# Understanding Ovarian Cancer Risk: Pathogenic *TP53* mutation burden in peritoneal fluid is associated with *BRCA* germline mutation status

Xin Ray Tee

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Committee:

Rosana Risques

Steve Salipante

Barbara Norquist

Program Authorized to Offer Degree:

Laboratory Medicine & Pathology

University of Washington

# Abstract

Understanding Ovarian Cancer Risk: Pathogenic *TP53* mutation burden in peritoneal fluid is associated with *BRCA* germline mutation status

Xin Ray Tee

Chair of the Supervisory Committee:

Rosana Risques

Department of Laboratory Medicine & Pathology

Individuals with germline *BRCA1* and *BRCA2* mutations (*BRCA* carriers) have an increased risk of developing high-grade serous ovarian cancer (HGSOC), the most common and deadliest subtype of ovarian cancer. While HGSOC is initiated by *TP53* mutations, these mutations are also now recognized as a common feature of normal human aging. We hypothesize that *BRCA* carriers harbor increased *TP53* mutation burden, underlying their predisposition to HGSOC. We used high-fidelity, high-depth duplex sequencing to analyze mutations in *TP53* in Paps and peritoneal fluid of 25 *BRCA1* carriers, 21 *BRCA2* carriers, and 11 *BRCA* non-carriers. Mutations were annotated and compared across groups, sample types, and with the *TP53* UMD HGSOC database. *TP53* coding and pathogenic mutation burden were associated with age in Paps and peritoneal fluid of all individuals. *BRCA1* carriers older than 45 years old had a significantly higher pathogenic mutation burden than *BRCA2* carriers and controls of the same age in peritoneal fluid, but not in Paps, Peritoneal fluid from *BRCA1* carriers older than 45 years old had the most significant enrichment of hotspot *TP53* mutations with 36.2% of all mutations

identified in hotspot codons compared to 7.7% expected by chance ($p = 5x10^{-8}$). 2.6% (6 out of 230) of peritoneal fluid *TP53* variants were also found in Paps but not in blood. *BRCA1* carriers harbor an excess of *TP53* pathogenic mutations in peritoneal fluid after age 45, coinciding with the age of increased HGSOC risk. Paps have no resolution to detect differences in *TP53* mutation burden due to germline status.

## Introduction

High-grade serous ovarian cancer (HGSOC), which encompasses cancer identified on the ovary, fallopian tube, and peritoneum, is the most common subtype of ovarian cancer as well as the deadliest, accounting for 70-80% of all ovarian cancer deaths [1]. This high mortality rate is likely due to the indistinct symptoms experienced during disease progression as well as the lack of an effective cancer screening test, leading to the majority of HGSOC being diagnosed at advanced stages [2]. The overall 5-year survival of advanced-stage disease is only 30% in contrast to the >90% survival for early-stage diagnosis [3]. Unfortunately, *BRCA1* and *BRCA2* germline mutation carriers (*BRCA* carriers) are at a disproportionately higher risk of developing HGSOC than the general population; with a cumulative 44% risk for *BRCA1* and 17% for *BRCA2* carriers [4] compared to ~1.4% risk for the general population [5]. To date, there is no good screening test for detecting HGSOC or its precursor lesions, even though patients who are diagnosed at early stages of HGSOC have a higher survival rate compared to late-stage patients [6]. The current clinical preventative recommendation for *BRCA1* carriers above age 35 and *BRCA2* carriers above age 40 is to undergo risk-reducing salpingo-oophorectomy, a prophylactic surgery that involves the surgical removal of the fallopian tube and ovaries. Prophylactic surgery was associated with up to 80% reduction in ovarian cancer risk in *BRCA* carriers [7], but the procedure leads to premature menopause, reliance on hormonal therapy, and loss of reproductive ability, causing patients to often postpone the procedure and leading to increasing of overall risk. Thus, there is a critical need to develop approaches for the early detection and prediction of ovarian cancer for the general population, but especially for *BRCA* carriers who could greatly benefit from a personalized estimate of cancer risk to develop prevention strategies tailored to their specific needs.

The majority of HGSOCs both in individuals with *BRCA* mutations and sporadic cases originate in the distal fallopian tube (fimbriae) [8, 9], which is in direct continuity with the peritoneal cavity. *TP53* is mutated in the vast majority of HGSOC, with percentages ranging from 80% to 97% of tumors, depending on the study [1, 10]. The earliest putative precursor lesion of HGSOC is called p53 foci (or p53 signatures), which are clusters of fallopian tube epithelial cells with p53 overexpression but are otherwise histologically normal [11]. These signatures may take a long time (up to 20 years or more), before they develop into serous tubal intraepithelial carcinoma (STIC), a recognized precursor of HGSOC [12], which in turn may

progress to invasive carcinoma in 6-7 years [13-15]. It is thought that the metastatic behavior of these cancer lesions is acquired during the development of STIC into HGSOC, but interestingly, not all HGSOCs arise from STIC lesions, even in high-risk women. STIC was observed only in 11-61% (mean 31%) of HGSOCs [16], and some STICs appear to be metastases from HGSOCs rather than the precursors [17]. To reconcile these results, a "precursor escape" hypothesis for HGSOC carcinogenesis has been proposed, where early p53 foci cells first disseminate into the peritoneal cavity, before seeding onto the ovaries and peritoneum to further develop into HGSOC [18]. Under this hypothesis, investigation of potentially exfoliated *TP53* mutant cells into the gynecological tract may provide an opportunity for early cancer detection and interception.

In the last decade, advances in NGS technologies have enabled the detection of rare mutant cancer cells in gynecological liquid biopsies including Paps [19-21], peritoneal fluid [22], and uterine lavage [23-25]. Our group leveraged the high fidelity of duplex sequencing [25, 26] to demonstrate that HGSOC mutations could be detected in the peritoneal fluid of most patients with HGSOC (94%) and at frequencies as low as one in 24,736 normal genomes, and in the large majority of samples from patients with HGSOC (94%) [22]. Subsequently, we demonstrated that HGSOC *TP53* driver mutations could also be detected in 80% of uterine lavages [25] and 33% of Paps [21], the latter in agreement with prior studies [20, 27]. Interestingly, all these studies also revealed the presence of abundant *TP53* mutations (not tumor mutations) that increased with age, tended to occur in common cancer hotspot codons, and were largely pathogenic, indicating that these mutations had been selected and clonally expanded in normal tissue. These mutations occurred in individuals with and without HGSOC, but their mutational burden was higher in those with HGSOC [21, 22], suggesting an association between *TP53* clonal expansions and the development of HGSOC. Collectively, these results are consistent with a growing body of literature demonstrating the presence of cancer mutations in non-cancerous tissue, which is now understood as a manifestation of somatic clonal evolution [28-30]. *TP53* mutations appear to be ubiquitous in clonal expansions of multiple human tissues, including blood, skin, esophagus, and endometrium [28, 31].

To date, however, it is unknown whether *TP53* clonal expansions are elevated in individuals at high risk for *TP53*-driven cancer. In this exploratory study, we aim to answer that question in the context of *BRCA1* and *BRCA2* germline mutation carriers. We hypothesize that

the risk of developing HGSOC is related to the burden of pathological *TP53* mutations in the gynecological tract and thus, the burden will be higher in *BRCA* carriers than in individuals without *BRCA* germline mutations. Based on our success in prior studies using high fidelity sequencing to measure very low frequency *TP53* mutations in gynecological biopsies [21, 22, 25], we propose to use duplex sequencing to quantify *TP53* mutation burden in Pap test and peritoneal fluid DNA from individuals with and without *BRCA* germline mutations. While Paps are expected to have less sensitivity, they are easily accessible samples that could be valuable clinically if proven informative. In addition, having two different sample types enables a more comprehensive characterization of *TP53* mutations in individuals without HGSOC, including the determination of shared clones between samples, which is critical to our understanding of the transit of *TP53* mutant clones in the gynecological tract.

## Results

*Ultra-deep sequencing identifies higher TP53 coding mutation frequency than non-coding mutation frequency in Paps and peritoneal fluid from patients that underwent prophylactic surgery*

This cohort included 57 patients that underwent prophylactic surgery (risk-reducing salpingo-oophorectomy, RRSO) due to *BRCA* germline mutations or familial history of cancer (Figure 1A), of which there were 25 *BRCA1* carriers, 21 *BRCA2* carriers, and 11 patients with no *BRCA1/2* mutation (*BRCA wt*), as determined by BROCA [32] or other commercially available sequencing panels (see Methods). The three groups of patients had a similar range of ages to account for age-association of somatic mutations [30]. Paps were collected before surgery, and peritoneal fluid, which involves a saline wash of the peritoneum, was collected from the same individuals during surgery (Figure 1B). Clinico-pathological information of all patients is summarized in Supplementary Table S1.

