

# Generalized correlation coefficient for genome-wide association analysis of cognitive ability in twins

Afsaneh Mohammadnejad<sup>1</sup>, Marianne Nygaard<sup>1</sup>, Shuxia Li<sup>1</sup>, Dongfeng Zhang<sup>2</sup>, Chunsheng Xu<sup>3</sup>, Weilong Li<sup>1</sup>, Jesper Lund<sup>1</sup>, Lene Christiansen<sup>1,4</sup>, Jan Baumbach<sup>5,6</sup>, Kaare Christensen<sup>1,7</sup>, Jacob v. B. Hjelmborg<sup>1</sup>, Qihua Tan<sup>1,7</sup>

<sup>1</sup>Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark

<sup>2</sup>Department of Epidemiology and Health Statistics, School of Public Health, Qingdao University, Qingdao, China

<sup>3</sup>Qingdao Center for Disease Control and Prevention, Qingdao, China

<sup>4</sup>Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen Ø, Denmark

<sup>5</sup>Computational Biomedicine, Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

<sup>6</sup>Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany

<sup>7</sup>Unit of Human Genetics, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

**Correspondence to:** Qihua Tan; email: [qtan@health.sdu.dk](mailto:qtan@health.sdu.dk)

**Keywords:** generalized correlation coefficient, GWAS, cognition, twins

**Received:** July 20, 2020

**Accepted:** September 29, 2020

**Published:** November 24, 2020

**Copyright:** © 2020 Mohammadnejad et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Despite a strong genetic background in cognitive function only a limited number of single nucleotide polymorphisms (SNPs) have been found in genome-wide association studies (GWASs). We hypothesize that this is partially due to mis-specified modeling concerning phenotype distribution as well as the relationship between SNP dosage and the level of the phenotype. To overcome these issues, we introduced an assumption-free method based on generalized correlation coefficient (GCC) in a GWAS of cognitive function in Danish and Chinese twins to compare its performance with traditional linear models. The GCC-based GWAS identified two significant SNPs in Danish samples (rs71419535,  $p = 1.47e-08$ ; rs905838,  $p = 1.69e-08$ ) and two significant SNPs in Chinese samples (rs2292999,  $p = 9.27e-10$ ; rs17019635,  $p = 2.50e-09$ ). In contrast, linear models failed to detect any genome-wide significant SNPs. The number of top significant genes overlapping between the two samples in the GCC-based GWAS was higher than when applying linear models. The GCC model identified significant genetic variants missed by conventional linear models, with more replicated genes and biological pathways related to cognitive function. Moreover, the GCC-based GWAS was robust in handling correlated samples like twin pairs. GCC is a useful statistical method for GWAS that complements traditional linear models for capturing genetic effects beyond the additive assumption.

## INTRODUCTION

Cognitive function is an important phenotype involving multiple mental abilities including learning, thinking, reasoning, remembering, problem-solving, decision-

making, and attention. Thus, cognitive function affects health and quality of life in the elderly. Recent twin and family studies have shown that the heritability of general cognitive function is more than 50% in adolescence through adulthood to older age [1],

suggesting a substantial genetic contribution to the phenotype. Despite the relatively large number of genome-wide association studies (GWASs) of cognition performed to date, a lot of the heritability is still unexplained. For instance, Savage et al., 2018 [2] reported the largest GWAS of intelligence to date ( $N = 269,867$ ), identifying 205 significantly associated genomic loci and  $h^2_{SNP} = 0.22$ . In addition, different limitations in current GWASs might be the reason for the situation due to the distribution of cognitive measurements as well as the complex relationship between SNP genotypes and cognitive performance. One of the assumptions in the popular linear models used for GWAS is normality of the phenotype of interest. Unfortunately, the distribution of cognitive measurements is typically skewed [3]. Also, the additive genetic effect has been the most popular assumption in current genetic association studies and the only genetic model addressed in studies using linear models. However, the pattern of genetic associations is likely more complex, including both linear (additive) and non-linear (non-additive) relationships [4, 5].

As a recent development, the generalized measure of dependency for analysis of omics data has been proposed [4–6]. The concept of a generalized measure of association comes from the rank correlation and information theory. The rank correlation was defined by Hoeffding's  $D$  by measuring the difference between the joint rank of two random variables  $x$  and  $y$  and the product of their marginal rank [7]. In 2012, Reshef [8] and his colleagues proposed the maximal information coefficient (MIC), which belongs to a larger class of maximal information-based nonparametric exploration (MINE), as a measure of association. MIC is able to not only measure two-variable dependence by calculating a score that is approximately equal to a coefficient of determination ( $R^2$ ) relative to the regression function, but also to detect non-linear relationships between variables. Recently, Murrell et al., 2016 [9] proposed an approach similar to MIC that additionally included the advantages of fast convergence with rising sample size and having more power to detect the relationships compared to MIC. Both MIC and Murrell's method follow desired properties of bivariate association, generality and equitability [9]. Murrell and his colleagues have suggested a new measure of bivariate association, which is the same as  $R^2$ , but powerful when the parametric form of association is unknown. A point of particular importance is that through applying this method both generality and equitability can be attained by estimating a generalized  $R^2$  based on density approximation. We believe that this method has advantages that merit it to be considered in GWAS as the associations between the SNP dosage and the phenotype are expected to be complex. More

importantly, this method does not rely on strict assumptions such as normality of the phenotype distribution as well as linearity of the genetic effect (or an additive genetic effect, which is assumed in most GWAS). Additionally, when related samples such as twins or family data are included in association studies, more model assumptions are imposed. Hence, the assumption-free or generalized measurements of association may be advantageous in genetic association studies.

In this study, we aimed to compare the performance of popular linear regression models and the generalized correlation coefficient (GCC) method to identify genetic variants, genes and pathways in Danish and Chinese GWASs of cognitive function.

## RESULTS

The study population comprised Danish and Chinese GWAS data. A description of the two samples is illustrated in Table 1. The means of cognitive function in Danish and Chinese data are 45.86 and 21.01 respectively. We validated the three models, GCC, Kinship, and mixed-linear model (LME), by estimating type I error rates after simulating genotypes for one SNP and a random phenotype (standard normal) for 1000 replications. The type I error rates estimated for GCC, Kinship and LME were 0.052, 0.052 and 0.050, respectively. This indicates that the three models are generally unbiased.

### Generalized correlation coefficient and linear models

The summary results for the top 30 SNPs from the GWASs performed in Danish and Chinese samples using each of the three models are shown in Supplementary Tables 1–6. The use of the GCC method resulted in the identification of two genome-wide significant ( $p < 5e-8$ ) SNPs in Danish samples: rs71419535 ( $p = 1.47e-08$ ) on chromosome 2 near the gene *THSD7B*, and rs905838 ( $p = 1.69e-08$ ) on chromosome 5 near the gene *CTD-2533K21.4*; and two genome-wide significant SNPs in Chinese samples: rs2292999 ( $p = 9.27e-10$ ) on chromosome 3 near the gene *ABCC5*, and rs17019635 ( $p = 2.50e-09$ ) on chromosome 4 near the gene *GRID2*. Circos plots for the genome-wide significant SNPs identified using GCC are shown in Figure 1. In contrast, no genome-wide significant SNPs were identified in either the Danish or Chinese sample when using Kinship or LME. The density plots of the three genotyped SNPs ( $p < 1e-06$ ) in the Danish and Chinese GWAS when using GCC are depicted in Figure 2. In general, if genotype and cognition are independent, the widest parts of the

**Table 1. Descriptive statistics of the 900 dizygotic twins and 272 single monozygotic twins included in the Danish GWAS and the 278 dizygotic twins included in the Chinese GWAS.**

GWAS Variables	Male	Female	Total
Danish sample	611	561	1172
Mean of age± sd (min, max)	66.51±6.05(56.4,79.88)	66.12±5.93(55.94,80.21)	66.32±6.0(55.94,80.21)
Mean of cognitive score± sd (min, max)	45.03±9.83(11.68,84.93)	46.74±9.94(21.16,83.69)	45.86±9.91(11.68, 84.93)
Chinese sample	141	137	278
Mean of age± sd (min, max)	51.04±7.04(41,70)	51.18±7.04(40,70)	51.11±7.03(40,70)
Mean of cognitive score± sd (min, max)	20.59±4.65(5,30)	21.44±4.77(5,29)	21.01±4.72(5,30)

density beans should be on the same level (height), around 0 on the Y-axis. The patterns in Figure 2 show clear deviations from independence as well as from an additive genetic effect. The QQ plots comparing the three models in both studies are shown in Figure 3A (Danish) and 3B (Chinese). Across the three models, GCC was able to detect non-random SNPs as compared with the kinship and LME models. The genomic inflation factors for GCC, Kinship and LME were 1.061, 0.995 and 0.998 for the Danish sample and 1.059, 1.013 and 1.033 for the Chinese sample. Supplementary Figure 1 shows QQ plots comparing the GCC GWAS results with the GWAS results obtained using each of the linear models. Both plots show a higher efficiency of GCC than of kinship and LME. Figure 4 compares SNP p-values from the GCC GWAS with SNP p-values from the kinship model GWAS in Danish (4A) and Chinese (4B) samples, with suggestively significant SNPs ( $p < 1e-05$ ) colored red if identified using GCC and green if identified using the kinship (linear) model. Similar comparisons of the GCC GWAS results with those of the LME model GWAS are shown in Supplementary Figure 2. Manhattan plots for all three models in both samples are shown in Supplementary Figure 3.

### Post-GWAS analysis in FUMA

For the GCC GWAS results in the Danish and Chinese samples, the post-GWAS analysis in FUMA identified 29 and 24 independent SNPs with  $p < 1e-6$ , 862 and 618 mapped genes, as well as 180 and 204 tagged SNPs that are in linkage disequilibrium (LD) ( $r^2 < 0.6$ ) with the independent significant SNPs, respectively. Most of the tagged SNPs identified in FUMA from the Danish GWAS results are positioned in intergenic and ncRNA intronic regions (Supplementary Figure 4). Similarly, most of the tagged SNPs identified from the Chinese GWAS results are positioned in intergenic and intronic regions, but with a smaller proportion in downstream regions (Supplementary Figure 5). In contrast, most of the tagged SNPs identified from the Kinship and LME GWAS results are in ncRNA-intronic and intergenic regions (Supplementary Figures 6–8).

### Gene-based test in MAGMA

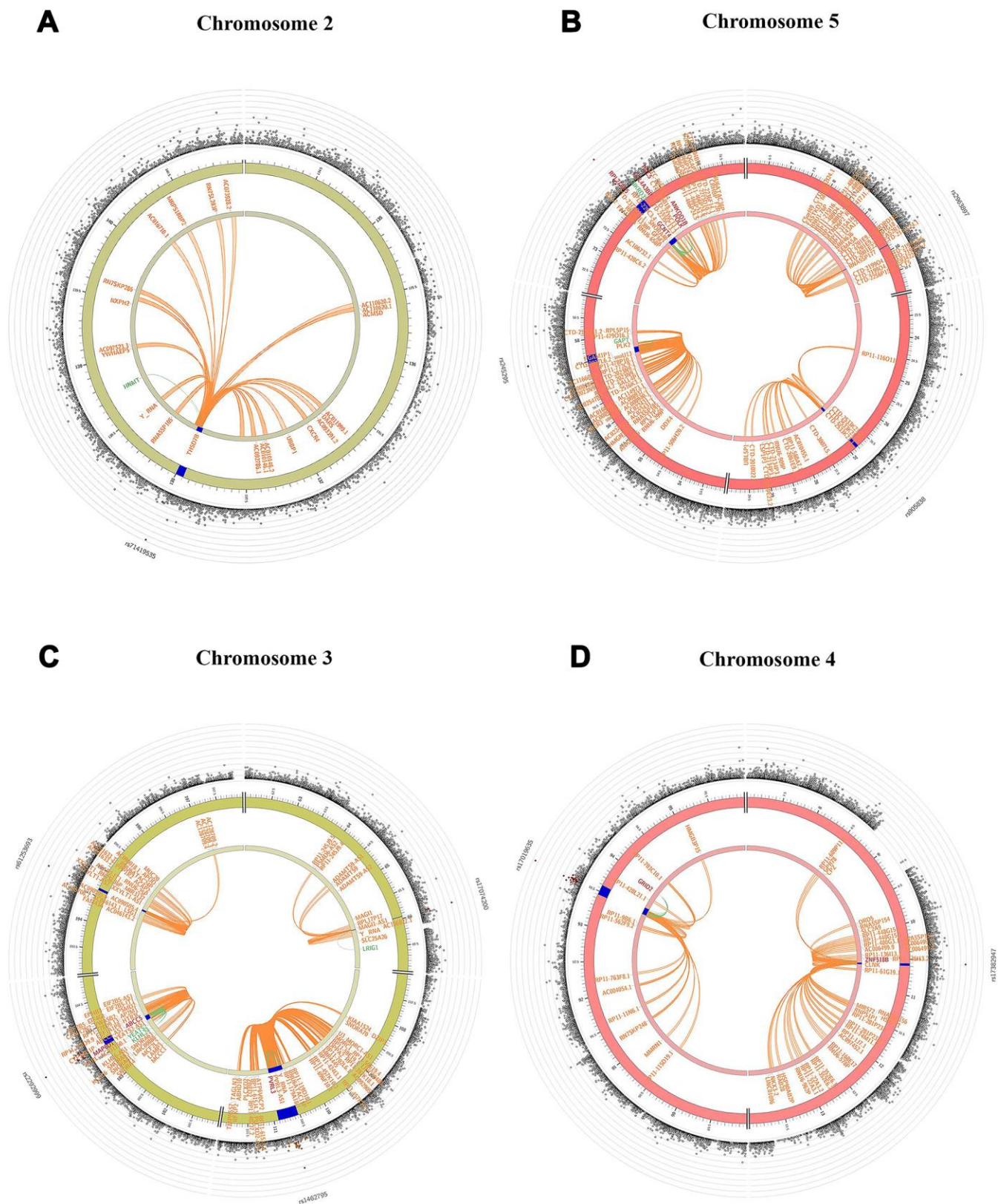
The MAGMA gene-set analysis was performed for all SNPs based on the results of the GWAS using each of the three modules. In the Danish and Chinese samples, respectively,  $1115/18787 \approx 5.93\%$  and  $1024/18918 \approx 5.41\%$  of genes from the GCC GWAS,  $1046/18787 \approx 5.57\%$  and  $1069/18918 \approx 5.65\%$  of genes from the kinship model GWAS, and  $1034/18787 \approx 5.50\%$  and  $1014/18918 \approx 5.36\%$  of genes from the LME model GWAS were identified with a  $p < 0.05$ .

Among the genes identified with a  $p < 0.05$  in the Danish sample, a total of 476, 50 and 54 genes were replicated in the Chinese sample in GCC GWAS results, LME model GWAS results, and Kinship model GWAS results respectively. A Venn diagram showing the number of overlapping genes for both Danish and Chinese samples is shown in Figure 5A, 5B.

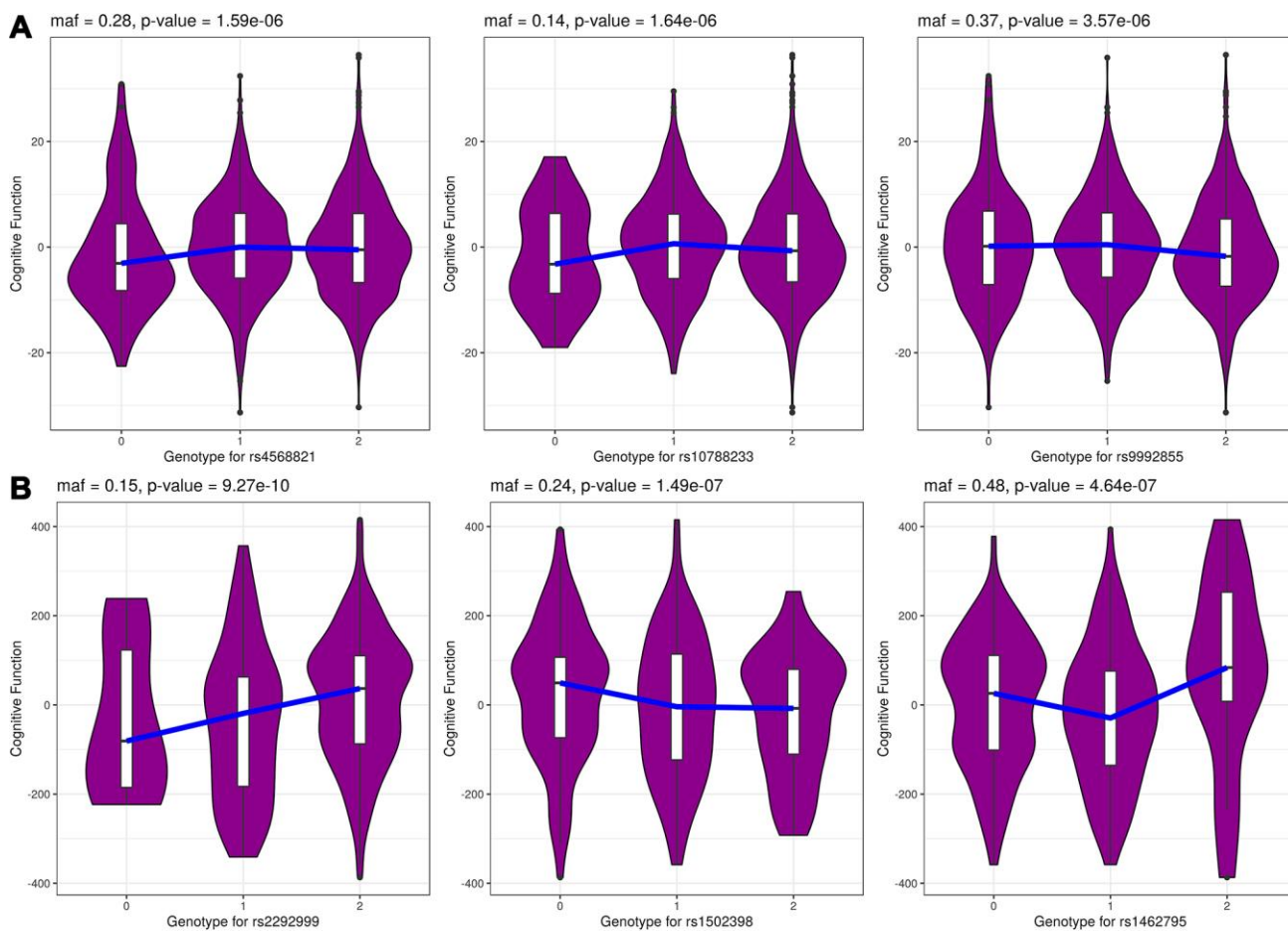
The two top genes identified from the discovery data GCC GWAS results were *CSMD1* ( $p = 5.63e-07$ , FDR = 0.01) and *PTPRD* ( $p = 1.68e-5$ , FDR = 0.16), which were both replicated in the Chinese data (Supplementary Table 7, 8). The two top genes identified from the Kinship model and LME model GWAS results in the Danish sample were *MNT* (Kinship model  $p = 1.70e-5$ , LME model  $p = 1.79e-5$ ) and *MS4A2* (Kinship model  $p = 8.3e-5$ , LME model  $p = 8.69e-5$ ), while the two top genes identified in the Chinese sample were *TEX26* (Kinship model  $p = 2.21e-7$ , LME model  $p = 4.12e-6$ ) and *WDPCP* (Kinship model  $p = 1.13e-6$ , LME model  $p = 1.83e-6$ ).

