Evolution of oncogene amplification across 86,000 cancer cell genomes

Jake June-Koo Lee[1,2,3]†*, Sohrab Salehi[1,2,4]†, Matthew A. Myers[1], Marc J. Williams[1], Melissa A. Yao[5,6], Duaa H. Al-Rawi[1,2], Jin Lee[2], Eric G. Sun[5,7,8], Kerstin Thol[1], Seongmin Choi[1], Eliyahu Havasov[1], Asher Preska Steinberg[1], Michelle Wu[1], Nicole Rusk[1], Caitlin Timmons[1,8], Cheryl Zi Jin Phua[1,8], Stephen Martis[1], Neeman Mohibullah[9], Parvathy Manoj[2], Esther Redin[2], Álvaro Quintanal-Villalonga[2], Pedram Razavi[2], Samuel Aparicio[10,11], Natasha Rekhtman[12], Viviane Tabar[5,13], Mark M. Awad[2,3], Helena A. Yu[2,3], Kenny Kwok Hei Yu[5,13], Andrea Ventura[5], Andrew McPherson[1], Charles M. Rudin[2,3], Sohrab P. Shah[1,3]*

Affiliations:

1. Halvorsen Center for Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
2. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
3. Weill Cornell Graduate School of Medical Sciences, Weill Cornell Medicine, New York, NY, USA.
4. Irving Institute for Cancer Dynamics, Columbia University, New York, NY, USA.
5. Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
6. Louis V. Gerstner Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
7. Tri-Institutional MD-PhD Program, Weill Cornell Medicine, Rockefeller University, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
8. Tri-Institutional PhD Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY, USA.
9. Integrated Genomics Operation, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
10. Department of Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, BC, Canada.
11. Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.
12. Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
13. Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

* Corresponding author. Email: shahs3@mskcc.org (S.P.S.); leej39@mskcc.org (J.J.L.)

† These authors contributed equally to this work.

Abstract:

High-level copy-number (CN) amplification (HLAMP) is a major mechanism of oncogene activation in human cancer. Despite progress in therapeutically targeting HLAMPs, the processes underlying HLAMP evolution are incompletely understood, leaving critical knowledge gaps in their etiology and mechanisms of therapeutic response. To address this, we analyzed using novel computational approaches, the evolutionary trajectories of HLAMP in single-cell whole-genome sequencing of 86,239 cancer cells from 94 patients and 8 experimental systems. We found that two common etiologies of oncogene amplification–intrachromosomal amplification (ICamp) and extrachromosomal circular DNA (ecDNA)–lead to distinct CN distributions across cells in amplitude, variance and clone specificity. Notably, ICamp events exhibited widespread subclonal specificity, indicating their dynamic evolution through numeric or structural modulation associated with clonal expansions. In contrast, ecDNAs exhibited higher cell-to-cell CN variation linked to structural rearrangements, whereby complex ecDNA architectures were resolved at single-nucleotide resolution in individual cell genomes. Through joint analysis of ecDNA structure and their cellular abundance, we observed both divergent and convergent modes of ecDNA evolution shaped by DNA damage, selection, and chromosomal re-integration. Finally, modeling single-cell distributions of ecDNAs substantially improved bulk genome-graph based predictions, revealing that ecDNA prevalence is tissue-type specific, rather than broadly shared across cancer types, with implications for ecDNA-directed therapy.

Main Text:

High-level copy-number amplification (HLAMP) is a common mechanism of oncogene activation in human cancers and often creates a targetable functional dependence. For example, anti-HER2 therapies in HER2-positive breast cancer[1] and other cancer types[2,3] have enabled effective precision cancer medicine in previously intractable tumors. In contrast to other primary oncogenic alterations, such as point mutations and gene fusions, HLAMPs often exhibit profound intratumoral heterogeneity. For example, *ERBB2* (encoding HER2) copy number can differ between primary and metastatic sites[4], across regions within a tumor[5], and among individual cells[6]. Such heterogeneity has been associated with poor clinical outcomes and therapy resistance[7], yet its mechanistic basis remains incompletely understood. One plausible explanation is amplification via extrachromosomal circular DNA (ecDNA), which lacks centromeres and therefore can undergo unequal segregation during cell division[8]. Recent studies indicate that non-Mendelian inheritance[9], together with unique biophysical properties of ecDNA[10], can drive clonal evolution and lead to heterogeneous cancer cell populations[11,12]. By contrast, intrachromosomal amplification (ICamp) is expected to segregate symmetrically with chromosomes. How ICamps evolve and contribute to observed cellular heterogeneity remains unclear. Recently, therapeutic approaches exploiting selective vulnerabilities of ecDNA-containing cancer cells have been proposed[13–15], underscoring the clinical importance of accurately distinguishing HLAMP mechanisms and characterizing their evolutionary dynamics.

To map the mechanisms underlying HLAMP etiology, evolution and their intratumoral heterogeneity in clinical tumor samples, we analyzed a large-scale single-cell whole-genome sequencing (scWGS) dataset consisting of >86,000 single cells genomes derived from >100 samples across a range of solid epithelial malignancies and experimental models. This enabled both direct quantification of cell-to-cell variability in copy number (CN)[16,17] and nucleotide resolution SV-resolved segment boundaries at HLAMP loci. From these data, we developed a probabilistic mixture model, `eicicle`, informed by evolutionary simulations to distinguish between ICamp and ecDNA based on single-cell CN distributions. We found that ICamps diversify at a clone level pervasively through two mechanisms: numeric modulation whereby the copy number architecture is maintained, but arm-level or chromosomal-level aneuploidies increase or decrease copy number in a clone-specific manner; and structural modulation whereby SV-driven mutational processes drive subclonal diversification. In ecDNAs we observed cases harboring multiple ecDNA species, and developed `ECADeMix`, a mixed-integer linear programming framework, to decompose ecDNA heteroplasmy at single-cell resolution through simultaneous inference of ecDNA structures and cellular prevalence reflective of evolutionary histories. Finally, our scWGS-based approach underscores limitations of bulk genome graph-based ecDNA

predictions and reveals notable tissue-type specificity[18] of ecDNAs, informing patient selection for emerging ecDNA-directed therapeutics.

Cell-to-cell copy number variance distinguishes ecDNA and ICamp

We analyzed direct library preparation (DLP+)-based scWGS data from 102 individuals (patients and experimental models), including 71 published[17,19–21] and 31 newly generated cases (**Fig. 1A, top** and **table S1**). These included clinical tumor samples and patient-derived xenografts from 94 patients across four cancer types: high-grade serous ovarian cancer (HGSOC; n=62), invasive breast cancer including hormone receptor-negative (HR−BC; n=10; one with *ERBB2* amplification) and -positive (HR+BC; n=5), lung cancer including small-cell (SCLC; n=8) and non-small cell (NSCLC; n=5; all with *EGFR* mutation), and glioblastoma multiforme (GBM; n=4). In addition, we studied 8 experimental models including cell lines (n=7) and a mouse model of

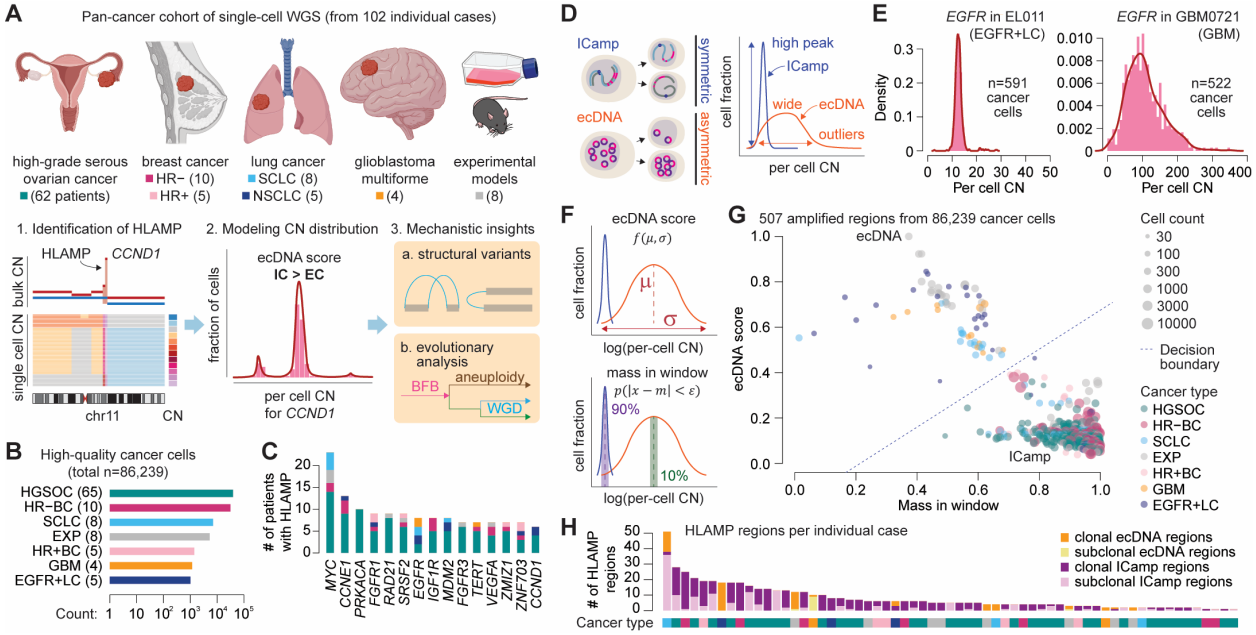

Fig. 1. Cell-to-cell copy-number variance in scWGS differentiates ecDNA-mediated amplifications from intrachromosomal amplifications. (**A**) Cancer types and case counts (top) and schematic illustration of study design (bottom). Organ and model illustrations are from BioRender. (**B**) Number of high-quality cancer cells analyzed (x-axis, log-scale) by tumor type (y-axis). (**C**) Frequently amplified oncogenes and their case counts. (**D**) Inheritance patterns for intrachromosomal amplification (ICamp) versus extrachromosomal circular DNA (ecDNA; **left**) and expected CN distributions across cells (right). (**E**) Representative examples of ICamp and ecDNA, from a patient with *EGFR*-mutated lung adenocarcinoma (left) and glioblastoma (right). (**F**) Conceptual illustration of two metrics used to classify HLAMP events. (**G**) Classification of 507 HLAMP regions from 86,239 cancer cells based on single-cell CN distribution. (**H**) Fraction of clonal and subclonal HLAMP regions by mechanism, across cases.

