353. doi:10.3758/BF03192900
- Chikahisa, S., Sei, H., Morishima, M., Sano, A., Kitaoka, K., Nakaya, Y., & Morita, Y. (2006). Exposure to music in the perinatal period enhances learning performance and alters BDNF/TrkB signaling in mice as adults. *Behav Brain Res*, *169*, 312-319. doi:10.1016/j.bbr.2006.01.021
- Clayton, D. F. (2013). The genomics of memory and learning in songbirds. *Annu Rev Genomics Hum Genet*, *14*, 45-65. doi:10.1146/annurev-genom-090711-163809
- Collingridge, G. L., Peineau, S., Howland, J. G., & Wang, Y. T. (2010). Long-term depression in the CNS. *Nat Rev Neurosci*, *11*, 459-473. doi:10.1038/nrn2867
- Coon, H., & Carey, G. (1989). Genetic and environmental determinants of musical ability in twins. *Behav Genet*, *19*, 183-193.
- Cravchik, A., Subramanian, G., Broder, S., & Venter, J. C. (2001). Sequence analysis of the human genome: implications for the understanding of nervous system function and disease. *Arch Neurol*, *58*, 1772-1778.
- Craxton, M. (2004). Synaptotagmin gene content of the sequenced genomes. *BMC Genomics*, *5*, 43. doi:10.1186/1471-2164-5-43
- Cummings, A. C., Torstenson, E., Davis, M. F., D'Aouost, L. N., Scott, W. K., Pericak-Vance, M. A., . . . Haines, J. L. (2013). Evaluating power and type 1 error in large pedigree analyses of binary traits. *PLoS one*, *8*, e62615. doi:10.1371/journal.pone.0062615

- Dai, J. Y., Kooperberg, C., Leblanc, M., & Prentice, R. L. (2012). Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*, *99*, 929-944. doi:10.1093/biomet/ass044
- Dastgheib, S. S., Layegh, P., Sadeghi, R., Foroughipur, M., Shoebibi, A., & Gorji, A. (2014). The Effects of Mozart's Music on Interictal Activity in Epileptic Patients: Systematic Review and Meta-analysis of the Literature. *Curr Neurol Neurosci*, *14*, 420. doi:10.1007/s11910-013-0420-x
- Davis, J. M., Searles, V. B., Anderson, N., Keeney, J., Raznahan, A., Horwood, L. J., . . . Sikela, J. M. (2015). DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores. *Hum Genet*, *134*, 67-75. doi:10.1007/s00439-014-1489-2
- Davis, S., Renaudineau, S., Poirier, R., Poucet, B., Save, E., & Laroche, S. (2010). The formation and stability of recognition memory: what happens upon recall? *Front Behav Neurosci*, *4*, 177. doi:10.3389/fnbeh.2010.00177
- de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*, *367*, 1921-1929. doi:10.1056/NEJMoa1206524
- de Quervain, D. J., & Papassotiropoulos, A. (2006). Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *P Natl Acad Sci U S A*, *103*, 4270-4274. doi:10.1073/pnas.0510212103
- de Quervain, D. J., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nat Neurosci*, *3*, 313-314. doi:10.1038/73873
- Debette, S., Ibrahim Verbaas, C. A., Bressler, J., Schuur, M., Smith, A., Bis, J. C., . . . Aging Research in Genomic Epidemiology, C. (2015). Genome-wide studies of verbal declarative memory in nondemented older people: the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. *Biol Psychiatry*, *77*, 749-763. doi:10.1016/j.biopsych.2014.08.027
- Decker, E. L., Skerka, C., & Zipfel, P. F. (1998). The early growth response protein (EGR-1) regulates interleukin-2 transcription by synergistic interaction with the nuclear factor of activated T cells. *J Biol Chem*, *273*, 26923-26930.
- Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency. *J Acoust Soc Am*, *125*, 2398-2403. doi:10.1121/1.3081389
- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol Bull*, *136*, 822-848. doi:10.1037/a0019749
- Dong, S., & Clayton, D. F. (2008). Partial dissociation of molecular and behavioral measures of song habituation in adult zebra finches. *Genes Brain Behav*, *7*, 802-809. doi:10.1111/j.1601-183X.2008.00423.x
- Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*, 1969-1972. doi:10.1126/science.291.5510.1969
- Drnevich, J., Replogle, K. L., Lovell, P., Hahn, T. P., Johnson, F., Mast, T. G., . . . Clayton, D. F. (2012). Impact of experience-dependent and -independent factors on gene expression in songbird brain. *P Natl Acad Sci USA*, *109 Suppl 2*, 17245-17252. doi:10.1073/pnas.1200655109
- Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*, *4*, 1184-1191. doi:10.1038/nprot.2009.97
- Earp, S. E., & Maney, D. L. (2012). Birdsong: is it music to their ears? *Front Evol Neurosci*, *4*, 14. doi:10.3389/fnevo.2012.00014
- Eda-Fujiwara, H., Satoh, R., Bolhuis, J. J., & Kimura, T. (2003). Neuronal activation in female budgerigars is localized and related to male song complexity. *Eur J Neurosci*, *17*, 149-154. doi:10.1046/j.1460-9568.2003.02414.x

