



Genome-Wide Homozygosity Mapping Reveals Genes Associated With Cognitive Ability in Children From Saudi Arabia

Sergey A. Kornilov^{1,2}, Mei Tan², Abdullah Aljughaiman³, Oxana Yu Naumova^{2,4} and Elena L. Grigorenko^{1,2,5,6*}

¹ Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX, United States, ² Department of Psychology, University of Houston, Houston, TX, USA, ³ Special Education Department, King Faisal University, Ahsaa, Saudi Arabia, ⁴ Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia, ⁵ Child Study Center, Yale University, New Haven, CT, USA, ⁶ Moscow State University for Psychology and Education, Moscow, Russia

OPEN ACCESS

Edited by:

Michelle Luciano,
University of Edinburgh,
United Kingdom

Reviewed by:

Juko Ando,
Keio University,
Japan
Pippa Ann Thomson,
Medical Research Council Institute
of Genetics and Molecular Medicine
(MRC), United Kingdom

*Correspondence:

Elena L. Grigorenko
elena.grigorenko@times.uh.edu

Specialty section:

This article was submitted to
Behavioral and Psychiatric Genetics,
a section of the journal
Frontiers in Genetics

Received: 23 October 2018

Accepted: 22 August 2019

Published: 18 September 2019

Citation:

Kornilov SA, Tan M, Aljughaiman A,
Naumova OY and Grigorenko EL
(2019) Genome-Wide Homozygosity
Mapping Reveals Genes Associated
With Cognitive Ability in Children
From Saudi Arabia.
Front. Genet. 10:888.
doi: 10.3389/fgene.2019.00888

Recent studies of the genetic foundations of cognitive ability rely on large samples (in extreme, hundreds of thousands) of individuals from relatively outbred populations of mostly European ancestry. Hypothesizing that the genetic foundation of cognitive ability depends on the broader population-specific genetic context, we performed a genome-wide association study and homozygosity mapping of cognitive ability estimates obtained through latent variable modeling in a sample of 354 children from a consanguineous population of Saudi Arabia. Approximately half of the sample demonstrated significantly elevated homozygosity levels indicative of inbreeding, and among those with elevated levels, homozygosity was negatively associated with cognitive ability. Further homozygosity mapping identified a specific run, inclusive of the *GRIA4* gene, that survived corrections for multiple testing for association with cognitive ability. The results suggest that in a consanguineous population, a notable proportion of the variance in cognitive ability in the normal range in children might be regulated by population-specific mechanisms such as patterns of elevated homozygosity. This observation has implications for the field's understanding of the etiological bases of intelligence and its variability around the world.

Keywords: intelligence, cognitive ability, genome-wide association study, homozygosity mapping, consanguinity, *GRIA4*

INTRODUCTION

Studies of the genetic foundations of cognitive ability published in the last decade have made significant progress in elucidating the complex etiology of cognitive ability as a human trait that is influenced by a multitude of genetic factors. Two recent meta-analyses are particularly noteworthy, as they signify a leap in sample size, to $n = 78,308$ (Sniekers et al., 2017) and $n = 300,486$ (Davies et al., 2018), which may generate enough statistical power to detect small-effect-size genetic loci that constitute some of these factors. The former (Sniekers et al., 2017) revealed 18 independent genome-wide significant loci associated with 336 single nucleotide polymorphisms (SNPs), with the majority being previously unreported, although many of the newly implicated genes (e.g., *MEF2C* and *SHANK3*) are known for their functions in neural development. The latter (Davies

et al., 2018) is the largest meta-analysis of cognitive ability to date (combining the CHARGE and COGENT consortia and the UK Biobank data); it implicated 148, including 58 newly reported, loci. The derived polygenic scores predicted up to 4.31% of the variance in cognitive ability, and several pathways, including the positive regulation of nervous system development, were overrepresented among the genes highlighted by this meta-genome-wide association study (GWAS).

These findings are supported by results from recent large-scale GWAS studies of educational attainment that indicate that this proxy phenotype for cognitive ability is also governed by substantial genetic control. Genome-wide significant findings in the studied cohorts have identified biologically plausible pathways for cognitive ability [e.g., among the 74 loci identified, *MEF2C* was also significantly associated with educational attainment (Okbay et al., 2016)], and derived polygenic scores explained up to 2% of the variance in the trait (Rietveld et al., 2013). Polygenic scores for educational attainment also predicted cognitive behavioral phenotypes across the lifespan in another longitudinal study (Belsky et al., 2016).

This steady stream of findings implicating multiple small-effect loci as contributors to the genetic control of cognitive ability puts emphasis on the substantial statistical power required for their reliable detection under the common variant hypothesis. Notably, around 50% of the estimated chip heritability [i.e., the proportion of variance in ability estimates cumulatively explained by the modeled genetic distances between individuals using a large number of genotyped markers (Plomin and Von Stumm, 2018)] can be attributed to variants located in genomic regions that are under negative selection (Hill et al., 2016). Concordantly, rare functional coding alleles, when associated with cognitive ability, are more likely to be detrimental rather than beneficial (Spain et al., 2016) to it.

The large-scale studies mentioned earlier were performed in samples drawn from populations of European ancestry, characterized by relatively low levels of homozygosity due to the low prevalence of consanguinity. Much less used are samples with genetic backgrounds characterized by excessive homozygosity. Given that cryptic parental relatedness and consanguinity have been found to be significantly negatively associated with cognitive ability (Howrigan et al., 2016), as well as severity of syndromic intellectual disability (Gandin et al., 2015) even in relatively outbred individuals, homozygosity can be considered an additional window into the genetic regulation of intelligence. This hypothesis is supported by the presence of significant negative associations between SNP-based inbreeding coefficients and a range of traits (including educational attainment) in a sample of approximately 5,500 unrelated Finnish individuals (Verweij et al., 2014) and, complementarily, the absence of association between runs of homozygosity (ROH) burden and cognitive ability in a UK sample of $n = 2,329$ unrelated individuals (Power et al., 2014b).

More recent and larger studies of cohorts where elevated autozygosity is not observed provide further evidence for the role of homozygosity in cognitive ability. In a well-powered study of the effects of homozygosity on 16 health-related traits in a sample of $n = 354,224$ individuals, elevated homozygosity

was associated with decreased cognitive ability across study cohorts (Joshi et al., 2015). Similarly, autosomal homozygosity burden was associated with lower general intelligence in the UK Biobank sample of almost 400,000 individuals (Johnson et al., 2018). Thus, the particular constellation of factors involved in the genetic regulation of cognitive ability might depend on a complex polygenic genomic background, that is, population-specific mechanisms such as increased homozygosity (i.e., the presence of long stretches of homozygous DNA reflecting parental relatedness that may have deleterious effects on gene function in a recessive fashion) potentially account for additional variance in cognitive ability that has not been captured by individual and even gene-based analyses in populations not enriched for homozygosity-related haplotypic variation.

The goal of this study was to examine patterns in the genetic foundation of cognitive ability in a sample of children from the Kingdom of Saudi Arabia (KSA), a country that historically has had a significant [i.e., up to 57.7% prevalence in one study (El-Hazmi et al., 1995)] and stable (Warsy et al., 2014) number of consanguineous marriages. The substantial negative effects of consanguinity and homozygosity on Mendelian traits and the elevated prevalence of rare recessive diseases and disorders are well-established in the Gulf countries, including in recent reports from Saudi Arabia (Alfares et al., 2017; Majid et al., 2017) and Qatar (Bener and Mohammad, 2017). To date, however, no study has directly examined genome-wide homozygosity patterns with respect to their contribution to cognitive ability in Middle Eastern countries (for a review of published studies, see Ceballos et al., 2018b), despite the rising awareness of the possible contribution of consanguinity-driven homozygosity not only to rare Mendelian disorders and intellectual disability in consanguineous populations but also to the etiology of complex traits and common disorders (Erzurumluoglu et al., 2016).

Given the recent findings from exome-sequencing studies in Saudi Arabia that documented a high diagnostic yield of variants associated with a plethora of disorders (Alfares et al., 2017; Monies et al., 2017), among them intellectual disability, we hypothesized that individual differences in cognitive ability in this population might be partly regulated by regions of homozygosity. We also hypothesized that these regions would be placed in-between common and rare variants on the continuum of effect sizes and that elevated levels of homozygosity (and allelic ROHs) would increase our statistical power to detect such effects even in a sample of modest size. This hypothesis is contextualized by 1) observations that long ROHs are enriched for deleterious variation that amplify the effects of mildly deleterious variants that exist in a homozygous state (Szpiech et al., 2013) and 2) recent successes in the homozygosity mapping of neurodevelopmental attention-deficit/hyperactivity disorder in a small sample of Saudi siblings with attention-deficit/hyperactivity disorder, which identified several new candidates for the disorder in a sample of limited size (Shinwari et al., 2017). Therefore, by capitalizing on the unique nature of the sample from the consanguineous population of KSA, recent advances in model-based identification of long tracts of homozygosity (Ceballos et al., 2018a; Ceballos et al., 2018b), and quantification of latent cognitive ability trait estimates via structural equation modeling,

we sought to perform a GWAS and homozygosity mapping of cognitive ability in a sample of children from Saudi Arabia using a genotyping platform with increased exonic coverage, enabling the identification of long genic ROHs.

MATERIALS AND METHODS

Participants

Participants for the genetic study were recruited from a larger sample enrolled in an epidemiological study of cognitive ability in children in Saudi Arabia. For this epidemiological study, a total of $n = 7,186$ children in the age range from 7.34 to 18.71 years ($M = 12.28$, $SD = 1.81$; 4,682 males and 2,504 females) were recruited from local schools in seven major regions of Saudi Arabia: Abha ($n = 605$), Khamis-Mushyat ($n = 585$), Tabuk ($n = 1,095$), Al-Jubail ($n = 1,004$), Jeddah ($n = 2,272$), and Al-Hassa ($n = 1,625$).

Using the data from the epidemiological study, $n = 942$ children of Saudi Arabian ancestry were identified for the genetic study by the school officials based on their performance on the cognitive ability battery administered as part of the epidemiological study. Proband selection was performed separately for each grade using the 85th percentile cut-off across the sum of all administered items, and the genetic study sample was therefore expected to be enriched for high general cognitive ability. Thus, although the data reported in this manuscript are based on the sample of children selected to represent the high tail of the cognitive ability distribution, the final cognitive ability estimates used for the genetic association were obtained after a careful examination of the psychometric properties of Aurora and appropriate revisions (i.e., exclusion of misfitting items, see discussion later). Thus, the resulting scores covered a broad range of the distribution in general intelligence, albeit with a notable shift toward higher scores (see **Supplementary Material**).

Saliva was collected from every consenting proband ($n = 354$) in the age range from 8.23 to 13.93 ($M = 11.33$, $SD = 1.36$; 273 or 77% were boys). These children comprised the main sample of the study reported here. All children and their parents were born in KSA to Saudi Arabian parents according to a self-report (see also **Supplementary Table S1** for a more detailed study sample demographics breakdown).

