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]. Beyond these, 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 state. 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],

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

33 Furthermore, the extent to which dormancy is enacted through transcriptional or genetic control
34 is unknown[5,25], as the mutational processes and genomic changes driving quiescence have not
35 been systematically profiled across cancer tissues. Understanding the evolutionary triggers and
36 molecular mechanisms that enable cancer cells to enter and maintain dormancy would enable
37 us to develop pharmacological strategies to selectively eradicate quiescent cancer cells or
38 prevent them from re-entering proliferative cycles.

39 To address these challenges, we have developed a new method to uniquely and reliably quantify
40 tumour dormancy using transcriptomic data, and employed it to characterise this phenomenon
41 in bulk and single cell datasets from a variety of solid tumours. We identify and validate mutational
42 constraints for dormancy emergence, hinting at potential new therapeutic targets that could
43 exploit this mechanism. We also demonstrate the relevance of dormancy to responses to a range
44 of compounds targeting cell cycle, kinase signalling and epigenetic mechanisms in single cell
45 datasets, and propose an expression signature that could be employed to detect tumour
46 dormancy-induced treatment resistance.

47
48 **RESULTS**

49 **Evaluating tumour dormancy from transcriptomic data**

50 We hypothesised that primary tumours contain varying numbers of quiescent cells, and this in
51 turn may determine their ability to overcome therapies targeting cycling cells. To capture this
52 elusive phenotype, we developed a computational framework that would allow us to quantify
53 dormancy signals in bulk and single cell sequenced cancer samples (Figure 1a). We employed

54 139 differentially expressed genes across five distinct forms of quiescence resulting from serum
55 starvation, contact inhibition, MEK inhibition, CDK4/6 inhibition and spontaneous quiescence, as
56 described by Min and Spencer[18], as markers of a generic quiescence state. The expression
57 levels of these markers were used to derive a score reflecting the relative level of dormancy
58 within individual tumours (see Methods).

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

72 We experimentally validated our methodology in nine lung adenocarcinoma cell lines. There was
73 a remarkably good correlation between our predicted quiescence levels and the fraction of
74 quiescent cells, estimated using quantitative, single-cell imaging of phospho-Ser807/811-Rb
75 (PRb, which labels proliferative cells[38]) and 24 hour EdU proliferation assays (Figure 1e-h). Cells
76 that were negative for either PRb or EdU were defined as quiescent (see Methods, Figure 1e-f).
77 Importantly, a G1 signature (Methods) was not associated with these experimental
78 measurements, proving our method indeed captures quiescence rather than a prolonged G1
79 state (Figure 1g-h). The quiescence correlations appeared robust to random removal of individual
80 genes from the signature, with no single gene having an inordinate impact on the score

81 (Supplementary Figure 1c-e). This provided further reassurance that our Z-score based
82 methodology is successful in capturing dormancy signals from bulk tumour data.

83 Having established a robust framework for quantifying tumour dormancy, we next profiled 8,005
84 primary tumour samples across 31 solid cancer tissues from The Cancer Genome Atlas (TCGA).
85 After accounting for potential confounding signals from the microenvironment by correcting for
86 tumour purity (see Methods, Supplementary Figure 1f-g), we observed an entire spectrum of
87 dormancy across tumours that ranged from highly dormant to highly proliferative (Figure 2a).
88 Cancers well known to be frequently dormant, such as glioblastoma[3,30], were amongst the
89 highest ranked in terms of quiescence, along with kidney and adrenocortical carcinomas (Figure
90 2b).

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

98 Lastly, we confirmed dependencies on the p53/p21/DREAM activation axis: tumours that were
99 proficient in *TP53* or the components of the DREAM complex, as well as those with higher p21
100 expression, had elevated dormancy levels across numerous tissues (Figure 2e, Supplementary
101 Figures 2e-f), although only 8 out of 139 dormancy genes are directly transcriptionally regulated
102 by p53[45]. Nevertheless, p53 proficiency appears to be a non-obligatory dependency of dormancy,
103 which is also observed to arise in p53 mutant scenarios in 21% of cases. p53 has also been
104 shown to play a role in preventing the occurrence of larger structural events and polyploidy[46-48],
105 potentially explaining the lower dormancy levels we observed in tumours that had undergone
106 whole genome duplication (Figure 2f).

107
108 **The genomic background of cancer dormancy**

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

120 Following the initial prioritisation of putative genomic triggers of dormancy, we sought to identify
121 those that could best distinguish dormant from fast proliferating tumours, while accounting for
122 tissue effects. An ensemble elastic net selection approach similar to the one described by Pich
123 et al[51] was applied for this purpose (Figure 3b, Methods). Our pan-cancer model identified tissue
124 type to be a major determinant of quiescence levels (Supplementary Figure 3a). It also uncovered
125 a reduced set of 60 genomic events linked with tumour dormancy levels, including SNVs and
126 copy number losses in 17 cancer genes, as well as amplifications of 13 cancer genes (Figure
127 3c). These events could then be successfully employed to predict quiescence in a separate test
128 dataset, thus internally validating our model (Supplementary Figure 3b).

129 Reassuringly, a functional *TP53*, lack of *MYC* amplification and lower mutation rates (Figure 3c)
130 were amongst the top ranked characteristics of dormant tumours, which also displayed less
131 aneuploidy. Additionally, we uncovered novel dependencies, such as *CEP89* and *LMNA*
132 amplifications observed in fast cycling tumours, or *ZMYM2* deletions prevalent in dormant
133 samples. p16 (*CDKN2A*) deletions, one of the frequent early events during cancer evolution[52,53],

134 were enriched in dormant tumours. *RB1* deletions and amplifications were both associated with
135 a reduction in tumour dormancy, which might reflect the dual role of RB1 in regulating proliferation
136 and apoptosis[54]. Overall, the genomic dependencies of dormancy mainly comprised genes
137 involved in cell cycle pathways, p53 regulation and ubiquitination (most likely of cell cycle
138 targets), TGF-β signalling mediators and RUNX2/3 regulation which have previously been shown
139 to play a role in controlling proliferation and cell cycle entry (Supplementary Figure 3c).

140 Our model also calls attention to the broader mutational processes associated with this cellular
141 state. Such processes showed fairly weak and heterogeneous correlations with dormancy within
142 individual cancer tissues (Supplementary Figure 2g), but their contribution becomes substantially
143 clearer pan-cancer once other genomic sources are accounted for. In particular, we identified an
144 association between dormancy and mutagenesis induced by the AID/APOBEC family of cytosine
145 deaminases as denoted by signature SBS2[49] (Figure 3c). As highlighted by Mas-Ponte and
146 Supek[55], APOBEC/AID driven mutations are likely to have deleterious functional consequences
147 as they tend to be directed towards early-replicating, gene-rich regions of the genome. The
148 authors reported an enrichment of APOBEC-generated mutations in several genes including
149 *ZMYM2*, which our pan-cancer model has linked with higher dormancy.

150 In turn, defective DNA mismatch repair, as evidenced by signatures SBS44, SBS20, SBS15,
151 SBS14 and SBS6[49], was enriched in fast cycling tumours (Figure 3c). Mismatch repair
152 deficiencies lead to hypermutation of single nucleotides and short tandem repeats due to
153 unrepaired DNA polymerase slippage in a phenomenon termed “microsatellite instability” (MSI).
154 Cancers that are particularly prone to MSI include colon, stomach and endometrial carcinomas[56],
155 where we could indeed observe differences in quiescence levels based on the MSI status (Figure
156 3d). Furthermore, dormant tumours were depleted of alterations across all DNA damage repair
157 pathways (Figure 3e).

158 Our measurements of dormancy also reflected expected cycling patterns across 27 MCF7
159 strains[57]: cell lines with longer doubling times exhibited higher quiescence (Figure 3f). This also

160 coincided with a depletion of *PTEN* mutations, a dependency highlighted by the pan-cancer
161 model.

162 When checking for dependencies in individual cancer tissues, 27 out of the 28 genes identified
163 by the model were significantly associated with dormancy in at least one tissue, most prominently
164 in breast, lung and liver cancers which also represent the largest studies within TCGA (Figure
165 3g, top panel). Most of these genomic insults were linked with a decrease in dormancy. In
166 external validation datasets these associations, including deletions in *PTEN* and *LRP1B* or
167 amplifications of *MYC, CEP89* or *STAG1,* featured most prominently in the largest cohort of
168 breast cancer samples (Figure 3g, bottom panel).

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

170 To gain more insight into the underlying biology of cellular dormancy, we sought to experimentally
171 validate associations highlighted by the pan-cancer model. We focused on the impact of *CEP89*
172 activity on quiescence-proliferation decisions due to the high ranking of this putative oncogene
173 in the model, the relatively unexplored links between *CEP89* and cell cycle control, as well as its
174 association with quiescence across a variety of cancer cell lines (Supplementary Figure 4a-c).
175 The function of CEP89 is not well characterised, however, the encoded protein has been
176 proposed to function as a centrosomal-associated protein[58,59]. Centrosomes function as major
177 microtubule-organising centres in cells, playing a key role in mitotic spindle assembly[60] as well
178 as in the mitotic entry checkpoint by providing a platform for concentrating and amplifying signals
179 of cell cycle proteins such as cyclin B/Cdk1[61]. Moreover, centrosomes function as sites of
180 ubiquitin-mediated proteolysis of cell cycle targets[62] and members of several growth signalling
181 pathways, such as Wnt and NF-kB, have been shown to localise at these structures[63,64]. Several
182 genetic interactions have also been reported between CEP89 and key cell cycle proteins,
183 including cyclin D2[65] (Figure 4a).

184 Our model linked *CEP89* amplification with fast cycling tumours (Figure 3c). Centrosome
185 amplification is a common feature of tumours with high proliferation rates and high genomic

186 instability[66]. This is reflected by the strong negative correlation we found between quiescence
187 levels and a previously reported centrosome amplification transcriptomic signature (CA20) in
188 TCGA (Figure 4b). Previous studies have shown that overexpression of centrosomal proteins
189 can lead to alterations in centriole structure[67,68] and we observed significantly higher CA20 scores
190 in tumours containing *CEP89* amplifications (Figure 4c). Furthermore, *CEP89* expression was
191 prognostic across multiple cancer tissues (Figure 4d) and linked with toxicity of several cancer
192 compounds in cell line models (Supplementary Figure 4d).

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

198 **Dormancy in breast cancer**

199 Due to the variation of dormancy levels across tissues and to increase power, we have taken a
200 pan-cancer approach to capture genomic features that are universally associated with dormancy.
201 However, we acknowledge that the extremely diverse genomic background for cancer
202 development in different tissues might mean that certain genetic alterations may lead to higher
203 or lower propensity of dormancy in a single tissue only. To investigate this further we focused on
204 breast cancer, the largest cancer dataset available from TCGA, where we still have adequate
205 power and balance of dormant/cycling phenotypes to uncover tissue-specific genomic markers
206 of dormancy. A combined ANOVA and random forest classification approach identified 30
207 genomic events that could distinguish dormant from fast cycling tumours with a mean accuracy
208 of 74% (Figure 5a). This model recovered the associations with the APOBEC mutational
209 signature SBS2, *TP53* mutation as well as *MYC*, *KLF6*, *LMNA*, *ETV6* and *RAD21* amplification
210 already seen in the pan-cancer model (Figure 5b). All additional associations validated in the
211 METABRIC cohort and were also seen in several other cancers, particularly bladder, lung and
212 lower grade glioma (Figure 5c).

