Differences in somatic TP53 Mutation Type in Breast Tumors by Race and Receptor Status

Nijole C Pollock
The Ohio State University
Johnny R Ramroop
The Ohio State University
Heather Hampel
The Ohio State University College of Medicine
Melissa A Troester
University of North Carolina at Chapel Hill Louis Round Wilson Special Collections Library: The University of North Carolina at Chapel Hill
Kathleen Conway
University of North Carolina at Chapel Hill
Jennifer J Hu
University of Miami School of Medicine
Jo L Freudenheim
University at Buffalo
Olufunmilayo I Olopade
University of Chicago Biological Sciences Division: University of Chicago Division of the Biological Sciences
Dezheng Huo
University of Chicago Biological Sciences Division: University of Chicago Division of the Biological Sciences
Elad Ziv
University of California San Francisco
Susan L Neuhausen
City of Hope Comprehensive Cancer Center Duarte
Patrick Stevens
The Ohio State University
Joseph McElroy
The Ohio State University
Amanda Ewart Toland (amanda.toland@osumc.edu)
The Ohio State University https://orcid.org/0000-0002-0271-1792

Research article

Keywords: TP53, gain-of-function, loss-of function, dominant negative activity, racial differences, breast cancer, somatic mutation, triple negative breast cancer, estrogen negative breast cancer

DOI: https://doi.org/10.21203/rs.3.rs-268719/v1

Abstract

Background: Somatic driver mutations in *TP53* are associated with triple negative breast cancer (TNBC) and poorer outcomes. Breast cancers arising in women of African ancestry (AA) are more likely to be TNBC and have somatic *TP53* mutations than non-Hispanic White (NHW) women. Missense driver mutations in *TP53* are known to have different functional effects including loss-of function (LOF) or gain-of-function (GOF) activity. A subset of variants exhibit dominant negative (DNE) activity over wildtype TP53 or other TP53 family members.

Methods: To determine if there were differences in *TP53* mutation type by race or TNBC status we identified published and unpublished datasets with somatic mutation data, race, age, and hormone receptor status. Mutations were classified as LOF, GOF or CpG Island hotspot by published literature or type of mutation (e.g. frameshift and nonsense mutations were considered to be LOF). We used Fisher's Exact test to assess for significant differences.

Results: We identified 96 independent breast cancer studies with race and somatic *TP53* mutation status available. The study comprised data from a total of 2964 women with somatic *TP53* mutations: 715 (24.1%) Asian, 258 (8.7%) AA, 1931 (65.2%) NHW, and 60 (2%) Latina. Of the total *TP53* mutations, 939 (31.7%) were GOF, 1739 (58.7%) were LOF, and the remaining 286 (9.6%) were not able to be classified. With respect to DNE activity, 1246 (42%) showed activity, 1190 (40.1%) showed no activity, and 528 (17.8%) were unknown status. The distribution of TP53 mutation type was similar by race and ethnicity. However, 35.8% of tumors from NHW individuals had GOF mutations compared to 29% in AA individuals (p=0.04). Mutations lacking DNE activity were positively associated with TNBC (OR=1.37, p=0.03) and estrogen receptor (ER) negative status (OR=1.38; p=0.005).

Conclusions: In summary, ER negative and TNBC breast tumors are less likely to have DNE+ *TP53* mutations which could reflect biological processes. Larger cohorts are needed to further elucidate some of these findings.

Introduction

Tumor protein 53, encoded by *TP53*, is a transcription factor with tumor suppressive activity that regulates target genes in response to cellular stress. Pathways regulated by TP53 include cell cycle check points, senescence, DNA repair, cell metabolism and apoptosis. Somatic mutations in *TP53* are the most common genetic abnormality in multiple cancer types. *TP53* somatic mutations are found in 40–60% of breast cancer cases (1–3) and are a negative prognostic factor, associated with more aggressive triple-negative breast cancers (TNBCs) and basal-like breast cancers (BLBCs) (4, 5).

Over eighty percent of *TP53* mutations are missense mutations with consequences that differ depending on the mutation position and amino acid change (6). *TP53* hotspot mutations make up approximately 30% of all *TP53* somatic variants and result in changes in the DNA-binding domain. Hotspot mutations frequently occur at positions containing a 5'-methylated cytosine, involving C to T transitions at CpG sites

(6, 7). Aging, environmental exposures, and other factors may impact the frequency of C to T transitions at these sites (8). Pathogenic somatic mutations in *TP53* often disrupt DNA binding capability, impair transcriptional activity and result in other loss-of-function (LOF) effects. However, a subset of missense somatic variants lead to new gain-of-function (GOF) activities. GOF activity is frequently mediated by the mutant protein binding to other tumor suppressive or oncogenic proteins or to novel regulatory regions (9). GOF mutations result in accelerated onset of tumors, metastasis, drug resistance and poorer survival outcomes (10). *TP53* missense mutations can also display dominant negative activity (DNE), in which one mutant TP53 protein disrupts the activity of the non-mutant protein during tetramerization. This effect is more common at hotspot mutation sites and may contribute to accelerated loss of heterozygosity and tumor progression (11). TP53 DNE can also impact the function of other TP53 family members including TP63 and TP73 (12). Because the importance of *TP53* mutations has been well-established for decades, there are abundant functional studies identifying LOF, GOF, and DNE activity for specific *TP53* mutations.