The study protocol and procedures were approved by the Internal Review Boards of Yale University, University of Houston, and King Faisal University (Saudi Arabia). Parents of children participating in the study and authorities at the appropriate educational institutes provided written informed consent in accordance with the Declaration of Helsinki.

Cognitive Ability Assessment

All children were administered a paper-and-pencil cognitive ability assessment in a group format in a controlled school environment. Specifically, we used the Aurora (Chart et al., 2009; Tan et al., 2009; Hein et al., 2015) assessment to obtain estimates of cognitive ability in the large sample of $n = 7,186$ children to adequately model the distributions of population parameters. The version of the instrument used for this study was developed

taking into account relevant cultural and linguistic factors (Tan et al., 2009) across a number of cultures, countries, and languages. In the absence of convergent or predictive (with respect to, for example, academic outcomes) validity data on the instrument in Saudi Arabia, we performed a careful psychometric analysis and construct validation of the instrument as reported later (more information is available from the authors upon request).

We used the data from the “general” cognitive ability module of Aurora that was designed to capture variability in children’s general cognitive ability and appropriately model domain-specific variation. The test battery includes nine subtests designed to fill a 3 (domain: verbal, numerical, and figural) \times 3 (task: classification, series, and analogies) test design matrix, with the number of items ranging from 8 to 22 per subtest (8 to 18 after item exclusion, see later). The order of the Aurora subtests within the administered testlets was randomized across data collection sites. Children were allowed 45 min to complete the assessment. Test administration was governed by trained assessors. If the child did not attempt a subtest, the data for that subtest were coded as missing.

The calculation of latent ability trait estimates was performed using a two-step approach. First, patterns of responses to test items were scored using item response theory. Specifically, estimates of subtest-level abilities were obtained by fitting a set of 2-parametric models to the binary response vectors as implemented in the R *mirt* (Chalmers, 2012) library, thereby estimating both item difficulty and item discrimination. Local fit indices and item characteristic curves were analyzed, and items were removed from the dataset when local misfit was diagnosed, and when removal resulted in the increase of variance in item response patterns explained by the item response theory model (see also **Supplementary Table S2** for the description of subtest-level parameters).

Second, we performed latent variable modeling of subtest-level ability estimates by fitting a set of alternative models to the data using EQS 6.1. software (Bentler, 2006). The models differed in the number of estimated latent trait variances and covariance, and corresponding pattern and error loadings. Overall, several models displayed good fit to the data, and the bifactor model (see **Figure 1**, also **Table S3**) was chosen as the final model based on theoretical considerations as well as recent data documenting the superior fit of the bifactor model to the data in most cases (Cucina and Byle, 2017). Briefly, the bifactor models postulates the existence of orthogonal specific ability factors that load on their respective items [i.e., in this case—domain-specific factors of verbal (Aurora-V), numerical (Aurora-N), and spatial/figural ability (Aurora-F)], while all items load on the common general factor (Aurora-G). Missing data were modeled using the robust pairwise covariance approach, and latent ability estimates were computed using the generalized least squares estimator.

Estimates of latent ability traits were adjusted for gender, age, and its quadratic term age^2 and the interaction between age and gender terms using the robust implementation of linear mixed modeling in the *robustlmm* (Koller, 2016) package for R. The models were fitted separately for each phenotype (Aurora-V, Aurora-N, Aurora-F, and Aurora-G) and included a school ID variable as a random effect to account for the nested structure

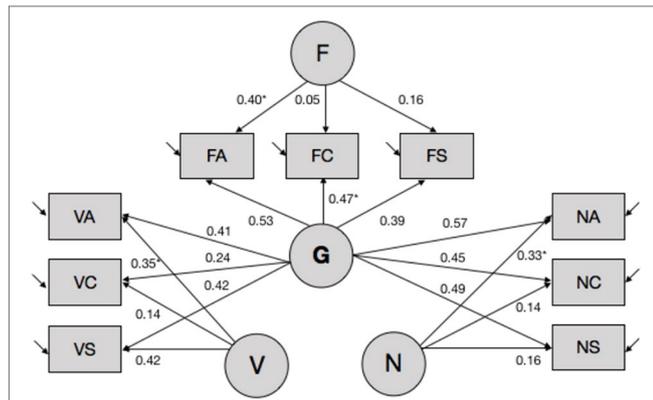


FIGURE 1 | Latent variable model of Aurora. *(s) denote coefficients that were fixed to 1 to enable model identification and parameter estimation. Other parameters were estimated freely. Covariances between latent variables were fixed to 0. Aurora-G, general cognitive ability; Aurora-V, verbal cognitive ability; Aurora-N, numerical cognitive ability; Aurora-F, figural (spatial) cognitive ability. Subtests: VA, verbal analogies; VC, verbal classification; VS, verbal series; NA, numerical analogies; NC, numerical classification; NS, numerical series; FA, figural analogies; FC, figural classification; FS, figural series.

of the data obtained for the total sample of children, thereby increasing the precision of the parameter estimates.

DNA Sample Collection, Processing, and Genotyping

All participating individuals provided a saliva sample using Oragene™ OG-250 (DNA Genotek, Inc) self-collection kits. Saliva samples were processed at Yale University. DNA was extracted using the manufacturer's protocol; DNA integrity was verified by agarose gel electrophoresis, and DNA concentration was determined with the Quant-iT dsDNA High-Sensitivity Assay Kit using a Qubit 2.0 fluorometer (ThermoFisher Scientific, Inc). Genotyping was performed by Illumina, Inc using the FastTrack genotyping service. To minimize batch effects, the placement of samples on individual plates was block-randomized with respect to geographical region/school, gender, age, and cognitive ability (using the median-split on Aurora-G scores).

Samples were genotyped using the HumanCoreExome v1.2 BeadChip microarray that included a total of 550k SNP markers. The HumanCoreExome array is specifically enriched for exonic SNPs and low minor allele frequency (MAF) SNPs. Allele calling was performed using Illumina's GenomeStudio. Individual SNPs were mapped onto the hg19 build; clustering results were examined visually, and clustering was adjusted manually when necessary.

All samples had a call rate > 95% and passed the reported vs. inferred gender check. Marker-based quality control included: removal of Illumina's insertion/deletion markers ($k = 9,497$), removal of markers with call rate < 95% ($k = 2,330$), and removal of markers with $MAF < 1\%$ ($k = 229,664$). Thus, the resulting dataset included $k = 285,340$ autosomal markers, with an estimated density of 1 SNP per 9.965 kb.

Genetic principal components using the additive genotypic model were estimated after the linkage disequilibrium pruning of

the dataset (with r^2 threshold of 0.50, and a window size of 50; $k = 171,399$ markers) using SNP & Variation Suite (GoldenHelix, Inc).

Homozygosity Mapping

Homozygosity mapping was performed using GARLIC (Szpiech et al., 2017). GARLIC detects ROH using a model-based approach and has an advantage over SNP counting methods with respect to the modeling of population-specific parameters, genotyping error rates, and microarray platform SNP density. Peri-centromeric and telomeric regions were excluded from the analyses using hg19 coordinates. The technical replicate-based genotyping error rate of 0.0019681673 was provided to the software with the *-weighted* flag enabled, and panel density-adjusted window size was automatically set to 62 SNPs.

In contrast to genotype counting methods such as implemented in PLINK, GARLIC utilizes the logarithm of odds score measure of homozygosity using a sliding window framework to infer ROHs, with distribution of logarithm of odds scores estimated through the use of a Gaussian kernel density estimator (KDE). KDE was computed using the data from all samples (i.e., KDE thinning was disabled). GARLIC uses a three-component mixture model to classify ROH calls into three categories. Sample ROHs were classified into Short (A; 75.43% of all calls, with an average length of 625 kb) ROHs reflecting homozygosity for ancestral haplotypes; Medium (B; 16.64%; length > 1.3 Mb; average length 1.9 Mb) ROHs; and Long (C; 7.93%, length > 4.48 Mb; average length 11.2 Mb) ROHs, reflecting recent population processes and consanguinity. ROH-driven homozygosity (f_{ROH}) values were estimated separately for each class of ROHs and cumulatively, representing the percent of the mappable autosomal genome covered by ROHs (i.e., proportion out of $9,965 \times 285,340$ bp). A SNP-based homozygosity index f_{SNP} was calculated for each sample using the pruned dataset in SNP & Variation Suite.

Copy Number Variant Calling

Copy number analysis was performed on 353 out of 354 samples after the exclusion of one sample with outlying SD of $\log R$ intensity values. To ensure the comparability of ROH and copy number variant (CNV) calls, probe intensity values, genotypes, and corresponding B allele frequencies were exported for $k = 285,340$ autosomal polymorphic markers that passed SNP-based quality control for the total sample. Probe intensity values were adjusted for genomic waves using the CNV quality control module of SNP & Variation Suite. Genomic wave-adjusted values were then winsorized using the *copynumber* (Nilsen et al., 2012) library for R, and missing values for no-call genotypes were imputed using the piecewise constant segmentation approach (*pcf*) implemented in *copynumber*.

We then used PennCNV (Wang et al., 2007) to perform CNV segmentation and calling. PennCNV uses Hidden Markov Model approach for evaluating copy number states in genome-wide microarray genotyping data. Analyses were performed with the following settings: CNV segments were required to have a minimum length of 10 SNPs and 100 kb; median-adjustment of BAF and probe intensity values was enabled. The population frequency of the B allele (**.pfb*) and the Hidden Markov model file, with the expected intensity values for different copy number states for the

HumanCoreExome panel that we used in the study, was downloaded from the PennCNV website (<http://penncnv.openbioinformatics.org>). Marker locations were verified against Illumina's manifest files.

A total of $k = 1,673$ CNVs ranging in size from 10 kb to 1.3 Mb ($M = 170$ kb, median = 123 kb, SD = 163 kb) were called in the 353 analyzed samples, with 476 or 28% being copy number gain and the remaining 1,197 or 72% being copy number loss events.

Statistical Analyses

A GWAS was performed using a linear mixed modeling approach implemented in GEMMA (Zhou and Stephens, 2012) using a standardized pairwise identity-by-state relatedness matrix to account for population structure and relatedness. P-values were estimated using the Wald method, and the threshold for genome-wide significance was estimated using the M_{eff} approach as implemented in simpleM (Gao et al., 2010) and multiplied by the number of tested phenotypes, with the resulting threshold value of $P = 5.29 \times 10^{-8}$ for four Aurora phenotypes (Aurora-V, Aurora-N, Aurora-F, and Aurora-G) combined.

