Identification of genes involved in the evolution of human intelligence through combination of inter-species and intra-species genetic variations

Mengjie Li¹, Wenting Zhang¹, Xiaoyi Zhou Correspond. ¹

¹ College of Life Sciences, Shanghai Normal University, Shanghai, China

Corresponding Author: Xiaoyi Zhou
Email address: zhouxy@shnu.edu.cn

Understanding the evolution of human intelligence is an important undertaking in the science of human genetics. A great deal of biological research has been conducted to search for genes which are related to the significant increase in human brain volume and cerebral cortex complexity during hominid evolution. However, genetic changes affecting intelligence in hominid evolution have remained elusive. We supposed that a subset of intelligence-related genes, which harbored intra-species variations in human populations, may also be evolution-related genes which harbored inter-species variations between humans (Homo sapiens) and great apes (including Pan troglodytes and Pongo abelii). Here we combined inter-species and intra-species genetic variations to discover genes involved in the evolution of human intelligence. Information was collected from published GWAS works on intelligence and a total of 549 genes located within the intelligence-associated loci were identified. The intelligence-related genes containing human-specific variations were detected based on the latest high-quality genome assemblies of three human’s closest species. Finally, we identified 40 strong candidates involved in human intelligence evolution. Expression analysis using RNA-Seq data revealed that most of the genes displayed a relatively high expression in the cerebral cortex. For these genes, there is a distinct expression pattern between humans and other species, especially in neocortex tissues. Our work provided a list of strong candidates for the evolution of human intelligence, and also implied that some intelligence-related genes may undergo inter-species evolution and contain intra-species variation.
Identification of genes involved in the evolution of human intelligence through combination of inter-species and intra-species genetic variations

Mengjie Li, Wenting Zhang, Xiaoyi Zhou*

College of Life Sciences, Shanghai Normal University, Shanghai, China

Corresponding Author:
Xiaoyi Zhou
No.100 Guilin Road, Shanghai, 200234, China
Email address: zhouxy@shnu.edu.cn

Abstract
Understanding the evolution of human intelligence is an important undertaking in the science of human genetics. A great deal of biological research has been conducted to search for genes which are related to the significant increase in human brain volume and cerebral cortex complexity during hominid evolution. However, genetic changes affecting intelligence in hominid evolution have remained elusive. We supposed that a subset of intelligence-related genes, which harbored intra-species variations in human populations, may also be evolution-related genes which harbored inter-species variations between humans (Homo sapiens) and great apes (including Pan troglodytes and Pongo abelii). Here we combined inter-species and intra-species genetic variations to discover genes involved in the evolution of human intelligence. Information was collected from published GWAS works on intelligence and a total of 549 genes located within the intelligence-associated loci were identified. The intelligence-related genes containing human-specific variations were detected based on the latest high-quality genome assemblies of three human’s closest species. Finally, we identified 40 strong candidates involved in human intelligence evolution. Expression analysis using RNA-Seq data revealed that most of the genes displayed a relatively high expression in the cerebral cortex. For these genes, there is a distinct expression pattern between humans and other species, especially in neocortex tissues. Our work provided a list of strong candidates for the evolution of human intelligence, and also implied that some intelligence-related genes may undergo inter-species evolution and contain intra-species variation.
Introduction

Humans have brains with significantly increased size and complexity compared to their ape counterparts (Rakic., 2009; Chenn et al., 2002; Lui et al., 2011). Corresponding alterations in intelligence have helped humans survive and create tools (Deary et al., 2012). Inspection of human genomic differences from our closest evolutionary relatives could help us to understand the intelligence-related genetic events during hominid evolution. The intelligence difference is thought to be derived from changes in genetics, owing to a small fraction of the 1% of sequence differences between the human genome and the chimpanzee genome, in which the human-specific gene insertions, deletions, and duplications played a critical role (Cheng et al., 2015).

Various approaches in molecular biology have been used to search for the human-specific genes and mutations therein that were related to the human intelligence. Several candidate genes involved in human intelligence evolution were identified based on human-ape comparative genomic analysis, gene expression profiling and other functional evidences. For example, in a very recent study, the information from gene expression profiling was integrated with the information from gene duplications in the ape and human lineages, which was then used to search for the human-specific genes that were highly expressed during human corticogenesis. In >35 candidates obtained through bioinformatics analysis, NOTCH2NL was functionally investigated – it was found that the gene was able to promote the expansion of cortical progenitors, serving as an important gene contributing to the evolution of the human brain (Fiddes et al., 2018; Suzuki et al., 2018). More recently, several human-specific genes have been identified and the critical genetic changes often occurred in gene regulation regions or resulted from the human-specific gene duplications; these include the NOTCH2NL gene, as well as FZD8, SRGAP2, ARHGAP11B, and TBC1D3, (Boyd et al., 2015; Dennis et al., 2012; Charrier et al., 2012; Florio et al., 2015; Ju et al., 2016).

As a highly heritable trait, intelligence has been intensively investigated using forward genetic approaches (Davies et al., 2015; Davies et al., 2016; Sniekers et al., 2017; Trampush et al., 2017; Zabaneh et al., 2018; Savage et al., 2018; Davies et al., 2018; Hill et al., 2016; Hill et al., 2019). Several genome-wide association studies (GWAS) and meta-analyses using very large human populations have been performed to identify causative genomic loci and genes underlying intelligence. Despite a significant enrichment in the nervous system, the functional links of the identified genes are diverse in which a wide variety of genes are involved (Davies et al., 2015; Davies et al., 2016; Sniekers et al., 2017; Trampush et al., 2017; Zabaneh et al., 2018; Savage et
This suggests that the evolution of human intelligence is a complicated process. While most causative genes may directly affect the central nervous system, genes from many related biological processes may be involved in the intelligence evolution as well.

Investigating the human-specific variations (genetic differences between humans and great apes) could provide key clues for understanding the process of the evolution of human intelligence. As the assembling gaps and errors in the previous reference genomes of the great apes (e.g., the human sequence guided assembling from short reads), the previous genomes were not qualified enough for the detection of complex structural variations (Prüfer et al., 2012; Scally et al., 2012; Prado-Martinez et al., 2013) such as tandem repeats, large-scale inversions, and duplications. These structural variations usually play important roles in human evolution (McLean et al., 2011). Hence, comparative genomic analysis from the complete genome sequences of both humans and great apes is needed to comprehensively identify the genetic variation. Recently, the high-quality genome sequences of three of human’s closest relatives, chimpanzee, orangutan, and gorilla, were generated from long-read sequencing (PacBio technology) and de novo assembling (Kronenberg et al., 2018; Gordon et al., 2016). The chromosome-level contiguous genome assemblies facilitate a deeper understanding of the genomic differences between these species.

Among a number of genomic differences between humans and apes, it is very important but technically difficult to know which variants are specific to intelligence. In order to further search out the candidate genes related to the evolution of human intelligence, we suppose that some intelligence genes may have both inter-species (between humans and the great apes) and intra-species (within humans) variations. For intra-species variations, we collected genomic loci identified by several sets of GWAS on human intelligence, and genes in these loci (termed as intelligence-associated-genes) could be regarded to contain intra-species intelligence differences. For inter-species variation, genomic differences between humans and the great apes, revealed by recent high-quality sequencing (Kronenberg et al., 2018; Gordon et al., 2016), were used and filtered. Hence, intelligence evolution was integrated by the overlap of intelligence-associated genes and human-specific variations. We found that many of the intelligence-associated genes, including tens of strong candidate genes related to the evolution of human intelligence, contained human-specific structural variations. Coupled with the expression profiling of the genes, this genome-wide analysis provided a useful resource for the evolutionary genetic studies on intelligence.
Materials & Methods

Preparation of bait genes

To exploit the genes related to human intelligence, six major works were collected from GWAS or GWAS meta-analyses on human intelligence with a large sample size from the last five years (various intelligence related phenotypes including general cognitive, reaction time and verbal-numerical reasoning). A total of 271 loci associated with intelligence in the human genome (Table S1) were identified in these six studies.

