Ongoing genome doubling promotes evolvability and immune dysregulation in ovarian cancer

Andrew McPherson[1,2,†,*], Ignacio Vázquez-García[1,2,3,*], Matthew A. Myers[1,2], Matthew Zatzman[1,2], Duaa Al-Rawi[1,2,4], Adam Weiner[1,2,5], Samuel Freeman[1,2], Neeman Mohibullah[6], Gryte Satas[1,2], Marc J. Williams[1,2], Nicholas Ceglia[1,2], Allen W. Zhang[7], Jun Li[8,9], Jamie L.P. Lim[1], Michelle Wu[1,2], Seongmin Choi[1,2], Eliyahu Havasov[1,2], Diljot Grewal[1,2], Hongyu Shi[1], Minsoo Kim[1,2,5], Roland Schwarz[10,11], Tom Kaufmann[11,12], Khanh Ngoc Dinh[3,13], Florian Uhlitz[1,2], Julie Tran[1,2], Yushi Wu[6], Ruchi Patel[6], Satish Ramakrishnan[6], DooA Kim[6], Justin Clarke[6], Hunter Green[14], Emily Ali[1,2,14], Melody DiBona[8,9], Nancy Varice[15], Ritika Kundra[16], Vance Broach[15], Ginger J. Gardner[15], Kara Long Roche[15], Yukio Sonoda[15], Oliver Zivanovic[15,17], Sarah H. Kim[15], Rachel N. Grisham[4,18], Ying L. Liu[4,18], Agnes Viale[16], Nicole Rusk[1,2], Yulia Lakhman[18,19], Lora H. Ellenson[14], Simon Tavaré[3], Samuel Aparicio[20], Dennis S. Chi[15], Carol Aghajanian[4], Nadeem R. Abu-Rustum[15,21], Claire F. Friedman[4,18], Dmitriy Zamarin[22], Britta Weigelt[14], Samuel F. Bakhoum[8,9], Sohrab P. Shah[1,2,†]

[1] Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[2] The Halvorsen Center for Computational Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[3] Irving Institute for Cancer Dynamics, Columbia University, New York, NY, 10027, USA
[4] Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[5] Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY, USA
[6] Integrated Genomics Operation, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
[7] Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada
[8] Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[9] Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[10] Institute for Computational Cancer Biology (ICCB), Center for Integrated Oncology (CIO), Cancer Research Center Cologne Essen (CCCE), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[11] BIFOLD - Berlin Institute for the Foundations of Learning and Data, Berlin, Germany
[12] Berlin Institute for Molecular Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
[13] Department of Statistics, Columbia University, New York, NY, 10027, USA
[14] Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[15] Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

[16] Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
[17] Department of OB/GYN, University Hospital Heidelberg, Germany
[18] Weill Cornell Medical Center, New York, NY 10065, USA
[19] Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
[20] Department of Molecular Oncology, BC Cancer, Vancouver, Canada
[21] Department of OB/GYN, Weill Cornell Medical College, New York NY
[22] Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

* equal contribution

† Corresponding authors:

Sohrab P. Shah (shahs3@mskcc.org)

Andrew McPherson (mcphera1@mskcc.org)

Keywords: whole-genome doubling, high-grade serous ovarian cancer; genomic instability; innate immunity; immune phenotypes; single cell whole genome

ABSTRACT

Whole-genome doubling (WGD) is a critical driver of tumor development and is linked to drug resistance and metastasis in solid malignancies. Here, we demonstrate that WGD is an ongoing mutational process in tumor evolution. Using single-cell whole-genome sequencing, we measured and modeled how WGD events are distributed across cellular populations within tumors and associated WGD dynamics with properties of genome diversification and phenotypic consequences of innate immunity. We studied WGD evolution in 65 high-grade serous ovarian cancer (HGSOC) tissue samples from 40 patients, yielding 29,481 tumor cell genomes. We found near-ubiquitous evidence of WGD as an ongoing mutational process promoting cell-cell diversity, high rates of chromosomal missegregation, and consequent micronucleation. Using a novel mutation-based WGD timing method, `doubleTime`, we delineated specific modes by which WGD can drive tumor evolution: (i) unitary evolutionary origin followed by significant diversification, (ii) independent WGD events on a pre-existing background of copy number diversity, and (iii) evolutionarily late clonal expansions of WGD populations. Additionally, through integrated single-cell RNA sequencing and high-resolution immunofluorescence microscopy, we found that inflammatory signaling and cGAS-STING pathway activation result from ongoing chromosomal instability and are restricted to tumors that remain predominantly diploid. This contrasted with predominantly WGD tumors, which exhibited significant quiescent and immunosuppressive phenotypic states. Together, these findings establish WGD as an evolutionarily 'active' mutational process that promotes evolvability and dysregulated immunity in late stage ovarian cancer.

INTRODUCTION

Whole-genome doubling (WGD) is found in >30% of solid cancers, leading to increased rates of metastasis, drug resistance and poor outcomes[1]. Often observed on a background of *TP53* mutation, genome doubling leads to increased chromosomal instability (CIN) and karyotypic diversification[2]. Several studies have reported that the fitness advantage of genome-doubled cells is conferred through its buffering effect on deleterious mutations[2–4]. *In vitro* studies indicate that genome doubling also leads to major phenotypic consequences, such as chromatin and epigenetic changes[5], replication stress[6] and cellular quiescence[5,6]. Previous studies of WGD in patient tumors used bulk whole-genome sequencing (WGS), which requires computational reconstruction of somatic evolutionary histories[7], making cellular diversity difficult to infer. These analyses have tended to cast WGD as an early event in tumor evolution[8], restricting its occurrence and mechanistic significance to an etiologic role[7]. However, live-cell analysis has previously suggested that errors in chromosome segregation often lead to cytokinesis failure and the ongoing generation of polyploid cells[9], suggesting WGD might be an ongoing process during tumor evolution. Recent reports from *in vitro* and PDX models have demonstrated that temporal and evolutionary dynamics of genome doubling can be captured at single-cell resolution[10,11]. However, in the patient setting, how dynamical properties of genome doubling drive evolution and phenotypic state changes at the time of clinical presentation remains understudied. We contend that applying single cell approaches to clinical samples therefore opens the opportunity to ask new questions of how genome doubling evolution drives genomic diversity and phenotypic cellular states in the patient context.

We used single-cell whole-genome sequencing to study WGD in the context of high-grade serous ovarian cancer (HGSOC), a tumor type with ubiquitous *TP53* mutation and frequent WGD. We analyzed 65 untreated HGSOC samples at the time of diagnosis from 40 patients with single-cell whole-genome sequencing (29,481 tumor cell genomes) and site-matched immunofluorescent staining for markers of micronuclei and DNA sensing, known byproducts of chromosome segregation defects. Using this multi-modal approach, we conclude WGD is an ongoing mutational process which promotes evolvability through cell-cell diversity, high rates of CIN, and pervasive co-occurrence of cells with heterogeneous ploidy states within the same tumor. We delineated three modes of WGD evolution: (i) early fixation of a single event, (ii) late fixation of multiple independent WGD events, and (iii) emergence of late WGD clones. By linking genomic measurements with cellular phenotypes in previously generated site-matched single-cell RNA sequencing data[13] we found that microenvironmental inflammatory signaling remains active in tumors that remain predominantly diploid in contrast to enriched quiescent and immunosuppressive states in predominantly WGD tumors. Our findings therefore point to WGD as a critical co-variate of inflammatory signaling and immunosuppression. Given our findings are derived from clinical samples at disease presentation,

113 we suggest our study should further motivate and inform novel therapeutic targeting of WGD and
114 CIN[12,13].

115 **RESULTS**

116 **Ovarian cancer patient cohort**

117 Surgical specimens (*n*=65) from treatment-naive HGSOC patients (*n*=40) were collected from
118 multiple sites during primary debulking surgery or laparoscopy, as previously described[14] (**Fig. 1A**,
119 **Extended Data Fig. 1A, Methods**). Patients were confirmed as advanced HGSOC by gynecologic
120 pathologists. Sampled sites included adnexa (i.e. ovary and fallopian tube), omentum, peritoneum,
121 bowel, and other intraperitoneal sites (**Extended Data Fig. 1B**). Clinical characteristics of all patients
122 are summarized in **Extended Data Fig. 1B** and **Supp. Tab. 1**. Somatic and germline driver
123 mutations were determined by MSK-IMPACT clinical sequencing, including ubiquitous somatic *TP53*
124 loss, somatic and germline *BRCA1/2* loss, somatic *CDK12* mutation and somatic *CCNE1*
125 amplification (**Extended Data Fig. 1B**)[14]. Mutational signatures derived from whole-genome
126 sequencing included homologous recombination-deficient (HRD)-Dup (*BRCA1* mutant-like) and
127 HRD-Del (*BRCA2* mutant-like) cases, as well as HR-proficient foldback inversion-bearing (FBI) and
128 tandem duplicator (TD) tumors (18 HRD-Dup, 8 HRD-Del, 13 FBI, 1 TD) using integrated point
129 mutation and structural variations, as previously described[11,14,15].

130 **Single-cell whole-genome sequencing and orthogonal phenotypic assays**

131 Tumor-derived single-cell suspensions were flow-sorted to remove $CD45^+$ immune cells, then
132 subject to single-cell whole-genome sequencing (scWGS) using the Direct Library Preparation
133 protocol[16] (DLP+, **Methods**, **Supp. Tab. 2**). A total of 53,005 single-cell whole-genomes (median
134 1,345 per patient) were generated with median coverage depth of 0.078 per cell and median
135 coverage breadth of 0.073 (**Extended Data Fig. 2A-B**) with 29,481 genomes admitted into analysis
136 following quality control (**Methods**). In addition, whole-slide H&E and immunofluorescence (IF)
137 images from adjacent formalin-fixed paraffin-embedded (FFPE) tissue sections were obtained for 37
138 out of 40 patients (**Supp. Tab. 2**). IF sections were assessed for DNA sensing mechanisms and
139 genome sequencing-independent readouts of chromosomal instability (DAPI, cGAS) (**Methods**). In
140 addition, we leveraged previously-generated single-cell RNA sequencing (scRNA-seq) data from
141 both $CD45^+$ and $CD45^-$ compartments of 32 patients (52 scRNA samples site matched to scWGS)[14],
142 enabling genotype-phenotype analyses of these tissues. Together the dataset comprises a single-
143 cell resolution multi-modal measurement of aneuploidies, genomic and chromosomal instability, and
144 their cell-intrinsic and tumor microenvironment phenotypic readouts in HGSOC patient samples.

145 **Whole-genome doubling at single-cell resolution**

146 Using the 29,481 high-quality single cell cancer genomes, we first investigated the distribution of
147 WGD states across our cohort and within each tumor. We inferred the number of WGD events in the
148 evolutionary history of each cell based on allele-specific copy-number profiles[17,18] (**Fig. 1B**). The
149 distribution of allele-specific copy number features showed clear separation between WGD states,
150 permitting assignment of per-cell WGD multiplicities of 0 (0×WGD, 46% of cells), 1 (1×WGD, 53%)
151 and 2 (2×WGD, 1%) (**Extended Data Fig. 2H-I**). The number of WGD events per cell correlated with
152 both cell size measured through the optical components of DLP+ (**Extended Data Fig. 2J**), and
153 mitochondrial copy number (**Extended Data Fig. 2K**), providing orthogonal validation based on
154 known correlates of nuclear genome scaling[16,19].

155 We then analyzed intra-patient WGD states at single-cell resolution, finding pervasive heterogeneity
156 in WGD states within tumors. For example, patient 045 (**Fig. 1C**) simultaneously harbored a minority
157 of 0×WGD (1%, **Fig. 1D**), a majority of 1×WGD (97%, **Fig. 1E**), and a minor fraction of 2×WGD cells
158 (2%, **Fig. 1F**). Surprisingly, tumors with co-existing WGD states were present in 36/40 patients,
159 including 31/32 of the patients with >200 tumor cells (**Fig. 1G**). In total, 5% of all tumor cells (*n*=1,481
160 cells), were part of non-dominant WGD states across the cohort (median of 2.6% of cells per patient;
161 **Extended Data Fig. 2L**). As each patient's tumor was typically dominated by a single WGD state
162 (the dominant state comprising >85% of cells for 38/40 patients), we dichotomized each tumor as
163 either Prevalent WGD: harboring >50% of 1×WGD or 2×WGD cells (26/40 patients); or Rare WGD:
164 with ≥50% 0×WGD cells (14/40 patients). Prevalent WGD patients comprised 60% of the cohort,
165 were older, and were enriched for FBI and HRD-Del mutation signature patients, consistent with
166 previous bulk genome sequencing studies[7,14] (**Extended Data Fig. 2M-P**). Thus, while average
167 signals corroborate previous bulk estimates of WGD prevalence across patients[17], single cell
168 analysis established that WGD exists as a distribution over 0×WGD, 1×WGD, 2×WGD cells within
169 tumors, with at least 2 co-existing WGD states observed in the majority of patients.

170 **Evolutionary histories of WGD clones**

171 We next analyzed evolutionary histories of WGD and non-WGD clones including timing the origin of
172 single or multiple WGD expansions within each patient. We developed `doubleTime`, a multi-step
173 computational approach that (i) uses a scWGS mutation caller to predict SNVs (`Articull`,
174 manuscript in prep.), (ii) identifies SNV clones using SBMClone[20], (iii) constructs a clone phylogeny,
175 (iv) places WGD events on branches of the phylogeny, and (v) infers mutation timing including the
176 relative timing of WGDs on the WGD branches (**Fig. 2A, Methods**). For 23 out of 25 Prevalent WGD
177 patients, a single ancestral WGD event was common to the dominant 1×WGD population of cells
178 (**Fig. 2B**). For two patients (025 and 045) we observed coexisting WGD clones from distinct WGD

events, consistent with lineage divergence in the ancestral diploid population followed by expansion of independent WGD clones (**Fig. 2C, Extended Data Fig. 3A-B**). Within both patients, the multiple WGD events were predicted to occur at approximately the same mutation time in the tumor's life history. Remarkably, for both of these patients, WGD clones coexisted in multiple anatomic sites. All WGD clones were present in both right adnexa and omentum of 025. For patient 045, the left adnexa harbored one of the three of the WGD clones whereas the right adnexa, omentum, and peritoneal tumors were mixtures of all three WGD clones. The remaining 14 patients did not show evidence of expanded WGD clones through SNV analysis (**Fig. 2D**), although all harbored small populations of 1×WGD cells (**Fig. 2D**).

We then investigated the evolutionary timing of WGD clonal expansions to determine if they fixed early in tumorigenesis, or whether clonal expansions of WGD cells occurred throughout their life histories. Prevalent WGD patients exhibited increased mutation time from fertilization to surgical resection (total C>T CpG burden) vs Rare WGD patients, similar to WGD vs non-WGD patients in previous bulk WGS analyses[7] (**Fig. 2E**). However, while bulk genome sequencing studies of ovarian tumors reported early acquisition of WGD[7], we found that later WGD clonal expansions inferred from scWGS were common, with 7/25 WGD events occurring more than 50% of the way through the tumors life history. For three of the late WGD patients (045, 075 and 081), WGD events approximately coincided with the most recent common ancestor (MRCA). (**Fig. 2B-C,E**). These same three patients, in addition to late WGD patient 125, all exhibited extant populations of 0×WGD cells: 17 (0.9%) 0×WGD cells for 045 (**Extended Data Fig. 3A**), 34 (3.9%) 0×WGD cells for 075 (**Extended Data Fig. 3B**), 14 (27%) 0×WGD cells for 125 (**Extended Data Fig. 3C**), and 38 0×WGD cells in the omentum sample of 081 which contained the site-specific WGD expansion (**Fig. 2B**). Thus, in these four patients, more recently expanded WGD clones co-existed with extant cells from the 0×WGD population from which they were derived.

The existence of a substantial fraction of subclonal WGD subpopulations (1×WGD in 0×WGD clones, 2×WGD in 1×WGD clones) across multiple clones was consistent with parallel and ongoing WGD. We investigated if these rare subclonal WGD cells shared common mutations, indicative of late WGD-associated clonal expansions (**Fig. 2B-D,F**). In patient 025, a small subpopulation of 43 2×WGD cells harbored 325 SNVs specific to the 2×WGD cells (**Extended Data Fig. 3D**). Subclonal WGD expansions in patients 031 (7 cells) and 006 (27 cells) were too small to be detected by SNV analysis, but could be identified by copy-number events shared across multiple subclonal WGD cells (**Extended Data Fig. 3E-F**). For other patients, subclonal WGD cells were evenly distributed across multiple clones, indicative of continual WGD across clonal populations. Together, quantifying the evolutionary history of WGD and chromosomal instability in single cells revealed distinct modes of ongoing WGD evolution: (i) diploid tumors with a background rate of unexpanded WGD cells, (ii)

214 tumors with evolutionary late WGD expansions including parallel expansion of multiple WGD clones,
215 and (iii) evolutionary early WGD tumors with a single dominant WGD clone.

216 **Post WGD genomic diversification**

217 Leveraging single-cell-resolution measurements, we next asked how WGD promotes genomic
218 diversification and evolvability. We first quantified cell-to-cell genomic heterogeneity using pairwise
219 nearest-neighbor copy-number distance (NND) for each cell (**Methods**, **Extended Data Fig. 4A**).
220 Mean NND increased with WGD multiplicity and was highest for subclonal WGD populations, with
221 1×WGD populations in Rare WGD patients exhibiting higher mean NND than 1×WGD populations
222 in Prevalent WGD patients, and 2×WGD cells in predominantly 1×WGD tumors exhibiting the highest
223 cell-cell diversity (**Fig. 3A**). Some Prevalent WGD patients exhibited surprising levels of diversity: in
224 8 patients, cells were on average different for 10% of the genome when compared with the most
225 similar cell. The empirical distribution of NND values had a heavy tail (**Extended Data Fig. 4B**) with
226 unexpected enrichment for highly divergent cells with very distinct copy-number profiles. We defined
227 divergent cells as those with NND greater than the 99th percentile of a Beta distribution fit to the
228 NND values (**Fig. 3B**). The CN profiles of divergent cells resembled the previously reported 'hopeful
229 monsters' found in colorectal cancer organoids[21], suggesting they may be the product of unstable
230 tetraploid cells undergoing multipolar mitosis (**Fig. 3C**). When compared with clonal CN profiles,
231 divergent cells harbored elevated rates of whole chromosome and chromosome arm loss, uniformly
232 distributed across the genome (**Extended Data Fig. 4C-D**), accompanied by a significant rate of arm
233 and chromosome nullisomy for both Rare and Prevalent WGD patients (**Extended Data Fig. 4E**).
234 Notably, these divergent cells were present in 39/40 patients (mean 2.8% of cells), with higher rates
235 in Prevalent WGD patients suggesting an increased propensity for abnormal mitoses (**Fig. 3D**).
236 Interestingly, the three patients with large clonal expansion of late WGD (081, 045 and 025), ranked
237 first, fourth and seventh highest in divergent cell fraction. Furthermore, patient 049 had the second-
238 highest divergent cell fraction and the third most recent clonal WGD behind two of the three WGDs
239 in 045. Overall, our data is concordant with previous evidence suggesting WGD cell populations
240 sustain a period of instability following WGD, which can result from an increase in the number of
241 centrosomes[22].

242 To investigate rates of chromosome missegregation events, we computed copy number alterations
243 in each cell (excluding divergent cells) accrued since its immediate ancestor in a phylogeny inferred
244 for each patient (**Fig. 3E, Extended Data Fig. 4F, Methods**). This enabled inference of rates of cell-
245 specific (and therefore most recent) copy number changes. The rate (counts per cell) of gains and
246 losses of whole chromosomes, chromosome arms, and segments (>15MB) increased with WGD
247 multiplicity across all event types (**Fig. 3F**). We recomputed a ploidy-adjusted version of the gain and
248 loss rates that accounted for the increased opportunity for copy-number events in higher ploidy cells.

249 The ploidy-adjusted rates (counts per cell per GB) showed similar increases, highlighting that the
250 rate differences were not merely a function of increased chromosome number but were instead
251 indicative of systemic changes in post-WGD cells (**Extended Data Fig. 4G, Methods**). 1×WGD
252 subpopulations had higher ploidy-adjusted rates of chromosome losses in Rare WGD patients than
253 Prevalent WGD patients, suggesting more recently emerging WGD cells were more prone to
254 chromosome loss events, or that early WGD populations had stabilized. For instance, ploidy-
255 adjusted chromosome losses were 4.1 times more abundant in Rare WGD 1×WGD cells compared
256 to Rare WGD 0×WGD cells (p=5.6×10^{-4}, Mann-Whitney U test), 2.3 times more abundant in
257 Prevalent WGD 1×WGD cells compared to Rare WGD 0×WGD cells (p=5.6×10^{-4}, Mann-Whitney U
258 test), and 1.8 times more abundant in Rare WGD 1×WGD cells compared to Prevalent WGD 1×WGD
259 cells (p=0.016, Mann-Whitney U test).

