Genomic hallmarks of cellular dormancy in cancer and therapeutic implications

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)

ABSTRACT

Therapy resistance in cancer is often driven by a subpopulation of cells that are temporarily arrested in a non-proliferative, quiescent or 'dormant' 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 dormancy 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 uncover novel genomic dependencies of this process, including the amplification of the centrosomal gene *CEP89* as a driver of dormancy impairment. Lastly, we demonstrate that dormancy underlies unfavourable responses to various therapies exploiting cell cycle, kinase signalling and epigenetic mechanisms in single cell data, and propose a signature of dormancy-linked therapeutic resistance to further study and clinically track this state.

INTRODUCTION

Quiescence is a widely conserved state of reversible proliferative arrest, essential for the normal development and homeostasis of eukaryotic organisms[1,2]. During quiescence, cells are halted in the G0 phase of the cell cycle, where they maintain minimal basal activity[2]. In cancer, this temporary cell cycle arrest, also termed cellular dormancy, has been proposed as one of the contributing factors of resistance to treatments that target actively dividing cells, such as chemotherapy[3-5]. Moreover, a variety of other pre-existing or acquired resistance scenarios have been explained by a drug-tolerant 'persister' cell state represented by slow cycling or entirely quiescent cells[6,7]. Their presence is associated with minimal residual disease, a major cause of relapse in the clinic[8]. As the neoplastic cells evolve, dormancy can also be employed as a mechanism to facilitate immune evasion[9,10] or adaptation to new environmental niches during metastatic seeding[11,12].

Despite increasing evidence for the importance of dormancy in cancer progression, it remains a poorly defined biological programme. Hampering our understanding is the fact that quiescence represents a number of heterogeneous states[1]. Canonically, cells can be forced into quiescence through serum starvation, mitogen withdrawal or contact inhibition[1]. Cells can also enter dormancy spontaneously in response to cell-intrinsic factors like replication stress[13-15]. This process is controlled by p53[16], which triggers the inhibition of cyclin-CDK complexes by activating p21[14]. This in turn allows the assembly of the DREAM complex - a key effector responsible for repression of cell-cycle dependent gene expression[17]. Min and Spencer[18] 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[19] or altered p38/ERK signalling[20].

Despite these advances, the identification of dormant 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[21,22], quiescent cells can be mistaken for other therapy resistant cell types such as senescent or cancer stem cells[6,23]. Unlike quiescent cells, which can readily resume their proliferative state, senescent cells are irreversibly arrested[18] while cancer stem cells have a high capacity for self-renewal and sit at the top of the differentiation hierarchy[24]. Biomarkers of quiescence that are sufficiently specific and robust to be clinically useful are clearly needed.

Furthermore, the extent to which dormancy is enacted through transcriptional or genetic control is unknown[5,25], as the mutational processes and genomic changes driving quiescence have not been systematically profiled across cancer tissues. Understanding the evolutionary triggers and molecular mechanisms that enable cancer cells to enter and maintain dormancy 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 tumour dormancy using transcriptomic data, and employed it to characterise this phenomenon in bulk and single cell datasets from a variety of solid tumours. We identify and validate mutational constraints for dormancy emergence, hinting at potential new therapeutic targets that could exploit this mechanism. We also demonstrate the relevance of dormancy 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 tumour dormancy-induced treatment resistance.

RESULTS

Evaluating tumour dormancy from transcriptomic data

We hypothesised that primary tumours contain varying numbers of quiescent cells, and this in turn may determine their ability to overcome therapies targeting cycling cells. To capture this elusive

51 phenotype, we developed a computational framework that would allow us to quantify dormancy
52 signals in bulk and single cell sequenced cancer samples (Figure 1a). We employed 139
53 differentially expressed genes across five distinct forms of quiescence resulting from serum
54 starvation, contact inhibition, MEK inhibition, CDK4/6 inhibition and spontaneous quiescence, as
55 described by Min and Spencer[18], as markers of a generic quiescence state. The expression levels of
56 these markers were used to derive a score reflecting the relative level of dormancy within
57 individual tumours (see Methods).

58 We validated multiple scoring methodologies[26-29] using seven single-cell and bulk datasets[30-36]
59 where actively proliferating and quiescent cells had been independently profiled (Extended Data
60 Figure 1a, Supplementary Table 1, Methods). A combined Z-score approach had the highest
61 accuracy in detecting signals of dormancy, with a 91% mean performance in classifying cells as
62 quiescent or cycling (Figure 1b, Extended Data Figure 1b). It also outperformed classical cell cycle
63 and dormancy markers, such as the expression of targets of the DREAM complex, CDK2, Ki67
64 and of mini-chromosome replication maintenance (MCM) protein complex genes - which are
65 involved in the initiation of eukaryotic genome replication (Figure 1c). Importantly, our approach
66 provided a good separation between quiescent and proliferating samples across a variety of cancer
67 types and models including cancer cell lines, 3D organoid cultures, circulating tumour cells and
68 patient-derived xenografts (Supplementary Table 1), thereby demonstrating its broad applicability.
69 Furthermore, the strength of the score appeared to reflect the quiescence depth, i.e. the time the
70 cells stay in quiescence[37] (Figure 1d).

71 We experimentally validated our methodology in nine lung adenocarcinoma cell lines. There was a
72 remarkably good correlation between our predicted quiescence levels and the fraction of quiescent
73 cells, estimated using quantitative, single-cell imaging of phospho-Ser807/811-Rb (PRb, which
74 labels proliferative cells[38]) and 24 hour EdU proliferation assays (Figure 1e-h). Cells that were
75 negative for either PRb or EdU were defined as quiescent (see Methods, Figure 1e-f). Importantly,

76 a G1 signature (Methods) was not associated with these experimental measurements, proving our
77 method indeed captures quiescence rather than a prolonged G1 state (Figure 1g-h). The quiescence
78 correlations appeared robust to random removal of individual genes from the signature, with no
79 single gene having an inordinate impact on the score (Extended Data Figure 1c-e). This provided
80 further reassurance that our Z-score based methodology is successful in capturing dormancy
81 signals from bulk tumour data.

82 Having established a robust framework for quantifying tumour dormancy, we next profiled 8,005
83 primary tumour samples across 31 solid cancer tissues from The Cancer Genome Atlas (TCGA).
84 After accounting for potential confounding signals from the microenvironment by correcting for
85 tumour purity (see Methods, Extended Data Figure 1f-g), we observed an entire spectrum of
86 dormancy across tumours that ranged from highly dormant to highly proliferative (Figure 2a).
87 Cancers well known to be frequently dormant, such as glioblastoma[3,30], were amongst the highest
88 ranked in terms of quiescence, along with kidney and adrenocortical carcinomas (Figure 2b).

89 As expected, tissues with lower stem cell division rates presented a greater propensity for
90 quiescence (Figure 2c)[39]. Importantly, we confirmed that our quiescence scores capture a cellular
91 state that is distinct from that of cancer stem cells, marked by signatures associated with high
92 telomerase activity and an undifferentiated state[40,41], and that of senescent cells, marked by the
93 Senescence-Associated Secretory Phenotype (SASP) and β-galactosidase activity[42-44] (Figure 2d,
94 Extended Data Figures 2a-d). Our scores also showed strong negative correlations with the
95 expression of proliferation markers (Figure 2d).

96 Lastly, we confirmed dependencies on the p53/p21/DREAM activation axis: tumours that were
97 proficient in *TP53* or the components of the DREAM complex, as well as those with higher p21
98 expression, had elevated dormancy levels across numerous tissues (Figure 2e, Extended Data
99 Figures 2e-f), although only 8 out of 139 dormancy genes are directly transcriptionally regulated
100 by p53[45]. Nevertheless, p53 proficiency appears to be a non-obligatory dependency of dormancy,

101 which is also observed to arise in p53 mutant scenarios in 21% of cases. p53 has also been shown
102 to play a role in preventing the occurrence of larger structural events and polyploidy[46-48],
103 potentially explaining the lower dormancy levels we observed in tumours that had undergone
104 whole genome duplication (Figure 2f).

105
106 **The genomic background of cancer dormancy**

107 After obtaining dormancy estimates for primary tumour samples, we set out to identify potential
108 genomic triggers which might create a favourable background for tumour dormancy. We identified
109 285 cancer driver genes that were preferentially altered (via mutations or copy number alterations)
110 either in dormant or fast proliferating tumours (Figure 3a). Reassuringly, this list included genes
111 previously implicated in driving cellular quiescence-proliferation decisions such as *TP53* and
112 *MYC*[16,19]. An additional 10 cancer genes showed signals of positive selection in the context of high
113 dormancy/proliferation (Supplementary Table 2). We also investigated associations with mutagenic
114 footprints of carcinogens (termed "mutational signatures"), which can be identified as trinucleotide
115 substitution patterns in the genome[49,50]. 15 mutational signatures were linked with dormancy either
116 within individual cancer studies or pan-cancer (Extended Data Figure 2g).

117 Following the initial prioritisation of putative genomic constraints of dormancy, we sought to
118 identify those events that could best distinguish dormant from fast proliferating tumours, while
119 accounting for tissue effects. An ensemble elastic net selection approach similar to the one
120 described by Pich et al[51] was applied for this purpose (Figure 3b, Methods). Our pan-cancer model
121 identified tissue type to be a major determinant of quiescence levels (Extended Data Figure 3a). It
122 also uncovered a reduced set of 60 genomic events linked with tumour dormancy levels, including
123 SNVs and copy number losses in 17 cancer genes, as well as amplifications of 13 cancer genes
124 (Figure 3c). These events could then be successfully employed to predict quiescence in a separate
125 test dataset, thus internally validating our model (Extended Data Figure 3b).

126 Reassuringly, a functional *TP53*, lack of *MYC* amplification and lower mutation rates (Figure 3c)
127 were amongst the top ranked characteristics of dormant tumours, which also displayed less
128 aneuploidy. Additionally, we uncovered novel dependencies, such as *CEP89* and *LMNA*
129 amplifications observed in fast cycling tumours, or *ZMYM2* deletions prevalent in dormant
130 samples. p16 (*CDKN2A*) deletions, one of the frequent early events during cancer evolution[52,53],
131 were enriched in dormant tumours. *RB1* deletions and amplifications were both associated with a
132 reduction in tumour dormancy, which might reflect the dual role of RB1 in regulating proliferation
133 and apoptosis[54]. Overall, the genomic dependencies of dormancy mainly comprised genes involved
134 in cell cycle pathways, p53 regulation and ubiquitination (most likely of cell cycle targets), TGF-b
135 signalling mediators and RUNX2/3 regulation which have previously been shown to play a role in
136 controlling proliferation and cell cycle entry[55] (Extended Data Figure 3c).

137 Our model also calls attention to the broader mutational processes associated with this cellular
138 state. Such processes showed fairly weak and heterogeneous correlations with dormancy within
139 individual cancer tissues (Extended Data Figure 2g), but their contribution becomes substantially
140 clearer pan-cancer once other genomic sources are accounted for. In particular, we identified an
141 association between dormancy and mutagenesis induced by the AID/APOBEC family of cytosine
142 deaminases as denoted by signature SBS2[49] (Figure 3c). As highlighted by Mas-Ponte and Supek[56],
143 APOBEC/AID driven mutations tend to be directed towards early-replicating, gene-rich regions of
144 the genome, inducing deleterious events on several genes including *ZMYM2*, which our pan-cancer
145 model has linked with higher dormancy.

146 In turn, defective DNA mismatch repair, as evidenced by signatures SBS44, SBS20, SBS15,
147 SBS14 and SBS6[49], was prevalent in fast cycling tumours (Figure 3c). Mismatch repair
148 deficiencies lead to hypermutation in a phenomenon termed "microsatellite instability" (MSI),
149 which has been linked with increased immune evasion[57]. Cancers particularly prone to MSI include
150 colon, stomach and endometrial carcinomas[58], where this state was indeed linked with lower

151 quiescence (Figure 3d). Furthermore, dormant tumours were depleted of alterations across all DNA
152 damage repair pathways (Figure 3e).

153 Our measurements of dormancy also reflected expected cycling patterns across 27 MCF7 strains[59]:
154 cell lines with longer doubling times exhibited higher quiescence (Figure 3f). This coincided with a
155 depletion of *PTEN* mutations, a dependency highlighted by the pan-cancer model.

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

163 **CEP89 amplification contributes to high tumour proliferation rates**

164 To gain more insight into the underlying biology of cellular dormancy, we sought to
165 experimentally validate associations highlighted by the pan-cancer model. We focused on the
166 impact of *CEP89* activity on quiescence-proliferation decisions due to the high ranking of this
167 putative oncogene in the model, the relatively unexplored links between *CEP89* and cell cycle
168 control, as well as its association with quiescence across a variety of cancer cell lines (Extended
169 Data Figure 4a-c). The function of *CEP89* is not well characterised, however, the encoded protein
170 has been proposed to function as a centrosomal-associated protein[60,61]. Centrosomes function as
171 major microtubule-organising centres in cells, playing a key role in mitotic spindle assembly[62] and
172 the mitotic entry checkpoint[63]. Moreover, centrosomes act as sites of ubiquitin-mediated
173 proteolysis of cell cycle targets[64], and members of several growth signalling pathways, such as Wnt

174 and NF-kB, localise at these structures[65,66]. Several genetic interactions have also been reported
175 between CEP89 and key cell cycle proteins, including cyclin D2[67] (Figure 4a).

176 Our model linked *CEP89* amplification with fast cycling tumours (Figure 3c). Centrosome
177 amplification is a common feature of tumours with high proliferation rates and high genomic
178 instability[68], and overexpression of centrosomal proteins can alter centriole structure[69,70]. Indeed,
179 *CEP89* amplified tumours presented elevated expression of a previously reported centrosome
180 amplification signature (CA20)[68] (Figure 4b), which was strongly anticorrelated with quiescence
181 levels (Figure 4c). Furthermore, *CEP89* expression was prognostic across multiple cancer tissues
182 (Figure 4d) and linked with toxicity of several cancer compounds in cell line models (Extended
183 Data Figure 4d).

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

189 **Dormancy in breast cancer**

190 Due to the variation of dormancy levels across tissues and to increase power, we have taken a pan-
191 cancer approach to capture genomic features that are universally associated with dormancy.
192 However, we acknowledge that the extremely diverse genomic background for cancer development
193 in different tissues might mean that certain genetic alterations may lead to higher or lower
194 propensity of dormancy in a single tissue only. To investigate this further we focused on breast
195 cancer, the largest cancer dataset available from TCGA, where we still have adequate power and
196 balance of dormant/cycling phenotypes to uncover tissue-specific genomic markers of dormancy.
197 A combined ANOVA and random forest classification approach identified 30 genomic events that

198 could distinguish dormant from fast cycling tumours with a mean accuracy of 74% (Figure 5a).
199 This model recovered the associations with the APOBEC mutational signature SBS2, *TP53*
200 mutation as well as *MYC*, *KLF6*, *LMNA*, *ETV6* and *RAD21* amplification already seen in the pan-
201 cancer model (Figure 5b). All additional associations validated in the METABRIC cohort and were
202 also seen in several other cancers, particularly bladder, lung and lower grade glioma (Figure 5c).

203 Since APOBEC/AID activity has been linked with dormancy in both the pan-cancer and the tissue-
204 specific models, we further investigated its role in breast cancer, where the associated mutation
205 signature SBS2 is relatively prevalent. Indeed, breast cancer samples enriched in APOBEC-
206 associated mutations, defined as mutations attributed to SBS2 and SBS13, presented significantly
207 higher quiescence levels (Figure 5d). Across breast cancer subtypes, the Luminal A subtype
208 showed the highest levels of dormancy, however, Her2+ tumours had the highest incidence of
209 APOBEC associated mutational events, suggesting this subtype may be the most susceptible to
210 APOBEC-linked quiescence (Figure 5e-f).

211 **Characterisation of individual quiescence programmes**

212 While we had previously examined a generic dormancy programme, cancer cells can enter
213 quiescence due to different stimuli[1] and this may inform its aetiology and manifestation. To
214 explore this, we re-scored tumours based on gene expression programmes specific to serum
215 starvation, contact inhibition, MEK inhibition, CDK4/6 inhibition or spontaneously occurring
216 quiescence as defined by Min and Spencer[18] (see Methods). We observed a good correlation
217 between the dormancy estimates representing individual quiescence programmes and the
218 expression of genes associated with the corresponding form of quiescence in the literature (Figure
219 6a-e, Extended Data Figure 5). CDK4/6 inhibition-induced quiescence levels were further
220 validated using external RNA-seq datasets from cancer cell lines and xenograft mice sequenced
221 before and after treatment with the CDK4/6 inhibitor Palbociclib[71,72] (Figure 6f, Supplementary

222 Table 4). Interestingly, we also observed significant differences in spontaneous quiescence scores
223 before and after treatment. Indeed, p21 activity has been linked with the Palbociclib mechanism of
224 action[73,74], and this analysis suggests potential similarities between CDK4/6 inhibition and p21-
225 dependent quiescence phenotypes.

