Genome-wide identification of human genes involved in intelligence evolution through comparative genomics analysis (#40721)

First submission

Guidance from your Editor

Please submit by **13 Oct 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

4 Figure file(s)
6 Table file(s)
1 Other file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- ✓ Clear, unambiguous, professional English language used throughout.
- ✓ Intro & background to show context. Literature well referenced & relevant.
- ✓ Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- ✓ Figures are relevant, high quality, well labelled & described.
- ✓ Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- ✓ Original primary research within Scope of the journal.
- ✓ Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- ✓ Rigorous investigation performed to a high technical & ethical standard.
- ✓ Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- ℹ Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- ✓ All underlying data have been provided; they are robust, statistically sound, & controlled.
- ✓ Speculation is welcome, but should be identified as such.
- ✓ Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Genome-wide identification of human genes involved in intelligence evolution through comparative genomics analysis

Mengjie Li [1] , **Wenting Zhang** [1] , **Xiaoyi Zhou** Corresp.

[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 from great apes to humans is an important undertaking in the science of human genetics. Recently, a great deal of biological research has been conducted to search for the human-specific genes and variations that have resulted in the significant increase in human intelligence over that of apes. It is very important, yet extremely difficult, to discover additional genes involved in the evolution of human intelligence and various approaches need to be taken to further explore the issue. We designed a new strategy to discover genes involved in the evolution of human intelligence. Information was collected from published GWAS works on intelligence and from these a total of 549 genes located within the associated loci were identified. The intelligence-related genes containing a human-specific variation were detected based on the latest high-quality genome assemblies of three great apes, including 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. However, there is a distinct expression pattern between humans and other species, especially in ~~the tissues of the~~ neocortex tissues. Our work may provide a list of strong candidates for the evolution of human intelligence, which may also imply that some intelligence-related genes may undergo inter-species evolution and contain intra-species variation. More importantly, the work provides a new method in searching for the key genes in human evolutionary genetics.

Genome-wide identification of human genes involved in intelligence evolution through comparative genomics analysis

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 from great apes to humans is an important undertaking in the science of human genetics. Recently, a great deal of biological research has been conducted to search for the human-specific genes and variations that have resulted in the significant increase in human intelligence over that of apes. It is very important, yet extremely difficult, to discover additional genes involved in the evolution of human intelligence and various approaches need to be taken to further explore the issue. We designed a new strategy to discover genes involved in the evolution of human intelligence. Information was collected from published GWAS works on intelligence and from these a total of 549 genes located within the associated loci were identified. The intelligence-related genes containing a human-specific variation were detected based on the latest high-quality genome assemblies of three great apes, including 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. However, there is a distinct expression pattern between humans and other species, especially in the tissues of the neocortex tissues. Our work may provide a list of strong candidates for the evolution of human intelligence, which may also imply that some intelligence-related genes may undergo inter-species evolution and contain intra-species variation. More importantly, the work provides a new method in searching for the key genes in human evolutionary genetics.

Introduction

The rapid change in intelligence from great apes to humans is one of the greatest mysteries in evolutionary genetics (Varki et al., 2008). Physically, humans have brains with significantly increased size and complexity with a large expansion of the neocortex ~~versus~~ their ape counterparts (Rakic., 2009; Chenn et al., 2002; Lui et al., 2011). Corresponding to the increased size of the neocortex, humans are much more intelligent than chimpanzees, although chimpanzees are able to learn to use some specific tools and undertake certain tasks. The human advantage in intelligence has helped this species compete with nature over the past few million years and survive, create tools, civilizations, and sciences (Deary et al., 2012). The superior intelligence 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 hominid-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 that have led to the remarkable leap in human intelligence. For example, in a very recent study, the information from gene expression profiling was integrated with the information from gene duplications in the hominid 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 and found to be able to expand 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 and variations have been identified in which the genetic changes often occurred in gene regulation regions or resulted from the hominid-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 the genomic loci and related genes underlying intelligence. Despite a significant enrichment in the nervous system, the functional links of the identified genes are diverse and 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 al., 2018). This suggests that the evolution of intelligence from great apes to humans should also be a complicated process and that the causative genes may be derived mainly from those in the central nervous system; however, genes from many related biological processes may be also involved.

