

# Real-world clinical multi-omics analyses reveal bifurcation of ER-independent and ER-dependent drug resistance to CDK4/6 inhibitors

Received: 29 January 2024

Accepted: 2 January 2025

Published online: 22 January 2025

 Check for updates

Zhengyan Kan<sup>1</sup>✉, Ji Wen<sup>1</sup>, Vinicius Bonato<sup>2</sup>, Jennifer Webster<sup>3</sup>, Wenjing Yang<sup>1</sup>, Vladimir Ivanov<sup>4</sup>, Kimberly Hyunjung Kim<sup>1</sup>, Whijae Roh<sup>1</sup>, Chaoting Liu<sup>1</sup>, Xinmeng Jasmine Mu<sup>1</sup>, Jennifer Lapira-Miller<sup>1</sup>, Jon Oyer<sup>1</sup>, Todd VanArsdale<sup>1</sup>, Paul A. Rejto<sup>1</sup> & Jadwiga Bienkowska<sup>1</sup>✉

To better understand drug resistance mechanisms to CDK4/6 inhibitors and inform precision medicine, we analyze real-world multi-omics data from 400 HR+/HER2- metastatic breast cancer patients treated with CDK4/6 inhibitors plus endocrine therapies, including 200 pre-treatment and 227 post-progression samples. The prevalences of *ESR1* and *RBI* alterations significantly increase in post-progression samples. Integrative clustering analysis identifies three subgroups harboring different resistance mechanisms: ER driven, ER co-driven and ER independent. The ER independent subgroup, growing from 5% pre-treatment to 21% post-progression, is characterized by down-regulated estrogen signaling and enrichment of resistance markers including *TP53* mutations, *CCNE1* over-expression and Her2/Basal subtypes. Trajectory inference analyses identify a pseudotime variable strongly correlated with ER independence and disease progression; and revealed bifurcated evolutionary trajectories for ER-independent vs. ER-dependent drug resistance mechanisms. Machine learning models predict therapeutic dependency on *ESR1* and *CDK4* among ER-dependent tumors and *CDK2* dependency among ER-independent tumors, confirmed by experimental validation.

The hormone receptor-positive (HR+), human epidermal growth factor receptor 2-negative (HER2-) subtype accounts for over 70% of breast cancer diagnoses<sup>1</sup>. Among those patients ~30% progress to metastatic disease and ~6% present with metastatic disease at diagnosis<sup>2</sup>. Cyclin-dependent kinase 4 and 6 inhibitors (CDK4/6i) such as palbociclib, ribociclib and abemaciclib have significantly improved clinical outcomes for HR+, HER2- locally advanced or metastatic breast cancer (mBC) when combined with endocrine therapies (ET)<sup>3</sup>.

However, intrinsic or acquired resistance to CDK4/6i plus ET represents a major unmet medical need. Since the FDA approval of Palbociclib in 2015, many clinical and pre-clinical studies have shed light on the molecular mechanisms driving CDK4/6i resistance and informed new drug development. Comparing the molecular profiles of tumor or liquid biopsies taken at pre-treatment and post-progression provides direct insights into drug resistance mechanisms in patients. However, tumor biopsy studies conducted to-date have been limited by the

<sup>1</sup>Oncology Research & Development, Pfizer Inc., San Diego, CA, USA. <sup>2</sup>BioStatistics, Pfizer Inc., San Diego, CA, USA. <sup>3</sup>Real World Evidence, Pfizer Inc., New York, NY, USA. <sup>4</sup>Global Biometrics & Data Management, Pfizer Inc., Cambridge, MA, USA. ✉e-mail: [zhengyan.kan@pfizer.com](mailto:zhengyan.kan@pfizer.com); [Jadwiga.R.Bienkowska@pfizer.com](mailto:Jadwiga.R.Bienkowska@pfizer.com)

number of samples, especially those collected post-progression. On the other hand, liquid biopsy studies such as sequencing of circulating tumor DNA (ctDNA) have been limited by the number of genes investigated and lower assay sensitivity compared to tumor biopsies. Nevertheless, the collective scientific effort has identified several potential drivers of resistance contributing to mBC progression on CDK4/6i plus ET in the clinical setting<sup>4</sup>.

Sequencing of ctDNA from patients after progression on ribociclib plus ET has identified *FGFR1* amplifications or activating mutations in 41% of cases<sup>5</sup>. ctDNA analysis of paired liquid-biopsy samples from the PALOMA-3 trial found that new *PIK3CA* and *ESR1* driver mutations frequently emerged after treatment in both the treatment arm (palbociclib plus fulvestrant) and the control arm (fulvestrant only). In contrast, *RBI* LOF mutations were detected only in the treatment arm in 4.7% of the patients<sup>6</sup>. Gene-panel sequencing of a cohort of real-world tumor-biopsies confirmed that *PIK3CA* and *ESR1* mutations are most enriched post CDK4/6 plus ET progression<sup>7</sup>, with *RBI* LOF in 5% of post-progression samples. A whole-exome sequencing (WES) study revealed eight distinct mechanisms of resistance to CDK4/6i including *RBI* loss, activating alterations in *AKT1*, *RAS*, *AURKA*, *CCNE2*, *ERBB2*, and *FGFR2*, and loss of estrogen receptor expression<sup>8</sup>. A large-panel gene expression analysis of tumor samples from the PALOMA-3 trial found that high *CCNE1* mRNA expression was associated with relative resistance to palbociclib<sup>9</sup>. A prospective study of palbociclib-treated patients revealed that post-treatment tumors were enriched in APO-BEC mutation signatures, and many switched to aggressive molecular subtypes with estrogen-independent characteristics. Upon disease progression, mutation frequency increased in *RBI*, *ESR1*, *PTEN*, and *KMT2C*<sup>10</sup>. In addition, *CDK4* or *CDK6* overexpression has been reported as a putative driver of CDK4/6i resistance in preclinical studies<sup>11</sup>. Recent studies are increasingly adopting the multi-omics and integrative approaches towards studying CDK4/6i resistance and identifying correlates of drug response<sup>12,13</sup>. However, prior studies mainly focused on univariate approaches that preclude combined tumor molecular attributes, hampering the development of precision medicine approaches tailored for patients who progress on CDK4/6i plus ET due to different resistance mechanisms.

With the growing catalog of molecular markers associated with CDK4/6i resistance, we aimed to address several inadequately addressed questions: (1) Which of the previously identified markers associate with resistance in the real-world treatment setting? (2) What is the biomarker prevalence in real-world pre-treatment and post-progression? (3) Do these biomarkers define tumor subgroups driven by distinct resistance mechanisms? (4) How do these resistance mechanisms evolve in patients' tumors? To address these questions, we assembled a real-world clinical genomic dataset from a cohort of HR+/HER2- metastatic BC patients who developed progressive disease (PD) under CDK4/6i plus ET from the de-identified Tempus database. We then retrospectively analyzed the multi-omics profiling data derived from pre-treatment and post-progression tumor biopsies taken from these 400 patients. Our analysis of genetic drivers and transcriptional programs provided a comprehensive list of biomarkers associated with resistance to CDK4/6i plus ET and estimated prevalence of these markers post-progression. Integrative machine-learning analysis classified tumors into molecularly distinct subgroups, including a subset of the most aggressive tumors driven by estrogen receptor (ER) independent mechanisms. Our analyses suggest distinct drivers of CDK4/6i resistance and actionable therapeutic strategies for the patient subgroups supported by real-world evidence.

## Results

### Overview of cohort and dataset

Our cohort consisted of 400 patients diagnosed with HR+/HER2- metastatic breast cancers who were treated by CDK4/6 inhibitors

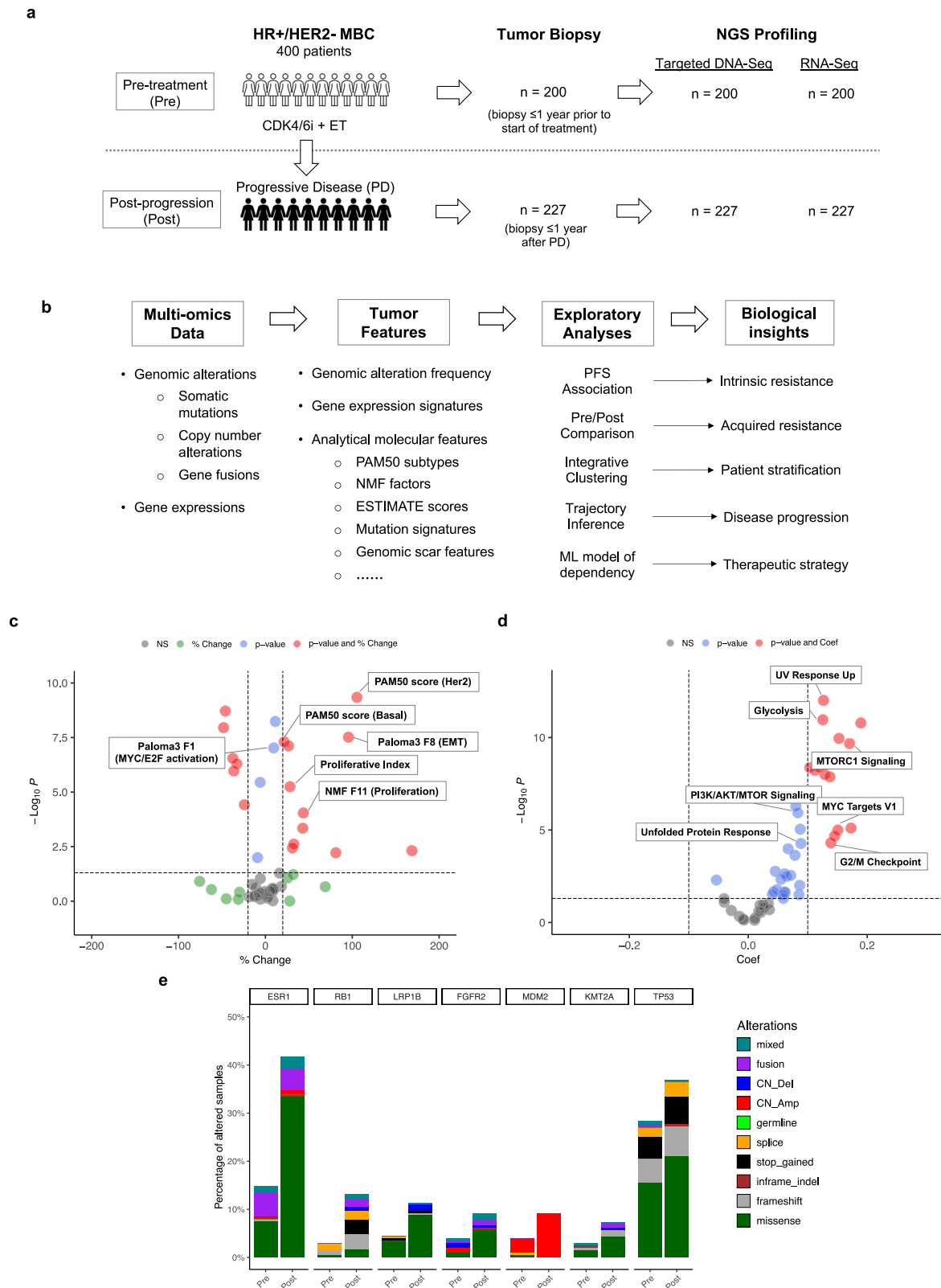
(CDK4/6i) plus endocrine therapies (ET) and developed progressive disease (PD). Pre-treatment (Pre) biopsies were taken from 200 patients within one year before starting the CDK4/6i treatment and post-progression (Post) biopsies from 227 patients were taken within one year following PD (Fig. 1a). Targeted DNA sequencing and RNA-Seq, using the Tempus xT and RS solid tumor assays, were performed on 427 tumor samples (200 Pre, 227 Post) including 27 longitudinal Pre/Post pairs.

The median age of the patients was 57 years old (interquartile range (IQR): 47–64) and median progression free survival (PFS) is 379 days (95% confidence interval: 341–433) (Table 1). All patients received CDK4/6i plus ET in the metastatic setting, with 80.2%, 21.8%, and 11.0% of patients receiving palbociclib, abemaciclib, and ribociclib respectively. The CDK4/6i were combined with letrozole, fulvestrant, and anastrozole as endocrine therapies for 38%, 24.5%, and 8% of patients. Most patients (58%) did not receive any prior treatment in the adjuvant or neoadjuvant settings. Only 39.2% of patients received prior treatment in the metastatic setting, including 26.8% who received prior endocrine therapies and 11.8% who received prior chemotherapies. The median duration between pre-treatment biopsy to the start of CDK4/6i treatment is 29 days (IQR: 17–54). The median duration between the end of CDK4/6i treatment to post-treatment biopsy is 31 days (IQR: 9.5–99.5). Among all biopsies, the most common tissue sites are liver (32.1%), breast (27.6%), and lymph nodes (10.1%).

### Systematic identification of molecular features associated with disease progression

We analyzed the multi-omics profiling data to derive three types of molecular features—genomic alteration frequencies, gene expression signatures, and analytical molecular features—to characterize different aspects of tumor intrinsic biology as well as tumor microenvironment (Fig. 1b). We then applied two systematic approaches, Pre/Post comparison, and baseline PFS association, to identify features associated with intrinsic and acquired drug resistance mechanisms that drive disease progression. Building upon these initial analyses, we performed exploratory analyses such as integrative clustering and trajectory inference to glean biological insights pertaining to patient stratification and disease progression.

We applied a suite of bioinformatics methods to calculate 63 analytically derived molecular features (See “Methods”) (Supplementary Data 1). Pre/Post comparison analyses found 13 of these features to be significantly enriched post-progression after FDR correction ( $p < 0.05$  and  $q < 5\%$ ) (Supplementary Data 2) and 4 features significantly associated with shorter PFS at baseline ( $p < 0.05$  and  $q < 5\%$ ) (Supplementary Data 3, Supplementary Fig. 1a). Since molecular mechanisms that drive disease progression could be more confidently detected as having association with both intrinsic and acquired resistance, we examined the convergence between significant hits from the two analyses. We found that 4 features were significant in both analyses: proliferative index, PAM50 correlation score for Her2 subtype and two latent expression factors identified by non-negative matrix factorization (NMF) analyses—Paloma3 MYC/E2F activation factor (F1) and de novo NMF proliferation factor (F11) (Fig. 1c). We also calculated expression signature scores for 50 Hallmark pathways to represent specific biological states or processes implicated in tumor biology<sup>14</sup> and identified 33 signatures significantly increased from Pre to Post ( $p < 0.05$  and  $q < 5\%$ ) (Supplementary Data 4). PFS association analysis found that 7 signatures were significantly associated with shorter PFS (Supplementary Data 3, Supplementary Fig. 1b). These seven signatures, related to tumor metabolism (glycolysis, PI3K/AKT/mTOR signaling, UV response up), cell cycle regulation (G2M checkpoint, MYC targets) and unfolded protein response pathways, were significant hits from both analyses (Fig. 1d).



