nature genetics

Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes

Yi Kan Wang[1,15], Ali Bashashati[1,15], Michael S Anglesio[2,3], Dawn R Cochrane[1], Diljot S Grewal[1,2], Gavin Ha[1,14], Andrew McPherson[1,2], Hugo M Horlings[1], Janine Senz[1], Leah M Prentice[1], Anthony N Karnezis[2], Daniel Lai[1], Mohamed R Aniba[1], Allen W Zhang[1,4,5], Karey Shumansky[1], Celia Siu[1], Adrian Wan[1], Melissa K McConechy[2], Hector Li-Chang[2], Alicia Tone[2,14], Diane Provencher[6–8], Manon de Ladurantaye[6,7], Hubert Fleury[6,7], Aikou Okamoto[9], Satoshi Yanagida[9], Nozomu Yanaihara[9], Misato Saito[9], Andrew J Mungall[10], Richard Moore[10], Marco A Marra[10,11], C Blake Gilks[2,12], Anne-Marie Mes-Masson[6,7,13], Jessica N McAlpine[3], Samuel Aparicio[1,2], David G Huntsman[1–3] & Sohrab P Shah[1,2,10]

We studied the whole-genome point mutation and structural variation patterns of 133 tumors (59 high-grade serous (HGSC), 35 clear cell (CCOC), 29 endometrioid (ENOC), and 10 adult granulosa cell (GCT)) as a substrate for class discovery in ovarian cancer. *Ab initio* clustering of integrated point mutation and structural variation signatures identified seven subgroups both between and within histotypes. Prevalence of foldback inversions identified a prognostically significant HGSC group associated with inferior survival. This finding was recapitulated in two independent cohorts (*n* = 576 cases), transcending *BRCA1* and *BRCA2* mutation and gene expression features of HGSC. CCOC cancers grouped according to APOBEC deamination (26%) and age-related mutational signatures (40%). ENOCs were divided by cases with microsatellite instability (28%), with a distinct mismatch-repair mutation signature. Taken together, our work establishes the potency of the somatic genome, reflective of diverse DNA repair deficiencies, to stratify ovarian cancers into distinct biological strata within the major histotypes.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

The major histotypes of ovarian cancer have distinct cellular morphologies and etiologies, as well as molecular, genetic, and clinical attributes[1–5]. HGSC accounts for 70% of epithelial ovarian carcinomas and for 90% of advanced-stage disease and mortality. Endometriosis-associated cancers account for approximately 20% of epithelial ovarian carcinomas, including ENOC and CCOC[6,7]. The major histotypes associate with distinct sets of recurrently mutated genes and aberrant mechanisms of DNA repair: *TP53* loss and profound genomic instability due to *BRCA1* and *BRCA2* defects are ubiquitous in HGSC[4,8,9]; CCOC and ENOC tumors harbor *ARID1A* loss-of-function mutations (occurring in approximately 50% and 30% of cases, respectively)[10,11] variously accompanied by loss of *PTEN*, mutation of *KRAS*, *CTNNB1*, *PIK3CA*, *PPP2R1A*, and *TERT* promoters[12–19], and additional properties such as microsatellite instability (MSI), mismatch-repair deficiency, and hypermutation[20]. Adult GCTs are rare (4–5% of all ovarian cases), phenotypically distinct non-epithelial ovarian carcinomas, unambiguously identified through a *FOXL2* mutation encoding p.Cys134Trp[21].

Although these histotypes are viewed as distinct diseases, they are still usually treated with surgery and combination platinum and taxane chemotherapy[22–27]. The use of PARP inhibitors in BRCA-deficient HGSC cancers represents both the first histotype-specific treatment and the first successful attempt to further stratify histotypes by meaningful treatment options[28–31]. However, in complex disease phenotypes, gene-based biomarkers offer limited representations of underlying biology and can be complemented by more global properties.

Whole-genome sequencing has shed light on intrinsic and extrinsic mutational processes via the analysis of substitution patterns and the localized sequence context of mutations[32,33]. Complementary insights have been gained from analysis of structural variation (SV) patterns reflective of double-strand break repair mechanisms operating in various tumor types exhibiting genomic instability[34,35], including

[1]Department of Molecular Oncology, BC Cancer Agency, Vancouver, British Columbia, Canada. [2]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. [3]Department of Gynecology and Obstetrics, University of British Columbia, Vancouver, British Columbia, Canada. [4]Graduate Bioinformatics Training Program, University of British Columbia, Vancouver, British Columbia, Canada. [5]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Vancouver, British Columbia, Canada. [6]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montreal, Quebec, Canada. [7]Institut du Cancer de Montréal, Montreal, Quebec, Canada. [8]Division of Gynecologic Oncology, Université de Montréal, Montreal, Quebec, Canada. [9]Department of Obstetrics and Gynecology, Jikei University School of Medicine, Tokyo, Japan. [10]Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. [11]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [12]Department of Pathology, Vancouver General Hospital, Vancouver, British Columbia, Canada. [13]Department of Medicine, Université de Montréal, Montreal, Quebec, Canada. [14]Present addresses: Dana-Farber Cancer Institute and the Broad Institute, Boston, Massachusetts, USA (G.H.) and Division of Gynecologic Oncology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada (A.T.). [15]These authors contributed equally to this work. Correspondence should be addressed to D.G.H. (dhuntsma@bccancer.bc.ca) or S.P.S. (sshah@bccrc.ca).

Received 13 March; accepted 28 March; published online 24 April 2017; doi:10.1038/ng.3849

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

patterns of evolution in HGSC[36]. The most common SVs include tandem duplication resulting from insertion of an adjacent identical segment, foldback inversion–forming localized inverted duplication caused by a breakage fusion bridge[34], interstitial deletion in which the ends of multiple breaks in a chromosome are rejoined with a segment removed, and interchromosomal translocations where the break ends are on different chromosomes. The relative proportions of structural alterations attributed to tandem duplication, foldback inversion, interstitial deletion, and other interchromosomal translocations provide context as a readout of specific DNA repair mechanisms operating in human cancers[37–39].

As such, the cancer genome is itself reflective of phenotypic properties. In the absence of complete genome sequences across the spectrum of ovarian cancer histotypes, the relative patterns of both mutation and SV signatures have not been elucidated. We asked how the consequences of aberrant DNA repair are reflected in the features of somatic genomes, and how these features could be used to further segregate patients into distinct molecular and biological classes. We posited that the genomic features of ovarian cancers could provide a substrate for class discovery beyond single-gene mutation status or expression profiles. We investigated the capacity of genomic characteristics inferred as consequences of aberrant DNA repair processes within and between HGSC, ENOC, CCOC, and GCT cancers to provide new substructure(s) for these diseases. Our results show that properties of the somatic genome on both the structural and point mutation scales are powerful, discriminant biomarkers between and within histotypes for subgroup discovery in ovarian cancer. In particular, we show across internal discovery and external test and validation HGSC cohorts (total $n = 576$) that a pattern of foldback inversions associated with high-level amplifications (HLAMPs) stratifies patients into prognostically significant groups, superseding BRCA mutation status and gene expression biomarkers. In aggregate, our findings present a novel stratification of ovarian cancer diseases, dividing the classical histotypes and identifying new biological strata that may inform contemporary and future therapeutic opportunities.

RESULTS

Cohort construction, clinical data, and sequencing

One hundred and thirty-three patients with ovarian cancer with histologically confirmed HGSC ($n = 59$), CCOC ($n = 35$), ENOC ($n = 29$), and GCT ($n = 10$) were included in this study (**Supplementary Table 1**). Clinical follow-up data (including overall and progression-free survival) for HGSC, ENOC, and CCOC cases were also recorded. *BRCA1* promoter methylation and *BRCA1* and *BRCA2* germline mutation status were determined for all patients with HGSC through hereditary cancer screening programs. MSI testing performed on all tumor DNA samples using five repeated loci confirmed MSI in 28% of ENOC cases ($n = 8$) (**Supplementary Table 1**), while MSI was low or absent in all other cases.

