Non-genetic control of cell plasticity and phenotypic cell states is mediated by cytoplasmic lncRNA nucleated signalling pathways

Running title: Cytoplasmic lncRNAs drive non-genetic phenotypic divergence

Shlyakhtina Y*[1], Bloechl B*[1,2,4], Moran KL*[1], Maslakova A[2,4], Johnson RB[2], Munro AF[3,4], Carragher NO[3,4] and Portal MM[#1,2,4]

*- Contributed equally to this work

[#]- Correspondence should be addressed to: Maximiliano.Portal@glasgow.ac.uk

Affiliations:

1- Cell Plasticity & Epigenetics Lab, Cancer Research UK – Manchester Institute, The University of Manchester, M20 4BX, Manchester, UK
2- Cell Plasticity & Epigenetics Lab, Cancer Research UK – Scotland Institute, University of Glasgow, G61 1BD, UK
3- Edinburgh Cancer Research, Institute of Genetics and Cancer, University of Edinburgh, EH4 2XR, UK
4- Cancer Research UK Scotland Centre, UK

Abstract

Cell plasticity, the ability that cells display to rapidly adapt to environmental cues, is thought to be encoded in non-genetic information reservoirs. Although the notion is widely acknowledged, the molecular details underlying this phenomenon remain largely concealed. Herein, we show that clonal cell populations inherently display multiple co-existing metastable gene expression states that co-segregate with various phenotypic outputs. Moreover, we provide primer evidence suggesting that transcriptome states are inherited, dynamically interconvert and determine phenotypic output upon a variety of biological cues, not as a result of transcriptional shifts, but rather through yet unidentified post-transcriptional mechanisms. Remarkably, among phenotypically divergent clonal cell populations enriched in subsets of transcriptome states, we identified a peri-nuclear cytoplasmic structure (Signal Integration Portal – SIP) where state-specific lncRNAs, proteins harbouring intrinsically disordered regions and various active signalling pathways converge. Herein, we propose that SIP-condensates act as nucleating reservoir of non-genetic information at the crossroads of cell plasticity and non-genetic heterogeneity where they integrate intra- and extracellular inputs thereby moulding phenotypic output.

Key words: cell plasticity, epigenetics, RNA sequencing, transcriptomics

38 **Introduction**

39

40 The wide field of genetics has revolutionised the manner in which we approach biological problems
41 and, even though its remarkable relevance cannot be contested, hundreds of observations over the
42 years suggest that the non-genetic compartment is of equal importance to define life as we know it
43 and, therefore, should not be neglected[1-3]. Along those lines, following the development and
44 refinement in next generation sequencing (NGS) technology and its application to single cell analysis,
45 we can finally begin to query the non-genetic compartment with unprecedented detail in a robust
46 and quantitative manner. It is in that context that cell plasticity – the ability of a single genotype to
47 produce a variety of phenotypes – has emerged as a central player in various and diverse fields of
48 study ranging from microbiology to cancer research[4-7]. Notably, within the plethora of non-genetic
49 information carriers present in cells, the study of the transcriptome has transformed basic and
50 biomedical research. In particular, the development of single-cell RNA-Sequencing (scRNA-Seq)
51 unravelled that even genetically homogeneous cell populations inherently display heterogeneous
52 transcriptome profiles[8].

53

54 Strikingly, though transcriptome heterogeneity is now considered to be prevalent in most biological
55 settings, our understanding of how transcriptome states are established, maintained and inherited,
56 and the molecular mechanisms underlying their potential to reshape into novel transcriptome
57 arrangements remains fragmentary. Although attempts to unravel the molecular details underlying
58 this phenomenon have been met with modest success[9-13], the field of system dynamics brought
59 forward a theoretical framework which anchored on Waddington's epigenetic landscape[14-16]. It
60 suggests that each metastable – attractor – state embedded in a gene regulatory network can be
61 represented as low-energy wells in a rugged quasi-potential landscape towards which unstable
62 transient states are attracted to[14,17,18]. Along those lines, and in partial agreement with the
63 aforementioned theoretical framework, our reported data showed that non-genetic cell

64 heterogeneity in terms of meta-stable transcriptome states is not randomly established but follows
65 precise phylo-(epi-)genetic linked inheritance patterns, generating traceable cell-ancestry trees
66 within isogenic and genetically stable experimental models (Barcode Decay Lineage tracing, BdLT-
67 Seq[8]). Moreover, intrapopulation lineage-linked transcriptome states are perpetuated for several
68 generations fuelling population heterogeneity and thus suggesting the existence of a "molecular
69 memory" retained through clonal propagation/divergence.
70
71 Herein, we show that by subcloning a genetically stable clonal founder cell population we isolate
72 ancestry trees that are enriched in specific subsets of transcriptome states which correlate with the
73 fate of individual cells in response to intracellular and extracellular cues. Moreover, we provide
74 primer evidence suggesting that for a given cellular population, the degree of phenotypic plasticity is
75 anchored to the transcriptome state of each individual cell and its associated lineage-linked
76 restricted plasticity. Finally, we describe the existence of a cytoplasmic compartment (Signal
77 Integration Portal - SIP) that nucleates long non-coding RNAs, proteins containing intrinsically
78 disordered regions (IDR) and active components of multiple signalling pathways that co-segregate
79 with subsets of transcriptome states and whose variable composition correlates with phenotypic
80 output and transcriptome-linked non-genetic heterogeneity.
81
82 Collectively, our data supports the notion that the perpetuation of non-genetically heterogeneous
83 cell populations in physiological and pathological conditions, as well as in response to biological
84 and/or therapeutic cues, is not only dictated by the overall cellular architecture (i.e. epithelial,
85 mesenchymal, tissue of origin) and molecular characteristics (i.e. genetic landscape, bulk
86 transcriptome profiles) of the population but also strongly relies on the non-genetic molecular
87 identity (SIP complexity and composition) and plasticity of its intrapopulation phylo-epigenetic
88 lineages.

Results

Clonal cell systems retain a molecular memory of past cell states

We have recently shown that by applying Barcode Decay Lineage tracing (BdLT-Seq) to a given cellular system of choice we can unravel metastable gene expression state dynamics in a lineage-linked manner[8]. BdLT-Seq data also suggested that the gene expression states observed within a cellular population are not randomly generated but rather constrained by unknown molecular mechanisms and could potentially drive phenotypic output.

Therefore, to unravel the molecular mechanisms underlying gene expression state transitions and its link with phenotypic output, we first evaluated a clonal founder population of HA1ER cells (F12 - HEK cells immortalised with hTERT/SV40ER and transformed by $HRAS^{G12V}$ [19]) and two stable subclones derived from it (1F8 and 1C9) previously reported as diverging at their transcriptome level[8]. Even though the three clonal populations (F12, 1F8 and 1C9) were essentially indistinguishable for a plethora of cellular parameters (cell proliferation, ploidy, cell cycle profile, morphology. Supplementary Fig. 1), scRNA-Seq (Chromium 10x) revealed that, after data processing to mitigate clustering artefacts due to cell cycle progression and abnormal mitochondrial counts (Proxy for cell death. Seurat v4.0), gene expression states dramatically differed between the founder population and its subclones (subclones 1C9 and 1F8; Figs. 1a, 2a, Supplementary Figs. 2a, 3a). Interestingly, those transcriptomic differences are further propagated down the lineage when the first subclones obtained from the founder clone are further subcloned, indicating a certain degree of inheritance based on the lineage-linked plastic capacity of the transcriptome states (Subclones 1F8, 3C5 and 1E8; Supplementary Fig. 4, left. Subclones 2A6, 2B12 and 1F5; Supplementary Fig. 4, right). Remarkably, whole exome (WES, 100x coverage) and whole genome sequencing (WGS, 30x coverage) for clones F12, 1F8 and 1C9 did not reveal any genetic difference between them (variance per site (nucleotide

115 position) <0.02, Supplementary Figs. 5, 6) thus suggesting that the observed gene expression state
116 divergence is encoded and propagated by non-genetic networks.
117
118 To further verify this possibility, we performed BdLT-Seq to determine if transcriptome plasticity was
119 differential between the founder clone F12 and subclones 1F8 and 1C9 and revealed that indeed, the
120 founder clone and its derived subclones display not only a divergent state composition but also
121 differential basal transcriptome plasticity (Fig. 1b, c and Supplementary Fig. 2b). For example, after
122 four days, the progeny of cells in cluster 2 of clone F12 could be traced to every cluster except
123 clusters 3, 7 and 12, with the majority transitioning to clusters 0 and 10. On the other hand, cells
124 from cluster 2 of subclone 1F8 mainly diverged to clusters 3, 7 and 11, whereas cells from cluster 2 of
125 subclone 1C9 could be traced primarily to clusters 0 and 1 but not to clusters 3, 10 and 11 (Fig. 1c,
126 Supplementary Fig. 2b).
127
128 Prompted by these results, we hypothesised that any isogenic and genetically stable cellular system
129 amenable to cloning could display similar divergent transcriptome behaviour when analysing
130 subclonal populations at the single-cell level and concordantly could be exploited to shed light onto
131 the molecular mechanism underlying the genesis/establishment and maintenance of metastable
132 transcriptome state dynamics. To pursue this hypothesis, we built three different clonal cellular
133 systems (HA1ER, described above; HA1E, HEK cells immortalised with hTERT/SV40ER; HepG2,
134 Hepatocellular carcinoma cells) by generating in the first step a clonal founder population and then
135 subcloning each founder population to obtain transcriptome state diverging subclones whilst
136 maintaining isogeneity and enabling lineage-linked comparisons (Fig. 1d).
137
138 After initial phenotypic characterisation of each clone (cell proliferation, ploidy, cell cycle profile and
139 morphology. Supplementary Figs. 1) and determination of their genetic makeup which did not show
140 any significant differences (WES for all clones and WGS for HA1ER clones F12, 1F8 and 1C9 and

141 HepG2 clones 1A, 14F and 44D, variance per site <0.02 for all, Supplementary Figs. 5, 6), the
142 transcriptome states of the founder clone and the four subclones from each system were analysed
143 by scRNA-Seq (10x Genomics Chromium 3', Fig. 2a-c). Strikingly, scRNA-Seq data analysis determined
144 that for the HA1ER system, 3 of the 4 subclones (1D12, 2A6 and 1F8) displayed transcriptome state
145 re-arrangements when contrasted to the founder population F12, whilst the remaining subclone
146 (1C9) exhibited similar transcriptome states as the founder albeit displaying altered state
147 representation (Fig. 2a, Supplementary Fig. 3a). Similarly, when populations from the HA1E lineage
148 were analysed, 3 out of the 4 subclones (A7A, A9H, B2F) also displayed vast transcriptome re-
149 arrangements when compared to the founder clone 10E whereas subclone B5F was the most
150 comparable to the parental clone (Fig. 2b, Supplementary Fig. 3b).
151
152 To further explore whether transcriptome complexity (state representation) would affect phenotypic
153 output throughout dynamic processes, we took advantage of the capability of HepG2 cells to activate
154 a TGFβ-dependent transcriptional programme driving the epithelial-to-mesenchymal transition
155 (EMT). Briefly, we treated the founder HepG2 population (1A) and the four established subclones
156 (32G, 14F, 48E and 44D) with TGFβ to promote EMT (Supplementary Fig. 7a) and evaluated the
157 representation of gene expression states by scRNA-Seq in vehicle-treated (-TGFβ) or 4 days TGFβ-
158 treated populations (+TGFβ). As observed in HA1E and HA1ER cells, when TGFβ–naïve HepG2 cells
159 were assessed, significant differences in the transcriptomic states between subclones could be
160 observed when compared to the founder clone 1A (Fig. 2c, Supplementary Figs. 3c, 8). For instance,
161 whereas in the naïve founder clone (1A) virtually no cells could be detected belonging to cluster 7,
162 subclone 14F and 48E displayed a pronounced increase in the cellular representation in this cluster
163 (Fig. 2c, Supplementary Fig. 3c). As expected, upon EMT induction, a considerable gene expression
164 shift was observed when either the founder clone or the subclones were subjected to TGFβ
165 treatment. Strikingly, the initial differences observed in the TGFβ–naïve subclones propagated into
166 the states identified upon TGFβ treatment. Notably, the subclones displaying an enrichment of cells

167 in cluster 7 in naïve conditions (14F and 48E) displayed a reduced representation of cells occupying
168 cluster 6 following EMT compared to the other clones (Fig. 2c, Supplementary Fig. 3c).
169
170 To take this further and given that the HepG2-based TGFβ-induced EMT model system also enables
171 us to assess the mesenchymal-to-epithelial transition (MET) by the removal of TGFβ from the
172 cultures, we took advantage of this possibility and explored transcriptomic state re-arrangement
173 upon EMT reversion using scRNA-seq (Supplementary Fig. 8). We observed that the founder clone 1A
174 and its most similar subclone at the transcriptome level, 44D, reverted to occupy similar
175 transcriptome states as the initial TGFβ–naïve population; however, their state representation was
176 significantly altered with cluster 6 becoming the most predominant (Supplementary Fig. 8). Notably,
177 clone 14F which was the most dissimilar from the founder clone 1A, took a completely different path
178 upon EMT reversion, as cluster 1 and 8 were overpopulated with only a minority of cells occupying
179 other states (Supplementary Fig. 8). Together, this data supports the notion that even when a
180 transcription factor mediated gene expression programme is enforced by external cues, the initial
181 cell states are largely inherited and may impact the phenotypic fate of a clonal population.
182
183 **Isogenic subpopulations enriched in different transcriptomic states are associated with distinct**
184 **biological phenotypes**
185
186 Following up on these observations and our previously reported data[8], we set out to evaluate several
187 phenotypic outputs relevant to each particular cellular system to compare the phenotypic responses
188 between subclonal populations enriched in distinct transcriptomic states. First, we took advantage of
189 the well-characterised HA1ER cell system that bears the capability to grow in 3D soft agar matrices,
190 where its growth in anchorage-independent conditions serves as *in-vitro* proxy for cellular
191 transformation[19-23].
192

193 Interestingly, though no major differences could be observed in the proliferation rates of the
194 analysed clones (1D12, 1C9, 2A6 and 1F8) and their founder clone (F12) in 2D growth
195 (Supplementary Fig. 1a), challenging these clonal populations to grow in anchorage-independent
196 conditions resulted in variable phenotypic outputs as subclone 1C9 (most similar to the founder
197 population F12 at the transcriptome state level), grew more efficiently under these conditions (Fig.
198 2d). Notably, HA1ER subclones 2A6 and 1F8 (most dissimilar at the transcriptome state level to
199 founder clone F12), have almost lost the capability to grow in 3D settings albeit being genetically
200 identical to the founder cell population (Fig. 2d, Supplementary Fig. 7c). Strikingly, subclone 1D12
201 grew to a similar extent to the parental clone F12 in 3D conditions despite being the most similar at
202 the transcriptome level to subclone 2A6 which failed to grow efficiently under the same conditions
203 (Fig. 2a, Supplementary Fig. 7c), thus suggesting that transcriptome organisation plays a role in
204 modulating phenotypic output but is not the sole determinant. Interestingly, when the second round
205 of cloning from the F12 lineage was analysed, the subclones derived from 1F8 and 2A6 (1E8, 3C5;
206 1F5, 2B12, respectively) did follow the phenotype displayed by their respective founder clone and
207 failed to grow in 3D soft agar (Supplementary Fig. 7c), thus indicating that phenotypic divergence is
208 propagated across several generations and is potentially restricted by the plasticity of the parental
209 clone.
210
211 To delve further into our findings, we took advantage of the fact that immortalised HA1E cells (hTERT
212 + SV40ER), are susceptible to transformation by overexpressing a RAS oncogene[19,23]. To evaluate
213 whether transcriptomic divergence can result in differences in transformation potential, we prepared
214 a multiplexed lentiviral library containing 26 RAS variants from the three main families H-RAS, N-RAS
215 and K-RAS (wild type and 7-8 mutants for each family; Supplementary Fig. 7f), transduced each of
216 the HA1E subclones (A7A, A9H, B2F and B5F) and their founder (10E) and then subjected each
217 population to anchorage-independent growth conditions in 3D soft agar to evaluate their
218 susceptibility to malignant transformation (Fig. 2e). Notably, subclones A9H and B5F grew to a similar

219 extent in 3D soft agar cultures as the founder clone (subclone B5F most similar at the transcriptome
220 level to the founder clone 10E. Fig. 2b, e). However, two subclones that dramatically diverge at the
221 level of gene expression states from the founder clone, A7A and B2F, either grew only to half the
222 extent of the founder or grew two times more efficiently (Fig. 2b, e) thus suggesting that the
223 transformation capacity of a given cell is somehow modulated by intrinsic molecular mechanisms
224 encoded within state-specific non-genetic networks.
225
226 Concordantly, when HepG2 clonal populations were analysed for their capability to grow in soft agar
227 in naïve conditions or upon TGFβ treatment, once again a divergence in their individual growth
228 capabilities could be observed as all the subclones analysed grew less efficiently than the founder
229 clone in naïve conditions, yet displaying variability among them (Supplementary Fig. 7d).
230 Interestingly, whereas most clones almost lost the ability to form colonies in soft agar upon TGFβ
231 treatment (1A, 32G, 44D), subclones 14F and 48E, which both display a divergent transcriptome
232 profile compared to the other three clones grew more efficiently in 3D soft agar compared to the
233 other subclones (Supplementary Fig. 7e).
234
235 Interestingly, when we subjected clonal populations of HA1ER and HepG2 and their corresponding
236 subclones to a transwell migration assay, once again we observed profound phenotypic variability
237 that did not segregate with the observed growth in 3D soft agar but was albeit different between the
238 subclones analysed and their respective founder clone (Fig. 2f, g, Supplementary Fig. 7b). Moreover,
239 as HA1ER cells are susceptible to trigger apoptosis upon treatment with Tumour necrosis factor-
240 related apoptosis-inducing ligand (TRAIL), we also tested our clonal HA1ER models for their capability
241 to withstand TRAIL challenge. Notably, although TRAIL treatment induced fractional killing in all
242 clones, the extent to which apoptosis was triggered differed between all subclones (1F8, 1C9, 2A6
243 and 1D12) and the founder clone (F12) (Fig. 2h). Collectively, our data reveals notable phenotypic

244 differences between genetically identical subclonal populations that appear to be profoundly
245 impinged by their underlying (state-specific) non-genetic makeup.
246
247 **Gene expression states are post-transcriptionally established and maintained**
248
249 Our data supports the notion that gene expression states are highly dynamic and shape phenotypic
250 output. These observations encouraged us to explore the underlying molecular mechanisms
251 contributing to transcriptome state dynamics, maintenance and divergence.
252 To address these, we set out to investigate whether changes in active transcription were driving
253 transcriptome divergence by initially analysing clones from the HA1ER systems (F12 lineage,
254 subclones 1F8 and 1C9). To do so, we performed SLAM-Seq[24,25] (shown to provide high-resolution
255 transcriptome dynamics data either at the transcriptional or RNA stability levels) to determine the
256 rate of active transcription and its contribution to steady-state RNA levels (defined as snapshot RNA
257 pool) in both clones. After sequencing, obtained reads were mapped to hg38 using a T>C aware
258 aligner (NextGenMap[26]) and further processed through the SLAM-DUNK pipeline[27]. Next, following
259 the determination of T-to-C conversion rates in each sequencing library as quality control (0.52-0.75
260 for all samples, robustly above background T-to-G conversion rate of 0.07, Supplementary Fig. 9a),
261 we proceeded to perform a pairwise comparison between both clones analysed (1F8 and 1C9) where
262 we applied the paired Beta binomial distribution (active transcription vs steady-state levels) to
263 determine whether differences at the nascent RNA level could be observed[25]. Strikingly, we did not
264 detect significant differences in the nascent RNA pool (FDR<0.05, FC>2) whilst at the steady-state
265 level, we observed 217 transcripts differentially expressed between both clones (FDR<0.01, FC>2;
266 Fig. 3a), suggesting that the observed differences in transcriptome state levels do not rely on
267 transcriptional processes, at least in the cellular system evaluated here.
268

269 Prompted by those results, we wondered whether our findings were a general phenomenon or if it
270 was an isolated case only valid for the 1F8 vs 1C9 comparison. Thus, we performed SLAM-Seq
271 analysis in the three clonal systems that we have previously shown (HA1ER, HA1E and HepG2) and
272 compared all subclones with their corresponding founder clone for each system (T-to-C conversion
273 rates: HA1ER 0.23-0.36, HA1E 0.18-0.34, HepG2 0.045-0.22, T-to-G background conversion rate: n.d.,
274 Supplementary Figs. 9b, 10). Strikingly, for most of the comparisons, no significant differences in
275 nascent transcription were observed (HA1ER: F12/1C9, HA1E: 10E/A7A, 10E/B5F, HepG2: 1A/14F,
276 1A/48E; Supplementary Fig. 10) whereas when transcriptional differences were detected (HA1ER:
277 F12/1F8, F12/2A6, F12/1D12, HA1E: 10E/A9H, 10E/B2F, HepG2: 1A/44D, 1A/32G), they only had a
278 slight-to-none impact on steady-state RNA levels and quite often the observed changes in steady-
279 state RNA (up or downregulation) did not follow the transcriptional trend (Supplementary Fig. 10b-d,
280 f-h, j-l).
281
282 To further validate our striking observation, we evaluated general chromatin organisation by
283 analysing chromatin accessibility by ATAC-Seq[28,29] in all the clonal systems herein presented (HA1E:
284 10E, A7A, A9H, B2F and B5F; HA1ER: F12, 1F8, 1C9, 2A6 and 1D12; HepG2: 1A, 32G, 14F, 48E and
285 44D). Notably, we did not detect substantial changes in chromatin accessibility between any of the
286 clones belonging to the same cellular system (HA1E: 178 598 peaks detected, HA1ER: 207 566 peaks
287 detected and HepG2: 232 711 peaks detected. Fig. 3b, Supplementary Fig. 11) thus supporting our
288 hypothesis that transcriptome state dynamics are not primarily controlled at the transcriptional level.
289
290 These data, obtained from three independent clonal cellular systems, suggest that steady-state RNA
291 levels and hence metastable gene expression states are not established/maintained by diverging
292 active transcription programmes but rather by a molecular machinery that operates at the post-
293 transcriptional level. Furthermore, our data from multiple systems supports the hypothesis that this
294 phenomenon, rather than an oddity, is a general biological feature.

