Recent selection is a major force driving cancer evolution

Langyu Gu^{1,*}, Guofen Yang^{2,*}

¹ State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, Guangdong, China

² Department of Gynecology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510060, Guangdong, China

*Correspondance:

Langyu Gu: guly6@mail.sysu.edu.cn

Guofen Yang: yangguof@mail.sysu.edu.cn

Abstract

Cancer is one of the most threatening diseases to humans. Understanding the evolution of cancer genes is helpful for therapy management. However, systematic investigation of the evolution of cancer driver genes is sparse. Using comparative genomic analysis, population genetics analysis and computational molecular evolutionary analysis, we detected the evolution of 568 cancer driver genes of 66 cancer types across the primate phylogeny (long timescale selection), and in modern human populations from the 1000 human genomics project (recent selection). We found that recent selection pressures, rather than long timescale selection, significantly affect the evolution of cancer driver genes in humans. Cancer driver genes related to morphological traits and local adaptation are under positive selection in different human populations. The African population showed the largest extent of divergence compared to other populations. It is worth noting that the corresponding cancer types of positively selected genes exhibited population-specific patterns, with the South Asian population possessing the least numbers of cancer types. This helps explain why the South Asian population usually has low cancer incidence rates. Population-specific patterns of cancer types whose driver genes are under positive selection also give clues to explain discrepancies of cancer incidence rates in different geographical populations, such as the high incidence rate of Wilms tumour in the African population and of Ewing's sarcomas in the European population. Our findings are thus helpful for understanding cancer evolution and providing guidance for further precision medicine.

Keywords:

cancer driver genes, positive selection, primates, modern human populations

Introduction

Cancer is one of the most threatening diseases with a high mortality rate in humans. Somatic mutations of cancer genes can improve the fitness of tumor cells by promoting their survival, growth, and proliferation and are often under positive selection at the cellular level (Martincorena et al. 2017). Unlike somatic mutations, germline mutations of cancer genes at the individual level are usually unwelcome since mutations can be deleterious to the organism and thus affect the fitness (Fischer et al. 2011; Shendure and Akey 2015). With this logic, germline mutations of cancer genes should be dominant by purifying selection. Surprisingly, studies continuously found that cancer genes are under positive selection by comparative genomics analysis (Kang and Michalak 2015; Vicens and Posada 2018), indicating their advantageous roles during evolution. Several hypotheses have been proposed to explain the evolutionary trade-off between positive selection and cancer risks, such as sexual selection, pathogen-host interactions, and genomic compensation (Nielsen et al. 2005; Crespi and Summers 2006; Vicens and Posada 2018). However, few studies detected positive selection considering variation both among sites and branches across phylogeny. The extent to which positive selection is specific to the human lineage remains unknown.

Positive selection detection across phylogeny usually detects selection on a large time scale, i.e., usually million years (long timescale selection) (Schaschl and Wallner 2020). In contrast, detection at the population level can reflect recent selection in modern human populations. Neolithic demographic transition and the migration out of Africa have led modern humans to experience significant changes in lifestyles and living environments over the last 100,000 years (Stewart and Stringer 2012; Scerri et al. 2019). Diversified living environments and food resources accompanied by new infectious diseases brought new selective pressures

to modern human populations (Benton et al. 2021). How humans interact and adapt to changing environments continues to attract attention. Thus, the last 100,000 years are one of the most interesting time scales in human history (Sabeti et al. 2002; Sabeti et al. 2007; Schaschl and Wallner 2020). Interestingly, many human disease-related genes have been reported under positive selection due to their pleiotropic adaptive functions, such as BRCA1, SPANX, and oncogenetic viruses (Burrows et al. 2004; Kouprina et al. 2004; Lou et al. 2014). Whether genes that cause cancer, the common disease in modern humans, also show a similar signature remains unknown.