113 liver cancer harboring engineered *Mdm2*-containing ecDNA[22] (n=1). Single-cell WGS data were
114 assessed for quality and jointly processed with matched normal bulk WGS data for somatic variant
115 calling, as described previously[20]. After removing normal cells and low-quality cells, the dataset
116 consisted of 86,239 high-quality cancer cells (**Fig. 1B**). To systematically analyze the generative
117 and evolutionary mechanisms of HLAMP[23], we integrated somatic structural variant (SV)
118 breakpoints detected at the pseudobulk level with genome-wide CN bins to generate SV-aware
119 single-cell CN profiles (**Methods**). Using pairwise CN distance between single cells, we identified
120 subclones and reconstructed their evolutionary relationships (**Methods**). Across the 102 cases, we
121 identified 507 HLAMP regions with i) CN at least 3 times the chromosome arm-level baseline in
122 one or more subclones and ii) a pseudobulk CN of at least 6. These HLAMP regions typically
123 included *bona fide* oncogenes known to be activated by HLAMP, such as *MYC*, *CCNE1*, *EGFR*,
124 and *MDM2* (**Fig. 1C**), in addition to loci devoid of known oncogenic drivers, often amplified in
125 conjunction with other HLAMP oncogene-containing regions. For each HLAMP region
126 (comprising one or more contiguous SV-aware segments), we evaluated CN distributions across
127 cells, determined generative mechanisms from SV features at their boundaries, and analyzed
128 clone-specific features to delineate evolutionary mechanisms driving subclonal diversification
129 (**Fig. 1A, bottom**).

130 First, we performed *in silico* simulation of single-cell CN distributions accounting for the mode of
131 inheritance of oncogene amplifications and evolutionary constraints (**Methods**) [9,12,24,25]. These
132 simulations showed that conditioned on symmetric cell division, as expected for ICamps, CN
133 distributions consist of prominent, narrow peaks reflecting cell populations that share one or a
134 discrete set of CN values (**Fig. 1D)**. By contrast, conditioned on asymmetric cell divisions, as
135 expected for ecDNA[12], broad, heavy-tail distribution with extreme outliers were observed,
136 reflecting CN variance across the cell population. Modulating key parameters further simulated
137 fitness consequences on ecDNA CN distributions: stronger positive selection (e.g., higher potency
138 of oncogene) shifted the modal CN to the right, while greater cell-death penalties (e.g., DNA
139 damage responses at extreme CN) shifted it to the left (**Fig. S1** and **Methods**). Consistent with
140 these simulations, we observed two distinct and quantitatively distinguishable classes of single-
141 cell CN distributions across HLAMP regions in our cohort. For example, in the *EGFR*-mutated
142 lung cancer case EL011, the *EGFR* allele harboring the E746-A750 microdeletion was highly
143 amplified and exhibited a narrow, high-CN peak from 591 cancer cells, consistent with ICamp
144 (**Fig. 1E, left** and **Fig. S2A**). In contrast, the surgically resected glioblastoma GBM0721 showed
145 a broad, heavy-tail distribution of wild-type *EGFR* CN across 522 cancer cells, consistent with
146 ecDNA (**Fig. 1E, right** and **Fig. S2B**). Motivated by these patterns, we developed a machine
147 learning-based classifier, `eicicle`, to distinguish ecDNA from ICamp (**Fig. 1F**). `Eicicle` includes
148 two metrics: *ecDNA score* which captures distribution width, and *mass-in-window* which
149 quantifies the fraction of cells near the modal CN (high in ICamp, low in ecDNA). These metrics

are derived from a probabilistic mixture model that conditionally models wide, heavy tailed distributions, representing ecDNA, and narrow distributions representing ICamp (**Methods**). The 507 HLAMP regions by these metrics were bimodally distributed (**Fig. 1G**). We performed a discriminant analysis to infer a decision boundary, designating 15 HLAMP regions with high ecDNA scores and low mass-in-window (left upper corner) as representative ecDNA and 6 regions with low ecDNA scores and high mass-in-window (right lower corner) as representative ICamp (**Fig. S3**). This boundary classified 72 HLAMP regions from 11 cases as ecDNA and 435 regions from 55 cases as ICamp. We then orthogonally validated classifier predictions across all available experimental models (5 ecDNA-positive and 3 ecDNA-negative by `eicicle`; **Fig. S4A-C**) using metaphase DNA FISH. In all models classified as ecDNA, we observed dispersed extrachromosomal fluorescent signals for the amplified oncogenes, consistent with ecDNA. In models classified as ICamp, we observed strong intrachromosomal clustered signals. In two models classified as ecDNA by `eicicle`, ecDNA coexisted with homogeneously staining regions (HSRs), indicating ongoing ecDNA evolution. Notably, HLAMP regions classified as ecDNA were largely from GBM, SCLC, and *EGFR*-mutated NSCLC. Despite numerous HLAMP events in triple-negative breast cancers and ovarian cancers, none exhibited a single-cell CN distribution consistent with ecDNA. Instead, they showed prominent peaks typical for ICamp. We then integrated ecDNA/ICamp classification with subclonal architectures for each case (**Fig. 1H**). Many ecDNA-mediated HLAMP regions were clonally amplified, with 11 of 72 (15%) classified as subclonal (amplified in <90% of the cells). In contrast, 143 of 435 (33%) ICamp regions were subclonal, indicating substantial intratumoral heterogeneity. Together, these results establish single-cell CN distributions derived from scWGS as an effective quantitative readout to distinguish ecDNA from ICamp, reflecting their distinct mechanistic modes of inheritance during cell division, and their impact on HLAMP evolution.

Evolution of ICamps via structural and numeric modulators

Among 435 HLAMP events classified as ICamp, 270 (62%) exhibited multiple CN peaks (**Fig. 2A**), indicating intratumoral heterogeneity (exemplified by case OV-044; **Fig. 2B**). In general, distinct HLAMP CN peaks aligned with phylogenetic clades based on genome-wide CN (in multi-peak cases, specific clones were preferentially concentrated in a single peak rather than distributed across multiple peaks ($p<1\times10^{-16}$, Stouffer-combined one-sided permutation)). In aggregate, 69% of subclones were restricted to one peak, consistent with clonal expansions subsequent to subclone-specific ICamp events. Reflecting this heterogeneity, ICamp cases showed significantly greater CN variance than stable genomic regions used as controls (*TP53* and *RB1*; mean standard deviation 3.32 vs. 0.575; $p=1.97\times10^{-72}$ by one-sided two-sample *t*-test), although it was significantly less than ecDNA (mean standard deviation: ecDNA 43.1, ICamp 3.32; $p=8.68\times10^{-20}$

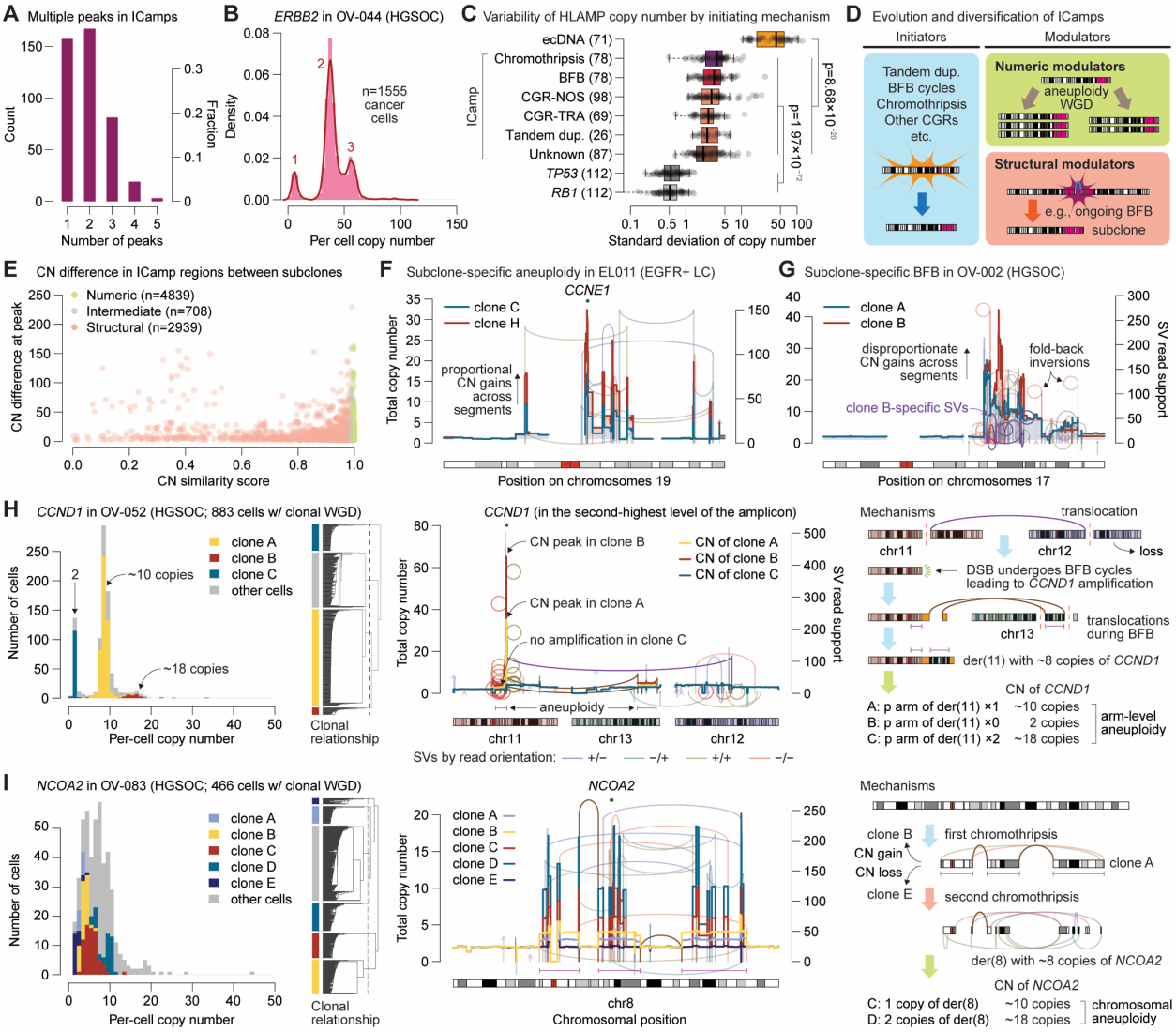

Fig. 2. Contrasting mechanisms of ICamp CN heterogeneity by numeric and structural modulators. (**A**) Prevalence of multi-peak CN distribution among ICamp cases. **(B)** An ICamp case exhibiting multi-peak CN distribution across cells, indicating subclones with distinct *ERBB2* CN. (**C**) CN variance of HLAMP events stratified by initiating mechanism (BFB, breakage-fusion bridge; CGR-NOS, complex genomic rearrangement, not otherwise specified; CGR-TRA, complex genomic rearrangement, translocation-rich). (**D**) Initiating and modulating mechanisms for ICamps. (**E**) Pairwise comparisons of HLAMP-associated arm-level CN profiles between subclones. **(F)** A classical example of aneuploidy-mediated doubling of *CCNE1* amplification in EL011. Compared to clone C, all the HLAMP segments exhibit proportional CN increase in clone H. **(G)** A structural modulation of 17q amplification in OV-002. Clone B exhibits further amplification of two segments (one amplifying *MIR21*) within the HLAMP segment in clone A. (**H**) Multiple CN peaks of *CCND1* amplification in OV-052 matching the subclonal phylogeny (left), clone-specific CNs and SVs indicative of BFB-mediated amplification and subsequent modulation by arm-level aneuploidy (middle), and schematics of the mechanism (right). (**I**) An ICamp event involving chromosome 8 in OV-083, initiated by chromothripsis, followed by a second chromothripsis and subsequent whole-chromosome aneuploidy.

