Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer

Andrew McPherson[1–4,13], Andrew Roth[1,3,13], Emma Laks[1], Tehmina Masud[1,4], Ali Bashashati[1], Allen W Zhang[1,3,5], Gavin Ha[1,3,6,7], Justina Biele[1], Damian Yap[1], Adrian Wan[1], Leah M Prentice[8], Jaswinder Khattra[1], Maia A Smith[1,3], Cydney B Nielsen[4], Sarah C Mullaly[1], Steve Kalloger[1], Anthony Karnezis[8], Karey Shumansky[1], Celia Siu[1], Jamie Rosner[1], Hector Li Chan[8], Julie Ho[8], Nataliya Melnyk[8], Janine Senz[8], Winnie Yang[8], Richard Moore[9], Andrew J Mungall[9], Marco A Marra[9], Alexandre Bouchard-Côté[10], C Blake Gilks[11], David G Huntsman[1,4,8], Jessica N McAlpine[12], Samuel Aparicio[1,4] & Sohrab P Shah[1,4,9]

We performed phylogenetic analysis of high-grade serous ovarian cancers (68 samples from seven patients), identifying constituent clones and quantifying their relative abundances at multiple intraperitoneal sites. Through whole-genome and single-nucleus sequencing, we identified evolutionary features including mutation loss, convergence of the structural genome and temporal activation of mutational processes that patterned clonal progression. We then determined the precise clonal mixtures comprising each tumor sample. The majority of sites were clonally pure or composed of clones from a single phylogenetic clade. However, each patient contained at least one site composed of polyphyletic clones. Five patients exhibited monoclonal and unidirectional seeding from the ovary to intraperitoneal sites, and two patients demonstrated polyclonal spread and reseeding. Our findings indicate that at least two distinct modes of intraperitoneal spread operate in clonal dissemination and highlight the distribution of migratory potential over clonal populations comprising high-grade serous ovarian cancers.

High-grade serous ovarian cancer (HGSOvCa) is the most lethal histotype of ovarian cancer, where up to 80% of patients suffer relapse after initial response to treatment, with many succumbing to treatment-resistant disease. HGSOvCa is characterized by widespread and early intraperitoneal spread to organs, such as the omentum and contralateral ovary. The genomes of HGSOvCa exhibit pervasive instability due to *TP53* loss in nearly all cases[1], accompanied by deficiencies in homologous recombination and other DNA repair pathways in the majority of cases[2]. Thus, combined dysregulation of apoptosis and DNA repair processes provides a mechanism for generating a substrate of genomically diverse clones[3–5] from a monoclonal etiology[6,7].

The peritoneal cavity lacks distinct physical barriers to ovarian cancer metastasis, resulting in early and widespread disease at distal peritoneal sites. The majority of disease spread in HGSOvCa is thought to be driven by cell mixing and invasion within a physical anatomical cavity[8]. Such 'local' spreading properties contrast with those for other solid malignancies, where hematological or lymphatic processes govern metastasis[9–12]. It is unknown what proportion of local spread in the peritoneum arises from monoclonal seeding and

expansion or from extensive cellular mixing of diverse polyclonal populations. As such, the interaction of differential migration and invasion capacities for divergent clones with the intraperitoneal fitness landscape is poorly understood. One model for local spread would suggest that microenvironmental niches induced by peritoneal organs introduce variable selective topologies for genomically diverse clones. Alternatively, if the composition of the peritoneal cavity were uniformly selective, this could yield a pattern consistent with multiple migrations and extensive clonal mixing.

Studying the properties of intraperitoneal spread for genomically defined clones in HGSOvCa requires a multiregion whole-genome sequencing approach. Whole-genome sequencing offers a robust experimental design to study evolution in cancer, primarily because orders of magnitude more mutations are accessible to phylogenetic modeling, with ascertainment of structural copy number alterations and rearrangement breakpoints as clonal marks. Single-nucleus sequencing in turn complements bulk whole-genome sequencing approaches for direct measurement of extant clonal genotypes in a sample of interest, facilitating tracking of clones in temporal or spatial dimensions[13–15].

We set out to resolve the modes of clonal spread to gain an improved understanding of pretreatment HGSOvCa disease progression. Clonal population profiling of spatially distinct intraperitoneal samples was performed, combining the experimental approaches of whole-genome and deep, targeted sequencing in all samples and direct sequencing of >1,500 individual nuclei across multiple samples from three patients. We established robust clonal genotypes and their phylogenetic relationships, gaining new insights into the mechanisms of tumor evolution, including mutation loss over evolutionary history and convergent evolution of the structural genome. We measured the abundance of clones across all sites for each patient, identifying properties of intraperitoneal spread across two dimensions: the clonal composition of individual samples and the migration capacity of individual clones. A minority of patients exhibited a high degree of polyphyletic clonal mixing and reseeding of clones at distal foci, whereas in the majority of cases clonal diversity emerged at a primary site followed by unidirectional, monoclonal seeding to distal intraperitoneal sites.

RESULTS

We collected 68 tumor samples (**Supplementary Table 1**) from seven patients (between 5 and 13 samples per patient) (**Table 1**), including samples from the ovary, omentum, fallopian tube, peritoneal sites, and other distant metastatic sites (**Fig. 1a**). For one patient (patient 7), samples were obtained at initial diagnosis, at first relapse after 13 months, and at second relapse after an additional 19 months. For the remaining patients, all samples were obtained during initial debulking surgery. Additional detail on sample acquisition, including the limitations of surgical sampling, is provided in the **Supplementary Note**.

Sample phylogenies and clonal divergence by mutation loss

Whole-genome sequencing (~30× coverage) was applied to 31 cryo-reserved tissues that yielded high-quality DNA extractions (discovery samples), in addition to matched normal blood from each patient (**Supplementary Table 2**). The remaining 37 formalin-fixed, paraffin-embedded specimens (archival samples) were used in targeted sequencing as described below. Analysis of the discovery samples resulted in a catalog of somatic genomic aberrations, including single-nucleotide variants (SNVs), allele-specific absolute copy number changes, and rearrangement breakpoints, for each patient (see the **Supplementary Note** for details on the analytical approaches). Between 3,577 and 16,987 SNVs and between 132 and 1,092 rearrangement breakpoints were identified per patient. A high proportion of events were specific to a subset of samples, ranging from

0.28 to 0.87 for SNVs and from 0.22 to 0.48 for rearrangement breakpoints (**Supplementary Table 3**), providing an initial quantification of the degree of clonal divergence between sites.

We inferred the evolutionary relationships between dominant clones in each sample (sample trees) as defined by both SNVs and rearrangement breakpoints. We implemented a probabilistic phylogenetic model based on the stochastic Dollo process[16,17] (**Supplementary Note**) accounting for genomic instability and associated deletion of chromosome segments containing point mutations. We computed maximum-likelihood sample tree topologies and rate of mutation loss parameters for SNVs (**Fig. 1a** and **Supplementary Figs. 1–3**) and rearrangement breakpoints (**Supplementary Figs. 4–6**). The SNV- and breakpoint-derived trees were consistent for all patients except patient 3, providing corroborating topologies from completely orthogonal features of the genome. The number of rearrangement breakpoints and the number of SNVs originating on each branch in each tree were correlated (Spearman $\rho = 0.88$, $P = 2.2 \times 10^{-16}$, patient 3 excluded; **Supplementary Fig. 1b** and **Supplementary Table 4**). However, numbers of mutation losses were only weakly correlated (Spearman $\rho = 0.37$, $P = 0.01$, patient 3 excluded; **Supplementary Fig. 1c** and **Supplementary Table 4**). In all patients, sample tree topologies indicated the presence of multiple ancestrally related clones with substantial divergences between samples.

