Unravelling the genetic architecture of musical rhythm

Maria Niarchou[1,2*], J. Fah Sathirapongsasuti[3], Nori Jacoby[4,5], Eamonn Bell[6], Evonne McArthur[2], Peter Straub[2], The 23andMe Research Team[3^], J. Devin McAuley[7], John A. Capra[8,9,10,2], Fredrik Ullén[11], Nicole Creanza[8], Miriam A. Mosing[11,12], David Hinds[3], Lea K. Davis[2,13,10,14,15*†], Reyna L. Gordon[16,17,2*†]

[1]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, UK

[2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, TN, USA

[3]23andMe, Inc., Sunnyvale, CA, USA

[4]Computational Auditory Perception Research Group, Max Planck Institute for Empirical Aesthetics, Germany

[5]Center for Science and Society, Columbia University, USA

[6]Music Department, Columbia University, USA

[7]Department of Psychology, Michigan State University, Michigan, USA

[8]Department of Biological Sciences, Vanderbilt University, TN, USA

[9]Department of Computer Science, Vanderbilt University, TN, USA

[10]Department of Biomedical Informatics, Vanderbilt University Medical Center, TN, USA

[11]Department of Neuroscience, Karolinska Institutet, Sweden

[12]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

[13]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, TN, USA

[14]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, TN, USA

[15]Department of Molecular Physiology and Biophysics, Vanderbilt University, TN, USA

[16]Department of Otolaryngology, Vanderbilt University Medical Center, TN, USA

[17]Department of Psychology, Vanderbilt University, TN, USA

^full list of authors in Supplementary Note

*Correspondence to: Reyna L. Gordon (reyna.gordon@Vanderbilt.Edu), Lea Davis (lea.k.davis@gmail.com), Maria Niarchou (maria.niarchou@vumc.org)

†Joint supervision

Abstract

While timing and rhythm-related phenotypes are heritable, the human genome variations underlying these traits are not yet well-understood. We conducted a genome-wide association study to identify common genetic variants associated with a self-reported musical rhythm phenotype in 606,825 individuals. Rhythm exhibited a highly polygenic architecture with sixty-eight loci reaching genome-wide significance ($p<5x10^{-8}$) and SNP-based heritability of 13%-16%. Polygenic scores for rhythm predicted the presence of musician-related keywords in the BioVU electronic health record biobank. Genetic associations with rhythm were enriched for genes expressed in brain tissues. Genetic correlation analyses revealed shared genetic architecture with several traits relevant to cognition, emotion, health, and circadian rhythms, paving the way to a better understanding of the neurobiological pathways of musicality.

Introduction

Rhythm is a fundamental aspect of music across cultures [1,2] and more broadly, rhythmic patterns provide sensori-motor structure to human interactions. Our tendency to perceive, create, manipulate, and appreciate rhythms in a variety of contexts (e.g., speech, music, movement) is part of what makes us human. Even very young children are sensitive to the social and linguistic signals carried by rhythm [3], thus it is not surprising that parents use rhythmic vocalizations and synchronous movement (e.g., lullabies and rocking) to interact with their infants from birth [4]. Moving in synchrony to a musical beat ("beat synchronization") appears to be a key feature of human musical experiences throughout the lifespan [5-7].

Although most people are able to effortlessly detect and synchronize with the beat even without musical training [6,7], there is substantial inter-individual variability (within cultures) in the extent to which individuals can perceive and produce musical rhythm accurately [8-10]. While the neuroimaging literature points to auditory-motor networks in the brain underlying rhythm perception and production [11], less is known about the genetic underpinnings that give rise to individual differences in these networks. Heritability estimates from family-based studies, using a variety of measures relevant for rhythmic ability, range from 21% [12] to 50% [13]. There is a gap in knowledge about genomic loci underlying variation in rhythm ability [14], in part due to the challenge of assessing the rhythm phenotype in a sample large enough to provide sufficient power to detect common variants with small effects, as expected for complex traits [15].

Summary of Approach.

We conducted a genome-wide association study (GWAS) to identify common genetic variants associated with a self-reported musical rhythm phenotype, i.e. “Can you clap in time with a musical beat?”, collected from 606,825 individuals participating in research with the personal genetics company 23andMe, Inc. We then validated this self-reported phenotype in a separate internet-based behavioural study conducted in 724 individuals and found that it was significantly correlated with rhythm perception (Online Methods). In the GWAS, a total of 68 independent SNPs surpassed the threshold for genome-wide significance ($p<5x10^{-08}$). In addition to determining which genes were implicated by these variants, we estimated how much of the total phenotypic variance could be explained by all variation across the genome (i.e., SNP-based heritability). We then further explored this heritability to test the hypothesis that variants associated with rhythm were enriched among genes expressed in brain compared to genes expressed in other tissues (e.g., muscle, adipose, etc.), and furthermore enriched in genes expressed in neurons compared to other brain cell types (e.g., oligodendrocyte, astrocyte).

Using an independent sample of 67,441 genotyped individuals from the Vanderbilt University Medical Center biobank, BioVU, we tested whether a cumulative sum of the genetic effects for rhythm detected in our GWAS (i.e., rhythm polygenic score), was significantly associated with an indication of musician status in the electronic health record (EHR). Because little is yet known about the relationship of the genomics of rhythm to other traits, we also performed exploratory genetic correlation analysis including 764 complex traits for which a well-powered GWAS has been performed and deposited in LDHub [16]. Finally, we evaluated the contribution to rhythm

of regions of the genome that have experienced significant human-specific evolutionary shifts (since the divergence of humans and chimpanzees from their last common ancestor, ~6 million years ago).

Results

Validating the self-reported rhythm phenotype

The study population is N=606,825 participants of European ancestry (59% females, mean age(SD)=52.09(18.5) years, who consented to participate in research with 23andMe, Inc. Data were available from individuals who answered the question "Can you clap in time with a musical beat?" The majority of participants answered 'Yes' (91.57%) and 8.43% answered 'No', which is slightly higher than the estimated population prevalence of poor rhythm at ~5% [17,18] (Table 1 and Supplementary Note). In light of prior work suggesting that human rhythm is a complex trait that can be quantified with both objective and self-report measures [10], we sought to validate the self-report question against an objective measure of rhythm perception. We conducted a phenotype validation study with a sample (N=724; mean age=36 years, SD=10.9; 46% females) recruited anonymously from Amazon's Mechanical Turk. Participants performed an objective musical rhythm perception test and were asked "Can you clap in time with a musical beat?" (details provided in Online Methods). In each of the 32 trials, participants had to judge whether a pair of rhythms were the same or different, following a standard procedure for assessing individual differences in musical perception ability [9] and utilizing rhythm sequences with simple (highly metrical) and complex (syncopated) rhythms [19]. Individuals who had better performance on discriminating musical rhythms were more likely to answer 'Yes' to the self-report synchronization question than those who answered 'No' (OR(95%CI)=1.94(1.28 to 3.01), p=0.002, McFadden's R^2=0.39 (i.e, we expect to see a 94% increase in the odds of answering 'Yes', for a standard

deviation increase in the rhythm discrimination test). In the remainder of the paper, the "rhythm" trait in our study refers to the self-reported beat synchronization phenotype.

GWAS results and heritability estimation

GWAS was conducted using logistic regression under an additive genetic model, while adjusting for age, sex, the top five principal components of ancestry in order to control for population stratification, and indicators for genotype platforms to account for batch effects. We excluded SNPs with Minor Allele Frequency (MAF) <0.01, low imputation quality (R^2<0.3) and indels, resulting in a final set of 8,288,851 SNPs for all subsequent analyses. Sixty-eight independent SNPs (after two rounds of LD pruning, first at r^2=0.6 and then at r^2=0.1, kb = 250) reached genome-wide significance ($p<5x10^{-8}$; Figure 1, Supplementary Table 1, Supplementary Figure 1**)**, from a total of 6,115 SNPs that passed the significance threshold.

Linkage Disequilibrium Score Regression (LDSC)[20] analyses revealed that heritability estimates on the liability scale ranged from 13% to 16% when adjusted for a range of estimated population prevalence of rhythm deficits (from 3.5% to 6.5% [17,18]) (Supplementary Table 2, Supplementary Note).The observed SNP-heritability explained 5% (se=0.0002) of the phenotypic variance in the rhythm trait, with an LD score regression intercept of 1.02 (se=0.01).

Gene-based analyses. Gene-based association analyses performed with MAGMA yielded 203 genes that surpassed the threshold of $p<3x10^{-6}$ (Supplementary Table 3).

