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).

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

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

116 **The spectrum of proliferation/quiescence capacity in solid primary tumours**

117 Having established a robust framework for quantifying tumour cell quiescence, we next profiled
118 8,005 primary tumour samples across 31 solid cancer tissues from The Cancer Genome Atlas

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

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

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

146 **The genomic background of cancer cell quiescence**

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

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

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

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

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

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

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

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

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

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

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

229 **Validation of CEP89 as a modulator of proliferation/quiescence capacity**

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

242 Our model linked *CEP89* amplification with fast cycling tumours (Figure 3c). Centrosome
243 amplification is a common feature of tumours with high proliferation rates and high genomic
244 instability[79], and overexpression of centrosomal proteins can alter centriole structure[80,81]. Indeed,
245 *CEP89* amplified tumours presented elevated expression of a previously reported centrosome
246 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

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

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

283 Overall, tumour cell quiescence appears to be beneficial for the long-term outcome of cancer patients
284 (Figure 6a, Supplementary Figure 7a). Indeed, such slow cycling, indolent tumours would have
285 higher chances of being eradicated earlier in the disease, which is consistent with reported worse
286 prognosis of patients with higher tumour cell proliferation rates and less stable, more mutated
287 genomes[86]. As expected, quiescence levels were increased in stage 1 tumours, although later stages
288 also exhibited this phenotype occasionally (Supplementary Figure 7b). However, outcomes do vary
289 depending on the quiescence subtype, with worse survival observed upon contact inhibition (Figure
290 6b). The outcomes also vary by tissue, with lung, colon or esophageal carcinoma patients displaying
291 significantly worse prognosis in the context of high tumour cell quiescence (Figure 6c).

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

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

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

326 Thus, quiescence appears to explain resistance to broad acting chemotherapy agents as well as
327 targeted molecular inhibitors of the Ras/MAPK signalling pathway, being either selected for, or
328 induced immediately upon treatment, and gradually waning over time as cells start re-entering the
329 cell cycle. Using massively multiplexed chemical transcriptomic data, we also analysed responses to
330 188 small molecule inhibitors in cell lines at single-cell resolution[91] (Supplementary Figure 8). We
331 observed a large increase in quiescence following treatment with compounds targeting cell cycle
332 regulation and tyrosine kinase signalling, consistent with our previous results, but also for
333 compounds modulating epigenetic regulation, e.g. histone deacetylase inhibitors – thus highlighting
334 the broad relevance of quiescence.

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

336 Our ability to probe the nature of quiescent phenotypes in scRNA-seq data using a defined
337 quiescence signature could aid the development of methods to selectively target quiescent drug-
338 resistant persister cells. However, a major challenge of scRNA-seq data analysis is the high
339 percentage of gene dropout, which could impact our ability to evaluate quiescence using the full 139
340 gene signature. The scRNA-seq datasets we analysed exhibited an average drop-out of 8.5 genes out
341 of the full gene signature. While our scoring method remains robust to such levels of dropout
342 (Supplementary Figure 1c-e), we also employed machine learning to reduce our initial list of 139
343 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.

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

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

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

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

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

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

421

422 **MATERIALS AND METHODS**

423 **Selection of quiescence marker genes**

424 *Generic quiescence markers:*

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

437 *Quiescence subtype-specific markers:*

438 Gene lists representing spontaneous quiescence, contact inhibition, serum starvation, CDK4/6
439 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:

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

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

495 and NCIH460 cells were plated at a density of 500 cells/well, and all other cell lines at a density of
496 1000/well, in 384well CellCarrier Ultra plates (PerkinElmer) in the relevant media. 24h later, 5 μM
497 EdU was added and cells were incubated for a further 24h before fixing in a final concentration of
498 4% formaldehyde (15 min, RT), permeabilization with PBS/0.5% Triton X-100 (15 min, RT) and
499 blocking with 2% BSA in PBS (60 min, RT). The EdU signal was detected using Click-iT chemistry,
500 according to the manufacturer's protocol (ThermoFisher). Cells were also labelled for phospho-
501 Ser807/811 Rb (PRb) using Rabbit mAb 8516 (CST) at 1:2000 in blocking solution, overnight at
502 4°C. Unbound primary antibody was washed three times in PBS and secondary Alexa-conjugated
503 antibodies were used to detect the signal (ThermoFisher, 1:1000, 1h at RT). Finally nuclei were
504 labelled with Hoechst 33258 (1 μg/ml, 15 min RT) before imaging on a high-content widefield
505 Operetta microscope, 20x N.A. 0.8. Automated image analysis (Harmony, PerkinElmer) was used
506 to segment and quantify nuclear signals in imaged cells. Quiescent cells were defined by the absence
507 of EdU or PRb staining, determined by quantification of their nuclear expression (Figure 1e-f).

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

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

526

527 **Multi-omics discovery cohort**

528 FPKM normalised RNA-sequencing expression data, copy number variation gene-level data, RPPA
529 levels for p27 as well as mutation annotation files aligned against the GRCh38 human reference
530 genome from the Mutect2 pipeline were downloaded using the *TCGABiolinks* R package [100] for
531 9,712 TCGA primary tumour samples across 31 solid cancer types. Haematological malignancies
532 were excluded as the quiescence markers were derived in epithelial cells and might not be equally
533 suited to capture this phenotype in blood. For patients with multiple samples available, one RNA-
534 seq barcode entry was selected for each individual patient resulting in 9,631 total entries. All
535 expression data were log-transformed for downstream analysis. During quiescence score calculation,
536 expression data for the primary tumour samples was scaled according to tumour purity estimates
537 reported by Hoadley et al[101] to account for potential confounding quiescence signals coming from
538 non-tumour cells in the microenvironment. Samples with purity estimates lower than 30% were
539 removed, leaving 8,005 samples for downstream analysis.

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

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

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

555 Aneuploidy scores and whole genome duplication events across TCGA samples were obtained from
556 Taylor et al[104]. Microsatellite instability status for uterine corpus endometrial carcinoma, as well as
557 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-

584 cancer or within individual TCGA studies. The *caret* R package was used to build an elastic net
585 regression model 1000 times on the training dataset of 3,753 TCGA primary tumour samples (80%
586 of the total dataset). Only samples with at least 50 mutations were used in the model, for which
587 mutational signatures could be reliably estimated. For each of the 1000 iterations, we randomly
588 selected 90% of the samples from the training dataset to build the model. Only features which were
589 included in all 1000 model iterations were selected for further analysis. To test the performance of
590 our approach, a linear regression model was built using the reduced list of genomic features and their
591 corresponding coefficients averaged across the 1000 elastic net regression model iterations. When
592 applying the resulting linear regression model on the internal validation dataset of 936 samples, we
593 found a strong correlation between the observed and predicted quiescence scores ($R = 0.73$, $p < 2.2e-$
594 16).