260 One of the phenotypic consequences of chromosome segregation errors is the formation of
261 micronuclei, which are chromosome or chromosome arm containing structures that are distinct from
262 the primary nucleus during interphase. Micronuclear envelopes are rupture-prone, often exposing
263 their enclosed genomic double-stranded DNA (dsDNA) to the cytoplasm[23–25], leading to activation of
264 innate immune signaling driven by the cytosolic dsDNA sensing cGAS-STING pathway[26]. We asked
265 whether the propensity of chromosome missegregation correlates with micronuclei formation via
266 high-resolution immunofluorescence microscopy of cGAS and DAPI staining on FFPE sections, site-
267 matched to scWGS datasets. We used a deep learning approach to automatically detect primary
268 nuclei (PN) and cGAS^{+} ruptured micronuclei (MN), enabling whole-slide quantification of MN-to-PN
269 ratios at scale (1,779,351 PN and 83,352 ruptured MN from 61 quality-filtered IF images of slides
270 obtained for 31 patients, **Fig. 3G**, **Methods**). Ruptured micronuclei per primary nucleus (MN rate)
271 ranged from 0.005 to 0.31 (median 0.05) and was 2 times higher in Prevalent WGD patients (p<0.01,
272 **Fig. 3H, Methods**), with MN rate showing modest correlation with rates of cell specific copy number
273 change (**Fig. 3I**). Thus, WGD-related copy number change associates with the formation of
274 micronuclei and raises the possibility that micronuclei are a vehicle for losses and segmental
275 amplifications in HGSOC[27].

276 Abnormalities in micronuclei have been proposed as a mechanism for the formation of complex
277 chromosomal rearrangements, chromothripsis and extrachromosomal DNA, all of which can lead to
278 elevated rates of oncogene amplification. We found that clonal (>90% cells) high-level amplification
279 (HL Amps, **Methods**, **Fig. 3J**) were more frequent in Prevalent WGD patients (p=0.028 Mann-
280 Whitney U test, **Fig. 3K**), including events amplifying *MDM2* (002), *CCNE1* (105), *ERBB2* (044, 051),
281 *CCND1* (065), and *CCND3* (083). Only 1 of the 14 clonal HL Amps was found in a Rare WGD patient,
282 involving the *CCNE1* gene in patient 004, corroborated by bulk sequencing[7] (**Extended Data Fig.
283 4H**). We classified HL Amps as low cancer cell fraction (CCF) if they occurred in less than 10% of
284 the patient cell population. Prevalent WGD was associated with a significantly higher number of low-

285 CCF HL Amps per cell compared to Rare WGD (p=0.022 Mann-Whitney U test, **Fig. 3L**), highlighting
286 that the mutational process that generates HL Amps may itself be increased in Prevalent versus
287 Rare WGD patients. Many of the low prevalence HL Amps are undetectable at a bulk level,
288 highlighting the need for single cell data to identify these events (**Fig. 3M**).

289 Taken together, multiple forms of cell-to-cell genome diversification, including chromosomal
290 missegregations, multipolar mitoses, ruptured micronuclei and HL Amps, all exhibited elevated rates
291 in Prevalent WGD patients, firmly linking WGD to increased cellular genomic diversification in
292 HGSOC.

293 **Evolutionary dynamics of WGD and non-WGD clones**

294 Given the increased rate of chromosomal instability associated with WGD, we next used a
295 phylogenetic approach to investigate the impact of this instability on tumor evolution and the extent
296 to which WGD promotes punctuated vs gradual evolutionary change (**Fig. 4A**, **Methods**). We first
297 focused on events predicted to be on the ancestral branches of each patient. These events were
298 divided into those inferred to occur (i) after WGD in the ancestral branches of prevalent WGD patients
299 (post-WGD) (ii) before WGD in ancestral branches of rare WGD patients (pre-WGD) or (iii) on the
300 ancestral branches of rare WGD patients (non-WGD) (**Fig. 4B**). Rates of losses and gains of arms
301 and chromosomes were significantly higher post-WGD relative to pre-WGD or non-WGD ancestral
302 branches. Thus WGD was associated not only with increased rates of CIN, but also increased
303 propensity for fixation of the changes resulting from CIN. Gains of arms and especially whole
304 chromosomes were rare pre-WGD or on non-WGD ancestral branches, and were significantly more
305 prevalent post-WGD (**Fig. 4B**). This highlights that commonly observed pseudo-triploid karyotypes
306 are unlikely to arise through incremental gains on a diploid background. Instead, triploidy in HGSOC
307 most likely results from WGD and both pre- and post-WGD losses.

308 To investigate punctuated vs gradual evolution post-WGD, we interrogated the post-WGD events for
309 all high-confidence clonal (>95% of cells) and subclonal (<95% of cells) WGD clones (**Fig. 4C**).
310 Comparing post-WGD events between clonal and subclonal WGD clones, we found clonal WGDs to
311 have significantly more post-WGD events of all types (**Fig. 4D**). For instance, clonal WGDs accrued
312 on average 3 times as many whole chromosome losses compared to subclonal WGDs. The number
313 of events post-WGD for some subclonal WGD was surprisingly low. For example, the large WGD
314 subclone (70% of cells) in patient 081 exhibited only two arm loss events post-WGD, while the many
315 cells with more highly divergent genomes post-WGD were unique, possibly indicative of their
316 comparable lack of fitness and inability to expand. In clonal WGDs, the number of chromosome
317 losses and arm gains and losses were significantly correlated with the age of the WGD as estimated
318 using mutations (**Methods**, **Extended Data Fig. 4J**). Patient 014 with a clonal WGD exemplifies

post-WGD evolution (**Fig. 4E**). In this patient, a single cell distinct from the majority of the cells in the patient shared several post-WGD copy number changes with the majority population, but lacked 2 focal HL Amps common to the remaining cells, and retained 4 copies post-WGD of 2q, 6q, 7, 8q, and 18, all of which had 3 copies in the majority of cells. Thus, this outlying cell represented an intermediate stage of evolution post-WGD, likely outcompeted by the other cells with additional focal HL Amps and arm and chromosome losses. Post-WGD divergent evolution with clone specific HL Amps and parallel allele-specific losses was also observed, as exemplified by patient 083 (**Fig. 4F**).

WGD-specific cell intrinsic and tumor microenvironment phenotypes

We next investigated phenotypic associations with WGD states to understand the cancer cell-intrinsic, stromal, and immune activation states found in HGSOC, leveraging previously published patient and site matched scRNA-seq data[14]. We first compared the fraction of cancer cells in the G1, S and G2/M phase of the cell cycle. Prevalent WGD samples exhibited a lower proportion of S-phase cells and a higher proportion of G1-phase cells, consistent with a slower proliferation rate and elongated G1 progression through the cell cycle (**Fig. 5A**, **Methods**)[28]. Pseudotime inference of cell cycle trajectories revealed divergent cell cycle progression in Prevalent vs Rare WGD tumors (**Fig. 5B,C**, **Extended Data Fig. 5A-E, Methods**). In particular, MCM complex genes involved in licensing of DNA replication origins at the G1/S transition (*MCM2*, *MCM6*) were expressed earlier in the cell cycle in Prevalent WGD tumors, together with factors involved in MCM complex loading such as *CDC6* (**Fig. 5D-E**), thus facilitating the replication of larger genomes. Mitotic cyclins (*CCNE1*) and genes involved in DNA repair (*BRCA2*, *MSH2*) had altered temporal order in association with WGD. We also observed correlation between cell cycle distribution and chromosomal missegregation event rates in a WGD-specific manner (**Fig. 5F**), where the fraction of cells in G1 was highly correlated with rates of chromosome losses and arm losses and gains in Rare WGD patients, but not in Prevalent WGD patients (**Fig. 5G, Extended Data Fig. 5G)**. This might be due to the well-documented G1 cell cycle arrest that occurs upon chromosome missegregation and which must be overcome for cells to tolerate CIN[29,30], an evolutionary milestone that is likely achieved by clones that have undergone WGD.

We next proceeded to investigate the association between WGD and cancer cell-intrinsic immune signaling. Cells in Prevalent WGD tumors showed a significant decrease in Type I (IFN-α/IFN-β) and Type II (IFN-γ) interferon, inflammatory pathways, complement and TNFa/NF-κB signaling (**Fig. 6A**). We investigated how chromosomal instability phenotypes encoded in a CIN gene expression signature[31] related to WGD state and found this was significantly higher in Prevalent WGD (**Fig. 6A**), likely due to the elevated missegregation rates as observed in scWGS. Interestingly, STING (*TMEM173*), an innate immune response gene activated by the presence of cytosolic DNA, was expressed at significantly lower levels in Prevalent WGD (**Fig. 6B**), suggesting STING expression

354 may be repressed in Prevalent WGD tumors to evade the immunostimulatory effects of CIN[32–34]. In
355 the context of Rare WGD, STING expression showed strong positive correlation with rates of
356 missegregation, especially chromosome losses (**Fig. 6C**). In addition, expression of E2F target
357 genes showed strong negative correlation with chromosome losses in Rare WGD (**Fig. 6D**). We
358 validated our findings in an hTERT-immortalized retinal pigment epithelial (RPE1) cell line
359 (**Methods**). In diploid RPE1 cells (**Extended Data Fig. 6D-E**), treatment with nocodazole and
360 reversine resulted in increasing levels of chromosome and arm losses and gains (**Extended Data**
361 **Fig. 6F**), in addition to concomitant increases in G1 cell fraction (**Extended Data Fig. G**) and STING
362 expression (**Extended Data Fig. 6H**). Next we compared non-WGD cells in a later passage with a
363 spontaneously arising WGD clone present in the same sample (**Extended Data Fig. 6i-J, Methods**).
364 We found no difference in cell cycle fractions (**Extended Data Fig. 6K**), but STING expression was
365 decreased in the WGD clone (**Extended Data Fig. 6L**). In the Rare WGD context, our results are
366 concordant with the hypothesis that CIN-associated cytosolic DNA activates NF-κB, which promotes
367 transcription of STING[35], and suppresses E2F targets[36,37], ultimately leading to G1 arrest or delay.
368 Critically, we note that this cascade only appears to hold in Rare WGD tumors, suggesting signal re-
369 wiring in Prevalent WGD tumors that enables highly chromosomally unstable tumor cells to adapt to
370 ongoing chromosome missegregation events, thereby evading anti-tumor immune surveillance, as
371 recently proposed[38].

372 We next analyzed the impact of WGD and chromosome missegregation on the tumor immune
373 microenvironment (TME). Consistent with the increased expression of IFN-stimulated genes (ISGs)
374 observed in the cancer cells of Rare WGD patients, we observed enrichment of $CXCL10^{+}CD274^{+}$
375 macrophages (M1.CXCL10), IFN-producing plasmacytoid (pDCs), and activated dendritic cells
376 (cDC1) in the microenvironments of Rare WGD tumors (**Fig. 6E-F**). All major cell types surveyed
377 showed significant enrichment of ISGs in Rare WGD tumors, indicating a pro-inflammatory immune
378 response (**Fig. 6G**). By contrast, the TME of Prevalent WGD tumors showed enrichment for
379 endothelial cells, pericytes, and cancer-associated fibroblasts (CAFs) (**Fig. 6F**), along with
380 suppression of ISG expression. Prevalent WGD tumors also showed slight enrichment of cytotoxic
381 $CD8^{+}$ T cells, possibly due to mutually exclusivity between cytotoxic $CD8^{+}$ T cells and
382 $CXCL10^{+}CD274^{+}$ macrophages across the cohort (**Extended Data Fig. 5C**). Notably, all major cell
383 types in the TME of Prevalent WGD tumors, except for endothelial cells, exhibited marked depletion
384 in cell cycle related gene expression, consistent with a pro-angiogenic yet immunosuppressive
385 microenvironment in WGD tumors (**Fig. 6H**).

386 **DISCUSSION**

387 Using single-cell whole-genome sequencing matched with scRNA and tissue-based quantification
388 of ruptured micronuclei, we illuminate the significant impact of WGD on tumor evolvability and identify
389 associations with cell cycle regulation, inflammatory signaling, angiogenesis and
390 immunosuppressive phenotypes. Through evolutionary timing of WGD, we find variation in clonal
391 WGD expansions from very early to late[7], in addition to both subclonal WGD expansions and multiple
392 independent WGD expansions that would be difficult to identify from bulk sequencing data[18].
393 Interestingly, for two patients with independent WGD clones, the WGD events were approximately
394 synchronous in the tumor's life history. The remaining patients could be classified into three groups:
395 those with small fractions of non-clonal WGD cells, those with a late-emerging non-dominant WGD
396 clone, and those with a dominant WGD clone. Importantly, we did not observe coexisting clones with
397 varying WGD timing. These findings suggest that the evolutionary history of WGD in HGSOC is
398 characterized by the rapid expansion of WGD clones, likely driven by changes in the fitness
399 landscape that favor their proliferation.

400 The established relationship between WGD and genomic diversification is especially evident in our
401 data, wherein we find ubiquitous presence of minor populations that have undergone additional
402 doublings, an increased rate of cell-specific aneuploidies post-WGD, and profoundly divergent cells
403 for which WGD has led to extreme instability. Evolutionary analysis of our data indicates that gradual
404 losses, rather than punctuated evolution, shape the post-WGD evolution of many WGD clones,
405 suggesting historical adaptation and tolerance for the high CIN levels associated with WGD. While
406 WGD was associated with increased $cGAS^+$ ruptured micronuclei, as expected given the higher
407 levels of CIN, prevalent WGD tumors showed decreased cell-intrinsic and cell-extrinsic interferon
408 signaling. In Rare WGD tumors, strong correlations between CIN and tumor phenotypes were
409 consistent with CIN-dependent G1 elongation and increased STING transcription indicating an active
410 cGAS-STING response in Rare WGD patients. This is in contrast with Prevalent WGD, which did not
411 exhibit CIN-associated cell cycle alterations or STING expression increases. Thus STING
412 transcriptional repression may be a prerequisite for clonal expansion of WGD. Furthermore, given
413 the very early timing of WGD in some patients, our results also suggest that deactivation of STING
414 may also be an early event in the evolutionary history of some HGSOC tumors, and may predate
415 WGD. Future investigations into therapeutic targeting of the cGAS-STING pathway should consider
416 WGD-specific abrogation of this pathway, as well as heterogeneity in WGD states.

417 Our results rely in part on the ability to accurately identify ploidy at single cell resolution. Several
418 lines of evidence support the veracity of our results. As shown, many of the cells that make up the
419 small fraction of subclonal WGD cells in each tumor are highly divergent and/or harbor homozygous
420 regions, and would be unlikely to be either a miscalled doublet or poor quality copy number. The less

421 divergent cells still show copy number changes in addition to a perfect doubling, given the
422 requirement of at least one 10MB or larger segment with copy number state 1, 3 or 5. Given the
423 expectation that WGD should be associated with additional post-WGD instability, we were surprised
424 to find that some of the low-prevalence and late-emerging WGD clones did not have large amounts
425 of post-WGD copy number change. This suggests that we may in fact be underestimating the number
426 of small WGD clones, as some of those clones may not be marked by common post-WGD changes.
427 Individual cells that have sustained perfect doublings and non-aberrant G2 phase cells would be
428 detected as half their true ploidy in this study. Future methods may allow isolation of these cell
429 populations, providing further insight into the dynamics of WGD.

430 Critically, we show that potentially targetable therapeutic vulnerabilities such as high-level oncogene
431 amplification preferentially occur on a WGD background, and therefore arise in the context of a low-
432 inflammation and immunosuppressive tumor microenvironment. Those tumors were primarily
433 composed of 1×WGD and 2×WGD cells with increased immunosuppressive properties. We
434 speculate that even if selective targeting of focal oncogene amplification[39] were successful,
435 immunosuppressive states may persist. As therapeutic stratification of patients by genomic
436 properties gains traction in HGSOC, our data introduces a critical covariate given that nearly every
437 tumor harbors WGD cells and multiple co-existent WGD states. Even with the modest cohort size
438 presented here, we anticipate that studying how expanded WGD clones intersect with homologous
439 recombination deficiency and impact responsiveness to anti-angiogenic therapies such as
440 bevacizumab, will advance the rational administration of therapeutic strategies for HGSOC[40,41].
441 Moreover, targeting the WGD process itself may be required to prevent emergence of newly acquired
442 WGD clones. The relevance of our findings to other tumor types remains unclear, although breast
443 PDX[11], *in vitro*[10] and pancreatic cancer mouse[42] studies suggest that WGD dynamics may be
444 pervasive across *TP53* mutant cancers with implications for diverse mechanisms of therapeutic
445 resistance[43]. Thus studying the role of WGD throughout the life history of a tumor should be
446 prioritized as a determinant of therapeutic response.

447 **Data availability**

448 Publicly accessible and controlled access data generated and analyzed in this study are documented
449 in Synapse (accession number syn25569736). Raw sequencing data for scWGS data will be
450 available from the NCBI Sequence Read Archive prior to publication. 10x 3' scRNA-seq is available
451 from the NCBI Gene Expression Omnibus (accession number GSE180661). scWGS copy number
452 heatmaps can be visualized on Synapse
453 (https://www.synapse.org/#!Synapse:syn51769919/datasets/). In addition, MEDICC2 trees and
454 SBMClone results are provided as supplementary file `spectrum-trees.html`. IF images will be
455 available from Synapse prior to publication.

Code availability

The pipeline to process DLP+ scWGS is available at https://github.com/mondrian-scwgs. `SIGNALS`[11] was used for most plotting and scWGS analysis and is available at https://github.com/shahcompbio/signals. `doubleTime` is available at https://github.com/shahcompbio/doubleTime.

Acknowledgements

This project was funded in part by Cycle for Survival supporting Memorial Sloan Kettering Cancer Center and the Halvorsen Center for Computational Oncology. S.P.S. holds the Nicholls Biondi Chair in Computational Oncology and is a Susan G. Komen Scholar. This work was funded in part by awards from the Ovarian Cancer Research Alliance (OCRA) Collaborative Research Development Grant [648007] and NIH R01 CA281928-01 to S.P.S., OCRA Ann Schreiber Mentored Investigator Award to I.V.-G. [650687], OCRA Liz Tilberis Award to D.Z., the Department of Defense Congressionally Directed Medical Research Programs to S.P.S., D.Z. and B.W [W81XWH-20-1-0565], the Seidenberg Family Foundation and the Cancer Research UK Cancer Grand Challenges Program to S.P.S. [C42358/A27460], the Marie-Josée and Henry R. Kravis Center for Molecular Oncology and the National Cancer Institute (NCI) Cancer Center Core Grant [P30-CA008748]. S.F.B. is funded by NIH/NCI grants [P50CA247749, DP5OD026395, R01CA256188, P30-CA008748], the Department of Defense Congressionally Directed Medical Research Program [BC201053], Burroughs Wellcome Fund (BWF), Josie Robertson Foundation, Pershing Square Sohn Alliance for Cancer Research, and Mary Kay Ash Foundation. B.W. is funded in part by Breast Cancer Research Foundation and NIH/NCI P50 CA247749 01 grants. D.Z. is funded by NIH grant R01 CA269382. A.C.W. is supported by NCI Ruth L. Kirschstein National Research Service Award for Predoctoral Fellows F31-CA271673. R.F.S. is a Professor at the Cancer Research Center Cologne Essen (CCCE) funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia. R.F.S. was partially funded by the German Ministry for Education and Research as BIFOLD - Berlin Institute for the Foundations of Learning and Data [01IS18025A and 01IS18037A].

Competing interests

B.W. reports grant funding by Repare Therapeutics paid to the institution, outside the submitted work, and employment of a direct family member at AstraZeneca. C.A. reports grants from Clovis, Genentech, AbbVie and AstraZeneca and personal fees from Tesaro, Eisai/Merck, Mersana Therapeutics, Roche/Genentech, Abbvie, AstraZeneca/Merck and Repare Therapeutics, outside the scope of the submitted work. C.A. reports clinical trial funding to the institution from Abbvie, AstraZeneca, and Genentech/Roche; participation on a data safety monitoring board or advisory board: AstraZeneca, Merck; unpaid membership of the GOG Foundation Board of Directors and the

NRG Oncology Board of Directors. C.F. reports research funding to the institution from Merck, AstraZeneca, Genentech/Roche, Bristol Myer Squibb, and Daiichi; uncompensated membership of a scientific advisory board for Merck and Genentech; and is a consultant for OncLive, Aptitude Health, Bristol Myers Squibb and Seagen, all outside the scope of this manuscript. D.S.C. reports membership of the medical advisory board of Verthermia Acquio Inc and Biom'up, is a paid speaker for AstraZeneca, and holds stock of Doximity, Moderna, and BioNTech. D.Z. reports institutional grants from Merck, Genentech, AstraZeneca, Plexxikon, and Synthekine, and personal fees from AstraZeneca, Xencor, Memgen, Takeda, Astellas, Immunos, Tessa Therapeutics, Miltenyi, and Calidi Biotherapeutics. D.Z. own a patent on use of oncolytic Newcastle Disease Virus for cancer therapy. N.A.-R. reports grants to the institution from Stryker/Novadaq and GRAIL, outside the submitted work. R.N.G. reports funding from GSK, Novartis, Mateon Therapeutics, Corcept, Regeneron, Clovis, Context Therapeutics, EMD Serono, MCM Education, OncLive, Aptitude Health and Prime Oncology, outside this work. S.F.B. owns equity in, receives compensation from, and serves as a consultant and the Scientific Advisory Board and Board of Directors of Volastra Therapeutics Inc. He also serves on the SAB of Meliora Therapeutics Inc. S.P.S. reports research funding from AstraZeneca and Bristol Myers Squibb, outside the scope of this work; S.P.S. is a consultant and shareholder of Canexia Health Inc. Y.L.L. reports research funding from AstraZeneca, GSK/Tesaro, Artios Pharma, and Tesaro Therapeutics outside this work. Y.L. reports serving as a consultant for Calyx Clinical Trial Solutions outside this work.