202 by one-sided two-sample *t*-test; **Fig. 2C**). To investigate the source of this variability, we classified
203 ICamp events by their generative mechanisms using boundary SV features (**Fig. S5A-F**; **Methods**)
204 and asked whether variability differed by initiating mechanism[23,26]. Chromothripsis- and breakage-
205 fusion-bridge (BFB) cycle-mediated ICamps showed greater variance than other groups combined,
206 but the absolute difference was modest (mean standard deviation 3.65 vs. 3.14; one-sided two-
207 sample *t*-test, p=0.0253; **Fig. 2C**), indicating that the initiating mutational processes may have only
208 a small impact on CN variance. To identify processes modulating ICamp CN during clonal
209 evolution, we compared CN profiles between subclones (**Fig. 2D-E**). We identified pairs of
210 subclones with nearly identical chromosomal or arm-level CN profiles that differed primarily in
211 amplitude, consistent with numeric modulation (4,839 of 8,486 region-subclone pairs, 57%; **Fig.**
212 **2F**); these patterns occurred most frequently in the context of whole-genome doubling (WGD;
213 2,033/4,839, 42%; **Fig. S6A**), and the rest from chromosome- or arm-specific aneuploidies (**Fig.**
214 **S6B)**. In other cases, subclones showed dissimilar CN profiles, often accompanied by subclone-
215 specific SVs, indicative of structural modulation (2,939/8,486, 35%; **Fig. 2G**). These involved
216 diverse mechanisms, including subclone-specific fold-back inversion (consistent with additional
217 BFB cycles), chromothripsis, and other complex genomic rearrangements (**Fig. S7A**).

218 As an illustrative example, HGSOC case OV-052 harbored a *CCND1* ICamp with two large and
219 one small CN peaks, indicative of a heterogeneous ICamp. These peaks aligned with three
220 subclones in the CN-based phylogeny (**Fig. 2H, left**). In the major clone A, the *CCND1*
221 amplification, adjacent to an inter-chromosomal rearrangement translocating the distal portion of
222 11q to 12q, exhibited multiple fold-back inversions at its borders, consistent with BFB cycles
223 further modifying locus subsequent to the translocation. Minor subclone B exhibited arm-level
224 aneuploidy that duplicated the *CCND1* amplicon, reflected by two-fold CN across segments from
225 the centromere through the distal amplicon (**Fig. 2H, middle and right**). Notably, minor clone C
226 lacked evidence of the *CCND1* amplification and showed a subclone-specific loss of the allele
227 with the amplification. Analysis of haplotype-informed allele-specific CN was consistent with
228 arm-level aneuploidy-mediated loss of the *CCND1* ICamp (**Fig. S6C**). Moreover, single-cell RNA
229 sequencing from adjacent tumor tissue identified a subclone with negligible cyclin D1 expression
230 and a genome-wide CN profile matching clone C, supporting the phenotypic impact of this
231 aneuploidy-mediated HLAMP loss (**Fig. S8A-D**). Notably, this subclone exhibited a distinct gene
232 expression program, with a high interferon alpha response and upregulation of antigen-
233 presentation machinery, features proposed to promote an immunosuppressive microenvironment
234 [27]. These results suggest that such bidirectional numeric modulation of ICamp could enhance
235 transcriptional adaptability under selection. We observed similar aneuploidy-mediated elimination
236 of ICamp events in 3 additional cases, often involving G1/S cyclin genes (**Fig. S6C**).

237 In another exemplary case (OV-083), we observed a chromothripsis-mediated HLAMP involving
238 multiple segments on chromosome 8 with substantial heterogeneity (**Fig. 2I, left**). Two minor
239 subclones (A and B) harbored three large CN-gain segments at different amplitudes above baseline,
240 with SVs interconnecting their boundaries, indicating chromothripsis (**Fig. 2I, middle and right**).
241 Two additional subclones (C and D) displayed focal, more fragmented HLAMP profiles within
242 those segments, consistent with a second chromothripsis event. Their shared CN contour but
243 differing amplitudes indicate numeric modulation via aneuploidy. Branched evolution after
244 chromothripsis and subsequent modulation by aneuploidy was observed in OV-004 as well (**Fig.
245 S7B**). Together, our analyses highlight dynamic evolution of ICamp through two principal routes[28]:
246 numeric modulation by aneuploidy and WGD, and structural remodeling via BFB, chromothripsis,
247 and other mutational processes. Notably, these changes occur as both gains and losses and often
248 define sizable subclones, reflective of non-negative fitness and a linkage to tumor evolution.

249

250 **ecDNA architecture resolved by cell-to-cell CN variance and genomic rearrangement**

251 We identified 72 HLAMP regions across 11 cases (6 patients and 5 experimental models) with
252 single-cell CN distributions consistent with ecDNA. Notably, genome-wide CN profiles in these
253 11 cases were relatively homogeneous across cells, with most intercellular differences confined to
254 HLAMP regions. Given the presence of multiple HLAMP segments in each ecDNA-positive case,
255 we initially leveraged within-case single-cell CN correlations to infer whether segment pairs were
256 co-located in the same ecDNA species or segregated independently. In GBM0721, two HLAMP
257 segments were identified: one encompassing *MDM4* (1q32.1) and another *EGFR* (7p11.2) (**Fig.
258 3A, left**). They showed a modest but significant correlation, with considerable variability across
259 the CN range (**Fig. 3A, right**; Pearson correlation=0.25; $p=6.5\times10^{-8}$). In contrast, in a surgically
260 resected, heavily pre-treated *EGFR*-mutated lung cancer case EL001 where we identified 18
261 ecDNA-amplified segments, *EGFR* (7q11.2) and *CCND1* (11q13.3) CNs were tightly and linearly
262 correlated despite their different chromosomes of origin (**Fig. 3B;** Pearson correlation=0.98;
263 $p=3.4\times10^{-58}$). In GBM0721, the *MDM4* and *EGFR* amplicons likely reside on separate ecDNA
264 species, consistent with independent head-to-tail SVs encircling each segment (**Fig. 3C**). The
265 correlation between *MDM4* and *EGFR* CN may reflect co-selection of a shared phenotype,
266 supported by single-cell RNA-seq from an adjacent tumor region: cells with high *MDM4* and
267 *EGFR* expression showed high proliferative features (e.g., mitotic spindle, Ki-67), whereas cells
268 with low expression of both showed elevated reactive oxygen species signaling (**Fig. S9A-C**). By
269 contrast, the linear *EGFR*/*CCND1* correlation in EL001 suggests a chimeric ecDNA species
270 carrying both oncogenes, supported by chromothripsis connecting segments harboring *EGFR* and
271 *CCND1* (**Fig. 3D**). Similarly, *CDK4* and *MDM2* CNs were also tightly correlated (**Fig. S10A**) with

