An elevated rate of whole-genome duplication events in cancers from Black patients

Leanne M. Brown[1]*, Ryan A. Hagenson[1]*, Jason M. Sheltzer[1†]

1. Yale University School of Medicine, New Haven, CT 06511, USA

*Equal contribution

† Corresponding author. Email: jason.sheltzer@yale.edu

ABSTRACT

In the United States, Black individuals have higher rates of cancer mortality than any other racial or ethnic group. The sources of these significant racial disparities are not fully understood, and may include social, environmental, and genetic factors that influence cancer onset, diagnosis, and treatment. Here, we examined genomic data from several large-scale cancer patient cohorts to search for racial associations in chromosome copy number alterations. We found that tumors from Black patients were significantly more likely to exhibit whole-genome duplications (WGDs), a genomic event that enhances metastasis and aggressive disease, compared to tumors from white patients. Among patients with WGD-positive cancers, there was no significant difference in survival between Black and white patients, suggesting that the increased incidence of WGD events could contribute to the disparities in patient outcome. Genomic analysis identified several somatic alterations associated with WGD events that were consistent between Black and white populations, indicating that the increase in WGD events may be driven by environmental or epigenetic factors in Black patients. In total, these findings identify a class of genomic alterations that may influence racial disparities in cancer patient outcome. As cancers that have undergone WGD events exhibit unique genetic vulnerabilities, therapies that selectively target WGD-positive cancers may be particularly effective at treating aggressive malignancies in Black patients.

INTRODUCTION

The unequal landscape of healthcare outcomes in the United States, particularly among Black patients, demands immediate attention. Black patients face higher rates of chronic disease, higher incidence of preventable illness, and higher all-cause mortality compared to white patients[1–4]. These disparities are particularly evident within the field of oncology. For many cancer types, Black patients are diagnosed with more advanced or aggressive cancers than white patients[5–7]. Although cancer outcomes have drastically improved over the past several decades, Black patients continue to have higher rates of cancer-related death[8–10]. Mortality disparities are exacerbated within certain cancer types, including breast and endometrial cancers, in which Black patients exhibit 41% and 21% higher mortality, respectively, compared to white patients[11–13]. Uncovering the mechanisms that mediate these unequal outcomes could help identify prevention or treatment strategies that may aid those populations most at risk.

The source of racial disparities in cancer outcomes is at present unresolved and may result from social, environmental, and genetic factors[14–20]. In the United States, Black individuals are more than twice as likely to live below the federal poverty line than white individuals, and lower socioeconomic status has been linked with decreased healthcare access and lower cancer screening rates[21–25]. In addition, people within economically-deprived communities have disproportionately higher rates of environmental carcinogen exposure, which may increase the risk of cancer development[25,26]. Within the healthcare system itself, Black patients received worse care compared to white patients in more than half of quality care metrics in the 2022 National Healthcare Quality and Disparity Report[27]. These findings extend to cancer-specific interventions, as Black patients experience more delays in chemotherapy induction for breast cancer treatment and are less likely to have adequate oncological resection for gastrointestinal cancers treated with curative intent surgery[28,29].

In addition to these social and environmental influences on cancer mortality, recent research has raised the possibility that genetic differences between patient populations could affect the development and/or progression of cancer. To date, most population-based cancer profiling efforts have sought to investigate variability in the prevalence of somatic point mutations between racial and ethnic groups[30–34]. These studies have uncovered certain differences in the frequency of mutations in common oncogenes and tumor suppressors that could impact disease pathogenesis. For example, Black patients have been found to exhibit higher rates of mutations in the tumor suppressor *TP53* compared to white patients, and *TP53* inactivation has consistently been linked with poor prognosis[35–40]. Similarly, Black patients with lung cancer have fewer mutations in the druggable oncogene *EGFR*, which may affect treatment options[41]. Uncovering population-based differences in mutation profiles can aid in clinical assessment and may shed light on strategies to ameliorate outcome disparities.

Comparatively less is known about population-based associations for other types of somatic alterations in cancer beyond single-nucleotide point (SNP) mutations. Notably, chromosomal copy number alterations (CNAs) are pervasive across tumor types and have been linked with disease progression, drug resistance, and poor patient outcomes[42–45]. One recent study examined the prevalence of arm-scale CNAs across a large cohort of cancer patients and found few significant differences between Black and white populations[46]. Population-based differences in many other classes of CNAs have not been previously investigated. One common type of chromosomal alteration is a whole-genome duplication (WGD) event, in which a cell's chromosome complement doubles. The causes and consequences of WGDs are poorly understood; mutations in *TP53* have been linked with WGD development but other genetic and environmental causes remain obscure[43,47,48]. WGDs enhance tumor adaptability and increase metastatic dissemination, potentially by enhancing tumor heterogeneity and

allowing cancers to sample a wider range of karyotypes[47,49–51]. Approximately 30% of tumors exhibit WGDs, but whether patient race or ethnicity is associated with these events is unknown[47].

RESULTS

Tumors from Black patients display an increased frequency of WGDs.

WGD events in cancer are associated with genomic instability and aggressive disease (Fig. 1A)[47,52]. We investigated the frequency of WGD events in cancers from Black and white patients from three different patient cohorts: the MSK-MET cohort (n=13,071 patients), The Cancer Genome Atlas (TCGA) (n=8,060 patients), and the Pan-cancer Analysis of Whole Genomes (PCAWG) (n=1,963 patients)[53–55]. We discovered that cancers from Black patients exhibited a significantly higher incidence of WGDs compared to cancers from white patients in each cohort (Fig. 1B-D). For the MSK-MET and TCGA cohorts, we conducted this analysis based on individuals self-reporting their race, while for the PCAWG dataset self-reported race was unavailable and we used inferred genetic ancestry instead (discussed in more detail below). Notably, CNAs in each of these datasets were detected using different genomic technologies: SNP arrays, targeted gene sequencing, and whole-genome sequencing, for TCGA, MSK-MET, and PCAWG, respectively[53–55]. As these findings were consistent across all three cohorts, we anticipate that they are robust and independent of any platform-specific artifacts.

The increase in WGD events in Black patients compared to white patients ranged from 11% in the TCGA cohort to 34% in the PCAWG cohort. Additionally, the overall rate of WGD events ranged from 26% in MSK-MET to 31% in PCAWG. However, within cancer types that were shared across datasets, the frequency of WGD events was highly correlated (R^2=0.79, $p<0.001$) (Fig. S1). For instance, thyroid cancers consistently displayed the lowest incidence of WGDs (0-4%) while ovarian cancers consistently displayed the highest incidence of WGDs (55-60%), in alignment with other reports[43,47]. The overall differences in WGD frequency between cohorts may reflect differences in the distribution of cancer types.

Next, we investigated whether WGD events were elevated in cancers in patients who identified as Asian. However, in the TCGA and MSK-MET cohorts, we did not detect a significant difference in WGD frequency between white and Asian patients (Fig. S2A). Similarly, tumors from men and women displayed equivalent frequencies of WGD events within each cohort (Fig. S2B).

Tumors from Black patients with specific cancer types display an increased frequency of WGDs.

We considered the possibility that the increased incidence of WGD events in our pan-cancer analysis of Black patients could result from differences in the representation of distinct cancer types between populations. However, Black patients have historically been underrepresented in genomic studies, which limits our statistical power to detect significant differences in every cancer lineage[56,57]. For that reason, we focused our analysis on cancer types for which we had genomic data from at least 60 Black patients in both the MSK-MET and TCGA datasets: breast cancer, endometrial cancer, and non-small cell lung cancer (NSCLC). In both cohorts, we detected a significant increase in WGD events in Black patients with breast and endometrial cancers. For NSCLC, we detected a significant increase in the MSK-MET cohort but not the TCGA cohort. The increase ranged from 32% in breast cancer (TCGA) to 202% in endometrial cancer (MSK-MET) (Fig. 1E). These results indicate that Black patients have a significantly higher incidence of WGD events both across cancers and within individual cancer types.

Analysis of WGD frequency by tumor stage and histological subtype.

WGD abundance varies between histological cancer subtypes and is more common in advanced malignancies[47]. Accordingly, we considered the possibility that differences in the prevalence of histological subtypes or differences in the stage at which tumors were diagnosed could produce the increase in WGD events in Black cancer patients that we observed. However, we determined that WGD events were still significantly more common in tumors from Black patients for many individual cancer stages and subtypes (Fig. S3-4). For instance, Black patients with either stage II or III breast tumors had a higher incidence of WGD events compared to white patients with similarly-staged tumors (Fig. S3B). Additionally, while the frequency of histological subtypes varied between Black and white patients, when we limited our analysis to only the most common histological subtypes of each cancer (breast cancer: invasive ductal carcinoma; endometrial cancer: endometrioid tumors, and NSCLC: adenocarcinoma), we still detected an increased frequency of WGD events among Black patients (Fig. S4). These results indicate that the increased incidence of WGDs in Black cancer patients is not simply due to differences in tumor stage or histological subtype at diagnosis.

Analysis of WGDs by Inferred Genetic Ancestry.

Recent studies have urged caution in the use of race when conducting population-based research, noting that widely used definitions of race represent artificial social constructs[58,59]. Instead, shared genetic ancestry may better reflect population differences in disease risk. Nonetheless, race may still be useful for investigating patterns of heath and disease in the US due to its association with the social determinants of health, including poverty, pollution, systemic racism, and lack of healthcare access[60]. According to the National Academy of Sciences' 2023 report *Using Population Descriptors in Genetics and Genomics Research*, "race… may be a useful population descriptor for researchers who wish to measure a consequential form of social status and affiliation… [R]ace may be a proxy for the experience of racism in health disparities studies." As described above, we found that both individuals who self-reported as Black or African American and individuals with African ancestry were more likely to exhibit WGD-positive cancers compared to individuals who self-reported as white or individuals with European ancestry (Fig. 1B-E). To extend this investigation, we re-analyzed the MSK-MET dataset using inferred genetic ancestry instead of racial self-reporting[34]. We found that the overall patterns of WGD were conserved: African ancestry was associated with a 16% increase in WGD events in a pan-cancer analysis and with a 32-135% increase in breast, endometrial, and NSCLC (Fig. S5A-B). We conclude that WGD events are associated with both African ancestry and Black identity.

Association of WGD events with TP53 mutations

We sought to uncover the somatic alterations associated with WGD events in Black and white patients. For this and subsequent analyses, we focused on the MSK-MET cohort, as this was the largest single patient cohort and had the best sequencing coverage of cancer-relevant genes. We constructed a logistic regression model linking somatic alterations with WGD status while correcting for cancer type (Fig. 2A and Table S1A). Consistent with previous results, the strongest predictor of WGD status was the presence of mutations in *TP53*[47]. In contrast, mutations in *KRAS*, *BRAF*, and *PTEN* were significantly associated with a reduced likelihood of WGD events. We repeated this analysis for Black and white patients, considered separately (Fig. 2B-C, Table S1B-C). In Black patients, the only significant features associated with WGD status with a q-value <0.1 were *TP53* mutations and amplifications of cyclin E (*CCNE1*). Cyclin E gains have recently been identified as a driver of WGDs, and both features were also significantly associated with WGD events in our logistic regression model of tumors from white patients (Fig. 2C)[48]. These results suggest that similar somatic alterations drive WGDs in both Black and white patients.

As the types of somatic alterations associated with WGD events in Black and white patients were similar, we considered the possibility that differences in frequencies of these alterations could influence the prevalence of WGDs. Consistent with previous observations, we found that Black patients were significantly more likely to harbor *TP53* mutations than white patients in both a pan-cancer analysis and in breast and endometrial cancers, but not NSCLC (Fig. 2D-E)[35–40]. Next, we separated patients based on *TP53* status. We observed that there was no significant difference in WGD frequency between Black and white patients with *TP53*-WT cancers. However, among patients with *TP53*-mutant cancers, Black patients showed a trend towards increased WGD events in a pan-cancer analysis (p=0.067) and a significant increase in the NSCLC cohort (p=0.011) (Fig 2F-G). These results suggest that the increased frequency of *TP53* mutations contributes to but cannot fully account for the increased incidence of WGD events in Black patients.