Gene-based association analyses were carried out with nominal P-values from the GEMMA analysis using the effective chi-square gene-based test as implemented in KGG v4 (Li et al., 2010). Analyses were limited to protein-coding genes, and gene boundaries were padded with 10-kb regions for SNP assignment. P-values were combined for the four phenotypes and adjusted for multiple comparisons jointly using the Benjamini–Hochberg correction (Benjamini and Hochberg, 1995). P-values were combined for the four phenotypes and adjusted for multiple comparisons jointly using the Benjamini–Hochberg (Benjamini and Hochberg, 1995) correction. ROH and CNV association analyses were performed using the linear model, $lm()$, function in R, while controlling for the top five genetic principal components and f_{SNP} homozygosity. P-values were adjusted using the Benjamini–Hochberg correction. Latent class analysis (LCA) of homozygosity was performed using the R *mclust* (Scrucca et al., 2016) library. The goal of the LCA as an unsupervised data mining technique is to perform a data dimensionality reduction by identifying latent clusters of individuals in the multidimensional space of indicator variables. LCA, also known as finite mixture modeling, is a person-centered (as opposed to variable-centered) analytical framework that models the classification of individuals into unobserved subpopulations (latent classes) via the joint modeling of observed distributions as representing mixtures of distributions of two or more subpopulations (Vermunt and Madigson, 2002). We used LCA in the current study to explore CNV and ROH burden values to quantify the broad genomic background, aiming to identify homogenous subgroups of children with lower and higher levels of CNV and ROH burdens.

Since the LCA results suggested the presence of several latent clusters of individuals that differed in size, we used complementary methods to evaluate the contribution of f_{ROH} and CNV burdens to cognitive ability. For the smallest classes, nonparametric linear regression with permutation testing was performed with the *lmPerm()* package. ROH and CNV burden analyses in larger classes of individuals were conducted using quantile regression as implemented in the *quantreg* library for R. We chose to utilize

this semi-parametric method because we hypothesized that the contribution of f_{ROH} might be uneven at different levels of cognitive ability. Therefore, the 25th, 50th, and 75th quantiles were chosen for the analyses aimed at examining the contribution of homozygosity and structural variant burden at conditional low, moderate, and high levels of cognitive ability. Note that, in the current sample, these quantiles corresponded to average-high, high, and very high levels of cognitive ability (relative to the larger normative sample of children). Standard errors and P-values were obtained using bootstrapping, with 10,000 bootstrap samples. R-based analyses were performed using the Microsoft OpenMP implementation of R (<http://mran.microsoft.com/>) in the R Studio Server IDE environment, installed on a Linux bioinformatics high-performance computing workstation. Additional information regarding specific procedures used for CNV association analyses, analyses of relatedness using patterns of identity by descent-sharing, and ancestral composition of the sample is provided in the **Supplementary Material**.

RESULTS

Single Nucleotide Polymorphism-Based Genome-Wide Association Study of Aurora-G Scores

SNP-based association analyses conducted using linear mixed modeling in GEMMA did not reveal any markers that would survive corrections for multiple testing (**Figure 2**). Using a suggestive association threshold of $P = 1 \times 10^{-5}$, we established a set of tentative association signals for 11 SNPs (summarized in **Table 1**). Several of these findings are noteworthy and will be briefly discussed here. First, a rare missense SNP, rs79027391, located in exon 1 of the human immunodeficiency virus type I enhancer binding protein 1 (*HIVEP1*; 6p24.1) gene, was strongly negatively associated with figural/spatial ability (Aurora-F). *HIVEP1* is a transcription factor that belongs to the ZAS family of proteins that bind to enhancer elements of viral as well as cellular genes. Mutations in a closely functionally related gene, *HIVEP2*, were recently shown to be associated with intellectual disability and developmental delay (Steinfeld et al., 2016).

Second, one SNP, rs9765600, was significantly negatively associated with numerical ability (Aurora-N) as well as verbal ability (Aurora-V). This SNP is located in an intron of proline-rich 16 also known as Largen (*PRR16*; located on chromosome 5q23.1), a cell size regulator that promotes cell size increase through the upregulation of messenger RNAs, in particular those involved in mitochondrial functions (Yamamoto and Mak, 2017).

Third, Aurora-N was also nominally associated with intronic rs2806225 located in the *DSP* (desmoplakin, 6p24.3) gene and rs1863004 located in the CCR4-NOT transcription complex subunit 4 (*CNOT4*; 7q33) gene. Desmoplakin is a critical component of desmosome structure and helps maintain the structural integrity of cell–cell junctions. Mutations in *DSP* are associated with abnormal cell–cell junctions and cause cardiomyopathies and keratodermas. In the brain, *DSP* is highly and specifically expressed in the dental gyrus, suggesting that it plays a role in hippocampal neurogenesis (Wang et al., 2015), possibly through interactions with cadherin family proteins.

CCR4-NOT complex is a highly conserved transcriptional regulator that is essential for neural development (Chen et al., 2015) and differentiation of neural stem cells and has been featured in studies of the 7q33 deletion syndrome (Lopes et al., 2018) that is characterized by intellectual disability.

Gene-Based Genome-Wide Association Study of Aurora-G Scores

Although our SNP-based analyses did not reveal genome-wide significant associations that would survive corrections for multiple testing, gene-based association analyses performed in KGG revealed a more pronounced pattern of results. Six genes survived corrections for multiple testing (Table 2; regional association plots are presented in Figure 3). Three of these extend and replicate

SNP-based findings: Aurora-N was associated with *PRR16* ($P = 1.00 \times 10^{-6}$), *CNOT4* ($p = 1.39 \times 10^{-6}$), and *DSP* ($p = 1.74 \times 10^{-8}$).

We also established a genome-wide significant association between general cognitive ability (Aurora-G) and two additional genes—*C3orf20* (chromosome 3 open reading frame 20 (3p25.1) at $P = 2.38 \times 10^{-7}$ and *NR2C2AP* (nuclear receptor 2C2 associated protein; 19p13.11) at $P = 4.03 \times 10^{-6}$. Although the function of the encoded protein is unknown, *C3orf20* maps onto the locus that has been associated with age of onset of sporadic Alzheimer's disease (Velez et al., 2016). Notably, this locus, as well as *C3orf20* itself, was recently implicated in recessive intellectual disability (Karimnejad et al., 2015) in three siblings of Iranian ancestry.

The second significant signal for Aurora-G comes from the locus containing the gene encoding NR2C2-associated protein (*NR2C2AP*), which is hypothesized to repress NR2C2-mediated

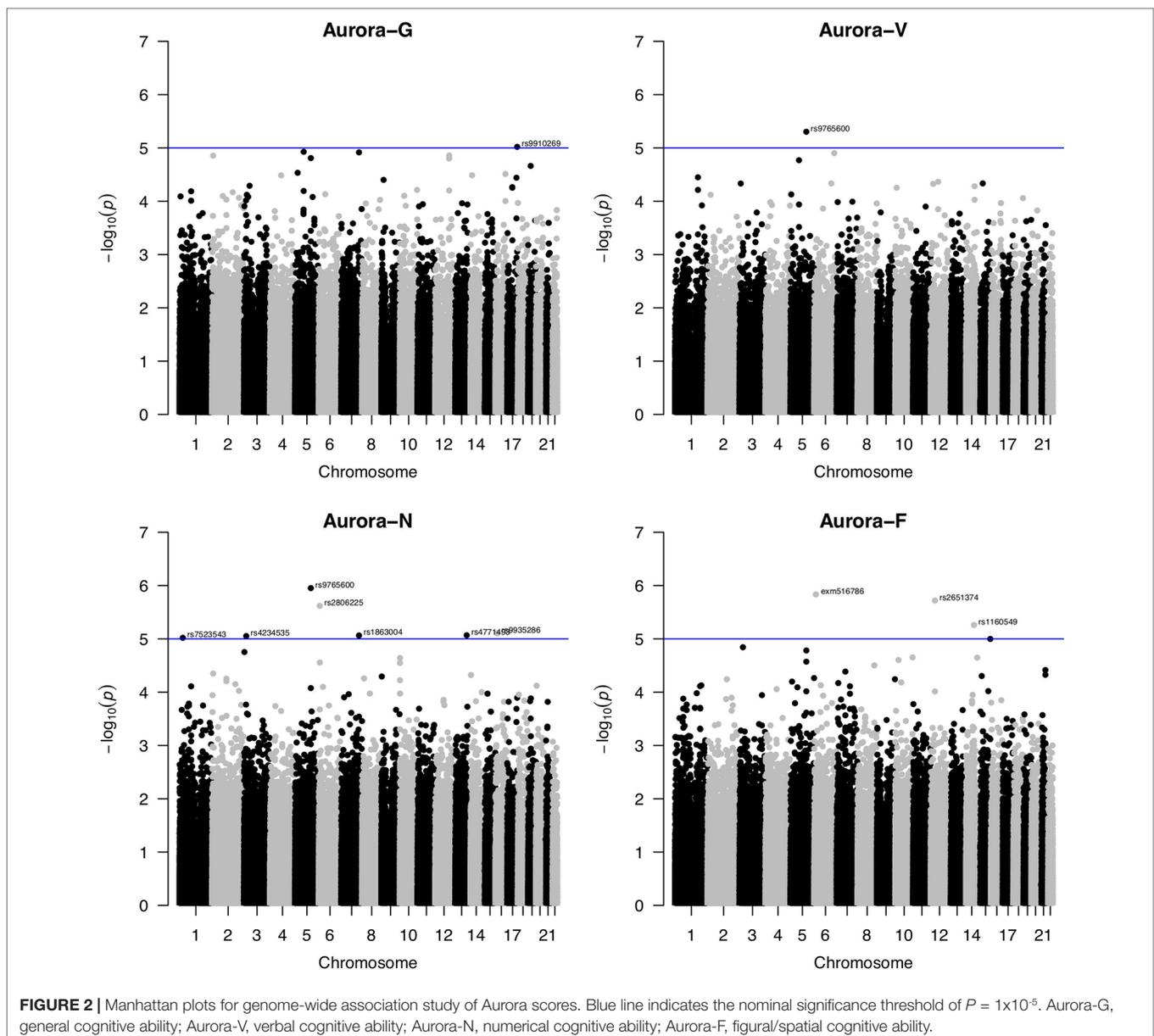


TABLE 1 | Suggestive SNP association signals for Aurora.

SNP	Chr	Position (bp)	Min/Maj	MAF	B	SE	P	Phenotype	Gene
rs7523543	1	25,186,716	G/A	0.37	-5.19	1.15	9.52×10^{-6}	Aurora-N	<i>LOC105376876</i> (intron)
rs4234535	3	16,162,533	A/G	0.28	-5.24	1.16	8.84×10^{-6}	Aurora-N	<i>GALNT15</i> (upstream of)
rs9765600	5	119,919,047	A/C	0.40	-5.18	1.04	1.12×10^{-6}	Aurora-N	<i>PRR16</i> (intron)
rs79027391	6	12,122,102	G/T	0.02	-20.01	4.08	1.46×10^{-6}	Aurora-F	<i>HIVEP1</i> (missense)
rs2806225	6	7,579,412	T/C	0.24	6.16	1.28	2.41×10^{-6}	Aurora-N	<i>DSP</i> (intron)
rs1863004	7	135,053,638	G/A	0.24	-5.73	1.27	8.60×10^{-6}	Aurora-N	<i>CNOT4</i> (intron)
rs2651374	12	32,582,858	C/T	0.46	-5.10	1.05	1.91×10^{-6}	Aurora-F	<i>FGD4</i> (intron)
rs4771493	13	105,927,432	C/T	0.32	5.24	1.16	8.57×10^{-6}	Aurora-N	<i>LOC105370344</i> (upstream of)
rs1160549	14	80,464,509	C/T	0.13	6.88	1.49	5.49×10^{-6}	Aurora-F	<i>NRXN3</i> (downstream from)
rs9935286	16	12,594,359	C/T	0.22	5.97	1.32	7.86×10^{-6}	Aurora-N	<i>SNX29</i> (intron)
rs9910269	17	75,049,645	T/C	0.24	-5.18	1.15	9.53×10^{-6}	Aurora-G	<i>SEC14L1</i> (upstream of)

Min, minor allele; Maj, major allele; MAF, minor allele frequency; B, effect size estimate; SE, standard error; P, Wald p value (uncorrected); Phenotypes: Aurora-G, general cognitive ability; Aurora-V, verbal cognitive ability; Aurora-N, numerical cognitive ability; Aurora-F, figural cognitive ability.