We performed *TP53* ultra-deep sequencing in Paps (mean duplex depth: 4913x, range: 2285x – 8736x) and peritoneal fluid (mean duplex depth: 9363x, range: 5905x – 18023x) of all 57 individuals. Duplex depth between groups and between samples within each group was comparable (Figure 1C, Suppl. Fig S1). Overall, we identified a total of 1104 *TP53* variants, with 674 in Paps and 430 in peritoneal fluid. All variants were present at very low frequency

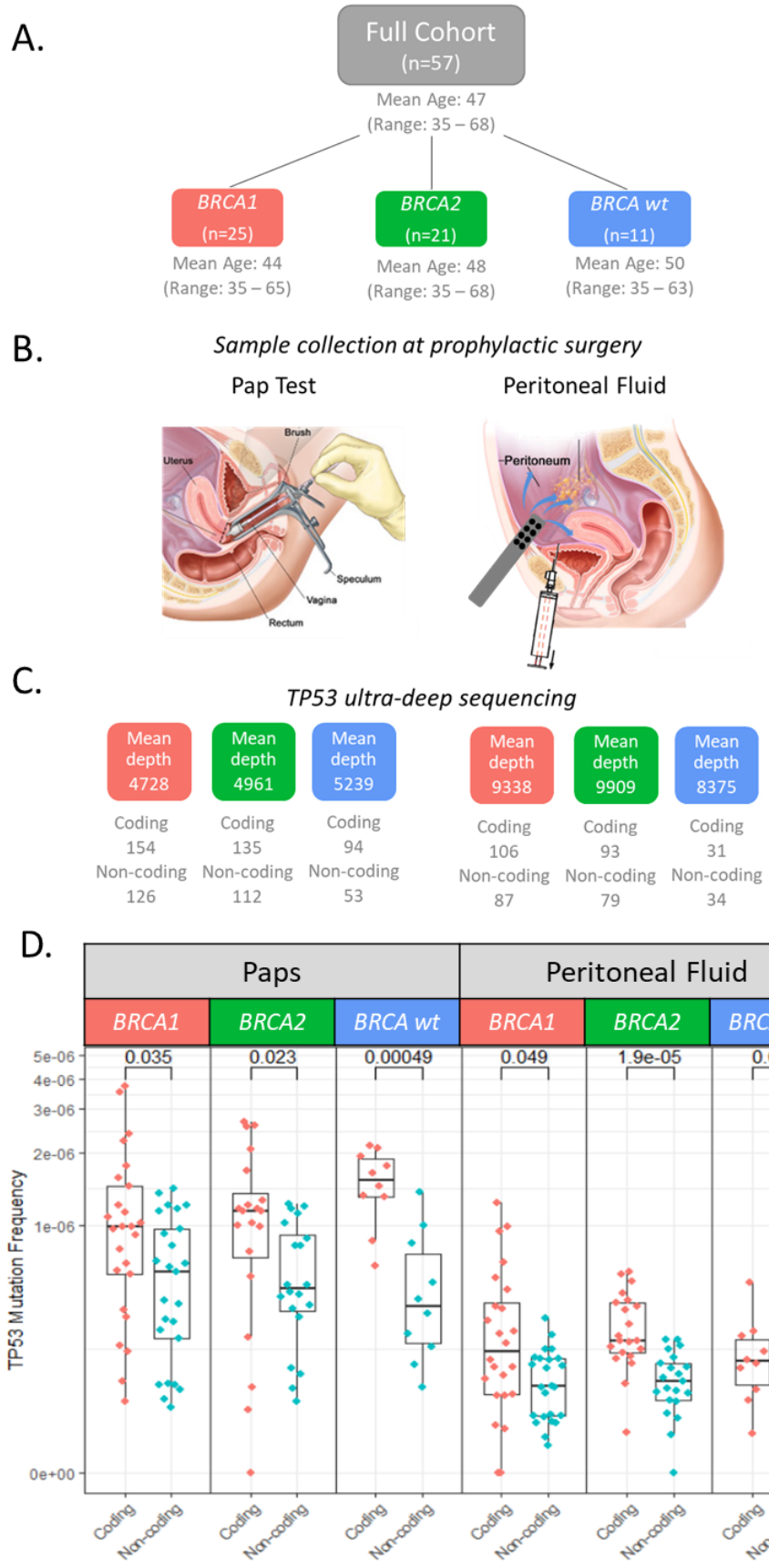



**Figure 1. Ultra-deep sequencing identifies abundant *TP53* mutations in Pap tests and peritoneal fluid from patients with and without *BRCA* germline mutations. A.** Cohort description. 57 patients undergoing risk-reducing salpingo-oophorectomy as prophylactic surgery for HGSOC were included in the study. Patients included *BRCA1* and *BRCA2* germline mutation carriers as well as patients without *BRCA* germline mutations (*BRCA wt*). **B.** Pap tests were collected pre-operatively (image from Winslow, 2009) and peritoneal fluid was collected at surgery (Image from Winslow, 2011). **C.** *TP53* ultra-deep sequencing revealed multiple *TP53* coding mutations and non-coding mutations in the 3 groups of patients and two sample types. **D.** Coding and non-coding mutation frequency plots for Pap tests and peritoneal fluid. For each sample, coding and non-coding mutation frequencies were calculated as the number of mutations divided by the total number of duplex nucleotides sequenced in coding and non-coding regions, respectively. Overlying box plots display the quartiles with whiskers extending up to 1.5x the interquartile range. p-values correspond to Mann-Whitney U tests.

(VAF<0.02) in all samples. After categorizing variants into coding and non-coding, mutation frequencies (MF) were calculated by dividing the number of coding and non-coding mutations by the total number of duplex nucleotides sequenced in the coding and non-coding regions, respectively (Figure 1C). Coding MF was higher than non-coding MF in both Paps and peritoneal fluids in the three groups of patients, reaching significance ($p<0.05$) in all comparisons except for the peritoneal fluids of the *BRCA* non-carriers (Figure 1D). The excess of *TP53* coding mutations indicates selection of functional changes as opposed to non-coding mutations. The list of all coding variants for Paps (n=383) and peritoneal fluids (n=230) are included in Supplementary Tables S2 & S3, respectively. Paps and peritoneal fluid MF values for each sample are shown in Supplementary Tables S4 & S5, respectively.

*Coding TP53 mutations in Paps and peritoneal fluid are found in most individuals and are frequently pathogenic*

We next investigated how *TP53* coding mutations were distributed across patients, and whether there were any associations with clinico-pathological variables (Figure 2A). A subset of patients had a history of breast cancer, most of whom underwent chemotherapy, but we did not observe higher *TP53* MF in association with prior breast cancer or chemotherapy in Paps or peritoneal fluids (Suppl. Fig S2A). We also inspected the association with *TP53* mutational burden (MB), which was calculated by dividing the number of duplex mutant reads by the total number of duplex nucleotides sequenced in coding regions. In duplex sequencing, each mutant duplex read corresponds to a single DNA molecule, thus the total number of mutant reads for a given variant is proportional to the size of the mutant clone. While MF measures the number of mutant variants, MB captures both the number of mutant variants and the size of mutant clones. Similar to MF, *TP53* MB was also not associated with a history of breast cancer or prior chemotherapy in any of the groups of patients (Suppl. Fig S2B). CA-125 measurements, a well-established ovarian cancer serum biomarker, were available for most patients. There was a single patient (P54) with CA-125 levels >35, the established clinical threshold, and this patient carried a high number of *TP53* mutations both in Paps and peritoneal fluid (Suppl. Table S2 & S3). Overall, however, there was no significant correlation between CA-125 levels and elevated *TP53* MF (Suppl. Fig S3A) or *TP53* MB (Suppl. Fig S3B).