213 Since APOBEC/AID activity has been linked with dormancy in both the pan-cancer and the
214 tissue-specific models, we further investigated its role in breast cancer, where the associated
215 mutation signature SBS2 is relatively prevalent. A significant correlation (Pearson R = 0.27, p =
216 3.5e-07) was observed between the fraction of mutations contributed by the SBS2 process and
217 dormancy in this cohort. Breast cancer samples enriched in APOBEC-associated mutations,
218 defined as mutations attributed to SBS2 and SBS13, presented significantly higher quiescence
219 levels (Figure 5d). Across breast cancer subtypes, the Luminal A subtype showed the highest
220 levels of dormancy, however, Her2+ tumours had the highest incidence of APOBEC associated
221 mutational events, suggesting this subtype may be the most susceptible to APOBEC-linked
222 quiescence (Figure 5e-f).

223 **Characterisation of individual quiescence programmes**

224 While we had previously examined a generic dormancy programme, cancer cells can enter
225 quiescence due to different stimuli[1] and this may inform its aetiology and manifestation. To
226 explore this, we re-scored tumours based on gene expression programmes specific to serum
227 starvation, contact inhibition, MEK inhibition, CDK4/6 inhibition or spontaneously occurring
228 quiescence as defined by Min and Spencer[18] (see Methods). We observed a good correlation
229 between the dormancy estimates representing individual quiescence programmes and the
230 expression of genes associated with the corresponding form of quiescence in the literature
231 (Figure 6a-e, Supplementary Figure 5). CDK4/6 inhibition-induced quiescence levels were further
232 validated using external RNA-seq datasets from cancer cell lines and xenograft mice sequenced
233 before and after treatment with the CDK4/6 inhibitor Palbociclib[69,70] (Figure 6f, Supplementary
234 Table 4). Interestingly, we also observed significant differences in spontaneous quiescence
235 scores before and after treatment. Indeed, p21 activity has been linked with the Palbociclib
236 mechanism of action[71,72], and this analysis suggests potential similarities between CDK4/6
237 inhibition and p21-dependent quiescence phenotypes.

238 Having validated our framework for quantifying stimulus-specific quiescence programmes, we
239 proceeded to estimate the dominant form of quiescence in different cancer types (Figure 6g). We

240 found a range of quiescence aetiologies across most tissues, while a minority of cancers were
241 dominated by a single form of quiescence, e.g. serum starvation in all dormant
242 pheochromocytomas and paragangliomas, contact inhibition in 88% of head and neck
243 carcinomas and CDK4/6 inhibition in 80% of dormant adrenocortical carcinomas. While we do
244 not wish to claim that the dormancy state will have necessarily been induced by the actual
245 predicted stimulus (impossible in the case of CDK4/6 or MEK inhibition, as the analysed samples
246 are all treatment-naïve), we suggest that the downstream signalling cascade may resemble that
247 triggered by such stimuli, e.g. via CDK4/6 or MEK loss of function mutations. Amongst these
248 states, spontaneous quiescence appeared most strongly dependent on p53 functionality
249 (Supplementary Table 3). This points to common transcriptional features of drug-tolerant
250 dormant cells in different cancer settings that could be employed in designing ways to eradicate
251 these cells in the future.

252 **Clinical relevance of tumour dormancy**

253 Overall, dormancy appears to be beneficial for the long-term outcome of cancer patients (Figure
254 7a, Supplementary Figure 6a). As expected, quiescence levels were increased in stage 1
255 tumours, although later stages also exhibited occasional elevated dormancy (Supplementary
256 Figure 6b). Indeed, such dormant, indolent tumours could have higher chances of being
257 eradicated earlier in the disease, which is consistent with reported worse prognosis of patients
258 with higher tumour cell proliferation rates and less stable, more mutated genomes[73]. However,
259 outcomes do vary depending on the quiescence subtype, with worse survival observed upon
260 contact inhibition (Figure 7b). The outcomes also vary by tissue, with lung, colon or esophageal
261 carcinoma patients displaying significantly worse prognosis in the context of dormancy (Figure
262 7c).

263 While dormancy may confer an overall survival advantage in most cancers, it can also provide a
264 pool of cells that are capable of developing resistance to therapy, as demonstrated in several
265 studies[34,74]. Using our methodology, we indeed observed an increase in quiescence levels in cell
266 lines following treatment with EGFR, BRAF and CDK4/6 inhibitors, as well as conventionally used

267 chemotherapies such as 5-Fluorouracil (5-FU) in multiple bulk RNA-seq datasets (Figure 7d).
268 Furthermore, the recent widespread availability of single-cell transcriptomics offers the
269 opportunity to investigate the impact of quiescence on such therapies with much greater
270 granularity than is allowed by bulk data. Using our quiescence signature and single-cell data from
271 RKO and SW480 colon cancer cell lines treated with 5-FU[75], we could observe quiescence-
272 proliferation decisions following conventional chemotherapy treatment. Within the p53 proficient
273 cell line RKO, the fraction of quiescent cells increased from 41% to 93% after treatment with a
274 low dose (10 μM) of 5-FU and persisted at higher doses (Figure 7e-f). In contrast, a comparable
275 increase was not observed in *TP53* mutant SW480 cells, further emphasizing the key role of p53
276 as a regulator of quiescence (Figure 7g-h). SW480 cells showed higher apoptotic activity
277 following treatment compared to RKO cells, particularly within actively cycling cells, further
278 corroborating that cells capable of entering quiescence may be less vulnerable to this therapy
279 (Supplementary Figure 6c-d).

280 Similarly, using single cell data from an *EGFR* mutant Non-Small Cell Lung Cancer (NSCLC) cell
281 line treated with the EGFR inhibitor Erlotinib[76], we predicted that 40% of cells were likely to exist
282 in a dormant state prior to treatment. EGFR inhibition led to a massive decrease in cell numbers
283 immediately after treatment, mostly due to proliferating cells dying off (Supplementary Figures
284 6e-f), while the proportion of quiescent cells increased to 96% at day 1, indicating an immediate
285 selective advantage for such cells (Figure 7i-j). Interestingly, these cells appear to gradually start
286 proliferating again in the following days during continuous treatment, with the percentage of
287 proliferating cells approaching pre-treatment levels by day 11 (Figure 7j). The same trend
288 captured by our signature could be observed upon KRAS and BRAF inhibition in different cell
289 line models (Supplementary Figure 6g-j, Supplementary Table 4)[34,76]. Furthermore, during the
290 first days of treatment the NSCLC cells that survived EGFR inhibition appeared to reside in a
291 state most resembling that induced by serum starvation (Figures 7k-l). Both EGFR kinase
292 inhibitors and serum starvation have been shown to trigger autophagy[77], which may explain the
293 convergence between this inhibitory trigger and the type of quiescence response. At day 11 most

294 of the remaining dormant cells appeared in a state similar to that preceding the treatment (Figure
295 7l).

296 Thus, quiescence appears to explain resistance to broad acting chemotherapy agents as well as
297 targeted molecular inhibitors of the Ras/MAPK signalling pathway, being either selected for, or
298 induced immediately upon treatment, and gradually waning over time as cells start re-entering
299 the cell cycle. Using massively multiplexed chemical transcriptomic data, we also analysed
300 responses to 188 small molecule inhibitors in cell lines at single-cell resolution[78] (Supplementary
301 Figure 7). We observed a large increase in quiescence following treatment with compounds
302 targeting cell cycle regulation and tyrosine kinase signalling, consistent with our previous results,
303 but also for compounds modulating epigenetic regulation, e.g. histone deacetylase inhibitors -
304 thus highlighting the broad relevance of quiescence.

305 These analyses showcase our ability to probe the nature of quiescent phenotypes using a defined
306 dormancy signature, which could aid the development of methods to selectively target quiescent
307 drug-resistant persister cells. However, a major challenge of scRNA-seq data analysis is the high
308 percentage of gene dropout, which could impact our ability to evaluate dormancy using the full
309 139 gene signature. The scRNA-seq datasets we employed to investigate responses to common
310 therapeutic agents exhibited an average drop-out of 10.6 genes out of the full gene signature.
311 While our scoring method remains robust to such levels of dropout (Supplementary Figure 1c-
312 e), we also employed machine learning to reduce our initial list of 139 markers of dormancy to a
313 robust 35-gene signature, comprised mainly of RNA metabolism and splicing-regulating factors,
314 but also of genes involved in cell cycle progression, ageing and senescence, which could be
315 applied to sparser datasets with larger levels of gene dropout (Methods, Figure 8a-b). The
316 optimised signature of dormancy performed similarly to the initial broadly defined programme in
317 distinguishing highly dormant and fast cycling tumours (Figure 8c), showed no dropout across
318 the scRNAseq datasets used in this study (Figure 8d), was similarly prognostic (p=0.006) and
319 showed comparable profiles of resistance to treatment (Figure 8e, Supplementary Figure 8). This

320 minimal expression signature could be employed to track and further study emerging
321 quiescence-induced resistance in a variety of therapeutic scenarios.

322
323 **DISCUSSION**

324 Despite its crucial role in cancer progression and resistance to therapies, dormancy has
325 remained poorly characterised due to the scarcity of suitable models and biomarkers for large-
326 scale tracking in the tissue or blood. Lack of proliferative markers such as Ki67 or CDK2[21,79] does
327 not uniquely distinguish quiescence from other cell cycle phases, e.g. G1 or S. This study
328 overcame this limitation by employing genes active in different forms of quiescence whose
329 patterns of expression reflect this specific state and are distinct from markers of a longer G1
330 phase, senescence or stemness. We demonstrate a robust method of profiling them to identify
331 dormancy in bulk and single cell RNA-seq data. Since our list of quiescence biomarkers is defined
332 using expression data from human epithelial cells, our approach could be limited in terms of
333 applicability to other cell types. However, carcinomas originating from epithelial tissues make up
334 85% of all malignancies. Moreover, our approach had high classification accuracy in a variety of
335 datasets encompassing solid as well as haematological cancers, thus demonstrating the
336 versatility of our method. More variable performance was observed when applied to
337 hematopoietic stem cells as it was not designed to capture signals in this context (Supplementary
338 Figure 1b).

339 We show that dormancy is pervasive across different solid cancers and generally associated with
340 more stable, less mutated genomes with intact DNA damage repair pathways. We also find a link
341 between APOBEC mutagenesis and higher levels of dormancy. We identify mutational events
342 affecting a variety of genes such as *PTEN*, *CEP89, CYLD*, *LMNA* that appear unfavourable to
343 dormancy development, thus potentially implicating them in influencing quiescence-proliferation
344 decisions. Among these, we propose *CEP89* amplification as novel biomarker of non-small cell
345 lung primary tumours with low quiescence capacity. Genomic alterations enriched in dormant
346 tumours, such as p16 or *ZMYM2* deletions, were more rare and predominantly captured pan-

347 cancer. Such targets should be further validated and could be exploited to either counteract
348 dormancy or induce it.