The inferred probability of losing a mutation on a branch ranged from 0.00094 (patient 10) to 0.012 (patient 9) for SNVs and from 0 (patients 2, 3, and 10) to 0.044 (patient 9) for rearrangement breakpoints. The proportion of SNV losses attributed to copy number changes ranged from 0.28 to 0.89 (**Fig. 1b**, **Supplementary Figs. 7–13**, and **Supplementary Table 3**), with patients 1 and 7 exhibiting the highest proportions at 0.87 and 0.89, respectively (**Supplementary Table 3**). The majority of lost SNVs were found in regions of divergent copy number across samples (see the example in **Fig. 1c**), suggesting that SNVs were likely removed by sample-specific or clade-specific segmental deletions.

Temporal attributes of mutation signatures in HGSOvCa

We next sought to use the sample trees to determine the mutational signatures (as defined by Alexandrov et al.[18]) operating at different points in the evolutionary histories of the tumors. We used a latent Dirichlet allocation (LDA) model[19] to learn the relative proportion of sample-specific SNVs attributed to each of a curated list of 30 signatures[20] (**Supplementary Fig. 14**, **Supplementary Table 5**, and **Supplementary Note**). Hierarchical clustering of the signature proportions indicated that sample-specific signatures were similar within each patient, with the brain metastasis in patient 7 being the lone

Table 1 Patients and samples included in the study

Patient	Samples (index)	Anatomical samples (index)	Samples (archival)	Anatomical samples (archival)	Age at diagnosis (years)	Stage	Recurrence	RFS (months)	Status[a]	OS (months)
1	6	ROv1–ROv4; Om1; SBwl	6	ApC1; LFTB4; LOvB2; RFTA16; ROvA4; SBwlE4	71.7	IIIC	No		NED	55
2	4	Om1, Om2; ROv1, ROv2	6	LOvD3; OmA2; OmB1; RFTC10; ROvC2; ROvC4	75.5	IIIC	Yes	11	DOD	45
3	4	ROv1, ROv2; Om1; LOv2	7	CDSB1; ClnE1; LFTC1; LOvC5; OmF2; RFTA2; ROvA7	69.2	IIIC	Yes	27	AWD	53
4	5	ROv1–ROv4; LPvS	4	LOvB2; LPvSC1; ROvA5; RPvSD1	52.9	IIIB	Yes	46	AWD	52
7[b]	3	LOv1; BrnM; RPvM	10	LOvA10; LOvA4; ROvC4; ROvC5; ROvC6; RUtD1; RUtD2; RUtD3; BrnMA1; BwlA6	47.2	IIIC	Yes	9	DOD	52
9	5	ROv1 LOv1, LOv2; Om1, Om2	0	NA	52.5	IIIC	Yes	6	DOD	31
10	4	ROv1–ROv4	4	LFTB2; OmC1; ROvA4; ROvA9	73.9	IIA	No		NED	49

All cases are of high-grade (grade III) serous histology, having undergone secondary pathology review. Stage refers to surgical staging. RFS, recurrence-free survival; OS, overall survival. The abbreviations for anatomical samples are defined in **Supplementary Table 1**. [a]Current disease status: AWD, alive with disease; NED, no evidence of disease; DOD, dead of disease. [b]Samples were acquired temporally: BrnM was acquired 14 months after the LOv, ROv, and RUt samples, and the RPvM and Bwl samples were acquired after an additional 19 months.
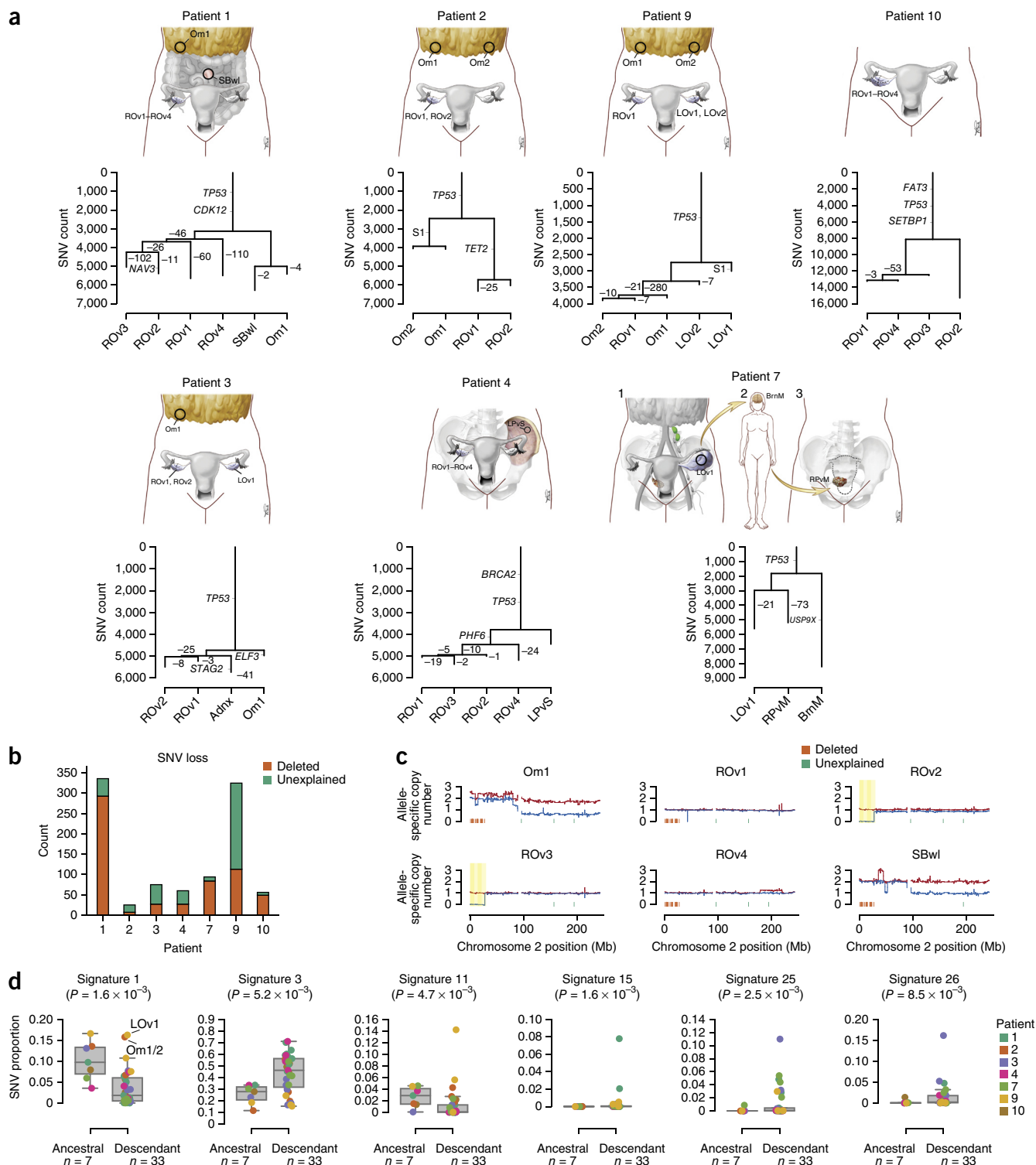
Figure 1 SNV phylogenies for seven patients with HGSOvCa. (**a**) Anatomical sites sampled for whole-genome sequencing in seven patients with HGSOvCa. Phylogeny was inferred from SNVs, with previously identified driver genes annotated on the branches. Branch lengths represent the number of SNVs originating on each branch. Branches are annotated with the number of SNVs lost along each branch. (**b**) Number of SNVs classified as deleted or unexplained by any identifiable deletion event. (**c**) Allele-specific copy number for the major (red) and minor (blue) alleles (*y* axis) are shown for chromosome 2 in all samples from patient 1. SNVs classified as deleted or unexplained by identifiable deletions are shown as orange and green ticks above the *x* axis. Only SNVs classified as lost are shown. A deletion event common to ROv2 and ROv3 is highlighted in yellow. (**d**) Box plots of the inferred proportion of SNVs attributed to each mutational signature for SNVs acquired on the ancestral versus descendant branches of each sample tree. Only signatures with a significant difference in the inferred proportion between the ancestral and descendant branches are shown (*P* < 0.01, Mann–Whitney *U* test). Branches with fewer than 100 mutations were excluded. Superimposed strip plot points representing each data point are colored by patient. Boxes extend from the third (Q3) to the first (Q1) quartile (IQR), with a line at the median; whiskers extend to Q3 + 1.5 IQR and Q1 − 1.5 IQR.