**Fig. 1 | Molecular features associated with disease progression.** Schematic overview of study design (a) and data analysis strategy (b). Volcano plots highlighting molecular features (c) and Hallmark gene signatures (d) significantly up regulated from Pre to Post and associated with shorter PFS at baseline. The y-axis shows the signed  $\log_{10}(p)$  representing the statistical significance of the change in feature value between Pre and Post, estimated by two-sided linear mixed-effects regression (LMER). % Change: percentage change in feature value in Post vs. Pre. Coef: effect size measured by regression coefficient from LMER analysis. The vertical dashed lines indicate

the cutoffs  $>10\%$  or  $<-10\%$ . The horizontal dashed line indicates the cutoff for  $p$  value  $< 0.05$ . Each signature is colored based on its statistical significance, with red indicating that both cutoffs are met and blue and green indicating that only one of the cutoffs is met.  $-\log_{10} P$ : statistical significance. NS: not significant. e Systematic Pre vs. Post Comparison of Genomic Alteration Frequencies identified 7 genes with one-tailed fisher exact test  $p < 0.05$  and alteration frequency in Post  $>5\%$  and covered in all sequencing panels. Source data are provided as a Source Data file.

**Table 1 | Summary of cohort characteristics**

	Pre (n = 200)	Post (n = 227)	All (n = 400)
Patient age, median (s.d.)			
Age at PD	59 (12.0)	55 (12.1)	57 (12.2)
CDK4/6 treatment, n (%)			
Palbociclib	156 (78.0%)	185 (81.5%)	321 (80.2%)
Abemaciclib	48 (24.0%)	47 (20.7%)	87 (21.8%)
Ribociclib	24 (12.0%)	24 (10.6%)	44 (11.0%)
Endocrine therapy, n (%)			
Letrozole	95 (47.5%)	95 (41.9%)	178 (44.5%)
Fulvestrant	65 (32.5%)	77 (33.9%)	134 (33.5%)
Anastrozole	19 (9.5%)	25 (11.0%)	40 (10.0%)
Exemestane	10 (5.0%)	9 (4.0%)	17 (4.2%)
Leuprolide	9 (4.5%)	15 (6.6%)	21 (5.2%)
Goserelin	6 (3.0%)	10 (4.4%)	15 (3.8%)
Tamoxifen	5 (2.5%)	13 (5.7%)	17 (4.2%)
Prior treatment in adjuvant or neoadjuvant settings, n (%)			
Chemotherapy	46 (23.0%)	61 (26.9%)	100 (25.0%)
Endocrine therapy	57 (28.5%)	87 (38.3%)	132 (33.0%)
All prior treatment	76 (38.0%)	105 (46.3%)	168 (42.0%)
Prior treatment in metastatic setting, n (%)			
Prior chemotherapy	21 (10.5%)	32 (14.1%)	47 (11.8%)
Prior endocrine therapy	55 (27.5%)	61 (26.9%)	107 (26.8%)
All prior treatment	76 (38.0%)	97 (42.7%)	157 (39.2%)
Biopsy site, n (%)			
Breast	83 (41.5%)	35 (15.4%)	118 (27.6%)
Liver	32 (16.0%)	105 (46.3%)	137 (32.1%)
Lymph node	25 (12.5%)	18 (7.9%)	43 (10.1%)
Bone	22 (11.0%)	14 (6.2%)	36 (8.4%)
Thorax	11 (5.5%)	10 (4.4%)	21 (4.9%)
Skin	7 (3.5%)	11 (4.8%)	18 (4.2%)
Lung	6 (3.0%)	6 (2.6%)	12 (2.8%)
Pleura	4 (2.0%)	8 (3.5%)	12 (2.8%)
Others	10 (5.0%)	20 (8.8%)	30 (7.0%)

### ***ESRI* and *RBI* genomic alterations represent major mechanisms of acquired drug resistance**

We compared genomic alteration frequencies (GAF) between pre-treatment and post-progression sub-cohorts for 611 genes and identified 7 genes with significant increases of GAF post-progression ( $p < 0.05$ , Fisher Exact Test–FET) (Supplementary Data 5, Fig. 1e). *ESRI* had the most significant increase in GAF between Pre and Post (15% vs. 41.9%,  $p = 5.41e-10$ ). *RBI* was ranked second with GAF increased from 3% pre-treatment to 13.2% post-progression ( $p = 8.5e-05$  FET). Only *ESRI* and *RBI* remained significant after FDR adjustment, with  $q = 4.48e-07$  and  $q = 0.035$  respectively. We calculated baseline PFS associations for 140 gene-level mutations, 41 copy number amplifications, and 14 deletions (Supplementary Data 3). After FDR adjustment, only TP53 mutation remained significant ( $q = 3.7e-3$ , HR = 2.12) (Supplementary Fig. 1c). *RBI* and *ESRI* mutations showed a non-significant trend towards association with shorter PFS with  $p = 5.4e-2$  ( $q = 0.22$ , HR = 2.53) and  $p = 0.35$  ( $q = 0.51$ , HR = 1.31) respectively (Supplementary Fig. 1d–e).

An acquired genomic alteration, defined as a somatic alteration detected in the post-progression samples but not in any pre-treatment samples from the same patient, could be selected through tumor clonal evolution due to its functional contribution to drug resistance.

We ranked the gene-level frequencies of acquired genomic alterations for 26 longitudinally paired samples and found that *ESRI* and *RBI* were among the top 10 genes ranked by acquired GAF, strengthening the evidence implicating these two genes in conferring resistance to CDK4/6i plus ET (Supplementary Fig. 2a). Compared to *RBI* and *TP53* mutations, *ESRI* mutations are more strongly associated with up-regulation of estrogen response signature, consistent with the understanding that *ESRI* mutations activate the estrogen signaling pathways (Supplementary Fig. 2b). On the other hand, *RBI* and *TP53* mutations are more strongly associated with markers of tumor growth and proliferation such as proliferation index and cyclin E1 gene expression (Supplementary Fig. 2c, d). Post-progression *ESRI* mutations are enriched in the ligand-binding domain (LBD) consisting of multiple hotspot mutations, such as Y537S, Y537N, D538G, known to mediate constitutive signaling (Supplementary Fig. 3a). *RBI* and *TP53* mutations are enriched in protein-truncating alterations, such as frameshift, nonsense or splice-site changes that tend to result in loss-of-function (LOF) (Supplementary Fig. 3b, c). Echoing previous observations such as the PALOMA-3 ctDNA study<sup>6</sup>, our results indicate that *ESRI* and *RBI* genomic alterations are two major mechanisms of acquired resistance to CDK4/6i plus ET<sup>6</sup>.

### **Cell cycle de-regulation and decreased ER signaling strongly associated with disease progression**

Cell cycle dysregulation through genetic alteration of cell cycle regulators such as *RBI* and *TP53* is a key mechanism for conferring CDK4/6i resistance<sup>4</sup>. Our analyses revealed various evidence suggesting that aggressive tumor growth and proliferation due to abnormal regulation of cell cycle is strongly linked to disease progression. While *RBI* was one of the top genes with increased GAF post-progression, GAF of *TP53* also increased from 28.5% to 37% ( $p = 3.9e-02$ , FET) (Fig. 1e). In addition, both *RBI* and *TP53* mutations were significantly associated with increased proliferation index (*RBI*:  $p = 6e-04$ , Wilcoxon rank sum test–WRST; *TP53*:  $p = 8.1e-06$ , WRST) and *CCNE1* gene expression (*RBI*:  $p = 3e-08$ , WRST; *TP53*:  $p = 2.4e-04$ , WRST) (Supplementary Fig. 2c, d).

The proliferative index, an 11-gene signature representing tumor growth and proliferation, significantly increased post-progression among all samples ( $p = 1.2e-4$ , WRST) and was significantly associated with shorter PFS ( $p = 3.1e-3$ ,  $q = 0.04$ , HR = 1.3, CPH) (Supplementary Fig. 4a, b). In addition, aggressive PAM50 subtypes such as Her2 (2% to 12.3%,  $p = 2.96e-5$ , Fisher Exact Test–FET) and Basal (1.5% to 7.1%,  $p = 8.0e-3$ , FET) increased from Pre to Post. In contrast, the less aggressive Luminal A subtype decreased from 70.5% Pre to 47.1% Post ( $p = 1.31e-06$ , FET) (Supplementary Fig. 4c). PAM50 correlation scores for Her2 ( $p = 4.1e-5$ ,  $q = 2.2e-03$ , HR = 1.41, CPH) and Basal ( $p = 1.4e-2$ ,  $q = 0.10$ , HR = 1.27, CPH) subtypes are also significantly associated with shorter PFS (Supplementary Fig. 4d, e). On the other hand, PAM50-score for Luminal A was associated with longer PFS ( $p = 7.7e-05$ ,  $q = 2.8e-03$ , HR = 0.71, CPH) (Supplementary Fig. 4f).

Cyclin E1 is a key cell cycle regulator of CDKs and the G1/S transition and *CCNE1* expression has been previously reported to be a predictive marker of CDK4/6 resistance by the PALOMA-3 clinical study<sup>9</sup>. NMF analysis of the PALOMA-3 gene expression data also identified the MYC/E2F Activation Factor (F1) as a predictive marker of CDK4/6i resistance<sup>15</sup>. In our cohort, *CCNE1* expression is significantly up-regulated post-progression ( $p = 2.5e-05$ , WRST) and associated with shorter PFS at baseline ( $p = 2.5e-3$ , HR = 1.29, CPH) (Supplementary Fig. 5a, b). Moreover, the MYC/E2F Activation factor significantly increased post-treatment in all samples ( $p = 6.9e-06$ , WRST) with a marginally significant increase among paired samples ( $p = 0.052$ , paired WRST) (Supplementary Fig. 5c). This factor is also significantly associated with shorter PFS ( $p = 1.1e-5$ ,  $q = 5.0e-04$ , HR = 1.49, CPH) (Supplementary Fig. 5d). Consistently, Hallmark signatures for E2F targets and MYC targets are also significantly up-regulated post-progression and associated with shorter PFS at baseline (Supplementary

Fig. 5e–h). Taken together, these results suggest that de-regulation of cell cycle manifesting in aggressive tumor growth and proliferation confers drug resistance.

While ER loss has been associated with CDK4/6i resistance<sup>8</sup>, its mechanism and correlation with other markers of CDK4/6i resistance remains poorly understood. Our analysis revealed that markers of decreased ER signaling are strongly associated with disease progression. The estrogen receptor- $\alpha$  (*ESR1*) gene expression is down-regulated after disease progression ( $p = 7.2e-13$ , WRST) and significantly associated with longer PFS pre-treatment ( $p = 2.6e-2$ , HR = 0.80, CPH) (Supplementary Fig. 6a, b). Progesterone receptor (*PGR*), a marker of estrogen signaling, is significantly down-regulated post-treatment ( $p = 9e-11$ , WRST) and associated with longer PFS pre-treatment ( $p = 4.7e-5$ , HR = 0.7, CPH) (Supplementary Fig. 6c, d). In addition, the estrogen response factor (F10) identified by PALOMA-3 data analysis significantly decreased in all samples ( $p = 5.4e-09$ , WRST) and paired samples ( $p = 0.00021$ , paired WRST) (Supplementary Fig. 6e). Moreover, the estrogen response factor is significantly associated with longer PFS ( $p = 6.9e-7$ ,  $q = 1.3e-04$ , HR = 0.61, CPH) (Supplementary Fig. 6f). Thus, decreased ER signaling could indicate a loss of dependence on ER signaling as the oncogenic driver which would confer resistance to endocrine treatments.

### Integrative multi-omics pattern clustering analysis identified molecularly distinct subgroups

A meta-analysis of six independent cohorts has reported mutual exclusivity of *ESR1* and *TP53* mutations in endocrine resistant, ER+ metastatic breast cancer<sup>16</sup>. In our cohort, *TP53* and *ESR1* mutations are significantly mutually exclusive in Post ( $p = 3.78e-4$ , logistic regression–LR), but were not detected as mutually exclusive in Pre ( $p = 0.299$ ). Interestingly, *TP53* mutations are also mutually exclusive with *GATA3* mutations in both Pre ( $p = 1.8e-2$ , LR) and Post ( $p = 2.4e-4$ , LR). *ESR1* and *RBI* mutations showed a non-significant trend of mutual exclusivity among post-progression samples ( $p = 0.126$ , LR).

To explore the possibility that patients may be stratified into segments whose cancers are driven by distinct drug resistance mechanisms, we performed integrative clustering analysis on selected multi-omics features using iClusterPlus<sup>17</sup> (see “Methods”). This analysis revealed five integrative clusters IC1–5 with distinct molecular features consisting of 13.35% ( $n = 57$ ), 29.7% ( $n = 127$ ), 22.7% ( $n = 97$ ), 25.1% ( $n = 107$ ) and 9.1% ( $n = 39$ ) of the cohort samples respectively (Fig. 2a, Supplementary Data 6). IC1 is significantly associated with shorter PFS ( $p = 3.43e-05$ , HR = 4.47) and increased from 5% pre-treatment to 20.7% post-progression (FET:  $p = 1.38e-06$ ). In contrast, IC4 is marginally associated with longer PFS ( $p = 8.5e-02$ , HR = 0.68) and decreased from 30% pre-treatment to 20.7% post-progression (FET:  $p = 3.3e-02$ ) (Fig. 2b, c). The PAM50 subtypes are differentially distributed vs. the integrative clusters. The non-luminal subtypes, Basal and Her2-like, are strongly enriched in IC1, of which 35.1% (20/57) was classified as Her2, 33.3% (19/57) was classified as Basal, and 22.8% (13/57) was classified as Luminal B. On the other hand, IC4 and IC5 are predominantly classified as Luminal A while IC2 and IC3 are predominantly classified as Luminal A or Luminal B (Supplementary Fig. 7a). IC3 is differentiated from IC2 in that it is enriched in liver as the tissue site of biopsy (52.6% vs. 31.7%) and has higher tumor purity than other clusters (Supplementary Fig. 7b, c). IC5 exhibits a distinctive tissue site distribution compared to IC4 and other clusters with an enrichment in pleura and lymph. (Supplementary Figs. 7b, c).

