Journal of Translational Medicine

RESEARCH

Open Access

Novel driver gene MDC1 confers homologous recombination repair deficiency and genomic instability in chemoresistant relapsing ovarian cancer

Michael Diaz[1†], Nicole Gull[1†], Pei-Chen Peng[1,2], Kruttika Dabke[1], Jessica Baker[1], Jennifer Sun[3], Juan Alvarado[4], Benjamin Bastin[4], Nimisha Mazumdar[1], Chintda Santiskulvong[4], Andrew J. Li[3,5], Beth Y. Karlan[6], B. J. Rimel[3,5], Sarah J. Parker[7,8], Simon A. Gayther[1,9] and Michelle R. Jones[1,5*]

Abstract

Background Nearly 80% of patients with High grade serous ovarian cancer (HGSOC) will experience recurrence within 5 years, but little is known about the mechanisms that drive this process.

Methods In this study we used whole genome sequencing to assess SNV and SV burdens. These were in turn used to estimate clonal dynamics, genomic scarring, and establish mutational patterns.

Results Mutational burdens and clonal compositions are established early and are maintained throughout recurrence. Using both next generation and ultra long read sequencing to analyze single nucleotide and structural variants (SVs) we discovered that although tumors from the same patient remained relatively stable, homologous recombination repair proficient (HRP) and homologous recombination repair deficient (HRD) tumors presented with distinct clonal profiles. SV signature analysis revealed three distinct classes: tumors defined by DNA losses, DNA gains, and copy number neutral changes. Each class displayed structural variation affecting distinct regions of the genome. Ultra long read sequencing validated most of the SVs identified in short read sequencing and identified additional SVs. A novel candidate driver gene involved in DNA repair, *MDC1*, was significantly mutated in patients with HRP tumors.

Conclusions The phenotype of high grade serous ovarian tumors, as defined by mutation and clonality profiles, is established early in disease development and remain largely unchanged through chemotherapy and recurrence. This, when considered with the significant inter-patient heterogeneity identified in HGSOC, demonstrates the need for personalized therapies based on tumor profiling. Loss of *MDC1* increases invasive properties in cell lines and may drive HRD in a subset of patients.

†Michael Diaz and Nicole Gull contributed equally to this work.

*Correspondence:
Michelle R. Jones
michelle.jones@csmc.edu

Full list of author information is available at the end of the article

Keywords High grade serous ovarian cancer, Clonality, Structural variants, Recurrence, Chemoresistance, Long read sequencing

Background

High grade serous ovarian cancer (HGSOC) is the deadliest subtype of the most lethal gynecologic malignancy in the US, epithelial ovarian cancer (EOC) [1]. Standard of care has remained largely the same for decades; maximal debulking surgery followed by combinatorial chemotherapy with taxane and platinum agents [2], with the more recent addition of poly-ADP ribose inhibitor (PARPi) treatment for a subset of patients [3]. Patients initially demonstrate a strong response to platinum based treatment, however > 80% cycle through disease relapse and chemotherapy and eventually develop progressive, chemoresistant tumors resulting in the patient succumbing to treatment-resistant disease [4].

HGSOC tumors are marked by a complex mix of copy number signatures, loss of DNA repair mechanisms, tumor suppressor loss, and oncogene activation leading to highly genomically unstable tumors [5]. This genomic instability leads to marked intratumoral and intertumoral heterogeneity. Both simple somatic mutations (SSMs), such as single nucleotide variants (SNVs) and indels (insertions/deletions), and structural variants (SV) impact the epigenomic and transcriptomic profiles encoded by the tumor genome. Previous reports that focused on SSMs identified few drivers but have established two broad categories of primary HGSOC tumors: those with and without a loss of the homologous recombination repair pathway; termed homologous recombination repair deficient (HRD) or homologous recombination proficient (HRP) [6]. How these classifications impact recurrent tumors is not currently known, as few studies have profiled primary and recurrent tumors from the same patient.

Known drivers of HRD include *BRCA1* and *BRCA2*, but loss of other genes involved in this pathway may lead to similar outcomes. Loss of *MDC1* may confer HRD outside of the standard *BRCA1/2* paradigm by inhibiting the formation of the *BRCA1/2* complex. *MDC1* serves as a docking platform to promote the localization of various DNA damage response components, including *BRCA1*, to DNA double-strand break (DSB) sites [7]. DSB leads to phosphorylation of histone H2AX (γ-H2AX) resulting in assembly of the MDC1-histone complex, which seems integral to retention of many repair proteins in these regions. Unsurprisingly, loss of *MDC1* impairs DSB repair [8]. *MDC1* variants have been observed in melanoma, bladder cancer, prostate cancer, and non-small cell lung cancer patient samples suggesting its loss may play a role in tumorigenesis [7].

It is hypothesized that tumors begin as a single clone that acquired a driver mutation and then progressively developed mutations leading to changes in fitness. Major mutational changes lead to establishment of new clones that carry passenger mutations from their founding clone. Until recently it was unknown if the local spread of HGSOC in the peritoneum arises from monoclonal or polyclonal seeding. Recent reports indicate that both of these mechanisms occur, with examples of polyclonal and monoclonal seeding observed in different metastases from a single patient [9, 10]. Previously published studies of primary tumors and/or metastases have included small cohorts and have not included matched chemo-naive and recurrent tumors from the same patient [9]. Whole exome data suggests clonal populations of chemoresistant cells develop early in tumor development, seeding metastatic sites and serving as the dormant cell population from which recurrent tumors arise, but this observation has not been confirmed in matched tumors [11].

Previous studies using next generation sequencing (NGS) have identified high levels of copy number alterations and a high burden of SVs in HGSOC [6, 12]. However, the recent development of direct molecule whole genome sequencing by Oxford Nanopore Technology (ONT) has facilitated the characterization of highly repetitive regions previously inaccessible by NGS, and preliminary studies have indicated these new strategies improve SV detection by up to 68% [13]. It is likely that complex high-level rearrangements have been undercharacterized in previous studies, which plays a role in our poor understanding of the temporal and structural genomic features of disease progression. Here, we study a cohort of 72 matched primary and recurrent tumors utilizing a combination of sequencing methods to temporally identify and validate the heterogeneous genomic features associated with mortality and clonal evolution.

Results

Landscape of simple somatic mutations

We identified a cohort of 32 HGSOC patients for whom paired chemo-naive, chemoresistant tumors, and germline DNA were available (See Supplementary Table 1, Additional File 1, Supplementary Fig. 1A, Additional File 2). We identified 11 patients who carry germline pathogenic variants in *BRCA1* or *BRCA2* (*BRCA1/2* carriers) while the remaining patients did not carry pathogenic mutations in either gene, nor any gene currently used for screening for breast and ovarian cancer. *BRCA1/2* carriers had longer overall survival, as previously shown [14]. SSMs were called from short read whole genome

sequencing (WGS) performed in epithelial enriched samples from each tumor. Five tumors lacking germline variants in *BRCA1/2* developed somatic SV or point mutations in these genes (Fig. 1A). Polygenic risk scores were not significantly different between *BRCA1/2* carriers and non-carriers (p = 0.689, Fig. 1A and Supplementary Table 1, Additional File 1). Chromothripsis was not detected in any of the germline *BRCA1/2* carriers and 37.5% (3/8) of patients with chromothriptic tumors bore chromothripsis in both the primary and recurrent tumor. *PIK3CA* amplifications were identified in 11 tumors from 9 patients (1 *BRCA1* carrier, 1 *BRCA2* carrier, 7 non-carriers) and were far more common than *PIK3CB* amplifications, which were only found in 3 tumors. *PIK3CA* and *PIK3CB* amplifications were evenly distributed among primary and recurrences, with 6 and 5 instances respectively. We observed tumors from 15 patients with somatic loss of *PTEN*, 15 with an amplification of *CCNE1*, and 6 with an amplification of *RB1* (2/6 due to intergenic copy number changes) (Fig. 1A and Supplementary Table 2, Additional File 1). All tumors in our cohort contained a deleterious mutation in the *TP53* gene, however not all mutations were initially detected using automated somatic variant calling methods. Upon manual review, each mutation was identified and annotated. In addition to *TP53*, we identified the genes *MDC1*, *MUC22*, and *TDG* as genes commonly mutated above a background rate (Fig. 1A and Supplementary Table 3, Additional File 1). In accordance with previous WGS profiling studies of HGSOC, there are few genes frequently mutated across the cohort [15].

Homologous recombination repair deficiency is stable throughout disease

Homologous recombination repair deficiency (HRD), as defined by germline deleterious variants, somatic mutations, or promoter hypermethylation in *BRCA1/2*, *RAD51C*, *RAD51D*, was consistent across primary and recurrent tumors from the same patient. Two patients developed a new HR status throughout the disease course (See Supplementary Table 2, Additional File 1). In the fourth recurrent tumor of patient 31 (a *BRCA1/2* non-carrier), a somatic in-frame deletion with a predicted moderate effect on *BRCA1* function was observed. Of the 11 *BRCA1/2* germline carriers, reversion was observed only in the third recurrent tumor of patient 3. Patient 3 likely regained *BRCA2* function as a 36 bp deletion was identified over the germline deleterious variant. This

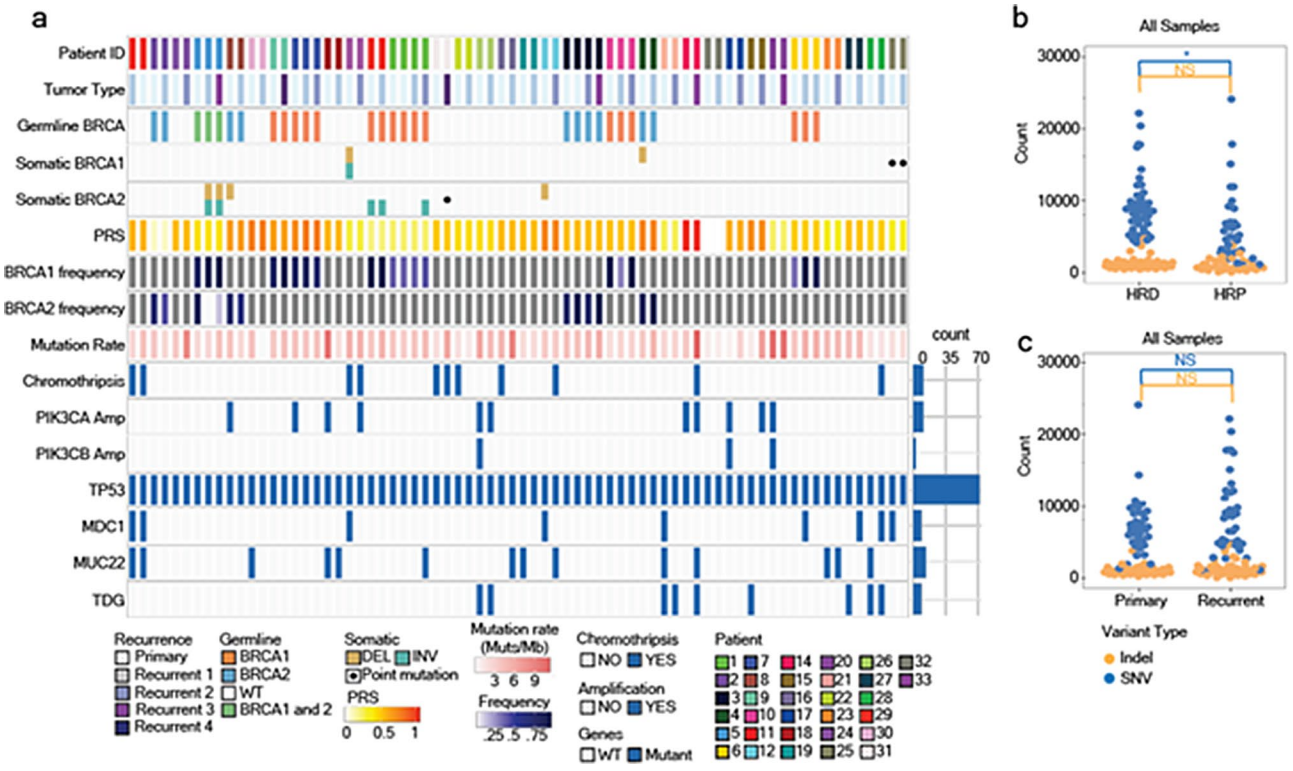

Fig. 1 The mutational profile of HGSOC is preserved through disease recurrence. **a**, Descriptive graphic of tumor mutational profile, including recurrent stage, germline and somatic BRCA1/2 variants, patient polygenic risk score (PRS), somatic BRCA1/2 mutation frequency, mutation rate and common mutational events and commonly mutated genes within our cohort. **b**, Number of SNV and indel per tumor, split by HRD status. **c**, Number of SNV and indel per patient, split by primary or recurrent status. For **b**-**c** blue bars indicate statistics for SNV while orange bars indicate statistics for indels. *p < 0.05, NS not significant, students T test

short deletion is predicted to restore the reading frame and is not predicted to be within a functional domain of the gene. Outside of these two patients, the methylation and mutation patterns of these genes were conserved.

HR status, but not recurrence, had a significant effect on the number of SNVs ($p = 0.0167$). While HRD and HRP tumors did not display significant differences in the number of indels ($p = 0.198$), HRP tumors did have a broader distribution of indel count (Fig. 1B). Recurrent tumors did not have a significantly higher SSM burden than primary tumors. However, HRD recurrent tumors have significantly more somatic SNVs than primary tumors ($p = 0.0314$; Fig. 1B, C, See Supplementary Fig. 1B, Additional File 2).