References

1. Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
2. Dewhurst, S. M. *et al.* Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov.* **4**, 175–185 (2014).
3. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).
4. Selmecki, A. M. *et al.* Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
5. Lambuta, R. A. *et al.* Whole-genome doubling drives oncogenic loss of chromatin segregation. *Nature* **615**, 925–933 (2023).
6. Gemble, S. *et al.* Author Correction: Genetic instability from a single S phase after whole-genome duplication. *Nature* **608**, E27 (2022).
7. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
8. Cheng, Z. *et al.* The Genomic Landscape of Early-Stage Ovarian High-Grade Serous Carcinoma. *Clin. Cancer Res.* **28**, 2911–2922 (2022).
9. Shi, Q. & King, R. W. Chromosome nondisjunction yields tetraploid rather than aneuploid cells in human cell lines. *Nature* **437**, 1038–1042 (2005).
10. Salehi, S. *et al.* Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* **595**, 585–590 (2021).
11. Funnell, T. *et al.* Single-cell genomic variation induced by mutational processes in cancer. *Nature* **612**, 106–115 (2022).
12. Zhang, R. *et al.* Discovery of potent, orally active KIF18A inhibitors targeting CIN-high cancer cells. *J. Clin. Orthod.* **40**, e15046–e15046 (2022).
13. Belmontes, B. *et al.* Abstract 516: Discovery and preclinical characterization of AMG 650, a first-in-class inhibitor of kinesin KIF18A motor protein with potent activity against chromosomally unstable cancers. *Cancer Res.* **83**, 516–516 (2023).
14. Vázquez-García, I. *et al.* Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* **612**, 778–786 (2022).
15. Funnell, T. *et al.* Integrated structural variation and point mutation signatures in cancer genomes using correlated topic models. *PLOS Computational Biology* vol. 15 e1006799 Preprint at https://doi.org/10.1371/journal.pcbi.1006799 (2019).
16. Laks, E. *et al.* Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing. *Cell* **179**, 1207–1221.e22 (2019).
17. Dentro, S. C. *et al.* Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184**, 2239–2254.e39 (2021).
18. Frankell, A. M. *et al.* The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023).
19. Kim, M. *et al.* Single-cell mtDNA dynamics in tumors is driven by coregulation of nuclear and mitochondrial genomes. *Nat. Genet.* **56**, 889–899 (2024).
20. Myers, M. A., Zaccaria, S. & Raphael, B. J. Identifying tumor clones in sparse single-cell mutation data. *Bioinformatics* **36**, i186–i193 (2020).
21. Bollen, Y. *et al.* Reconstructing single-cell karyotype alterations in colorectal cancer identifies punctuated and gradual diversification patterns. *Nat. Genet.* **53**, 1187–1195 (2021).
22. Sauer, C. M. *et al.* Molecular landscape and functional characterization of centrosome amplification in ovarian cancer. *Nat. Commun.* **14**, 6505 (2023).
23. Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
24. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).
25. Umbreit, N. T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, (2020).
26. Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).

27. Al-Rawi, D. H. & Bakhoum, S. F. Chromosomal instability as a source of genomic plasticity. *Curr. Opin. Genet. Dev.* **74**, 101913 (2022).
28. Liang, S., Wang, F., Han, J. & Chen, K. Latent periodic process inference from single-cell RNA-seq data. *Nat. Commun.* **11**, 1441 (2020).
29. Santaguida, S. *et al.* Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev. Cell* **41**, 638–651.e5 (2017).
30. Thompson, S. L. & Compton, D. A. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *J. Cell Biol.* **188**, 369–381 (2010).
31. Bakhoum, S. F. *et al.* Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467–472 (2018).
32. Konno, H. *et al.* Suppression of STING signaling through epigenetic silencing and missense mutation impedes DNA damage mediated cytokine production. *Oncogene* **37**, 2037–2051 (2018).
33. Amiji, M. M. & Milane, L. S. *Cancer Immunology and Immunotherapy: Volume 1 of Delivery Strategies and Engineering Technologies in Cancer Immunotherapy*. (Academic Press, 2021).
34. de Queiroz, N. M. G. P., Xia, T., Konno, H. & Barber, G. N. Ovarian Cancer Cells Commonly Exhibit Defective STING Signaling Which Affects Sensitivity to Viral Oncolysis. *Mol. Cancer Res.* **17**, 974–986 (2019).
35. Chen, L.-Y., Pang, X.-Y., Chen, C. & Xu, H.-G. NF-κB regulates the expression of STING via alternative promoter usage. *Life Sci.* **314**, 121336 (2023).
36. Penzo, M. *et al.* Sustained NF-kappaB activation produces a short-term cell proliferation block in conjunction with repressing effectors of cell cycle progression controlled by E2F or FoxM1. *J. Cell. Physiol.* **218**, 215–227 (2009).
37. Araki, K., Kawauchi, K. & Tanaka, N. IKK/NF-kappaB signaling pathway inhibits cell-cycle progression by a novel Rb-independent suppression system for E2F transcription factors. *Oncogene* **27**, 5696–5705 (2008).
38. Li, J. *et al.* Non-cell-autonomous cancer progression from chromosomal instability. *Nature* **620**, 1080–1088 (2023).
39. Au-Yeung, G., Mileshkin, L. & Bowtell, D. D. L. CCNE1 Amplification as a Therapeutic Target. *J. Clin. Oncol.* **41**, 1770–1773 (2023).
40. Tewari, K. S. *et al.* Final Overall Survival of a Randomized Trial of Bevacizumab for Primary Treatment of Ovarian Cancer. *J. Clin. Oncol.* **37**, 2317–2328 (2019).
41. Lorusso, D. *et al.* Updated progression-free survival and final overall survival with maintenance olaparib plus bevacizumab according to clinical risk in patients with newly diagnosed advanced ovarian cancer in the phase III PAOLA-1/ENGOT-ov25 trial. *Int. J. Gynecol. Cancer* **34**, (2024).
42. Baslan, T. *et al.* Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).
43. Hobor, S. *et al.* Mixed responses to targeted therapy driven by chromosomal instability through p53 dysfunction and genome doubling. *Nat. Commun.* **15**, 4871 (2024).
44. Zehir, A., Benayed, R., Shah, R. H., Syed, A. & Middha, S. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* (2017).
45. Bykov, Y., Kim, S. H. & Zamarin, D. Preparation of single cells from tumors for single-cell RNA sequencing. *Methods Enzymol.* **632**, 295–308 (2020).
46. Medina-Martínez, J. S. *et al.* Isabl Platform, a digital biobank for processing multimodal patient data. *BMC Bioinformatics* **21**, 549 (2020).
47. Lai, D., Ha, G. & Shah, S. HMMcopy: copy number prediction with correction for GC and mappability bias for HTS data. *R package version*.
48. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
49. Curtius, K., Wright, N. A. & Graham, T. A. An evolutionary perspective on field cancerization. *Nat. Rev. Cancer* **18**, 19–32 (2018).
50. McPherson, A. W. *et al.* ReMixT: clone-specific genomic structure estimation in cancer. *Genome Biol.* **18**, 140 (2017).

51. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
52. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019) doi:10.1101/861054.
53. McPherson, A., Shah, S. & Cenk Sahinalp, S. deStruct: Accurate Rearrangement Detection using Breakpoint Specific Realignment. *bioRxiv* 117523 (2017) doi:10.1101/117523.
54. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
55. Wang, Y. K. *et al.* Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **49**, 856–865 (2017).
56. Ranganath, R., Gerrish, S. & Blei, D. Black Box Variational Inference. in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (eds. Kaski, S. & Corander, J.) vol. 33 814–822 (PMLR, Reykjavik, Iceland, 22--25 Apr 2014).
57. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
58. Kaufmann, T. L. *et al.* MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. *Genome Biol.* **23**, 241 (2022).
59. Ross, E. M., Haase, K., Van Loo, P. & Markowetz, F. Allele-specific multi-sample copy number segmentation in ASCAT. *Bioinformatics* **37**, 1909–1911 (2021).
60. Fitch, W. M. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Syst. Zool.* **20**, 406–416 (1971).
61. Sankoff, D. Minimal Mutation Trees of Sequences. *SIAM J. Appl. Math.* **28**, 35–42 (1975).
62. Zhang, A. W. *et al.* Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat. Methods* **16**, 1007–1015 (2019).
63. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
64. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
65. Zatzman, M. *et al.* Widespread hypertranscription in aggressive human cancers. *Sci Adv* **8**, eabn0238 (2022).
66. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977–2000 (2002).
67. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
68. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* **1**, 417–425 (2015).
69. Hoffman, G. E. *et al.* Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. *bioRxiv* (2023) doi:10.1101/2023.03.17.533005.
70. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
71. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
72. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
73. Gao, T. *et al.* Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2023).

667 **METHODS**

668 **Experimental methods**

669 **Sample collection**

670 All enrolled patients were consented to an institutional biospecimen banking protocol and MSK-
671 IMPACT testing[44], and all analyses were performed per a biospecimen research protocol. All
672 protocols were approved by the Institutional Review Board (IRB) of Memorial Sloan Kettering Cancer
673 Center. Patients were consented following the IRB-approved standard operating procedures for
674 informed consent. Written informed consent was obtained from all patients before conducting any
675 study-related procedures. The study was conducted in accordance with the Declaration of Helsinki
676 and the Good Clinical Practice guidelines (GCP).

677 We collected fresh tumor tissues from 40 HGSOC patients at the time of upfront diagnostic
678 laparoscopic or debulking surgery. Ascites and tumor tissue from multiple metastatic sites, including
679 bilateral adnexa, omentum, pelvic peritoneum, bilateral upper quadrants, and bowel were procured
680 in a predetermined, systemic fashion (median of 4 primary and metastatic tissues per patient) and
681 were placed in cold RPMI for immediate processing. Blood samples were collected pre-surgery for
682 the isolation of peripheral blood mononucleated cells (PBMCs) for normal whole-genome
683 sequencing (WGS). The isolated cells were frozen and stored at -80°C. In addition, tissue was snap
684 frozen for bulk DNA extraction and tumor WGS. Tissue was also subjected to formalin fixation and
685 paraffin-embedding (FFPE) for histologic, immunohistochemical and multiplex immunophenotypic
686 characterization.

687 **Sample processing**

688 We profiled patient samples using five different experimental assays:

689 1. Viably frozen single-cell suspensions were derived from fresh tissue samples and processed
690 for single-cell whole-genome sequencing (scWGS) of 65 sites from 40 patients (~815 cells
691 per site, **Supp. Tab. 2**). $CD45^-$ cells were flow-sorted in samples with low tumor purity.
692 2. $CD45^+$ and $CD45^-$ flow-sorted cells were previously reported fresh tissue samples and
693 processed for single-cell RNA sequencing (scRNA-seq) of 123 sites from 32 patients (~6k
694 cells per site, **Supp. Tab. 2**).
695 3. For each specimen with scRNA-seq, site-matched FFPE tissue sections adjacent to the H&E
696 section were stained by multiplexed immunofluorescence (IF) for micronuclei and DNA
697 sensing mechanisms (83 tissue samples from 37 patients).

698 4. FDA-approved clinical sequencing of 468 cancer genes (MSK-IMPACT) was obtained on
699 DNA extracted from FFPE tumor and matched normal blood specimens for each patient
700 (**Extended Data Fig. 1B**).
701 5. Snap-frozen tissues were processed to obtain matched tumor-normal bulk whole-genome
702 sequencing (WGS) on a single representative site of 33 out of 40 patients with scWGS,
703 scRNA-seq and IF, to derive mutational processes from genome-wide single nucleotide and
704 structural variants.

705 **Single-cell DNA sequencing**

706 *Tissue dissociation*

707 Tumor tissue was immediately processed for tissue dissociation. Fresh tissue was cut into 1 mm
708 pieces and dissociated at 37°C using the Human Tumor Dissociation Kit (Miltenyi Biotec) on a
709 gentleMACS Octo Dissociator. After dissociation, single-cell suspensions were filtered and washed
710 with Ammonium-Chloride-Potassium (ACK) Lysing Buffer. Cells were stained with Trypan Blue and
711 cell counts and viability were assessed using the Countess II Automated Cell Counter
712 (ThermoFisher). For detailed protocol see Bykov et al., 2020[45]. Freshly dissociated cells were
713 processed for scRNA-seq as described in Vázquez-García et al., 2022[14]. Viably frozen dissociated
714 cells were stored for scWGS.

715 *Cell sorting*

716 Viably frozen dissociated cells used for scWGS were thawed and then stained with a mixture of
717 GhostRed780 live/dead marker (TonBo Biosciences) and Human TruStain FcX™ Fc Receptor
718 Blocking Solution (BioLegend). For samples with low tumor purity, the stained samples were then
719 optionally incubated and stained with Alexa Fluor® 700 anti-human CD45 Antibody (BioLegend).
720 Post staining, they were washed and resuspended in RPMI + 2% FCS and submitted for cell sorting.
721 The cells were sorted into CD45 positive and negative fractions by fluorescence assisted cell sorting
722 (FACS) on a BD FACSAria™ III flow cytometer (BD Biosciences). Positive and negative controls
723 were prepared and used to set up compensations on the flow cytometer. Cells were sorted into tubes
724 containing RPMI + 2% FCS for sequencing.

725 *Library preparation and sequencing*

726 Single-cell whole-genome library preparation was carried out as described in Laks et al., 2019[16].
727 Briefly, single cells were dispensed into nanowells with protease (Qiagen) and DirectPCR Cell lysis
728 reagent (Viagen). After overnight incubation cells are subjected to heat lysis and protease
729 inactivation followed by tagmentation in a tagmentation mix (14.335 nL TD Buffer, 3.5 nL TDE1, and

730 0.165 nL 10% Tween-20) at 55°C for 10 minutes. Once the tagmentation reaction was neutralized,
731 8 cycles of PCR followed. The indexed single-cell libraries were recovered from the nanowells by
732 centrifugation into a pool and sequenced on Illumina NovaSeq 6000.

733 **Immunofluorescence**

734 *Overview*

735 We profiled matched FFPE tissues with cGAS and DAPI immunofluorescence to quantify the rate of
736 micronuclei formation in tumors. The immunofluorescence detection of cGas was performed at the
737 Molecular Cytology Core Facility of Memorial Sloan Kettering Cancer Center using Discovery XT
738 processor (Ventana Medical Systems.Roche-AZ). Antigen retrieval was performed using ULTRA
739 Cell Conditioning (Ventana Medical Systems, 950-224). The tissue sections were blocked first for 30
740 minutes in Background Blocking reagent (Innovex, catalog#: NB306).

741 *Tissue staining*

742 For the cGAS staining, a mouse monoclonal cGAS antibody (LSBio, LS-C757990) was used in 1:200
743 dilution. The incubation with the primary antibody is done for 5 hours followed by biotinylated mouse
744 secondary (Vector Labs, MOM Kit BMK-2202) in 5.75µg/mL. Blocker D, Streptavidin- HRP and TSA
745 Alexa594 (Life Tech, cat#B40957) was applied for 16 minutes.

746 All slides were counterstained in 5µg/mL DAPI (Sigma D9542) for 5 minutes at room temperature,
747 mounted with anti-fade mounting medium Mowiol.

748 **RPE1 cell line experiments**

749 We explored the phenotypic effects of chromosomal instability and WGD in *TP53*-knockout RPE1
750 cells. *TP53*-knockout RPE-1 was a gift from the Maciejowski laboratory at the Memorial Sloan
751 Kettering Cancer Center (MSKCC). RPE-1 cells were cultured in DMEM (Corning) supplemented
752 with 10% fetal bovine serum (Sigma-Aldrich), 1% penicillin-streptomycin (Thermo Fisher) at 37°C
753 and 5% CO2. All cells were periodically tested for mycoplasma contamination.

754 *TP53*-/- RPE1 cells were treated with nocodazole, reversine and DMSO control to induce varying
755 levels of chromosomal instability, then subject to both 10X multiome sequencing and DLP+ scWGS.
756 For nocodazole treatment, RPE-1 cells were seeded at 20% confluence at the time of nocadazole
757 addition. Cells were treated with 100 ng/ml nocodazole (Sigma-Aldrich) or DMSO for 8hrs. After 8hrs,
758 cells were treated with three washes with phosphate buffered saline to remove the drug. After 48hrs
759 the cells were collected. For reversine (Cayman Chemical Company) treatment, cells were treated
760 at a concentration of 0.5 µM reversine for 48hrs. After 48hrs, cells were washed with three washes

761 with phosphate buffered saline to remove the drug. Cells were collected after 12hrs. 10,000 cells per
762 condition were collected for 10X Chromium Single Cell Multiome ATAC+Gene Expression according
763 to the manufacturer's protocol. Library preparation and sequencing were performed in MSKCC
764 Integrated Genomics Core. 1M cells per condition were subject to DLP+ as described above.

765 A spontaneously arising WGD subclone was observed as a minor population of the *TP53*-knockout
766 RPE1 cells (**Extended Data Fig. 6E**). After 30 additional passages (sample RPE-WGD), the WGD
767 subclone, as measured by DLP+, comprised 37% of the population, presenting the opportunity to
768 explore phenotypic differences between WGD and non-WGD cells. Sample RPE-WGD was subject
769 to DLP+ scWGS and 10X scRNA.

Sample	Treatment	Sequencing	Description
RPE-D	DMSO	10X Multiome/DLP+	non-WGD, low CIN control
RPE-Noco	Nocodazole	10X Multiome/DLP+	non-WGD, medium CIN
RPE-Rev	Reversine	10X Multiome/DLP+	non-WGD, high CIN
RPE-WGD	None	10X scRNA/DLP+	WGD/non-WGD mixed population

770 **Computational methods**

771 Computational analyses of multi-modal datasets were enabled by the Isabl platform[46].

772 **Single-cell DNA sequencing**

773 *Overview*

774 The single-cell DNA analysis pipeline is a suite of workflows for analyzing the single-cell data
775 generated by the DLP+ platform[16]. The workflow takes dual-indexed reads from Illumina paired-end
776 sequencing data as the input and performs various alignment and postprocessing tasks. The pipeline
777 is publicly available on GitHub (https://github.com/mondrian-scwgs/mondrian), which we run within
778 the Isabl framework[46].

779 *Alignment*

780 We use Trim Galore to remove adapters and FastQC to generate QC reports before running
781 alignment. The reads are then aligned with bwa-mem (with support for bwa-aln). The pipeline can
782 also perform local indel realignment with GATK's IndelAligner if required. PCR duplicates are marked
783 using Picard MarkDuplicates and alignment metrics are computed for each cell with Picard tools
784 CollectWgsMetrics and CollectInsertSizeMetrics. The pipeline also generates plots for each
785 alignment metric for a quick overview.

786 *Copy number segmentation*

787 Reads are tabulated for non-overlapping 500 kb regions. A modal regression normalization[16] is
788 performed to reduce GC bias. The pipeline then runs HMMcopy with 6 different ploidy settings and
789 the best fit is chosen automatically[47]. The pipeline also generates heatmaps with cell clustering, per-
790 cell copy-number profile and the modal regression fit for visualization.

791 *Quality control*

792 scWGS was first subjected to quality control and filtering to remove non-cancer cells, S-phase
793 replicating cells, low quality cells, and doublets, resulting in 29,481 high-quality cancer cell genomes
794 (**Extended Data Fig. 2C-D**). The quality control pipeline compiles the results from the total copy
795 number analysis and alignment, and we then use a random forest classifier to predict the quality of
796 each cell based on the alignment and HMMcopy metrics[16]. We then inferred allele-specific copy
797 number for each of these cells using `SIGNALS`[11]. Patient level average ploidy ranged from 1.6 to 4.5,
798 and LOH ranged from 0.2 to 0.8. Ploidy and LOH estimates were concordant with matching bulk
799 WGS and clinical panel sequencing by MSK-IMPACT, and losses and gains from scWGS coincided
800 with known drivers of HGSOC (**Extended Data Fig. 2E-G**). Thus at a pseudobulk level, the genomic
801 characteristics of our scWGS cohort matched those of both whole-genome and targeted bulk data.

802 *Haplotype-specific copy number*

803 In a matched normal sample we measure reference and alternate allele counts for SNPs from the
804 1000 Genomes Phase 2 reference panel. We use a binomial exact test to filter for SNPs
805 heterozygous in the normal sample. Using SHAPEIT[48] and the 1000 Genomes Phase 2 reference
806 panel, we compute haplotype blocks. Next we measure per-cell reference and alternate allele counts
807 for heterozygous SNPs in the tumor scWGS data.

