

SUPPLEMENTARY INFORMATION

A Pangenome Reference of 36 Chinese populations

Correspondence and requests for materials should be addressed to Dr. Shuhua Xu
(xushua@fudan.edu.cn)

Contents

S1.	Populations and Samples	3
S2.	Ethical statement.....	3
S3.	Establishment of Epstein-Barr Virus (EBV) immortalized B cell lines	4
S4.	DNA Extraction	4
S5.	DNA Quantification & Qualification.....	6
S6.	Library construction and sequencing.....	6
S7.	Hi-C sequencing	10
S8.	High-Fidelity (HiFi) sequencing.....	11
S9.	Oxford Nanopore (ONT) Sequencing.....	12
S10.	Genome Assembly and Quality Control.....	13
S11.	Alignment of Hifi reads to CHM13.....	13
S12.	Evaluate alignment quality of short-reads using CPC reference	14
S13.	Pangenome graph construction.....	14
S14.	Pangenome growth	15
S15.	Statistics on variants	15
S16.	Identification and visualization of novel variants.....	16
S17.	Complex structural variation detection.....	16
S18.	Alignment, variant identification and filtering	17
S19.	Population Genetic Analyses	18
S20.	Functional annotation and enrichment analysis of genes	18
S21.	Archaic introgression segments detection and functional annotation	19
S22.	References.....	- 21 -
S23.	Supplementary Tables.....	- 23 -
	Table S1. Information of the samples for sequencing in CPC.....	- 24 -
	Table S2 Summary statistics of the 116 CPC genome assemblies	- 25 -
	Table S3 A comparison of variant calling based on linear genome reference and graph genome reference.....	- 27 -
	Table S4 CPC-specific-CNV-related genes compared to HPRC assemblies.....	- 28 -
	Table S5 Overlapped CNV-related-genes between CPC and HPRC	- 34 -
	Table S6. Functional enrichment of genes affected by the novel SVs (≥ 1 kb) in the CPC assembly set.	- 36 -
	Table S7 Copy number of <i>RASA4(B)</i> genes in CPC samples	- 36 -
	Table S8 Counts of haplotypes with different <i>RASA4(B)</i> copy number in different populations	- 38 -

Table S9	Frequency distribution of <i>RASA4(B)</i> copy number in different populations.....	39 -
Table S10	Genes and Frequencies of CDS Regions Covered by CPC Ancient Introgression Segments.....	40 -
Table S11	Functional enrichment of genes in CDS which CPC-AIS located and these genes were found in at least 5 samples.....	45 -
Table S12	Functional enrichment of genes in CDS which CPC-AIS located and these genes were found in at least 1 sample.....	47 -
Table S13	Functional enrichment of genes in CDS where CPC-AIS are located and observed at least 10 samples.....	50 -
Table S14	Genes and Frequencies of CDS Regions Covered by CPC-specific Ancient Introgression Segments.....	51 -
Table S15	Functional enrichment of genes in CDS where CPC- specific-AIS are located.....	53 -
Table S16	GeneAnalytics results of genes in CDS where CPC-specific-AIS are located..	54 -
S24.	Supplementary Figures	55 -
Fig. S1	Geographical distribution of CPC core samples.	56 -
Fig. S2	Assembly pipeline and quality control of CPC samples.	57 -
Fig. S3	Evaluation of overall assembly correction of CPC genome assemblies.	58 -
Fig. S4	Evaluation of three types of assembly errors of CPC genome assemblies.....	59 -
Fig. S5	Novel sequences identified from CPC samples.....	60 -
Fig. S6	Overlapped CNV genes with higher frequency in CPC than that in two HPRC subsets.	61 -
Fig. S7	The count of CSVs in each sample.....	62 -
Fig. S8	The number of CSVs for each CSV type in each sample.....	63 -
Fig. S9	The count of CSVs with different shared sample numbers.	64 -
Fig. S10	The length distribution of the final 706 CSVs.....	65 -
Fig. S11	The number of CSVs in each CSV class.	66 -
Fig. S12	Functional enrichment of 2617 genes in CDS where CPC-AIS are located and observed in at least 5 samples.....	67 -
Fig. S13	Functional enrichment of genes in CDS where AIS are located and observed in at least 1 sample.....	68 -
Fig. S14	Functional enrichment of genes in CDS where AIS are located and observed in at least 10 samples.	69 -
Fig. S15	Functional enrichment of 1211 genes in CDS where CPC-AIS-specific are located.	70 -
Fig. S16	The proportion of properly paired reads of 10 East Asian samples from KGP. .	71 -
Fig. S17	The insert size average over 10 East Asian samples.	72 -
Fig. S18	Standard deviation of insertion size at whole genome level and repeats region.-	73 -
Fig. S19	Number of identified small variants and SVs inferred from the pangenome graph..	74 -
Fig. S20	The length distribution of SVs in the CPC reference graph Pan-genome.....	75 -
Fig. S21	The ancestral component Inference of CPC cohort.....	76 -
Fig. S22	Growth of non-reference sequences with haplotype addition.	77 -
Fig. S23	Comparison of SV distribution between CPC and HPRC reference Pangenomes. ..	-

S1. Populations and Samples

For the Phase I of CPC, we selected 68 samples from 731 individuals with genomes deep-sequenced using next-generation sequencing. It turned out that we applied a procedure similar to that of HPRC to select the representative samples of a subpopulation¹. After quality control of the HiFi data, eventually 58 samples were retained for all the further analyses in this study. These include Achang (n=1), Bai (n=1), Blang (n=2), Bouyei (n=1), Chosen (n=3), Daur (n=2), Deang (n=2), Dong (n=2), Drung (n=1), Evenki (n=1), Hezhen (n=1), Hui (n=1), Jingpo (n=1), Jino (n=1), Kazakh (n=1), Kyrgyz (n=1), Kinh (n=2), Lisu (n=1), Man (n=1), Miao (n=2), Mongol (n=3), Mongol-TH (n=1), Mosuo (n=3), Naxi (n=1), Oroqen (n=1), Qiang (n=3), Salar (n=2), She (n=3), Tu (n=2), Tujia (n=3), Uyghur (n=1), Wa (n=2), Yao (n=1), Yi (n=1), Yugur (n=2), and Zhuang (n=1).

S2. Ethical statement

Informed consent was obtained from all individual participants included in the study. The personal identifiers of all samples, if any existed, were stripped off before sequencing and analysis. All procedures were in accordance with the ethical standards of the Responsible Committee on Human Experimentation and the 1964 Helsinki Declaration, its later amendments (2000) or comparable ethical standards. The research content and procedures performed in studies involving human participants were approved by the Biomedical Research Ethics Committee of Shanghai Institutes for Biological Sciences (No. ER-SIBS-261408), the Biomedical Research Ethics Committee of Kunming Institute of Zoology, Chinese Academy of Sciences (No. SMKX-20180715-154), the Biomedical Research Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University (No. XJTU1AF2021LSK-051).

S3. Establishment of Epstein-Barr Virus (EBV) immortalized

B cell lines

Blood collection and lymphocyte separation

For each participant, 3ml peripheral blood were collected in vacutainer tubes containing EDTA. The blood was diluted with 3mL RPMI1640 and spun at 1200G for 20min at room temperature (24°C). The opaque lymphocyte layer was collected down to the erythrocyte pellet and transferred to another 15-ml centrifuge tube. The cells were washed with centrifugation in 7mL RPMI1640 at 800G for 8min and then the clear top plasma layer was discarded.

Immortalization of lymphocytes

Lymphocytes were shaken at 120RPM for 1 hour with 4 ml EBV stock from Epstein-Barr virus (EBV)-transformed B95-8 marmoset cell line. The cells were grown at 37°C in RPMI1640 medium, supplemented with 20% fetal bovine serum and cyclosporine A with a final concentration of 2µg/ml. Every 4 days of incubation, the cultures were examined and photographed by microscopy. The cell aggregates of proliferative lymphoblast cells indicated successful transformation. These LCLs are limited in their growth to minimize genetic changes that may occur during extended cell culture. The whole procedure was done in the Immortalize Cell Bank of Chinese Ethnic Groups hosted in the Institute of Medical Biology, CAMS.

S4. DNA Extraction

DNA preparation for NGS

Genomic DNA was extracted from the blood samples using QIAGEN DNeasy Blood & Tissue Kit. DNA concentrations were measured with the NanoDrop 2000 (Thermo Fisher Scientific), and sheared with Covaris S220 Sonicator (Covaris) to target of 500–600 base pairs (bp) average size. Fragmented DNA was purified using Sample Purification Beads (Illumina). Adapter-ligated libraries were

prepared with the TruSeq Nano DNA Sample Prep Kits (Illumina) according to Illumina-provided protocol. DNA concentrations of the resulting sequencing libraries were measured with the Qubit 2.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific). Quantities and sizes of the resulting sequencing libraries were analyzed using Agilent BioAnalyzer 2100 (Agilent). The libraries were used in cluster formation on an Illumina cBOT cluster generation system with HiSeq X HD PE Cluster Kits (illumina). Whole-genome sequencing, with a target coverage 30× for 150 bp paired-end reads, was performed in WuxiNextCODE at Shanghai using an Illumina HiSeq X following Illumina-provided protocols. Each sample was run on a unique lane with at least 90 GB PF data (Table S2) and reads data were quality controlled for ensuring that 80% of the bases achieved at least a base quality score of 30 (Table S2).

DNA preparation for TGS

Genomic DNA was extracted from the peripheral blood samples and cells line using the optimized Cetyl Trimethyl Ammonium Bromide (CTAB) based protocol. Take frozen finely powdered tissues in a new 50 mL Falcon tube and mixed with the pre-heated extraction buffer (10 ml) for one sample. Briefly, vortex for 30sec to ensure sample is fully mixed with buffer. Fine grind is a key to obtaining high DNA quantity with lesser artifacts of resin. Next the falcon tube was kept into the 65°C incubator or water bath and mix gently by inversion after every 10 min till 45 min. After incubation, place the tube at room temperature for five min to reach to room temperature environment. Centrifuge the 50ml falcon tube for 5 min at 3000×g on room temperature. For the Third-Generation Sequencing (TGS), the greater the genome size, lower is speed of initial centrifugation. Transfer 1ml of supernatant to each 2ml Eppendorf tubes already containing 1 ml of chloroform: isoamyl alcohol (24:1). Mix supernatant and chloroform: isoamyl alcohol by gentle inversions for 10 min and subsequently place the tube on ice for 10 min. Centrifuge the tube for 10 min at 5000 × g at 4°C. Transfer the upper aqueous phase into new 2ml tubes and add 5µl RNase A (10 mg/mL). Place the tube for 30 min at 37°C. After RNase treatment, add one volume of chloroform: isoamyl alcohol again and

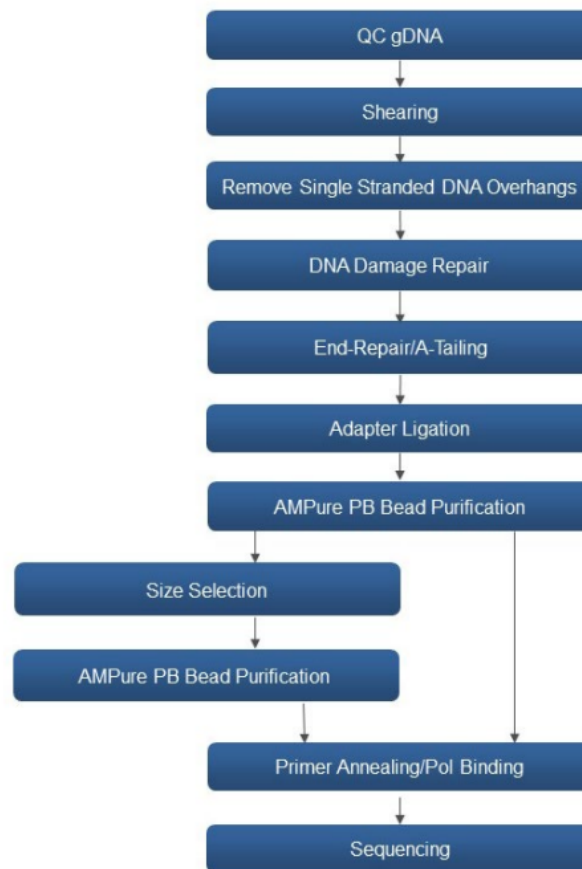
mix by inversions for 5 min. The tube is centrifuged for 10 min at $5000\times g$ at 4°C . Transfer the clear supernatant into new 2ml tube and add half volume of 5 M NaCl to the sample and mix gently by inversions. Add 2 volumes of cold 95% ethanol and mix gently by inversion. The tubes are incubated at for 45 min in -20°C freezer. It should not be for more time as some remaining phenolics and resin may also precipitate with DNA. After incubation, the tubes were centrifuged at $5000 \times g$ for 10 min at 4°C and the supernatant was gently removed. The pellet is washed two times with 1ml of 70% ethanol and the DNA is pellet by $5000 \times g$ at 4°C for only 5 min. The supernatant is discarded and the pellet is air-dried (10 min). The pellet is allowed to re-suspend in 50 μL of TE (10 mM Tris. HCl pH 8.0; 1 mM EDTA pH 8.0).

S5. DNA Quantification & Qualification

Total DNA was qualified and quantified as follows: (1) DNA purity and concentration were then examined using NanoDrop 2000; (2) Size distribution of and degradation degree DNA were measured by Pulsed-field Electrophoresis; (3) Accurate quantification of DNA were measured by Qubit Fluorometric Quantitation.

S6. Library construction and sequencing

Overview of the SMRTbell Express 2.0 Large-insert Library Workflow



The library of 15 kb was constructed using a SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, CA, USA). The construction includes **DNA shearing, damage repair, end repair, hairpin adapter ligation, size selection** and **purification of the library**. The process is as follows.

DNA shearing.

After examination of the quality of isolated DNA, dilute 2.0 µg gDNA into 200 µL 1X Elution Buffer to a final concentration of 10 ng/µL. Transfer gDNA to g-TUBE and centrifuge at 2400 x g (6000 rpm in the Eppendorf MiniSpin Plus) for 2 minutes. Then repeat spin until entire gDNA sample has passed through the orifice. (This may take 2-3 spins). Transfer the sheared gDNA to a fresh 1.5 mL Lo-bind microfuge tube. After that, concentrate DNA using AMPure PB beads to obtain the aimed DNA fragments for the construction of >15kb HiFi libraries.

DNA damage repair and end-repair/A-tailing

Use the following table to prepare the DNA damage repair reaction (Table 1). Firstly, pipette mix 10 times with wide-bore pipette tips. It is important to mix well. And spin down contents of tube with a quick spin in a microfuge. The incubate at 37°C for 30 minutes, then return the reaction to 4°C.

Use the following table to prepare the next end-repair/A-tailing processes (Table 2). Pipette mix 10 times with wide-bore pipette tips and spin down contents of tube with a quick spin in a microfuge. Then incubate at 20°C for 10 minutes and at 65°C for 30 minutes, then return the reaction to 4°C. The obtained samples are used to the next step.

Table 1 DNA damage repair reaction

Reagent (Reaction Mix 2)	Tube Cap Color	Volume	✓	Notes
Reaction Mix 1		55.0 µL		
DNA Damage Repair Mix v2		2.0 µL		
Total Volume		57.0 µL		

Table 2 End-Repair/A-tailing reaction

Reagent (Reaction Mix 3)		Volume	✓	Notes
Reaction Mix 2		57.0 µL		
End Prep Mix		3.0 µL		
Total Volume		60.0 µL		

Adapter ligation

Use the following table to prepare your reaction, adding the components below in the order listed. Pipette mix 10 times with wide-bore pipette tips, and spin down contents of tube with a quick spin in a microfuge. Then incubate at 20°C for 60 minutes, then return the reaction to 4°C.

Reagent (Reaction Mix 4)	Tube Cap Color	Volume	✓	Notes
Reaction Mix 3		60.0 µL		
Overhang Adapter v3		5.0 µL		
Ligation Mix		30.0 µL		
Ligation Additive		1.0 µL		
Ligation Enhancer		1.0 µL		
Total Volume		97.0 µL		

After quality control test, a SMRTbell library was obtained. The library was sequenced using a single 8 M SMAT Cell on the PacBio Sequel II/Sequel IIE platform (Pacific Biosciences, CA, USA).

Size selection and purification of SMRTbell Templates

When constructing large insert SMRTbell libraries for whole genome sequencing of complex organisms, it is beneficial to remove small insert SMRTbell templates by performing size selection with the BluePippin System (which collects fragments above a size cut-off threshold). And bring up the volume of eluted, size-selected DNA SMRTbell templates to 100 µL with 1X Elution Buffer to perform the purification of SMRTbell templates after size selection.

Anneal and Bind SMRTbell Library Templates and sequencing.

Use SMRT Link Sample Setup for instructions for primer annealing and polymerase binding. For primer annealing, use Sequencing Primer v4 (found in the SMRTbell Express Template Prep Kit v2). Sequencing Primer v4 is for diffusion loading only and cannot be used for MagBead loading. Wide bore pipette tips are required when constructing large insert libraries (>15 kb). The library was sequenced using a single 8 M SMAT Cell on the PacBio Sequel IIE platform (Pacific Biosciences, CA, USA).

Sequence quality checking and filtering

The PacBio SMRT-Analysis package (<https://www.pacb.com>) was used for the quality control of the raw polymerase reads; we removed the following types of polymerase reads: (1) Polymerase reads with length less than 50bp; (2) Polymerase

reads with the quality value lower than 0.8; (3) Polymerase reads containing the adaptor ligation to itself; (4) Removed the adaptor sequence in the polymerase reads. HiFi reads were generated by SMRTLink 9.0 software with parameters --min-passes=3 --min-rq=0.99.

S7. Hi-C sequencing

To anchor hybrid scaffolds onto the chromosome, genomic DNA was extracted for the Hi-C library from blood. We followed the standard protocol described previously with certain modifications (PMID: 19815776 and 25497547). In brief, we constructed the Hi-C library and obtained sequencing data via the Illumina Novaseq-6000 platform. About 1 ml blood sample was washed by PBS buffer and the pellet was resuspended and fixed in 2% formaldehyde in room temperature. Crosslinking was stopped by adding 0.125M glycine. The fixed sample was then resuspending and lysed in Hi-C lysis buffer. The purified nuclei were digested with 100 units of DpnII and marked by incubating with biotin-14-dATP. Biotin-14-dATP from non-ligated DNA ends was removed owing to the exonuclease activity of T4 DNA polymerase. The ligated DNA was sheared into 300-500 bp fragments, and then was blunt-end repaired and A-tailed, followed by purification through biotin-streptavidin-mediated pull down. Finally, the Hi-C libraries were quantified and sequenced using the Illumina Novaseq-6000 platform.

Genomic DNA preparation

DNA degradation and contamination were monitored on 1% agarose gels. DNA purity was checked using the NanoPhotometer®spectrophotometer (IMPLEN, CA, USA). DNA concentration was measured using Qubit®DNA Assay Kit in Qubit®2.0Fluorometer (LifeTechnologies, CA, USA).

Library preparation and Illumina sequencing

A total amount of 1.5µg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using Truseq Nano DNA HT Sample preparation Kit (Illumina USA) following manufacturer's

recommendations and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350bp, then DNA fragments were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR. These libraries constructed above were sequenced by Illumina NovaSeq-6000 platform and 150bp paired-end reads were generated with insert size around 350bp.

S8. High-Fidelity (HiFi) sequencing

Genomic DNA Preparation

High molecular weight genomic DNA was prepared by the CTAB method and followed by purification with QIAGEN® Genomic kit (Cat#13343, QIAGEN) for regular sequencing, according to the standard operating procedure provided by the manufacturer. The DNA degradation and contamination of the extracted DNA was monitored on 1% agarose gels. DNA purity was then detected using NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA), of which OD_{260/280} ranging from 1.8 to 2.0 and OD_{260/230} is between 2.0-2.2. At last, DNA concentration was further measured by Qubit® 4.0 Fluorometer (Invitrogen, USA).

Library preparation and sequencing

SMRTbell target size libraries were constructed for sequencing according to PacBio's standard protocol (Pacific Biosciences, CA, USA) using 15kb preparation solutions. The main steps for library preparation are: (1) gDNA shearing, (2) DNA damage repair, end repair and A-tailing, (3) ligation with hairpin adapters from the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences), (4) nuclease treatment of SMRTbell library with SMRTbell Enzyme Cleanup Kit, (5) size selection, and (6) binding to polymerase. Briefly, a total amount of 15 µg DNA per sample was used for the DNA library preparations. The genomic DNA sample was

sheared by g-TUBEs (Covaris, USA) according to the expected size of the fragments for the library. Single-strand overhangs were then removed, and DNA fragments were damage repaired, end repaired and A-tailing. Then the fragments ligated with the hairpin adaptor for PacBio sequencing. And the library was treated by nuclease with SMRTbell Enzyme Cleanup Kit and purified by AMPure PB Beads. Target fragments were screened by the BluePippin (Sage Science, USA). The SMRTbell library was then purified by AMPure PB beads, and Agilent 2100 Bioanalyzer (Agilent technologies, USA) was used to detect the size of library fragments. Sequencing was performed on a PacBio Sequel II instrument with Sequencing Primer V2 and Sequel II Binding Kit 2.0 in Grandomics.

S9. Oxford Nanopore (ONT) Sequencing

Genomic DNA preparation

Samples were collected, and high molecular weight genomic DNA was prepared by the CTAB method and followed by purification with QIAGEN® Genomic kit (Cat#13343, QIAGEN) for regular sequencing, according to the standard operating procedure provided by the manufacturer. Ultra-long DNA was extracted by the SDS method without purification step to sustain the length of DNA. The DNA degradation and contamination of the extracted DNA was monitored on 1% agarose gels. DNA purity was then detected using NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA), of which OD_{260/280} ranging from 1.8 to 2.0 and OD_{260/230} is between 2.0-2.2. At last, DNA concentration was further measured by Qubit® 4.0 Fluorometer (Invitrogen, USA).

Library preparation and sequencing

A total amount of 3-4 µg DNA per sample was used as input material for the ONT library preparations. After the sample was qualified, size-select (>50k and >100k) of long DNA fragments were performed using the PippinHT system (Sage Science, USA). Next, the ends of DNA fragments were repaired, and A-ligation reaction were conducted with NEBNext Ultra II End Repair/dA-tailing Kit

(Cat# E7546). The adapter in the SQK-LSK109 (Oxford Nanopore Technologies, UK) was used for further ligation reaction and DNA library was measured by Qubit® 4.0 Fluorometer (Invitrogen, USA).

S10. Genome Assembly and Quality Control

In our study, we collected 58 new HiFi samples covering 36 ethnic groups that represent genetic diversity in China. The 68 samples were sequenced on 2–5 SMRT Cells. The obtained subreads are converted into HiFi reads by the ccs-v6.3.0 in Pacbio tools with `--hifi-kinetics --min-passes 3 --min-length 50`. At the individual level, we combined HiFi reads from multiple cells and applied hifiasm v0.16.1² to perform both primary assembly and diploid assembly for 68 PacBio HiFi samples. As for 11 samples with Hi-C data, we ran hifiasm with following command:

```
hifiasm -o $sample.asm -h1 sample.r1.fastq.gz -h2 sample.r2.fastq.gz -t 96 $sample.ccs.fastq.gz.
```

As for the remaining 58 samples, we ran hifiasm with the command:

```
hifiasm -o $sample.asm -t 96 $sample.ccs.fastq.gz.
```

We used QUAST v5.2.0³ and inspector v1.2⁴ to assess the assembly quality of 68 primary assemblies and 136 haplotype assemblies. We removed 3 samples with primary assembly N50 < 20 Mb or contig number > 2000, and 7 samples with two haplotype assembly N50 < 10 Mb or contig number > 2000. Finally, 58 samples with 116 high-quality assemblies were retained for subsequent analyses. The assessment commands of QUAST and inspector in this study were as follows:

```
quast-lg.py $asm.fa -r $reference.fa -g $reference.gff.gz -o $asm/ --large --est-ref-size 3100000000 --no-icarus -t 64
```

```
inspector.py -c $asm.fa -r $sample.ccs.fastq.gz -o $asm/ --datatype hifi -t 64
```

S11. Alignment of Hifi reads to CHM13

The obtained hifi reads were aligned to the chm13.v2.0 reference genome using the preset parameters “- ax map-hifi” by minimap2⁵, and then sorted by

samtools sort⁶. The quality of reads in alignment-level was evaluated by samtools stats⁶. This BAM file will be used for the identification of conventional structural variation and complex structural variation.

S12. Evaluate alignment quality of short-reads using CPC reference

We selected samples from five East Asian populations of the 1000 Genomes Project (KGP) according to the standard of one male and one female per population. Fastq files of 10 samples were downloaded from ENA (study access: PRJEB31736). We first aligned the paired-end reads of each sample to the graph genomes of CPC and HPRC by VG giraffe, then converted them to linear space (CHM13) using VG surject, and then marked the duplicate reads by Picard's markduplicate. We obtained the final BAM file for each sample. Through the samtools stats, we calculated the alignment quality of each sample.

S13. Pangenome graph construction

We applied the Minigraph-Cactus pipeline (<https://github.com/ComparativeGenomicsToolkit/cactus/blob/master/doc/pangenome.md>) to construct the CPC pangenome variation graph in two version: GRCh38-based and CHM13-based. We first constructed the graph with SVs only larger than 50 bp using Minigraph (version 0.19)⁷. Starting from the reference assembly, either GRCh38 or CHM13, as the initial graph, the remaining reference genome and 122 CPC assemblies were sequentially mapped to the graph. Then, we used Minigraph-Cactus pipeline to add SNP-level variants to the graph. Firstly, centromeres and telomeres regions were identified and softmasked with dna-brnn⁸ to exclude the influence of highly repetitive sequence on subsequent analysis. Then, the assemblies were remapped to the minigraph to produce exact alignments between the contigs of input assembly and minigraph node sequences. Quality

control was applied in this step by excluding all softmasked sequences longer than 100 kb and all mappings with MAPQ<5. Next, we split the minigraph and assigned the contigs into different chromosomes, and applied Cactus base alignment⁹ separately, and the output HAL¹⁰ files were converted to the vg format with hal2vg (<https://github.com/ComparativeGenomicsToolkit/hal2vg>). Paths larger than 10kb that do not align to the underlying minigraph were removed, and GFAffix (<https://github.com/marschall-lab/GFAffix>) were used to normalize the graph. Finally, the graphs of chromosomes were combined into a whole-genome graph, indexed and exported to VCF with vg toolkit version 1.42 (<https://github.com/vgteam/vg>), and allele filtered were also applied for short-read mapping by removing all nodes traversed by fewer than 12 haplotype paths (minimum AF=10%) using `vg clip -d 12 -m 10000`.

S14. Pangenome growth

We covert the CPC graph to the vcf file by “vg deconstruct”, and calculated the haplotype depth of non-reference segments (presenting as SNPs and insertions in vcf file). The haplotype depth is defined as the frequency of non-reference sequences presenting in the assembled haplotypes. We counted the frequency of segments based on the GT information from the vcf file and calculated the segment length based on the ALT and REF sequences. We removed the super-large segments (> 3 Mb) from the results since they may be caused by graph constructing.

S15. Statistics on variants

We used vg deconstruct (<https://github.com/vgteam/vg>) to infer the variants and the genotype of assemblies from the graph, and variants with reference allele longer than 10 Mb were removed using vcfbub v1.0 (<https://github.com/pangenome/vcfbub>) because they are likely to be misjudged as deletions. Variants were classified as small variants (<50 bp) and SVs (≥ 50 bp), and then were converted to per-sample biallelic VCF files using bcftools v1.14¹¹

separately to apply individual and haplotype level statistics.

S16. Identification and visualization of novel variants

We used a 1-Mb window (step size = 500 kb) sliding along the chromosomes to identify the novel variants in CPC pangenome graph. Novel SVs (≥ 50 bp) were identified if the longest allele sequence at that site could not confidently match ($\leq 50\%$ coverage or $\leq 50\%$ identity) any HPRC variants. Similarly, small variants in CPC were determined as novel if the longest allele at that site did not overlap with any HPRC variants. We applied RIdeogram¹² to visualize the SV distribution on chromosomes. We extracted the subgraphs of SVs with gfabase v0.6.0 (<https://github.com/mlin/gfabase>) and then visualized the subgraphs with bandage v0.9.0¹³. Gene positions in the subgraphs were determined by aligning the sequences of genes to graph with GraphAligner v1.0.16¹⁴.

S17. Complex structural variation detection

We aligned HiFi reads of 58 samples against CHM13-T2T as reference by minimap2⁵ with parameter “-ax map-hifi”. We convert sam files to bam files and generate the corresponding index files by SAMtools⁶. Then we detected the complex structural variations (CSVs) by SVision¹⁵ with parameters “-s 5 --min_mapq 10 --min_sv_size 50 -qname --graph” and pretrained model of “svision-cnn-model.ckpt”. We selected CSVs (the term with <CSV> ALT column) from SVision results. To remove the coverage bias and get confidential CSV sets, we filtered CSVs with the ratio of supported read number to sequence coverage smaller than 25%. To get a nonredundant set of discovery CSVs from CSV calls from all samples, we adopted a merging strategy where HIFI032567D (with the most CSVs) served as the initial callset and new sites were added per sample. We excluded any CSVs in the sample that have 80% reciprocal overlap and the same type with an existing discovery CSV. This merging process yield 684 nonredundant CSVs. Samples were added by the descending order of CSV numbers. Each merged

CSV is represented by a single CSV call in a single sample, but it is annotated with its discovery sample along with the corresponding genomic coordinate.

S18. Alignment, variant identification and filtering

Per-individual sequence reads were aligned using ‘mem’ algorithm “bwa mem -M -R @RG\tID:name\tSM:name” in the Burrows-Wheeler Algorithm (BWA) version 0.7.17-r1188¹⁶ to the reference human genome (GRCh38), and then converted to BAM format, sorted by genomic position and indexed using SAMtools version 1.15.1⁶.

Picard toolkit version 2.21.9 (<http://broadinstitute.github.io/picard/>) was used to mark the potential duplicate reads inherited from library construction step, in which the amplified PCR errors can introduce the wrong variants in variants calling^{17,18}. The MarkDuplicates.jar in Picard was used for chromosome-wise duplicates marking per-individual.

BQSR module in the Genome Analysis Toolkit (GATK) version 4.1.7.0 was applied to reduce the base quality score bias from the sequencer, and the 1000 Genomes phase I INDELs and dbSNP database (version 151) were used as the training data sets. HaplotypeCaller module in GATK was used for SNPs and INDELs calling chromosome-wise simultaneously for each sample.

For population-based analyses, GATK GenotypeGVCFs module was applied to the GVCFs generated from previous step to call the variants for each chromosome of the combined samples.

The chromosome-wise raw variants were combined to genome-wide raw variants for population-based variants VCFs and individual-based GVCFs. GATK variants quality score recalibration (VQSR) module was used to filter the population based raw SNPs and INDELs, separately. Briefly speaking, VQSR used maximized sensitivity on these variants first, and then used some variants collections as training dataset to estimate the levels of specificity to filter these raw variants. For SNPs filtering, the variants collections contained HapMap 3.3

genotyping result, OMNI genotyping dataset, 1000 Genomes Phase I high confident SNPs and dbSNP dataset. For INDELs filtering, the variants collections contained Mills and 1000 Genomes Phase I gold standard INDELs.

S19. Population Genetic Analyses

To investigate the population diversity and coverage of CPC samples in the context of East Asia, we performed principal components analysis (PCA) at the individual level using SNPRelate v1.6.4¹⁹ and evaluated with the first two principal components. We highlighted HPRC and CPC samples with long-read sequencing data in the PC plot according to the linguistic classification.

To dissect diverse genetic components of 58 CPC core samples from different genetic background, we applied ADMIXTURE v1.3.0²⁰ to perform global ancestry inference using 36 populations corresponding to 58 CPC samples, 2 populations used in HPRC pangenome construction (CHS and KHV), and other KGP reference populations including African (YRI), South Asians (BEB, GIH, and ITU), and Europeans (CEU, FIN, and IBS). To reduce the bias of sample size, we performed ADMIXTURE analysis with 10 samples for each population. We first incorporate each HPRC and CPC sample to the corresponding population group, and further randomly incorporate remaining samples of each population to 10 samples. The ADMIXTURE analysis was replicated with 10 times from K=2 to K=12. The SNPs used in analyses were randomly selected from each 20 kb distance, and the final result was overall clustered from the results of 10 runs. The 58 CPC samples and 4 HPRC samples were labeled in the ADMIXTURE plot with short lines.

S20. Functional annotation and enrichment analysis of genes

We characterized gene functions and further conducted enrichment analysis based on several knowledgebases or databases embedded in clusterProfiler

v3.10.1²¹, including the Gene Ontology (GO)²², the Disease Ontology (DO)²³, and the Kyoto Encyclopedia of Genes and Genomes (KEGG)²⁴. For some genes of particular interest, we applied GeneAnalytics²⁵ online (geneanalytics.genecards.org) to obtain expression-based and function-based annotations. We also searched for the NHGRI-EBI Catalog of human genome-wide association studies (GWAS Catalog, <https://www.ebi.ac.uk/gwas/>) for possible phenotypic consequences of the genes.

S21. Archaic introgression segments detection and functional annotation

We performed *ArchaicSeeker* v2.010²⁶ (AS2) to identify archaic introgression segments (AIS) in the CPC and HPRC genomes. Prior to the analysis, we converted the genomic coordinates in the phased VCFs from human genome assembly GRCh38 to GRCh37 as required by AS2, using LiftoverVcf embedded in Picard toolkit version 1.117 (<http://broadinstitute.github.io/picard/>), and then removed the multi-allelic SNPs. We focused on the Altai Neanderthal-like or Denisovan-like AISs identified by AS2, and then used GFF3 (Generic Feature Format Version 3) of GRCh37 version to annotate these AIS. To detect the novel AIS in CPC compared to HPRC, we applied BEDtools²⁷ to remove the overlapping regions in the AISs identified in CPC and HPRC.

We performed functional enrichment analyses for genes affected by AISs located in the CDS regions using clusterProfiler v4.4.4²⁸, based on the Gene Ontology (GO)²² and the Kyoto Encyclopedia of Genes and Genomes (KEGG)²⁴. We set different frequency thresholds (appear in ≥ 2 samples, ≥ 5 samples, and ≥ 10 samples) for the AISs and then analyzed the affected gene sets independently. For some genes located around AISs with extremely high frequency (>40 samples), we applied GeneAnalytics²⁵ online (geneanalytics.genecards.org) to obtain expression-based and function-based annotations. We also referred to the

GeneCards database (<https://www.genecards.org/>) for possible phenotypic consequences of the genes.

S22. References

- 1 Liao, W.-W. *et al.* A Draft Human Pangenome Reference. *bioRxiv*, 2022.2007.2009.499321, doi:10.1101/2022.07.09.499321 (2022).
- 2 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175, doi:10.1038/s41592-020-01056-5 (2021).
- 3 Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142-i150, doi:10.1093/bioinformatics/bty266 (2018).
- 4 Chen, Y., Zhang, Y., Wang, A. Y., Gao, M. & Chong, Z. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biol* **22**, 312, doi:10.1186/s13059-021-02527-4 (2021).
- 5 Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, doi:10.1093/bioinformatics/btab705 (2021).
- 6 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 7 Li, H., Feng, X. & Chu, C. The design and construction of reference pangenome graphs with minigraph. *Genome Biol* **21**, 265, doi:10.1186/s13059-020-02168-z (2020).
- 8 Li, H. Identifying centromeric satellites with dna-brnn. *Bioinformatics* **35**, 4408-4410, doi:10.1093/bioinformatics/btz264 (2019).
- 9 Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246-251, doi:10.1038/s41586-020-2871-y (2020).
- 10 Hickey, G., Paten, B., Earl, D., Zerbino, D. & Haussler, D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341-1342, doi:10.1093/bioinformatics/btt128 (2013).
- 11 Danecck, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
- 12 Hao, Z. *et al.* RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the ideograms. *PeerJ Comput Sci* **6**, e251, doi:10.7717/peerj-cs.251 (2020).
- 13 Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350-3352, doi:10.1093/bioinformatics/btv383 (2015).
- 14 Rautiainen, M. & Marschall, T. GraphAligner: rapid and versatile sequence-to-graph alignment. *Genome Biol* **21**, 253, doi:10.1186/s13059-020-02157-2 (2020).
- 15 Lin, J. *et al.* SVision: a deep learning approach to resolve complex structural variants. *Nat Methods*, doi:10.1038/s41592-022-01609-w (2022).
- 16 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 17 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
- 18 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).