Tumor and matched normal DNA samples from each patient were subjected to whole-genome sequencing with median coverage of 51× and 37× for the tumor and normal samples, respectively. Somatic alterations on all scales were identified in the tumor genome of each case, including single-nucleotide variants (SNVs), small indels, copy number alterations (CNAs), and SVs (revalidation through PCR-based targeted amplicon sequencing was performed for selected SNVs and SVs; **Supplementary Tables 2** and **3**). Wide variation both within and between histotypes was observed for all event types (**Supplementary Fig. 1** and **Supplementary Table 4a**). For example, the number of SNV events (**Supplementary Fig. 2a** and **Supplementary Table 4b**) was distributed as follows: GCT: 568–2,471 per tumor, median 1,433; CCOC: 481–18,268 per tumor, median 2,693; ENOC: 1,165–596,135 per tumor, median 3,928; HGSC: 1,595–16,058 per tumor, median 4,812). We consequently investigated whether global patterns of somatic genomic variation could be exploited to stratify cases.

Within- and between-histotype stratification

We computed 20 genomic features for each case from somatic SNVs, indels, CNAs, and SVs, including 6 previously described[32] mutation signatures (**Supplementary Fig. 2b**,**c**), 4 additional SNV/indel properties, 7 SV features, and 3 CNA properties (detailed descriptions of the 20 features are provided in **Supplementary Fig. 3**). We determined the consensus coefficients ($C_{S.BC}$, $C_{S.MMR}$, etc.) for the six mutation signatures (described below) required to explain the SNV mutational repertoire in each sample (**Supplementary Fig. 1a**). CNA features were inferred as the proportion of the genome affected by amplification, deletion, and loss of heterozygosity (LOH). SV features were determined by the relative proportion of balanced rearrangement, deletion rearrangement, tandem duplication, foldback inversion, inversion, and unbalanced rearrangement among all SVs for each case.

Ab initio hierarchical clustering of the 133 patients with ovarian cancer based on the 20 genomic features (**Supplementary Table 5a**) identified seven major subgroups (**Fig. 1a**, **Supplementary Fig. 4a**,**b**, and **Supplementary Table 6**), with the optimal number of groups determined through marginal gain of explained variance (**Supplementary Fig. 4c** and **Supplementary Table 5b**). Linear discriminant analysis identified the dominant feature(s) distinguishing the tumors in each group (**Fig. 1b** and **Supplementary Table 7a**). Each of the three main histotypes stratified into subgroups, while all GCTs were grouped together (**Fig. 1c** and **Supplementary Table 7b**). Similar results were obtained whether clustering was performed on all histotypes together or HGSC and endometriosis-associated samples (CCOC and ENOC) were clustered independently (**Supplementary Figs. 5** and **6**). The seven groups were characterized as follows: G-BC: GCT tumors with mutation signature S.BC (associated with breast cancer and medulloblastoma); E-MSI: MSI ENOC tumors characterized by mutation signature S.MMR (reflective of mismatch-repair deficiency); Mixture: HGSC, CCOC, and ENOC cases without obvious discriminant features; C-APOBEC: CCOC cases characterized by mutation signature S.APOBEC (attributed to activity of the AID/APOBEC family of cytidine deaminases); C-AGE: CCOC cases characterized by mutation signature S.AGE (associated with age at diagnosis); H-FBI: HGSC cases with high prevalence of foldback inversion SVs; H-HRD: HGSCs with prevalence of duplications or deletion rearrangements and mutation signature S.HRD (reflective of homologous recombination deficiency).

A subgroup of HGSC characterized by foldback inversions

Two subgroups comprised primarily HGSC cases: H-FBI ($n = 24$, 41% of HGSC cases) and H-HRD ($n = 31$, 53% of HGSC cases; **Fig. 1c**). Relative importance analysis of H-FBI and H-HRD highlighted SV patterns, dominated by foldback inversions, as discriminant features between the HGSC groups; this was corroborated by a statistically higher proportion of foldback inversions over all SVs in H-FBI relative to H-HRD (mean 0.12 versus 0.04 of SVs; Student's *t* test, adjusted $P < 0.0001$; **Fig. 2a**). Further differences were found in mutation signatures: H-HRD exhibited higher $C_{S.HRD}$ than H-FBI (mean 0.42 versus 0.29; Student's *t* test, adjusted $P < 0.0001$) while $C_{S.AGE}$ was higher in H-FBI (mean 0.25 versus 0.06 in H-HRD; Student's *t* test, adjusted $P < 0.0001$; **Supplementary Table 8a**), and H-HRD exhibited a higher overall mutation load (**Supplementary Fig. 4d**).

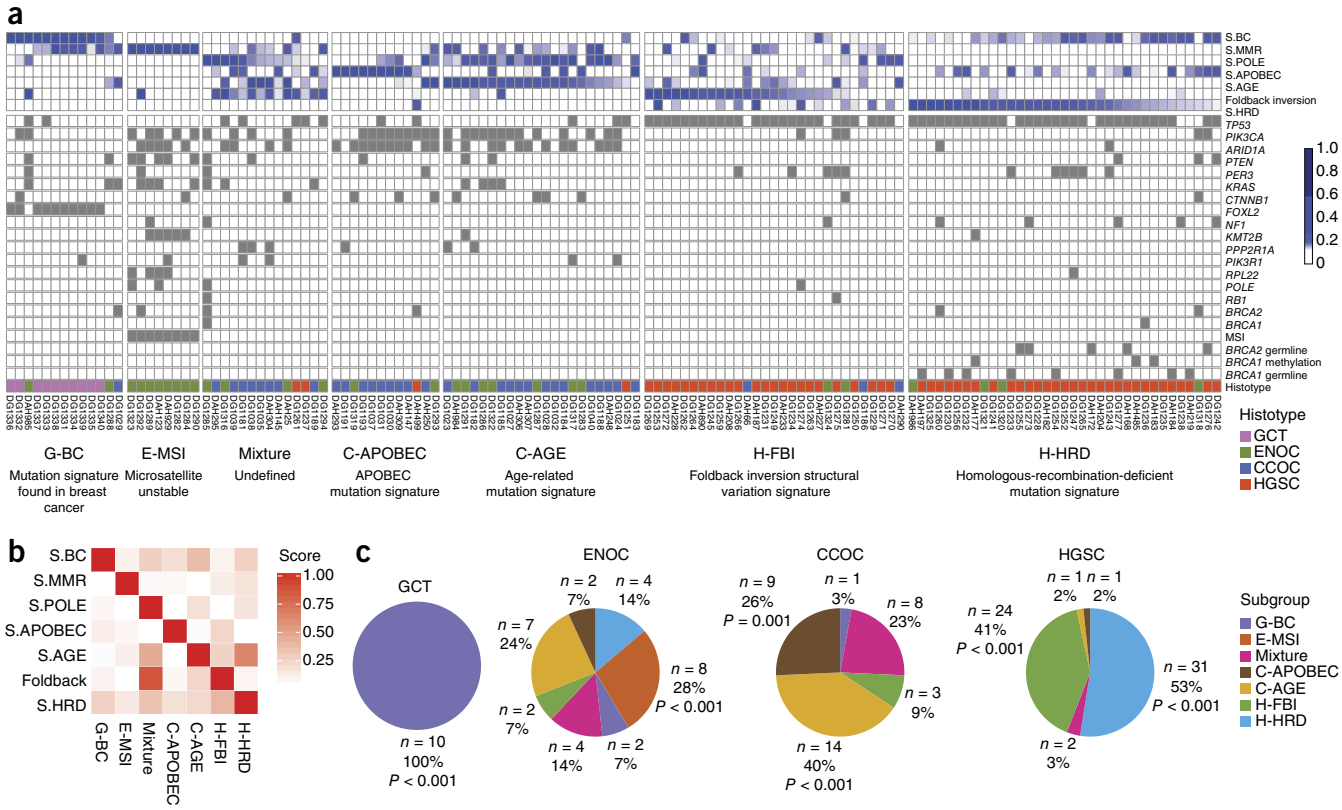

Figure 1 Integration of genomic features stratifies ovarian cancer histotypes. (**a**) Hierarchical clustering of 133 patients with ovarian cancer (columns) by integrating genomic features, including point mutation, copy number, and structural rearrangement profiles. The contributions of the seven dominant genomic features defining each cluster are shown (rows at top); box color corresponds to the quantile-normalized value for each genomic feature. Mutation status (presence, gray; absence, white) for significantly mutated genes (MutSigCV $q < 0.01$; **Supplementary Table 6**) and DNA repair genes across patients is displayed (bottom). Histotype is included as an annotation row (red, HGSC; blue, COCC; green, ENOC; purple, GCT). (**b**) Discriminant features defining each subgroup. The heat map shows the normalized t score for the discriminant features in each subgroup. (**c**) Contribution of genomic subgroup memberships to each histotype. The number (n) and proportion (%) of samples from each subgroup are shown. Enrichment of cases (per histotype) in subgroups was assessed using Fisher's exact test (corresponding adjusted P values are shown for significant enrichment, $P < 0.01$).