### Biological pathway analysis

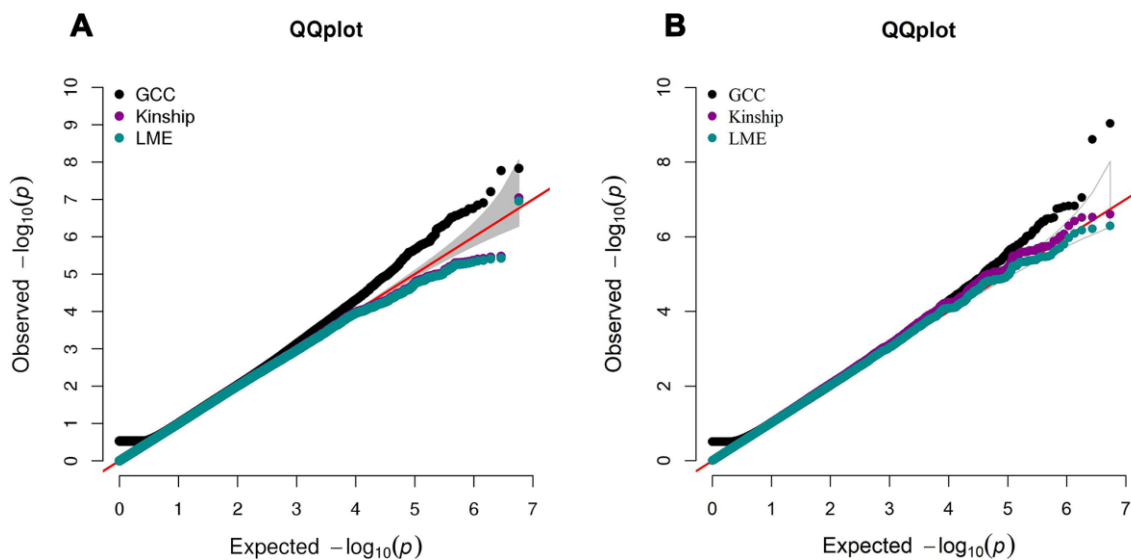
Finally, the results of the overrepresentation analysis performed in GSEA based on genes with  $p < 0.05$  in the gene-based analysis are shown in Supplementary Tables 9–15. The Significant (FDR < 0.05) KEGG pathways identified in the GSEA analysis for genes with  $p < 0.05$  found to overlap in gene-based analyses of both GCC and linear model results are illustrated in Table 2. By using



**Figure 1.** Circos plots indicating genes from genome-wide significant SNPs on chromosomes 2 (A), 5 (B) in Danish sample and 3 (C), 4 (D) in Chinese sample based on GCC model. The blue region shows the genomic risk region. Genes mapped by chromatin interaction, eQTL and both are displayed in orange, green and red respectively. The most outer layer shows a Manhattan plot only for SNPs with  $p < 0.05$  and SNPs are colored in red based on linkage disequilibrium (LD) patterns with the lead SNPs.



**Figure 2.** The genotype-specific density distribution for top 3 genotyped SNPs in Danish (A) and Chinese (B) samples from GCC model. The x-axis shows the SNP genotypes 0, 1 and 2 and y-axis shows cognitive function phenotype.

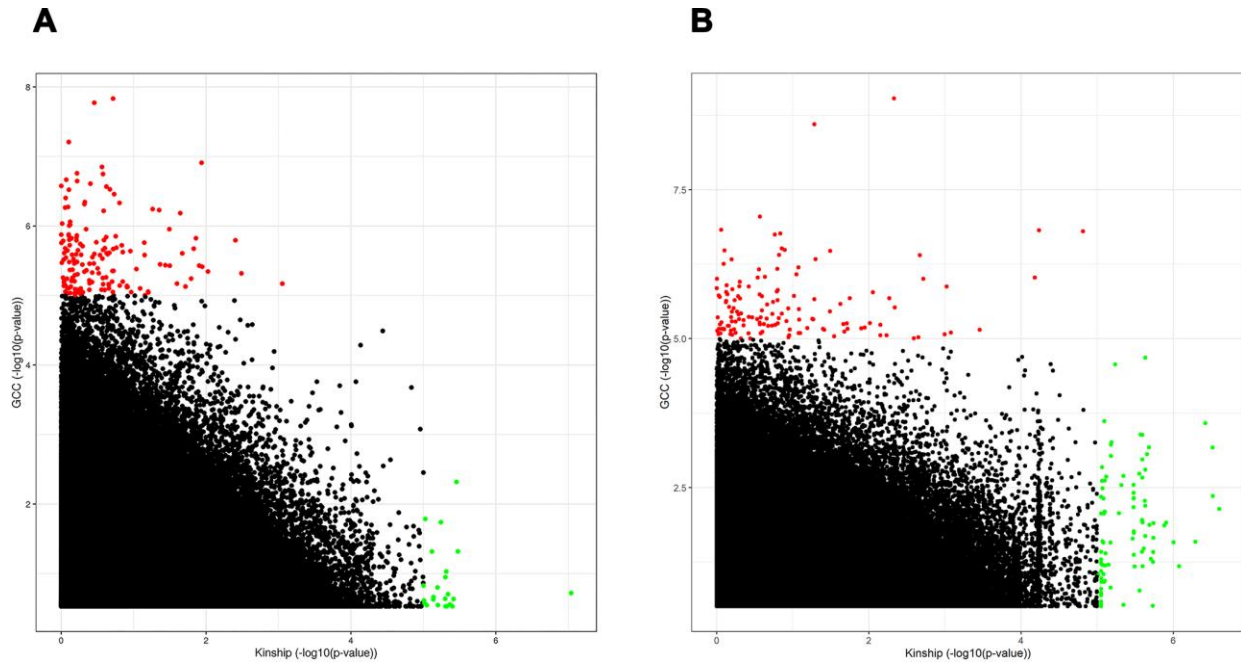


**Figure 3.** QQ plot comparing the performance of GCC, kinship and LME models in Danish (A) and Chinese (B) GWAS data. The left QQ plot is from Danish sample and the right QQ plot is from Chinese sample. In each plot, x-axis is the expected p-value and y-axis is the observed p-value from the GWAS.

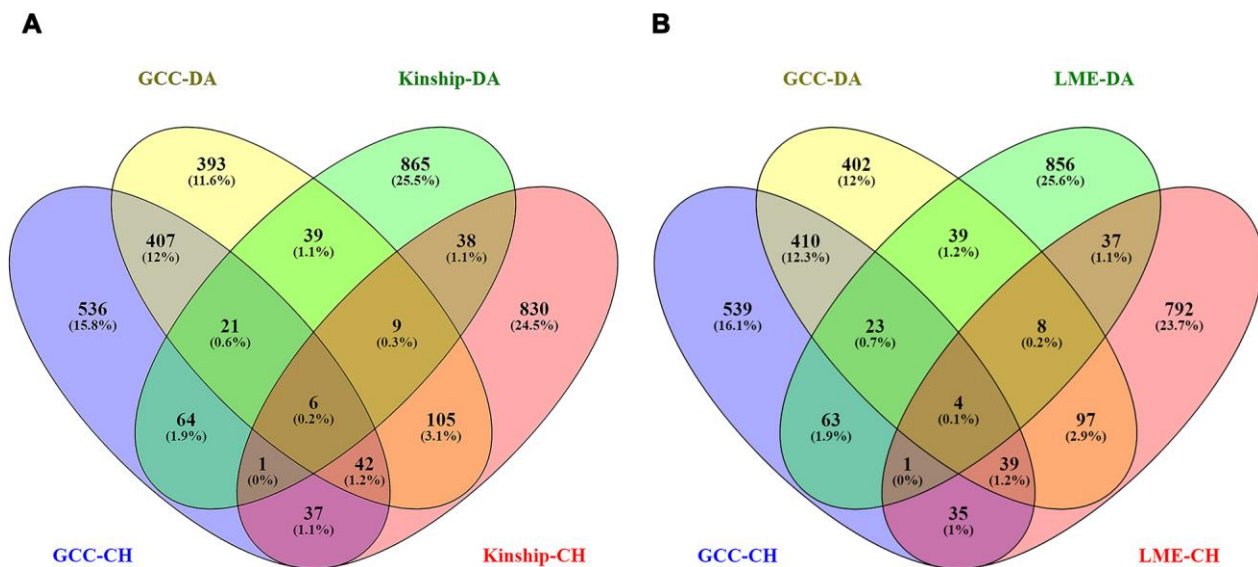
genes with  $p < 0.05$  overlapping in the GCC results from the two samples, 28 KEGG pathways with  $FDR < 0.05$  were found (Supplementary Table 15). In contrast, the 54 and 50 replicated genes from the Kinship and LME analyses were not adequate for pathways analysis.

## DISCUSSION

The literature on GWAS data analysis has been dominated by multiple assumptions (e.g. additive or dominant genetic effects, and normal distribution of



**Figure 4.** Scatter plot comparing the performance of SNPs in linear model to the GCC model in both Danish (A) and Chinese (B) samples. The x-axis and y-axis show  $-\log_{10}(p\text{-value})$  from Kinship and GCC models respectively.



**Figure 5.** The Venn diagrams showing the number of overlapped genes with  $p < 0.05$  among GCC and Kinship models from both samples in left plot (A) and GCC and LME models in the right plot (B). The total number of genes included are GCC-DA (GCC in Danish data): 1115, GCC-CH (GCC in Chinese data): 1024, Kinship-DA (Kinship in Danish data): 1046, Kinship-CH (Kinship in Chinese data): 1069, LME-DA (LME in Danish data): 1034 and LME-CH (LME in Chinese data): 1014.

**Table 2. Significant KEGG pathways (FDR< 0.05) overlapping between GCC, Kinship model and LME model GWAS results from both Danish and Chinese samples.**

Model	Gene set name	Description	P-value	FDR q-value
GCC-DA-CH*	KEGG_PATHWAYS_IN_CANCER	Pathways in cancer	1.08e-12	2.01e-10
GCC-DA-CH	KEGG_AXON_GUIDANCE	Axon guidance	1.16e-10	1.08e-08
GCC-DA-CH	KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2.27e-09	1.41e-07
GCC-DA-CH	KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	Vascular smooth muscle contraction	4.67e-09	2.17e-07
GCC-DA-CH	KEGG_CALCIIUM_SIGNALING_PATHWAY	Calcium signaling pathway	8.07e-09	3e-07
GCC-DA-CH	KEGG_FOCAL_ADHESION	Focal adhesion	1.51e-08	4.67e-07
GCC-DA-CH	KEGG_MAPK_SIGNALING_PATHWAY	MAPK signaling pathway	2.35e-07	6.26e-06
GCC-DA-CH	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	Hypertrophic cardiomyopathy (HCM)	6.33e-07	1.47e-05
GCC-DA-CH	KEGG_DILATED_CARDIOMYOPATHY	Dilated cardiomyopathy	1.64e-06	3.04e-05
GCC-DA-CH	KEGG_LONG_TERM_DEPRESSION	Long-term depression	4.43e-06	7.5e-05
GCC-DA-CH	KEGG_GNRH_SIGNALING_PATHWAY	GnRH signaling pathway	6.09e-06	9.44e-05
GCC-DA-CH	KEGG_ADHERENS_JUNCTION	Adherens junction	6.75e-06	9.65e-05
GCC-DA-CH	KEGG_BASAL_CELL_CARCINOMA	Basal cell carcinoma	2.34e-05	0.00029
GCC-DA-CH	KEGG_HEDGEHOG_SIGNALING_PATHWAY	Hedgehog signaling pathway	2.72e-05	0.000294
GCC-DA-CH	KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	Regulation of actin cytoskeleton	4.39e-05	0.000409
GCC-DA-CH	KEGG_GAP_JUNCTION	Gap junction	5.08e-05	0.00045
GCC-DA-CH	KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	Phosphatidylinositol signaling system	5.87e-05	0.000496
GCC-DA-CH	KEGG_CELL_ADHESION_MOLECULES_CAMS	Cell adhesion molecules (CAMs)	0.000116	0.000896
GCC-DA-CH	KEGG_ECM_RECEPTOR_INTERACTION	ECM-receptor interaction	0.000138	0.000989
GCC-DA-CH	KEGG_TYPE_II_DIABETES_MELLITUS	Type II diabetes mellitus	0.000347	0.00239
GCC-DA-CH	KEGG_MELANOGENESIS	Melanogenesis	0.000624	0.00387
Kin-LME-DA-CH	KEGG_FOCAL_ADHESION	Focal adhesion	0.000777	0.0361
Kin-LME-DA-CH**	KEGG_WNT_SIGNALING_PATHWAY	Wnt signaling pathway	0.001980	0.0462
Kin-LME-DA-CH	KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	T cell receptor signaling pathway	0.002090	0.0462

\*GCC-DA-CH: Pathways overlapping in GCC GWAS results from both Danish (DA) and Chinese (CH) samples.

\*\*Kin-LME-DA-CH: Pathways overlapping in Kinship model and LME model GWAS results from both Danish and Chinese samples.

phenotypes), which, if not fulfilled, could be speculated to be responsible for low statistical power, missing heritability and the lack of replication of GWAS findings. Our study, through model comparison and validation of GWAS results from two independent samples of different ethnicity, illustrates the strength of the GCC method over popular linear models, with more replicated GCC-based results and cognition-implicated SNPs and functional pathways.

Among the two genome-wide significant SNPs identified in the GCC-based GWAS in the Danish sample, rs71419535 is near *THSD7B*, which is associated with melanoma, metabolism, and Cutaneous Malignant 1 diseases. Variation in this gene has been shown to associate with multiple system atrophy, which is an adult-onset neurodegenerative disorder [10]. The other SNP, rs905838, is near *CTD-2533K21.4*, which is a novel transcript highly expressed in brain hippocampus.

Among the two genome-wide significant SNPs found in the GCC-based GWAS in the Chinese sample, rs2292999 on chromosome 3 is located near *ABCC5*, which is expressed in both brain and muscle tissue, and rs17019635 on chromosome 3 near *GRID2*, which is expressed in the brain. The *ABCC5* gene has previously been found to encode a general glutamate conjugate and analog transporter that can limit the brain levels of endogenous metabolites, drugs, and toxins [11]. *ABCC5* is additionally related to the ‘Blood-Brain Barrier’ and ‘Immune Cell Transmigration’ pathways. Glutamate is the principal excitatory neurotransmitter in the brain. It is crucial for learning and memory in everyday brain functions but causes excitotoxic damage in traumatic brain injury and stroke [11, 12]. The *GRID2* gene is highly expressed in brain and associated with depression. This gene plays a role in synapse organization between parallel fibers and Purkinje cells (a class of GABAergic neurons in the cerebellum in the brain). The implications of GABAergic neurotransmission in Alzheimer’s disease (AD) have been discussed [13–15], and in the processes of learning and memory, changes in GABAergic function could be an important factor in both early and later stages of AD pathogenesis [15]. Furthermore, the two top genes from the GCC-based GWAS found in both the Danish and Chinese samples, *CSMD1* and *PTPRD*, are very interesting; a variant in *CSMD1* was recently associated with cognitive function [16], and in a very recent study, the *PTPRD* gene was discussed as a druggable target, and a marker for and an important constituent of brain circuits of likely importance for major brain-based phenotypes [17].

Although some significant pathways were identified from the GWAS results generated using all models, more biologically meaningful pathways were found when using the GCC method. Importantly, more genes and pathways identified from the CGG GWAS results were replicated in the Chinese sample than was the case for the linear models. Some replicated pathways from the GCC model GWAS results were pathways in calcium signaling, Type II diabetes mellitus, long-term depression, melanogenesis, axon guidance, focal adhesion, MAPK-signaling, Hedgehog signaling, GnRH signaling, and some related to cardiomyopathy (Table 2). The replicated pathways from the linear model GWAS results were related to focal adhesion, Wnt signaling, and T cell receptor signaling (Table 2).

Calcium signaling and potassium channels have been reported in aging-related diseases, including cognitive function. Calcium signaling is a crucial messenger between synapse and nucleus as calcium transient in nucleus is needed for neuroadaptations to switch on the necessary genes [18]. Additionally, dysregulated

expression of calcium signaling genes has been shown to occur with progression of Alzheimer’s-type pathology in the aging brain [19]. Type II diabetes mellitus is discussed in numerous studies because of the presumed role of it in increasing the risk of cognitive impairment and dementia [20, 21]. Focal adhesion is involved in integrin adhesion, communication between the extracellular matrix and the actin cytoskeleton, and the regulation of many cell types. Loss of cell adhesion can lead to cell death and altered focal signaling has been linked to synaptic loss, which may cause AD [22, 23]. There is evidence that axon guidance might play a role in some brain disorders such as Parkinson’s and AD [24].

The Wnt signaling pathway is also important and many recent studies have discussed and reported this pathway in relation to aging and AD, where the loss of canonical Wnt signaling has been found to be involved in the pathogenesis of AD, and play a role in synaptic plasticity and maintenance in the adult brain [25, 26].

