

TP53 germline pathogenic variants in modern humans were likely originated during recent human history

Si Hoi Kou[†], Jiaheng Li[†], Benjamin Tam, Huijun Lei, Bojin Zhao, Fengxia Xiao and San Ming Wang^{*}

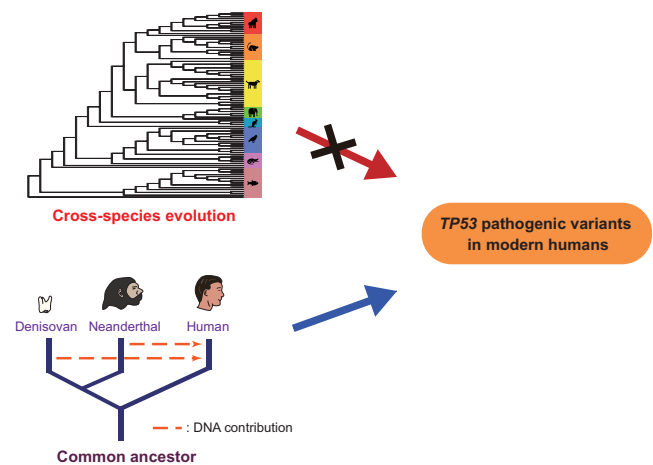
Ministry of Education Frontiers Science Center for Precision Oncology, Cancer Centre and Institute of Translational Medicine, Department of Public Health and Medical Administration, Faculty of Health Sciences, University of Macau, Macao SAR, China

Received March 01, 2023; Revised May 05, 2023; Editorial Decision May 11, 2023; Accepted May 16, 2023

ABSTRACT

TP53 is crucial for maintaining genome stability and preventing oncogenesis. Germline pathogenic variation in *TP53* damages its function, causing genome instability and increased cancer risk. Despite extensive study in *TP53*, the evolutionary origin of the human *TP53* germline pathogenic variants remains largely unclear. In this study, we applied phylogenetic and archaeological approaches to identify the evolutionary origin of *TP53* germline pathogenic variants in modern humans. In the phylogenetic analysis, we searched 406 human *TP53* germline pathogenic variants in 99 vertebrates distributed in eight clades of Primate, Euarchontoglires, Laurasiatheria, Afrotheria, Mammal, Aves, Sarcopterygii and Fish, but we observed no direct evidence for the cross-species conservation as the origin; in the archaeological analysis, we searched the variants in 5031 ancient human genomes dated between 45045 and 100 years before present, and identified 45 pathogenic variants in 62 ancient humans dated mostly within the last 8000 years; we also identified 6 pathogenic variants in 3 Neanderthals dated 44000 to 38515 years before present and 1 Denisovan dated 158 550 years before present. Our study reveals that *TP53* germline pathogenic variants in modern humans were likely originated in recent human history and partially inherited from the extinct Neanderthals and Denisovans.

GRAPHICAL ABSTRACT



INTRODUCTION

TP53 is one of the most important tumor suppressors. As the ‘guardian of the genome’, *TP53* detects intrinsic and extrinsic stress signals of DNA damage, hypoxia, mitophagy, telomere erosion, redox potential, epigenetic changes, etc., and responses by regulating global gene expression to maintain genome stability and cellular homeostasis (1,2). *TP53* is also a highly mutable gene that it is mutated in over half of human cancer types (3). The majority of the variants are located within its DNA-binding domain (DBD), and a group of high-frequency pathogenic variants (PVs) is known as ‘hotspot’ variants (4–6). The mutated *TP53* lose their function, leading to genome instability and promoting cancer development. While the majority of *TP53* variants are somatic, a part of the variants is germline that over 2000 *TP53* germline variants have been identified in human (<https://www.ncbi.nlm.nih.gov/clinvar/>, accessed 21 February 2022; <https://TP53.isb-cgc.org/>, accessed October 31, 2022). Many germline variants in *TP53* are PVs in causing different types of cancer (7–9) as represented by the

*To whom correspondence should be addressed. Tel: +853 8822 4836; Fax: +853 8822 2314; Email: sanmingwang@um.edu.mo

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Li-Fraumeni syndrome (LFS), a *TP53* germline PV-caused syndrome with various types of cancer developed at early age (10).

With the germline nature, knowledge for the evolutionary origin of *TP53* germline PVs will promote the deep understanding of the relationship between *TP53* germline variation and oncogenesis. The information will also help to identify specific targets for cancer prevention and treatment. Indeed, attempts have been made towards this direction (11–21). For example, it was observed that *TP53* was originated in unicellular eukaryotes (1,11,13,22); *TP53* is highly evolutionary conservation especially for the DNA-binding domain (1,12,18); *TP53* is under positive selection in humans (14,19,23,24). However, only a handful of *TP53* germline variants have been analyzed so far (Supplementary Table S1). Currently, there is no consensus for the evolutionary origin of *TP53* germline PVs in modern humans.

We hypothesized that there could be three potential origins for the *TP53* germline PVs in modern humans. One could be from the common ancestor of humans and other species through evolution conservation, the second one could be from human during its evolution process, and the third one could be the mixed origin of both cross-species conservation and human evolution. Taking the advantage of rich genome data from non-human species and ancient humans, we tested our hypothesis using the phylogenetic and archaeological approaches. Data from our study provided evidence to support the possibility that *TP53* germline PVs in modern humans were originated during recent human history and partially inherited from the extinct Neanderthals and Denisovans.

MATERIALS AND METHODS

Sources of human *TP53* variants

The current classification of genetic variants mainly follows the ACMG/AMG guidelines (25). The guidelines use a series of criteria evidence to interpret sequence variants in Mendelian disorders into five categories of ‘Pathogenic’, ‘Likely Pathogenic’, ‘Uncertain Significance’, ‘Likely Benign’, and ‘Benign’. In our study, the pathogenic variants (PVs) included both ‘Pathogenic’ and ‘Likely pathogenic’; the benign variants (BVs) included both ‘Benign’ and ‘Likely Benign’. A total of 1023 human *TP53* germline PVs and BVs from the ClinVar database were used for the study (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/clinvar.20220213.vcf.gz, accessed February 21, 2022). The annotation of *TP53* variants was based on the reference sequences: genome DNA version: hg38 NC.000017.11; cDNA: NM_000546.6; protein: NP_000537.3. The human genome position of *TP53* was chr17:7565097–7590856 (hg19 of Ensemble). The variants included single nucleotide variants and indel variants affecting less than 5 bases in order to maintain the accuracy of sequence alignment analysis. In our pilot test with indels of different sizes, we observe that the accuracy of sequence alignment with indels decreased following the increased size of indels. Therefore, we eliminated the indels with size >5 bases in our analysis, which accounted for 6.2% of the total of 2065 *TP53* germline variants under the five categories in ClinVar database. The variants with conflict classification,

variants of unknown significant (VUS) and unclassified variants were excluded from the study.