295

296 **Altered transcriptomic states and phenotypic divergence between clones are not defined by major**

297 **changes at the (phospho-) proteome level**

298

299 The fact that transcriptomes diverge between cell lineages and that the observed divergence is
300 accompanied by profound changes in phenotypic output raises the question whether major
301 contributors to phenotype such as protein composition (e.g. translation) and/or signalling pathway
302 status (e.g. MAPK phosphorylation) also diverge. To test these hypotheses, we first analysed the total
303 protein content of clones 1F8 and 1C9 from the HA1ER system by label-free mass-spectrometry (LC-
304 MS/MS) and performed a pairwise comparison between both phenotype-diverging subclones.
305 Notably, from a pool of 3809 proteins analysed, only a small fraction was differentially expressed
306 (FDR<0.01, FC>1.5, Fig. 4a) with 23 upregulated and 34 downregulated proteins, suggesting that,
307 even though diverging clonal populations differ notably in their transcriptomes, they do not display a
308 concordant difference in the proteome. Moreover, gene ontology analysis of the differentially
309 represented proteins did not reveal enrichment in any biological pathways (Reactome v3.7[30],
310 Supplementary Fig. 12a) that could guide further hypotheses to explain the observed phenotypic
311 output variability in, for example, such a complex phenotype as anchorage-independent growth.
312 Therefore, these seemingly negative data prompted us to explore whether the diverging phenotypes
313 could be explained by an imbalance in signalling pathway output and therefore we performed a
314 phospho-proteomic analysis (TMT[31]) in the 1F8 and 1C9 populations (FDR<0.05, FC>0.4, Fig. 4c).
315 Surprisingly, the obtained data including 305 detected phospho-sites across 141 proteins revealed
316 only a minimal change in the phospho-proteome between both clonal populations with 33 phospho-
317 sites displaying enhanced phosphorylation and 24 displaying reduced phosphorylation levels across
318 only 28 proteins at bulk levels (Fig. 4c). Taken together, the remarkably little variance in the
319 (phospho-) proteome between two subclonal populations which exhibit notable divergence in

320 various phenotypic outputs encouraged us to look deeper into the data to elucidate the molecular
321 mechanisms underlying this non-genetic phenotypic heterogeneity.
322
323 **Clonally diverging lncRNAs localise and converge into perinuclear structures**
324
325 To advance our understanding of the molecular mechanisms underlying phenotypic divergence in
326 clonal cell populations, we endeavoured into annotating our various high-throughput datasets
327 including SLAM-Seq, label-free mass spectrometry and phospho-proteomic data based on their
328 molecular features. Intriguingly, a large fraction of the differentially expressed transcripts found by
329 comparing 1F8 and 1C9 SLAM-Seq data were lncRNAs with approximately 30% of them
330 corresponding to this category (FC>2. Fig. 3c, d). Given the ascribed regulatory role of lncRNAs in
331 various cellular processes[32], we evaluated the lncRNA content in our scRNA-Seq datasets where we
332 validated our initial observations even across all three clonal systems in which the lncRNA content
333 ranged from ~30-64% between various conditions including clonal divergence and TGFβ-induced
334 EMT (HA1ER, HA1E and HepG2. DE FC>3 for each pair-wise comparison included. Fig. 3e,
335 Supplementary Fig. 9c).
336
337 Given the generality of our observations across the evaluated clonal lineages from the 3 different cell
338 types, we combined data obtained from the SLAM-Seq assays with the scRNA-Seq datasets to build a
339 composite lncRNA catalogue from each clonal system that could be experimentally queried. To do so,
340 a subset of clone-specific lncRNAs were first evaluated by reverse transcription followed by qPCR (RT-
341 qPCR), then by subcellular fractionation followed by RT-qPCR and finally by RNA-FISH followed by
342 super-resolution microscopy.
343
344 Notably, from the composite subset (SLAM-Seq + scRNA-Seq) of lncRNA candidates obtained from
345 the HA1ER system, we were able to validate ~45% (RT-qPCR) of the differentially expressed lncRNAs

346 tested that diverged between 1C9 and 1F8 (Fig. 3f, Supplementary Fig. 9d). Moreover, from the HA1E
347 and the HepG2 clonal systems, we validated ~65% of the diverging lncRNAs by the same means
348 (Supplementary Fig. 9e, f). Furthermore, we found that most of the lncRNA transcripts validated as
349 divergent among states/clones display cytoplasmic localisation (subcellular fractionation followed by
350 RT-qPCR) in both HA1ER and HepG2 clones (Fig. 3g, Supplementary Fig. 9g). Following up on these
351 data, we moved forward to assess the intracellular localisation of these cytoplasmic clonally
352 diverging transcripts by RNA-FISH (Zeiss Airyscan) and verified that all the evaluated lncRNA
353 transcripts from the HA1ER and HepG2 systems indeed localised to the cytoplasm and more
354 precisely, into perinuclear structures reminiscent of molecular condensates (Fig. 5, Supplementary
355 Fig. 12a, b, RNA-FISH specificity control in Supplementary Fig. 12c).
356
357 Given that our observations suggest that a large fraction of clonally diverging lncRNAs may modulate
358 phenotypic output in a non-nuclear-linked manner and that most of the differentially expressed
359 coding RNAs are not accompanied by increased/reduced protein levels, we hypothesised that most
360 of the identified divergent RNAs (coding and non-coding) act as molecular scaffolds.
361
362 **Intrinsically disordered regions (IDR)-containing proteins clonally diverge and localize with peri-**
363 **nuclear lncRNAs**
364
365 Our mass spectrometry data (LC-MS/MS and Phospho-proteomics) demonstrated that only a small
366 subset of proteins was differentially represented between 1F8 and 1C9 clones, suggesting that
367 phenotypic divergence at the protein level may rely on discrete features such as intracellular
368 localisation and scaffolding properties, among others. At the same time, we have shown that the
369 cytoplasmic accumulation of divergent lncRNAs resembles the deposition of molecular condensates,
370 a phenomenon for which a key role for lncRNAs has been previously described in the nucleus[33-37].
371 Therefore, as IDR-proteins have been shown to act as scaffolding molecules contributing to the

372 formation of biomolecular condensates[38-40], we hypothesised that these subsets of proteins could
373 play a role in the nucleation/localisation of the identified perinuclear lncRNAs. To evaluate this
374 possibility, we moved on to validate a subset of the identified differentially expressed proteins
375 between clones 1F8 and 1C9 from the HA1ER system prioritising the annotation of IDR-containing
376 proteins, to which the majority of the diverging proteins could be ascribed to (~63%, Fig. 4b,
377 AlphaFold2 annotation[41,42]).
378
379 To initially validate the observed differential protein expression between HA1ER clones 1F8 and 1C9,
380 we set up a CyTOF analysis panel for a subset of differentially expressed proteins (IDR-proteins JNK-
381 interacting protein 4 (JIP4), α-Synuclein (SYUA), and monocarboxylate transporter 4 (MOT4), myelin
382 expression factor 2 (MYEF2), heat shock protein beta-6 (HSPB6), sequestosome-1 (SQSTM) and non-
383 IDR-protein L-xylulose reductase (DCXR)) to determine the levels of each protein in the HA1ER
384 founder clone F12 and in 1F8 and 1C9 subclones in a comparative manner. Notably, all the analysed
385 proteins labelled subpopulations within the F12 clone (Supplementary Fig. 13d, e) suggesting that
386 clonal divergence could also be traced at the proteome level even when the differentially expressed
387 proteins are only a minimal fraction of the total proteome. Next, we assessed a subset of these
388 proteins (JIP4, SYUA, MOT4, DCXR, PD-L1 (positive control) and CD206 (negative control)) for their
389 divergent expression in subclones 1F8 and 1C9 together with their founder clone F12. Whilst none of
390 the included proteins displayed cell cycle-dependent expression (Supplementary Fig. 13f), clone 1F8
391 contained a major cellular subpopulation where candidate antibody signal (JIP4, SYUA, MOT4 and
392 DCXR) is predominant and that was nearly absent in clone 1C9 (Fig. 4d-f, Supplementary Fig. 13b, c).
393
394 Given that a large fraction of the clonally diverging proteins belongs to the IDR-containing family and
395 could potentially act as molecular scaffolds, we next expanded our candidate list to accommodate an
396 additional diverging IDR-protein, gamma1-adaptin (AP1G1), and moved forward to evaluate whether
397 the diverging IDR-proteins and the observed lncRNAs enriched in the cytoplasm could indeed co-

398 localise within the observed perinuclear condensates. Therefore, we performed RNA-FISH coupled to
399 immunocytochemistry (ICC) for a subset of diverging lncRNAs and IDR-proteins and analysed their
400 intracellular distribution by super-resolution microscopy (Zeiss Airyscan). Interestingly, we could
401 show that in clonal lineages of HA1ER cells diverging lncRNAs and IDR-proteins co-localise in the
402 perinuclear region (lncRNAs: DPP4-DT (clone 1F8) and EMP2-AS (clone 1C9). IDR-proteins: JIP4,
403 SYUA), and only for IDR-protein AP1G1 could no such co-localisation be detected with either lncRNA
404 (Fig. 5a, Supplementary Fig. 12a, RNA-FISH specificity control shown in Supplementary Fig. 12c).
405 Moreover, we also validated the co-localisation of IDR-protein JIP4 and several diverging lncRNAs
406 (PINCR and LINC00494 (clone 14F), SCAT8 and LINC2476-203 (clone 44D)) in the perinuclear region in
407 the HepG2 clonal model (Supplementary Fig. 12b), for which the observations were in agreement
408 with data obtained from the HA1ER system.
409
410 **The Signal Integration Portal (SIP)**
411
412 So far, our findings have revealed that a subset of diverging lncRNAs and a subset of IDR-containing
413 proteins co-localise into perinuclear cytoplasmic structures in a variety of clonally diverging cellular
414 models. However, none of our findings indicated a functional/effector link that could explain the
415 observed divergence in phenotypic output for a given complex cellular behaviour such as the
416 capability to grow in anchorage-independent conditions, migration or fractional killing. Therefore, we
417 hypothesised that cell-to-cell variability in phenotypic output relies on the re-arrangement of
418 signalling pathways following different configurations.
419
420 To further explore this concept, we contemplated the possibility that the peri-nuclear structures that
421 we have identified could nucleate signalling components that would alter phenotypic output. Given
422 that we have merely detected 28 statistically significant differentially phosphorylated proteins
423 between our flagship HA1ER clones 1F8 and 1C9, and for which phospho-site specific antibodies are

424 not available, we decided to test whether multiple activated signalling pathways would converge
425 within the observed peri-nuclear structures. To do so, we built a panel of phospho-site specific
426 antibodies directed towards multiple major signalling pathways including RAF/MEK/ERK, PI3K/AKT,
427 IκBα/NFκB and MKK4/p38, and performed RNA-FISH (targeting peri-nuclear localised diverging
428 lncRNA DPP4-DT) coupled to ICC (phospho-site specific; Supplementary Table) in clone 1F8.
429 Strikingly, after super-resolution imaging analysis, we observed that multiple active signalling
430 pathways (effector proteins or phospho-proteins) were indeed localised in the peri-nuclear
431 structures delineated by the lncRNA DPP4-DT (pARAF, pBRAF, pMEK1/2, pERK1/2, pRAC1, pMKK4, p-
432 p38, pPI3K, pAKT, pPAK1/2, pIκBα and pNFκB. Fig. 5b, Supplementary Fig. 14a). Intriguingly, rather
433 unrelated active signalling pathways coalesce into these cytoplasmic condensates suggesting that the
434 observed structures could function as a signal organiser to determine phenotypic output and thus
435 we termed this compartment the Signal Integration Portal (SIP).
436
437 To further verify the existence of the SIP and taking advantage of the capability of the HepG2 model
438 to undergo TGFβ-induced EMT, we decided to evaluate if members of the TGFβ signalling pathway
439 would converge into SIP structures delineated by HepG2-specific diverging lncRNA PINCR in clone
440 14F by RNA-FISH coupled to ICC. To do so, we included members of the canonical TGFβ pathway
441 (SMAD2 and SMAD3) in addition to members of the non-canonical pathway also reported to become
442 active by the presence of TGFβ (RAC1 and ROCK1) and evaluated their total levels and those of their
443 corresponding phosphorylated active forms. As expected for TGFβ-naïve conditions, PINCR localised
444 to the peri-nuclear region in a punctuated pattern across most of the cells (Fig. 5c, Supplementary
445 Fig. 14b) whilst most of the components of the TGFβ signalling pathway and their phosphorylated
446 counterparts displayed a diffuse cytoplasmic localisation (Fig. 5c, Supplementary Fig. 14b).
447
448 However, upon TGFβ treatment, PINCR (proxy for SIP particles) co-localised with active members of
449 the TGFβ signalling cascade including pSMAD2, pSMAD3 and pRAC1 (Fig. 5c, Supplementary Fig. 14b)

450 suggesting that signalling pathway components are brought together to the SIP to integrate
451 biological cues and fine-tune phenotypic output. Interestingly, whilst ROCK1 was observed within SIP
452 condensates, we did not detect pROCK1 (Tyr913) suggesting that either this phospho-protein is not
453 recruited to the SIP or the epitope is masked within SIP-condensates (Supplementary Fig. 14b).
454
455 Altogether, we showed that in phenotypically diverging clonal populations of fully committed and
456 differentiated cellular models (HA1ER and HepG2) lncRNAs, IDR-containing proteins and active
457 signalling effectors from multiple signalling cascades coalesce into a peri-nuclear structure which we
458 termed Signal Integration Portal (SIP) that sits at the crossroads of phenotypic divergence nucleating
459 inputs from the nucleus (lncRNA), the cytoplasm (IDR-proteins and signalling cascades) and the
460 environment (signalling cascades).
461
462 **Remodelling cell states by altering the SIP lncRNA composition**
463
464 We have already shown that diverging molecules from different families (lncRNAs, IDR-proteins,
465 signalling proteins) converge into peri-nuclear SIP structures which may be involved in the
466 establishment/propagation of transcriptome states and in dictating phenotypic output. Therefore, to
467 further our claims, we hypothesised that altering the SIP composition may result in transcriptome re-
468 arrangements. To do so, we designed antisense oligonucleotides (G, gapmers) targeting a subset of
469 SIP-associated lncRNAs identified in the diverging HA1ER clones 1F8 and 1C9 (DPP4-DT, AL161733.1
470 and AC008464.1) and in HepG2 clones 14F and 44D (PINCR and SCAT8). Importantly, to strengthen
471 our claims, we designed our assays to include a non-targeting control gapmer and two gapmers
472 targeting different regions per target transcript (when possible) to evaluate knockdown effects in a
473 confident manner. After verifying successful knock-down of the evaluated lncRNAs from either
474 system by RT-qPCR (Fig. 6e, Supplementary Fig. 15c), cells from each experimental condition were
475 processed for scRNA-Seq (10x Genomics Chromium 3'). After sequencing, obtained data was pre-

476 processed to mitigate clustering artefacts due to cell cycle progression and abnormal mitochondrial
477 counts and clustered using Seurat. Strikingly, upon knock-down of the assessed lncRNAs we observed
478 transcriptome re-arrangements relative to the non-targeting control (ctrl) that ranged from the loss
479 of a single transcriptome state whilst largely preserving transcriptome heterogeneity (PINCR-G:
480 cluster 8; HepG2. Supplementary Fig. 15b, g), the reduced representation of certain states and the
481 emergence of novel ones whilst preserving a certain degree of the original transcriptome states
482 (AL161733.1: cluster 10 reduced, clusters 6 and 9 novel for G1, clusters 0, 11 and 12 novel for G2;
483 HA1ER. SCAT8: cluster 8 reduced, cluster 10 novel for G1, cluster 5 novel/increased for G2; HepG2.
484 Fig. 6a, Supplementary Fig. 15b, f, g) to a complete reprogramming of the metastable transcriptomic
485 state landscape (AC008464.1-G and DPP4-DT-G1 and -G2; HA1ER. Fig. 6a, Supplementary Fig. 15a, f).
486 Notably, while both gapmers targeting the same lncRNA (DPP4-DT and AL161733.1) led to distinct
487 transcriptome state re-arrangements which might partly be due to differences in knockdown
488 efficiency, we detected specific transcriptomic markers concordantly upregulated by both gapmers,
489 indicating common intracellular effects (Fig. 6e, Supplementary Fig. 15d). These findings suggest that
490 SIP-bound lncRNA levels harbour the capability to remodel the transcriptome landscape and hence
491 modulate phenotypic output.
492
493 **Cell state remodelling is accompanied by profound changes in signalling**
494
495 To further explore the phenotypic consequences of SIP disruption and transcriptome remodelling in
496 the HA1ER model, we evaluated whether major signalling pathways would be altered upon a given
497 lncRNA knockdown. To do so, we took advantage of a pathway-specific antibody panel (180
498 antibodies; Supplementary Table) and evaluated signalling pathway remodelling by Reverse Phase
499 Protein Array (RPPA[43]) in the HA1ER clone 1F8 (increased expression of lncRNAs AL161733.1 and
500 DPP4-DT when compared to 1C9 clone). Upon lncRNA knockdown (~50-94% KD when compared to
501 the non-targeting control, Fig. 6e), we first validated transcriptome remodelling by evaluating a

502 subset of identified expression markers that become upregulated under these conditions by RT-qPCR
503 (MDM2 and CDKN1A. Supplementary Fig. 15d, e) and then proceeded to run the RPPA assay.
504
505 Strikingly, upon lncRNA AL161733.1 knockdown, from the 155 antibodies for which we observed a
506 robust signal, we found 41 upregulated signals whilst 16 were downregulated with any of the
507 antisense gapmers when contrasted to the non-targeting control (≥20% increase/decrease to control)
508 and which showed the same trend in both gapmers (Fig. 6b bottom). Moreover, when knock-down of
509 lncRNA DPP4-DT was evaluated, 50 signals were upregulated whilst 25 were downregulated with any
510 of the gapmers with the same trend for both gapmers (Fig. 6b top, ≥20% increase/decrease to
511 control). In ~90 % of identified up- or down regulated targets both gapmers showed the same trend,
512 indicating that similar intracellular effects at the signalling pathway level are triggered by both
513 gapmers targeting the same lncRNA.
514
515 Notably, as we hypothesised that the lncRNAs evaluated are involved in the integration of signalling
516 information by nucleating pathways, we annotated each of the assessed antibodies with their
517 respective pathways (Reactome, Top50 Terms, FDR<0.01). For this, we extracted the upregulated and
518 downregulated antibody signals with specific changes for lncRNA AL161733.1 and lncRNA DPP4-DT
519 knockdowns, as well as the list with overlapping changes upon knock-down of lncRNA AL161733.1
520 and lncRNA DPP4-DT (Fig. 6c, d, Supplementary Fig. 15h). Outstandingly, separate knock-down of
521 both lncRNAs led to a modulation of common pathways (overlap), such as upregulation of signalling
522 by ALK in cancer, PI3K/AKT signalling and signalling by LTK, whereas the AP-2 (TFAP2) family of
523 transcriptional regulators, several developmental lineages (mammary gland, integumentary system)
524 and signalling by KIT in disease, amongst others, were downregulated (Fig. 6c, Supplementary Fig.
525 15h), suggesting that both lncRNAs contribute to a general mechanism of signalling pathway control.
526 However, knocking down lncRNA AL161733.1 also resulted in the modulation of specific pathways
527 such as the upregulation of NOTCH signalling and downregulation of signalling by TGFβ family

528 members and signalling by WNT (Fig. 6c, d, Supplementary Fig. 15h), whilst knocking down lncRNA
529 DPP4-DT resulted in the modulation of a completely different subset of pathways, including the
530 deregulation of multiple transcription-regulating pathways (Fig. 6c, d, Supplementary Fig. 15h), thus
531 suggesting the existence of an exquisite mechanism of signalling pathway control where individual
532 lncRNAs modulate a defined set of pathways in a specific manner.
533
534 Finally, the observation that major – unrelated – signalling pathways were indeed altered upon
535 lncRNA knockdown suggests that by solely adjusting SIP composition through the reduction of SIP-
536 linked lncRNA levels, both transcriptome re-arrangement and, as hypothesised, major remodelling of
537 signalling pathways occurs.
538
539 **Cell state remodelling is accompanied by signalling pathway reorganisation/redistribution and**
540 **altered phenotypic output**
541
542 Our data suggest that SIP-bound lncRNAs are major contributors to phenotypic output as altering SIP
543 composition using lncRNA knockdown results in signalling pathway rewiring. Therefore, if SIP
544 complexity is at the crossroad of transcriptome heterogeneity, signalling organisation/integration and
545 phenotypic output, altering SIP composition should result in profound changes in phenotypic output.
546 Therefore, to evaluate this hypothesis, we performed a series of phenotypic assays where we took
547 advantage of the efficient gapmer knockdown and its observed consequences on transcriptome
548 rearrangement and signalling pathway rewiring (Fig. 6, Supplementary Fig. 15).
549
550 First, given the profound changes in signalling pathway organisation upon DPP4-DT knockdown in the
551 1F8 clone from the HA1ER system, we set out to evaluate whether signalling pathway components
552 would relocalise within the cell under these conditions and performed RNA-FISH followed by ICC for
553 a subset of SIP-associated signalling proteins (pRAC1, pPAK1, pMKK4 and pMEK1/2). Strikingly, upon

554 DPP4-DT knockdown in the HA1ER 1F8 clone, we could evidence redistribution of pRAC1 from SIP
555 condensates to the nuclei, pPAK1 accumulation in the cytoplasm, a pronounced build-up of pMKK4
556 in the perinuclear area accompanied by the loss of distinct punctate structures and the coalescence
557 of pMEK1/2 into a different kind of perinuclear accumulation reminiscent of biomolecular
558 condensates (Fig. 7a, Supplementary Fig. 16) thus supporting the notion that SIP-condensates
559 contribute to signal integration by acting as a signal organising hub. Encouraged by these results, we
560 further set out to evaluate if SIP-linked lncRNA knockdown would affect, what we believe to be one
561 of the most extreme phenotypes under SIP influence, transformation potential. Therefore, we
562 knocked down lncRNA AC008464.1 in clone 1C9 of the HA1ER system (Fig. 7c), which displayed
563 increased clonogenicity when compared to its parental clone F12 or sibling clone 1F8 as established
564 in previous soft agar assays (Fig. 2d, Supplementary Fig. 7c). Obtained data revealed that altering SIP-
565 linked lncRNA AC008464.1 levels reduced clonogenic potential of clone 1C9 to ~53% when compared
566 to non-targeting control gapmer treated cells (Fig. 7c) suggesting that the clonogenic potential of
567 clone 1C9 is severely impaired when SIP composition is compromised.
568
569 Taken together, we propose the Signal Integration Portal (SIP) as a peri-nuclear molecular entity
570 where biological information from various non-genetic compartments including cytoplasmic lncRNA
571 content, intrinsically disordered region-containing proteins and active signalling pathways converge
572 (Fig. 7d) to modulate phenotypic output, thus controlling cell plasticity and the degree of non-
573 genetic heterogeneity for a given cellular system.
574
575 **Discussion**
576
577 Cell plasticity, an inherent characteristic of all known organisms, is one of the most widespread
578 phenomena observed in biology. However, though a myriad of hypotheses have been brought
579 forward, very little is known about the molecular mechanisms underlying this phenomenon.