Investigating the human-specific variations (genetic differences between human and the great apes) would provide key clues for understanding the process of the evolution of human intelligence. However, the previous reference genomes of great apes (e.g., the human sequence guided assembling from short reads) are 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. However, 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 human and great apes is needed to comprehensively mine 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* assembly (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. These differences are responsible for all phenotypic differences between humans and apes but it is 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 gene may have both inter-species (between human and the great apes) and intra-species (within human) variations. We collected genomic loci identified by several ~~sets of~~ GWAS on human intelligence. Genes in these loci (termed as intelligence-associated-genes) could be related to the intra-species intelligence differences. DNA sequence variations between the human genome and ~~that~~ of the great ape, revealed by recent high-quality sequencing (Kronenberg et al., 2018; Gordon et al., 2016), were considered to be the inter-species variations. 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 contained human-specific structural variations, including tens of strong candidate genes related to the evolution of human intelligence. 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

Identification of candidate genes from GWAS on human intelligence

To exploit the genes related to human intelligence, six major works were collected from GWAS or meta-analyses on human intelligence with a large sample size ~~in the recent~~ five years (various intelligence related phenotypes including general cognitive, reaction time, verbal-numerical reasoning) which identified 271 loci associated with intelligence in the human genome (Table S1). The six GWAS studies include (i) meta-analyses ~~for~~ 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 for calculated Spearman's *g* or a primary measure of fluid intelligence (*n*=78,308, Sniekers et al., 2017); (iv) GWAS using human populations with extremely high intelligence (*n*=1,238, Trampush et al., 2017); v) meta-analysis and gene-based analysis for human cognition using 24 cohorts (*n*=35,298, 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 ANNOVAR annotations) were integrated with ~~the~~ redundancies (the same gene identified in more than one study) removed. We take these genes as "bait genes" (Table S2), which are candidates for human intelligence-related genes.

It should be noted here 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) and GWAS have relatively low power in discovering the associations in the ~~23th~~ chromosome, the potential genes related to intelligence in the sex chromosomes were not included in our "bait genes".

Comparative genomics analysis for the intelligence genes

Previous studies (McLean et al., 2011; Kronenberg et al., 2018; Dennis et al., 2017) have identified a number of human-specific variants that were included in our work including exon gain and loss, STR, indels, hCONDEL, HSDs, and large structural 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 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 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~~s of~~ human-specific variations from previous reports (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 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, thus generating the "prey genes" (Table1 and Table S3-S5).

Investigation of HIEGs

Most "prey" genes contained the human-specific variation in introns ~~with the variation~~ far away from the exon-intron junction site. This may not affect~~ed~~ the gene functions, so only the genes that contained variations in the coding regions were considered, named ~~as~~ 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-specific transcripts and their most similar principle transcripts in humans and great apes are shown in Figure 3. The human principle isoform, the variant isoform, the chimpanzee isoform, the gorilla isoform, and the orangutan isoform of the *PCCB* gene are ENST00000469217 (NM_001178014), ENST00000466072, ENSPTRT00000028811, ENSGGOT00000005484 and ENSPPYT00000016431, respectively. The human principle isoform, the variant isoform, the chimpanzee isoform, the gorilla isoform, and the orangutan isoform of the *STAU1* gene are ENST00000371856, ENST00000340954, ENSPTRT00000050938, ENSGGOT00000048652 and ENSPPYT00000012903, 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 ~~was~~ downloaded from the Human Protein Atlas (Http://www.proteinatlas.org), including 102 samples of 37 tissues, in which TPM (transcripts per million) ~~was~~ used for the evaluation of expression levels. In order to compare the expression levels between humans and great apes, the RNA-Seq data (NCBI ID: GSE100796) from 107 samples of 8 brain regions from humans, chimpanzees, gorillas, and gibbons ~~was~~ used (Xu et al., 2018). The brain tissues included five neocortical areas, namely the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (DPFC), the ventrolateral prefrontal cortex (VPFC), the premotor cortex (PMC), the primary visual cortex (V1C), and three other brain tissues including the hippocampus (HIP), the striatum, and the cerebellum (CB). Hierarchical cluster analysis was applied using the RPKM (reads per kilobase per million) of 39 genes in 8 brain tissues of humans and three primates and was displayed using the software MeV4.2 (http://www.tm4.org/mev.html).

Results

A strategy for searching for genes in human intelligence evolution

The study strategy that was used for detecting the candidate genes in the evolution of human intelligence is briefly described in Figure 1. Hundreds of genetic loci have been identified by GWAS in the last ten years and studies have been conducted using large population data for human intelligence and its related traits (e.g., general cognitive ability, reaction time, and verbal-numerical reasoning). We take the genes implicated by these genetic loci as “bait genes” which are related to human intelligence. In the meantime, comparative genomics analyses between the human genome and the great ape genomes (including those of the chimpanzee, gorilla, and orangutan) were used to identify kinds of specific human variations. The human-specific variations are taken as "variation ponds" that are related to human evolution. The "baits" and "ponds" were then integrated to detect intelligence-related genes with human-specific variations as "prey genes" which are related to the evolution of human intelligence. Furthermore, based on the potential influence of the sequence variations and other gene function studies, the strong candidates in the "prey genes" ~~was~~ highlighted as potential human intelligence evolution related genes (HIEGs). The expression profiles of HIEGs were also investigated.

