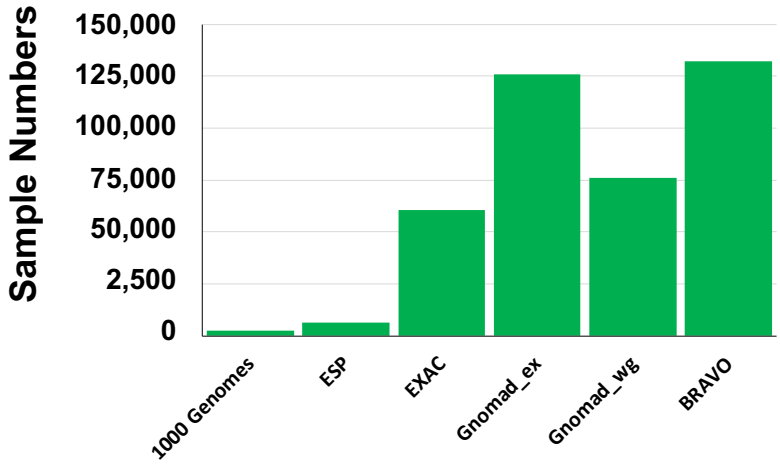
8SNPs in the coding region of the TP53 gene: gazing at the needles in a haystack of pathogenic variants

