

Evolution of core archetypal phenotypes in progressive high grade serous ovarian cancer

Aritro Nath

City of Hope

Patrick A. Cosgrove

City of Hope

Benjamin Copeland

City of Hope

Hoda Mirsafian

City of Hope

Elizabeth L. Christie

Peter MacCallum Cancer Centre

Lance Pflieger

City of Hope

Sumana Majumdar

City of Hope

Mihaela C. Cristea

City of Hope

Ernest S. Han

City of Hope

Stephen J. Lee

City of Hope

Edward W. Wang

City of Hope

Sian Fereday

Peter MacCallum Cancer Centre

Nadia Traficante

Peter MacCallum Cancer Centre

Ravi Salgia

City of Hope

Theresa Werner

University of Utah

Adam L. Cohen

University of Utah

Philip Moos

University of Utah

Jeffrey T. Chang

University of Texas Health Science Center at Houston

David D. L. Bowtell

Peter MacCallum Cancer Centre

Andrea H. Bild (✉ abild@coh.org)

City of Hope

Research Article

Keywords: single-cell RNA sequencing, ovarian cancer evolution, archetype analysis, metabolism and proliferation

Posted Date: January 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-141400/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The evolution of resistance in high-grade serous ovarian cancer (HGSOC) cells following chemotherapy is only partially understood. To uncover phenotypic changes associated with chemotherapy resistance, we profiled single-cell RNA-sequencing (scRNA-seq) transcriptomes of HGSOC tumors collected longitudinally during patient treatment. Analysis of scRNA-seq data from two independent patient cohorts revealed that HGSOC is driven by three core archetypal phenotypes, defined as oncogenic tasks that describe the majority of the transcriptome variation. A multi-task learning approach to identify the biological tasks of each archetype identified metabolism and proliferation, cellular defense response, and DNA repair signaling. The metabolism and proliferation archetype evolved during treatment and was enriched in cancer cells from patients that received multiple-lines of treatment and had elevated tumor burden indicated by CA-125 levels. The emergence of archetypes was not consistently associated with specific whole-genome driver mutations. However, archetypes were closely associated with subclonal populations at the single-cell level, indicating that subclones within a tumor often specialize in unique biological tasks. Our study reveals the core archetypes found in progressive HGSOC and shows consistent enrichment of subclones with the metabolism archetype as resistance is acquired to multiple lines of therapy.

Introduction

Transcriptional dysregulation is a hallmark feature and a driver of evolution in human cancers (1). As one of the deadliest forms of gynecological malignancy, the survival rates for high-grade serous ovarian cancer (HGSOC) have remained poor over the past few decades (2). Despite initial responsiveness to platinum-based chemotherapy and the introduction of novel combination therapeutic interventions (3), the development of resistance over the course of treatment remains a major challenge in the clinical management of HGSOC (4, 5). Thus, characterizing the key transcriptional changes in HGSOC tumor evolution is critical for understanding tumor progression and resistance to cancer therapy (6, 7).

A majority of HGSOCs arise from the fallopian tubes (8) and are characterized by somatic alterations leading to the loss of function of the tumor suppressor gene *TP53* (9, 10) and regulators components of homologous recombination (HR) DNA-damage repair pathway, including *BRCA1* and *BRCA2* (11). Whole-genome sequencing (WGS) analyses have revealed several key genomic mechanisms of acquired resistance, such as somatic alterations in the multi-drug resistance gene *ABCB1*, and secondary somatic mutations alterations in HR genes and the protection of stalled replication forks (12–14). However known mechanisms explain only a fraction of resistance drivers (12). Therefore, focusing on transcriptional changes could help improve our understanding of chemoresistance, especially in cases where obvious single gene alterations are not detectable. Further, the number of critical signaling pathways important for HGSOC cell growth and survival is unknown; therefore, therapeutic regimens may miss important oncogenic traits and enable progression.

Rapid developments in single-cell RNA-sequencing (scRNA-seq) technologies have enabled the investigation of intratumor heterogeneity and evolution at the cellular level (15–17). Longitudinal analysis of tumors in response to drug treatment using scRNA-seq combined with genomic sequencing has been utilized to understand the ecology and evolution of tumors along with phenotypic mechanisms that could be harnessed as potential drug targets in resistant tumors (18). Key questions in HGSOV that remain to be addressed include: 1) the number of key phenotypic features in progressive tumors, 2) the biological processes underlying progression, 3) how changes in the number of cells specializing in specific phenotypes contribute to progression, and 4) how genetically distinct subclonal populations impact phenotypic diversity.

Recent developments in cancer evolutionary theory suggest tumor cells can evolve to exhibit a range of phenotypes under selective pressure like chemotherapy (19, 20). However, every cell in the tumor can only exist in a transcriptional state that is optimal at performing a single phenotypic task owing to metabolic and spatial constraints (20). Thus, elucidating the biological tasks associated with transcriptional specialists in chemoresistant HGSOV could help in developing new therapeutic strategies targeting these emergent phenotypes. To identify the number and biological function of tasks associated with HGSOV cancer cell progression, we employed a method that uses the Pareto optimization concept, which states that there is a combination of tasks that dominate an organism's fitness (20). The approach defines a polygon, where the number of vertices reflects the number of tasks describing the data. The cells at the edges of the polygon (termed "archetype") specialize in that specific biological task (21). Based on these principles, we projected the scRNA-seq profiles from the HGSOV samples on to archetypes to determine the number of driver phenotypes in the data, the biological features of those archetypal phenotypes, and to study if cells specialize in specific archetypal tasks during progression. Finally, these archetypes are evaluated together with genetic alterations to identify the potential link between somatic alterations and phenotypic state.

In this study, we used malignant ascites and pleural effusion samples from 9 HGSOV patients, collected over months to years of treatment, to perform scRNA-seq and WGS analysis. We also performed scRNA-seq analysis of an independent cohort of unmatched 8 pre-treatment and 7 post-treatment samples to study longitudinal patterns of transcriptomic heterogeneity. Our results show enrichment of an archetype associated with elevated metabolic activity, driven by oxidative phosphorylation or glycolysis, and proliferation in post-treatment patients compared to treatment-naïve patients of the validation cohort. Further, cell defense and DNA repair describe two additional key archetypal phenotypes in HGSOV. While consistent genomic alterations did not define the archetypes, sub-clonal clusters inferred from scRNA-seq profiles were associated with the enrichment of the metabolic archetype as cancer cells

progress on therapy. Finally, we validated the metabolic activities of tumor cells derived from patient samples *in vitro* to confirm a pattern of increased metabolic activity in post-treatment samples.

Results

Temporal transcriptomic diversity of HGSOC cells

To study the landscape of genetic and transcriptomic heterogeneity of ovarian cancer cells in response to chemotherapy, we obtained 25 malignant ascites or pleural effusion samples from 9 HGSOC patients (Figure 1a, Supplementary Tables 1-3). Samples were collected over the course of treatment, with initial samples obtained at the time of surgery or before the commencement of therapy in 5 of the 9 patients, and early in treatment for the remainder. Following initial debulking surgery, patients received adjuvant platinum-taxane based chemotherapy as a first-line treatment followed by 3 to 7 lines of chemotherapy over the course of their disease progression (Fig. 1a and Supplementary Tables 1-3). The samples were processed to isolate nuclei or whole cells to perform single-cell RNA sequencing (scRNA-seq), whole-genome sequencing (WGS), and establish *in vitro* cell lines for metabolic assays (Figure 1b and Materials and Methods). In addition, we obtained 8 pre-treatment and 7 post-treatment malignant ascites or pleural samples as a validation cohort for the scRNA-seq analyses (Supplementary Table 4).

We analyzed the transcriptomes of approximately 37,000 high-quality cells or nuclei using scRNA-seq. Preliminary clustering of the scRNA-seq data resulted in the separation of cells by patients (Supplementary Figure 1). Following batch-correction with canonical correlation analysis (CCA) (22), unsupervised clustering resulted in 8 clusters representing individual cell types instead of patient identity (Figure 1c). Reference-based prediction of cell types (23) revealed large clusters of predominantly malignant epithelial cells (clusters 1-2, 7), confirmed by the expression of the epithelial marker (EPCAM) and the tumor biomarker MUC16(CA-125) (Supplementary Figure 2). Despite prior immune depletion, we also detected smaller but distinct clusters of immune cells, including CD4+/CD8+ T cells (cluster 3), monocytes, natural killer (NK) cells (cluster 4), and fibroblasts (cluster 5). In total, approximately 27,000 cells were identified as malignant epithelial cells that were distributed across multiple clusters reflecting the heterogeneity within the population (Figure 1c).

Transcriptional evolution of ovarian cancer cells is associated with core biological tasks

To interrogate transcriptional heterogeneity in progressive HGSOC, we applied an approach that accounts for tumor evolution and the use of tasks to enhance fitness (20). Our goal was to determine how many archetypal phenotypes are found in HGSOC and how these tasks evolve as patients receive therapy and become resistant. We utilized a Pareto task inference method that relies on the principle convex hull algorithm to identify core archetypes (24). Briefly, the method attempts to identify a polygon that can

best enclose the principal component projection of the gene expression data. The vertices of this polygon are inferred as the core archetypes. Analyses were limited to scRNA-seq profiles obtained using the 10X platform (patients 4-9) with sufficient numbers of malignant epithelial cells available for projection on to the archetypes. To determine the shape of a polygon that can best enclose the data, we performed simulations with varying number of vertices (Supplementary Figure 3). Based on this simulation, we determined that a three-vertex triangle was sufficient to enclose the data (Figure 2a).

To determine the distinct biological tasks associated with each archetype, we implemented a multi-task learning approach based on group-lasso (see methods for details) that applied the hallmark pathways and genes to cells located on each archetypal vertices. Hierarchical clustering analyses with pathway coefficients show three distinct clusters linked to the archetypes (Figure 2b, Supplementary Figures 4-5). Three broad tasks are associated with these archetypes, including metabolism and proliferation (MAP), cellular defense response (CDR), and DNA damage repair (DDR) (Figure 2c). The MAP archetype was defined based on the enrichment of multiple key metabolism phenotypes including glycolysis, oxidative phosphorylation, and proliferation pathways associated with cell cycle and E2F genes, G2M checkpoint genes, and MYC targets (Supplementary Figures 4-5). Association of key genes indicative of proliferation (*MKI67*) and glycolysis (*GAPDH*) (25) supported the classification of this archetype. The CDR archetype was defined based on enrichment of the interferon-gamma response pathway (26) and the enrichment of multiple downstream pathways and genes related to the activation of cellular defense response, including canonical JAK/STAT (27), WNT/beta-catenin (27), PI3K/AKT (28) and MTORC1 (29) signaling pathways, as well as cell cytokine and immunogenic signaling genes. Finally, the DDR archetype was derived based on enrichment of apoptosis, P53, and TNFA (30) signaling pathways along with key genes such as *ATM* (31) and *CHEK1* (32).

We also performed archetype analysis on an independent validation cohort of 8 pre-treatment and 7 post-treatment malignant ascites or pleural samples. All pre-treatment samples were from treatment naïve patients. Post-treatment patients received an average of 5 lines of treatment, including chemotherapies and targeted therapies (Supplemental Table 4). Dimensionality reduction and clustering of the cells from the early (pre-treatment)/late (post-treatment) cohort resulted in a large malignant cell cluster of epithelial origin and smaller immune cell and fibroblast clusters (Figure 2e), confirmed by expression of individual markers (Supplementary Figure 6). As with the initial longitudinal cohort, we performed the Pareto task inference analysis on the validation cohort samples revealing three major archetypes (Figure 2e, Supplementary Figure 7a). Group-lasso coefficients of the pathways associated with three archetypes validated the three classes predicted in the longitudinal cohort (Figure 2f, Supplementary Figure 8). By projecting the archetypes separately on the early pre-treatment (top panel) and late post-treatment (bottom panel) cohorts, we observe a clear transition between the archetypes with a shift towards the

metabolism and proliferation archetype (A2) after therapy (Figure 2e, 2f), with both cell defense and DNA repair archetypes decreasing over time.

Next, we confirmed the presence of the archetypes by resolving the scRNA-seq profiles of each patient from the longitudinal cohort. Simulations confirmed a three-vertex polygon was sufficient to enclose the complete data for each patient (Supplementary Figure 9). The biological phenotypes associated with these three clusters were consistent with the phenotypes associated with the archetypes identified in the integrated dataset when performed separately for each patient (Figure 3a, 3b, Supplementary Figure 10).

We then evaluated the patterns of shifts in the populations of specialist cells, defined as cells close to a vertex representing one of the key archetypes (MAP, CDR, or DDR) during treatment of our initial patient cohort (Figure 3c, Supplementary Table 5). At most time points, cancer cells were present that specialized in each of the three key archetypes, with three of five patients having an enrichment in either the MAP or CDR specialists. In the case of patient 4, all three archetypes were present at the three-time points. MAP was the principal archetype at the first time point (42%), and with most cells specializing in the CDR (35%) or MAP (24%) archetypes at the last time point. The proportions of MAP archetype were higher at time 1 and 3 compared to time 2, which coincide with the lower CA-125 levels of the patient while on treatment during time 2 (Figure 1a). The relative proportion of the specialists in patient 5 did not change over time, with MAP remaining the critical archetype at the last time point (46%). Patient 6 also showed a pattern of MAP archetype enrichment that show similar trends as CA-125 burden, with the highest levels at time 1 (30%) and time 3 (32%) compared to time 2 (11%). In the case of patient 7, CDR emerged as the core archetype at the second time point (81%). Although the two samples for patient 7 were collected across a gap of more than 3 years, the CA-125 levels were relatively low compared to the late time points of other patients (Supplementary Table 3), thus explaining the exceptional pattern of MAP specialists observed in this patient. The archetypes of patient 8 were mostly dominated by non-specialist cells (78%) at the last time point. However, this patient also showed an increase in the proportion of MAP specialists from 2 to 5% between the first and last time point, with the highest proportion of MAP specialists (16%) coinciding with the highest CA-125 levels for this patient at time 2 (Figure 1a, Supplementary Table 3). In the case of patient 9, the proportion of MAP specialists increased progressively over time from 9% to 22%, again following the CA-125 levels for this patient. Thus, the patterns of the shift in the MAP archetype derived from the scRNA-seq data agree well with the levels of tumor marker over time.