The six GWAS studies included: (i) meta-analyses of general cognitive function ($n=53,949$, Davies et al., 2015); (ii) GWAS of cognitive function and educational attainment ($n=112,151$, Davies et al., 2016); (iii) meta-analyses of calculated Spearman’s $g$ or a primary measure of fluid intelligence ($n=78,308$, Sniekers et al., 2017); (iv) meta-analysis and gene-based analysis of human cognition using 24 cohorts ($n=35,298$, Trampush et al., 2017); v) GWAS using human populations with extremely high intelligence ($n=1,238$, Zabaneh et al., 2018); (vi) a recent meta-analysis of 14 independent epidemiological cohorts with intelligence assessed ($n=269,867$, Savage et al., 2018).

All the independent, significantly associated SNPs (IndSigSNPs) nearest genes (based on ANNOVAR annotations) were integrated with redundancies (the same gene identified in more than one study) removed. We took these 549 genes as “bait genes” (Table S2), which were candidates for human intelligence-related genes.

It should be noted that some GWAS works underlying human intelligence that have been published very recently may be not included in this study. This would not affect our analyses because this study aimed to provide some candidate genes for human intelligence evolution and could not identify all related genes at one time. In addition, because genotype imputation was not performed for the X chromosome in some cohorts (e.g., the UKB cohort including 195,653 samples with the assessed phenotype verbal and mathematical reasoning), the potential genes related to intelligence in the X chromosome were not included in our “bait genes”.

Preparation of variation pond

The new genome assemblies of chimpanzees, gorillas and orangutans have become fully available recently through single-molecule, real-time (SMRT) long-read sequencing, which improved the resolution of large and complex genomic regions. We incorporated the human-specific structural variations (from intermediate size to large size) into a “variation pond”,...
including exon gain/loss, short tandem repeats (STRs), insertion/deletion (indels) of more than 50 bp, and inversions. Moreover, considering the important role of the human-specific segmental duplications (HSDs, >1 kb sequence with >90% similarity, indicating the large recent duplication events, Bailey et al., 2001) in new gene function and human evolution (Dennis et al., 2016), the HSDs identified recently from the comparative genomic information of both macaque and mouse (Dennis et al., 2017) were added into the “pond”. The “baits” and “ponds” were then integrated to detect whether there were human-specific variations hit by “bait genes”, thus generating the “prey genes” (Table S3-S5), which refer to the intelligence-related genes with human-specific variations. The human-specific variation that was located within each of the 549 genes was left for further analysis using a window of 1 Kb (the locations of the variations and genes on the chromosomes were divided by 1000 for the computation efficiency).

The genome sequences of three great apes were downloaded from the NCBI database. The reference genome sequences from Pan troglodytes (chimpanzee), Pongo abelii (Orangutan) and Gorilla (Western Lowland Gorilla) were downloaded from: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/880/755/GCF_002880755.1_Clint_PTRv2, ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/880/775/GCF_002880775.1_Susie_PABv2, and ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/905/GCF_000151905.2_gorGor4, respectively.

Localized alignments of the target gene sequences were performed to filter out the false positive human-specific variations from previous reports (McLean et al., 2011; Kronenberg et al., 2018; Dennis et al., 2017). The local sequences of the human genome were retrieved from chromosomes found in the GRCh38 version, and were then aligned with the great ape genomes using BLASTn (ncbi-blast-2.2.28+ version) with the parameters “-evalue 1E-50 -dust no”. The human-specific variation that was undetectable with a local BLAST was then removed for subsequent analyses.

**Investigation of HIEGs**

Most “prey” genes contained the human-specific variation in introns, far away from the exon-intron junction site. These variants are less likely to affect the gene functions, so only genes that contained variations in the coding regions were considered, named human intelligence evolution genes (HIEGs). These HIEGs included all the genes containing exon-gain/loss (2 genes), hCONDEL (28 genes), or HSD (1 gene), and genes with exon-located indels (8 genes) or STRs (4 genes). 40 non-redundant genes were finally identified, which were associated with human
intelligence and which carried significant human-specific variations.

The gene transcript information was obtained from Ensembl Release 95 (http://asia.ensembl.org/index.html). For exon alignment, all transcript isoforms of one gene were compared both intra-species (within human) and inter-species (between human and the great apes) in order to confirm the specificity of the new transcript in humans. The human principal isoform, the human variant isoform, the chimpanzee isoform, the gorilla isoform, and the orangutan isoform of the PCCB gene are ENST00000469217 (NM_001178014), ENST00000466072, ENSPRT0000028811, ENSGGOT0000005484 and ENSPPYT0000016431, respectively. The human principal isoform, the human variant isoform, the chimpanzee isoform, the gorilla isoform, and the orangutan isoform of the STAU1 gene are ENST00000371856, ENST00000340954, ENSPRT0000050938, ENSGGOT0000048652 and ENSPPYT0000012903, respectively.

The analysis for protein domain was performed using the Simple Modular Architecture Research Tool (SMART) in the normal mode (http://smart.embl-heidelberg.de). The protein accession numbers in humans, chimpanzees, gorillas, and orangutans for KMT2D are NP_003473, XP_016778992, XP_018894141, XP_024112209, respectively. The protein accession numbers in human, chimpanzee, gorilla, and orangutan for TRIOBP are NP_001034230, XP_016794633, XP_004063488, and H2P4B5 (UniProt), respectively. We performed multiple alignments using the software Constraint-based Multiple Alignment Tool (COBALT, ftp.ncbi.nlm.nih.gov/pub/agarwala/cobalt).

**Expression analysis for the candidates for intelligence evolution**

We assessed the expression patterns of the HIEGs using transcriptome data for humans and their closest relatives. The HPA RNA-seq data were downloaded from the Human Protein Atlas (http://www.proteinatlas.org), including 102 samples of 37 tissues, in which TPM (transcripts per million) were used for the evaluation of expression level. In order to compare the expression levels between humans and great apes, the RNA-Seq data (NCBI ID: GSE100796) from 107 samples in 8 brain regions of humans, chimpanzees, gorillas, and gibbons were used (Xu et al., 2018). The brain tissues included five neocortical areas and three other brain tissues. The five neocortical areas are the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (DPFC), the ventrolateral prefrontal cortex (VPFC), the premotor cortex (PMC), the primary visual cortex (V1C). The three other brain tissues are the hippocampus (HIP), the striatum, and
Hierarchical cluster analysis was applied using the RPKM (reads per kilobase per million) of 39 genes in 8 brain tissues of humans and the other three primates, and was displayed using the software MeV4.2 (http://www.tm4.org/mev.html).

**Results**

**Identification of candidate genes for human intelligence evolution**

The strategy that was used for detecting the candidate genes in the evolution of human intelligence is briefly described (Fig. 1). Hundreds of genetic loci for human intelligence and its related traits have been identified by GWAS using large population data. We integrated six high-quality GWAS works and large-scale GWAS meta-analyses from the last five years. This enabled the identification of a total of 271 loci in the human genome underlying intelligence-related phenotypes (Table S1). A total of 549 human genes, distributed across the whole human genome, were found to be located within the 271 intelligence-associated loci (Fig. 2, Table S2). We take the 549 genes as “bait genes”, which constituted intelligence-related genes from GWAS, underlying the intra-species intelligence-related variations related to intelligence in human populations.

In the meantime, the recently released long-read sequence assembly of chimpanzee, orangutan, and gorilla provided a large number of high-quality sequence differences between the human genome and the great ape genomes. We incorporated the human-specific structural variations, including exon gain and loss, STR, indels, hCONDEL, HSDs, and large structural variations, into a “variation pond” which are related to human evolution.

The “bait genes” and “variation pond” were then integrated to detect whether there were “bait genes” hit by the human-specific variation. After putting the “bait genes” into the “variation pond”, we found 406 sequence variations physically located within 213 genes related to intelligence, which is considered to be “prey genes” (Fig. 2). The 213 “prey genes” linked the inter-species and intra-species variations and may be related to the evolution of human intelligence. Additionally, the potential effects of the human-specific variation through local sequence comparisons and gene structure analyses were carefully checked. Following in-depth analyses, there were 40 strong candidate genes that were identified as containing human-specific variations, probably changing either the coding or the expression of intelligence related genes,
which were named as human intelligence evolution genes (HIEGs, Table 1).