TNBCs, which are negative for estrogen receptor (ER) and progesterone receptor (PR) expression and lack HER2 amplification, and BLBCs have poorer prognoses compared to ER-positive breast cancer subtypes (13). BLBCs are often triple negative and are defined by expression of basal cytokeratins (14). TNBCs and BLBCs are more likely to have somatic *TP53* mutations than other types of breast cancer such as Luminal A and Luminal B subtypes (15). Interestingly, racial differences are observed between different breast cancer subtypes; TNBC and BLBC occur more frequently in women of African Ancestry (AA) (28–30%) or Latina ethnicity (17.5%) compared to non-Hispanic White (NHW) (12–15%) women (16–20). AA women have 2–3 fold higher prevalence of BLBC as defined by the PAM50-gene set predictor relative to NHW women and a 42% higher breast cancer mortality rate compared to NHW women (21–23). In addition, AA women have a higher risk of tumor recurrence than NHW women (hazard ratio, 2.22; CI; 1.05–4.67)(24).

As TNBCs are more common in breast tumors from AA women than NHW women, and *TP53* mutations are more frequently observed in TNBC and BLBC than other subtypes, it is not surprising that the proportion of breast tumors with *TP53* mutations is 1.5 to 1.6-fold higher in AA than NHW women (24–26). While there has been extensive research about *TP53* somatic mutation frequency overall and by race, there has been little investigation to determine if there are differences by *TP53* mutation type. Given that *TP53* mutation effects can impact prognosis, mutation type is an important consideration. Because of the differences in outcomes between AA women and NHW women, even after accounting for subtype differences, and the literature supporting different outcomes for GOF versus LOF *TP53* mutations, we hypothesized that there would be frequency differences in types of *TP53* mutations across racial and ethnic groups. To test this hypothesis, we compared the racial distribution of *TP53* mutation type in breast cancer using existing published and unpublished datasets.

Methods

Summary of data

This study was approved by the Ohio State Cancer Institutional Review Board.
Data for this study were ascertained from multiple sources including The Ohio State University Total Cancer Care repository, existing data in publicly accessible databases (TCGA, IARC, dbGAP), existing data contained in publications, and unpublished data ascertained during follow-up with authors. A description of all studies included in this paper are detailed in Additional File 1.

Study inclusion and exclusion criteria

For inclusion in our analyses, a study was required to have *TP53* sequence data from tumor DNA for at least exons 5–8 using any method such as targeted sequencing (Sanger or panel-based next-generation sequencing), exome sequencing or whole genome sequencing. We excluded studies that used immunohistochemistry or other non-DNA-sequencing based methods to infer *TP53* mutation genotype. All likely invasive stages, grades, and morphologies of primary breast tumor were considered. Non-invasive ductal carcinoma in-situ (DCIS) tumors were excluded. Data were annotated with race and ethnicity by the original authors or were from homogeneous populations. If available, data describing patient age, tumor grade, stage, receptor status, and morphology were collected. Studies were excluded if there was no ancestry information or individual level mutation data available. Studies also were excluded if there was a high degree of admixture within a population (e.g. Brazil) or if they represented a unique population that was underpowered to detect difference (e.g. 10 individuals from Egyptian ancestry). Individuals were excluded from study if their breast tumors did not carry a somatic *TP53* mutation. We considered studies of any design that fit this criteria, including population and clinical-based studies.

Data from unpublished datasets

Previously unpublished *TP53* somatic mutation data and self-reported race and ethnicity were ascertained from The Ohio State University's Total Cancer Care repository for 143 women (Additional File 2). Other previously or partially unpublished data included 163 individuals with *TP53* mutant breast tumors enrolled through The Western New York Exposures and Breast Cancer (WEB) study, 24 through The Sylvester Comprehensive Cancer Center, and 21 through The City of Hope Comprehensive Cancer Center. Patients were enrolled through protocols approved by their respective institutional review boards.

Data from publicly accessible databases

We identified individuals and studies in databases with somatic mutation information that met inclusion criteria. From the International Agency for Research on Cancer (IARC)(27), 1254 individuals from 78 studies met inclusion criteria, as well as 333 from The Cancer Genome Atlas (TCGA), and 637 from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (28, 29)(Additional File1).

Data from literature review

To obtain additional data from previously published work, we conducted a literature search for studies in which individual level *TP53* data and race/ethnicity information were available. A PubMed search using "*TP53* and race" identified 277 articles. We excluded articles: 1) already captured from database search (e.g. studies of TCGA data or previously reported in IARC); 2) of cancers other than breast; 3) on germline

variants or polymorphisms in *TP53*; and 4) where study inclusion were not met. For studies without individual level race and/or ethnicity information, we contacted authors by e-mail to request this information. All studies included are listed in Additional File 1.

Categorizing TP53 missense variants

TP53 studies have included different variants and focused on different aspects of p53 function, so we used a standardized approach to evaluate findings from IARC and other *TP53* literature. All missense variant annotations were based on existing functional studies with cell culture, yeast, or animal experiments; we did not consider *in-silico* testing alone for inclusion. However, in cases of mutations with uncharacterized function, we utilized PHANTM (Broad Institute) to exclude variants with a predicted function close to wild-type p53 (maximum PHANTM score < 0, ~50 variants)(11). We excluded well-established germline polymorphisms, such as p.P72R, and mutations with activity comparable to p53 wild type. *TP53* mutations were categorized by function (GOF or LOF) and dominant negative activity (DNE + or DNE-) as two separate criteria.