The top two genes are: *CCSER1,* in proximity to genes previously associated with musicality [21], and *VRK2* (converging with the top locus identified in our SNP-based association analyses).

We also examined potential replication of genetic associations with musicality in humans from prior reports (28 genes were selected including 26 reported in meta-analysis by [21], and additionally, *GATA2* and *PCDH7* [22] and *UGT8* [23]. Although none of the genes reached statistical significance (Supplementary Table 4, Supplementary Note), several are located near *CCSER1* in the 4q22-24 region.

Heritability Partitioning

One advantage to SNP-based heritability estimation is the ability to partition heritability according to SNP-annotations, which provides insight into the types of genetic variation that contribute most to rhythm. To determine whether heritability is enriched for specific functional categories of SNP annotations, stratified LDSC [24] was used to partition heritability (Supplementary Table 5). We hypothesized that SNPs falling into regions of open chromatin (i.e., accessible to transcriptional machinery), and regions with human-specific variation, would be enriched for rhythm-associated variation. We found enrichment in regions conserved in mammals (regions of the genome identified by Lindblad-Toh et al. 2013 as being under purifying selection) (enrichment=15.8, $p=1.19 \times 10^{-12}$) and in functional categories involved in acetylation of histone H3 at lysine 9 (H3K9ac) (enrichment=8.0, $p=1.85 \times 10^{-8}$) and monomethylation of histone H3 at lysine 4 (H3K4me3) (enrichment=1.29, $p=2.16 \times 10^{-5}$), supporting associations mediated by

effects on gene regulation. Enrichment was also found in the 'Repressed' category of chromatin states (enrichment=0.87, p=0.0002), and for introns. We also examined whether genes expressed in specific cell-types show enrichment among rhythm-associated variants as described in [24]: and found that genes expressed specifically in neurons contributed significantly to trait heritability (coefficient=1.19 x 10^{-9}, p=0.037) conditional to the other annotations (Supplementary Table 6).

Gene set analyses

Using FUMA [25], we performed a gene-property analysis where the average expression of the genes per tissue type (using GTEx gene expression panels in 53 tissue types [26] was added as a covariate in the model. As predicted, gene associations were significantly enriched in brain tissue compared to non-brain tissues (Figure 2). To further examine potential biological pathways associated with rhythm, we performed MAGMA gene-set analyses as implemented by FUMA [25]. Two gene-sets out of 10,678 achieved statistical significance after Bonferroni correction (Supplementary Table 7). The top associated gene-sets with rhythm were: Negative regulation of transcription from RNA polymerase II promoter (p=8.6 x 10^{-07}), gene-set from the Gene Ontology project [27,28](i.e., any process that includes glucose and decreases the rate or frequency of transcription from an RNA polymerase II promoter) and Negative regulation of gene expression (2.9x10^{-6}).

Human Accelerated Region and Neanderthal Introgression Stratified Heritability Analyses

Given previous hypotheses about the origins of rhythm[6,7,29], we evaluated the contribution of regions of the human genome that have experienced significant human-specific shifts in evolutionary pressure using stratified LDSC [20,24]. In particular, we analyzed the contribution to rhythm heritability from variants in genomic loci that are conserved across non-human species, but have elevated substitution rate on the human lineage [30]. Many of these human accelerated regions (HARs) play roles in human-specific traits[31], including cognition [32]. The heritability of rhythm is enriched 2.26-fold in variants in perfect linkage disequilibrium with HARs (p = 0.14). However, given the small number of variants in these regions and the enrichment of HARs in functional regions of the genome, it is difficult to explicitly link these shifts to rhythm. Nonetheless, two of the variants most strongly associated with rhythm (rs14316, rs1464791) fall within HARs, and the rs1464791 variant is near *GBE1*, a gene associated with a range of traits including body-mass index (BMI) [33] and cognitive deficits [34].

We also evaluated the contribution of genetic variants detected in the Neanderthal genome present in modern Eurasians due to interbreeding (hereafter "Neanderthal variants") to the heritability of the rhythm phenotype. Eurasian genomes contain ~1.5-4% of DNA as a result from interbreeding with Neanderthals around 50,000 years ago. Heritability of rhythm was significantly depleted among Neanderthal variants (1.97-fold depletion, P = 0.001). However, Neanderthal ancestry is significantly depleted in functional genomic regions overall [35], therefore, the depletion of rhythm heritability in these regions is likely the result of the overall depletion for Neanderthal ancestry in functional regions of the genome. This is supported by a non-significant τ_c^*, illustrating that Neanderthal vs. human variants do not provide unique heritability when

conditioned on a broad set of regulatory elements [36](Supplementary Table 8, Online Methods).

Proof-of-concept of the genetics of musicality in a health care context

As a proof-of-concept that genetics of rhythm are more widely tied to the biology of musicality, we further examined whether the contribution of the common alleles associated with rhythm en masse (also known as polygenic scores (PGS)) predict the presence of keywords indicating "musician" status in clinical documentation collected in the electronic health record (see Supplementary Note for details). In a sample of 67,441 individuals in Vanderbilt's BioVU, we identified 864 individuals with the keyword "musician" (or other closely related keywords for musical instruments) present in the EHR that we compared with 66,577 without any mention of "musician" keywords in their EHR. We found evidence that the PGS for rhythm was significantly higher among individuals with the "musician" keywords in their chart (OR per SD increase in PGS, 1.30, 95%CI:1.20-1.38, $p<2.5 \times 10^{-13}$, Nagelkerke's R2=1%) (Supplementary Table 9, Figure 3), confirming our hypothesis that the rhythm phenotype assessed in our study captures a dimension of musicality.

Rhythm beyond the contribution of intelligence

In light of previous work linking rhythm and IQ [17,37], we used multi-trait conditional joint analysis [38] (mtcojo) to remove shared genetic effects between intelligence and

rhythm. This analysis generated a new set of summary statistics of rhythm in which betas, standard errors and p-values were adjusted based on the intelligence summary statistics from [39]. Using FUMA as described above, we identified 66 independent, genome-wide significant loci in the conditioned GWAS, all of which were within 5kb of the loci in the unadjusted rhythm summary statistics (Supplementary Table 10, Supplementary Figure 2). We also compared effect estimates in 47 independent, genome-wide significant SNPs available from both the unadjusted rhythm and IQ GWAS datasets; all that were in common between these two datasets remained significant at the GWAS threshold, and their effect estimates were not changed (Supplementary Table 11). Also, the genetic correlation between the IQ GWAS dataset and rhythm was not significant (rg=-0.003(standard error=0.02), p=0.88). Similarly, the estimates of the heritability in the liability scale remained the same (13% to 16%). These findings indicate that our results are largely driven by associations with rhythm rather than cognitive ability.

Table 2 shows the rhythm-related loci that are also present in the GWAS catalogue after adjusting for genetic effects shared with IQ (for a full list of loci see Supplementary Table 12).

Cross-trait analyses

To determine if rhythm shares genetic architecture with other traits, we tested genetic correlations [20] between rhythm and all 764 available traits in LDHub (v.1.9.2) using LDscore regression. This method is designed to show whether there is shared genetic variation linked to a particular trait (here, our rhythm trait) and traits measured in

other samples/studies. There were 31 statistically significant genetic correlations ($p<6.5 \times 10^{-5}$) between rhythm and other traits after adjusting for multiple comparisons (Figure 4, Supplementary Table 13).

As expected, processing speed measured as 'mean time to correctly identify matches' was negatively correlated with rhythmic ability ($rg=-0.16$, $p=3.22 \times 10^{-13}$) (i.e., faster processing speed was associated with having rhythm). Educational qualifications (O' levels/GCSEs or equivalent) ($rg=0.16$, $p=4.6 \times 10^{-7}$), evening chronotype ($rg=0.09$, $p=3.8 \times 10^{-5}$) and tinnitus ($rg=0.20$, $p=6.7 \times 10^{-6}$) were all positively associated with rhythm. While falling short of the correction for multiple testing, exposure to loud music was also correlated with a similar point estimate ($rg=0.20$, $p=2.0 \times 10^{-4}$) and could be due to a relationship between tinnitus and loud music exposure in the UKBB ($rg=0.30$, $p=4.8 \times 10^{-6}$) [36,40].