Exon gain and loss on the intelligence related genes

There were only two genes classified as the “prey genes” with exon gain or loss events, PCCB and STAU1 (Table 1). The PCCB gene encoding the propionyl-CoA carboxylase subunit beta, was located in the locus on chromosome 3 identified by GWAS for intelligence ($P = 1.956 \times 10^{-8}$, Savage et al., 2018). Human PCCB contained 15 protein coding isoforms according to the Ensembl release 95. The principal isoforms generated proteins of 539 aa, 559 aa and 570 aa, respectively, all of which have identical isoforms in the chimpanzee. However, a new transcript variant was human-specific. Compared with the principal isoform, the variant lost a 60-bp exon (exon 4 of principal isoform) but gained another new 60-bp exon (exon 11 of the variational isoform). The gained 60-bp exon did not appear in any transcripts of the PCCB of the three great apes (Fig. 3A). This human-specific transcript variant can be detected in the human cerebral cortex at a lower level than that of the principal isoform (transcriptome data were obtained from the Human Protein Atlas, HPA). However, the relative cerebral cortex expression level of the PCCB variational transcript was higher than that of the principal transcript (relative to the average expression level of the transcript in 37 tissues) (Fig. 3B and 3C). These results suggested the generally low expression level and the relative enrichment of the PCCB variant in the cerebral cortex, implying a potential role of the human-specific PCCB transcript variant in the cerebral cortex. Although there were no reports about the mechanism of PCCB function in neuron system development, mutations in PCCB are one of the major causes of the genetic disease propionic academia (PA). Neurological complications, such as intellectual disability, brain structural abnormalities, optic neuropathy, and cranial nerve abnormalities are significant symptoms of PA (Schreiber et al., 2012). Moreover, there were reports that patients carrying PCCB mutations exhibited intellectual disabilities (Witters et al., 2016).

STAU1 was also located in a locus identified by GWAS for intelligence (Savage et al., 2018). STAU1 encoding the double-stranded RNA-binding protein which regulates RNA metabolism. There was a total of 10 protein-coding isoforms of human STAU1. Compared with the longest principal isoform, a human-specific insertion resulted in a gain of 123-bp exon (exon 2, located within the 5′UTR of the gene) in one transcript variant. The isoform, with the addition of a 123-bp exon, was not detected in any transcripts of chimpanzee, gorilla or orangutan (Fig. 3D). The human-specific isoform was expressed in many human tissues. In the human cerebral cortex, the
expression of the new isoform was equivalent to ~21% of that of the principal isoform of the
gene (Fig. 3E and 3F). Previous functional studies found that STAU1 plays a role on mRNA
transport in neuronal dendrites (Broadus et al., 1998).

STR variations on the intelligence related genes

In the variation pond, there were a total of 1,465 human-specific STR contractions and 4,921
human-specific STR expansions. These STRs hit 100 “bait genes” (26 genes containing the STR
contractions and 74 genes containing the STR expansions, Table S3). Most of the human-
specific STR variations were located within the intron regions or intergenic regions. Only four
human-specific STRs were located within the exonic regions, highlighted as HIEGs (Table1). Among
them, three STR expansions were located within the exon of non-protein coding
isoforms (processed transcripts) of three genes (ARIH2, STAB1 and TSNARE1), and one STR
contraction was located within the 39th exon of the KMT2D gene.

KMT2D, also known as MLL2, encoded an H3K4 histone methyltransferase made up of
5537 amino acids. GWAS for intelligence detected one associated locus ($P = 2.518 \times 10^{-14}$, Savage
et al., 2018) on chromosome 12 containing the candidate gene KMT2D. Compared with the
KMT2D sequence in the great apes, the STR region contracted for 60-bp in the c.11838 of human
KMT2D. This region was also polymorphic among the human population. The contraction led to
a 16-aa discontinuous deletion for the human KMT2D protein from the p.3958 position (Fig. 3G).
The peptide coding by this STR region was in the coiled coil region for KMT2D, which could
affect the protein structure through wrapping the hydrophobic residues and forming an
amphipathic surface (Mason et al., 2004). Mutations in KMT2D are the main cause of the genetic
disease Kabuki syndrome, and several mutations in the 39th exon have been found in Kabuki
patients. Kabuki syndrome affected mental capabilities and most of these patients showed
various levels of intellectual disability (Lehman et al., 2017).

Indels on the intelligence related genes

Among 5,894 human-specific deletions and 11,899 human-specific insertions in the “variation
pond”, we found 94 “prey genes” with 148 deletions and 144 genes with 298 insertions (Table
S4). Furthermore, it was found that there were 8 insertions and 1 deletion for the exonic regions,
as highlighted by the HIEGs (Table 1). The insertion in ARIH2 was also identified as STR
expansion which has been described above. Insertions in PDE4D, NRXN1, EXOC4, FUT8, and
ZNF584, and the deletion in SLC27A5 affect the lengths of the non-protein coding isoforms (processed transcripts) of the six genes. Moreover, GNB5 and TRIOBP carried an insertion in the 3’UTR region and an insertion in the exon region, respectively.

The insertion in TRIOBP resulted in a gain of 675-bp coding region (c.887-1560) in the 5th exon of the longest isoform when compared with that in the chimpanzee. This variation resulted in a 238-aa discontinuous insertion in the region p.296-811 of the human TRIOBP protein (Fig. 3H). The GWAS for intelligence ($P=3.582\times10^{-8}$, Savage et al., 2018) and the GWAS for underlying brain ventricular volume also identified the gene TRIOBP as a strong candidate (Vojinovic et al., 2018). Biochemistry experiments have shown that TRIOBP could physically interact with TRIO, which is an important protein involved in neural tissue development (Seipel et al., 2001). Mutations in TRIOBP could cause autosomal recessive deafness-28 (DFNB28), and surprisingly, several causal mutations were located within the human-specific insertion regions (e.g., R347X and Q297X) (Shahin et al., 2006).

We also searched for human conserved deletions (hCONDELs) near the “bait genes” that had been identified (McLean et al., 2011; Kronenberg et al., 2018). These sequences are lost in the human genome but are highly conserved among other species (including the great apes, the macaque, and the mouse). In total, 28 genes were identified as containing the hCONDELs. Although none of the 28 hCONDELs were located in the coding regions, they were all taken into HIEGs (Table 1), as the high lineage specificity of hCONDELs serving as important signs of intelligence evolution. The list included NRXN1, GRID2 and GRIA4, which were all involved in neurotransmission and the formation of synaptic contacts.

The GNB5 gene in the intelligence-associated locus on chromosome 15 ($P=2.47\times10^{-11}$, Savage et al., 2018), which encoded the beta subunit of the heterotrimeric GTP-binding proteins (G proteins). Aligned with the GNB5 sequences of the humans and the great apes, we found that there was a 292-bp human-specific insertion in the 3’UTR of the gene and a 1,472-bp hCONDEL in the intron (2.7 kb distance to the third exon) of the gene (Fig. 2). GNB5 was expressed in the brain tissues and participated in neurotransmitter signaling (e.g., through the dopamine D2 receptor) (Xie et al., 2012). In human populations, mutations in GNB5 have been reported to cause several diseases affecting intelligence, including language delay, ADHD/cognitive impairment with or without cardiac arrhythmia, and intellectual developmental disorder with cardiac arrhythmia (Lodder et al., 2016). The human-specific variation in GNB5 might participate in the evolution of human intelligence as well.
Large structural variations on the intelligence related genes

The inversion variation is the rearrangement in which the genomic segment is reversed. Based on the previous report, there were a total of 625 inversions in our “variation pond” ranging from 9 kb to 8.4 Mb in size. Among these, 31 of them hit the “bait genes” (Table S5). However, none of these genes were located in the breakpoint of the human-specific inversions.

There were 218 human-specific duplications (HSD) the length of which were more than 5-kb (Prado-Martinez et al., 2013). Among them, one 24.6-kb HSD was detected to be overlapped with the AFF3 gene. The AFF3 also contained an hCONDEL around the intron-exon junction regions (31-bp distance to the exon). The region around AFF3 has been identified to be an intelligence-associated locus in found in two distinct meta-analyses conducted on intelligence ($P = 1.56 \times 10^{-8}$, Sniekers et al., 2017 or $P = 3.41 \times 10^{-10}$, Savage et al., 2018), but no functional reports on neuron development are available.

