Differential expression of *Nanog* 3UTR and coding region mRNA components defines an embryonic stem cell niche

Ze Yang[1], Shaoyi Ji[1], Leonardi Gozali[1], Mujib Ullah[1], Sophie Parsa[1], Barry Behr[2], and Mary A. Hynes*[1]

Stanford University Department of Biology[1] and Department of Obstetrics and Gynecology[2] and the Stanford IVF laboratory, Stanford University, Stanford California 94305 USA

Correspondence: mhynes@stanford.edu

Summary

The known molecular heterogeneity among pluripotent embryonic stem cells (ESCs) in culture may be maladaptive or a beneficial determinant of normal biology. In other systems such as hair and intestine, normal and dynamic heterogeneity among stem cells is necessary for balancing stem cell division and quiescence in order to maintain the stem cell pool. In ESCs, *Sox2, Oct4* and *Nanog* (the *SON* genes) critically regulate pluripotency but their role in stem cell heterogeneity or pool maintenance is not clear. We previously reported an additional level of gene expression, the unexpected differential expression of mRNA coding region (CDS) and cognate 3′untranslated region (3UTR) components, often revealing novel developmental patterns. Here we show dramatic, patterned, concentric, differential expression of pluripotent gene 3UTR and CDS mRNA components in both undifferentiated ESCs in culture and in developing human and mouse embryos. This differential expression reveals robust pluripotent ESC heterogeneity that provides the underpinnings of a previously unrecognized ESC cell niche structure. A hallmark of the niche is the presence of cells expressing high levels of *Nanog* 3UTR but only low or no *Nanog* CDS at structure perimeters, We used a novel bioinformatic approach combining the 3UTR to CDS ratio and gene expression, and identified unique ESC clusters. High *Nanog* 3UTR cells have lower cell cycle activity and higher quiescence than other ESCs and express a distinct complement of cellular histone marks. These quiescent perimeter cells exchange with interior domains, opening the possibility that they act as cell replenishment reservoirs in the niche or the embryo. Functionally, the *Nanog* 3UTR, independent of its CDS, is critical, as its selective deletion decreases ESC cell division while increasing trophectoderm markers. Lastly, self-organization of ESCs according to pluripotent gene differential 3UTR and CDS expression is modified by activin signaling, indicating extrinsic modulation of the niche structure. Pluripotent embryonic stem cells, actively display regulated, dynamic differential expression of *SON* gene mRNA components, allowing the formation of a heterogenous, ordered, stem cell niche bordered by quiescent hub stem cells, thus illuminating the role of *SON* genes in pluripotent embryonic stem cell pool maintenance. More generally this suggests that differential mRNA component expression may widely define cell types that exist in non-permanent, different states in either development or disease.

Keywords: Embryonic stem cell, niche, quiescent, organoid, hESC/mESC, 3UTR, *Nanog*, *Oct4*, Activin, heterogeneity, histone marks

Introduction

Due to their unlimited differentiation capacity embryonic stem cells (ESCs) have traditionally been considered to be homogeneous cells (Keller, 2005) however, heterogeneity, even among such undifferentiated cells, has recently been observed by marker expression (Bauwens et al., 2008; Peerani et al., 2007; Weinberger et al., 2016; Yang et al., 2021) and single cell transcriptomics (Kolodziejczyk et al., 2015). Such heterogeneity may be maladaptive, arising from errors acquired during long passaging, or may underly necessary and important biological functions. In other tissues stem cells show ample heterogeneity, with these differences allowing the formation stem cell niches, which act to regulate the balance between proliferation and quiescence, to ensure maintenance of a stem cell pool (Krieger and Simons, 2015). How and if embryonic stem cells form niches is not clear, and the absence of clear systematic heterogeneity would preclude formation. Here we piece together these parts of the puzzle showing that embryonic stem cells exist in a state of non-permanent and systematic heterogeneity, allowing them to spontaneously form robust niche structures while remaining pluripotent.

The key pluripotent genes *Sox2*, *Oct4* and *Nanog* (the "SON" genes) sit at the apex of pluripotency regulation for both mouse and human ESCs (Chambers and Tomlinson, 2009; Festuccia et al., 2013; Krieger and Simons, 2015; Saunders et al., 2013), however, while various molecular interactions among their protein products have been documented (Boyer et al., 2005; Chambers and Tomlinson, 2009; Li and Belmonte, 2017; Li, 2017; Loh et al., 2006; Rodda et al., 2005), how they contribute to ESC heterogeneity or early patterning events remains uncertain.

We and others recently discovered an additional level of mRNA regulation, and therefore gene expression: the unexpected but widespread differential expression of the 3' untranslated region (3UTR) and cognate coding region (CDS) mRNA components. This occurs across the organism in non-random and tightly regulated patterns, with prominent differential mRNA component expression during development and in transitional events (Ji et al., 2021; Kocabas et al., 2015). How these different transcripts arise molecularly is incompletely understood (see Discussion), but their existence led us to hypothesize that an additional dimension for SON gene dynamics, and perhaps ESC heterogeneity and biology, might lie in the spatial relationship of differentially expressed 3UTR and CDS mRNA components for SON genes, and functions of their 3UTRs operating independently of their CDS, or vice versa.

Here we show that ESCs grown on patterned substrates as two dimensional organoids self-organize in a remarkable fashion into patterns of SON gene mRNA component expression, revealing a robust, concentric, ESC niche. A similar pattern of expression is observed in vivo as well, in both mouse and human blastocysts. A prominent feature is that cells at the perimeter in both organoids and embryos express high levels of *Nanog* 3UTR with little or no *Nanog* CDS. Using a novel bioinformatic method combining gene 3UTR/CDS ratio with expression level, we show that cells with high *Nanog* 3UTR expression are more quiescent, with and lower cell cycle activity. They also express a unique panel of chromatin marks. When the organoid niche is divided into zones based on SON gene 3UTR/CDS ratio, each zone uniquely expresses a panel of active or repressive histone marks suggestive of (from outside to inside) a quiescent layer, a stem cell layer and a pre-differentiation state. Functionally the *Nanog* 3UTR is critical, as its

deletion while retaining the CDS results in compromised colony formation, decreased cell division and a shift from pluripotency to a trophectoderm lineage. Finally cells within the niche move dynamically, and the niche pattern can be modified by activin signaling molecules, suggesting that self-sorting of pluripotent ESCs according to SON mRNA component expression is modulated and/or refined by extrinsic signaling.

Thus differential pluripotent gene mRNA component expression reveals robust pluripotent ESC heterogeneity, which then provides the underpinnings of a previously unrecognized ESC cell niche structure. By "switching" expression of their 3UTR and CDS components for the critical pluripotent genes, ESCs change their identities, transcriptomes and epigenetic markers and can self-sort. Importantly this is neither a differentiated state, nor is it permanent. ESCs can, and do transition between the hub cell semi-quiescent state and actively dividing progenitor state, with the isolated Nanog 3UTR playing a critical role in stem cell division and colony maintenance.

Results

Mouse and human embryonic stem cells self-organize according to pluripotent gene mRNA 3UTR/CDS ratio.

We first examined the expression of 3UTR and CDS mRNA components for each SON gene (*Sox2, Oct4,* and *Nanog)* in untreated embryonic stem cells (ESCs), by *in situ* hybridization using RNA specific probes (Table S1). In these studies, unless otherwise noted, the 3UTR is labeled red, the CDS green, and cells expressing both are yellow. Untreated pluripotent mouse or human ESCs grown as 2D organoids on patterned substrates (circles of different diameters) for 72 hours showed unexpected and dramatic sorting and self-organization, as evidenced by the 3UTR/CDS mRNA ratio for each SON gene, with particularly marked patterning of the *Nanog* 3UTR and CDS (Fig. 1A-C; for controls see (Ji *et al.*, 2021; Kocabas *et al.*, 2015). Two points are striking: first, each gene shows differential expression of its 3UTR and CDS components; and second, differential expression occurs in concentric circles (Fig. 1A-D). Perimeter cells show a distinct high-level expression of *Nanog* 3UTR compared to CDS, highlighting again that some 3UTR transcripts exist separate from their CDS. From the *Nanog* 3UTR/CDS pattern, we delimit four concentric zones (Fig. 1D; Table 1; zone boundaries were more diffuse for mESCs than for hESCs). All zone-1 cells show high nuclear *Nanog* 3UTR with no or low *Nanog* (nuclear) CDS. This zone appears to have two sub-zones; outer zone-1 is a single-cell layer oriented orthogonally to other cells (Figure S1A,A'), with no *Oct4* or *Sox2* mRNAs (zones are further defined in Fig. 4 and Figure S4). Inner zone-1 cells may express very low levels of nuclear *Nanog* CDS, and low levels of *Oct4* and *Sox2* mRNA components. In zone-2, all *Nanog*, *Oct4* and *Sox2* 3UTR and CDS mRNA components are robustly expressed, with the *Nanog* CDS and all *Oct4* and *Sox2* mRNA components cytoplasmic, and each showing a higher 3UTR/CDS ratio in outer, compared to inner, cells (Fig.1B,C). In zone-3, cells express detectable *Nanog* and *Sox2* CDS, but not *Nanog* 3UTR nor *Oct4* mRNA components. Zone-4, cells are largely devoid of SON mRNA components, except for a peppering of cells expressing high *Nanog* 3UTR, primarily in the largest organoids (Fig. 5A).

Our finding of high levels of *Nanog* 3UTR in perimeter cells, with very low or no CDS, in both mouse and human ESCs, prompted us focus principally on *Nanog* and the role of the isolated *Nanog* 3UTR.

The 2D organoids were formed on patterned (Cytoo^{r}) chips, each of which contains a mix of circles of different diameters, ranging from 80 to 1000 μm. Regardless of organoid size, we observed that the width of the *Nanog* 3UTR plus CDS pattern is about ~100 μm, filling the entire structure of the smallest circles (Fig. 1E), but ringing only the perimeter in large circles. In cells grown in bulk culture, dispersed cells show high *Nanog* 3UTR (Fig. 1F), but, interestingly, as they become aggregated, the cells are significantly more likely to co-express *Nanog* 3UTR and CDS. In addition, most bulk culture cells appeared to express one or both of some SON mRNA element, suggesting that zone-4, in which cells do not express SON mRNA components, is not prominent in cells that are migrating randomly in a plate (not shown).

These results show that pluripotent hESC and mESCs are heterogenous, and that *Nanog* and other SON gene mRNA components show stereotyped differential expression in pluripotent ESC organoids (Table 1).

How is this pattern generated? We note first that *Nanog* 3UTR border cells can be detected in the absence of high *Nanog* CDS expressing cells but not vice versa, suggesting that border cells arise first (Fig. 2A, Figure S1B,C; arrows). Secondly, we examined whether the *Nanog* mRNA 3UTR/CDS ratio is predetermined and constant, by examining the ratio of *Nanog* 3UTR cells to total cells across circles. A constant ratio would suggest predetermination. Instead, regardless of organoid size (all confluent on a single chip), high *Nanog* 3UTR cells consistently mark the perimeters, comprising a much greater percentage of all cells in small circles compared to larger circles (Fig. 2B). This suggest that ESCs determine their SON 3UTR/CDS ratio on a cell-by-cell basis as circumstances dictate.