The metabolic and proliferative archetype is enriched in late-stage resistant HGSOV cancer compared to treatment naïve cancer cells

We next compared the specialist populations in the unpaired treatment-naïve & post-treatment validation cohorts. We observed MAP specialists in only two out of 8 treatment-naïve samples, while all 7 post-treatment samples showed an enrichment of MAP specialists as the principal archetype (Figure 4a). On average, this reflected an 11% shift towards the MAP archetype in post-treatment samples, while both DDR and CDR decreased by 5-10% in the post-treatment samples (Figure 4b). While the enrichment of the MAP archetype correlated with the tumor burden in the longitudinal cohort, the dramatic increase in the MAP specialists in the post-treatment validation cohort suggests multiple lines of chemotherapy may also contribute to this shift. To experimentally validate the observed shift in the MAP archetype, we derived multiple primary cancer cells from patients 4 and 8 and tested the metabolic capacity changes over time. We created early passage primary patient cell lines using ascites samples from the two patients. In both cases, the late samples were obtained at an advanced stage after several lines of treatment. These serial cell lines displayed an increased basal ATP production capacity over time, with the majority of the energy production was contributed by the glycolytic pathway in patient 4, and both oxidative phosphorylation and glycolysis in patient 8 (Figure 4c). Lastly, to test the relative metabolic potentials in the cancer cells from our independent validation cohort, we also created cell lines from two pre-treatment patients (patients 16 and 17) and two post-treatment cohort patients (patients 21 and 23). We compared the ATP production rates for these four unmatched samples and found an overall increase in ATP production in the late treatment samples, contributed by both the glycolytic and oxidative phosphorylation pathways (Figure 4d).

Temporal evolution of genomic variants in progressive HGSOV

In order to test for association between genetic variants and archetype during tumor evolution, we next performed whole-genome sequencing analysis of germline and malignant samples from the longitudinal cohort patients. We observed an average of 12,000 SNVs and small indels in our samples along with an average tumor mutation burden of about 2.5/MB, with 6 out of 9 patients displaying an increase in mutation burden over time (Figure 5a, Supplementary Table 6). An average of 800 structural variants including indels (>25bp) and breakpoints were observed in our samples (Figure 5b, Supplementary Table 6). In addition, we also observed copy number gains or losses in 1 to 8 cancer genes per sample (Figure 5c). The non-synonymous SNVs, splice-site variants, indels, structural variants (breakpoints), protein interaction variants, and copy number variants affecting cancer genes are shown in Figure 5d.

To determine the potential pathogenicity of the non-synonymous SNVs and small indels, we searched for potential drivers by comparing the mutations in our samples with the IntoGen list of predicted and validated driver mutations. We also genotyped and predicted the impact of variants affecting homologous recombination (HR) genes in the germline samples. Pathogenic germline mutations in *BRIP1* in patient 1 and *BRCA1* in patient 8 were previously shown to contribute to the deficiency of the HR pathway (12, 14). Additional HR variants in germline samples were predicted to be either benign or

common SNPs of unknown significance (Supplementary Table 7). Somatic missense and splice-site mutations in only one gene, *TP53*, were found to be potential drivers. Truncal *EPHA3* mutations in patients 5 and 8, and *RHOA* in patient 9 were determined to be high impact based on SnpEff annotations, however, predicted as benign or passenger by IntoGen (Supplementary Table 8). We found frequent copy number gains of the *MYC* and *IGF2BP2* oncogenes, both associated with progressive ovarian cancers (33–35). Similarly, copy number gains were observed in the *PIK3CA* (36), *ERBB2* (37) and *SOX2* (38) oncogenes, each reported to be potentially associated with chemotherapeutic resistance in ovarian cancers (Figure 4d, Supplementary Table 9). In addition, we also observed a copy number loss of tumor suppressor genes *NF1* in patient 1, and *RB1* in patient 7 (Figure 4d, Supplementary Table 9).

We next mapped the major acquired genomic events that tracked with the progression of the longitudinal cohort patients (Figure 6). In patient 1, where the samples were profiled 1474 days apart, a copy number loss of *NF1* along with a gain of *IGF2B2* were observed at time 2 (Figure 6a). In the case of patient 2, we did not observe acquired copy number gains or losses in cancer genes. However, a breakpoint in *ESR1* exon was acquired at the second time point (Figure 6b). In the case of patient 3, the first sample was collected after the patient had already received first-line chemotherapy. We observed acquired copy number gains in *AKT2* oncogenes, along with a pathogenic *ABCB1-SLC25A40* fusion at the second time point (Figure 6c) that had previously been reported.

Patient 4 did not show acquired CNVs at the second time point. Concurrent with an increase in the CDR archetype, we detected a passenger missense mutation in *LTR1* exon and a breakpoint in *ERBB4* exon of unknown significance (Figure 6d). The CNV and SV profiles of patients 5, where the MAP was the critical archetype at all points, showed no variants that affected cancer genes. This patient acquired a splice-site *COL2A1* and missense *TPR1* passenger mutations at times 2 and 3 respectively (Figure 6e). The SNV, CNV and SVs in the samples from patient 6 profiled in a short period were truncal (Figure 6f). Several key oncogenes were amplified in the second time point for patient 7, including *PIK3CA*, *KRAS*, and *SOX2*, along with a loss of *RB1* copy (Figure 6g). These acquired driver mutations in patient 7 correspond to the emergence of CDR as the principal archetype at the second time point. *PIK3CA*, *PIK3CB*, and *IGF2BP2* were amplified at the second time point for patient 8, along with acquired breakpoints in the exons of *ARID1B* and *FEN1*. In contrast with patient 7, the second time point in the patient was associated with a decrease in CDR and an increase in the MAP archetypes. Overall, the relatively larger number of acquired events affecting similar pathways (PI3K/MAPK) in both patients 7 and 8 appeared to coincide with the long gap between the sample collection times but did not correspond to the evolution of the same archetypes (Supplementary Figure 12). Patient 9, where the samples were profiled within a short time span and showed consistent enrichment of the CDR and MAP archetypes, did not acquire CNVs. However, a passenger missense variant in *NUP98* was identified at time 2 along with a breakpoint in an exon of *DMNT* at time 3.

To evaluate the association between key genomic variants and archetypes, we compared the proportion of specialists across samples grouped by the presence of a mutation (Figure 7). Samples grouped by pathogenic *TP53* mutations, present in 21 out of 24 samples, did not show any significant difference in the proportion of specialists for any archetype (Figure 7a). Similarly, we did not observe a significant difference in the proportion of specialists in samples grouped by *MYC* gain in 14 samples (Figure 7b), *IGF2BP2* gain in 10 samples (Figure 7c), *PIK3CA* gain in 6 samples (Figure 7d) or *ERBB2* gain in 7 samples (Figure 7e). However, we found that *SOX2* gain detected in 6 samples was associated with a significantly higher proportion of CDR specialists ($P = 0.02$) (Figure 7f). Other archetype specialists were not significantly different.

Single-cell sub-clones are associated with emergent archetypes

As shown above, driver genomic alterations that were acquired as a function of time or in response to chemotherapy could not completely explain the observed shift of all archetypes or development of therapeutic resistance over time. Therefore, we next investigated whether the sub-clonal architecture of the single cells might be associated with the emergence of the archetypes. We determined the sub-clonal structure of the longitudinal cohort scRNA-seq samples using the InferCNV method, assuming that the transcriptional heterogeneity at the single-cell level was driven by alterations that resulted in a change in expression levels of contiguous genes along the chromosomes (39) (Supplementary Figure 12). Overall, we found that the archetypes were significantly associated with specific inferred sub-clonal clusters in most patients (Supplementary Figure 13).

The CDR archetype was associated with the sub-clone cluster 1 in patient 4, displaying an enrichment at the later time points compared to the initial time point (Figure 8a), while the MAP archetype was associated with sub-clone cluster 2, and DDR with sub-clone cluster 3 (Supplementary Figure 13). In patient 5, the DDR archetype was associated with sub-clone cluster 1 and remained the core archetype through-out the study (Figure 8b, Supplementary Figure 13). Patient 6 MAP archetypes enriched at later time points were associated with sub-clone cluster 3 (Figure 8c, Supplementary Figure 13). The MAP archetype present in the first time point of patient 7 was enriched in the sub-clone cluster 3, while CDR that became the core archetype the later time point was linked to cluster 2 (Figure 8d, Supplementary Figure 13). In patient 8, the MAP archetype was enriched at the later time points and tracked with sub-clone cluster 2 (Figure 8e, Supplementary Figure 13). Patient 9 showed a shift towards the MAP archetype at the later time points and was associated with sub-clone cluster 3 (Figure 8f). Thus, we observed that specific sub-clonal clusters were associated with the key MAP archetype in most patients. However, every cluster could not be mapped to an observed whole-genome amplification or deletion event. Therefore, further resolution of the genetic, epigenetic, and regulatory driver events behind the sub-

clonal evolution of the HGSOC tumors could help elucidate the mechanism of archetype shifts in these cells.

Discussion

With the emergence of scRNA-seq technologies, it is now possible to study the patterns of transcriptional evolution at the cellular level. Understanding the patterns of transcriptional heterogeneity at the single-cell level may help elucidate the mechanisms of chemoresistance in HGSOCs, especially in cases where driver single-gene genomic alterations could not be detected. HGSOCs presents a unique challenge, where the genetic heterogeneity is generally driven by SVs and CNVs, rather than single-gene driver mutations affecting cancer-related genes. This observation was confirmed in our WGS analysis of the temporal samples from the longitudinal cohort patients, where only truncal *TP53* driver mutations were detected across most patients.

Our study utilized the theory of multi-task evolution to characterize transcriptional heterogeneity over time. This theory suggests that tumors perform various biological tasks, however each cell within the tumor is optimized to perform only a specific task, which can evolve under selective pressure. By identifying the transcriptional task-specialists in HGSOC cells and defining the core phenotypes that evolve during the course of treatment, we may be able to identify therapeutic targets against those critical phenotypes. A key finding of our study was the identification of three major archetypes or transcriptional specialists that could describe the gene expression of HGSOC cells. These included the metabolism and proliferation archetype, the DNA repair archetype, and the cell defense archetype (Figure 2b, c). We found that the metabolism and proliferation archetype (MAP) was predominantly observed in the longitudinal cohort samples where patients had previously received chemotherapy and showed elevated tumor burden (CA-125 levels). In addition, the post-treatment samples showed a dramatic enrichment of cells specializing in the MAP archetype when compared with treatment-naïve samples in the validation cohort. In previous reports, ovarian cancer cell lines have been characterized to show metabolic reprogramming of cancer cells that supported survival, promoted the development of chemoresistance and disease progression (40, 41). Thus, our scRNA-seq models lend support to these *in vitro* observations by demonstrating a shift towards a high-metabolism archetype in post-treatment patient-derived samples.

Clinical interest to target this metabolic phenotype has garnered the attention of many investigators seeking to utilize combination therapies for more effective treatment options. In our WGS analyses, acquired SVs and CNVs affecting the *MYC* oncogene were detected in about 1/3 samples from the original 9 patients of the longitudinal cohort. Increased *MYC* expression potentially contributes to the increased metabolic phenotype of ovarian cancers through increased glycolysis mediated by LDH as well as glutamine addiction in *MYC*-driven cancers (42). Thus, drugs such as BRD4 inhibitors that target the upstream pathways regulating *MYC* may be attractive candidates to control the metabolism and growth

of chemoresistant cancer cells. Indeed, a small molecular BRD4 inhibitor, JQ1, has been shown to inhibit cell proliferation, induce apoptosis as well as increase sensitivity to cisplatin in ovarian cancer cells (43, 44). Concurrent with a shift towards a high-metabolism and EMT archetype in the post-treatment samples, we also observed an increase in the proportion of non-specialist cells at the late time points (Figure 6b, 6h). Previous studies have shown evidence linking EMT with both ovarian cancer progression and acquired chemotherapeutic resistance (45). In addition, the activation of the EMT program is closely associated with increased plasticity, reprogramming of metabolism, and metastatic progression of cancer cells (46, 47). Recent studies show that the activation of the EMT program may be regulated via epigenetic mechanisms instead of somatic variants (48–50). In addition, aberrant ovarian cancer cell metabolism was recently shown to be regulated by microRNAs using *in vivo* models (51). Thus, the concurrent shift towards a high-metabolism archetype and enrichment of the EMT pathway, along with an increase in plasticity, may be driven by epigenetic mechanisms and serve as potential therapeutic targets (52).