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

597 **Gene enrichment and network analysis**

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

603 **Gene lists**

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

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

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

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

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

616 **Validation of the genomic features of quiescence**

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

624 **Genomic dependency modelling in breast cancer**

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

630 **Survival analysis**

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

634 **Treatment response scRNA-seq and bulk RNA-seq data**

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

639

640 **Quiescence subtype determination**

641 *TCGA cohort studies:*

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

651 *Single-cell RNA seq treatment response datasets:*

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

661 **Optimisation of the quiescence signature**

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

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

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

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

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

676 *(3) External validation of candidate subsets*

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

689 The optimised gene signature is provided in Supplementary Table 6.

690 **Statistical analysis**

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

694 **Data availability**

695 The results published here are in part based upon data generated by the TCGA Research Network:
696 https://www.cancer.gov/tcga, METABRIC (https://ega-archive.org/studies/EGAS00000000083),
697 MSK-IMPACT (https://www.mskcc.org/msk-impact), or deposited at cBioPortal
698 (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).

728 5. Rehman, S.K. *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive
729 Chemotherapy. *Cell* **184**, 226-242.e21 (2021).
730 6. Phan, T.G. & Croucher, P.I. The dormant cancer cell life cycle. *Nat Rev Cancer* **20**, 398-411
731 (2020).
732 7. Cabanos, H.F. & Hata, A.N. Emerging Insights into Targeted Therapy-Tolerant Persister Cells
733 in Cancer. *Cancers (Basel)* **13**(2021).
734 8. Sharma, S.V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell
735 subpopulations. *Cell* **141**, 69-80 (2010).
736 9. Swayden, M., Chhouri, H., Anouar, Y. & Grumolato, L. Tolerant/Persister Cancer Cells and
737 the Path to Resistance to Targeted Therapy. *Cells* **9**(2020).
738 10. Xue, J.Y. *et al.* Rapid non-uniform adaptation to conformation-specific KRAS(G12C)
739 inhibition. *Nature* **577**, 421-425 (2020).
740 11. Aissa, A.F. *et al.* Single-cell transcriptional changes associated with drug tolerance and
741 response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
742 12. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of
743 WNT. *Cell* **165**, 45-60 (2016).
744 13. Ribas, A. Adaptive Immune Resistance: How Cancer Protects from Immune Attack. *Cancer*
745 *Discov* **5**, 915-9 (2015).
746 14. Sosa, M.S., Bragado, P. & Aguirre-Ghiso, J.A. Mechanisms of disseminated cancer cell
747 dormancy: an awakening field. *Nat Rev Cancer* **14**, 611-22 (2014).
748 15. Barkan, D. *et al.* Metastatic growth from dormant cells induced by a col-I-enriched fibrotic
749 environment. *Cancer Res* **70**, 5706-16 (2010).
750 16. Masago, K., Fujita, S. & Yatabe, Y. Targeting minimal residual disease after surgery with
751 molecular targeted therapy: the real path to a cure? *J Thorac Dis* **10**, S1982-S1985 (2018).
752 17. Coller, H.A., Sang, L. & Roberts, J.M. A new description of cellular quiescence. *PLoS Biol* **4**,
753 e83 (2006).
754 18. Rittershaus, E.S., Baek, S.H. & Sassetti, C.M. The normalcy of dormancy: common themes in
755 microbial quiescence. *Cell Host Microbe* **13**, 643-51 (2013).
756 19. Miles, S., Bradley, G.T. & Breeden, L.L. The budding yeast transition to quiescence. *Yeast* **38**,
757 30-38 (2021).
758 20. Marescal, O. & Cheeseman, I.M. Cellular Mechanisms and Regulation of Quiescence. *Dev*
759 *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).

791 36. Barbie, D.A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers
792 require TBK1. *Nature* **462**, 108-12 (2009).
793 37. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray
794 and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
795 38. Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**,
796 404 (2018).
797 39. Aulestia, F.J. *et al.* Quiescence status of glioblastoma stem-like cells involves remodelling of
798 Ca(2+) signalling and mitochondrial shape. *Sci Rep* **8**, 9731 (2018).
799 40. Kurppa, K.J. *et al.* Treatment-Induced Tumor Dormancy through YAP-Mediated
800 Transcriptional Reprogramming of the Apoptotic Pathway. *Cancer Cell* **37**, 104-122.e12
801 (2020).
802 41. Ebinger, S. *et al.* Characterization of Rare, Dormant, and Therapy-Resistant Cells in Acute
803 Lymphoblastic Leukemia. *Cancer Cell* **30**, 849-862 (2016).
804 42. Jordan, N.V. *et al.* HER2 expression identifies dynamic functional states within circulating
805 breast cancer cells. *Nature* **537**, 102-106 (2016).
806 43. Buczacki, S.J.A. *et al.* Itraconazole targets cell cycle heterogeneity in colorectal cancer. *J Exp*
807 *Med* **215**, 1891-1912 (2018).
808 44. Grigore, F. *et al.* BRAF inhibition in melanoma is associated with the dysregulation of histone
809 methylation and histone methyltransferases. *Neoplasia* **22**, 376-389 (2020).
810 45. Vairo, G. *et al.* Bcl-2 retards cell cycle entry through p27(Kip1), pRB relative p130, and
811 altered E2F regulation. *Mol Cell Biol* **20**, 4745-53 (2000).
812 46. Fujimaki, K. *et al.* Graded regulation of cellular quiescence depth between proliferation and
813 senescence by a lysosomal dimmer switch. *Proc Natl Acad Sci U S A* **116**, 22624-22634
814 (2019).
815 47. Stallaert, W. *et al.* The structure of the human cell cycle. *Cell Syst* **13**, 103 (2022).
816 48. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be
817 explained by the number of stem cell divisions. *Science* **347**, 78-81 (2015).
818 49. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an
819 association with cancer stemness and proliferation. *Nature Communications* **12**, 139 (2021).
820 50. Malta, T.M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic
821 Dedifferentiation. *Cell* **173**, 338-354.e15 (2018).