We next analyzed SON protein localization in organoids. Surprisingly, there are high levels of Nanog protein in zone-1 (in cells with high *Nanog* 3UTR but either low nuclear, or no, CDS) as well as zone-2 (in *Nanog* 3UTR/CDS co-expressing cells) (Fig. 2D and Figure S2B,C-C"; and see discussion), but not zone-3, where high *Nanog* CDS cells reside (Fig. 2D; white arrow). Oct4 protein is localized to zone-2, in both high *Oct4* 3UTR and high CDS cells (Fig. 2E). The nuclear localization of cognate proteins in high 3UTR-only cells is consistent with our previous findings for some transcription factors (Ji *et al.*, 2021), where we speculated that 3UTRs might sometimes act to stabilize or sequester cognate proteins in the nucleus (see further below). Thus, of the SON mRNA components, only the *Nanog* 3UTR is prominently expressed in zone-1, while all are expressed in zone-2, highlighting a considerable transition in cell phenotype between these two zones. These distributions are summarized in Table 1.

Table 1. Zonal expression of the SON genes

Zone/character	***Nanog***	***Oct4***	***Sox2***
Zone-1 outer	High 3UTR, no CDS	No	No
Zone-1 inner	High 3UTR, low nuclear or no CDS	Low or no	Low or no
Zone-2	Hi 3UTR and CDS	High 3UTR and CDS	High 3UTR and CDS
Zone-3	Moderate CDS, no 3UTR	No	Moderate CDS
Zone-4	Sporadic high 3UTR	No or very low	Sporadic

Differential expression of SON cognate 3UTR and CDS mRNA components in mouse and human blastocysts *in vivo*.

The *in vivo* relevance of SON differential 3UTR/CDS expression was assessed by examining developing mouse and human embryos. Strikingly, as in ESCs cultures, for both mouse and human harvested blastocysts we observed differential 3UTR/CDS expression for both *Nanog* and *Oct4* (Fig. 2G-L; *Sox2* expression is more overlapping, Figure S2D). Dramatically, for both *Oct4* and *Nanog*, high 3UTR cells populate blastocyst perimeters, while high CDS cells colonize the Oct4 protein positive, interior cells of the inner cell mass (ICM) (Figure S2F-F"'). This expression pattern is highly prominent around the time of the first cell fate decision between trophectoderm and embryo proper (at ~24-36 cells) (Rossant, 2007) (Fig. 2G-L). Differential SON 3UTR/CDS expression is also observed on a cell-by-cell basis in younger embryos, but with less obvious spatial patterning (Figure S2E). This *in vivo* pattern is remarkably reminiscent of the pattern in small 2D organoids that are less 250 um in diameter, which, intriguingly, is similar to the diameter of the *in vivo* ICM (~100-200 uM).

<u>Mouse and human *Nanog* high 3UTR expressing cells define distinct cell populations</u>

Our finding that cells expressing high *Nanog* 3UTR and low or no CDS reliably form the border in 2D organoids and the embryo, raises the possibility that the *Nanog* 3UTR, expressed in the absence of the CDS, has as a distinct biological role. As a first step toward exploring this, we analyzed published mouse (Kolodziejczyk *et al*., 2015) and human (Chu et al., 2016) single cell (sc) RNA seq data (Fig. 3), clustering with tSNE. Superimposing the 3UTR/CDS ratios for various genes (see Methods), starting with *Nanog*, we found that for both mouse (mESC) and human ESCs (hESC), cells expressing a high *Nanog* 3UTR/CDS ratio clustered distinctly (Fig. 3B,F). In hESCs, undifferentiated cells (containing cells from both H1 and H9 lines), form a large, distinct group comprised of two clusters, termed 2 and 0 (Fig. 3A). Clusters 2 and 0 are not distinguished by *Nanog* expression levels (Fig. 3B; left), but, remarkably are distinguished by differential high *Nanog* 3UTR vs high *Nanog* CDS expression (Fig. 3B; right, Table 2), with cluster 2 cells expressing high *Nanog* 3UTR, and cluster 0 cells more likely to express high *Nanog* CDS. Both clusters contain both H1 and H9 cells (Fig. 3A). Thus, high *Nanog* 3UTR cells, even from different ESC lines, are more transcriptionally similar to each other than they are to high *Nanog* CDS ESCs.

Mouse ES cells clustered similarly (Fig. 3E-G). Here, cells were grown in different media types; of cells grown in 2i or a2i media (both similar to the medium used in our experiments), high *Nanog* 3UTR cells congregate in cluster 4 (Fig. 3F; left). Using differential expression to probe for genes that might influence, or be influenced by, isolated *Nanog* 3UTR expression, we found that compared to other cells grown in 2i media, cluster 4 high *Nanog* 3UTR cells are enriched for transcripts involved in cell cycle and chromosome organization (Fig. 3G and Figure S3C, and see below). In contrast cluster 0 (high *Nanog* CDS) is enriched for cell development, differentiation and signal transduction (Fig. 3G and Figure S3C; Table 2). Like mESCs, high *Nanog* 3UTR hESCs are similarly enriched for transcripts involved in DNA processing and cell cycle/DNA replication (Fig. 3D, Figure S3B, Table 2).

Because *Nanog* 3UTR cells populate organoid peripheries, and co-segregate with future trophectoderm in the embryo, we asked if mESC cluster 4 (high *Nanog* 3UTR) cells are truly pluripotent, or possibly trophectoderm-like. While distinct, the transcriptome of cluster 4 cells

indicates high pluripotency, and shows no significant expression of differentiation markers (Fig. 3H).

Further examining cluster 4; there is a distinct subcluster of cells (denoted 4a), that expresses high *Nanog* 3UTR and no *Oct4* mRNA- these cells likely represent outer zone-1 cells. The remaining cluster 4 cells do express *Oct4* (denoted 4b), and likely represent inner zone-1 cells (Fig. 3I, Table 2). We find that the *Nanog* 3UTR/CDS ratio changes dramatically across clusters 4a, 4b and 0 (zone-2) and asked then, if Nanog signaling differs as well. Three critical Nanog effector genes, *Zfp42*, *Cldn4* and *Crxos* (Boroviak et al., 2015), also show dramatic 3UTR/CDS changes, with high 3UTR or no expression in cluster 4a, to high CDS in cluster 0 (Fig. 3I). In contrast, they do not show large changes in absolute expression levels (Fig. 3I; left with blue dots).

These data highlight that examination of the ratio of 3UTR to CDS can identify new cell clusters, or trends across clusters, and suggest that Nanog signaling is distinct in high *Nanog* 3UTR cells compared to high *Nanog* CDS cells (Table 2). These trends would not be apparent by examining transcript levels without distinguishing 3UTR and CDS expression. These studies add to our knowledge of the different zones by showing that high *Nanog* 3UTR cells both in mouse and in multiple hESC cell lines are a transcriptionally distinct population of cells within the pool of undifferentiated ESCs. (Table 2)

Table 2 Transcriptome analysis of *Nanog* cells

Zone/character	***Nanog***	**hESC**	**mESC**	**mESC Nanog downstream gene mRNA**	**Prominent GO hESC**	**Prominent GO mESC**	**Pluripotent**
Inner zone-1	High 3UTR, no CDS	Cluster 2	Cluster 4a	High 3UTR	DNA processing and cell cycle/DNA replication	Cell cycle and chromosome organization	Yes
Outer zone-1	High 3UTR, low nuclear or no CDS		Cluster 4b	variable			Yes
Zone-2	Hi 3UTR and CDS	Cluster 0	Cluster 0	High CDS	Epithelial cell proliferation Heart development	Cell development, differentiation and signal transduction	Yes

Histone modification identifies differential transcriptional activity in high *Nanog* 3UTR cells versus other cells in 2D organoids

We next performed gene ontology (GO) analysis of transcripts in mouse and human ESCs expressing high *Nanog* 3UTR, and found enrichment of the GO categories “chromosome organization” and “cell cycle”. This prompted us to examine whether either of these biological processes contributes to regulation of, or is an effector of, the isolated *Nanog* 3UTR, using immunohistochemistry to examine markers of these processes. First, Ki67, a marker of proliferation (or of all cells not in G0) is observed in all zone-1 and most zone-2, but fewer zone-3 and 4 cells (Fig. 4C,D), consistent with the fact that both zone-1 and 2 cells are undifferentiated. Next, we examined “active” (H3K9ac, H3K27ac, H3K4me3), “repressive” (H3K9me3, H3k27me3), “differentiation” (Hp1a) and “proliferation” arrest (H4K20me3) histone marks. Chromosomes can be tightly packed and inaccessible to transcription, or loose and accessible, largely governed by the covalent modification of histone protein tails by such active

or repressive marks. We first examined these marks in SON genes themselves. In H1 hESCs, active marks are observed 5' to each SON gene, but not to control genes (Figure S4A). In addition, we observe many active marks both within the CDS, and 5' to the 3UTR start (Figure S4A). As we were interested in the state of the cells in 2D organoids in relation to zonal SON 3UTR/CDS expression, we examined histone mark and other marker expression across the organoids, as a cellular measure of transcriptional activity (Risom et al., 2022; Vogelauer et al., 2000). Remarkably, most showed non-random, concentric expression in the 2D organoids (Fig. 4A), largely in register with the *Nanog* UTR/CDS zonal domains (Fig. 4B,C, Figure S4 B-J', and Table 3).

Table 3. Histone marks in *Nanog* cells

Zone/ character	***Nanog***	**Active Marks**	**Lineage maintenance**	**Repressive**	**Adherens-junctions**	**Tumor supp or different.**
Outer zone-1	High 3UTR, no CDS	H3K27ac and H3K9ac	H3K9me3 high		No	No
Inner zone-1	High 3UTR, low nuclear or no CDS		H3K9me3 low	H3K27me3 subset	No	No
Zone-2	Hi 3UTR and CDS	H3k9ac and H3k4me3	No		Yes, high	No
Zone-3	Moderate CDS, no 3UTR	No	No	No	Moderate	Yes high
Zone4	Sporadic high 3UTR	No	Sporadic	Hp1a, and proliferation arrest mark, H4K20me3	No	Yes high

Zone-1 (high *Nanog* 3UTR) cells express active marks (H3k9ac and H3K27ac), as well as repressive marks H3K9me3 (alternatively known to play a role in maintaining lineage fidelity) (Nicetto and Zaret, 2019), and some H3k27me3 (in inner zone-1) (Fig. 4A; as shown. H3k27ac and HeK27me3 are largely non-overlapping (Figure S4)). In contrast in zone-2, where all SON mRNA components are highly expressed, most repressive marks are diminished, (with some residual H3k27me3), and an additional active mark, H3k4me3, is robustly expressed. Adherens-junctions are necessary for stem cell proliferation and the adherens-junctions marker ZO1 is highly expressed in zone-2 (but not zone-1) (Fig. 4A-C). Cells in zone-3 and 4 are markedly different: in zone-4 all of the aforementioned markers are turned off, while beginning at zone-3, the differentiation and proliferation arrest marks (Hp1a and H4K20me3 respectively) (Fraga et al., 2005; Van Den Broeck et al., 2008) are robustly expressed (Fig. 4A-C, Supp Fig. 4B-D', Table 3). Thus, as summarized in Fig. 4D, we find an outside-in pattern in undifferentiated ESC organoids, with distinct cell types, containing discrete (active or repressive) transcriptional profiles, taking up residence in stereotypical locations.

To assess a possible functional relationship between histone marks and *Nanog* mRNA expression, 2D organoids were grown in the presence of Cpth6, an inhibitor of H3k9ac, an active mark expressed in zone-1-3 (Fig. 4A,D). Inhibition of H3k9ac disrupts zonal *Nanog* 3UTR/CDS expression, significantly increasing *Nanog* CDS while decreasing high *Nanog* 3UTR cells (Fig. 4E). In comparison, A485, a general inhibitor of CBP/p300, known also to affect *Nanog* expression (Ebrahimi et al., 2019) caused the opposite effect: a significant increase in high *Nanog* 3UTR cells, and a decrease in *Nanog* CDS cells (Fig. 4E).