Expression profiling analysis of the HIEGs

The transcriptome data of 37 human tissues from the Human Protein Atlas database (Uhlén et al, 2015) was used to investigate the tissue expression patterns of the 40 highlighted HIEGs (Fig. 4A). Of these there were 23 genes with higher expression levels in the cerebral cortex than their average expression levels in all 37 tissues. Furthermore, there were 12 genes with more than twofold expression levels in the cerebral cortex versus their average expression levels (Fig. 4A). These were regarded as genes that were potentially involved in the development of the cerebral cortex. Except FUT8, the remaining 11 genes are the hCONDEL-containing genes. The expression patterns of the 11 genes exhibited three types. Four genes, including GRIA4, NRXN1, CADM2, and CALN, showed the most significant high expressions in the cerebral cortex and low expressions in the other tissues. SGCZ, DCC, and GRID2 showed only relatively high expressions in the cerebral cortex compared with the low expressions in all other tissues. NCAM1, FBXL17, FUT8 and GNB5 were generally expressed in all tissues (Fig. 4A).

A transcriptome dataset sampling eight brain regions (five neocortical areas, hippocampus, striatum, and cerebellum) of both humans and four primate species (chimpanzees, gorillas, gibbon, and macaques) was reported previously (Xu et al., 2018), which enabled a comprehensive interspecies comparison. This dataset was used to examine whether the genes containing human-specific variation showed expression changes in the cortex between humans
and the great apes. The expression data for 39 of the 40 HIEGs, with the exception of *STAU1*, could be found in the transcriptome dataset. Among them, the expression levels of the genes *AFF3*, *SKAP1*, *REEP3*, *DCC*, and *SGCZ* in the human neocortical areas were much lower than those in the neocortical areas of the great apes (fold change <0.5), while the expressions of *STAB1* in the human neocortex were much higher than those in the great apes (fold change= 4.8).

Hierarchical clustering was also performed for the 39 gene expressions in all samples. We found that 5 neocortex tissues in the same species could be always clustered (that is, one clade for one species), while CB, STR, and HIP were generally clustered based on their tissue types (Fig4 B). This result suggested that the expression profiles of the HIEGs in the neocortex tissues displayed a strong species specificity, which was in contrast to the profiles in these non-neocortex tissues (e.g., CB). Taken together, the intelligence-related genes, containing the human-specific variations, expressed differently in the neocortex tissues between the humans and the great apes.

## Discussion

The understanding of human intelligence from the view of evolutionary genetics is an important scientific undertaking. Science magazine posted 125 scientific questions which were driving basic scientific research in various disciplines (Kennedy et al., 2005); among the top 25 questions, human evolution was addressed with the question: “What Genetic Changes Made Us Uniquely Human?” (Culotta, 2005). Obviously, genetic changes related to intelligence is one of the key steps in human evolution, making us uniquely *Homo sapiens*. However, it is still very challenging to identify the causative genes that are responsible for the intelligence difference between human and the other species in Hominidae. With both the human genome sequence and those of great apes becoming available, it is possible to pinpoint the genetic changes underlying the phenotypic differences between humans and the great apes. A subset of genes that were related to the human-ape intelligence differences may also contain intraspecific allelic variation underlying the variation of intelligence levels in human populations. Hence, in this work we integrated both the latest GWAS information on intelligence and the latest advances in the great ape genomics, aiming to mine the gene clues to understand the evolution of human intelligence.

We found several strong candidates, for example, the genes *TRIOBP* and *GNB5* contain human-specific variations and have the genetic evidence to be involved in the development of intelligence (Vojinovic et al. 2018; Xie et al., 2012; Lodder et al., 2016), although more in-depth molecular evidence and validations are needed in future experiments. Although our work did
find some candidate genes involved in human intelligence evolution, this strategy has its shortcomings. Some of intelligence-related genes with inter-species (between human and the great apes) variation may contain no intra-specific variation in human populations and cannot be identified by GWAS. Hence, the human genes involved in intelligence evolution but without intra-species variations in human populations were missed using our strategy.

We identified a total of 40 HIEGs, based on the location of the human-specific variations. Except for the hCONDELs as their high lineage specificity and importance in evolution, all the other human-specific variations in the HIEGs were located within the exons or putative regulatory regions. However, some of them were located within the exon of non-protein coding isoforms. There were five genes carried human-specific variations in exons or UTRs of protein coding isoforms. In this paper only five HIEGs were addressed. It was notable that two of the 5 genes were related to mRNA decay. One candidate gene with an exon gain, Staufen1 (STAU1), were reported to be involved in the transport, relocation, translation of mRNA and mRNA decay (Paul et al., 2018). Loss of STAU1 function in mice resulted in impaired mRNA transport and reduced synapse formation (Vessey et al., 2008). Another candidate gene, KMT2D harbored a human-specific STR contraction. Several truncating mutations within KMT2D resulting in mRNA degradation through nonsense-mediated mRNA decay, contributing to protein haploinsufficiency (Micale et al., 2014). It is still unclear whether there are any functional links between the two genes for mRNA processes in brain developments.

The associated loci from GWAS in humans often contain several candidate genes due to genetic mapping resolution, which is one of the difficulties in our bioinformatics analyses. To avoid artificial bias, all the candidate genes around the associated loci were included in the collection of “bait genes”, although usually for each associated locus only a single gene is causative. Consequently, a total of 549 human genes were included for the 271 intelligence-associated loci and “bait genes” contain many false positives. We cannot clearly distinguish the true one with the highly linked one, because other information (e.g., based on expression profiles or the distance to lead SNPs) is often misleading. As a result, HIEGs must contain many unrelated genes, although the intelligence related genes involved in human evolution have been partly enriched. Further experiments and analyses may include the validation of gene functions (whether and how these genes could influence the development of human brains) and the assessment of the effects of the human-specific variation (whether and how these sequence variations could influence the gene coding or the gene expression patterns).
There are already several findings of the human-chimpanzee differences altering the
development of the neocortex to date. The knowledge from the works of gene functional studies
and evolutionary genetics studies greatly enhanced our understanding of intelligence and the
brain (Boyd et al., 2015; Dennis et al., 2012; Charrier et al., 2012; Florio et al., 2015; Ju et al.,
2016). Certainly, the known genes (e.g., NOTCH2NL, FZD8, SRGAP2, ARHGAP11B, and
TBC1D3) are only a small proportion of the whole gene set that encapsulates the vast differences
in brain size and intelligence between great apes and humans, leaving many remaining gaps in
our knowledge. More integrated approaches incorporating genetics, genomics, bioinformatics,
and development biology will be needed in future works.

Conclusion

GWAS has identified hundreds of genes associated with intelligence variation in human
populations. Through inter-species genome comparisons between human and great apes, we
found that a small proportion of intelligence-related genes also contained human-specific
variation. Through integrated approaches, especially the careful checking of sequence
alignments and gene annotations, we identified 40 candidates in which human-specific variation
may have effects on gene coding or expressions. Transcriptome-wide comparison between
humans and four primate species for the 40 candidate genes suggests that several of them
displayed a different expression pattern among these species. The results implied that at least a
few of the intelligence-related genes may contain both intra-species variation and inter-species
variation. The intra-species variation underlies the small variation of intelligence levels among
different human individuals while the inter-species variation controlled the large genetic
differences of intelligence between the great apes and humans. This work may provide a list of
candidate genes to be used in subsequent studies.

Acknowledgments

We thank Professor Xuehui Huang from Shanghai Normal University for providing computer
servers.
References

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11:1005-1017.

Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, Wray GA, Silver DL. 2015. Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. Current Biology 25:772-779.

Broadus J, Fuerstenberg S, Doe CQ. 1998. Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. Nature 391:792-795.

Charrier C, Joshi K, Coutinho-Budd J, Kim JE, Lambert N, de Marchena J, Jin WL, Vanderhaegen P, Ghosh A, Sassa T, Polleux F. 2012. Inhibition of SRGAP2 Function by Its Human-Specific Paralogs Induces Neoteny during Spine Maturation. Cell 149:923-935.