822 51. Dimri, G.P. *et al.* A biomarker that identifies senescent human cells in culture and in aging
823 skin in vivo. *Proceedings of the National Academy of Sciences* **92**, 9363 (1995).
824 52. Debacq-Chainiaux, F., Erusalimsky, J.D., Campisi, J. & Toussaint, O. Protocols to detect
825 senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of senescent
826 cells in culture and in vivo. *Nat Protoc* **4**, 1798-806 (2009).
827 53. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker
828 development. *PLOS Biology* **18**, e3000599 (2020).
829 54. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956 (2017).
830 55. Andreassen, P.R., Lohez, O.D., Lacroix, F.B. & Margolis, R.L. Tetraploid state induces p53-
831 dependent arrest of nontransformed mammalian cells in G1. *Molecular biology of the cell*
832 **12**, 1315-1328 (2001).
833 56. Di Leonardo, A. *et al.* DNA rereplication in the presence of mitotic spindle inhibitors in
834 human and mouse fibroblasts lacking either p53 or pRb function. *Cancer Res* **57**, 1013-9
835 (1997).
836 57. Ganem, N.J. *et al.* Cytokinesis failure triggers hippo tumor suppressor pathway activation.
837 *Cell* **158**, 833-848 (2014).
838 58. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**,
839 94-101 (2020).
840 59. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-
841 421 (2013).
842 60. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat Genet* **51**, 1732-1740
843 (2019).
844 61. Miyazono, K., Maeda, S. & Imamura, T. Coordinate regulation of cell growth and
845 differentiation by TGF-beta superfamily and Runx proteins. *Oncogene* **23**, 4232-7 (2004).
846 62. Cibis, H., Biyanee, A., Dörner, W., Mootz, H.D. & Klempnauer, K.H. Characterization of the
847 zinc finger proteins ZMYM2 and ZMYM4 as novel B-MYB binding proteins. *Sci Rep* **10**, 8390
848 (2020).
849 63. Maley, C.C. & Reid, B.J. Natural selection in neoplastic progression of Barrett's esophagus.
850 *Semin Cancer Biol* **15**, 474-83 (2005).
851 64. Bardeesy, N. & DePinho, R.A. Pancreatic cancer biology and genetics. *Nat Rev Cancer* **2**, 897-
852 909 (2002).

853 65. Indovina, P., Pentimalli, F., Casini, N., Vocca, I. & Giordano, A. RB1 dual role in proliferation
854 and apoptosis: cell fate control and implications for cancer therapy. *Oncotarget* **6**(2015).
855 66. Mas-Ponte, D. & Supek, F. DNA mismatch repair promotes APOBEC3-mediated diffuse
856 hypermutation in human cancers. *Nat Genet* **52**, 958-968 (2020).
857 67. Kloor, M. & von Knebel Doeberitz, M. The Immune Biology of Microsatellite-Unstable
858 Cancer. *Trends Cancer* **2**, 121-133 (2016).
859 68. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P.J. A molecular portrait of
860 microsatellite instability across multiple cancers. *Nature Communications* **8**, 15180 (2017).
861 69. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug
862 response. *Nature* **560**, 325-330 (2018).
863 70. Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic
864 implications. *World J Clin Oncol* **5**, 412-24 (2014).
865 71. Jakobsen, L. *et al.* Novel asymmetrically localizing components of human centrosomes
866 identified by complementary proteomics methods. *Embo j* **30**, 1520-35 (2011).
867 72. Sillibourne, J.E. *et al.* Assessing the localization of centrosomal proteins by PALM/STORM
868 nanoscopy. *Cytoskeleton (Hoboken)* **68**, 619-27 (2011).
869 73. Bettencourt-Dias, M. & Glover, D.M. Centrosome biogenesis and function: centrosomics
870 brings new understanding. *Nature Reviews Molecular Cell Biology* **8**, 451-463 (2007).
871 74. Grallert, A. *et al.* Removal of centrosomal PP1 by NIMA kinase unlocks the MPF feedback
872 loop to promote mitotic commitment in S. pombe. *Curr Biol* **23**, 213-22 (2013).
873 75. Wigley, W.C. *et al.* Dynamic association of proteasomal machinery with the centrosome.
874 *The Journal of cell biology* **145**, 481-490 (1999).
875 76. Itoh, K., Jenny, A., Mlodzik, M. & Sokol, S.Y. Centrosomal localization of Diversin and its
876 relevance to Wnt signaling. *J Cell Sci* **122**, 3791-8 (2009).
877 77. Kfoury, Y. *et al.* Ubiquitylated Tax targets and binds the IKK signalosome at the centrosome.
878 *Oncogene* **27**, 1665-1676 (2008).
879 78. Lin, A., Wang, R.T., Ahn, S., Park, C.C. & Smith, D.J. A genome-wide map of human genetic
880 interactions inferred from radiation hybrid genotypes. *Genome Res* **20**, 1122-32 (2010).
881 79. de Almeida, B.P., Vieira, A.F., Paredes, J., Bettencourt-Dias, M. & Barbosa-Morais, N.L. Pan-
882 cancer association of a centrosome amplification gene expression signature with genomic
883 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).

916 95. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data
917 using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
918 96. Amezquita, R.A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nature Methods*
919 **17**, 137-145 (2020).
920 97. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality
921 control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**,
922 1179-1186 (2017).
923 98. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis of
924 single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
925 99. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of
926 anticancer drug sensitivity. *Nature* **483**, 603-7 (2012).
927 100. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of
928 TCGA data. *Nucleic Acids Res* **44**, e71 (2016).
929 101. Hoadley, K.A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000
930 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
931 102. Zhang, P. *et al.* Germline and Somatic Genetic Variants in the p53 Pathway Interact to Affect
932 Cancer Risk, Progression, and Drug Response. *Cancer Res* **81**, 1667-1680 (2021).
933 103. Wiecek, A.J., Jacobson, D.H., Lason, W. & Secrier, M. Pan-Cancer Survey of Tumor Mass
934 Dormancy and Underlying Mutational Processes. *Frontiers in Cell and Developmental*
935 *Biology* **9**, 1820 (2021).
936 104. Taylor, A.M. *et al.* Genomic and Functional Approaches to Understanding Cancer
937 Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).
938 105. Moon, K.R. *et al.* Visualizing structure and transitions in high-dimensional biological data.
939 *Nature biotechnology* **37**, 1482-1492 (2019).
940 106. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing
941 batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*
942 **28**, 882-3 (2012).
943 107. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*
944 **171**, 1029-1041.e21 (2017).
945 108. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration
946 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

970 **FIGURE LEGENDS**

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

989

990 **Figure 2: Pan-cancer evaluation of proliferative heterogeneity and linked tumour hallmarks.**
991 **(a)** PHATE plot illustrating the wide spectrum of proliferative to slow cycling/quiescent states across
992 8,005 primary tumour samples from TCGA. Each sample is coloured according to the relative
993 quiescence level. **(b)** Variation in tumour quiescence levels across different cancer tissues. **(c)**
994 Correlation between mean quiescence scores and stem cell division estimates for various tissue types.
995 **(d)** Correlating tumour quiescence scores with cancer cell stemness (Stemness Index), telomerase
996 activity (EXTEND score), p21 activity (CDKN1A) and the expression of several commonly used
997 proliferation markers. The Pearson correlation coefficient is displayed. RC – replication complex.
998 **(e)** Consistently higher levels of quiescence are detected in samples with functional p53. **(f)** Lower
999 quiescence scores are observed in tumours with one or two whole genome duplication events.
1000 Wilcoxon rank-sum test p-values are displayed in boxplots, *$p<0.05$; **$p<0.01$; ***$p<0.001$;
1001 ****$p<0.0001$.

