





















The genomic and immune landscape of long-term survivors of high-grade serous ovarian cancer

Received: 6 January 2022

Accepted: 17 October 2022

Published online: 01 December 2022

 Check for updates

Dale W. Garsed ^{1,2}✉, Ahwan Pandey ¹, Sian Fereday ^{1,2}, Catherine J. Kennedy ^{3,4,5}, Kazuaki Takahashi^{1,6}, Kathryn Alsop^{1,2}, Phineas T. Hamilton⁷, Joy Hendley¹, Yoke-Eng Chiew^{3,4,5}, Nadia Traficante^{1,2}, Pamela Provan ^{3,4,5}, Dinuka Ariyaratne ¹, George Au-Yeung^{1,2}, Nicholas W. Bateman ^{8,9}, Leanne Bowes^{1,10}, Alison Brand^{4,5}, Elizabeth L. Christie ^{1,2}, Julie M. Cunningham ¹¹, Michael Friedlander¹², Bronwyn Grout¹³, Paul Harnett^{5,14}, Jillian Hung^{3,4}, Bryan McCauley¹⁵, Orla McNally^{1,10,16}, Anna M. Piskorz¹⁷, Flurina A. M. Saner ^{1,18}, Robert A. Vierkant ¹⁵, Chen Wang ¹⁹, Stacey J. Winham ¹⁹, Paul D. P. Pharoah ^{20,21}, James D. Brenton ^{17,21}, Thomas P. Conrads ^{8,22}, George L. Maxwell^{8,22}, Susan J. Ramus ^{23,24}, Celeste Leigh Pearce²⁵, Malcolm C. Pike²⁶, Brad H. Nelson ^{7,27,28}, Ellen L. Goode²⁹, Anna DeFazio ^{3,4,5,30,31} & David D. L. Bowtell ^{1,2,31}✉

Fewer than half of all patients with advanced-stage high-grade serous ovarian cancers (HGSCs) survive more than five years after diagnosis, but those who have an exceptionally long survival could provide insights into tumor biology and therapeutic approaches. We analyzed 60 patients with advanced-stage HGSC who survived more than 10 years after diagnosis using whole-genome sequencing, transcriptome and methylome profiling of their primary tumor samples, comparing this data to 66 short- or moderate-term survivors. Tumors of long-term survivors were more likely to have multiple alterations in genes associated with DNA repair and more frequent somatic variants resulting in an increased predicted neoantigen load. Patients clustered into survival groups based on genomic and immune cell signatures, including three subsets of patients with *BRCA1* alterations with distinctly different outcomes. Specific combinations of germline and somatic gene alterations, tumor cell phenotypes and differential immune responses appear to contribute to long-term survival in HGSC.

Patients diagnosed with advanced HGSC have a 5-year survival rate of 41%¹, and fewer than 15% survive more than 10 years². Treatment usually consists of debulking surgery followed by adjuvant platinum and paclitaxel-based chemotherapy, or increasingly, neoadjuvant chemotherapy and interval debulking surgery³. Tumor stage and the extent of surgical removal are important clinical predictors of patient survival^{4,5}.

HGSC has the highest frequency of germline alterations in homologous recombination DNA repair genes including *BRCA1* and *BRCA2*^{6–8} and is among the most chromosomally unstable of any cancer type⁹ with near ubiquitous somatic *TP53* alterations^{7,10}. Prognostic biomarkers include gene expression-based molecular subtype¹¹, *CCNE1* gene amplification¹², tumor immune cell infiltration¹³, and tumor DNA repair

A full list of affiliations appears at the end of the paper. ✉ e-mail: dale.garsed@petermac.org; david.bowtell@petermac.org

status^{14–16}. Approximately 50% of HGSC have defects in homologous recombination-mediated DNA repair pathway genes^{7,14} and homologous recombination defective cancers show increased sensitivity to platinum and inhibitors of poly(ADP-ribose) polymerase 1 (PARPi)^{17–20}. Patients with germline *BRCA1* or *BRCA2* mutations have a longer 5-year survival than noncarriers^{6,21}, although this survival advantage is lost in patients with *BRCA1* mutations over time²².

A subgroup of patients with HGSC with apparently poor prognosis disease at presentation have a remarkable response to treatment and extraordinary long-term survival, including a small number with incomplete removal of macroscopic disease following surgery²³. The extent to which known clinical, immune and molecular biomarkers can explain exceptionally long survival in HGSC is unclear. Here, we genomically characterize patients with HGSC who have survived more than 10 years after diagnosis.

Results

Long-term survivor cohort

We accessed Australian and United States ovarian cancer biobanks with detailed, longitudinal clinical follow-up data collection to ascertain 60 long-term survivors (overall survival (OS) greater than 10 years from diagnosis). These were compared with patients from our earlier study¹⁴ that included 34 short-term survivors (OS < 2 years) and 32 moderate-term survivors (OS ≥ 2 and < 10 years; Extended Data Fig. 1a). All patients had advanced-stage HGSC (stage IIIC–IV), and 70% (42/60) of long-term survivors were alive at last follow-up, including 72% (43/60) with macroscopic residual disease at the conclusion of primary surgery, which is a well-accepted adverse prognostic factor (Extended Data Fig. 1b and Supplementary Table 1). Among the long-term survivors with residual disease were a subset with no disease recurrence (51%, 22/43), indicating an exceptional response to primary treatment.

Pervasive DNA repair pathway alterations

We analyzed data from whole-genome sequencing (WGS; mean coverage 64× tumor and 40× normal DNA), RNA sequencing (RNA-seq; average 115 million paired reads) and methylome analysis on 126 cases: primary tumor samples from 60 long-term survivor patients, and leveraging existing sequencing data from our previous study¹⁴, 7 of the included long-term survivors, 34 short-term survivors and 32 moderate-term survivors (Supplementary Tables 2–4).

The combined germline and somatic homologous recombination alteration rate in long-term survivors (76.7%, 46/60) and moderate-term survivors (78.1%, 25/32) were similar but were higher compared with short-term survivors (38.2%, 13/34; $P = 0.0012$; Fig. 1a). These included germline mutations in *BRCA1*, *BRCA2*, *BRIPI*, *PALB2* and *RADS1C*, somatic mutations in *BRCA1*, *BRCA2*, *ATM*, *CDK12*, *PTEN*, *RADS1B*, *RADS1C* and *RADS1D* (Supplementary Tables 5 and 6) and promoter methylation in *BRCA1* and *RADS1C*. Consistent with previous findings^{24,25}, *CCNE1* gene amplification was largely mutually exclusive with *BRCA1* (false discovery rate adjusted P value (P_{adj}) = 0.0169; co-occurrence in 1/126 primary tumors, 0.79%) and *BRCA2* alterations (P_{adj} = 0.4554; co-occurrence in zero cases; Supplementary Note and Fig. 1b) and more prevalent in short-term survivors (Fig. 1a). The tumors of six long-term survivors showed *CCNE1* amplification (Fig. 1b), an unexpected finding, as it is an established poor prognostic marker associated with primary platinum resistance^{12,14}. Inference of immune cell subsets from transcript data²⁶ revealed enrichment of activated CD4 memory T cells (P_{adj} = 0.0050) and CD8 T cells (P_{adj} = 0.0100) in *CCNE1* amplified tumors in long-term survivors ($n = 6$) compared to short-term survivors ($n = 11$; Supplementary Note).

We found examples of multiple co-occurring mutations in genes involved in chromosome stability and DNA repair (Fig. 1b), most commonly due to structural variants that interrupted open reading frames. Long-term survivors had a higher proportion of tumors (28.3%, 17/60)

with three or more altered DNA repair genes compared to moderate- (15.6%, 5/32) and short-term survivors (5.9%, 2/34; $P = 0.0224$; Fig. 1c). Patients whose tumors exhibited three or more DNA repair gene alterations had longer OS (median OS, 11.2 years) compared to those with two (median OS, 5.8 years), one (median OS 8.6 years) or no DNA repair gene alterations (median OS, 2.2 years; $P = 0.0136$; Fig. 1d). DNA repair pathway alterations were ranked from highest to lowest cancer cell fraction in each sample, finding that the majority were clonal in the first- (95.4%, 83/87) and second-ranked alterations (72.1%, 31/43), whereas in tumor samples with more than two DNA repair alterations, the third and fourth alterations were more likely to be subclonal (Supplementary Note).

Homologous recombination-deficient (HRD) tumors rely on error-prone DNA repair such as non-homologous end joining, which generates distinct mutational scars²⁷. We confirmed the functional impact of homologous recombination alterations using CHORD²⁸, integrating base substitution, small-scale insertion and deletion (indel), and structural rearrangement signatures to classify tumor genomes as *BRCA1*-type HRD, *BRCA2*-type HRD, or homologous recombination proficient (Fig. 1b). Among tumors considered to be HRD (CHORD score > 0.5), for almost all (97.1%, 67/69), we identified the likely homologous recombination gene alteration driving the signature. Although generally either the *BRCA1*-type or *BRCA2*-type score was dominant, some tumors showed evidence of a mixture of both signatures, and in some cases, this finding could be attributed to two or more altered homologous recombination genes (Fig. 1b). MMAY00758 was a long-term survivor patient who experienced >11 years progression-free and had three homologous recombination pathway gene alterations, with a germline *RADS1C* missense mutation and somatic structural variants in *BRCA1* and *BRIPI* and evidence of both *BRCA2*-type (0.68) and *BRCA1*-type (0.27) HRD scores (Fig. 1b; for additional examples, see Supplementary Note). Mutations in *CDK12* have been postulated to contribute to an HRD phenotype²⁹; however, an assessment of mutational scarring indicated that these tumors were homologous recombination intact (Fig. 1b).

Recurrent mutations in long-term survivors

We confirmed our previous findings¹⁴ of ubiquitous *TP53* mutations, infrequent nonsynonymous single-nucleotide variants (SNVs) and indels in other cancer-associated genes, and common somatically acquired structural variants and copy-number alterations (Supplementary Data 1–3, Supplementary Table 7 and Supplementary Note), including disruption of *RBI*, *NFI*, *PTEN* and *RADS1B*. *MYH9*, *EZH2*, *ARID1B*, *TBLIX1*, *ARID1A*, *YWHAE*, *CREBBP*, *RHOA*, *ATRX*, *AXIN1* and *STAG1* were identified as also disrupted by gene breakage ($P_{\text{adj}} < 0.1$; Supplementary Data 2). Despite frequent gene breakage in HGSC, only two recurrent in-frame gene fusions were identified (*USP7-CARHSP1* and *KIF1B-PGD*), both at a frequency of 1.6% (2/126 primary tumors; Supplementary Note).

Somatic alterations were enriched in specific cancer-associated genes among the survival groups (Extended Data Fig. 2). Given the limited number of independent HGSC whole genomes, we used more readily available gene expression information^{7,30} to validate findings. Among tumor suppressor genes frequently inactivated in long-term survivors, low mRNA expression of *ARID1B*, *RBI* and *NFI* was associated with longer OS (Supplementary Table 8; $P < 0.05$). *ARID1A* and *ARID1B*, two cancer-associated genes involved in switch/sucrose non-fermentable (SWI/SNF) signaling, were both commonly disrupted by structural variants and had a combined somatic alteration rate of 30% (18/60) in long-term survivors, compared to 15.6% (5/32) and 17.6% (6/34) in moderate- and short-term survivors, respectively (Extended Data Fig. 2a). We also noted copy-number variants (CNVs) involving cytokines (for example, *CXCL9* and *IFNG*) occurred at different frequencies in the survival cohorts (Supplementary Note).

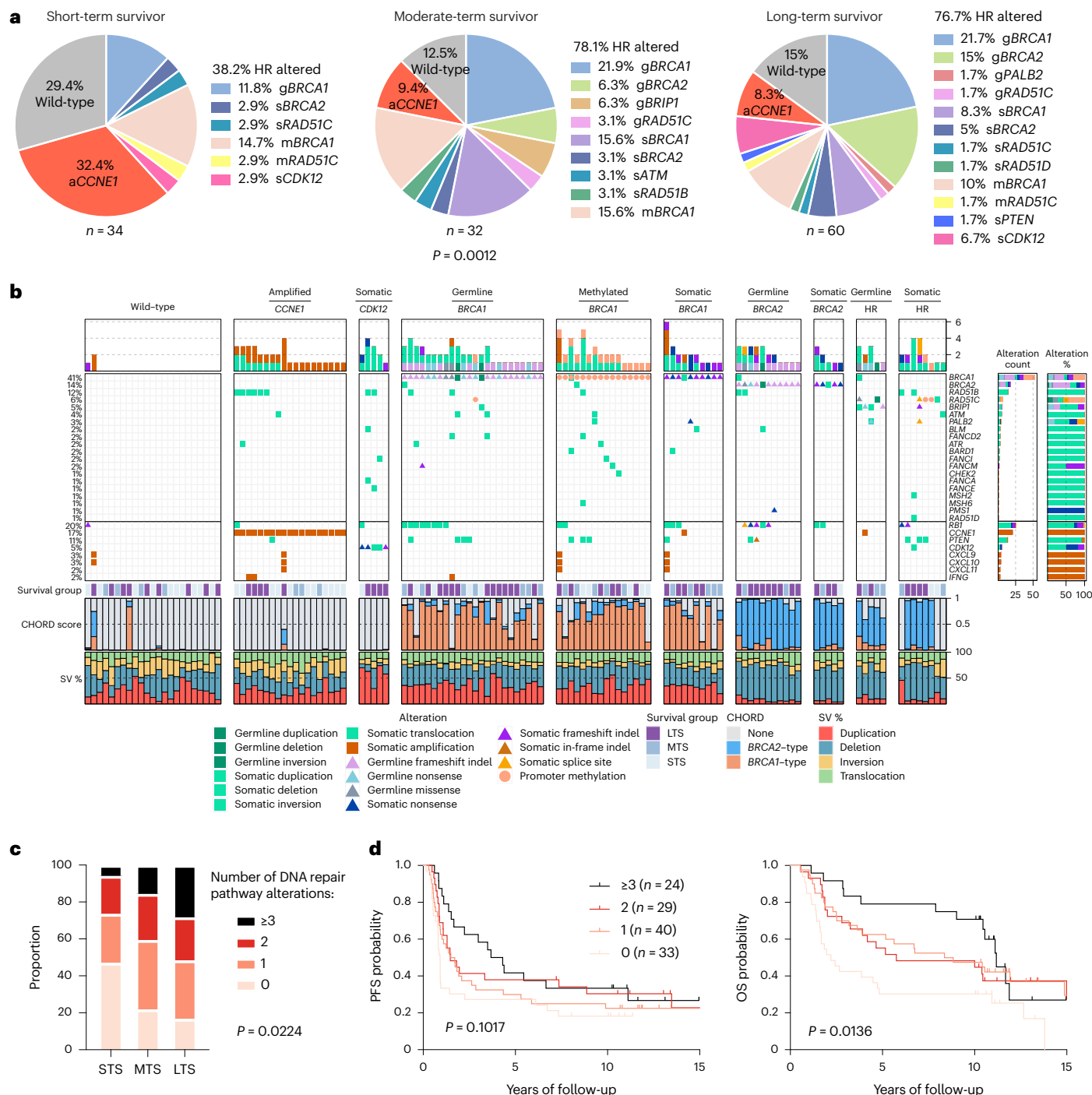


Fig. 1 | HGSCs with multiple altered DNA repair pathway genes are associated with long-term survival. **a**, Proportion of patients affected by homologous recombination (HR) DNA repair pathway gene alterations and *CCNE1* gene amplification (aCCNE1) in each survival group. Homologous recombination alterations include pathogenic germline (g) or somatic (s) mutations, and *BRCA1* or *RAD51C* promoter methylation (m) as indicated. One alteration is counted for patients with more than one change, prioritizing alterations by variant allele frequency and/or by evidence of genomic scarring associated with the candidate driver alteration. Differences in proportions of homologous recombination-altered, *CCNE1* amplified and wild-type tumors between survival groups were assessed by chi-square. **b**, Bars at the top represent the number of alterations in each listed gene per patient. Pathogenic germline and somatic alterations in homologous recombination pathway genes are shown, as well as alterations in other DNA repair associated genes, immune genes and *CCNE1*. Each patient (column) is annotated with survival group (LTS, long-term survivor; MTS,

moderate-term survivor; STS, short-term survivor). Bars indicate the level of homologous recombination deficiency (HRD) in each primary tumor sample, measured as probabilities of *BRCA1*-type (orange) HRD, *BRCA2*-type (blue) HRD, or homologous recombination proficiency (none, gray), as predicted by CHORD²⁸. Bar plots at the bottom indicate the proportion of total detected structural variants (SV) classified as duplications, deletions, inversions or interchromosomal translocations. Samples are grouped by the primary gene alteration identified in each patient. Alteration count and proportion of alteration types per gene are shown as bar plots on the right. **c**, Proportion of patients with 0, 1, 2 or 3 more DNA repair pathway alterations by survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor). Differences in proportions between groups were assessed by chi-square. **d**, Kaplan–Meier analysis of progression-free survival (PFS) (left) and OS in patients (right) with 0, 1, 2 or 3 more DNA repair pathway alterations. P values calculated by Mantel–Cox log-rank test.

Disease recurrence in long-term survivors

Tumor collection during recurrence and long-term survival are both uncommon events, but we obtained samples from four patients at relapse (Fig. 2a). Tumor-specific somatic alterations indicated that in all patients, samples were consistent with recurrence of their primary tumor rather than a new malignancy (Fig. 2b,c and Supplementary Note). *BRCA1/2* reversion mutations commonly impart acquired treatment resistance³¹ but were not detected in the relapse samples of three patients with *BRCA1* mutations.

Patient MWMH00552 experienced 13.5 years of disease-free remission before progressing rapidly and dying 18 months later from progressive disease. Remarkably, although a large deletion over *RBI* was found in the primary tumor, the emergent clone at recurrence lacked this deletion and the patient experienced relatively short duration responses to subsequent chemotherapy (Fig. 2c,d). Patient MAOC00944 had two different *RBI* deletions in their primary and relapse sample (Fig. 2d) and was still alive at last follow-up (>10 years) despite brain metastases. These findings support our observations here and previously²³ that co-occurrence of *RBI* and *BRCA1* mutations are associated with favorable response and outcome.

Patient MWMH00758 experienced cycles of recurrence and remission following an initial progression-free interval of >6 years. The primary tumor had amplification of chemokines (*CXCL9*, *CXCL10* and *CXCL11*) and was classified as the C2/immunoreactive molecular subtype¹¹ (Fig. 2c), and this classification was maintained at first and second relapse, consistent with the notion that the favorable immune subtype imparted a degree of tumor control over a lengthy period. Patient MAOC01893 had a solitary recurrence removed 3 years after diagnosis and then remained progression-free and was alive at last follow-up, 15 years after diagnosis (Fig. 2a). Of these four cases, the genomic alterations in the recurrent tumor in MAOC01893 was most similar to the primary tumor (Fig. 2b). However, although the primary tumor showed amplification of *CXCL9*, *CXCL10* and *CXCL11* and was C2/immunoreactive molecular subtype¹¹, consistent with good outcome, the recurrence was C1/mesenchymal molecular subtype and lacked amplification of the chemokine genes (Fig. 2c). This finding suggests that the recurrence represented a treatment-resistant clone with features associated with worse outcome and the patient may have experienced a substantial clinical benefit from surgical removal.

Mutational signatures across survival groups

To identify genomic variation patterns that define survival subgroups, we evaluated the contribution of previously identified genome-wide mutational signatures, including base substitution signatures³², indel signatures³² and ovary-specific rearrangement signatures³³ (that is, Ovary_A to Ovary_G; Methods and Supplementary Tables 9 and 10). Based on the most prominent 27 signatures (mean relative exposure >0.04 across all 126 samples), we performed unsupervised clustering of primary tumor samples. Samples segregated into seven clusters (SIG.1–7; Fig. 3) with distinct molecular phenotypes (summarized in Extended Data Fig. 3a), with SIG.1, SIG.4, SIG.6 and SIG.7 associated with longer survival (progression-free survival, $P = 0.0044$; OS, $P < 0.0001$; Extended Data Fig. 3b,c).

Clusters SIG.1 ($n = 14$), SIG.2 ($n = 25$) and SIG.3 ($n = 13$) were characterized by the tandem duplication (>100 kb) phenotype^{34,35} associated with loss-of-function *CDK12* mutations and *CCNE1* amplification ($P_{\text{adj}} = 0.0004$; Extended Data Figs. 3d and 4a). Tumors in these groups were homologous recombination proficient ($P_{\text{adj}} < 0.0001$), and patients were older at diagnosis ($P_{\text{adj}} = 0.0020$; Extended Data Figs. 4b and 5a). Despite having homologous recombination proficient tumors with a high frequency of *CCNE1* amplification (43%, 6/14), features typically associated with poor outcomes, cluster SIG.1 mostly comprised of long-term survivors (64%, 9/14; median OS not reached). Tumors in cluster SIG.1 had a high number of duplications ($P_{\text{adj}} < 0.0001$) and inversions ($P_{\text{adj}} = 0.0029$), a higher mutation

burden and neoantigen count ($P_{\text{adj}} < 0.0001$), somatic alterations in *RAD51B* (50%, 7/14; $P_{\text{adj}} = 0.0276$) and enrichment of indel signatures 1 and 2 ($P_{\text{adj}} < 0.0001$; Extended Data Fig. 5b), both thought to be caused by slippage during DNA replication and associated with defective DNA mismatch repair³². By contrast, SIG.2 and SIG.3 tumors had fewer mutations and structural variants, the lowest predicted neoantigen burdens and a lack of DNA repair gene alterations and were associated with the shortest survival (median OS, 1.7 and 2.4 years, respectively). SIG.3 tumors were also enriched for rearrangement signature Ovary_D (unknown driver; $P_{\text{adj}} = 0.0002$) and double-base substitution (DBS) signature 11 (unknown etiology; $P_{\text{adj}} < 0.0001$), whereas SIG.2 tumors were enriched for single-base substitution (SBS) signature 5 (unknown etiology; $P_{\text{adj}} < 0.0001$), indel signature 4 (unknown etiology; $P_{\text{adj}} < 0.0001$) and DBS signature 7 ($P_{\text{adj}} < 0.0001$), thought to be associated with defective DNA mismatch repair.