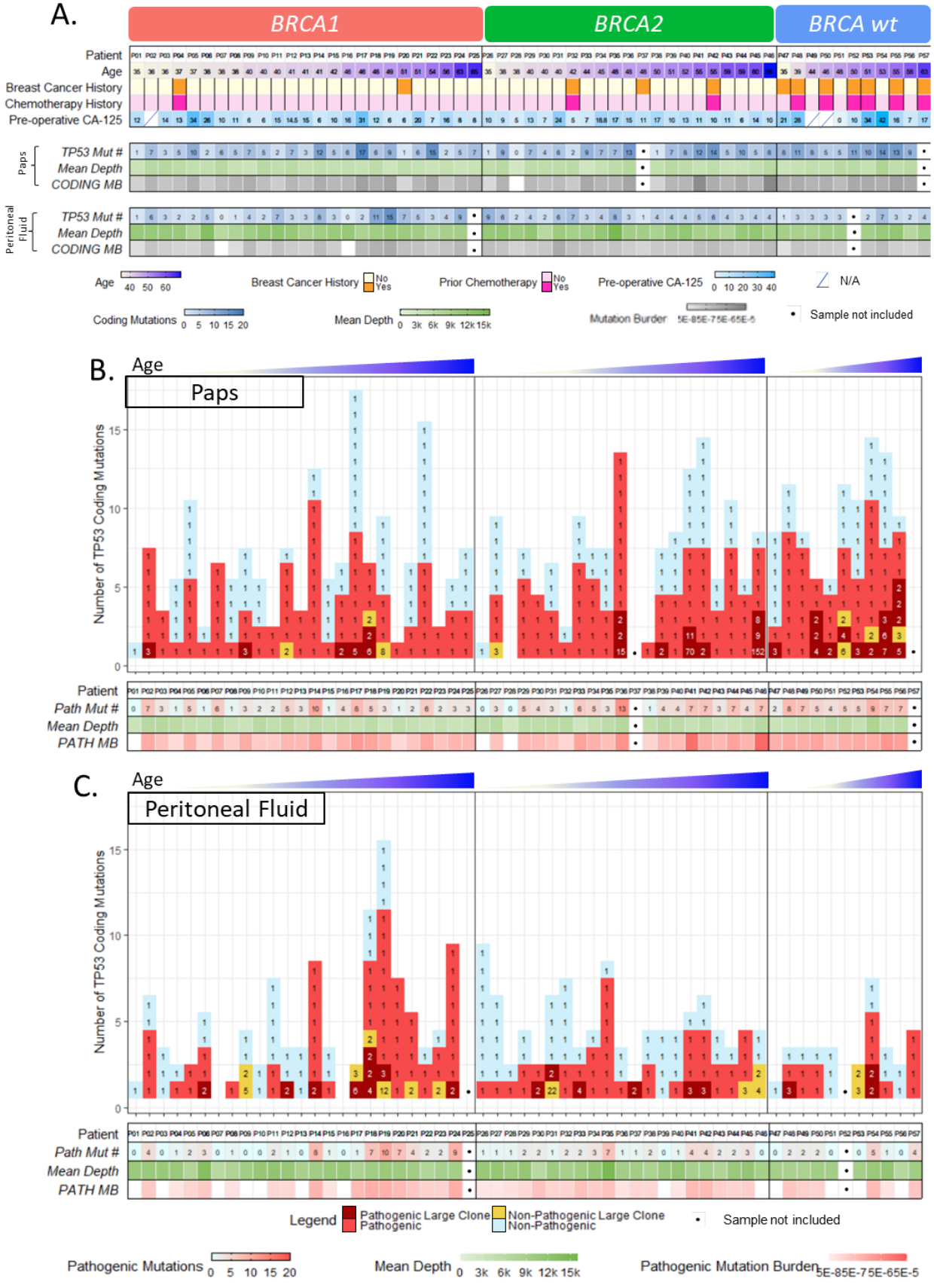

A.
BRCA1
BRCA2
BRCA wt
Patient
Age
Breast Cancer History
Chemotherapy History
Pre-operative CA-125
Paps
TP53 Mut #
Mean Depth
CODING MB
Peritoneal Fluid
TP53 Mut #
Mean Depth
CODING MB
Age 40 50 60
Breast Cancer History No Yes
Prior Chemotherapy No Yes
Pre-operative CA-125 0 10 20 30 40
N/A
Coding Mutations 0 5 10 15 20
Mean Depth 0 3k 6k 9k 12k 15k
Mutation Burder 5E-85E-75E-65E-5
Sample not included
B.
Age
Paps
Number of TP53 Coding Mutations
Patient
Path Mut #
Mean Depth
PATH MB
C.
Age
Peritoneal Fluid
Number of TP53 Coding Mutations
Patient
Path Mut #
Mean Depth
PATH MB
Legend
Pathogenic Large Clone
Pathogenic
Non-Pathogenic Large Clone
Non-Pathogenic
Sample not included
Pathogenic Mutations 0 5 10 15 20
Mean Depth 0 3k 6k 9k 12k 15k
Pathogenic Mutation Burden 5E-85E-75E-65E-5

**Figure 2. Coding *TP53* mutations in Paps and peritoneal fluid are found in most individuals and are frequently pathogenic and present as large clones. A.** Graphical representation of relevant clinical information (top) and mutational information from Paps (middle) and peritoneal fluid (bottom). Each column corresponds to a patient. Patients are grouped by germline mutation and sorted by ascending age. Mutational information for Paps and peritoneal fluid includes mutation counts (number of mutations indicated inside of squares), depth, and coding mutation burden (MB). Depth indicates average duplex depth for all positions sequenced. Coding MB was calculated as the total number of coding mutant duplex reads divided by the total number of duplex nucleotides sequenced in coding regions. Missing samples are indicated with a black dot, and missing clinical information is indicated by a diagonal line in the square. **B.** Catalog of *TP53* mutations identified in Pap DNA. Each square corresponds to a single mutation. The number in the square indicates the number of duplex reads carrying the mutation. Mutations are color-coded by pathogenicity and the number of mutant duplex reads >1, which is indicative of larger clones. Pathogenicity is defined by the Seshat algorithm [33]. *TP53* pathogenic mutation burden (Path MB) was calculated as the total number of pathogenic mutant duplex reads divided by the total number of duplex nucleotides sequenced in coding regions. **C.** Catalog of *TP53* mutations identified in peritoneal fluid DNA. Legends as in B.

To gain further insight into the nature of the *TP53* mutations identified in Paps and peritoneal fluid, we next classified them into pathogenic and non-pathogenic based on the Seshat algorithm for predicted pathogenicity [33]. In Paps, we observed pathogenic mutations in nearly all patients, without obvious differences between groups (Figure 2B). In peritoneal fluids, however, pathogenic mutations were less abundant than in Paps and appeared to cluster in *BRCA1* mutation carriers of older age (Figure 2C). In addition, large clones (2 or more mutant duplex reads) of pathogenic and non-pathogenic mutations also appeared to be more abundant in peritoneal fluid from *BRCA1* mutation carriers of older age. To enable a formal quantification of these observations, we calculated Pathogenic *TP53* MB for each sample by considering only duplex reads from pathogenic mutations. This metric integrates the information of pathogenicity and mutation burden and was used together with *TP53* MB for further analyses.

*TP53 mutation burden in Paps and peritoneal fluid is associated with age*

Next, we investigated the association between *TP53* MB and Pathogenic MB with age. When considering all patients, MB was associated with age in Paps ($p=0.0023$ by Spearman's rank correlation test) and peritoneal fluids ($p=0.028$ by Spearman's rank correlation test) (Figure 3A). Similar trends were observed when separating by patient groups, although associations were not always significant (Figure 3B). Pathogenic MB similarly increased with age in Paps and peritoneal fluid of all patients (Suppl. Fig S4A) and across some groups of patients (Suppl. Fig S4B).

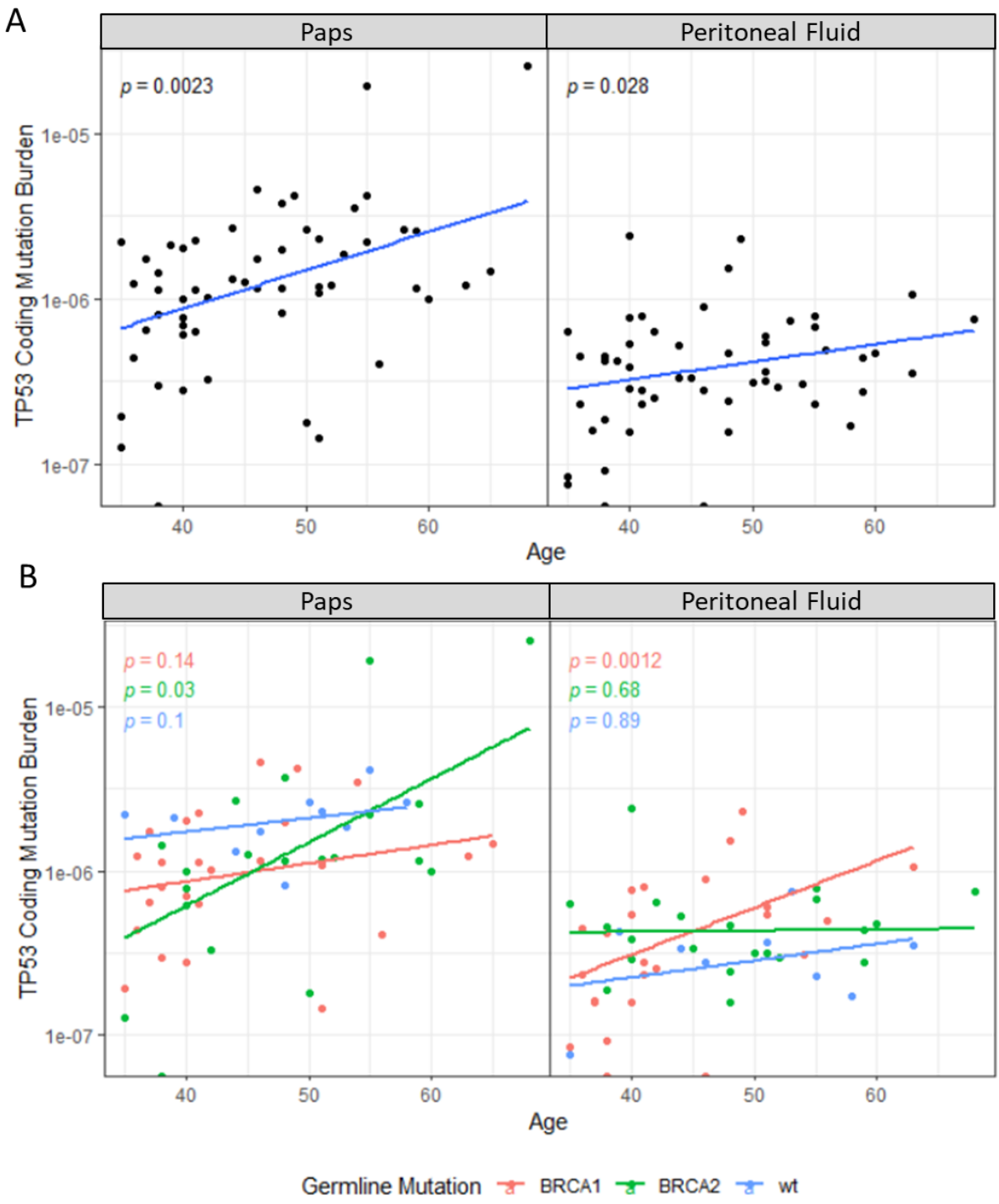



**Figure 3. *TP53* mutation burden is associated with age in both Paps and peritoneal fluids. A.** Age is correlated with *TP53* coding mutation burden identified in Paps (left) and peritoneal fluids (right). *TP53* mutation burden was calculated as the total number of mutant duplex reads divided by the total number of duplex nucleotides sequenced in the *TP53* coding region. p-values were calculated with Spearman's rank correlation test. **B.** Correlation plot of age and *TP53* mutation burden color-coded by *BRCA1* carriers, *BRCA2* carriers, and *BRCA* non-carriers (*wt*) for Paps (left) and peritoneal fluids (right). p-values were calculated with Spearman's rank correlation test.

*Pathogenic TP53 mutation burden is higher in the peritoneal fluid of BRCA1 carriers older than 45 years*

To get a better understanding of the differences in *TP53* MB and Pathogenic *TP53* MB across groups, we next plotted these two variables for each patient by ascending age (Figure 4A and Suppl. Fig. S5A). We observed that, in peritoneal fluid, Pathogenic *TP53* MB experienced a sharp increase around age 45 (Figure 4A). Interestingly, the decade between 40 to 50 years of age is when the risk for ovarian cancer starts to increase in *BRCA1* carriers, according to epidemiological studies [4]. Thus, we decided to use the cut-off age of 45 years to determine differences in *TP53* MB across groups. We observed that *TP53* MB in peritoneal fluid is increased in *BRCA1* carriers older than 45 years of age compared to *BRCA2* carriers and non-carriers, but none of the comparisons reached statistical significance (Suppl Fig. S5B). For pathogenic *TP53* MB in peritoneal fluid, however, *BRCA1* mutation carriers showed significantly higher levels than non-carriers ($p=0.029$) and *BRCA2* carriers ($p=0.036$) (Figure 4B). In Paps, the levels of Pathogenic *TP53* MB were overall higher than in peritoneal fluids but showed no significant differences between groups (Figure 4B). These results indicate that around the age of 45 when the risk of ovarian cancer increases for *BRCA*1 carriers [4], there is an increase of Pathogenic *TP53* MB in peritoneal fluid, which might underlie the high risk of ovarian cancer progression in this group of individuals.