TABLE 2 | Genome-wide significant gene-based associations for Aurora.

Gene	k SNPs	Gene-based P	Adjusted P	Chr	Size (bp)	Phen
<i>DSP</i>	18	1.74×10^{-8}	0.00119789	6p24.3	45,143	Aurora-N
<i>C3orf20</i>	21	2.38×10^{-7}	0.00819244	3p25.1	98,812	Aurora-G
<i>PRR16</i>	24	1.00×10^{-6}	0.02294800	5q23.1	223,055	Aurora-N
<i>CNOT4</i>	8	1.39×10^{-6}	0.02392329	7q33	148,333	Aurora-N
<i>MUC17</i>	22	1.98×10^{-6}	0.02726222	7q22.1	38,788	Aurora-F
<i>NR2C2AP</i>	2	4.03×10^{-6}	0.04624022	19p13.11	2,027	Aurora-G

k SNPs, number of SNPs assigned to the gene for the association analysis; Adjusted p, p-value adjusted using the Benjamini-Hochberg correction for multiple comparisons (applied for all phenotypes and genes). Phen, phenotype.

transactivation through the suppression of binding between NR2C2 and TR4-response elements in target genes. The signal for NR2C2AP is driven by two SNPs assigned to this gene—rs10424365 ($B = -4.25$, $SE = 1.44$, $P = 0.003319076$; $MAF = 13.4\%$) and rs34282046 ($B = -16.19$, $SE = 3.76$, $P = 2.17 \times 10^{-5}$; $MAF = 1.8\%$). Both SNPs were also assigned to regulatory factor X associated ankyrin containing protein (*RFXANK*), a transmembrane protein from the MHC class II family that is involved in MHCII expression, is regulated by class IIa histone deacetylases (Wang et al., 2005), and was highlighted as a transcriptional regulator of neural stem/progenitor cells in mice (Kawase et al., 2014); rs34282046 is a nonsynonymous mutation that affects 9 alternative *RFXANK* transcripts, yet is predicted to be benign by PolyPhen2 (Score = 0.000).

Finally, Aurora-F was associated with mucin 17 (*MUC17*, cell surface associated; 7q22.1) at $P = 1.98 \times 10^{-6}$. The protein encoded by this gene is mostly found in epithelial cells of the intestinal tract. *MUC17* is one of the over 100 genes recently reported to be included in the interstitial deletion of the 7q22.1 syndrome, which is characterized by structural brain abnormalities and intellectual disability (Katz et al., 2016). We did not establish any genome-wide significant associations for Aurora-V.

Genome-Wide Surrogate Runs of Homozygosity Association Analysis

ROHs identified by GARLIC were subjected to two complementary analyses: homozygosity burden analyses using f_{ROH} estimates and

ROH association analyses using a binary matrix of surrogate ROHs (sROHs) that accounted for the partial overlaps of ROHs in the sample. The analyses were performed using custom R scripts. The identification of surrogate ROHs followed the algorithm parallel to that employed by the CNVRuler (Kim et al., 2012) software for the detection of overlapping CNV segments; the two approaches produced an identical set of ROH regions. Given the methodological considerations of superficial signal similarity between ROHs and hemizygous deletions (Mcquillan et al., 2008), we conservatively restricted the set of ROHs in the association analyses to those that did not overlap with copy number variable regions detected by PennCNV for the entire study sample (see Figure 4).

$K = 30,733$ unique sROHs outside of CNV regions were called, and 19,214 had a frequency above 1% (at least 4 out of 354 individuals in the sample). sROH frequency was inversely correlated with its length at $\rho = -0.15$ ($p < 0.001$). To reduce the number of multiple comparisons and increase power, further analyses focused specifically on $k = 11,992$ genic sROHs that we hypothesized were more likely to have an interpretable pronounced phenotypic effect than intergenic sROHs.

Four loci exhibited a total of 15 sROH segments with nominal association P values < 0.0001 (see Table 3). A region on chromosome 11q22.3 survived corrections for multiple testing (for Aurora-F, $P_{adj} = 0.02093395$ and 0.03569284 for the two sROHs) and was simultaneously negatively associated with Aurora-G and positively—with Aurora-F and Aurora-N. Effect

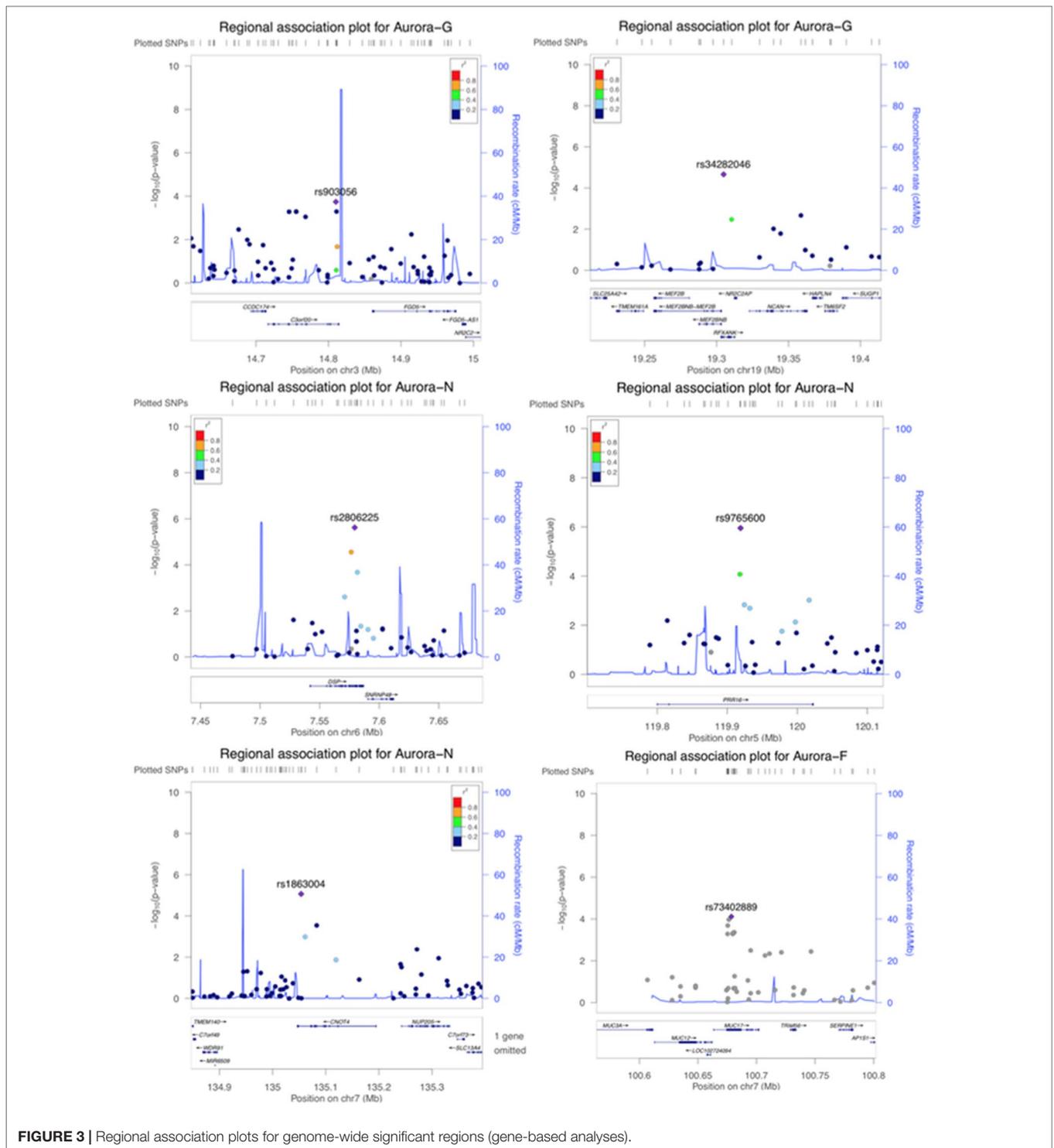
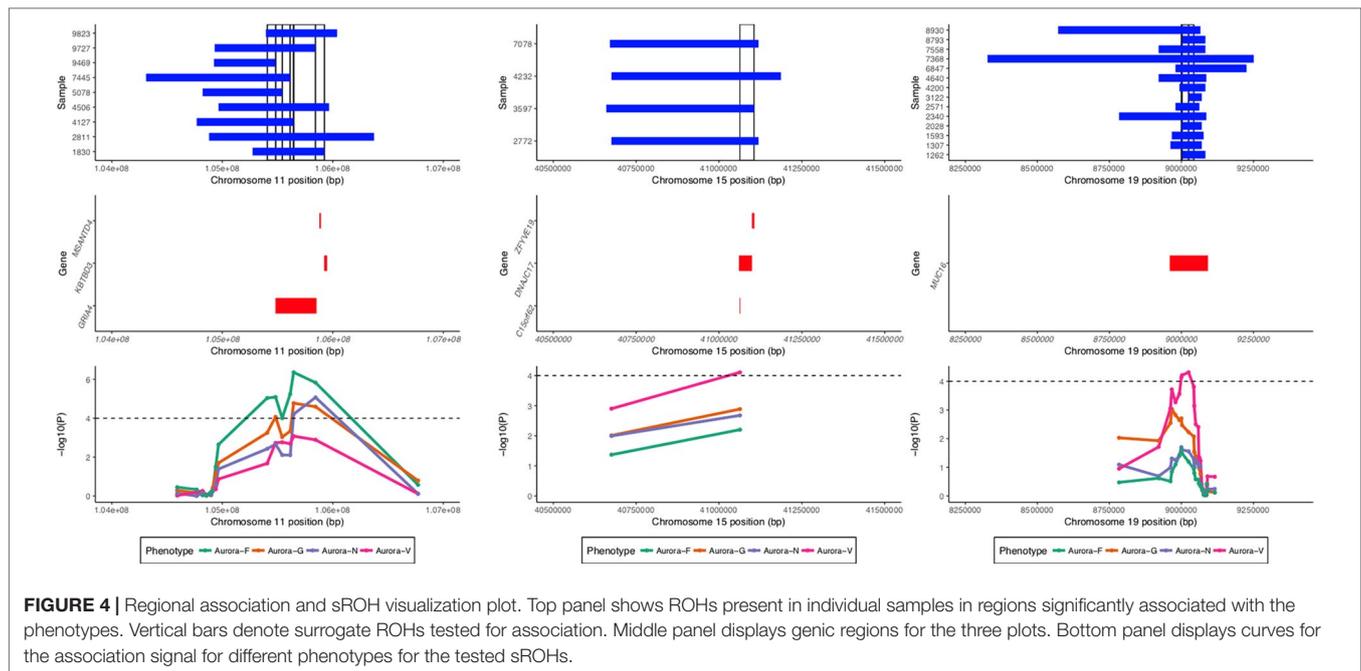


FIGURE 3 | Regional association plots for genome-wide significant regions (gene-based analyses).

sizes exceeded 20 standard IQ points, and allele frequencies ranged from 1.12 to 2.54%, with the cumulative length of the affected region estimated at around 796 kb. This region contains several important genes, most notably glutamate ionotropic receptor alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid type subunit 4 (*GRIA4*), an L-glutamate receptor gene that

is highly expressed in the central nervous system and is involved in excitatory synaptic transmission. We further interrogated the sample for the presence of possible coding mutations in *GRIA4* by performing genotype imputation using BEAGLE (Browning and Browning, 2009) and a custom reference panel constructed using whole-genome DNA sequencing data from n = 108 individuals



of Qatari ancestry ($k = 9,763,073$ markers) (Rodriguez-Flores et al., 2016). Using this dataset, we were able to impute five coding SNPs in the study sample (rs57375419, rs7933212, rs56091908, rs116000289, and rs61736488). All of the ROH carriers in the study sample were homozygous for the reference alleles of the imputed SNPs.