We next examined the effect of *Nanog* 3UTR deletion *(N-U*ko) on histone mark expression. Deletion was carried out in mESCs using a CRISPR knockout approach, with dual

(labeled) guide RNAs (gRNAs) targeted to each end of 3UTRs (Fig. 4F, and see Fig. 5 and Figure S5 for details). Histone mark expression was quantitated by FACS analysis of double gRNA expressing cells. H3k9ac and H3k9me3, which are normally expressed in zone-1 or zone 1-3 cells, respectively (Fig. 4A), were highly increased after *Nanog* 3UTR knockout (*N-U*ko), (Fig. 4G). In comparison H4k20me3, which is restricted to zones 3 and 4, increased only slightly after *N-U*ko. Deletion of either the *Oct4* or *Sox2* 3UTRs caused fewer, and different changes (Fig. 4G).

In summary, differential SON 3UTR/CDS expression not only marks spatial domains but also correlates with, and is modified by, global gene transcription states as well. Each SON gene mRNA is expressed in multiple distinct forms, with the relative levels of its 3UTR and CDS mRNA components being non-randomly "switched". The expression of each mRNA component is thus differentially and independently regulated, resulting in a self-organized ESC structure in organoids, with a similar pattern in the early embryo. The zonal relationship of these cells, schematized in Fig. 4D, suggests a stem cell niche, where external and internal cells serve different roles (Herrera et al., 2021; Knight et al., 2018).

***Nanog* 3UTR knockout causes colony stasis and patterning changes**

While high *Nanog* 3UTR expression distinguishes organoid boundaries, and represents a particular class of ESC, whether or not the isolated *Nanog* 3UTR has a biological role, distinct from the CDS, is not known. In perimeter cells, the *Nanog* 3UTR is prominently expressed in the nucleus (Fig 1A,2D, Figure S1A,A'). Such expression is also found in a smattering of cells in zone-4 with high *Nanog* 3UTR and low or no CDS cells, mostly in cells of larger organoids (1000 µm, and some 500 um) (Fig. 5A,A'). Based on Ki67 expression and mitotic figures visualized by Dapi, these zone-4 cells appear to be actively dividing (Fig. 5A; yellow arrows and quant) with over 94% of the M phase cells in this region expressing high *Nanog* 3UTR. In these cells the *Nanog* 3UTR can be seen tightly localized to either side of the mitotic spindle (Figure 5A). Like perimeter cells, zone-4 high *Nanog* 3UTR cells also specifically express H3k9me3, the lineage specific maintenance mark (Nicetto and Zaret, 2019), but not H3k4me3 (a zone-2,3 marker), nor high levels of *Oct4* (Fig. 5B,C, Figure S5A,A'), suggesting they have similar characteristics to zone-1 high *Nanog* 3UTR cells. Given the prominent location of high *Nanog* 3UTR in mitotic figures, along with high representation of the cell cycle regulation GO category in both mESC and hESC high *Nanog* 3UTR cells by scRNA seq, we examined the role of the isolated *Nanog* 3UTR in cell proliferation and colony kinetics using *Nanog 3UTR* knock-out *(N-U*ko) cells. *Sox2* and *Oct4* 3UTR knockout cells (*Sox2-U*ko, *Oct4-U*ko) were used for comparison. Significant and specific deletion of the 3UTR or the CDS was verified by genomic PCR and/or T7 digestion (Figure S5B, and not shown).

Cells with a *Nanog* 3UTR deletion (*N-U*ko) maintain *Nanog* CDS and protein expression (Fig. 5D and Figure S5E,F). For colony formation, FACS-sorted double gRNA positive cells were plated at clonal density for 5 days; *N-U*ko produced a significant decrease in both the size and number of knockout colonies, compared to *S9-U*ko, *Hd10-U*ko, or vector control (Vect ctl) (Fig. 5E). To confirm the function of the *Nanog* 3UTR itself, separate from the CDS, rescue was carried out by concomitant co-overexpression (oe) of a CRISPR resistant *Nanog* 3UTR* (Fig. 5F; schema), which rescued colony formation in *N-U*ko relative to control (Fig. 5G; quant). Thus, *Nanog* 3UTR expression, independent of the *Nanog* CDS, restores robust colony formation. Proliferation and death were further assessed by 24 hour live image analysis of single

colonies. We found a significant decrease in the number of cell divisions per cell in *N-U*ko compared to control, (Fig. 5H; representative pictures), and significant rescue of the number of cells with one division, by *Nanog* 3UTR overexpression (*N-U*oe*) (Fig. 5H; quantitation below right). In addition to fewer divisions, the latency to first division is significantly increased (Fig. 5I; left), and this is rescued to control levels by *N-U*oe* (Fig. 5I, right). Cell death was not changed after *N-U*ko (Figure S5H). Finally, we examined patterning after *N-U*ko by quantitative PCR (qPCR) measurement of pluripotency and germ layer markers. Even when grown continuously in pluripotency media, *N-U*ko cells showed decreased pluripotency, and upregulation of trophectoderm markers, but not markers of endoderm, mesoderm or ectoderm (Fig. 5J; and not shown). In contrast, *Nanog* CDS knock-out *(N-Cko)* cells did not show the same effects (Fig. 5J).

While *Nanog* 3UTR overexpression (*N-U*oe*) did not increase colony size (Fig. 5G), we further examined its effects on 3D organoid patterning where we observed a decrease in Sox2 protein in *N-Uoe** cells (Figure S6B,C), and further effects on downstream patterning (Figure S6D-H). Overexpression of *Nanog* CDS (*N-C*oe), in contrast, did not decrease Sox2 protein levels (Figure S6C). These studies confirm that the *Nanog* 3UTR is essential for proper cell division and that it can be effective in cis, restoring 3UTR deletion or affecting patterning after OE.

High *Nanog* 3UTR cells show distinct signatures that are associated with niche hub cells

In stem cell niches, the stem cells are bordered by “hub” support, or more quiescent cells (Lilly et al., 2011), often depicted as oriented orthogonally to the other cells (Pagella et al., 2015; Spradling et al., 2001). Niches are also governed by short range cues that promote stemness, in many cases through activation and inhibition via the activin signaling pathway (Raz and Yamashita, 2021). Our data above, showing stereotypical localization of distinct cell types, with perimeter high *Nanog* 3UTR cells oriented orthogonal to the other cells and containing distinct histone marks, suggest a niche structure, with high *Nanog* 3UTR cells as “hub” cells. To further evaluate this possibility, we examined both cell cycle and quiescence markers (Cheung and Rando, 2013) in high *Nanog* 3UTR cells compared other cells. Remarkably, High *Nanog* 3UTR cells (mESC) differ distinctly from all other ES cells on both counts: cell cycle control genes (across the board) are downregulated (Fig. 6A), as are genes specific to prevent quiescence (Fig. 6B).

We also examined morphogenic signaling pathways. Of select morphogen ligands, only activin family members, (e.g. *Gdf3*) are significantly expressed in any mESC cluster (Fig. 6C and Figure S7A,B). We therefore examined multiple activin signaling molecule genes (Andersson et al., 2008; Levine and Brivanlou, 2006) across clusters, using clusters 4a, 4b and 0 as surrogates for inner and outer zone-1, and zone-2 respectively. While none differ in absolute expression across these clusters (Figure S7C), many show dramatic 3UTR/CDS ratio changes (Fig. 6D, E; summary), indicating an asymmetry in activin signaling between zones 1 and 2. To test the functional relevance of activin signaling, broad spectrum activin receptor inhibitors (A8301 or SB431542 (the latter not shown)), or an activin agonist were added to developing 2D organoids for the last 48h of the 72h culture period. Both activation and inhibition of this path disrupts the *Nanog* mRNA pattern in distinct ways, compared to control DMSO treated cultures (Fig. 6F).

High *Nanog* 3UTR cells may thus represent a type of hub cell, forming a niche boundary providing structure, organization and stability. One puzzle, however, is that zone-1 high *Nanog* 3UTR cells are Ki67 positive, indicating that they have divided within the previous 48 hours. Traditionally, stem cell pools or hub cells are not highly mitotic, but slowly cycling or dormant (Lilly *et al.*, 2011). In certain niches, however, particularly cancer stem cell niches, hub cells may transdifferentiate to stem cells as needed (Herrera *et al.*, 2021; Raz and Yamashita, 2021). We hypothesized that in ES cells, zone-2 stem cells may be in exchange with zone-1 perimeter high *Nanog* 3UTR cells, with proliferation more likely to occur in zone-2 than zone-1. To test this, we used a combination of time lapse analysis and pulse EDU labeling. Tracking of GFP-labeled cell movement in near-confluent 2D organoids for 24-48h showed that more than 30% of GFP labeled cells transitioned outward from the zone-2 region to the perimeter. Very few actively migrated inwards, however one third of cells that had migrated to the edge were then pushed inward by cell movement by 48h, causing mixing. Thus, there is ample movement between zone-1 and zones-2,3. We next used EDU labeling, comparing the results of EDU pulse without chase (immediate fixation), to EDU pulse followed by a 6hr chase. Without chase, a significantly greater percentage of cells in zones 2 and 3 incorporated EDU compared to zone-1 (Fig. 6H) suggesting greater proliferation in zones 2 and 3 compared to 1. When pulsed, then chased for 6 hours (with time for cells to move in the absence of further EDU incorporation), this significance was erased (Fig. 6H; quantitation), suggesting that EDU positive cells move out of zone-2 or -3 into zone 1 during the 6hr chase.

In summary, as schematized in Fig. 6I, in 2D organoids and early embryos, our data are consistent with a model in which ESCs spontaneously form niche-like structures, with perimeter high *Nanog* 3UTR cells representing hub cells, and more interior zone-2 cells that are positive for all SON components, activating histone marks and adherens-junctions markers, representing the transit amplifying cells (Hsu et al., 2014) of the stem cell zone. Zones-3 and -4, marked by tumor suppressor and differentiation marks, appear largely to represent partially or fully differentiated stem cell spin-offs, while still retaining a reservoir of hub cells (high *Nanog* 3UTR/H3k4me3 positive cells).

Discussion

Pluripotent ESCs self-organize as a niche structure that resembles the in-vivo embryo

Here we show dramatic, patterned, concentric, differential expression of critical pluripotent gene 3UTR and CDS mRNA components in undifferentiated ESCs, and in developing human and mouse embryos. This differential expression reveals robust pluripotent ESC heterogeneity that provides the underpinnings of a previously unrecognized ESC cell niche structure. We find that ESCs dynamically modulate their SON 3UTR/CDS ratios while in a pluripotent state, suggesting that such differential pluripotent gene mRNA component expression is critical at a time when pluripotency, by definition, must be in delicate balance with differentiation. Such dynamic, differentially expressed pluripotent gene mRNA component expression provides a molecular mechanism to allow the crucial stem cell heterogeneity necessary to maintain a stem cell population, in the absence of any cells initiating irreversible programs of differentiation. This work pieces together parts of the puzzle in that embryonic stem cells exist in a state of non-permanent, undifferentiated but systematic, heterogeneity.