808 *Cell filtering*

809 We established stringent filters to maximize the removal of problematic cells without sacrificing
810 sensitivity to rare interesting populations including those representing cell specific WGD.

811 *Removal of low-quality cells*

812 We removed cells with quality score <0.75. The quality score was computed using the classifier
813 presented in Laks et al., 2019[16].

814 *Removal of suspect high-ploidy cells*

815 We restricted analysis to cells with high confidence ploidy calls. Absolute ploidy is unidentifiable from
816 the sequencing data of an individual cell, thus we take a parsimony approach and assume the true
817 ploidy to be the lowest ploidy value that provides a reasonable fit to the data. One failure mode in
818 the automatic determination of ploidy by HMMCopy occurs when HMMCopy converges on a solution
819 with double the true ploidy driven by the overfitting of isolated outlier bins. Such cells are
820 characterized by mostly even copy number states except for isolated bins with odd copy number. To
821 remove such potential artifacts we required there to be at least one segment >10MB in length with
822 copy number 1, 3 or 5. Cells with no segment >10MB in length with copy number 1, 3, or 5 were
823 removed from further analysis.

824 *Removal of doublets*

825 We applied several orthogonal approaches to remove doublets from the DLP data. First, under the
826 assumption that chromosome 17 LOH should be clonal in ovarian cancer, we removed tumor cells
827 that lacked LOH of chromosome 17. Then, we used a combination of mutation-based features to
828 manually identify tumor-normal doublets, including LOH (much lower than typical tumor cells),
829 proportion of SNVs with alternate reads (higher than typical normal cells), and copy-number profiles
830 that were similar to tumor cells with the addition of 2 copies across the genome. Finally, 2 raters
831 separately reviewed the brightfield image of each cell in the clear microfluidic nozzle prior to
832 deposition in the microwell array for sequencing, and flagged any images that appeared to contain
833 more than 1 cell. Any cell whose image was flagged by at least 1 reviewer was removed from
834 analysis. Additional details on these approaches are described in the Supplementary Note.

835 *Removal of S-phase cells*

836 It is necessary to remove S-phase cells before downstream analysis as the observed HMMcopy
837 profiles of these cells reflect a mixture of both somatic (heritable) copy number and transient doubling
838 of replicated genomic loci. We nominated S-phase cells through a combination of features known to
839 correlate with S-phase cells. As we aimed to isolate the high-quality G1/2-phase cells for downstream
840 analysis, we did not need to distinguish between S-phase cells and low quality cells (i.e. noisy
841 HMMcopy profiles due to other factors such as under-tagmentation prior to sequencing or incomplete
842 cell lysis).

843 We first computed the following three features for each cell:

844 **1)** The Spearman correlation between HMMcopy state profile for a cell-of-interest and the RepliSeq
845 replication timing profile from MCF-7 cells. S-phase cells will have higher correlations than G1/2-
846 phase cells.

847 **2)** The number of HMMcopy breakpoints per cell (number of adjacent loci with different integer copy
848 number state). S-phase cells have more breakpoints than G1/2-phase cells.

849 **3)** The median breakpoint prevalence across all HMMcopy breakpoints. This statistic is calculated
850 by first computing the mean prevalence of each breakpoint across all cells belonging to said patient.
851 Then, for each cell-of-interest, we subset to only the genomic loci with detected breakpoints in that
852 cell and calculate the median of the mean breakpoint prevalences for said loci. S-phase cells have
853 low median breakpoint frequency scores as they have lots of rare breakpoints.

854 All three features varied widely across patients due to each patient's unique number, positioning,
855 and heterogeneity of somatic copy number alteration. Thus we employed a strategy of examining
856 each feature's distribution across all cells in a patient, manually inspecting outlier cells, and selecting
857 custom thresholds for each patient. We employ a filtering approach whereby cells are called as S-
858 phase if any two of the three features are beyond the threshold. This conservative strategy ensures
859 that all remaining cells are truly in G1/2-phase and therefore have HMMcopy profiles that accurately
860 reflect somatic copy number.

861 *Removal of normal cells*

862 After copy number calling, we identified normal cells as those cells with copy number state average
863 between 1.95 and 2.05 and standard deviation less than 0.5. We removed these normal cells from
864 further analysis. We also manually inspected cells with aneuploidy slightly outside this range but
865 much less than tumor cells in the same sample, and removed these "aberrant normal" cells (see
866 Supplementary Note for examples). These cells typically did not share SNVs with the tumor cells
867 and may correspond to other epithelial cells affected by field cancerization[49] or immune/stromal cells
868 with rare chromosomal aberrations.

869 *Comparison with bulk copy number*

870 We use WGS copy number inferred by *ReMixT*[50] to validate the average ploidy in the MSK
871 SPECTRUM cohort. Similarly, we use IMPACT copy number inferred by FACETS[51] for additional
872 orthogonal validation.

873 *Detecting WGD in single cells using allele-specific copy number*

874 WGD events were identified in single cells based on the allele-specific copy number state previously

875 described for bulk WGS[18]. We computed two metrics from SIGNALS results: fraction of the genome
876 with ≥2 copies for the major allele (*FM2*), and fraction of the genome ≥3 copies for the major allele
877 (*FM3*). Similar to results in bulk WGS, a clear separation can be seen between subpopulations using
878 each metric (**Extended Data Fig. 3H,I**). We classified any cell with FM2 > 0.5 as having undergone
879 at least 1 WGD, and any cell with FM3 > 0.5 as having undergone at least 2 WGD.

880 *Patient level WGD classifications*

881 Patients were classified as Prevalent WGD if the fraction of cells classified as 1WGD≥1 exceeded
882 50% of the cells sequenced for that patient. The remaining patients were classified as Rare WGD.

883 *Subclonal WGD classification*

884 We classified cells within each patient as comprising a subclonal WGD subpopulation if they were
885 predicted to have 1 more WGD than the number of WGDs for the majority population. However, for
886 patients 081 and 125, a significant fraction of cells were predicted to be 0×WGD (>25%), with the
887 remaining cells 1×WGD. For these patients, we considered the 1×WGD to be the subclonal WGD
888 population.

889 *Variant calling*

890 *SNV calling*

891 Since the low per-cell coverage in scWGS is insufficient to resolve variants at nucleotide resolution,
892 we merge all the single cells together to create a pseudo-bulk genome for each library. We run the
893 Mutect2 variant caller[52] on the merged data across all libraries from each patient. We compute the
894 reference and alternate counts for each cell at variant loci that are detected by either caller over all
895 libraries.

896 *SV calling*

897 We employed a similar approach for breakpoint calling by creating pseudo-bulk libraries, then
898 running deStruct[53] and Lumpy[54] on each library. Only consensus SVs detected by both methods
899 were retained, where an SV from both methods were considered consensus if their coordinates were
900 within 200 bp and their orientations matched. The SV calls were further post-processed as described
901 in a previous study[55].

902 *Filtering somatic variant calls using evolutionary constraints*

903 Standard variant callers can produce artifactual calls on scWGS data, since its low insert sizes can
904 result in incorrect alignments that appear to represent somatic variants. To address this, we
905 developed a label propagation classifier to identify artifacts based on read-level features. To train
906 this classifier, we leveraged the principle that distinct copy-number clones should not share subclonal
907 variants to annotate high-confidence true and high-confidence artifact variants in a subset of
908 samples. We then applied this classifier and trained it on high-confidence correct and artifactual calls
909 based on manually labeled clones from a subset of patients, then applied it to all variants from all
910 patients.

911 *SBMClone*

912 We applied SBMClone[20] to the filtered somatic variants for each patient. SBMClone was run 10 times
913 on each patient with different random initializations, and the solution with the highest likelihood was
914 kept.

915 *Evolutionary histories of SNV clones using* `doubleTime`

916 We developed `doubleTime`, a method for computing evolutionary histories of the SNV clones in
917 each patient, including accurate placement of WGD events in the clonal phylogeny of each patient.
918 Our approach involved three major steps. First, we constructed a clonal phylogeny relating the clones
919 identified by SBMClone. Second, we assigned WGD events to branches in the clonal phylogeny. For
920 each pair of WGD clones, we assessed whether those clones arised from a single common WGD or
921 two independent WGD. Given this information we were able to unambiguously assign WGD events
922 to branches throughout each patient's clonal phylogeny. Third, we used a probabilistic model to
923 assign SNVs to branches of the clonal phylogeny, including assignment before and after WGD
924 events on WGD branches. We describe each of the three steps in additional detail below.

925 *Perfect phylogenies of SNV clones*

926 We reconstructed phylogenetic trees with SBMClone clones as leaves using a binarized version of
927 the implicit block structure inferred by SBMClone. We first computed a density matrix D where each
928 row corresponds to a clone (i.e., cell block), each column corresponds to an SNV cluster (i.e., SNV
929 block) and each entry $D_{i,j}$ contains the number of pairs *(a,b)* in which cell *a* in clone *i* has at least
930 one alternate read covering SNV *b* in cluster *j*, divided by the total number of possible pairs (i.e., the
931 size of clone *i* times the size of cluster *j*). We then computed a binary matrix B by rounding up those
932 entries of D that exceeded a density of 0.01, removed empty columns, and attempted to infer a
933 phylogenetic tree by applying the perfect phylogeny algorithm. Matrices B that did not permit a

934 perfect phylogeny were manually modified with the minimum number of changes required to permit
935 a perfect phylogeny – this typically occurred when mutations shared between two or more clones
936 had been lost due to a deletion in a subset of the clones.

937 *Discerning independent from shared WGD*

938 To identify cases in which sequenced WGD cells arose from distinct WGD events, we analyzed
939 SNVs from the single-cell DNA sequencing data. Specifically, for each patient, we focused
940 exclusively on those regions that exhibited copy-neutral loss of heterozygosity (cnLOH; i.e., major
941 copy number 2 and minor copy number 0) among nearly all (≥ 90%) tumor cells with a single WGD.
942 Given a candidate bipartition of the 1 WGD cells, under the infinite sites assumption, each cnLOH
943 SNV then fits into one of the following categories:

944 - 2 mutant copies in both clones (shared pre-WGD and pre-divergence)
945 - 1 mutant copy in one clone (private post-divergence)
946 - 0 mutant copies (false positive variant)
947 - 1 mutant copy in both clones (shared post-WGD and pre-divergence)
948 - 2 mutant copies in both clones (post-WGD and post-divergence)

949 The last two categories of SNVs present evidence for or against multiple independent WGD events.
950 SNVs that are shared at 1 variant copy (VAF ~0.5) would suggest that the two sets of cells underwent
951 the same ancestral WGD event, as they share mutations that must have followed the WGD.
952 Conversely, SNVs that are private at two variant copies (VAF ~1) would suggest that the two sets of
953 cells underwent distinct WGD events, as they have private mutations that preceded the WGD.
954 Specifically, we considered the following hypotheses:

955 1. Single-WGD: Shared 1-copy SNVs are allowed, but private 2-copy SNVs are not allowed.
956 2. Multiple-WGD: Shared 1-copy SNVs are not allowed, but private 2-copy SNVs are allowed.

957 To evaluate the relative strength of these hypotheses, we developed a likelihood ratio test that
958 compared the probability of observing the given variant counts for cnLOH SNVs under these two
959 hypotheses: for each patient, we evaluated P(Multiple-WGD)/P(Single-WGD) using a simple
960 binomial model of read counts. We then tested the significance of this likelihood ratio by generating
961 an empirical distribution: we fixed the SNV read counts and their best-fitting variant copy numbers
962 under the Single-WGD hypothesis and resampled alternate counts.

963 *Assigning SNVs to branches and estimating branch lengths*

964 From the previous steps, we are given a tree relating the clones detected by SBMclone. We place
965 WGD events on branches such that all Prevalent WGD patients had a WGD event placed on the

root of the tree, except those in which independent WGD events had been identified (patients 025 and 045) or WGD only affected a subset of clones (patient 081), in which case those specific events were placed further down the tree. We used a probabilistic model to assign SNVs to branches and estimate branch lengths based on read count evidence for SNVs in each clone. For WGD branches, the model assigns SNVs as occurring before or after the WGD, and estimates the length of the branch before and after the WGD. This strategy effectively splits each branch with a WGD event into two unique positions in the tree, meaning that the total number of positions in the tree to which an SNV can be assigned is equal to the number of branches determined by SBMclone + the number of branches with WGD events.

For this analysis, we considered only those SNVs in regions where for each SBMClone clone, over 80% of cells shared the same copy-number state. We further restricted analysis to SNVs in regions with allele-specific copy-number states whose multiplicity (i.e., variant copy number, or the number of copies of the genome containing the SNV) and thus expected VAF could be uniquely determined by the combination of tree placement and WGD status (i.e., whether or not the clone was affected by an ancestral WGD event). Specifically, we analyzed regions with the following copy-number states across all clones:

- 1:0 in both WGD and non-WGD clones
- 1:1 in both WGD and non-WGD clones
- 2:0 in WGD clones, 1:0 in non-WGD clones
- 2:1 in WGD clones, 1:1 in non-WGD clones
- 2:2 in WGD clones, 1:1 in non-WGD clones

In each of these scenarios, we assume that the WGD and copy-number events immediately following the WGD account for the differences in copy number between WGD and non-WGD clones. Note that the only patient in the cohort with different WGD status for different leaves was patient 081, so for nearly all patients we analyzed only those SNVs with clonal copy-number states (matching the above listed states depending on WGD status). The multiplicity for an SNV on a particular allele placed on a particular branch of the tree was as follows:

- 0, if the corresponding allele had 0 copies
- Equal to the allele-specific copy number of the allele in the clone, if the SNV occurred pre-WGD and the leaf was affected by WGD
- Equal to 1 otherwise

Each SNV is assigned to a tree position by fitting the observed total and alternative counts of said SNV to the expected VAFs for all clones. SNVs are assigned to positions in the tree using a Dirichlet-Categorical distribution, and a Beta-Binomial emission model is used to relate observed SNV counts

1000 to expected VAFs. The model is implemented in Pyro and fit using black box variational inference[56].
1001 Note that when computing branch lengths, we only use C>T SNVs at CpG sites as these SNVs have
1002 been reported to correspond most closely to chronological age[57].

1003 To account for the differences in genome size and copy-number heterogeneity between different
1004 patients with varying amounts of aneuploidy, we normalize the number of C>T CpG SNVs on each
1005 branch by the number of bases being considered. First, we computed the effective genome length
1006 of each clone as the total size of the bins considered to be "clonal" for a valid copy-number state as
1007 defined above, with each bin weighted by its total copy number. Then, for the internal nodes of the
1008 tree, we assumed that the only copy-number changes to these bins were directly coupled to WGD
1009 events. Thus, for post-WGD branches, the genome length was identical to that of the leaves; and for
1010 pre-WGD branches, the genome length was computed using the correspondence described above
1011 between pre- and post-WGD copy numbers.

1012 *Estimating pre- and post-WGD changes in WGD subpopulations*

1013 We use a maximum parsimony based method to estimate pre- and post-WGD changes from
1014 estimated ancestral and descendent copy-number profiles. We proceed independently for each bin.
1015 Let *x* be the ancestral copy number state and *y* the descendent copy number state, and assume *y*
1016 is produced by a combination of pre-WGD CN change followed by WGD followed by post-WGD CN
1017 change. We can relate *x* and *y* using,

1018 $$y = 2(x + b) + a$$

1019 where *b* represents pre-WGD CN change and a post-WGD CN change. Let the *cost* of any given *a*
1020 and *b* be $|a| + |b|$. Conveniently, every combination of *x* and *y* results in a unique *a* and *b* that
1021 minimize this cost. Thus, for each *x* and *y* we compute the associated *b* and *a* as the pre- and post-
1022 WGD changes and $|a| + |b|$ as the cost of those changes.

1023 *Measures of diversity and heterogeneity*

1024 We compute the "percent genome different" for a pair of cells as follows. First, we compute the bin
1025 level difference in total copy number and identify consecutive segments of changed and unchanged
1026 bins. We then remove segments less than or equal to 2 MB in size (i.e., affecting fewer than four
1027 consecutive 500-kb bins). Finally, we count the number of bins for which the two genomes have
1028 different total copy number and divide by the total number of bins considered.

1029 *Classification of divergent cells*

1030 We call divergent cells as outliers of the nearest neighbor distance (NND), using "percent genome
1031 different" as the distance metric. For each index cell we compute its nearest neighbor as the other
1032 cell in the population within minimum percent genome different. The nearest neighbor distance for
1033 each cell is thus the percent genome different with respect to its neighbor cell. We then fit a beta
1034 distribution to the NND values of all cells in the cohort, and call divergent cells as those cells that
1035 have NND values in the 99th percentile of the beta distribution fit to the data.

1036 *Cell phylogenies using MEDICC2*

1037 We derived estimates of chromosome missegregation rates per cell for each patient from copy-
1038 number phylogenies inferred with MEDICC2[58]. First, we refined single-cell haplotype-specific copy-
1039 number profiles by applying the dynamic programming formulation from asmultipcf[59] on GC-
1040 corrected read counts and phased B-allele frequencies for each bin. Using this method, we identified
1041 segment boundaries across all cells for each patient and then summarized the number of copies of
1042 each segment and allele in each cell by rounding. Next, we ran MEDICC2[58] on these refined
1043 haplotype-specific single-cell copy numbers, which infers a tree (with single cells corresponding to
1044 leaves), copy-number profiles for the ancestral internal nodes of the tree, and copy-number events
1045 for each branch of the tree. We used the –wgd-x2 flag for MEDICC2 which represents WGD as an
1046 actual doubling of all copy-number segments in the genome, rather than the default behavior of
1047 adding 1 to all segments. We then computed missegregation rates by counting the number of inferred
1048 chromosome-/arm-level gains and losses on the terminal branches of the tree (i.e., the number of
1049 cell-specific events) and dividing by the total number of cells in the tree.

1050 *Reconstruction of ancestral copy number*

1051 To infer the ancestral haplotype-specific copy-number profiles associated with internal nodes of the
1052 cell phylogeny, we use a maximum parsimony approach that treats each bin independently and aims
1053 to minimize the total number of changes on the tree. Specifically, the parsimony score for each
1054 branch is the sum across bins and across both haplotypes of the absolute difference in copy number
1055 between the parent and the child. Transitions from 0 to any other copy number are given a score of
1056 infinity to prevent gain from 0 copies. The score for a WGD branch (assumed known from MEDICC2)
1057 is the sum of two parsimony scores: the score for copy-number changes between the parent and an
1058 intermediate genome, and the score for copy-number changes between a doubled version of the
1059 intermediate genome and the child (this intermediate genome is described above in *Estimating pre-*
1060 *and post-WGD changes in WGD subpopulations*). The state of each bin at each branch in the tree
1061 is chosen to minimize this parsimony score using the Sankoff algorithm[60,61]. We assume that the

1062 MEDICC2 placement of WGD on branches of the phylogeny is correct in all but two patients: for
1063 patients 025 and 045, we adjusted WGD placement to be concordant with SNV evidence suggesting
1064 independent clonal origin of multiple WGD clones.

1065 *Classifying event from CN differences*

1066 Given a phylogenetic tree where both leaves and internal nodes are labeled by haplotype-specific
1067 copy-number profiles, we identify the copy-number events on each branch using a greedy approach.
1068 First, we identify the differences between the parent haplotype-specific copy-number profile and the
1069 child copy-number profile. Then, for each chromosome and haplotype, we aim to explain the copy-
1070 number differences between parent and child using events that are as large as possible:

1071 1. If more than 90% of bins in the chromosome are altered in the same direction, we call a
1072 chromosome gain or loss that accounts for a change of one copy for all bins in the
1073 chromosome.
1074 2. If no chromosome gain or loss is found, but 90% of the bins in one of the two arms is altered
1075 in the same direction, we call an arm-level gain or loss that accounts for a change of 1 copy
1076 for all bins in the chromosome arm.
1077 3. If no chromosome- or arm-level gain or loss is found, we call a gain or loss of the largest
1078 contiguous segment that has a change in the same direction.

1079 We then adjust the copy number difference by the selected event, and repeat until all copy-number
1080 changes between parent and child have been accounted for. Note that if all but a few of the bins of
1081 a chromosome are gained (or lost), our method will first predict a chromosome gain (or loss), then
1082 an additional small segment loss (or gain) to account for the few bins that were predicted as
1083 unchanged. We have selected this approach as we consider a whole chromosome (or arm) change
1084 to be more parsimonious if most of a chromosome's (or arm's) bins are altered. Our approach is also
1085 more robust to bin level noise than a strategy that requires 100% of the bins to be altered.

1086 For branches with WGD, we compute the intermediate pre-doubling profile that would result in the
1087 fewest copy-number changes (see *Estimating pre- and post-WGD changes in WGD subpopulations*
1088 above). Using our bin-independent parsimony model, we can compute the optimal intermediate
1089 profile analytically. We then perform the event calling procedure described above twice: once on the
1090 differences between the parent and the intermediate pre-WGD profile, and again between the
1091 doubled intermediate profile and the child.