Although the importance of the evolution of cancer genes began to attract attention, systematic investigation of these topics in a broader view is sparse. Existing studies have mainly focused on a few cancer genes of specific cancer types or limited species across phylogeny (Huttley et al. 2000; Wildman et al. 2003; Nielsen et al. 2005; Lou et al. 2014). Besides, studies usually did not differentiate cancer driver genes and passenger genes, which could be under different selection pressures (Martincorena et al. 2017). An in-depth comparative analyses of cancer genes of different cancer types across phylogeny will give us a comprehensive picture of cancer gene evolution.

To obtain a deep understanding of the evolutionary of cancer, we detected the evolution of 568 cancer driver genes of 66 cancer types across the primate phylogeny (long timescale selection), as well as in modern human populations retrieved from the 1000 human genomics project (recent selection). We aim to answer the following question: does long timescale selection or recent selection play an important role in the evolution of cancer driver genes in humans? Our findings will have profound implications for understanding the evolution of cancer driver genes and provide clues for further precision medicine.

Results

Postive selection detection across the primate phylogeny

Transcript IDs of 568 cancer driver genes across 66 cancer types were retrieved from the literature published in Nature Reviews Cancer recently (Martínez-Jiménez et al. 2020). To understand the evolutionary dynamics of 568 cancer driver genes across the primate phylogeny, we obtained orthologous sequences in 24 primates using Oma inference. We further detected positive selection in different lineages under the branch-site model using PAML. Branches showing unusually large ω values ($\omega > 100$) could be due to positive selection but could also be because sequences are too divergent so that dS is saturated. We thus removed 92 genes with foreground branches showing unusually large ω values for downstream PAML analysis. We obtained two genes that are detected under positive selection in the human lineage. ZNF626 (ENST00000601440) belongs to Zinc finger protein coding genes with significant roles in regulatory evolution and local adaptation (Perdomo-Sabogal and Nowick 2019; Jovanovic et al. 2021). The RGPD3 gene (ENST00000409886) is related to human skull shape and face morphology determination (Wu et al. 2019). These two genes are both involved in limited cancer types, i.e., The RGPD3 gene is involved in cutaneous melanoma of the skin (CM) and the thyroid adenocarcinoma (THCA), and ZNF626 gene is involved in the nasopharyngeal cancer NSPH (Table 1).

Table 1 Genes under positive selection in the human lineage

Ensembl	gene name	main function	cancer type	branch-site model			
				model A	model B	2ln(A-B)	<i>W</i>
ENST00000601440	ZNF626	regulatory evolution	NSPH	-6898.8	-6906.9	16.32	8.67
ENST00000409886	RGPD3	skull shape and face morphology	CM, THCA	-12854	-12881	53.66	70.6

Statistical significance was detected after Bonferroni's correction.

NSPH, nasopharyngeal cancer; CM, cutaneous melanoma of the skin; THCA, thyroid adenocarcinoma.

In contrast, many genes are under positive selection in ancestral lineages across the primate phylogeny (Fig. 1, Supplementary File 1).

Figure 1 Genes under positive selection using a branch-site model in PAML across the primate phylogeny. Target foreground branches were labeled. The number of genes under positive selection was given in the target foreground branch in red. The species tree was retrieved from the literature (Upham et al. 2019). The unrooted tree was used for PAML analysis.

Positive selection detection in modern human populations

We inferred SNPs that were under positive selection based on the following criteria: 1) the absolute value of LD-based test statistics $iHS > 2$ in the population detected. 2) SNPs with $iHS > 2$ also showed high pairwise F_{st} (> 0.20) between the detected population and the other population. We thus found 77 cancer driver genes under positive selection in total in different populations (Supplementary File 2).