A key outstanding question emerging from our study is the mechanisms that could explain the observed shift in archetypes over time. We found that driver somatic mutations were not associated with the emergence of archetypes across the patients. This does not completely rule out the potential role of genetic mechanisms in archetypal evolution, as evidenced by the close association of archetype shifts with specific subclones. However, resolving the specific somatic mutations in single-cell subclonal populations is quite challenging due to the low depth of coverage and sparsity of the scRNA-seq data. Improvements in scDNA-seq technologies and development of analytical methods to resolve somatic mutations in single cells may help bridge this gap in knowledge (53). On the other hand, transcriptional evolution could be also driven by non-genetic mechanisms, including epigenetic alterations and acquired changes in the non-coding transcriptome of single cells (54, 55). Complete characterization of such mechanisms would require new technologies to simultaneously profile and study such changes.

Overall, our study shows compelling evidence of tumor evolution driven by a shift towards a high-metabolism biological state in HGSOCs after receiving multiple lines of chemotherapy in two separate patient cohorts. Additionally, we found a correlation between an increase in plasticity and a shift towards the high-metabolism archetypes that may be driven by the activation of the EMT program. *In vitro* assays confirm the shift towards a high-metabolism state in the post-treatment samples, suggesting the potential of new avenues of therapy that specifically target the metabolic pathways in chemoresistant HGSOCs.

Materials And Methods

Sample Collection and Processing

Ascites or pleural effusions were drained and collected from nine ovarian cancer patients longitudinally over the course of patient treatment. Samples were collected under IRB # 07047 & 17334 (City of Hope),

41030 and 89989 (University of Utah), or HREC # 01/60, 16/161 by the Australian Ovarian Cancer Study (AOCS) which were analyzed under HREC # 15/84 (Peter MacCallum Cancer Centre). Malignant fluids were centrifuged at 500 x g for 5 minutes to pellet cells. Red blood cells were removed by lysis in Tris-Ammonium Chloride Buffer (17mM Tris, pH 7.4, 135mM Ammonium Chloride) and incubated for 5 minutes in a 37°C water bath. Cells were then centrifuged at 500 x g for 5 minutes at room temperature and repeated until red blood cells were removed. Cells were washed in 1x PBS (Gibco, Cat # 10010) before frozen viably in 50% RPMI-1640 (Gibco, Cat # 11875) + 40% Fetal Bovine Serum (FBS, Sigma, Cat# 12306C) + 10% DMSO (Fisher Scientific, Cat# D2650). Ascites fluid collected by the Australian Ovarian Cancer Study (AOCS) was centrifuged at 1500rpm for 10 minutes at 4°C. Red blood cells were removed by incubation in ice-cold lysis buffer (14.4uM NH₄Cl, 1uM NH₄HCO₃) at room temperature for 10 minutes. Cells were centrifuged at 1500rpm for 10 minutes at 4°C, washed in 10% FBS in 1x PBS, and centrifuged again. Cells were frozen viably in FBS + 10% DMSO.

Cancer Cell Isolation

Frozen viable ascites or pleural effusion cells were thawed, centrifuged at 300 x g, and resuspended in 1x PBS to determine, concentration, viability, and cancer cell purity by trypan blue staining. In some cases, cancer cells were purified by Miltenyi Biotec QuadroMACS by negative selection of CD45+ (CD45 MicroBeads, Miltenyi Biotec, Cat# 130-045-801), CD90+ (CD90 MicroBeads, Miltenyi Biotec, Cat# 130-096-253), and podoplanin expressing cells (biotinylated anti-podoplanin antibody, Biolegend, Cat# 3370015 and Miltenyi Biotec anti-biotin MicroBeads Cat# 130-105-637) using LD columns (Miltenyi Biotec, Cat# 130-042-901) according to manufacturer instructions. The samples were processed using the StemCell EasyRights EasySep column-free magnet to remove CD45+ (EasySep CD45 Depletion Kit II, Cat# 17898), and/or dead cells (EasySep Dead Cell Removal (annexin V) Kit, cat# 17899) as appropriate. To isolate cancer cells using StemCell EasySep antibody Kits, cells were centrifuged and resuspended in 1x PBS + 2% FBS + 1mM Calcium Chloride (G-Biosciences, Cat# R040) to a concentration of <10⁸ cells per 2mL total volume and transferred to a round bottom 5mL FACS tube. Sequentially Dead Cell Removal Cocktail (50µL/mL sample), and Biotin Selection Cocktail (50µL/mL sample) was added and incubated at room temperature for 3 minutes, followed by CD45 Depletion Cocktail (50 µL/mL sample) and incubated at room temperature for 5 minutes. StemCell RapidSphere magnet beads were added (75µL/mL for CD45 RapidSpheres and 100µL/mL for Dead Cell RapidSpheres) and incubated at room temperature for 3 minutes off the magnet. Cell samples were then incubated on EasyEight magnet for 5 minutes, collected supernatant, and repeated additional EasyEight magnet column cleanup. Collected cells were then centrifuged and resuspended in 1x PBS and maintained at 4°C.

Nuclei Isolation

After cancer cell isolation, patient samples that did not dissociate into single-cell suspensions, or had a high proportion of cancer cell clusters, were then processed to isolate single nuclei suspensions. To isolate nuclei, cells were resuspended in (4:1) Lysis Buffer (10mM Tris-HCl, pH 7.8 (Teknova, Cat# T1078), 146mM NaCl (Alfa Aesar, Cat# J60434AK), 1mM CaCl₂ (G-Biosciences, Cat# R040), 21mM MgCl₂ (G-Biosciences, Cat# R004), 0.05% BSA (EMD Millipore, Cat # 12661525), 0.2% Igepal CA-630 (MP Biomedicals, Cat# 198596), DNase/RNase free water (Gibco, Cat# 10977)) : DAPI Buffer (106mM MgCl₂, 50 µg/mL 4', 6-diamidino-2-phenylindole (DAPI, Invitrogen, Cat# D1306), 5mM Ethylenediaminetetraacetic acid (EDTA, Quality Biological Inc., Cat# E522100ML), DNase/RNase free water)) supplemented with fresh 0.2 U/µL SUPERase•In RNase Inhibitor (Invitrogen, Cat# AM2694). Cells were incubated for 15 minutes at 4°C to release nuclei. The lysate was then filtered through a 40 µm mesh filter (Falcon, Cat# 352340) collecting nuclei in flow-through. All downstream nuclei processing utilized Eppendorf LoBind DNA tubes to prevent nuclei loss. Nuclei were centrifuged 500 x g, at 4°C, for 5 minutes and washed two times with 500 µL of 1x PBS + 1% BSA + 0.2U/µL SUPERase•In RNase Inhibitor. Nuclei were resuspended in 1x PBS + 1% BSA + 0.2U/µL SUPERase•In RNase Inhibitor at a target of 1,000 cell/µL, re-filtered using a 40 µm mesh filter, and counted on a hemocytometer by DAPI fluorescence using an Invitrogen Countess equipped with DAPI filter cube and maintained at 4°C.

Single-cell RNA Sequencing (scRNA-seq)

Single-cell RNA-Sequencing (scRNA-seq) was performed on single cell or single nuclei suspensions using either the Takara Bio ICELL8 Single-Cell System or the 10X Genomics Chromium to prepare cDNA sequencing libraries. Samples processed on the ICELL8 Single-Cell System (Takara Bio) were prepared using the SMARTer ICELL8 3' DE Reagent Kit V2 (Takara Bio, Cat # 640167) from isolated nuclei. DAPI stained nuclei were diluted to a concentration of 60,000 cell/mL in 1x PBS + 1% BSA + 1x Second Diluent + 0.2U SUPERase•In RNase Inhibitor and dispensed onto the ICELL8 3' DE Chip (Takara Bio, Cat# 640143) using the ICELL8 MultiSample NanoDispenser. Single nuclei candidates were selected using the ICELL8 imaging system with CellSelect Software (Takara Bio) selecting for DAPI positive nuclei and reverse transcription and sequencing library preparation was performed according to manufacturer instructions. ICELL8 cDNA sequencing libraries were sequenced at a depth of 200K reads/cell on Illumina HiSeq 2500, read #1= 26nt, and read #2= 100nt.

Samples processed on the 10X Genomics Chromium were processed using the Chromium Single Cell 3' V3 Kit (10X Genomics, Cat # 1000075) using whole cells or isolated nuclei. Single cells or nuclei were diluted to a target of 1,000 cell/µL in 1x PBS (whole cells) or 1x PBS + 1.0% BSA + 0.2U/µL SUPERase•In RNase Inhibitor to generate GEM's prepared at a target of 5,000 cells per sample. Barcoding, reverse transcription, and library preparation were performed according to manufacturer instructions. 10X Genomics generated cDNA libraries were sequenced on Illumina HiSeq 2500 or NovaSeq 6000 instruments using 150 cycle paired-end sequencing at a depth of 10K reads/cell. scRNA-seq was

performed at the Integrative Genomics Core at City of Hope, Fulgent Genetics, and the High Throughput Genomics Core at Huntsman Cancer Institute (HCI) of University of Utah.

Genomic DNA Isolation and Whole Genome Sequencing (WGS)

Genomic DNA was isolated using the QIAamp DNA Micro Kit (Qiagen, Cat # 56304) according to manufacturer instructions for isolated cancer cells and nuclei suspensions from scRNA-seq, as well as patient-matched buffy coat for germline DNA. Germline DNA was also isolated from patient matched isolated peripheral lymphocytes using the salting-out method. Briefly, lymphocytes were resuspended in nuclei lysis buffer (0.1M Tris pH8, 2mM EDTA pH8, 0.1M NaCl, proteinase K & SDS), and incubated at 56°C for 1hr followed by 37°C for 3hrs. Saturated salt solution (~6M NaCl) was added to lysed cells, which were centrifuged at 14000rpm for 15min at 4°C after vigorous mixing. The supernatant was transferred to ice-cold 100% ethanol and the tubes were rocked gently until the DNA precipitated. After overnight incubation in ethanol at -20°C, DNA was rinsed twice by placing in 70% ethanol, centrifugation, and removing the ethanol. DNA was air-dried and resuspended in sterile water. WGS DNA libraries were prepared using either NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), KAPA Hyper Prep PCR Free Library Prep Kit (Roche), or Nextera DNA Flex Library Prep Kit (Illumina), and sequencing performed on Illumina NovaSeq 6000 instruments at 150 cycles and paired-end sequencing to read depth of 40-60X coverage. Sequencing was performed at Admera Health, Fulgent Genetics, and the High Throughput Genomics Core at HCI of University of Utah.

Cell Culture

To create stable patient-derived primary cell lines, frozen patient ascites were processed and then immediately placed in media as specified below. All cells were maintained in RPMI 1640 (Gibco; cat # 11875085) supplemented with 10% heat-inactivated FBS (Sigma, cat # 12306C) and 1% antibiotic/antimycotic solution (Gibco; cat #15240062) in uncoated filter top polystyrene flasks and maintained at 37°C in 5% CO₂, patient cells were additionally kept in 5% O₂ hypoxic simulated humidified air.

Metabolic Assays

ATP production rates were assayed with the XF Real-Time ATP Rate Assay Kit (Agilent, cat # 103592-100) as per the manufactures' instructions. Briefly, cells were plated down in the Seahorse XF96 cell culture microplates at 10000 cells/well/80 µL and placed back in 37°C, 5% CO₂ incubator. After 24 hours cells were washed in assay media made up from Seahorse XF RPMI Media, pH 7.4 (Agilent, cat # 103576-100) containing 10 mM glucose (Agilent, Cat# 103577-100), 1 mM pyruvate (Agilent, cat # 103578-100), and 2 mM L-Glutamine (Agilent, Cat# 103579-100) and incubated for 1 h in a non-CO₂ incubator at 37°C before a final wash in the assay media. The Seahorse XFe96 analyzer was calibrated and the assay run using a

standard XF Real-Time ATP Rate template created using the WAVE software (V2.6.1) and assay standard drug injections were used of 1.5 μ M Oligomycin in port A and 0.5 μ M Rotenone/Antimycin A in port B.

Results for each well were normalized by cell count using 1 μ g/mL Hoechst that was added to port B with the Rotenone/Antimycin A cocktail and injected automatically, then visualized by imaging the wells at 4X on the Cytation5 multimode cell imager (BioTek) and analyzed with GEN5 software (BioTek; V3.0.5) for cell count. If multiple plates were needed for comparison, OAW42 cells were plated down at 5000 cells/well in triplicate 24 hours before the assay for environmental variable normalization between plates. Analysis for the ATP rate assay was performed using the Agilent ATP report generator as per manufacturer's recommendations.

Cell Growth and Viability Assays

Cell viability of the matched samples from patient 4 (2 samples) and patient 8 (3 samples) was assessed by the CellTiter-Glo 2.0 cell viability assay (Promega; cat # G9241) as per the manufacturer's instructions. Briefly, 1000 cells/well were plated in triplicate, flat-clear bottom 96 well plate with media as previously described. After 12 days the cells were equilibrated to room temperature for 30 mins and then equal volumes of the CellTiter-Glo reagent to media was added to each well and placed on an orbital shaker for 2 mins, then allowed to incubate for a further 10 mins at room temperature and luminescence was read on a plate reader (Tecan infinite M1000). The growth of the cells in the 96 well plates was also assessed by imaging each well every 24 hours in a Cytation5 multimode cell imaging system (BioTek). Specifically, a phase-contrast image was taken with a 4 x 4 montage, and then the GEN5 software (BioTek, V3.0.5) was used to stitch the image together and cell analysis calculated the cell count of each well. Both cell growth and viability were plotted with GraphPad (Prism V8.4.3).