Notably, IC1 is differentiated from other clusters by significant down-regulation of estrogen signaling markers such as estrogen receptor (*ESR1*) and progesterone receptor (*PGRI*) expression (Fig. 2d). IC1 is also strongly associated with markers linked to cell cycle progression and CDK4/6i resistance, including cyclin E1 up-regulation (Fig. 2d), increased PALOMA-3 NMF factor F1 (MYC/E2F activation)

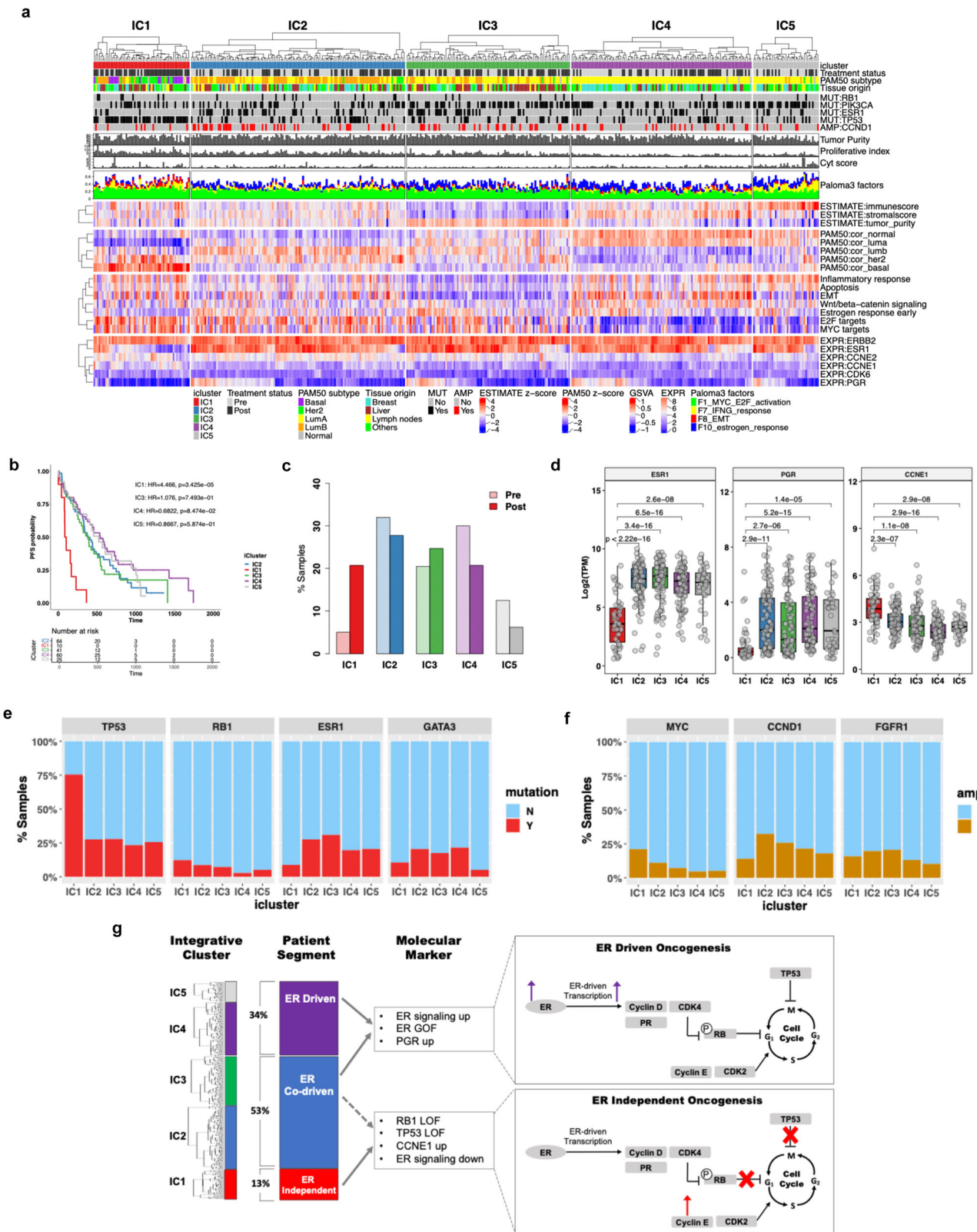
factor, decreased F10 (Estrogen response) factor (Supplementary Fig. 7d), increased Hallmark E2F/MYC targets and decreased estrogen response signatures (Supplementary Fig. 7e). Furthermore, these five clusters of tumors harbored distinct pattern of genomic alterations suggesting that different oncogenic drivers contributed to their disease development. *TP53* mutation is significantly enriched in IC1 (75.4% vs. 24.6%, FET:  $p = 1.32e-12$ ) whereas *ESR1* mutation prevalence is significantly lower in IC1 compared to other clusters (8.8% vs. 25.4%, FET:  $p = 0.004$ ) (Fig. 2e). *RBI* mutation is enriched in IC1 (12.3% vs. 6.2%, FET:  $p = 0.1$ ) but under-represented in IC4 (2.8% vs. 8.4%, FET:  $p = 0.05$ ). Likewise, *MYC* amplification is enriched in IC1 (21.1% vs. 7.6%, FET:  $p = 0.003$ ) but under-represented in IC4 (4.7% vs. 10.9%, FET:  $p = 0.057$ ) (Fig. 2f). Hence, the mutual exclusivity patterns among mutations of *TP53*, *ESR1*, *RBI*, and *GATA3* could be attributed to tumor stratification that result from distinctive drug resistance mechanisms mediated by these cancer genes.

To investigate whether IC1 tumors had lower ER levels prior to treatment or lost ER signaling during the course of treatment, we curated immunohistochemistry (IHC) results in an independent set of tumor specimens taken from patients in our cohort. We then evaluated the longitudinal change in ER levels over three clinical settings—primary tumor (Primary), metastatic & pre-treatment (MetPre) and metastatic & post-treatment (MetPost)—while grouping patients by integrative clusters or PAM50 subtypes based on tumor expression profiles of their Post samples (Supplementary Fig. 8a, b). A significant longitudinal loss of ER was only observed among patients whose tumors were classified as HER2 ( $p = 4e-04$ , Kruskal Wallis test - KWT) or IC1 ( $p = 5.4e-02$ , KWT) after disease progression, suggesting that treatment induced ER loss in these subsets of patients.

IC1 tumors are characterized by both loss of ER signaling and drug resistance markers such as cyclin E1 over-expression that are indicative of aggressive tumor growth and proliferation. Thus, we hypothesize that IC1 represents a molecular state that is “ER-independent”, indicating loss of dependence on ER signaling as the oncogenic driver with increasing reliance on alternative mechanisms such as cell cycle de-regulation through *RBI* loss-of-function (LOF) (Fig. 2g). This molecular state is likely to be resistant to both CDK4/6i and endocrine therapies. In contrast, IC4 and IC5 seem to be driven by ER signaling, coined as “ER-driven”, based on higher estrogen response signatures, frequent *ESR1* gain-of-function (GOF) mutations along with a lack of *RBI* mutations relative to IC1. It is noteworthy that tumors classified as IC2 or IC3 share molecular characteristics with both ER-independent tumors, such as *RBI* LOF mutation, and ER-dependent tumors, such as ER GOF mutations and high estrogen response signatures. Hence, we hypothesize that IC2/3 tumors are “ER co-driven”, driven by both ER signaling and alternate mechanisms such as cell cycle de-regulation. The main distinction between IC2 and IC3 appears to be tumor external factors since IC3 is enriched in liver biopsies with relatively high tumor purity compared to IC2. This observation promoted us to further delineate molecular heterogeneity by focusing on tumor intrinsic factors.

### ER-independent tumors harbored heterogeneous tumor intrinsic mechanisms implicated in drug resistance

To investigate the biological mechanisms that uniquely contributed to ER-independence and associated drug resistance, we performed systematic differential expression (DE) analyses to compare integrative clusters IC1–5 and PAM50 subtypes (Supplementary Data 7). We deconvoluted bulk tumor expression profiles using BayesPrism<sup>18</sup> to derive cancer-specific expression (CSE) profiles and gene signature scores for delineating differences in tumor intrinsic biology. Bulk tumor expression analyses had linked increased cell cycle regulatory and tumor metabolic signatures and decreased estrogen response markers to disease progression (Supplementary Fig. 3). Consistently, CSE analysis revealed that IC1 has strongly decreased estrogen



response signature and increased cell cycle regulatory signatures such as E2F targets and MYC targets and tumor metabolic signatures such as glycolysis and mTORC1 signaling compared to IC2-5, affirming that IC1 is the most aggressive of the four clusters. Moreover, CSE analysis revealed that tumor intrinsic inflammatory response, JAK/STAT, and epithelial-mesenchymal-transition (EMT) signatures are up-regulated in IC1 (Fig. 3a, b). EMT enables cancer cells to acquire stem cell-like

phenotype and is believed to confer resistance to various therapeutic agents in multiple cancers<sup>19</sup>. JAK/STAT and inflammatory signaling were found to drive lineage plasticity and antiandrogen resistance in castration-resistant prostate cancer<sup>20</sup>. Taken together, up-regulation of these tumor intrinsic signatures in IC1 (IC1-UP) indicated that ER-independent tumors harbored heterogeneous tumor intrinsic molecular mechanisms that contributed to drug resistance, including cell

**Fig. 2 | Integrative clustering analysis identified molecularly distinct subgroups.** **a** Heatmap showing distinct patterns of molecular features of five integrative clusters IC1-5. Panels of selected multi-omics features include gene expressions, hallmark signatures (GSVA), PAM50 subtype scores, ESTIMATE scores for tumor microenvironment, projection to Paloma3 NMF factors, gene-level genomic alteration status and other tumor characteristics. icluster: integrative cluster. Cyt score: cytolytic activity score. TMB: tumor mutation burden. Treatment status: Pre/Post. **b** KM plot comparing the PFS of patients classified into the five IC clusters. **c** Changes in the prevalence of IC clusters Pre vs. Post. **d** Distributions of *ESRI*, *PGR* and *CCNE1* gene expressions vs. IC clusters, with sample sizes  $n = 57$  (IC1), 127 (IC2), 97 (IC3), 107 (IC4), and 39 (IC5). Statistical significance was determined

using two-sided Wilcoxon rank sum test. Distributions of *TP53*, *ESRI*, *RBI*, *GATA3* mutation statuses (**e**) and *MYC*, *CCND1* and *FGFR1* amplification statuses (**f**) vs. IC clusters. **g** Diagram illustrating the hypothesis of stratifying HR+/HER2- mBC patients into three segments differentiated by the dependency on ER signaling as the oncogenic mechanism. The dotted line indicates that the ER co-driven tumor segment shares some characteristics with the ER independent segment, such as harboring higher RB1 mutation frequency than ER-driven tumors. For all box-and-whisker plots, the box is bounded by the first and third quartile with a horizontal line at the median and whiskers extend to the maximum and minimum value. Source data are provided as a Source Data file.

cycle dysregulation<sup>4</sup>, metabolic reprogramming<sup>21,22</sup>, EMT<sup>19</sup>, and inflammatory signaling<sup>20</sup>.

Non-luminal PAM50 subtypes such as Her2-like and Basal-like are typically associated with HER2-positive and triple negative breast cancers. These non-Luminal subtypes only accounted for 11.9% (51/427) of our HR+/HER2- cohort but are significantly associated with disease progression (Supplementary Fig. 4c, e) and strongly enriched in the ER-independent tumors (Supplementary Fig. 7b). In a longitudinal study of Palbociclib resistance, we have previously reported that tumors frequently switched from Luminal subtypes at baseline to Her2-like subtype post-treatment along with decreased expression of estrogen-response genes<sup>10</sup>. Lineage plasticity is known to cause drug resistance by enabling tumor conversion into a different subtype and loss of dependency on the original oncogenic driver<sup>23</sup>. Loss of luminal lineage factors along with ER expression has been linked to endocrine resistance in ER+ breast cancers<sup>24</sup>. A recent study has demonstrated that cell plasticity enables a distinct subset of luminal breast tumors expressing lower levels of ER to evolve and gain basal-like features<sup>25</sup>. Echoing these findings, we also observed in our cohort that Basal-like tumors are down-regulated in luminal lineage markers such as *FOXAI*, *SPDEF*, and *KRT18* while up-regulating basal lineage markers including *KRT5*, *KRT14*, and *KRT17* (Supplementary Fig. 9a-b). In addition, Her2-like tumors are strongly up-regulated in *ERBB2* signaling along with higher *ERBB2* gene expression when compared to other subtypes (Supplementary Fig. 9c, d). The non-Luminal tumors are characterized by higher levels of IC1-UP drug resistance signatures and lower levels of tumor intrinsic estrogen response signature when compared to Luminal tumors. Tumor metabolic signatures are higher in both Her2 and Basal to a similar extent while inflammatory and EMT signatures are significantly higher in Basal than in Her2 (Fig. 3c).