Next, we investigated whether different classes of *TP53* mutations could underlie the different prevalence of WGD events within Black and white populations. However, the overall distribution of classes of *TP53* mutations (e.g., missense vs nonsense) was similar between Black and white patients (Fig 2H-I). Additionally, the same set of p53-inactivating point mutations, including the R175, R248, and R273 mutations, were commonly observed among both Black and white patients (Fig. 2J). We conclude that somatic genetic alterations are unable to fully account for the increased prevalence of WGD events among Black patients, and this difference may alternatively be a consequence of epigenetic or environmental factors.

Association between Race, WGD Status, and Patient Outcome

We sought to determine whether the increased incidence of WGD events in Black patients was linked with racial disparities in patient outcome. Consistent with previous observations, we found that Black cancer patients exhibited a significantly shorter overall survival time following diagnosis compared to white patients (Fig. 3A)[8–10]. Similarly, WGD events were also associated with worse patient outcomes (Fig. 3B). Interestingly, among the subset of patients with WGD-positive tumors, there was no significant difference in survival time between Black and white patients (Fig. 3C). However, within the WGD-negative patient subset, a significant difference in survival between Black and white patients was observed (Fig. 3D). These results are consistent with prior reports demonstrating that WGD events drive metastases and aggressive disease, and suggest that the increased incidence of WGD events in Black patients may be linked with worse overall survival[45,53,61]. We speculate that within WGD-negative patients, WGD events may be occurring post-diagnosis, or additional genetic and environmental factors may be contributing to these disparate outcomes.

We repeated this analysis within the MSK-MET breast, endometrial, and NSCLC cohorts. We observed that Black patients with endometrial cancer, but not breast or NSCLC, had significantly shorter survival times compared to white patients within this study (Fig. S6A-C). Racial disparities within breast and NSCLC outcomes have been reported in other analyses, and we speculate that the MSK-MET cohort sizes may not be sufficiently large to detect survival differences that are moderate overall[11–13]. Within endometrial cancer, we observed that patients with WGD-positive tumors had worse outcomes (Fig. S6D). Consistent with our pan-cancer analysis, within the subset of tumors that were WGD-positive there was no significant difference between Black and white patient outcomes, while a significant difference remained apparent within the WGD-negative subset (Fig S4E-F).

Finally, we further divided the pan-cancer survival analysis based on *TP53* status. As expected, *TP53*-mutant tumors had significantly worse overall survival (Fig. S7A). In general, further subdividing patients based on *TP53* status, in addition to race and WGD status, minimized race-based differences in survival time (Fig. S7B-I). In

total, these results suggest that the increased incidence of *TP53* mutations and WGD events contribute to the worse overall outcomes for Black cancer patients, although our findings do not rule out the influence of additional social, environmental, and genetic factors.

Race and WGD status are associated with metastases to the same anatomic sites

Next, we used additional patient information that was collected as part of the MSK-MET dataset to further explore the clinical correlates of race and WGD status. Within this cohort, Black patients were diagnosed and underwent surgery at earlier ages compared to white patients, and Black patients were also more likely to die younger (Fig. 4A). Similarly, WGD-positive tumors were associated with younger diagnoses, younger age at surgery, and younger death in all patients, regardless of race (Fig. 4B). Next, we examined microsatellite instability in each patient, which is a genomic state characterized by the accumulation of point mutations in repetitive sequences[62,63]. We found that tumors from Black patients were significantly less likely to exhibit high microsatellite instability (MSI-H) compared to tumors from white patients, and MSI-H status was also less common in WGD-positive tumors (Fig. 4C-D). MSI-H status has been linked with favorable outcomes, and the under-representation of MSI-H cancers among tumors from both Black patients and patients with WGD-positive disease could represent another factor that contributes to differences in patient survival[64,65].

Finally, we compared the frequency of metastatic dissemination to different anatomic sites between patient populations. Consistent with established disparities in overall outcomes, we observed that Black cancer patients had a greater incidence of metastatic disease compared to white patients. Notably, we found that Black patients had significantly higher rates of metastases to regional lymph nodes, distant lymph nodes, intra-abdominal space, the male and female genitourinary system, and skin (Fig. 4E). Intriguingly, metastases to each of these sites except the male genitourinary system was also more common among WGD-positive tumors compared to WGD-negative tumors (Fig. 4F). In total, these findings suggest that several differences in the clinical presentation of cancers in Black and white populations could be related to the increased incidence of WGDs among Black patients.

DISCUSSION

In this work, we found that Black or African American cancer patients exhibited a significantly greater incidence of WGD events compared to white cancer patients. This discrepancy was detectable in both a pan-cancer analysis and in several individual cancer subtypes. Historically, most research on genetic differences between cancer patient populations has focused on single-nucleotide point mutations; our work demonstrates the existence of significant, outcome-associated differences in patterns of chromosomal alterations as well[30–33]. We speculate that analyzing other types of genetic or epigenetic alterations in cancer (e.g., methylation patterns, smaller CNAs, intratumoral heterogeneity, etc.) may reveal additional informative differences.

Our work is consistent with a previous report that documented an increased incidence of WGD events among prostate cancer patients in sub-Saharan Africa[66]. Interestingly, the cancer types in which Black patients exhibit frequent WGD events (breast, endometrial, prostate, and NSCLC) are also among those that have been recognized as exhibiting the most significant racial disparities in patient incidence or outcome[13,19,20,33]. We found that, among WGD-positive cancer patients, there were no differences in overall survival times, suggesting that WGDs may represent one mechanism underlying the disparate outcomes that have been demonstrated within the American healthcare system.

Our genetic analysis of tumor sequencing data revealed a strong association between *TP53* mutations and WGD events in both Black and white patients. Consistent with previous results, we found that Black patients exhibited a higher incidence of *TP53* mutations overall[35,36,38]. However, WGD events were still more common among Black patients with *TP53*-mutant cancers compared to white patients. Our work did not identify any significant differences in the genetic drivers of WGDs associated with patient race, suggesting that these events may be influenced by epigenetic or environmental factors. For instance, African Americans in the US are more likely to live in areas with higher rates of carcinogenic air pollution, which could affect the development of WGDs in lung cancer[25,26]. Additional work will be required to explore the source(s) of the disparate rates of WGDs between patient populations.

Finally, the increase in WGD events among Black patients that we have documented has the potential to influence patient staging and treatment. We found that WGD-positive tumors were more likely to have spread to regional or distant lymph nodes, which may warrant additional surveillance and interventions among Black patients. More broadly, chromosomal alterations are strongly associated with patient outcome, and analyzing tumor karyotypes as part of a standard pathological workup may improve our ability to preemptively detect aggressive disease[44,67–71]. Additionally, recent research has demonstrated that WGD-positive cancer cells harbor unique genetic vulnerabilities. For instance, alterations in the mitotic apparatus resulting from WGD events cause cells to become dependent on the mitotic kinesin KIF18A, which is otherwise dispensable in diploid cells[72,73]. AMG650 is a small molecular inhibitor of KIF18A that has entered Phase I clinical trials, underscoring the recent progress toward selectively targeting WGD-positive cancers[74]. Therapies designed to selectively target WGD-positive tumors may be particularly effective in Black cancer patients and could serve to ameliorate the disparate racial outcomes in cancer mortality.

MATERIALS AND METHODS

Data Acquisition

Mutational, copy number, sample, and patient data were downloaded from the cBioPortal datahub (https://github.com/cBioPortal/datahub)[75,76]. Whole genome duplication determinations were sourced from the original cohort study (MSK-MET and PCAWG) or from subsequent analyses (TCGA)[43,47,55]. A sample was considered to have undergone WGD based on processing by FACETS (MSK-MET), consensus across 10 different methods (PCAWG), or ABSOLUTE (TCGA)[55,77,78]. Genetic ancestry determinations for the MSK-MET cohort were graciously provided by Kanika Arora, following the method from Arora, et al. (2022)[34]. Regional lymph node metastasis data were sourced from Nguyen, et al. (2022)[53].

Data Harmonization and Cleaning

Data was harmonized across cohorts by including only primary lesions with known WGD status within sample data, considering only genomic aberrations for genes found in the IMPACT-505 geneset within mutational and copy number data, and maintaining patients with known sex, and self-reported race and/or inferred ancestry within patient data, depending on availability. In MSK-MET, patients with self-reported race of Asian, Black, or white and inferred ancestry of either African or European were maintained separately; in TCGA, patients with self-reported race of Asian, Black, or white were maintained; in PCAWG, patients with inferred ancestry of African or European were maintained. To be included in our analysis, genomic aberrations had to occur in ≥200 patients across racial groups and ≥2% of patients within a racial group. We focused mutational analysis on

nonsynonymous mutations and trichotomized copy number levels to neutral, loss, and gain. As the cohorts we analyzed do not use the same identifiers for cancer type, we considered all cancer types regardless of identifier in pan-cancer analyses, while standardizing across cohorts for comparisons by cancer type. Standardization was as follows: breast cancer were samples listed as BRCA (TCGA) or Breast Cancer (MSK-MET); endometrial cancer were samples listed as UCEC (TCGA), Endometrioid Adenocarcinoma (MSK-MET), Serous Carcinoma (MSK-MET); non-small cell lung cancer (NSCLC) were samples listed as LUAD (TCGA), LUSC (TCGA), or Non-Small Cell Lung Cancer (MSK-MET). A positive determination for regional lymph node metastasis was made according to the presence of "regional_lymph" in the "met_site_mapped" variable from Nguyen, et al. (2022)[53]. The code used to perform this analysis is available at https://github.com/sheltzer-lab/wgd_disparities.

Software Versions

Analyses were performed using R (version 4.2.2)[79], PRISM (version 9.4.1), and Python (version 3.10.8)[80]. R packages: car (version 3.1-2)[81], EnhancedVolcano (version 1.16.0)[82], ggplot2 (version 3.4.2)[83], gt (version 0.9.0)[84], gtsummary (version 1.7.2)[85], maftools (version 2.14.0)[86], openxlsx (version 4.2.5.2)[87], reshape (version 1.4.4)[88], tidyverse (version 2.0.0)[89]. Python packages: lifelines (version 0.27.4)[90], matplotlib (version 3.6.2)[91], and pandas (version 1.5.2)[92].

Whole-genome Duplication Frequency Analysis of MSK-MET, TCGA, PCAWG

The frequency of WGD was calculated as the number of WGD-positive samples over the total sample count within a self-reported racial group, inferred ancestry, or self-reported sex at a pan-cancer level and additional subset by cancer type; testing for association was done via two-sided Pearson's Chi-squared test. WGD frequencies by cancer type were compared between MSK-MET and TCGA via Pearson correlation.

Whole-genome Duplication Frequency Analysis by Metastatic Status and Cancer Stage

The frequency of WGD was calculated as the number of WGD-positive samples over the total sample count within a self-reported racial group by metastatic status (MSK-MET) and stage (TCGA) for each cancer type: testing for association was done via two-sided Pearson's Chi-squared test. Available staging information in TCGA was utilized which includes pathological staging for breast cancer and NSCLC and clinical staging for endometrial cancer. Metastatic status and cancer staging distributions were summarized within each racial group by cancer type; testing for association was done via two-sided Pearson's Chi-squared test and Fisher's exact test across racial groups.

Whole-genome Duplication Frequency Analysis by Histological Subtype

Shared cancer types between MSK-MET and TCGA were aggregated. For breast cancer, IDC includes MSK-MET histological designations HR+/HER2+ Ductal Carcinoma, HR+/HER2- Ductal Carcinoma, HR-/HER2+ Ductal Carcinoma, Ductal triple negative breast cancer (TNBC) and TCGA histological designations Ductal Luminal A, Ductal Luminal B, Ductal HER2-enriched, Ductal Basal-like, Ductal Normal-like. ILC includes MSK-MET histological designations HR+ Lobular Carcinoma and TCGA histological designations Lobular Luminal A, Lobular Luminal B, Lobular HER2-enriched, Lobular Basal-like, Lobular Normal-like. For endometrial cancer, shared cancer types between MSK-MET and TCGA included endometrioid and serous subtypes. For NSCLC, shared cancer types between MSK-MET and TCGA included adenocarcinoma and squamous cell carcinoma. The frequency of WGD was calculated as the number of WGD as the number of WGD-positive samples over the total sample count within a self-reported racial group by share histological subtype for each cancer type; testing for association was done via two-sided Pearson's Chi-squared test across racial groups.