*TP53 mutations in Paps and peritoneal fluids resemble ovarian cancer mutations*

Next, we compared the type, pathogenicity, and spectrum of *TP53* mutations in Paps and peritoneal fluids with the catalog of *TP53* mutations identified in HGSOC from the UMD database (n=7700). The UMD database is used by the Seshat tool [33] and includes the Catalog of Somatic Mutation in Cancer (COSMIC) data as well as other large catalogs of cancer mutations (see Methods). In addition, we also compared mutation type, pathogenicity, and spectrum with an in-silico model that emulated all possible substitutional mutation events in the *TP53* coding region (n=3546) (See Methods). This model represents a situation where all substitutions are possible with no selective pressure, whereas the cancer data summarizes the features of selected *TP53* mutations in HGSOC.

We observed that the distribution of mutation types in Paps and peritoneal fluid more closely resembled the distribution in cancer than in the absence of selection (Figure 5A). For

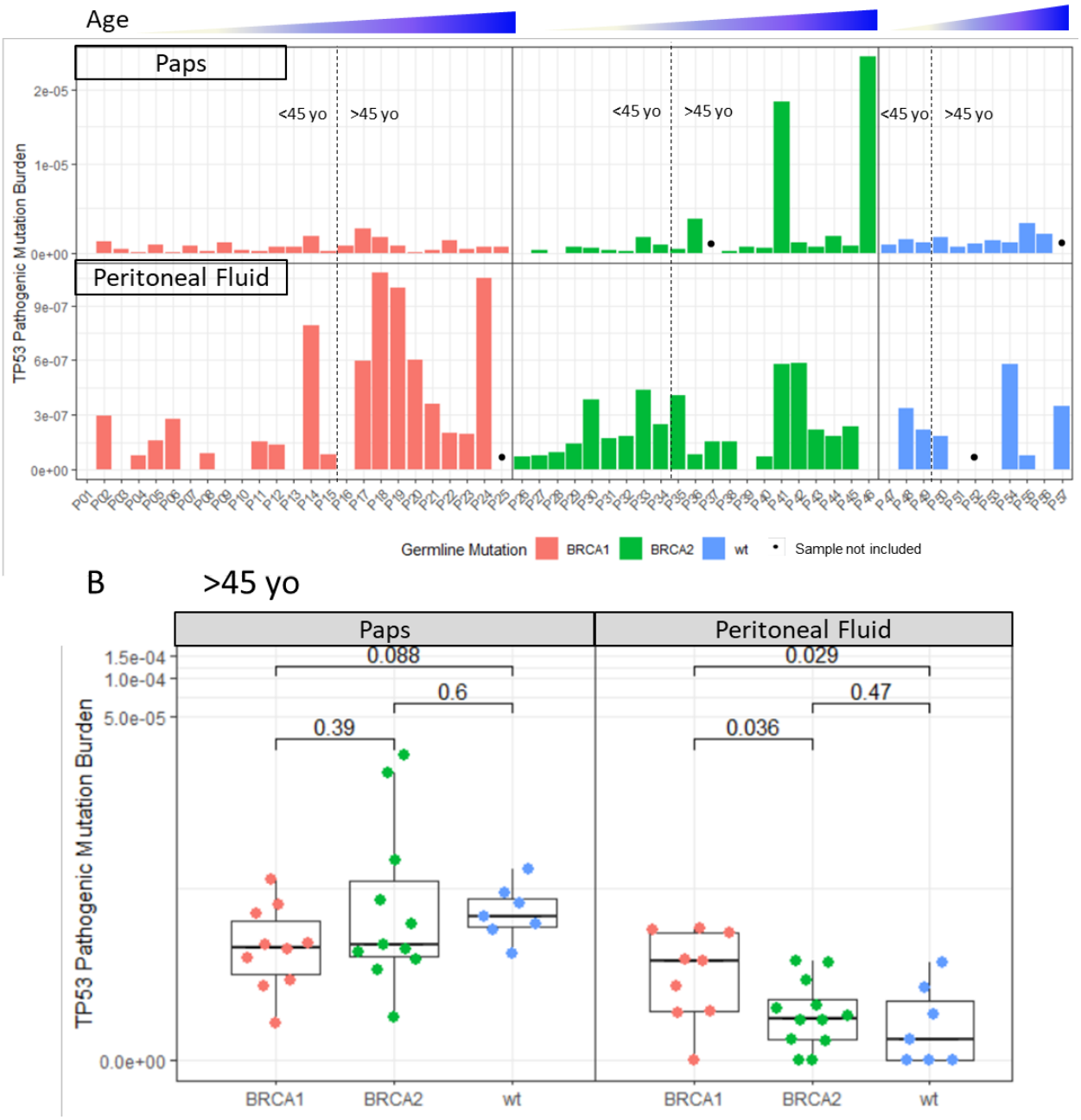


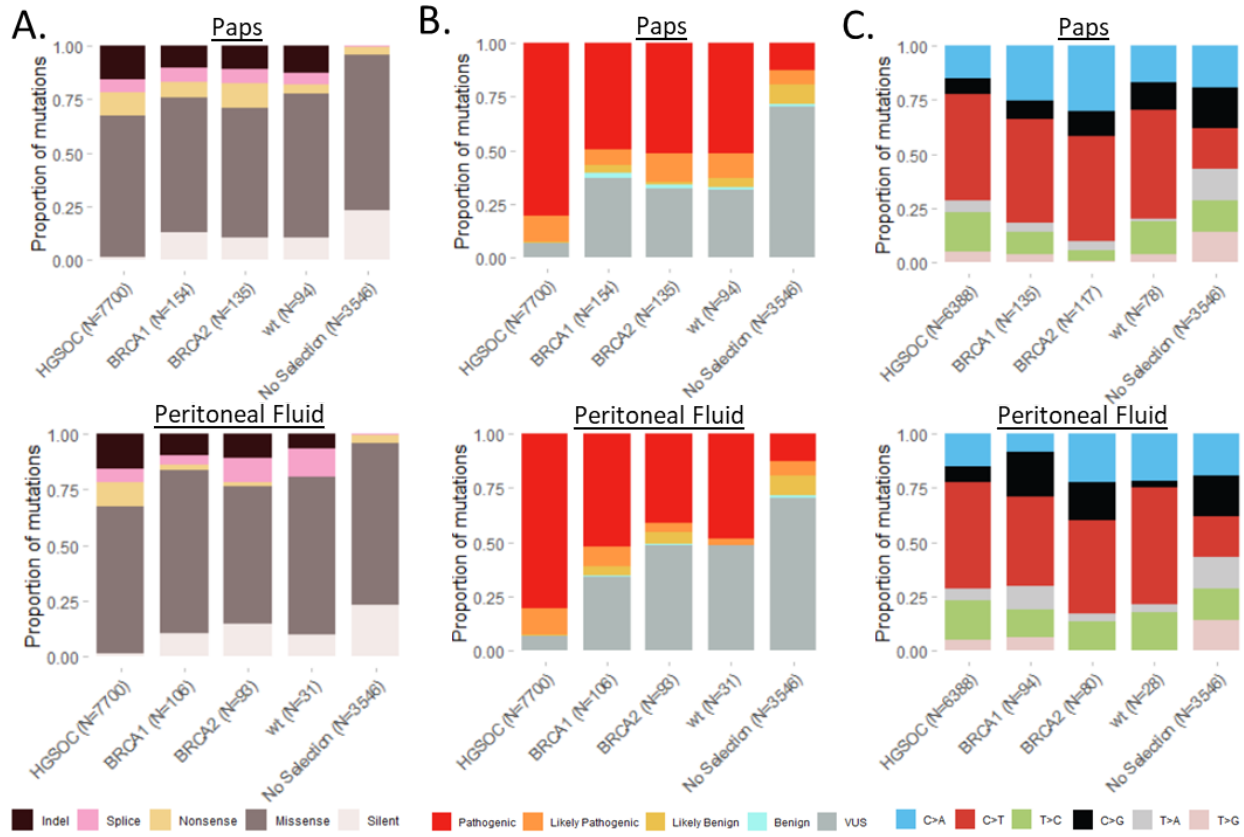

**Figure 4. *TP53* pathogenic mutation burden is increased in *BRCA1* carriers older than 45 years of age. A.** *TP53* pathogenic mutation burden of Paps (top) and peritoneal fluid (bottom) plotted per individual. Each column corresponds to a patient. Patients are grouped by germline mutation and sorted by ascending age. *TP53* pathogenic mutation burden was calculated as the total number of pathogenic mutant duplex reads divided by the total number of duplex nucleotides sequenced in coding regions. Pathogenicity was defined by the Seshat algorithm [33]. Dashed lines indicate age 45 cutoffs. Missing samples are indicated with a black dot. **B.** Comparison of *TP53* pathogenic mutation burden between *BRCA1* carriers, *BRCA2* carriers, and *BRCA* non-carriers (*wt*) older than 45 years in Paps (left) and peritoneal fluid (right). Overlying box plots display the quartiles with whiskers extending up to 1.5x the interquartile range. p-values correspond to Mann-Whitney U tests.