MDC1 as a novel candidate driver gene in HGSOC

MDC1 was identified as one of the most significantly mutated genes in our analysis of SSMs (Fig. 1A). 9 tumors from 8 patients had SSMs in the PST region of *MDC1*, encoded by exon 10 of the MDC1 gene, all of which were predicted to have a moderate variant effect (Fig. 2A). Five primary (2 HRD, 3 HRP) and 4 recurrent tumors (1 HRD,

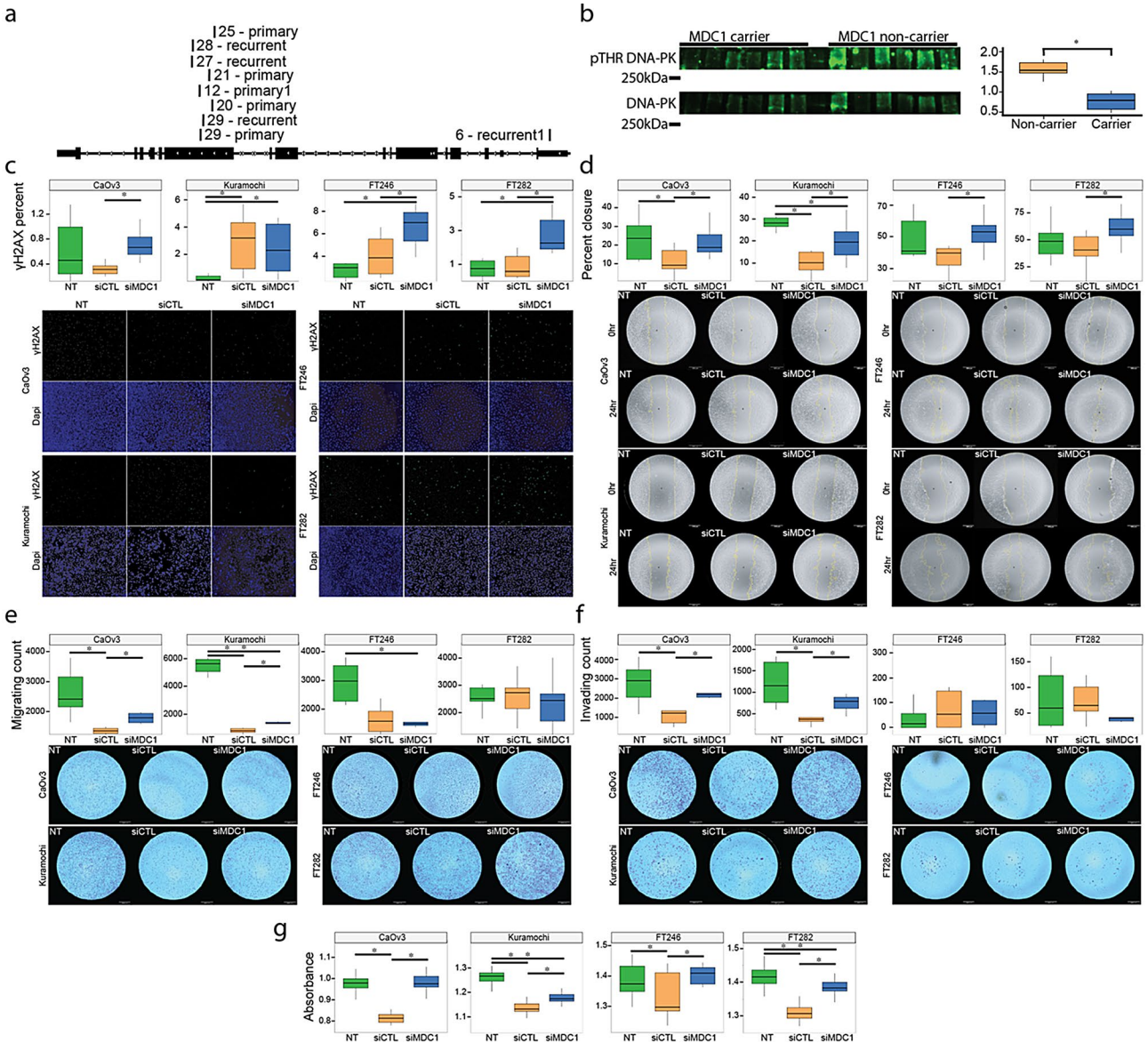

Fig. 2 MDC1 knockdown increases migration and invasion. **a**, Diagram of variants identified in our cohort. **b**, Left, blots of total DNA-PK and phosphorylation on threonine residue 2609. Right, relative phosphorylation in each sample. **c**, (top) Percentage of γ-H2AX positive cells, $n = 6$. (bottom) Representative images. **d**, (top) Wound closure percentage, $n = 9$. (bottom) Representative images taken at 0 h post scratch and 24 h after addition of growth media. **e**, (top) Count of migrated cells through 8μM pores, $n = 6$. (bottom) Representative images. **f**, (top) Count of cells that invaded through a Matrigel membrane, $n = 6$. (bottom) Representative images. **g**, MTT assay read at 570 nm, $n = 24$. *$p < 0.05$, NS not significant, students T test

3 HRP) were identified as *MDC1* somatic mutation carriers (See Supplementary Table 3, Additional File 1). Several *MDC1* carrier tumors have regions of chromothripsis and complex allele specific copy number profiles, irrespective of chromothripsis. *MDC1* mutants displayed a lower mutational rate than *BRCA1/2* non-carrier tumors, similar to *BRCA1/2* carrier tumors (See Supplementary Table 2, Additional File 1). Four of the 5 primary tumors did not initially have an *MDC1* mutation identified in the matched recurrent tumor (See Supplementary Table 3, Additional File 1). Manual review of aligned sequencing reads revealed the variant was in fact present in all tumors from the same patient, but at a sub-clonal allele frequency that was lower than the thresholds used in our somatic variant calling pipeline, underscoring potential missed subclonal variants (VAF < 0.05, See Supplementary Table 3, Additional File 1). This low allele frequency may have been the result of intra-tumoral heterogeneity and likely contributed to previous studies of HGSOC not identifying this gene as being frequently mutated. Within the Pan-Cancer Analysis of Whole Genomes (PCAWG) ovarian cancer study [16] depth was bimodal with nodes at 38 and 60x where there was sufficient power to detect subclonal drivers that were above 30% CCF [17]. We performed somatic sequencing at an average depth of 60x and all the *MDC1* variants identified here had a VAF between 9 and 30%, below the PCAWG threshold. Mutations were confirmed using Illumina Ampliseq targeted sequencing at an average depth of 5000x. Most somatic mutations displayed a lower VAF in the Ampliseq panel than WGS, and all somatic mutations in MDC1 were detected with this second DNA sequencing biochemical assay (See Supplementary Table 4, Additional File 1).

MDC1 loss increases invasion and migration in invasive ovarian cancer

MDC1 has been shown to increase autophosphorylation in the MDC1 binding partner DNA-PK [18], which has functions in the DNA damage response, non-homologous end joining, and HR [19]. Phosphorylation at TYR2609 leads to a conformational change necessary for DNA end processing [18, 19]. To assess the impact that *MDC1* variants have on MDC1 function we selected 6 *MDC1* variant carrier and 6 non-carrier samples, based on available tissue. In line with previous reports of MDC1 function, the tumors bearing MDC1 variants displayed a reduction in relative pTYR2609 versus MDC1 non-carriers (Fig. 2B). Phosphorylation of the TYR2609 residue is vital to double strand break repair, further suggesting that loss of MDC1 is a potential driver for HRD and tumorigenesis in these tumors.

We previously profiled the proteome of these tumors using mass spectrometry [20], and used those data to measure MDC1 protein expression. *MDC1* demonstrated consistent protein expression across all tumor samples. *MDC1* was not differentially expressed at the transcript or protein level between primary and recurrent tumors or between HRD and HRP tumors. Thus, it is unlikely that the reduction in DNA-PK autophosphorylation is due to a decrease in MDC1 abundance and is instead due to decreased function. However, it was distinctly upregulated in tumor tissues when contrasted with tumor-adjacent stromal samples, which may contribute to chemoresistance through restoration of DNA repair pathways. MDC1 upregulation was replicated using publicly available proteomics data sets [21] (log2FC = 0.72; FDR = 5.27E-07, See Supplementary Table 5, Additional File 1).

To assess functional outcomes of *MDC1* loss, four cell lines (two ovarian cancer lines, CaOv3 and Kuramochi, and two fallopian tube lines, FT246 and FT282) were transfected with small interfering RNA targeting *MDC1* or a non-targeting control and assayed (See Supplementary Table 6, Additional File1). Quantitative PCR confirmed *MDC1* knockdown in all lines (See Supplementary Fig. 2A, Additional File 2 and Supplementary Table 6, Additional File1). We performed RNAseq on all lines, and cells lacking *MDC1* displayed no significantly upregulated genes. The genes downregulated alongside MDC1 were *TTC33*, and *LMLN* (See Supplementary Fig. 2B, Additional File 2). These proteins respectively have an unknown function and negatively regulate cell proliferation [22].

MDC1 functions to repair double strand breaks through recruitment via γH2AX phosphorylation [23] and naturally its loss led to an increase in γH2AX phosphorylation, with the exception of the Kuramochi line (Fig. 2C). The Kuramochi line bears a heterozygous BRCA2 variant, as well as one of the MDC1 variants observed in our tumor cohort, suggesting this cell line may be deficient in some MDC1 signaling. Although there was increased γH2AX phosphorylation in CaOv3, it is worth noting that the foci were far dimmer and less frequent in this line than the normal fallopian tube lines, likely due to competing mechanisms that suppress DSB repair. MDC1 loss led to increased cell migration across a flat surface in all lines (Fig. 2D) and increased migration through 8μ meter pores in response to FBS in only the ovarian cancer lines (Fig. 2E). Knockdown of *MDC1* led to increased invasion through a Matrigel membrane in both cancer cell lines, but not within the fallopian tube lines (Fig. 2F). Increased invasion may be the result of increased metabolism, as MDC1 knockdown led to an increase in MTT absorbance in all lines, with a muted effect seen in the Kuramochi line (Fig. 2G). These results suggest that *MDC1* enhances migration and invasion through increased metabolism, but MDC1 loss is not sufficient to confer invasion through a basement membrane

in lines that lack invasive properties. These results are in line with the RNAseq data.

Homologous recombination deficiency drives tumor mutational profiles

Copy number profiles appeared relatively unaffected by recurrence as hierarchical clustering often placed tumors from the same individual adjacently (Fig. 3A). HRD tumors have a higher count of SVs overall ($p = 0.00404$) with significant increases in SV burden for deletions, tail to tail inversions, and translocations ($p = 2.72 \times 10^{-3}$, 0.0387, 0.0238, 2.22×10^{-3} respectively; Fig. 3B, C,D, See Supplementary Table 2, Additional File 1). There was no significant difference in the count of any individual class of SV between primary and recurrent tumors (Fig. 3B). Mean SV size across each SV class showed no difference across recurrent status (See Supplementary Fig. 3A, Additional File 2, Supplementary Table 7, Additional File 1, Fig. 3E). Foldback inversions were relatively rare and not significantly affected by HRD or recurrent status (Fig. 3B, C, D). The proportion of SSM and SVs shared by both the primary and recurrent tumor was not different based on HRD status (SSM HRP-25%, HRD-37% $p = 0.079$; SV HRP-25%, HRD-34% $p = 0.292$), however the proportion of SV unique to the primary tumor was increased in HRP tumors (SSM HRP-36%, HRD-29% p=0.403, SV HRP-44%, HRD-35% $p = 0.448$) (Fig. 3F). There was less variability in the proportion of shared events across HRD tumors than HRP tumors (Fig. 3F, Supplementary Table 8, Additional File 1).

Homologous recombination status drives SV distribution

Enrichment of SVs in hotspots across the genome was also influenced by HR status. HRP and HRD tumors displayed at least 5% of large SVs in chromosomes 1, 2, 3, 5, 7, and 8 although HRP samples possessed a greater abundance in chromosomes 5, 7, 19, 20, and 22 (Fig. 3E). The largest disparity between groups in relative SV abundance was found on chromosome 19, in which 7.4% of all large SV were found in HRP tumors versus 1.2% in HRD tumors (See Supplementary Table 9, Additional File 1, Supplementary Fig. 3B, Supplementary Fig. 4A, Additional File 2). Copy number profiles from chromosomes 19 and 20 also display extreme gains and losses within the same tumor, particularly in the HRP tumors (Fig. 3A, Supplementary Fig. 4B, Additional File 2). HRP tumors tend to accumulate fewer SV >10 Mb but have a high concentration of large SV in chromosomes 19 and 20 where 14% of all large SVs were found, with at least one variant greater than 10 Mb found in 17/30 tumors. By contrast, the HRD tumors tend to have large, evenly distributed SVs that are relatively absent from chromosomes 19 and 20, where 2.7% of all large SVs were found (Fig. 3E, Supplementary Table 9, Additional File 1, Supplementary Fig. 3B, 4A, Additional File 2). Few regions are commonly impacted by SVs in the HRD tumors, however the q arm of chromosome 17 does contain a high proportion of SV which are frequently head-to-head inversions. The abundance of SVs was not unique to a particular patient's tumors, as nearly 35% of all HRD tumors (15/42) possessed an SV of at least 1 MB within chromosome 17 (Fig. 3E, Supplementary Table 9, Additional File 1, Supplementary Fig. 4A, Additional File 2).

Fewer changes were observed in the size and distribution of SV across recurrence. (See Supplementary Fig. 4A, C, Additional File 2). SV counts unique to recurrent tumors were not significantly more abundant (p = 0.661) or different in size (for size groups 1–100 bp, 100 bp-1 kb, 1 kb-5 Mb, 5 Mb-10 Mb, and > 10 Mb, p = 0.564, 0.829, 0.624, 0.408, 0.884 respectively; See Supplementary Table 10, Additional File 1). Recurrent tumors did bear a high percent of very large SVs in chr 3 (17.4% of total; See Supplementary Fig. 3B, Additional File 2). The most commonly deleted gene in this region was *ROBO2* which bore 75 deletions of at least 5 Mb found across 29 tumors. Loss of *ROBO2* is associated with poor prognosis in pancreatic ductal carcinoma [24]. Other frequently deleted genes include *SHQ1* (70 deletions across 26 tumors), which inhibits prostate cancer growth and metastasis [25], and *PPP4R2* (70 deletions across 26 tumors), which is recurrently deleted in acute myeloid leukemia and is required for efficient double strand break repair [26] (See Supplementary Table 10, Additional File 1). Another marked difference in SV distribution between primary and recurrent tumors was observed in chr 17q where large SVs were mostly identified in primary tumors, not recurrent. Primary tumors were more likely to harbor large SV in chromosomes 7, 9, 10, and 17, while recurrent tumors were more likely to harbor large SVs in chromosomes 3, 4, 12, 19, and 20 (See Supplementary Fig. 3B, 4 C, Additional File 2).