Finally, Aurora-V was positively associated with sROH located on chromosome 15q15.1 and encompassing DnaJ heatshock protein family (Hsp40) member C17 (*DNAJC17*), also involved in autoimmune disorder lecolicitis, as well as with chromosome 15 open reading frame 62 (*C15orf62*), and zinc finger FYVE-type containing 19 (*ZFYVE19*).

Contribution of Genome-Wide Homozygosity and CNV Burden to Cognitive Ability

In order to capture inter-individual variation in patterns of homozygosity and structural variant burdens, we conducted a LCA using $f_{ROH(A)}$, $f_{ROH(B)}$, $f_{ROH(C)}$, f_{SNP} , and CNV burden values as indicator variables. Analysis of BIC values for the estimated models for k latent mixture components ranging from 1 to 10 suggested that the best fit is provided by the five-component model that postulates the presence of five latent classes of individuals (additional sensitivity analyses using reduced variable groupings and alternative formulations of models converged on a five-mixture model; data not shown). The respective model displayed EVE geometric notation (ellipsoidal distribution, equal volume, and variable shape among latent classes); for details, see Browne and Nicholas (Browne and McNicholas, 2014) and **Supplementary Material**. Model fit was also evaluated using sequential bootstrap likelihood ratio test with 1,000 bootstrap samples: the analyses suggested that

model fit improved with increases in number of estimated mixture components from 1 to 2 ($p = 0.000999001$), from 2 to 3 ($p = 0.00099001$), from 3 to 4 ($p = 0.001998002$), and from 4 to 5 ($p = 0.000999001$) but not from 5 to 6 ($p = 0.946053946$), suggesting that the $k = 5$ mixture components model indeed provides best fit to the data.

The classes ranged in size from 11 to 181 individuals and could heuristically (based on the results of permutation-based analysis of variance for indicator variables; data not shown) be labeled according to **Table 4**, which provides relative qualifiers for estimated SNP and ROH homozygosity and CNV burden. Overall, 51% of the sample was classified into Class 3, characterized by low levels of homozygosity and CNV burden. Classes 1 and 2 had moderate levels of homozygosity but were characterized by increased CNV burden. Classes 4 and 5, on the other hand, represented individuals with low CNV burden but dramatically elevated f_{ROH} homozygosity driven by class B and C ROHs, respectively.

Notably, although the five latent classes did not differ with respect to average estimated cognitive ability (all P 's > 0.6667 ; see **Supplementary Material**), they differed in patterns of association between cognitive ability and f_{ROH} and CNV burdens (**Figure 5**). For the purpose of the analysis, we combined Classes 1 and 2 into one group. Since the primary focus of our analyses was on general cognitive ability, Aurora-G was used as the main phenotype for the association analyses. All analyses controlled for the first five principal components estimated from genotyping data.

Given the small sample size for Classes 1 + 2, we used simple linear regression with permutation testing to evaluate the contribution of f_{ROH} and CNV burden to cognitive ability in this group. Analyses indicated that although the model with just PC covariates was not statistically significant, $F_{(5,30)} = 1.791$, $P =$

TABLE 3 | Nominally significant (at $P < .0001$) associations between sROHs and Aurora.

sROH	B	SE	T	P	Adjusted P	Chr	Start	End	Length	Carriers	Gene	Phen
sROH_17583	22.20	4.93	4.50	9.19×10^{-6}	0.073441885	11	105,405,895	105,482,024	76,130	9/354	GRI44	Aurora-F
sROH_17584	-20.89	5.26	-3.98	8.53×10^{-5}	0.292376267	11	105,482,025	105,541,743	59,719	8/354	GRI44	Aurora-G
sROH_17584	23.64	5.22	4.53	8.20×10^{-6}	0.073441885	11	105,482,025	105,541,743	59,719	8/354	GRI44	Aurora-F
sROH_17585	22.10	5.61	3.94	9.89×10^{-6}	0.316240857	11	105,541,744	105,614,894	73,151	7/354	GRI44	Aurora-F
sROH_17586	27.68	6.01	4.61	5.69×10^{-6}	0.073441885	11	105,614,895	105,645,648	30,754	6/354	GRI44	Aurora-F
sROH_17587	-28.75	6.59	-4.36	1.69×10^{-5}	0.115640555	11	105,645,649	105,844,065	198,417	5/354	GRI44	Aurora-G
sROH_17587	26.82	6.61	4.05	6.18×10^{-5}	0.269457038	11	105,645,649	105,844,065	198,417	5/354	GRI44	Aurora-N
sROH_17587	33.59	6.52	5.15	4.36×10^{-7}	0.020933945	11	105,645,649	105,844,065	198,417	5/354	GRI44	Aurora-F
sROH_17588	33.18	7.34	4.52	8.52×10^{-6}	0.073441885	11	105,844,066	105,923,564	79,499	4/354	KBTD3, GRI44, MSANTD4	Aurora-N
sROH_17588	-31.42	7.37	-4.27	2.57×10^{-5}	0.153851824	11	105,844,066	105,923,564	79,499	4/354	KBTD3, GRI44, MSANTD4	Aurora-G
sROH_17588	35.79	7.31	4.90	1.49×10^{-6}	0.035692842	11	105,844,066	105,923,564	79,499	4/354	KBTD3, GRI44, MSANTD4	Aurora-F
sROH_20937	29.54	7.39	4.00	7.77×10^{-5}	0.286862153	15	41,063,245	41,105,658	42,414	4/354	DNAJC17, C15orf62, ZFYVE19	Aurora-V
sROH_23251	-18.06	4.50	-4.01	7.30×10^{-5}	0.286862153	19	8,999,675	9,001,832	2,158	11/354	MUC16	Aurora-V
sROH_23252	-16.82	4.15	-4.05	6.18×10^{-5}	0.269457038	19	9,001,833	9,024,870	23,038	13/354	MUC16	Aurora-V
sROH_23253	-16.46	4.00	-4.11	4.85×10^{-5}	0.258308827	19	9,024,871	9,043,020	18,150	14/354	MUC16	Aurora-V

sROH, unique surrogate ROH identifier; B, effect size (in standardized scores units); SE, standard error; t, t-value for regression coefficient. Adjusted P—p-value adjusted using the Benjamini–Hochberg correction. Adjusted P-values < 0.05 are shown in **bold**. Phen, phenotype.

0.145, $R^2 = 0.1015$, adding fROH and CNV burden variables drastically improved its fit, $F_{(9,26)} = 3.395$, $P = 0.006991$, $R^2 = 0.3812$, with fROH and CNV burden explaining an additional 28% of variance in general cognitive ability in this sample, with statistically significant negative contributions from $f_{ROH(C)}$, $B = -2.21$, $P = 0.044$, and CNV burden, $B = -49.57$, $P = .001$. Thus, in this subgroup of individuals with moderate ROH and substantial CNV burdens, both variables accounted for over a quarter of the variation in cognitive ability. However, sensitivity analyses removing one outlying data point rendered the CNV burden effect nonsignificant while preserving the $f_{ROH(C)}$ effect.

Notably, in other groups, we did not establish significant contributions of ROH and CNV burdens to cognitive ability, using linear regression fit for the entire distribution of ability values (data not shown). Given the larger sample size of the combined Classes 4 and 5 and Class 3, we further investigated the relationship between ROH and CNV burdens and cognitive ability at different segments of its distribution, thereby accounting for possible nonlinear effects. Thus, quantile regressions were fit separately for Class 3 ($n = 181$) and Classes 4 and 5 ($n = 137$). Parameter estimates can be found in the **Supplementary Material**.

This analysis revealed that for Class 3, the only significant effect was observed for $f_{ROH(B)}$ at the 50th quantile ($B = 6.22$, $t = -2.51$, $P = 0.01314$), with higher homozygosity due to medium-length ROHs being associated with lower cognitive ability. At the same time, for Classes 4 and 5, it was the $f_{ROH(C)}$ burden that acted as a significant negative predictor of Aurora-G at the 25% quantile ($B = -1.28$, $SE = 0.60$, $t = -2.11$, $P = .03664$) but not 50% quantile ($B = -.14$, $SE = 0.42$, $t = -0.32$, $P = .15322$) and 75% quantiles ($B = -.14$, $SE = 0.42$, $t = -0.32$, $P = .15322$). In Classes 4 and 5, CNV burden was also a strong negative predictor of cognitive ability at the 75-th quantile ($B = -214.64$, $SE = 74.96$, $t = -2.86$, $P = 0.00491$).

Thus, our regression analyses performed separately for three groups of individuals suggested that in the presence of substantial f_{ROH} and CNV burdens, these variables are strongly predictive of general cognitive ability, with decreases in Aurora-G scores estimated at 2.21 standardized IQ points for each 1% increase in $f_{ROH(C)}$ and almost 50 points for each 1% increase in CNV burden. In the class characterized by low levels of CNV and ROH burdens, homozygosity burden due to medium-length ROHs predicted ability negatively in the average ability range. In the classes of individuals characterized by high levels of CNV and ROH burdens, homozygosity effects were due to consanguinity-related long ROHs, and effects were pronounced in the lower and higher but not the median portion of the cognitive ability distribution.

DISCUSSION

In this study, we examined the genetic underpinnings of cognitive ability in a sample of children from the consanguineous population of Saudi Arabia. First and foremost, we demonstrated that the homozygosity burden in this population is significant, with approximately half of the sample estimated to have f_{ROH}

TABLE 4 | Heuristic description of five latent classes estimated using ROH and CNV burden values.