226 Having validated our framework for quantifying stimulus-specific quiescence programmes, we
227 proceeded to estimate the dominant form of quiescence in different cancer types (Figure 6g). We
228 found a range of quiescence aetiologies across most tissues, while a minority of cancers were
229 dominated by a single form of quiescence, e.g. serum starvation in all dormant
230 pheochromocytomas and paragangliomas, contact inhibition in 88% of head and neck carcinomas
231 and CDK4/6 inhibition in 80% of dormant adrenocortical carcinomas. While we do not wish to
232 claim that the dormancy state will have necessarily been induced by the actual predicted stimulus
233 (impossible in the case of CDK4/6 or MEK inhibition, as the analysed samples are all treatment-
234 naïve), we suggest that the downstream signalling cascade may resemble that triggered by such
235 stimuli, e.g. via CDK4/6 or MEK loss of function mutations. Amongst these states, spontaneous
236 quiescence appeared most strongly dependent on p53 functionality (Supplementary Table 3). This
237 points to common transcriptional features of drug-tolerant dormant cells in different cancer settings
238 that could be employed in designing ways to eradicate these cells in the future.

239 **Clinical relevance of tumour dormancy**

240 Overall, dormancy appears to be beneficial for the long-term outcome of cancer patients (Figure
241 7a, Extended Data Figure 6a). As expected, quiescence levels were increased in stage 1 tumours,
242 although later stages also exhibited occasional elevated dormancy (Extended Data Figure 6b).
243 Indeed, such dormant, indolent tumours could have higher chances of being eradicated earlier in
244 the disease, which is consistent with reported worse prognosis of patients with higher tumour cell
245 proliferation rates and less stable, more mutated genomes[75]. However, outcomes do vary depending

246 on the quiescence subtype, with worse survival observed upon contact inhibition (Figure 7b). The
247 outcomes also vary by tissue, with lung, colon or esophageal carcinoma patients displaying
248 significantly worse prognosis in the context of dormancy (Figure 7c).

249 While dormancy may confer an overall survival advantage in most cancers, it can also provide a
250 pool of cells that are capable of developing resistance to therapy[34,76]. Using our methodology, we
251 indeed observed an increase in quiescence levels in cell lines following treatment with EGFR,
252 BRAF and CDK4/6 inhibitors, as well as conventionally used chemotherapies such as 5-
253 Fluorouracil (5-FU) in multiple bulk RNA-seq datasets (Figure 7d). Furthermore, the recent
254 widespread availability of single-cell transcriptomics offers the opportunity to investigate the
255 impact of quiescence on such therapies with much greater granularity than is allowed by bulk data.
256 Using our quiescence signature and single-cell data from RKO and SW480 colon cancer cell lines
257 treated with 5-FU[77], we could observe quiescence-proliferation decisions following conventional
258 chemotherapy treatment. Within the p53 proficient cell line RKO, the fraction of quiescent cells
259 increased from 41% to 93% after treatment with a low dose (10 μM) of 5-FU and persisted at
260 higher doses (Figure 7e-f). In contrast, a comparable increase was not observed in *TP53* mutant
261 SW480 cells, further emphasizing the key role of p53 as a regulator of quiescence (Figure 7g-h).
262 SW480 cells showed higher apoptotic activity following treatment compared to RKO cells,
263 particularly within actively cycling cells, further corroborating that cells capable of entering
264 quiescence may be less vulnerable to this therapy (Extended Data Figure 6c-d).

265 Similarly, using single cell data from an *EGFR* mutant Non-Small Cell Lung Cancer (NSCLC) cell
266 line treated with the EGFR inhibitor Erlotinib[78], we predicted that 40% of cells were likely to exist
267 in a dormant state prior to treatment. EGFR inhibition led to a massive decrease in cell numbers
268 immediately after treatment, mostly due to proliferating cells dying off (Extended Data Figures 6e-
269 f), while the proportion of quiescent cells increased to 96% at day 1, indicating an immediate
270 selective advantage for such cells (Figure 7i-j). Interestingly, these cells appear to gradually start

271 proliferating again in the following days during continuous treatment, with the percentage of
272 proliferating cells approaching pre-treatment levels by day 11 (Figure 7j). The same trend captured
273 by our signature could be observed upon KRAS and BRAF inhibition in different cell line models
274 (Extended Data Figure 6g-j, Supplementary Table 4)[34,78]. Furthermore, during the first days of
275 treatment the NSCLC cells that survived EGFR inhibition appeared to reside in a state most
276 resembling that induced by serum starvation (Figures 7k-l). Both EGFR kinase inhibitors and
277 serum starvation have been shown to trigger autophagy[79], which may explain the convergence
278 between this inhibitory trigger and the type of quiescence response. At day 11 most of the
279 remaining dormant cells appeared in a state similar to that preceding the treatment (Figure 7l).

280 Thus, quiescence appears to explain resistance to broad acting chemotherapy agents as well as
281 targeted molecular inhibitors of the Ras/MAPK signalling pathway, being either selected for, or
282 induced immediately upon treatment, and gradually waning over time as cells start re-entering the
283 cell cycle. Using massively multiplexed chemical transcriptomic data, we also analysed responses
284 to 188 small molecule inhibitors in cell lines at single-cell resolution[80] (Extended Data Figure 7).
285 We observed a large increase in quiescence following treatment with compounds targeting cell
286 cycle regulation and tyrosine kinase signalling, consistent with our previous results, but also for
287 compounds modulating epigenetic regulation, e.g. histone deacetylase inhibitors - thus highlighting
288 the broad relevance of quiescence.

289 These analyses showcase our ability to probe the nature of quiescent phenotypes using a defined
290 dormancy signature, which could aid the development of methods to selectively target quiescent
291 drug-resistant persister cells. However, a major challenge of scRNA-seq data analysis is the high
292 percentage of gene dropout, which could impact our ability to evaluate dormancy using the full 139
293 gene signature. The scRNA-seq datasets we analysed exhibited an average drop-out of 8.5 genes
294 out of the full gene signature. While our scoring method remains robust to such levels of dropout
295 (Extended Data Figure 1c-e), we also employed machine learning to reduce our initial list of 139

296 markers of dormancy to a robust 35-gene signature, comprised mainly of RNA metabolism and
297 splicing-regulating factors, but also of genes involved in cell cycle progression, ageing and
298 senescence, which could be applied to sparser datasets with larger levels of gene dropout (Methods,
299 Figure 8a-b). The optimised signature of dormancy performed similarly to the initial broadly
300 defined programme in distinguishing highly dormant and fast cycling tumours (Figure 8c), showed
301 an average dropout of only 0.5 genes across the scRNA-seq datasets used in this study (Figure 8d),
302 was similarly prognostic (p=0.004) and showed comparable profiles of resistance to treatment
303 (Figure 8e, Extended Data Figure 8). This minimal expression signature could be employed to
304 track and further study emerging quiescence-induced resistance in a variety of therapeutic
305 scenarios.

306
307 **DISCUSSION**

308 Despite its crucial role in cancer progression and resistance to therapies, dormancy has remained
309 poorly characterised due to the scarcity of suitable models and biomarkers for large-scale tracking
310 in the tissue or blood. Lack of proliferative markers such as Ki67 or CDK2[21,81] does not uniquely
311 distinguish quiescence from other cell cycle phases, e.g. G1 or S. This study overcame this
312 limitation by employing genes active in different forms of quiescence whose patterns of expression
313 are distinct from markers of a longer G1 phase, senescence or stemness. We demonstrate a robust
314 method for profiling this signature to identify dormancy in bulk and single cell RNA-seq data, its
315 versatility evidenced by high classification accuracies across a variety of solid cancer datasets.
316 More variable performance was observed when applied to hematopoietic stem cells as it was not
317 designed to capture signals in this context (Extended Data Figure 1b).

318 We show that dormancy is pervasive across different solid cancers and generally associated with
319 more stable, less mutated genomes with intact DNA damage repair pathways. We also find a link
320 between APOBEC mutagenesis and higher levels of dormancy. We identify mutational events

321 affecting a variety of genes such as *PTEN*, *CEP89, CYLD*, *LMNA* that appear unfavourable to
322 dormancy development, thus potentially implicating them in influencing quiescence-proliferation
323 decisions. Among these, we propose *CEP89* amplification as novel biomarker of non-small cell
324 lung primary tumours with low quiescence capacity. Genomic alterations enriched in dormant
325 tumours, such as p16 or *ZMYM2* deletions, were more rare and predominantly captured pan-cancer.
326 Such targets should be further validated and could be exploited to either counteract dormancy or
327 induce it.

328 Our finding that primary tumour dormancy confers a favourable prognosis could raise a potential
329 limitation of our approach in distinguishing between slow-dividing and drug-tolerant persister cell
330 phenotypes[6] in bulk transcriptome data. Nevertheless, our methodology is directly extendable to
331 single cell data, where we demonstrate a clear association between existent/emerging quiescence
332 and resistance to multiple cell cycle, kinase signalling and epigenetic targeting regimens in a
333 variety of cancers. Quiescent cells utilise several mechanisms to achieve drug resistance, including
334 upregulation of stress-induced pathways such as anti-apoptotic BCL-2 signalling[82], anti-ROS
335 programmes[18] or immune evasion[10]. Further studies are needed to elucidate if these mechanisms
336 are specific to individual cancer tissues or individual quiescence types.

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

345

346 **ONLINE METHODS**

347 **Selection of quiescence marker genes**

348 *Generic quiescence markers:*

349 Differential expression analysis results comparing cycling immortalised, non-transformed human
350 epithelial cells and cells in five different forms of quiescence (spontaneous quiescence, contact
351 inhibition, serum starvation, CDK4/6 inhibition and MEK inhibition) were obtained from Min and
352 Spencer[18]. A total of 195 genes were differentially expressed in all five forms of quiescence under
353 an adjusted p-value cut-off of 0.05. This gene list, reflective of a generic quiescence phenotype,
354 was subjected to the following refinement and filtering steps: (1) selection of genes with a
355 unidirectional change of expression across all five forms of quiescence; (2) removal of genes
356 involved in other cell cycle stages included in the "KEGG_CELL_CYCLE" gene list deposited at
357 MSigDB; (3) removal of genes showing low standard deviation and low levels of expression within
358 the TCGA dataset, or which showed low correlation with the pan-cancer expression of the
359 transcriptional targets of the DREAM complex, the main effector of quiescence, in TCGA.

360 *Quiescence subtype-specific markers:*

361 Gene lists representing spontaneous quiescence, contact inhibition, serum starvation, CDK4/6
362 inhibition and MEK inhibition programmes were obtained using genes differentially expressed in
363 each individual quiescence form using an adjusted p-value cutoff of 0.05. The gene lists were
364 subjected to filtering steps 2 and 3 described above. Following the refinement steps, 10 upregulated
365 and 10 downregulated genes with highest log2 fold changes were selected for each quiescence
366 type.

367

368 **Dormancy quantification**

369 The *GSVA* R package was used to implement the combined z-score[26], ssGSEA[27] and GSVA[28] gene
370 set enrichment methods. For the above three methods a separate score was obtained for genes
371 upregulated in quiescence and genes downregulated in quiescence, following which a final
372 quiescence score was obtained by subtracting the two scores. The singscore single-sample gene
373 signature scoring method[29] was implemented using the *singscore* R package. In addition to these,
374 we also calculated a mean scaled quiescence score based on the refined list of genes upregulated
375 and downregulated in quiescence, as well as a curated housekeeping genes from the
376 "HSIAO_HOUSEKEEPING_GENES" list deposited at MSigDB, as follows:

377 $$\mathrm{QS} = \frac{\frac{1}{\mathrm{n}}\sum G_U - \frac{1}{\mathrm{n}}\sum G_D}{\frac{1}{\mathrm{n}}\sum G_H}$$

378 QS = mean scale quiescence score

379 G_U = expression of genes upregulated in quiescence

380 G_D = expression of genes downregulated in quiescence

381 G_H = expression of housekeeping genes

382 n = number of genes in each gene set

383 Quiescence scores for the TCGA cohort were derived from expression data scaled by tumour purity
384 estimates. The pan-cancer TCGA samples were also classified into "high" or "low" quiescence
385 groups based on k-means clustering (k=2) on the expression data of 139 quiescence biomarker
386 genes, following the removal of tissue-specific expression differences using the *ComBat* function
387 from the *sva* R package[83].

388 **Measuring the depth of quiescence**

389 We employed the GSE124109 dataset from Fujimaki et al[37] where rat embryonic fibroblasts were
390 transcriptomically profiled as they moved from shallow to deep quiescence over time in the
391 absence of growth signals. The derived quiescence scores using our combined z-score
392 methodology increased from shallow to deep quiescence.

393 **Validation of quiescence scoring methodologies**

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

395 Datasets were obtained from the ArrayExpress and Gene Expression Omnibus (GEO) databases
396 though the following GEO Series accession numbers: GSE83142, GSE75367, GSE137912,
397 GSE139013, GSE90742 and E-MTAB-4547. Quality control analysis was standardised using the
398 *SingleCellExperiment*[84] and *scater*[85] R packages. Normalisation was performed using the *scran*[86] R
399 package.

400 *Bulk RNA-sequencing validation datasets:*

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

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

408

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

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

413 Cell lines were obtained from ATCC or Sigma and regularly checked for mycoplasma. A549 and
414 NCIH460 were cultured in DMEM (Gibco). NCIH358, NCIH1299 and NCIH1563 were
415 maintained in RPMI-1640 (Gibco) supplemented with 5mM sodium pyruvate and 0.5% glucose.
416 NCIH1944, NCIH1666, NCIH1650 and L23 were grown in RPMI-1640 ATCC formulation
417 (Gibco). A427 were cultured in EMEM (ATCC). A549, NCIH460, H358, NCIH1299, NCIH1563,
418 A427 were supplemented with 10% heat inactivated FBS. NCIH1666 with 5% heat-inactivated
419 FBS and all other cell lines with 10% non-heat inactivated FBS. All cell lines had penicillin-
420 streptomycin (Gibco) added to 1%. Cells were maintained at 37°C and 5% CO_2. To calculate the
421 quiescent fraction, A549 and NCIH460 cells were plated at a density of 500 cells/well, and all
422 other cell lines at a density of 1000/well, in 384well CellCarrier Ultra plates (PerkinElmer) in the
423 relevant media. 24h later, 5 µM EdU was added and cells were incubated for a further 24h before
424 fixing in a final concentration of 4% formaldehyde (15 min, RT), permeabilization with PBS/0.5%
425 Triton X-100 (15 min, RT) and blocking with 2% BSA in PBS (60 min, RT). The EdU signal was
426 detected using Click-iT chemistry, according to the manufacturer's protocol (ThermoFisher). Cells
427 were also labelled for phospho-Ser807/811 Rb (PRb) using Rabbit mAb 8516 (CST) at 1:2000 in
428 blocking solution, overnight at 4°C. Unbound primary antibody was washed three times in PBS
429 and secondary Alexa-conjugated antibodies were used to detect the signal (ThermoFisher, 1:1000,
430 1h at RT). Finally nuclei were labelled with Hoechst 33258 (1 µg/ml, 15 min RT) before imaging
431 on a high-content widefield Operetta microscope, 20x N.A. 0.8. Automated image analysis
432 (Harmony, PerkinElmer) was used to segment and quantify nuclear signals in imaged cells.

433 Quiescent cells were defined by the absence of EdU or PRb staining, determined by quantification
434 of their nuclear expression (Figure 1e-f).

435 The quiescence scores for cancer cell lines were calculated using corresponding log-transformed
436 RPKM normalised bulk RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE)
437 database[87].

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

454 **Multi-omics discovery cohort**

455 FPKM normalised RNA-sequencing expression data, copy number variation gene-level data, as
456 well as mutation annotation files aligned against the GRCh38 human reference genome from the

457 Mutect2 pipeline were downloaded using the *TCGABiolinks* R package [88] for 9,712 TCGA primary
458 tumour samples across 31 solid cancer types. Haematological malignancies were excluded as the
459 quiescence markers were derived in epithelial cells and might not be equally suited to capture this
460 phenotype in blood. For patients with multiple samples available, one RNA-seq barcode entry was
461 selected for each individual patient resulting in 9,631 total entries. All expression data were log-
462 transformed for downstream analysis. During quiescence score calculation, expression data for the
463 primary tumour samples was scaled according to tumour purity estimates reported by Hoadley et
464 al[89] to account for potential confounding quiescence signals coming from non-tumour cells in the
465 microenvironment. Samples with purity estimates lower than 30% were removed, leaving 8,005
466 samples for downstream analysis.

