Cytoplasmic lncRNAs nucleate signalling pathways to define metastable state dynamics and determine phenotypic output

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

*- 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 concept is widely acknowledged, the molecular mechanisms underlying this phenomenon remain largely undiscovered. Herein, we describe that clonal cell populations inherently display multiple co-existing metastable gene expression states that co-segregate with various phenotypic outputs. We show that transcriptome states are readily inherited and dynamically interconvert; not due to a shift in transcriptional programmes, but rather through yet unidentified post-transcriptional regulatory 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 containing intrinsically disordered regions and diverse active signalling pathways converge. Herein, we propose the SIP as a nucleating reservoir of non-genetic information at the crossroads of cell plasticity and non-genetic heterogeneity where it integrates intra- and extracellular signalling pathways buffering phenotypic output.

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

Introduction

The wide field of genetics has revolutionised the manner in which we approach biological problems and, even though its remarkable relevance cannot be contested, hundreds of observations over the years suggest that the non-genetic compartment is of equal importance to define life as we know it and, therefore, should not be neglected[1-3]. Along those lines, following the revolution in next generation sequencing (NGS) and its application to single cell analysis, we can finally begin to query the non-genetic compartment with unprecedent detail in a robust and quantitative manner. It is in that context that cell plasticity has emerged as a central player in various and diverse fields of study ranging from microbiology to cancer research[4-7]. Notably, within the plethora of non-genetic information carriers present in cells, the study of the transcriptome both in bulk and at the single-cell level revolutionised basic and biomedical research. In particular, the development of single-cell RNA-Sequencing (scRNA-Seq) unravelled that even genetically homogeneous cell populations display heterogeneous transcriptome profiles[8]. Strikingly, although transcriptome heterogeneity is now considered to be prevalent in most biological settings, our understanding of how transcriptome states are established, maintained and inherited, and the molecular mechanisms underlying their potential to reshape into novel transcriptome arrangements, remains fragmentary. In that regard, our reported data shows that non-genetic cell heterogeneity in terms of transcriptomes is not randomly established but follows precise phylo-(epi-)genetic linked inheritance patterns, generating traceable cell-ancestry trees within isogenic and genetically stable experimental models (Barcode Decay Lineage tracing, BdLT-Seq[8]). Moreover, intrapopulation lineage-linked transcriptome states are perpetuated for several generations fuelling population heterogeneity and thus suggesting the existence of a “molecular memory” retained through clonal propagation/divergence.

Herein, we show that by subcloning a genetically stable clonal founder cell population we isolate ancestry trees that are enriched in specific subsets of transcriptome states which correlate with the fate of individual cells in response to intracellular and extracellular cues. Moreover, we provide primer evidence suggesting that for a given cellular population, the degree of phenotypic plasticity is anchored to the transcriptome state of each individual cell and its associated lineage-linked restricted plasticity. Finally, we describe the existence of a cytoplasmic compartment (Signal Integration Portal - SIP) that nucleates long non-coding RNAs, proteins containing intrinsically disordered regions (IDR) and active components of multiple signalling pathways that co-segregate with subsets of transcriptome states and whose variable composition correlates with phenotypic output and transcriptome-linked non-genetic heterogeneity.

Collectively, our data supports the notion that the perpetuation of non-genetically heterogeneous cell populations in physiological and pathological conditions, as well as in response to biological and/or therapeutic cues, is not only dictated by the overall cellular (i.e. epithelial, mesenchymal, tissue of origin) and molecular (i.e. genetic landscape, bulk transcriptome profiles) characteristics of the population but also strongly relies on the non-genetic molecular identity (SIP complexity and composition) and plasticity of its intrapopulation phylo-epigenetic 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 could 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 explore the molecular mechanism underlying gene expression state transitions and its link with phenotypic output, we first evaluated a clonal founder population of HA1ER cells (F12 - HEK cells immortalized with hTERT/SV40 and transformed by $HRAS^{G12V}$ [9]) and two stable subclones derived from it (1F8 and 1C9) previously reported as diverging at their transcriptome level. Strikingly, 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 Figure 1), scRNA-Seq (Chromium 10x) revealed that, after data processing to mitigate clustering artifacts due to cell cycle progression and abnormal mitochondrial counts (proxy for cell death. Seuratv4.0), gene expression states dramatically differed between the founder population and its subclones (subclones 1C9 and 1F8; Figure 1A, 2A and Supplementary Figure 2, 3A). Interestingly, those differences are further propagated down the lineage when the first subclones obtained from the founder cell are further subcloned (Subclones 1F8, 3C5 and 1E8; Supplementary Figure 4, left. Subclones 2A6, 2B12 and 1F5; Supplementary Figure 4, right). Remarkably, whole exome sequencing (WES) for clones F12, 1F8 and 1C9 did not reveal any difference (Supplementary Figure 5A) thus suggesting that the observed gene expression state divergence is encoded and propagated by non-genetic networks.

To further verify this possibility, we performed BdLT-Seq to determine if transcriptome plasticity was differential between the founder clone F12 and subclones 1F8 and 1C9 and revealed that indeed, the founder clone and its derived subclones display not only a divergent state composition but also differential basal transcriptome plasticity (Figure 1B, C and Supplementary Figure 2B). For example, after four days, the progeny of cells in cluster 2 of clone F12 could be traced to every cluster except clusters 3, 7 and 12, with the majority transitioning to clusters 0 and 10. On the other hand, cells from cluster 2 of subclone 1F8 mainly diverged to clusters 3, 7 and 11, whereas cells from cluster 2 of subclone 1C9 could be traced primarily to clusters 0 and 1 but not to clusters 3, 10 and 11 (Figure 1C, Supplementary Figure 2B). Prompted by these results, we hypothesised that any isogenic and genetically stable cellular system amenable to cloning could display similar transcriptome behaviour when analysed at the single-cell level and concordantly could be exploited to shed light onto the molecular mechanism underlying the establishment/genesis of metastable transcriptome state dynamics. To pursue this hypothesis, we built three different clonal cellular systems (HA1ER, described above; HA1E, HEK cells immortalized with hTERT/SV40; HepG2, Hepatocellular carcinoma cells) by generating in the first step a clonal founder population and then subcloning each founder population to obtain transcriptome state diverging subclones whilst maintaining isogeneity (Figure 1D) and enabling lineage-linked comparisons.

After an initial molecular and phenotypic assessment of each clone (WES, cell proliferation, ploidy, cell cycle profile and morphology. Supplementary Figures 1 and 5), the transcriptome states of the founder clone and the four subclones from each system were analysed by scRNA-Seq (Chromium 10X, Figure 2A-C). Strikingly, scRNA-Seq data analysis determined that for the HA1ER system, 3 of the 4 subclones obtained from the founder clone (F12) display massive transcriptome state re-arrangements when contrasted to the founder population (1D12, 2A6 and 1F8) whilst the remaining one (1C9) exhibited

similar transcriptome states as the founder albeit displaying altered state representation (Figure 2A and Supplementary Figure 3A). Similarly, when populations from the HA1E lineage were analysed, 3 out of the 4 subclones (A7A, A9H, B2F) also displayed massive transcriptome re-arrangements when compared to the founder HA1E clone whereas subclone B5F was the most comparable to the parental founder clone 10E (Figure 2B, Supplementary Figure 3B).

To further explore transcriptome plasticity and inheritance in dynamic processes, we took advantage of the capability of HepG2 cells to activate a TGFβ-dependent transcriptional programme driving the epithelial to mesenchymal transition (EMT). Briefly, we treated the founder HepG2 population (1A) and the four generated subclones (32G, 14F, 48E and 44D) with TGFβ to induce EMT (Supplementary Figure 6A) and evaluated the representation of gene expression states by scRNA-Seq in vehicle-treated (-TGFβ) or 4-days TGFβ-treated populations. As observed in HA1E and HA1ER cells, when naïve HepG2 cells are assessed, dramatic differences between subclones can be observed when compared to the founder clone (Figure 2C and Supplementary Figures 3C and 7). For instance, whereas in the naïve founder clone (1A) virtually no cells can be detected belonging to cluster 7, subclone 14F and 48E display a pronounced increase in the representation in the same cluster (Figure 2C and Supplementary Figure 3C). Notably, upon EMT induction, a massive gene expression shift is observed when the founder and the subclones are subjected to TGFβ treatment. Strikingly, the initial differences observed in the naïve subclones propagate into the states identified upon TGFβ treatment. As example, the subclones displaying cells in cluster 7 in naïve conditions (14F and 48E) display a reduced representation of cells occupying cluster 6 following EMT compared to the other clones (Figure 2C).
To take this further and given that the HepG2-based TGFβ-induced EMT model system also enables to the exploration of the mesenchymal-to-epithelial transition (MET) just by the subsequent removal of TGFβ from the cultures, we took advantage of this possibility and explored transcriptomic state re-arrangement upon EMT reversion (Supplementary Figure 7). We observed that whilst the founder clone 1A and its most similar subclone at the transcriptome level, 44D, revert to occupy the same initial transcriptome states as the naïve population, however, their state representation is significantly altered with cluster 6 becoming the most predominant (Supplementary Figure 7). Notably, clone 14F which was the most dissimilar from the founder 1A, takes a completely different path upon reverting through MET, as cluster 1 and 8 are now overpopulated with only a minority of cells occupying other states (Supplementary Figure 7). Together, this data supports the notion that the initial cell states, even when a transcription factor mediated transcriptional gene expression programme is enforced by external cues, are largely inherited and dictates the phenotypic fate of a clonal population.

Transcriptome divergence encompasses phenotypic output in a constrained manner

Prompted by these observations and our previously reported data[8], we set out to evaluate several phenotypic outputs relevant to each particular cellular system analysed. In the first instance, we took advantage of the well characterized HA1ER cell system that bears the capability to grow in 3D soft agar matrices, where its growth in anchorage-independent conditions is a proxy for cellular transformation. Interestingly, although no major differences can be observed in the proliferation of the analysed clones (1D12, 1C9, 2A6 and 1F8) and its founder (F12) in 2D growth (Supplementary Figure 1A), challenging these clonal populations to grow in 3D anchorage-independent conditions results in a divergent phenotypic output as clone 1C9, based on observed transcriptomic states the most similar to the founder F12 population, grows more efficiently in these conditions (Figure 2D). Notably, at the transcriptome level the most dissimilar HA1ER clones analysed, 2A6 and 1F8, have almost lost the capability to grow in 3D settings albeit being genetically identical to the founder cellular population (Figure 2D, Supplementary Figure 6C). Strikingly, the growth of clone 1D12 is similar to the parental F12 founder population in 3D conditions despite being the most similar at the transcriptome level to clone 2A6 which fails to grow efficiently under the same conditions (Figure 2A, Supplementary Figure 6C), thus suggesting that transcriptome organization plays a role in modulating phenotypic output but

is not the sole determinant. Interestingly, when the second round of cloning from the F12 lineage was analysed, the subclones derived from 1F8 and 2A6 (1E8 and 3C5, and 1F5 and 2B12, respectively) did follow the phenotype displayed by their respective founder clone and failed to grow on 3D (Supplementary Figure 6C), thus indicating that phenotypic divergence is propagated across several generations.

To delve further into our findings, we took advantage of the fact that immortalized HA1E cells (hTERT + SV40), are susceptible to become transformed by overexpressing a RAS oncogene[9,10]. To evaluate whether transcriptomic divergence can result in differences in transformation potential, we prepared a multiplexed lentiviral library containing 26 RAS variants from the three main families H-RAS, N-RAS and K-RAS (wild type and 7-8 mutants for each family; Supplementary Figure 6F), transduced each of the HA1E clones (A7A, A9H, B2F and B5F) and their founder (10E) and then subjected each population to anchorage-independent growth conditions to evaluate their susceptibility to malignant transformation (Figure 2E). Notably, clones A9H and B5F grow to a similar extent in 3D cultures with the clone B5F being the most similar at the transcriptome level to the founder clone 10E from all the analysed clones (Figure 2B, E). However, two subclones that dramatically diverge at the level of gene expression states from the founder clone, A7A and B2F, either grow only to half the extent of the founder or grow two times more efficiently (Figure 2B, E) thus suggesting that the transformation capacity of a given cell could be somehow modulated by intrinsic molecular mechanisms hidden within transcriptome states. Concordantly, when HepG2 clonal populations were analysed for the capability to grow in soft agar in naïve conditions or upon TGFβ treatment, once again a divergence in their individual capabilities could be observed as all the clones analysed grew less efficiently than the founder clone in naïve conditions (Supplementary Figure 6D), yet displaying variability among them. Interestingly, whereas most clones almost lost the ability to form colonies in soft agar upon TGFβ treatment (1A, 32G, 44D), the clonogenicity of subclones 14F and 48E, which both display a divergent transcriptome profile compared to the other three clones, was significantly increased (Supplementary Figure 6E).

Interestingly, when we subjected clonal populations of HA1ER and HepG2 and their corresponding subclones to a transwell migration assay, once again we observed profound phenotypic variability that does not seem to segregate with the observed growth in 3D but is albeit different between the subclones analysed and their respective founder clone (Figure 2G, H and Supplementary Figure 6B). Moreover, as HA1ER cells are susceptible to trigger apoptosis upon treatment with TRAIL (Tumor necrosis factor-related apoptosis-inducing ligand), we also tested our clonal models for their capability to withstand TRAIL challenge. Notably, although TRAIL treatment induces fractional killing in all clones, once again, a phenotypic divergence among all subclones (1F8, 1C9, 2A6 and 1D12) and the founder clone (F12) could be observed in the extent to which apoptosis is triggered (Figure 2F).

Gene expression states do not rely on active transcription but rather are post-transcriptionally established and maintained

Our data supports that gene expression states are highly dynamic and display several layers of organization shaping phenotypic output. These observations prompted us to investigate which are the molecular mechanisms contributing to transcriptome state dynamics, maintenance and divergence.
To address these, we set out to investigate whether changes in active transcription were driving transcriptome divergence by initially analysing clones from the HA1ER systems, in particular the transcriptome state divergent clones 1F8 and 1C9 from the F12 lineage. To do so, we performed SLAM-Seq[11,12] to determine the rate of active transcription compared to steady state RNA levels by metabolically labelling nascent RNAs with 4-thiouridine (4-SU pulse for 15 minutes), and then converting the modified/unmodified RNA into strand-specific RNA sequencing libraries. After quality control and determination of T-to-C conversion rate in the sequencing libraries (Supplementary Figure

8A), data was mapped to hg38 using a T>C aware aligner (NextGenMap[13]) and further processed through the SLAM-DUNK pipeline[14]. Next, to determine the levels of nascent transcripts and their corresponding steady state RNAs and to perform a pairwise comparison between both clones 1F8 and 1C9 in a quantitative manner, we applied the paired Beta binomial distribution to establish significant differences at the nascent RNA level. Strikingly, we did not detect any significant difference at the nascent RNA level (FDR<0.05, FC>2) whereas a significant difference was observed at the steady state level with 217 transcripts differentially expressed between both clones (FDR<0.01, FC>2; Figure 3A), thus suggesting that transcriptomic state dynamics are not under the control of active transcriptional processes in the compared cellular systems.

To further validate our striking observation, we performed CUT&RUN experiments[15] to assess informative histone modifications that correlate with transcriptional activity and chromatin organization such as H3K27Ac (active enhancers), H3K27me3 (facultative heterochromatin) and CTCF (overall chromatin organization). Surprisingly, we did not detect major alterations in chromatin architecture between clones 1F8 and 1C9 at any of the levels analysed (Supplementary Figure 9), thus supporting our hypothesis that indeed transcriptome state dynamics are controlled at the post-transcriptional level. Prompted by those results, we wondered whether our findings were a general phenomenon or if it was an isolated case only valid for the 1F8 vs 1C9 comparison. Thus, we performed SLAM-Seq analysis in the three clonal systems that we have previously shown here, HA1ER, HA1E and HepG2 and compared all subclones with their corresponding founder for each system (Supplementary Figures 8B and 10). Strikingly, for most of the comparisons, no significant differences in nascent transcription were observed (HA1ER: F12/1C9, HA1E: 10E/A7A, 10E/B5F, HepG2: 1A/14F, 1A/48E; Supplementary Figure 10) whereas when transcriptional differences were detected (HA1ER: F12/1F8, F12/2A6, F12/1D12, HA1E: 10E/A9H, 10E/B2F, HepG2: 1A/44D, 1A/32G), they only had a slight-to-none impact in steady state RNA levels and quite often the observed changes in steady state RNA (up or down regulated) did not follow the transcriptional trend (Supplementary Figure 10B-D). These data, obtained from multiple clonal cellular systems, suggest that steady state RNA levels and hence metastable gene expression states are not established/maintained by diverging active transcription programmes but rather by a molecular machinery that operates at the post-transcriptional level. Furthermore, our data supports the hypothesis that this phenomenon is a general biological feature rather than an oddity.

Neither protein content nor protein phosphorylation are differentially represented between diverging clones

The fact that transcriptomes diverge throughout lineages and that the observed divergence is accompanied by profound changes in phenotypic output prompts the question whether major contributors to phenotype such as protein composition (e.g. translation) and/or signalling pathway status (e.g. ERK1/2 phosphorylation) also diverge. To test these hypotheses, we analysed total protein content for 1F8 and 1C9 clones by label free mass spectrometry in-bulk and performed a pairwise comparison between both phenotype-diverging subclones. Notably, this data showed that only a small fraction of proteins was differentially expressed (FDR<0.01, FC>1.5, Figure 4A) with 23 upregulated and 34 downregulated proteins, suggesting that, even though diverging clonal populations differ in their transcriptomes, they do not display a concordant difference in translation. Regardless of the meagre differences observed at this level, gene ontology analysis did not reveal enrichment in any biological pathways (Reactome v3.7, Supplementary Figure 11A) that could guide further hypotheses to explain the observed phenotypic divergence.
These seemingly negative data prompted us to explore whether the diverging phenotypes could be explained by an imbalance in signalling pathway components or signalling capability and therefore we performed a phospho-proteomic analysis (TMT) in the 1F8 and 1C9 populations (FDR<0.05, FC>0.4, Figure 4C). Surprisingly, data suggested a minimal change in the phospho-proteome between both

clonal populations with 33 phospho-sites displaying enhanced phosphorylation and 24 displaying reduced phosphorylation levels across only 28 proteins at bulk levels (Figure 4C and Supplementary Table). Once again, the data obtained failed to explain such an important divergence in phenotypic output and therefore prompted us to look deeper into the data.