In the literature, there is evidence regarding the role of MAPK-signaling, long-term depression, Hedgehog signaling, GnRH signaling, melanogenesis, T cell receptor pathways, cancer, vascular smooth muscle cells and hypertrophic cardiomyopathy (HCM) on cognitive function, Alzheimer’s, Parkinson’s and Huntington’s diseases, and age-related neurodegenerative disorders [19, 27–34].

Moreover, worth mentioning are the pathways ‘neuroactive ligand receptor interaction’ and ‘Alzheimer’s disease’ that were identified among the 28 KEGG pathways identified from the 476 genes overlapping between the gene-based analysis of Danish and Chinese GCC-based GWAS results (Supplementary Table 15). The neuroactive ligand receptor interaction is important as it comprises genes involved in transmission across chemical synapses [19], and cognitive decline is the earliest sign of AD.

In Figure 4, it is indicated that the statistical significance or the p-values for the same SNPs identified by the GCC and the kinship models are very different. This can potentially be explained by the fact that the linear model is unable to capture SNPs with non-additive effects, while SNPs with an additive effect are best fitted by the linear model. Based on this, we believe that to detect SNPs with an additive effect and SNPs with a non-additive effect both the GCC and a linear model should be applied on the same data when performing a GWAS, i.e. with GCC as a complementary approach to the conventional linear model. As shown in Figure 4, the number of SNPs with



a non-additive effect is relatively large compared to the number of SNPs with an additive effect in both the Danish and Chinese samples. This finding is highly important because it suggests that a considerable number of significant SNPs could have been missed in GWASs performed so far due to the search being limited to SNPs with an additive genetic effect.

A limitation of the GCC method, and different from the conventional linear models that report regression coefficients with a direction of effect (+ or -), is that the association parameter from GCC,  $A$ , has no direction, which is understandable because direction of effect does not make sense in case of a non-linear relationship. We argue that, as a hypothesis-free or agnostic approach, the primary goal is to identify reliably significant SNP markers. The pattern of association for specific markers can be examined empirically by plotting the data to reveal the various relationships not limited to minus or plus. For SNPs not captured by the linear model, the linear model results can still provide reference information about the direction of the GCC correlation. However, as an assumption-free approach, GCC is inherently robust in handling correlated or structured samples in association testing. The QQ plots in Figure 3 all show that the statistical significance estimated by GCC is not affected by the twin correlation in our samples, with genomic inflation factors comparable to those found for kinship and mixed effect models, which are models specifically dealing with the twin correlation in our samples. In fact, the null distribution for estimating GCC is calculated based on the marginal distributions of the correlated twin samples. As a result, the association assessment by GCC is conditional on the correlated structure in the twin samples, ensuring unbiased estimates of statistical significance. Moreover, unlike the linear models, GCC is insensitive to outliers or extreme phenotype values. All of the above features make GCC a valuable approach for GWAS of complex phenotypes.

## CONCLUSIONS

As a generalized measurement of association, GCC is capable of capturing different patterns of genotype-phenotype correlation, not limited to additive genetic effects. As an assumption-free approach, GCC is robust in dealing with correlated samples, e.g. twins, and skewed phenotype distributions, e.g. cognition, as compared to the linear models. By exemplifying and illustrating the advantages and the necessity of GCC in GWAS, this study could serve as a precedent for application and further development of assumption-free methods to complement current GWAS on cognitive traits and on other complex traits as well.

## MATERIALS AND METHODS

### Participants and cognitive measurements

The Danish sample comprises 1172 participants aged 50-80 years, including 440 same-sex dizygotic (DZ) (238 male, 202 female) twins, 460 (230 male, 230 female) opposite-sex DZ twins and 272 (143 male, 129 female) single monozygotic (MZ) twins recruited by the Danish Twin Registry as a part of the study of Middle-Aged Danish Twins (MADT) (<https://pubmed.ncbi.nlm.nih.gov/31544734/>). General cognitive functioning was assessed by a cognitive composite score computed from five brief cognitive tests evaluating verbal fluency, attention and working memory (digits forward and digits backward), and memory (immediate and delayed word recall) [35]. Scores obtained from each cognitive test were first standardized using the means and standard deviations of the MADT participants who were 46-50-year-old born from 1949-1952, and then the 5 scores were summed to obtain the composite score. If individuals had two or more scores missing, the cognitive composite score was coded as not available. If individuals had just one missing score, their composite score was computed by prorating the 4 scores they had (i.e., by multiplying by 5/4) [36] (Table 1). The 5 test scores were given equal weight and at most one missing item was allowed. The final cognitive composite score was standardized to a mean of 50 and an SD of 10. For the Danish sample, written informed consent was obtained from all participants and the study was approved by the Regional Committees on Health Research Ethics for Southern Denmark (S-VF-19980072).

Participants in the Chinese sample were 278 dizygotic (DZ) twins aged 40-70 years, including 41 male, 39 female, and 59 opposite sex twin pairs, who completed the Montreal Cognitive Assessment (MoCA) questionnaire to measure their cognitive function and had blood samples taken for DNA extraction and genotyping (Table 1). All voluntary twin subjects gave their informed consent and then completed the Montreal Cognitive Assessment (MoCA) questionnaire and physical examination under the supervision of physicians at the Qingdao CDC. Each twin pair was interviewed face to face by the same well-trained and experienced investigator. Details about sample collection and ethical approval have been described elsewhere [37]. For the Chinese sample, ethical approval was obtained from the Institutional Review Boards of Qingdao Center for Disease Control and Prevention (Qingdao CDC). The study was performed in accordance with the ethical principles of the Helsinki Declaration. All individuals provided a written consent prior to completing the questionnaire and venous blood collection for DNA extraction at Qingdao CDC.

## Genotyping and SNP imputation

Genome-wide SNP genotyping for all the twins in Danish data was conducted using Illumina PsychArray (Illumina, San Diego, California, USA). Genotyping was conducted by the SNP&SEQ Technology Platform, Science for Life Laboratory, Uppsala, Sweden (<http://snpseq.medsci.uu.se/genotyping/snp-services/>). Pre-imputation quality control (QC) removed SNPs on genotype call rate < 98%, Hardy-Weinberg equilibrium (HWE)  $p < 10^{-6}$ , and MAF = 0, and individuals on sample call rate < 99%, relatedness and gender mismatch. Pre-phasing and imputation to the 1000 Genomes phase 3 reference panel was performed using IMPUTE2 [2]. The Chinese data genotyping was implemented using the Illumina Infinium Omni2.5Exome-8v1.2BeadChip platform (Illumina, San Diego, California, USA) which contains 2,608,742 SNPs. Before imputation, QC was conducted in PLINK 2 [38] to remove SNPs with a minor allele frequency (MAF) < 0.05, a HWE  $p < 1e-4$ , a call-rate < 0.98 as well as SNPs with strand issues. Imputation was done via the Michigan imputation server [39], using the 1000 Genomes phase 3 reference panel. In post-imputation QC for both datasets, SNPs were filtered according to the following thresholds: 1) a minimum imputation INFO score (information metric) of > 0.6, and 2) a MAF of at least 5%. This resulted in 5,779,266 and 5,432,814 genotyped and imputed SNPs to be included in the statistical analysis in the Danish and Chinese samples, respectively. Only autosomal SNPs were analyzed in both studies.

## Statistical analysis

### Generalized correlation coefficient and parametric linear models

We performed the association test of cognitive scores with imputed allele dosages using the mixed-effect kinship model (Kinship) from the Kinship2 package [40], the mixed-linear model (LME) from the *lme4* package [41], and the GCC model from the *matie* package [9] in R. In the association tests, the cognitive measurements were adjusted for age and sex. Normality of the cognitive score phenotypes is not a requirement in GCC analysis, whereas it is crucial in both linear methods. Hence, in the Chinese data we used Box-Cox transformation to ensure the normality of data for all three models in order to make an unbiased comparison among the models. The kinship model calculates a kinship matrix and integrates it in the covariance matrix of the genetic data, whereas the linear mixed-effect model corrects for the twin correlation in the sample by including random effects in the model.

The GCC is a generalized measure of association between variables. The association strength, which is called  $A$ , ranges from 0 (in case variables are

independent) to 1 (in the case of perfect association between variables).  $A$  is the square of the correlation coefficient and can be considered as the proportion of variance in one variable explained by another variable or a number of variables. As the explained proportion of variance is one minus the unexplained proportion of variance, then the proportion of explained variance,  $R^2$ , can be expressed as:  $1 - \sigma_{error}^2 / \sigma_{total}^2$ , where  $\sigma_{total}^2$  and  $\sigma_{error}^2$  are average square deviations from a flat “null” model and a deterministic “alternative” model. Based on the assumption of normality of observations from least square regression, the deviation from a point on the regression line can be expressed as a probability density [9]. In a different way, the null model in GCC assumes that the variables are independent so the joint density for two variables  $x$  and  $y$  is the product of the marginal density estimation as  $P(x)P(y)$ . The alternative model is considered as a mixture distribution of both dependent and independent components. It calculates a generalized  $R^2$  based on the probability density ratio of null and alternative models [9].

$$R^2 = 1 - \prod_i \left( \frac{P(x_i, y_i | null)}{P(x_i, y_i | alt)} \right)^{2/n} \quad (1)$$

To calculate a p-value for assessing the significance of the test, i.e. to check the departure of observations from independence, it calculates a cross-validation likelihood for both the null and the alternative model. Equation (2) shows the calculation of the cross-validation likelihood statistics (CVLRS):

$$CVLRS = -2 \log \left( \frac{Lcv(Null)}{Lcv(Alt)} \right), \quad (2)$$

where  $Lcv(Null)$  and  $Lcv(Alt)$  are cross-validation likelihoods for the null and alternative models [9].

## Post-GWAS analysis

### Gene-based test and functional mapping in FUMA

Genomic risk loci were defined from the SNP-based GWAS results of the GCC, kinship and LME model analyses using functional mapping and annotation of genetic associations (FUMA) [42]. As part of post-GWAS analyses, a gene-based test was performed in MAGMA [43] integrated in FUMA. From GWAS summary statistic results, it tests the joint distribution of all SNP p-values. It maps SNPs to genes (window threshold = 10kb) based on their genomic location, using 1000 Genomes Phase3 as a reference, and LD within and between genes is determined. Furthermore, we investigated eQTL mapping, using the GTEx portal (<https://gtexportal.org/home/>), for the significant SNPs from the GCC-based GWAS to detect tissue-

specific gene expression associated with the identified genetic.

### **Gene-set enrichment analysis (GSEA)**

To further investigate the biological function of the genes identified in the gene-based analysis from MAGMA, we used the top genes with  $p < 0.05$  identified using the results of analyses with each of the three models in both Danish and Chinese samples as input for GSEA using KEGG pathways ( $N = 186$ ). A hypergeometric test is used in GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) to test the enrichment of genes against the KEGG pathways from MSigDB gene-sets. Results with FDR  $q$ -value  $< 0.05$  are reported.

### **Abbreviations**

SNPs: single nucleotide polymorphisms; GWASs: genome-wide association studies; GCC: generalized correlation coefficient; MIC: maximal information coefficient; MINE: maximal information-based nonparametric exploration; LD: linkage disequilibrium; AD: Alzheimer's disease; FDR: false discovery rate; DZ: dizygotic; MZ: monozygotic; MADT: Middle-Aged Danish Twins; MoCA: Montreal Cognitive Assessment; MAF: minor allele frequency; HWE: Hardy-Weinberg equilibrium; LME: mixed-linear model; GSEA: Gene-set enrichment analysis.

### **AUTHOR CONTRIBUTIONS**

AM, QT, MN, JH conceived the study. AM performed data analysis. AM, MN, QT drafted the manuscript. LC and MN contributed for sample collections in the Danish data and DZ contributed for sample collections in the Chinese data. All authors contributed in the manuscript, commented, and approved the submitted manuscript.

### **CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

### **FUNDING**

This study was jointly supported by the Independent Research Fund Denmark (grant number DFF – 6110-00114) and PhD scholarship grant from the Faculty of Health Science, University of Southern Denmark. Genotyping of the Danish twins was conducted by the SNP&SEQ Technology Platform, Science for Life Laboratory, Uppsala, Sweden (<http://snpseq.medsci.uu.se/genotyping/snp-services/>) and supported by NIH R01 AG037985 (Pedersen).

## **REFERENCES**

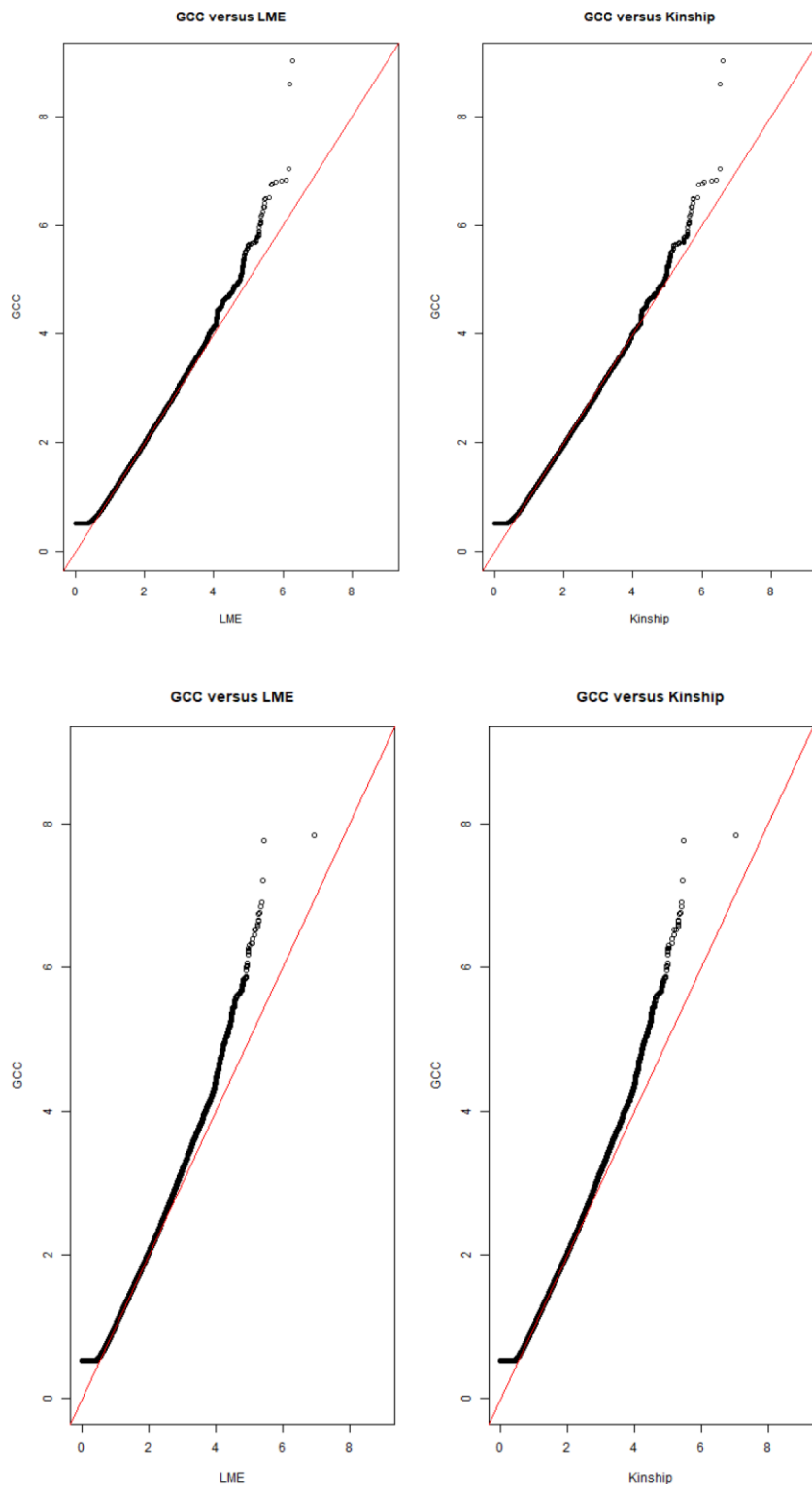
1. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, Hagenaars SP, Ritchie SJ, Marioni RE, Fawns-Ritchie C, Liewald DCM, Okely JA, Ahola-Olli AV, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun.* 2018; 9:2098. <https://doi.org/10.1038/s41467-018-04362-x> PMID:29844566
2. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, Nagel M, Awasthi S, Barr PB, Coleman JR, Grasby KL, Hammerschlag AR, Kaminski JA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet.* 2018; 50:912–19. <https://doi.org/10.1038/s41588-018-0152-6> PMID:29942086
3. Malek-Ahmadi M, Mufson EJ, Perez SE, Chen K. Statistical considerations for assessing cognition and neuropathology associations in preclinical Alzheimer's disease. *Biostat Epidemiol.* 2017; 1: 92–104. <https://doi.org/10.1080/24709360.2017.1342186>
4. Tan Q, Tepel M, Rasmussen LM, Hjelmborg JB. Generalized measure of dependency for analysis of omics data. *J Data Mining Genomics Proteomics.* 2015; 7:2153–0602. <https://doi.org/10.4172/2153-0602.1000183>.
5. Tan Q, Thomassen M, Burton M, Mose KF, Andersen KE, Hjelmborg J, Kruse T. Generalized correlation coefficient for non-parametric analysis of microarray time-course data. *J Integr Bioinform.* 2017; 14:20170011. <https://doi.org/10.1515/jib-2017-0011> PMID:28753536
6. Speed T. Mathematics. A correlation for the 21st century. *Science.* 2011; 334:1502–03. <https://doi.org/10.1126/science.1215894> PMID:22174235
7. de Siqueira Santos S, Takahashi DY, Nakata A, Fujita A. A comparative study of statistical methods used to identify dependencies between gene expression signals. *Brief Bioinform.* 2014; 15:906–18. <https://doi.org/10.1093/bib/bbt051> PMID:23962479
8. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. *Science.* 2011; 334:1518–24. <https://doi.org/10.1126/science.1205438> PMID:22174245