1092 *Estimating rates of cell specific events*

1093 We explored controlling for the "opportunity" for each cell to mis-segregate by dividing the number
1094 of copy-number events for each cell by the number of chromosomes (for chromosome-level
1095 missegregations) or arms (for arm-level missegregations) in the inferred parent node of each cell in
1096 the tree (i.e., the source of the terminal branch). This yields a rate of missegregation events per cell
1097 and per parental copy. For shorter "segmental" copy-number events, we divided the number of
1098 events in each cell by its parent's genome length to control for opportunity. While the resulting rate
1099 is not comparable to segment- and arm-level rates, it makes the cell-specific segmental rates more
1100 comparable between cells and across patients.

1101 *Detection of focal high-level amplifications in single cells*

1102 To detect focal high-level amplification in single cells, we used a two-stage approach compiling a set
1103 of potential amplified segments, then re-called amplification of those segments in individual cells.
1104 We first identified all contiguous segments with copy number exceeding 3× ploidy per cell. We then
1105 merged per cell segments to generate a set of amplified segments for the patient tumor cell
1106 population as a whole, and merged adjacent amplified segments if the boundaries of those segments
1107 were closer than 2MB. Only amplified segments larger than 500kb (1 bin) were considered further.
1108 Given a set of amplification segments predicted per patient, we then computed the average copy
1109 number for each cell within each segment, as well as the average copy number for the 5MB on either
1110 side of each segment. A focal high-level amplification was called in an individual cell if the average
1111 copy number of the amplification segment was greater than 3× ploidy and greater than 3× the
1112 average copy number in the boundary segments.

1113 *Enumerating events on ancestral branches*

1114 We computed gains and losses of chromosomes and chromosome arms for three classes of event
1115 timing. Events were classified as non-WGD if they were predicted to occur on the root branch of a
1116 Rare WGD patient, pre-WGD if they were predicted to occur prior to the WGD event on the root
1117 branch of a Prevalent WGD patient, and post-WGD if they were predicted to occur prior to the WGD
1118 event on the root branch of a Prevalent WGD patient. Patients 025, 045, and 081 were omitted from
1119 this analysis as their WGD history precludes this categorization of copy-number events.

1120 *Calculating post-WGD changes in WGD clones*

1121 We cataloged all high confidence WGD clones detected in our cohort. This included all predicted
1122 WGD clades with at least 20 cells in the MEDICC2 phylogenies. In addition, we included two small
1123 WGD clones from patient 006 and 031 (**Extended Data Fig. 3E-F**). Counts of post-WGD events

1124 were calculated from ancestral reconstruction on MEDICC2 trees as described above (see section
1125 *Reconstruction of ancestral copy number*).

1126 **Single-cell RNA sequencing**

1127 *Cell type assignment*

1128 Using scRNA-seq of CD45$^{+/-}$ sorted cells we assigned major cell types using supervised clustering
1129 with CellAssign[62], as described in Vázquez-García et al., 2022[14].

1130 *InferCNV and scRNA-seq derived copy number clonal decomposition*

1131 InferCNV (version 1.3.5) was used for identifying large-scale copy number alterations in ovarian
1132 cancer cells identified by CellAssign[63,64]. For each patient, 3,200 non-cancer cells annotated by
1133 CellAssign were randomly sampled from the cohort and used as the set of reference "normal" cells.
1134 After subtracting out reference expressions in non-cancer cells, chromosome-level smoothing, and
1135 de-noising, we derived a processed expression matrix which represents copy number signals.
1136 Cancer cell subclusters are identified by ward.D2 hierarchical clustering and "random_trees" partition
1137 method using p-value < 0.05.

1138 *WGD classification*

1139 Identification of WGD cells from scRNA data is technically challenging, as inferred copy number from
1140 expression data is typically noisy, allele-specific markers are sparse, and as shown in our scWGS
1141 analysis, the prevalence of non WGD cells in Prevalent WGD cases, and WGD cells in Rare WGD
1142 cases is generally low, confounding identification of non-clonal ploidy populations within samples.
1143 We reasoned that due to the high concordance between scWGS and scRNA derived copy number,
1144 even between non site-matched patient samples (**Extended Data Fig. 5A**), and the typically clonal
1145 nature of WGD, WGD status could be propagated to all available patient matched scRNA samples
1146 for the purposes of transcriptional phenotyping analysis. Furthermore, within-sample absolute
1147 normalization of UMI counts between tumor and non-tumor cells showed a significant increase in
1148 overall transcript counts per cell in Prevalent versus Rare WGD patients (**Extended Data Fig. 5B**),
1149 which was highly concordant with established estimates of transcriptional changes in WGD versus
1150 non-WGD samples in bulk RNA[65]. Thus, we concluded that site-matched scRNA data effectively
1151 captures WGD transcriptional phenotypes. Any analyses correlating scWGS derived missegregation
1152 rates to transcriptional phenotypes were restricted to site matched samples with at least 20 cells in
1153 both DLP and scRNA.

1154 *Cell cycle analysis*

1155 We identified circular trajectories linked to cell cycle progression in cancer cells using Cyclum[28].
1156 Across the cohort, 10,000 cancer cells annotated by CellAssign were randomly sampled across
1157 tumors and used for cell cycle trajectory inference. Pseudotime inference was run on the scaled cell-
1158 by-gene matrix, limiting genes to cell cycle markers included in cell cycle GO terms (GO:0007049).
1159 Discretization of the continuous pseudotime trajectories was accomplished using a three-component
1160 Gaussian mixture model. Discrete cell cycle phase information was computed using Seurat's
1161 CellCycleScoring function, excluding samples with fewer than 20 malignant cells. Smoothed
1162 pseudotime trajectories of cell cycle-related genes previously reported in the literature[66] were then
1163 evaluated to interpret phase-specific gene activity and phase transitions as a function of pseudotime
1164 (**Fig. 5D**).

1165 *Differential gene and pathway activity*

1166 Pathways were curated from the single-cell hallmark metaprograms[67], the 50 hallmark pathways[68],
1167 or CIN-associated gene signatures manually curated from literature, including inflammatory signaling
1168 and ER stress[31,38], and scored in single cells using Seurat's 'AddModuleScore' function. Due to the
1169 hierarchical nature of the data, with multiple samples from patients, we used generalized estimating
1170 equations (GEE) on sample mean gene or pathway expression levels, adding tumor site (adnexa vs
1171 non-adnexa) as a covariate in the model, and restricting analysis to samples with at least 20 cells in
1172 order to compare WGD states. *P*-values were adjusted for multiple testing using FDR. In parallel, we
1173 also performed differential expression analysis using a pseudobulked generalized linear mixed
1174 model (DREAMLET[69]), accounting for random patient and fixed tumor site effects, and performed
1175 gene set enrichment analysis (GSEA) with the same set of pathways.

1176 *Differential cell type abundance*

1177 To determine cell populations that were differentially abundant between rare WGD and prevalent
1178 WGD samples we utilized miloR v1.8.1[70], setting 'prop' to 0.2, and using 'tumor_megasite' (adnexa
1179 vs non-adnexa) as a contrast in the differential abundance testing. To obtain significance values for
1180 each cell population, we ran permutation tests by swapping the sample WGD status labels 1,000
1181 times, and computing the proportion of tests in which the resulting non-permuted median log2-fold
1182 change was more extreme than the permuted median values for each cell type.

1183 **Immunofluorescence**

1184 *Regions of interest*

1185 We defined regions of interest (ROIs) containing tumor on IF images by delineating regions with
1186 tumor foci, and contrasting these with images of the IF-adjacent H&E section. ROI annotations were
1187 drawn in QuPath. To ensure that complex tissue regions within ROIs used for analysis only included
1188 tumor, we classified regions of tumor, stroma, vasculature and glass within each ROI. We trained a
1189 pixel classifier with examples of tumor, stroma, vasculature and glass from each of the ROIs and
1190 slides using the IF-adjacent H&E section.

1191 *Segmentation of primary nuclei and micronuclei*

1192 Whole-slide IF images stained with cGAS, ENPP1 and DAPI were analyzed to characterize primary
1193 nuclei (PN) and micronuclei (MN) within ROIs. Segmentation of PN was carried out in QuPath v0.3.0
1194 using the StarDist algorithm on the DAPI channel[71]. We used a segmentation model pre-trained on
1195 single-channel DAPI images ("dsb2018_heavy_augment.pb"). Applying the PN segmentation model
1196 across all ROIs yielded 1,779,351 PN in tumor regions. Segmented PN ranged between 5 μm^2 and
1197 100 μm^2 in size, with a minimum fluorescence intensity of 1 a.u. The cell membrane for each PN
1198 was approximated using a cell expansion of 3 µm of the nuclear boundary.

1199 Micronuclei were detected by StarDist segmentation of cGAS spots. We trained a new segmentation
1200 model on single-channel cGAS images using a U-Net architecture. We manually annotated $cGAS^+$
1201 MN in a set of 256px x 256px tiles encompassing tumor regions across all slides. We created training
1202 and test sets using a 70:30 split, resulting in a training set of 70 tiles and a test set of 30 tiles. To
1203 ensure that the model generalized across patients and samples, we applied augmentation to the
1204 training set by applying random rotations, flips, and intensity changes. We monitored the loss
1205 function during model training and saved the trained model with frozen weights.

1206 This allows for whole slide quantification and cell-level annotation of PN and MN. Nuclear
1207 segmentation was also carried out using StarDist on the DAPI channel. Each MN was assigned to
1208 the closest PN. MN were excluded if they were >10µm from the centroid of the closest nucleus, had
1209 area >10μm^2 or probability <0.75.

1210 *Validation of micronuclei segmentation*

1211 We have evaluated our method on a test dataset with held-out MN labels, showing good performance
1212 of predicted MN segmentations with high average precision and F1 scores (IoU < 0.5). We
1213 quantitatively evaluated the segmentation performance on the test data by considering $cGAS^+$ MN
1214 objects in the ground truth to be correctly matched if there are predicted objects with overlap. We

1215 used the intersection-over-union (IoU) as an overlap criterion, demonstrating good performance with
1216 a chosen IoU threshold > 0.5.

1217 *Micronuclei rates*

1218 Micronucleus rupture rates were estimated based on the number of $cGAS^+$ MN and PN segmented
1219 within tumor ROIs. The rate of micronuclei rupture was estimated by localization of $cGAS^+$ MN
1220 neighboring PN. MN rate was calculated as the fraction of PN with 1 or more MN. Applying the MN
1221 segmentation model across all ROIs yielded 83,352 $cGAS^+$ MN in tumor regions, with a mean MN
1222 area of 2 μm^2, ranging between 1 μm^2 and 10 μm^2, and a minimum object probability of 0.75. To
1223 overcome batch effects, we used within-batch MN rate Z-score for downstream comparisons.

1224 *Statistical comparisons of micronuclei rates*

1225 For comparing MN rate between prevalent and rare WGD, we used generalized estimating equations
1226 (GEE). We used binary Prevalent vs Rare WGD as the dependent variable with binomial distribution
1227 and Z-score MN rate as the independent variable, adding patient as a group variable in the model.
1228 Reported effect size of WGD was calculated from the coefficient of Z-score MN rate in the learned
1229 model. For correlation between gain and loss rates and MN rate, we used a mixed linear model with
1230 Z-score MN rate as the dependent variable, gain or loss rate as the independent variable, and patient
1231 as a group variable.

1232 **Multi-modal sample matching**

1233 For integrative genotype-phenotype analyses, we utilized scRNA-seq data patient-matched with
1234 scWGS to profile cell type-specific abundance and gene/pathway activity changes in the context of
1235 WGD (**Figure 6**). Given the clonally dominant nature of each sample's WGD status, we reasoned
1236 that tumor cells identified in scRNA-seq within each patient would likewise be mostly clonal WGD or
1237 not, allowing for direct comparisons across all tumor cells in each patient. Indeed, site-matched
1238 scWGS and scRNA-seq derived estimates of copy number were highly concordant (**Extended Data**
1239 **Fig. 6A**), with UMI count ratios between tumor and normal cells being significantly elevated in
1240 Prevalent WGD compared to Rare WGD cases as expected (**Extended Data Fig. 6B**).

1241 **Mutational signatures**

1242 We analyzed mutational signatures by integrating SNVs and structural variations detected by either
1243 bulk WGS or scWGS in a unified probabilistic approach called multi-modal correlated topic models
1244 (MMCTM)[15].

1245 For bulk WGS samples, we obtained signature labels in the MSK SPECTRUM cohort (*n*=40) using
1246 MMCTM, as presented in Vázquez-García et al., 2022[14]. Mutational signatures for cases without
1247 bulk WGS data were assigned based on mutational signatures inferred from scWGS. For scWGS
1248 samples, we obtained signature labels in the MSK SPECTRUM cohort (*n*=40) using a ridge classifier
1249 with default regularization strength (α=1.0). This classifier was trained on the integrated SNV and
1250 SV signature probabilities, which were obtained using MMCTM[11] from HGSOC bulk whole genomes
1251 (*n*=170)[11].

1252 Consensus mutational signatures were preferentially derived based on: (i) MMCTM signatures
1253 derived from bulk WGS, and (ii) MMCTM signatures from scWGS. Mutational signatures for cases
1254 without bulk WGS data (006, 044, 046, 071) or inconclusive bulk WGS assignments (004, 045, 080,
1255 081) were resolved based on scWGS.

1256 **Analysis of RPE1 cell line experiments**

1257 *10X scRNA pre-processing*

1258 Raw 10X sequencing data were aligned using CellRanger (version 7.0.0), which also performed
1259 barcode filtering and unique molecular identifier (UMI) gene counting using the 10X GRCh38
1260 reference transcriptome.

1261 *10X Multiome pre-processing*

1262 Raw 10X sequencing data were aligned to the 10X GRCh38 reference transcriptome using
1263 CellRanger ARC (version 2.0.2). CellRanger ARC also performed barcode filtering and unique
1264 molecular identifier (UMI) gene counting to generate feature-barcode matrices for both RNA and
1265 ATAC modalities.

1266 *scATAC copy number analysis*

1267 Copy number was inferred from the scATAC component of the 10X multiome data for RPE-D, RPE-
1268 Noco and RPE-Rev samples. Blacklist filtered fragments were first counted in 10MB genome bins.
1269 Bins with GC content of less than 30% were removed prior to performing GC correction using modal
1270 regression[16]. Cells with more than 5% of their bins containing NA values after GC modal correction
1271 were removed from subsequent analysis. GC corrected counts were smoothed using the DNACopy
1272 R package (v1.73.0) 'smooth.CNA' function, setting 'smooth.region'=4. Smoothed counts were
1273 mean-normalized per cell prior to clustering using Seurat (v5)[72]. For visualization, mean-normalized
1274 and smoothed counts were scaled binwise to emphasize copy differences between clusters.

1275 *scRNA copy number analysis*

1276 Copy number was inferred from 10X scRNA for the RPE-WGD sample using Numbat (v1.4.0)[73] to
1277 preprocess and smooth expression counts . Smoothed counts were then rebinned to 500Kb, bins,
1278 reduced to 50 dimensions by PCA, and then clustered using Leiden clustering at 1.0 resolution on a
1279 SNN graph.

1280 *Identification of WGD subclones*

1281 A spontaneously arising WGD subclone was observed in all DLP+ samples, characterized by gain
1282 of 1p and loss of 1q, 2q, 4q and 21 (**Extended Data Fig. 6E**). The same WGD clone was evident
1283 copy number inferred from both scATAC for RPE-D, RPE-Noco and RPE-Rev (**Extended Data Fig.**
1284 **6D**) and scRNA for RPE-WGD (**Extended Data Fig. 6I**). For RPE-D, RPE-Noco and RPE-Rev, our
1285 aim was to characterize the phenotypic impact of CIN in non-WGD cells. Thus we excluded scRNA
1286 cells in the scATAC inferred WGD cluster from further analysis. For RPE-WGD we aimed to
1287 characterize the phenotypic differences between WGD and non-WGD cells. We thus used the
1288 scRNA based copy number clusters to label cells as either WGD or non-WGD in that sample.

1289 *Estimating rates of cell specific events from DLP+*

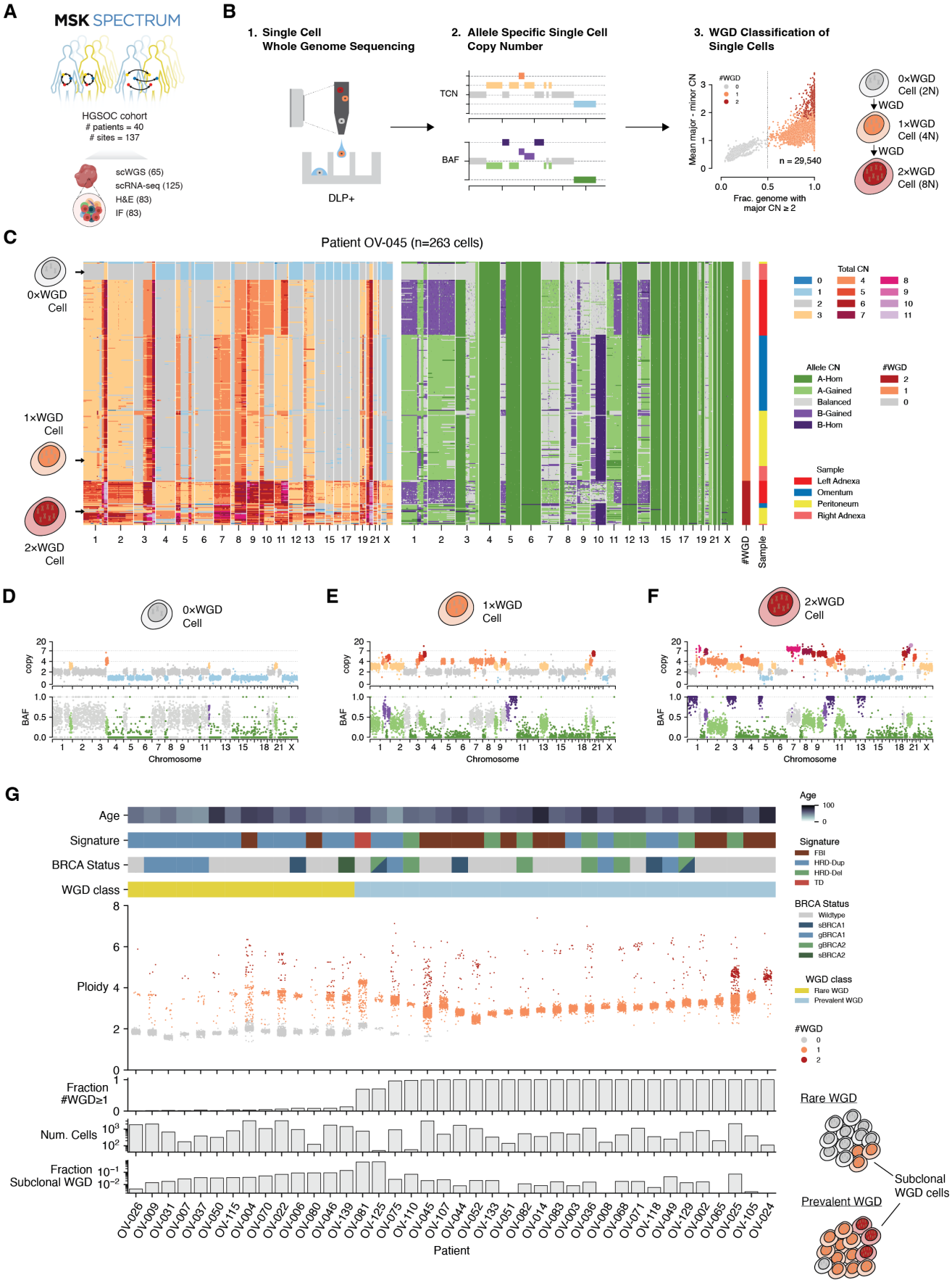
1290 We inferred cell-specific rates of copy number change from DLP+ data using similar methods to
1291 those applied to the patient data. We first removed low quality and cycling cells as described above.
1292 For RPE-D, RPE-Noco and RPE-Rev we removed cells with ploidy > 2.5, thereby removing the WGD
1293 clone and other WGD cells. We then used MEDICC2 to infer a phylogeny independently for each
1294 sample, computed cell specific changes and classified those changes into chromosome, arm, and
1295 segment as described above.

1296 **FIGURES**

Figure 1	Whole-genome doubling is a dynamic mutational process
Figure 2	Evolutionary timing of WGD events from single nucleotide variants
Figure 3	Impact of WGD on rates of chromosomal instability at single-cell resolution
Figure 4	Modes of evolution post WGD
Figure 5	Cell cycle progression in the context of whole genome doubling
Figure 6	Tumor cell phenotypes and microenvironment remodeling in the context of whole genome doubling
Extended Data Figure 1	Study and cohort overview
Extended Data Figure 2	Quality control of scWGS data and WGD inference
Extended Data Figure 3	Non-WGD subclones and subclonal WGD
Extended Data Figure 4	Single cell measurements of chromosomal instability
Extended Data Figure 5	Cell cycle progression in the context of whole genome doubling
Extended Data Figure 6	Tumor cell phenotypes and microenvironment remodeling in the context of whole genome doubling

1297

Figure 1: Whole genome duplication is a dynamic mutational process



1298 **Figure 1. Whole genome doubling is a dynamic mutational process**

1299 **A.** Overview of the MSK SPECTRUM cohort and specimen collection workflow.