Among these 77 genes, ten genes have been reported under positive selection, including host-pathogen interaction-related genes (PARP4(Daugherty et al. 2014; Gossmann and Ziegler 2014), SH2B3(Coenen et al. 2009; Zhernakova et al. 2010; Aaron 2011)), neuron-related genes (FAT1(Sato and Kawata 2018), NIN(Montgomery and Mundy 2012)), pigmentation-related genes (HERC2(Wilde et al. 2014; Cheng et al. 2017), EGFR(Lao et al. 2007)), metabolic-related genes (UGT2B17(Xue et al. 2008)), the well-known tumor suppressor gene TP53(Levine 2020), and the two genes mentioned above, i.e., Skull shape and facial morphology-related gene RGPD3 (Wu et al. 2019) and regulatory evolution-related ZNF gene ZNF626(Perdomo-Sabogal and Nowick 2019; Jovanovic et al. 2021). Gene enrichment analysis revealed that these 77 positively selected genes were involved in immune-related pathways, p53 pathways, neuron-related pathways, angiogenesis, the Insulin/IGF pathway, the PDGF pathway, the FGF pathway, the VEGF pathway, the EGF pathway, and the Wnt signaling pathway (Fig.2).

Figure 2 Pathway enrichment of positively selected cancer driver genes in modern human populations. % of gene list in the category is calculated for each testing list as: genes for the category/total genes in the list * 100. The red bar represents the reference human genome. The blue bar represents the target positively selected genes. It displays only results for FDR $p < 0.05$.

Cancer driver genes are under positive selection in different populations (Fig. 3, Supplementary File 2). There are 27 genes corresponding to 39 cancer types under positive selection in the African population, 26 genes corresponding to 47 cancer types under positive selection in the East Asian population, 35 genes corresponding to 61 cancer types under positive selection in the European population, and 6 genes corresponding to 11 cancer types under positive selection in the South Asian population. Note that the African population shows the largest extent of divergence with other populations. On the one hand, among 27

positively selected genes in the African population, there are 15 genes showing high F_{st} compared with the European population, 24 genes showing high F_{st} compared with the East Asian population, 17 genes showing high F_{st} compared with the South Asian population. On the other hand, other populations showed large numbers of positively selected genes with high F_{st} compared with the African population. For example, there are 32 out of 35 positively selected genes in the European population showing high F_{st} compared with the African population; 26 out of 26 positively selected genes in the East Asian population showing high F_{st} compared with the African population; and 6 out of 11 positively selected genes showing high F_{st} compared with the African population.

When took account the corresponding cancer types of positively selected cancer driver genes in each population, we found that the European population has the largest numbers of cancer types (61 cancer types), followed by East Asian (47 cancer types) and African (39 cancer types) populations. The South Asian population has the lowest numbers of cancer types (11 cancer types). These cancer types also exhibited population-specific patterns. For example, driver genes of cancer types of WT (Wilms tumor), ALL (acute lymphoblastic leukaemia) and NSCLC (non small cell lung cancer) are under positive selection in the African population but not in the East Asian population. Driver genes of cancer types of SG (salivary gland cancer) and BCNET (small intestine cancer neuroendocrine) are under positive selection in the South Asian population but not in the African population. Driver genes of cancer types of SG, DLBCL (diffuse large B cell lymphoma) and THCA (thyroid adenocarcinoma) are under positive selection in the South Asian population but not in the East Asian population. More details can be found in Fig. 3. Abbreviations of cancer types can be found in the Supplementary File 3.

Figure 3 Genes under positive selection in modern human populations, i.e., the African population, the European population, the East Asian population, and the South Asian population. Arrows showed positively selected genes with $iHS > 2$ in the local population and $F_{st} > 0.20$ between populations. There are 27 genes corresponding to 39 cancer types under positive selection in the African population, 26 genes corresponding to 47 cancer types under positive selection in the East Asian population, 35 genes corresponding to 61 cancer types under positive selection in the European population, and 6 genes corresponding to 11 cancer types under positive selection in the South Asian population. The African population showed the largest extent of divergence with other populations. The European population has the largest numbers of cancer types (61 cancer types) whose driver genes are under positive selection, followed by East Asian (47 cancer types) and African (39 cancer types) populations. The South Asian population has the lowest numbers of cancer types (11 cancer types). The corresponding cancer types whose driver genes are under positive selection showing population-specific patterns are given in boxes. Abbreviations of cancer types can be found in the Supplementary File 3.

