Genomic hallmarks and therapeutic implications of cancer cell quiescence

Anna J. Wiecek[1], Stephen J. Cutty[2], Daniel Kornai[1], Mario Parreno-Centeno[1], Lucie E. Gourmet[1], Guidantonio Malagoli Tagliazucchi[1], Daniel H. Jacobson[1,3], Ping Zhang[4], Lingyun Xiong[4], Gareth L. Bond[5], Alexis R. Barr[2,6], Maria Secrier[1,*]

[1] UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, UK

[2] Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK

[3] UCL Cancer Institute, Paul O'Gorman Building, University College London, London, UK

[4] Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

[5] Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, UK

[6] Cell Cycle Control Team, MRC London Institute of Medical Sciences (LMS), London, UK

* To whom correspondence should be addressed (m.secrier@ucl.ac.uk)

Running title: Hallmarks of cancer quiescence and therapeutic implications

Keywords: cancer cell quiescence, genomics, drug resistance, multi-omics integration, machine learning

Financial support:
AJW and DHJ were supported by MRC DTP grants (MR/N013867/1). MPC was supported by an Academy of Medical Science Springboard award (SBF004\1042). GMT was supported by a Wellcome Seed Award in Science (215296/Z/19/Z). MS was supported by a UKRI Future Leaders Fellowship (MR/T042184/1). Work in MS's lab was supported by a BBSRC equipment grant (BB/R01356X/1) and a Wellcome Institutional Strategic Support Fund (204841/Z/16/Z). ARB and SC are supported by a CRUK CDF (C63833/A25729) and work in ARB's lab is supported by MRC core-funding to the London Institute of Medical Sciences (MC-A658-5TY60).

Corresponding author:
Maria Secrier
UCL Genetics Institute
Department of Genetics, Evolution and Environment
University College London
Darwin Building, Gower Street
London WC1E 6BT
Email: m.secrier@ucl.ac.uk

Conflict of interest: None declared.

ABSTRACT

Therapy resistance in cancer is often driven by a subpopulation of cells that are temporarily arrested in a non-proliferative, quiescent state, which is difficult to capture and whose mutational drivers remain largely unknown. We developed methodology to uniquely identify this state from transcriptomic signals and characterised its prevalence and genomic constraints in solid primary tumours. We show quiescence preferentially emerges in the context of more stable, less mutated genomes which maintain *TP53* integrity and lack the hallmarks of DNA damage repair deficiency, while presenting increased APOBEC mutagenesis. We employ machine learning to uncover novel genomic dependencies of this process, and validate the role of the centrosomal gene *CEP89* as a modulator of proliferation/quiescence capacity. Lastly, we demonstrate that quiescence underlies unfavourable responses to various therapies exploiting cell cycle, kinase signalling and epigenetic mechanisms in single cell data, and propose a signature of quiescence-linked therapeutic resistance to further study and clinically track this state.

STATEMENT OF SIGNIFICANCE

We developed a robust transcriptomic signature of cellular quiescence, and employed it to systematically characterise proliferation/quiescence decisions in solid primary cancers and the genomic events influencing them. We propose CEP89 as a novel target whose suppression increases quiescence. Our expression signature of quiescence could be employed to track resistance to multiple anti-cancer compounds in a drug-tolerant persister cell setting.

INTRODUCTION

Tumour proliferation is one of the main hallmarks of cancer development[1], and has been extensively studied. While most of the cells within the tumour have a high proliferative capacity, occasionally under stress conditions some cells will become arrested temporarily in the G0 phase of the cell cycle, in a reversible state called 'quiescence' where they maintain minimal basal activity[2]. It has been proposed that this state enables cells to become resistant to anti-cancer compounds that target actively dividing cells, such as chemotherapy[3-5]. Moreover, a drug-tolerant 'persister' cell state represented by slow cycling or entirely quiescent cells[6-9] has been observed in a variety of pre-existing or acquired resistance scenarios, also in the context of targeted therapies[10,11]. As neoplastic cells evolve, quiescence can also be employed as a mechanism to facilitate immune evasion[12,13] or adaptation to new environmental niches during metastatic seeding[14,15]. In the context of disseminated tumour cells, quiescence can facilitate minimal residual disease, a major cause of relapse in the clinic[16].

Although quiescence is a widely conserved cellular state, essential for the normal development and homeostasis of eukaryotes[2,17], and has been extensively studied in a variety of organisms including bacteria and yeast[18,19], its role in cancer is still poorly defined. Hampering our understanding is the fact that it represents a number of heterogeneous states[17,20]. Canonically, cells can be forced into quiescence through serum starvation, mitogen withdrawal or contact inhibition[17]. Cells can also undergo cell cycle arrests spontaneously in response to cell-intrinsic factors like replication stress[21-23]. This process is controlled by p53[24], which triggers the inhibition of cyclin-CDK complexes by activating p21[22]. This in turn allows the assembly of the DREAM complex - a key effector responsible for repression of cell-cycle dependent gene expression[25]. Min and Spencer[26] recently demonstrated a much broader systemic coordination of 198 genes underlying distinct types of quiescence by profiling the transcriptomes of cells that entered this state either spontaneously or upon different stimuli. Additionally, proliferation-quiescence decisions can be impacted by oncogenic changes such as *MYC* amplification[27] or altered p38/ERK signalling[28].

Despite these advances, the identification of quiescent cells within tumours presents an ongoing challenge due to their scarcity and lack of universal, easily measurable markers for the activation and maintenance of this state. As they are often defined by a lack of proliferative markers[29,30], quiescent tumour cells can be mistaken for other therapy resistant cell types such as senescent or cancer stem cells[6,31]. Unlike quiescent cells, which can readily resume their proliferative state, senescent cells are irreversibly arrested[26] while cancer stem cells have a high capacity for self-renewal and sit at the top of the differentiation hierarchy[32]. Biomarkers of quiescence that are sufficiently specific and robust to be clinically useful are clearly needed.

Furthermore, our understanding of how cancer evolution is shaped by proliferation-quiescence decisions is limited. The proliferative heterogeneity of cancer cell populations has been previously described and linked with FAK/AKT1 signalling[33], but the constraints and consequences of these cell state switches have not been systematically profiled across cancer tissues. The extent to which tumour cell quiescence is enacted through transcriptional or genetic control is unknown[5,34], and neither are the mutational processes and genomic events modulating this state. Understanding the evolutionary triggers and molecular mechanisms that enable cancer cells to enter and maintain quiescence would enable us to develop pharmacological strategies to selectively eradicate quiescent cancer cells or prevent them from re-entering proliferative cycles.

To address these challenges, we have developed a new method to uniquely and reliably quantify quiescence in cancer using transcriptomic data, and employed it to characterise this phenomenon in bulk and single cell datasets from a variety of solid tumours. We describe the spectrum of proliferation/quiescence decisions in primary tumours, which reflects a range of stress adaptation mechanisms during the course of cancer development from early to advanced disease. We identify and validate mutational constraints for the emergence of quiescence, hinting at potential new therapeutic targets that could exploit this mechanism. We also demonstrate the relevance of quiescence to responses to a range of compounds targeting cell cycle, kinase signalling and epigenetic mechanisms in single cell datasets, and propose an expression signature that could be employed to detect treatment resistance induced by quiescent tumour cells.

RESULTS

Evaluating tumour cell quiescence from transcriptomic data

We hypothesised that primary tumours contain varying numbers of quiescent cells, which reflect evolutionary adaptations to cellular stress and may determine their ability to overcome antiproliferative therapies. To capture this elusive phenotype, we developed a computational framework that would allow us to quantify quiescence signals in bulk and single cell sequenced cancer samples (Figure 1a). To define a signature of quiescence, we focused on genes that have been shown by Min and Spencer[26] to be specifically activated or inactivated during quiescence that arises spontaneously or as a response to serum starvation, contact inhibition, MEK inhibition or CDK4/6 inhibition. The activity of 139 of these genes changed in a coordinated manner across all these five distinct forms of quiescence, likely representing generic transcriptional consequences of this cellular state. The expression levels of these markers were used to derive a score reflecting the relative abundance of quiescent cells within individual tumours (see Methods, Supplementary Table 1).

To validate this signature and select the optimal method to score quiescence in individual samples amongst different enrichment/rank-based scoring methodologies[35-38], we used seven single-cell and bulk datasets[10,39-44] where actively proliferating and quiescent cells had been independently isolated and sequenced (Supplementary Figure 1a, Supplementary Table 2, Methods). A combined Z-score approach had the highest accuracy in detecting signals of quiescence, with a 91% mean performance in classifying cells as quiescent or cycling (Figure 1b, Supplementary Figure 1b). Our signature reflected an expected increase in p27 protein levels, which are elevated in quiescence[45] (Supplementary Figure 1c). It also outperformed classical cell cycle and arrest markers, such as the expression of targets of the DREAM complex, CDK2, Ki67 and of mini-chromosome replication maintenance (MCM) protein complex genes - which are involved in the initiation of eukaryotic genome replication (Figure 1c). Importantly, our approach provided a good separation between quiescent and proliferating samples across a variety of cancer types and models including cancer cell lines, 3D organoid cultures, circulating tumour cells and patient-derived xenografts (Supplementary Table 2), thereby demonstrating its broad applicability. Furthermore, the strength of the score appeared to reflect the duration of arrest in quiescence[46] (Figure 1d).

We further experimentally validated our methodology in nine lung adenocarcinoma cell lines. We estimated the fraction of quiescent cells in each of these cell lines using quantitative, single-cell imaging of phospho-Ser807/811-Rb (PRb, which labels proliferative cells[47]) and 24 hour EdU proliferation assays (Figure 1e-h). Cells that were negative for either PRb or EdU were defined as quiescent (see Methods, Figure 1e-f). There was a remarkably good correlation between our predicted quiescence levels and the fraction of quiescent cells, as assessed by lack of EdU incorporation (which happens during S phase) but particularly by lack of Rb phosphorylation. Phosphorylation and inactivation of the retinoblastoma protein is often used to define the boundary between G0 and G1, and was specifically shown to distinguish the G0 state recently by Stallaert et al[47]. Furthermore, a G1 signature (Methods) was not associated with these experimental measurements, suggesting our method recovers a state more similar to G0 arrest rather than a prolonged G1 state (Figure 1g-h). The quiescence correlations appeared robust to random removal of individual genes from the signature, with no single gene having an inordinate impact on the score (Supplementary Figure 1d-f). This provided further reassurance that our Z-score based methodology is successful in capturing quiescence signals from bulk tumour data.

The spectrum of proliferation/quiescence capacity in solid primary tumours

Having established a robust framework for quantifying tumour cell quiescence, we next profiled 8,005 primary tumour samples across 31 solid cancer tissues from The Cancer Genome Atlas

(TCGA). After accounting for potential confounding signals of quiescent non-tumour cells from the microenvironment by correcting for tumour purity (see Methods, Supplementary Figure 1g-h), we observed an entire spectrum of fast proliferating to slowly cycling tumours, with the latter presenting stronger quiescence-linked signals (Figure 2a). While we acknowledge that no tumour would be entirely quiescent and we cannot identify individual quiescent cells within the tumour, this analysis does capture a broad range of phenotypes reflecting varying proliferation and cell cycle arrest rates, which suggests that quiescence is employed to different extents by tumours as an adaptive mechanism to various extrinsic and intrinsic stress factors. Cancers known to be frequently dormant, such as glioblastoma[3,39], were amongst the highest ranked in terms of quiescence, along with kidney and adrenocortical carcinomas (Figure 2b). This is likely explained by the innate proliferative capacity of the respective tissues. Indeed, tissues with lower stem cell division rates presented a greater propensity for quiescence (Figure 2c)[48].

Importantly, we confirmed that our quiescence scores capture a cellular state that is distinct from that of cancer stem cells, marked by signatures associated with high telomerase activity and an undifferentiated state[49,50], and that of senescent cells, marked by the Senescence-Associated Secretory Phenotype (SASP) and β-galactosidase activity[51-53] (Figure 2d, Supplementary Figures 2a-d). Our scores also showed strong negative correlations with the expression of proliferation markers (Figure 2d).

Lastly, we confirmed expected dependencies on the p53/p21/DREAM activation axis: tumours that were proficient in *TP53* or the components of the DREAM complex, as well as those with higher p21 expression, had elevated quiescence levels across numerous tissues (Figure 2e, Supplementary Figures 2e-f), although only 8 out of 139 quiescence genes are directly transcriptionally regulated by p53[54]. Nevertheless, p53 proficiency appears to be a non-obligatory dependency of quiescence, which is also observed to arise in p53 mutant scenarios in 21% of cases. p53 has also been shown to play a role in preventing the occurrence of larger structural events and polyploidy[55-57], potentially explaining the lower quiescence levels we observed in tumours that had undergone whole genome duplication (Figure 2f).

The genomic background of cancer cell quiescence

Cancer evolution is often driven by a variety of genomic events, ranging from single base substitutions to larger scale copy number variation and rearrangements of genomic segments. It is reasonable to expect that certain mutations accumulated by the cancer cells might enable a more proliferative phenotype, impairing the ability of cells to enter quiescence, or – on the contrary –

might favour quiescence as a temporary adaptive mechanism to extreme levels of stress. Having obtained quiescence estimates for primary tumour samples, we set out to identify potential genomic triggers or constraints that may shape proliferation-quiescence decisions in cancer. We identified 285 cancer driver genes that were preferentially altered (via mutations or copy number alterations) either in slow cycling or fast proliferating tumours (Figure 3a). Reassuringly, this list included genes previously implicated in driving cellular quiescence-proliferation decisions such as *TP53* and *MYC*[24,27]. An additional 10 cancer genes showed signals of positive selection in the context of high quiescence/proliferation (Supplementary Table 3). We also investigated associations with mutagenic footprints of carcinogens (termed "mutational signatures"), which can be identified as trinucleotide substitution patterns in the genome[58,59]. 15 mutational signatures were linked with quiescence either within individual cancer studies or pan-cancer (Supplementary Figure 2g).