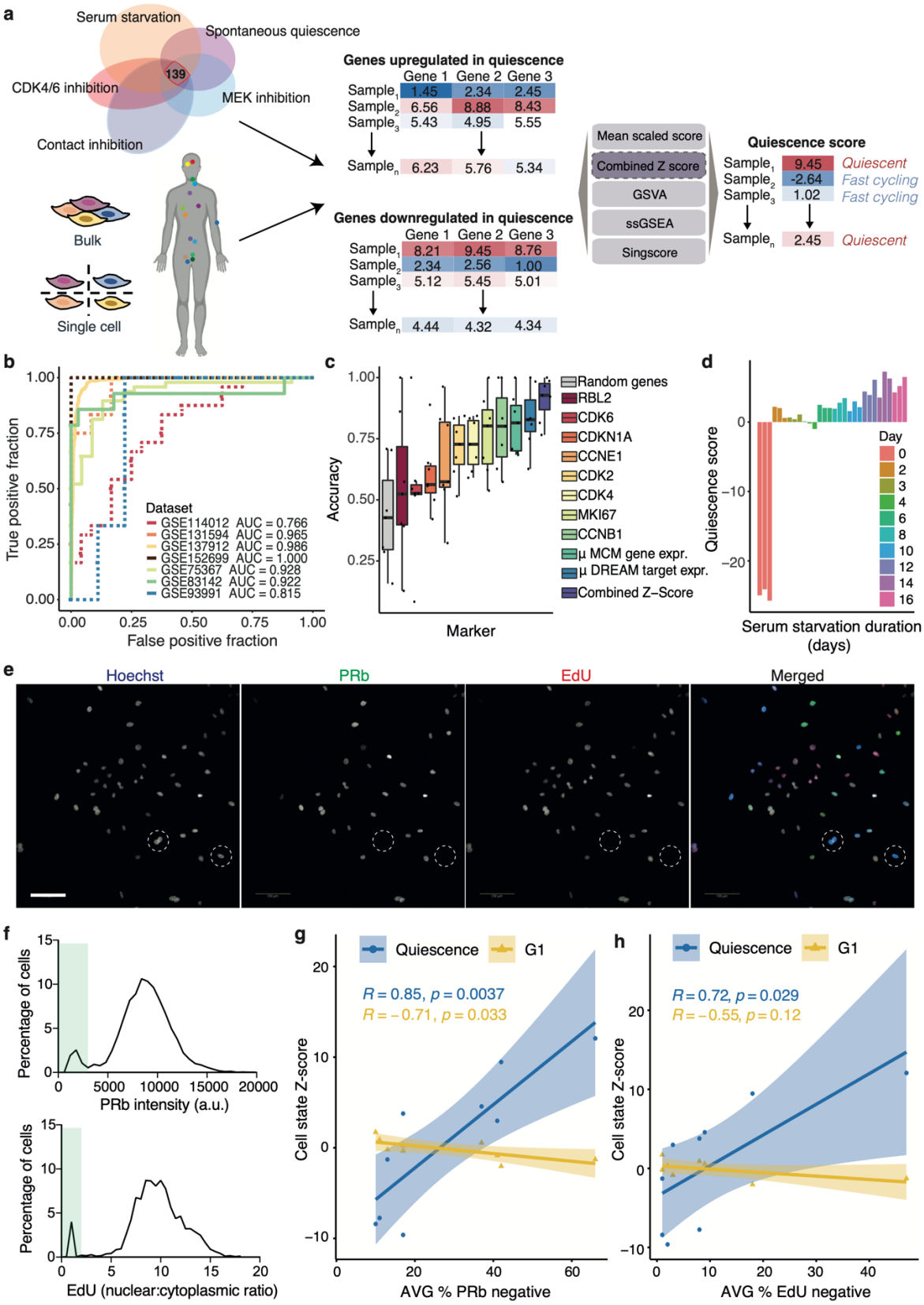
1002

1003 **Figure 3: Genomic landscape of proliferation/quiescence decisions in cancer. (a)** Cancer drivers
1004 with mutations or copy number alterations depleted pan-cancer in a quiescence context. Features
1005 further selected by the pan-cancer model are highlighted. **(b)** Schematic of the ensemble elastic net
1006 modelling employed to prioritise genomic changes associated with tumour quiescence. **(c)** Genomic
1007 events significantly associated with quiescence, ranked according to their importance in the model
1008 (highest to lowest). Each point depicts an individual tumour sample, coloured by the value of the
1009 respective feature. For discrete variables purple indicates the presence of the feature and green its
1010 absence. The Shapley values indicate the impact of individual feature values on the quiescence score
1011 prediction. **(d)** Quiescence levels are significantly reduced in microsatellite unstable (MSI) samples
1012 in stomach adenocarcinoma (STAD) and uterine corpus endometrial carcinoma (UCEC), with the
1013 same trend (albeit not significant) shown in colon adenocarcinoma (COAD). Wilcoxon rank-sum
1014 test *$p<0.05$; **$p<0.01$. **(e)** Genomic alterations are depleted across DNA repair pathways during
1015 quiescence. Odds ratios of mutational load on pathway in quiescence are depicted, along with
1016 confidence intervals. CS=chromosome segregation; p53=p53 pathway; UR=ubiquitylation response;
1017 CPF=checkpoint factors; TM=telomere maintenance; CR=chromatin remodelling; TLS=translesion
1018 synthesis; NHEJ=non-homologous end joining; NER=nucleotide excision repair; MMR=mismatch
1019 repair; FA=Fanconi Anaemia; BER=base excision repair. **(f)** Quiescence is increased in cell lines
1020 with slow doubling time across MCF7 strains, which also show lower prevalence of PTEN
1021 mutations. **(g)** Tissue-specific changes in quiescence between samples with/without quiescence-
1022 associated deletions (blue), amplifications (red) and SNVs (brown) within the TCGA cohort (top)
1023 and external validation datasets (bottom).
1024
1025 **Figure 4: CEP89 amplification is associated with lower tumour quiescence. (a)** Network
1026 illustrating CEP89 interactions with cell cycle genes (from GeneMania). The edge colour indicates
1027 the interaction type, with green representing genetic interactions, orange representing predicted
1028 interactions and purple indicating pathway interactions. The edge width illustrates the interaction
1029 weight. **(b)** CA20 scores are significantly increased in TCGA primary tumours containing a CEP89
1030 amplification. **(c)** Pan-cancer relationship between CA20 scores and tumour quiescence across the
1031 TCGA cohort. **(d)** Cox proportional hazards analysis estimates of the log hazards ratio for the impact
1032 of CEP89 expression on patient prognosis within individual cancer studies, after adjusting for tumour
1033 stage. Patients with high expression of CEP89 show significantly worse prognosis within ACC,
1034 LUSC, LIHC, KIRC and STAD, but significantly better prognosis within HNSC, PAAD and KIRP
1035 studies. **(e)** Western blot showing depletion of Cep89 protein 48h after siRNA transfection of NCI-
1036 H1299 cells. Mock is lipofectamine only, NTC is Non-targeting control siRNA. B-actin is used as a