580 Paradoxically, years of research centred around the genetic compartment could not explain the
581 plastic behaviour displayed by cells and organisms overall. However, a multitude of examples
582 challenging the role of the genome as a master controller underlying this phenomenon have come to
583 light including bet-hedging in bacteria[44,45], protozoan development/evolution[46-49] and the emergence
584 of resistance to anti-cancer therapy[50-54]. Remarkably, the advent of multimodal single-cell
585 technologies opened a new dimension of biological insight that fuelled the in-depth exploration of
586 the non-genetic compartment and has enabled the study of the molecular determinants underlying
587 cell plasticity[8,55-60]. In that regard, we and others ventured into exploring transcriptome content as a
588 determinant of cell plasticity in mammalian cellular systems and found that indeed steady-state
589 transcriptome heterogeneity harbours sufficient biological information to drive divergent phenotypic
590 output even in fully controlled clonal cellular systems, thus fuelling molecular and functional
591 heterogeneity within populations[8,47,61]. Strikingly, in all the models that we have tested so far, the
592 contribution of active transcription to transcriptome heterogeneity is negligible and does not
593 support a model where transcriptome divergence is instructed by changes in gene expression
594 programmes (as defined by alterations in chromatin structure resulting in different subsets of genes
595 being actively transcribed). Instead, transcriptome plasticity seems to rely on post-transcriptional
596 regulatory modules wherein a subset of cytoplasmic lncRNAs and intrinsically disordered region-
597 containing proteins (IDR-proteins) play a key role throughout phenotypic divergence and are
598 enriched in subsets of cellular states. Both observations challenge the status quo as, to begin with,
599 lncRNA function has been ascribed mainly to the nucleus where they, in principle, would contribute
600 to genome control (active transcription, chromatin architecture, among others)[32,62] whilst our data
601 supports a key role for lncRNAs within the cytoplasm. Consequently, our data does not support a
602 direct role for lncRNAs in genome control as the primary driver of transcriptome divergence, whilst
603 highlighting marked differences between nascent and steady-state RNA pools. The latter observation
604 comes as no surprise as it has been acknowledged for decades that the contribution of active

605 transcription to steady-state RNA levels is rather modest and that major players in the regulation of
606 transcriptome content operate at a post-transcriptional level[63-66].
607 An important aspect of our study stems from the fact that to explore non-genetic divergence and
608 evolution we required - first and foremost - to characterise the level of genetic divergence in each
609 cellular population to ultimately derive genetically stable cellular models that would allow us to
610 evaluate the non-genetic compartment avoiding genetic variation confounding factors. In that
611 context, HA1E and HA1ER cell systems were an ideal choice as previous reports and our own
612 observations confirmed that these cellular models accumulate over time negligible genetic
613 alterations[67]. Together with our HepG2 clonal system, for which genetic stability has also been
614 verified, our catalogue of isogenic cell systems comprises non-transformed (HA1E) and transformed
615 cells originating from distant tissue-of-origin (HA1ER, kidney and HepG2, liver) and therefore
616 displaying pronounced differences in their non-genetic state repertoire. Our coherent observations in
617 unrelated cellular systems suggest that our findings can be juxtaposed to any cellular system and are
618 a biological general feature rather than an odd occurrence.
619
620 Following this line of thought, we argue that cells thriving in vivo could be constantly rewiring their
621 non-genetic landscape to accommodate their individual needs and/or the needs of the tissue and
622 that cell plasticity is a pervasive feature across organismal biology. It follows that cell plasticity as
623 defined in our report, may be wired to adapt to a myriad of both local and systemic signals acting on
624 the tissue of origin (e.g. differentiation signals, immune challenge, therapeutic pressure) and
625 respond in turn by expressing a variety of phenotypes to increase population fitness and thus
626 securing their perpetuation. Importantly, our findings corroborate the concept that cell plasticity may
627 underlie a fundamental constrain for therapeutic development as cell plasticity is ever-present and
628 often overlooked in bulk analyses, resulting in a large panoply of phenotypic outputs often obscuring
629 data interpretation. For instance, whilst some of the cellular states observed in a given cellular
630 system or tissue could be therapy susceptible, the possibility to find intrinsic resistant states already

631 present within the cellular population is almost a certainty. Moreover, as many state-of-the-art anti-
632 cancer therapies are based on genotoxic attack, the risk associated with generating multiple levels of
633 cell plasticity – a palette of metastable states associated with each drug-induced genotype –
634 complicates the outlook further.
635
636 Strikingly, our data supports a model where cell plasticity is not infinite but defined by a limited
637 number of predominant states that are characterised by state-specific SIP condensates, which
638 prompts the question about how metastable states are inherited and propagated throughout the
639 progeny. An obvious model would immediately consider the asymmetric inheritance of (SIP-)
640 components as a centre piece underlying cell plasticity. In that regard, asymmetric cell division and
641 asymmetry in the inheritance of RNA molecules and organelles has been widely documented as
642 contributing to phenotypic output and stem-cell differentiation[68-71], key mechanisms to keep in mind
643 whilst delving further into SIP biology.
644
645 Interestingly, our model suggests that transcriptome organisation and divergence could result or be
646 directed by a re-organisation of the molecular players involved in the process and is mainly
647 controlled by their intracellular localisation. This exquisite molecular mechanism would ensure the
648 generation of molecular and cellular heterogeneity by minimal energetic means and could provide
649 the basis to maintain a basal level of phenotypic divergence within clonal populations. Moreover, it
650 brings to light the observation that identity (i.e. cell type) and phenotypic output are not necessarily
651 coupled and that cells, through the means of cell plasticity, can dramatically follow entirely different
652 paths despite their identity. This concept opens a new dimension of biological insight as for instance
653 in cancer research, most of the single cell analyses are based on cell type annotation (using hallmark
654 markers) and not on what the cells are indeed doing in their context. Therefore, there is a clear need
655 to further characterise and distinguish between cell identity and cell activity (expression of the
656 phenotype) as it could completely alter the current view underlying how cellular systems operate.

657

658 Finally, further characterisation of SIP components and architecture, perhaps linked with a distinct
659 set of phenotypes of interest, is expected to reveal suitable targets for SIP disruption or SIP
660 reorganisation (i.e. lncRNAs, IDR-proteins, others) which in turn would allow the targeted fine-tuning
661 of cellular states towards a desired phenotype (e.g. anti-cancer drug susceptible). However, it is our
662 view based on observations made by us and others over the years[1,2,8,18,53,72-74], that cell plasticity
663 cannot be reduced/removed from a system, as plastic cell phenotypes and the dynamic intracellular
664 molecular landscape impinges a fundamental biological property through which cell populations
665 adapt when subjected to intra-/extracellular environmental changes. Therefore, the identification of
666 molecular switches that would allow the controlled, targeted and non-reversible transition to a
667 distinct – advantageous – phenotype might be a cumbersome task to achieve.

668

669 **Materials & Methods**

670

671 **Cell lines**

672 Immortalised HA1E (hTERT and SV40 ER) and tumorigenic HA1ER cells (hTERT, SV40ER and HRAS-
673 G12V) from stepwise tumorigenesis models generated from normal human embryonic kidney cells
674 were obtained from Prof. Weinberg (Broad Institute of MIT and Harvard, Cambridge, USA[19]). HepG2
675 cells were obtained from ATCC (ATCC, cat. HB-8065). HA1E and HA1ER cells were cultured in DMEM
676 (1g/L glucose) (Gibco, cat. 21885025) supplemented with 10% heat-inactivated foetal calf serum
677 (FCS, Gibco) and 100µg/ml hygromycin (hTERT selection, Thermo Fisher Scientific, cat. 10687010),
678 400µg/ml G418 (SV40-ER selection, Thermo Fisher Scientific, cat. 10131035), 0.5µg/ml puromycin
679 (HRAS-G12V selection, Thermo Fisher Scientific, cat. A1113803) and 50 µg/ml gentamicin. HepG2
680 cells were cultured in DMEM high glucose (4.5g/L glucose) (Thermo Fisher, cat. 41966029)
681 supplemented with 10% heat-inactivated FCS (Gibco) and 50U/ml Penicillin-Streptomycin (Thermo
682 Fisher, cat. 15070063). Single cells were isolated by flow cytometric cell sorting (BD FACSAria III; BD

683 Biosciences) and expanded in standard 2D culture conditions to establish clonal populations. LentiX
684 293T were obtained from Prof. Lacaud (Cancer Research UK Manchester Institute, Manchester, UK)
685 and cultured in DMEM (1g/L glucose) (Gibco, cat. 21885025) supplemented with 10% heat-
686 inactivated FCS (Gibco). HA1E and HA1ER cells were plated at a confluency of 1 x
687 10^6 cells/6010mm^2 surface with 10ml fresh medium, passed every 48 hours and kept in culture for a
688 maximum of 15 passages. HepG2 cells were plated at a confluency of 1.5 x
689 10^6 cells/6010mm^2 surface with 10ml fresh medium, passed every 48 hours and kept in culture for a
690 maximum of 15 passages. All cell lines were grown at 37°C with 5% carbon dioxide (CO_2) saturation.
691 Mycoplasma contamination was analysed by reverse transcription followed by quantitative
692 Polymerase Chain Reaction (RT-qPCR) (VenorGeM qEP, cat. 11-9100).
693
694 **Lineage tracing libraries (LTv-BC-H2B-GFP)**
695 The episomal vector GBX (EBNA1/OriP) was a gift from Linzhao Cheng (Addgene, cat. 64123) that
696 was engineered to build a low-copy propagating vector in bacterial hosts (*bom* and *rop*) and with the
697 potential to express H2B-GFP fusion protein in mammalian systems (LTv-H2B-GFP). Using this
698 template, Barcode decay Lineage Tracing (BdLT-Seq) plasmids were generated by introducing a
699 molecular barcode into the 3'UTR of H2B-GFP sequence. Briefly, the 3'UTR of the LTv-H2B-GFP
700 vector was PCR amplified using a primer that contains a 12 nucleotide long random sequence split
701 into three segments by intervening nucleotides (NNNNGCGNNNNTGANNNN). Next, PCR amplicons
702 were cloned into the LTv-H2B-GFP vectors using HindIII and XbaI restriction sites (LTv-BC-H2B-GFP
703 vectors). The generated high-complexity library was propagated using electrocompetent MegaX
704 DH10B T1 bacteria strain (Invitrogen, cat. C640003). The final LTv-BC-H2B-GFP lineage tracing
705 episome library contained approximately 1.2 x 10^6 unique barcodes and was already validated and
706 used in our previous study[8].
707
708 **Transfection with the LTv-BC-H2B-GFP episome libraries**

709 Transfection of HA1ER cells with LTv-BC-H2B-GFP lineage tracing episome library was performed
710 following standard direct transfection protocol using Lipofectamine LTX reagent (Invitrogen, cat.
711 153385000). The expression of H2B-GFP was verified by fluorescent microscopy (Thermo Fisher
712 Scientific, cat. AMEX1000) and flow cytometry (BD LSRFortessa X-20 Cell Analyzer, BD Biosciences)
713 24- or 48-hours post-transfection depending on experimental set up (usually 25-35% of H2B-GFP
714 positive cells). Flow cytometry data was analysed using FlowJo software (BD, version 10.8.0).
715
716 **BdLT-Seq**
717 A detailed BdLT-Seq protocol can be found in our published protocol[75]. HA1ER cells (clones F12, 1F8
718 and 1C9) were transfected with LTv-BC-H2B-GFP episome library as described. Forty-eight hours
719 post-transfection GFP positive cells were sorted using BD FACSAria III Cell Sorter (BD Biosciences) or
720 BD InFlux Cell Sorter (BD Biosciences) aiming for >94% purity after sorting. Cells were expanded in
721 standard culture conditions for five days – Phase I – a barcode decay phase in which most of the
722 episomes are lost upon cell division. The second barcode decay phase – Phase II - begins
723 approximately 5-7 days after transfection where the episomes stabilize however still displaying a
724 characteristic episome loss. To perform BdLT-Seq (only during phase II), cells were trypsinised
725 (Invitrogen, cat. 25300104), collected by centrifugation at 230 rcf, 20°C, and plated at 2,000 to 3,000
726 cells per well (0.32cm^2) according to experimental conditions. Five to six days later (~ 80,000 cells),
727 cells from each clonal population were split into two samples: sample N1 ("Day 0") was stored in
728 liquid nitrogen (90% FCS, 10% DMSO) whereas sample N2 ("Day 4") was expanded in standard
729 culture conditions for four additional days. Four days later, cells were stored in liquid nitrogen -
730 sample N2 ("Day 4") (90% FCS, 10% DMSO). The day before single-cell RNA sequencing, cells were
731 returned to standard 2D culture for 24 hours. The following day, samples were processed for scRNA-
732 Seq sequencing (10x Genomics Chromium, Single-Cell RNA-Seq System, version 3.1, cat. PN-
733 1000269) coupled to lineage tracing (BdLT-Seq). Briefly, cells were collected by trypsinisation,
734 washed in 5% FCS PBS and resuspended in 5% FCS PBS. Hashtag antibodies were used to specifically

735 label individual samples for multiplexing purposes. For that, cells were incubated with the hashtag
736 antibodies coupled to a specific oligonucleotide (HTO) (BioLegend, cat. A0251-A0260) at a ratio of
737 0.5 µg per 50,000 cells (stock 10 µg/ml) for 30 min, on ice. Cells were then washed three times using
738 0.04% BSA PBS and counted using Neubauer counting chamber (Marienfeld, cat. 0640030). Samples
739 were then combined according to the experimental setting and 20,000 cells per channel (for the
740 expected outcome of 10,000 cells/channel) were loaded onto 3 channels of the 10x Chromium
741 platform (10x Genomics, cat. 1000204). Samples were processed following standard 10x Chromium
742 protocol (10x Genomics, version 3.1, cat. PN-1000269) up until cDNA amplification reaction. cDNA
743 amplification reaction was supplemented with the HTO additive primer and followed by a cDNA
744 clean-up separating the HTO fraction from the cell cDNA using standard methodologies (Beckman
745 Coulter, AMPure XP, cat. A63880). All size selected fractions (cell cDNA and HTO cDNA) were
746 assessed by Fragment Analyser 5200 (Agilent, cat. M5310AA). Next, HTO depleted cDNA fraction
747 (cell cDNA) was used to capture H2B-GFP cDNA. For that, 5µl of Dynabeads MyOne Silane beads
748 (Invitrogen, cat. 37002D) were washed twice in 500µl RLT lysis buffer (Qiagen, cat. 79216),
749 resuspended in 140µl RLT buffer (3.5x volume of initial sample) and added to 40µl of the HTO
750 depleted cDNA fraction. Then 180µl of absolute ethanol was added to the reaction (4.5x volume of
751 initial sample), mixed well and incubated for 15 minutes at room temperature in a ThermoMixer C
752 (300rpm, Eppendorf, cat. EP5382000015). The supernatant was discarded, and the beads were
753 washed three times in 80% ethanol, dried for 3-4 minutes at room temperature, and eluted in 5µl
754 nuclease free water (10-15 minutes at room temperature, ThermoMixer C at 300rpm). To capture
755 H2B-GFP cDNA, a set of 4 sequence-specific biotinylated probes targeting H2B-GFP was used. The
756 probes were hybridised H2B-GFP cDNA using xGen Hybridization and Wash Kit (IDT, cat. 1080577)
757 following manufacturer's instructions. Briefly, 8.5µl of Gen 2x Hybridization buffer was combined
758 with 2.7µl of Gen Hybridization enhancer, 5µl cDNA fraction and 2µl of the probes (4 pmoles, pool of
759 4 probes), mixed, incubated for 10 minutes at room temperature, followed by the incubation in a
760 thermocycler: 30 seconds at 95°C and 16 hours at 65°C (lid temperature 105°C). Following

761 hybridisation, the H2B-GFP cDNA fractions were pulled down using streptavidin beads. Twenty
762 microliters of streptavidin beads were washed twice using 100µl xGen 1x Bead wash buffer and then
763 resuspended in 17µl bead resuspension buffer (8.5µl xGen 2x Hybridization buffer, 2.7µl Gen
764 Hybridization buffer enhancer, 5.8µl nuclease free water). Seventeen microliters of beads from the
765 previous step were added to the hybridisation reaction and incubated in a thermocycler at 65°C for
766 45 minutes (lid temperature 70°C) mixing the reaction every 10 minutes. After incubation, the
767 samples were washed following manufacturer's instructions. Briefly, 100µl of 65°C wash buffer 1
768 was added to the sample and carefully mixed. The supernatant was transferred into a new tube and
769 kept on ice (cDNA for 10x gene expression library – GEx fraction). Beads were then washed by
770 adding 150µl of 65°C stringent buffer, mixed well and incubated in a thermocycler for 5 minutes at
771 56°C (lid temperature 70°C). Next, the supernatant was discarded, and the previous step was
772 repeated for a total of two washes. 150µl of wash buffer 1 was added to the beads, incubated for 2
773 minutes at room temperature, supernatant was then discarded, 150µl of wash buffer 2 was added
774 to the beads and incubated for 2 minutes at room temperature. After discarding the supernatant,
775 150µl of wash buffer 3 was added to the beads and incubated for 2 minutes at room temperature.
776 To elute H2B-GFP cDNA from the beads, 100µl of 100mM NaOH 0.1mM EDTA was added to the
777 complex beads/cDNA and incubated 15 minutes at room temperature in a ThermoMixer C (300
778 rpm). The supernatant was transferred into a new tube and 10µl 1M HCl, 10µl 1M Tris-HCl (pH 7.5)
779 was added. Both H2B-GFP cDNA fractions (lineage tracing fraction) as well as the fraction that
780 contained the rest of the cDNA for the 10x gene expression library (GEx fraction) were purified using
781 Dynabeads MyOne Silane beads (Invitrogen, cat. 37002D) following the protocol described above.
782 10x GEx libraries were generated using 120ng of the cDNA fraction following 10x Chromium protocol
783 (Fragmentation, End repair and Library construction steps). H2B-GFP and HTO libraries were
784 generated by amplifying either H2B-GFP or HTO cDNA using primers that specifically recognise H2B-
785 GFP or HTO on the 5'-end and Read 1 on the 3'-end, generating standard Illumina paired-end
786 sequencing constructs containing P5 and P7 (Supplementary Table 1). Final sequencing library

787 quality was assessed by Fragment Analyser 5200 (Agilent, cat. M5310AA). Libraries were sequenced
788 on a NovaSeq 6000 version 1.5 (Illumina, cat. 20013850). Sequencing depth: GEx library –
789 35000/50000 read pairs per cell, HTO library – 2000/3000 read pairs per cell, GFP and mCherry
790 library – 10000/12000 read pairs per cell. Sequencing type: paired-end, dual indexing. Sequencing
791 read: Read 1 – 28 cycles, i7 Index – 10 cycles, i5 Index – 10 cycles, Read 2 – 90 cycles.
792
793 **10x Gene expression libraries and sequencing**
794 10x libraries for HA1E (10E, A7A, A9H, B2F and B5F clones), HA1ER (F12, 1D12, 1C9, 2A6, 1F8, 3C5, 1E8,
795 2B12 and 1F5 clones) and HepG2 (1A, 32G, 14F, 48E and 44D clones) and for cells following gapmer-
796 mediated knock-down were prepared following manufacturer's instructions. In order to perform a
797 direct comparison of scRNA-Seq output avoiding mathematical integration cell preparations from each
798 cell system analysed were multiplexed using HTOs (BioLegend, cat. A0251-A0260) prior to library
799 construction. Libraries were sequenced on a NovaSeq 6000 version 1.5 (Illumina, cat. 20013850).
800 Sequencing depth: GEx library – 35000/50000 read pairs per cell, HTO library – 2000/3000 read pairs
801 per cell. Sequencing type: paired-end, dual indexing. Sequencing read: Read 1 – 28 cycles, i7 Index –
802 10 cycles, i5 Index – 10 cycles, Read 2 – 90 cycles.
803
804 **Cell cycle analysis**
805 Cells were plated at a confluency of 1 x 10^6 cells/6010mm^2 surface. Twenty-four to 48 hours later,
806 cells were collected, fixed and permeabilised with methanol. Cell cycle progression was assessed
807 using DAPI nuclear counterstain (Sigma-Aldrich, cat. MBD0015) and analysed by flow cytometry (BD
808 LSRFortessa X-20 Cell Analyzer, BD Biosciences or Attune Nxt, Thermo Fisher Scientific).
809
810 **Bright field imaging**

811 Cells were plated at a confluency of 1-1.5 10^6 cells/10cm Petri dishes (6010mm^2 surface). Twenty-
812 four hours later, cells were imaged using a fluorescent microscope (Thermo Fisher Scientific, cat.
813 AMEX1000).
814
815 **Cell proliferation**
816 Cells were plated in standard 2D culture conditions at a defined number per well (HA1ER: 50,000
817 cells per 9.6cm^2, HA1E: 120,000 cells per 9.6cm^2, HepG2: 150,000 cells per 9.6cm^2) and over a period
818 of 4 days, three wells per clone were harvested each day and counted manually using a Neubauer
819 counting chamber.
820
821 **Anchorage-independent growth in 3D soft agar**
822 0.6% and 1% Difco agar noble solutions were prepared (BD Biosciences, cat. 214230) and autoclaved.
823 Agar solutions were melted and equilibrated to 42°C in a water bath together with 2X growth
824 medium (HepG2: DMEM high glucose (Sigma-Aldrich cat. D7777), 0.7% sodium bicarbonate (Sigma-
825 Aldrich, cat. S8761), 20% FCS and 100U/ml Penicillin-Streptomycin (Thermo Fisher, cat. 15070063);
826 HA1ER/HA1E: DMEM (Sigma-Aldrich cat. D5523), 0.7% sodium bicarbonate, 20% FCS and 100 µg/ml
827 gentamicin (Thermo Fisher, cat. 15750037)). The lower 0.5% agar layer was generated by mixing the
828 1% agar solution 1:1 with the 2X medium and 1.5ml was added to each well of 6-well plates. After
829 solidification of the bottom layer, the top 0.3% agar layer was prepared by mixing the 0.6% agar
830 solution 1:1 with the 2X medium containing a fixed number of cells in a single-cell suspension
831 (HA1ER: 10,000 cells/well, HA1E: 75,000 cells/well, HepG2: 22,500 cells/well) and 1.5ml was added
832 per well. Once the top layer was set, 1ml of normal growth medium was added per well and 0.5ml of
833 growth medium was added to each well every 3-4 days. For TGFβ-treated HepG2 samples, the
834 growth medium was supplemented with 5ng/ml TGFβ. Colony growth was assessed by crystal violet
835 staining after 2-3 weeks (HA1ER/HA1E: 3 weeks). For this, wells were fixed and stained with 0.1%
836 crystal violet (Sigma-Aldrich, cat. C6158) and 10% ethanol for 30min, followed by multiple washes

837 over several days. Finally, plates were imaged on Optronics GelCount at 600dpi and processed for
838 colony counting using ImageJ (version 2.0.0).
839
840 **Transwell assay**
841 Serum-free growth medium was used to preincubate transwells (Corning, cat. 3422) at 37°C during
842 preparation of cells. Cells were trypsinised and resuspended in serum-free growth medium at a
843 concentration of $1x10^6$ cells/ml. For TGFβ-treated HepG2 samples, medium was supplemented with
844 5ng/ml TGFβ. 0.5ml of complete growth medium (+5ng/ml TGFβ for TGFβ-treated HepG2 samples)
845 was added to each well of a 24well plate and transwells were transferred to these wells, ensuring no
846 bubbles were present between the transwell membrane and the lower medium. Finally, 100 µl of the
847 cell suspension (100,000 cells) was added to the upper chamber of the transwell and incubated for
848 72 hours at 37°C. A cotton swab was used to remove non-migrated cells in the upper chamber,
849 before migrated cells were fixed and permeabilised in ice-cold methanol for 10 minutes. Cells were
850 stained in PBS supplemented with 10ug/ml DAPI (Sigma-Aldrich, cat. MBD0015) for 15min at room
851 temperature and imaged using a 4x objective (EVOS FL Cell Imaging System). Adobe Photoshop was
852 used to stitch together individual images covering the entire well and migrated cells were quantified
853 using ImageJ.
854
855 **TRAIL treatment and PARP labelling**
856 Cells were plated at a confluency of 1×10^6 cells/6010mm^2 surface. Twenty-four hours later, HA1ER-
857 derived clone F12 and subclones 1F8 and 1C9 were challenged with recombinant human TRAIL
858 (rhTRAIL, 1µg/ml, R&D Systems, cat. 375-LT-010) for 4 hours. TRAIL-induced apoptosis was
859 determined by assessing the percentage of cleaved PARP positive cells using flow cytometry (BD
860 LSRFortessa X-20 Cell Analyzer, BD Biosciences) and FlowJo software (BD, version 10.8.0) was used
861 for data analysis. Briefly, TRAIL-treated cells were collected and fixed with ice-cold methanol.
862 Methanol fixated cells were washed once with 1x PBS and incubated in blocking solution (0.5% BSA