Gene-based co-association identified tumors harboring *BRCA1* somatic ($n = 1$) or germline ($n = 8$) mutations, methylation of the *BRCA1* promoter ($n = 3$), and *BRCA2* somatic ($n = 1$) or germline ($n = 5$) mutations in the H-HRD group (**Fig. 1**). Analysis of a set of 61 genes related to homologous recombination (HR) pathways identified variants in 31 HR-related genes (point mutations in 18 genes and gene breakage in 20 genes; **Supplementary Table 9a**,**b**), including group-specific *RAD51B* gene breakage in H-FBI cases (7/24, 29%) and *NF1* disruption in H-HRD cases (7/31, 23%) (**Supplementary Table 9c**). Statistically over-represented focal copy number amplification of *CCNE1* (19q21) was observed in H-FBI, whereas H-HRD implicated *MECOM* (3q26.2), *MYC* (8q24.21), and *CCND1* (11q13.3) as over-represented regions of focal copy number amplification (**Fig. 2b** and **Supplementary Table 10a**,**b**). Focal loss in *PTEN* (10q23.31) was uniquely identified in H-FBI, and *RB1* focal copy number deletion (false discovery rate (FDR) $q < 0.0001$, $n = 29$, 94% of samples) was significant in H-HRD (**Fig. 2c**, **Supplementary Fig. 7**, and **Supplementary Table 10a**,**b**). In aggregate, these results identify distinct biological subgroups within HGSC, separated by mutational processes, in particular foldback inversion events. Co-associated genomic disruptions of specific genes and pathways could implicate divergent biological properties for H-FBI and H-HRD cases.

Inferior prognosis in HGSC informed by foldback inversions

We next compared the survival response to uniform platinum-based chemotherapy in the two HGSC subgroups. The overall and progression-free survival rates of the H-FBI subgroup were poor relative to those of the H-HRD subgroup (log-rank test, $P = 0.0053$ and 0.0232; **Fig. 2d**). Outcome associations were independent of BRCA status, which was established by excluding the 15 cases with *BRCA1* or *BRCA2* somatic or germline mutations and recomputing the survival curves (log-rank test, $P = 0.024$ and 0.037; **Fig. 2e**). We next isolated the foldback inversion proportion as a feature and stratified cases into two groups (High FBI and Low FBI) on the basis of the median value. Both overall and progression-free survival were significantly worse for the High FBI group, establishing that foldback inversion proportion could be used as a single prognostic feature (log-rank test, $P = 0.0187$ and 0.0286; **Supplementary Fig. 8a**). We then validated this approach in an external cohort from the International Cancer Genome Consortium (ICGC)[40]. The 82 ICGC primary HGSC tumors were stratified into two equal-sized groups according to the prevalence of foldback inversions (**Supplementary Table 11a**,**b**). We note that the ICGC samples were sequenced independently, SVs were predicted with an orthogonal computational method, and the cohort was not controlled for uniform treatment protocols. Despite these variations, the ICGC group with high prevalence of foldback inversions

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

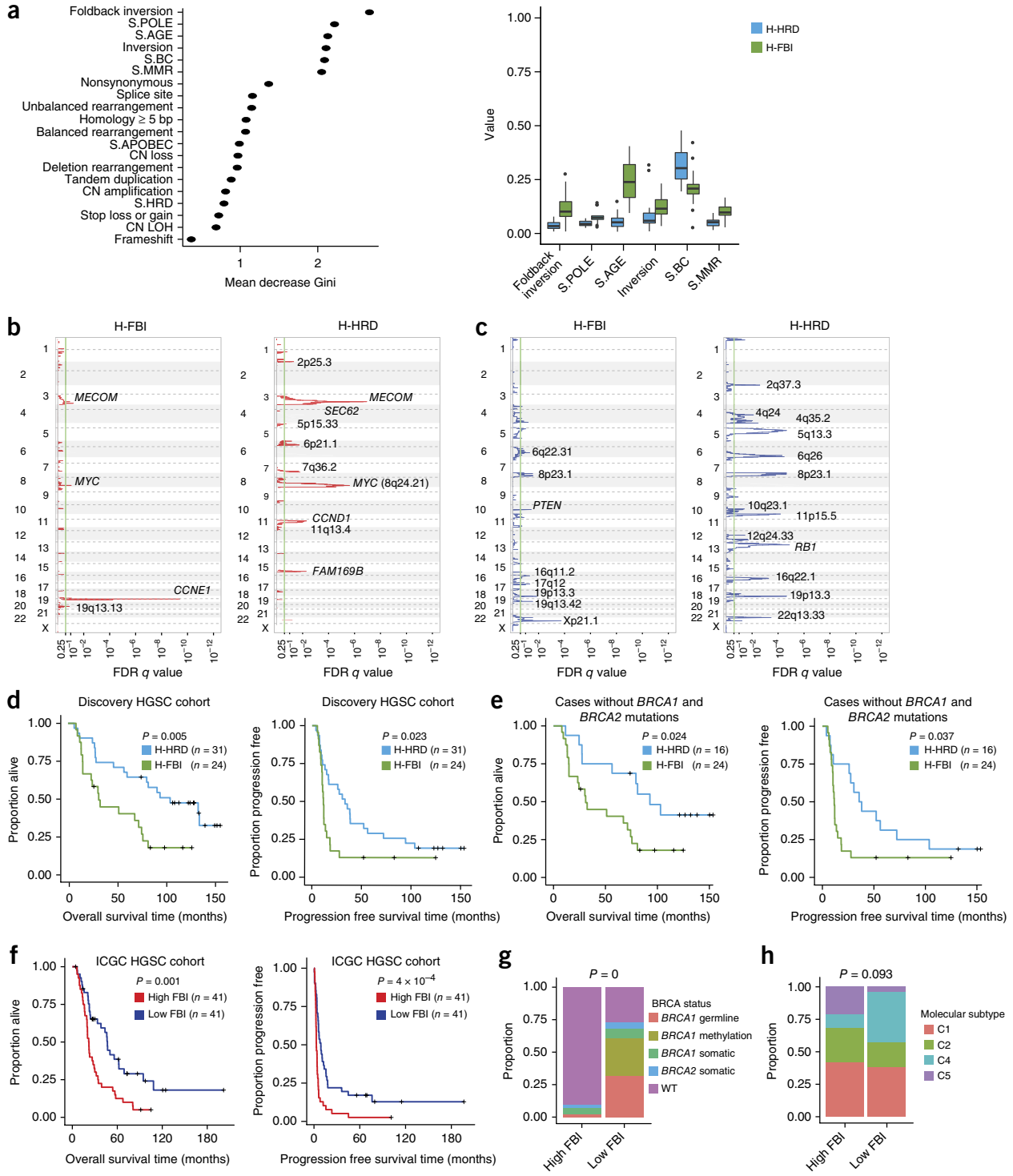

Figure 2 Foldback inversion profile stratifies patients with HGSC. (**a**) Importance of the genomic features segregating the H-HRD and H-FBI subgroups of HGSC tumors. Left, genomic features (*y* axis) sorted in descending order of average Gini score (*x* axis), reflecting the importance of features in stratifying the subgroups. Right, box plot showing the distribution of the top six genomic features contributing to the differences between H-HRD ($n = 31$) and H-FBI ($n = 24$). Boxes extend from the third (Q3) to the first (Q1) quartile (IQR), with the line at the median; whiskers extend to Q3 + 1.5 IQR and Q1 − 1.5 IQR. Points beyond the ends of the whiskers are outliers. CN, copy number. (**b**,**c**) GISTIC profiles showing significant focal copy number amplifications (**b**) and deletions (**c**) for the H-HRD and H-FBI subgroups, with significantly highly amplified and deleted regions ($q < 0.05$) annotated. (**d**) Kaplan–Meier plots of overall (left) and progression-free (right) survival for the H-HRD and H-FBI subgroups of HGSC. *P* values were calculated by log-rank test. (**e**) Kaplan–Meier plots of overall (left) and progression-free (right) survival for the H-HRD ($n = 16$) and H-FBI ($n = 24$) subgroups of HGSC excluding cases bearing *BRCA1* and *BRCA2* germline and somatic mutations. *P* values were calculated by log-rank test. (**f**) Kaplan–Meier plots presenting overall (left) and progression-free (right) survival of ICGC HGSC cases in the High FBI ($n = 41$) and Low FBI ($n = 41$) subgroups. *P* values were calculated by log-rank test. (**g**,**h**) Bar plots show the distribution of BRCA-mutant cases (**g**) and gene-expression-defined molecular subgroups (**h**) between the High FBI and Low FBI subgroups. *P* values were calculated by Pearson's χ^2 test. WT, wild type.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