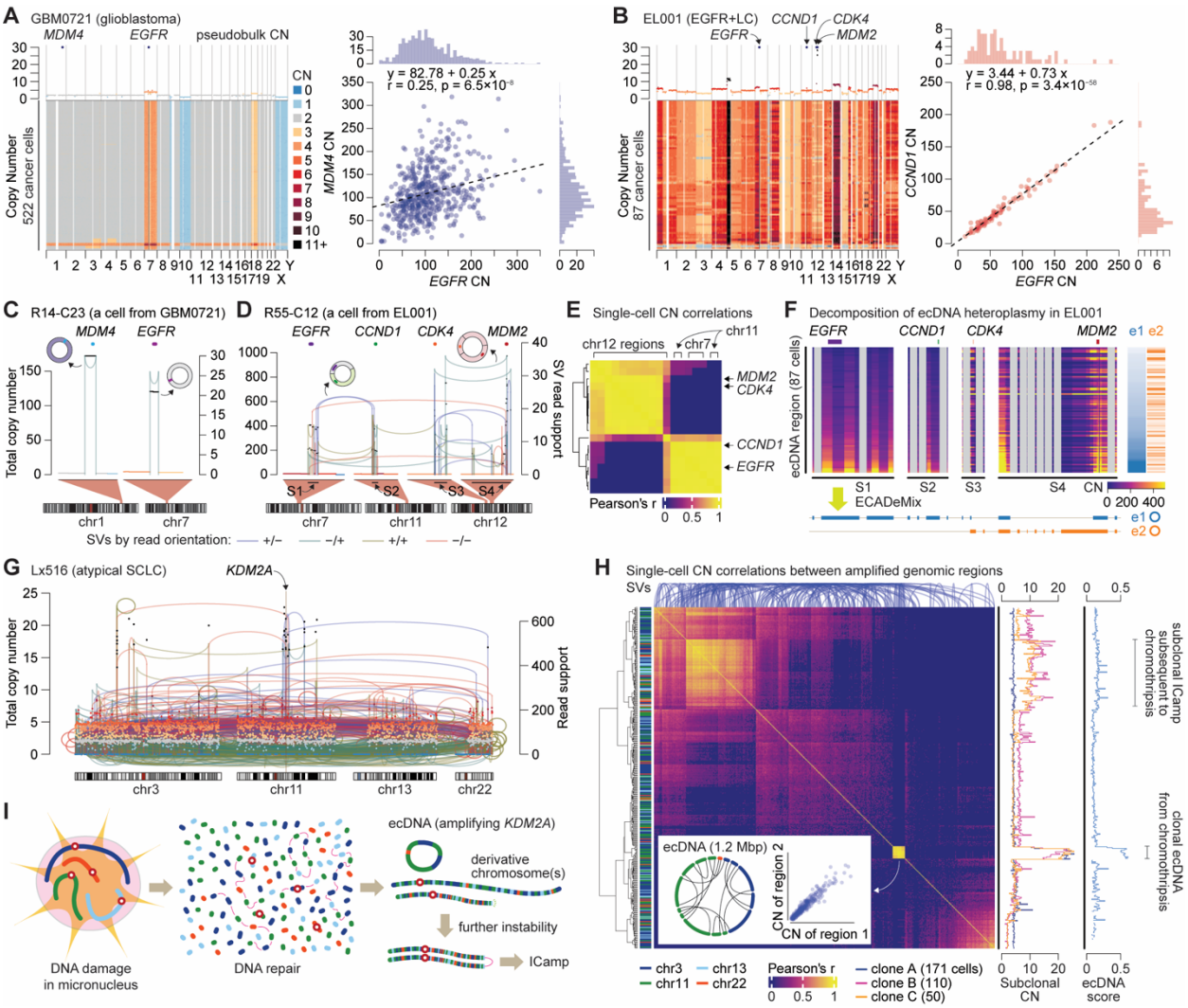

Fig. 3. Resolving mechanisms of segregation of co-amplified oncogenes in ecDNA in single cells. (**A**, **B**) Genome-wide CN profiles (left) and CN correlations between key oncogenes (right) for GBM0721 (**A**) and EL001 (**B**). (**C**, **D**) Single cell CN/SV profiles indicating two distinct simple ecDNAs in GBM0721 (**C**) and two distinct hybrid ecDNAs in EL001 (**D**). (**E**) Hierarchical clustering of HLAMP regions in EL001 based on pairwise CN correlations reveals two ecDNA species co-amplifying oncogenes. (**F**) Decomposition of ecDNA heteroplasmy in EL001 by `ECADeMix`. HLAMP segments used (S1-S4) are indicated in **D**. (**G**) Pseudobulk CN/SV profile from Lx516 showing multi-chromosomal chromothripsis leading to HLAMP of *KDM2A*. (**H**) Clustering based on pairwise CN correlation identifies a hybrid ecDNA and a subclonal ICamp event. Rearrangements between segments are illustrated as arcs on the top. Chromosomal origins in the left heatmap. Subclonal CN and ecDNA scores on the right side. Inset plots show the reconstructed ecDNA structure (left) and a representative CN correlation between two segments that comprise the ecDNA. (**I**) Schematic illustration of HLAMP mechanisms in Lx516.

SV links, implying another ecDNA species co-amplifying these oncogenes. Recognizing that the copy number of segments on the same ecDNA species will co-vary across cells, we computed

287 pairwise correlations among all HLAMP regions in EL001 and performed hierarchical clustering
288 (**Fig. 3E; Methods**). This revealed two largely mutually exclusive correlation clusters—one
289 comprising regions from chromosomes 7 and 11, and another with chromosome 12 regions—
290 consistent with two distinct ecDNA species each harboring two genes.

291 Motivated by these observations, we developed `ECADeMix` (Extra-Chromosomal Amplicon De-
292 Mixing) to systematically identify the structure of ecDNA species and their heteroplasmy across
293 cells. Briefly, analogous to heteroplasmy analyses in mitochondrial DNA in single cells[29],
294 `ECADeMix` jointly infers integer CN profiles for each ecDNA species and their multiplicities per
295 cell by decomposing single-cell CN profiles of ecDNA regions: here to 10-Kbp resolution (**Fig.**
296 **S10-S12**; **Methods**). `ECADeMix` identified two ecDNA species in each of GBM0721 and EL001,
297 confirming our previous conclusions and additionally revealing that in EL001 the *MDM2* locus
298 was also present on the *EGFR*/*CCND1* ecDNA (**Fig. 3F**), a feature that could not be resolved from
299 correlation analyses. Furthermore, `ECADeMix` identified ecDNA species co-amplifying multiple
300 oncogenes in two other cases (*MYC* and *CDX2* in COLO320-DM and *NKX2-1* and *FOXA1* in
301 EL003) as well as those in which multiple distinct clone-specific species separately amplified the
302 same oncogene (e.g., GBM0510, discussed below). Using `ECADeMix`, we identified a total of 23
303 ecDNA species across 11 cases, including 6 cases with multiple species (2-11 each). This included
304 surgically resected brain metastases of atypical SCLC characterized by an exceptional multi-
305 chromosomal chromothripsis involving four chromosomes, resulting in ecDNA-mediated
306 HLAMP (**Fig. 3G**). Given the extreme complexity of this case, we questioned how many ecDNA
307 species had formed through this process. Notably, `ECADeMix` identified only a single ecDNA
308 species conjoining 13 regions originating from 3 chromosomes (**Fig. S10D**). Accordingly,
309 pairwise CN correlations among all 391 amplified regions in this case (152,881 pairs) revealed a
310 distinct cluster of near-perfect correlations encompassing all 13 regions that formed a full circle in
311 structural reconstruction (**Fig. 3H)**. Additional regions showed strong but more modest
312 correlations (left upper part of the heatmap); their CN profiles and ecDNA scores are consistent
313 with subclonal ICamp events arising from chromothriptic derivative chromosome(s), likely
314 reflecting ongoing instability (**Fig. 3I**). Together, these analyses illustrate that cell-to-cell CN
315 variance and pairwise segment correlations enable deconvolution of complex ecDNA architectures.
316 Further, `ECADeMix` analysis of scWGS unveils new evolutionary models for ecDNA through robust
317 estimates of both presence and prevalence of ecDNA species in each cell, enabling an algorithmic
318 approach to dissect complex ecDNA-mediated cancer evolution.

319

320 **ecDNA evolution at single-nucleotide resolution**

321 We next integrated SV features and amplicon species information at the single-cell level to
322 elucidate how ecDNA drives cancer evolution. The high coverage of ecDNA regions enabled
323 single cell-level, nucleotide-level analysis of SVs (median per-cell coverage 11.09X, range
324 0.9522X−48.11X). In 5 of 11 ecDNA-positive cases (Lx33, COLO320-DM, GBM39-DM, NCI-
325 H69, NCI-H524, 2765_2), we detected a dominant ecDNA species accompanied by subpopulation
326 of cells sharing lower-frequency variant forms, consistent with branched (divergent) evolution that
327 followed initial clonal expansion by the ancestral ecDNA (**Fig. S11A-C**). For example, the NCI-
328 H69 SCLC cell line harbored a chromothripsis-derived ecDNA amplifying *MYCN* (**Fig. 4A, top**
329 **left**) with more than half of cells carrying a parental ecDNA species (e0) containing all
330 chromothriptic HLAMP segments. Using `ECADeMix`, we identified three variant ecDNA forms (e1-
331 e3) exhibiting internal CN losses in distinct regions, consistent with derivative forms of e0 (**Fig.**
332 **4A, bottom left**). These variant ecDNAs coexisted with e0 in some cells, providing further support
333 for their derivation from the parental molecule. Single-cell SV analysis identified specific
334 rearrangements that account for the variant-specific internal deletions (**Fig. 4A, right**). We
335 considered whether these deletions reflect active exclusion of deleterious genetic elements; in
336 NCI-H69 and NCI-H524, the two samples where the variant ecDNAs are most prevalent, this
337 seems unlikely as the deletions predominantly affect gene deserts or already-truncated genes (**Fig.**
338 **S12A-B**). Instead, the data are more consistent with ongoing ecDNA remodeling shaped by
339 stochastic DNA damage and repair[30], with potential purifying selection preserving segments
340 essential for oncogenic function, such as *MYCN* (**Fig. 4B**). In addition, we identified cells
341 exhibiting unique CN patterns in ecDNA regions, suggesting ongoing branched evolution (**Fig.**
342 **S12C**).

343 In contrast, GBM0510 provides evidence of convergent evolution[31] through ecDNA. In this tumor,
344 `ECADeMix` revealed that three oncogenes—*TERT* (5p15.33), *EGFR* (7p11.2), and *CDK4*
345 (12q14.1)—were each amplified on separate ecDNA species (**Fig. 4C, top left** and **Fig. S13A**).
346 *TERT* and *CDK4* ecDNAs were subclonal, present in 47% and 46% of cells, respectively, and were
347 structurally uniform within positive cells: *TERT* ecDNA arose from a local chromothripsis event
348 in the subtelomeric region 5p region, and *CDK4* ecDNA formed by a simple head-to-tail
349 rearrangement (**Fig. 4C, bottom left**). In contrast, *EGFR* amplification was clonal, but partitioned
350 into three mutually exclusive ecDNA forms (e1-e3) across cells. The largest and most prevalent
351 (e1) was formed by a simple head-to-tail rearrangement, whereas e2 and e3 were generated by
352 distinct, complex rearrangements (**Fig. 4C, right**). The amplicon intervals for e1-e3 overlap only
353 at the *EGFR* locus, and no SVs are shared among them, supporting independent origins. Single-
354 nucleotide variant (SNV) profiles (**Fig. S13**) indicate that the subclones defined by the three *EGFR*
355 ecDNA forms share most SNVs and harbor few private SNVs, consistent with relatively late
356 emergence of these ecDNA forms[32]. Taken together, these features support convergent evolution
357 through recurrent, independent acquisition of *EGFR*-amplifying ecDNAs. Compared with the