Alignment of vertebrate genome sequences

Human *TP53* variants were compared to the *TP53* genomic sequences of 99 vertebrate species in eight clades of Primate, Euarchontoglires, Laurasiatheria, Afrotheria, Mammal, Aves, Sarcopterygii and Fish through UCSC Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgGateway>), using the human *TP53* reference genomic sequence as control. The genome position of each *TP53* variant was entered into UCSC genome browser to obtain the sequences at the sites of different vertebrate species. Sequence alignment with 100 species was performed following the procedures in Li *et al.* (26). Briefly, ‘Multiz Alignments of 100 Vertebrates’ in the UCSC genome browser were used for the alignment (https://genome.ucsc.edu/cgi-bin/hgc?hgsid=1317652321_HIKV8m9ZVbi7yloKxK6oNsv2TTwa&db=hg38&c=chr17&l=7669691&r=7669692&o=7669691&t=7669692&g=multiz100way&i=multiz100way). Multiple sequence alignment was generated by Lastz (27) and Multiz (28) through comparing the genomic sequences of different species with human genome sequences. The alignments were based on the phylogenetic distance of reference genomes to adjust the scoring matrix and parameters for pairwise alignment and chaining (28,29). High-scoring chains were placed along with the genome and low-scoring chains were used to fill in the gaps. Alignment results were generated by the combined chains. The ‘Getbase’ program was used to collect the aligned data from ‘Multiz Alignments of 100 Vertebrates’ in the UCSG genome browser (<https://github.com/Skyllette14/GetBase>). The results of sequence alignment were divided into single-base and multiple-base alignments. The single-base alignment was based on the base position in other species, and the results were compared to the counterpart human variants; the multiple-base alignment used the same approach as single-base alignment but performed multiple times. The alignments were checked manually to verify the results. The phylogenetic tree model of 100 vertebrates, including the human, was based on the comparative genomics in the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1318537207_Wnfs9hCE2mtJjhZwL3lO1TMzxNNh&db=hg38&c=chr17&g=cons100way) and was generated by PhyloFit program (30) in PHAST package (<http://compugen.cshl.edu/phast/>). The alignment and statistical figures were generated by GraphPad Prism (version 8.0).

Ancient human genomic data analysis

Ancient human genome sequences were collected from ‘Allen Ancient DNA Resource’ (version 50.0, <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>, accessed 10 October 2021), and related publications through searching the keywords ‘ancient human genome’ in PubMed and Google scholar. A total of 5047 ancient individuals were included in the study, including

5031 human genome data dated between 45045 and 100 years before present (BP), 12 Neanderthal genome data dated between 120000 and 38515 years BP, and 4 Denisovan genome data dated between 158500 and 69650 years BP. Each set of sequence data was checked using mapDamage 2.0 (version 2.1.1) (31), a base recalibration tool to remove false variants generated by deamination in ancient DNA. SAMtool (32) was used to call the ancient human SNV and indel variants to generate the VCF file (<http://www.htslib.org/>). The called variants were annotated by using the ANNOVAR (33) program (<https://annovar.openbioinformatics.org/en/latest/>). The annotated variants were compared manually between human *TP53* PVs and BVs to obtain information from the related ancient carriers. The location and dated age of ancient human individuals were based on the information from 'Allen Ancient DNA Resource' and original publications. The distribution map for the ancient individuals was generated by MATLAB (version R2022a).

RESULTS

Phylogenetic analysis of human *TP53* germline PVs

We used the 1023 *TP53* germline variants including 406 PVs and 617 BVs from the ClinVar database for the phylogenetic analysis (Supplementary Table S2). We searched the PVs in the *TP53* sequences of 99 vertebrates distributed in eight clades of Primate, Euarchontoglires, Laurasiatheria, Afrotheria, Mammal, Aves, Sarcopharyngii, and Fish. We identified 70 (17.2%, 70 in 406) human PVs shared with 54 species (Supplementary Table S3A, B). We also used BVs as the control in the mapping analysis and identified 528 (85.6%, 528 in 617) human BVs shared with 99 vertebrates (Supplementary Table S3C, D). The shared variants had the following features:

PVs

- Low sharing rate. Although 70 human PVs (61 SNV and 9 indel variants) shared with other species, 48.6% (34 in 70) shared only with 1 species.
- Species sharing human PVs not following evolutionary tree. These species sharing human PVs were mostly distributed in the clades of Aves, Sarcopharyngii, and Fish distal to humans (Figure 1 and Supplementary Table S3A, 3B). For example, c.437G > A (p.Trp146Ter) was shared with 30 species from Zebrafish in Fish to Rock pigeon in Aves; c.1010G > A (p.Arg337His), a *TP53* founder variant in the Brazilian population, shared with 10 species in Euarchontoglires, Laurasiatheria, Aves, Sarcopharyngii and Fish. None of the shared PVs were in the species within Primate (Figure 1) except 2 PVs of c.537T > A (p.His179Gln) and c.455dup (p.Pro153fs), shared between humans and Gibbon with the divergent time of 33 million years (34).

BVs

- High sharing rate. Human BVs were continuously shared in all 99 vertebrate species, although the shared numbers

decreased when closer to the humans (Figure 2 and Supplementary Table S3C, D). c.408A > G (p.Gln136 =) had the highest sharing number of 91 species from Lamprey in Fish to Orangutan in Primate; c.493C > A (p.Gln165Lys) had the second highest sharing number of 87 species from Lamprey in Fish to Squirrel in Euarchontoglires. Squirrel in Euarchontoglires had the highest sharing number of 102 human BVs, and Tasmanian devil in Mammal had the second highest sharing number of 100 human BVs. Synonymous SNV was the major type of shared SNV BVs (49.6%, 257 in 518), and all shared indel BVs were intronic variants.

- High sharing with species in Primate. Multiple human BVs were shared highly with the species in Primate (Figure 2). For example, c.376 – 283T > C was shared from Squirrel monkey to Chimp, c.1101 – 221G > A was shared from Bushbaby to Chimp, and c.408A > G (p.Gln136 =) was shared from Bushbaby to Orangutan. Bushbaby in Primate had the highest sharing number of 71 human BVs.

Data from the phylogenetic analysis showed that human *TP53* germline PVs were lowly shared across vertebrate species. More importantly, the species sharing human PVs did not follow the evolutionary tree but were mostly distal to humans. Therefore, evolution conservation was unlikely the source of *TP53* germline PVs in modern humans. This is in contrast to the *TP53* germline BVs in modern humans, which were more shared with vertebrate species including those in Primate following evolutionary tree.

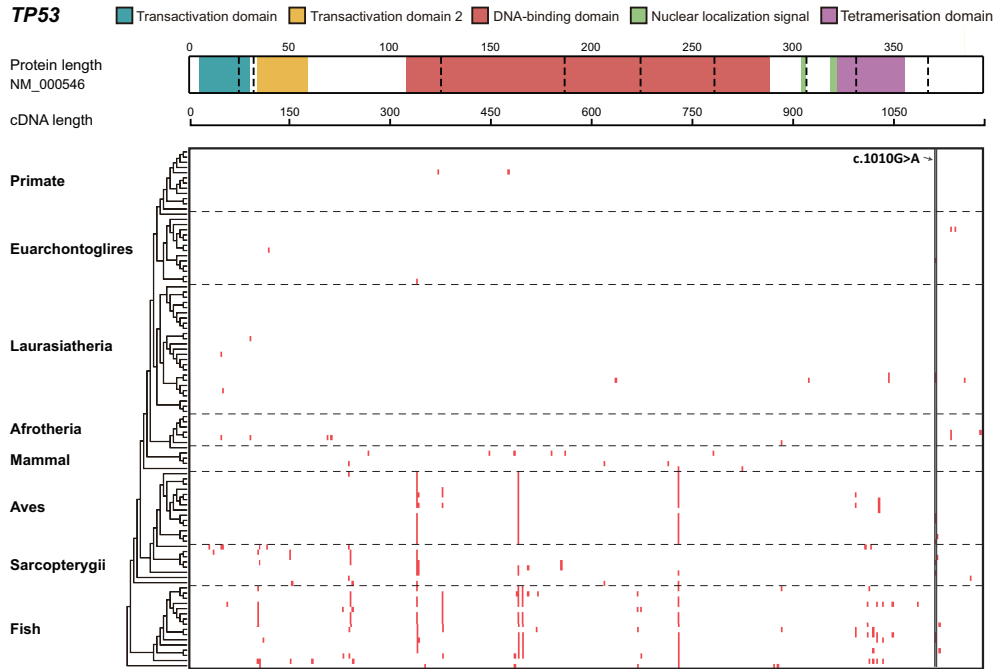
Archaeological analysis of human *TP53* germline variants in ancient humans