Preparation of bait genes

We integrated six high quality GWAS works and the large-scale meta-analyses from the last five years that were found through publication searches. This enabled the identification of a total of 271 associated loci in the human genome underlying intelligence related phenotypes (Table S1). 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 cannot identify all related genes at one time. According to the human gene annotation information, a total of 549 human genes were found to be located within the 271 associated loci. The 549 human genes, used as "bait genes" (Table S2), were distributed across the whole human genome as indicated by red dots in Fig. 2.

The "baits" constituted intelligence-related genes from GWAS underlying intra-species intelligence variations in human populations. A subset of these may be involved in the evolution of intelligence from great apes to humans. Hence, the comparative genomics analyses between great apes and humans were further added to examine whether there was a human-specific variation hit by our "bait genes".

Human-specific variation on the intelligence related genes

The nearly compete genome sequences from chimpanzees, gorillas, and orangutans have become fully available recently through single-molecule, real-time (SMRT) long-read sequencing, providing a large number of high-quality sequence differences between the human genome and great ape genomes. The new genome assemblies had improved the resolution of large and complex regions. We incorporated the human-specific structural variations (from intermediate size to large size) into a "variation pond", including exon gain/lost, short tandem repeats (STRs), insertion/deletion (indels) of more than 50bp, and inversions. Moreover, considering the important role of the human-specific segmental duplications (HSDs, >1kb 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 genomic information of both the macaque and mouse (Dennis et al., 2017) were added into the "pond". 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 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 during the evolution of the human lineage, *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* =1.956×10-9, 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, comparative genomics analysis showed human-specific variation led to the generation of a new transcript variant. 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 (note: transcriptome data was obtained from the Human Protein Atlas, HPA). However, the relative expression level in the cerebral cortex, relative to the average expression level of this gene in 37 tissues, of the *PCCB* variant was higher than the principal transcript (Fig. 3B), implying a potential role of the human-specific *PCCB* transcript variant in the cerebral cortex. Although there were few reports on neuron system development for *PCCB*, 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 were 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, was not detected in any transcripts of the chimpanzee, gorilla, or orangutan. The isoform, with the addition of a 123-bp exon, is a new isoform in the human transcriptome (Fig. 3C). 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. 3D). 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 "prey genes" (26 genes containing the STR contractions and 74 genes containing the STR expansions, Table S3). Most of the human-specific STR variation was 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 noncoding isoforms 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*, encodes an H3K4 histone methyltransferase in humans made up of 5537 amino acids. GWAS for intelligence detected one associated locus (P =2.518×10^{-14}, Savage et al., 2018) on chromosome 12 containing the candidate gene *KMT2D*. Compared with the *KMT2D* sequence in great apes, the STR region contracted for 60-bp in the c.11838, and 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. 3E). The

STR region is 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 affects mental capabilities and most of these patients show 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 7 insertions and 1 deletion for the exonic regions, as highlighted by the HIEGs (Table 1). Insertions in *PDE4D*, *NRXN1*, *EXOC4*, *FUT8*, and *ZNF584*, and the deletion in *SLC27A5* affect the lengths of the noncoding isoforms of the six genes, while one insertion in *TRIOBP* resulted in a gain of 675-bp coding regions (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. 3F). 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 gene involved with neural tissue development (Seipel et al., 2001). Mutations in *TRIOBP* can cause autosomal recessive deafness-28 (DFNB28), which is a genetic disease, and surprisingly, several causal mutations have been located within the human-specific insertion regions (e.g., R347X and Q297X) (Shahin et al., 2006). Hence, the human-specific variation in *TRIOBP* is probably also involved in the evolution of human intelligence.

We also searched for human conserved deletions (hCONDELs) near the "bait genes" that had been previously 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 great apes, the macaque, and the mouse). In total, 28 "prey genes" were identified as containing the hCONDEL and all 28 genes were taken into HIEGs with the high lineage specificity of hCONDEL (Table 1) 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 associated locus on chromosome 15 (*P*= 2.47×10^{-11}, Savage et al., 2018), which is responsible for encoding the beta subunit of heterotrimeric GTP-binding proteins (G proteins), was found to be one of the HIEGs. *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). Aligned with the genome sequences of 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). 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 a rearrangement in which a 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) of > 5 kb that were also reported (Prado-Martinez et al., 2013). Among them, one 24.6-kb HSD was detected to be overlapped with the *AFF3* gene. The *AFF3* 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 associated locus in found in two distinct meta-analyses conducted on intelligence (*P*= 1.56×10^{-8} [16] or *P*= 3.41×10^{-10}, Savage et al., 2018), but no functional reports on neuron development are available.