Logistic Regression of Genomic Aberrations to WGD

In MSK-MET, in order to determine which genomic aberrations correlated with increased rates of WGD, we built multivariate logistic regression models in R for Black patients only, white patients only, and all Black and white patients (i.e., all patients). Genomic aberrations that did not occur frequently enough for inclusion in either of the Black-only or white-only models were masked in the "all patients" model. Models included correction for cancer type and removal of covariates in two stages: 1) removal of aliased covariates (i.e., those with perfect correlation to another covariate), and 2) recursive removal of the covariate with the highest variance inflation factor (VIF) until all covariates had a VIF≤4. Significance of covariates were tested by Wald test followed by multiple hypothesis correction with Benjamini & Hochberg's method[93].

TP53 Analysis

Nonsynonymous mutations in *TP53* were analyzed across racial groups, WGD status, and cancer type. Sample containing at least one nonsynonymous mutation in *TP53* were consider *TP53*-mutant, while the remaining samples were considered *TP53*-WT. The frequency of *TP53* mutation was calculated as the proportion of *TP53*-mutant samples over the total sample count within a racial group at a pan-cancer level and additionally subset by cancer type and/or WGD status; testing for association was done via two-sided Pearson's Chi-squared test across racial groups. Variant classification distributions were summarized within each racial group at a pan-cancer level and additionally subset by cancer type; testing for association was done via two-sided Pearson's Chi-squared test across racial groups. Location of *TP53* mutations were summarized within each racial group and visually compared across racial groups.

Survival Analysis

Survival analysis was performed in Python, using the packages: lifelines, matplotlib, and pandas. Significance was tested via logrank test.

Clinical Correlates of WGD and Patient Race

Equivalence of ages at key clinical event times (diagnosis, death, sequencing, and surgical procedure) were compared across racial groups and WGD status via Wilcoxon rank-sum test. All ages except for diagnosis were directly reported by MSK-MET, while age at diagnosis was inferred based on survival status, whereby dead patients' age at diagnosis was calculated as overall survival in months, while for living patients' age at diagnosis was calculated from age at last contact minus overall survival in months. Microsatellite instability status was determined using designated Stable and Instable determinations; the frequency of microsatellite instability (MSI-H) was determined as the number of Instable samples divided by the total number of Stable and Instable samples within a racial group or WGD status; testing for association was done via two-sided Pearson's Chi-squared test. Staging information for MSK-MET dataset remained incomplete. The frequency of metastasis location was determined by the proportion of samples with recorded metastasis divided by the total number of samples from a racial group or WGD status; testing for association was done via two-sided Pearson's Chi-squared test. For those metastases occurring in only a subset of patients (e.g., Female Genital and Male Genital), only samples contained within the same subset were considered in our calculations.

Data Visualization

Scientific illustrations were assembled using BioRender. Graphs and scatterplots were generated using Graphpad Prism.

ACKNOWLEDGMENTS

We thank John Kunstman, Sajid Khan, Jun Lu, Marc Vittoria, Kanika Arora, Elizabeth Godfrey, Daniel Kerekes, and members of the Sheltzer Lab for helpful feedback on this manuscript. Research in the Sheltzer Lab is supported by NIH grants R01CA237652 and R01CA276666, Department of Defense grant W81XWH-20-1-068, an American Cancer Society Research Scholar Grant, a Breast Cancer Alliance Young Investigator Award, a sponsored research agreement from Ono Pharmaceuticals, and a sponsored research agreement from Meliora Therapeutics.

DECLARATION OF INTERESTS

J.M.S. has received consulting fees from Merck, Pfizer, Ono Pharmaceuticals, and Highside Capital Management, is a member of the advisory boards of Tyra Biosciences, BioIO, and the Chemical Probes Portal, and is a co-founder of Meliora Therapeutics.

FIGURE LEGENDS

Figure 1. WGDs are more common among Black cancer patients and those with African ancestry. (A) A schematic of the whole-genome duplication process. WGDs produce a cell with a doubled chromosome complement (typically 4N). WGD events are associated with metastasis and disease progression. Loss of the tumor suppressor *TP53* promotes WGDs; other causes of WGDs are largely unknown. **(B-D)** The frequency of WGDs in Black and white patients from three different cohorts: **(B)** MSK-MET, **(C)** TCGA, and **(D)** PCAWG. Note that in the PCAWG dataset, a patient's self-reported race was not available, and instead inferred genetic ancestry was used (African (AFR) vs European (EUR)). **(E)** The frequency of WGD events in Black and white patients with either breast cancer, endometrial cancer, or NSCLC, in the MSK-MET and TCGA cohorts. Statistical testing was performed via two-tailed Pearson's Chi-squared test. Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure 2. Genetic analysis of WGD events in Black and white cancer patients. **(A)** A volcano plot displaying genetic alterations associated with an increased or decreased likelihood of WGD events across both Black and white patients in the MSK-MET cohort. Acronyms: Mutant (Mut). **(B)** A table displaying 10 events exhibiting the strongest correlation with WGD events among Black patients. **(C)** A table displaying 10 events exhibiting the strongest correlation with WGD events among white patients. **(D)** A bar graph displaying the frequency of *TP53* mutations across Black and white cancer patients. **(E)** A bar graph displaying the frequency of *TP53* mutations across Black and white patients with either breast cancer, endometrial cancer, or NSCLC. **(F)** A bar graph displaying the frequency of WGD events among Black and white cancer patients, divided based on *TP53* status. **(G)** A bar graph displaying the frequency of WGD events among Black and white patients with either breast cancer, endometrial cancer, or NSCLC, divided based on *TP53* status. **(H)** A bar graph displaying the distribution of different types of *TP53* mutations in Black and white cancer patients. **(I)** A bar graph displaying the distribution of different types of *TP53* mutations in Black and white patients with either breast cancer, endometrial cancer, or NSCLC. **(J)** A lollipop plot displaying the sites of TP53 mutations in tumors from either white (top) or Black (bottom) cancer patients. Statistical testing was performed via Wald test (A-C) with correction by Benjamini-Hochberg's method (B-C) and two-tailed Pearson's Chi-squared test (D-I). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure 3. Pan-cancer survival analysis in Black and white cancer patients. **(A)** A Kaplan-Meier plot displaying survival of Black and white patients in the MSK-MET cohort. **(B)** A Kaplan-Meier plot displaying survival of cancer patients based on WGD status. **(C)** A Kaplan-Meier plot displaying survival of WGD-positive Black and white cancer patients. **(D)** A Kaplan-Meier plot displaying survival of WGD-negative Black and white cancer patients. Statistical testing was performed via logrank test (A-D).

Figure 4. Clinical correlates of WGD status in Black and white cancer patients. **(A)** A table displaying the demographics of Black and white cancer patients in the MSK-MET cohort. **(B)** A table displaying the demographics of cancer patients based on WGD status. **(C)** A bar graph displaying the frequency of microsatellite instability (MSI-H) among tumors from Black and white cancer patients. (**D)** A bar graph displaying the frequency of microsatellite instability (MSI-H) among cancer patients based on WGD status. **(E)** The frequency of metastatic dissemination to different anatomic sites, divided by patient race. Locations written in red are significantly more likely among Black patients, no locations were more likely among white patients. For Ovary and Female Genital locations, frequency of metastasis represents female patients only. For Male Genital location, frequency of metastasis represents male patients only. **(F)** The frequency of metastatic dissemination to different anatomic sites based on WGD status. Locations written in red are significantly more likely among WGD-positive patients, no locations were more likely among WGD-negative patients. For Ovary and Female Genital locations, frequency of metastasis represents female patients only. For Male Genital location, frequency of metastasis represents male patients only. Acronyms: Central Nervous System (CNS), Lymph Node (LN), Peripheral Nervous System (PNS), Urinary Tract (UT). Statistical testing was performed via Wilcoxon rank-sum test (A-B) and two-tailed Pearson's Chi-squared test (C-F). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Frequency of WGD events by cancer type. (A) A table showing the frequency of WGD events in different cancer types across the MSK-MET, TCGA, and PCAWG patient cohorts. **(B)** A scatterplot showing the correlation between the frequency of WGD events within cancer types between the TCGA and MSK-MET datasets. Correlation was conducted via Pearson correlation.

Figure S2. WGD status among Asian patients and patient sex. (A) A bar graph displaying the frequency of WGD events among tumors from Asian and white cancer patients. **(B)** A bar graph displaying the frequency of WGD events among tumors from female and male cancer patients. Statistical testing was performed via two-tailed Pearson's Chi-squared test (A-B). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure S3. WGD status separated based on metastatic status and tumor stage. (A) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by metastatic status in the MSK-MET breast cancer cohort. (B) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by pathological stage in TCGA breast cancer cohort. (C) A bar graph displaying the distribution of metastatic status by race within the MSK-MET breast cancer cohort. (D) A bar graph displaying the distribution of pathological stage by race within the TCGA breast cancer cohort. (E) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by metastatic status in the

MSK-MET endometrial cancer cohort. (F) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by clinical stage in TCGA endometrial cancer cohort. (G) A bar graph displaying the distribution of metastatic status by race within the MSK-MET endometrial cancer cohort. (H) A bar graph displaying the distribution of clinical stage by race within the TCGA endometrial cancer cohort. (I) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by metastatic status in the MSK-MET NSCLC cancer cohort. (J) A bar graph displaying the frequency of WGD events among tumors from Black and white patients stratified by pathological stage in TCGA NSCLC cancer cohort. (K) A bar graph displaying the distribution of metastatic status by race within the MSK-MET NSCLC cancer cohort. (L) A bar graph displaying the distribution of pathological stage by race within the TCGA NSCLC cancer cohort. Statistical testing was performed via Fisher's exact test (A, B, E, F, I, J) and Pearson's Chi-squared test (C, D, G, H, K, L). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure S4. WGD status separated based on histological subtypes. (A) A bar graph displaying the frequency of WGD events among tumors from Black and white patients from shared breast cancer histological subtypes in MSK-MET and TCGA cohorts. (B) A bar graph displaying the frequency of WGD events among tumors from Black and white patients from shared endometrial cancer histological subtypes in MSK-MET and TCGA cohorts. (C) A bar graph displaying the frequency of WGD events among tumors from Black and white patients from shared endometrial cancer histological subtypes in MSK-MET and TCGA cohorts. Acronyms: Invasive/Infiltrating Ductal Carcinoma (IDC), Invasive/Infiltrating Carcinoma (ILC). Statistical testing was performed via Pearson's Chi-squared test (A-C). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure S5. Analysis of WGD status based on inferred genetic ancestry. (A) A bar graph displaying the frequency of WGD events among tumors from cancer patients with inferred African ancestry (AFR) and European ancestry (EUR). **(B)** A bar graph displaying the frequency of WGD events among patients with either breast cancer, endometrial cancer, or NSCLC based on inferred genetic ancestry. Statistical testing was performed via two-tailed Pearson's Chi-squared test (A-B). Statistical significance: NS $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure S6. Survival analysis by cancer type in Black and white cancer patients. **(A-C)** A Kaplan-Meier plot displaying survival of Black and white patients with **(A)** breast cancer, **(B)** endometrial cancer, and **(C)** NSCLC. **(D)** A Kaplan-Meier plot displaying survival of endometrial cancer patients based on WGD status. **(E)** A Kaplan-Meier plot displaying the survival of WGD-positive Black and white endometrial cancer patients. **(F)** A Kaplan-Meier plot displaying the survival of WGD-negative Black and white endometrial cancer patients. Statistical testing was performed via logrank test (A-F).

Figure S7. Association between WGD events, *TP53* status, patient race, and patient survival. **(A)** A Kaplan-Meier plot displaying the survival of cancer patients based on *TP53* mutation status. **(B)** A Kaplan-Meier plot displaying the survival of *TP53*-WT Black and white cancer patients. **(C)** A Kaplan-Meier plot displaying the survival of *TP53*-WT cancer patients based on WGD status. **(D)** A Kaplan-Meier plot displaying the survival of *TP53*-WT, WGD-positive Black and white cancer patients. **(E)** A Kaplan-Meier plot displaying the survival of *TP53*-WT, WGD-negative Black and white cancer patients. **(F)** A Kaplan-Meier plot displaying the survival of *TP53*-mutant Black and white cancer patients. **(G)** A Kaplan-Meier plot displaying the survival of *TP53*-mutant cancer patients based on WGD status. **(H)** A Kaplan-Meier plot displaying the survival of *TP53*-mutant, WGD-

positive Black and white cancer patients. **(I)** A Kaplan-Meier plot displaying the survival of *TP53*-mutant, WGD-negative Black and white cancer patients. Statistical testing was performed via logrank test (A-I).