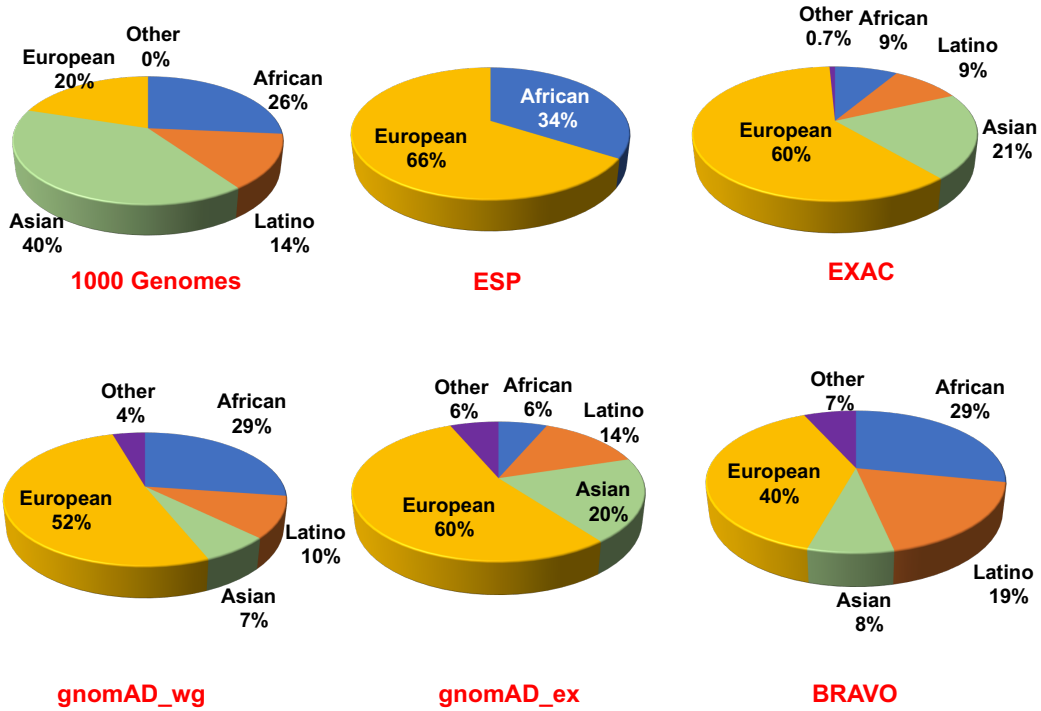
Thierry Soussi

Supplementary document

Contents

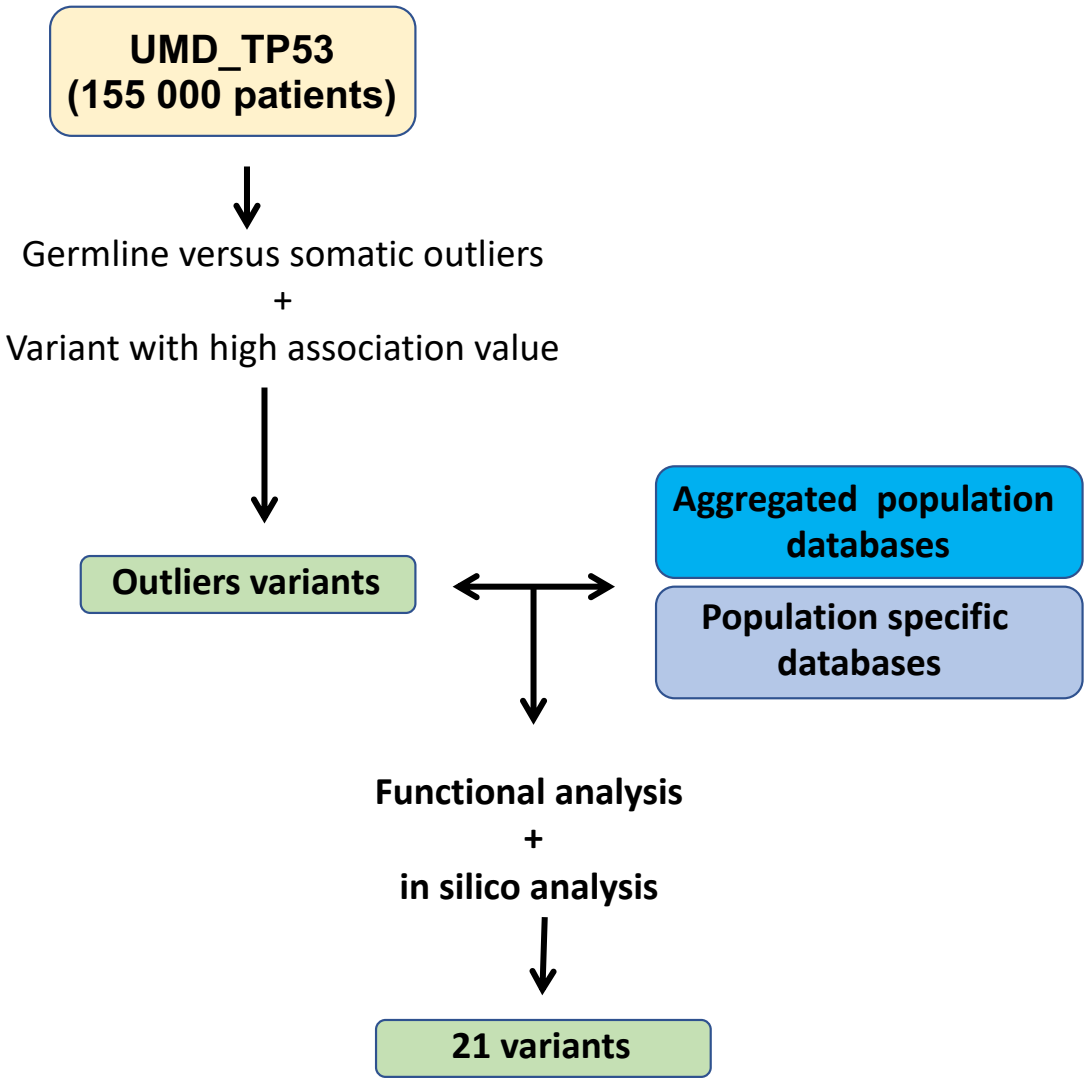
Supplementary Figure S1: Population data. 2
Supplementary Figure S2: Flow chart of the strategy used to identify new benign TP53 variants. 3
Methods: identification of putative TP53 SNPs: a summary. 4
Supplementary Figure S3: Occurrence of TP53 variant associations. 5
Supplementary Figure S4: Germline versus somatic (GVS) ratio of TP53 variants. 6
Asian variants are not associated with an increased risk of cancer. 7
Supplementary Figure S5: Prevalence of the four Asian-specific TP53 SNPs in three case-control cohort studies in breast, colorectal and pancreatic cancer. 7
Supplementary Figure S6: African specific *TP53* variants. 8
Supplementary Figure S7: *TP53* variants in mixed populations. 9
Supplementary Figure S8: prevalence of eight *TP53* SNPs in a case-control cohort study in breast cancer. 10





Supplementary Figure S1: Population data.