Diverging lncRNAs localise and converge into perinuclear structures

To advance our understanding of the underlying molecular mechanisms of phenotypic divergence in clonal cell populations, we endeavoured into annotating our various datasets including SLAM-Seq data, label-free mass spectrometry and phospho-proteomic data on their molecular features. Intriguingly, a large fraction of the differentially expressed transcripts found by comparing 1F8 and 1C9 SLAM-Seq data are indeed lncRNAs with approximately 30% of them corresponding to this category (Figure 3B,C). This observation prompted us to evaluate the lncRNA content in our single-cell datasets where we validated our initial observations even across all three clonal systems where we observed the lncRNA content ranging from ~30-64% between various conditions including clonal divergence and TGFβ-induced EMT (HA1ER, HA1E and HepG2. Figure 3D and Supplementary Figure 8C).

Given the generality of our observations across different clonal lineages, we combined data obtained from the SLAM-Seq assays with the scRNA-Seq datasets to build a composite lncRNA catalogue from every system analysed. To do so, a subset of diverging lncRNAs were first evaluated by reverse transcription followed by qPCR (RT-qPCR), then by subcellular fractionation followed by RT-qPCR and finally by RNA-FISH followed by super-resolution microscopy. Notably, from a composite subset of lncRNA candidates, we were able to validate ~45% of the differentially expressed lncRNAs that diverged between 1C9 and 1F8 by RT-qPCR (Figure 4E and Supplementary Figure 8D). Moreover, from the HA1E and the HepG2 clonal systems, we validated ~65% of the diverging lncRNAs by the same means (Supplementary Figure 8E, F). Strikingly, we found that most of the lncRNA transcripts tested and annotated as divergent display cytoplasmic localization in both HA1ER and HepG2 clones (Figure 3F and Supplementary Figure 8G). Prompted by these data, we moved forward to assess the intracellular localization of these cytoplasmic diverging transcripts by RNA-FISH (Zeiss Airyscan) and verified that clonally diverging lncRNA transcripts from HA1ER and HepG2 indeed localize to the cytoplasm and more precisely, into perinuclear structures with all the lncRNAs analysed displaying the same intracellular pattern (Figure 5 and Supplementary Figure 12), suggesting that a large fraction of diverging lncRNAs may somehow interact to modulate phenotypic output in a non-nuclear-linked manner. Moreover, given our data suggesting that most of the differentially expressed coding RNAs are not accompanied by increased/reduced translation of their corresponding transcripts, we hypothesised that most of those RNAs act as molecular scaffolds.

Intrinsically disordered regions-containing proteins (IDR-proteins) clonally diverge and localize with diverging peri-nuclear lncRNAs

Mass spectrometry data demonstrated that only a small subset of proteins was differentially expressed between 1F8 and 1C9 clones thus suggesting that phenotypic divergence may rely on a reduced number of features (e.g localization, scaffolding properties, disordered regions, etc). To explore this possibility, we annotated the differentially expressed proteins into structured proteins or IDR-proteins using AlphaFold2[16,17]. Strikingly, ~63% of the differentially expressed proteins between clones 1F8 and 1C9 displayed intrinsically disordered regions with a major subset showing highly unstructured domains (Figure 4D). To validate the observed divergent protein expression between clones 1F8 and 1C9, we set up a CyTOF analysis for a subset of differentially expressed proteins in F12 (IDR proteins JNK-interacting protein 4 (JIP4), α-Synuclein (SYUA), and monocarboxylate transporter 4 (MOT4), myelin expression factor 2 (MYEF2), heat shock protein beta-6 (HSPB6), sequestosome-1 (SQSTM) and non-IDR-protein L-xylulose reductase (DCXR)). Notably, all the analysed proteins were differentially

386 expressed and labelled subpopulations within the F12 population (Figure 4D and Supplementary
387 Figure 11B, D, E). Next, we assessed a subset of these proteins (JIP4, SYUA, MOT4, DCXR, PD-L1
388 (positive control), CD206 (negative control)) for their divergent expression in subclones 1F8 and 1C9.
389 Whilst none of them displayed cell cycle-dependent expression (Supplementary Figure 11F), strikingly,
390 clone 1F8 contains a major cellular subpopulation where candidate antibody signal (JIP4, SYUA, MOT4
391 and DCXR) is predominant and that is nearly absent in clone 1C9 (Figure 4D and Supplementary Figure
392 11 B, C). This validates our candidate markers as clonally diverging proteins and suggests that clonal
393 divergence can also be traced at the proteome level even when the differentially expressed proteins
394 are only a minimal fraction of the total proteome. Given that a large fraction of the clonally diverging
395 proteins belongs to the IDR-containing family and could potentially act as molecular scaffolds, we next
396 expanded our candidate list to accommodate an additional diverging IDR-protein gamma1-adaptin
397 (AP1G1) and explored the hypothesis of a potential link between the diverging IDR-proteins and the
398 diverging lncRNAs enriched in the cytoplasm. Therefore, we performed RNA-FISH coupled to ICC for a
399 subset of diverging lncRNAs and IDR-proteins and analysed their intracellular distribution by super-
400 resolution microscopy (Zeiss Airyscan). Interestingly, we show that in clonal lineages of HA1ER
401 diverging lncRNAs and IDR-proteins co-localize in the perinuclear region (lncRNAs: DPP4-DT (1F8) and
402 EMP2-AS (1C9), IDR-proteins: JIP4, SYUA: Figure 5A and Supplementary Figure 12A), and only for IDR-
403 protein gamma1-adaptin could no such co-localisation be detected with either lncRNA. Moreover, we
404 also validated the co-localisation of IDR-protein JIP4 and several diverging lncRNAs (PINCR and
405 LINC00494 (14F), SCAT8 and LINC2476-203 (44D)) in the perinuclear region in the HepG2 clonal model
406 (Supplementary Figure 12B), compatible with our previous observations in the HA1ER system, thus
407 suggesting a functional link between both compartments.
408
409
410 **The Signal Integration Portal (SIP)**
411
412 We have already verified that a subset of diverging lncRNAs and a subset of IDR-containing proteins
413 co-localise into perinuclear cytoplasmic structures in a variety of clonally diverging cellular models.
414 However, none of our observations indicated a functional/effector link that could explain the observed
415 divergence in phenotypic output for a given complex cellular behaviour such as the capability to grow
416 in anchorage-independent conditions, migration or fractional killing. Therefore, we hypothesised that
417 cellular plasticity in phenotypic output relies on the re-arrangement of existing signalling pathways
418 following different configurations. To further explore this concept, we contemplated the possibility
419 that the peri-nuclear structures that we have identified could nucleate signalling components that
420 would alter phenotypic output. Given that we have merely detected 28 statistically significant
421 differentially phosphorylated proteins between our flagship clones 1F8 and 1C9, and for which
422 phospho-site specific antibodies are not available, we decided to test whether multiple activated
423 signalling pathways would converge within the observed peri-nuclear structures. To do so, we built a
424 panel of phospho-site specific antibodies directed towards multiple major signalling pathways
425 including RAF/MEK/ERK, PI3K/AKT, IκBα/NFκB and MKK4/p38, and performed RNA-FISH (targeting
426 peri-nuclear localized diverging lncRNA DPP4-DT) coupled to ICC (phospho-site specific;
427 Supplementary Table) in clone 1F8. Strikingly, after super-resolution imaging analysis, we observed
428 that multiple active signalling pathways, in the form of phospho-proteins, indeed co-localise and
429 converge into the peri-nuclear structures delineated by the lncRNA DPP4-DT (pARAF, pBRAF, pMEK1/2,
430 pERK1/2, pRAC1, pMKK4, p-p38, pPI3K, pAKT, pPAK1/2, pIκBα and pNFκB. Figure 5B and
431 Supplementary Figure 12 C). Intriguingly, rather unrelated active signalling pathways coalesce into
432 these cytoplasmic particles suggesting that the observed structures could function as a signal organizer
433 to determine phenotypic output and thus we termed this compartment the Signal Integration Portal
434 (SIP). To further verify the existence of the SIP and taking advantage of the capability of the HepG2
435 model to undergo TGFβ-induced EMT, we decided to evaluate if members of the TGFβ signalling
436 pathway would converge into SIP structures delineated by HepG2-specific diverging lncRNA PINCR in

437 clone 14F by RNA-FISH coupled to ICC. To do so, we included members of the canonical TGFβ pathway
438 (SMAD2 and SMAD3) in addition to members of the non-canonical pathway also reported to become
439 active by the presence of TGFβ signalling (RAC1 and ROCK1) and evaluated their total levels and their
440 corresponding phosphorylated active forms. As expected, PINCR localised to the peri-nuclear region
441 in a punctuated pattern across most of the cells in TGFβ-naïve conditions (Figure 5C and
442 Supplementary Figure 12D). Notably, most of the components of the TGFβ signalling pathway and their
443 phosphorylated counterparts analysed displayed a diffuse cytoplasmic localization in naïve conditions
444 (Figure 5C and Supplementary Figure 12D). However, upon TGFβ treatment, PINCR (proxy for SIP
445 particles) co-localised with active members of the TGFβ signalling cascade including pSMAD2, pSMAD3
446 and pRAC1 (Figure 5C and Supplementary Figure 12D). Interestingly, whilst ROCK1 was detected in SIP
447 structures, we did not detect pROCK1 (Tyr913) suggesting that either this particular phospho-protein
448 is not recruited to the SIP or is masked/absent in HepG2 cells (Supplementary Figure 12D).
449
450 Altogether, we showed that in phenotype-diverging clonal populations of fully committed and
451 differentiated cellular models (HA1ER and HepG2) lncRNAs, IDR-containing proteins and active
452 signalling effectors from multiple signalling cascades coalesce into a peri-nuclear structure which we
453 termed Signal Integration Portal (SIP) that sits at the crossroads of phenotypic divergence nucleating
454 inputs from the nucleus (lncRNA), the cytoplasm (IDR-proteins and signalling cascades) and the
455 environment (signalling cascades).
456
457 **Remodelling cell states by altering the SIP lncRNA composition**
458
459 We have already shown that diverging SIP components coalesce into peri-nuclear structures that
460 appear to contribute to the establishment/propagation of transcriptome states. Therefore, to further
461 our claims, we hypothesised that altering the SIP composition may result in transcriptome re-
462 arrangements. To do so, we designed antisense oligonucleotides (G, gapmer technology) targeting a
463 subset of lncRNAs identified in the diverging HA1ER clones 1F8/1C9 (DPP4-DT, AL161733.1 and
464 AC008464.1) and in HepG2 clones 14F and 44D (PINCR and SCAT8). After verifying successful knock-
465 down of the evaluated lncRNAs from either system by RT-qPCR (Supplementary Figure 13D), cells from
466 each experimental condition were processed for scRNA-Seq (Chromium10x). After sequencing,
467 obtained data was pre-processed to mitigate clustering artifacts due to cell cycle progression and
468 abnormal mitochondrial counts and clustered using Seurat. Strikingly, we observed transcriptome re-
469 arrangements that range from the removal of a transcriptome state whilst largely preserving
470 transcriptome heterogeneity (PINCR-G: cluster 8; HepG2. Supplementary Figure 13B, H), the reduced
471 representation of certain states and the emergence of novel ones whilst preserving a certain degree
472 of the original transcriptome states (AL161733.1: cluster 10 reduced, clusters 6 and 9 novel for G1,
473 clusters 0, 11 and 12 novel for G2; HA1ER. SCAT8: cluster 8 reduced, cluster 10 novel for G1, cluster 5
474 novel/increased for G2; HepG2. Fig. 6A and Supplementary Figure 13G, H) to a complete
475 reprogramming of the metastable state landscape (AC008464.1-G and DPP4-DT-G1 and -G2; HA1ER.
476 Fig. 6A and Supplementary Figure 13A, G) thus suggesting that SIP-bound lncRNAs levels harbour the
477 capability to remodel the transcriptome landscape and hence modulate phenotypic output.
478
479 **Cell state remodelling is accompanied by profound changes in signalling**
480
481 To further explore the phenotypic consequences of SIP disruption and transcriptome remodelling, we
482 evaluated whether major signalling pathways would be altered upon a given lncRNA knockdown (2
483 gapmers targeting different regions per target) in the HA1ER model. To do so, we took advantage of a
484 pathway-specific antibody panel (180 antibodies; Supplementary Table) and evaluated signalling
485 pathway remodelling by Reverse Phase Protein Array (RPPA[18]) in the 1F8 diverging clone (increased
486 expression of lncRNAs AL161733.1 and DPP4-DT). Upon efficient lncRNA knockdown (~50-94% KD
487 when compared to the non-targeting control, Supplementary Figure 13E), we validated transcriptome

488 remodelling by evaluating a subset of expression markers by RT-qPCR (Supplementary Figure 13C, F)
489 and then proceeded to run the RPPA assay. Strikingly, upon lncRNA AL161733.1 knockdown, from the
490 180 antibodies tested we found 50 upregulated antibody signals whilst 23 were downregulated with
491 both antisense gapmers (Figure 6B bottom, ≥20% increase/decrease to control) when contrasted to
492 the non-targeting gapmer. Moreover, when knock-down of lncRNA DPP4-DT was evaluated, 55 signals
493 were upregulated whilst 39 were downregulated with both gapmers (Figure 6B top, ≥20%
494 increase/decrease to control). Notably, as we hypothesise that the lncRNAs evaluated are indeed
495 involved in integrating signalling information by nucleating pathways, we annotated each of the
496 assessed antibodies with their respective pathways (Reactome, Top50 Terms, FDR<0.01). For this, we
497 extracted the upregulated and downregulated antibody signals with specific changes for lncRNA
498 AL161733.1 and lncRNA DPP4-DT knockdowns, as well as the list with overlapping changes upon
499 knock-down of lncRNA AL161733.1 and lncRNA DPP4-DT (Figure 6C, D). Outstandingly, separate knock-
500 down of both lncRNAs lead to a modulation of common pathways (overlap list), such as upregulation
501 of signalling by ALK in cancer, PI3K/AKT signalling, signalling by LTK and signalling by RTKs (Figure 6C,
502 Supplementary Figure 13I) suggesting that both lncRNAs contribute to a general mechanism of
503 signalling pathway control. However, knocking down lncRNA AL161733.1 resulted in the modulation
504 of specific pathways such as upregulation of cytokine signalling in the immune system, TRIF-mediated
505 TLR4 signalling and the Notch-HLH transcription pathway (Figure 6C, Supplementary Figure 13I)
506 whereas knocking down lncRNA DPP4-DT resulted in the modulation of a complete divergent subset
507 of pathways, including the upregulation of multiple transcription-regulating pathways (Figure 6C,
508 Supplementary Figure 13I), thus suggesting the existence of an exquisite mechanism of signalling
509 pathway control where individual lncRNAs modulate a defined set of pathways in a specific manner.
510 Finally, the observation that major – unrelated – signalling pathways were indeed altered upon lncRNA
511 knockdown suggests that by solely adjusting SIP composition by reducing the levels of SIP-linked
512 lncRNAs, both transcriptome re-arrangement and, as hypothesised, major remodelling of post-
513 translational signalling pathways occurs.
514
515 Taken altogether, we propose the Signal Integration Portal (SIP) as a peri-nuclear molecular entity
516 where biological information from various non-genetic compartments including cytoplasmic lncRNA
517 content, intrinsically disordered region containing proteins and active signalling pathways converge
518 (Figure 6E) to modulate phenotypic output thus controlling cell plasticity and the degree of non-
519 genetic heterogeneity for a given cellular system.
520
521
522 **Discussion**
523
524 Cell plasticity, an inherent characteristic of all known organisms, is one of the most widespread
525 phenomena observed in biology. However, although a myriad of hypotheses have been put forward,
526 very little is known about the mechanisms underlying this phenomenon. Paradoxically, years of
527 research that centred on the genetic compartment could not explain the plastic behaviour displayed
528 by cells and organisms overall whilst a multitude of examples challenging the role of the genome as a
529 master controller underlying this phenomenon have come to light including bet-hedging[19,20],
530 protozoan development/evolution[21-24], all the way to the emergence of resistance to therapy in
531 cancer[25-29]. Remarkably, the advent of multimodal single-cell technologies opened a new dimension of
532 biological insight that fuelled the in-depth exploration of the non-genetic compartment and has
533 enabled the study of the molecular determinants underlying cell plasticity[8,30-35]. In that regard, we and
534 others ventured into exploring transcriptome content as a determinant of cell plasticity in mammalian
535 cellular systems and found that indeed steady-state transcriptome heterogeneity harbours sufficient
536 biological information to drive phenotypic output and to diverge even in fully controlled clonal cellular
537 systems, thus fuelling molecular and functional heterogeneity within populations[8,22,36]. Strikingly, in all
538 the models that we have tested so far, the contribution of active transcription to transcriptome

heterogeneity is negligible and does not support a model where transcriptome divergence is prompted by changes in gene expression programmes (as defined by alterations in chromatin structure resulting in different subsets of genes being actively transcribed). Instead, transcriptome plasticity seems to rely on post-transcriptional regulatory modules wherein a subset of cytoplasmic lncRNAs and intrinsically disordered region-containing proteins (IDR-proteins) play a key role throughout phenotypic divergence and are enriched in subsets of cellular states. Both observations challenge the status quo as, to begin with, lncRNA function has been ascribed mainly to the nucleus where they, in principle, would contribute to genome control (active transcription, chromatin architecture, among others)[37,38] whilst our data supports a key role for lncRNAs within the cytoplasm. Consequently, the direct link between lncRNAs and genome control as the main mechanism underlying transcriptome divergence falls flat and highlights the profound differences that exist between the actively transcribed RNAs and the steady-state RNA pools. The later observation comes as no surprise as it has been acknowledged for decades that the contribution of active transcription to steady-state RNA levels is rather modest and that major players in the regulation of transcriptome content operate at a post-transcriptional level[39-42].