9. Murrell B, Murrell D, Murrell H. Discovering general multidimensional associations. *PLoS One*. 2016; 11:e0151551.  
<https://doi.org/10.1371/journal.pone.0151551>  
PMID:[26991498](https://pubmed.ncbi.nlm.nih.gov/26991498/)
10. Sailer A, Scholz SW, Nalls MA, Schulte C, Federoff M, Price TR, Lees A, Ross OA, Dickson DW, Mok K, Mencacci NE, Schottlaender L, Chelban V, et al, and European Multiple System Atrophy Study Group, and the UK Multiple System Atrophy Study Group. A genome-wide association study in multiple system atrophy. *Neurology*. 2016; 87:1591–98.  
<https://doi.org/10.1212/WNL.0000000000003221>  
PMID:[27629089](https://pubmed.ncbi.nlm.nih.gov/27629089/)
11. Jansen RS, Mahakena S, de Haas M, Borst P, van de Wetering K. ATP-binding cassette subfamily C member 5 (ABCC5) functions as an efflux transporter of glutamate conjugates and analogs. *J Biol Chem*. 2015; 290:30429–40.  
<https://doi.org/10.1074/jbc.M115.692103>  
PMID:[26515061](https://pubmed.ncbi.nlm.nih.gov/26515061/)
12. Danbolt NC. Glutamate uptake. *Prog Neurobiol*. 2001; 65:1–05.  
[https://doi.org/10.1016/s0301-0082\(00\)00067-8](https://doi.org/10.1016/s0301-0082(00)00067-8)  
PMID:[11369436](https://pubmed.ncbi.nlm.nih.gov/11369436/)
13. Wu C, Sun D. GABA receptors in brain development, function, and injury. *Metab Brain Dis*. 2015; 30:367–79.  
<https://doi.org/10.1007/s11011-014-9560-1>  
PMID:[24820774](https://pubmed.ncbi.nlm.nih.gov/24820774/)
14. Li Y, Sun H, Chen Z, Xu H, Bu G, Zheng H. Implications of GABAergic neurotransmission in Alzheimer’s disease. *Front Aging Neurosci*. 2016; 8:31.  
<https://doi.org/10.3389/fnagi.2016.00031>  
PMID:[26941642](https://pubmed.ncbi.nlm.nih.gov/26941642/)
15. Govindpani K, Calvo-Flores Guzmán B, Vinnakota C, Waldvogel HJ, Faull RL, Kwakowsky A. Towards a better understanding of GABAergic remodeling in Alzheimer’s disease. *Int J Mol Sci*. 2017; 18:1813.  
<https://doi.org/10.3390/ijms18081813>  
PMID:[28825683](https://pubmed.ncbi.nlm.nih.gov/28825683/)
16. Athanasiu L, Giddaluru S, Fernandes C, Christoforou A, Reinvang I, Lundervold AJ, Nilsson LG, Kauppi K, Adolfsson R, Eriksson E, Sundet K, Djurovic S, Espeseth T, et al. A genetic association study of CSMD1 and CSMD2 with cognitive function. *Brain Behav Immun*. 2017; 61:209–16.  
<https://doi.org/10.1016/j.bbi.2016.11.026>  
PMID:[27890662](https://pubmed.ncbi.nlm.nih.gov/27890662/)
17. Uhl GR, Martinez MJ. PTPRD: neurobiology, genetics, and initial pharmacology of a pleiotropic contributor to brain phenotypes. *Ann N Y Acad Sci*. 2019; 1451:112–29.  
<https://doi.org/10.1111/nyas.14002>  
PMID:[30648269](https://pubmed.ncbi.nlm.nih.gov/30648269/)
18. Bading H. Nuclear calcium signalling in the regulation of brain function. *Nat Rev Neurosci*. 2013; 14:593–608.  
<https://doi.org/10.1038/nrn3531>  
PMID:[23942469](https://pubmed.ncbi.nlm.nih.gov/23942469/)
19. Tan Q, Heijmans BT, Hjelmborg JV, Soerensen M, Christensen K, Christiansen L. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int J Epidemiol*. 2016; 45:1146–58.  
<https://doi.org/10.1093/ije/dyw132> PMID:[27498152](https://pubmed.ncbi.nlm.nih.gov/27498152/)
20. Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P. Risk of dementia in diabetes mellitus: a systematic review. *Lancet Neurol*. 2006; 5:64–74.  
[https://doi.org/10.1016/S1474-4422\(05\)70284-2](https://doi.org/10.1016/S1474-4422(05)70284-2)  
PMID:[16361024](https://pubmed.ncbi.nlm.nih.gov/16361024/)
21. Saedi E, Gheini MR, Faiz F, Arami MA. Diabetes mellitus and cognitive impairments. *World J Diabetes*. 2016; 7:412–22.  
<https://doi.org/10.4239/wjd.v7.i17.412> PMID:[27660698](https://pubmed.ncbi.nlm.nih.gov/27660698/)
22. Caltagarone J, Jing Z, Bowser R. Focal adhesions regulate abeta signaling and cell death in Alzheimer’s disease. *Biochim Biophys Acta*. 2007; 1772:438–45.  
<https://doi.org/10.1016/j.bbadis.2006.11.007>  
PMID:[17215111](https://pubmed.ncbi.nlm.nih.gov/17215111/)
23. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet*. 2007; 3:e98.  
<https://doi.org/10.1371/journal.pgen.0030098>  
PMID:[17571925](https://pubmed.ncbi.nlm.nih.gov/17571925/)
24. Palomer E, Buechler J, Salinas PC. Wnt signaling deregulation in the aging and Alzheimer’s brain. *Front Cell Neurosci*. 2019; 13:227.  
<https://doi.org/10.3389/fncel.2019.00227>  
PMID:[31191253](https://pubmed.ncbi.nlm.nih.gov/31191253/)
25. Tapia-Rojas C, Inestrosa NC. Loss of canonical Wnt signaling is involved in the pathogenesis of Alzheimer’s disease. *Neural Regen Res*. 2018; 13:1705–10.  
<https://doi.org/10.4103/1673-5374.238606>  
PMID:[30136680](https://pubmed.ncbi.nlm.nih.gov/30136680/)
26. Kim EK, Choi EJ. Pathological roles of MAPK signaling pathways in human diseases. *Biochim Biophys Acta*. 2010; 1802:396–405.  
<https://doi.org/10.1016/j.bbadis.2009.12.009>  
PMID:[20079433](https://pubmed.ncbi.nlm.nih.gov/20079433/)
27. Lam RW, Kennedy SH, McIntyre RS, Khullar A. Cognitive dysfunction in major depressive disorder: effects on psychosocial functioning and implications for treatment. *Can J Psychiatry*. 2014; 59:649–54.

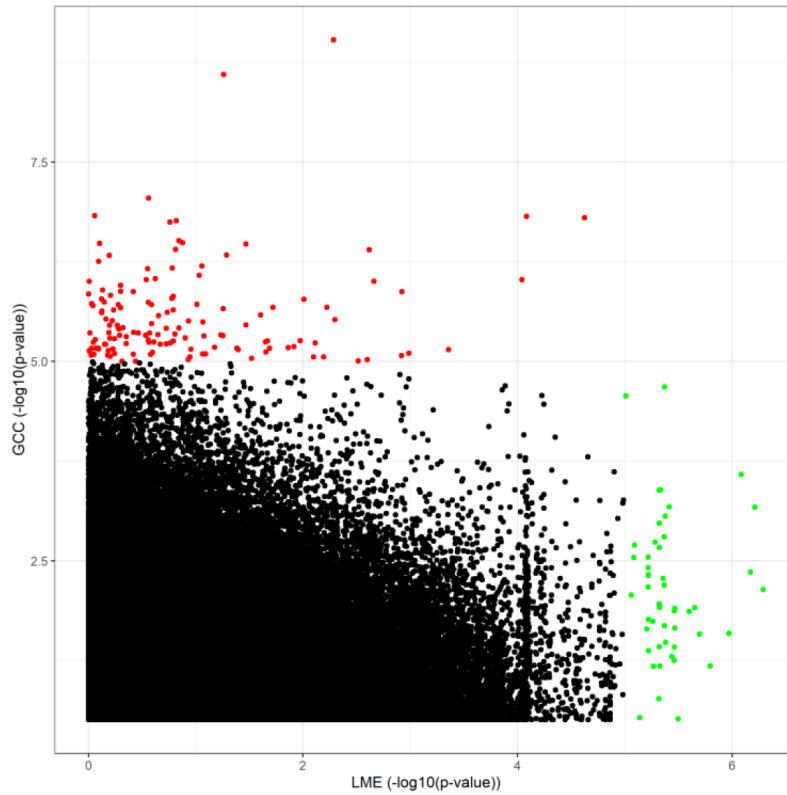
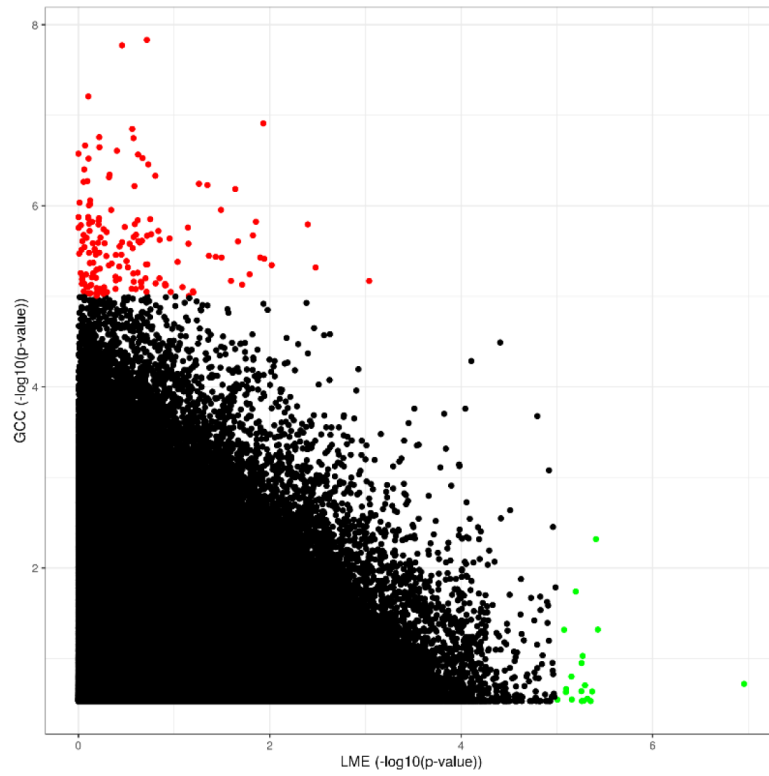
- <https://doi.org/10.1177/070674371405901206>  
PMID:[25702365](https://pubmed.ncbi.nlm.nih.gov/25702365/)
28. Sasai N, Toriyama M, Kondo T. Hedgehog signal and genetic disorders. *Front Genet.* 2019; 10:1103.  
<https://doi.org/10.3389/fgene.2019.01103>  
PMID:[31781166](https://pubmed.ncbi.nlm.nih.gov/31781166/)
29. Wang L, Chadwick W, Park SS, Zhou Y, Silver N, Martin B, Maudsley S. Gonadotropin-releasing hormone receptor system: modulatory role in aging and neurodegeneration. *CNS Neurol Disord Drug Targets.* 2010; 9:651–60.  
<https://doi.org/10.2174/187152710793361559>  
PMID:[20632963](https://pubmed.ncbi.nlm.nih.gov/20632963/)
30. Pillaiyar T, Namasivayam V, Manickam M, Jung SH. Inhibitors of melanogenesis: an updated review. *J Med Chem.* 2018; 61:7395–418.  
<https://doi.org/10.1021/acs.jmedchem.7b00967>  
PMID:[29763564](https://pubmed.ncbi.nlm.nih.gov/29763564/)
31. Marin I, Kipnis J. Learning and memory ... and the immune system. *Learn Mem.* 2013; 20:601–06.  
<https://doi.org/10.1101/lm.028357.112>  
PMID:[24051097](https://pubmed.ncbi.nlm.nih.gov/24051097/)
32. Pendergrass JC, Targum SD, Harrison JE. Cognitive impairment associated with cancer: a brief review. *Innov Clin Neurosci.* 2018; 15:36–44.  
PMID:[29497579](https://pubmed.ncbi.nlm.nih.gov/29497579/)
33. Shaoxun W, Feng J, Ya G, George B, Richard R, Fan F. Abstract TP556: Role of Vascular Smooth Muscle Cells in Diabetes-related Vascular Cognitive Impairment. *Stroke.* 2020 (Suppl 1); 50: ATP556.  
[https://doi.org/10.1161/str.50.suppl\\_1.TP556](https://doi.org/10.1161/str.50.suppl_1.TP556)
34. Ben Hamouda I, Tougourti MN, Hamza M. [Hypertrophic cardiomyopathy: a rare cause of vascular dementia. A case report]. *Tunis Med.* 2002; 80:420–23.  
PMID:[12611354](https://pubmed.ncbi.nlm.nih.gov/12611354/)
35. McGue M, Christensen K. The heritability of level and rate-of-change in cognitive functioning in Danish twins aged 70 years and older. *Exp Aging Res.* 2002; 28:435–51.  
<https://doi.org/10.1080/03610730290080416>  
PMID:[12227922](https://pubmed.ncbi.nlm.nih.gov/12227922/)
36. Petersen I, McGue M, Tan Q, Christensen K, Christiansen L. Change in depression symptomatology and cognitive function in twins: a 10-year follow-up study. *Twin Res Hum Genet.* 2016; 19:104–11.  
<https://doi.org/10.1017/thg.2016.3> PMID:[26879845](https://pubmed.ncbi.nlm.nih.gov/26879845/)
37. Xu C, Zhang D, Tian X, Wu Y, Pang Z, Li S, Tan Q. Genetic and environmental basis in phenotype correlation between physical function and cognition in aging Chinese twins. *Twin Res Hum Genet.* 2017; 20:60–65.  
<https://doi.org/10.1017/thg.2016.98> PMID:[28105965](https://pubmed.ncbi.nlm.nih.gov/28105965/)
38. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015; 4:7.  
<https://doi.org/10.1186/s13742-015-0047-8>  
PMID:[25722852](https://pubmed.ncbi.nlm.nih.gov/25722852/)
39. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, et al, and Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016; 48:1279–83.  
<https://doi.org/10.1038/ng.3643> PMID:[27548312](https://pubmed.ncbi.nlm.nih.gov/27548312/)
40. Therneau T. The lme4 function. Rochester, MN: Mayo Clinic. 2012. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.398.7032>
41. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. *J Stat Softw.* 2014; 67:1–48.  
<https://doi.org/10.18637/jss.v067.i01>
42. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017; 8:1826.  
<https://doi.org/10.1038/s41467-017-01261-5>  
PMID:[29184056](https://pubmed.ncbi.nlm.nih.gov/29184056/)
43. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015; 11:e1004219.  
<https://doi.org/10.1371/journal.pcbi.1004219>  
PMID:[25885710](https://pubmed.ncbi.nlm.nih.gov/25885710/)

## SUPPLEMENTARY MATERIALS

### Supplementary Figures

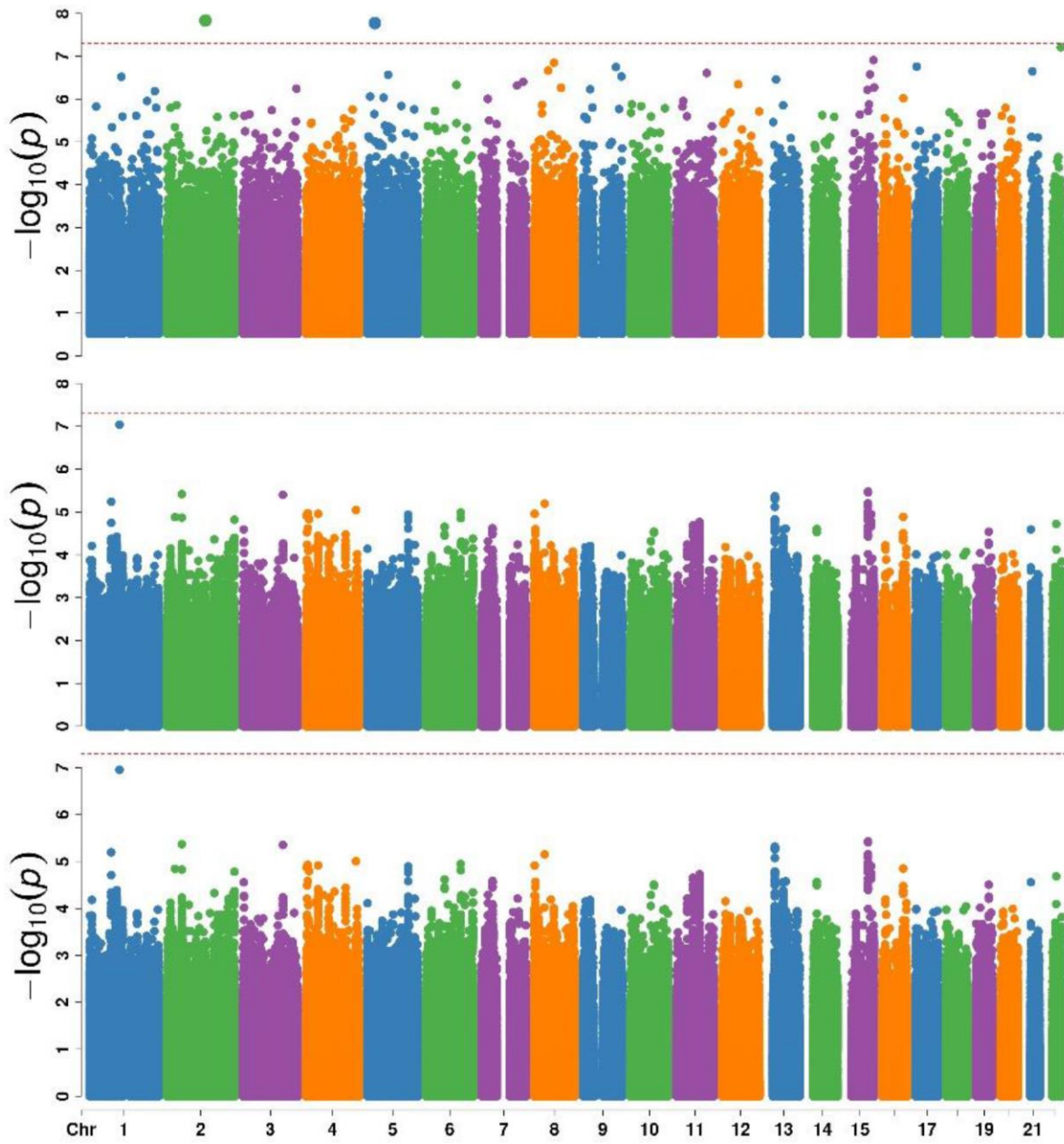


**Supplementary Figure 1. QQ Plots comparing the GCC model to the kinship and LME models in Danish (upper plot) and Chinese (lower plot) data.** The x-axis in both the left and right plots shows  $-\log_{10}$  of p-values from GWAS using the LME and Kinship models, respectively. The y-axis in both plots shows  $-\log_{10}$  of p-values from GWAS using the GCC model.

