Clonal transcriptomics identifies mechanisms of chemoresistance and empowers rational design of combination therapies.

Sophia A Wild[1]†, Ian G Cannell[1]†, Katarzyna Kania[1], Ashley Nicholls[1], Dario Bressan[1], CRUK IMAXT Grand Challenge Team[1], Gregory J Hannon[1]*, Kirsty Sawicka[1]*

[1] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge. CB2 0RE. UK.

† these authors contributed equally to this work

* to whom correspondence should be addressed Greg.Hannon@cruk.cam.ac.uk or Kirsty.Sawicka@cruk.cam.ac.uk

Abstract

Tumor heterogeneity is thought to be a major barrier to successful cancer treatment due to the presence of drug resistant clonal lineages. However, identifying the characteristics of such lineages that underpin resistance to therapy has remained challenging. Here we utilize clonal transcriptomics with WILD-seq; **W**holistic **I**nterrogation of **L**ineage **D**ynamics by **seq**uencing, in mouse models of triple-negative breast cancer (TNBC) to understand response and resistance to therapy, including BET bromodomain inhibition and taxane-based chemotherapy. This analysis revealed oxidative stress protection by NRF2 as a major mechanism of taxane resistance and led to the discovery that our tumor models are collaterally sensitive to asparagine deprivation therapy using the clinical stage drug L-asparaginase after frontline treatment with docetaxel. In summary, clonal transcriptomics with WILD-seq identifies mechanisms of resistance to chemotherapy that are also operative in patients and pin points asparagine bioavailability as a druggable vulnerability of taxane resistant lineages.

Introduction

Intra-tumoral heterogeneity (ITH) is thought to underlie tumor progression and resistance to therapy by providing a reservoir of phenotypically diverse clonal lineages on which selective pressures from the microenvironment or therapeutic intervention exert their effects (Bhang et al., 2015; Turajlic & Swanton, 2016). Inference of clonal composition from bulk sequencing has elucidated the breadth of ITH across tumor types and suggests that often rare pre-existing clones can resist therapy-induced killing to drive relapse (Dentro et al., 2021; Ding et al., 2012; Gerlinger et al., 2012; Jamal-Hanjani et al., 2014; Landau et al., 2013). However, such methods are limited by their inability to characterize such resistant clones beyond genotype and how their properties change over time and in response to therapy. Recently, several lineage tracing approaches have emerged that are able to link clonal identity with gene expression by utilizing expressed genetic barcodes that are read-out by single cell RNA sequencing (Biddy et al., 2018; Gutierrez et al., 2021; Quinn et al., 2021; Simeonov et al., 2021; Weinreb et al., 2020; Yang et al., 2022). These powerful methods allow deconvolution of complex mixtures of clones while simultaneously providing a gene expression profile of those cells that can indicate the pathways on which they depend. However, to date in solid tumors these technologies have mostly been used to study drug response *in vitro* (Gutierrez et al., 2021; Oren et al., 2021) or metastatic dissemination *in vivo* (Quinn et al., 2021; Simeonov et al., 2021; Yang et al., 2022) and have not been utilized to study therapeutic response in immune-competent models.

A thorough understanding of the biomarkers of sensitivity and mechanisms of resistance to chemotherapy is essential if we are to improve patient outcomes. Most existing combination cancer therapies are not rationally designed but were instead empirically optimized to avoid overlapping toxicities. More recently alternative therapeutic strategies have emerged including synthetic lethality, drug synergy (Al-Lazikani et al., 2012; O'Neil et al., 2017) and collateral sensitivity (Mueller et al., 2021; Pluchino et al., 2012; Zhao et al., 2016) that aim to leverage selective vulnerabilities of tumor cells while minimizing toxicity. Of particular promise is collateral sensitivity, in which as a tumor becomes resistant to one drug it comes at the cost of sensitivity to a second drug. Since many modern clinical trials occur in the context of neo-adjuvant chemotherapy, the identification of frontline therapy-induced collateral sensitivities to second line therapy would have the potential to be rapidly translated into improved outcomes for patients.

Here we develop WILD-seq (**W**holistic **I**nterrogation of **L**ineage **D**ynamics by **seq**uencing), an accessible and adaptable platform for lineage tracing at the single-cell transcriptomic level that facilitates *in vivo* analysis of clonal dynamics and apply it to the study of syngeneic triple negative breast cancer (TNBC) mouse models. Our optimized pipeline ensures recurrent representation of clonal lineages across animals and samples, facilitating analysis of clonal dynamics under the selective pressure of therapeutic intervention. Importantly, analysis of response of TNBC models to frontline taxane-based chemotherapy revealed an enrichment of clones with high levels of NRF2 signaling, implicating defense against oxidative damage as a major determinant of resistance to chemotherapy. Building on the work of others (LeBoeuf et al., 2020) we show that these NRF2-high, taxane-resistant, lineages are collaterally sensitive to asparagine deprivation with L-asparaginase and that they adapt to this second line intervention by up-regulating *de novo* asparagine synthesis through asparagine synthetase (Asns). Together these data indicate that high levels of NRF2 signaling, which is also observed in patients following neo-adjuvant chemotherapy, promotes both resistance to chemotherapy and sensitivity to asparagine deprivation and warrant the exploration of L-asparaginase as a therapeutic modality in solid tumors.

Results

Establishment of an expressed barcode system to simultaneously detect clonal lineage and gene expression.

WILD-seq uses a lentiviral library to label cells with an expressed, heritable barcode that enables identification of clonal lineage in conjunction with single cell RNA sequencing. The WILD-seq construct comprises a zsGreen transcript which harbours in its 3' untranslated region (UTR) a barcode consisting of two 12 nucleotide variable regions separated by a constant linker (Fig. 1a). Each variable region is separated from any other sequence in the library by a Hamming distance of 5 to allow for library preparation and sequencing error correction and our library contains over 1.5 million unique barcodes. The barcode is appropriately positioned relative to the polyadenylation signal to ensure its capture and sequencing by standard oligo-dT single cell sequencing platforms.

The standard WILD-seq pipeline is illustrated in Figure 1b. A heterogeneous cell line is transduced with a barcode library at low multiplicity of infection (MOI) to ensure that each cell receives a maximum of one barcode. An appropriate size pool of barcoded clones is selected and stabilized in culture. Empirically, we have found a pool established from 750 individual clones works well to provide effective representation of the diversity within the cell lines used herein while also enabling recurrent representation of the same clones across animals and experiments. Once stabilized in culture, the pool of WILD-seq clones can be analyzed directly by single cell sequencing or injected into a recipient animal for *in vivo* tumor growth. WILD-seq single cell sequencing libraries can be prepared using a standard oligo-dT based protocol and addition of an extra PCR amplification step can be used to increase coverage of the barcode region and aid cell lineage assignment.

We first established a WILD-seq clonal pool from the mouse 4T1 cell line, a triple negative mammary carcinoma model that can be orthotopically implanted into the mammary fat pad of a BALB/c syngeneic host, which we have previously shown to be heterogeneous with distinct sub-clones having unique biological properties (Wagenblast et al., 2015). We performed single cell sequencing of the *in vitro* WILD-seq pool (Fig. 1c) and *in vivo* tumors derived from this clonal pool (Fig. 1d). Over the course of our studies, we injected multiple cohorts of mice with our WILD-seq 4T1 pool as detailed in Supplementary Table 1, some of which were subjected to a specific drug regime. All tumors were harvested at humane endpoint, as determined by tumor volume unless otherwise stated and immediately dissociated for single cell sequencing.

For the purpose of characterising the baseline properties of our clones, we performed an in-depth transcriptomic analysis of all tumors from untreated and vehicle-treated animals. A WILD-seq barcode and thereby clonal lineage could be unambiguously assigned to 30-60% of cells per sample within the presumptive tumor cell/mammary epithelial cell cluster. 132 different WILD-seq barcodes were observed *in vitro* and in total 94 different WILD-seq barcodes were observed across our *in vivo* tumor samples. Our *in vivo* tumor samples comprised both tumor cells and host cells of the tumor microenvironment including cells of the innate and adaptive immune system, enabling simultaneous profiling of the tumor and its microenvironment (Fig. 1d). Clustering was performed after removal of reads mapping to the WILD-seq vector, to avoid any influence of the WILD-seq transcript on clustering, and the WILD-seq barcode assignment subsequently overlaid onto these data. The tumor cell clusters

were clearly identifiable by the high expression of the barcode transcript. Occasionally a barcode was observed in cells which clustered according to their transcriptome outside of the main tumor cluster. Since this could be the result of sequencing or technical error causing a mismatch between the WILD-seq barcode and the cell of origin, only barcoded cells that clustered within the main tumor/mammary epithelium cell cluster were included in our analysis.

We reproducibly observed the same clonal populations across animals and independent experiments which is critical to our ability to examine the effects of different interventions and treatments (Fig. 1d, 1e). The relative abundance of clones was similar in tumors grown in NOD scid gamma (NSG) immunodeficient and BALB/c immunocompetent mice but was drastically different to that found in the *in vitro* cell pool from which they were established (Fig. 1e, Supplementary Table 2), suggesting that clones that show greatest fitness in cell culture do not necessarily show fitness *in vivo*. Therefore, *in vitro* clonal lineage tracking experiments are likely to capture a different collection of clones and have the potential to identify sensitive or resistance clones that are not represented *in vivo*. Pseudo-bulk analysis of the major clonal lineages revealed that the composition of the tumor microenvironment has a dramatic effect on the transcriptome of the tumor cells for all clones (Fig. 1f). Comparison of *in vitro* culture, tumors from NSG mice, and tumors from BALB/c mice by principal component analysis (PCA), showed clear separation of the tumor cells depending on their environment, with differences in interferon gamma signaling, TNF-alpha signaling, and cell cycle being most prominent between cells grown *in vivo* and *in vitro* (PC1, Fig. 1g). Differences in gene expression between tumors growing in immunocompetent and immunodeficient hosts were related to changes in the expression of extracellular matrix proteins and changes in interferon gamma and Il-2 signaling, consistent with the differences in T-cell abundance (PC2, Fig. 1g). These data highlight the importance of the host immune system in sculpting the transcriptome and provide cautionary context for the analysis of tumor gene expression in immune-compromised hosts. Although there were large differences between clonal gene expression patterns across hosts the clones showed consistent differences in gene expression across all settings, reflective of intrinsic clonal properties, with the biggest variation in gene expression across the clones being related to their position along the epithelial-mesenchymal transition (EMT) axis (PC3, Fig. 1g). In particular, Clone 679 is the most distinct and the most mesenchymal of the clones.

To further characterize the major clones in our tumors, we performed gene set expression analysis using AUCell (Aibar et al., 2017) to identify pathways that are enriched in cells of a specific clonal lineage. Analysis was performed across four independent experiments each with three vehicle-treated animals and for the majority of clones we were able to identify distinct gene expression signatures that were reproducible across animals and experiments (Fig. 1h, Supplementary Table 4, Supplementary Table 5).

Simultaneous detection of changes in clonal abundance, gene expression, and tumor microenvironment in response to BET bromodomain inhibition with WILD-seq.

Having established that we can repeatedly observe the same clonal lineages and their gene expression programs across animals and experiments, we next sought to perturb the system. We chose the BET bromodomain inhibitor JQ1 for our proof-of-principle experiments to assess the ability of the WILD-seq system to simultaneously measure changes in clonal abundance, gene expression and the tumor microenvironment that occur following therapeutic intervention. JQ1 competitively binds to acetylated lysines, displacing BRD4 and thereby repressing transcription at specific loci. A large number of studies have indicated that BET inhibitors may be beneficial in the treatment of hematological malignancies and solid tumors including breast cancer, possibly by inhibiting certain key proto-oncogenes such as MYC (G. Jiang et al., 2020).

Treatment of our 4T1 WILD-seq tumor-bearing mice with JQ1 caused an initial suppression of tumor growth but with only a small overall effect on time to humane endpoint (Fig. 2a). Tumors treated with JQ1 or vehicle alone were harvested at endpoint, dissociated and subjected to single cell sequencing (Fig. 2b). Two independent experiments were performed, each with 3 mice per condition.

We first explored whether JQ1 had any effect on the tumor microenvironment. The most striking difference we observed was a change in abundance among the cells belonging to the T-cell compartment. To analyze this further, we computationally extracted these cells from the single cell data, reclustered them and performed differential abundance testing using Milo (Fig. 2c). Milo detects sets of cells that are differentially abundant between conditions by modeling counts of cells in neighborhoods of a KNN graph (Dann et al., 2021). When applied to our reclustered T-cells, Milo identified a significant decrease in abundance in cytotoxic T-cells, as identified by their expression of Cd8a and Cd8b1,

following JQ1 treatment. A significant change was observed in both of our experiments although the magnitude of the effect was greater in experiment A.

We next examined the effect of JQ1 treatment on the transcriptome of the tumor cells. Differential expression analysis was performed for each clonal lineage and experiment independently. As expected, given its mode of action, we identified significant down-regulation of a wide range of genes with consistent changes across clonal lineages (Fig 2d, supplementary table 6). Among the repressed genes, were a number of genes related to interferon (IFN) signaling and antigen processing and presentation (Fig. 2d, 2e), including GBP2 which is strongly induced by IFN gamma, the MHC class II protein, Cd74, and B2m, a component of the MHC class I complex. JQ1 has previously been reported to directly inhibit transcription of IFN-response genes (Gibbons et al., 2019; Gusyatiner et al., 2021) suggesting this may be due to the direct action of JQ1 within our tumor cells, however JQ1-dependent changes to the tumor microenvironment may also influence these expression pathways.

Our barcoded 4T1 clones showed varied sensitivity to JQ1, with treatment causing reproducible changes to clonal proportions within the tumor (Fig. 2f, 2g, Supplementary Table 2). In particular, one of the most abundant clones, clone 473, is highly sensitive to JQ1 treatment. In contrast, 3 clones were identified as being the most resistant to JQ1 treatment, clones 93, 439 and 264. These clones which together make up less than 5% of the tumor in vehicle treated mice constitute on average 12.8% of the JQ1-treated tumors. To examine baseline transcriptomic signatures of JQ1-sensitivity and resistance, we identified gene sets whose expression in vehicle-treated tumors was highly correlated with response (Figs. 2h, 2i, Supplementary Table 7). Interestingly, interferon signaling which is significantly attenuated in our JQ1-treated tumors is highly correlated with sensitivity to JQ1, suggesting a possible higher dependence of the sensitive clones on these pathways. Conversely resistance is associated with higher levels of unfolded protein response and mTOR signaling consistent with a known role of mTOR-mediated autophagy in resistance to JQ1 (Luan et al., 2019), and cytotoxic synergy between PI3K/mTOR inhibitors and BET inhibitors (Lee et al., 2015; Stratikopoulos et al., 2015).

Clonal transcriptomic correlates of response and resistance to taxane chemotherapy in the 4T1 mammary carcinoma model.

Our studies with JQ1 exemplify the ability of the WILD-seq system to simultaneously measure *in vivo* the effect of therapeutic intervention on clonal dynamics, gene expression and the tumor microenvironment. However, we were interested in using our system to investigate a chemotherapeutic agent currently in use in the clinic. We therefore treated our 4T1 WILD-seq tumor-bearing mice with docetaxel as a representative taxane, a class of drugs which are routinely used to treat triple negative breast cancer patients. As with JQ1, docetaxel treatment resulted in an initial, modest reduction in tumor growth followed by relapse (Fig. 3a). Comparison of vehicle and docetaxel (DTX) treated tumors revealed differential response of clonal lineages to treatment (Figs. 3b, 3c, 3d, Supplementary Table 2) with clone 679 being the most resistant and clone 238 the most sensitive to chemotherapy.