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

469 *TP53* functional status was assessed based on somatic mutation and copy number alterations as
470 described in Zhang et al[90]. *TP53* mutation and copy number for the TCGA tumours were
471 downloaded from cBioPortal (http://www.cbioportal.org). Tumours with *TP53* oncogenic
472 mutations (annotated by OncoKB) and copy-number alterations (GISTIC score ≤ -1) were assigned
473 as *TP53* mutant and CNV loss. Tumours without these *TP53* alterations were assigned as TP53
474 wild type. The effects of the *TP53* mutation status on quiescence score were then determined with
475 a linear model approach with the quiescence score as a dependent variable and mutational status as
476 an independent variable. The *P* values were FDR-adjusted.

477 APOBEC mutagenesis enriched samples were determined through pan-cancer clustering of
478 mutational signature contributions as described in Wiecek et al[91]. The APOBEC mutagenesis
479 cluster was defined as the cluster with highest mean SBS2 and SBS13 contribution. This was
480 repeated 100 times and only samples which appeared in the APOBEC cluster at least 50 times were
481 counted as being APOBEC enriched.

482 Aneuploidy scores and whole genome duplication events across TCGA samples were obtained
483 from Taylor et al[92]. Microsatellite instability status for uterine corpus endometrial carcinoma, as
484 well as stomach and colon adenocarcinoma samples were obtained from Cortes-Ciriano et al[58].
485 Telomerase enzymatic activity "EXTEND" scores were obtained from Noureen et al[40]. Expression-
486 based cancer cell stemness indices were obtained from Malta et al[41]. Centrosome amplification
487 transcriptomic signature (CA20) scores were obtained from Almeida et al[68].

488 **PHATE dimensionality reduction**

489 The *phateR* R package[93] was used to perform the dimensionality reduction with a constant seed for
490 reproducibility. The *ComBat* function from the *sva* R package[94] was used to remove tissue-specific
491 expression patterns from the TCGA RNA-seq data.

492 **Cancer stem cell division estimates**

493 The mean stem cell division estimates for different cancer types used in this study were obtained
494 from Tomasetti and Vogelstein[39].

495 **Positive selection analysis**

496 Genes positively selected specifically in samples clustered into low or high dormancy groups were
497 identified based on dN/dS analysis using the *dNdScv* R package[95], run with default parameters.
498 Genes showing signals of positive selection in either the high dormancy or low dormancy samples
499 which encoded olfactory receptors were discarded from downstream analysis.

500 **Mutational signature estimation**

501 Mutational signature contributions were inferred as described in Wiecek et al[91].

502

503 **Ensemble elastic net regression model**

504 The COSMIC database was used to source a list of 723 known drivers of tumorigenesis (Tiers
505 1+2). 285 oncogenes and tumour suppressors from a curated list showed a significant enrichment
506 or depletion of mutations or copy number variants in dormant samples either pan-cancer or within
507 individual TCGA studies.

508 To classify dormant from fast proliferating tumours, the 286 genes were used as input features for
509 an ensemble elastic net regression model along the tumour mutational rate, whole-genome
510 doubling estimates, ploidy, aneuploidy scores, 10 positively selected genes and 15 mutational
511 signatures, which showed a significant correlation with quiescence scores either pan-cancer or
512 within individual TCGA studies. The *caret* R package was used to build an elastic net regression
513 model 1000 times on the training dataset of 3,753 TCGA primary tumour samples (80% of the total
514 dataset). Only samples with at least 50 mutations were used in the model, for which mutational
515 signatures could be reliably estimated. For each of the 1000 iterations, we randomly selected 90%
516 of the samples from the training dataset to build the model. Only features which were included in
517 all 1000 model iterations were selected for further analysis. To test the performance of our
518 approach, a linear regression model was built using the reduced list of genomic features and their
519 corresponding coefficients averaged across the 1000 elastic net regression model iterations. When
520 applying the resulting linear regression model on the internal validation dataset of 936 samples, we
521 found a strong correlation between the observed and predicted quiescence scores ($R = 0.73$, $p <$
522 $2.2e-16$).

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

525

526 **Gene enrichment and network analysis**

527 Gene set enrichment analysis was carried out using the *ReactomePA* R package, as well as
528 GeneMania[96] and ConsensusPathDB[97]. Interactions between CEP89 and other cell cycle
529 components were inferred using the list of cell cycle genes provided by cBioPortal and GeneMania
530 to reconstruct the expanded network with direct interactors (*STAG1*, *CCND2*, *STAT3*). Networks
531 were visualised using Cytoscape[98].

532 **Gene lists**

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

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

537 Genes associated with the senescence-associated secretory phenotype were obtained from Basisty
538 et al[44]. Lists of genes making up the various DNA damage repair pathways were derived from
539 Pearl et al[99].

540 Genes associated with contact inhibition were obtained from the curated "contact inhibition" gene
541 ontology term. Genes associated with serum starvation were obtained from the curated
542 "REACTOME_CELLULAR_RESPONSE_TO_STARVATION" list deposited at MSigDB. MEK
543 inhibition was assessed based on the activity of the MAPK pathway as determined using an
544 expression signature (MPAS) consisting of 10 downstream MAPK transcripts[100].

545 **Validation of genomic features of dormancy**

546 For elastic net model feature validation, RNA-seq data was downloaded for six cancer studies from
547 cBioPortal[101], along with patient-matched whole-genome, whole-exome and targeted sequencing

548 data. The 6 datasets used comprise breast cancer (SMC[102] and METABRIC[103]), paediatric Wilms'
549 tumor (TARGET[104]), bladder cancer, prostate adenocarcinoma and sarcoma (MSKCC[105-107])
550 studies. The data were processed and analysed in the same manner as the TCGA data. RNA-seq
551 data for 27 MCF7 cell line strains, alongside cell line growth rates and targeted mutational
552 sequencing data were obtained from Ben-David et al[59].

553 **Genomic dependency modelling in breast cancer**

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

559 **Survival analysis**

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

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

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

568

569 **Quiescence subtype determination**

570 *TCGA cohort studies:*

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

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

581 The quiescence subtype of individual single cells was inferred by mapping such individual cells
582 onto the reference dataset of MCF10A cells reflecting different forms of quiescence obtained from
583 Min and Spencer[18]. The *ComBat* R package was used to remove the study batch effect between the
584 expression data to be classified and the reference bulk RNA-seq data. PCA dimensionality
585 reduction analysis was then used on the combined datasets using the *prcomp* R function. For each
586 patient sample or single-cell expression data entry, a k-nearest neighbour algorithm classification
587 was performed using the *knn* function from the *class* R package. During the classification the three
588 nearest reference bulk RNA-seq data points were considered, with two nearest neighbours with
589 identical class needed for classification.

590

591

592 **Optimisation of dormancy signature**

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

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

597 We collected three high confidence single cell expression datasets separating quiescent from
598 dormant cells. A random forest model was trained on each dataset separately to predict the
599 quiescence state of a given cell based on the expression levels of the 139 genes. The Gini
600 indices corresponding to each gene in the model were normalised to a range of values between
601 0 and 1, which would reflect how important an individual gene was for determining quiescence
602 state relative to the other 138 genes. The procedure was repeated 1000 times for each of the
603 three datasets, and the average Gini coefficients across iterations were stored.

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

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

608 *(3) External validation of candidate subsets*

609 The gene combinations in (2) were tested for their ability to predict quiescence. For this, a
610 separate validation dataset was utilized, which contained gene expression levels for the 139
611 genes in the 10 lung cancer cell lines previously employed for experimental validation, along
612 with the quiescence state of the lines as inferred by PRb and EdU staining. For each gene
613 subset, a combined z-score of quiescence was calculated from the expression levels as
614 described previously. The correlations between this z-score and the two experimental
615 measurements of quiescence were used to establish the ability of a gene combination to predict

616 quiescence. Among the top performing subsets, a 35 gene signature with a mean correlation of
617 78% between predicted and measured quiescence levels in the test data (p=0.016) showed the
618 highest correlation with PRb measurements capturing short-lived quiescence, the more
619 common state observed in single cell treatment datasets. Therefore, this signature was deemed
620 to achieve the best trade-off between gene numbers and signal capture.

621 The optimised gene signature is available at https://github.com/secrierlab/CancerDormancy

622 **Data visualization and basic statistics**

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

626 **Data availability**

627 The results published here are in part based upon data generated by the TCGA Research Network:
628 https://www.cancer.gov/tcga, METABRIC (https://ega-archive.org/studies/EGAS00000000083),
629 MSK-IMPACT (https://www.mskcc.org/msk-impact), or deposited at cBioPortal
630 (https://www.cbioportal.org/) and GEO (https://www.ncbi.nlm.nih.gov/geo/). All data comply with
631 ethical regulations, with approval and informed consent for collection and sharing already obtained
632 by the relevant consortia.

633 **Code availability**

634 All code developed for the purpose of this study can be found at the following repository:
635 https://github.com/secrierlab/CancerDormancy

636

637 **FUNDING**

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

645 **ACKNOWLEDGEMENTS**

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

648 **AUTHOR CONTRIBUTIONS**

649 MS designed the study and supervised the computational analyses. ARB designed and supervised
650 the experimental validation in cell lines. GB supervised the analysis of p53 functional association.
651 AJW developed the dormancy scoring methodology and performed all computational analyses to
652 validate and apply it in bulk and single cell datasets, as well as link it to genomic features. SC
653 performed the experimental validation of quiescence prevalence and CEP89 association with
654 quiescence in cell lines. DK performed the inference of the minimal signature of dormancy
655 applicable in single cell data. MPC performed the random forest modelling and feature selection in
656 breast cancer. LG performed the positive selection analysis. GMT wrote the code for batch effect
657 correction and for PCA mapping of single cell data on a reference dataset. DHJ performed the
658 APOBEC enrichment classification. PZ and LX performed the quiescence comparison of p53 wild
659 type and mutated cancers. MS, AJW, ARB and SC wrote the manuscript, with contributions from
660 all other authors. All authors read and approved the manuscript.

661 **COMPETING INTEREST STATEMENT**

662 None declared.

663 **REFERENCES**

664 1. Coller, H.A., Sang, L. & Roberts, J.M. A new description of cellular quiescence. *PLoS Biol* **4**, e83
665 (2006).
666 2. van Dijk, D. *et al.* Slow-growing cells within isogenic populations have increased RNA polymerase
667 error rates and DNA damage. *Nat Commun* **6**, 7972 (2015).
668 3. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy.
669 *Nature* **488**, 522-6 (2012).
670 4. Puig, I. *et al.* TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence.
671 *J Clin Invest* **128**, 3887-3905 (2018).
672 5. Rehman, S.K. *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive
673 Chemotherapy. *Cell* **184**, 226-242.e21 (2021).
674 6. Phan, T.G. & Croucher, P.I. The dormant cancer cell life cycle. *Nat Rev Cancer* **20**, 398-411 (2020).
675 7. Cabanos, H.F. & Hata, A.N. Emerging Insights into Targeted Therapy-Tolerant Persister Cells in
676 Cancer. *Cancers (Basel)* **13**(2021).
677 8. Masago, K., Fujita, S. & Yatabe, Y. Targeting minimal residual disease after surgery with molecular
678 targeted therapy: the real path to a cure? *J Thorac Dis* **10**, S1982-S1985 (2018).
679 9. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT.
680 *Cell* **165**, 45-60 (2016).
681 10. Ribas, A. Adaptive Immune Resistance: How Cancer Protects from Immune Attack. *Cancer Discov*
682 **5**, 915-9 (2015).
683 11. Sosa, M.S., Bragado, P. & Aguirre-Ghiso, J.A. Mechanisms of disseminated cancer cell dormancy:
684 an awakening field. *Nat Rev Cancer* **14**, 611-22 (2014).
685 12. Barkan, D. *et al.* Metastatic growth from dormant cells induced by a col-I-enriched fibrotic
686 environment. *Cancer Res* **70**, 5706-16 (2010).
687 13. Arora, M., Moser, J., Phadke, H., Basha, A.A. & Spencer, S.L. Endogenous Replication Stress in
688 Mother Cells Leads to Quiescence of Daughter Cells. *Cell Rep* **19**, 1351-1364 (2017).
689 14. Barr, A.R. *et al.* DNA damage during S-phase mediates the proliferation-quiescence decision in the
690 subsequent G1 via p21 expression. *Nat Commun* **8**, 14728 (2017).
691 15. Heldt, F.S., Barr, A.R., Cooper, S., Bakal, C. & Novák, B. A comprehensive model for the
692 proliferation-quiescence decision in response to endogenous DNA damage in human cells. *Proc*
693 *Natl Acad Sci U S A* **115**, 2532-2537 (2018).
694 16. Itahana, K. *et al.* A role for p53 in maintaining and establishing the quiescence growth arrest in
695 human cells. *J Biol Chem* **277**, 18206-14 (2002).
696 17. Sadasivam, S. & DeCaprio, J.A. The DREAM complex: master coordinator of cell cycle-dependent
697 gene expression. *Nature Reviews Cancer* **13**, 585-595 (2013).
698 18. Min, M. & Spencer, S.L. Spontaneously slow-cycling subpopulations of human cells originate from
699 activation of stress-response pathways. *PLoS Biol* **17**, e3000178 (2019).
700 19. García-Gutiérrez, L., Delgado, M.D. & León, J. MYC Oncogene Contributions to Release of Cell
701 Cycle Brakes. *Genes (Basel)* **10**(2019).
702 20. Aguirre-Ghiso, J.A., Estrada, Y., Liu, D. & Ossowski, L. ERK(MAPK) activity as a determinant of
703 tumor growth and dormancy; regulation by p38(SAPK). *Cancer Res* **63**, 1684-95 (2003).
704 21. Spencer, S.L. *et al.* The proliferation-quiescence decision is controlled by a bifurcation in CDK2
705 activity at mitotic exit. *Cell* **155**, 369-83 (2013).
706 22. Miller, I. *et al.* Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence.
707 *Cell Rep* **24**, 1105-1112.e5 (2018).
708 23. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem cells.
709 *Nature* **414**, 105-11 (2001).