Following the initial prioritisation of putative genomic constraints of quiescence, we employed machine learning to identify those events that could best distinguish slow cycling tumours with higher abundance of quiescent cells from fast proliferating ones, while accounting for tissue effects. An ensemble elastic net selection approach similar to the one described by Pich et al[60] was applied for this purpose (Figure 3b, Methods). Our pan-cancer model identified tissue type to be a major determinant of quiescence levels (Supplementary Figure 3a). It also uncovered a reduced set of 60 genomic events linked with proliferation/quiescence switches, including SNVs and copy number losses in 17 cancer genes, as well as amplifications of 13 cancer genes (Figure 3c). These events could then be successfully employed to predict quiescence in a separate test dataset, thus internally validating our model (Supplementary Figure 3b). Thus, while these events are not necessarily causative, the link is strong enough to be identifying quiescent states from genomic data alone. Such events may also pinpoint cellular vulnerabilities that could be exploited therapeutically.

Overall, the genomic dependencies of quiescence mainly comprised genes involved in cell cycle pathways, p53 regulation and ubiquitination (most likely of cell cycle targets), TGF-b signalling mediators and RUNX2/3 regulation which have previously been shown to play a role in controlling proliferation and cell cycle entry[61] (Supplementary Figure 3c). Invariably, this analysis has captured several events that are well known to promote cellular proliferation in cancer: this is expected and confirms the validity of our model. It was reassuring that a functional *TP53*, lack of *MYC* amplification and lower mutation rates (Figure 3c) were amongst the top ranked characteristics of highly quiescent tumours, which also displayed less aneuploidy. However, our analysis has also uncovered novel dependencies of quiescence-proliferation decisions that have not been reported previously, such as *CEP89* and *LMNA* amplifications observed in fast cycling tumours, or *ZMYM2*

deletions prevalent in samples with high levels of quiescence. ZMYM2 has recently been described as a novel binding partner of B-MYB and has been shown to be important in facilitating the G1/S cell cycle transition[62]. p16 (*CDKN2A*) deletions, one of the frequent early events during cancer evolution[63,64], were enriched in quiescent tumours. *RB1* deletions and amplifications were both associated with a reduction in tumour cell quiescence, which might reflect the dual role of RB1 in regulating proliferation and apoptosis[65].

Our model also calls to attention to the broader mutational processes associated with this cellular state. Such processes showed fairly weak and heterogeneous correlations with quiescence within individual cancer tissues (Supplementary Figure 2g), but their contribution becomes substantially clearer pan-cancer once other genomic sources are accounted for. In particular, we identified an association between quiescence and mutagenesis induced by the AID/APOBEC family of cytosine deaminases as denoted by signature SBS2[58] (Figure 3c). As highlighted by Mas-Ponte and Supek[66], APOBEC/AID driven mutations tend to be directed towards early-replicating, gene-rich regions of the genome, inducing deleterious events on several genes including *ZMYM2*, which our pan-cancer model has linked with higher quiescence.

In turn, defective DNA mismatch repair, as evidenced by signatures SBS44, SBS20, SBS15, SBS14 and SBS6[58], was prevalent in fast cycling tumours (Figure 3c). Mismatch repair deficiencies lead to hypermutation in a phenomenon termed "microsatellite instability" (MSI), which has been linked with increased immune evasion[67]. Cancers particularly prone to MSI include colon, stomach and endometrial carcinomas[68], where this state was indeed linked with lower quiescence (Figure 3d). Furthermore, quiescent tumours were depleted of alterations across all DNA damage repair pathways (Figure 3e).

Our measurements of quiescence also reflected expected cycling patterns across 27 MCF7 strains[69]: cell lines with longer doubling times exhibited higher quiescence (Figure 3f). This coincided with a depletion of *PTEN* mutations, a dependency highlighted by the pan-cancer model.

When checking for dependencies in individual cancer tissues, 27 out of the 28 genes identified by the model were significantly associated with quiescence/proliferation decisions in at least one tissue, most prominently in breast, lung and liver cancers which also represent the largest studies within TCGA (Figure 3g, top panel). Most of these genomic insults were linked with a decrease in quiescence. In external validation datasets these associations, including deletions in *PTEN* and *LRP1B* or amplifications of *MYC, CEP89* or *STAG1,* featured most prominently in the largest cohort of breast cancer samples (Figure 3g, bottom panel). These results highlight the fact that although a

pan-cancer approach is suited to capture genomic events that are universally associated with quiescence, certain genetic alterations may facilitate a higher or lower propensity of quiescence in a single tissue only.

Indeed, when building a tissue-specific breast cancer model of quiescence using a combined ANOVA and random forest classification approach (Supplementary Figure 4a), we not only recovered the associations with the *TP53, MYC*, *KLF6*, *LMNA*, *ETV6* and *RAD21* events already seen in the pan-cancer model (Supplementary Figure 4b), but identified additional events which validated in the METABRIC cohort and were also seen in several other cancers, e.g. bladder, lung and lower grade glioma (Supplementary Figure 4c). Notably, the APOBEC mutational signature SBS2 was the strongest genomic signal linked with quiescence in breast cancer (Supplementary Figure 4b,d) and was most prevalent in Her2+ tumours, although the Luminal A subtype showed the highest levels of quiescence overall, as expected given its well-known lower proliferative capacity[70] (Supplementary Figure 4e,f).

Validation of CEP89 as a modulator of proliferation/quiescence capacity

To gain more insight into the underlying biology of tumour cell quiescence, we sought to experimentally validate associations highlighted by the pan-cancer model. We focused on the impact of *CEP89* activity on quiescence-proliferation decisions due to the high ranking of this putative oncogene in the model, the relatively unexplored links between *CEP89* and cell cycle control, as well as its negative(?) association with quiescence across a variety of cancer cell lines (Supplementary Figure 5a-c). The function of *CEP89* is not well characterised, however, the encoded protein has been proposed to function as a centrosomal-associated protein[71,72]. Centrosomes function as major microtubule-organising centres in cells, playing a key role in mitotic spindle assembly[73] and the mitotic entry checkpoint[74]. Moreover, centrosomes act as sites of ubiquitin-mediated proteolysis of cell cycle targets[75], and members of several growth signalling pathways, such as Wnt and NF-kB, localise at these structures[76,77]. Several genetic interactions have also been reported between CEP89 and key cell cycle proteins, including cyclin D2[78] (Figure 4a).

Our model linked *CEP89* amplification with fast cycling tumours (Figure 3c). Centrosome amplification is a common feature of tumours with high proliferation rates and high genomic instability[79], and overexpression of centrosomal proteins can alter centriole structure[80,81]. Indeed, *CEP89* amplified tumours presented elevated expression of a previously reported centrosome amplification signature (CA20)[79] (Figure 4b), which was strongly anticorrelated with quiescence

levels (Figure 4c). Furthermore, *CEP89* expression was prognostic across multiple cancer tissues (Figure 4d) and linked with toxicity of several cancer compounds in cell line models (Supplementary Figure 5d).

We validated this target in the lung adenocarcinoma cell line NCI-H1299 showing high levels of *CEP89* amplification. Cep89 depletion via siRNA knockdown caused a consistent decrease in cell number, in the absence of any detectable cell death, and an increase in the fraction of quiescent cells as measured by PRb and EdU assays (Figure 4e-f). Thus, we propose *CEP89* as a novel cell proliferation regulator that may be exploited in certain scenarios to control tumour growth.

Characterisation of individual quiescence programmes

While we had previously examined a generic quiescence programme, cancer cells can enter this state due to different stimuli[17] and this may inform its aetiology and manifestation. To explore this, we re-scored tumours based on gene expression programmes specific to serum starvation, contact inhibition, MEK inhibition, CDK4/6 inhibition or spontaneously occurring quiescence as defined by Min and Spencer[26] (see Methods). We observed a good correlation between the estimates representing individual quiescence programmes and the expression of genes associated with the corresponding form of quiescence in the literature (Figure 5a-e, Supplementary Figure 6). CDK4/6 inhibition-induced quiescence levels were further validated using external RNA-seq datasets from cancer cell lines and xenograft mice sequenced before and after treatment with the CDK4/6 inhibitor Palbociclib[82,83] (Figure 5f, Supplementary Table 4). Interestingly, we also observed significant differences in spontaneous quiescence scores before and after treatment. Indeed, p21 activity has been linked with the Palbociclib mechanism of action[84,85], and this analysis suggests potential similarities between CDK4/6 inhibition and p21-dependent quiescence phenotypes.

Having validated our framework for quantifying stimulus-specific quiescence programmes, we proceeded to estimate the dominant form of quiescence in different cancer types (Figure 5g). We found a range of quiescence aetiologies across most tissues, while a minority of cancers were dominated by a single form of quiescence, e.g. serum starvation in all quiescent pheochromocytomas and paragangliomas, contact inhibition in 88% of head and neck carcinomas and CDK4/6 inhibition in 80% of quiescent adrenocortical carcinomas. While we do not wish to claim that the state of cell cycle arrest will have necessarily been induced by the actual predicted stimulus (impossible in the case of CDK4/6 or MEK inhibition, as the analysed samples are all treatment-naïve), we suggest that the downstream signalling cascade may resemble that triggered by such stimuli, e.g. via CDK4/6 or MEK loss of function mutations. Amongst these states, spontaneous quiescence appeared most

strongly dependent on p53 functionality (Supplementary Table 5). This points to common transcriptional features of drug-tolerant quiescent cells in different cancer settings that could be employed in designing ways to eradicate these cells in the future.

Role of quiescence in driving therapeutic resistance in cancer uncovered from single cell data

Overall, tumour cell quiescence appears to be beneficial for the long-term outcome of cancer patients (Figure 6a, Supplementary Figure 7a). Indeed, such slow cycling, indolent tumours would have higher chances of being eradicated earlier in the disease, which is consistent with reported worse prognosis of patients with higher tumour cell proliferation rates and less stable, more mutated genomes[86]. As expected, quiescence levels were increased in stage 1 tumours, although later stages also exhibited this phenotype occasionally (Supplementary Figure 7b). However, outcomes do vary depending on the quiescence subtype, with worse survival observed upon contact inhibition (Figure 6b). The outcomes also vary by tissue, with lung, colon or esophageal carcinoma patients displaying significantly worse prognosis in the context of high tumour cell quiescence (Figure 6c).

While quiescence may confer an overall survival advantage in most cancers, it can also provide a pool of cells that are capable of developing resistance to therapy[10,87]. Using our methodology, we indeed observed an increase in quiescence levels in cell lines following treatment with EGFR, BRAF and CDK4/6 inhibitors, as well as conventionally used chemotherapies such as 5-Fluorouracil (5-FU) in multiple bulk RNA-seq datasets (Figure 6d).

Furthermore, the recent widespread availability of single-cell transcriptomics offers the opportunity to investigate the impact of quiescence on such therapies with much greater granularity than is allowed by bulk data. Using our quiescence signature and single-cell data from RKO and SW480 colon cancer cell lines treated with 5-FU[88], we could observe quiescence-proliferation decisions following conventional chemotherapy treatment. Within the p53 proficient cell line RKO, the fraction of quiescent cells increased from 41% to 93% after treatment with a low dose (10 μM) of 5-FU and persisted at higher doses (Figure 6e-f). In contrast, a comparable increase was not observed in *TP53* mutant SW480 cells, further emphasizing the key role of p53 as a regulator of quiescence (Figure 6g-h). This implies that although *TP53* mutations confer a more aggressive tumour phenotype, *TP53* wild-type tumour cells are more likely to be capable of entering a quiescent "persistent" state associated with drug-resistance. SW480 cells showed higher apoptotic activity following treatment compared to RKO cells, particularly within actively cycling cells, further corroborating that cells capable of entering quiescence may be less vulnerable to this therapy (Supplementary Figure 7c-d).

Similarly, using single cell data from an *EGFR* mutant Non-Small Cell Lung Cancer (NSCLC) cell line treated with the EGFR inhibitor Erlotinib[89], we predicted that 40% of cells were likely to exist in a quiescent state prior to treatment. EGFR inhibition led to a massive decrease in cell numbers immediately after treatment, mostly due to proliferating cells dying off (Supplementary Figures 7e-f), while the proportion of quiescent cells increased to 96% at day 1, indicating an immediate selective advantage for such cells (Figure 6i-j). Interestingly, these cells appear to gradually start proliferating again in the following days during continuous treatment, with the percentage of proliferating cells approaching pre-treatment levels by day 11 (Figure 6j). The same trend captured by our signature could be observed upon KRAS and BRAF inhibition in different cell line models (Supplementary Figure 7g-j, Supplementary Table 4)[10,89]. Furthermore, during the first days of treatment the NSCLC cells that survived EGFR inhibition appeared to reside in a state most resembling that induced by serum starvation (Figures 6k-l). Both EGFR kinase inhibitors and serum starvation have been shown to trigger autophagy[90], which may explain the convergence between this inhibitory trigger and the type of quiescence response. At day 11 most of the remaining quiescent cells appeared in a state similar to that preceding the treatment (Figure 6l).