As phylogenetic analysis did not find evidence to support evolution conservation as the direct origin of human *TP53* germline PVs, we then performed an archaeological analysis to investigate whether PVs could be originated from humans themselves. We searched the 406 *TP53* germline PVs of modern humans in the genome sequences of 5031 ancient human cases, 12 Neanderthal cases and 4 Denisovan cases, and used the 617 *TP53* germline BVs of modern humans as the control. We identified 48 of 406 (11.8%) PVs shared with 62 ancient human cases, 3 Neanderthals and 1 Denisovan. We also identified 153 of 617 (24.8%) BVs shared with 2931 ancient human cases, 6 Neanderthals and 1 Denisovan. The shared variants had the following features:

PVs

- Short history of the shared PVs. Of the 62 ancient human individuals sharing 45 modern human PVs, 61 were dated between 7889 and 450 years BP (Figure 3) except c.734G > A (p.Gly245Asp), the oldest PV in a northeastern Mongolia carrier dated 34425 years BP. c.993G > A (p.Gln331 =) and c.993 + 1G > A were the youngest variants shared in an individual in Turkey dated 450 years BP (Table 1).
- High sharing rate. Eleven (24.4%) shared PVs were present in multiple ancient human individuals. For example, c.637C > T (p.Arg213Ter) had 6 ancient human

A



B

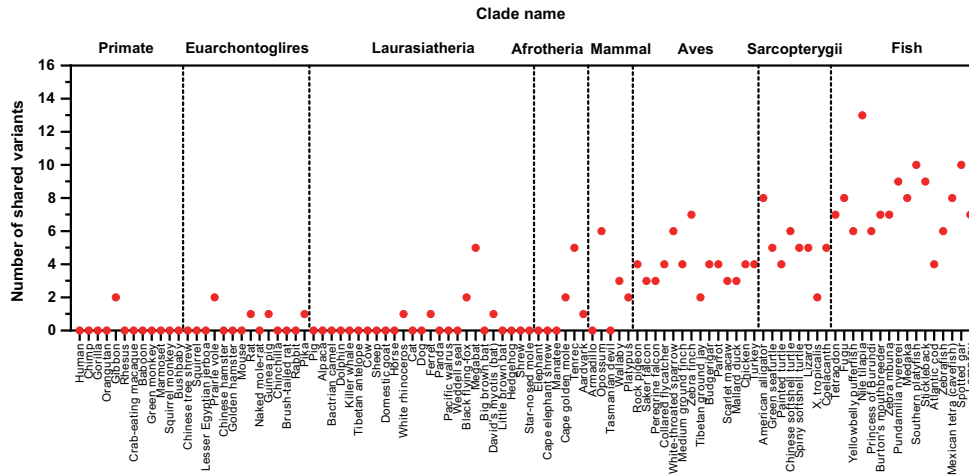
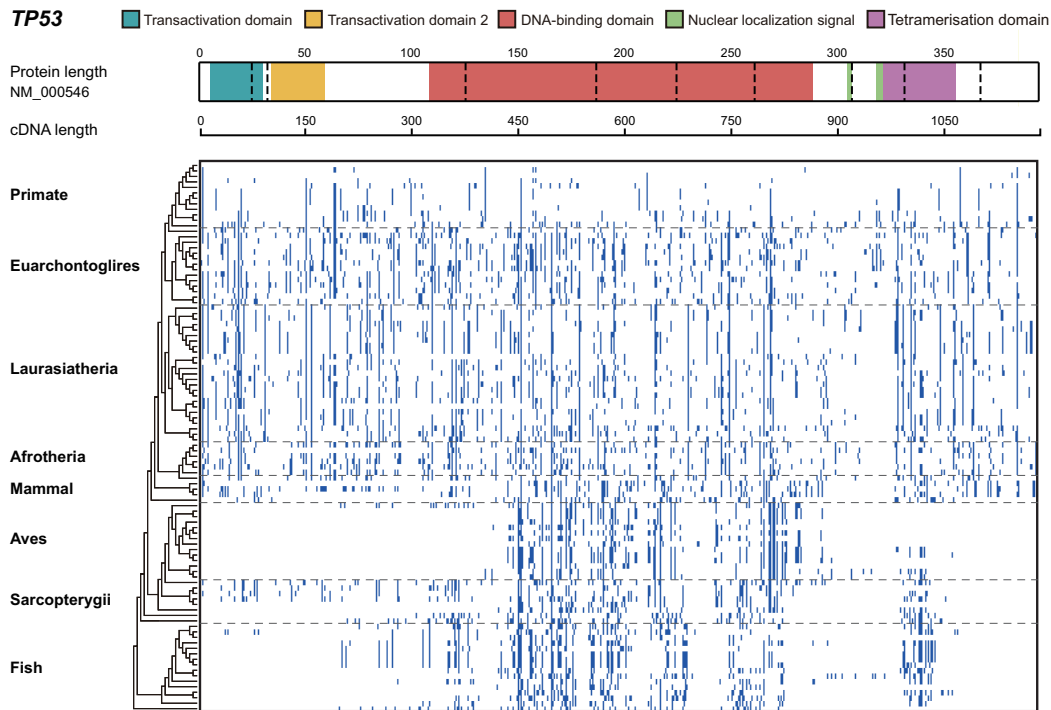


Figure 1. Cross-species analysis of *TP53* PVs. (A) Distribution of the shared human *TP53* PVs in 8 clades. It shows the number of human *TP53* PVs shared in eight clades: Primate, Euarchontoglires, Laurasiatheria, Afrotheria, Mammal, Aves, Sarcopterygii and Fish. The *TP53* coding region with functional domain information is at the top. Red cell: the PV shared between humans and other species. c.1010G > A: the *TP53* founder PV in the Brazilian population. (B) Distribution of the shared human *TP53* PVs in 99 vertebrates. X-axis from left to right: human and 99 vertebrates in 8 clades; Y-axis: number of human *TP53* PVs shared with non-human vertebrates.

- carriers dated 7144, 4050, 2600, 2255, 2160 and 850 years BP (Figure 3A and Table 1).
- Centered in the DNA-binding domain. Thirty of 45 (66.7%) shared PVs were in the DNA-binding domain (Figure 4A).
 - Hotspot PVs present in ancient humans. For the *TP53* hotspot PVs of c.733G > A (p.Gly245Ser) and c.743G > A (p.Arg248Gln), the oldest carriers for each PV were dated 3415 and 2200 years BP respectively, accounted for 4.4% (2 in 45) of the PVs shared in the

- ancient humans. c.733G > A (p.Gly245Ser) had 1 ancient carrier in Kazakhstan dated 3415 years BP, 1 ancient carrier in Ukraine dated 2550 years BP; c.743G > A (p.Arg248Gln) was present in a British carrier dated 2200 years BP.
- Present in Neanderthals and Denisovans. Four *TP53* PVs including c.473G > A (p.Arg158His), c.517G > A (p.Val173Met), hotspot c.524G > A (p.Arg175His), c.542G > A (p.Arg181His) were identified in 3 Neanderthals dated 44000, 42540 and 38515 years BP, and