1037 loading control. **(f)** Graphs show that Cep89 depletion in NCI-H1299 cells leads to a reduction in
1038 nuclear number and an increase in the fraction of quiescent cells, measured by an increase in the
1039 percentage of EdU negative (24h EdU pulse) and Phospho-Ser 807/811 Rb negative cells. One-Way
1040 ANOVA, *p<0.05, **p<0.01. Mean of n=3.
1041
1042 **Figure 5: Pan-cancer characterisation of individual quiescence programmes. (a-e)** Comparison
1043 of correlation coefficients between quiescence programme scores and **(a)** mean expression of CDK4
1044 and CDK6, **(b)** mean expression of curated contact inhibition genes, **(c)** a transcriptional MAPK
1045 Pathway Activity Score (MPAS), **(d)** mean expression of curated serum starvation genes, and **(e)**
1046 CDKN1A expression, across TCGA cancers. The correlations expected to be strongest (either
1047 negative or positive) are denoted by an asterisk. The generic quiescence score refers to scores
1048 calculated using the original list of 139 genes differentially expressed across all 5 forms of
1049 quiescence. **(f)** Comparison of quiescence programme scores measured in cancer cell lines before
1050 (grey) and after (red) Palbociclib treatment across three validation studies. Datasets used for
1051 validation are denoted by their corresponding GEO series accession number. **(g)** Predicted
1052 quiescence type composition of samples estimated to be highly quiescent across individual cancer
1053 types. The same colour legend as in **(a)** is applied. Gray bars represent the proportion of samples for
1054 which the quiescence type could not be estimated.
1055
1056 **Figure 6: Impact of tumour cell quiescence on patient prognosis and treatment response. (a)**
1057 Disease-specific survival based on tumour quiescence for patients from TCGA within 15 years of
1058 follow up. Patients with high levels of quiescence in primary tumours showed significantly better
1059 prognosis than patients with low quiescence. **(b-c)** Hazard ratio ranges illustrating the impact of
1060 different forms of quiescence **(b)** and different tissues **(c)** on patient prognosis, after taking into
1061 account potential confounding factors. Values above 0 indicate significantly better prognosis in the
1062 context of high tumour quiescence. **(d)** Change in quiescence scores inferred from bulk RNA-seq
1063 across breast, pancreatic, colorectal and skin cancer cells in response to treatment with the CDK4/6
1064 inhibitor Palbociclib, 5-FU or the BRAF inhibitor Vemurafenib. **(e-f)** UMAP plot illustrating the
1065 response of the TP53-proficient RKO colorectal cancer cell line to various 5-FU doses and the
1066 corresponding proportions of cells predicted to be quiescent/proliferating. Each dot is an individual
1067 cell, coloured according to its quiescence level. **(g-h)** The same as previous, but for the TP53-
1068 deficient SW480 cell line. **(i-j)** UMAP plot illustrating the response of individual PC9 NSCLC cells
1069 to the EGFR inhibitor Erlotinib across several time points and the corresponding proportion of cells
1070 predicted to be quiescent/proliferating. **(k)** Principal component analysis illustrating the

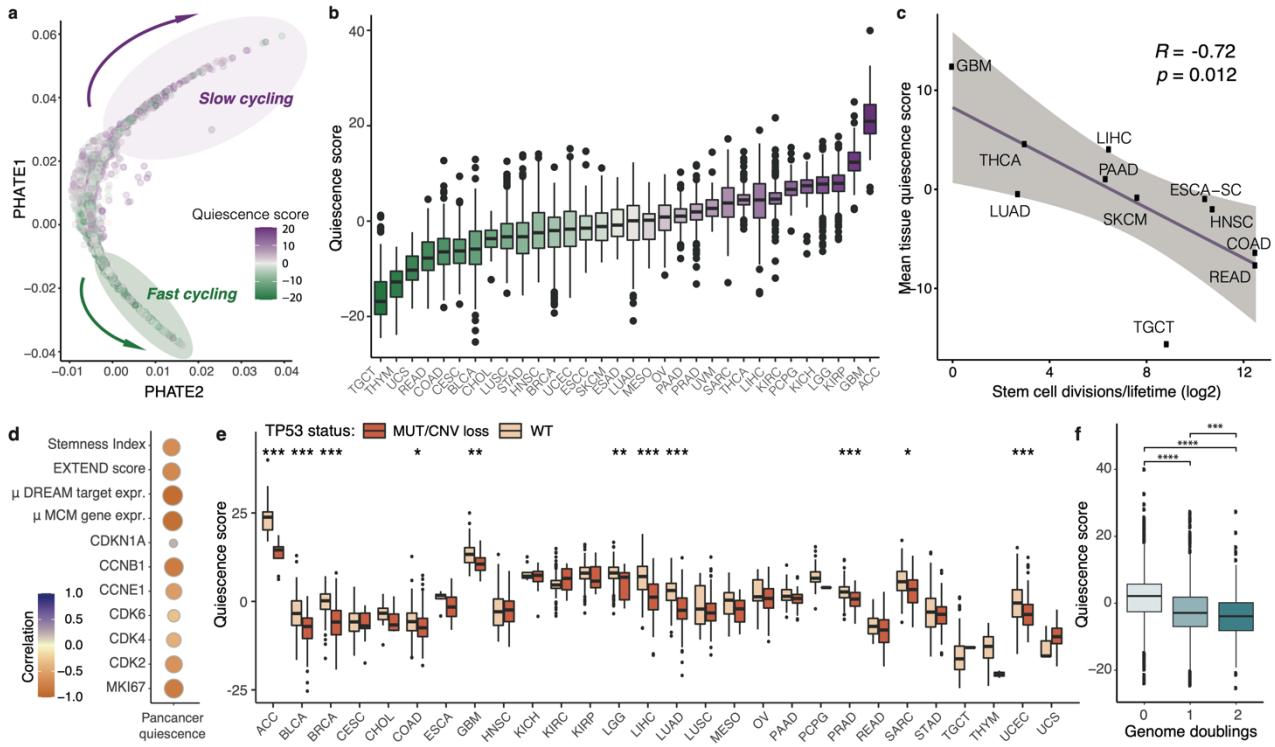
1071 superimposition of scRNA-seq profiles (circles) of quiescent NSCLC cells before/after EGFR
1072 inhibition onto the bulk RNA-seq reference data (triangles) for MCF10A cells occupying various
1073 quiescence states. **(l)** The proportion of NSCLC cells in **(k)** predicted to occupy different quiescence
1074 states across several time points.
1075
1076 **Figure 7: Optimisation of quiescence signature for use in scRNA-seq data. (a)** Methodology for
1077 refining the gene signature of quiescence: random forest classifiers are trained to distinguish
1078 quiescent from cycling tumours on three high confidence datasets; Gini index thresholding is
1079 optimised to prioritise a final list of 35 genes. **(b)** Gini index variation, correlation with
1080 experimentally measured quiescence via EdU and PRb staining assays, and corresponding p-values
1081 are plotted as the number of genes considered in the model is increased. The red dotted line indicates
1082 the threshold chosen for the final solution of 35 genes. The black dotted line indicates the threshold
1083 for p-value significance. **(c)** Additional external validation of the 35 gene signature acting as a
1084 classifier of quiescent and proliferation cells in single cell and bulk datasets. **(d)** Dropout in single
1085 cell data by gene signature. The percentage of genes out of the 35 (red) and 139 (grey) gene lists with
1086 reported expression across the single-cell RNA-seq datasets analysed in this study. **(e)** Proportion of
1087 cycling and quiescent cells estimated in single cell datasets of p53 wild type and mutant lines treated
1088 with 5FU, as well as cells treated with EGFR inhibitors. Data as in Figure 7.
1089
1090
1091
1092