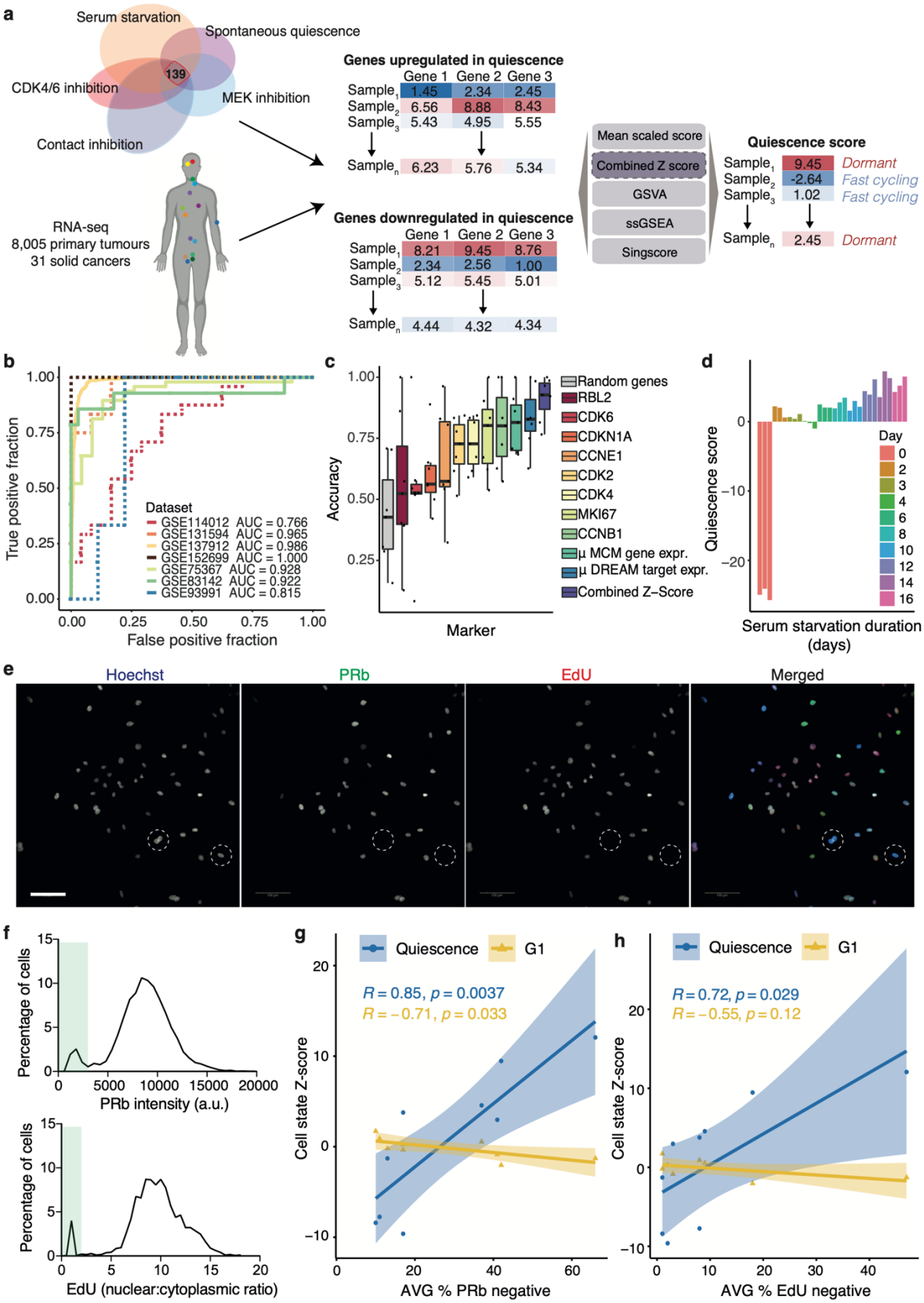
24. Kleinsmith, L.J. & Pierce, G.B., Jr. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *Cancer Res* **24**, 1544-51 (1964).
25. Turati, V.A. *et al.* Chemotherapy induces canalization of cell state in childhood B-cell precursor acute lymphoblastic leukemia. *Nature Cancer* (2021).
26. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).
27. Barbie, D.A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-12 (2009).
28. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
29. Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018).
30. Aulestia, F.J. *et al.* Quiescence status of glioblastoma stem-like cells involves remodelling of Ca(2+) signalling and mitochondrial shape. *Sci Rep* **8**, 9731 (2018).
31. Kurppa, K.J. *et al.* Treatment-Induced Tumor Dormancy through YAP-Mediated Transcriptional Reprogramming of the Apoptotic Pathway. *Cancer Cell* **37**, 104-122.e12 (2020).
32. Ebinger, S. *et al.* Characterization of Rare, Dormant, and Therapy-Resistant Cells in Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 849-862 (2016).
33. Jordan, N.V. *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102-106 (2016).
34. Xue, J.Y. *et al.* Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature* **577**, 421-425 (2020).
35. Buczacki, S.J.A. *et al.* Itraconazole targets cell cycle heterogeneity in colorectal cancer. *J Exp Med* **215**, 1891-1912 (2018).
36. Grigore, F. *et al.* BRAF inhibition in melanoma is associated with the dysregulation of histone methylation and histone methyltransferases. *Neoplasia* **22**, 376-389 (2020).
37. Fujimaki, K. *et al.* Graded regulation of cellular quiescence depth between proliferation and senescence by a lysosomal dimmer switch. *Proc Natl Acad Sci U S A* **116**, 22624-22634 (2019).
38. Stallaert, W. *et al.* The structure of the human cell cycle. *bioRxiv*, 2021.02.11.430845 (2021).
39. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78-81 (2015).
40. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an association with cancer stemness and proliferation. *Nature Communications* **12**, 139 (2021).
41. Malta, T.M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354.e15 (2018).
42. Dimri, G.P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences* **92**, 9363 (1995).
43. Debacq-Chainiaux, F., Erusalimsky, J.D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of senescent cells in culture and in vivo. *Nat Protoc* **4**, 1798-806 (2009).
44. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLOS Biology* **18**, e3000599 (2020).
45. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956 (2017).
46. Andreassen, P.R., Lohez, O.D., Lacroix, F.B. & Margolis, R.L. Tetraploid state induces p53-dependent arrest of nontransformed mammalian cells in G1. *Molecular biology of the cell* **12**, 1315-1328 (2001).
47. Di Leonardo, A. *et al.* DNA rereplication in the presence of mitotic spindle inhibitors in human and mouse fibroblasts lacking either p53 or pRb function. *Cancer Res* **57**, 1013-9 (1997).
48. Ganem, N.J. *et al.* Cytokinesis failure triggers hippo tumor suppressor pathway activation. *Cell* **158**, 833-848 (2014).
49. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
50. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
51. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat Genet* **51**, 1732-1740 (2019).

765 52. Maley, C.C. & Reid, B.J. Natural selection in neoplastic progression of Barrett's esophagus. *Semin*
766 *Cancer Biol* **15**, 474-83 (2005).
767 53. Bardeesy, N. & DePinho, R.A. Pancreatic cancer biology and genetics. *Nat Rev Cancer* **2**, 897-909
768 (2002).
769 54. Indovina, P., Pentimalli, F., Casini, N., Vocca, I. & Giordano, A. RB1 dual role in proliferation and
770 apoptosis: cell fate control and implications for cancer therapy. *Oncotarget* **6**(2015).
771 55. Miyazono, K., Maeda, S. & Imamura, T. Coordinate regulation of cell growth and differentiation by
772 TGF-beta superfamily and Runx proteins. *Oncogene* **23**, 4232-7 (2004).
773 56. Mas-Ponte, D. & Supek, F. DNA mismatch repair promotes APOBEC3-mediated diffuse
774 hypermutation in human cancers. *Nat Genet* **52**, 958-968 (2020).
775 57. Kloor, M. & von Knebel Doeberitz, M. The Immune Biology of Microsatellite-Unstable Cancer.
776 *Trends Cancer* **2**, 121-133 (2016).
777 58. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P.J. A molecular portrait of
778 microsatellite instability across multiple cancers. *Nature Communications* **8**, 15180 (2017).
779 59. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response.
780 *Nature* **560**, 325-330 (2018).
781 60. Jakobsen, L. *et al.* Novel asymmetrically localizing components of human centrosomes identified
782 by complementary proteomics methods. *Embo j* **30**, 1520-35 (2011).
783 61. Sillibourne, J.E. *et al.* Assessing the localization of centrosomal proteins by PALM/STORM
784 nanoscopy. *Cytoskeleton (Hoboken)* **68**, 619-27 (2011).
785 62. Bettencourt-Dias, M. & Glover, D.M. Centrosome biogenesis and function: centrosomics brings
786 new understanding. *Nature Reviews Molecular Cell Biology* **8**, 451-463 (2007).
787 63. Grallert, A. *et al.* Removal of centrosomal PP1 by NIMA kinase unlocks the MPF feedback loop to
788 promote mitotic commitment in S. pombe. *Curr Biol* **23**, 213-22 (2013).
789 64. Wigley, W.C. *et al.* Dynamic association of proteasomal machinery with the centrosome. *The*
790 *Journal of cell biology* **145**, 481-490 (1999).
791 65. Itoh, K., Jenny, A., Mlodzik, M. & Sokol, S.Y. Centrosomal localization of Diversin and its
792 relevance to Wnt signaling. *J Cell Sci* **122**, 3791-8 (2009).
793 66. Kfoury, Y. *et al.* Ubiquitylated Tax targets and binds the IKK signalosome at the centrosome.
794 *Oncogene* **27**, 1665-1676 (2008).
795 67. Lin, A., Wang, R.T., Ahn, S., Park, C.C. & Smith, D.J. A genome-wide map of human genetic
796 interactions inferred from radiation hybrid genotypes. *Genome Res* **20**, 1122-32 (2010).
797 68. de Almeida, B.P., Vieira, A.F., Paredes, J., Bettencourt-Dias, M. & Barbosa-Morais, N.L. Pan-
798 cancer association of a centrosome amplification gene expression signature with genomic alterations
799 and clinical outcome. *PLOS Computational Biology* **15**, e1006832 (2019).
800 69. Kohlmaier, G. *et al.* Overly long centrioles and defective cell division upon excess of the SAS-4-
801 related protein CPAP. *Curr Biol* **19**, 1012-8 (2009).
802 70. Tang, C.-J.C., Fu, R.-H., Wu, K.-S., Hsu, W.-B. & Tang, T.K. CPAP is a cell-cycle regulated
803 protein that controls centriole length. *Nature Cell Biology* **11**, 825-831 (2009).
804 71. Hafner, M. *et al.* Multiomics Profiling Establishes the Polypharmacology of FDA-Approved
805 CDK4/6 Inhibitors and the Potential for Differential Clinical Activity. *Cell Chem Biol* **26**, 1067-
806 1080.e8 (2019).
807 72. Salvador-Barbero, B. *et al.* CDK4/6 Inhibitors Impair Recovery from Cytotoxic Chemotherapy in
808 Pancreatic Adenocarcinoma. *Cancer Cell* **37**, 340-353.e6 (2020).
809 73. Guiley, K.Z. *et al.* p27 allosterically activates cyclin-dependent kinase 4 and antagonizes
810 palbociclib inhibition. *Science* **366**(2019).
811 74. Pack, L.R., Daigh, L.H., Chung, M. & Meyer, T. Clinical CDK4/6 inhibitors induce selective and
812 immediate dissociation of p21 from cyclin D-CDK4 to inhibit CDK2. *Nat Commun* **12**, 3356
813 (2021).
814 75. Zhu, X. *et al.* The prognostic and predictive potential of Ki-67 in triple-negative breast cancer.
815 *Scientific Reports* **10**, 225 (2020).
816 76. Brown, J.A. *et al.* TGF-β-Induced Quiescence Mediates Chemoresistance of Tumor-Propagating
817 Cells in Squamous Cell Carcinoma. *Cell Stem Cell* **21**, 650-664.e8 (2017).
818 77. Park, S.R. *et al.* Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-
819 Fluorouracil-Induced DNA Damage. *Cell Rep* **32**, 108077 (2020).

78. Aissa, A.F. *et al.* Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
79. Tan, X., Thapa, N., Sun, Y. & Anderson, R.A. A kinase-independent role for EGF receptor in autophagy initiation. *Cell* **160**, 145-60 (2015).
80. Srivatsan, S.R. *et al.* Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45-51 (2020).
81. Moser, J., Miller, I., Carter, D. & Spencer, S.L. Control of the Restriction Point by Rb and p21. *Proc Natl Acad Sci U S A* **115**, E8219-e8227 (2018).
82. Minassian, L.M., Cotechini, T., Huitema, E. & Graham, C.H. Hypoxia-Induced Resistance to Chemotherapy in Cancer. *Adv Exp Med Biol* **1136**, 123-139 (2019).
83. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
84. Amezquita, R.A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nature Methods* **17**, 137-145 (2020).
85. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017).
86. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
87. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-7 (2012).
88. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71 (2016).
89. Hoadley, K.A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
90. Zhang, P. *et al.* Germline and Somatic Genetic Variants in the p53 Pathway Interact to Affect Cancer Risk, Progression, and Drug Response. *Cancer Res* **81**, 1667-1680 (2021).
91. Wiecek, A.J., Jacobson, D.H., Lason, W. & Secrier, M. Pan-Cancer Survey of Tumor Mass Dormancy and Underlying Mutational Processes. *Frontiers in Cell and Developmental Biology* **9**, 1820 (2021).
92. Taylor, A.M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).
93. Moon, K.R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology* **37**, 1482-1492 (2019).
94. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-3 (2012).
95. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e21 (2017).
96. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-20 (2010).
97. Kamburov, A. *et al.* ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res* **39**, D712-7 (2011).
98. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
99. Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B. & Pearl, F.M.G. Therapeutic opportunities within the DNA damage response. *Nature Reviews Cancer* **15**, 166-180 (2015).
100. Wagle, M.C. *et al.* A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. *NPJ Precis Oncol* **2**, 7 (2018).
101. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
102. Kan, Z. *et al.* Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat Commun* **9**, 1725 (2018).
103. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-52 (2012).

875 104. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic
876 landscape of Wilms tumor. *Nat Genet* **49**, 1487-1494 (2017).
877 105. Iyer, G. *et al.* Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder
878 cancer. *J Clin Oncol* **31**, 3133-40 (2013).
879 106. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22
880 (2010).
881 107. Barretina, J. *et al.* Subtype-specific genomic alterations define new targets for soft-tissue sarcoma
882 therapy. *Nat Genet* **42**, 715-21 (2010).
883

884 **FIGURES**



885

886 **Figure 1: Methodology for quantifying cancer dormancy. (a)** Workflow for evaluating

887 dormancy from RNA-seq data. 139 genes differentially expressed in multiple forms of quiescence

888 were employed to score dormancy across cancer tissues. **(b)** Receiver operating characteristic
889 (ROC) curves illustrating the performance of the z-score methodology on separating actively
890 proliferating and quiescent cells in seven single-cell (continuous curves) and bulk RNA-seq (dotted
891 curves) datasets. AUC = area under the curve. **(c)** Compared classification accuracies of the z-score
892 approach and classic cell proliferation markers across the seven single-cell/bulk RNA-seq
893 validation datasets. **(d)** Quiescence levels of embryonic fibroblast cells under serum starvation for
894 various amounts of time. Replicates are depicted in the same colour. **(e)** Representative images of
895 lung cancer cell lines immunostained and analysed to detect the quiescent fraction. Hoechst (labels
896 all nuclei) is in blue, PRb in green and EdU in red in merged image. White dashed circles highlight
897 quiescent cells that are negative for both PRb and EdU signals. Scale bar: 100 µm. **(f)** Graphs show
898 single cell quantification of PRb and EdU intensities taken from images and used to define the cut-
899 off to calculate the quiescent fraction (QF; green boxes). Images in **(e)** and graphs in **(f)** are taken
900 from the A549 cell line. **(g-h)** Correlation between theoretical estimates of a quiescence or G1 state
901 and the fraction of cells entering quiescence in nine lung adenocarcinoma cell lines, as assessed
902 through **(g)** phospho-Rb assays and **(h)** EdU assays. Mean of n=3 is shown for the average
903 percentage of quiescent cells.
904

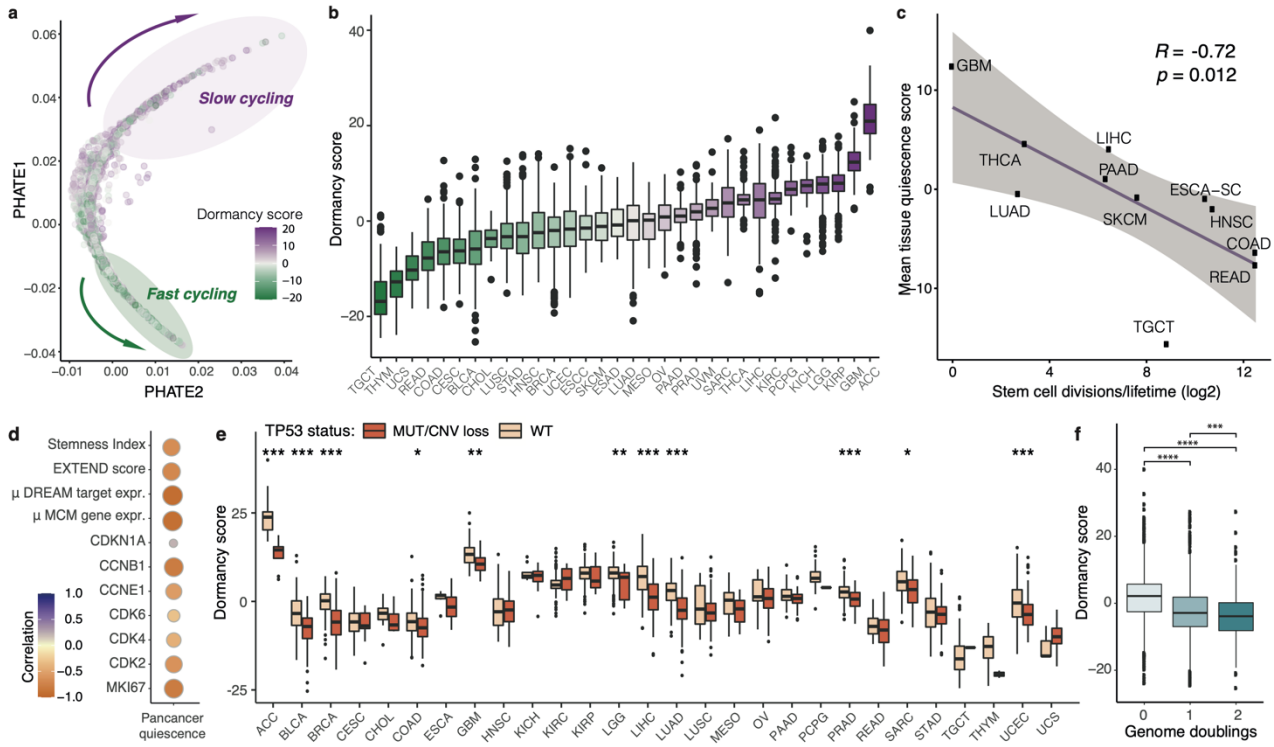


Figure 2: Pan-cancer evaluation of dormancy. (a) PHATE plot illustrating the wide spectrum of dormancy-proliferation states across 8,005 primary tumour samples from TCGA. Each sample is coloured according to the relative dormancy level. **(b)** Variation in tumour dormancy levels across different cancer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