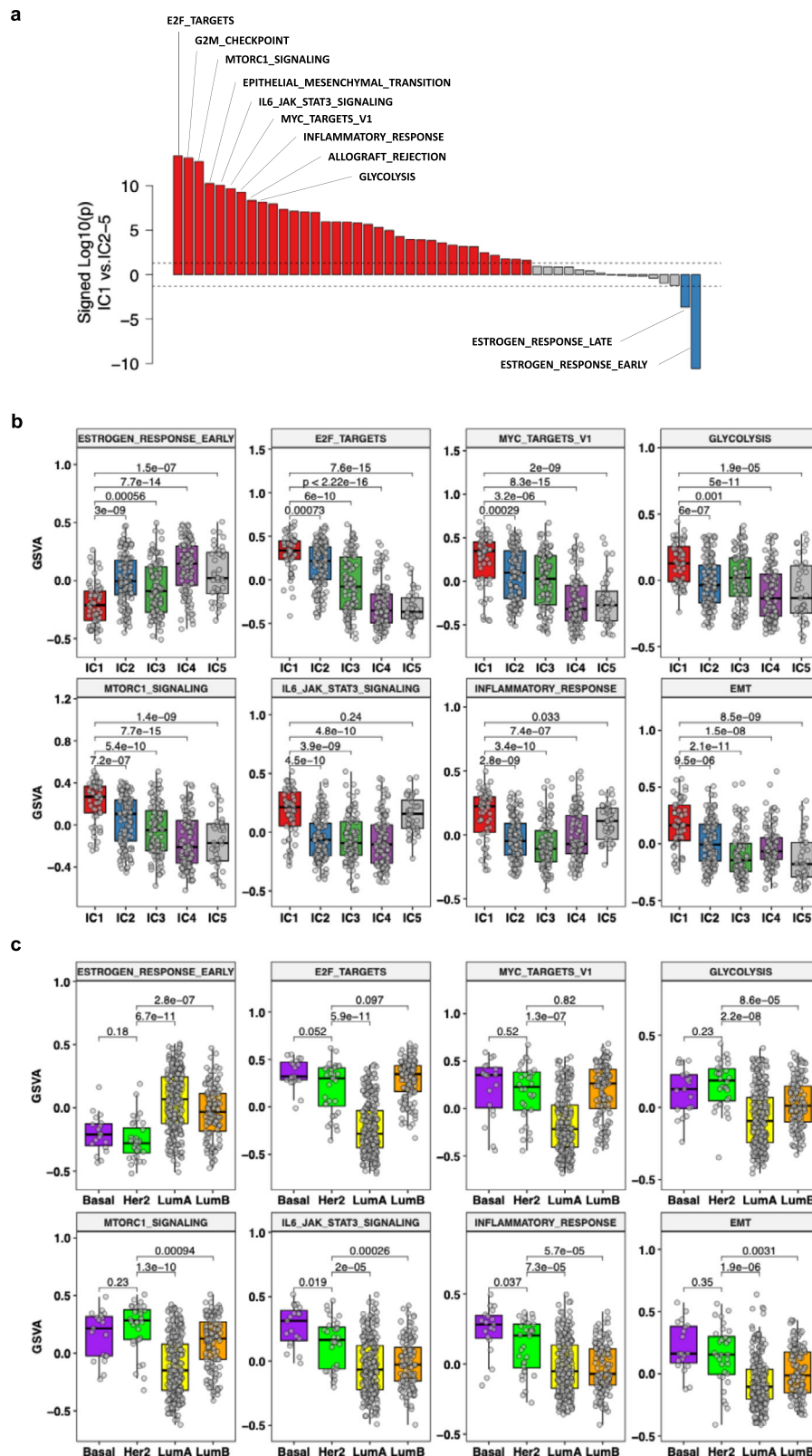
### Pseudotime analysis identified a latent disease progression trajectory associated with ER independence

We noticed the opposite trends of increasing drug resistance signatures and decreasing estrogen response signatures from Luminal subtypes associated with IC4-5 to non-luminal subtypes associated with IC1, suggesting that drug resistance evolved through a series of molecular states along with decreasing dependence on estrogen signaling (Fig. 3b, c). To better delineate this latent process of disease progression, we performed trajectory inference analyses which are often used in single-cell transcriptomics to determine quantitative variables that measure the relative progression of cells through a dynamic biological process. We postulated that tumor expression profiles from a large cohort of patients allow us to sample the landscape of possible molecular states of the disease that emerge and evolve under treatment and model dynamic progression of drug resistance. To explore the latent dimensions of disease progression, we used Monocle 3<sup>26</sup> to define a pseudotime variable (PT) to measure progression along a latent process for each tumor sample from the CSE profiles (Supplementary Data 8, Supplementary Fig. 10a). We separated the post-progression samples into two time points—during treatment (During) and post treatment (Post) and saw that PT longitudinally increased from Pre, During to Post ( $p = 1.9e-08$ )

(Supplementary Fig. 10b, c). We also observed an increasing trend of PT from Luminal A to Luminal B and then to Her2 and Basal ( $p = 0.089$ , Mann-Kendall Trend Test - MKT), consistent with the prior knowledge that non-Luminal subtypes are more aggressive with worse clinical outcome than Luminal subtypes (Supplementary Fig. 10d). Comparing tumors classified into different integrative clusters, PT increased from IC4 to IC2 and IC3 and reached the highest values in IC1 ( $p = 0.027$ ), consistent with the hypothesis that IC1 represent a subgroup of tumors that are aggressive and drug-resistant (Supplementary Fig. 11a). PT was also significantly anti-correlated with estrogen signaling markers including *ESRI* (Pearson correlation  $r = -0.42$ ,  $p = 2e-12$ ), *PGR* gene expression ( $r = -0.4$ ,  $p = 4.7e-10$ ) and the Estrogen response factor F10 ( $r = -0.61$ ,  $p < 2.2e-16$ ) (Supplementary Fig. 11b-d). On the other hand, PT was significantly correlated with cell growth and proliferation markers including *CCNE1* ( $r = 0.51$ ,  $p < 2.2e-16$ ) and *MYC2/E2F* activation factor ( $r = 0.53$ ,  $p < 2.2e-16$ ) (Supplementary Fig. 11e, f). Among the 50 Hallmark tumor intrinsic signatures, PT is significantly anti-correlated with the tumor intrinsic estrogen response signature ( $r = -0.36$ ,  $p = 2.66e-10$ ) and positively correlated with cell cycle regulation, tumor metabolism, inflammatory response and EMT-related signatures that are enriched in ER independent tumors and implicated in drug resistance (Fig. 4a). Finally, PT is significantly associated with genomic alterations contributing to faster disease progression including *TP53* mutation ( $p = 1.9e-07$ ), *RBI* mutation ( $p = 6.8e-03$ ), and *MYC* amplification ( $p = 9.9e-03$ ) (Supplementary Fig. 11g). These observations support the hypothesis that the latent pseudotime variable is a descriptor of the dynamic and continuous evolution of ER independence as a drug resistance mechanism.

### Trajectory inference analysis revealed bifurcation of ER driven vs. ER independent trajectories towards drug resistance

To examine the possibility of multiple trajectories of disease progression, we used the elastic principal graph (EPG) algorithm<sup>27</sup> and inferred a topology comprised of 3 branches A-C and 21 nodes, where each branch consists of multiple nodes that are mapped to individual samples (Fig. 4b). We found that tumor samples mapped to branch C are enriched in IC1 while those mapped to branches A and B are enriched with IC4 and IC2-3 respectively (Fig. 4c). Like IC1, branch C is enriched in Her2 and Basal subtypes. In addition, branch A is enriched in Luminal A while branch B is enriched in Luminal B (Supplementary Fig. 12a). Branch C increased from 13.8% Pre to 43.9% Post while branch A decreased from 53.9% to 24.3% (Supplementary Fig. 12b). Branch C is down regulated in ER signaling markers such as *ESRI*, *PGR* expression (Fig. 4d), estrogen response signature (Supplementary Fig. 12c) and the PALOMA3 estrogen response factor (Supplementary Fig. 12d). On the other hand, branch C is up regulated in cell growth and proliferation markers including *CCNE1* expression (Fig. 4d), E2F and *MYC* targets signatures (Supplementary Fig. 12c) and the PALOMA3 *MYC/E2F* activation factor (Supplementary Fig. 12d). There is a trend of increasing pseudotime across branches A, B and C (Supplementary Fig. 12e). Branch C also resembles IC1 with an enrichment of *TP53* mutations (61.8% vs. 24.6%,  $p = 9.25e-08$ , FET) and fewer *ESRI* mutations (7.4% vs. 18.3%,  $p = 0.03$ , FET) (Fig. 4e). Branch A resembles IC4 in



having a lower prevalence of *RBI* mutations (1.9% vs. 11.3%,  $p = 3.4 \times 10^{-3}$ , FET). Branch B harbored more frequent *ESR1* (22.9% vs. 11.9%,  $p = 2.7 \times 10^{-2}$ , FET) and *GATA3* mutations (30.1% vs. 13.6%,  $p = 2.2 \times 10^{-3}$ , FET), which are enriched in the ER driven and co-driven clusters. It also harbored frequent oncogenic amplifications, including *CCND1* (31.3% vs. 16.4%,  $p = 8.9 \times 10^{-3}$ , FET), *MYC* (18.1% vs. 8.5%,  $p = 3.6 \times 10^{-2}$ , FET), and *FGFR1* (22.9% vs. 8.0%,  $p = 1.3 \times 10^{-3}$ , FET) (Fig. 4f). Based on these molecular

association patterns, we assigned EPG branches A-C to three distinct molecular states of ER dependence - branch A to ER driven, B to ER co-driven and C to ER-independent (Fig. 4b). If branch A is considered the root of the tree corresponding to the ER driven and relatively drug sensitive molecular state, then the bifurcated branches B and C represent two alternative mechanisms of drug resistance states where tumors in branch B frequently enhanced ER signaling through *ESR1*



**Fig. 3 | Differential expression patterns of tumor intrinsic gene signatures associated with ER independence.** **a** Waterfall plot showing the differential expression of 50 Hallmark signatures in IC1 vs. IC2-5. The y-axis shows the statistical significance as signed  $\log_{10}(p)$  values representing the statistical significance of the difference in mean signature scores between two groups. The dotted lines represent the significance cutoff ( $p = 0.05$ ). Each signature is colored based on its statistical significance, with red indicating up-regulation in IC1 and blue indicating up-regulation in IC2-5. P-value is determined by Wilcoxon rank sum test. Distribution

of IC1 up-regulated signatures vs. IC1-5 clusters (**b**) and PAM50 subtypes (**c**). Sample sizes for IC1-5 clusters are 57 (IC1), 127 (IC2), 97 (IC3), 107 (IC4), and 39 (IC5). Sample sizes for PAM50 subtypes are 19 (Basal), 32 (Her), 248 (LumA), and 121 (LumB). Statistical significance was determined using two-sided Wilcoxon rank sum test. For all box-and-whisker plots, the box is bounded by the first and third quartile with a horizontal line at the median and whiskers extend to the maximum and minimum value. Source data are provided as a Source Data file.

gain-of-function mutations while tumors in branch C bypassed ER signaling as oncogenic driver.

Defining a trajectory as a sequential ordering of nodes, two distinct trajectories emerge through EPG analysis with one spanning branches A to B (A-B) and another one spanning branches A and C (A-C) (Fig. 4b). Echoing the trend of increasing pseudotime from branches A to C (Supplementary Fig. 12e), we observed increasing fractions of post-progression samples (During or Post) across the nodes spanned by trajectories A-B and A-C (Supplementary Fig. 13a, b). We also observed increasing fractions of ER co-driven (IC2-3) samples along trajectory A-B compared to increasing fractions of ER independent (IC1) samples along trajectory A-C (Supplementary Fig. 13c, d). Moreover, PT exhibited significant trends continuously increasing across the nodes spanned by both trajectories (A-B:  $p = 1.7e-4$ , A-C:  $p = 2.0e-5$ , MKT), suggesting that both trajectories represent disease progression (Supplementary Fig. 13e, f). On the other hand, there is a significant decreasing trend of *ESR1* expression ( $p = 5.1e-5$ , MKT) through trajectory A-C whereas no significant change of *ESR1* expression was observed along trajectory A-B ( $p = 0.92$ , MKT) (Supplementary Fig. 13g, h), suggesting distinctive molecular dynamics in these tumors during disease progression.

We evaluated the statistical significance of the monotonic trends for various CSE signatures along the two trajectories (Supplementary Data 8). Among IC1 up-regulated signatures, tumor growth and proliferation signatures such as MYC targets exhibited an increasing trend in both trajectories (Supplementary Fig. 14a). On the other hand, tumor metabolic (glycolysis), inflammatory (JAK/STAT) and KRAS signaling signatures only increased along the trajectory A-C (Supplementary Fig. 14b–d). Furthermore, we observed strong increasing trends from branches A to C for stem cell division and HIPPO/YAP pathway signatures in contrast to a decreasing trend from branches A to B, implicating cellular reprogramming and dedifferentiation in the transition into ER independence (Supplementary Fig. 14e, f).

### Machine learning models predicted changes in target gene dependencies in ER-independent tumors

Integrative clustering analyses classified 77.0% (161/209) of post-progression tumors as ER driven or ER co-driven, suggesting that ER signaling remained a dependency for most of the post-progression tumors and therefore should be targeted by the next generation of therapies. In addition, we have previously reported that palbo-resistant breast cancer cell lines up-regulated cyclin E1 expression and became sensitive to CDK2/4/6i (PF3600) due to its *CDK2* inhibitory activity<sup>15</sup>. Since IC1 also over-expressed *CCNE1* compared to other clusters, *CDK2* inhibition could be an effective treatment against ER-independent tumors.

To examine this therapeutic hypothesis, we used the elastic net algorithm to develop machine learning models (ENRS) to predict dependency on four drug target genes—*ESR1*, *CDK2*, *CDK4*, and *CDK6* (Fig. 5a). The models were trained on CRISPR-KO screen and gene expression data from a pan-cancer panel of cell lines from the DepMap project<sup>28</sup> (see “Methods”). Using tumor CSE as input, each ENRS model estimates a score for a target gene similar to the CERES dependency score<sup>29</sup>, where lower score indicates a higher likelihood that inhibiting the gene function would kill the cancer cell (Supplementary Data 9). In