SV locations were annotated with the Gencode v34 [27] to determine the genes affected by each SV. Chromosomes 1, 2, and 8 were the few chromosomes that possessed at least 5% of large SVs across all SV groups (See Supplementary Fig. 3B, Additional File 2). Genes commonly affected by SVs shorter than 5 MB were rarely observed. The most commonly disrupted gene was *CNTNAP2*, which was deleted in 20/72 tumors. This may be a function of its size, as *CNTNAP2* is approximately 2.3 Mb in length (See Supplementary Table 11, Additional File 1). GO analysis applied to the genes affected by deleterious SVs (deletion/ inversion/ duplication within an exon) that were shorter than 5 Mb and appearing in at least 4 tumors (the most stringent filter that will produce GO results) revealed that SVs in HRP tumors were enriched in genes from lipid metabolism pathways ($p = 9.83 \times 10^{-4}$) while HRD tumors had SVs enriched in genes

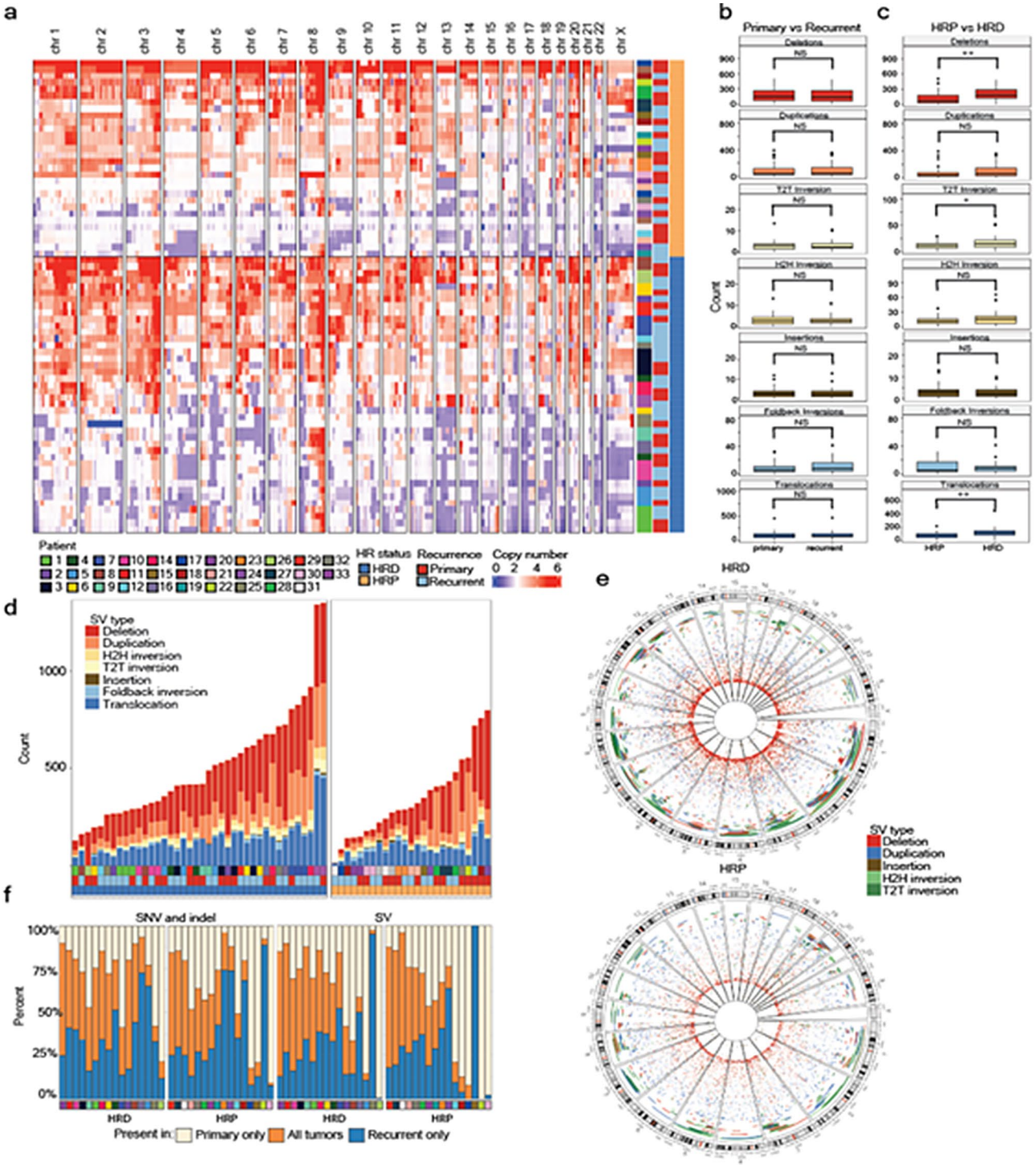

Fig. 3 Homologous recombination status determines structural variant burden and distribution in the genome. **a**, Averaged copy number intervals in a 1 Mb window split by HR status. Rows are clustered by copy number profiles. **b**, SV counts per tumor, split by primary and recurrent status. **c**, SV counts per tumor, split by HR status. **d**, Stacked bar charts depicting the total SV burden shown in c for each tumor, split by HR status. **e**, Circos plots depicting the location and size of each SV. The circular X axis shows location, while the Y shows the log size of the SV, relative to the chromosome in which it is shown. **f**, Shared and unique percentages across primary and recurrent tumors for SNV and indel (top), and SV (bottom). *p < 0.05, NS not significant, students T test

in the cell adhesion pathway ($p = 1.24 \times 10^{-6}$). GO was applied to the genes affected by deleterious SVs in recurrent samples found in at least 5 tumors (the most stringent filter that will produce GO results), which revealed that recurrent samples have an enrichment of SVs affecting genes from immune response pathways, driven by the two most commonly affected genes(*SLA2* and *PTK2*; See Supplementary Fig. 5, Additional File 2, Supplementary Table 12, Additional File 1). Genes affected by deleterious SVs in more than 3 primary tumor were too minimal to produce GO results.

Clonal evolution occurs early in disease development

Clonal abundance (Fig. 4A) changed very little through the disease course in most patients, even after primary debulking surgery followed by various lines of

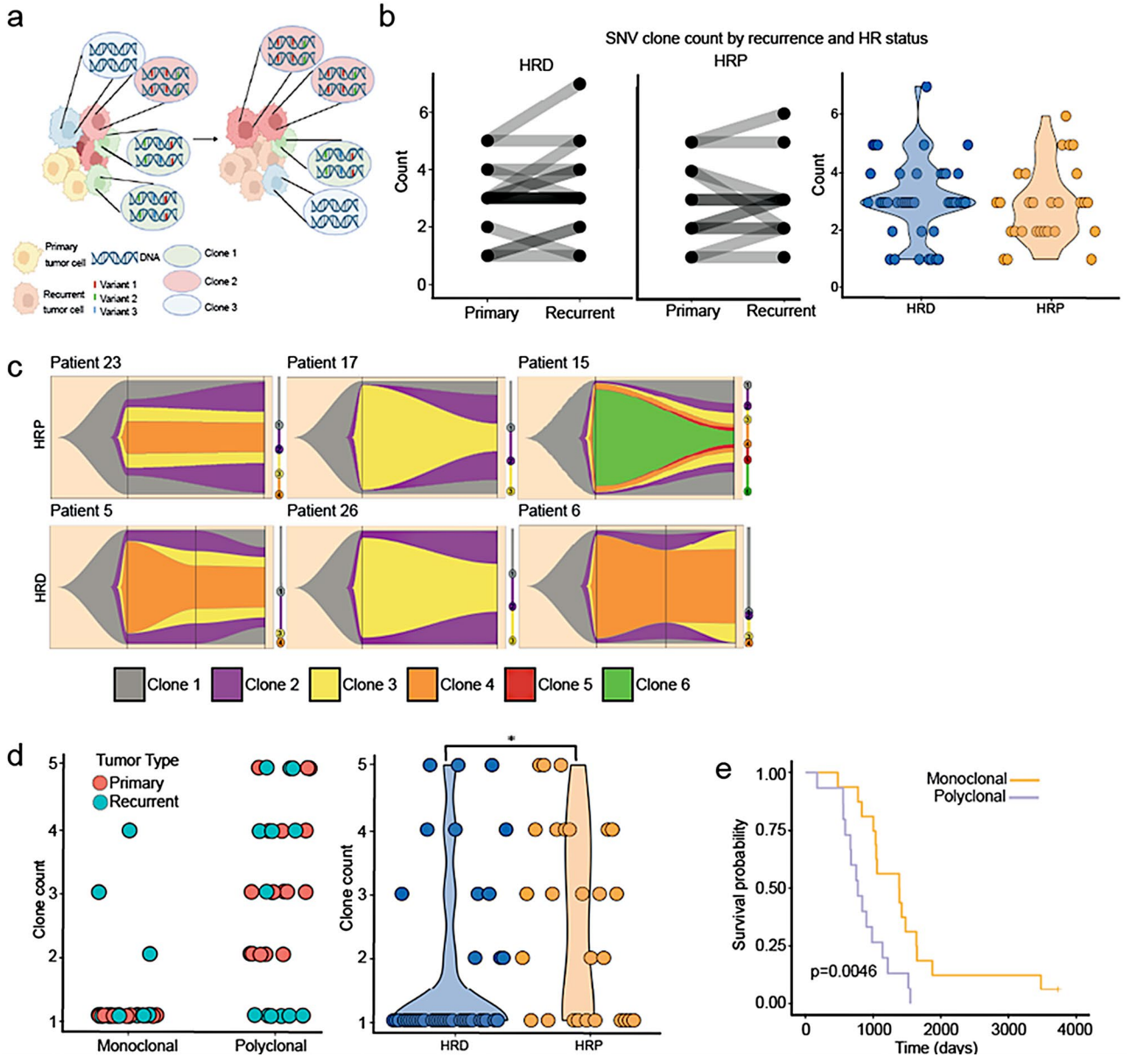

Fig. 4 Clonal profiles change little across recurrence. **a**, Schematic modeling how clonality is measured. Clonal populations will have groups of variants that all increase or decrease in unison as the clone grows or recedes. Measuring the mutation dynamics therefore is a measure of clone dynamics. **b**, (left) SNV based clone counts, split by HR status. Primary and recurrent tumors are connected by transparent lines. (right) Clone counts per tumor, split by HR status. **c**, Representative fishplots depicting clonal abundances at indicated times of sampling. Where present, vertical black lines denote a sample collection. Phylogenic trees showing clonal evolution are to the right of each plot. Relative distances between clone nodes on the phylogenic tree indicate the quantity of SNV that separate each clone. **d**, (left) Primary and recurrent clone counts based on SV dynamics. Tumors are stratified by whether the primary tumor was monoclonal or polyclonal. (right) SV based clone counts, split by HR status. **e**, Kaplan-Meier plot depicting overall survival of patients with monoclonal and polyclonal primary tumors. *p < 0.05, NS not significant, students T test

chemotherapy (Fig. 4B, C). Clonal proportions remained relatively stable; new clones were rarely identified in the recurrent tumors, with the same clonal structure persisting through treatment in most patients (19/27; Fig. 4C, See Supplementary Fig. 6, Additional File 2). In a majority of cases (18/27), the clone that comprised the greatest abundance in the primary tumor continued to comprise the greatest abundance in the recurrent tumor(s), while only 7 patients saw elimination of even a single clone during the course of chemotherapy (Fig. 4C, See Supplementary Fig. 6, Additional File 2). These trends were consistent across both HRP and HRD tumors, with no correlation seen between the number of SSMs and clonal burdens (See Supplementary Fig. 7A, Additional File 2). A subset of tumors (13/32) displayed regions of extremely high allele-specific copy number (> 5), preventing the estimation of clonality based on all SSMs within these tumors. An average of 4322 SSMs (~ 35% of the total SSMs in each tumor) were excluded from clonality estimates in each of these 13 tumors (See Supplementary Table 2, Additional File 1). These tumors display extreme heterogeneity in the copy number affecting each variant, which is likely a function of increased clonal dynamics and complexity, akin to the adaptive evolutionary state described in [28].

Using SV data to generate clonality revealed the same trends as SNV based dynamics; few tumors display development of additional clones or purifying selection after chemotherapy, remission, and recurrence. Monoclonal primary tumors tend to give rise to monoclonal recurrent tumors (12/15), while polyclonal primary tumors tend to give rise to polyclonal recurrent tumors (7/13) (Fig. 4D). SV breakpoint cancer cell fractions (CCFs) ranged from 0.09 to 1 and did not correlate with any given type of SV (See Supplementary Fig. 7B, Additional File 2), suggesting little bias was introduced in the cluster determination. Although SVs were highly conserved in some patients (i.e. patient 20 who shared 72% of SV) this was not indicative of low clone count, but rather the conservation of clones (4 clones in primary vs. 5 in recurrent). Despite a high correlation between the percent of SNV and SV shared between a patients' primary and recurrent tumors (R = 0.61, p = 2.3 × 10^{-11}), there was weak correlation between clonal composition defined by SNV and SV, mostly driven by tumors with one clone (R = 0.31, p = 0.012; See Supplementary Fig. 7C, D, Additional File 2). Unlike with SNV based analysis, HRD status did affect clonal composition as determined by SV. HRP tumors are more clonally diverse based on SV data, as HRP tumors were more likely to be polyclonal (17/29 polyclonal tumors are HRP) while HRD tumors tend to be monoclonal (29/41 monoclonal tumors are HRD, See Supplementary Table 13, Additional File 1). Tumors showed a very strong correlation between clone count and subclonal SV percentage, which likely drove the high subclonal percentages seen in HRP tumors (See Supplementary Fig. 7C, E, Additional File 2). In our cohort, we noted both an increase in total clones and subclonal variant count in the HRP tumors (Fig. 4D, Supplementary Fig. 7C-E, Additional File 2). Patients with HRP tumors tend to have a shorter progression-free survival in the clinic [29], suggesting the tumors are more proliferative which is affirmed by increased subclonal SV and clone counts. Patients with polyclonal primary tumors had a significantly shorter progression-free survival than patients with monoclonal tumors, irrespective of HR status (Fig. 4E). This trend was more apparent in the HRP tumors, where polyclonal tumors led to a substantially shorter overall survival, which may be influenced by treatment options (See Supplementary Fig. 8, Additional File 2).