Single-cell RNA sequencing analysis

Raw scRNA-seq data were pre-processed in the Bioinformatics ExperT SYstem (BETSY) (56) using the Cell Ranger v2.1.1 pipeline for 10X data, aligned to the hg19 reference genome using the STAR aligner (57), followed by extraction of read counts using featureCounts (58). The resulting count matrix of cells was used for downstream analysis using the R package Seurat v3 (59). High-quality cells were identified based on the following criteria: a minimum of 1000 total number of expressed genes per cell, a minimum of 2000 UMIs per cell, and a percentage of mitochondrial genes less than 25%. Counts matrix from individual patient samples were normalized and integrated using the canonical correlation analysis (CCA) algorithm for batch correction (59). This was followed by a principal component analysis of the variable genes in the integrated dataset, clustering using unsupervised graph-based clustering and dimensionality

reduction using uniform manifold approximation (UMAP) or t-distributed stochastic neighbor embedding (T-SNE).

The cell-type identities of the clusters were determined using a two-step approach. A first pass prediction was performed using the SingleR reference-based classification approach (23) using references based on the ENCODE (60) and HPMC (61) datasets. Next, the individual markers corresponding to predicted cell-types were mapped on to the clusters to confirm their classification. Additionally, we classified malignant epithelial cells and normal cells by inferring chromosomal copy number aberrations (CNAs) from the scRNA-seq data using the method described by Patel et al. (16). The copy numbers were inferred using the R package InferCNV, using predicted fibroblasts as reference. For pathway enrichment, raw counts were first normalized using the method proposed by Rizzo et al. (62). Then, a single sample gene set enrichment scores were calculated for hallmark (63) and curated molecular signature (64) gene sets using the GSVA package for R (65).

Whole-genome sequencing analysis

Germline and tumor WGS sequencing raw reads were pre-processed using the Bioinformatics Expert SYstem (BETSY) to add read-groups, mark duplicates, perform indel realignment, base quality recalibration, sorting and indexing, and alignment to the hg19 reference genome using BWA MEM to generate BAM files. Allele-specific CNVs calls, along with ploidy and cellularity estimates were called from the BAM files using Sequenza (66) or Facets (67) CNV callers using the corresponding germline BAM files of that patient as reference. For each sample, the CNV calls were z-transformed (allele-specific copy number – mean sample copy number / standard deviation of sample copy number) and rounded to the nearest integer for comparison. Copy number alterations were defined as z-transformed copy numbers of ≥ 2 for gains and ≤ -2 for losses.

Germline variants in homologous repair genes (*ATM*, *ATR*, *CHEK1*, *CHEK2*, *BRCA1*, *BRCA2*, *BARD1*, *BRIP1*, *FAM175A*, *MRE11A*, *NBN*, *PALB2*, *RAD51C*, *RAD51D*) along with *TP53* and *RB1* were determined by genotyping the germline BAM files using GATK, platypus, varscan and freebayes. Variants detected by at least two callers and with a VAF ≥ 0.05 were retained and annotated using SnpEff (68) to determine non-synonymous variants. Somatic SNVs and small insertions or deletions were determined from the BAM files using strelka (69), mutect2 (70) and muse (71) variant callers. Genes with a variant allele frequency ≥ 0.05 determined by at least two callers were retained for further analyses after adjusting for cellularity as determined from the CNV callers. Non-synonymous variants were first determined using SnpEff. Cancer genes were defined based on the list of cancer census genes from COSMIC (72). Potential driver mutations were defined based on the list of known or predicted drivers in the IntoGen database (73). Structural variants, including insertions, deletions, and breakpoints were called and annotated using SvABA (74). CNV and SVs were visualized as circos plots using the R package RCircos (75).

Archetype analysis and biological task classification

We analyzed the HGSOC scRNA-seq transcriptomes intending to identify distinct biological tasks that each of the cells need to perform and face evolutionary trade-offs (24). Based on the theory proposed by Shoval et al. (21), we seek to represent the transcriptome datasets as a Pareto-optimal situation by identifying that encloses the data with the vertex of the polygon representing task-specific archetypes. For this analysis, we used the first 5 principal components of the CCA-normalized scRNA-seq data from the longitudinal cohort, individual patient samples from the longitudinal and the early (treatment naïve), or late (multiline treatment) cohorts. We used the ParTI package for R (<https://github.com/vitkl/ParetoTI>) to first determine the minimum number of vertices required to enclose the transcriptome data based on the principal convex hull algorithm (76). Simulations with an increasing number of vertices revealed three vertices (triangle) were sufficient to enclose the data in each case, with additional components resulting in no gain in the proportion of variance explained by the resulting polygon. Subsequently, the polygon fit and archetype scores, or standardized Euclidean distance of each cell to the nearest vertex, were determined. For each archetype, specialist cells were defined as cells above the 80th percentile of archetype scores, while cells that did not meet this criterion for any archetype were classified as non-specialists. The evolution of the archetypes was represented as the percentage of specialists at each time point using the R package fishplot (77).

To determine the biological tasks that described each archetype, we used a gaussian multi-task or multinomial model with the set of archetype scores as the outcome variable and the hallmark gene set enrichment scores or CCA-normalized gene expression of each cell as the set of predictors. The multi-task model was fit using a group-lasso penalty using the R package glmnet (78). Briefly, 10-fold internal cross-validation was performed with a lasso penalty ($\alpha = 1$) to determine the multi-task model error over varying penalty parameter (λ) values. The contribution (coefficients) of each pathway to the model based on the fraction of deviance explained by the was also assessed over varying levels of degrees of freedom. Top pathway phenotypes contributing to the model were used to define the phenotypes associated with each archetype. Subsequently, the model was fit using a λ value within one standard error of the minimum. The group-lasso coefficients of each hallmark pathway were then analyzed using hierarchical clustering and correlation analyses to determine clusters of related pathways that were associated with each archetype. Further, the identities of the archetypes were validated based on repeated clustering pattern of the pathway coefficients determined using multi-task learning analysis of the individual patient archetypes, co-clustering of coefficients from related pathways, and expression levels of key genes that were available in the normalized scRNA-seq dataset.

Data availability and software

Single-cell RNA-Seq data generated and analysed during this study are available from the GEO database under accession GSE158722 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158722>)

The following secure token has been created to allow review of record GSE158722 while it remains in private status:kzonuiyutvkzlgp

Whole genome sequencing and raw scRNA-seq data are available under controlled access from dbGaP. The BETSY software environment (56) used for the bioinformatic analyses is available at <https://github.com/jefftc/changlab>. Custom pipelines for the pre-processing of scRNA-seq, WGS, and gene set enrichment analyses with BETSY, and Seurat, archetype, and multi-task learning analyses with R are available at <https://github.com/U54Bioinformatics>. Analyses with R-packages were performed in R-Studio (version 1.2.5033; R version 3.6.3.).

Declarations

Acknowledgements

This study was supported by a US National Cancer Institute U54 grant (U54CA209978) awarded to A.H.B., J.T.C. and D.D.L.B. D.D.L.B is supported by National Health and Medical Research Council of Australia (NHMRC) grants APP1092856 and APP1117044. E.L.C is supported by NHMRC grants APP1124309 and APP1161198. The authors acknowledge additional support from Border Ovarian Cancer Awareness Group, the Peter MacCallum Cancer Foundation, Wendy Taylor and Arthur Coombs and family. The Australian Ovarian Cancer Study (AOCS) was supported by the U.S. Army Medical Research and Materiel Command under DAMD17-01-1-0729. The authors wish to thank Fred Adler, Mark Smithson and members of City of Hope U54 community for their invaluable feedback and support.

Author contributions

Conceptualization: A.N., A.H.B., D.D.L.B.

Data curation: P.A.C., E.L.C., S.M., J.T.C.

Formal analysis: A.N., P.A.C., H.M., B.C., L.P.

Funding acquisition: A.H.B., D.D.L.B.

Investigation: A.N., H.M., B.C., P.A.C.

Methodology: A.N., A.H.B., J.T.C.

Project administration: A.H.B., P.A.C, D.D.L.B.

Resources: M.C.C., E.S.H., S.J.L., E.W.W., S.F., N.T., R.S., T.W., A.L.C., P.M., A.H.B., D.D.L.B.

Software: J.T.C., A.N., H.M., L.P.

Supervision: A.H.B., D.D.L.B.

Visualization: A.N., H.M.

Writing – original draft: A.N., P.A.C., B.C.

Writing – review & editing: A.H.B., D.D.L.B.

Declarations

Samples were collected with proper consent and ethical compliance under IRB # 07047 & 17334 (City of Hope), 41030 and 89989 (University of Utah), or HREC # 01/60, 16/161 by the Australian Ovarian Cancer Study (AOCS) which were analyzed under HREC # 15/84 (Peter MacCallum Cancer Centre).

Competing interests: The authors declare no competing interests

References

1. J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional Addiction in Cancer. *Cell*. **168**, 629–643 (2017).
2. R. Cress, Y. Chen, C. Morris, M. Petersen, G. Leiserowitz, Characteristics of Long-Term Survivors of Epithelial Ovarian Cancer. *Obstetrics & Gynecology*. **126**, 491–497 (2015).
3. K. Moore, N. Colombo, G. Scambia, B.-G. Kim, A. Oaknin, M. Friedlander, A. Lisysanskaya, A. Floquet, A. Leary, G. S. Sonke, C. Gourley, S. Banerjee, A. Oza, A. González-Martín, C. Aghajanian, W. Bradley, C. Mathews, J. Liu, E. S. Lowe, R. Bloomfield, P. DiSilvestro, Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med*. **379**, 2495–2505 (2018).
4. D. D. Bowtell, S. Böhm, A. A. Ahmed, P.-J. Aspúria, R. C. Bast, V. Beral, J. S. Berek, M. J. Birrer, S. Blagden, M. A. Bookman, J. D. Brenton, K. B. Chiappinelli, F. C. Martins, G. Coukos, R. Drapkin, R. Edmondson, C. Fotopoulou, H. Gabra, J. Galon, C. Gourley, V. Heong, D. G. Huntsman, M. Iwanicki, B. Y. Karlan, A. Kaye, E. Lengyel, D. A. Levine, K. H. Lu, I. A. McNeish, U. Menon, S. A. Narod, B. H. Nelson, K. P. Nephew, P. Pharoah, D. J. Powell, P. Ramos, I. L. Romero, C. L. Scott, A. K. Sood, E. A. Stronach, F. R. Balkwill, Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat. Rev. Cancer*. **15**, 668–679 (2015).
5. S. L. Cooke, J. D. Brenton, Evolution of platinum resistance in high-grade serous ovarian cancer. *The Lancet Oncology*. **12**, 1169–1174 (2011).
6. M.-C. W. Lee, F. J. Lopez-Diaz, S. Y. Khan, M. A. Tariq, Y. Dayn, C. J. Vaske, A. J. Radenbaugh, H. J. Kim, B. M. Emerson, N. Pourmand, Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proceedings of the National Academy of Sciences*. **111**, E4726–E4735 (2014).