Strikingly, our model suggests that transcriptome organization and divergence could result or be directed by a re-organization of the molecular players involved in the process and is mainly controlled by their intracellular localization. This exquisite molecular mechanism would ensure the generation of molecular and cellular heterogeneity by minimal energetic means and could provide the basis to maintain a basal level of phenotypic divergence within clonal populations. Moreover, it brings to light the observation that identity (i.e. cell type) and phenotypic output are not necessarily coupled and that cells, through the means of cell plasticity, can dramatically follow entirely different paths despite their identity. This concept opens a new dimension of biological insight as for instance in cancer research, most of the single cell analyses are based on cell type annotation (using hallmark markers) and not on what the cells are indeed doing in their context. Therefore, there is a clear need to further characterize and distinguish between cell identity and cell activity (expression of the phenotype) as it could completely alter the current view underlying how cellular systems operate.

Acknowledgements

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

Contributions

M.M.P. conceived the research project. Y.S., B.B., K.L.M. and M.M.P. designed and performed experiments. A.M. performed nuclear/cytoplasmic fractionation assay on HA1ER cells. M.M.P. performed computational analysis. N.O.C. set up the RPPA platform and A.F.M. performed RPPA assays. Y.S., B.B., K.L.M. and M.M.P. wrote the Material & Methods section and prepared figures. M.M.P. wrote the manuscript. Y.S., B.B., K.L.M., A.M., A.F.M., N.O.C. and M.M.P. edited the final manuscript.

Conflict of interest

N.O.C. is a co-founder, shareholder and management consultant for PhenoTherapeutics Ltd. All other authors declare on competing interests.

Materials & Methods

Cell lines

Immortalised HA1E (hTERT and SV40 ER) and tumorigenic HA1ER cells (hTERT, SV40ER and HRAS-G12V) from stepwise tumorigenesis models generated from normal human embryonic kidney cells were obtained from Prof. Weinberg (Broad Institute of MIT and Harvard, Cambridge, USA[9]). HepG2 cells were obtained from ATCC (ATCC, cat. HB-8065). HA1E and HA1ER cells were cultured in DMEM (1g/L glucose) (Gibco, cat. 21885025) supplemented with 10% heat-inactivated foetal calf serum (FCS, Gibco) and 100µg/ml hygromycin (hTERT selection, Thermo Fisher Scientific, cat. 10687010), 400µg/ml G418 (SV40-ER selection, Thermo Fisher Scientific, cat. 10131035), 0.5µg/ml puromycin (HRAS-G12V selection, Thermo Fisher Scientific, cat. A1113803) and 50 µg/ml gentamicin. HepG2 cells were cultured in DMEM high glucose (4.5g/L glucose) (Thermo Fisher, cat. 41966029) supplemented with 10% heat-inactivated FCS (Gibco) and 50U/ml Penicillin-Streptomycin (Thermo Fisher, cat. 15070063). Single cells were isolated by flow cytometric cell sorting (BD FACSAria III; BD Biosciences) and expanded in standard 2D culture conditions to establish clonal populations. LentiX 293T were obtained from Prof. Lacaud (Cancer Research UK Manchester Institute, Manchester, UK) and cultured in DMEM (1g/L glucose) (Gibco, cat. 21885025) supplemented with 10% heat-inactivated FCS (Gibco). HA1E and HA1ER cells were plated at a confluency of 1 x 10^6 cells/6010mm^2 surface with 10ml fresh medium, passed every 48 hours and kept in culture for a maximum of 15 passages. HepG2 cells were plated at a confluency of 1.5 x 10^6 cells/6010mm^2 surface with 10ml fresh medium, passed every 48 hours and kept in culture for a maximum of 15 passages. All cell lines were grown at 37°C with 5% carbon dioxide (CO_2) saturation. Mycoplasma contamination was analysed by reverse transcription followed by quantitative Polymerase Chain Reaction (RT-qPCR) (VenorGeM qEP, cat. 11-9100).

Lineage tracing libraries (LTv-BC-H2B-GFP)

The episomal vector GBX (EBNA1/OriP) was a gift from Linzhao Cheng (Addgene, cat. 64123) that was engineered to build a low-copy propagating vector in bacterial hosts (*bom* and *rop*) and with the potential to express H2B-GFP fusion protein in mammalian systems (LTv-H2B-GFP). Using this template, Barcode decay Lineage Tracing (BdLT-Seq) plasmids were generated by introducing a molecular barcode into the 3'UTR of H2B-GFP sequence. Briefly, the 3'UTR of the LTv-H2B-GFP vector was PCR amplified using a primer that contains a 12 nucleotide long random sequence split into three segments by intervening nucleotides (NNNNGCGNNNNTGANNNN). Next, PCR amplicons were cloned into the LTv-H2B-GFP vectors using HindIII and XbaI restriction sites (LTv-BC-H2B-GFP vectors). The generated high-complexity library was propagated using electrocompetent MegaX DH10B T1 bacteria strain (Invitrogen, cat. C640003). The final LTv-BC-H2B-GFP lineage tracing episome library contained approximately 1.2 x 10^6 unique barcodes and was already validated and used in our previous study[8].

Transfection with the LTv-BC-H2B-GFP episome libraries

Transfection of HA1ER cells with LTv-BC-H2B-GFP lineage tracing episome library was performed following standard direct transfection protocol using Lipofectamine LTX reagent (Invitrogen, cat. 153385000). The expression of H2B-GFP was verified by fluorescent microscopy (Thermo Fisher Scientific, cat. AMEX1000) and flow cytometry (BD LSRFortessa X-20 Cell Analyzer, BD Biosciences) 24- or 48-hours post-transfection depending on experimental set up (usually 25-35% of H2B-GFP positive cells). Flow cytometry data was analysed using FlowJo software (BD, version 10.8.0).

634 **BdLT-Seq**

635 A detailed BdLT-Seq protocol can be found in our published protocol[43]. HA1ER cells (clones F12, 1F8
636 and 1C9) were transfected with LTv-BC-H2B-GFP episome library as described. Forty-eight hours post-
637 transfection GFP positive cells were sorted using BD FACSAria III Cell Sorter (BD Biosciences) or BD
638 InFlux Cell Sorter (BD Biosciences) aiming for >94% purity after sorting. Cells were expanded in
639 standard culture conditions for five days – Phase I – a barcode decay phase in which most of the
640 episomes are lost upon cell division. The second barcode decay phase – Phase II - begins approximately
641 5-7 days after transfection where the episomes stabilize however still displaying a characteristic
642 episome loss. To perform BdLT-Seq (only during phase II), cells were trypsinised (Invitrogen, cat.
643 25300104)**,** collected by centrifugation at 230 rcf, 20°C, and plated at 2,000 to 3,000 cells per well
644 (0.32cm^2) according to experimental conditions. Five to six days later (~ 80,000 cells), cells from each
645 clonal population were split into two samples: sample N1 ("Day 0") was stored in liquid nitrogen (90%
646 FCS, 10% DMSO) whereas sample N2 ("Day 4") was expanded in standard culture conditions for four
647 additional days. Four days later, cells were stored in liquid nitrogen - sample N2 ("Day 4") (90% FCS,
648 10% DMSO). The day before single-cell RNA sequencing, cells were returned to standard 2D culture
649 for 24 hours. The following day, samples were processed for scRNA-Seq sequencing (10x Genomics
650 Chromium, Single-Cell RNA-Seq System, version 3.1, cat. PN-1000269) coupled to lineage tracing
651 (BdLT-Seq). Briefly, cells were collected by trypsinisation, washed in 5% FCS PBS and resuspended in
652 5% FCS PBS. Hashtag antibodies were used to specifically label individual samples for multiplexing
653 purposes. For that, cells were incubated with the hashtag antibodies coupled to a specific
654 oligonucleotide (HTO) (BioLegend, cat. A0251-A0260) at a ratio of 0.5 µg per 50,000 cells (stock 10
655 µg/ml) for 30 min, on ice. Cells were then washed three times using 0.04% BSA PBS and counted using
656 Neubauer counting chamber (Marienfeld, cat. 0640030). Samples were then combined according to
657 the experimental setting and 20,000 cells per channel (for the expected outcome of 10,000
658 cells/channel) were loaded onto 3 channels of the 10x Chromium platform (10x Genomics, cat.
659 1000204). Samples were processed following standard 10x Chromium protocol (10x Genomics,
660 version 3.1, cat. PN-1000269) up until cDNA amplification reaction. cDNA amplification reaction was
661 supplemented with the HTO additive primer and followed by a cDNA clean-up separating the HTO
662 fraction from the cell cDNA using standard methodologies (Beckman Coulter, AMPure XP, cat.
663 A63880). All size selected fractions (cell cDNA and HTO cDNA) were assessed by Fragment Analyser
664 5200 (Agilent, cat. M5310AA). Next, HTO depleted cDNA fraction (cell cDNA) was used to capture H2B-
665 GFP cDNA. For that, 5µl of Dynabeads MyOne Silane beads (Invitrogen, cat. 37002D) were washed
666 twice in 500µl RLT lysis buffer (Qiagen, cat. 79216), resuspended in 140µl RLT buffer (3.5x volume of
667 initial sample) and added to 40µl of the HTO depleted cDNA fraction. Then 180µl of absolute ethanol
668 was added to the reaction (4.5x volume of initial sample), mixed well and incubated for 15 minutes at
669 room temperature in a ThermoMixer C (300rpm, Eppendorf, cat. EP5382000015). The supernatant
670 was discarded, and the beads were washed three times in 80% ethanol, dried for 3-4 minutes at room
671 temperature, and eluted in 5µl nuclease free water (10-15 minutes at room temperature,
672 ThermoMixer C at 300rpm). To capture H2B-GFP cDNA, a set of 4 sequence-specific biotinylated
673 probes targeting H2B-GFP was used. The probes were hybridized H2B-GFP cDNA using xGen
674 Hybridization and Wash Kit (IDT, cat. 1080577) following manufacturer's instructions. Briefly, 8.5µl of
675 Gen 2x Hybridization buffer was combined with 2.7µl of Gen Hybridization enhancer, 5µl cDNA
676 fraction and 2µl of the probes (4 pmoles, pool of 4 probes), mixed, incubated for 10 minutes at room
677 temperature, followed by the incubation in a thermocycler: 30 seconds at 95°C and 16 hours at 65°C
678 (lid temperature 105°C). Following hybridization, the H2B-GFP cDNA fractions were pulled down using
679 streptavidin beads. Twenty microliters of streptavidin beads were washed twice using 100µl xGen 1x
680 Bead wash buffer and then resuspended in 17µl bead resuspension buffer (8.5µl xGen 2x Hybridization
681 buffer, 2.7µl Gen Hybridization buffer enhancer, 5.8µl nuclease free water). Seventeen microliters of
682 beads from the previous step were added to the hybridization reaction and incubated in a
683 thermocycler at 65°C for 45 minutes (lid temperature 70°C) mixing the reaction every 10 minutes.

After incubation, the samples were washed following manufacturer's instructions. Briefly, 100µl of 65°C wash buffer 1 was added to the sample and carefully mixed. The supernatant was transferred into a new tube and kept on ice (cDNA for 10x gene expression library – GEx fraction). Beads were then washed by adding 150µl of 65°C stringent buffer, mixed well and incubated in a thermocycler for 5 minutes at 56°C (lid temperature 70°C). Next, the supernatant was discarded, and the previous step was repeated for a total of two washes. 150µl of wash buffer 1 was added to the beads, incubated for 2 minutes at room temperature, supernatant was then discarded, 150µl of wash buffer 2 was added to the beads and incubated for 2 minutes at room temperature. After discarding the supernatant, 150µl of wash buffer 3 was added to the beads and incubated for 2 minutes at room temperature. To elute H2B-GFP cDNA from the beads, 100µl of 100mM NaOH 0.1mM EDTA was added to the complex beads/cDNA and incubated 15 minutes at room temperature in a ThermoMixer C (300 rpm). The supernatant was transferred into a new tube and 10µl 1M HCl, 10µl 1M TrisHCl (pH 7.5) was added. Both H2B-GFP cDNA fractions (lineage tracing fraction) as well as the fraction that contained the rest of the cDNA for the 10x gene expression library (GEx fraction) were purified using Dynabeads MyOne Silane beads (Invitrogen, cat. 37002D) following the protocol described above. 10x GEx libraries were generated using 120ng of the cDNA fraction following 10x Chromium protocol (Fragmentation, End repair and Library construction steps). H2B-GFP and HTO libraries were generated by amplifying either H2B-GFP or HTO cDNA using primers that specifically recognise H2B-GFP or HTO on the 5'-end and Read 1 on the 3'-end, generating standard Illumina paired-end sequencing constructs containing P5 and P7 (Supplementary Table 1). Final sequencing library quality was assessed by Fragment Analyser 5200 (Agilent, cat. M5310AA). Libraries were sequenced on a NovaSeq 6000 version 1.5 (Illumina, cat. 20013850). Sequencing depth: GEx library – 35000/50000 read pairs per cell, HTO library – 2000/3000 read pairs per cell, GFP and mCherry library – 10000/12000 read pairs per cell. Sequencing type: paired-end, dual indexing. Sequencing read: Read 1 – 28 cycles, i7 Index – 10 cycles, i5 Index – 10 cycles, Read 2 – 90 cycles.

10x Gene expression libraries and sequencing

10x libraries for HA1E (10E, A7A, A9H, B2F and B5F clones), HA1ER (F12, 1D12, 1C9, 2A6, 1F8, 3C5, 1E8, 2B12 and 1F5 clones) and HepG2 (1A, 32G, 14F, 48E and 44D clones) and for cells following gapmer-mediated knock-down were prepared following manufacturer's instructions. In order to perform a direct comparison of scRNA-Seq output avoiding mathematical integration cell preparations from each cell system analysed were multiplexed using HTOs (BioLegend, cat. A0251-A0260) prior to library construction. Libraries were sequenced on a NovaSeq 6000 version 1.5 (Illumina, cat. 20013850). Sequencing depth: GEx library – 35000/50000 read pairs per cell, HTO library – 2000/3000 read pairs per cell. Sequencing type: paired-end, dual indexing. Sequencing read: Read 1 – 28 cycles, i7 Index – 10 cycles, i5 Index – 10 cycles, Read 2 – 90 cycles.

Cell cycle analysis

Cells were plated at a confluency of 1 x 10^6 cells/6010mm^2 surface. Twenty-four to 48 hours later, cells were collected, fixed and permeabilised with methanol. Cell cycle progression was assessed using DAPI nuclear counterstain (Sigma-Aldrich, cat. MBD0015) and analysed by flow cytometry (BD LSRFortessa X-20 Cell Analyzer, BD Biosciences or Attune Nxt, Thermo Fisher Scientific).

Bright field imaging

Cells were plated at a confluency of 1-1.5 10^6 cells/10cm Petri dishes (6010mm^2 surface). Twenty-four hours later, cells were imaged using a fluorescent microscope (Thermo Fisher Scientific, cat. AMEX1000).

Cell proliferation

Cells were plated in standard 2D culture conditions at a defined number per well (HA1ER: 50,000 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 of 4

days, three wells per clone were harvested each day and counted manually using a Neubauer counting chamber.

Anchorage-independent growth in 3D soft agar

0.6% and 1% Difco agar noble solutions were prepared (BD Biosciences, cat. 214230) and autoclaved. Agar solutions were melted and equilibrated to 42°C in a water bath together with 2X growth medium (HepG2: DMEM high glucose (Sigma-Aldrich cat. D7777), 0.7% sodium bicarbonate (Sigma-Aldrich, cat. S8761), 20% FCS and 100U/ml Penicillin-Streptomycin (Thermo Fisher, cat. 15070063); HA1ER/HA1E: DMEM (Sigma-Aldrich cat. D5523), 0.7% sodium bicarbonate, 20% FCS and 100 µg/ml gentamicin (Thermo Fisher, cat. 15750037)). The lower 0.5% agar layer was generated by mixing the 1% agar solution 1:1 with the 2X medium and 1.5ml was added to each well of 6-well plates. After solidification of the bottom layer, the top 0.3% agar layer was prepared by mixing the 0.6% agar solution 1:1 with the 2X medium containing a fixed number of cells in a single-cell suspension (HA1ER: 10,000 cells/well, HA1E: 75,000 cells/well, HepG2: 22,500 cells/well) and 1.5ml was added per well. Once the top layer was set, 1ml of normal growth medium was added per well and 0.5ml of growth medium was added to each well every 3-4 days. For TGFβ-treated HepG2 samples, the growth medium was supplemented with 5ng/ml TGFv. Colony growth was assessed by crystal violet staining after 2-3 weeks (HA1ER/HA1E: 3 weeks, For this, wells were fixed and stained with 0.1% crystal violet (Sigma-Aldrich, cat. C6158) and 10% ethanol for 30min, followed by multiple washes over several days. Finally, plates were imaged on Optronics GelCount at 600dpi and processed for colony counting using ImageJ (version 2.0.0).