1093 **FIGURES**



1094

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



1096

1097 **Figure 2: Pan-cancer evaluation of proliferative heterogeneity and linked tumour hallmarks**

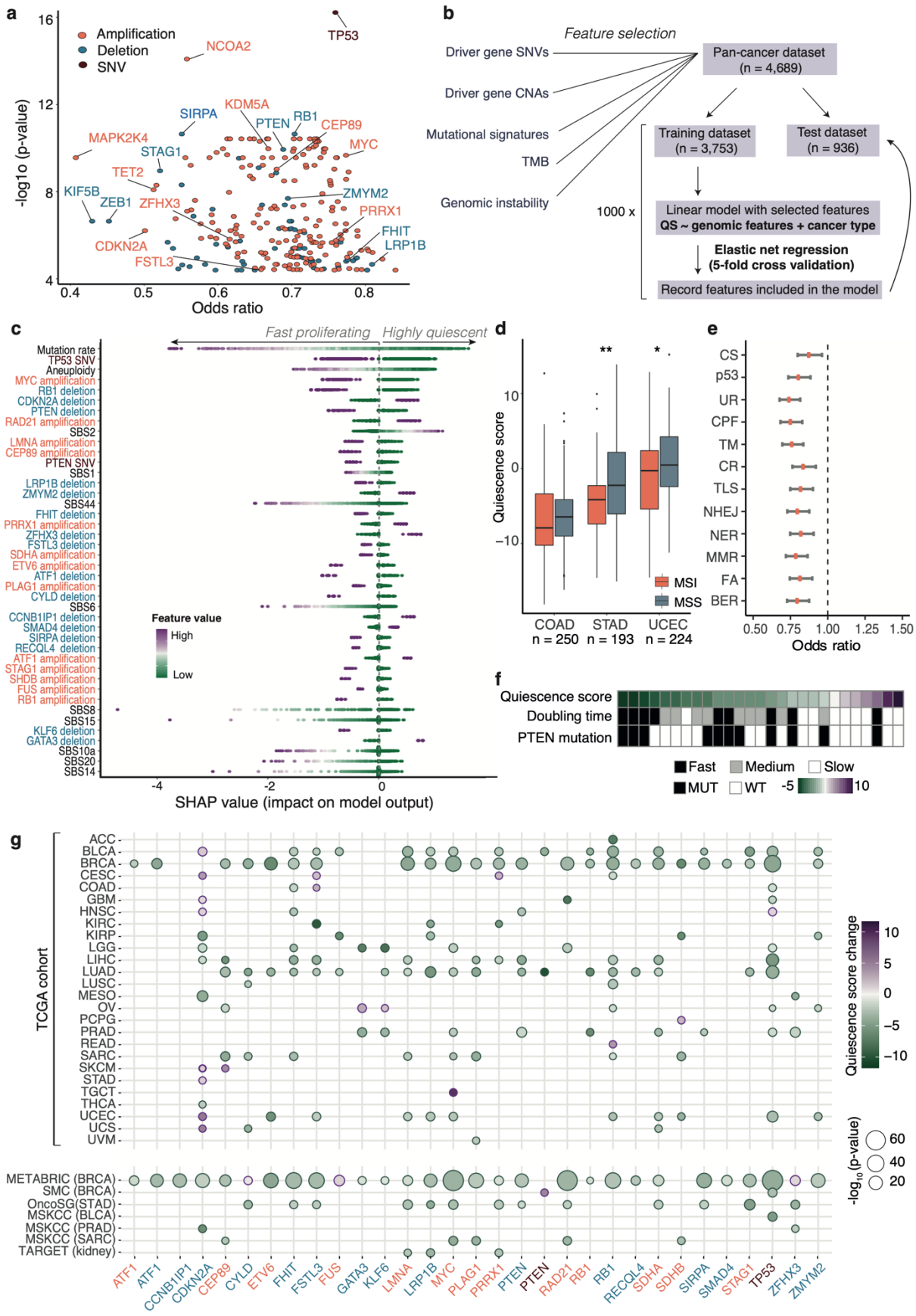