Figure 2 Evidence of samples composed of divergent clonal mixtures. (**a–c**) Predicted copy number for major (red) and minor (blue) alleles for each sample of chromosome 2 from patient 2 (**a**), chromosome 2 from patient 3 (**b**), and chromosome 3 from patient 9 (**c**). Ticks above the *x* axis show the positions of SNVs present in each sample. SNV ticks are colored orange if they can be explained by a clonally dominant deletion; otherwise, they are colored green. In **a**, a subclonally deleted chromosomal segment is highlighted in yellow. In **c**, SNVs are evenly distributed across the chromosome (in addition to other chromosomes) and are not associated with identifiable deletion events. (**d**) Length of the reference genome predicted as having a subclonal copy number state for each sample in each patient.

exception. Mutations were predominately generated by signature 1, the age-associated signature; signature 3, associated with failure of DNA double-strand break repair by homologous recombination; and signatures 5 and 8, common signatures of unknown etiology.
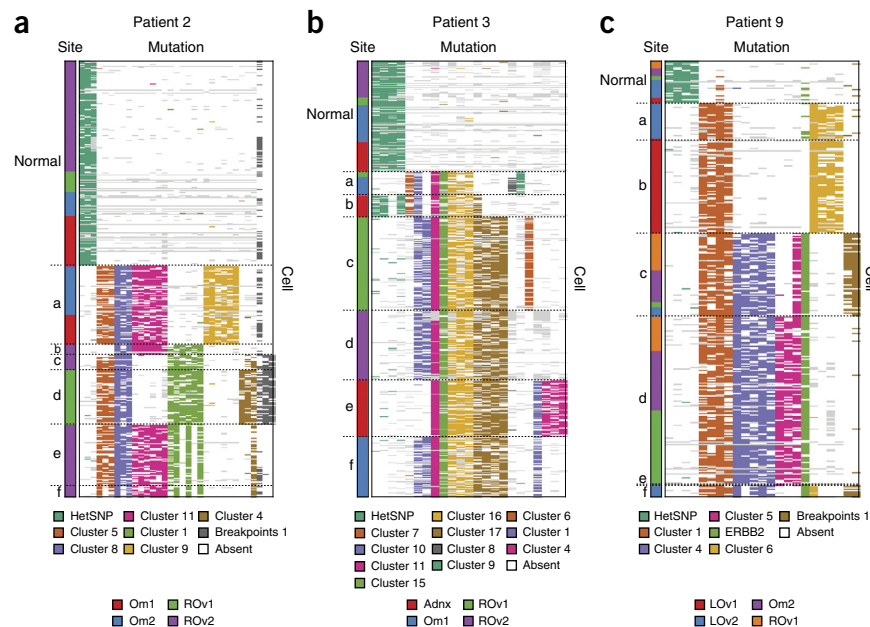
We characterized the relative evolutionary timing of each signature, applying LDA to SNVs partitioned by their phylogenetic origin in the sample tree (**Supplementary Fig. 15** and **Supplementary Table 6**). Hierarchical clustering of branch-specific signature proportions resulted in two major clusters, one of which contained all ancestral branches. Two signatures were highly differentiated in ancestral and descendant SNVs. Signature 3 (homologous recombination deficiency) generated a higher proportion of descendant SNVs, and signature 1 (age associated) generated a higher proportion of ancestral SNVs (Mann–Whitney *U* test, *P* < 0.01; **Fig. 1d**), consistent with the activation of homologous recombination deficiency after the expansion of ancestral clones. Four additional signatures, including signatures 15 and 26 (associated with defective DNA mismatch repair), were significantly associated with SNVs on descendant branches in a minority of patients. Two descendant branches in patients 2 and 9 exhibited a high proportion of signature 1 SNVs and a corresponding low proportion of signature 3 SNVs (**Fig. 1d**), suggesting that clones in patients 2 and 9 had either intact or only partially deficient homologous recombination.

Evidence of clonal mixtures in specific samples

Multiple lines of evidence nominated samples from patients 2, 3, and 9 as mixtures of divergent tumor lineages. For all three patients, more than 50% of lost mutations could not be explained by sample-specific deletion (**Fig. 1b**). For patients 2 and 3, densely spaced sets of SNVs predicted to be lost coincided with clone-specific copy number changes (**Fig. 2a,b**). By contrast, lost SNVs in patient 9 were distributed

uniformly throughout the genome (**Fig. 2c** and **Supplementary Fig. 12**), without convincing corroborating copy number changes. Moreover, the top four samples with respect to length of the genome called as having a subclonal copy number state included a sample each from patients 2, 3, and 9 (**Fig. 2d** and **Supplementary Table 7**), supportive of multiple divergent clones being present in samples from these patients.

We performed single-nucleus sequencing on patients 2, 3, and 9 (*n* = 1,680 cells in total; **Table 2**) with the aim of determining the co-occurrence of breakpoints and PyClone SNV clusters to be used as clonal marks (Online Methods). Single-cell sequencing, in combination with the single-cell genotyper (SCG) model[21], produced clusters of cells (clones) defined by common patterns of presence and absence for clonal marks (clonal genotypes). The coexistence of multiple clones was evident in patient 2 (right ovary 2 sample (ROv2; genotype clusters b, c, e, and f)) (**Fig. 3a**), patient 3 (adnexa sample (Adnx; genotype clusters b and e), omentum 1 sample (Om1; genotype clusters a and f), and ROv1 (genotype clusters a and c)) (**Fig. 3b**), and patient 9 (ROv1 (genotype clusters c and d) and Om2 (genotype clusters c and d)) (**Fig. 3c**). Clonal genotypes were used to infer clone phylogenies, ancestral genotypes, and the relative order of mutation acquisition on the basis of a stochastic Dollo-like model

Table 2 Single-nucleus sequencing statistics

Patient	Samples[a]	Cells[b]	Breakpoints[c]	SNVs[d]	Median coverage	Mean coverage
2	4	588	5	43	391	1,273
3	4	672	12	84	52	355
9	5	420	5	43	375	2,775

[a]Number of samples sequenced. [b]Number of nuclei sequenced. [c]Number of breakpoint targets. [d]Number of SNV targets.