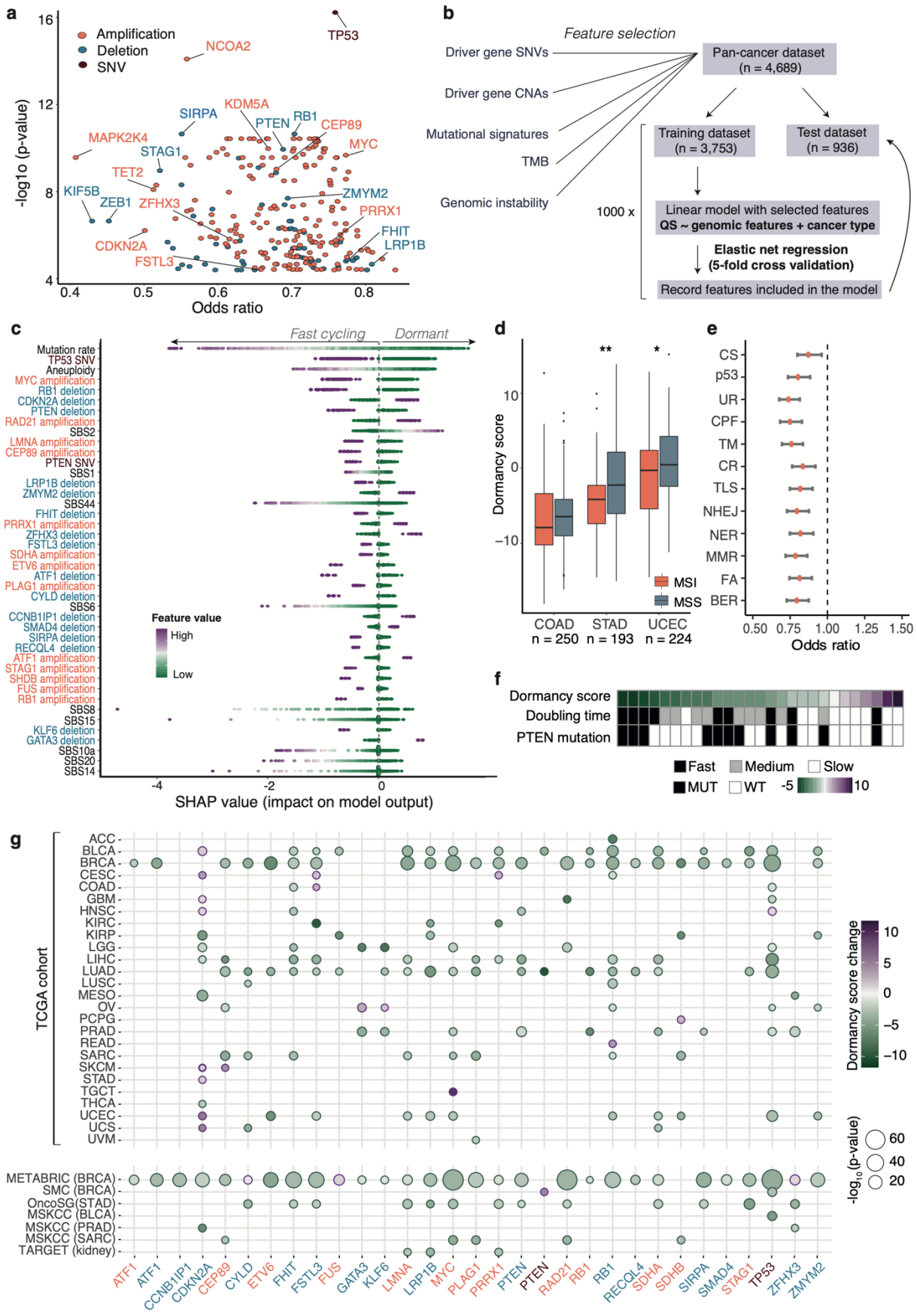


Figure 3: Pan-cancer genomic landscape of dormancy. (a) Cancer drivers with mutations or copy number alterations depleted pan-cancer in a dormancy context. Features further selected by

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

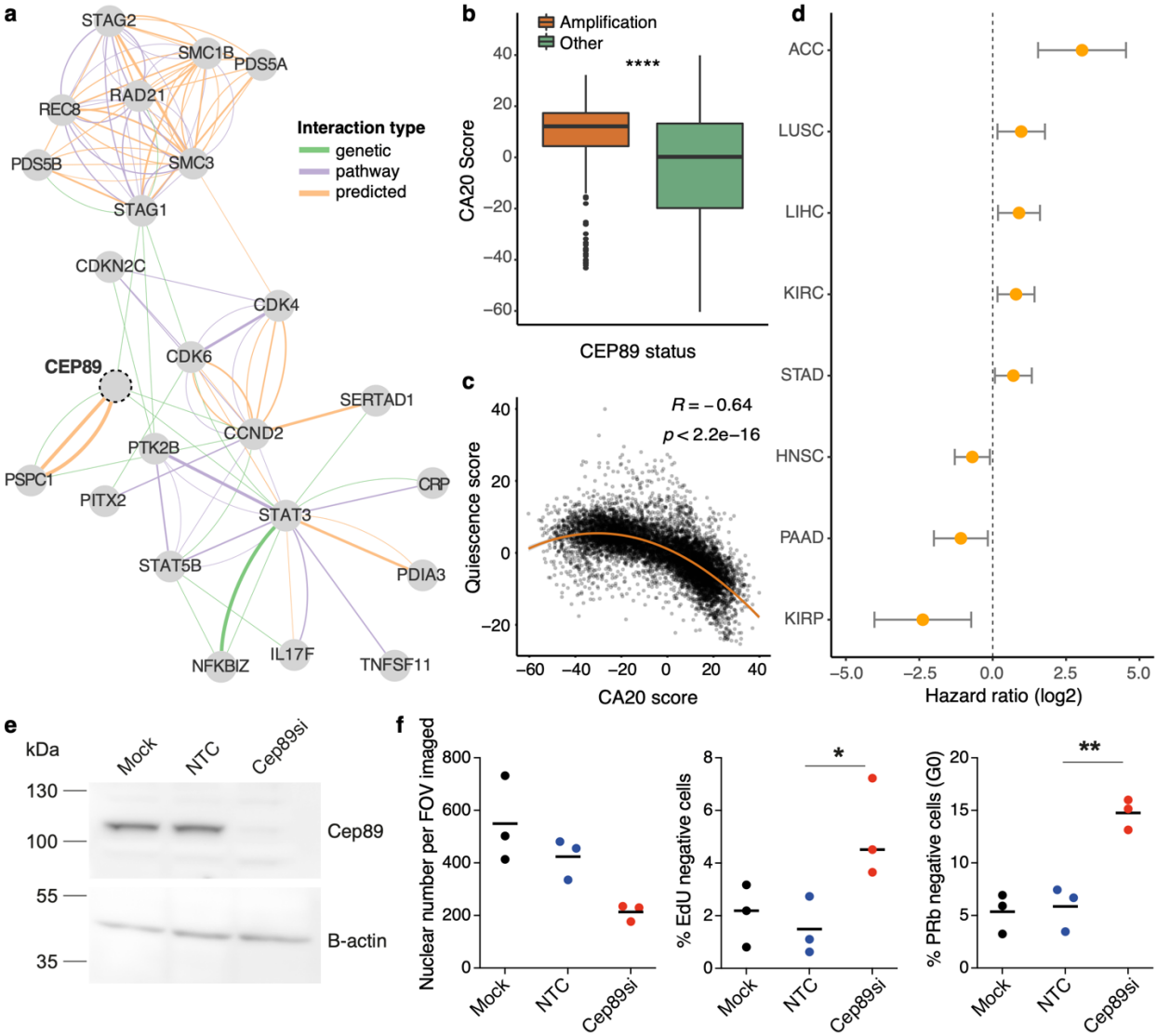


Figure 4: CEP89 amplification is associated with lower tumour dormancy. (a) Network illustrating CEP89 interactions with cell cycle genes (from GeneMania). The edge colour indicates the interaction type, with green representing genetic interactions, orange representing predicted interactions and purple indicating pathway interactions. The edge width illustrates the interaction weight. **(b)** CA20 scores are significantly increased in TCGA primary tumours containing a CEP89 amplification. **(c)** Pan-cancer relationship between CA20 scores and tumour quiescence across the TCGA cohort. **(d)** Cox proportional hazards analysis estimates of the log hazards ratio for the impact of CEP89 expression on patient prognosis within individual cancer studies, after adjusting for tumour stage. Patients with high expression of CEP89 show significantly worse prognosis within ACC, LUSC, LIHC, KIRC and STAD, but significantly better prognosis within HNSC,

950 PAAD and KIRP studies. **(e)** Western blot showing depletion of Cep89 protein 48h after siRNA
951 transfection of NCI-H1299 cells. Mock is lipofectamine only, NTC is Non-targeting control
952 siRNA. B-actin is used as a loading control. **(f)** Graphs show that Cep89 depletion in NCI-H1299
953 cells leads to a reduction in nuclear number and an increase in the fraction of quiescent cells,
954 measured by an increase in the percentage of EdU negative (24h EdU pulse) and Phospho-Ser
955 807/811 Rb negative cells. One-Way ANOVA, *$p<0.05$, **$p<0.01$. Mean of n=3.

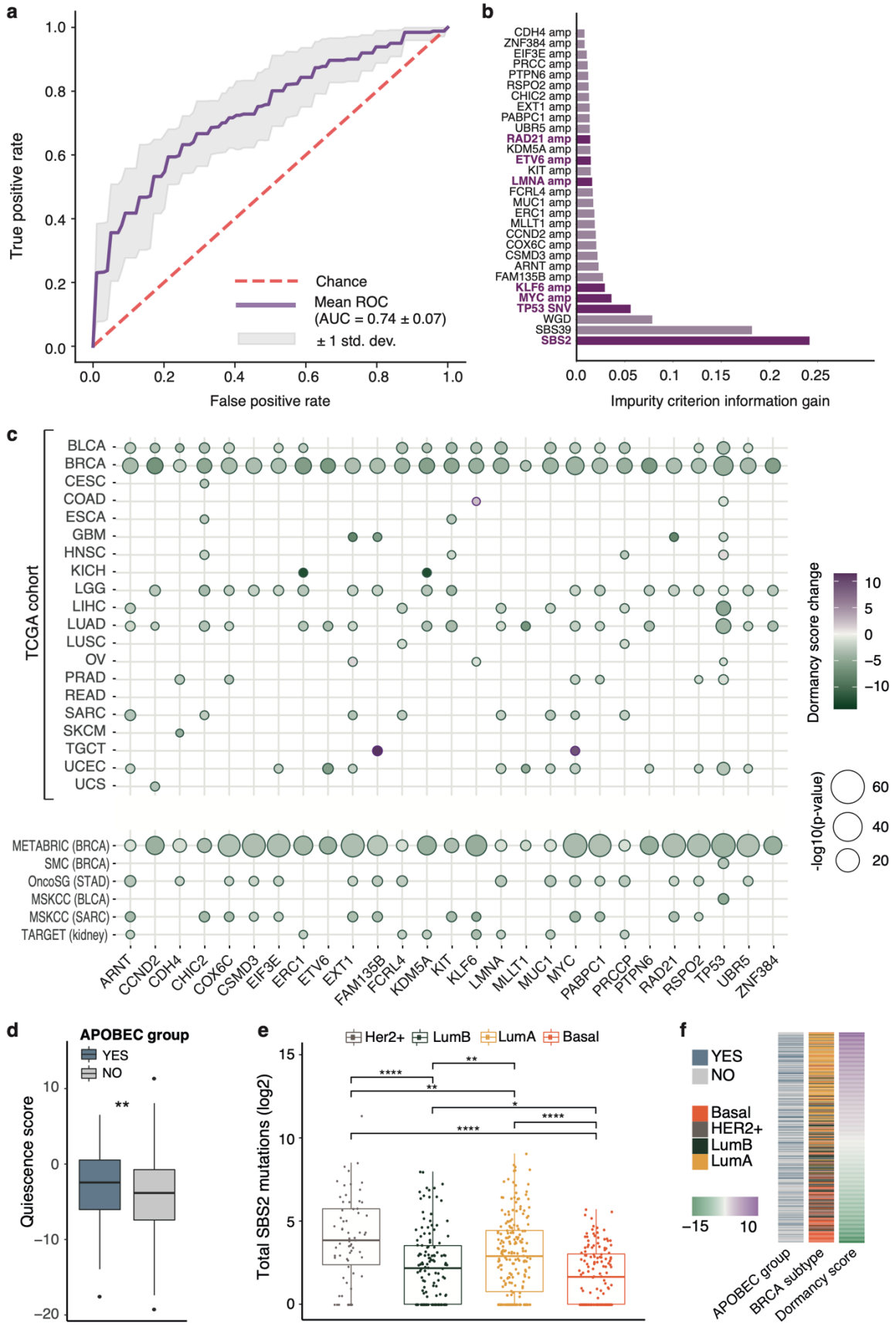


Figure 5: Genomic landscape of breast cancer dormancy. (a) ROC curve illustrating the mean performance of the random forest model for dormancy prediction in breast cancer based on 9

959 randomisations. **(b)** Impurity criterion information gain for the top ranked genomic events
960 predictive of dormancy in the TCGA BRCA cohort, modelled using random forests. Features also
961 appearing in the pan-cancer model are highlighted in dark purple. WGD = whole-genome
962 doubling. **(c)** Tissue-specific changes in dormancy estimates between samples with and without
963 dormancy-associated amplifications (red) and SNVs (purple), identified by the breast cancer
964 specific random forest model, within the TCGA cohort (top panel) and external validation datasets
965 (bottom panel). **(d)** Dormancy scores are significantly higher in APOBEC mutation-enriched breast
966 cancer samples. **(e)** APOBEC-linked SBS2 mutational burden compared between breast cancer
967 subtypes. **(f)** Association between dormancy scores, APOBEC enrichment and breast cancer
968 subtypes in TCGA.
969
970
971
972