Transwell assay

Serum-free growth medium was used to preincubate transwells (Corning, cat. 3422) at 37°C during preparation of cells. Cells were trypsinised and resuspended in serum-free growth medium at a concentration of $1x10^6$ cells/ml. For TGFβ-treated HepG2 samples, medium was supplemented with 5ng/ml TGFβ. 0.5ml of complete growth medium (+5ng/ml TGFβ for TGFβ-treated HepG2 samples) was added to each well of a 24well plate and transwells were transferred to these wells, ensuring no bubbles were present between the transwell membrane and the lower medium. Finally, 100 µl of the cell suspension (100,000 cells) was added to the upper chamber of the transwell and incubated for 72 hours at 37°C. A cotton swab was used to remove non-migrated cells in the upper chamber, before migrated cells were fixed and permeabilised in ice-cold methanol for 10 minutes. Cells were stained in PBS supplemented with 10ug/ml DAPI (Sigma-Aldrich, cat. MBD0015) for 15min at room temperature and imaged using a 4x objective (EVOS FL Cell Imaging System). Adobe Photoshop was used to stitch together individual images covering the entire well and migrated cells were quantified using ImageJ.

TRAIL treatment and PARP labelling

Cells were plated at a confluency of 1×10^6 cells/6010mm^2 surface. Twenty-four hours later, HA1ER-derived clone F12 and subclones 1F8 and 1C9 were challenged with recombinant human TRAIL (rhTRAIL, 1µg/ml, R&D Systems, cat. 375-LT-010) for 4 hours. TRAIL-induced apoptosis was determined by assessing the percentage of cleaved PARP positive cells using flow cytometry (BD LSRFortessa X-20 Cell Analyzer, BD Biosciences) and FlowJo software (BD, version 10.8.0) was used for data analysis. Briefly, TRAIL-treated cells were collected and fixed with ice-cold methanol. Methanol fixated cells were washed once with 1x PBS and incubated in blocking solution (0.5% BSA PBS) for 1 hour at 20°C. Immunolabelling of cleaved PARP was performed using fluorescently labelled Cleaved PARP Alexa 647 antibody (1:200 dilution, Cell Signalling, cat. 6987) for 1 hour at 20°C, followed by two washes in blocking solution (0.5% BSA PBS).

TGFβ treatment

Prior to TGFβ treatment, HepG2 cells were plated in a 6-well plate at 150T cells per well. The next day, TGFβ was diluted in 10 µl medium per well and added to reach a final concentration of 5 ng/ml. As vehicle control, the same volume of water was diluted in 10 µl medium per well and added. Cells were

incubated at 37 °C for 48h, before exchanging the growth medium supplemented again with 5 ng/ml TGFβ. After additional incubation for 48h at 37 °C, cells were collected for analysis or plated without TGFβ for withdrawal experiments.

Western blot

Cells were lysed in RIPA buffer and whole cell extracts were quantified by Bradford assay to homogenize protein concentration across samples prior to protein separation by 10% SDS PAGE (including PageRuler Plus Prestained Protein Ladder, Thermo Fisher Scientific, cat. 26619) and transfer to nitrocellulose membranes. Membranes were incubated in 5% milk in PBS-T (0.1% Tween in PBS) for 1 h at 20°C and then washed 3 times in PBS-T for 5min. The membranes were incubated with antibodies against N-cadherin (1:2000) or α-Tubulin (1:8000) in 5% milk at 4°C overnight. Membranes were washed 3 times in PBS-T for 5 min and then incubated with the corresponding secondary antibody (1:10,000) for 1 h. Finally, membranes were washed 3 times in PBS-T for 5min and imaged on a LI-COR Odyssey imaging system.

Generation of a lentiviral library containing 26 RAS variants with allocated molecular barcodes in their H2B-GFP reporter

As described previously[8], gBlock DNA fragment encoding the H2B-GFP fusion containing a NdeI site in the 3'UTR and matching restriction sites at the ends for further downstream cloning into pLEX_307 (gift from David Root; Addgene, cat. 41392) (KpnI and SbfI) was amplified by PCR and cloned into the pGEM-T Easy vector system. Following verification of the construct by Sanger sequencing, this H2B-GFP fragment was further sub-cloned into the pLEX vector via One Shot ccdB Survival 2 T1 bacteria strain (Invitrogen, cat. A10460). To introduce a molecular barcode into the 3'UTR of the H2B-GFP gene, a gBlock DNA fragment with compatible restriction sites (NdeI and SbfI) encoding a modified version of the plasmid sequence downstream of the H2B-GFP stop codon was amplified by PCR using a primer with 12 random nucleotide barcodes split into three segments by intervening nucleotides (NNNNGCGNNNNTGANNNN), and further cloned into the pGEM-T Easy vector. Pooled plasmid DNA purification was performed (Qiagen Plasmid Midi Kit cat. 12143), the plasmid mix was digested using NdeI and SbfI for sub-cloning the barcoded fragment into the previously generated pLEX-H2B-GFP plasmid, and in-house generated electrocompetent One Shot ccdB Survival 2 T1 bacteria were transformed. Approximately 50,000 bacteria colonies were pooled for plasmid DNA purification of the barcoded pLEX-H2B-GFP plasmid mix. RAS variants were obtained from the Target Accelerator Pan-Cancer Mutant Collection (Addgene, cat. 1000000103). Individual liquid cultures of bacteria expressing pDONR plasmids encoding each of the 26 RAS variants included in our assay were pooled according to their OD_{600} to achieve a balanced representation, and then pDONR-RAS plasmid mix was purified using Qiagen Plasmid Midi Kit (cat. 12143). Following this, the pDONR-RAS plasmid mix was used to shuttle the oncogene encoding sequences into the previously generated pLEX-H2B-GFP plasmid mix by recombination using Gateway LR Clonase II Plus enzyme (Invitrogen, cat. 12538120) following the manufacturer's instructions. A unique molecular barcode was assigned to each RAS variant by (Sanger) sequencing the barcode region in the 3'UTR of H2B-GFP as well as the RAS variant region. Finally, liquid cultures of the bacteria clones containing the 26 final lentiviral transfer vectors (pLEX-H2B-GFP-BC-RAS: HRAS-A11D, HRAS-G12A, HRAS-G12C, HRAS-G12D, HRAS-Q61K, HRAS-Q61L, HRAS-Q61R, HRAS-wt, KRAS-G12A, KRAS-G12C, KRAS-G12D, KRAS-G12V, KRAS-G12S, KRAS-D33E, KRAS-A59G, KRAS-E62K, KRAS-wt, NRAS-G12A, NRAS-G12C, NRAS-Q61H, NRAS-Q61K, NRAS-Q61L, NRAS-Q61R, NRAS-Y64D, NRAS-wt) were pooled based on their optical density and plasmid DNA was purified, rendering the transfer vector mix for lentivirus production.

Determination of the functional lentiviral titer and transduction of HA1E cells with 26 RAS variants

Increasing volumes of lentivirus stocks (0-500 µl) were added to 150,000 HA1E cells and supplemented with polybrene (Sigma, cat. H9268-5G) to a final concentration of 2 µg/ml. The percentage of GFP

835 positive cells was determined 48 hours post-transduction by flow cytometry (BD LSRFortessa X-20 Cell
836 Analyzer, BD Biosciences) and FlowJo software (BD, version 10.8.0) was used for data analysis. The
837 functional viral titer, calculated as the number of viral particles (i.e. transducing units (TU)) capable to
838 transduce a particular cell line per volume was calculated using the following formula:
839

841 $$\text{functional titer}\left[\frac{\text{TU}}{\mu\text{l}}\right] = \frac{\text{cell number x frequency of GFP + cells [\%]}}{\text{volume of virus [}\mu\text{l]}}$$

840

842 Based on this, HA1E cells were transduced with the pooled lentiviral library containing 26 RAS variants
843 at a multiplicity of infection (MOI) of 0.05, supplemented with polybrene at a final concentration of
844 2µg/ml. Forty-eight hours post-transduction, FACS was used to collect GFP-positive cells (BD InFlux Cell
845 sorter or FACSAria Fusion Cell Sorter, both BD Biosciences) with a purity of >90% and sorted cells were
846 re-plated in standard growing conditions.
847
848

849 **Single-cell data and BdLT-Seq analysis**

850 10x Chromium single-cell data was analysed using Seurat v4.0[44]. HTO-labelled cells were demultiplexed
851 using the HTODEMUX module of Seurat and further processed for quality control and cell filtering
852 taking into account nFeature_RNA, nCount_RNA and mitochondrial content. Further pre-processing
853 included the mathematical regression of cell cycle related gene expression levels to remove cell cycle
854 related bias in cluster processing. Similarly, variance in cluster detection due to high levels of
855 mitochondrial transcripts present per cell was also regressed prior to final clustering. HTO and BdLT-
856 Seq data was processed using CITE-seq-count[45] prior to demultiplexing or barcode counting. BdLT-Seq
857 barcode reads were pre-processed to remove sequencing artefacts, flanking and intervening
858 sequences using cutadapt v3.7[46]. Clean barcodes were then processed using CITE-seq-count to
859 generate lineage tracing files. A custom Node script[8] was used to determine lineage relationships
860 based on a similarity score considering the number of similar barcodes between two cells and data
861 obtained by these means was used to build a directional lineage tree. Lineage visualization was
862 generated using CIRCOS[47]. All data analysed across the manuscript was pre-processed/processed using
863 R scripts. Final image processing and output was obtained using Adobe Illustrator.
864

865 **Whole Exome Sequencing**

866 Founder clones and thereof derived subclones from each cellular system were evaluated to assess
867 possible genetic divergence at the exome level (WES, Whole Exome Sequencing. 100x coverage).
868 Briefly, cells were grown as described above and processed for DNA purification using the QIAmp DNA
869 mini kit (Qiagen. Cat 51304). High quality DNA Samples were then either processed for exome capture
870 using the SureSelect Human All Exon V6 (Agilent, cat 5190-8872) and then processed for DNBSeq
871 compatible library construction (BGI) or directly streamlined into DNBSeq library construction (BGI).
872 WES sequencing was performed following a 100PE sequencing protocol in a DNBSeq instrument (BGI).
873 WES was mapped to hg38 from each clonal lineage analysed, piled-up using mpileup[48] (-q1 -Q30) and
874 further processed through VarScan Somatic pipeline[49] (--min-coverage 50). Chromosome level data
875 visualization was generated using chromo_plot from the vcfR package[50].
876

877 **SLAM-Seq and SLAM-DUNK**

878 SLAM-Seq was used to determine the fraction of nascent RNAs in contrast to steady state RNA levels
879 for each clonal lineage (at least 3 biological triplicates). Briefly, the founder clone and its corresponding
880 subclones from each system were pulsed with 10µM 4-thiouridine (4-SU. Sigma, cat. 4509) for 15 min
881 (1F8 vs 1C9) or 45 minutes (5 clones of HA1ER, HA1E and HepG2) and collected in Trizol (Invitrogen,
882 cat. 15596-026). After RNA quality control assessment (Fragment Analyzer 5200, Agilent, cat.
883 M5310AA) and mass determination (Qubit, Invitrogen), 5 µg of RNA from each sample was incubated
884 in 4-SU conversion buffer (10 mM IAA in H_2O, 50 mM sodium phosphate pH 8, 0.15mM DTT, 50% DMSO

885 (v/v) and H_2O to a final volume of 50 ul) for 15 min at 50°C on a heat block. Immediately after, the
886 conversion reaction was stopped by addition of 2 µl of 1M DTT and EtOH precipitated over-night at -
887 80°C (adding 200µl H_2O, 2 µl glycogen (20 mg/ml), 30 µl 3M pH 5.2sodium acetate and 900 µ l ice-cold
888 absolute EtOH). Retrieved RNA was further processed to build an Illumina compatible strand-specific
889 RNA sequencing library using either NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB,
890 cat. E7760S, for 1F8 vs 1C9) or QuantSeq 3' mRNA-Seq V2 Library Prep Kit FWD with UDI (Lexogen, all
891 clonal systems) kits following manufacturer's instructions. SLAM-Seq sequencing data was processed
892 using a T>C aware mapper (NextGenMap[13]) and processed following SLAM-DUNK pipeline[14]
893 (NextGenMap align to hg38 reference genome, mapping quality filter, snp annotation filtering and
894 modified reads count). Final count tables were further processed using the beta binomial test to
895 determine transcriptional differences from steady state RNA levels (FDR>0.01). Data visualization was
896 generated using heatmap.2 (R v4.2.0).
897
898 **CUT&RUN**
899 HA1ER anchoring clone F12 and two diverging subclones 1F8 and 1C9 were plated at a $1x10^6$/10cm
900 plate 24 h prior to the experiment. The following day cells were collected by trypsinization, spun down
901 at 1100rpm for 4 minutes, washed in ice-cold PBS and resuspended in 5ml of freshly prepared swelling
902 buffer (25mM HEPES pH7.8, 1.5mM MgCl2, 10mM KCl, 0.5 % NP40 (Igepal CA630), 1mM DTT and
903 Protease inhibitors) and further incubated for 5 minutes on ice. Nuclei were then spun down at 1100
904 rpm for 5 minutes and the resuspended in 5ml of freshly prepared MNase buffer (50mM Tris HCl pH
905 7.6, 0.2% Triton X100, 5mM Na butyrate and Protease inhibitors). Next, nuclei were spun down at 1100
906 rpm for 5 minutes and the resuspended in 5ml of MNase buffer. Nuclei were then counted in a
907 Neubauer chamber and $1x10^6$ nuclei per condition were spun down at 1100 rpm for 5 minutes. Each
908 nuclei sample was then resuspended in 100 µl of MNase buffer and after the addition of targeting
909 antibodies (H3K27me3, H3K27Ac and CTCF. 5 µg each) they were incubated for 1h at 20°C in a
910 thermomixer comfort pipetting up and down every 10 minutes. After incubation, each sample was
911 washed twice in 1 ml of MNase buffer followed by spin down at 1100 rpm for 5 minutes. Then, nuclei
912 were resuspended in 100 µl of MNase buffer containing 2.5 µl of pA/G-MNase (in-house prepared)
913 and then incubated for 30 min at 20°C in a thermomixer comfort mixing the nuclei pipetting up and
914 down from time to time. After this incubation, each sample was washed twice in 1 ml of MNase buffer
915 followed by spin down at 1100 rpm for 5 minutes. Next, samples were resuspended in 100 µl of MNase
916 buffer supplemented with 2 µl of 50 mM $CaCl_2$ to activate the MNase and incubated for 15 min at 37°C
917 in a thermomixer comfort. Immediately after the reaction is stopped by adding 20 µl of freshly
918 prepared STOP buffer (50 mM EDTA and 50 mM Tris HCl pH7.6) and incubating the samples for 5 min
919 at 20°C. After this, samples were spun down at 10.000rpm for 5 min and the supernatant was
920 recovered. Samples were then processed for Phenol:Chloroform extraction and purified DNA was used
921 to build Illumina compatible sequencing libraries using the NEB next ultra II directional library prep kit
922 starting with End repair and A tailing following manufacturer's instructions. Sequencing libraries were
923 mapped to hg38 using Bowtie2[51] and MACS2 v2.2.7.1[52] was used to identified peaks (FDR<0.01). A
924 catalogue containing all peaks identified on each sample (clones 1F8 and 1C9) for each antibody used
925 (H3K27me3, H3K27Ac and CTCF) was built to extract genomic coordinates used in heatmap
926 visualization (Deeptools v3.5.1[53]).
927
928 **Label-free proteomic analysis (LC-MS-MS)**
929 Cell lysis was performed in 5% SDS, 50 mM TEAB followed by protein content measurement using the
930 2D quant kit (Cytiva, cat. 80-6483-56). Then, 200 µg of protein from each sample was reduced with
931 5mm TCEP at 55°C for 15 min followed by alkylation with 15mM MMTS for 10 min at 20°C. Samples
932 were then loaded onto STRAP minis (Protifi, cat. C02-mini) and digested with 20 µg sequencing grade
933 trypsin (Sigma, cat. TRYPSEQ-RO) for 18 h at 37°C. Following STRAP elution, peptide digests were dried
934 to completeness in a vacuum centrifuge at 45°C before resuspension at 800ng/µl on 0.2% formic acid.
935