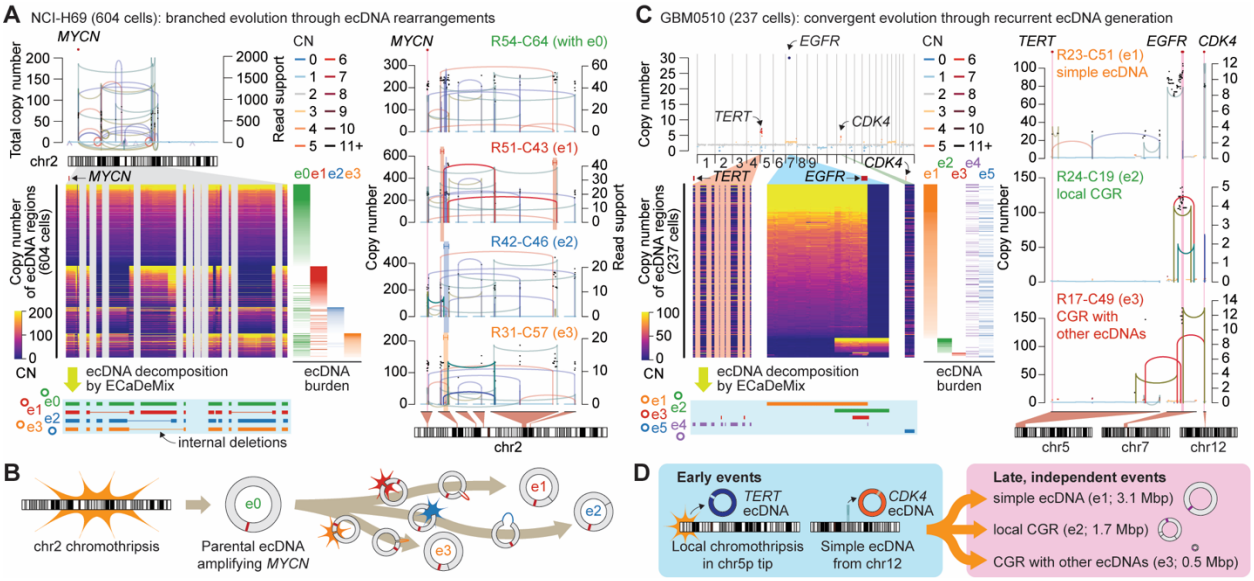

Fig. 4. Modes of ongoing ecDNA evolution. (**A**) Chromothripsis-derived ecDNA amplifying *MYCN* in the NCI-H69 SCLC cell line (top left); single-cell CN profiles of the amplicon and ecDNA heteroplasmy delineated by `ECADeMix` (bottom left). CN/SV profiles from 4 cancer cells with distinct ecDNA subspecies (right). e0 is the likely parental ecDNA; e1-e3 are daughter ecDNAs derived from e0 via internal rearrangements. (**B**) Schematic illustration of branched evolution via ecDNA rearrangement. (**C**, **D**) Pseudobulk genome-wide CN profile of glioblastoma GBM0510, exhibiting multiple ecDNA-mediated HLAMPs (**C,** top left); single-cell CN profiles of the amplicons and decomposed ecDNA species (**C,** bottom left). *EGFR* is amplified by three distinct ecDNAs (e1, e2, and e3) in a mutually exclusive manner across cells, forming distinct subclones (single-cell examples in the right panel of **C**). ecDNAs amplifying *TERT* and *CDK4* are monotypic, likely reflecting temporal order among the events (**D**).

multiple *EGFR* ecDNA species, *TERT* and *CDK4* ecDNAs were structurally uniform, consistent with earlier acquisition, whereas their subclonality may reflect stochastic loss from unequal ecDNA segregation (**Fig. 4D**). Similarly, in GBM0721 we identified three mutually exclusive ecDNA species amplifying *EGFR*, one of which predominated (**Fig. S13B**). In contrast, we found only a single ecDNA species for *MDM4*, again supporting a late acquisition of ecDNA-mediated *EGFR* amplification in GBM. Thus, these findings–resolved to both single-nucleotide and single-cell resolution–are consistent with ecDNAs in human cancers mediating both divergent and convergent evolution, resulting in dynamic cell-to-cell diversification over time.

ecDNA status by scWGS relative to bulk genome-graph predictions

We next asked how single-cell analyses of ecDNA species and their evolutionary trajectories compare to bulk genome-graph based predictions. To contrast our single cell modeling, we compared results from Amplicon Architect (AA), which has been widely used for identifying

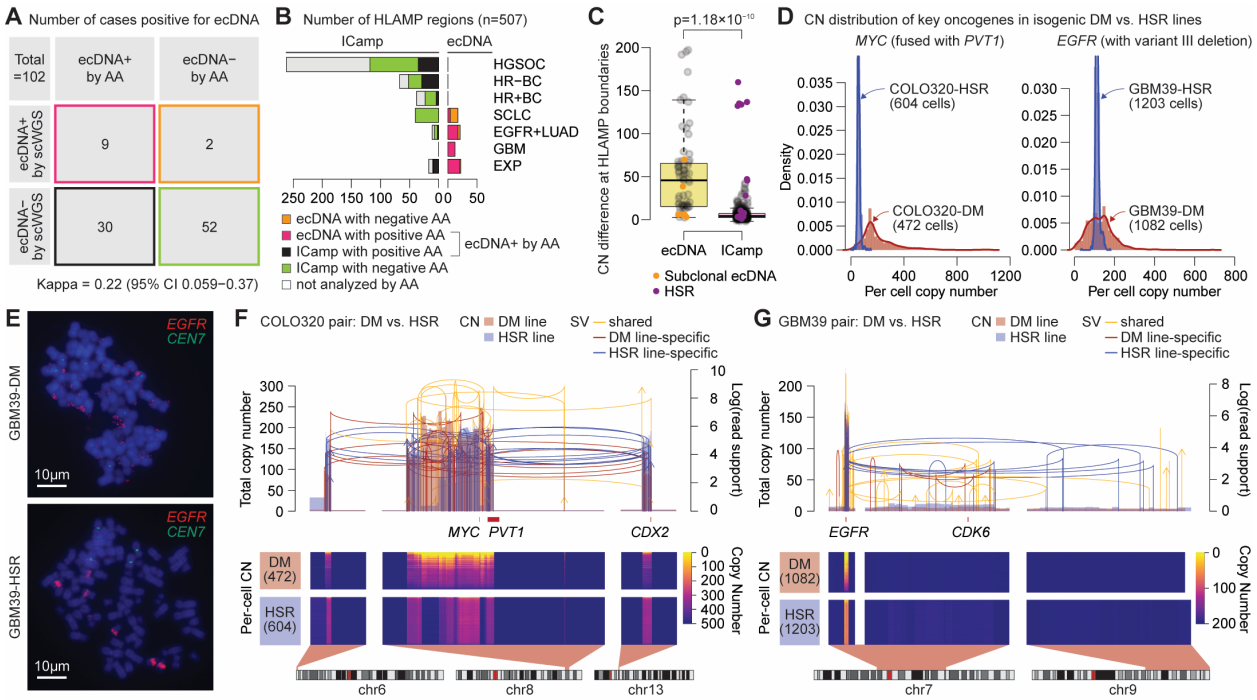

Fig. 5. scWGS distinguishes current ecDNA from historical bulk-level footprints. (**A**) Case-level comparison of ecDNA predictions from bulk genome-graph methods (Amplicon Architect/Classifier) versus scWGS. (**B**) HLAMP region-level comparison divided by cancer type. (**C**) Junctional CN difference profiles in ecDNA and ICamp regions. Statistical significance was determined by Student's *t* test. (**D**) CN distributions in isogenic paired cell lines: ecDNA-mediated double minutes (COLO320-DM and GBM39-DM) versus homogeneously staining regions (COLO320-HSR and GBM39-HSR). (**E**) Cytogenetic features of GBM39 isogenic pair exhibiting numerous variable sized, dispersed signals consistent with *EGFR*-bearing ecDNA (top) and intrachromosomal clustered amplification manifesting as homogeneously staining region (HSR). (**F**, **G**) Pseudobulk CN/SV profiles (top) and single-cell CN profiles (bottom) for DM versus HSR lines reveal genomic rearrangements that occurred after the most recent common ancestor and prior to HSR formation in COLO320 (**F**) and GBM39 (**G**) isogenic pairs.

ecDNA from bulk sequencing data[33] (**Fig. 5A, Supplementary Methods**), applied to pseudobulk libraries. HLAMP regions were matched to AA amplicons, and classifications were compared. Critically, in these pseudobulk sequencing analyses, the CN distribution over cells is not accessible. Accordingly, we focused comparison on two features–circular constructs from SVs *and* CN distributions over cells to determine how single cell information could augment bulk analyses. At the case level, 9 out of 11 ecDNA-positive cases by single-cell CN distribution were also called ecDNA-positive by AA (81% concordance). However, among 82 cases without a CN distribution over cells that was consistent with ecDNA, AA called 30 (37%) as ecDNA-positive. At the HLAMP region level, stratified by cancer type (**Fig. 5B**), discrepancies were mainly in HGSOC and hormone receptor-negative (HR−) breast cancers: these regions exhibited one or more narrow, high CN peaks in single cells, indicating ICamp (median ecDNA score=0.129,

405 range=0.040−0.380). In contrast, ecDNA-positive cases by AA in *EGFR*-mutated NSCLC, SCLC,
406 and GBM mostly showed CN distributions consistent with ecDNAs, indicating tissue-type
407 specificity of ecDNA prevalence. Review of AA genome graphs suggested several potential
408 sources of discrepancies (**Fig. S14A-C**). In ecDNA-positive cases by scWGS, the CN transition
409 between unamplified and amplified segments were typically abrupt, exhibiting high junctional
410 copy number in pseudobulk profiles (**Fig. 5C** and **Fig. S14A**)[34], with notable exceptions in cases
411 with subclonal ecDNAs. However, in many HLAMP cases in HGSOC and HR− breast cancers,
412 despite SVs arranged in cyclic paths, pseudobulk CN profiles showed stepwise CN increases
413 across extended intervals (**Fig. S14B**) consistent with progressive ICamp via multiple
414 rearrangements, rather than ecDNA[35].

415 Three HLAMP cases exhibited circular paths in AA genome graphs and abrupt CN transitions in
416 pseudobulk profiles, yet their single-cell CN distributions were consistent with ICamp (**Fig. 5C**).
417 Two of these were isogenic cell line models exhibiting homogeneously staining regions (HSRs) in
418 DNA FISH (COLO320-HSR and GBM39-HSR; **Fig. 5D-E**)[36,37]. Similar to their ecDNA-positive
419 counterparts (COLO320-DM and GBM39-DM, respectively), these HSR cell lines retain the SV
420 breakpoint footprint of the initiating ecDNA (**Fig. 5F** and **5G**), indicating that the HSR was formed
421 by chromosomal integration of ecDNAs[38]. The third case was NCI-H82, a SCLC cell line, which
422 was labeled as ecDNA-positive in a previous study using AA[39]. The bulk genome graph showed a
423 cyclic path amplifying *MYC* with abrupt CN transitions, but the single-cell CN distribution was
424 sharply peaked, consistent with ICamp (**Fig. S14C**). DNA FISH confirmed *MYC* HSR formation
425 in this line without evidence of ecDNA (**Fig. S4B**).

426 Last, we analyzed genomic rearrangements associated with chromosomal integration of ecDNA
427 in isogenic pairs. In COLO320-HSR, the core amplicon contained several internal deletions
428 indicated by large, abrupt CN transitions, consistent with branched evolution from the parental
429 ecDNA prior to integration (**Fig. 5F**). In GBM39-HSR, HSR line-specific SVs included a
430 chromothripsis event linking the *EGFR* amplicon to the subtelomeric 9q region (**Fig. 5G**). These
431 HSR line-specific SVs inform mutational processes between the shared ecDNA ancestor and HSR
432 formation but do not, by themselves, identify the integrated chromosomes. Orthogonal mapping
433 from previous studies places the HSR on chrX for COLO320-HSR[40]. Together, these analyses
434 show that single-cell CN distributions can distinguish current ecDNAs from historical ecDNA
435 footprints integrated into chromosomes, refine bulk genome-graph interpretations, and place
436 HLAMP events within their evolutionary context.