863 PBS) for 1 hour at 20°C. Immunolabelling of cleaved PARP was performed using fluorescently
864 labelled Cleaved PARP Alexa 647 antibody (1:200 dilution, Cell Signalling, cat. 6987) for 1 hour at
865 20°C, followed by two washes in blocking solution (0.5% BSA PBS).
866
867 **TGFβ treatment**
868 Prior to TGFβ treatment, HepG2 cells were plated in a 6-well plate at 150T cells per well. The next
869 day, TGFβ was diluted in 10 µl medium per well and added to reach a final concentration of 5 ng/ml.
870 As vehicle control, the same volume of water was diluted in 10 µl medium per well and added. Cells
871 were incubated at 37 °C for 48h, before exchanging the growth medium supplemented again with 5
872 ng/ml TGFβ. After additional incubation for 48h at 37 °C, cells were collected for analysis or plated
873 without TGFβ for withdrawal experiments.
874
875 **Western blot**
876 Cells were lysed in RIPA buffer and whole cell extracts were quantified by Bradford assay to
877 homogenize protein concentration across samples prior to protein separation by 10% SDS PAGE
878 (including PageRuler Plus Prestained Protein Ladder, Thermo Fisher Scientific, cat. 26619) and
879 transfer to nitrocellulose membranes. Membranes were incubated in 5% milk in PBS-T (0.1% Tween
880 in PBS) for 1 h at 20°C and then washed 3 times in PBS-T for 5min. The membranes were incubated
881 with antibodies against N-cadherin (1:2000) or α-Tubulin (1:8000) in 5% milk at 4°C overnight.
882 Membranes were washed 3 times in PBS-T for 5 min and then incubated with the corresponding
883 secondary antibody (1:10,000) for 1 h. Finally, membranes were washed 3 times in PBS-T for 5min
884 and imaged on a LI-COR Odyssey imaging system.
885
886 **Generation of a lentiviral library containing 26 RAS variants with allocated molecular barcodes in**
887 **their H2B-GFP reporter**

888 As described previously[8], gBlock DNA fragment encoding the H2B-GFP fusion containing a NdeI site
889 in the 3'UTR and matching restriction sites at the ends for further downstream cloning into pLEX_307
890 (gift from David Root; Addgene, cat. 41392) (KpnI and SbfI) was amplified by PCR and cloned into the
891 pGEM-T Easy vector system. Following verification of the construct by Sanger sequencing, this H2B-
892 GFP fragment was further sub-cloned into the pLEX vector via One Shot ccdB Survival 2 T1 bacteria
893 strain (Invitrogen, cat. A10460). To introduce a molecular barcode into the 3'UTR of the H2B-GFP
894 gene, a gBlock DNA fragment with compatible restriction sites (NdeI and SbfI) encoding a modified
895 version of the plasmid sequence downstream of the H2B-GFP stop codon was amplified by PCR using
896 a primer with 12 random nucleotide barcodes split into three segments by intervening nucleotides
897 (NNNNGCGNNNNTGANNNN), and further cloned into the pGEM-T Easy vector. Pooled plasmid DNA
898 purification was performed (Qiagen Plasmid Midi Kit cat. 12143), the plasmid mix was digested using
899 NdeI and SbfI for sub-cloning the barcoded fragment into the previously generated pLEX-H2B-GFP
900 plasmid, and in-house generated electrocompetent One Shot ccdB Survival 2 T1 bacteria were
901 transformed. Approximately 50,000 bacteria colonies were pooled for plasmid DNA purification of
902 the barcoded pLEX-H2B-GFP plasmid mix. RAS variants were obtained from the Target Accelerator
903 Pan-Cancer Mutant Collection (Addgene, cat. 1000000103). Individual liquid cultures of bacteria
904 expressing pDONR plasmids encoding each of the 26 RAS variants included in our assay were pooled
905 according to their OD_{600} to achieve a balanced representation, and then pDONR-RAS plasmid mix was
906 purified using Qiagen Plasmid Midi Kit (cat. 12143). Following this, the pDONR-RAS plasmid mix was
907 used to shuttle the oncogene encoding sequences into the previously generated pLEX-H2B-GFP
908 plasmid mix by recombination using Gateway LR Clonase II Plus enzyme (Invitrogen, cat. 12538120)
909 following the manufacturer's instructions. A unique molecular barcode was assigned to each RAS
910 variant by (Sanger) sequencing the barcode region in the 3'UTR of H2B-GFP as well as the RAS variant
911 region. Finally, liquid cultures of the bacteria clones containing the 26 final lentiviral transfer vectors
912 (pLEX-H2B-GFP-BC-RAS: HRAS-A11D, HRAS-G12A, HRAS-G12C, HRAS-G12D, HRAS-Q61K, HRAS-Q61L,
913 HRAS-Q61R, HRAS-wt, KRAS-G12A, KRAS-G12C, KRAS-G12D, KRAS-G12V, KRAS-G12S, KRAS-D33E,

914 KRAS-A59G, KRAS-E62K, KRAS-wt, NRAS-G12A, NRAS-G12C, NRAS-Q61H, NRAS-Q61K, NRAS-Q61L,
915 NRAS-Q61R, NRAS-Y64D, NRAS-wt) were pooled based on their optical density and plasmid DNA was
916 purified, rendering the transfer vector mix for lentivirus production.
917
918 **Determination of the functional lentiviral titre and transduction of HA1E cells with 26 RAS variants**
919 Increasing volumes of lentivirus stocks (0-500 µl) were added to 150,000 HA1E cells and
920 supplemented with polybrene (Sigma, cat. H9268-5G) to a final concentration of 2 µg/ml. The
921 percentage of GFP positive cells was determined 48 hours post-transduction by flow cytometry (BD
922 LSRFortessa X-20 Cell Analyzer, BD Biosciences) and FlowJo software (BD, version 10.8.0) was used
923 for data analysis. The functional viral titre, calculated as the number of viral particles (i.e. transducing
924 units (TU)) capable to transduce a particular cell line per volume was calculated using the following
925 formula:

$$\text{functional titer}\left[\frac{\text{TU}}{\mu\text{l}}\right] = \frac{\text{cell number x frequency of GFP + cells [\%]}}{\text{volume of virus [}\mu\text{l]}}$$

926 Based on this, HA1E cells were transduced with the pooled lentiviral library containing 26 RAS
927 variants at a multiplicity of infection (MOI) of 0.05, supplemented with polybrene at a final
928 concentration of 2µg/ml. Forty-eight hours post-transduction, FACS was used to collect GFP-positive
929 cells (BD InFlux Cell sorter or FACSAria Fusion Cell Sorter, both BD Biosciences) with a purity of >90%
930 and sorted cells were re-plated in standard growing conditions.
931
932 **Single-cell data and BdLT-Seq analysis**
933 10x Chromium single-cell data was analysed using Seurat v4.0[76]. HTO-labelled cells were
934 demultiplexed using the HTODEMUX module of Seurat and further processed for quality control and
935 cell filtering taking into account nFeature_RNA, nCount_RNA and mitochondrial content. Further pre-
936 processing included the mathematical regression of cell cycle related gene expression levels to
937 remove cell cycle related bias in cluster processing. Similarly, variance in cluster detection due to high

938 levels of mitochondrial transcripts present per cell was also regressed prior to final clustering. HTO
939 and BdLT-Seq data was processed using CITE-seq-count[77] prior to demultiplexing or barcode
940 counting. BdLT-Seq barcode reads were pre-processed to remove sequencing artefacts, flanking and
941 intervening sequences using cutadapt v3.7[78]. Clean barcodes were then processed using CITE-seq-
942 count to generate lineage tracing files. A custom Node script[8] was used to determine lineage
943 relationships based on a similarity score considering the number of similar barcodes between two
944 cells and data obtained by these means was used to build a directional lineage tree. Lineage
945 visualisation was generated using CIRCOS[79]. All data analysed across the manuscript was pre-
946 processed/processed using R scripts. Final image processing and output was obtained using Adobe
947 Illustrator.
948
949 **Whole Exome Sequencing**
950 Founder clones and thereof derived subclones from each cellular system were evaluated to assess
951 possible genetic divergence at the exome level (WES, Whole Exome Sequencing. 100x coverage).
952 Briefly, cells were grown as described above and processed for DNA purification using the QIAmp
953 DNA mini kit (Qiagen. Cat 51304). High quality DNA Samples were then either processed for exome
954 capture using the SureSelect Human All Exon V6 (Agilent, cat 5190-8872) and then processed for
955 DNBSeq compatible library construction (BGI) or directly streamlined into DNBSeq library
956 construction (BGI). WES sequencing was performed following a 100PE sequencing protocol in a
957 DNBSeq instrument (BGI). WES was mapped to hg38 from each clonal lineage analysed, piled-up
958 using mpileup[80] (-q1 -Q30) and further processed through VarScan Somatic pipeline[81] (--min-
959 coverage 50). Chromosome level data visualisation was generated using chromo_plot from the vcfR
960 package[82].
961
962 **Whole Genome Sequencing**

963 Founder clones (HA1ER, clone F12 and HepG2 clone 1A) and thereof derived subclones from each
964 cellular system (HA1ER, clones 1F8 and 1C9; HepG2, clones 14F and 44D) were evaluated to assess
965 possible genetic divergence at the whole genome level (WGS, Whole Genome Sequencing. 30x
966 coverage). Briefly, cells were grown as described and processed for DNA purification using the QIAmp
967 DNA mini kit (Qiagen. Cat 51304). High quality DNA Samples were then directly streamlined into
968 DNBSeq library construction (BGI). WGS sequencing was performed following a 100PE sequencing
969 protocol in a DNBSeq instrument (BGI). WGS was mapped to hg38 from each clonal lineage analysed,
970 piled-up using mpileup[80] (-q1 -Q30) and further processed through VarScan Somatic pipeline[81] (--
971 min-coverage 50). Chromosome level data visualisation was generated using chromo_plot from the
972 vcfR package[82].
973
974 **SLAM-Seq and SLAM-DUNK**
975 SLAM-Seq was used to determine the fraction of nascent RNAs in contrast to steady-state RNA levels
976 for each clonal lineage (at least 3 biological triplicates). Briefly, the founder clone and its
977 corresponding subclones from each system were pulsed with 10µM 4-thiouridine (4-SU. Sigma, cat.
978 4509) for 15 min (1F8 vs 1C9) or 45 minutes (5 clones of HA1ER, HA1E and HepG2) and collected in
979 TRIzol (Invitrogen, cat. 15596-026). After RNA quality control assessment (Fragment Analyzer 5200,
980 Agilent, cat. M5310AA) and mass determination (Qubit, Invitrogen), 5 µg of RNA from each sample
981 was incubated in 4-SU conversion buffer (10 mM IAA in H_2O, 50 mM sodium phosphate pH 8,
982 0.15mM DTT, 50% DMSO (v/v) and H_2O to a final volume of 50 µl) for 15 min at 50°C on a heat block.
983 Immediately after, the conversion reaction was stopped by addition of 2 µl of 1M DTT and EtOH
984 precipitated over-night at -80°C (adding 200µl H_2O, 2 µl glycogen (20 mg/ml), 30 µl 3M pH 5.2sodium
985 acetate and 900 µ l ice-cold absolute EtOH). Retrieved RNA was further processed to build an
986 Illumina compatible strand-specific RNA sequencing library using either NEBNext Ultra II Directional
987 RNA Library Prep Kit for Illumina (NEB, cat. E7760S, for 1F8 vs 1C9) or QuantSeq 3′ mRNA-Seq V2
988 Library Prep Kit FWD with UDI (Lexogen, all clonal systems) kits following manufacturer's

989 instructions. Pooled libraries were sequenced on a NovaSeq 6000 (3' 150bp) aiming to sequence
990 30M reads per sample. SLAM-Seq sequencing data was processed using a T>C aware mapper
991 (NextGenMap[26]) and processed following SLAM-DUNK pipeline[27] (NextGenMap align to hg38
992 reference genome, mapping quality filter, snp annotation filtering and modified reads count). Final
993 count tables were further processed using the beta binomial test to determine transcriptional
994 differences from steady-state RNA levels (FDR<0.05, FC>2). Data visualisation was generated using
995 heatmap.2 (R v4.2.0).
996
997 **Chromatin accessibility (ATAC-Seq)**
998 ATAC-seq was performed for all clones evaluated herein (HA1E: 10E, A7A, A9H, B2F and B5F; HA1ER:
999 F12, 1F8, 1C9, 2A6 and 1D12; HepG2: 1A, 32G, 14F, 48E and 44D) following the Omni-ATAC protocol[83]
1000 but using an in-house prepared Tn5 transposase[84] . Briefly, 75,000 cells were processed for ATAC-Seq
1001 as follows: HA1E and HA1ER clones were seeded at 3.5×10^5 cells per 6-cm dish whilst HepG2 clones
1002 at 5.25×10^5 cells and grown for 2 days. Next, cells from every system were rinsed twice with 4ml
1003 PBS and then trypsinised and collected by centrifugation (250 g, 4 min), resuspended in PBS, placed
1004 on ice and counted in a Neubauer chamber (Marienfeld, cat. 0640030). Aliquots containing 75,000
1005 cells were transferred to pre-chilled DNA low-binding tubes, centrifuged (350 g, 5 min, 4 °C; swing-
1006 bucket rotor with FACS-tube adaptor, Eppendorf 5810 R) and resuspended in 75 µl ice-cold ATAC Lysis
1007 Buffer (9.7 mM Tris-HCl pH 7.5, 9.7 mM NaCl, 2.91 mM $MgCl_2$, 0.1% (wt/vol) IGEPAL CA-630, 0.1%
1008 (wt/vol) Tween-20, 0.01 % (wt/vol) digitonin). After 3 min on ice, lysates were diluted with 1 ml ice-
1009 cold ATAC Wash Buffer (9.9 mM Tris-HCl, 9.9 mM NaCl, 2.97 mM $MgCl_2$, 0.1% Tween-20), mixed by
1010 inversion and nuclei were collected by centrifugation (500 g, 10 min, 4 °C). Next, cell pellets were
1011 gently resuspended in 75 µl ice-cold transposition buffer (10 mM Tris-HCl pH 7.6, 5 mM $MgCl_2$, 10 %
1012 (vol/vol) dimethylformamide, 33 % PBS, 0.1% Tween-20, 0.01% digitonin, 1.25 µM loaded Tn5) and
1013 incubated at 37°C in a Thermomixer comfort at 1000 rpm for 30 min. Transposition reactions were
1014 stopped by the addition of 375 µl of DNA Binding Buffer (DNA Clean & Concentrator-5 kit, Zymo

1015 Research D4013) and further processed for transposed DNA enrichment using DNA Clean &
1016 Concentrator-5 kit (Zymo Research D4013). Finally, transposed DNA was eluted in 21 µl Elution Buffer
1017 into DNA low-binding tubes. Next, for multiplexing purposes, barcoding PCR reactions were
1018 conducted for each sample. Each PCR reaction contained 25 µl NEBNext Ultra II Q5 2× Master Mix
1019 (M0544S, New England Biolabs), 2.5 µl 5 µM i5 index primer, 2.5 µl 5 µM i7 index primer, and 20 µl
1020 tagmented DNA and were amplified as follows: 72°C 5 min; 98 °C 30 s; 13 cycles (HA1ER, HepG2) or
1021 14 cycles (HA1E) of 98°C 10 s, 65°C 1 min 15 s; final extension 65°C 3 min, hold 4°C.
1022
1023 Barcoded PCR products (sequencing-ready library) were subjected to size selection using Mag-Bind
1024 TotalPure NGS beads (range 150 – 1000 nt. M1378-00, Omega Bio-tek, Inc.) and then assessed on an
1025 Agilent 4200 TapeStation using D1000 ScreenTape and D1000 Reagents (35–1,000 bp detection
1026 range). Validated libraries were then sequenced in a NextSeq2000 apparatus generating a 50PE reads
1027 for each sample. Obtained reads were trimmed (Tn5 adaptor sequence: CTGTCTCTTATACACATCT)
1028 using cutadapt[78] prior to mapping to hg38 using Bowtie2[85]. After mapping peak calling was
1029 performed (FDR<0.05) using MACS2 v2.2.7.1[86] pipeline. Finally, a catalogue containing all peaks
1030 identified on every sample from each cellular system analysed was built to extract genomic
1031 coordinates used for heatmap visualisation (Deeptools v3.5.1[87]).
1032
1033 **Label-free proteomic analysis (LC-MS-MS)**
1034 Cell lysis was performed in 5% SDS, 50 mM TEAB followed by protein content measurement using the
1035 2D quant kit (Cytiva, cat. 80-6483-56). Then, 200 µg of protein from each sample was reduced with
1036 5mm TCEP at 55°C for 15 min followed by alkylation with 15mM MMTS for 10 min at 20°C. Samples
1037 were then loaded onto STRAP minis (Protifi, cat. C02-mini) and digested with 20 µg sequencing grade
1038 trypsin (Sigma, cat. TRYPSEQ-RO) for 18 h at 37°C. Following STRAP elution, peptide digests were
1039 dried to completeness in a vacuum centrifuge at 45°C before resuspension at 800ng/µl on 0.2%
1040 formic acid.

1041

1042 LC-MS was performed in technical quintuplicate in a RSLC HPLC system (mobile phase A: water 0.1%
1043 formic acid; mobile phase B: acetonitrile) coupled to an Orbitrap Lumos mass-spectrometer via an
1044 EasySpray source (Thermo Fisher Scientific). Peptides were injected directly onto an EasySpray
1045 column (50 cm long, 7 µm ID pepmap C18) at 180 nl/min in 1% mobile phase B and were separated
1046 at 180 nl/min with a gradient of 1-24% over 90 min. LC eluent was sprayed directly into the MS
1047 source at an ion spray voltage of 1.9Kv. The MS was operated in data dependent mode where an MS1
1048 survey scan was acquired in the Orbitrap detector at 120K resolution (at m/z 200) in the mass range
1049 of 350-1200m/z at a target value of 2e5 ions and a maximum fill time of 40ms. All available charged
1050 ions were selected for MS2 by HCD at a normalised collision energy of 35% with product ions
1051 detected in the Orbitrap detector at a resolution of 30K, 54ms max fill time and a target value of 1e5
1052 ions. A maximum cycle time of 1.2s was allowed.

1053

1054 **TMT-proteomics**

1055 Cell lysis was performed in 5% SDS, 50mM TEAB followed by protein content measurement using the
1056 2D quant kit (Cytiva, cat. 80-6483-56). Then, 200 µg of protein from each sample was reduced with 5
1057 mm TCEP at 55 °C for 15 min followed by alkylation with 15 mM MMTS for 10 min at 20°C. Samples
1058 were then loaded onto STRAP minis (Protifi, cat. C02-mini) and digested with 20 µg sequencing grade
1059 trypsin (Sigma, cat. TRYPSEQ-RO) for 18 h at 37°C. Following STRAP elution, peptide digests were
1060 dried to completeness in a vacuum centrifuge at 45°C. Peptide digests were resuspended in 500 mM
1061 TEAB at a concentration of 2 µg /µl and TMT 10-plex labelled according to the manufacturer's
1062 instructions (Thermo Fisher Scientific). Reactions were quenched by addition of 8 µl 5%
1063 hydroxylamine prior to pooling. Samples were acidified to pH 2.7 by addition of trifluoracetic acid
1064 before revers phase solid phase extraction.

1065 Pooled, acidified peptides were loaded onto C18 RP cartridges (Sep-Pak 500mg tC18, Waters, cat.
1066 186004619), washed with 10 ml 0.1 % formic acid before elution in 500 µl 60% acetonitrile, 0.1%

1067 formic acid. 10% of the pooled sample was aspirated off, the remaining 90% of peptide was split
1068 equally into 2 tubes and all aliquots were dried to completeness in a vacuum centrifuge at 45°C.
1069 Approximately a 10 % aliquot (200 µg) was reserved for proteomics profiling and both 45% aliquot
1070 (900 µg) were resuspended for subsequent enrichment either by TiO2 (Thermo Fisher cat. A32993)
1071 or FeNTA enrichment (Thermo Fisher cat. A32992). Following elution from either TiO2 or FeNTA
1072 columns, samples were dried to completeness in a vacuum centrifuge at 20°C followed by
1073 resuspension in 5 µl of 0.05% TFA and were ready for LC-MS-MS analysis.
1074 LC-MS was performed in technical quintuplicate in a RSLC HPLC system (mobile phase A: water 0.1%
1075 formic acid; mobile phase B: acetonitrile) coupled to an Orbitrap Lumos mass-spectrometer via an
1076 EasySpray source (Thermo Fisher Scientific). Following, peptides/phospho-peptides were injected
1077 directly onto an EasySpray column (50 cm long, 75 µm ID pepmap C18) at 180 nl/min in 1% mobile
1078 phase B. Peptides were separated at 180 nl/min with a gradient of 1-24% over 90 min. LC eluent was
1079 sprayed directly into the MS source at an ion spray voltage of 1.9 Kv. The MS was operated in data
1080 dependent mode where an MS1 survey scan was acquired in the Orbitrap detector at 120 K
1081 resolution (at m/z 200) in the mass ranges of 350-1200 m/z at a target value of 2e5 ions and a
1082 maximum fill time of 40 ms. All available charged ions were selected for selected for MS2 by HCD at a
1083 normalised collision energy of 42%. Product ions were detected in the Orbitrap detector at a
1084 resolution of 50K, 86 ms max fill time and a target value of 1e5 ions. A maximum cycle time of 4s was
1085 allowed.
1086
1087 **Label-free proteomics data processing**
1088 Raw data was imported into Progenesis QI proteomics pipeline (Waters Nonlinear Dynamics) where
1089 peak detection was performed at maximum sensitivity. Protein relative quantitation was performed
1090 using the top 3 method. Proteins were identified by extraction of a mascot generic file (MGF) that
1091 was exported to Mascot (Matrix science) with the following search criterion. Species human,
1092 specificity was trypsin allowing up to 2 missed cleavages, precursor mass accuracy 10ppm, product

1093 ion sensitivity 20 ppm, fixed modification Methylthio (C) variable modifications oxidation (M)
1094 deamidation (NQ) searched against Uniprot_Swissprot_2022 using a decoy FDR method limited at
1095 1%. Protein identities were imported back into Progenesis QI and data was further processed in R for
1096 visualisation purposes.
1097
1098 **TMT proteomic analysis (PEAKS)**
1099 Raw data was imported into Peaks Studio 10.6 (Bioinformatics Solutions Inc.) where both protein
1100 identification and quantitation were performed using the default TMT 10 plex template. Database
1101 search evaluated human acquired data against the Uniprot_Sprot_2022.fasta database using fixed
1102 modification of methylthio (C) variable modifications oxidation (M) deamidation (NQ) and the
1103 following advance settings: merge scans [DDA], correct precursor [DDA], mass only, associate
1104 features with chimera scan [DDA], filter features charge between 2 and 4 10 ppm precursor mass
1105 tolerance, fragment ion 0.1 Da, digest mode specific trypsin, max missed cleavages per peptide 2,
1106 estimated FDR by decoy-fusion, quant TMT 10 plex, MS2 quant tolerance 20 ppm, FDR threshold 1%.
1107
1108 **TMT phospho-proteomic analysis (PEAKS)**
1109 Raw data was imported into Peaks Studio 10.6 (Bioinformatics Solutions Inc.) where both protein
1110 identification and quantitation were performed using the default TMT 10 plex template. Database
1111 search evaluated human acquired data against the Uniprot_Sprot_2022.fasta database using fixed
1112 modification of methylthio (C), variable modifications oxidation (M), deamidation (NQ),
1113 phosphorylation (STY) and the following advance settings: merge scans [DDA], correct precursor
1114 [DDA], mass only, associate features with chimera scan [DDA], filter features charge between 2 and 4
1115 10ppm precursor mass tolerance, fragment ion 0.1Da, digest mode specific trypsin, max missed
1116 cleavages per peptide 2, estimated FDR by decoy-fusion, quant TMT 10 plex, MS2 quant tolerance
1117 20ppm, FDR threshold 1%. Differentially represented phosphorylation sites between subclones were
1118 identified by applying the beta binomial distribution (FDR<0.05, FC>0.4) using countData package (R).