349 Our finding that primary tumour dormancy confers a favourable prognosis could raise a potential
350 limitation of our approach in distinguishing between slow-dividing and persister cell phenotypes[6]
351 in bulk transcriptome data. Nevertheless, our methodology is directly extendable to single cell
352 data, where we demonstrate a clear association between existent/emerging quiescence and
353 resistance to multiple cell cycle, kinase signalling and epigenetic targeting regimens in a variety
354 of cancers. Quiescent cells utilise a variety of mechanisms to achieve drug resistance, including
355 upregulation of stress-induced pathways such as anti-apoptotic BCL-2 signalling[80], anti-ROS
356 programmes[18] or immune evasion[10]. Further studies are needed to elucidate if these
357 mechanisms are specific to individual cancer tissues or individual quiescence types.

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

367
368 **METHODS**

369 **Selection of quiescence marker genes**

370 *Generic quiescence markers:*

371 Differential expression analysis results comparing cycling immortalised, non-transformed human
372 epithelial cells and cells in five different forms of quiescence (spontaneous quiescence, contact

373 inhibition, serum starvation, CDK4/6 inhibition and MEK inhibition) were obtained from Min and
374 Spencer[18]. A total of 195 genes were differentially expressed in all five forms of quiescence under
375 an adjusted p-value cut-off of 0.05. This gene list, reflective of a generic quiescence phenotype,
376 was subjected to the following refinement and filtering steps: (1) selection of genes with a
377 unidirectional change of expression across all five forms of quiescence; (2) removal of genes
378 involved in other cell cycle stages included in the "KEGG_CELL_CYCLE" gene list deposited at
379 MSigDB; (3) removal of genes showing low standard deviation and low levels of expression within
380 the TCGA dataset, or which showed low correlation with the pan-cancer expression of the
381 transcriptional targets of the DREAM complex, the main effector of quiescence, in TCGA.

382 *Quiescence subtype-specific markers:*

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

389 **Dormancy quantification**

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

398 $$QS = \frac{\frac{1}{n}\sum G_U - \frac{1}{n}\sum G_D}{\frac{1}{n}\sum G_H}$$

399 QS = mean scale quiescence score

400 G_U = expression of genes upregulated in quiescence

401 G_D = expression of genes downregulated in quiescence

402 G_H = expression of housekeeping genes

403 n = number of genes in each gene set

404 Quiescence scores for the TCGA cohort were derived from expression data scaled by tumour
405 purity estimates. The pan-cancer TCGA samples were also classified into “high” or “low”
406 quiescence groups based on k-means clustering (k=2) on the expression data of 139 quiescence
407 biomarker genes, following the removal of tissue-specific expression differences using the
408 *ComBat* function from the *sva* R package[81].

409 **Measuring the depth of quiescence**

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

414 **Validation of quiescence scoring methodologies**

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

416 Datasets were obtained from the ArrayExpress and Gene Expression Omnibus (GEO) databases
417 though the following GEO Series accession numbers: GSE83142, GSE75367, GSE137912,
418 GSE139013, GSE90742 and E-MTAB-4547. Quality control analysis was standardised using the

419 *SingleCellExperiment*[82] and *scater*[83] R packages. Normalisation was performed using the *scran*[84]
420 R package.

421 *Bulk RNA-sequencing validation datasets:*

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

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

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

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

433 Cell lines were obtained from ATCC or Sigma and regularly checked for mycoplasma. A549 and
434 NCIH460 were cultured in DMEM (Gibco). NCIH358, NCIH1299 and NCIH1563 were maintained
435 in RPMI-1640 (Gibco) supplemented with 5mM sodium pyruvate and 0.5% glucose. NCIH1944,
436 NCIH1666, NCIH1650 and L23 were grown in RPMI-1640 ATCC formulation (Gibco). A427 were
437 cultured in EMEM (ATCC). A549, NCIH460, H358, NCIH1299, NCIH1563, A427 were
438 supplemented with 10% heat inactivated FBS. NCIH1666 with 5% heat-inactivated FBS and all
439 other cell lines with 10% non-heat inactivated FBS. All cell lines had penicillin-streptomycin
440 (Gibco) added to 1%. Cells were maintained at 37°C and 5% CO_2. To calculate the quiescent
441 fraction, A549 and NCIH460 cells were plated at a density of 500 cells/well, and all other cell
442 lines at a density of 1000/well, in 384well CellCarrier Ultra plates (PerkinElmer) in the relevant
443 media. 24h later, 5μM EdU was added and cells were incubated for a further 24h before fixing in

444 a final concentration of 4% formaldehyde (15 min, RT), permeabilization with PBS/0.5% Triton
445 X-100 (15 min, RT) and blocking with 2% BSA in PBS (60 min, RT). The EdU signal was detected
446 using Click-iT chemistry, according to the manufacturer's protocol (ThermoFisher). Cells were
447 also labelled for phospho-Ser807/811 Rb (PRb) using Rabbit mAb 8516 (CST) at 1:2000 in
448 blocking solution, overnight at 4°C. Unbound primary antibody was washed three times in PBS
449 and secondary Alexa-conjugated antibodies were used to detect the signal (ThermoFisher,
450 1:1000, 1h at RT). Finally nuclei were labelled with Hoechst 33258 (1 μg/ml, 15 min RT) before
451 imaging on a high-content widefield Operetta microscope, 20x N.A. 0.8. Automated image
452 analysis (Harmony, PerkinElmer) was used to segment and quantify nuclear signals in imaged
453 cells. Quiescent cells were defined by the absence of EdU or PRb staining, determined by
454 quantification of their nuclear expression (Figure 1e-f).

455 The quiescence scores for cancer cell lines were calculated using corresponding log-transformed
456 RPKM normalised bulk RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE)
457 database[85].

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

471 (CST 7074P2 and CST 7076P2, respectivelt) diluted 1:2000 in blocking buffer for 1h at RT.
472 Membranes were washed three times again and signal detected using Clarity ECL solution
473 (BioRad) and scanned on an Amersham ImageQuant 800 analyser.

474 **Multi-omics discovery cohort**

475 FPKM normalised RNA-sequencing expression data, copy number variation gene-level data, as
476 well as mutation annotation files aligned against the GRCh38 human reference genome from the
477 Mutect2 pipeline were downloaded using the *TCGABiolinks* R package [86] for 9,712 TCGA
478 primary tumour samples across 31 solid cancer types. Haematological malignancies were
479 excluded as the quiescence markers were derived in epithelial cells and might not be equally
480 suited to capture this phenotype in blood. For patients with multiple samples available, one RNA-
481 seq barcode entry was selected for each individual patient resulting in 9,631 total entries. All
482 expression data were log-transformed for downstream analysis. During quiescence score
483 calculation, expression data for the primary tumour samples was scaled according to tumour
484 purity estimates reported by Hoadley et al[87] to account for potential confounding quiescence
485 signals coming from non-tumour cells in the microenvironment. Samples with purity estimates
486 lower than 30% were removed, leaving 8,005 samples for downstream analysis.

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

489 *TP53* functional status was assessed based on somatic mutation and copy number alterations
490 as described in Zhang et al[88]. *TP53* mutation and copy number for the TCGA tumours were
491 downloaded from cBioPortal (http://www.cbioportal.org). Tumours with *TP53* oncogenic
492 mutations (annotated by OncoKB) and copy-number alterations (GISTIC score ≤ -1) were
493 assigned as *TP53* mutant and CNV loss. Tumours without these *TP53* alterations were assigned
494 as TP53 wild type. The effects of the *TP53* mutation status on quiescence score were then
495 determined with a linear model approach with the quiescence score as a dependent variable and
496 mutational status as an independent variable. The *P* values were FDR-adjusted.

497 APOBEC mutagenesis enriched samples were determined through pan-cancer clustering of
498 mutational signature contributions as described in Wiecek et al[89]. The APOBEC mutagenesis
499 cluster was defined as the cluster with highest mean SBS2 and SBS13 contribution. This was
500 repeated 100 times and only samples which appeared in the APOBEC cluster at least 50 times
501 were counted as being APOBEC enriched.

502 Aneuploidy scores and whole genome duplication events across TCGA samples were obtained
503 from Taylor et al[90]. Microsatellite instability status for uterine corpus endometrial carcinoma, as
504 well as stomach and colon adenocarcinoma samples were obtained from Cortes-Ciriano et al[56].
505 Telomerase enzymatic activity "EXTEND" scores were obtained from Noureen et al[40].
506 Expression-based cancer cell stemness indices were obtained from Malta et al[41]. Centrosome
507 amplification transcriptomic signature (CA20) scores were obtained from Almeida et al[66].

508 **PHATE dimensionality reduction**

509 The *phateR* R package[91] was used to perform the dimensionality reduction with a constant seed
510 for reproducibility. The *ComBat* function from the *sva* R package[92] was used to remove tissue-
511 specific expression patterns from the TCGA RNA-seq data.

512 **Cancer stem cell division estimates**

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

515 **Positive selection analysis**

516 Genes positively selected specifically in samples clustered into low or high dormancy groups
517 were identified based on dN/dS analysis using the *dNdScv* R package[93], run with default
518 parameters. Genes showing signals of positive selection in either the high dormancy or low
519 dormancy samples which encoded olfactory receptors were discarded from downstream
520 analysis.

521 **Mutational signature estimation**

522 Mutational signature contributions were inferred as described in Wiecek et al[89].

523 **Ensemble elastic net regression model**

524 The COSMIC database was used to source a list of 723 known drivers of tumorigenesis (Tiers
525 1+2). 285 oncogenes and tumour suppressors from a curated list showed a significant
526 enrichment or depletion of mutations or copy number variants in dormant samples either pan-
527 cancer or within individual TCGA studies.

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

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

545

546 **Gene enrichment and network analysis**

547 Gene set enrichment analysis was carried out using the *ReactomePA* R package, as well as
548 GeneMania[94] and ConsensusPathDB[95]. Interactions between CEP89 and other cell cycle
549 components were inferred using GeneMania. Networks were visualised using Cytoscape[96].

550 **Gene lists**

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

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

555 Genes associated with the senescence-associated secretory phenotype were obtained from
556 Basisty et al[44]. Lists of genes making up the various DNA damage repair pathways were derived
557 from Pearl et al[97].

558 Genes associated with contact inhibition were obtained from the curated "contact inhibition" gene
559 ontology term. Genes associated with serum starvation were obtained from the curated
560 "REACTOME_CELLULAR_RESPONSE_TO_STARVATION" list deposited at MSigDB. MEK
561 inhibition was assessed based on the activity of the MAPK pathway as determined using an
562 expression signature (MPAS) consisting of 10 downstream MAPK transcripts[98].

563 **Validation of genomic features of dormancy**

564 For elastic net model feature validation, RNA-seq data was downloaded for 6 cancer studies from
565 cBioPortal[99], along with patient-matched whole-genome, whole-exome and targeted sequencing
566 data. The 6 datasets used comprise breast cancer (SMC[100] and METABRIC[101]), paediatric Wilms'
567 tumor (TARGET[102]), bladder cancer, prostate adenocarcinoma and sarcoma (MSKCC[103-105])
568 studies. The data were processed and analysed in the same manner as the TCGA data. RNA-

569 seq data for 27 MCF7 cell line strains, alongside cell line growth rates and targeted mutational
570 sequencing data were obtained from Ben-David et al[57].