WORKS CITED

1. Caraballo, C. *et al.* Temporal Trends in Racial and Ethnic Disparities in Multimorbidity Prevalence in the United States, 1999-2018. *Am. J. Med.* **135**, 1083-1092.e14 (2022).
2. National Academies of Sciences, E. *et al.* The State of Health Disparities in the United States. in *Communities in Action: Pathways to Health Equity* (National Academies Press (US), 2017).
3. Rich, J. A. Primary Care for Young African American Men. *J. Am. Coll. Health* **49**, 183–186 (2001).
4. Williams, D. R. & Mohammed, S. A. Discrimination and racial disparities in health: evidence and needed research. *J. Behav. Med.* **32**, 20–47 (2009).
5. Heller, D. R., Nicolson, N. G., Ahuja, N., Khan, S. & Kunstman, J. W. Association of Treatment Inequity and Ancestry With Pancreatic Ductal Adenocarcinoma Survival. *JAMA Surg.* **155**, e195047 (2020).
6. Sherman, M. E. & Devesa, S. S. Analysis of racial differences in incidence, survival, and mortality for malignant tumors of the uterine corpus. *Cancer* **98**, 176–186 (2003).
7. Megwalu, U. C. & Ma, Y. Racial Disparities in Oropharyngeal Cancer Stage at Diagnosis. *Anticancer Res.* **37**, 835–839 (2017).
8. O'Keefe, E. B., Meltzer, J. P. & Bethea, T. N. Health disparities and cancer: racial disparities in cancer mortality in the United States, 2000-2010. *Front. Public Health* **3**, 51 (2015).
9. Aizer, A. A. *et al.* Lack of reduction in racial disparities in cancer-specific mortality over a 20-year period. *Cancer* **120**, 1532–1539 (2014).
10. Islami, F. *et al.* American Cancer Society's report on the status of cancer disparities in the United States, 2021. *CA. Cancer J. Clin.* **72**, 112–143 (2022).
11. Yedjou, C. G. *et al.* Health and Racial Disparity in Breast Cancer. *Adv. Exp. Med. Biol.* **1152**, 31–49 (2019).
12. Ryan, B. M. Lung cancer health disparities. *Carcinogenesis* **39**, 741–751 (2018).

13. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA. Cancer J. Clin.* **72**, 7–33 (2022).

14. D'Arcy, M. *et al.* Race-associated biological differences among Luminal A breast tumors. *Breast Cancer Res. Treat.* **152**, 437–448 (2015).

15. Beebe-Dimmer, J. L. *et al.* Racial Differences in Risk of Prostate Cancer Associated With Metabolic Syndrome. *Urology* **74**, 185–190 (2009).

16. Dess, R. T. *et al.* Association of Black Race With Prostate Cancer–Specific and Other-Cause Mortality. *JAMA Oncol.* **5**, 975 (2019).

17. Olson, S. H. *et al.* The Impact of Race and Comorbidity on Survival in Endometrial Cancer. *Cancer Epidemiol. Biomarkers Prev.* **21**, 753–760 (2012).

18. Cote, M. L., Ruterbusch, J. J., Olson, S. H., Lu, K. & Ali-Fehmi, R. The Growing Burden of Endometrial Cancer: A Major Racial Disparity Affecting Black Women. *Cancer Epidemiol. Biomarkers Prev.* **24**, 1407–1415 (2015).

19. Chornokur, G., Dalton, K., Borysova, M. & Kumar, N. Disparities at Presentation, Diagnosis, Treatment and Survival in African American Men, Affected by Prostate Cancer. *The Prostate* **71**, 985–997 (2011).

20. Park, E. R., Japuntich, S. J., Traeger, L., Cannon, S. & Pajolek, H. Disparities Between Blacks and Whites in Tobacco and Lung Cancer Treatment. *The Oncologist* **16**, 1428–1434 (2011).

21. Williams, D. R. & Rucker, T. D. Understanding and Addressing Racial Disparities in Health Care. *Health Care Financ. Rev.* **21**, 75–90 (2000).

22. Coughlin, S. S., King, J., Richards, T. B. & Ekwueme, D. U. Cervical Cancer Screening among Women in Metropolitan Areas of the United States by Individual-Level and Area-Based Measures of Socioeconomic Status, 2000 to 2002. *Cancer Epidemiol. Biomarkers Prev.* **15**, 2154–2159 (2006).

23. Stormacq, C., Van den Broucke, S. & Wosinski, J. Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. *Health Promot. Int.* **34**, e1–e17 (2019).

24. Beech, B. M., Ford, C., Thorpe, R. J., Bruce, M. A. & Norris, K. C. Poverty, Racism, and the Public Health Crisis in America. *Front. Public Health* **9**, 699049 (2021).

25. Tannenbaum, S. L., Koru-Sengul, T., Zhao, W., Miao, F. & Byrne, M. M. Survival Disparities in Non–Small Cell Lung Cancer by Race, Ethnicity, and Socioeconomic Status. *Cancer J.* **20**, 237–245 (2014).

26. Larsen, K., Rydz, E. & Peters, C. E. Inequalities in Environmental Cancer Risk and Carcinogen Exposures: A Scoping Review. *Int. J. Environ. Res. Public. Health* **20**, 5718 (2023).

27. *2022 National Healthcare Quality and Disparities Report*. (Agency for Healthcare Research and Quality (US), 2022).

28. Green, A. K. *et al.* Racial disparities in chemotherapy administration for early-stage breast cancer: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **172**, 247–263 (2018).

29. Bakkila, B. F. *et al.* Evaluation of Racial Disparities in Quality of Care for Patients With Gastrointestinal Tract Cancer Treated With Surgery. *JAMA Netw. Open* **5**, e225664 (2022).

30. Zhang, W., Edwards, A., Flemington, E. K. & Zhang, K. Racial disparities in patient survival and tumor mutation burden, and the association between tumor mutation burden and cancer incidence rate. *Sci. Rep.* **7**, 13639 (2017).

31. Nassar, A. H., Adib, E. & Kwiatkowski, D. J. Distribution of *KRAS* ^{G12C} Somatic Mutations across Race, Sex, and Cancer Type. *N. Engl. J. Med.* **384**, 185–187 (2021).

32. Bollig-Fischer, A. *et al.* Racial Diversity of Actionable Mutations in Non–Small Cell Lung Cancer. *J. Thorac. Oncol.* **10**, 250–255 (2015).

33. Cote, M. L. *et al.* Racial Differences in Oncogene Mutations Detected in Early-Stage Low-Grade Endometrial Cancers. *Int. J. Gynecol. Cancer* **22**, 1367–1372 (2012).

34. Arora, K. *et al.* Genetic Ancestry Correlates with Somatic Differences in a Real-World Clinical Cancer Sequencing Cohort. *Cancer Discov.* **12**, 2552–2565 (2022).

35. Yuan, J. *et al.* Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell* **34**, 549-560.e9 (2018).

36. Guttery, D. S. *et al.* Racial differences in endometrial cancer molecular portraits in The Cancer Genome Atlas. *Oncotarget* **9**, 17093–17103 (2018).

37. Hebert-Magee, S. *et al.* The combined survival effect of codon 72 polymorphisms and p53 somatic mutations in breast cancer depends on race and molecular subtype. *PLOS ONE* **14**, e0211734 (2019).

38. Ademuyiwa, F. O., Tao, Y., Luo, J., Weilbaecher, K. & Ma, C. X. Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians. *Breast Cancer Res. Treat.* **161**, 491–499 (2017).

39. Munro, A. J., Lain, S. & Lane, D. P. P53 abnormalities and outcomes in colorectal cancer: a systematic review. *Br. J. Cancer* **92**, 434–444 (2005).

40. Keenan, T. *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. Off. J. Am. Soc. Clin. Oncol.* **33**, 3621–3627 (2015).

41. Bauml, J. *et al.* Frequency of EGFR and KRAS mutations in patients with non small cell lung cancer by racial background: Do disparities exist? *Lung Cancer* **81**, 347–353 (2013).

42. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).

43. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).

44. Lukow, D. A. & Sheltzer, J. M. Chromosomal instability and aneuploidy as causes of cancer drug resistance. *Trends Cancer* **8**, 43–53 (2022).

45. Yang, S. Y. C., Pugh, T. J. & Oza, A. M. Double Trouble: Whole-Genome Doubling Distinguishes Early from Late Ovarian Cancer. *Clin. Cancer Res.* **28**, 2730–2732 (2022).

46. Carrot-Zhang, J. *et al.* Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* **37**, 639-654.e6 (2020).

47. Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).

48. Zeng, J., Hills, S. A., Ozono, E. & Diffley, J. F. X. Cyclin E-induced replicative stress drives p53-dependent whole-genome duplication. *Cell* **186**, 528-542.e14 (2023).

49. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).

50. Frankell, A. M. *et al.* The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023).

51. Prasad, K. *et al.* Whole-Genome Duplication Shapes the Aneuploidy Landscape of Human Cancers. *Cancer Res.* **82**, 1736–1752 (2022).

52. Lambuta, R. A. *et al.* Whole-genome doubling drives oncogenic loss of chromatin segregation. *Nature* 1–9 (2023) doi:10.1038/s41586-023-05794-2.

53. Nguyen, B. *et al.* Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563-575.e11 (2022).

54. The Cancer Genome Atlas homepage; http://cancergenome.nih.gov/abouttcga.

55. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

56. Spratt, D. E. *et al.* Racial/Ethnic Disparities in Genomic Sequencing. *JAMA Oncol.* **2**, 1070–1074 (2016).

57. Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L. & Bonham, V. L. Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice. *Health Aff. (Millwood)* **37**, 780–785 (2018).

58. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research *et al. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. 26902 (National Academies Press, 2023). doi:10.17226/26902.

59. Why Nature is updating its advice to authors on reporting race or ethnicity. *Nature* **616**, 219–219 (2023).

60. Ashing, K. T., Jones, V., Bedell, F., Phillips, T. & Erhunmwunsee, L. Calling Attention to the Role of Race-Driven Societal Determinants of Health on Aggressive Tumor Biology: A Focus on Black Americans. *JCO Oncol. Pract.* **18**, 15–22 (2022).

61. Kikutake, C. & Suyama, M. Pan-cancer analysis of whole-genome doubling and its association with patient prognosis. *BMC Cancer* **23**, 619 (2023).

62. Li, K., Luo, H., Huang, L., Luo, H. & Zhu, X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int.* **20**, 16 (2020).

63. Yamamoto, H. & Imai, K. Microsatellite instability: an update. *Arch. Toxicol.* **89**, 899–921 (2015).

64. Popat, S., Hubner, R. & Houlston, R. S. Systematic review of microsatellite instability and colorectal cancer prognosis. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **23**, 609–618 (2005).

65. Zhu, L. *et al.* Microsatellite instability and survival in gastric cancer: A systematic review and meta-analysis. *Mol. Clin. Oncol.* **3**, 699–705 (2015).

66. Jaratlerdsiri, W. *et al.* African-specific molecular taxonomy of prostate cancer. *Nature* **609**, 552–559 (2022).

67. Hieronymus, H. *et al.* Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *eLife* **7**, e37294 (2018).

68. Gómez-Rueda, H., Martínez-Ledesma, E., Martínez-Torteya, A., Palacios-Corona, R. & Trevino, V. Integration and comparison of different genomic data for outcome prediction in cancer. *BioData Min.* **8**, 32 (2015).

69. Van Dijk, E. *et al.* Chromosomal copy number heterogeneity predicts survival rates across cancers. *Nat. Commun.* **12**, 3188 (2021).

70. Smith, J. C. & Sheltzer, J. M. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *eLife* **7**, e39217 (2018).

71. Girish, V. *et al.* Oncogene-like addiction to aneuploidy in human cancers. *Science* **381**, eadg4521 (2023).

72. Quinton, R. J. *et al.* Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature* **590**, 492–497 (2021).

73. Cohen-Sharir, Y. *et al.* Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. *Nature* **590**, 486–491 (2021).

74. Govindan, R. *et al.* Trial in progress: A phase 1, multicenter, open-label, dose-exploration and dose-expansion study evaluating the safety, tolerability, pharmacokinetics, and efficacy of AMG650 in subjects with advanced solid tumors. *J. Clin. Oncol.* **39**, TPS5600–TPS5600 (2021).

75. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2**, 401–404 (2012).

76. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, (2013).

77. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131–e131 (2016).

78. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

79. R Core Team. *R: A language and environment for statistical computing*. https://www.R-project.org/ (2022).

80. Van Rossum, G. & Drake, F. L. *Python 3 reference manual*. (CreateSpace, 2009).

81. Fox, J. & Weisberg, S. *An R companion to applied regression*. (SAGE, 2019).

82. Blighe, K. EnhancedVolcano. (2018) doi:10.18129/B9.BIOC.ENHANCEDVOLCANO.

83. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing : Imprint: Springer, 2016). doi:10.1007/978-3-319-24277-4.

84. Iannone, R. *et al. gt: Easily create presentation-ready display tables*. https://CRAN.R-project.org/package=gt (2023).

85. Sjoberg, D., D., Whiting, K., Curry, M., Lavery, J., A. & Larmarange, J. Reproducible Summary Tables with the gtsummary Package. *R J.* **13**, 570 (2021).

86. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).

87. Schauberger, P. & Walker, A. *openxlsx: Read, write and edit xlsx files*. https://CRAN.R-project.org/package=openxlsx (2023).

88. Wickham, H. Reshaping Data with the **reshape** Package. *J. Stat. Softw.* **21**, (2007).

89. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

90. Davidson-Pilon, C. lifelines: survival analysis in Python. *J. Open Source Softw.* **4**, 1317 (2019).

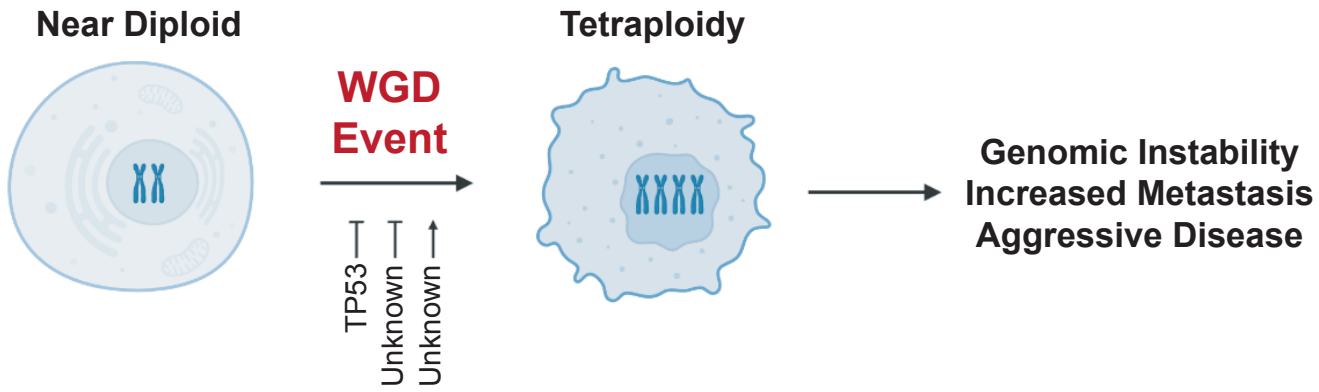
91. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

92. Team, T. P. D. pandas-dev/pandas: Pandas. (2023) doi:10.5281/ZENODO.3509134.

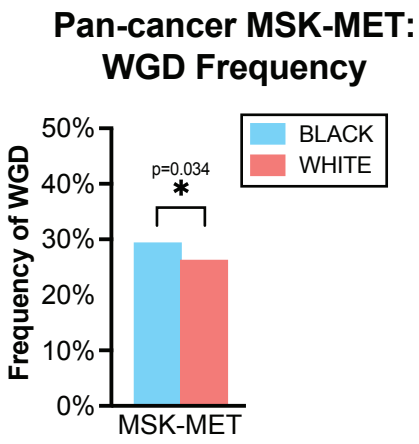
93. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

Figure 1

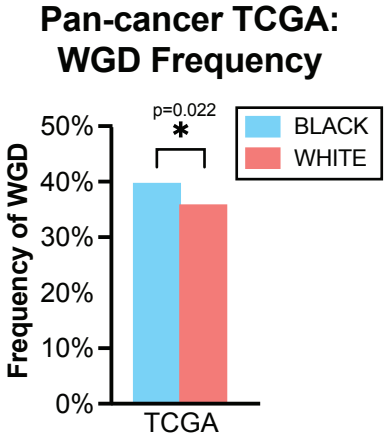
A



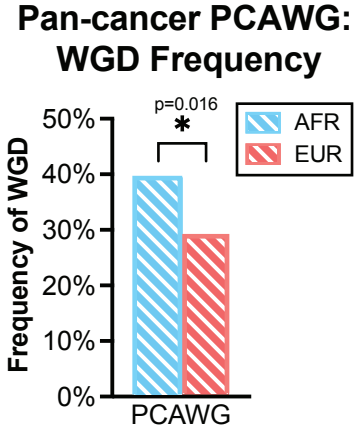
B



C



D



E

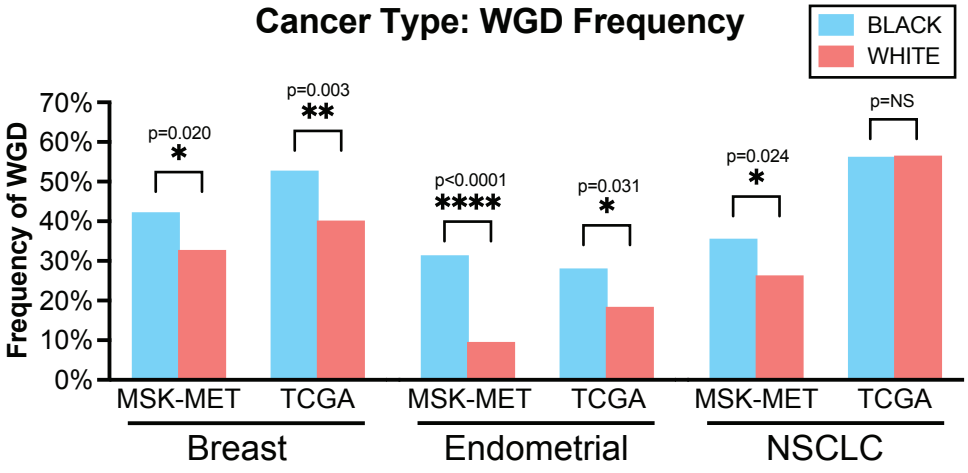
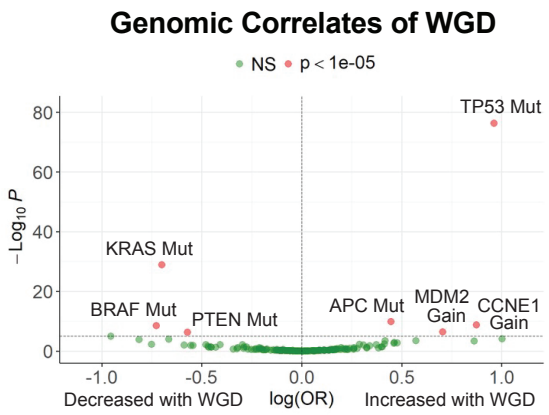


Figure 2

A



B

Genomic Correlates of WGD: Black Patients

Gene	Event	OR[†]	95% CI[†]	q-value
TP53	Mut	4.19	2.71, 6.47	**<0.001**
CCNE1	Gain	3.76	1.6, 8.87	0.089
GATA3	Mut	3.98	1.42, 11.18	0.2
STK11	Mut	7.40	1.59, 34.36	0.2
EGFR	Mut	2.52	1.2, 5.28	0.2
ERBB2	Gain	4.71	1.3, 17.04	0.2
PIK3C2G	Mut	4.68	1.08, 20.23	0.4
KIT	Mut	2.52	0.95, 6.72	0.5
FGFR1	Gain	2.24	0.95, 5.29	0.5
RAD21	Gain	4.02	0.81, 19.86	0.5

[†] OR = Odds Ratio, CI = Confidence Interval

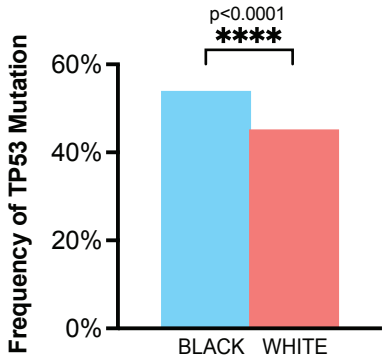
C

Genomic Correlates of WGD: White Patients

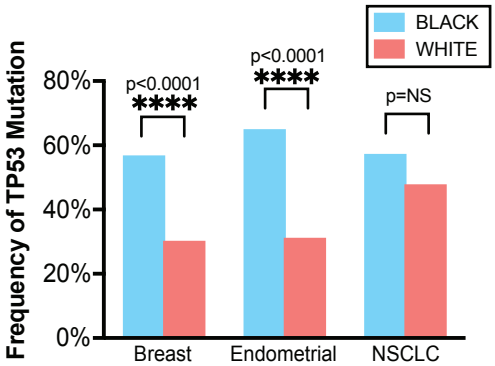
Gene	Event	OR[†]	95% CI[†]	q-value
TP53	Mut	2.62	2.37, 2.9	**<0.001**
APC	Mut	1.56	1.36, 1.79	**<0.001**
CCNE1	Gain	2.39	1.8, 3.18	**<0.001**
MDM2	Gain	2.03	1.55, 2.66	**<0.001**
CCND1	Gain	3.69	1.85, 7.35	**0.003**
HGF	Mut	1.77	1.3, 2.41	**0.004**
MYC	Gain	1.52	1.21, 1.91	**0.005**
CDKN2A	Loss	2.37	1.47, 3.8	**0.005**
TCF7L2	Mut	1.59	1.2, 2.12	**0.016**
RECQL4	Gain	1.62	1.2, 2.19	**0.019**