We described function as GOF or LOF. GOF mutations resulted in significantly different activity from both TP53 null and TP53 WT proteins such as novel transcript activity, TP73 interference, growth advantage, and facilitation of oncogene activity. LOF mutations had evidence of protein truncation, loss of tetramerization, or activity comparable to TP53 null. When we found reports of both LOF and GOF activities, but not direct contradictions for the same TP53 function, we categorized variants as GOF. Variants with limited functional data available and PHANTM prediction scores that differed from wild-type were annotated as unknown. DNE + variants were those with published evidence that the TP53 mutant protein interfered with TP53 WT function in heterozygous cells. DNE + mutants formed heterotetramers with TP53 WT units and changed TP53 WT function, causing a dominant GOF or LOF effect. Transcript-truncating mutations, such as nonsense, splicing, frameshifts, and large deletions, were assumed to be LOF without DNE. Hotspot codons were defined as those at positions 175, 245, 248, 249, 273 and 282 and CpG hotspot mutations were defined as C to T transitions at those codons plus R158H and P152L as described (7)

For tumors with multiple *TP53* somatic mutations, we considered the sum of multiple predicted effects. If any mutations were DNE+, the tumor was considered DNE+. GOF and LOF mutations were prioritized over unknown or functional mutations. Tumors with both GOF and LOF mutations were called GOF/LOF.

Statistical analyses

A Fisher's exact test for count data was used for comparisons between mutation categories (GOF/LOF, DNE/not DNE or CPG hotspot/not hotspot) and race, TNBC status, and ER status. For comparisons of mutation categories and age, a Welch Two Sample t-test was used. Analyses were run in R version 3.6.3 (2020-02-29) (30). A comparison-wise p-value of 0.05 was considered significant.

Results

Characteristics of study population

The study population is summarized in Table 1. We included somatic *TP53* mutation data from 2964 breast cancers from 96 studies for analysis (Additional File 1). Patients were categorized into 4 racial/ethnic groups. The study population was 65.2% NHW (n = 1931), 24.1% Asian (n = 715), and 8.7% AA (n = 258). Two percent (n = 60) of patients had Hispanic or Latina ethnicity with European or undefined race (n = 47 [1.6%] and n = 13 [0.4%], respectively). Populations excluded from analysis due to low representation included Pacific Islander, Ashkenazi Jewish, Southwest Asian/North African, Indian Asian, and Latina AA women.

Age at diagnosis data were available for 1969 patients. Across the study population, ages ranged from 21 to 96 years, with a median age of 54 years and an average age of 55 years. By racial/ethnic group, median ages were 49 for AAs, 47 for Asians, 56 for NHW, and 52 years for Latina women.

Only a subset of tumors had receptor data available. ER status was available for 1481 tumors, with 47.5% ER+ (n = 704) and 52.5% ER- (n = 777). A smaller subset had additional tumor information. TNBC status was available on 1221 tumors with 36% classified as TNBC (n = 439) and 64% as non-TNBC (n = 782). Data were collected for morphology, grade, and stage, but were not used for analysis due to low availability across the datasets and the high number of categories.

Table 1
Study Population Demographics and Tumor Characteristics

Study Population Demographics	NHW n (%)	AA n (%)	Asian n (%)	Latina n (%)	Total n (%)
Individuals[1]	1931 (65%)	258 (9%)	715 (24%)	60 (2.0%)	2964
Breast Cancer dx age available[2]	1461 (76%)	162 (63%)	286 (40%)	60 (100%)	1969 (66%)
Age range (years)	21–96	24–84	26–90	31–72	21–90
Median age (years)	56	49	47	52	
p-value	[Ref]	< 0.0001	< 0.0001	0.0161	
Tumor Characteristics[2]					
ER+	553 (49%)	54 (36%)	60 (42%)	37 (62%)	704 (47.5%)
ER-	575 (51%)	95 (64%)	84 (58%)	23 (38%)	777 (52.5%)
No ER data	803	109	571	0	1483
Non-TNBC	660 (64%)	47 (51%)	31 (69%)	44 (77%)	782 (64%)
TNBC	366 (36%)	46 (49%)	14 (31%)	13 (23%)	439 (36%)
No TNBC data	905	165	670	3	1743

NHW, non-Hispanic White, AA, African ancestry; n, number; % percentage; ref, reference population; dx, diagnosis

[1] Percentages were calculated for participants in the study for each racial/ethnic group category

[2] Percentages for age at diagnosis and tumor characteristics were calculated only including individuals for whom marker information was available within each racial/ethnic group category

Characteristics of mutations

Fifty-four tumors had more than one *TP53* mutation, for a total of 3024 *TP53* mutations across all 2964 patients (Additional File 3, Additional File 4). A majority of mutations were missense (65%, n = 1972). We identified 829 distinct mutations, including 427 missense, 63 nonsense, 209 frameshift, 58 in-frame insertion/deletions, 2 large insertion/deletions, and 56 unique splicing changes (Additional Files 3 and 4). The most frequent missense mutations were p.R175H (n = 138, 4.7%), p.R248Q (n = 104, 3.4%), p.R273H (n = 87, 2.9%), and p.R248W (n = 73, 2.5%), all known hotspot mutations (Fig. 1). Overall, mutations at reported hotspot sites (R175, G245, R248, R273 and R282) accounted for 20% of mutations (n = 616). More specifically, mutations of CpG nucleotides at hotspots accounted for 17.5% of mutations (n = 530)

For functional classification, tumors were analyzed based on the net effect of all *TP53* mutations per tumor (n = 2964). Overall, we characterized 939 mutations as GOF and 1739 LOF; 286 mutations did not have sufficient information for classification. Evaluating mutation dominant negative activity, the breakdown of mutation types was 1246 DNE+, 1190 DNE-, and 528 uncharacterized.