Correlating the clones' baseline transcriptomic profiles with response to docetaxel, revealed a major role for EMT in modulating sensitivity and resistance to taxane-based therapy. The 4T1 clones which are most sensitive to docetaxel are characterized by high expression of E-Cadherin regulated genes and low Zeb1 activity consistent with a more epithelial phenotype (Figs. 3e, 3f, Supplementary Table 8). These observations are in agreement with previous studies that have implicated EMT, and its associated endowment of cancer stem cell-like characteristics, as a mechanism of resistance to cytotoxic chemotherapies like docetaxel in cell culture and patients (Bhola et al., 2013; Creighton et al., 2009; Gupta et al., 2009). Resistance to docetaxel was correlated with up-regulation of multiple gene sets (Figs. 3e, 3f, Supplementary Table 8). This included genes whose expression is elevated in non-responders to docetaxel in human breast cancer patients (Honma et al., 2008) demonstrating the relevance of findings arising from this approach. Interestingly, we also identify metabolic reprogramming as a potential mechanism of docetaxel resistance with higher expression of genes related to glycogen and glutathione metabolism being correlated with resistance to docetaxel (Fig, 3e).

Clonal transcriptomic signatures of response and resistance to taxane chemotherapy in the D2A1 mammary carcinoma model.

To explore the general applicability of WILD-seq to other models, we utilized a second triple negative mammary carcinoma model, D2A1-m2 (hereafter referred to as D2A1). Similar to the 4T1 cell line, this line was derived from a mouse mammary tumor in a BALB/c mouse and can be orthotopically implanted into the mammary fat pad of immunocompetent, syngeneic hosts (Jungwirth et al., 2017).

We established a WILD-seq D2A1 clonal pool by transducing the D2A1 cell line with our WILD-seq barcode library. These barcoded cells were orthotopically implanted into a cohort of mice, half of which were treated with docetaxel, while the remaining animals received vehicle alone. Docetaxel treatment caused an initial reduction in tumor growth with subsequent relapse (Fig. 4a). We performed single cell RNA sequencing of three tumors per condition and assigned the tumor cells to a distinct clonal lineage based on the presence of the WILD-seq barcode (Fig. 4b). In total 103 different WILD-seq barcodes were observed *in vivo* with a dramatic shift in relative clonal abundance on docetaxel treatment (Fig. 4d, Supplementary Table 3). Unlike our 4T1 breast cancer model, variation between clonal lineages was no longer dominated by the EMT status of the clones and all clones exhibited a more mesenchymal-like phenotype consistent with the fact that this was a subline of D2A1 selected for its metastatic properties (Fig. 4c). This provides us with a distinct yet complementary system to investigate chemotherapy resistance with the potential to reveal alternative mechanisms than EMT status.

We identified 3 clones which were acutely sensitive to docetaxel, clones 118, 2874 and 1072. Together these constitute on average 37% of the vehicle-treated tumors but only 1.3% of the docetaxel-treated tumors (Fig. 4d). To understand the properties of these clones, we analyzed the baseline gene expression characteristics of clones in vehicle-treated tumors. The gene expression of cells from a clone of interest was compared to all tumor cells to which a WILD-seq barcode could be assigned from the same sample, and clonal signatures identified that were significantly enriched across animals. Specific gene expression signatures were identifiable for all clones analyzed, some of which were unique to a single clone while others overlapped across the sensitive clones (Fig. 4e, Supplementary Table 9, Supplementary Table 10). For example, clone 1072 shows elevated levels of expression of cell cycle related pathways, such as E2F-target genes (Fig 4f), indicating that aberrant cell cycle control in these cells that could increase their susceptibility to an antimitotic cancer drug (Fig. 4f), interestingly high levels of E2F-targets have recently been shown to be associated with response to chemotherapy in breast cancer patients (Sammut et al., 2021).

Three clones robustly increased their relative abundance within the tumor following docetaxel treatment, clones 1197, 751 and 1240, which despite making up less than 1% of the vehicle-treated tumors together constituted more than 20% of the docetaxel-treated tumors (Fig. 4d). Due to the low abundance of cells in vehicle-treated samples, cells belonging to all 3 clones were pooled to analyse their baseline gene expression profiles (Fig. 4g). Among the gene sets differentially expressed between resistant and sensitive clones, were a number of breast cancer amplicons indicating that there may be specific copy number variations associated with these clones (Figs. 4g, 4h). However single cell DNA sequencing data would be required to confirm the presence of specific genetic traits within our clones. Interestingly, gains in 8q24 (Han et al., 2010), 20q11 (Voutsadakis, 2021) and loss of 16q (Höglander et al., 2018) have previously been reported to be associated with resistance to taxane-based chemotherapy in agreement with our findings. Highly upregulated within all 3 of our resistant clones were genes related to the NRF2 pathway, even in the absence of docetaxel treatment (Figs. 4g, 4h). NRF2 activation has been linked to cancer progression and metastasis and has been suggested to confer resistance to chemotherapy (Homma et al., 2009; T. Jiang et al., 2010; Konstantinopoulos et al., 2011; Romero et al., 2017; Shibata et al., 2008; Singh et al., 2006).

Delineating the contribution of clonal abundance to gene expression changes upon drug treatment.

Prior to the advent of single cell sequencing, the majority of studies relied on bulk RNA-seq or microarray analysis of gene expression to interrogate the effect of chemotherapeutic interventions. While informative, these studies cannot differentiate between changes in bulk gene expression that arise due to clonal selection and changes that are induced within a clonal lineage as the result of drug exposure. Even with single cell sequencing, definitive identification of the same clonal population across treatment conditions is impractical. Our method alleviates these difficulties by enabling the direct comparison of clones of the same lineage under different conditions.

To examine the relative contribution of clonal selection and transcriptional reprogramming to changes in gene expression upon chemotherapy, we compared analysis of gene expression within each clone individually to a combined analysis of all pooled tumor cells (Fig. 5, Supplementary Table 11). Consistent with their mode of action, docetaxel had relatively little effect on the transcriptome of individual clones while JQ1 caused substantial changes to the transcriptome predominantly down-regulating gene expression. Genes were identified under all treatments that were altered within the tumor as a whole but as a result of clonal selection rather than intra-clonal changes in gene expression, with the biggest effects being observed with docetaxel treatment in D2A1 tumors, in agreement with this condition inducing the largest changes in relative clonal abundance. To confirm that changes in

281 gene expression detected in bulk tumor analysis but not the clonal analysis could be attributed to
282 differences in clonal sensitivity to chemotherapy, we analyzed baseline expression of these genes
283 across the major clonal populations (Fig. 5b). As expected, we found that genes up-regulated only in
284 bulk tumor analysis had significantly higher expression in clones resistant to chemotherapy (that
285 increase in abundance with treatment) and genes only down-regulated in bulk tumor analysis had
286 significantly lower expression in these resistant clonal lineages.

287 Among the genes that change in expression within the tumor as a whole as a result of clonal selection
288 upon docetaxel treatment, we identified a number of genes related to glutathione synthesis and
289 conjugation including Mgst2, Esd and Gclm, that may endow resistant clones with greater ability to
290 resolve reactive oxygen species (ROS) induced by docetaxel (Alexandre et al., 2007). Of note, we also
291 observed that in 4T1 tumors, Epcam was significantly reduced in expression in the bulk tumor but was
292 not changed within the individual clonal populations. This suggests that rather than inducing an EMT
293 within the tumor cells, docetaxel is selecting clones of a pre-existing more mesenchymal phenotype.

294 **Convergent WILD-seq analysis across models identifies redox defense as a mediator of taxane**
295 **resistance and amino asparagine deprivation as a means to target resistant clones.**

296 To examine if there were any shared mechanisms of taxane resistance across our 4T1 and D2A1 WILD-
297 seq clones, we looked for genes that were enriched in resistant clonal lineages in both models. We
298 identified 47 overlapping resistance genes (Fig. 6a, Supplementary Table 12). These genes were
299 significantly enriched in pathways related to resolution of oxidative stress including the NRF2 pathway
300 and glutathione-mediated detoxification (Fig. 6b).

301 Importantly, these genes were also enriched in human patients following combined anthracycline and
302 taxane-based therapy, highlighting the potential clinical significance of our findings (Fig. 6c). Gene
303 expression data from a previously published study with paired pre-neo adjuvant chemotherapy (NAC)
304 core needle biopsies and post-chemotherapy surgical samples (Vera-Ramirez et al., 2013) was re-
305 analyzed using GSVA (Hänzelmann et al., 2013) to determine the effect of chemotherapy on a gene
306 set composed of our 47 shared resistance genes (Fig. 6c) as well as NRF2-targets as determined by
307 ChIP enrichment analysis (CHEA) (Lachmann et al., 2010) (Fig. 6d). Expression of both these gene
308 sets was significantly increased after chemotherapy, which our data would suggest is the result of
309 outgrowth of resistant clonal lineages with increased propensity to withstand taxane-induced oxidative
310 stress.

311 Given these findings, we hypothesized that combining taxane-based chemotherapy with a drug
312 specifically targeting resistant clones with high Nrf2 signaling would provide a highly effective treatment
313 regime. To test this hypothesis, we leveraged the finding that tumors with constitutively active Nrf2, due
314 to mutation in the negative regulator Keap1, have metabolic vulnerabilities that arise from their high
315 antioxidant production (Romero et al., 2017), including dependency of glutamine (Romero et al., 2017)
316 and a general dependency on exogenous non-essential amino acids (NEAA) including asparagine
317 (LeBoeuf et al., 2020). This metabolic dependency can be targeted therapeutically by L-asparaginase
318 (ASNase from *E.coli*), which is used in the clinical management of acute lymphoblastic leukemia (*ALL*)
319 (Batool et al., 2016), and catalyzes the conversion asparagine to aspartic acid and ammonia (Chan et
320 al., 2019).

321 To ascertain whether docetaxel resistant clones were collaterally sensitive to ASNase, we treated D2A1
322 WILD-seq tumors initially with docetaxel to select for resistant clones and then began daily treatment
323 with L-asparaginase one week later. This dosing regime was chosen as we found that with the dose of
324 docetaxel used in this study co-administration of the 2 drugs or treatment with ASNase immediately
325 following docetaxel was poorly tolerated. As shown in Figure 6e, treatment with ASNase arrested tumor
326 growth and led to a ~40% increase in time to endpoint (relative to vehicle) in this highly aggressive
327 model, although the tumors did acquire resistance and regrew after approximately one week of
328 treatment. Importantly, ASNase alone had no significant effect on tumor growth, indicative of a
329 docetaxel-induced effect (Fig. 6f). To determine the response of individual clonal lineages to ASNase
330 treatment, we performed single cell sequencing on vehicle treated tumors (day 21), as well as docetaxel
331 treated tumors before the start of ASNase treatment (day 21) and after 4 doses of ASNase (day 25).
332 As before, our docetaxel-resistant clones, 751, 1197 and 1240, which have high levels of Nrf2 signaling
333 all exhibited a dramatic increase in their abundance with docetaxel treatment (Fig. 6g). Excitingly,
334 clones 751 and 1197 were sensitive to ASNase returning to baseline levels. Clone 1240 decreased in
335 abundance in 2 of the 3 mice analyzed so is likely to also be sensitive to ASNase but more data is
336 required to confirm its response. As predicted, our Nrf2-high resistant clones were selectively targeted

by amino acid deprivation as other clones such as 2323 were unchanged in their relative abundance (Fig. 6g).

To confirm the mechanism of action of L-asparaginase and identify potential mechanisms of resistance to this drug that might cause the relapse observed, we analyzed the transcriptomic effects of ASNase administration. Genes which consistently changed in expression after ASNase treatment across clonal lineages are shown in Figure 6h. Many of the genes found to be differentially expressed in our tumor cells following L-asparaginase treatment are either directly related to protein synthesis (Eif3c, Gars, Eif3g, Eif5a) or are consistent with changes in gene expression reported in cell lines following amino acid deprivation including Atf5, Atf3, Jun, Fos, Egr1 and Asns (Fu et al., 2011; Pan et al., 2003; Pohjanpelto & Hölttä, 1990; Shan et al., 2010). Of specific interest is the up-regulation of asparagine synthetase (Asns) which catalyzes the *de novo* biosynthesis of L-asparagine from L-aspartate. In acute lymphoblastic leukemia (*ALL*), low levels of ASNS resulting in a dependence on extra-cellular asparagine are considered an important biomarker for L-asparaginase treatment. Moreover, the importance of ASNS overexpression in conferring asparaginase resistance has been well documented and is frequently seen in *ALL* patients that develop drug-resistant forms of the disease following treatment with ASNase (reviewed in (Richards & Kilberg, 2006)). In our experiments, this adaptation to asparaginase is observed across all clones analyzed suggesting a general resistance mechanism and supporting the clinical utility of an Asns inhibitor, if one were to be developed, as third line treatment in this context.

In summary, these data support the notion that WILD-seq can identify causal mechanisms of drug resistance *in vivo*, that can be leveraged to inform new combination therapies. Since the redox defense signatures we identified are detectable in patients after neo-adjuvant chemotherapy (NAC), one can envisage an approach whereby patients receiving NAC have the surgical tumor specimen profiled for NRF2 gene signatures and those with high levels receive a post-operative course of L-asparaginase.

Discussion

Tumor heterogeneity is thought to underlie drug resistance through the selection of clonal lineages that can preferentially survive therapy. However, identifying the features of such lineages, so that they can be targeted therapeutically, has been challenging due the lack of understanding of their molecular characteristics and the lack of animal models to prospectively test therapeutic interventions and combinations thereof. To overcome these challenges, we utilized WILD-seq, a system that leverages expressed barcodes, population bottle necking, syngeneic mouse models and single cell RNA-seq to link clonal lineage to the transcriptome. Among the existing methods for coupling lineage tracing with single cell transcriptomic profiling, the majority use either lentiviral delivery of a genetic barcode similar to that used here or CRISPR/Cas9-mediate mutations for clonal lineage identification (Biddy et al., 2018; Gutierrez et al., 2021; Quinn et al., 2021; Simeonov et al., 2021; Weinreb et al., 2020). We chose to avoid CRISPR/Cas9-based lineage labeling as induction of DNA damage could have an impact on the transcriptome and the sensitivity of the cells to therapeutic agents (Haapaniemi et al., 2018; L. Jiang et al., 2021). Our approach is unique in that we purposefully bottleneck our clonal population to achieve a balance between maximizing clonal diversity and minimizing variation in clonal representation across replicate animals and experiments. It is this feature that allows us to robustly call clonal gene expression signatures and differential clonal abundance before and after therapeutic intervention and it is this in turn that allows us to identify relevant drug resistance mechanisms *in vivo*.

We find that the abundance of clones in cell culture and *in vivo* differ greatly, with the most abundant clones *in vitro* being lowly represented *in vivo* and vice versa thus providing a cautionary note when analyzing drug response *in vitro*. Moreover, WILD-seq of 4T1 tumors revealed that the relative immune competence of the host profoundly sculpts the transcriptome of clonal lineages and, as exemplified by JQ1, therapeutic interventions can impact the tumor microenvironment and its interaction with tumor cells, effects that would be missed *in vitro* and in immunocompromised hosts. We utilized WILD-seq to analyze sensitivity and resistance to taxane chemotherapy in two syngeneic, triple negative, mammary carcinoma models highlighting both known and new pathways of resistance (Marine et al., 2020). Resistance to cancer therapies can arise due to clonal selection or through adaptive reprograming of the epigenome and transcriptome of individual clones. Our data with docetaxel treatment in 4T1 and D2A1 indicate that over the time frames we have examined clonal selection is the dominant force driving resistance to chemotherapy with gene expression signatures, such as EMT and Nrf2 signaling, being present in clones at baseline that are then selected for during therapy. However, depending on the

392 mode of action of specific drugs, transcriptional reprogramming may also induce therapeutic resistance
393 and such mechanisms can also be effectively identified with the WILD-seq platform. Indeed, up-
394 regulation of Asns, detected across clonal lineages after L-asparaginase provides an example of *de*
395 *novo* acquisition of a resistance phenotype.