1119

1120 **CyTOF**

1121 HA1ER cells (clones F12, 1C9 and 1F8; $3x10^6$ each) were pulsed for 30 min with IdU following
1122 manufacturer's instructions (Maxpar cell cycle panel kit, cat. 201313) and then collected by
1123 trypsinisation. Cells were next spun-down and resuspended in 1ml 1X Maxpar Fix I Buffer (Fluidigm,
1124 cat. 201065) for 10 min at 20°C. Then cells were spun down and resuspended in 7 ml Maxpar Cell
1125 Staining Buffer (Fluidigm, cat. 201068), mixed gently and further incubated for 10 min at 20°C. Next,
1126 cells were spun down once again and after repeating this washing step, cells were resuspended in 50
1127 µl of Maxpar Cell Staining buffer and then permeabilised with ice-cold methanol for 15 min on ice.
1128 Cells were then washed twice with 2 ml Maxpar Cell Staining Buffer +EDTA (Maxpar Cell Staining
1129 Buffer supplemented with 5mM EDTA) and spun down for 5 min at 800g. Immediately after, 1 µl of
1130 each metal isotope-labelled specific antibody (MOT4, JIP4, SYUA, DCXR, SQSTM, MYEF2, Hsp20, pRb,
1131 CyclinB1, Ki67, pH3 (Supplementary Table), using Maxpar X8 Antibody Labeling Kit (Fluidigm, cat.
1132 201144A, 201167A, 201155A, 201171A, 201176A, 201147A, 201159A)) was added to Maxpar Cell
1133 Staining Buffer +EDTA and a final volume of 50 µl was added per sample and incubated for 30 min at
1134 20°C. Following incubation, cells were washed twice with 2 ml Maxpar Cell Staining Buffer +EDTA and
1135 further spun down at 800 g for 4 min. Immediately after, supernatant was removed and 1ml of 4%
1136 PFA and 1 µl Iridium were added to each sample and stored at 4°C for 16 h. Next day, samples were
1137 washed in 2 ml of Maxpar Cell Staining Buffer +EDTA and spun down. Finally, samples were washed
1138 twice in Maxpar Cell Acquisition Plus Solution (Fluidigm, cat. 201244). Samples were analysed in a
1139 Helios apparatus and data was processed using flowCore v2.16.0[88], CATALYST v1.28.0[89] and visualised
1140 in R v4.2.0.

1141

1142 **Gapmer-mediated knock-down**

1143 Antisense LNA gapmers were designed against lncRNA transcript sequences (Ensembl release 105[90])
1144 using the Qiagen Antisense LNA GapmeR custom builder (Supplementary Table). One day prior to

1145 transfection, 100T cells (HA1ER) or 150T cells (HepG2) were seeded into a 6well plate in 2ml growth
1146 medium. Next day, around 450 µl of medium were removed from each well for a final volume of
1147 1.5ml per well. 40 µM gapmer was mixed with 75 µl OptiMEM (Invitrogen, cat. 31985062) per well to
1148 be transfected. The X-tremeGENE 360 transfection reagent (Roche, cat. 8724105001) was
1149 equilibrated to 20°C and briefly vortexed before immediately adding 2.5 µl per well to be transfected
1150 to the gapmer mix. After thorough mixing and incubation at for 30min at 20°C, 75 µl of the
1151 transfection mix was added dropwise to each well of cells (final gapmer concentration in 1.5ml: 26
1152 nM). Following gentle mixing, cells were incubated at 37°C. The growth medium was exchanged 24h
1153 post-transfection. Cells were used in RT-qPCR, scRNA-seq or RPPA assay 48h post-transfection, only
1154 for DPP-DT-G2 cells were used 24h post-transfection.
1155
1156 **Nuclear-Cytoplasmic fractionation for HA1ER clones**
1157 HA1ER clones F12, 1F8 and 1C9 were seeded at 1×10^6 onto a 10 cm culture dish and were grown to
1158 ~70% confluency. Cells were rinsed twice with 10 ml ice-cold PBS and trypsinised with 1 ml of 0.25%
1159 trypsin. Cells were pelleted by centrifugation at 220g for 3 minutes at 4°C and then washed twice in 1
1160 ml ice-cold PBS, counted, and 2×10^6 cells were then taken for further fractionation. Pellets were
1161 resuspended in 100 µl of hypotonic buffer (20 mM Tris pH 7.6, 0.1 mM EDTA, 2 mM $MgCl_2$, 0.5 mM
1162 NaF, and 0.5 mM Na_3VO_4) supplemented with 1x cOmplete EDTA-free Protease Inhibitor Cocktail
1163 (Roche, cat. 11873580001), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 260 U/ml RiboLock
1164 RNase Inhibitor (Thermo Fisher Scientific, cat. EO0381), and transferred to a fresh 2ml DNase/RNase-
1165 free DNA-low binding tube. Cells were then incubated for 2 minutes at 20°C, then for 10 minutes on
1166 ice to induce hypotonic swelling, followed by addition of 10 µl of 10% IGEPAL CA-630 (Sigma-Aldrich,
1167 cat. I8896) and pipetted up-and-down approximately 20 times. Nuclei, unbroken, or partially lysed
1168 cells were pelleted at 550g for 3 minutes at 4°C and then 80% of the supernatant was taken from the
1169 top (containing the cytoplasmic fraction), transferred to a fresh tube. Next, 150 µl of hypotonic
1170 buffer was added to the pellet and remaining cytoplasmic extract; the pellet was resuspended and

1171 once again spun down. 80% of the supernatant was taken from the top and combined with the
1172 cytoplasmic fraction. Nuclei pellets were then resuspended in hypotonic buffer in a volume equal to
1173 the final cytoplasmic fraction (373 µl). 20 ng of spike-in eGFP transcript for HA1ER cell system (*in-*
1174 *vitro* transcribed from an eGFP encoding plasmid. HighYield T7 RNA Synthesis Kit, Jena Bioscience,
1175 cat. NT-101) was added to each fraction, and total RNA was extracted using 1 ml of TRIzol
1176 (Invitrogen, cat. 15596-026). 773 µl of the aqueous phase from each fraction was mixed with 25.5 µl
1177 of 3M NaOAc, 1.5 µl of 5 mg/ml linear acrylamide (Invitrogen, cat. AM9520), 1 volume of
1178 isopropanol, and RNA was precipitated overnight at -20°C. Pellets were washed with 75% ice-cold
1179 ethanol, dried, and re-dissolved in 12 µl of nuclease-free water. RNA concentrations were measured
1180 using Nanodrop.
1181
1182 **Nuclear-Cytoplasmic fractionation for HepG2 clones**
1183 HepG2 clones 1A, 14F and 44D were seeded at 1.5×10^6 onto a 10 cm culture dish and were grown to
1184 ~70% confluency. Cells were harvested and resuspended in ice-cold PBS and from thereon kept on
1185 ice. For each sample, 2×10^6 cells were transferred to a fresh tube, pelleted and resuspended in 380 µl
1186 ice-cold hypotonic lysis buffer (HLB, 10mM Tris-HCl pH7.5, 10mM NaCl, 3mM $MgCl_2$, 10% glycerol,
1187 0.3% IGEPAL CA-630) with 100U Protector RNase inhibitor (Roche, cat. 3335399001) and incubated
1188 on ice for 10min. Following brief vortexing and pelleting by centrifugation, the supernatant was
1189 transferred to a fresh tube (cytoplasmic fraction). The pelleted nuclei were washed twice in 380 µl
1190 HLB and finally resuspended in 380 µl (nuclear fraction). 1 ng of a custom spike-in plasmid was added
1191 to each sample to allow for normalisation during RT-qPCR analysis. After adding 1ml TRIzol to each
1192 fraction, RNA was extracted according to the manufacturer's instructions. To ensure quantitative
1193 comparisons between samples in RT-qPCR, care was taken to treat samples equally during RNA
1194 extraction, e.g. collecting equal volume of aqueous phase for each sample.

1195 **Reverse transcription followed by qPCR**

1196 Total RNA was isolated using TRIzol (Invitrogen, cat. 15596-026) following manufacturer's
1197 instructions. Reverse transcription was performed using QuantiTect Reverse Transcription Kit
1198 (Qiagen, cat. 205311) following the manufacturer's instructions with 1 or 2µg RNA input. qPCR was
1199 performed using PowerUp SYBR Green Master Mix (Invitrogen, cat. A25742) in a QuantStudio3 Real
1200 Time PCR System (Thermo Fisher Scientific, cat. A28136). Data analysis was conducted using Real-
1201 Time qPCR Connect Data Analysis Tool (Thermo Fisher Scientific Connect Platform).

1202

1203 **FISH probes design and production**

1204 FISH probe amplicon regions between 260-300bp were determined based on transcript sequences
1205 (Ensembl release 105[90]) and primers to amplify these regions were then designed using the Connect
1206 OligoPerfect tool (Thermo Fisher). RNA purified from cells known to express target RNAs at high level
1207 was reverse transcribed with QuantiTect reverse transcription kit (Qiagen, cat. 205311) using oligo-dT
1208 primers following the manufacturer's recommendations. cDNA was then amplified by the Expand
1209 high fidelity PCR system (Roche, cat. 11732641001) using specific primers and PCR-products were
1210 then resolved by agarose gel electrophoresis. Bands of the expected molecular weights were purified
1211 and quantified using the Qubit dsDNA BR assay kit (Invitrogen, ca. Q32850). 1µg amplicon DNA was
1212 then digested using Lambda exonuclease (NEB, cat. M0262) at 37°C for 1 hour, followed by heat
1213 inactivation at 75°C for 10 minutes. Digestion products were then isolated on streptavidin beads and
1214 then washed twice with 1X B+W buffer (5mM Tris-HCl pH7.5, 1mM EDTA, 2M NaCl). Next, beads
1215 were resuspended in 25µl amplification mix 1 (1µl dNTP mix (10mM dATP, 10mM dCTP, 10mM dGTP,
1216 7.5mM dTTP and 2.5mM aminoallyl dUTP (Thermo Scientific, cat. R1101)), 2µl 20µM random
1217 hexamers, 22µl H_2O) before the addition of 25µl amplification mix 2 (0.75µl Expand high fidelity
1218 enzyme mix (Roche, cat. 11732641001), 5µl 10X Expand high fidelity buffer with 15mM $MgCl_2$
1219 (Roche, cat. 11732641001) and 19.25µl H_2O). Beads were then washed twice with H_2O and
1220 resuspended in 25µl H_2O. 15µl labelling buffer (25 mg/ml sodium bicarbonate (Sigma-Aldrich, cat.
1221 S5761) was added to the resuspended beads before addition of 10µl dye (Alexa Fluor 555/647

1222 Reactive Dye (Invitrogen, cat. A32757, A32756) dissolved in 40 µl DMSO). Samples were vortexed
1223 briefly and incubated at room temperature for 1 hour blocked from light. Beads were then washed
1224 three times in H_2O and FISH probes were released through chemical denaturation by resuspending
1225 beads in 100µl filtered denaturation buffer (100mM NaOH, 0.1mM EDTA pH8.0) and incubating at
1226 room temperature for 15 min blocked from light. The supernatant was separated from the beads
1227 using a magnetic separator and collected into a fresh tube. Probes were then precipitated from the
1228 supernatant using 1/10 vol sodium acetate, 2µl Ethachinmate (Nippon Gene, cat. 318-01793) and
1229 2.5X vol 100% ethanol and then eluted in 20µl H_2O. Probes stocks were stored at -80°C.
1230
1231 **RNA-FISH**
1232 Cells were grown to 70% confluency on glass coverslips in 24-well plates and treated according to
1233 experimental design. Cells were then washed twice with 1X ice-cold PBS and fixed with 3%
1234 formaldehyde (Thermo Scientific, cat. 28906) and 4% sucrose in 1X PBS for 15 minutes at 20°C.
1235 Following fixation, cells were washed three times in 1X PBS and permeabilised with 0.25% Triton-X in
1236 1X PBS for 10 minutes at 20°C, followed by three washes in 1X PBS. Fluorescently labelled RNA-FISH
1237 probes were diluted in formamide (10µl total volume per coverslip) and denatured at 75°C for 7
1238 minutes. 10µl 2X hybridisation buffer (4X SSC, 40% dextran sulphate (Sigma-Aldrich, cat. D8906),
1239 2mg/ml BSA) per coverslip was then added to each probe dilution. 20µl probe mix was hybridised to
1240 each sample of cells in a Thermobrite 110/120 VAC hybridisation chamber (Abbott, cat. 07J91-010),
1241 using the incubation program as follows: 65°C for 5 minutes, 37°C for 16 hours. Following
1242 hybridisation, samples were washed three times in 50% formamide/2X SSC at 42°C for 5 minutes
1243 each, three times in 2X SSC at 42°C for 5 minutes each and three times in 1X SSC at 42°C for 5
1244 minutes each. Samples were then counterstained with DAPI and washed twice in 1X SSC, before
1245 mounting in ProLong Gold Antifade Mountant (Molecular Probes, cat. P36930). Cells were analysed
1246 on a Zeiss LSM880 microscope equipped with Airyscan module.
1247

1248 **RNA-FISH coupled to ICC**

1249 Cells were grown to 70% confluency on glass coverslips in 24-well plates and treated according to
1250 experimental design. Cells were then washed once with 1X PBS and fixed with 3% formaldehyde
1251 (Thermo Scientific, cat. 28906) and 4% sucrose in 1X PBS for 15 minutes at room temperature.
1252 Following fixation, cells were washed three times in 1X PBS and permeabilised with 0.25% Triton-X in
1253 1X PBS for 10 minutes at room temperature, followed by three washes in 1X PBS. Cells were blocked
1254 in blocking buffer (1% BSA in 1X PBS for at least 2 hours at 20°C, before incubation with 1:200
1255 primary antibody in blocking buffer overnight at 4°C. Cells were then washed twice with 0.1% Tween-
1256 20 in 1X PBS and incubated with 1:1000 secondary antibody in blocking buffer for 1 hour at 20°C,
1257 blocked from light. Cells were then washed twice with blocking buffer and then crosslinked again
1258 with 3% formaldehyde and 4% sucrose in 1X PBS for 10 minutes at 20°C. Next, cells were washed
1259 twice with 2X SSC (Sigma-Aldrich, cat. S6639) / 50% formamide (Sigma-Aldrich, cat. F9037). Next,
1260 fluorescently labelled RNA-FISH probes were diluted in formamide (10µl total volume per coverslip)
1261 and denatured at 75°C for 7 minutes. 10µl 2X hybridisation buffer (4X SSC, 40% dextran sulphate
1262 (Sigma-Aldrich, cat. D8906), 2mg/ml BSA) per coverslip was then added to each probe dilution. 20µl
1263 probe mix was hybridised to each sample of cells in a Thermobrite 110/120 VAC hybridisation
1264 chamber (Abbott, cat. 07J91-010), using the incubation program as follows: 65°C for 5 minutes, 37°C
1265 for 16 hours. Following hybridisation, samples were washed three times in 50% formamide/2X SSC at
1266 42°C for 5 minutes each, three times in 2X SSC at 42°C for 5 minutes each and three times in 1X SSC
1267 at 42°C for 5 minutes each. Samples were then counterstained with DAPI and washed twice in 1X
1268 SSC, before mounting in ProLong Gold Antifade (Invitrogen, cat. P36930). Cells were visualised with a
1269 Zeiss LSM880 microscope equipped with Airyscan module. Images were obtained using ZEN software
1270 (Zeiss). ImageJ version 2.0.0 was used to split channel information. Adobe Photoshop 2025 was used
1271 to prepare pseudocoloured panels.

1272

1273 **Reverse Phase Protein Array (RPPA)**

1274 Cells were transfected and cultured as described in the section 'Gapmer-mediated knock-down'. To
1275 prepare samples for RPPA, cells were kept on ice and washed twice with ice-cold PBS. Next, lysis
1276 buffer (1 % Triton X-100 (v/v), 0.05 M HEPES pH7.4, 1 mM EGTA ph7.5-8, 0.15 M NaCl, 1.5 mM $MgCl_2$,
1277 1 mM Na_3VO_4, 10 mM TSPP, 112.5 mM NaF, 10 % glycerol, 1 tablet/10ml cOmplete Mini EDTA-free
1278 protease inhibitor (Roche, cat. 11836170001), 1 tablet/10ml PhosStop phosphatase inhibitor (Roche,
1279 cat. 04906837001) was added and cells were scraped and collected into a tube. After a 20min
1280 incubation on ice, samples were centrifuged at 16,900 rcf for 10min at 4°C and the supernatant
1281 protein lysate was collected and stored at -20°C. Protein concentration was determined by Bradford
1282 assay. Samples were prepared for printing by addition of 4X Sample buffer (40% Glycerol, 8% SDS,
1283 0.25 M Tris-HCl, 10% β-mercaptoethanol) and heat denaturation at 95°C for 5 min. All samples were
1284 diluted to 1.5 mg/ml (D1), 0.75 mg/ml (D2), 0.375 mg/ml (D3) and 0.1875 mg/ml (D4) using PBS
1285 containing 10% glycerol. Array spotting was carried out with the Quanterix 2470 Arrayer platform
1286 using 185 µM pins. Each sample was spotted at four dilutions (D1-D4) onto single pad ONCYTE
1287 SuperNOVA nitrocellulose slides (Grace Bio-Labs, cat. 705278) at a 500 µM spot-to-spot distance and
1288 in 3 technical replicates. 180 signalling pathway markers (Supplementary Table) were profiled in a
1289 standard RPPA assay as described briefly below[43]. After spotting, slides were incubated with antigen
1290 retrieval solution (1x Reblot strong. Merck Millipore, cat. 2504) for 10 min before being placed in a
1291 microfluidic structure to individually address the arrays with primary or secondary antibody
1292 solutions. Following blocking buffer (Superblock T20. Thermo Scientific, cat. 37536) for 10 min the
1293 detection of marker antibodies was performed in a two-step sequential assay. First the array is
1294 incubated with primary analyte-specific antibody in blocking buffer for 60 min at room temperature
1295 followed by the removal of excess antibody by washing arrays with 0.1% PBS-Tween followed by a
1296 further incubation with blocking buffer and PBS-T washes, and incubation with secondary antibody
1297 (Dylight-800-labeled anti-species antibodies diluted 1:2500 in Superblock T20) for 30 min. After
1298 further washing and slide drying, the arrays were imaged in the Innopsys Innoscan 710 scanner
1299 (Innopsys). Blank signals were determined by omitting the primary antibody and instead incubating

1300 the array with Superblock T20 alone, followed by secondary antibody incubation. Sample loading on
1301 arrays (for normalisation) was determined by staining one slide with fast green protein dye and
1302 scanning at 800 nm. Microarray images were analysed using Mapix software (Innopsys) with feature
1303 (spot) diameter of the grid set to 270 µm. Average signal intensities was determined for each
1304 individual feature and the median background from the adjacent area was subtracted from each
1305 feature signal leading to a net signal per feature. Fluorescence intensity for each feature (spot) on
1306 the array was measured and a test was performed for linear fit through the 4-point dilution series for
1307 all samples on all arrays using a flag system where R2 > 0.9 (green flag) is deemed good, >0.75
1308 (amber flag) is deemed acceptable and <0.75 (red flag) is poor and was excluded from data analysis.
1309 The median values from the 4-point dilution series were calculated and used as a measure of
1310 fluorescence intensity. This median fluorescence intensity for each sample and target was
1311 normalised to the median fluorescence intensity of the fast green protein stain (total protein) and
1312 thereafter averaged across the technical triplicates. The final (phospho-) protein levels were
1313 averaged across both biological replicates and a threshold of ≥20% increase/decrease compared to
1314 the mean of the non-targeting control gapmer was applied to determine differential levels.
1315
1316 **Statistics and reproducibility**
1317 All histograms displayed in the manuscript show mean value ± standard deviation (SD) or
1318 propagated error with the number of replicates and, if performed, the method to determine
1319 statistical significance indicated in the figure legends.
1320 As part of scRNA-Seq data pre-processing, cells found to be displaying multiple HTO (hashtag oligo)
1321 labelling after demultiplexing were excluded from further analysis. Moreover, to attenuate
1322 artefactual clustering, cells displaying high levels of mitochondrial transcripts and abnormal read
1323 counts were masked. Further processing included the mathematical regression of the datasets based
1324 on cell cycle-related genes expression to remove cell cycle-related bias in cluster mapping. Random

sampling was applied to BdLT-Seq data libraries to optimise the number of barcodes needed to build lineage trees and to reduce computational burden.

All attempts of replication were successful. No statistical method was used to predetermine the sample size. For the analysis of the label-free MS in Fig. 3a, one sample from 1F8 (1_A) and one sample from 1C9 (1_A) were excluded from the analysis due to poor performance in the MS, presumably linked to a technical issue related to stabilisation/optimisation of the instrument. Given the fragile nature of the soft agar in the clonogenicity assays, some damaged wells were excluded from the analysis. The experiments were not randomised and the investigators were not blinded during experiments and outcome assessment.

Data availability: All data used to generate figures in this study is provided as Source Data. Sequencing data have been deposited in the Gene Expression Omnibus (GEO) database under the accession code GSE223496. WES and WGS have been deposited the Short Read Archive (SRA) and can be retrieved under the accession code PRJNA1219504 and PRJNA1279408. The mass spectrometry proteomics and phospho-proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[91] partner repository with the dataset identifiers PDX063769, PXD064047, 10.6019/PDX063769 and 10.6019/PXD064047.

Code availability: All data were analysed using publicly available, benchmarked, and validated pipelines (R v4.2.0, Seurat v4.0, CITE-seq-count v1.4.5, cutadapt v3.7, CIRCOS, Bowtie2 v.2.5.4, MACS2 v2.2.7.1, Deeptools v3.5.1, samtools v1.15.1, bcftools, VarScan v2.4.4, vcfR v1.15.0, SLAM-DUNK v0.4.3, NextGenMap v0.5.5, FlowCore v2.16.0, CATALYST v1.28.0). Data analysis rationale is described in the Methods section. For any data analysis-related inquiries, please contact M.M.P. (Maximiliano.Portal@glasgow.ac.uk).

Acknowledgements

1351 We thank Valeria Pavet for helpful discussions and critical reading of the manuscript. We are also
1352 grateful to Richard Marais for his support. We also thank Naveed Khan for IT assistance. Y.S. was a
1353 Cancer Research UK—Manchester Institute postdoctoral fellow, B.B. and K.L.M. are recipients of a
1354 Cancer Research UK—Manchester Institute PhD studentship, A.M. is a Cancer Research UK-Scotland
1355 Institute postdoctoral fellow and M.M.P. was a Cancer Research UK Fellow and is currently a CRUK
1356 Scotland Institute Fellow and Senior Lecturer of the University of Glasgow. We acknowledge the
1357 HTPU microarray service at the Institute of Genetics and Cancer at the University of Edinburgh for
1358 their technical support with RPPA. We thank the Core Facilities at the CRUK Manchester Institute
1359 (Molecular Biology Core, Flow Cytometry, Biological Mass Spectrometry and Visualisation, Irradiation
1360 & Analysis) and Finance and Logistics Services at Cancer Research UK Manchester Institute and
1361 Scotland Institute. This research was funded by CRUK-MI core grant number C5759/A27412 and
1362 CRUK-SI core grant number A31287.