- Elston, R. C., & Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Hum Hered*, *21*, 523-542.
- Emanuele, E., Boso, M., Cassola, F., Broglia, D., Bonoldi, I., Mancini, L., . . . Politi, P. (2010). Increased dopamine DRD4 receptor mRNA expression in lymphocytes of musicians and autistic individuals: bridging the music-autism connection. *Neuro Endocrinol Lett*, *31*, 122-125.
- Encode Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57-74. doi:10.1038/nature11247
- Ewens, W. J., & Spielman, R. S. (1995). The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet*, *57*, 455-464.
- Fancourt, D., Ockelford, A., & Belai, A. (2014). The psychoneuroimmunological effects of music: a systematic review and a new model. *Brain Behav Immun*, *36*, 15-26. doi:10.1016/j.bbi.2013.10.014
- Fancourt, D., & Williamson, A. (2016). Attending a concert reduces glucocorticoids, progesterone and the cortisol/DHEA ratio. *Public Health*, *132*, 101-104. doi:10.1016/j.puhe.2015.12.005
- Fisher, S. E., & Ridley, M. (2013). Evolution. Culture, genes, and the human revolution. *Science*, *340*, 929-930. doi:10.1126/science.1236171
- Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends Genet*, *25*, 166-177. doi:10.1016/j.tig.2009.03.002
- Forstmeier, W., Burger, C., Temnow, K., & Deregnacourt, S. (2009). The genetic basis of zebra finch vocalizations. *Evolution*, *63*, 2114-2130. doi:10.1111/j.1558-5646.2009.00688.x
- Freedman, M. L., Monteiro, A. N., Gayther, S. A., Coetzee, G. A., Risch, A., Plass, C., . . . Mills, I. G. (2011). Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*, *43*, 513-518. doi:10.1038/ng.840
- Friend, R. S. (1939). Influences of heredity and musical environment on the scores of kindergarten children on the Seashore measures of musical ability. *Journal of Applied Psychology*, *23*, 347-357. doi:10.1037/h0061211
- Fry, D. B. (1948). An experimental study of tone deafness. *Speech*, *1*, 4-11.
- Fukui, H., & Toyoshima, K. (2013). Influence of music on steroid hormones and the relationship between receptor polymorphisms and musical ability: a pilot study. *Front Psychol*, *4*, 910. doi:10.3389/fpsyg.2013.00910
- Fukui, H., & Toyoshima, K. (2014). Chill-inducing music enhances altruism in humans. *Front Psychol*, *5*, 1215. doi:10.3389/fpsyg.2014.01215
- Fukui, H., & Yamashita, M. (2003). The effects of music and visual stress on testosterone and cortisol in men and women. *Neuro Endocrinol Lett*, *24*, 173-180.
- Gerra, G., Zaimovic, A., Franchini, D., Palladino, M., Giucastro, G., Reali, N., . . . Brambilla, F. (1998). Neuroendocrine responses of healthy volunteers to 'techno-music': relationships with personality traits and emotional state. *Int J Psychophysiol*, *28*, 99-111.
- Gertz, E. M., Hiekkalinna, T., Digabel, S. L., Audet, C., Terwilliger, J. D., & Schäffer, A. A. (2014). PSEUDOMARKER 2.0: efficient computation of likelihoods using NOMAD. *BMC bioinformatics*, *15*, 47. doi:10.1186/1471-2105-15-47
- Gervain, J., Vines, B. W., Chen, L. M., Seo, R. J., Hensch, T. K., Werker, J. F., & Young, A. H. (2013). Valproate reopens critical-period learning of absolute pitch. *Front Syst Neurosci*, *7*, 102. doi:10.3389/fnsys.2013.00102
- Gordon, E. E. (1979). Primary measures of music audiation. Chicago, USA: GIA Publications.
- Gordon, E. E. (2007). *Learning Sequences in Music: A Contemporary Music Learning Theory*. Chicago, USA: GIA Publications.
- Gourmaud, S., Paquet, C., Dumurgier, J., Pace, C., Bouras, C., Gray, F., . . . Hugon, J. (2015). Increased levels of cerebrospinal fluid JNK3 associated with amyloid pathology: links to cognitive decline. *J Psychiatry Neurosci*, *40*, 151-161.
- Granot, R. Y., Frankel, Y., Gritsenko, V., Lerer, E., Gritsenko, I., Bachner-Melman, R., . . . Ebstein, R. P. (2007). Provisional evidence that the arginine vasopressin 1a receptor gene is associated with musical memory. *Evol Hum Behav*, *28*, 313-318.



- Granot, R. Y., Uzefovsky, F., Bogopolsky, H., & Ebstein, R. P. (2013). Effects of arginine vasopressin on musical working memory. *Front Psychol*, *4*, 712. doi:10.3389/fpsyg.2013.00712
- Grass, J. A., Jing, H., Kim, S. I., Martowicz, M. L., Pal, S., Blobel, G. A., & Bresnick, E. H. (2006). Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. *Mol Cell Biol*, *26*, 7056-7067. doi:10.1128/MCB.01033-06
- Gray, K. A., Yates, B., Seal, R. L., Wright, M. W., & Bruford, E. A. (2015). Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res*, *43*, D1079-1085. doi:10.1093/nar/gku1071
- Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (1999). Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component. *Am J Hum Genet*, *65*, 911-913. doi:10.1086/302541
- Gregersen, P. K., Kowalsky, E., Lee, A., Baron-Cohen, S., Fisher, S. E., Asher, J. E., . . . Li, W. (2013). Absolute pitch exhibits phenotypic and genetic overlap with synesthesia. *Hum Mol Genet*, *22*, 2097-2104. doi:10.1093/hmg/ddt059
- Grillet, N., Schwander, M., Hildebrand, M. S., Sczaniecka, A., Kolatkar, A., Velasco, J., . . . Muller, U. (2009). Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. *Am J Hum Genet*, *85*, 328-337. doi:10.1016/j.ajhg.2009.07.017
- Hall, J., Trent, S., Thomas, K. L., O'Donovan, M. C., & Owen, M. J. (2015). Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. *Biol Psychiatry*, *77*, 52-58. doi:10.1016/j.biopsych.2014.07.011
- Hallam, S., Rogers, L., & Creech, A. (2008). Gender differences in musical instrument choice. *Int J Music Educ*, *26*, 7-19. doi:10.1177/0255761407085646
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, *22*, 1760-1774. doi:10.1101/gr.135350.111
- Haugas, M., Lillevali, K., Hakanen, J., & Salminen, M. (2010). Gata2 is required for the development of inner ear semicircular ducts and the surrounding perilymphatic space. *Dev Dyn*, *239*, 2452-2469. doi:10.1002/dvdy.22373
- Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat Neurosci*, *6*, 663-668. doi:10.1038/nn1080
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., . . . Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, *39*, 311-318. doi:10.1038/ng1966
- Heller, K. A., Perleth, C., & Lim, T. K. (2005). The Munich model of giftedness designed to identify and promote gifted students. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (Vol. 2, pp. 147-170). New York, USA: Cambridge University Press.
- Henry, M. J., & McAuley, J. D. (2010). On the Prevalence of Congenital Amusia. *Music Percept*, *27*, 413-418. doi:10.1525/mp.2010.27.5.413
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*, 486-502. doi:10.1016/j.neuron.2012.10.011
- Hertel, N., Redies, C., & Medina, L. (2012). Cadherin expression delineates the divisions of the postnatal and adult mouse amygdala. *J Comp Neurol*, *520*, 3982-4012. doi:10.1002/cne.23140
- Heston, J. B., & White, S. A. (2015). Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *J Neurosci*, *35*, 2885-2894. doi:10.1523/JNEUROSCI.3715-14.2015
- Hiekkalinna, T., Goring, H. H. H., Lambert, B., Weiss, K. M., Norrgrann, P., Schaffer, A. A., & Terwilliger, J. D. (2012). On the statistical properties of family-based association tests in datasets containing both pedigrees and unrelated case-control samples. *European Journal of Human Genetics*, *20*, 217-223. doi:10.1038/ejhg.2011.173
- Hiekkalinna, T., Schäffer, A. A., Lambert, B., Norrgrann, P., Göring, H. H., & Terwilliger, J. D. (2011). PSEUDOMARKER: a powerful program for joint linkage and/or linkage disequilibrium analysis on mixtures of singletons and related individuals. *Hum Hered*, *71*, 256-266. doi:10.1159/000329467