396 Applying WILD-seq to examine docetaxel response across two TNBC models afforded the opportunity
397 to overlap resistance genes for the same drug across models and remove model-specific effects. These
398 analyses uncovered a critical role for redox defense in docetaxel resistance that also appears to be
399 operative in human breast cancer patients after chemotherapy. Having identified a primary cause of
400 resistance, we next sought to explore the possibility of collateral sensitivity. Collateral sensitivity, first
401 described for antibiotics (Imamovic & Sommer, 2013; Pluchino et al., 2012; Roemhild & Andersson,
402 2021) is the phenomenon by which resistance to one drug comes at the cost of sensitivity to a second
403 drug. In the context of cancer and taxanes, collateral sensitivity has the distinct advantage over other
404 therapeutic strategies of maintaining the initial first line therapy and only modifying subsequent
405 therapies. We took advantage of previous findings linking constitutive Nrf2 signaling, via Keap1 loss, to
406 a dependency on exogenous non-essential amino acids (LeBoeuf et al., 2020) and thereby sensitivity
407 to L-asparaginase. Application of L-asparaginase led to an initial cessation of tumor growth followed by
408 regrowth 6 days later. WILD-seq of docetaxel treated tumors before and after L-asparaginase treatment
409 confirmed the specific suppression of Nrf2 high clones and also revealed a compensatory, clone
410 agnostic, up-regulation of asparagine synthetase (Asns), which likely drives relapse in these tumors
411 given the importance of ASNS to L-asparaginase resistance in *ALL* (Richards & Kilberg, 2006).
412 Interestingly, we have previously shown that asparagine bioavailability regulates EMT and metastatic
413 progression in breast cancer models (Knott et al., 2018). Thus, asparagine deprivation, which has not
414 been extensively explored in breast cancer, may present multiple benefits to patients and the utility of
415 L-asparaginase, a clinical stage drug, in this setting warrants further investigation.

416 This study highlights the challenges of tackling tumor heterogeneity therapeutically. Even though we
417 can effectively suppress the induction of docetaxel resistant clones by administration of L-asparaginase
418 the tumors still adapt to this intervention and regrow, most likely due to transcriptionally shifting their
419 metabolism towards *de novo* asparagine synthesis. Nevertheless, hope still remains since there are
420 only three avenues by which cells can supply themselves with asparagine (1) uptake of extra-cellular
421 asparagine which is effectively shut-off by ASNase (2) *de novo* synthesis through Asns or (3) catabolism
422 of existing proteins. If we could effectively force tumors to depend on synthesis through Asns, we could
423 then deprive them of that additional dependency if Asns-directed therapeutics were to be developed.
424 This concept of steering clonal evolution with drugs towards a predictable and irreconcilable,
425 therapeutically targetable, dependency may provide a general approach to achieving durable
426 therapeutic responses for which tractable models of tumor evolution, such as those described here, are
427 essential predictive components.
428
429 **References**

430 Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G.,
431 Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., Oord, J. van den, Atak, Z. K., Wouters, J., &
432 Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature*
433 *Methods*, *14*(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

434 Alexandre, J., Hu, Y., Lu, W., Pelicano, H., & Huang, P. (2007). Novel Action of Paclitaxel against
435 Cancer Cells: Bystander Effect Mediated by Reactive Oxygen Species. *Cancer Research*, *67*(8),
436 3512–3517. https://doi.org/10.1158/0008-5472.can-06-3914

437 Al-Lazikani, B., Banerji, U., & Workman, P. (2012). Combinatorial drug therapy for cancer in the post-
438 genomic era. *Nature Biotechnology*, *30*(7), 679–692. https://doi.org/10.1038/nbt.2284

439 Batool, T., Makky, E. A., Jalal, M., & Yusoff, M. M. (2016). A Comprehensive Review on l-
440 Asparaginase and Its Applications. *Applied Biochemistry and Biotechnology*, *178*(5), 900–923.
441 https://doi.org/10.1007/s12010-015-1917-3

Bhang, H. C., Ruddy, D. A., Radhakrishna, V. K., Caushi, J. X., Zhao, R., Hims, M. M., Singh, A. P., Kao, I., Rakiec, D., Shaw, P., Balak, M., Raza, A., Ackley, E., Keen, N., Schlabach, M. R., Palmer, M., Leary, R. J., Chiang, D. Y., Sellers, W. R., … Stegmeier, F. (2015). Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nature Medicine*, *21*(5), 440–448. https://doi.org/10.1038/nm.3841

Bhola, N. E., Balko, J. M., Dugger, T. C., Kuba, M. G., Sánchez, V., Sanders, M., Stanford, J., Cook, R. S., & Arteaga, C. L. (2013). TGF-β inhibition enhances chemotherapy action against triple-negative breast cancer. *Journal of Clinical Investigation*, *123*(3), 1348–1358. https://doi.org/10.1172/jci65416

Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T., & Morris, S. A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. *Nature*, *564*(7735), 219–224. https://doi.org/10.1038/s41586-018-0744-4

Buschmann, T. (2017). DNABarcodes: an R package for the systematic construction of DNA sample tags. *Bioinformatics*, btw759. https://doi.org/10.1093/bioinformatics/btw759

Chan, W.-K., Horvath, T. D., Tan, L., Link, T., Harutyunyan, K. G., Pontikos, M. A., Anishkin, A., Du, D., Martin, L. A., Yin, E., Rempe, S. B., Sukharev, S., Konopleva, M., Weinstein, J. N., & Lorenzi, P. L. (2019). Glutaminase Activity of L-Asparaginase Contributes to Durable Preclinical Activity against Acute Lymphoblastic Leukemia. *Molecular Cancer Therapeutics*, *18*(9), 1587–1592. https://doi.org/10.1158/1535-7163.mct-18-1329

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128–128. https://doi.org/10.1186/1471-2105-14-128

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., Rimm, D. L., Wong, H., Rodriguez, A., Herschkowitz, J. I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M. C., Renshaw, L., Larionov, A. A., Faratian, D., Hilsenbeck, S. G., … Chang, J. C. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proceedings of the National Academy of Sciences*, *106*(33), 13820–13825. https://doi.org/10.1073/pnas.0905718106

Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D., & Marioni, J. C. (2021). Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nature Biotechnology*, 1–9. https://doi.org/10.1038/s41587-021-01033-z

Dentro, S. C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A. G., Yu, K., Rubanova, Y., Macintyre, G., Demeulemeester, J., Vázquez-García, I., Kleinheinz, K., Livitz, D. G., Malikic, S., Donmez, N., Sengupta, S., Anur, P., Jolly, C., Cmero, M., … Loo, P. V. (2021). Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell*, *184*(8), 2239-2254.e39. https://doi.org/10.1016/j.cell.2021.03.009

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., … DiPersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, *481*(7382), 506–510. https://doi.org/10.1038/nature10738

Fu, L., Balasubramanian, M., Shan, J., Dudenhausen, E. E., & Kilberg, M. S. (2011). Auto-activation of c-JUN Gene by Amino Acid Deprivation of Hepatocellular Carcinoma Cells Reveals a Novel c-JUN-mediated Signaling Pathway*. *Journal of Biological Chemistry*, *286*(42), 36724–36738. https://doi.org/10.1074/jbc.m111.277673

Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S.,

McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., … Swanton, C. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *The New England Journal of Medicine*, *366*(10), 883–892. https://doi.org/10.1056/nejmoa1113205

Gibbons, H. R., Mi, D. J., Farley, V. M., Esmond, T., Kaood, M. B., & Aune, T. M. (2019). Bromodomain inhibitor JQ1 reversibly blocks IFN-γ production. *Scientific Reports*, *9*(1), 10280. https://doi.org/10.1038/s41598-019-46516-x

Gupta, P. B., Onder, T. T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R. A., & Lander, E. S. (2009). Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. *Cell*, *138*(4), 645–659. https://doi.org/10.1016/j.cell.2009.06.034

Gusyatiner, O., Bady, P., Pham, M. D. T., Lei, Y., Park, J., Daniel, R. T., Delorenzi, M., & Hegi, M. E. (2021). BET inhibitors repress expression of Interferon-stimulated genes and synergize with HDAC inhibitors in glioblastoma. *Neuro-Oncology*, *23*(10), noab115-. https://doi.org/10.1093/neuonc/noab115

Gutierrez, C., Al'Khafaji, A. M., Brenner, E., Johnson, K. E., Gohil, S. H., Lin, Z., Knisbacher, B. A., Durrett, R. E., Li, S., Parvin, S., Biran, A., Zhang, W., Rassenti, L., Kipps, T. J., Livak, K. J., Neuberg, D., Letai, A., Getz, G., Wu, C. J., & Brock, A. (2021). Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. *Nature Cancer*, *2*(7), 758–772. https://doi.org/10.1038/s43018-021-00222-8

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z

Han, S., Park, K., Shin, E., Kim, H.-J., Kim, J. Y., Kim, J. Y., & Gwak, G. (2010). Genomic change of chromosome 8 predicts the response to taxane-based neoadjuvant chemotherapy in node-positive breast cancer. *Oncology Reports*, *24*(1), 121–128.

Hänzelmann, S., Castelo, R., & Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, *14*(1), 7–7. https://doi.org/10.1186/1471-2105-14-7

Höglander, E. K., Nord, S., Wedge, D. C., Lingjærde, O. C., Silwal-Pandit, L., Gythfeldt, H. vdL, Vollan, H. K. M., Fleischer, T., Krohn, M., Schlitchting, E., Borgen, E., Garred, Ø., Holmen, M. M., Wist, E., Naume, B., Loo, P. V., Børresen-Dale, A.-L., Engebraaten, O., & Kristensen, V. (2018). Time series analysis of neoadjuvant chemotherapy and bevacizumab-treated breast carcinomas reveals a systemic shift in genomic aberrations. *Genome Medicine*, *10*(1), 92. https://doi.org/10.1186/s13073-018-0601-y

Hollern, D. P., Swiatnicki, M. R., & Andrechek, E. R. (2018). Histological subtypes of mouse mammary tumors reveal conserved relationships to human cancers. *PLOS Genetics*, *14*(1), e1007135. https://doi.org/10.1371/journal.pgen.1007135

Homma, S., Ishii, Y., Morishima, Y., Yamadori, T., Matsuno, Y., Haraguchi, N., Kikuchi, N., Satoh, H., Sakamoto, T., Hizawa, N., Itoh, K., & Yamamoto, M. (2009). Nrf2 Enhances Cell Proliferation and Resistance to Anticancer Drugs in Human Lung Cancer. *Clinical Cancer Research*, *15*(10), 3423–3432. https://doi.org/10.1158/1078-0432.ccr-08-2822

Honma, K., Iwao-Koizumi, K., Takeshita, F., Yamamoto, Y., Yoshida, T., Nishio, K., Nagahara, S., Kato, K., & Ochiya, T. (2008). RPN2 gene confers docetaxel resistance in breast cancer. *Nature Medicine*, *14*(9), 939–948. https://doi.org/10.1038/nm.1858

Imamovic, L., & Sommer, M. O. A. (2013). Use of Collateral Sensitivity Networks to Design Drug Cycling Protocols That Avoid Resistance Development. *Science Translational Medicine*, *5*(204), 204ra132-204ra132. https://doi.org/10.1126/scitranslmed.3006609

534 Jamal-Hanjani, M., Hackshaw, A., Ngai, Y., Shaw, J., Dive, C., Quezada, S., Middleton, G., Bruin, E.
535 de, Quesne, J. L., Shafi, S., Falzon, M., Horswell, S., Blackhall, F., Khan, I., Janes, S., Nicolson,
536 M., Lawrence, D., Forster, M., Fennell, D., … Swanton, C. (2014). Tracking Genomic Cancer
537 Evolution for Precision Medicine: The Lung TRACERx Study. *PLoS Biology*, *12*(7), e1001906.
538 https://doi.org/10.1371/journal.pbio.1001906

539 Jiang, G., Deng, W., Liu, Y., & Wang, C. (2020). General mechanism of JQ1 in inhibiting various
540 types of cancer. *Molecular Medicine Reports*, *21*(3), 1021–1034.
541 https://doi.org/10.3892/mmr.2020.10927

542 Jiang, L., Ingelshed, K., Shen, Y., Boddul, S. V., Iyer, V. S., Kasza, Z., Sedimbi, S., Lane, D. P., &
543 Wermeling, F. (2021). CRISPR/Cas9-induced DNA damage enriches for mutations in a p53-linked
544 interactome: implications for CRISPR-based therapies. *Cancer Research*, *82*(1),
545 canres.1692.2021. https://doi.org/10.1158/0008-5472.can-21-1692

546 Jiang, T., Chen, N., Zhao, F., Wang, X.-J., Kong, B., Zheng, W., & Zhang, D. D. (2010). High Levels
547 of Nrf2 Determine Chemoresistance in Type II Endometrial Cancer. *Cancer Research*, *70*(13),
548 5486–5496. https://doi.org/10.1158/0008-5472.can-10-0713

549 Jungwirth, U., Weverwijk, A. van, Melake, M. J., Chambers, A. F., Gao, Q., Fivaz, M., & Isacke, C. M.
550 (2017). Generation and characterisation of two D2A1 mammary cancer sublines to model
551 spontaneous and experimental metastasis in a syngeneic BALB/c host. *Disease Models &*
552 *Mechanisms*, *11*(1), dmm031740. https://doi.org/10.1242/dmm.031740

553 Knott, S. R. V., Wagenblast, E., Khan, S., Kim, S. Y., Soto, M., Wagner, M., Turgeon, M.-O., Fish, L.,
554 Erard, N., Gable, A. L., Maceli, A. R., Dickopf, S., Papachristou, E. K., D'Santos, C. S., Carey, L.
555 A., Wilkinson, J. E., Harrell, J. C., Perou, C. M., Goodarzi, H., … Hannon, G. J. (2018). Asparagine
556 bioavailability governs metastasis in a model of breast cancer. *Nature*, *554*(7692), 378–381.
557 https://doi.org/10.1038/nature25465

558 Konstantinopoulos, P. A., Spentzos, D., Fountzilas, E., Francoeur, N., Sanisetty, S., Grammatikos, A.
559 P., Hecht, J. L., & Cannistra, S. A. (2011). Keap1 Mutations and Nrf2 Pathway Activation in
560 Epithelial Ovarian Cancer. *Cancer Research*, *71*(15), 5081–5089. https://doi.org/10.1158/0008-
561 5472.can-10-4668

562 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S.,
563 Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G.
564 W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server
565 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

566 Lachmann, A., Xu, H., Krishnan, J., Berger, S. I., Mazloom, A. R., & Ma'ayan, A. (2010). ChEA:
567 transcription factor regulation inferred from integrating genome-wide ChIP-X experiments.
568 *Bioinformatics*, *26*(19), 2438–2444. https://doi.org/10.1093/bioinformatics/btq466