1363

1364 **Contributions**

1365 M.M.P. conceived the research project. Y.S., B.B., K.L.M. and M.M.P. designed and performed
1366 experiments. A.M. performed nuclear/cytoplasmic fractionation assay on HA1ER cells and ATAC-Seq
1367 on all clonal systems. R.B.J. prepared recombinant Tn5 for ATAC-Seq experiments. M.M.P. performed
1368 computational analysis. N.O.C. set up the RPPA platform and A.F.M. performed RPPA assays. Y.S.,
1369 B.B., K.L.M. and M.M.P. wrote the Material & Methods section and prepared figures. M.M.P. wrote
1370 the manuscript. Y.S., B.B., K.L.M., A.M., R.B.J., A.F.M., N.O.C. and M.M.P. edited the final manuscript.

1371

1372 **Conflict of interest**

1373 N.O.C. is a co-founder, shareholder and management consultant for PhenoTherapeutics Ltd. All other
1374 authors declare no competing interests.

1375

1376 **Legend to Figures:**

1377

1378 **Fig. 1. Non-genetic clonal divergence dynamics predicts transcriptomic state bottlenecks. a.** HA1ER
1379 genetic anchor (F12) and two diverging subclones (1F8 and 1C9) were subjected to BdLT-Seq coupled
1380 to scRNA-Seq and traced for 4 days. UMAP plots depict transcriptome states at day 0 and 4 of
1381 tracing. Individual states are depicted in colour and numbered. **b.** Chord diagrams representing
1382 transcriptome state dynamics for cells belonging to a particular state/cluster at the beginning of
1383 tracing and their divergence after 4 days for clone F12 and subclones 1F8 and 1C9. Detected clusters
1384 are depicted (C0 to C12) and integrate the collapsed behaviour of all cells that belong to each gene
1385 expression state. Origin clusters are shown in different colours (Day 0) and chords represent the
1386 endpoint cluster association (Day 4). **c.** Chord diagrams representing individual examples of
1387 transcriptome divergence. C2 and C10 divergence is depicted. Lineage relationship of the clones is
1388 depicted for **a**, **b** and **c**. **d.** Schemes representing transcriptome divergence conceptual framework.

1389

1390 **Fig. 2: Clonal cell systems retain a molecular memory of past cells states. a**, **b** and **c**. Genetic
1391 anchors from the HA1ER (F12), HA1E (10E) and HepG2 (1A) and divergent subclones from each cell
1392 system (HA1ER – 1F8, 1C9, 2A6 and 1D12; HA1E – A7A, A9H, B2F and B5F; HepG2 – 32G, 14F, 48E
1393 and 44D) were subjected to scRNA-Seq. UMAP plots represent transcriptome states and their
1394 divergence among analysed subclones. Individual states are depicted in colour and numbered. **c.**
1395 HepG2 clonal system was subjected to either vehicle (-TGFβ) or TGFβ treatment (+TGFβ) and
1396 processed for scRNA-Seq. UMAP plots represents clonal divergence among clones and between
1397 conditions. Lineage relationship is depicted for **a**, **b** and **c**. **d.** Clonogenicity assay in HA1ER clonal
1398 system (F12, 1F8 and 1C9). Histogram represents the number of colonies per cm^2 normalised to F12
1399 (genetic anchor). Histograms represent the mean value ± standard deviation (SD) of 3 independent
1400 experiments performed in 4-6 replicates. Statistical significance between 1F8 and 1C9 was assessed
1401 by two-tailed unpaired t-test with Welch correction. Individual data points are shown. **e.**
1402 Clonogenicity assay in HA1E clonal system (10E, A7A, A9H, B2F and B5F) upon RAS variants

1403 transduction. Histogram represents the number of colonies per cm^2 normalised to 10E (genetic
1404 anchor). Histograms represent the mean value ± standard deviation (SD) of 4 (3 for A9H)
1405 independent experiments performed in triplicate. Statistical significance between all subclones was
1406 assessed by ordinary one-way ANOVA followed by Tukey's multiple comparisons test. Individual data
1407 points are shown. **f.** HA1ER clonal system (F12, 1F8, 1C9, 2A6 and 1D12) was subjected to transwell
1408 migration assay. Histograms depict the mean area covered by migrated cells ± standard deviation
1409 (SD) of 3 independent experiments performed in triplicate/duplicate. Statistical significance of the
1410 subclones compared to F12 was assessed by ordinary one-way ANOVA followed by Dunnett's
1411 multiple comparisons test. Individual data points are shown. **g.** HepG2 clonal system (1A, 32G, 14F,
1412 48E and 44D) was subjected to transwell (migration) assay. Histograms depict the mean area covered
1413 by migrated cells ± standard deviation (SD) of 3 independent experiments performed in triplicate.
1414 Statistical significance of the subclones compared to 1A was assessed by ordinary one-way ANOVA
1415 followed by Dunnett's multiple comparisons test. Individual data points are shown. Clonal
1416 relationship is depicted for all panels. **h.** Histograms depicting TRAIL-induced apoptosis (1 µg/ml
1417 rhTRAIL for 2h) in HA1ER clonal system (F12, 1F8, 1C9, 2A6 and 1D12) as determined by cleaved
1418 PARP staining analysed by flow cytometry. Histograms show the mean value ± standard deviation
1419 (SD) of 3 biological replicates (~20,000 cells per condition). Statistical significance of the subclones
1420 compared to F12 was assessed by one-way ANOVA followed by Dunnett's multiple comparisons test.
1421 Individual data points are shown.
1422
1423 **Fig. 3. LncRNAs are major contributors to clonal divergence. a.** Diverging clones 1F8 and 1C9
1424 (HA1ER system) were subjected to SLAM-Seq analysis. Heatmap represents differentially expressed
1425 transcripts at the steady-state level (FDR<0.01; FC>2). No differences were observed in active
1426 transcription (FDR<0.05). Scale bar is depicted. Results represent 2 independent experiments
1427 performed in triplicate. **b.** ATAC-Seq was performed on HA1ER clones 1F8 and 1C9 to evaluate
1428 accessible chromatin regions. A representative heatmap displaying the identified peaks as

1429 determined by MACS using FDR<0.05 for each clone is depicted. Peaks distance ± 1Kb. Scale bar is
1430 depicted. **c.** Differentially expressed transcripts between 1F8 and 1C9 extracted from SLAM-Seq data
1431 in **a** were annotated based on their coding capacity. Histogram represents the fraction of coding vs
1432 non-coding transcripts. **d.** Differentially expressed non-coding RNAs were annotated (g:Profiler
1433 version e111_eg58_p18_f463989d[92], Ensembl release 105[90], LNCipedia v5.2[93]). Pie chart depicts the
1434 fraction of differentially expressed non-coding RNAs falling into known categories. **e.** Pairwise
1435 comparison of differentially expressed transcripts (FC>3 up-regulated) in scRNA-Seq datasets of
1436 HA1ER clones depicted in Fig. 2a and their annotation on coding vs non-coding potential. **f.** A subset
1437 of identified differentially expressed lncRNA transcripts between HA1ER clones 1F8 vs 1C9 were
1438 validated by reverse transcription followed by qPCR (RT-qPCR). Histograms depict AL161733.1, DPP4-
1439 DT, EMP2-AS and AC008464.1 transcript levels normalised to GAPDH levels. Results represent the
1440 mean ± standard deviation (SD) of 2 independent experiments performed in technical triplicates.
1441 Statistical significance was assessed by two-tailed unpaired Student's t-test. Individual data points
1442 are depicted. **g.** A subset of differentially expressed lncRNA transcripts between 1F8 vs 1C9 were
1443 analysed for their intracellular localisation by subcellular fractionation (nucleus (nuc) vs cytoplasm
1444 (cyt)) and their levels evaluated by RT-qPCR. Histograms depict RNA levels of each fraction upon
1445 normalisation to the spike-in (*in-vitro* transcribed partial eGFP transcript). Data for AL161733.1,
1446 DPP4-DT, EMP2-AS and AC008464.1 is depicted. MALAT1 (enriched in nucleus), GAPDH intron
1447 (enriched in nucleus) and GAPDH exon-exon junction (enriched in cytoplasm) are shown as
1448 fractionation controls. Data represents the mean ± standard deviation (SD) of 3-4 independent
1449 experiments performed in technical triplicates. Statistical significance was assessed by two-tailed
1450 unpaired Student's t-test. Individual data points are depicted.
1451
1452 **Fig. 4. IDR-containing proteins show clonal divergence. a.** Clones 1F8 and 1C9 from the HA1ER
1453 system were subjected to label-free LC-MS/MS. Heatmaps depict the identified differentially
1454 expressed proteins (FDR<0.01; FC>1.5). Data was pooled from 2 independent experiments containing

1455 4 and 5 biological replicates each. Scale bar is depicted. **b.** Differentially expressed proteins identified
1456 in A were annotated considering their intrinsically disordered region (IDR) potential (AlphaFold2[16,17]).
1457 Histogram represents the fraction of IDR vs non-IDR proteins identified as differentially expressed. **c.**
1458 Clones from **a** were analysed by TMT-based phospho-proteomics. Heatmap (left) represents
1459 differentially represented phospho-sites between 1F8 and 1C9 (FDR<0.05; FC>0.4). Data was
1460 obtained from 5 biological replicates. Accompanying heatmap (right) displays total protein levels for
1461 differentially phosphorylated proteins identified. Scale bars are depicted. Letters at the end of a
1462 phosphorylation site represent the detection of a different peptide. **d.** A subset of differentially
1463 expressed IDR-containing proteins (1F8 vs 1C9) were analysed at a single-cell level by mass
1464 cytometry (CyTOF). Levels of MOT4, JIP4 and SYUA (all IDR-containing) were evaluated using specific
1465 antibodies in F12, 1F8 and 1C9. DCXR is also shown as a non-IDR diverging protein control. Scale bar
1466 is depicted. **e.** Heatmap represents the clustered expression of the antibody panel (MOT4, JIP4, SYUA
1467 and DCXR. PD-L1 and CD206 are included as a positive and negative control respectively) in collective
1468 data from all clones. The fraction of cells analysed in each cluster is depicted. Scale bar is depicted. **f.**
1469 Heatmaps represent the individual analysis for each antibody in the CyTOF panel (MOT4, JIP4, SYUA
1470 and DCXR) in each clone (F12, 1F8 and 1C9). Clustered expression levels are depicted. Scale bar is
1471 depicted.
1472
1473 **Fig. 5. Cytoplasmic lncRNAs co-localize with IDR-containing proteins and activated signalling**
1474 **pathways. a.** HA1ER clones 1F8 and 1C9 were subjected to immunocytochemistry for JIP4, SYUA and
1475 AP1G1 coupled to RNA-FISH targeting DPP4-DT or EMP2-AS lncRNAs and analysed by Airyscan
1476 microscopy. Upper panels depict individual channels (DAPI, JIP4 and DPP4-DT) and a merged image
1477 (JIP4/DPP4-DT) whilst lower panels depict overlay images for JIP4/EMP2-AS, SYUA/DPP4-DT,
1478 SYUA/EMP2-AS, AP1G1/DPP4-DT and AP1G1/EMP2-AS. Inset is displayed for the upper panel. Scale
1479 bar 10 µm. **b.** HA1ER clone 1F8 was subjected to immunocytochemistry for activated signalling
1480 pathways coupled to RNA-FISH targeting DPP4-DT lncRNA. Upper panels depict individual channels

1481 (DAPI, pERK1/2 and DPP4-DT) and a merged image (pERK1/2/DPP4-DT) whilst the lower panels
1482 depict merged images for DPP4-DT/pBRAF, DPP4-DT/pMEK1/2, DPP4-DT/pRAC1, DPP4-DT/pPI3K,
1483 DPP4-DT/pAKT, DPP4-DT/pMKK4, DPP4-DT/p-p38, DPP4-DT/pPAK1/2, DPP4-DT/pNFκB and DPP4-
1484 DT/pIκBα. Inset is displayed for the upper panel. Scale bar 10 µm. **c.** HepG2 clone 14F was either
1485 treated with TGFβ or with vehicle (+TGFβ or -TGFβ respectively) and subjected to
1486 immunocytochemistry for either active (pSMAD2, pSMAD3 and pRAC1) or their corresponding non-
1487 active (SMAD2, SMAD3 and RAC1) TGFβ-dependent signalling proteins coupled to RNA-FISH
1488 targeting PINCR lncRNA. Upper panel depicts individual channels (DAPI, pSMAD3 and PINCR) and an
1489 overlaid image (pSMAD3/PINCR) whilst the lower panels display SMAD3/PINCR, pSMAD2/PINCR,
1490 SMAD2/PINCR, pRAC1/PINCR and RAC1/PINCR. Inset is displayed for the upper panel. Scale bar 10
1491 µm.
1492
1493 **Fig. 6. SIP destabilisation promotes transcriptome and signalling pathway re-organisation. a.**
1494 HA1ER clone 1F8 was subjected to antisense mediated knock-down (Gapmer) for DPP4-DT or
1495 AL161733.1 lncRNAs and processed for scRNA-Seq after 48h (24h for DPP4-DT-G2). UMAP plots
1496 depict transcriptome states in control (non-targeting control gapmer) and upon knockdown (DPP4-
1497 DT-G1, DPP4-DT-G2, AL161733.1-G1 and AL161733.1-G2). Two alternative antisense molecules were
1498 used targeting different regions of the target transcript (G1 and G2). Individual states are depicted in
1499 colour and numbered. **b.** HA1ER clone 1F8 was subjected to antisense mediated knock-down
1500 (Gapmer) for DPP4-DT or AL161733.1 lncRNAs and further processed for Reverse Phase Protein
1501 Arrays (RPPA) after 48h (24h for DPP4-DT-G2) to evaluate signalling integrity. Two alternative
1502 antisense molecules were used targeting different regions of the target transcript (G1 and G2).
1503 Heatmaps display differentially represented signalling pathway members (≥20% increase/decrease to
1504 control) upon either DPP4-DT or AL161733.1 knockdown. Two biological replicates were processed
1505 for each condition. Scale bars are depicted **c.** Pathway analysis (Reactome) on HA1ER clone 1F8 upon
1506 DPP4-DT or AL161733.1 lncRNAs knockdown (FDR<0.01). Top 10 upregulated pathways are shown

1507 for antibody signals specifically upregulated upon DPP4-DT (DPP4-DT-specific) or AL161733.1 knock-
1508 down (AL161733.1-specific), as well as for upregulated signalling pathway members detected in both
1509 knock-downs (overlap). **d.** Venn diagrams depict the intersection between the Top 50 enriched terms
1510 (Reactome, Up- or downregulated) from DPP4-DT and AL161733.1 lncRNAs knockdown datasets for
1511 specific or overlapping signalling pathway members identified. **e.** lncRNA knockdown was evaluated
1512 for cells processed for scRNA-Seq shown in **a** (left) or for cells subjected to RPPA shown in **b** (right).
1513 Transcript levels of lncRNAs DPP4-DT and AL161733.1 were determined by reverse transcription
1514 followed by qPCR (RT-qPCR) and were normalised to GAPDH. Histograms represent the mean ±
1515 propagated error of 3 technical replicates. Individual data points are depicted.
1516
1517 **Fig. 7. SIP disruption results in intracellular relocalisation of signalling pathways and alters**
1518 **phenotypic output. a.** HA1ER clone 1F8 was subjected to antisense-mediated knock-down for DPP4-
1519 DT lncRNA (DPP4-DT-G1) or non-targeting control gapmer (ctrl) and further processed for
1520 immunocytochemistry for activated signalling pathways (pRAC1, pPAK1/2, pMKK4 and pMEK1/2)
1521 coupled to RNA-FISH targeting lncRNA DPP4-DT lncRNA. Panels depict overlay images of signalling
1522 protein (green) and DPP4-DT (red) and are accompanied by an inset. Bar 10 µm. **b.** Cells from HA1ER
1523 clone 1C9 were subjected to antisense-mediated knock-down (gapmer) targeting lncRNA
1524 AC008464.1 or non-targeting control (ctrl) and further processed for clonogenic soft agar assay.
1525 Clonogenic potential is presented as the number of colonies per cm^2 upon knock-down normalised
1526 to the control. Histograms represent the mean value ± standard deviation (SD) of 3 independent
1527 experiments performed in 4-6 replicates. Individual data points are shown. **c.** Knock-down of lncRNA
1528 AC008464.1 prior to clonogenic assay shown in **b** was evaluated by reverse transcription followed by
1529 qPCR (RT-qPCR) normalised to GAPDH. Histograms represent the mean ± SD of 3 independent
1530 experiments. Individual data points are depicted. **d.** Toy diagram depicting the link between cellular
1531 state, SIP composition and phenotype. Identified components are depicted on the right.
1532

1533 **References**

1534 1 Moran, K. L., Shlyakhtina, Y. & Portal, M. M. The role of non-genetic information in
1535 evolutionary frameworks. *Crit Rev Biochem Mol Biol* **56**, 255-283 (2021).
1536 https://doi.org/10.1080/10409238.2021.1908949
1537 2 Shlyakhtina, Y., Moran, K. L. & Portal, M. M. Genetic and Non-Genetic Mechanisms
1538 Underlying Cancer Evolution. *Cancers (Basel)* **13** (2021).
1539 https://doi.org/10.3390/cancers13061380
1540 3 Yona, A. H., Frumkin, I. & Pilpel, Y. A relay race on the evolutionary adaptation
1541 spectrum. *Cell* **163**, 549-559 (2015). https://doi.org/10.1016/j.cell.2015.10.005
1542 4 Aissa, A. F. *et al.* Single-cell transcriptional changes associated with drug tolerance
1543 and response to combination therapies in cancer. *Nat Commun* **12**, 1628 (2021).
1544 https://doi.org/10.1038/s41467-021-21884-z
1545 5 Kashima, Y. *et al.* Single-Cell Analyses Reveal Diverse Mechanisms of Resistance to
1546 EGFR Tyrosine Kinase Inhibitors in Lung Cancer. *Cancer Res* **81**, 4835-4848 (2021).
1547 https://doi.org/10.1158/0008-5472.CAN-20-2811
1548 6 Ma, P. *et al.* Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated
1549 heterogeneous cellular states. *Cell* **186**, 877-891 e814 (2023).
1550 https://doi.org/10.1016/j.cell.2023.01.002
1551 7 McNulty, R. *et al.* Probe-based bacterial single-cell RNA sequencing predicts toxin
1552 regulation. *Nat Microbiol* **8**, 934-945 (2023). https://doi.org/10.1038/s41564-023-
1553 01348-4
1554 8 Shlyakhtina, Y., Bloechl, B. & Portal, M. M. BdLT-Seq as a barcode decay-based
1555 method to unravel lineage-linked transcriptome plasticity. *Nat Commun* **14**, 1085
1556 (2023). https://doi.org/10.1038/s41467-023-36744-1
1557 9 Chen, C. M. & Yu, R. A multi-step completion process model of cell plasticity. *Brief*
1558 *Bioinform* **26** (2025). https://doi.org/10.1093/bib/bbaf165
1559 10 Rockne, R. C. *et al.* State-Transition Analysis of Time-Sequential Gene Expression
1560 Identifies Critical Points That Predict Development of Acute Myeloid Leukemia.
1561 *Cancer Res* **80**, 3157-3169 (2020). https://doi.org/10.1158/0008-5472.CAN-20-0354
1562 11 Rukhlenko, O. S. *et al.* Control of cell state transitions. *Nature* **609**, 975-985 (2022).
1563 https://doi.org/10.1038/s41586-022-05194-y
1564 12 Zhou, P., Wang, S., Li, T. & Nie, Q. Dissecting transition cells from single-cell
1565 transcriptome data through multiscale stochastic dynamics. *Nat Commun* **12**, 5609
1566 (2021). https://doi.org/10.1038/s41467-021-25548-w
1567 13 Mojtahedi, M. *et al.* Cell Fate Decision as High-Dimensional Critical State Transition.
1568 *PLoS Biol* **14**, e2000640 (2016). https://doi.org/10.1371/journal.pbio.2000640
1569 14 Huang, S. The molecular and mathematical basis of Waddington's epigenetic
1570 landscape: a framework for post-Darwinian biology? *Bioessays* **34**, 149-157 (2012).
1571 https://doi.org/10.1002/bies.201100031
1572 15 Waddington, C. H. *The Strategy of the Genes*. (George Allen and Unwin, London, UK.,
1573 1957).
1574 16 Ferrell, J. E., Jr. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr*
1575 *Biol* **22**, R458-466 (2012). https://doi.org/10.1016/j.cub.2012.03.045
1576 17 Huang, S. Non-genetic heterogeneity of cells in development: more than just noise.
1577 *Development* **136**, 3853-3862 (2009). https://doi.org/10.1242/dev.035139

1578 18 Pisco, A. O. & Huang, S. Non-genetic cancer cell plasticity and therapy-induced
1579 stemness in tumour relapse: 'What does not kill me strengthens me'. *Br J Cancer* **112**,
1580 1725-1732 (2015). https://doi.org/10.1038/bjc.2015.146
1581 19 Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements.
1582 *Nature* **400**, 464-468 (1999). https://doi.org/10.1038/22780
1583 20 Liu, Y., Li, X., Zhang, H. & Huang, Y. MIAT inhibits proliferation of cervical cancer cells
1584 through regulating miR-150-5p. *Cancer Cell Int* **20**, 242 (2020).
1585 https://doi.org/10.1186/s12935-020-01338-0
1586 21 Macpherson, I. & Montagnier, L. Agar Suspension Culture for the Selective Assay of
1587 Cells Transformed by Polyoma Virus. *Virology* **23**, 291-294 (1964).
1588 https://doi.org/10.1016/0042-6822(64)90301-0
1589 22 Pires, M. M., Hopkins, B. D., Saal, L. H. & Parsons, R. E. Alterations of EGFR, p53 and
1590 PTEN that mimic changes found in basal-like breast cancer promote transformation
1591 of human mammary epithelial cells. *Cancer Biol Ther* **14**, 246-253 (2013).
1592 https://doi.org/10.4161/cbt.23297
1593 23 Rangarajan, A., Hong, S. J., Gifford, A. & Weinberg, R. A. Species- and cell type-
1594 specific requirements for cellular transformation. *Cancer Cell* **6**, 171-183 (2004).
1595 https://doi.org/10.1016/j.ccr.2004.07.009
1596 24 Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat*
1597 *Methods* **14**, 1198-1204 (2017). https://doi.org/10.1038/nmeth.4435
1598 25 Matsushima, W. *et al.* Sequencing cell-type-specific transcriptomes with SLAM-ITseq.
1599 *Nat Protoc* **14**, 2261-2278 (2019). https://doi.org/10.1038/s41596-019-0179-x
1600 26 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate
1601 read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791 (2013).
1602 https://doi.org/10.1093/bioinformatics/btt468
1603 27 Neumann, T. *et al.* Quantification of experimentally induced nucleotide conversions
1604 in high-throughput sequencing datasets. *BMC Bioinformatics* **20**, 258 (2019).
1605 https://doi.org/10.1186/s12859-019-2849-7
1606 28 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition
1607 of native chromatin for fast and sensitive epigenomic profiling of open chromatin,
1608 DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
1609 https://doi.org/10.1038/nmeth.2688
1610 29 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for
1611 Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-
1612 21 29 29 (2015). https://doi.org/10.1002/0471142727.mb2129s109
1613 30 Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* **52**,
1614 D672-D678 (2024). https://doi.org/10.1093/nar/gkad1025
1615 31 Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for
1616 comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-
1617 1904 (2003). https://doi.org/10.1021/ac0262560
1618 32 Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and
1619 recommendations. *Nat Rev Mol Cell Biol* **24**, 430-447 (2023).
1620 https://doi.org/10.1038/s41580-022-00566-8
1621 33 Guo, Q., Shi, X. & Wang, X. RNA and liquid-liquid phase separation. *Noncoding RNA*
1622 *Res* **6**, 92-99 (2021). https://doi.org/10.1016/j.ncrna.2021.04.003