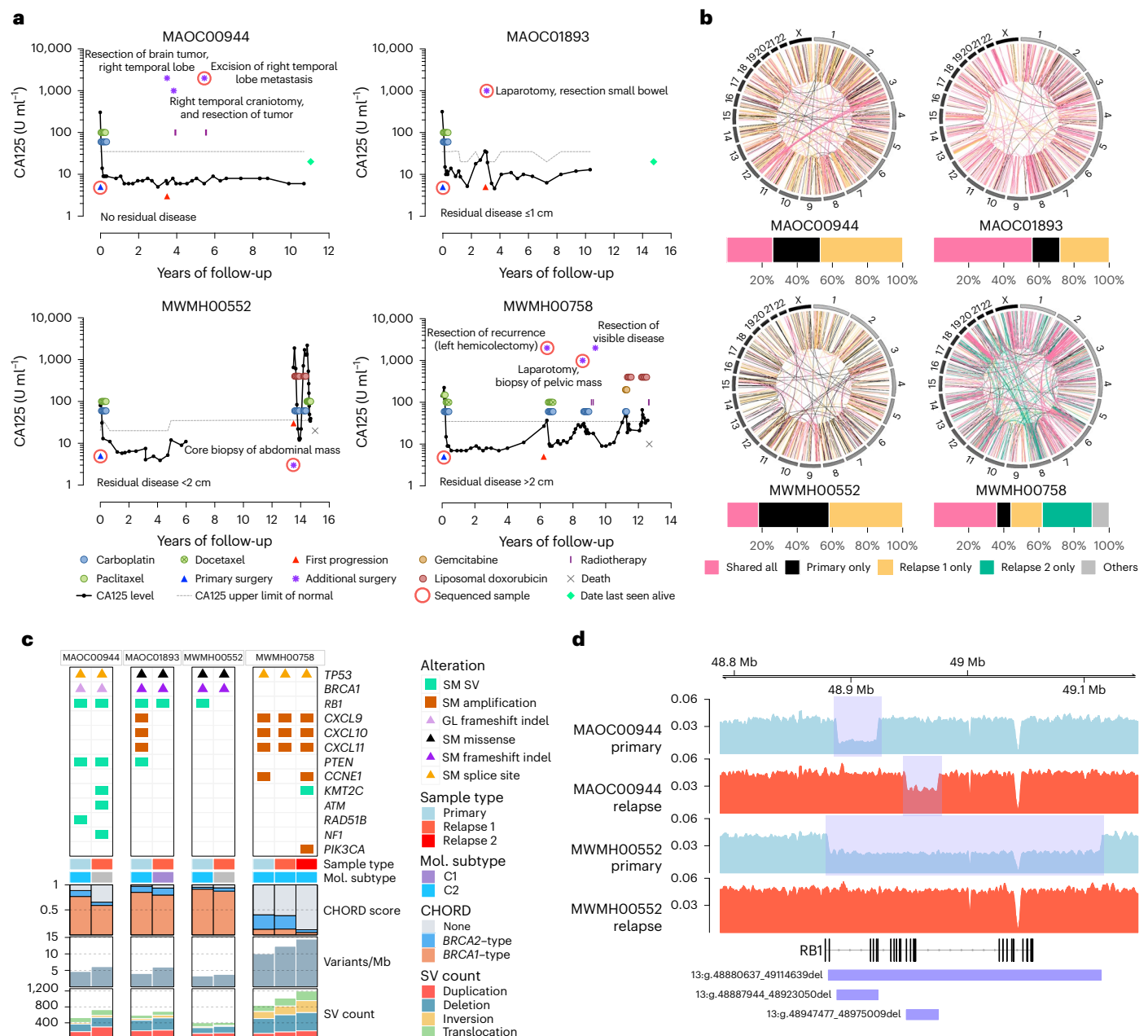
Cluster SIG.4 tumors ($n = 27$) were highly enriched for *BRCA2*-type nonclustered 1- to 100-kb deletions (consistent with rearrangement signature Ovary_A; $P_{\text{adj}} < 0.0001$), DBS signature 4 (unknown etiology; $P_{\text{adj}} < 0.0001$), indel signature 6 (associated with HRD; $P_{\text{adj}} < 0.0001$) and alterations in *BRCA2* ($P_{\text{adj}} < 0.0001$), as well as multiple other DNA repair pathway genes ($P_{\text{adj}} < 0.0001$; Fig. 3 and Extended Data Figs. 3–5). Cluster SIG.4 genomes had the highest median mutation burden and neoantigen count and almost all (25/27, 93%) had a high *BRCA2*-type CHORD score ($P_{\text{adj}} < 0.0001$). Consistent with the previously identified *BRCA2*/deletion HGSC subgroup³⁶, cluster SIG.4 had the highest survival rate (21/27 (78%) long-term survivors), with a median OS of 11.9 years.

BRCA1-altered tumor subgroups with differential outcomes

Clusters SIG.5 ($n = 22$), SIG.6 ($n = 9$) and SIG.7 ($n = 16$) were characterized by *BRCA1* alterations (100%, 100% and 94%, respectively, $P_{\text{adj}} < 0.0001$), but the three groups had distinctly different survival outcomes (Extended Data Figs. 3 and 4). All showed nonclustered 1- to 100-kb tandem duplications, enrichment of rearrangement signature Ovary_G, and *BRCA1*-type HRD scores ($P_{\text{adj}} < 0.0001$; Fig. 3 and Extended Data Figs. 3–5), consistent with their *BRCA1* mutational status. Of the *BRCA1* groups, cluster SIG.7 had the highest proportion of long-term survivors (75%, 12/16; median OS, 10.4 years), followed by SIG.6 (56%, 5/9; median OS, not reached) and SIG.5 (27%, 6/22; median OS, 4.5 years).

In a subset analysis considering only the *BRCA1*-altered clusters, the key mutational signatures driving these clusters were DBS signature 2 in SIG.5 ($P_{\text{adj}} = 0.0050$), rearrangement signature Ovary_A in SIG.6 ($P_{\text{adj}} = 0.0092$) and SBS signature 40 (unknown etiology) in SIG.7 ($P_{\text{adj}} = 0.0092$; Supplementary Note). DBS signature 2 is proposed to be associated with tobacco smoking and/or exposure to acetaldehyde, which is a constituent of cigarette smoke but also a metabolite of alcohol³². Self-reported smoking history was available for 84.9% (107/126) of cases, and across the seven mutational signature clusters, SIG.5 had the highest frequency of smokers (66.7%, 12/18; $P_{\text{adj}} = 0.5092$; Extended Data Fig. 4b). We compared the relative contribution of all mutational signatures between never-smokers ($n = 60$) and ever-smokers ($n = 47$), and the most predominant mutational signature in smokers was DBS signature 2 ($P_{\text{adj}} = 0.5490$; Supplementary Note). Of all the mutational signature clusters, SIG.5 had the youngest age of diagnosis ($P_{\text{adj}} = 0.0020$; Extended Data Fig. 5a), consistent with these patients being at a higher risk of developing cancer due to combined *BRCA1* deficiency and a history of smoking.

The prominence of rearrangement signature Ovary_A in cluster SIG.6 indicates there is a mixture of *BRCA1* and *BRCA2* deficiency in this subgroup; this was corroborated by a higher prevalence of *BRCA2*-type nonclustered 1- to 100-kb deletions in SIG.6 relative to SIG.5 and SIG.7 (Extended Data Fig. 3d) and the detection of both *BRCA1*-type and *BRCA2*-type HRD scores in SIG.6 tumors (Fig. 3). Tumors with combined *BRCA1* and *BRCA2* loss of function may have greater sensitivity to platinum chemotherapy. Indeed, despite cluster SIG.6 having a high



samples, structural variants that were shared only by two tumor samples were classified as “others”. **c**, Somatic (SM) and germline (GL) alterations in primary and relapse tumor samples in genes of interest (rows). Each sample (column) is annotated with sample type (primary or relapse) and molecular (Mol.) subtype¹¹ if RNA was available. Bars indicate the level of HRD in each tumor sample, measured as probabilities of *BRCA1*-type (orange) HRD, *BRCA2*-type (blue) HRD, or homologous recombination proficient (none, gray), as predicted by CHORD²⁸. Bar plots at the bottom indicate somatic mutation burden in variants per megabase (Mb) and SV counts in each sample. **d**, BigWig tracks of DNA sequencing coverage (y axis) in two paired primary (shaded blue) and relapse (shaded red) tumor samples showing the locations (x axis) of deletions (blue rectangles) identified in *RB1*.

proportion (56%, 5/9) of suboptimal residual disease (>1 cm) following surgical cytoreduction ($P_{\text{adj}} = 0.6131$), this *BRCA1* subgroup had the longest progression-free survival (median 9.9 years, $P = 0.0044$), indicating SIG.6 tumors were particularly platinum chemosensitive (Extended Data Figs. 3 and 4).

Patterns of DNA methylation

To determine whether tumor DNA methylation profiles were associated with exceptional outcomes, we performed consensus clustering of the 1% most variable CpG sites (number of probes = 3,645) across all 126 primary tumors. Compared to mutational signatures, differential DNA

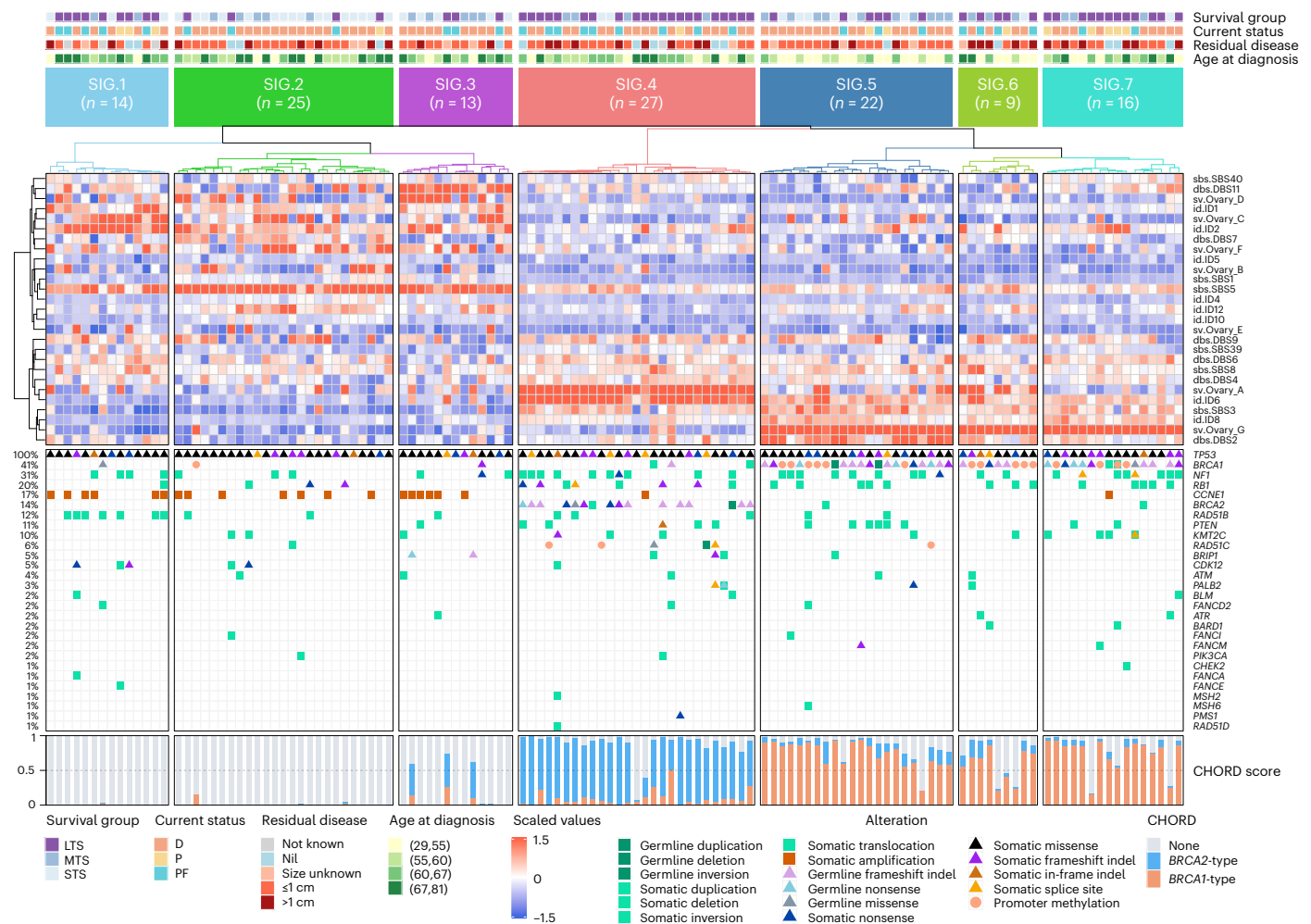


Fig. 3 | Long-term survivor tumor genomes are characterized by distinct molecular phenotypes. Heatmap of mutational signatures following consensus clustering based on proportions of mutational signature exposures in each primary tumor sample ($n = 126$). Patient (column) scaled z-scores are shown in the heatmap. Seven mutational signature clusters were identified (SIG.1–SIG.7), and each patient (column) is annotated with survival group (LTS, long-term

survivor; MTS, moderate-term survivor; STS, short-term survivor), status at last follow-up (D, dead; P, progressed and alive; PF, progression-free and alive), residual disease and age at diagnosis (in years; quartiles). Bars represent the *BRCA1*-type (orange) and *BRCA2*-type (blue) HRD (CHORD²⁵) scores of each tumor sample, and germline and somatic alterations affecting genes of interest are shown in the bottom panel.

methylation patterns were less discriminatory, with the five distinct methylation clusters (MET.1, $n = 46$; MET.2, $n = 14$; MET.3, $n = 17$; MET.4, $n = 19$; MET.5, $n = 30$) showing moderate to weak associations with progression-free ($P = 0.1949$) and OS ($P = 0.0587$; Extended Data Fig. 6). The strongest genomic difference between methylation clusters was *BRCA1* alteration status, with enrichment of *BRCA1*-altered tumors (72%, 33/46) in cluster MET.1 ($P_{\text{adj}} < 0.0001$; Supplementary Note). Patients in MET.1 had a relatively poor survival (median OS, 5.7 years) compared to MET.2 (median OS, 11.9 years), the most closely related cluster in the dendrogram. The most striking differences between the two groups were the proportion of smokers (MET.1 64% (23/36) vs. MET.2 8.3% (1/12), $P_{\text{adj}} = 0.0225$) and younger age of MET.1 patients ($P_{\text{adj}} = 0.0059$). Therefore, the MET.1 cluster also identifies a subset of *BRCA1*-altered tumors associated with smoking and a relatively poor survival, largely overlapping with cluster SIG.5 (Supplementary Note).

Tumor mutation burden and immune transcriptional patterns

Ovarian cancer was among the first documented examples of an association between lymphocytic infiltration and survival³⁷, an observation confirmed in large patient cohorts^{13,38} and individual case reports³⁹. We therefore characterized mutational burden as a driver of immune

response within the survival groups (Fig. 4 and Supplementary Note). Consistent with a previous report⁴⁰, long-term survivor tumor samples had a higher tumor mutation burden (median of 4.66 mutations/Mb) compared to short-term (3.27 mutations/Mb) and moderate-term survivors (3.25 mutations/Mb, $P = 0.0003$), and concordantly, long-term survivors had the highest number of predicted neoantigens ($P < 0.0001$; Fig. 4e). Both moderate- and long-term survivor tumors had more structural variants compared to short-term survivors ($P = 0.0012$; Fig. 4e). Tumor neoantigen count was more strongly associated with better survival (hazard ratio (HR) = 0.71, 95% confidence interval (CI) = 0.56–0.91, $P = 0.0069$) compared with the number of mutations (HR: 0.75, 95% CI: 0.59–0.96, $P = 0.0202$) and structural variants (HR: 0.79, 95% CI: 0.63–1.0, $P = 0.0482$; Fig. 4f).

We compared primary tumor gene expression profiles between survival groups using fast gene set enrichment analysis (FGSEA⁴¹) and observed significantly perturbed MSigDB hallmark gene sets⁴² (FGSEA $P_{\text{adj}} < 0.05$; Extended Data Fig. 7a). The top five enriched gene sets between long-term survivors and short-term survivors were E2F targets (overexpressed or 'up' in long- versus short-term survivors), epithelial mesenchymal transition (down), allograft rejection (up), interferon gamma response (up) and G₂M checkpoint (up; FGSEA $P_{\text{adj}} < 0.0001$;

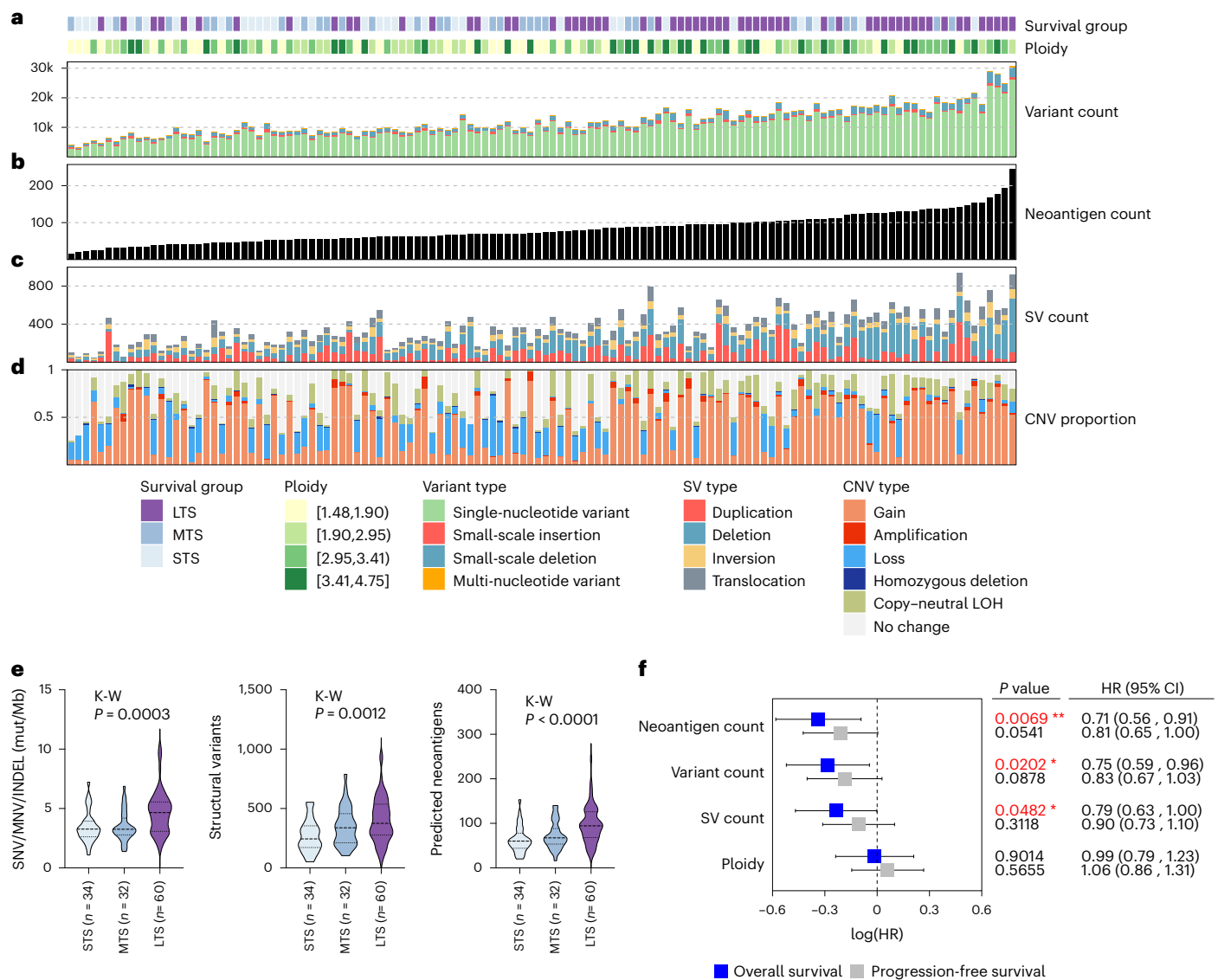


Fig. 4 | Elevated somatic mutation burden in long-term survivors. **a**, Bars indicate the total variant count of each primary tumor sample ($n = 126$), including SNVs, small-scale insertions and deletions (INDELs) and multinucleotide variants (MNVs). **b**, Tumor samples are ordered left to right from fewest to largest number of neoantigens (black bars). **c**, Bars indicate total number of large-scale structural variants (SVs) in primary tumors, including duplications, deletions, inversions and interchromosomal translocations. **d**, Bars indicate the proportion of each tumor genome affected by CNVs, including regions of gain, amplification, loss, homozygous deletion and copy-neutral loss of heterozygosity (LOH). **e**, Violin plots represent the tumor mutation burden,

structural variant count and predicted neoantigen counts of tumor genomes in each survival group. Dashed lines represent the median and dotted lines represent the lower and upper quartiles. Kruskal–Wallis (K-W) test P values are reported. STS, short-term survivor; MTS, moderate-term survivor; LTS, long-term survivor; mut/Mb, mutations per megabase. **f**, Forest plot indicates the HR (squares) and 95% CI (whiskers) for progression-free and OS ($n = 126$ patients) calculated using a univariate Cox proportional hazard regression model based on genomic features as indicated; P values < 0.05 are colored red (* $P < 0.05$, ** $P < 0.01$) and were not adjusted for multiple comparisons.