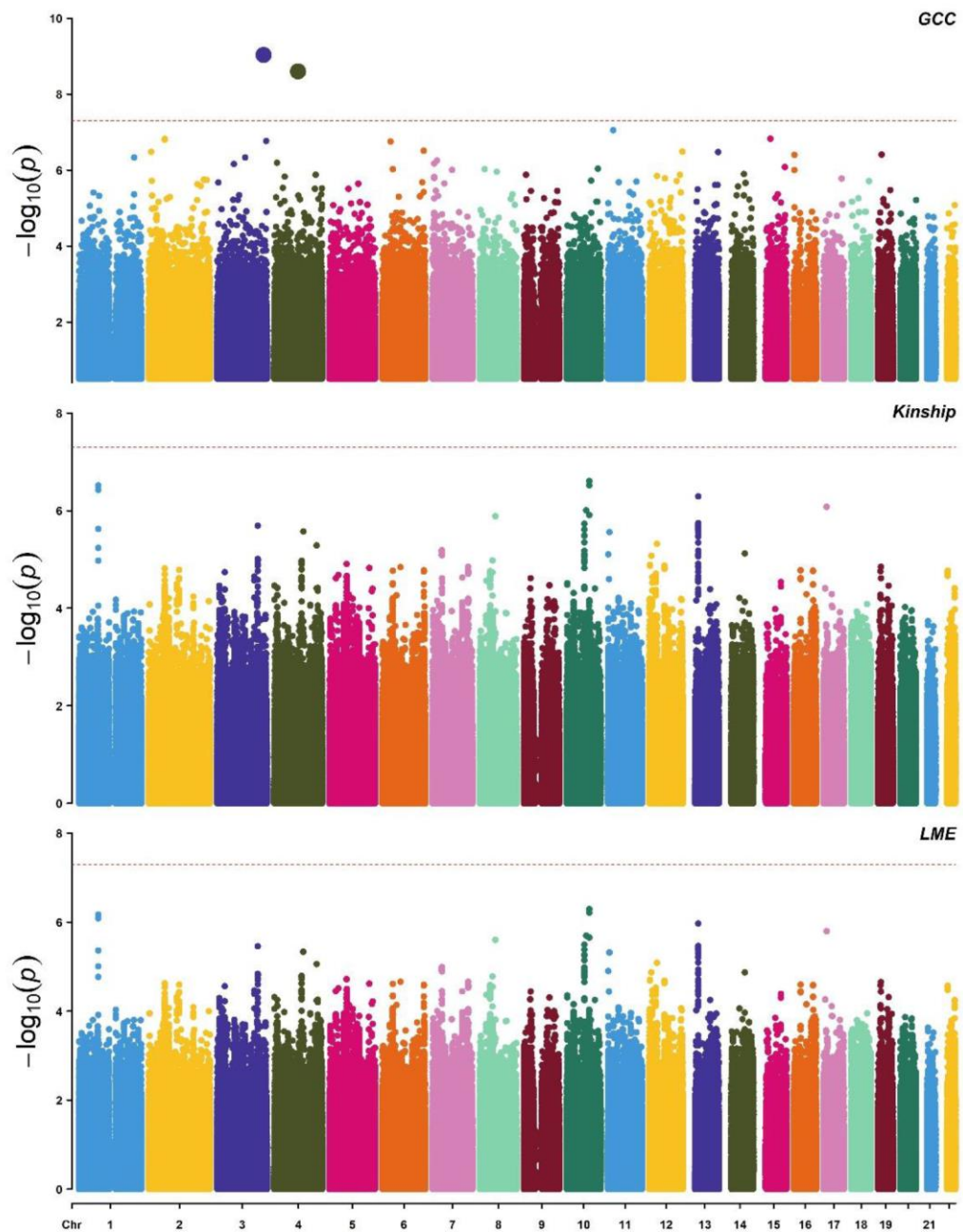


**Supplementary Figure 2. Scatter plot comparing SNP p-values from the GCC GWAS to SNP p-values from the LME model GWAS in both Danish (upper plot) and Chinese (lower plot) data.** The x-axis in both plots shows  $-\log_{10}$  of SNP p-values from the LME GWAS. The y-axis in both plots shows  $-\log_{10}$  SNP p-values from the GCC model GWAS.

**A**

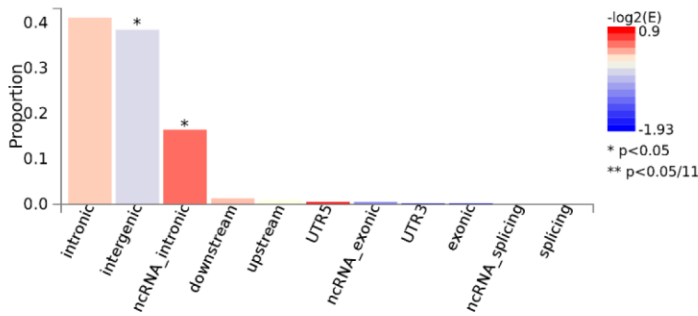




**B**

**Supplementary Figure 3.** Manhattan plots of the GCC, Kinship and LME GWAS results for the discovery data (A) and the replication data (B). The Manhattan plots in each plot from top to bottom are GCC GWAS, Kinship GWAS and LME GWAS, respectively.

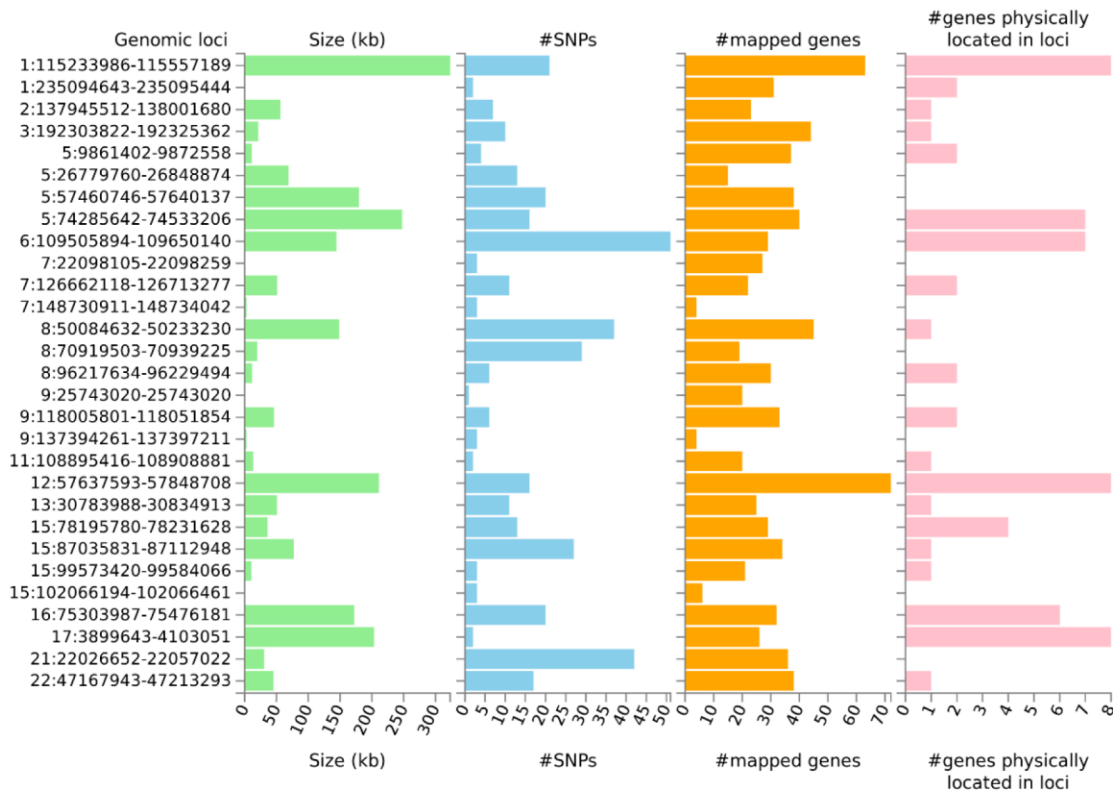
### A Functional consequences of SNPs on genes



### B Summary of SNPs and mapped genes

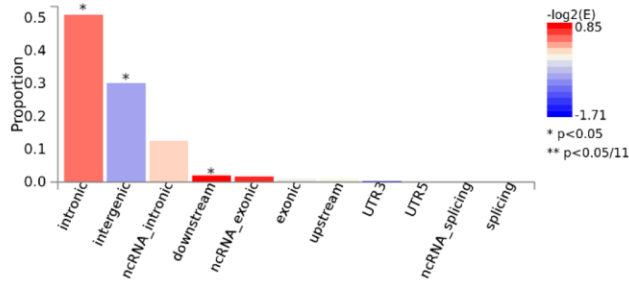
#Genomic risk loci	29
#Lead SNPs	29
#Ind. Sig. SNPs	29
#candidate SNPs	399
#candidate GWAS tagged SNPs	219
#mapped genes	862

### C Summary per genomic risk locus



**Supplementary Figure 4. Post-GWAS summary from FUMA of SNPs and mapped genes for the GCC GWAS in the Danish sample.** (A) The histogram shows the proportion of SNPs that are in LD with independent significant SNPs, which have corresponding functional annotation assigned by ANNOVAR. Bars are colored log<sub>2</sub> (enrichment) related to all SNPs in the reference panel. (B) The table shows summary of the number of SNPs and genes. (C) The histogram shows summary of results per genomic risk loci and genomic loci which could contain more than one independent lead SNPs.

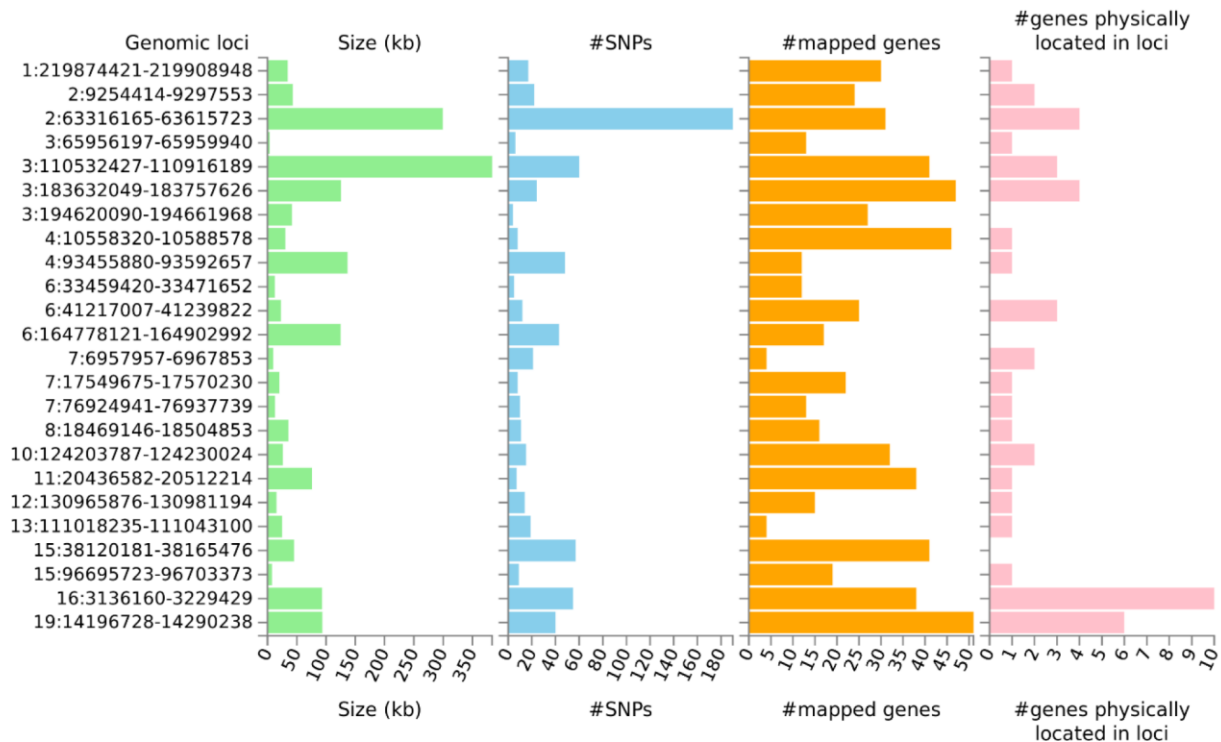
## A Functional consequences of SNPs on genes



## B summary of SNPs and mapped genes

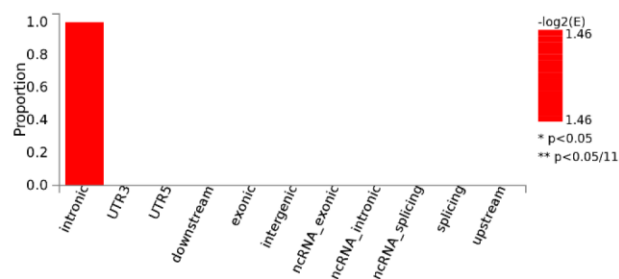
#Genomic risk loci	24
#Lead SNPs	24
#Ind. Sig. SNPs	24
#candidate SNPs	705
#candidate GWAS tagged SNPs	501
#mapped genes	618

## C Summary per genomic risk loci



**Supplementary Figure 5. Post-GWAS summary from FUMA of SNPs and mapped genes for the GCC GWAS in the Chinese sample.** (A) The histogram shows the proportion of SNPs that are in LD with independent significant SNPs which have corresponding functional annotation assigned by ANNOVAR. Bars are colored log<sub>2</sub> (enrichment) related to all SNPs in the reference panel. (B) The table shows summary of the number of SNPs and genes. (C) The histogram shows summary of results per genomic risk loci and genomic loci which could contain more than one independent lead SNPs.

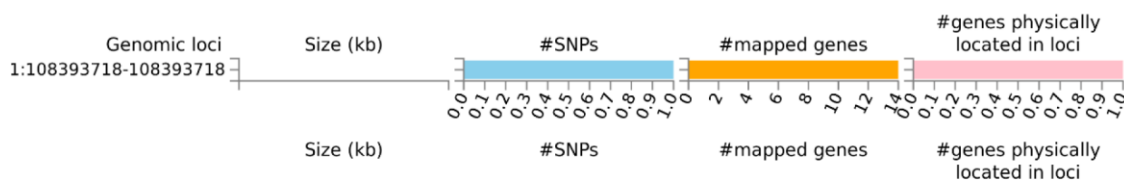
### A Functional consequences of SNPs on genes



### B Summary of SNPs and mapped genes

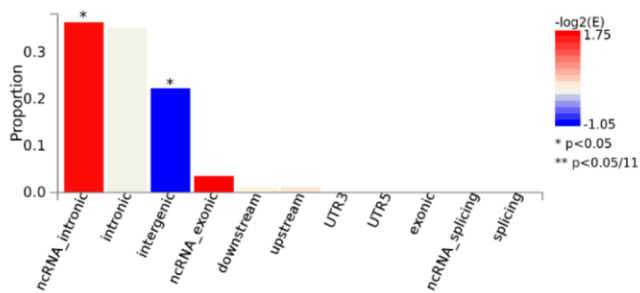
#Genomic risk loci	1
#Lead SNPs	1
#Ind. Sig. SNPs	1
#candidate SNPs	1
#candidate GWAS tagged SNPs	1
#mapped genes	14

### C Summary per genomic risk loci



**Supplementary Figure 6. Post-GWAS summary from FUMA of SNPs and mapped genes for the Kinship and LME GWAS in the Danish sample.** (A) The histogram shows the proportion of SNPs that are in LD with independent significant SNPs, which have corresponding functional annotation assigned by ANNOVAR. Bars are colored log<sub>2</sub>(enrichment) related to all SNPs in the reference panel. (B) The table shows summary of the number of SNPs and genes. (C) The histogram shows summary of results per genomic risk loci and genomic loci which could contain more than one independent lead SNPs.

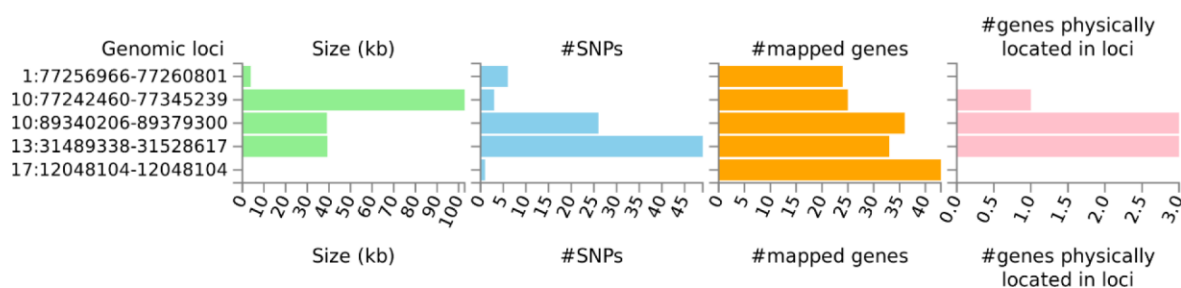
### A Functional consequences of SNPs on genes



### B Summary of SNPs and mapped genes

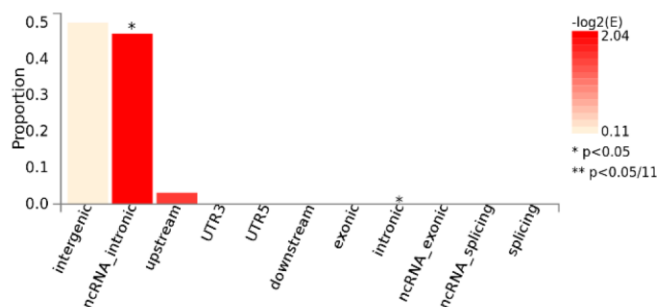
#Genomic risk loci	5
#Lead SNPs	5
#Ind. Sig. SNPs	5
#candidate SNPs	85
#candidate GWAS tagged SNPs	77
#mapped genes	161

### C Summary per genomic risk locus



**Supplementary Figure 7. Post-GWAS summary from FUMA of SNPs and mapped genes for the Kinship GWAS in the Chinese sample.** (A) The histogram shows the proportion of SNPs that are in LD with independent significant SNPs, which have corresponding functional annotation assigned by ANNOVAR. Bars are colored log<sub>2</sub> (enrichment) related to all SNPs in the reference panel. (B) The table shows summary of the number of SNPs and genes. (C) The histogram shows summary of results per genomic risk loci and genomic loci which could contain more than one independent lead SNPs.