Latent Class	N	f_{SNP}	f_{ROH}		f_{ROH}		CNV burden		M	SD
			M	SD	M	SD	M	SD		
1	11	Moderate	0.040	0.030	Moderate	3.26%	1.22%	High	0.218%	0.182%
2	25	Moderate	0.032	0.039	Moderate	4.97%	2.56%	Moderate	0.096%	0.035%
3	181	Low	0.008	0.032	Low	1.79%	0.79%	Low	0.018%	0.016%
4	69	High	0.064	0.056	High	7.48%	4.18%	Low	0.008%	0.008%
5	68	High	0.086	0.032	High	6.98%	2.55%	Low	0.022%	0.016%

burden values above 3%. Even individuals in Class 3 characterized by relatively low levels of f_{ROH} still had an estimated f_{ROH} at 1.79%; this is over 10 times larger than the lowest average level of f_{ROH} across European populations presented in a recent study (Howrigan et al., 2016) that evaluated contributions of genome-wide homozygosity to cognitive ability. At the same time, 49% of the individuals in our sample were identified as demonstrating moderate to high levels of inbreeding, with the average f_{ROH} estimated in Class 4 at 7.48% of the autosomal genome, i.e., over 12 times larger than the highest f_{ROH} value for populations studied previously.

We established that individuals with low and high levels of CNV and ROH burden did not differ with respect to their average cognitive ability. The absence of group differences in cognitive ability between children with low and high homozygosity burden estimates in the sample is intriguing. It is possible that the potential performance gap between these classes of children is masked by the advantageous effects of the high socioeconomic status of children born into the

arranged, consanguineous marriages of the Saudi traditional families, which are well supported both financially and via social services (Warsy et al., 2014). Nonetheless, when limited to the analysis within this subpopulation, f_{ROH} and CNV burdens negatively predicted cognitive ability at its lowest and highest distributional tails in the subsample with pronounced CNV and ROH burden, with the longest C-class ROHs driving the effect. This class of ROHs is hypothesized to arise through parental relatedness and has been noted to contain a disproportionate number of detrimental homozygous variants (Szpiech et al., 2013), directly linking consanguinity in the studied population with variation in children's cognitive ability.

Homozygosity mapping in the current study also identified a subgroup of nine individuals carrying a ROH that encompasses the *GRIA4* gene. Out of nine carriers of ROHs in this region, the vast majority (8/9) were assigned to groups with pronounced ROH burden, and the presence of ROHs in this region was associated with lower general cognitive ability and higher levels of the figural ability, effectively indexing disparate cognitive profile with lower estimated general ability (ranging from 78 to 122 for carriers, $M = 93.89$; $SD = 18.93$) but higher spatial ability (ranging from 88 to 145 for carriers, $M = 121.80$; $SD = 19.41$). Given the higher reliability of the general cognitive ability estimate due to its reliance on nine indicator variables (as opposed to three residual indicators for domain factors), we believe this effect should be interpreted as capturing variance related primarily to general cognitive ability.

GRIA4 encodes a subunit of the alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors that underlie excitatory synaptic transmission and activity-dependent synaptic plasticity (Bettler and Fakler, 2017) and has been implicated in a range of human phenotypes, including schizophrenia (Marshall et al., 2017), substance use, and major depression. In a recent exome sequencing study of individuals with intellectual disability, *de novo* heterozygous pathogenic variants in *GRIA4* were found in five unrelated individuals with the condition (Martin et al., 2017). Although the functional impact of the identified ROHs on *GRIA4* function in the Saudi population is unknown, these genome-wide significant findings nonetheless suggest that this gene plays a role in the regulation of cognitive traits in the population. Notably, *GRIA4* was highlighted as a significant predictor of cognitive ability in the largest meta-GWAS of cognitive ability to date (Davies et al., 2018).

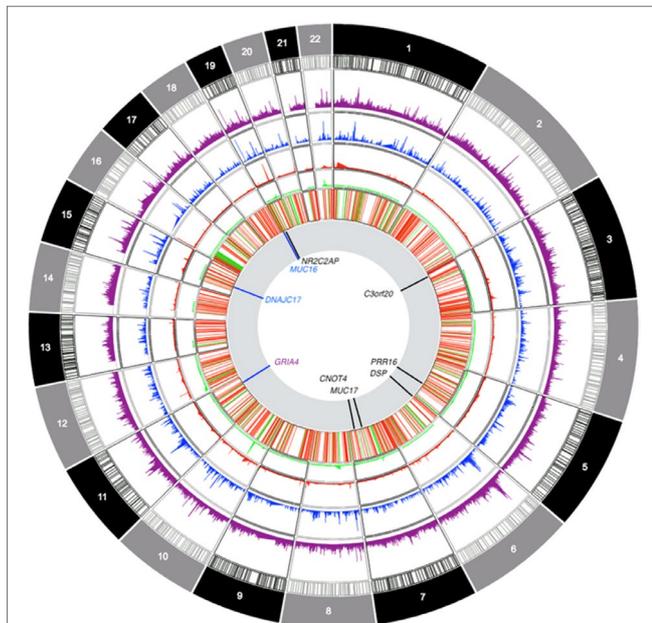


FIGURE 5 | Contribution of f_{ROH} and CNV burden variables to general cognitive ability (Aurora-G) in four latent classes of individuals. Top panel presents data for Class1 + Class2, bottom panel—for Class4 + Class5.

Of note is that, homozygosity mapping also identified a locus on chromosome 19 harboring the *MUC16* gene as being nominally associated with verbal ability in this sample. Although this locus did not survive corrections for multiple testing, this finding is noteworthy for two reasons. First, it corresponded to the ROH hotspot on chromosome 19 clearly visible in **Figure 4**: 14 out of 354 (or 4%) individuals had a ROH in this region. Although increased regional homozygosity has been previously associated with positive selection, in our sample, the average verbal general cognitive ability among carriers of ROH was estimated at 86.97 IQ points, slightly

above the -1 SD cutoff, with 8 out of 14 individuals displaying scores at or below 85 standard points. Second, another related gene, *MUC17*, was genome-wide significantly associated with figural/spatial ability in this sample. The lead SNP from this gene was rs73402889, a missense S>I mutation in exon 3 of *MUC17* that is predicted to be tolerated by SIFT. This SNP also is a significant expression QTL for *CLDN15* that encodes Claudin-15, which is an integral membrane protein that is involved in the structure and maintenance of tight junctions and cell polarity. Both *MUC16* and *MUC17* proteins belong to the class of transmembrane mucins containing SEA domains.

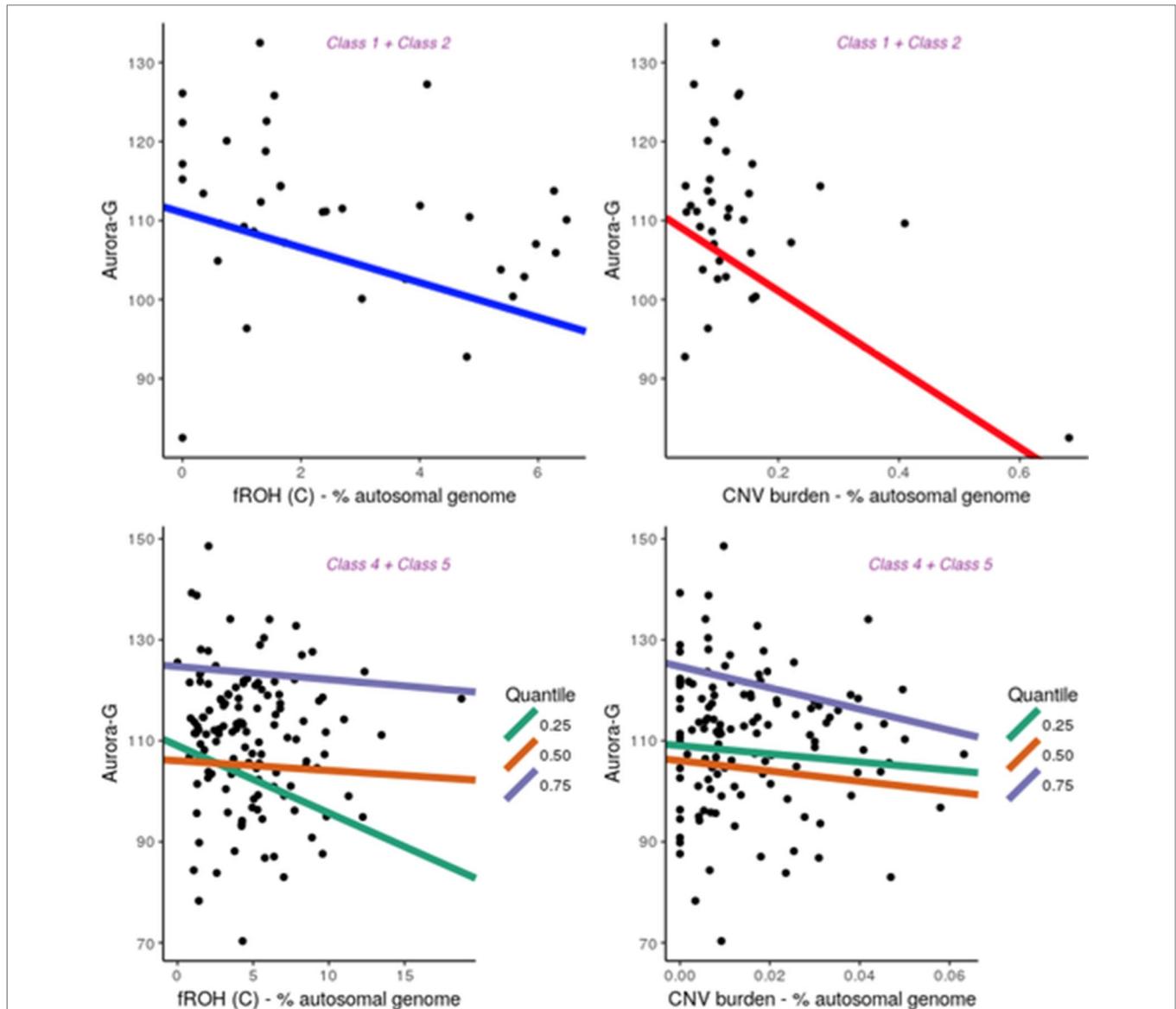


FIGURE 6 | Circos-style plot of genome-wide association study data integrating gene-based and ROH-based findings. Starting from the outside, tracks denote: frequency of sROH segments prior to CNV filtering, frequency of sROHs outside CNV regions, frequency of CNV loss events, frequency of CNV gain events, joint distribution of CNV segments (regions excluded from ROH analyses), association results. The last track highlights genomic locations with findings that survived multiple corrections for gene-based association tests (in black) and findings at $p < 0.0001$ for sROH association analyses (blue) as well as the finding from the *GRIA4* (purple) region that survived corrections for multiple testing.