1623 34 Chujo, T., Yamazaki, T. & Hirose, T. Architectural RNAs (arcRNAs): A class of long
1624 noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim Biophys Acta*
1625 **1859**, 139-146 (2016). https://doi.org/10.1016/j.bbagrm.2015.05.007
1626 35 Luo, J. *et al.* LncRNAs: Architectural Scaffolds or More Potential Roles in Phase
1627 Separation. *Front Genet* **12**, 626234 (2021).
1628 https://doi.org/10.3389/fgene.2021.626234
1629 36 Yamazaki, T. *et al.* Functional Domains of NEAT1 Architectural lncRNA Induce
1630 Paraspeckle Assembly through Phase Separation. *Mol Cell* **70**, 1038-1053 e1037
1631 (2018). https://doi.org/10.1016/j.molcel.2018.05.019
1632 37 Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA
1633 is essential for the structure of paraspeckles. *Mol Cell* **33**, 717-726 (2009).
1634 https://doi.org/10.1016/j.molcel.2009.01.026
1635 38 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates:
1636 organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
1637 https://doi.org/10.1038/nrm.2017.7
1638 39 Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence
1639 domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).
1640 https://doi.org/10.1016/j.cell.2012.04.017
1641 40 Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates
1642 environmentally responsive membraneless organelles. *Mol Cell* **57**, 936-947 (2015).
1643 https://doi.org/10.1016/j.molcel.2015.01.013
1644 41 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*
1645 **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2
1646 42 Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure
1647 coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**, D368-D375
1648 (2024). https://doi.org/10.1093/nar/gkad1011
1649 43 Macleod, K. G., Serrels, B. & Carragher, N. O. Reverse Phase Protein Arrays and Drug
1650 Discovery. *Methods Mol Biol* **1647**, 153-169 (2017). https://doi.org/10.1007/978-1-
1651 4939-7201-2_10
1652 44 Morawska, L. P., Hernandez-Valdes, J. A. & Kuipers, O. P. Diversity of bet-hedging
1653 strategies in microbial communities-Recent cases and insights. *WIREs Mech Dis* **14**,
1654 e1544 (2022). https://doi.org/10.1002/wsbm.1544
1655 45 O'Dea, R. E., Noble, D. W. A., Johnson, S. L., Hesselson, D. & Nakagawa, S. The role of
1656 non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet
1657 hedging and epigenetic traps. *Environ Epigenet* **2**, dvv014 (2016).
1658 https://doi.org/10.1093/eep/dvv014
1659 46 Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity--a mutation-independent
1660 driving force for the somatic evolution of tumours. *Nat Rev Genet* **10**, 336-342
1661 (2009). https://doi.org/10.1038/nrg2556
1662 47 Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-
1663 wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-
1664 547 (2008). https://doi.org/10.1038/nature06965
1665 48 Charlebois, D. A. Effect and evolution of gene expression noise on the fitness
1666 landscape. *Phys Rev E Stat Nonlin Soft Matter Phys* **92**, 022713 (2015).
1667 https://doi.org/10.1103/PhysRevE.92.022713

1668 49 Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression:
1669 from theories to phenotypes. *Nat Rev Genet* **6**, 451-464 (2005).
1670 https://doi.org/10.1038/nrg1615
1671 50 Hangauer, M. J. *et al.* Drug-tolerant persister cancer cells are vulnerable to GPX4
1672 inhibition. *Nature* **551**, 247-250 (2017). https://doi.org/10.1038/nature24297
1673 51 Knoechel, B. *et al.* An epigenetic mechanism of resistance to targeted therapy in T
1674 cell acute lymphoblastic leukemia. *Nat Genet* **46**, 364-370 (2014).
1675 https://doi.org/10.1038/ng.2913
1676 52 Kurata, T. *et al.* Effect of re-treatment with gefitinib ('Iressa', ZD1839) after
1677 acquisition of resistance. *Ann Oncol* **15**, 173-174 (2004).
1678 https://doi.org/10.1093/annonc/mdh006
1679 53 Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell
1680 subpopulations. *Cell* **141**, 69-80 (2010). https://doi.org/10.1016/j.cell.2010.02.027
1681 54 Yano, S. *et al.* Retreatment of lung adenocarcinoma patients with gefitinib who had
1682 experienced favorable results from their initial treatment with this selective
1683 epidermal growth factor receptor inhibitor: a report of three cases. *Oncol Res* **15**,
1684 107-111 (2005).
1685 55 Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue
1686 of origin. *Nat Biotechnol* **33**, 503-509 (2015). https://doi.org/10.1038/nbt.3209
1687 56 Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to
1688 embryonic stem cells. *Cell* **161**, 1187-1201 (2015).
1689 https://doi.org/10.1016/j.cell.2015.04.044
1690 57 Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and
1691 combinatorial labeling. *Nat Methods* **9**, 743-748 (2012).
1692 https://doi.org/10.1038/nmeth.2069
1693 58 Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA
1694 profiling by sequential hybridization. *Nat Methods* **11**, 360-361 (2014).
1695 https://doi.org/10.1038/nmeth.2892
1696 59 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual
1697 Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
1698 https://doi.org/10.1016/j.cell.2015.05.002
1699 60 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin
1700 state. *Nat Biotechnol* **33**, 1165-1172 (2015). https://doi.org/10.1038/nbt.3383
1701 61 Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming.
1702 *Nature* **564**, 219-224 (2018). https://doi.org/10.1038/s41586-018-0744-4
1703 62 Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding
1704 RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96-118 (2021).
1705 https://doi.org/10.1038/s41580-020-00315-9
1706 63 Barbieri, E. *et al.* Rapid and Scalable Profiling of Nascent RNA with fastGRO. *Cell Rep*
1707 **33**, 108373 (2020). https://doi.org/10.1016/j.celrep.2020.108373
1708 64 Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat Rev*
1709 *Genet* **13**, 246-259 (2012). https://doi.org/10.1038/nrg3160
1710 65 Yamada, T. & Akimitsu, N. Contributions of regulated transcription and mRNA decay
1711 to the dynamics of gene expression. *Wiley Interdiscip Rev RNA* **10**, e1508 (2019).
1712 https://doi.org/10.1002/wrna.1508

1713 66 Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA
1714 modifications. *Nat Rev Mol Cell Biol* **18**, 31-42 (2017).
1715 https://doi.org/10.1038/nrm.2016.132
1716 67 Zimonjic, D., Brooks, M. W., Popescu, N., Weinberg, R. A. & Hahn, W. C. Derivation of
1717 human tumor cells in vitro without widespread genomic instability. *Cancer Res* **61**,
1718 8838-8844 (2001).
1719 68 Chao, S., Yan, H. & Bu, P. Asymmetric division of stem cells and its cancer relevance.
1720 *Cell Regen* **13**, 5 (2024). https://doi.org/10.1186/s13619-024-00188-9
1721 69 Knoblich, J. A. Mechanisms of asymmetric stem cell division. *Cell* **132**, 583-597
1722 (2008). https://doi.org/10.1016/j.cell.2008.02.007
1723 70 Shlyakhtina, Y., Moran, K. L. & Portal, M. M. Asymmetric Inheritance of Cell Fate
1724 Determinants: Focus on RNA. *Noncoding RNA* **5** (2019).
1725 https://doi.org/10.3390/ncrna5020038
1726 71 Zion, E. H., Chandrasekhara, C. & Chen, X. Asymmetric inheritance of epigenetic
1727 states in asymmetrically dividing stem cells. *Curr Opin Cell Biol* **67**, 27-36 (2020).
1728 https://doi.org/10.1016/j.ceb.2020.08.003
1729 72 Mulas, C., Chaigne, A., Smith, A. & Chalut, K. J. Cell state transitions: definitions and
1730 challenges. *Development* **148** (2021). https://doi.org/10.1242/dev.199950
1731 73 Raghavan, S. *et al.* Microenvironment drives cell state, plasticity, and drug response
1732 in pancreatic cancer. *Cell* **184**, 6119-6137 e6126 (2021).
1733 https://doi.org/10.1016/j.cell.2021.11.017
1734 74 Russo, M. *et al.* Cancer drug-tolerant persister cells: from biological questions to
1735 clinical opportunities. *Nat Rev Cancer* **24**, 694-717 (2024).
1736 https://doi.org/10.1038/s41568-024-00737-z
1737 75 Shlyakhtina, Y., Bloechl, B., Moran, K. L. & Portal, M. M. Protocol to study the
1738 inheritance and propagation of non-genetically encoded states using barcode decay
1739 lineage tracing. *STAR Protoc* **5**, 102809 (2024).
1740 https://doi.org/10.1016/j.xpro.2023.102809
1741 76 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587
1742 e3529 (2021). https://doi.org/10.1016/j.cell.2021.04.048
1743 77 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single
1744 cells. *Nat Methods* **14**, 865-868 (2017). https://doi.org/10.1038/nmeth.4380
1745 78 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1746 reads. *EMBnet.journal* **17**, 10-12 (2011).
1747 https://doi.org/https://doi.org/10.14806/ej.17.1.200
1748 79 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics.
1749 *Genome Res* **19**, 1639-1645 (2009). https://doi.org/10.1101/gr.092759.109
1750 80 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
1751 https://doi.org/10.1093/gigascience/giab008
1752 81 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration
1753 discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576 (2012).
1754 https://doi.org/10.1101/gr.129684.111
1755 82 Knaus, B. J. & Grunwald, N. J. vcfr: a package to manipulate and visualize variant call
1756 format data in R. *Mol Ecol Resour* **17**, 44-53 (2017). https://doi.org/10.1111/1755-
1757 0998.12549

83 Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc* **17**, 1518-1552 (2022). https://doi.org/10.1038/s41596-022-00692-9

84 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040 (2014). https://doi.org/10.1101/gr.177881.114

85 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012). https://doi.org/10.1038/nmeth.1923

86 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org/10.1186/gb-2008-9-9-r137

87 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016). https://doi.org/10.1093/nar/gkw257

88 Ellis B, H. P., Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M, Finak G. flowCore: Basic structures for flow cytometry data. R package version 2.18.0. (2024).

89 Crowell H, Z. V., Chevrier S, Robinson M CATALYST: Cytometry dATa anALYSis Tools. R package version 1.30.0. (2024). <https://github.com/HelenaLC/CATALYST>.

90 Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res* **49**, D884-D891 (2021). https://doi.org/10.1093/nar/gkaa942

91 Perez-Riverol, Y. *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res* **53**, D543-D553 (2025). https://doi.org/10.1093/nar/gkae1011

92 Kolberg, L. *et al.* g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207-W212 (2023). https://doi.org/10.1093/nar/gkad347

93 Volders, P. J. *et al.* LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res* **47**, D135-D139 (2019). https://doi.org/10.1093/nar/gky1031

94 Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947, doi:10.1093/nar/gky1015 (2019)

Figure 1

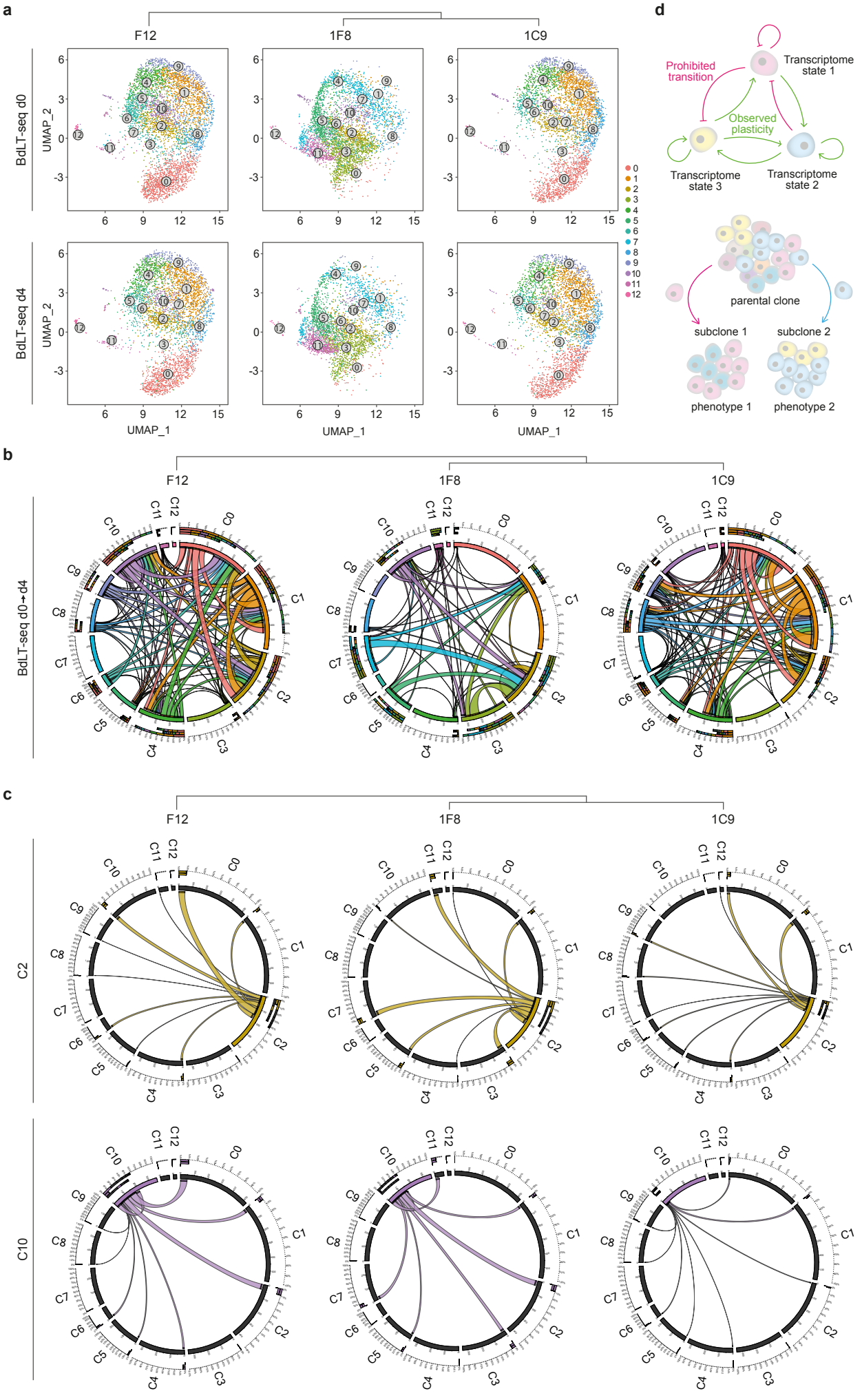

a
F12
1F8
1C9
BdLT-seq d0
BdLT-seq d4
UMAP_2
UMAP_1
0
1
2
3
4
5
6
7
8
9
10
11
12
d
Prohibited transition
Transcriptome state 1
Observed plasticity
Transcriptome state 3
Transcriptome state 2
parental clone
subclone 1
subclone 2
phenotype 1
phenotype 2
b
BdLT-seq d0→d4
c
C2
C10

Figure 2

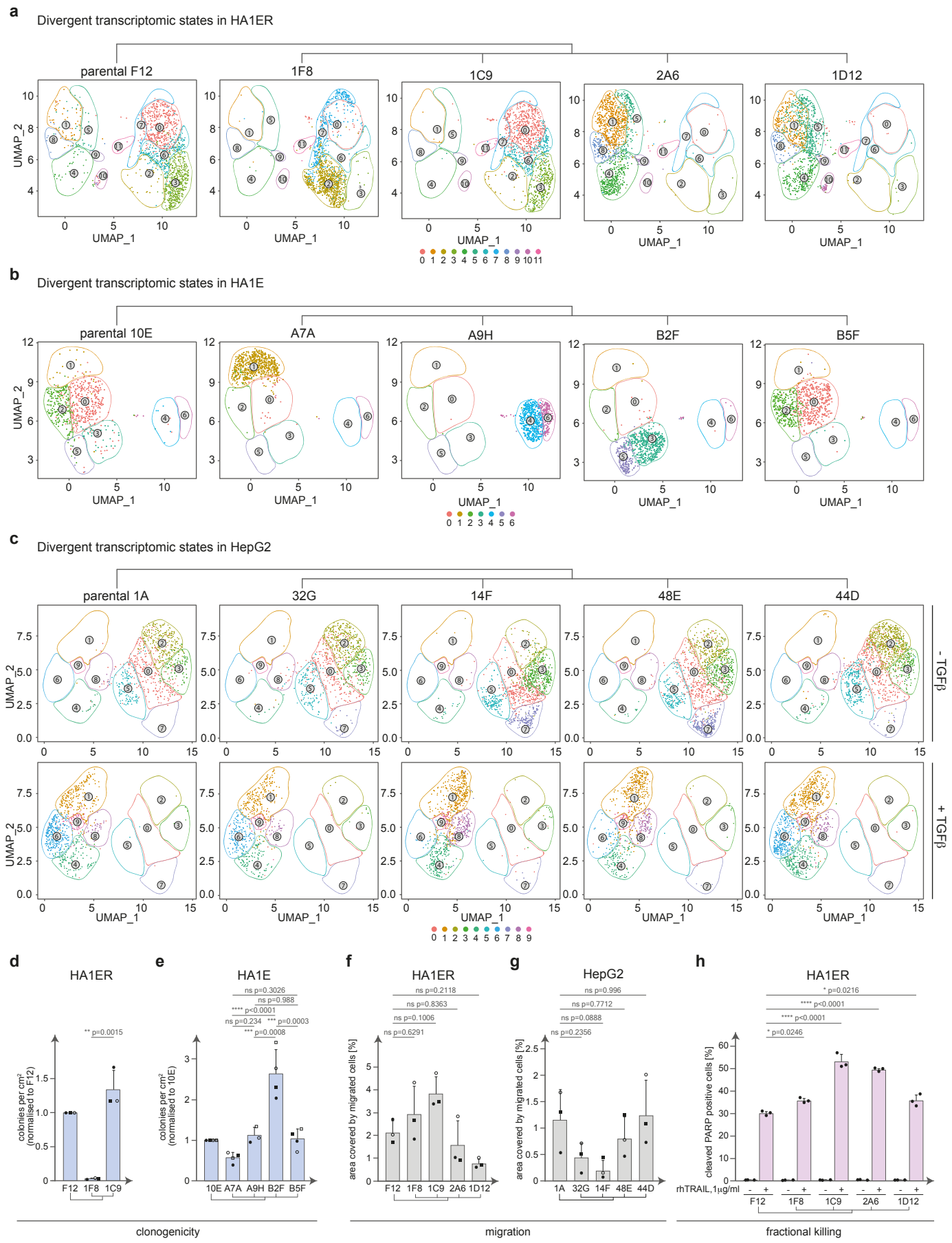

a
Divergent transcriptomic states in HA1ER
parental F12
1F8
1C9
2A6
1D12
UMAP_1
UMAP_2
b
Divergent transcriptomic states in HA1E
parental 10E
A7A
A9H
B2F
B5F
c
Divergent transcriptomic states in HepG2
parental 1A
32G
14F
48E
44D
- TGFβ
+ TGFβ
d
HA1ER
colonies per cm² (normalised to F12)
** p=0.0015
F12 1F8 1C9
e
HA1E
colonies per cm² (normalised to 10E)
ns p=0.3026
ns p=0.988
**** p<0.0001
ns p=0.234
*** p=0.0003
*** p=0.0008
10E A7A A9H B2F B5F
f
HA1ER
area covered by migrated cells [%]
ns p=0.2118
ns p=0.8363
ns p=0.1006
ns p=0.6291
F12 1F8 1C9 2A6 1D12
g
HepG2
area covered by migrated cells [%]
ns p=0.996
ns p=0.7712
ns p=0.0888
ns p=0.2356
1A 32G 14F 48E 44D
h
HA1ER
cleaved PARP positive cells [%]
* p=0.0216
**** p<0.0001
**** p<0.0001
* p=0.0246
rhTRAIL,1µg/ml
F12 1F8 1C9 2A6 1D12
clonogenicity
migration
fractional killing

Figure 3

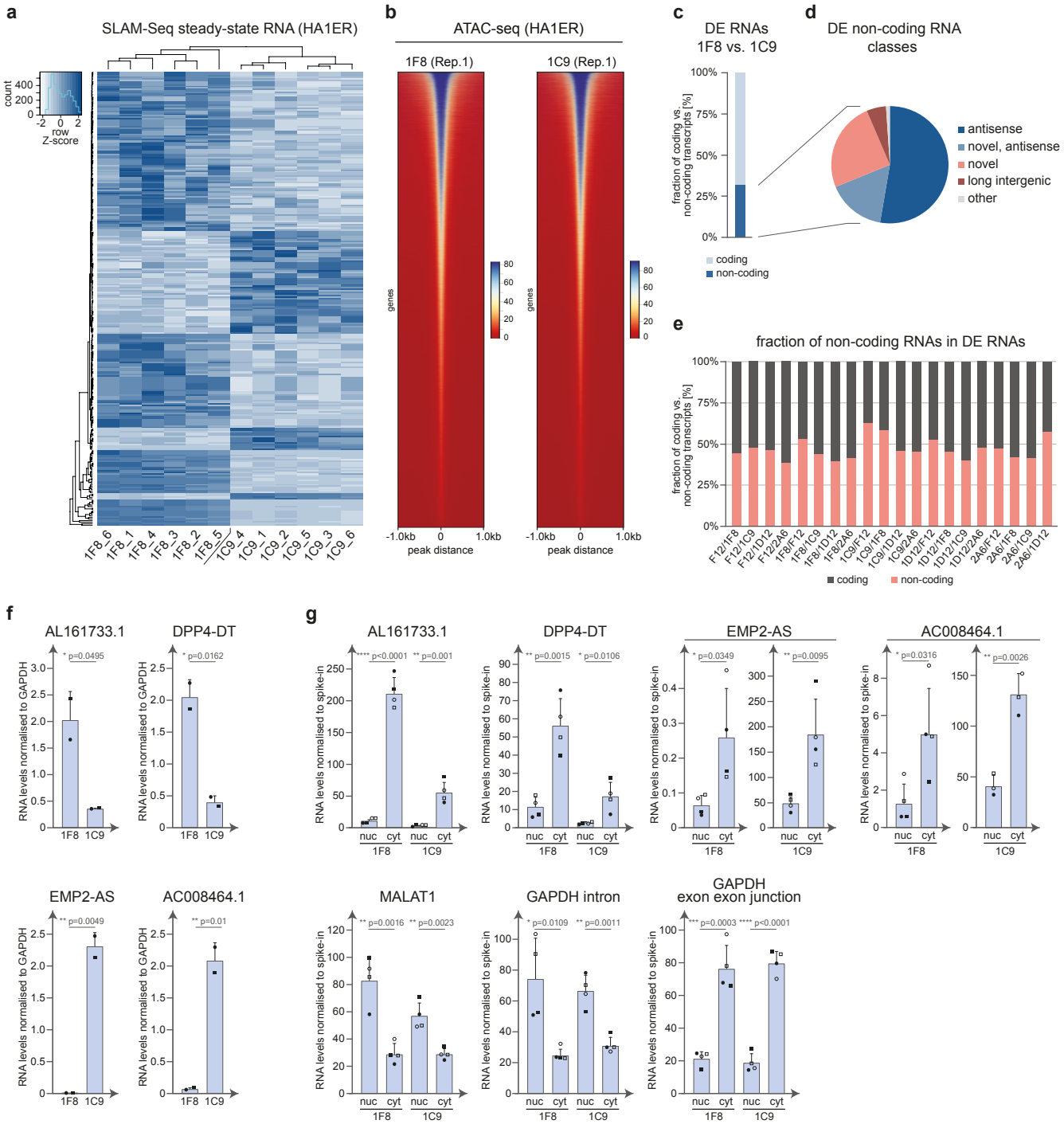

a
SLAM-Seq steady-state RNA (HA1ER)
count
row Z-score
1F8_6 1F8_1 1F8_4 1F8_3 1F8_2 1F8_5 1C9_4 1C9_1 1C9_2 1C9_5 1C9_3 1C9_6
b
ATAC-seq (HA1ER)
1F8 (Rep.1)
1C9 (Rep.1)
genes
-1.0kb 0 1.0kb
peak distance
c
DE RNAs 1F8 vs. 1C9
fraction of coding vs. non-coding transcripts [%]
coding
non-coding
d
DE non-coding RNA classes
antisense
novel, antisense
novel
long intergenic
other
e
fraction of non-coding RNAs in DE RNAs
F12/1F8 F12/1C9 F12/1D12 F12/2A6 1F8/F12 1F8/1C9 1F8/1D12 1F8/2A6 1C9/F12 1C9/1F8 1C9/1D12 1C9/2A6 1D12/F12 1D12/1F8 1D12/1C9 1D12/2A6 2A6/F12 2A6/1F8 2A6/1C9 2A6/1D12
f
AL161733.1
DPP4-DT
EMP2-AS
AC008464.1
RNA levels normalised to GAPDH
1F8 1C9
g
RNA levels normalised to spike-in
nuc cyt
MALAT1
GAPDH intron
GAPDH exon exon junction