Discussion

Using comparative genomic analysis, population genetic analysis, and computational molecular evolutionary analysis across the primate phylogeny and in modern human populations, we found that recent selection, rather than long timescale selection, has significant effects on the evolution of cancer driver genes in humans. Cancer driver genes related to morphological traits and local adaptation are under positive selection in different human populations. The corresponding cancer types of these positively selected genes showed population-specific patterns, providing clues to explain cancer incidence rates discrepancy in different human populations. These findings are helpful for understanding cancer evolution and providing clues for further precision medicine.

A few studies have already conducted a molecular evolutionary analysis of cancer genes across primates and found many cancer genes are under positive selection in primates (Lou et al. 2014). We also found tens of hundreds of genes under positive selection in primates (Fig. 1). This indicates that these genes indeed play advantageous roles during the early evolution of primates. The genetic background of many cancer genes can be traced back to before human lineage itself. For example, some cancer genes are essential biological genes controlling basic biology, such as DNA damage repair, cell division (O'Connell 2010; Wu et al. 2020; Zahir et al. 2020), or genes associated with adaptation (Muller 2017; Shi et al. 2019). As we mentioned in the introduction section, several hypotheses have already been proposed to explain this phenomenon.

However, when we focused on positive selection in the human lineage under a branch-site model, we found only two cancer driver genes under positive selection (Fig. 1). The reason could be that their harmful effects on human

evolution outweighed their advantageous roles in the early evolution of primates. Indeed, massive changes occurred during the evolution of human beings, such as upright walking (Niemitz 2010; Gruss and Schmitt 2015; Zirkle and Lovejoy 2019) and brain development (Kanton et al. 2019), which can induce different selection pressures in the evolution of the human lineage. In addition, nonhuman primates might have other cancer genetic bases or anticancer mechanisms, which are not rare in mammals, including mole rats, elephants and whales (Seluanov et al. 2018). Comparisons among species provide a snapshot of selected events that have been unfolding over long timescales. The few numbers of cancer driver genes under positive selection in the human lineage suggests that long timescale positive selection plays a minor role in driving the evolution of human cancer.

In contrast, 77 cancer driver genes are under recent positive selection in modern human populations. Most of these genes are related to host-pathogen interactions, neuron development, pigmentation and metabolism, which are involved in human-environmental interactions (Lao et al. 2007; Xue et al. 2008; Coenen et al. 2009; Zhernakova et al. 2010; Aaron 2011; Montgomery and Mundy 2012; Daugherty et al. 2014; Gossmann and Ziegler 2014; Wilde et al. 2014; Cheng et al. 2017; Sato and Kawata 2018; Levine 2020). Many genes showed population-specific positive selection patterns (Fig. 3), suggesting their adaptations to different environmental conditions. Significant effects of condition changes in modern human populations instead of long timescales evolution are also found in the evolution of testis-related genes (Schaschl and Wallner 2020). Food consumptions, environmental pollution, living stress and culture affecting diseases in modern human populations have increasingly attracted attention (review in (Benton et al. 2021)). The numbers of cancer driver genes under positive selection show geographical differences among populations, with the African population showing the largest extent of divergence with other populations. These suggest that cancer driver genes are involved in local

adaptation and could be related to the migration *out of Africa* during modern human evolution (Stewart and Stringer 2012; Scerri et al. 2019).

It is worth noting that the corresponding cancer types whose driver genes are under positive selection show population-specific patterns (Fig. 3). Wilms tumor (WT) is more common in Africa than in East Asia (Leslie et al. 2021), which is consistent with our findings that driver genes of WT are under positive selection in the African population but not in the East Asian population (Fig. 3). Another case is Ewing's sarcomas (EWS), which occurs more popular in the European population than in African or Asian populations. Genetic component is believed to be a key factor for developing this cancer (Worch et al. 2010). Our study gives a support of this opinion, giving that driver genes of EWS are found under positive selection in the European population, but not in African and Asian populations (Fig. 3). Noticeably, the South Asian population has the lowest numbers of cancer types whose driver genes are under positive selection (11 cancer types). This gives a good explanation why the South Asian population usually has low cancer incidence rate (Tran et al. 2018). Our findings that varied cancer driver genes are under different selection pressures in different populations can thus help explain geographical disparities of cancer incidence rates, although other factors such as environmental conditions and living styles are also important (Bahnassy et al. 2020). These results again, suggest that recent positive selection is a major force driving the evolution of cancer in humans.