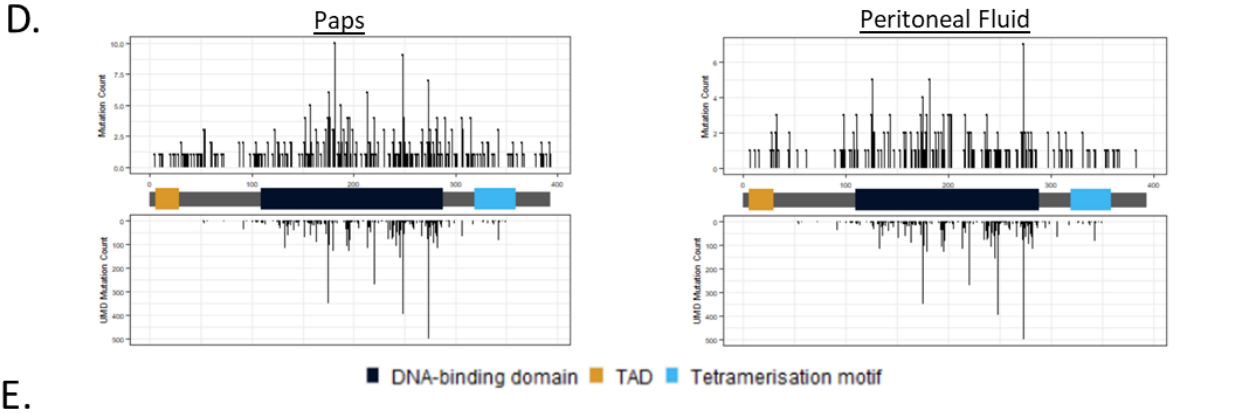
both sample types, the 3 groups of patients carried substantial amounts of indels, splice, and/or nonsense mutations, in proportions approaching the observed distribution of *TP53* mutation types in cancer.

The distribution of pathogenic mutations in the samples of the study was an intermediate between what was expected without selection and what is seen in cancer (Figure 5B). Similarly, the spectrum distribution of *TP53* mutations in Paps and peritoneal fluids was also an intermediate between the extremes of no selection and cancer, with an expected enrichment of "C>T" mutational events, a large portion of which may be age-related [34] (Figure 5C).

To gain further insight into the similarity of Paps and peritoneal fluid *TP53* mutations with HGSOC mutations, we plotted the location of all the identified substitutions along the coding region of the gene and compared this with the distribution of HGSOC *TP53* mutations in the UMD database. As expected, HGSOC mutations clustered in hotspot codons in the DNA binding domain coding region (exons 5 to 8). Paps and peritoneal fluid mutations also tended to cluster in the same areas (Figure 5D).

Given the age effect on *TP53* clonal expansions, we next compared the proportion of *TP53* substitution mutations occurring at hotspot codons (hotspot mutations) in Paps and peritoneal fluids with the catalog of *TP53* mutations identified in HGSOC from the UMD database and the "No Selection" model (Figure 5E) among individuals younger and older than 45 years of age. Hotspot codons were defined as codons with at least 1% of reported substitution in HGSOC from the UMD (See Methods). Among Paps, we observed a significant difference between the proportion of hotspot mutations with the "No Selection" model, without obvious differences between groups among individuals older or younger than 45 (Figure 5E). In peritoneal fluids, however, among individuals younger than 45 years of age, only *BRCA1* carriers had a significant difference in the proportion of *TP53* hotspot mutations compared to the "No Selection" model (23.4% of hotspot mutations in *BRCA1* carriers <45yo vs 7.7% expected by chance, p=0.001 by Fisher's exact test). Among individuals 45 years and older, proportions of hotspot mutations were significantly different from the "No Selection" model for individuals of all groups, but the highest percentage and most significant comparison was observed for *BRCA1* carriers (36.2% of hotspot mutations in *BRCA1* carriers >45yo vs 7.7% expected by chance, $p=5x10^{-8}$ by Fisher's exact test) (Figure 5E). Overall, these results agree with prior observations


A.
Paps
Peritoneal Fluid
B.
Paps
Peritoneal Fluid
C.
Paps
Peritoneal Fluid
Proportion of mutations
HGSOC (N=7700)
BRCA1 (N=154)
BRCA2 (N=135)
wt (N=94)
No Selection (N=3546)
HGSOC (N=6388)
BRCA1 (N=135)
BRCA2 (N=117)
wt (N=78)
BRCA1 (N=106)
BRCA2 (N=93)
wt (N=31)
BRCA1 (N=94)
BRCA2 (N=80)
wt (N=28)
Indel
Splice
Nonsense
Missense
Silent
Pathogenic
Likely Pathogenic
Likely Benign
Benign
VUS
C>A
C>T
T>C
C>G
T>A
T>G


D.
Paps
Peritoneal Fluid
Mutation Count
UMD Mutation Count
DNA-binding domain
TAD
Tetramerisation motif

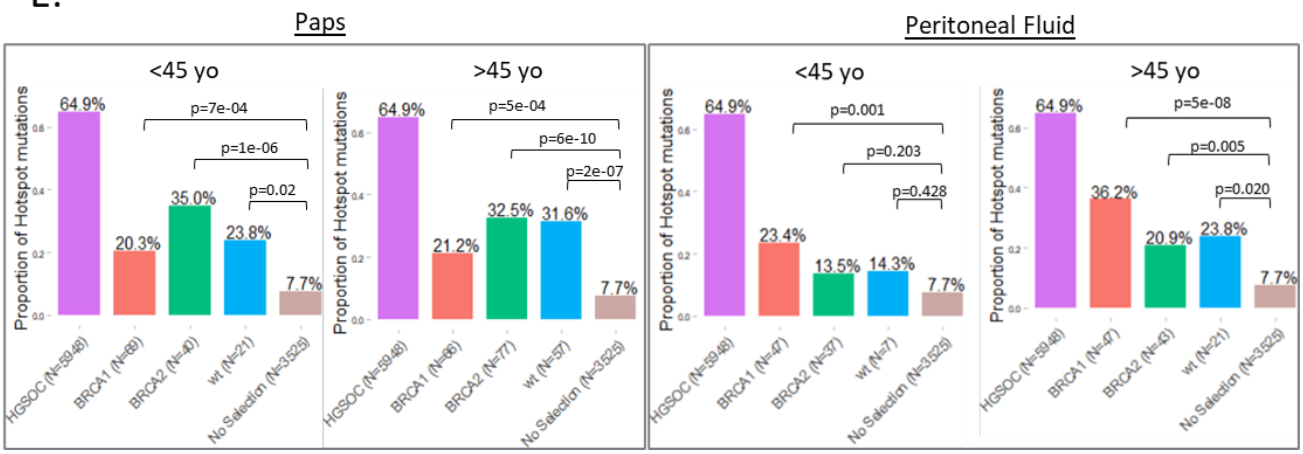

E.
Paps
Peritoneal Fluid
<45 yo
>45 yo
Proportion of Hotspot mutations
64.9%
20.3%
35.0%
23.8%
7.7%
p=7e-04
p=1e-06
p=0.02
21.2%
32.5%
31.6%
p=5e-04
p=6e-10
p=2e-07
23.4%
13.5%
14.3%
p=0.001
p=0.203
p=0.428
36.2%
20.9%
23.8%
p=5e-08
p=0.005
p=0.020
HGSOC (N=5948)
BRCA1 (N=69)
BRCA2 (N=40)
wt (N=21)
No Selection (N=3525)
BRCA1 (N=66)
BRCA2 (N=77)
wt (N=57)
BRCA1 (N=47)
BRCA2 (N=37)
wt (N=7)
BRCA2 (N=43)

**Figure 5. *TP53* mutations in Paps and peritoneal fluid resemble cancer mutations. A.** Distribution of mutation type in Paps (top) and peritoneal fluids (bottom) of *BRCA1* carriers, *BRCA2* carriers, and *BRCA* non-carriers compared to *TP53* mutations identified in HGSOC (UMD database, n=7700) and all possible mutations in the gene ("No selection" model, n=3546, see Methods). **B.** Distribution of pathogenicity of all *TP53* mutations in Paps (top) and peritoneal fluids (bottom) for the same groups as in A. **C.** Mutational spectrum comparisons of *TP53* coding nucleotides for the same groups as in A. **D.** *TP53* mutational distribution maps for Paps and peritoneal fluids compared to *TP53* cancer mutations. Upper panels show the location by codon number of *TP53* mutations identified in Paps (left) and peritoneal fluid (right), with the mutational count on the y-axis. Bottom panels show the codon location of *TP53* HGSOC mutations from the UMD database (n=7700). Indels are excluded from plots as they may span multiple codons. *TP53 g*ene domains are color-coded and indicated in the legend. **E.** Proportion of substitution mutations identified at hotspots codons in Paps and peritoneal fluid of individuals younger than 45 years of age (left panel) and older than 45 years of age (right panel). Hotspots codons were defined as codons accounting for at least 1% of reported substitutions in HGSOC (UMD database, n=5948, see Methods). p-values correspond to Fisher's exact tests.

of clustering of *TP53* mutations from peritoneal fluid and uterine lavage in hotspot codons [22, 25], supporting the notion that *TP53* clonal expansions are similar to those driving cancer even in individuals without cancer. Our data expands these prior results by revealing that the similarity to cancer is higher in the peritoneal fluid of individuals carrying a *BRCA1* germline mutation.

*TP53 mutations clones are shared across multiple sample types*

A unique feature of this study is the analysis of two different, but related gynecological samples obtained from the same individuals at identical time points. Thus, we aimed to investigate the overlap of *TP53* mutant clones between samples, which could potentially be indicative of early dissemination of mutant clones into the gynecological tract and the peritoneum. A total of 12 shared mutations between Paps and peritoneal fluid were identified, 9 variants that originated from 6 *BRCA1* carriers (25% of the *BRCA1* cohort), 2 variants that originated from 2 *BRCA2* individuals (10% of the *BRCA2* cohort), and 1 variant from 1 one the controls (10% of the *BRCA* non-carrier cohort) (Figure 6A). While these mutations could indicate the exfoliation of *TP53* mutant clones of tubal epithelial origin, they could also have originated in leukocytes, reflecting clonal hematopoiesis [35].