Bar chart: Number of individuals in the most common datasets used for population analyses. 1000 genomes (2,504 individuals) ESP: NHLBI Go Exome Sequencing Project (6,500 individuals); ExAC: Exome Aggregation Consortium (60,500 individuals); gnomAD: The Genome Aggregation Database whole exome v 2.1.1 (_ex) or whole genome c 3.1.2 (_wg); (125,748 and 76,156 individuals respectively); BRAVO: data from TOPMed (Trans-Omics for Precision Medicine) version 8. (132,345 individuals);
The six pie charts show the breakdown of subpopulations in each global dataset.



Supplementary Figure S2: Flow chart of the strategy used to identify new benign TP53 variants.

Methods: identification of putative TP53 SNPs: a summary.

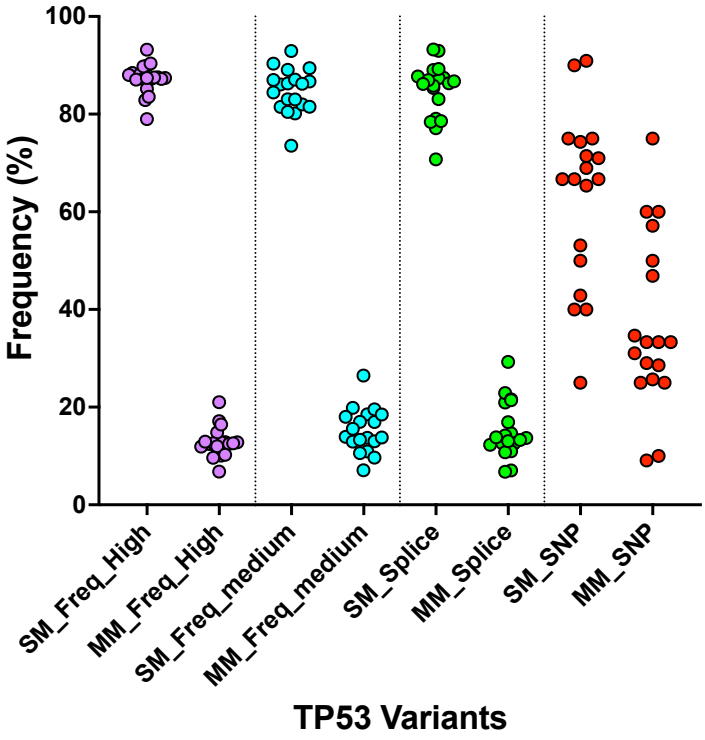
Several lines of evidence led to the identification of 19 novel *TP53* SNPs, including ten variants in the DNA binding domain where the majority of cancer driver mutations are localized. First, it was observed that many *TP53* variants included in the UMD_*TP53* database were frequently found in tumors carrying more than one of them. That situation seemed unusual as, except for several hematological neoplasia, only one *TP53* driver variant is normally found per tumor (**Supplementary Figure S3)** (1). Indeed, driver variants, whether hot spot, moderately frequent or splice, were only infrequently (range 5–15% of the cases) associated with a second *TP53* variant. For these 19 new variants, this frequency ranged between 20 and 80%, suggesting that the second mutation, known to be pathogenic in most cases, is co-selected by the driver event.