571 **Genomic dependency modelling in breast cancer**

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

577 **Survival analysis**

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

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

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

586 **Quiescence subtype determination**

587 *TCGA cohort studies:*

588 Samples with evidence of dormancy characterised by a generic quiescence score > 0 were
589 further subclassified based on the most likely form of quiescence exhibited, among CDK4/6
590 inhibition, contact inhibition, MEK inhibition, spontaneous quiescence or serum starvation, using
591 subtype-specific expression signatures. We opted for a conservative approach and classed each

592 dormant sample into a specific quiescence subtype if the quiescence score for the corresponding
593 programme was higher than one standard deviation of the distribution across the TCGA cohort,
594 and if the score was significantly higher than for the remaining programmes when assessed using
595 a Student's t test. Samples which could not be classified into any of the five quiescence states
596 characterised in this study were classified as "uncertain".

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

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

607 **Optimisation of dormancy signature**

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

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

quiescence state relative to the other 138 genes. The procedure was repeated for each of the three datasets.

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

Genes were placed in the candidate subset if their importance metric was above a given threshold in at least one of the datasets. By gradually increasing the threshold from 0 to 1, 40 different gene combinations were produced.

(3) External validation of candidate subsets

The gene combinations in (2) were tested for their ability to predict quiescence. For this, a separate validation dataset was utilized, which contained gene expression levels for the 139 genes in the 10 lung cancer cell lines previously employed for experimental validation, along with the quiescence state of the lines as inferred by PRb and EdU staining. For each gene subset, a combined z-score of quiescence was calculated from the expression levels as described previously. The correlations between this z-score and the three experimental measurements of quiescence were used to establish the ability of a gene combination to predict quiescence. The best performing subset with a mean correlation of 80% between predicted and measured quiescence levels in the test data (p=0.01) comprised 35 genes.

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

Data visualization and basic statistics

Graphs were generated using the ggplot2 and ggpubr R packages. Groups were compared using the Student's t test, Wilcoxon rank-sum test or ANOVA, as appropriate.

Code

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

640 **FUNDING**

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

648 **ACKNOWLEDGEMENTS**

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

651 The results published here are in part based upon data generated by the TCGA Research
652 Network: https://www.cancer.gov/tcga.

653 **AUTHOR CONTRIBUTIONS**

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

665 manuscript, with contributions from all other authors. All authors read and approved the
666 manuscript.

667 **COMPETING INTEREST STATEMENT**

668 None declared.

669 **REFERENCES**

670 1. Coller, H.A., Sang, L. & Roberts, J.M. A new description of cellular quiescence. *PLoS Biol*
671 **4**, e83 (2006).
672 2. van Dijk, D. *et al.* Slow-growing cells within isogenic populations have increased RNA
673 polymerase error rates and DNA damage. *Nat Commun* **6**, 7972 (2015).
674 3. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after
675 chemotherapy. *Nature* **488**, 522-6 (2012).
676 4. Puig, I. *et al.* TET2 controls chemoresistant slow-cycling cancer cell survival and tumor
677 recurrence. *J Clin Invest* **128**, 3887-3905 (2018).
678 5. Rehman, S.K. *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive
679 Chemotherapy. *Cell* **184**, 226-242.e21 (2021).
680 6. Phan, T.G. & Croucher, P.I. The dormant cancer cell life cycle. *Nat Rev Cancer* **20**, 398-
681 411 (2020).
682 7. Cabanos, H.F. & Hata, A.N. Emerging Insights into Targeted Therapy-Tolerant Persister
683 Cells in Cancer. *Cancers (Basel)* **13**(2021).
684 8. Masago, K., Fujita, S. & Yatabe, Y. Targeting minimal residual disease after surgery with
685 molecular targeted therapy: the real path to a cure? *J Thorac Dis* **10**, S1982-S1985
686 (2018).
687 9. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of
688 WNT. *Cell* **165**, 45-60 (2016).
689 10. Ribas, A. Adaptive Immune Resistance: How Cancer Protects from Immune Attack.
690 *Cancer Discov* **5**, 915-9 (2015).

691 11. Sosa, M.S., Bragado, P. & Aguirre-Ghiso, J.A. Mechanisms of disseminated cancer cell
692 dormancy: an awakening field. *Nat Rev Cancer* **14**, 611-22 (2014).
693 12. Barkan, D. *et al.* Metastatic growth from dormant cells induced by a col-I-enriched fibrotic
694 environment. *Cancer Res* **70**, 5706-16 (2010).
695 13. Arora, M., Moser, J., Phadke, H., Basha, A.A. & Spencer, S.L. Endogenous Replication
696 Stress in Mother Cells Leads to Quiescence of Daughter Cells. *Cell Rep* **19**, 1351-1364
697 (2017).
698 14. Barr, A.R. *et al.* DNA damage during S-phase mediates the proliferation-quiescence
699 decision in the subsequent G1 via p21 expression. *Nat Commun* **8**, 14728 (2017).
700 15. Heldt, F.S., Barr, A.R., Cooper, S., Bakal, C. & Novák, B. A comprehensive model for the
701 proliferation-quiescence decision in response to endogenous DNA damage in human
702 cells. *Proc Natl Acad Sci U S A* **115**, 2532-2537 (2018).
703 16. Itahana, K. *et al.* A role for p53 in maintaining and establishing the quiescence growth
704 arrest in human cells. *J Biol Chem* **277**, 18206-14 (2002).
705 17. Sadasivam, S. & DeCaprio, J.A. The DREAM complex: master coordinator of cell cycle-
706 dependent gene expression. *Nature Reviews Cancer* **13**, 585-595 (2013).
707 18. Min, M. & Spencer, S.L. Spontaneously slow-cycling subpopulations of human cells
708 originate from activation of stress-response pathways. *PLoS Biol* **17**, e3000178 (2019).
709 19. García-Gutiérrez, L., Delgado, M.D. & León, J. MYC Oncogene Contributions to Release
710 of Cell Cycle Brakes. *Genes (Basel)* **10**(2019).
711 20. Aguirre-Ghiso, J.A., Estrada, Y., Liu, D. & Ossowski, L. ERK(MAPK) activity as a
712 determinant of tumor growth and dormancy; regulation by p38(SAPK). *Cancer Res* **63**,
713 1684-95 (2003).
714 21. Spencer, S.L. *et al.* The proliferation-quiescence decision is controlled by a bifurcation in
715 CDK2 activity at mitotic exit. *Cell* **155**, 369-83 (2013).
716 22. Miller, I. *et al.* Ki67 is a Graded Rather than a Binary Marker of Proliferation versus
717 Quiescence. *Cell Rep* **24**, 1105-1112.e5 (2018).

718 23. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer
719 stem cells. *Nature* **414**, 105-11 (2001).
720 24. Kleinsmith, L.J. & Pierce, G.B., Jr. MULTIPOTENTIALITY OF SINGLE EMBRYONAL
721 CARCINOMA CELLS. *Cancer Res* **24**, 1544-51 (1964).
722 25. Turati, V.A. *et al.* Chemotherapy induces canalization of cell state in childhood B-cell
723 precursor acute lymphoblastic leukemia. *Nature Cancer* (2021).
724 26. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T. & Lee, D. Inferring pathway activity toward
725 precise disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).
726 27. Barbie, D.A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven
727 cancers require TBK1. *Nature* **462**, 108-12 (2009).
728 28. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for
729 microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
730 29. Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics*
731 **19**, 404 (2018).
732 30. Aulestia, F.J. *et al.* Quiescence status of glioblastoma stem-like cells involves remodelling
733 of Ca(2+) signalling and mitochondrial shape. *Sci Rep* **8**, 9731 (2018).
734 31. Kurppa, K.J. *et al.* Treatment-Induced Tumor Dormancy through YAP-Mediated
735 Transcriptional Reprogramming of the Apoptotic Pathway. *Cancer Cell* **37**, 104-122.e12
736 (2020).
737 32. Ebinger, S. *et al.* Characterization of Rare, Dormant, and Therapy-Resistant Cells in
738 Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 849-862 (2016).
739 33. Jordan, N.V. *et al.* HER2 expression identifies dynamic functional states within circulating
740 breast cancer cells. *Nature* **537**, 102-106 (2016).
741 34. Xue, J.Y. *et al.* Rapid non-uniform adaptation to conformation-specific KRAS(G12C)
742 inhibition. *Nature* **577**, 421-425 (2020).
743 35. Buczacki, S.J.A. *et al.* Itraconazole targets cell cycle heterogeneity in colorectal cancer.
744 *J Exp Med* **215**, 1891-1912 (2018).

745 36. Grigore, F. *et al.* BRAF inhibition in melanoma is associated with the dysregulation of
746 histone methylation and histone methyltransferases. *Neoplasia* **22**, 376-389 (2020).
747 37. Fujimaki, K. *et al.* Graded regulation of cellular quiescence depth between proliferation
748 and senescence by a lysosomal dimmer switch. *Proc Natl Acad Sci U S A* **116**, 22624-
749 22634 (2019).
750 38. Stallaert, W. *et al.* The structure of the human cell cycle. *bioRxiv*, 2021.02.11.430845
751 (2021).
752 39. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues
753 can be explained by the number of stem cell divisions. *Science* **347**, 78-81 (2015).
754 40. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an
755 association with cancer stemness and proliferation. *Nature Communications* **12**, 139
756 (2021).
757 41. Malta, T.M. *et al.* Machine Learning Identifies Stemness Features Associated with
758 Oncogenic Dedifferentiation. *Cell* **173**, 338-354.e15 (2018).
759 42. Dimri, G.P. *et al.* A biomarker that identifies senescent human cells in culture and in aging
760 skin in vivo. *Proceedings of the National Academy of Sciences* **92**, 9363 (1995).
761 43. Debacq-Chainiaux, F., Erusalimsky, J.D., Campisi, J. & Toussaint, O. Protocols to detect
762 senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of
763 senescent cells in culture and in vivo. *Nat Protoc* **4**, 1798-806 (2009).
764 44. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging
765 biomarker development. *PLOS Biology* **18**, e3000599 (2020).
766 45. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943-3956 (2017).
767 46. Andreassen, P.R., Lohez, O.D., Lacroix, F.B. & Margolis, R.L. Tetraploid state induces
768 p53-dependent arrest of nontransformed mammalian cells in G1. *Molecular biology of the*
769 *cell* **12**, 1315-1328 (2001).
770 47. Di Leonardo, A. *et al.* DNA rereplication in the presence of mitotic spindle inhibitors in
771 human and mouse fibroblasts lacking either p53 or pRb function. *Cancer Res* **57**, 1013-
772 9 (1997).

773 48. Ganem, N.J. *et al.* Cytokinesis failure triggers hippo tumor suppressor pathway activation.
774 *Cell* **158**, 833-848 (2014).
775 49. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature*
776 **578**, 94-101 (2020).
777 50. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**,
778 415-421 (2013).
779 51. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat Genet* **51**, 1732-1740
780 (2019).
781 52. Maley, C.C. & Reid, B.J. Natural selection in neoplastic progression of Barrett's
782 esophagus. *Semin Cancer Biol* **15**, 474-83 (2005).
783 53. Bardeesy, N. & DePinho, R.A. Pancreatic cancer biology and genetics. *Nat Rev Cancer*
784 **2**, 897-909 (2002).
785 54. Indovina, P., Pentimalli, F., Casini, N., Vocca, I. & Giordano, A. RB1 dual role in
786 proliferation and apoptosis: cell fate control and implications for cancer therapy.
787 *Oncotarget* **6**(2015).
788 55. Mas-Ponte, D. & Supek, F. DNA mismatch repair promotes APOBEC3-mediated diffuse
789 hypermutation in human cancers. *Nat Genet* **52**, 958-968 (2020).
790 56. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P.J. A molecular portrait of
791 microsatellite instability across multiple cancers. *Nature Communications* **8**, 15180
792 (2017).
793 57. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug
794 response. *Nature* **560**, 325-330 (2018).
795 58. Jakobsen, L. *et al.* Novel asymmetrically localizing components of human centrosomes
796 identified by complementary proteomics methods. *Embo j* **30**, 1520-35 (2011).
797 59. Sillibourne, J.E. *et al.* Assessing the localization of centrosomal proteins by
798 PALM/STORM nanoscopy. *Cytoskeleton (Hoboken)* **68**, 619-27 (2011).
799 60. Bettencourt-Dias, M. & Glover, D.M. Centrosome biogenesis and function: centrosomics
800 brings new understanding. *Nature Reviews Molecular Cell Biology* **8**, 451-463 (2007).