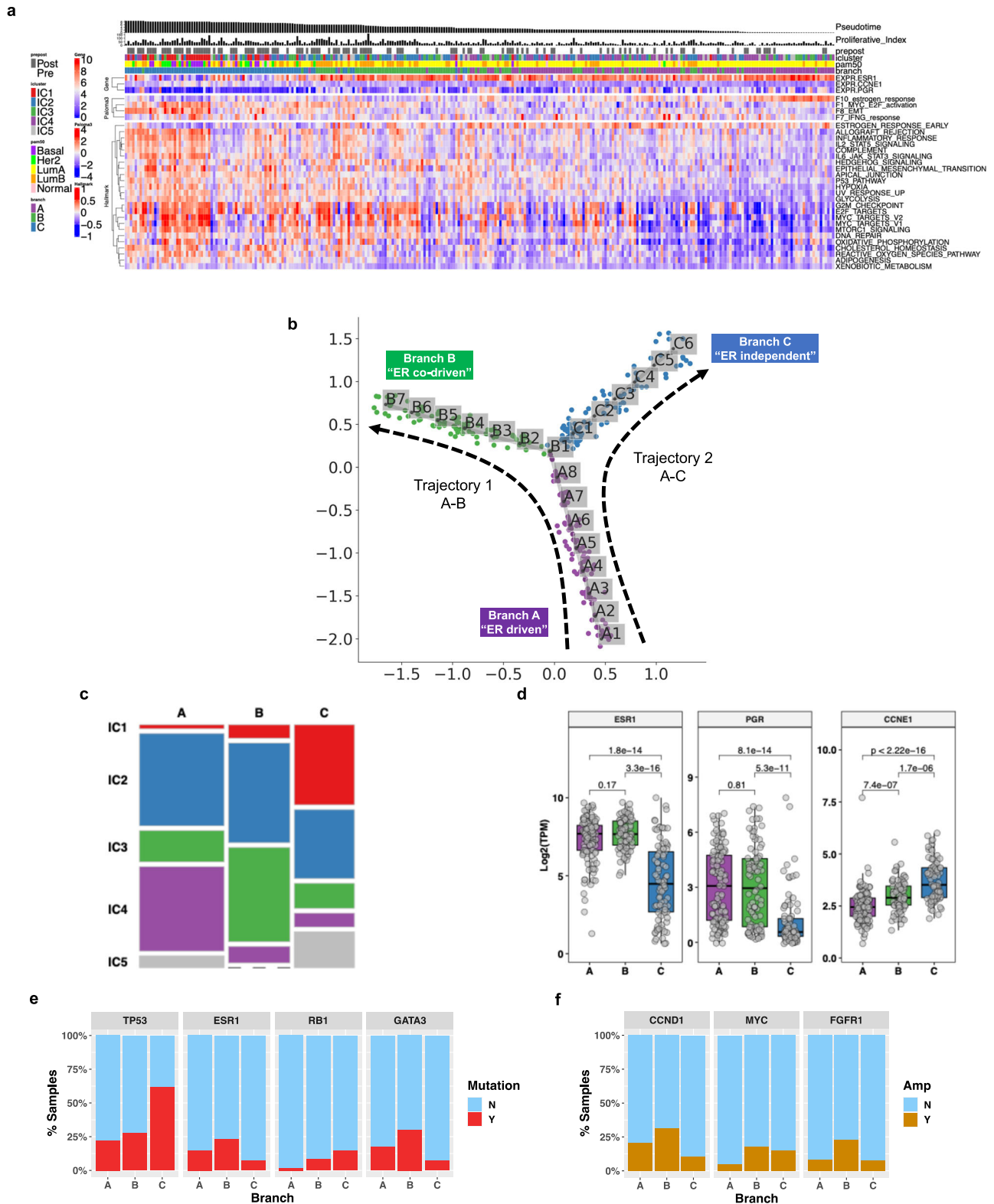
support of our hypothesis, the models predicted increased dependency on *CDK2* along with decreased dependency on *ESR1* and *CDK4* in IC1 compared to IC2-5 (Fig. 5b). Consistent with this observation, tumor dependency on *CDK2* was predicted to increase at Post vs. Pre in contrast to decreased dependency on *ESR1* and *CDK4* (Fig. 5c). Predicted *CDK6* dependency decreased in IC1 and post-progression but the change is less drastic than *CDK4* or *ESR1* (Fig. 5b, c). These results suggest the need for a differentiated therapeutic approach to address the relative shift in cancer dependency from *ESR1* and *CDK4* in ER-driven tumors towards *CDK2* in ER-independent tumors.

### Experimental validation of therapeutic hypotheses in genetically modified cell lines

The ER driven and co-driven tumor subgroups (IC2-4) frequently harbored *ESR1* gain-of-function (GOF) mutations. In contrast, the ER-independent tumor subgroup is enriched in *RBI* and *TP53* mutations with low prevalence of *ESR1* mutations (Fig. 2g). These oncogenic mutations may be used to create preclinical models that harbor similar therapeutic vulnerability as the corresponding tumor archetypes in the clinical setting although these models cannot fully recapitulate the complex genomic profiles of individual tumors. Like the dependency pattern observed in ER-dependent tumors, perturbation response models predicted increased dependency on *ESR1* and *CDK4* compared to wild types in *ESR1* mutant tumors (Fig. 5d), suggesting that *ESR1* activating mutations do not confer resistance to CDK4 inhibition or ER ablation. The ER-Y537S mutation has been shown to promote constitutive ER signaling and induce resistance to fulvestrant<sup>30</sup>. To validate our in-silico prediction, we created an isogenic ER GOF mutant (Y537S) cell line from MCF7 and compared treatment effects by fulvestrant and palbociclib on the *ESR1* mutant and wild-type cell lines using an in vitro competitive growth assay. The ER GOF mutant cells maintained similar sensitivity to palbociclib as the parental MCF cells (Fig. 6a) but were resistant to fulvestrant (Fig. 6b), suggesting that post-treatment tumors that developed drug resistance through acquired *ESR1* mutations are still sensitive to CDK4 inhibition. Perturbation response models also predicted that *RBI* and *TP53* mutant tumors harbored increased dependency on *CDK2* when compared to the wild-type tumors (Fig. 5d). An isogenic RB LOF mutant cell line was created from the Luminal breast cancer cell line T47D, which endogenously harbored a *TP53* mutation L194F. We then performed colony formation assays and found that shRNA mediated *CDK2* knockdown resulted in greater reduction in cell viability in RB LOF mutant T47D compared to the parental cells (Fig. 6c, e), supporting the causality between *RBI* loss-of-function and increased *CDK2* dependency.

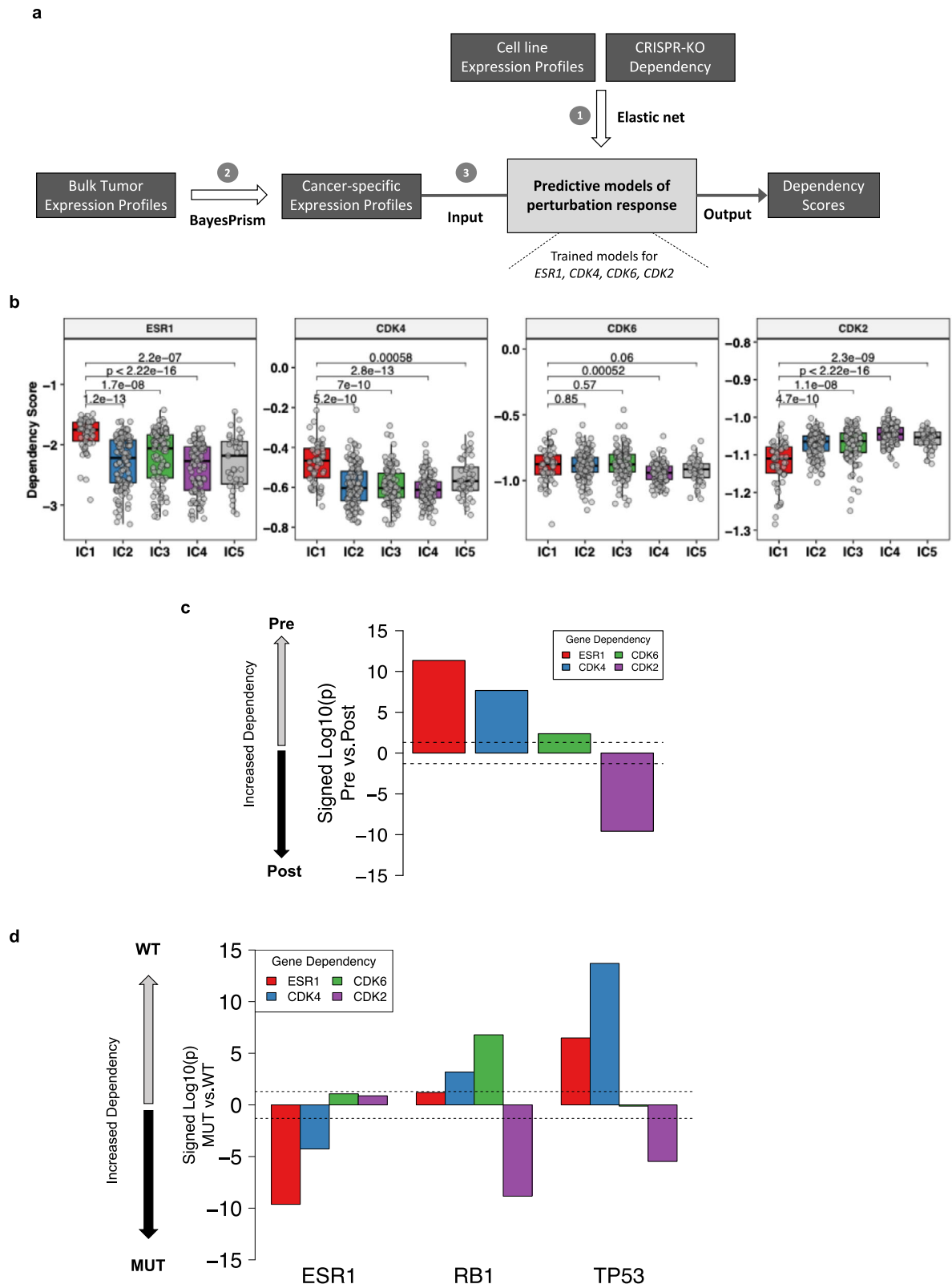
### Discussion

Our retrospective analysis of a real-world multi-omics dataset from metastatic BC patients who received CDK4/6i plus ET confirmed significant increases in the prevalence of *ESR1* and *RBI* mutations among post-progression tumors. Integrative analysis further identified three tumor subgroups—ER driven, ER co-driven and ER independent. The ER-independent subgroup is down-regulated in estrogen receptor mediated signaling and enriched CDK4/6i resistance markers including *RBI* LOF and *TP53* LOF mutation. In contrast, the ER-driven and co-driven subgroups harbored fewer *TP53* mutations but more frequent *ESR1*



**Fig. 4 | Trajectory inference analysis revealed a bifurcation of drug resistance mechanisms. a** Heatmap showing the tumor intrinsic expression patterns of 23 Hallmark gene signatures correlated with PT (Pearson correlation  $|r| > 0.35$ ). Shown in the top two panels are *ESR1*, *PGR* and *CCNE1* gene expression (Gene) and four Paloma3 NMF factors (Paloma3). Samples are sorted by PT value shown in the top bar plot, and Proliferative Index is shown in the bar plot below. **b** Elastic Principal Graph (EPG) analysis identified a topology comprised of 3 branches and 21 nodes. Dotted lines represent inferred trajectory 1 (A-B) and trajectory 2 (A-C).

**c** Distributions of EPG branches vs. integrative clusters. **d** Distributions of *ESR1*, *PGR*, and *CCNE1* expression vs. EPG branches, with sample sizes  $n = 118$  (A), 88 (B), and 84 (C). Statistical significance was determined using two-sided Wilcoxon rank sum test. Distributions of *TP53*, *ESR1*, *RBI*, *GATA3* mutation statuses (**e**) and *MYC*, *CCND1* and *FGFR1* amplification statuses (**f**) vs. branches. For all box-and-whisker plots, the box is bounded by the first and third quartile with a horizontal line at the median and whiskers extend to the maximum and minimum value. Source data are provided as a Source Data file.



mutations along with up-regulation of ER signaling signatures. Computational modeling and experimental validation demonstrated that tumors with RB LOF mutations harbored stronger dependency on *CDK2* whereas ER GOF tumors harbored stronger dependency on *ESR1* and *CDK4*. Altogether, our results suggest that precision medicine strategies are necessary for treating different patient segments whose tumors harbor distinct drug resistance mechanisms.

Our study revealed a dichotomy between ER GOF and RB LOF, two major mechanisms for acquired drug resistance. The standard of care is a combination of both CDK4/6i and ET. On the one hand, ER GOF confers resistance to ET but not to CDK4/6 inhibition, but on the other hand, RB LOF confers resistance to CDK4/6 inhibition but not to ET. Thus, one could infer that resistance to either single agent is sufficient to hamper the efficacy of the combination treatment and lead to

**Fig. 5 | Drug target gene dependencies changed in ER-independent vs. ER-driven tumors.** **a** Workflow for applying elastic net models trained on CRISPR loss-of-function knockout (KO) screen data in cell lines to predict gene-level dependency using tumor gene expression profiles. **b** Distributions of predicted dependency scores for the four drug target genes *ESR1*, *CDK4*, *CDK6*, and *CDK2* vs. integrative clusters, with sample sizes  $n = 57$  (IC1), 127 (IC2), 97 (IC3), 107 (IC4), and 39 (IC5). A lower score indicates stronger dependency and greater sensitivity to gene knockout. Statistical significance was determined using two-sided Wilcoxon rank sum test. **c** Relative change in dependency scores for the four drug target

genes in Post vs. Pre tumors. **d** Relative change in dependency scores for the four drug target genes vs. the mutation statuses of *ESR1*, *RBI*, and *TP53*. The y-axis shows the signed  $\log_{10}(p)$  values representing the statistical significance of the change in gene-level dependency scores between two subgroups, determined by two-sided Wilcoxon rank sum test. The dotted lines represent the significance cutoff ( $p < 0.05$ ). MUT: mutated. WT: wild type. For all box-and-whisker plots, the box is bounded by the first and third quartile with a horizontal line at the median and whiskers extend to the maximum and minimum value. Source data are provided as a Source Data file.

disease progression. Even though ET could be effective as monotherapy, our findings suggest that inhibitions of both resistance mechanisms would be the most effective treatment strategy for this indication. Based on the integrative clustering analysis, the vast majority (95%) of the pre-treatment tumors are dependent on ER signaling. Even in the post-treatment setting, most tumors (79%) retain dependency on the ER signaling pathway. Thus, treatment strategies targeting this pathway with improved therapies, including the next generation of anti-ER and CDK4/6 inhibitors, ought to improve outcomes for patients naïve to CDK4/6i plus ET and most patients who progressed on CDK4/6i plus ET. At the same time, we must take note that tumors can adopt alternative drug resistance mechanisms that no longer rely upon ER signaling.

Studies of endocrine resistance indicate that tumors under the pressure of anti-estrogen therapies can evolve via distinct paths including reactivating ER $\alpha$  signaling or bypassing ER $\alpha$  signaling altogether. Some cancers that lose ER dependency become driven by orthogonal signaling pathways such as HER2 or mitogen-activated protein kinases (MAPKs) while others lose ER $\alpha$  expression and show evidence of lineage plasticity<sup>24</sup>. An important insight from our study is that down-regulation of ER signaling is a common characteristic for a segment of patients whose tumors are enriched in various markers associated with CDK4/6i resistance and disease progression. The prevalence of this segment is very low before treatment (5%), indicating that it may have evaded detection previously due to a lack of molecular profiling data from post-treatment tumor samples. Our analyses further suggested that CDK2 inhibition may be more effective against ER-independent tumors when compared to ER driven or co-driven tumors. CDK2 is the canonical partner of cyclin E1, overexpression of which could lead to aberrant CDK2 activation and G1/S transition. CDK2 blockade has been shown to be effective in aggressive cancers such as high-grade serous ovarian cancers that harbored frequent amplification of *CCNE1* and overexpression of *MYC*<sup>31</sup>. Previously we also demonstrated the efficacy of a CDK2/4/6 inhibitor against palbociclib resistant models characterized by up-regulation of cyclin E1 and the MYC/E2F activation expression factor<sup>15</sup>, which was also observed for the ER independent tumors in our cohort (Fig. 2d, Supplementary Fig. 7c). Hence, our results indicate that patients who developed drug resistance via ER independent mechanisms could benefit from CDK2 inhibitors currently in clinical development.