SV signatures reveal three groups of tumors

Five *de novo* SV signatures were identified within our dataset (Fig. 5A, Supplementary Fig. 9, Additional File 2). Signatures 1 and 2 were defined by small and large duplications, respectively. Signatures 3 and 4 were defined by clustered (which may also be classed as copy neutral insertions) and unclustered translocations, respectively. Signature 5 was defined by small deletions (See Supplementary Fig. 9, Additional File 2). Each tumor presented a mixture of SV signatures, so an unbiased hierarchical bayesian clustering method was used to cluster the tumors. We observed closer intra-patient tumor clustering than inter-patient, as previously reported in the cohort using other genomic data types [30] (p = 3.22 × 10^{-14}, See Supplementary Fig. 10A, Additional File 2). As seen in previous work [28, 31], three distinct groups emerged from the resulting dendrogram, tumors defined by signatures 1 and 2 (referred to as duplicators), tumors defined by signatures 3 and 4 (referred to as translocators), and tumors defined by signature 5 (referred to as deletors) (Fig. 5A, See Supplementary Table 14, Additional File 1).

All groups showed significant differences in SV counts with duplicators containing the highest counts, followed by deletors, and translocators (mean 539, 401, 167, p = 0.062 del vs. dup; p = 4.23 × 10^{-7} del vs. trans, p = 1.37 × 10^{-4} dup vs. trans; See Supplementary Fig. 10B, C, Additional File 2). As expected, the deletor group contained the highest percentage of deletions, significantly more than any other group (1.67 × 10^{-7}, 8.24 × 10^{-11}, duplicator and translocator respectively). Duplicators contained the highest percentage of duplications (p = 1.95 × 10^{-8}, 3.06 × 10^{-5}, deletor and translocator respectively; See Supplementary Table 15, Additional File 1). Translocators contained the highest percentage of foldback inversions (FBI) and translocations (FBI p = 0.018, 0.016, TRA p = 0.007, 0.046, deletor and duplicator respectively,

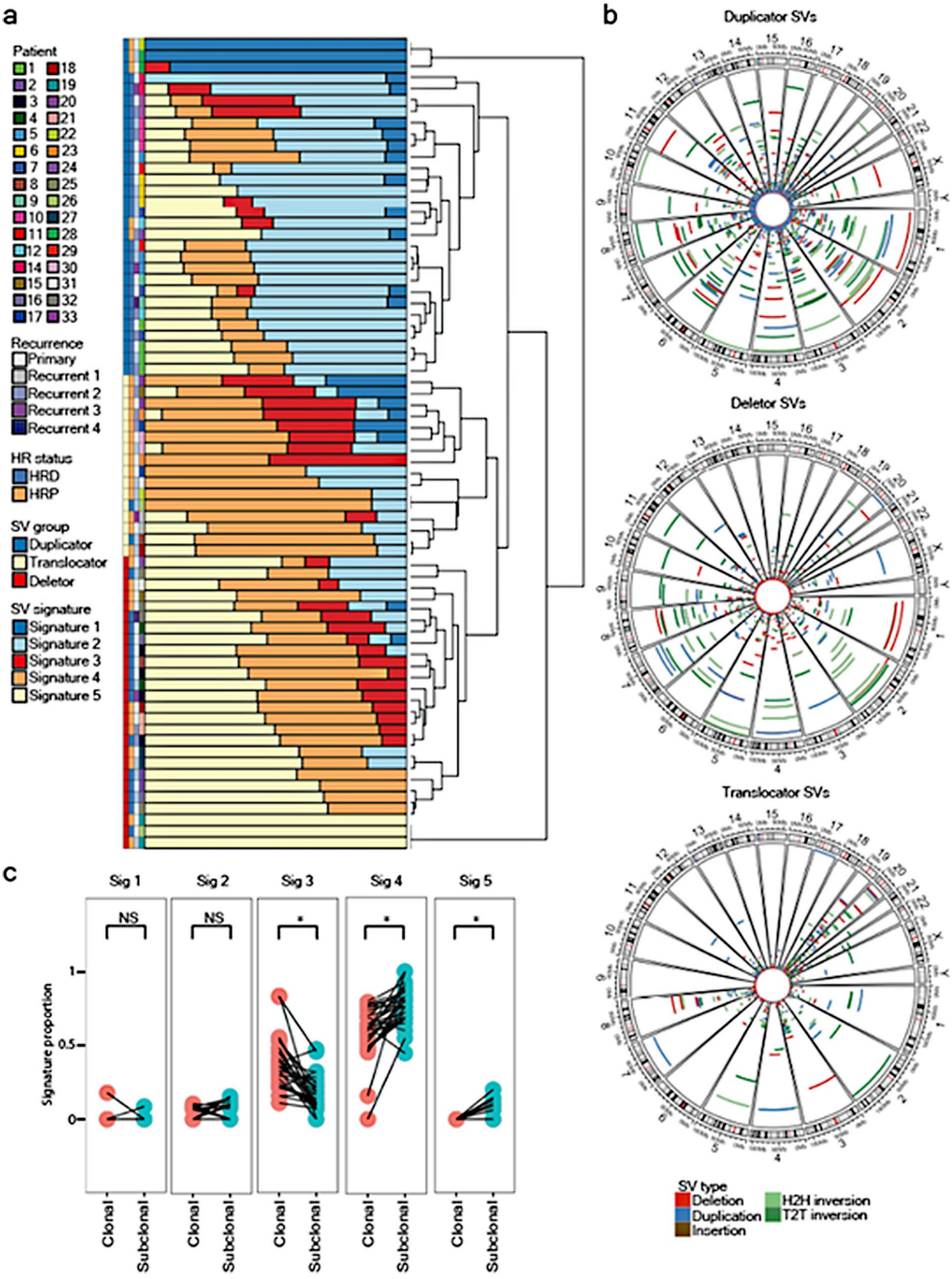

Fig. 5 Structural variant signatures identify three distinct SV mutational mechanisms. **a**, Five de novo structural variant signatures applied to each tumor in the cohort and clustered by signature composition. **b**, Genomic distribution of all SV stratified by SV signature group. **c**, SV signature proportions in clonal or subclonal populations. *$p < 0.05$, NS not significant, students T test

See Supplementary Fig. 10B, Additional File 2). Similar to previous SV signature analysis these groups were largely, but not completely, defined by germline deleterious variants in *BRCA1* (duplicators, 16/26 with *BRCA1* muts), *BRCA2* (deletors 9/23 *BRCA2* mut), and *BRCA* non-carriers (translocators 13/18 non-carrier) [32, 33]. Only three tumors from carriers of deleterious variants in *BRCA1/2* did not cluster within their predominant group. Consistent with previous reports, the translocators, largely defined by a lack of *BRCA1/2* deleterious variants, had the highest abundance of FBI and lowest patient survivorship (See Supplementary Fig. 10D, Additional File 2) [34]. Patient survivorship was highest in the deletor and duplicator groups, in line with previous reports of carriers of *BRCA1* and *BRCA2* pathogenic variants having increased survival [35] (See Supplementary Fig. 10D, Additional File 2).

SV signatures were largely conserved across recurrences. In only four cases did a primary tumor cluster in a different group than the patients paired recurrent tumor, and in 3/4 cases the group switching was between the two groups that were polyclonal (deletors and translocators). The fourth case, patient 24, transitioned from duplicator to deletor. Interestingly, this tumor pair showed an expansion of clones which is consistent with the expansion of SV and SV signature alteration seen in this patient (See Supplementary Table 13, Additional File 1, Supplementary Table 14, Additional File 1). None of the patients with tumors in multiple groups had altered HR status, and both tumors from the patient with altered HR status (patient 31) belonged to the deletor group. In line with clonality, this lack of group alteration seems to suggest that the tumor's mutational profile is established early in tumor development and conserved throughout treatment.

We noted a correlation between the SV signature of a tumor and clonality status. While deletor patients had an even dispersion of monoclonal and polyclonal primary tumors (7 and 5 respectively), duplicators tended to be monoclonal (6/9), and translocators tended to be polyclonal (5/7). Carriers of polyclonal primary tumors had a shorter progression-free survival overall, but this did not consistently apply to the SV groupings. Although the number of cases was low, the progression-free survival in the duplicator and translocator groups was not affected by the presence of multiple clones. The deletor group exhibited drastically shorter overall survival intervals if a patient had a polyclonal tumor ($p < 0.0001$, See Supplementary Fig. 10E, Additional File 2). Additionally, the shorter overall survival seen in polyclonal tumors was primarily driven by the deletor group, in which the overall survival was more than double the length for patients presenting monoclonal tumors, as defined by SV, versus polyclonal (~750 vs. ~1500 days, See Supplementary Fig. 8, Supplementary Fig. 10E, Additional File 2, $p = 7.9 \times 10^{-4}$).

The distribution of SVs across the genome was unique to each SV signature group (Fig. 5B, See Supplementary Table 16, Additional File 1). As the translocator group was heavily composed of HRP tumors, the SV distribution closely matched the HRP SV profile, while the deletors and duplicators bore aspects of the HRD profile. Translocators lack large SV in chr 9 and 10, while having an abundance in chr 19 and 20. Deletors display the most abundant large SV in chr 9 while lacking almost all large SV in chr 14, 17, 18, and 19. Duplicators display all the chr 17 q arm inversions seen in the primary/HRD group while lacking SV in chr 19 and 20 (Fig. 5B, Supplementary Fig. 4A-C, Additional File 2).

Some SV signatures are defined by clonal or subclonal SVs

The five identified *de novo* SV signature proportions were measured in each polyclonal tumor ($n = 29$) using subclonal or clonal SVs separately (12 HRD, 17 HRP). Signature 3 (defined by clustered translocations) was identified in all 29 tumors and was significantly higher when clonal SVs were used ($p = 2.2 \times 10^{-6}$). Signature 4 (defined by unclustered translocations) was also identified in all 29 tumors but was a significantly higher proportion in each tumor when subclonal SVs were used ($p = 1.95 \times 10^{-5}$). The enrichment of this signature may be due to increased accuracy in identifying translocation breakpoints, allowing these variants to pass the stricter filtering needed to identify clonality. Signature 5 (defined by small deletions) was identified in 8 tumors and was only present when subclonal SVs were used ($p = 6.06 \times 10^{-3}$). No significant differences in signature proportions were identified for the remaining signatures (Fig. 5C, Supplementary Fig. 11, Additional File 2).

Long read sequencing validates most short read calls while elucidating complex rearrangements

Five tumors were selected, based on available tissue, to perform high molecular weight DNA extraction and ultra-long ONT sequencing. ONT sequencing yielded an average output of 80.06 gigabases, with an average N50 of 55.82 kb per tumor sample. Using phased reads, an average of 1696 (median 615) SVs were called in each sample with a median phase block N50 of 11 Mb. An average of 32 SV clusters occurred per sample, as determined by Severus [36] (See Supplementary Table 17, Additional File 1, Supplementary Table 18, Additional File 1). Figure 6A depicts one such cluster in the recurrent tumor of Patient 14, where chromosomes 3, 8, 17, and X display multiple breakpoints generating new derivative chromosomes. The phased long read sequencing allowed us to identify the haplotypes that each breakpoint was present in and identify these new derivative chromosomes

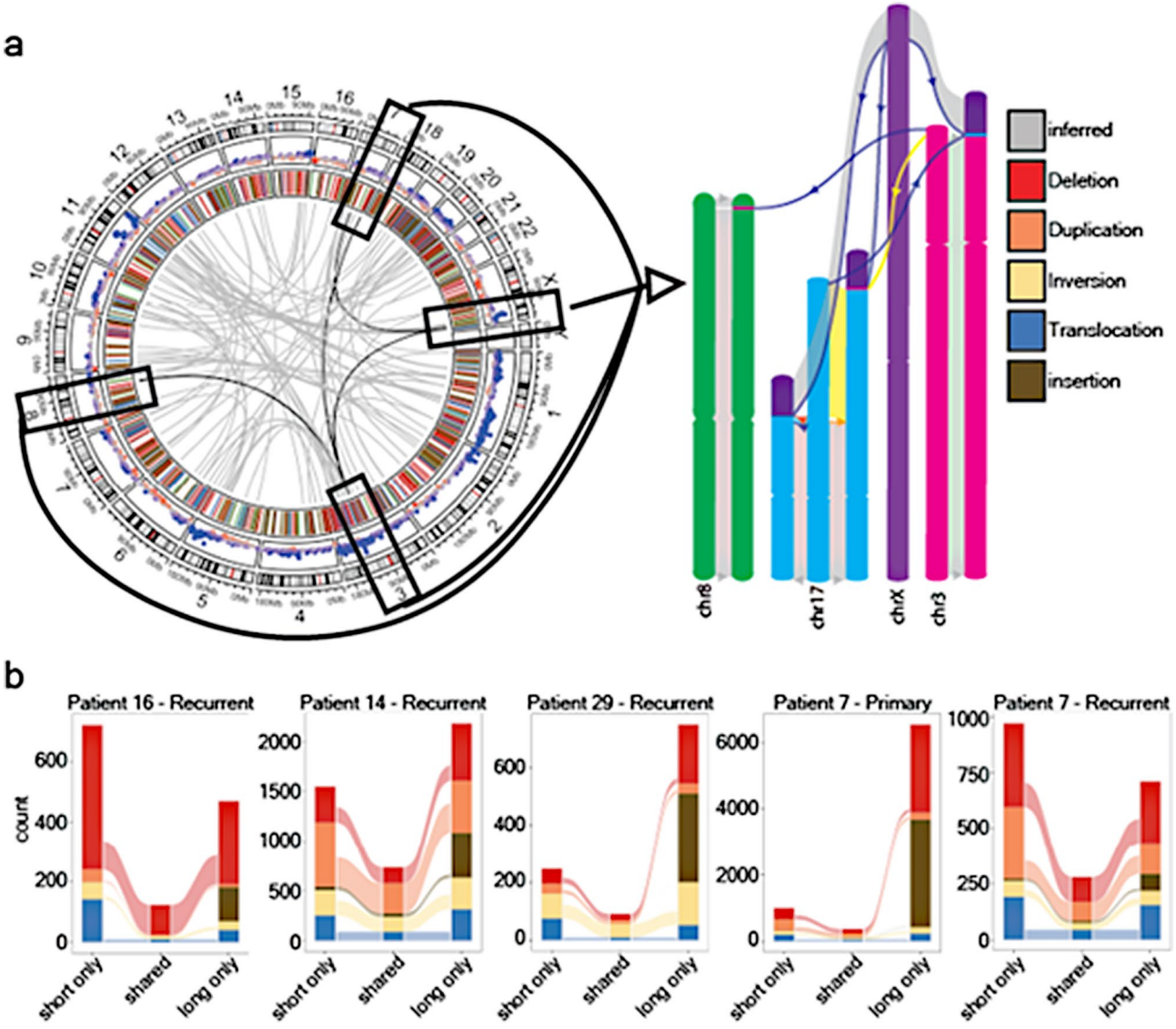

Fig. 6 Haplotagged long reads identify complex chromosomal rearrangements. **a**, SVs identified with long read sequencing from the recurrent tumor from patient 14. (Left) Circos plot showing copy number and structural variants generated from short read sequencing for a representative SV cluster. Translocations have been colored black and boxes placed around other SVs involved in the cluster. (Right) Diagram depicting derivative chromosomes generated as a result of a single SV cluster, as they are observed in the tumor using phased long read sequencing data. Colored arrows depict the type of SV and the donor chromosome. **b**, Alluvial plot showing the quantity of variants shared amongst short and long read sequencing

arising from complex structural variants. Seven of the 14 breakpoints in the cluster lie in long interspersed nuclear elements (LINES) while both of the chromosome 3 to chromosome 8 translocation breakpoints lie in short interspersed nuclear elements (SINES). When regions ± 200 bp around each breakpoint were analyzed, we found that each breakpoint pair shared at least one region of 10 bp in length with exact homology (See Supplementary Fig. 12, Additional File 2), giving some insight into the mechanism of donor site selection.