7. J. Fan, H.-O. Lee, S. Lee, D. Ryu, S. Lee, C. Xue, S. J. Kim, K. Kim, N. Barkas, P. J. Park, W.-Y. Park, P. V. Kharchenko, Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* **28**, 1217–1227 (2018).
8. J. S. Berek, C. Crum, M. Friedlander, Cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynecology & Obstetrics.* **119**, S118–S129 (2012).
9. H. Zhang, T. Liu, Z. Zhang, S. H. Payne, B. Zhang, J. E. McDermott, J.-Y. Zhou, V. A. Petyuk, L. Chen, D. Ray, S. Sun, F. Yang, L. Chen, J. Wang, P. Shah, S. W. Cha, P. Aiyetan, S. Woo, Y. Tian, M. A. Gritsenko, T. R. Clauss, C. Choi, M. E. Monroe, S. Thomas, S. Nie, C. Wu, R. J. Moore, K.-H. Yu, D. L. Tabb, D. Fenyö, V. Bafna, Y. Wang, H. Rodriguez, E. S. Boja, T. Hiltke, R. C. Rivers, L. Sokoll, H. Zhu, I.-M. Shih, L. Cope, A. Pandey, B. Zhang, M. P. Snyder, D. A. Levine, R. D. Smith, D. W. Chan, K. D. Rodland, S. A. Carr, M. A. Gillette, K. R. Klauser, E. Kuhn, D. R. Mani, P. Mertins, K. A. Ketchum, R. Thangudu, S. Cai, M. Oberti, A. G. Paulovich, J. R. Whiteaker, N. J. Edwards, P. B. McGarvey, S. Madhavan, P. Wang, D. W. Chan, A. Pandey, I.-M. Shih, H. Zhang, Z. Zhang, H. Zhu, L. Cope, G. A. Whiteley, S. J. Skates, F. M. White, D. A. Levine, E. S. Boja, C. R. Kinsinger, T. Hiltke, M. Mesri, R. C. Rivers, H. Rodriguez, K. M. Shaw, S. E. Stein, D. Fenyö, T. Liu, J. E. McDermott, S. H. Payne, K. D. Rodland, R. D. Smith, P. Rudnick, M. Snyder, Y. Zhao, X. Chen, D. F. Ransohoff, A. N. Hoofnagle, D. C. Liebler, M. E. Sanders, Z. Shi, R. J. C. Slebos, D. L. Tabb, B. Zhang, L. J. Zimmerman, Y. Wang, S. R. Davies, L. Ding, M. J. C. Ellis, R. R. Townsend, Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell.* **166**, 755–765 (2016).
10. A. A. Ahmed, D. Etemadmoghadam, J. Temple, A. G. Lynch, M. Riad, R. Sharma, C. Stewart, S. Fereday, C. Caldas, A. deFazio, D. Bowtell, J. D. Brenton, Driver mutations in *TP53* are ubiquitous in high grade serous carcinoma of the ovary: *TP53* mutation in high-grade pelvic serous carcinoma. *J. Pathol.* **221**, 49–56 (2010).
11. P. T. Kroeger, R. Drapkin, Pathogenesis and heterogeneity of ovarian cancer: *Current Opinion in Obstetrics and Gynecology.* **29**, 26–34 (2017).
12. The Australian Ovarian Cancer Study Group, A.-M. Patch, E. L. Christie, D. Etemadmoghadam, D. W. Garsed, J. George, S. Fereday, K. Nones, P. Cowin, K. Alsop, P. J. Bailey, K. S. Kassahn, F. Newell, M. C. J. Quinn, S. Kazakoff, K. Quek, C. Wilhelm-Benartzi, E. Curry, H. S. Leong, A. Hamilton, L. Mileskin, G. Au-Yeung, C. Kennedy, J. Hung, Y.-E. Chiew, P. Harnett, M. Friedlander, M. Quinn, J. Pyman, S. Corder, P. O'Brien, J. Leditschke, G. Young, K. Strachan, P. Waring, W. Azar, C. Mitchell, N. Traficante, J. Hendley, H. Thorne, M. Shackleton, D. K. Miller, G. M. Arnau, R. W. Tohill, T. P. Holloway, T. Semple, I. Harliwong, C. Nourse, E. Nourbakhsh, S. Manning, S. Idrisoglu, T. J. C. Bruxner, A. N. Christ, B. Poudel, O. Holmes, M. Anderson, C. Leonard, A. Lonie, N. Hall, S. Wood, D. F. Taylor, Q. Xu, J. L. Fink, N. Waddell, R. Drapkin, E. Stronach, H. Gabra, R. Brown, A. Jewell, S. H. Nagaraj, E. Markham, P. J. Wilson, J. Ellul, O. McNally, M. A. Doyle, R. Vedururu, C. Stewart, E. Lengyel, J. V. Pearson, N. Waddell, A. deFazio, S. M. Grimmond, D. D. L. Bowtell, Whole-genome characterization of chemoresistant ovarian cancer. *Nature.* **521**, 489–494 (2015).
13. A. E. Freimund, J. A. Beach, E. L. Christie, D. D. L. Bowtell, Mechanisms of Drug Resistance in High-Grade Serous Ovarian Cancer. *Hematology/Oncology Clinics of North America.* **32**, 983–996 (2018).

14. E. L. Christie, S. Pattnaik, J. Beach, A. Copeland, N. Rashoo, S. Fereday, J. Hendley, K. Alsop, S. L. Brady, G. Lamb, A. Pandey, A. deFazio, H. Thorne, A. Bild, D. D. L. Bowtell, Multiple ABCB1 transcriptional fusions in drug resistant high-grade serous ovarian and breast cancer. *Nat Commun.* **10**, 1295 (2019).
15. D. A. Lawson, K. Kessenbrock, R. T. Davis, N. Pervolarakis, Z. Werb, Tumour heterogeneity and metastasis at single-cell resolution. *Nat Cell Biol.* **20**, 1349–1360 (2018).
16. 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. Suva, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* **344**, 1396–1401 (2014).
17. W. Chung, H. H. Eum, H.-O. Lee, K.-M. Lee, H.-B. Lee, K.-T. Kim, H. S. Ryu, S. Kim, J. E. Lee, Y. H. Park, Z. Kan, W. Han, W.-Y. Park, Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun.* **8**, 15081 (2017).
18. S. W. Brady, J. A. McQuerry, Y. Qiao, S. R. Piccolo, G. Shrestha, D. F. Jenkins, R. M. Layer, B. S. Pedersen, R. H. Miller, A. Esch, S. R. Selitsky, J. S. Parker, L. A. Anderson, B. K. Dalley, R. E. Factor, C. B. Reddy, J. P. Boltax, D. Y. Li, P. J. Moos, J. W. Gray, L. M. Heiser, S. S. Buys, A. L. Cohen, W. E. Johnson, A. R. Quinlan, G. Marth, T. L. Werner, A. H. Bild, Combating subclonal evolution of resistant cancer phenotypes. *Nat Commun.* **8**, 1231 (2017).
19. C. A. Aktipis, A. M. Boddy, R. A. Gatenby, J. S. Brown, C. C. Maley, Life history trade-offs in cancer evolution. *Nat Rev Cancer.* **13**, 883–892 (2013).
20. J. Hausser, U. Alon, Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nat Rev Cancer.* **20**, 247–257 (2020).
21. O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, U. Alon, Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. *Science.* **336**, 1157–1160 (2012).
22. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* **36**, 411–420 (2018).
23. D. Aran, A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte, M. Bhattacharya, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* **20**, 163–172 (2019).
24. Y. Hart, H. Sheftel, J. Hausser, P. Szekely, N. B. Ben-Moshe, Y. Korem, A. Tendler, A. E. Mayo, U. Alon, Inferring biological tasks using Pareto analysis of high-dimensional data. *Nat Methods.* **12**, 233–235 (2015).
25. C. Tristan, N. Shahani, T. W. Sedlak, A. Sawa, The diverse functions of GAPDH: Views from different subcellular compartments. *Cellular Signalling.* **23**, 317–323 (2011).
26. F. Castro, A. P. Cardoso, R. M. Gonçalves, K. Serre, M. J. Oliveira, Interferon-Gamma at the Crossroads of Tumor Immune Surveillance or Evasion. *Front. Immunol.* **9**, 847 (2018).

27. A. Majoros, E. Platanitis, E. Kernbauer-Hölzl, F. Rosebrock, M. Müller, T. Decker, Canonical and Non-Canonical Aspects of JAK–STAT Signaling: Lessons from Interferons for Cytokine Responses. *Front. Immunol.* **8** (2017), doi:10.3389/fimmu.2017.00029.
28. D. Rosner, V. Stoneman, T. Littlewood, N. McCarthy, N. Figg, Y. Wang, G. Tellides, M. Bennett, Interferon- γ Induces Fas Trafficking and Sensitization to Apoptosis in Vascular Smooth Muscle Cells via a PI3K- and Akt-Dependent Mechanism. *The American Journal of Pathology.* **168**, 2054–2063 (2006).
29. X. Su, Y. Yu, Y. Zhong, E. G. Giannopoulou, X. Hu, H. Liu, J. R. Cross, G. Rättsch, C. M. Rice, L. B. Ivashkiv, Interferon- γ regulates cellular metabolism and mRNA translation to potentiate macrophage activation. *Nat Immunol.* **16**, 838–849 (2015).
30. B. N. Chau, T.-T. Chen, Y. Y. Wan, J. DeGregori, J. Y. J. Wang, Tumor Necrosis Factor Alpha-Induced Apoptosis Requires p73 and c-ABL Activation Downstream of RB Degradation. *MCB.* **24**, 4438–4447 (2004).
31. A. Marechal, L. Zou, DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harbor Perspectives in Biology.* **5**, a012716–a012716 (2013).
32. Z. Nikitaki, C. E. Hellweg, A. G. Georgakilas, J.-L. Ravanat, Stress-induced DNA damage biomarkers: applications and limitations. *Front. Chem.* **3** (2015), doi:10.3389/fchem.2015.00035.
33. V. V. Baker, M. P. Borst, D. Dixon, K. D. Hatch, H. M. Shingleton, D. Miller, c-myc amplification in ovarian cancer. *Gynecologic Oncology.* **38**, 340–342 (1990).
34. J. M. Reyes-Gonzalez, G. N. Armaiz-Pena, L. S. Mangala, F. Valiyeva, C. Ivan, S. Pradeep, I. M. Echevarria-Vargas, A. Rivera-Reyes, A. K. Sood, P. E. Vivas-Mejia, Targeting c-MYC in Platinum-Resistant Ovarian Cancer. *Molecular Cancer Therapeutics.* **14**, 2260–2269 (2015).
35. X. Liu, H. Ye, L. Li, W. Li, Y. Zhang, J.-Y. Zhang, Humoral Autoimmune Responses to Insulin-Like Growth Factor II mRNA-Binding Proteins IMP1 and p62/IMP2 in Ovarian Cancer. *Journal of Immunology Research.* **2014**, 1–7 (2014).
36. I. G. Campbell, S. E. Russell, D. Y. H. Choong, K. G. Montgomery, M. L. Ciavarella, C. S. F. Hooi, B. E. Cristiano, R. B. Pearson, W. A. Phillips, Mutation of the *PIK3CA* Gene in Ovarian and Breast Cancer. *Cancer Res.* **64**, 7678–7681 (2004).
37. N. Iqbal, N. Iqbal, Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Molecular Biology International.* **2014**, 1–9 (2014).
38. Y. Wen, Y. Hou, Z. Huang, J. Cai, Z. Wang, SOX2 is required to maintain cancer stem cells in ovarian cancer. *Cancer Sci.* **108**, 719–731 (2017).
39. S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, B. E. Bernstein, Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell.* **171**, 1611-1624.e24 (2017).
40. Tumor metabolism regulating chemosensitivity in ovarian cancer. *genesandcancer* (2018), doi:10.18632/genesandcancer.176.