A



B

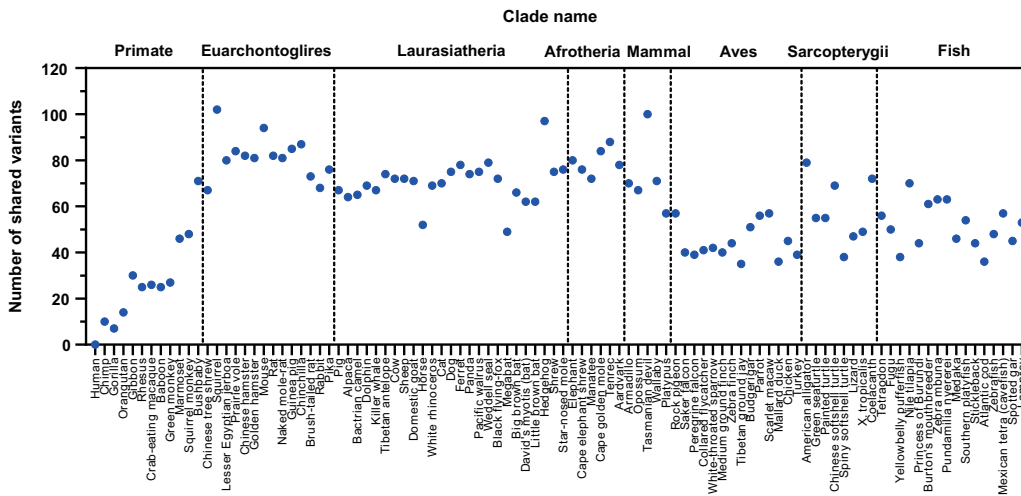


Figure 2. Cross-species analysis of *TP53* BVs. (A) Distribution of the shared human *TP53* BVs in 8 clades. It shows the number of human *TP53* BVs shared in eight clades: Primate, Euarchontoglires, Laurasiatheria, Afrotheria, Mammal, Aves, Sarcopterygii and Fish. Blue cell: human BVs shared with the species in different clades. (B) Distribution of the shared human *TP53* BVs in 99 vertebrates. X-axis from left to right: human and 99 vertebrates in 8 clades; Y-axis: number of human *TP53* BVs shared with non-human vertebrates.

2 PVs of c.574C > T (p.Gln192Ter) and c.661G > T (p.Glu221Ter) were identified in 1 Denisovan dated 158550 years BP (Table 1). All PVs shared with Neanderthals and Denisovans were located within the DNA-binding domain.

BVs

- High sharing rate. A total of 150 BVs were identified in 2931 ancient human individuals dated 45045 to 100

years BP, of which 74 (49.3%) BVs were shared with multiple individuals, 10 (6.7%) BVs were shared with more than 100 carriers (Supplementary Table S4). For example, c.376 – 91G > A and c.672 + 62A > G were present in 2422 and 1403 ancient humans. These two variants were also present in Neanderthals and Denisovans.

- Distributed across *TP53* coding region. Only 27 (18.0%) of the 150 shared BVs were within the DNA-binding domain (Figure 4B).

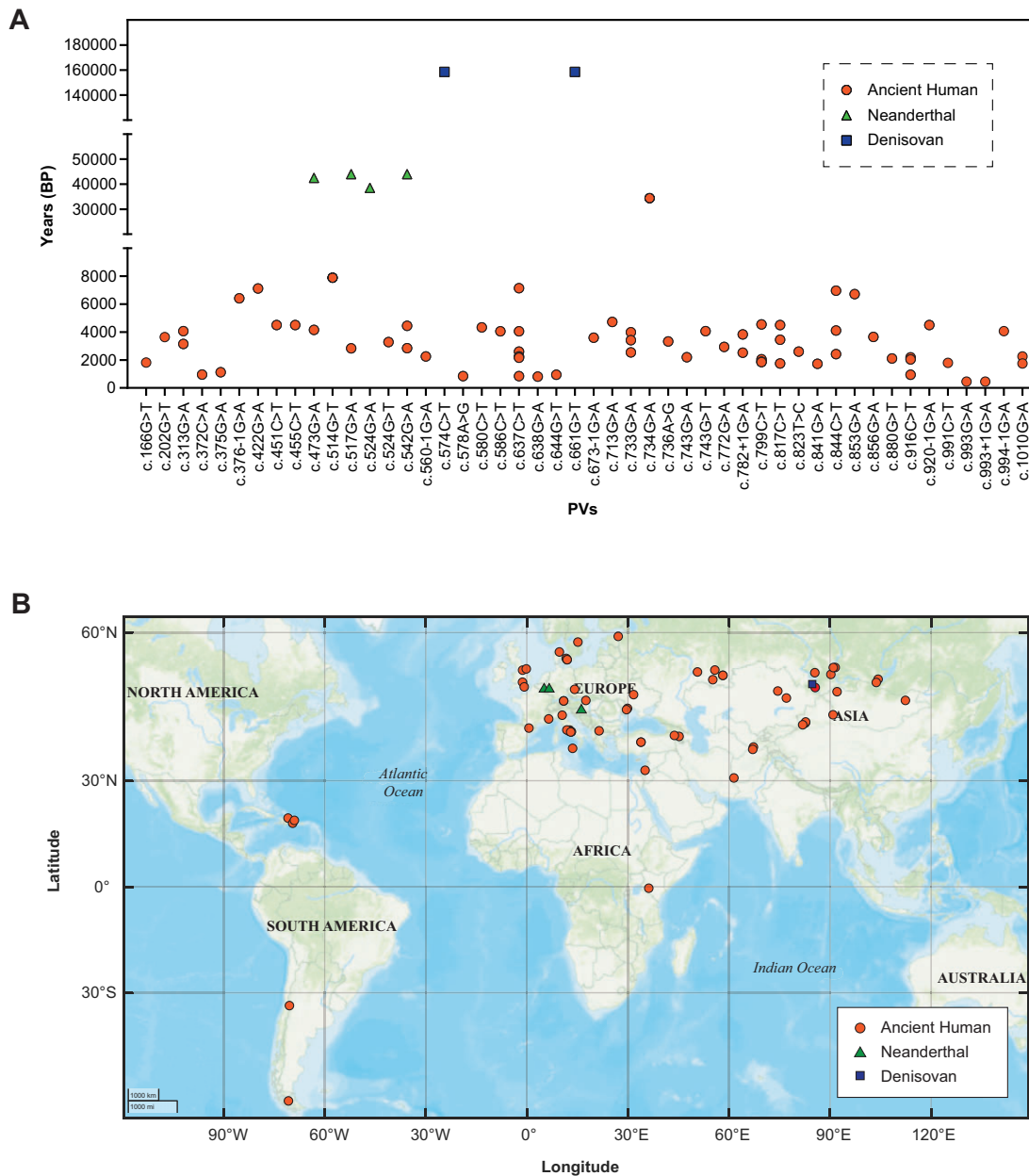


Figure 3. Temporal and geographic distribution of human *TP53* PVs shared in ancient individuals. (A) Temporal distribution of *TP53* PVs identified in ancient individuals. X-axis: PVs; Y-axis: Years before present (BP). (B) Geographic distribution of *TP53* PVs identified in ancient individuals. Orange dots: ancient human; green triangle: Neanderthals; blue rectangle: Denisovan.

- Highly prevalent in ancient humans. For example, codon 72 had 3 variants of c.215C > A (p.Pro72His), c.215C > G (p.Pro72Arg), c.216C > T (p.Pro72 =). c.215C > G (p.Pro72Arg) was present in 1113 ancient individuals, all were dated within 10000 years BP except 11 carriers dated 45045 to 10093 years BP. Further, only 6 of the 1113 c.215C > G (p.Pro72Arg) ancient carriers were in Africa.
- High prevalent *TP53* BVs were often not shared with other vertebrate species. For example, c.74 + 38C > G, c.215C > G (p.Pro72Arg), c.376 – 91G > A and c.376 – 125T > C had more than 100 ancient carriers but they

were not present in other vertebrate species (Supplementary Table S5).

- BVs arose earlier than PVs. The oldest BV carrier with five BVs of c.74 + 38C > G, c.215C > G (p.Pro72Arg), c.376 – 91G > A, c.376 – 283T > C, c.672 + 62A > G was in Omsk Oblast, Russia dated 45045 years BP; the second oldest carrier was in Tianyuan Cave, China dated 40000 years BP; the third oldest carrier was in Belgium dated 34795 years BP; the fourth oldest carrier was in northeastern Mongolia dated 34425 years BP. In comparison, c.734G > A (p.Gly245Asp) was the oldest PV dated 34425 years BP.