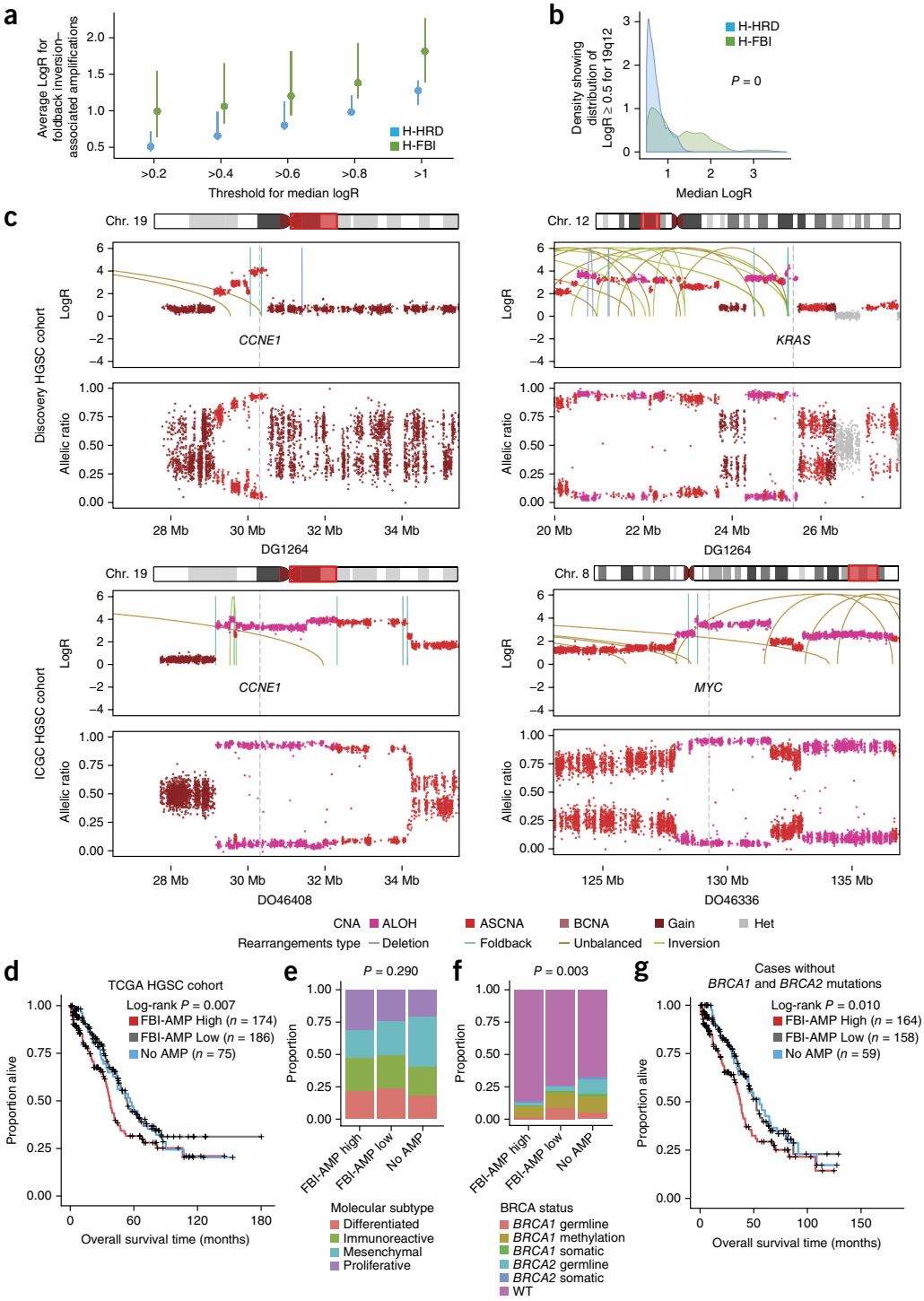

Figure 3 Association between foldback inversions and high-level amplifications, with validation on TCGA data. (**a**) Bar plot showing the lower quartile, median, and upper quartile of mean average LogR computed from foldback inversion–associated copy number amplifications in the H-FBI and H-HRD subgroups at different LogR thresholds from 0.2 to 1. (**b**) Distribution of LogR in the 19q12 amplified regions in the H-HRD and H-FBI subgroups. The *P* value was calculated by two-sample Kolmogorov–Smirnov test. (**c**) Examples of focal HLAMPs colocalized with foldback inversions in *CCNE1* and *KRAS* in our discovery HGSC cohort and in *CCNE1* and *MYC* in the ICGC HGSC cohort. In each example, the following are shown: a chromosome ideogram highlighting the region of interest (top); a LogR plot overlaying rearrangement events (shown with arcs) on CNA segments (middle); and an allelic ratio plot showing the corresponding LOH profile in each region (bottom). ALOH, amplified LOH; ASCNA, allele-specific copy number amplification; BCNA, balanced copy number amplification; Gain, copy number gain; Het, diploid heterozygous. (**d**) Kaplan–Meier plots for TCGA HGSC samples ($n = 435$) in three subgroups: (i) cases without any amplification events (No AMP); (ii) cases enriched in foldback inversion–associated amplifications (FBI-AMP High), and (iii) cases depleted of foldback inversion–associated amplifications (FBI-AMP Low). The *P* value was calculated by log-rank test. (**e**,**f**) Bar plots show the distribution of molecular subgroups (**e**) and BRCA-mutant cases (**f**) in the No AMP, FBI-AMP High, and FBI-AMP Low subgroups. *P* values were calculated by Pearson's χ^2 test. (**g**) Kaplan–Meier plots for the No AMP, FBI-AMP High, and FBI-AMP Low subgroups excluding BRCA-mutant cases. The *P* value was calculated by log-rank test.