Association of race, tumor characteristics, and age with mutation type

To determine if there were associations between race and type of mutation, we conducted Fisher's exact test for racial and ethnic ancestry by mutation categories. No significant associations were identified overall for GOF/LOF status (p = 0.15), DNE (p = 0.62), mutation hotspots (p = 0.32) or CpG sites (p = 0.52) and race (Table 2). However, association of GOF/LOF status and race was significant when comparing GOF versus LOF in NHW and AA patients only, with NHW patients more likely to have GOF mutations (35.8% versus 29.2%, respectively, p = 0.04). We additionally tested whether there was any association between ER or TNBC status and mutation type (Table 3). We identified a significant association between DNE and TNBC (p = 0.03) and related ER status (p = 0.005). ER- tumors and TNBCs were less likely to have *TP53* somatic mutations that were DNE+. We did not identify associations between ER and GOF/LOF (p = 0.5076), with mutation hotspots (p = 0.15) or with CpG hotspots (p = 0.24). Interestingly, patients with hotspot mutations were slightly younger on average, with a mean age of 53.6 years versus 55.0 years for patients with non-hotspot mutations, at a level approaching significance (p = 0.065)(Fig. 2). We did not identify significant associations with age and GOF/LOF, with a mean age of 54.5 years for GOF and 55.0 years for LOF (p = 0.52). We also found no significant association between age and DNE; the mean age was 54.5 years for DNE+, and 55.0 years for DNE- (p = 0.49).

Table 2
TP53 Mutation Frequency by Type and Population

	Populations				**Total**
Mutation type	NHW (n = 1931)	Asian (n = 715)	AA (n = 258)	Latina (n = 60)	n = 2964
GOF	623 (32.2%)	222 (31%)	70 (27.1%)	24 (40%)	939 (31.7%)
LOF	1116 (57.8%)	419 (58.6%)	170 (65.9%)	34 (56.7%)	1739 (58.7%)
Unknown function	192 (10%)	74 (10.3%)	18 (7%)	2 (3.3%)	286 (9.6%)
p-value	0.15				
DNE+	829 (43%)	286 (40%)	105 (40.7%)	26 (43.3%)	1246 (42%)
DNE-	767 (39.7%)	280 (39.2%)	116 (45%)	27 (45%)	1190 (40.1%)
Unknown	335 (17.3%)	149 (20.2%)	37 (14.3%)	7 (11.7%)	528 (17.8%)
p-value	0.62				
Hotspot	416 (21.5%)	140 (19.6%)	44 (17%)	13 (21.7%)	613 (20.7%)
Non-hotspot	1515 (78.5%)	575 (80.4%)	214 (83%)	47 (78.3%)	2351 (79.3%)
p-value	0.32				
Hotspot CpG	353 (18.3%)	125 (17.5%)	39 (15.1%)	13 (21.7%)	530 (17.9%)
Non-hotspot CpG	1578 (81.7%)	590 (82.5%)	219 (84.9%)	47 (78.3%)	2434 (82.1%)
p-value	0.52				

n, number; GOF, gain of function mutation; LOF, loss of function mutation; DNE+, dominant negative activity present; DNE-, no dominant negative activity; NHW, non-Hispanic White, AA, African Ancestry

Table 3
TP53 Mutation Frequency by Age and Tumor Characteristics

	Tumor Subtypes				
Mutation type	**ER-** (n = 777)	**ER+** (n = 704)	**TNBC** (n = 439)	**Non-TNBC** (n = 782)	**Mean age**
GOF	216 (27.8%)	216 (30.7%)	126 (28.7%)	230 (29.4%)	54.5
LOF	514 (66.1%)	424 (60.2%)	291 (66.3%)	485 (62%)	55.0
Unknown	47 (6%)	64 (9.1%)	22 (5%)	67 (8.6%)	NA
p-value	0.10		0.51		0.52
DNE+	288 (37.1%)	297 (42.2%)	162 (36.9%)	317 (40.5%)	54.5
DNE-	384 (49.4%)	285 (40.5%)	226 (51.5%)	331 (42.3%)	55.0
Unknown	105 (13.5%)	122 (17.3%)	51 (11.6%)	134 (17.1%)	NA
p-value	0.0045		0.029		0.49
Hotspot	143 (18.4%)	151 (21.4%)	81 (18.5%)	164 (21%)	53.6
Non-hotspot	634 (81.6%)	553 (78.6%)	358 (81.5%)	618 (79%)	55.0
p-value	0.15		0.20		0.065
Hotspot CpG	126 (16.2%)	131 (18.6%)	75 (17.1%)	135 (17.3%)	54.0
Non-hotspot CpG	651 (83.8%)	573 (81.4%)	364 (82.9%)	647 (82.7%)	55.0
p-value	0.24		1.0		0.30

n, number; GOF, gain of function mutation; LOF, loss of function mutation; DNE+, dominant negative activity present; DNE-, no dominant negative activity; ER, Estrogen receptor; TNBC, Triple negative breast cancer

Discussion

The goal of our study was to determine if the type of *TP53* somatic mutation (GOF or LOF, DNE- or DNE+, hotspot status, and CpG nucleotide position) varied in frequency between patients of different ancestry. Considering that the overall rate of somatic *TP53* mutations in breast cancer differs by race (15–18), this

is an important concern for study of *TP53*-mutant breast tumors and differences in outcomes and treatment response by race. We identified a modest difference between AA and NHW individuals, with NHWs slightly more likely to have GOF mutations. DNE- *TP53* somatic mutations were associated with TNBC and ER.