Additionally, we identified significant genetic correlations between rhythm and hand grip strength ($rg(left)=0.18$, $se=0.02$, $p=3.6 \times 10^{-16}$, $rg(right)=0.16$, $se=0.02$, $p=6.91 \times 10^{-15}$), smoking including 'ever smoked' ($rg=0.16$, $p=2.5 \times 10^{-11}$) and 'past tobacco smoking' ($rg=-0.15$, $p=4.6 \times 10^{-10}$) as well as with peak expiratory flow from both the UKBiobank ($rg=0.15$, $p=2.11 \times 10^{-9}$) and a second independent GWA study ($rg=0.11$, $p=6.6$, 10^{-8}) and several other lung-related phenotypes (Supplementary Table 13). Given that the majority of these traits come from the UKBiobank, it is also possible that their genetic correlations with rhythm, may be a function of their correlation with each other, as some degree of phenotypic correlation is also expected.

Recent studies illustrate the potential for very subtle residual population substructure to influence some polygenic analyses [41] including genetic correlations. Therefore, we also adjusted the rhythm associations for SNP-loadings on the first principal component of ancestry estimated from 1KG European populations. We then used these SNP estimates of ancestry to adjust the rhythm GWAS results which yielded no change in the genetic correlations results [41] (Supplementary Table 14 and Supplementary Note).

Although we cannot determine potential causality, we conducted MR analyses using the GSMR [38] to examine whether there are significant bi-directional relationships between rhythm and processing speed, handgrip strength, and chronotype (Supplementary Note). We found significant bidirectional relationships for all traits in the analysis (Supplementary Table 15).

Sensitivity Analysis of Chromosome 17 locus for chromosomal inversions and Parkinson's Disease

Given that the genome-wide significant locus (lead SNP rs4792891) on chromosome 17q21 is located within a well-established inversion region that may also be associated with local population substructure [42], we conducted additional analyses focused on the region. The inversion was not associated with local ancestry within our study sample (Supplementary Table 16), suggesting that the association between this locus and rhythm is not likely to be due to local population confounding.

In addition, we sought to explore the potential effect of Parkinson's disease (PD) phenotype on this Microtubule Associated Protein Tau (*MAPT)* locus (17q21). Taking

into account that PD patients may have difficulty discriminating beat-based rhythms [43], and also that PD patients are over-represented in the 23andMe database, it was possible that the inclusion of PD patients in the sample may account for these associations. The associations between the independent SNP in the locus, rs4792891, and rhythm remained after removing PD patients from the sample, indicating that this MAPT association with rhythm is not driven by PD cases (Supplementary Note**,** Supplementary Table 17).

Discussion

This study demonstrates that common genetic variation plays a role in a musical rhythm trait, complementing prior evidence of innate human rhythm sensitivity [6,7]. Based on a self-reported beat synchronization phenotype that was validated with an objective measure of rhythm perception, the present large-scale study (606,825 participants from 23andMe) is a significant first step towards well-powered genomic evidence of a musicality phenotype. Sixty-eight independent SNPs (Supplementary Table 1) surpassed the threshold for genome-wide significance, with the top-associated locus mapped to *VRK2-FANCL* (rs848293, $p=9.2 \times 10^{-18}$), a protein kinase with multiple spliced isoforms expressed in brain that was previously associated with behavioural and psychiatric phenotypes (i.e., depression, neuroticism and schizophrenia [44-46] developmental delay) [47], indicating a biological connection between rhythm and neurodevelopment.

The total SNP-based heritability of our rhythm trait on the liability scale ranged from 13 to 16%, in line with both estimates of other complex traits (e.g., asthma [48]) and previously reported heritability estimates of musical rhythm abilities reported in twins [13]. Enrichment of heritability of rhythm in multiple brain tissues, notably cerebellum, basal ganglia, and cortex, likely reflects the genetic contribution to subcortical-cortical networks underlying musical rhythm perception and production [11]. Indeed, brain structures associated with rhythm include basal ganglia [49-51], cerebellum [52,53] and thalamus [54]. Furthermore, we found heritability enriched in genes expressed in neuronal cell types and in SNPs and genes responsible for expression regulation; taken together,

these results suggest that genomic loci that influence rhythm are enriched for effects on the brain and mediated by regulation of gene expression.

Initial clues about the evolution of rhythm traits in humans may be indicated by the occurrence of two of the rhythm-associated loci in human-accelerated regions (HARS) of the genome. In particular, rs1464791 is an eQTL that regulates expression of *GBE1* in multiple tissues including adrenal gland and muscle [26]. It is too early to tell whether the overlap between rhythm-associated loci and those two HARS support evolutionary theories about music (e.g., moving to a beat in synchrony during joint music-making and temporally coordination movements has been posited to have a selection effect in modern humans by enhancing group social cohesion and mother-infant bonding [1,55].

The genetic architecture of rhythm remained virtually unchanged after conditioning the analyses on known GWAS markers of intelligence, in line with twin studies showing specific genetic effects of rhythmic aptitude, over and above common genetic influences on rhythm and intelligence [17,56]. Furthermore, 30 loci do not appear to have existing genome-wide significant associations with other traits in the current literature, and thus may represent genomic regions newly associated to some aspect of musicality. At the same time, the other 36 loci coincided with robust associations in the GWAS catalogue for a variety of cognitive, neuropsychological, and health traits (Table 2, Supplementary Table 12), indicating that rhythm shares genetic architecture with many other traits. We replicated previous findings implicating location 4q22.1 in musicality-related traits [12,23] (*CCSER1* was the top-associated gene in our MAGMA analysis) but did not find support for previous gene associations from prior candidate-

gene, linkage, and GWAS studies with relatively small samples [21], potentially due to well-known methodological problems with these methods particularly when applied to small samples [57].

Positive genetic correlations between rhythm and faster processing speed aligned with prior phenotypic and behavioural genetic studies of cognition, sensory processing, and musicality [17,56,58,59]. The correlation between rhythm and chronotype opened up the possibility of a relationship between musical traits and evening chronotype, complimenting evidence of insomnia in musicians [60].

We found positive genetic correlations with tinnitus, which could be driven by exposure to loud music (this latter correlation with rhythm was just above the significance threshold after multiple-test corrections); both commonly occur among musicians and may lead to hearing loss [61] highlighting the importance of estimating the prevalence of professional musicians within the study sample in future GWAS of rhythm (this information was not available in the current sample). Unexpected genetic correlations included associations of rhythm with better lung capacity, previous smoker phenotypes, and greater handgrip strength. In light of recent evidence that lung function is genetically related to motor function, processing speed, and cognition in older adults [62], it is possible that rhythm shares common biology with a constellation of traits. These lines of research may have clinical-translational implications: for example, a recent intervention study found that music listening improved handgrip strength in older adults [63]. We also uncovered shared genetic effects between musical rhythm and biological rhythms including circadian chronotypes and breathing-related phenotypes.

More broadly, the genetic correlations between rhythm and other complex traits were relatively modest, suggesting that the present phenotype is not primarily confounded/co-occurring with any particular trait we examined. There are no large-sample GWAS data for major processes fundamental to beat synchronization [7,11]: auditory processing, sensori-motor synchronization, locomotion, or temporal processing as a component of general timing abilities [64], for which we may expect greater genetic correlations with rhythm in future studies.

The primary limitation of our study is the self-reported assessment of rhythm. Although our independent phenotypic validation study indicated that an individual's self-assessment of beat synchronization is related to their objectively-measured rhythm perception abilities, the self-report itself is not an objective assessment of rhythm. Nevertheless, previous studies of other health traits based on self-report have effectively replicated associations from studies using validated assessments, indicating that a powerful sample size can overcome limitations arising from phenotyping error [65]. The selection of the self-report beat synchronization phenotype was made because it theoretically relates to fundamental components of rhythm including motor periodicity, beat extraction, meter perception, and auditory-motor entrainment (see [7] and Glossary in Supplementary Note). Nevertheless, the phenotype available in our GWAS dataset did not allow us to separate the rhythm phenotype into those component factors, and the prevalence of individuals with musical training in the sample was not established. However, given the result that polygenic score for rhythm predicted the presence of musician keywords in an electronic health record-linked biobank, it is likely that we have indeed captured a robust aspect of musicality. These results are promising for future

large-scale genomic interrogations using comprehensive music phenotyping yielding continuous musicality variables (whether questionnaire-based [10,66] or objective aptitude-based [13]). Even without continuous measures of rhythm, here we have identified biology potentially differentiating rhythm deficits [67] from typical rhythm development. Once GWAS results are available from other heritable musicality traits such as pitch discrimination and music training [14], the field will be able to test for moderate genetic correlations between rhythm and other musical traits as predicted by family-based studies [12,13,68]. Another important area of inquiry will be to investigate musicality and cross-trait correlations in populations of non-European ancestry, hence capturing the spectra of musicality, a human universal, in a wider range of ethnic, cultural and socio-economic contexts.