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, Rocchi M, Eichler EE. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications Nature 437:88-93.

Chenn A, Walsh CA. 2002. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. Science 297:365-369.

Culotta E. 2005. What genetic changes made us uniquely human? Science 309: 91.

Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, Hofer E, Ibrahim-Verbaas CA, Kirin M, Lahti J, van der Lee SJ, Le Hellard S, Liu T, Marioni RE, Oldmeadow C, Postmus I, Smith AV, Smith JA, Thalamuthu A, Thomson R, Vitart V, Wang J, Yu L, Zgaga L, Zhao W, Boxall R, Harris SE, Hill WD, Liewald DC, Luciano M, Adams H, Ames D, Amin N, Amouyel P, Assareh AA, Au R, Becker JT, Beiser A, Berr C, Bertram L, Boerwinkle E, Buckley BM, Campbell H, Corley J, De Jager PL, Dufouil C, Eriksson JG, Espeseth T, Faul JD, Ford I; Generation Scotland, Gottesman RF, Griswold ME, Gudnason V, Harris TB, Heiss G, Hofman A, Holliday EG, Huffman J, Kardia SL, Kochan N, Knopman DS, Kwok JB, Lambert JC, Lee T, Li G, Li SC, Loitfelder M, Lopez OL, Lundervold AJ, Lundqvist A, Mather KA, Mirza SS, Nyberg L, Oostra BA, Palotti A, Papenberg G, Pattie A, Petrovic K, Polasek O, Psaty BM, Redmond P, Reppermund S, Rotter JI, Schmidt H, Schuur M, Schofield PW, Scott RJ, Steen VM, Stott DJ, van Swieten JC, Taylor KD, Trollor J, Trompet S, Uitterlinden AG, Weinstein G, Widen E, Windham...
BG, Jukema JW, Wright AF, Wright MJ, Yang Q, Amieva H, Attia JR, Bennett DA, Brodaty H, de Craen AJ, Hayward C, Ikram MA, Lindenberger U, Nilsson LG, Porteous DJ, Räikkönen K, Reinvang I, Rudan I, Sachdev PS, Schmidt R, Schofield PR, Srikanth V, Starr JM, Turner ST, Weir DR, Wilson JF, van Duijn C, Launer L, Fitzpatrick AL, Seshadri S, Mosley TH Jr, Deary IJ. 2015. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Molecular Psychiatry 20:183-192.

Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, Hagemanars SP, Ritchie SJ, Marioni RE, Fawns-Ritchie C, Liewald DCM, Okely JA, Ahola-Olli AV, Barnes CLK, Bertram L, Bis JC, Burdick KE, Christoforou A, DeRosse P, Djurovic S, Espey T, Giakoumaki S, Giddaluru S, Gustavson DE, Hayward C, Hofer E, Ikram MA, Karlsson R, Knowles E, Lahti J, Leber M, Li S, Mather KA, Melle I, Morris D, Oldmeadow C, Palviainen T, Payton A, Pazoki R, Petrovic K, Reynolds CA, Sargurupremraj M, Scholz M, Smith JA, Smith AV, Terzikhan N, Thalamuthu A, Trompet S, van der Lee SJ, Ware EB, Windham BG, Wright MJ, Yang J, Yu J, Ames D, Amin N, Amouyel P, Andreassen OA, Armstrong NJ, Assareh AA, Attia JR, Attix D, Avramopoulos D, Bennett DA, Böhmer AC, Boyle PA, Brodaty H, Campbell H, Cannon TD, Cirulli ET, Congdon E, Conley ED, Corley J, Cox SR, Dale AM, Dehghan A, Dick D, Dickinson D, Eriksson JG, Evangelou E, Faul JD, Ford I, Freimer NA, Gao H, Giegling I, Gillespie NA, Gordon SD, Gottesman RF, Griswold ME, Gudnason V, Harris TB, Hartmann AM, Hatzimanolis A, Heiss G, Holliday EG, Joshi PK, Kähönen M, Kardia SLR, Karlsson I, Kleineidam L, Knopman DS, Kochan NA, Konte B, Kwok JB, Le Hellard S, Lee T, Lehtimäki T, Li SC, Lill CM, Liu T, Koini M, London E, Longstreth WT Jr, Lopez OL, Loukola A, Luck T, Lundervold AJ, Lundquist A, Lyytikäinen LP, Martin NG, Montgomery GW, Murray AD, Need AC, Noordam R, Nyberg L, Ollier W, Papenberg G, Pattie A, Polasek O, Poldrack RA, Psaty BM, Reppermund S, Riedel-Heller SG, Rose RJ, Rotter JI, Roussos P, Rovio SP, Saba Y, Sabb FW, Sachdev PS, Satizabal CL, Schmid M, Scott RJ, Scult MA, Simino J, Slagboom PE, Smyrnis N, Soumaré A, Stefanis NC, Stott DJ, Straub RE, Sundet K, Taylor AM, Taylor KD, Tzoulaki I, Tzourio C, Uitterlinden A, Vitart V, Voineskos AN, Kaprio J, Wagner M, Wagner H, Weinhold L, Wen KH, Widen E, Yang Q, Zhao W, Adams HHH, Arking DE, Bilder RM, Bitsios P, Boerwinkle E, Chiba-Falek O, Corvin A, De Jager PL, Debette S, Donohoe G, Elliott P, Fitzpatrick AL, Gill M, Glahn DC, Hägg S, Hansell NK, Hariri AR,
Ikram MK, Jukema JW, Vuoksimaa E, Keller MC, Kremen WS, Launer L, Lindenberger U, Palotie A, Pedersen NL, Pendleton N, Porteous DJ, Räikkönen K, Raitakari OT, Ramirez A, Reinvang I, Rudan I, Dan Rujescu, Schmidt R, Schmidt H, Schofield PW, Schofield PR, Starr JM, Steen VM, Trollor JN, Turner ST, Van Duijn CM, Villringer A, Weinberger DR, Weir DR, Wilson JF, Malhotra A, McIntosh AM, Gale CR, Seshadri S, Mosley TH Jr, Bressler J, Lencz T, Deary IJ. 2018. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun 9:2098.

Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, Ritchie SJ, Luciano M, Fawns-Ritchie C, Lyall D, Cullen B, Cox SR, Hayward C, Porteous DJ, Evans J, McIntosh AM, Gallacher J, Craddock N, Pell JP, Smith DJ, Gale CR, Deary IJ. 2016. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112151). Mol Psychiatry 21:758-767.

Deary IJ. 2012. Intelligence. Annu Rev Psychol 63:453-482.

Dennis MY, Eichler EE. 2016. Human adaptation and evolution by segmental duplication. Curr Opin in Genet & Dev 41:44-52.

Dennis MY, Harshman L, Nelson BJ, Penn O, Cantsilieris S, Huddleston J, Antonacci F, Penewit K, Denman L, Raja A, Baker C, Mark K, Malig M, Janke N, Espinoza C, Stessman HAF, Nuttle X, Hoekzema K, Lindsay-Graves TA, Wilson RK, Eichler EE. 2017. The evolution and population diversity of human-specific segmental duplications. Nat Ecol Evol 1:69.

Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian S, Malig M, Kotkiewicz H, Curry CJ, Shafer S, Shaffer LG, de Jong PJ, Wilson RK, Eichler EE. 2012. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell 149:912-922.

Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, Lorig-Roach R, Field AR, Haeussler M, Russo L, Bhaduri A, Nowakowski TJ, Pollen AA, Dougherty ML, Nuttle X, Addor MC, Zwolinski S, Katzman S, Kriegstein A, Eichler EE, Salama SR, Jacobs FMJ, Haussler D. 2018. Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell 173:1356-1369.

Florio M, Albert M, Taverna E, Namba T, Brandl H, Lewitus E, Haffner C, Sykes A, Wong FK, Peters J, Guhr E, Klemroth S, Prüfer K, Kelso J, Naumann R, Nüsslein I, Dahl A,
Lachmann R, Pääbo S, Huttner WB. 2015. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. Science 347:1465-1470.

Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW, Dunn C, Baker C, Armstrong J, Dickhans M, Paten B, Shendure J, Wilson RK, Haussler D, Chin CS, Eichler EE. 2016. Long-read sequence assembly of the gorilla genome. Science 352: aae0344.