936 LC-MS was performed in technical quintuplicate in a RSLC HPLC system (mobile phase A: water 0.1%
937 formic acid; mobile phase B: acetonitrile) coupled to an Orbitrap Lumos mass-spectrometer via an
938 EasySpray source (Thermo Fisher Scientific). Peptides were injected directly onto an EasySpray column
939 (50 cm long, 7 µm ID pepmap C18) at 180 nl/min in 1% mobile phase B and were separated at 180
940 nl/min with a gradient of 1-24% over 90 mins. LC eluent was sprayed directly into the MS source at an
941 ion spray voltage of 1.9Kv. The MS was operated in data dependent mode where an MS1 survey scan
942 was acquired in the Orbitrap detector at 120K resolution (at m/z 200) in the mass range of 350-
943 1200m/z at a target value of 2e5 ions and a maximum fill time of 40ms. All available charged ions were
944 selected for MS2 by HCD at a normalised collision energy of 35% with product ions detected in the
945 Orbitrap detector at a resolution of 30K, 54ms max fill time and a target value of 1e5 ions. A maximum
946 cycle time of 1.2s was allowed.
947
948 **TMT-proteomics**
949 Cell lysis was performed in 5% SDS, 50mM TEAB followed by protein content measurement using the
950 2D quant kit (Cytiva, cat. 80-6483-56). Then, 200 µg of protein from each sample was reduced with 5
951 mm TCEP at 55 °C for 15 min followed by alkylation with 15 mM MMTS for 10 min at 20°C. Samples
952 were then loaded onto STRAP minis (Protifi, cat. C02-mini) and digested with 20 µg sequencing grade
953 trypsin (Sigma, cat. TRYPSEQ-RO) for 18 h at 37°C. Following STRAP elution, peptide digests were dried
954 to completeness in a vacuum centrifuge at 45°C. Peptide digests were resuspended in 500 mM TEAB
955 at a concentration of 2 ug/µl and TMT 10-plex labelled according to the manufacturer's instructions
956 (Thermo Fisher Scienific). Reactions were quenched by addition of 8 µl 5% hydroxylamine prior to
957 pooling. Samples were acidified to pH 2.7 by addition of trifuoroacetic acid before revers phase solid
958 phase extraction.
959 Pooled, acidified peptides were loaded onto C18 RP cartridges (Sep-Pak 500mg tC18, Waters, cat.
960 186004619), washed with 10 ml 0.1 % formic acid before elution in 500 µl 60% acetonitrile, 0.1% formic
961 acid. 10% of the pooled sample was aspirated off, the remaining 90% of peptide was split equally into
962 2 tubes and all aliquots were dried to completeness in a vacuum centrifuge at 45°C. Approximately a
963 10 % aliquot (200 µg) was reserved for proteomics profiling and both 45% aliquot (900 µg) were
964 resuspended for subsequent enrichment either by TiO2 (Thermo Fisher cat. A32993) or FeNTA
965 enrichment (Thermo Fisher cat. A32992). Following elution from either TiO2 or FeNTA columns,
966 samples were dried to completeness in a vacuum centrifuge at 20°C followed by resuspension in 5 µl
967 of 0.05% TFA and were ready for LC-MS-MS analysis.
968 LC-MS was performed in technical quintuplicate in a RSLC HPLC system (mobile phase A: water 0.1%
969 formic acid; mobile phase B: acetonitrile) coupled to an Orbitrap Lumos mass-spectrometer via an
970 EasySpray source (Thermo Fisher Scientific). Following, peptides/phospho-peptides were injected
971 directly onto an EasySpray column (50 cm long, 75 µm ID pepmap C18) at 180 nl/min in 1% mobile
972 phase B. Peptides were separated at 180 nl/min with a gradient of 1-24% over 90 min. LC eluent was
973 sprayed directly into the MS source at an ion spray voltage of 1.9 Kv. The MS was operated in data
974 dependent mode where an MS1 survey scan was acquired in the Orbitrap detector at 120 K resolution
975 (at m/z 200) in the mass ranges of 350-1200 m/z at a target value of 2e5 ions and a maximum fill time
976 of 40 ms. All available charged ions were selected for selected for MS2 by HCD at a normalised collision
977 energy of 42%. Product ions were detected in the Orbitrap detector at a resolution of 50K, 86 ms max
978 fill time and a target value of 1e5 ions. A maximum cycle time of 4s was allowed.
979
980 **Label-free proteomics data processing**
981 Raw data was imported into Progenesis QI proteomics pipeline (Waters Nonlinear Dynamics) where
982 peak detection was performed at maximum sensitivity. Protein relative quantitation was performed
983 using the top 3 method. Proteins were identified by extraction of a mascot generic file (MGF) that was
984 exported to Mascot (Matrix science) with the following search criterion. Species human, specificity
985 was trypsin allowing up to 2 missed cleavages, precursor mass accuracy 10ppm, product ion sensitivity
986 20 ppm, fixed modification Methylthio (C) variable modifications oxidation (M) deamidation (NQ)

searched against Uniprot_Swissprot_2022 using a decoy FDR method limited at 1%. Protein identities were imported back into Progenesis QI and data was further processed in R for visualization purposes.

TMT proteomic analysis (PEAKS)

Raw data was imported into Peaks Studio 10.6 (Bioinformatics Solutions Inc.) where both protein identification and quantitation were performed using the default TMT 10 plex template. Database search evaluated human acquired data against the Uniprot_Sprot_2022.fasta database using fixed modification of methylthio (C) variable modifications oxidation (M) deamidation (NQ) and the following advance settings: merge scans [DDA], correct precursor [DDA], mass only, associate features with chimera scan [DDA], filter features charge between 2 and 4 10 ppm precursor mass tolerance, fragment ion 0.1 Da, digest mode specific trypsin, max missed cleavages per peptide 2, estimated FDR by decoy-fusion, quant TMT 10 plex, MS2 quant tolerance 20 ppm, FDR threshold 1%.

TMT phospho-proteomic analysis (PEAKS)

Raw data was imported into Peaks Studio 10.6 (Bioinformatics Solutions Inc.) where both protein identification and quantitation were performed using the default TMT 10 plex template. Database search evaluated human acquired data against the Uniprot_Sprot_2022.fasta database using fixed modification of methylthio (C), variable modifications oxidation (M), deamidation (NQ), phosphorylation (STY) and the following advance settings: merge scans [DDA], correct precursor [DDA], mass only, associate features with chimera scan [DDA], filter features charge between 2 and 4 10ppm precursor mass tolerance, fragment ion 0.1Da, digest mode specific trypsin, max missed cleavages per peptide 2, estimated FDR by decoy-fusion, quant TMT 10 plex, MS2 quant tolerance 20ppm, FDR threshold 1%.

CyTOF

HA1ER cells (clones F12, 1C9 and 1F8; $3x10^6$ each) were pulsed for 30 min with IdU following manufacturer's instructions (Maxpar cell cycle panel kit, cat. 201313) and then collected by trypsinization. Cells were next spun-down and resuspended in 1ml 1X Maxpar Fix I Buffer (Fluidigm, cat. 201065) for 10 min at 20°C. Then cells were spun down and resuspended in 7 ml Maxpar Cell Staining Buffer (Fluidigm, cat. 201068), mixed gently and further incubated for 10 min at 20°C. Next, cells were spun down once again and after repeating this washing step, cells were resuspended in 50 µl of Maxpar Cell Staining buffer and then permeabilized with ice-cold methanol for 15 min on ice. Cells were then washed twice with 2 ml Maxpar Cell Staining Buffer +EDTA (Maxpar Cell Staining Buffer suppmenented with 5mM EDTA) and spun down for 5 min at 800g. Immediately after, 1 µl of each metal isotope-labelled specific antibody (MOT4, JIP4, SYUA, DCXR, SQSTM, MYEF2, Hsp20, pRb, CyclinB1, Ki67, pH3 (Supplementary Table), using Maxpar X8 Antibody Labeling Kit (Fluidigm, cat. 201144A, 201167A, 201155A, 201171A, 201176A, 201147A, 201159A)) was added to Maxpar Cell Staining Buffer +EDTA and a final volume of 50 µl was added per sample and incubated for 30 min at 20°C. Following incubation, cells were washed twice with 2 ml Maxpar Cell Staining Buffer +EDTA and further spun down at 800 g for 4 mins. Immediately after, supernatant was removed and 1ml of 4% PFA and 1 µl Iridium were added to each sample and stored at 4°C for 16 h. Next day, samples were washed in 2 ml of Maxpar Cell Staining Buffer +EDTA and spun down. Finally, samples were washed twice in Maxpar Cell Acquisiition Plus Solution (Fluidigm, cat. 201244). Samples were analysed in a Helios apparatus and data was processed using flowCore v2.16.0[54], CATALYST v1.28.0[55] and visualized in R v4.2.0.

Gapmer-mediated knock-down

Antisense LNA gapmers were designed against lncRNA transcript sequences (Ensembl release 105[56]) using the Qiagen Antisense LNA GapmeR custom builder (Supplementary Table). One day prior to transfection, 100T cells (HA1ER) or 150T cells (HepG2) were seeded into a 6well plate in 2ml growth medium. Next day, around 450 µl of medium were removed from each well for a final volume of 1.5ml

per well. 40 µM gapmer was mixed with 75 µl OptiMEM (Invitrogen, cat. 31985062) per well to be transfected. The X-tremeGENE 360 transfection reagent (Roche, cat. 8724105001) was equilibrated to 20°C and briefly vortexed before immediately adding 2.5 µl per well to be transfected to the gapmer mix. After thorough mixing and incubation at for 30min at 20°C, 75 µl of the transfection mix was added dropwise to each well of cells (final gapmer concentration in 1.5ml: 26 nM). Following gentle mixing, cells were incubated at 37°C. The growth medium was exchanged 24h post-transfection. Cells were used in RT-qPCR, scRNA-seq or RPPA assay 48h post-transfection, only for DPP-DT-G2 cells were used 24h post-transfection.

Nuclear-Cytoplasmic fractionation for HA1ER clones

HA1ER clones F12, 1F8 and 1C9 were seeded at 1×10^6 onto a 10 cm culture dish and were grown to ~70% confluency. Cells were rinsed twice with 10 ml ice-cold PBS and trypsinized with 1 ml of 0.25% trypsin. Cells were pelleted by centrifugation at 220g for 3 minutes at 4°C and then washed twice in 1 ml ice-cold PBS, counted, and 2×10^6 cells were then taken for further fractionation. Pellets were resuspended in 100 µl of hypotonic buffer (20 mM Tris pH 7.6, 0.1 mM EDTA, 2 mM $MgCl_2$, 0.5 mM NaF, and 0.5 mM Na_3VO_4) supplemented with 1x cOmplete EDTA-free Protease Inhibitor Cocktail (Roche, cat. 11873580001), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 260 U/ml RiboLock RNase Inhibitor (Thermo Fisher Scientific, cat. EO0381), and transferred to a fresh 2ml DNase/RNase-free DNA-low binding tube. Cells were then incubated for 2 minutes at 20°C, then for 10 minutes on ice to induce hypotonic swelling, followed by addition of 10 µl of 10% IGEPAL CA-630 (Sigma-Aldrich, cat. I8896) and pipetted up-and-down approximately 20 times. Nuclei, unbroken, or partially lysed cells were pelleted at 550g for 3 minutes at 4°C and then 80% of the supernatant was taken from the top (containing the cytoplasmic fraction), transferred to a fresh tube. Next, 150 µl of hypotonic buffer was added to the pellet and remaining cytoplasmic extract; the pellet was resuspended and once again spun down. 80% of the supernatant was taken from the top and combined with the cytoplasmic fraction. Nuclei pellets were then resuspended in hypotonic buffer in a volume equal to the final cytoplasmic fraction (373 µl). 20 ng of spike-in eGFP transcript for HA1ER cell system (in-vitro transcribed from an eGFP encoding plasmid. HighYield T7 RNA Synthesis Kit, Jena Bioscience, cat. NT-101) was added to each fraction, and total RNA was extracted using 1 ml of TRIzol (Invitrogen, cat. 15596-026). 773 µl of the aqueous phase from each fraction was mixed with 25.5 µl of 3M NaOAc, 1.5 µl of 5 mg/ml linear acrylamide (Invitrogen, cat. AM9520), 1 volume of isopropanol, and RNA was precipitated overnight at -20°C. Pellets were washed with 75% ice-cold ethanol, dried, and re-dissolved in 12 µl of nuclease-free water. RNA concentrations were measured using Nanodrop.

Nuclear-Cytoplasmic fractionation for HepG2 clones

HepG2 clones 1A, 14F and 44D were seeded at 1.5×10^6 onto a 10 cm culture dish and were grown to ~70% confluency. Cells were harvested and resuspended in ice-cold PBS and from thereon kept on ice. For each sample, 2×10^6 cells were transferred to a fresh tube, pelleted and resuspended in 380 µl ice-cold hypotonic lysis buffer (HLB, 10mM Tris-HCl pH7.5, 10mM NaCl, 3mM $MgCl_2$, 10% glycerol, 0.3% IGEPAL CA-630) with 100U Protector RNase inhibitor (Roche, cat. 3335399001) and incubated on ice for 10min. Following brief vortexing and pelleting by centrifugation, the supernatant was transferred to a fresh tube (cytoplasmic fraction). The pelleted nuclei were washed twice in 380 µl HLB and finally resuspended in 380 µl (nuclear fraction). 1 ng of a custom spike-in plasmid was added to each sample to allow for normalization during RT-qPCR analysis. After adding 1ml TRIzol to each fraction, RNA was extracted according to the manufacturer's instructions. To ensure quantitative comparisons between samples in RT-qPCR, care was taken to treat samples equally during RNA extraction, e.g. collecting equal volume of aqueous phase for each sample.

Reverse transcription followed by qPCR

Total RNA was isolated using TRIzol (Invitrogen, cat. 15596-026) following manufacturer's instructions. Reverse transcription was performed using QuantiTect Reverse Transcription Kit (Qiagen, cat. 205311)

following the manufacturer's instructions with 1 or 2µg RNA input. qPCR was performed using PowerUp SYBR Green Master Mix (Invitrogen, cat. A25742) in a QuantStudio3 Real Time PCR System (Thermo Fisher Scientific, cat. A28136). Data analysis was conducted using Real-Time qPCR Connect Data Analysis Tool (Thermo Fisher Scientific Connect Platform).

FISH probes design and production

FISH probe amplicon regions between 260-300bp were determined based on transcript sequences (Ensembl release 105[56]) and primers to amplify these regions were then designed using the Connect OligoPerfect tool (Thermo Fisher). RNA purified from cells known to express target RNAs at high level was reverse transcribed with QuantiTect reverse transcription kit (Qiagen, cat. 205311) using oligo-dT primers following the manufacturer's recommendations. cDNA was then amplified by the Expand high fidelity PCR system (Roche, cat. 11732641001) using specific primers and PCR-products were then resolved by agarose gel electrophoresis. Bands of the expected molecular weights were purified and quantified using the Qubit dsDNA BR assay kit (Invitrogen, ca. Q32850). 1µg amplicon DNA was then digested using Lambda exonuclease (NEB, cat. M0262) at 37°C for 1 hour, followed by heat inactivation at 75°C for 10 minutes. Digestion products were then isolated on streptavidin beads and then washed twice with 1X B+W buffer (5mM Tris-HCl pH7.5, 1mM EDTA, 2M NaCl). Next, beads were resuspended in 25µl amplification mix 1 (1µl dNTP mix (10mM dATP, 10mM dCTP, 10mM dGTP, 7.5mM dTTP and 2.5mM aminoallyl dUTP (Thermo Scientific, cat. R1101)), 2µl 20µM random hexamers, 22µl H_2O) before the addition of 25µl amplification mix 2 (0.75µl Expand high fidelity enzyme mix (Roche, cat. 11732641001), 5µl 10X Expand high fidelity buffer with 15mM $MgCl_2$ (Roche, cat. 11732641001) and 19.25µl H_2O). Beads were then washed twice with H_2O and resuspended in 25µl H_2O. 15µl labelling buffer (25 mg/ml sodium bicarbonate (Sigma-Aldrich, cat. S5761) was added to the resuspended beads before addition of 10µl dye (Alexa Fluor 555/647 Reactive Dye (Invitrogen, cat. A32757, A32756) dissolved in 40 µl DMSO). Samples were vortexed briefly and incubated at room temperature for 1 hour blocked from light. Beads were then washed three times in H_2O and FISH probes were released through chemical denaturation by resuspending beads in 100µl filtered denaturation buffer (100mM NaOH, 0.1mM EDTA pH8.0) and incubating at room temperature for 15 min blocked from light. The supernatant was separated from the beads using a magnetic separator and collected into a fresh tube. Probes were then precipitated from the supernatant using 1/10 vol sodium acetate, 2µl Ethachinmate (Nippon Gene, cat. 318-01793) and 2.5X vol 100% ethanol, and then eluted in 20µl H_2O. Probes stocks were stored at -80°C.

RNA-FISH

Cells were grown to 70% confluency on glass coverslips in 24-well plates and treated according to experimental design. Cells were then washed twice with 1X ice-cold PBS and fixed with 3% formaldehyde (Thermo Scientific, cat. 28906) and 4% sucrose in 1X PBS for 15 minutes at 20°C. Following fixation, cells were washed three times in 1X PBS and permeabilised with 0.25% Triton-X in 1X PBS for 10 minutes at 20°C, followed by three washes in 1X PBS. Fluorescently labelled RNA-FISH probes were diluted in formamide (10µl total volume per coverslip) and denatured at 75°C for 7 minutes. 10µl 2X hybridisation buffer (4X SSC, 40% dextran sulphate (Sigma-Aldrich, cat. D8906), 2mg/ml BSA) per coverslip was then added to each probe dilution. 20µl probe mix was hybridised to each sample of cells in a Thermobrite 110/120 VAC hybridisation chamber (Abbott, cat. 07J91-010), using the incubation program as follows: 65°C for 5 minutes, 37°C for 16 hours. Following hybridisation, samples were washed three times in 50% formamide/2X SSC at 42°C for 5 minutes each, three times in 2X SSC at 42°C for 5 minutes each and three times in 1X SSC at 42°C for 5 minutes each. Samples were then counterstained with DAPI and washed twice in 1X SSC, before mounting in ProLong Gold Antifade Mountant (Molecular Probes, cat. P36930). Cells were analyzed on a Zeiss LSM880 microscope equipped with Airyscan module.