Online Methods

Study sample

We obtained genome-wide association study summary statistics from the personal genetics company 23andMe, Inc. Phenotypic status was based on responses to online surveys in which individuals self-reported “Yes” (cases) or “No” (controls) to the question ‘Can you clap in time with a musical beat?”. Individuals who responded “I’m not sure” were excluded from our genomic study. The GWAS included a total of 555,660 cases and 51,165 controls (total N=606,825, mean age(SD)=52.09(18.5), prevalence=92%). Specifically, 10.4% of the individuals were 30 years old or younger, 24.4% were between 30 and 45 years old, 27.1% were between 45 and 60 years old and 38.1% were older than 60 years old (Table 1). All individuals provided informed consent according to 23andMe’s human subject protocol, which is reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com).

Phenotype validation study

Overview. To validate the rhythm phenotype used in the genetic study, we conducted a separate internet-based study in N=724 participants from Amazon’s Mechanical Turk. The experiment was designed to determine if self-reported rhythm abilities measured with the question used in the GWAS (i.e., ‘Can you clap in time with a musical beat?’) would be associated with objective performance on a task of rhythm abilities. The Beat-based advantage paradigm was selected as a rhythm discrimination test due to its design of stimuli with simple and complex meter [69] and prior history investigating

436 individual differences in rhythm perception in a variety of brain and behavioural studies
437 in adults and children with typical and atypical development [19,43,70,71], as well as
438 feasibility for internet-based adaptation. The questionnaire (self-report questions) was
439 administered prior to the perception task, to avoid biasing participant self-report
440 responses by how they perceived they performed on the objective test.
441
442 *Participants*
443 We recruited 724 participants anonymously from Amazon Mechanical Turk. The study
444 received ethical approval from the Columbia University Institutional Review Board.
445 Participants (333 females) were 18-73 years old (mean = 36.1 years, SD=10.9) with 0-
446 45 years of self-reported musical experience (mean 3.7 years, SD=5.8).
447
448 *Stimuli*
449 Stimuli consisted of 32 rhythms drawn from prior work [19,69]; half were “simple” rhythms
450 (strong beat-based metrical structure and generally easier to discriminate) and half were
451 “complex” rhythms (weaker metrical structure due to syncopation and generally more
452 challenging to discriminate). Each rhythm was presented as a pure tone in one of 6
453 frequencies (294, 353, 411, 470, 528, and 587 Hz, selected at random), and one of 4
454 durations (ISI of 220, 230, 240, and 250 ms). Each trial consisted of 3 rhythms
455 separated by 1500 ms of silence. As in prior work, the two first presentations were
456 always identical, and in half of the trials (counterbalanced) the third rhythm was also
457 identical (standard condition); in the other trials the rhythm was slightly different (deviant
458 condition).

459

460 *Procedure*

461 Amazon Mechanical Turk (M-Turk) participants were invited to participate in an
462 experiment where they would "listen to sounds and answer questions". To simulate the
463 user environment within 23andMe where research participants answer a series of
464 unrelated questions about health and other traits, we asked participants to provide
465 answers for a series of randomly presented questions on a variety of other topics
466 (presented at random order; see methods), such as "Do you have wisdom teeth?".
467 Among these questions we embedded two rhythm-related questions: the target
468 question: "Can you clap in time with a musical beat?" and an additional question, "Do
469 you have a good sense of rhythm?". After answering these questions, participants
470 passed a test for usage of headphones [72]. This test checks whether participants can
471 hear sounds that are presented through headphones, and guarantees good listening
472 conditions as well as the ability to follow instructions. Participants that passed the
473 headphone test were invited to perform the rhythm perception task (Supplementary
474 Figure 3).

475

476 Participants received 8 training trials that were selected from rhythms that were not part
477 of the test set, and then performed 32 rhythm perception task trials. In all trials (practice
478 and task) participants received feedback regarding their performance ("correct" and
479 "incorrect"), and each correct trial resulted in adding a small monetary bonus.
480 Participants were paid for their performance about $1.60-$2.00 depending on their
481 performance, and the duration of the test was about 16-18 minutes. Participants who

did not pass the headphone test received $0.20 for about one minute of answering the initial questions and performing the headphone test. Participant demographic data was collected after the rhythm test.

Phenotype Validation Results

654 (90.3%), 25 (3.5%) and 45 (6.2%) participants answered “yes”, “no,” and “I’m not sure” to the target question, “Can you clap in time to a musical beat”. Regarding the self-report question ‘Do you have a good sense of rhythm?’, 503(67%) answered ‘Yes’, 102(14%) answered ‘No’ and 117(16%) answered ‘I don’t know’. N=488 answered Yes to both questions, while 166 answered Yes to the Clap to Beat question and 15 answered Yes to the sense of rhythm question, resulting in a total tetrachoric correlation between these two self-report questions of r=0.73.

Responses to the rhythm discrimination perception test were analysed using signal detection theory [73], as in [19]; this method is appropriate for discrimination tasks where the participant has to categorize stimuli along some dimension; the resulting *d’* values the strength of detection of the signal relative to noise. *d’* values were calculated on the 32 test trials (16 simple rhythm trials and 16 complex rhythm trials) and are reported in Supplementary Table 18. As expected from prior work [19,70], individuals scored better in the simple rhythms than the complex rhythms ($t(724)=11.11$, $p<2.2 \times 10^{-16}$, Cohen’s $d=0.58$ (Supplementary Figure 4).

To examine whether the self-report of rhythm ability was related to the objective performance on the rhythm discrimination/perception test (see task performance in

relation to responses to self-report, shown in Supplementary Figure 5a), we performed a logistic regression analysis in which the self-report rhythm question (Yes vs. No) was the outcome and the rhythm discrimination test performance (standardized *d'* scores mean = 0, SD = 1) was the predictor. Covariates included age at time of assessment, education, and sex. Individuals with higher performance in the rhythm discrimination test (total *d'*) were more likely to answer that they can clap to the beat (OR(95%CI)=1.94(1.28 to 3.01), p=0.002, McFadden's R^2=0.39), indicating there is approximately a 94% increase in the odds of answering 'Yes', per standard deviation increase in the rhythm discrimination test. We did not include 'I'm not sure' in the regression, because this answer is not included in the phenotype assessment of the genetic study. Because the simple rhythms have a strong metrical structure and are known to facilitate detection and synchronization of the beat [19], we also tested whether performance on the simple rhythm trials predicted self-reported beat synchronization (i.e., those who responded Yes to the clap-to-beat question). As above, we found that individuals with higher scores on the simple rhythm trials were more likely to answer that they can clap to the beat (OR(95%CI=1.99(1.36-2.90), p<0.001, McFadden's R^2=0.40 (Supplementary Figure 5b). Taken together, these results suggest that the "clap to the beat" self-report phenotype is a broad representation of musical rhythm ability, potentially capturing aspects both of rhythm perception ability and of self-perceived beat synchronization ability.

Genotypes and QC

23andMe dataset

The National Genetics Institute (NGI) performed the DNA extraction and genotyping on saliva samples. Overall, there were five genotyping platforms and subjects were genotyped on only one of them. The v1 and v2 platforms had variants of the Illumina HumanHap550+ BeadChip, including approximately 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The v3 platform had variants of the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with the v2 array, with a total of about 950,000 SNPs. The v4 platform covered about 570,000 SNPs, providing extra coverage of lower-frequency coding variation. The v5 platform, in current use, is based on an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. In cases where samples did not reach the 98.5% call rate, the sample was re-genotyped. When analyses failed repeatedly, then customers were re-contacted by 23andMe customer service to provide additional samples.

23andMe restricted participants to a set of unrelated individuals of European ancestry as determined through an analysis of local ancestry [74]. Relatedness was defined using a segmental identity-by-descent (IBD) estimation algorithm [75]. Imputation was conducted by combining the May 2015 release of 1000 Genomes Phase 3 haplotypes [76] with the UK10K imputation reference panel [77] to create a single unified imputation reference panel. Phasing was conducted using an internally-developed tool, Finch, which uses the Beagle graph-based haplotype phasing algorithm [78] for platforms V1 to V4 while for the V5 platform a similar approach was used with a new phasing algorithm, Eagle2 [79]. SNPs with a Hardy-Weinberg $p<10^{-20}$, or a call rate of <90% were flagged.

SNPs were also flagged if they were only genotyped on their 'V1' and/or 'V2' platforms due to small sample size and also if SNPs had genotype date effects. Finally, SNPs were also flagged if they had probes matching multiple genomic positions in the reference genome [75-79].