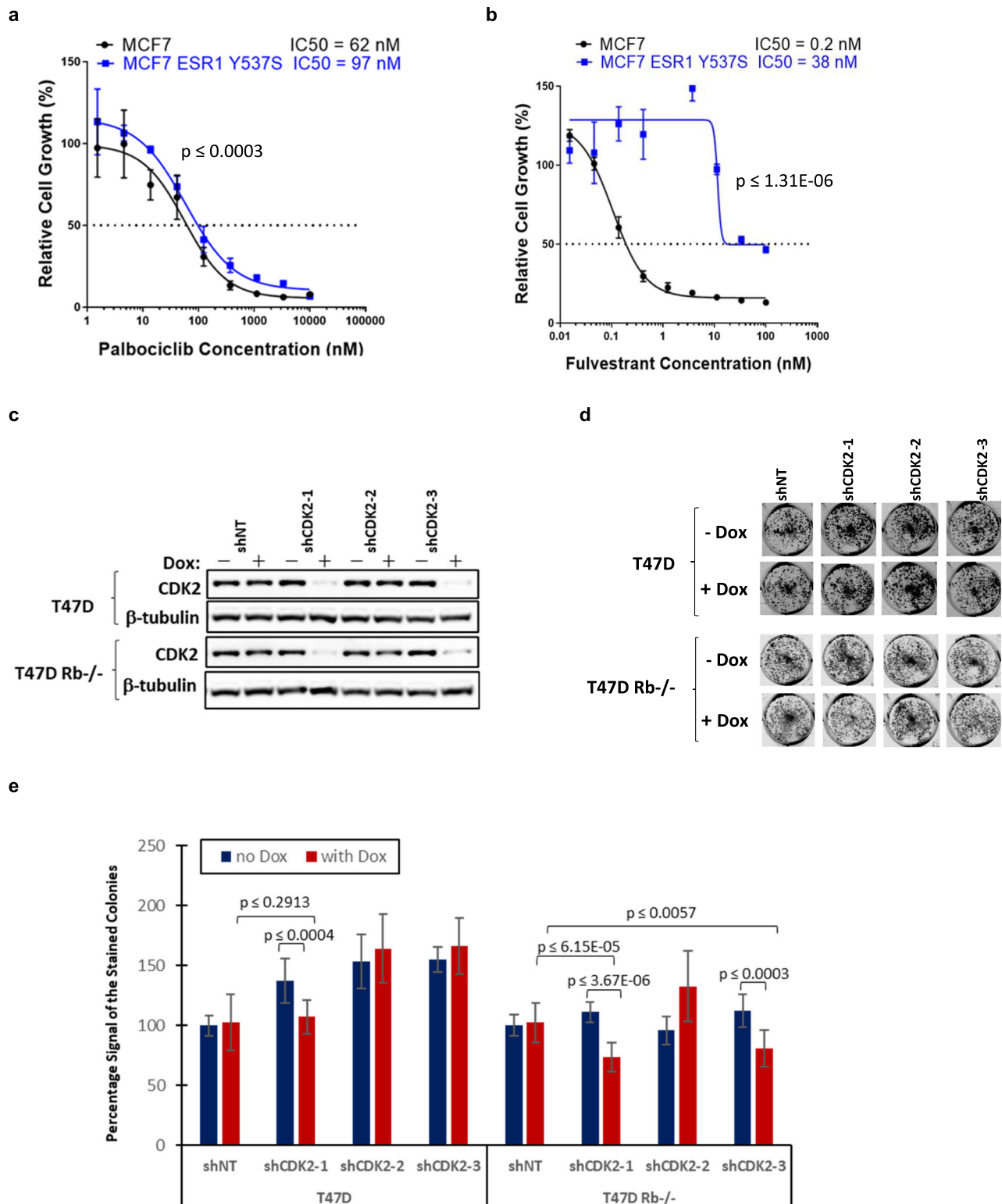
We noticed a distinctive segregation of intrinsic breast cancer subtypes based on the PAM50 classification vs. the three tumor subgroups with different dependency on ER signaling. While Basal and Her2 subtypes are enriched in ER-independent tumors (IC1), Luminal B is enriched in ER co-driven (IC2-3) and Luminal A is enriched in the ER driven subgroups (IC4-5). The basal-like subset of the ER-independent tumors up-regulate tumor intrinsic EMT and inflammatory signatures. Moreover, the ER-independent subgroup, similar to Her2-like and Basal-like subtypes, is enriched in *TP53* mutations but has lower frequency of *ESR1* mutations. The evolution of resistance towards increased ER signaling driven by ER GOF mutations aligns well with the paradigm of punctuated evolution caused by selection of drug-resistant mutations. However, it is less clear what mechanisms drive cancer cells towards the ER-independent state, and why *TP53*

mutations are associated with this molecular state and are mutually exclusive with ER GOF.

Intrinsic breast tumor subtypes are believed to originate from epithelial subpopulations segregated along the normal differentiation hierarchy of mammary epithelial cells<sup>32</sup>. The basal-like tumors harbor similar expression profile as the luminal progenitor cells while Her2 and Luminal subtypes are mapped to different subsets within the Luminal lineage, with Luminal A most resembling the mature luminal cells. By co-opting mechanisms of dedifferentiation, trans-differentiation and epithelial-to-mesenchymal transition, cancer cell plasticity enables cancer cells to transition across distinct cell states, such as from a differentiated state to a stem cell-like state, and contributes to tumor initiation, progression and drug resistance<sup>33</sup>. In breast cancer, induced *TP53* loss has been shown to promote cell plasticity and predispose mouse mammary luminal cells to developing mammary tumors with stem-like cellular state<sup>34</sup>. A multi-omics study of longitudinally paired primary and metastatic tumors reported that *TP53* mutation was associated with intrinsic subtype switching from Luminal A/B to HER2 in metastatic breast cancers<sup>35</sup>. It has also been demonstrated that luminal breast tumors expressing lower levels of ER represent a distinct subset characterized by basal-like features that arose from luminal cells through cellular plasticity<sup>25</sup>. Hence, it is possible that under the selective pressure of CDK4/6i plus ET, *TP53* mutations predispose luminal breast cancers to evolving into a non-luminal molecular state through reversal of the normal epithelial differentiation trajectory.

We applied machine learning to estimate cancer-specific expression profiles and model evolution of cancer-specific resistance to CDK4/6i plus ET. Subsequently, trajectory inference analyses revealed a latent descriptor variable of disease progression called pseudotime and two bifurcated evolutionary paths towards drug resistance—ER gain-of-function vs. ER independence. Our analysis indicated that ER-independence is a continuum strongly correlated with the pseudotime of disease evolution driven by CDK4/6i plus ET. Moreover, we identified a major disease evolution trajectory spanning two branches corresponding to increasing ER-independence and pseudotime. This sequential set of trajectory nodes approximates the continuous evolution through molecular states and resembles cancer cell developmental plasticity states in reverse order. Thus, ER-independence could result from reprogramming of luminal cancer cells into stem-like cells that become resistant to therapies targeting the ER signaling pathway through cancer cell plasticity.

This real-world clinical genomics study has added to the body of knowledge regarding the molecular landscape of the HR+/HER2-metastatic breast cancers in the post CDK4/6i plus ET setting. Our observations such as the frequency of resistance mechanisms may be limited in generalizability as the patients included in our cohort were selected for having a perceived benefit from genomic profiling and thus may not represent the broader population with HR+/HER2- mBC. Our study does not have a control arm, so it is difficult to ascertain whether any reported marker association is treatment-specific or prognostic. This real-world cohort is also heterogeneous in prior treatment status including both treatment naïve patients as well as patients who had failed one or more lines of therapies in the metastatic



**Fig. 6 | Experimental validation of therapeutic hypotheses.** Dose response curves evaluating the viability of MCF7 and MCF7 ESR1 Y537S cells after exposure to Palbociclib (a) or Fulvestrant (b) at different concentrations for 7 days. Cells treated with DMSO were used as control for normalization. Data shown on the y-axis are mean  $\pm$  SD ( $n = 3$ ). P-values were calculated by one-tailed student's t-test. c Western blot analysis was performed to determine the expression of CDK2 in the presence of 100 ng/ml Doxycycline. Cells were harvested at 4 days after the Doxycycline treatment and lysates were immunoblotted with the indicated antibodies. d Cell

Colony formation assay was conducted in the 12-well plates in the absence or presence of 50 ng/ml Doxycycline. Crystal violet staining of cells was performed after 14 days of cell growth. Three shRNAs against CDK2 and one non-targeted shRNA as control were tested. e Quantitative analysis of cells colony formation by scanning the intensities of crystal violet stained cell colonies using LI-COR Odyssey CLx Imaging System. Data shown on the y-axis are mean  $\pm$  SD ( $n = 12$ ) of six replicates in two independent experiments. P-values were calculated by one-tailed Student's t-test. Source data are provided as a Source Data file.

setting. Most of the patients in this cohort had de novo metastatic disease, which is higher than expected based on the epidemiology of breast cancer. Due to the inclusion of biopsies up to 1 year prior to CDK4/6i treatment or after disease progression, intervening therapies could alter the tumor molecular profiles in ways that would confound the Pre/Post comparison analyses intended for studying CDK4/6i resistance mechanisms. We should seek to replicate our findings in future studies with larger sample sizes and more focused treatment settings. Regardless of these caveats, our work has derived unique insights into the dichotomy of ER dependent vs. independent mechanisms of CDK4/6i resistance, highlighted the need for the next generation of anti-ER and cell cycle inhibitors and provided guidance for precision medicine strategies.

## Methods

### Patient enrollment and sample collection

This retrospective analysis used clinical and molecular real-world health data from the Tempus Database, which includes longitudinal structured and unstructured de-identified data from geographically diverse oncology practices ([www.tempus.com](http://www.tempus.com)). Real-world data included 400 breast cancer samples profiled with the Tempus xT assay and whole-transcriptome RNA sequencing, as previously described<sup>36–40</sup>.

### Genomic assessment

Tumor biopsy and matched normal samples were sequenced using the Tempus xT assay, a targeted NGS-based panel that detects single-nucleotide variants, insertions and/or deletions, and copy number variants in 648 cancer-related genes, as well as chromosomal rearrangements in 22 genes with high sensitivity and specificity<sup>36</sup>. Sample processing, library construction, sequencing, and variant classifications were performed by Tempus as previously described<sup>38,39</sup>.

### Gene expression data collection and normalization

Tempus RNA whole-transcriptome sample processing, library construction, and sequencing with a minimum depth of 30 million reads per sample were performed by Tempus as previously described<sup>39</sup>. Transcript level pseudo-alignment and quantification to the Ensembl GRCh37 Release 75 (July 2019) reference was performed using Kallisto (version 0.44). The transcript-level abundance was normalized using all mapped transcripts per million (TPM) obtained from the Kallisto pseudoalignment. The gene-level abundance was obtained by summing transcripts independent of the Ensembl label for all genes with at least one annotated protein coding transcript covered by the RNAseq assay.

### Detection of gene fusion

Gene fusions were determined from DNA and RNA sequencing data as previously described<sup>39</sup>. DNaseq data alignment utilized BWA and mapping to GRCh37, while RNAseq data utilized STAR aligner and mapping to GRCh38. DNA rearrangements were called if read support included at least 50 supporting reads, with a minimum of 2 of those reads being discordant pairs, and a minimum of 30 split reads. RNA rearrangements were called if  $\geq 1$  high-quality spanning reads (i.e., a significant fraction of the read was present on both the 5' and 3' sides of the breakpoint).

### Tumor mutational burden

TMB was calculated by dividing the number of non-synonymous mutations by the megabase size of the panel, as previously described<sup>39</sup>. All non-silent somatic coding mutations, including missense, indel and stop-loss variants with coverage greater than 100X and an allele frequency  $>5\%$  were included as non-synonymous mutations.

### Germline variant calls

Germline variant calls were determined utilizing the Tempus bioinformatics variant-calling algorithm and variant allele frequency<sup>39</sup>.

Germline alterations were classified using specific classification criteria and matched against the Tempus knowledge database, which includes recommendations by the American College of Medical Genetics, National Comprehensive Cancer Network Genetic/Familial High-Risk Assessment Guidelines, and other data sources<sup>41</sup>.

### Tumor purity

The tumor purity of each sample was determined utilizing two methods, a bioinformatics algorithm and expert pathology slide review. The final tumor purity estimate was determined via the Tempus bioinformatics copy number algorithm (CONA) for solid tumor samples and CNVkit for liquid biopsy samples. The final tumor percentage was determined by a pathologist based on the macrodissected specimen slide and takes into consideration the final tumor purity estimate.

### Genomic and molecular features

We applied a suite of multi-omics data analysis tools to compute 63 molecular features in 9 categories that characterize different aspects of breast cancer biology: (1) 5 summary features such as tumor mutation burden (TMB) and proliferative index; (2) 6 features based on PAM50 subtype classification; (3) 6 features for estimated immune cell fractions; (4) 4 features from tumor microenvironment and purity analysis; (5) 5 features based on mutation signature analysis; (6) 5 features based on genomics scar analysis; (7) 6 features for germline pathogenic mutation statuses; (8) 12 gene-expression factors identified by de novo NMF analysis of gene expression profiles; (9) 14 gene-expression factors identified by a previous NMF analysis of the PALOMA-3 gene expression profiles<sup>15</sup>.