Niches were first described for hematopoietic cells and then in *Drosophila* (Li and Xie, 2005; Lin, 2002; Schofield, 1978; Weissman, 1994), and possess specialized cell populations residing in stereotypical locations in relation to the other cells (Li and Xie, 2005). Our studies reveal that there is a highly stereotyped molecular pattern that arises spontaneously in organoids and that is also evident in embryos, forming the underpinnings of robust structures. We propose that in these niches, the perimeter cells that express high nuclear *Nanog* 3UTR with no or low expression of other SON mRNA components represent niche hub cells (Knight *et al.*, 2018) (Li and Xie, 2005). They exhibit a unique histone mark signature, and show lowered cell cycle activity and higher quiescence than other ESCs – all markers of hub cells and are in exchange with interior cells. Moving further interior within the niche, (in zone-2) we find dividing stem cells, or transit amplifying cells. They express Ki67, Oct4 protein, all SON mRNA components, high levels of adherens-junctions markers, which are necessary for stem cell division, and primarily activating histone marks. In contrast, cells in organoid centers (zones-3 and -4) express differentiation marks, suggesting some partial or pre-differentiation along the niche radial axis. Interestingly, however, within zones 3 and 4 there remains a reservoir of dividing, high *Nanog* 3UTR cells.

In certain niches, signaling and interchangeable movement between hub cells and internal cells occurs (Raz and Yamashita, 2021). We find such movement in ESC 2D organoids, with many cells mixing between zones 2 and 1, suggesting that they change from expressing both high *Nanog* 3UTR/and CDS to just high *Nanog* 3UTR. This movement might in addition, provide an explanation for the fact that Nanog protein expression is as high in *Nanog* 3UTR zone-1 cells as it is in zone-2, the high *Nanog* CDS cells: an intriguing possibility is that cells synthesize Nanog protein in zone 2 (in cells expressing high *Nanog* CDS) and then move to zone-1, losing *Nanog* CDS while still maintaining Nanog protein.

We additionally find that the ESC 2D organoid niche structure is regulated by morphogenic signaling, specifically by activin. This aligns with previous findings that extrinsic signals act on undifferentiated embryonic stem cells (Davey and Zandstra, 2006), and that in particular activin/nodal signaling modulates Nanog levels (Peerani *et al.*, 2007; Vallier et al., 2009). Activin signaling is asymmetrically higher in the epiblast than the trophectoderm (Herrera *et al.*, 2021), suggesting that a check on activin activity may prolong pluripotency. Activin also plays a critical role in hub cell interchange with stem cells in other systems, as secretion of Follistatin (an activin inhibitor) from the internal cells is necessary for hub cell survival (Raz and Yamashita, 2021). The differential 3UTR/CDS expression of various activin signaling molecules in zones 1-3-like cells, together with the marked effects of activin modulating drugs on *Nanog* mRNA pattern, suggests that under normal circumstances, tight regulation of this pathway with its many activators and inhibitors, may refine and tighten organoid patterning, and that differences in the 3UTR/CDS ratio for each of molecules in this pathway itself, may additionally add to this refinement.

Further, we find that niche regions are marked by concentric expression of unique combinations of histone marks. The exact role of the histone marks in creating the niche structure is not clear, but deletion of the *Nanog* 3UTR causes upregulation of histone marks that would normally be expressed in high *Nanog* 3UTR cells, suggesting that the histone mark is upregulated as a first step towards the replenishment of the deleted cells. Consistent with this, blockade of the zone-1 histone H3k9ac decreases the number of high *Nanog* 3UTR cells. Both lines of inquiry suggest that histone marker expression influences SON mRNA component expression.

The isolated *Nanog* 3UTR as a long non-coding RNA

By Crispr knockout, we show that the isolated *Nanog* 3UTR plays a biological role in ESCs. While not essential for short-term viability, it is essential for robust, healthy colony formation. When expressed separately from the *Nanog* CDS, the *Nanog* 3UTR is found at high levels in the cell nucleus and straddling the mitotic spindle of dividing cells. It plays an important role in regulating the ability of ESCs to self-perpetuate and maintain the integrity of a colony since in its absence, cells fail to proliferate normally, and begin to differentiate towards the trophectoderm lineage. Our data also reveal a range of *Nanog* 3UTR to CDS ratios. Cells with high *Nanog* 3UTR and little or no CDS (outer zone-1) are not highly proliferative. In contrast, cells with small amounts of CDS, such as those in outer zone-2, and in the interior of large organoids, are highly proliferative, as are cells with similar levels of *Nanog* 3UTR and CDS. Cells with no detectable 3UTR, but high CDS, such as those in zone-3, do not proliferate. Thus, the *Nanog* 3UTR/CDS ratio, and – based on our deletion studies – the 3UTR in particular, appear to play an important role in proliferation-quiescent kinetics, an essential balance for the prevention of stem cell exhaustion and maintenance of the stem cell pool (Bradford et al., 1997).

Generation of isolated 3UTR or CDS mRNA components

How differential expression of SON gene 3UTR and CDS occurs at a molecular level is still only partially understood (Ji *et al.*, 2021; Malka et al., 2017; Mercer et al., 2011; Sudmant et al., 2018). Cap analysis gene expression (CAGE) shows that for a number of genes (Ji *et al.*, 2021; Mercer *et al.*, 2011) including *Nanog*, CAGE tag marks are found at the 5' end of the 3UTR transcript, thus these transcripts are capped (Figure S5H). For some genes, transcription factor binding regions and CpG islands reside upstream of 3UTR peaks (Ji *et al.*, 2021). For other genes in which genomic elements do not correlate well with 3UTR peaks we find a high correlation of mRNA stabilization and destabilization (Rabani et al., 2017) motifs and 3UTR peaks (unpublished observations). This suggests that some isolated 3UTRs may be transcribed *de novo*, while others may be preferentially stabilized compared to their CDS.

Mouse vs human ESCs and embryos

In these studies we have examined both mouse and human ESCs. While there are some slight differences, we find that in 2D organoids, and in blastocysts, they largely show very similar SON mRNA component expression patterns, similar responses to exogenous factors, similar correspondence with histone marks, and similar transcriptional profiles.

While we focused most extensively here on the *Nanog* 3UTR, we also found that all SON gene 3UTRs can cause changes in gene expression, colony formation and/or patterning, suggesting that each plays a distinct and important role in mouse and human ESC colony and niche formation (Figs. 4J, 5E,I,J, Figures S2A, S4K, S6 E,F, S7D). While there are some similar effects of different SON 3UTR knockouts, this is not unexpected, as all SONs are known to act dose-dependently in embryonic stem cells (Boyer *et al.*, 2005), with perturbation in either direction causing a loss of pluripotency. Thus, these 2D organoid niche structures will be useful for further study of early embryonic development.

In sum

Beyond the specific role that each SON gene 3UTR may play, differential 3UTR/CDS expression allows ESCs to distinguish themselves from one another without "extinguishing" the core genes necessary for pluripotency, and thus losing pluripotency. Cell sorting is well known to play a critical role in multiple aspects of development and patterning (Krens and Heisenberg, 2011), and differential marker expression is a critical step in this process. We also previously described distinct 3UTR/CDS expression patterns for important regulatory genes in both hair and neural stem cell niches (Ji *et al.*, 2021), suggesting this may be a common mechanism to help generate heterogeneity while maintaining pluripotency. This principal also has wider implications, as our increased understanding of differential mRNA component expression is likely to positively impact our understanding of basic biological and disease states across tissues and cells.

In sum, we demonstrate three important findings: first, illumination of a dramatic and dynamic differential 3UTR and cognate CDS expression in ESCs that reveals a detailed stem cell niche structure; second, a marked but systematic heterogeneity of undifferentiated human and mouse embryonic stem cells; and finally, demonstration of an important and relevant biological role for an isolated 3UTR transcript. Future work will focus on the precise roles of the isolated 3UTRs for all three SON genes, as well as interactions between the isolated 3UTRs of SON genes, their cognate proteins, and other potential binding partners.

Acknowledgements; we thank Melanie O'Rourke for technical assistance, Arif Kocabas, Carolyn Sook Park and Ali Brivanlou for pilot studies on embryonic stem cells, and Marc Tessier-Lavigne, Richard Axel, Jeffrey Friedman, Mary Beth Hatten and for critical reading of the manuscript.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mary Hynes (mhynes@stanford.edu).

This study did not generate new unique reagents.

Author contributions; Z. Yang together with M. Hynes, conceived of all conceptual ideas and experiments. Z. Yang carried out all wet lab experiments within, and was instrumental in figure and manuscript preparation, as well as overseeing and coordinating all support research staff. S. Ji was instrumental in early CRISPR knockout studies, and performed all the bioinformatic analyses, with help from S. Parsa. She also contributed to figure design and paper editing. L. Gozali contributed to early conceptual ideas as well as providing research support. M. Ullah provided research support and he and S. Parsa contributed to manuscript discussions. B. Behr provide human blastocysts and training in their handling.

Declaration of interests- the authors declare no competing interests.

<u>Fig. 1</u> Mouse and human embryonic stem cells self-organize according to SON mRNA 3UTR/CDS ratio.
All; For all *SON* and other genes, 3UTR; red, CDS; green, unless otherwise noted. Throughout the manuscript- high 3UTR (vice versa for high CDS) means greater 3UTR expression than CDS, it does not mean there is no CDS. (A-C) expression of SON 3UTR/CDS in 2D organoids left; mouse ESC (mESC), middle; human ESC (hESC). 72h culture of untreated embryonic stem cells grown as 500 μm, 2D organoids, right; zonal expression schemas. (D) Organoids were divided into 4 zones based on *Nanog* 3UTR/CDS expression (and see Fig. 4 and Figure S4B-J'). (E) *Nanog* 3UTR/CDS in 80-1000 μm Cytoo^{r} chip circles. Outer ~100 μm; *Nanog* 3UTR plus CDS pattern, interior cells; Dapi positive (and see Figures 2B and S4H). The *Nanog* CDS sometimes appears to be graded and diffusing, and is possibly extracellular (A). (F) *Nanog* 3UTR/CDS in bulk-culture, untreated mESCs, top; low power, bottom; higher power. *Nanog* high CDS cells are found primarily in aggregates (X^2, $p<0.01$; right), with high *Nanog* 3UTR, low/no CDS in unaggregated cells. Scale bar; (A-C) 60, (D) 66, (E) 100, (F) 120 μm.

<u>Fig. 2</u> Differential expression of SON cognate 3UTR and CDS mRNA components in mouse and human blastocysts in vivo. (A) *Nanog* 3UTR occupies all borders before CDS expression is initiated (white arrows and Figure S1B,C), and *Nanog* 3UTR in border cells is nuclear (see Figure S1A,A') (B) quantification of *Nanog* 3UTR cells; red, Dapi; blue, (Dapi positive cells are visible throughout the circle, but dimmer in center; and Figure S4H). On one chip, the ratio of high *Nanog* 3UTR/total cells is quantified, and different, in different sized circles; graph insert. (C) In 1000 μm organoids *Oct4* 3UTR/CDS expression undergoes symmetry breaking (van den Brink et al., 2014), and differential 3UTR/CDS expression is polarized and distinct, n=8 (and Figure S2A). (D-F) SON protein (blue) with cognate 3UTR/CDS expression in hESCs (similar finding in mESCs; not shown). Separate channels, as marked, shown on side. The Nanog and Sox2 antibodies were verified for specificity (Figure S2B). Sox2 protein, surprisingly, is highest in cells without high levels of *Sox2* mRNA components (in both hESC (shown) and mESC), and in 2D and 3D cultures, using two independent antibodies (not shown). (G-L) mouse and human blastocyst embryos, examined for SON mRNA and proteins, as marked. The ICM marker Oct4 protein, is highly expressed in *Nanog* high 3UTR/high CDS cells (Figure S2F). (L) Oct4 protein is detected in most, but not every cell in the embryo. Scale bar; (A,C) 120, (B,D,E) 100, (H-L) 18 μm. And see Figure S1 and S2.