437

438 **Discussion**

439 Leveraging scWGS, we quantified cell-to-cell variability in HLAMP across four major cancer
440 types and well studied experimental models, and linked this variability to specific generative and
441 modulatory mechanisms. In contrast to bulk genome-graph predictions, single-cell CN
442 distributions of HLAMP events distinguish episomal ecDNA from ICamp, including HSRs.
443 Although cyclic genome graphs in bulk WGS can nominate candidate ecDNAs, many such
444 candidates—particularly in HGSOC and HR− breast cancers—show narrow, sharply peaked CN
445 distributions consistent with ICamp. In these tumor types, HLAMP heterogeneity appears largely
446 shaped by ICamp evolution through numeric modulation and structural remodeling, which
447 nonetheless can generate substantial intratumoral heterogeneity. Notably, the subclones lacking
448 oncogene amplification due to aneuploidy-mediated loss warrant further clinical investigation, as
449 selection of such subclones could lead to early resistance to oncogene-directed targeted therapies[41].

450 In contrast to enrichment-based single-cell methods that selectively profile ecDNA[42–44], scWGS
451 enables genome-wide characterization of HLAMP evolution in single cancer cells, including both
452 ecDNAs and ICamps. Compared with other genome-wide single-cell profiling methods such as
453 single-cell ATAC[45] and RNA sequencing[46], which are being explored for ecDNA detection,
454 DLP+-based scWGS or related technologies[47,48] provide unbiased CN profiles and SV information
455 with high CN fidelity and nucleotide-level coverage at HLAMP loci in single cells, both of which
456 are crucial for defining the mechanisms and evolution of HLAMP events. Heuristically, in the
457 analyses presented here, we required at least 30 high-quality single-cell CN profiles for ecDNA
458 detection. However, we suggest further algorithmic improvements and context-specific events
459 with ultra-high copies could be resolved to fewer than 30 cells.

460 Our results suggest that ecDNA is prevalent in select cancer types, in contrast to recent studies
461 based on bulk genome-graph predictions[49,50]. In cancer types where ecDNA has been extensively
462 described, namely GBM and SCLC, our scWGS-based analysis confirms a high prevalence of
463 ecDNA. We also identified 3 mutually exclusive ecDNA species in 2 of 5 *EGFR*-mutated NSCLCs,
464 amplifying key oncogenes including *EGFR*, *MDM2*, and *NKX2-1*. The contribution of ecDNAs to
465 the biology of *EGFR*-mutated lung cancer and to responses to targeted therapies will be of clinical
466 interest. By contrast, we did not observe any HLAMP event with a CN distribution consistent with
467 ecDNA among 72 patients with HGSOC or HR− breast cancers. Although cyclic paths were
468 identified around HLAMP regions in many cases, their single-cell CN distributions and SV
469 features were instead consistent with ICamps arising from multi-step rearrangement processes.
470 This striking paucity of ecDNA suggests that some tissues could be less permissive of ecDNAs,
471 contributing to tissue-specific differences in ecDNA prevalence across human cancers. In addition,
472 previous studies indicated that anticancer therapies, such as chemotherapy or radiotherapy, can
473 induce ecDNA formation[51,52], potentially reshaping ecDNA prevalence in patients with advanced
474 cancer receiving therapies. In line with this, a clinical study in patients with platinum-resistant

475 HGSOC reported double minutes in cancer cells from patients' ascites[53], however the tissue
476 specificity of therapy initiated ecDNA needs further investigation.

477 Motivated by longstanding efforts to develop ecDNA-directed therapeutics[13–15,54,55] and growing
478 recognition of the profound impact of ecDNA on cancer evolution, there is increasing interest in
479 detecting ecDNAs in patient biospecimens. Conventional cytogenetic techniques like FISH require
480 metaphase cells to accurately detect ecDNAs, but these are not often available for clinical tumor
481 samples. Instead, high-throughput sequencing of clinical tumors has become an integral part of
482 oncology care, and the use of such assays to detect ecDNA has been proposed[8]. Our findings
483 highlight intrinsic challenges to this approach. A key genomic hallmark of prevalent ecDNA is
484 pronounced cell-to-cell CN variability, which cannot be resolved from single-sample bulk
485 sequencing. Genome graph-based approaches instead depend on sensitive detection of SVs from
486 WGS and cannot reliably distinguish extant ecDNA from its vestigial footprints after chromosomal
487 integration. Given the dynamic nature of HLAMP evolution, including the widespread modulation
488 of ICamp events observed in our study, scalable scWGS, or targeted single cell approaches
489 whereby cellular distributions can be read out on specific loci, may offer a more promising path
490 for clinical ecDNA detection and monitoring in the future.

491 In conclusion, our analysis of >86,000 single cell cancer genomes defines interpretable CN
492 distributions of HLAMP events consistent with distinct ICamp and ecDNA generative processes
493 and elucidates how these mechanisms diversify cancer cell populations. These dynamics extend
494 beyond ecDNA and are pervasive in tumors with ICamp, underscoring ongoing genome
495 remodelling. Single-cell genomic profiling of longitudinal clinical specimens, integrated with
496 orthogonal readouts of gene expression[48], would enable direct measurement of how therapy
497 reshapes HLAMP evolution. The derivative single-cell readouts of architecture, abundance and
498 clonal selection in cancer cells will therefore help to identify actionable dependencies and inform
499 therapeutic strategies through an evolutionary lens.

500 **Bibliography:**

501 1. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of Chemotherapy plus a Monoclonal
502 Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N Engl J*
503 *Med*. 2001;344(11):783-792.

504 2. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with
505 chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric
506 or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled
507 trial. *Lancet*. 2010;376(9742):687-697.

508 3. Yagisawa M, Taniguchi H, Satoh T, et al. Trastuzumab deruxtecan in advanced solid tumors
509 with human epidermal growth factor receptor 2 amplification identified by plasma cell-free
510 DNA testing: A multicenter, single-arm, phase II basket trial. *J Clin Oncol*.
511 2024;42(32):3817-3825.

512 4. Regitnig P, Schippinger W, Lindbauer M, Samonigg H, Lax SF. Change of HER-2/neu
513 status in a subset of distant metastases from breast carcinomas. *J Pathol*. 2004;203(4):918-
514 926.

515 5. Seol H, Lee HJ, Choi Y, et al. Intratumoral heterogeneity of HER2 gene amplification in
516 breast cancer: its clinicopathological significance. *Mod Pathol*. 2012;25(7):938-948.

517 6. Jordan NV, Bardia A, Wittner BS, et al. HER2 expression identifies dynamic functional
518 states within circulating breast cancer cells. *Nature*. 2016;537(7618):102-106.

519 7. Filho OM, Viale G, Stein S, et al. Impact of HER2 heterogeneity on treatment response of
520 early-stage HER2-positive breast cancer: Phase II neoadjuvant clinical trial of T-DM1
521 combined with pertuzumab. *Cancer Discov*. 2021;11(10):2474-2487.

522 8. Weiser NE, Watkins TBK, Chang HY, Mischel PS. A guide to extrachromosomal DNA:
523 Cancer's dynamic circular genome. *Cancer Discov*. 2025;15(6):1105-1114.

524 9. Turner KM, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification
525 drives tumour evolution and genetic heterogeneity. *Nature*. 2017;543(7643):122-125.

526 10. Hung KL, Yost KE, Xie L, et al. ecDNA hubs drive cooperative intermolecular oncogene
527 expression. *Nature*. 2021;600(7890):731-736.

528 11. Snuderl M, Fazlollahi L, Le LP, et al. Mosaic amplification of multiple receptor tyrosine
529 kinase genes in glioblastoma. *Cancer Cell*. 2011;20(6):810-817.

530 12. Lange JT, Rose JC, Chen CY, et al. The evolutionary dynamics of extrachromosomal DNA
531 in human cancers. *Nat Genet*. 2022;54(10):1527-1533.

532 13. Clarke TL, Tang R, Chakraborty D, et al. Histone lysine methylation dynamics control
533 *EGFR* DNA copy-number amplification. *Cancer Discov*. 2020;10(2):306-325.

534 14. Tang J, Weiser NE, Wang G, et al. Enhancing transcription-replication conflict targets

535 ecDNA-positive cancers. *Nature*. 2024;635(8037):210-218.

536 15. Dharanipragada P, Zhang X, Liu S, et al. Blocking Genomic Instability Prevents Acquired
537 Resistance to MAPK Inhibitor Therapy in Melanoma. *Cancer Discov*. 2023;13(4):880-909.

538 16. Laks E, McPherson A, Zahn H, et al. Clonal decomposition and DNA replication states
539 defined by scaled single-cell genome sequencing. *Cell*. 2019;179(5):1207-1221.e22.

540 17. Funnell T, O'Flanagan CH, Williams MJ, et al. Single-cell genomic variation induced by
541 mutational processes in cancer. *Nature*. 2022;612(7938):106-115.

542 18. Haigis KM, Cichowski K, Elledge SJ. Tissue-specificity in cancer: The rule, not the
543 exception. *Science*. 2019;363(6432):1150-1151.

544 19. Salehi S, Kabeer F, Ceglia N, et al. Clonal fitness inferred from time-series modelling of
545 single-cell cancer genomes. *Nature*. 2021;595(7868):585-590.

546 20. McPherson A, Vázquez-García I, Myers MA, et al. Ongoing genome doubling shapes
547 evolvability and immunity in ovarian cancer. *Nature*. 2025;644(8078):1078-1087.

548 21. Williams MJ, Vázquez-García I, Tam G, et al. Tracking clonal evolution during treatment in
549 ovarian cancer using cell-free DNA. *Nature*. 2025;647(8090):757-765.

550 22. Pradella D, Zhang M, Gao R, et al. Engineered extrachromosomal oncogene amplifications
551 promote tumorigenesis. *Nature*. Published online December 18, 2024:1-10.

552 23. Lee JJK, Jung YL, Cheong TC, et al. ERα-associated translocations underlie oncogene
553 amplifications in breast cancer. *Nature*. 2023;618(7967):1024-1032.

554 24. Doob JL. Topics in the theory of Markoff chains. *Trans Amer Math Soc*. 1942;52(1):37-64.

555 25. Gillespie DT. A general method for numerically simulating the stochastic time evolution of
556 coupled chemical reactions. *J Comput Phys*. 1976;22(4):403-434.

557 26. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer
558 genomes. *Nature*. 2020;578(7793):112-121.

559 27. Luo Y, Xia Y, Liu D, et al. Neoadjuvant PARPi or chemotherapy in ovarian cancer informs
560 targeting effector Treg cells for homologous-recombination-deficient tumors. *Cell*.
561 2024;187(18):4905-4925.e24.