RNA-FISH coupled to ICC

Cells were grown to 70% confluency on glass coverslips in 24-well plates and treated according to experimental design. Cells were then washed once with 1X PBS and fixed with 3% formaldehyde (Thermo Scientific, cat. 28906) and 4% sucrose in 1X PBS for 15 minutes at room temperature. Following fixation, cells were washed three times in 1X PBS and permeabilised with 0.25% Triton-X in 1X PBS for 10 minutes at room temperature, followed by three washes in 1X PBS. Cells were blocked in blocking buffer (1% BSA in 1X PBSfor at least 2 hours at 20°C, before incubation with 1:200 primary antibody in blocking buffer overnight at 4°C. Cells were then washed twice with 0.1% Tween-20 in 1X PBS and incubated with 1:1000 secondary antibody in blocking buffer for 1 hour at 20°C, blocked from light. Cells were then washed twice with blocking buffer and then crosslinked again with 3% formaldehyde and 4% sucrose in 1X PBS for 10 minutes at 20°C. Next, cells were washed twice with 2X SSC (Sigma-Aldrich, cat. S6639) / 50% formamide (Sigma-Aldrich, cat. F9037). Next, fluorescently labelled RNA-FISH probes were diluted in formamide (10µl total volume per coverslip) and denatured at 75°C for 7 minutes. 10µl 2X hybridisation buffer (4X SSC, 40% dextran sulphate (Sigma-Aldrich, cat. D8906), 2mg/ml BSA) per coverslip was then added to each probe dilution. 20µl probe mix was hybridised to each sample of cells in a Thermobrite 110/120 VAC hybridisation chamber (Abbott, cat. 07J91-010), using the incubation program as follows: 65°C for 5 minutes, 37°C for 16 hours. Following hybridisation, samples were washed three times in 50% formamide/2X SSC at 42°C for 5 minutes each, three times in 2X SSC at 42°C for 5 minutes each and three times in 1X SSC at 42°C for 5 minutes each. Samples were then counterstained with DAPI and washed twice in 1X SSC, before mounting in ProLong Gold Antifade (Invitrogen, cat. P36930). Cells were visualised with a Zeiss LSM880 microscope equipped with Airyscan module.

Reverse Phase Protein Array (RPPA)

Cells were transfected and cultured as described in the section 'Gapmer-mediated knock-down'. To prepare samples for RPPA, cells were kept on ice and washed twice with ice-cold PBS. Next, lysis buffer (1.9 % Triton X-100 (v/v), 0.25M HEPES pH7.4, 0.25M EGTA ph7.5-8, 0.5M NaCl, 0.5M $MgCl_2$, 100 mM Na_3VO_4, 100mM TSPP, 1M NaF, 10 % glycerol, 1 tablet/10ml cOmplete Mini EDTA-free protease inhibitor (Roche, cat. 11836170001), 1 tablet/10ml PhosStop phosphatase inhibitor (Roche, cat. 04906837001) was added and cells were scraped and collected into a tube. After a 20min incubation on ice, samples were centrifuged at 16,900 rcf for 10min at 4°C and the supernatant protein lysate was collected and stored at -20°C. Protein concentration was determined by Bradford assay. Samples were prepared for printing by addition of 4X Sample buffer (40% Glycerol, 8% SDS, 0.25 M Tris-HCl, 10% β-mercaptoethanol) and heat denaturation at 95°C for 5 min. All samples were 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 containing 10% glycerol. Array spotting was carried out with the Quanterix 2470 Arrayer platform using 185 µM pins. Each sample was spotted at four dilutions (D1-D4) onto single pad ONCYTE SuperNOVA nitrocellulose slides (Grace Bio-Labs, cat. 705278) at a 500 µM spot-to-spot distance and in 3 technical replicates. 180 signalling pathway markers (Supplementary Table) were profiled in a standard RPPA assay as described briefly below[18]. After spotting, slides were incubated with antigen retrieval solution (1x Reblot strong. Merck Millipore, cat. 2504) for 10 min before being placed in a microfluidic structure to individually address the arrays with primary or secondary antibody solutions. Following blocking buffer (Superblock T20. Thermo Scientific, cat. 37536) for 10 min the detection of marker antibodies was performed in a two-step sequential assay. First the array is incubated with primary analyte-specific antibody in blocking buffer for 60 mins at room temperature followed by the removal of excess antibody by washing arrays with 0.1% PBS-Tween followed by a further incubation with blocking buffer and PBS-T washes, and incubation with secondary antibody (Dylight-800-labeled anti-species antibodies diluted 1:2500 in Superblock T20) for 30 mins. After further washing and slide drying, the arrays were imaged in the Innopsys Innoscan 710 scanner (Innopsys). Blank signals were determined by omitting the primary antibody and instead incubating the array with Superblock T20 alone, followed by secondary antibody incubation. Sample loading on arrays (for normalisation) was determined by staining one slide with fast green protein dye and scanning at 800 nm. Microarray images were

analyzed using Mapix software (Innopsys) with feature (spot) diameter of the grid set to 270µm. Average signal intensities was determined for each individual feature and the median background from the adjacent area was subtracted from each feature signal leading to a net signal per feature. Fluorescence intensity for each feature (spot) on the array was measured and a test was performed for linear fit through the 4-point dilution series for all samples on all arrays using a flag system where R2 > 0.9 (green flag) is deemed good, >0.75 (amber flag) is deemed acceptable and <0.75 (red flag) is poor and was excluded from data analysis. The median values from the 4-point dilution series were calculated and used as a measure of fluorescence intensity.

Statistics and reproducibility

All histograms displayed in the manuscript show mean value ± standard deviation (SD) or propagated error with the number of replicates and, if performed, the method to determine statistical significance indicated in the figure legends.

As part of scRNA-Seq data pre-processing, cells found to be displaying multiple HTO (hashtag oligo) labelling after demultiplexing were excluded from further analysis. Moreover, to attenuate artefactual clustering, cells displaying high levels of mitochondrial transcripts and abnormal read counts were masked. Further processing included the mathematical regression of the datasets based 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. 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 database under the accession code "GSEXXXXX". Mass spectrometry data have been deposited in PRIDE database under the accession code "XXXXXX".

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).

Legend to Figures:

Figure 1. Non-genetic clonal divergence dynamics predicts transcriptomic state bottlenecks. A. HA1ER genetic anchor (F12) and two diverging subclones (1F8 and 1C9) were subjected to BdLT-Seq coupled to scRNA-Seq and traced for 4 days. UMAP plots depict transcriptome states at day 0 and 4 of tracing. Individual states are depicted in colour and numbered. **B.** Chord diagrams representing transcriptome state dynamics for cells belonging to a particular state/cluster at the beginning of tracing and their divergence after 4 days for clone F12 and subclones 1F8 and 1C9. Detected clusters are depicted (C0 to C12) and integrate the collapsed behaviour of all cells that belong to each gene expression state. Origin clusters are shown in different colours (Day 0) and chords represent the endpoint cluster association (Day 4). **C.** Chord diagrams representing individual examples of

transcriptome divergence. C2 and C10 divergence is depicted. Lineage relationship of the clones is depicted for **A**, **B** and **C**. **D.** Schemes representing transcriptome divergence conceptual framework.

Figure 2: Clonal cell systems retain a molecular memory of past cells states. A, **B** and **C**. Genetic anchors from the HA1ER (F12), HA1E (10E) and HepG2 (1A) and divergent subclones from each cell system (HA1ER – 1F8, 1C9, 2A6 and 1D12; HA1E – A7A, A9H, B2F and B5F; HepG2 – 32G, 14F, 48E and 44D) were subjected to scRNA-Seq. UMAP plots represent transcriptome states and their divergence among analysed subclones. Individual states are depicted in colour and numbered. **C.** HepG2 clonal system was subjected to either vehicle (-TGFβ) or TGFβ treatment (+TGFβ) and processed for scRNA-Seq. UMAP plots represents clonal divergence among clones and between conditions. Lineage relationship is depicted for **A**, **B** and **C**. **D.** Clonogenicity assay in HA1ER clonal system (F12, 1F8 and 1C9). Histogram represents the number of colonies per cm^2 normalized to F12 (genetic anchor). Histograms represent the mean value ± standard deviation (SD) of 3 independent experiments performed in 4-6 replicates. Statistical significance between 1F8 and 1C9 was assessed by two-tailed unpaired t-test with Welch correction. Individual data points are shown. **E.** Clonogenicity assay in HA1E clonal system (10E, A7A, A9H, B2F and B5F) upon RAS variants transduction. Histogram represents the number of colonies per cm^2 normalized to 10E (genetic anchor). Histograms represent the mean value ± standard deviation (SD) of 4 (3 for A9H) independent experiments performed in triplicate. Statistical significance between all subclones was assessed by Brown-Forsythe and Welch ANOVA followed by Dunnett's T3 multiple comparisons test. Individual data points are shown. **F.** Histograms depicting TRAIL-induced apoptosis (1 µg/ml rhTRAIL for 2h) in HA1ER clonal system (F12, 1F8, 1C9, 2A6 and 1D12) as determined by cleaved PARP staining analysed by flow cytometry. Histograms show the mean value ± standard deviation (SD) of 3 biological replicates (~20,000 cells per condition). Statistical significance was assessed by one-way ANOVA followed by Dunnett's multiple comparisons test. Individual data points are shown. **G.** HA1ER clonal system (F12, 1F8, 1C9, 2A6 and 1D12) was subjected to transwell migration assay. Histograms depict the mean area covered by migrated cells ± standard deviation (SD) of 3 independent experiments performed in triplicate/duplicate. Statistical significance of the subclones compared to F12 was assessed by Brown-Forsythe and Welch ANOVA followed by Dunnett's T3 multiple comparisons test. Individual data points are shown. **H.** HepG2 clonal system (1A, 32G, 14F, 48E and 44D) was subjected to transwell (migration) assay. Histograms depict the mean area covered by migrated cells ± standard deviation (SD) of 3 independent experiments performed in triplicate. Statistical significance of the subclones compared to 1A was assessed by ordinary one-way ANOVA followed by Dunnett's multiple comparisons test. Individual data points are shown. Clonal relationship is depicted for all panels.

Figure 3. LncRNAs are major contributors to clonal divergence. A. Diverging clones 1F8 and 1C9 (HA1ER system) were subjected to SLAM-Seq analysis. Heatmap represents differentially expressed transcripts at the steady-state level (FDR<0.01; FC>2). No differences were observed in active transcription (FDR<0.05). Scale bar is depicted. Results represent 2 independent experiments performed in triplicate. **B.** Differentially expressed transcripts between 1F8 and 1C9 were annotated based on their coding capacity. Histogram represents the fraction of coding vs non-coding transcripts. **C.** Differentially expressed non-coding RNAs were annotated (g:Profiler version e111_eg58_p18_f463989d[57], Ensembl release 105[56], LNCipedia v5.2[58]). Pie chart depicts the fraction of differentially expressed non-coding RNAs falling into known categories. **D.** Pairwise comparison of differentially expressed transcripts (>3FC up-regulated) in scRNA-Seq datasets of HA1ER clones depicted in Figure 2A and their annotation on coding vs non-coding potential. **E.** A subset of identified differentially expressed lncRNA transcripts between 1F8 vs 1C9 were validated by reverse transcription followed by qPCR (RT-qPCR). Histograms depict AL161733.1, DPP4-DT, EMP2-AS and AC008464.1 transcript levels normalized to GAPDH levels. Results represent the mean ± standard deviation (SD) of 2 independent experiments performed in technical triplicates. Statistical significance was assessed by two-tailed unpaired Student's t-test. Individual data points are depicted. **F.** A subset of differentially

expressed lncRNA transcripts between 1F8 vs 1C9 were analysed for their intracellular localization by subcellular fractionation (nucleus (nuc) vs cytoplasm (cyt)) and their levels evaluated by RT-qPCR. Histograms depict RNA levels of each fraction upon normalization to the spike-in (in-vitro transcribed partial eGFP transcript). Data for AL161733.1, DPP4-DT, EMP2-AS and AC008464.1 is depicted. MALAT1 (enriched in nucleus), GAPDH intron (enriched in nucleus) and GAPDH exon-exon junction (enriched in cytoplasm) are shown as fractionation controls. Data represents the mean ± standard deviation (SD) of 3-4 independent experiments performed in technical triplicates. Statistical significance was assessed by two-tailed unpaired Student's t-test. Individual data points are depicted.

Figure 4. IDR-containing proteins show clonal divergence. A. Clones 1F8 and 1C9 from the HA1ER system were subjected to label-free LC-MS/MS. Heatmaps depict the identified differentially expressed proteins (FDR<0.01; FC>1.5). Data was pooled from 2 independent experiments containing 4 and 5 biological replicates each. Scale bar is depicted. **B.** Differentially expressed proteins identified in A were annotated considering their intrinsically disordered region (IDR) potential (AlphaFold2[16,17]). Histogram represents the fraction of IDR vs non-IDR proteins identified as differentially expressed. **C.** Clones from A were analysed by TMT-based phospho-proteomics. Heatmap (left) represents differentially represented phospho-sites between 1F8 and 1C9 (FDR<0.05; FC>0.4). Data was obtained from 5 biological replicates. Accompanying heatmap (right) displays total protein levels for differentially phosphorylated proteins identified. Scale bars are depicted. **D.** A subset of differentially expressed IDR-containing proteins (1F8 vs 1C9) were analysed at a single-cell level by mass cytometry (CyTOF). Levels of MOT4, JIP4 and SYUA (all IDR-containing) were evaluated using specific antibodies in F12, 1F8 and 1C9. DCXR is also shown as a non-IDR diverging protein control. Scale bar is depicted. **E.** Heatmap represents the clustered expression of the antibody panel (MOT4, JIP4, SYUA and DCXR. PD-L1 and CD206 are included as a positive and negative control respectively) in collective data from all clones. The fraction of cells analysed in each cluster is depicted. Scale bar is depicted. **F.** Heatmaps represent the individual analysis for each antibody in the CyTOF panel (MOT4, JIP4, SYUA and DCXR) in each clone (F12, 1F8 and 1C9). Clustered expression levels are depicted. Scale bar is depicted.

Figure 5. Cytoplasmic lncRNAs co-localize with IDR-containing proteins and activated signalling pathways. A. HA1ER clones 1F8 and 1C9 were subjected to immunocytochemistry for JIP4, SYUA and AP1G1 coupled to RNA-FISH targeting DPP4-DT or EMP2-AS lncRNAs and analysed by Airyscan microscopy. Upper panels depict individual channels (DAPI, JIP4 and DPP4-DT) and a merged image (JIP4/DPP4-DT) whilst lower panels depict overlay images for JIP4/EMP2-AS, SYUA/DPP4-DT, SYUA/EMP2-AS, AP1G1/DPP4-DT and AP1G1/EMP2-AS. Inset is displayed for the upper panel. Scale bar 10 µm. **B.** HA1ER clone 1F8 was subjected to immunocytochemistry for activated signalling pathways coupled to RNA-FISH targeting DPP4-DT lncRNA. Upper panels depict individual channels (DAPI, pERK1/2 and DPP4-DT) and a merged image (pERK1/2/DPP4-DT) whilst the lower panels depict merged images for 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 and DPP4-DT/pIκBα. Inset is displayed for the upper panel. Scale bar 10 µm. **C.** HepG2 clone 14F was either treated with TGFβ or with vehicle (+TGFβ or -TGFβ respectively) and subjected to immunocytochemistry for either active (pSMAD2, pSMAD3 and pRAC1) or their corresponding non-active (SMAD2, SMAD3 and RAC1) TGFβ-dependent signalling proteins coupled to RNA-FISH targeting PINCR lncRNA. Upper panel depicts individual channels (DAPI, pSMAD3 and PINCR) and an overlayed image (pSMAD3/PINCR) whilst the lower panels display SMAD3/PINCR, pSMAD2/PINCR, SMAD2/PINCR, pRAC1/PINCR and RAC1/PINCR. Inset is displayed for the upper panel. Scale bar 10 µm.

Figure 6. SIP destabilization promotes transcriptome and signalling pathways re-organization. A. HA1ER clone 1F8 was subjected to antisense mediated knock-down (Gapmer) for DPP4-DT or AL161733.1 lncRNAs and processed for scRNA-Seq after 48h (24h for DPP4-DT-G2). UMAP plots depict transcriptome states in control (non-targeting control gapmer) and upon knockdown (DPP4-DT-G1,

DPP4-DT-G2, AL161733.1-G1 and AL161733.1-G2). Two alternative antisense molecules were used targeting different regions of the target transcript (G1 and G2). Individual states are depicted in colour and numbered. **B.** HA1ER clone 1F8 was subjected to antisense mediated knock-down (Gapmer) for DPP4-DT or AL161733.1 lncRNAs and further processed for Reverse Phase Protein Arrays (RPPA) after 48h (24h for DPP4-DT-G2) to evaluate signalling integrity. Two alternative antisense molecules were used targeting different regions of the target transcript (G1 and G2). Heatmaps display differentially represented signalling pathway members (≥20% increase/decrease to control) upon either DPP4-DT or AL161733.1 knockdown. Two biological replicates were processed for each condition. Scale bars are depicted **C.** Pathway analysis (Reactome) on HA1ER clone 1F8 upon DPP4-DT or AL161733.1 lncRNAs knockdown (FDR<0.01). Top 10 upregulated pathways are shown for antibody signals specifically upregulated upon DPP4-DT (DPP4-DT-specific) or AL161733.1 knock-down (AL161733.1-specific), as well as for upregulated signalling pathway members detected in both knock-downs (overlap). **D.** Venn diagrams depict the intersection between the Top 50 enriched terms (Reactome – Up- or downregulated) from DPP4-DT and AL161733.1 lncRNAs knockdown datasets for specific or overlapping signalling pathway members identified. **E.** Toy diagram depicting the link between cellular state, phenotype and SIP composition. Identified components are depicted.

References

1 Moran, K. L., Shlyakhtina, Y. & Portal, M. M. The role of non-genetic information in evolutionary frameworks. *Crit Rev Biochem Mol Biol* **56**, 255-283, doi:10.1080/10409238.2021.1908949 (2021).

2 Shlyakhtina, Y., Moran, K. L. & Portal, M. M. Genetic and Non-Genetic Mechanisms Underlying Cancer Evolution. *Cancers (Basel)* **13**, doi:10.3390/cancers13061380 (2021).

3 Yona, A. H., Frumkin, I. & Pilpel, Y. A relay race on the evolutionary adaptation spectrum. *Cell* **163**, 549-559, doi:10.1016/j.cell.2015.10.005 (2015).

4 Aissa, A. F. *et al.* Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat Commun* **12**, 1628, doi:10.1038/s41467-021-21884-z (2021).

5 Kashima, Y. *et al.* Single-Cell Analyses Reveal Diverse Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors in Lung Cancer. *Cancer Res* **81**, 4835-4848, doi:10.1158/0008-5472.CAN-20-2811 (2021).