A genetic disease can occur as a byproduct of evolution. For instance, the seemingly human-specific disease of schizophrenia and the greater human susceptibility to Alzheimer's disease may be a byproduct of the human specialization for higher cognitive function (Vamathevan et al. 2008). This can also be applied to cancer to humans. Cancer driver genes under recent positive

selection in human populations we found here suggest that pleiotropic effects should be under consideration during precision cancer medicine.

Materials and Methods

Orthologous sequences inference

Transcript IDs of 568 cancer driver genes across 66 cancer types were retrieved from the literature published in Nature Reviews Cancer recently (Martínez-Jiménez et al. 2020). The corresponding sequences retrieved from Ensembl Biomart were used as queries in a tblastx search with $e^{-0.001}$ against each species's related genome-wide coding sequences (cds). Cds of the first top tblastx hit of each queried transcript were extracted for the corresponding species. Oma was used to perform further orthologous sequence inference with default parameters. A species tree was retrieved from the literature (Upham et al. 2019), and *Mus musculus* was used as the outgroup.

Positive selection detection across the phylogeny of primates

The rates ω of nonsynonymous (dN) to synonymous (dS) substitutions can vary over sites and time. Positive selection detection was based on the rates ω calculated using a branch-site model with codeml in PAML. The branch-site model considers the variation in ω among sites and across branches to detect positive selection affecting sites along the target lineage (foreground branch). If a nonsynonymous mutation is less likely to be tolerated during evolution, ω will be $\ll 1$, meaning that purifying selection takes effects. If there is no selection pressure (neutral selection), ω will be equal to 1. If changes of nonsynonymous

mutations have beneficial effects and are favored by selection, ω will be $\gg 1$, indicating that they are under positive selection.

Multiple sequence alignment was conducted using MAFFT within T-coffee. An unrooted species tree retrieved from the literature (Upham et al. 2019) was used for the calculation. Nine branches were set as foreground branches to run the branch-site model separately (Fig. 1). Bonferroni correction was used for multiple test corrections. Parameter settings and statistical significance tests were the same as in the literature (Gu and Xia 2019).

Positive selection detection in modern human populations

The integrated haplotype score method (iHS) was applied to identify recently positively selected genes in different human populations. The principle of this method is that under the condition of neutral evolution, the high-frequency rise rate of the selected allele is faster than expected. The long-range association with the surrounding loci will not have enough time to be eliminated by recombination. Therefore, by measuring the association attenuation between core haplotypes and alleles at different distances, i.e., extended haplotype homozygosity (EHH), we can detect signals of recent positive selection in favor of variants that have not yet reached fixation. One positive selection feature is finding a core haplotype combining high frequency and high EHH compared with other core haplotypes at the site. Therefore, various core haplotypes at one locus can be used as internal controls to adjust the heterogeneity of the local recombination rate.

Phased genetic data of the 1000 genomes project of modern human populations were retrieved from <ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>. We analyzed larger geographical areas, including Africa, East Asia, Europe, and

South Asia. Details can be found in the Supplementary File 4. PLINK and VCFtools were used to process variant call format (VCF) files for all chromosomes. Only biallelic SNPs with minor allele frequency (MAF)>0.05 and indels removed were considered in this study. The human genome GTF file from Ensembl was used to retrieve genomic location of cancer genes. The integrated haplotype score test (iHS) implemented in selscan was used to detect genome-wide positive selection with default parameters. The unstandardized iHS scores were normalized in frequency bins across the entire genome using the script *norm* implemented in the selscan program. Pairwise Fst was calculated for each SNP using the Weir & Cockerham Fst calculation implemented in VCFtools.