Figure 3 Single-nucleus sequencing of patients with samples containing mixtures of divergent lineages. (**a**–**c**) Nucleus × target matrices for patients 2 (**a**), 3 (**b**), and 9 (**c**) with nuclei clustered by genotype using the SCG model. Matrix entries are colored by variant cluster (PyClone cluster, breakpoint, or heterozygous SNP). Nuclei are annotated with the sample from which they originated (colored "site" indicators), in addition to their genotype clusters (labels to the left of the site indicators). HetSNPs represent loci predicted in the whole-genome sequencing data to be heterozygous in the normal sample that were predicted to be homozygous reference in tumors; these SNPs act as indicators of normal cells. For patient 3, HetSNPs also include SNPs predicted to be encompassed by subclonal LOH events. Cluster labels correspond to those in **Supplementary Table 16**.

(**Supplementary Fig. 16**, **Supplementary Table 8**, and **Supplementary Note**).

Intraperitoneal sample diversity and migration capacity

In total, we inferred 44 clonal genotypes across the seven patients, ranging in number from 3 (patient 9) to 9 (patient 1) per patient (**Supplementary Fig. 16** and **Supplementary Table 8**), resolved by either single-cell sequencing (patients 2, 3, and 9) or inferred from deep sequencing (patients 1, 4, 7, and 10; Online Methods). We applied targeted deep sequencing of mutations comprising the clonal genotypes on both discovery and archival samples to determine their prevalence conditioned on the PyClone clustering of the discovery samples (**Supplementary Note**). For each patient, we elucidated the pattern of spread for each clone and a robust estimate of the clonal composition of each sample (**Supplementary Table 9**).

We categorized samples as clonally pure or polyclonal; polyclonal samples were further classified into samples where multiple clones were from within the same phylogenetic clade (monophyletic) or were from different clades (polyphyletic) (**Fig. 4** and **Supplementary Fig. 17**). The majority of samples were monophyletic or clonally pure; however, each patient exhibited at least one polyphyletic sample (**Fig. 5a**), including LOvB2 in patient 1, ROv2 in patient 2, ROv2 and ROv4 in patient 4, ROvA7 in patient 3, and ROvA9 in patient 10. Polyphyletic samples are indicative of an ancestral population having undergone branched clonal evolution or of the site having been seeded either serially or simultaneously (polyclonal migration) by multiple clones. For patients with a single polyphyletic sample (patients 2, 7, and 9), the ancestral population model is more parsimonious with respect to the number of migration events needed to explain the data.

For four of the six patients with primary ovary and non-ovarian intraperitoneal discovery samples (patients 1, 2, 4, and 9), one of the two major clades of the clone phylogeny dominated the putative primary ovary site, with the divergent clade dominating the peritoneal and contralateral ovary sites (**Fig. 5b**). For patients 1 and 2, the primary ovary samples were pure for the primary clade, whereas, for patients 4 and 9, the primary ovary samples contained minor populations of the peritoneal clones. For patients 4 and 9, analysis of competing migration scenarios (**Supplementary Note**) was most parsimonious with migrating clones originating within and then disseminating away from the primary ovary site (**Supplementary Figs. 18–24**). Thus, for four of the six patients, we did not observe clonal reseeding from peritoneal to primary sites. For patient 7, the

minimum-event migration scenario nominated RUtD3 as the pre-treatment site in which the observed tumor clones originated, with subsequent metastasis to the brain and other peritoneal sites.

To characterize the distribution of clonal composition over all samples, we calculated two parameters for each sample: the sample mixture entropy (the relative balance of the clonal proportions of the clones in a sample) and the intraperitoneal and between-clade divergence of the clones comprising the sample (**Fig. 5a** and **Supplementary Note**). The distribution across samples was patterned by three broad categories: samples that were clonally pure (**Fig. 5a**, circles), samples with multiple clones present but with relatively low entropy, consistent with monophyletic topologies (**Fig. 5a**, triangles), and samples with balanced mixtures from polyphyletic clones (**Fig. 5a**, stars). The spectrum of samples over the space of this distribution can be interpreted as follows: samples in the upper right region are clonally diverse and composed of clones from distinct clades in the clone phylogeny, whereas samples in the lower left region are clonally pure or are formed from closely related clones. Notably, every patient exhibited at least one sample with high entropy and high clone divergence (**Fig. 5a**). However, the shape of the sample distribution indicates that intraperitoneal spread in the cohort is not characterized by repeat reseeding of multiple clones. Most of the density in **Figure 5a** is found in the lower left region, where clones are characterized by either clonally pure or monophyletic topologies. Thus, only a minority of specific sites (often primary ovary sites) are permissive to coexisting clones (whether from multiple migrations or branched evolution), which seed the remainder of the intraperitoneal sites composed of clonally pure or chain topologies. Exceptions to this trend were found in patients 3 and 10, for whom samples comprised multiple polyphyletic samples in the upper right region of the distribution (**Fig. 5a**, purple and brown stars). As such, for patients 3 and 10, the existence of multiple samples composed of divergent lineages is concordant with polyclonal reseeding or polyclonal migration, in contrast to the scenarios in patients 1, 2, 4, and 9. Furthermore, these two patients ranked highest in the number of predicted migration events per site (**Fig. 5d**).

We next calculated the spreading and relative within-sample prevalence characteristics of individual clones. We plotted the number of