(1) The proliferative index was calculated as the geometric mean of gene expression (TPM) using an 11-gene signature including *BIRC5*, *CCNB1*, *CDC20*, *NUF2*, *CEP55*, *NDC80*, *MKI67*, *PTTG1*, *RRM2*, *TYMS*, and *UBE2C*. The cytolytic activity score was calculated as the geometric mean of *GZMA* and *PRFI* expressions<sup>42</sup>. The tumor mutational burden (TMB) was calculated as the number of protein-altering mutations in each sample, including essential splice site, frameshift, in-frame indel, missense, nonsense, and stop-loss mutations. (2) PAM50 classification was performed using the “intrinsic.cluster.predict” function in the R package Genefu v2.14.0. For each tumor sample, the function calculated the Spearman correlation coefficients between the PAM50 gene expressions in the tumor and the PAM50 gene expression centroids of five PAM50 subtypes: Basal, Her2, Luminal A, Luminal B, and Normal-like. The tumor was then classified as the molecular subtype with the highest correlation. Differences in technology and normalization method between the microarray data used for training the original PAM50 classifiers and the Tempus RNA-Seq data could lead to inaccurate PAM50 classifications<sup>43</sup>. To adjust for this bias, we employed the subgroup-specific gene-centering method<sup>44,45</sup>. The ER+/HER2-subgroup-specific gene centering columns derived from the UNC232 training set, used for training the original PAM50 classifier, were used to correct the PAM50 gene expressions in our samples before running the PAM50 classifier. The UNC232 training set and the code used for correction were downloaded from the website (<https://unclineberger.org/peroulab/algorithms/>). (3) The estimated proportion of immune cell infiltrates compared to tumor and stroma cells, as well as the relative proportion of immune cell subtypes were determined utilizing RNA sequencing data as previously described<sup>39</sup>. (4) Tumor purity, stromal and immune cell fractions were calculated using ESTIMATE v1.0.13<sup>46</sup> based on gene expression profiles. (5)

The panel sequencing mutational signature analysis was performed using SigMA v1.0<sup>47</sup>, by setting tumor type to breast, and `check_msi` to true, the NNLS output was extracted using function `get_sig_exps` for downstream analysis. (6) The calculated genomic scar features include large-scale transitions (LST), telomeric allelic instability (TAI), and loss of heterozygosity (HRD-LOH) using allele-specific copy number, tumor purity, and ploidy inferred by

FACETS<sup>10,48</sup>. The HRD index was derived as the unweighted sum of LST, TAI, and HRD-LOH. (7) The germline variants were filtered by focusing on 19 breast cancer related genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CDHL*, *CHEK2*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *NBN*, *NF1*, *PALB2*, *PMS2*, *PTEN*, *RADS1C*, *RADS1D*, *STK11*, *TP53*). In addition, we required the variants to be annotated as “Pathogenic” or “Pathogenic/Likely Pathogenic” in Clinvar database version 20220213. (8) The NMF algorithm factorizes the gene expression matrix  $V$  ( $g$  genes by  $s$  samples) into two non-negative matrices  $W$  and  $H$ , where  $W$  is gene factor matrix ( $g$  genes by  $k$  factors) representing the different expression pattern of the  $k$  factors, and  $H$  is sample factor matrix ( $k$  factors by  $s$  samples) representing the contribution of each of  $k$  factors in each sample<sup>49</sup>. We removed 5 samples with tissue origin from blood or bone marrow or classified as normal-like subtype before NMF analysis. NMF was then performed on gene expression TPM matrix with the R package “NMF” v0.21.0 with “brunet” algorithm. To determine the optimal  $k$ , we used the copnetic coefficient and chose  $k = 12$  to balance the coefficient score and number factors derived<sup>50</sup>. (9) We projected the 14 NMF gene expression factors derived from PALOMA-3 HTG data<sup>15</sup> onto the gene expression profiles using metagene projection method<sup>51</sup>. To maintain consistency with PALOMA-3 NMF analysis, the gene expression values for each case were first rank transformed before projection.

To evaluate treatment induced expression changes in key pathways implicated in tumor biology, we calculated gene signature scores for 50 Hallmark genesets from the molecular signatures database (MsigDB) v7.2<sup>14</sup> using the gene set variation analysis (GSVA) R package v1.32.0<sup>52</sup>. Linear mixed effect regression (LMER) analysis was then performed to identify numerical features and expression signatures with significant changes Pre vs. Post after adjusting for tumor purity and metastatic tissue sites. Two-tailed fisher’s exact test and chi-squared test were performed to identify significant changes in categorical features in Pre vs. Post. We also calculated PFS associations for molecular features and gene expression signatures at baseline to identify those associated with disease progression.

### Comparison analysis of genomic alteration frequencies

We selected 611 genes with DNA-Seq coverage  $\geq 400$  samples for genomic alteration analyses including somatic mutations, copy number amplification and deletion and gene fusions. A total of 546 genes were mutated, with 140 genes mutated  $\geq 10$  samples. 290 genes harbored copy number amplifications with 141 genes amplified  $\geq 10$  samples and 209 genes harboring copy number deletions with 14 genes deleted in  $\geq 10$  samples. Gene fusions were detected in 85 genes with ESRI and CCNE1 harboring gene fusions in  $\geq 10$  samples. Gene expression for 11,724 genes was analyzed after filtering out genes with low expression or variation.

To systematically compare the gene level alteration frequency difference between Post and Pre samples, we aggregated the genomic alteration events at the gene level per sample for each gene including mutations, copy number variation (CNV), germline and fusion events. The following filtering criteria were applied before frequency calculation. First, the mutations identified as false positives by further manual review were removed. Second, only the CNVs filtered by Tempus threshold (copy number  $\geq 7$  or  $=0$ ) were kept. Third, germline pathogenic mutation for selected genes after Clinvar filtering was included. Forth, for structural variations (SV) identified from RNASeq, any SVs with  $<5$  rearr\_num\_reads were removed due to low evidence. Fifth, for SVs identified from DNA, we first removed any SVs that have the exact identical breakpoints in different patients, as it is almost impossible to have somatic SV with identical genomic breakpoint in different patients. We then removed any DNA SVs with  $<3$  discordant reads support or SVs with the discordant reads found in normal samples. For the remaining DNA SV candidates, we manually reviewed to remove any SVs that have softclip reads evidence in normal samples at the breakpoints. We applied a

one-tailed Fisher exact test to test alteration enrichment in Post samples (Supplementary Data 5).

To calculate genomic alterations frequency for paired samples, a cohort of 26 patients with DNA sequencing available from paired Pre and Post samples were used. One pair was excluded due to missing the matched normal for the Post tumor sample. The same filtering criteria and frequency calculation were repeated for the 26 patients (Supplementary Data 5). We also calculated Post-specific genomic alteration frequency based on the 26 paired patients. A genomic alteration was deemed Post-specific if it was detected in the Post but not in Pre samples (Supplementary Data 5).

### PFS association analysis of clinico-genomic variables

Univariate Cox Proportional Hazard regression analysis was performed to identify clinico-genomic data variables and Hallmark gene signatures significantly associated ( $p < 0.05$ ) with PFS at pre-CDK using the *survival* R package v3.5-5, adjusting for line-of-treatment and patient age. Estimates of hazard ratios reflect the change in risk of progression by changing 1 unit of the continuous variables whereas for categorical variables the hazard ratio reflects the category risk compared to its most frequent category. FDR (q-value) was calculated using the Benjamini and Hochberg method. All statistical analyses were performed using R version 4.3.0, and  $p < 0.05$  was considered to be statistically significant.

### Integrative multi-omics clustering analysis

The objective of this analysis is to apply an unsupervised machine learning approach on different types of molecular features to examine whether our cohort can be stratified into subgroups that harbor distinct molecular mechanisms of drug resistance. To represent different aspects of breast cancer biology, we selected 67 multi-omics features including 32 genomic alteration statuses, 18 analytical features, 10 gene signatures and 7 gene-level expressions (Supplementary Data 6). We performed integrative clustering on these features using iClusterPlus<sup>17</sup> while varying the numbers of clusters ( $k$ ) from 2 to 8. To identify the optimal  $k$ , we examined the Bayesian information criteria (BIC) and log-likelihood based deviance ratio which tend to increase with larger  $k$ . We chose the median of 5 clusters to balance having too few clusters that cannot sufficiently capture molecular heterogeneity vs. too many clusters that splinter into smaller subsets that are difficult for biological interpretation. For the 5 clusters (IC1-5), we then performed association analyses with the full set of molecular features to evaluate whether any clusters exhibited distinctive association patterns vs. disease progression. We excluded IC5 as it consisted of diverse tissue origins for the tumor biopsies that likely confounded bulk expression-based features and caused its separate classification.

### Cell-type deconvolution of bulk RNA-seq data using BayesPrism

Cell-type deconvolution of the bulk RNA-seq data was performed using BayesPrism<sup>18</sup>. The raw count matrix from the Tempus Breast Cancer RNA-seq data was used as the input bulk expression matrix and the raw count matrix from a breast cancer scRNA-seq dataset<sup>53</sup> was used as the input reference matrix. The cells with low library size  $<1000$  UMI were excluded and the cell types with less than 50 cells were excluded. A total of 96,831 single cells passed BayesPrism input reference scRNA-seq data criteria. For cell type and cell subtype labels, 17 broad cell type annotations were used, and 117 more refined cell subtype annotations were used. Among 117 more refined cell subtype annotations, 93 malignant cell clusters were included as cell subtype annotation after intra-patient malignant cell clustering with log-normalized reads. The “run.Ted” function was run with the ‘scRNA’ option and mitochondrial and ribosomal protein genes were removed as suggested by the BayesPrism authors. Normalized cancer cell-specific expression matrix from the output was used for downstream cancer cell-specific analyses. A limitation of this approach is that we used the scRNA-seq data

derived from primary breast cancers<sup>8</sup> to deconvolve the bulk expression profiles from metastatic breast cancers in the Tempus cohort. However, there should be robust deconvolution due to the stronger similarity in expression profiles between primary and metastatic tumors when compared to that between tumor and non-tumor compartments in the tumor microenvironment.

### Differential expression analysis

The expression features including individual gene expressions or gene signature scores were compared between Pre and Post, to identify features with significant differences. The expression feature changes over treatment time were modeled using linear mixed-effects regression model

$$y = \beta_0 + \beta_1 t + \beta_2 p + \beta_3 o + Z d + \varepsilon \quad (1)$$

Where  $y$  is the observed expression value for a specific feature across all samples,  $\beta_0$  is the intercept,  $\beta_1$  is the fixed effect of treatment time  $t$  on expression,  $\beta_2$  is the effect of tumor purity  $p$ ,  $\beta_3$  estimate the differences due to different tissue origin  $o$ ,  $Z$  is a binary design matrix indicating if the sample are from the same patient or not,  $d$  is a vector of random effect for donor patients, which is normally distributed with mean zero and represents the deviation from the overall mean of the mean feature value for each donor patient, and  $\varepsilon$  is random errors. The significance of  $\beta_1$  was assessed by Chi-squared tests to evaluate the association between treatment time  $t$  and expression features. The *lmerTest* R package v3.1-2 was used for the mixed-effects regression model fitting. To test cancer specific expression feature difference (individual gene expressions or gene signature scores) across PAM50 subtypes, or IC clusters, a similar approach was applied. For PAM50 subtype, the mixed-effects regression model is formulated as

$$y = \beta_0 + \beta_{basal} I_{basal} + \beta_{her2} I_{her2} + \beta_{lumA} I_{lumA} + \beta_{lumB} I_{lumB} + \beta_2 p + \beta_3 o + Z d + \varepsilon \quad (2)$$

where  $I_{basal}$ ,  $I_{her2}$ ,  $I_{lumA}$ , and  $I_{lumB}$  are indicator variables for each subtype, the regression coefficients  $\beta_{basal}$ ,  $\beta_{her2}$ ,  $\beta_{lumA}$ , and  $\beta_{lumB}$  are fixed effects for each subtype. For IC clusters, the mixed-effects regression model is formulated as

$$y = \beta_0 + \beta_{IC1} I_{IC1} + \beta_{IC2} I_{IC2} + \beta_{IC3} I_{IC3} + \beta_{IC4} I_{IC4} + \beta_2 p + \beta_3 o + Z d + \varepsilon \quad (3)$$

where  $I_{IC1}$ ,  $I_{IC2}$ ,  $I_{IC3}$ , and  $I_{IC4}$  are indicator variables for each IC cluster, the regression coefficients  $\beta_{IC1}$ ,  $\beta_{IC2}$ ,  $\beta_{IC3}$ , and  $\beta_{IC4}$  are fixed effects for each IC clusters.

### Predictive modeling of cancer dependencies

We trained elastic-net based models (ENRS) to predict gene dependency scores in cancer cell lines. For training of gene dependency models, DepMap 21Q1 CERES scores were used as response variable and DepMap 20Q3 gene expression profiles were used as input data<sup>54</sup>. To derive a robust prediction model, we first applied elastic net and cross validation to the training data to look for the optimal hyperparameters using R package *glmnet* v4.0-2. The default range of hyperparameter  $\lambda$  was determined by *glmnet*. A sequence of values between 0.1 and 1 with increment 0.1 was used for hyperparameter  $\alpha$  values. The optimal  $\alpha$  and  $\lambda$  values with minimum mean cross-validated error were then used to select robust gene expression features by bootstrap method. Specifically, we randomly collected the same number of observations with replacement from training data and repeat the process 100 times, each time we fit a new elastic net model with the bootstrap samples and optimal  $\alpha$  and  $\lambda$  values. The bootstrapping frequency for each gene was defined as frequency of elastic net models with the gene selected. To include only robust relevant features and increase model generalizability, we kept only gene

features with bootstrapping frequency  $>0.9$ . We then derived the final linear regression model using selected gene features by ordinary least square regression (OLS). The model performance was evaluated using Pearson correlation coefficient (PCC) between the responses and predictions, and only models with  $PCC \geq 0.3$  was used for predictions. If a model has  $PCC < 0.3$ , we also tried to get lower confident models by relaxing the cutoff to include more genes with bootstrapping frequency  $>0.8$ ,  $>0.7$ ,  $>0.6$ , or  $>0.5$ , until the model reached  $PCC \geq 0.3$ , or else it was considered unreliable to predict perturbation response from gene expression. For clinical tumor analysis, cancer-specific expression profiles (CSE) were used as input to the model to closely match the expression profile of cancer cell lines and minimize the confounder effect of various non-tumor cells in the tumor microenvironment. The model then used each tumor CSE to estimate a score for each of the four target genes similar to the CERES dependency score<sup>29</sup>, where lower score indicates a higher likelihood that eliminating the gene function would reduce cancer cell viability.

### Pseudotime and trajectory inference analysis using Monocle

We used Monocle 3<sup>26</sup> to construct a tree-like trajectory of all tumors, ordered samples along this trajectory, and assigned a numeric value referred to as pseudotime to each sample to indicate how far it progresses along this dynamic process. The gene-sample cancer-specific expression matrix was used as the input for Monocle 3. We excluded liver biopsies from subsequent trajectory inference analyses to remove the confounding effect due to the distinct transcriptional profiles of liver tissues. Principal Components Analysis was run to select the top 50 PCs. The top 2000 most variable genes were used to construct the pseudotime trajectory. Dimensionality reduction was implemented using `reduceDimension()` which uses the UMAP algorithm. Samples were then clustered using `cluster_cells()`, based a community detection technique called “Leiden algorithm”<sup>55</sup> and visualized in the UMAP space. Our data resulted in 1 cluster which indicates that there is no batch effect from different data batches nor tumor origins. `LearnGraph()` function was then applied to learn a principal graph that resides in the same UMAP space and output the paths samples took as they progressed. Lastly, Monocle 3 used the principal graph as a guide and computed the pseudotime of each sample as its geodesic distance back to one root node in the trajectory. The root nodes were selected based on the UMAP graph region enriched with pre-treatment samples.