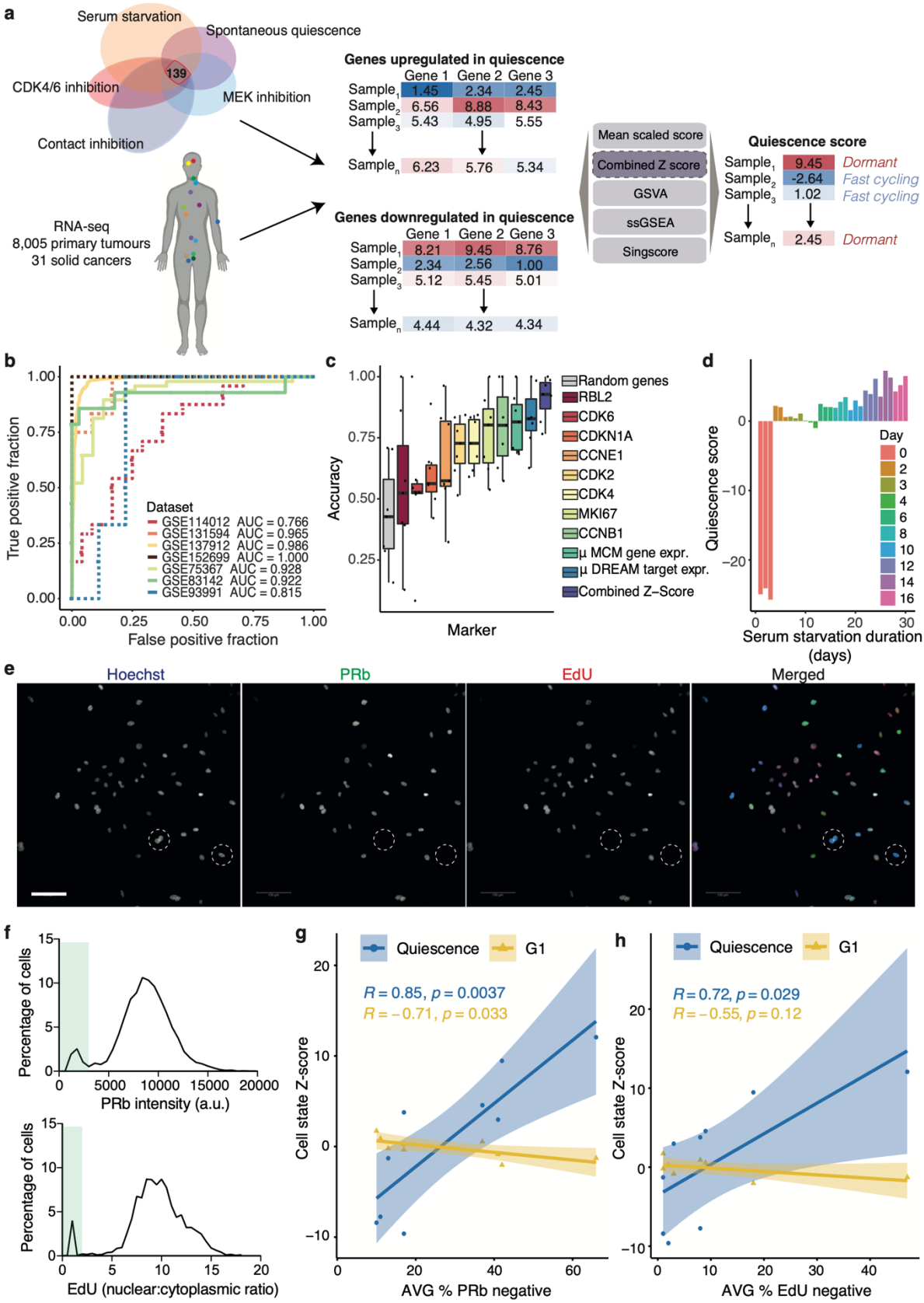
801 61. Grallert, A. *et al.* Removal of centrosomal PP1 by NIMA kinase unlocks the MPF feedback
802 loop to promote mitotic commitment in S. pombe. *Curr Biol* **23**, 213-22 (2013).
803 62. Wigley, W.C. *et al.* Dynamic association of proteasomal machinery with the centrosome.
804 *The Journal of cell biology* **145**, 481-490 (1999).
805 63. Itoh, K., Jenny, A., Mlodzik, M. & Sokol, S.Y. Centrosomal localization of Diversin and its
806 relevance to Wnt signaling. *J Cell Sci* **122**, 3791-8 (2009).
807 64. Kfoury, Y. *et al.* Ubiquitylated Tax targets and binds the IKK signalosome at the
808 centrosome. *Oncogene* **27**, 1665-1676 (2008).
809 65. Lin, A., Wang, R.T., Ahn, S., Park, C.C. & Smith, D.J. A genome-wide map of human
810 genetic interactions inferred from radiation hybrid genotypes. *Genome Res* **20**, 1122-32
811 (2010).
812 66. de Almeida, B.P., Vieira, A.F., Paredes, J., Bettencourt-Dias, M. & Barbosa-Morais, N.L.
813 Pan-cancer association of a centrosome amplification gene expression signature with
814 genomic alterations and clinical outcome. *PLOS Computational Biology* **15**, e1006832
815 (2019).
816 67. Kohlmaier, G. *et al.* Overly long centrioles and defective cell division upon excess of the
817 SAS-4-related protein CPAP. *Curr Biol* **19**, 1012-8 (2009).
818 68. Tang, C.-J.C., Fu, R.-H., Wu, K.-S., Hsu, W.-B. & Tang, T.K. CPAP is a cell-cycle
819 regulated protein that controls centriole length. *Nature Cell Biology* **11**, 825-831 (2009).
820 69. Hafner, M. *et al.* Multiomics Profiling Establishes the Polypharmacology of FDA-Approved
821 CDK4/6 Inhibitors and the Potential for Differential Clinical Activity. *Cell Chem Biol* **26**,
822 1067-1080.e8 (2019).
823 70. Salvador-Barbero, B. *et al.* CDK4/6 Inhibitors Impair Recovery from Cytotoxic
824 Chemotherapy in Pancreatic Adenocarcinoma. *Cancer Cell* **37**, 340-353.e6 (2020).
825 71. Guiley, K.Z. *et al.* p27 allosterically activates cyclin-dependent kinase 4 and antagonizes
826 palbociclib inhibition. *Science* **366**(2019).

827 72. Pack, L.R., Daigh, L.H., Chung, M. & Meyer, T. Clinical CDK4/6 inhibitors induce selective
828 and immediate dissociation of p21 from cyclin D-CDK4 to inhibit CDK2. *Nat Commun* **12**,
829 3356 (2021).
830 73. Zhu, X. *et al.* The prognostic and predictive potential of Ki-67 in triple-negative breast
831 cancer. *Scientific Reports* **10**, 225 (2020).
832 74. Brown, J.A. *et al.* TGF-β-Induced Quiescence Mediates Chemoresistance of Tumor-
833 Propagating Cells in Squamous Cell Carcinoma. *Cell Stem Cell* **21**, 650-664.e8 (2017).
834 75. Park, S.R. *et al.* Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-
835 Fluorouracil-Induced DNA Damage. *Cell Rep* **32**, 108077 (2020).
836 76. Aissa, A.F. *et al.* Single-cell transcriptional changes associated with drug tolerance and
837 response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
838 77. Tan, X., Thapa, N., Sun, Y. & Anderson, R.A. A kinase-independent role for EGF receptor
839 in autophagy initiation. *Cell* **160**, 145-60 (2015).
840 78. Srivatsan, S.R. *et al.* Massively multiplex chemical transcriptomics at single-cell
841 resolution. *Science* **367**, 45-51 (2020).
842 79. Moser, J., Miller, I., Carter, D. & Spencer, S.L. Control of the Restriction Point by Rb and
843 p21. *Proc Natl Acad Sci U S A* **115**, E8219-e8227 (2018).
844 80. Minassian, L.M., Cotechini, T., Huitema, E. & Graham, C.H. Hypoxia-Induced Resistance
845 to Chemotherapy in Cancer. *Adv Exp Med Biol* **1136**, 123-139 (2019).
846 81. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression
847 data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
848 82. Amezquita, R.A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nature*
849 *Methods* **17**, 137-145 (2020).
850 83. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: pre-processing, quality
851 control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*
852 **33**, 1179-1186 (2017).
853 84. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis
854 of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).

855 85. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of
856 anticancer drug sensitivity. *Nature* **483**, 603-7 (2012).
857 86. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of
858 TCGA data. *Nucleic Acids Res* **44**, e71 (2016).
859 87. Hoadley, K.A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of
860 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
861 88. Zhang, P. *et al.* Germline and Somatic Genetic Variants in the p53 Pathway Interact to
862 Affect Cancer Risk, Progression, and Drug Response. *Cancer Res* **81**, 1667-1680 (2021).
863 89. Wiecek, A.J., Jacobson, D.H., Lason, W. & Secrier, M. Pan-Cancer Survey of Tumor
864 Mass Dormancy and Underlying Mutational Processes. *Frontiers in Cell and*
865 *Developmental Biology* **9**, 1820 (2021).
866 90. Taylor, A.M. *et al.* Genomic and Functional Approaches to Understanding Cancer
867 Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).
868 91. Moon, K.R. *et al.* Visualizing structure and transitions in high-dimensional biological data.
869 *Nature biotechnology* **37**, 1482-1492 (2019).
870 92. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for
871 removing batch effects and other unwanted variation in high-throughput experiments.
872 *Bioinformatics* **28**, 882-3 (2012).
873 93. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues.
874 *Cell* **171**, 1029-1041.e21 (2017).
875 94. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration
876 for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-20
877 (2010).
878 95. Kamburov, A. *et al.* ConsensusPathDB: toward a more complete picture of cell biology.
879 *Nucleic Acids Res* **39**, D712-7 (2011).
880 96. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of
881 biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).