Our finding that *TP53* mutations without DNE activity were associated with TNBC ($p = 0.03$) and ER status ($p = 0.005$) is novel. From previous studies, ER- and TNBC have a higher frequency of *TP53* mutations compared to ER + tumors (13). In this study, that only includes *TP53*-mutant tumors, we observed a higher proportion of ER- and TNBC tumors overall compared to unselected populations. This is consistent with previous studies that identified *TP53* somatic mutations in 40–60% of all breast tumors versus ~ 85% of TNBC (1–3, 5). There has been some debate about the significance of mutant *TP53* DNE versus GOF activity, as many common somatic mutations, including hotspot mutations, are both DNE + and GOF. Overall the importance of DNE + versus GOF is highly dependent on the specific mutation type, genetic and cellular context, loss of heterozygosity, and the phenotype evaluated (31). It is thus of great interest that the association with receptor status was only significant for DNE; there was not a significant association with receptor status or GOF/LOF. Thus mutant *TP53* DNE may be a more important component of tumor subtype, though functional studies are needed to better understand this phenomenon.

Our cohort included somatic *TP53* mutation data from TCGA, METABRIC, and IARC databases, studies identified for inclusion from literature, and 351 previously unpublished cases (Additional File 1). The frequency of hotspot mutations observed in our study (20%) was slightly lower than previous studies finding that 28% of *TP53* mutations occurred at mutation hotspots (7). We observed that 36% of tumors from NHW individuals had GOF mutations compared to 29% in AA individuals ($p = 0.04$). This is opposite of what we expected to find as missense/GOF variants have been associated with poorer prognosis or worse outcomes in previous studies (9). We considered that this effect may be an artifact of more NHW patients sequenced with earlier technology, such as Sanger, which could bias the *TP53* mutation detected to the exons more likely to have GOF mutations. However, there was no difference in use of Sanger vs NGS between these population groups, with 43.6% of NHW patients sequenced with Sanger, compared to 43.8% of AA patients. There also was no notable difference in the number of exons sequenced; 67.3% of NHW patients had at least exons 2–11 sequenced, compared to 68.6% of AA patients. Additionally, there was no difference in the percentage of unclassified variants between groups (7% in AA versus 9.9% in NHW for GOF/LOF, 14.3% in AA versus 17.3% in NHW for DNE). Thus, this difference does not appear to be due to technological differences in mutation detection or in mutation classification. Further studies of larger numbers of AA and NHW women are warranted to confirm this finding.

Participants with hotspot mutations were younger than those with non-hotspot mutations, with a mean age of 53.61 in hotspots versus 55.04 in non-hotspots, but this was not statistically significant ($p = 0.065$). Age did not correlate with DNE or GOF/LOF. This finding is somewhat unexpected. Susceptibility to hotspot mutations is likely due to properties of the genetic sequence being vulnerable to mutation, rather than purely selective growth advantage of tumor cells (7). A high proportion of hotspot mutations

are CpG sites, a feature of mutation signature 1, which correlates with age, so it would seem more likely for somatic hotspot mutations at CpG sites to be associated with later age at diagnosis (8). However, a correlation for breast cancer has not yet been reported in the literature of which we are aware. (1). In Li-Fraumeni syndrome, germline *TP53* missense mutations have been associated with an earlier age of breast cancer diagnosis compared to frameshift and nonsense mutations (32, 33). Furthermore, in one study, age of diagnosis was not associated with hotspot mutations in ovarian cancer (34). We did not have sufficient numbers of individuals with stage information to evaluate whether there was a correlation with age, stage and hotspot status. Data from this study and others suggest that age is not the most critical contributor to *TP53* hotspot mutations at CpG sites, warranting additional study.

Strengths of this study include the large number of women included for study. Previous studies characterizing *TP53* mutation types have not focused on race or ancestry and have relied on TCGA or IARC datasets alone. This study incorporated multiple sources including previously unpublished data. We limited the dataset to only include tumors with *TP53* somatic mutations and only included participants with race or ancestry data. There are a number of limitations to this study. Many of the studies used self-reported race and ethnicity information, which may not reflect genetic ancestry, and may have been categorized differently by study depending on data collection method, such as distinguishing between NHW and Ashkenazi Jewish ethnicity. There may be differences in *TP53* mutation types between ethnic groups within a racial group, such as between individuals of European ancestry from Greece and Finland. For countries that are predominantly one racial group and for which detailed racial information was not available from the study authors, we made an assumption that the individuals in that study were very likely to be of that racial group (e.g. Norway and European ancestry; China and Asian ancestry). Some older studies that used single-strand conformation polymorphism followed by Sanger to identify mutations may have missed some mutations. A small number of studies only performed analyses of exons 4 through 8 which could miss more LOF (splice, nonsense and frameshift) variants that occur in other exons compared to GOF or DNE-associated missense variants that predominantly map to these exons. Because of the mixed data sources, this study did not include large copy number losses or exon specific deletions as variants for study resulting in fewer *TP53* LOF variants being included. As mutation signatures resulting in specific mutation types could have a basis in biological or environmental differences between races, we may have missed racial specific differences in mutations resulting in large DNA gains or losses. Finally, classifications of variants as GOF/ LOF and DNE were made based on studies in the literature. For some variants there was discordant information; for this study, we used the classifications from studies that were larger (tested more variants) or studies that included a larger number of different assays. It is possible that some of the rarer missense variants included were misclassified or may act differently in humans than in the system tested (e.g. yeast).