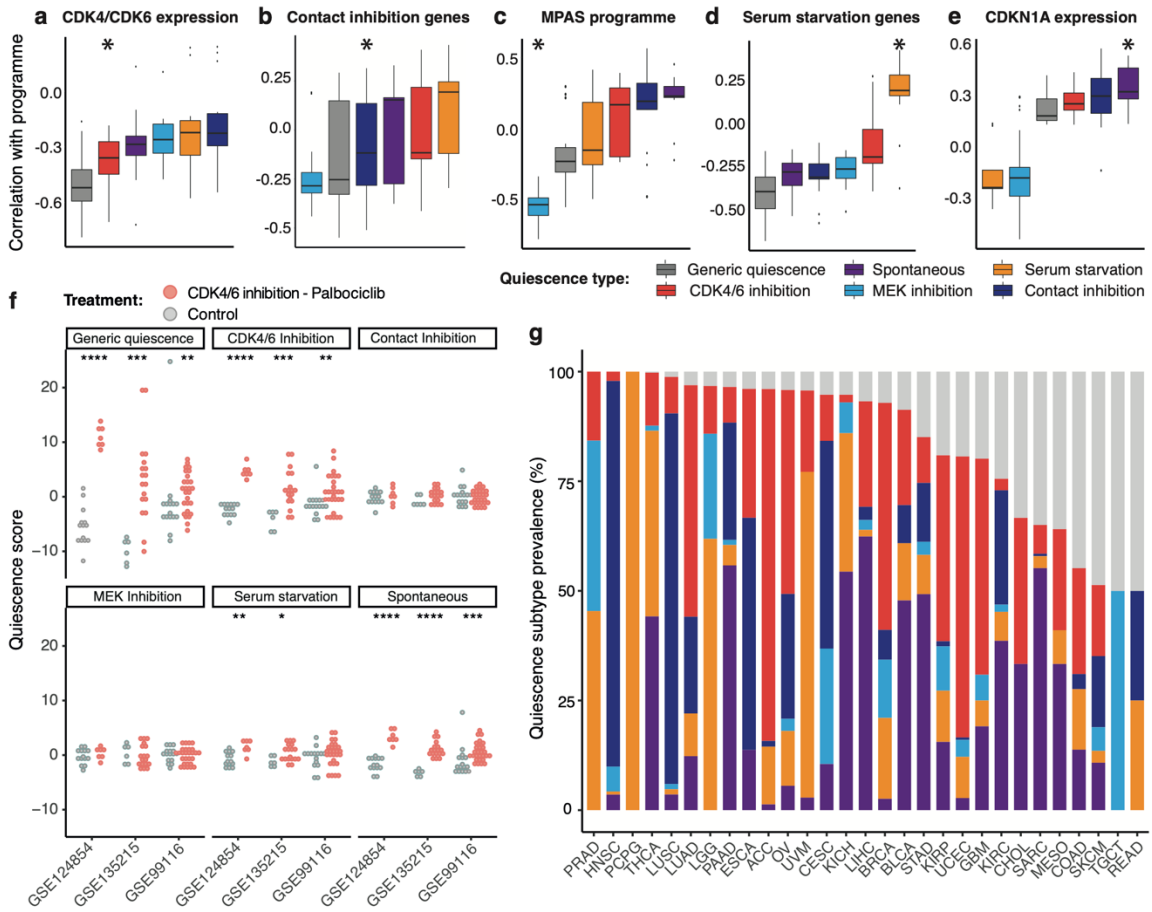
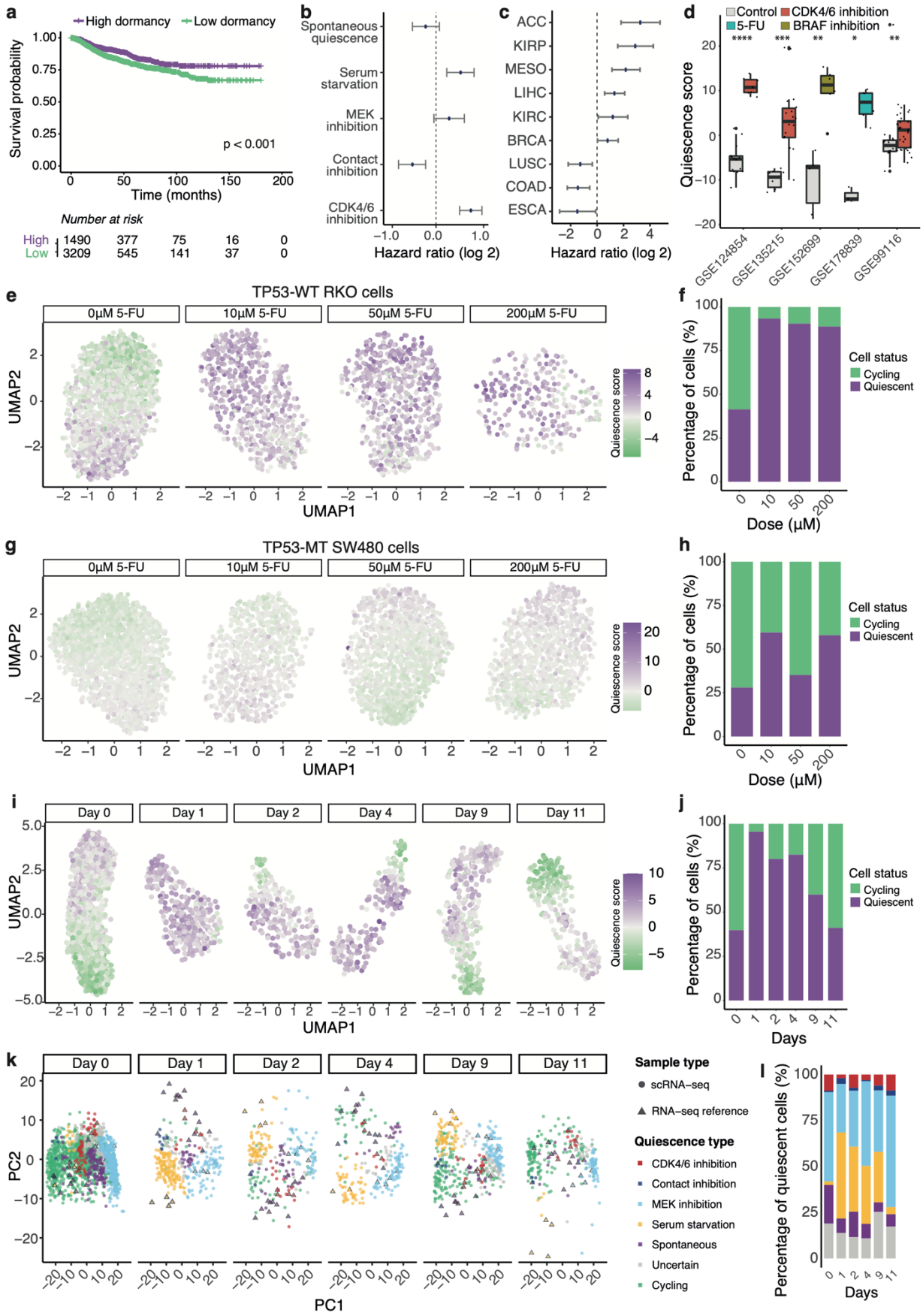


Figure 6: Pan-cancer characterisation of individual quiescence programmes. (a-e) Comparison of correlation coefficients between quiescence programme scores and **(a)** mean expression of CDK4 and CDK6, **(b)** mean expression of curated contact inhibition genes, **(c)** a transcriptional MAPK Pathway Activity Score (MPAS), **(d)** mean expression of curated serum starvation genes, and **(e)** CDKN1A expression, across TCGA cancers. The correlations expected to be strongest (either negative or positive) are denoted by an asterisk. The generic quiescence score refers to scores calculated using the original list of 139 genes differentially expressed across all 5 forms of quiescence. **(f)** Comparison of quiescence programme scores measured in cancer cell lines before (grey) and after (red) Palbociclib treatment across three validation studies. Datasets used for validation are denoted by their corresponding GEO series accession number. **(g)** Predicted quiescence type composition of samples estimated to be dormant across individual cancer types. The same colour legend as in **(a)** is applied. Gray bars represent the proportion of samples for which the quiescence type could not be estimated.



987

988 **Figure 7: Impact of cellular dormancy on patient prognosis and treatment response. (a)**

989 Disease-specific survival based on tumour dormancy for patients from TCGA within 15 years of

990 follow up. Patients with high primary tumour dormancy showed significantly better prognosis than
991 patients with low dormancy. **(b-c)** Hazard ratio ranges illustrating the impact of different forms of
992 quiescence **(b)** and different tissues **(c)** on patient prognosis, after taking into account potential
993 confounding factors. Values above 0 indicate significantly better prognosis in the context of high
994 tumour dormancy. **(d)** Change in quiescence scores inferred from bulk RNA-seq across breast,
995 pancreatic, colorectal and skin cancer cells in response to treatment with the CDK4/6 inhibitor
996 Palbociclib, 5-FU or the BRAF inhibitor Vemurafenib. **(e-f)** UMAP plot illustrating the response of
997 the TP53-proficient RKO colorectal cancer cell line to various 5-FU doses and the corresponding
998 proportions of cells predicted to be quiescent/proliferating. Each dot is an individual cell, coloured
999 according to its dormancy level. **(g-h)** The same as previous, but for the TP53-deficient SW480
1000 cell line. **(i-j)** UMAP plot illustrating the response of individual PC9 NSCLC cells to the EGFR
1001 inhibitor Erlotinib across several time points and the corresponding proportion of cells predicted to
1002 be quiescent/proliferating. **(k)** Principal component analysis illustrating the superimposition of
1003 scRNA-seq profiles (circles) of quiescent NSCLC cells before/after EGFR inhibition onto the bulk
1004 RNA-seq reference data (triangles) for MCF10A cells occupying various quiescence states. **(l)** The
1005 proportion of NSCLC cells in **(k)** predicted to occupy different quiescence states across several
1006 time points.

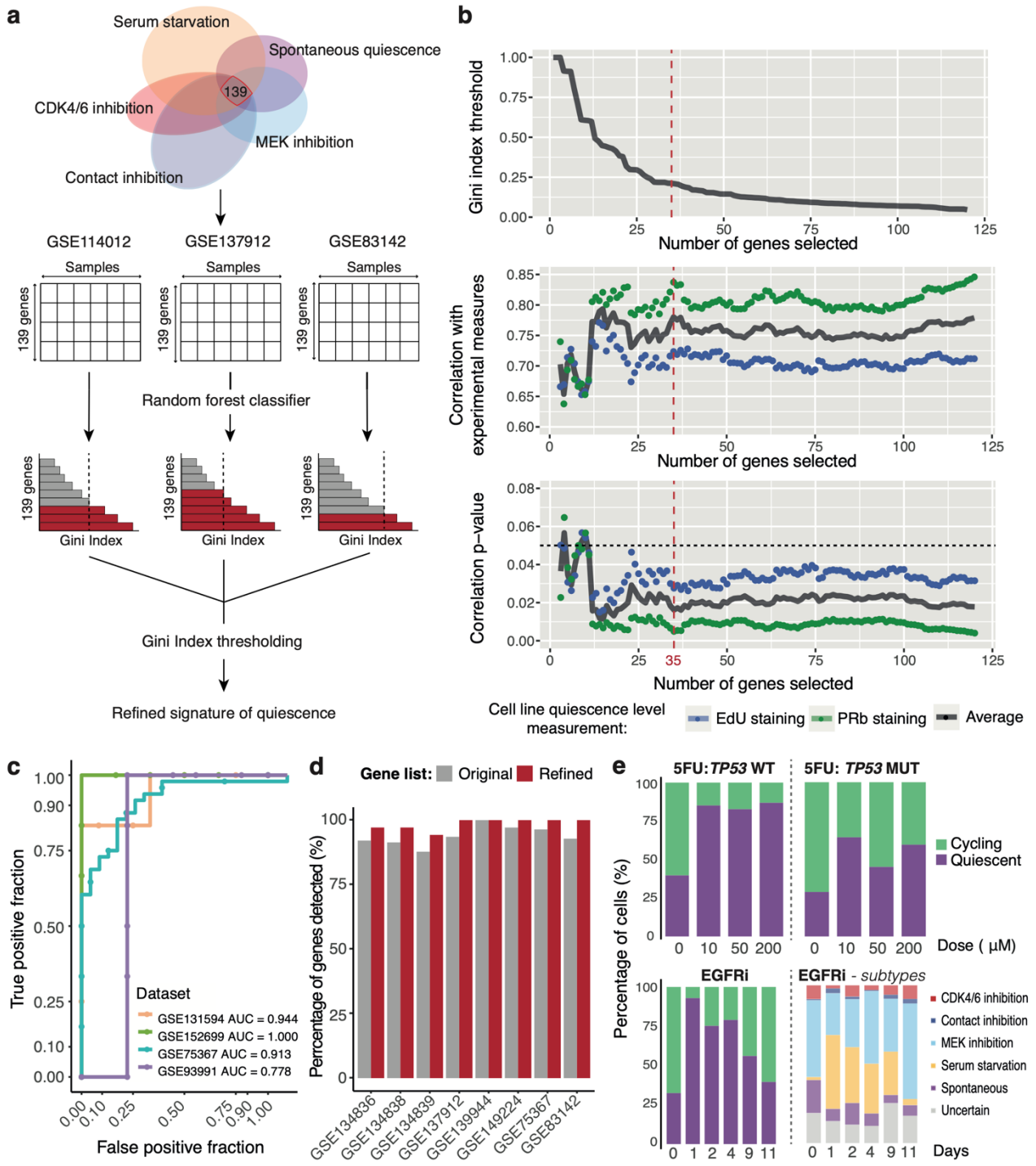
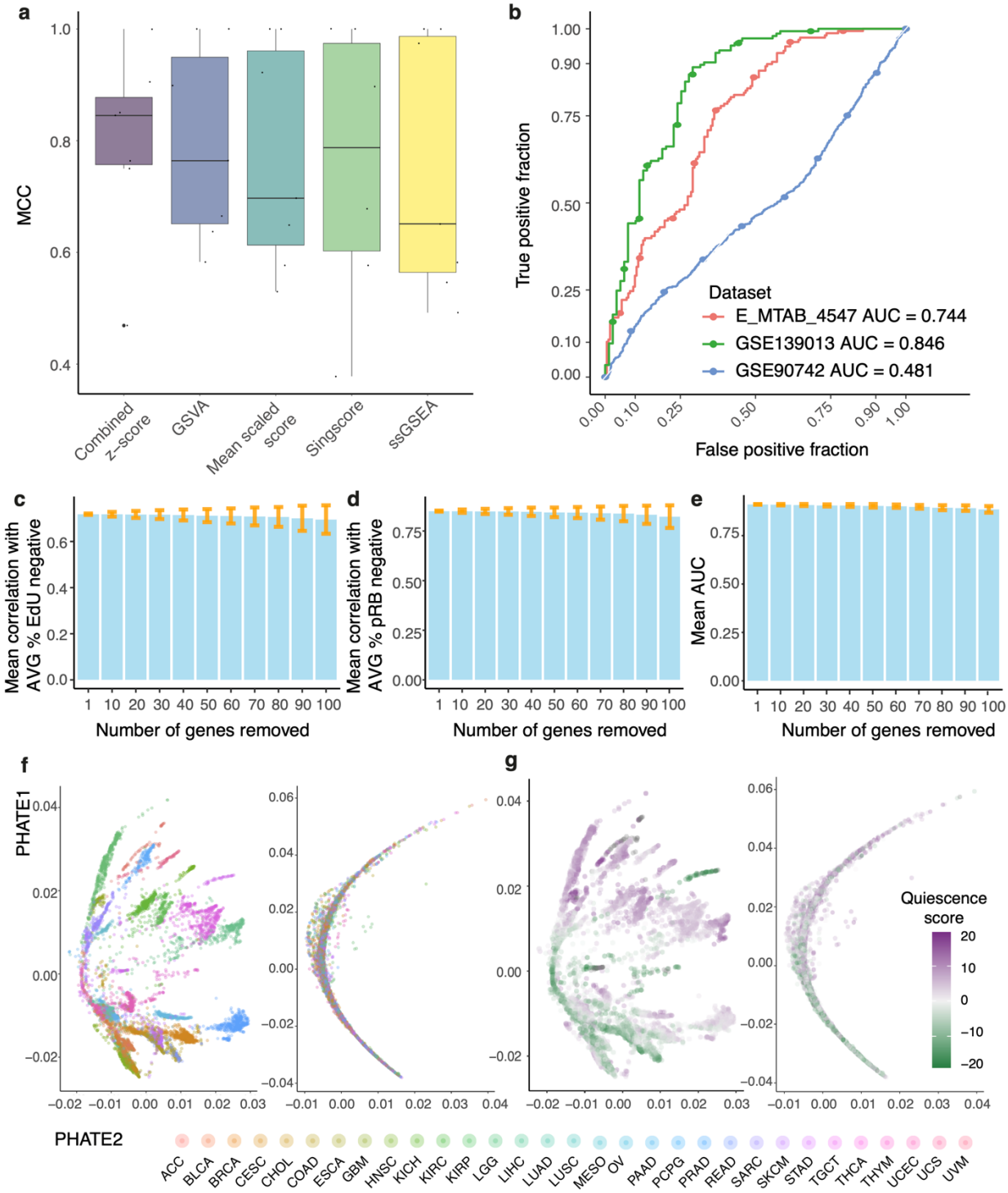


Figure 8: Optimisation of dormancy signature for use in scRNA-seq data. (a) Methodology for refining the gene signature of dormancy: random forest classifiers are trained to distinguish dormant from cycling tumours on three high confidence datasets; Gini index thresholding is optimised to prioritise a final list of 35 genes. **(b)** Gini index variation, correlation with experimentally measured quiescence via EdU and PRb staining assays, and corresponding p-values are plotted as the number of genes considered in the model is increased. The red dotted line

1014 indicates the threshold chosen for the final solution of 35 genes. The black dotted line indicates the
1015 threshold for p-value significance. **(c)** Additional external validation of the 35 gene signature acting
1016 as a classifier of quiescent and proliferation cells in single cell and bulk datasets. **(d)** Dropout in
1017 single cell data by gene signature. The percentage of genes out of the 35 (red) and 139 (grey) gene
1018 lists with reported expression across the single-cell RNA-seq datasets analysed in this study. **(e)**
1019 Proportion of cycling and quiescent cells estimated in single cell datasets of p53 wild type and
1020 mutant lines treated with 5FU, as well as cells treated with EGFR inhibitors. Data as in Figure 7.