GWAS

GWAS was conducted using logistic regression under an additive genetic model, while adjusting for age, sex, the top five principal components of ancestry in order to control for population stratification, and indicators for genotype platforms to account for batch effects. We excluded SNPs with Minor Allele Frequency (MAF) <0.01, low imputation quality (R^2<0.3) and indels, resulting in a final set of 8,288,851 SNPs for all subsequent analyses.

Statistical analyses

FUMA-based analyses. The FUMA [25] web application was used on the Genome-Wide Association summary statistics to identify the SNPs that were independent in our analysis with a genome-wide significant P-value (<5 x 10^{-8}) that are in approximate linkage disequilibrium (LD) with each other at r^2<0.1 and to generate Manhattan and Quantile-Quantile plots and the SNP functional annotations.

Gene analysis and gene-set analysis was performed with MAGMA (v1.07) using FUMA (v1.3.4) and the association analysis summary statistics. Gene expression analysis was obtained from GTEx v7 (https://www.gtexportal.org/home/) integrated by FUMA [80]. More specifically, the gene expression values were log2 transformed average RPKM per

574 tissue type after winsorization at 50 based on GTEx RNA-seq data. Tissue expression
575 analysis was performed for 53 tissue types where the result of gene analysis was tested
576 for one side while conditioning on average expression across all tissue types.
577
578 *LD score regression and genetic correlations.* SNP-heritability was computed with LD
579 Score regression software [24], and heritability estimates were adjusted to the liability
580 scale based on population prevalence of rhythm deficits of 3.5%-6.5% (Supplementary
581 Table 2, Supplementary Note). We then partitioned heritability of rhythm by functional
582 category and investigated cell-type-specific enrichments using stratified LD score
583 regression as per [24]. The Bonferroni-corrected p-value was $0.05/1015=4.9 \times 10^{-5}$.
584 The set of human accelerated regions (HARs) was taken from [30]. All variants in
585 perfect LD ($r^2 = 1.0$ in 1000 Genomes European individuals) with variants in HARs were
586 considered in the analysis. Similarly, variants tagging Neanderthal introgressed
587 haplotypes were defined as in [81]. All variants in perfect LD with a Neanderthal tag SNP
588 were considered Neanderthal variants. For each set, we performed stratified LDSC
589 (v1.0.0) with European LD scores and the baseline LD-score annotations v2.1. The
590 heritability enrichment is defined as the proportion of heritability explained by SNPs in
591 the annotation divided by the proportion of SNPs in the annotation. Standard effect size
592 (τ_c^*), which quantifies the effects unique to the annotation, is the proportionate change in
593 per-SNP heritability associated with a one standard deviation increase in the value of
594 the annotation, conditional on other annotations in the baseline v2.1 model [82].
595 Genetic correlations between rhythm and other complex traits were estimated
596 using LDSC through LD Hub v1.9.0 (http://ldsc.broadinstitute.org/ldhub/) [16] and publicly

597 available GWAS summary statistics. 764 traits were examined and the Bonferroni
598 corrected p-value threshold for significance was $0.05/764=6.5 \times 10^{-5}$. To examine
599 whether the genetic correlations are influenced by residual population stratification, we
600 adjusted the rhythm GWAS summary statistics for the SNP PC-loadings of all top 10
601 PCs. PC loadings were generated from the 1000 Genomes Project because individual-
602 level genotype data was unavailable on the analysed sample [83], following [41].
603
604 We used the gsmr R-package (gcta version:v1.92.1beta6) to implement Generalised
605 Summary-data-based Mendelian Randomization to test for causal genetic associations
606 [38]; see Supplementary Note.
607
608 *Conditional analyses.*
609 To control for pleiotropy between cognition and rhythm abilities(*23*) and identify genetic
610 effects of rhythm traits above and beyond those shared with IQ, we ran a multi-trait
611 conditional and joint analysis (mtCOJO) [38], conditioning on intelligence using GWAS
612 summary statistics from [39].
613
614

Acknowledgments: We are grateful to 23andMe participants for their contribution to the study and to Nancy Cox for suggestions and insight throughout the process. We would also like to thank Michaela Novakovic, Yune Lee, and Duane Watson for input during previous stages of the project and Vanderbilt Trans-Institutional Programs for sparking the initial collaboration. Members of the 23andMe Research Team are: Michelle Agee, Stella Aslibekyan, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sarah L. Elson, Kipper Fletez-Brant, Pierre Fontanillas, Nicholas A. Furlotte, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Steven J. Pitts, G. David Poznik, Anjali J. Shastri, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, and Xin Wang.

Funding:

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number DP2HD098859 and by the National Institute On Deafness And Other Communication Disorders of the National Institutes of Health under Award Number K18DC017383. JAC was supported by the National Institutes of Health (R35GM127087). The content is solely the responsibility of the authors and does not

necessarily represent the official views of the National Institutes of Health. MN was supported by the Wellcome Trust (110222/Z/15/Z).

The dataset used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at https://victr.vanderbilt.edu/pub/biovu/. Also, The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 07/26/19 and dbGaP accession number phs000424.vN.pN on 07/26/19 .

Author contributions

Conceptualization of study: Reyna Gordon, Lea Davis

Study design of GWAS and design of other genomic analyses:

Lea Davis, Reyna Gordon, J. Fah Sathirapongsasuti, Maria Niarchou, Tony Capra, David Hinds

Data collection of genomic data: J. Fah Sathirapongsasuti, The 23andMe Research Team, David Hinds

Genome-wide Association analysis: J. Fah Sathirapongsasuti, The 23andMe Research Team

Post-association QC and generation of figures: Peter Straub and Maria Niarchou

Post-GWAS analyses (heritability, gene-based analyses, gene set analyses, LD correlations, GSMR, mtcojo, PGS in BioVU):

Maria Niarchou, Reyna Gordon, Peter Straub, Lea Davis

HARS and Neanderthal introgression analyses and interpretation: Evonne McArthur, John A Capra

Sensitivity Analyses of Chromosome 17 inversion and Parkinson's Disease: J. Fah Sathirapongsasuti

Estimation of population prevalence of rhythm deficits: Miriam A Mosing and Reyna Gordon

Phenotype validation study design and materials: Reyna Gordon, J. Devin McAuley, Nori Jacoby

Phenotype validation data collection: Nori Jacoby, Eamonn Bell

Phenotype validation study data analysis: Eamonn Bell, Nori Jacoby, Peter Straub, Maria Niarchou, Reyna Gordon

Interpretation of data, writing, editing, and reviewing drafts: All authors

Project Supervision: Reyna Gordon, Lea Davis, David Hinds

Competing interests: JFS, DH, and members of the 23andMe Research Team are employees of 23andMe, Inc., and hold stock or stock options in 23andMe. All other authors declare no competing interests.

Data and materials availability: The full GWAS summary statistics for the 23andMe dataset will be made available through 23andMe to qualified researchers under an agreement that protects the privacy of the 23andMe participants. Please visit

687 research.23andme.com/collaborate/#publication for more information and to apply to
688 access the data. The data and code from the phenotype validation study, and all of the
689 code from the post-GWAS analyses, are also available upon reasonable request.
690

References

1 Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci U S A* **112**, 8987-8992, doi:10.1073/pnas.1414495112 (2015).

2 Trehub, S. E. Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences* **112**, 8809-8810, doi:10.1073/pnas.1510724112 (2015).

3 Nazzi, T., Bertoncini, J. & Mehler, J. Language discrimination by newborns: toward an understanding of the role of rhythm. *J Exp Psychol Hum Percept Perform* **24**, 756-766 (1998).

4 Cirelli, L. K., Trehub, S. E. & Trainor, L. J. Rhythm and melody as social signals for infants. *Annals of the New York Academy of Sciences* **1423**, 66-72 (2018).

5 McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M. & Miller, N. S. The time of our lives: life span development of timing and event tracking. *J Exp Psychol Gen* **135**, 348-367, doi:10.1037/0096-3445.135.3.348 (2006).

6 Patel, A. D. & Iversen, J. R. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in systems neuroscience* **8**, 57 (2014).

7 Kotz, S. A., Ravignani, A. & Fitch, W. T. The Evolution of Rhythm Processing. *Trends in Cognitive Sciences* **22**, 896-910, doi:https://doi.org/10.1016/j.tics.2018.08.002 (2018).

8 Grahn, J. A. & McAuley, J. D. Neural bases of individual differences in beat perception. *NeuroImage* **47**, 1894-1903 (2009).

9 Law, L. N. & Zentner, M. Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PloS one* **7**, e52508 (2012).

10 Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one* **9**, e89642 (2014).

11 Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M. & Fitch, W. T. Finding the beat: a neural perspective across humans and non-human primates. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140093, doi:10.1098/rstb.2014.0093 (2015).