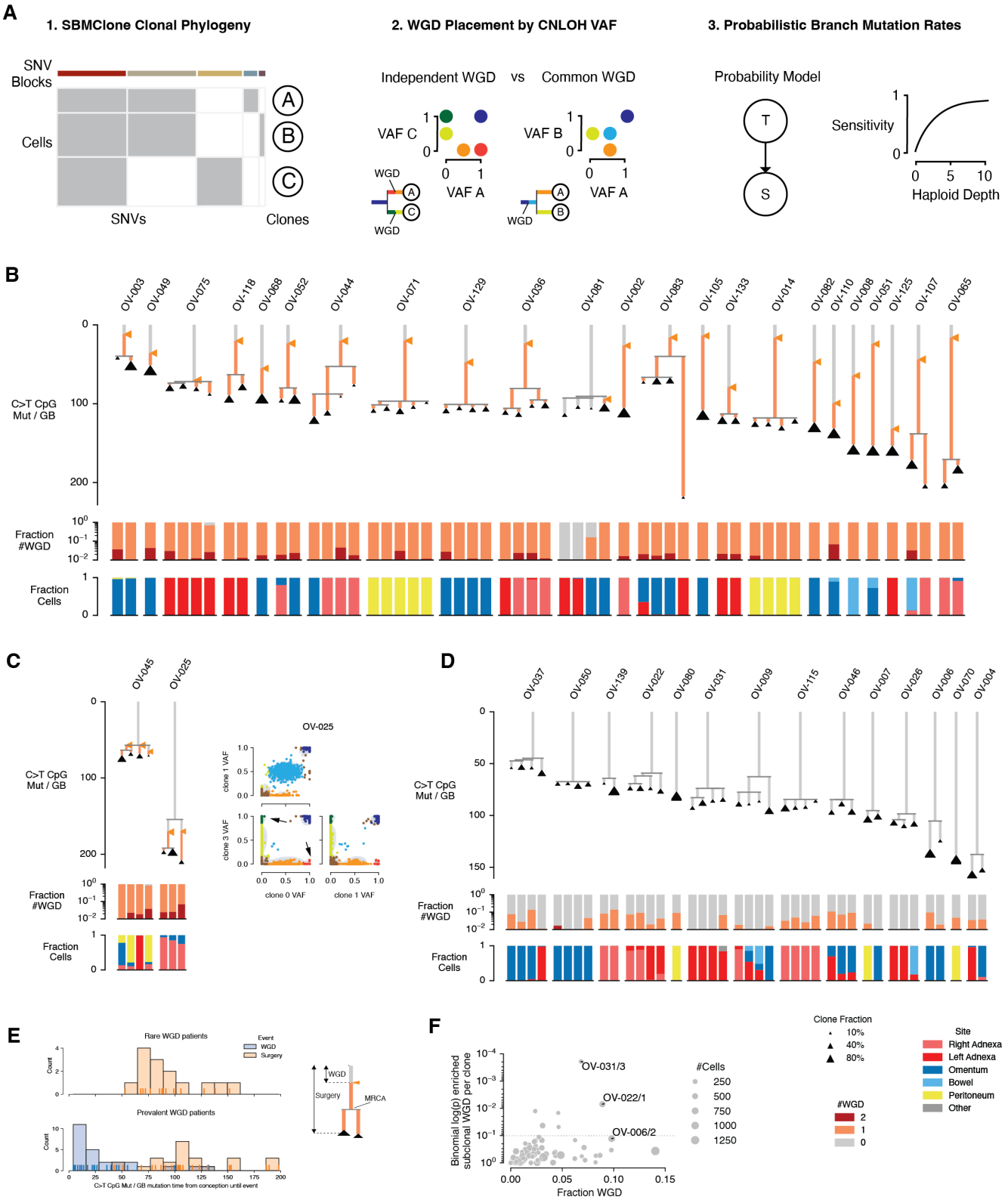
1300 **B.** Study design for analyzing cellular ploidy and WGD in single cells using scWGS with the DLP+
1301 protocol. Right-hand plot shows classification of WGD multiplicity in cancer cells (# WGD=0, 1, or 2)
1302 using fraction of the genome with major CN ≥ 2 (x-axis) vs mean allele CN difference (y-axis).

1303 **C.** Heatmap of total (left) and allele specific (right) copy number for patient 045, with predicted #WGD
1304 and site of resection for each cell annotated. The dominant 1×WGD population was downsampled
1305 from 1,857 to 200 cells, and the full 0×WGD and 2×WGD populations numbering 18 and 44 cells
1306 respectively are shown.

1307 **D-F.** Example 0×WGD, 1×WGD, and 2×WGD cells from patient 045.

1308 **G.** Distribution of cell ploidy (middle y-axis) of individual cells for each tumor, colored by # WGD. Age
1309 at diagnosis, mutation signature, *BRCA1/2* mutation status, and WGD class are annotated at top; %
1310 WGD and number of cells per patient are annotated at bottom.
1311

Figure 2: Evolutionary timing of WGD events from single nucleotide variants



1312 **Figure 2. Evolutionary timing of WGD events from single nucleotide variants**

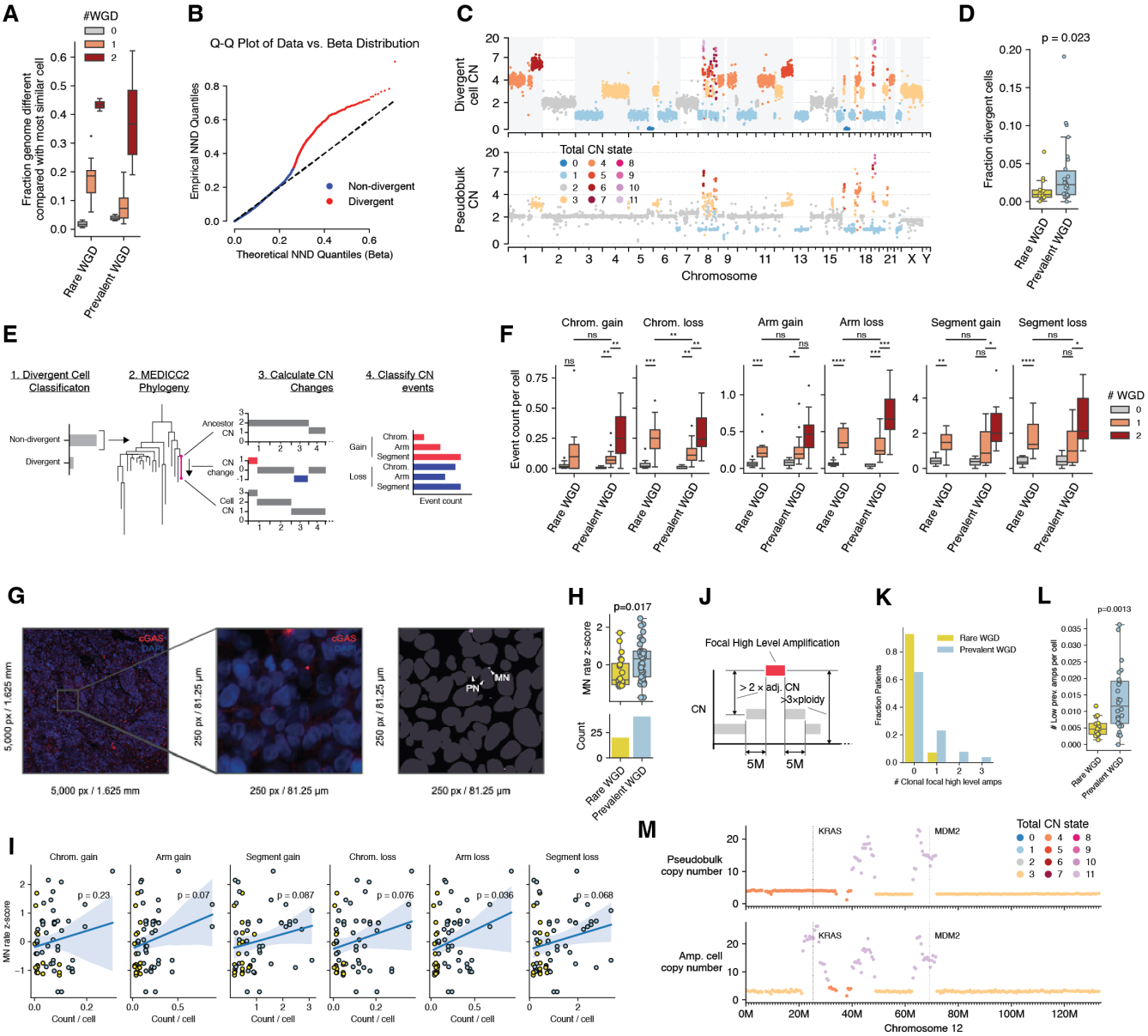
1313 **A.** Schematic of the approach for timing WGDs in SNV clones. SBMClone is used to infer clones
1314 based on SNVs, and a phylogeny is constructed from presence/absence patterns of SNVs across
1315 SNV clones (left). For each pair of WGD clones, independence of the WGD is determined through
1316 analysis of the SNV VAF in clonal cnLOH regions (center). Predictions of independent vs common
1317 WGD are used to place WGD events in the tree. A probabilistic method is used to assign SNVs to
1318 the tree including placing SNVs before or after WGD events (right). The method models the
1319 relationship between depth of coverage and SNV sensitivity to account for clones of differing size.

1320 **B-D.** Clone phylogenies for the 39 patients for which the SNV based method could be applied.
1321 Length of branches show the number of age-associated SNVs (C to T at CpG) assigned to each
1322 branch, adjusted for coverage-depth-related reduction in SNV sensitivity. Clone sizes as a fraction
1323 of the patient's total sequenced cells are shown by the size of the triangle for each leaf. Clonal WGD
1324 events are represented as orange triangles at the predicted location along WGD branches, and
1325 branches are colored according to the number of WGD at that point in the evolutionary history. The
1326 fractions of each clone with each #WGD state and from each sampled site are shown below each
1327 clone tree. Each patient is annotated with mutation signature and age at diagnosis.

1328 **E.** Histogram and rug plot showing the sensitivity-adjusted age-associated SNV count for WGD and
1329 diagnosis events for rare WGD (top) and prevalent WGD (bottom) patients.

1330 **F.** Fraction of +1 WGD cells within each clone (x axis) and log binomial *p*-value for the test that a
1331 clone has a greater fraction of +1 WGD cells than the overall +1 WGD fraction for the patient.
1332

Figure 3: Impact of WGD on rates of chromosomal instability at single-cell resolution



1333 **Figure 3: Impact of WGD on rates of chromosomal instability at single-cell resolution**

1334 **A.** Divergence as measured by nearest neighbor distance, where distance is represented as the
1335 fraction of the genome with different CN. NND is calculated for each population of cells within each
1336 patient. Boxplots show the mean NND for each WGD population within each patient.

1337 **B.** QQ plot of a beta fit (x-axis) vs empirical (y-axis) quantiles of NND values for all cells in the cohort.
1338 Divergent cells, defined as outliers (>99 percentile) of the beta distribution, are shown in red.

1339 **C.** CN profile of an example divergent cell from patient 004 (top) compared to pseudobulk CN of all
1340 cells for that patient (bottom). Shaded regions show differences between cell and pseudobulk CN.

1341 **D.** Fraction of divergent cells in Rare vs Prevalent WGD patients.

1342 **E.** Method for computing cell specific events in non-divergent cells.

1343 **F.** Event counts per cell for loss and gain of chromosomes, arms, and large segments, split by #WGD
1344 state and Prevalent vs Rare WGD patient status. Mann-Whitney U test significance is annotated as
1345 'ns': $5.0\times10^{-2} < p \leq 1.0$, '*': $1.0\times10^{-2} < p \leq 5.0\times10^{-2}$, '**': $1.0\times10^{-3} < p \leq 1.0\times10^{-2}$, '***': $1.0\times10^{-4} < p \leq$
1346 1.0×10^{-3}, '****': $p \leq 1.0\times10^{-4}$.

1347 **G.** Left: Low-magnification IF image of FFPE tumor section from a representative HGSOC patient,
1348 stained with DAPI (DNA) and anti-cGAS antibody. Middle: High-magnification inset. Right: cGAS
1349 segmentation mask of MN in the foreground and DAPI segmentation mask of PN in the background.

1350 **H.** Z-scored MN rate split by Prevalent vs Rare WGD patient status.

1351 **I.** Patient CN event counts per cell for loss and gain of chromosomes, arms, and large segments (x-
1352 axis) compared with slide specific Z-scored MN rate (y-axis). Points are colored by Prevalent vs
1353 Rare WGD patient status.

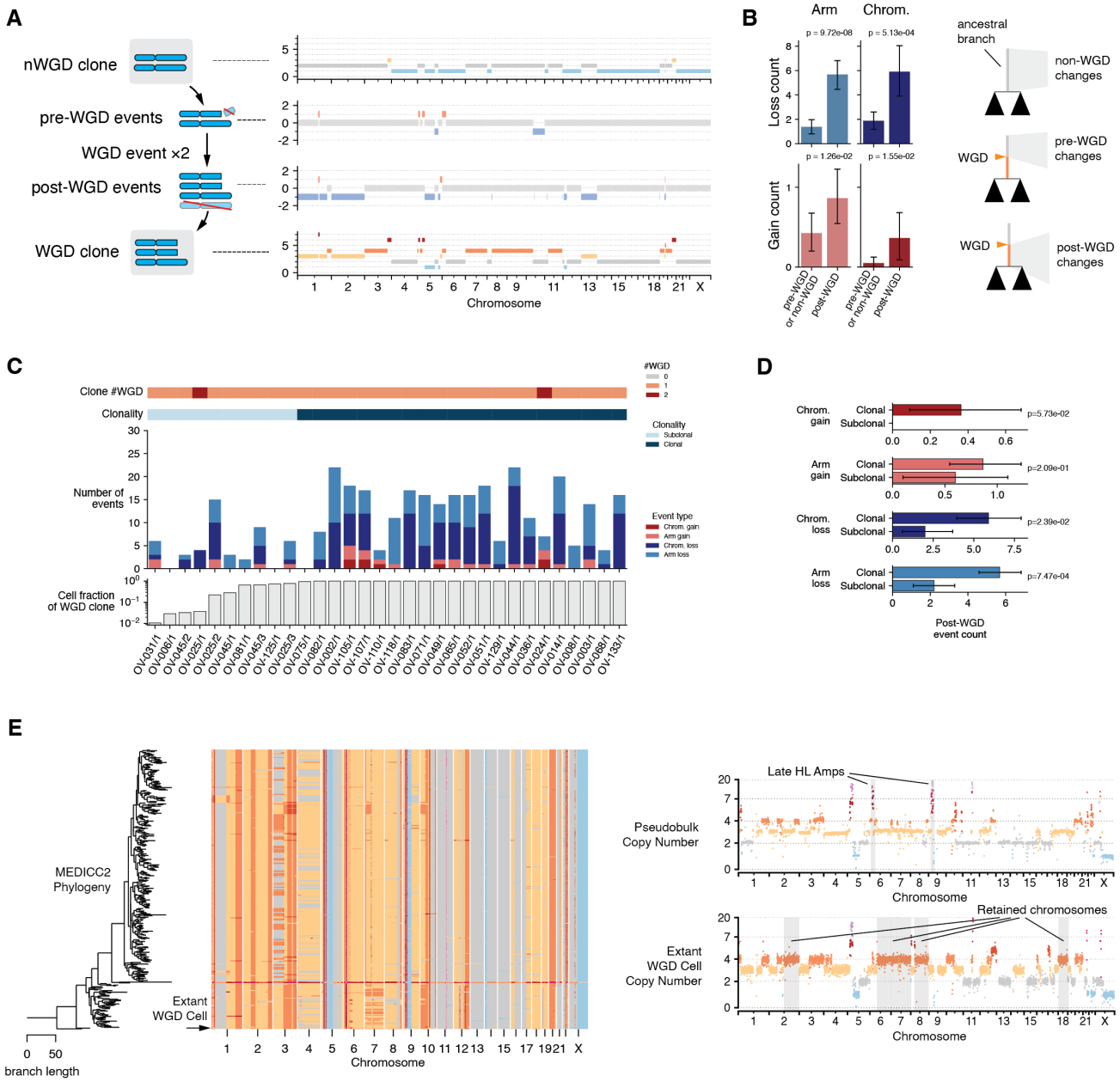
1354 **J.** Diagram defining focal high-level amplification.

1355 **K.** Count of clonal focal high-level amplifications per patient split by Prevalent vs Rare WGD.

1356 **L.** Count per cell of low-prevalence focal high-level amplifications split by Prevalent vs Rare WGD.
1357 Low prevalence was defined as occurring in 2 or more cells but less than 10% of the patient cell
1358 population.

1359 **M.** Example low prevalence focal high-level amplification found in patient 002 (bottom) and not
1360 detectable in the pseudobulk copy number of the same patient.

Figure 4: Modes of evolution post WGD



1361 **Figure 4. Modes of evolution post WGD**

1362 **A.** Pre- and post-WGD events illustrated for the ancestral branch of patient 044.

1363 **B.** Counts of ancestral arm and chromosome events detected across the cohort, grouped into pre-
1364 or post-WGD.

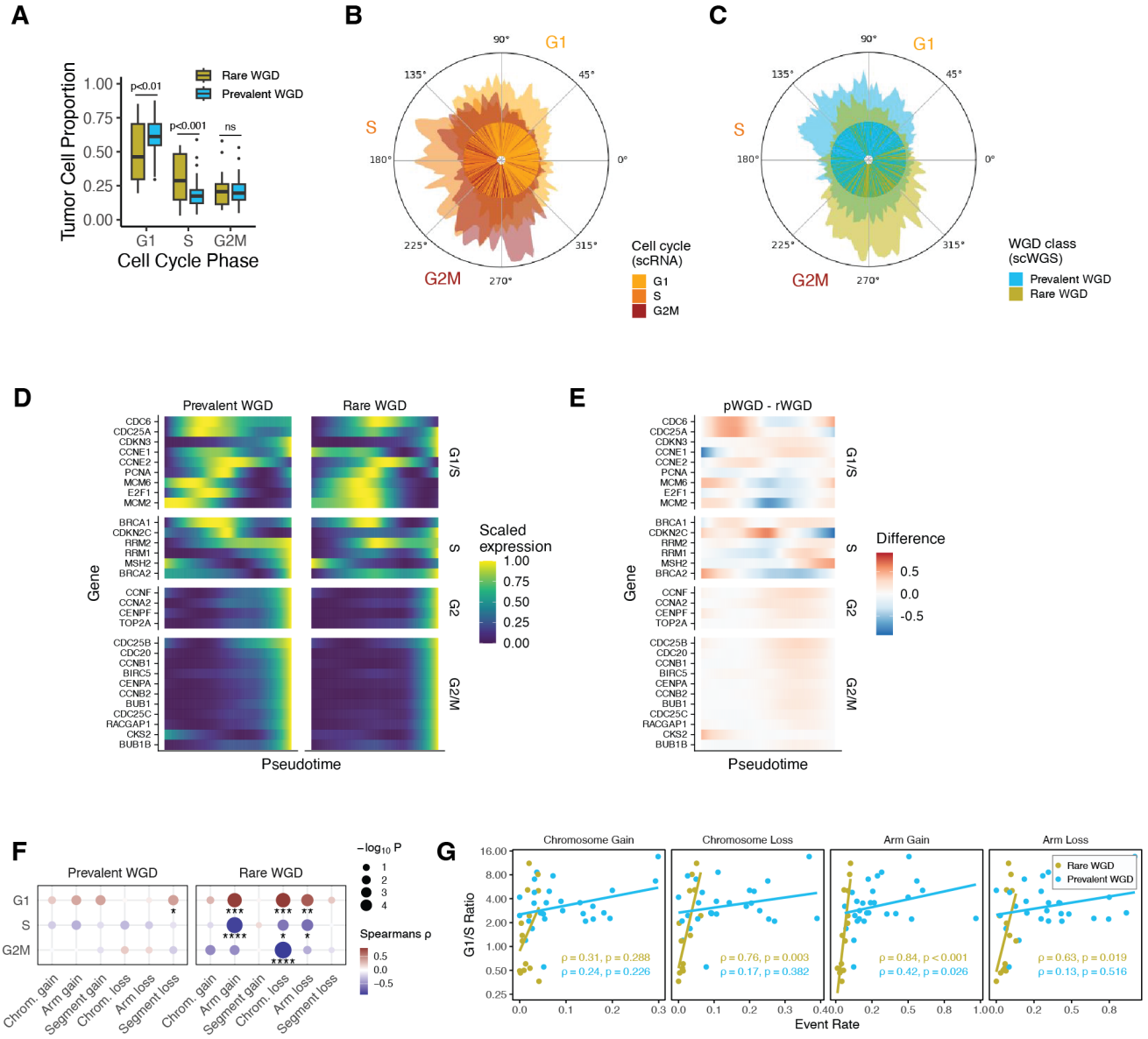
1365 **C.** Counts of arm and chromosome events occurring post-WGD for all high-confidence clonal and
1366 subclonal WGD events detected across the cohort, split by clonality of the WGD (cell fraction
1367 threshold 0.99).