Table 1. TP53 PVs identified in ancient individuals

Year (BP) ^a	Variant			dbSNP155	Domain ^b	Carriers	Ref. ^c
	cDNA	Protein	Type				
Ancient Human							
34425	c.734G > A	p.Gly245Asp	Nonsynonymous SNV	rs121912656	DBD	2	26, 37
7889	c.514G > T	p.Val172Phe	Nonsynonymous SNV	rs1131691043	DBD	1	27
7144	c.637C > T	p.Arg213Ter	Stopgain	rs397516436	DBD	6	3, 8, 14, 22, 39, 49
7123	c.422G > A	p.Cys141Tyr	Nonsynonymous SNV	rs587781288	DBD	1	6
6960	c.844C > T	p.Arg282Trp	Nonsynonymous SNV	rs28934574	DBD	3	22, 38, 44
6713	c.853G > A	p.Glu285Lys	Nonsynonymous SNV	rs112431538	DBD	1	6
6415	c.376 – 1G > A	-	Splicing	rs868137297	-	1	41
4725	c.713G > A	p.Cys238Tyr	Nonsynonymous SNV	rs730882005	DBD	1	44
4545	c.799C > T	p.Arg267Trp	Nonsynonymous SNV	rs55832599	DBD	4	3, 20, 21, 60
4500	c.817C > T	p.Arg273Cys	Nonsynonymous SNV	rs121913343	DBD	3	3, 44
4500	c.451C > T	p.Pro151Ser	Nonsynonymous SNV	rs28934874	DBD	1	2
4500	c.455C > T	p.Pro152Leu	Nonsynonymous SNV	rs587782705	DBD	1	2
4500	c.920 – 1G > A	-	Splicing	rs587781702	-	1	2
4450	c.542G > A	p.Arg181His	Nonsynonymous SNV	rs397514495	DBD	2	1, 6
4330	c.580C > T	p.Leu194Phe	Nonsynonymous SNV	rs587780071	DBD	1	44
4160	c.473G > A	p.Arg158His	Nonsynonymous SNV	rs587782144	DBD	1	44
4071	c.313G > A	p.Gly105Ser	Nonsynonymous SNV	rs1060501195	-	2	2
4071	c.743G > T	p.Arg248Leu	Nonsynonymous SNV	rs11540652	DBD	1	2
4069	c.994 – 1G > A	-	Splicing	rs587782272	-	1	6
4050	c.586C > T	p.Arg196Ter	Stopgain	rs397516435	DBD	1	54
3826	c.782 + 1G > A	-	Splicing	rs1555525429	-	2	2, 49
3650	c.856G > A	p.Glu286Lys	Nonsynonymous SNV	rs786201059	DBD	1	20
3640	c.202G > T	p.Glu68Ter	Stopgain	rs869312782	-	1	51
3595	c.673 – 1G > A	-	Splicing	rs878854073	-	1	2
3415	c.733G > A	p.Gly245Ser	Nonsynonymous SNV	rs28934575	DBD	2	20, 44
3328	c.736A > G	p.Met246Val	Nonsynonymous SNV	rs483352695	DBD	1	2
3280	c.524G > T	p.Arg175Leu	Nonsynonymous SNV	rs28934578	DBD	1	44
2937	c.772G > A	p.Glu258Lys	Nonsynonymous SNV	rs121912652	DBD	1	61
2840	c.517G > A	p.Val173Met	Nonsynonymous SNV	rs876660754	DBD	1	20
2600	c.823T > C	p.Cys275Arg	Nonsynonymous SNV	rs1057519983	DBD	1	3
2254	c.560 – 1G > A	-	Splicing	rs1202793339	-	1	20
2250	c.1010G > A	p.Arg337His	Nonsynonymous SNV	rs121912664	Tet	2	3, 61
2200	c.743G > A	p.Arg248Gln	Nonsynonymous SNV	rs11540652	DBD	1	49
2181	c.916C > T	p.Arg306Ter	Stopgain	rs121913344	NLS	3	3, 22, 49
2113	c.880G > T	p.Glu294Ter	Stopgain	rs1057520607	-	1	49
1814	c.166G > T	p.Glu56Ter	Stopgain	rs1597375294	TAD	1	3
1790	c.991C > T	p.Gln331Ter	Stopgain	rs1597359130	Tet	1	20
1724	c.841G > A	p.Asp281Asn	Nonsynonymous SNV	rs764146326	DBD	1	20
1125	c.375G > A	p.Thr125=	Synonymous SNV	rs55863639	DBD	1	8
960	c.372C > A	p.Cys124Ter	Stopgain	rs1555526478	DBD	1	3
950	c.644G > T	p.Ser215Ile	Nonsynonymous SNV	rs587782177	DBD	1	50
850	c.578A > G	p.His193Arg	Nonsynonymous SNV	rs786201838	DBD	1	3
806	c.638G > A	p.Arg213Gln	Nonsynonymous SNV	rs587778720	DBD	1	8
450	c.993G > A	p.Gln331=	Synonymous SNV	rs11575996	Tet	1	6
450	c.993 + 1G > A	-	Splicing	rs11575997	-	1	6
Neanderthal							
44000	c.517G > A	p.Val173Met	Nonsynonymous SNV	rs876660754	DBD	1	67
44000	c.542G > A	p.Arg181His	Nonsynonymous SNV	rs397514495	DBD	1	67
42540	c.473G > A	p.Arg158His	Nonsynonymous SNV	rs587782144	DBD	1	67
38515	c.524G > A	p.Arg175His	Nonsynonymous SNV	rs28934578	DBD	1	67
Denisovan							
158550	c.574C > T	p.Gln192Ter	Stopgain	rs866380588	DBD	1	73
158550	c.661G > T	p.Glu221Ter	Stopgain	rs786201592	DBD	1	73

^aDated age of the oldest carrier.

^bDBD: DNA binding domain; Tet: tetramerization domain, NLS: nuclear localization signal; TAD: transactivation domain.

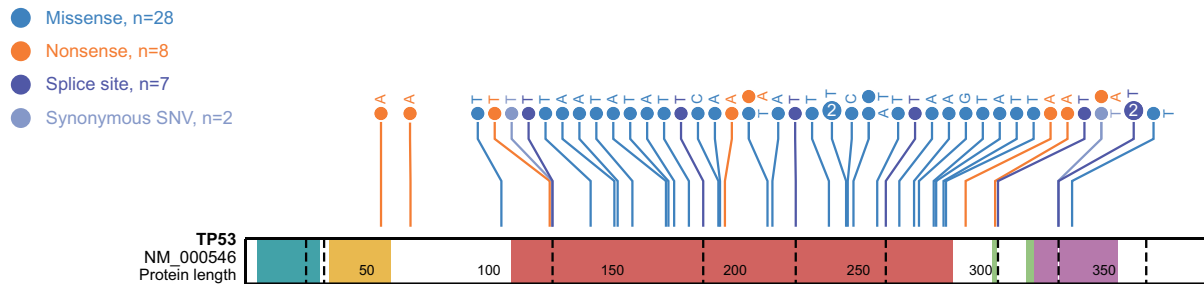
^cReferences were listed in Supplementary Table 6.

• Many BVs were present in Neanderthals and Denisovans. There were 12 BVs identified in 6 Neanderthals and 3 in 1 Denisovan, of which only 2 were in the DNA-binding domain (Supplementary Table S4). c.376 – 91G > A and c.672 + 62A > G were present in 2422 and 1403 ancient human individuals, they were also present in Neanderthals and Denisovans; c.215C > G (p.Pro72Arg) was present in 1113 an-

cient human individuals, it was also present in a Denisovan.

Data from the archaeological analysis demonstrated that human TP53 germline PVs were likely originated from humans themselves during the recent human evolution history, and partially inherited from the extinct Neanderthals and Denisovans.