12 Pulli, K. *et al.* 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 (2008).

13 Ullén, F., Mosing, M. A., Holm, L., Eriksson, H. & Madison, G. Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Personality and Individual Differences* **63**, 87-93 (2014).

14 Gingras, B., Honing, H., Peretz, I., Trainor, L. J. & Fisher, S. E. Defining the biological bases of individual differences in musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140092, doi:doi:10.1098/rstb.2014.0092 (2015).

15 Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk of complex disease. *Current opinion in genetics & development* **18**, 257-263 (2008).

16 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).
17 Mosing, M. A., Verweij, K. J., Madison, G. & Ullén, F. The genetic architecture of correlations between perceptual timing, motor timing, and intelligence. *Intelligence* **57**, 33-40 (2016).
18 Peretz, I. & Vuvan, D. T. Prevalence of congenital amusia. *European Journal of Human Genetics* **25**, 625 (2017).
19 Grahn, J. A. & Brett, M. Rhythm and beat perception in motor areas of the brain. *Journal of cognitive neuroscience* **19**, 893-906 (2007).
20 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
21 Oikkonen, J., Onkamo, P., Järvelä, I. & Kanduri, C. Convergent evidence for the molecular basis of musical traits. *Scientific Reports* **6**, 39707, doi:10.1038/srep39707
https://www.nature.com/articles/srep39707 - supplementary-information (2016).
22 Oikkonen, J. *et al.* A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular psychiatry* **20**, 275 (2015).
23 Park, H. *et al.* 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 (2012).
24 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228 (2015).
25 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
26 Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
27 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature genetics* **25**, 25 (2000).
28 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338, doi:10.1093/nar/gky1055 (2019).
29 Mithen, S. J. *The Singing Neanderthals: The origins of music, language, mind, and body.*, (Harvard University Press, 2005).
30 Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130025, doi:10.1098/rstb.2013.0025 (2013).
31 Hubisz, M. J. & Pollard, K. S. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr Opin Genet Dev* **29**, 15-21, doi:10.1016/j.gde.2014.07.005 (2014).
32 Doan, R. N. *et al.* Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* **167**, 341-354 e312, doi:10.1016/j.cell.2016.08.071 (2016).

33 Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75, doi:10.1016/j.ajhg.2018.11.008 (2019).

34 Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature communications* **9**, 2098 (2018).

35 Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357, doi:10.1038/nature12961 (2014).

36 Couth, S., Mazlan, N., Moore. D., Munro, K. & Dawes, P. Hearing difficulties and tinnitus in construction, agricultural, music and finance industries: contributions of demographic, health and lifestyle factors. *Trends in Hearing* (in press).

37 Mosing, M. A., Madison, G., Pedersen, N. L. & Ullén, F. Investigating cognitive transfer within the framework of music practice: Genetic pleiotropy rather than causality. *Developmental Science* **19**, 504-512 (2016).

38 Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 224, doi:10.1038/s41467-017-02317-2 (2018).

39 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics* **50**, 912 (2018).

40 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).

41 Bhatia, G. *et al.* Correcting subtle stratification in summary association statistics. *bioRxiv*, 076133 (2016).

42 Steinberg, K. M. *et al.* Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet* **44**, 872-880, doi:10.1038/ng.2335 (2012).

43 Grahn, J. A. & Brett, M. Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex* **45**, 54-61 (2009).

44 Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* **50**, 6-11, doi:10.1038/s41588-017-0013-8 (2018).

45 Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920-927, doi:10.1038/s41588-018-0151-7 (2018).

46 Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).

47 Lévy, J. *et al.* Molecular and clinical delineation of 2p15p16. 1 microdeletion syndrome. *American Journal of Medical Genetics Part A* **173**, 2081-2087 (2017).

48 Vicente, C. T., Revez, J. A. & Ferreira, M. A. R. Lessons from ten years of genome-wide association studies of asthma. *Clinical & Translational Immunology* **6**, e165, doi:10.1038/cti.2017.54 (2017).

49 Grahn, J. A. & Rowe, J. B. Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. *The Journal of Neuroscience* **29**, 7540-7548, doi:10.1523/jneurosci.2018-08.2009 (2009).

50 Grahn, J. A. & Rowe, J. B. Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex* **23**, 913-921 (2012).

51 Kung, S.-J., Chen, J. L., Zatorre, R. J. & Penhune, V. B. Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *Journal of Cognitive Neuroscience* **25**, 401-420 (2013).

52 Bengtsson, S. L. *et al.* Listening to rhythms activates motor and premotor cortices. *Cortex* **45**, 62-71, doi:https://doi.org/10.1016/j.cortex.2008.07.002 (2009).

53 Teki, S., Grube, M., Kumar, S. & Griffiths, T. D. Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *The Journal of Neuroscience* **31**, 3805-3812, doi:10.1523/jneurosci.5561-10.2011 (2011).

54 McAuley, J. D., Henry, M. J. & Tkach, J. Tempo mediates the involvement of motor areas in beat perception. *Annals of the New York Academy of Sciences* **1252**, 77-84 (2012).

55 Dissanayake, E. If music is the food of love, what about survival and reproductive success? *Musicae scientiae* **12**, 169-195 (2008).

56 Mosing, M. A., Pedersen, N. L., Madison, G. & Ullén, F. Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PLoS One* **9**, e113874 (2014).

57 Border, R. *et al.* No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry* **176**, 376-387 (2019).

58 Zuk, J., Benjamin, C., Kenyon, A. & Gaab, N. Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one* **9**, e99868 (2014).

59 Medina, D. & Barraza, P. Efficiency of attentional networks in musicians and non-musicians. *Heliyon* **5**, e01315 (2019).

60 Vaag, J., Saksvik-Lehouillier, I., Bjørngaard, J. H. & Bjerkeset, O. Sleep difficulties and insomnia symptoms in norwegian musicians compared to the general population and workforce. *Behavioral sleep medicine* **14**, 325-342 (2016).

61 Jansen, E., Helleman, H., Dreschler, W. & de Laat, J. Noise induced hearing loss and other hearing complaints among musicians of symphony orchestras. *International archives of occupational and environmental health* **82**, 153-164 (2009).

62 Finkel, D. & Bravell, M. Genetic influences on lung function contribute to subsequent age changes in motor and cognitive function. *Behavioral Genetics* (2019).

63 van den Elzen, N. *et al.* The Power of Music: Enhancing Muscle Strength in Older People. *Healthcare (Basel)* **7**, doi:10.3390/healthcare7030082 (2019).

64 Haegens, S. & Golumbic, E. Z. Rhythmic facilitation of sensory processing: a critical review. *Neuroscience & Biobehavioral Reviews* **86**, 150-165 (2018).

65 Tung, J. Y. *et al.* Efficient replication of over 180 genetic associations with self-reported medical data. *PloS one* **6**, e23473 (2011).

66 Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J. & Rodriguez-Fornells, A. Individual differences in music reward experiences. *Music Perception: An Interdisciplinary Journal* **31**, 118-138 (2013).

67 Sowiński, J. & Dalla Bella, S. Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia* **51**, 1952-1963 (2013).

68 Oikkonen, J. *et al.* A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular Psychiatry* **20**, 275, doi:10.1038/mp.2014.8 https://www.nature.com/articles/mp20148 - supplementary-information (2014).

69 Povel, D.-J. & Essens, P. Perception of temporal patterns. *Music Perception: An Interdisciplinary Journal* **2**, 411-440 (1985).

70 Gordon, R. L. *et al.* Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science* **18**, 635-644 (2015).

71 Wieland, E. A., McAuley, J. D., Dilley, L. C. & Chang, S.-E. Evidence for a rhythm perception deficit in children who stutter. *Brain Lang* **144**, 26-34 (2015).

72 Woods, K. J., Siegel, M. H., Traer, J. & McDermott, J. H. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* **79**, 2064-2072 (2017).

73 Macmillan, N. A. & Creelman, C. D. *Detection theory: A user's guide*. (Psychology press, 2004).

74 Durand, E. Y., Do, C. B., Mountain, J. L. & Macpherson, J. M. Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*, 010512, doi:10.1101/010512 (2014).

75 Henn, B. M. *et al.* Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* **7**, e34267 (2012).

76 Consortium, G. P. A global reference for human genetic variation. *Nature* **526**, 68 (2015).

77 consortium, U. K. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82 (2015).

78 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**, 1084-1097 (2007).

79 Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nature genetics* **48**, 811 (2016).

80 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: Functional mapping and annotation of genetic associations. *bioRxiv*, doi:10.1101/110023 (2017).