Conclusions

In this large study based on data from the literature and databases, we found that somatic *TP53* mutation types did not differ by race overall, but GOF mutations were more common in NHW women when compared to AA women. We uncovered a modest association between DNE- and ER- and triple

negative tumor receptor status. Functional studies are needed to understand this phenomenon. Additional tumor sequencing data from more racially diverse cohorts are needed to follow-up on these findings.

Abbreviations

AA
African ancestry
BLBC
Basal-like breast cancer
DNE
Dominant negative
ER
Estrogen receptor
GOF
Gain of function
LOF
Loss of function
NHW
non-Hispanic White
PR
Progesterone receptor
TCC
Total Cancer Care
TNBC
Triple negative breast cancer
TP53
Tumor protein 53

Declarations

Acknowledgements:

We thank Lara Sucheston-Campbell, Philip Tsichlis, and Michael Freitas for the initial ideas leading to this study. The Ohio State University Comprehensive Cancer Center (OSU CCC) Total Cancer Care from the Biospecimen Services Shared Resource provided mutational data on breast cancer cases. The results published here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Authors' contributions

NP, JR, and AET conceived the study idea. NP and JR performed literature and database reviews. NP curated *TP53* mutation status. PS downloaded data from publicly available databases. JM performed statistical analyses. HH, MT, KC, JJH, JLF, OIO, DH, EZ, and SN provided data for this study. NP, JR, and AET wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by R01 CA215151. The OSU CCC TCC is supported in part by National Cancer Institute grant P30 CA016058. JR was supported by a Pelotonia Fellowship. KCD was supported by The UNC Breast Cancer SPORE Grant CA58223 from the National Cancer Institute. OIO and DH were supported by grants from National Institutes of Health (P20 CA233307, R01 MD013452) and Breast Cancer Research Foundation. For the City of Hope Hispanic breast cancer cases: sample collection and data were collected under support from NIH R01 CA184585 (SLN and EZ); tumor sequencing was supported by the National Institute on Minority Health and Health Disparities; and tumor blocks were pulled and slides cut in the Pathology Lab Core supported by the National Cancer Institute of the National Institutes of Health under award number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data:

All data are available in this manuscript and Additional materials (Files 1-4) or in the original publications (Additional File 5).

Ethics Approval and Consent to Participate:

This study was approved by the OSU Cancer Institutional Review Board. Participants from the OSU CCC Total Cancer Care provided informed consent for this study. The City of Hope Institutional Review Board and the University of Chicago IRB approved study for participants enrolled at their respective sites. The remainder of data for this study was existing de-identified data or available from previous publications.

Consent for publication:

Not applicable.

Conflicts of Interest:

HH is on the scientific advisory board for Invitae Genetics, Promega, and Genome Medical and has stock/stock options in Genome Medical and GI OnDemand. None of these are direct conflicts with this study of somatic *TP53* mutations in breast cancer.

References

1. Olivier M, Langerød A, Carrieri P, Bergh J, Klaar S, Eyfjord J, et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. Clin Cancer Res. 2006.
2. Pitt JJ, Riester M, Zheng Y, Yoshimatsu TF, Sanni A, Oluwasola O, et al. Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. Nat Commun. 2018.
3. Zhang G, Wang Y, Chen B, Guo L, Cao L, Ren C, et al. Characterization of frequently mutated cancer genes in Chinese breast tumors: a comparison of Chinese and TCGA cohorts. Ann Transl Med. 2019.
4. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A. 2005.
5. Prat A, Adamo B, Cheang MCU, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. Oncologist [Internet]. 2013/02/12. 2013;18(2):123–33. Available from: https://pubmed.ncbi.nlm.nih.gov/23404817.
6. Leroy B, Fournier JL, Ishioka C, Monti P, Inga A, Fronza G, et al. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. Nucleic Acids Res. 2013 Jan;41(Database issue):D962-9.
7. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death and Differentiation. 2018.
8. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015 Dec;47(12):1402–7.
9. Yamamoto S, Iwakuma T. Regulators of Oncogenic Mutant TP53 Gain of Function. Cancers (Basel). 2018 Dec;11(1).
10. Alexandrova EM, Mirza SA, Xu S, Schulz-Heddergott R, Marchenko ND, Moll UM. p53 loss-of-heterozygosity is a necessary prerequisite for mutant p53 stabilization and gain-of-function in vivo. Cell Death Dis. 2017 Mar;8(3):e2661.
11. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet [Internet]. 2018;50(10):1381–7. Available from: https://doi.org/10.1038/s41588-018-0204-y.
12. Billant O, Léon A, Le Guellec S, Friocourt G, Blondel M, Voisset C. The dominant-negative interplay between p53, p63 and p73: A family affair. Oncotarget. 2016 Oct;7(43):69549–64.
13. Scott LC, Mobley LR, Kuo T-M, Il'yasova D. Update on triple-negative breast cancer disparities for the United States: A population-based study from the United States Cancer Statistics database, 2010 through 2014. Cancer. 2019 Oct;125(19):3412–7.
14. Toft DJ, Cryns VL. Minireview. Basal-like breast cancer: from molecular profiles to targeted therapies. Mol Endocrinol. 2011 Feb;25(2):199–211.
15. Dumay A, Feugeas JP, Wittmer E, Lehmann-Che J, Bertheau P, Espié M, et al. Distinct tumor protein p53 mutants in breast cancer subgroups. Int J Cancer. 2013.