There were significant differences identified in the number of SVs in short read and long read sequencing. 45% of all SVs, 54% of deletions, 29% of duplications, 54% of inversions, 60% of insertions, and 30% of translocations identified by short reads were confirmed with long reads as the same variant with 80% reciprocal overlap (Fig. 6B, See Supplementary Table 17, Additional File 1, Supplementary Fig. 13A, B, Additional File 2). Short reads identified many more duplications >5 Mb, whereas long reads identified far more insertions which were often mistyped as duplications by short reads (See Supplementary Fig. 13C, Additional File 2). As expected, we identified increasing counts of SVs as we acquired more reads from each tumor, highlighting the abundance of subclonal rearrangements that are only present on a small number of reads (See Supplementary Table

17, Additional File 1). Although most SVs were replicated, long read sequencing identified several additional tumor suppressor gene mutations missed by short read sequencing. In patient 29 we identified a 133 Mb deletion covering *VHL* (19% VAF), in patient 7's primary tumor a 351 bp deletion in *NF1* (24% VAF) and 335 bp insertion in *RB1* intron (20% VAF), and in patient 16's recurrent tumor a 41 Mb deletion covering *RB1* (25% VAF) (See Supplementary Table 18, Additional File 1).

Known *BRCA1/2* deleterious variants and SVs detected on chromosomes 8 and X using short read sequencing were investigated further using the long read sequencing data. Patient 7 is a heterozygous carrier of a *BRCA1* deleterious variant (See Supplementary Table 1, Additional File 1). While short read sequencing did not suggest a reversion, long read phased data showed reads mapped to the deletion haplotype that bore *BRCA1* reversion, suggesting clones with a reversion may have arisen (See Supplementary Fig. 14A, Additional File 2). Chromosome 8q amplification, which harbors the *MYC* locus, was one of the few copy number profiles shared across many of the tumors in our cohort (Fig. 3A). Fourteen 1 kb-1 Mb duplications with ≥ 2 fold changes in relative depth were verified at this locus (See Supplementary Fig. 14B, Additional File 2), twelve of which were also detected in short read sequencing and shared exact breakpoints with the long read data. However, 4/12 shared duplications were identified as templated insertions and 6 of the 12 copy number gains were tandem duplications (See Supplementary Table 19, Additional File 1). None of these SV fully explained the increased copy numbers observed in the *MYC* locus. The progressive copy number gains without accompanying SVs suggests we need greater depth to capture the full range of genomic instability at this locus, or that the increased depth is the result of ploidy alterations. Chromosome X displayed frequent large scale copy number losses based on short read data, as shown in Fig. 3A. Two tumors used for long read sequencing displayed copy number < 2 across the X chromosome with short read sequencing. Both tumors were profiled at ≥ 20X depth with ONT long reads, and had sufficient coverage for phasing analysis [37]. Neither tumor was found to have loss of heterozygosity on the X chromosome, rather SVs were interspersed across the chromosome and affected both haplotypes. Interestingly, both tumors were found to have "haplobiased" X chromosomes [37], with a majority of haplotype 1 SVs occurring in the first 60 Mb, few SVs found in the 60–100 Mb region, and a majority of haplotype2 SVs occurring the 100–155 Mb regions (See Supplementary Table 20, Additional File 1).

Discussion

Patients diagnosed with HGSOC are presented with limited treatment options and a majority relapse with progressive tumors. Deeper understanding of the molecular signatures that underlie mutational changes that contribute to genomic instability is imperative to improve patient outcomes. Biopsies of recurrent tumors are rarely collected [11], making this cohort a highly valuable and novel resource. Previous reports have focused on utilizing ascites as a proxy for recurrent HGSOC tumors, and have relied on SSM analysis to investigate origin and resistance mechanisms [17, 38]. The low rate of *TP53* mutations detected in ascites samples serves as a molecular indicator of their unsuitability to represent solid recurrent HGSOC [39]. Thus, understanding the role of SVs in primary tumors and matched recurrent tumors has not been investigated due to the limited availability of paired chemo-naive primary and recurrent tumors collected at time of relapse. In the present study, we examined a cohort of 32 HGSOC patients with paired chemo-naïve and chemoresistant tumors, as well as germline DNA.

Our analyses are consistent with previous reports where few genes, outside of *TP53*, are frequently mutated across HGSOC tumors [15]. However, likely due to the increased sequencing depth used in our study we identified a novel commonly mutated gene, *MDC1.* Loss of *MDC1* led to increased genomic instability, metabolism, and migration and invasion in the cancer lines, implicating a role in tumor progression. The cell line with a known *BRCA2* and *MDC1* variants did not display enhanced DNA damage after *MDC1* KO and had muted increases in migration, invasion, and metabolism, suggesting these variants lead to reduced MDC1 function in this line. Additional studies will be required to determine the extent each variant plays in the results observed. Eight of the nine tumors with identified *MDC1* mutations had no germline deleterious variants in *BRCA1/2*. Further experiments will be necessary to understand if loss of MDC1 can explain how these tumors became genomically unstable. As HR deficiency is associated with chromothripsis [40], the higher incidence of chromothripsis observed in *MDC1* mutants implies that *MDC1* variants contribute to HR deficiency. Similar to *BRCA1/2* loss, the inactivation of *MDC1* induces sensitivity to DNA damage [23, 41], which could explain why we observed a higher proportion of *MDC1* loss in primary tumors (See Supplementary Table 3, Additional File 1). Loss of *MDC1* did not alter patient survivorship (See Supplementary Fig. 15, Additional File 2), although this may be impacted by sample size and the same patient often displaying a differential *MDC1* burden across tumors (See Supplementary Table 3, Additional File 1). Importantly, mutations in *MDC1* were almost exclusively located in the PST repeat domain region (Fig. 2A), which specifically binds to the

DNA-dependent protein kinase (DNA-PK) and regulates autophosphorylation of DNA-PK [18]. DNA-PK autophosphorylation was reduced in *MDC1* carriers suggesting impaired DSB repair due to impaired DNA-PK binding ability.

The PST domain of MDC1 is multifaceted, required for loading of RAD51 to mediate HR in interphase [42], is responsible for γH2AX independent DSB recruitment of 53BP1 [23], and promotes HR and BRCA1 foci formation [43]. Additional studies will be required to determine which PST domain function is responsible for the genomic outcomes observed in carriers in our cohort and if PST domain mutations confer the same aggression and DNA damage response as gene knockdown in functional experiments. Whether PARP inhibition affects survival in patients with SSM in *MDC1* in the same manner as HGSOC *BRCA1/2* variant carriers needs to be investigated further, although in vitro studies in other cancers (including breast cancer) suggest MDC1 loss does sensitize cells to PARP inhibitors [44, 45]. Our cohort was selected based on the availability of paired primary and recurrent surgical tissues in a Biobank, which likely represents a patient population with very aggressive recurrent disease, as they underwent debulking surgery at recurrence. In this study we identified SSM in *MDC1* in 8 of 32 patients, representing 25% of the cohort. Given the relatively high frequency of SSM in *MDC1* observed in this study, if future studies are able to confirm a functional role for the loss of MDC1 in HRD and establish the frequency of such mutations in HGSOC it may be of significant benefit to include MDC1 in somatic testing panels that identify patients likely to benefit from PARP inhibition. There are currently 4 commercially available diagnostic panels for establishing HR deficiency in HGSOC that use targeted sequencing [46] where MDC1 could be included in the panel with relative ease; MyChoice CdxPlus [47], OncoDEEP [48], Sophia DDM Dx HRD CE-IVD [49] and Foundation One CDx [50]. Given that PARP inhibition is one of the few therapies available for HGSOC with clear evidence to increase patient survival, identifying additional women who would benefit from treatment represents an important opportunity for the clinical translation and improved patient outcomes. These results also may represent an approach to continue to identify sub-populations of HGSOC patients who benefit from personalized therapy regimes based on molecular biomarkers. Somatic profiling studies in increasingly large sample sizes may identify additional molecular vulnerabilities that can be targeted by existing therapies, offering some alternative clinical strategies for increasing survival in a highly lethal disease.

Outside of *MDC1*, homologous recombination was stable throughout disease. Only two patients had an altered disease status (1 reversion, 1 in-frame deletion) across the course of disease. Although the reversion rate is estimated to be up to 26% in HGSOC [51], we did not observe this in our cohort, possibly due to sample size. Similar to Gull et al. 2022 [14], somatic mutation profiles were largely conserved throughout the disease course from primary to recurrent tumors. SSM and SV burden were not significantly different in tumors from different time points in disease course from the same patient. Interestingly, large SVs affecting chromosomes 19 and 20 in HRP tumors were not unique to either primary or recurrent tumors. This implies these variants may be important for both tumorigenesis and maintenance. Large SVs on chromosome 17 were identified in almost 40% of HRD tumors (Fig. 3E). These SV are predominantly inversions, making it difficult to infer their functional effect, but suggesting that altering expression of genes in this region may aid in tumor and clonal expansion, but an outright loss is not beneficial. Inversions have also been shown to be associated with DNA methylation changes [6, 15] suggesting carriers of these inversions may have unique responses to environmental changes or therapeutic intervention. The proportion of shared events across HRD tumors was lower than HRP tumors, suggesting that the mutational landscape is relatively conserved across variant types, likely due to a preservation of clonal dynamics.

An assumption of a single initiating cancer clone that gives rise to a diverse polyclonal tumor that is subjected to purifying selection is the standard dogma for many cancer types [52, 53]. However, this mechanism was not observed in this study, wherein we observed sustained clonal and subclonal populations throughout chemotherapy, surgical resection, and tumor regrowth. This suggests that the drivers for chemoresistance are present within the primary tumor and chemotherapy does not drive the development of new clones, but rather filters out the clones that were chemo-sensitive. We did observe that subclonal SV tend to be deletions while clonal SV tend to be duplications (Fig. 5C). Duplications are less likely to lead to cell death [54] while deletions, particularly in DNA repair genes, lead to genomic instability [55, 57].These deletions are believed to be random and reflect the evolutionary process of increasing fitness.

Additionally, high grade serous ovarian tumors are frequently polyclonal, and their persistence suggests tumor clones act synergistically to facilitate tumor growth. Previous studies with smaller cohorts have seen polyclonal seeding of metastatic sites [9, 28] which was also observed in our cohort. Patients with polyclonal primary tumors (as defined by SV) tended to have polyclonal recurrent tumors with similar clonal compositions, and also displayed a shorter progression-free survival than patients initially presenting with monoclonal tumors (Fig. 4E). A more diverse set of clones would set up the tumor

to handle chemotherapeutic insults better, which may partially explain the higher mortality seen in HRP tumor bearing patients [58]. Consistent with our cohort, high levels of intratumoral heterogeneity (i.e. increased clone counts) have been shown to diminish immune responses and survival intervals [59].

It is possible that our study may have missed some of the complexity of the tumor clones due to single site sampling which cannot resolve spatial complexity and subclonal variants often not passing QC thresholds. Previous studies using ascites have noted more complex clonality [9]. A particular clone may have been absent from the sampling site but was relatively abundant in other regions. While sampling may have led to an underrepresentation of minor clones in some samples, which may have contributed to the inability to estimate some clonal structures, it is unlikely that this would have been the case in the majority of samples. Sampling sites were highly enriched for neoplastic tissue in both primary and recurrent samples, removing stromal contamination as a possible confounding factor in clone estimates. Multi-site sampling would likely have led to an increase in overall clone count but we do not believe that the lack of clonal evolution is heavily influenced by the study design.

Structural variant signatures, largely re-capitulating the genomic impacts of the presence of loss of *BRCA1/2* function, were also associated with reduced survival. Previous publications have identified the relationship between *BRCA1* loss leading to increased duplications, and *BRCA2* loss leading to increased deletions and their phenotypic outcomes [33]. However, some *BRCA1/2* non-carrier patients share similar SV and SV signatures to these two groups and potentially have similar molecular profiles. *BRCA1/2* non-carrier patients that display deletor/duplicator signatures may benefit from the same interventions as *BRCA1* and *BRCA2* mutation carriers.