569 Landau, D. A., Carter, S. L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M. S., Sougnez,
570 C., Stewart, C., Sivachenko, A., Wang, L., Wan, Y., Zhang, W., Shukla, S. A., Vartanov, A.,
571 Fernandes, S. M., Saksena, G., Cibulskis, K., Tesar, B., Gabriel, S., … Wu, C. J. (2013). Evolution
572 and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell*, *152*(4), 714–726.
573 https://doi.org/10.1016/j.cell.2013.01.019

574 LeBoeuf, S. E., Wu, W. L., Karakousi, T. R., Karadal, B., Jackson, S. R., Davidson, S. M., Wong, K.-
575 K., Koralov, S. B., Sayin, V. I., & Papagiannakopoulos, T. (2020). Activation of Oxidative Stress
576 Response in Cancer Generates a Druggable Dependency on Exogenous Non-essential Amino
577 Acids. *Cell Metabolism*, *31*(2), 339-350.e4. https://doi.org/10.1016/j.cmet.2019.11.012

578 Lee, D. H., Qi, J., Bradner, J. E., Said, J. W., Doan, N. B., Forscher, C., Yang, H., & Koeffler, H. P.
579 (2015). Synergistic effect of JQ1 and rapamycin for treatment of human osteosarcoma.
580 *International Journal of Cancer*, *136*(9), 2055–2064. https://doi.org/10.1002/ijc.29269

581 Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P.
582 (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*(12), 1739–1740.
583 https://doi.org/10.1093/bioinformatics/btr260

584 Luan, W., Pang, Y., Li, R., Wei, X., Jiao, X., Shi, J., Yu, J., Mao, H., & Liu, P. (2019). Akt/mTOR-
585 Mediated Autophagy Confers Resistance To BET Inhibitor JQ1 In Ovarian Cancer. *OncoTargets*
586 *and Therapy*, *12*, 8063–8074. https://doi.org/10.2147/ott.s220267

587 Marine, J.-C., Dawson, S.-J., & Dawson, M. A. (2020). Non-genetic mechanisms of therapeutic
588 resistance in cancer. *Nature Reviews Cancer*, *20*(12), 743–756. https://doi.org/10.1038/s41568-
589 020-00302-4

590 Mueller, H. S., Fowler, C. E., Dalin, S., Moiso, E., Udomlumleart, T., Garg, S., Hemann, M. T., & Lees,
591 J. A. (2021). Acquired resistance to PRMT5 inhibition induces concomitant collateral sensitivity to
592 paclitaxel. *Proceedings of the National Academy of Sciences*, *118*(34), e2024055118.
593 https://doi.org/10.1073/pnas.2024055118

594 O'Neil, N. J., Bailey, M. L., & Hieter, P. (2017). Synthetic lethality and cancer. *Nature Reviews*
595 *Genetics*, *18*(10), 613–623. https://doi.org/10.1038/nrg.2017.47

596 Oren, Y., Tsabar, M., Cuoco, M. S., Amir-Zilberstein, L., Cabanos, H. F., Hütter, J.-C., Hu, B.,
597 Thakore, P. I., Tabaka, M., Fulco, C. P., Colgan, W., Cuevas, B. M., Hurvitz, S. A., Slamon, D. J.,
598 Deik, A., Pierce, K. A., Clish, C., Hata, A. N., Zaganjor, E., … Regev, A. (2021). Cycling cancer
599 persister cells arise from lineages with distinct programs. *Nature*, *596*(7873), 576–582.
600 https://doi.org/10.1038/s41586-021-03796-6

601 Pan, Y., Chen, H., Siu, F., & Kilberg, M. S. (2003). Amino Acid Deprivation and Endoplasmic
602 Reticulum Stress Induce Expression of Multiple Activating Transcription Factor-3 mRNA Species
603 That, When Overexpressed in HepG2 Cells, Modulate Transcription by the Human Asparagine
604 Synthetase Promoter*. *Journal of Biological Chemistry*, *278*(40), 38402–38412.
605 https://doi.org/10.1074/jbc.m304574200

606 Pluchino, K. M., Hall, M. D., Goldsborough, A. S., Callaghan, R., & Gottesman, M. M. (2012).
607 Collateral sensitivity as a strategy against cancer multidrug resistance. *Drug Resistance Updates*,
608 *15*(1–2), 98–105. https://doi.org/10.1016/j.drup.2012.03.002

609 Pohjanpelto, P., & Hölttä, E. (1990). Deprivation of a single amino acid induces protein synthesis-
610 dependent increases in c-jun, c-myc, and ornithine decarboxylase mRNAs in Chinese hamster
611 ovary cells. *Molecular and Cellular Biology*, *10*(11), 5814–5821.
612 https://doi.org/10.1128/mcb.10.11.5814-5821.1990

613 Quinn, J. J., Jones, M. G., Okimoto, R. A., Nanjo, S., Chan, M. M., Yosef, N., Bivona, T. G., &
614 Weissman, J. S. (2021). Single-cell lineages reveal the rates, routes, and drivers of metastasis in
615 cancer xenografts. *Science*, *371*(6532), eabc1944. https://doi.org/10.1126/science.abc1944

616 Richards, N. G. J., & Kilberg, M. S. (2006). Asparagine Synthetase Chemotherapy. *Annual Review of*
617 *Biochemistry*, *75*(1), 629–654. https://doi.org/10.1146/annurev.biochem.75.103004.142520

618 Roemhild, R., & Andersson, D. I. (2021). Mechanisms and therapeutic potential of collateral sensitivity
619 to antibiotics. *PLoS Pathogens*, *17*(1), e1009172. https://doi.org/10.1371/journal.ppat.1009172

620 Romero, R., Sayin, V. I., Davidson, S. M., Bauer, M. R., Singh, S. X., LeBoeuf, S. E., Karakousi, T.
621 R., Ellis, D. C., Bhutkar, A., Sánchez-Rivera, F. J., Subbaraj, L., Martinez, B., Bronson, R. T.,
622 Prigge, J. R., Schmidt, E. E., Thomas, C. J., Goparaju, C., Davies, A., Dolgalev, I., …
623 Papagiannakopoulos, T. (2017). Keap1 loss promotes Kras-driven lung cancer and results in
624 dependence on glutaminolysis. *Nature Medicine*, *23*(11), 1362–1368.
625 https://doi.org/10.1038/nm.4407

626 Sammut, S.-J., Crispin-Ortuzar, M., Chin, S.-F., Provenzano, E., Bardwell, H. A., Ma, W., Cope, W.,
627 Dariush, A., Dawson, S.-J., Abraham, J. E., Dunn, J., Hiller, L., Thomas, J., Cameron, D. A.,
628 Bartlett, J. M. S., Hayward, L., Pharoah, P. D., Markowetz, F., Rueda, O. M., … Caldas, C. (2021).
629 Multi-omic machine learning predictor of breast cancer therapy response. *Nature*, 1–10.
630 https://doi.org/10.1038/s41586-021-04278-5

631 Shan, J., Lopez, M.-C., Baker, H. V., & Kilberg, M. S. (2010). Expression profiling after activation of
632 amino acid deprivation response in HepG2 human hepatoma cells. *Physiological Genomics*, *41*(3),
633 315–327. https://doi.org/10.1152/physiolgenomics.00217.2009

634 Shibata, T., Kokubu, A., Gotoh, M., Ojima, H., Ohta, T., Yamamoto, M., & Hirohashi, S. (2008).
635 Genetic Alteration of Keap1 Confers Constitutive Nrf2 Activation and Resistance to Chemotherapy
636 in Gallbladder Cancer. *Gastroenterology*, *135*(4), 1358-1368.e4.
637 https://doi.org/10.1053/j.gastro.2008.06.082

638 Simeonov, K. P., Byrns, C. N., Clark, M. L., Norgard, R. J., Martin, B., Stanger, B. Z., Shendure, J.,
639 McKenna, A., & Lengner, C. J. (2021). Single-cell lineage tracing of metastatic cancer reveals
640 selection of hybrid EMT states. *Cancer Cell*, *39*(8), 1150-1162.e9.
641 https://doi.org/10.1016/j.ccell.2021.05.005

642 Singh, A., Misra, V., Thimmulappa, R. K., Lee, H., Ames, S., Hoque, M. O., Herman, J. G., Baylin, S.
643 B., Sidransky, D., Gabrielson, E., Brock, M. V., & Biswal, S. (2006). Dysfunctional KEAP1–NRF2
644 Interaction in Non-Small-Cell Lung Cancer. *PLoS Medicine*, *3*(10), e420.
645 https://doi.org/10.1371/journal.pmed.0030420

646 Stratikopoulos, E. E., Dendy, M., Szabolcs, M., Khaykin, A. J., Lefebvre, C., Zhou, M.-M., & Parsons,
647 R. (2015). Kinase and BET Inhibitors Together Clamp Inhibition of PI3K Signaling and Overcome
648 Resistance to Therapy. *Cancer Cell*, *27*(6), 837–851. https://doi.org/10.1016/j.ccell.2015.05.006

649 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius,
650 M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7),
651 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

652 Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich,
653 A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment
654 analysis: A knowledge-based approach for interpreting genome-wide expression profiles.
655 *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550.
656 https://doi.org/10.1073/pnas.0506580102

657 Turajlic, S., & Swanton, C. (2016). Metastasis as an evolutionary process. *Science*, *352*(6282), 169–
658 175. https://doi.org/10.1126/science.aaf2784

659 Vera-Ramirez, L., Sanchez-Rovira, P., Ramirez-Tortosa, C. L., Quiles, J. L., Ramirez-Tortosa, Mc., &
660 Lorente, J. A. (2013). Transcriptional Shift Identifies a Set of Genes Driving Breast Cancer
661 Chemoresistance. *PLoS ONE*, *8*(1), e53983. https://doi.org/10.1371/journal.pone.0053983

662 Voutsadakis, I. A. (2021). Chromosome 20q11.21 Amplifications in Colorectal Cancer. *Cancer*
663 *Genomics & Proteomics*, *18*(3 Suppl), 487–496. https://doi.org/10.21873/cgp.20274

664 Wagenblast, E., Soto, M., Gutiérrez-Ángel, S., Hartl, C. A., Gable, A. L., Maceli, A. R., Erard, N.,
665 Williams, A. M., Kim, S. Y., Dickopf, S., Harrell, J. C., Smith, A. D., Perou, C. M., Wilkinson, J. E.,
666 Hannon, G. J., & Knott, S. R. V. (2015). A model of breast cancer heterogeneity reveals vascular
667 mimicry as a driver of metastasis. *Nature*, *520*(7547), 358–362.
668 https://doi.org/10.1038/nature14403

669 Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., & Klein, A. M. (2020). Lineage tracing on
670 transcriptional landscapes links state to fate during differentiation. *Science*, *367*(6479), eaaw3381.
671 https://doi.org/10.1126/science.aaw3381

672 Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A.,
673 Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set
674 Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3), e90. https://doi.org/10.1002/cpz1.90

675 Yang, D., Jones, M. G., Naranjo, S., Rideout, W. M., Min, K. H. (Joseph), Ho, R., Wu, W., Replogle, J.
676 M., Page, J. L., Quinn, J. J., Horns, F., Qiu, X., Chen, M. Z., Freed-Pastor, W. A., McGinnis, C. S.,
677 Patterson, D. M., Gartner, Z. J., Chow, E. D., Bivona, T. G., … Weissman, J. S. (2022). Lineage
678 tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*, *185*(11), 1905-
679 1923.e25. https://doi.org/10.1016/j.cell.2022.04.015

680 Zhao, B., Sedlak, J. C., Srinivas, R., Creixell, P., Pritchard, J. R., Tidor, B., Lauffenburger, D. A., &
681 Hemann, M. T. (2016). Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution. *Cell*,
682 *165*(1), 234–246. https://doi.org/10.1016/j.cell.2016.01.045

683

684 **Acknowledgments**

685 We would like to thank all members of the Hannon and IMAXT labs at the CRUK Cambridge Institute for
686 valuable discussions and advice, in particular Clare Rebbeck for assistance with home office animal
687 licensing. We would also like to thank CRUK Cambridge Institute's core facilities, in particular members
688 of the genomics core for assistance with sequencing library preparation and sequencing, members of
689 the Biological Resource Unit (BRU) for animal husbandry and members of the flow cytometry core for
690 assistance with cell sorting. This work was supported by Cancer Research
691 UK [C14303/A21143, C14303/A19926, C14303/A19927, C9545/A24042]. G.J.H. is a Royal Society
692 Wolfson Research Professor.

693 **Author contributions**

694 Conceptualization- SAW, KS, IGC, GJH. Data Curation- SAW, IGC, KS. Formal Analysis- SAW, IGC,
695 KS. Funding Acquisition- KS, GJH. Investigation- SAW, IGC, KK. Methodology- SAW, IGC, KS. Project
696 Administration- SAW, KS, IGC, GJH, Software- SAW, KS. Supervision- KS, IGC, GJH.
697 Visualization- SAW, KS, IGC. Writing- Original Draft- SAW, KS, IGC. Writing- reviewing and
698 editing- SAW, KS, IGC, GJH.

699 **Data availability**

700 Single cell RNA-Seq data are being deposited in the gene expression omnibus (GEO) and will be made
701 available upon publication. All other data are available from the corresponding authors upon reasonable
702 request.

703 **Methods**

704

705 **Cell lines and culture**

706 The mouse mammary tumor cell lines 4T1 (ATCC) and D2A1-m2 (kind gifted from Clare Isacke's lab)
707 and the 293FT (Thermo Fisher Scientific) packaging cell line for virus production were cultivated in
708 DMEM high glucose (Gibco), supplemented with 10% heat-inactivated fetal bovine serum (Gibco) and
709 50 U/mL penicillin-streptomycin (Gibco).

710

711 **Virus production**

712 The WILD-seq library was packaged using 293FT lentivirus packaging cells. Cells were plated on 15
713 cm adherent tissue culture plates (Corning) one day before transfection at a confluency of ~70%.
714 Lentiviral particles were produced by co-transfecting 293FT cells with the transfer plasmid and standard
715 third-generation packaging vectors pMDL (12.5 µg), CMV-Rev (6.25 µg) and VSV-G (9 µg) using the
716 calcium-phosphate transfection method (Invitrogen). The transfection mixture was added to the
717 packaging cells along with 100 mM chloroquine (Sigma-Aldrich). After 16-18 h, media was replaced for
718 fresh growth media. Viral supernatant was collected 48 h after transfection and filtered through a 45µm
719 filter. The viral supernatant was applied directly to cells or stored at 4°C for short-term storage or -80°C
720 for long-term storage. When necessary, virus was concentrated using ultracentrifugation. Lentiviral titre
721 was determined by serial dilutions and measurements of fluorescence via flow cytometry.