The second clue leading to the identification of these SNPs came from the analysis of germline variants in the UMD_*TP53* database. Some of these variants were found disproportionally more frequently as germline rather than somatic events. Because many of these variants were not associated with any hereditary cancer, it seemed possible that they were genuine SNPs. The germline-to-somatic (GVS) ratio for each *TP53* variant in the UMD_*TP53* database was evaluated to detect any enrichment in germline mutations. This ratio was very low (range 0.01 to 0.1) for most cancer-associated *TP53* variants, including pathogenic variants found in hereditary cancer, but far higher (range 0.5 to 15) for the 19 variants discussed here **(Supplementary Figure S4)**. Additionally, this ratio was very low or null for variants specifically identified in tumors linked to environmental carcinogen exposure, such as liver cancer associated with aflatoxin exposure (p.Arg249Ser) or lung cancer in smokers (p.Val157Phe).

The Cancer Genome Atlas (TCGA) released a genetic analysis of 10,255 patient samples from 32 different cancer types. Germline variants were extensively curated and removed to provide an accurate picture of the somatic events occurring in these tumors. Not surprisingly, the two common *TP53* SNPs, rs1042522;p.Pro72Arg and rs1800371;p.Pro47Ser, were absent from the 3,786 patients with *TP53* mutations. However, all but one of the 19 *TP53* variants discussed here were also absent from this somatic dataset. The only exception, rs563378859;p.Arg156Cys, was found in a single brain tumor expressing two other *TP53* variants.

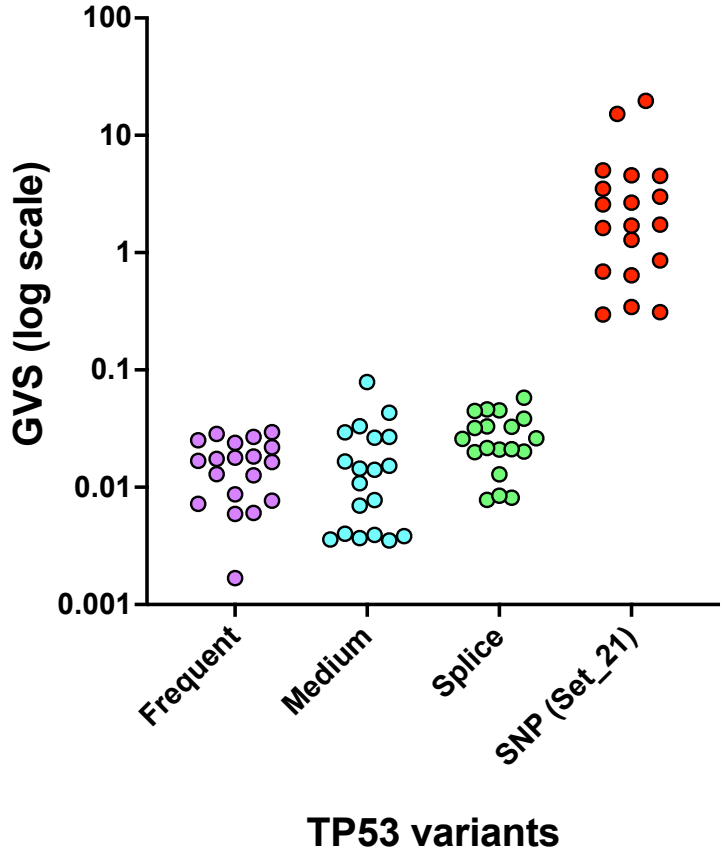
Finally, a population analysis of *TP53* variants in the UMD_*TP53* database revealed the population-specific origin of several *TP53* variants, such as rs587780728;p.Asp49His or rs201753350;p.Val31Ile in the Japanese population or rs368771578; p.Tyr107His in the African population. Although mostly defined as somatic, it seems likely that all these variants are indeed misidentified germline variants.

These observations prompted an extensive analysis of all *TP53* variants included in not only the widely-used population databases such as gnomAD or EXAC but also in the population-specific repositories not currently included in the global databases (2). From that analysis, 19 new *TP53* variants were found to be potential benign SNPs and are discussed in the manuscript. An analysis of the two well-characterized *TP53* SNPs, rs1042522;p.Pro72Arg and rs1800371;p.Pro47Ser is also included.