Figure 4

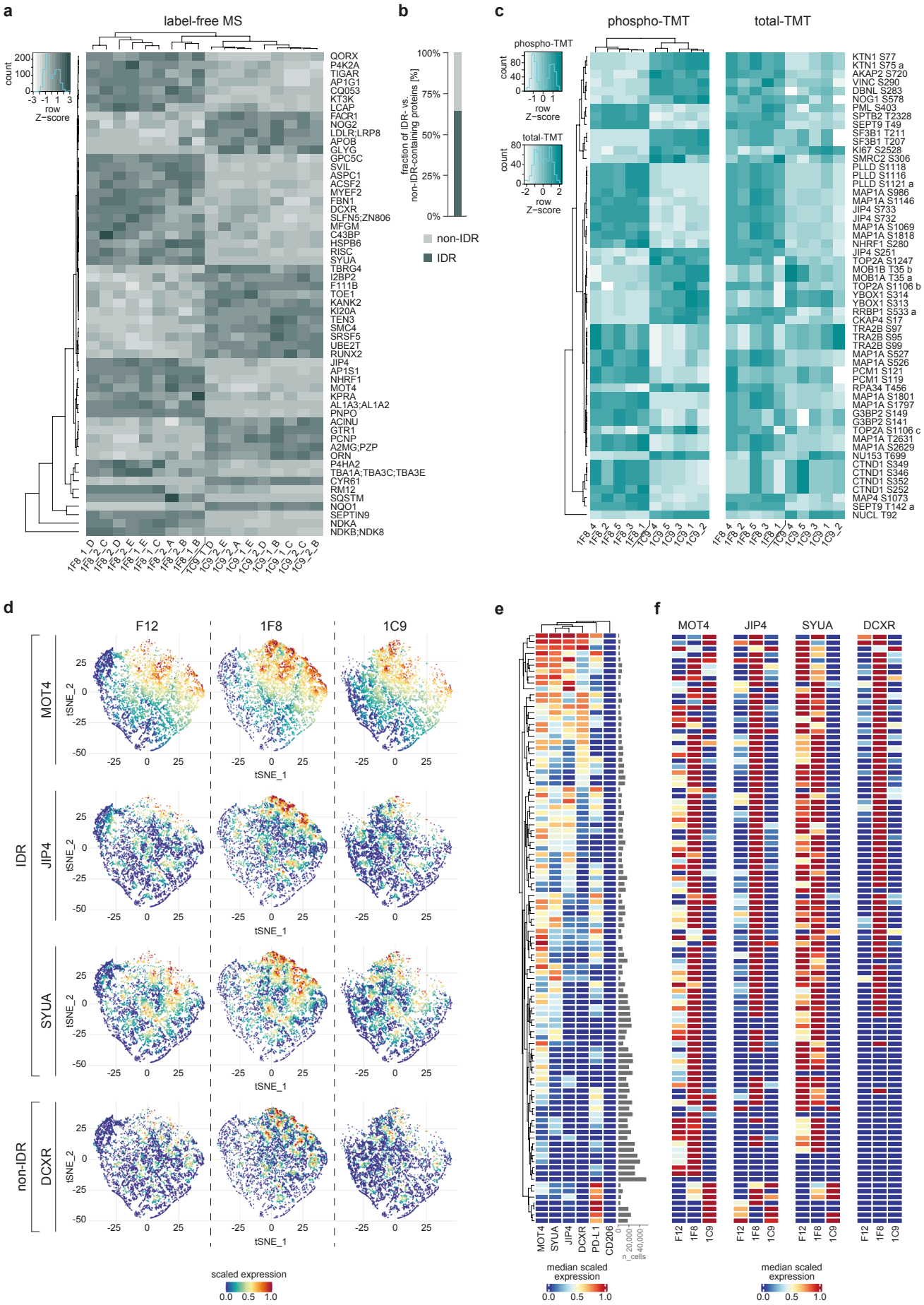

a
label-free MS
count
row Z-score
QORX
P4K2A
TIGAR
AP1G1
CQ053
KT3K
LCAP
FACR1
NOG2
LDLR;LRP8
APOB
GLYG
GPC5C
SVIL
ASPC1
ACSF2
MYEF2
FBN1
DCXR
SLFN5;ZN806
MFGM
C43BP
HSPB6
RISC
SYUA
TBRG4
I2BP2
F111B
TOE1
KANK2
KI20A
TEN3
SMC4
SRSF5
UBE2T
RUNX2
JIP4
AP1S1
NHRF1
MOT4
KPRA
AL1A3;AL1A2
PNPO
ACINU
GTR1
PCNP
A2MG;PZP
ORN
P4HA2
TBA1A;TBA3C;TBA3E
CYR61
RM12
SQSTM
NQO1
SEPTIN9
NDKA
NDKB;NDK8
b
fraction of IDR- vs. non-IDR-containing proteins [%]
non-IDR
IDR
c
phospho-TMT
total-TMT
KTN1 S77
KTN1 S75 a
AKAP2 S720
VINC S290
DBNL S283
NOG1 S578
PML S403
SPTB2 T2328
SEPT9 T49
SF3B1 T211
SF3B1 T207
KI67 S2528
SMRC2 S306
PLLD S1118
PLLD S1116
PLLD S1121 a
MAP1A S986
MAP1A S1146
JIP4 S733
JIP4 S732
MAP1A S1069
MAP1A S1818
NHRF1 S280
JIP4 S251
TOP2A S1247
MOB1B T35 b
MOB1A T35 a
TOP2A S1106 b
YBOX1 S314
YBOX1 S313
RRBP1 S533 a
CKAP4 S17
TRA2B S97
TRA2B S95
TRA2B S99
MAP1A S527
MAP1A S526
PCM1 S121
PCM1 S119
RPA34 T456
MAP1A S1801
MAP1A S1797
G3BP2 S149
G3BP2 S141
TOP2A S1106 c
MAP1A T2631
MAP1A S2629
NU153 T699
CTND1 S349
CTND1 S346
CTND1 S352
CTND1 S252
MAP4 S1073
SEPT9 T142 a
NUCL T92
d
F12
1F8
1C9
IDR
non-IDR
MOT4
JIP4
SYUA
DCXR
tSNE_1
tSNE_2
scaled expression
e
MOT4
SYUA
JIP4
DCXR
PD-L1
CD206
n_cells
median scaled expression
f
MOT4
JIP4
SYUA
DCXR
median scaled expression

Figure 5

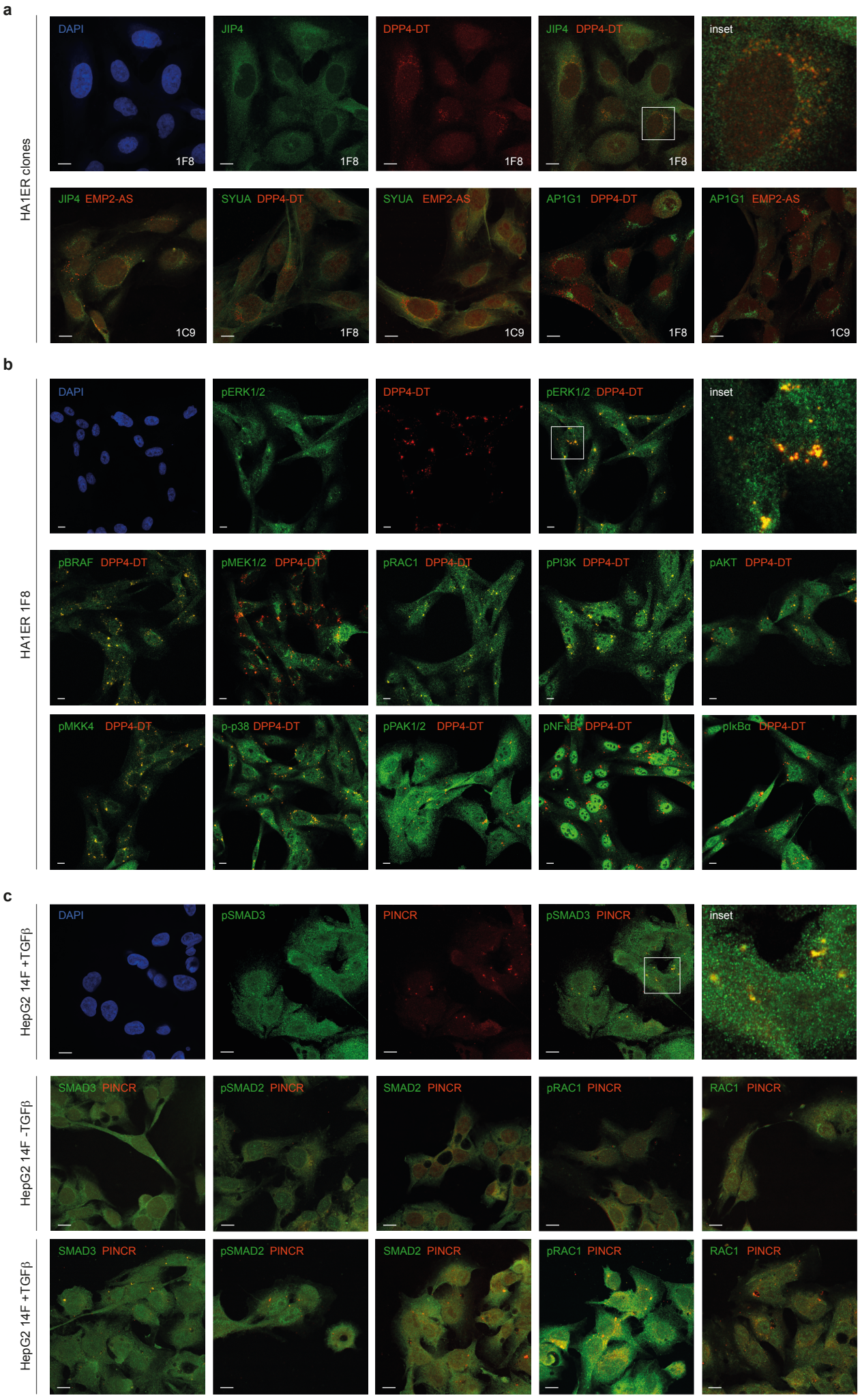

a
HA1ER clones
DAPI
JIP4
DPP4-DT
JIP4 DPP4-DT
inset
1F8
JIP4 EMP2-AS
SYUA DPP4-DT
SYUA EMP2-AS
AP1G1 DPP4-DT
AP1G1 EMP2-AS
1C9
b
HA1ER 1F8
pERK1/2
pERK1/2 DPP4-DT
pBRAF DPP4-DT
pMEK1/2 DPP4-DT
pRAC1 DPP4-DT
pPI3K DPP4-DT
pAKT DPP4-DT
pMKK4 DPP4-DT
p-p38 DPP4-DT
pPAK1/2 DPP4-DT
pNFκB DPP4-DT
pIκBα DPP4-DT
c
HepG2 14F +TGFβ
pSMAD3
PINCR
pSMAD3 PINCR
HepG2 14F -TGFβ
SMAD3 PINCR
pSMAD2 PINCR
SMAD2 PINCR
pRAC1 PINCR
RAC1 PINCR

Figure 6

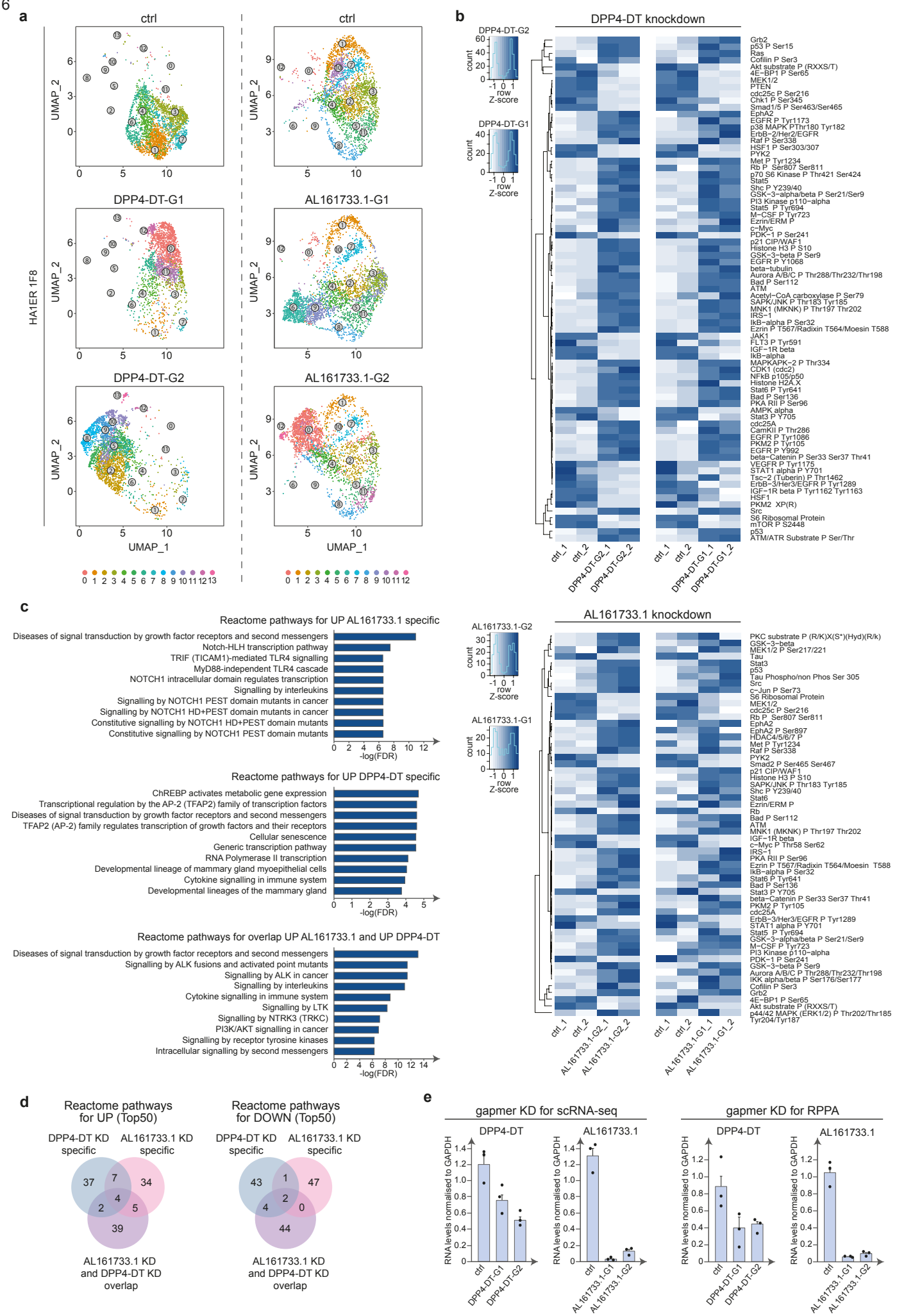
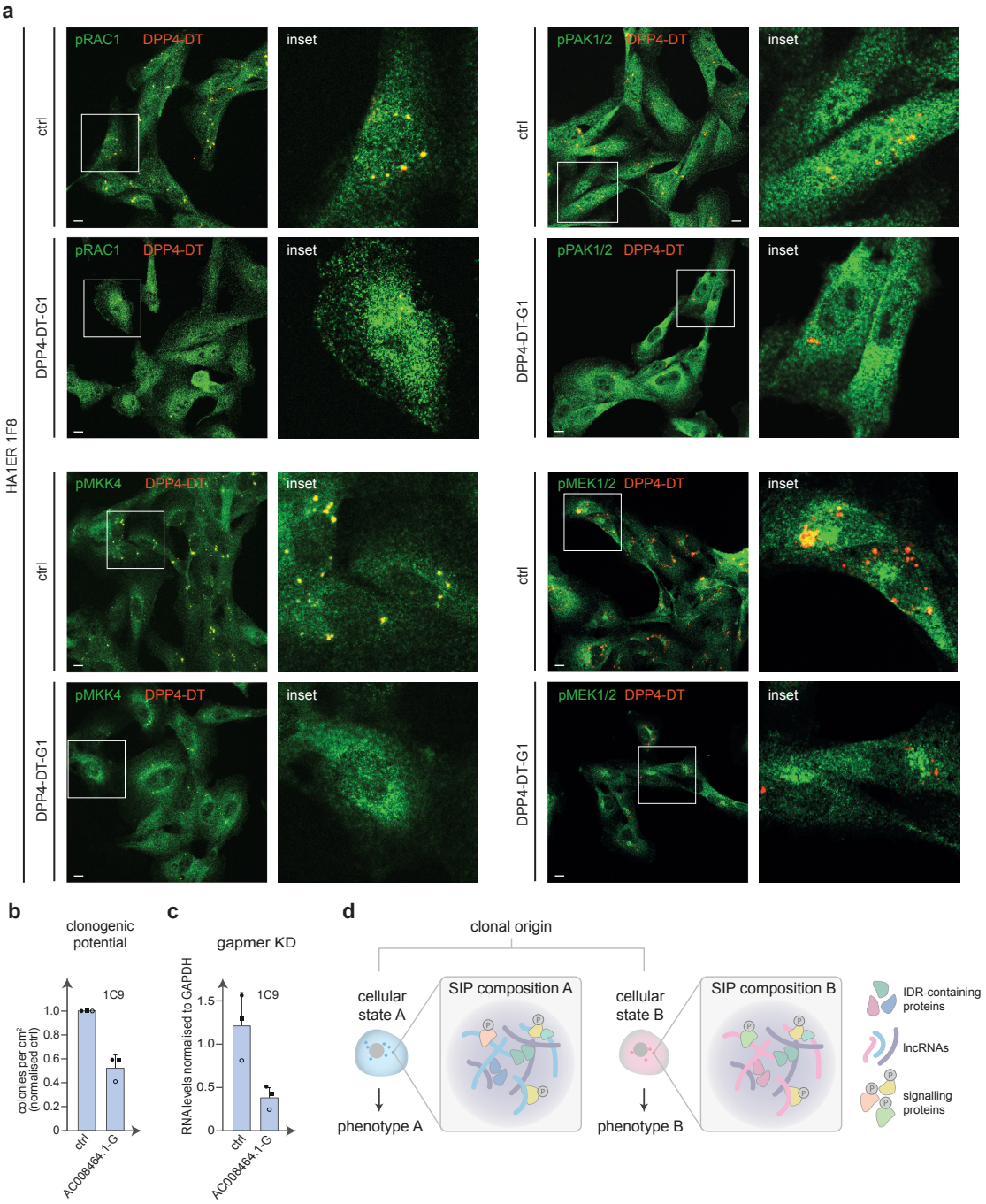

a
ctrl
ctrl
DPP4-DT-G1
AL161733.1-G1
DPP4-DT-G2
AL161733.1-G2
HA1ER 1F8
UMAP_2
UMAP_1
b
DPP4-DT knockdown
DPP4-DT-G2
DPP4-DT-G1
count
row Z-score
ctrl_1
ctrl_2
DPP4-DT-G2_1
DPP4-DT-G2_2
DPP4-DT-G1_1
DPP4-DT-G1_2
c
Reactome pathways for UP AL161733.1 specific
Diseases of signal transduction by growth factor receptors and second messengers
Notch-HLH transcription pathway
TRIF (TICAM1)-mediated TLR4 signalling
MyD88-independent TLR4 cascade
NOTCH1 intracellular domain regulates transcription
Signalling by interleukins
Signalling by NOTCH1 PEST domain mutants in cancer
Signalling by NOTCH1 HD+PEST domain mutants in cancer
Constitutive signalling by NOTCH1 HD+PEST domain mutants
Constitutive signalling by NOTCH1 PEST domain mutants
-log(FDR)
Reactome pathways for UP DPP4-DT specific
ChREBP activates metabolic gene expression
Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors
Diseases of signal transduction by growth factor receptors and second messengers
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors
Cellular senescence
Generic transcription pathway
RNA Polymerase II transcription
Developmental lineage of mammary gland myoepithelial cells
Cytokine signalling in immune system
Developmental lineages of the mammary gland
Reactome pathways for overlap UP AL161733.1 and UP DPP4-DT
Diseases of signal transduction by growth factor receptors and second messengers
Signalling by ALK fusions and activated point mutants
Signalling by ALK in cancer
Signalling by interleukins
Cytokine signalling in immune system
Signalling by LTK
Signalling by NTRK3 (TRKC)
PI3K/AKT signalling in cancer
Signalling by receptor tyrosine kinases
Intracellular signalling by second messengers
AL161733.1 knockdown
AL161733.1-G2
AL161733.1-G1
ctrl_1
ctrl_2
AL161733.1-G2_1
AL161733.1-G2_2
AL161733.1-G1_1
AL161733.1-G1_2
d
Reactome pathways for UP (Top50)
DPP4-DT KD specific
AL161733.1 KD specific
AL161733.1 KD and DPP4-DT KD overlap
Reactome pathways for DOWN (Top50)
e
gapmer KD for scRNA-seq
gapmer KD for RPPA
DPP4-DT
AL161733.1
RNA levels normalised to GAPDH
ctrl
DPP4-DT-G1
DPP4-DT-G2
AL161733.1-G1
AL161733.1-G2

Figure 7


a
HA1ER 1F8
ctrl
DPP4-DT-G1
pRAC1 DPP4-DT
pPAK1/2 DPP4-DT
pMKK4 DPP4-DT
pMEK1/2 DPP4-DT
inset
b
clonogenic potential
1C9
colonies per cm² (normalised ctrl)
ctrl
AC008464.1-G
c
gapmer KD
RNA levels normalised to GAPDH
d
clonal origin
cellular state A
phenotype A
SIP composition A
cellular state B
phenotype B
SIP composition B
IDR-containing proteins
lncRNAs
signalling proteins