Extended Data Fig. 7a; Supplementary Data 4). Tumors in long-term survivors had an increased expression of cell proliferation-related genes *PCNA* and *MKI67* (Extended Data Fig. 7b), indicating that tumor cells in long-term survivors may have exceptionally deregulated cell cycle progression and increased proliferation. This is consistent with our previous finding that Ki-67-positivity of tumor cell nuclei, a marker of proliferation, is significantly higher in patients with prolonged progression-free survival and OS²³.

We used an established deconvolution method²⁶ to estimate the abundance of immune cell types from bulk RNA-seq data (Supplementary Tables 11, 12). CIBERSORTx absolute scores correlated with CD8⁺ T cell density, both in the tumor epithelium ($P < 0.0001$) and the

stromal compartment ($P = 0.0002$), which were previously quantified by immunohistochemistry²³ in a subset of primary tumors ($n = 54$; Supplementary Note). Scores of the most prominent cell populations (detected in $>5\%$ of samples) were treated as a continuous variable and correlated with survival. Activated memory CD4 T cells (HR: 0.44, 95% CI: 0.23–0.85, $P = 0.0144$) and plasma cells (HR: 0.68, 95% CI: 0.49–0.95, $P = 0.0249$) were associated with improved OS, whereas resting mast cells (HR: 1.44, 95% CI: 1.15–1.78, $P = 0.0013$) and M2 macrophages (HR: 1.46, 95% CI: 1.09–1.96, $P = 0.0119$) were associated with an increased risk of death (Extended Data Fig. 7c).

Unsupervised clustering of primary tumor samples using the computationally estimated immune cell densities identified five patient

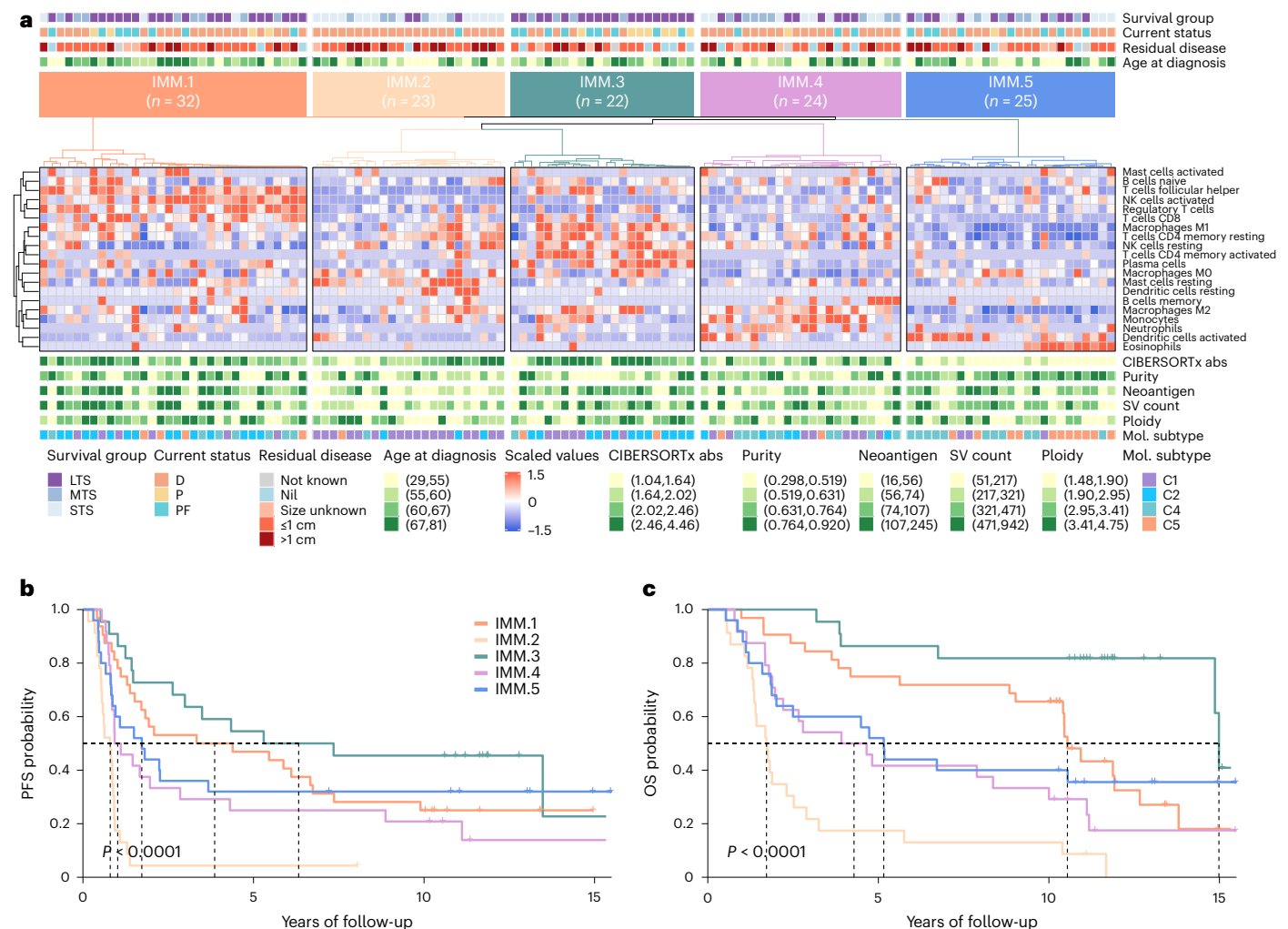


Fig. 5 | Immune phenotypes of long-term survivors. a, Heatmap of scaled abundance of immune cell types following consensus clustering based on CIBERSORTx²⁶ estimated absolute abundance of immune cell types from bulk RNA-seq data of each primary tumor ($n = 126$). Five immune clusters were identified (IMM.1–IMM.5), and each patient (column) is annotated with survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor), status at last follow-up (D, dead; P, progressed and alive; PF, progression-free and alive), residual disease size and age at diagnosis (quartiles).

CIBERSORTx absolute (abs) immune scores, tumor purity, neoantigen counts, structural variant (SV) counts and ploidy estimates are shown as quartiles. Tumor samples are also classified according to molecular subtype¹¹ (C1, mesenchymal; C2, immunoreactive; C4, differentiated; C5, proliferative). **b, c**, Kaplan–Meier analysis of progression-free survival (b) and OS (c) in patients stratified by immune clusters. P values were calculated by Mantel–Cox log-rank test, and dotted lines indicate median survival.

groups associated with differential survival outcomes (IMM.1–5; Fig. 5). Patients in cluster IMM.3 ($n = 22$) had the longest progression-free survival (median 6.3 years, $P < 0.0001$) and OS (median 15.0 years, $P < 0.0001$), with tumor samples enriched for plasma cells, activated memory CD4 T cells, M1 macrophages and resting natural killer (NK) cells ($P_{\text{adj}} < 0.0001$; Extended Data Fig. 8a). Cluster IMM.1 patients ($n = 32$) had the second longest OS (median 10.5 years) and were particularly enriched for CD8 T cells, activated NK cells, regulatory T cells and follicular helper T cells ($P_{\text{adj}} < 0.0001$). Tumor genomes in clusters IMM.1 and IMM.3 had the highest and second highest neoantigen burden respectively ($P_{\text{adj}} = 0.0007$; Extended Data Fig. 8b). By contrast, samples from cluster IMM.2 ($n = 23$) have the lowest neoantigen burden, the shortest OS (median 1.7 years), and were enriched for resting mast cells and dendritic cells ($P_{\text{adj}} < 0.05$). Concordantly, samples in cluster IMM.2 were predominantly classified as the C1/mesenchymal molecular subtype (78.3%, 18/23), whereas samples in clusters IMM.1 and IMM.3 were predominantly the C2/immunoreactive molecular subtype (46.9%, 15/32) and 54.5% (12/22) respectively; $P_{\text{adj}} < 0.0001$;

Extended Data Fig. 9a). Consistent with having an active immune response, IMM.1 and IMM.3 tumors had higher densities of CD8⁺ T cells in the tumor epithelium compared to other clusters ($P_{\text{adj}} = 0.0281$; Extended Data Fig. 8b).

Although the two long-term survival immune clusters IMM.1 and IMM.3 were characterized by elevated HRD scores (scarHRD mean, $P_{\text{adj}} = 0.0439$; BRCA2-type CHORD, $P_{\text{adj}} = 0.0859$; Extended Data Fig. 8b), no particular DNA repair gene alteration was associated with the immune clusters (Extended Data Fig. 9b). In a subgroup analysis, differences in immune cell composition were observed between the BRCA1-altered mutational signature clusters: the most notable being elevated expression of the activated NK cell signature in clusters SIG.6 and SIG.7 compared to SIG.5 (SIG.5 versus SIG.6 $P_{\text{adj}} = 0.0760$, SIG.5 versus SIG.7 $P_{\text{adj}} = 0.0580$; Supplementary Note). Concordantly, immune cluster IMM.1, which is enriched with the activated NK cell signature, was the dominant immune cluster in both SIG.6 (44.4%, 4/9) and SIG.7 (43.8%, 7/16) and the least abundant immune cluster in SIG.5 tumors (13.6%, 3/22; $P = 0.4688$; Fig. 6a and Supplementary Note).

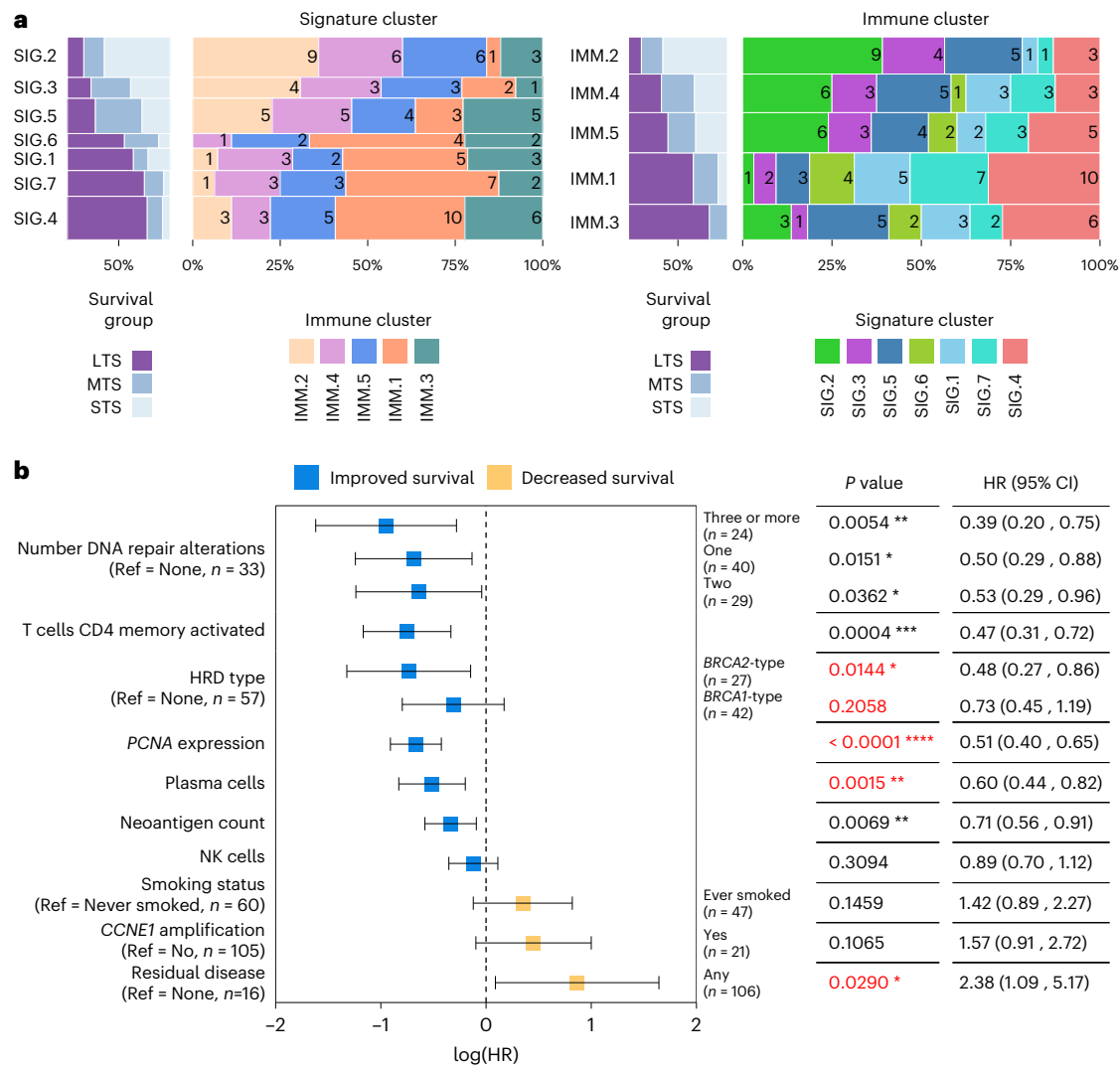


Fig. 6 | Key features associated with exceptional survival in HGSC. a, Spine plots showing the proportion of samples shared between the mutational signature clusters (left) and the immune clusters (right), with the number of overlapping samples indicated inside the colored bars. Also shown is the proportion of survival groups (STS, short-term survivor; MTS, moderate-term survivor; LTS, long-term survivor) in each cluster. Each of the two types of clusters is ordered horizontally and vertically by the overall proportion of LTS. The height of the bars indicates the number of samples in the cluster. **b**, Forest plot illustrates the HR (squares) and 95% CI (whiskers) for OS calculated using

a univariate Cox proportional hazard regression model based on selected features; results were not adjusted for multiple comparisons (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). Features are sorted top to bottom by smallest to largest HR and P values less than 0.05 in multivariable model are colored red. Complete univariable and multivariable results, including feature associations with progression-free survival, are provided in Supplementary Tables 13, 14. Ref indicates the reference used for categorical features, and n indicates the number of samples in the categorical group.

Predictors of long-term survival

We considered the key features identified in this study, finding seven were individually associated with OS (Fig. 6b and Supplementary Table 13; univariable Cox regression model), including the number of DNA repair gene alterations (three or more HR: 0.39, 95% CI: 0.20–0.75, $P = 0.0054$), activated CD4 memory T cells (HR: 0.47, 95% CI: 0.31–0.72, $P = 0.0004$), *BRCA2*-type HRD (HR: 0.48, 95% CI: 0.27–0.86, $P = 0.0144$), *PCNA* expression (HR: 0.51, 95% CI: 0.40–0.65, $P < 0.0001$), plasma cells (HR: 0.60, 95% CI: 0.44–0.82, $P = 0.0015$), neoantigen count (HR: 0.71, 95% CI: 0.56–0.91, $P = 0.0069$) and residual disease (HR: 2.38, 95% CI: 1.09–5.17, $P = 0.0290$). When combined in a multivariable regression model, four features were statistically associated with OS, including HRD type (*BRCA2*-type HR: 0.33, 95% CI: 0.17–0.66, $P = 0.0018$; *BRCA1*-type HR: 0.45, 95% CI: 0.25–0.82, $P = 0.0086$), *PCNA* expression (HR: 0.50, 95% CI: 0.38–0.67, $P < 0.0001$), plasma cells (HR: 0.54, 95% CI:

0.37–0.78, $P = 0.0011$) and residual disease (HR: 3.15, 95% CI: 1.37–7.21, $P = 0.0067$; Supplementary Table 14).

Discussion

Cancer studies have focused on the determinants of treatment failure (that is, primary and acquired drug therapy resistance), with comparatively little attention on those patients who exceed expectations, despite their potential to provide therapeutic insights^{23,43,44}. By accessing samples and data from studies that collectively include over 3,800 patients with HGSC, we were able to perform whole-genome characterization of HGSC in 60 exceptional survivors.

Conceivably, exceptional survival in HGSC may be determined by a dominant rare event or by the interaction of multiple factors that are individually common but due to chance are infrequent in combination. Our finding of the association of survival with a variety of factors

involving the patient's genome, tumor somatic mutational profile, and immune response strongly supports the latter explanation and is consistent with a diversity of molecular and clinical pathways to long-term survival in HGSC^{23,40}.

Synthetic lethality induced by PARPi on a *BRCA* altered background provides a potent example of the clinical effect of simultaneously targeting DNA repair processes⁴⁵. In a genetic parallel, we found that co-occurring alterations in DNA repair pathway genes are associated with long-term survival. In some instances, this was associated with evidence of both *BRCA1*-type and *BRCA2*-type HRD within a tumor, often due to structural variants that inactivate homologous recombination pathway genes. Indeed, inactivation of homologous recombination pathway genes by structural variants is a common and perhaps unappreciated source of HRD in HGSC^{14,46,47}. It is plausible that multiple genetic defects in DNA repair may render tumor cells exceptionally sensitive to chemotherapy, as recently reported in an exceptional responder with metastatic breast cancer⁴⁴, or perhaps impede the development of resistance.

We previously reported²³, and validated here, how co-occurrence of *RBI* and *BRCA1* or *BRCA2* loss-of-function mutations is associated with long-term survival. Interestingly, *RBI* has a non-canonical function in homologous recombination DNA repair⁴⁸ in addition to cell cycle regulation. Our comparison of primary and relapse tumor samples in a small number of patients provided additional evidence for a key role of co-occurring *RBI* and *BRCA1* mutations in exceptional response and good outcome.

One of the strongest associations with long-term survival were markers of enhanced proliferation. While enhanced proliferation is generally associated with aggressive cancer phenotypes, faster replication may also render cells more susceptible to chemotherapy. Higher proliferative counts in long-term survivors could also relate to a reduced ability of the cells to enter a quiescent state, which has been associated with the development of treatment resistance in lung⁴⁹ and ovarian cancer⁵⁰. A subset of long-term survivors had *CCNE1* amplification and evidence of enhanced immune activity, suggesting that an engaged tumor-immune microenvironment can overcome the poor primary treatment response typically associated with *CCNE1* amplification and homologous recombination proficiency.

The determinants of long-term survival in HGSC are complex, and progress will depend on detailed discovery studies⁴³ and validation of specific findings in large, clinically annotated patient cohorts with long follow-up⁵¹. Refining comparisons⁴³ will also be key, as exemplified by the three subsets of patients, all with *BRCA1* mutations but with distinctly different survival outcomes. Clusters SIG.5 and MET.1 identify a subgroup of more aggressive *BRCA1*-driven tumors associated with a younger age of onset, whose unique genomic signature associated with tobacco or alcohol exposure, relatively lower NK cell infiltration and/or less frequent compounded HRD may drive a diminution in survival. A recent study found that cigarette smoking is associated with worse survival among women with germline *BRCA1/2* mutations compared to noncarriers⁵². Furthermore, a large cohort study of asymptomatic individuals found that NK cell activity decreases in smokers in a dose-dependent manner⁵³, indicating a plausible link between smoking-associated NK cell deficit and an elevated risk of malignancy, particularly in *BRCA1* mutation carriers. Collectively our molecular data support the observation that survival outcomes in women with *BRCA1*-altered HGSC may be influenced by prior mutagen exposure, a potentially modifiable risk factor.

We have identified distinct HGSC subgroups separated by mutational processes, DNA methylation and immune response, and found that differential outcomes may be associated with compounding lifestyle-related exposures, surgical outcomes, anti-tumor immune activity, cell cycle deregulation and/or disruption of multiple DNA repair pathways. Although most of our patients predate the introduction of PARPi, given that response to platinum is predictive of PARPi

sensitivity¹⁹, our findings may also provide insights into long-term PARPi response.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-022-01230-9>.

References

1. Millstein, J. et al. Prognostic gene expression signature for high-grade serous ovarian cancer. *Ann. Oncol.* **31**, 1240–1250 (2020).
2. Hoppenot, C., Eckert, M. A., Tienda, S. M. & Lengyel, E. Who are the long-term survivors of high grade serous ovarian cancer? *Gynecol. Oncol.* **148**, 204–212 (2018).
3. Fagö-Olsen, C. L. et al. Does neoadjuvant chemotherapy impair long-term survival for ovarian cancer patients? A nationwide Danish study. *Gynecol. Oncol.* **132**, 292–298 (2014).
4. Chi, D. S. et al. What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecol. Oncol.* **103**, 559–564 (2006).
5. Horowitz, N. S. et al. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *J. Clin. Oncol.* **33**, 937–943 (2015).
6. Alsop, K. et al. *BRCA* mutation frequency and patterns of treatment response in *BRCA* mutation-positive women with ovarian cancer: A report from the Australian ovarian cancer study group. *J. Clin. Oncol.* **30**, 2654–2663 (2012).
7. The Cancer Genome Atlas Research Network. Integrated genomic analysis of ovarian cancer. *Nature* **474**, 609–615 (2011).
8. Walsh, T. et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108**, 18032–18037 (2011).
9. Ciriello, G. et al. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
10. Ahmed, A. A. et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221**, 49–56 (2010).
11. Tothill, R. W. et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
12. Etemadmoghadam, D. et al. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clin. Cancer Res.* **15**, 1417–1427 (2009).
13. Hwang, W. T., Adams, S. F., Tahirovic, E., Hagemann, I. S. & Coukos, G. Prognostic significance of tumor-infiltrating T cells in ovarian cancer: A meta-analysis. *Gynecol. Oncol.* **124**, 192–198 (2012).
14. Patch, A. M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
15. Wang, Y. K. et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **49**, 856–864 (2017).
16. Macintyre, G. et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).
17. Pennington, K. P. et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin. Cancer Res.* **20**, 764–775 (2014).

18. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
19. Fong, P. C. et al. Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J. Clin. Oncol.* **28**, 2512–2519 (2010).
20. Swisher, E. M. et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **18**, 75–87 (2017).
21. Bolton, K. L. et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA* **307**, 382–390 (2012).
22. Candido-dos-Reis, F. J. et al. Germline mutation in BRCA1 or BRCA2 and ten-year survival for women diagnosed with epithelial ovarian cancer. *Clin. Cancer Res.* **21**, 652–657 (2015).
23. Garsed, D. W. et al. Homologous recombination DNA repair pathway disruption and retinoblastoma protein loss are associated with exceptional survival in high-grade serous ovarian cancer. *Clin. Cancer Res.* **24**, 569–580 (2018).
24. Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22**, 398–406 (2012).
25. Etemadmoghadam, D. et al. Synthetic lethality between CCNE1 amplification and loss of BRCA1. *Proc. Natl Acad. Sci. USA* **110**, 19489–19494 (2013).
26. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
27. Miller, R. E. et al. ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Ann. Oncol.* **31**, 1606–1622 (2020).
28. Nguyen, L., W. M. Martens, J., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**, 1–12 (2020).
29. Joshi, P. M., Sutor, S. L., Huntoon, C. J. & Karnitz, L. M. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. *J. Biol. Chem.* **289**, 9247–9253 (2014).
30. Anaya, J. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *Peer J. Comp. Sci.* **2**, e67 (2016).
31. Norquist, B. et al. Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. *J. Clin. Oncol.* **29**, 3008–3015 (2011).
32. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
33. Degasperi, A. et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat. Cancer* **1**, 249–263 (2020).
34. Popova, T. et al. Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. *Cancer Res.* **76**, 1882–1891 (2016).
35. Wu, Y. M. et al. Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. *Cell* **173**, 1770–1782.e1714 (2018).
36. Funnell, T. et al. Integrated structural variation and point mutation signatures in cancer genomes using correlated topic models. *PLoS Comput. Biol.* **15**, 1–24 (2019).
37. Zhang, L. et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
38. Ovarian Tumor Tissue Analysis (OTTA) Consortium. Dose-response association of CD8⁺ tumor-infiltrating lymphocytes and survival time in high-grade serous ovarian cancer. *JAMA Oncol.* **3**, e173290 (2017).
39. Jiménez-Sánchez, A. et al. Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient. *Cell* **170**, 927–938.e920 (2017).
40. Yang, S. Y. C. et al. Landscape of genomic alterations in high-grade serous ovarian cancer from exceptional long- and short-term survivors. *Genome Med* **10**, 81 (2018).
41. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/060012> (2021).
42. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
43. Saner, F. A. M. et al. Going to extremes: determinants of extraordinary response and survival in patients with cancer. *Nat. Rev. Cancer* **19**, 339–348 (2019).
44. Wheeler, D. A. et al. Molecular features of cancers exhibiting exceptional responses to treatment. *Cancer Cell* **39**, 38–53.e37 (2021).
45. Moore, K. et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.* **379**, 2495–2505 (2018).
46. Ewing, A. et al. Structural Variants at the BRCA1/2 loci are a common source of homologous repair deficiency in high-grade serous ovarian carcinoma. *Clin. Cancer Res.* **27**, 3201–3214 (2021).
47. Swisher, E. M. et al. Characterization of patients with long-term responses to rucaparib treatment in recurrent ovarian cancer. *Gynecol. Oncol.* **163**, 490–497 (2021).
48. Velez-Cruz, R. et al. RB localizes to DNA double-strand breaks and promotes DNA end resection and homologous recombination through the recruitment of BRG1. *Genes Dev.* **30**, 2500–2512 (2016).
49. Fan, W. et al. MET-independent lung cancer cells evading EGFR kinase inhibitors are therapeutically susceptible to BH3 mimetic agents. *Cancer Res.* **71**, 4494–4505 (2011).
50. Cole, A. NFATC4 promotes quiescence and chemotherapy resistance in ovarian cancer. *JCI Insight* **5**, e131486 (2020).
51. Sieh, W. et al. Hormone-receptor expression and ovarian cancer survival: an Ovarian Tumor Tissue Analysis consortium study. *Lancet Oncol.* **14**, 853–862 (2013).
52. Gersekowski, K. et al. Germline BRCA variants, lifestyle and ovarian cancer survival. *Gynecol. Oncol.* **165**, 437–445 (2022).
53. Jung, Y. S. et al. Impact of smoking on human natural killer cell activity: A large cohort study. *J. Cancer Prev.* **25**, 13–20 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

¹Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ²Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, Australia. ³The Westmead Institute for Medical Research, Sydney, New South Wales, Australia. ⁴Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia. ⁵The University of Sydney, Sydney, New South Wales, Australia. ⁶Department of Obstetrics and Gynecology, The Jikei University School of Medicine, Tokyo, Japan. ⁷The Deeley Research Centre, BC Cancer, Victoria, British Columbia, Canada. ⁸Women's Health Integrated Research Center, Gynecologic Cancer Center of Excellence, Uniformed Services University and Walter Reed National Military Medical Center, Bethesda, MD, USA. ⁹Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA. ¹⁰The Royal Women's Hospital, Parkville, Victoria, Australia. ¹¹Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ¹²Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia. ¹³Sydney, New South Wales, Australia. ¹⁴Crown Princess Mary Cancer Centre, Westmead Hospital, Sydney, New South Wales, Australia. ¹⁵Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. ¹⁶Department of Obstetrics and Gynaecology, The University of Melbourne, Parkville, Victoria, Australia. ¹⁷Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. ¹⁸Department of Obstetrics and Gynecology, Bern University Hospital and University of Bern, Bern, Switzerland. ¹⁹Division of Computational Biology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. ²⁰Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²¹Department of Oncology, University of Cambridge, Cambridge, UK. ²²Women's Health Integrated Research Center, Women's Service Line, Inova Health System, Falls Church, VA, USA. ²³School of Clinical Medicine, Faculty of Medicine and Health, University of NSW, Sydney, New South Wales, Australia. ²⁴Adult Cancer Program, Lowy Cancer Research Centre, University of NSW, Sydney, New South Wales, Australia. ²⁵Department of Epidemiology and Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA. ²⁶Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ²⁷Department of Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada. ²⁸Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada. ²⁹Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. ³⁰The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW, Sydney, New South Wales, Australia. ³¹These authors contributed equally: Anna DeFazio, David D. L. Bowtell. ✉e-mail: dale.garsed@petermac.org; david.bowtell@petermac.org

Methods

Study participants and patient samples

This project was conducted with approval from the Peter MacCallum Cancer Centre Human Research Ethics Committee, the Western Sydney Local Health District Human Research Ethics Committee and the Mayo Clinic Institutional Review Board. The study population consisted of women diagnosed with epithelial ovarian cancer between 1980 and 2019, enrolled in the Australian Ovarian Cancer Study (AOCS), the Gynaecological Oncology Biobank at Westmead Hospital (Sydney) and the Mayo Clinic. Participation in these studies was voluntary (patients were not compensated), and written informed consent was provided by all participants. Women with histologically confirmed high-grade serous ovarian carcinoma and survival time available ($n = 3,824$) were considered for the study.

Inclusion criteria. Cases were selected as follows: (i) histologically confirmed high-grade (grade 2 or 3) serous ovarian, fallopian or peritoneal carcinoma; (ii) International Federation of Gynecology and Obstetrics stage IIIc or IV disease; (iii) primary treatment incorporating a platinum-based agent; (iv) fresh-frozen tumor obtained during primary debulking surgery and matched blood samples available or previously analyzed¹⁴. Survival categories were defined as follows: (i) short-term survivors had died less than 2 years from diagnosis, (ii) moderate-term survivors had survived at least 2 years since diagnosis but died before 10 years and (iii) long-term survivors had an OS of at least 10 years after diagnosis (Extended Data Fig. 1a). This definition of long-term survival is consistent with previous studies^{23,54}. To confirm high-grade serous carcinoma, all eligible cases underwent pathology review as previously described²³.

Clinical definitions. Progression-free survival was defined as the time between histological diagnosis and disease progression, as determined by imaging or CA125 serum levels according to the Gynecological Cancer Intergroup criteria, or death. OS was defined as the time interval between histological diagnosis and death (all causes) or date of last follow-up. Never-smokers were those participants who had self-reported never smoking (or having smoked less than 100 cigarettes in their lifetime) before diagnosis.

Cohorts. Sequencing of 73 patients was previously described¹⁴ (7 long-term survivors, 34 short-term survivors and 32 moderate-term survivors) as part of the International Cancer Genome Consortium (ICGC) Ovarian Cancer project. Additional sequencing of samples from 53 long-term survivors was performed here as part of the Multidisciplinary Ovarian Cancer Outcomes Group (MOCOG) study. Genomic data from the ICGC and MOCOG cohorts were uniformly processed and analyzed for the current study. The analysis cohort consisted of 126 female patients with HGSC (Extended Data Fig. 1a), diagnosed between the ages of 29 and 81 years (Extended Data Fig. 1b and Supplementary Table 1).

Biospecimens. Normal DNA was isolated from peripheral lymphocytes or lymphoblastoid cell lines using the salting out method, the QIAamp DNA Blood Mini Kit (QIAGEN) or the FlexiGene DNA Kit (QIAGEN) using the AutoGen FlexSTAR+ instrument according to the manufacturer's instructions. Tumor DNA was extracted from fresh frozen cryosectioned tumor tissue using either the DNeasy Blood & Tissue Kit (QIAGEN), the AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN) or the Gentra Puregene Kit (QIAGEN) according to the manufacturer's instructions. Tumor RNA was extracted from fresh frozen cryosectioned tumor tissue using the mirVana miRNA Isolation Kit (Ambion/Life Technologies), the AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN) or the RNeasy Mini Kit (QIAGEN) using the QIAcube automated system according to the manufacturer's instructions. DNA was quantified using the Qubit dsDNA BR Assay (Invitrogen), the Lunatic spectrometer (Unchained

Labs) and the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). RNA quality and quantity were assessed using the Bioanalyzer RNA 6000 Nano assay (Agilent) and the NanoDrop Spectrophotometer (Thermo Fisher Scientific).

Molecular assays

Single-nucleotide polymorphism (SNP) arrays and quality control. Tumor and matched normal DNA was assayed with the Infinium OmniExpress-24 BeadChip arrays, arrays scanned and data processed using Genotyping module 2.0.3 software in GenomeStudio 2.0.3 to calculate logR ratios and B-allele frequencies according to manufacturer's instructions (Illumina) at the Australian Genome Research Facility (AGRF; Melbourne, Australia). HYSYS⁵⁵ was used to confirm correspondence of normal and tumor DNA, and tumor cellularity was assessed using qPure⁵⁶ and ASCAT⁵⁷, based on B-allele frequencies for ~67k common probes between HumanOmni2.5-8v1_A, HumanOmni25M-8v1-1_B, InfiniumOmniExpress-24v1-2_A1 and InfiniumOmniExpress-24v1-3_A1 SNP array platforms. Tumor DNA samples with estimated tumor cellularity >40% proceeded to WGS and methylation arrays. B-allele frequencies were also used to visually inspect profiles across tumor and germline samples.

Methylation arrays. Quality assessment was performed by QuantiFluor (Promega) and 500 ng tumor DNA was bisulfite converted with the EZ DNA Methylation kit (Zymo Research) and assayed using the Infinium MethylationEPIC BeadChip arrays according to manufacturer's instructions (Illumina) at the AGRF.

WGS. Sequence libraries were generated from tumor and matched normal genomic DNA using the KAPA HyperPrep PCR-free library preparation kit (Roche) according to manufacturer's instructions. Sequencing was carried out by the Kinghorn Centre for Clinical Genomics Sequencing Laboratory (Sydney, Australia) on the HiSeq X Ten System (Illumina) to a minimum base coverage of 30-fold for normal DNA and 60-fold for tumor DNA samples.

RNA-seq. Quality assessment was performed using the Bioanalyzer RNA 6000 Nano assay (Agilent), finding a median RNA integrity number of 9.0 (range 4.7 to 10). Libraries were generated using Illumina Stranded mRNA Prep and 150-bp paired-end sequencing was performed to a minimum of 100 million reads on Illumina NovaSeq 6000 instruments at the AGRF in accordance with the manufacturer's instructions.

Processing of whole-genome sequence data

FASTQ files were assessed for sequencing quality using FASTQC (v0.11.8) and for contaminants using FastQ Screen⁵⁸ (v0.11.4). The files were trimmed of adapters, low-quality bases and N content using fastq-mcf from ea-utils (v1.05). Sequence data were mapped to the human genome reference GRCh37b37 using BWA mem⁵⁹ (v0.7.17-r1188), producing BAM files. BAM files were then sorted, lanes merged and duplicates marked using Picard Tools (v2.17.3). Bases were recalibrated using GATK⁶⁰ BaseRecalibrator (v4.0.10.1). Coverage was calculated using GATK DepthOfCoverage (v3.8-1-0-gf15c1c3ef), and metrics such as insert size distribution, OxoG, base quality, GC bias and quality distribution were generated using Picard Tools (v2.17.3). GATK HaplotypeCaller (v4.0.10.1) was used on germline BAMs to generate Genomic Variant Call Format files, which were used as the 'Panel of Normal' for Mutect2 variant calling. Tumor purity and ploidy was estimated using FACETS⁶¹ (v0.6.1).

Variant detection

Germline variant calling. Germline base substitution and INDEL variants were called using VarDictJava (v1.5.7 with $-r = 2 - Q = 10 - f = 0.1$) for genes of interest (Supplementary Table 5).

Somatic base substitution and INDEL calling. Four variant calling tools were used to call somatic base substitutions and INDELS, as follows: Mutect2 (ref. ⁶⁰) (v4.0.11.0 with defaults), VarDictJava⁶² (v1.5.7 with $-r = 2 - Q = 10 - V = 0.05 - f = 0.01$), Strelka2 (ref. ⁶³) (v2.9.9 with defaults), and VarScan2 (ref. ⁶⁴) (SAMtools⁶⁵ v1.9 for mpileup and VarScan2 v2.4.3 with $-min\text{-coverage } 7 - min\text{-var-freq } 0.05 - min\text{-freq-for-hom } 0.75 - p\text{-value } 0.99 - somatic\text{-p-value } 0.05 - strand\text{-filter } 0$). Variant calls from all four tools were then decomposed (that is, multi-allelic to biallelic) and normalized (that is, left trimmed) using vt⁶⁶ (v0.57721). The passing variants for each caller were then processed using GATK ReadBackPhasing (v3.8-1-0-gf15c1c3ef with $-phase\text{-QualityThresh } 10 - enableMergePhasedSegregatingPolymorphismsToMNP - min_base_quality_score } 10 - min_mapping_quality_score } 10 - maxGenomicDistanceForMNP } 2$). The main purpose of running this tool was to combine contiguous SNVs to multinucleotide variants (for example, a DBS). The variant call format (VCF) files per caller were then merged using GATK CombineVariants (v3.8-1-0-gf15c1c3ef with $-genotypeMergeOptions UNIQIFY -priority Strelka2, Mutect2, VarScan2, VarDictJava$). The combined VCF was split and left trimmed using vt. Any variants that failed all callers were excluded. The VCF was annotated for homopolymers and tandem repeats using GATK VariantAnnotator (v3.8-1-0-gf15c1c3ef with $-reference_window_stop } 1000 - A HomopolymerRun - A TandemRepeatAnnotator$). High-confidence variants were those that passed at least two callers, had at least one variant containing read in each strand, were not in the Duke and DAC blacklisted regions and were not in the list of Frequently mutated Genes (FLAGS⁶⁷).

Structural variant detection. Four callers were used to identify somatic structural variants: Manta⁶⁸ + BreakPointInspector (v1.5.0), GRIDSS⁶⁹ (v2.0.1), Smoove (v0.2.2) and SvABA⁷⁰ (v134). Structural variant calls were separated into germline and somatic VCFs. For each germline/somatic VCF from the four callers, a custom R script used the StructuralVariantAnnotation and rtracklayer⁷¹ libraries to merge the SVs and generate a combined VCF from the four callers. A value of 10 was used for the maxgap parameter along with the strand orientation of the method findBreakpointOverlaps to identify common structural variants across the callers. Structural variants were annotated as duplication, deletion, inversion or translocation using a simple event type classifier provided by the GRIDSS package. Breakpoints called by at least two callers were deemed high confidence.

CNV detection. SNP pileup frequencies on common SNPs (dbSNP build 151, reference = GRCh37.p13, N = 37,906,831) were generated for tumor and normal BAMs. Pileups were generated using the snp-pileup tool (with $-pseudo\text{-snps } 100 - min\text{-map-quality } 10 - min\text{-base-quality } 10 - max\text{-depth } 5000 - min\text{-read-counts } 15,0$) as provided by the developers of FACETS⁶¹ (v0.6.1). The pileups were then used for cnv_facets (v0.13.0), which is a convenience tool for FACETS that executes all necessary steps to generate a VCF from the BAMs. Various values of pre-processing and processing cvals along with nbhd-snp were used for the analysis. The settings with the most robust CNV calls and purity agreement with the SNP array data were used for further analysis. The settings used for FACETS were ($-nbhd\text{-snp} = 500 - cval = 50 - 1000 - depth = 15 - 5000$).

Whole-genome duplication and whole-genome loss. Whole-genome duplication percentages were assessed using previously established methods⁷². Briefly, the percentage genome with a major copy number (MCN) of greater than or equal to two was calculated. The same method was applied to assess whole-genome loss, where percentage genome with a total copy number of less than or equal to one was calculated.

Annotation of variants in genes of interest

High-confidence base substitutions and INDELS were filtered to remove (1) variants with less than four supporting reads and/or variants without bidirectional read support, (2) all silent (synonymous) mutations

with no prior evidence of being pathogenic, (3) common variants with a global minor allele frequency > 0.001 in the Genome Aggregation Database (gnomAD) v2.1.1 (<https://gnomad.broadinstitute.org/>) and (4) variants previously found to be benign or low clinical significance in one or more mutation databases (<https://www.ncbi.nlm.nih.gov/clinvar/>, <https://brcaexchange.org/>). Structural variants that were detected within a gene footprint were considered truncating if the variant was (1) a translocation that breaks the gene anywhere between the translation start site and the first base of the final coding exon, (2) a deletion, duplication, or inversion that spans one or more exons (unless it only spans the final coding exon) or (3) a deletion, duplication or inversion that results in a frameshift within an exon (unless it is within the final coding exon). CNVs were considered pathogenic if (1) a region of homozygous deletion (gene level copy number = 0) spans the whole gene or a coding exon (unless it only spans the final coding exon), or (2) a region of amplification (gene level copy number ≥ 7) spans the whole gene.

Evidence of mutation was sought from both WGS and RNA-seq data, and manual review of germline and somatic variants in genes of interest was carried out using Integrative Genomics Viewer⁷³. Manually curated genes and pathogenic variants with supporting evidence are listed in Supplementary Tables 5–7. Mutations reported in Supplementary Tables 5–7 were only those deemed pathogenic, that is truncating mutations (nonsense, splice site, frameshift, deletions, duplications, inversions and translocations that disrupt the coding transcript) and missense variants previously reported as pathogenic or likely pathogenic in curated mutation databases (<https://tp53.isb-cgc.org/>, <https://www.ncbi.nlm.nih.gov/clinvar/>, <https://brcaexchange.org/>)

Homologous recombination pathway analysis

In addition to pathogenic germline and somatic mutations in genes involved with homologous recombination and DNA repair (Supplementary Tables 5 and 6), the promoter methylation status of *BRCA1* and *RADSIC* was determined in tumor samples using methylation array data and gene expression (Supplementary Methods). Tumor samples with multiple potential driver gene alterations were assigned to a primary alteration category in the following order of preference: (1) germline mutation in homologous recombination gene (*BRCA1*, *BRCA2*, *BRIPI1*, *PALB2*, *RADSIC*, or *RADSID*), (2) somatic promoter methylation of *BRCA1* or *RADSIC*, (3) somatic mutation in homologous recombination gene (*BRCA1*, *BRCA2*, *BRIPI1*, *PALB2*, *RADSIC*, or *RADSID*), (4) somatic *CDK12* mutation, (5) somatic *CCNE1* amplification, (6) somatic mutation in putative homologous recombination gene (*BARD1*, *BLM*, *CHEK2*, *FANCA*, *FANCD2*, *FANCE*, *FANCI*, *FANCM*, *PTEN*, *ATM*, *ATR*, or *RADS1B*), (7) somatic mutation in mismatch repair gene (*MSH2*, *MSH6*, *PMS1*, or *PMS2*), (8) wild-type (no germline or somatic homologous recombination alteration, *CDK12* mutation, *CCNE1* amplification, or mismatch repair mutation). Where multiple potential driver mutations were identified, the variant allele frequency and/or mutational signatures were used to assign the likely driver.