- 19 Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328, doi:10.1093/bioinformatics/bts606 (2012).
- 20 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).
- 21 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 22 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
- 23 Schriml, L. M. *et al.* Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res* **40**, D940-946, doi:10.1093/nar/gkr972 (2012).
- 24 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).
- 25 Ben-Ari Fuchs, S. *et al.* GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data. *OMICS* **20**, 139-151, doi:10.1089/omi.2015.0168 (2016).
- 26 Yuan, K. *et al.* Refining models of archaic admixture in Eurasia with ArchaicSeeker 2.0. *Nat Commun* **12**, 6232, doi:10.1038/s41467-021-26503-5 (2021).
- 27 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 28 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

s23. Supplementary Tables

Table S1. Information of the samples for sequencing in CPC

Sample ID	Population	Gender	Sequencing Platform	Latitude	Longitude	Source
HIF032682D	Achang	M	PacBio HiFi, NGS	24.43	98.59	CPC
HIF032585D	Bai	M	PacBio HiFi, NGS	26.11	99.95	CPC
HIF032069D	Blang	F	PacBio HiFi, NGS	21.96	100.45	CPC
HIF032373D	Blang	M	PacBio HiFi, NGS	21.96	100.45	CPC
HIF032487D	Bouyei	F	PacBio HiFi, NGS	24.98	105.81	CPC
HIF032018D	Chosen	M	PacBio HiFi, NGS	42.91	129.51	CPC
HIF032306D	Chosen	F	PacBio HiFi, NGS	42.91	129.51	CPC
RY05-CNCM0001	Chosen	M	PacBio HiFi, ONT, Hi-C, NGS	43.12	128.91	CPC
HIF032668D	Daur	M	PacBio HiFi, tell-seq, NGS	49.21	119.74	CPC
HIF032698D	Daur	F	PacBio HiFi, NGS	49.21	119.74	CPC
HIF032292D	Deang	F	PacBio HiFi, NGS	24.43	98.59	CPC
HIF032462D	Deang	M	PacBio HiFi, NGS	24.43	98.59	CPC
HIF032473D	Dong	M	PacBio HiFi, NGS	25.78	109.60	CPC
HIF032510D	Dong	F	PacBio HiFi, NGS	25.78	109.60	CPC
HIF032450D	Drung	M	PacBio HiFi, NGS	27.74	98.67	CPC
HIF032706D	Evenki	M	PacBio HiFi, NGS	49.21	119.74	CPC
HIF032007D	Kinh	F	PacBio HiFi, NGS	21.55	107.97	CPC
RY06-CNKG0001	Kyrgyz	F	PacBio HiFi, ONT, Hi-C, NGS	39.15	75.95	CPC
RY07-CNKF0001	Kazakh	F	PacBio HiFi, ONT, Hi-C, NGS	44.91	82.07	CPC
RY08-CNZH0001	Zhuang	F	PacBio HiFi, ONT, Hi-C, NGS	23.74	106.92	CPC
RY09-CNHUM0001	Hui	M	PacBio HiFi, ONT, Hi-C, NGS	39.02	106.37	CPC
HIF032513D	Kinh	M	PacBio HiFi, NGS	21.55	107.97	CPC
HIF032685D	Hezhen	M	PacBio HiFi, NGS	47.64	132.51	CPC
HIF032440D	Jingpo	M	PacBio HiFi, NGS	24.43	98.59	CPC
HIF032429D	Jino	M	PacBio HiFi, NGS	22.01	100.80	CPC
HIF032731D	Lisu	F	PacBio HiFi, NGS	25.85	98.85	CPC
HIF032289D	Man	M	PacBio HiFi, NGS	40.28	123.29	CPC
HIF032167D	Miao	M	PacBio HiFi, NGS	28.17	109.20	CPC
RY04-CNMHM0001	Miao	M	PacBio HiFi, ONT, Hi-C, NGS	26.58	109.70	CPC
HIF032164D	Mongol	M	PacBio HiFi, NGS	39.82	109.96	CPC
RY10-CNMG0001	Mongol	F	PacBio HiFi, ONT, Hi-C, NGS	41.60	119.35	CPC
HIF032607D	Mongol	F	PacBio HiFi, NGS	41.70	110.43	CPC
HIF032604D	Mongol-TH	F	PacBio HiFi, NGS	24.11	102.77	CPC
HIF032161D	Mosuo	F	PacBio HiFi, NGS	27.28	100.85	CPC
HIF032422D	Mosuo	M	PacBio HiFi, tell-seq, NGS	27.28	100.85	CPC
HIF032501D	Mosuo	M	PacBio HiFi, NGS	27.28	100.85	CPC
HIF032693D	Naxi	M	PacBio HiFi, tell-seq, NGS	26.82	100.24	CPC
HIF032335D	Orogen	U*	PacBio HiFi, tell-seq, NGS	50.60	123.73	CPC
HIF032420D	Qiang	M	PacBio HiFi, NGS	31.68	103.85	CPC
HIF032567D	Qiang	F	PacBio HiFi, NGS	31.68	103.85	CPC
HIF032662D	Qiang	M	PacBio HiFi, tell-seq, NGS	31.68	103.85	CPC
HIF032453D	Salar	M	PacBio HiFi, NGS	35.85	102.49	CPC
HIF032586D	Salar	F	PacBio HiFi, NGS	35.85	102.49	CPC
HIF032302D	She	F	PacBio HiFi, NGS	27.09	119.65	CPC
HIF032591D	She	M	PacBio HiFi, NGS	27.09	119.65	CPC
RY03-CNSHM0001	She	M	PacBio HiFi, ONT, Hi-C, NGS	22.73	114.25	CPC
HIF032349D	Tu	M	PacBio HiFi, NGS	36.84	101.96	CPC
HIF032454D	Tu	F	PacBio HiFi, NGS	36.84	101.96	CPC
HIF032097D	Tujia	F	PacBio HiFi, NGS	29.49	109.41	CPC
HIF032103D	Tujia	M	PacBio HiFi, NGS	29.49	109.41	CPC
RY11-CNTJM0001	Tujia	M	PacBio HiFi, ONT, Hi-C, NGS	30.01	108.12	CPC
HIF032529D	Wa	F	PacBio HiFi, NGS	22.65	99.60	CPC
HIF032711D	Wa	M	PacBio HiFi, NGS	22.65	99.60	CPC
HIF032182D	Yi	M	PacBio HiFi, NGS	26.82	100.24	CPC
HIF032566D	Yugur	M	PacBio HiFi, NGS	38.84	99.61	CPC
HIF032692D	Yugur	F	PacBio HiFi, NGS	38.84	99.61	CPC
RY01-CNYAM0001	Yao	M	PacBio HiFi, ONT, Hi-C, NGS	24.82	111.28	CPC
RY02-CNUGM0001	Uyghur	M	PacBio HiFi, ONT, Hi-C, NGS	39.47	76.00	CPC
Han	Han-N	M	ONT, NGS	31.81	117.22	CPC
LX	Han-S	M	ONT, NGS	23.79	108.77	CPC
HZZ1	Han-C	F	ONT, NGS	34.76	113.66	CPC
HZZ2	Han-C	M	ONT, NGS	34.76	113.66	CPC
Xinjiang	Uyghur	M	ONT, NGS	41.21	85.33	CPC
Zang	Tibetan	M	ONT, NGS	31.68	88.17	CPC
Mos	Mosuo	M	ONT, NGS	27.72	100.79	CPC
HG00438	Han-S	F	PacBio HiFi, ONT, Hi-C, BioNano, NGS	23.14	113.23	HPRC
HG00621	Han-S	M	PacBio HiFi, ONT, Hi-C, BioNano, NGS	23.14	113.23	HPRC
HG00673	Han-S	M	PacBio HiFi, ONT, Hi-C, BioNano, NGS	23.14	113.23	HPRC
HX1	Han-S	M	PacBio CLR, BioNano, 10X Genomics, NGS	23.79	108.77	Shi et al., 2016
NH1	Han-N	M	PacBio CLR, BioNano, NGS	36.77	117.05	Du et al., 2019
TJ1	Tujia	M	PacBio CLR, BioNano, Hi-C, 10X Genomics, NGS	30.65	111.31	Lou et al., 2022
ZF1	Tibetan	M	PacBio CLR, BioNano, Hi-C, 10X Genomics, NGS	29.28	88.90	Ouzhuluobu et al., 2020

Table S2 Summary statistics of the 116 CPC genome assemblies

Genome	Ethnicity	Contig N50 (Mb)	Contig Number	Assembly Size (Gb)
HIFI032097D-H1	Tujia	49.18	470	3.02
HIFI032097D-H2	Tujia	33.25	484	3.04
HIFI032103D-H1	Tujia	33.10	644	3.02
HIFI032103D-H2	Tujia	30.71	699	3.02
RY11-CNTJM0001-H1	Tujia	42.61	718	2.89
RY11-CNTJM0001-H2	Tujia	40.81	579	3.12
RY10-CNMGF0001-H1	Mongol	42.62	533	3.06
RY10-CNMGF0001-H2	Mongol	43.44	437	3.01
HIFI032164D-H1	Mongol	20.94	845	3.03
HIFI032164D-H2	Mongol	21.57	801	3.03
HIFI032604D-H1	Mongol-TH	42.74	541	3.02
HIFI032604D-H2	Mongol-TH	27.40	528	3.02
RY04-CNMHM0001-H1	Miao	87.30	461	2.90
RY04-CNMHM0001-H2	Miao	81.84	397	3.06
HIFI032167D-H1	Miao	17.60	1192	3.00
HIFI032167D-H2	Miao	11.71	1269	3.00
RY01-CNYAM0001-H1	Yao	32.31	620	3.06
RY01-CNYAM0001-H2	Yao	39.83	496	2.93
RY03-CNSHM0001-H1	She	52.23	585	2.88
RY03-CNSHM0001-H2	She	40.08	476	3.07
HIFI032302D-H1	She	26.21	733	3.01
HIFI032302D-H2	She	28.69	681	3.02
HIFI032513D-H1	Gin	36.39	690	3.01
HIFI032513D-H2	Gin	25.28	712	3.02
HIFI032007D-H1	Gin	23.40	962	3.02
HIFI032007D-H2	Gin	13.88	1106	3.02
HIFI032018D-H1	Chosen	35.60	813	2.98
HIFI032018D-H2	Chosen	28.86	752	3.01
HIFI032306D-H1	Chosen	39.26	590	3.03
RY05-CNCSM0001-H1	Chosen	58.98	533	3.06
RY05-CNCSM0001-H2	Chosen	83.61	451	2.89
HIFI032306D-H2	Chosen	25.23	644	3.01
HIFI032692D-H1	Yugur	44.18	613	3.04
HIFI032692D-H2	Yugur	40.25	549	3.04
HIFI032566D-H1	Yugur	39.04	505	2.98
HIFI032566D-H2	Yugur	36.57	517	3.00
RY02-CNUGM0001-H1	Uyghur	31.63	643	2.90
RY02-CNUGM0001-H2	Uyghur	24.85	555	3.07
HIFI032069D-H1	Blang	37.33	681	3.02
HIFI032069D-H2	Blang	27.19	761	3.01
HIFI032161D-H1	Mosuo	29.69	703	3.03
HIFI032161D-H2	Mosuo	24.78	749	3.02
HIFI032182D-H1	Yi	45.14	493	3.04
HIFI032182D-H2	Yi	32.50	485	2.99
HIFI032289D-H1	Man	57.76	502	3.02
HIFI032289D-H2	Man	37.28	546	3.00
HIFI032292D-H1	Deang	19.43	1084	3.02
HIFI032292D-H2	Deang	16.80	1082	3.01
HIFI032335D-H1	Oroqen	17.67	1286	3.05
HIFI032335D-H2	Oroqen	12.54	1568	3.04
HIFI032349D-H1	Tu	19.03	982	3.03
HIFI032349D-H2	Tu	25.41	782	3.00
HIFI032454D-H1	Tu	27.39	728	3.03
HIFI032454D-H2	Tu	24.78	642	3.02
HIFI032373D-H1	Blang	38.87	691	3.00
HIFI032373D-H2	Blang	23.97	680	3.01
HIFI032420D-H1	Qiang	36.22	484	3.04
HIFI032420D-H2	Qiang	38.83	434	2.97
HIFI032422D-H1	Mosuo	21.66	824	3.02
HIFI032422D-H2	Mosuo	14.36	926	3.02
HIFI032429D-H1	Jino	26.67	877	3.01
HIFI032429D-H2	Jino	17.02	843	3.01
HIFI032440D-H1	Jingpo	33.69	568	3.04
HIFI032440D-H2	Jingpo	36.14	519	2.97
HIFI032450D-H1	Drung	59.52	559	3.03
HIFI032450D-H2	Drung	40.89	565	3.02

Genome	Ethnicity	Contig N50 (Mb)	Contig Number	Assembly Size (Gb)
HIFI032453D-H1	Salar	51.43	458	3.02
HIFI032453D-H2	Salar	49.27	389	2.97
HIFI032462D-H1	Deang	19.76	1039	3.04
HIFI032462D-H2	Deang	11.66	1075	3.00
HIFI032473D-H1	Dong	43.19	593	3.01
HIFI032473D-H2	Dong	29.94	558	3.03
HIFI032487D-H1	Bouyei	41.02	881	3.03
HIFI032487D-H2	Bouyei	16.29	955	3.00
HIFI032501D-H1	Mosuo	29.87	763	3.01
HIFI032501D-H2	Mosuo	22.31	787	3.02
HIFI032510D-H1	Dong	28.41	807	3.03
HIFI032510D-H2	Dong	22.02	705	3.02
HIFI032529D-H1	Wa	33.34	640	3.02
HIFI032529D-H2	Wa	33.83	615	3.00
HIFI032567D-H1	Qiang	30.63	862	3.04
HIFI032567D-H2	Qiang	14.68	986	3.01
HIFI032585D-H1	Bai	34.36	893	3.00
HIFI032585D-H2	Bai	20.79	818	3.03
HIFI032586D-H1	Salar	44.02	577	3.03
HIFI032586D-H2	Salar	27.70	639	3.03
HIFI032591D-H1	She	24.11	848	3.05
HIFI032591D-H2	She	21.87	901	3.00
HIFI032607D-H1	Mongol	38.11	497	3.04
HIFI032607D-H2	Mongol	32.01	515	3.02
HIFI032662D-H1	Qiang	34.99	1164	3.04
HIFI032662D-H2	Qiang	18.74	1128	2.97
HIFI032668D-H1	Daur	23.52	817	3.01
HIFI032668D-H2	Daur	24.92	749	3.02
HIFI032682D-H1	Achang	37.94	671	3.01
HIFI032682D-H2	Achang	23.29	690	2.99
HIFI032685D-H1	Hezhen	19.23	1004	3.01
HIFI032685D-H2	Hezhen	26.91	784	3.04
HIFI032693D-H1	Naxi	86.68	508	3.06
HIFI032693D-H2	Naxi	70.62	316	2.93
HIFI032698D-H1	Daur	37.78	561	3.04
HIFI032698D-H2	Daur	37.73	536	3.02
HIFI032706D-H1	Evenki	38.47	698	3.00
HIFI032706D-H2	Evenki	40.79	588	3.02
HIFI032711D-H1	Wa	57.22	550	3.02
HIFI032711D-H2	Wa	42.91	498	2.99
HIFI032731D-H1	Lisu	25.51	757	3.02
HIFI032731D-H2	Lisu	23.61	740	3.04
RY06-CNKGf0001-H1	Kyrgyz	68.01	399	3.00
RY06-CNKGf0001-H2	Kyrgyz	69.00	294	3.04
RY07-CNkZF0001-H1	Kazakh	51.74	529	3.01
RY07-CNkZF0001-H2	Kazakh	69.84	415	3.05
RY08-CNzHF0001-H1	Zhuang	52.55	521	3.03
RY08-CNzHF0001-H2	Zhuang	52.72	463	3.03
RY09-CNhUM0001-H1	Hui	51.71	563	2.88
RY09-CNhUM0001-H2	Hui	49.71	487	3.07

Table S3 A comparison of variant calling based on linear genome reference and graph genome reference

Sample ID	Population	Gender	NGS				TGS	
			#SNV	#Indel	#Small	#SV	#Small	#SV
HIFI032682D	Achang	M	3,981,870	934,745	4,916,615	NA	4,599,222	26,541
HIFI032585D	Bai	M	3,946,080	903,023	4,849,103	NA	4,578,607	26,513
HIFI032069D	Blang	F	3,967,098	914,684	4,881,782	6,653	4,606,504	26,705
HIFI032373D	Blang	M	3,945,193	922,679	4,867,872	6,485	4,571,708	26,631
HIFI032487D	Bouyei	F	3,951,056	909,839	4,860,895	6,434	4,557,349	26,489
HIFI032018D	Chosen	M	3,958,448	922,895	4,881,343	6,958	4,585,641	26,538
HIFI032306D	Chosen	F	3,956,333	919,731	4,876,064	6,442	4,599,628	26,934
HIFI032668D	Daur	M	3,941,864	919,041	4,860,905	NA	4,562,386	26,384
HIFI032698D	Daur	F	4,004,782	922,601	4,927,383	NA	4,631,431	27,105
HIFI032292D	Deang	F	3,971,930	913,793	4,885,723	6,616	4,644,571	26,661
HIFI032462D	Deang	M	3,956,954	909,444	4,866,398	6,392	4,602,115	26,413
HIFI032473D	Dong	M	3,930,632	913,711	4,844,343	NA	4,584,493	26,871
HIFI032510D	Dong	F	3,953,361	923,789	4,877,150	6,733	4,602,817	26,869
HIFI032450D	Drung	M	3,869,859	897,694	4,767,553	6,420	4,487,380	26,093
HIFI032706D	Evenki	M	3,956,037	926,899	4,882,936	6,602	4,601,577	27,049
HIFI032007D	Kinh	F	3,964,429	901,121	4,865,550	6,264	4,627,015	27,046
HIFI032513D	Kinh	M	3,945,516	919,825	4,865,341	6,344	4,582,526	26,533
HIFI032685D	Hezhen	M	3,932,237	924,893	4,857,130	8,154	4,584,970	26,739
HIFI032440D	Jingpo	M	3,964,992	919,729	4,884,721	6,860	4,589,200	26,329
HIFI032429D	Jino	M	3,894,987	883,763	4,778,750	6,317	4,529,639	26,174
HIFI032731D	Lisu	F	3,938,758	911,987	4,850,745	NA	4,566,520	26,344
HIFI032289D	Man	M	3,933,484	913,460	4,846,944	6,779	4,630,959	26,889
HIFI032167D	Miao	M	3,943,574	914,567	4,858,141	6,428	4,578,903	26,449
HIFI032164D	Mongol	M	3,971,878	919,973	4,891,851	6,491	4,604,914	26,844
HIFI032607D	Mongol	F	3,989,424	920,192	4,909,616	6,291	4,653,140	27,110
HIFI032604D	Mongol-TH	F	3,991,529	921,907	4,913,436	NA	4,647,405	26,859
HIFI032161D	Mosuo	F	3,986,438	921,085	4,907,523	6,550	4,613,822	27,035
HIFI032422D	Mosuo	M	3,962,229	913,173	4,875,402	6,571	4,587,426	26,332
HIFI032501D	Mosuo	M	3,941,005	908,076	4,849,081	6,609	4,570,766	26,674
HIFI032693D	Naxi	M	3,973,593	923,681	4,897,274	NA	4,570,682	27,079
HIFI032335D	Oroqen	U*	3,975,892	933,232	4,909,124	6,800	4,584,312	26,710
HIFI032420D	Qiang	M	3,962,567	923,202	4,885,769	6,498	4,594,111	26,676
HIFI032567D	Qiang	F	3,963,390	918,122	4,881,512	NA	4,613,808	26,786
HIFI032662D	Qiang	M	3,939,887	922,339	4,862,226	NA	4,575,002	26,367
HIFI032453D	Salar	M	3,985,636	924,962	4,910,598	6,923	4,650,093	26,809
HIFI032586D	Salar	F	4,009,785	940,648	4,950,433	NA	4,647,321	27,275
HIFI032302D	She	F	3,994,318	934,364	4,928,682	6,999	4,599,200	26,811
HIFI032591D	She	M	3,951,041	908,420	4,859,461	6,225	4,584,528	26,403
HIFI032349D	Tu	M	3,927,367	916,009	4,843,376	8,645	4,615,686	26,942
HIFI032454D	Tu	F	3,990,219	920,184	4,910,403	6,773	4,657,164	27,089
HIFI032097D	Tujia	F	3,975,408	920,010	4,895,418	6,805	4,631,580	26,835
HIFI032103D	Tujia	M	3,961,654	917,803	4,879,457	6,747	4,571,844	26,623
HIFI032529D	Wa	F	3,967,728	929,913	4,897,641	6,604	4,602,395	26,593
HIFI032711D	Wa	M	3,926,544	908,897	4,835,441	6,613	4,548,274	26,349
HIFI032182D	Yi	M	3,951,309	920,145	4,871,454	6,843	4,579,802	26,744
HIFI032566D	Yugur	M	3,987,602	918,758	4,906,360	NA	4,629,655	26,820
HIFI032692D	Yugur	F	4,033,224	945,721	4,978,945	6,826	4,667,689	27,293

Table S4 CPC-specific-CNV-related genes compared to HPRC assemblies

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
KCNJ18	115	x	x	GOLGA6L4	11	x	o	H2AB1	5	x	x
MTRNR2L6	115	x	x	HNRNPCL2	11	x	x	H3-2	5	x	o
OR8U1	101	x	x	LIMS4	11	x	o	HBA2	5	x	x
FRG2C	74	x	x	USP17L24	11	x	x	MRPL20	5	x	x
GOLGA8G	56	x	x	PRAMEF15	10	x	x	PRAMEF2	5	x	o
POTEH	56	x	x	RIMBP3C	10	x	x	PRAMEF26	5	x	x
FOXD4L4	51	x	x	TBC1D3E	10	x	x	PRAMEF4	5	x	o
OR2A7	51	x	x	TRIM49	10	x	x	PRPF31	5	x	x
PGA4	49	x	x	TRIM49D2	10	x	x	SPDYE13	5	x	o
GOLGA6B	46	x	x	ELOA3D	9	x	x	SYT15	5	x	o
GAGE12E	43	x	x	GOLGA8N	9	x	o	TAF11L13	5	x	x
UPK3BL2	42	x	x	GOLGA8O	9	x	x	TAF11L3	5	x	x
OR4F4	40	x	x	GOLGA8R	9	x	x	TAF11L6	5	x	x
POLR2J2	40	x	x	H3C13	9	x	x	TAF11L7	5	x	x
GOLGA6L6	39	x	x	H3C14	9	x	x	TAF11L8	5	x	x
OR4K5	37	x	x	OR4N4	9	x	x	TBC1D3K	5	x	x
CTAGE4	36	x	x	OR52E5	9	x	x	TP53TG3D	5	x	x
SPANXA2	35	x	x	TAF11L2	9	x	x	TSPY1	5	x	o
OR2A42	34	x	x	USP17L8	9	x	x	USP17L17	5	x	x
RFPL4AL1	34	x	x	CBWD3	8	x	x	USP17L19	5	x	x
CCL4L2	33	x	x	GAGE12H	8	x	x	ANKRD20A2P	4	x	x
CLEC18B	33	x	x	H2AC19	8	x	x	AURKAIP1	4	x	x
CTAGE8	33	x	x	LRRC37A2	8	x	x	BOLA1	4	o	x
NPIPB15	33	x	x	OR4C46	8	x	x	C2orf92	4	x	x
NBPF19	30	x	o	OR52E4	8	x	x	CD99	4	x	x
DEFB103B	28	x	x	POTEB3	8	x	x	CHRFAM7A	4	x	o
SLX1B	28	x	x	SERF1B	8	x	x	CRLF2	4	x	x
XAGE1B	28	x	x	TAF11L10	8	x	x	CSAG3	4	x	x
GAGE12B	27	x	x	TCP11X2	8	x	x	CT45A3	4	x	x
OR4M1	27	x	o	USP17L18	8	x	x	CTAG2	4	x	x
USP17L7	27	x	x	CT45A5	7	x	x	DGCR6	4	x	x
NOTCH2NLA	24	x	x	CTAG1A	7	x	o	GOLGA6C	4	x	x
OR11H1	24	x	x	ELOA2	7	x	x	GPR148	4	x	x
USP17L23	24	x	x	ELOA3	7	x	x	HNRNPCL3	4	x	x
POTEM	22	x	x	ELOA3B	7	x	x	KIR2DL3	4	x	x
CFC1	21	x	o	F8A1	7	x	x	MAGEA6	4	x	x
LGALS9C	21	x	x	GAGE2A	7	x	x	NBPF15	4	x	x
NPIPA9	21	x	x	H2AC20	7	x	x	OR4A5	4	x	x
SPAG11B	21	x	x	H2AC21	7	x	x	OR4C15	4	x	x
SPDYE11	21	x	x	MAGED4B	7	x	x	OR4K15	4	x	x
MBD3L3	19	x	x	NPIPA8	7	x	x	OR51A2	4	x	x
POTEG	19	x	x	NTAN1	7	x	o	OR5W2	4	x	x
USP17L22	19	x	x	PPIAL4D	7	x	x	PGA5	4	x	x
BOLA2-SMG1P6	18	x	x	PRAMEF10	7	x	o	POTEJ	4	x	x
DEFA1B	18	x	x	PSAPL1	7	x	x	RBMV1E	4	x	o
NPIPA7	18	x	x	RRN3	7	x	o	SCG5	4	x	x
TRIM64B	18	x	x	SSU72P2	7	x	x	SMN2	4	x	x
GOLGA8F	17	x	x	SSU72P5	7	x	x	SSX2	4	x	o
PNMA6A	17	x	o	TRIM77	7	x	x	TAF11L9	4	x	x
TBC1D3B	17	x	x	USP17L3	7	x	x	USP17L13	4	x	x
AMY1C	16	x	x	ABC7-42404400C24.1	6	x	o	ZAP70	4	x	x
CTAGE6	16	x	x	ACTR1B	6	x	x	ZNF468	4	x	x
ELOA3CP	16	x	x	CEL	6	x	x	ZNF92	4	x	x
GAGE12G	16	x	x	CENPVL2	6	x	x	AC007326.13	3	x	x
HCAR2	16	x	x	COX5B	6	x	x	ADGRE1	3	x	x
OR4M2B	16	x	x	FAM25C	6	x	x	ALKBH4	3	x	x
ZDHHC11B	16	x	x	H2BC21	6	x	x	ANKRD20A4P	3	x	o
POLR2J3	15	x	x	MBD3L2B	6	x	x	ANKRD36B	3	o	x
SPANXB1	15	x	x	NPIPA2	6	x	x	ANKRD65	3	x	x
USP17L20	15	x	x	NPIPB7	6	x	x	C4B	3	x	x
TBC1D3G	14	x	x	NXF2B	6	x	x	C8orf33	3	x	x
FCGR3B	13	x	x	PRAMEF8	6	x	o	CATSPER2	3	x	o
GOLGA8K	13	x	x	PWWP4	6	x	x	CBWD5	3	x	x
NPIPB12	13	x	o	SPATA31D4	6	x	x	DDTL	3	x	o
EIF3C	12	x	o	TAF11L4	6	x	x	DEFA5	3	x	o
FOXD4L3	12	x	x	TBC1D3H	6	x	x	ERV3-1	3	x	o
GSTT4	12	x	o	TMEM191B	6	x	o	FAHD2A	3	x	o
H4C15	12	x	x	ZNF658	6	x	o	FAM110C	3	x	x
OR11H12	12	x	x	ARHGAP11A	5	x	x	FAM178B	3	x	o
OR4M2	12	x	x	CCNL2	5	x	x	FAM236B	3	x	x
OR4N2	12	x	o	CXorf51B	5	x	x	FAM236D	3	x	x
OR4N4C	12	x	x	FAHD2B	5	x	o	FAM246A	3	x	x
AMY1B	11	x	x	GAGE12F	5	x	x	FAM83G	3	x	x
FAM90A23P	11	x	x	GOLGA6L9	5	x	x	FAM86B2	3	x	x

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
FAM86B2	3	x	x	ASL	2	x	x	MTRNR2L8	2	x	x
FCGR3A	3	x	x	ATAD3C	2	x	x	MYO19	2	x	x
FOXD4L6	3	x	x	BFAR	2	x	o	NAALAD2	2	x	x
GRAPL	3	x	x	BIRC7	2	x	x	NBPF11	2	x	x
GTPBP6	3	x	x	BSPH1	2	x	o	NBPF4	2	x	x
GZMM	3	x	x	BTN3A1	2	x	x	NDUFAF8	2	x	o
HIC2	3	x	o	C17orf99	2	x	x	NECAP2	2	o	x
HSPA6	3	x	x	CBWD6	2	o	x	NMB	2	x	o
IZUMO2	3	x	o	CCDC116	2	x	x	NOP53	2	x	x
LILRA5	3	x	x	CCDC74B	2	x	x	NOTCH2NLC	2	x	o
LIMS3	3	x	o	CCR5	2	x	x	NUPR1	2	x	o
LRWD1	3	x	x	CCRL2	2	x	x	NUTM2B	2	x	x
MAGEA2	3	x	x	CDC34	2	x	x	OR11H2	2	x	x
MAGEA3	3	x	x	CDK11B	2	x	o	OR2A4	2	x	x
MAGEH1	3	x	x	CHORDC1	2	x	x	OR2F1	2	x	x
MPV17L2	3	x	x	CLDN23	2	x	x	OR2F2	2	x	x
NDUFA3	3	x	x	CLEC4G	2	x	x	OR2V1	2	x	x
OR10AG1	3	x	x	CLN3	2	x	o	OR2V2	2	x	x
OR2T29	3	x	x	CNNM3	2	x	o	OR4K17	2	x	x
OR2Z1	3	x	x	CNNM4	2	x	o	OR56A3	2	x	x
OR4A15	3	x	x	CRX	2	x	o	OR56A5	2	x	x
OR4A16	3	x	x	CRYM	2	x	x	OR5AS1	2	x	x
OR4C12	3	x	x	CSAG1	2	x	x	OR5J2	2	x	x
OR4C13	3	x	x	CT47A10	2	x	x	OR5T2	2	x	x
OR4C16	3	x	x	CT55	2	x	o	OR5T3	2	x	x
OR4K13	3	x	x	CYP21A2	2	x	o	OR6B1	2	x	x
OR4K14	3	x	x	CYP2C9	2	x	x	OR8H2	2	x	x
OR4L1	3	x	x	DAP3	2	x	x	OR8H3	2	x	x
OR5D13	3	x	x	DDIT4L	2	x	x	OR8I2	2	x	x
OR5D14	3	x	x	DEFA4	2	x	x	OR8J3	2	x	x
OR5D16	3	x	x	DEFA6	2	x	x	OR8K5	2	x	x
OR5D18	3	x	x	DEFB1	2	x	x	PCDHB16	2	x	o
OR5D3P	3	x	x	DHFR2	2	x	x	PCDHB8	2	x	x
OR5F1	3	x	x	DHRS11	2	x	x	PDXDC1	2	x	o
OR5I1	3	x	x	DHRS2	2	x	x	PGBD2	2	x	x
OR5L1	3	x	x	DHRS4L2	2	x	x	PIAS2	2	x	x
OR5L2	3	x	x	DMRTC1B	2	x	o	PIGW	2	x	x
ORA12	3	x	x	EHD2	2	x	x	PRRT4	2	x	x
OSCAR	3	x	x	EIF2AK3	2	x	o	PTCHD1	2	x	x
PRAG1	3	x	x	EIF3CL	2	x	o	RIMBP3	2	x	x
PRAMEF27	3	x	o	ELSPBP1	2	x	x	RNPC3	2	x	x
PRAMEF9	3	x	x	EMB	2	x	o	RP1-321E8.5	2	x	x
PRODH	3	x	x	EML2	2	x	x	RP11-315A19.1	2	x	o
RP11-69H14.10	3	x	x	EPHA8	2	x	x	RP11-353J17.5	2	x	o
RPIA	3	x	o	ETDC	2	x	x	RP11-435I10.4	2	x	o
SCGB1C1	3	x	x	FAM72B	2	x	o	RP11-736I24.5	2	x	x
SEMA4C	3	x	o	FAM72C	2	x	o	RP4-583P15.14	2	x	x
SLC16A8	3	x	o	FAM72D	2	x	x	RP4-583P15.15	2	x	x
SLC5A10	3	x	o	FAM90A9P	2	x	x	RPL35A	2	x	x
SMPDL3B	3	x	x	FAM9A	2	x	x	RPS10	2	x	x
SPDYE6	3	x	x	FAN1	2	x	x	RPS12	2	x	x
SULT2A1	3	x	o	FGFBP1	2	x	x	RPS17	2	x	x
TAF11L1	3	x	x	FOLH1	2	x	x	RPS9	2	x	x
TBC1D3D	3	x	x	FOXD4L5	2	x	x	SCCPDH	2	x	x
TBC1D3L	3	x	x	FOXI3	2	x	o	SDF2L1	2	x	x
TCAF2C	3	x	x	GGNBP2	2	x	x	SEC22B	2	x	x
TFPT	3	x	x	GGT2	2	x	x	SELENOW	2	x	o
TNFRSF6B	3	x	x	GOLGA6L22	2	x	x	SGCA	2	x	x
TRIM43B	3	x	x	GSPT2	2	x	x	SH2D2A	2	x	x
TRIM49C	3	x	x	GUSB	2	x	x	SHOX	2	x	x
TRIM51GP	3	x	x	H2AB2	2	x	x	SLC2A4RG	2	x	x
TSEN34	3	x	o	H3Y1	2	x	x	SPANXN5	2	x	x
USP17L10	3	x	x	HNRNPCL1	2	x	o	SPDYE21	2	x	x
USP17L25	3	x	x	HSFX2	2	x	o	SSU72P3	2	x	x
USP51	3	x	o	IGHV3OR16-17	2	x	x	SSU72P4	2	x	x
VAMP7	3	x	x	IGSF6	2	x	o	STH	2	x	x
XKR8	3	x	x	IL27	2	x	o	STRC	2	x	o
YTHDF1	3	x	x	IL9R	2	x	x	SYNGR2	2	x	x
ZNF117	3	x	x	KANSL1	2	x	x	TAF11L12	2	x	x
ZNF14	3	x	x	LILRB3	2	x	x	TAF11L14	2	x	x
ACTL9	2	x	x	LILRB5	2	x	o	TBC1D3I	2	x	o
AMY2B	2	x	x	LIME1	2	x	x	TEX37	2	x	o
ANKRD36	2	x	o	LMAN2L	2	x	o	TEX38	2	x	x
ANKRD39	2	x	x	MAGEA12	2	x	x	TMC8	2	x	x
ANKS4B	2	x	x	MAGEA2B	2	x	x	TMEM106B	2	x	x
ANTXR1	2	x	o	MAPT	2	x	x	TMEM131	2	x	x
ANXA8L1	2	x	x	MRM1	2	x	x	TMSB15A	2	x	x
APOBR	2	x	o	MRM3	2	x	x	TPRX1	2	x	o
ARFRP1	2	x	x	MT4	2	x	x	TPSB2	2	x	o
ARHGAP11B	2	x	x	MTMR10	2	x	x	TRIM49B	2	x	x

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
TRIM64C	2	x	x	BTN3A2	1	x	x	COL1A1	1	x	x
TRIML1	2	x	o	BTN3A3	1	x	x	COL6A1	1	x	o
USP17L1	2	x	x	C11orf42	1	x	x	COL6A2	1	x	o
USP17L11	2	x	x	C13orf46	1	x	x	COL9A3	1	x	x
USP17L15	2	x	x	C15orf39	1	x	x	COX7A1	1	x	x
USP17L4	2	x	x	C17orf64	1	x	x	CPLANE2	1	x	x
VWA1	2	x	x	C1QTNF12	1	x	x	CRCP	1	x	x
VWDE	2	x	x	C2	1	x	x	CRHR1	1	x	x
WDR73	2	x	o	C22orf39	1	x	x	CRIP1	1	x	o
YDJC	2	x	x	C2CD4C	1	x	x	CRIP2	1	x	o
YY1AP1	2	x	x	C4orf48	1	x	x	CSAG2	1	x	x
ZBED1	2	x	x	C8orf58	1	x	x	CSNK2A2	1	x	x
ZGPAT	2	x	x	C8orf74	1	x	x	CSTF1	1	x	x
ZNF138	2	x	x	CA4	1	x	x	CT45A7	1	x	x
ZNF273	2	x	x	CABP5	1	x	x	CTD-2587H24.4	1	x	x
ZNF333	2	x	x	CACNA1B	1	x	x	CTD-3051L14.15	1	x	x
ZNF33B	2	x	x	CALML5	1	x	x	CTLA4	1	x	x
ZNF705G	2	x	x	CALY	1	x	x	CUEDC2	1	x	x
ZNHIT3	2	x	x	CAPNS1	1	x	x	CYHR1	1	x	x
ZSCAN2	2	x	o	CAPNS2	1	x	x	CYP2A6	1	x	o
ZSWIM9	2	x	x	CASP1	1	x	x	CYP46A1	1	x	x
ABHD17A	1	x	x	CASP5	1	x	x	DARS2	1	x	x
ABT1	1	x	o	CASS4	1	x	x	DBNDD1	1	x	x
AC006486.9	1	x	x	CBR1	1	x	x	DCAF16	1	x	x
AC008079.12	1	x	x	CCDC115	1	x	x	DCTD	1	x	x
AC068533.7	1	x	x	CCDC125	1	x	x	DCUN1D4	1	x	o
ACADVL	1	x	x	CCDC17	1	x	x	DDX41	1	x	x
ACAP3	1	x	x	CCKBR	1	x	x	DDX51	1	x	x
ACBD4	1	x	x	CCL11	1	x	x	DECR2	1	x	x
ACP5	1	x	x	CCL2	1	x	x	DEF8	1	x	x
ACSM2B	1	x	x	CCL3	1	x	x	DEFA3	1	x	x
ACTRT2	1	x	x	CCL7	1	x	x	DEFB105A	1	x	o
ADAM30	1	x	x	CCN5	1	x	x	DEFB126	1	x	x
ADAMTS10	1	x	x	CCNB1	1	x	x	DGCR2	1	x	x
ADAMTS15	1	x	x	CCT6A	1	x	x	DGKE	1	x	x
AGPAT2	1	o	x	CCT8L2	1	x	x	DHH	1	x	x
AGPAT5	1	x	x	CD177	1	x	x	DHODH	1	x	x
AJUBA	1	x	x	CD1B	1	x	x	DHRS4	1	x	x
AKAP17A	1	x	x	CD1E	1	x	x	DHX36	1	x	x
AKAP5	1	x	x	CD276	1	x	x	DNAAF3	1	x	x
AL928654.7	1	x	o	CD2BP2	1	o	x	DOCK3	1	x	x
ALG10	1	x	o	CDC16	1	x	x	DPF2	1	x	x
ALG10B	1	x	x	CDC45	1	x	x	DXO	1	x	x
ALPG	1	x	x	CDK7	1	x	x	DYRK4	1	x	x
ALPI	1	x	x	CDKL4	1	x	x	E2F4	1	x	x
ALPK3	1	x	x	CDR2	1	x	x	EAPP	1	x	x
ALPP	1	x	x	CEACAM6	1	x	x	EBAG9	1	x	x
AMER2	1	x	x	CEBPD	1	x	x	ECEL1	1	x	x
AMER3	1	x	x	CELA3B	1	x	x	EDEM2	1	x	x
AMY2A	1	x	x	CELSR2	1	x	o	EEF1A2	1	x	x
ANGEL1	1	x	x	CENPH	1	x	x	EEF2K	1	x	x
ANKRD20A1	1	x	x	CENPJ	1	x	x	EFNA2	1	x	x
AP000471.4	1	x	x	CENPL	1	x	x	EGFL7	1	o	x
AQP12B	1	x	o	CENPVL3	1	x	x	EHMT2	1	x	x
ARGLU1	1	x	x	CES1	1	x	o	EID3	1	x	x
ARL14EPL	1	x	x	CES3	1	x	x	EIF5A2	1	x	x
ARMCX5	1	x	x	CFAP126	1	x	x	ELMO3	1	x	x
ASS1	1	x	x	CFAP20	1	x	x	ELOB	1	x	x
ASXL1	1	x	o	CFAP410	1	x	x	ENSG00000283886	1	x	x
ATF7	1	x	x	CFAP45	1	x	x	EPHB3	1	x	o
ATF7-NPFF	1	x	x	CFB	1	x	x	EPN1	1	x	o
ATP12A	1	x	x	CFHR1	1	x	x	EPPIN	1	x	x
ATP13A1	1	x	x	CFHR4	1	x	x	ERF	1	x	x
ATP13A2	1	x	x	CFL2	1	x	x	ERI1	1	x	x
ATP6V0E2	1	o	x	CGB2	1	x	x	ERLEC1	1	x	x
AURKA	1	x	x	CGRRF1	1	x	x	ERVV-1	1	x	x
B3GALT6	1	x	x	CH17-159N18.5	1	x	o	ESS2	1	x	x
B4GALT7	1	x	x	CHAC2	1	x	x	EXOC3L1	1	x	x
BAHD1	1	x	x	CHAMP1	1	x	x	EYA3	1	x	x
BBS5	1	x	o	CHCHD2	1	x	x	F3	1	x	x
BDH1	1	x	x	CHMP1A	1	x	x	FAM107A	1	x	x
BDP1	1	x	o	CHMP2A	1	x	x	FAM153A	1	x	x
BEST2	1	x	x	CHRA1	1	x	x	FAM160A2	1	x	x
BHLHA9	1	x	x	CHST14	1	x	x	FAM193B	1	x	x
BHMT	1	o	x	CINP	1	x	x	FAM209A	1	x	x
BICRA	1	x	x	CKMT1B	1	x	x	FAM209B	1	x	x
BRD9	1	x	x	CLDN5	1	x	x	FAM210B	1	x	x
BTN1A1	1	x	x	CLTCL1	1	x	x	FAM240C	1	x	x
BTN2A1	1	x	x	CNGA4	1	x	x	FAM25G	1	x	x
BTN2A2	1	x	x	CNIH1	1	x	x	FAM3A	1	x	x

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
FAM50A	1	x	x	IGLL5	1	x	x	MRPS36	1	x	x
FAM83E	1	x	x	IMP4	1	x	x	MTHFD1	1	x	x
FAM90A26	1	x	x	IMPDH1	1	x	x	MTLN	1	x	o
FASTKD1	1	x	o	INSYN1	1	x	x	MTMR6	1	x	x
FBXL15	1	x	x	INTS14	1	x	x	MTRNR2L12	1	x	x
FBXO42	1	x	x	ISLR	1	x	x	MYBPHL	1	x	x
FBXW10	1	x	x	ISLR2	1	x	o	MYEOV	1	x	o
FCER1A	1	x	x	IZUMO1	1	x	x	MZF1	1	x	x
FCGR1A	1	x	o	JMY	1	x	x	MZT2A	1	x	x
FCGR1B	1	x	x	JOSD2	1	x	x	NAPA	1	x	x
FGF21	1	x	x	JUP	1	x	x	NAT8L	1	x	x
FGFBP2	1	x	x	KAT14	1	x	x	NBPF10	1	x	x
FLII	1	x	x	KCNF1	1	x	x	NBPF14	1	x	x
FMC1	1	x	x	KCNK15	1	x	x	NBPF20	1	x	x
FNDCC10	1	x	x	KCNT1	1	x	x	NBPF6	1	x	x
FOXD3	1	x	x	KCTD5	1	x	x	NELFE	1	x	x
FOXL3	1	x	x	KCTD7	1	x	x	NEU1	1	x	x
FRG2B	1	x	x	KIF5B	1	x	x	NFAT5	1	x	x
FRMD8	1	x	x	KLF13	1	x	x	NFILZ	1	x	x
FSCB	1	x	x	KLHDC7A	1	x	x	NFKB2	1	x	x
FUOM	1	x	x	KLHL20	1	x	x	NKD2	1	x	x
FUT1	1	x	x	KLHL40	1	x	x	NKTR	1	x	x
FUZ	1	x	x	KLHL41	1	x	o	NME8	1	x	x
FZD2	1	x	x	KLK6	1	x	x	NMI	1	x	x
G6PD	1	x	x	KLK7	1	x	x	NOB1	1	x	o
GABRD	1	x	o	KPTN	1	x	x	NOBOX	1	x	x
GADD45B	1	x	x	KRT18	1	x	x	NOC3L	1	x	x
GALE	1	x	x	LAGE3	1	x	x	NOP14	1	x	x
GAS8	1	x	x	LAIR1	1	x	o	NOTCH2NLR	1	x	o
GCNT7	1	x	x	LAPTM5	1	x	x	NPFF	1	x	x
GDF10	1	x	x	LCE2D	1	x	x	NPHP1	1	x	o
GDF2	1	x	x	LCN15	1	x	o	NPIPB3	1	x	x
GDPD3	1	x	x	LHFPL4	1	x	x	NPIPB6	1	x	x
GET4	1	x	x	LIG1	1	x	x	NPIPB9	1	x	x
GGT5	1	x	x	LILRB2	1	x	x	NPTN	1	x	x
GGTLC2	1	x	x	LINC02210-CRHR1	1	x	x	NQO1	1	x	x
GGTLC3	1	x	o	LINC02218	1	x	x	NRGN	1	x	x
GMFB	1	x	x	LMBR1L	1	x	x	NSUN3	1	x	x
GMP	1	x	x	LMNB2	1	x	x	NSUN5	1	x	o
GOLGA6D	1	x	x	LMNTD2	1	x	x	NUP58	1	x	x
GPER1	1	x	x	LONRF1	1	x	x	NUPR2	1	x	x
GPR12	1	x	x	LOXL1	1	x	x	NUTM2D	1	x	x
GPR42	1	x	x	LPAR2	1	x	x	NUTM2G	1	x	x
GPR65	1	x	x	LPCAT2	1	x	x	OCLN	1	x	x
GPRASP1	1	x	x	LRP10	1	x	x	ODAPH	1	x	x
GPRASP2	1	x	x	LRRC56	1	x	x	ODF2L	1	x	x
GREM1	1	x	x	LRRC66	1	x	x	ODF3L2	1	x	x
GSC2	1	x	x	LRRC75B	1	x	x	OPN1MW3	1	x	o
GSTA3	1	x	x	LRRFIP2	1	x	x	OR10A6	1	x	x
GSTA5	1	x	x	LSM2	1	x	x	OR13F1	1	x	x
GSX1	1	x	o	LSM8	1	x	x	OR13G1	1	x	x
GTF2IRD2B	1	x	x	LSS	1	x	o	OR13J1	1	x	x
GTF3C6	1	x	x	LYPLA2	1	x	x	OR1E1	1	x	x
H2AC11	1	x	x	MAGEA1	1	x	x	OR1F1	1	x	x
H2BC18	1	x	o	MAGEA11	1	x	x	OR1G1	1	x	x
H2BW2	1	x	x	MAP3K12	1	x	x	OR1I1	1	x	x
H3Y2	1	x	x	MAPK1	1	x	x	OR2A12	1	x	x
HAUS4	1	x	x	MAPK3	1	x	x	OR2A14	1	x	x
HBA1	1	x	x	MARVELD2	1	x	x	OR2A2	1	x	x
HBQ1	1	x	x	MATN1	1	x	x	OR2A25	1	x	x
HDGFL1	1	x	x	MBD3L5	1	x	x	OR2A5	1	x	x
HDHD2	1	x	x	MC1R	1	x	x	OR2T3	1	x	x
HEXIM1	1	x	x	MCCC2	1	x	o	OR2T34	1	x	x
HEXIM2	1	x	x	MCM4	1	x	x	OR4A47	1	x	x
HHATL	1	x	x	MCRIP2	1	x	x	OR4B1	1	x	x
HHIPL1	1	x	x	METTL26	1	x	x	OR4C3	1	x	x
HIC1	1	x	x	METTL2B	1	x	x	OR4C5	1	x	x
HILPDA	1	x	x	METTL9	1	x	o	OR4F15	1	x	x
HIRA	1	x	x	MFHAS1	1	x	x	OR4F6	1	x	x
HMGNA4	1	x	x	MIEF2	1	x	x	OR4N5	1	x	x
HRAS	1	x	x	MIER2	1	x	x	OR4S1	1	x	x
HSBP1L1	1	x	x	MMP14	1	x	x	OR4X1	1	x	x
HSPA1A	1	x	x	MMP2	1	x	x	OR4X2	1	x	x
HSPA1B	1	x	x	MMP9	1	x	x	OR52L1	1	x	x
HSPA1L	1	x	x	MOK	1	x	x	OR56A4	1	x	x
HSPA2	1	x	x	MPG	1	x	x	OR5AP2	1	x	x
HTR6	1	x	x	MPV17L	1	x	o	OR5AR1	1	x	x
HYPK	1	x	x	MRPL28	1	x	x	OR5M1	1	x	x
IER3IP1	1	x	x	MRPL40	1	x	x	OR5M10	1	x	x
IFNL3	1	x	x	MRPL52	1	x	x	OR5M11	1	x	x