Hill WD, Davies G, Harris SE, Hagenaars SP; neuroCHARGE Cognitive Working group, Liewald DC, Penke L, Gale CR, Deary IJ. 2016. Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. Transl Psychiatry 6:e980.

Hill WD, Marioni RE, Maghzian O, Ritchie SJ, Hagenaars SP, McIntosh AM, Gale CR, Davies G, Deary IJ. 2019. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Mol Psychiatry 24:169-181.

Ju XC, Hou QQ, Sheng AL, Wu KY, Zhou Y, Jin Y, Wen T, Yang Z, Wang X, Luo ZG. 2016. The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. Elife 5: e18197

Kennedy D, Norman C. 2005. What don't we know? Science 309:75.

Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, Underwood JG, Nelson BJ, Chaisson MJP, Dougherty ML, Munson KM, Hastie AR, Dickhans M, Hormozdiari F, Lorusso N, Hoekzema K, Qiu R, Clark K, Raja A, Welch AE, Sorensen M, Baker C, Fulton RS, Armstrong J, Graves-Lindsay TA, Denli AM, Hoppe ER, Hsieh P, Hill CM, Pang AWC, Lee J, Lam ET, Dutcher SK, Gage FH, Warren WC, Shendure J, Haussler D, Schneider VA, Cao H, Ventura M, Wilson RK, Paten B, Pollen A, Eichler EE. 2018. High-resolution comparative analysis of great ape genomes. Science. 360

Lehman N, Mazery AC, Visier A, Baumann C, Lachesnais D, Capri Y, Toutain A, Odent S, Mikaty M, Goizet C, Taupiac E, Jacquemont ML, Sanchez E, Schaefer E, Gatinois V, Faivre L, Minot D, Kayirangwa H, Sang KLQ, Boddart N, Bayard S, Lacombe D, Moutton S, Toutou I., Rio M, Amiel J, Lyonnet S, Sanlaville D, Picot MC, Geneviève D. 2017. Molecular, clinical and neuropsychological study in 31 patients with Kabuki syndrome and KMT2D mutations. Clin Genet 92:298-305.

Lodder EM, De Nittis P, Koopman CD, Wiszniewski W, De Souza CFM, Lahrouchi N, Guex N, Napolioni V, Tessadori F, Beekman L, Nannenberg EA, Boualla L, Blom NA, de Graaff W, Kamermans M, Cocciadiferro D, Malerba N, Mandriani B, Akdemir ZHC, Fish
RJ, Eldomery MK, Ratbi I, Wilde AAM, de Boer T, Simonds WF, Neerman-Arbez M, Sutton VR, Kok F, Lupski JR, Reymond A, Bezzina CR, Bakkers J, Merla G. 2016. GNB5 Mutations Cause an Autosomal-Recessive Multisystem Syndrome with Sinus Bradycardia and Cognitive Disability. Am J Hum Genet 99: 786-786.

Lui JH, Hansen DV, Kriegstein AR. 2011. Development and Evolution of the Human Neocortex. Cell 146: 18-36.

Mason JM, Arndt KM. 2004. Coiled coil domains: stability, specificity, and biological implications. Chembiochem 5:170-176.

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, Wenger AM, Bejerano G, Kingsley DM. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471:216-219.

Micale L, Augello B, Maffeo C, Selicorni A, Zucchetti F, Fusco C, De Nittis P, Pellico MT, Mandriani B, Fischetto R, Boccone L, Silengo M, Biamino E, Perria C, Sotgiu S, Serra G, Lapi E, Neri M, Ferlini A, Cavaliere ML, Chiurazzi P, Monica MD, Scarano G, Faravelli F, Ferrari P, Mazzanti L, Pilotta A, Patricelli MG, Bedeschi MF, Benedictini F, Prontera P, Toschi B, Salviati L, Melis D, Di Battista E, Vancini A, Garavelli L, Zelante L, Merla G. 2014. Molecular Analysis, Pathogenic Mechanisms, and Readthrough Therapy on a Large Cohort of Kabuki Syndrome Patients. Human Mutat 35:841-850.

Paul S, Dansithong W, Figueroa KP, Scoles DR, Pulst SM. 2018. Staufen1 links RNA stress granules and autophagy in a model of neurodegeneration. Nat Commun 9:3648.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prüfer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprubí C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, Siegismund H, Blanc J, Agueda-Calpena L, Gut M, Fulton L, Tishkoff SA, Mullikin JC, Wilson RK, Gut IG, Gonder MK, Ryder OA, Hahn BH, Navarro A, Akey JM, Bertranpetit J, Reich D, Mailund T, Schierup MH, Hvilsom C, Andrés AM, Wall JD, Bustamante CD, Hammer MF, Eichler EE, Marques-Bonet T. 2013. Great ape genetic diversity and population history. Nature 499:471-475.
Prufer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, Knight JR, Mullikin JC, Meader SJ, Ponting CP, Lunter G, Higashino S, Hobolth A, Dutheil J, Karakoç E, Alkan C, Sajjadian S, Catacchio CR, Ventura M, Marques-Bonet T, Eichler EE, André C, Atencia R, Mugisha L, Junhold J, Patterson N, Siebauer M, Good JM, Fischer A, Ptak SE, Lachmann M, Symer DE, Mailund T, Schierup MH, Andrés AM, Kelso J, Pääbo S. 2012. The bonobo genome compared with the chimpanzee and human genomes. Nature 486:527-531.

Rakic P. 2009. Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci 10:724-735.

Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, Nagel M, Awasthi S, Barr PB, Coleman JRI, Grasby KL, Hammerschlag AR, Kaminski JA, Karlsson R, Krapohl E, Lam M, Nygaard M, Reynolds CA, Trampush JW, Young H, Zabaneh D, Hägg S, Hansell NK, Karlsson IK, Linnarsson S, Montgomery GW, Muñoz-Manchado AB, Quinlan EB, Schumann G, Skene NG, Webb BT, White T, Arking DE, Avramopoulos D, Bilder RM, Bitsios P, Burdick KE, Cannon TD, Chiba-Falek O, Christoforou A, Cirulli ET, Congdon E, Corvin A, Davies G, Deary IJ, DeRosse P, Dickinson D, Djurovic S, Donohoe G, Conley ED, Eriksson JG, Espeseth T, Freimer NA, Giakoumaki S, Giegling I, Gill M, Glahn DC, Hariri AR, Hatzimanolias A, Keller MC, Knowles E, Koltaí D, Konte B, Lahti J, Le Hellard S, Lencz T, Liewald DC, London E, Lundervold AJ, Malhotra AK, Melle I, Morris D, Need AC, Ollier W, Palotie A, Payton A, Pendleton N, Poldrack RA, Rääkkönen K, Reinvang I, Roussos P, Rujescu D, Sabb FW, Scult MA, Smeland OB, Smyrnis N, Starr JM, Steen VM, Stefanis NC, Straub RE, Sundet K, Tiemeier H, Voinikos AN, Weinberger DR, Widen E, Yu J, Abecasis G, Andreassen OA, Breen G, Christiansen L, Debrabant B, Dick DM, Heinz A, Hjerling-Leffler J, Ikram MA, Kendler KS, Martin NG, Medland SE, Pedersen NL, Plomin R, Polderman TJC, Ripke S, van der Sluis S, Sullivan PF, Vrieze SI, Wright MJ, Posthuma D 2018. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet 50:912-919.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA, Gu Y, Heath P, Heger A, Karakoc E, Kolb-Kokocinski A, Laird GK, Lunter G, Meader S, Mort M, Mullikin JC,
Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K, Simpson JT, Stenson PD, Turner DJ, Vigilant L, Vilella AJ, Whitener W, Zhu B, Cooper DN, de Jong P, Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA, Ryder OA, Searle SM, Warren WC, Wilson RK, Schierup MH, Rogers J, Tyler-Smith C, Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483: 169-175.

Schreiber J, Chapman KA, Summar ML, Ah Mew N, Sutton VR, MacLeod E, Stagni K, Ueda K, Franks J, Island E, Matern D, Peña L, Smith B, Urv T, Venditti C, Chakarapani A, Gropman AL. 2012. Neurologic considerations in propionic acidemia. Mol Genet Metab 105:10-15.