Expression profiling analysis of the HIEGs

The transcriptome data from 37 human tissues in 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 expressions in the cerebral cortex than its average expression in all 37 tissues. Furthermore, there were 12 genes with more than twofold expressions in the cerebral cortex versus its average expression levels. These were regarded as genes that were potentially involved in the development of the cerebral cortex. Except the case for *FUT8*, the remaining 11 genes are the hCONDEL-containing genes. Four genes, including

GRIA4, *NRXN1*, *CADM2*, and *CALN,* were highly expressed in brain tissues but had low expressions in the other tissues; *SGCZ*, *DCC,* and *GRID2* showed relatively high expressions in brain tissues with low expressions in all other tissues; and *NCAM1*, *FBXL17*, *FUT8* and *GNB5* were expressed well in all tissues.

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 very recently reported (Xu et al., 2018), which enabled a comprehensive interspecies comparison. This dataset was used to examine whether the human-specific variation led to the expression changes in the cortex of HIEGs between humans and great apes. The expression data of the other 39 genes, with the exception of *STAU1,* can 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 great apes (fold change <0.5), while the expressions of *STAB1* in the human neocortex were much higher than those in great apes (fold change= 4.8). Hierarchical clustering was also performed for the 39 gene expressions in all samples. We found that 5 neocortices 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, it was possible to determine that the human-specific variation in the intelligence-related genes may have effects on the changes in expression in the neocortex tissues.

Discussion

The understanding of human intelligence from the view of evolutionary genetics is an important scientific undertaking. Science magazine posted 125 scientific questions to cover over the next quarter-century which were selected to cover various disciplines as a way to celebrate its 125th anniversary (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, a genetic rise in intelligence over the last million years is a key step in human evolution, making us uniquely Homo sapiens. The high-quality genome assemblies of three great apes were completed recently and the improved sequence contiguity enabled more comprehensive and accurate discoveries with complex variations. With both the human genome sequence and those of our closest relatives becoming available, it is possible to pinpoint the genetic changes underlying the phenotypic differences between humans and great apes. However, it is still very challenging to study the causative genetic changes that are responsible for the rise in human intelligence because the molecular mechanisms controlling human intelligence are largely unknown. One possible solution is to utilize the genetic findings on intelligence from GWAS. A subset of genes with that reveals the differences in the evolution of intelligence in the human-chimpanzee 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 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.

The associated loci from GWAS in humans often contain several candidate genes, 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 “baits”, although usually for each associated locus only a single one is causative. Consequently, a total of 549 human genes were included for the 271 associated loci and “bait genes” contain many false positives. However, 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 intelligence the development of intelligence in 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 ~~levels~~ from great apes to humans, leaving many remaining gaps in our knowledge. More integrated approaches incorporating genetics, genomics, bioinformatics, and development biology will be needed in future works.

One candidate gene with an exon gain in human evolution, Staufen1 (*STAU1*), which is involved in the transport, relocation, translation of mRNA and mRNA decay is known to regulate the post-transcription phase (Paul et al., 2018). However, the loss of *STAU1* function in mice resulted in impaired mRNA transport and reduced synapse formation (Vessey et al., 2008). Another candidate gene, *KMT2D* with STR contractions, showed a number of truncating mutations within *KMT2D* resulting in mRNA degradation through nonsense-mediated mRNA decay, contributing to protein haploinsufficiency (Micale et al., 2014). It is unclear whether there are any functional links between the two genes for mRNA processes in brain developments.

Conclusion

GWAS has identified hundreds of genes associated with intelligence variation in human populations. Through inter-species genome comparisons with great apes, we found a small proportion of intelligence-related genes that also contained a human-specific variation which were detected in multiple high-quality genome assemblies of humans and its closest relatives. Through integrated analytical approaches, especially the careful checking of sequence alignments and gene annotations, we identified 40 strong 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 variations and inter-species variations. The intra-species variation underlies the small variation of intelligence levels for different human individuals while the inter-species variation controlled the large genetic differences of intelligence between great apes and humans. This work may provide a list of candidate genes to be used in subsequent studies as well as a new route for discovering genes that are important in the study of human intelligence evolution.

Acknowledgments

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

servers.

References

Bailey JA, Yavor AM, Massa HF, Trask BJ, and 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, et al. 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, and 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, et al. 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, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications Nature 437:88-93.

Chenn A, and 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, et al. 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, et al. 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, et al. 2016. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112151). Mol Psychiatry 21:758-767.

Deary IJ. 2016. Intelligence. Annu Rev Psychol 63**:**453-482.

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

Dennis MY, Harshman L, Nelson BJ, Penn O, Cantsilieris S, Huddleston J, et al. 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, et al. 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, et al. 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, et al. 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, et al. 2016. Long-read sequence assembly of the gorilla genome. Science 352: aae0344.