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
OR5M3	1	x	x	RABGEF1	1	x	x	SOHLH1	1	x	o
OR5M8	1	x	x	RAD17	1	x	x	SOS2	1	x	x
OR5M9	1	x	x	RAD51AP1	1	x	x	SOX7	1	x	x
OR5T1	1	x	x	RASSF7	1	x	x	SPACA5B	1	x	o
OR8H1	1	x	x	RBM23	1	x	x	SPATA18	1	x	o
OR8J1	1	x	x	RBM28	1	x	x	SPATA21	1	o	x
OR8K1	1	x	x	RBM43	1	x	x	SPATA31A3	1	x	x
OR8K3	1	x	x	RBMY1D	1	x	o	SPATA31D3	1	x	x
OR8U3	1	x	x	RDH16	1	x	x	SPATA33	1	x	x
ORM2	1	x	x	REEP6	1	x	x	SPDYE16	1	x	x
OTOA	1	x	o	REM2	1	x	x	SPIDR	1	x	x
P2RX1	1	x	x	REXO4	1	x	x	SPN	1	x	x
P2RX3	1	x	x	RFLNB	1	x	x	SPPL2C	1	x	x
P3H4	1	x	x	RGCC	1	x	x	SPRR2D	1	x	x
PABPC3	1	x	x	RGPD4	1	x	x	SRRM2	1	x	x
PARN	1	x	o	RHBDF1	1	x	x	SS18L2	1	x	x
PARP8	1	x	x	RHOF	1	x	x	SSRP1	1	x	x
PCNA	1	x	x	RIF1	1	x	x	SSX4B	1	x	o
PCSK4	1	x	x	RILPL2	1	x	x	STAR	1	x	x
PDCD1	1	x	x	RNASE2	1	x	x	STING1	1	x	x
PDF	1	x	x	RNASE3	1	x	x	STK19	1	x	x
PDLIM7	1	x	x	RNF112	1	x	x	STKLD1	1	x	x
PDX1	1	x	o	RNF17	1	x	x	STOML1	1	x	x
PET117	1	x	x	RNF223	1	x	x	SULT1A2	1	x	o
PEX26	1	x	x	RNF227	1	x	x	SUMF2	1	x	x
PGLYRP1	1	x	x	ROGDI	1	x	x	SURF1	1	x	x
PGM5	1	x	x	RP1-37C10.8	1	x	x	SURF2	1	x	x
PHKG1	1	x	x	RP11-152F13.10	1	x	x	SURF4	1	x	x
PIM3	1	x	x	RP11-298I3.5	1	x	x	SYDE1	1	x	x
PITHD1	1	x	x	RP11-343C2.11	1	x	x	SZRD1	1	x	x
PIWIL3	1	x	x	RP11-364B14.3	1	x	x	TACSTD2	1	x	x
PKP3	1	x	x	RP11-426L16.10	1	x	x	TAF9	1	x	x
PLA2G10	1	x	o	RP11-473M10.3	1	x	x	TARBP2	1	x	x
PLA2G4C	1	x	x	RP11-49K24.6	1	x	x	TARM1	1	x	x
PLCD3	1	x	x	RP11-566K11.2	1	x	x	TAS2R13	1	x	x
PLCXD1	1	x	x	RP11-724O16.1	1	x	o	TBC1D21	1	x	x
PLGLB2	1	x	x	RP11-792A8.5	1	x	x	TBC1D3F	1	x	x
PLIN4	1	x	x	RP11-872D17.8	1	x	x	TCFL5	1	x	x
PLIN5	1	x	x	RPEL1	1	x	x	TDRD1	1	x	x
PLK5	1	x	x	RSC1A1	1	x	x	TEDC1	1	x	o
PLXNA3	1	x	x	RTF2	1	x	x	TEX13A	1	x	x
PML	1	x	x	RTP5	1	x	x	TEX50	1	x	x
PNMA6F	1	x	x	RUSF1	1	x	o	THEG	1	x	x
POLR3E	1	x	x	SAMD10	1	x	x	THY1	1	x	o
POU4F1	1	x	x	SAPCD1	1	x	x	TIGD3	1	x	x
PPDPF	1	x	x	SARS1	1	x	o	TIMM13	1	x	x
PPIAL4C	1	x	x	SBDS	1	x	x	TIMM22	1	x	x
PPIAL4G	1	x	x	SBK1	1	x	o	TKTL1	1	x	x
PPIAL4H	1	x	x	SDF4	1	x	x	TMED9	1	x	x
PPIL2	1	x	x	SDR42E2	1	x	x	TMEM121	1	x	o
PPM1F	1	x	x	SEC11A	1	x	o	TMEM129	1	x	x
PPP1R36	1	x	x	SEC22C	1	x	x	TMEM158	1	x	x
PPP1R3B	1	x	x	SERINC4	1	x	x	TMEM18	1	x	x
PRAMEF1	1	x	x	SERPINA12	1	x	x	TMEM191C	1	x	x
PRAMEF14	1	x	x	SERPINA4	1	x	x	TMEM211	1	x	x
PRAMEF25	1	x	x	SERPINA5	1	x	x	TMEM230	1	x	x
PRAMEF5	1	x	o	SETD4	1	x	x	TMEM240	1	x	x
PRAP1	1	x	x	SFXN4	1	x	x	TMEM248	1	x	x
PRB1	1	x	x	SGCB	1	x	o	TMEM252	1	x	x
PRDM7	1	x	x	SHC2	1	x	x	TNFAIP6	1	x	x
PRDX3	1	x	x	SHMT1	1	x	x	TNFRSF18	1	x	x
PRG2	1	x	x	SKIV2L	1	x	x	TNFRSF4	1	x	x
PRG3	1	x	x	SLC10A3	1	x	x	TNKS1BP1	1	x	x
PRH2	1	x	x	SLC25A1	1	x	x	TNNI3	1	x	x
PRKDC	1	x	x	SLC25A45	1	x	x	TOP3A	1	x	x
PRMT5	1	x	x	SLC27A5	1	x	x	TOP3B	1	x	x
PRR33	1	x	x	SLC28A1	1	x	x	TP53TG3F	1	x	x
PRSS27	1	x	x	SLC30A5	1	x	x	TPGS1	1	x	x
PSD	1	x	x	SLC35E2B	1	x	x	TPH1	1	x	x
PSG6	1	x	x	SLC43A3	1	x	x	TRARG1	1	x	x
PSG7	1	x	x	SLC44A4	1	x	x	TRIM25	1	x	x
PSG8	1	x	x	SLC52A3	1	x	x	TRIM28	1	x	x
PSPH	1	x	x	SLC5A2	1	x	o	TRIM74	1	x	x
PSRC1	1	x	o	SLC6A2	1	x	x	TRIML2	1	x	x
PTK6	1	x	o	SLC8A2	1	x	x	TRIP13	1	x	o
PTPN18	1	x	x	SMCR8	1	x	x	TRIR	1	x	x
PTPN20	1	x	x	SMIM10L2B	1	x	x	TRMT9B	1	x	x
PUSL1	1	x	x	SMIM27	1	x	x	TRPM1	1	x	x
QPR1	1	x	x	SMPD4	1	x	x	TSPAN3	1	x	x
RAB12	1	x	x	SNURF	1	x	x	TSSK2	1	x	x

Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS	Gene	CPC Count (n=116)	HPR C.EA S	HPR C.nE AS
TTYH1	1	x	x	VSTM1	1	x	x	ZNF281	1	x	x
TUBA3C	1	x	x	VWA3A	1	x	o	ZNF324	1	x	x
TUBA8	1	x	x	VWCE	1	x	x	ZNF446	1	x	x
TUBAL3	1	x	x	WDR1	1	x	x	ZNF449	1	x	o
TUBB3	1	x	x	WFIKN1	1	x	x	ZNF496	1	x	x
TVP23B	1	x	x	WWP2	1	x	x	ZNF511	1	x	x
UBE2M	1	x	x	XAGE5	1	x	x	ZNF541	1	x	x
UBL4A	1	x	x	XKR5	1	x	x	ZNF558	1	x	x
UCN2	1	x	x	XXbac-BPG116M5.17	1	x	x	ZNF565	1	x	x
UCN3	1	x	x	YPEL1	1	x	x	ZNF592	1	x	x
UFD1	1	x	x	ZBTB1	1	x	x	ZNF679	1	x	x
UGT2B28	1	x	o	ZBTB12	1	x	x	ZNF680	1	x	x
UPF3A	1	x	x	ZBTB25	1	x	x	ZNF705D	1	x	o
USF1	1	x	x	ZBTB45	1	x	x	ZNF716	1	x	o
USP17L12	1	x	x	ZBTB47	1	x	x	ZNF735	1	x	x
USP17L2	1	x	x	ZCCHC18	1	x	x	ZNF736	1	x	x
USP17L27	1	x	x	ZFAND2A	1	x	x	ZNF750	1	x	o
USP17L29	1	x	x	ZFP14	1	x	o	ZNF75D	1	x	o
USP2	1	x	o	ZFP82	1	x	o	ZNF761	1	x	x
USP32	1	x	x	ZIM3	1	x	x	ZNF799	1	x	x
VASH1	1	x	x	ZNF101	1	x	x	ZNF813	1	x	x
VCPKMT	1	x	x	ZNF107	1	x	x	ZNF839	1	x	x
VGLL3	1	x	x	ZNF146	1	x	x	ZP2	1	x	x
VPS37D	1	x	x	ZNF200	1	x	o				
VPS4A	1	x	x	ZNF217	1	x	x				

Table S5 Overlapped CNV-related-genes between CPC and HPRC

Gene	CPC Freque ncy (n=116)	HPRC Freque ncy (n=88)	HPRC.E AS Freque ncy (n=8)	HPRC.n EAS Freque ncy (n=80)	Gene	CPC Freque ncy (n=116)	HPRC Freque ncy (n=88)	HPRC.E AS Freque ncy (n=8)	HPRC.n EAS Freque ncy (n=80)
CYP2D6	0.26	0.06	0.00	0.05	TSEN34	0.03	0.02	0.00	0.03
CFC1	0.18	0.06	0.00	0.06	RPIA	0.03	0.02	0.00	0.03
PNMA6A	0.15	0.03	0.00	0.04	CATSPER2	0.03	0.02	0.00	0.03
GOLGA6L4	0.09	0.02	0.00	0.03	OR4Q3	0.25	0.25	0.13	0.25
TSPY8	0.16	0.10	0.13	0.10	NBPF19	0.26	0.26	0.00	0.29
LIMS4	0.09	0.05	0.00	0.05	MTLN	0.01	0.01	0.00	0.01
CTAG1A	0.06	0.01	0.00	0.01	COL6A2	0.01	0.01	0.00	0.01
POTEB2	0.14	0.09	0.13	0.06	SBK1	0.01	0.01	0.00	0.01
ARHGEF35	0.07	0.02	0.13	0.01	RUSF1	0.01	0.01	0.00	0.01
GOLGA8N	0.08	0.03	0.00	0.04	ZFP82	0.01	0.01	0.00	0.01
GOLGA6L9	0.04	0.00	0.00	0.00	MCCC2	0.01	0.01	0.00	0.01
ZNF658	0.05	0.01	0.00	0.01	NOB1	0.01	0.01	0.00	0.01
NOTCH2NLB	0.10	0.07	0.13	0.06	DEFB105A	0.01	0.01	0.00	0.01
CD99	0.03	0.00	0.00	0.00	LCN15	0.01	0.01	0.00	0.01
CRLF2	0.03	0.00	0.00	0.00	MPV17L	0.01	0.01	0.00	0.01
FAHD2B	0.04	0.01	0.00	0.01	EPHB3	0.01	0.01	0.00	0.01
PRAMEF4	0.04	0.01	0.00	0.01	ZNF200	0.01	0.01	0.00	0.01
PRAMEF2	0.04	0.01	0.00	0.01	UGT2B28	0.01	0.01	0.00	0.01
PRAMEF10	0.06	0.03	0.00	0.04	FASTKD1	0.01	0.01	0.00	0.01
VAMP7	0.03	0.00	0.00	0.00	AGPAT2	0.01	0.01	0.13	0.00
GTPBP6	0.03	0.00	0.00	0.00	ZNF750	0.01	0.01	0.00	0.01
CHRFAM7A	0.03	0.01	0.00	0.01	BDP1	0.01	0.01	0.00	0.01
SSX2	0.03	0.01	0.00	0.01	SOHLH1	0.01	0.01	0.00	0.01
BOLA1	0.03	0.01	0.13	0.00	PDX1	0.01	0.01	0.00	0.01
H3-2	0.04	0.02	0.00	0.03	ISLR2	0.01	0.01	0.00	0.01
SYT15	0.04	0.02	0.00	0.03	H2BC18	0.01	0.01	0.00	0.01
PRAMEF8	0.05	0.03	0.00	0.04	GSX1	0.01	0.01	0.00	0.01
ZBED1	0.02	0.00	0.00	0.00	ASXL1	0.01	0.01	0.00	0.01
IL9R	0.02	0.00	0.00	0.00	VWA3A	0.01	0.01	0.00	0.01
SHOX	0.02	0.00	0.00	0.00	PSRC1	0.01	0.01	0.00	0.01
NTAN1	0.06	0.05	0.00	0.05	RP11-724O16.1	0.01	0.01	0.00	0.01
RRN3	0.06	0.05	0.00	0.05	NOTCH2NLR	0.01	0.01	0.00	0.01
PRAMEF27	0.03	0.01	0.00	0.01	LSS	0.01	0.01	0.00	0.01
SULT2A1	0.03	0.01	0.00	0.01	FCGR1A	0.01	0.01	0.00	0.01
DEFA5	0.03	0.01	0.00	0.01	SGCB	0.01	0.01	0.00	0.01
FAM178B	0.03	0.01	0.00	0.01	AQP12B	0.01	0.01	0.00	0.01
ANKRD20A4P	0.03	0.01	0.00	0.01	GABRD	0.01	0.01	0.00	0.01
USP51	0.03	0.01	0.00	0.01	USP2	0.01	0.01	0.00	0.01
LIMS3	0.03	0.01	0.00	0.01	BBS5	0.01	0.01	0.00	0.01
IZUMO2	0.03	0.01	0.00	0.01	KLHL41	0.01	0.01	0.00	0.01
ERV3-1	0.03	0.01	0.00	0.01	PTK6	0.01	0.01	0.00	0.01
ANKRD36B	0.03	0.01	0.13	0.00	SULT1A2	0.01	0.01	0.00	0.01
FAHD2A	0.03	0.01	0.00	0.01	SARS1	0.01	0.01	0.00	0.01
SEMA4C	0.03	0.01	0.00	0.01	CELSR2	0.01	0.01	0.00	0.01
SLC16A8	0.03	0.01	0.00	0.01	EPN1	0.01	0.01	0.00	0.00
SLC5A10	0.03	0.01	0.00	0.01	SLC5A2	0.01	0.01	0.00	0.01
RRP7A	0.11	0.10	0.13	0.10	ATP6V0E2	0.01	0.01	0.13	0.00
RGPD5	0.04	0.03	0.13	0.03	ZNF705D	0.01	0.01	0.00	0.01
AKAP17A	0.01	0.00	0.00	0.00	PRAMEF5	0.01	0.01	0.00	0.01
PLCXD1	0.01	0.00	0.00	0.00	NPHP1	0.01	0.01	0.00	0.01
CD2BP2	0.01	0.00	0.00	0.00	SPATA21	0.01	0.01	0.13	0.00
CRX	0.02	0.01	0.00	0.01	EGFL7	0.01	0.01	0.13	0.00
NECAP2	0.02	0.01	0.13	0.00	OPN1MW3	0.01	0.01	0.00	0.01
LILRB5	0.02	0.01	0.00	0.01	ZFP14	0.01	0.01	0.00	0.01
PCDHB16	0.02	0.01	0.00	0.01	TRIP13	0.01	0.01	0.00	0.01
FAM72C	0.02	0.01	0.00	0.01	BHMT	0.01	0.01	0.13	0.00
CNNM4	0.02	0.01	0.00	0.01	ABT1	0.01	0.01	0.00	0.01
CBWD6	0.02	0.01	0.13	0.00	ZNF716	0.01	0.01	0.00	0.01
HNRNPCL1	0.02	0.01	0.00	0.01	THY1	0.01	0.01	0.00	0.01
WDR73	0.02	0.01	0.00	0.01	SPATA18	0.01	0.01	0.00	0.01
EIF3CL	0.02	0.01	0.00	0.01	SEC11A	0.01	0.01	0.00	0.01
FAM72B	0.02	0.01	0.00	0.01	DCUN1D4	0.01	0.01	0.00	0.01
PDXDC1	0.02	0.01	0.00	0.01	PARN	0.01	0.01	0.00	0.01
CDK11B	0.02	0.01	0.00	0.01	ALG10	0.01	0.01	0.00	0.01
TPRX1	0.02	0.01	0.00	0.01	EIF2AK3	0.02	0.02	0.00	0.03
TRIML1	0.02	0.01	0.00	0.01	NOTCH2NLC	0.02	0.02	0.00	0.03
NMB	0.02	0.01	0.00	0.01	CLN3	0.02	0.02	0.00	0.03
IL27	0.02	0.01	0.00	0.01	APOBR	0.02	0.02	0.00	0.03
BSPH1	0.02	0.01	0.00	0.01	RP11-315A19.1	0.02	0.02	0.00	0.03
RP11-435I10.4	0.02	0.01	0.00	0.01	FOXI3	0.02	0.02	0.00	0.03
HSFX2	0.02	0.01	0.00	0.00	IGSF6	0.02	0.02	0.00	0.03
NDUF8	0.02	0.01	0.00	0.01	HIC2	0.03	0.03	0.00	0.04
ZSCAN2	0.02	0.01	0.00	0.01	OR4N2	0.10	0.11	0.00	0.11
LMAN2L	0.02	0.01	0.00	0.01	FAM156B	0.03	0.05	0.13	0.04
SELENOW	0.02	0.01	0.00	0.01	SSX4B	0.01	0.02	0.00	0.03
ANKRD36	0.02	0.01	0.00	0.01	CES1	0.01	0.02	0.00	0.03
CNNM3	0.02	0.01	0.00	0.01	CYP2A6	0.01	0.02	0.00	0.03
NUPR1	0.02	0.01	0.00	0.01	AL928654.7	0.01	0.02	0.00	0.01
OR4M1	0.23	0.23	0.00	0.24	SPACA5B	0.01	0.02	0.00	0.03

Gene	CPC Freque ncy (n=116)	HPRC Freque ncy (n=88)	HPRC.E AS Freque ncy (n=8)	HPRC.n EAS Freque ncy (n=80)	Gene	CPC Freque ncy (n=116)	HPRC Freque ncy (n=88)	HPRC.E AS Freque ncy (n=8)	HPRC.n EAS Freque ncy (n=80)
ZNF449	0.01	0.02	0.00	0.03	SPDYE2	0.28	0.74	0.38	0.75
PLA2G10	0.01	0.02	0.00	0.03	ARL17A	0.16	0.65	0.50	0.64
ATF7IP2	0.01	0.02	0.13	0.01	NBPF1	0.47	0.98	0.75	0.98
CH17-159N18.5	0.01	0.02	0.00	0.03	DEFB107A	0.46	0.98	0.75	0.98
MYEOV	0.01	0.02	0.00	0.03	NOMO1	0.44	0.98	0.75	0.98
COL6A1	0.01	0.02	0.00	0.03	OR2A1	0.43	0.98	0.75	0.98
LAIR1	0.01	0.02	0.00	0.03	GAGE2E	0.18	0.74	0.50	0.74
CRIP1	0.01	0.02	0.00	0.01	SLX1A	0.37	0.98	0.75	0.98
OTOA	0.01	0.02	0.00	0.03	OR4F17	0.35	0.98	0.75	0.98
ZNF75D	0.01	0.02	0.00	0.03	RASA4	0.33	0.98	0.75	0.98
TEDC1	0.01	0.02	0.00	0.01	UPK3BL1	0.32	0.98	0.75	0.98
XAGE3	0.01	0.02	0.13	0.01	GPAT2	0.04	0.77	0.63	0.78
TMEM191B	0.05	0.07	0.00	0.08	DEFB4A	0.23	0.98	0.75	0.98
ABC7-420440C24.1	0.05	0.07	0.00	0.06	PRR23D1	0.23	0.98	0.75	0.98
CYP21A2	0.02	0.03	0.00	0.04	DEFB104A	0.22	0.98	0.75	0.98
CT55	0.02	0.03	0.00	0.04	DEFB106A	0.22	0.98	0.75	0.98
CORO1A	0.02	0.03	0.25	0.01	PRR20A	0.22	0.98	0.75	0.98
EMB	0.02	0.03	0.00	0.04	GOLGA6L10	0.17	0.98	0.75	0.98
TPSB2	0.02	0.03	0.00	0.04	DEFA1	0.17	0.98	0.75	0.98
TEX37	0.02	0.03	0.00	0.04	C4A	0.16	0.98	0.75	0.98
BFAR	0.02	0.03	0.00	0.04	BOLA2	0.16	0.98	0.75	0.98
DDTL	0.03	0.05	0.00	0.05	MBD3L2	0.14	0.98	0.75	0.98
OR9G1	0.49	0.51	0.38	0.53	RFPL4A	0.13	0.98	0.75	0.98
METTL9	0.01	0.03	0.00	0.04	AMY1A	0.12	0.98	0.75	0.98
TMEM121	0.01	0.03	0.00	0.03	KIR2DL1	0.12	0.98	0.75	0.98
CRIP2	0.01	0.03	0.00	0.01	NPY4R	0.11	0.98	0.75	0.98
RCC1L	0.01	0.03	0.13	0.03	ORM1	0.10	0.98	0.75	0.98
STRC	0.02	0.05	0.00	0.05	GOLGA6A	0.10	0.98	0.75	0.98
RBM1Y1B	0.02	0.05	0.13	0.04	CLEC18A	0.09	0.98	0.75	0.98
TSPY1	0.04	0.08	0.00	0.09	GOLGA6L1	0.09	0.98	0.75	0.98
NSUN5	0.01	0.05	0.00	0.05	LRRC37A	0.08	0.98	0.75	0.98
RBM1Y1A1	0.02	0.06	0.13	0.05	POTEB	0.08	0.98	0.75	0.98
RP11-353J17.5	0.02	0.06	0.00	0.06	SPAG11A	0.06	0.98	0.75	0.98
ANTXR1	0.02	0.06	0.00	0.05	SIGLEC14	0.04	0.98	0.75	0.98
TBC1D3I	0.02	0.07	0.00	0.08	CXorf49	0.04	0.98	0.75	0.98
NPIPB12	0.11	0.17	0.00	0.19	GRAP	0.03	0.98	0.75	0.98
GGTLC3	0.01	0.07	0.00	0.08	ETDA	0.03	0.98	0.75	0.98
RBM1Y1D	0.01	0.07	0.00	0.08	TRIM43	0.03	0.98	0.75	0.98
WASH6P	0.02	0.08	0.13	0.06	TBC1D3	0.03	0.98	0.75	0.98
DMRTC1B	0.02	0.08	0.00	0.09	GYPA	0.03	0.98	0.75	0.98
GSTT4	0.10	0.17	0.00	0.19	CT45A1	0.03	0.98	0.75	0.98
OR4K1	0.32	0.39	0.25	0.39	PICK1	0.03	0.98	0.75	0.98
OR4K2	0.31	0.39	0.25	0.39	CKMT1A	0.03	0.98	0.75	0.98
PDPR	0.39	0.47	0.25	0.48	FAM90A1	0.03	0.98	0.75	0.98
SPDYE2B	0.16	0.25	0.25	0.25	LILRA6	0.03	0.98	0.75	0.98
RBM1Y1E	0.03	0.13	0.00	0.14	H4C14	0.03	0.98	0.75	0.98
SPDYE13	0.04	0.14	0.00	0.15	LILRA4	0.03	0.98	0.75	0.98
TSPY10	0.05	0.15	0.25	0.14	TRIM51	0.03	0.98	0.75	0.98
EIF3C	0.10	0.20	0.00	0.23	FRG2	0.03	0.98	0.75	0.98
GPRIN2	0.82	0.93	0.75	0.93	POLR2J	0.03	0.98	0.75	0.98
TSPY9P	0.08	0.19	0.25	0.19	SMN1	0.02	0.98	0.75	0.98
FRG1	0.02	0.15	0.13	0.15	PRAMEF11	0.02	0.98	0.75	0.98
MRGPRX1	0.12	0.26	0.13	0.28	FCGR2A	0.02	0.98	0.75	0.98
DEFB105B	0.20	0.34	0.38	0.34	TRIM48	0.02	0.98	0.75	0.98
ANKRD20A3P	0.01	0.16	0.00	0.15	ANKRD23	0.02	0.98	0.75	0.98
TPTE	0.42	0.58	0.38	0.60	TRIM64	0.02	0.98	0.75	0.98
TSPY4	0.03	0.19	0.25	0.19	CGB1	0.01	0.98	0.75	0.98
CCL3L1	0.28	0.47	0.63	0.44	SMIM10L2A	0.01	0.98	0.75	0.98
PRAMEF18	0.40	0.58	0.13	0.60	CCL4	0.01	0.98	0.75	0.98
TSPY3	0.01	0.19	0.25	0.19	FAM86B1	0.01	0.98	0.75	0.98
CLPS	0.34	0.55	0.63	0.53	GGT1	0.01	0.98	0.75	0.98
RP11-577H5.5	0.27	0.48	0.25	0.48	ZDHHC11	0.01	0.98	0.75	0.98
SULT1A4	0.13	0.34	0.38	0.34	NPIPA1	0.01	0.98	0.75	0.98
RP11-514P8.7	0.24	0.49	0.38	0.49	SPATA31D1	0.01	0.98	0.75	0.98
SULT1A3	0.37	0.64	0.25	0.66	PLG	0.01	0.98	0.75	0.98
RP11-514P8.6	0.30	0.57	0.38	0.58	OPN1LW	0.01	0.98	0.75	0.98
DUSP22	0.56	0.83	0.63	0.83	NUTM2A	0.01	0.98	0.75	0.98
BOLA2B	0.42	0.70	0.38	0.73	XAGE1A	0.01	0.98	0.75	0.98
LINC02203	0.02	0.41	0.63	0.36	PSG1	0.01	0.98	0.75	0.98
NBPF9	0.02	0.44	0.25	0.45	ANXA8	0.01	0.98	0.75	0.98

Table S6. Functional enrichment of genes affected by the novel SVs ($\geq$ 1kb) in the CPC assembly set.

Database	Term ID	Description	Number of targeted genes		Odds ratio	FDR-corrected p-value
			In the query gene set	In the background gene set		
GO:BP	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	49	167	3.74	1.00×10^{-12}
GO:BP	GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	61	270	2.88	7.91×10^{-11}
GO:BP	GO:0043547	positive regulation of GTPase activity	62	405	1.95	4.96×10^{-4}
GO:BP	GO:0035249	synaptic transmission, glutamatergic	22	90	3.11	0.002
GO:BP	GO:0043087	regulation of GTPase activity	66	481	1.75	0.006
GO:BP	GO:0050808	synapse organization	55	384	1.82	0.009
GO:BP	GO:0034330	cell junction organization	42	287	1.86	0.045
GO:BP	GO:0034332	adherens junction organization	25	139	2.29	0.045
GO:BP	GO:0060078	regulation of postsynaptic membrane potential	25	139	2.29	0.045
GO:CC	GO:0097060	synaptic membrane	71	431	2.10	1.05×10^{-6}
GO:CC	GO:0045211	postsynaptic membrane	54	322	2.14	2.54×10^{-5}
GO:CC	GO:0098984	neuron to neuron synapse	50	342	1.86	0.003
GO:CC	GO:0099572	postsynaptic specialization	49	340	1.84	0.004
GO:CC	GO:0014069	postsynaptic density	46	316	1.86	0.004
GO:CC	GO:0032279	asymmetric synapse	46	320	1.83	0.005
GO:CC	GO:0098978	glutamatergic synapse	48	350	1.75	0.009
GO:CC	GO:0098802	plasma membrane receptor complex	29	184	2.01	0.018
GO:CC	GO:0030673	axolemma	6	15	5.10	0.041
GO:CC	GO:0098878	neurotransmitter receptor complex	12	54	2.83	0.041
GO:CC	GO:0005913	cell-cell adherens junction	19	110	2.20	0.041
GO:CC	GO:0045121	membrane raft	40	304	1.68	0.041
GO:CC	GO:0031838	haptoglobin-hemoglobin complex	5	11	5.80	0.041
GO:CC	GO:0098857	membrane microdomain	40	305	1.67	0.041
GO:MF	GO:0005096	GTPase activator activity	46	279	2.12	6.45×10^{-4}
GO:MF	GO:0030695	GTPase regulator activity	49	309	2.04	6.45×10^{-4}
GO:MF	GO:0060589	nucleoside-triphosphatase regulator activity	52	353	1.89	0.002
GO:MF	GO:0048020	CCR chemokine receptor binding	13	43	3.88	0.003
GO:MF	GO:0004115	3',5'-cyclic-AMP phosphodiesterase activity	7	15	5.99	0.012
GO:MF	GO:0008066	glutamate receptor activity	9	27	4.28	0.021
GO:MF	GO:0030276	clathrin binding	13	57	2.93	0.047

Table S7 Copy number of *RASA4(B)* genes in CPC samples

Assembly	Ethnic group	RASA4(B) Dosage	Assembly	Ethnic group	RASA4(B) Dosage
HG00438.1	Han	5copy	HIFI032487D.1	Bouyei	4copy
HG00438.2	Han	5copy	HIFI032501D.1	Mosuo	3copy+partial
HG00621.1	Han	4copy+partial	HIFI032501D.2	Mosuo	4copy
HG00621.2	Han	1copy+partial	HIFI032510D.1	Dong	4copy
HIFI032007D.1	Gin	1copy+partial	HIFI032510D.2	Dong	2copy
HIFI032007D.2	Gin	6copy	HIFI032513D.1	Gin	3copy+partial
HIFI032018D.1	Chosen	4copy+partial	HIFI032513D.2	Gin	3copy+partial
HIFI032018D.2	Chosen	4copy	HIFI032529D.2	Wa	2copy
HIFI032097D.1	Tujia	3copy	HIFI032566D.1	Yugur	2copy
HIFI032097D.2	Tujia	2copy	HIFI032566D.2	Yugur	3copy+partial
HIFI032103D.1	Tujia	3copy+partial	HIFI032567D.1	Qiang	2copy+partial
HIFI032103D.2	Tujia	2copy	HIFI032567D.2	Qiang	2copy
HIFI032161D.1	Mosuo	5copy	HIFI032585D.1	Bai	2copy
HIFI032161D.2	Mosuo	3copy+partial	HIFI032585D.2	Bai	6copy
HIFI032167D.1	Miao	6copy	HIFI032586D.1	Salar	2copy
HIFI032289D.1	Man	6copy	HIFI032586D.2	Salar	2copy
HIFI032289D.2	Man	4copy+partial	HIFI032591D.1	She	2copy
HIFI032292D.1	Deang	4copy+partial	HIFI032591D.2	She	2copy
HIFI032292D.2	Deang	2copy+partial	HIFI032604D.1	Mongol-TH	5copy
HIFI032302D.1	She	2copy	HIFI032604D.2	Mongol-TH	2copy
HIFI032302D.2	She	5copy	HIFI032607D.1	Mongol	2copy
HIFI032306D.1	Chosen	3copy+partial	HIFI032607D.2	Mongol	5copy
HIFI032306D.2	Chosen	3copy	HIFI032662D.1	Qiang	2copy
HIFI032335D.1	Oroqen	4copy	HIFI032662D.2	Qiang	2copy
HIFI032335D.2	Oroqen	1copy+partial	HIFI032668D.1	Daur	5copy
HIFI032349D.1	Tu	1copy+partial	HIFI032682D.1	Achang	3copy+partial
HIFI032349D.2	Tu	3copy+partial	HIFI032682D.2	Achang	5copy
HIFI032373D.2	Blang	4copy+partial	HIFI032685D.1	Hezhen	2copy
HIFI032420D.1	Qiang	3copy+partial	HIFI032685D.2	Hezhen	2copy
HIFI032420D.2	Qiang	4copy+partial	HIFI032692D.1	Yugur	3copy+partial
HIFI032422D.1	Mosuo	1copy+partial	HIFI032692D.2	Yugur	3copy+partial

HIFI032422D.2	Mosuo	3copy+partial	HIFI032693D.1	Naxi	4copy
HIFI032429D.1	Jino	3copy	HIFI032693D.2	Naxi	1copy+partial
HIFI032440D.1	Jingpo	1copy+partial	HIFI032698D.1	Daur	4copy
HIFI032450D.1	Drung	5copy	HIFI032698D.2	Daur	1copy+partial
HIFI032450D.2	Drung	2copy	HIFI032706D.1	Evenki	3copy+partial
HIFI032453D.1	Salar	2copy	HIFI032706D.2	Evenki	5copy
HIFI032453D.2	Salar	3copy+partial	HIFI032711D.1	Wa	5copy
HIFI032454D.2	Tu	5copy	HIFI032711D.2	Wa	2copy+partial
HIFI032462D.1	Deang	3copy+partial	HIFI032731D.1	Lisu	4copy+partial
HIFI032462D.2	Deang	3copy+partial	HIFI032731D.2	Lisu	4copy+partial
HIFI032473D.1	Dong	4copy+partial	RY06.1	Kyrgyz	2copy
RY01.1	Yao	3copy+partial	RY06.2	Kyrgyz	2copy
RY01.2	Yao	5copy	RY07.1	Kazakh	2copy
RY02.1	Uyghur	4copy+partial	RY07.2	Kazakh	3copy+partial
RY03.1	She	3copy+partial	RY08.1	Zhuang	2copy
RY03.2	She	4copy	RY08.2	Zhuang	2copy
RY04.1	Miao	1copy+partial	RY09.1	Hui	2copy+partial
RY04.2	Miao	4copy	RY09.2	Hui	3copy+partial
RY05.1	Chosen	4copy	RY10.1	Mongol	3copy+partial
RY05.2	Chosen	3copy	RY10.2	Mongol	5copy
			RY11.1	Tujia	3copy+partial
			RY11.2	Tujia	3copy+partial

Table S8 Counts of haplotypes with different *RASA4(B)* copy number in different populations

	1copy+partial	2copy	2copy+partial	3copy	3copy+partial	4copy	4copy+partial	5copy
Austro-Asiatic	1	1	2		4		2	1
Turkic	1	7			6		1	1
Hmong-Mien	1	3			2	2		2
Mongolic	1	1			1	1		3
Sino-Tibetan	4	8	2	2	9	2	4	6
Koreanic				2	1	2	1	
Tungusic	1	2			1	1	1	1
Tai-Kadai		3				2	1	

Note: reference assembly GRCh38 is with 2 copy of *RASA4(B)* and CHM13 is with 2 copy and a 14.9 Kb fragment of *RASA4(B)*.