882 97. Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B. & Pearl, F.M.G. Therapeutic
883 opportunities within the DNA damage response. *Nature Reviews Cancer* **15**, 166-180
884 (2015).
885 98. Wagle, M.C. *et al.* A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically
886 relevant biomarker in multiple cancer types. *NPJ Precis Oncol* **2**, 7 (2018).
887 99. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring
888 multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
889 100. Kan, Z. *et al.* Multi-omics profiling of younger Asian breast cancers reveals distinctive
890 molecular signatures. *Nat Commun* **9**, 1725 (2018).
891 101. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours
892 reveals novel subgroups. *Nature* **486**, 346-52 (2012).
893 102. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic
894 landscape of Wilms tumor. *Nat Genet* **49**, 1487-1494 (2017).
895 103. Iyer, G. *et al.* Prevalence and co-occurrence of actionable genomic alterations in high-
896 grade bladder cancer. *J Clin Oncol* **31**, 3133-40 (2013).
897 104. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**,
898 11-22 (2010).
899 105. Barretina, J. *et al.* Subtype-specific genomic alterations define new targets for soft-tissue
900 sarcoma therapy. *Nat Genet* **42**, 715-21 (2010).
901

902 **FIGURES**



903

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

905 dormancy in RNA-seq cancer samples. 139 genes differentially expressed in multiple forms of
906 quiescence were employed to score dormancy across TCGA cancers. **(b)** Receiver operating
907 characteristic (ROC) curves illustrating the performance of the z-score methodology on
908 separating actively proliferating and quiescent cells in seven publicly available single-cell
909 (continuous curves) and bulk RNA-seq (dotted curves) datasets. AUC = area under the curve.
910 **(c)** Compared classification accuracies of the z-score approach and classic cell proliferation
911 markers across the seven single-cell/bulk RNA-seq validation datasets. **(d)** Quiescence levels of
912 embryonic fibroblast cells under serum starvation for various amounts of time. Replicates are
913 depicted in the same colour. **(e)** Representative images of lung cancer cell lines immunostained
914 and analysed to detect the quiescent fraction. Hoechst (labels all nuclei) is in blue, PRb in green
915 and EdU in red in merged image. White dashed circles highlight quiescent cells that are negative
916 for both PRb and EdU signals. Scale bar: 100 μm. **(f)** Graphs show single cell quantification of
917 PRb and EdU intensities taken from images and used to define the cut-off to calculate the
918 quiescent fraction (QF; green boxes). Images in **(e)** and graphs in **(f)** are taken from the A549
919 cell line. **(g-h)** Correlation between theoretical estimates of a quiescence or G1 state and the
920 fraction of cells entering quiescence in nine lung adenocarcinoma cell lines, as assessed through
921 **(g)** phospho-Rb assays and **(h)** EdU assays. Mean of n=3 is shown for the average percentage
922 of quiescent cells.
923
924
925
926
927

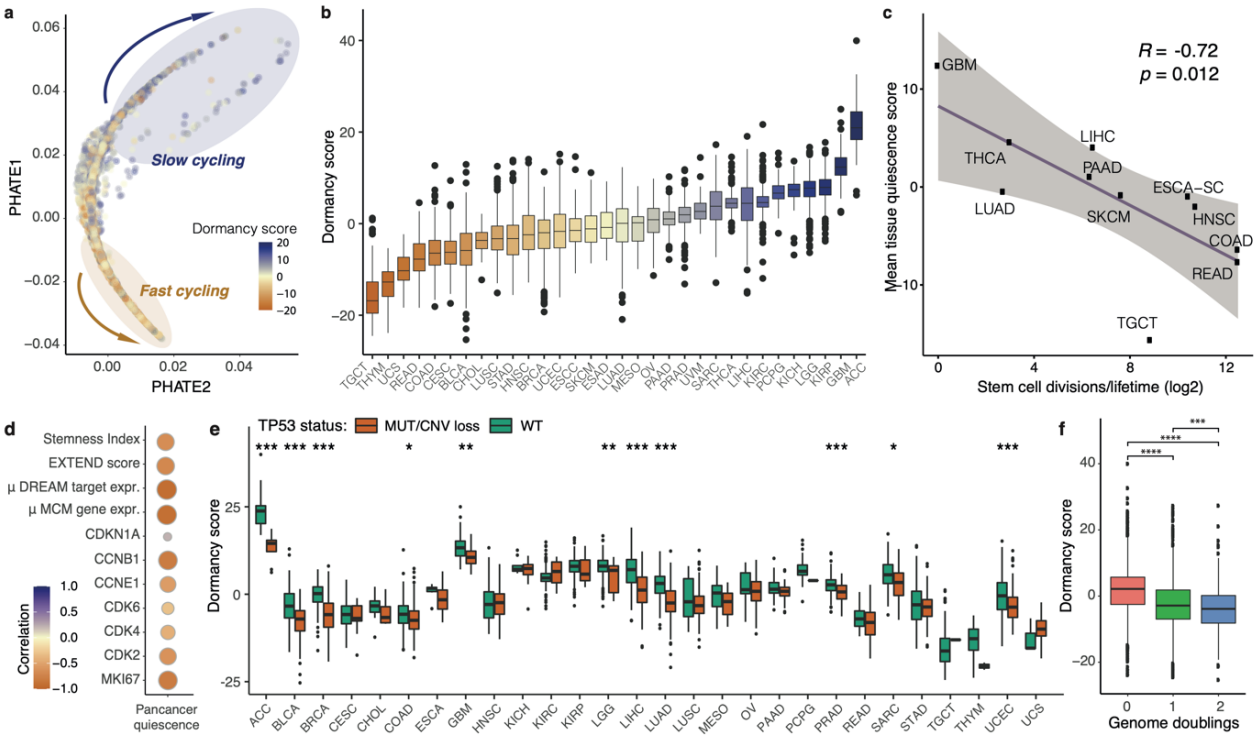
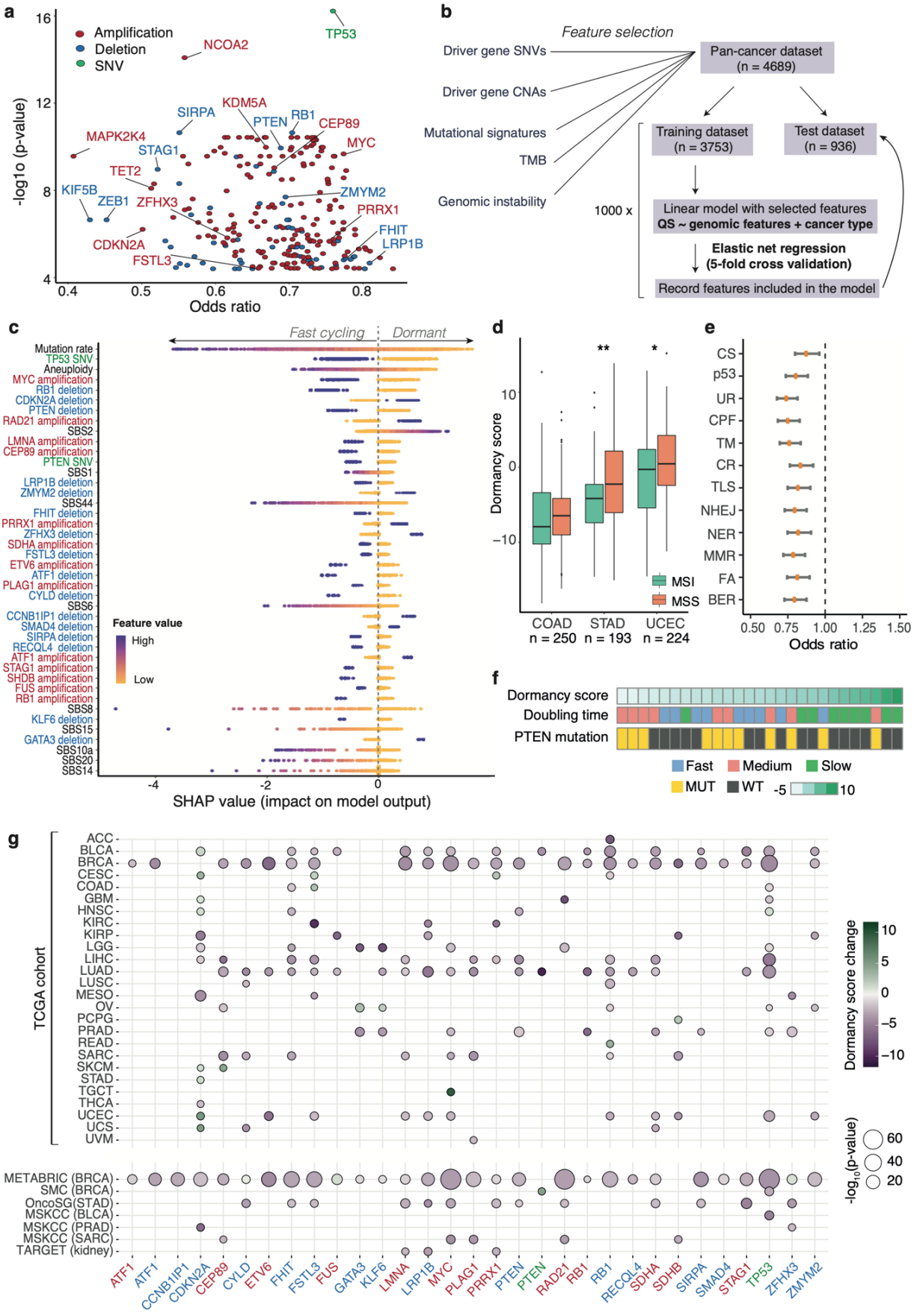


Figure 2: Pan-cancer evaluation of dormancy. (a) PHATE plot illustrating the wide spectrum of dormancy-proliferation states across 9,712 primary tumour samples from TCGA. Each sample is coloured according to the relative dormancy level. **(b)** Variation in tumour dormancy levels across different cancer 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.



941

942 **Figure 3: Pan-cancer genomic landscape of dormancy. (a)** Cancer drivers with mutations or

943 copy number alterations depleted pan-cancer in a dormancy context. Features further selected

944 by the pan-cancer model are highlighted. **(b)** Schematic of the ensemble elastic net modelling
945 employed to prioritise genomic changes associated with tumour dormancy. **(c)** Genomic events
946 significantly associated with dormancy, ranked according to their importance in the model (from
947 highest to lowest contribution). Each point represents an individual primary tumour sample of the
948 training dataset, coloured by the corresponding value of each feature. For discrete variables
949 purple indicates the presence of the feature and yellow indicates the absence of the feature. The
950 Shapley values indicate the impact of individual feature values on the quiescence score
951 prediction. **(d)** Dormancy levels are significantly reduced in microsatellite unstable (MSI) samples
952 in stomach adenocarcinoma (STAD) and uterine corpus endometrial carcinoma (UCEC), with
953 the same trend (albeit not significant) shown in colon adenocarcinoma (COAD). Wilcoxon rank-
954 sum test *$p<0.05$; **$p<0.01$. **(e)** Genomic alterations are depleted across DNA repair pathways
955 during dormancy. Odds ratios of mutational load on pathway in dormancy are depicted, along
956 with confidence intervals. CS, chromosome segregation; p53, p53 pathway; UR, ubiquitylation
957 response; CPF, checkpoint factor; TM, telomere maintenance; CR, chromatin remodelling; TLS,
958 translesion synthesis; NHEJ, non-homologous end joining; NER, nucleotide excision repair;
959 MMR, mismatch repair; FA, Fanconi Anaemia pathway; BER, base excision repair. **(f)** Dormancy
960 is increased in cell lines with slow doubling time across MCF7 strains, which also show lower
961 prevalence of PTEN mutations. **(g)** Tissue-specific changes in quiescence between samples
962 with/without dormancy-associated deletions (blue), amplifications (red) and SNVs (green) within
963 the TCGA cohort (top) and external validation datasets (bottom).