- Hilliard, A. T., Miller, J. E., Fraley, E. R., Horvath, S., & White, S. A. (2012). Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron*, *73*, 537-552. doi:10.1016/j.neuron.2012.01.005
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *P Natl Acad Sci U S A*, *106*, 9362-9367. doi:10.1073/pnas.0903103106
- Hoffmann, L. A., Saravanan, V., Wood, A. N., He, L., & Sober, S. J. (2016). Dopaminergic Contributions to Vocal Learning. *J Neurosci*, *36*, 2176-2189. doi:10.1523/JNEUROSCI.3883-15.2016
- Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality Introduction. *Philos T R Soc B*, *370*, 5-12. doi:10.1098/rstb.2014.0088
- Horvath, S., Xu, X., & Laird, N. M. (2001). The family based association test method: strategies for studying general genotype--phenotype associations. *Eur J Hum Genet*, *9*, 301-306. doi:10.1038/sj.ejhg.5200625
- Howe, M. J., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: reality or myth? *Behav Brain Sci*, *21*, 399-407; discussion 407-342.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009a). The effects of musical training on structural brain development: a longitudinal study. *Ann N Y Acad Sci*, *1169*, 182-186. doi:10.1111/j.1749-6632.2009.04852.x
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009b). Musical training shapes structural brain development. *J Neurosci*, *29*, 3019-3025. doi:10.1523/JNEUROSCI.5118-08.2009
- International HapMap, C. (2003). The International HapMap Project. *Nature*, *426*, 789-796. doi:10.1038/nature02168
- Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Ann N Y Acad Sci*, *1016*, 749-777. doi:10.1196/annals.1298.038
- Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., & Nottebohm, F. (1998). For whom the bird sings: context-dependent gene expression. *Neuron*, *21*, 775-788.
- Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *Neuroimage*, *57*, 1572-1579. doi:10.1016/j.neuroimage.2011.05.061
- Kanduri, C., & Järvelä, I. (2016). GenRank: an R/Bioconductor package for prioritization of candidate genes. Retrieved from <http://dx.doi.org/10.1101/048264>
- Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A. K., Lähdesmäki, H., & Järvelä, I. (2015). The effect of music performance on the transcriptome of professional musicians. *Sci Rep*, *5*, 9506. doi:10.1038/srep09506
- Kanduri, C., Raijas, P., Ahvenainen, M., Philips, A. K., Ukkola-Vuoti, L., Lähdesmäki, H., & Järvelä, I. (2015). The effect of listening to music on human transcriptome. *PeerJ*, *3*, e830. doi:10.7717/peerj.830
- Karma, K. (1994). Auditory and visual temporal structuring: How important is sound to musical thinking? *Psychol music*, *22*, 20-30.
- Karma, K. (2007). Musical aptitude definition and measure validation: Ecological validity can endanger the construct validity of musical aptitude tests. *Psychomusicology*, *19*, 79-90. doi:10.1037/h0094033
- Kavanagh, D. H., Tansey, K. E., O'Donovan, M. C., & Owen, M. J. (2015). Schizophrenia genetics: emerging themes for a complex disorder. *Mol Psychiatry*, *20*, 72-76. doi:10.1038/mp.2014.148
- Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, *15*, 246. doi:10.1186/1471-2164-15-246

- Keri, S. (2009). Genes for psychosis and creativity: a promoter polymorphism of the neuregulin 1 gene is related to creativity in people with high intellectual achievement. *Psychol Sci*, *20*, 1070-1073. doi:10.1111/j.1467-9280.2009.02398.x
- Khandekar, M., Brandt, W., Zhou, Y., Dagenais, S., Glover, T. W., Suzuki, N., . . . Engel, J. D. (2007). A Gata2 intronic enhancer confers its pan-endothelia-specific regulation. *Development*, *134*, 1703-1712. doi:10.1242/dev.001297
- Kimpo, R. R., & Doupe, A. J. (1997). FOS is induced by singing in distinct neuronal populations in a motor network. *Neuron*, *18*, 315-325. doi:10.1016/S0896-6273(00)80271-8
- Kobayashi-Osaki, M., Ohneda, O., Suzuki, N., Minegishi, N., Yokomizo, T., Takahashi, S., . . . Yamamoto, M. (2005). GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene. *Mol Cell Biol*, *25*, 7005-7020. doi:10.1128/MCB.25.16.7005-7020.2005
- Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Front Psychol*, *2*, 110. doi:10.3389/fpsyg.2011.00110
- Koelsch, S., Maess, B., Grossmann, T., & Friederici, A. D. (2003). Electric brain responses reveal gender differences in music processing. *Neuroreport*, *14*, 709-713. doi:10.1097/01.wnr.0000065762.60383.67
- Koelsch, S., & Skouras, S. (2014). Functional centrality of amygdala, striatum and hypothalamus in a "small-world" network underlying joy: an fMRI study with music. *Hum Brain Mapp*, *35*, 3485-3498.
- Kong, A., & Cox, N. J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet*, *61*, 1179-1188. doi:10.1086/301592
- Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *J Behav Med*, *27*, 623-635.
- Kruglyak, L., Daly, M. J., Reeve-Daly, M. P., & Lander, E. S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*, *58*, 1347-1363.
- Kurz, A., Wöhr, M., Walter, M., Bonin, M., Auburger, G., Gispert, S., & Schwarting, R. K. (2010). Alpha-synuclein deficiency affects brain Foxp1 expression and ultrasonic vocalization. *Neuroscience*, *166*, 785-795. doi:10.1016/j.neuroscience.2009.12.054
- Kyaga, S., Landen, M., Boman, M., Hultman, C. M., Langstrom, N., & Lichtenstein, P. (2013). Mental illness, suicide and creativity: 40-year prospective total population study. *J Psychiatr Res*, *47*, 83-90. doi:10.1016/j.jpsychires.2012.09.010
- Lahti, L., Achim, K., & Partanen, J. (2013). Molecular regulation of GABAergic neuron differentiation and diversity in the developing midbrain. *Acta Physiol (Oxf)*, *207*, 616-627. doi:10.1111/apha.12062
- Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*, 519-523. doi:10.1038/35097076
- Laird, N. M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet*, *7*, 385-394. doi:10.1038/nrg1839
- Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature*, *470*, 187-197. doi:10.1038/nature09792
- Lander, E. S., & Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *P Natl Acad Sci U S A*, *84*, 2363-2367.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860-921. doi:10.1038/35057062
- Lange, K., Papp, J. C., Sinsheimer, J. S., Sripracha, R., Zhou, H., & Sobel, E. M. (2013). Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics*, *29*, 1568-1570. doi:10.1093/bioinformatics/btt187
- Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*, *95*, 5-23. doi:10.1016/j.ajhg.2014.06.009