Table S9 Frequency distribution of *RASA4(B)* copy number in different populations

	1copy+partial	2copy	2copy+partial	3copy	3copy+partial	4copy	4copy+partial	5copy
Austro-Asiatic	0.0833	0.0833	0.1667		0.3333		0.1667	0.0833
Turkic	0.0625	0.4375			0.3750		0.0625	0.0625
Hmong-Mien	0.0909	0.2727			0.1818	0.1818		0.1818
Mongolic	0.1429	0.1429			0.1429	0.1429		0.4286
Sino-Tibetan	0.1053	0.2105	0.0526	0.0526	0.2368	0.0526	0.1053	0.1579
Koreanic				0.3333	0.1667	0.3333	0.1667	
Tungusic	0.1250	0.2500			0.1250	0.1250	0.1250	0.1250
Tai-Kadai		0.5000				0.3333	0.1667	

Table S10 Genes and Frequencies of CDS Regions Covered by CPC
Ancient Introgression Segments

Frequency	Genes
60	PALM2AKAP2
58	TXNDC8
57	HHAT
56	TXN,TBC1D1
55	POU2F3,NLRC5,CTNNA3
53	NRXN3
52	SIPA1L2
51	SNED1,C9orf152
50	SVEP1,STAB2,CSMD2
49	OAF
48	KRT82,IL17RA,COL27A1,BOD1
47	UTP25,SYT14,MACROD2,IRF6,ENSG00000289700
46	SERTAD4,PTPN14,EEF1A2,CAPN11,ANKS1B
45	ZFP37,OSBPL10,LRRTM3,KIAA1217,HELZ2
44	ZMYND10,USH2A,TUSC2,TMEM115,SMYD2,SLC38A3,SEMA3B,RASSF1,NPRL2,NAA80,LSMEM2,KRT85,KRT84,KLF3,KCNQ2,IFRD2,HYAL3,HYAL2,HYAL1,GNAI2,ENSG00000272104,DNAJC12,CYB561D2,CACNA2D2
43	TRIM62,SEMA3F,HSD11B1,GNAT1,GMEB2,ENSG00000277971,CTNND2,CHST15,C12orf42,AK2
42	ZFAND1,TRAF3IP3,SLC10A5,PTTG2,ONECUT2,EPC2,CNTN5,C1orf74
41	RHPN2,RBM19,PTPRN2,G0S2,CDH23,AZIN2
40	UTRN,TSPAN11,SASH1,MUSK,KRT75,KRT6C,KRT6B,KRT6A,KLHL22,FKBP15,CROCC2,CAMK2D
39	TENM3,TBX5,SLC6A5,SLC35F3,SIRT1,SCARF2,NCOR2,MYBPC1,HERC4,CERS6
38	TMTC1,RIMBP2,PRDM16,PPDPF,NXN,NT5DC3,NELL1,JHY,HSP90B1,ENSG00000289325,DDX1,CRYBG1,AGBL1
37	ZNF362,WWOX,SLC35F4,SH3RF2,PTGR1,OSTN,LAMB3,KRT81,HECW2,GRK5,GLP1R,EYA4,ETFBKMT,CPNE4,BORCS5
36	ZBTB43,NAV2,MED15,KLF10,IL17F,FAT1,DKK3,CAMK1G,ABCB11
35	UTS2B,TDG,SPATA2,SMLR1,SLC31A2,SEMA3C,RNF114,RAB18,PTPN3,PPP3R1,PCSK6,MEI4,GLT8D2,FAIM2,ENSG00000273398,DNAAF10,CRTAM,CCDC80,C14orf132,C12orf73,ANTXR1,AMN1
34	ZNF444,UBASH3B,TACC2,SYK,SLC9A8,SAYSD1,RNF19B,RANBP3L,NWD2,MAP3K1,KRT5,ITGB6,GLB1,GALP,FKSG48,ENSG00000275740,DNAH11,CPNE7,COL14A1,CCDC50,ADAMTS7
33	ZNF483,WASHC3,TPPO,ST6GAL1,ST20-MTHFS,SMIM35,SLC17A6,RBMS1,PNO1,PMCH,PGM2,PARBP,OSMR,NUP37,NRIP1,NEBL,MYPN,MTHFS,FXDY6-FXYD2,FRMD3,EPB41L2,CTSH,CSTA,COL13A1,CLYBL,ADRA1D,AC024580.1
32	ZNRF4,TWIST2,TLR6,TLR1,STK17B,PLA2G2A,LMX1B,LHFPL3,KCNK1,DNAH7,COL12A1,BTBD16
31	TMEM121B,TIAM1,SPATA13,SLC39A10,ROR2,PTCHD3,MROH7-TTC4,MORF4L1,LY75-CD302,LY75,LMX1A,KRT86,KRT83,GABBR2,FZD10,EPB41L4B,ENSG00000287694,DROSHA,CSNK1A1,CNTNAP4,CMTM8,CHRNB4,BCL2A1,BAZ2B,A3GALT2
30	TRA2B,RELL1,PYDC2,PRKCH,PANK1,NFKBIE,NEDD9,MROH7,MAGI2,LDLRAD3,ITPR2,IGF1,HTATIP2,FHL2,DAPK1,CHFR,CDH15,CACNA2D3,ARHGAP42
29	SSR1,SSPN,SMOX,SLC22A31,SKP2,RASGRF1,PASK,NPAS3,NIM1K,NADK2,KCNH1,IMPG1,IL10RA,GRM1,GPD1L,GOLGA3,GHR,ENSG00000284299,ECPAS,DGKB,CSMD1,CHD5,CCDC102B,C1QTNF9B,ATP2C1,ANKLE2
28	ZNF778,TRIM39-RPP21,SLC35B2,SLC31A1,RPP21,REEP3,PRPF4,PLXNA2,PLPP3,PLD5,OSBPL6,MYO3B,MRPS15,LEPROTL1,JMJD1C,HSP90AB1,GRIN2B,GNPTAB,FYB2,FXYD6,EGR2,DRAM1,DOK6,CSF3R,CDG26,CALB2,ARHGEF37,ANKRD22
27	WDR31,UBE2H,TLL1,SYCP3,STT3B,SRRM4,SLC25A3,ROS1,PRMT3,PLK4,PLA2R1,P2RX6,OSCP1,OR2K2,NRBF2,NDST1,NALCN,MYOCD,MMP17,MFSD8,MACROH2A2,LIPM,LARP1B,KSR2,ITGA6,IRAG2,HTR4,HSPA4L,HK1,GRID2,ESRRG,ENSG00000285837,C-PNE2,CNRP1,CHPT1,ASTN2,APAF1,ABHD18
26	ZNF131,WARS2,SLC25A31,RGS10,PPP1R7,PLAC8L1,PDSS2,P2RX2,OVCH1,NNT,MMP20,MBOAT4,MANSC1,KIF16B,KAZN,INTU,IKBIP,GAS7,FMN1,FAM102A,FAH,CGNL1,CDC5L,BSPRY,ANO7,ANO5,ANKRD26,ACBD5
25	ZFHX3,TUNAR,TMOD1,TMEM63B,TIMM21,THAP4,SLC28A3,SERPINB13,NUDT7,MRPL14,MECOM,LIFR,KIF20B,INTS13,GLDN,GCLC,FOS,FBXO15,ENSG00000273167,EGF,CPLX4,COL25A1,CAPN8,CAPN2,C5orf22,C2orf49,ATG4B,ARRDC4,ARID5B
24	YME1L1,USP47,TRIM32,TLR10,TGFBRAP1,SPATS1,RAB3C,PXYLP1,POU4F3,PLA2G5,PIP5KL1,MCPH1,MASTL,LIMD1,KRT2,KCNQ5,ING5,HEXB,GRM7,GRIN3A,FOXO6,EPHX1,ELOVL6,DYSF,DDIT2,CHRNA5,CHRNA3,B9D1,ATOH8,ANXA2R,ADQ,ABLM3
23	ZFAND6,TUT4,TTN,TNFRSF19,LE3,STRA8,SLC24A4,RNF39,RBM27,RASSF8,PRRT1,PPT2-EGFL8,PPT2,PPP1R11,POLR1H,NPAS2,MYL7,MMP7,LARGE1,KALRN,HMX2,GCK,FGD3,EPN2,ENSG00000285085,DNAJC25-GNG10,DNAJC25,DNAJC13,DNAH8,D2HGDH,CTXND1,CDKN2B,BUB3,ARHGEF16,ARHGAP44,ARHGAP26,ANK2
22	ZNF727,ZNF722,ZNF354B,WRAP73,TRIM71,TRIM31,TPRG1L,TMEM54,SYNPO,SLC4A10,SLC16A12,SLC15A4,RNF122,PSMA4,PREP,PGRMC2,PCLO,OR10P1,MCM3,LARS1,KIAA1958,IL2RA,IL17RB,HYKK,HPCA,EYS,ENSG00000285330,DHRS3,DECR1,CPE,CD200R1L,CD200R1,CALB1,CACNA1D,C8orf48,ARSL,AMPH,ADGB,ACSF3
21	ZNF860,ZNF787,UNC5B,TMEM132C,TMED3,THAP7,SYDE2,STAMBPL1,SRMS,SMARCAL1,SLC9A9,SINHCFA,SELENOP,RIN3,RELN,PXMP2,PTK6,PTGIR,PPARGC1B,POLE,POLD2,PGAM5,PDZD2,PKD1,NEK2,MTERF4,LZTR1,KRT74,KRT1,JAKMIP2,IFIH1,HSD17B3,FSTL5,FNDC11,FAM241B,DUSP16,CRTC3,CGAS,CELA1,CCDC152,C20orf27,AKAP7,AEBP1,ADAMTS9,ADAMTS18,AC002472.1
20	ZNF454,ZFP69B,ZFP2,USP6NL,USP53,TUBGCP6,TRPC6,TMPRSS4,TMEM132D,TINAGL1,SUCLG2,SLC28A1,RTN4IP1,RASL10B,RAD51B,PRDM5,PPP2R5E,OTUD1,NELL2,NDUFB1,KRT72,KCNQ1,HDAC10,GPR31,GAS2L2,GAR1,GALNT12,FAT2,ESPNL,DYNC2I1,DNAI3,DDX60L,DBX2,CPSF2,CDKN2A,APP,ANAPC4,ALPK3

Frequency	Genes
19	ZNF891,ZNF592,YARS1,VSTM2B,TNXB,TNS2,TMEM248,SYNPO2,STK32A,SRP9,SPINK4,SORL1,SMAD3,SELENOO,SCMH1,S100PBP,RALGPS1,QRSL1,PYGL,PRKCQ,PMP22,PI4K2B,NPHP3-ACAD11,NPHP3,NEPRO,MYO16,MRPS35,MN1,MGAT5,KRT73,KIAA1522,IMPA1,IGSF8,IGFBP6,HTR1B,HEYL,HECW1,GTPBP8,GRM6,GRID1,GLT1D1,GARNL3,GALNT18,FNDC5,FBRSL1,ENG,DPYSL3,DDX43,CWF19L2,CFI,BIN2,ARHGEF10L
18	WDR91,ULK4,UGT3A2,TRAK1,TP73,TMEM151B,TLE1,TC1E1,TCL1A,TBC1D4,TBATA,SYNPR,STRIT1,SPRYD3,SOBP,SNAI1,SLFN1,SLC30A10,RNF183,PYCR2,PTCH1,PPFIBP1,PLEKHH1,PLEKHA1,PCSK2,PBX3,NTRK2,MYO6,MME,LEFTY2,KRT77,CKNA6,JAG1,HDLBP,GNG10,GCG,GALNT6,FGF10,FAP,ESYT2,ENSG00000255835,DUSP26,DHRS9,CTPS1,CCDC27,C4orf19,ATRN1,AIFM2,ACTA2,ACADS
17	ZNF84,ZNF605,ZNF26,ZNF140,ZCCHC4,ZBTB34,WDR72,UBA5,TTC4,TTC29,TTC22,TP63,TP53I11,TMEM267,TM2D3,TBR1,TAS1R2,TARS3,SP140,SMAD7,SLAMF6,SF3B5,SENP6,SEC62,SCML4,SAMD7,SAMD4A,RSPRY1,RPUSD1,ROBO2,RBM17,QRFP,PSME3IP1,PRKAG2,PFKFB3,PDZRN4,PCDH7,PARS2,PAPPA,OEOP,NME8,NALF1,MYOZ2,MS4A13,MROH2B,MAPK12,IQGAP1,IFT140,HTRA1,HSDL2,HMGCS1,HDHD5,GABRB3,GABRA5,GAB4,ENSG00000283952,ENSG00000272442,EFCAB13,CXADR,COL24A1,CCN1,C7,ATXN7L1,AP2B1,ANO6,ACAD11,AARS2
16	XPA,XIRP2,WSCD1,VASH2,UGCG,TSTD2,TRIM39,TNFSF8,TMEM88,TMEM63A,TM4SF18,THSD7B,TELO2,SPP2,SMIM38,SLFN14,SLC52A3,SHOC1,SEPTIN2,SELENOK,SCARA5,RP13-279N23,2,RHBDL2,RBFOX1,RASSF4,PSTK,PRKRA,PPP1R26,PPM1H,PLEKHG1,PJVK,PAIP1,OR6B3,ONECUT1,OMA1,NOSTRIN,NLRP5,NHSL1,NFYB,NCBP1,NCAPG2,NA38,MFGE8,MEGF6,MATN2,MARCHF7,MAP3K5,LUZP2,LMBRD2,LIP1,LINGO2,LEMD1,IFFO2,HSPA8,HIVEP1,HCFC2,GJA9,FXD2,FLT4,FKBPL,FKBP7,FILIP1,FARP2,ENSG00000274944,DIS3L,CYB5D1,CTSL,CLMP,CLDN11,CD302,CCL28,C5orf34,BMP10,ATP1A4,ATF6B,ANGPTL2,ALPK1,AGBL4,ADARB2,ACTR8,ABCG1
15	ZNF774,ZNF684,ZFP69,XXYLT1,WNK2,VNN1,VEGFA,USP5,UBE2F-SCLY,UBE2F,UBD,TSPAN18,TRABD,TNC,TMPRSS13,TM7SF3,SUSD3,STK25,SLC41A2,SLC17A8,SKIL,SIGLEC1,SETD9,SCEL,RRAGC,RNASET2,REP15,RARG,PUS7,PSMD14,PRKCI,PRKAA2,PIGH,PEX12,OR21P,NLRP7,NKAIN1,NDPIP1,MYCBP,MMP8,MMP27,MIER3,MAPKAP1,ING2,IL15RA,IKZF5,HMGB4,H3-5,FZD3,FPGS,FOX3,FJX1,FGFR1OP2,FAM114A2,EXO5,ENSG00000285218,ENSG00000283877,ENSG00000249141,EEF1A1,DRI1,CH1,DNAI7,CXCL12,COX7A2,COMMD2,CLEC4E,CLEC4D,CDC43,CD74,CD36,CD226,CCND2,CAND1,BRD1,ATG5,ARMS2,ARHGEF3,ANKUB1,ALDH1L2,ACKR4
14	ZNF823,ZNF740,ZNF248,WDR64,WDR49,TRIT1,TRGV10,TMEM72,TMEM268,TMEM14A,TMCO4,TEKT3,TGDF1,TBC1D32,SUSD4,SPSB4,SOAT2,SLC38A1,SHANK2,SGC2,SEC23IP,SCLY,SATB1,RTP3,RPL30,RICTOR,RASGEF1A,RAMP1,PTCD3,PSD3,POLR1A,PLCZ1,PDE4D,PCNX2,NTRK3,NOTCH4,MYO1G,MYMX,MCMBP,MAPK11,MAN1A1,LRRC2,LIPA,LEFTY1,KRT71,KLF18,KLF17,KCNH7,ITGB7,INPP5F,IMMT,IFNE,GSTA2,GPSM3,GPR142,GPBP1,GM2A,GKN1,GAL3ST2,FOXO2,FBXL3,FAT4,FAM78B,FAM120A,ERI3,ENSG00000288711,ELAC2,DSCAML1,DOCK2,DMAPI1,DIRAS2,DDAH1,CTDP1,CSGALNACT2,C5AD,CPB1,CAM3,COL19A1,CLN5,CEEC3A,CEP43,CDK9,CCSER1,CCR4,CCDC69,CACNA2D1,BLACAT1,BAG3,ATXN3,ARID3A,ADAMTS14,ACBD3
13	ZNF11,ZNF717,ZNF484,ZNF25,ZCRB1,YIPF1,VEZT,TYW1,TMEM204,TMED4,SUGCT,STAT1,SPOCK3,SLC47A1,SHC2,SH3TC2,SESTD1,RRH,RRBP1,RET,RCAN1,PTPRF,PRKN,PRICKLE1,PREX1,PRDM2,PPHLN1,PLK2,PHF2,PDLIM5,PALM,OLIG3,OGDH,NOPC,HAP1,NFIA,NEUROG3,NEU4,NECTIN1,MYNN,MTMR6,MSR1,MGST1,LRRC34,LRRC23,LGMN,KRAS,KCNB1,IFNA1,IFIT1B,IFIT1,HS12B,GRK7,GRB10,GAPT,GALNT9,FYB1,FRMD4A,FRK,FOXPA,FASTKD1,FAM120AOS,ETFRF1,ENSG00000285558,ENSG00000282218,ENO2,EIF4B,DYNC2H1,DIO1,DDX56,DDX27,COPS9,CNOT11,CDH20,CCR6,CCM2,C1orf174,C1orf63,C1orf143,BSX,AVP,ATN1,ARHGAP33,ARHGAP31,ARHGAP25,ANO4,ADCY1,ACTRT3,ACOT11,ABHD12B
12	ZNF782,ZNF510,ZNF441,ZNF22,ZGRF1,ZBTB32,ZBED4,WWTR1,WSB2,WDR27,VGLL2,UNC79,UNC13C,U2AF1L4,TWIST1,TUSC3,TTI2,TRGV11,TRGJP1,TPCN2,TMEM30A,TMEM247,TINCR,TCOF1,TACSTD2,STAC,SLFN12L,SLC66A1,SLC36A2,SGSM1,SERPINI1,SCG3,SABF2,RTN4R,RPL3L,RNF17,RNF149,PTPDC1,PSMA2,PSENEN,PRTFDC1,PROSER3,PRMT8,PRKCZ,PRKCE,PPIG,POP4,PLBD2,PHOSPHO2,PHC3,PDE11A,PCYOX1L,OTOS,OR2H2,NUTM2G,NUP58,NLRP4,NDUFA12,MT01,MS4A1B,MRPL5,MRPL35,MRPL32,MMP10,MMP1,MCUB,MAK16,LYPD6B,LRT3,LRCOL1,LPCAT2,LIN37,LEXM,LDLRAD4,LDB2,LARP7,KMT2B,KLHL41,KISS1R,KHDC3L,KCNQ4,KCNK5,ITPR3,ITGA11,IL5RA,IL17B,IGFLR1,IFIT5,IBA57,HSPB6,GUK1,GPR160,GPC6,GNAT3,GJC2,GABBR1,FUT10,FERMT2,FERD3L,FCER1A,FAM120B,EYA2,ERICH5,ERCC6L2,EPRS1,EPB41L3,ENSG00000288715,ENSG00000285733,ENSG00000267120,ENSG00000261732,ENSG00000256646,ENSG00000251569,ENSG00000188223,ELMO1,DPPA5,DL1,DHRS12,DDX60,CXCR5,CTSV,CREL2,CRAMP1,COX8C,COQ6,COMMD9,CITED4,CFAP54,CFAP210,CENPJ,C1orf35,BTBD7,BFSP1,BBS5,BACH2,ASCL1,ARF1,ANKRD27,ANKRD11,ALG12,AGAP1,ADA2,ACAP2
11	ZNF491,ZNF385B,ZNF365,ZFYVE26,ZCCHC17,XCR1,WNT3A,WDFY2,VPS53,UQCC2,TRPM3,TRIP11,TPI1,TP53BP2,TOMM40,TNR1,TNFSF15,TMCO3,TIPIN,TIAL1,TBRG4,TAF1,STAT4,STAR3D3NL,SPTL1,SPSB2,SPRR2G,SPC25,SNRNP40,SLC13A3,SIDT1,SH2D4A,SFRP4,SARAF,SAMD5,RTKN2,RFX2,RFC5,RDH12,RAI14,PYGO1,PLXDC2,PIGF,PDCC10,PCSK9,PAX7,PACRG,P3R3URF-PIK3R3,OBSCN,NLRP8,NEIL3,NECTIN2,NACAD,MSA48,MRO,MMP28,MMP2,MLN,MEX3C,MARCHF1,MAPK8IP3,MAP7,MAP2K1,MALRD1,LTF,LRRK2,LRRFIP1,LIPN,LIN52,LELP1,KRTAP5-7,KRT78,KLHL9,KIRREL3,KDSR,JPT2,IP6K3,IGDCC3,IFNA8,IFNA5,IFNA2,IFIT3,IFIT2,ICE1,HSPA5,HMX3,GRPEL2,GKN2,GAS2,GA PVD1,FKBP5,FHD03,FBH1,EXO1,EPDR1,EP400,ENTPD5,DPP6,DDX6,DCUN1D2,DBX1,CYP7A1,CYP4B1,CYL2,CRIPT,CRACR2A,CPXM2,COL6A6,COA7,CNTNAP2,CHI3L1,CHDH,CCDC185,CASP6,CAPRIN2,C1orf52,C1orf50,BBOF1,ATP2B4,ARHGEF12,ARG2,ANXA6,ANKRD16,ALDH6A1,ADTRP,ACTL8,ABC85
10	ZNF440,ZNF439,ZMYND8,XAF1,WNT9A,VTI1B,UTP20,ULK2,UBXN2B,TSPAN1,TRPC7,TRIM44,TRIM42,TRGJP2,TRGJP,TRGJ2,TRGJ1,TRGC2,TRGC1,TRAPPC4,TRAF3IP1,TPST1,TLCD5,TDFP1,TARP,TAAR1,ST8SIA3,ST3GAL5,SNCAIP,SMCO4,SLC37A4,SLC29A1,SLC25A25,SLC15A5,SERINC2,SEC16B,SCN4B,SCN1A,SCAI,RPS25,RPS14,RNF7,RAPGEF4,RABGEF1,RABEPK,PTGES2,PPP6C,PLA2G12A,PIWIL1,PHACTR2,PDE7B,PCDH9,PANX2,P3R3URF,P3H3,OR6C4,OR10J3,OAS3,OAS2,NAI1,NEK11,NAIF1,MYO1B,MYH13,MTX2,MOV10L1,MLNR,MLC1,MAPRE1,LYZL6,LYSMD2,LRRC31,LMAN1,LEMD2,LCP1,KRT76,KCTD7,IFNA6,IFNA13,HHPIL2,GZMH,GCA,G6PC2,FRG2C,FOXR1,FNDC3A,FRD2,FBLN5,FAM83B,FABP3,ENSG00000284461,ENSG00000249319,EGLFAM,DU SP10,DPY19L3,DMBT1,DLX2,DDX11,DAB1,CYTIP,CRYL1,CRCP,CENATAC,CDH22,CDADC1,CCSER2,CCDC25,CAMK1D,CAB39L,C11orf96,C10orf90,BCL2L14,BCL10,BANK1,ATP8A2,ANKRD44,ALDH3A1,AKNA,ADAMTS16,ACTRT2,AC114494.1
9	ZHX2,YAF2,VXS2,VEPH1,UPK1B,ULK1,UGT3A1,TTC21B,TRPV3,TRGV9,TRGV8,TPST2,TMEM260,TMA16,TKTL2,TDRD9,TAF15,STXBP6,SPATA31E1,SMIM1,SLC6A20,SLC36A1,SLC22A2,SLC17A5,SHISAL2A,SEPSECS,SDS,RTP5,RNF112,RFTN1,RDM1,RDH11,RBM47,RBM44,RASA3,R3HDM4,PRSS54,PRSS37,PRR16,PROX1,PRODH2,PLXNB2,PIGR,PDGFD,PDE4B,PDCC1,PCNX4,ORM2,ORM1,OR9A4,OR9A1P,OR6C76,OR6C75,OR6C70,OR6C68,OR6C65,OR6C2,NUDT16,NOC4L,NEDD1,NCKAP1L,MYOZ3,MYBPH,MMP13,MIX23,MFAP4,MAPK7,MAP3K20,LUC7L,LRR1Q4,LRRC9,LRRC63,LRP8,LMBRD1,KRT79,KLRD1,IPO8,IGSF21,HOXC4,HEATR9,H BZ,HBOQ1,HBM,HBA2,HBA1,GINS4,GCM1,GALNT13,GABRB2,FAM24A,FAM240C,FAM234A,FAM184B,FAM162A,FAM107B,EPMD2A,ENSG00000287725,ENSG00000258466,ELOVL5,EBF3,DMBX1,DLGAP1,DLG2,DISC1,DHRS7,DGKH,DFFB,DEPP1,DDX50,DDX21,DAPL1,DAOT1,CYP4A11,CSNK2A3,CORO2B,CNIH3,CLEC5A,CFAP97D2,CES5A,CCR1,CCL5,CCDC175,CCDC113,CARD19,CAMKK2,CACNG6,C1orf94,C10orf88,BTN3A3,BTN3A2,BTN3A1,BTN2A2,BTG3,BOK,BLM,BCL11B,B4GAL74,ATP1B3,ATP1A2,ARNTL2,ALG10,ADORA1,ADGRV1,ADGRA3,ABCD4
8	ZNF735,ZNF713,ZNF679,ZNF598,ZNF597,ZC2HC1B,ZAP70,YAE1,VSIG10,USP44,USP36,USP24,USF2,TYSND1,TSPAN33,TRMO,TMEM30B,TDDR12,TBL3,TATDN2,TACR2,SYNGR3,SV2C,STAT2,SRPK2,SPINK6,SPINK5,SNX5,SMIM12,SMAD6,SLC7A9,SLC7A2,SLC39A5,SLC35C2,SLC2A8,SLC1A3,SLC3A3,SHPRH,SEZ6L,SERPIND1,SEC63,SACM1L,RUNX2,RTN1,RHCG,RGS9BP,RGS5,REEP1,RAB32,PUS1,PSEN2,PLCH1,PKIB,PI4KA,PHLDB2,PHC2,PDE8A,PDE5A,PDCC5,PCED1B,PCDH20,PBK,PAN2,OSTM1,OR10G3,NUDT19,NR3C1,PNW,NOXO1,NMNA2T,NME3,NIPSNAP2,NINJ1,NEUROG2,NEDD4,NCAM2,NCALD,NAV3,NAGK,NADSYN1,NAAG60,NAAG30,MYCL,MUC19,MSRB2,MRPS34,MRPS17,MPHOSPH10,MLIP,MAI3,MGME1,METAP2,MED21,MCEE,MAPK9,MAPK4,MAGOH,L RRC47,LNPEP,LASP1,KMT2E,KLHL23,KIFBP,KHDC1L,KHDC1,KCTD16,INTS7,INTS6,IL23A,IGSF11,IFNA7,IFNA4,IFNA2,IFNA17,IFNA16,IFNA14,IFNA10,GXYLT1,GRHL2,GNB3,GNB1L,GLS,GFER,FLG,FCAMR,FBXO45,FBXO30,FAT3,FANCF,FAM161B,FABP2,EVA1C,ESCO2,ENSG00000289360,ENSG00000288914,ENSG00000272410,ENSG00000259164,ENSG00000249773,ENSG00000244255,ENSG00000243501,ENSG00000197991,ENSG00000144785,EME2,ELMO2,DTL,DSTN,DMXL2,DLG5,DDX51,DAAM1,CZIB,CYTH1,C S,CPT2,COQ10A,COL6A3,CNTN4,CNPY2,CNOT10,CMAS,CIPC,CHD1L,CFB,CEP89,CD42BPA,CDQ16,CD109,CCDC180,CCDC141,C4orf3,C2,C1orf116,C12orf71,BUD13,BRF1,BCL9L,BARX1,ASTE1,ASGR2,ARHGEF28,AQP12B,AQP12A,APOF,APLF,ANO3,ANKRD52,ALKBH3,AK1,AGFG1,ADRA1A,ADCY10,AC005606.1,ABHD4

Frequency	Genes
7	ZPR1,ZNF79,ZNF678,ZNF334,ZNF322,ZDHHHC8,ZDHHHC22,WDR86,WDR53,WASHC4,VWA3B,USP2,UPF3A,UBQLN1,UBE2C,TTC23 L,TRMT2A,TRGV5,TRGV4,TRGV3,TRGV2,TRGV1,TRDN,TRAC,TRABD2B,TNNC2,TM4SF1,TFDP2,TRDR7,TCP11L2,TCERG1L,TBP L1,TBC1D17,TAGLN3,SYT16,SULT1E1,SULT1C4,STK38L,STK19,STAM,ST8SIA2,SPTAN1,SPINK13,SP8,SNX21,SNRPA1,SNAP47,SNAP29,SMOC2,SMO,SLCO3A1,SLC6A2,SLC39A11,SLC35A5,SLC2A12,SLAMF1,SKIV2L,SIK3,SELENOS,SCGB3A2,SCGB2B2,SB DS,RPL12,RPH3A1,RNF5,RGL4,RBM22,RBM20,RASEF,RANBP1,PRRX2,PROKR1,PRIM2,PPT1,PPP2R2B,POLR3B,PNKP,PLEKHA 3,PLA2G2D,PIP4K2C,PIGM,PHF21B,PEX3,PERP,PDGFRB,PALLD,PAK5,ORC5,OR5AN1,OR10J4,OR10J1,OMD,OGN,OAT,NUGGC,NUDT14,NTMT1,NRDC,NOL8,NME4,NLRP11,NIBAN2,NELFE,NDRG4,NAB1,MS4A12,MMP3,MLPH,MKRN1,MFSD5,METAP1D,MDN 1,MAG,LSR,LOXHD1,LMNTD1,LIMS1,LAX1,LAMP5,L3HYPDH,KIF5A,KCNIP4,JMJD4,JKAMP,JAG2,IRX6,IQSEC3,IFNGR1,IFNB1,IDN K,IARS1,HSD17B12,HMGNA4,HEMGN,HAMP,GRIK3,GPR162,GPR141,GPR135,GPM6A,GOT2,GOLGA1,GLO1,GLE1,GCC2,GAREM1 ,FUCA2,FOX1,FOCAD,FECH,FBXO9,FBXO38,ETV6,ETNK1,ENSG00000289503,ENSG00000288646,ENSG00000288529,ENSG000 00286135,ENSG00000284695,ENSG00000284686,ENSG00000283536,ENOX1,EGFL8,ECM2,DXO,DTX3,DPH6,DGCR8,DECR2,DCT N4,DAOA,CYP4Z1,CYP4X1,CYP21A2,CRKL,CREBBP,COQ8A,CLDN8,CEP104,CENPP,CDCP1,CD82,CD22,CCR3,CASP8AP2,CAP1 ,CADPS,C9orf50,C5orf46,C4B,C4A,C19orf12,C10orf120,BTN2A1,BTN1A1,BTLA,BTBD6,ATP1B1,ATG3,ASPN,ASB6,ARSI,ARHGEF25 ,APPL2,APOC3,APOA5,APOA4,APOA1,ANKRD31,ALDH4A1,AL645922.1,AKT1S1,AKAP13,AKAP12,AJAP1,AGPAT1,AGER,ADAMTS L3,ABTB1,ABT1,ABHD10
6	ZRANB1,ZNF335,ZNF302,ZC3H11A,ZBTB8A,ZBTB38,ZBED6,YIPF5,WFDCC3,WDR11,VIPR1,UNC45B,UMAD1,UHRF1BP1L,UBXN7,U BE3D,TXNDC15,TVP23C-CDRT4,TSPYL5,TSPAN8,TRMT61A,TRIM40,TRIM29,TRIM26,TRIM15,TRIM10,TMCO1,TM9SF2,TIMP2,TEX55,TCL1B,TANK,TAF1A,T AAR2,STOX2,STIMATE-MUSTN1,STIMATE,ST6GALNAC4,SPRR2A,SPINT4,SPINK9,SPINK7,SPICE1,SPECC1,SLIT2,SLC8A3,SLC5A8,SLC39A12,SLC36A3,SLC35G3,SLC22A3,SLC12A5,SH2D3C,SFMBT1,SEC61A2,SEC24B,SDK2,SCN2B,RWDD4,RPS6KC1,RPRM,RNASE2,RINT1,RFT1,R D3,RBM45,RAX,PRSS38,POLK,POC5,PLEKHF1,PLAAT1,PCIF1,PCBD2,PBX2,PAXIP1,PARP1,PAMR1,PAK1,SH3GLB2,SH3BGR1,OR5A1,OR4D6,OR4D10,OR10J5,OLFML2B,OCSTAMP,NRF1,NPY1R,NMUR2,NEB,MYO5C,MVB12B,MUSTN1,MTAP,MRPL3,MOB3 B,MMP9,MMP12,MLLT3,MISP,MGLL,MGAM2,MFSD2A,METTL17,MEOX1,MARK3,MARCOL,MAP3K21,MACC1,LTV1,LRRC52,LPAR1 ,LOXL2,LAMC2,KRTAP5-9,KRTAP5-8,KRTAP5-10,KIAA0895,KDM3A,KCNJ10,JPH2,JAML,ITIH4,ITIH3,ITGAE,ISM2,IL7R,IFNW1,HMHB1,HMGC,RHERC1,HACD4,GPR35,GPR161,G OLIM4,GOLGA7,GNG12,GNAI1,GLI3,GDAP1L1,GCNT4,GABRA6,FXYD4,FRMD1,FPR2,FOXN3,FOXG1,FCRLB,FCRLA,FAM240A,F1 3A1,EPHB2,ENSG00000289763,ENSG00000289653,ENSG00000288644,ENSG00000285943,ENSG00000272305,ENSG0000025455 3,ENSG00000251012,ENSG00000243696,DYNC2I2,DYMKO,DPD4,DNTTIP1,DNAH5,DLK1,DHCR7,DERA,DCTN6,DCTN2,DCBLD1,D APK2,CYPA42,CTBP2,CREBL2,COL28A1,COL16A1,COLB,CNOT4,CMTM7,CLBA1,CKB,CERT1,CERS3,CEP170B,CELF2,CDRT4,C DKN2AIP,CDCA4,CCRL2,CCR5,CCR2,CCL3,CATSPER3,CASQ1,CAPSL,CAPN13,CAGE1,BRPF3,BANF2,BAK1,AZIN1,ATP13A5,AR PP19,ANKMY1,ANKDD1B,ANGEL2,AKAP11,ADCK1,ADA,ACYP2,ACTBL2,ACCSL,ACCS
5	ZSWIM3,ZSWIM1,ZSCAN2,ZNF829,ZNF80,ZNF599,ZNF568,ZNF487,ZNF420,ZNF181,ZFYVE28,ZFYVE19,ZBTB8B,XRCC4,WWC2,W DR73,WDR38,WDR20,VPS16,USP3,UPK2,UNC5A,UCK2,UBR7,UAP1,TTYH2,TTC7B,TTC16,TRMT5,TRAV1-2,TRAM2,TOR2A,TNFRSF1B,TNFRSF10A,TMTC4,TMOD2,TMEM251,TMEM196,TMEM131,TMEM117,TM4SF19-DYNLT2B,TM4SF19,TLN2,TLCD3A,TIE1,TEX53,TEX48,TESK2,TEKT1,TCF7L2,TBX3,TBC1D16,TASOR2,SYN3,SYCE1L,SULT1C3,S TXBP1,STX11,ST8SIA1,ST6GALNAC6,SRGAP3,SPSB3,SPRR2F,SPRR2E,SPRR2B,SPINK14,SPATA25,SP4,SP3,SMPDL3A,SMPD3, TMX31,SMAP2,SLFN11,SLC51A,SLC38A6,SLC38A11,SLC30A1,SLC29A3,SLC16A13,SLC16A11,SKA1,SH3GLB2,SH3BGR12,SETD6,SEC11A,SCYGR1,SCN2A,SCHIP1,SAR1A,RWDD2A,RPS6KA2,RPS12,RPL39L,RNLS,RNASEK-C17orf49,RNASEK,RLF,RIPOR2,RFX3,RCS1,RASGRP3,RANBP2,RAB37,R3HCC1,PTX3,PTRH1,PTK7,PTK2B,PRSS57,PRSS41,P ROM1,PRMT7,PRDX1,PRDM1,PPP1R9A,PPP1R14D,PPA1,POP1,POMT2,PLTP,PLSCR1,PLCXD2,PLCH2,PLAGL1,PLA2G2E,PIK3R 3,PIGN,PGM3,PDE6H,PCYT1A,PCED1A,P2RX4,OVOL2,OR5A2,OR4E2,OR4E1,NUP188,NUBP2,NUAK1,NTM,NMB,NKX6-1,NKX1-2,NGB,NFE2L2,NEURL2,NCAPG,NASP,MTAP,MRM3,MPZL3,MPC2,MON1B,MOK,MOAP1,MNAT1,MMACHC,MKNK1,MICOS10,NBL1,MICOS10,MGST3,MEGF10,MED6,ME2,ME1,MDGA2,MAST2,MAPK10,MAP4K2,MAP2K7,MAB21L4,LYZL4,LRSAM1,LRRC8E,L RRC1,LRP2,LRIT2,LRIT1,LMF1,LINC00672,LGR6,LECT2,LORL1,KRTAP27-1,KRTAP26-1,KRTAP25-1,KRTAP24-1,KRTAP23-1,KRTAP13-2,KLHL30,KLHL25,KIF27,KCNJ9,ITIH1,ITGB2,IRS2,IRAK2,IQCC-SCHIP1,IPP,INSR,IMPACT,IL22RA2,IL20,IKBKE,HUNK,HSP90AA1,HS3ST2,HRH4,HNRNP,HELH,HACD1,GREB1,GPX2,GPR42,GP R19,GPR176,GPR132,GPBP1L1,GON7,GOLGA6L4,GLRX5,GLRX3,GLOD4,GLIS1,GKAP1,GHRL,GEMIN4,GDI2,GALNT17,FZD8,FIT M2,FGF9,FFAR3,FBXL22,FAM98A,FAM110A,FABP7,FAAP24,ENSG00000288674,ENSG00000288208,ENSG00000285446,ENSG000 00280148,ENSG00000272741,ENSG00000267360,ENSG00000259066,ENSG00000258989,ENSG00000257524,ENSG00000251184,ENSG00000249209,ENSG00000214558,ENPEP,EMC1,ELP3,ELMOD3,EFCAB10,EEF1E1-BLOC1S5,EEF1E1,EDN2,EDDM13,EBF2,DYNLT2B,DYNC11L1,DUSP5,DUSP28,DUSP23,DSE,DRD3,DPYSL2,DOP1A,DNAJC17,DN AH14,DLX1,DCAF6,DCAF16,CXCL13,CTSA,CPNE8,CPA2,CCOLCA2,COL15A1,CNTN3,CNOT1,CMTM6,CLIC5,CILK1,CHURC1-FNTB,CHURC1,CFAP157,CDHR1,CD84,CD4,CCL23,CCL20,CCDC17,CCDC163,CCDC138,CBLB,CAPG,CANT1,C5orf49,C5orf15,C1 7orf49,C10orf67,C10orf105,BTBD17,BTBD11,BICD2,BEGAIN,BCL6B,ATP6V1C1,ATP5PO,ATP2B2,ART4,ARRDC3,ARN2,ANP32B,A NKRD30A,ANGPT4,ANAPC5,ALOX12,AL109659.1,AKR1A1,AIG1,AHNAK2,AGXT,AFAP1L2,ADAT2,ADAM20,ACTR6,ACOT8
4	ZNF804A,ZNF763,ZNF701,ZNF700,ZNF69,ZNF438,ZEB2,ZDHHHC19,ZBTB40,WWC1,WHRN,WBP4,VNN2,VIPR2,UQCRH,ULK3,UBR4 ,TTC30B,TRMT1L,TRIM35,TREH,TRAV1-1,TRAJ9,TRAJ7,TRAJ6,TRAJ5,TRAJ4,TRAJ3,TRAJ26,TRAJ25,TRAJ24,TRAJ23,TRAJ22,TRAJ21,TRAJ20,TRAJ2,TRAJ19,TRAJ18,TR AJ17,TRAJ16,TRAJ14,TRAJ13,TRAJ12,TRAJ11,TRAJ10,TRAJ1,TRAF6,TRAF5,TPCN1,TMEM69,TMEM26,TMEM209,TMEM183A,TM EM167A,TMEM127,TMEM120B,TMED8,TLE7,TINAG,TIGIT,THEM5,TEX38,TENT5A,TECPR2,TCF7L1,TBC1D5,TARBP1,TAGLN,TAB 1,SZT2,SYT2,SYNGR1,SYNE3,SWT1,SUSD2,SULT1C2,STRN3,STRAP,STIL,STARD7,SSMEM1,SRPK1,SRGAP2,SQOR,SPRR4,SP RR3,SPRR2D,SPRR1B,SPRR1A,SPIRE1,SPATA5,SOX8,SNX19,SNRNP200,SMCO3,SLC8B1,SLC38A7,SLC35F1,SLC2A9,SLC26A8,SLC25A36,SLC18B1,SLC15A1,SKA3,SIM1,SIDT2,SERPINA6,SERPINA2,SERPINA1,SEMA6D,SEC61B,SEC13,SCYL2,SCGB2A,SC GB1D4,SCGB1A1,SCAMP2,RSU1,RPP25,RPL35,RNPEPL1,RNF2,RNF166,RNASE3,RMDN2,RIN2,RIMS3,RIMS2,RIMOC1,RHOJ,SC OH,RECQL,RALA,RAD54L,RABIF,QSOX2,PYROXD1,PURB,PTPRA,PRR5L,PRIMA1,PPP1R2,PPM1A,PIPA,PPARD,PON3,PON2,PO N1,POMGNT1,PLXNC1,PLEKHG4B,PLD4,PIK3R4,PHLDB1,PHGR1,PGLYRP4,PGLYRP3,PEF1,PCSK7,PCSK5,PATJ,PAQR8,PAFAH 1B2,OXCT1,OSBPL1A,ORC4,ORAI1,OR8S1,OR5BS1P,OR13A1,OR10G2,ODF3L2,OASL,NUDT5,NSUN4,NLRP13,NFE2,NFATC1,NE URL3,NCAPH,NBL1,N4BP2,MYCBP2,MTHFD1L,MTARC1,MRT04,MRPS26,MRPL57,MRPL2,MPL,MPI,MORN3,MCOB3C,MIXL1,MFSD 4A,METTL8,MED8,MAPK14,MAP3K7,LURAP1,LT44H,LRRC41,LRP1B,LORICRIN,LIN9,LHX4,LHX3,LHFP15,LCE5A,LCA5,LBR,LALB A,KRT8,KRT39,KNSTRN,KLHL14,KLHL12,KLC4,KIF19,KDM2B,JAZF1,IVNS1ABP,IVD,ITPR1P1,ITPKB,IQCF5,IQCF3,IQCF2,IPPK,IL9 ,IL24,IL17REL,IL17A,IFTAP,HY1,HNRNPA1,HEATR5A,HCRT1R1,HCN2,HAS2,HAL,H2AZ2,GSTA4,GRIP1,GRIK2,GRAMD2B,GPX7,GP R26,GPC5,GOPC,GOLM1,GNRH2,GNB5,GALNTL6,GALNT3,GALNS,FXYPD7,FXYPD1,FSTL4,FMOD,FLT1,FGF2,FFAR1,FCMR,FBXO4 ,FAM219B,FAM177B,FAAH,ESR1,ERAP1,ENSG00000288716,ENSG00000288684,ENSG00000288614,ENSG00000288564,ENSG000 00286231,ENSG00000285749,ENSG00000283228,ENSG00000267699,ENSG00000267179,ENSG00000249624,ENAH,ELOV1L1,ELF1 ,ELAC1,EIF2AK4,EHFC1,EFCAB14,EEP1D,EDIL3,EDEM3,DYNC11I1,DYNC11H1,DOCK1,DMRTB1,DISP2,DHTKD1,DEPDCA,DENN5 B,DEF6,DDX58,DCUN1D5,DAZL,DAW1,CYP4F8,CUL7,CRCT1,CPA4,COX5A,COPZ1,COL3A1,COBLL1,CMPK1,CLPSL1,CLIP1,CIAO 1,CEP295NL,CCDC20,CD9,CCSAP,CCL15-CCL14,CCL15,CCL14,CCDC88C,CCDC149,CBX5,CASZ1,CAPN15,CACNB4,CABIN1,C1orf87,C12orf60,C12orf54,BTBD9,BROX,BPIF A3,BOD1L2,BIRC3,BCL2L10,BAALC,B4GALT5,ATPAF1,ATPB1,ATPB1G1,ASRGL1,ASCC3,ARPC5L,ARMC12,ARL5A,ARHGDI1B,A RFGEF3,APCS,AP5S1,ANP32D,ANGPT2,AMER2,ALG2,ALG10B,AKR7L,AKR7A3,AKR7A2,AK4P3,AIDA,AHRR,AGTR1,AGPS,ACKR3 ,AAK1
3	ZNHIT6,ZNF850,ZNF839,ZNF790,ZNF777,ZNF76,ZNF662,ZNF618,ZNF567,ZNF529,ZNF461,ZNF410,ZNF382,ZNF367,ZNF345,ZER1, ZDHHHC12,ZC3H11B,YBX3,XPO7,WEE2,WDR7,VDR,VDAC1,VAT1L,VANGL2,UST,UHMK1,TXNDC16,TTC39A,TTC33,TTC30A,TTC3, TSTD3,TSPAE,TRPC22D2,TRPV1,TRIML2,TRIML1,TRIM9,TRIM66,TRAPPC2L,TPD52L1,TNIP1,TNFAIP8,TNFAIP3,TMPPE,TMOD3, TMEM82,TMEM163,TMEM123,TMEM106C,TKT,THTPA,TGFB1,TFRC,TCF4,TBRG1,TBC1D8,TBC1D13,TAX1BP1,TATDN3,TAS2R5,T AS2R4,TAS2R38,TAS2R3,TAL1,TAFA3,SYNE1,SVIP,SUMF2,SUMF1,STXB5P,STX18,STK32C,STAB1,SSBP1,SRGN,SPX,SPRR5,SP OUT1,SPNS3,SPATA6,SPATA45,SOX5,SNX30,SNRNP2B,SNAI3,SMIM29,SMIM21,SMCO1,SMARCB1,SMAD4,SLX4IP,SLITRK1,SLC 1B1,SLC7A10,SLC6A3,SLC5A7,SLC5A12,SLC47A2,SLC2A6,SLC2A13,SLC2A1,SLC25A34,SLC25A15,SLC19A2,SLC17A3,SLC17A2,SLC17A1,SKOR1,SIGLEC8,SET,SERPINI2,SERPINA3,SERPINA10,SEPTIN9,SDHA,SD2,SDC3,SDC2,SCUBE3,SCGB2A1