samples harboring each clone as a function of the maximum prevalence observed across all samples and identified clones possessing three phenotypes with respect to spread in the peritoneal cavity: clones with site specificity, clones with the capacity to colonize and become dominant, and high-mobility clones that never achieve dominance (**Fig. 5c**). Clones in these three groupings were distributed non-randomly
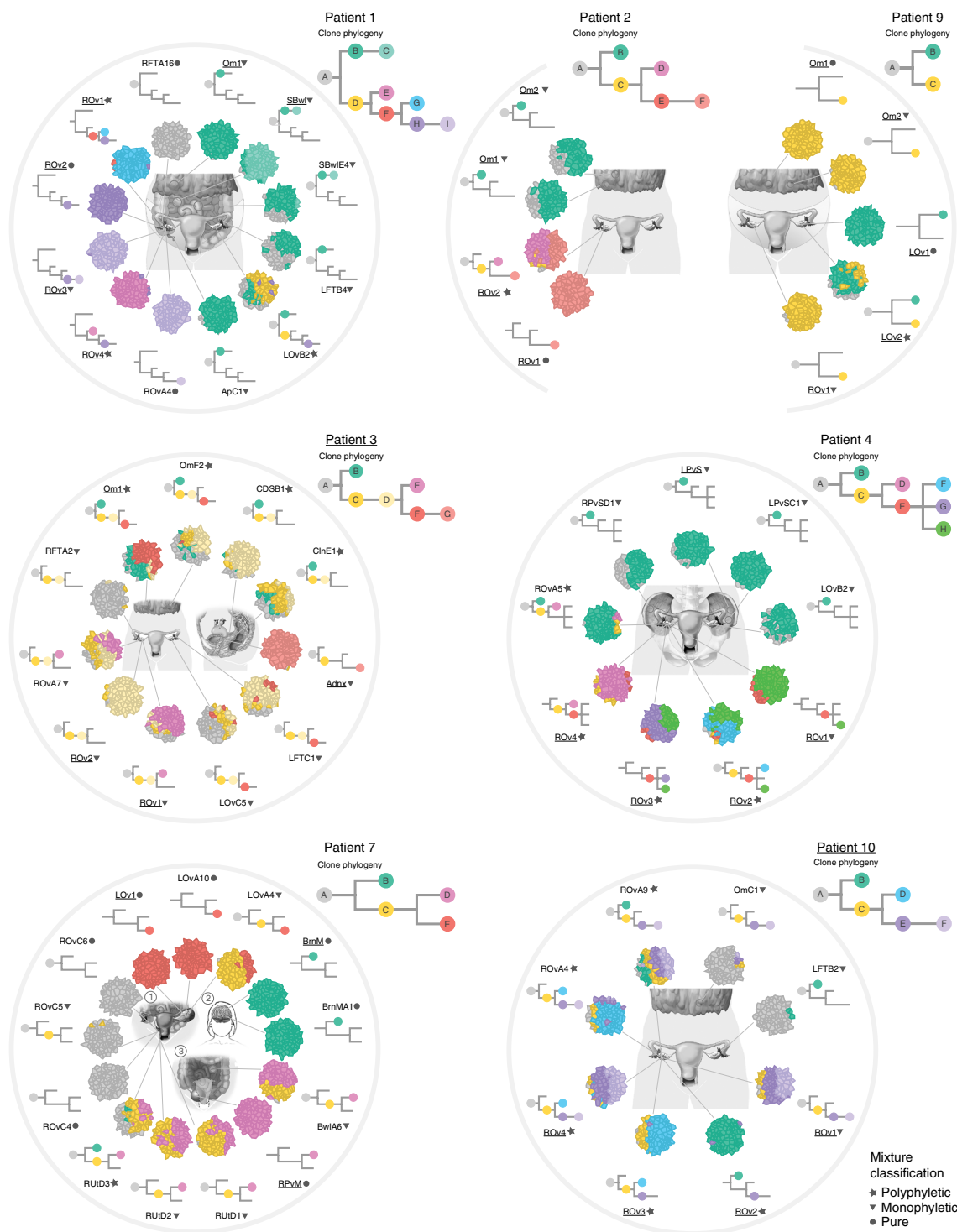
Figure 4 Clone phylogenies and sample clone mixtures. A clone phylogeny is shown for each patient, with distinct clonal genotypes color-coded. The clonal composition of each sample is depicted as a schematic tumor, with the coloring of 'cells' proportional to the prevalence of each clone. Samples are arrayed in a circle around the anatomical diagram, with constituent clones from the overall clone phylogeny represented in an outer circle. Index samples sequenced by whole-genome sequencing are underlined. Pure, monophyletic, and polyphyletic samples are indicated by circles, triangles, and stars, respectively, and correspond to data points in **Figure 5a**. Samples obtained at the first, second, and third time points from patient 7 are indicated by number. Patients with prevalence patterns consistent with multiple reseeding events and/or extensive mixing (patients 3 and 10) are underlined. Fully quantitative prevalences are given in **Supplementary Table 9**.
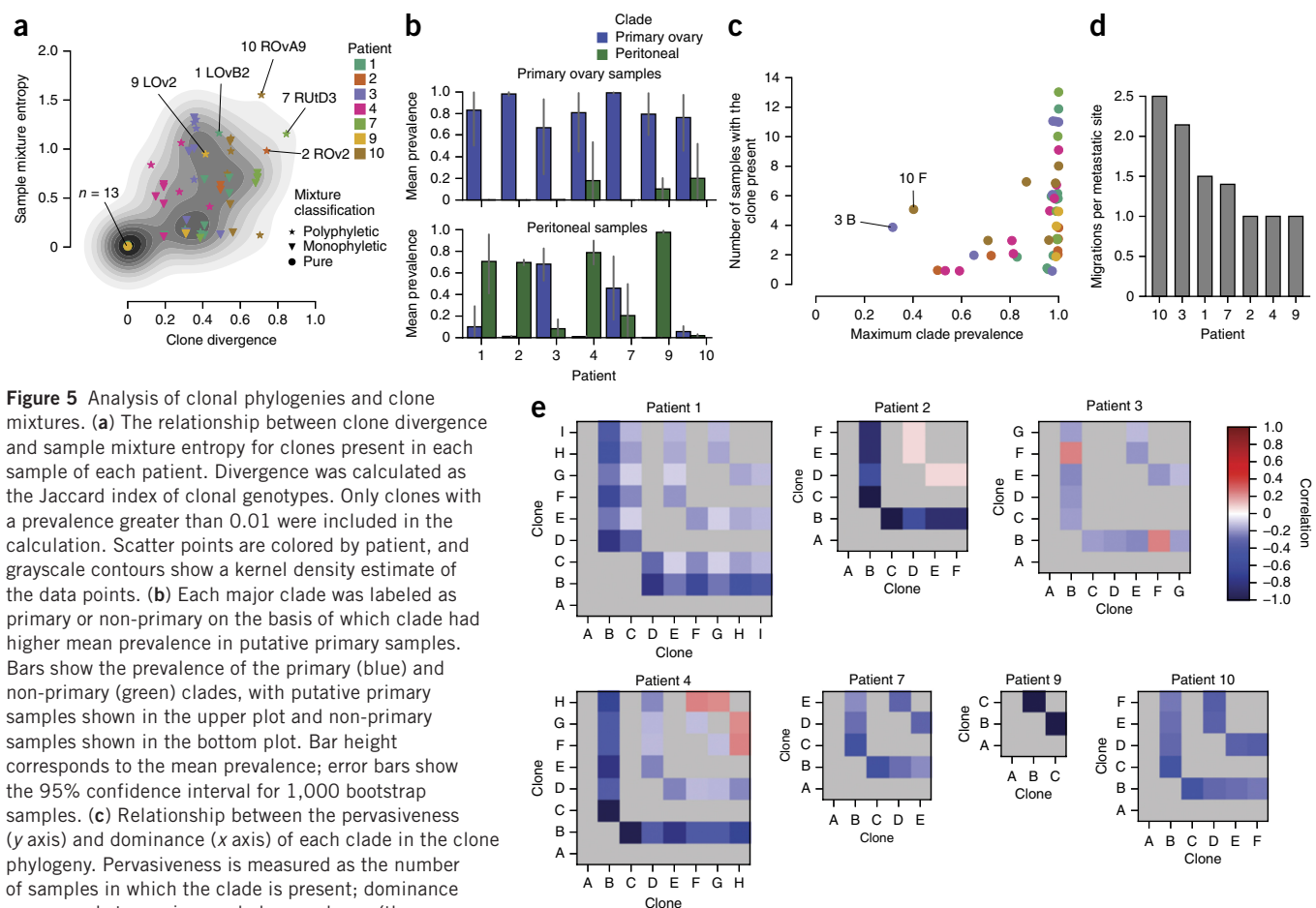
Figure 5 Analysis of clonal phylogenies and clone mixtures. (**a**) The relationship between clone divergence and sample mixture entropy for clones present in each sample of each patient. Divergence was calculated as the Jaccard index of clonal genotypes. Only clones with a prevalence greater than 0.01 were included in the calculation. Scatter points are colored by patient, and grayscale contours show a kernel density estimate of the data points. (**b**) Each major clade was labeled as primary or non-primary on the basis of which clade had higher mean prevalence in putative primary samples. Bars show the prevalence of the primary (blue) and non-primary (green) clades, with putative primary samples shown in the upper plot and non-primary samples shown in the bottom plot. Bar height corresponds to the mean prevalence; error bars show the 95% confidence interval for 1,000 bootstrap samples. (**c**) Relationship between the pervasiveness (*y* axis) and dominance (*x* axis) of each clade in the clone phylogeny. Pervasiveness is measured as the number of samples in which the clade is present; dominance corresponds to maximum clade prevalence (the sum of the prevalence for a clone and its descendants) across samples in which the clade is present. Scatter points are colored by patient. (**d**) Bar plot of the number of migration events per site for each patient's most parsimonious migration scenario. (**e**) Correlations of sample prevalences for clone pairs. Only correlations for clone pairs from divergent lineages are shown; those for clones with an ancestral–descendant relationship were not considered (gray).