Transmembrane mucins are highly expressed in epithelial cells. In the developing brain, MUC16 and MUC17 are both expressed at 10 weeks post-conception in the choroid plexus (Expression Atlas; <https://www.ebi.ac.uk/gxa/>), a network of blood vessels and cells involved in the production of cerebrospinal fluid. It is possible that transmembrane mucins contribute to cognitive ability in this population through their involvement in mucin-type O-glycosylation, a form of posttranslational protein modification that promotes vascular integrity (Herzog et al., 2014) and are involved in a range of developmental processes. Studies suggest that O-glycosylation is involved in the stabilization of early brain vasculature and that subventricular zones containing neural progenitor cells are particularly susceptible to O-glycosylation deficiencies (Xia et al., 2004).

The genes highlighted in this study (see **Figure 6** for a summary) did not harbor pathogenic mutations in the clinical exome study of a primarily pediatric cohort of 454 phenotypically heterogeneous probands (Alfares et al., 2017) or an even larger study of over 1,000 families also with heterogeneous and suspected Mendelian phenotypes (Monies et al., 2017) from Saudi Arabia, as well as a smaller cohort of 149 children and young adult probands from Qatar (Yavarna et al., 2015). This is not contrary to our findings, as we selected our sample for higher cognitive ability scores in the population sample (i.e., the opposite of the selection criteria for the abovementioned studies). Yet, at this point, it is important to attempt a replication of these findings in an unselected sample to verify the generalizability of the obtained results. Interestingly, our findings do overlap with genes identified in recent meta-analyses of cognitive ability: in addition to the *GRIA4* finding, we established genome-wide significant associations for *CNOT4* and *PRR16*, both of which are highly expressed in the brain and have been previously associated with cognitive ability (Davies et al., 2018).

Saudi Arabian researchers are actively examining the effects of consanguinity on the prevalence and etiology of Mendelian traits, and recent studies have been extremely fruitful in identifying coding variants responsible for fairly severe phenotypes, including intellectual disability. The results of our study are in concert with the growing realization of the importance of ROH in a variety of traits and conditions (Lencz et al., 2007; Yang et al., 2010; Keller et al., 2012; Gamsiz et al., 2013; Power et al., 2014a; Power et al., 2014b; Verweij et al., 2014; Gandin et al., 2015; Ghani et al., 2015; Mezzavilla et al., 2015; Howrigan et al., 2016) and highlight ROHs as population-specific alleles that form the genetic foundation for individual differences in cognitive ability. This complex trait regulation from genetic variants that can harbor multiple deleterious variants in the homozygous forms suggests that ROHs represent an important class of genetic variants and that theoretical developments in human trait genetics should attempt to incorporate consanguinity and ROHs [for example, from the infinitesimal model (Barton et al., 2017)] into their models. It has been noted that age-related Flynn effects (i.e., substantial and sustained population-level increases in cognitive ability over time) are smaller in Arab

countries (Bakhiet et al., 2018); it is possible that in addition to cultural environmental factors (such as educational standards) highlighted by others, such as disproportionately small cognitive ability gains over time can be attributed to population-specific genomic background characterized by elevated homozygosity as a substantial risk factor.

Admittedly, the key limitation of the present study lies in its small sample size (in fact, it is 1,000 times smaller than the most recent publication on the genetics of cognitive ability) and the absence of a replication sample. We would like to note, however, that this study yielded a pronounced pattern of interpretable results, both in terms of quantifying the extent of homozygosity burden in children in Saudi Arabia and highlighting its substantial contribution to the genetic foundation of cognitive ability in this consanguineous population. Pathway enrichment analyses (**Supplementary Material**) indicated two overrepresented pathways (cadherin and Wnt-signaling) and also highlighted calcium ion binding and nervous system development as overrepresented molecular function and biological process terms. Homozygosity mapping implicated *GRIA4* and *MUC16* as possible regulators of cognitive ability in this sample, with the majority of individuals harboring the ROHs encompassing these genes being assigned to the high homozygosity burden latent classes, thereby linking consanguinity, homozygosity, and cognitive ability. To the best of our knowledge, this is the first investigation of the genetic foundation of cognitive ability of a highly consanguineous population of Saudi Arabia, and replication datasets are lacking at the moment.

Future studies could utilize recently published clinical microarray and sequencing data produced by the multicenter clinical team in KSA, which contains a sizeable number of individuals with neurodevelopmental disorders (Alfares et al., 2017; Monies et al., 2017). Given the demonstrated power of the approach taken by this study, future studies should also attempt further characterizing the contribution of homozygosity burden to cognitive ability in highlighted consanguineous populations, such as populations from Central and South Asia (Pemberton and Rosenberg, 2014), as homozygosity and CNV burdens were associated with variation in cognitive ability, specifically in groups where the genetic load of these factors was substantial.

DATA AVAILABILITY

The data generated for this study can be found in the dbGAP NCBI repository using accession number phs001884.v1.p1 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001884.v1.p1).

ETHICS STATEMENT

The study protocol and procedures were approved by the Internal Review Boards of Yale University and University of Houston. Parents of children participating in the study and

authorities at the appropriate educational institutes provided written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

EG, MT, and AA conceptualized and designed the study and the study assessments and supervised the data collection. ON and SK processed the biological specimens. SK, ON, and EG planned and carried out the analyses. SK, MT, AA, ON, and EG drafted, revised, and approved the final manuscript.

FUNDING

This research was supported by the Ministry for Higher Education, Kingdom of Saudi Arabia. Grantees undertaking such projects

REFERENCES

- Alfares, A., Alfadhel, M., Wani, T., Alsahli, S., Alluhaydan, I., Al Mutairi, F., et al. (2017). A multicenter clinical exome study in unselected cohorts from a consanguineous population of Saudi Arabia demonstrated a high diagnostic yield. *Mol. Genet. Metab.* 121, 91–95. doi: 10.1016/j.ymgme.2017.04.002
- Bakhiet, S. F. A., Dutton, E., Ashaer, K. Y. A., Essa, Y. A. S., Blahmar, T. A. M., Hakami, S. M., et al. (2018). Understanding the Simber effect: why is age-dependent increase in children's cognitive ability smaller in Arab countries than in Britain. *Pers. Individual Diff.* 122, 38–42. doi: 10.1016/j.paid.2017.10.002
- Barton, N. H., Etheridge, A. M., and Veber, A. (2017). The infinitesimal model: definition, derivation, and implications. *Theor. Popul. Biol.* 118, 50–73. doi: 10.1016/j.tpb.2017.06.001
- Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., et al. (2016). The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol. Sci.* 27, 957–972. doi: 10.1177/0956797616643070
- Bener, A., and Mohammad, R. R. (2017). Global distribution of consanguinity and their impact on complex diseases: genetic disorders from an endogamous population. *Egypt. J. Med. Hum. Genet.* 18, 315–320. doi: 10.1016/j.ejmhg.2017.01.002
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Ser. B-Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Bettler, B., and Fakler, B. (2017). Ionotropic AMPA-type glutamate and metabotropic GABA_B receptors: determining cellular physiology by proteomes. *Curr. Opin. Neurobiol.* 45, 16–23. doi: 10.1016/j.conb.2017.02.011
- Browne, R. P., and McNicholas, P. D. (2014). Estimating common principal components in high dimensions. *Adv. Data Anal. Classification* 8, 217–226. doi: 10.1007/s11634-013-0139-1
- Browning, B. L., and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. doi: 10.1016/j.ajhg.2009.01.005
- Ceballos, F. C., Hazelhurst, S., and Ramsay, M. (2018a). Assessing runs of homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* 19, 106. doi: 10.1186/s12864-018-4489-0
- Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., and Wilson, J. F. (2018b). Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* 19, 220. doi: 10.1038/nrg.2017.109
- Chalmers, R. P. (2012). mirt: a multidimensional item response theory package for the R environment. *J. Stat. Software* 48, 1–29. doi: 10.18637/jss.v048.i06
- Chart, H., Grigorenko, E. L., and Sternberg, R. J. (2009). “Identification: the aurora battery,” in *Critical Issues and Practices in Gifted Education: What the Research Says*. Eds. J. A. Plucker and C. M. Callahan (Waco TX US: Prufrock Press), 281–301.
- Chen, L., Chu, C., Kong, X. Y., Huang, T., and Cai, Y. D. (2015). Discovery of new candidate genes related to brain development using protein interaction information. *Plos One* 10, e0118003. doi: 10.1371/journal.pone.0118003
- Cucina, J., and Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *J. Intell.* 5, 1–21. doi: 10.3390/jintelligence5030027
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* 9, 2098. doi: 10.1038/s41467-018-04362-x
- El-Hazmi, M. A., Al-Swailem, A. R., Warsy, A. S., Al-Swailem, A. M., Sulaimani, R., and Al-Meshari, A. A. (1995). Consanguinity among the Saudi Arabian population. *J. Med. Genet.* 32, 623–626. doi: 10.1136/jmg.32.8.623
- Erzurumluoglu, A. M., Shihab, H. A., Rodriguez, S., Gaunt, T. R., and Day, I. N. M. (2016). Importance of genetic studies in consanguineous populations for the characterization of novel human gene functions. *Ann. Hum. Genet.* 80, 187–196. doi: 10.1111/ahg.12150
- Gamsiz, E. D., Viscidi, E. W., Frederick, A. M., Nagpal, S., Sanders, S. J., Murtha, M. T., et al. (2013). Intellectual disability is associated with increased runs of homozygosity in simplex autism. *Am. J. Hum. Genet.* 93, 103–109. doi: 10.1016/j.ajhg.2013.06.004
- Gandin, I., Faletra, F., Faletra, F., Carella, M., Pecile, V., Ferrero, G. B., et al. (2015). Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. *Genet. Med.* 17, 396–399. doi: 10.1038/gim.2014.118
- Gao, X. Y., Becker, L. C., Becker, D. M., Starmer, J. D., and Province, M. A. (2010). Avoiding the high bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* 34, 100–105. doi: 10.1002/gepi.20430
- Ghani, M., Reitz, C., Cheng, R., Vardarajan, B. N., Jun, G., Sato, C., et al. (2015). Association of long runs of homozygosity with alzheimer disease among African American Individuals. *Jama Neurol.* 72, 1313–1323. doi: 10.1001/jamaneurol.2015.1700
- Hein, S., Tan, M., Aljughaiman, A., and Grigorenko, E. L. (2015). Gender differences and school influences with respect to three indicators of general intelligence: evidence from saudi arabia. *J. Educ. Psychol.* 107, 486–501. doi: 10.1037/a0037519
- Herzog, B. H., Fu, J., and Xia, L. (2014). Mucin-type O-glycosylation is critical for vascular integrity. *Glycobiology* 24, 1237–1241. doi: 10.1093/glycob/cwu058

ACKNOWLEDGMENTS

We thank Dr. Alaa Ayoub from Gulf University, Bahrain, for his assistance with data collection and Mrs. Maria Lee from Yale University, USA, for her assistance with the processing of the biological specimens for the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00888/full#supplementary-material>