Thus, quiescence appears to explain resistance to broad acting chemotherapy agents as well as targeted molecular inhibitors of the Ras/MAPK signalling pathway, being either selected for, or induced immediately upon treatment, and gradually waning over time as cells start re-entering the cell cycle. Using massively multiplexed chemical transcriptomic data, we also analysed responses to 188 small molecule inhibitors in cell lines at single-cell resolution[91] (Supplementary Figure 8). We observed a large increase in quiescence following treatment with compounds targeting cell cycle regulation and tyrosine kinase signalling, consistent with our previous results, but also for compounds modulating epigenetic regulation, e.g. histone deacetylase inhibitors – thus highlighting the broad relevance of quiescence.

Tumour cell quiescence signature for use in single-cell transcriptomics data

Our ability to probe the nature of quiescent phenotypes in scRNA-seq data using a defined quiescence signature could aid the development of methods to selectively target quiescent drug-resistant persister cells. However, a major challenge of scRNA-seq data analysis is the high percentage of gene dropout, which could impact our ability to evaluate quiescence using the full 139 gene signature. The scRNA-seq datasets we analysed exhibited an average drop-out of 8.5 genes out of the full gene signature. While our scoring method remains robust to such levels of dropout (Supplementary Figure 1c-e), we also employed machine learning to reduce our initial list of 139 markers of quiescence to a robust 35-gene signature, comprised mainly of RNA metabolism and

splicing-regulating factors, but also of genes involved in cell cycle progression, ageing and senescence, which could be applied to sparser datasets with larger levels of gene dropout (Methods, Figure 7a-b, Supplementary Table 6). The optimised signature of quiescence performed similarly to the initial broadly defined programme in distinguishing highly quiescent and fast cycling tumours (Figure 7c), showed an average dropout of only 0.5 genes across the scRNA-seq datasets used in this study (Figure 7d), was similarly prognostic (p=0.004) and showed comparable profiles of resistance to treatment (Figure 7e, Supplementary Figure 9). This minimal expression signature could be employed to track and further study emerging quiescence-induced resistance in a variety of therapeutic scenarios.

DISCUSSION

Despite its crucial role in cancer progression and resistance to therapies, tumour cell quiescence has remained poorly characterised due to the scarcity of suitable models and biomarkers for large-scale tracking in the tissue or blood. The lack of proliferative markers such as Ki67 or CDK2[29,92] does not uniquely distinguish quiescence from other cell cycle phases, e.g. G1 or S. Furthermore, these and other reliable markers of G0 arrest such as p27 or p130[45] are best captured at protein level, which is much more sparsely measured, and expression does not accurately reflect their activity. This study overcame this limitation by employing genes active in different forms of quiescence whose patterns of expression are distinct from markers of a longer G1 phase, senescence or stemness. We have extensively validated our method and signature in single cell datasets and cancer cell lines, and have demonstrated that it can reliably and robustly capture signals of quiescence both in bulk tissue as well as in single cells. The versatility of this signature is evidenced by high classification accuracies across a variety of solid cancer datasets. More variable performance was observed when applied to hematopoietic stem cells as it was not designed to capture signals in this context (Supplementary Figure 1b). While we cannot exclude that the patterns captured may also occasionally reflect cell cycle arrest in G1 or G2, this broad signature would still capture phenotypes resulting from intrinsic or extrinsic cellular stress that reflect temporary tumour adaptation during the course of cancer evolution or upon treatment with drugs. Thus, studying such states is relevant for identifying vulnerabilities that could be exploited at different time points during the course of cancer treatment.

We show that quiescence is pervasive across different solid cancers and generally associated with more stable, less mutated genomes with intact DNA damage repair pathways. We also find a link between APOBEC mutagenesis and higher levels of quiescence. We identify mutational events affecting a variety of genes such as *PTEN*, *CEP89, CYLD*, *LMNA* that appear unfavourable to cell cycle arrest, thus potentially implicating them in influencing quiescence-proliferation decisions.

Among these, we propose and validate *CEP89* as a novel modulator of quiescence capacity in non-small cell lung cancer. As such, CEP89 could be targeted to induce and actively maintain quiescence in a scenario where this is favourable, possibly in combination with other cancer therapies. Neoplastic events enriched in quiescent tumours, such as p16 or *ZMYM2* deletions, could mark elevated genomic stress that renders cells more prone to cell cycle arrest. Such targets should be further validated and could be exploited to either counteract quiescence or induce it.

These large-scale genomic associations with quiescence are only currently feasible in bulk datasets. However, bulk sequenced data has a major limitation in capturing an average signal across all cells within the tumour, which prevents individual cell state identification and counting. Our subsequent exploration of single cell datasets across 193 therapeutic scenarios complements this analysis and illustrates the power of applying our signature in single cells.

Our signature of quiescence is prognostic and marks primary tumours with a lower proliferative capacity before treatment, but we also clearly demonstrate that it can be employed to track resistance to multiple cell cycle, kinase signalling and epigenetic targeting regimens, where it often appears as a short-lived phenotype. In this setting, vulnerabilities of quiescent cells could be exploited for combination therapies. Quiescent cells utilise several mechanisms to achieve drug resistance, including upregulation of stress-induced pathways such as anti-apoptotic BCL-2 signalling[93], anti-ROS programmes[26] or immune evasion[13]. Further studies are needed to elucidate which of these mechanisms are specifically employed on a case-by-case basis. Moreover, a key open question remains: does quiescence drive resistance in a Darwinian fashion, as a pre-existing population that is selected for upon drug treatment, or is it instead an acquired phenotype? Our single cell analyses cannot exclude either scenario, but the optimised signature we propose for single cell data makes it tractable to a variety of further future studies in this area.

Our findings contribute to the understanding of the aetiology and genetic context of quiescence in cancer. This is particularly relevant to identifying new anti-proliferative targets, but also for the detection and eradication of drug tolerant persister cells, which have been frequently, although not always, observed to be driven by slow cycling or entirely quiescent[7,8]. Importantly, the state of cancer quiescence that we have studied here is distinct from that of disseminated tumour cells causing clinical dormancy and cancer relapse, often after many years from the treatment of the primary tumour[6,94]. Here, we have focused on understanding how tumours make proliferation and quiescence decisions during the earlier stages of cancer development, within the treatment-naïve primary tumour and as an immediate response to anti-cancer therapies. However, since the dormancy of disseminated tumour cells is fundamentally enabled through a long but temporary cell cycle arrest, we believe our

findings of the fundamental processes linked with quiescence could in the future help inform a better characterisation of dormant tumour cells when combined with specific microenvironmental signatures that are critical for enabling that process.

Overall, our study provides, for the first time, a pan-cancer view of cellular quiescence and its evolutionary constraints, underlying novel mutational dependencies which could be exploited in the clinic. We propose a quiescence signature which can be robustly measured in bulk tissue or single cells, and could inform therapeutic strategies or risk of relapse. This signature could be assessed in the clinic to track rapidly emerging resistance, e.g. through liquid biopsies or targeted gene panels. We hope these insights can be used as building blocks for future studies into the different regulators of quiescence, including epigenetics and microenvironmental interactions, as well as the mechanisms by which it enables therapeutic resistance both in solid and haematological malignancies.

MATERIALS AND METHODS

Selection of quiescence marker genes

Generic quiescence markers:

Differential expression analysis results comparing cycling immortalised, non-transformed human epithelial cells and cells in five different forms of quiescence (spontaneous quiescence, contact inhibition, serum starvation, CDK4/6 inhibition and MEK inhibition) were obtained from Min and Spencer[26]. A total of 195 genes were differentially expressed in all five forms of quiescence under an adjusted p-value cut-off of 0.05. This gene list, reflective of a generic quiescence phenotype, was subjected to the following refinement and filtering steps: (1) selection of genes with a unidirectional change of expression across all five forms of quiescence; (2) removal of genes involved in other cell cycle stages included in the "KEGG_CELL_CYCLE" gene list deposited at MSigDB; (3) removal of genes showing low standard deviation and low levels of expression within the TCGA dataset, or which showed low correlation with the pan-cancer expression of the transcriptional targets of the DREAM complex, the main effector of quiescence, in TCGA. The resulting 139-gene signature is presented in Supplementary Table 1.

Quiescence subtype-specific markers:

Gene lists representing spontaneous quiescence, contact inhibition, serum starvation, CDK4/6 inhibition and MEK inhibition programmes were obtained using genes differentially expressed in

each individual quiescence form using an adjusted p-value cutoff of 0.05. The gene lists were subjected to filtering steps 2 and 3 described above. Following the refinement steps, 10 upregulated and 10 downregulated genes with highest log2 fold changes were selected for each quiescence type.

Quantification of tumour cell quiescence

The *GSVA* R package was used to implement the combined z-score[35], ssGSEA[36] and GSVA[37] gene set enrichment methods. For the above three methods a separate score was obtained for genes upregulated in quiescence and genes downregulated in quiescence, following which a final quiescence score was obtained by subtracting the two scores. The singscore single-sample gene signature scoring method[38] was implemented using the *singscore* R package. In addition to these, we also calculated a mean scaled quiescence score based on the refined list of genes upregulated and downregulated in quiescence, as well as a curated housekeeping genes from the "HSIAO_HOUSEKEEPING_GENES" list deposited at MSigDB, as follows:

$$QS = \frac{\frac{1}{n}\sum G_U - \frac{1}{n}\sum G_D}{\frac{1}{n}\sum G_H}$$

QS = mean scale quiescence score

G_U = expression of genes upregulated in quiescence

G_D = expression of genes downregulated in quiescence

G_H = expression of housekeeping genes

n = number of genes in each gene set

Quiescence scores for the TCGA cohort were derived from expression data scaled by tumour purity estimates. The pan-cancer TCGA samples were also classified into "high" or "low" quiescence groups based on k-means clustering (k=2) on the expression data of 139 quiescence biomarker genes, following the removal of tissue-specific expression differences using the *ComBat* function from the *sva* R package[95].

Measuring the duration of quiescence

We employed the GSE124109 dataset from Fujimaki et al[46] where rat embryonic fibroblasts were transcriptomically profiled as they moved from short to long-term quiescence in the absence of

466 growth signals. The derived quiescence scores using our combined z-score methodology increased
467 from short to longer-term quiescence.

468 **Validation of quiescence scoring methodologies**

469 *Single-cell RNA-sequencing validation datasets:*

470 Datasets were obtained from the ArrayExpress and Gene Expression Omnibus (GEO) databases
471 though the following GEO Series accession numbers: GSE83142, GSE75367, GSE137912,
472 GSE139013, GSE90742 and E-MTAB-4547. Quality control analysis was standardised using the
473 *SingleCellExperiment*[96] and *scater*[97] R packages. Normalisation was performed using the *scran*[98] R
474 package.

475 *Bulk RNA-sequencing validation datasets:*

476 Datasets were obtained from the GEO database through the following GEO Series accession
477 numbers: GSE93391, GSE114012, GSE131594, GSE152699, GSE124854, GSE135215, GSE99116
478 and GSE124109. GSE114012 count data were normalised to TPM values using the *GeoTcgaData* R
479 package. All normalised datasets were log-transformed before further analysis.

480 The accuracy with which the quiescence scoring methods could separate proliferating and quiescent
481 samples within the validation datasets was determined by calculating the area under the curve of the
482 receiver operating characteristic (ROC) curves, using the *plotROC* R package.

483 **Experimental validation in lung adenocarcinoma cell lines**

484 The average fraction of cancer cells spontaneously entering quiescence was estimated for nine lung
485 adenocarcinoma cell lines (NCIH460, A549, NCIH1666, NCIH1944, NCIH1563, NCIH1299,
486 NCIH1650, H358, L23) using EdU and phospho-Rb staining proliferation assays.

487 Cell lines were obtained from ATCC or Sigma and regularly checked for mycoplasma. A549 and
488 NCIH460 were cultured in DMEM (Gibco). NCIH358, NCIH1299 and NCIH1563 were maintained
489 in RPMI-1640 (Gibco) supplemented with 5mM sodium pyruvate and 0.5% glucose. NCIH1944,
490 NCIH1666, NCIH1650 and L23 were grown in RPMI-1640 ATCC formulation (Gibco). A427 were
491 cultured in EMEM (ATCC). A549, NCIH460, H358, NCIH1299, NCIH1563, A427 were
492 supplemented with 10% heat inactivated FBS. NCIH1666 with 5% heat-inactivated FBS and all
493 other cell lines with 10% non-heat inactivated FBS. All cell lines had penicillin-streptomycin (Gibco)
494 added to 1%. Cells were maintained at 37°C and 5% CO_2. To calculate the quiescent fraction, A549

and NCIH460 cells were plated at a density of 500 cells/well, and all other cell lines at a density of 1000/well, in 384well CellCarrier Ultra plates (PerkinElmer) in the relevant media. 24h later, 5 μM EdU was added and cells were incubated for a further 24h before fixing in a final concentration of 4% formaldehyde (15 min, RT), permeabilization with PBS/0.5% Triton X-100 (15 min, RT) and blocking with 2% BSA in PBS (60 min, RT). The EdU signal was detected using Click-iT chemistry, according to the manufacturer's protocol (ThermoFisher). Cells were also labelled for phospho-Ser807/811 Rb (PRb) using Rabbit mAb 8516 (CST) at 1:2000 in blocking solution, overnight at 4°C. Unbound primary antibody was washed three times in PBS and secondary Alexa-conjugated antibodies were used to detect the signal (ThermoFisher, 1:1000, 1h at RT). Finally nuclei were labelled with Hoechst 33258 (1 μg/ml, 15 min RT) before imaging on a high-content widefield Operetta microscope, 20x N.A. 0.8. Automated image analysis (Harmony, PerkinElmer) was used to segment and quantify nuclear signals in imaged cells. Quiescent cells were defined by the absence of EdU or PRb staining, determined by quantification of their nuclear expression (Figure 1e-f).