in patients. Patients 3 and 10 were enriched for subdominant clones that were present at multiple intraperitoneal sites—ROvA9 and LFTB2 for clone B of patient 10 and CDSB1, Om1, ClnE2, and ROvA7 for clone B of patient 3 (**Fig. 4**)—consistent with multiple reseeding events and extensive intraperitoneal mixing. Furthermore, the prevalence of clone B in patient 3 was found to be correlated across samples with the prevalence of clone F, suggesting that these polyphyletic clones migrated together or reseeded multiple sites (**Fig. 5e**). By contrast, all pairwise correlations of clones for patients 1, 7, and 9 were negative (**Fig. 5e**), suggesting that the majority of sites were populated by a single clone or a monophyletic mixture. The spreading and invasion capacity of clones was decidedly non-uniform, indicating that there is widespread variation in the propensity of clones to mix and invade. Together, the variation in sample diversity and clonal migration patterns define two distinct modes of intraperitoneal mixing across patients, corresponding to patients with a single site permissive of clonal diversity and monoclonal seeding and patients exhibiting extensive polyclonal mixing and reseeding.

Mutation of ovarian cancer genes in divergent clones

We next analyzed the distribution of mutations affecting the known driver genes *TP53*, *BRCA1* and *BRCA2*, *RB1*, *NF1*, *PTEN*, and *RAD51B* across the clonal phylogenies to determine the timing of acquisition and clone specificity of these events. *TP53* mutation was detected in all clones (**Supplementary Table 10**), accompanied by loss of heterozygosity (LOH) of chromosome 17 in all whole-genome sequencing discovery samples (**Supplementary Figs. 7–13**), indicating an ancestral role for *TP53* mutation, consistent with previous studies[1–3,22]. Additional ancestral changes included mutations of *CDK12* in patient 1, *BRCA2* (stop codon) in patient 4, and *FAT3* in patient 10 (**Fig. 1a** and **Supplementary Table 10**). We observed many clade-specific (and, therefore, evolutionarily 'late') driver events, including clone-specific homozygous segmental deletions affecting *CDKN2A* and *WWOX* in the right ovary clones of patient 1, *MAP2K4* in the ROv2 clone of patient 10, and *LRP1B* in the non-ROv2 clones of patient 10 (**Supplementary Figs. 25–29**). In patient 4, homozygous deletion of *ANKRD11* was detected exclusively in the LPvS sample. In patient 7, homozygous deletion of *NF1* was specific to right utero-sacral samples from time point 1 and brain metastasis samples from time point 2 (**Supplementary Fig. 30**). Clone-specific focal high-level amplification of *ERBB2* was identified in patient 9 (**Supplementary Figs. 31–33**), with single-nucleus sequencing confirming that the clone harboring this amplification was present in the primary ovary site and ubiquitous in all peritoneal metastases (**Fig. 3c**).

Convergent evolution of the structural genome

Finally, we determined the presence of convergent evolution resulting from copy number alterations in specific clones. Five patients showed
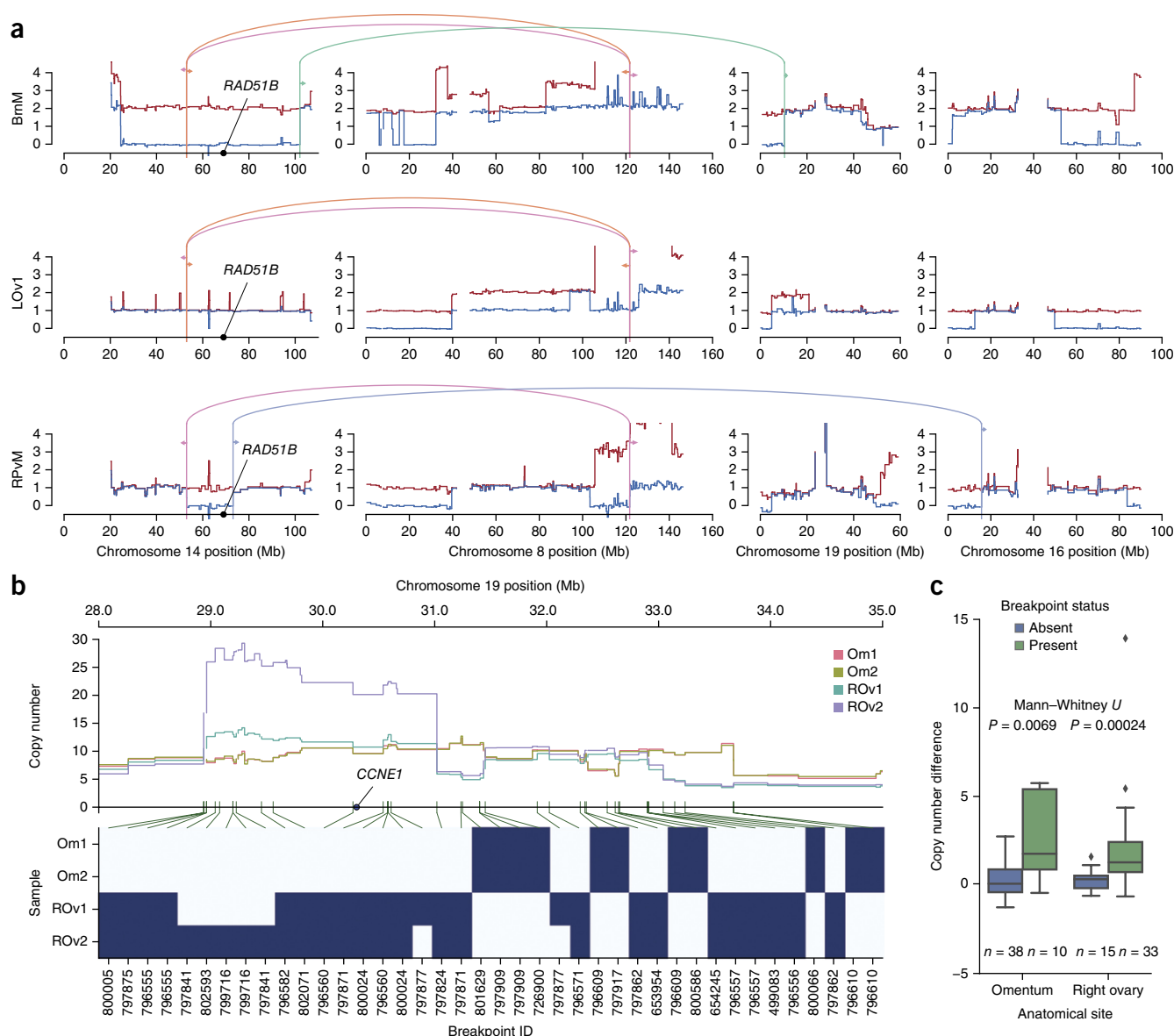
Figure 6 Examples of convergent evolution of copy number. (**a**) Copy numbers for major (red) and minor (blue) alleles for chromosome 14 of patient 7, with the positions of experimentally validated breakpoints indicated (colored arcs). The brain metastasis (BrnM) and right pelvic mass (RPvM) recurrences exhibit LOH of distinct segments of chromosome 14, as corroborated by the presence of distinct sets of breakpoints. (**b**) The top plot shows total copy number for omentum (Om1, Om2) and right ovary (ROv1, ROv2) samples in patient 2. The bottom plot shows the presence (blue) or absence (white) of rearrangement breakpoints, as identified by analysis of discordant paired-end reads, with break ends in the *CCNE1* amplified region. Green lines show correspondence between rearrangement breakpoints in the heat map and the positions of break ends, highlighting the association between breakpoints and copy number transitions. (**c**) Break ends are associated with significantly greater copy number changes in samples in which they are predicted to be present versus those in which they are absent. Shown is the distribution of the difference in copy number between segments on the right and left of each break end, correcting for break end orientation. A greater copy number difference indicates a greater copy number change induced by the breakpoint. For each break end in each sample, we predicted whether the break end existed in that sample. We then compared the copy number difference in each sample for break ends that were predicted to be absent (left) or present (right) in those samples.

evidence of the independent acquisition of breakpoints leading to LOH events or amplifications affecting known cancer-related genes. The total length of the segments classified as LOH in any sample ranged from 1,100 to 1,600 Mb across patients (**Supplementary Table 11**). In patients for whom analysis was not confounded by divergent mixtures, between 1.5 and 7.2% of LOH segments, by length, could not have originated on a single branch of the clone phylogeny and are thus the result of convergent LOH (**Supplementary Table 11**).