6 Ma, P. *et al.* Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877-891 e814, doi:10.1016/j.cell.2023.01.002 (2023).

7 McNulty, R. *et al.* Probe-based bacterial single-cell RNA sequencing predicts toxin regulation. *Nat Microbiol* **8**, 934-945, doi:10.1038/s41564-023-01348-4 (2023).

8 Shlyakhtina, Y., Bloechl, B. & Portal, M. M. BdLT-Seq as a barcode decay-based method to unravel lineage-linked transcriptome plasticity. *Nat Commun* **14**, 1085, doi:10.1038/s41467-023-36744-1 (2023).

9 Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464-468, doi:10.1038/22780 (1999).

10 Rangarajan, A., Hong, S. J., Gifford, A. & Weinberg, R. A. Species- and cell type-specific requirements for cellular transformation. *Cancer Cell* **6**, 171-183, doi:10.1016/j.ccr.2004.07.009 (2004).

11 Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**, 1198-1204, doi:10.1038/nmeth.4435 (2017).

12 Matsushima, W. *et al.* Sequencing cell-type-specific transcriptomes with SLAM-ITseq. *Nat Protoc* **14**, 2261-2278, doi:10.1038/s41596-019-0179-x (2019).

13 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791, doi:10.1093/bioinformatics/btt468 (2013).
14 Neumann, T. *et al.* Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* **20**, 258, doi:10.1186/s12859-019-2849-7 (2019).
15 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).
16 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).
17 Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**, D368-D375, doi:10.1093/nar/gkad1011 (2024).
18 Macleod, K. G., Serrels, B. & Carragher, N. O. Reverse Phase Protein Arrays and Drug Discovery. *Methods Mol Biol* **1647**, 153-169, doi:10.1007/978-1-4939-7201-2_10 (2017).
19 Morawska, L. P., Hernandez-Valdes, J. A. & Kuipers, O. P. Diversity of bet-hedging strategies in microbial communities-Recent cases and insights. *WIREs Mech Dis* **14**, e1544, doi:10.1002/wsbm.1544 (2022).
20 O'Dea, R. E., Noble, D. W. A., Johnson, S. L., Hesselson, D. & Nakagawa, S. The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environ Epigenet* **2**, dvv014, doi:10.1093/eep/dvv014 (2016).
21 Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet* **10**, 336-342, doi:10.1038/nrg2556 (2009).
22 Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-547, doi:10.1038/nature06965 (2008).
23 Charlebois, D. A. Effect and evolution of gene expression noise on the fitness landscape. *Phys Rev E Stat Nonlin Soft Matter Phys* **92**, 022713, doi:10.1103/PhysRevE.92.022713 (2015).
24 Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451-464, doi:10.1038/nrg1615 (2005).
25 Hangauer, M. J. *et al.* Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature* **551**, 247-250, doi:10.1038/nature24297 (2017).
26 Knoechel, B. *et al.* An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. *Nat Genet* **46**, 364-370, doi:10.1038/ng.2913 (2014).
27 Kurata, T. *et al.* Effect of re-treatment with gefitinib ('Iressa', ZD1839) after acquisition of resistance. *Ann Oncol* **15**, 173-174, doi:10.1093/annonc/mdh006 (2004).
28 Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69-80, doi:10.1016/j.cell.2010.02.027 (2010).
29 Yano, S. *et al.* Retreatment of lung adenocarcinoma patients with gefitinib who had experienced favorable results from their initial treatment with this selective epidermal growth factor receptor inhibitor: a report of three cases. *Oncol Res* **15**, 107-111 (2005).
30 Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* **33**, 503-509, doi:10.1038/nbt.3209 (2015).
31 Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).
32 Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods* **9**, 743-748, doi:10.1038/nmeth.2069 (2012).
33 Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* **11**, 360-361, doi:10.1038/nmeth.2892 (2014).

34 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).
35 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-1172, doi:10.1038/nbt.3383 (2015).
36 Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224, doi:10.1038/s41586-018-0744-4 (2018).
37 Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* **24**, 430-447, doi:10.1038/s41580-022-00566-8 (2023).
38 Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96-118, doi:10.1038/s41580-020-00315-9 (2021).
39 Barbieri, E. *et al.* Rapid and Scalable Profiling of Nascent RNA with fastGRO. *Cell Rep* **33**, 108373, doi:10.1016/j.celrep.2020.108373 (2020).
40 Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* **13**, 246-259, doi:10.1038/nrg3160 (2012).
41 Yamada, T. & Akimitsu, N. Contributions of regulated transcription and mRNA decay to the dynamics of gene expression. *Wiley Interdiscip Rev RNA* **10**, e1508, doi:10.1002/wrna.1508 (2019).
42 Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18**, 31-42, doi:10.1038/nrm.2016.132 (2017).
43 Shlyakhtina, Y., Bloechl, B., Moran, K. L. & Portal, M. M. Protocol to study the inheritance and propagation of non-genetically encoded states using barcode decay lineage tracing. *STAR Protoc* **5**, 102809, doi:10.1016/j.xpro.2023.102809 (2024).
44 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
45 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868, doi:10.1038/nmeth.4380 (2017).
46 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12, doi:https://doi.org/10.14806/ej.17.1.200 (2011).
47 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).
48 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
49 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).
50 Knaus, B. J. & Grunwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* **17**, 44-53, doi:10.1111/1755-0998.12549 (2017).
51 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
52 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
53 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
54 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).
55 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>.
56 Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res* **49**, D884-D891, doi:10.1093/nar/gkaa942 (2021).

57 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, doi:10.1093/nar/gkad347 (2023).

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

59 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

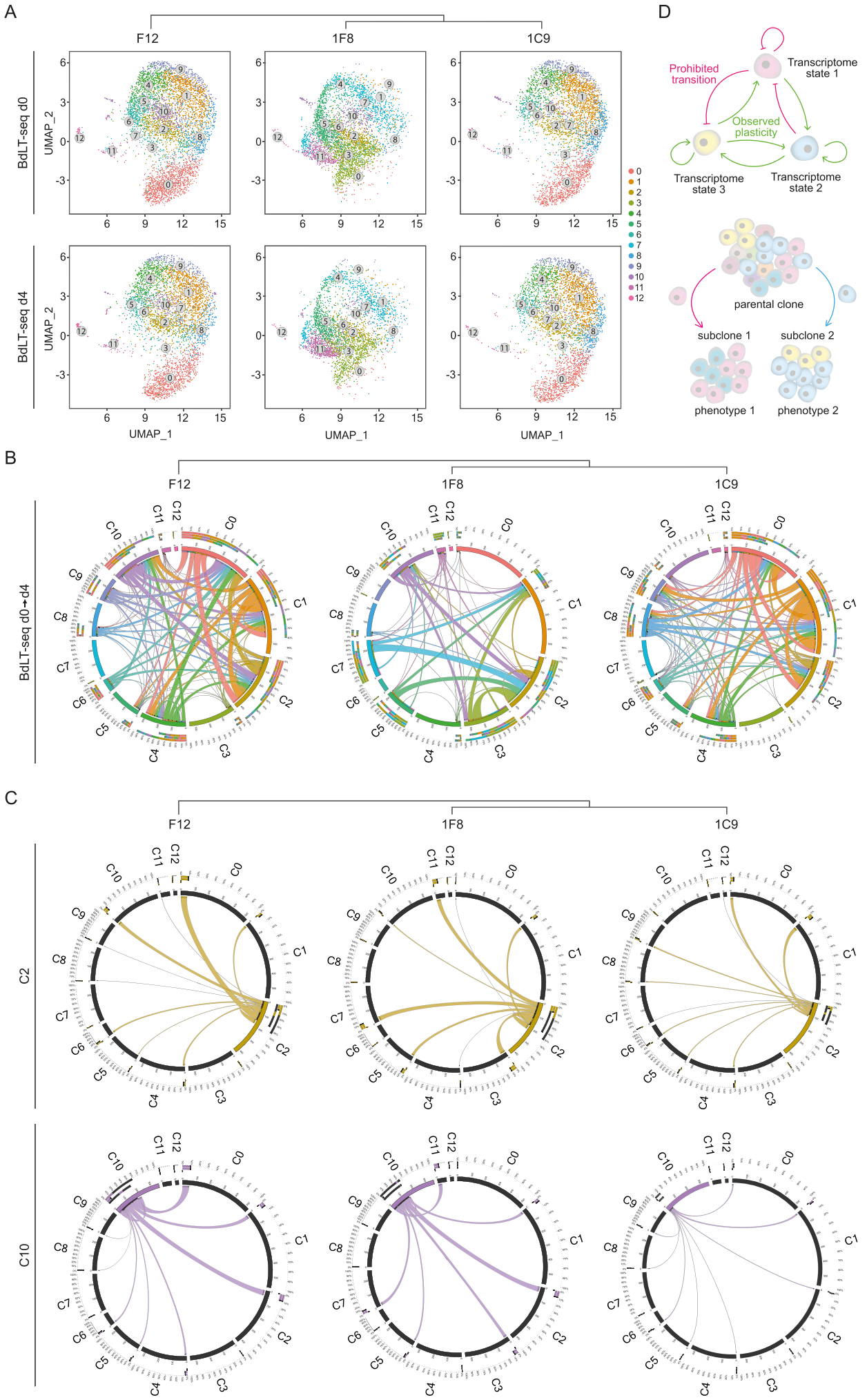

A
F12
1F8
1C9
BdLT-seq d0
BdLT-seq d4
UMAP_2
UMAP_1
0
1
2
3
4
5
6
7
8
9
10
11
12
D
Prohibited transition
Transcriptome state 1
Observed plasticity
Transcriptome state 3
Transcriptome state 2
parental clone
subclone 1
subclone 2
phenotype 1
phenotype 2
B
BdLT-seq d0→d4
C
C2
C10
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12

Figure 2

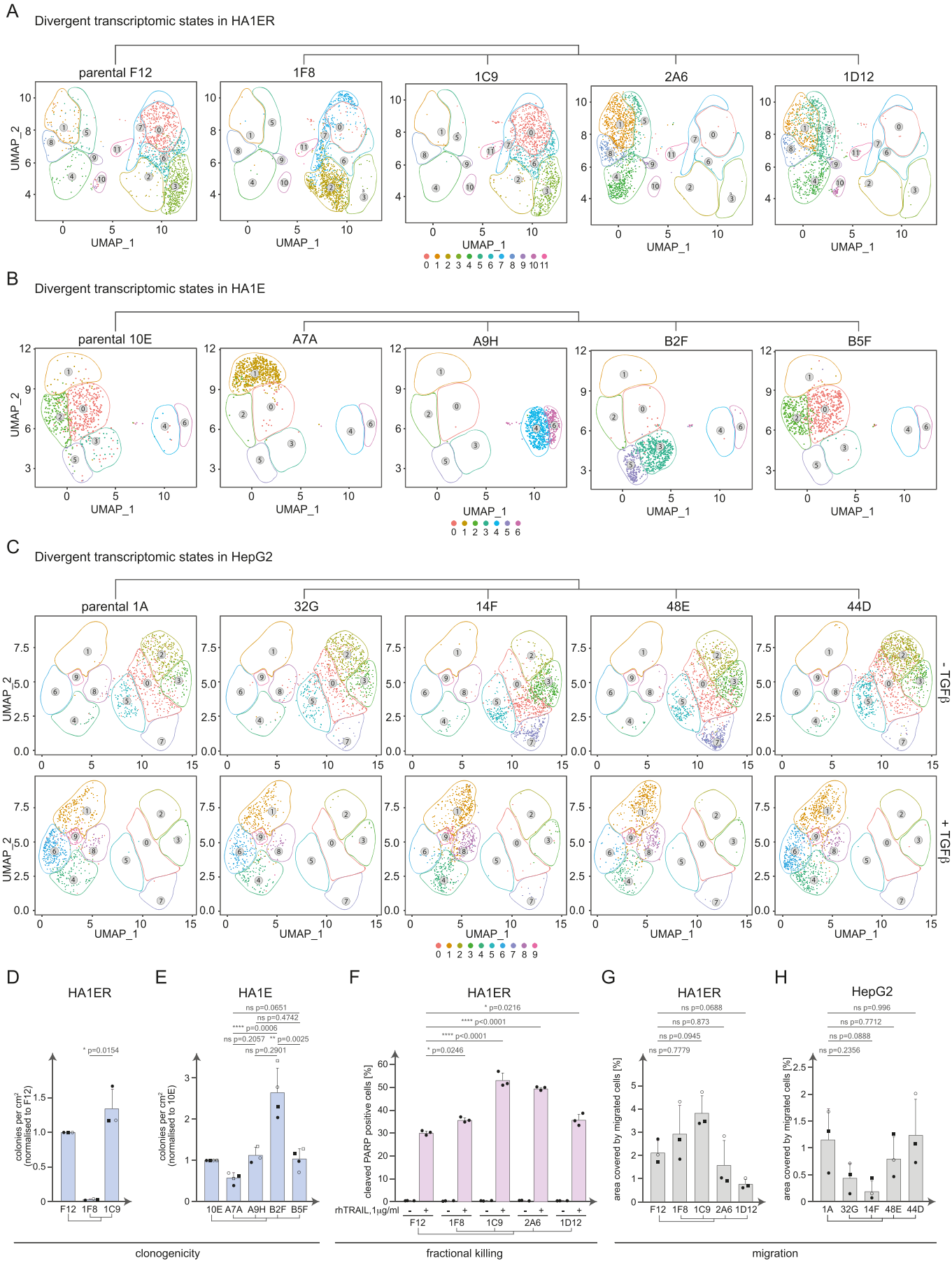

A
Divergent transcriptomic states in HA1ER
parental F12
1F8
1C9
2A6
1D12
UMAP_2
UMAP_1
0 1 2 3 4 5 6 7 8 9 10 11
B
Divergent transcriptomic states in HA1E
parental 10E
A7A
A9H
B2F
B5F
0 1 2 3 4 5 6
C
Divergent transcriptomic states in HepG2
parental 1A
32G
14F
48E
44D
- TGFβ
+ TGFβ
0 1 2 3 4 5 6 7 8 9
D
HA1ER
colonies per cm² (normalised to F12)
* p=0.0154
F12 1F8 1C9
E
HA1E
colonies per cm² (normalised to 10E)
ns p=0.0651
ns p=0.4742
**** p=0.0006
ns p=0.2057
** p=0.0025
ns p=0.2901
10E A7A A9H B2F B5F
F
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
G
HA1ER
area covered by migrated cells [%]
ns p=0.0688
ns p=0.873
ns p=0.0945
ns p=0.7779
F12 1F8 1C9 2A6 1D12
H
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
clonogenicity
fractional killing
migration

Figure 3

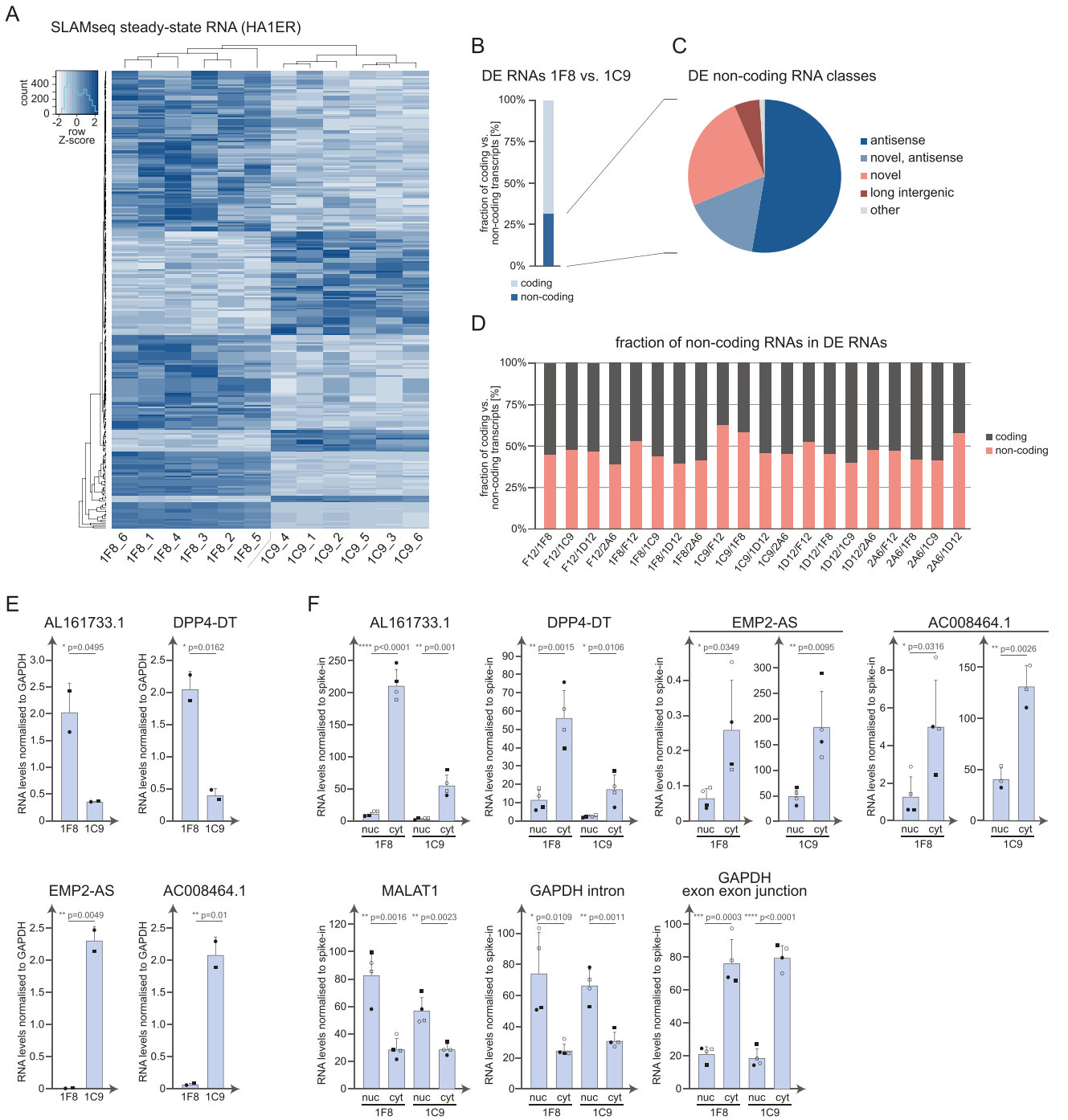

A
SLAMseq steady-state RNA (HA1ER)
count
row Z-score
1F8_6 1F8_1 1F8_4 1F8_3 1F8_2 1F8_5 1C9_4 1C9_1 1C9_2 1C9_5 1C9_3 1C9_6
B
DE RNAs 1F8 vs. 1C9
fraction of coding vs. non-coding transcripts [%]
coding
non-coding
C
DE non-coding RNA classes
antisense
novel, antisense
novel
long intergenic
other
D
fraction of non-coding RNAs in DE RNAs
fraction of coding vs. non-coding transcripts [%]
coding
non-coding
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
E
AL161733.1
DPP4-DT
RNA levels normalised to GAPDH
* p=0.0495
* p=0.0162
1F8 1C9
F
AL161733.1
DPP4-DT
EMP2-AS
AC008464.1
RNA levels normalised to spike-in
**** p<0.0001
** p=0.001
** p=0.0015
* p=0.0106
* p=0.0349
** p=0.0095
* p=0.0316
** p=0.0026
nuc cyt
1F8 1C9
EMP2-AS
AC008464.1
** p=0.0049
** p=0.01
MALAT1
GAPDH intron
GAPDH exon exon junction
** p=0.0016
** p=0.0023
* p=0.0109
** p=0.0011
*** p=0.0003
**** p<0.0001