Frequency	Genes
	,SCGB1D2,SCGB1D1,SCARA3,SAP18,RSL24D1,RSC1A1,RRP9,RRP36,RPUSD2,RPS24,RPS2,RPL37,RPL36AL,RPL29,RPAP3,RNF34,RNF168,RNF151,RNASE1,RMDN3,RGS7,RGS6,RGL1,RFX4,RESF1,RCOR3,RASSF2,RAPGEF3,RAG2,RAG1,RAD51,RAB4A,Q TRT2,PTPRB,PTGR2,PTGER4,PTF1A,PSPH,PSMG2,PRTG,PRNP,PRKAA1,PRDM4,PRDM13,PPP4R4,PPP2R5D,POLR3A,POLE2,P LEKHM2,PLEKHJ1,PLCXD3,PKN3,PIWIL4,PHYHD1,PHKG1,PHF5A,PEX7,PEX6,PDCC6-AHRR,PDCC6,PCBP4,PARP3,PANX3,PABPN1L,NUPR2,NUDT6,NTNG1,NTN4,NSL1,NR3C2,NPM2,NNAT,NME7,NLGN1,NISCH,ND UFB10,NC51,NCOA5,NCF2,NAF1,MXI1,MUC13,MTRF1,MRPS31,MRE11,MPC1,MMP11,MLLT1,MIS18BP1,MIOS,MIGA2,MICU3,MG AT2,MEIOB,MEA1,MBD1,MAPK6,MAPK13,MAP2K5,MAML2,LZTFL1,LZYL1,LSM14A,LRRC8A,LRRC14B,LRR1,LRP3,LPIN1,LIPK,LH FPL4,LGI4,LGALS3BP,LEO1,LDLRAP1,LBHD2,LARP1,LAPTM4B,KYAT1,KRTAP5-11,KRT80,KRBOX1,KNL1,KMT2A,KLHDC3,KIN,KIF11,KCTD9,KCNMA1,JRKL,JPH4,JDP2,ISG20L2,IQCH,IQCD,INIP,IL1RAP,IDE,HTT ,HSD17B4,HOOK1,HMGA1,HLX,HIBADH,HHEX,HFM1,HASPIN,HAGH,HABP4,GULP1,GSTZ1,GSAP,GRB14,GPR83,GPR62,GPR137 C,GPC1,GOLT1B,GNRH1,GNMT,GLRA1,GFPT2,GFOD1,GCHFR,GBA3,GASK1A,GADD45G,G3BP1,FYCO1,FXYD5,FXYD3,FOXO1,F LVCR1,FLG2,FGF17,FGD4,FBXO39,FARS2,FANCM,FANCC,FAHD1,EXOC3L4,ERP27,ERCC2,ENSG00000289722,ENSG000002890 25,ENSG00000286112,ENSG00000286001,ENSG00000285827,ENSG00000285526,ENSG00000285269,ENSG00000273291,ENSG0 0000258653,ENSG00000257065,ENDOU,ENDOG,EHD3,EGR3,DOT1L,DOLK,DOK2,DOCK5,DOCK10,DNAJC16,DNAI2,DNAH9,DNA AF2,DLCL1,DISP1,DIPK1C,DERL3,DDI2,DCP1A,DCAF17,DAP,CYP2J2,CYBRD1,CYB5R1,CXXC1,CXCR6,CFRN,CR1L,COTL1,COLE C12,COL6A5,COL2A1,COG2,CNPY3,CLPS,CINP,CHSY3,CHSY1,CHST8,CHST14,CHCHD2,CFAP57,CEP76,CENPF,CDK1 8,CD34,CD163L1,CCT6A,CCR9,CCNC,CCL18,CCL16,CCDC82,CCDC32,CCDC191,CCDC181,CCDC177,CCDC127,CBFA2T3,CASP9 ,CARD6,CACNG8,CACFD1,C6,C3orf33,C15orf62,BTG2,BTBD2,BSG,BPIFA2,BLZF1,BLCAP,BIRC2,BICD1,BHLHE41,BCLAF1,BCAT1, BATF3,BAHD1,B3GALT5,ATXN7L3B,ATP2A3,ATP12A,ATOX1,ASCL4,ASB1,ARPC5,ARHGFE7,ARHGAP10,APOBEC4,AP4S1,AP1G 2,ANKS6,ANKRD50,AMDHD1,ALK,AKAP6,AKAIN1,AGPAT4,AGMAT,AFAP1L1,ADRB2,ADIPOR1,ADGRB3,ADCY9,ADAMTS5,ADAM TS13,ADAM12,ACY1,ACP3,ACO2,ABHD14B,ABHD14A-ACY1,ABHD14A,ABCC4,AADACL3
2	ZNF865,ZNF827,ZNF746,ZNF616,ZNF615,ZNF37A,ZNF33A,ZNF217,ZNF189,ZMIZ2,ZFYVE21,ZFXH2,ZDHHC20,ZC3H12A,ZBTB41, ZBTB22,ZBTB2,ZBBX,XRCC3,XPR1,XNDC1N-ZNF705EP-ALG1L9P,XDH,WIPF3,WDR89,WDR46,VWF,VWA8,VPS52,VIPAS39,VAV1,VARS1,USP45,URM1,URGCP-MRPS24,URGCP,URB1,UCHL5,UBOX5,UBE2T,UBE2D4,UBAP1L,U2SURP,TXNRD2,TXNL1,TXLNB,TUFT1,TTC36,TTBK1,TSPAN31,TSPAN17,TSMF,TSEN2,TRUB2,TRPC4,TRMT44,TRIM38,TRAV39,TRAF3IP2,TPH2,TNS4,TNKS1BP1,TNFRSF11A,TMEM63C,TMEM 275,TMEM25,TMEM192,TMEM179,TMEM121,TMEFF2,TMCC3,TLR5,TIAM2,TIA1,THUMP3,TGFB2,TGFA,TSE9,TESPA1,TED C1,TDRKH,TCERG1,TBX15,TBX1,TASP1,TAPBP,SWI5,STK33,STIM2,STAT6,ST8SIA5,ST18,SSRP1,SRSF6,SPRTN,SPNS1,SPG7,S PARC,SOC54,SNX27,SNRPG,SNIP1,SMIM41,SMIM40,SMIM34,SMIM11,SMC3,SMC2,SMARCC2,SLCO1B3-SLCO1B7,SLCO1B3,SLCO1A2,SLC9C1,SLC6A7,SLC6A6,SLC5A4,SLC51B,SLC45A3,SLC36A4,SLC35B3,SLC27A4,SLC25A29,SLC2 5A21,SLC2A5,SLC22A4,SLC22A18,SLC1A2,SLC16A10,SLC11A1,SKOR2,SKI,SIX6,SIX1,SIGMAR1,SHPK,SH3GL3,SH2D6,SGPP1, SFTPC,S2F3A1,S3F3A2,SETD5,SERPINB6,SEPTIN14,SENP5,SEMA4D,SEMA3A,SEL1L2,SDHAF4,SCYL3,SCYGR4,SCYGR3,SCYGR 2,SCYGR10,SCN8A,SCN3A,SCGB3A1,SCG2,SAMD15,RTCB,RSP01,RPS6KA5,RPS29,RPS18,RPP25L,RPL13,RPA3,RORC,RNF41, RNF26,RNF11,RNASE12,RNASE11,RMND1,RIIAD1,RIC8B,RHEX,RGS3,RGS13,RGS1,RGR,RGL2,RFX8,RFXT,RFTN2,RFPL3,RFPL 2,REV3L,REPS1,RELCH,RD3L,RASSF9,RASL12,RAP1GAP2,RABEP2,RAB6C,RAB3IP,RAB27A,RAB19,PWP1,PTPRK,PTPRG,PTPA ,PROSER1,PRND,PRLR,PRKCD,PRELP,PRELID2,PPP1R13B,PPP1R12B,POPDC3,POLR1F,PLSCR4,PLD5P1,PLCB4,PLBD1,PLA2 G4A,PKD1L1,PIP5K1C,PIGZ,PIEZO1,PHYHIP,PHLDA2,PGR,PGAP6,PGAP1,PFDN6,PEX14,PEX10,PEL12,PEBP4,PDZD3,PDHA2,PD E3A,PDE1A,PCYOX1,PCGF3,PCDH18,PBOV1,PARP16,PAH,P4HA2,P2RX5-TAX1BP3,P2RX5,P2RX3,OXT,OTUB2,OS9,ORTE24,OR6C6,OR6C3,OR5V1,OR5AK3P,OR5AK2,OR13D1,OR13C9,OR13C8,OR13C5, 0R13C2,OR12D3,OPTC,OPN4,OLFML2A,OA3Z,NXPH1,NUS1,NTS,NTPCR,NSG1,NRG3,NR4A1,NR1H4,NPS,NPBWR1,NXORED1,N OS1AP,NOL10,NLRX1,NIPSNAP3B,NIPSNAP3A,NIN,NHLRC3,NFYC,NFU1,NFATC2IP,NEMF,NDUFAF5,NDFIP2,NCSTN,NCBP2AS2 ,NCBP2,NABP2,MYRFL,MYO5A,MYLK,MYL3,MYH8,MYH4,MYH2,MYH1,MTFMT,MTA1,MSRB1,MRSP6,MRPS24,MRPL9,MRPL28,M RPL22,MRGPRG,MPZL1,MORC3,MOB4,MNS1,MIPOL1,MIF,MICU2,MGP,MGAM,MFSD4B,MFSD1,METTL1,MED12L,MEAF6,ME3,M AST4,MARK1,MARCHF9,MAPK8,MAPK1IP1L,MAP4K3,MAN1B1,MAGI1,LYN,LTPB1,LSM2,LRRK1,LRRC55,LRP5,LPP,LIX1,LIPJ,LIP F,LINGO4,LINC00696,LHFPL6,LDB3,LAT,L3MBTL1,L1TD1,KRTDAP,KRTAP12-4,KRTAP12-3,KRTAP12-2,KRTAP12-1,KRTAP10- 9,KRTAP10-11,KRTAP10-10,KRT18,KLHDC2,KLHDC1,KLF6,KLF12,KLC1,KIFAP3,KIF1A,KIAA1614,KIAA0586,KDM4C,KCTD3,KCNE2,KBTBD13,JSRP1,JCAD, JAGN1,IVL,ITPR1,ITIH2,IRF2,IRF1,IQCF1,IMPDH1,IL11RA,IL10RB,IGSF9,IGLL1,IFNAR1,IFI27,IER5L,IBTK,HSPE1-MOB4,HSPE1,HSPD1,HSPB8,HSPA13,HSD17B7,HS3ST6,HOXA9,HOXA7,HOXA6,HOXA5,HOXA4,HOXA3,HOXA2,HNRNPA3,HIST1 H4B,HINFP,HILPDA,HIF1A,HHIP,HFE,HERPUD1,H4C1,H3C3,H3C2,H3C10,H3C1,H3-3A,H2BC3,H2BC13,H2AC4,H2AC14,H2AC13,H1-2,H1-1,GYS2,GUCY2C,GTF3C4,GTF3C3,GSTT2B,GSTT2,GRM4,GRIN1,GRAMD1C,GPR151,GPR107,GPHB5,GP2,GORAB,GOLGA5,GNL 2,GLYATL2,GLYATL1,GLYAT,GFPT1,GEMIN5,GCSAM,CNT2,GAS2L3,GALT,GAK,FSHB,FOXDI1,FNBP1,FLRT3,FLJ00388,FIGLA,GF 20,FER,FEM1B,FCER2,FBXO42,FBN2,FAXDC2,FAM219A,FAM184A,FAM136A,FAM135A,FAF1,F13B,FLT2,EXOC8,EXOC1L,EXO C1,EVX1,EVL,ESF1,ERC2,EPHA7,EPAS1,ENSG00000289697,ENSG00000289258,ENSG00000288520,ENSG00000288000,ENSG00 000287908,ENSG00000287856,ENSG00000285762,ENSG00000285064,ENSG00000262304,ENSG00000259171,ENSG00000259060 ,ENSG00000258728,ENSG00000258052,ENSG00000257921,ENSG00000257341,ENSG00000257184,ENSG00000257062,ENSG000 00256500,ENSG00000251357,ENSG00000235007,ENSG00000203546,ENSG00000187186,ENHO,ENDOD1,ELOVL4,ELK4,ELAVL4, EIF5,EIF4E1B,EGLN1,EFEMP1,EEF1AKMT3,EDNRB,ECT2L,ECHDC3,E2F7,DUSP7,DTWD2,DTD2,DRD5,DPPY5,DPAGT1,DOL PP1,DOCK9,DNPH1,DNALI1,DNAI1,DMRTA2,DMKN,DEUP1,DENND2B,DEGS2,DEFB108B,DDX31,DDTL,DDT,DDR2,DCTN3,DCCC,D AXD,AB2IP,CYP7B1,CYP27B1,CYP1B1,CUL9,CUBN,CTSE,CTNS,CTDSP2,CRY1,CRP,CRIP2,CRIP1,CROT,CRA2D,CRACDL,CRA CD,CPLANE2,CPA5,CPA1,COQ4,COQ10B,COPS8,COLQ,COL4A1,COL20A1,COA8,CNTN6,CNTN1,CNTLN,CNOT8,CNOT6,CNKS3R 3,CNGB3,CNGB3,CLU,CLN6,CLDN24,CLDN22,CIDE,CCHRM2,CHMP7,CHIT1,CHD2,CHAF1B,CFFR5,CFFR3,CFFR2,CFFR1,CFH,CE P83,CEP41,CEP350,CEP164,CELF3,CELA2B,CDX1,CDT1,CDKN2C,CDK4,CDH13,CDC73,CDCT7,CD6,CDH3,CD300LF,CD19,CCNL1,CCNB1IP1,CCL27,CCL19,CCDC83,CCDC68,CCDC197,CCDC179,CCDC170,CCDC153,CCDC150,CCDC146,CBL,CATSPER B,CAST,CAPNS2,CABCOCO1,C9orf43,C5orf63,C4BPB,C4BPA,C3orf52,C3,C2orf66,C2CD5,C2CD4D,C2CD2L,C22orf42,C21orf140,C 1QTNF5,C1orf185,C1orf134,C14orf39,C14orf180,C10orf99,BVES,BTD,BMP1,BMF,BEND4,BBOX1,BAZ1A,BAG5,BAAT,B4GAL T6,B4G ALNT1,B3GALT4,B3GALT2,AXIN1,AVIL,ATRIN,ATP6V1C2,ATP5MJ,ATP23,ATG101,ASPM,ASPG,ASPA,ARPIN-AP3S2,ARL14EP,ARID3C,ARHGFE19,ARHGAP22,ARFGAP1,ARF6,AQP3,APRT,APOLD1,APCD1L,AP3S2,AP3D1,APOEP,ANXA4, ANKRD9,ANKRD66,ANKRD34C,ANKRD28,ANG,AMIGO2,AMH,ALX4,ALDH1A3,AL589765.1,AKR1C8,AHSA1,AHNAK,AGMO,AGAP2, ADGRA1,ADD2,ADAMTSL1,ACOX3,ACBD6,ABI3BP,ABCG4,ABCA1,AAGAB,AADACL4
1	ZW10,ZSCAN9,ZSCAN31,ZSCAN26,ZSCAN23,ZSCAN16,ZSCAN12,ZP4,ZNF83,ZNF814,ZNF670-ZNF695,ZNF670,ZNF669,ZNF649,ZNF613,ZNF609,ZNF582,ZNF579,ZNF577,ZNF536,ZNF534,ZNF524,ZNF521,ZNF518B,ZNF518A,Z NF507,ZNF485,ZNF418,ZNF417,ZNF391,ZNF350,ZNF32,ZNF318,ZNF30,ZNF256,ZNF239,ZNF236,ZNF219,ZNF184,ZNF165,ZNF133, ZNF124,ZKSCAN8P1,ZKSCAN8,ZKSCAN4,ZKSCAN3,ZKSCAN2,ZHX3,ZDHHC23,ZDHHC14,ZCCHC2,ZC3H12C,ZC3H10,ZBTB9,ZBT B39,ZBTB20,ZBTB12,ZBED9,XRN1,XPO6,XPO4,XPNPEP1,XNDC1N,XKR3,WTIP,WRAP53,WNT5B,WLS,WFIKKN1,WDR90,WDR88, WDR37,WDR25,WDR24,WDR1,WASF3,WAPL,VWCE,VWA7,VSTM2A,VPS37A,VPS18,VPS13D,VPS11,VIM,VHL,VAPB,VAMP8,VAM P5,USP8,USP50,USP28,USP10,USF3,URI1,UPP2,UPF2,UNC5C,UNC119B,UHRF1,UGT1A9,UGT1A8,UGT1A7,UGT1A6,UGT1A5,UG T1A4,UGT1A3,UGT1A10,UGT1A1,UBR2,UBE2V1,UBE2R2,UBAP2,UBAP1,UBAC2,UBAC1,UBA2,UAP1L1,TYMP,TXNDC9,TWF1,TUF M,TUBB,TLL8,TTK,TTC6,TTCS,TTCT2,TSPLY4,TSPLYL1,TSPAN9,TSPAN15,TSNAX-DISC1,TSGA10,TSC1,TRPM8,TRPM7,TRPM5,TRPC1,TRNT1,TRIP4,TRIAP1,TRHDE,TREML4,TRDV3,TRDJ4,TRDJ3,TRDJ2,TRDJ1, TRDD3,TRDC,TRBV20OR9-2,TRAV9-2,TRAV8-6,TRAV2,TRAV14DV4,TRAV12-3,TRAPPC3L,TPTE2,TPM1,TPBG,TP53,TOX2,TOPI1,TOMT,TOMM20L,TOGARAM2,TOGARAM1,TNS1,TNRC6C,TNIP3,TNIK,TNFSF1 4,TNFRSF14,TNFAIP8L3,TNFAIP2,TMPRSS7,TMPRSS3,TMPRSS2,TMPRSS15,TMIE,TMEM53,TMEM40,TMEM255B,TMEM253,TME M250,TMEM242,TMEM239,TMEM233,TMEM216,TMEM178B,TMEM178A,TMEM138,TMEM132B,TMEM126B,TMEM126A,TMEM125,T MEM108,TKFC,TIMP3,TIMM9,THUMP2D,THRAP3,THBS2,THBS1,TGM6,TGFB1R1,TEX49,TEX264,TESMIN,TERT,TDRD6,TCP11,TC N1,TCF21,TBP,TBC1D22A,TASOR,TANGO2,TAF45,TAC3,SYTL2,SYT9,SYNJ2BP-COX16,SYNJ2BP,SYNGAP1,SYCE3,SVIL,SUSD6,SUSD5,SURF4,SURF4,SURF1,SULT6B1,STYK1,STX3,STX2,STUB1,STRA 6,STPG2,STON2,STOML3,STMP1,STKLD1,STK40,STK32B,STK17A,STIM1,ST8SIA6,ST3GAL4,SRSF9,SRSF5,SRP54,SRGAP1,SRF, SRD5A3,SPTB,SPRY1,SPPL3,SPPL2A,SPOPL,SPNS2,SPINT1,SPG21,SPEF2,SPEF1,SPECC1L-ADORA2A,SPECC1L,SPDYE1,SPATA46,SPATA22,SPATA17,SPATA12,SPACA9,SOX18,SORCS2,SORBS1,SOGA3,SOC5,SOC53,

Frequency	Genes
	SNX22,SNX13,SNX10,SNX1,SNW1,SNAP25,SMURF1,SMNDC1,SMIM32,SMG7,SMCO2,SMARCD1,SMAD5,SLIT3,SLIRP,SLC6A19,SLC6A18,SLC5A9,SLC5A1,SLC49A4,SLC44A4,SLC43A2,SLC41A1,SLC39A8,SLC39A7,SLC38A9,SLC37A3,SLC37A1,SLC35E3,SLC35D3,SLC35C1,SLC26A9,SLC25A35,SLC25A27,SLC25A12,SLC22A7,SLC1A4,SLC19A3,SLC16A7,SLC13A5,SLC13A4,SLC12A6,SLC10A1,SLAMF8,SLAIN1,SKIDA1,SIM2,SIGLEC11,SHOX2,SHC3,SH3D21,SH2B1,SGPP2,SGCA,SFXN5,SFXN1,SFMBT2,SETD7,SESN3,SERPINB12,SERINC1,SEMA7A,SEMA4F,SECISBP2,SEC61G,SEC24A,SEC16A,SDHAF3,SDHAF2,SCTR,SCP2D1,SCO2,SCN3B,S,CN10A,SCLT1,SCARF1,SCAMP5,SBF1,SAV1,SARNP,SAPCD2,SAMM50,SAMD14,SALL3,SALL2,S1PR4,RYR2,RXYLT1,RXRG,RXR,B,RUFY4,RUBCNL,RTL10,RTL1,RSRC1,RSPO3,RRP1B,RPS8,RPS27L,RPS10-NUDT3,RPS10,RPL7A,RPL41,RPL38,RPL31,RPL27A,RPL26L1,RPL10L,RPGRIP1L,ROPN1,ROBO4,ROBO3,RO60,RNH1,RNGTT,RNF220,RNF214,RNF181,RNF121,RNF113B,RNF10,RNASE9,RITA1,RIPOR3,RIOK2,RING1,RIMKL8,RILP,RIDA,RHOV,RHOT2,RHOG,RHBDL1,RGS21,RGS18,RGS17,RGMA,RFLNA,REXO4,REV1,RELB,RDX,RDH5,RDH10,RBPMS2,RBP4,RBM26,RBM15B,RASSF7,RASSF5,RASAL1,RARRES1,RAP1B,RANGRF,RAF1,RAD54L2,RACGAP1,RABL6,RAB8B,RAB40C,RAB3GAP2,RAB36,RAB31,RAB29,RAB27B,RAB20,RAB15,RAB11FIP3,R3HDM1,QSOX1,QSER1,QPCT,PXT1,PXDN,PWWP2B,PUS7L,PUM1,PTTG1IP,PTPRR,PTPRM,PTPRH,PTPN6,PTPN1,PTH1R,PTGIS,PTDSS2,PTCHD4,PTBP1,PSMD1,PSMB1,PSMA6,PSAT1,PRXL2B,PRSS50,PRSS3,PRSS23,P,RSS16,PRSS12,PRRT4,PRRT3,PRRG4,PRR5-ARHGAP8,PRR35,PRR3,PRPF8,PRPF39,PRORP,PROC,PRKD3,PRDM6,PRB3,PPP6R3,PPP6R2,PPP5C,PPP4R3B,PPP4R1,PPP2R3C,PPP1R9B,PPP1R32,PPP1R18,PPP1R10,PPM1L,PPIL1,PIIB,PPCDC,POU6F2,POU4F2,POP5,POM12L2,POLRMT,POLR3G,POLR3D,POLG,PNPT1,PNPLA3,PNP,PNKD,PNISR,PMEL,PM20D1,PLXNA4,PLS1,PLPPR5,PLPPR3,PLOD1,PLEKHD1,PLEK,PLCL1,PLCG1,PLCB1,PLB1,PLA2G7,PKHD1,PIWIL2,PIK3AP1,PIGO,PIF1,PICALM,PIBF1,PHRF1,PHPT1,PHOX2A,PHIP,PHF10,PHF1,PGM1,PGBD5,PGBD1,PGAP2,PGA5,PGA4,PGA3,PEX19,PEPD,PEDS1-UBE2V1,PEDS1,PEA15,PDXP,PDXK,PDIA6,PDE6A,PDE1B,PDCC2L,PDCC2,PCOLCE2,PCLAF,PCDH1,PCARE,PAX3,PATL1,PARVB,PARP2,PAQR9,PALD1,PACS2,PA2G4,P3H2,P2RY14,OTULINL,OTUD4,OSBPL8,OSBPL3,OSBPL11,OSBP2,OSBP,ORMDL2,OR6Q1,OR6C74,OR5C1,OR5AU1,OR52M1,OR4D9,OR4D11,OR3A3,OR3A1,OR2B8P,OR2B6,OR2B2,OR1R1P,OR1L6,OR1L4,OR1L3,OR1E2,OR1E1,OR10V1,ODF3B,ODF2,OA2Z,NUP35,NUP205,NUP107,NUMA1,NUDT3,NUDT2,NUCKS1,NUBPL,NUAK2,NT5DC1,NRRO,S,NRP2,NRM,NRG1,NR2C1,NPY5R,NPEPPS,NPDC1,NOM1,NKAPL,NHLRC4,NHLH1,NFKB1,NETO1,NEMP1,NEMP4,NDUFAF7,NDUFA10,NDST3,NDRG2,NCL,NCAPH2,NCAM1,NBPF3,NAB2,NAA35,MYT1L,MYSM1,MYORG,MYO5B,MYO1A,MYO10,MYL6B,MYL6,MYL10,MVD,MTUS1,MTRR,MTNR1B,MTMR7,MTIF3,MTDH,MTCL1,MSX1,MSI1,MRS2,MRPS18B,MRPL50,MRPL46,MRPL30,MRPL21,MRPL16,MRGPRE,MRAP,MPPED2,MPHOSPH8,MNX1,MMP21,MMP19,MMEL1,MLLT10,MLF1,MLEC,MKRN2,MKKS,MKI67,MITD1,MIOX,MGAT4C,MGAT4A,MFAP5,METTL7B,METTL6,METTL2B,METTL26,METRN,MERTK,MEP1B,MEP1A,MED28,MED22,ME D10,MDM2,MDFIC2,MDFI,MDC1,MCRIP2,MCM9,MCAM,MBP,MBOAT1,MAS1L,MARS2,MAPKAPK3,MAPKAPK2,MAPK8IP2,MAPK8IP1,MAPP271,MAP4K5,MAP3K19,MANF,MANBA,MAMDC4,MAGOH8,MAF,MAD2L1,M1AP,LZTS3,LYZL2,LYSMD3,LYG2,LYG1,LXN,LSAMP,LRTOMT,LRRMT1,LRRCS8,LRRCS6,LRRCS5,LRRCS10,LRFN2,LPIN3,LPGAT1,LPAR3,LONRF2,LNPK,LMNTD2,LMF2,LMAN1L,LIPT1,LHX9,LGI2,LGALS3,LGALS1,LNCN9,LARS2,LAP3,LAMTOR1,LACTB,LACRT,KRTAP10-5,KRTAP10-4,KRTAP10-12,KRT19,KRT15,KRBA1,KNCN,KLK14,KLHL3,KLHL28,KLHDC8A,KLHDC7B,KLF5,KLF14,KIRREL1,KIFC1,KIF6,KIF4B,KIF2C,KIF24,KIAA2013,KIAA1109,KIAA0408,KDM6B,KCTD15,KCNMB4,KCNK2,KCNJ6,KCNH5,KCNF1,KCNE4,KCNE1,KCNAB1,KATNBL1,KANSL2,KAAG1,JUN,JMJD8,JADE1,ITIH5,ITGB1,ITGA7,ITGA1,ISLR,IRAK4,IRAK1BP1,IQCF6,IQCA1,INPP5A,INO80,INHBB,ILKAP,IL31RA,I L21,IL2,IL18BP,IL17RD,IL17D,IKZF3,IKZF2,IGHMBP2,IGFL1,IGFALS,IFT74,IFT43,IFI27L2,IFI27L1,IER3,IARS2,HYUO1,HTR6,HTR3B,HTR3A,HTR2B,HSPBAP1,HSPA1L,HSPA1B,HSPA1A,HSF2BP,HSF2,HSD17B8,HS3ST3B1,HRAS,HGPCAL1,HOXD9,HOXD13,HOXD12,HOXD11,HOXD10,HOXA1,HMGCLL1,HMGB2,HMCN1,HMBS,HLA-E,HLA-DRB5,HK2,HIBCB,HGD,HESX1,HEMK1,HECA,HDGFL1,HDDC2,HDAC11,HAT1,HACL1,H4C9,H4C13,H4C12,H4C11,H3C12,H3C11,H2BC17,H2BC15,H2BC14,H2BC12L,H2BC12,H2BC11,H2AX,H2AC17,H2AC16,H2AC15,H2AC12,H2AC11,H1-5,GYS1,GYPC,GTSF1,GTFF3A,GSGL2,GSC,GRXCR2,GRP,GRM2,GRK1,GRHL1,GPX6,GPX5,GPX3,GPRC5C,GPR65,GPR33,GPR32,GPR183,GPR18,GPR171,GPR158,GPR156,GPR137B,GPR12,GPI,GPD1,GPATCH2L,GPATCH2,GPATCH1,GPAT4,GNL1,GN2,GNAO1,GMD5,GLT6D1,GLRX2,GLRA3,GLDC,GK5,GK2,GJB6,GJA10,GGCX,GFRA4,GFRA2,GFM1,GDF11,GCH1,GBGT1,GATC,GAS1,GARRE1,GARIN1B,GARIN1A,GALNT7,GALC,GAL,GABRA1,GABARAPL1,FZD4,FUT7,FURIN,FSTL3,FSTL1,FSIP2,FSIP1,FSCB,FRZB,FRY,FRS2,FRMD5,FREM2,FOX1,FOX1A,FMO4,FMN2,FLYWCH2,FLYWCH1,FLOT1,FLJ14816,FKBP3,FIZ1,FIGN,FIG4,FGR,FGL1,FGFR2,FGF12,FGD6,FFAR4,FES,FBXO48,FBXO34,FBXO21,FBXL17,FBXL16,FBN3,FBLIM1,FAXC,FASTKD5,FASTKD3,FAS,FAR P1,FANCI,FANCD2OS,FANCD2,FAM24B,FAM214A,FAM205A,FAM180A,FAM177A1,FAM167A,FAM151A,FAM131B,F5,EXOC3L2,EXOC2,EXV2,EVC,EVA1B,ESYT1,ESRRB,ESD,ERMARD,ERGC1,ERFE,ERBB4,ERBB3,ENTPD2,ENSG00000288712,ENSG00000288695,ENSG00000288645,ENSG00000288623,ENSG00000286131,ENSG00000286088,ENSG00000285982,ENSG00000285471,ENSG00000285304,ENSG00000285245,ENSG00000284776,ENSG00000283321,ENSG00000279073,ENSG00000277611,ENSG00000276302,ENSG00000273155,ENSG00000269825,ENSG00000269476,ENSG00000266979,ENSG00000266953,ENSG00000260234,ENSG00000260170,ENSG00000259316,ENSG00000258790,ENSG00000258465,ENSG00000258311,ENSG00000257411,ENSG00000257390,ENSG00000256591,ENSG00000255330,ENSG00000249240,ENSG00000241962,ENSG00000241690,ENSG00000206549,ENSG00000170846,ENSG00000111780,EN1,EMILIN3,EMC4,EMC3,ELOVL7,EIF5B,EIF2S3B,EIF2AK2,EHMT2,EGFR,EFNB3,EEF1AKMT1,EDRF1,EDF1,EDARADD,EBNA1BP2,EAF1,DYRK3,DYNLT2,DYNLL1,DYNC1I2,DTD1,DST,DSPP,DRD1,DYPSL4,DPY19L2,DOP1B,DOK1,DOCK3,DNAJC14,DNAJC10,DNAJB6,DMXL1,DMRTA1,DLX6,DLX5,DLL4,DLK2,DLGAP5,DLG1,DIS3,DHX34,DHX16,DHX15,DHRS7C,DHCR24,DEPDC7,DEPDC1B,DENND6B,DENND1B,DEFB131B,DDX54,DDX46,DDX4,DDX24,DDB1,DCSTAMP,P,DCLK3,DCLK1,DCDC2,DCD,DCBLD2,DCAF8,DCAF12,DCAF1,DBI,DAPP1,DACT1,DACH1,CYP8B1,CYP39A1,CYP2C9,CYP24A1,CYB561A3,CUZD1,CUTA,CTU2,CTNNB1,CTNNB1,CSRNP3,CSNK1G1,CRY2,CRTAP,CRMP1,CRIP3,CREG2,CREG1,CREBZF,CRB1,CPT1B,CPT1A,CPSF7,CPSF6,CPQ,CPNE5,CPM,CPLX3,CPAMD8,COX6A1,COX16,COQ5,COQ3,COPA,COL9A1,COL5A1,COL4A2,COL1A1,COL11A2,COG6,COCH,COA1,CNTNAP5,CNTFR,CNR1,CNOT2,CNDP2,CNBD1,CLRN1,CLPTM1,CLPTM1,CLPSL2,CLPB,CLNK,CLEC1B,CLEC19A,CLEC12B,CLEC12A,CLDN6,CLDN17,CLCA4,CLASP2,CKS2,CKAP4,CIZ1,CISH,CIAO2A,CHST2,CHODL,CHL1,CHKB-CPT1B,CHKB,CHD6,CHD3,CHAC1,CH25H,CFHR4,CFAP77,CFAP73,CFAP61,CFAP57,CFAP300,CFAP299,CERCAM,CEP55,CENPB,CEBPZOS,CEBPZ,CEBPB,CYDL,CDKN1C,CDKN1B,CDH24,CDH18,CDCA2,CDC42BPB,CDC25B,CD80,CD63,CD200,CD163,CCT2,CCT2,CNT1,CCNJ,CCND1,CCN5,CCN1,CCL21,CCDC9B,CCDC91,CCDC89,CCDC85A,CCDC8,CCDC74A,CCDC73,CCDC7,CCDC66,CCDC60,CCDC33,CCDC198,CCDC190,CCDC183,CCDC148,CCDC14,CCBE1,CC2D2B,CBR4,CBLIF,CAVIN3,CAV3,CATSPERE,CATIP,CASP4,CAND2,CAMSAP1,CAMLG,CAMK2A,CALHM6,CALHM5,CACNA1S,CACNA1I,CABS1,CABP1,CA6,CA12,C9orf24,C6orf89,C6orf62,C6orf136,C6orf120,C5orf24,C5AR2,C5AR1,C4orf51,C4orf33,C3orf18,C2orf88,C2orf76,C2orf72,C2orf15,C1R,C1orf53,C1orf226,C1orf210,C1GALT1,C15orf61,C14orf28,C11orf53,BSND,BRK1,BPNT1,BPIFC,BORCS6,BOLL,BMPR1B,BMP5,BLVRA,BLOC1S1,BLNK,BLID,BEST4,BEST3,BEND5,BCR,BCHE,BBS12,BBS10,BBLN,BACE2,AVPR1B,ATXN2L,ATR,ATP6V0A4,ATP5F1C,ATP4B,ATP4A,ATP2A1,ATL2,ATL1,ATG14,ATAT1,ASIC1,ASF1A,ASB7,ASB4,ARVCF,ARSA,ARPP21,ARPIN,ARMH1,ARMC3,ARL15,ARID1B,ARHGFEF40,ARHGAP8,ARHGAP5,ARHGAP21,ARHGAP20,ARCN1,AQP7,AQP6,AQP5,AQP2,APOD,APOC4-APOC2,APOC4,APOC2,APLN,APH1B,AP00695.1,AOAH,ANXA10,ANKS1A,ANKRD10,ANKDD1A,ANK3,ANAPC15,ANAPC10,AMOTL1,ALS2CL,ALOX15,ALDOB,AKT1,AKR1C4,AKR1C1,AK8,AK7,AJM1,AGA,AFF3,AF131215.5,ADORA2A,ADM2,ADH7,ADGRF5,ADD3,ADAMTS2,ADAM3,ADAM21,ADAD1,ACTR3B,ACTN1,ACTB,ACSL3,ACOT6,ACOT13,ACMSD,ACKR2,ACAN,ACADS,ACAA2,ABRAXAS2,ABRACL,ABHD2,ABHD17C,ABCF1,ABCC1,ABCC4,ABCA2,A2ML1

Table S11 Functional enrichment of genes in CDS which CPC-AIS located and these genes were found in at least 5 samples

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	17/1992	20/18800	2.10E-14	1.23E-10	1.17E-10	17
BP	GO:0033139	regulation of peptidyl-serine phosphorylation of STAT protein	17/1992	23/18800	1.36E-12	3.99E-09	3.77E-09	17
BP	GO:0042501	serine phosphorylation of STAT protein	17/1992	27/18800	7.47E-11	1.47E-07	1.38E-07	17
BP	GO:0043330	response to exogenous dsRNA	23/1992	49/18800	1.26E-10	1.86E-07	1.75E-07	23
BP	GO:0033138	positive regulation of peptidyl-serine phosphorylation	35/1992	107/18800	5.17E-10	5.13E-07	4.84E-07	35
BP	GO:0043331	response to dsRNA	24/1992	56/18800	5.22E-10	5.13E-07	4.84E-07	24
BP	GO:0033135	regulation of peptidyl-serine phosphorylation	41/1992	144/18800	2.05E-09	1.72E-06	1.63E-06	41
BP	GO:0009913	epidermal cell differentiation	55/1992	230/18800	4.87E-09	3.59E-06	3.39E-06	55
BP	GO:0018209	peptidyl-serine modification	71/1992	338/18800	1.18E-08	7.73E-06	7.30E-06	71
BP	GO:0031424	keratinization	28/1992	85/18800	2.29E-08	1.23E-05	1.17E-05	28
BP	GO:0030216	keratinocyte differentiation	43/1992	167/18800	2.31E-08	1.23E-05	1.17E-05	43
BP	GO:0002323	natural killer cell activation involved in immune response	16/1992	32/18800	2.72E-08	1.34E-05	1.26E-05	16
BP	GO:0018105	peptidyl-serine phosphorylation	65/1992	315/18800	9.78E-08	4.43E-05	4.18E-05	65
BP	GO:0008544	epidermis development	70/1992	355/18800	2.04E-07	8.58E-05	8.10E-05	70
BP	GO:0030198	extracellular matrix organization	61/1992	307/18800	9.30E-07	0.000361	0.000341	61
BP	GO:0045109	intermediate filament organization	22/1992	68/18800	1.02E-06	0.000361	0.000341	22
BP	GO:0043062	extracellular structure organization	61/1992	308/18800	1.04E-06	0.000361	0.000341	61
BP	GO:0045229	external encapsulating structure organization	61/1992	310/18800	1.30E-06	0.000427	0.000403	61
BP	GO:0030574	collagen catabolic process	16/1992	42/18800	2.68E-06	0.000832	0.000785	16
BP	GO:0045104	intermediate filament cytoskeleton organization	25/1992	88/18800	2.88E-06	0.000847	0.0008	25
BP	GO:0006816	calcium ion transport	76/1992	424/18800	3.02E-06	0.000848	0.000801	76
BP	GO:0045103	intermediate filament-based process	25/1992	89/18800	3.60E-06	0.000964	0.00091	25
BP	GO:0046425	regulation of receptor signaling pathway via JAK-STAT	27/1992	105/18800	9.27E-06	0.002373	0.002241	27
BP	GO:0042100	B cell proliferation	26/1992	100/18800	1.09E-05	0.00267	0.002521	26
BP	GO:0043588	skin development	56/1992	296/18800	1.19E-05	0.002791	0.002635	56
BP	GO:1902305	regulation of sodium ion transmembrane transport	20/1992	67/18800	1.23E-05	0.002791	0.002635	20
BP	GO:2000649	regulation of sodium ion transmembrane transporter activity	18/1992	58/18800	1.84E-05	0.004015	0.003791	18
BP	GO:1904892	regulation of receptor signaling pathway via STAT	28/1992	116/18800	2.29E-05	0.004818	0.004549	28
BP	GO:0051607	defense response to virus	54/1992	290/18800	2.68E-05	0.005434	0.00513	54
BP	GO:0140546	defense response to symbiont	54/1992	291/18800	2.95E-05	0.005798	0.005474	54
BP	GO:0019722	calcium-mediated signaling	41/1992	202/18800	3.22E-05	0.006119	0.005777	41
BP	GO:0070588	calcium ion transmembrane transport	57/1992	314/18800	3.46E-05	0.006367	0.006012	57
BP	GO:0019932	second-messenger-mediated signaling	55/1992	303/18800	4.72E-05	0.008416	0.007946	55
BP	GO:0010959	regulation of metal ion transport	68/1992	403/18800	7.15E-05	0.012394	0.011702	68
BP	GO:0009615	response to virus	66/1992	392/18800	9.70E-05	0.016317	0.015406	66
BP	GO:0007626	locomotory behavior	38/1992	192/18800	0.000108	0.017159	0.016201	38
BP	GO:0048839	inner ear development	38/1992	192/18800	0.000108	0.017159	0.016201	38
BP	GO:0032963	collagen metabolic process	24/1992	101/18800	0.000111	0.017277	0.016312	24
BP	GO:0006814	sodium ion transport	46/1992	249/18800	0.000124	0.018709	0.017664	46
BP	GO:0002028	regulation of sodium ion transport	22/1992	90/18800	0.000136	0.020078	0.018957	22
BP	GO:0071287	cellular response to manganese ion	7/1992	13/18800	0.000143	0.020514	0.019368	7
BP	GO:0002286	T cell activation involved in immune response	26/1992	116/18800	0.000165	0.023108	0.021817	26
BP	GO:0034765	regulation of ion transmembrane transport	76/1992	476/18800	0.000175	0.023981	0.022641	76

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	17/1992	20/18800	2.10E-14	1.23E-10	1.17E-10	17
BP	GO:0033139	regulation of peptidyl-serine phosphorylation of STAT protein	17/1992	23/18800	1.36E-12	3.99E-09	3.77E-09	17
BP	GO:0042501	serine phosphorylation of STAT protein	17/1992	27/18800	7.47E-11	1.47E-07	1.38E-07	17
BP	GO:0043330	response to exogenous dsRNA	23/1992	49/18800	1.26E-10	1.86E-07	1.75E-07	23
BP	GO:0035725	sodium ion transmembrane transport	35/1992	177/18800	0.000201	0.026931	0.025427	35
BP	GO:0006874	cellular calcium ion homeostasis	73/1992	456/18800	0.000215	0.028093	0.026524	73
BP	GO:0032412	regulation of ion transmembrane transporter activity	47/1992	263/18800	0.000236	0.030178	0.028493	47
BP	GO:0032943	mononuclear cell proliferation	52/1992	300/18800	0.000248	0.030687	0.028973	52
BP	GO:1904062	regulation of cation transmembrane transport	59/1992	352/18800	0.00025	0.030687	0.028973	59
BP	GO:0007259	receptor signaling pathway via JAK-STAT	34/1992	173/18800	0.000278	0.033368	0.031505	34
BP	GO:0022898	regulation of transmembrane transporter activity	48/1992	273/18800	0.000301	0.035417	0.033439	48
BP	GO:0046651	lymphocyte proliferation	51/1992	296/18800	0.000328	0.037854	0.03574	51
BP	GO:0072503	cellular divalent inorganic cation homeostasis	77/1992	494/18800	0.00034	0.038506	0.036355	77
BP	GO:0042391	regulation of membrane potential	68/1992	425/18800	0.000353	0.03919	0.037001	68
BP	GO:0043583	ear development	40/1992	219/18800	0.000421	0.044537	0.04205	40
BP	GO:0097696	receptor signaling pathway via STAT	35/1992	184/18800	0.000433	0.044537	0.04205	35
BP	GO:0001655	urogenital system development	58/1992	352/18800	0.000439	0.044537	0.04205	58
BP	GO:1904645	response to amyloid-beta	15/1992	55/18800	0.000441	0.044537	0.04205	15
BP	GO:2000650	negative regulation of sodium ion transmembrane transporter activity	7/1992	15/18800	0.000442	0.044537	0.04205	7
BP	GO:0048525	negative regulation of viral process	21/1992	91/18800	0.000446	0.044537	0.04205	21
BP	GO:0055074	calcium ion homeostasis	73/1992	468/18800	0.000468	0.045981	0.043413	73
CC	GO:0043025	neuronal cell body	82/2084	482/19594	1.15E-05	0.005963	0.005485	82
CC	GO:0045095	keratin filament	26/2084	102/19594	1.71E-05	0.005963	0.005485	26
MF	GO:0005132	type I interferon receptor binding	16/2031	17/18410	6.96E-15	7.62E-12	7.18E-12	16
MF	GO:0030280	structural constituent of skin epidermis	23/2031	37/18410	1.11E-13	6.10E-11	5.74E-11	23
MF	GO:0001637	G protein-coupled chemoattractant receptor activity	11/2031	26/18410	4.57E-05	0.012493	0.011768	11
MF	GO:0004950	chemokine receptor activity	11/2031	26/18410	4.57E-05	0.012493	0.011768	11
MF	GO:0016493	C-C chemokine receptor activity	10/2031	23/18410	7.69E-05	0.016829	0.015853	10
MF	GO:0004896	cytokine receptor activity	24/2031	97/18410	0.000106	0.018575	0.017497	24
MF	GO:0019957	C-C chemokine binding	10/2031	24/18410	0.000119	0.018575	0.017497	10
MF	GO:0008237	metallopeptidase activity	38/2031	189/18410	0.000176	0.024113	0.022714	38
MF	GO:0046873	metal ion transmembrane transporter activity	71/2031	428/18410	0.00029	0.035286	0.033238	71
MF	GO:0004222	metalloendopeptidase activity	25/2031	111/18410	0.000374	0.040932	0.038557	25
KEGG	hsa04622	RIG-I-like receptor signaling pathway	23/897	70/8164	7.47E-07	0.000244	0.000221	23
KEGG	hsa05167	Kaposi sarcoma-associated herpesvirus infection	42/897	194/8164	1.04E-05	0.001229	0.001112	42
KEGG	hsa04974	Protein digestion and absorption	27/897	103/8164	1.13E-05	0.001229	0.001112	27
KEGG	hsa04630	JAK-STAT signaling pathway	35/897	162/8164	5.85E-05	0.004782	0.004325	35
KEGG	hsa05162	Measles	31/897	139/8164	8.19E-05	0.005353	0.004842	31
KEGG	hsa05417	Lipid and atherosclerosis	41/897	215/8164	0.000269	0.013155	0.0119	41
KEGG	hsa04620	Toll-like receptor signaling pathway	24/897	104/8164	0.000295	0.013155	0.0119	24
KEGG	hsa04623	Cytosolic DNA-sensing pathway	17/897	63/8164	0.000322	0.013155	0.0119	17
KEGG	hsa05163	Human cytomegalovirus infection	42/897	225/8164	0.000366	0.013287	0.012019	42
KEGG	hsa04151	PI3K-Akt signaling pathway	59/897	354/8164	0.000624	0.019527	0.017664	59
KEGG	hsa05161	Hepatitis B	32/897	162/8164	0.000657	0.019527	0.017664	32
KEGG	hsa05152	Tuberculosis	34/897	180/8164	0.001039	0.028317	0.025615	34
KEGG	hsa04724	Glutamatergic synapse	24/897	114/8164	0.001205	0.030302	0.02741	24

Table S12 Functional enrichment of genes in CDS which CPC-AIS located and these genes were found in at least 1 sample.