- Hill, W. D., Davies, G., Harris, S. E., Hagenaars, S. P., Liewald, D. C., Penke, L., et al. (2016). Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl. Psychiatry* 6, e980. doi: 10.1038/tp.2016.246
- Howrigan, D. P., Simonson, M. A., Davies, G., Harris, S. E., Tenesa, A., Starr, J. M., et al. (2016). Genome-wide autozygosity is associated with lower general cognitive ability. *Mol. Psychiatry* 21, 837–843. doi: 10.1038/mp.2015.120
- Johnson, E. C., Evans, L. M., and Keller, M. C. (2018). Relationships between estimated autozygosity and complex traits in the UK Biobank. *PLoS Genet.* 14, e1007556. doi: 10.1371/journal.pgen.1007556
- Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nütle, T., et al. (2015). Directional dominance on stature and cognition in diverse human populations. *Nature* 523, 459. doi: 10.1038/nature14618
- Kariminejad, A., Nafissi, S., Nilipoor, Y., Tavassoli, A., Van Veldhoven, P. P., Bonnard, C., et al. (2015). Intellectual disability, muscle weakness and characteristic face in three siblings: a newly described recessive syndrome mapping to 3p24.3-p25.3. *Am. J. Med. Genet. Part A* 167, 2508–2515. doi: 10.1002/ajmg.a.37248
- Katz, O. L., Krantz, I. D., and Noon, S. E. (2016). Interstitial deletion of 7q22.1q31.1 in a boy with structural brain abnormality, cardiac defect, developmental delay, and dysmorphic features. *Am. J. Med. Genet. Part C-Semin. Med. Genet.* 172, 92–101. doi: 10.1002/ajmg.c.31485
- Kawase, S., Kuwako, K., Imai, T., Renault-Mihara, F., Yaguchi, K., Itoharu, S., et al. (2014). Regulatory factor X transcription factors control *musashi1* transcription in mouse neural stem/progenitor cells. *Stem Cells Dev.* 23, 2250–2261. doi: 10.1089/scd.2014.0219
- Keller, M. C., Simonson, M. A., Ripke, S., Neale, B. M., Gejman, P. V., Howrigan, D. P., et al. (2012). Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *Plos Genet.* 8, 425–435. doi: 10.1371/journal.pgen.1002656
- Kim, J. H., Hu, H. J., Yim, S. H., Bae, J. S., Kim, S. Y., and Chung, Y. J. (2012). CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics* 28, 1790–1792. doi: 10.1093/bioinformatics/bts239
- Koller, M. (2016). *robustlmm*: an R package for robust estimation of linear mixed-effects models. *J. Stat. Software* 75, 1–24. doi: 10.18637/jss.v075.i06
- Lencz, T., Lambert, C., Derosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *PNAS* 104, 19942–19947. doi: 10.1073/pnas.0710021104
- Li, M. X., Sham, P. C., Cherny, S. S., and Song, Y. Q. (2010). A knowledge-based weighting framework to boost the power of genome-wide association studies. *Plos One* 5, e14480. doi: 10.1371/journal.pone.0014480
- Lopes, F., Torres, F., Lynch, S. A., Jorge, A., Sousa, S., Silva, J., et al. (2018). The contribution of 7q33 copy number variations for intellectual disability. *Neurogenetics* 19, 27–40. doi: 10.1007/s10048-017-0533-5
- Majid, A., Ali, A. O., Saif, A. S., Fuad, A. M., Moeenaldeen, A., Zuhair, R., et al. (2017). Expanded newborn screening program in Saudi Arabia: incidence of screened disorders. *J. Paediatr. Child Health* 53, 585–591. doi: 10.1111/jpc.13469
- Marshall, C. R., Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W. T., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* 49, 27–35. doi: 10.1038/ng.3725
- Martin, S., Chamberlin, A., Shinde, D. N., Hempel, M., Strom, T. M., Schreiber, A., et al. (2017). De Novo variants in *GRIA4* lead to intellectual disability with or without seizures and gait abnormalities. *Am. J. Hum. Genet.* 101, 1013–1020. doi: 10.1016/j.ajhg.2017.11.004
- Mcquillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007
- Mezzavilla, M., Vozzi, D., Badii, R., Alkowiari, M. K., Abdulhadi, K., Giroto, G., et al. (2015). Increased rate of deleterious variants in long runs of homozygosity of an inbred population from Qatar. *Hum. Heredity* 79, 14–19. doi: 10.1159/000371387
- Monies, D., Abouelhoda, M., Alsayed, M., Alhassnan, Z., Alotaibi, M., Kayyali, H., et al. (2017). The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum. Genet.* 136, 921–939. doi: 10.1007/s00439-017-1821-8
- Nilsen, G., Liestol, K., Van Loo, P., Moen Vollan, H. K., Eide, M. B., Rueda, O. M., et al. (2012). Copynumber: efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* 13, 591. doi: 10.1186/1471-2164-13-591
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533, 539–539+. doi: 10.1038/nature17671
- Pemberton, T. J., and Rosenberg, N. A. (2014). Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. *Hum. Heredity* 77, 37–48. doi: 10.1159/000362878
- Plomin, R., and Von Stumm, S. (2018). The new genetics of intelligence. *Nat. Rev. Genet.* 19, 148–159. doi: 10.1038/nrg.2017.104
- Power, R. A., Keller, M. C., Ripke, S., Abdellaoui, A., Wray, N. R., Sullivan, P. F., et al. (2014a). A recessive genetic model and runs of homozygosity in major depressive disorder. *Am. J. Med. Genet. Part B-Neuropsychiatr. Genet.* 165, 157–166. doi: 10.1002/ajmg.b.32217
- Power, R. A., Nagoshi, C., Defries, J. C., Plomin, R., and Control, W. T. C. (2014b). Genome-wide estimates of inbreeding in unrelated individuals and their association with cognitive ability. *Eur. J. Hum. Genet.* 22, 386–390. doi: 10.1038/ejhg.2013.155
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340, 1467–1471. doi: 10.1126/science.1235488
- Rodriguez-Flores, J. L., Fakhro, K., Agosto-Perez, F., Ramstetter, M. D., Arbiza, L., Vincent, T. L., et al. (2016). Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Res.* 26, 151–162. doi: 10.1101/gr.191478.115
- Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *R J* 8, 289–317. doi: 10.32614/RJ-2016-021
- Shinwari, J. M. A., Al Yemni, E. A. A., Alnaemi, F. M., Abebe, D., Al-Abdulaziz, B. S., Al Mubarak, B. R., et al. (2017). Analysis of shared homozygosity regions in Saudi siblings with attention deficit hyperactivity disorder. *Psychiatric Genet.* 27, 131–138. doi: 10.1097/YPG.0000000000000173
- Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence (vol 49, 2017). *Nat. Genet.* 49, 1558–1558, pg 1107. doi: 10.1038/ng1017-1558c
- Spain, S. L., Pedrosa, I., Kadeva, N., Miller, M. B., Iacono, W. G., McGue, M., et al. (2016). A genome-wide analysis of putative functional and exonic variation associated with extremely high intelligence (vol 21, 2016). *Mol. Psychiatry* 21, 1152–1152, pg 1152. doi: 10.1038/mp.2015.145
- Steinfeld, H., Cho, M. T., Retterer, K., Person, R., Schaefer, G. B., Danylchuk, N., et al. (2016). Mutations in *HIVEP2* are associated with developmental delay, intellectual disability, and dysmorphic features. *Neurogenetics* 17, 158–163. doi: 10.1007/s10048-016-0479-z
- Szpiech, Z. A., Blant, A., and Pemberton, T. J. (2017). GARLIC: genomic autozygosity regions likelihood-based Inference and classification. *Bioinformatics* 33, 2059–2062. doi: 10.1093/bioinformatics/btx102
- Szpiech, Z. A., Xu, J. S., Pemberton, T. J., Peng, W. P., Zollner, S., Rosenberg, N. A., et al. (2013). Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* 93, 90–102. doi: 10.1016/j.ajhg.2013.05.003
- Tan, M., Aljughaiman, A., Elliott, J., Kornilov, S., Ferrando-Prieto, M., Bolden, D., et al. (2009). “Considering language, culture, and cognitive abilities: the international translation and adaptation of the Aurora assessment battery” in *Multicultural Psychoeducational Assessment*, ed. E.L. Grigorenko. NY: Springer, 443–468.
- Velez, J. I., Lopera, F., Patel, H. R., Johar, A. S., Cai, Y. P., Rivera, D., et al. (2016). Mutations modifying sporadic alzheimer’s disease age of onset. *Am. J. Med. Genet. Part B-Neuropsychiatr. Genet.* 171, 1116–1130. doi: 10.1002/ajmg.b.32493
- Vermunt, J. K., and Madigson, J. (2002). “Latent class cluster analysis,” in *Applied Latent Class Analysis*, eds. J. Hagenaars and A. McCutcheon (Cambridge: Cambridge University Press), 89–106S. doi: 10.1017/CBO9780511499531.004

- Verweij, K. J. H., Abdellaoui, A., Vejjola, J., Sebert, S., Koiranen, M., Keller, M. C., et al. (2014). The association of genotype-based inbreeding coefficient with a range of physical and psychological human traits. *Plos One* 9, e103102. doi: 10.1371/journal.pone.0103102
- Wang, A. H., Gregoire, S., Zika, E., Xiao, L., Li, C. S., Li, H. W., et al. (2005). Identification of the ankyrin repeat proteins ANKRA and RFXANK as novel partners of class IIa histone deacetylases. *J. Biol. Chem.* 280, 29117–29127. doi: 10.1074/jbc.M500295200
- Wang, H., Warner-Schmidt, J., Varela, S., Enikolopov, G., Greengard, P., and Flajolet, M. (2015). Norbin ablation results in defective adult hippocampal neurogenesis and depressive-like behavior in mice. *PNAS* 112, 9745–9750. doi: 10.1073/pnas.1510291112
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., et al. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665–1674. doi: 10.1101/gr.6861907
- Warsy, A. S., Al-Jaser, M. H., Albdass, A., Al-Daihan, S., and Alanazi, M. (2014). Is consanguinity prevalence decreasing in Saudis?: a study in two generations. *Afr. Health Sci.* 14, 314–321. doi: 10.4314/ahs.v14i2.5
- Xia, L., Ju, T., Westmuckett, A., An, G., Ivanciu, L., Mcdaniel, J. M., et al. (2004). Defective angiogenesis and fatal embryonic hemorrhage in mice lacking core 1-derived O-glycans. *J. Cell Biol.* 164, 451–459. doi: 10.1083/jcb.200311112
- Yamamoto, K., and Mak, T. W. (2017). Mechanistic aspects of mammalian cell size control. *Dev. Growth Differ.* 59, 33–40. doi: 10.1111/dgd.12334
- Yang, T. L., Guo, Y., Zhang, L. S., Tian, Q., Yan, H., Papasian, C. J., et al. (2010). Runs of homozygosity identify a recessive locus 12q21.31 for human adult height. *J. Clin. Endocrinol. Metab.* 95, 3777–3782. doi: 10.1210/jc.2009-1715
- Yavarna, T., Al-Dewik, N., Al-Mureikhi, M., Ali, R., Al-Mesaifri, F., Mahmoud, L., et al. (2015). High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. *Hum. Genet.* 134, 967–980. doi: 10.1007/s00439-015-1575-0
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–U136 doi: 10.1038/ng.2310

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Kornilov, Tan, Aljughaiman, Naumova and Grigorenko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.