<u>Fig. 3 Mouse and human *Nanog* high 3UTR cells define distinct cell populations</u>
(A-D) scRNA seq of (1146) hESC cells. (A) tSNE clustering of hESC colored by (A) cell type and culture condition as marked. Undifferentiated cells from both H1 and H9 hESC lines form two closely aligned clusters, cluster 2; blue outline and cluster 0; purple outline. (B) Superimposed expression level of *Nanog*. Left; TPM; right, *Nanog* 3UTR/CDS ratio (C) Superimposed expression level of *Oct4*. Left; TPM; right, *Oct4* 3UTR/CDS ratio. (E-G) scRNA seq of (864) mESC cells. (E) tSNE clustering of mESC colored by cell type and by culture condition (by symbol as marked). Cells from 2i media (with some a2i, both similar to the

medium used here) fall into four clusters; 2,4,7 and 0. (F) Superimposed expression level of *Nanog;* left and *Oct4;* right. High *Nanog* 3UTR cells populate cluster 4; high *Oct4* 3UTR cells populate cluster 0. (D,G) GO function enrichment analysis (additional categories Figure S3) (D) hESC for cluster 2 vs others; left, for cluster 0 vs others; right. (G) mESC for cluster 4 vs others; left, for cluster 0 vs others; right. Superimposed *Nanog* or *Oct4* 3UTR/CDS ratio in each cell shown as; high 3UTR; red dots, high CDS; green dots, no expression; gray dots, for denoted gene. (H) Heat map showing the expression of select genes for pluripotency, differentiated germ layer markers, or other cell types (Guttman et al., 2011; Kopp et al., 2008; Loh *et al.*, 2006), in cells from clusters 2,4,7,0 (grown in 2i or a2i media). Cells in all clusters are pluripotent and show low levels of other markers. Scale and categories on right. (I) examination of *Nanog* effector genes (Festuccia et al., 2012; Loh *et al.*, 2006; Shi et al., 2006) by transcript level (left), and *Nanog* 3UTR/CDS ratio (right), for each gene. Cluster 4 was divided into two subregions, cluster 4a; with no *Oct4* expression possibly representing outer zone-1 cells, and cluster 4b; high *Nanog* 3UTR plus *Oct4* CDS, representing inner zone-1cells, and cluster 0, representing zone 2 cells (and see Table 2). Differences in the 3UTR/CDS ratio for *Nanog* effector genes across clusters is represented by the larger colored circles. Red; most cells show high 3UTR, yellow; most cells show equal 3UTR and CDS, gray; most cells show no expression. Two circles indicate a mixed population with a large proportion of those two cell types. And see Figure S3.

Fig. 4 *Nanog* CDS/3UTR zonal expression correlates with transcriptional activity as identified by histone modification.

(A) Zonal alignment of histone and other marks with *Nanog* 3UTR and CDS expression. Top left; *Nanog* 3UTR; red, CDS; green, Dapi; white. Others: white, protein as marked, Dapi; blue. Prominent expression zone(s) for each marker listed above; mark characteristic listed inside. (B) Hp1a is restricted to zones 3 and 4 and is non-overlapping with either *Oct4* 3UTR or CDS. (C) *Nanog* 3UTR/CDS and various markers; as marked (other markers in Figure S4 B-E'). (D) representative images of hESC 2D organoids after 72h growth in presence with indicated inhibitor; quantitation on right; green bars; *Nanog* CDS, red bars; *Nanog* 3UTR (similar results for mESC; not shown). (E) schematic summarizing marker expression in various zones, (*) indicates most prominent expression zone. Purple shaded ovals indicate expression levels (darkest is highest). Green shaded zones are active, lavender; repressive.) (F) schematic of CRISPR knockout in mESCs for the *Nanog*, *Oct4*, *Sox2*, or other 3UTR's (CDS; green box, 3UTR; red line). Dual-tagged 5' and 3' gRNAs (see Figure S5C for *Nanog* gRNA locations; others with similar maps; not shown), sparing the poly A addition site, were used to delete each 3UTR. (G) Histone marker expression was then measured in FACs sorted doubly-labeled (yellow) gRNA cells; quantitation of histone marker in labeled cells, normalized to Vect control. Scale bar; (A,B) 1000, (C) top; 220, bottom; 75, (E) 180 μm. And see Figure S4.

Fig. 5 *Nanog* 3UTR knockout causes colony stasis and patterning changes

(A-C) Large, 1000 μm organoid. (A,A') 94% of mitotic figures, identified by Ki67 and Dapi signature (white lines marked by yellow arrows A'; yellow arrows), are also highly positive for *Nanog* 3UTR expression; red. (A') *Nanog* CDS is also sometimes observed in these cells.

Nearby, *Nanog* 3UTR negative, non-dividing cells (blue arrows), show higher levels of *Nanog* CDS; green; Dapi; white. (B) Same section as A'; all high *Nanog* 3UTR cells express H3K9me3, the lineage maintenance mark expressed in perimeter high *Nanog* 3UTR cells; other nearby Dapi positive cells do not. (C) Sister culture showing that Dapi identified, dividing cells do not express high levels of *Oct4* 3UTR or CDS or of H3k4me3 (and Figure S5A). (D) *Nanog* 3UTR and CDS expression in vector control (Vect Ctl), *N-U*ko, or *Oct4* 3UTR knockout (*O4-U*ko) mESC cells. Cells were grown as 3D organoids for 5 days (see methods; and Figure S6D). Vect control cells show differential high *Nanog* 3UTR and high *Nanog* CDS cells (quantified below as % of explant for each marker). After *N-U*ko, high *Nanog* CDS cells are significantly increased (as % of explant) ($p < 0.01$), and high *Nanog* 3UTR cells decreased ($p < 0.01$). After *O4-U*ko, *Nanog* 3UTR and CDS expression as a % of the explant is not changed. All knockouts were verified by PCR (Figure S5B), and non-overlapping gRNA pairs (Figure S5C; and not shown). (E) Colony formation of FACs sorted 3UTR knockout cells. *N-U*ko and *S2-U*ko produce significantly smaller size single colonies than Vect Ctl colonies ($p < 0.01$); quantitation below. (F) Schematic for 3UTR knockout and rescue. CRISPR resistant (*) 3UTR overexpression constructs (e.g. *N-U*OE*) were co-infected into double gRNA infected mESCs. (G) FACS sorted double gRNA positive 3UTR knockout cells plus and minus plv vector control, or *N*-UOE* grown as single cell colonies for 9d. *N-U*ko and *N-U*ko plus plv vector control OE colonies were significantly decreased in size ($p<0.001$), but restored to normal, after co-*N*-UOE* (see quantitation on right). (H) representative pictures of 24 h live imaged Vect Ctl or *N-U*ko cells; * show divided cells after 24 hours; (100%) of Vect Ctl, 20% of *N-U*ko cells. Quantitation using Chi-square ($\chi 2$) analysis of the number of cells/total cells, showing 0,1 or 2 divisions (below; left); after *N-U*ko fewer cells divide twice and more divide only once. Cells that divide once are decreased to control levels after co-*N-U*ko + *N*-UOE* (right). (I) Latency in time (left) to first division is significantly increased after *N-U*ko or *O4*-Uko, (and also after *Nanog* CDS (*N-C*ko), or *Oct4* (*O4-C*ko) compared to Vect Ctl. (Right) time to first division was restored by *N*-UOE* ($p < 0.05$, vs. Vect ctl, with similar findings for *O4*-Uko). (J) rtPCR of marker expression (as noted), in FACs sorted knockout cells. Critical trophectoderm (Loh *et al.*, 2006) markers are significantly upregulated in *N-U*ko vs Vect ctl ($p <0.01$), but not after *Nanog* CDS deletion (*N*-Ckd). The majority of endoderm (as well as mesoderm or ectoderm; not shown) (Kopp *et al.*, 2008; Loh *et al.*, 2006), markers were not elevated after *N-U*ko. Scale bar; (A,B) 1000, (C) top; 220, bottom; 75, (D) 180 μm. And Figure S4.

<u>Fig 6. High *Nanog* 3UTR cells show distinct signatures that are associated with niche hub cells.</u> (A,B) Heat map expression analysis of (A) cell cycle or (B) quiescent marker genes in cells in mESC clusters 2, 4, 0 and 7. Dark purple bars; high expression, yellow bars; low expression. Cell cycle control genes (for both G2M and G1S) are downregulated in high *Nanog* 3UTR cells (cluster 4) compared to other clusters (0,2, and 7). "Cell cycle regulated" genes were not as affected. (B) likewise, genes that suppress quiescence were downregulated in mESC cluster 4, high *Nanog* 3UTR cells. Surprisingly all mESCs showed high levels for genes that enhance quiescence. (C,D) gene expression for morphogenic ligands in mESC (and Figure S7A-C) (C) tSNE analysis for *Activin*, *Wnt* and *Hedgehog* ligand(s) expression. (D) Overlay of 3UTR/CDS expression ratio for select *Activin* family receptors, inhibitors and down-steam signaling molecules in each cell. Large dots show 3UTR/CDS for majority of cells in each zone, as

marked. Gray; no expression. Zoom of selected gene expression in clusters of interest, 4a, 4b and 0. (E) Summary schematic of Activin pathway molecules in zones as marked. (F) Representative images of 2D organoids after 72 growth in presence of *Alk* receptor inhibitor (A 83-01), or 40ng/ml of activin agonist; quantitation on right (SB43154) showed same result, and similar effects seen using mESC; not shown). The *Nanog* CDS is increased by the inhibitor, while the *Nanog* 3UTR is decreased by Activin addition (p values as marked). (G) representative movie frames from 24 h time lapse of GFP positive hESC cells in developing 2D organoids, measured for 24 hours from 48h to 72h in culture. Cells were measured for movement between zone-2 (>8 cell diameters from the edge) to the edge, or vice versa. Orange bar shows distance of cell to the edge at start, blue bar shows current distance of cell to edge. Quantification below. (H) Edu incorporation into the various zones in developing 2D organoids with fixation 0 or 6 hours after a 15 min pulse. EDU; red, Dapi; white. Quantitation on right shows significantly more incorporation into zones-2 and 3 compared to zone-1, upon immediate fixation. After 6h, and time for movement, this difference is erased. (I) Summary schematic for organoid niche structure. Scale bar; (F) 100, (G) 30, (H) 45 um. And see Figure S6.