Multiple DNA repair pathway alterations. To determine the number of DNA repair alterations per sample, all independent germline and somatic alterations were tallied in the following gene sets: (1) homologous recombination pathway genes, (2) putative homologous recombination pathway genes, (3) mismatch repair genes and (4) *CDK12* and *RBI*.

Homologous recombination deficiency. Homologous recombination deficiency was estimated in tumor samples using CHORD²⁸ and scarHRD⁷⁴.

Mutational signatures

Mutational signatures were generated for high-confidence variants as described above. Variants for each sample were converted into catalogs

or categories of mutational spectra for SBSs, DBSs and INDELS using the R package ICAMS v2.0.10.9001 (<https://github.com/steverozen/ICAMS>) and the function 'VCFsToCatalogs'. Each type of mutational catalog contains a number of contexts based on the COSMIC definitions³², namely 96 contexts for SBSs, 78 contexts for DBSs and 84 contexts for INDELS. This provides a sample by mutation context matrix per SBS, DBS and INDEL type. Structural variant signature catalogs consisting of 32 contexts were generated using the R package signature.tools.lib³³ (v0.0.0.9000) and the function 'bedpeToRearrCatalogue'. The SBS, DBS and INDEL catalogs were then fit to the COSMIC Mutational Signatures v3.2 database (<https://cancer.sanger.ac.uk/signatures/>), and the structural variant catalogs were fit to the ovary-specific rearrangement signatures (<https://signal.mutationalsignatures.com/>). Further details on mutational signature fitting and clustering are described in the Supplementary Methods.

Neoantigen prediction

HLA types were generated using HLA-VBSeq⁷⁵ (v11_22_2018) for neoantigen prediction as follows. Unmapped reads and reads mapped to HLA regions were extracted using jvarkit samviewwithmate (ec2c236) and converted to FASTQ files using samtools view (v1.9) along with Picard Tools SamToFastq (v2.17.3); these were then mapped using BWA mem (v0.7.17-r1188) to the HLA v2 database based on IMGT/HLA Database⁷⁶ release 3.31.0 and Japanese HLA reference dataset for HLA estimation. The HLA types were fed into pVACtools⁷⁷ pVACseq (v1.3.5) to identify and construct neoantigens from the high-confidence variants. Briefly, high-confidence coding variants in VCF format were annotated using the VEP (v92.4) plugin 'Downstream', which provides the predicted downstream protein sequence and the change in length relative to the reference protein, and the plugin 'Wildtype', which includes the transcript protein sequence in the annotation. RNA read counts for the annotated variants were generated using bam_readcount_helper.py and added to the VCF using vatools vcf-readcount-annotator. Normalized transcripts per million were added using vatools vcf-expression-annotator. Finally, pVACseq was run against this final annotated VCF for both MHC Class I and MHC Class II predictions.

RNA-seq data processing and quality control

Initial quality control checks on raw FASTQ files were performed using FastQC (v0.11.8). Reads were trimmed for low quality, adapters, N content and poly(A) tails using fastq-mcf (v1.05), and contamination assessed using FastQScreen (v0.11.4). Reads were mapped to the human reference GRCh37.92 using the STAR⁷⁸ two pass method (v2.6.0b). Mapped reads were sorted using Picard Tools (v2.17.3). Additional quality control after mapping was performed using Picard Tools CollectRnaSeqMetrics (v2.17.3) and RSeQC⁷⁹ (v2.6.4). Counts were generated on the Ensembl release GRCh37.92 gene annotation using HTSeq⁸⁰ (v0.10.0). Counts were generated on the exons only using the 'intersection-nonempty' mode.

Raw counts data were filtered to only include protein coding genes. To remove lowly expressed genes, the data were converted to CPM (counts per million = number of reads mapped to a gene $\times 10^6$ /total number of mapped reads), and only genes where at least 10 samples had a CPM of greater than 0.5 were kept for further processing. The data were normalized using the trimmed mean of M values (TMM) method in edgeR⁸¹ and batch effects removed using the removeBatchEffect function of limma⁸². Further details on batch correction and expression analyses are provided in the Supplementary Methods.

Methylation data processing and quality control

Methylation data quality control assessment and processing were performed using the R package minfi⁸³ (v1.32.0). Probes failing detection ($P > 0.01$), SNP positions and cross-reactive probes (as collected in https://github.com/sirselim/illumina450k_filtering) were excluded.

Data were normalized using the minfi function 'preprocessFunnorm' (Functional normalization as described previously⁸⁴), and beta values were generated. Probes were annotated to the Ensembl release GRCh37.92 gene transfer format annotation. Beta values for samples from EPIC and 450k arrays were combined to contain shared probes and batch corrected using the 'ComBat' function in the R package sva⁸⁵ (v3.34.0).

Statistical analyses

Differences in proportions of categorical variables between groups were assessed by the chi-square or Fisher's exact test as appropriate. Continuous variables were evaluated using either a Kruskal–Wallis test or a Mann–Whitney test. The Kaplan–Meier methodology was applied to estimate and plot progression-free and OS probabilities and the corresponding time to event were compared between groups using the log-rank (Mantel–Cox) test. For display purposes, the x axis in Kaplan–Meier plots is capped at 15 years. Outcomes were assessed using univariable and multivariable Cox proportional hazards models, for continuous and categorical features, using the 'coxph' function of the R package survival (v3.2-7) with default parameters. Continuous variables were scaled and centered across the cohort using the R function 'scale'. Correlations between continuous variables were assessed by Spearman correlation. All statistical tests were two sided and considered significant when $P < 0.05$. The Benjamini–Hochberg procedure was applied to correct P values for the impact of multiple testing, with false discovery rate-adjusted P values denoted by P_{adj} . R (v3.6.3) and Prism (v9.2.0) were used for statistical analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

ICGC datasets: Previously published WGS and RNA-seq data generated as part of the ICGC Ovarian Cancer project¹⁴ are available from the European Genome-phenome Archive (EGA) repository (<https://ega-archive.org>) as a single BAM file for each sample type (tumor/normal) under the accession code EGAD00001000877. Due to the sensitive nature of these patient data sets, access is subject to approval from the ICGC Data Access Compliance Office (<https://docs.icgc.org/download/data-access/>), an independent body who authorizes controlled access to ICGC sequencing data. ICGC SNP array and methylation data sets have been deposited into the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession code GSE65821, without access restrictions. ICGC gene count level transcriptomic data have been deposited into the GEO under accession code GSE209964.

MOCOG datasets: WGS, RNA-seq and SNP array data from long-term survivors generated as part of the MOCOG study have been deposited in the EGA repository under accession code EGAS00001005984. WGS and RNA-seq data are available as raw FASTQ files for each sample type (tumor/normal) and SNP array data are available as raw signal intensity files in text format for each sample type (tumor/normal). Access to patient sequence data can be gained for academic use through application to the independent Data Access Committee (dac@petermac.org). Responses to data requests will be provided within two weeks. Information on how to apply for access is available at the EGA under accession code EGAS00001005984. The MOCOG cohort raw methylation data sets have been submitted to the GEO under accession code GSE211687, with no access restrictions.

Uniformly processed somatic variant data from the ICGC and MOCOG cohorts have been deposited in Synapse under accession code syn34616347, and processed expression and methylation data from both cohorts have been submitted into the GEO under accession code GSE211687, without access restrictions.

Population frequencies of genetic variants can be accessed via the Genome Aggregation Database (gnomAD) at <https://gnomad.broadinstitute.org/>. Supporting evidence for pathogenicity of genomic alterations can be accessed via ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), BRCA Exchange (<https://brcaexchange.org/>) and the TP53 Database (<https://tp53.isb-cgc.org/>). The Ensembl ranked order of severity of variant consequences is available at: https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html. Precomputed TCGA ovarian serous cystadenocarcinoma survival analysis data can be downloaded from OncoLnc (<http://www.oncolnc.org/>). Mutational signature reference databases can be accessed via COSMIC (<https://cancer.sanger.ac.uk/signatures/>) and Signal (<https://signal.mutationalsignatures.com/>). The LM22 signature matrix used for immune cell deconvolution can be downloaded at <https://cibersortx.stanford.edu/>. The COSMIC Cancer Gene Census can be accessed at <https://cancer.sanger.ac.uk/census>. MSigDB hallmark gene sets can be accessed at <https://www.gsea-msigdb.org/gsea/msigdb/>. Illumina methylation probes that were filtered out due to poor performance (for example, cross-reactive or nonspecific probes) can be found at https://github.com/sirselim/illumina450k_filtering. Germline polymorphic sites for reference and variant allele read counts used in FACETS analysis can be found at ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/common_all_20180423.vcf.gz. The gene transfer format used for annotation and RNA-seq counts is available at <ftp://ftp.ensembl.org/pub/grch37/release-92/>. All other data are available within the article and its supplementary information files.

Code availability

No custom code or software was used in the data analyses. All results can be replicated using publicly available tools and software. The tools and versions used are fully described in the Methods and Supplementary Information.

References

54. Cress, R. D., Chen, Y. S., Morris, C. R., Petersen, M. & Leiserowitz, G. S. Characteristics of long-term survivors of epithelial ovarian cancer. *Obstet. Gynecol.* **126**, 491–497 (2015).
55. Schröder, J., Corbin, V. & Papenfuss, A. T. HYSYS: Have you swapped your samples? *Bioinformatics* **33**, 596–598 (2017).
56. Song, S. et al. qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS One* **7**, 5–11 (2012).
57. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).
58. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* **7**, 1338 (2018).
59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
60. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
61. Shen, R. & Seshan, V. E. FACETS: Allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, 1–9 (2016).
62. Lai, Z. et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **44**, e108 (2016).
63. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
64. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
65. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
66. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
67. Shyr, C. et al. FLAGS, frequently mutated genes in public exomes. *BMC Med. Genomics* **7**, 64 (2014).
68. Chen, X. et al. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
69. Cameron, D. L. et al. GRIDSS: Sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* **27**, 2050–2060 (2017).
70. Wala, J. A. et al. SvABA: Genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).
71. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
72. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
73. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
74. Sztupinszki, Z. et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* **4**, 16 (2018).
75. Nariai, N. et al. HLA-VBSeq: Accurate HLA typing at full resolution from whole-genome sequencing data. *BMC Genomics* **16**, 1–6 (2015).
76. Robinson, J. et al. IPD-IMGT/HLA Database. *Nucleic Acids Res.* **48**, D948–D955 (2020).
77. Hundal, J. et al. PVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol. Res.* **8**, 409–420 (2020).
78. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
79. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
80. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
81. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
82. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
83. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
84. Fortin, J.-P. et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
85. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

Acknowledgements

We thank P. Webb, K. Byth, R. Lupat, J. Ellul and the Peter MacCallum Cancer Centre Research Computing Facility for their contributions to the study. This work was supported by the U.S. Army Medical Research and Materiel Command Ovarian Cancer Research Program (Award No. W81XWH-16-2-0010 and W81XWH-21-1-0401), the National Health and Medical Research Council of Australia (1092856, 1117044 and 2008781 to D.D.L.B., and 1186505 to D.W.G.), and the U.S. National Cancer Institute (P30CA046592 for C.L.P. and P30CA008748 for M.C.P.). This research was made

possible by generous support from the Border Ovarian Cancer Awareness Group, the Garvan Research Foundation, the Graf Family Foundation, Mrs Margaret Rose AM, Arthur Coombs and family, and the Piers K Fowler Fund. The Australian Ovarian Cancer Study (AOCS) gratefully acknowledges the cooperation of participating institutions in Australia and the contribution of study nurses, research assistants and all clinical and scientific collaborators. The complete AOCS Group can be found at www.aocstudy.org. We would like to thank all of the women who participated in the study. AOCS was supported by the U.S. Army Medical Research and Materiel Command (DAMD17-01-1-0729), The Cancer Council Victoria, Queensland Cancer Fund, The Cancer Council New South Wales, The Cancer Council South Australia, The Cancer Council Tasmania, The Cancer Foundation of Western Australia and the National Health and Medical Research Council of Australia (NHMRC; ID199600, ID400413, ID400281). AOCS gratefully acknowledges additional support from Ovarian Cancer Australia and the Peter MacCallum Cancer Foundation. We thank all the women who participated in the GynBiobank and gratefully acknowledge the Departments of Gynaecological Oncology, Medical Oncology and Anatomical Pathology at Westmead Hospital, Sydney. The Gynaecological Oncology Biobank at Westmead was funded by the NHMRC (ID310670, ID628903), the Cancer Institute NSW (12/RIG/1-17, 15/RIG/1-16) and the Department of Gynaecological Oncology, Westmead Hospital, and acknowledges financial support from the Sydney West Translational Cancer Research Centre, funded by the Cancer Institute NSW (15/TRC/1-01). E.L.C. was supported by NHMRC grant APP1161198. F.A.M.S. was supported by a Swiss National Foundation EarlyPostdoc Fellowship (P2BEP3-172246), Swiss Cancer Research Foundation grant BIL KFS-3942-08-2016 and a Professor Dr Max Cloëtta and Uniscientia Foundation grant. A.M.P. and J.D.B. were supported by Cancer Research UK (A22905). B.H.N. was supported by the BC Cancer Foundation, Canada's Networks of Centres of Excellence (BioCanRx), Genome BC and the Canada Foundation for Innovation. D.D.L.B. was supported by the U.S. National Cancer Institute U54 program (U54CA209978).

Author contributions

D.W.G., S.F., A.D. and D.D.L.B. conceived the project. C.J.K., K.A., N.T., G.A.-Y., A.B., M.F., P.R.H., O.M., E.L.G., A.D. and D.D.L.B. provided patient samples and clinical information, and D.W.G., J. Hendley, Y.C. and E.L.C.

prepared patient samples. D.W.G., A.P., S.F., C.J.K., K.A., D.A., L.B., J.M.C., J. Hung, B.M., R.A.V., C.W., S.J.W. and E.L.G. acquired data. A.P. managed and processed genomic data and performed data analysis together with D.W.G., S.F., K.T., P.P., D.A., E.L.C., B.M., F.A.M.S., R.A.V., C.W. and S.J.W. D.W.G., S.F., K.A., C.L.P., M.C.P. and D.D.L.B. coordinated the study and interpreted the results along with A.P., K.T., P.T.H., N.W.B., E.L.C., B.G., A.M.P., F.A.M.S., P.D.P.P., J.D.B., T.P.C., G.L.M., S.J.R., B.H.N. and A.D. D.W.G., A.D. and D.D.L.B. supervised the work and wrote the manuscript together with A.P. and K.T. All authors discussed the results and read the manuscript.

Competing interests

S.F., K.A., N.T. and A.D. received grant funding from AstraZeneca for unrelated work. G.A.-Y. received grant funding from AstraZeneca and Roche-Genentech for unrelated work. M.F. declares honoraria for advisory boards AstraZeneca, GSK, Incyclix, Lilly, MSD, Novartis and Takeda; consultancy for AstraZeneca, Eisai and Novartis; speaker's fee and travel from AstraZeneca; speaker's fee from ACT Genomics; and institutional research funding from AstraZeneca, BeiGene, Novartis; all for unrelated work. J.D.B. received funding from Aprea and Clovis Oncology for unrelated work. D.D.L.B. received funding from AstraZeneca, Genentech-Roche and BeiGene for unrelated work. The remaining authors declare no competing interests.

Additional information

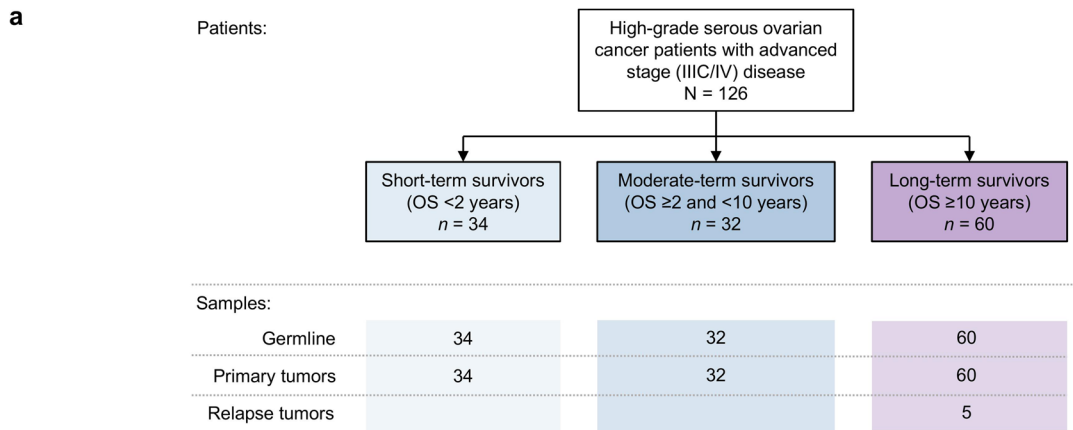
Extended data is available for this paper at <https://doi.org/10.1038/s41588-022-01230-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01230-9>.

Correspondence and requests for materials should be addressed to Dale W. Garsed or David D. L. Bowtell.

Peer review information *Nature Genetics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

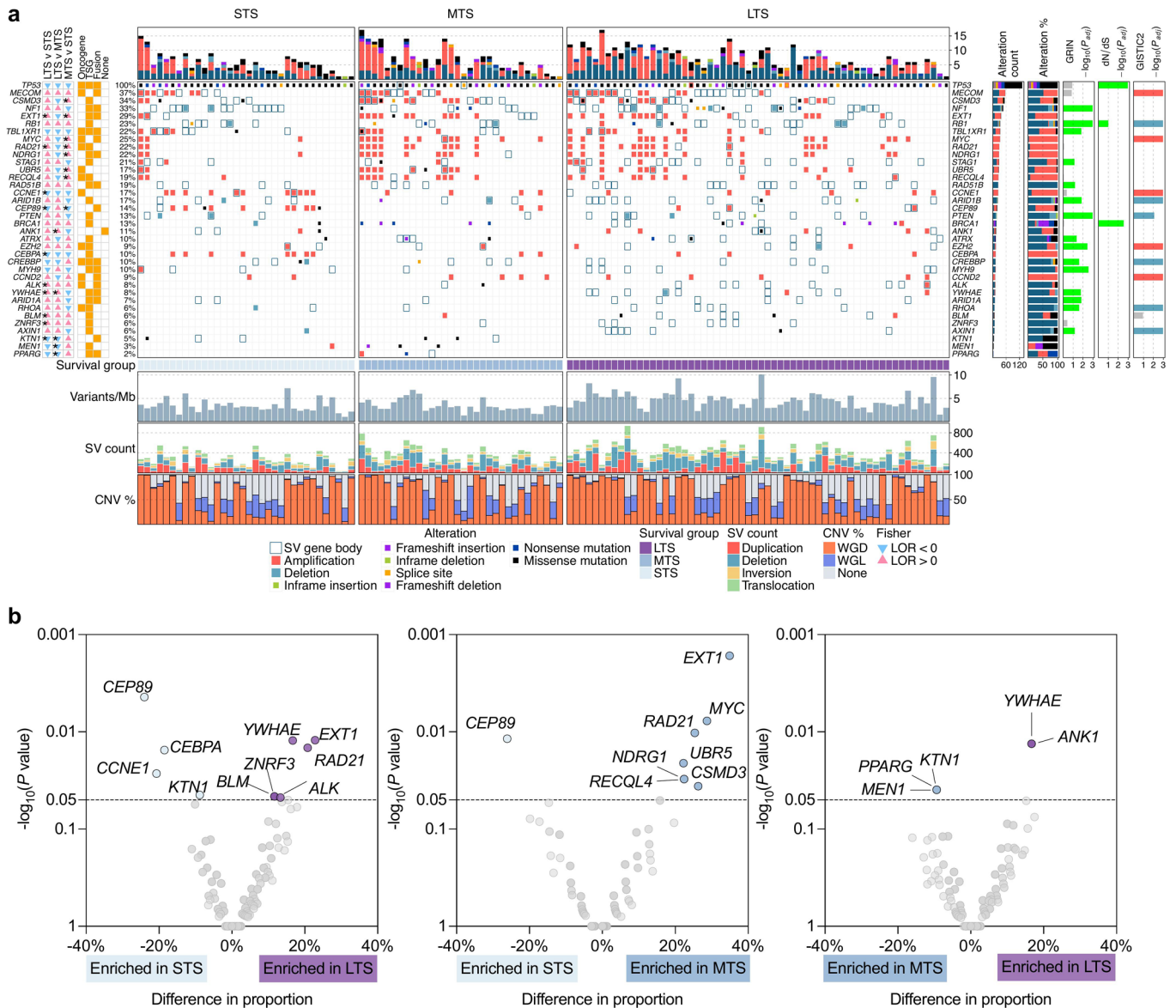


b

Characteristics	Short-term survivors (n = 34) n (%)	Moderate-term survivors (n = 32) n (%)	Long-term survivors (n = 60) n (%)	P
Age at diagnosis (years)				
Median	60.5	57.0	60.0	0.2222 ^a
Range	45–78	39–73	29–81	
Primary site				
Ovary	22 (65)	28 (88)	51 (85)	0.1010 ^b
Peritoneum	11 (32)	3 (9)	8 (13)	
Fallopian tube	1 (3)	1 (3)	1 (2)	
Grade				
2	5 (15)	3 (9)	10 (17)	0.6335 ^b
3	29 (85)	29 (91)	50 (83)	
FIGO stage				
IIIC	31 (91)	28 (88)	49 (82)	0.4243 ^b
IV	3 (9)	4 (13)	11 (18)	
Residual disease				
Nil	1 (3)	2 (6)	13 (22)	0.0227 ^b
≤1 cm	18 (53)	21 (66)	24 (40)	
>1 cm	13 (38)	7 (22)	13 (22)	
Size unknown	2 (6)	2 (6)	6 (10)	
Not known	0 (0)	0 (0)	4 (7)	
Neoadjuvant therapy				
Yes	3 (9)	1 (3)	0 (0)	0.0640 ^b
No	31 (91)	31 (97)	60 (100)	
Smoking				
Never smoked	19 (56)	11 (34)	30 (50)	0.1433 ^b
Ever smoked	13 (38)	16 (50)	18 (30)	
Not known	2 (6)	5 (16)	12 (20)	
Current status				
Alive with no progression	0 (0)	0 (0)	29 (48)	<0.0001 ^b
Alive with progression	0 (0)	0 (0)	13 (22)	
Dead	34 (100)	32 (100)	18 (30)	
Progression-free survival from diagnosis (months)				
Median	8.0	12.6	161.8	<0.0001 ^c
Range	1.84–13.31	4.93–79.99	5.61–247.5	
Overall survival from diagnosis (months)				
Median	16.6	46.9	179.8	<0.0001 ^c
Range	6.05–23.64	24.13–108.3	120.0–247.5	