[†] OR = Odds Ratio, CI = Confidence Interval

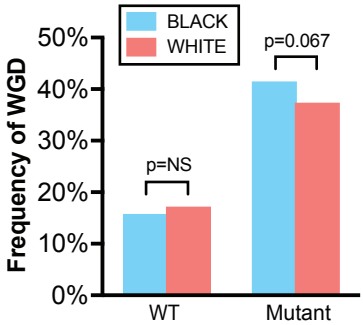
D



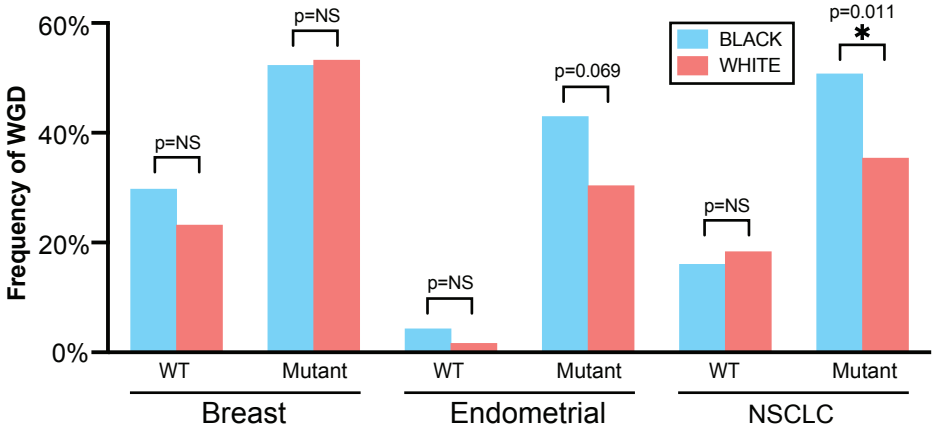
E



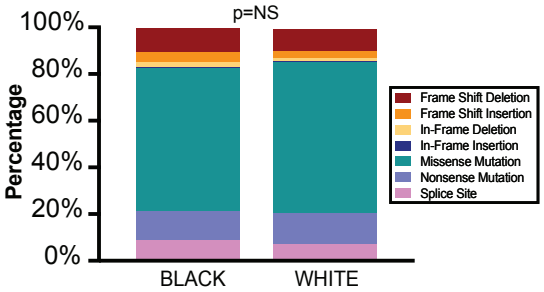
F



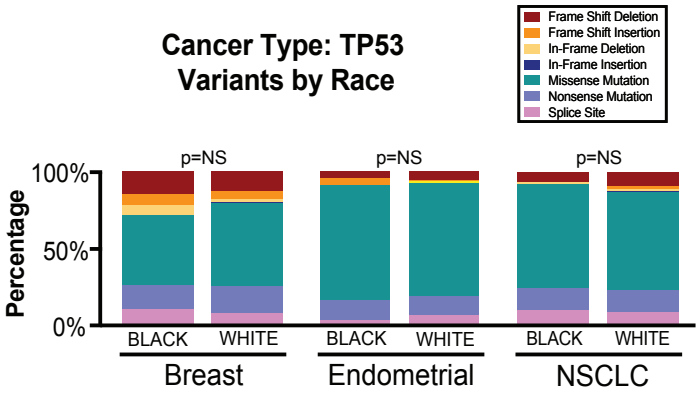
G



H



I



J

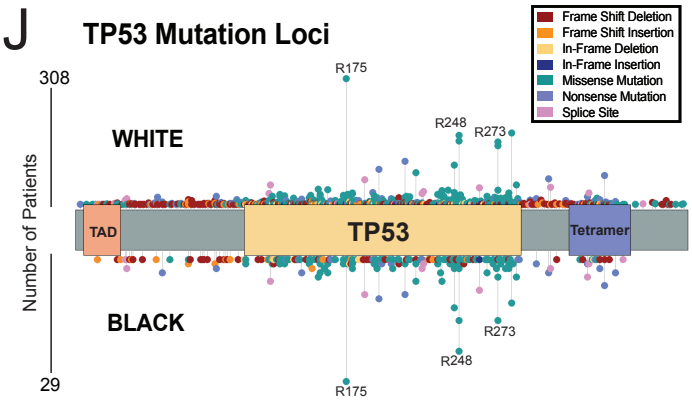


Figure 3

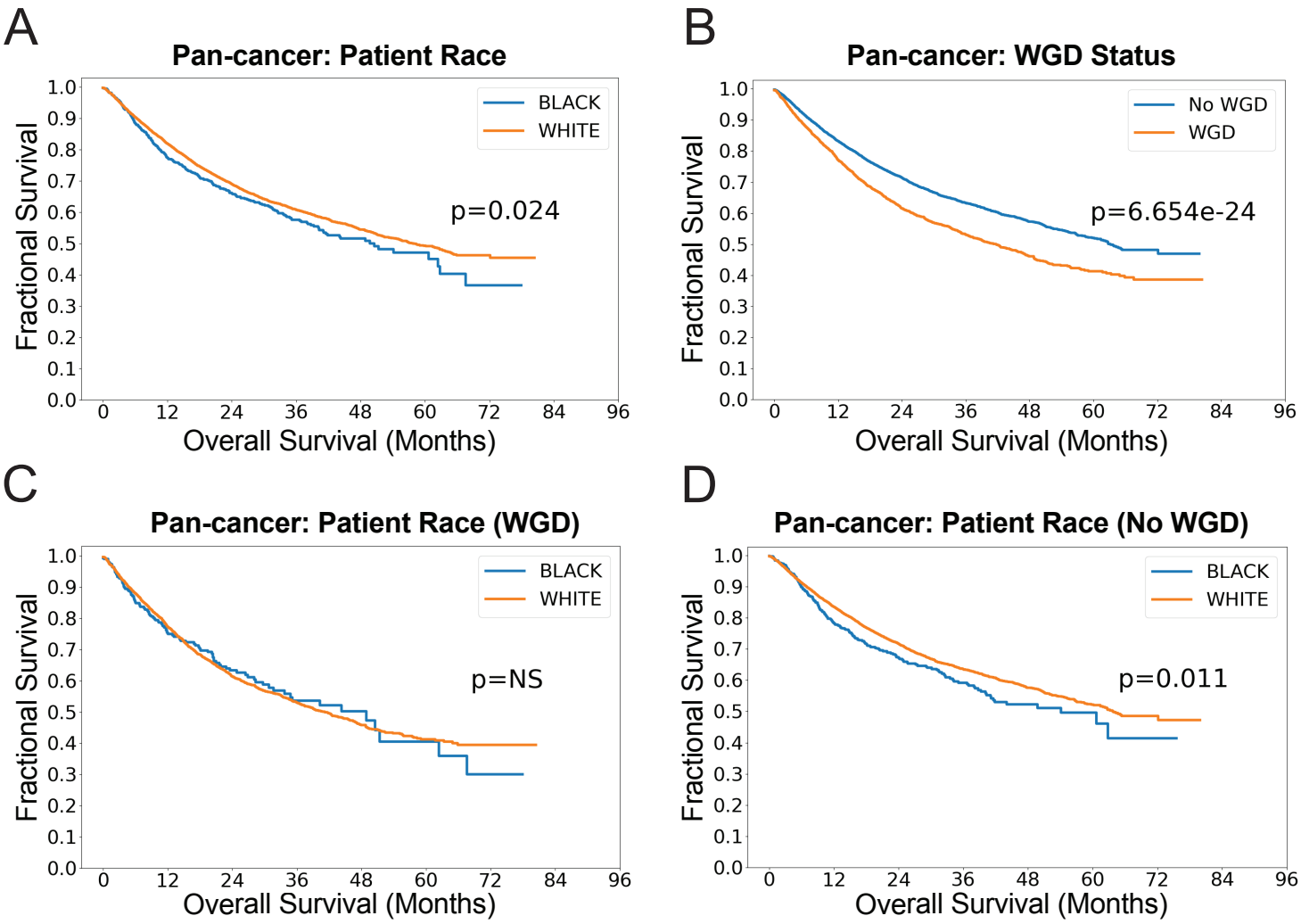

A
Pan-cancer: Patient Race
BLACK
WHITE
p=0.024
Fractional Survival
Overall Survival (Months)
B
Pan-cancer: WGD Status
No WGD
WGD
p=6.654e-24
Fractional Survival
Overall Survival (Months)
C
Pan-cancer: Patient Race (WGD)
BLACK
WHITE
p=NS
Fractional Survival
Overall Survival (Months)
D
Pan-cancer: Patient Race (No WGD)
BLACK
WHITE
p=0.011
Fractional Survival
Overall Survival (Months)

Figure 4

A

Pan-cancer: Demographics by Patient Race

Characteristic	BLACK, N = 959[1]	WHITE, N = 12,112[1]	p-value[2]
Age at Diagnosis	61 (52, 69)	65 (55, 72)	<0.001
Age at Death	63 (54, 71)	68 (59, 75)	<0.001
Age at Sequencing	61 (52, 69)	65 (55, 72)	<0.001
Age at First Metastasis	61 (51, 67)	64 (55, 71)	<0.001
Age at Surgical Procedure	61 (51, 68)	64 (54, 71)	<0.001

[1] Median (IQR)

[2] Wilcoxon rank sum test

B

Pan-cancer: Demographics by WGD Status

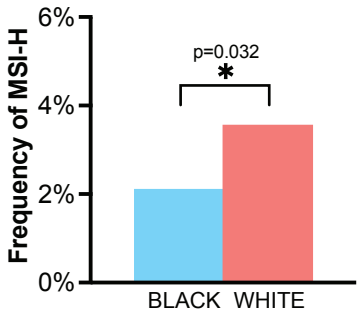
Characteristic	No WGD, N = 9,621[1]	WGD, N = 3,450[1]	p-value[2]
Age at Diagnosis	65 (55, 72)	64 (54, 72)	0.002
Age at Death	68 (59, 75)	67 (57, 74)	0.043
Age at Sequencing	65 (55, 72)	64 (54, 72)	0.004
Age at First Metastasis	64 (55, 71)	63 (53, 71)	0.002
Age at Surgical Procedure	64 (54, 71)	63 (53, 71)	0.002

[1] Median (IQR)

[2] Wilcoxon rank sum test

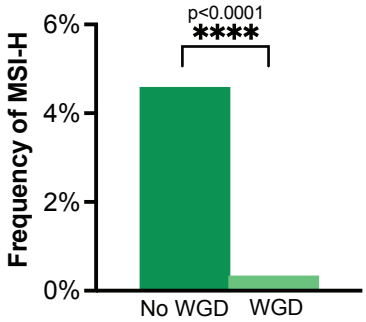
C

Pan-cancer: MSI-H Frequency by Patient Race



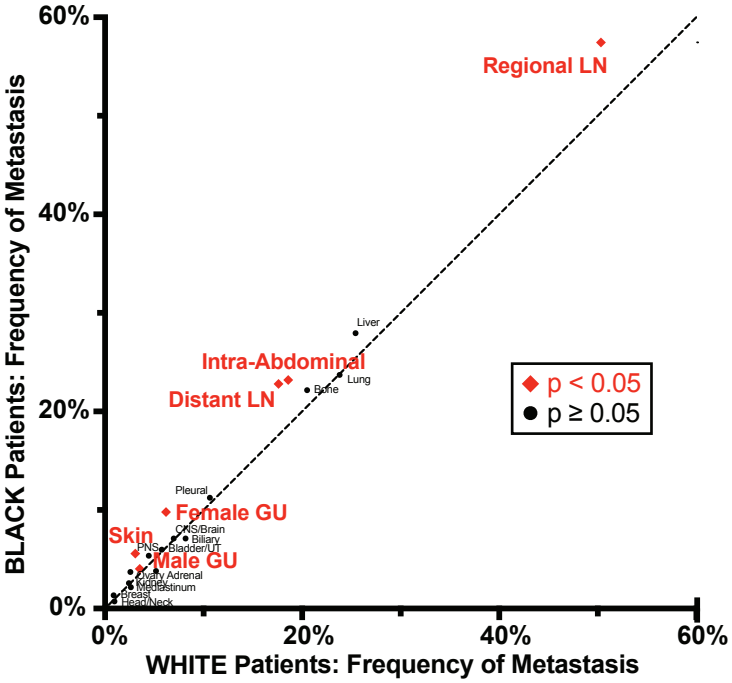
D

Pan-cancer: MSI-H Frequency by WGD Status



E

Pan-cancer: Metastasis Location by Patient Race



F

Pan-cancer: Metastasis Location by WGD Status

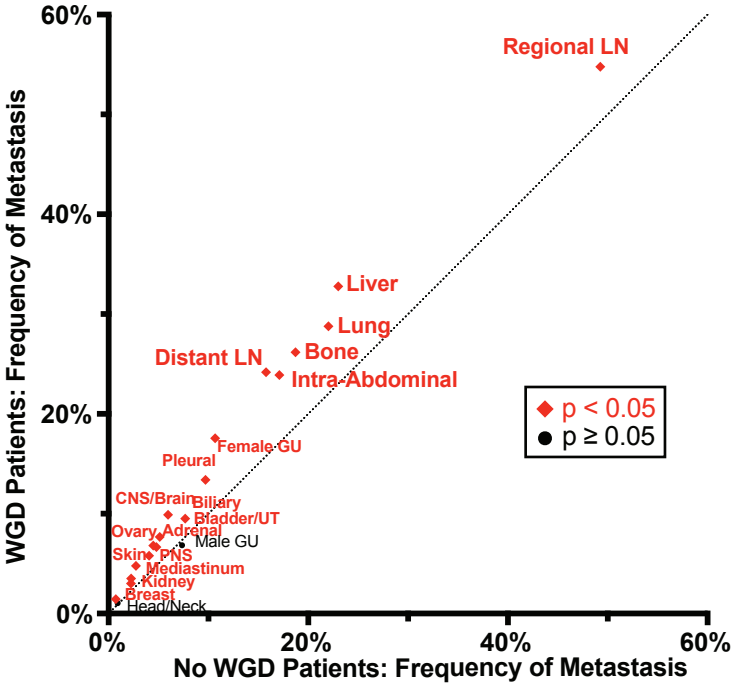


Figure S1

A

Cancer Type: WGD Frequencies			
Cancer Type	**MSK-MET**	**TCGA**	**PCAWG**
Ovarian	54.8%	59.6%	59.1%
Esophagogastric	39.7%	45.6%	60.3%
Sarcoma	37.1%	46.3%	45.2%
Breast	33.7%	42.4%	47.7%
Colorectal	31.2%	44.5%	47.9%
NSCLC	26.7%	56.5%	69.6%
Head/Neck	24.5%	38.8%	40.0%
Hepatobiliary	21.1%	31.0%	33.3%
Pancreatic	13.3%	23.8%	33.7%
Endometrial	11.8%	20.4%	63.2%
Prostate	12.2%	4.8%	4.8%
Renal	11.8%	14.5%	14.6%
Thyroid	3.8%	1.8%	0.0%
Matched cancer types with at least 30 patients			