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0031424	keratinization	42/3538	85/18800	1.67E-10	8.98E-07	8.46E-07	42
BP	GO:0033141	positive regulation of peptidyl-serine phosphorylation of STAT protein	17/3538	20/18800	2.87E-10	8.98E-07	8.46E-07	17
BP	GO:0030216	keratinocyte differentiation	65/3538	167/18800	9.89E-10	1.68E-06	1.58E-06	65
BP	GO:0033139	regulation of peptidyl-serine phosphorylation of STAT protein	18/3538	23/18800	1.07E-09	1.68E-06	1.58E-06	18
BP	GO:0009913	epidermal cell differentiation	80/3538	230/18800	6.31E-09	7.68E-06	7.23E-06	80
BP	GO:0042501	serine phosphorylation of STAT protein	19/3538	27/18800	7.36E-09	7.68E-06	7.23E-06	19
BP	GO:0046942	carboxylic acid transport	90/3538	279/18800	4.68E-08	4.19E-05	3.94E-05	90
BP	GO:0043330	response to exogenous dsRNA	26/3538	49/18800	7.87E-08	6.16E-05	5.80E-05	26
BP	GO:0015849	organic acid transport	97/3538	318/18800	2.91E-07	0.000193	0.000182	97
BP	GO:0006814	sodium ion transport	80/3538	249/18800	3.08E-07	0.000193	0.000182	80
BP	GO:0015711	organic anion transport	107/3538	361/18800	3.55E-07	0.000202	0.00019	107
BP	GO:0043588	skin development	91/3538	296/18800	4.57E-07	0.000237	0.000223	91
BP	GO:0018209	peptidyl-serine modification	101/3538	338/18800	4.92E-07	0.000237	0.000223	101
BP	GO:0043331	response to dsRNA	27/3538	56/18800	5.67E-07	0.000242	0.000228	27
BP	GO:0006816	calcium ion transport	121/3538	424/18800	5.80E-07	0.000242	0.000228	121
BP	GO:0018105	peptidyl-serine phosphorylation	94/3538	315/18800	1.29E-06	0.000505	0.000475	94
BP	GO:0008544	epidermis development	103/3538	355/18800	1.73E-06	0.000614	0.000578	103
BP	GO:0033135	regulation of peptidyl-serine phosphorylation	51/3538	144/18800	1.81E-06	0.000614	0.000578	51
BP	GO:0033138	positive regulation of peptidyl-serine phosphorylation	41/3538	107/18800	1.86E-06	0.000614	0.000578	41
BP	GO:0070588	calcium ion transmembrane transport	93/3538	314/18800	2.09E-06	0.000654	0.000616	93
BP	GO:0030198	extracellular matrix organization	91/3538	307/18800	2.58E-06	0.000769	0.000724	91
BP	GO:0043062	extracellular structure organization	91/3538	308/18800	2.99E-06	0.000852	0.000802	91
BP	GO:0045229	external encapsulating structure organization	91/3538	310/18800	4.02E-06	0.001094	0.00103	91
BP	GO:0006688	glycosphingolipid biosynthetic process	18/3538	33/18800	4.74E-06	0.001237	0.001164	18
BP	GO:0006874	cellular calcium ion homeostasis	124/3538	456/18800	6.12E-06	0.001534	0.001443	124
BP	GO:0006687	glycosphingolipid metabolic process	26/3538	60/18800	1.08E-05	0.002611	0.002458	26
BP	GO:0002323	natural killer cell activation involved in immune response	17/3538	32/18800	1.39E-05	0.003199	0.003011	17
BP	GO:0055074	calcium ion homeostasis	125/3538	468/18800	1.43E-05	0.003199	0.003011	125
BP	GO:0019722	calcium-mediated signaling	63/3538	202/18800	1.54E-05	0.003315	0.00312	63
BP	GO:0072503	cellular divalent inorganic cation homeostasis	130/3538	494/18800	2.11E-05	0.004395	0.004137	130
BP	GO:0035725	sodium ion transmembrane transport	55/3538	177/18800	5.78E-05	0.011431	0.010759	55
BP	GO:2000649	regulation of sodium ion transmembrane transporter activity	24/3538	58/18800	5.84E-05	0.011431	0.010759	24
BP	GO:0010959	regulation of metal ion transport	107/3538	403/18800	7.44E-05	0.014124	0.013293	107
BP	GO:0042100	B cell proliferation	35/3538	100/18800	9.38E-05	0.017284	0.016268	35
BP	GO:0035330	regulation of hippo signaling	12/3538	21/18800	0.000104	0.01837	0.01729	12
BP	GO:1902305	regulation of sodium ion transmembrane transport	26/3538	67/18800	0.000107	0.01837	0.01729	26
BP	GO:0070085	glycosylation	70/3538	244/18800	0.000109	0.01837	0.01729	70
BP	GO:0046425	regulation of receptor signaling pathway via JAK-STAT	36/3538	105/18800	0.000122	0.020048	0.018868	36
BP	GO:0060326	cell chemotaxis	86/3538	315/18800	0.000132	0.021175	0.01993	86
BP	GO:0097396	response to interleukin-17	8/3538	11/18800	0.000149	0.021917	0.020628	8
BP	GO:0097398	cellular response to interleukin-17	8/3538	11/18800	0.000149	0.021917	0.020628	8
BP	GO:0097400	interleukin-17-mediated signaling pathway	8/3538	11/18800	0.000149	0.021917	0.020628	8
BP	GO:0051480	regulation of cytosolic calcium ion concentration	95/3538	356/18800	0.00015	0.021917	0.020628	95

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0009100	glycoprotein metabolic process	101/3538	386/18800	0.00021	0.029822	0.028068	101
BP	GO:0048705	skeletal system morphogenesis	65/3538	228/18800	0.000227	0.031655	0.029794	65
BP	GO:0043406	positive regulation of MAP kinase activity	38/3538	116/18800	0.000235	0.031854	0.029981	38
BP	GO:0032412	regulation of ion transmembrane transporter activity	73/3538	263/18800	0.000239	0.031854	0.029981	73
BP	GO:0038093	Fc receptor signaling pathway	21/3538	52/18800	0.000251	0.032426	0.030519	21
BP	GO:0019932	second-messenger-mediated signaling	82/3538	303/18800	0.000254	0.032426	0.030519	82
BP	GO:0006865	amino acid transport	45/3538	145/18800	0.000268	0.033515	0.031543	45
BP	GO:0051924	regulation of calcium ion transport	70/3538	251/18800	0.000273	0.033515	0.031543	70
BP	GO:0015718	monocarboxylic acid transport	38/3538	117/18800	0.000285	0.034358	0.032337	38
BP	GO:0002028	regulation of sodium ion transport	31/3538	90/18800	0.000318	0.036536	0.034387	31
BP	GO:0009247	glycolipid biosynthetic process	26/3538	71/18800	0.000318	0.036536	0.034387	26
BP	GO:0001573	ganglioside metabolic process	13/3538	26/18800	0.000321	0.036536	0.034387	13
BP	GO:0009615	response to virus	101/3538	392/18800	0.00038	0.042201	0.039719	101
BP	GO:0045109	intermediate filament organization	25/3538	68/18800	0.000384	0.042201	0.039719	25
BP	GO:1901019	regulation of calcium ion transmembrane transporter activity	31/3538	91/18800	0.000397	0.042369	0.039877	31
BP	GO:0043507	positive regulation of JUN kinase activity	18/3538	43/18800	0.000405	0.042369	0.039877	18
BP	GO:0006664	glycolipid metabolic process	34/3538	103/18800	0.000416	0.042369	0.039877	34
BP	GO:0051928	positive regulation of calcium ion transport	38/3538	119/18800	0.000417	0.042369	0.039877	38
BP	GO:0071875	adrenergic receptor signaling pathway	15/3538	33/18800	0.000422	0.042369	0.039877	15
BP	GO:1905039	carboxylic acid transmembrane transport	46/3538	152/18800	0.000433	0.042369	0.039877	46
BP	GO:0051607	defense response to virus	78/3538	290/18800	0.000433	0.042369	0.039877	78
BP	GO:0022898	regulation of transmembrane transporter activity	74/3538	273/18800	0.000469	0.042811	0.040293	74
BP	GO:0019221	cytokine-mediated signaling pathway	121/3538	486/18800	0.000471	0.042811	0.040293	121
BP	GO:0098657	import into cell	66/3538	238/18800	0.000475	0.042811	0.040293	66
BP	GO:0006486	protein glycosylation	63/3538	225/18800	0.000478	0.042811	0.040293	63
BP	GO:0043413	macromolecule glycosylation	63/3538	225/18800	0.000478	0.042811	0.040293	63
BP	GO:0140546	defense response to symbiont	78/3538	291/18800	0.000484	0.042811	0.040293	78
BP	GO:0003333	amino acid transmembrane transport	33/3538	100/18800	0.000505	0.042811	0.040293	33
BP	GO:1904892	regulation of receptor signaling pathway via STAT	37/3538	116/18800	0.000505	0.042811	0.040293	37
BP	GO:1903825	organic acid transmembrane transport	46/3538	153/18800	0.000508	0.042811	0.040293	46
BP	GO:1903509	liposaccharide metabolic process	34/3538	104/18800	0.000508	0.042811	0.040293	34
BP	GO:0071880	adenylate cyclase-activating adrenergic receptor signaling pathway	13/3538	27/18800	0.000513	0.042811	0.040293	13
BP	GO:0032409	regulation of transporter activity	81/3538	305/18800	0.00052	0.042817	0.040299	81
BP	GO:0038095	Fc-epsilon receptor signaling pathway	12/3538	24/18800	0.000546	0.044393	0.041782	12
BP	GO:0031398	positive regulation of protein ubiquitination	38/3538	121/18800	0.0006	0.048131	0.0453	38
CC	GO:0045121	membrane raft	94/3730	326/19594	1.04E-05	0.003156	0.00291	94
CC	GO:0098857	membrane microdomain	94/3730	327/19594	1.19E-05	0.003156	0.00291	94
CC	GO:0045095	keratin filament	38/3730	102/19594	1.25E-05	0.003156	0.00291	38
CC	GO:0005938	cell cortex	87/3730	310/19594	6.44E-05	0.012205	0.011254	87
CC	GO:0043025	neuronal cell body	124/3730	482/19594	0.000158	0.019912	0.018361	124
CC	GO:0045177	apical part of cell	111/3730	424/19594	0.000165	0.019912	0.018361	111
CC	GO:0030286	dynein complex	22/3730	54/19594	0.000184	0.019912	0.018361	22
CC	GO:0062023	collagen-containing extracellular matrix	110/3730	429/19594	0.000411	0.038923	0.03589	110
CC	GO:0030315	T-tubule	20/3730	51/19594	0.000643	0.047592	0.043884	20
CC	GO:0005775	vacuolar lumen	51/3730	174/19594	0.000669	0.047592	0.043884	51
CC	GO:0016324	apical plasma membrane	93/3730	358/19594	0.000702	0.047592	0.043884	93
CC	GO:0043202	lysosomal lumen	32/3730	97/19594	0.000753	0.047592	0.043884	32

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
MF	GO:0030280	structural constituent of skin epidermis	28/3640	37/18410	3.36E-13	4.00E-10	3.66E-10	28
MF	GO:0005132	type I interferon receptor binding	16/3640	17/18410	7.36E-11	4.37E-08	4.01E-08	16
MF	GO:0038024	cargo receptor activity	33/3640	79/18410	6.17E-06	0.002444	0.00224	33
MF	GO:0004707	MAP kinase activity	11/3640	15/18410	1.10E-05	0.003265	0.002992	11
MF	GO:0004896	cytokine receptor activity	37/3640	97/18410	2.10E-05	0.004986	0.004569	37
MF	GO:0005044	scavenger receptor activity	21/3640	47/18410	9.18E-05	0.012946	0.011862	21
MF	GO:0016493	C-C chemokine receptor activity	13/3640	23/18410	0.000106	0.012946	0.011862	13
MF	GO:0015179	L-amino acid transmembrane transporter activity	25/3640	61/18410	0.000116	0.012946	0.011862	25
MF	GO:0001637	G protein-coupled chemoattractant receptor activity	14/3640	26/18410	0.000117	0.012946	0.011862	14
MF	GO:0004950	chemokine receptor activity	14/3640	26/18410	0.000117	0.012946	0.011862	14
MF	GO:0004712	protein serine/threonine/tyrosine kinase activity	120/3640	446/18410	0.000137	0.012946	0.011862	120
MF	GO:0008514	organic anion transmembrane transporter activity	58/3640	186/18410	0.000137	0.012946	0.011862	58
MF	GO:0030169	low-density lipoprotein particle binding	11/3640	18/18410	0.000142	0.012946	0.011862	11
MF	GO:0015293	symporter activity	47/3640	144/18410	0.000176	0.014917	0.013669	47
MF	GO:0019957	C-C chemokine binding	13/3640	24/18410	0.00019	0.015063	0.013802	13
MF	GO:0008237	metallopeptidase activity	58/3640	189/18410	0.000221	0.016412	0.015038	58
MF	GO:0046943	carboxylic acid transmembrane transporter activity	51/3640	164/18410	0.000362	0.025297	0.023179	51
MF	GO:0005342	organic acid transmembrane transporter activity	51/3640	165/18410	0.000424	0.028034	0.025688	51
MF	GO:0005496	steroid binding	34/3640	100/18410	0.000575	0.035983	0.032971	34
MF	GO:0008509	anion transmembrane transporter activity	86/3640	315/18410	0.000697	0.041413	0.037947	86
MF	GO:0106310	protein serine kinase activity	96/3640	360/18410	0.000832	0.046144	0.042281	96
MF	GO:0017080	sodium channel regulator activity	17/3640	40/18410	0.000854	0.046144	0.042281	17
KEGG	hsa04622	RIG-I-like receptor signaling pathway	32/1614	70/8164	7.30E-07	0.000242	0.000221	32
KEGG	hsa04974	Protein digestion and absorption	39/1614	103/8164	1.44E-05	0.001775	0.001616	39
KEGG	hsa05167	Kaposi sarcoma-associated herpesvirus infection	63/1614	194/8164	1.60E-05	0.001775	0.001616	63
KEGG	hsa04620	Toll-like receptor signaling pathway	37/1614	104/8164	0.000112	0.008745	0.007958	37
KEGG	hsa05162	Measles	46/1614	139/8164	0.000132	0.008745	0.007958	46
KEGG	hsa05161	Hepatitis B	51/1614	162/8164	0.00024	0.011908	0.010835	51
KEGG	hsa05417	Lipid and atherosclerosis	64/1614	215/8164	0.000251	0.011908	0.010835	64
KEGG	hsa04936	Alcoholic liver disease	45/1614	142/8164	0.000464	0.019236	0.017504	45
KEGG	hsa00604	Glycosphingolipid biosynthesis - ganglio series	9/1614	15/8164	0.000709	0.026153	0.023798	9
KEGG	hsa05171	Coronavirus disease - COVID-19	66/1614	232/8164	0.000802	0.026621	0.024224	66
KEGG	hsa04217	Necroptosis	48/1614	159/8164	0.001021	0.029579	0.026915	48
KEGG	hsa04621	NOD-like receptor signaling pathway	54/1614	184/8164	0.001069	0.029579	0.026915	54
KEGG	hsa04630	JAK-STAT signaling pathway	48/1614	162/8164	0.001585	0.036821	0.033505	48
KEGG	hsa04024	cAMP signaling pathway	62/1614	221/8164	0.001658	0.036821	0.033505	62
KEGG	hsa04657	IL-17 signaling pathway	31/1614	94/8164	0.001664	0.036821	0.033505	31

Table S13 Functional enrichment of genes in CDS where CPC-AIS are located and observed at least 10 samples

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0031424	keratinization	24/1215	85/18800	4.44E-10	2.18E-06	2.10E-06	24
BP	GO:0009913	epidermal cell differentiation	42/1215	230/18800	8.08E-10	2.18E-06	2.10E-06	42
BP	GO:0030216	keratinocyte differentiation	34/1215	167/18800	1.81E-09	3.27E-06	3.14E-06	34
BP	GO:0045109	intermediate filament organization	20/1215	68/18800	5.74E-09	7.76E-06	7.45E-06	20
BP	GO:0045104	intermediate filament cytoskeleton organization	22/1215	88/18800	2.76E-08	2.99E-05	2.87E-05	22
BP	GO:0045103	intermediate filament-based process	22/1215	89/18800	3.45E-08	3.11E-05	2.98E-05	22
BP	GO:0008544	epidermis development	51/1215	355/18800	6.46E-08	4.99E-05	4.79E-05	51
BP	GO:0043588	skin development	42/1215	296/18800	1.29E-06	0.000869	0.000834	42
BP	GO:0033135	regulation of peptidyl-serine phosphorylation	25/1215	144/18800	5.37E-06	0.003038	0.002917	25
BP	GO:0071287	cellular response to manganese ion	7/1215	13/18800	5.62E-06	0.003038	0.002917	7
BP	GO:0030574	collagen catabolic process	12/1215	42/18800	9.06E-06	0.00445	0.004271	12
BP	GO:0018105	peptidyl-serine phosphorylation	41/1215	315/18800	1.48E-05	0.006206	0.005958	41
BP	GO:0033138	positive regulation of peptidyl-serine phosphorylation	20/1215	107/18800	1.49E-05	0.006206	0.005958	20
BP	GO:0018209	peptidyl-serine modification	43/1215	338/18800	1.64E-05	0.006325	0.006072	43
BP	GO:0007605	sensory perception of sound	25/1215	156/18800	2.26E-05	0.008129	0.007803	25
BP	GO:0010042	response to manganese ion	8/1215	22/18800	4.20E-05	0.014182	0.013613	8
BP	GO:0050954	sensory perception of mechanical stimulus	26/1215	177/18800	7.21E-05	0.022934	0.022015	26
BP	GO:0030198	extracellular matrix organization	38/1215	307/18800	9.00E-05	0.025391	0.024373	38
BP	GO:0001655	urogenital system development	42/1215	352/18800	9.13E-05	0.025391	0.024373	42
BP	GO:0043062	extracellular structure organization	38/1215	308/18800	9.64E-05	0.025391	0.024373	38
BP	GO:1902229	regulation of intrinsic apoptotic signaling pathway in response to DNA damage	10/1215	38/18800	0.000109	0.025391	0.024373	10
BP	GO:0045229	external encapsulating structure organization	38/1215	310/18800	0.000111	0.025391	0.024373	38
BP	GO:0007626	locomotory behavior	27/1215	192/18800	0.000113	0.025391	0.024373	27
BP	GO:0048839	inner ear development	27/1215	192/18800	0.000113	0.025391	0.024373	27
BP	GO:0072001	renal system development	38/1215	312/18800	0.000126	0.027337	0.026242	38
BP	GO:0006814	sodium ion transport	32/1215	249/18800	0.000157	0.032677	0.031367	32
BP	GO:0010975	regulation of neuron projection development	48/1215	431/18800	0.000163	0.032677	0.031367	48
BP	GO:0043583	ear development	29/1215	219/18800	0.000187	0.03594	0.034499	29
BP	GO:0045747	positive regulation of Notch signaling pathway	11/1215	48/18800	0.000193	0.03594	0.034499	11
CC	GO:0045095	keratin filament	21/1275	102/19594	2.07E-06	0.001306	0.001212	21
CC	GO:0043202	lysosomal lumen	17/1275	97/19594	0.00016	0.049453	0.045878	17
CC	GO:0043025	neuronal cell body	52/1275	482/19594	0.000235	0.049453	0.045878	52
MF	GO:0030280	structural constituent of skin epidermis	19/1238	37/18410	2.56E-13	2.54E-10	2.37E-10	19

Table S14 Genes and Frequencies of CDS Regions Covered by CPC-specific Ancient Introgression Segments.

Frequency	Genes
50	CSMD2
48	IL17RA,COL27A1
46	ANKS1B
43	CTNND2
41	PTPRN2
40	TSPAN11
39	SLC35F3,NCOR2
38	RIMBP2,AGBL1
37	SLC35F4,HECW2,ETFBKMT
36	NAV2
35	PCSK6,AMN1
33	COL13A1
31	TMEM121B,TIAM1,FZD10
30	TRA2B,LDLRAD3
29	NPAS3,CSMD1
28	MYO3B
27	MMP17
26	KAZN
24	KCNQ5,GRM7,B9D1
23	EPN2,ARHGAP26,ANK2
22	EYS
21	TMEM132C,TMED3,SLC9A9,SINHCAF,RELN,FSTL5
20	TMEM132D,KCNQ1,DYNC2I1,APP
19	SORL1,PYGL,MGAT5
18	ULK4,TRAK1,NTRK2,ESYT2
17	RPUSD1,PRKAG2,PFKFB3,HDHD5,GABRB3,GABRA5,GAB4
16	NLRP5,NCAPG2,MAP3K5,CLMP,ALPK1
15	H3-5
14	SLC38A1,SHANK2,PCNX2,LIPA,CCSER1,CACNA2D1
13	SLC47A1,SHC2,PTPRF,PRKN,GALNT9
12	ZGRF1,WDR27,TMEM247,SGSM1,RTN4R,PRTFDC1,PRMT8,LDLRAD4,ITGA11,EPB41L3,ENSG00000285733,ELMO1,ADA2
11	TRPM3,SLC13A3,SH2D4A,RFX2,RAI14,PACRG,MRO,KIRREL3,ICE1,FKBP5,EP400,DPP6,CRACR2A,CAPRIN2,ABCB5
10	ULK2,PIWIL1,PHACTR2,PDE7B,MYH13,MLNR,GZMH,FNDC3A,DDX11,CRYL1,CDADC1,CAMK1D,CAB39L,C10orf90,ANKRD44,ALDH3A1,ADAMTS16
9	ULK1,TRPV3,RNF112,RBM47,NOC4L,MFAP4,MAPK7,KLRD1,IPO8,IGSF21,GABRB2,DLGAP1,DLG2,DISC1,DGKH,CORO2B,CNIH3,CAMKK2,C1orf94
8	SNX5,SEZ6L,PUS1,PLCH1,NEUROG2,MLIP,MGME1,MAPK4,KCTD16,DDX51,AQP12B,AQP12A
7	ZDHHC8,USP2,TRMT2A,TBC1D17,SULT1C4,ST8SIA2,SP8,SLC39A11,RPH3AL,RBM20,RANBP1,PRRX2,PNKP,PHF21B,PEX3,NUGGC,NTMT1,MLPH,LIMS1,KCNIP4,GRIK3,GCC2,GAREM1,FUCA2,ETV6,DGCR8,CDCP1,C9orf50,ASB6,AKT1S1,AKAP12,ABTB1
6	ZBTB8A,UNC45B,TXNDC15,TSPYL5,STIMATE-MUSTN1,STIMATE,SPECC1,SLIT2,SLC39A12,SLC35G3,SFMBT1,SDK2,RNASE2,RFT1,RAX,PCBD2,PAXIP1,P2RX7,OLFML2B,MUSTN1,MOB3B,MGLL,MGAM2,METTL17,MEOX1,ITIH4,ITIH3,HERC1,GPR35,GOLIM4,GNG12,FRMD1,FOXN3,FCRLB,FCRLA,F13A1,ENSG00000272305,ENSG00000254553,ENSG00000243696,DNAH5,CREBL2,CCL3,CATSPER3,BRPF3,AKAP11,ADA,ACYP2
5	ZBTB8B,XRCC4,USP3,UAP1,TMTC4,TLN2,TIE1,TBC1D16,SULT1C3,SMPD3,SMIM31,SLFN11,RPS6KA2,RIPOR2,RCSD1,RANBP2,RAB37,PRSS41,PRMT7,POP1,PLSCR1,PLCH2,P2RX4,OVOL2,NUAK1,NTM,MDGA2,MAP4K2,LRP2,LMF1,ITIH1,ITGB2,INSR,IL20,H3S3T2,GREB1,GPR19,GPR176,GLRX5,GLIS1,GALNT17,FBXL22,FAM110A,ENSG00000258989,ENSG00000249209,ENPEP,ELP3,EDDM13,CXCL13,CNTN3,CLIC5,CD4,CCL20,CCDC138,C5orf49,C5orf15,ATP5PO,ATP2B2,ARNT2,ANGPT4,ANAPC5,AIG1,AFAP1L2,ADAT2
4	ZNF804A,ZNF438,ZBTB40,VIPR2,TMEM167A,TMEM127,TMEM120B,THEM5,TARBP1,TAB1,SZT2,SYNGR1,SYNE3,SUSD2,SULT1C2,STRN3,STARD7,SRPK1,SQOR,SOX8,SNRNP200,SLC2A9,SLC26A8,SCGB2A2,SCGB1D4,SCGB1A1,RNASE3,PPARD,PLXNC1,PGLYRP4,PGLYRP3,PCSK5,ORA1,ODF3L2,OASL,NEURL3,NCAPH,MTHFD1L,MTARC1,MPL,MORN3,MED8,MAPK14,LORICRIN,LHFPL5,KDM2B,ITPR1L1,HYI,HEATR5A,GPC5,GALNTL6,FXDY7,FXDY1,ENSG00000286231,ENSG00000249624,ELOVL1,EIF2AK4,E DIL3,DYNC111,DMRTB1,DENND5B,DEF6,CYP4F8,CLPSL1,CLIP1,CIAO1,CDC20,CD9,CCDC149,CABIN1,BPIFA3,ASRGL1,ARMC12,AMER2,AK4P3

Frequency	Genes
3	ZNF777,ZNF76,ZNF662,ZNF618,VDR,VDAC1,UHMK1,TSPEAR,TRPV1,TRIM9,TRIM66,TPD52L1,TNIP1,TNFAIP8,TMEM82,TMEM163,TMEM106C,STK32C,STAB1,SPNS3,SMIM29,SMARCB1,SLX4IP,SLITRK1,SLC5A7,SLC47A2,SLC2A1,SLC25A34,SIGLEC8,SEPTIN9,SCUBE3,SCGB2A1,SCGB1D2,SCGB1D1,RSC1A1,RPS24,RPL36AL,RNF34,RGL1,RESF1,POLR3A,POLE2,PLEKHM2,PIWIL4,NTN1,LRG1,NR3C2,NISCH,NCF2,MUC13,MTRF1,MRE11,MPC1,MMP11,MIS18BP1,MGAT2,MAPK13,LRR1,LGI4,LBHD2,KRBOX1,KCTD9,KCMA1,ISG20L2,HTT,HSD17B4,HMGA1,HLX,GPR83,GNRH1,GFOD1,GASK1A,FXDY3,FGD4,FANCM,FANCC,EXOC3L4,ENSG00000285526,ENSG00000273291,EGR3,DOCK5,DNAJC16,DNAH9,DNAAF2,DIPK1C,DERL3,DDI2,DAP,COLEC12,COL2A1,CLPS,CHSY3,CCL18,CASP9,C3orf33,BTBD2,BPIFA2,BICD1,BCAT1,B3GALT5,ATP2A3,ATP12A,ARPC5,ARHGAP10,APOBEC4,AP4S1,AKAIN1,AGMAT,AFAP1L1
2	ZNF746,ZBTB41,ZBTB2,XPR1,WIPF3,WDR89,VWF,TUFT1,TRAV39,TMEM192,TMEM121,TMEFF2,TIA1,TGFB2,TESPA1,TEDC1,TDKHK,TBX15,STK33,STIM2,SPRTN,SPNS1,SPARC,SNX27,SNRPG,SMIM34,SMIM11,SLC9C1,SLC6A6,SLC22A18,SLC11A1,SKI,SHPK,SGPP1,SF3B1,SERPINB6,SCN8A,SCG2,RPS29,RORC,RNASE12,RNASE11,RMND1,RIIAD1,RGS1,RFTN2,RAP1GAP2,RABEP2,RAB6C,PTPRG,PIP5K1C,PHLDA2,PGAP1,PEX10,PELI2,PEBP4,PDHA2,PCYOX1,PCGF3,PAH,ORTE24,OR6C6,OR13D1,OR13C9,OR13C8,OR13C5,OR13C2,OA23,NXPH1,NOS1AP,NOL10,NIPSNAP3B,NIPSNAP3A,NFATC2IP,NEMF,MYH8,MYH4,MYH2,MYH1,MTA1,MRPS6,MRPL9,MOB4,MIF,MGAM,MAPK8,MAP4K3,MAN1B1,LRRK1,LRP5,LPP,LINGO4,LHFPL6,LAT,KRTDAP,KRTAP12-4,KRTAP12-3,KRTAP12-2,KRTAP12-1,KRTAP10-9,KRTAP10-11,KRTAP10-10,KLHDC2,KLHDC1,KLF12,KIAA1614,KCNE2,ITPR1,IL10RB,IGLL1,IFNAR2,IFNAR1,IBTK,HSPE1-MOB4,HSPE1,HSPD1,HSD17B7,HIF1A,HERPUD1,GTFC3,GSTT2B,GSTT2,GRM4,GRIN1,GPR107,GP2,GCSAM,GCNT2,GAK,FOXDI,FNBP1,FEM1B,FCER2,FBXO42,FAM136A,F13B,EXOC8,EXOC1L,EXOC1,ENSG00000289697,ENSG00000287856,ENSG00000285762,ENSG00000262304,ENSG00000259171,ENSG00000259060,ENSG00000257341,ENSG00000251357,ENSG00000203546,ELAVL4,EGLN1,DTD2,DRD5,DPYS,DPP7,DMKN,DENND2B,DDX31,DDTL,DDT,DDR2,DCC,CTNS,CRY1,CRIP2,CRIP1,CRADD,CACDL,CPLANE2,CQO10B,CNTN6,CNTLN,CNGB3,CNGB3,CLU,CLN6,CFHR5,CFHR3,CFHR2,CFHR1,CFH,CEP83,CEP164,CELF3,CELA2B,CDC7,CD53,CD300LF,CD19,CCNB1IP1,CCDC68,CCDC170,CCDC150,C3orf52,C3,C2orf66,C2CD4D,C21orf140,C1orf134,BMF,B4GALT6,ATP6V1C2,ASPM,ARPIN-AP3S2,ARHGEF19,ARHGAP22,ARF6,APOLD1,APCDD1L,AP3S2,AOPEP,ANKRD34C,ANG,AL589765.1,AHNAK,AGMO,ABCA1
1	ZNF582,ZNF518B,ZNF219,ZNF133,ZKSCAN2,ZHX3,ZC3H12C,XRN1,XPO4,XKR3,WRAP53,WNT5B,WLS,WDR1,VWCE,VPS18,VAPB,USP10,URI1,UNC5C,UHRF1,UGT1A9,UGT1A8,UGT1A7,UGT1A6,UGT1A5,UGT1A4,UGT1A3,UGT1A10,UGT1A1,UBR2,UBE2V1,UAP1L1,TYMP,TUFM,TTC5,TSPYL4,TSPYL1,TSPAN9,TSNAX-DISC1,TSC1,TRPM8,TRPM5,TRPC1,TRNT1,TRDV3,TRDJ4,TRDJ3,TRDJ2,TRDJ1,TRDD3,TRDC,TRAV9-2,TRAV8-6,TRAV14DV4,TRAV12-3,TPTE2,TPBG,TP53,TOP1,TOGARAM2,TOGARAM1,TNS1,TNFSF14,TNFRSF14,TNFAIP8L3,TNFAIP2,TMPRSS2,TMPRSS15,TMEM53,TMEM255B,TMEM253,TMEM250,TMEM233,TMEM216,TMEM178A,TMEM138,TMEM132B,TKFC,THUMP2,THRAP3,THBS1,TGM6,TESMIN,TCP11,TCF21,TBP,TBC1D22A,SYT9,SYNJ2BP-COX16,SYNJ2BP,SYCE3,SVIL,SUSD5,STX2,STPG2,STON2,STK40,STK32B,SRD5A3,SPNS2,SPINT1,SPATA46,SPATA17,SPACA9,SOX18,SOGA3,SNX10,SNAP25,SMG7,SMARCD1,SLC49A4,SLC43A2,SLC39A8,SLC35E3,SLC19A3,SLC12A6,SKIDA1,SIM2,SH3D21,SH2B1,SGCA,SFXN5,SFXN1,SETD7,SEC24A,SEC16A,SDHAF2,SCTR,SCO2,SCN3B,SCN10A,SCLT1,SBF1,SAPCD2,SAMD14,RUFY4,RUBCNL,RSPO3,RRP1B,RPS8,RPS10-NUDT3,RPS10,RPL27A,RPGRIP1L,RNF220,RNF113B,RNASE9,RIMKLB,RIDA,RHOV,RGS21,RGS18,RFLNA,RELB,RDX,RDH10,RAP1B,RACGAP1,RABL6,RAB27B,RAB20,RAB11FIP3,QSER1,PXT1,PWWP2B,PTTG1P,PTPRH,PTPN6,PTGIS,PTCHD4,PSMD1,PSMB1,PRXL2B,PRSS12,PRRG4,PRR5-ARHGAP8,PRPF39,PROC,PPP6R3,PPP6R2,PPP5C,PPP1R9B,PPP1R32,PPM1L,PPIL1,POLR3G,POLG,PNP,PNKD,PMEL,PLXNA4,PLS1,PLOD1,PLCL1,PLCG1,PKHD1,PHPT1,PHIP,PHF10,PGM1,PGA5,PGA4,PGA3,PEPD,PEDS1-UBE2V1,PEDS1,PDXP,PDXK,PDIA6,PCDC2,PCOLCE2,PCDH1,PCARE,PARP2,PAQR9,OTUD4,OSBPL3,OSBP2,OR6Q1,OR6C74,OR5AU1,ODF3B,NUP35,NUP107,NUDT3,NUBPL,NT5DC1,NRROS,NRP2,NPDC1,NFKB1,NETO1,NEK4,NDUFA10,NDST3,NDRG2,NCN1,NCAPH2,NBPF3,MYO10,MVD,MTUS1,MTIF3,MSX1,MRPL46,MPHOSPH8,MMP21,MMEL1,MLLT10,MKRN3,MKKS,MIOX,MGAT4A,MFAP5,MERTK,MEP1B,MED10,MDM2,MDFIC2,MARS2,MAPK8IP2,MAP7D1,MAP3K19,MANBA,MAMDC4,LYZL2,LYSMD3,LSAMP,LRRTM1,LRFN2,LPIN3,LONRF2,LNPK,LMF2,LHX9,LGI2,LGALS1,LCN9,LARS2,KRTAP10-5,KRTAP10-4,KRTAP10-12,KRT19,KRT15,KRBA1,KLK14,KLHL3,KLHL28,KLHDC7B,KIF2C,KIAA2013,KIAA0408,KCNH5,KCNF1,KCNE1,KCNAB1,KATNBL1,JADE1,ITIH5,ITGA1,IRAK1BP1,INPP5A,INO80,INHBB,IL17D,IKZF3,IKZF2,IFT74,IFT43,HTR2B,HSPBAP1,HSF2BP,HPCAL1,HOXD9,HOXD13,HOXD12,HOXD11,HOXD10,HMCN1,HLA-DRB5,HK2,HECA,HDCC2,HDAC11,H2BC12L,GYS1,GTFC3A,GRP,GRK1,GRHL1,GPX3,GPR33,GPR32,GPR158,GPD1,GPATCH2,GN2,GNAO1,GLT6D1,GK2,GJB6,GFRA2,GBGT1,GAL,GABARAP1,FUT7,FURIN,FSIP2,FSIP1,FRMD5,FOX11,FMO4,FKBP3,FIG4,FGF,R,FGL1,FGFR2,FES,FBXO21,FBXL17,FBLIM1,FARP1,FANCI,FAM131B,EXOC3L2,EXX2,EVA1B,ESD,ERMARD,ERGIC1,ENTPD2,ENSG00000288695,ENSG00000285304,ENSG00000285245,ENSG00000279073,ENSG00000260234,ENSG00000260170,ENSG00000256591,ENSG00000255330,EMILIN3,EMC4,EFNB3,EEF1AKMT1,EDRF1,EDF1,DYNLT2,DRD1,DPYSL4,DMRTA1,DLL4,DHX15,DHCR24,DEPDC7,DENND1B,DDX46,DDB1,DACH1,CYP8B1,CYP24A1,CYB5B1A3,CTNNB1,CRMP1,CRB1,CPT1B,CPT1A,CPSF7,CPNE5,CPM,CPAMD8,COX16,COL4A2,COL1A1,CNTNAP5,CLRN1,CLPTM1L,CLPTM1,CLPSL2,CLNK,CLEC1B,CLEC12B,CLEC12A,CLDN6,CLASP2,CIZ1,CHODL,CHL1,CHKB-CPT1B,CHKB,CHD6,CHAC1,CFHR4,CFAP77,CFAP61,CFAP300,CFAP299,CEBPB,CDYL,CDKN1C,CDKN1B,CDH24,CDCA2,CDC42BPB,CD200,CCNT2,CCND1,CCN5,CCDC8,CCDC73,CCDC198,CCDC190,CCDC183,CCBE1,CAVIN3,CATSPERE,CATIP,CASP4,CAMSA1,CAMLG,CABS1,C6orf89,C6orf120,C5orf24,C4orf51,C4orf33,C2orf72,C1orf53,C1orf226,C1GALT1,C14orf28,BSND,BOLL,BMPR1B,BEST4,BBLN,BACE2,ATXN2L,ATR,ATP4A,ATP2A1,ATL2,ASIC1,ARSA,ARPIN,ARMH1,ARID1B,ARHGEF40,ARHGAP8,ARHGAP5,ARHGAP21,ARHGAP20,AQP6,AQP5,AQP2,APOC4-APOC2,APOC4,APOC2,AOAH,ANKS1A,ANKRD10,AMOTL1,AKT1,AK8,AJM1,AF131215.5,ADM2,ADAM21,ACTN1,ACTB,ACMSD,ACKR2,ACAN,ABHD17C,ABCA4,ABCA2