722

723 **WILD-seq library design and cloning**
724 The pHSW8 lentiviral backbone was constructed using a four-way Gibson Assembly (NEB) by inserting
725 a reverse expression cassette, consisting of a PGK promoter, the zsGreen ORF, a cloning site for high-
726 diversity barcode libraries and a synthetic polyA signal, into an empty pCCL-c-MNDU3-X backbone
727 (#81071 Addgene). To generate the WILD-seq library, a barcode cassette was introduced at the cloning
728 site within the pHSW8 lentiviral backbone, using PCR (Q5 High-Fidelity DNA Polymerase, NEB) and
729 Gibson Assembly (NEB), such that it is expressed within the 3'UTR of the zsGreen transcript.
730

Name	Sequence
Assembly_Fwd	5'-AAACTCTTGAGTGAACTCCAGTGATTTTGAACCAAGCGATTCAAAGTTCT-3'
Assembly_Rev	5'-ccttgccctgaTAACTGGAGGCAGTAATTTACAGCCATGCGCTCGTTTAC-3'
BarcodeOligo_Fwd	5'-TGAACCAAGCGATTCAAAGTTCTATCCG**NNNNNNNNNNNN**tgcatcggttaaccgatgca-3'
BarcodeOligo_Rev	5'-ATGCGCTCGTTTACTATACGAT**NNNNNNNNNNNN**tgcatcggttaaccgatgca-3'

731
732 The barcode library was designed by generating 12 nt variable sequences using the R package
733 DNABarcodes (Buschmann, 2017) and a set Hamming distance of 5. The resulting pool of sequences
734 was then purchased as a custom oligo pool (Twist Bioscience). Reverse complement oligos
735 (BarcodeOligo_Fwd/Rev) each containing a specific PCR handle, a 12-bp variable region and 20-bp
736 constant linker were annealed and amplified by PCR for 20 cycles (using Assembly_Fwd/Rev primers).
737 The amplified barcode library was column purified (Gel extraction kit, Qiagen) and the vector backbone
738 was prepared by digestion with SwaI (NEB). WILD-seq barcodes were inserted into the lentiviral vector
739 backbone through Gibson Assembly (NEB), concentrated and transformed into 10b electrocompetent
740 *E.coli* cells (NEB).
741
742 **Bottlenecking strategy and characterisation of WILD-seq pools**
743 4T1 or D2A1-m2 cells were infected with WILD-seq library at low MOI (~ 0.2-0.3). Two days after
744 infection, the desired number of zsGreen positive cells, ranging from 10 to 1250 cells, were collected
745 and cultured for two weeks to allow for the pool of clones to stabilize. Different pooling strategies were
746 tested, the ultimate WILD-seq pool was generated from three independent pools each established from
747 250 sorted cells, maintained separately and mixed in equal proportions immediately prior to injection.
748
749 **Library complexity analysis**
750 WILD-seq barcodes of the lentiviral library were amplified using a one-step PCR protocol. 1 ng plasmid
751 was used as template in four separate PCR reactions to account for PCR biases and errors. All
752 reactions were pooled, concentrated and purified on a column and then sequenced on one lane of
753 HiSeq4000. Reads that contained the WILD-seq barcode motif were identified and extracted from the
754 FASTQ files. Detected WILD-seq barcode were filtered based on a 90th percentile cut-off. The resulting
755 whitelist was further filtered for barcodes that contain the common linker region.
756
757 **Whitelist generation of WILD-seq barcodes**
758 To generate a comprehensive whitelist of expressed barcodes in each pool, RNA was extracted from
759 WILD-seq transduced cells (High Pure RNA isolation kit, Roche) and reverse transcribed using the
760 Superscript IV reverse transcription kit (Invitrogen) and a target site-specific primer with a unique
761 molecular identifier (UMI) and an Illumina sample index. cDNA was amplified by PCR (Q5 High-Fidelity
762 DNA Polymerase, NEB) using primers (RTWhitelist_Fwd/Rev) containing Illumina-compatible adapters.
763 Alternatively, 1 μg of gDNA was extracted from WILD-seq transduced cells (Blood&Cell Culture DNA
764 Kit, Qiagen) and the barcode amplified by PCR using primers containing Illumina-compatible adapters
765 (gDNAWhitelist_Fwd/Rev). PCR products were purified via gel extraction (Qiagen) and quantified by
766 Qubit. The library was sequenced on an Illumina MiSeq with a custom sequencing primer for Read1
767 (CustomRead1).
768

Name	Sequence
RT Primer	5'-CAAGCAGAAGACGGCATACGAGAT**NNNNNN**GTGACTGGAG TTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAAGCGATTCAAAGTTCTATCCG-3'
RTWhitelist_Rev	5'-CAAGCAGAAGACGGCATACGA-3'
RTWhitelist_Fwd	5'-AATGATACGGCGACCACCGAGATCTACACCAGCAGTATGCATG CGCTCGTTTACTATACGAT-3'
gDNAWhitelist_Fwd	5'-AATGATACGGCGACCACCGAGATCTACACCAGCAGTATGCATGC GCTCGTTTACTATACGAT-3'
gDNAWhitelist_Rev	5'-CAAGCAGAAGACGGCATACGAGAT**NNNNNN**GTGACT GGAGTTCAGACGTGTGCTCTTCCGATCCAAGCGATTCAAAGTTCTATCCG-3'
CustomRead1 Primer	5'-CCAGCAGTATGCATGCGCTCGTTTACTATACGAT-3'

Reads from the RT-PCR barcode library that contained the WILD-seq barcode motif were identified and the number of unique UMIs supporting each barcode was calculated. If barcode sequences amplified from gDNA were also available an additional filtering step was included and any barcodes not also detected in the gDNA library excluded from the whitelist. Based on UMI counts, the top 90th percentile of detected barcodes were taken and collapsed for PCR and sequencing errors using hierarchical clustering and combining sequences with a Hamming distance less than 5.

Single cell library preparation

Tumor tissues were collected, minced and dissociated using the gentleMACS Octo Dissociator (Miltenyi Biotec) and the relevant kit (Tumor Dissociation Kit mouse). Tissues were process into single cell suspensions following manufacturer's instructions and filtered through 70 µm filters (Miltenyi) to remove any remaining larger particles from single cell suspension after dissociation. The cell suspension was concentrated and filtered again through a 70 µm filter. Three million live cells were sorted based on live-dead staining with propidium iodide to remove dead cells and debris, pelleted and resuspended in 1 mL phosphate-buffered saline with 0.04% bovine serum albumin (Sigma Aldrich). Cells were counted with a hemocytometer to ensure accurate concentration. The final single cell suspension was diluted as required and NGS libraries were prepared using Chromium Single Cell 3' Reagent Kit (v3.1 Chemistry Dual Index, user guide reference: CG000315) with no modifications.

Enrichment library preparation

To enrich for WILD-seq barcodes, the amplified cDNA libraries were further amplified with WILD-seq-specific primers containing Illumina-compatible adapters and sample indices:

Name	Sequence
Enrich_Fwd	5'-AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNNACACTCTTTCCCTACACGACGCTC-3'
Enrich_Rev	5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCAGCCATGCGCTCGTTTACTATAC-3'
"N" denotes sample indices	

1 µL amplified cDNA library was used as template in a 29-cycle PCR reaction using KAPA HiFi HotStart ReadyMix (Roche). To avoid possible PCR-induced library biases, six reactions were run in parallel. All reactions were combined, purified by columns (Gel purification kit, Qiagen) and quantified by Qubit. Gene expression libraries and barcode enrichment libraries were pooled in an approximately 10:1 molar ratio and libraries were sequenced on the NovaSeq platform (Illumina).

Animals and *in vivo* dosing

All mouse experiments were performed under the Animals (Scientific Procedures) Act 1986 in accordance with UK Home Office licenses (Project License # PAD85403A) and approved by the Cancer Research UK (CRUK) Cambridge Institute Animal Welfare and Ethical Review Board. Female six to eight week-old BALB/c were purchased from The Charles River Laboratory. 60,000 tumor cells were resuspended in 50 µL of a 1:1 mixture of PBS and growth-factor reduced Matrigel (Corning). All orthotopic injections were performed into the fourth mammary gland. Primary tumor volume was measured using the formula $V=0.5(LxW^2)$, in which W is the with and L is length of the primary tumor.

Tumor-bearing mice were treated with either vehicle or with different drugs from seven days post transplantation. All drugs were administered via intraperitoneal injection. For JQ1 treatment, animals were dosed 75 mg/kg JQ1 (dissolved in DMSO and diluted 1:10 in 10% β-cyclodextrin) 5 days/week (5 consecutive days followed by 2 days off) until tumors reached endpoint. For docetaxel treatment, animals were dosed at 12.5 mg/kg docetaxel (dissolved in 1:1 mixture of ethanol and Kolliphor and diluted 1:4 in saline) 3 times/week, except when L-asparaginase was to be administered concurrently and then the dose was reduced to 10 mg/kg. For L-asparaginase treatment, mice were administered 100 µL of 60 U L-asparagine (Abcam) diluted in saline. Vehicle-treated mice were sacrificed 21 days post tumor transplantation and treated animals were sacrificed when tumor volumes reached that of vehicle treated animals at 21 days unless otherwise stated.

scRNA-seq analysis

scRNA-seq libraries generated by the 10X Chromium platform were processed using CellRanger version 3.0.1. Reads were aligned to a custom reference genome that was created by adding the sequence of the zsGreen-WILD-seq barcode transgene as a new chromosome to the mm10 mouse

genome. The gene expression matrices generated were then analyzed with the Seurat R package (Stuart et al., 2019) using a standard pipeline. Briefly, datasets were first filtered based on the number of unique genes detected per cell (typical accepted range 200-10000 genes) and the percentage of reads that map to the mitochondrial genome (< 12 %). Reads which mapped to the zsGreen-WILD-seq barcode transgene were removed from the count matrix to prevent these driving cell clustering. Normalisation was performed using sctransform, including cell cycle regression. Differential abundance of cell subtypes was performed using Milo (Dann et al., 2021).

Clonal barcode assignment to single cell data

Extraction of WILD-seq barcodes from scRNA-seq data: Reads mapping to the zsGreen-WILD-seq barcode transgene and containing the full barcode sequence (20nt constant linker with a 12 nt variable region on either side) were extracted from the BAM file produced by Cell Ranger and mapped using Bowtie to a whitelist of barcodes expressed in the WILD-seq cell pool. A WILD-seq clonal barcode was assigned to a cell if there were at least 2 independent reads which matched the barcode to the cell and more than 50% of barcode mapped reads from the cell supported the assignment.

Extraction of WILD-seq barcodes from PCR enrichment data: Reads from the PCR barcode enrichment were processed separately using the UMI-tools to extract 10X cell barcodes and UMIs from the raw read files. The sequence corresponding to the full barcode sequence (20nt constant linker with a 12 nt variable region on either side) was extracted from each read and then mapped to the WILD-seq clonal barcode whitelist using Bowtie. A WILD-seq clonal barcode was assigned to a cell if there were at least 10 UMIs which matched the barcode to the cell and at least twice as many UMIs supporting this assignment compared to the next best.

WILD-seq barcode assignment: The WILD-seq clonal barcode assignment from these 2 pipelines was then compared. If the assignment from the transcriptomic analysis and the PCR enrichment analysis were in agreement the barcode was assigned. On the rare occasion the assignment didn't match a clonal barcode was not assigned. If a cell was assigned a WILD-seq barcode by only one method, a further more stringent filtering step was included. For WILD-seq barcodes assigned only from the 10X scRNA-seq dataset but not the PCR-enrichment, the minimum number of UMIs required to support the assignment was increased to 5 and for WILD-seq barcodes assigned only from the PCR-enrichment but not the 10X scRNA-seq dataset, the minimum number of UMIs required to support the assignment was increased to 30.

Differential gene expression

Differential gene expression was determined using the FindMarkers function in Seurat with a Wilcoxon rank sum test to identify differentially expressed genes. For differential expression of groups of genes, we used the AUCell R package (Aibar et al., 2017) which enables analysis of the relative expression of a gene set (i.e. gene signature or pathway) across all the cells in single-cell RNA-seq data using the "Area Under the Curve" (AUC) to calculate the enrichment of the input geneset within the expressed genes for each cell. An AUCell score was calculated for each tumor cell for every gene set in the MSigDB C2 collection (Liberzon et al., 2011; Subramanian et al., 2005) that contained more than 20 genes with detectable expression in our data. AUCell scores were compared across clones or conditions using a Wilcoxon rank sum test and p-values were adjusted for multiple comparison using the Benjamini-Hochberg correction method.

To generate baseline transcriptomic signatures for each clone in vehicle-treated tumors, comparisons were made between the clone of interest and all assigned tumor cells from the same sample (in the case of D2A1 tumors) or the same experiment (in the case of 4T1 tumors). Samples/experiments were included if they contained at least 20 cells assigned to the clone of interest. To define consistently enriched/depleted signatures, p-values from comparisons within each sample/experiment were combined using the Fisher's method.

Patient data analysis

Microarray gene expression data was downloaded from GSE28844 (Vera-Ramirez et al., 2013). A single probe for each gene was selected based on the highest median expression. Gene set expression per patient sample was calculated using GSVA (Hänzelmann et al., 2013).