- Levitin, D. J. (2006). *This is your brain on music – the science of a human obsession*. Penguin Book, New York.
- Levitin, D. J. (2012). What does it mean to be musical? *Neuron*, 73, 633-637. doi:10.1016/j.neuron.2012.01.017
- Li, M.-X., Yeung, J. Y., Cherny, S., & Sham, P. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet*, 131, 747-756. doi:10.1007/s00439-011-1118-2
- Li, X., Wang, Q., He, K., Li, Z., Chen, J., Li, W., . . . Shi, Y. (2014). Common variants in the CDH7 gene are associated with major depressive disorder in the Han Chinese population. *Behav Genet*, 44, 97-101. doi:10.1007/s10519-014-9645-y
- Liew, C. C., Ma, J., Tang, H. C., Zheng, R., & Dempsey, A. A. (2006). The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med*, 147, 126-132. doi:10.1016/j.lab.2005.10.005
- Lin, J., Yan, X., Wang, C., Guo, Z., Rolfs, A., & Luo, J. (2012). Anatomical expression patterns of delta-protocadherins in developing chicken cochlea. *J Anat*, 221, 598-608. doi:10.1111/j.1469-7580.2012.01568.x
- Liu, X., Bulgakov, O. V., Darrow, K. N., Pawlyk, B., Adamian, M., Liberman, M. C., & Li, T. (2007). Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells. *P Natl Acad Sci U S A*, 104, 4413-4418. doi:10.1073/pnas.0610950104
- Liu, X., Ong, R. T., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., . . . Teo, Y. Y. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *Am J Hum Genet*, 92, 866-881. doi:10.1016/j.ajhg.2013.04.021
- Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *P Natl Acad Sci U S A*, 107, 961-968. doi:10.1073/pnas.0912629107
- Madsen, P., & Jensen, J. (2013). A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models (Version 6). Tjele, Denmark: Danish Institute of Agricultural Sciences.
- Mangiamele, L. A., & Burmeister, S. S. (2008). Acoustically evoked immediate early gene expression in the pallium of female tungara frogs. *Brain Behav Evol*, 72, 239-250. doi:10.1159/000171481
- Mascheretti, S., Facoetti, A., Giorda, R., Beri, S., Riva, V., Trezzi, V., . . . Marino, C. (2015). GRIN2B mediates susceptibility to intelligence quotient and cognitive impairments in developmental dyslexia. *Psychiatr Genet*, 25, 9-20. doi:10.1097/YPG.0000000000000068
- McCraty, R., Atkinson, M., Rein, G., & Watkins, A. D. (1996). Music enhances the effect of positive emotional states on salivary IgA. *Stress Medicine*, 12, 167-175.
- McGee, J., Goodyear, R. J., McMillan, D. R., Stauffer, E. A., Holt, J. R., Locke, K. G., . . . Richardson, G. P. (2006). The very large G-protein-coupled receptor VLGR1: a component of the ankle link complex required for the normal development of auditory hair bundles. *J Neurosci*, 26, 6543-6553. doi:10.1523/JNEUROSCI.0693-06.2006
- Mello, C. V., Vicario, D. S., & Clayton, D. F. (1992). Song presentation induces gene expression in the songbird forebrain. *P Natl Acad Sci U S A*, 89, 6818-6822.
- Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brunn*, IV, 3-47.
- Meng, B., Zhu, S., Li, S., Zeng, Q., & Mei, B. (2009). Global view of the mechanisms of improved learning and memory capability in mice with music-exposure by microarray. *Brain Res Bull*, 80, 36-44. doi:10.1016/j.brainresbull.2009.05.020
- Merriam, A. P. (1964). *The Anthropology of Music*. Evanston, US: Northwestern University Press.
- Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., . . . Kent, W. J. (2013). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*, 41, D64-69. doi:10.1093/nar/gks1048
- Millan, M. J., Agid, Y., Brune, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., . . . Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov*, 11, 141-168. doi:10.1038/nrd3628