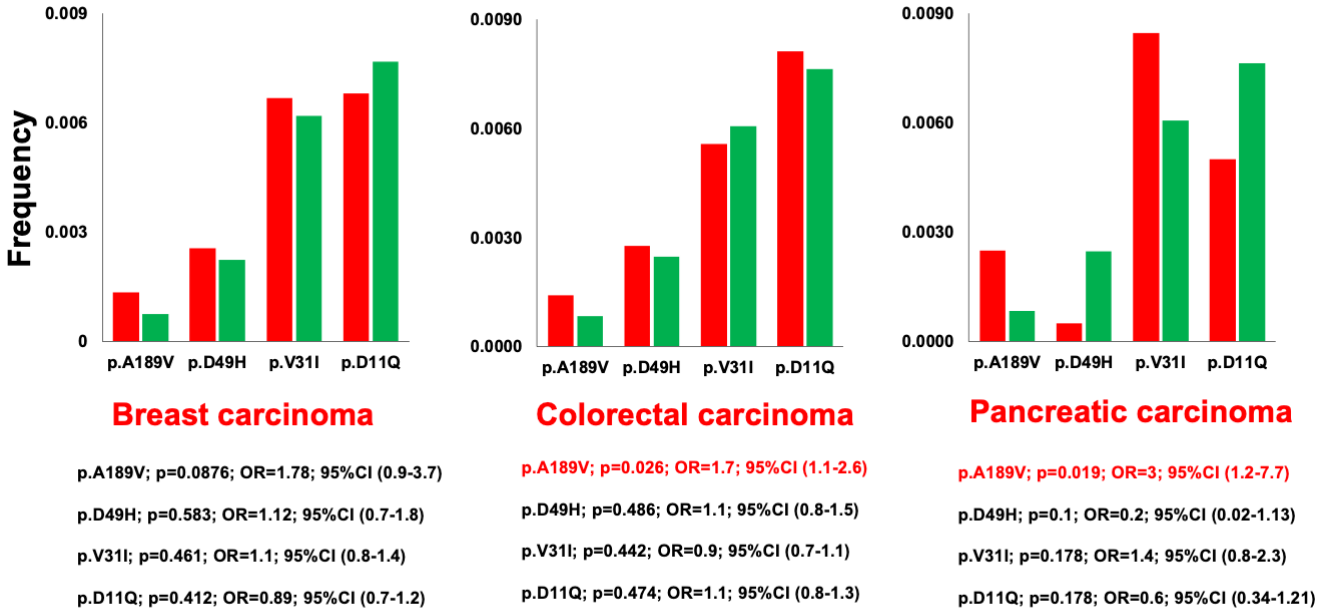
Supplementary Figure S3: Occurrence of TP53 variant associations.

For each TP53 variant, the frequency of tumors found only with that variant (SM) or found associated with at least one other TP53 variant (MM) have been evaluated. **Freq_High**: missense TP53 variants found more 800 times in the database (range occurrence in UMD_TP53 800-6000); **medium**: missense TP53 variants found at medium frequency (range occurrence in UMD_TP53 200-250); **Splice**: TP53 variants localized in splice donor or acceptor sequences (+1/+1 or -1/-2) (range occurrence in UMD_TP53 100-300); **SNP**: eighteen TP53 SNP variants. Variants rs1042522 (p.Pro72Arg) and rs1800371 (p.Pro47Ser) were not included as they are bona fide SNPs always excluded from the UMD_TP53 database. Variant rs144386518; p.Pro58Arg was only found twice in multimutated tumors. In most cases the second TP53 variant is pathogenic. In several tumors known to express some of these variants as a unique mutation, reanalysis of the normal DNA confirmed the germline origin of the variant (2).



Supplementary Figure S4: Germline versus somatic (GVS) ratio of TP53 variants.