Hill WD, Davies G, Harris SE, Hagenaars SP, Liewald DC, Penke L, et al. 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, et al. 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, et al. 2016. The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. Elife 5: e18197

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

Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, et al. 2018. High-resolution comparative analysis of great ape genomes. Science. 360

Lehman N, Mazery AC, Visier A, Baumann C, Lachesnais D, Capri Y, et al. 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, et al. 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, and Kriegstein AR. 2011. Development and Evolution of the Human Neocortex. Cell 146: 18-36.

Mason JM, and 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, et al. 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, et al. 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, and 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, et al. 2013. Great ape genetic diversity and population history. Nature 499:471-475.

Prufer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, et al. 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, et al. 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, et al. 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, et al. 2012. Neurologic considerations in propionic acidemia. Mol Genet Metab 105:10-15.

Seipel K, O'Brien SP, Iannotti E, Medley QG, and 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, et al. 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, et al. 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, et al. 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, et al. 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, Oksvold P, Mardinoglu A, et al. 2015. Proteomics. Tissue-based map of the human proteome. Science 347:1260419.

Varki A, Geschwind DH, and Eichler EE. 2008. Explaining human uniqueness: genome interactions with environment, behaviour and culture. Nat Rev Genet 9:749-763.

Vessey JP, Macchi P, Stein JM, Mikl,M, Hawker KN, Vogelsang P, et al. 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, et al. 2018. Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. Nat Commun 9:3945.

Witters P, Debbold E, Crivelly K, Vande Kerckhove K, Corthouts K, Debbold B, et al. 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, et al. 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, et al. 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, et al. 2018. A genome-wide association study for extremely high intelligence. Mol Psychiatry 23:1226-1232.

Figure 1

Strategy of the genome-wide analysis forintelligence associated genes containing human-specific variations

Grey areas indicated the result used from the previous studies including six GWAS works on intelligence and the comparative genomics analysis between human and great apes.

Figure 1

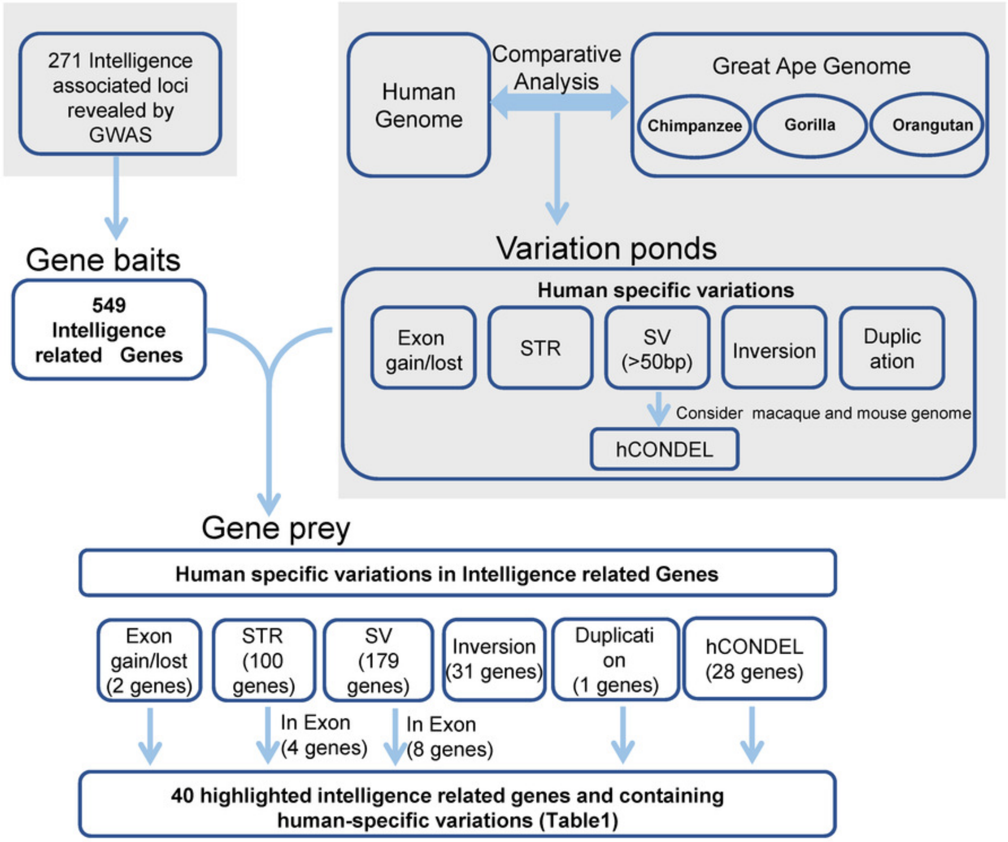

Figure 1: Strategy of the genome-wide analysis for intelligence associated genes containing human-specific variations.

Grey areas indicated the result used from the previous studies including six GWAS works on intelligence and the comparative genomics analysis between human and great apes.