Extended Data Fig. 1 | Patient cohort. **a**, Overview of patients ($n = 126$) and tumor samples analyzed in this study. In addition to paired germline and primary tumor samples in all patients, 5 relapse tumor samples were also analyzed from 4 long-term survivor patients. OS, overall survival. **b**, Clinical characteristics

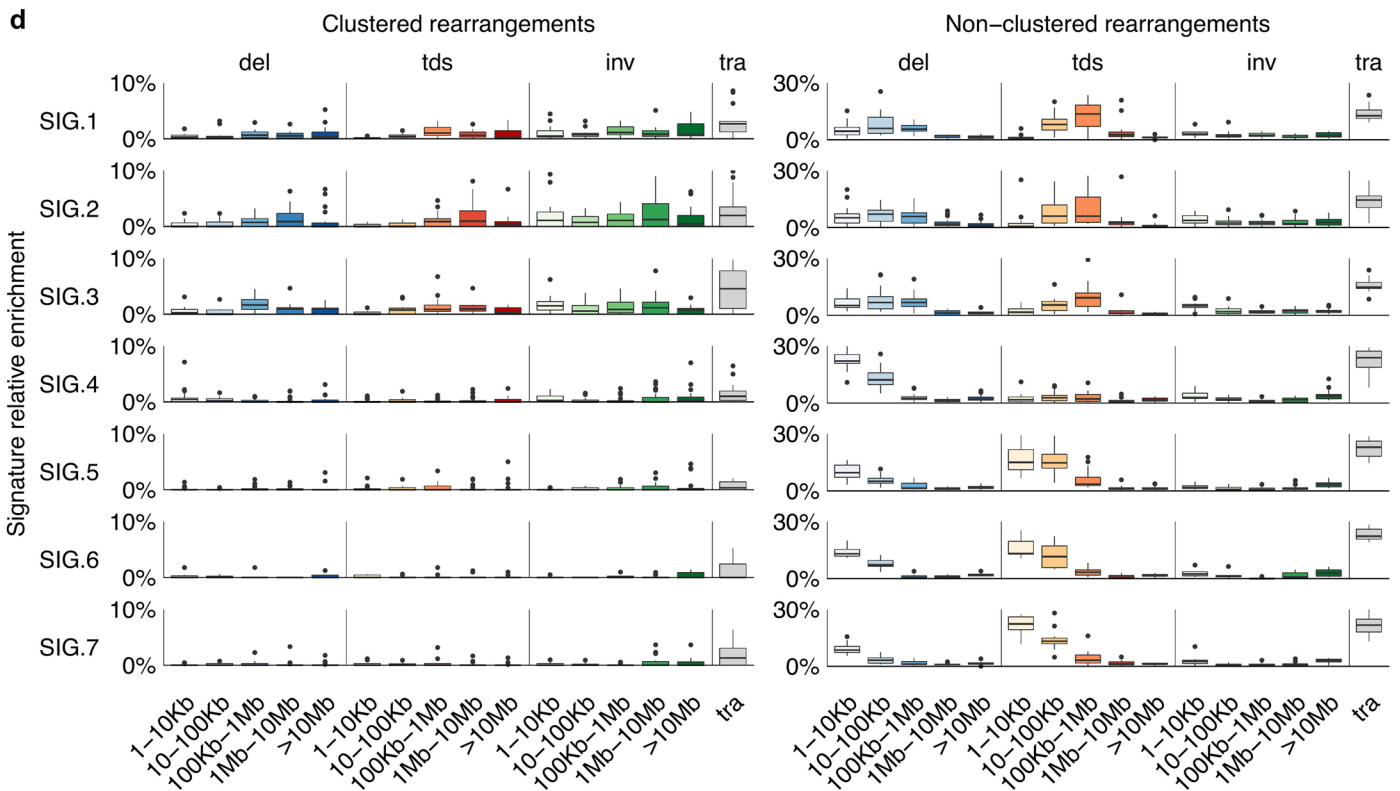
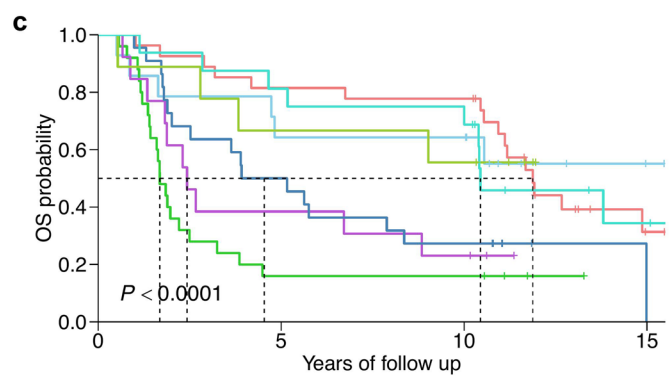
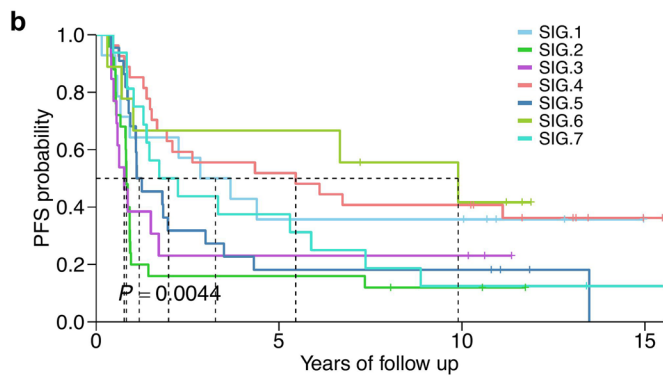
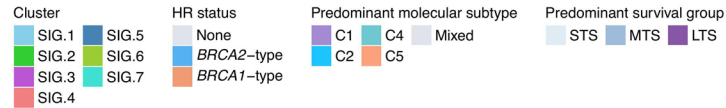
of patients by survival group. All patients received primary platinum therapy. ^aKruskal–Wallis, ^bChi-square, or ^clog-rank Mantel–Cox test P values comparing survival groups reported.



Extended Data Fig. 2 | Frequently altered cancer genes across survival groups. a, Overview of somatic alterations in driver genes detected by GRIN, dNdScv, GISTIC, and/or in cancer-associated genes (COSMIC Cancer Gene Census) that are enriched in a survival group relative to another survival group. From left: two-sided Fisher’s test of the difference in proportions of altered samples between survival groups, triangles and color indicate direction of the log odds ratio (LOR; blue = down, pink = up), asterisks indicate P value < 0.05 (see Supplementary Table 6 for P values), P values were not adjusted for multiple comparisons; role of gene in COSMIC Cancer Gene Census (TSG, tumor suppressor gene); genomic alterations split by survival groups, bars at the top indicate the number of alterations in each listed gene per patient; bar plot of the number of samples with an alteration (alteration type indicated by color); bar plots showing the proportion of alteration types per gene; P values were calculated using the genomic random interval (GRIN) statistical model (one-sided) for recurrent structural variants (SV) (see Supplementary Data 2 for GRIN P values), the dNdScv likelihood-ratio test (two-sided) for recurrent base substitutions and small-scale deletions and insertions (see Supplementary Data 1 for dNdScv P values), and GISTIC2 permutation-of-markers test (one-sided) for recurrent copy-number variants (CNV) with red indicating

amplification and blue indicating deletion (see Supplementary Data 3 for GISTIC2 P values), P values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure (dNdScv, GISTIC2) or the robust false discovery rate procedure (GRIN) and are shown as negative \log_{10} P values and capped at 0.001 for display purposes. Each patient (column) is annotated with survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor). Below the alterations are bar plots indicating somatic mutation burden in variants per megabase (Mb); SV count including duplications, deletions, inversions and intrachromosomal rearrangements; and the proportion of the tumor genome that is duplicated (WGD) or lost (WGL). **b**, Pairwise comparison of the alteration frequencies between survival groups for genes in the COSMIC Cancer Gene Census. The difference in relative alteration frequency is shown on the x-axis and the P value (Fisher’s test, two-sided) is shown on the y-axis. Symbols of genes with P values < 0.05 are displayed. Multiple hypothesis correction was not applied in this analysis as adjusted P values were all greater than 0.1. Alterations in this analysis included non-synonymous mutations, homozygous deletions, amplifications and structural variants in coding genes that are expressed.

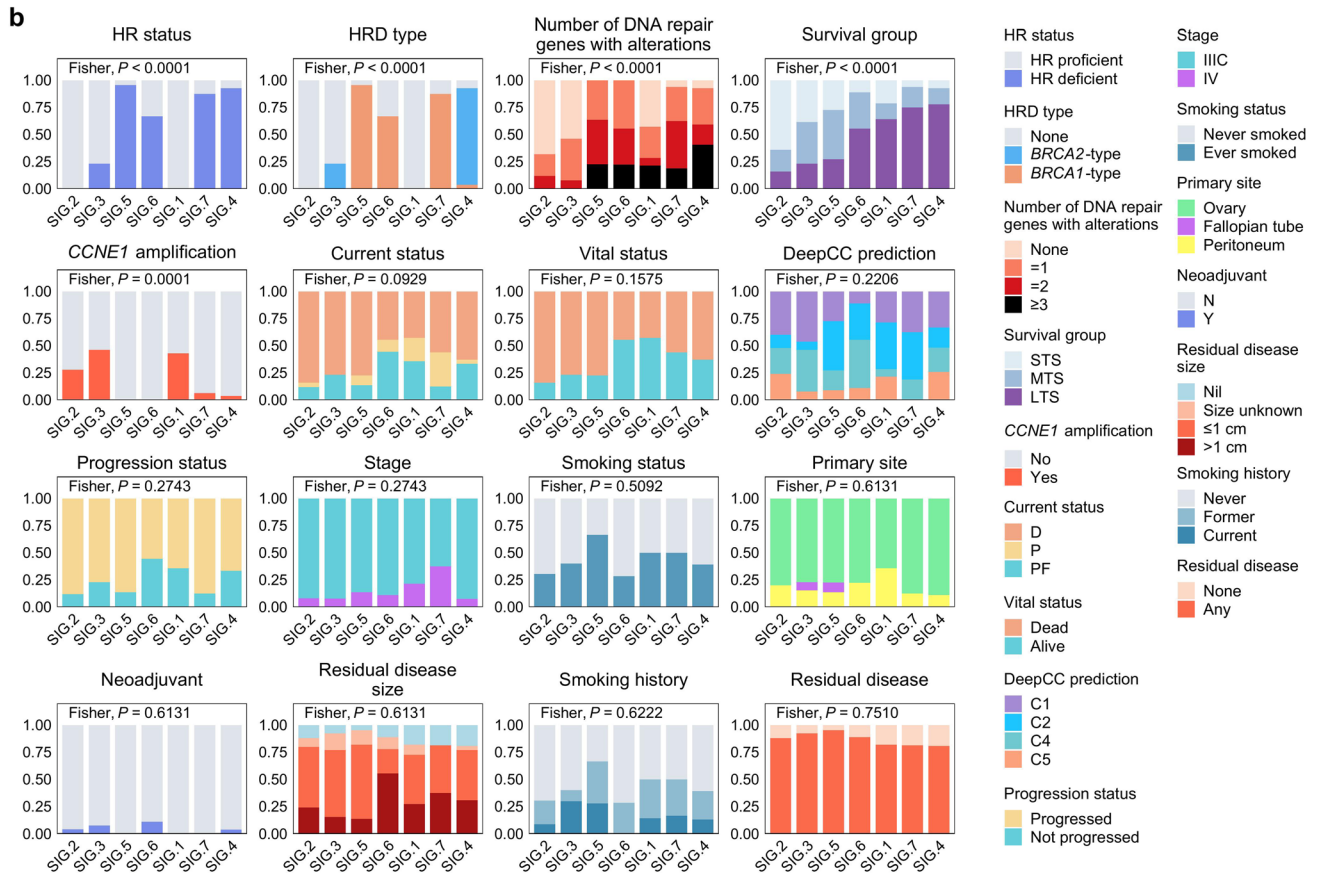
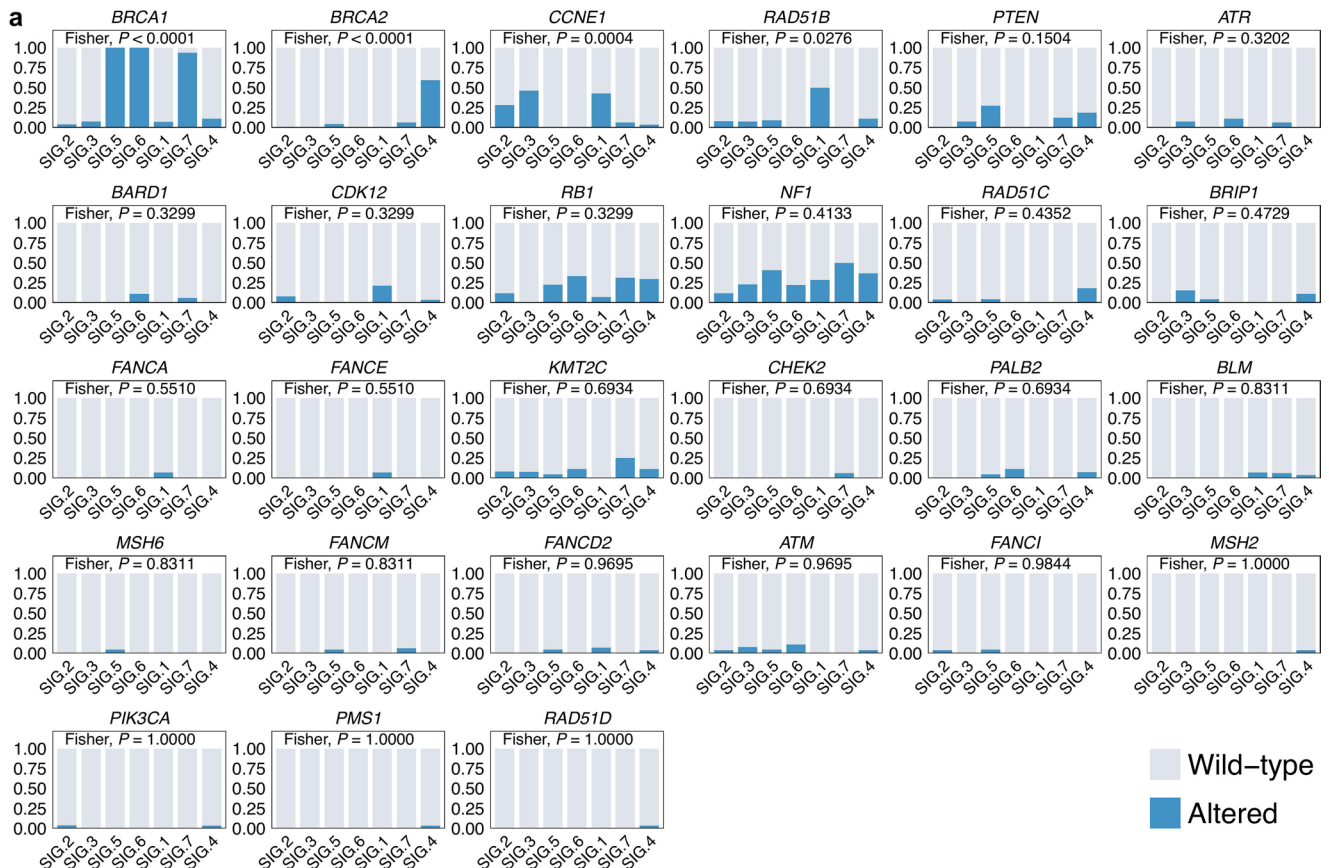
Cluster	Gene drivers	Age at diagnosis (median, years)	HR status	SV burden	SV (clustered/not-clustered)	Predominant SV type	% genome LOH	% genome loss	Ploidy	Predominant molecular subtype	Neoantigen burden	Number DNA repair alterations	Proportion alive at last follow-up	Overall survival (median, years)	Progression-free survival (median, years)	Proportion LTS	Predominant survival group
SIG.2 (n = 25)	CCNE1 amplification (28%)	63	Proficient	Low	Clustered	INV	Low	Medium	Medium	C1	Low	Low	0.16	1.68	0.83	0.16	STS
SIG.3 (n = 13)	CCNE1 amplification (46%)	66	Proficient	Low	Clustered	INV	Medium	Low	High	C1	Low	Low	0.23	2.43	0.77	0.23	MTS
SIG.5 (n = 22)	BRCA1 (100%) PTEN (27%) NF1 (41%)	53.5	BRCA1-type	Medium	Not-clustered	DUP	Medium	High	Low	C2	Medium	High	0.23	4.54	1.18	0.27	MTS
SIG.6 (n = 9)	BRCA1 (100%) RB1 (33%)	54	BRCA1-type	Medium	Not-clustered	DUP	High	High	Low	C4	Medium	High	0.56	Not reached	9.90	0.56	LTS
SIG.1 (n = 14)	CCNE1 amplification (43%) RAD51B (50%) CDK12 (21%)	63.5	Proficient	Medium	Mixed	DUP INV	Low	Low	High	C2	High	Medium	0.57	Not reached	3.26	0.64	LTS
SIG.7 (n = 16)	BRCA1 (94%), NF1 (50%), RB1 (31%) and KMT2C (25%)	59	BRCA1-type	High	Not-clustered	DUP	High	Medium	Medium	C2	High	High	0.44	10.45	1.98	0.75	LTS
SIG.4 (n = 27)	BRCA2, RAD51C, PALB2, BRIP1, PTEN, NF1, RB1 and/or RAD51D alterations (93%)	60	BRCA2-type	High	Not-clustered	DEL	Medium	Medium	Medium	Mixed	High	High	0.37	11.88	5.46	0.78	LTS



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Key features of mutational signature clusters and associated survival outcomes. **a**, Summary of the key clinical and genomic features of each mutational signature cluster. Clusters are ordered top to bottom by lowest to highest proportion of long-term survivors (LTS) in each cluster. HR, homologous recombination; LOH, loss-of-heterozygosity; SV, structural variant; MTS, moderate-term survivor; STS, short-term survivor; DUP, duplications; DEL, deletions; INV, inversions. **b**, Kaplan–Meier analysis of progression-free and **c**, overall survival in patients stratified by signature clusters. *P* values calculated by Mantel–Cox log-rank test and dotted lines indicate median survival. **d**,