A



B

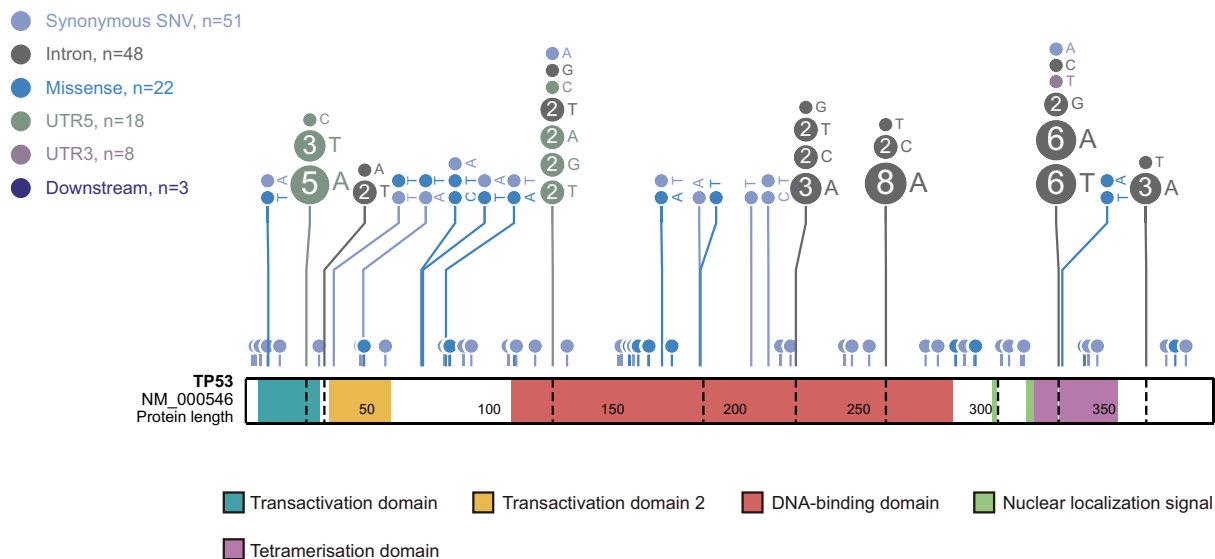


Figure 4. Distribution of the shared *TP53* PVs and BVs in different functional domains. (A) Domain distribution of 45 PVs shared between modern and ancient humans. (B) Domain distribution of 150 BVs shared between modern and ancient humans. The figure shows that the shared PVs were centered in the DNA-binding domain, whereas the shared BVs were distributed across the entire coding region.

DISCUSSION

By using phylogenetic and archaeological approaches, our study reveals the evolutionary origins of *TP53* germline PVs in modern humans that they were likely originated during recent human history within the last 8000 years. In contrast, the human *TP53* germline BVs were more originated from both cross-species conservation and human evolution. The key driver force for the difference between PVs and BVs lays on the evolution selection. *TP53* PVs are deleterious, they are under negative selection to be suppressed in the population; whereas *TP53* BVs can be beneficial for survival and reproduction, therefore, they can be selected and conserved across species. Consistent with the observation that modern humans inherited a part of genomic materials from the extinct Neanderthals and Denisovans, we also observed that a part of *TP53* germline PVs and BVs in modern humans was present in Neanderthals and Denisovans (35–37).

Our phylogenetic analysis showed that human PVs were highly shared in distal species but not in closer species. What could be the explanation for this sharing pattern? A study observed the sharing of multiple human deleterious variants in mice (38). Upon evaluating multiple related theories including ‘Founder effect’, ‘Fixations of slightly deleterious mutations’, ‘Relaxed selection on late-onset phenotypes’ and ‘Compensatory changes’, the authors considered that the ‘Compensation theory’ could better explain the phenomenon, which states that ‘compensatory mutations at other sites of the same or a different protein render the deleterious mutations neutral’, implying that the PVs in humans may not be pathogenic in the distal species. While the compensation theory may also be used to explain why human *TP53* PVs were shared in distal species, we consider it may not explain well for the high-degree *TP53* PVs shared between human and distal species, such as the species in Fish. We consider an alternative explanation that the *TP53*

PVs shared between human and the distal species may reflect a possibility that *TP53* was highly mutated in Fish. During the evolution process from Fish to Amphibian to Mammal to human, many of the PVs were eliminated by evolution selection due to their deleterious effects. The PVs shared between human and distal species were the survived PVs after the selection.

Of the 62 *TP53* PVs carriers in ancient humans, 61 were in Europe and Asia but only one was in Africa. This confirms that human *TP53* variants were arisen after the latest 'out of Africa' migration (22,39), and is consistent with the observation that *TP53* variation in the modern African population has the lowest prevalence of 0.07% than the prevalence in non-African populations (0.12% in South Asian, 0.15% in East Asian, 0.16% in Latin American, and 0.24–0.28% in European) (40). The positive selection of human *TP53* explains why *TP53* germline PVs are at high prevalence in modern humans (22).

Genetic contribution from extinct hominins to modern humans has been increasingly studied. While most studies focused on the biological function involving basic physiology like immunity, high altitude adaptation, hair color, metabolism, etc. (41–47), there have been rare information for the medical relevant issues such as cancer genetic predisposition as shown by our data.

c.1010G > A (p.Arg337His) is a *TP53* founder variant in the Southern Brazilian population for LFS (48), but it has limited penetrance due to modifications by other genetic variants or metabolic cofactors. It is located at a CpG motif in *TP53*, a region highly mutable due to high spontaneous deamination (36). Haplotype analysis indicated that it was originated from Caucasian/Portuguese-Iberic origin 2000 years BP (49). Our study identified two carriers, one in Mongolia dated 2250 years BP and another in Italy dated 1750 years BP, suggesting that c.1010G > A (p.Arg337His) could have occurred in more ancient ethnic human populations either through genetic connection or occurred spontaneously by chance (50).

What can be the explanation for the high prevalence of *TP53* germline PVs in modern humans, regardless of their deleterious effects and selection pressure? *TP53* germline PVs can cause LFS, with cancers development at young age before reaching reproduction age. Therefore, *TP53* germline PVs would be expected to be rare in human population. However, of the spectrum of human *TP53* germline PVs, not all cause LFS. Our analysis showed that of the 406 *TP53* PVs included in our study, only 95 (23.4%) were solely derived from LFS. The selection pressure on these PVs not contributing to LFS may not be negatively selected as heavy as those causing LFS. Another possibility can be that certain *TP53* PVs could be beneficial rather than deleterious. This is a situation similar to *BRCA1* PVs, in which the PVs can increase reproduction (51), enhance pathogen immunity (52), promote neural development (53), and gene expression (54) at the cost of cancer development at the post-reproduction age in the PV carriers. The cancer is lethal for the PV carriers but it doesn't affect the population propagation.

Multiple *TP53* hotspot PVs including c.524G > A (p.Arg175His), c.733G > A (p.Gly245Ser) are highly present in modern humans (3,4). The hotspot variants

were often C to T transition that occurs more often than other types of base substitution. This may explain why the c.524G > A (p.Arg175His) was the oldest hotspot variant in Neanderthals dated 38515 years BP; c.733G > A (p.Gly245Ser) were present in ancient humans with 2 carriers dated 3415 to 2550 years BP. It may be possible that their high prevalence in ancient humans may simply be the consequence of their high probability of occurrence coupled with their poor antigenic properties. This can also be the case for the Brazilian founder c.1010G > A (p.Arg337His).

Codon 72 was highly variable in ancient humans, with three *TP53* BVs of c.215C > A (p.Pro72His), c.215C > G (p.Pro72Arg), c.216C > T (p.Pro72 =), with c.215C > G (p.Pro72Arg) as the most frequent one (Supplementary Table S4). Allele frequency of c.215C > G (p.Pro72Arg) in the African modern population (0.2 to 0.4) is lower than in non-African modern populations (0.5 to 0.8) (24,55). None of the codon 72 variants were PVs, suggesting that codon 72 variants could be beneficial. For example, c.215C > G (p.Pro72Arg) enhances human adaptation to lower temperature (23) by increasing leukemia inhibitory factor (LIF) expression to protect the implantation of fertilized eggs (16,22,56), and enhances innate immunity (57). Its presence in Denisovan suggests that c.215C > G (p.Pro72Arg) could be inherited from Denisovans to ancient and modern humans. Therefore, c.215C > G (p.Pro72Arg) can be a beneficial variant.