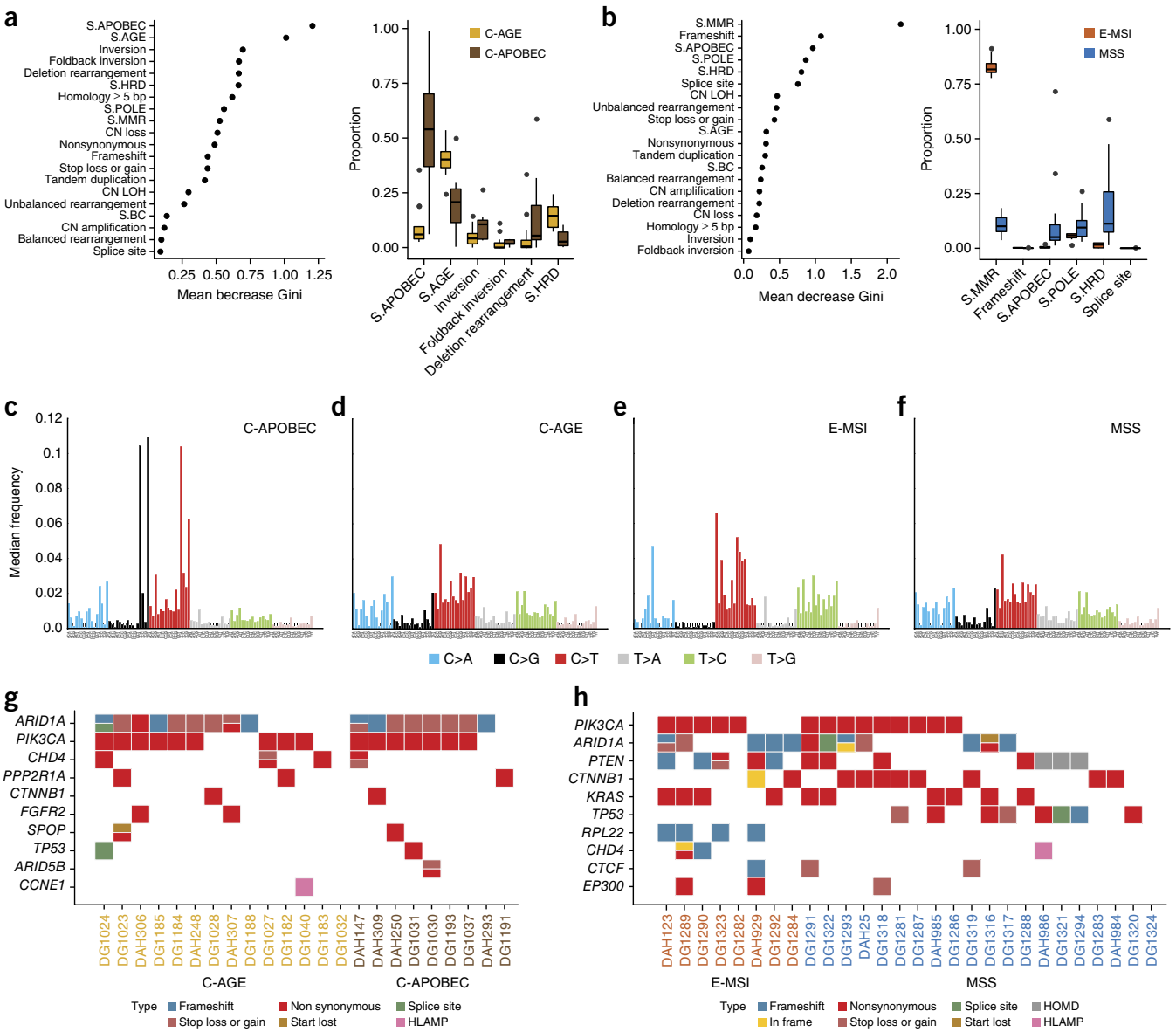

Figure 4 Stratification of endometriosis-associated tumors. (**a**,**b**) Importance of the genomic features segregating the C-APOBEC ($n = 9$) and C-AGE ($n = 14$) subgroups of CCOC (**a**) and the E-MSI ($n = 8$) and MSS ($n = 20$) subgroups of ENOC (**b**). Left, genomic features (*y* axis) sorted in descending order of average Gini score (*x* axis), reflecting the importance of features in stratifying subgroups. Right, box plot showing the distribution of the top six genomic features contributing to the differences between the two subgroups. Boxes extend from the third (Q3) to the first (Q1) quartile (IQR), with the line at the median; whiskers extend to Q3 + 1.5 IQR and Q1 − 1.5 IQR. Points beyond the ends of the whiskers are outliers. (**c**–**f**) Aggregated mutation signature patterns of tumors in the CCOC and ENOC subgroups C-APOBEC (**c**), C-AGE (**d**), E-MSI (**e**), and MSS (**f**); the 96 possible mutation context combinations are plotted along the *x* axis. (**g**,**h**) The oncoplots show frequent mutations in CCOC C-AGE versus C-APOBEC tumors (**g**) and in ENOC E-MSI versus MSS tumors (**h**). Cases are colored according to subgrouping.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

(High FBI, $n = 41$) was associated with poor outcome relative to the group with low prevalence (Low FBI, $n = 41$; log-rank test, $P < 0.001$ for both overall and progression-free survival; **Fig. 2f**). Enrichment of *BRCA1* and *BRCA2* mutants was observed in the Low FBI group (χ^2 test, $P = 0$; **Fig. 2g**), including cases with *BRCA1* somatic (3/5, 60%; **Supplementary Table 11a**) or germline (13/14, 93%) mutations, methylation of the *BRCA1* promoter (12/12, 100%), or *BRCA2* somatic mutations (2/3, 67%), consistent with the observation of prevalence for cases with *BRCA1* or *BRCA2* mutations in the H-HRD subgroup of the discovery cohort. Notably, H-HRD and H-FBI were not associated with previously described molecular subtypes[9,41] in the discovery cohort (χ^2 test, $P = 0.6129$) nor in the ICGC cohort (χ^2 test, $P = 0.0928$; **Fig. 2h** and **Supplementary Table 11c**), indicating that foldback inversion is independent of gene-expression-based subgroups.

Colocalization of foldback inversions and amplifications

Foldback inversions in HGSC tumor genomes were characterized by short distances between two breakpoints (median 2,329 bp; **Supplementary Fig. 8b**). Two short homologous sequences indicative of microhomology were observed on both strands at the breakpoints of foldback inversions (for example, see **Supplementary Fig. 8c–f**), suggesting that microhomology-mediated end joining (MMEJ) DNA repair mechanisms[42,43] could be operating in these tumors. As foldback inversions are a consequence of breakage-fusion-bridge

cycles leading to amplification, we then examined whether HLAMP were associated with foldback inversions in H-HRD and H-FBI cases. We compared the average log ratio (LogR) of the CNA profiles at the breakpoints of foldback inversions in the two subgroups at increasing levels of LogR (**Fig. 3a**). We found significantly greater LogR values in H-FBI than in H-HRD at HLAMP events (median LogR > 1; Mann–Whitney–Wilcoxon test, adjusted $P = 0.0099$; **Supplementary Table 12**). We then examined regions of significant recurrent amplifications across H-HRD and H-FBI HGSC cases (**Fig. 2b**) and computed structural rearrangement types associated with these events. The most prominent signal from recurrent amplifications was found in the 19q21 region harboring *CCNE1*. The distribution of LogR across 19q21 events was accordingly higher in the H-FBI group than in the H-HRD group (**Fig. 3b**). Furthermore, 19q21 amplification events showed a co-occurrence of copy number state change and breakpoints for foldback inversions at the focal aberrant regions of *CCNE1* in H-FBI cases and in ICGC High FBI cases. HLAMPs at 12p12.1 (*KRAS*) and 8q24.21 (*MYC*) also showed precise localization of foldback inversions to the large copy number transition points of HLAMP events in the H-FBI discovery cohort and High FBI ICGC cases (**Fig. 3c** and **Supplementary Fig. 9**). In addition, we sequenced the genomes of a pair of previously characterized HGSC cell lines (TOV1369 and OV1369 (R2)) derived from two temporally sampled primary and relapse specimens from the same patient before and after chemotherapy[44] and identified foldback inversion–associated HLAMPs. The proportion of foldback inversions colocalized with HLAMPs of the primary and relapse cell lines fell squarely within the H-HRD and H-FBI distributions from the discovery cohort (**Supplementary Fig. 10**), which is anecdotally commensurate with a shift before and after chemotherapy in half-maximal inhibitory concentration (IC_{50}) values (carboplatin (average ±s.d.): 5.64 ± 1.29 μM for TOV1369 and 9.8 ± 1.00 μM for OV1369 (R2)).