562 28. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*.
563 1998;396(6712):643-649.

564 29. Kim M, Gorelick AN, Vàzquez-García I, et al. Single-cell mtDNA dynamics in tumors is
565 driven by coregulation of nuclear and mitochondrial genomes. *Nat Genet*. 2024;56(5):889-
566 899.

567 30. Ilić M, Zaalberg IC, Raaijmakers JA, Medema RH. Life of double minutes: generation,
568 maintenance, and elimination. *Chromosoma*. 2022;131(3):107-125.

569 31. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution
570 revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883-892.

571 32. Ozawa T, Riester M, Cheng YK, et al. Most human non-GCIMP glioblastoma subtypes
572 evolve from a common proneural-like precursor glioma. *Cancer Cell*. 2014;26(2):288-300.

573 33. Deshpande V, Luebeck J, Nguyen NPD, et al. Exploring the landscape of focal
574 amplifications in cancer using AmpliconArchitect. *Nat Commun*. 2019;10(1):392.

575 34. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a
576 single catastrophic event during cancer development. *Cell*. 2011;144(1):27-40.

577 35. Notta F, Chan-Seng-Yue M, Lemire M, et al. A renewed model of pancreatic cancer
578 evolution based on genomic rearrangement patterns. *Nature*. 2016;538(7625):378-382.

579 36. Alitalo K, Schwab M, Lin CC, Varmus HE, Bishop JM. Homogeneously staining
580 chromosomal regions contain amplified copies of an abundantly expressed cellular
581 oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. *Proc*
582 *Natl Acad Sci U S A*. 1983;80(6):1707-1711.

583 37. Nathanson DA, Gini B, Mottahedeh J, et al. Targeted therapy resistance mediated by
584 dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*. 2014;343(6166):72-
585 76.

586 38. Rosswog C, Bartenhagen C, Welte A, et al. Chromothripsis followed by circular
587 recombination drives oncogene amplification in human cancer. *Nat Genet*. Published online
588 November 15, 2021:1-13.

589 39. Zhang J, Jin Y, Lin H, et al. Amplification of extrachromosomal *MYC* paralogs shapes
590 immunosuppressive tumor microenvironment in small cell lung cancer. *Clin Cancer Res*.
591 Published online May 13, 2025:OF1-OF14.

592 40. Nichols A, Choi Y, Norman RX, et al. Chromosomal tethering and mitotic transcription
593 promote ecDNA nuclear inheritance. *Mol Cell*. 2025;85(15):2839-2853.e8.

594 41. Chen W, Gupta A, Mai N, et al. Trastuzumab Deruxtecan resistance via loss of HER2
595 expression and binding. *Cancer Discov*. 2026;16(2):235-249.

596 42. Hung KL, Luebeck J, Dehkordi SR, et al. Targeted profiling of human extrachromosomal
597 DNA by CRISPR-CATCH. *Nat Genet*. 2022;54(11):1746-1754.

598 43. Chapman OS, Luebeck J, Sridhar S, et al. Circular extrachromosomal DNA promotes tumor
599 heterogeneity in high-risk medulloblastoma. *Nat Genet*. Published online November 9,
600 2023. doi:10.1038/s41588-023-01551-3

601 44. Chen JP, Diekmann C, Wu H, et al. scCircle-seq unveils the diversity and complexity of
602 extrachromosomal circular DNAs in single cells. *Nat Commun*. 2024;15(1):1768.

603 45. Jiang R, Lu Z, Li F, et al. scCirclehunter delineates ecDNA-containing cells using single-

604 cell ATAC-seq, with a focus on glioblastoma. *Cell Discov*. 2025;11(1):98.

605 46. Hendriksen JD, Locallo A, Schlotmann BC, et al. Resolving clonal evolution and selection
606 of extrachromosomal DNA at single-cell resolution. *Genome Biol*. 2026;27(1):10.

607 47. Wang K, Kumar T, Wang J, et al. Archival single-cell genomics reveals persistent subclones
608 during DCIS progression. *Cell*. 2023;186(18):3968-3982.e15.

609 48. Wang K, Ye R, Bai S, et al. Coalescing single-cell genomes and transcriptomes to decode
610 breast cancer progression. *Cell*. 2025;188(22):6355-6369.e16.

611 49. Kim H, Nguyen NP, Turner K, et al. Extrachromosomal DNA is associated with oncogene
612 amplification and poor outcome across multiple cancers. *Nat Genet*. 2020;52(9):891-897.

613 50. Bailey C, Pich O, Thol K, et al. Origins and impact of extrachromosomal DNA. *Nature*.
614 2024;635(8037):193-200.

615 51. Coquelle A, Toledo F, Stern S, Bieth A, Debatisse M. A new role for hypoxia in tumor
616 progression. *Mol Cell*. 1998;2(2):259-265.

617 52. Shoshani O, Brunner SF, Yaeger R, et al. Chromothripsis drives the evolution of gene
618 amplification in cancer. *Nature*. 2021;591(7848):137-141.

619 53. Raymond E, Faivre S, Weiss G, et al. Effects of hydroxyurea on extrachromosomal DNA in
620 patients with advanced ovarian carcinomas. *Clin Cancer Res*. 2001;7(5):1171-1180.

621 54. Von Hoff DD, McGill JR, Forseth BJ, et al. Elimination of extrachromosomally amplified
622 MYC genes from human tumor cells reduces their tumorigenicity. *Proc Natl Acad Sci U S*
623 *A*. 1992;89(17):8165-8169.

624 55. Shimizu N, Itoh N, Utiyama H, Wahl GM. Selective entrapment of extrachromosomally
625 amplified DNA by nuclear budding and micronucleation during S phase. *J Cell Biol*.
626 1998;140(6):1307-1320.

627 56. J. S. Medina-Martínez, J. E. Arango-Ossa, M. F. Levine, Y. Zhou, G. Gundem, A. L. Kung,
628 E. Papaemmanuil, Isabl Platform, a digital biobank for processing multimodal patient data.
629 *BMC Bioinformatics* **21**, 549 (2020).

630 57. G. Ha, A. Roth, D. Lai, A. Bashashati, J. Ding, R. Goya, R. Giuliany, J. Rosner, A. Oloumi,
631 K. Shumansky, S.-F. Chin, G. Turashvili, M. Hirst, C. Caldas, M. A. Marra, S. Aparicio, S.
632 P. Shah, Integrative analysis of genome-wide loss of heterozygosity and monoallelic
633 expression at nucleotide resolution reveals disrupted pathways in triple-negative breast
634 cancer. *Genome Res.* **22**, 1995–2007 (2012).

635 58. D. L. Cameron, J. Baber, C. Shale, J. E. Valle-Inclan, N. Besselink, A. van Hoeck, R.
636 Janssen, E. Cuppen, P. Priestley, A. T. Papenfuss, GRIDSS2: comprehensive
637 characterisation of somatic structural variation using single breakend variants and structural
638 variant phasing. *Genome Biol.* **22**, 202 (2021).

59. C. Chiang, R. M. Layer, G. G. Faust, M. R. Lindberg, D. B. Rose, E. P. Garrison, G. T. Marth, A. R. Quinlan, I. M. Hall, SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).

60. D. W. Garsed, O. J. Marshall, V. D. A. Corbin, A. Hsu, L. Di Stefano, J. Schröder, J. Li, Z.-P. Feng, B. W. Kim, M. Kowarsky, Ben Lansdell, R. Brookwell, O. Myklebost, L. Meza-Zepeda, A. J. Holloway, F. Pedeutour, K. H. A. Choo, M. A. Damore, A. J. Deans, A. T. Papenfuss, D. M. Thomas, The Architecture and Evolution of Cancer Neochromosomes. *Cancer Cell* **26**, 653–667 (2014).

61. E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, N. D. Goodman, Pyro: Deep universal probabilistic programming. *Journal of machine learning research* **20**, 1–6 (2019).

62. M. Hoffman, D. M. Blei, C. Wang, J. Paisley, Stochastic Variational Inference, *arXiv [stat.ML]* (2012)pp. 1303–1347.

63. J. Luebeck, E. Huang, F. Kim, T. Liefeld, B. Dameracharla, R. Ahuja, D. Schreyer, G. Prasad, M. Adamaszek, R. Kenkre, T. Agashe, D. Torvi, T. Tabor, M. Giurgiu, S. Kim, H. Kim, P. Bailey, R. G. W. Verhaak, V. Deshpande, M. Reich, P. S. Mischel, J. Mesirov, V. Bafna, AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer genomes, *Bioinformatics* (2024). https://www.biorxiv.org/content/10.1101/2024.05.06.592768v1.

64. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

65. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv [stat.ML]* (2018). http://arxiv.org/abs/1802.03426.

66. I. Vázquez-García, F. Uhlitz, N. Ceglia, J. L. P. Lim, M. Wu, N. Mohibullah, J. Niyazov, A. E. B. Ruiz, K. M. Boehm, V. Bojilova, C. J. Fong, T. Funnell, D. Grewal, E. Havasov, S. Leung, A. Pasha, D. M. Patel, M. Pourmaleki, N. Rusk, H. Shi, R. Vanguri, M. J. Williams, A. W. Zhang, V. Broach, D. S. Chi, A. Da Cruz Paula, G. J. Gardner, S. H. Kim, M. Lennon, K. Long Roche, Y. Sonoda, O. Zivanovic, R. Kundra, A. Viale, F. N. Derakhshan, L. Geneslaw, S. Issa Bhaloo, A. Maroldi, R. Nunez, F. Pareja, A. Stylianou, M. Vahdatinia, Y. Bykov, R. N. Grisham, Y. L. Liu, Y. Lakhman, I. Nikolovski, D. Kelly, J. Gao, A. Schietinger, T. J. Hollmann, S. F. Bakhoum, R. A. Soslow, L. H. Ellenson, N. R. Abu-Rustum, C. Aghajanian, C. F. Friedman, A. McPherson, B. Weigelt, D. Zamarin, S. P. Shah, Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* **612**, 778–786 (2022).

67. A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).

68. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

Data Availability: Source code for the computational models and scripts used for analysis in this manuscript can be found at https://github.com/shahcompbio/hlamp. Newly generated raw scWGS data will be available from the NCBI dbGaP prior to publication. Previously published raw scWGS datasets are available through European Genome-Phenome Archive (references 17 and 19: EGAS00001006343 and EGAS00001003190) and dbGaP under the National Institute of Health (references 20 and 21: phs002857.v3.p1).