Recent reports have indicated that short reads are capable of capturing a vast majority of the complexity reported in long reads, and that the main advantage of long read sequencing is phasing [60]. It was unclear if these results would be replicated in genomically unstable HGSOC tumors, and although our sample size does not allow us to comment on HGSOC processes we did identify some trends within the sequencing strategies. Short read sequencing successfully identified many of the large SVs that accompanied copy number changes but was less accurate at calling > 300 bp deletions that had a less pronounced copy number effect. Insertions, templated and non-templated, are particularly difficult for short reads to identify, which led to overcalling duplications based on local sequencing depth from short read data (See Supplementary Fig. 13B, C, Additional File 2). As expected, most short (> 100 bp) deletions are accurately captured. Around half of the deletions were confirmed by long reads, suggesting that the structural variant signature groups dominated by deletions likely represent true genomic processes. Interestingly, the most replicated >5 Mb SV type was inversions, which are not typically associated with copy number changes, and therefore have fewer lines of evidence in short read sequencing (See Supplementary Fig. 13B, Additional File 2). While just over half of the short read variants were validated by long read sequencing, it is important to note that these data were generated from adjacent tissue punches, collected at different times, with different extraction methods. Given the high level of conservation of somatic variants across time points in the disease course, even after many treatments with chemotherapy, it seems unlikely that sampling variation at a single time point would significantly contribute to the large proportion of variants uniquely identified by long read sequencing, although it cannot be entirely ruled out. It is also feasible that multiple short variants on separate haplotypes were incorrectly called as a single long variant in short read data, inflating the number of unique variants identified with long read sequencing data.

While our study does represent one of the larger cohorts of matched primary and recurrent tumors with genomic profiling, we are still underpowered to comment on the effect that individual chemotherapy regimens may have had on the molecular profile of tumors. As patients' disease progressed additional lines of treatment were added, making the treatment regimen beyond the first recurrence extremely heterogenous across the cohort. To separate out the variants that drive recurrence and those associated with individual chemotherapy agents, we will need a larger cohort. Still, we have shown that inter-tumor heterogeneity is pronounced in HGSOC and that tumors from the same patient more closely resemble one another than any other, highlighting a scenario by which each patient develops their own largely private set of somatic mutations, albeit through shared mutational mechanisms. This underscores the need for individualized patient care based on tumor profiling to identify alternate therapies that may target specific molecular vulnerabilities in each patient.

Conclusions

Clonal profiles of HGSOC tumors are established prior to chemotherapy, despite the extensive clonal profiles previously identified. Structural variant signatures were also remarkably stable, and we highlighted the improved SV identification resulting from long read sequencing. Additionally, MDC1 loss was identified as a potential mechanism by which seemingly HRP tumors gain extensive genomic alterations.

Methods

Specimen acquisition and preparation

Fresh-frozen matched primary (n = 34) and recurrent (n = 38) high-grade serous ovarian cancer specimens and DNA from thirty-three women diagnosed with high grade serous adenocarcinoma were included. While 11 participants were known carriers of deleterious *BRCA1* and/or *BRCA2* germline variants, the remaining participants were not carriers of deleterious variants in any known high or moderate effect breast or ovarian cancer risk genes. All patients were diagnosed with stage III or stage IV disease and underwent primary optimal surgical cytoreduction (to less than 1 cm residual disease) prior to administration of combination chemotherapy with platinum and taxane between the years of 1990 and 2014. At the time of surgical debulking, prior to chemotherapy, germline DNA was collected from circulating blood and primary tumor samples were fresh-frozen shortly after collection [14]. At patient relapse, where tumor debulking was indicated, recurrent tumor samples were also fresh-frozen shortly after collection. For each patient, detailed clinical data were available including clinical genetic testing results, dates of original diagnosis and each subsequent recurrence, treatments administered throughout their disease course, operative and pathology reports, and other clinicopathologic variables including other cancer diagnoses and comorbidities. These details can be found in Supplementary Table 1, Additional File 1. Tumors were annotated and areas of pure neoplastic tissue were extracted as explained in the "Protein extraction" methods. The tumor samples were sequenced at a depth of 80x (mean 67x; See Supplementary Table 2, Additional File 1). Stranded total RNA sequencing (RNA-Seq) was performed and generated a mean of 335 million reads per library, and whole genome bisulfite sequencing (WGBS) was performed to capture methylation at a mean depth of 30x, as previously reported [14].

Germline DNA was extracted from whole blood drawn into 4mL EDTA treated collection tubes at the time of debulking surgery. DNA was extracted with the Qiagen DNEasy Blood & Tissue Kit (Qiagen, MD, USA) and quantitated with the Quant-IT dsDNA Broad Range kit on a QuBit (Thermo Fisher Sci, CA, USA). Fresh-frozen tumors were embedded in optimal cutting temperature (OCT) compound, bisected and mounted and two slides were made for hematoxylin and eosin (H&E) staining. All slides were reviewed by a single pathologist to identify regions enriched for epithelial carcinoma (avoiding regions enriched for stroma). These regions were sampled in a single core of approximately 50 mg collected on dry ice. Each core was divided into three pieces, one of which was used for genomic DNA (gDNA) extraction using the Macherey-Nagel Nucleospin DNA Kit (Machery Nagel, Germany). Remaining tissue was used for gene expression profiling and proteomics profiling ([14] and [20]). DNA samples were quantified using the QuBit (Thermo Fisher Sci, CA, USA) to measure the content of double stranded DNA and quality was confirmed using the Agilent TapeStation (Agilent, CA, USA).

Library preparation

gDNA was diluted in water to a concentration of 50ng/ul and a total of 1ug of DNA was shipped to Fulgent Therapeutics (Temple, CA, USA) for whole genome sequencing. DNA samples were sheared using the Covaris m220 to fragment size 400–500 bp, which was confirmed with the Agilent TapeStation. Agencourt AMPure Beads were used for fragmented DNA purification and size selection. The NEBNext® Ultra™ II DNA Library Prep kit | NEB and Illumina TruSeq Index Adapters were used to generate PCR-free sequencing libraries following manufacturers protocols.

Whole genome sequencing

Libraries were sequenced on either a single lane (germline DNA) or two lanes (tumor DNA) of the HiSeqX (Illumina, CA, USA). Fastqc and Picard Tools (*CollectAlignmentSummaryMetrics*, *CollectWGSMetrics*, *CollectGcBiasMetrics*) were used to generate QC reports that were visualized in MultiQC. Samples with insufficient read depth were excluded and an additional sample aliquot was repeated using the methods described above. Sequencing reads were aligned to hg38 using a custom BWA-GATK pipeline in the Cancer Genomics Cloud, in which indels were realigned with IndelRealigner, base quality estimated, and base quality recalibrated with BaseRecalibrator [61]. Coverage metrics, GC bias, and mismatch rate and chimera rates were collected with picardtools [62], coverage was calculated using bedtools genomecov [63]. These metrics are summarized in Supplementary Table 2, Additional File 1. Matching germline and somatic samples were verified using PLINK [64].

Amplicon sequencing

Somatic mutations in *MDC1*, *TP53*, *KRAS*, *BRCA1*, *BRCA2*, *RB1*, *MYC*, and *CCNE1* were sequenced using targeted amplicon sequencing. Tumor derived DNA used for short read Illumina sequencing was aliquoted (50ul of DNA at 18ng/ul) and mixed with 10ul of 2x Ampliseq custom primer pool (REF:20020495, lot:10841836, Illumina, CA, USA), and 5ul of Ampliseq HiFi mix were added per sample. The Ampliseq custom primer pools contained 100ng of each primer set (Pool 1: 347 primer pairs, Pool 2: 342 primer pairs). Libraries were quantified and pooled and sequenced on the Illumina NovaSeqX + to an average of 5 M reads and 9562x coverage. Sequencing reads were aligned with BWA [65] and variants called in tumor only mode using Mutect2 [61].

Long read sequencing

An additional punch was collected from epithelial enriched regions of the same fresh frozen specimen and DNA extracted using the Monarch® HMW DNA Extraction Kit for Tissue (New England Biolabs, MA, USA) and quantified using the Qubit method described above. Samples were profiled for fragment analysis on the FemtoPulse (Agilent, CA, USA). Ultra long sequencing libraries were generated using the Ultra Long read Kit (ULK114) for tumor DNA, and the Ligation Sequencing Kit V14 (SQK-LSK114) for germline samples and sequenced on the ONT Promethion FlowCell vR10.4.1 on the Promethion P24 (Oxford Nanopore Technologies, Oxford, United Kingdom). Germlines were processed on as many flow cells as possible to achieve 100Gb of data. Each tumor derived library was sequenced on a single flow cell. Tumor derived libraries with less than 50GB of data generated from a single sequencing run, corresponding to ~ 15x coverage, were excluded. Reads were aligned to hg38 using the map-ont option in minimap2 [66]. Clair3 [67] identified SNVs in the long read data, which was used in haplotagging the bam file with Longphase [68]. N50 of phase blocks was determined using a script provided in the longphase github. Haplotagged somatic and germline bam files were used by Severus [36] to identify somatic structural variants and variant clusters. Severus default parameters of at least 3 supporting reads, SVs of > 50 bp, and a minimum VAF of 0.05 were used. Shared SVs between ultra long and short read sequencing datasets were determined by intersecting SVs called in each dataset and retaining only variants called as the same SV type with > 80% reciprocal overlap between each dataset.

Cell lines and culture conditions

Kuramochi cells purchased from JCRB and CaOv3 cells purchased from ATCC were grown and maintained in RPMI 1640 (Thermo Fisher) and supplemented with 10% FBS (Thermo Fisher) and 100 U/mL Penicillin-Streptomycin (Gibco). FT246 and FT282 were donated by Cedars Sinai and maintained in DMEM F12 50/50 (Corning) supplemented with 10% FBS (Thermo Fisher), 2mM L-Glutamine (Thermo Fisher), and 100 U/mL Penicillin-Streptomycin (Gibco). Small interfering RNA targeting MDC1 and a negative human control were purchased from Genscript and transfected in 12 well plates with cell specific media and FBS conditions (5%FBS for Kuramochi, 1% FBS FT282, 0% for CaOv3 and FT246) using lipofectamine RNAiMAX (Thermo Fisher) according to manufacturers protocol. Cells were transfected for 48 h before fixing for γH2AX analysis, scratching for the wound healing assay, or re-seeding for transwell migration and invasion assays. For RNA analysis cells were allowed to grow for 24 h post transfection in growth media before RNA isolation to mimic the protocols of the other assays. For the wound healing assays a scratch was made in the transfected well using a sterile 1mL pipette tip after the 48 h incubation. Images were taken at the time of scratch and transfection media replaced with growth media. 24 h later a second image was taken and relative migration assessed using the ImageJ Wound Healing Tool (RRID: SCR_025260). For the Kuramochi cell line the scratch was made 24 h after transfection to allow for 48 h of migration time without allowing re-expression of MDC1. For γH2AX staining the transfected wells were fixed using 4% formalin for 15 min after the 48 h transfection incubation. Wells were blocked and the antibody diluted according to manufacturers protocol. Cells were stained overnight at 4 C with 1:200 anti-Phospho-serine 139 Histone H2A.X Rabbit monoclonal antibody conjugated with Alexa Fluor® 488 (clone 20E3, #9719 Cell Signaling Technologies), washed, stained for 5 min with 300nM DAPI (Thermo Fisher), washed, and imaged using the fluorescence capture on an ECHO revolve microscope. Images were analyzed using the positive cell detection feature of the QuPath software [69]. When reseeding was necessary 25 thousand FT246 or 100 thousand of the other cell lines were seeded into 300ul or 500ul of serum free, 1X pen/strep, media in transwell inserts for the migration and invasion assays, respectively. Transwells were placed in 750ul of growth media and 24 h after re-seeding transwell inserts were removed and fixed with 4% formalin for 30 min at room temperature. Transwells were stained with 1:20 Giemsa stain for 4 h and images were taken using the bright-field capture on an ECHO revolve microscope. For MTT ((3-[5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assay cells were seeded 48 h after transfection at the same density used for migration into individual wells of a 96 well plate and assayed according to manufacturer's protocols (Millipore Sigma). Briefly, 24 h after seeding cells were grown in MTT for four hours before stabilizing overnight. Colorimetric changes were assessed using a microplate spectrophotometer (Agilent Technologies) reading each well at 570 nm.

Protein extraction

A 3 mm core of approximately 50 mg was extracted from a region of pure epithelial carcinoma from the intact frozen tumor as previously stated, with care not to include any of the surrounding stroma. These pure neoplastic tissue cores were used for subsequent proteomic and genomic analyses. We also extracted regions adjacent to the pure neoplastic tissues which were either a mix of tumor and stromal samples (labeled as 50% stromal content) or stromal samples (labeled as 100% stromal content). Frozen tumors and adjacent stromal samples were thawed on ice, 200–400 µL lysis buffer

(6 M Urea + 0.1% Rapigest) was added, and tissues were homogenized with a polytron. Proteins were extracted by high-pressure barocycling on a Pressure BioSciences instrument, model 2320EXT (PBI is Easton, MA) at room temperature, ramping pressure to 45 kPSI, holding for 50 s, returning to atmospheric pressure for 10 s, and repeating for 60 cycles. Protein concentration was assayed using the Pierce BCA assay. Three data sets were prepared, a pool of all available protein lysates (pooled homogenate acquired at 2, 4, 8 µg, each with three technical replicates), a pilot set of tumor samples from 6 HGSOC tumors with either 2 or 3 technical replicates each ($n = 14$) and a clinical data set of tumor and tumor adjacent stromal samples from 16 individual patients, 4 of whom had more than one primary tumor or adjacent stroma sample available for profiling ($n = 41$).

Proteomics data acquisition

Data acquisition was performed as described previously in [70] and performed by others on the same publicly available dataset [71]. Database searches for both the internal and downloaded external datasets utilized human protein sequences defined in a FASTA database of Swiss-Prot-reviewed, Human canonical reviewed proteome containing 20,406 protein sequences that were downloaded July 2019 and appended with Biognosys indexed retention time (iRT) peptide sequence (Biognosys, Schlieren, Switzerland) and randomized decoy sequences appended. A final, combined consensus spectral library containing all peptide identifications made between the internal and external datasets was compiled. Decoy sequences were appended using the transproteomic pipeline with Spectrast library generation and conversion to TraML format as described previously [72]. Peptide identification was performed as previously described where tumor purity was confirmed [72, 73]. MS data and supplementary analysis files were submitted to ProteomeXchange via the PRIDE database [74] under the identifiers PXD023012 for the dilution series, PXD022996 for the small pilot data set, and PXD023040 for the patient tumor proteomic data set.