To further investigate this, for individuals who had shared clones between Paps and peritoneal fluid (N=9), we duplex sequenced blood at a depth comparable to Paps (mean depth: 5553x, range: 2632x – 7983x) (Suppl. Table S6). Within each individual, the percent of mutations shared across any two pairs of sample types was small (4-20%) (Figure 6B). Overall, out of 54 peritoneal fluid mutations identified in 9 individuals, only 6 were identified in Paps (11%), 4 were identified in blood (7%) and 6 (11%) were identified in Paps and blood, (Figure 6A & 6B), indicating that clonal hematopoiesis may have contributed minimally to mutations

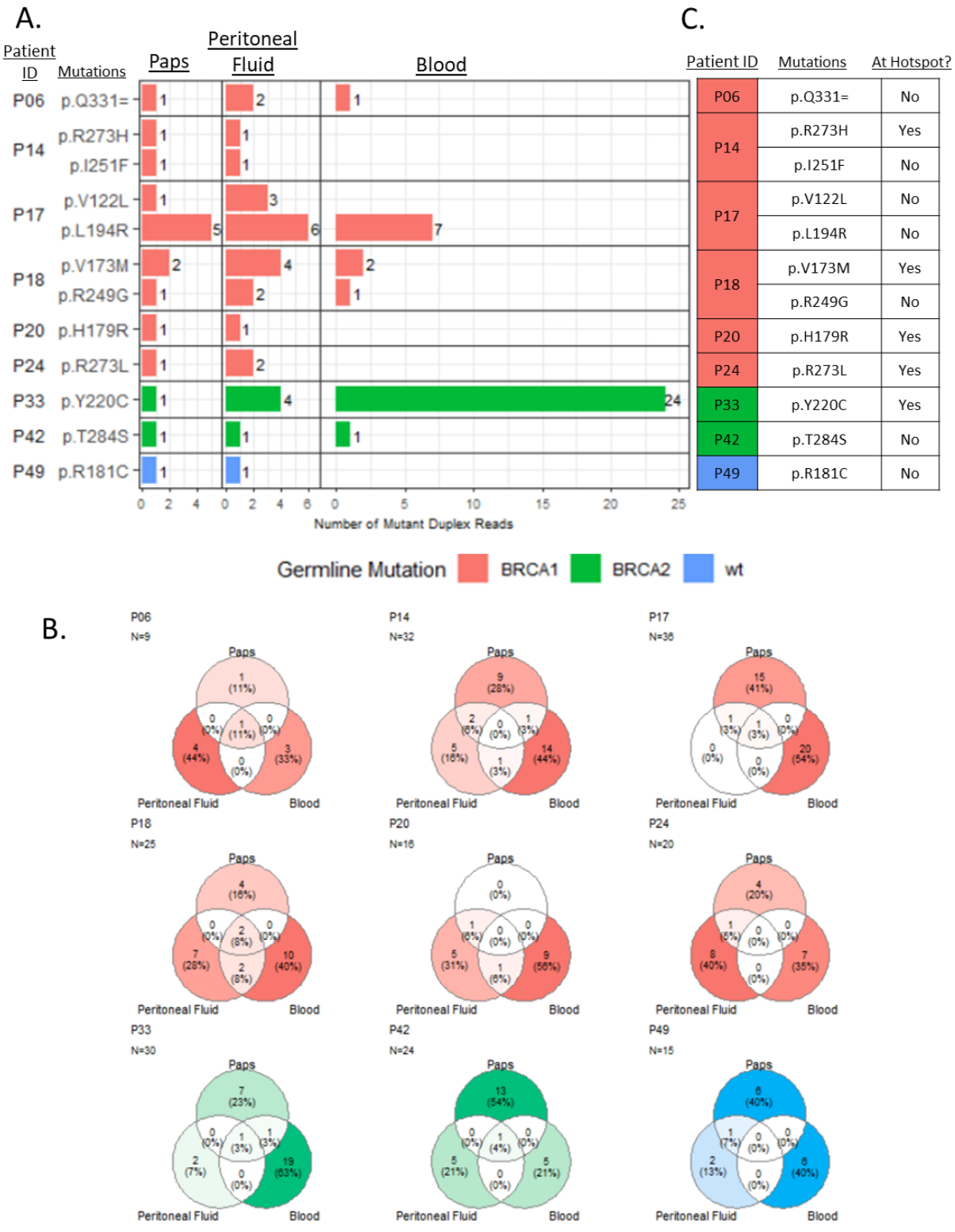



**Figure 6. Shared *TP53* mutations in Paps and peritoneal fluids. A.** *TP53* mutations found in Paps and peritoneal fluids from the same individuals are listed, grouped by patient and germline mutation group. For each mutation, the number of mutant duplex reads is labeled right of each bar and also presented on the x-axis. For individuals with

shared mutations in Paps and peritoneal fluid, blood DNA was also sequenced and mutations corresponding to those found in Paps and peritoneal fluids are shown. Empty rows indicate that the specific mutation was not found in the blood of that individual. Patient IDs are listed on the left. **B.** Venn diagrams indicating the number and percentages of *TP53* mutations shared in Paps, peritoneal fluid, and blood for each patient listed in A. Colors correspond to germline mutation status with the saturation of color proportional to the number of mutations. **C.** Hotspot status of the codon position of each *TP53* mutation identified in Paps and peritoneal fluid, as listed in A. Hotspots were defined as codons accounting for at least 1% of reported substitutions in HGSOC (UMD database, n=5948, see Methods).

identified in gynecological samples. However, 6 out of the 12 overlapping mutations were only identified in Paps and peritoneal fluid, suggesting that shared mutations were derived from the same *TP53* mutant clone found in both sample types. Alternatively, parallel evolution resulting in the independent selection and clonal expansion of the same mutation could also explain these findings.

To explore this possibility, we checked whether the overlapping mutations were occurring in hotspots (Figure 6C). Three of the 6 overlapping mutations not found in blood were hotspots, which could arguably be explained by parallel evolution. That still leaves 3 mutations, 2 in *BRCA1* carriers and 1 in a non-carrier, that are rare but shared between Paps and peritoneal fluids. These results suggest that exfoliated *TP53* mutant cells in patients without malignancy might be able to travel within the peritoneal and uterine cavity, but additional studies with larger sample types and simultaneous sequencing of fallopian tube epithelium are needed to fully explore this possibility.

## Discussion

By using ultra-deep, ultra-accurate sequencing of *TP53* in peritoneal fluid and Paps from individuals at high risk of HGSOC, we were able to demonstrate an excess of pathogenic *TP53* clonal expansion in the peritoneal fluid of *BRCA1* carriers over the age of 45. *TP53* mutations were prevalent in both Paps and peritoneal fluids; associated with age in both sample types, and resembled mutations found in HGSOC. However, the peritoneal fluid of *BRCA1* carriers showed an excess of pathogenic *TP53* mutations after age 45, which coincides with the age of increased risk of HGSOC in the population of *BRCA1* carriers according to epidemiological studies [4]. While our study cannot prove causality, the association between pathogenic *TP53* clonal expansions and *BRCA1* germline mutation provides a plausible link to understanding the susceptibility to HGSOC in this population at risk.

Results from our prior studies have indicated that patients with HGSOC carry higher *TP53* mutational burden in peritoneal fluid [22] and Pap test DNA [21]. Although the main contributor to the higher MB is the cancer clone itself, these biopsies also contained *TP53* mutations not found in the cancer clone, indicating the presence of multiple somatic pathogenic clones in these patients. The results in this study reveal that, while *TP53* pathogenic clones are common in the peritoneal fluid of healthy individuals, they are more abundant in *BRCA1* mutation carriers above 45 years of age compared to *BRCA2* carriers and non-carriers, suggesting that these clones might be related to the increased risk of HGSOC in these individuals. Interestingly, the excess of pathogenic *TP53* mutant clones in peritoneal fluid from *BRCA* carriers coincides with the age where the increased risk of HGSOC in *BRCA1* carriers becomes apparent [4]. These results highlight the compounding effect of genetic risk and age in enhancing clonal expansions of pathogenic clones and suggest that these clones might be related to the increased risk of HGSOC in these patients.

One main unanswered question from this study is the origin of shared clones observed in both Paps and peritoneal fluids of these individuals. We have demonstrated that a minority of these clones (~18%) correspond to *TP53* mutations found in blood. *TP53* mutant clones in leukocytes are recognized as part of clonal hematopoiesis of indeterminate potential (CHIP) [35], and studies have revealed that if more sensitive sequencing is used, as done in this study, mutant clones are identified in many more individuals and at younger ages [36]. However, the vast majority of peritoneal fluid mutations (82%) were not identified in blood. Notably, a subset of mutations (N=6) was identified in Paps and peritoneal fluid but not in blood. Three mutations occurred in hotspots, suggesting parallel evolution as a possible origin, but this still leaves 3 less common *TP53* mutations found in both sample types. One possible explanation of the origin of these rare mutations is putative early precursors of HGSOC (p53 foci) that exfoliated from the fallopian tube and onto the peritoneal and uterine cavity. p53 foci have been identified in fallopian tubes of *BRCA* carriers without HGSOC in as many as 40.8% of *BRCA1* carriers and 35.5 % in *BRCA2* carriers [37]. However, these findings are likely underestimates, since the identification of p53 foci relies on visual examination of p53-stained tubal epithelium, with pathogenic alterations found as both overexpression and null expression of p53; the latter being especially challenging to detect. Given the increased frequency and burden of *TP53* mutations in the peritoneal fluid of *BRCA1* carriers and the pathogenesis of HGSOC [38], we postulate that

these shared clones are likely to be *TP53* mutant clones that have exfoliated from the tubal epithelium into the peritoneum and uterine cavity. Similarly, the "precursor theory" also postulates that p53 foci exfoliate into the peritoneum and down the gynecological tract, before seeding and further developing into HGSOC [18]. Prior studies have also indicated that the time interval of development from precursor *TP53* mutant clones may take up to decades before developing into STICs [12], further supporting the idea that these rare mutations may potentially be p53 foci.