964

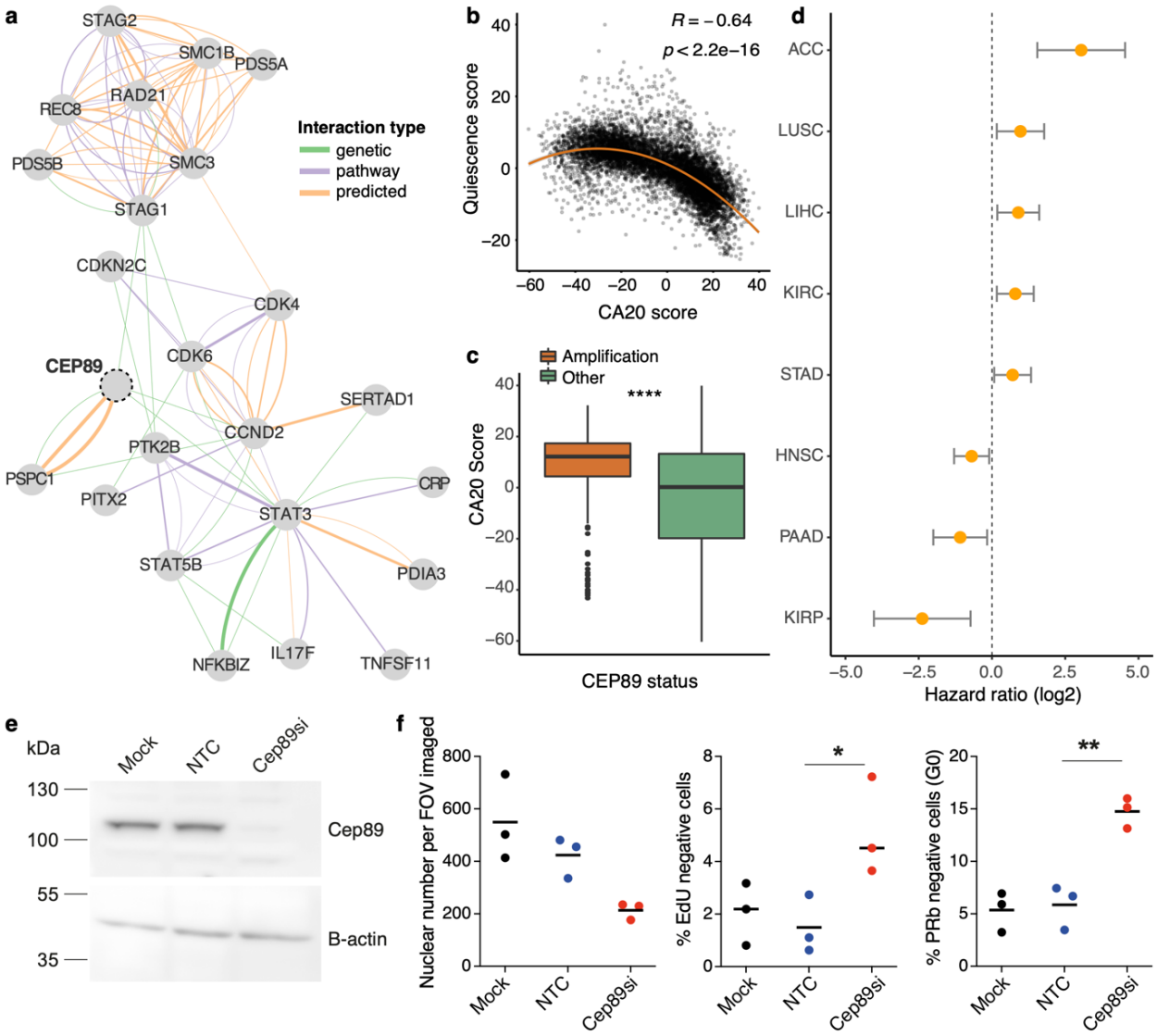


Figure 4: CEP89 amplification is associated with lower tumour dormancy. (a) Network illustrating CEP89 interactions with cell cycle genes (from GeneMania). The edge colour indicates the interaction type, with green representing genetic interactions, orange representing predicted interactions and purple indicating pathway interactions. The edge width illustrates the interaction weight. **(b)** Pan-cancer relationship between CA20 scores and tumour quiescence across the TCGA cohort. **(c)** CA20 scores are significantly increased in TCGA primary tumours containing a CEP89 amplification. **(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 considering tumour stage. Patients with high expression of CEP89 show significantly worse prognosis within ACC, LUSC, LIHC, KIRC and STAD, but significantly better prognosis within HNSC, PAAD and KIRP studies. **(e)** Western blot showing depletion of Cep89

977 protein 48h after siRNA transfection of NCI-H1299 cells. Mock is lipofectamine only, NTC is Non-
978 targeting control siRNA. B-actin is used as a loading control. **(f)** Graphs show that Cep89
979 depletion in NCI-H1299 cells leads to a reduction in nuclear number and an increase in the
980 fraction of quiescent cells, measured by an increase in the percentage of EdU negative (24h EdU
981 pulse) and Phospho-Ser 807/811 Rb negative cells. One-Way ANOVA, *$p<0.05$, **$p<0.01$. Mean
982 of n=3.

983

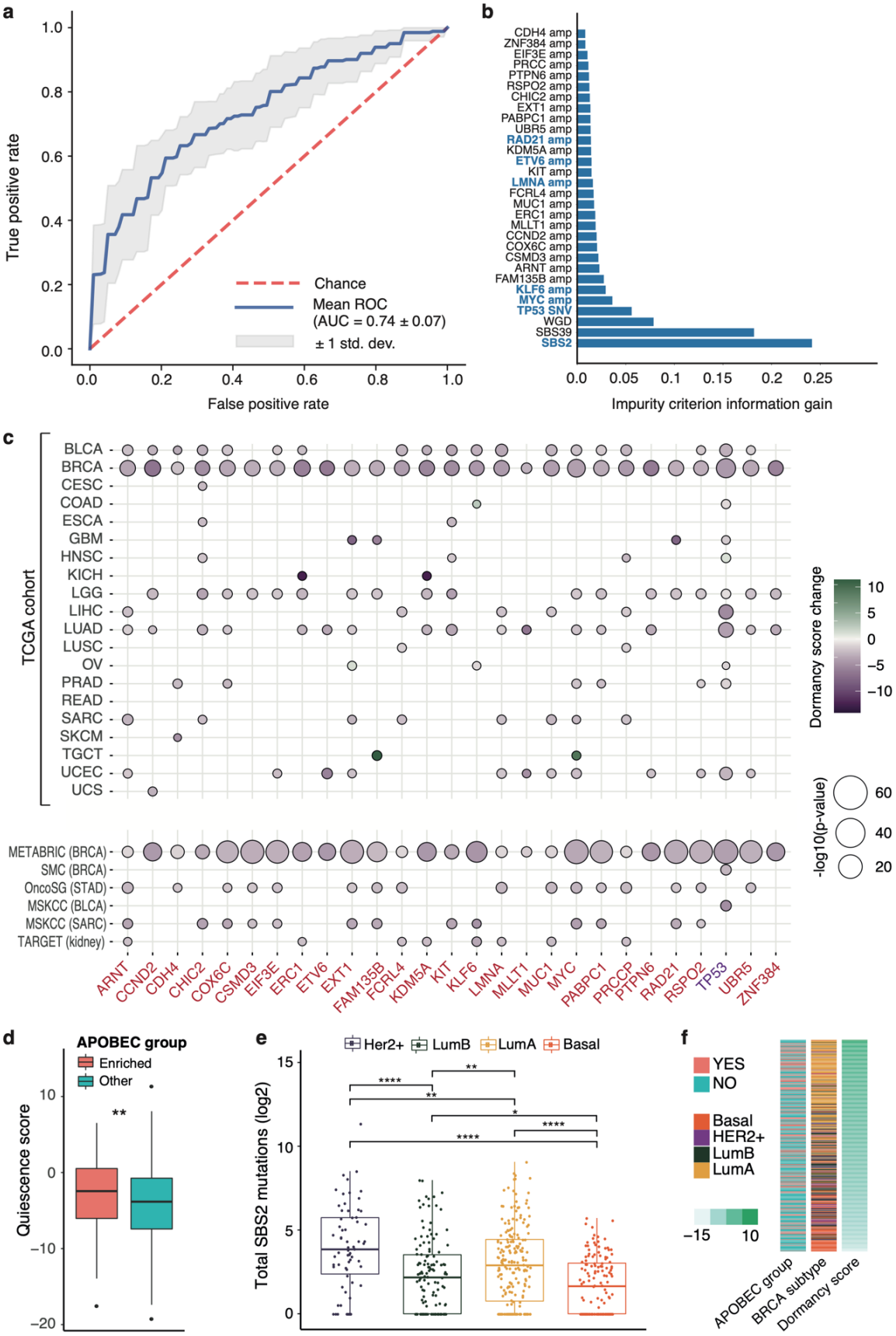


Figure 5: Genomic landscape of breast cancer dormancy. (a) ROC curve illustrating the mean performance of the random forest model for dormancy prediction in breast cancer based

on 9 randomisations. **(b)** Impurity criterion information gain for the top ranked genomic events predictive of dormancy in the TCGA BRCA cohort, modelled using random forests. Features also appearing in the pan-cancer model are highlighted in blue. **(c)** Tissue-specific changes in dormancy estimates between samples with and without dormancy-associated amplifications (red) and SNVs (purple), identified by the breast cancer specific random forest model, within the TCGA cohort (top panel) and external validation datasets (bottom panel). **(d)** Dormancy scores are significantly higher in APOBEC mutation-enriched breast cancer samples. **(e)** APOBEC-linked SBS2 mutational burden compared between breast cancer subtypes. **(f)** Association between dormancy scores, APOBEC enrichment and breast cancer subtypes in TCGA.