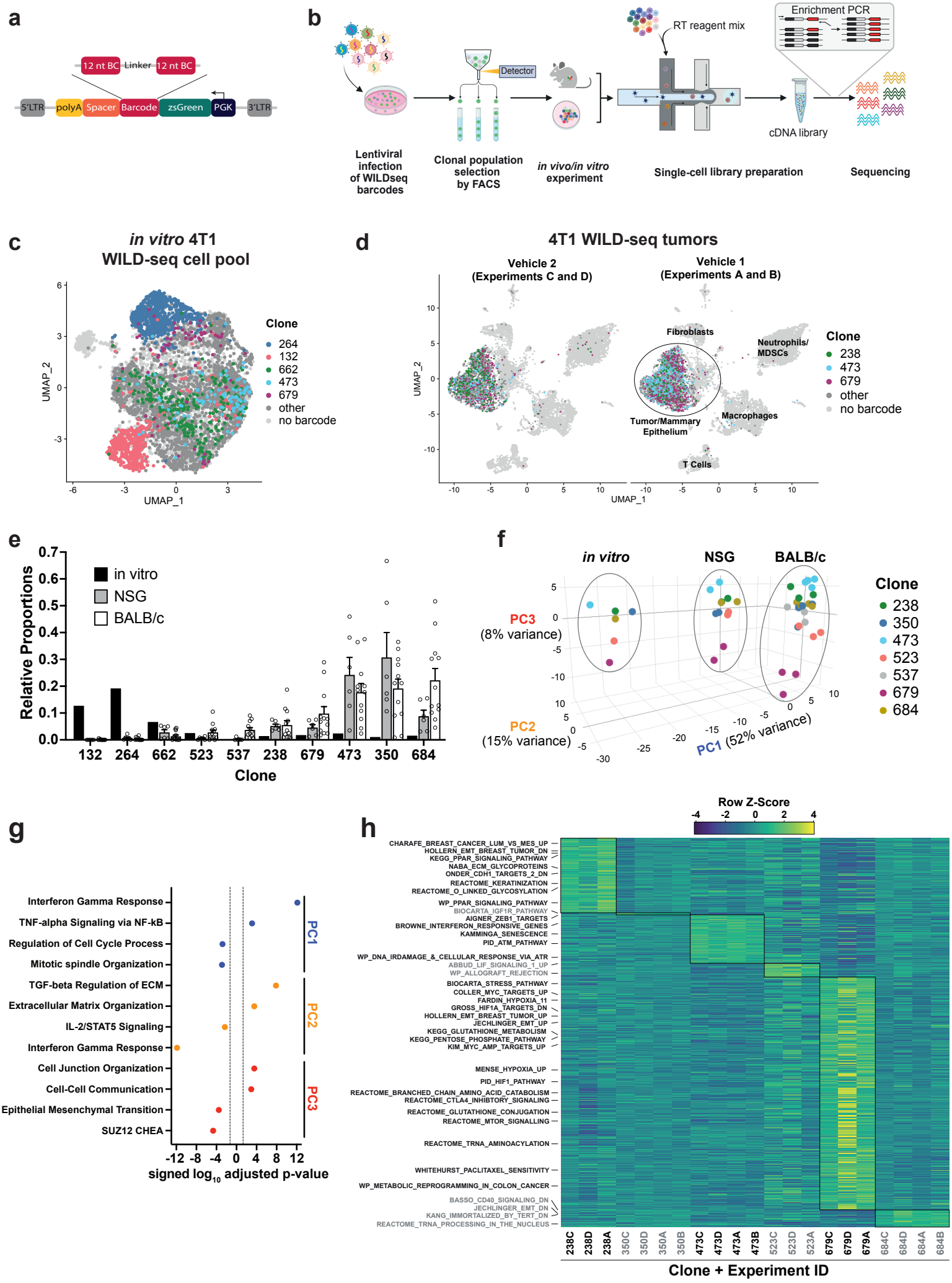

Figure 1. WILD-seq

Figure 1. Establishment of an expressed barcode system to simultaneously detect clonal lineage and gene expression from single cells *in vivo*.
a. Lentiviral construct design. A PGK promoter drives expression of a transcript encoding zsGreen harboring a WILD-seq barcode sequence in the 3'UTR. A spacer sequence and polyadenylation signal ensure that that the barcode is detectable as part of a standard oligo dT single cell RNA library preparation and sequencing pipeline. The barcode cassette comprises 2 distinct 12 nucleotide barcode sequences separated by a constant 20 nucleotide linker region. The library of barcode sequences was designed with Hamming distance 5 to allow for sequencing error correction. **b. Schematic of WILD-seq method.** Tumor cells are infected with the WILD-seq lentiviral library and an appropriate size population of zsGreen positive cells isolated, each of which will express a single unique WILD-seq barcode. This WILD-seq barcoded, heterogenous cell pool is then subjected to an intervention of interest (such as *in vivo* treatment of the implanted pool with a therapeutic agent) and subsequently analyzed by single cell RNA sequencing using the 10X Genomics platform. An additional PCR amplification step is included that specifically enriches for the barcode sequence to increase the number of cells to which a WILD-seq barcode can be conclusively assigned. **c. scRNA-seq of *in vitro* 4T1 WILD-seq cell pool.** UMAP plot of *in vitro* cultured 4T1 WILD-seq cells. Cells for which a WILD-seq clonal barcode is identified are shown as dark grey or colored spots. Cells which belong to five selected clonal lineages are highlighted. **d. scRNA-seq of 4T1 WILD-seq tumors.** UMAP plots of vehicle-treated 4T1 WILD-seq tumors generated by injecting the 4T1 WILD-seq pool into the mammary fatpad of BALB/c mice. Four independent experiments were performed each involving injection into 3 separate host animals. Six animals from experiments A and B received vehicle 1 (10% DMSO, 0.9% β-cyclodextrin) and six animals from experiments C and D received vehicle 2 (12.5% ethanol, 12.5% Kolliphor). **e. Clonal representation.** Proportion of tumor cells assigned to each clonal lineage based on the WILD-seq barcode (n = 1 for *in vitro* cultured cells, n = 6 for tumors from NSG mice, n = 12 for vehicle-treated tumors from BALB/c mice). Selected clones from the most abundant lineages are plotted. Data represents mean ± SEM. **f. Principal component analysis of clonal transcriptomes.** Pseudo-bulk analysis was performed by summing counts for all tumor cells expressing the same WILD-seq clonal barcode within an independent experiment. For *in vivo* tumor samples each point represents the combined cells from 3 animals. Principal component analysis of normalized pseudo-bulk count data showed separation of samples by origin with PC1 and PC2 and separation by clonality with PC3. **g. Transcriptomic programs associated with principal components.** The top/bottom 50 gene loadings of PC1, PC2 and PC3 were analyzed using Enrichr (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). **h. Clonal transcriptomic signatures from vehicle-treated BALB/c tumors.** An AUCell score (Aibar et al., 2017) enrichment was calculated for each clone and for each experiment by comparing cells of a specific clonal lineage of interest to all assigned tumor cells within the same experiment. All gene sets which showed consistent and statistically significant enrichment in one of the six most abundant clones across experiments are illustrated.

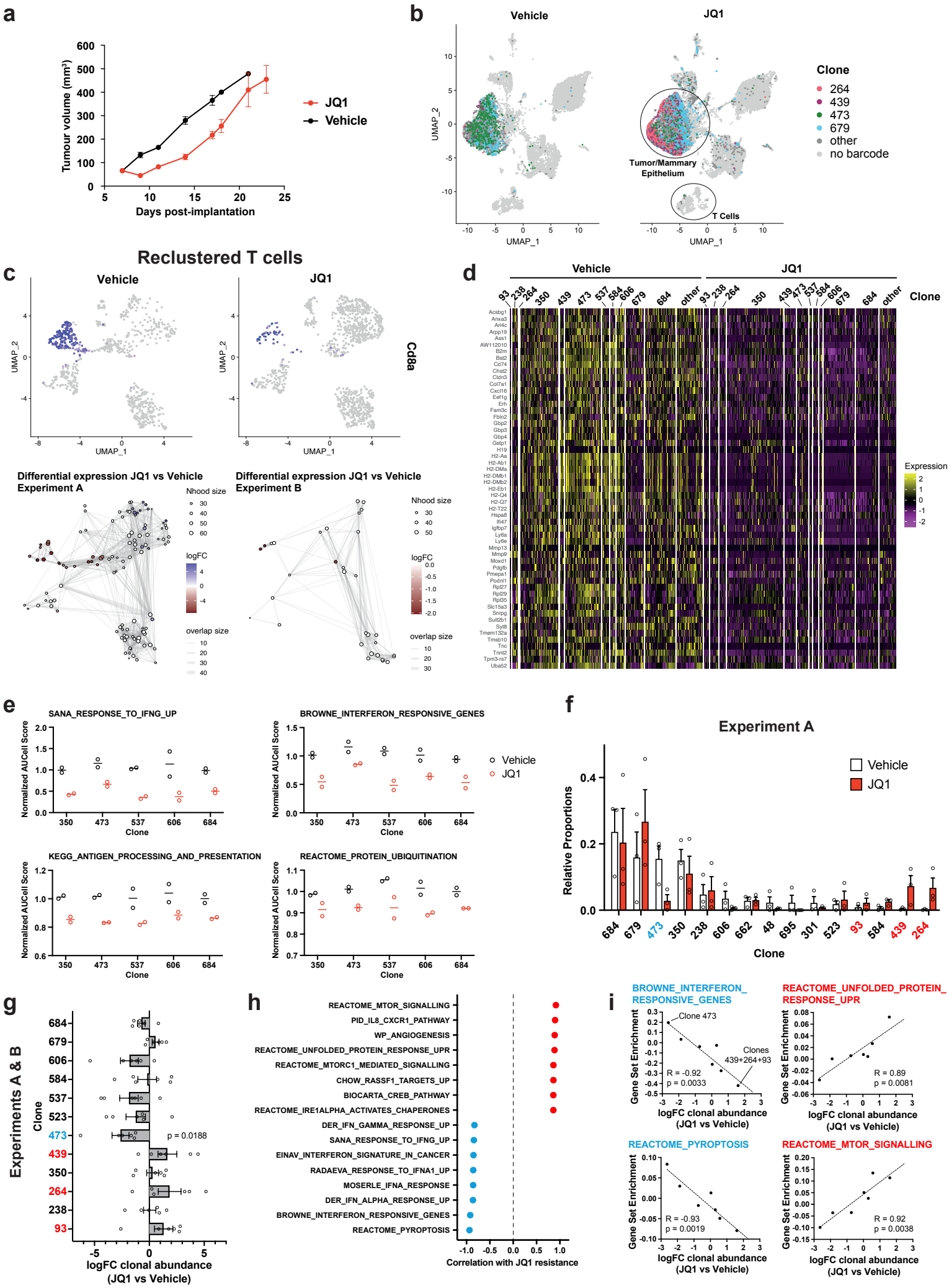

Figure 2. WILD-seq

Figure 2. Simultaneous detection of changes in clonal abundance, gene expression, and tumor microenvironment in response to BET bromodomain inhibition with WILD-seq.
a. Tumor growth curves with JQ1 treatment. 4T1 WILD-seq tumors were treated with the BET bromodomain inhibitor JQ1 or vehicle from 7 days post-implantation until endpoint (n = 4 mice per condition). Data represents mean ± SEM. **b. scRNA-seq of JQ1-treated 4T1 WILD-seq tumors.** UMAP plots of vehicle- or JQ1-treated 4T1 WILD-seq tumors. Combined cells from 2 independent experiments, each with 3 mice per condition are shown. Cells for which a WILD-seq clonal barcode is identified are shown as dark grey or colored spots. Cells which belong to four selected clonal lineages are highlighted. **c. JQ1-treatment results in a reduction in Cd8+ tumor-associated T-cells.** Cells belonging to the T-cell compartment were computationally extracted from the single cell data and reclustered. Upper panels show combined UMAP plots from experiments A and B with Cd8a expression per cell illustrated enabling identification of the Cd8+ T cell cluster. Lower panels show neighborhood graphs of the results from differential abundance testing using Milo (Dann et al., 2021). Colored nodes represent neighbourhoods with significantly different cell numbers between conditions (FDR < 0.05) and the layout of nodes is determined by the position of the neighborhood index cell in the UMAP panel above. Experiments A and B were analyzed separately due to differences in cell numbers. **d. Differential gene expression between JQ1- and vehicle-treated tumor cells.** Single cell heatmap of expression for genes which are significantly and consistently down-regulated across clonal lineages (combined fisher p-value < 0.05 and mean logFC < -0.2 for both experiments).1600 cells are represented (400 per experiment/condition), grouped according to their clonal lineage. **e. Differential gene set expression between JQ1 and vehicle-treated tumor cells.** Median AUCell score per experiment/condition for selected gene sets. The 5 clonal lineages with the highest representation across experiments are shown. **f. Clonal representation.** Proportion of tumor cells assigned to each clonal lineage in experiment A based on the WILD-seq barcode (n = 3 tumors per condition). Clones which make up at least 2% of the assigned tumor cells under at least one condition are plotted. The most sensitive clone 473 is highlighted in blue and the most resistant clones 93, 439, 264 are highlighted in red. Data represents mean ± SEM. **g. Clonal response to JQ1-treatment.** Log_2 fold change in clonal proportions upon JQ1 treatment across experiments A and B. Fold change was calculated by comparing each JQ1-treated sample with the mean of the 3 corresponding vehicle-treated samples from the same experiment. p-value calculated by one-sample t-test vs a theoretical mean of 0. Data represents mean ± SEM. **h. and i. Correlation of JQ1-response with baseline clonal transcriptomic signatures.** Clonal gene set enrichment scores for vehicle-treated tumors were calculated by comparing cells of a specific clonal lineage of interest to all assigned tumor cells within the same experiment. Correlation between these scores and JQ1-treatment response (mean log_2 fold change clonal proportion JQ1 vs vehicle) was then calculated for each gene set. Selected gene sets with the highest positive or negative correlation values (Pearson correlation test) are shown. A positive correlation indicates a higher expression in resistant clones, whereas a negative correlation indicates a higher expression in sensitive clones. Resistant clonal lineages identified by barcodes 93, 264 and 439 were combined for the purpose of this analysis to have enough cells for analysis within the vehicle-treated samples.

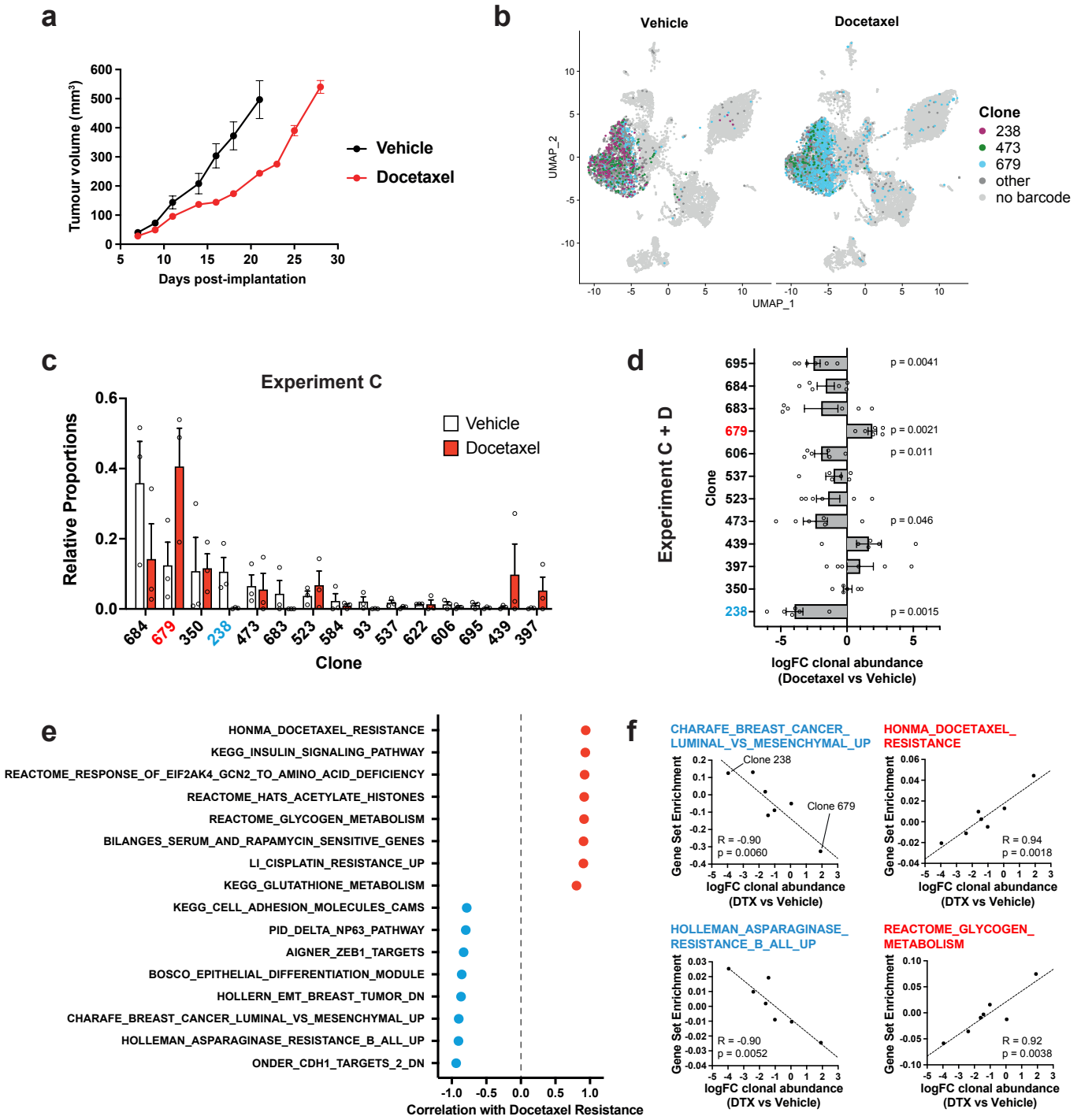

Figure 3. WILD-seq

Figure 3. Clonal transcriptomic correlates of response and resistance to taxane chemotherapy in the 4T1 mammary carcinoma model.

a. Tumor growth curves with docetaxel treatment. 4T1 WILD-seq tumors were treated with docetaxel or vehicle (12.5% ethanol, 12.5% Kolliphor) from 7 days post-implantation for 2 weeks (n = 5 mice per condition). Dosing regimen was 12.5 mg/Kg docetaxel three times per week. Data represents mean ± SEM. **b. scRNA-seq of docetaxel-treated 4T1 WILD-seq tumors.** UMAP plots of vehicle- or docetaxel-treated 4T1 WILD-seq tumors. Combined cells from 2 independent experiments, each with 3 mice per condition are shown. Cells for which a WILD-seq clonal barcode is identified are shown as dark grey or colored spots. Cells which belong to three selected clonal lineages are highlighted. **c. Clonal representation.** Proportion of tumor cells assigned to each clonal lineage in experiment C based on the WILD-seq barcode (n = 3 tumors per condition). Clones which make up at least 2% of the assigned tumor cells under at least one condition are plotted. The most sensitive clone 238 is highlighted in blue and the most resistant clone 679 is highlighted in red. Data represents mean ± SEM. **d. Clonal response to docetaxel-treatment.** Log_2 fold change in clonal proportions upon docetaxel treatment across experiments C and D. Fold change was calculated by comparing each docetaxel-treated sample with the mean of the 3 corresponding vehicle-treated samples from the same experiment. p-values calculated by one-sample t-test vs a theoretical mean of 0. Data represents mean ± SEM. **e. and f. Correlation of docetaxel-response with baseline clonal transcriptomic signatures.** Clonal gene set enrichment scores for vehicle-treated tumors were calculated by comparing cells of a specific clonal lineage of interest to all assigned tumor cells within the same experiment. Correlation between these scores and docetaxel-treatment response (mean log_2 fold change clonal proportion docetaxel vs vehicle) was then calculated for each gene set. Selected gene sets with the highest positive or negative correlation values (Pearson correlation test) are shown. A positive correlation indicates a higher expression in resistant clones, whereas a negative correlation indicates a higher expression in sensitive clones.