Methods

Fluorescence in situ hybridization (ISH)

Two color fluorescence ISH was performed as previously described (Ji *et al.*, 2021; Kocabas *et al.*, 2015), using the TSA Plus Cyanine 3 & Fluorescein (NEL753001KT, PerkinElmer, Waltham, MA) according to the manufacturer's instructions. Briefly, cells in culture were fixed with 4% paraformaldehyde (PFA) in PBS for 15 min, treated with proteinase K (1ug/ml) for 5 min, acetylation (1% triethanolamine and 0.25% acetic anhydride) for 10 min, prehybridization (50% formamide, 5X SSC, 5X Denhardt's solution, 0.5 mg/ml Herring sperm DNA and 250 μg/ml Yeast tRNA) for 10 min, then hybridized with fluorescein labeled CDS and Digoxygenin (DIG) labeled probes 3UTR probes diluted in prehybridization buffer (0.5 ng/ul) at 56°C for 3 hours, proper washes were carried out accordingly between each steps. Specimens were then sequentially post-stained with fluorescein or Cy3 chromogens, respectively, and the corresponding Alexa Fluor 647 secondary antibodies and Hoechst 33342 for DNA staining, and mounted with Fluoromount-G mounting medium (SouthernBiotech catlog #0100-01). Antibodies are listed in Table S1, ISH probes in Table S2.

mESC and hESC Cell culture

Strain 129 mouse embryonic stem cells (mESCs) were obtained from previous studies (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2747754/) cultured in E14 plus complete medium (Glasgow Minimum Essential medium(Sigma-Aldrich G5154-6) supplemented with 10% heat inactivated FBS (Fisher Scientific SH3007103), 0.1 mM non-essential amino acids (Thermo Fisher Scientific 11140050), 1 mM sodium pyruvate (Thermo Fisher Scientific 11360070), 1x Pen/Strep (Thermo Fisher Scientific 15140122), 1x β-mercaptoethanol (55uM) (Thermo Fisher Scientific 21985023), 1,000 units/ml Leukemia Inhibitory Factor (LIF) (Millipore Sigma ESG1106), and freshly added glutamax (Thermo Fisher Scientific catalog

#35050061) (final concentration of 1X), 1 uM PD03259010 (Reprocell 04-0006-10) and 3 uM CHIR99021 (Cellagen Technology C2447-50)), and used within one week. mESCs are maintained and split at 1:6 ratio with 2x Trypsin every two days, re-plated in 6 well plates precoated with 0.1% gelatin (Sigma Millipore ES-006-B) for 30 minutes at 37C.
H7 human embryonic stem cells (hESCs) were incubated at 37 C, 5% CO2 in mTeSR1 plus (STEMCELL Technologies 85850) complete medium (mTeSR1 basal medium with supplement made according to manufacture protocol). hESCs are obtained from WiCell (RRID:CVCL_9772) maintained and split from around 90% confluence with 1x versene (Fisher Scientific 15-040-066) for 5 minutes at RT and then split with 1:8 ratio into 6 well plates pre-coated with 1:200 diluted geltrex in DMEM/F12 medium (Thermo Scientific A1413201) for 30 minutes at 37C. All cells are incubated at 37 C with 5% CO2.

Lentivirus production

Lentiviruses of the sgRNAs, rtTA, and pLV overexpression constructs were obtained by co-transfection of library plasmids with two viral packaging plasmids psPAX2(RRID:Addgene_12260) and pMD2.G (RRID:Addgene_ 12259) into HEK293T using jetPRIME transfection reagents (VWR 89129-926) according to manufactory standard protocol. Human embryonic kidney (HEK293) cell line, originally obtained from ATCC (ATCC Cat# CRL-3216, RRID:CVCL_0063), was grown in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific 11960069) supplemented with 10% fetal bovine serum (FBS; Gibco10437028), and 1X Penicillin Streptomycin Glutamine (Fisher Scientific 10-378-016)

Micropattern cell culture

Micropatten CYTOO (Arena A, CYTOO, France) chip cell culture was performed as previously described us (Ji *et al*., 2021), and others (Deglincerti et al., 2016; Martyn et al., 2019; Morgani et al., 2018) who used the chips to examine differentiated cells in germ layers. Patterns of 1000, 500, 250, 120 and 80 um circles are dispersed over the full surface. Briefly, for mESC, 700 ul drops of 20 ug/ml mouse Laminin (Sigma, L20202) in PBS without calcium and magnesium (PBS-/-) were spotted onto parafilm sitting in a 15 cm tissue culture plate. After one wash in PBS-/-, chips were inverted on top of the drops for 2h at 37°C. They were then washed 5x with PBS-/- and a single cell suspension of mESC cells (2 x 106) was evenly plated onto chips grown in 6-well plates for three days in complete E14 medium before fixation and ISH. For hESC, 700 ul drops of 5 ug/ml human Laminin (LN521)(Biolamina) in PBS with calcium and magnesium (PBS+/+) in were spotted onto parafilm in a 15 cm tissue culture plate. After one wash in PBS+/+, chips were inverted on top of the drops for 2h at 37°C, then washed 5x with PBS+/+ and a single cell suspension of mESC cells (2 x 106) was evenly plated onto chips, grown in 6-well plates for three days in complete mTeSR medium before fixation and ISH.

Mouse and human embryos

Human blastocysts originated from the Stanford Fertility and Reproductive Health Center in excess of reproductive need. The Blastocysts were donated to the RENEW Biobank at Stanford University (IRB10466) after In Vitro Fertilization, embryo culture and vitrification on Day 5 or 6. As approved by a SCRO protocol (SCRO-795), vitrified Blastocysts were warmed using Kitazato vitrification/warming protocol, allowing 1-2hrs for re-expansion prior to them being fixed.

Mouse embryos were obtained from Stanford transgenic, knockout and tumor model center. Briefly, Zygotes were made from in vitro fertilization (IVF), and then zygotes were transferred to 30 µl drops of culture medium covered with sterile mineral oil on a culture dish and placed in culture incubator at 37°C under 5% CO2 until embryos reach morula or blastocyst stages.
ISH and antibody stain was carried out with methods described above with the modification that each embryo was processed within a series of single droplets of 15ul hybridization and wash solutions while situated on a microscope slide.

Cas9 mESC stable cell line and CRISPR gRNAs Knockout
gRNAs targeting different gene CDS and 3UTR regions were designed individually using IDT Custom Alt-R CRISPR-Cas9 guide RNA design tool, specific gRNAs were selected for both their on-target potential score and off-target risk score. gRNAs were synthesized by IDT, gRNAs then were cloned into modified pL-CRISPR.EFS.GFP construct (RRID:Addgene_57818) with removal of Cas9, and GFP was replaced by tdTomato for 3' targeting gRNAs. 3UTRs and Cas9 resistant rescue *Nanog* 3UTR were cloned into pLKO-H2BmRFP1 construct, a gift from Dr. Fuchs lab at the Rockefeller University. The Cas9 expressing construct (RRID:Addgene_52962) and rtTA construct (RRID:Addgene_61472) was obtained from addgene.
A Cas9 mESC stable cell line was made by lentiviral infection of mESCs with lentiCas9-Blast, screened for single colonies by blasticidin treatment (10 ug/ml) and evaluated by PCR. Lentivirus was made according to lentivirus production section, and two different gRNAs with a total of 2x10e7 IU, along with polybrene, were used to infect the cells for knockout/knockdown. Knockout efficiency examined by PCR of genomic DNA with primers targeting the deleted region after lentivirus infection. For rescue experiment, equal amounts of Cas9 resistant *Nanog* 3UTR virus, along with 5x10e6 IU rtTA virus, were delivered at the same time as the knockout gRNA virus.

Colony formation assay
mESCs infected with gRNA pairs carrying either EGFP or tdTomato, were sorted after 4 days by flow cytometry for double positive cells. Cells were collected and plated at 1000 cells /24-well plate well and grown for an additional five days before fixation for imaging and quantification.

Live imaging of Cultured Cells
mESCs were grown in 24-well plates in E14 plus complete medium. Time-lapse imaging was carried out using a computer-assisted fluorescence microscope (NIKON) equipped with an objective lens, LED and a CCD camera. For DIC and fluorescent imaging, images were acquired every 20 minutes for 24 hours using Nikon NIS-elements AR software, exported as mp4 video files and analyzed in Fiji (Fiji, RRID:SCR_002285).

Inhibitors and activin
Activin/NODAL/TGF-β pathway inhibitors A 83-01 (Cayman Chemical Company 9001799) inhibit ALK5, ALK4, and ALK7. Histone modification inhibitors CPTH6 (Cayman Chemical Company 19828) inhibits the lysine acetyltransferase activity of Gcn5 and pCAF, and A-485 (Cayman Chemical Company 24119) that inhibits p300/CBP. All were dissolved in DMSO. The dose response for their specific effects measured in bulk cultured mESCs and hESCs, by either

qPCR or immunofluorescent staining. The inhibitors and activin (STEMCELL Technologies 78001) are used at optimized concentration (10uM for A 83-01 and CPTH6, 33 uM for A-485, 40 ng/ml for activin). Inhibitors are added to Cytoo chip cultured cells 24 hours after plating cells for 48 hours followed by fixation and in situ analysis.

qPCR

RNA was extracted using the Qiagen RNeasy plus mini kit (Qiagen 74134) according to the manufacturer's instruction. The 2 µg of total RNA was reverse transcribed to cDNA using the SuperScript IV (Thermo Fisher Scientific, 18091050) according to the manufacturer's instruction. And then qPCR assays were performed on the LightCycler 480 II (Roche) in 10µL reaction volume containing 6µL of SYBR green Master mix (Fisher Scientific A25918), 2µL of diluted cDNA (20 ng) and 1µL (300 nmol) each of a gene-specific forward and reverse primer. The following standard PCR reaction conditions were used for all transcripts: 95 °C 2 min; 40 cycles of 95 °C 10 s, 60 °C 1 min; 1 cycle of 95 °C 15 s, 60 °C 15 s, 95 °C 15 s. All PCR sample reactions had four replicates. The experiment was repeated four times to ensure reproducibility. For data analysis, we adopted the $2^{-\Delta\Delta}$ threshold cycle (Ct) method. The Ct of each target gene was first normalized to the Ct of GAPDH in each sample and then to the corresponding value in the control sample and expressed as log2.

Quantiation

Quantitative analyses were done as described in the text or figure legends. Statistical tests include Student t-test and X square analyses. $P < 0.05$ = *, $p < 0.01$=**, $p<0.001$= ***. For FACS analysis Y axis is forward scatter, and X axis is measured protein level; gRNAs infected mESCs were sorted and stained for Nanog protein 8 days after infection, Nanog protein was stained by antibody, examined by FACS, analyzed in FlowJo with a selection of single cells and GFP+tdT+ cells first.

Bioinformatic analyses

3UTR/CDS ratio quantification

Raw data for hESC and mESC scRNA seq are downloaded from GEO database under accession number GSE75748 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75748) and ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-2600. Quantification of 3UTR and CDS expression are as described in (Kocabas *et al.*, 2015). To obtain expression matrices for 3UTR and CDS respectively, we quantified gene average coverage for 3UTR or CDS separately using bedtools (v2.17.0) function coveragebed with the alignment file (BAM). Genes with coverage <= 5 mapped reads of both CDS and UTR are removed. CDS or 3UTR expression were normalized to their library size (total mapped reads) and quantified as transcripts per base pair.

The ratio of 3UTR/CDS is defined as:

$$\text{3'UTR/CDS ratio} = \frac{\text{UTR expression}}{\text{UTR expression + CDS expression}}$$

Read processing and mapping
Reads were mapped against hg38 or mm10 using STAR (v2.7.3a) with default parameters. Gene expression was quantified using RSEM (v1.3.1) as expected raw counts and TPM with default parameters.

Cell clustering
We applied R package "Seurat" to analye single-cell data and excluded poor quality cells as described in (Chu *et al.*, 2016; Kolodziejczyk *et al.*, 2015). Cells are clustered with tSNE and the number of clusters was automatically chosen by Seurat.

GO category enrichment
Genes enriched for each cluster were chosen as p_value >= 0.5 and log2_fold_change>=0.5. GO category enrichment for enriched genes are done using R package "Enrichplot" with "Biological Process".