Figure 2

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

The intelligence associated genes (bait genes) are 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 2

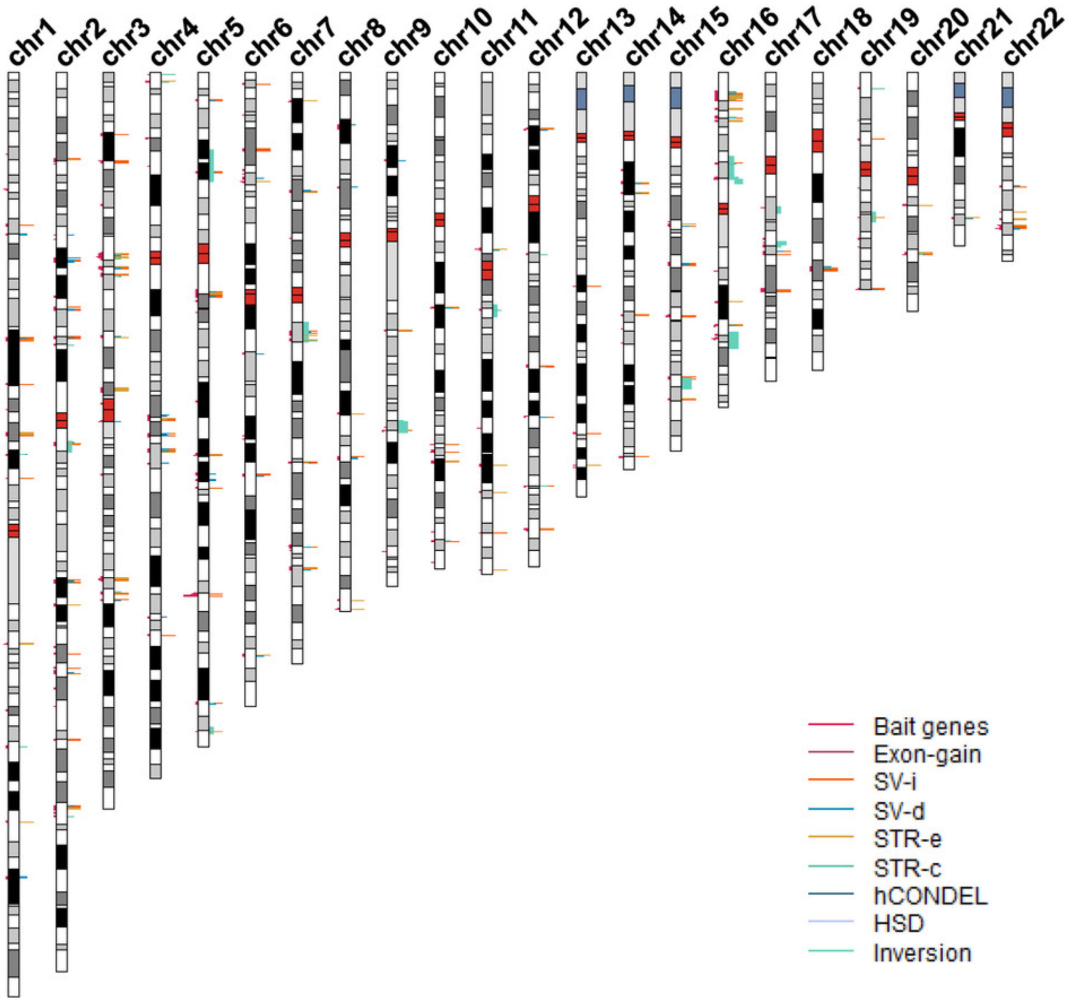

Figure 2: The distribution of intelligence associated genes and their located human specific variations in the 22 autosomes of the human genome.
The intelligence associated genes (bait genes) are on the left side of the chromosome bars. The human specific variations around the intelligence associated 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 candidates involved in humanintelligence evolution

(A) Gene structures of PCCB in human and great ape genomes. ~~O~~range box represents the gained exon(also indicated by red arrow). The mutations in patients with intellectual disability are labeled. ~~S~~olid box represent~~s~~ coding exon, ~~and~~ hollow box represent UTR; (B) The expression level of PCCB ~~of the principle~~ isoform and ~~the~~ variant isoform. Left panel: TPM of transcript in cerebral cortex. Right panel: TPM in cerebral cortex relative to 37 human tissues. (C) Gene structures of STAU1. (D) The expression level of STAU1. (E) The schematic representation of KMT2D and TRIOBP proteins. 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 were hided amino acid.