- Mockel, M., Rocker, L., Stork, T., Vollert, J., Danne, O., Eichstadt, H., . . . Hochrein, H. (1994). Immediate physiological responses of healthy volunteers to different types of music: cardiovascular, hormonal and mental changes. *Eur J Appl Physiol Occup Physiol*, *68*, 451-459.
- Montealegre-Z, F., Jonsson, T., Robson-Brown, K. A., Postles, M., & Robert, D. (2012). Convergent evolution between insect and mammalian audition. *Science*, *338*, 968-971. doi:10.1126/science.1225271
- Mori, C., & Wada, K. (2015). Audition-independent vocal crystallization associated with intrinsic developmental gene expression dynamics. *J Neurosci*, *35*, 878-889. doi:10.1523/JNEUROSCI.1804-14.2015
- Morley, A. P., Narayanan, M., Mines, R., Molokhia, A., Baxter, S., Craig, G., . . . Craig, I. (2012). AVPR1A and SLC6A4 polymorphisms in choral singers and non-musicians: a gene association study. *PLoS one*, *7*, e31763. doi:10.1371/journal.pone.0031763
- Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullen, F. (2014). Practice does not make perfect: no causal effect of music practice on music ability. *Psychol Sci*, *25*, 1795-1803. doi:10.1177/0956797614541990
- Mosing, M. A., Pedersen, N. L., Madison, G., & Ullen, F. (2014). Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PLoS one*, *9*, e113874. doi:10.1371/journal.pone.0113874
- Mosing, M. A., Verweij, K. J. H., Madison, G., & Ullén, F. (2016). The genetic architecture of correlations between perceptual timing, motor timing, and intelligence. *Intelligence*, *57*, 33-40. doi:10.1016/j.intell.2016.04.002
- Murcray, C. E., Lewinger, J. P., & Gauderman, W. J. (2009). Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*, *169*, 219-226. doi:10.1093/aje/kwn353
- Murugan, M., Harward, S., Scharff, C., & Mooney, R. (2013). Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron*, *80*, 1464-1476. doi:10.1016/j.neuron.2013.09.021
- Nousiainen, A., & Ranta, M. (2015, November 7th). Sopraano sai aivoinfarktinkin ja menetti puhekykynsä, mutta hän laulaa aivan kuin ennen. *Kuukausi-ilite, Helsingin Sanomat*.
- Nozawa, D., Suzuki, N., Kobayashi-Osaki, M., Pan, X., Engel, J. D., & Yamamoto, M. (2009). GATA2-dependent and region-specific regulation of Gata2 transcription in the mouse midbrain. *Genes Cells*, *14*, 569-582. doi:10.1111/j.1365-2443.2009.01289.x
- Nusbaum, E. C., Silvia, P. J., Beaty, R. E., Burgin, C. J., Hodges, D. A., & Kwapil, T. R. (2014). Listening Between the Notes: Aesthetic Chills in Everyday Music Listening. *Psychol Aesthet Creat*, *8*, 104-109. doi:10.1037/a0034867
- Nuytemans, K., Theuns, J., Cruts, M., & Van Broeckhoven, C. (2010). Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Hum Mutat*, *31*, 763-780. doi:10.1002/humu.21277
- O'Connell, J. R., & Weeks, D. E. (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*, *63*, 259-266. doi:10.1086/301904
- O'Neill, S. A. (1997). Gender and Music. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 319). New York, US: Oxford University Press.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539-542. doi:10.1038/nature17671
- Olson, C. R., Hodges, L. K., & Mello, C. V. (2015). Dynamic gene expression in the song system of zebra finches during the song learning period. *Dev Neurobiol*, *75*, 1315-1338. doi:10.1002/dneu.22286
- Ostroveanu, A., van der Zee, E. A., Eisel, U. L., Schmidt, M., & Nijholt, I. M. (2010). Exchange protein activated by cyclic AMP 2 (Epac2) plays a specific and time-limited role in memory retrieval. *Hippocampus*, *20*, 1018-1026. doi:10.1002/hipo.20700
- Otsuka, Y., Yanagi, J., & Watanabe, S. (2009). Discriminative and reinforcing stimulus properties of music for rats. *Behav Process*, *80*, 121-127. doi:10.1016/j.beproc.2008.10.009
- Ott, J., & Hoh, J. (2000). Statistical approaches to gene mapping. *Am J Hum Genet*, *67*, 289-294. doi:10.1086/303031



- Park, H., Lee, S., Kim, H. J., Ju, Y. S., Shin, J. Y., Hong, D., . . . Seo, J. S. (2012). Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *J Med Genet*, *49*, 747-752. doi:10.1136/jmedgenet-2012-101209
- Parker, J., Tsagkogeorga, G., Cotton, J. A., Liu, Y., Provero, P., Stupka, E., & Rossiter, S. J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, *502*, 228-231. doi:10.1038/nature12511
- Patel, A. D. (2014). The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol*, *12*, e1001821. doi:10.1371/journal.pbio.1001821
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet*, *2*, e190. doi:10.1371/journal.pgen.0020190
- Peltonen, P. (2013). *Ympäristötekijöiden yhteys musikaalisuuden kehittymiseen ja ilmenemiseen*. (Master's thesis), University of Helsinki. Retrieved from <http://hdl.handle.net/10138/41847>
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., . . . Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*, 499-502. doi:10.1038/nature05295
- Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., . . . Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *P Natl Acad Sci U S A*, *107*, 4758-4763. doi:10.1073/pnas.0909074107
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*, 1-32. doi:10.1016/j.cognition.2005.11.004
- Peretz, I., Cummings, S., & Dube, M. P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am J Hum Genet*, *81*, 582-588. doi:10.1086/521337
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol*, *56*, 89-114. doi:10.1146/annurev.psych.56.091103.070225
- Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., . . . Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*, 1256846. doi:10.1126/science.1256846
- Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci*, *26*, 4970-4982. doi:10.1523/JNEUROSCI.3771-05.2006
- Power, R. A., Steinberg, S., Bjornsdottir, G., Rietveld, C. A., Abdellaoui, A., Nivard, M. M., . . . Stefansson, K. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci*, *18*, 953-955. doi:10.1038/nn.4040
- Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*, *20*, R208-215. doi:10.1016/j.cub.2009.11.055
- Profita, J., & Bidder, T. G. (1988). Perfect pitch. *Am J Med Genet*, *29*, 763-771. doi:10.1002/ajmg.1320290405
- Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H., & Järvelä, I. (2008). Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *J Med Genet*, *45*, 451-456. doi:10.1136/jmg.2007.056366
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*, 559-575. doi:10.1086/519795
- Qian, W., Deng, L., Lu, D., & Xu, S. (2013). Genome-wide landscapes of human local adaptation in Asia. *PLoS one*, *8*, e54224. doi:10.1371/journal.pone.0054224
- Qu, S., Olafsrud, S. M., Meza-Zepeda, L. A., & Saatcioglu, F. (2013). Rapid gene expression changes in peripheral blood lymphocytes upon practice of a comprehensive yoga program. *PLoS one*, *8*, e61910. doi:10.1371/journal.pone.0061910
- Rajakulendran, S., Roberts, J., Koltzenburg, M., Hanna, M. G., & Stewart, H. (2013). Deletion of chromosome 12q21 affecting KCNC2 and ATXN7L3B in a family with neurodevelopmental delay and ataxia. *J Neurol Neurosurg Psychiatry*, *84*, 1255-1257. doi:10.1136/jnnp-2012-304555