Seipel K, O'Brien SP, Iannotti E, Medley QG, Streuli M. 2001. Tara, a novel F-actin binding protein, associates with the Trio guanine nucleotide exchange factor and regulates actin cytoskeletal organization. J Cell Sci 114:389-399.

Shahin H, Walsh T, Sobe T, Abu Sa'ed J, Abu Rayan A, Lynch ED, Lee MK, Avraham KB, King MC, Kanaan M. 2006. Mutations in a novel isoform of TRIOBP that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss. Am J Hum Genet 78:144-152.

Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, Taskesen E, Hammerschlag AR, Okbay A, Zabaneh D, Amin N, Breen G, Cesari D, Chabris CF, Iacono WG, Ikram MA, Johannesson M, Koellinger P, Lee JJ, Magnusson PKE, McGue M, Miller MB, Ollier WER, Payton A, Pendleton N, Plomin R, Rietveld CA, Tiemeier H, van Duijn CM, Posthuma D. 2017. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet 49:1107-1112.

Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, Detours V, Vanderhaeghen P. 2018. Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell 173:1370-1384e1316.

Trampush JW, Yang MLZ, Yu J, Knowles E, Davies G, Liewald DC, Starr JM, Djurovic S, Melle I, Sundet K, Christoforou A, Reinvang I, DeRosse P, Lundervold AJ, Steen VM, Espeseth T, Räikkönen K, Widen E, Palotie A, Eriksson JG, Giegling I, Konte B, Roussos P, Giakoumaki S, Burdick KE, Payton A, Ollier W, Horan M, Chiba-Falek O, Attix DK, Need AC, Cirulli ET, Voineskos AN, Stefanis NC, Avramopoulos D, Hatzimanolis A, Arking DE,
Smyrnis N, Bilder RM, Freimer NA, Cannon TD, London E, Poldrack RA, Sabb FW, Congdon E, Conley ED, Scult MA, Dickinson D, Straub RE, Donohoe G, Morris D, Corvin A, Gill M, Hariri AR, Weinberger DR, Pendleton N, Bitsios P, Rujescu D, Lahtti J, Le Hellard S, Keller MC, Andreassen OA, Deary IJ, Glahn DC, Malhotra AK, Lencz T. 2017. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. Mol Psychiatry 22:1651-1652.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Osvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. 2015. Proteomics. Tissue-based map of the human proteome. Science 347:1260419.

Vessey JP, Macchi P, Stein JM, Mikl M, Hawker KN, Vogelsang P, Wieczorek K, Vendra G, Riefler J, Tübing F, Aparicio SA, Abel T, Kiebler MA. 2008. A loss of function allele for murine Staufen1 leads to impairment of dendritic Staufen1-RNP delivery and dendritic spine morphogenesis. Proc Natl Acad Sci USA 105:6374-16379.

Vojinovic D, Adams HH, Jian X, Yang Q, Smith AV, Bis JC, Teumer A, Scholz M, Armstrong NJ, Hofer E, Saba Y, Luciano M, Bernard M, Trompet S, Yang J, Gillespie NA, van der Lee SJ, Neumann A, Ahmad S, Andreassen OA, Ames D, Amin N, Arfanakis K, Bastin ME, Becker DM, Beiser AS, Beyer F, Brodaty H, Bryan RN, Bülow R, Dale AM, De Jager PL, Deary IJ, DeCarli C, Fleischman DA, Gottesman RF, van der Grond J, Gudnason V, Harris TB, Homuth G, Knoopman DS, Kwok JB, Lewis CE, Li S, Loeffler M, Lopez OL, Maillard P, El Marroun H, Mather KA, Mosley TH, Muetzel RL, Nauck M, Nyquist PA, Panizzon MS, Pausova Z, Psaty BM, Rice K, Rotter JI, Royle N, Satizabal CL, Schmidt R, Schofield PR, Schreiner PJ, Sidney S, Stott DJ, Thalamuthu A, Uitterlinden AG, Valdés Hernández MC, Vernooij MW, Wen W, White T, Witte AV, Wittfeld K, Wright MJ, Yanek LR, Tiemeier H, Kremen WS, Bennett DA, Jukema JW, Paus T, Wardlaw JM, Schmidt H, Sachdev PS, Villringer A, Grabe HJ, Longstreth WT, van Duijn CM, Launer LJ, Seshadri S, Ikram MA, Fornage M. 2018. Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. Nat Commun 9:3945.
Andersson H, Vannieuwenborg L, Geuens S, Baumgartner M, Kozicz T, Settles L, Morava E. 2016. Autism in patients with propionic acidemia. Mol Genet Metab 119:317-321.

Xie K, Ge S, Collins VE, Haynes CL, Renner KJ, Meisel RL, Lujan R, Martemyanov KA. 2012. Gβ5-RGS complexes are gatekeepers of hyperactivity involved in control of multiple neurotransmitter systems. Psychopharmacology (Berl) 219:823-834.

Xu C, Li Q, Efimova O, He L, Tatsumoto S, Stepanova V, Oishi T, Udono T, Yamaguchi K, Shigenobu S, Kakita A, Nawa H, Khaitovich P, Go Y. 2018. Human-specific features of spatial gene expression and regulation in eight brain regions. Genome Res 28:1097-1110.

Zabaneh D, Krapohl E, Gaspar HA, Curtis C, Lee SH, Patel H, Newhouse S, Wu HM, Simpson MA, Putallaz M, Lubinski D, Plomin R, Breen G. 2018. A genome-wide association study for extremely high intelligence. Mol Psychiatry 23:1226-1232.
Figure 1

Strategy for searching for genes in human intelligence evolution

Grey areas indicated results from the previous studies including six GWAS works on intelligence and the comparative genomics analysis between humans and great apes.
Manuscript to be reviewed

- 271 intelligence associated loci revealed by GWAS
- Gene baits
  - 549 Intelligence related Genes
- Comparative Analysis
- Human Genome
- Great Ape Genome
  - Chimpanzee
  - Gorilla
  - Orangutan
- Variation ponds
  - Exon gain/lost
  - STR
  - SV (>50bp)
  - Inversion
  - Duplication
  - Consider macaque and mouse genome
  - hCONDEL
- Human specific variations in Intelligence related Genes
  - Exon gain/lost (2 genes)
  - STR (100 genes)
  - SV (179 genes)
  - Inversion (31 genes)
  - Duplication (1 gene)
  - hCONDEL (28 genes)
- In Exon (4 genes)
- In Exon (8 genes)
- 40 highlighted intelligence related genes and containing human-specific variations (Table1)
Figure 2

The distribution of intelligence associated genes and their located human specific variations in the 22 autosomes of the human genome

The intelligence associated genes are indicated on the left side of the chromosome bars. The human specific variations around the intelligence associated loci are indicated by lines of colors on the right side of chromosome bars. The centromere regions are indicated by red boxes. STR-c: STR contraction; STR-e: STR expansion; SV-d: deletions; SV-i: insertions; HSD: human segmental duplications; and hCONDEL: human conserved deletions
Figure 3

Four candidate genes involved in human intelligence evolution

(A) Gene structures of PCCB in human and great ape genomes. The orange box represents the gained exon (also indicated by red arrow). The mutations in patients with intellectual disability are labeled. The solid boxes represent coding exons, and the hollow box represent UTRs; (B) TPM of PCCB principal isoform and variant isoform in cerebral cortex; (C) TPM of PCCB principal isoform and variant isoform in cerebral cortex relative to 37 human tissues; (D) Gene structures of STAU1; (E) TPM of STAU1 principal isoform and variant isoform in cerebral cortex; (F) TPM of STAU1 principal isoform and variant isoform in cerebral cortex relative to 37 human tissues; (G) The schematic representation of KMT2D protein; (H) The schematic representation of TRIOBP protein. The protein sequence alignments of regions with human-specific variation are showed. The numbers indicated the position of amino acids, and the numbers in square brackets between sequences indicated the hided amino acids.
Figure 4