16. Newman LA, Stark A, Chitale D, Pepe M, Longton G, Worsham MJ, et al. Association Between Benign Breast Disease in African American and White American Women and Subsequent Triple-Negative Breast Cancer. JAMA Oncol. 2017 Aug;3(8):1102–6.
17. Sineshaw HM, Gaudet M, Ward EM, Flanders WD, Desantis C, Lin CC, et al. Association of race/ethnicity, socioeconomic status, and breast cancer subtypes in the National Cancer Data Base (2010–2011). Breast Cancer Res Treat. 2014 Jun;145(3):753–63.
18. Martínez ME, Gomez SL, Tao L, Cress R, Rodriguez D, Unkart J, et al. Contribution of clinical and socioeconomic factors to differences in breast cancer subtype and mortality between Hispanic and non-Hispanic white women. Breast Cancer Res Treat. 2017 Nov;166(1):185–93.
19. Sturtz LA, Melley J, Mamula K, Shriver CD, Ellsworth RE. Outcome disparities in African American women with triple negative breast cancer: a comparison of epidemiological and molecular factors between African American and Caucasian women with triple negative breast cancer. BMC Cancer. 2014 Feb;14:62.
20. Hines LM, Risendal B, Byers T, Mengshol S, Lowery J, Singh M. Ethnic disparities in breast tumor phenotypic subtypes in Hispanic and non-Hispanic white women. J Womens Health (Larchmt). 2011 Oct;20(10):1543–50.
21. Doepker MP, Holt SD, Durkin MW, Chu CH, Nottingham JM. Triple-Negative Breast Cancer: A Comparison of Race and Survival. Am Surg. 2018 Jun;84(6):881–8.
22. Troester MA, Sun X, Allott EH, Geradts J, Cohen SM, Tse C-K, et al. Racial Differences in PAM50 Subtypes in the Carolina Breast Cancer Study. J Natl Cancer Inst. 2018 Feb;110(2):176–82.
23. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA Cancer J Clin. 2016;66(1):31–42.
24. Keenan T, Moy B, Mroz EA, Ross K, Niemierko A, Rocco JW, et al. Comparison of the genomic landscape between primary breast cancer in African American versus white women and the association of racial differences with tumor recurrence. J Clin Oncol. 2015.
25. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, et al. Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. JAMA Oncol. 2017.
26. Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. Cancer Cell. 2020 May;37(5):639–54.e6.
27. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. Hum Mutat. 2016.
28. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016 May;7:11479.

29. Rueda OM, Sammut S-J, Seoane JA, Chin S-F, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. Nature. 2019 Mar;567(7748):399–404.
30. R Core Team. (2020). R: A language and environment for statistical computing. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020.
31. Stein Y, Aloni-Grinstein R, Rotter V. Mutant p53 oncogenicity – dominant-negative or gain-of-function? Carcinogenesis. 2020.
32. Bougeard G, Sesboüé R, Baert-Desurmont S, Vasseur S, Martin C, Tinat J, et al. Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. Vol. 45, Journal of medical genetics. England; 2008. p. 535–8.
33. Zerdoumi Y, Aury-Landas J, Bonaïti-Pellié C, Derambure C, Sesboüé R, Renaux-Petel M, et al. Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. Hum Mutat. 2013 Mar;34(3):453–61.
34. Tuna M, Ju Z, Yoshihara K, Amos CI, Tanyi JL, Mills GB. Clinical relevance of TP53 hotspot mutations in high-grade serous ovarian cancers. Br J Cancer [Internet]. 2020;122(3):405–12. Available from: https://doi.org/10.1038/s41416-019-0654-8.