The quiescence scores for cancer cell lines were calculated using corresponding log-transformed RPKM normalised bulk RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE) database[99].

CEP89 was depleted by ON-Target siRNA Pool from Horizon. NCI-H1299 cells were reverse transfected in 384 well plates with 20nM of Non-targeting control (NTC) or CEP89-targeting siRNA using Lipofectamine RNAiMax (ThermoFisher), according to the manufacturer's instructions. Cells were left for 24h, before 5 μM EdU was added for the final 24h and then cells were processed as above to determine the quiescent fraction. To determine the level of Cep89 depletion by western blot, cells were reverse transfected with siRNA in 24 well plates. 48h after transfection, cells were lysed directly in 1x SDS sample buffer with 1mM DTT (ThermoFisher). Samples were separated on pre-cast 4-20% Tris-Glycine gels, transferred to PVDF using the iBlot2 system and membranes blocked in blocking buffer (5% milk in TBS) for 1h at RT. The membrane was then cut and the upper half was incubated in 1:1000 Cep89 antibody (Sigma, HPA040056), the bottom half in B-actin antibody 1:2000 (CST; 3700S) diluted in blocking buffer overnight at 4'C. Membranes were washed three times in TBS-0.05% TritonX-100 before being incubated in secondary anti-rabbit (Cep89) or anti-mouse (B-actin) HRP conjugated antibodies (CST 7074P2 and CST 7076P2, respectively) diluted 1:2000 in blocking buffer for 1h at RT. Membranes were washed three times again and signal detected using Clarity ECL solution (BioRad) and scanned on an Amersham ImageQuant 800 analyser.

Multi-omics discovery cohort

FPKM normalised RNA-sequencing expression data, copy number variation gene-level data, RPPA levels for p27 as well as mutation annotation files aligned against the GRCh38 human reference genome from the Mutect2 pipeline were downloaded using the *TCGABiolinks* R package [100] for 9,712 TCGA primary tumour samples across 31 solid cancer types. Haematological malignancies were excluded as the quiescence markers were derived in epithelial cells and might not be equally suited to capture this phenotype in blood. For patients with multiple samples available, one RNA-seq barcode entry was selected for each individual patient resulting in 9,631 total entries. All expression data were log-transformed for downstream analysis. During quiescence score calculation, expression data for the primary tumour samples was scaled according to tumour purity estimates reported by Hoadley et al[101] to account for potential confounding quiescence signals coming from non-tumour cells in the microenvironment. Samples with purity estimates lower than 30% were removed, leaving 8,005 samples for downstream analysis.

The mutation rates of all TCGA primary tumour samples were determined by log-transforming the total number of mutations in each sample divided by the length of the exome capture (38Mb).

TP53 functional status was assessed based on somatic mutation and copy number alterations as described in Zhang et al[102]. *TP53* mutation and copy number for the TCGA tumours were downloaded from cBioPortal (http://www.cbioportal.org). Tumours with *TP53* oncogenic mutations (annotated by OncoKB) and copy-number alterations (GISTIC score $\leq$ -1) were assigned as *TP53* mutant and CNV loss. Tumours without these *TP53* alterations were assigned as TP53 wild type. The effects of the *TP53* mutation status on quiescence score were then determined with a linear model approach with the quiescence score as a dependent variable and mutational status as an independent variable. The *P* values were FDR-adjusted.

APOBEC mutagenesis enriched samples were determined through pan-cancer clustering of mutational signature contributions as described in Wiecek et al[103]. The APOBEC mutagenesis cluster was defined as the cluster with highest mean SBS2 and SBS13 contribution. This was repeated 100 times and only samples which appeared in the APOBEC cluster at least 50 times were counted as being APOBEC enriched.

Aneuploidy scores and whole genome duplication events across TCGA samples were obtained from Taylor et al[104]. Microsatellite instability status for uterine corpus endometrial carcinoma, as well as stomach and colon adenocarcinoma samples were obtained from Cortes-Ciriano et al[68]. Telomerase

558 enzymatic activity "EXTEND" scores were obtained from Noureen et al[49]. Expression-based cancer
559 cell stemness indices were obtained from Malta et al[50]. Centrosome amplification transcriptomic
560 signature (CA20) scores were obtained from Almeida et al[79].

561 **PHATE dimensionality reduction**

562 The *phateR* R package[105] was used to perform the dimensionality reduction with a constant seed for
563 reproducibility. The *ComBat* function from the *sva* R package[106] was used to remove tissue-specific
564 expression patterns from the TCGA RNA-seq data.

565 **Cancer stem cell division estimates**

566 The mean stem cell division estimates for different cancer types used in this study were obtained
567 from Tomasetti and Vogelstein[48].

568 **Positive selection analysis**

569 Genes positively selected specifically in samples clustered into low or high quiescence groups were
570 identified based on dN/dS analysis using the *dNdScv* R package[107], run with default parameters.
571 Genes showing signals of positive selection in either the highly quiescent or fast proliferating
572 samples which encoded olfactory receptors were discarded from downstream analysis.

573 **Mutational signature estimation**

574 Mutational signature contributions were inferred as described in Wiecek et al[103].

575 **Machine learning of quiescence-linked features via ensemble elastic net regression models**

576 The COSMIC database was used to source a list of 723 known drivers of tumorigenesis (Tiers 1+2).
577 285 oncogenes and tumour suppressors from a curated list showed a significant enrichment or
578 depletion of mutations or copy number variants in quiescent samples either pan-cancer or within
579 individual TCGA studies.

580 To classify highly quiescent from fast proliferating tumours, the 286 genes were used as input
581 features for an ensemble elastic net regression model along the tumour mutational rate, whole-
582 genome doubling estimates, ploidy, aneuploidy scores, 10 positively selected genes and 15
583 mutational signatures, which showed a significant correlation with quiescence scores either pan-

cancer or within individual TCGA studies. The *caret* R package was used to build an elastic net regression model 1000 times on the training dataset of 3,753 TCGA primary tumour samples (80% of the total dataset). Only samples with at least 50 mutations were used in the model, for which mutational signatures could be reliably estimated. For each of the 1000 iterations, we randomly selected 90% of the samples from the training dataset to build the model. Only features which were included in all 1000 model iterations were selected for further analysis. To test the performance of our approach, a linear regression model was built using the reduced list of genomic features and their corresponding coefficients averaged across the 1000 elastic net regression model iterations. When applying the resulting linear regression model on the internal validation dataset of 936 samples, we found a strong correlation between the observed and predicted quiescence scores ($R = 0.73$, $p < 2.2e-16$).

SHAP values for the linear regression model used to predict quiescence scores were obtained using the *fastshap* R package.

Gene enrichment and network analysis

Gene set enrichment analysis was carried out using the *ReactomePA* R package, as well as GeneMania[108] and ConsensusPathDB[109]. Interactions between CEP89 and other cell cycle components were inferred using the list of cell cycle genes provided by cBioPortal and GeneMania to reconstruct the expanded network with direct interactors (*STAG1*, *CCND2*, *STAT3*). Networks were visualised using Cytoscape[110].

Gene lists

Genes associated with the G1 phase of the cell cycle were obtained from the curated "REACTOME_G1_PHASE " list deposited at MSigDB.

Genes associated with apoptosis were obtained from the curated "HALLMARK_APOPTOSIS" list deposited at MSigDB.

Genes associated with the senescence-associated secretory phenotype were obtained from Basisty et al[53]. Lists of genes making up the various DNA damage repair pathways were derived from Pearl et al[111].

Genes associated with contact inhibition were obtained from the curated "contact inhibition" gene ontology term. Genes associated with serum starvation were obtained from the curated

"REACTOME_CELLULAR_RESPONSE_TO_STARVATION" list deposited at MSigDB. MEK inhibition was assessed based on the activity of the MAPK pathway as determined using an expression signature (MPAS) consisting of 10 downstream MAPK transcripts[112].

Validation of the genomic features of quiescence

For elastic net model feature validation, RNA-seq data was downloaded for six cancer studies from cBioPortal[113], along with patient-matched whole-genome, whole-exome and targeted sequencing data. The 6 datasets used comprise breast cancer (SMC[114] and METABRIC[115]), paediatric Wilms' tumor (TARGET[116]), bladder cancer, prostate adenocarcinoma and sarcoma (MSKCC[117-119]) studies. The data were processed and analysed in the same manner as the TCGA data. RNA-seq data for 27 MCF7 cell line strains, alongside cell line growth rates and targeted mutational sequencing data were obtained from Ben-David et al[69].

Genomic dependency modelling in breast cancer

An ANOVA-based feature importance classification was used and identified 30 genomic features most discriminative of samples with lower and higher than average quiescence scores. A random forest model was then built using the identified features and correctly classified samples according to their quiescence state with a mean accuracy of 74% across five randomly sampled test datasets from the cohort.

Survival analysis

Multivariate Cox Proportional Hazards analysis was carried out using the *coxph* function from the *survival* R package. The optimal quiescence score cut-off value of 2.95 was determined using the *surv_cutpoint* function.

Treatment response scRNA-seq and bulk RNA-seq data

Datasets have been obtained from the GEO database through the following GEO Series accession numbers: GSE134836, GSE134838, GSE134839, GSE137912, GSE149224, GSE124854, GSE135215, GSE99116, GSE152699, GSE178839 and GSE139944. The *umap* R package was used for dimensionality reduction with constant seed for reproducibility.

Quiescence subtype determination

TCGA cohort studies:

Samples with evidence of quiescence characterised by a generic quiescence score > 0 were further subclassified based on the most likely form of quiescence exhibited, among CDK4/6 inhibition, contact inhibition, MEK inhibition, spontaneous quiescence or serum starvation, using subtype-specific expression signatures. We opted for a conservative approach and classed each quiescent sample into a specific quiescence subtype if the quiescence score for the corresponding programme was higher than one standard deviation of the distribution across the TCGA cohort, and if the score was significantly higher than for the remaining programmes when assessed using a Student's t test. Samples which could not be classified into any of the five quiescence states characterised in this study were classified as "uncertain".

Single-cell RNA seq treatment response datasets:

The quiescence subtype of individual single cells was inferred by mapping such individual cells onto the reference dataset of MCF10A cells reflecting different forms of quiescence obtained from Min and Spencer[26]. The *ComBat* R package was used to remove the study batch effect between the expression data to be classified and the reference bulk RNA-seq data. PCA dimensionality reduction analysis was then used on the combined datasets using the *prcomp* R function. For each patient sample or single-cell expression data entry, a k-nearest neighbour algorithm classification was performed using the *knn* function from the *class* R package. During the classification the three nearest reference bulk RNA-seq data points were considered, with two nearest neighbours with identical class needed for classification.

Optimisation of the quiescence signature

We investigated if a subset of the 139 quiescence-related genes could act as a more reliable marker of quiescence that would bypass dropout issues in single cell data. This was performed in three steps:

(1) Assessment of individual importance as quiescence marker for a given gene

We collected three high confidence single cell expression datasets separating quiescent from proliferating cells. A random forest model was trained on each dataset separately to predict the quiescence state of a given cell based on the expression levels of the 139 genes. The Gini indices corresponding to each gene in the model were normalised to a range of values between 0 and 1, which would reflect how important an individual gene was for determining quiescence state

relative to the other 138 genes. The procedure was repeated 1000 times for each of the three datasets, and the average Gini coefficients across iterations were stored.

(2) Prioritisation of gene subsets based on cumulative importance in the model

Genes were placed in the candidate subset if their importance metric was above a given threshold in at least one of the datasets. By gradually increasing the threshold from 0 to 1, different gene combinations were produced.

(3) External validation of candidate subsets

The gene combinations in (2) were tested for their ability to predict quiescence. For this, a separate validation dataset was utilized, which contained gene expression levels for the 139 genes in the 10 lung cancer cell lines previously employed for experimental validation, along with the quiescence state of the lines as inferred by PRb and EdU staining. For each gene subset, a combined z-score of quiescence was calculated from the expression levels as described previously. The correlations between this z-score and the two experimental measurements of quiescence were used to establish the ability of a gene combination to predict quiescence. Among the top performing subsets, a 35 gene signature with a mean correlation of 78% between predicted and measured quiescence levels in the test data ($p=0.016$) showed the highest correlation with PRb measurements capturing short-lived quiescence, the more common state observed in single cell treatment datasets. Therefore, this signature was deemed to achieve the best trade-off between gene numbers and signal capture.

The optimised gene signature is provided in Supplementary Table 6.

Statistical analysis

Groups were compared using a two-sided Student's t test, Wilcoxon rank-sum test or ANOVA, as appropriate. P-values were adjusted for multiple testing where appropriate using the Benjamini-Hochberg method. Graphs were generated using the ggplot2 and ggpubr R packages.

Data availability

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga, METABRIC (https://ega-archive.org/studies/EGAS00000000083), MSK-IMPACT (https://www.mskcc.org/msk-impact), or deposited at cBioPortal (https://www.cbioportal.org/) and GEO (https://www.ncbi.nlm.nih.gov/geo/). All data comply with

699 ethical regulations, with approval and informed consent for collection and sharing already obtained
700 by the relevant consortia.