Pathway enrichment analysis

The pathway enrichment analysis of target positively selected cancer driver genes was conducted using the Gene Ontology Resource (<http://geneontology.org/>) with the human genome as the background. Fisher's exact test was conducted, and false discovery rates (<0.05) were calculated to enrich significant pathways in the Panther pathways database.

Acknowledgements

We thank Dr. Juan Felipe Ortiz, Dr. Haosen Li, Dr. Cheng Huang and Prof. Canwei Xia for the support of data analyses. This work was supported by the Fundamental Research Funds for the Central Universities (20lgpy109) to LG and the National Natural Science Foundation of China (81772769) to GY; Data analyses was supported by National Supercomputer Center in Guangzhou, China. The English language was edited by Elsevier Language Editing Services.

Author Contributions

LG and GY designed the study. LG did the analysis. LG and GY wrote the manuscript.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statements

Detail software and corresponding parameters were given in the Methods.

Supplementary Files

Supplementary File 1 Genes under positive selection under branch-site model with PAML across the primate phylogeny.

Supplementary File 2 Recent positive selection detection within human populations and Fst comparisons between human populations.

Supplementary File 3 Abbreviations of 66 cancer types.

Supplementary File 4 IDs of population data retrieved from the 1000 human genomes project.

References

Aaron L. 2011. The last two millennias echo-catastrophes are the driving forces for the potential genetic advantage mechanisms in celiac disease. *Med. Hypotheses*. 77:773–776.

Bahnassy AA, Abdellateif MS, Zekri A-RN. 2020. Cancer in Africa: Is It a Genetic or Environmental Health Problem? *Front. Oncol*. 10:2508.

Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA. 2021. The influence of evolutionary history on human health and disease. *Nat. Rev. Genet*. 22:269–283.

Burrows JM, Bromham L, Woolfit M, Piganeau G, Tellam J, Connolly G, Webb N, Poulsen L, Cooper L, Burrows SR, et al. 2004. Selection pressure-driven evolution of the Epstein-

430 Barr virus-encoded oncogene LMP1 in virus isolates from Southeast Asia. *J. Virol.*
431 78:7131–7137.

432 Cheng X, Xu C, DeGiorgio M. 2017. Fast and robust detection of ancestral selective sweeps.
433 *Mol. Ecol.* 26:6871–6891.

434 Coenen MJH, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, van Leeuwen
435 M, Brouwer E, Boezen MH, Postma DS, et al. 2009. Common and different genetic
436 background for rheumatoid arthritis and coeliac disease. *Hum. Mol. Genet.* 18:4195–
437 4203.

438 Crespi BJ, Summers K. 2006. Positive selection in the evolution of cancer. *Biol. Rev. Camb.*
439 *Philos. Soc.* 81:407–424.

440 Daugherty MD, Young JM, Kerns JA, Malik HS. 2014. Rapid evolution of PARP genes
441 suggests a broad role for ADP-ribosylation in host-virus conflicts. *PLoS Genet.*
442 10:e1004403.

443 Fischer A, Greenman C, Mustonen V. 2011. Germline fitness-based scoring of cancer
444 mutations. *Genetics* 188:383–393.

445 Gossman TI, Ziegler M. 2014. Sequence divergence and diversity suggests ongoing
446 functional diversification of vertebrate NAD metabolism. *DNA Repair (Amst).* 23:39–48.

447 Gruss LT, Schmitt D. 2015. The evolution of the human pelvis: changing adaptations to
448 bipedalism, obstetrics and thermoregulation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*
449 370:20140063.

450 Gu L, Xia C. 2019. Cluster expansion of apolipoprotein D (ApoD) genes in teleost fishes. *BMC*
451 *Evol. Biol.* 19:9.