- Redies, C., Hertel, N., & Hubner, C. A. (2012). Cadherins and neuropsychiatric disorders. *Brain Res*, *1470*, 130-144. doi:10.1016/j.brainres.2012.06.020
- Rogenmoser, L., Elmer, S., & Jancke, L. (2015). Absolute pitch: evidence for early cognitive facilitation during passive listening as revealed by reduced P3a amplitudes. *J Cogn Neurosci*, *27*, 623-637. doi:10.1162/jocn\_a\_00708
- Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philos Trans R Soc Lond B Biol Sci*, *370*, 20140097. doi:10.1098/rstb.2014.0097
- Ross, D. A., Olson, I. R., Marks, L. E., & Gore, J. C. (2004). A nonmusical paradigm for identifying absolute pitch possessors. *J Acoust Soc Am*, *116*, 1793-1799.
- Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hear Res*, *308*, 71-83. doi:10.1016/j.heares.2013.08.016
- Saarikivi, K., Putkinen, V., Tervaniemi, M., & Huotilainen, M. (2016). Cognitive flexibility modulates maturation and music-training-related changes in neural sound discrimination. *Eur J Neurosci*, *44*, 1815-1825. doi:10.1111/ejn.13176
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., . . . Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*, 913-918. doi:10.1038/nature06250
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci*, *14*, 257-262. doi:10.1038/nn.2726
- Sanyal, T., Kumar, V., Nag, T. C., Jain, S., Sreenivas, V., & Wadhwa, S. (2013). Prenatal loud music and noise: differential impact on physiological arousal, hippocampal synaptogenesis and spatial behavior in one day-old chicks. *PLoS one*, *8*, e67347. doi:10.1371/journal.pone.0067347
- Särkämö, T., Altenmüller, E., Rodriguez-Fornells, A., & Peretz, I. (2016). Editorial: Music, Brain, and Rehabilitation: Emerging Therapeutic Applications and Potential Neural Mechanisms. *Front Hum Neurosci*, *10*, 103. doi:10.3389/fnhum.2016.00103
- Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soynila, S., Mikkonen, M., . . . Hietanen, M. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, *131*, 866-876. doi:10.1093/brain/awn013
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *P Natl Acad Sci U S A*, *112*, 8987-8992. doi:10.1073/pnas.1414495112
- Scharff, C., & White, S. A. (2004). Genetic components of vocal learning. *Ann N Y Acad Sci*, *1016*, 325-347. doi:10.1196/annals.1298.032
- Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Ann N Y Acad Sci*, *1337*, 170-177. doi:10.1111/nyas.12627
- Scherzer, C. R., Grass, J. A., Liao, Z., Pepivani, I., Zheng, B., Eklund, A. C., . . . Schlossmacher, M. G. (2008). GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein. *P Natl Acad Sci U S A*, *105*, 10907-10912. doi:10.1073/pnas.0802437105
- Schubloom, H. E., & Woolley, S. C. (2016). Variation in social relationships relates to song preferences and EGR1 expression in a female songbird. *Dev Neurobiol*, *76*, 1029-1040. doi:10.1002/dneu.22373
- Schulreich, S., Heussen, Y. G., Gerhardt, H., Mohr, P. N., Binkofski, F. C., Koelsch, S., & Heekeren, H. R. (2014). Music-evoked incidental happiness modulates probability weighting during risky lottery choices. *Front Psychol*, *4*, 981. doi:10.3389/fpsyg.2013.00981
- Scott, J. G., & Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J Stat Plan Infer*, *136*, 2144-2162. doi:10.1016/j.jspi.2005.08.031
- Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). *Seashore Measures of Musical Talents (Manual)*. New York, NY, USA: Psychological Corp.
- Seesjärvi, E., Särkämö, T., Vuoksima, E., Tervaniemi, M., Peretz, I., & Kaprio, J. (2016). The Nature and Nurture of Melody: A Twin Study of Musical Pitch and Rhythm Perception. *Behav Genet*, *46*, 506. doi:10.1007/s10519-015-9774-y



- Shah, K., & Lahiri, D. K. (2014). Cdk5 activity in the brain - multiple paths of regulation. *J Cell Sci*, *127*, 2391-2400. doi:10.1242/jcs.147553
- Shetake, J. A., Engineer, N. D., Vrana, W. A., Wolf, J. T., & Kilgard, M. P. (2012). Pairing tone trains with vagus nerve stimulation induces temporal plasticity in auditory cortex. *Exp Neurol*, *233*, 342-349. doi:10.1016/j.expneurol.2011.10.026
- Shinozuka, K., Ono, H., & Watanabe, S. (2013). Reinforcing and discriminative stimulus properties of music in goldfish. *Behav Process*, *99*, 26-33. doi:10.1016/j.beproc.2013.06.009
- Shuter, R. (1966). Hereditary and environmental factors in musical ability. *Eugen Rev*, *58*, 149-156.
- Shuter-Dyson, R., & Gabriel, C. (1981). *The psychology of musical ability* (2nd ed.). London ; New York: Methuen.
- Simonyan, K., Horwitz, B., & Jarvis, E. D. (2012). Dopamine regulation of human speech and bird song: a critical review. *Brain Lang*, *122*, 142-150. doi:10.1016/j.bandl.2011.12.009
- Singh, T. D., Basham, M. E., Nordeen, E., & Nordeen, K. (2000). Early sensory and hormonal experience modulate age-related changes in NR2B mRNA within a forebrain region controlling avian vocal learning. *J Neurobiol*, *44*, 82-94.
- Skouras, S., Gray, M., Critchley, H., & Koelsch, S. (2014). Superficial amygdala and hippocampal activity during affective music listening observed at 3 T but not 1.5 T fMRI. *Neuroimage*, *101*, 364-369. doi:10.1016/j.neuroimage.2014.07.007
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behav Brain Res*, *291*, 244-252. doi:10.1016/j.bbr.2015.05.026
- Sovansky, E. E., Wieth, M. B., Francis, A. P., & McIlhagga, S. D. (2016). Not all musicians are creative: Creativity requires more than simply playing music. *Psychol music*, *44*, 25-36. doi:10.1177/0305735614551088
- St Pourcain, B., Whitehouse, A. J., Ang, W. Q., Warrington, N. M., Glessner, J. T., Wang, K., . . . Smith, G. D. (2013). Common variation contributes to the genetic architecture of social communication traits. *Mol Autism*, *4*, 34. doi:10.1186/2040-2392-4-34
- Stanton, H. M. (1922). The inheritance of specific musical capacities. *Psychological Monographs*, *31*, 157-204. doi:10.1037/h0093168
- Storz, J. F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol*, *14*, 671-688. doi:10.1111/j.1365-294X.2005.02437.x
- Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear Res*, *308*, 109-121. doi:10.1016/j.heares.2013.08.004
- Sullivan, P. F., Fan, C., & Perou, C. M. (2006). Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet*, *141B*, 261-268. doi:10.1002/ajmg.b.30272
- Theusch, E., Basu, A., & Gitschier, J. (2009). Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *Am J Hum Genet*, *85*, 112-119. doi:10.1016/j.ajhg.2009.06.010
- Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philos T R Soc B*, *370*, 25-38. doi:10.1098/rstb.2014.0089
- Trehub, S. E. (2001). Musical predispositions in infancy. *Ann N Y Acad Sci*, *930*, 1-16.
- Tremere, L. A., Jeong, J. K., & Pinaud, R. (2009). Estradiol shapes auditory processing in the adult brain by regulating inhibitory transmission and plasticity-associated gene expression. *J Neurosci*, *29*, 5949-5963. doi:10.1523/JNEUROSCI.0774-09.2009
- Tylee, D. S., Kawaguchi, D. M., & Glatt, S. J. (2013). On the outside, looking in: a review and evaluation of the comparability of blood and brain "-omes". *Am J Med Genet B Neuropsychiatr Genet*, *162B*, 595-603. doi:10.1002/ajmg.b.32150
- Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., & Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS one*, *4*, e5534. doi:10.1371/journal.pone.0005534
- Ukkola-Vuoti, L., Oikkonen, J., Onkamo, P., Karma, K., Raijas, P., & Järvelä, I. (2011). Association of the arginine vasopressin receptor 1A (AVPR1A) haplotypes with listening to music. *J Hum Genet*, *56*, 324-329. doi:10.1038/jhg.2011.13