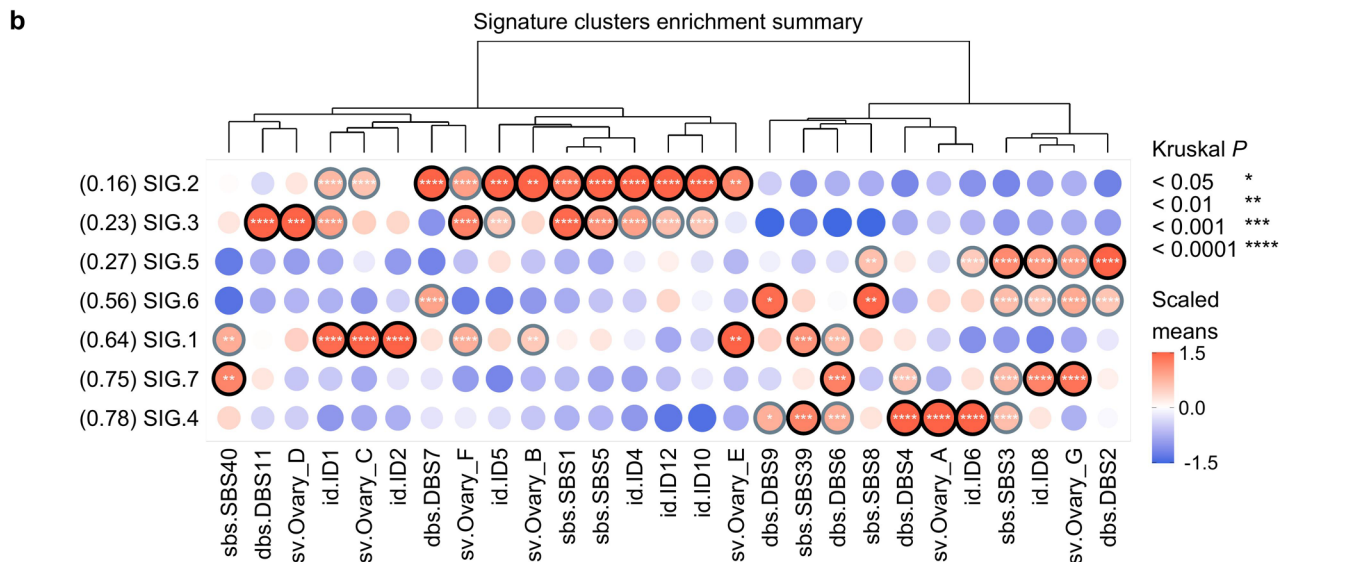
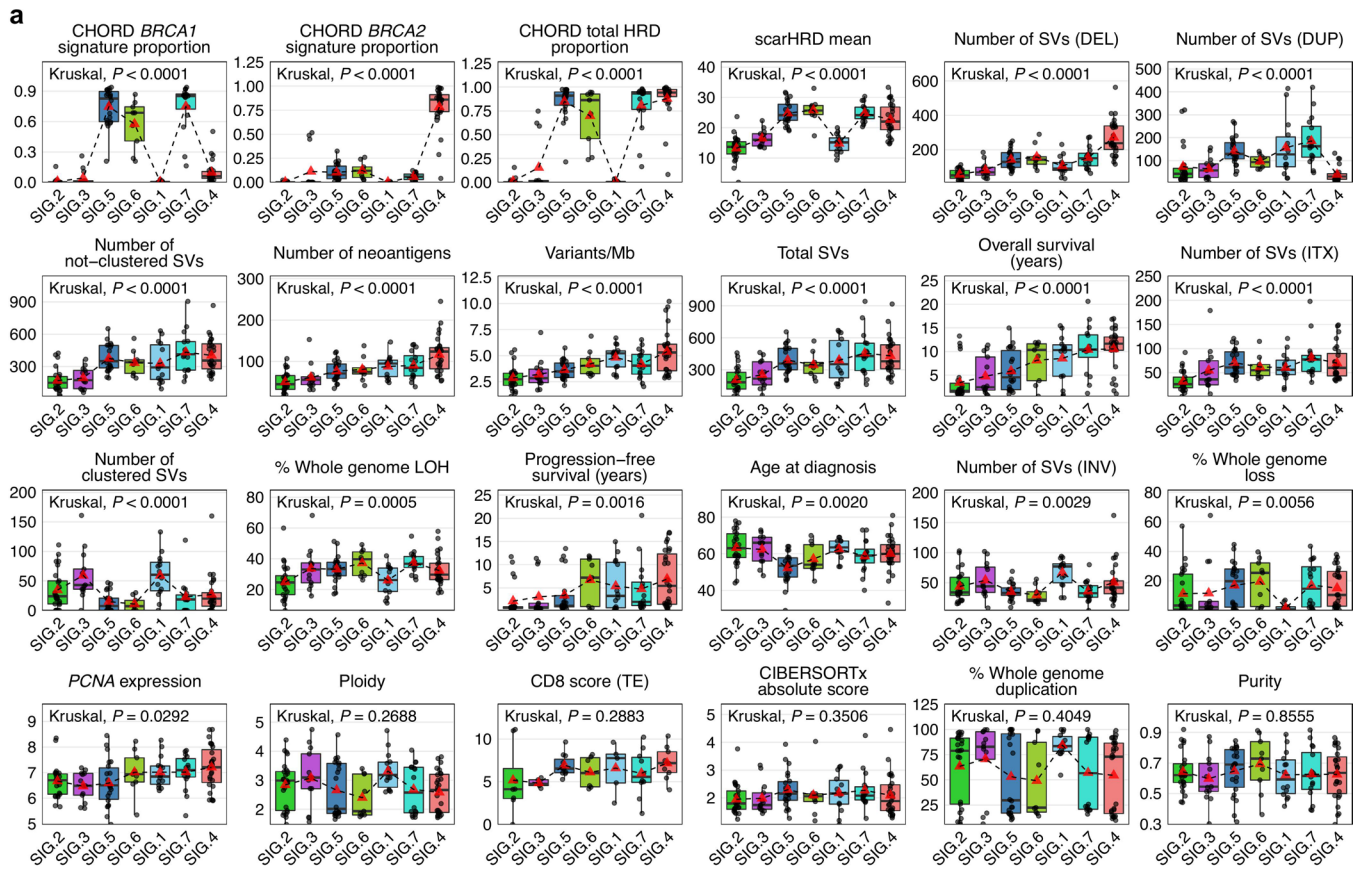
Boxplots summarize the proportion (y-axis) of clustered and nonclustered rearrangements by size (x-axis) and type, for each mutational signature cluster (SIG.1 *n* = 14, SIG.2 *n* = 25, SIG.3 *n* = 13, SIG.4 *n* = 27, SIG.5 *n* = 22, SIG.6 *n* = 9, SIG.7 *n* = 16); boxes show the interquartile range (25–75th percentiles), central lines indicate the median, whiskers show the smallest/largest values within 1.5 times the interquartile range and values outside it are shown as individual data points. Del, deletions; tds, tandem duplications; inv, inversions; tra, interchromosomal translocations; Kb, kilobase; Mb, megabase.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Categorical features of mutational signature clusters. **a**, Proportion of patients affected by gene alterations per mutational signature cluster. Genes are ordered by significance using Fisher's exact test (two-sided) and clusters are ordered by the proportion of long-term survivors. **b**, Proportion of patients with categorical features per cluster. Features are ordered by significance using Fisher's exact test (two-sided) and the clusters are arranged by the proportion of long-term survivors. The Fisher's test *P* values displayed in **(a)** and **(b)** are Benjamini-Hochberg adjusted *P* values. Features include homologous

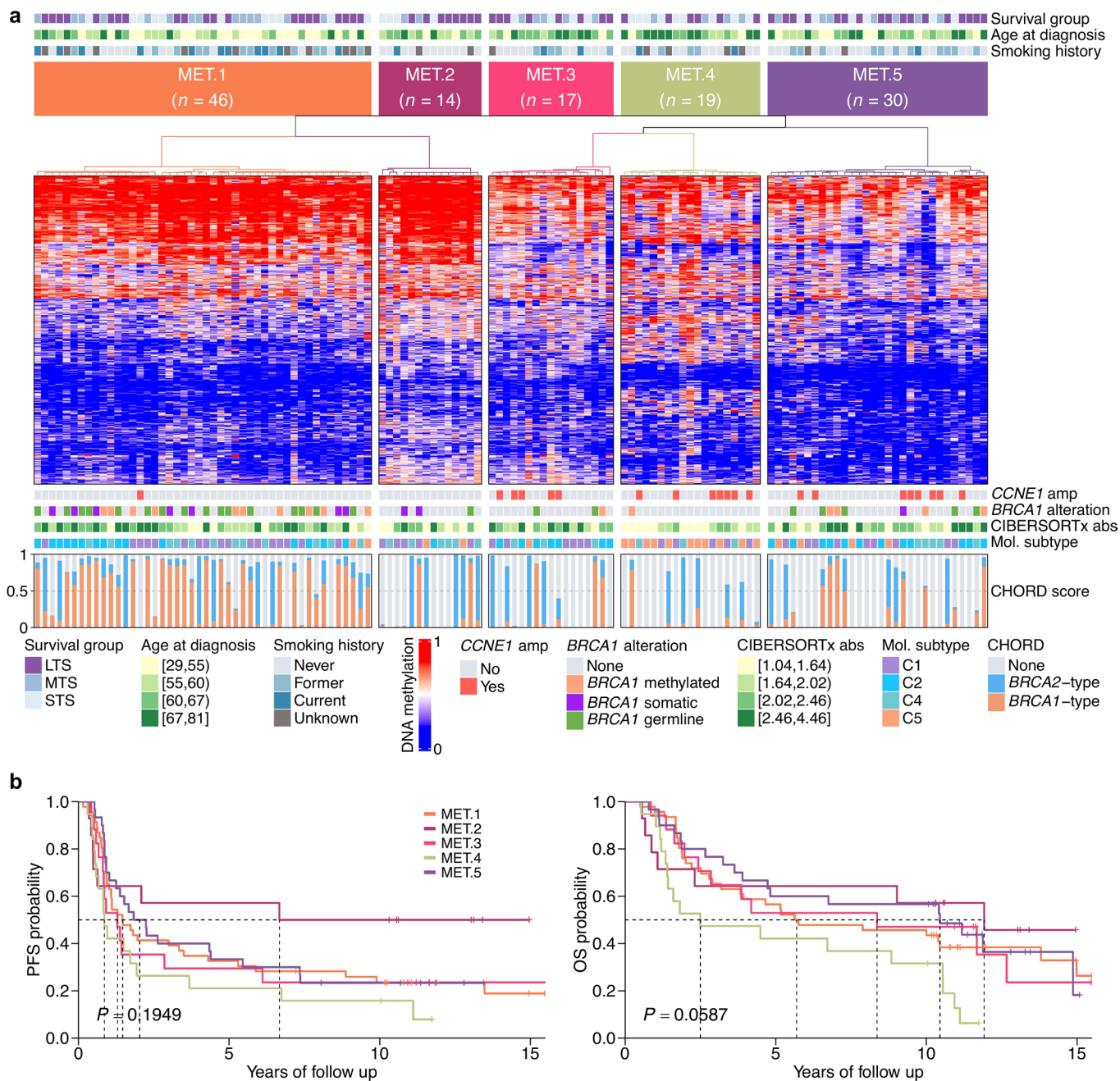
recombination (HR) status, homologous recombination deficiency (HRD) type, number of DNA repair pathway alterations, survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor), status at last follow-up (D, dead; P, progressed and alive; PF, progression-free and alive), self-reported smoking status, DeepCC molecular subtype (C1, mesenchymal; C2, immunoreactive; C4, differentiated; C5, proliferative), and neoadjuvant treatment (Y, yes; N, no).



Extended Data Fig. 5 | Clinical and genomic features of mutational signature clusters.

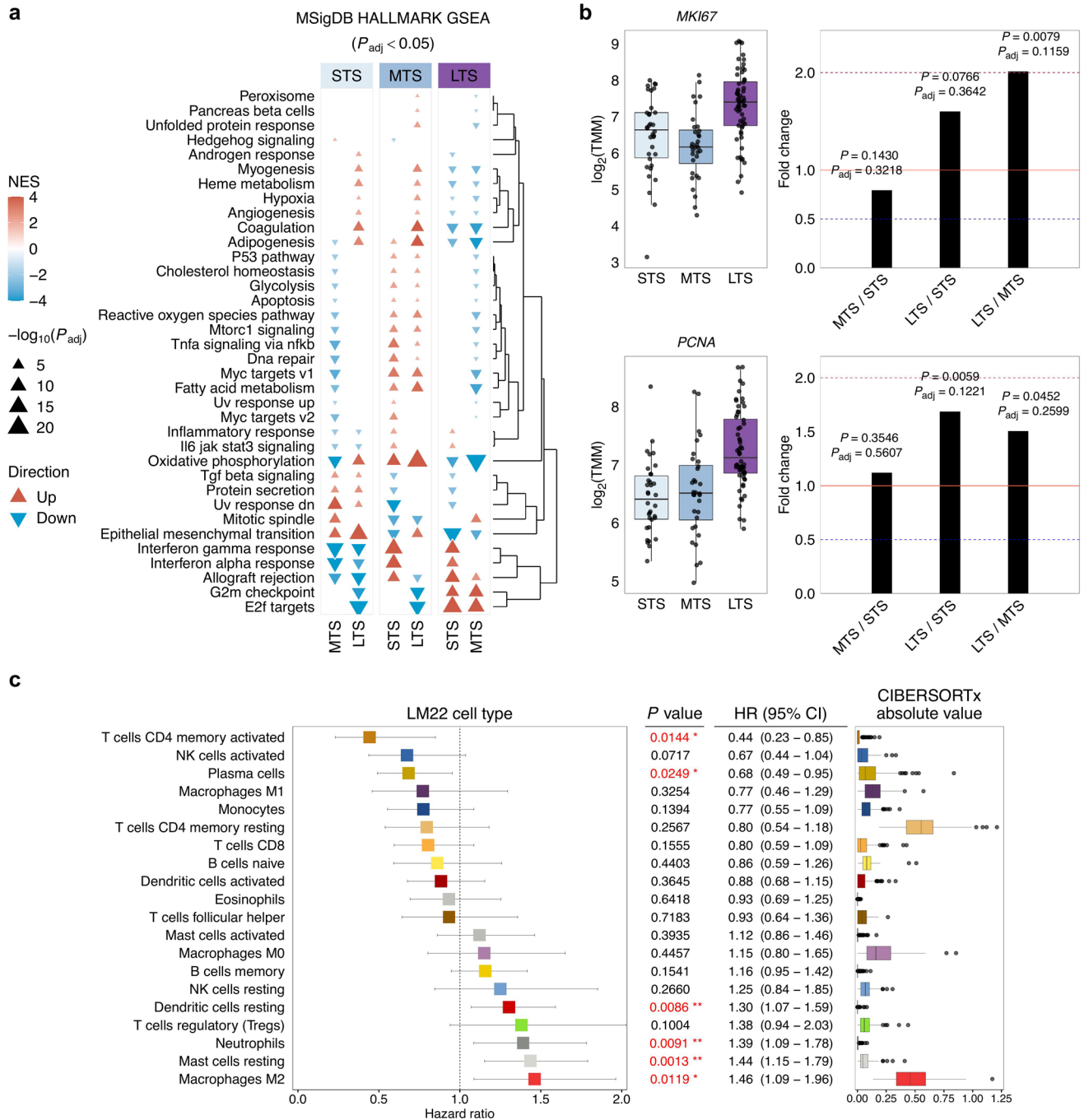
a, Boxplots summarize numerical, clinical and genomic features by mutational signature cluster; points represent each sample, boxes show the interquartile range (25–75th percentiles), central lines indicate the median, whiskers show the smallest/largest values within 1.5 times the interquartile range, red triangles indicate the mean, and dotted lines join the means of each cluster to visualize the trend. The Kruskal–Wallis test P values displayed are Benjamini–Hochberg adjusted P values. Features are ordered by their significance and clusters are ordered by the proportion of long-term survivors. CD8 scores were available for $n = 54$ primary tumors as previously measured by immunohistochemistry²³ and scored as density of CD8+ T cells (average cells/mm², y axis) in the tumor epithelium (TE). HRD, homologous recombination

deficiency; DEL, deletions; DUP; duplications; SV, structural variants; Mb, megabase; ITX, intrachromosomal rearrangements; LOH, loss-of-heterozygosity; INV, inversions. **b**, Bubble plot summary of mutational signature enrichment across signature clusters. The dendrogram is reused from the signature clustering (Fig. 3) to order the mutational signature types (columns). Mutational signature clusters (rows) are sorted by the proportion of long-term survivors in each cluster, indicated in brackets. The color and size of bubbles indicate the z-score scaled values of the mean signature exposure per cluster. Bubbles with a z-score of greater than or equal to 1 have a black border and bubbles with a z-score of greater than 0.5 but less than 1 have a gray border. Bordered bubbles have asterisks filled in to indicate Kruskal–Wallis test P values adjusted for multiple testing using Benjamini–Hochberg correction.



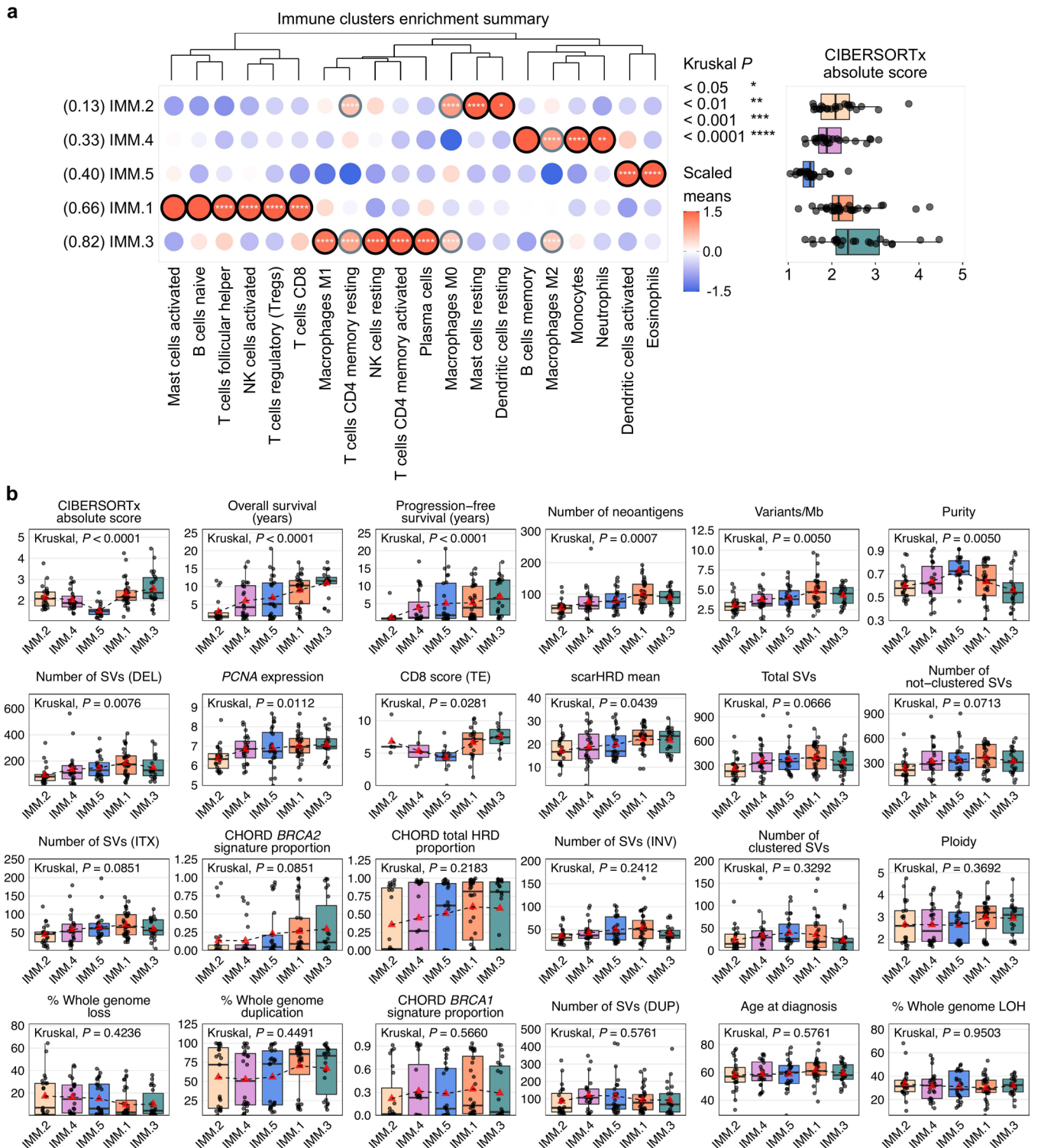
Extended Data Fig. 6 | DNA methylation clustering of primary tumor genomes. **a**, Heatmap of methylation data following consensus clustering of primary tumors (columns) based on the standardized CpG probe intensities (M-values) of the 1% most variable CpG probes (rows; number of probes = 3,645) across all primary tumor samples ($n = 126$). The heatmap scale shows the beta values. Five methylation clusters were identified (MET.1–MET.5), and each patient (column) is annotated with survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor), age at diagnosis (quartiles), and self-reported smoking history. Tumor samples are also classified

according to *CCNE1* amplification (amp) status, *BRCA1* alteration status, CIBERSORTx absolute (abs) immune scores (quartiles), and molecular subtype¹¹ (C1, mesenchymal; C2, immunoreactive; C4, differentiated; C5, proliferative). Bars in the bottom panel represent the *BRCA1* (orange) and *BRCA2* (blue) type homologous recombination deficiency (CHORD²⁵) scores of each tumor sample. **b**, Kaplan–Meier analysis of progression-free (PFS) and overall survival (OS) in patients stratified by methylation clusters. P values calculated by Mantel–Cox log-rank test and dotted lines indicate median survival in years since diagnosis.



Extended Data Fig. 7 | Transcriptional phenotypes in long-term survivors. **a**, Clustered heatmap summarizing gene set enrichment analysis (GSEA) using the hallmark Molecular Signatures Database (MSigDB) gene sets. Direction and color of triangles relate to the normalized enrichment score (NES) as generated by FGSEA. P values (two-sided) were calculated using the FGSEA default Monte Carlo method; the size of the triangles corresponds to the negative \log_{10} Benjamini-Hochberg adjusted P value (P_{adj}). The columns are split by survival groups (STS, short-term survivor; MTS, moderate-term survivor; LTS, long-term survivor), with the direction of enrichment denoted by the group in the heading (numerator) versus the two other groups labeled below. **b**, Boxplots summarize expression of *MKI67* and *PCNA* proliferation gene markers across the survival groups (left; STS $n = 34$, MTS $n = 32$, LTS $n = 60$); points represent each sample, boxes show the interquartile range (25–75th percentiles), central lines indicate the median, and whiskers show the smallest/largest values within 1.5

times the interquartile range. Differential expression analysis was performed using DESeq2 to determine fold change (right) of gene expression between survival groups (two-tailed Wald test, both unadjusted P values and Benjamini-Hochberg adjusted P values (P_{adj}) are shown). **c**, Forest plot (left) indicates the hazard ratio (HR, squares) and 95% confidence interval (CI; whiskers) for overall survival calculated using a univariate Cox proportional hazard regression model based on the LM22 immune cell types detected by CIBERSORTx analysis ($n = 126$ patients). Cell types are arranged by HR. P values < 0.05 are colored red ($*P < 0.05$, $**P < 0.01$) and were not adjusted for multiple comparisons. Absolute enrichment scores per cell type across the cohort are shown in boxplots (right); boxes show the interquartile range (25–75th percentiles), central lines indicate the median, whiskers show the smallest/largest values within 1.5 times the interquartile range and values outside it are shown as individual data points.