Expression profiling of the 40 HIEGs

(A) The expression levels of the 40 genes in 37 human tissues (data from the HPA). The sort order of the genes from top to bottom is based on the ratio of the expression in cortex to the average in 37 tissues, which is indicated behind each gene name. (B) Hierarchical clustering of the expressions of 39 HIGEs in the 8 brain tissues of human, chimpanzee, gorilla, and gibbon. ACC: anterior cingulate cortex; DPFC: dorsolateral prefrontal cortex; VPFC: ventrolateral prefrontal cortex; PMC: premotor cortex; V1C: primary visual cortex; HIP: hippocampus; striatum; CB: cerebellum.
Table 1 (on next page)

The highlighted human intelligence evolution related candidates (HIEGs)
# Table 1: The highlighted human intelligence evolution related candidates (HIEGs).

| Gene   | CHR | Gene loci                  | Related SNP                      | Variation Type | Variation Position       | Variation Length | Genetic Disorders                  |
|--------|-----|----------------------------|----------------------------------|----------------|--------------------------|------------------|-------------------------------------|
| KDM4A  | 1   | 43650158-43705515          | rs2842188[14]                    | hCONDEL        | 43656932                 | 466              |                                     |
| NRXN1  | 2   | 49918505-51225575          | rs7557525[14]                    | hCONDEL        | 50146470                 | 3405             | Pitt Hopkins like syndrome2         |
|        |     |                            |                                  |                | 50146625                 | 958              |                                     |
|        |     |                            |                                  |                 | 50827590-50829782        | 2193             |                                     |
| AFF3   | 2   | 99545419-100142739         | rs71413877[14]                   | hCONDEL        | 100080237-100104859      | 24623            |                                     |
|        |     |                            |                                  |                | 99554281                 | 1250             |                                     |
| THSD7B | 2   | 136765545-137677717        | rs2558096[14]                    | hCONDEL        | 137333693                | 1212             |                                     |
| ARIH2  | 3   | 48918821-48986382          | rs13096357                       | insertion      | 48968205-48968380        | 176              |                                     |
|        |     |                            |                                  |                |                          |                  |                                     |
| STAB1  | 3   | 52495338-52524495          | rs4687625[14]                    | STR expansion   | 52509368-52509454        | 87               |                                     |
| SFMBT1 | 3   | 52903572-53046750          | rs4687625[14]                    | hCONDEL        | 52951656                 | 4034             |                                     |
| FOXP1  | 3   | 70952817-71583993          | rs11720523[14]                   | hCONDEL        | 71162689                 | 1760             | Mental retardation                  |
| CADM2  | 3   | 84958981-86074429          | rs6770622[14]                    | hCONDEL        | 85947229                 | 1042             |                                     |
| PCCB   | 3   | 136250306-136337896        | rs9853960[14]                    | exon_gain      | 136326325-136326385      | 60               | Propionic acidemia                  |
| TFDP2  | 3   | 141944428-142149544        | rs10804681[14]                   | hCONDEL        | 142144111                | 2011             |                                     |
| GRID2  | 4   | 92303622-93810157          | rs1972860[14]                    | hCONDEL        | 92649012                 | 135              | Spinocerebellar ataxia              |
| BANK1  | 4   | 101411286-102074812        | rs13107325[14]                   | hCONDEL        | 101422878                | 3650             |                                     |
| TTC29  | 4   | 146706638-146945882        | rs6840804[14]                    | hCONDEL        | 146796204                | 9671             |                                     |
| PDE4D  | 5   | 58969038-60522120          | rs34426618[14]                   | insertion      | 60429841-60432132        | 2292             |                                     |
| Gene     | CHR | Gene loci                  | Related SNP          | Variation_Type | Variation_Position | Variation Length | Genetic Disorders                                      |
|----------|-----|----------------------------|----------------------|----------------|-------------------|-----------------|-------------------------------------------------------|
| PAM      | 5   | 102753981-103031105         | rs76160968[14]       | hCONDEL        | 102883977         | 3868            | Neurodevelopmental disorder                           |
| FBXL17   | 5   | 107859035-108382098         | rs1438660            | hCONDEL        | 108119106         | 959             |                                                       |
|          |     | rs12187824[14]             |                      |                |                   |                 |                                                       |
| CALN1    | 7   | 71779491-72447151          | rs56150095[14]       | hCONDEL        | 72221700          | 2994            |                                                       |
| SND1     | 7   | 127652180-128092609        | rs4731392[14]        | hCONDEL        | 127808446         | 749             |                                                       |
| EXOC4    | 7   | 133253073-134066589        | rs1362739[11]        | insertion      | 13389352-133895456| 6105          |                                                       |
|          |     | rs4728302[14]             |                      |                |                   |                 |                                                       |
| SGCZ     | 8   | 14089864-15238339          | rs13253386[14]       | hCONDEL        | 14090971          | 277             |                                                       |
| TSNARE1  | 8   | 142212080-142403240        | rs4976976[14]        | STR expansion  | 142326108-142326158| 51            |                                                       |
| REEP3    | 10  | 63521363-63625123          | rs2393967[14]        | hCONDEL        | 63593088          | 554             |                                                       |
| GRIA4    | 11  | 105609994-105982092        | rs7116046[14]        | hCONDEL        | 105754761         | 3804            |                                                       |
|          |     | rs7116046[14]             |                      |                |                   |                 |                                                       |
| NCAM1    | 11  | 112961247-113278436        | rs2885208[14]        | hCONDEL        | 113152794         | 160             |                                                       |
| RERG     | 12  | 15107783-15348675          | rs55754731[14]       | hCONDEL        | 15107134          | 69              |                                                       |
| KMT2D    | 12  | 49018975-49059774          | rs1054442[14]        | STR contraction| 49032866          | 60              | Kabuki syndrome                                       |
|          |     | rs146865992[14]           |                      |                |                   |                 |                                                       |
| PRKD1    | 14  | 29576479-30191898          | rs971681[14]         | hCONDEL        | 29869703          | 1515            | Congenital heart defects and ectodermal dysplasia       |
|          |     | rs971681[14]              |                      |                |                   |                 |                                                       |
| FUT8     | 14  | 65410592-65744121          | 14:66113725_C_A[10] | insertion      | 65457887-65458201 | 315            | Glycosylation disorder                                 |
|          |     | rs75322822[14]            |                      |                |                   |                 |                                                       |
| GNB5     | 15  | 52115105-52191369          | rs7172979[14]        | hCONDEL        | 52177036          | 1472            | Language delay and ADHD/cognitive                       |
| RTF1     | 15  | 41408408-41483563          | rs75322822[14]       | hCONDEL        | 41430384          | 1965            |                                                       |
|          |     | rs7172979[14]             |                      | insertion      | 52121612-52121903 | 292            |                                                       |
| Gene   | Chromosome | Chromosome Region | SNPs/Alleles | Database | Data Presence | P-value | Position |
|--------|------------|------------------|--------------|----------|---------------|---------|----------|
| SKAP1  | 17         | 48133440-48430275| rs12928404[11] hCONDEL 48286479 | 343       |
|        |            |                  | rs12928404   hCONDEL 48259161 | 53        |
| DCC    | 18         | 52340172-53535903| rs71367283 hCONDEL 52358208 | 566       |
|        |            |                  | rs6508220[14] |           |               |         |          |
| ZNF584 | 19         | 58401504-58418327| rs73068339[14] insertion 58404219-58406377 | 2159      |
| SLC27A5| 19         | 58479512-58512413| rs73068339[14] deletion 58490956 | 3235      |
| PHF20  | 20         | 35771974-35950381| rs78084033[14] hCONDEL 35797866 | 3808      |
| STAU1  | 20         | 49113339-49188367| rs6019535[14] exon_gain 49179121-49179244 | 124       |
| DDX27  | 20         | 49219295-49244077| rs6019535[14] hCONDEL 49221110 | 809       |
| TRIOBP | 22         | 37696988-37776556| rs4396807[14] insertion 37723443-37724117 | 675       |
| EP300  | 22         | 41091786-41180079| rs4821995[14] hCONDEL 41135508 | 2279      |

*: the insertion was also identified as STR expansion

Intellectual developmental disorder

Corpus callosum agenesis

Nonsyndromic deafness

Rubinstein-Taybi syndrome2