Quantification and statistical analysis

Mutation calling

A consensus pipeline for calling indels and SNV was constructed using Mutect2 [61], Strelka2 [75], VarScan2 [76], and Somatic Sniper [77], in which a variant must appear in at least two separate callers to be retained. Variants with fewer than 5 supporting reads were removed. Further filtering was applied to SNV wherein a base quality score of lower than 20, those located in regions of the lowest 1% mappability score [78], and variants within intervals included in the Encode Blacklist [79] were removed. Samples that did not have a somatic driver mutation idenifited in the automated variant calling and consensus pipeline in *TP53* were manually inspected in IGV [80] according to best practices [81].

Short read SVs: GRIDSS

The GRIDSS [82] suite was utilized to generate SV and copy number calls genome-wide. This analysis suite includes several analysis tools, including Gripss (a filtered version of the GRIDSS SV calls), Cobalt (which determines read depths), Amber (which determines BAF), Purple (PURity and PLoidy Estimator), and Linx (grouping and clustering SV events) [82]. A workflow chaining these individual programs was designed within the Cancer Genomics Cloud that incorporated Gripss v 2.10.2, Cobalt v 1.1, Amber 1.9, Purple v 2.51, and Linx v 1.14. Tools were run according to standard settings.

PCGR: Mutation rate and chromothripsis

We utilized the dockerized version of PCGR (version 0.9.2) [83] to estimate the mutation rate and MSI status of each tumor sample in our cohort. Chromothripsis was determined via Shatterseek [84], in which regions of at least three oscillating copy number profiles are colocalized with at least four nested structural variants are marked as likely chromothriptic. The output of Shatterseek was manually curated for accuracy of copy number oscillation and SV events, with only those meeting the copy number and SV criteria being retained.

Significantly mutated genes

Genes mutated at a significant rate above the background mutation rate were identified using ActiveDriverWGS [85]. A list of all somatic variants was cross referenced with a table of all coding regions from the Gencode v34 (gencode.v34.annotation.gtf), which was downloaded from the UCSC table browser. Mutated genes appearing with at least eight observed mutations and a p value < 0.005, and an FDR rate lower than 0.005, were determined to be significantly mutated. Mutations were then manually inspected to verify the surrounding regions were not subject to misalignment or high discrepancy regions [81].

PRS

The polygenic risk score (PRS) of an individual i was calculated by a linear function of $PRS_i = \Sigma_i X_{ij}\ \beta_j$, where X_{ij} represents the number of effect alleles at the jth SNP for the ith individual, and β_j represents the log odds ratio of the jth SNP. Genotypes are denoted as X, with values 0, 1, or 2, indicating dosage of the effect allele. The PRS model for overall EOC comprises 15 SNPs, which were obtained from [86], along with their corresponding log odds ratio. To calculate PRS for each patient, we used PLINK 2.0 [87].

PIK3CA/B amplification

Copy number profiles for genes were generated by utilizing b allele frequencies analyzed through the Purple function in the GRIDSS suite [82]. If a gene contained a copy number that rounded up to 6 or higher it was deemed an "amplification". *CCNE1* and *RB1* copy number profiles above 3.5 were deemed an "amplification".

Whole genome duplication

Copy number profiles generated by PURPL (within the GRIDSS suite) were used to determine if whole genome duplication had occurred. Whole genome duplication was defined as >70% of the genome displaying a copy number profile above 4.

Shared SNV, indel, and SVs

SNV and indel calls generated by the consensus pipeline were intersected with bedtools [63] intersect function with exact matching to determine quantities of private and shared variants. SVs were intersected with a minimum overlap of 70% between both SV for them to be counted as the same variant. Translocations with both breakpoints within 10 bp of another translocations breakpoints were merged and considered as a single event.

Homologous recombination status prediction

HRD was determined by considering several lines of genomic evidence; germline carrier status of deleterious variants in *BRCA1/*2, somatic promoter hypermethylation of *BRCA1/2*, *RAD51C*, or *RAD51D*. Although these markers are well established in the literature they do not fully capture homologous repair deficiencies. CHORD score was determined using the Classifier of HOmologous Recombination Deficiency (CHORD) program, but ultimately not used. SNV and INDEL information extracted from Mutect2 VCF files and SV information extracted from gripss VCF files, the filtered high-confidence output from GRIDSS, were used as inputs. Of the 5 tumors containing *BRCA1/2* promoter methylation and the 4 tumors with *RAD51C* promoter methylation, only two were determined to be HRD by CHORD. Germline deleterious variants and somatic mutations in *BRCA1* did not lead to CHORD HRD determination for 13/22 *BRCA1* germline and somatic mutation carriers (See Supplementary Table 1, Supplementary Table 2, Additional File 1). In light of the literature linking methylation and BRCA1/2 variants to HRD we opted to use all lines of sequencing evidence rather than an HRD tool. As we relied on bulk sequencing we cannot parse out any HRD spatial heterogeneity, which may be present in some samples.

Clonality analysis

To determine allele and clone specific copy number alterations along with clone proportions we utilized HATCHet [88]. This gave clonal composition and proportions for each mutation present within the sequencing data. The clonal matrix was cross referenced with the consensus file to remove the mutations that were more likely to be false positives. DeCifER [89] determined the clone that each variant belonged to and the descendant cell fraction (DCF) of each variant. Per DeCiFer's documentation, only mutations that appeared in both the primary and recurrent tumor were considered for clonal evolution. Unique state trees were generated for each tumor sample where possible. ClonEvol [90] interpreted the clonal structure, as defined by DeCifER, and the mean DCF of each clone, to generate clonal evolutionary structure. Fishplots for each tumor were generated by providing ClonEvol with DeCifER clonality results. True cluster DCFs were used for SNV estimations, in the cases when no consensus model could be found the point estimate DCF was used to guide fishplot generation. Resolving the clonal structure was not possible for 6/33 samples, which may be due in part to single site sampling. In order to determine how SV abundance and CCF of SV breakpoints affected clonal composition, we used SVclone [91]. Currently SVclone cannot accurately profile the temporal progression of SVs, therefore tumors from the same patient were analyzed separately.

SV signatures

Structural variant calls generated by GRIDSS were further analyzed by the "Palimpsest" [92] R package. The package determined if SVs were clustered by initiating events or occurred in an unclustered manner. Total SV counts and SV sizes were utilized to generate denovo SV signatures. Each tumor was analyzed for signature content and assigned a signature content value for each signature identified in the cohort. Within the data set there were several tumors that only displayed a single signature while others displayed up to four different signatures per tumor. Tumors were clustered with a euclidean clustering algorithm and the resulting dendrograms were used to determine SV signature groups.

Clonality defined by SVs and subclonal signatures

Structural variant calls generated by GRIDSS, in the VCF format, along with bam files were input in the SVclone [91] program. Read information present in the bams was used to adjust VAF for each breakpoint. After a filtering step, the SVs were clustered according to CCF, copy number intervals, and tumor purity. The Ccube clustering method was applied to the CCF adjusted breakpoints to determine cluster assignment for each SV. A joint SV and SNV clustering method was applied to the sample.

SVclone does not currently support multi-tumor analysis, and other standard tools, such as pyclone [93] or DeCifER [89], do not officially support SV clonal analysis. Therefore, we did not analyze temporal evolution of clones as defined by their SV profiles. Variants assigned to clone 1 were determined to be clonal while variants below a VAF of 80% or assigned to any other clones were designated as subclonal. We applied the five denovo signatures found in our total SV dataset to both the clonal and subclonal SV datasets. A total of 24 tumors were analyzed for subclonal and clonal SV signatures. Signature abundances were compared between clonal and subclonal signatures, with a p value < 0.05 considered significant. SVs were visualized as circos plots generated using the circlize R package [94]. SVs are arranged according to the percent of the chromosome that they affect.

Pathogenic structural variants

Svanna [95] was used to annotate SVs with functional annotations to predict deleterious effects on coding regions at SV sites. SVanna generates a ranked list based on intersections with predicted functional domains of coding regions, and membership in the top 10 is highly correlated with pathogenicity of a structural variant. To further select for highly pathogenic SV we retained only the top 5 predicted pathogenic SVs from each tumor. Predicted pathogenic SVs were filtered to include SVs ranked in the top 10 with a Svanna score > 2500. This cutoff represented an inflection point in the distribution of the scores across the annotated SVs in the study.

Statistical analysis

Statistical tests were performed in R using rstatix [96]. All reported statistics were two tailed students T tests, unless otherwise stated. Survivorship analysis was performed using Kaplan-Meier plots that were generated using the "survival" R package [97].

GO analysis

GOrilla [98] was used to annotate gene ontology of coding regions impacted by SVs. To create the input, all structural variants were intersected with the Gencode v34 gene annotations, available through the UCSC genome browser's [99] table browser [78]. This produced a data table with genes affected and the type of structural variant affecting them. To keep likely functional effects separate, we separated duplications with breakpoints within the coding region from duplications covering the entire gene. The list of genes affected by SV was sorted by the number of unique occurrences each gene was impacted by a given SV type across each tumor in our cohort.

Copy number depth profiles

Read depth information was extracted from the bam files using mosdepth [100]. The tool measured depth from BAM files in genomic windows of 10 kb. Relative depth was determined by dividing the depth at a given region within the tumor sample by the total depth of the same region within the corresponding normal.

SNV Count: genomic region cartoon

MDC1 gene region figure was produced using pyGenomeTracks [101] using python version 3.7and RefSeq Gene hg38 gtf. Genome was plotted from chromosome 6 position 30,695,807 to 30,720,000.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-026-07951-3.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We would like to acknowledge the generous contributions made by the patients who participated in this study.

Author contributions

MD, NG, SJP, SAG, and MRJ were involved in the study design, BJR, AJL, and BYK were involved in sample collection. MD, NG, P-CP, KD, JB, JS, JA, BB, NM, and CS were involved in sample processing, data acquisition and management. MD, NG, P-CP, SAG, and MRJ were involved in data analysis and interpretation. MD, NG, and MRJ were involved in manuscript writing. MD, NG, SAG and MRJ made the decision to submit for publication.

Funding

This work was supported by the Cedars-Sinai Medical Center Precision Healthy Initiative, NIH/NCI R01CA211575 (S.A.G). P-C.P. is supported by National Cancer Institute (K99CA256519). S.A.G is supported through kind donation from the Barth Family as Chair of Cancer Genetics. The study sponsor(s) did not contribute to the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Data availability

Project data will be made available prior to acceptance of the article. Code to generate most figures as well as unique code required for this study will be uploaded and maintained on Github by the Lead Contact, Michelle Jones: https://github.com/Jones-Lab-CSMC. All whole genome and targeted sequencing data will be made publicly available in SRA upon publication of the manuscript under bioproject PRJNA1214483. Gene expression (RNA-Seq) and methylation (WGBS) data are publicly available at the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under GSE316442 and GSE202245. MS data and supplementary analysis files were submitted to ProteomeXchange via the PRIDE database [74] under the identifiers PXD023012 for the dilution series, PXD022996 for the small pilot data set, and PXD023040 for the patient tumor proteomic data set.

Declarations

Ethics approval and consent to participate

Study participants were identified from the Cedars Sinai Medical Center Women's Cancer Program Biorepository, which was established and maintained under IRB Protocol# 901 and samples profiled under IRB Protocol #41922. Each protocol was reviewed and approved by the Institutional

Review Board and conducted according to the principles of the Declaration of Helsinki. All patients gave informed consent to participate in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[2]Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[3]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[4]Applied Genomics, Computation and Translation Core, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[5]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[6]Department of Obstetrics and Gynecology, Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, UCLA, Los Angeles, CA 90048, USA
[7]Proteomic and Metabolomics Core Facility at the Board of Governors Innovation Center, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[8]Department of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[9]Center for Inherited Oncogenesis, Department of Medicine, UT Health San Antonio, San Antonio, TX 78229, USA

Received: 30 September 2025 / Accepted: 22 February 2026
Published online: 06 March 2026