Figure 3: Genomic landscape of proliferation/quiescence decisions in cancer

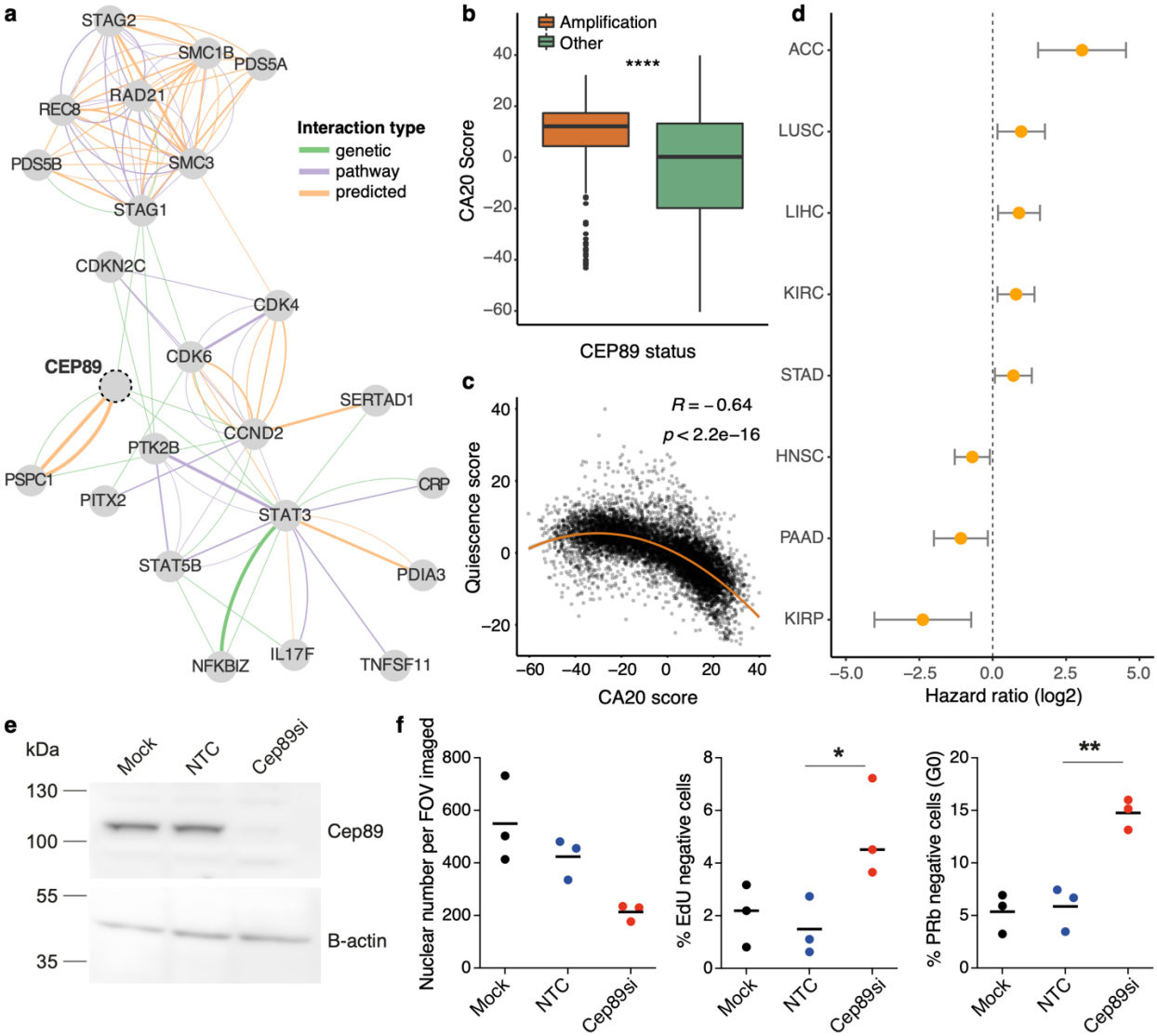


Figure 4: CEP89 amplification is associated with lower tumour quiescence

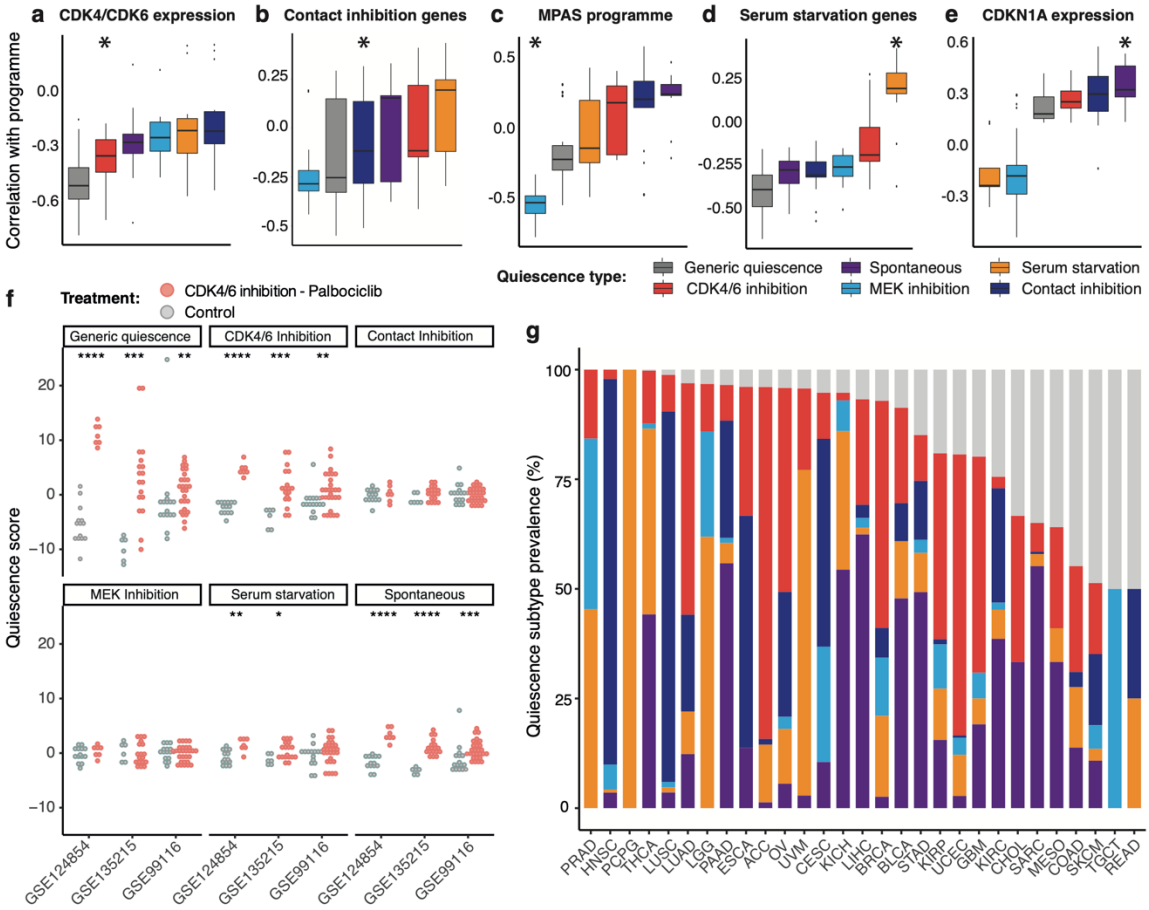


Figure 5: Pan-cancer characterisation of individual quiescence programmes