- Ullen, F., Hambrick, D. Z., & Mosing, M. A. (2015). Rethinking Expertise: A Multifactorial Gene-Environment Interaction Model of Expert Performance. *Psychol Bull*, *142*, 427-446. doi:10.1037/bul0000033
- Ullen, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Pers Individ Differ*, *63*, 87-93. doi:10.1016/j.paid.2014.01.057
- Umlauf, M. G. (2008). 22q11 deletion syndrome: Is that what they used to call...? *Perspect Psychiatr C*, *44*, 259-266. doi:10.1111/j.1744-6163.2008.00185.x
- van Bokhoven, H., Celli, J., van Reeuwijk, J., Rinne, T., Glaudemans, B., van Beusekom, E., . . . Brunner, H. G. (2005). MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome. *Nat Genet*, *37*, 465-467. doi:10.1038/ng1546
- van der Heijden, M. J., Oliari Araghi, S., van Dijk, M., Jeekel, J., & Hunink, M. G. (2015). The Effects of Perioperative Music Interventions in Pediatric Surgery: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PloS one*, *10*, e0133608. doi:10.1371/journal.pone.0133608
- VanderArk, S. D., & Ely, D. (1993). Cortisol, biochemical, and galvanic skin responses to music stimuli of different preference values by college students in biology and music. *Percept Mot Skills*, *77*, 227-234. doi:10.2466/pms.1993.77.1.227
- Velho, T. A., Lu, K., Ribeiro, S., Pinaud, R., Vicario, D., & Mello, C. V. (2012). Noradrenergic control of gene expression and long-term neuronal adaptation evoked by learned vocalizations in songbirds. *PloS one*, *7*, e36276. doi:10.1371/journal.pone.0036276
- Velho, T. A., Pinaud, R., Rodrigues, P. V., & Mello, C. V. (2005). Co-induction of activity-dependent genes in songbirds. *Eur J Neurosci*, *22*, 1667-1678.
- Vetter, D., Barth, J., Uyulmaz, S., Uyulmaz, S., Vonlanthen, R., Belli, G., . . . Clavien, P.-A. (2015). Effects of Art on Surgical Patients: A Systematic Review and Meta-analysis. *Ann Surg*, *262*, 704-713.
- Vieland, V. J. (1998). Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage. *Am J Hum Genet*, *63*, 947-954. doi:10.1086/302076
- Vieland, V. J., Huang, Y., Seok, S. C., Burian, J., Catalyurek, U., O'Connell, J., . . . Valentine-Cooper, W. (2011). KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Hum Hered*, *72*, 276-288. doi:10.1159/000330634
- Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *Am J Hum Genet*, *90*, 7-24. doi:10.1016/j.ajhg.2011.11.029
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu Rev Genet*, *47*, 97-120. doi:10.1146/annurev-genet-111212-133526
- Wada, K., Howard, J. T., McConnell, P., Whitney, O., Lints, T., Rivas, M. V., . . . Jarvis, E. D. (2006). A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. *P Natl Acad Sci U S A*, *103*, 15212-15217. doi:10.1073/pnas.0607098103
- Wang, C. (1985). Measures of creativity in sound and music. Retrieved from <http://www.uky.edu/~cecilia/MCSM/mcsm.htm>
- Wang, C. T., Grishanin, R., Earles, C. A., Chang, P. Y., Martin, T. F., Chapman, E. R., & Jackson, M. B. (2001). Synaptotagmin modulation of fusion pore kinetics in regulated exocytosis of dense-core vesicles. *Science*, *294*, 1111-1115. doi:10.1126/science.1064002
- Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Kunstner, A., . . . Wilson, R. K. (2010). The genome of a songbird. *Nature*, *464*, 757-762. doi:10.1038/nature08819
- Wason, J. M., & Dudbridge, F. (2012). A general framework for two-stage analysis of genome-wide association studies and its application to case-control studies. *Am J Hum Genet*, *90*, 760-773. doi:10.1016/j.ajhg.2012.03.007
- Watanabe, S., & Sato, K. (1999). Discriminative stimulus properties of music in Java sparrows. *Behav Process*, *47*, 53-57.
- Webster, P. (1983). *Refinement of a measure of musical imagination in young children and a comparison to aspects of musical aptitude*. Paper presented at the Loyola Symposium on Creativity, New Orleans, LA.