It is interesting to note that the pattern of the evolutionary origin of human *TP53* germline PVs is similar to that of human *BRCA* PVs, which were also originated from ancient humans but not through cross-species evolution conservation (26). The coincidence suggests the possibility that PVs in many if not all tumor suppressor genes in modern humans could arise during recent human evolution process.

Our study used the *TP53* germline variants from the ClinVar database. The *TP53* database (formerly the International Agency for Research on Cancer, IARC) also provides *TP53* germline variant data (<https://TP53.isb-cgc.org/>, accessed 31 October 2022). ClinVar contained 2065 and the *TP53* database contained 552 germline variants. Of the 552 variants, 336 (60.9%) were included in ClinVar. Therefore, we used the ClinVar data in our study. In the archeological analysis, our study identified only 48 (11.8%) of the 406 PVs used in the study. The lower number is likely related to the smaller size of the ancient individuals although nearly all ancient samples currently available were included in our study. When more ancient samples are available, the number of shared *TP53* PVs should likely increase.

It is important to indicate that the presence of a variant in an ancient carrier does not necessarily imply that the variant has been fixed in the human lineage for that length of time, as the variant at the same position could occur and eliminated repetitively over generations, with each appearance as an independent event. This is particular for the hotspot variants in *TP53*, which often have high tendency of spontaneous occurrence (4,9). While the haplotype test will provide a definitive answer, we were unable to provide the haplotype evidence to definitely prove that every *TP53* PVs in ancient humans were the origins of the PVs in modern humans. This is due largely to the limitation of ancient human genome data that many ancient human genome

sequences had poor quality, gaps, and lower genome coverage. It is technically difficult to identify reliable *TP53* haplotypes shared between modern and ancient human genomes. While it may specifically affect the interpretation for the hotspot *TP53* PVs as they have high tendency of spontaneous occurrence, only 2, c.733G > A (p.Gly245Ser) and c.743G > A (p.Arg248Gln), of the 45 ancient PVs identified in our study were the classical hotspot variants.

In summary, our study reveals that *TP53* germline PVs were likely originated during recent human history. As *TP53* germline PVs are strongly correlated with cancer risk (58), elucidating their evolution origin should deepen our understanding of the relationship between *TP53* variation and cancer.

DATA AVAILABILITY

The vertebrate genome data were collected from public resources and the details were shown in the Methods and Materials. The ancient human genome data were collected from public resources, and the related references were listed in Supplementary Table S6. Public programs and software used in this study were described in detail in the Methods and Materials. Any additional information for the data reported in this paper is available to request from the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

ACKNOWLEDGEMENTS

We are thankful to the Information and Communication Technology Office, University of Macau for providing the High-Performance Computing Cluster resource and facilities for the study. We thank Mrs. Stephanie Rachell Andaluz Almandoz Ramos for the English editing of the manuscript. *Author contributions:* S.H.K. was responsible for data analysis, methodology, visualization and contributed to manuscript writing. J.L. participated in methodology, software development, data analysis and manuscript revision. B.T., H.L., B.Z. and F.X. contributed to data analysis and manuscript revision. S.M.W. was responsible for conceptualization, supervision, methodology, manuscript revision and funding acquisition.

FUNDING

Macau Science and Technology Development Fund [085/2017/A2, 0077/2019/AMJ, 0032/2022/A1]; University of Macau [SRG2017-00097-FHS, MYRG2019-00018-FHS, 2020-00094-FHS]; Faculty of Health Sciences, University of Macau [FHSIG/SW/0007/2020P, MOE Frontiers Science Center for Precision Oncology pilot grants, and a startup fund to S.M.W.]; B.T. is the recipient of the University of Macau Postdoctoral Fellowship Class A of the Macao Talent Program and FDCT postdoctorate fellowship [UMMTP-FDCT/0027/APD/2021].

Conflict of interest statement. None declared.

REFERENCES

- Levine, A.J. (2020) p53: 800 million years of evolution and 40 years of discovery. *Nat. Rev. Cancer*, **20**, 471–480.
- Levine, A.J. and Oren, M. (2009) The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer*, **9**, 749–758.
- Freed-Pastor, W.A. and Prives, C. (2012) Mutant p53: one name, many proteins. *Genes Dev.*, **26**, 1268–1286.
- Baugh, E.H., Ke, H., Levine, A.J., Bonneau, R.A. and Chan, C.S. (2018) Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ.*, **25**, 154–160.
- Monti, P., Menichini, P., Speciale, A., Cutrona, G., Fais, F., Taiana, E., Neri, A., Bomben, R., Gentile, M., Gattei, V. *et al.* (2020) Heterogeneity of TP53 mutations and P53 protein residual function in cancer: does it matter? *Front. Oncol.*, **10**, 593383.
- Xu, J., Qian, J., Hu, Y., Wang, J., Zhou, X., Chen, H. and Fang, J.Y. (2014) Heterogeneity of Li-Fraumeni syndrome links to unequal gain-of-function effects of p53 mutations. *Sci. Rep.*, **4**, 4223.
- Guha, T. and Malkin, D. (2017) Inherited TP53 mutations and the Li-Fraumeni syndrome. *Cold Spring Harb. Perspect. Med.*, **7**, a026187.
- Lu, W.J., Amatruda, J.F. and Abrams, J.M. (2009) p53 ancestry: gazing through an evolutionary lens. *Nat. Rev. Cancer*, **9**, 758–762.
- Olivier, M., Hollstein, M. and Hainaut, P. (2010) TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.*, **2**, a001008.
- Li, F.P., Fraumeni, J.F. Jr, Mulvihill, J.J., Blattner, W.A., Dreyfus, M.G., Tucker, M.A. and Miller, R.W. (1988) A cancer family syndrome in twenty-four kindreds. *Cancer Res.*, **48**, 5358–5362.
- Bartas, M., Brazda, V., Cerven, J. and Pecinka, P. (2019) Characterization of p53 Family homologs in evolutionary remote branches of Holozoa. *Int. J. Mol. Sci.*, **21**, 6.
- Belyi, V.A. and Levine, A.J. (2009) One billion years of p53/p63/p73 evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 17609–17610.
- Biscotti, M.A., Barucca, M., Carducci, F., Forconi, M. and Canapa, A. (2019) The p53 gene family in vertebrates: evolutionary considerations. *J. Exp. Zool. B Mol. Dev. Evol.*, **332**, 171–178.
- Emam, M., Machado, J.P. and Antunes, A. (2020) Evolutionary genomics of mammalian lung cancer genes reveals signatures of positive selection in APC, RB1 and TP53. *Genomics*, **112**, 4722–4731.
- Glazko, G.V., Koonin, E.V. and Rogozin, I.B. (2004) Mutation hotspots in the p53 gene in tumors of different origin: correlation with evolutionary conservation and signs of positive selection. *Biochim. Biophys. Acta*, **1679**, 95–106.
- Hu, W., Feng, Z., Atwal, G.S. and Levine, A.J. (2008) p53: a new player in reproduction. *Cell Cycle*, **7**, 848–852.
- Jegga, A.G., Inga, A., Menendez, D., Aronow, B.J. and Resnick, M.A. (2008) Functional evolution of the p53 regulatory network through its target response elements. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 944–949.
- Lane, D.P., Cheok, C.F., Brown, C., Madhumalar, A., Ghadessy, F.J. and Verma, C. (2010) Mdm2 and p53 are highly conserved from placozoans to man. *Cell Cycle*, **9**, 540–547.
- Passow, C.N., Bronikowski, A.M., Blackmon, H., Parsai, S., Schwartz, T.S. and McGaugh, S.E. (2019) Contrasting patterns of rapid molecular evolution within the p53 network across mammal and sauropsid lineages. *Genome Biol Evol.*, **11**, 629–643.
- Qiang Huang, L.Y., Levine, A.J., Nussinov, R. and Ma, B. (2014) Dipeptide analysis of p53 mutations and evolution of p53 family proteins. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics*, **1844**, 198–206.
- Zhang, Q., Balourdas, D.I., Baron, B., Senitzki, A., Haran, T.E., Wiman, K.G., Soussi, T. and Joerger, A.C. (2022) Evolutionary history of the p53 family DNA-binding domain: insights from an Alvinella pompejana homolog. *Cell Death. Dis.*, **13**, 214.
- Belyi, V.A., Ak, P., Markert, E., Wang, H., Hu, W., Puzio-Kuter, A. and Levine, A.J. (2010) The origins and evolution of the p53 family of genes. *Cold Spring Harb. Perspect. Biol.*, **2**, a001198.
- Shi, H., Tan, S.J., Zhong, H., Hu, W., Levine, A., Xiao, C.J., Peng, Y., Qi, X.B., Shou, W.H., Ma, R.L. *et al.* (2009) Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. *Am. J. Hum. Genet.*, **84**, 534–541.
- Sucheston, L., Witonsky, D.B., Hastings, D., Yildiz, O., Clark, V.J., Di Rienzo, A. and Onel, K. (2011) Natural selection and functional genetic variation in the p53 pathway. *Hum. Mol. Genet.*, **20**, 1502–1508.

25. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405–424.
26. Li, J., Zhao, B., Huang, T., Qin, Z. and Wang, S.M. (2022) Human BRCA pathogenic variants were originated during recent human history. *Life Sci. Alliance*, **5**, e202101263.
27. Chiaromonte, F., Yap, V.B. and Miller, W. (2002) Scoring pairwise genomic sequence alignments. *Pac. Symp. Biocomput.*, **7**, 115–126.
28. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D. *et al.* (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.*, **14**, 708–715.
29. Armstrong, J., Fiddes, I.T., Diekhans, M. and Paten, B. (2019) Whole-genome alignment and comparative annotation. *Annu. Rev. Anim. Biosci.*, **7**, 41–64.
30. Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W. *et al.* (2001) Resolution of the early placental mammal radiation using bayesian phylogenetics. *Science*, **294**, 2348–2351.
31. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P.L. and Orlando, L. (2013) mapDamage2.0: fast approximate bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, **29**, 1682–1684.
32. 1000 Genome Project Data Processing Subgroup, Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and samtools. *Bioinformatics*, **25**, 2078–2079.
33. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
34. Glazko, G.V. and Nei, M. (2003) Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.*, **20**, 424–434.
35. Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C. *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.
36. Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M.R., Pugach, I., Ko, A.M., Ko, Y.C., Jinam, T.A., Phipps, M.E. *et al.* (2011) Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.*, **89**, 516–528.
37. Hajdinjak, M., Fu, Q., Hubner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I. *et al.* (2018) Reconstructing the genetic history of late Neanderthals. *Nature*, **555**, 652–656.
38. Gao, L. and Zhang, J. (2003) Why are some human disease-associated mutations fixed in mice? *Trends Genet.*, **19**, 678–681.
39. Gomez, F., Hirbo, J. and Tishkoff, S.A. (2014) Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb. Perspect. Biol.*, **6**, a008524.
40. de Andrade, K.C., Mirabello, L., Stewart, D.R., Karlins, E., Koster, R., Wang, M., Gapstur, S.M., Gaudet, M.M., Freedman, N.D., Landi, M.T. *et al.* (2017) Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. *Hum. Mutat.*, **38**, 1723–1730.
41. Abi-Rached, L., Jobin, M.J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F.A. *et al.* (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*, **334**, 89–94.
42. Mendez, F.L., Watkins, J.C. and Hammer, M.F. (2013) Neanderthal origin of genetic variation at the cluster of OAS immunity genes. *Mol. Biol. Evol.*, **30**, 798–801.
43. Huerta-Sanchez, E., Jin, X., Asan, B., Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M.Z., Somel, M. *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, **512**, 194–197.
44. Ding, Q.L., Hu, Y., Xu, S.H., Wang, J.C. and Jin, L. (2014) Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. *Mol. Biol. Evol.*, **31**, 683–695.
45. Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N. and Reich, D. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507**, 354–357.
46. Khrameeva, E.E., Bozek, K., He, L., Yan, Z., Jiang, X., Wei, Y.N., Tang, K., Gelfand, M.S., Prufer, K., Kelso, J. *et al.* (2014) Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat. Commun.*, **5**, 3584.
47. Vernot, B. and Akey, J.M. (2014) Resurrecting surviving neandertal lineages from modern human genomes. *Science*, **343**, 1017–1021.
48. Giacomazzi, J., Graudenz, M.S., Osorio, C.A., Koehler-Santos, P., Palmero, E.I., Zagonel-Oliveira, M., Michelli, R.A., Scapulatempo Neto, C., Fernandes, G.C., Achatz, M.I. *et al.* (2014) Prevalence of the TP53 p.R337H mutation in breast cancer patients in Brazil. *PLoS One*, **9**, e99893.
49. Paskulin, D.D., Giacomazzi, J., Achatz, M.I., Costa, S., Reis, R.M., Hainaut, P., Santos, S.E. and Ashton-Prolla, P. (2015) Ancestry of the Brazilian TP53 c.1010G>A (p.Arg337His, R337H) founder mutation: clues from haplotyping of short tandem repeats on chromosome 17p. *PLoS One*, **10**, e0143262.
50. Bar-Sade, R.B., Kruglikova, A., Modan, B., Gak, E., Hirsh-Yechezkel, G., Theodor, L., Novikov, I., Gershoni-Baruch, R., Risel, S., Papa, M.Z. *et al.* (1998) The 185delAG BRCA1 mutation originated before the dispersion of Jews in the diaspora and is not limited to Ashkenazim. *Hum. Mol. Genet.*, **7**, 801–805.
51. Smith, K.R., Hanson, H.A. and Hollingshaus, M.S. (2013) BRCA1 and BRCA2 mutations and female fertility. *Curr. Opin. Obstet. Gynecol.*, **25**, 207–213.
52. Lou, D.I., McBee, R.M., Le, U.Q., Stone, A.C., Wilkerson, G.K., Demogins, A.M. and Sawyer, S.L. (2014) Rapid evolution of BRCA1 and BRCA2 in humans and other primates. *BMC Evol. Biol.*, **14**, 155.
53. Pao, G.M., Zhu, Q., Perez-Garcia, C.G., Chou, S.J., Suh, H., Gage, F.H., O'Leary, D.D. and Verma, I.M. (2014) Role of BRCA1 in brain development. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E1240–E1248.
54. Rosen, E.M., Fan, S. and Ma, Y. (2006) BRCA1 regulation of transcription. *Cancer Lett.*, **236**, 175–185.
55. Doffe, F., Carbonnier, V., Tissier, M., Leroy, B., Martins, I., Mattsson, J.S.M., Micke, P., Pavlova, S., Pospisilova, S., Smardova, J. *et al.* (2021) Identification and functional characterization of new missense snps in the coding region of the TP53 gene. *Cell Death Differ.*, **28**, 1477–1492.
56. Hu, W. and Feng, Z. (2019) The role of p53 in reproduction, an unexpected function for a tumor suppressor. *J. Mol. Cell Biol.*, **11**, 624–627.
57. Lodhi, N., Singh, R., Rajput, S.P. and Saquib, Q. (2021) SARS-CoV-2: understanding the transcriptional regulation of ACE2 and TMPRSS2 and the role of single nucleotide polymorphism (SNP) at codon 72 of p53 in the Innate Immune response against virus infection. *Int. J. Mol. Sci.*, **22**, 8660.
58. Kato, S., Han, S.Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. and Ishioka, C. (2003) Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8424–8429.