### Disease trajectory analysis using Elastic Principal Graphs

Elastic Principal Graphs (EPG) method originates from principal manifolds introduced by Hastie et al<sup>56</sup>. Principal manifolds are the surfaces passing through the densest area of the multi-dimensional cloud of data points. Multiple manifolds form tree-like structures characterized by branching topology, so called principal tree. Elastic Principal Graphs can be used to approximate such principal trees. Elastic Principal Graphs are data approximators which consists of the undirected graph embedded in the high dimensional space, thus defining the manifold surface. The graphs are constructed to minimize mean squared distance from the data points to the nodes of the graph similar to k-means clustering. However, the embedded graph provides an additional tree-like structure contrary to k-means. The graph is modeled as a system of connected strings, which connect the nodes of the graph allowing the segments of the graph to bend and stretch. The energy of the graph is defined as a function of both elastic energy of the graph and k-mean-like penalty between the nodes and data points. This creates the optimization problem to minimize the total energy which can be solved analytically. To find the optimal graph structure, first, topological grammar which defines the graph operations is used to generate the set of possible graph structures. Secondly, the space of generated graph structures is explored with the optimization algorithm which directs the search in a manner similar to the gradient-



decent search. After removing liver biopsies, we performed trajectory analyses on cancer-specific expression profiles of 298 samples. First, we selected 5000 most variable genes for which 40 principal components were computed. The principal tree was computed using the ClinTrajan and ElPiGraph packages for Python<sup>57</sup>. We varied the number of nodes for ElPiGraph from 20 to 100 and observed that the learned principal trees look similar, and the larger numbers of nodes reveal more bifurcations. Given the limited sample size, we chose the principal tree with 20 nodes. We also applied tree pruning to remove branches consisting of only one node.

### Cell Lines and Culture

ER + BC cell lines (MCF7 and T47D) were obtained from the American Type Culture Collection (ATCC). T47D Rb<sup>-/-</sup> cells were acquired from UT Southwestern. MCF7 ESRI Y537S was a subclone of MCF7 that CRISPR knock-in ESRI with Y537S mutation. All cells were cultured with RPMI-1640 media supplemented with 10% FBS and penicillin-streptomycin in a humidified incubator at 37 °C with 5% CO<sub>2</sub>. Tet-inducible Lentiviral Expression of shRNA System were used to knockdown human CDK2 in T47D and T47D Rb<sup>-/-</sup> cells. The cells of T47D (or T47D Rb<sup>-/-</sup>)/shCDK2 and /shNT were generated using shRNA against CDK2 (target sequence: shCDK2-1: 5'-CCGAGA-GATCTCTCTGCTTAA-3'; shCDK2-2: 5'-ACGACCCTAACAAGCGGA TTT-3'; shCDK2-3: 5'-GCCTTCTACACGTTAGATT-3'), LUC (target sequence: 5'-CGCTGAGTACTCGAATGTC-3') as non-target control (shNT) and selected using 1.5 µg/ml Puromycin and CDK2 protein knockdown was validated using Western blotting.

### Cell Proliferation Assay

Cells were seeded in 96-well plates and treated with serial dilution of Palbociclib or Fulvestrant for 7 days in an incubator at 37 °C with 5% CO<sub>2</sub>. The CyQUANT™ Direct Cell Proliferation Assay (ThermoFisher Scientific) was performed according to manufacturer's recommendations and the fluorescence signal was read on Tecan Spark Multi-mode Microplate Reader (Tecan) at 485 nm excitation wavelength/535 nm emission wavelength. Relative cell growth was normalized to DMSO alone treatment, and IC<sub>50</sub> curves were generated using Graph-Pad Prism® software. The results shown represent average of readings from three wells per data point.

### Western blotting

Cells were lysed with Cell Lysis Buffer (Invitrogen™, ThermoFisher Scientific) containing protease and phosphatase inhibitor cocktail (Sigma-Aldrich), followed by centrifugation at 13,000 rpm at 4 °C for 10 min. The protein concentration of the lysates was determined using Bradford Assay (Bio-Rad Laboratories). Equal amounts of protein from each cell lysate were denatured, separated on SDS-PAGE and transferred to nitrocellulose membranes, blocked in 5% non-fat milk in 1x TBST and reacted with antibodies against CDK2 (Cell Signaling Technology), β-tubulin (Cell Signaling Technology). The membranes were then washed with TBST (Cell Signaling Technology), incubated with HRP-conjugated anti-rabbit IgG (Cell Signaling Technology) secondary antibodies, and the target proteins were detected with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific).

### Colony formation assay

T47D/shNT, T47D/shCDK2, T47D Rb<sup>-/-</sup>/shNT, and T47D Rb<sup>-/-</sup>/shCDK2 cells were seeded in 12-well plates at the density of 5000 cells per well, maintained in growth medium in the absence or presence of 50 ng/ml Doxycycline. Colonies were allowed to form for 2 weeks, and the media were replenished every 3–4 days. After 2 weeks, the colonies were fixed with methanol, stained with 0.1% crystal violet (Sigma-Aldrich). The images of crystal violet-stained colonies were scanned on LICOR Odyssey CLx Imaging system. The quantification of image intensities was analyzed with Image Studio Software.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Tempus deidentified data used in this research including clinical outcome and tumor sequencing data was collected in a real-world health care setting and is subject to controlled access for privacy and proprietary reasons. When possible, derived data supporting the findings of this study have been made available within the paper and its supplementary data. Restrictions apply to the availability of additional data, which were used under license for this study. Source data are provided with this paper. Tempus may make access to further data available pending a signed data use agreement. Requests for access should be sent to publication.inquiry@tempus.com. Source data are provided with this paper.

### Code availability

Code for data cleaning and analysis is provided as part of the replication package. It is available at: <https://doi.org/10.6084/m9.figshare.27085321>

### References

- Howlader, N., et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J. Natl. Cancer Inst.* **106**, dju055 (2014).
- Reinert, T. & Barrios, C. H. Optimal management of hormone receptor positive metastatic breast cancer in 2016. *Ther. Adv. Med Oncol.* **7**, 304–320 (2015).
- Eggersmann, T. K., Degenhardt, T., Gluz, O., Wuerstlein, R. & Harbeck, N. CDK4/6 inhibitors expand the therapeutic options in breast cancer: palbociclib, ribociclib and abemaciclib. *BioDrugs* **33**, 125–135 (2019).
- Pandey, K. et al. Molecular mechanisms of resistance to CDK4/6 inhibitors in breast cancer: a review. *Int J. Cancer* **145**, 1179–1188 (2019).
- Formisano, L. et al. Aberrant FGFR signaling mediates resistance to CDK4/6 inhibitors in ER+ breast cancer. *Nat. Commun.* **10**, 1373 (2019).
- O'Leary, B. et al. The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 Trial. *Cancer Discov.* **8**, 1390–1403 (2018).
- Rao, X., et al. Clinical and genomic characteristics of patients with HR+, HER2- metastatic breast cancer following progression on a CDK4 and 6 inhibitor. *Clin. Cancer Res.* **29**, 3372–3383 (2023).
- Wander, S. A. et al. The genomic landscape of intrinsic and acquired resistance to cyclin-dependent kinase 4/6 inhibitors in patients with hormone receptor-positive metastatic breast cancer. *Cancer Discov.* **10**, 1174–1193 (2020).
- Turner, N. C. et al. Cyclin E1 expression and palbociclib efficacy in previously treated hormone receptor-positive metastatic breast cancer. *J. Clin. Oncol.* **37**, 1169–1178 (2019).
- Park, Y. H. et al. Longitudinal multi-omics study of palbociclib resistance in HR-positive/HER2-negative metastatic breast cancer. *Genome Med.* **15**, 55 (2023).
- Yang, C. et al. Acquired CDK6 amplification promotes breast cancer resistance to CDK4/6 inhibitors and loss of ER signaling and dependence. *Oncogene* **36**, 2255–2264 (2017).
- Gomez Tejada Zanudo, J. et al. Exemestane plus everolimus and palbociclib in metastatic breast cancer: clinical response and genomic/transcriptomic determinants of resistance in a phase I/II trial. *Nat. Commun.* **15**, 2446 (2024).
- Johnson, B. E. et al. An omic and multidimensional spatial atlas from serial biopsies of an evolving metastatic breast cancer. *Cell Rep. Med* **3**, 100525 (2022).

14. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
15. Freeman-Cook, K. et al. Expanding control of the tumor cell cycle with a CDK2/4/6 inhibitor. *Cancer Cell* **39**, 1404–1421 e1411 (2021).
16. Li, Z. et al. Mutual exclusivity of ESR1 and TP53 mutations in endocrine resistant metastatic breast cancer. *NPJ Breast Cancer* **8**, 62 (2022).
17. Mo, Q. et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl. Acad. Sci. USA* **110**, 4245–4250 (2013).
18. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).
19. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).
20. Chan, J. M. et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**, 1180–1191 (2022).
21. Lv, L. et al. Relationship between metabolic reprogramming and drug resistance in breast cancer. *Front. Oncol.* **12**, 942064 (2022).
22. Payne, K. K. Cellular stress responses and metabolic reprogramming in cancer progression and dormancy. *Semin. Cancer Biol.* **78**, 45–48 (2022).
23. Quintanal-Villalonga, A. et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).
24. Will, M., Liang, J., Metcalfe, C. & Chandarlapaty, S. Therapeutic resistance to anti-oestrogen therapy in breast cancer. *Nat. Rev. Cancer* **23**, 673–685 (2023).
25. Mohamed, G. A. et al. Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits. *Breast Cancer Res.* **25**, 23 (2023).
26. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
27. Albergante, L., et al. Robust and scalable learning of complex intrinsic dataset geometry via ELPIGraph. *Entropy* **22**, 296 (2020).
28. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 e516 (2017).
29. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
30. Harrod, A. et al. Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. *Oncogene* **36**, 2286–2296 (2017).
31. Gerosa, R. et al. Cyclin-dependent kinase 2 (CDK2) inhibitors and others novel CDK inhibitors (CDKi) in breast cancer: clinical trials, current impact, and future directions. *Crit. Rev. Oncol. Hematol.* **196**, 104324 (2024).
32. Fu, N. Y., Nolan, E., Lindeman, G. J. & Visvader, J. E. Stem cells and the differentiation hierarchy in mammary gland development. *Physiol. Rev.* **100**, 489–523 (2020).
33. Perez-Gonzalez, A., Bevant, K. & Blanpain, C. Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat. Cancer* **4**, 1063–1082 (2023).
34. Tao, L., Xiang, D., Xie, Y., Bronson, R. T. & Li, Z. Induced p53 loss in mouse luminal cells causes clonal expansion and development of mammary tumours. *Nat. Commun.* **8**, 14431 (2017).
35. Aftimos, P. et al. Genomic and transcriptomic analyses of breast cancer primaries and matched metastases in AURORA, the Breast International Group (BIG) molecular screening initiative. *Cancer Discov.* **11**, 2796–2811 (2021).
36. Tempus Labs, I. Tempus xT gene panel list and validation. <https://www.tempus.com/oncology/genomic-profiling/xt-xr/> (2023).
37. Fernandes, L. E. et al. Real-world evidence of diagnostic testing and treatment patterns in US patients with breast cancer with implications for treatment biomarkers from RNA sequencing data. *Clin. Breast Cancer* **21**, e340–e361 (2021).
38. Beaubier, N. et al. Clinical validation of the tempus xT next-generation targeted oncology sequencing assay. *Oncotarget* **10**, 2384–2396 (2019).
39. Beaubier, N. et al. Integrated genomic profiling expands clinical options for patients with cancer. *Nat. Biotechnol.* **37**, 1351–1360 (2019).
40. Gao, Y. et al. Comparison of the tumor immune microenvironment and checkpoint blockade biomarkers between stage III and IV non-small cell lung cancer. *Cancer Immunol. Immunother.* **72**, 339–350 (2023).
41. Yap, T. A. et al. Prevalence of germline findings among tumors from cancer types lacking hereditary testing guidelines. *JAMA Netw. Open* **5**, e2213070 (2022).
42. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
43. Prat, A. & Parker, J. S. Standardized versus research-based PAM50 intrinsic subtyping of breast cancer. *Clin. Transl. Oncol.* **22**, 953–955 (2020).
44. Fernandez-Martinez, A. et al. Survival, pathologic response, and genomics in CALGB 40601 (Alliance), a neoadjuvant phase III trial of paclitaxel-trastuzumab with or without lapatinib in HER2-positive breast cancer. *J. Clin. Oncol.* **38**, 4184–4193 (2020).
45. Zhao, X., Rodland, E. A., Tibshirani, R. & Plevritis, S. Molecular subtyping for clinically defined breast cancer subgroups. *Breast Cancer Res* **17**, 29 (2015).
46. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
47. Gulhan, D. C., Lee, J. J., Melloni, G. E. M., Cortes-Ciriano, I. & Park, P. J. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nat. Genet.* **51**, 912–919 (2019).
48. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
49. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
50. Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA* **101**, 4164–4169 (2004).
51. Tamayo, P. et al. Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc. Natl. Acad. Sci. USA* **104**, 5959–5964 (2007).
52. Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
53. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
54. Pacini, C. et al. Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021).
55. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
56. Hastie, T. & Stuetzle, W. Principal curves. *J. Am. Stat. Assoc.* **84**, 502–516 (1989).
57. Golovenkin, S. E., et al. Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data. *Gigascience* **9**, gja128 (2020).

## Acknowledgements

The authors would like to thank Tempus for providing the study dataset and technical support. The authors also would like to thank the following Pfizer colleagues for providing valuable feedback: Feng Liu, Yuan Liu, Lynn McRoy, Clifford Restaino and Won-Chul Lee.

## Author contributions

J.B. and P.A.R. conceived and designed the study. J.W. performed clinical data QC and analysis. Z.Y.K., J.W., V.B., W.Y., V.I., W.R. and X.J.M. performed computational analyses. K.H.K., C.L., J.L.M., J.O. and T.V. performed experimental validations. Z.Y.K., J.W., V.B., W.Y., V.I., W.R., C.L., J.L.M. contributed to methodology development and provided method descriptions. Z.Y.K., J.W., V.B., W.Y., V.I., K.H.K. and C.L. provided tables and figures. Z.Y.K., J.B. and P.A.R. wrote and revised the manuscript.

## Competing interests

The authors declare no competing non-financial Interests but the following competing financial interests. All authors were employees of Pfizer Inc. at the time the work was performed.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-55914-x>.

**Correspondence** and requests for materials should be addressed to Zhengyan Kan or Jadwiga Bienkowska.

**Peer review information** *Nature Communications* thanks Vessela Kristensen, Charlotte Ng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025