The limited resolution of Paps in finding differences between *BRCA* carriers and non-carriers might be due to two factors. First, if the origin of enhanced *TP53* clonal expansions in *BRCA* carriers is the fallopian tube, the signal from Paps may be lower due to sampling at a distal site. Second, multiple studies have characterized the endometrium as a tissue with high levels of clonal expansions of common cancer genes (e.g., *KRAS, PIK3CA,* and *TP53*) [39-41], which are likely to have obscured the signal from the fallopian tubes. In fact, we did observe on average 4 times higher levels of *TP53* mutation burden in Paps compared to peritoneal fluid. We were interested in characterizing *TP53* mutations in Paps because it is an easily accessible sample whereas peritoneal fluid is surgically collected and thus not a viable sample type to monitor *TP53* clonal expansions in *BRCA* carriers. An alternative biopsy that could capture this high-risk *TP53* signature is uterine lavage [42] because it allows for sampling of clones in the uterine cavity and could be explored in the future as a possible biomarker for *TP53* clonal expansion quantification in individuals at high risk of HGSOC.

One potential confounder of our findings is that some patients who had a breast cancer history had undergone chemotherapy prior to collection of Pap test and prophylactic surgery. Prior studies have shown increased mutation frequency in leukocytes of patients that underwent prior chemotherapy [43, 44]. While we did not see an association between prior chemotherapy and prior breast cancer with *TP53* mutational frequency and burden in this study, this is likely due to the smaller number of patients in this group. In addition, the *BRCA* non-carriers are not truly low risk for HGSOC, as they also underwent prophylactic surgery for other familial risks, therefore a study of a less-invasive sample type, which includes true low-risk controls may provide a further understanding of *TP53* clonal expansions and HGSOC risk. Lastly, analysis of blood samples was only performed for individuals with overlapping clones in Paps and

peritoneal fluids. While leukocyte *TP53* mutations contributed minimally to the burden identified in Paps and peritoneal fluid in this smaller cohort, a more detailed analysis of *TP53* mutations in blood may be of interest to fully determine the full contribution of leukocyte clones to these gynecological samples and the role of CHIP in this high-risk population.

In conclusion, this exploratory study has revealed increased pathogenic *TP53* mutations in the peritoneal fluid of individuals who are older than 45 and have *BRCA1* germline mutations. These findings may help inform our understanding of the increased risk of HGSOC in *BRCA* carriers. Further studies of *TP53* mutations in the fallopian tube of *BRCA* carriers are warranted to explore the origin of *TP53* clonal expansions observed in the peritoneal fluid and to provide a further understanding of the enhanced risk of HGSOC in susceptible individuals.

## Methods

*Patients and Sample Selection*

This study included a cohort of 57 individuals who underwent prophylactic salpingo-oophorectomy at the University of Washington due to inherited mutations in *BRCA1* (N=25), *BRCA2* (N=21), or other familial history of cancer (N=11). Women were enrolled at diagnosis under an IRB- approved protocol at the University of Washington and consented to the collection of clinical information and tissue specimens including blood, peritoneal fluid, and a pre-operative Pap test. Paps were collected in the OR prior to the start of surgery with an endocervical cytobrush (Thinprep, Hologic, MA, USA) according to manufacturer's protocol. Blood was also drawn and processed prior to the surgery. Peritoneal fluids (a saline wash of the peritoneum) were collected during surgery. Both peritoneal fluid and Paps were centrifuged at 2700 rpm for 15 minutes and cell pellets were stored in 1.7mL of media at -80°C at the University of Washington Gynecologic Oncology Tissue Bank.

Sectioning and extensively examining the fimbriated end (SEE-FIM) was performed for the whole cohort to determine the presence of ovarian cancer precursor lesions. *BRCA* and other known ovarian cancer genes' germline mutation status was determined with commercial tests or the in-house BROCA sequencing panel [32]. CA-125 levels were collected prior to surgery. Inclusion criteria included a negative SEE-FIM and availability of peritoneal fluid and Paps with

sufficient DNA yield for analysis (200ng for Pap smears and blood, >400ng for peritoneal fluids).

Exclusion criteria included prior cancer history other than breast cancer, malignant histologies, endometriosis (including endometriotic cysts), and pelvic inflammatory disease (PID). The 3 groups of patients were age-matched to account for age-association of somatic mutations. Clinico-pathological information was recorded for each patient in Supplementary Table S1.

*DNA Extraction*

Paps and peritoneal fluid cell pellets were centrifuged at 3000g for 5 minutes, and DNA was extracted from the cell pellet using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) with 2-hour proteinase K incubation at 37°C and adding RNase A. DNA was quantified with a Qubit HS dsDNA kit (ThermoFisher Scientific).

Post collection, blood was centrifuged at 2700rpm for 15 minutes, and leukocytes (buffy coat layer) were isolated for DNA extraction. Leukocytes were treated with RBC lysis buffer before an overnight proteinase K incubation at 37°C before DNA extraction.

Post extraction, Paps, peritoneal fluid, and blood DNA Integrity Number (DIN) number of genomic DNA was collected using Agilent 4200 TapeStation Genomic tapes to determine the quality of DNA. All samples had DIN values above 5.

*Duplex Sequencing Library Preparation*

Duplex sequencing libraries were prepared using commercially available kits (TwinStrand Biosciences, Seattle, WA). For peritoneal fluid libraries, >400ng (range: 350-500ng) (Supplementary Table S1) of genomic DNA was used; for Pap smear and blood libraries, 200ng of genomic DNA was used. Library preparation involved sonication, end-repair, A-tailing, and ligation to duplex sequencing adapters. After PCR amplification of libraries, capture of *TP53* exons was performed with two rounds of hybridization with 120bp biotinylated probes (*TP53* human panel v1.0 from TwinStrand Duplex Sequencing, Seattle, WA, target size 3,000 bp). Library fragment size was confirmed using Agilent 4200 TapeStation with HS D1000 tapes, and libraries were quantified using the Qubit dsDNA HS Assay kit (ThermoFisher Scientific).

Libraries were then diluted and pooled for sequencing, allocating 5.6 million reads per peritoneal fluid libraries, and 2.8 million reads per Pap smear and blood libraries. Libraries were sequenced using 2 x 150bp paired-end kits on an Illumina Miseq on-site, an Illumina HiSeq 4000 at Genewiz (Azenta Life Sciences, South Plainfield, NJ), or an Illumina NovaSeq 6000 at UW Medicine Virology (Seattle, WA).

*Duplex Sequencing Analysis*

Sequencing reads were analyzed as previously described [45] using pipeline v2.1.3 from https://github.com/Kennedy-Lab-UW/Duplex-Seq-Pipeline. VCF files were converted to MAF files, which were then post-processed with R version 4.1.2 [47].

Raw reads were demultiplexed and then grouped to produce single-strand consensus sequences (SSCS) for each DNA strand. Then, complementary strands from the same DNA molecule were collapsed into a duplex consensus sequence (DCS), which were then aligned to the human genome reference hg38 (GRCh38), end-trimmed (clip 15), locally realigned, and overlap-trimmed. Variant calling was done through a samstools mpileup-based variant caller, and the VCF file outputs were then converted to MAF files using Vcf2Maf script (https://github.com/mskcc/vcf2maf) with VEP version 104. For each mutation, Variant Allele Frequency (VAF) was calculated as the number of mutant duplex reads divided by the total duplex depth sequenced at the given position. Single Nucleotide Polymorphisms (SNP) were identified based on VAF (>0.4) and were used to confirm the absence of cross-contamination. Mutations with VAF>0.4 were excluded for further analysis.

For each sample, we quantified coding and non-coding mutations and the total number of coding and non-coding nucleotides sequenced. “Coding” included nucleotides in the canonical *TP53* coding exons plus 2bp boundary nucleotides to capture splice site mutations, and “non-coding” included all the remaining nucleotides in the target regions. *TP53* areas prone to sequencing artifacts (located in exons 4 and 11) were masked. Only nucleotides sequenced at depth >1,000x were included in the analysis, which effectively included all coding positions and a subset of intronic positions. The capture probes also covered alternative exons, but they were considered non-coding for analysis. For each sample, the overall duplex depth was calculated as the average depth of all on-target coding nucleotides sequenced. Upon initial inspection, two peritoneal fluid sample had non-homogenous depth (much lower than others) and was discarded

from further analysis due to non-homogenous sampling. Two Pap samples had obvious clusters on 5' end of reads, which usually indicate mutational artifacts likely arising from damage from sonication, but captured due to end-repair and a-tailing [46], and were also discarded from further analysis. Post exclusion, the mean depth for peritoneal fluid for 55 samples was 9363x (range: 5905x – 18023x and the depth for Paps for 55 samples was 4913x (range: 2285x – 8736x) (Supplementary Figure S1).

*Mutation analysis*

Outputs from Duplex Sequencing Pipeline (MAF file, denoms file) were processed using R scripts to further annotate mutations, quantify mutational frequency (MF) and mutational burden (MB), and generate plots. Mutations were also annotated based on mutation type (missense, nonsense, splice, insertion, or deletion, synonymous, intronic) and mutation spectrum (C>A, C>G, C>T, T>A, T>C, and T>G) (Supplementary Table S2 & S3**)**. MF and MB were generated to normalize variation in sequencing depth between samples (Supplementary Table S4 & S5**)**. For each sample, MF was calculated independently for coding and non-coding regions as the number of mutant positions divided by the total number of duplex nucleotides sequenced in the corresponding regions. In duplex sequencing, each mutant duplex read corresponds to a single DNA mutant molecule, thus the total number of mutant duplex reads per position is informative of the size of the mutant clone. To capture this information, for each sample we also calculated MB as the total number of mutant duplex reads in coding positions divided by the total number of duplex nucleotides sequenced in the coding region.

*Characterization of TP53 mutations using Seshat*

All coding *TP53* variants identified in peritoneal fluid, Paps, and blood (n =) were submitted to Seshat on May 18th, 2022 (https://p53.fr/*TP53*-database/seshat ), a web service that annotates and characterizes *TP53* mutations based on data derived from the UMD *TP53* database [33]. Seshat long files were used to extract information about the frequency of mutations in the UMD cancer database and their pathogenicity.