Clone-specific rearrangements leading to LOH of *RAD51B* were identified in three patients. For patient 10, multiple *RAD51B*

rearrangements were detected in samples composed of descendant clones but were absent from the two samples (LFTB2 and OmC1) dominated by the ancestral (or ancestrally divergent) clone. In patient 3, *RAD51B* was affected by two rearrangements, including one event common to all samples and a ROv2-specific event. Patient 7 exhibited convergent evolution toward LOH of chromosome 14, with two distinct events independently deleting the reference allele of a SNP affecting a splice site of *RAD51B* (**Fig. 6a**). Both *RAD51B* LOH clones coexisted in the RUtD3 sample at time point 1, with clone B eventually dominating the brain metastasis at time point 2, clone D dominating

the right pelvic samples at time point 3, and *RAD51B*-heterozygous clones extinguished at both the later time points.

Five patients exhibited chromosome-arm-level amplification affecting known oncogenes: *CCNE1* in patients 2, 9, and 10; *KRAS* in patient 3; and *MYC* in patient 7. The predicted numbers of copies were highly variable among discovery samples (between 10 and 20 copies of *CCNE1* in patient 2, for example, as confirmed by FISH analysis) (**Supplementary Figs. 34** and **35**, and **Supplementary Table 12**). Unbalanced intrachromosomal rearrangements were the predominant class of events (**Supplementary Figs. 36–42**). The distributions of rearrangement breakpoints throughout the samples were indicative of a dynamic process of progressive amplification conferring distinct sets of breakpoints to specific clones. For patient 3, the presence/absence profiles of breakpoints associated with *KRAS* amplification were moderately heterogeneous across samples, although shared clonal mixtures in this patient prevented unambiguous assignment of breakpoints to clones (**Supplementary Fig. 43**). In patient 7, several breakpoints associated with *MYC* amplification were specific to a subset of samples (**Supplementary Fig. 44**). In patient 2, the distribution of breakpoints found within the *CCNE1*-amplified region recapitulated the distinct clonal lineages in right ovary and omentum sites identified by maximum likelihood for both SNVs and breakpoints (**Fig. 6b** and **Supplementary Figs. 2** and **5**). For all 28 breakpoints within the *CCNE1*-amplified region (**Supplementary Table 13** and **Supplementary Note**), the posterior probability of originating on the ancestral branch was <0.001, indicating that acquisitions were sample specific and, therefore, occurred independently (**Supplementary Figs. 45–47**). Furthermore, break ends within the *CCNE1*-amplified region coincided with significantly greater copy number differences in samples in which they were predicted to be present, as compared to samples in which they were predicted to be absent (**Fig. 6c**), confirming that sample-specific breakpoints induced sample-specific copy number changes (**Supplementary Note**).

DISCUSSION

Our findings challenge the notion that the intraperitoneal spreading of cell populations should result in widespread clonal mixing within a local potential space, as only a minority of patients exhibited patterns consistent with extensive migration and reseeding of polyclonal mixtures at multiple sites. This observation implies that the selection potential of the peritoneal cavity is likely non-uniform because of specific microenvironmental properties. Despite the lack of treatment-induced selective pressures, established clones present at diagnosis may have been selected by the intrinsic topology of the intraperitoneal fitness landscape. Intriguingly, we found that clones could be categorized by three modes of spreading behavior, providing a rationale for linking genotypic mechanisms to the differential spreading capacity of individual clones. Further study on larger series will be needed to confirm the generalizability of these findings. Ultimately, comparisons of primary and relapse samples over time would show whether clones with a high degree of spreading capacity could dominate in the presence of combined constraints from microenvironmental and therapeutic selective pressures. For example, in patient 7, we studied two recurrences. Notably, diagnostic specimens already contained clones with properties of intrinsic resistance before therapeutic intervention, leading to brain metastasis and right pelvic mass recurrences, respectively, at the two later time points. Whether such diverse pretreatment samples could be selectively targeted remains an open question.

Together, the properties of HGSOvCa evolution and the distinct modes of local intra-abdominal spread present an opportunity to study how these patterns relate to treatment outcomes. We note that the properties uncovered in our analysis have anecdotal association with clinical outcomes. Patients whose divergent clones were dominated by the age-related mutational signature (and not the mutational signature corresponding to homologous recombination deficiency; patients 2 and 9) had the poorest survival outcomes, and both died of disease (**Table 1**). Patients 2 and 9 were also characterized by monoclonal seeding patterns, exhibiting the fewest migrations per site. By contrast, the patients with the most migrations per site (patients 1, 3, and 10) did not succumb to disease, with patients 1 and 10 showing no evidence of disease after treatment within the follow-up period. Our study design is constrained by the limitations of surgical sampling and is likely characterizing a subset of the clonal diversity present in each patient. Nevertheless, with a median of $n = 10$ samples per patient, we always detected a sample representing diversity, allowing for interpretation of spatial clonal dynamics in the other sites. Furthermore, we note that most conclusions about ovarian cancer biology are derived from studying single specimens. Our observations and those of others[5,23] emphasize that the analysis of single samples provides a dramatically restricted view of tumor biology. The analytical and experimental approaches presented here will inform the study of larger series, including rapid autopsies[2,12], where clones will have escaped numerous therapeutic interventions. In addition, preclinical model system[24] and clinical trial studies assaying response to pharmacologic and biologic agents will be informed by the evolutionary properties reported here.

Our results identify several properties of HGSOvCa evolution, including loss of mutations due to copy number deletion, convergent evolution of the structural genome, and temporal 'activation' of mutational processes. Loss of mutations and structural convergence will affect analytical approaches for reconstructing evolutionary histories from multiregion sequencing experiments. Many genomically unstable tumors like HGSOvCa will likely exhibit properties that violate the assumptions imposed by perfect and persistent phylogeny. Ignoring these additional complexities will inevitably lead to the erroneous inference of both sample trees and the temporal ordering of mutations if convergence and mutation loss are not taken into account. Notably, our observation of convergent amplification of driver genes such as *CCNE1* raises new hypotheses that clade-specific clones harboring independently acquired amplifications would respond differently to intervention.

In summary, we have presented a detailed view of HGSOvCa clonal dynamics within the context of intraperitoneal local spread. Our results provide an impetus for dissecting the specific biological mechanisms underlying the interaction of fitness properties and the evolutionary selection governing the invasive capacity of ovarian cancers.