701 **Code availability**

702 All code developed for the purpose of this study can be found at the following repository:
703 https://github.com/secrierlab/CancerCellQuiescence

704 **ACKNOWLEDGEMENTS**

705 We would like to thank Prof Chris Barnes for the very helpful discussions and input on the findings
706 of the study.

707 **AUTHOR CONTRIBUTIONS**

708 MS designed the study and supervised the computational analyses. ARB designed and supervised
709 the experimental validation in cell lines. GB supervised the analysis of p53 functional association.
710 AJW developed the quiescence scoring methodology and performed all computational analyses to
711 validate and apply it in bulk and single cell datasets, as well as link it to genomic features. SC
712 performed the experimental validation of quiescence prevalence and CEP89 association with
713 quiescence in cell lines. DK performed the inference of the minimal signature of quiescence
714 applicable in single cell data. MPC performed the random forest modelling and feature selection in
715 breast cancer. LG performed the positive selection analysis. GMT wrote the code for batch effect
716 correction and for PCA mapping of single cell data on a reference dataset. DHJ performed the
717 APOBEC enrichment classification. PZ and LX performed the quiescence comparison of p53 wild
718 type and mutated cancers. MS, AJW, ARB and SC wrote the manuscript, with contributions from all
719 other authors. All authors read and approved the manuscript.

720 **REFERENCES**

721 1. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
722 2. van Dijk, D. *et al.* Slow-growing cells within isogenic populations have increased RNA
723 polymerase error rates and DNA damage. *Nat Commun* **6**, 7972 (2015).
724 3. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after
725 chemotherapy. *Nature* **488**, 522-6 (2012).
726 4. Puig, I. *et al.* TET2 controls chemoresistant slow-cycling cancer cell survival and tumor
727 recurrence. *J Clin Invest* **128**, 3887-3905 (2018).

5. Rehman, S.K. *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell* **184**, 226-242.e21 (2021).
6. Phan, T.G. & Croucher, P.I. The dormant cancer cell life cycle. *Nat Rev Cancer* **20**, 398-411 (2020).
7. Cabanos, H.F. & Hata, A.N. Emerging Insights into Targeted Therapy-Tolerant Persister Cells in Cancer. *Cancers (Basel)* **13**(2021).
8. Sharma, S.V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69-80 (2010).
9. Swayden, M., Chhouri, H., Anouar, Y. & Grumolato, L. Tolerant/Persister Cancer Cells and the Path to Resistance to Targeted Therapy. *Cells* **9**(2020).
10. Xue, J.Y. *et al.* Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature* **577**, 421-425 (2020).
11. Aissa, A.F. *et al.* Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
12. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165**, 45-60 (2016).
13. Ribas, A. Adaptive Immune Resistance: How Cancer Protects from Immune Attack. *Cancer Discov* **5**, 915-9 (2015).
14. Sosa, M.S., Bragado, P. & Aguirre-Ghiso, J.A. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nat Rev Cancer* **14**, 611-22 (2014).
15. Barkan, D. *et al.* Metastatic growth from dormant cells induced by a col-I-enriched fibrotic environment. *Cancer Res* **70**, 5706-16 (2010).
16. Masago, K., Fujita, S. & Yatabe, Y. Targeting minimal residual disease after surgery with molecular targeted therapy: the real path to a cure? *J Thorac Dis* **10**, S1982-S1985 (2018).
17. Coller, H.A., Sang, L. & Roberts, J.M. A new description of cellular quiescence. *PLoS Biol* **4**, e83 (2006).
18. Rittershaus, E.S., Baek, S.H. & Sassetti, C.M. The normalcy of dormancy: common themes in microbial quiescence. *Cell Host Microbe* **13**, 643-51 (2013).
19. Miles, S., Bradley, G.T. & Breeden, L.L. The budding yeast transition to quiescence. *Yeast* **38**, 30-38 (2021).
20. Marescal, O. & Cheeseman, I.M. Cellular Mechanisms and Regulation of Quiescence. *Dev Cell* **55**, 259-271 (2020).

760 21. Arora, M., Moser, J., Phadke, H., Basha, A.A. & Spencer, S.L. Endogenous Replication Stress
761 in Mother Cells Leads to Quiescence of Daughter Cells. *Cell Rep* **19**, 1351-1364 (2017).
762 22. Barr, A.R. *et al.* DNA damage during S-phase mediates the proliferation-quiescence decision
763 in the subsequent G1 via p21 expression. *Nat Commun* **8**, 14728 (2017).
764 23. Heldt, F.S., Barr, A.R., Cooper, S., Bakal, C. & Novák, B. A comprehensive model for the
765 proliferation-quiescence decision in response to endogenous DNA damage in human cells.
766 *Proc Natl Acad Sci U S A* **115**, 2532-2537 (2018).
767 24. Itahana, K. *et al.* A role for p53 in maintaining and establishing the quiescence growth arrest
768 in human cells. *J Biol Chem* **277**, 18206-14 (2002).
769 25. Sadasivam, S. & DeCaprio, J.A. The DREAM complex: master coordinator of cell cycle-
770 dependent gene expression. *Nature Reviews Cancer* **13**, 585-595 (2013).
771 26. Min, M. & Spencer, S.L. Spontaneously slow-cycling subpopulations of human cells originate
772 from activation of stress-response pathways. *PLoS Biol* **17**, e3000178 (2019).
773 27. García-Gutiérrez, L., Delgado, M.D. & León, J. MYC Oncogene Contributions to Release of
774 Cell Cycle Brakes. *Genes (Basel)* **10**(2019).
775 28. Aguirre-Ghiso, J.A., Estrada, Y., Liu, D. & Ossowski, L. ERK(MAPK) activity as a determinant
776 of tumor growth and dormancy; regulation by p38(SAPK). *Cancer Res* **63**, 1684-95 (2003).
777 29. Spencer, S.L. *et al.* The proliferation-quiescence decision is controlled by a bifurcation in
778 CDK2 activity at mitotic exit. *Cell* **155**, 369-83 (2013).
779 30. Miller, I. *et al.* Ki67 is a Graded Rather than a Binary Marker of Proliferation versus
780 Quiescence. *Cell Rep* **24**, 1105-1112.e5 (2018).
781 31. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem
782 cells. *Nature* **414**, 105-11 (2001).
783 32. Kleinsmith, L.J. & Pierce, G.B., Jr. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA
784 CELLS. *Cancer Res* **24**, 1544-51 (1964).
785 33. Dey-Guha, I. *et al.* A mechanism for asymmetric cell division resulting in proliferative
786 asynchronicity. *Mol Cancer Res* **13**, 223-30 (2015).
787 34. Turati, V.A. *et al.* Chemotherapy induces canalization of cell state in childhood B-cell
788 precursor acute lymphoblastic leukemia. *Nature Cancer* (2021).
789 35. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T. & Lee, D. Inferring pathway activity toward precise
790 disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).

36. Barbie, D.A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-12 (2009).
37. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
38. Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018).
39. Aulestia, F.J. *et al.* Quiescence status of glioblastoma stem-like cells involves remodelling of Ca(2+) signalling and mitochondrial shape. *Sci Rep* **8**, 9731 (2018).
40. Kurppa, K.J. *et al.* Treatment-Induced Tumor Dormancy through YAP-Mediated Transcriptional Reprogramming of the Apoptotic Pathway. *Cancer Cell* **37**, 104-122.e12 (2020).
41. Ebinger, S. *et al.* Characterization of Rare, Dormant, and Therapy-Resistant Cells in Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 849-862 (2016).
42. Jordan, N.V. *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102-106 (2016).
43. Buczacki, S.J.A. *et al.* Itraconazole targets cell cycle heterogeneity in colorectal cancer. *J Exp Med* **215**, 1891-1912 (2018).
44. Grigore, F. *et al.* BRAF inhibition in melanoma is associated with the dysregulation of histone methylation and histone methyltransferases. *Neoplasia* **22**, 376-389 (2020).
45. Vairo, G. *et al.* Bcl-2 retards cell cycle entry through p27(Kip1), pRB relative p130, and altered E2F regulation. *Mol Cell Biol* **20**, 4745-53 (2000).
46. Fujimaki, K. *et al.* Graded regulation of cellular quiescence depth between proliferation and senescence by a lysosomal dimmer switch. *Proc Natl Acad Sci U S A* **116**, 22624-22634 (2019).
47. Stallaert, W. *et al.* The structure of the human cell cycle. *Cell Syst* **13**, 103 (2022).
48. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78-81 (2015).
49. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an association with cancer stemness and proliferation. *Nature Communications* **12**, 139 (2021).
50. Malta, T.M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354.e15 (2018).

51. Dimri, G.P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences* **92**, 9363 (1995).
52. Debacq-Chainiaux, F., Erusalimsky, J.D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of senescent cells in culture and in vivo. *Nat Protoc* **4**, 1798-806 (2009).
53. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLOS Biology* **18**, e3000599 (2020).
54. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956 (2017).
55. Andreassen, P.R., Lohez, O.D., Lacroix, F.B. & Margolis, R.L. Tetraploid state induces p53-dependent arrest of nontransformed mammalian cells in G1. *Molecular biology of the cell* **12**, 1315-1328 (2001).
56. Di Leonardo, A. *et al.* DNA rereplication in the presence of mitotic spindle inhibitors in human and mouse fibroblasts lacking either p53 or pRb function. *Cancer Res* **57**, 1013-9 (1997).
57. Ganem, N.J. *et al.* Cytokinesis failure triggers hippo tumor suppressor pathway activation. *Cell* **158**, 833-848 (2014).
58. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
59. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
60. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat Genet* **51**, 1732-1740 (2019).
61. Miyazono, K., Maeda, S. & Imamura, T. Coordinate regulation of cell growth and differentiation by TGF-beta superfamily and Runx proteins. *Oncogene* **23**, 4232-7 (2004).
62. Cibis, H., Biyanee, A., Dörner, W., Mootz, H.D. & Klempnauer, K.H. Characterization of the zinc finger proteins ZMYM2 and ZMYM4 as novel B-MYB binding proteins. *Sci Rep* **10**, 8390 (2020).
63. Maley, C.C. & Reid, B.J. Natural selection in neoplastic progression of Barrett's esophagus. *Semin Cancer Biol* **15**, 474-83 (2005).
64. Bardeesy, N. & DePinho, R.A. Pancreatic cancer biology and genetics. *Nat Rev Cancer* **2**, 897-909 (2002).

65. Indovina, P., Pentimalli, F., Casini, N., Vocca, I. & Giordano, A. RB1 dual role in proliferation and apoptosis: cell fate control and implications for cancer therapy. *Oncotarget* **6**(2015).
66. Mas-Ponte, D. & Supek, F. DNA mismatch repair promotes APOBEC3-mediated diffuse hypermutation in human cancers. *Nat Genet* **52**, 958-968 (2020).
67. Kloor, M. & von Knebel Doeberitz, M. The Immune Biology of Microsatellite-Unstable Cancer. *Trends Cancer* **2**, 121-133 (2016).
68. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P.J. A molecular portrait of microsatellite instability across multiple cancers. *Nature Communications* **8**, 15180 (2017).
69. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325-330 (2018).
70. Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol* **5**, 412-24 (2014).
71. Jakobsen, L. *et al.* Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *Embo j* **30**, 1520-35 (2011).
72. Sillibourne, J.E. *et al.* Assessing the localization of centrosomal proteins by PALM/STORM nanoscopy. *Cytoskeleton (Hoboken)* **68**, 619-27 (2011).
73. Bettencourt-Dias, M. & Glover, D.M. Centrosome biogenesis and function: centrosomics brings new understanding. *Nature Reviews Molecular Cell Biology* **8**, 451-463 (2007).
74. Grallert, A. *et al.* Removal of centrosomal PP1 by NIMA kinase unlocks the MPF feedback loop to promote mitotic commitment in S. pombe. *Curr Biol* **23**, 213-22 (2013).
75. Wigley, W.C. *et al.* Dynamic association of proteasomal machinery with the centrosome. *The Journal of cell biology* **145**, 481-490 (1999).
76. Itoh, K., Jenny, A., Mlodzik, M. & Sokol, S.Y. Centrosomal localization of Diversin and its relevance to Wnt signaling. *J Cell Sci* **122**, 3791-8 (2009).
77. Kfoury, Y. *et al.* Ubiquitylated Tax targets and binds the IKK signalosome at the centrosome. *Oncogene* **27**, 1665-1676 (2008).
78. Lin, A., Wang, R.T., Ahn, S., Park, C.C. & Smith, D.J. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res* **20**, 1122-32 (2010).
79. de Almeida, B.P., Vieira, A.F., Paredes, J., Bettencourt-Dias, M. & Barbosa-Morais, N.L. Pan-cancer association of a centrosome amplification gene expression signature with genomic alterations and clinical outcome. *PLOS Computational Biology* **15**, e1006832 (2019).