References

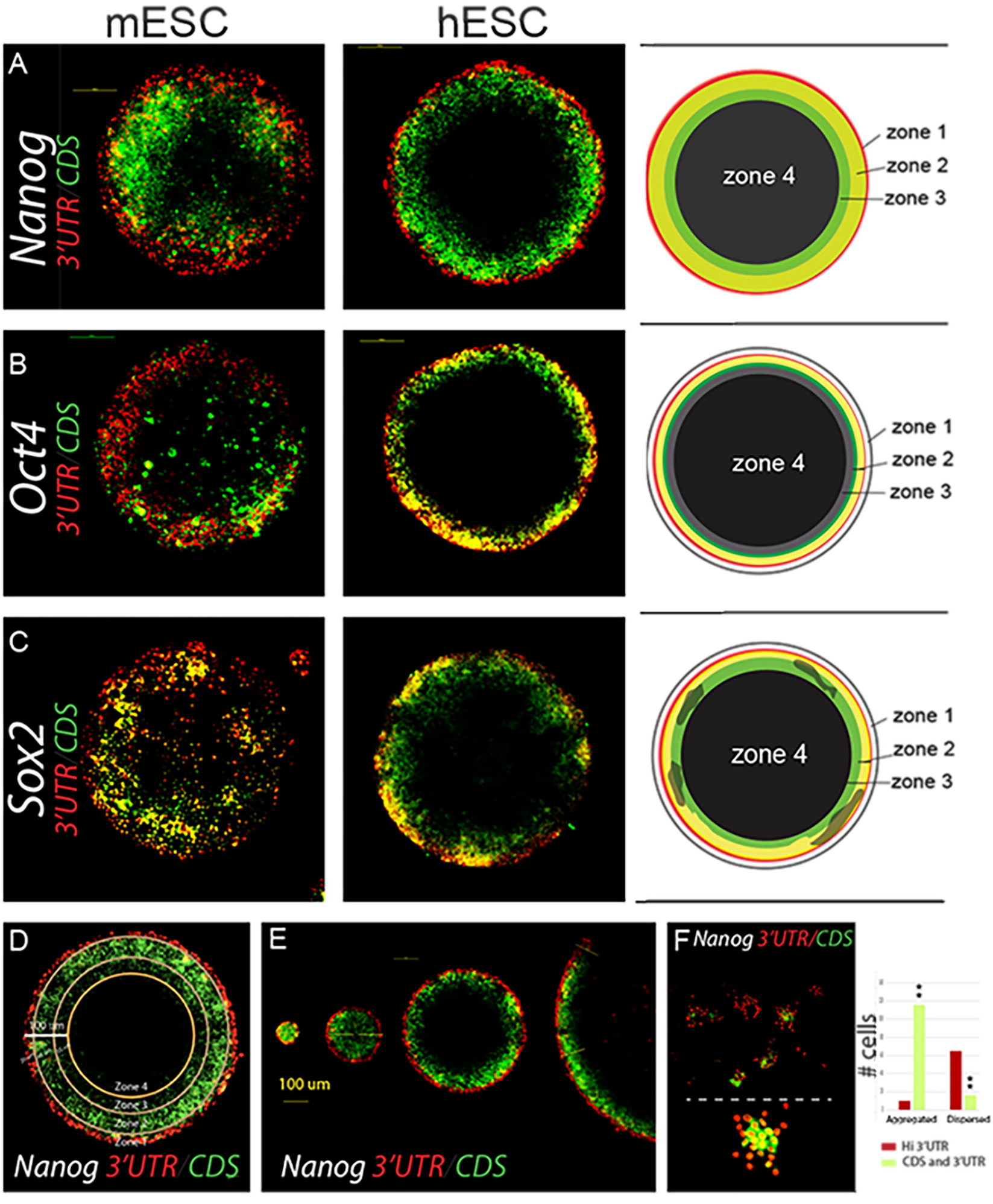
Andersson, O., Korach-Andre, M., Reissmann, E., Ibanez, C.F., and Bertolino, P. (2008). Growth/differentiation factor 3 signals through ALK7 and regulates accumulation of adipose tissue and diet-induced obesity. Proceedings of the National Academy of Sciences of the United States of America *105*, 7252-7256. 10.1073/pnas.0800272105.
Bauwens, C.L., Peerani, R., Niebruegge, S., Woodhouse, K.A., Kumacheva, E., Husain, M., and Zandstra, P.W. (2008). Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. Stem Cells *26*, 2300-2310. 10.1634/stemcells.2008-0183.
Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev Cell *35*, 366-382. 10.1016/j.devcel.2015.10.011.
Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956. 10.1016/j.cell.2005.08.020.
Bradford, G.B., Williams, B., Rossi, R., and Bertoncello, I. (1997). Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. Exp Hematol *25*, 445-453.
Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311-2322. 10.1242/dev.024398.
Chu, L.F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D.T., Choi, J., Kendziorski, C., Stewart, R., and Thomson, J.A. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol *17*, 173. 10.1186/s13059-016-1033-x.
Davey, R.E., and Zandstra, P.W. (2006). Spatial organization of embryonic stem cell responsiveness to autocrine gp130 ligands reveals an autoregulatory stem cell niche. Stem Cells *24*, 2538-2548. 10.1634/stemcells.2006-0216.

Deglincerti, A., Etoc, F., Guerra, M.C., Martyn, I., Metzger, J., Ruzo, A., Simunovic, M., Yoney, A., Brivanlou, A.H., Siggia, E., and Warmflash, A. (2016). Self-organization of human embryonic stem cells on micropatterns. Nat Protoc *11*, 2223-2232. 10.1038/nprot.2016.131.
Ebrahimi, A., Sevinc, K., Gurhan Sevinc, G., Cribbs, A.P., Philpott, M., Uyulur, F., Morova, T., Dunford, J.E., Goklemez, S., Ari, S., et al. (2019). Bromodomain inhibition of the coactivators CBP/EP300 facilitate cellular reprogramming. Nat Chem Biol *15*, 519-528. 10.1038/s41589-019-0264-z.
Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477-490. 10.1016/j.stem.2012.08.002.
Festuccia, N., Osorno, R., Wilson, V., and Chambers, I. (2013). The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states. Curr Opin Genet Dev *23*, 504-511. 10.1016/j.gde.2013.06.003.
Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature genetics *37*, 391-400. 10.1038/ng1531.
Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-300. 10.1038/nature10398.
Herrera, S.C., Sainz de la Maza, D., Grmai, L., Margolis, S., Plessel, R., Burel, M., O'Connor, M., Amoyel, M., and Bach, E.A. (2021). Proliferative stem cells maintain quiescence of their niche by secreting the Activin inhibitor Follistatin. Dev Cell *56*, 2284-2294 e2286. 10.1016/j.devcel.2021.07.010.
Hsu, Y.C., Li, L., and Fuchs, E. (2014). Transit-amplifying cells orchestrate stem cell activity and tissue regeneration. Cell *157*, 935-949. 10.1016/j.cell.2014.02.057.
Ji, S., Yang, Z., Gozali, L., Kenney, T., Kocabas, A., Jinsook Park, C., and Hynes, M. (2021). Distinct expression of select and transcriptome-wide isolated 3UTRs suggests critical roles in development and transition states. PLoS One *16*, e0250669. 10.1371/journal.pone.0250669.
Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes Dev *19*, 1129-1155. 10.1101/gad.1303605.
Knight, G.T., Lundin, B.F., Iyer, N., Ashton, L.M., Sethares, W.A., Willett, R.M., and Ashton, R.S. (2018). Engineering induction of singular neural rosette emergence within hPSC-derived tissues. Elife *7*. 10.7554/eLife.37549.
Kocabas, A., Duarte, T., Kumar, S., and Hynes, M.A. (2015). Widespread Differential Expression of Coding Region and 3' UTR Sequences in Neurons and Other Tissues. Neuron *88*, 1149-1156. 10.1016/j.neuron.2015.10.048.
Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Buhler, M., Liu, P., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell *17*, 471-485. 10.1016/j.stem.2015.09.011.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903-911. 10.1634/stemcells.2007-0951.
Krens, S.F., and Heisenberg, C.P. (2011). Cell sorting in development. Curr Top Dev Biol *95*, 189-213. 10.1016/B978-0-12-385065-2.00006-2.
Krieger, T., and Simons, B.D. (2015). Dynamic stem cell heterogeneity. Development (Cambridge, England) *142*, 1396-1406. 10.1242/dev.101063.
Levine, A.J., and Brivanlou, A.H. (2006). GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. Development (Cambridge, England) *133*, 209-216. 10.1242/dev.02192.
Li, L., and Xie, T. (2005). Stem cell niche: structure and function. Annu Rev Cell Dev Biol *21*, 605-631. 10.1146/annurev.cellbio.21.012704.131525.
Li, M., and Belmonte, J.C. (2017). Ground rules of the pluripotency gene regulatory network. Nat Rev Genet *18*, 180-191. 10.1038/nrg.2016.156.
Li, Y.Q. (2017). Networks of Transcription Factors for Oct4 Expression in Mice. DNA Cell Biol *36*, 725-736. 10.1089/dna.2017.3780.
Lilly, A.J., Johnson, W.E., and Bunce, C.M. (2011). The haematopoietic stem cell niche: new insights into the mechanisms regulating haematopoietic stem cell behaviour. Stem Cells Int *2011*, 274564. 10.4061/2011/274564.
Lin, H. (2002). The stem-cell niche theory: lessons from flies. Nat Rev Genet *3*, 931-940. 10.1038/nrg952.
Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440. 10.1038/ng1760.
Malka, Y., Steiman-Shimony, A., Rosenthal, E., Argaman, L., Cohen-Daniel, L., Arbib, E., Margalit, H., Kaplan, T., and Berger, M. (2017). Post-transcriptional 3 -UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. Nat Commun *8*, 2029. 10.1038/s41467-017-02099-7.
Martyn, I., Brivanlou, A.H., and Siggia, E.D. (2019). A wave of WNT signaling balanced by secreted inhibitors controls primitive streak formation in micropattern colonies of human embryonic stem cells. Development *146*. 10.1242/dev.172791.
Mercer, T.R., Wilhelm, D., Dinger, M.E., Solda, G., Korbie, D.J., Glazov, E.A., Truong, V., Schwenke, M., Simons, C., Matthaei, K.I., et al. (2011). Expression of distinct RNAs from 3' untranslated regions. Nucleic Acids Res *39*, 2393-2403. 10.1093/nar/gkq1158.
Morgani, S.M., Metzger, J.J., Nichols, J., Siggia, E.D., and Hadjantonakis, A.K. (2018). Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. Elife *7*. 10.7554/eLife.32839.
Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr Opin Genet Dev *55*, 1-10. 10.1016/j.gde.2019.04.013.
Pagella, P., Neto, E., Lamghari, M., and Mitsiadis, T.A. (2015). Investigation of orofacial stem cell niches and their innervation through microfluidic devices. Eur Cell Mater *29*, 213-223. 10.22203/ecm.v029a16.
Peerani, R., Rao, B.M., Bauwens, C., Yin, T., Wood, G.A., Nagy, A., Kumacheva, E., and Zandstra, P.W. (2007). Niche-mediated control of human embryonic stem cell self-renewal and differentiation. EMBO J *26*, 4744-4755. 10.1038/sj.emboj.7601896.