81 Vernot, B. *et al.* Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235-239, doi:10.1126/science.aad9416 (2016).

82 Hujoel, M. L. A., Gazal, S., Hormozdiari, F., van de Geijn, B. & Price, A. L. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. *Am J Hum Genet* **104**, 611-624, doi:10.1016/j.ajhg.2019.02.008 (2019).

919 83 The Genomes Project, C. *et al.* A global reference for human genetic variation.
920 *Nature* **526**, 68, doi:10.1038/nature15393
921 https://www.nature.com/articles/nature15393 - supplementary-information (2015).
922

FIGURES

Figure 1. Manhattan plot of GWAS results for rhythm. Results are shown for 606,825 research participants with 23andMe who responded Yes (N=555,660) vs. No (N=51,165) to the question “Can you clap in time to a musical beat?”, controlling for age, sex, top 5 PC’s for ancestry, and genotype platform. The x-axis shows chromosomal position and the y-axis shows -log10 p-values). 68 loci surpassed the threshold for genome-wide significance of $p<5x10^{-8}$ (dotted horizontal line). For illustration purposes, we only included 500,000 SNPs with p<0.1

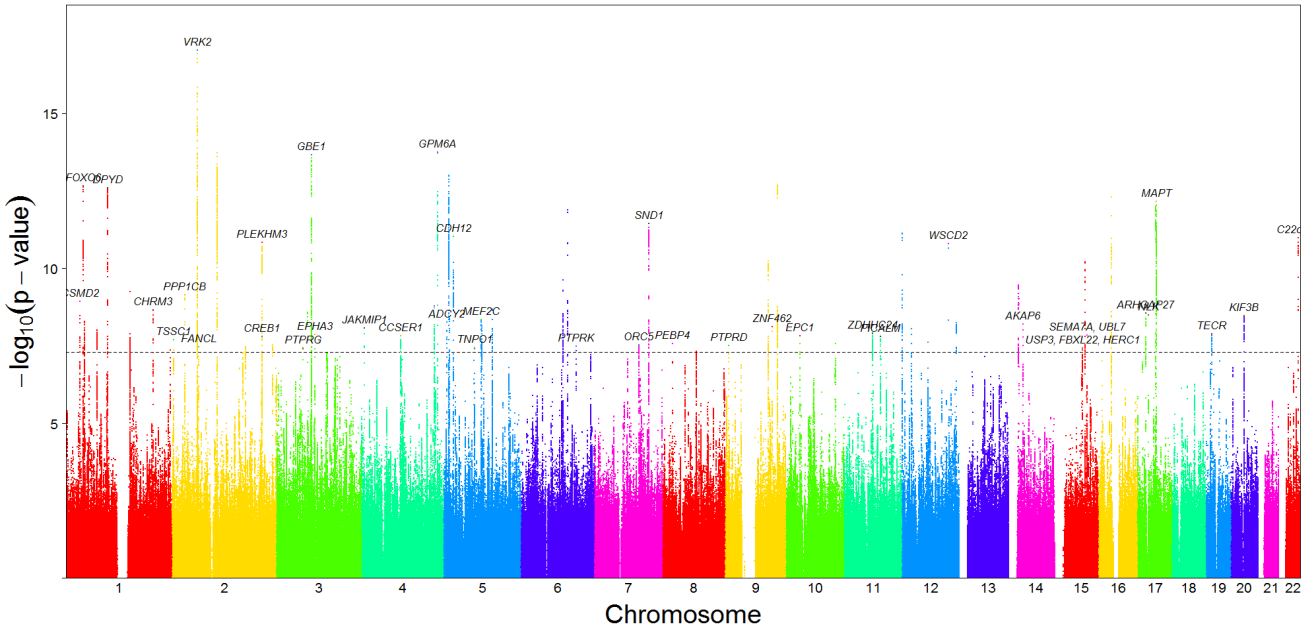


Figure 2. Gene expression profiles for genes associated with rhythm in 53 tissue types from the GTEx database. Gene property analysis was conducted in FUMA on the MAGMA gene-based results, in which the average expression per gene was added as a co-variate to the model. Associations with rhythm were significantly enriched in brain tissue compared to other tissues (-log-10 p-values are on y-axis, with type on x-axis).

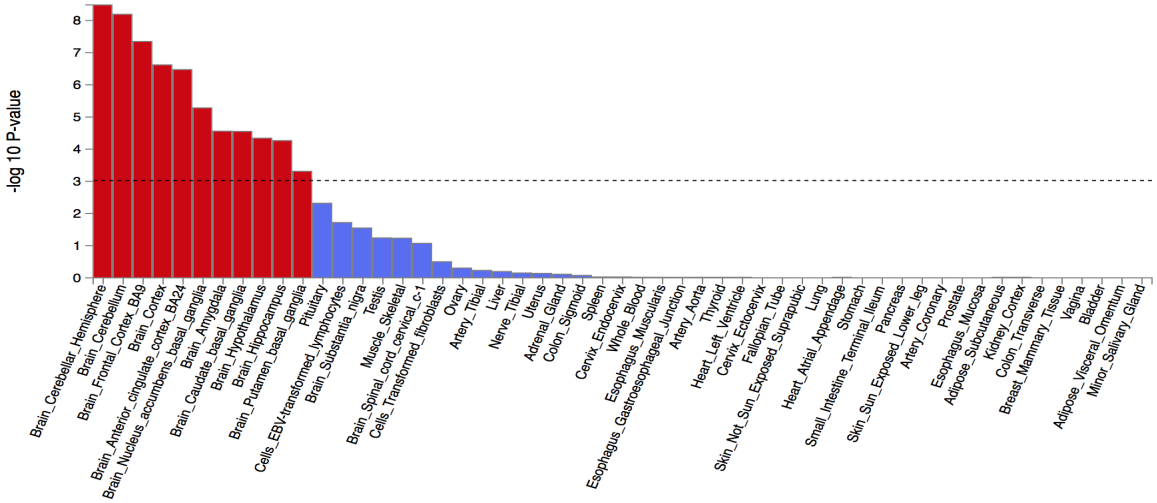


Figure 3. Polygenic score of rhythm predicts presence of musician-related keywords in electronic health record (EHR) biobank. With weights derived from the 23andMe GWAS, we applied polygenic scores (PGS) for rhythm to genomic data from N=864 individuals in Vanderbilt's BioVU whose EHR contained "musician" and related keywords, and compared to a control sample of N=66,577 (See Supplementary Note). PGS-rhythm were higher for musician vs. controls (OR=1.3) but did not predict 3 negative-control traits, shown here.

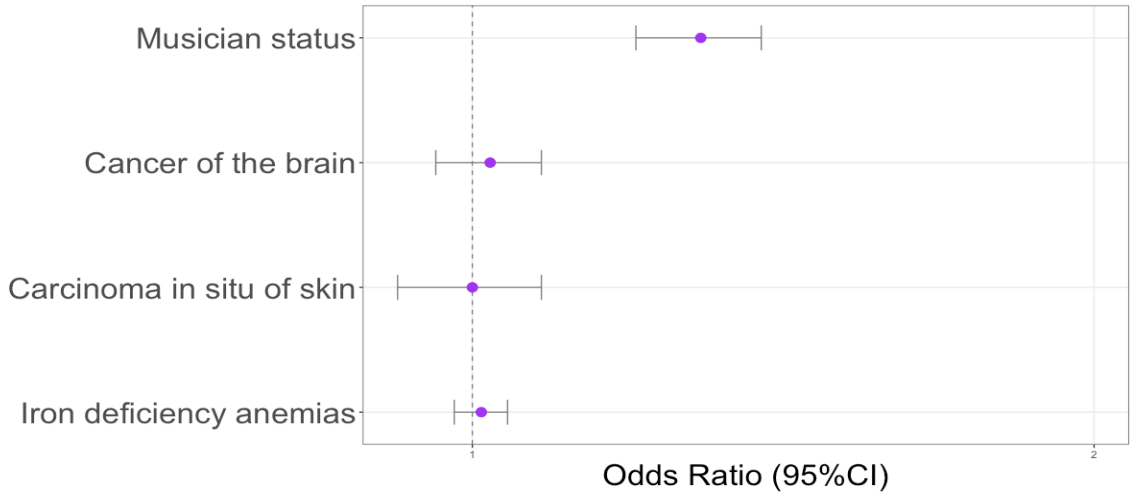
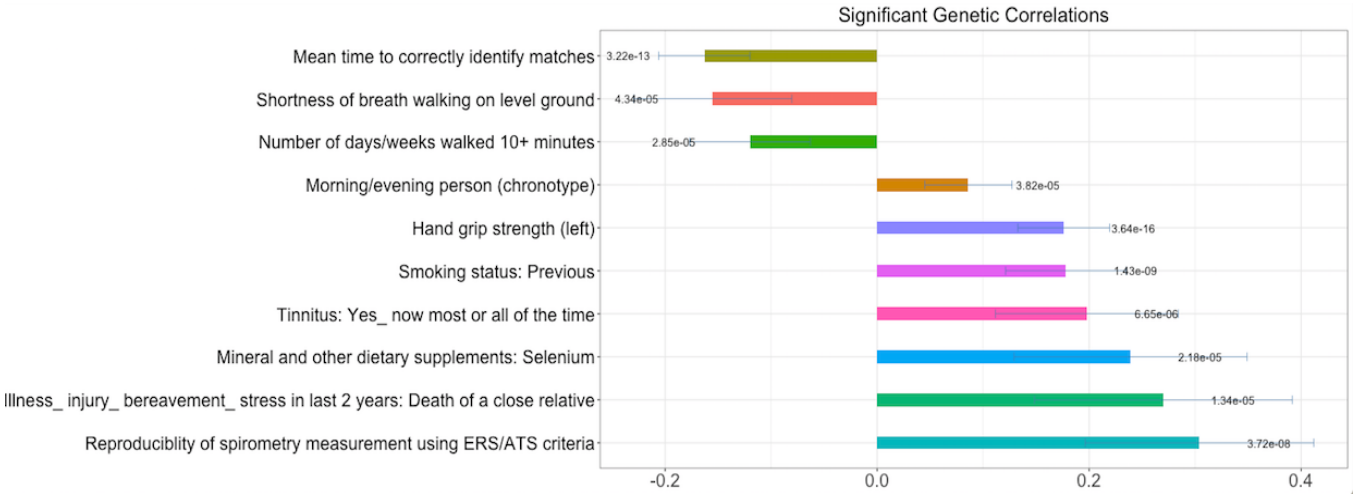