Table S15 Functional enrichment of genes in CDS where CPC- specific-AIS are located

Database	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count
BP	GO:0052697	xenobiotic glucuronidation	9/1040	11/18800	2.33E-10	1.22E-06	1.17E-06	9
BP	GO:0009812	flavonoid metabolic process	10/1040	16/18800	1.52E-09	3.98E-06	3.80E-06	10
BP	GO:0052695	cellular glucuronidation	9/1040	20/18800	4.52E-07	0.000791	0.000754	9
BP	GO:0006063	uronic acid metabolic process	9/1040	22/18800	1.21E-06	0.001271	0.001211	9
BP	GO:0019585	glucuronate metabolic process	9/1040	22/18800	1.21E-06	0.001271	0.001211	9
BP	GO:0007612	learning	23/1040	145/18800	4.69E-06	0.0041	0.00391	23
BP	GO:0008306	associative learning	16/1040	80/18800	6.50E-06	0.004868	0.004642	16
BP	GO:0050954	sensory perception of mechanical stimulus	25/1040	177/18800	1.51E-05	0.00943	0.008992	25
BP	GO:0007605	sensory perception of sound	23/1040	156/18800	1.62E-05	0.00943	0.008992	23
BP	GO:0061458	reproductive system development	45/1040	436/18800	4.31E-05	0.022613	0.021561	45
BP	GO:0048608	reproductive structure development	44/1040	433/18800	7.53E-05	0.035878	0.034209	44
BP	GO:0051928	positive regulation of calcium ion transport	18/1040	119/18800	9.35E-05	0.040846	0.038946	18
CC	GO:0005938	cell cortex	37/1107	310/19594	1.44E-05	0.009247	0.008447	37
CC	GO:0016324	apical plasma membrane	39/1107	358/19594	6.76E-05	0.021721	0.019842	39
CC	GO:0045177	apical part of cell	43/1107	424/19594	0.000152	0.025284	0.023096	43
CC	GO:0000145	exocyst	6/1107	16/19594	0.000157	0.025284	0.023096	6
MF	GO:0015020	glucuronosyltransferase activity	10/1062	34/18410	1.44E-05	0.013533	0.012885	10
MF	GO:0016758	hexosyltransferase activity	27/1062	198/18410	2.81E-05	0.013533	0.012885	27
MF	GO:0015267	channel activity	50/1062	489/18410	6.16E-05	0.015408	0.014672	50
MF	GO:0022803	passive transmembrane transporter activity	50/1062	490/18410	6.49E-05	0.015408	0.014672	50
MF	GO:0008194	UDP-glycosyltransferase activity	21/1062	144/18410	7.99E-05	0.015408	0.014672	21
MF	GO:0046873	metal ion transmembrane transporter activity	44/1062	428/18410	0.000148	0.023747	0.022611	44
MF	GO:0005216	ion channel activity	44/1062	442/18410	0.000301	0.035199	0.033516	44
MF	GO:0002039	p53 binding	12/1062	66/18410	0.000348	0.035199	0.033516	12
MF	GO:0044325	transmembrane transporter binding	18/1062	127/18410	0.000361	0.035199	0.033516	18
MF	GO:0099106	ion channel regulator activity	19/1062	138/18410	0.000365	0.035199	0.033516	19
MF	GO:0016757	glycosyltransferase activity	30/1062	271/18410	0.000478	0.0419	0.039896	30
MF	GO:0016247	channel regulator activity	19/1062	143/18410	0.000574	0.046097	0.043892	19
KEGG	hsa00053	Ascorbate and aldarate metabolism	10/454	30/8164	2.81E-06	0.000888	0.000805	10
KEGG	hsa00040	Pentose and glucuronate interconversions	10/454	35/8164	1.34E-05	0.00211	0.001911	10
KEGG	hsa00982	Drug metabolism - cytochrome P450	13/454	72/8164	0.000142	0.014998	0.013589	13
KEGG	hsa00983	Drug metabolism - other enzymes	13/454	80/8164	0.000421	0.025621	0.023214	13
KEGG	hsa00140	Steroid hormone biosynthesis	11/454	61/8164	0.000464	0.025621	0.023214	11
KEGG	hsa00860	Porphyrin metabolism	9/454	43/8164	0.000486	0.025621	0.023214	9
KEGG	hsa04610	Complement and coagulation cascades	13/454	85/8164	0.000767	0.034615	0.031363	13
KEGG	hsa00980	Metabolism of xenobiotics by cytochrome P450	12/454	78/8164	0.001149	0.042856	0.03883	12
KEGG	hsa05131	Shigellosis	26/454	247/8164	0.00123	0.042856	0.03883	26
KEGG	hsa05204	Chemical carcinogenesis - DNA adducts	11/454	69/8164	0.001356	0.042856	0.03883	11

Table S16 GeneAnalytics results of genes in CDS where CPC-specific-AIS are located

Score	Disease	Matched Genes	Matched Genes with Genetic Associations
61.02	Colorectal Cancer	AKT1, CCND1, CTNNB1, DCC, FANCC, FANCI, FGFR2, PKHD1, SLC9A9, TP53, UHRF1, BEST4, MTRF1, SCGB2A2, ACTB, CASP9, CCL20, CD4, CDCP1, CDKN1B, CLU, CYP24A1, DLL4, FES, FURIN, FZD10, GRP, HIF1A, HK2, INSR, KRT19, LGALS1, LRP5, MAPK13, MAPK14, MAPK8, MDM2, MIF, MMP11, MRE11, MTA1, NFKB1, NRP2, NUA1, POLE2, PPAR, SAPCD2, SEPTIN9, SLC2A1, SLIT2, SPARC, TGFB2, THBS1, TIAM1, TMEFF2, TOP1, TPBG, TYMP, UGT1A1, UGT1A6, UGT1A7, UGT1A9, VDR, WNT5B, FCRLA, FUT7, GPATCH2, LGI4, SLC11A1, UNC5C, ACAN, BICD1, CAB39L, CCDC150, CD200, CLEC12B, COL1A1, CORO2B, DENND5B, F13A1, GAL, KAZN, LHX9, LORICRIN, MAP4K2, MFAP4, NPAS3, NRROS, OLFML2B, RGS1, SLC6A6, SULT1C2, SZT2, VWF, ZBTB8B	AKT1, CCND1, CTNNB1, DCC, FANCC, FANCI, FGFR2, PKHD1, SLC9A9, TP53, UHRF1
60.18	Breast Cancer	ABCA1, AKT1, AOEPE, FANCC, FANCM, FGFR2, MRE11, SF3B1, SLC22A18, TP53, KRT19, ASPM, KRT15, SCGB1D2, SCGB2A2, ACTB, ATR, CASP9, CCDC170, CCND1, CCNT2, CDKN1B, CEBPB, CLU, CTNNB1, CYP24A1, DLL4, GNRH1, GPATCH2, GREB1, HIF1A, HMGA1, INSR, LGALS1, LRP5, MAPK14, MAPK4, MAPK7, MAPK8, MDM2, MMP11, MMP17, MTA1, MTUS1, NCOR2, NFKB1, NR3C2, NRP2, NTRK2, PAXIP1, PHLDA2, PLCG1, PPP5C, PRMT7, PTPRG, RABL6, RELB, SLC2A1, SPARC, SUS2, TAB1, TGFB2, THBS1, TOP1, TPD52L1, TSC1, TYMP, UHRF1, VDR, WNT5B, BCAT1, CHL1, COL1A1, CYP4F8, DMKN, FNDC3A, FSIP1, KLRD1, MGME1, MLLT10, OLFML2B, PDE7B, PNP, PPP6R3, QSER1, RAI14, RIDA, RPH3AL, ULK4, VAPB, WLS	ABCA1, AKT1, AOEPE, FANCC, FANCM, FGFR2, MRE11, SF3B1, SLC22A18, TP53, KRT19
47.78	Schizophrenia	CNTN6, CSMD1, CTNND2, DLG2, DPP6, HECW2, IRAK1BP1, LPP, NPAS3, PHIP, PRKN, RELN, RTN4R, ULK4, ZNF804A, DISC1, GABRB2, AKT1, APP, CHL1, CNIH3, CSMD2, CTNNB1, DLGAP1, DRD1, DRD5, EGR3, GABRA5, GABRB3, GRIK3, GRIN1, GRM4, GRM7, ITIH3, LRRTM1, NBPF3, NOS1AP, NTNG1, NTRK2, PAH, PPP1R9B, SLC2A1, SNAP25, ST8SIA2, TBP, TP53, TSNAX-DISC1, UHMK1, VIPR2, ZDHHC8	CNTN6, CSMD1, CTNND2, DLG2, DPP6, HECW2, IRAK1BP1, LPP, NPAS3, PHIP, PRKN, RELN, RTN4R, ULK4, ZNF804A, DISC1, GABRB2
47.08	Nervous System Disease	ARSA, ASPM, CLN6, CTNNB1, GNAO1, HECW2, HSD17B4, PNKP, POLG, RMND1, SRD5A3, TMEM216, ABCA1, ABCA4, ACTB, AKT1, APP, ATP12A, ATP4A, C3, CASP9, CCL3, CCND1, CD19, CD4, CDKN1B, CFH, CFHR2, CHL1, CLU, CRB1, ELAVL4, GJB6, GRIN1, HIF1A, HTT, IFNAR1, ITIH4, ITPR1, MAP3K5, MAPK14, MAPK8, MDM2, MERTK, NTRK2, PRKN, SCN8A, SLC2A1, SMARCB1, SNAP25, SORL1, TBP, TGFB2, THBS1, TP53, TRPV1, VDAC1, VWF	ARSA, ASPM, CLN6, CTNNB1, GNAO1, HECW2, HSD17B4, PNKP, POLG, RMND1, SRD5A3, TMEM216
44.04	Lung Cancer	AKT1, CTNNB1, FGFR2, PRKN, SLC22A18, TP53, CLIC5, GJB6, KRT15, PEBP4, SLC2A1, ARHGAP5, CASP9, CAVIN3, CCND1, CD9, CDCP1, CDKN1B, CDKN1C, COL4A2, DDR2, EFN3, ELAVL4, EPB41L3, FGL1, GRP, HIF1A, HK2, IFNAR1, ITGA11, KRT19, LGALS1, MAPK14, MAPK8, MDM2, MIF, NFKB1, PLCG1, SCGB1A1, SCUBE3, SEZ6L, SLFN11, SPARC, SRPK1, THBS1, TOP1, TYMP, UGT1A1, VDAC1, ASPM, CDC20, SUS2, TCF21, CAB39L, CCN5, COL1A1, CRIP1, CXCL13, FANCI, FGR, FXYD1, GPX3, KIF2C, MFAP4, MMP11, SLC39A8, UHRF1	AKT1, CTNNB1, FGFR2, PRKN, SLC22A18, TP53
40.55	Alzheimer Disease, Familial, 1	APP, UNC5C, ANKS1B, CAMKK2, DYNC111, GALNT17, KCNIP4, PAQR9, ABCA1, ABCA2, ACTB, AKT1, BACE2, CASP4, CASP9, CCL3, CFH, CLU, CTNND2, DHCR24, DRD1, FURIN, GAL, GRIN1, HTT, INSR, ITIH4, LRP2, MAP3K5, MAPK14, MAPK8, NCF2, NTRK2, P2RX7, PRKN, RELN, RTN4R, SLC5A7, SNAP25, SORL1, TBP, VDAC1, CAPRIN2, TNS1, AHNAK, CATSPERE, CRMP1, CSMD2, DDX46, DENND5B, DGCR8, FXYD1, FZD10, GPR158, ITGA11, ITGB2, ITIH3, KCNQ1, LRRTM1, NRP2, NTNG1, SLC2A1, TBC1D16, TPD52L1	APP, UNC5C
40.52	Retinitis Pigmentosa	ABCA4, CLRN1, CNGB3, CRB1, EYS, MERTK, PCARE, SNRNP200, ACTB, AKT1, APP, B9D1, CCND1, CEP164, CEP83, CFH, CLU, CNGA3, CNTLN, CTNNB1, CYP24A1, DNAAF2, DNAH5, ETV6, GABARAPL1, GJB6, GRK1, HIF1A, HTT, IFT43, IFT74, KLHL3, MDM2, MKKS, PKHD1, POLG, PRKN, RDH10, RPGRIP1L, SCLT1, SLC2A1, SLC6A6, SRD5A3, TBP, TMEM138, TMEM216, TMPRSS2, TP53, TRNT1, TSC1, ULK1, ZBTB8B	ABCA4, CLRN1, CNGB3, CRB1, EYS, MERTK, PCARE, SNRNP200

s24. Supplementary Figures

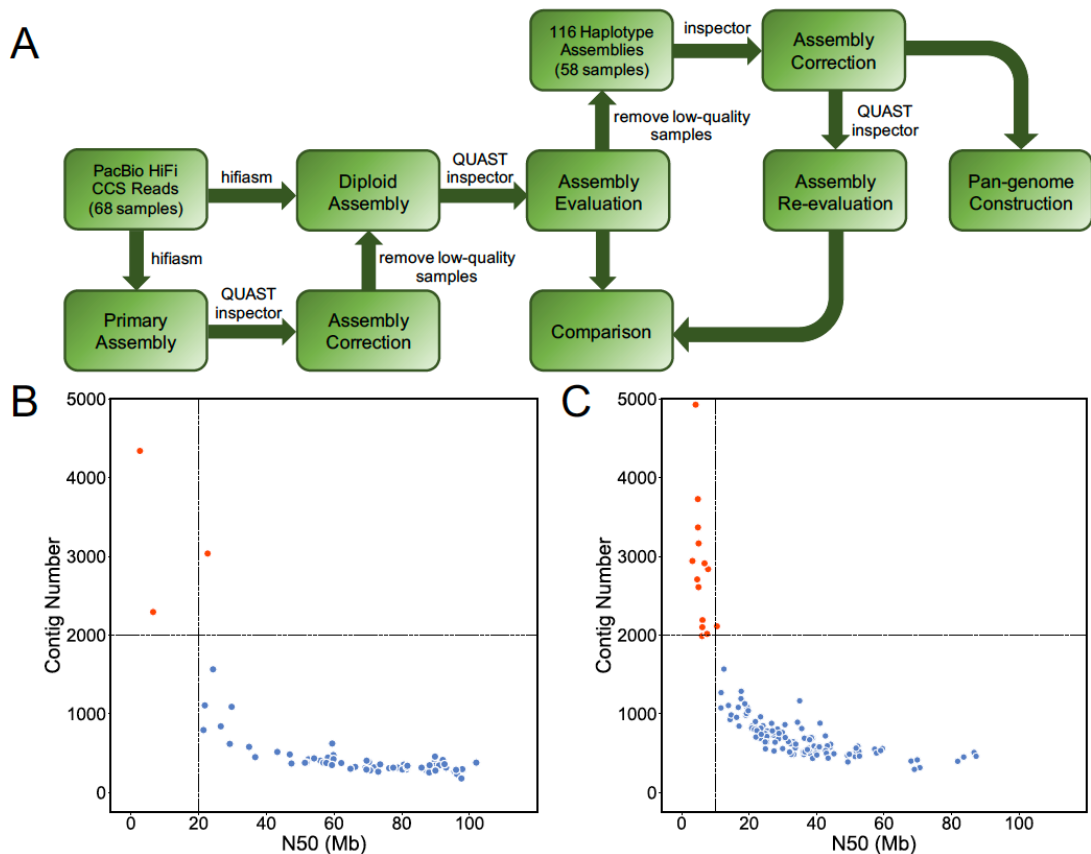


Fig. S2 | Assembly pipeline and quality control of CPC samples.

(A) Flowchart showing the quality control, assembly, and correction of CPC samples used for construction of pan-genome. (B) Quality control of primary assemblies of 68 CPC samples, 3 samples with N50 less than 20 Mb or contig number larger than 2000 were removed in subsequent analyses. (C) Quality control of diploid assemblies of 65 CPC samples, 7 samples of 14 haplotype assemblies with N50 less than 10 Mb or contig number larger than 2000 were removed in subsequent analyses.

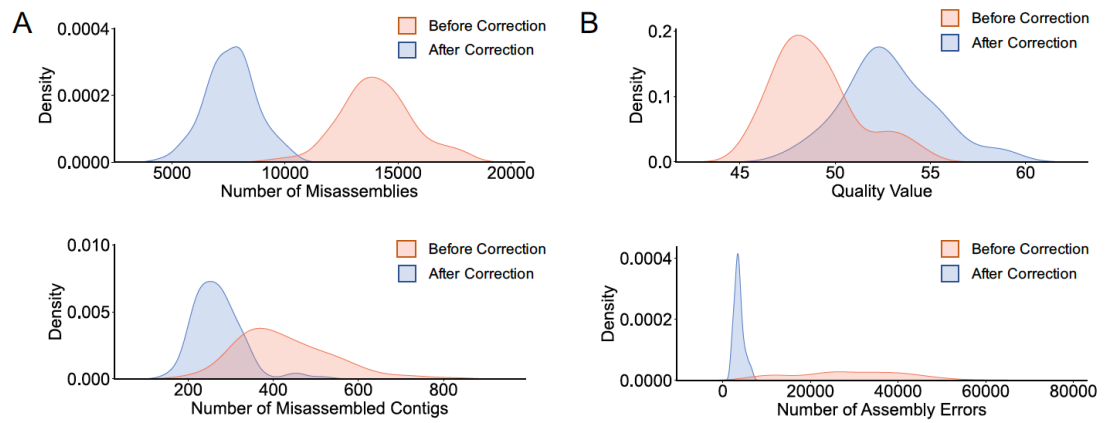


Fig. S3 | Evaluation of overall assembly correction of CPC genome assemblies.

Comparison of overall assembly errors and quality before and after assembly polishing, evaluated by (A) QUAST and (B) inspector.

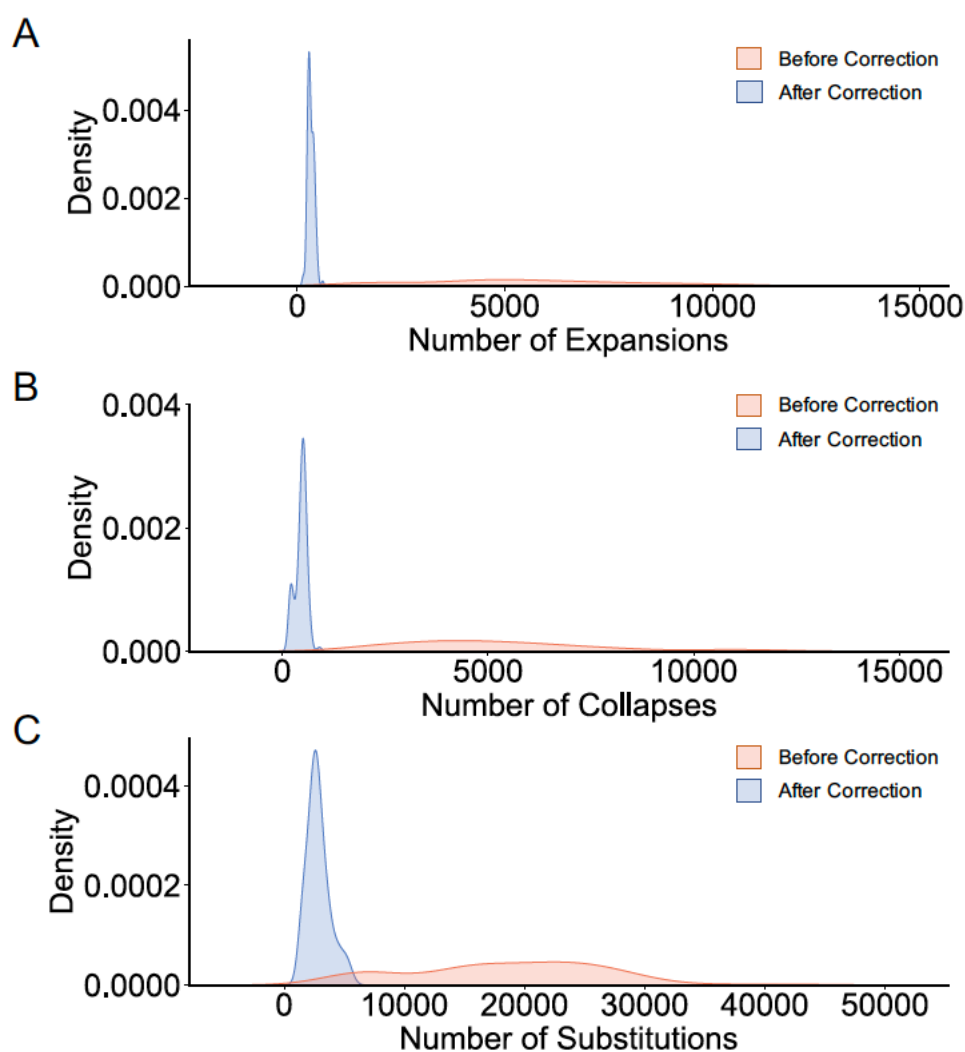


Fig. S4 | Evaluation of three types of assembly errors of CPC genome assemblies.

Comparison of three types of assembly errors, including (A) expansions, (B) collapses, and (C) substitutions, before and after assembly polishing evaluated by inspector.

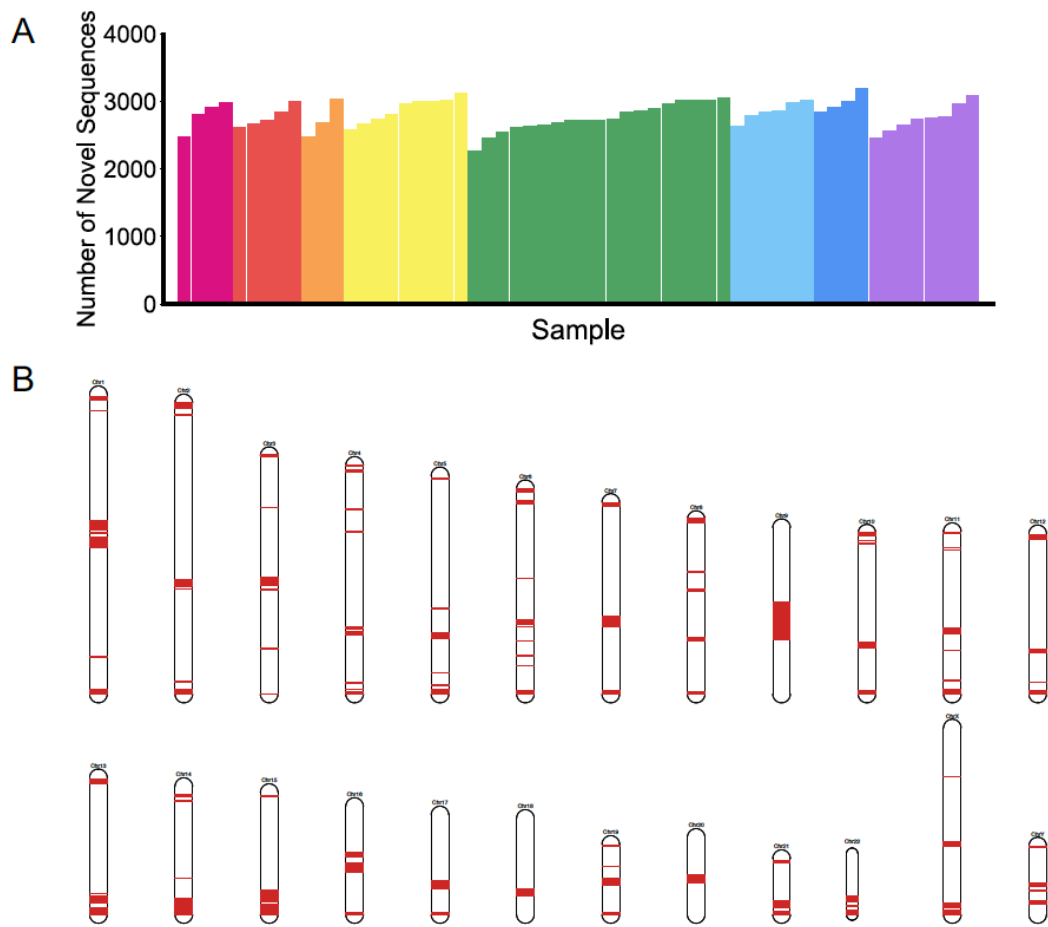


Fig. S5 | Novel sequences identified from CPC samples.

(A) Number of identified novel sequences (insertions larger than 1 kb) of CPC samples by aligning two phased assemblies of each sample to the T2T-CHM13 reference. (B) The 115 novel sequence hotspots on the chromosomes.

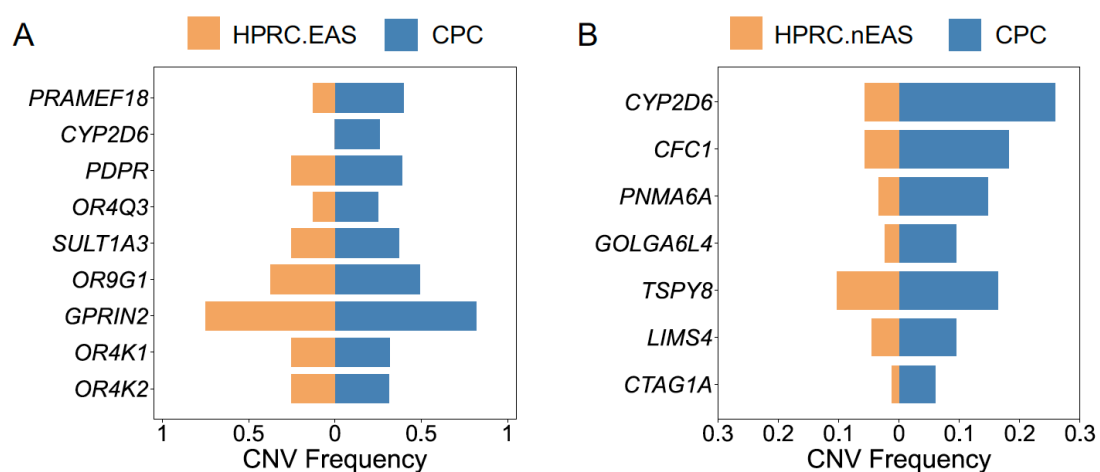


Fig. S6 | Overlapped CNV genes with higher frequency in CPC than that in two HPRC subsets.

Overlapped CNV genes showing higher frequency (> 5%) in CPC assemblies than in (A) HPRC.EAS and (B) HPRC.nEAS. HPRC.EAS: East Asian in HPRC; HPRC.nEAS: non-East Asian in HPRC.

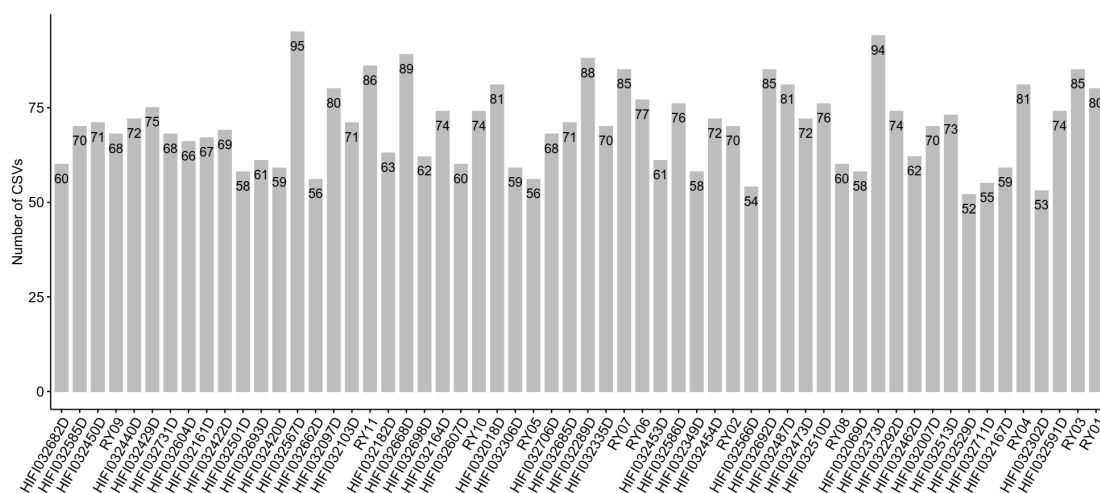


Fig. S7 | The count of CSVs in each sample.

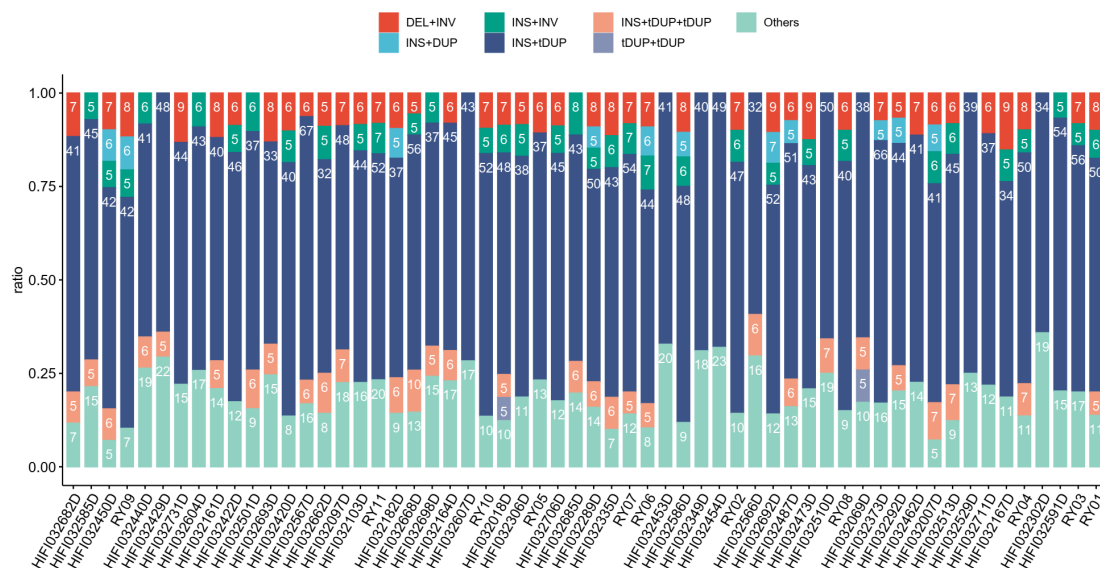


Fig. S8 | The number of CSVs for each CSV type in each sample.

The type of INS+IDUP is the most abundant CSV across all samples. Others indicate the CSV types with count smaller than five in each sample.

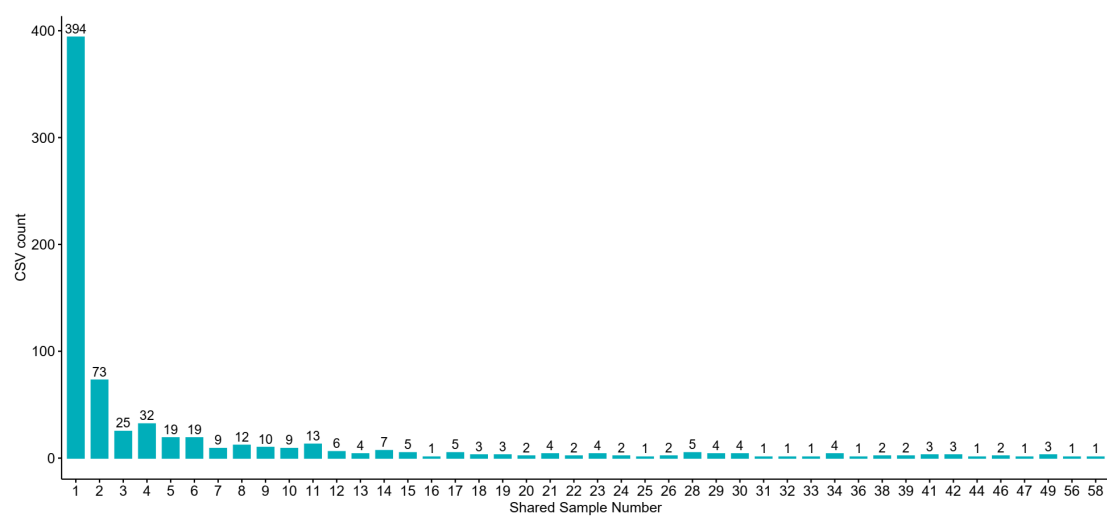


Fig. S9 | The count of CSVs with different shared sample numbers.

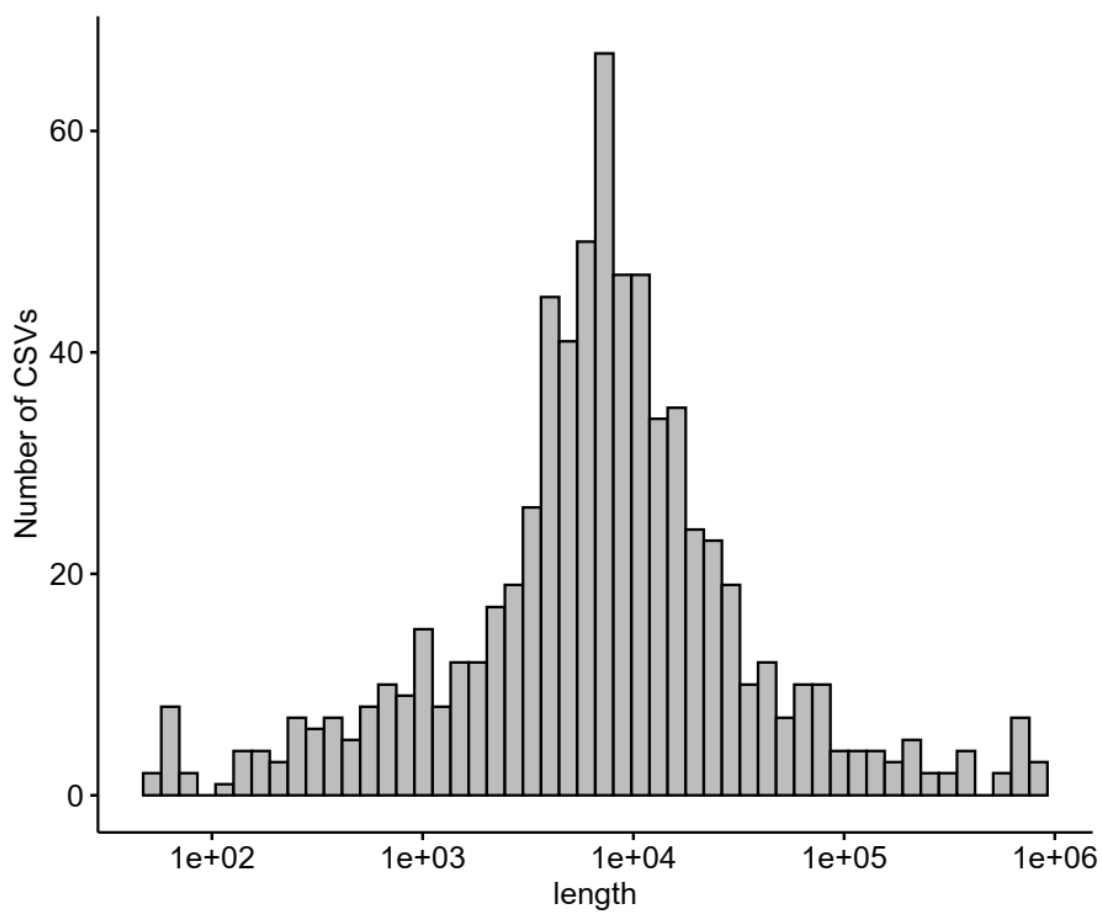


Fig. S10 | The length distribution of the final 706 CSVs.

We merged CSVs of all 58 samples by considering 80% reciprocal overlap and the same type, and obtained 706 CSVs in total.

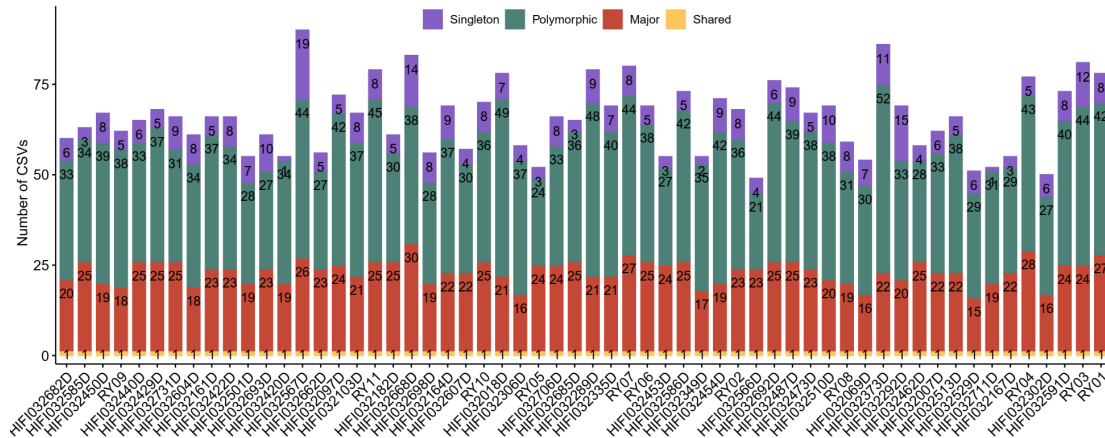


Fig. S11 | The number of CSVs in each CSV class.

We classified the CSVs by the shared sample numbers into Shared (present in all samples), Major (present in $\geq 50\%$ samples but not all), Polymorphic (present in more than 1 but $< 50\%$ of all samples), and Singleton (present in only one sample).

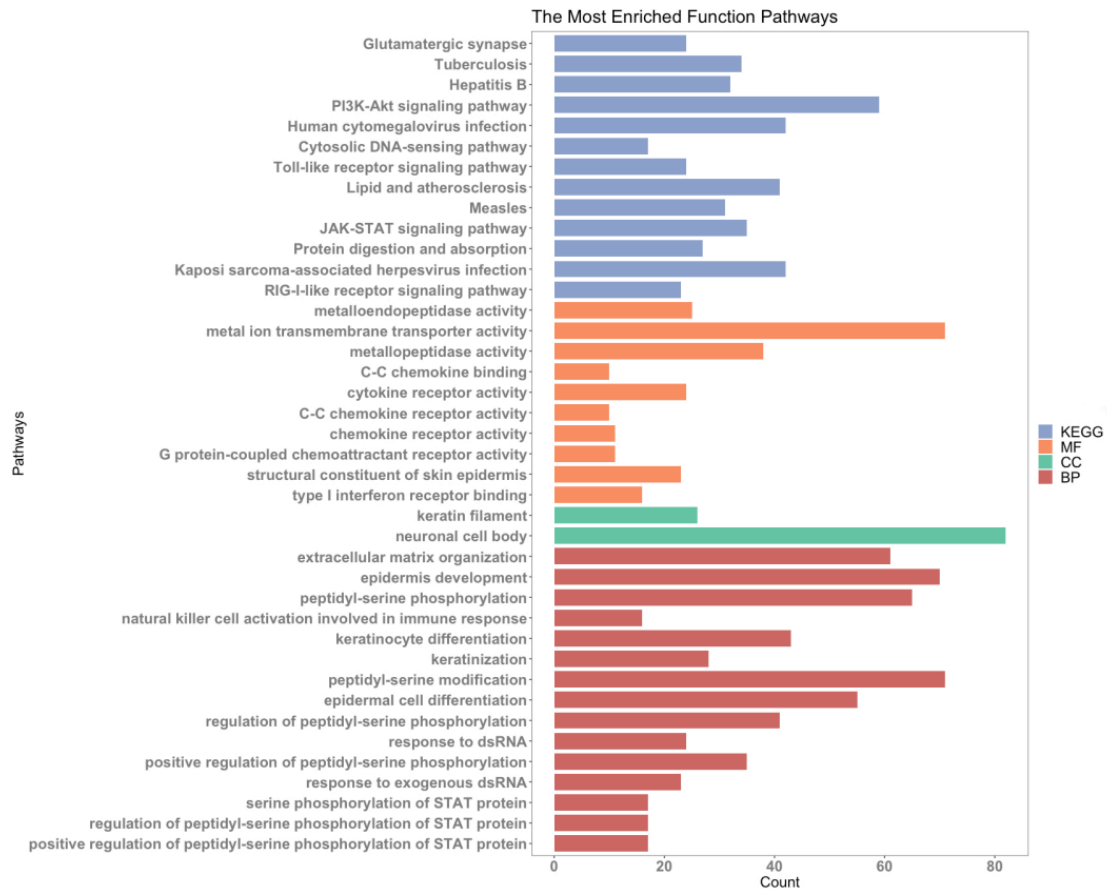


Fig. S12 | Functional enrichment of 2617 genes in CDS where CPC-AIS are located and observed in at least 5 samples.