Figure 3

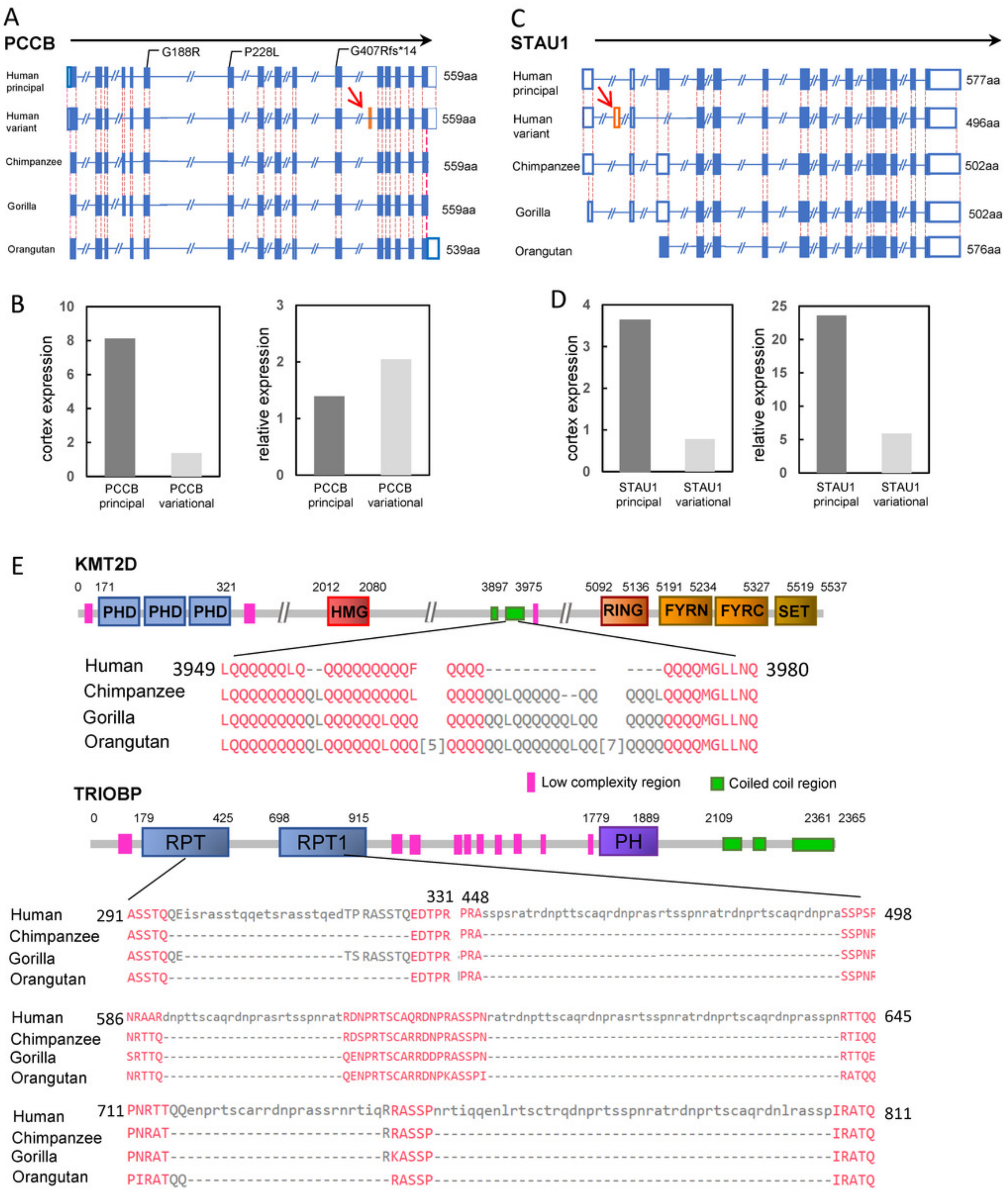

Figure: 3 Four candidate genes involved in human intelligence evolution.
(A) Gene structures of *PCCB* in human and great ape genomes. Orange box represents the gained exon(also indicated by red arrow). The mutations in patients with intellectual disability are labeled. Solid box represents coding exon, and hollow box represent UTR; (B) The expression level of PCCB of the principle isoform and the variant isoform. Left panel: TPM of transcript in cerebral cortex. Right panel: TPM in cerebral cortex relative to 37 human tissues. (C) Gene structures of STAU1. (D) The expression level of STAU1. (E) The schematic representation of KMT2D and TRIOBP proteins. 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 were hided amino acid.