References

1. Jervis S, Song H, Lee A, Dicks E, Tyrer J, Harrington P, et al. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. J Med Genet. 2014;51:108–13.
2. Lisio M-A, Fu L, Goyeneche A, Gao Z-H, Telleria C. High-grade serous ovarian cancer: basic sciences, clinical and therapeutic standpoints. Int J Mol Sci. 2019;20.
3. O'Malley DM, Krivak TC, Kabil N, Munley J, Moore KN. PARP inhibitors in ovarian cancer: A review. Target Oncol. 2023;18:471–503.
4. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009;59:225–49.
5. Rankin EB. Genomics and molecular mechanisms of high grade serous ovarian cancer: the 12th Biennial Rivkin Center Ovarian Cancer Research Symposium. Int J Gynecol Cancer. 2019;29(Suppl 2):s7–11.
6. Wang YK, Bashashati A, Anglesio MS, Cochrane DR, Grewal DS, Ha G, et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. Nat Genet. 2017;49:856–65.
7. Xie R, Yan Z, Jing J, Wang Y, Zhang J, Li Y, et al. Functional defects of cancer-associated MDC1 mutations in DNA damage repair. DNA Repair (Amst). 2022;114:103330.
8. Mailand N, Bekker-Jensen S, Faustrup H, Melander F, Bartek J, Lukas C, et al. RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell. 2007;131:887–900.
9. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet. 2016;48:758–67.
10. Eckert MA, Pan S, Hernandez KM, Loth RM, Andrade J, Volchenboum SL, et al. Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube. Cancer Discov. 2016;6:1342–51.
11. Lambrechts S, Smeets D, Moisse M, Braicu EI, Vanderstichele A, Zhao H, et al. Genetic heterogeneity after first-line chemotherapy in high-grade serous ovarian cancer. Eur J Cancer. 2016;53:51–64.
12. Burdett NL, Willis MO, Alsop K, Hunt AL, Pandey A, Hamilton PT, et al. Multiomic analysis of homologous recombination-deficient end-stage high-grade serous ovarian cancer. Nat Genet. 2023;55:437–50.
13. Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science. 2021;372.
14. Gull N, Jones MR, Peng P-C, Coetzee SG, Silva TC, Plummer JT, et al. DNA methylation and transcriptomic features are preserved throughout disease recurrence and chemoresistance in high grade serous ovarian cancers. J Exp Clin Cancer Res. 2022;41:232.
15. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.
16. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.
17. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell. 2021;184:2239–e225439.
18. Lou Z, Chen BP-C, Asaithamby A, Minter-Dykhouse K, Chen DJ, Chen J. MDC1 regulates DNA-PK autophosphorylation in response to DNA damage. J Biol Chem. 2004;279:46359–62.
19. Cui X, Yu Y, Gupta S, Cho Y-M, Lees-Miller SP, Meek K. Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. Mol Cell Biol. 2005;25:10842–52.
20. Dabke K, Gull N, Kreimer S, Peng P-C, Diaz M, Cheikh BB, et al. The mutational landscape defines the proteome and spatial organization of tumor, stroma and immune cells in ovarian cancer. BioRxiv. 2025. https://doi.org/10.1101/2025.06.01.655822
21. Li Y, Dou Y, Da Veiga Leprevost F, Geffen Y, Calinawan AP, Aguet F, et al. Proteogenomic data and resources for pan-cancer analysis. Cancer Cell. 2023;41:1397–406.
22. Xiang P, Yeung YT, Wang J, Wu Q, Du R, Huang C, et al. miR-17-3p promotes the proliferation of multiple myeloma cells by downregulating P21 expression through LMLN inhibition. Int J Cancer. 2021;148:3071–85.
23. Salguero I, Belotserkovskaya R, Coates J, Sczaniecka-Clift M, Demir M, Jhujh S, et al. MDC1 PST-repeat region promotes histone H2AX-independent chromatin association and DNA damage tolerance. Nat Commun. 2019;10:5191.
24. Ding C, Li Y, Wang S, Xing C, Chen L, Zhang H, et al. ROBO2 hampers malignant biological behavior and predicts a better prognosis in pancreatic adenocarcinoma. Scand J Gastroenterol. 2021;56:955–64.
25. Liu H, Xie S, Fang F, Kalvakolanu DV, Xiao W. SHQ1 is an ER stress response gene that facilitates chemotherapeutics-induced apoptosis via sensitizing ER-stress response. Cell Death Dis. 2020;11:445.
26. Herzig JK, Bullinger L, Tasdogan A, Zimmermann P, Schlegel M, Teleanu V, et al. Protein phosphatase 4 regulatory subunit 2 (PPP4R2) is recurrently deleted in acute myeloid leukemia and required for efficient DNA double strand break repair. Oncotarget. 2017;8:95038–53.
27. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47:D766–73.
28. Lahtinen A, Lavikka K, Virtanen A, Li Y, Jamalzadeh S, Skorda A, et al. Evolutionary states and trajectories characterized by distinct pathways stratify patients with ovarian high grade serous carcinoma. Cancer Cell. 2023;41:1103–e111712.
29. Feng Z, Shao D, Cai Y, Bi R, Ju X, Chen D, et al. Homologous recombination deficiency status predicts response to platinum-based chemotherapy in Chinese patients with high-grade serous ovarian carcinoma. J Ovarian Res. 2023;16:53.
30. Gull N, Jones MR, Peng P-C, Coetzee SG, Chedraoui Silva T, Plummer JT, et al. DNA methylation landscapes of matched primary and recurrent high grade serous ovarian cancers are preserved throughout disease progression and chemoresistance. BioRxiv. 2020. https://doi.org/10.1101/2020.08.25.267161
31. George J, Alsop K, Etemadmoghadam D, Hondow H, Mikeska T, Dobrovic A, et al. Nonequivalent gene expression and copy number alterations in high-grade serous ovarian cancers with BRCA1 and BRCA2 mutations. Clin Cancer Res. 2013;19:3474–84.
32. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. Nature. 2020;578:112–21.
33. Vázquez-García I, Uhlitz F, Ceglia N, Lim JLP, Wu M, Mohibullah N, et al. Ovarian cancer mutational processes drive site-specific immune evasion. Nature. 2022;612:778–86.

34. Funnell T, Zhang AW, Grewal D, McKinney S, Bashashati A, Wang YK, et al. Integrated structural variation and point mutation signatures in cancer genomes using correlated topic models. PLoS Comput Biol. 2019;15:e1006799.
35. Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA. 2012;307:382–90.
36. Keskus A, Bryant A, Ahmad T, Yoo B, Aganezov S, Goretsky A et al. Severus: accurate detection and characterization of somatic structural variation in tumor genomes using long reads. medRxiv. 2024. https://doi.org/10.1101/2024.03.22.24304756
37. Sakamoto Y, Miyake S, Oka M, Kanai A, Kawai Y, Nagasawa S, et al. Phasing analysis of lung cancer genomes using a long read sequencer. Nat Commun. 2022;13:3464.
38. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015;521:489–94.
39. Provencher DM, Lounis H, Fink D, Drouin P, Mes-Masson AM. Discordance in p53 mutations when comparing ascites and solid tumors from patients with serous ovarian cancer. Tumour Biol. 1997;18:167–74.
40. Ratnaparkhe M, Wong JKL, Wei P-C, Hlevnjak M, Kolb T, Simovic M, et al. Defective DNA damage repair leads to frequent catastrophic genomic events in murine and human tumors. Nat Commun. 2018;9:4760.
41. Foray N, Randrianarison V, Marot D, Perricaudet M, Lenoir G, Feunteun J. Gamma-rays-induced death of human cells carrying mutations of BRCA1 or BRCA2. Oncogene. 1999;18:7334–42.
42. Heyza JR, Mikhova M, Perez GI, Broadbent DG, Schmidt JC. The PST repeat region of MDC1 is a tunable multivalent chromatin tethering domain. BioRxiv. 2025. https://doi.org/10.1101/2025.01.10.632395
43. Xie A, Hartlerode A, Stucki M, Odate S, Puget N, Kwok A, et al. Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. Mol Cell. 2007;28:1045–57.
44. Wang Z, Zuo W, Zeng Q, Qian Y, Li Y, Liu C, et al. Loss of NFBD1/MDC1 disrupts homologous recombination repair and sensitizes nasopharyngeal carcinoma cells to PARP inhibitors. J Biomed Sci. 2019;26:14.
45. Sottnik JL, Shackleford MT, Nesiba CS, Richer AL, Fleischmann Z, Swartz JM, et al. Co-regulator activity of Mediator of DNA Damage Checkpoint 1 (MDC1) is associated with DNA repair dysfunction and PARP inhibitor sensitivity in lobular carcinoma of the breast. BioRxiv. 2025. https://doi.org/10.1101/2023.10.29.564555
46. Quesada S, Penault-Llorca F, Matias-Guiu X, Banerjee S, Barberis M, Coleman RL, et al. Homologous recombination deficiency in ovarian cancer: Global expert consensus on testing and a comparison of companion diagnostics. Eur J Cancer. 2025;215:115169.
47. Fumagalli C, Betella I, Ranghiero A, Guerini-Rocco E, Bonaldo G, Rappa A, et al. In-house testing for homologous recombination repair deficiency (HRD) testing in ovarian carcinoma: a feasibility study comparing AmoyDx HRD Focus panel with Myriad myChoiceCDx assay. Pathologica. 2022;114:288–94.
48. Lopacinska-Joergensen J, Oliveira DVNP, Poulsen TS, Hoegdall CK, Hoegdall EV. Somatic Variants in DNA Damage Response Genes in Ovarian Cancer Patients Using Whole-exome Sequencing. Anticancer Res. 2023;43:1891–900.
49. Pozzorini C, Andre G, Coletta T, Buisson A, Bieler J, Ferrer L, et al. GIInger predicts homologous recombination deficiency and patient response to PARPi treatment from shallow genomic profiles. Cell Rep Med. 2023;4:101344.
50. Milbury CA, Creeden J, Yip W-K, Smith DL, Pattani V, Maxwell K, et al. Clinical and analytical validation of FoundationOne®CDx, a comprehensive genomic profiling assay for solid tumors. PLoS ONE. 2022;17:e0264138.
51. Tobalina L, Armenia J, Irving E, O'Connor MJ, Forment JV. A meta-analysis of reversion mutations in BRCA genes identifies signatures of DNA end-joining repair mechanisms driving therapy resistance. Ann Oncol. 2021;32:103–12.
52. Wang J, Li B, Luo M, Huang J, Zhang K, Zheng S, et al. Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance. Signal Transduct Target Ther. 2024;9:83.
53. Zhang S, Xiao X, Yi Y, Wang X, Zhu L, Shen Y, et al. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. Signal Transduct Target Ther. 2024;9:149.
54. Schoeler S, Szibor R, Gellerich FN, Wartmann T, Mawrin C, Dietzmann K, et al. Mitochondrial DNA deletions sensitize cells to apoptosis at low heteroplasmy levels. Biochem Biophys Res Commun. 2005;332:43–9.
55. Maxwell KN, Wubbenhorst B, Wenz BM, De Sloover D, Pluta J, Emery L, et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun. 2017;8:319.
56. Gaudet MM, Kuchenbaecker KB, Vijai J, Klein RJ, Kirchhoff T, McGuffog L, et al. Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. PLoS Genet. 2013;9:e1003173.
57. Tiwari V, Wilson DM. DNA damage and associated DNA repair defects in disease and premature aging. Am J Hum Genet. 2019;105:237–57.
58. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481:306–13.
59. Wolf Y, Bartok O, Patkar S, Eli GB, Cohen S, Litchfield K, et al. UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. Cell. 2019;179:219–e23521.
60. Choo Z-N, Behr JM, Deshpande A, Hadi K, Yao X, Tian H, et al. Most large structural variants in cancer genomes can be detected without long reads. Nat Genet. 2023;55:2139–48.
61. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
62. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/. Accessed 23 Jul 2019.
63. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
64. PLINK. https://zzz.bwh.harvard.edu/plink/profile.shtml. Accessed 5 Jun 2022.
65. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
66. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
67. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nat Comput Sci. 2022;2:797–803.
68. Lin J-H, Chen L-C, Yu S-C, Huang Y-T. LongPhase: an ultra-fast chromosome-scale phasing algorithm for small and large variants. Bioinformatics. 2022;38:1816–22.
69. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7:16878.
70. Parker SJ, Rost H, Rosenberger G, Collins BC, Malmström L, Amodei D, et al. Identification of a Set of Conserved Eukaryotic Internal Retention Time Standards for Data-independent Acquisition Mass Spectrometry. Mol Cell Proteom. 2015;14:2800–13.
71. Thomas SN, Friedrich B, Schnaubelt M, Chan DW, Zhang H, Aebersold R. Orthogonal Proteomic Platforms and Their Implications for the Stable Classification of High-Grade Serous Ovarian Cancer Subtypes. iScience. 2020;23:101079.
72. Parker SJ, Venkatraman V, Van Eyk JE. Effect of peptide assay library size and composition in targeted data-independent acquisition-MS analyses. Proteomics. 2016;16:2221–37.
73. Parker SJ, Chen L, Spivia W, Saylor G, Mao C, Venkatraman V, et al. Identification of putative early atherosclerosis biomarkers by unsupervised deconvolution of heterogeneous vascular proteomes. J Proteome Res. 2020;19:2794–806.
74. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, et al. PRIDE: the proteomics identifications database. Proteomics. 2005;5:3537–45.
75. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15:591–4.
76. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25:2283–5.
77. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012;28:311–7.
78. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004;32 Database issue:D493-6.
79. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. Sci Rep. 2019;9:9354.
80. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinf. 2013;14:178–92.
81. Barnell EK, Ronning P, Campbell KM, Krysiak K, Ainscough BJ, Sheta LM, et al. Standard operating procedure for somatic variant refinement of sequencing data with paired tumor and normal samples. Genet Med. 2019;21:972–81.
82. Cameron DL, Baber J, Shale C, Papenfuss AT, Valle-Inclan JE, Besselink N et al. GRIDSS, PURPLE, LINX: Unscrambling the tumor genome via integrated

analysis of structural variation and copy number. BioRxiv. 2019. https://doi.org/10.1101/781013
83. Nakken S, Fournous G, Vodák D, Aasheim LB, Myklebost O, Hovig E. Personal Cancer Genome Reporter: variant interpretation report for precision oncology. Bioinformatics. 2018;34:1778–80.
84. Cortés-Ciriano I, Lee JJ-K, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet. 2020;52:331–41.
85. Zhu H, Uusküla-Reimand L, Isaev K, Wadi L, Alizada A, Shuai S, et al. Candidate Cancer Driver Mutations in Distal Regulatory Elements and Long-Range Chromatin Interaction Networks. Mol Cell. 2020;77:1307–e132110.
86. Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A, et al. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. J Med Genet. 2018;55:546–54.
87. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:s13742-015-0047–8.
88. Zaccaria S, Raphael BJ. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. Nat Commun. 2020;11:4301.
89. Satas G, Zaccaria S, El-Kebir M, Raphael BJ. Decifering the elusive cancer cell fraction in tumor heterogeneity and evolution. BioRxiv. 2021. https://doi.org/10.1101/2021.02.27.429196
90. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and visualization in cancer sequencing. Ann Oncol. 2017;28:3076–82.
91. Cmero M, Yuan K, Ong CS, Schröder J, PCAWG Evolution and Heterogeneity Working Group, Corcoran NM, et al. Inferring structural variant cancer cell fraction. Nat Commun. 2020;11:730.
92. Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. Bioinformatics. 2018;34:3380–1.
93. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods. 2014;11:396–8.
94. Circular Visualization in R. https://jokergoo.github.io/circlize_book/book/. Accessed 28 Sep 2023.
95. Danis D, Jacobsen JOB, Balachandran P, Zhu Q, Yilmaz F, Reese J, et al. SvAnna: efficient and accurate pathogenicity prediction of coding and regulatory structural variants in long-read genome sequencing. Genome Med. 2022;14:44.
96. Kassambara A. Pipe-Friendly Framework for Basic Statistical Tests [R package rstatix version 0.7.2]. 2023.
97. Therneau TM. survival-package. A Package for Survival Analysis in R. 2023. https://CRAN.R-project.org/package=survival. Accessed 28 Sep 2023.
98. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10:48.
99. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.
100. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34:867–8.
101. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun. 2018;9:189.

Publisher's note