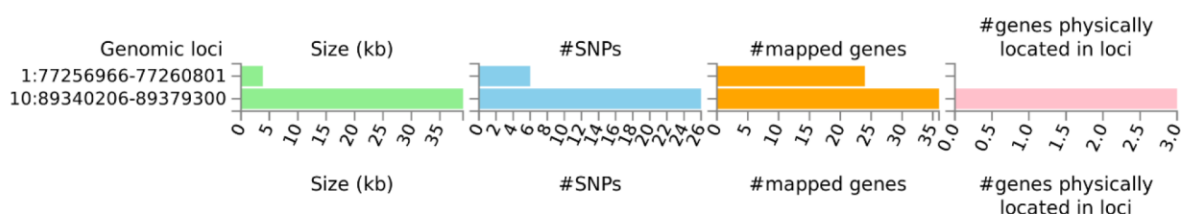
### A Functional consequences of SNPs on genes



### B Summary of SNPs and mapped genes

#Genomic risk loci	2
#Lead SNPs	2
#Ind. Sig. SNPs	2
#candidate SNPs	32
#candidate GWAS tagged SNPs	30
#mapped genes	60

### C Summary per genomic risk locus



**Supplementary Figure 8. Post-GWAS summary from FUMA of SNPs and mapped genes for the LME GWAS in the Danish sample.** (A) The histogram shows the proportion of SNPs that are in LD with independent significant SNPs, which have corresponding functional annotation assigned by ANNOVAR. Bars are colored log<sub>2</sub> (enrichment) related to all SNPs in the reference panel. (B) The table shows summary of the number of SNPs and genes. (C) The histogram shows summary of results per genomic risk loci and genomic loci which could contain more than one independent lead SNPs.

## Supplementary Tables

**Supplementary Table 1. Summary statistics for the top 30 SNPs identified in the GWAS of cognitive function in Danish samples using the GCC model. A and rawA are association scores from the matie R package.**

Chr:Position	SNP ID	A1	A2	MAF*	INFO*	A	rawA	P-value
2:137996342	rs71419535	C	T	0.248	0.994	0.0920	0.0397	1.47e-08
5:26779974	rs905838	G	A	0.401	0.967	0.0916	0.0394	1.69e-08
22:47168581	rs6008973	A	G	0.453	0.963	0.0879	0.0368	6.19e-08
15:99584066	rs8023294	G	A	0.068	0.912	0.0858	0.0354	1.23e-07
8:70935439	rs56065347	A	G	0.168	0.992	0.0854	0.0352	1.42e-07
17:4103051	rs62064368	T	A	0.051	0.900	0.0848	0.0347	1.74e-07
9:118044633	rs1100324	G	A	0.500	0.990	0.0847	0.0347	1.79e-07
8:50093277	rs1509086	T	C	0.473	0.999	0.0841	0.0343	2.16e-07
21:22035477	rs2826435	A	C	0.161	0.951	0.0840	0.0342	2.25e-07
11:108908881	rs11604258	A	G	0.132	0.993	0.0837	0.0340	2.46e-07
15:87071809	rs4516199	G	A	0.478	0.939	0.0835	0.0339	2.66e-07
5:74310301	rs115636606	G	A	0.077	0.992	0.0834	0.0338	2.71e-07
9:137397211	rs7859114	T	C	0.146	0.653	0.0831	0.0337	2.97e-07
1:115369112	rs6689326	T	A	0.460	0.996	0.0831	0.0336	3.01e-07
13:30798807	rs861727	G	C	0.333	0.940	0.0826	0.0333	3.49e-07
7:148734042	rs4483075	T	C	0.105	0.820	0.0822	0.0331	3.96e-07
12:57835965	rs2943693	C	T	0.222	0.972	0.0818	0.0328	4.53e-07
6:109522751	rs1271693	T	C	0.298	0.997	0.0817	0.0327	4.66e-07
7:126707558	rs3808123	T	C	0.419	0.998	0.0816	0.0326	4.84e-07
15:102066194	rs2084151	G	C	0.452	0.994	0.0813	0.0324	5.33e-07
8:96229494	rs74306623	G	A	0.162	0.916	0.0812	0.0324	5.43e-07
3:192325362	rs7610664	T	C	0.158	0.900	0.0810	0.0323	5.71e-07
9:25743020	rs10491876	G	A	0.129	0.647	0.0809	0.0322	5.90e-07
15:78231397	rs12901519	A	G	0.280	0.889	0.0808	0.0322	6.06e-07
1:235095444	rs2802929	A	G	0.358	0.679	0.0806	0.0320	6.53e-07
5:9872558	rs2963897	A	G	0.339	0.650	0.0797	0.0314	8.71e-07
5:57640137	rs245295	A	T	0.094	0.990	0.0795	0.0313	9.24e-07
16:75377514	rs9939080	A	G	0.20	0.749	0.0793	0.0312	9.60e-07
7:22098259	rs1833111	G	A	0.3550	0.749	0.0792	0.0312	9.88e-07
1:207421048	rs7535029	A	G	0.320	0.996	0.0789	0.0309	1.11e-06

\*MAF = Minor allele Frequency,

\*INFO = Information metric score from the imputation.

**Supplementary Table 2. Summary statistic for the top 30 SNPs from Kinship model in Danish GWAS.**

Chr:Position	SNPID	A1	A2	MAF*	INFO*	BETA	SE	Z	P-value
1:108393718	rs72707744	G	A	0.1012	0.7436	-3.6972	0.6919	-5.34	9.11e-08
15:80099089	rs34903151	T	C	0.105	0.922	3.0052	0.6464	4.65	3.34e-06
15:80099202	rs59367695	G	A	0.106	0.925	2.9949	0.6456	4.64	3.50e-06
2:53701324	rs903227	C	T	0.179	0.959	-2.3474	0.508	-4.62	3.82e-06
3:143762041	rs4839678	G	A	0.1656	0.7519	-2.6011	0.5639	-4.61	3.98e-06
13:26639396	rs7491901	A	G	0.4421	0.9979	1.8016	0.392	4.6	4.30e-06
13:26642629	rs7321426	C	T	0.4415	0.9973	1.7984	0.3923	4.58	4.57e-06
13:26641424	rs9512064	C	G	0.4416	0.9977	1.7949	0.3922	4.58	4.72e-06
13:26640931	rs9553716	T	A	0.4418	0.9976	1.7931	0.3922	4.57	4.84e-06
13:26634144	rs9512063	T	C	0.4416	0.9984	1.7918	0.3923	4.57	4.94e-06
13:26638368	rs1853720	C	A	0.4415	0.9987	1.7913	0.3923	4.57	4.97e-06
13:26636704	rs7317113	C	T	0.4415	0.9981	1.7908	0.3923	4.56	5.00e-06
1:78623626	rs17391694	C	T	0.1320	1	2.532	0.5582	4.54	5.74e-06
15:80119888	rs16971397	A	G	0.117	0.969	2.7288	0.6043	4.52	6.31e-06
8:37669354	rs10089977	T	C	0.2318	0.6758	2.4104	0.5341	4.51	6.39e-06
15:80103298	rs66588428	C	G	0.1103	0.941	2.8405	0.6332	4.49	7.26e-06
15:80100919	rs7183155	A	T	0.1098	0.932	2.8574	0.6372	4.48	7.31e-06
13:26637596	rs4770916	C	T	0.442	1	1.7614	0.3936	4.48	7.62e-06
4:178681212	rs10222887	G	T	0.3944	0.9902	1.6841	0.3793	4.44	8.99e-06
15:80151536	rs12912711	G	A	0.112	0.981	2.7315	0.6165	4.43	9.41e-06
15:80123923	rs9888692	G	A	0.113	0.992	2.7027	0.6108	4.43	9.64e-06
15:80122034	rs71409256	G	A	0.113	0.993	2.701	0.6108	4.42	9.80e-06
15:80143118	rs8033289	G	A	0.112	0.986	2.7199	0.6156	4.42	9.96e-06
15:80131092	rs7174937	G	A	0.113	0.990	2.7092	0.6134	4.42	1.00e-05
15:80141806	rs36060920	C	G	0.112	0.9868	2.7178	0.6155	4.42	1.01e-05
15:80140678	rs36107316	C	T	0.113	0.987	2.7176	0.6157	4.41	1.01e-05
6:124643640	rs2626123	G	C	0.2500	0.8012	-2.1333	0.4836	-4.41	1.03e-05
4:6196861	rs11729302	C	G	0.1158	0.6010	-3.3278	0.7559	-4.4	1.07e-05
15:80137560	exm1181778	T	C	0.113	1	2.6806	0.6094	4.4	1.09e-05
8:1886823	rs747689	C	T	0.3534	0.7651	1.9831	0.45	4.4	1-10e-05

\*MAF = Minor allele Frequency,

\*INFO = Information metric score from the imputation.



**Supplementary Table 3. Summary statistic for the top 30 SNPs from LME model in Danish GWAS.**

Chr:Position	SNP ID	A1	A2	MAF*	INFO*	Coef	tval	SE	P-value
1:108393718	rs72707744	G	A	0.1012	0.7436	-3.6972	-5.3437	0.6919	1.10e-07
15:80099089	rs34903151	T	C	0.105	0.922	3.0052	4.649	0.6464	3.73e-06
15:80099202	rs59367695	G	A	0.106	0.925	2.9949	4.6391	0.6456	3.91e-06
2:53701324	rs903227	C	T	0.179	0.959	-2.3474	-4.6206	0.508	4.27e-06
3:143762041	rs4839678	G	A	0.1656	0.7519	-2.6011	-4.6124	0.5639	4.43e-06
13:26639396	rs7491901	A	G	0.4421	0.9979	1.8016	4.5964	0.392	4.78e-06
13:26642629	rs7321426	C	T	0.4415	0.9973	1.7984	4.5836	0.3923	5.08e-06
13:26641424	rs9512064	C	G	0.4416	0.9977	1.7949	4.5767	0.3922	5.25e-06
13:26640931	rs9553716	T	A	0.4418	0.9976	1.7931	4.5716	0.3922	5.37e-06
13:26634144	rs9512063	T	C	0.4416	0.9984	1.7918	4.5675	0.3923	5.48e-06
13:26638368	rs1853720	C	A	0.4415	0.9987	1.7913	4.5659	0.3923	5.52e-06
13:26636704	rs7317113	C	T	0.4415	0.9981	1.7908	4.5649	0.3923	5.54e-06
1:78623626	rs17391694	C	T	0.1320	1	2.532	4.5357	0.5582	6.35e-06
15:80119888	rs16971397	A	G	0.117	0.969	2.7288	4.5158	0.6043	6.97e-06
8:37669354	rs10089977	T	C	0.2318	0.6758	2.4104	4.5131	0.5341	7.04e-06
15:80103298	rs66588428	C	G	0.1103	0.941	2.8405	4.4859	0.6332	8.00e-06
15:80100919	rs7183155	A	T	0.1098	0.932	2.8574	4.4845	0.6372	8.05e-06
13:26637596	rs4770916	C	T	0.442	1	1.7614	4.4756	0.3936	8.39e-06
4:178681212	rs10222887	C	T	0.3944	0.9902	1.6841	4.4401	0.3793	9.87e-06
15:80151536	rs12912711	G	A	0.112	0.981	2.7315	4.4303	0.6166	1.03e-05
15:80123923	rs9888692	G	A	0.113	0.992	2.7027	4.4251	0.6108	1.06e-05
15:80122034	rs71409256	G	A	0.113	0.993	2.7009	4.4216	0.6108	1.07e-05
15:80143118	rs8033289	G	A	0.112	0.986	2.7198	4.4179	0.6156	1.09e-05
15:80131092	rs7174937	G	A	0.113	0.990	2.7092	4.417	0.6134	1.10e-05
15:80141806	rs36060920	C	G	0.112	0.9868	2.7178	4.4159	0.6155	1.10e-05
15:80140678	rs36107316	C	T	0.113	0.987	2.7176	4.4142	0.6157	1.11e-05
6:124643640	rs2626123	G	C	0.2500	0.8012	-2.1333	-4.4114	0.4836	1.13e-05
4:6196861	rs11729302	C	G	0.1158	0.6010	-3.3278	-4.4021	0.756	1.17e-05
15:80137560	exm1181778	T	C	0.113	1	2.6806	4.3991	0.6094	1.19e-05
4:43193184	rs62299190	T	C	0.1139	0.8934	2.7868	4.3958	0.634	1.21e-05

\*MAF = Minor allele Frequency,

\*INFO = Information metric score.

**Supplementary Table 4. Summary statistic for the top 30 SNPs from GCC in Chinese GWAS. A and rawA are association scores from matie R package.**

Chr: Position	SNP ID	A1	A2	MAF*	INFO*	A	rawA	P-value
3:183681056	rs2292999	T	G	0.15108	1	0.2501	0.1697	9.27e-10
4:93505196	rs17019635	A	G	0.4479	0.99973	0.2436	0.1628	2.50e-09
11:20440691	rs12362967	A	G	0.2532	0.99503	0.2188	0.1373	8.95e-08
15:38153188	rs1502398	A	T	0.23565	0.99978	0.2151	0.1336	1.49e-07
2:63447200	rs13417792	G	T	0.05039	0.99921	0.2150	0.1335	1.51e-07
2:63439916	rs62177776	C	A	0.0543	0.91907	0.2146	0.1332	1.58e-07
3:194620090	rs61253693	G	A	0.23285	0.99638	0.2140	0.1326	1.71e-07
6:33459420	rs11757081	A	T	0.08714	0.98874	0.2138	0.1323	1.78e-07
6:164790026	rs10945984	C	T	0.45905	0.97115	0.2097	0.1282	3.07e-07
12:130968399	rs59068957	G	A	0.15953	0.94953	0.2093	0.1278	3.25e-07
2:9278695	rs7590864	T	A	0.10432	0.99966	0.2091	0.1276	3.31e-07
13:111036022	rs9515200	G	C	0.44155	0.9936	0.2090	0.1275	3.36e-07
19:14241153	rs8112885	A	G	0.10304	0.99419	0.2079	0.1264	3.91e-07
16:3181996	rs11649450	G	A	0.1781	0.90386	0.2078	0.1263	3.95e-07
3:110692754	rs1462795	T	C	0.46763	1	0.2066	0.1251	4.64e-07
1:219903142	rs4120823	C	T	0.31726	0.98683	0.2066	0.1251	4.65e-07
7:17554180	rs7808662	G	A	0.25336	0.9957	0.2052	0.1237	5.54e-07
4:10558320	rs17382947	A	G	0.09376	0.98675	0.2042	0.1227	6.38e-07
7:6964606	rs58404784	C	T	0.11034	0.76589	0.2037	0.1222	6.74e-07
3:65958585	rs17074200	A	T	0.266	0.99904	0.2036	0.1221	6.87e-07
15:96698614	rs727895	T	G	0.47377	0.99825	0.2021	0.1207	8.30e-07
10:124213046	rs61871747	C	T	0.45393	0.9945	0.2014	0.1199	9.13e-07
6:41222926	rs1872245	C	T	0.34691	0.99671	0.2012	0.1197	9.39e-07
8:18469146	rs77159425	A	G	0.11372	0.98384	0.2011	0.1197	9.43e-07
7:76927290	rs76326531	G	T	0.19772	0.98887	0.2008	0.1193	9.88e-07
16:3181963	rs1053871	T	C	0.1788	0.9017	0.2008	0.1193	9.91e-07
8:67171047	rs74193350	T	C	0.12023	0.9973	0.1999	0.1184	1.11e-06
14:69078998	rs7153168	C	T	0.16727	0.99998	0.1988	0.1174	1.27e-06
9:8011093	rs4740923	G	A	0.05204	0.91471	0.1985	0.1170	1.33e-06
4:164850409	rs55899995	A	G	0.05412	0.99699	0.1985	0.1170	1.33e-06

\*MAF = Minor allele Frequency,

\*INFO = Information metric score.