1368 **D.** Boxplots summarizing C annotated with *p*-values (Mann-Whitney U test).

1369 **E.** MEDICC2 tree and copy number for patient 014 (left). The outgroup cell is shown with missing
1370 focal HL Amps (bottom right) compared with the majority of cells represented as a pseudobulk (upper
1371 right) for which there have been additional losses and focal HL Amps.

1372 **F.** CN (top) and allelic imbalance (bottom) for two divergent WGD clones from patient 083 (left, right)
1373 with shared WGD origin. Late and divergent HL Amps and losses of distinct alleles are highlighted.
1374

Figure 5: Cell cycle progression in the context of whole-genome doubling



1375 **Figure 5. Cell cycle progression in the context of whole genome doubling**

1376 **A.** Proportion of cancer cells (y-axis) grouped by cell cycle phase (x-axis) in Prevalent WGD vs Rare
1377 WGD tumors (color).

1378 **B-C.** Cell cycle pseudotime inference in cancer cells. Inner ring shows cell cycle pseudotime in
1379 cancer cells and outer ring shows smoothed density estimate. B: Cell cycle assignment. C:
1380 Pseudotime grouped by Prevalent WGD (cyan) and Rare WGD (yellow) tumors.

1381 **D.** Scaled expression of phase-specific genes in Prevalent vs Rare WGD tumors as a function of cell
1382 cycle pseudotime.

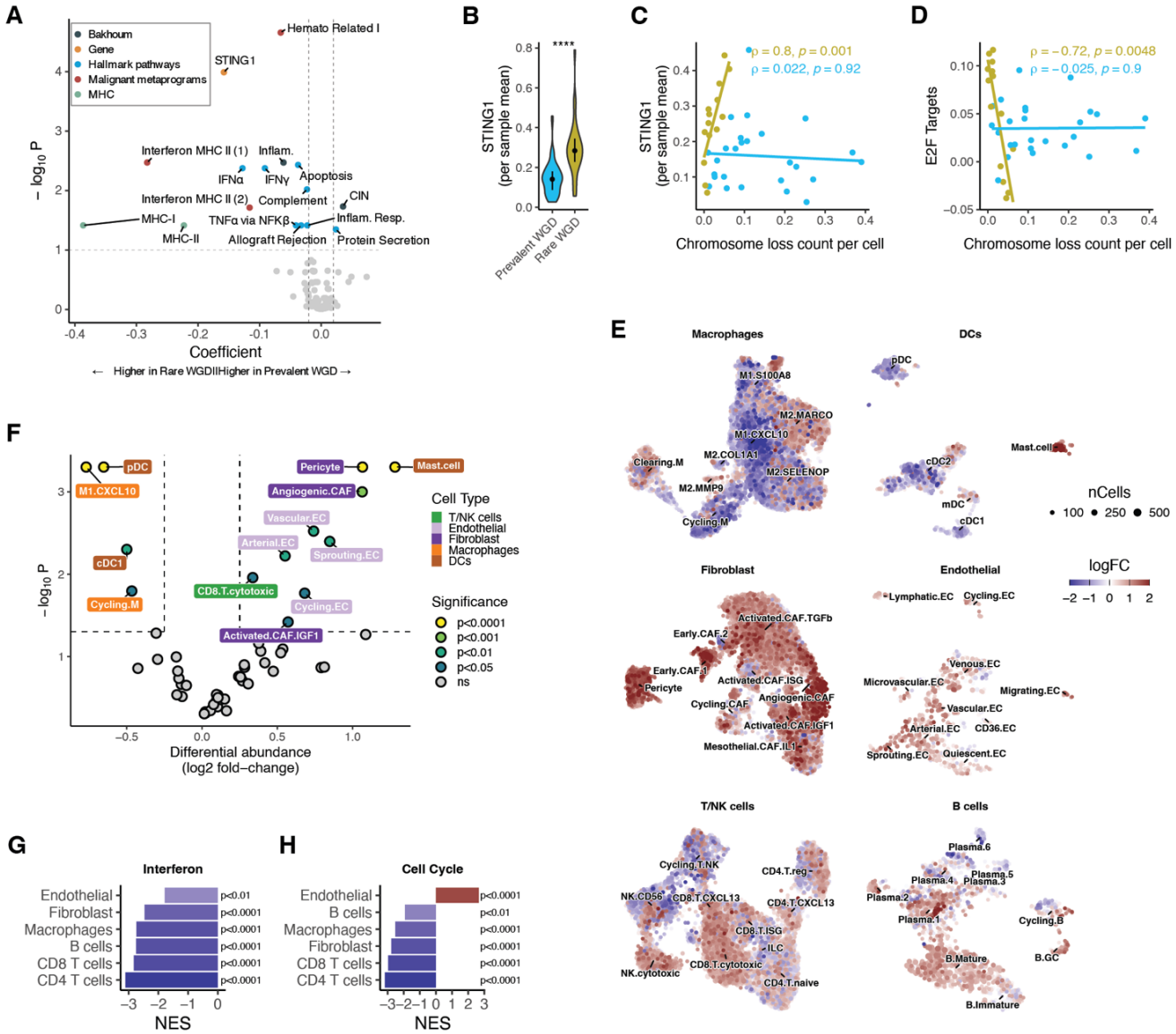
1383 **E.** Differences in scaled gene expression of phase-specific genes in Prevalent vs Rare WGD tumors
1384 as a function of cell cycle pseudotime.

1385 **F.** Dotplot of correlations between missegregation rates derived from scWGS and cell cycle phase
1386 from scRNA in site-matched samples.

1387 **G.** Scatter plot of G1/S cell count ratios (y-axis) by rates (counts per cell) of large chromosomal
1388 changes (x-axis) split by Rare and Prevalent WGD (color). Regression coefficients and significance
1389 results are shown separately for Rare and Prevalent WGD.

1390

Figure 6: Tumor cell phenotypes and microenvironment remodeling in the context of whole-genome doubling



1391 **Figure 6. Tumor cell phenotypes and microenvironment remodeling in the context of whole**
1392 **genome doubling**

1393 **A.** Scatter plot depicting regression coefficients (x-axis) and significance (y-axis) for selected genes
1394 and pathways in Prevalent versus Rare WGD tumor cells.

1395 **B.** Per-sample mean expression of STING1 in Prevalent and Rare WGD samples.

1396 **C.** Scatter plot of STING1 gene expression (y-axis) by rate (counts per cell) of chromosomal losses
1397 (x-axis) split by Rare and Prevalent WGD (color). Regression coefficients and significance results
1398 are shown separately for Rare and Prevalent WGD patients.

1399 **D.** Scatter plot of hallmark E2F module score (y-axis) by rate (counts per cell) of chromosomal losses
1400 (x-axis) split by Rare and Prevalent WGD (color). Regression coefficients and significance are shown
1401 separately for Rare and Prevalent WGD patients.

1402 **E.** UMAP showing differential cell state enrichment in Prevalent versus Rare WGD samples in
1403 different TME cell types.

1404 **F.** Differential cell-type abundance testing results for cell types in Prevalent versus Rare WGD
1405 samples.

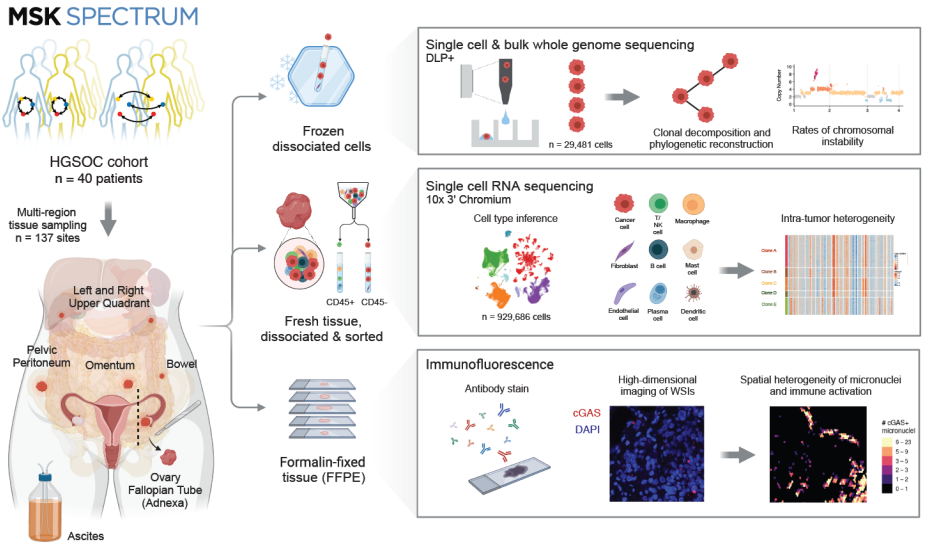
1406 **G.** Normalized enrichment scores (NES) for the interferon pathway across TME cell types.

1407 **H.** Normalized enrichment scores (NES) for the cell cycle pathway across TME cell types.

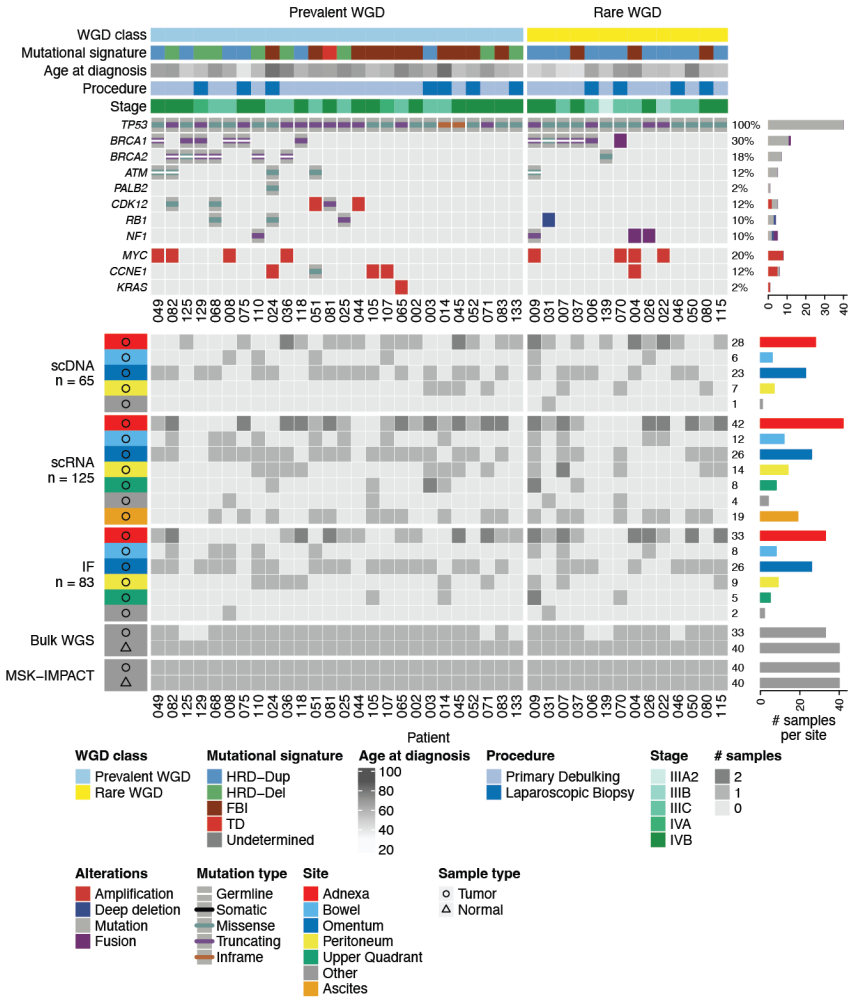
1408

Extended Data Figure 1: Study and cohort overview

A



B

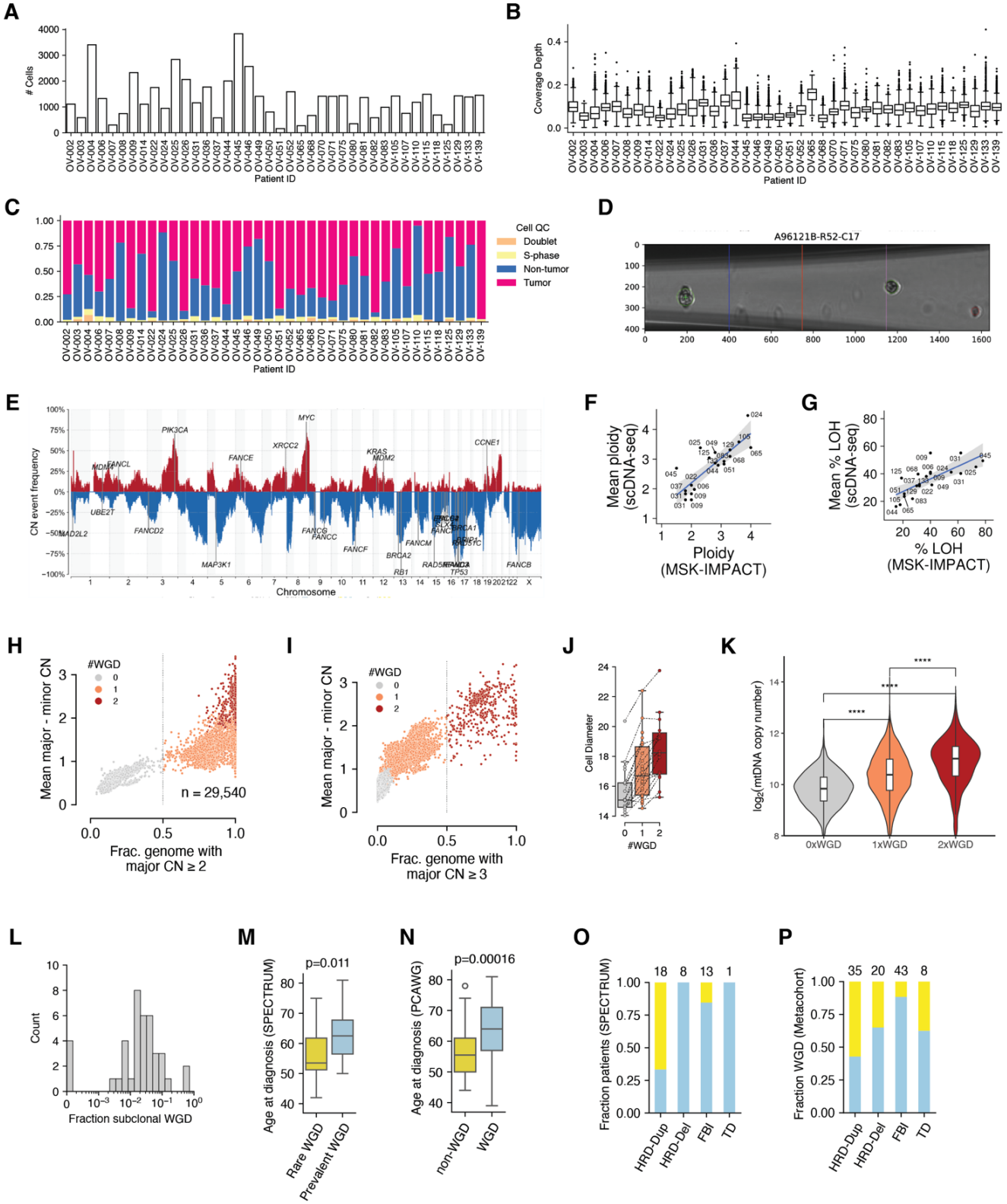


1409 **Extended Data Figure 1. Study and cohort overview**

1410 **A.** Schematic of the MSK SPECTRUM specimen collection workflow including primary debulking
1411 surgery or laparoscopic biopsy, single-cell suspensions for scWGS and scRNA-seq, and biobanking
1412 of snap-frozen and FFPE tissue samples.

1413 **B.** Cohort overview. Top panel: Oncoprint of selected somatic and germline mutations per patient
1414 and cohort-wide prevalence. Single nucleotide variants (SNVs), indels, and fusions shown are
1415 detected by targeted panel sequencing (MSK-IMPACT). Focal amplifications and deletions are
1416 detected by single-cell whole genome sequencing (scWGS). Patient data include WGD class,
1417 mutational signature subtype, patient age, staging following FIGO Ovarian Cancer Staging
1418 guidelines, and type of surgical procedure. Bottom panel: Sample and data inventory indicating
1419 number of co-registered multi-site datasets: single-cell whole genome sequencing, single-cell RNA
1420 sequencing, H&E whole-slide images, immunofluorescence, bulk WGS and bulk MSK-IMPACT.
1421

Extended Data Figure 2: Quality control of scWGS data and WGD inference



1422 **Extended Data Figure 2. Quality control of scWGS data and WGD inference**

1423 **A.** Number of high-quality cells generated per patient.

1424 **B.** Distributions of per-cell coverage depth per patient.

1425 **C.** Fraction of cells called as tumor, non-tumor, doublet, and S-phase for each patient.

1426 **D.** Example doublet identified from an image taken during DLP+ sequencing.

1427 **E.** Frequency of gains (red, above the horizontal) and losses (blue, below the horizontal) among all
1428 single-cell genomes in the cohort, with known drivers genes annotated.

1429 **F.** Ploidy (mean copy number) for each patient in the SPECTRUM cohort as measured by MSK
1430 IMPACT (x-axis) and scWGS (y-axis).

1431 **G.** Fraction of the genome with loss of heterozygosity (LOH) for each patient in the SPECTRUM
1432 cohort as measured by MSK IMPACT (x-axis) and scWGS (y-axis).

1433 **H.** Shown for all quality-filtered cells in the cohort is the mean difference between major and minor
1434 copy number (y-axis) versus the fraction of the genome with major copy number ≥ 2 (x-axis), with
1435 cells colored by #WGD state. The dashed line at 0.5 denotes the decision boundary for 0 vs 1
1436 WGDs.

1437 **I.** Shown for all quality filtered cells in the cohort is the mean difference between major and minor
1438 copy number (y-axis) versus the fraction of the genome with major copy number ≥ 3 (x-axis), with
1439 cells colored by #WGD state. The dashed line at 0.5 denotes the decision boundary for 1 vs 2 WGDs.

1440 **J.** Cell diameter measured from DLP+ images. Each point is the mean cell diameter within a given
1441 patient for 0×, 1× or 2×WGD cells. Points representing cells from the same patient are connected by
1442 dashed lines. Boxplots show the distribution of means for each WGD state.

1443 **K.** Distribution of mitochondrial DNA copy number (log2) inferred from scWGS in 0×, 1×, and 2×WGD
1444 cells.

1445 **L.** Distribution over patients of the fraction of cells within each patient with subclonal WGD, i.e., 1
1446 more WGD than the dominant population for that patient.

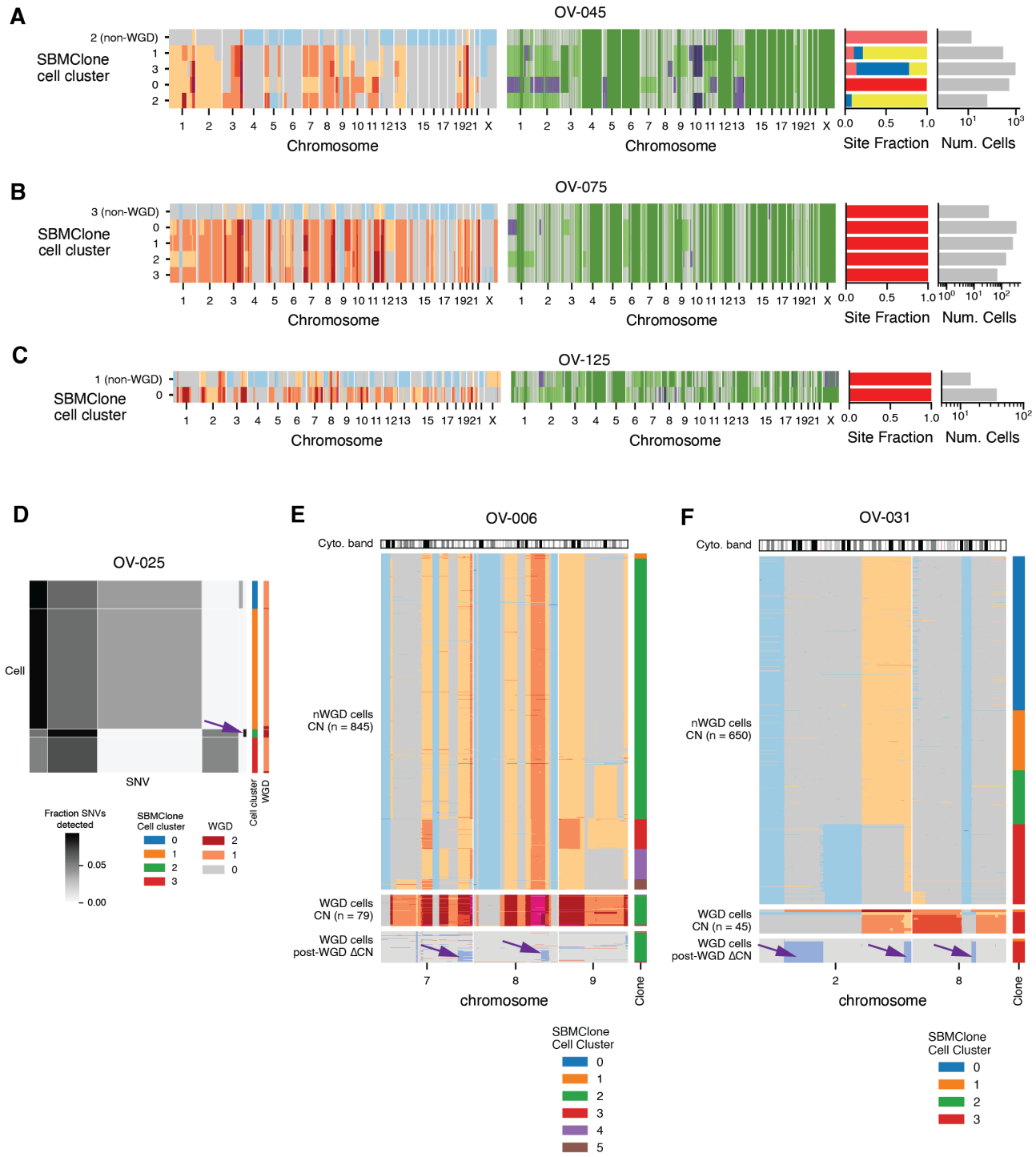
1447 **M.** Age at diagnosis for patients in the SPECTRUM cohort split by Prevalent vs Rare WGD.