884 80. Kohlmaier, G. *et al.* Overly long centrioles and defective cell division upon excess of the SAS-
885 4-related protein CPAP. *Curr Biol* **19**, 1012-8 (2009).
886 81. Tang, C.-J.C., Fu, R.-H., Wu, K.-S., Hsu, W.-B. & Tang, T.K. CPAP is a cell-cycle regulated
887 protein that controls centriole length. *Nature Cell Biology* **11**, 825-831 (2009).
888 82. Hafner, M. *et al.* Multiomics Profiling Establishes the Polypharmacology of FDA-Approved
889 CDK4/6 Inhibitors and the Potential for Differential Clinical Activity. *Cell Chem Biol* **26**, 1067-
890 1080.e8 (2019).
891 83. Salvador-Barbero, B. *et al.* CDK4/6 Inhibitors Impair Recovery from Cytotoxic Chemotherapy
892 in Pancreatic Adenocarcinoma. *Cancer Cell* **37**, 340-353.e6 (2020).
893 84. Guiley, K.Z. *et al.* p27 allosterically activates cyclin-dependent kinase 4 and antagonizes
894 palbociclib inhibition. *Science* **366**(2019).
895 85. Pack, L.R., Daigh, L.H., Chung, M. & Meyer, T. Clinical CDK4/6 inhibitors induce selective and
896 immediate dissociation of p21 from cyclin D-CDK4 to inhibit CDK2. *Nat Commun* **12**, 3356
897 (2021).
898 86. Zhu, X. *et al.* The prognostic and predictive potential of Ki-67 in triple-negative breast
899 cancer. *Scientific Reports* **10**, 225 (2020).
900 87. Brown, J.A. *et al.* TGF-β-Induced Quiescence Mediates Chemoresistance of Tumor-
901 Propagating Cells in Squamous Cell Carcinoma. *Cell Stem Cell* **21**, 650-664.e8 (2017).
902 88. Park, S.R. *et al.* Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-
903 Fluorouracil-Induced DNA Damage. *Cell Rep* **32**, 108077 (2020).
904 89. Aissa, A.F. *et al.* Single-cell transcriptional changes associated with drug tolerance and
905 response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
906 90. Tan, X., Thapa, N., Sun, Y. & Anderson, R.A. A kinase-independent role for EGF receptor in
907 autophagy initiation. *Cell* **160**, 145-60 (2015).
908 91. Srivatsan, S.R. *et al.* Massively multiplex chemical transcriptomics at single-cell resolution.
909 *Science* **367**, 45-51 (2020).
910 92. Moser, J., Miller, I., Carter, D. & Spencer, S.L. Control of the Restriction Point by Rb and p21.
911 *Proc Natl Acad Sci U S A* **115**, E8219-e8227 (2018).
912 93. Minassian, L.M., Cotechini, T., Huitema, E. & Graham, C.H. Hypoxia-Induced Resistance to
913 Chemotherapy in Cancer. *Adv Exp Med Biol* **1136**, 123-139 (2019).
914 94. Aguirre-Ghiso, J.A. & Sosa, M.S. Emerging Topics on Disseminated Cancer Cell Dormancy
915 and the Paradigm of Metastasis. *Annual Review of Cancer Biology* **2**, 377-393 (2018).

95. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).

96. Amezquita, R.A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nature Methods* **17**, 137-145 (2020).

97. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017).

98. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).

99. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-7 (2012).

100. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71 (2016).

101. Hoadley, K.A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).

102. Zhang, P. *et al.* Germline and Somatic Genetic Variants in the p53 Pathway Interact to Affect Cancer Risk, Progression, and Drug Response. *Cancer Res* **81**, 1667-1680 (2021).

103. Wiecek, A.J., Jacobson, D.H., Lason, W. & Secrier, M. Pan-Cancer Survey of Tumor Mass Dormancy and Underlying Mutational Processes. *Frontiers in Cell and Developmental Biology* **9**, 1820 (2021).

104. Taylor, A.M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).

105. Moon, K.R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology* **37**, 1482-1492 (2019).

106. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-3 (2012).

107. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e21 (2017).

108. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-20 (2010).

947 109. Kamburov, A. *et al.* ConsensusPathDB: toward a more complete picture of cell biology.
948 *Nucleic Acids Res* **39**, D712-7 (2011).
949 110. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
950 interaction networks. *Genome Res* **13**, 2498-504 (2003).
951 111. Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B. & Pearl, F.M.G. Therapeutic opportunities
952 within the DNA damage response. *Nature Reviews Cancer* **15**, 166-180 (2015).
953 112. Wagle, M.C. *et al.* A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically
954 relevant biomarker in multiple cancer types. *NPJ Precis Oncol* **2**, 7 (2018).
955 113. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring
956 multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
957 114. Kan, Z. *et al.* Multi-omics profiling of younger Asian breast cancers reveals distinctive
958 molecular signatures. *Nat Commun* **9**, 1725 (2018).
959 115. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals
960 novel subgroups. *Nature* **486**, 346-52 (2012).
961 116. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic
962 landscape of Wilms tumor. *Nat Genet* **49**, 1487-1494 (2017).
963 117. Iyer, G. *et al.* Prevalence and co-occurrence of actionable genomic alterations in high-grade
964 bladder cancer. *J Clin Oncol* **31**, 3133-40 (2013).
965 118. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**,
966 11-22 (2010).
967 119. Barretina, J. *et al.* Subtype-specific genomic alterations define new targets for soft-tissue
968 sarcoma therapy. *Nat Genet* **42**, 715-21 (2010).
969

FIGURE LEGENDS

Figure 1: Methodology for quantifying cancer cell quiescence. (a) Workflow for evaluating quiescence from RNA-seq data. 139 genes differentially expressed in multiple forms of quiescence were employed to score quiescence across cancer tissues. **(b)** Receiver operating characteristic (ROC) curves illustrating the performance of the z-score methodology on separating actively proliferating and quiescent cells in seven single-cell (continuous curves) and bulk RNA-seq (dotted curves) datasets. AUC = area under the curve. **(c)** Compared classification accuracies of the z-score approach and classic cell proliferation markers across the seven single-cell/bulk RNA-seq validation datasets. **(d)** Quiescence levels of embryonic fibroblast cells under serum starvation for various amounts of time. Replicates are depicted in the same colour. **(e)** Representative images of lung cancer cell lines immunostained and analysed to detect the quiescent fraction. Hoechst (labels all nuclei) is in blue, PRb in green and EdU in red in merged image. White dashed circles highlight quiescent cells that are negative for both PRb and EdU signals. Scale bar: 100 μm. **(f)** Graphs show single cell quantification of PRb and EdU intensities taken from images and used to define the cut-off to calculate the quiescent fraction (QF; green boxes). Images in **(e)** and graphs in **(f)** are taken from the A549 cell line. **(g-h)** Correlation between theoretical estimates of a quiescence or G1 state and the fraction of cells entering quiescence in nine lung adenocarcinoma cell lines, as assessed through **(g)** phospho-Rb assays and **(h)** EdU assays. Mean of n=3 is shown for the average percentage of quiescent cells.

Figure 2: Pan-cancer evaluation of proliferative heterogeneity and linked tumour hallmarks. **(a)** PHATE plot illustrating the wide spectrum of proliferative to slow cycling/quiescent states across 8,005 primary tumour samples from TCGA. Each sample is coloured according to the relative quiescence level. **(b)** Variation in tumour quiescence levels across different cancer tissues. **(c)** Correlation between mean quiescence scores and stem cell division estimates for various tissue types. **(d)** Correlating tumour quiescence scores with cancer cell stemness (Stemness Index), telomerase activity (EXTEND score), p21 activity (CDKN1A) and the expression of several commonly used proliferation markers. The Pearson correlation coefficient is displayed. RC – replication complex. **(e)** Consistently higher levels of quiescence are detected in samples with functional p53. **(f)** Lower quiescence scores are observed in tumours with one or two whole genome duplication events. Wilcoxon rank-sum test p-values are displayed in boxplots, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

Figure 3: Genomic landscape of proliferation/quiescence decisions in cancer. (a) Cancer drivers with mutations or copy number alterations depleted pan-cancer in a quiescence context. Features further selected by the pan-cancer model are highlighted. **(b)** Schematic of the ensemble elastic net modelling employed to prioritise genomic changes associated with tumour quiescence. **(c)** Genomic events significantly associated with quiescence, ranked according to their importance in the model (highest to lowest). Each point depicts an individual tumour sample, coloured by the value of the respective feature. For discrete variables purple indicates the presence of the feature and green its absence. The Shapley values indicate the impact of individual feature values on the quiescence score prediction. **(d)** Quiescence levels are significantly reduced in microsatellite unstable (MSI) samples in stomach adenocarcinoma (STAD) and uterine corpus endometrial carcinoma (UCEC), with the same trend (albeit not significant) shown in colon adenocarcinoma (COAD). Wilcoxon rank-sum test *$p<0.05$; **$p<0.01$. **(e)** Genomic alterations are depleted across DNA repair pathways during quiescence. Odds ratios of mutational load on pathway in quiescence are depicted, along with confidence intervals. CS=chromosome segregation; p53=p53 pathway; UR=ubiquitylation response; CPF=checkpoint factors; TM=telomere maintenance; CR=chromatin remodelling; TLS=translesion synthesis; NHEJ=non-homologous end joining; NER=nucleotide excision repair; MMR=mismatch repair; FA=Fanconi Anaemia; BER=base excision repair. **(f)** Quiescence is increased in cell lines with slow doubling time across MCF7 strains, which also show lower prevalence of PTEN mutations. **(g)** Tissue-specific changes in quiescence between samples with/without quiescence-associated deletions (blue), amplifications (red) and SNVs (brown) within the TCGA cohort (top) and external validation datasets (bottom).

Figure 4: CEP89 amplification is associated with lower tumour quiescence. (a) Network illustrating CEP89 interactions with cell cycle genes (from GeneMania). The edge colour indicates the interaction type, with green representing genetic interactions, orange representing predicted interactions and purple indicating pathway interactions. The edge width illustrates the interaction weight. **(b)** CA20 scores are significantly increased in TCGA primary tumours containing a CEP89 amplification. **(c)** Pan-cancer relationship between CA20 scores and tumour quiescence across the TCGA cohort. **(d)** Cox proportional hazards analysis estimates of the log hazards ratio for the impact of CEP89 expression on patient prognosis within individual cancer studies, after adjusting for tumour stage. Patients with high expression of CEP89 show significantly worse prognosis within ACC, LUSC, LIHC, KIRC and STAD, but significantly better prognosis within HNSC, PAAD and KIRP studies. **(e)** Western blot showing depletion of Cep89 protein 48h after siRNA transfection of NCI-H1299 cells. Mock is lipofectamine only, NTC is Non-targeting control siRNA. B-actin is used as a

loading control. **(f)** Graphs show that Cep89 depletion in NCI-H1299 cells leads to a reduction in nuclear number and an increase in the fraction of quiescent cells, measured by an increase in the percentage of EdU negative (24h EdU pulse) and Phospho-Ser 807/811 Rb negative cells. One-Way ANOVA, *p<0.05, **p<0.01. Mean of n=3.

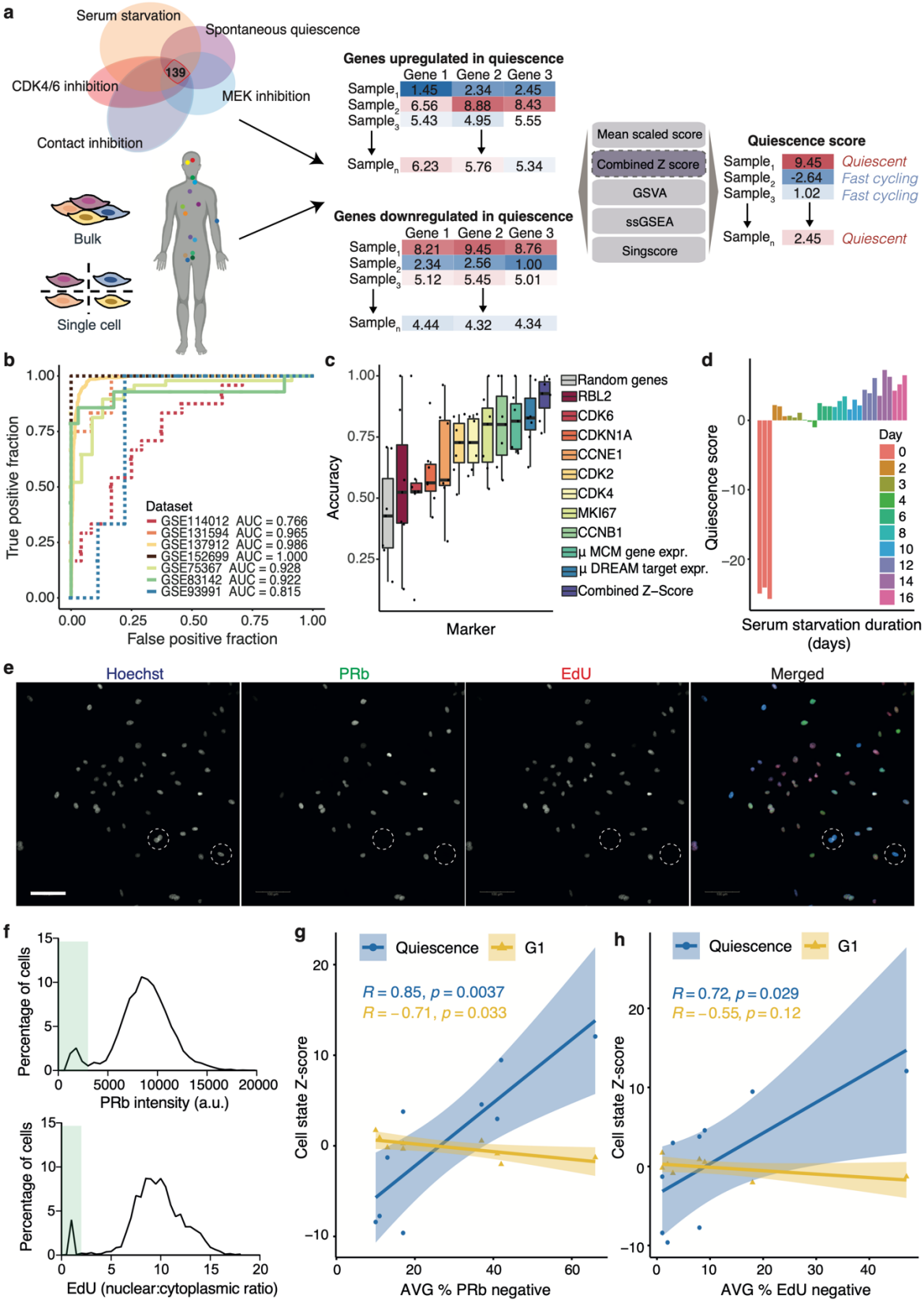
Figure 5: Pan-cancer characterisation of individual quiescence programmes. (a-e) Comparison of correlation coefficients between quiescence programme scores and **(a)** mean expression of CDK4 and CDK6, **(b)** mean expression of curated contact inhibition genes, **(c)** a transcriptional MAPK Pathway Activity Score (MPAS), **(d)** mean expression of curated serum starvation genes, and **(e)** CDKN1A expression, across TCGA cancers. The correlations expected to be strongest (either negative or positive) are denoted by an asterisk. The generic quiescence score refers to scores calculated using the original list of 139 genes differentially expressed across all 5 forms of quiescence. **(f)** Comparison of quiescence programme scores measured in cancer cell lines before (grey) and after (red) Palbociclib treatment across three validation studies. Datasets used for validation are denoted by their corresponding GEO series accession number. **(g)** Predicted quiescence type composition of samples estimated to be highly quiescent across individual cancer types. The same colour legend as in **(a)** is applied. Gray bars represent the proportion of samples for which the quiescence type could not be estimated.