Prognostic stratification of HGSC in TCGA

We next tested the prognostic relevance of the association between foldback inversion and copy number amplification in The Cancer Genome Atlas (TCGA) ovarian serous cystadenocarcinoma (OV) cohort[9]. Exome-capture sequence and outcome data were available for 435 patients. We analyzed copy number profiles to identify regions harboring amplifications (copy number LogR ≥ 1) for each patient (excluding cases ($n = 75$) with zero copy number events with LogR ≥1) and profiled SVs in these regions for broad inclusion. We then determined the subset of amplification events colocalized with foldback inversions. The average LogR was then computed for each case over amplified regions associated with foldback inversions, which as a score stratified 360 cases into two groups. Cases with a prevalence of foldback inversion–associated amplifications with high LogR (average LogR greater than the median score; $n = 174$) were labeled as FBI-AMP High, with the remainder of cases (average LogR less than or equal to the median score; $n = 186$) labeled as FBI-AMP Low (**Supplementary Table 13a**). Overall survival for FBI-AMP High cases was poor relative to that for FBI-AMP Low and No AMP cases (log-rank test, $P = 0.007$; **Fig. 3d**). Consistent with the discovery HGSC and ICGC cohorts, no association was observed between previously defined molecular subtypes[9,41] and the three groups (χ^2 test, $P = 0.290$; **Fig. 3e** and **Supplementary Table 13b**). Enrichment of BRCA mutants was identified in the No AMP and FBI-AMP Low subgroups (χ^2 test, $P = 0.003$; **Fig. 3f**). However, as for the discovery cohort, excluding BRCA mutants did not have an impact on differences in overall survival among the three groups (log-rank test, $P = 0.010$; **Fig. 3g**), establishing in a large series that foldback inversions associated with amplification events transcend both BRCA mutation status and gene-expression-based molecular subgroups.

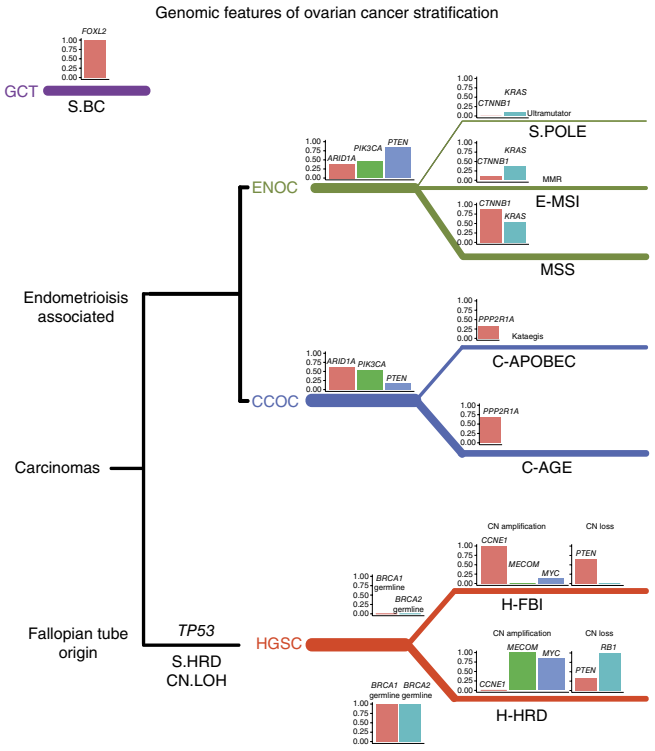

Figure 5 Overview of ovarian tumor subgroupings by the genomic consequences of aberrant DNA repair. GCT is characterized by a unique mutation signature identified in breast cancer (S.BC) and prevalence of *FOXL2* somatic mutations. Within the ovarian carcinomas, LOH profile and HR deficiency signature (S.HRD) distinguish HGSC from non-serous histotypes. Endometriosis-associated ovarian cancer histotypes are associated with *ARID1A*, *PIK3CA*, and *PTEN* somatic mutations. Bar plots show the proportions of cases harboring mutations in a specific gene seen in a subgroup. For example, 83% of the *PTEN*-mutant cases are seen in ENOC, whereas 17% are seen in CCOC. Mutation load and mismatch-repair signature (S.MMR) identify three subgroups of ENOC: ultramutator, MSI, and MSS. The MSS subgroup is associated with a high proportion of *CTNNB1* and *KRAS* mutations. The APOBEC and age-related mutation signatures (S.APOBEC and S.AGE) stratify CCOC into two subgroups. 67% of the *PPP2R1A*-mutant cases are seen in the age-related CCOC group. HGSC splits into two groups: one enriched in foldback inversions and one characterized by other types of rearrangements. The prevalence of *BRCA1* and *BRCA2* germline mutations and the significance of focal copy number amplifications and losses in each HGSC subgroup are shown in the bar plots. The thickness of the branch lines corresponds to the relative sample size of the subgroups in each histotype.

Mutational signatures in CCOC and ENOC

In contrast to HGSC, endometriosis-associated cancers (CCOC and ENOC) were primarily stratified on the basis of SNV mutational signatures (**Fig. 4a**,**b**). We identified two major subgroups of CCOC: C-APOBEC ($n = 9$, 26%) and C-AGE ($n = 14$, 40%). C-APOBEC was characterized by the S.APOBEC mutational signature (mean $C_{S.APOBEC} = 0.55$ versus 0.10; Student's t test, adjusted $P = 0.02$), while C-AGE was characterized by the S.AGE mutation signature (mean $C_{S.AGE} = 0.40$ versus 0.20; Student's t test, adjusted $P = 0.03$; **Fig. 4c**,**d** and **Supplementary Table 8b**). Kataegis analysis[32,33] detected regions of localized hypermutation with clusters of C>T and C>G mutations

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

in eight CCOC cases, five of which (55.6%) were members of C-APOBEC (**Supplementary Table 5a**). No statistical difference in the prevalence of either *ARID1A* or *PIK3CA* mutations was observed between the C-APOBEC and C-AGE subgroups (8/9, 89% versus 12/14, 86%; **Fig. 4g**). A HLAMP at 20q13.2 encompassing *ZNF217* was identified in both C-APOBEC ($n = 3$, 33%) and C-AGE ($n = 8$, 57%) cancers (FDR $q = 0.01$ and <0.0001, respectively; **Supplementary Fig. 4e** and **Supplementary Table 10c**,**d**). Focally deleted regions (FDR $q < 0.1$) including 1p35.3 (encoding *ARID1A*; $n = 10$, 29%), 8p22 ($n = 8$, 23%), and 18q22.2 ($n = 16$, 46%) were observed in CCOC tumors (**Supplementary Table 10h**).

For ENOC tumors, a clear subgroup (labeled as E-MSI) was delineated by S.MMR and corresponded directly to the MSI-positive cases (**Figs. 1a** and **4b**,**e**,**f**). The remaining microsatellite-stable (MSS) tumors were distributed across the remaining six groups, indicating a high degree of heterogeneity and lack of informative discriminant genomic features. Across all ENOC tumors, gene-based analysis recapitulated known mutation patterns. Homozygous deletion in *PTEN* ($n = 3$) was mutually exclusive with *PIK3CA* mutation in ENOC tumors (**Fig. 4h**), and *CTNNB1* mutation ($n = 10$) was mutually exclusive with *KRAS* mutation ($n = 10$). *TP53* was frequently mutated in MSS cases (8/20, 40%), while *RPL22* was frequently mutated in E-MSI cases (4/8, 50%). We found significantly more frameshift indels in E-MSI cases than in MSS ENOC tumors (on average, 0.2% versus 0.08% of all mutations; Student's t test, adjusted $P = 0.005$; **Supplementary Table 8c**), consistent with mismatch-repair deficiency resulting in an increased prevalence of small indels at nucleotide repeats[32]. We examined the recurrent focal somatic CNAs in ENOC tumors (**Supplementary Table 10**) and identified focally amplified regions at 3q26.2 encoding *MECOM* and at 8q24.21 encoding *MYC* and deleted regions at 10q23.31 that encoded *PTEN* in MSS cases. Consistent with the known depletion of copy number events in hypermutated cancers, no focal CNAs were observed in MSI cases (**Supplementary Fig. 4f** and **Supplementary Table 10e**,**f**). Finally, we estimated the number of immunogenic epitopes generated in the patients with ENOC and observed significantly higher counts in MSI than in MSS cases (Kruskal–Wallis rank-sum test, $P = 0.0023$; **Supplementary Fig. 6e**), suggesting higher rates of neoantigen generation in ENOC MSI cases.