Figure 4

Expression profiling analysis of 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

Figure 4

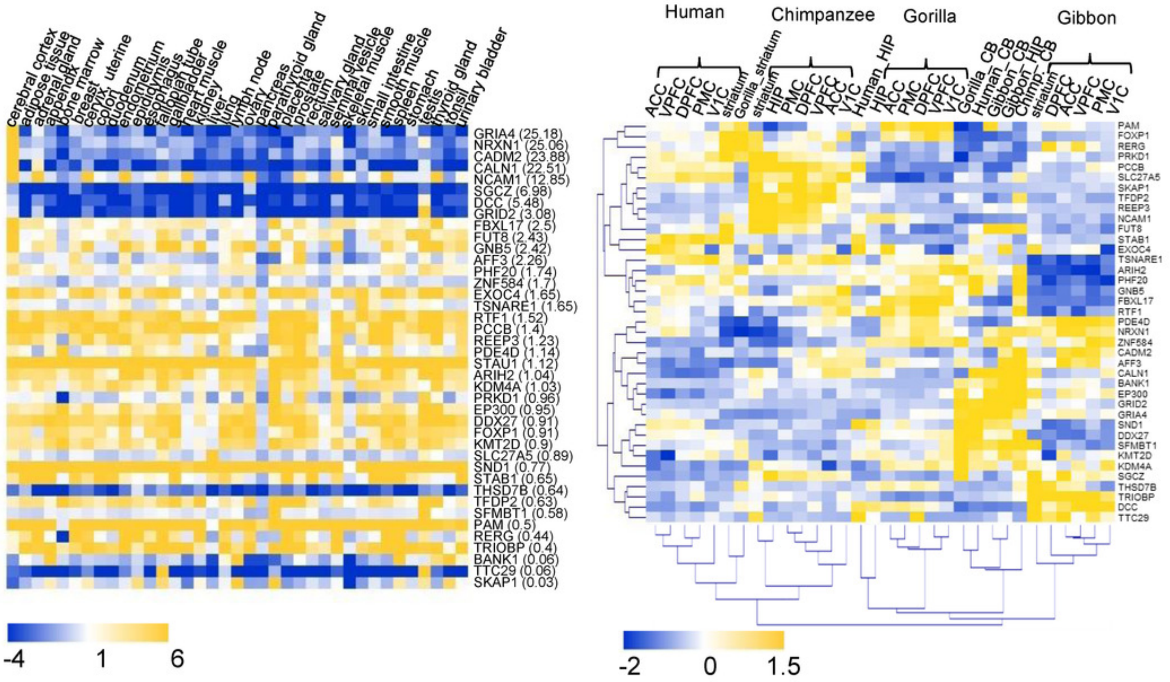

Figure: 4 Expression profiling analysis of 40 HIEGs.

(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)

Table1 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
				hCONDEL	50146625	958	
				insertion	50827590-50829782	2193	
AFF3	2	99545419-100142739	rs71413877[14]	HSD	100080237-100104859	24623	
			rs13010010[11]	hCONDEL	99554281	1250	
THSD7B	2	136765545-137677717	rs2558096[14]	hCONDEL	137333693	1212	
ARIH2	3	48918821-48986382	rs13096357	insertion	48968205-48968380	176	
			rs2352974	(STR_expansion)			
			rs73078367[14]				
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	71460722	449	Mental retardation
				hCONDEL	71162689	1760	
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	
			rs13107325	hCONDEL	101990081	4838	
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	
FBXL17	5	107859035-108382098	rs1438660 rs12187824[14]	hCONDEL	108119106	959	
CALN1	7	71779491-72447151	rs56150095[14]	hCONDEL	72221700	2994	
SND1	7	127652180-128092609	rs4731392[14]	hCONDEL	127808446	749	
EXOC4	7	133253073-134066589	rs1362739[11] rs4728302[14]	insertion	133889352-133895456	6105	
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	Neurodevelopmental disorder
NCAM1	11	112961247-113278436	rs2885208[14]	hCONDEL	113152794	160	
RERG	12	15107783-15348675	rs55754731[14]	hCONDEL	15107134	69	
KMT2D	12	49018975-49059774	rs1054442 rs146865992[14]	STR_constraction	49032866	60	Kabuki syndrome
PRKD1	14	29576479-30191898	rs971681[14]	hCONDEL	29869703	1515	Congenital heart defects and ectodermal dysplasia
FUT8	14	65410592-65744121	14:66113725_C_A[10]	insertion	65457887-65458201	315	Glycosylation disorder
RTF1	15	41408408-41483563	rs75322822[14]	hCONDEL	41430384	1965	
GNB5	15	52115105-52191369	rs7172979[14]	hCONDEL insertion	52177036 52121612-52121903	1472 292	Language delay and ADHD/cognitive

							impairment; Intellectual developmental disorder
SKAP1	17	48133440-48430275	rs12928404[11]	hCONDEL	48286479	343	
			rs12928404	hCONDEL	48259161	53	
DCC	18	52340172-53535903	rs71367283 rs6508220[14]	hCONDEL	52358208	566	Corpus callosum agenesis
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	Nonsyndromic deafness
EP300	22	41091786-41180079	rs4821995[14]	hCONDEL	41135508	2279	Rubinstein-Taybi syndrome2