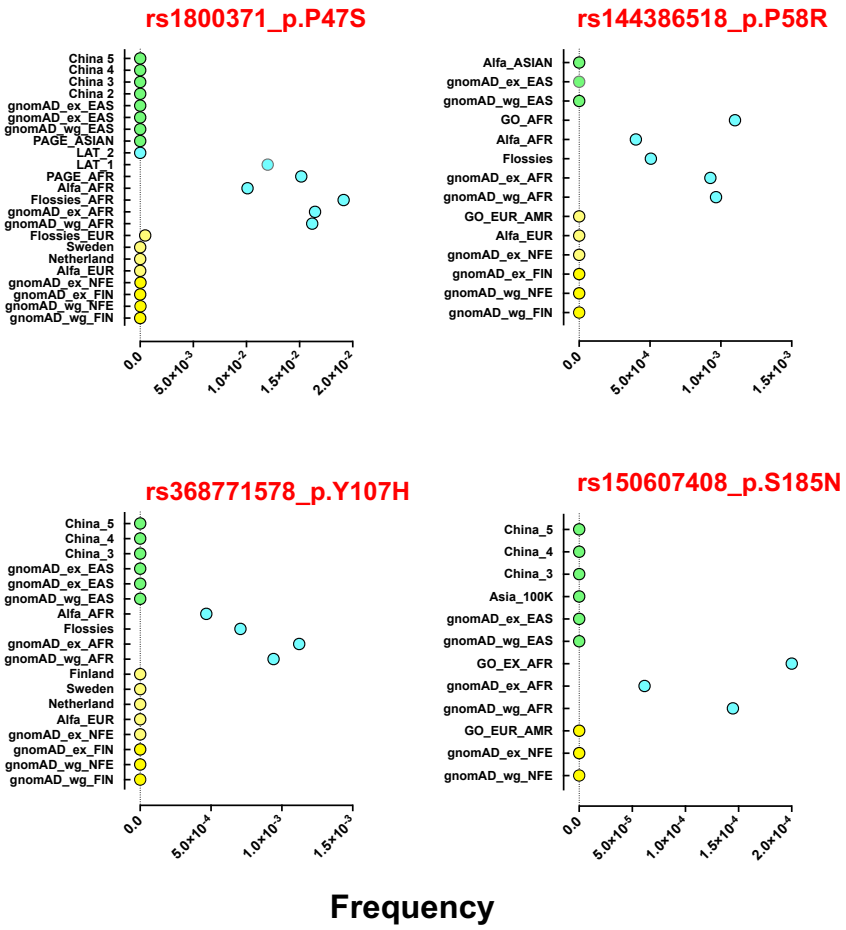
Nineteen variants taken from four different subsets of the database were analyzed. **Frequent**: missense TP53 variants found more 800 times in the UMD_TP53 database (range occurrence 800-6000); **Medium**: missense TP53 variants found at medium frequency (range occurrence 200-250); **Splice**: TP53 variants localized in splice donor or acceptor sequences (+1/+1 or -1/-2) (range occurrence 100-300); **SNP**: nineteen TP53 SNPs variants. Variants rs1042522 (p.Pro72Arg) and rs1800371 (p.Pro47Ser) were not included as they are bona fide SNPs always excluded from the UMD_TP53 database.



Supplementary Figure S5: Prevalence of the four Asian-specific TP53 SNPs in three case-control cohort studies in breast (left panel), colorectal (central panel) and pancreatic (right panel) cancer.
The odds ratio for each variant is shown below the bar charts. Only variant p.A189V shows a significant association with a higher risk of colorectal and pancreatic carcinomas but the low prevalence of this variant and the weak significance for it indicates that a larger cohort study will be needed to confirm the association (3–5).