DISCUSSION

Our results have several biological and clinical implications. We suggest that stratification of ovarian carcinomas by genomic characteristics provides an etiological model for ovarian cancer histotypes and a framework with which to research and treat patients with ovarian cancer. **Figure 5** shows a summary of our findings with specific illustrative examples, depicting divisions within histotypes as specific pathways to tumor progression that have potential implications for therapeutic options. HGSC cancers are thought to originate in the fallopian tube with early evolutionary acquisition of *TP53* mutation and LOH of chromosome 17 (refs. 45,46). Our results suggest a subsequent divergence whereby tumors acquire contrasting properties of double-strand break DNA repair processes. Some cases (H-HRD) exhibited tandem-duplication- and/or unbalanced-rearrangement-induced amplifications and had increased proportions of deletions and LOH across their genomes, while another distinct group (H-FBI) exhibited foldback inversions co-associated with HLAMPs. As foldback inversions with microhomology are reflective of active MMEJ processes, we suggest that these HGSC tumors may have increased capacity to repair events induced by genotoxic chemotherapy. As such, these cancers may not be responsive to PARP inhibitors and, in the independent cohorts presented here, show evidence of poor response to cisplatin.

Poor survival outcome in cases with foldback inversions was seen in three independent cohorts and by three independent methods. This was true of all non-BRCA-mutant cases. This may indicate that BRCA testing alone is insufficient as a biomarker for directing patients onto specific interventions. Rather, we suggest that the genome itself reflects the DNA repair and mutational processes that were active in its evolutionary history, providing highly discriminant features and potent signals of biological variation for patient stratification.

Several potential therapeutic opportunities for ovarian cancer across the spectrum of histotypes are implicated by our study. Recent progress in elucidating mechanisms that activate and repress MMEJ[47] and dependency of HR-deficient cells on MMEJ[48] implicate DNA polymerase θ (Polθ) as a promising therapeutic target for HGSC. We suggest that foldback inversion attributes may signal a vulnerability by targeting MMEJ through Polθ inhibition. Whether this could be mediated through sensitization to cisplatin or PARP inhibitors, or as an alternative therapeutic modality, remains an open question. Nonetheless, the foldback inversion mutation signature reported here could provide a critical link to identifying patients who may benefit from targeting MMEJ.

Endometriosis-associated tumors have a common etiological origin and often harbor *ARID1A* and *PIK3CA* mutations. However, these cancers grouped according to non-overlapping mutational processes. Approximately one-third of patients with ENOC exhibited MSI (E-MSI) with an accompanying mutation signature reflective of mismatch-repair deficiency and a high proportion of frameshift indels. These cases harbored approximately tenfold more coding mutations and showed evidence of generating neoantigens at a higher rate than other ENOC tumors. Recent successful application of the PD-1 blockade compound pembrolizumab in mismatch-repair-deficient colorectal and non-colorectal cancers[49] signals that mismatch-repair-deficient ENOC cancers could be candidates for immunotherapy, if not cured by surgical resection. The remaining ENOC cases lacked a strong genomic signature and thus remain as a therapeutic challenge.

Twenty-six percent of CCOC cases exhibited a mutational profile consistent with APOBEC-related mutational processes (C-APOBEC). APOBEC signature association was independent of *ARID1A* and *PIK3CA* mutation status, suggesting that these cases may have a unique etiology unrelated to known driver mutation status. APOBEC-mediated deamination has been implicated as a clonal diversity–generating mechanism. As such, the APOBEC mutational process has been proposed as a therapeutic target to prevent ongoing clonal evolution in disease progression[50]. Targeting APOBEC enzymes remains at very early stages in development; however, our results identify a subset of CCOC cases that could be candidates for APOBEC targeting. Although ENOC and CCOC share an etiological origin in endometriosis, their SNV mutation spectra indicate divergence along distinct mutational processes and pathways within and between cancers that share histological characteristics.

Our results show (i) a stratification of the major histotypes of ovarian cancer, leveraging the somatic genome as a biomarker, and (ii) that the consequences of aberrant DNA repair processes are written in the genome sequences of ovarian tumors. While analysis of larger cohorts may yet identify additional strata across the ovarian histotypes as a function of genomic patterns, future clinical trials and development of diagnostics and prognostics will likely derive benefit from assaying the genomic properties identified here. We anticipate that these features, along with histotype and gene-specific abnormalities, will

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

direct oncologists and patients toward more effective therapeutic options and illuminate vulnerabilities to exploit in developing future intervention strategies.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We wish to acknowledge generous long-term funding support from the BC Cancer Foundation, supporting the research program of S.P.S. and OvCaRe. The authors graciously thank the Gray Family Ovarian Clear Cell Carcinoma Research Resource, which provided funding critical to this project. Additional funding was provided by a Terry Fox Research Institute New Investigator grant to S.P.S. and a Canadian Cancer Society Research Institute Impact grant to D.G.H. and S.P.S. The CRCHUM Ovarian Cancer Tumour Bank was supported by the Réseau de Recherche sur le Cancer, Fonds de Recherche Québec Santé, affiliated with the Canadian Tumour Repository Network. S.P.S. is a Michael Smith Foundation for Health Research (MSFHR) scholar and holds a Canadian Institutes for Health Research (CIHR) Foundation grant. S.P.S. and D.G.H. hold Canada Research Chairs. Y.K.W. is an MSFHR postdoctoral fellow. Support for the development of software used in this study was provided by the Genome Canada and Genome British Columbia Bioinformatics and Computational Biology program. Finally, the authors wish to acknowledge the funding support to S.P.S. from the Discovery Frontiers: Advancing Big Data Science in Genomics Research program (grant RGPGR/448167-2013, 'The Cancer Genome Collaboratory'), which is jointly funded by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, Genome Canada, and the Canada Foundation for Innovation and with in-kind support from the Ontario Research Fund of the Ministry of Research, Innovation, and Science.

AUTHOR CONTRIBUTIONS

Y.K.W. and A.B. were the research project leaders and led and designed all data analysis. M.S.A. performed molecular subtype classification of HGSC samples. M.S.A., D.R.C., and H.M.H. collected CCOC, ENOC, and GCT samples and conducted experiments. D.S.G., G.H., A.M., D.L., M.R.A., A.W.Z., K.S., and C.S. performed data analysis and bioinformatics. L.M.P., J.S., M.K.M., and A.W. performed sample preparation and validation experiments. A.O., S.Y., N.Y., M.S., A.T., D.P., M.d.L., and H.F. were responsible for specimens and clinical data from Tokyo and Montreal. R.M., A.J.M., and M.A.M. performed library construction and genome sequencing. C.B.G., H.M.H., A.N.K., and H.L.-C. reviewed all specimens for histological and molecular pathology. J.N.M. and A.-M.M.-M. performed cohort design, HGSC sample selection, and tumor banking. S.A. oversaw experimental design and genome sequencing for validation. Y.K.W., A.B., and S.P.S. wrote the manuscript. D.G.H. conceived the project, provided oversight, and edited the manuscript. S.P.S. conceived and oversaw the project and is the senior responsible author.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