B

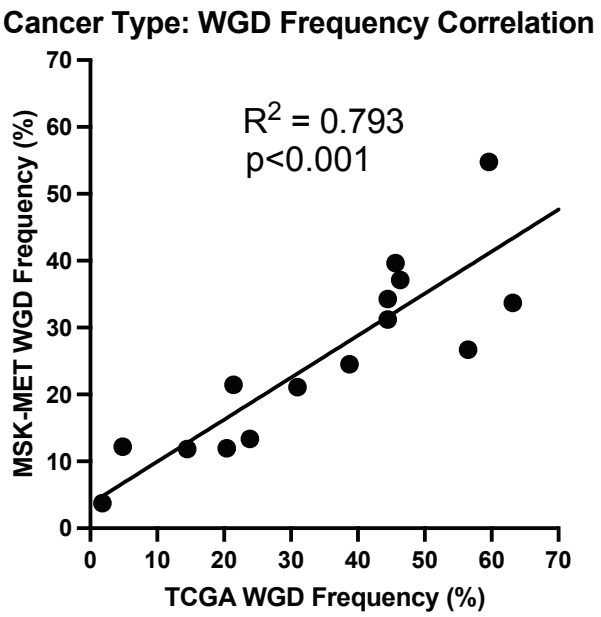


Figure S2

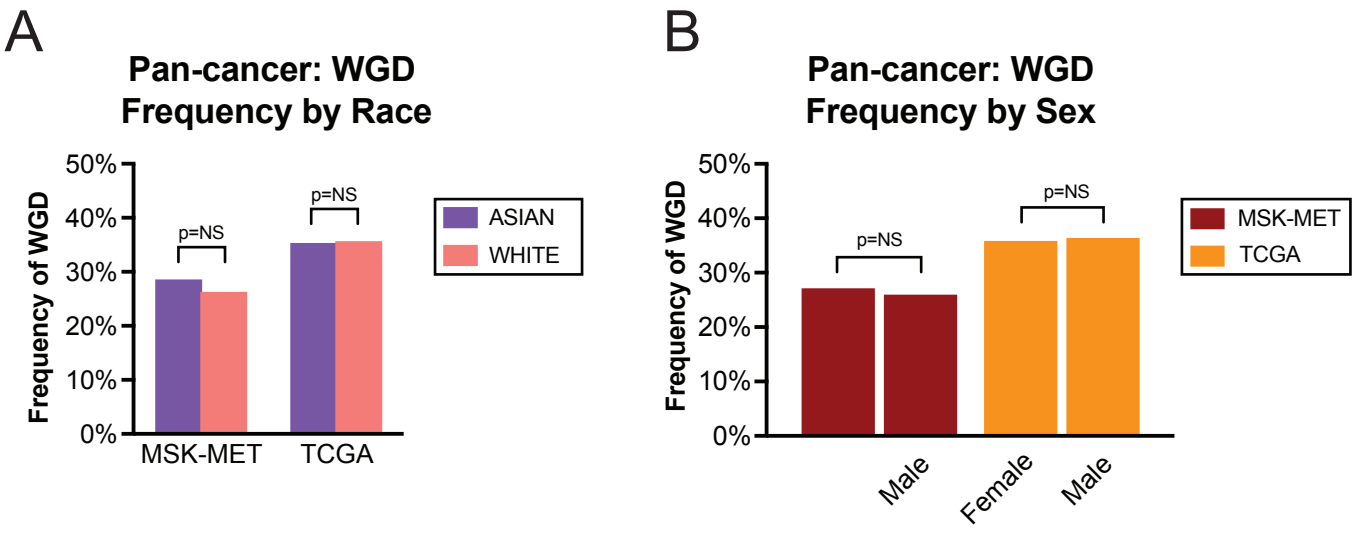

A
Pan-cancer: WGD
Frequency by Race
Frequency of WGD
50%
40%
30%
20%
10%
0%
p=NS
p=NS
MSK-MET
TCGA
ASIAN
WHITE
B
Pan-cancer: WGD
Frequency by Sex
Frequency of WGD
50%
40%
30%
20%
10%
0%
p=NS
p=NS
Male
Female
Male
MSK-MET
TCGA

Figure S3

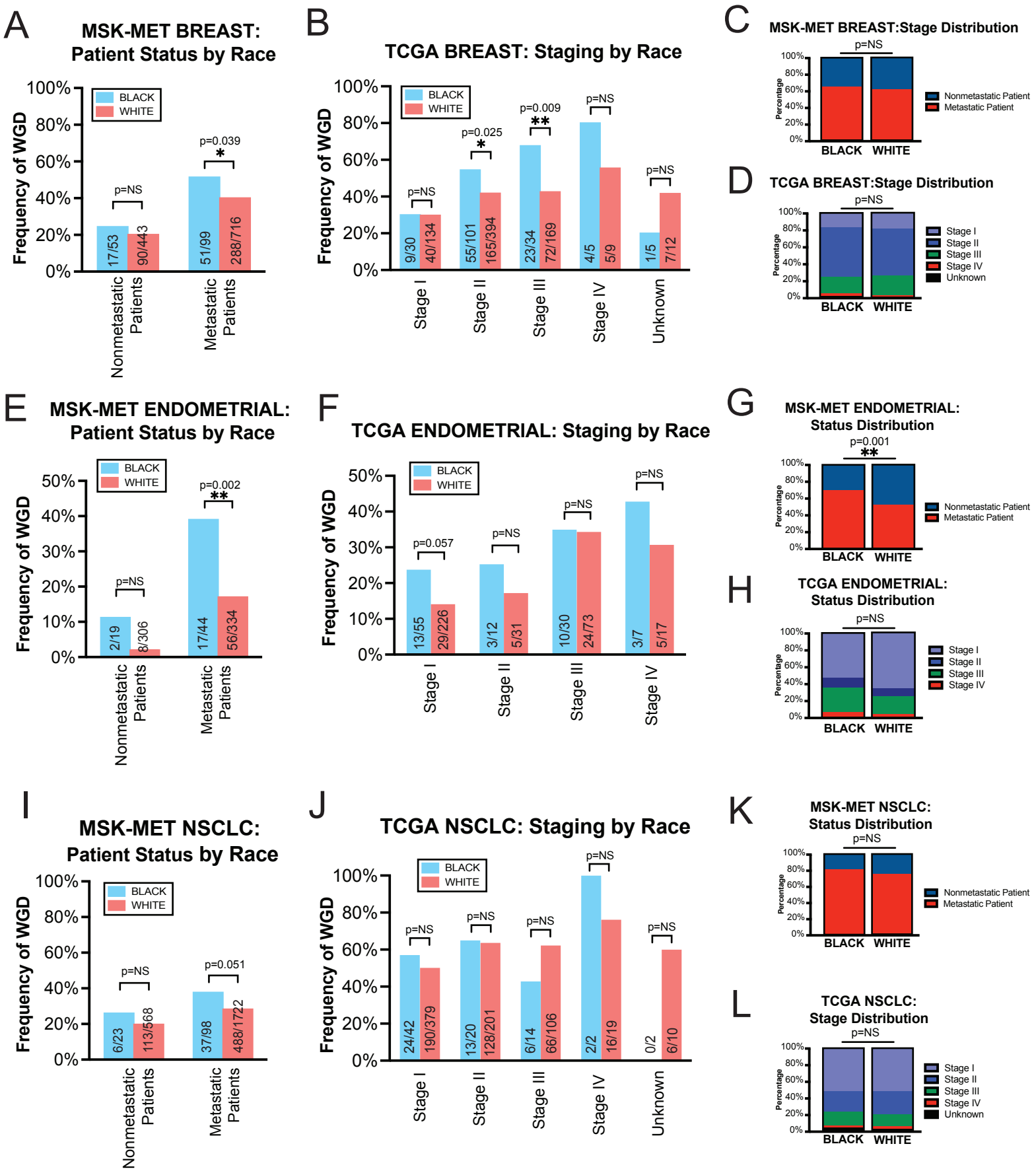

A
MSK-MET BREAST: Patient Status by Race
Frequency of WGD
BLACK
WHITE
p=NS
p=0.039
*
17/53
90/443
51/99
288/716
Nonmetastatic Patients
Metastatic Patients
B
TCGA BREAST: Staging by Race
p=NS
p=0.025
*
p=0.009
**
p=NS
p=NS
9/30
40/134
55/101
165/394
23/34
72/169
4/5
5/9
1/5
7/12
Stage I
Stage II
Stage III
Stage IV
Unknown
C
MSK-MET BREAST:Stage Distribution
p=NS
Percentage
Nonmetastatic Patient
Metastatic Patient
BLACK
WHITE
D
TCGA BREAST:Stage Distribution
p=NS
Stage I
Stage II
Stage III
Stage IV
Unknown
E
MSK-MET ENDOMETRIAL: Patient Status by Race
p=NS
p=0.002
**
2/19
8/306
17/44
56/334
F
TCGA ENDOMETRIAL: Staging by Race
p=0.057
p=NS
p=NS
p=NS
13/55
29/226
3/12
5/31
10/30
24/73
3/7
5/17
G
MSK-MET ENDOMETRIAL: Status Distribution
p=0.001
**
H
TCGA ENDOMETRIAL: Status Distribution
p=NS
I
MSK-MET NSCLC: Patient Status by Race
p=NS
p=0.051
6/23
113/568
37/98
488/1722
J
TCGA NSCLC: Staging by Race
p=NS
p=NS
p=NS
p=NS
p=NS
24/42
190/379
13/20
128/201
6/14
66/106
2/2
16/19
0/2
6/10
K
MSK-MET NSCLC: Status Distribution
p=NS
L
TCGA NSCLC: Stage Distribution
p=NS

Figure S4

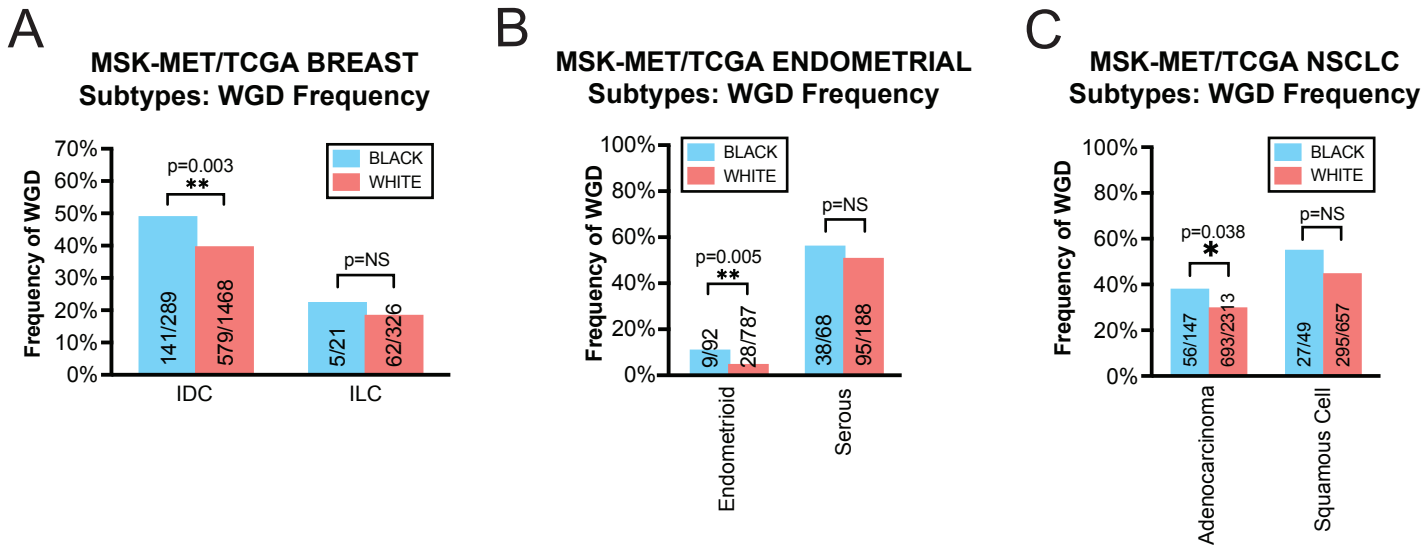
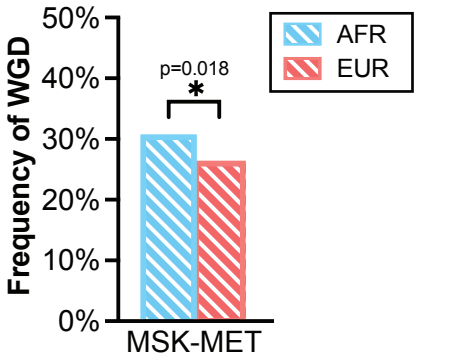