1021

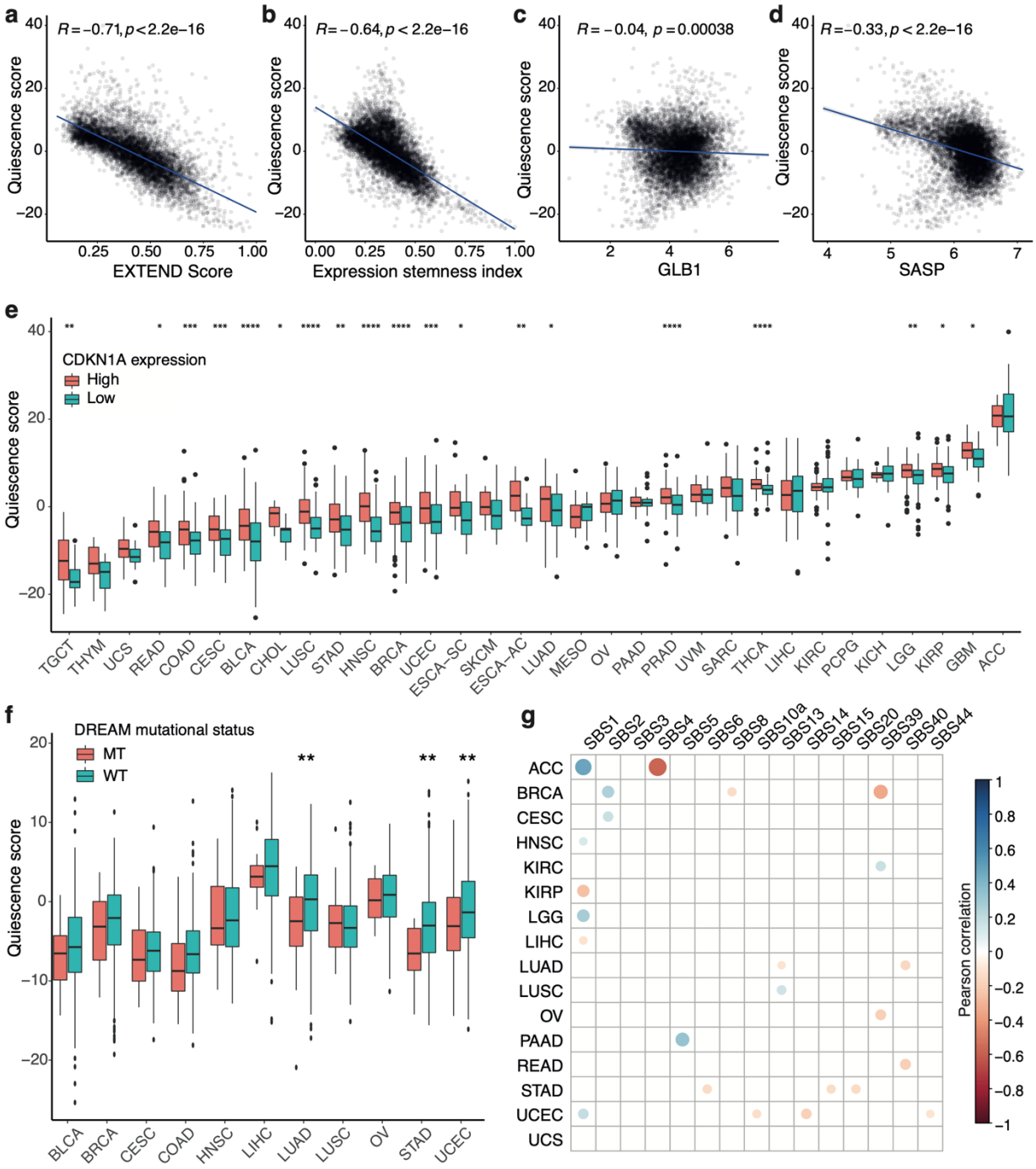
1022 **EXTENDED DATA**



1023

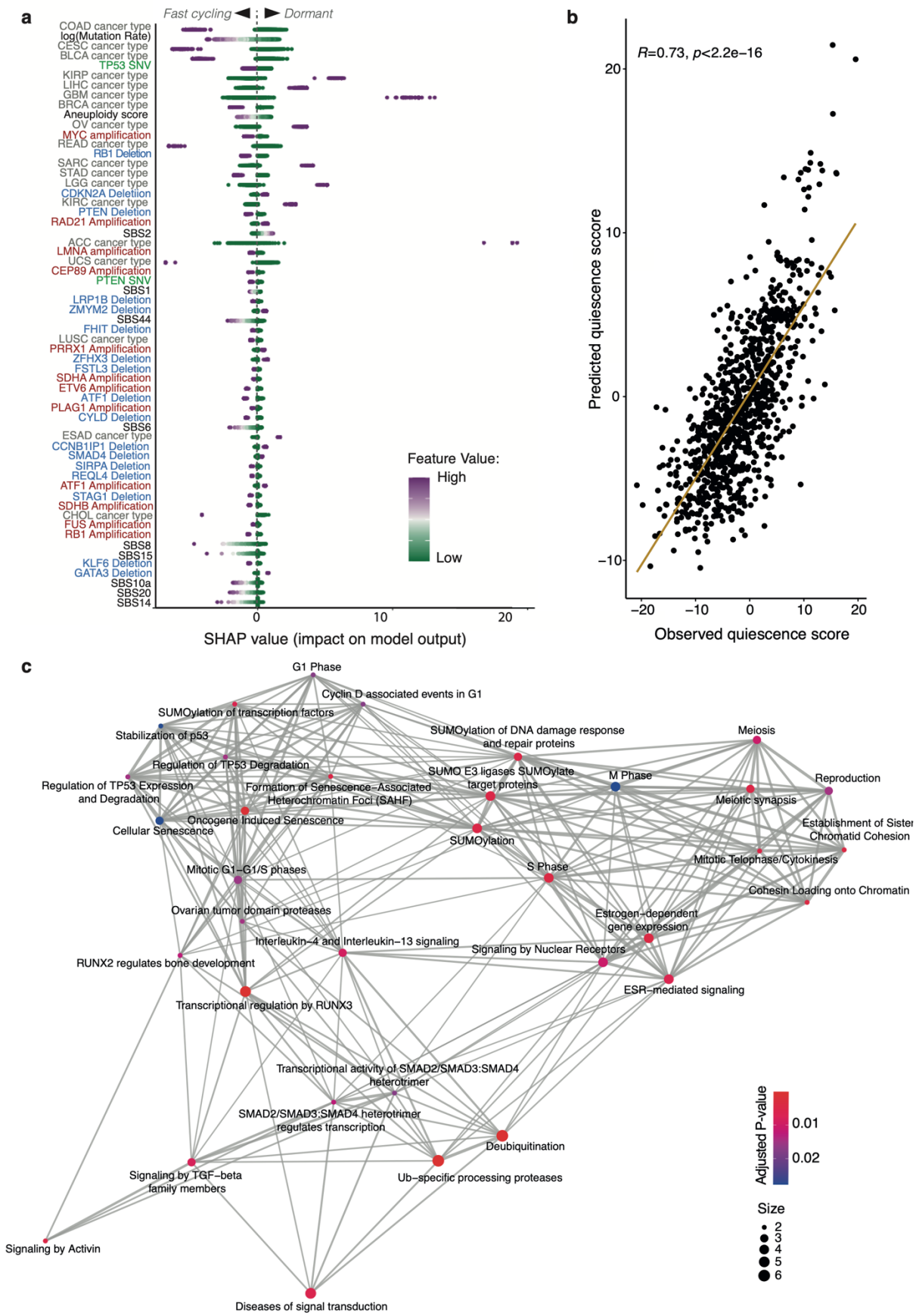
1024 **Extended Data Figure 1: Quiescence evaluation and validation. (a)** Classification accuracies
1025 measured using the Matthews correlation coefficient (MCC) comparing the combined z-score
1026 quiescence scoring method to other scoring methodologies across the seven single-cell and bulk
1027 RNA sequencing validation datasets. **(b)** Receiver operating characteristic (ROC) curves

1028 illustrating the accuracy of the combined z-score methodology when separating dormant and
1029 proliferating hematopoietic stem cells from three publicly available single-cell RNA-seq datasets.
1030 Datasets used are denoted by their corresponding GEO series accession number and the area under
1031 the curve (AUC) for each dataset is specified. **(c)** Mean correlation between computationally
1032 inferred dormancy scores after randomly removing variable number of genes from the dormancy
1033 signature and the fraction of cells entering quiescence in nine lung adenocarcinoma cell lines, as
1034 assessed through an EdU staining assay. **(d)** Same as **(c)** but with experimental levels of quiescence
1035 measured using a phospho-Rb staining assay. (e) Mean classification accuracy across 7 bulk and
1036 single-cell RNA-seq validation datasets after randomly removing variable number of genes from
1037 the dormancy signature. **(f-g)** PHATE plots displaying the first two principal components obtained
1038 from dimensionality reduction analysis across 8,005 primary tumour TCGA samples, based on the
1039 expression of 139 genes differentially expressed across different quiescence programmes before
1040 (left) and after (right) removal of tissue specific expression patterns. Each primary tumour sample
1041 is coloured by their corresponding tissue of origin **(f)** or the corresponding quiescence score **(g)**.



Extended Data Figure 2: Expression and genomic patterns of dormant tumours. (a-d) Correlation between quiescence scores and **(a)** telomerase activity inferred via the "EXTEND" score, **(b)** stemness expression index, **(c)** the expression of the β galactosidase-encoding gene GLB1 and **(d)** the expression of the Senescence-Associated Secretory Phenotype (SASP) gene expression. **(e)** Consistently higher levels of quiescence are detected in samples with high CDKN1A expression (encoding p21, upper quartile of the expression distribution) compared to

1049 samples with low expression (lower quartile of the expression distribution) across TCGA cancer
1050 types. **(f)** Consistently higher levels of dormancy are detected in samples with a functional
1051 DREAM complex. TCGA studies where mutations within one of the DREAM components could
1052 be detected in at least 10 patients are shown. **(g)** Correlation between quiescence levels and
1053 mutational signature contributions across TCGA cancer types. Only significant correlations are
1054 displayed (Pearson correlation p-value < 0.05).

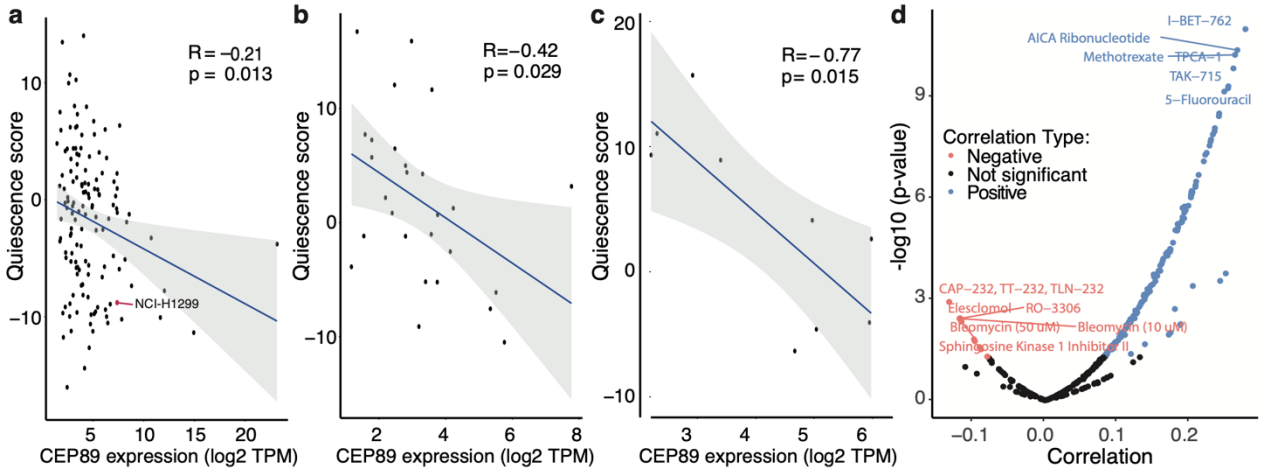


1055

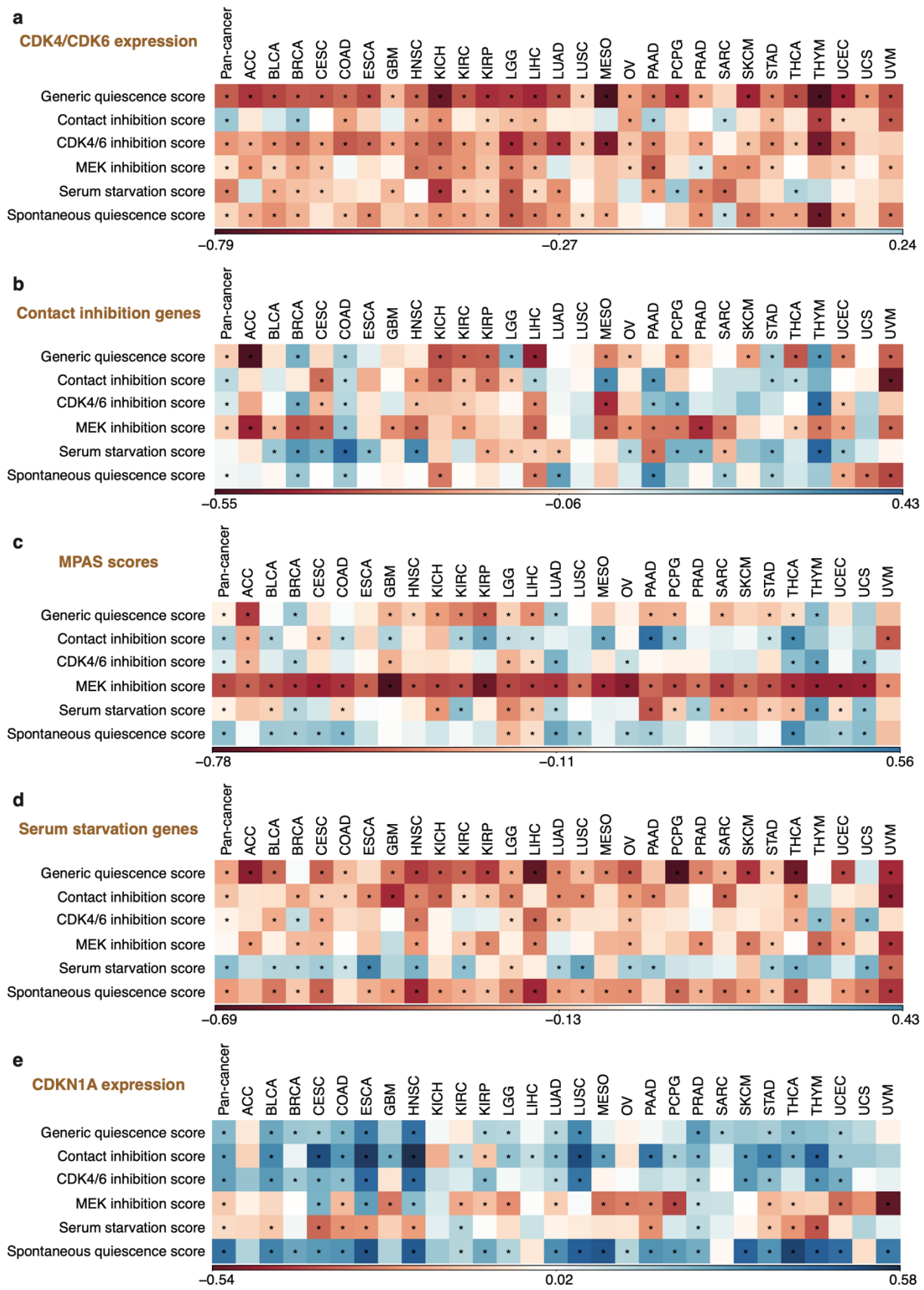
1056 **Extended Data Figure 3: Modelling and validating a pan-cancer classifier of dormancy. (a)**

1057 Genomic features and cancer types significantly associated with dormancy, ranked according to

their importance in the pan-cancer linear regression model (from highest to lowest contribution). The genomic features highlighted are identical to Figure 3c, but cancer types are shown here as well. Each point corresponds to one sample and is coloured according to the original value of the respective feature. For discrete variables purple indicates the presence of the feature and green indicates the absence of the feature. For cancer type variables, purple indicates the patient belongs to the specified TCGA cancer type group. The Shapley values indicate whether the quiescence scores predicted by the ensemble elastic net regression model for each patient are increased or decreased depending on the cancer type of each individual as well as specific SNVS, CNV and other genomic features. **(b)** Correlation between the observed quiescence scores within the test dataset (x axis) and scores predicted by the linear regression model using the selected pan-cancer features (y axis). **(c)** Pathways significantly altered in association with tumour dormancy.