452 Huttley GA, Eastal S, Southey MC, Tesoriero A, Giles GG, McCredie MR, Hopper JL, Venter
453 DJ. 2000. Adaptive evolution of the tumour suppressor BRCA1 in humans and
454 chimpanzees. Australian Breast Cancer Family Study. *Nat. Genet.* 25:410–413.

455 Jovanovic VM, Sarfert M, Reyna-Blanco CS, Indrischek H, Valdivia DI, Shelest E, Nowick
456 K. 2021. Positive Selection in Gene Regulatory Factors Suggests Adaptive Pleiotropic
457 Changes During Human Evolution. *Front. Genet.* 12:662239.

458 Kang L, Michalak P. 2015. The evolution of cancer-related genes in hominoids. *J. Mol. Evol.*
459 [Internet] 80:37–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25249249>

460 Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchís-Calleja F, Guijarro P, Sidow L, Fleck

461 JS, Han D, et al. 2019. Organoid single-cell genomic atlas uncovers human-specific
462 features of brain development. *Nature* 574:418–422.

463 Kouprina N, Mullokandov M, Rogozin IB, Collins NK, Solomon G, Otstot J, Risinger JI,
464 Koonin E V, Barrett JC, Larionov V. 2004. The SPANX gene family of cancer/testis-
465 specific antigens: rapid evolution and amplification in African great apes and hominids.
466 *Proc. Natl. Acad. Sci. U. S. A.* 101:3077–3082.

467 Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M. 2007. Signatures of positive
468 selection in genes associated with human skin pigmentation as revealed from analyses of
469 single nucleotide polymorphisms. *Ann. Hum. Genet.* 71.

470 Leslie SW, Sajjad H, Murphy PB. 2021. Wilms Tumor. StatPearls Publishing Available from:
471 <http://www.ncbi.nlm.nih.gov/pubmed/28723033>

472 Levine AJ. 2020. p53: 800 million years of evolution and 40 years of discovery. *Nat. Rev.*
473 *Cancer* 20:471–480.

474 Lou DI, McBee RM, Le UQ, Stone AC, Wilkerson GK, Demogines AM, Sawyer SL. 2014.
475 Rapid evolution of BRCA1 and BRCA2 in humans and other primates. *BMC Evol. Biol.*
476 14:155.

477 Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton
478 MR, Campbell PJ. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues.
479 *Cell* 171:1029–1041.e21.

480 Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, Mularoni
481 L, Pich O, Bonet J, Kranas H, et al. 2020. A compendium of mutational cancer driver
482 genes. *Nat. Rev. Cancer* 20:555–572.

483 Montgomery S, Mundy N. 2012. Positive selection on NIN, a gene involved in neurogenesis,
484 and primate brain evolution. *Genes. Brain. Behav.* 11.

485 Muller AWJ. 2017. Cancer is an adaptation that selects in animals against energy dissipation.
486 *Med. Hypotheses* 104:104–115.

487 Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A,
488 Tanenbaum DM, Civello D, White TJ, et al. 2005. A scan for positively selected genes in
489 the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.

490 Niemitz C. 2010. The evolution of the upright posture and gait--a review and a new synthesis.
491 *Naturwissenschaften* 97:241–263.

492 O'Connell MJ. 2010. Selection and the cell cycle: positive Darwinian selection in a well-
493 known DNA damage response pathway. *J. Mol. Evol.* 71:444–457.

494 Perdomo-Sabogal Á, Nowick K. 2019. Genetic Variation in Human Gene Regulatory Factors
495 Uncovers Regulatory Roles in Local Adaptation and Disease. *Genome Biol. Evol.*
496 11:2178–2193.

497 Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko
498 J V, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the
499 human genome from haplotype structure. *Nature* 419:832–837.

500 Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH,
501 McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of
502 positive selection in human populations. *Nature* 449:913–918.

503 Sato DX, Kawata M. 2018. Positive and balancing selection on SLC18A1 gene associated with
504 psychiatric disorders and human-unique personality traits. *Evol. Lett.* 2:499–510.