A
MSK-MET/TCGA BREAST
Subtypes: WGD Frequency
Frequency of WGD
70%
60%
50%
40%
30%
20%
10%
0%
p=0.003
**
p=NS
BLACK
WHITE
141/289
579/1468
5/21
62/326
IDC
ILC
B
MSK-MET/TCGA ENDOMETRIAL
Subtypes: WGD Frequency
Frequency of WGD
100%
80%
60%
40%
20%
0%
BLACK
WHITE
p=0.005
**
p=NS
9/92
28/787
38/68
95/188
Endometrioid
Serous
C
MSK-MET/TCGA NSCLC
Subtypes: WGD Frequency
Frequency of WGD
100%
80%
60%
40%
20%
0%
BLACK
WHITE
p=0.038
*
p=NS
56/147
693/2313
27/49
295/657
Adenocarcinoma
Squamous Cell

Figure S5

A


Pan-cancer: WGD Frequency
by Inferred Genetic Ancestry
Frequency of WGD
50%
40%
30%
20%
10%
0%
p=0.018
*
AFR
EUR
MSK-MET

B

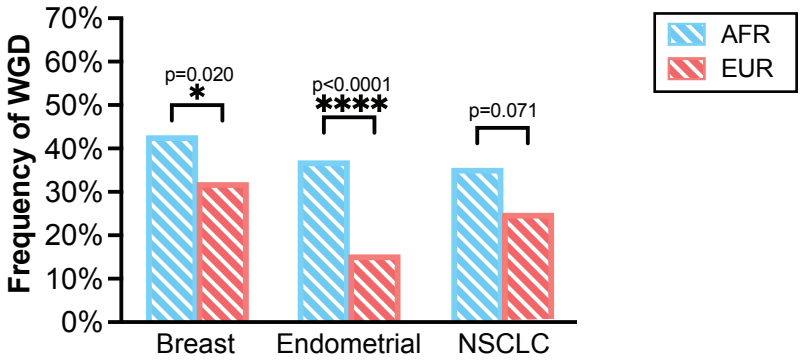

Cancer Type: WGD Frequency
by Inferred Genetic Ancestry
Frequency of WGD
70%
60%
50%
40%
30%
20%
10%
0%
p=0.020
*
p<0.0001
****
p=0.071
AFR
EUR
Breast
Endometrial
NSCLC

Figure S6

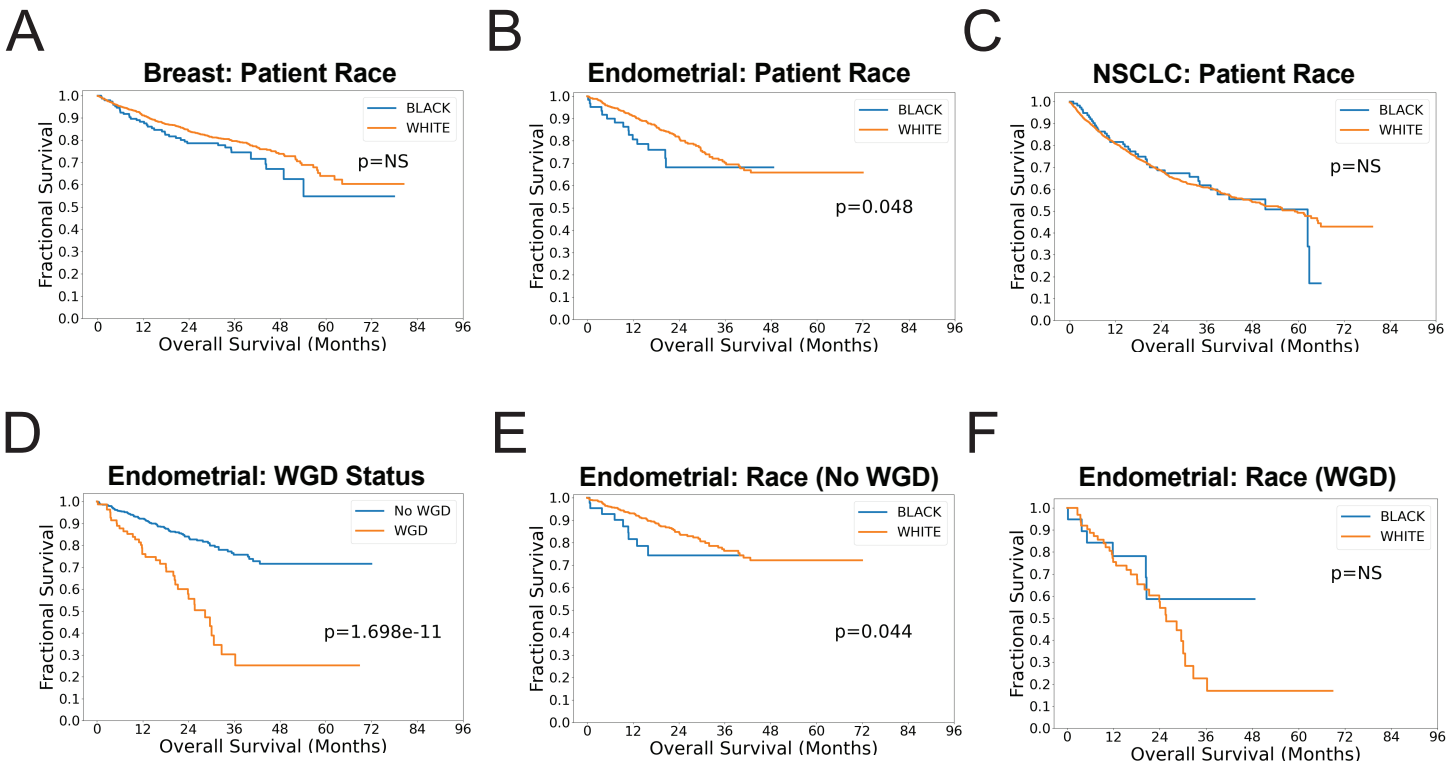
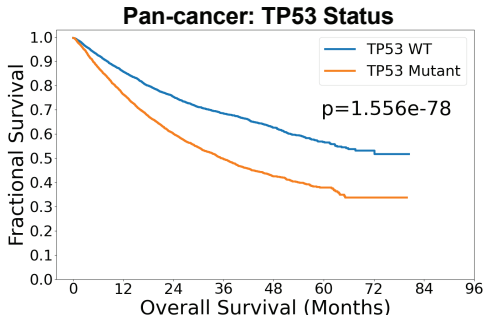
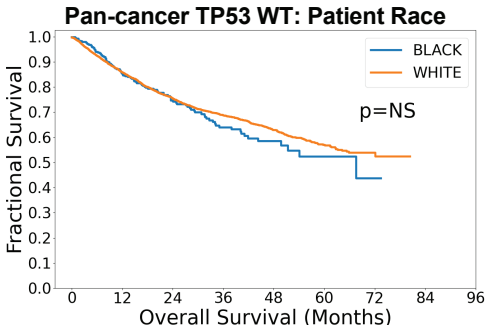

A
Breast: Patient Race
BLACK
WHITE
p=NS
Fractional Survival
Overall Survival (Months)
B
Endometrial: Patient Race
BLACK
WHITE
p=0.048
Fractional Survival
Overall Survival (Months)
C
NSCLC: Patient Race
BLACK
WHITE
p=NS
Fractional Survival
Overall Survival (Months)
D
Endometrial: WGD Status
No WGD
WGD
p=1.698e-11
Fractional Survival
Overall Survival (Months)
E
Endometrial: Race (No WGD)
BLACK
WHITE
p=0.044
Fractional Survival
Overall Survival (Months)
F
Endometrial: Race (WGD)
BLACK
WHITE
p=NS
Fractional Survival
Overall Survival (Months)

Figure S7

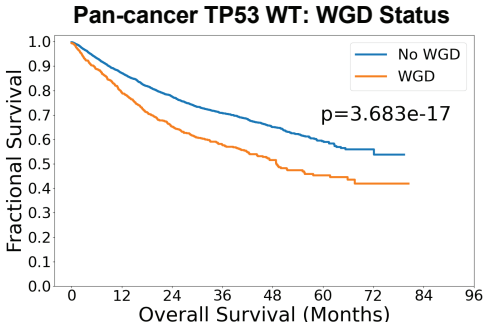
A



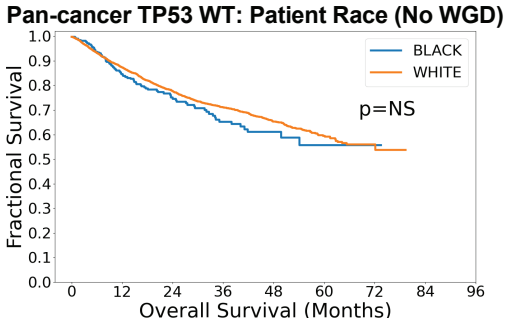
B



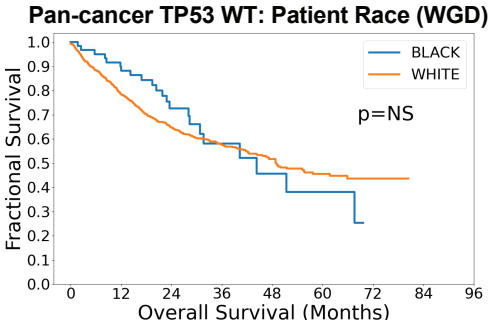
C



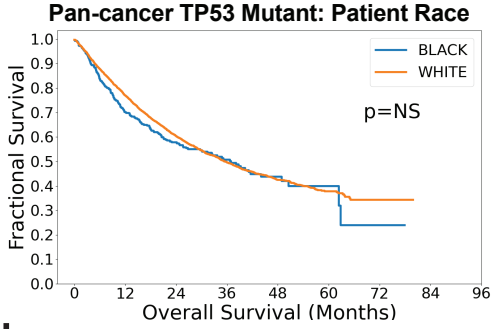
D



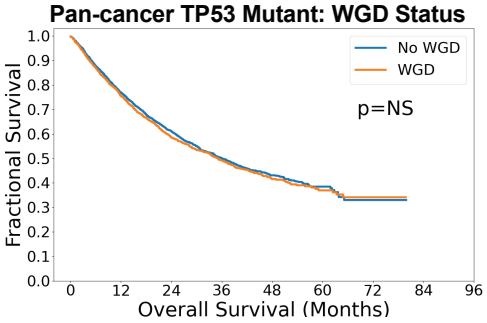
E



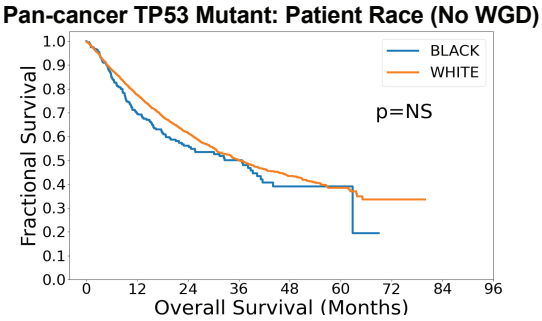
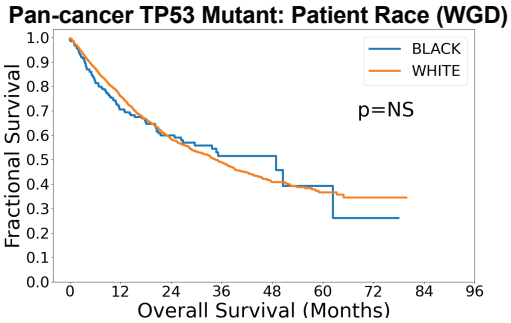
F



G



H



I