Figure 4. Selected results from the cross-trait genetic correlations analysis. Rhythm summary statistics were significantly genetically correlated with several health, cognitive, emotion, and circadian phenotypes in our analysis comparing rhythm to traits available in LDHub. Full results are presented in Supplementary Table 13. The y-axis is the rg correlation, the bars represent standard errors and the p-values are also indicated.



Tables

Table 1. Number of participants in each rhythm phenotype group, broken down by sex and age range in the GWAS sample.

Phenotype group (response to Clap-to-beat question)	**Total N**	**Males**	**Females**	**0 to 30 years old**	**30 to 45 years old**	**45 to 60 years old**	**60 years old and over**
Yes (cases)	555660	226188	329472	57898	135168	150939	211655
No (controls)	51165	23998	27167	5186	12909	13312	19758

975 Table 2. Genomic loci associated with rhythm after adjusting for intelligence and traits
976 with which they have been previously associated (loci in pink do not exist in GWAS
977 catalog)
978

SNPs after IQ cojo	chr	p-value	GWAS catalog (yes/no)	GWAS catalog
rs34762587	1	2.21584E-13	no	
rs2635634	5	9.96339E-12	no	
rs12913592	15	6.59873E-11	no	
rs2061843	1	1.08739E-09	no	
rs4443239	4	1.5082E-09	no	
rs1901739	5	2.15026E-09	no	
rs8079923	17	3.11892E-09	no	
rs7501911	17	3.99553E-09	no	
rs2284901	14	7.46306E-09	no	
rs10978661	9	9.75689E-09	no	
rs4263335	4	9.25137E-09	no	
rs7939759	11	1.45292E-08	no	
rs9710427	19	1.26321E-08	no	
rs12638746	3	1.56044E-08	no	
rs2505344	10	1.6844E-08	no	
rs6548147	2	2.0951E-08	no	
rs10885458	12	2.56522E-08	no	
rs10877461	10	2.61599E-08	no	
rs1996148	8	2.90685E-08	no	
rs12056186	7	2.93356E-08	no	
rs7856850	9	3.38293E-08	no	
rs4704043	5	3.49958E-08	no	
rs43182	3	4.50456E-08	no	
rs7875397	9	2.21728E-13	no	
rs2018545	14	3.58572E-10	no	
rs7002174	8	4.75329E-08	no	
rs34863893	5	8.60406E-10	no	
rs7297439	12	5.07484E-09	no	
rs72826882	17	2.58837E-09	no	
rs7625774	3	3.12812E-09	no	
rs848293	2	1.01057E-17	yes	depression, neuroticism, schizophrenia
rs62340585	4	2.2504E-14	yes	neuroticism, schizophrenia, highest math class taken
rs10168817	2	1.83859E-14	yes	reaction time, worry, pneumonia, phonic sneeze reflex, reaction time, HIV-1 viral set point
rs10779987	3	2.04673E-14	yes	hand grip strength
rs10875125	1	3.43636E-13	yes	educational attainment, schizophrenia, ASD, ADHD, MDD
rs4792891	17	7.06239E-13	yes	general cognitive ability, depressed affect, neuroticism
rs1468701	7	2.72542E-12	yes	intelligence
rs9626920	22	9.91369E-12	yes	self-reported risk-taking behaviour, smoking status, dupuytren's disease
rs764299	2	1.47E-11	yes	educational attainment, highest math class taken, cognitive performance
rs1426371	12	1.34602E-11	yes	extraversion, worry
rs7586405	2	7.19E-10	yes	heel bone mineral density
rs55678522	1	2.81431E-09	yes	glioma, non-glioblastoma glioma, highest math class taken, general cognitive ability, educational attainment, intelligence
rs6087848	20	3.75408E-09	yes	inflammatory bowel disease, Crohn's disease, cognitive performance,
rs13163173	5	6.08339E-09	yes	platelet count, intelligence, depressive symptoms, subjective well-being, depressed affect, neuroticism, BMI, conscientiousness, depression

rs67264739	5	5.89821E-09	yes	adolescent idiopathic scoliosis
rs1596431	5	5.77642E-09	yes	intelligence
rs12909047	15	1.46255E-08	yes	caffeine metabolism (plasma 1,3, 7, -trimethylxanthine (caffeine)level)
rs67816799	4	1.64433E-08	yes	educational attainment
rs10932201	2	1.55342E-08	yes	systolic blood pressure
rs526904	11	1.54544E-08	yes	Alzheimer's disease, family history of Alzheimer's disease,
rs13197257	6	2.77659E-08	yes	reaction time, educational attainment, general cognitive ability, highest math class taken
rs11692449	2	3.2277E-08	yes	immature fraction of reticulocytes
rs62014217	15	3.1548E-08	yes	atrial fibrillation, urinary albumin excretion
rs476141	1	4.68634E-08	yes	diabetic retinopathy
rs7715357	5	1.30965E-13	yes	smoking status
rs11865086	16	5.94369E-13	yes	blood protein levels, menarche(age at onset)
rs1536057	6	1.35469E-12	yes	platelet distribution width, smoking status (ever vs. never smokers), smoking initiation (ever regular vs. never regular), anxiety/tension(special factor of neuroticism)
rs1972582	12	7.90101E-12	yes	chronic kidney disease, glomerular filtration rate(creatinine), blood metabolite levels, chronotype
rs3780420	9	6.39377E-11	yes	quantitative traits, platelet distribution width, plateletcrit, platelet count
rs9385269	6	1.96812E-10	yes	educational attainment, cognitive performance, intelligence, regular attendance at a pub or social club, Tourette's syndrome, educational attainment, bipolar disorder, highest math class taken, self-reported math ability, cognitive function, alcohol consumption (drinks per week), general risk tolerance, extremely high intelligence, autism spectrum disorder, QT interval, risk taking tendency
rs16837903	1	4.74528E-10	yes	monocyte chemoattractant protein-1 levels, lung function
rs778353	2	3.01299E-08	yes	heel bone mineral density
rs10497357	2	3.42377E-08	yes	longitudinal change in brain amyloid plaque burden
rs2467452	12	9.8837E-09	yes	automobile speeding propensity, chronotype, morning person
rs6684973	1	1.01877E-08	yes	schizophrenia, depression, smoking initiation, smoking status, alcohol consumption self-reported math ability, highest math class taken
rs2819336	1	4.59116E-09	yes	smoking cessation, smoking initiation, cognitive ability, intelligence, general cognitive ability, cognitive performance, age of smoking initiation, hypertension risk in short sleep duration, menarche (age of onset), red blood cell count, educational attainment, self-reported math ability, highest math class taken, attention deficit hyperactivity disorder, male-pattern baldness

979