Figures

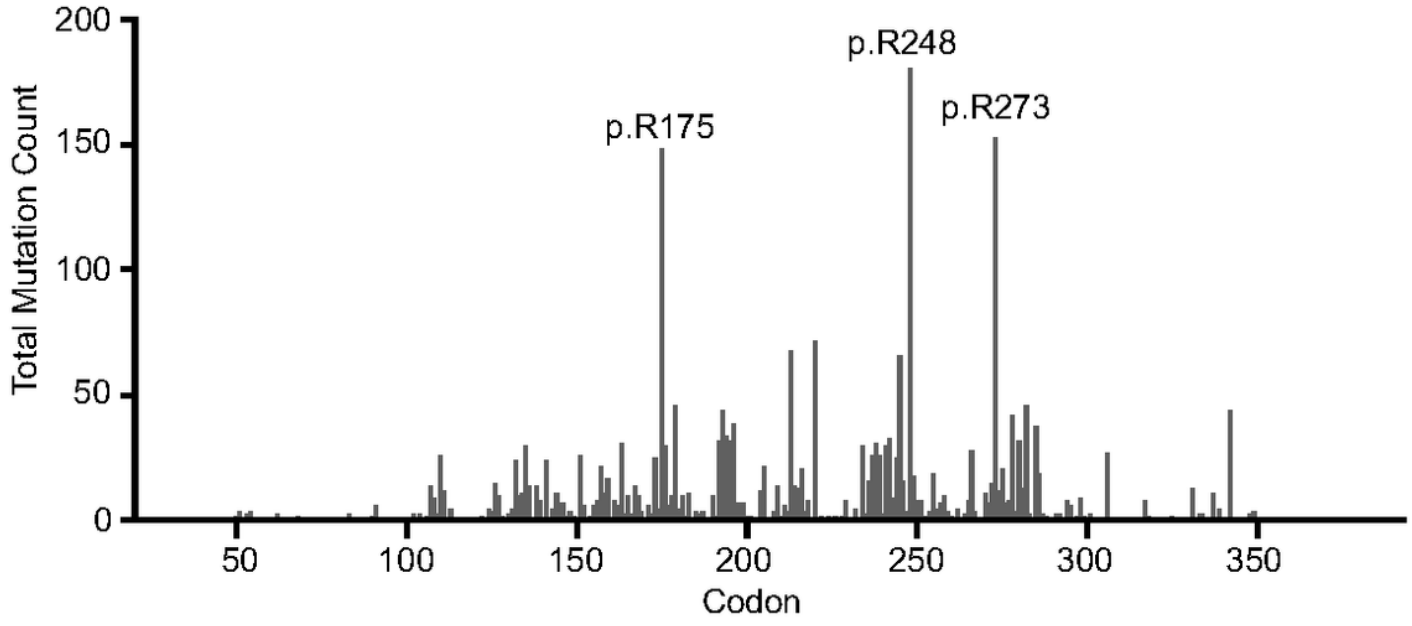

Figure 1

Mutation frequency by codon The frequency of somatic TP53 mutations by codon is plotted. Intronic mutations affecting splicing are not included.

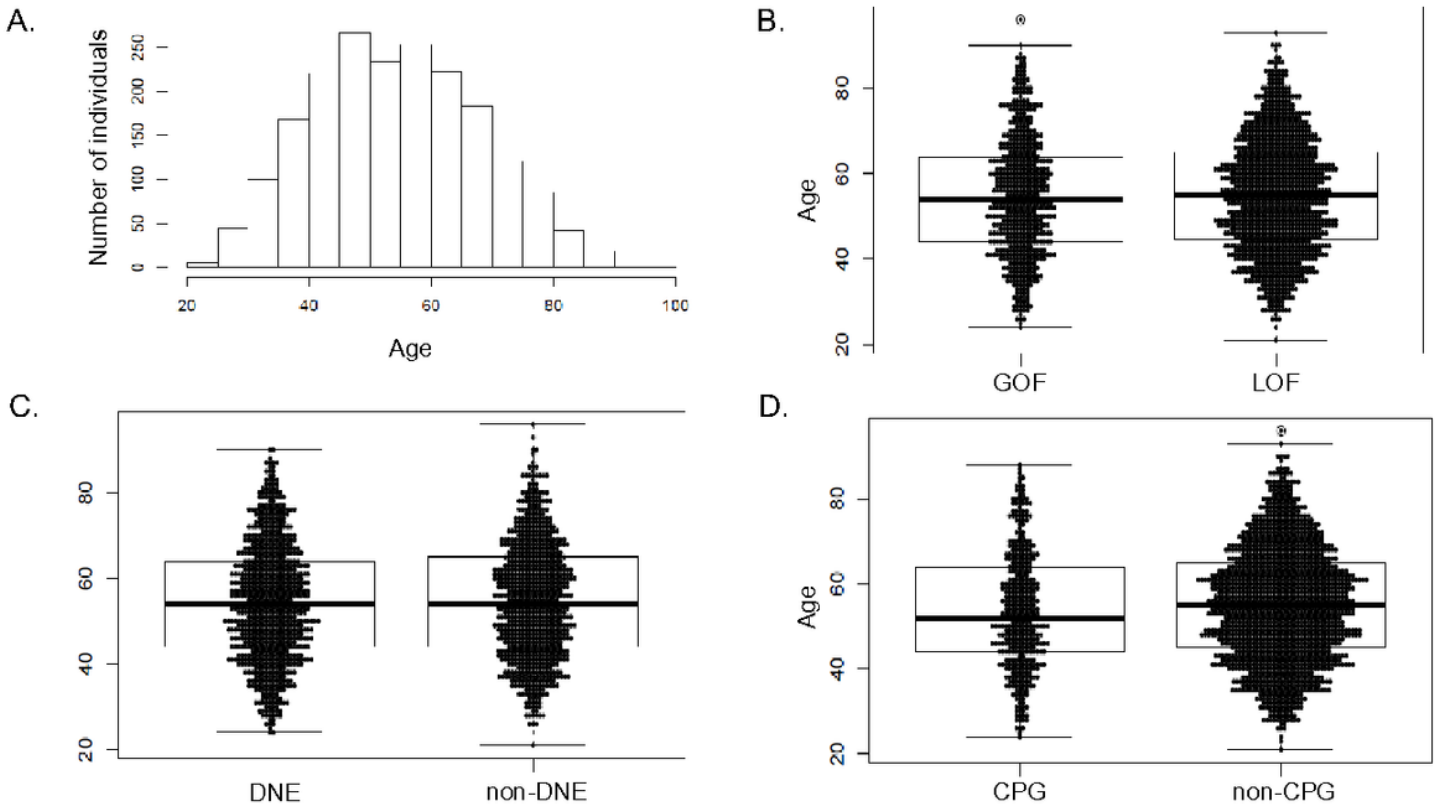

Figure 2

Association of age of breast cancer diagnosis with TP53 mutation characteristics The frequency of age of breast cancer diagnosis of all individuals included in the study was plotted by histogram (A). Age of diagnosis was not significantly associated with TP53 GOF versus LOF, p-value =0.5 (B) or dominant negative effect, p-value 0.49 (C). Individuals with a TP53 mutation at a CpG hotspot were slightly, but not significantly younger at age of diagnosis (53.6 years versus 55.0 years), p-value 0.065 (D). P-values were calculated using a Welch two sample T-test. P-values < 0.05 were considered significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalInformationFile1TableListofStudies.xlsx
- AdditionalInformationFile2TableOSUTCCParticipants.pdf
- AdditionalInformationFile3Table3.pdf
- AdditionalInformationFile4Table4.pdf
- AdditionalInformationFile5SupplementaryReferences.pdf