1448 **N.** Age at diagnosis for patients in the PCAWG ovarian cohort split by WGD vs non-WGD.

1449 **O.** Fraction of Prevalent and Rare WGD patients in the SPECTRUM cohort for each mutational
1450 signature.

1451 **P.** Fraction non-WGD and WGD patients in the Ovarian Metacohort for each mutation signature.
1452

Extended Data Figure 3: Non-WGD subclones and subclonal WGD



1453 **Extended Data Figure 3: Non-WGD subclones and subclonal WGD**

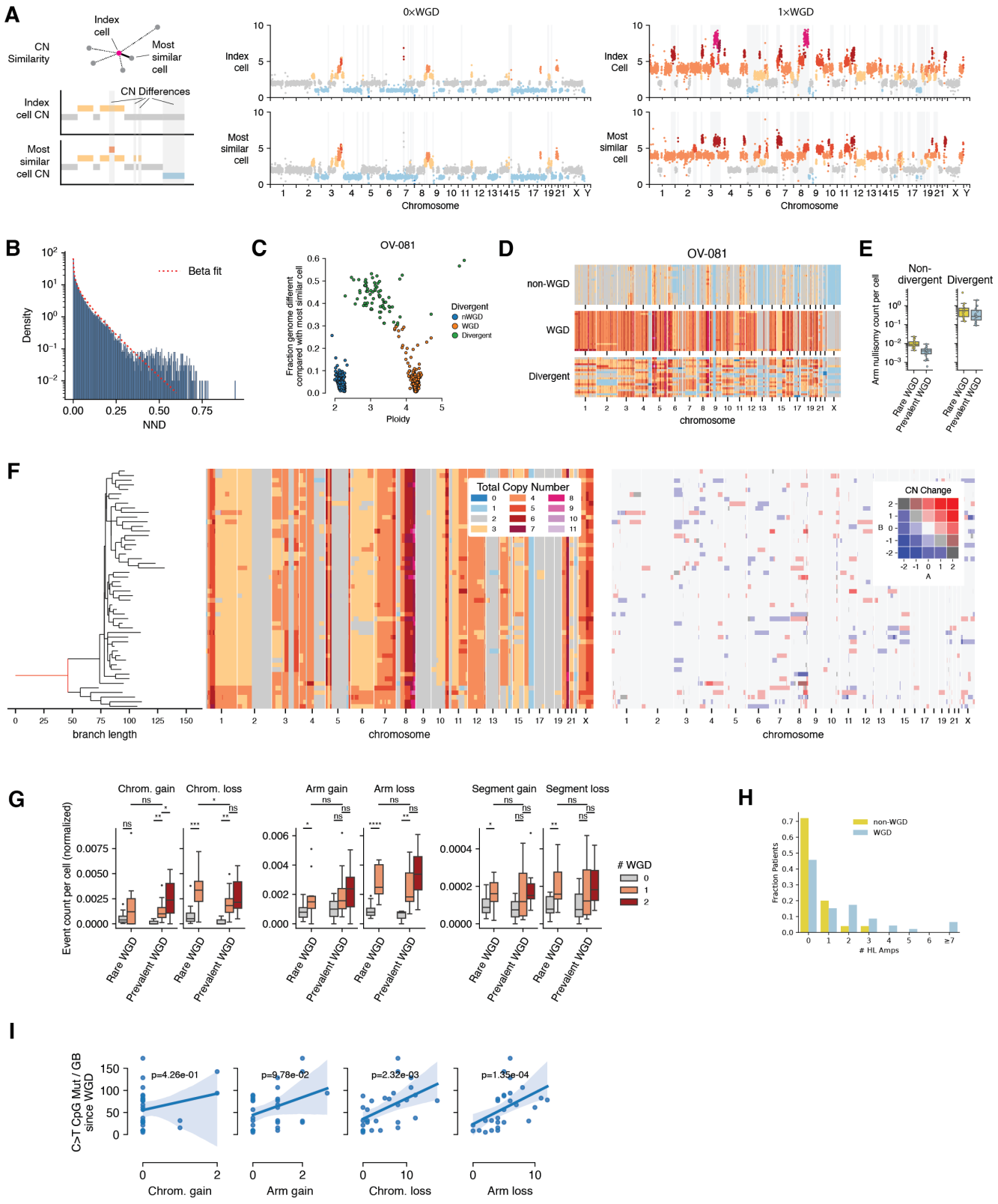
1454 **A-C.** 0×WGD subpopulations in patients 045 (A), 075 (B) and 125 (C). Shown for each patient is the
1455 total (left) and allele specific (middle) copy number for each clone (y-axis). At right are the fraction of
1456 cells from that clone found in each anatomic site (left) and the number of cells for each clone (right).

1457 **D.** SBMClone block density matrix for patient 025 showing the proportion of SNVs detected for each
1458 clone (y-axis) and SNV block (x-axis). The SBMClone cluster and WGD status of each cell are shown
1459 on the right. The 2×WGD clone in patient 025 is distinguished by clone-specific SNVs (arrow).

1460 **E.** Copy number for chromosomes 7, 8, and 9 for cells in patient 006, separated into non-WGD cells
1461 (top), WGD cells (middle), and inferred post-WGD changes in WGD cells (bottom). The cell order is
1462 the same for the middle and bottom plots. Arrows indicate shared post-WGD changes that represent
1463 a WGD subclone.

1464 **F.** Copy number for chromosomes 2 and 8 for cells in patient 031, separated into non-WGD cells
1465 (top), WGD cells (middle), and inferred post-WGD changes in WGD cells (bottom). The cell order is
1466 the same for the middle and bottom plots. Arrows indicate shared post-WGD changes that represent
1467 a WGD subclone. .
1468

Extended Data Figure 4: Single cell measurement of chromosomal instability



1469 **Extended Data Figure 4: Single cell measurement of chromosomal instability**

1470 **A.** Schematic of nearest neighbor difference (NND) using fraction of the genome different as a
1471 distance measure (left). Shown are the nearest neighbors and regions of the genome that are
1472 different for a 0×WGD cell (middle) and a 1×WGD cell (right).

1473 **B.** Empirical distribution of NND for all cells, and beta distribution fit (red).

1474 **C**. NND (y-axis) by ploidy (x-axis) for cells from patient 081. Color indicates #WGD and divergent
1475 status.

1476 **D**. Copy-number profiles for example 0×WGD (top), 1×WGD (middle) and divergent (bottom) cells
1477 from patient 081.

1478 **E.** Arm nullisomy rates (counts per cell) for divergent and non-divergent cells in rare and prevalent
1479 WGD patients. Shown is the distribution of mean rates per population in each patient.

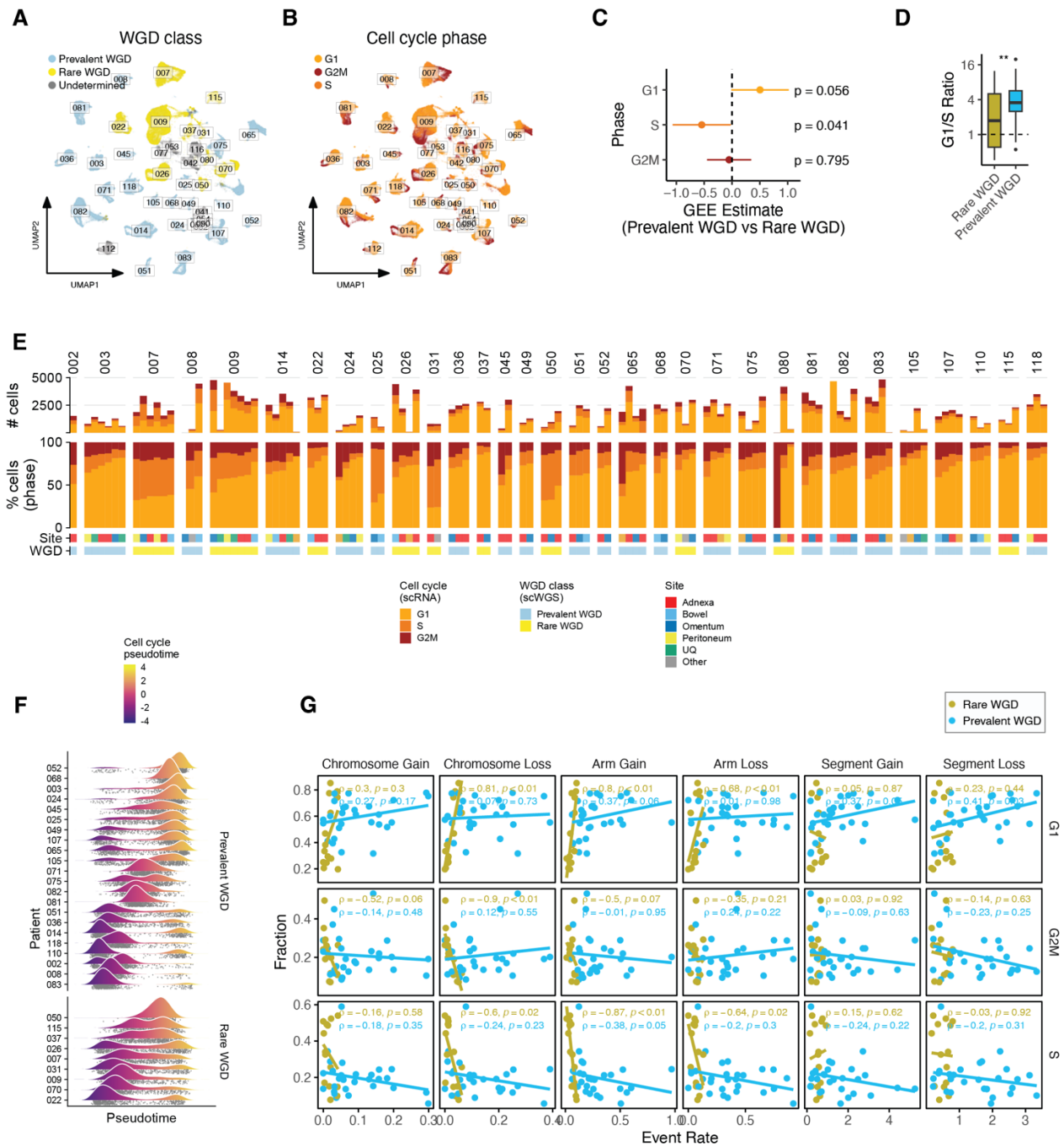
1480 **F.** MEDICC2 phylogeny (left) total copy number (center) and inferred cell specific copy number
1481 changes (right) for patient 110.

1482 **G.** Rates of chromosome, arm, and segment losses and gains (counts per cell) normalized for
1483 increased or decreased opportunity for an event based on genomic content in each cell's ancestor.
1484 MWU significance is annotated as 'ns': $5.0\times10^{-2} < p <= 1.0$, '*': $1.0\times10^{-2} < p <= 5.0\times10^{-2}$, '**': $1.0\times10^{-3} < p <= 1.0\times10^{-2}$, '***': $1.0\times10^{-4} < p <= 1.0\times10^{-3}$, '****': $p <= 1.0\times10^{-4}$.
1485

1486 **H** Number of focal high level amplifications per patient detected in the PCAWG ovarian cohort, split
1487 by WGD vs non-WGD.

1488 **I.** Number of post-WGD chromosome and arm gains and losses (x-axis) compared to the mutation
1489 time in C>T CpG counts (y-axis) measured since the WGD event.
1490

Extended Data Figure 5: Cell cycle progression in the context of whole-genome doubling



1491 **Extended Data Figure 5: Cell cycle progression in the context of whole genome doubling**

1492 **A.** UMAP of cancer cells colored by Rare vs Prevalent WGD patient labels.

1493 **B.** UMAP of cancer cells colored by inferred cell cycle state.

1494 **C.** Coefficients (x-axis) of a Generalized Estimation Equation (GEE) fit to the difference in cancer
1495 cell cycle fractions between Rare and Prevalent WGD samples, corrected for patient effects.
1496 Significance of WGD effect on cell cycle fractions are shown at right.

1497 **D.** Distribution of G1/S cancer cell cycle ratios for Rare and Prevalent WGD samples.

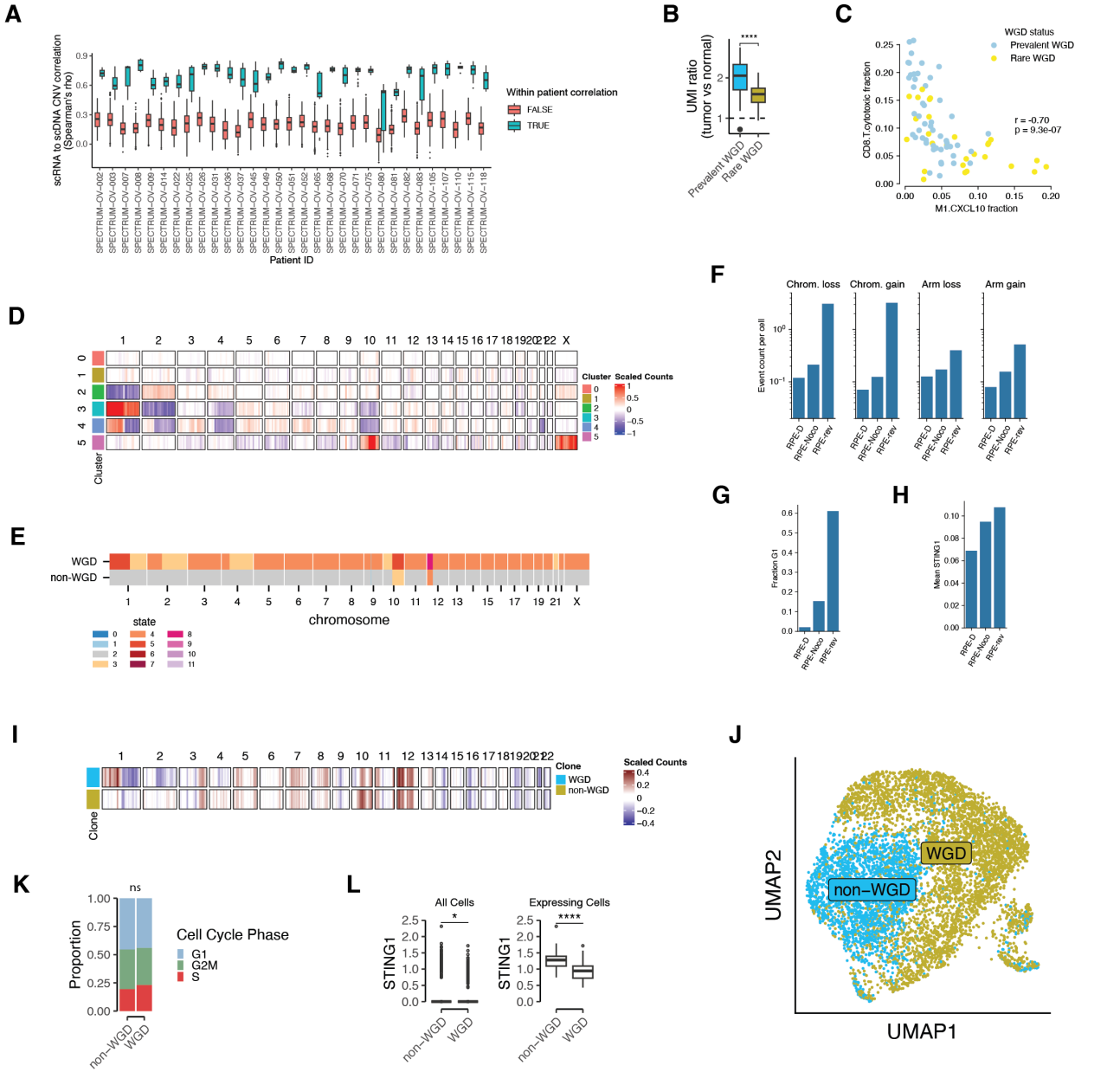
1498 **E.** Absolute and relative compositions of cell cycle fractions in $CD45^-$ sorted samples based on
1499 scRNA-seq. Samples are separated by patient and ordered by proportion of S-phase cells out of all
1500 cancer cells.

1501 **F.** Distribution of cell cycle pseudotime estimates over all cells for each patient, separated into
1502 Prevalent WGD (top) and Rare WGD (bottom).

1503 **G.** Correlation between the fraction of cancer cells in G1, S and G2M phase (y-axis) and rates
1504 (counts per cell) of chromosome, arm, and segment losses and gains (x-axes).

1505

Extended Data Figure 6: Tumor cell phenotypes and microenvironment remodeling in the context of whole genome doubling



1506 **Extended Data Figure 6: Tumor cell phenotypes and microenvironment remodeling in the**
1507 **context of whole genome doubling**

1508 **A.** Correlation between DLP+ and scRNA based copy number. Data points for box plots are scRNA
1509 inferCNV copy number clusters. The y-axis shows correlation between each DLP+ sbmclone cluster
1510 and each scRNA copy number cluster from the same patient (blue). As a comparator we show the
1511 same correlation computed with each DLP+ SBMClone cluster from any other patient (red).

1512 **B.** Ratio of cancer cell UMI counts to fibroblast and endothelial cell UMI counts, averaged within
1513 each patient. Patients are grouped by Rare vs Prevalent WGD.

1514 **C.** Cytotoxic $CD8^+$ T cells (y-axis) and *CXCL10*$^+$*CD274*$^+$ Macrophages (x-axis) as fractions of $CD45^+$
1515 cells across $CD45^+$ samples. Points are colored by the WGD class of the patient from which the
1516 sample originated.

1517 **D.** Copy number inferred from scATAC for RPE1 cells across treatment conditions.

1518 **E.** Clone copy number inferred from DLP for RPE1 cells across treatment conditions. Two clones
1519 were identified: one WGD and one non-WGD.

1520 **F.** Chromosome and arm loss and gain events per cell for non-WGD RPE1 cells treated with DMSO
1521 control (RPE-D), nocodazole (RPE-noco) and reversine (RPE-rev).

1522 **G.** Fraction of non-WGD RPE1 cells within G1 phase (y-axis) for each treatment condition: DMSO
1523 control (RPE-D), nocodazole (RPE-noco) and reversine (RPE-rev).

1524 **H.** Average STING1 expression (y-axis) for non-WGD RPE1 cells by treatment condition (x-axis):
1525 DMSO control (RPE-D), nocodazole (RPE-noco) and reversine (RPE-rev).

1526 **I.** WGD and non-WGD copy number clones inferred from scRNA-seq of sample RPE-WGD.

1527 **J.** Expression UMAP from scRNA-seq of sample RPE-WGD with cells colored by assignment to the
1528 WGD and non-WGD clones.

1529 **K.** Cell cycle fractions for WGD and non-WGD clones in the RPE-WGD sample.

1530 **L.** Expression of STING1 across all cells (left) and in cells with positive expression (right) in the
1531 RPE-WGD sample.

1532

1533 **TABLES**

Supplementary Table 1

Clinical overview of the MSK SPECTRUM patient cohort. Data include patient age at diagnosis, staging following FIGO Ovarian Cancer Staging guidelines, type of surgical procedure, WGD class, and mutational signature subtype.

Supplementary Table 2

Sample inventory. Metadata associated with scWGS, scRNA-seq, H&E, IF, bulk tumor and normal WGS, and tumor and normal MSK-IMPACT datasets.

1534