URLs. BC Cancer Agency Cancer Management Guidelines, http://www.bccancer.bc.ca/health-professionals/professional-resources/cancer-management-guidelines; Dollo source code, https://bitbucket.org/dranew/dollo; interactive visualization of patient-specific clonal maps, http://compbio.bccrc.ca/research/tumour-evolution/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequencing is available for download from the European Genome-phenome Archive (EGA) under study accession EGAS00001000547.

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

AUTHOR CONTRIBUTIONS
A.M. and A.R. performed algorithm development and software implementation and led the data analysis. G.H., A.W.Z., K.S., C.S., J.R., and A.B. performed data analysis and bioinformatics. E.L., T.M., J.B., D.Y., A.W., and J.K. performed single-nucleus sequencing. L.M.P., S.K., J.S., and W.Y. performed sample preparation and validation experiments. M.A.S. and C.B.N. performed data visualization. A.K., H.L.C., J.H., and N.M. performed immunohistochemistry and FISH analyses. R.M., A.J.M., and M.A.M. performed library construction and genome sequencing. C.B.G. and D.G.H. analyzed molecular and histological pathology. A.B.-C. contributed to algorithm development. J.N.M. performed surgery and tumor banking and was a clinical leader. S.C.M. edited the manuscript. A.M., A.R., S.A., and S.P.S. wrote the manuscript. S.A. oversaw experimental design and single-cell sequencing. S.P.S. is the senior responsible author and conceived the project and provided oversight.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. 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).
2. Patch, A.-M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
3. 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).
4. Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24**, 1881–1893 (2014).
5. Schwarz, R.F. *et al.* Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med.* **12**, e1001789 (2015).
6. Jacobs, I.J. *et al.* Clonal origin of epithelial ovarian carcinoma: analysis by loss of heterozygosity, *p53* mutation, and X-chromosome inactivation. *J. Natl. Cancer Inst.* **84**, 1793–1798 (1992).
7. Khalique, L. *et al.* The clonal evolution of metastases from primary serous epithelial ovarian cancers. *Int. J. Cancer* **124**, 1579–1586 (2009).
8. Lengyel, E. Ovarian cancer development and metastasis. *Am. J. Pathol.* **177**, 1053–1064 (2010).
9. Campbell, P.J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
10. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
11. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
12. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
13. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
14. Gawad, C., Koh, W. & Quake, S.R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl. Acad. Sci. USA* **111**, 17947–17952 (2014).
15. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
16. Alekseyenko, A.V., Lee, C.J. & Suchard, M.A. Wagner and Dollo: a stochastic duet by composing two parsimonious solos. *Syst. Biol.* **57**, 772–784 (2008).
17. Ryder, R.J. & Nicholls, G.K. Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. *J. R. Stat. Soc. Ser. C Appl. Stat.* **60**, 71–92 (2011).
18. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
19. Blei, D.M., Ng, A.Y. & Jordan, M.I. Latent dirichlet allocation. *J. Machine Learning Res.* **3**, 993–1022 (2003).
20. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
21. Roth, A. *et al.* Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat. Methods* http://dx.doi.org/10.1038/nmeth.3867 (2016).
22. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
23. Hoogstraat, M. *et al.* Genomic and transcriptomic plasticity in treatment-naive ovarian cancer. *Genome Res.* **24**, 200–211 (2014).
24. Cooke, S.L. & Brenton, J.D. Evolution of platinum resistance in high-grade serous ovarian cancer. *Lancet Oncol.* **12**, 1169–1174 (2011).

ONLINE METHODS

Patient consent and surgery. Ethical approval was obtained from the University of British Columbia (UBC) Ethics Board. Women undergoing debulking surgery (primary or recurrent) for carcinoma of ovarian, peritoneal, and/or fallopian tube origin were approached for informed consent for the banking of tumor tissue. Cases of HGSOvCa where more than one sample was collected in different anatomical locations (for example, different locations within the ovary and omentum) or where material was available at different time points (for example, at primary surgery and at recurrence) were chosen for this analysis. Clinicopathological and outcome data were collected by chart review. Consistent with practices at UBC and the British Columbia Cancer Agency, all patients with high-grade serous cancer are referred to the hereditary cancer clinic and offered genetic testing for *BRCA1* and *BRCA2* mutations[25,26].

Specimen preservation and histological evaluation. In cases identified as having high-grade serous histology, multiple tissue samples were obtained from the primary ovarian tumor and metastatic sites where adequate tumor volume permitted. When the ovary was pathologically enlarged, samplings were taken from up to five different areas, with an effort made to equally space sampling sites while staying within the bounds of grossly apparent tumor tissue. Each sampling was cut into three pieces, yielding two end pieces for cryovials and a middle portion placed in 10% buffered formalin. All paraffin-embedded blocks, including formalin-fixed tumor samples and molecularly fixed fallopian tubes, were sectioned and stained with hematoxylin and eosin before expert histopathological review (C.B.G.) to confirm the presence of high-grade serous carcinoma.

Library construction and sequencing. Samples for a total of ten patients were submitted for library construction and sequencing. Sample size was determined by the availability of resectable, cryopreserved tissue and also by DNA quality. Samples from patients 5 and 6 were excluded because of low tumor cellularity, as indicated by the absence of SNVs and copy number changes. Patient 8 was excluded as contaminated because of the high number of somatic variants present in dbSNP. For all tumor and normal samples, DNA extraction was followed by library construction and sequencing using Illumina HiSeq 2000 whole-genome shotgun v3 chemistry with paired-end 100-bp reads. Samples were sequenced to an average of 30× coverage (**Supplementary Table 2**).

Targeted deep sequencing. To resolve the clonal composition of each sample, we performed targeted deep sequencing of 300 loci per patient with multiplex PCR amplicons sequenced to a median of 7,700× coverage (**Supplementary Tables 10**, **14**, and **15**). These data were used to confirm the predicted presence/absence patterns of SNVs and breakpoints from whole-genome sequencing and to infer the clonal prevalence of SNV clusters using PyClone[27] (**Supplementary Table 16**). We defined a genotype as a set of present or absent PyClone SNV clusters and breakpoints. The set of aberrations present at a cellular prevalence of >50% in a sample provides an estimate of the genotype of the clone dominant in that sample. Furthermore, the sample tree provides an estimate of the phylogeny relating dominant clones. Under these assumptions, we estimated clonal genotypes and clone phylogenies.

Targeted single-cell sequencing. Individual nuclei from each sample were subjected to multiplex PCR reactions on Fluidigm Access arrays with libraries sequenced on Illumina MiSeq machines as previously reported[13]. We designed primers for 48 loci for patients 2 and 9 and for 96 loci for patient 3 (**Supplementary Table 17**). We sequenced 1,680 cells in total to a median depth of 100× per locus (**Table 2** and **Supplementary Tables 18** and **19**). Resulting data produced read counts for each targeted locus in each sequenced nucleus. Nuclei were clustered, and clonal genotypes were inferred using the SCG model[21] version 0.3.0; see the **Supplementary Note** for parameter settings. In each patient, heterozygous SNPs predicted to be rendered homozygous reference in tumors from whole-genome sequencing analysis were used to distinguish normal nuclei from tumor nuclei.

Additional methods are described in detail in the **Supplementary Note**.

25. Schrader, K.A. *et al*. Germline *BRCA1* and *BRCA2* mutations in ovarian cancer: utility of a histology-based referral strategy. *Obstet. Gynecol.* **120**, 235–240 (2012).
26. McAlpine, J.N. *et al*. *BRCA1* and *BRCA2* mutations correlate with *TP53* abnormalities and presence of immune cell infiltrates in ovarian high-grade serous carcinoma. *Mod. Pathol.* **25**, 740–750 (2012).
27. Roth, A. *et al*. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).