- Weir, B. S., & Hill, W. G. (2002). Estimating F-statistics. *Annu Rev Genet*, *36*, 721-750. doi:10.1146/annurev.genet.36.050802.093940
- Weisman, R. G., Balkwill, L.-L., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., & Sturdy, C. B. (2010). Absolute pitch in boreal chickadees and humans: Exceptions that test a phylogenetic rule. *Learn Motiv*, *41*, 156-173. doi:10.1016/j.lmot.2010.04.002
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., . . . Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*, *42*, D1001-1006. doi:10.1093/nar/gkt1229
- White, E. J., Hutka, S. A., Williams, L. J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. *Front Syst Neurosci*, *7*, 90. doi:10.3389/fnsys.2013.00090
- White, S. A. (2010). Genes and vocal learning. *Brain Lang*, *115*, 21-28. doi:10.1016/j.bandl.2009.10.002
- Whitney, O., Pfenning, A. R., Howard, J. T., Blatti, C. A., Liu, F., Ward, J. M., . . . Jarvis, E. D. (2014). Core and region-enriched networks of behaviorally regulated genes and the singing genome. *Science*, *346*, 1256780. doi:10.1126/science.1256780
- Willett, R. T., & Greene, L. A. (2011). Gata2 is required for migration and differentiation of retinorecipient neurons in the superior colliculus. *J Neurosci*, *31*, 4444-4455. doi:10.1523/JNEUROSCI.4616-10.2011
- Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting Factors Lead to Absolute Pitch Acquisition That Is Maintained in a "Fixed Do" Environment. *Music Percept*, *29*, 285-296. doi:10.1525/MP.2012.29.3.285
- Wing, H. D. (1941). A factorial study of musical tests. *B J Psychol - Gen Sect*, *31*, 341-355.
- Wozniak, R. J., Keles, S., Lugus, J. J., Young, K. H., Boyer, M. E., Tran, T. M., . . . Bresnick, E. H. (2008). Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. *Mol Cell Biol*, *28*, 6681-6694. doi:10.1128/MCB.01061-08
- Xu, J., Yu, L., Cai, R., Zhang, J., & Sun, X. (2009). Early auditory enrichment with music enhances auditory discrimination learning and alters NR2B protein expression in rat auditory cortex. *Behav Brain Res*, *196*, 49-54. doi:10.1016/j.bbr.2008.07.018
- Yang, E. J., Lin, E. W., & Hensch, T. K. (2012). Critical period for acoustic preference in mice. *P Natl Acad Sci U S A*, *109 Suppl 2*, 17213-17220. doi:10.1073/pnas.1200705109
- Yoder, K. M., Phan, M. L., Lu, K., & Vicario, D. S. (2015). He hears, she hears: are there sex differences in auditory processing? *Dev Neurobiol*, *75*, 302-314. doi:10.1002/dneu.22231
- Zatorre, R. J. (2003). Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nat Neurosci*, *6*, 692-695. doi:10.1038/nn1085
- Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: singing a different tune. *PLoS Biol*, *10*, e1001372. doi:10.1371/journal.pbio.1001372
- Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *P Natl Acad Sci U S A*, *110 Suppl 2*, 10430-10437. doi:10.1073/pnas.1301228110
- Zhou, Y., Yamamoto, M., & Engel, J. D. (2000). GATA2 is required for the generation of V2 interneurons. *Development*, *127*, 3829-3838.
- Zondervan, K. T., & Cardon, L. R. (2007). Designing candidate gene and genome-wide case-control association studies. *Nat Protoc*, *2*, 2492-2501. doi:10.1038/nprot.2007.366

## Recent Publications in this Series

**37/2016 Alexandre Borrel**

Development of Computational Methods to Predict Protein Pocket Druggability and Profile Ligands using Structural Data

**38/2016 Gonçalo Barreto**

Innate Immunity in Osteoarthritis: The Role of Toll-Like Receptors and Cartilage Derived Mediators in the Disease Progression

**39/2016 Alexey Yukin**

Animal Models of Early Brain Disorders: Behavioural and Electrophysiological Approaches

**40/2016 Satu Valo**

Western Diet and Genetic Predisposition as Risk Factors of Colon Cancer

**41/2016 Riccardo Siligato**

Hormonal Regulation of Primary and Secondary Growth in the Root of *Arabidopsis thaliana*

**42/2016 Janne Tynell**

Virus-Host Interactions of Emerging Respiratory Pathogens

**43/2016 Katri Korpela**

Intestinal Microbiota Development in Childhood: Implications for Health and Disease

**44/2016 Mikko Muona**

Identification of New Genetic Syndromes with Epilepsy by Whole-Exome Sequencing

**45/2016 Ana Cathia Dias Magalhães**

On the Mechanisms of Neural Development in the Ventral Telencephalon

**46/2016 Petri Arvilommi**

Treatment, Adherence and Disability in Bipolar Disorder

**47/2016 Annika Thomson**

Finnish Pretrial Male Firesetters: Mortality, Suicidality, Psychopathy and Morbidity of Schizophrenia

**48/2016 Sami Svanbäck**

Toward Accurate High-Throughput Physicochemical Profiling using Image-Based Single-Particle Analysis

**49/2016 Katarzyna Leskinen**

The Roles of YbeY, RfaH, and Hfq in Gene Regulation and Virulence of *Yersinia enterocolitica* O:3

**50/2016 Arjan van Adrichem**

Discovery of Small Molecule Modulators of Ras Superfamily Proteins Studies of MgcRacGAP and Ras

**51/2016 Meri Kekäle**

Chronic Myeloid Leukemia Patients' Adherence to Tyrosine Kinase Inhibitors in Finland – A Journey of Eighty-six Patients

**52/2016 Jaakko Kopra**

Endogenous GDNF as a Regulator of Midbrain Dopamine Neurons

**53/2016 Vasili Stegajev**

Histamine Receptor 4 (H4R) In pathogenesis of Sjögren's Syndrome

**54/2016 Katriina Viitasalo**

Shift Work and Cardiovascular Risk Factors — Prevention among Airline Employees

**55/2016 Michal Stepniewski**

Computational Studies on Membrane Proteins and Membrane-Drug Interactions