Figure 6: Impact of tumour cell quiescence on patient prognosis and treatment response. (a) Disease-specific survival based on tumour quiescence for patients from TCGA within 15 years of follow up. Patients with high levels of quiescence in primary tumours showed significantly better prognosis than patients with low quiescence. **(b-c)** Hazard ratio ranges illustrating the impact of different forms of quiescence **(b)** and different tissues **(c)** on patient prognosis, after taking into account potential confounding factors. Values above 0 indicate significantly better prognosis in the context of high tumour quiescence. **(d)** Change in quiescence scores inferred from bulk RNA-seq across breast, pancreatic, colorectal and skin cancer cells in response to treatment with the CDK4/6 inhibitor Palbociclib, 5-FU or the BRAF inhibitor Vemurafenib. **(e-f)** UMAP plot illustrating the response of the TP53-proficient RKO colorectal cancer cell line to various 5-FU doses and the corresponding proportions of cells predicted to be quiescent/proliferating. Each dot is an individual cell, coloured according to its quiescence level. **(g-h)** The same as previous, but for the TP53-deficient SW480 cell line. **(i-j)** UMAP plot illustrating the response of individual PC9 NSCLC cells to the EGFR inhibitor Erlotinib across several time points and the corresponding proportion of cells predicted to be quiescent/proliferating. **(k)** Principal component analysis illustrating the

superimposition of scRNA-seq profiles (circles) of quiescent NSCLC cells before/after EGFR inhibition onto the bulk RNA-seq reference data (triangles) for MCF10A cells occupying various quiescence states. **(l)** The proportion of NSCLC cells in **(k)** predicted to occupy different quiescence states across several time points.

Figure 7: Optimisation of quiescence signature for use in scRNA-seq data. (a) Methodology for refining the gene signature of quiescence: random forest classifiers are trained to distinguish quiescent from cycling tumours on three high confidence datasets; Gini index thresholding is optimised to prioritise a final list of 35 genes. **(b)** Gini index variation, correlation with experimentally measured quiescence via EdU and PRb staining assays, and corresponding p-values are plotted as the number of genes considered in the model is increased. The red dotted line indicates the threshold chosen for the final solution of 35 genes. The black dotted line indicates the threshold for p-value significance. **(c)** Additional external validation of the 35 gene signature acting as a classifier of quiescent and proliferation cells in single cell and bulk datasets. **(d)** Dropout in single cell data by gene signature. The percentage of genes out of the 35 (red) and 139 (grey) gene lists with reported expression across the single-cell RNA-seq datasets analysed in this study. **(e)** Proportion of cycling and quiescent cells estimated in single cell datasets of p53 wild type and mutant lines treated with 5FU, as well as cells treated with EGFR inhibitors. Data as in Figure 7.

1093 **FIGURES**



1094

1095 **Figure 1: Methodology for quantifying cancer cell quiescence**

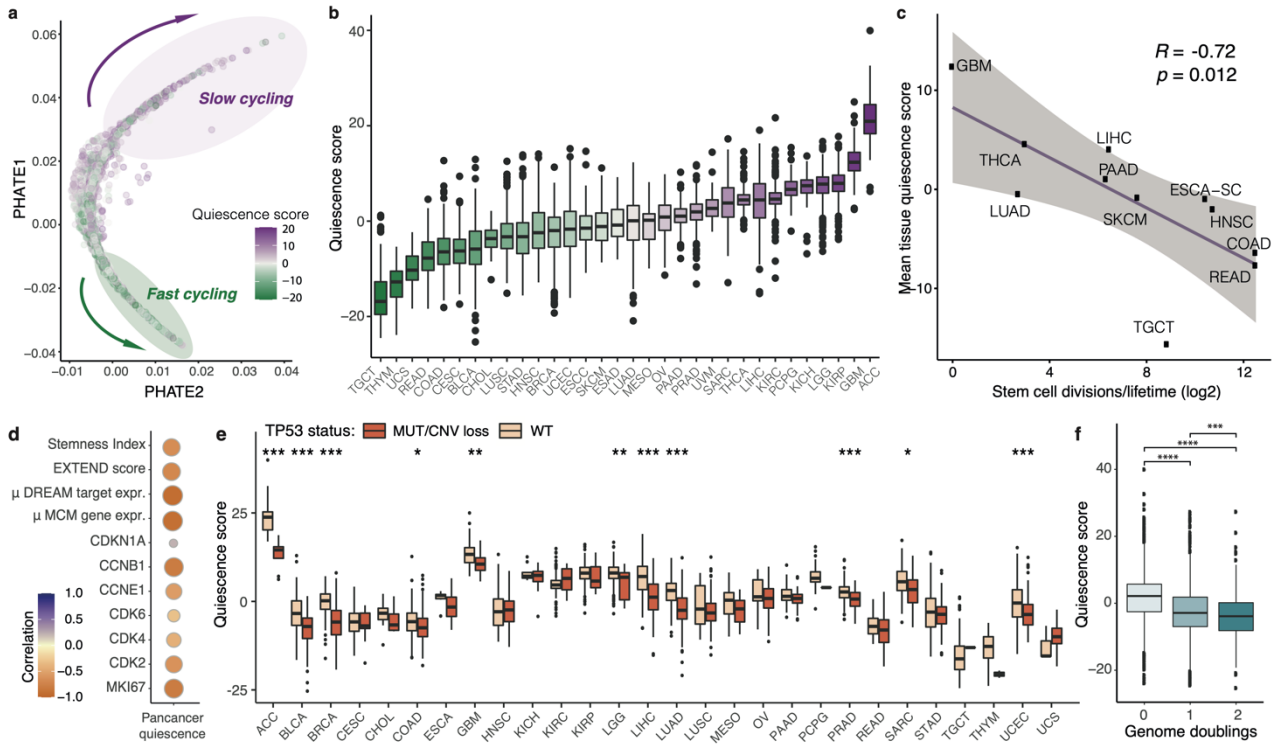


Figure 2: Pan-cancer evaluation of proliferative heterogeneity and linked tumour hallmarks