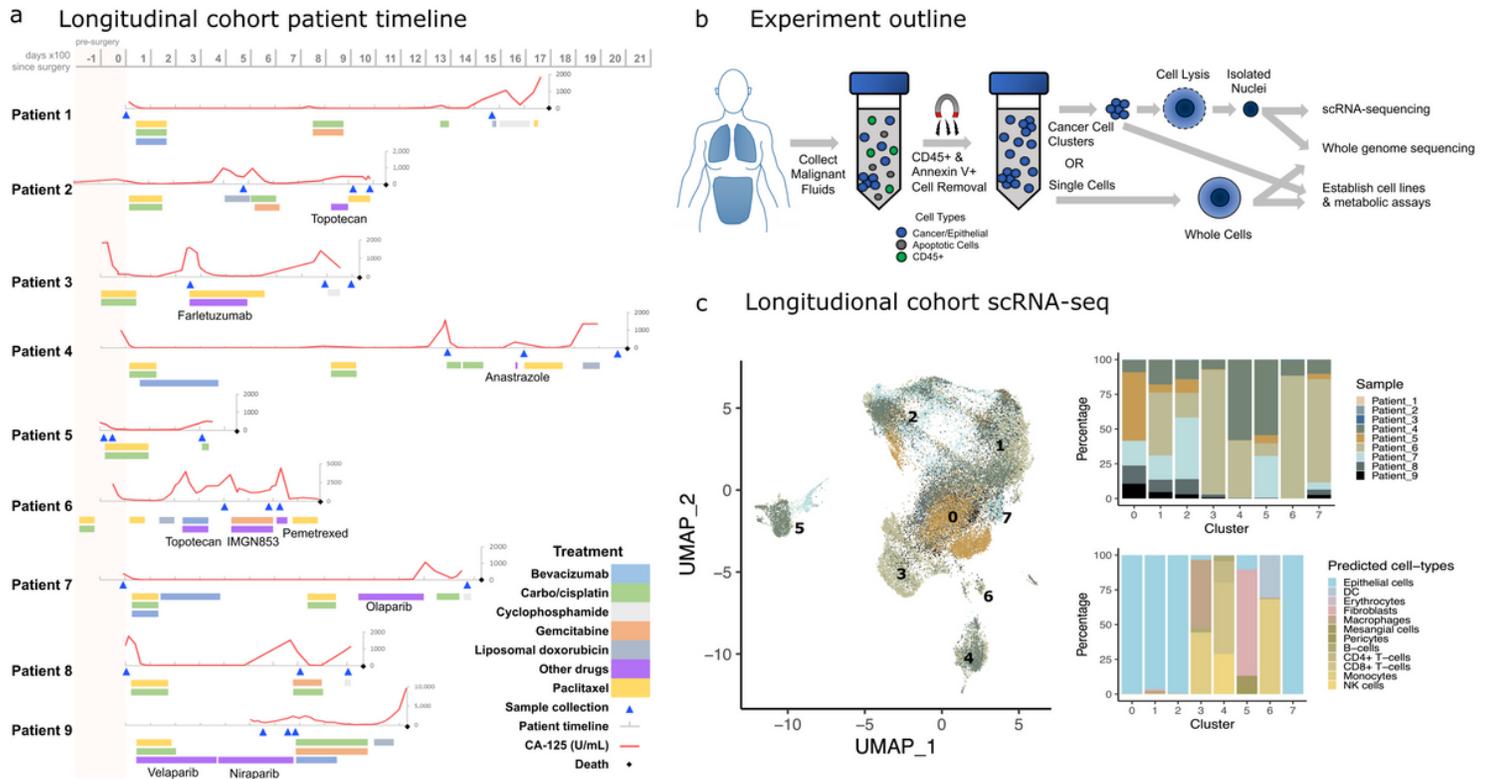
41. S. Dar, J. Chhina, I. Mert, D. Chitale, T. Buekers, H. Kaur, S. Giri, A. Munkarah, R. Rattan, Bioenergetic Adaptations in Chemoresistant Ovarian Cancer Cells. *Sci Rep.* **7**, 8760 (2017).
42. C. V. Dang, MYC, microRNAs and glutamine addiction in cancers. *Cell Cycle.* **8**, 3243–3245 (2009).
43. H. Qiu, A. L. Jackson, J. E. Kilgore, Y. Zhong, L. L.-Y. Chan, P. A. Gehrig, C. Zhou, V. L. Bae-Jump, JQ1 suppresses tumor growth through downregulating LDHA in ovarian cancer*. *Oncotarget.* **6**, 6915–6930 (2015).
44. T. Bagratuni, N. Mavrianou, N. G. Gavalas, K. Tzannis, C. Arapinis, M. Liontos, M. I. Christodoulou, N. Thomakos, D. Haidopoulos, A. Rodolakis, E. Kastritis, A. Scorilas, M. A. Dimopoulos, A. Bamias, JQ1 inhibits tumour growth in combination with cisplatin and suppresses JAK/STAT signalling pathway in ovarian cancer. *European Journal of Cancer.* **126**, 125–135 (2020).
45. N. Loret, H. Denys, P. Tummers, G. Berx, The Role of Epithelial-to-Mesenchymal Plasticity in Ovarian Cancer Progression and Therapy Resistance. *Cancers.* **11**, 838 (2019).
46. C. L. Chaffer, B. P. San Juan, E. Lim, R. A. Weinberg, EMT, cell plasticity and metastasis. *Cancer Metastasis Rev.* **35**, 645–654 (2016).
47. M. Sciacovelli, C. Frezza, Metabolic reprogramming and epithelial-to-mesenchymal transition in cancer. *FEBS J.* **284**, 3132–3144 (2017).
48. T. Kiesslich, M. Pichler, D. Neureiter, Epigenetic control of epithelial-mesenchymal-transition in human cancer. *Molecular and Clinical Oncology.* **1**, 3–11 (2013).
49. L. Li, W. Li, Epithelial–mesenchymal transition in human cancer: Comprehensive reprogramming of metabolism, epigenetics, and differentiation. *Pharmacology & Therapeutics.* **150**, 33–46 (2015).
50. W. L. Tam, R. A. Weinberg, The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med.* **19**, 1438–1449 (2013).
51. B. R. Muys, J. F. Sousa, J. R. Praça, L. F. de Araújo, A. A. Sarshad, D. G. Anastasakis, X. Wang, X. L. Li, G. A. de Molfetta, A. Ramão, A. Lal, D. O. Vidal, M. Hafner, W. A. Silva, miR-450a Acts as a Tumor Suppressor in Ovarian Cancer by Regulating Energy Metabolism. *Cancer Res.* **79**, 3294–3305 (2019).
52. S. Pareek, A. Nath, R. S. Huang, MicroRNA targeting energy metabolism in ovarian cancer: a potent contender for future therapeutics. *Ann Transl Med.* **7**, S299–S299 (2019).
53. L. J. Luquette, C. L. Bohrson, M. A. Sherman, P. J. Park, Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat Commun.* **10**, 3908 (2019).
54. A. Brock, H. Chang, S. Huang, Non-genetic heterogeneity – a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet.* **10**, 336–342 (2009).
55. N. Wang, J. Zheng, Z. Chen, Y. Liu, B. Dura, M. Kwak, J. Xavier-Ferrucio, Y.-C. Lu, M. Zhang, C. Roden, J. Cheng, D. S. Krause, Y. Ding, R. Fan, J. Lu, Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat Commun.* **10**, 95 (2019).
56. X. Chen, J. T. Chang, Planning bioinformatics workflows using an expert system. *Bioinformatics*, btw817 (2017).

57. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
58. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).
59. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888-1902.e21 (2019).
60. The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).
61. N. A. Mabbott, J. Baillie, H. Brown, T. C. Freeman, D. A. Hume, An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*. **14**, 632 (2013).
62. D. Risso, F. Perraudeau, S. Gribkova, S. Dudoit, J.-P. Vert, A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun*. **9**, 284 (2018).
63. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*. **1**, 417–425 (2015).
64. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. **27**, 1739–1740 (2011).
65. S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. **14**, 7 (2013).
66. F. Favero, T. Joshi, A. M. Marquard, N. J. Birkbak, M. Krzystanek, Q. Li, Z. Szallasi, A. C. Eklund, Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology*. **26**, 64–70 (2015).
67. R. Shen, V. E. Seshan, FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. **44**, e131–e131 (2016).
68. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly*. **6**, 80–92 (2012).
69. S. Kim, K. Scheffler, A. L. Halpern, M. A. Bekritsky, E. Noh, M. Källberg, X. Chen, Y. Kim, D. Beyter, P. Krusche, C. T. Saunders, Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. **15**, 591–594 (2018).
70. K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. **31**, 213–219 (2013).
71. Y. Fan, L. Xi, D. S. T. Hughes, J. Zhang, J. Zhang, P. A. Futreal, D. A. Wheeler, W. Wang, MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol*. **17**, 178 (2016).
72. Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, S. A. Forbes, The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. **18**, 696–705

(2018).

73. A. Gonzalez-Perez, C. Perez-Llamas, J. Deu-Pons, D. Tamborero, M. P. Schroeder, A. Jene-Sanz, A. Santos, N. Lopez-Bigas, IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*. **10**, 1081–1082 (2013).
74. J. A. Wala, P. Bandopadhyay, N. F. Greenwald, R. O'Rourke, T. Sharpe, C. Stewart, S. Schumacher, Y. Li, J. Weischenfeldt, X. Yao, C. Nusbaum, P. Campbell, G. Getz, M. Meyerson, C.-Z. Zhang, M. Imielinski, R. Beroukhim, SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res*. **28**, 581–591 (2018).
75. H. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*. **14**, 244 (2013).
76. M. Mørup, L. K. Hansen, Archetypal analysis for machine learning and data mining. *Neurocomputing*. **80**, 54–63 (2012).
77. C. A. Miller, J. McMichael, H. X. Dang, C. A. Maher, L. Ding, T. J. Ley, E. R. Mardis, R. K. Wilson, Visualizing tumor evolution with the fishplot package for R. *BMC Genomics*. **17**, 880 (2016).
78. J. H. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. **33**, 1–22 (2010).

Figures



Longitudinal cohort sample collection and scRNA-seq analysis a. Timeline of patients included in the study. Grey timeline shows days x100, with 0 referring to the day primary surgery was performed. The red lines indicate CA-125 levels, while the triangles along the timelines indicate the time points at which malignant fluid samples were obtained. Colored bars below the patient timeline indicates drug treatment received. b. The malignant fluid samples were processed to remove immune and apoptotic cells, and processed for whole-genome sequencing, in vitro metabolic assays, and scRNA-seq. c. Dimensionality reduction (UMAP) projections and clustering of the integrated, approximately 37,000 high-quality cells profiled using iCell8 (patients 1-3) or 10X (patients 4-9) scRNA-seq platforms following CCA normalization. The numbers indicate clusters obtained following unsupervised clustering. Stacked bar plots on the right showing the distribution of various predicted cell types across clusters and the diversity of samples distributed within each cluster.

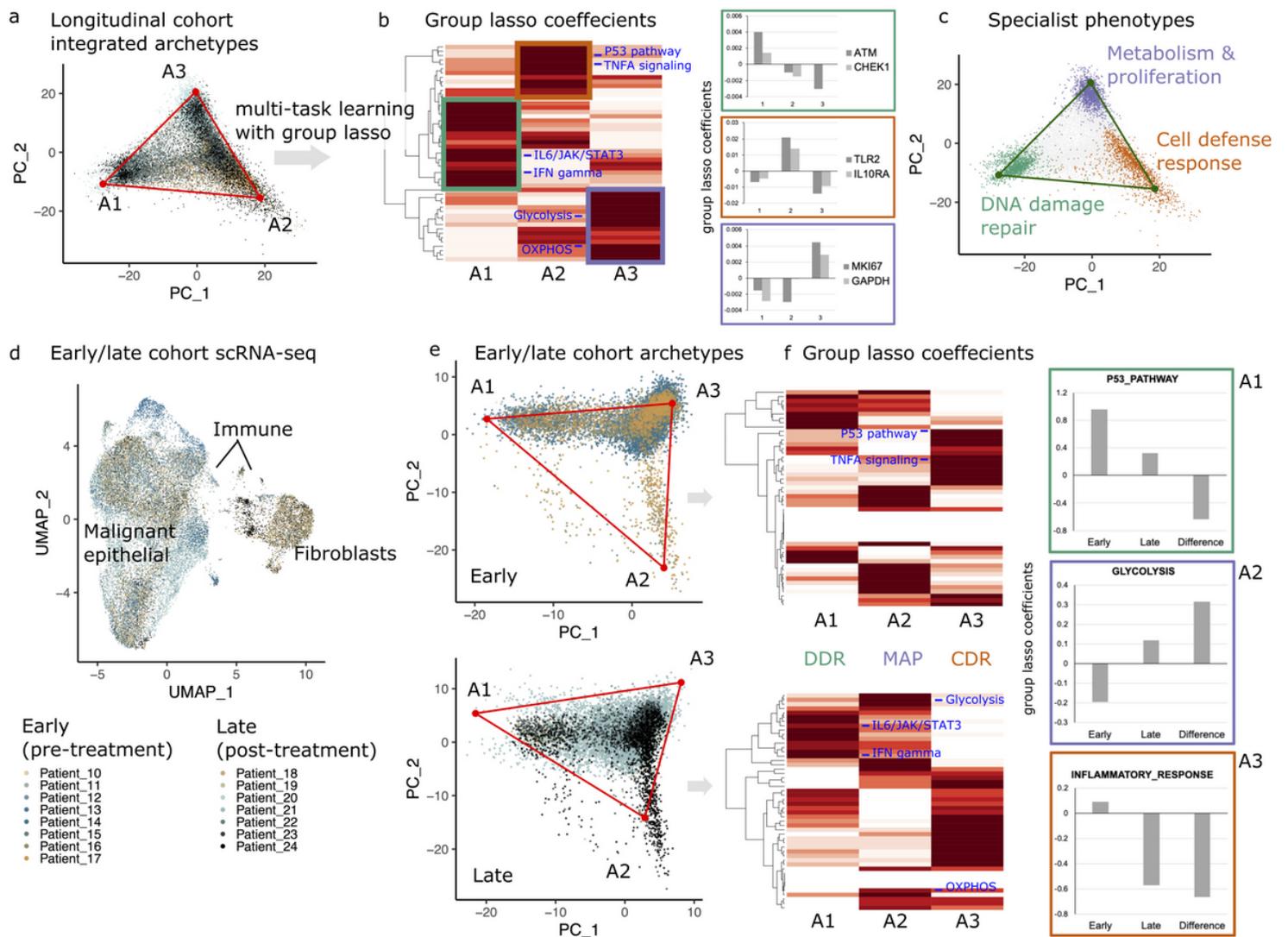


Figure 2

Archetype analysis to investigate shifts in biological tasks over time. Cells from the integrated longitudinal cohort are projected on the first two principal components of scRNA-seq data. Each vertex identified using the Pareto inference algorithm represents an archetype specializing in a specific

biological task b. Heatmap showing the hierarchical clustering of group-lasso coefficients for various hallmark pathways associated with each archetype. The three colored boxes display three hierarchical clusters of related phenotypes that reveal the identity of the biological task associated with each archetype. The principal pathways that define the archetypes are indicated on the heatmap in blue, with the green box indicating DNA damage repair (DDR), the red box indicating cellular defense response (CDR), and the blue box indicating metabolism and proliferation (MAP) archetypes. The inset bar plots show the group lasso coefficient of key genes corresponding to the three archetypes from the scRNA-seq data of the longitudinal cohort patients c. Three major biologicals tasks were predicted using the multi-task learning approach. Specialist cells are colored according to the closest archetype (cells in the 95th percentile), whereas non-specialists distal from all three archetypes are shown in grey. d. Dimensionality reduction (UMAP) projections and clustering of the integrated 27.7K high-quality cells, including 21.5K malignant epithelial cells, from the early/late (pre-/post-treatment) cohort profiled using 10X platforms followed by CCA normalization. Patients 10-17 were treatment -naïve while patients 18-24 received multiple lines of treatment. Major predicted cell types are annotated to show separate clusters with malignant epithelial cells, immune cells, and fibroblasts. e. Archetypes annotated on the first two principal components of the early/late cohorts. The samples from each cohort are shown separately to clearly display the population shifts in the archetypes. f. Hierarchical clustering of group-lasso coefficients for archetypes determined for early and late cohort patients, with principal pathways defining the archetypes indicated in blue. The inset bar plots show group lasso coefficients of key pathways and their difference between the pre-/post-treatment early and late cohorts, representing the shift in the archetype populations between the two cohorts.