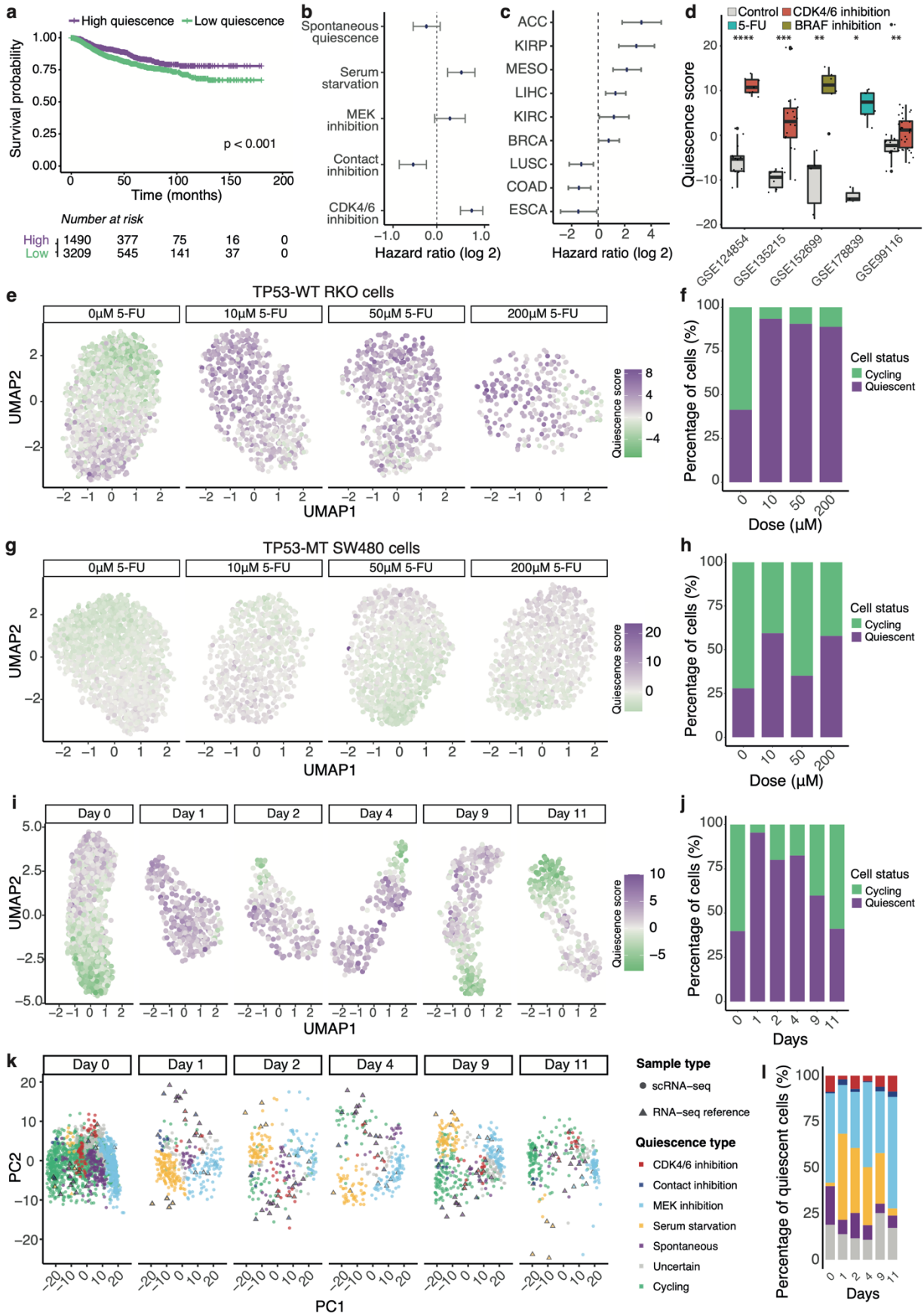


Figure 6: Impact of tumour cell quiescence on patient prognosis and treatment response

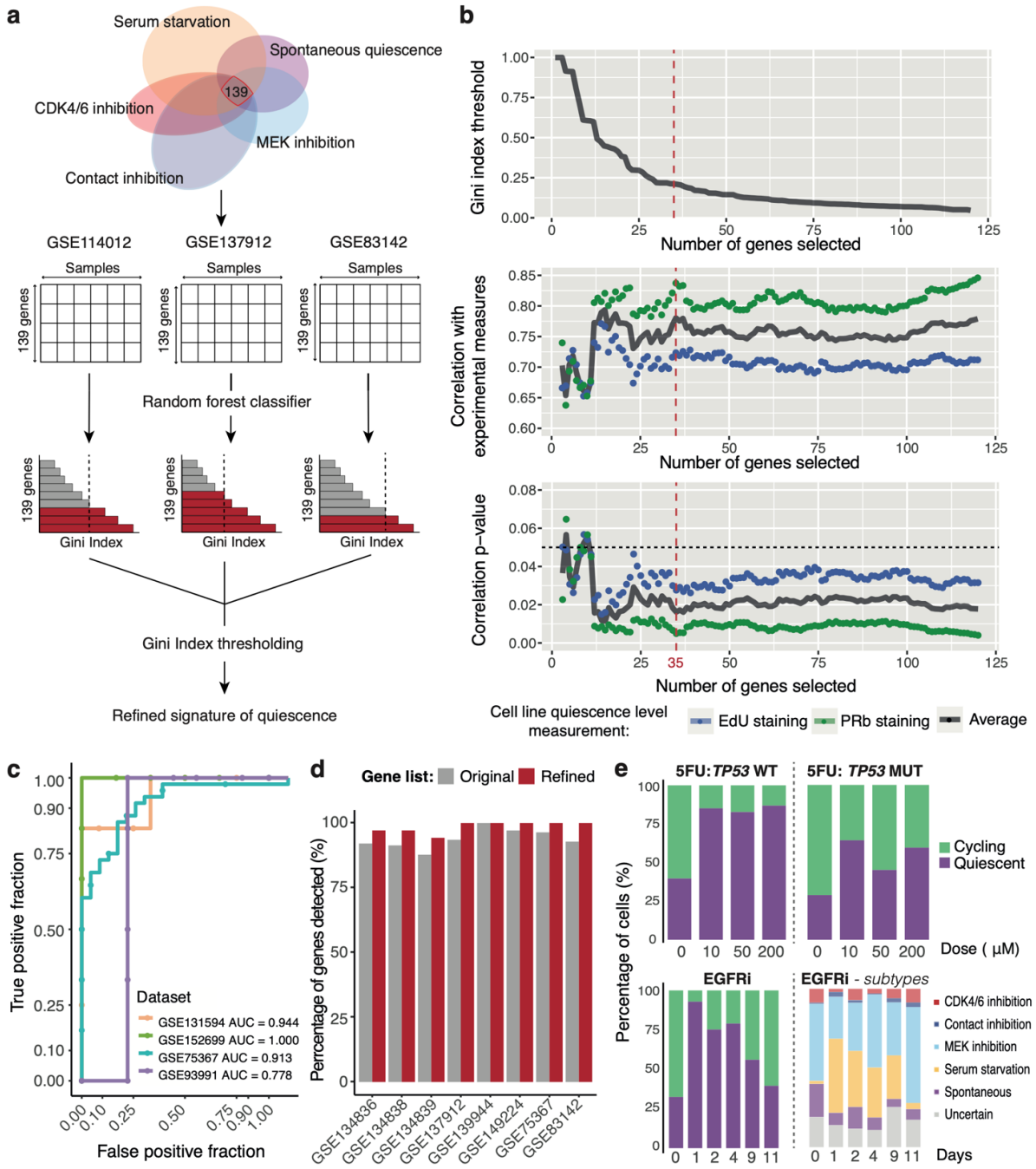


Figure 7: Optimisation of quiescence signature for use in scRNA-seq data