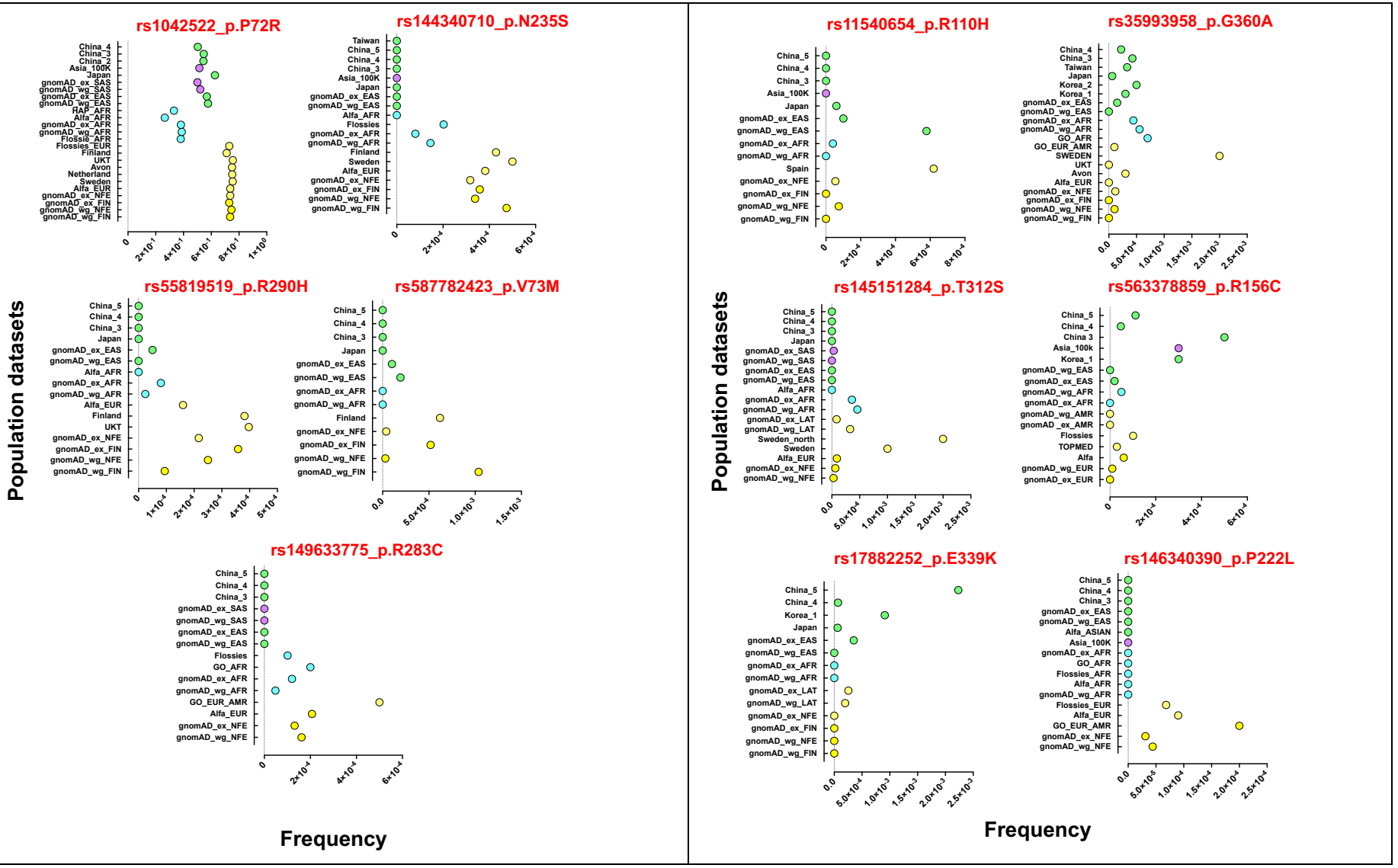
Asian variants are not associated with an increased risk of cancer.

Three Japanese studies have focused on identifying germline pathogenic variants associated with an increased risk of breast, colorectal or pancreatic cancer (3–5). In all three, patient cases (1,007, 12,503 and 7,051 for pancreatic, colorectal and breast cancer respectively) were sequenced for various predisposition genes including *TP53*, and variant frequencies were compared to a same control population (23,705 Japanese individuals). For rs587780728;p.D49H, rs201753350;p.V31I and rs201382018;p.D11Q, AF was statistically similar in both cases and controls **(Supplementary Figure S5)**.
For variant rs121912665;p.A189V, an association was found with colorectal and pancreatic cancers but the statistical significance was low (p=0.026 and p= 0.019 respectively). Although this variant, located in the DNA binding domain of *TP53*, has been shown to be fully functional (see below), it is possible that a slight defect remains undiscovered. Nevertheless, due to its low frequency compared to the three other variants, larger cohorts will be needed to fully uncover the effects of rs121912665;p.A189V.