For pathogenicity comparisons, mutations called by Seshat as "Benign", "Likely benign" and "VUS" were labeled as Non-pathogenic and all other mutations were called Pathogenic (this included the categories of "Pathogenic", "Likely pathogenic", indels, and complex mutations for

which Seshat could not determine pathogenicity). For each sample, Pathogenic MF and MB were calculated as mutant positions or mutant reads (respectively) with pathogenic mutations divided by the total number of duplex nucleotides sequenced for further sample comparisons.

In addition, to emulate spectrums of mutation with no selective pressure, an in-silico model that emulated all possible substitutional mutation events in *TP53* mutation events in *TP53* coding region (n=3546) was created and submitted to Seshat for mutation type and pathogenicity calling for comparisons.

*UMD TP53 cancer database*

Universal Mutation Database (UMD) *TP53* database data [33] (version 1.8 accessed date: July 21, 2021), containing *TP53* mutations reported in the HGSOC tumors was used for mutation type, pathogenicity, mutational spectrum, and mutational distribution comparisons. Specifically, for the mutational spectrum plots, only substitution mutations were included in that analysis.

UMD was also used to determine the codon locations of substitutions reported in HGSOC in *TP53*. A histogram of HGSOC substitution mutation location in *TP53* was then compared with the histogram of substitution mutations identified in samples.

Lastly, UMD was also used to determine HGSOC hotspot codons, which were defined as codons with at least 1% of reported substitutions in HGSOC. Hotspot codons include codons 132, 135, 151, 157, 163, 173, 175, 176, 179, 193, 195, 196, 213, 220, 234, 237, 238, 241, 244, 245, 248, 266, 272, 273, 275, 278, 281, 282, 306 and 342 (Supplementary Table S7).

*Statistical analysis*

Comparisons of MF and MB across groups were performed by Mann-Whitney U test, and correlations were tested with Spearman's rank test. Associations between categorical variables were tested with Fisher's exact test. All tests were two-sided at $\alpha$ level (type 1 error rate) of 0.05. Statistical analyses were performed with R version 4.1.1 [47] and SPSS version 26 [48].

## References


1. Bowtell, D.D., et al., *Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer.* Nat Rev Cancer, 2015. **15**(11): p. 668-79.
2. Temkin, S.M., et al., *Ovarian Cancer Prevention in High-risk Women.* Clin Obstet Gynecol, 2017. **60**(4): p. 738-757.
3. *Ovary Cancer, SEER 5-Year Relative Survival Rates, 2012-2018*. 2022.
4. Kuchenbaecker, K.B., et al., *Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers.* JAMA, 2017. **317**(23): p. 2402-2416.
5. Pearce, C.L., et al., *Population distribution of lifetime risk of ovarian cancer in the United States.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(4): p. 671-676.
6. Narod, S., *Can advanced-stage ovarian cancer be cured?* Nat Rev Clin Oncol, 2016. **13**(4): p. 255-61.
7. Finch, A.P., et al., *Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation.* J Clin Oncol, 2014. **32**(15): p. 1547-53.
8. Crum, C.P., et al., *Lessons from BRCA: the tubal fimbria emerges as an origin for pelvic serous cancer.* Clin Med Res, 2007. **5**(1): p. 35-44.
9. Ducie, J., et al., *Molecular analysis of high-grade serous ovarian carcinoma with and without associated serous tubal intra-epithelial carcinoma.* Nat Commun, 2017. **8**(1): p. 990.
10. Ghezelayagh, T.S., et al., *Characterizing TP53 mutations in ovarian carcinomas with and without concurrent BRCA1 or BRCA2 mutations.* Gynecol Oncol, 2021. **160**(3): p. 786-792.
11. Crum, C.P., *Intercepting pelvic cancer in the distal fallopian tube: theories and realities.* Mol Oncol, 2009. **3**(2): p. 165-70.
12. Kuhn, E., et al., *TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma--evidence supporting the clonal relationship of the two lesions.* J Pathol, 2012. **226**(3): p. 421-6.
13. Labidi-Galy, S.I., et al., *High grade serous ovarian carcinomas originate in the fallopian tube.* Nat Commun, 2017. **8**(1): p. 1093.
14. Wu, R.C., et al., *Genomic landscape and evolutionary trajectories of ovarian cancer precursor lesions.* J Pathol, 2019. **248**(1): p. 41-50.
15. Gerstung, M., et al., *The evolutionary history of 2,658 cancers.* Nature, 2020. **578**(7793): p. 122-128.
16. Chen, F., et al., *Serous tubal intraepithelial carcinomas associated with high-grade serous ovarian carcinomas: a systematic review.* BJOG, 2017. **124**(6): p. 872-878.
17. Eckert, M.A., et al., *Genomics of Ovarian Cancer Progression Reveals Diverse Metastatic Trajectories Including Intraepithelial Metastasis to the Fallopian Tube.* Cancer Discov, 2016. **6**(12): p. 1342-1351.
18. Soong, T.R., et al., *The fallopian tube, "precursor escape" and narrowing the knowledge gap to the origins of high-grade serous carcinoma.* Gynecol Oncol, 2019. **152**(2): p. 426-433.
19. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9530-5.
20. Wang, Y., et al., *Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers.* Sci Transl Med, 2018. **10**(433).

21. Krimmel-Morrison, J.D., et al., *Characterization of TP53 mutations in Pap test DNA of women with and without serous ovarian carcinoma.* Gynecol Oncol, 2020. **156**(2): p. 407-414.
22. Krimmel, J.D., et al., *Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues.* Proc Natl Acad Sci U S A, 2016. **113**(21): p. 6005-10.
23. Nair, N., et al., *Genomic Analysis of Uterine Lavage Fluid Detects Early Endometrial Cancers and Reveals a Prevalent Landscape of Driver Mutations in Women without Histopathologic Evidence of Cancer: A Prospective Cross-Sectional Study.* PLoS Med, 2016. **13**(12): p. e1002206.
24. Maritschnegg, E., et al., *Lavage of the Uterine Cavity for Molecular Detection of Mullerian Duct Carcinomas: A Proof-of-Concept Study.* J Clin Oncol, 2015. **33**(36): p. 4293-300.
25. Salk, J.J., M.W. Schmitt, and L.A. Loeb, *Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations.* Nat Rev Genet, 2018. **19**(5): p. 269-285.
26. Schmitt, M.W., et al., *Detection of ultra-rare mutations by next-generation sequencing.* Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14508-13.
27. Kinde, I., et al., *Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers.* Sci Transl Med, 2013. **5**(167): p. 167ra4.
28. Kakiuchi, N. and S. Ogawa, *Clonal expansion in non-cancer tissues.* Nat Rev Cancer, 2021. **21**(4): p. 239-256.
29. Martincorena, I., *Somatic mutation and clonal expansions in human tissues.* Genome Med, 2019. **11**(1): p. 35.
30. Risques, R.A. and S.R. Kennedy, *Aging and the rise of somatic cancer-associated mutations in normal tissues.* PLoS Genet, 2018. **14**(1): p. e1007108.
31. Yizhak, K., et al., *RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues.* Science, 2019. **364**(6444).
32. Walsh, T., et al., *Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(44): p. 18032-7.
33. Tikkanen, T., et al., *Seshat: A Web service for accurate annotation, validation, and analysis of TP53 variants generated by conventional and next-generation sequencing.* Hum Mutat, 2018. **39**(7): p. 925-933.
34. Alexandrov, L.B., *Understanding the origins of human cancer.* Science, 2015. **350**(6265): p. 1175.
35. Mitchell, S.R., J. Gopakumar, and S. Jaiswal, *Insights into clonal hematopoiesis and its relation to cancer risk.* Curr Opin Genet Dev, 2021. **66**: p. 63-69.
36. Jaiswal, S. and B.L. Ebert, *Clonal hematopoiesis in human aging and disease.* Science, 2019. **366**(6465).
37. Norquist, B.M., et al., *The molecular pathogenesis of hereditary ovarian carcinoma: alterations in the tubal epithelium of women with BRCA1 and BRCA2 mutations.* Cancer, 2010. **116**(22): p. 5261-71.
38. Shih, I.M., Y. Wang, and T.L. Wang, *The Origin of Ovarian Cancer Species and Precancerous Landscape.* Am J Pathol, 2021. **191**(1): p. 26-39.

39. Moore, L., et al., *The mutational landscape of normal human endometrial epithelium.* Nature, 2020. **580**(7805): p. 640-646.
40. Suda, K., et al., *Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium.* Cell Rep, 2018. **24**(7): p. 1777-1789.
41. Yamaguchi, M., et al., *Spatiotemporal dynamics of clonal selection and diversification in normal endometrial epithelium.* Nat Commun, 2022. **13**(1): p. 943.
42. Salk, J.J., et al., *Ultra-Sensitive TP53 Sequencing for Cancer Detection Reveals Progressive Clonal Selection in Normal Tissue over a Century of Human Lifespan.* Cell Rep, 2019. **28**(1): p. 132-144 e3.
43. Gercel-Taylor, C., J.J. Scobee, and D.D. Taylor, *Effect of chemotherapy on the mutation frequency of ovarian cancer cells at the HPRT locus.* Anticancer Res, 2005. **25**(3B): p. 2113-7.
44. Kubota, M., et al., *Cancer chemotherapy and somatic cell mutation.* Mutat Res, 2000. **470**(2): p. 93-102.
45. Kennedy, S.R., et al., *Detecting ultralow-frequency mutations by Duplex Sequencing.* Nat Protoc, 2014. **9**(11): p. 2586-606.
46. Abascal, F., et al., *Somatic mutation landscapes at single-molecule resolution.* Nature, 2021. **593**(7859): p. 405-410.
47. Team, R.C. (2021). R: A language and environment for statistical computing. (R Foundation for Statistical Computing).
48. StataCorp (2019). Stata Statistical Software: Release 16. (StataCorp LLC).