GO and KEGG databases were used here. Differential color means differential database (blue, BP: GO Biological process; orange, CC: GO cellular component; red, MF: GO metabolic function; green, KEGG: KEGG database). The FDR-corrected p-values of these terms are all less than 0.05.

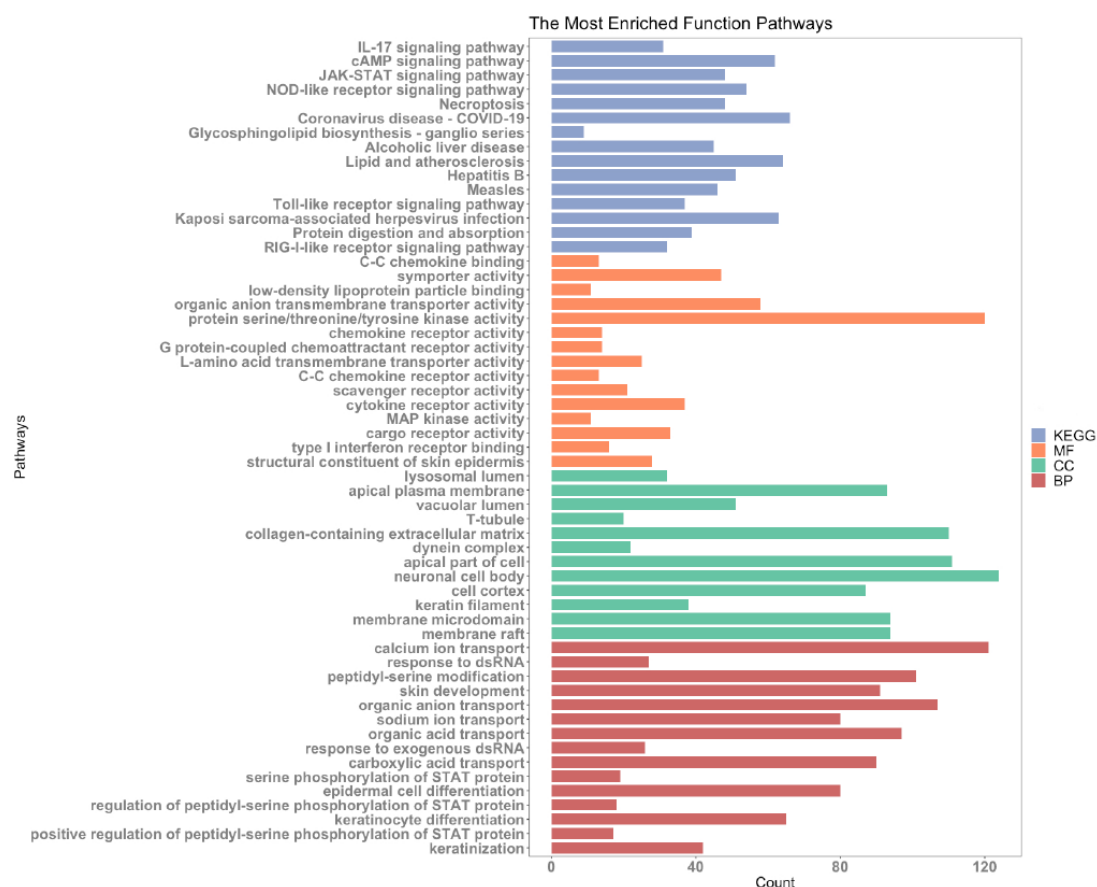


Fig. S13 | Functional enrichment of genes in CDS where AIS are located and observed in at least 1 sample.

GO and KEGG databases were used here. Differential color means differential database (blue, BP: GO Biological process; orange, CC: GO cellular component; red, MF: GO metabolic function; green, KEGG: KEGG database). The FDR-corrected p-values of these terms are all less than 0.05.

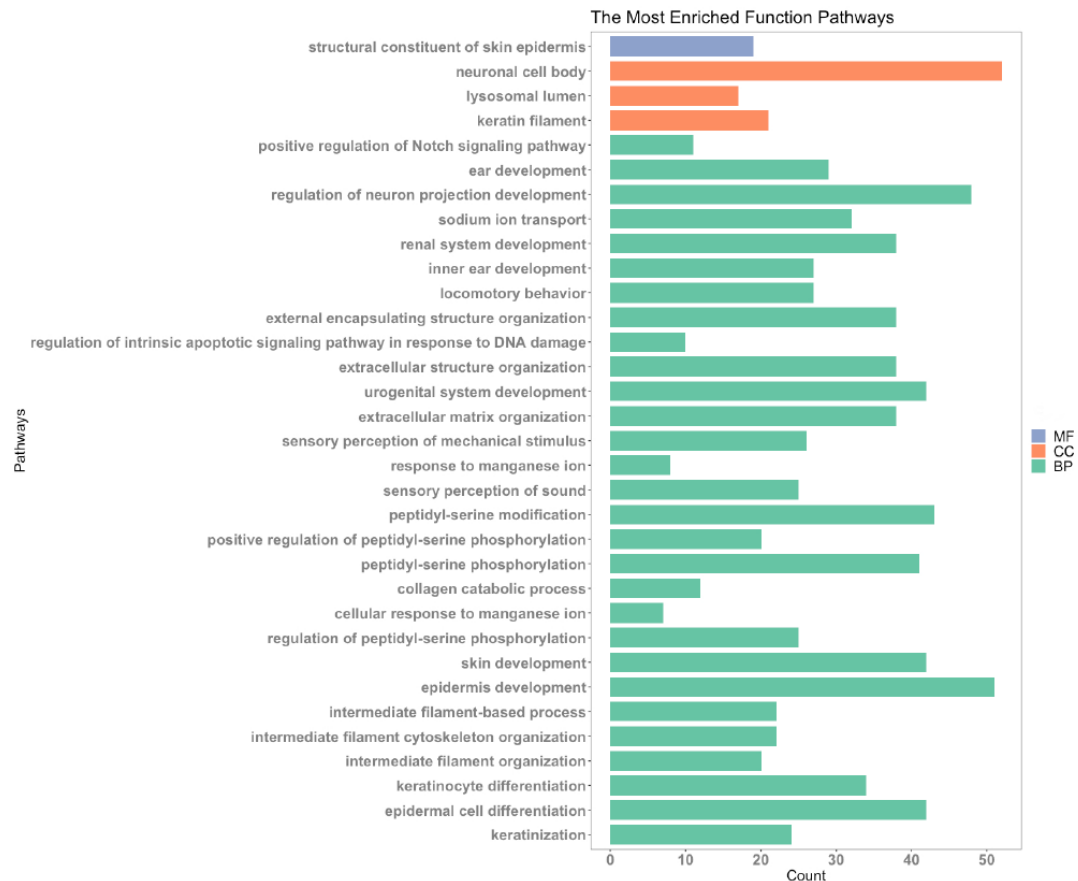


Fig. S14 | Functional enrichment of genes in CDS where AIS are located and observed in at least 10 samples.

GO and KEGG databases were used here. Differential color means differential database (blue, BP: GO Biological process; orange, CC: GO cellular component; red, MF: GO metabolic function; green, KEGG: KEGG database). The FDR-corrected p-values of these terms are all less than 0.05.

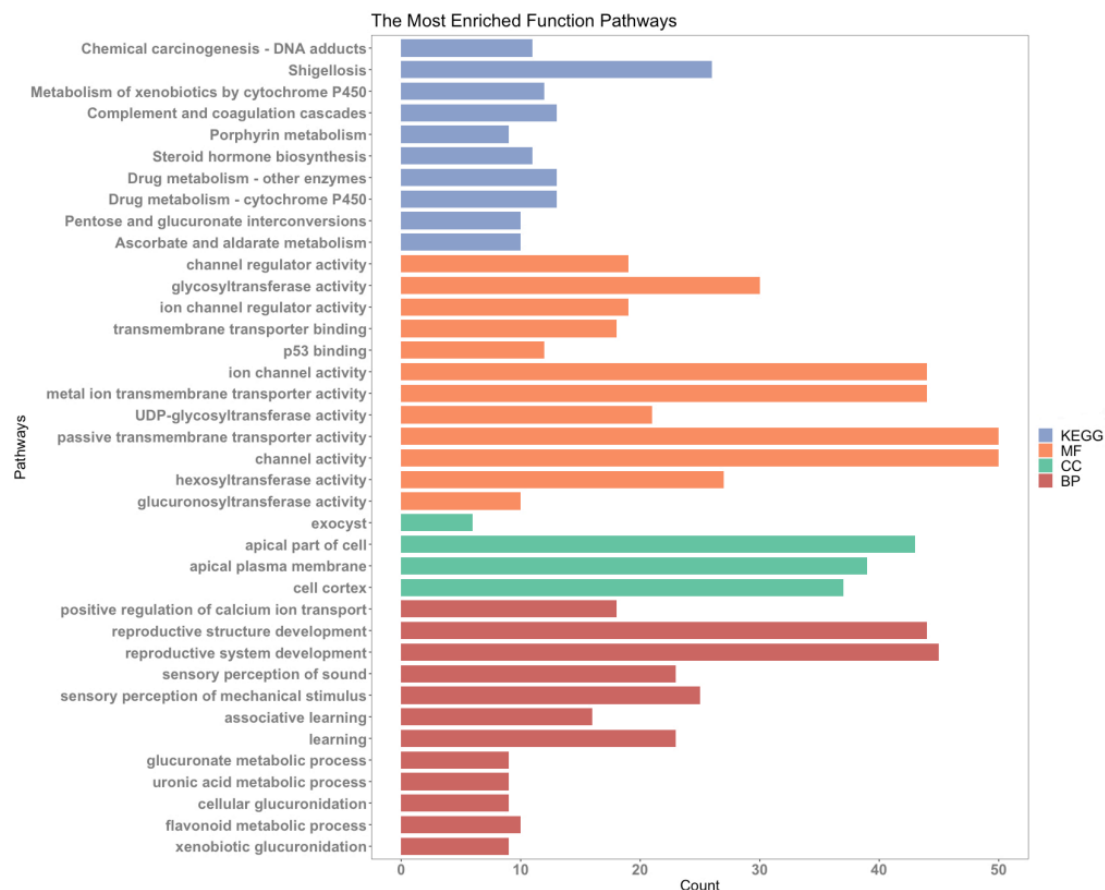


Fig. S15 | Functional enrichment of 1211 genes in CDS where CPC-AIS-specific are located.

GO and KEGG databases were used here. Differential color means differential database (blue, BP: GO Biological process; orange, CC: GO cellular component; red, MF: GO metabolic function; green, KEGG: KEGG database). The FDR-corrected p-values of these terms are all less than 0.05.

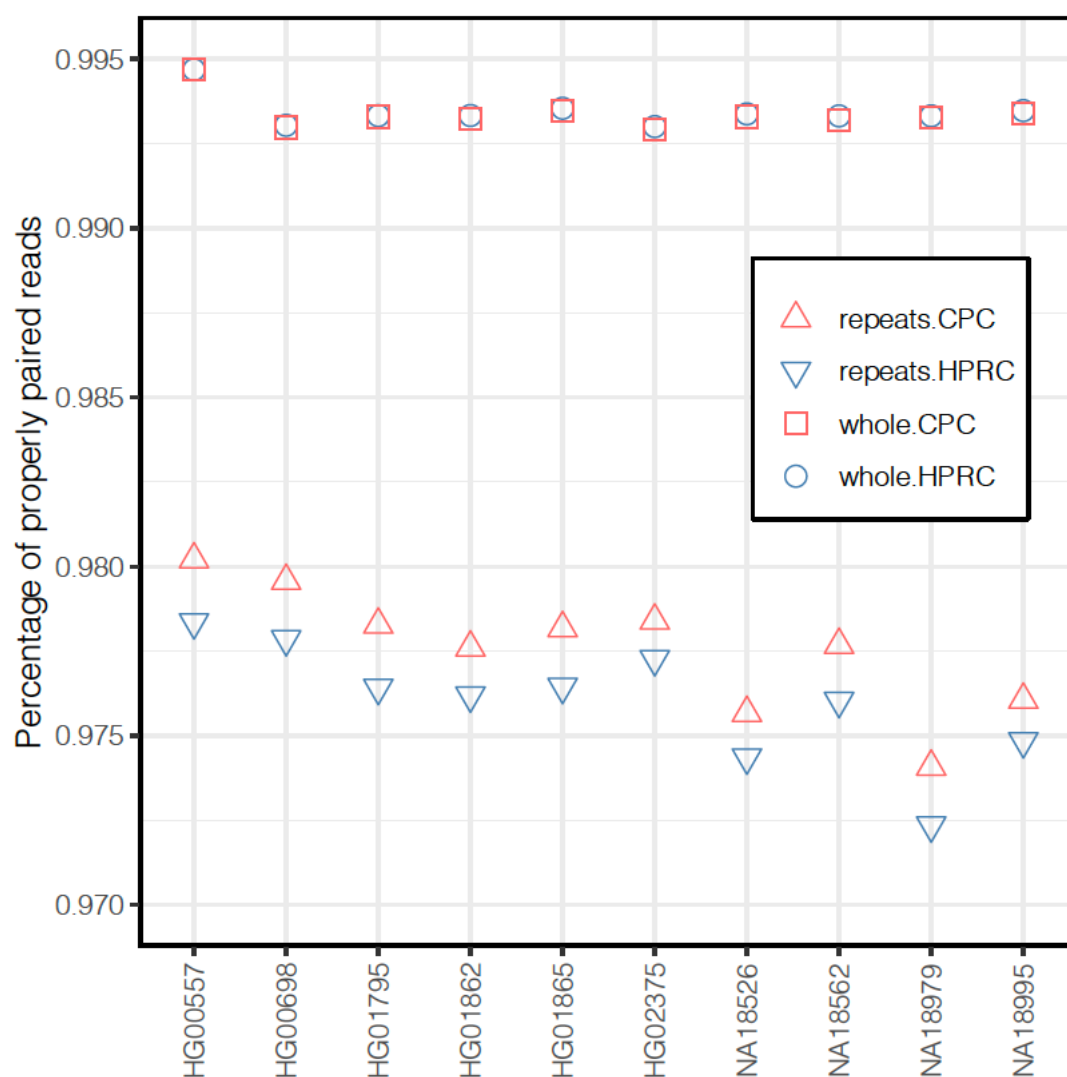


Fig. S16 | The proportion of properly paired reads of 10 East Asian samples from KGP.

The results were obtained by dividing the number of properly paired reads by the number of mapped reads. The effect of using HPRC reference or CPC reference at the genome-wide level is equivalent. However, in repeats region, a higher proportion of properly paired can be obtained by using CPC reference.

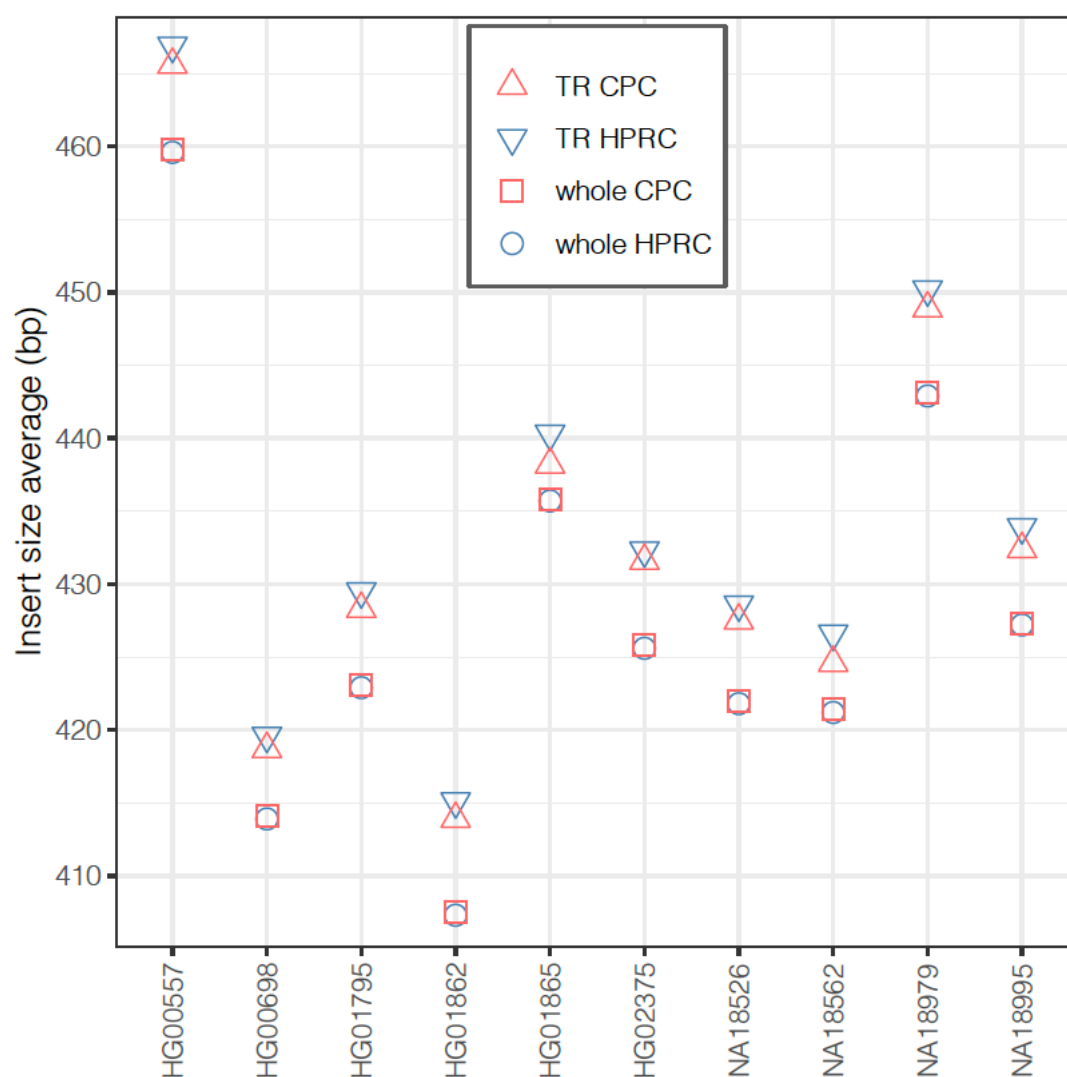


Fig. S17 | The insert size average over 10 East Asian samples.

In repeats region, the insert size obtained by using CPC reference alignment is closer to the genome-wide level.

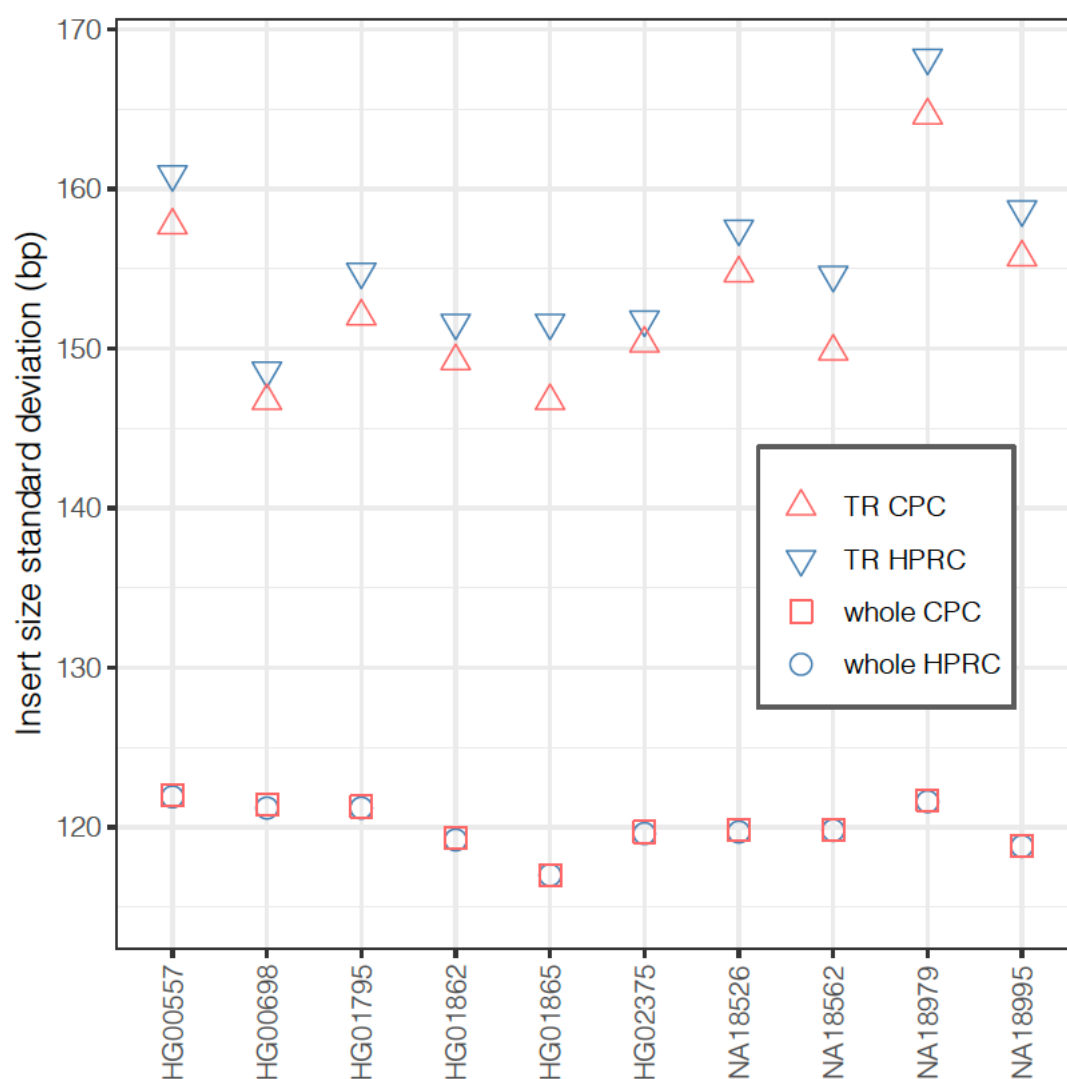


Fig. S18 | Standard deviation of insertion size at whole genome level and repeats region.

In repeats region, the standard deviation of insert size obtained by using CPC reference alignment is closer to the genome-wide level.

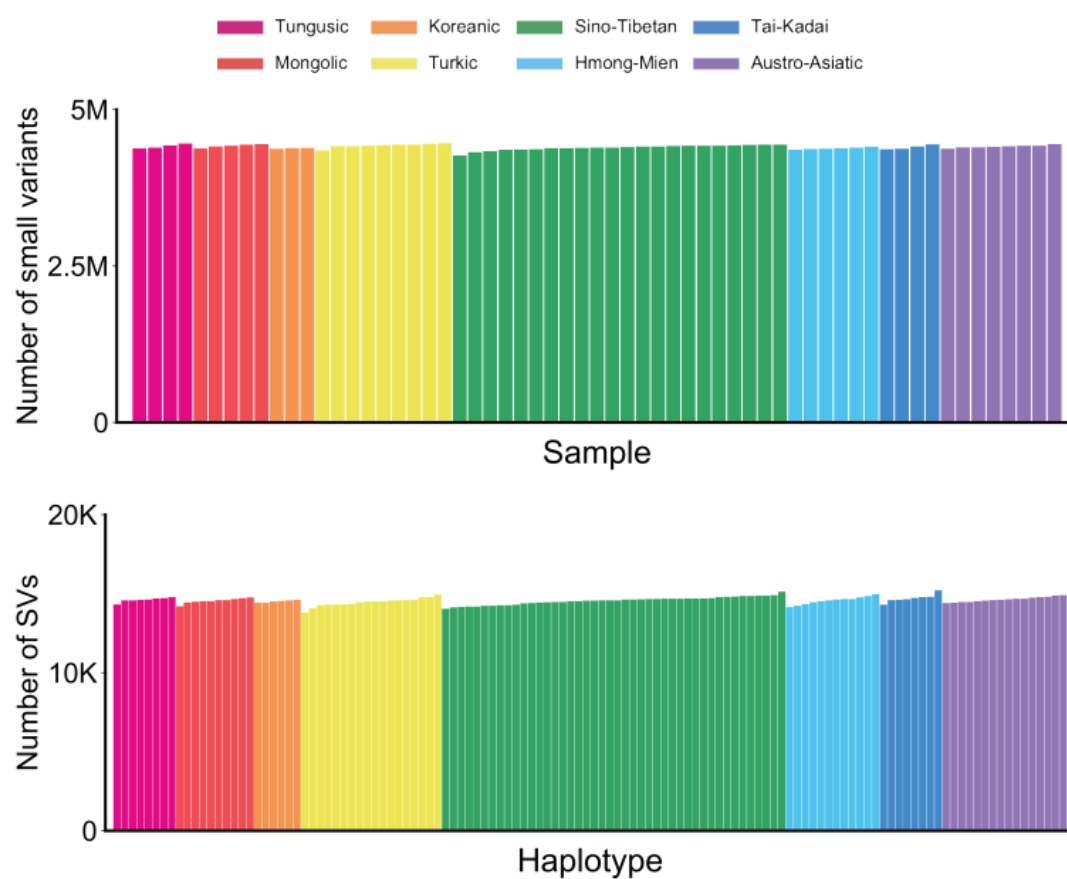


Fig. S19 | Number of identified small variants and SVs inferred from the pangenome graph.

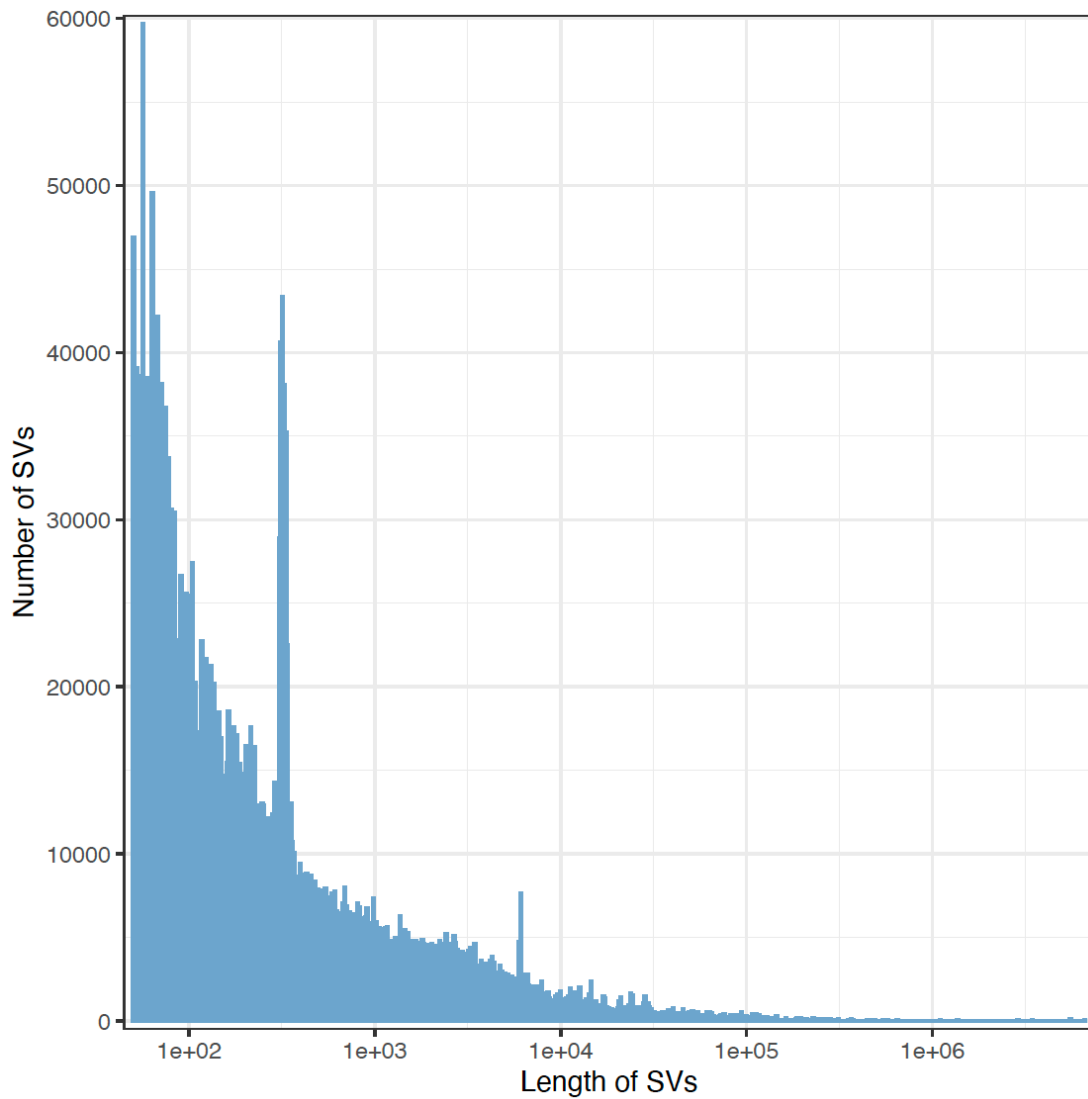


Fig. S20 | The length distribution of SVs in the CPC reference graph Pan-genome.

On CPC pan genome, the average length of SV is 3.2 kb, and the median length is 178 bp, which is higher than the length of short-reads. In addition, we observe modes at 300 bp peak for Alu, and 6 kb peak for L1.

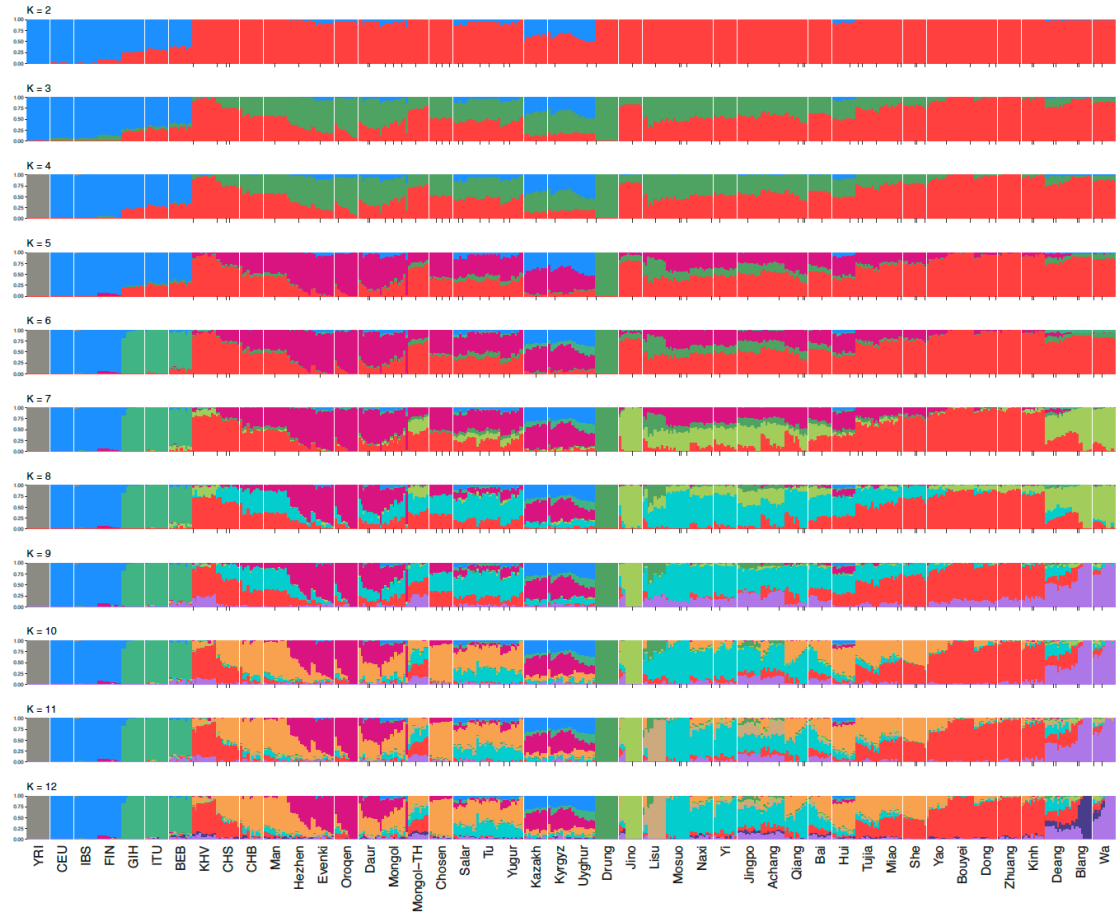


Fig. S21 | The ancestral component Inference of CPC cohort.

We randomly selected 10 samples from each population of CPC cohort after excluding low-quality data and related samples, and infer the ancestral composition of each sample by ADMIXTURE under the assumption that the number of ancestral components (K) is 2-12. The 58 CPC and 4 HPRC samples used in the construction of pan genome are marked with short vertical lines in the picture.

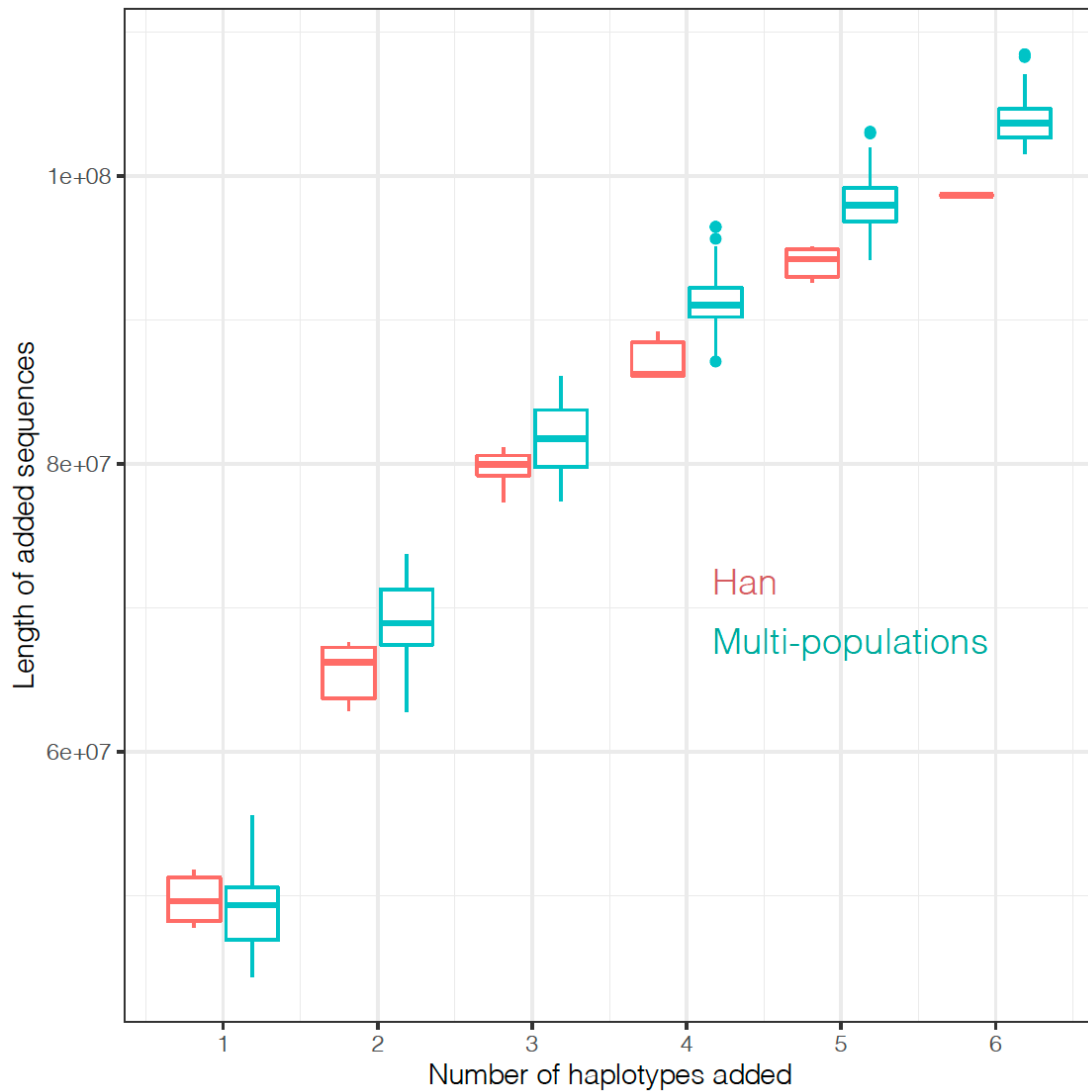


Fig. S22 | Growth of non-reference sequences with haplotype addition.

The cumulative length of non-reference (GRCh38) sequences with the addition of haplotypes in the CPC graph genome. The multi-population data in CPC cohort have a faster growth compared with the single Han population.

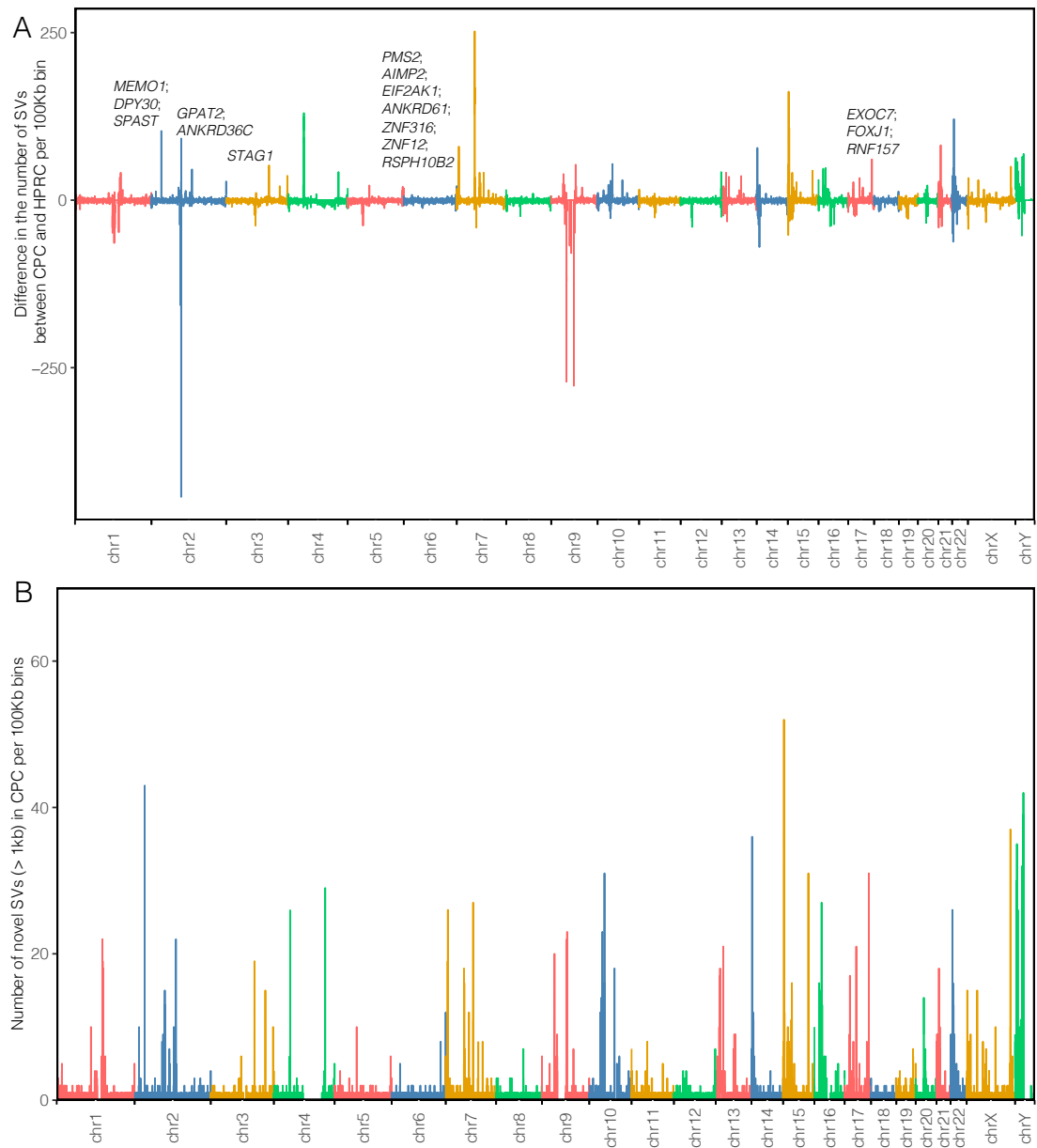


Fig. S23 | Comparison of SV distribution between CPC and HPRC reference Pangenomes.

(A) The quantity difference of SV between CPC and HPRC genomes per 100kb bin. (B) The number of CPC novel SV per 100kb bin.