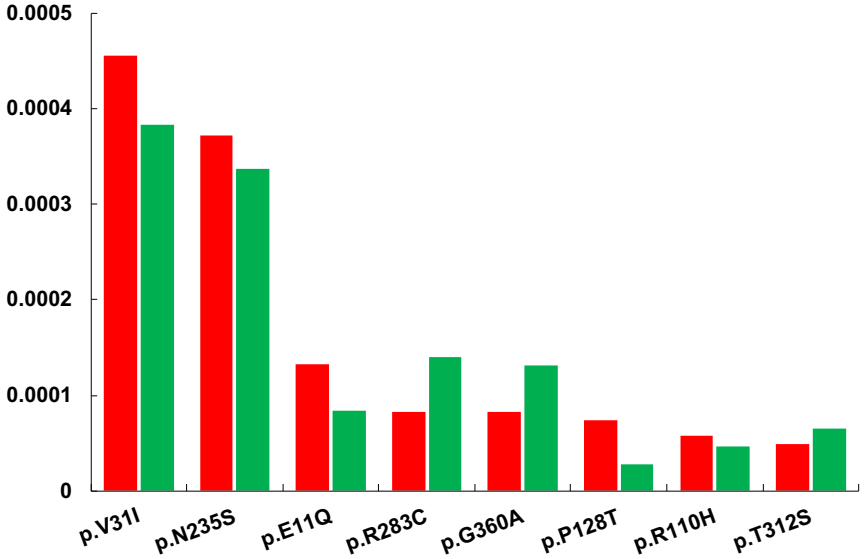
Supplementary Figure S6: African specific *TP53* variants.

Frequency of the four African-specific *TP53* SNPs in various populations. The sources of the different populations are provided in Supplementary Table S1. Population color codes: yellow: European; blue: African; green: East Asian; purple: South Asian.



Supplementary Figure S7: *TP53* variants in mixed populations.

The sources of the different populations are provided in **Supplementary Table S1**. Population color codes: yellow: European; blue: African; green: East Asian; purple: South Asian.



p.V31I; p=0.408; OR=1.2; 95%CI (0.8-1.8)
p.N235S; p=0.655; OR=1.1; 95%CI (0.7-1.7)
p.E11Q; p=0.278; OR=1.6; 95%CI (0.7-3.6)
p.R283C; p=0.195; OR=0.6; 95%CI (0.3-1.3)
p.G360A; p=0.267; OR=0.6; 95%CI (0.3-1.4)
p.P128T; p=0.143; OR=2.7; 95%CI (0.7-9.8)
p.R110H; p=0.716; OR=1.2; 95%CI (0.4-3.9)
p.T312S; p=0.267; OR=0.8; 95%CI (0.3-2.3)

Supplementary Figure S8: prevalence of eight *TP53* SNPs in a case-control cohort study in breast cancer.

The odds ratio for each variant is shown below the bar graph. Samples from 53,461 controls and 60,466 breast cancer patients were included upon completion of all quality control steps. Individuals with an Asian background made up 16% of this cohort (6).

References

1. Donehower LA, Soussi T, Korkut A et al. Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Rep* 2019;28:1370-1384.e5.
2. Doffe F, Carbonnier V, Tissier M et al. Identification and functional characterization of new missense SNPs in the coding region of the TP53 gene. *Cell Death Differ* 2021;28:1477-92.
3. Mizukami K, Iwasaki Y, Kawakami E et al. Genetic characterization of pancreatic cancer patients and prediction of carrier status of germline pathogenic variants in cancer-predisposing genes. *EBioMedicine* 2020;60:103033.
4. Fujita M, Liu X, Iwasaki Y et al. Population-based Screening for Hereditary Colorectal Cancer Variants in Japan. *Clin Gastroenterol Hepatol* 2020S1542-3565(20)31664.
5. Momozawa Y, Iwasaki Y, Parsons MT et al. Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. *Nat Commun* 2018;9:4083.
6. Breast CAC, Dorling L, Carvalho S et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 2021;384:428-39.