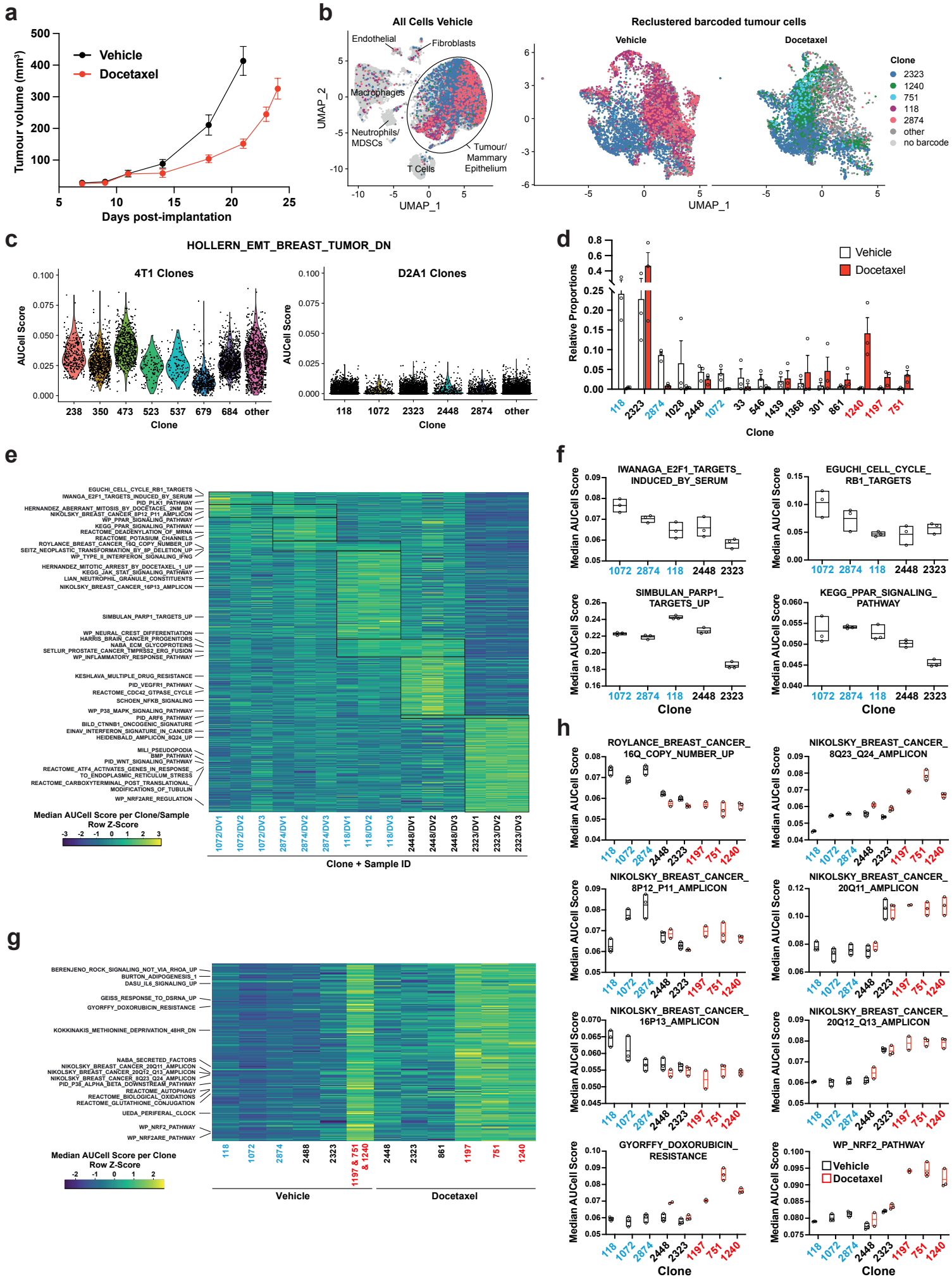

Figure 4. WILD-seq

Figure 4. Clonal transcriptomic signatures of response and resistance to taxane chemotherapy in the D2A1 mammary carcinoma model. a. D2A1 WILD-seq tumor growth curves with docetaxel treatment. D2A1 WILD-seq tumors were treated with docetaxel or vehicle from 7 days post-implantation for 2 weeks (n = 5 vehicle-treated mice, n = 4 docetaxel-treated mice). Data represents mean ± SEM. **b. scRNA-seq of docetaxel-treated D2A1 WILD-seq tumors.** UMAP plots of vehicle-treated D2A1 WILD-seq D2A1 tumors and reclustered barcoded-tumor cells from vehicle- and docetaxel-treated tumors. Combined cells from 3 mice per condition are shown. Cells for which a WILD-seq clonal barcode is identified are shown as dark grey or colored spots. Cells which belong to five selected clonal lineages are highlighted. **c. Comparison of EMT status of major 4T1 and D2A1 WILD-seq clones.** Violin plot of AUCell scores from vehicle-treated tumor cells generated using the HOLLERN_EMT_BREAST_TUMOR_DN (Hollern et al., 2018) gene set, a set of genes that have low expression in murine mammary tumors of mesenchymal histology. 4T1 WILD-seq clones exhibit varying levels of expression of this geneset whereas D2A1 WILD-seq clones have consistently low levels of expression of these genes. **d. Clonal representation.** Proportion of tumor cells assigned to each clonal lineage based on the WILD-seq barcode (n = 3 tumors per condition). Clones which make up at least 2% of the assigned tumor cells under at least one condition are plotted. The most sensitive clones to docetaxel treatment 118, 2874 and 1072 are highlighted in blue and the most resistant clones 1240, 1197 and 751 are highlighted in red. Data represents mean ± SEM. **e. Clonal transcriptomic signatures from vehicle-treated tumors.** Heatmap of median AUCell scores per sample for each of the five most abundant clones. All gene sets which showed consistent and statistically significant enrichment (combined fisher p-value < 0.01 & mean $\log_2$ enrichment > 0.1) in at least one of these clones are illustrated. **f. Selected gene sets whose expression is associated with sensitivity to docetaxel.** Median AUCell scores per sample for each of the five most abundant clones is plotted. **g. Transcriptomic signatures associated with resistance to docetaxel.** For vehicle-treated tumors, resistant clonal lineages identified by barcodes 1197, 751 and 1240 were combined to have enough cells for analysis. Gene sets with significantly enriched expression in these resistant clones in vehicle-treated tumors were determined (adjusted p-value < 0.01 & $\log_2$ enrichment > 0.1). A heatmap of median AUCell scores per clone, per condition of these resistance-associated gene sets is plotted. **h. Selected gene sets whose expression is enriched or depleted in resistant clones.** Median AUCell scores per clone, per sample are plotted for samples with at least 20 cells per clone. Due to changes in clonal abundance with treatment some clones can only be assessed under vehicle- or docetaxel-treated conditions.

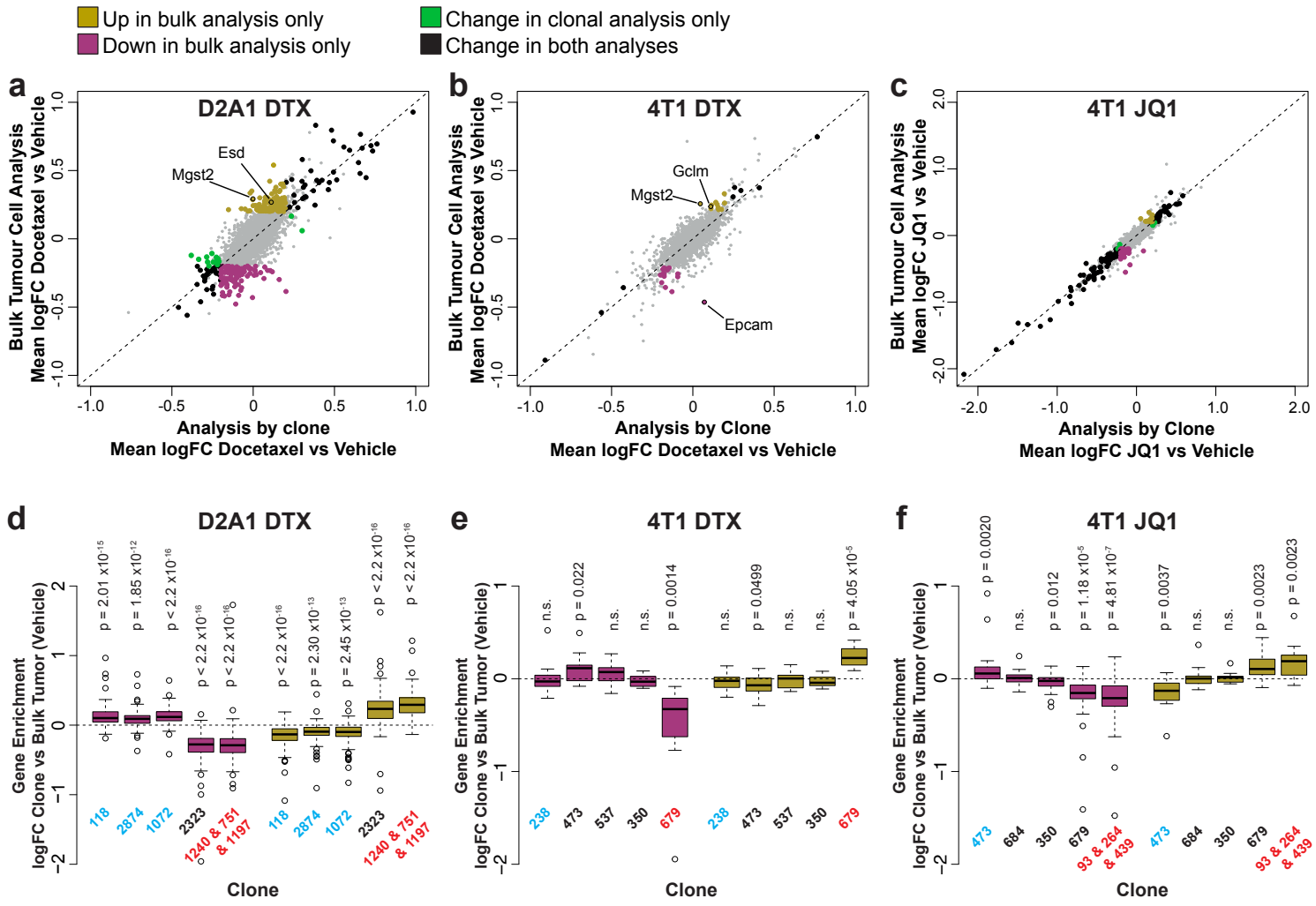

Figure 5. WILD-seq

Figure 5. Delineating the contribution of clonal abundance to gene expression changes upon drug treatment. a. Comparison of differential gene expression analysis in bulk tumor cells and intra-clonal changes in gene expression. Differential gene expression was performed for all barcoded tumor cells irrespective of clonal lineage comparing chemotherapy-treated and vehicle-treated cells (bulk tumor cell analysis). Alternatively differential gene expression was performed for each individual clone separately and the results combined to identify genes which robustly undergo intra-clonal changes in expression (analysis by clone). Whereas bulk tumor cell analysis will identify changes in overall gene expression due to both changes in clonal abundance and changes within the cells, analysis by clone enables us to delineate exclusively induced cellular changes in gene expression. Log_2 fold change in expression as determined by each of these analysis methods is plotted. Genes with significant changes in expression with chemotherapy (p-value < 0.05, logFC < -0.2 or > 0.2) are highlighted based on the method under which they were identified. Genes identified as significantly changing by one method only met neither logFC nor p-value cutoffs in the alternative method. **b. Changes in gene expression that are identified by bulk tumor cell analysis only can attributed to changes in clonal abundance.** The expression of genes which were identified as differential expressed after chemotherapy only in the bulk tumor cell analysis was assessed across clonal lineages at baseline. Baseline gene enrichment for each clone was determined as described previously by comparing cells of a specific clonal lineage to all barcoded tumor cells within the same vehicle-treated sample or experiment. Gene enrichment values for all genes with differential expression only in the bulk tumor cell analysis were plotted. As expected, genes down-regulated in bulk analysis have lower expression in resistant clones, whereas genes up-regulated in bulk analysis are enriched in resistant clones. p-values represent a one sample t-test vs a theoretical mean of 0.