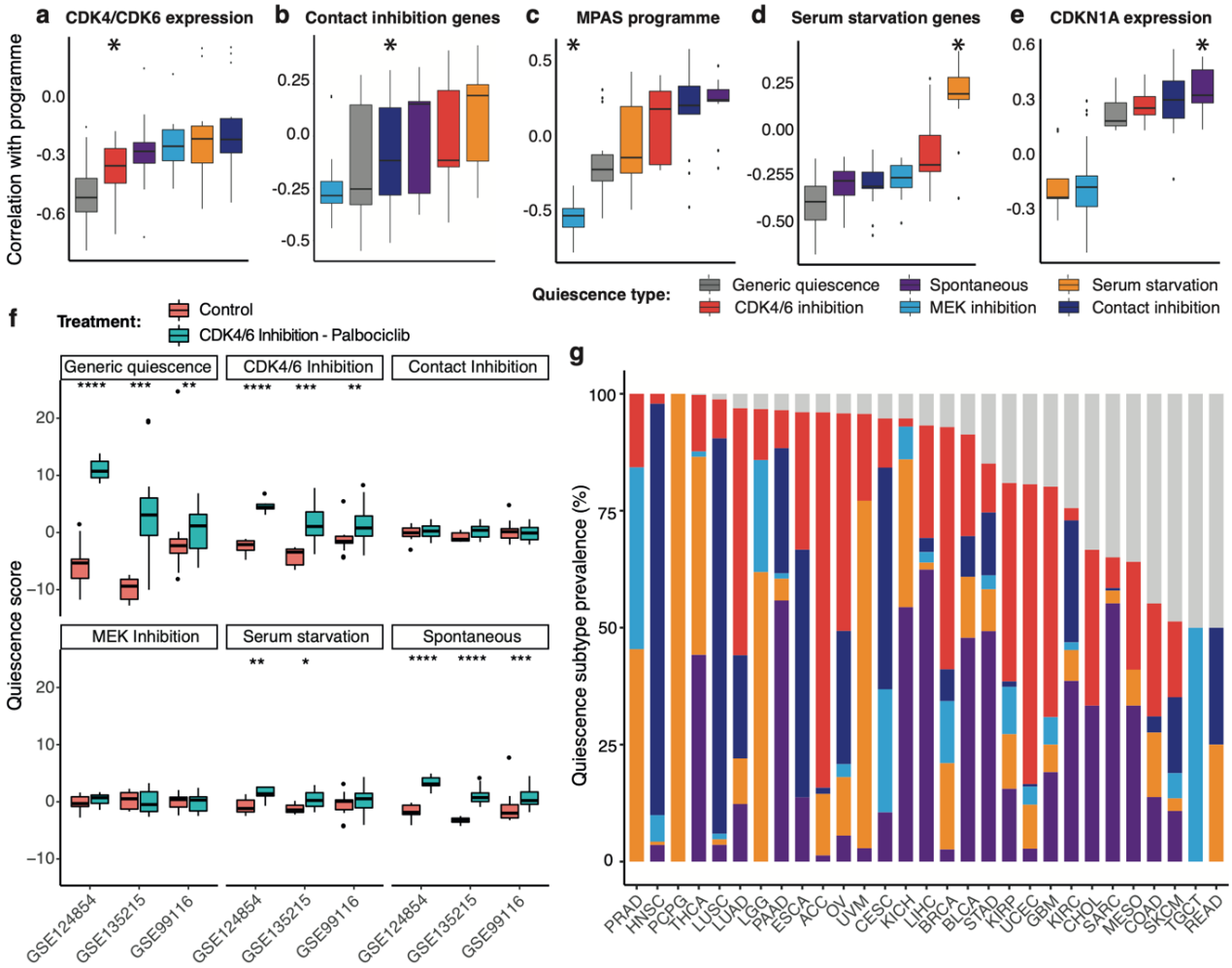


Figure 6: Pan-cancer characterisation of individual quiescence programmes. (a-e) Comparison of correlation coefficients between quiescence programme scores and **(a)** mean expression of CDK4 and CDK6, **(b)** mean expression of curated contact inhibition genes, **(c)** a transcriptional MAPK Pathway 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 (red) and after (blue) 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.

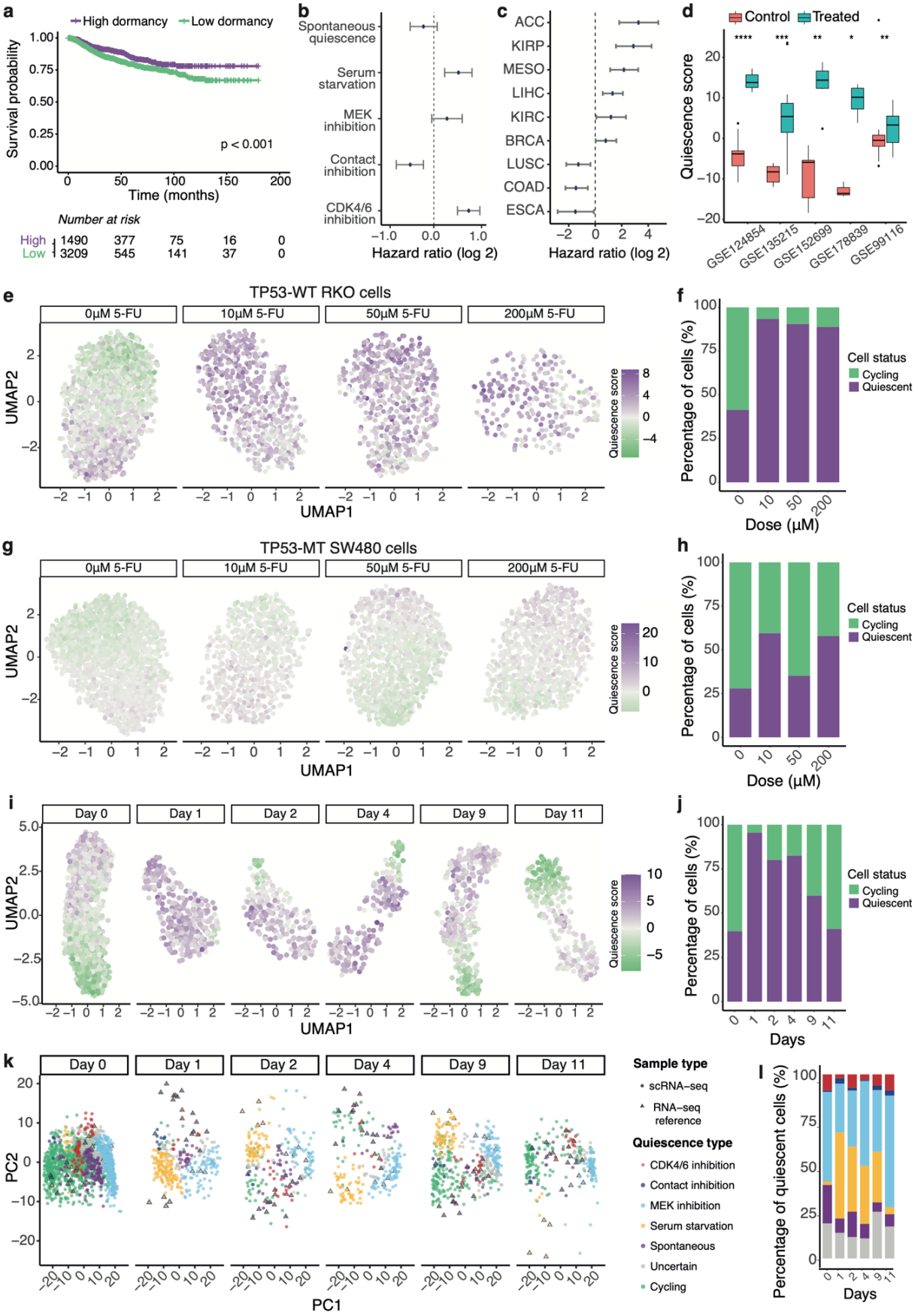


Figure 7: Impact of cellular dormancy on patient prognosis and treatment response. (a) Disease-specific survival based on tumour dormancy for patients from TCGA within 15 years of

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

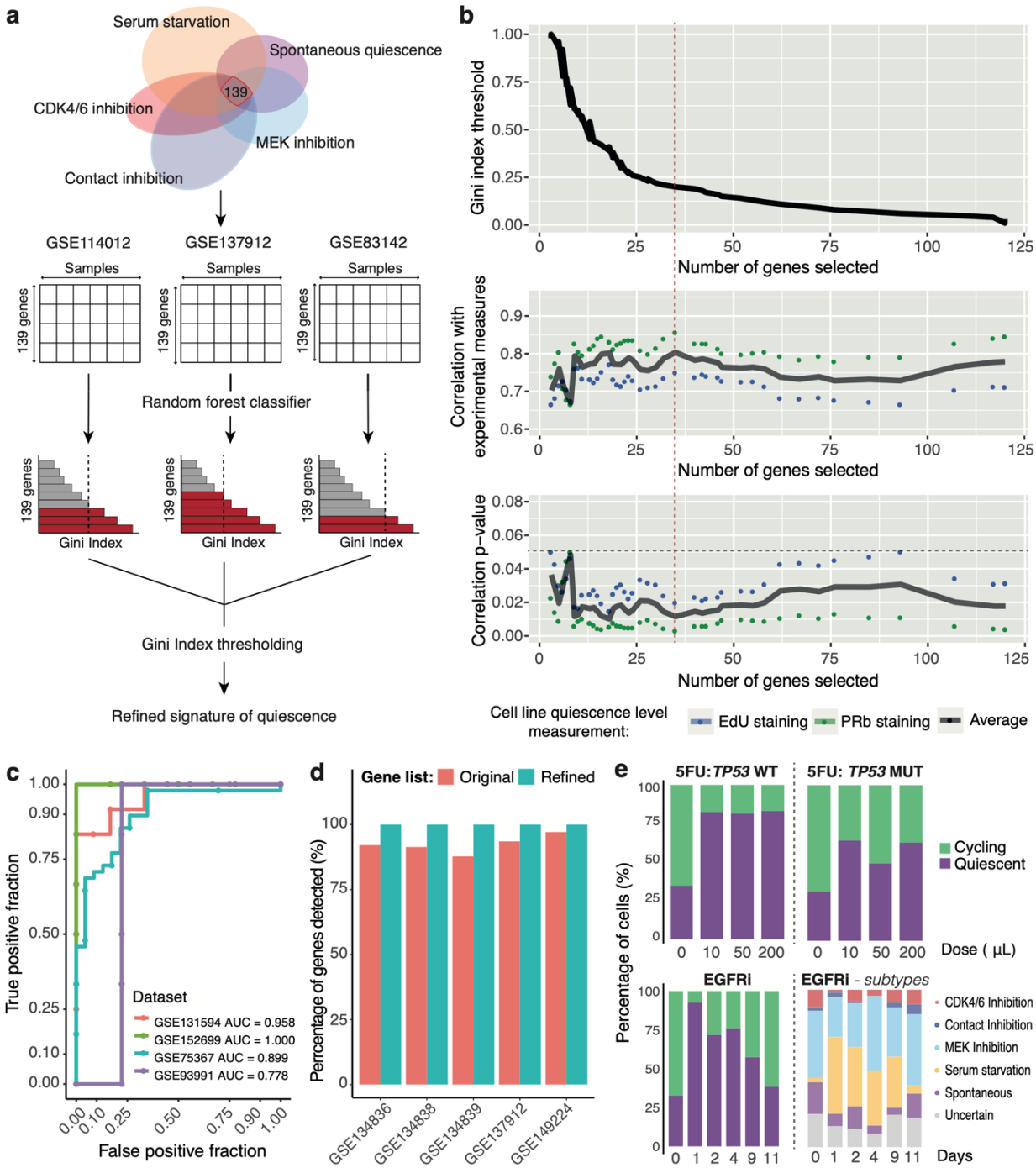


Figure 8: Optimisation of dormancy signature for use in scRNA-seq data. (a) Methodology for reducing the gene signature of dormancy: random forest classifiers are trained to distinguish dormant from cycling tumours on three high confidence datasets; Gini index thresholding is optimised to prioritise 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 indicates the threshold chosen for the final solution of 35 genes. The black dotted line

1043 indicates the threshold for p-value significance. **(c)** Additional external validation of the 35 gene
1044 signature acting as a classifier of quiescent and proliferation cells in single cell and bulk datasets.
1045 **(d)** Dropout in single cell data by gene signature. The percentage of genes out of the 35 (blue)
1046 and 139 (red) gene lists with reported expression across five single-cell RNA-seq datasets. **(e)**
1047 Proportion of cycling and quiescent cells estimated in single cell datasets of p53 wild type and
1048 mutant lines treated with 5FU, as well as cells treated with EGFR inhibitors. Data as in Figure 7.