Acknowledgments: We thank Gryte Satas, Ignacio Vazquez-Garcia, Samuel Freeman, Matthew Zatzman, Alessandro Grande, Giovanni Santacatterina, and the rest of the Shah, McPherson, and Lee labs for discussions and technical advice. We thank Dr. Paul Mischel (Stanford University) for generously providing the GBM39 cell lines used in this study.

Funding: This work was supported by the Halvorsen Center for Computational Oncology and Cycle for Survival supporting Memorial Sloan Kettering Cancer Center. S.P.S. holds the Nicholls Biondi Chair in Computational Oncology and is a Susan G. Komen Scholar. Additional support from the Department of Defense Congressionally Directed Medical Research Program award [W81XWH-20-1-0565], an Ovarian Cancer Research Alliance (OCRA) Collaborative Research Development Grant [648007], an NIH R01 CA281928-01, and by the Seidenberg Family Foundation funds to SPS. Additional support was provided through the MSK Core Grant (P30 CA008748), the National Cancer Institute of the National Institute of Health (K08CA301011 to JJL; K99CA277562 to SS; R01CA282913 to AV), the Burroughs Wellcome Fund Career Award for Medical Scientists (to JJL), the Cancer Grand Challenge Initiative (to AV), the American Cancer Society Discovery Boost Grant (to AV), the Mark Foundation for Cancer Research ASPIRE Grant (to AV).

Author Contributions: JJL and SPS conceived the project and supervised study design. JJL and SS designed experiments and performed primary computational analyses, with conceptual input from MJW and AM. SS led development of the amplification classifier `eicicle`. MAM led development of the ecDNA deconvolution algorithm `ECADeMix`. SS performed single-cell RNA sequencing analyses. MJW conducted allele-specific copy-number analyses. MY, DHA, EGS, and AV performed cytogenetic experiments. MY and JL conducted cell culture experiments. PM, ER, AQV, and CMR generated data from patient-derived xenograft models. PR, SA, N. Rekhtman, VT, MMA, HAY, KKHY, and CMR generated data from clinical tumor samples. JJL, SS, MAM,

MJW, KT, CT, CZJP, SM, and AM performed secondary computational analyses. SC, EH, and APS established and maintained computational analysis pipelines. MW and NM supervised single-cell whole-genome sequencing procedures. JJL, SS, MAM, N. Rusk, and SPS wrote the original draft, and all authors revised and approved the final manuscript.

Competing Interests: AQV is a current employee of AstraZeneca. PR reports research funding from Grail, Novartis, AstraZeneca, EpicSciences, Invitae/ArcherDx, Biotheranostics, Tempus, NeoGenomics, Biovica, Guardant, Personalis, Myriad and consulting or advisory role for Novartis, AstraZeneca, Pfizer, Lilly/Loxo, Prelude Therapeutics, Epic Sciences, Daiichi-Sankyo, Foundation Medicine, Inivata, Natera, Tempus, SAGA Diagnostics, Paige.ai, Guardant, and Myriad, outside the scope of this work. SA is cofounder and shareholder of Genome Therapeutics and reports advisory roles for Chordia Therapeutics and Sangamo Therapeutics, outside the scope of this work. MMA reports research funding from Amgen, AstraZeneca, Bristol Myers Squibb, Genentech and Lilly, consulting or advisory roles for Affini-T, AstraZeneca, Blueprint Medicines Corporation, D3Bio, EMD Serono, Gritstone, Iovance, Merck, Merus, Mirati, Novartis, Pfizer and Synthekine, and serves on the Data Safety Monitoring Boards for Apollomics and Bristol Myers Squibb, outside the scope of this work. HAY reports consulting or advisory role for Takeda, Taiho, Black Diamond, BMS, AbbVie, Amgen, AstraZeneca, Daiichi Sankyo, Ipsen, and Pfizer and serves on the Data and Safety Monitoring Board for Janssen and Mythic Therapeutics, outside the scope of this work. CMR reports consulting or advisory role for Genentech/Roche, AstraZeneca, Amgen, Jazz Pharmaceuticals, Earli, AbbVie, Daiichi Sankyo/UCB Japan, Merck, Auron Therapeutics, DISCO and research funding from Merck, Genentech/Roche, Daiichi Sankyo, outside the scope of this work. SPS reports research funding from AstraZeneca and Bristol Myers Squibb, outside the scope of this work.

736 **Figure Legends**

737 **Fig. 1. Cell-to-cell copy-number variance in scWGS differentiates ecDNA-mediated**
738 **amplifications from intrachromosomal amplifications.** (**A**) Cancer types and case counts (top)
739 and schematic illustration of study design (bottom). Organ and model illustrations are from
740 BioRender. (**B**) Number of high-quality cancer cells analyzed (x-axis, log-scale) by tumor type (y-
741 axis). (**C**) Frequently amplified oncogenes and their case counts. (**D**) Inheritance patterns for
742 intrachromosomal amplification (ICamp) versus extrachromosomal circular DNA (ecDNA; **left**)
743 and expected CN distributions across cells (right). (**E**) Representative examples of ICamp and
744 ecDNA, from a patient with *EGFR*-mutated lung adenocarcinoma (left) and glioblastoma (right).
745 (**F**) Conceptual illustration of two metrics used to classify HLAMP events. (**G**) Classification of
746 507 HLAMP regions from 86,239 cancer cells based on single-cell CN distribution. (**H**) Fraction
747 of clonal and subclonal HLAMP regions by mechanism, across cases.

748

749 **Fig. 2. Contrasting mechanisms of ICamp CN heterogeneity by numeric and structural**
750 **modulators.** (**A**) Prevalence of multi-peak CN distribution among ICamp cases. **(B)** An ICamp
751 case exhibiting multi-peak CN distribution across cells, indicating subclones with distinct *ERBB2*
752 CN. (**C**) CN variance of HLAMP events stratified by initiating mechanism (BFB, breakage-fusion
753 bridge; CGR-NOS, complex genomic rearrangement, not otherwise specified; CGR-TRA,
754 complex genomic rearrangement, translocation-rich). (**D**) Initiating and modulating mechanisms
755 for ICamps. (**E**) Pairwise comparisons of HLAMP-associated arm-level CN profiles between
756 subclones. **(F)** A classical example of aneuploidy-mediated doubling of *CCNE1* amplification in
757 EL011. Compared to clone C, all the HLAMP segments exhibit proportional CN increase in clone
758 H. **(G)** A structural modulation of 17q amplification in OV-002. Clone B exhibits further
759 amplification of two segments (one amplifying *MIR21*) within the HLAMP segment in clone A.
760 (**H**) Multiple CN peaks of *CCND1* amplification in OV-052 matching the subclonal phylogeny
761 (left), clone-specific CNs and SVs indicative of BFB-mediated amplification and subsequent
762 modulation by arm-level aneuploidy (middle), and schematics of the mechanism (right). (**I**) An
763 ICamp event involving chromosome 8 in OV-083, initiated by chromothripsis, followed by a
764 second chromothripsis and subsequent whole-chromosome aneuploidy.

765

766 **Fig. 3. Resolving mechanisms of segregation of co-amplified oncogenes in ecDNA in single**
767 **cells.** (**A**, **B**) Genome-wide CN profiles (left) and CN correlations between key oncogenes (right)
768 for GBM0721 (**A**) and EL001 (**B**). (**C**, **D**) Single cell CN/SV profiles indicating two distinct simple
769 ecDNAs in GBM0721 (**C**) and two distinct hybrid ecDNAs in EL001 (**D**). (**E**) Hierarchical
770 clustering of HLAMP regions in EL001 based on pairwise CN correlations reveals two ecDNA

771 species co-amplifying oncogenes. (**F**) Decomposition of ecDNA heteroplasmy in EL001 by
772 `ECADeMix`. HLAMP segments used (S1-S4) are indicated in **D**. (**G**) Pseudobulk CN/SV profile
773 from Lx516 showing multi-chromosomal chromothripsis leading to HLAMP of *KDM2A*. (**H**)
774 Clustering based on pairwise CN correlation identifies a hybrid ecDNA and a subclonal ICamp
775 event. Rearrangements between segments are illustrated as arcs on the top. Chromosomal origins
776 in the left heatmap. Subclonal CN and ecDNA scores on the right side. Inset plots show the
777 reconstructed ecDNA structure (left) and a representative CN correlation between two segments
778 that comprise the ecDNA. (**I**) Schematic illustration of HLAMP mechanisms in Lx516.

779

780 **Fig. 4. Modes of ongoing ecDNA evolution.** (**A**) Chromothripsis-derived ecDNA amplifying
781 *MYCN* in the NCI-H69 SCLC cell line (top left); single-cell CN profiles of the amplicon and
782 ecDNA heteroplasmy delineated by `ECADeMix` (bottom left). CN/SV profiles from 4 cancer cells
783 with distinct ecDNA subspecies (right). e0 is the likely parental ecDNA; e1-e3 are daughter
784 ecDNAs derived from e0 via internal rearrangements. (**B**) Schematic illustration of branched
785 evolution via ecDNA rearrangement. (**C**, **D**) Pseudobulk genome-wide CN profile of glioblastoma
786 GBM0510, exhibiting multiple ecDNA-mediated HLAMPs (**C,** top left); single-cell CN profiles
787 of the amplicons and decomposed ecDNA species (**C,** bottom left). *EGFR* is amplified by three
788 distinct ecDNAs (e1, e2, and e3) in a mutually exclusive manner across cells, forming distinct
789 subclones (single-cell examples in the right panel of **C**). ecDNAs amplifying *TERT* and *CDK4* are
790 monotypic, likely reflecting temporal order among the events (**D**).

791

792 **Fig. 5. scWGS distinguishes current ecDNA from historical bulk-level footprints.** (**A**) Case-
793 level comparison of ecDNA predictions from bulk genome-graph methods (Amplicon
794 Architect/Classifier) versus scWGS. (**B**) HLAMP region-level comparison divided by cancer type.
795 (**C**) Junctional CN difference profiles in ecDNA and ICamp regions. Statistical significance was
796 determined by Student's *t* test. (**D**) CN distributions in isogenic paired cell lines: ecDNA-mediated
797 double minutes (COLO320-DM and GBM39-DM) versus homogeneously staining regions
798 (COLO320-HSR and GBM39-HSR). (**E**) Cytogenetic features of GBM39 isogenic pair exhibiting
799 numerous variable sized, dispersed signals consistent with *EGFR*-bearing ecDNA (top) and
800 intrachromosomal clustered amplification manifesting as homogeneously staining region (HSR).
801 (**F**, **G**) Pseudobulk CN/SV profiles (top) and single-cell CN profiles (bottom) for DM versus HSR
802 lines reveal genomic rearrangements that occurred after the most recent common ancestor and
803 prior to HSR formation in COLO320 (**F**) and GBM39 (**G**) isogenic pairs.