**Supplementary Table 5. Summary statistic for the top 30 SNPs from Kinship model in Chinese GWAS.**

Chr: Position	SNP ID	A1	A2	MAF*	INFO*	Beta	SE	z	P-value
10:89340206	rs10749553	A	G	0.37741	0.93624	-70.5629	13.6776	-5.16	2.48e-07
1:77259107	rs11577464	A	G	0.18385	0.9976	-87.3978	17.0651	-5.12	3.03e-07
10:89341315	rs10749555	A	C	0.37628	0.94323	-69.8189	13.6345	-5.12	3.04e-07
1:77260333	rs61783213	C	T	0.18491	0.96408	-86.9847	17.1274	-5.08	3.80e-07
13:31508724	rs1331848	A	G	0.44958	0.99976	64.3429	12.8125	5.02	5.12e-07
17:12048104	rs12944210	A	C	0.07151	0.98875	-126.6781	25.7148	-4.93	8.38e-07
10:77242460	rs117082223	G	A	0.0603	0.63757	169.5016	34.6389	4.89	9.91e-07
10:89340727	rs10887729	A	T	0.46505	0.93468	-66.7767	13.7705	-4.85	1.24e-06
8:61573242	rs56274310	C	G	0.15584	0.93985	87.7518	18.1382	4.84	1.31e-06
13:31513622	rs7319162	T	C	0.44242	0.9997	61.1840	12.8201	4.77	1.82e-06
13:31513145	rs4943681	G	A	0.44245	1	61.1749	12.8188	4.77	1.82e-06
13:31512404	rs2872194	A	G	0.44248	0.99981	61.1677	12.8205	4.77	1.83e-06
13:31512486	rs2313006	T	C	0.44256	0.99939	61.1571	12.8222	4.77	1.85e-06
10:70280398	rs78573780	A	T	0.31538	0.96562	-70.4836	14.7817	-4.77	1.86e-06
13:31509070	rs7321378	A	T	0.44332	0.9977	60.8599	12.7927	4.76	1.96e-06
3:161523263	rs1436738	G	A	0.06914	0.96809	117.1424	24.6759	4.75	2.06e-06
13:31509870	rs7984339	G	A	0.44357	0.99829	-60.9480	12.8439	-4.75	2.08e-06
13:31507727	rs1002036	G	T	0.43885	0.99999	-62.8348	13.2757	-4.73	2.21e-06
13:31528505	rs59815783	C	T	0.44042	0.96004	61.7178	13.0697	4.72	2.33e-06
13:31514975	rs61947599	A	G	0.44402	0.99701	60.7065	12.8565	4.72	2.34e-06
13:31513931	rs2225505	T	A	0.44419	0.99913	60.6492	12.8448	4.72	2.34e-06
13:31513425	rs7325107	A	G	0.44424	1	60.6235	12.8406	4.72	2.34e-06
1:77257353	rs4949747	A	T	0.19922	0.96797	-79.0279	16.7488	-4.72	2.38e-06
10:70365455	rs73276500	A	C	0.32545	0.95417	-66.5443	14.1210	-4.71	2.45e-06
13:31526533	rs12864826	T	C	0.42596	0.99859	61.1288	12.9912	4.71	2.53e-06
13:31522359	rs4943792	C	T	0.42601	0.9986	61.1191	12.9917	4.7	2.55e-06
13:31515563	rs28582187	T	A	0.42624	0.99983	61.0835	12.9876	4.7	2.56e-06
13:31509744	rs4941921	G	A	0.42626	0.99999	61.0794	12.9875	4.7	2.56e-06
13:31515301	rs61580577	G	A	0.42626	1	61.0794	12.9875	4.7	2.56e-06
13:31517404	rs12870193	G	A	0.42617	0.99911	61.0942	12.9908	4.7	2.57e-06

\*MAF = Minor allele Frequency,

\*INFO = Information metric score.

**Supplementary Table 6. Summary statistic for the top 30 SNPs from LME model in Chinese GWAS.**

Chr: Position	SNP ID	A1	A2	MAF*	INFO*	Coef	SE	t-value	P-value
10:89340206	rs10749553	A	G	0.37741	0.93624	-70.6351	13.7296	-5.1447	5.11e-07
10:89341315	rs10749555	A	C	0.37628	0.94323	-69.8916	13.6862	-5.1067	6.14e-07
1:77259107	rs11577464	A	G	0.18385	0.9976	-87.2301	17.1337	-5.0911	6.73e-07
1:77260333	rs61783213	C	T	0.18491	0.96408	-86.8156	17.1957	-5.0487	8.23e-07
13:31508724	rs1331848	A	G	0.44958	0.99976	64.2952	12.8663	4.9972	1.07e-06
17:12048104	rs12944210	A	C	0.07151	0.98875	-126.7369	25.8196	-4.9085	1.60e-06
10:77242460	rs117082223	G	A	0.0603	0.63757	169.3622	34.7934	4.8677	2.02e-06
10:89340727	rs10887729	A	T	0.46505	0.93468	-66.8073	13.8211	-4.8337	2.23e-06
8:61573242	rs56274310	C	G	0.15584	0.93985	87.5579	18.2082	4.8087	2.52e-06
10:70280398	rs78573780	A	T	0.31538	0.96562	-70.5498	14.8339	-4.7560	3.19e-06
13:31513622	rs7319162	T	C	0.44242	0.9997	61.1277	12.8743	4.7480	3.41e-06
13:31513145	rs4943681	G	A	0.44245	1	61.1187	12.8730	4.7478	3.41e-06
13:31512404	rs2872194	A	G	0.44248	0.99981	61.1114	12.8747	4.7466	3.43e-06
13:31512486	rs2313006	T	C	0.44256	0.99939	61.1009	12.8763	4.7452	3.45e-06
3:161523263	rs1436738	G	A	0.06914	0.96809	117.3295	24.7618	4.7383	3.46e-06
13:31509070	rs7321378	A	T	0.44332	0.9977	60.7993	12.8465	4.7328	3.65e-06
13:31509870	rs7984339	G	A	0.44357	0.99829	-60.8889	12.8979	-4.7208	3.85e-06
10:70365455	rs73276500	A	C	0.32545	0.95417	-66.5908	14.1752	-4.6977	4.18e-06
13:31507727	rs1002036	G	T	0.43885	0.99999	-62.6854	13.3314	-4.7021	4.19e-06
13:31528505	rs59815783	C	T	0.44042	0.96004	61.6639	13.1247	4.6983	4.27e-06
13:31514975	rs61947599	A	G	0.44402	0.99701	60.6452	12.9104	4.6974	4.28e-06
13:31513931	rs2225505	T	A	0.44419	0.99913	60.5882	12.8987	4.6972	4.28e-06
13:31513425	rs7325107	A	G	0.44424	1	60.5626	12.8944	4.6968	4.29e-06
1:77257353	rs4949747	A	T	0.19922	0.96797	-78.8652	16.8166	-4.6897	4.38e-06
4:115490688	rs112169253	A	C	0.05673	0.78317	-143.0388	30.5930	-4.6755	4.64e-06
13:31526533	rs12864826	T	C	0.42596	0.99859	61.0632	13.0485	4.6797	4.72e-06
13:31522359	rs4943792	C	T	0.42601	0.9986	61.0534	13.0491	4.6788	4.74e-06
13:31515563	rs28582187	T	A	0.42624	0.99983	61.0180	13.0449	4.6775	4.77e-06
13:31509744	rs4941921	G	A	0.42626	0.99999	61.0140	13.0449	4.6772	4.77e-06
13:31515301	rs61580577	G	A	0.42626	1	61.0140	13.0449	4.6772	4.77e-06

\*MAF = Minor allele Frequency,

\*INFO = Information metric score

**Supplementary Table 7. Top 20 genes from the MAGMA gene-based analysis of the Danish GCC GWAS results.**

<b>SYMBOL</b>	<b>CHR</b>	<b>START</b>	<b>STOP</b>	<b>NSNPS</b>	<b>NPARAM</b>	<b>N</b>	<b>ZSTAT</b>	<b>P-value</b>	<b>FDR</b>
CSMD1	8	2782875	4862494	12973	488	1172	4.8682	5.63e-07	0.011
PTPRD	9	8304246	10622723	7726	379	1172	4.148	1.68e-05	0.16
RBFOX1	16	6059095	7773340	7791	369	1172	4.0358	2.72e-05	0.17
CDH13	16	82650408	83840204	5249	284	1172	3.7244	9.79e-05	0.26
WVOX	16	78123310	79256564	5054	278	1172	3.5371	0.000202	0.26
CNTN4	3	2130497	3109645	3493	256	1172	3.4814	0.000249	0.26
FHIT	3	59725036	61247133	4957	279	1172	3.3964	0.000341	0.26
DPP6	7	153574182	154695995	3037	244	1172	3.3432	0.000414	0.26
TMEM132D	12	129546270	130398211	2565	223	1172	3.3348	0.000427	0.26
NPAS3	14	33394139	34283382	1963	214	1172	3.2302	0.000618	0.26
CDH4	20	59817482	60525673	2223	187	1172	3.2198	0.000641	0.26
MACROD2	20	13966015	16043842	4335	233	1172	3.1694	0.000764	0.26
OPCML	11	132274871	133412414	2616	209	1172	3.1024	0.00096	0.26
NTM	11	131230373	132216716	2400	186	1172	3.0947	0.000985	0.26
DLGAP1	18	3486030	4465335	2480	224	1172	3.0724	0.00106	0.26
KAZN	1	14915200	15454539	1511	143	1172	3.0629	0.0011	0.26
CTNNA2	2	79402357	80885905	3473	215	1172	3.03	0.00122	0.26
CNTNAP2	7	145803453	148128090	6261	209	1172	3.008	0.00131	0.26
SORCS2	4	7184265	7754554	2406	210	1172	3.0043	0.00133	0.26
ASIC2	17	31330105	32511983	3092	217	1172	2.9859	0.00141	0.26

**Supplementary Table 8. Top 20 genes from GCC model in Chinese data from MAGMA.**

<b>SYMBOL</b>	<b>CHR</b>	<b>START</b>	<b>STOP</b>	<b>NSNPS</b>	<b>NPARAM</b>	<b>N</b>	<b>ZSTAT</b>	<b>P-value</b>	<b>FDR</b>
CSMD1	8	2782875	4862494	12560	474	278	4.1834	1.44e-05	0.27
PTPRD	9	8304246	10622723	7351	367	278	3.778	7.91e-05	0.29
RBFOX1	16	6059095	7773340	7664	395	278	3.6857	0.000114	0.29
NPAS3	14	33394139	34283382	1986	220	278	3.6352	0.000139	0.29
WVOX	16	78123310	79256564	5118	317	278	3.3459	0.00041	0.29
CDH13	16	82650408	83840204	4127	267	278	3.2772	0.000524	0.29
CTNNA2	2	79402357	80885905	3820	267	278	3.1567	0.000798	0.29
KSR2	12	117880817	118416788	1435	158	278	3.1535	0.000807	0.29
DLGAP1	18	3486030	4465335	2422	226	278	3.1423	0.000838	0.29
DAB1	1	57450451	59022406	3461	211	278	3.142	0.000839	0.29
FRMD4A	10	13675706	14514141	2954	215	278	3.1174	0.000912	0.29
TMEM132D	12	129546270	130398211	2659	238	278	3.1167	0.000915	0.29
SORCS2	4	7184265	7754554	2084	206	278	3.1155	0.000918	0.29
MACROD2	20	13966015	16043842	4120	248	278	3.1149	0.00092	0.29
NELL1	11	20681117	21607227	2475	198	278	3.0328	0.00121	0.29
RYR2	1	237195505	238007288	2183	165	278	3.0075	0.00132	0.29
TMEM132C	12	128741948	129202460	1095	137	278	2.8947	0.0019	0.29
TEX26	13	31496840	31559639	77	7	278	2.8887	0.00193	0.29
LDLRAD4	18	13207497	13662754	985	114	278	2.8763	0.00201	0.29
MAGI2	7	77636393	79092890	4052	224	278	2.872	0.00204	0.29

**Supplementary Table 9. The 49 significant (FDR< 0.05) KEGG pathways identified in the GSEA overrepresentation analysis based on genes found to have a p-value <0.05 in the gene-based analysis of GCC GWAS results in the Danish discovery sample.**

Gene set names[# Genes (K)]	Description	#genes in overlap(K)	P-value	FDR q-value
KEGG_PATHWAYS_IN_CANCER [325]	Pathways in cancer	37	1.08 e <sup>-12</sup>	2.01 e <sup>-10</sup>
KEGG_AXON_GUIDANCE [129]	Axon guidance	21	1.16 e <sup>-10</sup>	1.08 e <sup>-8</sup>
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_AR_CARDIOMYOPATHY_ARVC [74]	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	15	2.27 e <sup>-9</sup>	1.41 e <sup>-7</sup>
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION [115]	Vascular smooth muscle contraction	18	4.67 e <sup>-9</sup>	2.17 e <sup>-7</sup>
KEGG_CALCIIUM_SIGNALING_PATHWAY [177]	Calcium signaling pathway	22	8.07 e <sup>-9</sup>	3 e <sup>-7</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	23	1.51 e <sup>-8</sup>	4.67 e <sup>-7</sup>
KEGG_MAPK_SIGNALING_PATHWAY [267]	MAPK signaling pathway	25	2.35 e <sup>-7</sup>	6.26 e <sup>-6</sup>
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM [83]	Hypertrophic cardiomyopathy (HCM)	13	6.33 e <sup>-7</sup>	1.47 e <sup>-5</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	17	1.46 e <sup>-6</sup>	3.02 e <sup>-5</sup>
KEGG_DILATED_CARDIOMYOPATHY [90]	Dilated cardiomyopathy	13	1.64 e <sup>-6</sup>	3.04 e <sup>-5</sup>
KEGG_LONG_TERM_DEPRESSION [70]	Long-term depression	11	4.43 e <sup>-6</sup>	7.5 e <sup>-5</sup>
KEGG_GNRH_SIGNALING_PATHWAY [101]	GnRH signaling pathway	13	6.09 e <sup>-6</sup>	9.44 e <sup>-5</sup>
KEGG_ADHERENS_JUNCTION [73]	Adherens junction	11	6.75 e <sup>-6</sup>	9.65 e <sup>-5</sup>
KEGG_O_GLYCAN_BIOSYNTHESIS [30]	O-Glycan biosynthesis	7	1.71 e <sup>-5</sup>	2.27 e <sup>-4</sup>
KEGG_BASAL_CELL_CARCINOMA [55]	Basal cell carcinoma	9	2.34 e <sup>-5</sup>	2.9 e <sup>-4</sup>
KEGG_HEDGEHOG_SIGNALING_PATHWAY [56]	Hedgehog signaling pathway	9	2.72 e <sup>-5</sup>	2.94 e <sup>-4</sup>
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION [116]	Leukocyte transendothelial migration	13	2.77 e <sup>-5</sup>	2.94 e <sup>-4</sup>
KEGG_LONG_TERM_POTENTIATION [70]	Long-term potentiation	10	2.84 e <sup>-5</sup>	2.94 e <sup>-4</sup>
KEGG_ERBB_SIGNALING_PATHWAY [87]	ErbB signaling pathway	11	3.7 e <sup>-5</sup>	3.62 e <sup>-4</sup>
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON [213]	Regulation of actin cytoskeleton	18	4.39 e <sup>-5</sup>	4.09 e <sup>-4</sup>
KEGG_GAP_JUNCTION [90]	Gap junction	11	5.08 e <sup>-5</sup>	4.5 e <sup>-4</sup>
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM [76]	Phosphatidylinositol signaling system	10	5.87 e <sup>-5</sup>	4.96 e <sup>-4</sup>
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY [79]	Fc epsilon RI signaling pathway	10	8.2 e <sup>-5</sup>	6.63 e <sup>-4</sup>
KEGG_CELL_ADHESION_MOLECULES_CAMS [133]	Cell adhesion molecules (CAMs)	13	1.16 e <sup>-4</sup>	8.96 e <sup>-4</sup>
KEGG_NON_SMALL_CELL_LUNG_CANCER [54]	Non-small cell lung cancer	8	1.38 e <sup>-4</sup>	9.89 e <sup>-4</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	10	1.38 e <sup>-4</sup>	9.89 e <sup>-4</sup>
KEGG_TYPE_II_DIABETES_MELLITUS [47]	Type II diabetes mellitus	7	3.47 e <sup>-4</sup>	2.39 e <sup>-3</sup>
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS [96]	Fc gamma R-mediated phagocytosis	10	4.16 e <sup>-4</sup>	2.76 e <sup>-3</sup>
KEGG_INSULIN_SIGNALING_PATHWAY [137]	Insulin signaling pathway	12	5.73 e <sup>-4</sup>	3.68 e <sup>-3</sup>
KEGG_MELANOGENESIS [101]	Melanogenesis	10	6.24 e <sup>-4</sup>	3.87 e <sup>-3</sup>
KEGG_SMALL_CELL_LUNG_CANCER [84]	Small cell lung cancer	9	6.5 e <sup>-4</sup>	3.9 e <sup>-3</sup>
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION [42]	Aldosterone-regulated sodium reabsorption	6	1.15 e <sup>-3</sup>	6.69 e <sup>-3</sup>
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY [75]	B cell receptor signaling pathway	8	1.32 e <sup>-3</sup>	7.44 e <sup>-3</sup>
KEGG_CARDIAC_MUSCLE_CONTRACTION [78]	Cardiac muscle contraction	8	1.7 e <sup>-3</sup>	9.31 e <sup>-3</sup>
KEGG_PURINE_METABOLISM [159]	Purine metabolism	12	2.09 e <sup>-3</sup>	1.08 e <sup>-2</sup>
KEGG_NOTCH_SIGNALING_PATHWAY [47]	Notch signaling pathway	6	2.09 e <sup>-3</sup>	1.08 e <sup>-2</sup>
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION [272]	Neuroactive ligand-receptor interaction	17	2.23 e <sup>-3</sup>	1.12 e <sup>-2</sup>
KEGG_GLIOMA [65]	Glioma	7	2.47 e <sup>-3</sup>	1.21 e <sup>-2</sup>
KEGG_GLYCEROLIPID_METABOLISM [49]	Glycerolipid metabolism	6	2.59 e <sup>-3</sup>	1.24 e <sup>-2</sup>
KEGG_CHEMOKINE_SIGNALING_PATHWAY [189]	Chemokine signaling pathway	13	3.14 e <sup>-3</sup>	1.44 e <sup>-2</sup>
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE [22]	Glycosaminoglycan biosynthesis - chondroitin sulfate	4	3.18 e <sup>-3</sup>	1.44 e <sup>-2</sup>
KEGG_PROSTATE_CANCER [89]	Prostate cancer	8	3.9 e <sup>-3</sup>	1.73 e <sup>-2</sup>
KEGG_VIBRIO_CHOLERAE_INFECTION [54]	Vibrio cholerae infection	6	4.25 e <sup>-3</sup>	1.84 e <sup>-2</sup>
KEGG_TIGHT_JUNCTION [132]	Tight junction	10	4.62 e <sup>-3</sup>	1.95 e <sup>-2</sup>
KEGG_ENDOCYTOSIS [181]	Endocytosis	12	5.93 e <sup>-3</sup>	2.42 e <sup>-2</sup>

KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY [137]	Natural killer cell mediated cytotoxicity	10	5.99 e <sup>-3</sup>	2.42 e <sup>-2</sup>
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM [77]	Glycerophospholipid metabolism	7	6.37 e <sup>-3</sup>	2.52 e <sup>-2</sup>
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_GANGLIO_SERIES [15]	Glycosphingolipid biosynthesis - ganglio series	3	8.12 e <sup>-3</sup>	3.15 e <sup>-2</sup>
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY [108]	T cell receptor signaling pathway	8	1.22 e <sup>-2</sup>	4.63 e <sup>-2</sup>



**Supplementary Table 10. The 10 KEGG pathways overrepresented by genes in Kinship model from discovery sample with FDR < 0.05.**

<b>Gene set names[# Genes (K)]</b>	<b>Description</b>	<b>#genes in overlap(K)</b>	<b>P-value</b>	<b>FDR q-value</b>
KEGG_AXON_GUIDANCE [129]	Axon guidance	13	3.41 e <sup>-5</sup>	6.35 e <sup>-3</sup>
KEGG_PPAR_SIGNALING_PATHWAY [69]	PPAR signaling pathway	9	7.43 e <sup>-5</sup>	6.91 e <sup>-3</sup>
KEGG_CHEMOKINE_SIGNALING_PATHWAY [189]	Chemokine signaling pathway	14	4.67 e <sup>-4</sup>	2.89 e <sup>-2</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	14	7.77 e <sup>-4</sup>	3.61 e <sup>-2</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	8	1.57 e <sup>-3</sup>	4.62 e <sup>-2</sup>
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM [34]	Cysteine and methionine metabolism	5	1.76 e <sup>-3</sup>	4.62 e <sup>-2</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	11	1.98 e <sup>-3</sup>	4.62 e <sup>-2</sup>
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY [108]	T cell receptor signaling pathway	9	2.09 e <sup>-3</sup>	4.62 e <sup>-2</sup>
KEGG_TIGHT_JUNCTION [132]	Tight junction	10	2.46 e <sup>-3</sup>	4.62 e <sup>-2</sup>
KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS [53]	Amyotrophic lateral sclerosis (ALS)	6	2.49 e <sup>-3</sup>	4.62 e <sup>-2</sup>

**Supplementary Table 11. The 12 KEGG pathways overrepresented by genes in LME model from discovery sample with FDR < 0.05.**

Gene set names[# Genes (K)]	Description	#genes in overlap(K)	P-value	FDR q-value
KEGG_AXON_GUIDANCE [129]	Axon guidance	13	3.01 e <sup>-5</sup>	5.6 e <sup>-3</sup>
KEGG_PPAR_SIGNALING_PATHWAY [69]	PPAR signaling pathway	9	6.77 e <sup>-5</sup>	6.29 e <sup>-3</sup>
KEGG_CHEMOKINE_SIGNALING_PATHWAY [189]	Chemokine signaling pathway	14	4.14 e <sup>-4</sup>	2.57 e <sup>-2</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	14	6.91 e <sup>-4</sup>	3.21 e <sup>-2</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	8	1.45 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM [34]	Cysteine and methionine metabolism	5	1.67 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	11	1.8 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY [108]	T cell receptor signaling pathway	9	1.92 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_TIGHT_JUNCTION [132]	Tight junction	10	2.25 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS [53]	Amyotrophic lateral sclerosis (ALS)	6	2.34 e <sup>-3</sup>	4.35 e <sup>-2</sup>
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION [116]	Leukocyte transendothelial migration	9	3.13 e <sup>-3</sup>	4.92 e <sup>-2</sup>
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY [75]	B cell receptor signaling pathway	7	3.18 e <sup>-3</sup>	4.92 e <sup>-2</sup>

**Supplementary Table 12. The 35 KEGG pathways overrepresented by genes from GCC in replication sample with FDR < 0.05.**

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	p-value	FDR q-value
KEGG_PATHWAYS_IN_CANCER [325]	Pathways in cancer	34	9.48 e <sup>-12</sup>	1.76 e <sup>-9</sup>
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_AR_CARDIOMYOPATHY_ARVC [74]	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	16	7 e <sup>-11</sup>	6.51 e <sup>-9</sup>
KEGG_AXON_GUIDANCE [129]	Axon guidance	19	1.26 e <sup>-9</sup>	7.79 e <sup>-8</sup>
KEGG_CALCIIUM_SIGNALING_PATHWAY [177]	Calcium signaling pathway	22	1.74 e <sup>-9</sup>	8.11 e <sup>-8</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	21	7.39 e <sup>-8</sup>	2.75 e <sup>-6</sup>
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION [115]	Vascular smooth muscle contraction	15	3.5 e <sup>-7</sup>	1.08 e <sup>-5</sup>
KEGG_DILATED_CARDIOMYOPATHY [90]	Dilated cardiomyopathy	13	6.39 e <sup>-7</sup>	1.7 e <sup>-5</sup>
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM [83]	Hypertrophic cardiomyopathy (HCM)	12	1.7 e <sup>-6</sup>	3.95 e <sup>-5</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	12	1.94 e <sup>-6</sup>	4 e <sup>-5</sup>
KEGG_TYPE_II_DIABETES_MELLITUS [47]	Type II diabetes mellitus	8	2.72 e <sup>-5</sup>	5.06 e <sup>-4</sup>
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION [272]	Neuroactive ligand-receptor interaction	20	3.78 e <sup>-5</sup>	6.39 e <sup>-4</sup>
KEGG_CELL_ADHESION_MOLECULES_CAMS [133]	Cell adhesion molecules (CAMs)	13	4.95 e <sup>-5</sup>	7.68 e <sup>-4</sup>
KEGG_GNRH_SIGNALING_PATHWAY [101]	GnRH signaling pathway	11	6.95 e <sup>-5</sup>	9.95 e <sup>-4</sup>
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SY_SYSTEM [76]	Phosphatidylinositol signaling system	9	1.65 e <sup>-4</sup>	2.19 e <sup>-3</sup>
KEGG_PURINE_METABOLISM [159]	Purine metabolism	13	3 e <sup>-4</sup>	3.72 e <sup>-3</sup>
KEGG_SMALL_CELL_LUNG_CANCER [84]	Small cell lung cancer	9	3.54 e <sup>-4</sup>	4.11 e <sup>-3</sup>
KEGG_LONG_TERM_DEPRESSION [70]	Long-term depression	8	4.82 e <sup>-4</sup>	5.07 e <sup>-3</sup>
KEGG_CHEMOKINE_SIGNALING_PATHWAY [189]	Chemokine signaling pathway	14	4.91 e <sup>-4</sup>	5.07 e <sup>-3</sup>
KEGG_HEDGEHOG_SIGNALING_PATHWAY [56]	Hedgehog signaling pathway	7	6.25 e <sup>-4</sup>	6.12 e <sup>-3</sup>
KEGG_MAPK_SIGNALING_PATHWAY [267]	MAPK signaling pathway	17	7.44 e <sup>-4</sup>	6.92 e <sup>-3</sup>
KEGG_CARDIAC_MUSCLE_CONTRACTION [78]	Cardiac muscle contraction	8	9.96 e <sup>-4</sup>	8.82 e <sup>-3</sup>
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON [213]	Regulation of actin cytoskeleton	14	1.56 e <sup>-3</sup>	1.32 e <sup>-2</sup>
KEGG_GLYCEROLIPID_METABOLISM [49]	Glycerolipid metabolism	6	1.7 e <sup>-3</sup>	1.37 e <sup>-2</sup>
KEGG_GAP_JUNCTION [90]	Gap junction	8	2.51 e <sup>-3</sup>	1.9 e <sup>-2</sup>
KEGG_TIGHT_JUNCTION [132]	Tight junction	10	2.55 e <sup>-3</sup>	1.9 e <sup>-2</sup>
KEGG_ADHERENS_JUNCTION [73]	Adherens junction	7	3 e <sup>-3</sup>	2.12 e <sup>-2</sup>
KEGG_BASAL_CELL_CARCINOMA [55]	Basal cell carcinoma	6	3.08 e <sup>-3</sup>	2.12 e <sup>-2</sup>
KEGG_NICOTINATE_AND_NICOTINAMIDE_METABOLISM [24]	Nicotinate and nicotinamide metabolism	4	3.26 e <sup>-3</sup>	2.17 e <sup>-2</sup>
KEGG_ALZHEIMERS_DISEASE [165]	Alzheimer's disease	11	4.28 e <sup>-3</sup>	2.75 e <sup>-2</sup>
KEGG_MELANOGENESIS [101]	Melanogenesis	8	5.09 e <sup>-3</sup>	3.16 e <sup>-2</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	10	6.3 e <sup>-3</sup>	3.78 e <sup>-2</sup>
KEGG_O_GLYCAN_BIOSYNTHESIS [30]	O-Glycan biosynthesis	4	7.44 e <sup>-3</sup>	4.3 e <sup>-2</sup>
KEGG_ERBB_SIGNALING_PATHWAY [87]	ErbB signaling pathway	7	7.87 e <sup>-3</sup>	4.3 e <sup>-2</sup>
KEGG_HEMATOPOIETIC_CELL_LINEAGE [87]	Hematopoietic cell lineage	7	7.87 e <sup>-3</sup>	4.3 e <sup>-2</sup>
KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY [67]	Adipocytokine signaling pathway	6	8.15 e <sup>-3</sup>	4.33 e <sup>-2</sup>

**Supplementary Table 13. The 16 KEGG pathways overrepresented by genes in Kinship model in replication sample with FDR < 0.05.**

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	p-value	FDR q-value
KEGG_PATHWAYS_IN_CANCER [325]	Pathways in cancer	28	7.2 e <sup>-8</sup>	1.34 e <sup>-5</sup>
KEGG_MAPK_SIGNALING_PATHWAY [267]	MAPK signaling pathway	23	1.04 e <sup>-6</sup>	9.63 e <sup>-5</sup>
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY [108]	T cell receptor signaling pathway	13	6.57 e <sup>-6</sup>	4.07 e <sup>-4</sup>
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_AR_CARDIOMYOPATHY_ARVC [74]	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	2.74 e <sup>-5</sup>	1.06 e <sup>-3</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	17	2.85 e <sup>-5</sup>	1.06 e <sup>-3</sup>
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY [75]	B cell receptor signaling pathway	9	1.75 e <sup>-4</sup>	5.43 e <sup>-3</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	13	2.09 e <sup>-4</sup>	5.56 e <sup>-3</sup>
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY [137]	Natural killer cell mediated cytotoxicity	12	3.28 e <sup>-4</sup>	7.63 e <sup>-3</sup>
KEGG_CELL_CYCLE [124]	Cell cycle	11	5.16 e <sup>-4</sup>	1.07 e <sup>-2</sup>
KEGG_MELANOMA [71]	Melanoma	8	6.12 e <sup>-4</sup>	1.14 e <sup>-2</sup>
KEGG_OLFACTORY_TRANSDUCTION [389]	Olfactory transduction	22	9.05 e <sup>-4</sup>	1.53 e <sup>-2</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	8	1.85 e <sup>-3</sup>	2.87 e <sup>-2</sup>
KEGG_PARKINSONS_DISEASE [128]	Parkinson's disease	10	2.38 e <sup>-3</sup>	3.41 e <sup>-2</sup>
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450 [70]	Metabolism of xenobiotics by cytochrome P450	7	2.66 e <sup>-3</sup>	3.54 e <sup>-2</sup>
KEGG_DRUG_METABOLISM_CYTOCHROME_P450 [72]	Drug metabolism - cytochrome P450	7	3.13 e <sup>-3</sup>	3.88 e <sup>-2</sup>
KEGG_ADHERENS_JUNCTION [73]	Adherens junction	7	3.38 e <sup>-3</sup>	3.93 e <sup>-2</sup>

**Supplementary Table 14. 12 KEGG pathways overrepresented by genes in LME model in replication sample with FDR <0.05.**

<b>Gene Set Name [# Genes (K)]</b>	<b>Description</b>	<b># Genes in Overlap (k)</b>	<b>p-value</b>	<b>FDR q-value</b>
KEGG_PATHWAYS_IN_CANCER [325]	Pathways in cancer	24	3.26 e <sup>-6</sup>	6.07 e <sup>-4</sup>
KEGG_MAPK_SIGNALING_PATHWAY [267]	MAPK signaling pathway	20	1.7 e <sup>-5</sup>	1.15 e <sup>-3</sup>
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY [108]	T cell receptor signaling pathway	12	1.86 e <sup>-5</sup>	1.15 e <sup>-3</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	16	4.98 e <sup>-5</sup>	2.32 e <sup>-3</sup>
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_AR_CARDIOMYOPATHY_ARVC [74]	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	9	1.01 e <sup>-4</sup>	3.75 e <sup>-3</sup>
KEGG_WNT_SIGNALING_PATHWAY [150]	Wnt signaling pathway	12	4.41 e <sup>-4</sup>	1.37 e <sup>-2</sup>
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY [75]	B cell receptor signaling pathway	8	6 e <sup>-4</sup>	1.6 e <sup>-2</sup>
KEGG_PARKINSONS_DISEASE [128]	Parkinson's disease	10	1.54 e <sup>-3</sup>	3.59 e <sup>-2</sup>
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450 [70]	Metabolism of xenobiotics by cytochrome P450	7	1.91 e <sup>-3</sup>	3.94 e <sup>-2</sup>
KEGG_DRUG_METABOLISM_CYTOCHROME_P450 [72]	Drug metabolism - cytochrome P450	7	2.25 e <sup>-3</sup>	3.95 e <sup>-2</sup>
KEGG_ADHERENS_JUNCTION [73]	Adherens junction	7	2.43 e <sup>-3</sup>	3.95 e <sup>-2</sup>
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY [137]	Natural killer cell mediated cytotoxicity	10	2.55 e <sup>-3</sup>	3.95 e <sup>-2</sup>

**Supplementary Table 15. The 28 KEGG pathways overrepresented by overlapped genes in GCC models in both studies with FDR < 0.05.**

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	p-value	FDR q-value
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_AR_CARDIOMYOPATHY_ARVC [74]	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	2.83 e <sup>-8</sup>	2.11 e <sup>-6</sup>
KEGG_PATHWAYS_IN_CANCER [325]	Pathways in cancer	19	3.4 e <sup>-8</sup>	2.11 e <sup>-6</sup>
KEGG_CALCIIUM_SIGNALING_PATHWAY [177]	Calcium signaling pathway	14	5.51 e <sup>-8</sup>	2.56 e <sup>-6</sup>
KEGG_AXON_GUIDANCE [129]	Axon guidance	12	8.33 e <sup>-8</sup>	3.1 e <sup>-6</sup>
KEGG_FOCAL_ADHESION [199]	Focal adhesion	13	1.46 e <sup>-6</sup>	4.53 e <sup>-5</sup>
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM [83]	Hypertrophic cardiomyopathy (HCM)	8	9.37 e <sup>-6</sup>	2.49 e <sup>-4</sup>
KEGG_DILATED_CARDIOMYOPATHY [90]	Dilated cardiomyopathy	8	1.71 e <sup>-5</sup>	3.97 e <sup>-4</sup>
KEGG_CELL_ADHESION_MOLECULES_CAMS [133]	Cell adhesion molecules (CAMs)	9	4.58 e <sup>-5</sup>	9.47 e <sup>-4</sup>
KEGG_ECM_RECEPTOR_INTERACTION [84]	ECM-receptor interaction	7	8.71 e <sup>-5</sup>	1.62 e <sup>-3</sup>
KEGG_LONG_TERM_DEPRESSION [70]	Long-term depression	6	2.39 e <sup>-4</sup>	3.71 e <sup>-3</sup>
KEGG_LONG_TERM_POTENTIATION [70]	Long-term potentiation	6	2.39 e <sup>-4</sup>	3.71 e <sup>-3</sup>
KEGG_GNRH_SIGNALING_PATHWAY [101]	GnRH signaling pathway	7	2.76 e <sup>-4</sup>	3.85 e <sup>-3</sup>
KEGG_TYPE_II_DIABETES_MELLITUS [47]	Type II diabetes mellitus	5	2.9 e <sup>-4</sup>	3.85 e <sup>-3</sup>
KEGG_BASAL_CELL_CARCINOMA [55]	Basal cell carcinoma	5	6.06 e <sup>-4</sup>	7.45 e <sup>-3</sup>
KEGG_SMALL_CELL_LUNG_CANCER [84]	Small cell lung cancer	6	6.41 e <sup>-4</sup>	7.45 e <sup>-3</sup>
KEGG_GAP_JUNCTION [90]	Gap junction	6	9.22 e <sup>-4</sup>	1.01 e <sup>-2</sup>
KEGG_TIGHT_JUNCTION [132]	Tight junction	7	1.36 e <sup>-3</sup>	1.41 e <sup>-2</sup>
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON [213]	Regulation of actin cytoskeleton	9	1.48 e <sup>-3</sup>	1.45 e <sup>-2</sup>
KEGG_ADHERENS_JUNCTION [73]	Adherens junction	5	2.18 e <sup>-3</sup>	2.03 e <sup>-2</sup>
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION [272]	Neuroactive ligand-receptor interaction	10	2.32 e <sup>-3</sup>	2.05 e <sup>-2</sup>
KEGG_PHOSPHATIDYLIINOSITOL_SIGNALING_SYSTEM [76]	Phosphatidylinositol signaling system	5	2.6 e <sup>-3</sup>	2.2 e <sup>-2</sup>
KEGG_NOTCH_SIGNALING_PATHWAY [47]	Notch signaling pathway	4	2.76 e <sup>-3</sup>	2.23 e <sup>-2</sup>
KEGG_CARDIAC_MUSCLE_CONTRACTION [78]	Cardiac muscle contraction	5	2.91 e <sup>-3</sup>	2.26 e <sup>-2</sup>
KEGG_ERBB_SIGNALING_PATHWAY [87]	ErbB signaling pathway	5	4.66 e <sup>-3</sup>	3.4 e <sup>-2</sup>
KEGG_ALZHEIMERS_DISEASE [165]	Alzheimer's disease	7	4.75 e <sup>-3</sup>	3.4 e <sup>-2</sup>
KEGG_HEDGEHOG_SIGNALING_PATHWAY [56]	Hedgehog signaling pathway	4	5.2 e <sup>-3</sup>	3.58 e <sup>-2</sup>
KEGG_MAPK_SIGNALING_PATHWAY [267]	MAPK signaling pathway	9	6.57 e <sup>-3</sup>	4.36 e <sup>-2</sup>