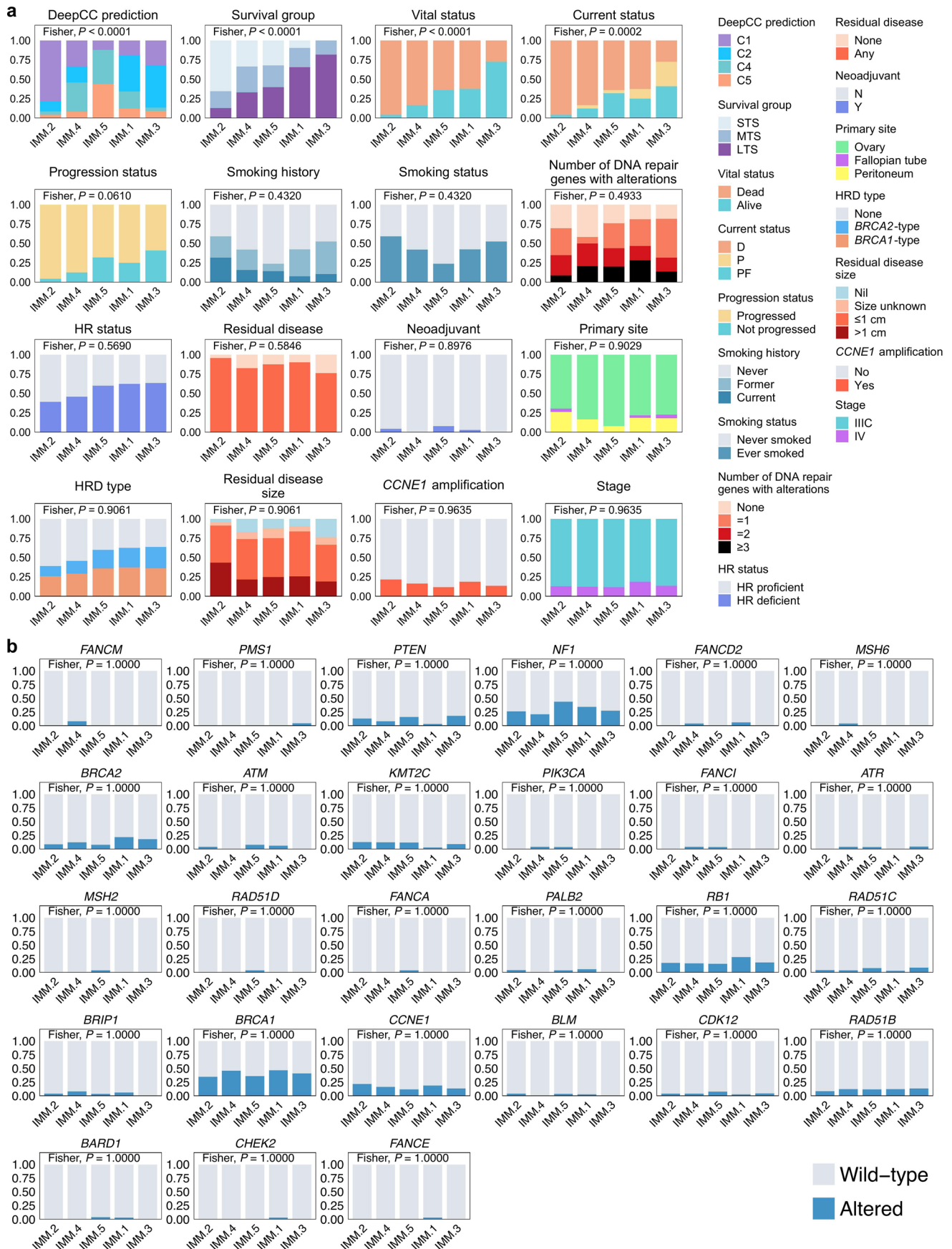


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Genomic and clinical features of immune clusters.

a, A condensed bubble plot of the various LM22 cell types used for the immune clustering (IMM.1 $n = 32$, IMM.2 $n = 23$, IMM.3 $n = 22$, IMM.4 $n = 24$, IMM.5 $n = 25$). The dendrogram is reused from the immune clustering (Fig. 5a) to order the cell types. Immune clusters (rows) are sorted by the proportion of long-term survivors indicated in brackets. The color and size of bubbles indicate z-score scaled values of the mean abundance of cell types per cluster. Bubbles with a z-score of greater than or equal to 1 have a black border, and those with a z-score of greater than 0.5 but less than 1 have a gray border. Asterisks indicate Kruskal–Wallis test P values adjusted for multiple testing using Benjamini–Hochberg correction. Boxplots (right) summarize CIBERSORTx absolute scores of each cluster; points represent each sample, boxes show the interquartile range (25–75th percentiles), central lines indicate the median, and whiskers show the smallest/largest values within 1.5 times the interquartile range. **b**, Boxplots

summarize numerical, clinical and genomic features by immune cluster (IMM.1 $n = 32$, IMM.2 $n = 23$, IMM.3 $n = 22$, IMM.4 $n = 24$, IMM.5 $n = 25$); points represent each sample, boxes show the interquartile range (25–75th percentiles), central lines indicate the median, whiskers show the smallest/largest values within 1.5 times the interquartile range, red triangles indicate the mean, and dotted lines join the means of each cluster to visualize the trend. The Kruskal–Wallis test P values displayed are Benjamini–Hochberg adjusted. Features are ordered by their significance and clusters are ordered by the proportion of long-term survivors. CD8 scores were available for $n = 54$ primary tumors as previously measured by immunohistochemistry²³ and scored as density of CD8+ T cells (average cells/mm², y axis) in the tumor epithelium (TE). HRD, homologous recombination deficiency; DEL, deletions; DUP, duplications; SV, structural variants; Mb, megabase; ITX, intrachromosomal rearrangements; LOH, loss-of-heterozygosity; INV, inversions.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Categorical features of immune clusters. a, Proportion of patients with categorical features per cluster. Features are ordered by significance using Fisher's exact test (two-sided) and the clusters are arranged by the proportion of long-term survivors. Features include homologous recombination (HR) status, homologous recombination deficiency (HRD) type, number of DNA repair pathway alterations, survival group (LTS, long-term survivor; MTS, moderate-term survivor; STS, short-term survivor), status at last follow-up (D, dead; P, progressed and alive; PF, progression-free and alive),

self-reported smoking status, DeepCC molecular subtype (C1, mesenchymal; C2, immunoreactive; C4, differentiated; C5, proliferative), and neoadjuvant treatment (Y, yes; N, no). **b**, Proportion of patients affected by gene alterations per immune cluster. Genes are ordered by significance using Fisher's exact test (two-sided) and clusters are ordered by the proportion of long-term survivors. The Fisher's test *P* values displayed in **(a)** and **(b)** are Benjamini-Hochberg adjusted *P* values.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

WGS

Paired-end Whole Genome Sequencing (WGS) at 150bp was performed on HiSeq X Ten System machines.

Software used for WGS processing:

- FASTQC (v0.11.8)
- FastQ Screen (v0.11.4)
- ea-utils (v1.05)
- BWA mem (v0.7.17-r1188)
- Picard Tools (v2.17.3)
- GATK (v4.0.10.1)
- GATK (v3.8-1-0-gf15c1c3ef)
- FACETS (v0.6.1)

RNA-seq

Libraries were generated using Illumina Stranded mRNA Prep and 150 bp paired-end sequencing was performed on Illumina NovaSeq 6000 instruments.

Software used for RNAseq processing:

- FASTQC (v0.11.8)
- FastQ Screen (v0.11.4)
- ea-utils (v1.05)
- STAR(v2.6.0b)
- Picard Tools (v2.17.3)
- RSeQC (v2.6.4)

Methylation

Tumor DNA was bisulfite converted with the EZ DNA Methylation kit (Zymo Research) and assayed using the Infinium MethylationEPIC

BeadChip arrays according to manufacturer's instructions (Illumina).

Software used for Methylation data processing:

- minfi82 955 (v1.32.0)
- sva (v3.34.0)

SNP array

Tumor and matched normal DNA was assayed with the Infinium OmniExpress-24 BeadChip arrays, arrays scanned and data processed using Genotyping module 2.0.3 software in GenomeStudio 2.0.3 to calculate logR ratios and B-allele frequencies according to manufacturer's instructions (Illumina).

Software used for SNP array data processing:

- HYSYS github version b730498 (<https://github.com/PapenfussLab/HaveYouSwappedYourSamples>)
- qPure v1.1 (<https://sourceforge.net/projects/qpure/>)
- ASCAT v2.5.2 (<https://github.com/VanLoo-lab/ascats>)

Data analysis

WGS

- VarDictJava (v1.5.7 with `-r=2 -Q=10 -V =0.05 -f=0.01`)
- Mutect2 (v4.0.11.0 with defaults)
- Strelka (v2.9.9 with defaults)
- VarScan2 (v2.4.3 with `--min-coverage 7 --min-var-freq 0.05 801 --min-freq-for-hom 0.75 --p-value 0.99 --somatic-p-value 0.05 --strand-filter 0 & SAMtools v1.9 for mpileup`)
- vt (v0.57721)
- GATK ReadBackPhasing (v3.8-1-0-gf15c1c3ef with `--phaseQualityThresh 10 --enableMergePhasedSegregatingPolymorphismsToMNP --min_base_quality_score 10 min_mapping_quality_score 10 --maxGenomicDistanceForMNP 2`)
- GATK CombineVariants (v3.8-1-0-gf15c1c3ef with `-genotypeMergeOptions UNIQUIFY -priority Strelka2, Mutect2, VarScan2, VarDictJava`)
- GATK VariantAnnotator (v3.8-1-0-gf15c1c3ef with `--reference_window_stop 1000 -A HomopolymerRun -A TandemRepeatAnnotator`)
- BCFtools (v1.9)
- bam-readcount (v0.8.0 with `-w 0 --min-mapping-quality 10 --min-base-quality 10 --max-count 100,000,000`)
- Ensembl Variant Effect Predictor (VEP v92.4)
- Manta + BreakPointInspector (v1.5.0)
- GRIDSS (v2.0.1)
- Smoove (v0.2.2)
- SvABA (v134)
- StructuralVariantAnnotation (v1.3.1 with `maxgap = 10, ignore.strand = FALSE`)
- InteractionSet (v1.14.0)
- rtracklayer (v1.46.0)
- snp-pileup (v1.0 with `--pseudo-snps 100 --min-map830 quality 10 --min-base-quality 10 --max-depth 5000 --min-read-counts 15,0`)
- cnv_facets (v0.13.0 with `--nbhd snp=500 --cval=50 1000 --depth=15 5000`)
- CHORD (v2.00)
- csaw (v1.20.0)
- ICAMS v2.0.10.9001
- rstatix (v0.7.0)
- signature.tools.lib (v0.0.0.9000 with `SignatureFit_withBootstrap [method = "KLD", nboot = 100, randomSeed = 42, threshold_percent = 2, threshold_p.value = 0.05]`)
- ConsensusClusterPlus (v1.50.0 with `maxK = 10, reps=1000, pltem=0.9, pFeature=0.9, clusterAlg="pam", distance="pearson", innerLinkage="ward.D2", finalLinkage="ward.D2", seed=12345678`)
- seriation (v1.3.0 with `method="OLO"`)
- HLA-VBSeq (v11_22_2018)
- Jvarkit samviewwithmate (ec2c236)
- samtools (v1.9)
- deepTools (v3.0.0 with parameters `--binSize 10 --minMappingQuality 10 --normalizeUsing CPM --skipNonCoveredRegions --samFlagExclude 1024 --outFileFormat bigwig`)
- pVACTools pVACseq (v1.3.5)
- dNdScv (v0.0.1.0 with `refdb = "hg19", sm = "192r_3w", max_muts_per_gene_per_sample = Inf, max_coding_muts_per_sample = Inf, use_indel_sites= FALSE`)
- GRIN (v1.4)
- GISTIC2 (v2.0.23 with `-savegene 1 -maxspace 1000 -ta 0.1 -td 0.1 -rx 0 -cap 3 -broad 0 -twoside 1 -res 0.05 -genegistic 0 -v 10`)
- maftools (v2.2.10)
- Mutalyzer (v2.0.35)
- R (v3.6.3)
- Prism (v9.2.0)

RNA-seq

- HTSeq (v0.10.0 with `mode = "intersection-nonempty"`)
- edgeR (v3.28.1)
- limma (v3.42.2)
- DeepCC (v0.1.1)
- DESeq2 (v1.26.0)
- fgSEA (v1.15.1)
- CIBERSORTx web version as at 05/21/2020 (<https://cibersortx.stanford.edu/>)
- ConsensusClusterPlus (v1.50.0)
- Arriba (v1.1.0)
- R (v3.6.3)

Methylation

- limma (v3.42.2)

- R (v3.6.3)

Feature correlation
- polycor (v0.8-1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

ICGC dataset: Previously published WGS and RNA-seq data generated as part of the ICGC Ovarian Cancer project¹⁴ are available from the European Genome-phenome Archive (EGA) repository (<https://ega-archive.org>) as a single bam file for each sample type (tumor/normal), under the accession code EGAD00001000877 (<https://ega-archive.org/datasets/EGAD00001000877>). Due to the sensitive nature of these patient datasets, access is subject to approval from the ICGC Data Access Compliance Office (<https://docs.icgc.org/download/data-access/>), an independent body who authorizes controlled access to ICGC sequencing data. ICGC SNP array and methylation data sets have been deposited into the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession code GSE65821 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65821>), without access restrictions. ICGC gene count level transcriptomic data has been deposited into the GEO under accession code GSE209964 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209964>).

MOCOG dataset: WGS, RNA-seq and SNP array data from long-term survivors generated as part of the MOCOG study have been deposited in the EGA repository under accession code EGAS00001005984 (<https://ega-archive.org/studies/EGAS00001005984>). WGS and RNA-seq data are available as raw FASTQ files for each sample type (tumor/normal) and SNP array data are available as raw signal intensity files in text format for each sample type (tumor/normal). Access to patient sequence data can be gained for academic use through application to the independent Data Access Committee (dac@petermac.org). Responses to data requests will be provided within two weeks. Information on how to apply for access is available at the EGA under accession code EGAS00001005984. The MOCOG cohort raw methylation data sets have been submitted to the GEO under accession code GSE211687 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211687>), with no access restrictions.

Uniformly processed somatic variant data from the ICGC and MOCOG cohorts is deposited in Synapse under accession code syn34616347 (<https://www.synapse.org/#!Synapse:syn34616347>), and processed expression and methylation data from both cohorts has been submitted into the GEO under accession code GSE211687 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211687>), without access restrictions.

Population frequencies of genetic variants can be accessed via the Genome Aggregation Database (gnomAD) at <https://gnomad.broadinstitute.org/>. Supporting evidence for pathogenicity of genomic alterations can be accessed via ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), BRCA Exchange (<https://brcaexchange.org/>) and the TP53 Database (<https://tp53.isb-cgc.org/>). The Ensembl ranked order of severity of variant consequences is available at: https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html. Precomputed TCGA ovarian serous cystadenocarcinoma survival analysis data can be downloaded from OncoLnc (<http://www.oncolnc.org/>). Mutational signature reference databases can be accessed via COSMIC (<https://cancer.sanger.ac.uk/signatures/>) and Signal (<https://signal.mutationalsignatures.com/>). The LM22 signature matrix used for immune cell deconvolution can be downloaded here: <https://cibersortx.stanford.edu/>. The COSMIC Cancer Gene Census can be accessed here: <https://cancer.sanger.ac.uk/census>. MSigDB hallmark gene sets can be accessed here: <https://www.gsea-msigdb.org/gsea/msigdb/>. Illumina methylation probes that were filtered out due to poor performance (e.g. cross reactive or non-specific probes) can be found here: https://github.com/sirselim/illumina450k_filtering. Germline polymorphic sites for reference and variant allele read counts used in FACETS analysis can be found at ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/common_all_20180423.vcf.gz. The GTF used for annotation and RNA-seq counts is available here: <ftp://ftp.ensembl.org/pub/grch37/release-92/>. All other data are available within the article and its supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. A limited number of cases were available given the rarity of high-grade serous ovarian cancer long-term survivors, with fresh frozen tumor tissue and matched germline samples available.
Data exclusions	Two tumor samples were flagged as outliers with a high somatic mutation rate (>20 mutations/Mb): AOCs-076 due to cross sample contamination and AOCs-166 due to a germline mutation in the mismatch repair gene PMS2. The two samples were therefore excluded from further analyses.
Replication	Due to the rarity of tumor tissue from long-term survivors and the high cost of genomic assays, WGS, RNA-sequencing, SNP arrays and methylation arrays were performed once for each sample. Supporting evidence for variants detected by WGS were sought in orthogonal datasets, such as SNP array, clinical sequencing, RNA-sequencing or DNA sequencing of matched samples (e.g. germline and tumor) from the same patient (Supplementary Tables 3, 4 and 5). CCNE1 amplification status determined by WGS data was verified by FISH and immunohistochemistry data available from a previous study (Aziz et al 2018). CIBERSORTx immune scores were verified with CD8+ T cell scores determined by multi-color immunohistochemistry and automated image scoring from our previous study (Garsed et al 2018), which confirmed the identification of the two long-term survival immune clusters (IMM.1 and IMM3). Mutational signature analyses were informed

by and/or consistent with previous analyses of breast, ovarian and pan-cancer genomic datasets (Alexandrov et al 2013, Patch et al 2015, Popova et al 2016, Alexandrov et al 2020, and Degasperi et al 2020). DNA methylation clustering identified a BRCA1-altered subset (MET.1) that overlapped with the poor outcome BRCA1 mutational signature subgroup (SIG.5), providing an independent marker of this group. To our knowledge, this is the first WGS, DNA methylation and RNA-seq analysis of HGSC long-term survivors, so there is currently no equivalent dataset available in which to validate the genomic and immune signatures, however where possible we sought validation of findings in an independent dataset (The Cancer Genome Atlas Research Network et al 2011).

Randomization Randomization was not applicable as this study did not involve an intervention. Long-term survivors were compared with patients with similar known survival predictors: histology, grade, age at diagnosis, stage, treatment, etc. (Extended Data Fig. 1b).

Blinding Blinding was not undertaken. Genomic and clinical data was uniformly processed and analyzed across all samples/cases.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Patients were diagnosed with advanced stage (FIGO Stage IIIC/IV), high-grade serous cancer (ovarian, fallopian tube or primary peritoneal carcinoma) confirmed by histopathology review, received primary chemotherapy incorporating a platinum-based compound, and met survival group criteria; either short-term survivors (death within 2 years); moderate-term survivors (death ≥ 2 and < 10 years); or long-term survivors (overall survival ≥ 10 years). Population characteristics are provided in Extended Data Fig. 1 and Supplementary Table 1. All 126 patients are female and were aged 29-81 years.

Recruitment

Patients were recruited to the population-based Australian Ovarian Cancer Study (AOCS), Peter MacCallum Cancer Centre (Melbourne, Australia), the Gynaecological Oncology Biobank (GynBiobank), Westmead Hospital (Sydney, Australia) or the Mayo Clinic (Rochester, USA) under Research Ethics Committee approval for each study. Written informed consent was provided by all patients. Participation was voluntary, with no compensation provided to participants.

1. Patients over the age of 18 (and under 80 for AOCS) with invasive epithelial ovarian cancer (or primary peritoneal or fallopian tube cancer) were recruited through major cancer treatment centers across Australia (AOCS and GynBiobank) and in the USA (the Mayo Clinic), generally at the time of initial diagnosis.
2. Surgical resection samples: Tissue samples for research were taken from tissue removed during surgery and were excess to diagnostic requirements. Sampling of surgical tissue was undertaken by a trained pathologist to ensure that samples taken for the purposes of research did not affect histopathological assessment. Samples were snap-frozen and stored for future research, with patient consent.
3. Surgical, pathology, systemic treatment and clinical outcome data was collected from medical records during regular longitudinal clinical follow-up. Cases were selected for the current study based on pre-specified criteria, including advanced stage disease (FIGO Stage IIIC/IV), histology (HGSC), primary treatment (platinum-based chemotherapy), availability of biospecimens for analysis (snap frozen tumor tissue and blood), and survival within the three specified survival groups (death within 2 years; death ≥ 2 and < 10 years; overall survival ≥ 10 years). This may represent a potential bias towards patients with tumors of sufficient size, and a high proportion of tumor cells within the research sample, to enable analyses. AOCS recruited patients < 80 years old, which could result in an increased proportion of long-term survivors compared to studies without an upper age limit, because the chance of surviving > 10 years if diagnosed over 80 would be lower (due to older age), however ages are consistent between survival groups and not likely to impact the results. No other sources of selection bias were identified.

Ethics oversight

The Peter MacCallum Cancer Centre Human Research Ethics Committee (HREC), the Western Sydney Local Health District HREC, and the Mayo Clinic Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.