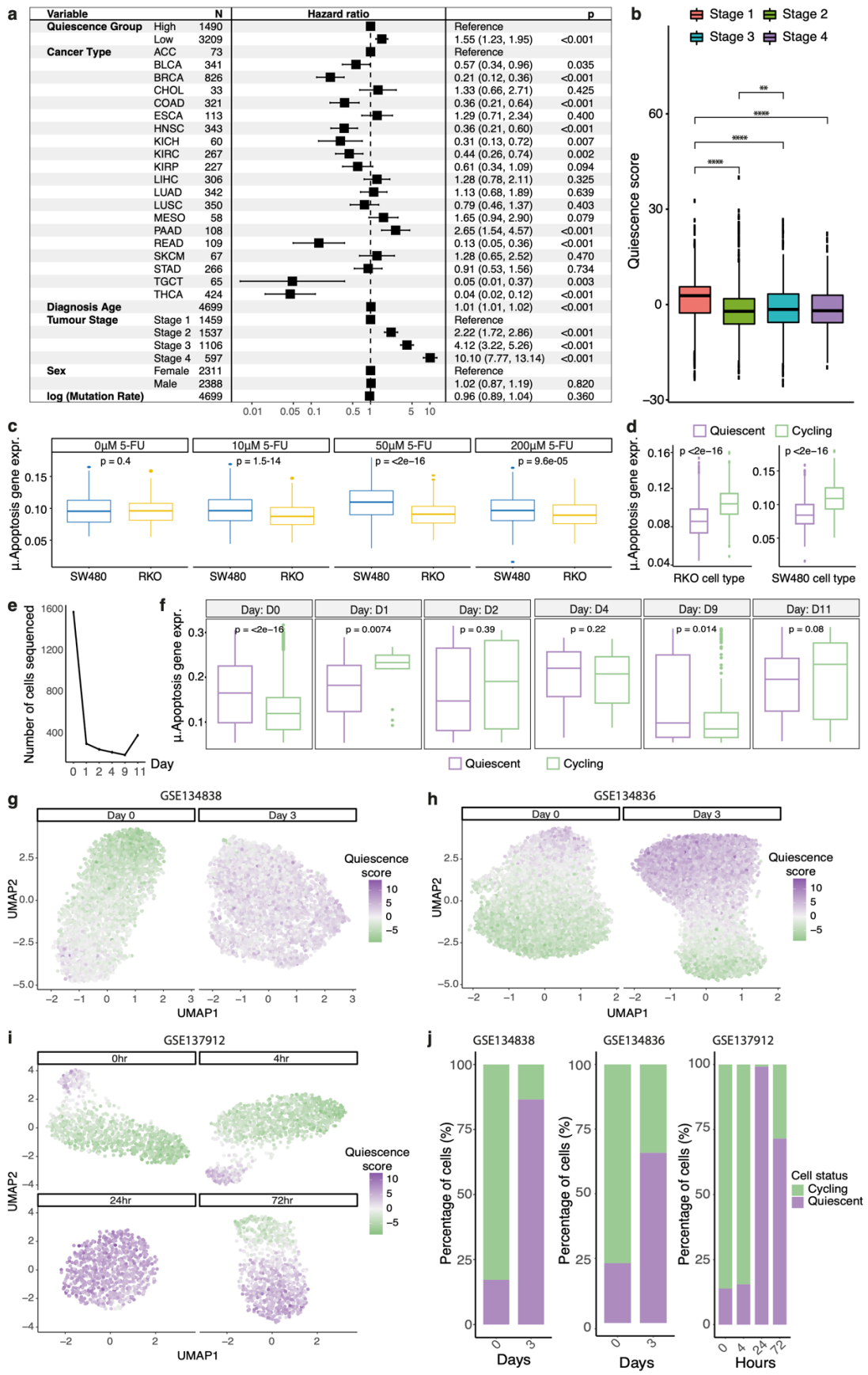


Extended Data Figure 4: CEP89 expression is associated with quiescence and has prognostic value. (a-c) CEP89 expression is negatively correlated with quiescence scores in lung **(a)**, pancreas **(b)** and thyroid **(c)** cancer cell lines. The NCI-H1299 cell line where CEP89 activity was tested is highlighted in red. **(d)** Compounds with selective toxicity on cell lines with high CEP89 expression. The volcano plot shows the Pearson correlation coefficients and the corresponding p-values for associations between CEP89 expression and Area Under the dose-response Curve (AUC) for each compound across cell lines form the CCLE.



Extended Data Figure 5: Quiescence subtype validation. Correlation between subtype-specific quiescence estimates and **(a)** mean expression of CDK4/6, **(b)** the mean expression of curated

1081 contact inhibition genes, **(c)** MPAS scores, **(d)** mean expression of curated serum starvation genes
1082 and **(e)** CDKN1A expression, across individual TCGA cancer studies and pan-cancer. The colour
1083 gradient indicates the strength of the association (positive or negative).

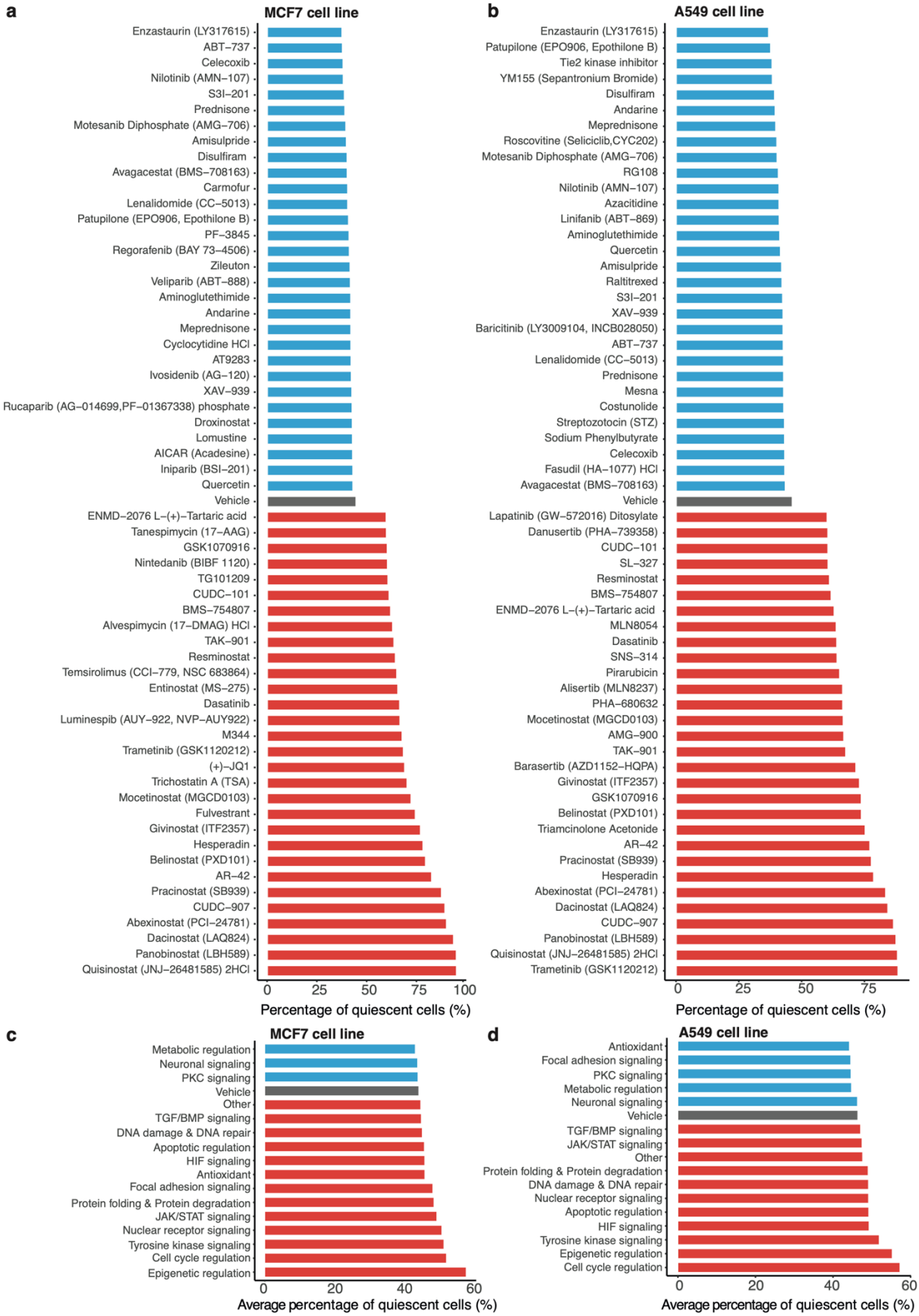


1084

1085 **Extended Data Figure 6: Tumour dormancy relevance to clinical outcome. (a)** Multivariate

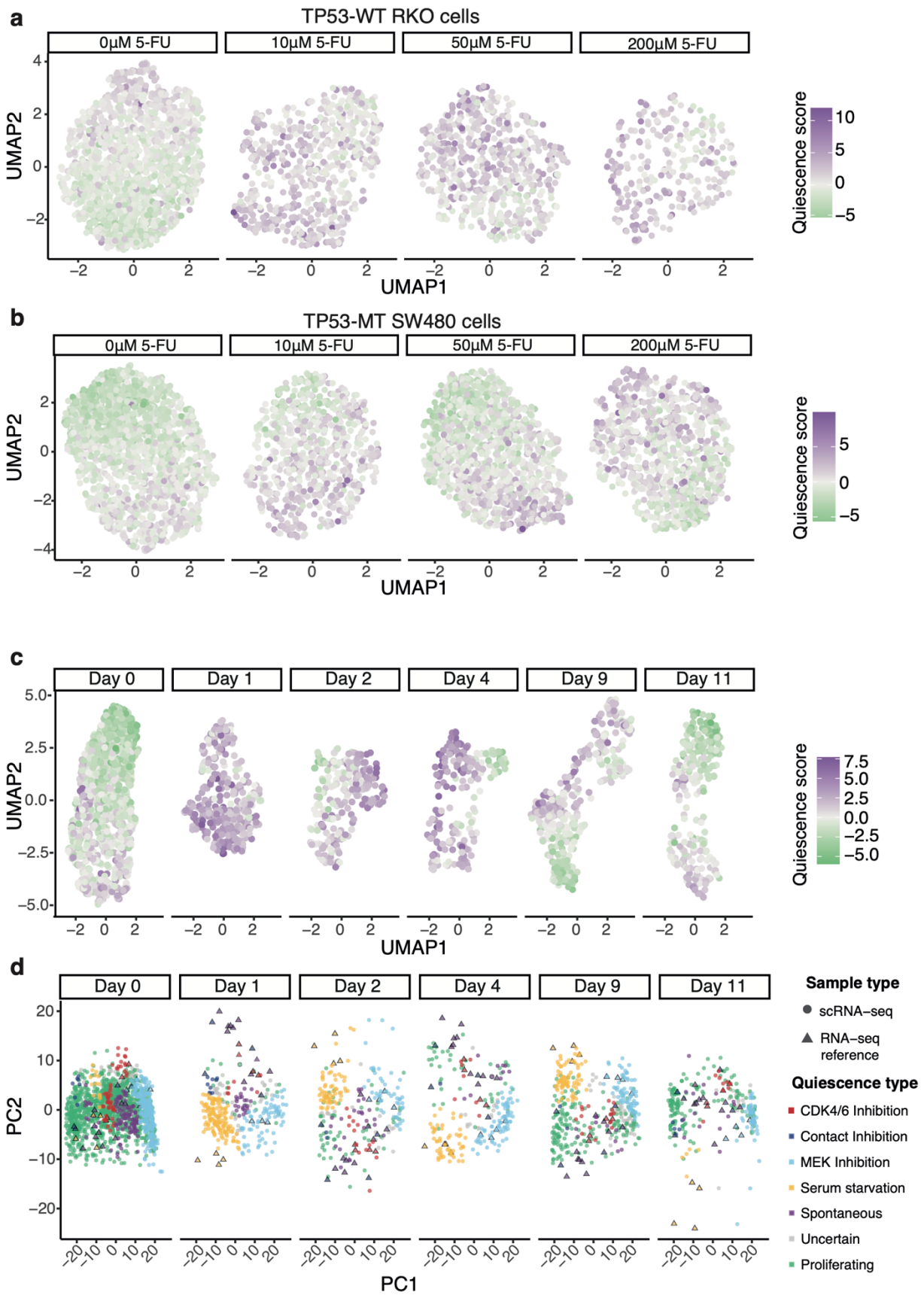
1086 Cox proportional hazards analysis of the disease specific survival in the TCGA cohort, modelled

1087 on tumour dormancy (binarized by optimised survival cut-off into high, where score > 2.95; and
1088 low, where score <= 2.95) and potential cofounding factors. **(b)** Quiescence scores in TCGA
1089 cancers compared across tumour stages. **(c)** The average expression of apoptosis genes compared
1090 between TP53 wild type RKO and TP53 mutant SW480 colorectal cancer cell lines in response to
1091 various 5-FU treatment doses. **(d)** The average expression of apoptosis genes compared between
1092 quiescent and cycling cells within the TP53 WT RKO and TP53 MT SW480 cell lines. **(e)** The
1093 number of NSCLC cells sequenced during the EGFR inhibition treatment time course. **(f)** The
1094 average expression of apoptosis genes compared between quiescent and cycling cells at different
1095 time points during EGFR inhibition treatment. There is a notable increase in apoptotic activity of
1096 cycling cells at day 1. **(g)** Response of melanoma cancer cells to BRAF inhibition treatment across
1097 several time points, illustrated using UMAP plots coloured by quiescence score of individual cells.
1098 **(h)** Response of NSCLC cells to EGFR inhibition via Erlotinib across several time points,
1099 illustrated using UMAP plots coloured by quiescence score of individual cells. **(i)** Response of lung
1100 adenocarcinoma cell lines to KRAS inhibition treatment across several time points, illustrated
1101 using UMAP plots coloured by quiescence score of individual cells. **(j)** Percentage of cells
1102 illustrated in **(g-i)** predicted to be quiescent across the treatment time course.



Extended Data Figure 7: Quiescence dynamics upon various treatment modalities. (a-b) Percentage of quiescent MCF7 **(a)** and A549 **(b)** cells following 24-hour treatment with selected

1106 compounds. Treatments highlighted in blue and red indicate the 30 compounds which showed the
1107 highest decrease and increase in quiescence levels, respectively, in comparison to the control,
1108 shown in grey. **(c-d)** The average percentage of quiescent MCF7 **(c)** and A549 **(d)** cells following a
1109 24-hour treatment with compounds affecting specific pathways. Pathways shown in blue and red
1110 indicate an average decrease and increase in quiescence respectively in comparison to the control,
1111 shown in grey.



Extended Data Figure 8: Application of reduced dormancy signature to scRNA-seq data. (a-c) UMAP plot illustrating the response of **(a)** the TP53-proficient RKO colorectal cancer cell line

1115 to various 5-FU doses, **(b)** the TP53-deficient SW480 colorectal cancer cell line to various 5-FU
1116 doses, and **(c)** the P9 NSCLC cell line to the EGFR inhibitor Erlotinib across several timepoints.
1117 Each dot is an individual cell and it is coloured according to its dormancy level, calculated using
1118 the optimised 35 gene signature. **(d)** Principal component analysis illustrating the superimposition
1119 of scRNA-seq profiles (circles) of quiescent NSCLC cells, inferred using the optimised 35 gene
1120 signature, before and after EGFR inhibition onto the bulk RNA-seq reference data (triangles) for
1121 MCF10A cells occupying various quiescence states. The NSCLC cell line quiescence type was
1122 predicted using a k-nearest neighbour algorithm based on the bulk RNA-seq reference.

1123