1. Köbel, M. *et al.* Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. *PLoS Med.* **5**, e232 (2008).
2. Vaughan, S. *et al.* Rethinking ovarian cancer: recommendations for improving outcomes. *Nat. Rev. Cancer* **11**, 719–725 (2011).
3. Risch, H.A. *et al.* Population *BRCA1* and *BRCA2* mutation frequencies and cancer penetrances: a kin-cohort study in Ontario, Canada. *J. Natl. Cancer Inst.* **98**, 1694–1706 (2006).
4. 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).
5. Berns, E.M. & Bowtell, D.D. The changing view of high-grade serous ovarian cancer. *Cancer Res.* **72**, 2701–2704 (2012).
6. Anglesio, M.S., Carey, M.S., Köbel, M., Mackay, H. & Huntsman, D.G. Clear cell carcinoma of the ovary: a report from the first Ovarian Clear Cell Symposium, June 24th, 2010. *Gynecol. Oncol.* **121**, 407–415 (2011).
7. Munksgaard, P.S. & Blaakaer, J. The association between endometriosis and ovarian cancer: a review of histological, genetic and molecular alterations. *Gynecol. Oncol.* **124**, 164–169 (2012).
8. 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).
9. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
10. Wiegand, K.C. *et al. ARID1A* mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363**, 1532–1543 (2010).
11. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
12. Obata, K. *et al.* Frequent *PTEN/MMAC* mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. *Cancer Res.* **58**, 2095–2097 (1998).
13. Wu, R., Zhai, Y., Fearon, E.R. & Cho, K.R. Diverse mechanisms of β-catenin deregulation in ovarian endometrioid adenocarcinomas. *Cancer Res.* **61**, 8247–8255 (2001).
14. Campbell, I.G. *et al.* Mutation of the *PIK3CA* gene in ovarian and breast cancer. *Cancer Res.* **64**, 7678–7681 (2004).
15. Kuo, K.-T. *et al.* Frequent activating mutations of *PIK3CA* in ovarian clear cell carcinoma. *Am. J. Pathol.* **174**, 1597–1601 (2009).
16. Kurman, R.J. & Shih, IeM. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer—shifting the paradigm. *Hum. Pathol.* **42**, 918–931 (2011).
17. McConechy, M.K. *et al.* Subtype-specific mutation of *PPP2R1A* in endometrial and ovarian carcinomas. *J. Pathol.* **223**, 567–573 (2011).
18. Nissenblatt, M. Endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **364**, 482–483, author reply 484–485 (2011).
19. Wu, R.-C. *et al.* Frequent somatic mutations of the telomerase reverse transcriptase promoter in ovarian clear cell carcinoma but not in other major types of gynaecological malignancy. *J. Pathol.* **232**, 473–481 (2014).
20. Niskakoski, A. *et al.* Distinct molecular profiles in Lynch syndrome–associated and sporadic ovarian carcinomas. *Int. J. Cancer* **133**, 2596–2608 (2013).
21. Shah, S.P. *et al.* Mutation of *FOXL2* in granulosa-cell tumors of the ovary. *N. Engl. J. Med.* **360**, 2719–2729 (2009).
22. Piccart, M.J. *et al.* Randomized intergroup trial of cisplatin–paclitaxel versus cisplatin–cyclophosphamide in women with advanced epithelial ovarian cancer: three-year results. *J. Natl. Cancer Inst.* **92**, 699–708 (2000).
23. Rauh-Hain, J.A. & Penson, R.T. Potential benefit of sunitinib in recurrent and refractory ovarian clear cell adenocarcinoma. *Int. J. Gynecol. Cancer* **18**, 934–936 (2008).
24. McAlpine, J.N. *et al. HER2* overexpression and amplification is present in a subset of ovarian mucinous carcinomas and can be targeted with trastuzumab therapy. *BMC Cancer* **9**, 433 (2009).
25. Anglesio, M.S. *et al.* IL6–STAT3–HIF signaling and therapeutic response to the angiogenesis inhibitor sunitinib in ovarian clear cell cancer. *Clin. Cancer Res.* **17**, 2538–2548 (2011).
26. Farley, J.H., Gibson, S.J. & Monk, B.J. American Society of Clinical Oncology 2012 annual meeting update: summary of selected gynecologic cancer abstracts. *Gynecol. Oncol.* **126**, 319–324 (2012).
27. Anglesio, M.S. *et al.* Molecular characterization of mucinous ovarian tumours supports a stratified treatment approach with HER2 targeting in 19% of carcinomas. *J. Pathol.* **229**, 111–120 (2013).
28. Ledermann, J. *et al.* Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N. Engl. J. Med.* **366**, 1382–1392 (2012).
29. Ledermann, J. *et al.* Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol.* **15**, 852–861 (2014).
30. Mirza, M.R. *et al.* Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N. Engl. J. Med.* **375**, 2154–2164 (2016).
31. 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).
32. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
33. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
34. Campbell, P.J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
35. Sudmant, P.H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
36. Ng, C.K.Y. *et al.* The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer. *J. Pathol.* **226**, 703–712 (2012).
37. Sasaki, S. *et al.* Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* **22**, 3792–3798 (2003).
38. Yang, L. *et al.* Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919–929 (2013).
39. Hermetz, K.E. *et al.* Large inverted duplications in the human genome form via a fold-back mechanism. *PLoS Genet.* **10**, e1004139 (2014).
40. Patch, A.-M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
41. 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).

42. Ottaviani, D., LeCain, M. & Sheer, D. The role of microhomology in genomic structural variation. *Trends Genet.* **30**, 85–94 (2014).
43. Sfeir, A. & Symington, L.S. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* **40**, 701–714 (2015).
44. Le'tourneau, I.J. *et al.* Derivation and characterization of matched cell lines from primary and recurrent serous ovarian cancer. *BMC Cancer* **12**, 1 (2012).
45. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231**, 21–34 (2013).
46. Schwarz, R.F. *et al.* Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med.* **12**, e1001789 (2015).
47. Mateos-Gomez, P.A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 (2015).
48. Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518**, 258–262 (2015).
49. Le, D.T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
50. Swanton, C., McGranahan, N., Starrett, G.J. & Harris, R.S. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Patient cohort description. Ovarian cancer cases (n = 133) were selected from the OvCaRe gynecological tissue bank (Vancouver, Canada; n = 65), the CRCHUM Ovarian Cancer Tumour Bank (Montreal, Canada; n = 55), and the Anatomical Pathology archives at the Jikei University School of Medicine (Tokyo, Japan; n = 13) (**Supplementary Table 1**). Patient consent, or waiver of consent, was approved by the respective institutional research ethics boards. The BC Cancer Agency or University of British Columbia Research Ethics Board approved the overall project processes.

HGSC cases in the OvCaRe and CRCHUM tumour banks were selected according to the following criteria: (i) treatment with platinum–taxane-based therapy; (ii) relapse within 12 months (365 d) or at least 4.5 years (1,642.5 d) of follow-up data; and (iii) at least 50% tumor content by hematoxylin and eosin staining and expert pathology review. All cases were re-reviewed by expert pathologists to confirm the diagnosis of HGSC. Germline *BRCA1* and *BRCA2* mutation status was determined for all patients through hereditary cancer screening programs. The design for case selection for the discovery cohort was engineered to amplify biological differences by selecting cases from the extremes of the outcome distribution.

For CCOC, ENOC, and GCT cohorts, OvCaRe cases were reviewed, including frozen material, by at least two expert gynecopathologists (H.M.H., A.N.K., H.L.-C., and C.B.G.) before inclusion in the sequencing cohort, and C.B.G. provided confirmation on the final selected cohort. Frozen hematoxylin and eosin sections from Tokyo were also used for evaluation along with representative hematoxylin and eosin images and review done at the Jikei School of Medicine.

For ENOC, samples DAH985 and DG1288 were recurrent and both were from cases treated with chemotherapy after the first surgery. DAH123 was an untreated sample, corresponding to a metastasis from a primary endometrial tumor. All HGSC, GCT, and CCOC tumors and the remaining ENOC tumors were primary tumor samples.

Clinical data and follow-up were provided by the Cheryl Brown Ovarian Cancer Outcomes Unit (for OvCaRe samples) or the respective providing institutions for Montreal and Tokyo cases.

Library construction and sequencing. Frozen specimens with >50% tumor cellularity (based on initial slide review) were used for cryosectioning and subsequent nucleic acid extraction. Patient tumor and normal blood samples were derived from fresh-frozen primary, untreated tumor specimens collected at diagnosis during standard-of-care debulking surgery. Germline DNA was obtained from the peripheral blood buffy coat for all specimens except 13 from Tokyo, where non-cancer frozen tissue was used as a germline source. DNA was extracted from matched normal (blood) and tumor (frozen tissue) samples using the QIAamp Blood and Tissue DNA kit (Qiagen) and quantified using a Qubit fluorometer and reagents (high-sensitivity assay). Three lanes of Illumina HiSeq 2500 v4 chemistry for normal samples and five lanes for tumor samples were obtained. The PCR-free protocol was adopted to eliminate PCR-induced bias and improve coverage across the genome.

Sequencing analysis. Whole-genome sequencing analysis was performed to identify somatic alterations on all scales in the tumor genome of each case, including SNVs, small indels, CNAs, and SVs. Revalidation through PCR-based targeted amplicon sequencing was performed for selected SNVs and SVs. All sequencing analyses were completed using workflows implemented in the Kronos workflow assembler for genome analytics and informatics[51].

Additional details on the analyses performed in this study, including whole-genome sequence analysis and validation of findings on the TCGA and ICGC HGSC cohorts, can be found in the **Supplementary Note**.

Data availability. All sequencing data are available at the European Genome-phenome Archive (EGA) under study accession EGAS00001002390.

51. Taghiyar, M.J. *et al.* Kronos: a workflow assembler for genome analytics and informatics. Preprint at *bioRxiv* http://dx.doi.org/10.1101/040352 (2016).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/ng.3849