Figure 4

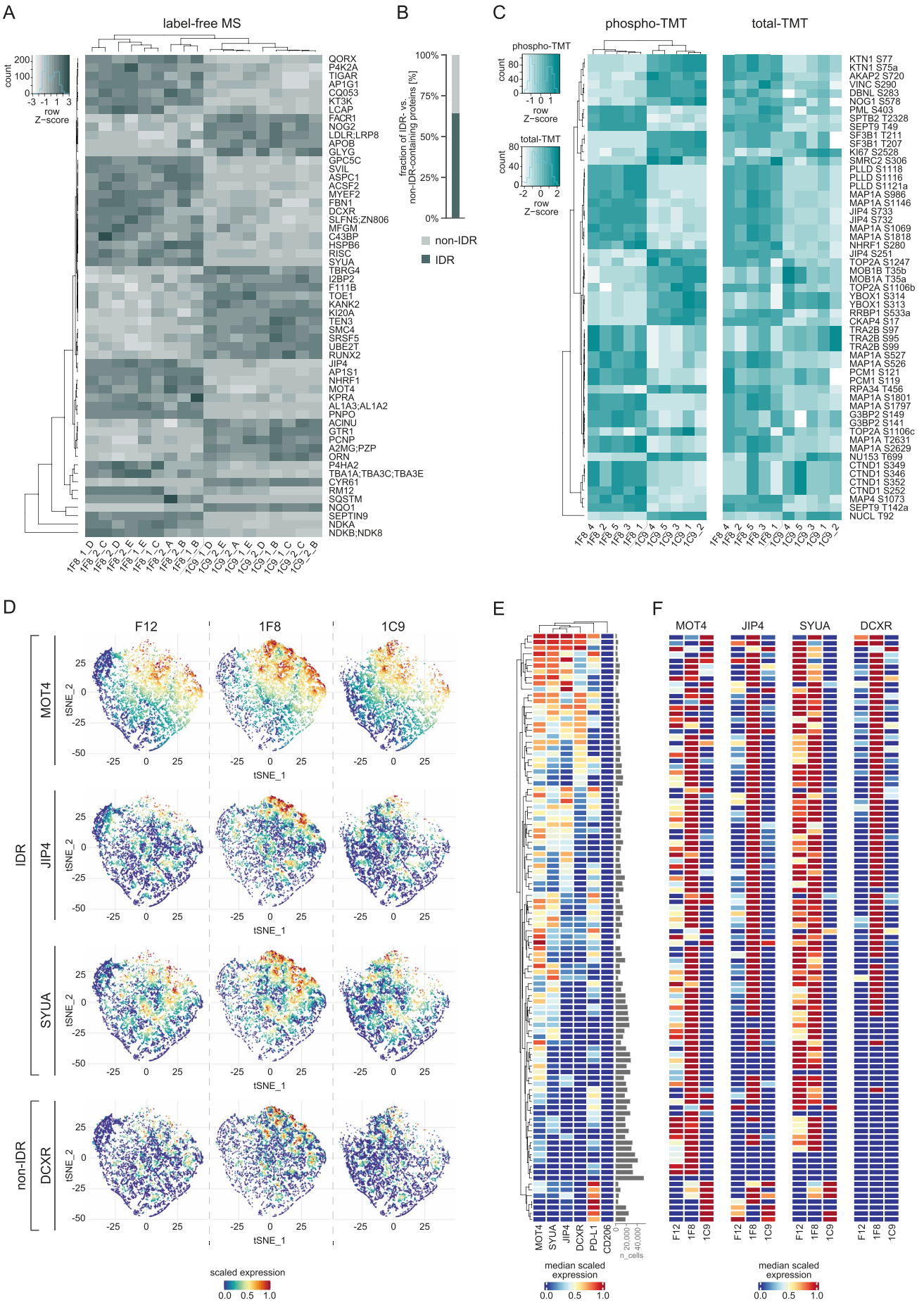

A
label-free MS
count
row Z-score
QORX
P4K2A
TIGAR
AP1G1
CQ053
KT3K
LCAP
FACR1
NOG2
LDLR;LRP8
APOB
GLYG
GPC5C
SVIL
ASPC1
ACSF2
MYEF2
FBN1
DCXR
SLFN5;ZN806
MFGM
C43BP
HSPB6
RISC
SYUA
TBRG4
I2BP2
F111B
TOE1
KANK2
KI20A
TEN3
SMC4
SRSF5
UBE2T
RUNX2
JIP4
AP1S1
NHRF1
MOT4
KPRA
AL1A3;AL1A2
PNPO
ACINU
GTR1
PCNP
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 S75a
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 S1121a
MAP1A S986
MAP1A S1146
JIP4 S733
JIP4 S732
MAP1A S1069
MAP1A S1818
NHRF1 S280
JIP4 S251
TOP2A S1247
MOB1B T35b
MOB1A T35a
TOP2A S1106b
YBOX1 S314
YBOX1 S313
RRBP1 S533a
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 S1106c
MAP1A T2631
MAP1A S2629
NU153 T699
CTND1 S349
CTND1 S346
CTND1 S352
CTND1 S252
MAP4 S1073
SEPT9 T142a
NUCL T92
D
F12
1F8
1C9
MOT4
JIP4
SYUA
DCXR
IDR
non-IDR
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

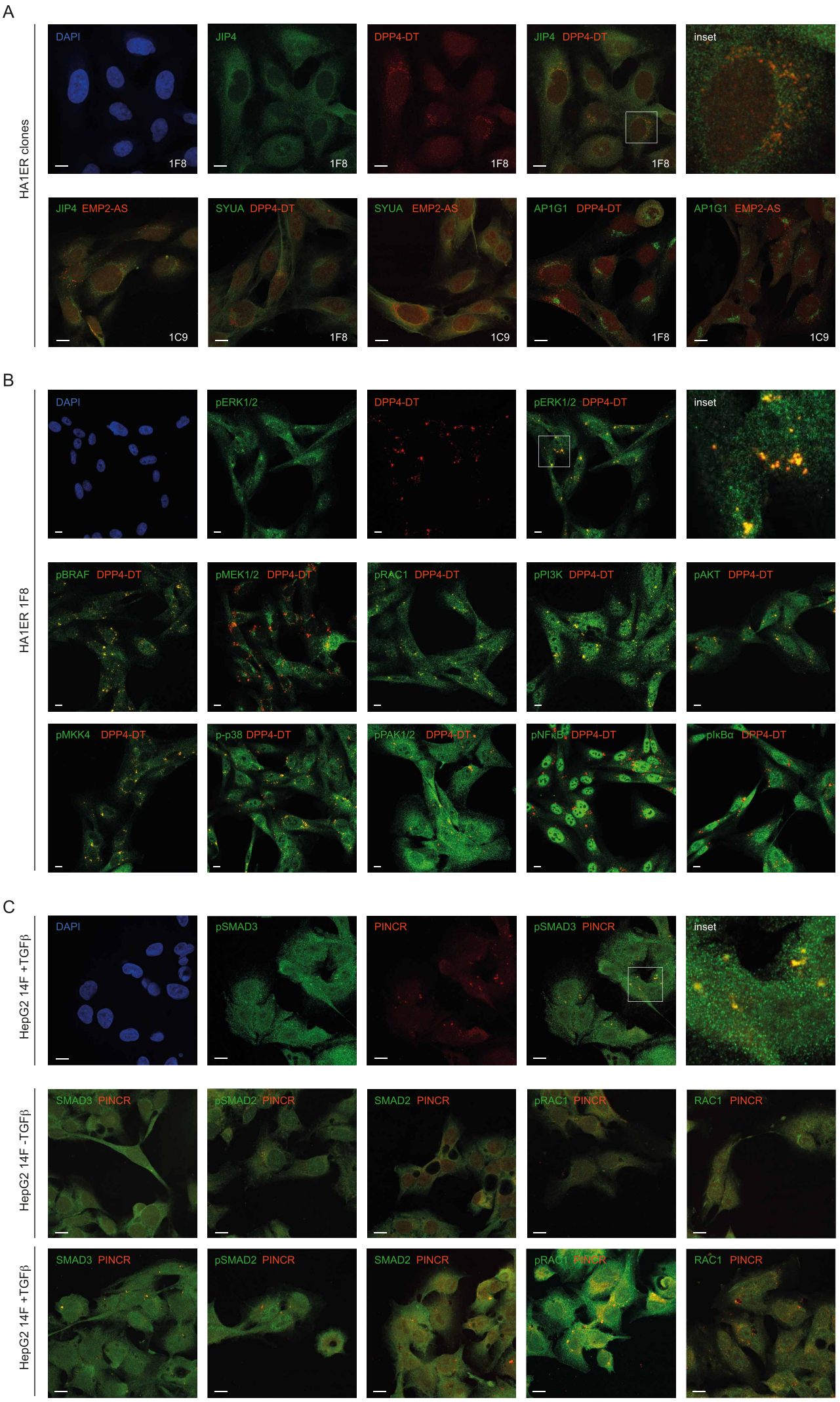
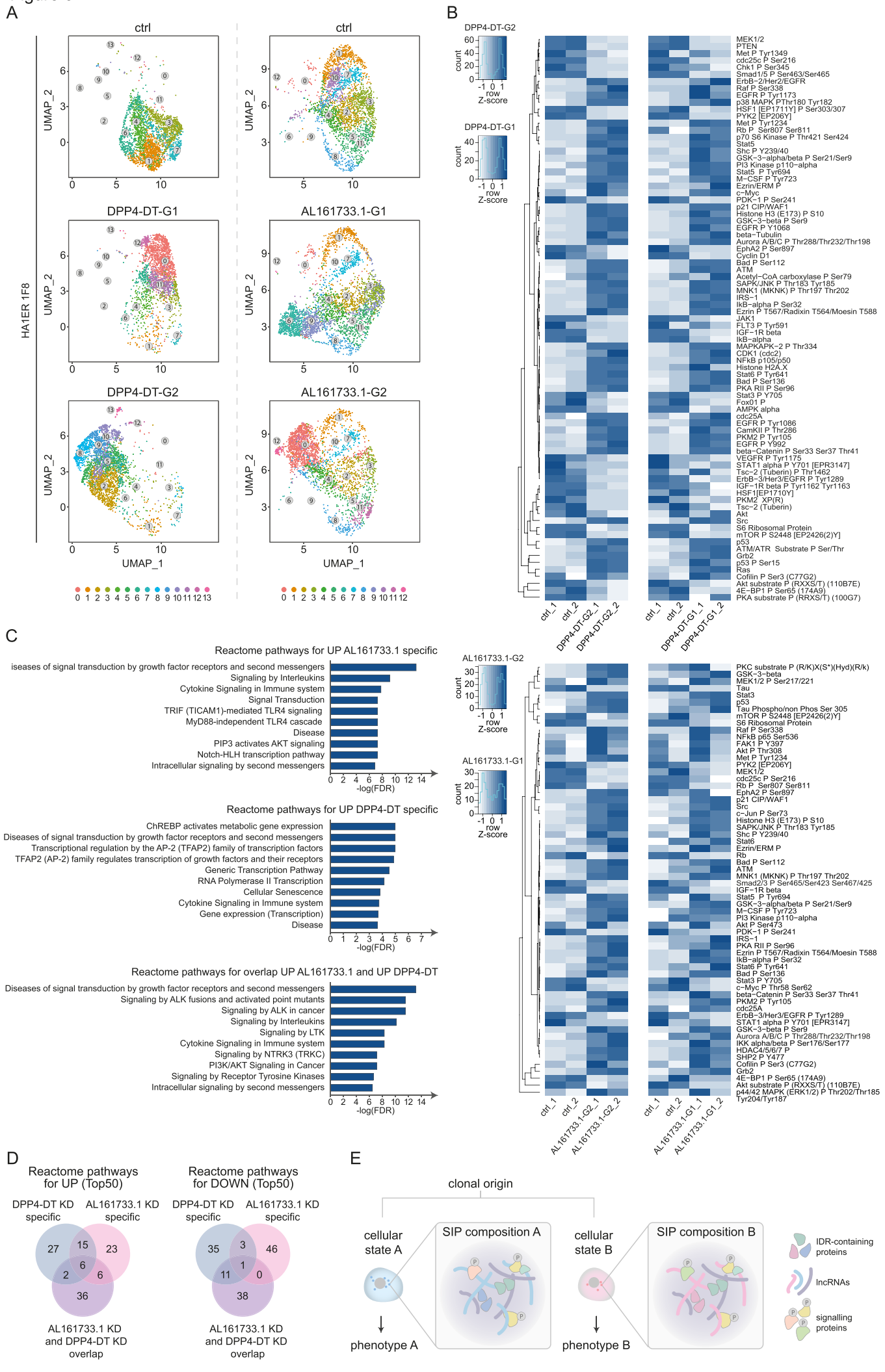

A
HA1ER clones
DAPI
1F8
JIP4
1F8
DPP4-DT
1F8
JIP4 DPP4-DT
1F8
inset
JIP4 EMP2-AS
1C9
SYUA DPP4-DT
1F8
SYUA EMP2-AS
1C9
AP1G1 DPP4-DT
1F8
AP1G1 EMP2-AS
1C9
B
HA1ER 1F8
DAPI
pERK1/2
DPP4-DT
pERK1/2 DPP4-DT
inset
pBRAF DPP4-DT
pMEK1/2 DPP4-DT
pRAC1 DPP4-DT
pPI3K DPP4-DT
pAKT DPP4-DT
pMKK4 DPP4-DT
p-p38 DPP4-DT
pPAK1/2 DPP4-DT
pNFκB DPP4-DT
pIκBα DPP4-DT
C
HepG2 14F +TGFβ
DAPI
pSMAD3
PINCR
pSMAD3 PINCR
inset
HepG2 14F -TGFβ
SMAD3 PINCR
pSMAD2 PINCR
SMAD2 PINCR
pRAC1 PINCR
RAC1 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-G2
DPP4-DT-G1
AL161733.1-G2
AL161733.1-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
AL161733.1-G2_1
AL161733.1-G2_2
AL161733.1-G1_1
AL161733.1-G1_2
C
Reactome pathways for UP AL161733.1 specific
iseases of signal transduction by growth factor receptors and second messengers
Signaling by Interleukins
Cytokine Signaling in Immune system
Signal Transduction
TRIF (TICAM1)-mediated TLR4 signaling
MyD88-independent TLR4 cascade
Disease
PIP3 activates AKT signaling
Notch-HLH transcription pathway
Intracellular signaling by second messengers
-log(FDR)
Reactome pathways for UP DPP4-DT specific
ChREBP activates metabolic gene expression
Diseases of signal transduction by growth factor receptors and second messengers
Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors
Generic Transcription Pathway
RNA Polymerase II Transcription
Cellular Senescence
Cytokine Signaling in Immune system
Gene expression (Transcription)
Disease
-log(FDR)
Reactome pathways for overlap UP AL161733.1 and UP DPP4-DT
Diseases of signal transduction by growth factor receptors and second messengers
Signaling by ALK fusions and activated point mutants
Signaling by ALK in cancer
Signaling by Interleukins
Signaling by LTK
Cytokine Signaling in Immune system
Signaling by NTRK3 (TRKC)
PI3K/AKT Signaling in Cancer
Signaling by Receptor Tyrosine Kinases
Intracellular signaling by second messengers
-log(FDR)
D
Reactome pathways for UP (Top50)
DPP4-DT KD specific
AL161733.1 KD specific
27
15
23
6
2
6
36
AL161733.1 KD and DPP4-DT KD overlap
Reactome pathways for DOWN (Top50)
DPP4-DT KD specific
AL161733.1 KD specific
35
3
46
1
11
0
38
AL161733.1 KD and DPP4-DT KD overlap
E
clonal origin
cellular state A
SIP composition A
phenotype A
cellular state B
SIP composition B
phenotype B
IDR-containing proteins
lncRNAs
signalling proteins