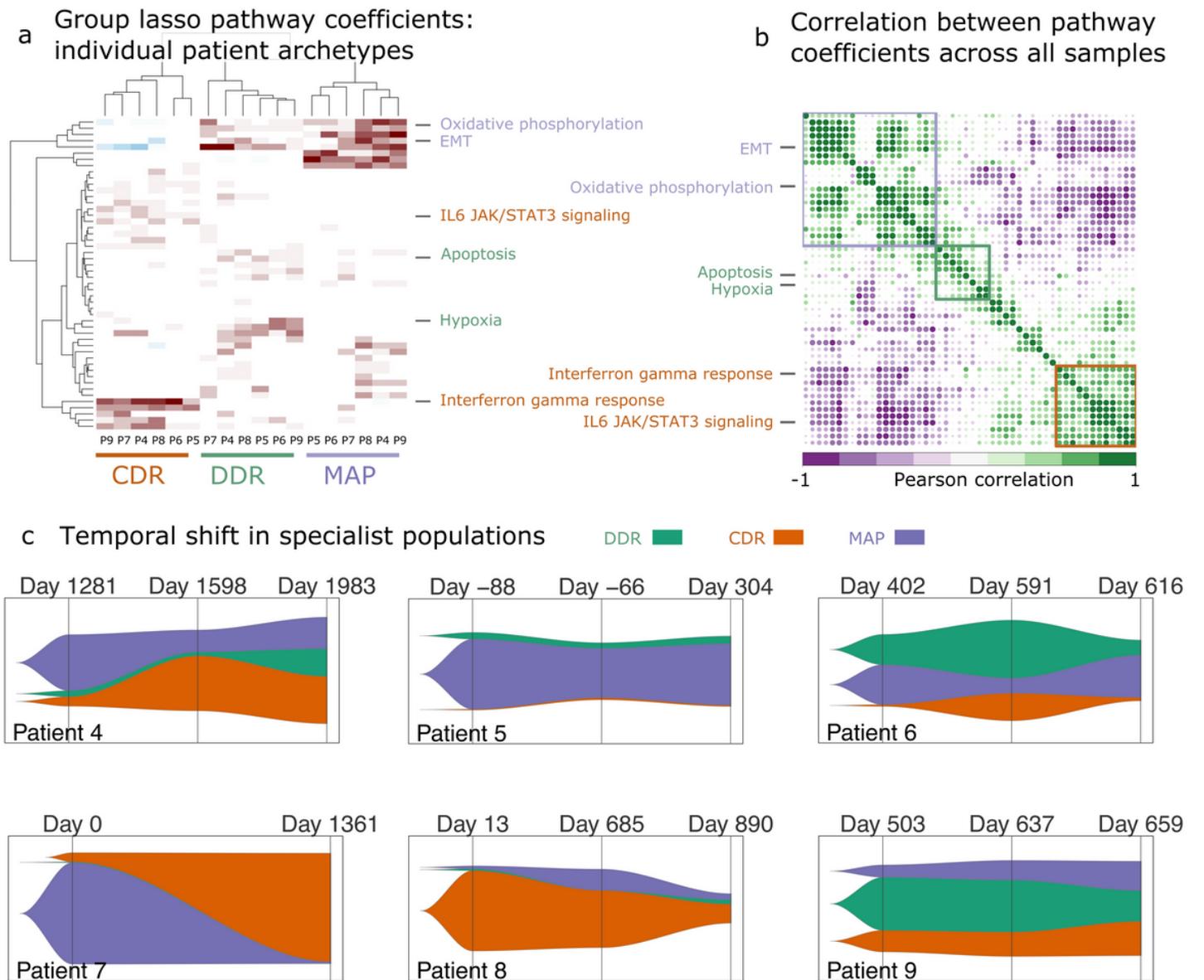


Figure 3

Archetype evolution in individual patients. a. Hierarchical clustering of group-lasso coefficients for archetypes determined for each longitudinal cohort patient across all time points. b. Correlation plot of pathway scores across all samples with three major clusters indicated using colored boxes. Key pathways that defined the biological phenotypes associated with the archetypes in the integrated dataset are shown for both the hierarchical clusters and the correlation plots c. Fishplots displaying the temporal shift in the population of task specialists for each patient. The three major tasks are represented with their respected abbreviations (DDR: DNA damage repair; MAP: metabolism and proliferation; CDR: cell defense response)

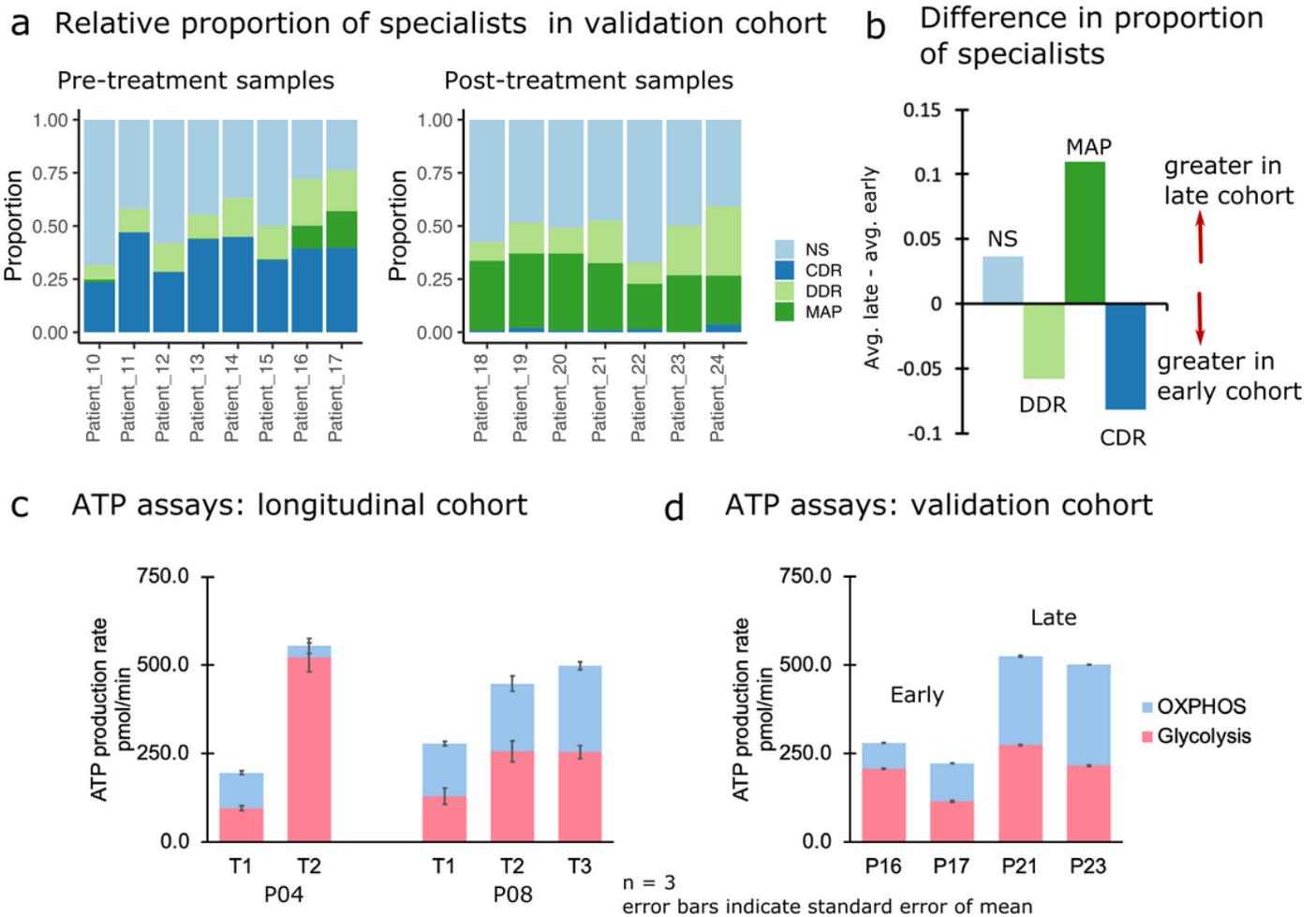


Figure 4

Metabolic shift and plasticity a. Stacked bar plots showing the proportion of specialists at biological tasks and non-specialists in the early/late patient cohorts b. Change in the proportion of task specialists and non-specialists between the early and late cohorts c and d. ATP assays of cell lines developed using the malignant cells isolated from patients 4 and 8 (c) and patients 16, 17, 20, and 23 (d).

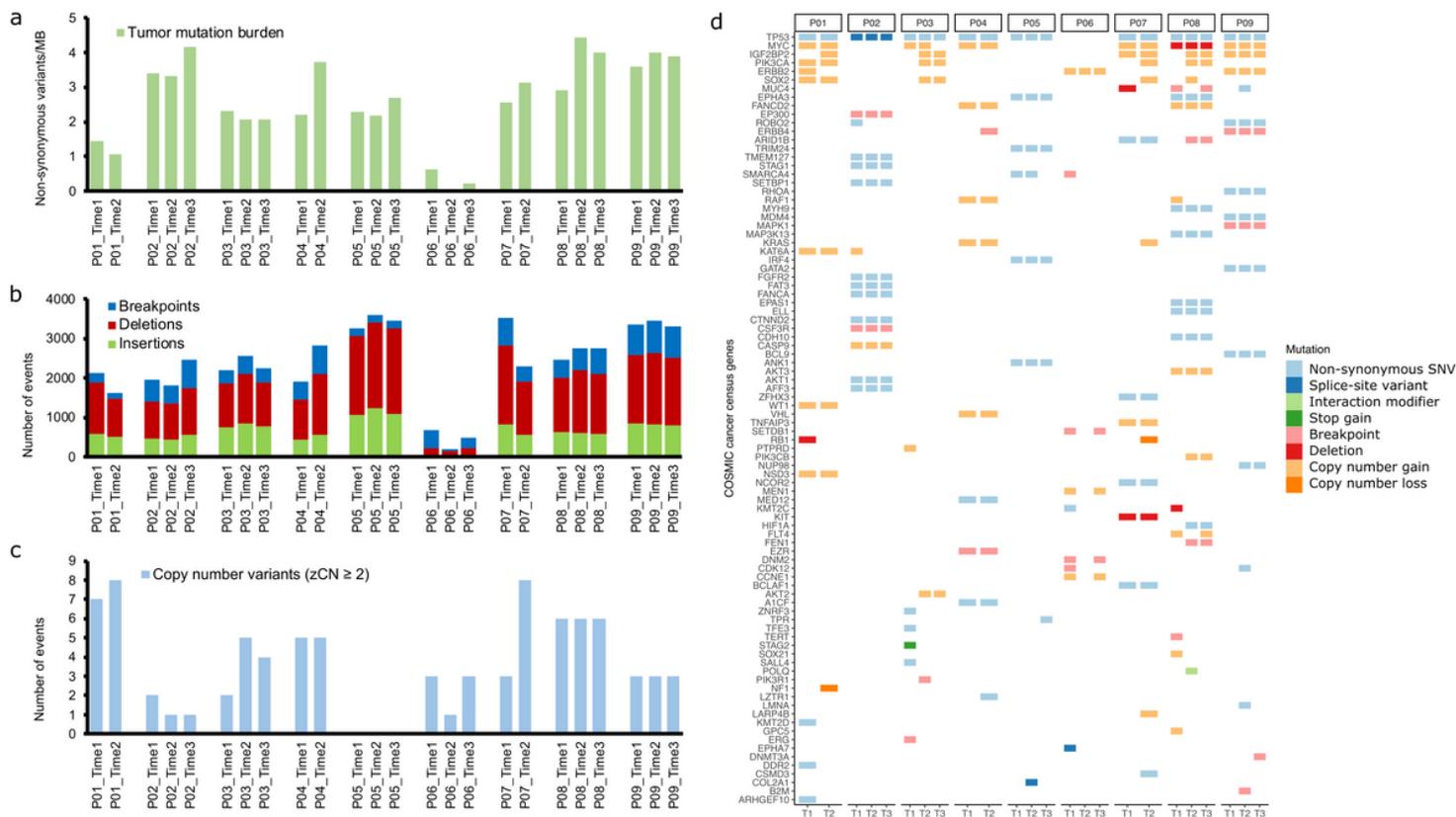


Figure 5

Overview of somatic WGS variants in the longitudinal cohort. a. Tumor mutation burden (non-synonymous SNVs and small indels/MB) in the longitudinal cohort samples. b. Stacked bar plots showing the frequency of SVs, including indels (> 25bp) and breakpoints in each sample. c. The number of genes with copy number alterations in each sample. The plot displays number of genes that show a z-transformed copy number ≥ 2 (gain) or ≤ -2 (loss) d. OncoPrint displaying somatic variants affecting COSMIC cancer census genes. The plot displays SNVs and small indels with VAF > 0.05, large indels and breakpoints affecting exons, and CNVs with z-transformed copy number ≥ 2 (gain) or ≤ -2 (loss).