505 Scerri EML, Chikhi L, Thomas MG. 2019. Beyond multiregional and simple out-of-Africa
506 models of human evolution. *Nat. Ecol. Evol.* 3:1370–1372.

507 Schaschl H, Wallner B. 2020. Population-specific, recent positive directional selection
508 suggests adaptation of human male reproductive genes to different environmental
509 conditions. *BMC Evol. Biol.* 20:27.

510 Seluanov A, Gladyshev VN, Vijg J, Gorbunova V. 2018. Mechanisms of cancer resistance in
511 long-lived mammals. *Nat. Rev. Cancer* 18:433–441.

512 Shendure J, Akey JM. 2015. The origins, determinants, and consequences of human mutations.
513 *Science* 349:1478–1483.

514 Shi Y, Fan S, Wu M, Zuo Z, Li X, Jiang L, Shen Q, Xu P, Zeng L, Zhou Y, et al. 2019.
515 YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. *Nat.*
516 *Commun.* 10:4892.

517 Stewart JR, Stringer CB. 2012. Human evolution out of Africa: the role of refugia and climate
518 change. *Science* 335:1317–1321.

519 Tran HN, Udaltsova N, Li Y, Klatsky AL. 2018. Low Cancer Risk of South Asians: A Brief
520 Report. *Perm. J.* 22:17–095.

521 Upham NS, Esselstyn JA, Jetz W. 2019. Inferring the mammal tree: Species-level sets of
522 phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.*

523 17:e3000494.

524 Vamathevan JJ, Hasan S, Emes RD, Amrine-Madsen H, Rajagopalan D, Topp SD, Kumar V,
525 Word M, Simmons MD, Foord SM, et al. 2008. The role of positive selection in
526 determining the molecular cause of species differences in disease. *BMC Evol. Biol.* 8:273.

527 Vicens A, Posada D. 2018. Selective Pressures on Human Cancer Genes along the Evolution
528 of Mammals. *Genes (Basel)*. 9.

529 Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M, Hollfelder N,
530 Potekhina ID, Schier W, Thomas MG, et al. 2014. Direct evidence for positive selection
531 of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad.*
532 *Sci. U. S. A.* 111:4832–4837.

533 Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M. 2003. Implications of natural
534 selection in shaping 99.4% nonsynonymous DNA identity between humans and
535 chimpanzees: enlarging genus *Homo*. *Proc. Natl. Acad. Sci. U. S. A.* 100:7181–7188.

536 Worch J, Matthay KK, Neuhaus J, Goldsby R, DuBois SG. 2010. Ethnic and racial differences
537 in patients with Ewing sarcoma. *Cancer* 116:983–988.

538 Wu W, Zhai G, Xu Z, Hou B, Liu D, Liu T, Liu W, Ren F. 2019. Whole-exome sequencing
539 identified four loci influencing craniofacial morphology in northern Han Chinese. *Hum.*
540 *Genet.* 138:601–611.

541 Wu Z, Li S, Tang X, Wang Y, Guo W, Cao G, Chen K, Zhang M, Guan M, Yang D. 2020.
542 Copy Number Amplification of DNA Damage Repair Pathways Potentiates Therapeutic
543 Resistance in Cancer. *Theranostics* 10:3939–3951.

544 Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP, et al.
545 2008. Adaptive evolution of UGT2B17 copy-number variation. *Am. J. Hum. Genet.*
546 83:337–346.

547 Zahir N, Sun R, Gallahan D, Gatenby RA, Curtis C. 2020. Characterizing the ecological and
548 evolutionary dynamics of cancer. *Nat. Genet.* 52:759–767.

549 Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, de Kovel CGF,
550 Franke L, Oosting M, Barisani D, et al. 2010. Evolutionary and functional analysis of
551 celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am. J.*
552 *Hum. Genet.* 86:970–977.

553 Zirkle D, Lovejoy CO. 2019. The hominid ilium is shaped by a synapomorphic growth

554 mechanism that is unique within primates. Proc. Natl. Acad. Sci. U. S. A. 116:13915–
555 13920.
556