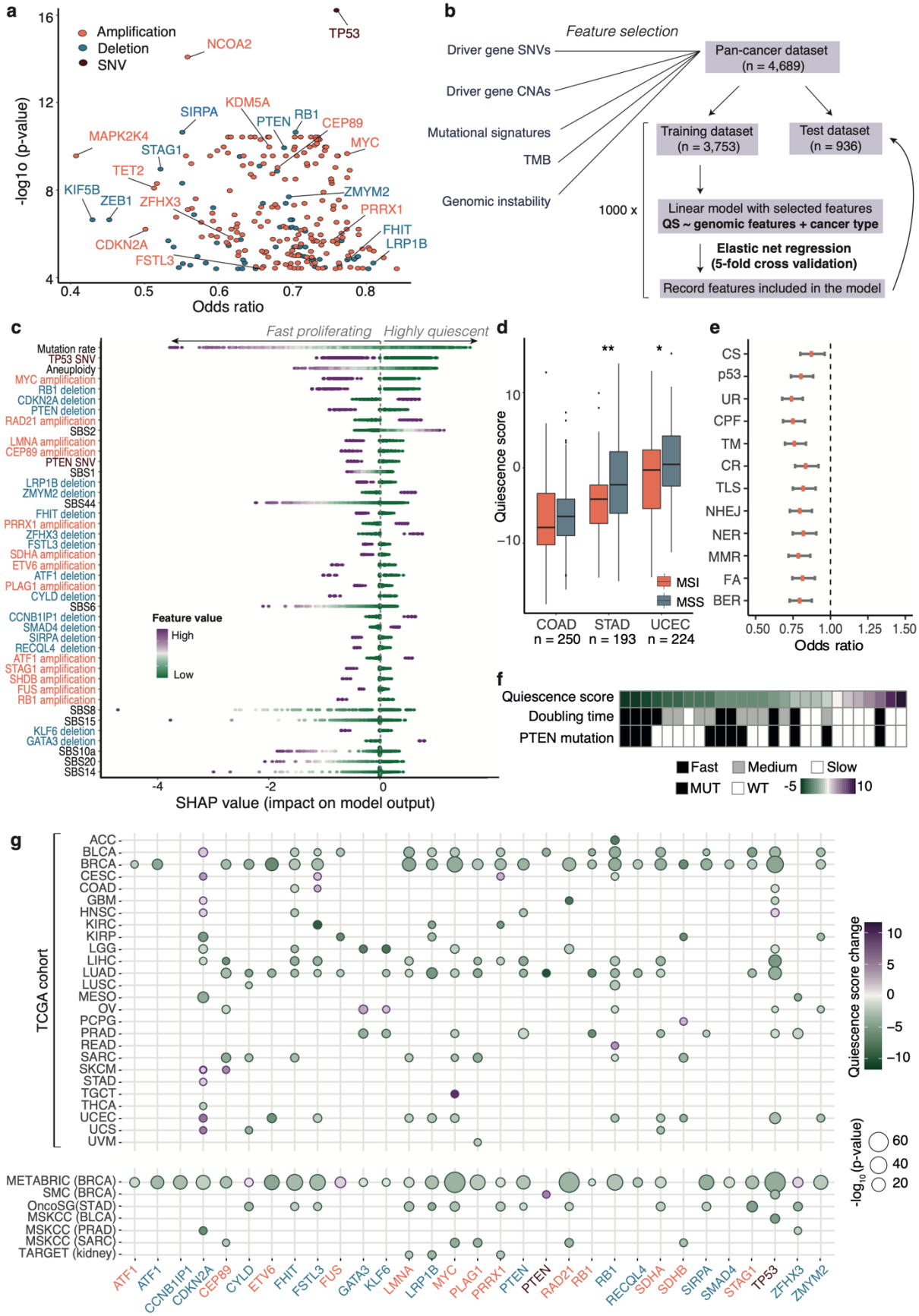


Figure 3: Genomic landscape of proliferation/quiescence decisions in cancer

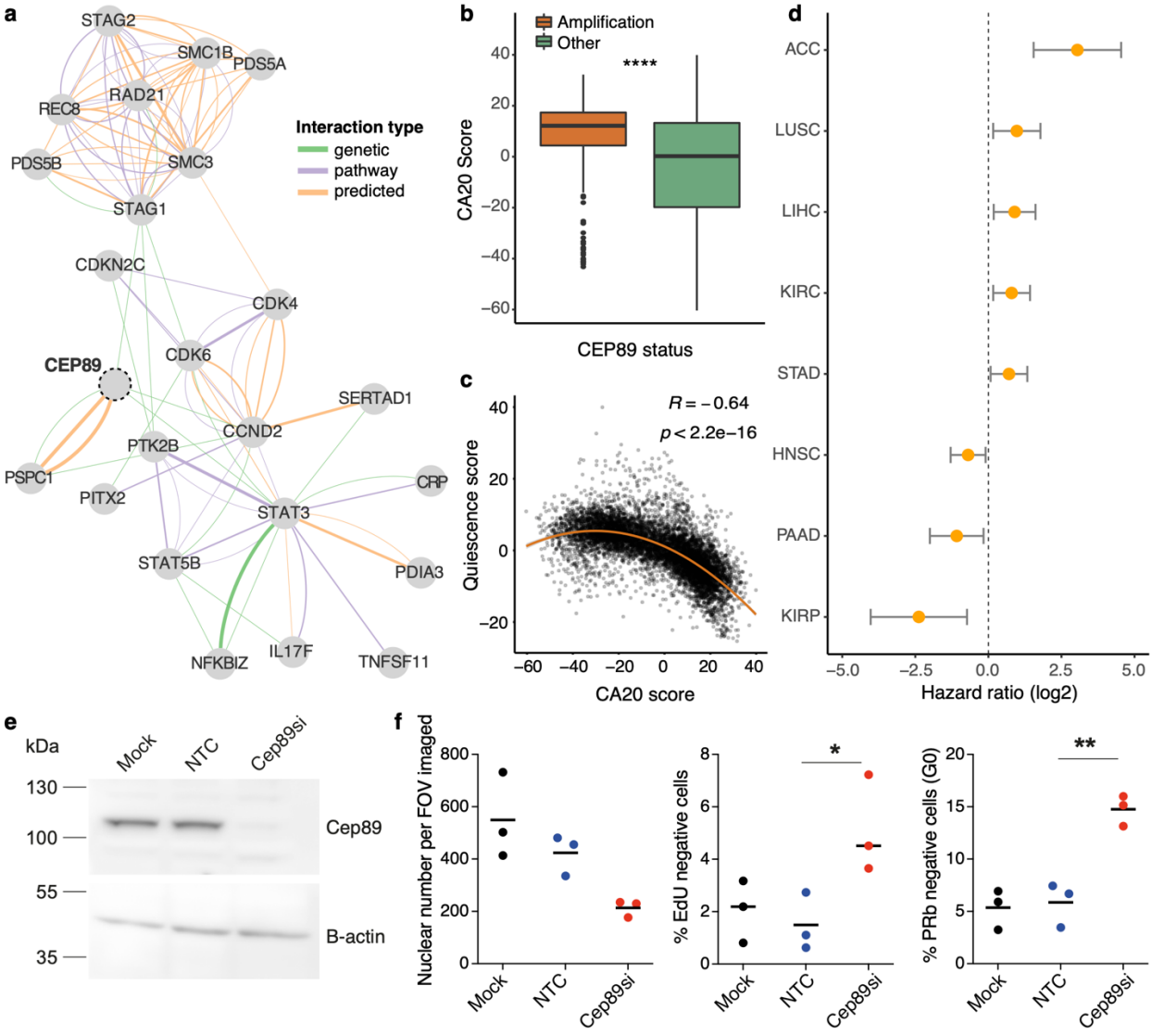


Figure 4: CEP89 amplification is associated with lower tumour quiescence

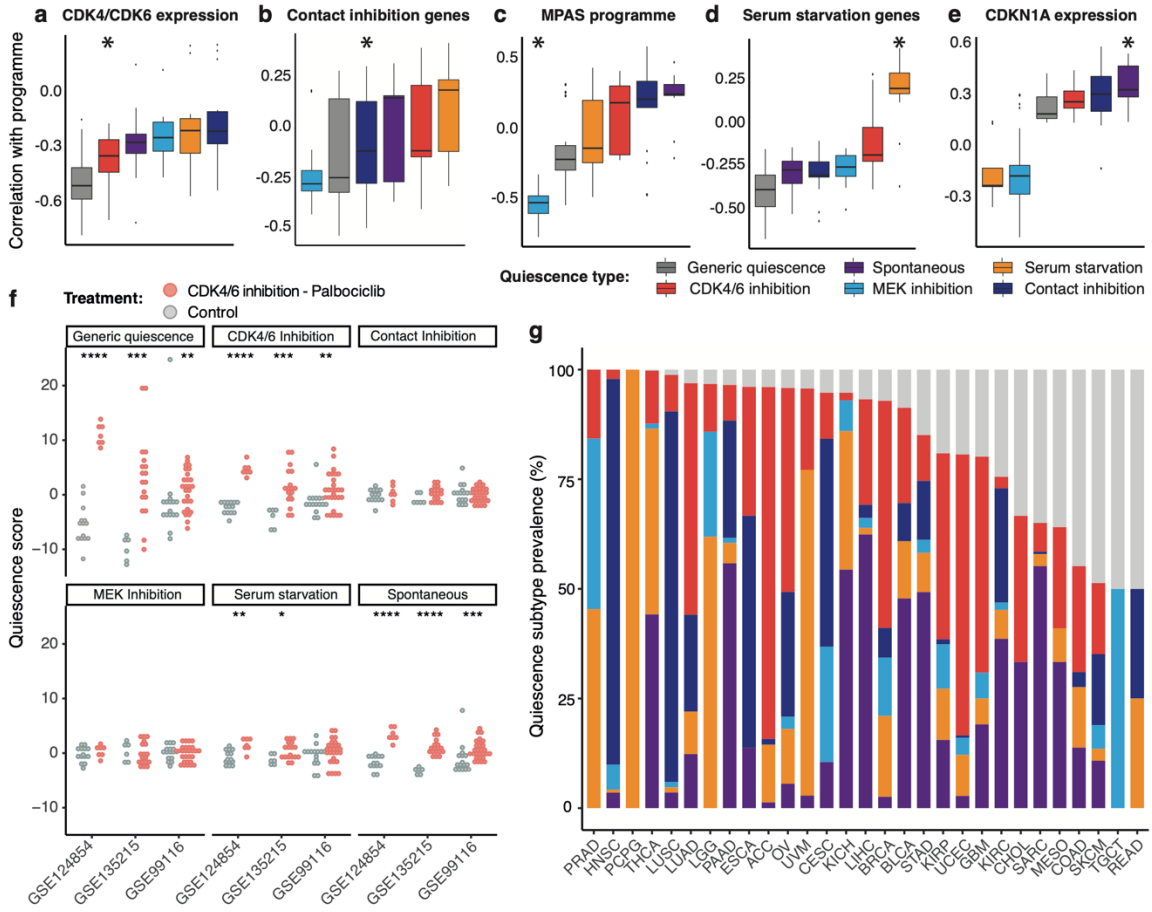


Figure 5: Pan-cancer characterisation of individual quiescence programmes

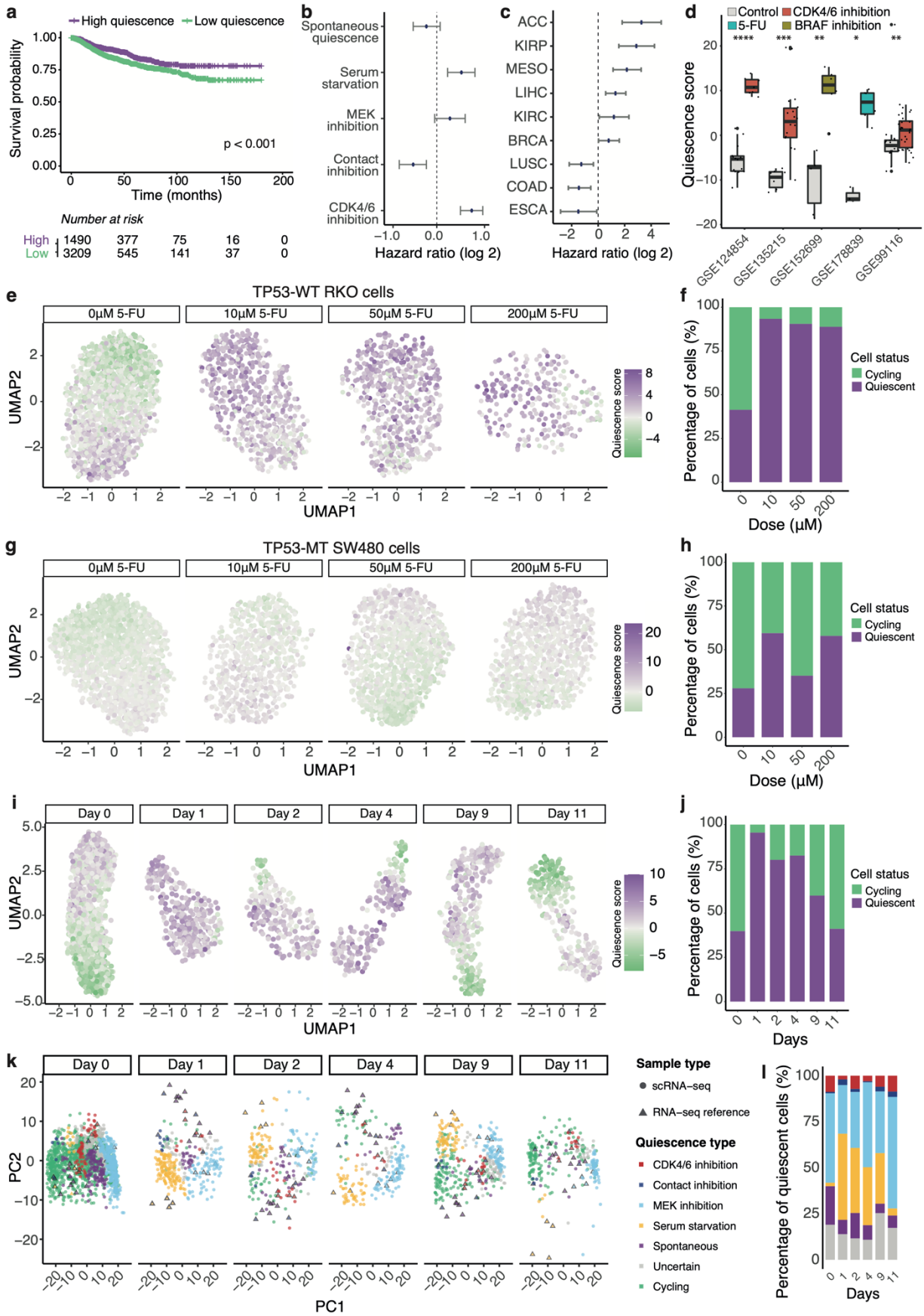


Figure 6: Impact of tumour cell quiescence on patient prognosis and treatment response

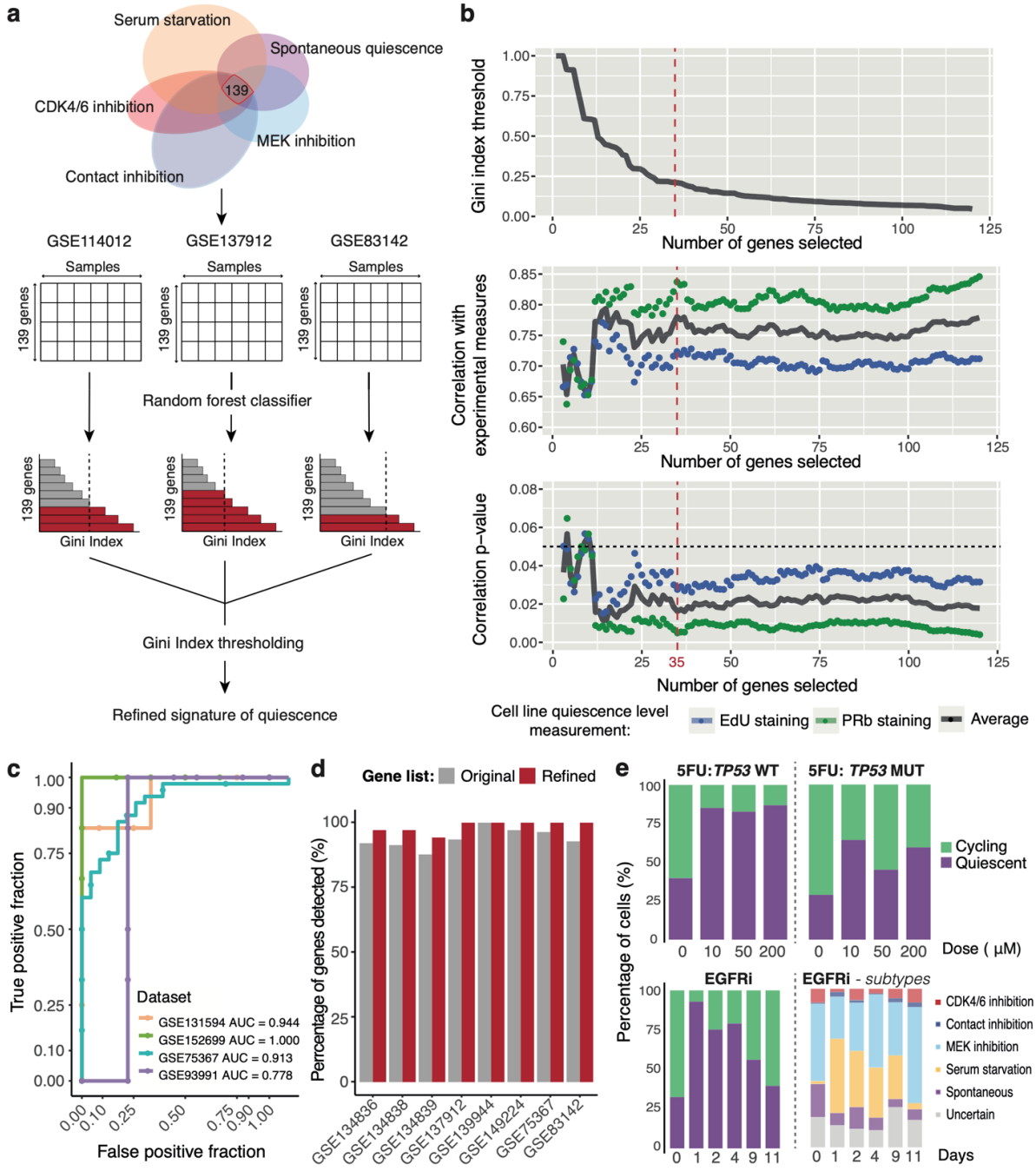


Figure 7: Optimisation of quiescence signature for use in scRNA-seq data