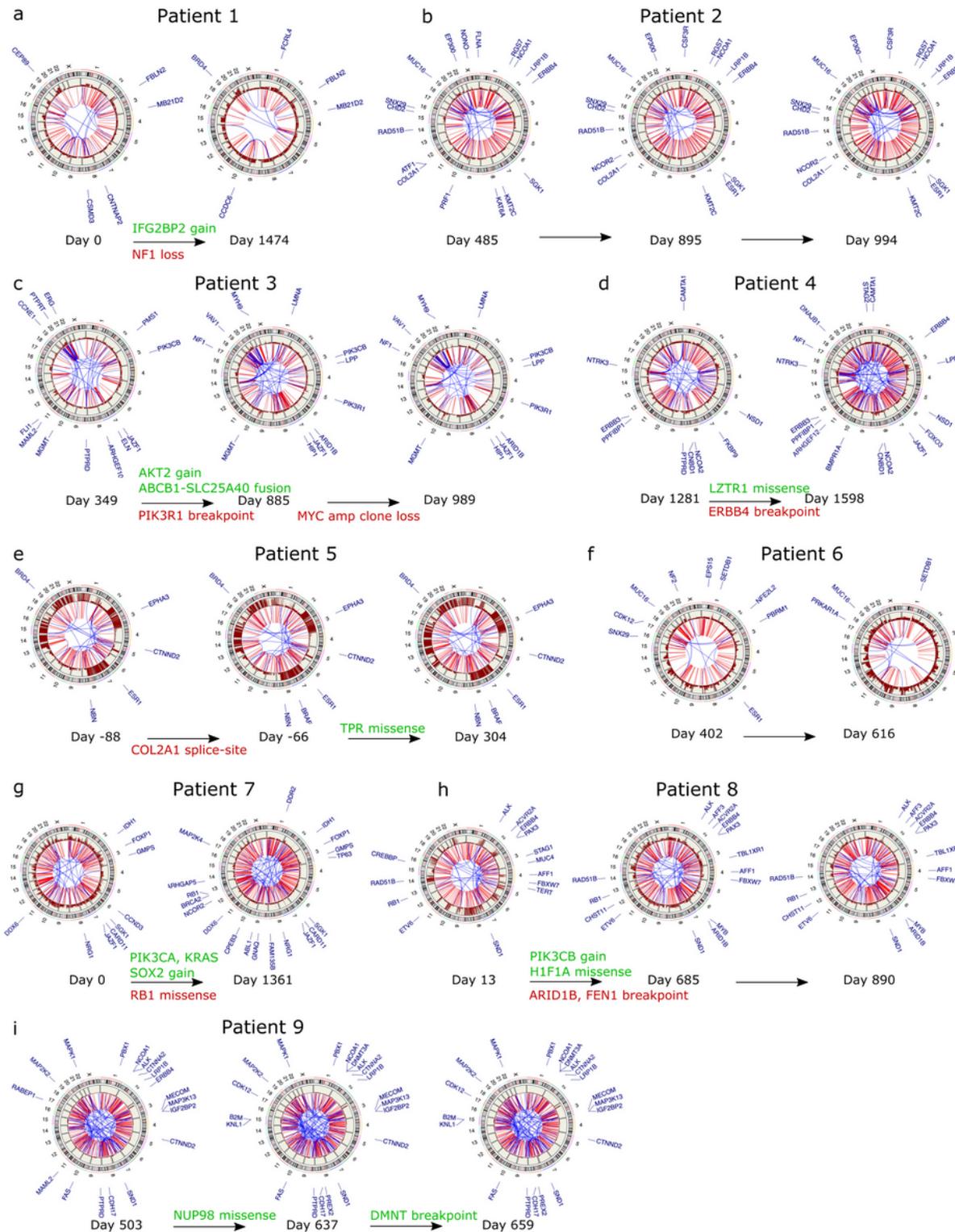


Figure 6

Temporal genomic characteristics of ovarian cancer. a-i Circos plots displaying breakpoints and CNVs along chromosomes. The breakpoints are displayed in blue, if within chromosomes, or red, if between chromosomes. Copy number gains along the outer track are shown in dark red, with losses indicated in blue and neutral copies in dark grey. The annotations at the periphery of the plots for the first time point of each patient indicates CNVs and SVs (breakpoints) affecting the exons of COSMIC cancer census

genes. The plots for the subsequent time points are annotated to show only acquired CNVs and SVs, with truncal breakpoints shown in light grey.

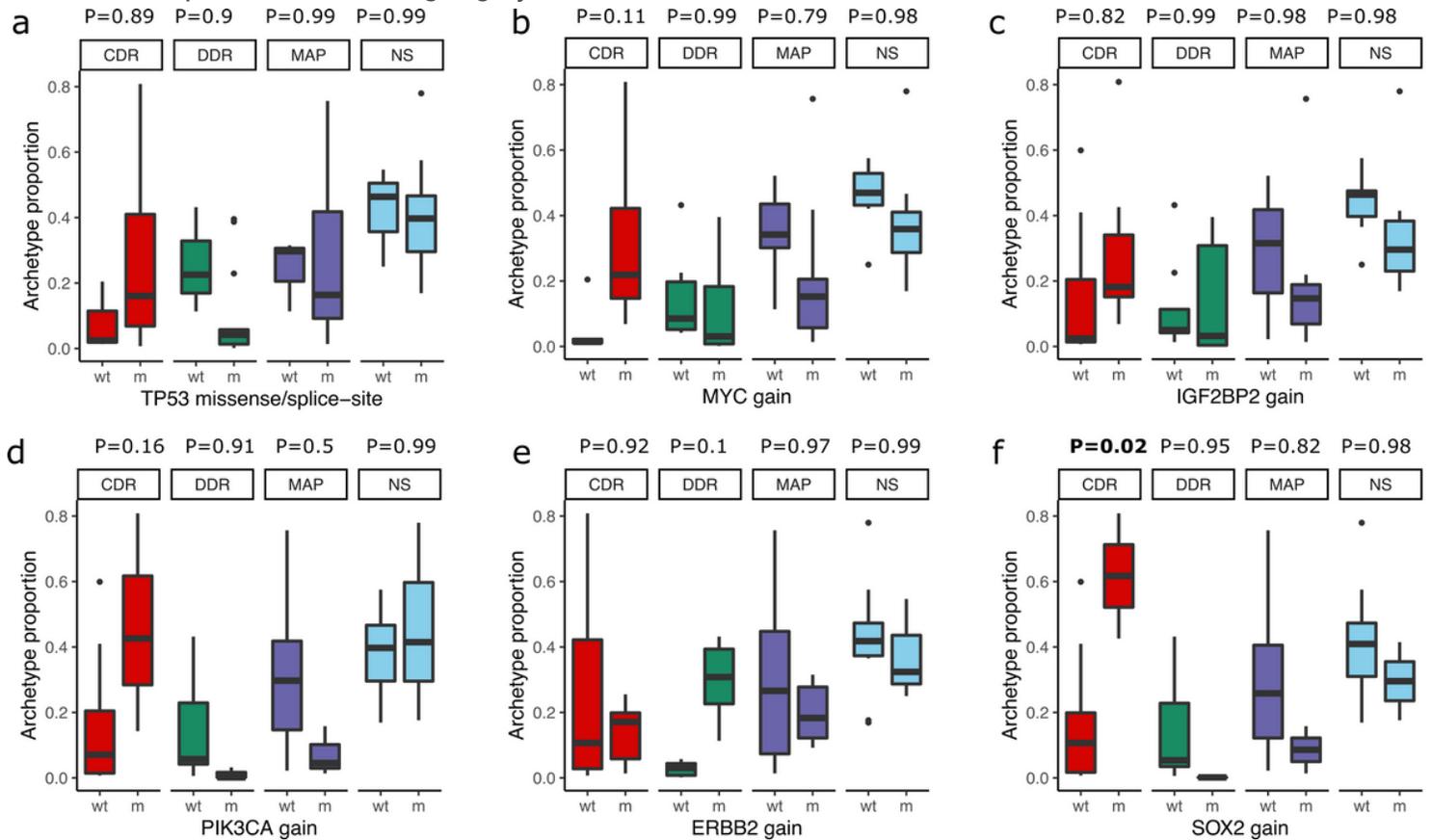


Figure 7

Archetype proportions in samples grouped by mutations in key genes. a-f. Boxplots showing the distribution of the proportion of each archetype specialist grouped by mutations that are considered pathogenic and appeared in at least 25% of the samples (6 out of 24). The boxes are colored by archetype with red indicating CDR, green showing CDR, dark blue indicating MAP and light blue showing non-specialists. The p-value annotations above each archetype label indicates adjusted pairwise p-values from Tukey post-hoc analysis of the ANOVA model of the proportion of specialists with mutation status as interaction term. The “wt” group label indicates samples with wild-type (or copy-neutral) gene or while “m” represents samples with mutated (or amplified) gene. Significant p-values are indicated in bold.

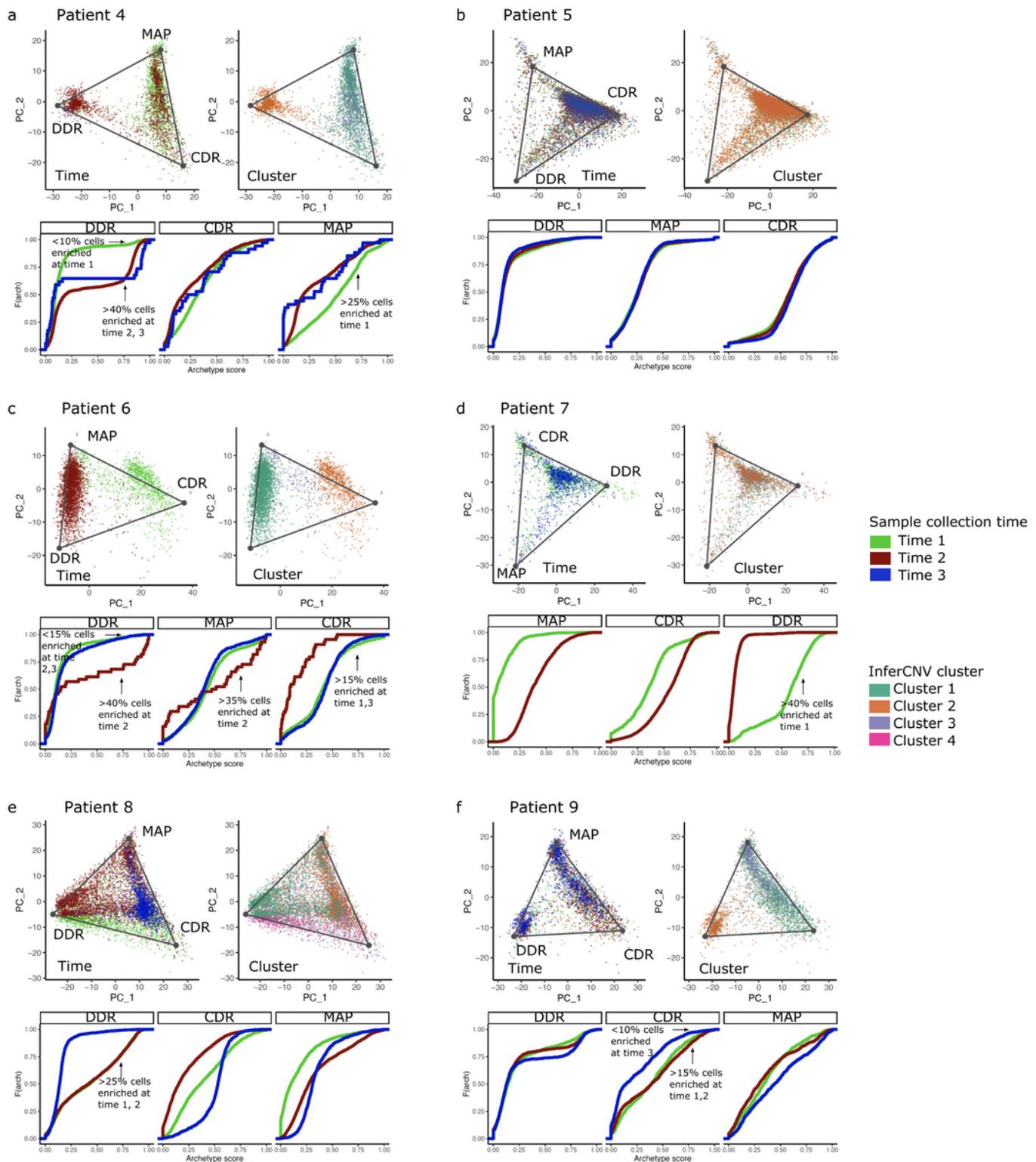


Figure 8

Mapping archetypes to time and inferred subclones. a-f. The top two panels display archetypes on the first two principal components of each longitudinal cohort patient. The panel on the left shows archetype projections on cells colored based on sample collection time, while the right panel shows cells colored based on clusters identified using the inferred CNV algorithm. The empirical cumulative distribution

functions of the archetype scores at each time point. Shifts in enriched archetypes, defined by an archetype score > 0.8 , are annotated on the plots.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable14OvarianCohortMetadata.xlsx](#)
- [SupplementaryTable5Arcspecialistproportions.xlsx](#)
- [SupplementaryTable610WGSanalysisissummary.xlsx](#)
- [SupplementaryFigure1UMAPbeforeCCA.pdf](#)
- [SupplementaryFigure2LongitudinalscRNAcellmarker.pdf](#)
- [SupplementaryFigure3Integratedarcvariance.pdf](#)
- [SupplementaryFigure4MTLCVCoefs.pdf](#)
- [SupplementaryFigure5MTLCoefficientsClustering.pdf](#)
- [SupplementaryFigure6ValidationscRNAmarkers.pdf](#)
- [SupplementaryFigure7ValidationArcSims.pdf](#)
- [SupplementaryFigure8MTLValidationClust.pdf](#)
- [SupplementaryFigure9SplitArcK.pdf](#)
- [SupplementaryFigure10SplitPatientMTLCoefsandCorr.pdf](#)
- [SupplementaryFigure11Geneticsandarcs.pdf](#)
- [SupplementaryFigure12InferCNVprofiles.pdf](#)
- [SupplementaryFigure13ArcbyInferHMM.pdf](#)