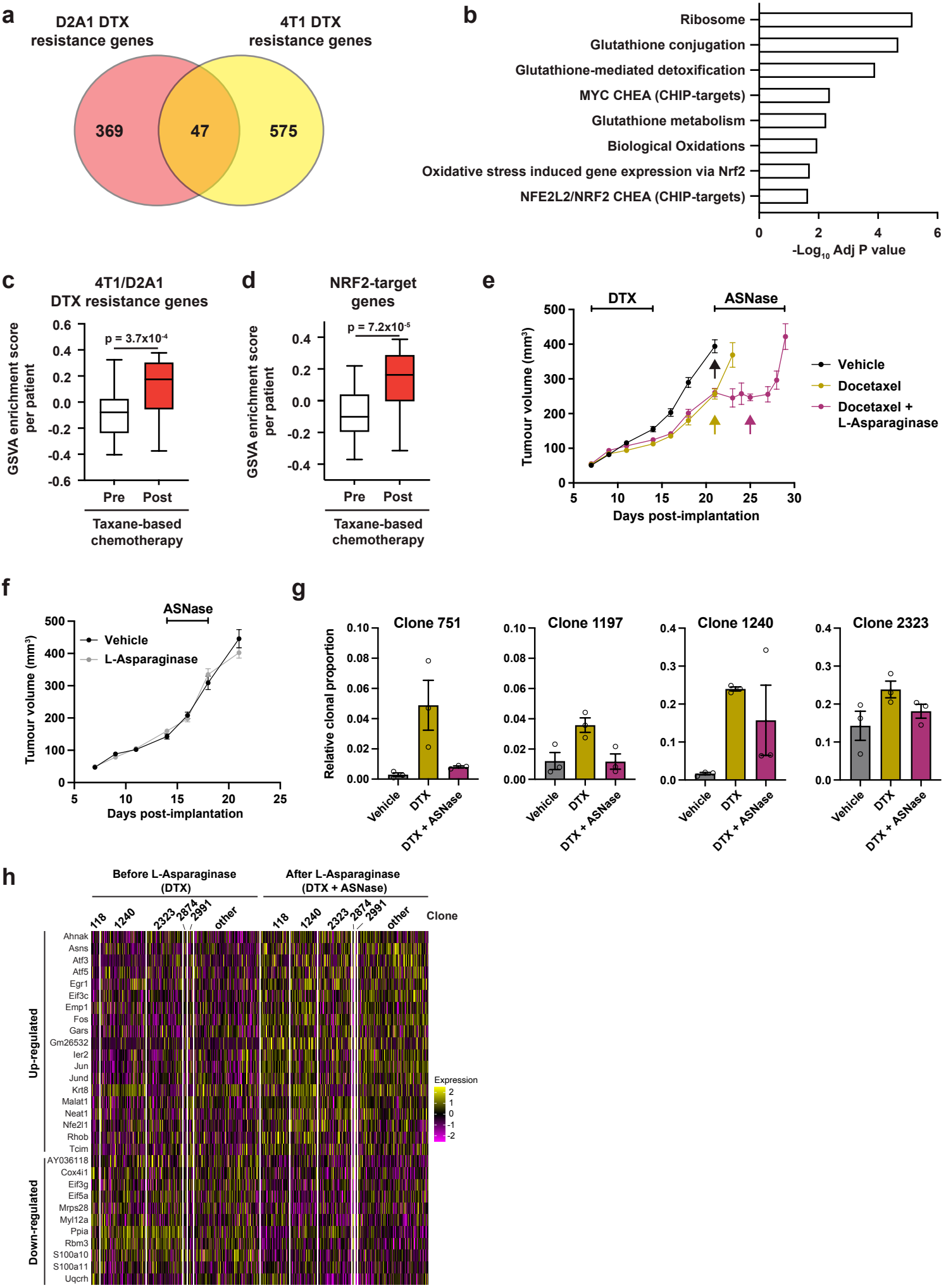

Figure 6. WILD-seq

Figure 6. Taxane-resistant clones have elevated NRF2 signaling and are sensitive to asparagine deprivation. a. Overlap of genes associated with resistance between the D2A1 and 4T1 WILD-seq models. 4T1 resistance genes were defined as those that were significantly enriched in resistant clone 679 but not in sensitive clone 238 ($p < 0.05$). D2A1 resistance genes were defined as those that were significantly enriched in combined resistant clones 1240, 751 and 1197 but not in sensitive clones 118, 2874 or 1072 ($p < 0.05$). In all cases, resistance genes were defined from vehicle treated tumors **b. Gene set enrichment analysis of common resistance genes.** Gene set enrichment was performed using Enrichr for the human orthologs of the 47 common resistance genes identified in Fig. 6a. Adjusted p-values for a subset of significant gene sets are plotted. **c. Expression of our identified resistance genes is increased in human breast tumors following taxane-based chemotherapy.** Expression of our 47 common resistance genes was assessed in human breast cancer samples taken before and after taxane-based neoadjuvant chemotherapy (GSE28844). GSVA enrichment scores for our gene set was calculated for samples from 28 patients for which matched pre- and post-treatment gene expression data were available. Patients received one of three taxane-containing treatment regimens; Regimen A: Epirubicin 90 mg/m^2-Cyclophosphamide 600 mg/m^2, 3 cycles bi-weekly and *Paclitaxel* 150 mg/m^2-Gemcitabine 2500 mg/m^2, 6 cycles bi-weekly ± weekly Herceptin 4 mg/Kg during the first week, 2 mg/Kg for the remaining 11 cycles. Regimen B: Doxorubicin 60 mg/m^2-Pemetrexed 500 mg/m^2, 4 cycles tri-weekly and *Docetaxel* 100 mg/m^2, 4 cycles tri-weekly. Regimen C: Doxorubicin 60 mg/m^2-Cyclophosphamide 600 mg/m^2, 4 cycles tri-weekly and *Docetaxel* 100 mg/m^2, 4 cycles tri-weekly. Expression of our common resistance gene set was significantly increased after chemotherapy in human samples. p-value calculated by paired t-test. **d. NRF2-target genes are upregulated in human patients following neoadjuvant chemotherapy.** GSVA enrichment scores for NRF2-target genes (NFE2L2 CHEA consensus CHIP-targets) were calculated for samples from 28 patients in the GSE28844 dataset for which pre- and post-treatment gene expression data were available. p-values calculated by paired t-test. **e. Docetaxel-resistant tumors are collaterally sensitive to L-asparaginase.** D2A1 WILD-seq tumors were treated with 3 doses of 12.5 mg/kg docetaxel (days 7,9,11 post-implantation) and 1 dose of 10 mg/kg docetaxel (day 14 post-implantation). From day 21 mice were treated daily with L-asparaginase. Arrows indicate timepoints of tumor collection for single-cell sequencing. Measurements are combined from 2 independent experiments. Due to sample collection at timepoints indicated the number of animals is reduced beyond this. Vehicle n = 15 mice, docetaxel n = 14 mice (reduced to 5 mice after day 21), docetaxel + L-asparaginase n = 13 mice (reduced to 4 mice from day 25). In addition, 2 mice reached humane endpoint (due to weight loss following docetaxel treatment but prior to administration of L-asparaginase) one in the DTX only arm at day 18 and one in the DTX+L-Asp arm at day 21. Data represents mean ± SEM. **f. L-asparaginase alone does not affect tumor growth.** D2A1 WILD-seq tumors were treated with L-asparaginase or vehicle for 5 consecutive days from day 14 post-implantation. n = 10 mice per condition. Data represents mean ± SEM. **g. Taxane-resistant clones are sensitive to L-asparaginase.** Relative clonal abundance in vehicle-treated (day 21), docetaxel-treated (day 21) and docetaxel and L-asparaginase-treated (day 25) D2A1 WILD-seq tumors is shown for 3 taxane-resistant clones (751, 1197, 1240) and 1 neutral clone (2323). Clonal proportions were calculated from single cell sequencing data of 3 tumors per condition. Data represents mean ± SEM. **h. Gene expression changes in tumor cells after L-asparaginase treatment.** Heatmap for genes which are most significantly and consistently differentially expressed across clonal lineages after treatment with L-asparaginase. 2400 cells are represented (400 per sample), grouped according to clonal lineage.

Description of supplementary tables

Supplementary Table 1. Overview of single cell RNA-seq samples generated.

Supplementary Table 2. Number and proportion of tumor cells assigned to each clonal barcode for all 4T1 WILD-seq sample.

Supplementary Table 3. Number and proportion of tumor cells assigned to each clonal barcode for all D2A1 WILD-seq sample.

Supplementary Table 4. 4T1 WILD-seq baseline gene enrichment signatures for major clones. Differential gene expression analysis was performed for each clone by comparing cells from a clonal lineage of interest to all assigned tumor cells within the same experiment. Only vehicle-treated samples were included in the analysis. Experiments were included in the analysis if they contained at least 20 cells assigned to the clone and clones were analyzed if they were represented by at least 20 cells in at least 3 of the 4 experiments. Differential gene expression was performed using Seurat FindMarkers function and Wilcoxon Rank Sum test. Fisher's method was used to combine p-values from separate experiments. Analysis for each clone is provided as a separate tab.

Supplementary Table 5. 4T1 WILD-seq baseline gene set enrichment signatures for major clones. Differential gene set expression analysis was performed for each clone by comparing cells from a clonal lineage of interest to all assigned tumor cells within the same experiment. All gene sets from the Molecular Signatures Database C2 curated gene set collection were included in the analysis that contained more than 20 genes detectable in our single cell data. Only vehicle-treated samples were included in the analysis. Experiments were included in the analysis if they contained at least 20 cells assigned to the clone and clones were analyzed if they were represented by at least 20 cells in at least 3 of the 4 experiments. Gene set expression analysis was performed using AUCell and differential expression was calculated using Wilcoxon Rank Sum test. Tables show median AUCell score per experiment for each gene set, enrichment in AUCell score relative to all assigned tumor cells within the same experiment ($\log_2$(median AUCell score clone of interest/median AUCell score all clones)) and adjusted p-value from Wilcoxon Rank Sum test of AUCell scores from clone of interest vs AUCell scores from all assigned tumors cells from the same experiment. Fisher's method was used to combine p-values from separate experiments. Analysis for each clone is provided as a separate tab. A final tab 'Data_for_Fig1h' provides the matrix of AUCell enrichment values used for the heatmap plotted in figure 1h compiled from individual analyses.

Supplementary Table 6. Differential expression analysis JQ1 vs Vehicle. Differential gene expression analysis was performed by comparing cells from the same clonal lineage treated with JQ1 or vehicle within the same experiment. Five clones were included in the analysis (clones 350, 473, 537, 606 and 684) for which there were at least 20 cells per condition across both experiments. Fisher's method was used to combine p-values from different clones within the same experiment. Gene level differential expression was performed using Seurat FindMarkers function and Wilcoxon Rank Sum test. These data are provided under the 'FindMarkers_JQ1vsVeh' tab. Gene set level differential expression was performed using AUCell and differential expression was calculated using Wilcoxon Rank Sum test. These data are provided under the 'AUCell_JQ1vsVeh' tab. The 'Median_norm_AUCell_Scores' tab provides a summary of the median normalised AUCell scores for each clone, condition and experiment used in the preparation of figure 2e. Normalization to enable comparison across separate experiments was performed by dividing by the median AUCell score for all vehicle-treated tumor cells assigned to any clonal lineage from the same experiment.

Supplementary Table 7. Correlation of clonal gene expression with JQ1 response. To determine genes and gene sets whose expression correlates with JQ1 response, the correlation between baseline gene and geneset enrichment values for the major clones as defined in supplementary tables 4 and 5 and the log fold change in clonal abundance between JQ1 and vehicle-treated samples was calculated using the Pearson correlation test. The Pearson correlation coefficient is provided for each gene and gene set.

Supplementary Table 8. Correlation of clonal gene expression with docetaxel response. To determine genes and gene sets whose expression correlates with docetaxel response, the correlation between baseline gene and geneset enrichment values for the major clones as defined in supplementary tables 4 and 5 and the log fold change in clonal abundance between JQ1 and vehicle-treated samples was calculated using the Pearson correlation test. The Pearson correlation coefficient is provided for each gene and gene set.

Supplementary Table 9. D2A1 WILD-seq baseline gene enrichment signatures for major clones. Differential gene expression analysis was performed for each clone by comparing cells from a clonal lineage of interest to all assigned tumor cells within the same sample. Only vehicle-treated samples were included in the analysis. Clones were included in the analysis if there were at least 20 cells assigned to that clone in all three vehicle samples (DV1, DV2 and DV3). Differential gene expression was performed using Seurat FindMarkers function and Wilcoxon Rank Sum test. Fisher's method was used to combine p-values from separate samples. Analysis for each clone is provided as a separate tab. In addition, analysis is included for the combined resistant clones 751 1197 and 1240. Due to their low representation in vehicle-treated samples cells assigned to these clones from all three vehicle-treated samples were combined for gene expression analysis and compared to all assigned tumor cells from the three samples.

Supplementary Table 10. D2A1 WILD-seq baseline gene set enrichment signatures for major clones. Differential gene set expression analysis was performed for each clone by comparing cells from a clonal lineage of interest to all assigned tumor cells within the same sample. All gene sets from the Molecular Signatures Database C2 curated gene set collection were included in the analysis that contained more than 20 genes detectable in our single cell data. Only vehicle-treated samples were included in the analysis. Clones were included in the analysis if there were at least 20 cells assigned to that clone in all three vehicle samples (DV1, DV2 and DV3). Gene set expression analysis was performed using AUCell and differential expression was calculated using Wilcoxon Rank Sum test. Tables show median AUCell score per sample for each gene set, enrichment in AUCell score relative to all assigned tumor cells within the same experiment ($\log_2$(median AUCell score clone of interest/median AUCell score all clones)) and adjusted p-value from Wilcoxon Rank Sum test of AUCell scores from clone of interest vs AUCell scores from all assigned tumors cells from the same sample. Fisher's method was used to combine p-values from separate samples. Analysis for each clone is provided as a separate tab. In addition, analysis is included for the combined resistant clones 751 1197 and 1240. Due to their low representation in vehicle-treated samples cells assigned to these clones from all three vehicle-treated samples were combined for gene expression analysis and compared to all assigned tumor cells from the three samples. The tab 'Data for Fig4e and 4f' provides the matrix of median AUCell scores used for the heatmap plotted in figure 4e compiled from individual analyses. The tab 'Data for Fig4h' provides median AUCell scores per sample for clones of interest for all samples and conditions where at least 20 cells per clone were present. Selected data from this table was plotted in figure 4h.

Supplementary Table 11. Comparison of differential gene expression analysis in bulk tumor cells and intra-clonal changes in gene expression. For each treatment condition (docetaxel/D2A1, docetaxel/4T1 and JQ1/4T1) differential expression analysis was performed between barcoded tumor cells from drug-treated and vehicle-treated animals from the same experiment. Analysis was performed either by using cells from a single clonal lineage (analysis by clone) or all barcoded tumor cells irrespective of clonal lineage (bulk tumor cell analysis). Differential gene expression was performed using Seurat FindMarkers function and Wilcoxon Rank Sum test. Log_2 fold change and adjusted p-value are provided for each comparison. For the analysis by clone, the mean logFC of all individual clonal comparisons is given (mean.logFC.clonal) and Fisher's method was used to combine p-values (fisher.combined.pvalue.clonal). Genes were classified as significantly changed in clonal analysis only, bulk analysis only or both analysis methods based on significance cutoffs of p-value < 0.05 and logFC < -0.2 or > 0.2. Genes identified as significantly changing by one method only met neither logFC nor p-value cutoffs in the alternative method. For analysis of WILD-seq 4T1 data, analysis was performed separately for the 2 experiments and genes had to meet significance cutoffs in both experiments.

Supplementary Table 12. Overlap of docetaxel resistance markers in 4T1 and D2A1 cell lines. 4T1 resistance genes were defined as those that were significantly enriched in resistant clone 679 but not in sensitive clone 238 ($p < 0.05$). D2A1 resistance genes were defined as those that were significantly enriched in combined resistant clones 1240, 751 and 1197 but not in sensitive clones 118, 2874 or 1072 ($p < 0.05$). Overlap of these lists revealed 47 common genes. These are listed along with their human orthologs.

Supplementary Table 13. Number and proportion of tumor cells assigned to each clonal barcode for docetaxel and L-asparaginase combination experiment.

Supplementary Table 14. Differential expression analysis for L-Asparaginase treatment. Differential gene expression analysis was performed by comparing cells from the same clonal lineage between each DTX+Asp sample and the combined DTX only samples. To ensure there were sufficient cells across all samples, five major clones (118, 1240, 2323, 2874 and 2991) were included

in the analysis. Differential expression analysis was performed using Seurat FindMarkers function and Wilcoxon Rank Sum test. Fisher's method was used to combine p-values from different clones within the same comparison. When selecting genes of interest, mean fold change between DTX+Asp samples and vehicle (also calculated on a per clone basis using abundant clones) was used as an additional cutoff and is included in the table. The most significantly and consistently differentially expressed genes are indicated in the final column 'Meets.cutoffs?'.