Rabani, M., Pieper, L., Chew, G.L., and Schier, A.F. (2017). A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol Cell *68*, 1083-1094 e1085. 10.1016/j.molcel.2017.11.014.
Raz, A.A., and Yamashita, Y.M. (2021). Stem cell niche signaling goes both ways. Dev Cell *56*, 2267-2268. 10.1016/j.devcel.2021.08.003.
Risom, T., Glass, D.R., Averbukh, I., Liu, C.C., Baranski, A., Kagel, A., McCaffrey, E.F., Greenwald, N.F., Rivero-Gutierrez, B., Strand, S.H., et al. (2022). Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. Cell *185*, 299-310 e218. 10.1016/j.cell.2021.12.023.
Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem *280*, 24731-24737. 10.1074/jbc.M502573200.
Rossant, J. (2007). Stem cells and lineage development in the mammalian blastocyst. Reprod Fertil Dev *19*, 111-118. 10.1071/rd06125.
Saunders, A., Faiola, F., and Wang, J. (2013). Concise review: pursuing self-renewal and pluripotency with the stem cell factor Nanog. Stem Cells *31*, 1227-1236. 10.1002/stem.1384.
Schofield, R. (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. Blood Cells *4*, 7-25.
Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. J Biol Chem *281*, 23319-23325. 10.1074/jbc.M601811200.
Spradling, A., Drummond-Barbosa, D., and Kai, T. (2001). Stem cells find their niche. Nature *414*, 98-104. 10.1038/35102160.
Sudmant, P.H., Lee, H., Dominguez, D., Heiman, M., and Burge, C.B. (2018). Widespread Accumulation of Ribosome-Associated Isolated 3' UTRs in Neuronal Cell Populations of the Aging Brain. Cell reports *25*, 2447-2456 e2444. 10.1016/j.celrep.2018.10.094.
Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339-1349. 10.1242/dev.033951.
van den Brink, S.C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.K., Nowotschin, S., Turner, D.A., and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. Development (Cambridge, England) *141*, 4231-4242. 10.1242/dev.113001.
Van Den Broeck, A., Brambilla, E., Moro-Sibilot, D., Lantuejoul, S., Brambilla, C., Eymin, B., and Gazzeri, S. (2008). Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. Clin Cancer Res *14*, 7237-7245. 10.1158/1078-0432.CCR-08-0869.
Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. (2000). Global histone acetylation and deacetylation in yeast. Nature *408*, 495-498. 10.1038/35044127.
Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol *17*, 155-169. 10.1038/nrm.2015.28.
Weissman, I.L. (1994). Developmental switches in the immune system. Cell *76*, 207-218. 10.1016/0092-8674(94)90329-8.

Yang, S., Cho, Y., and Jang, J. (2021). Single cell heterogeneity in human pluripotent stem cells. BMB Rep *54*, 505-515.


mESC
hESC
A
Nanog
3'UTR/CDS
zone 1
zone 2
zone 3
zone 4
B
Oct4
3'UTR/CDS
zone 1
zone 2
zone 3
zone 4
C
Sox2
3'UTR/CDS
zone 1
zone 2
zone 3
zone 4
D
Zone 4
Zone 3
Nanog 3'UTR/CDS
E
100 um
Nanog 3'UTR/CDS
F Nanog 3'UTR/CDS
cells
Aggregated
Dispersed
Hi 3'UTR
CDS and 3'UTR

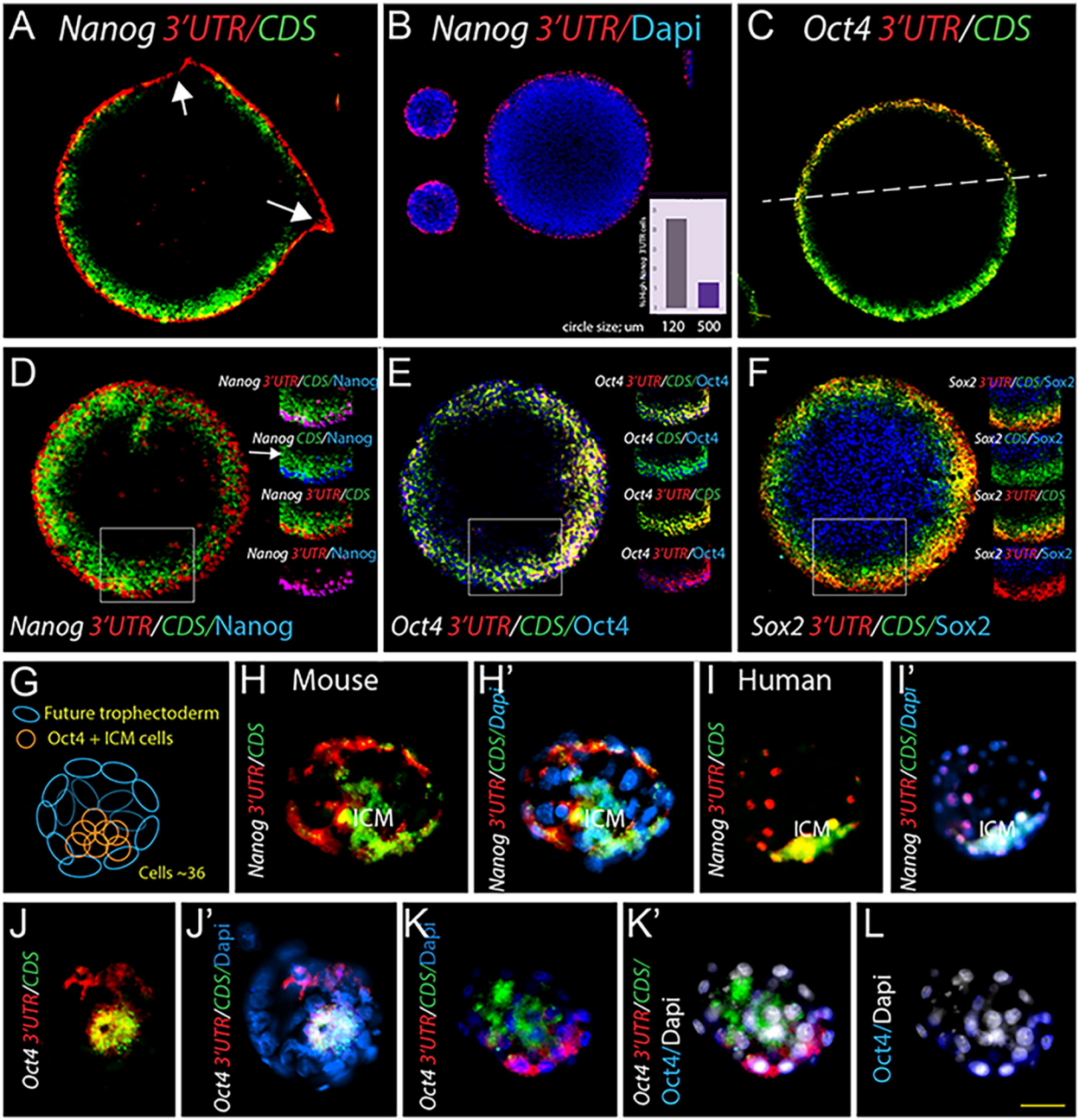

A Nanog 3'UTR/CDS
B Nanog 3'UTR/Dapi
circle size; um 120 500
C Oct4 3'UTR/CDS
D
Nanog 3'UTR/CDS/Nanog
Nanog CDS/Nanog
Nanog 3'UTR/CDS
Nanog 3'UTR/Nanog
Nanog 3'UTR/CDS/Nanog
E
Oct4 3'UTR/CDS/Oct4
Oct4 CDS/Oct4
Oct4 3'UTR/CDS
Oct4 3'UTR/Oct4
Oct4 3'UTR/CDS/Oct4
F
Sox2 3'UTR/CDS/Sox2
Sox2 CDS/Sox2
Sox2 3'UTR/CDS
Sox2 3'UTR/Sox2
Sox2 3'UTR/CDS/Sox2
G
Future trophectoderm
Oct4 + ICM cells
Cells ~36
H Mouse
Nanog 3'UTR/CDS
ICM
H'
Nanog 3'UTR/CDS/Dapi
ICM
I Human
Nanog 3'UTR/CDS
ICM
I'
Nanog 3'UTR/CDS/Dapi
ICM
J
Oct4 3'UTR/CDS
J'
Oct4 3'UTR/CDS/Dapi
K
Oct4 3'UTR/CDS/Dapi
K'
Oct4 3'UTR/CDS/
Oct4/Dapi
L
Oct4/Dapi

Fig 2

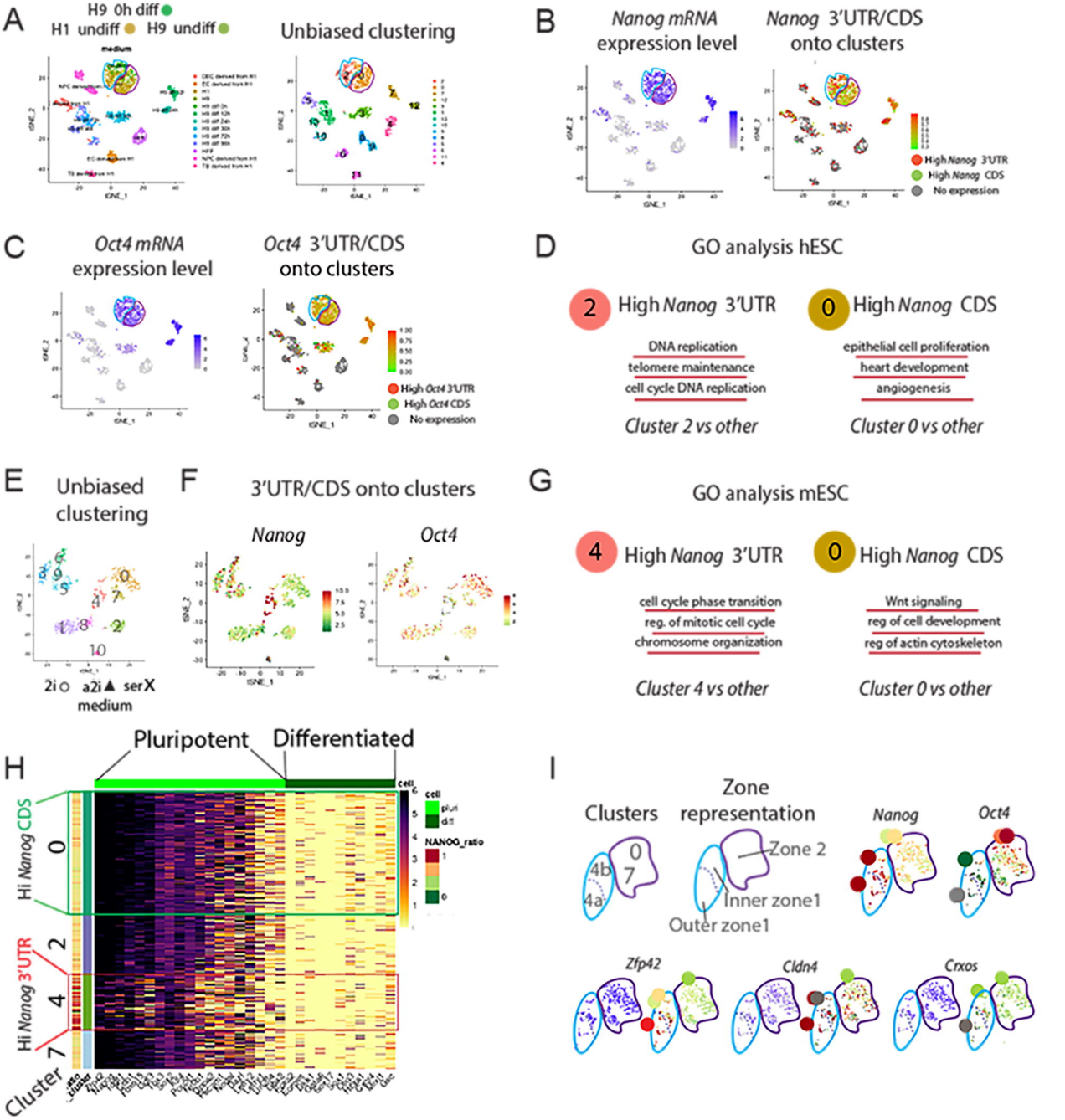
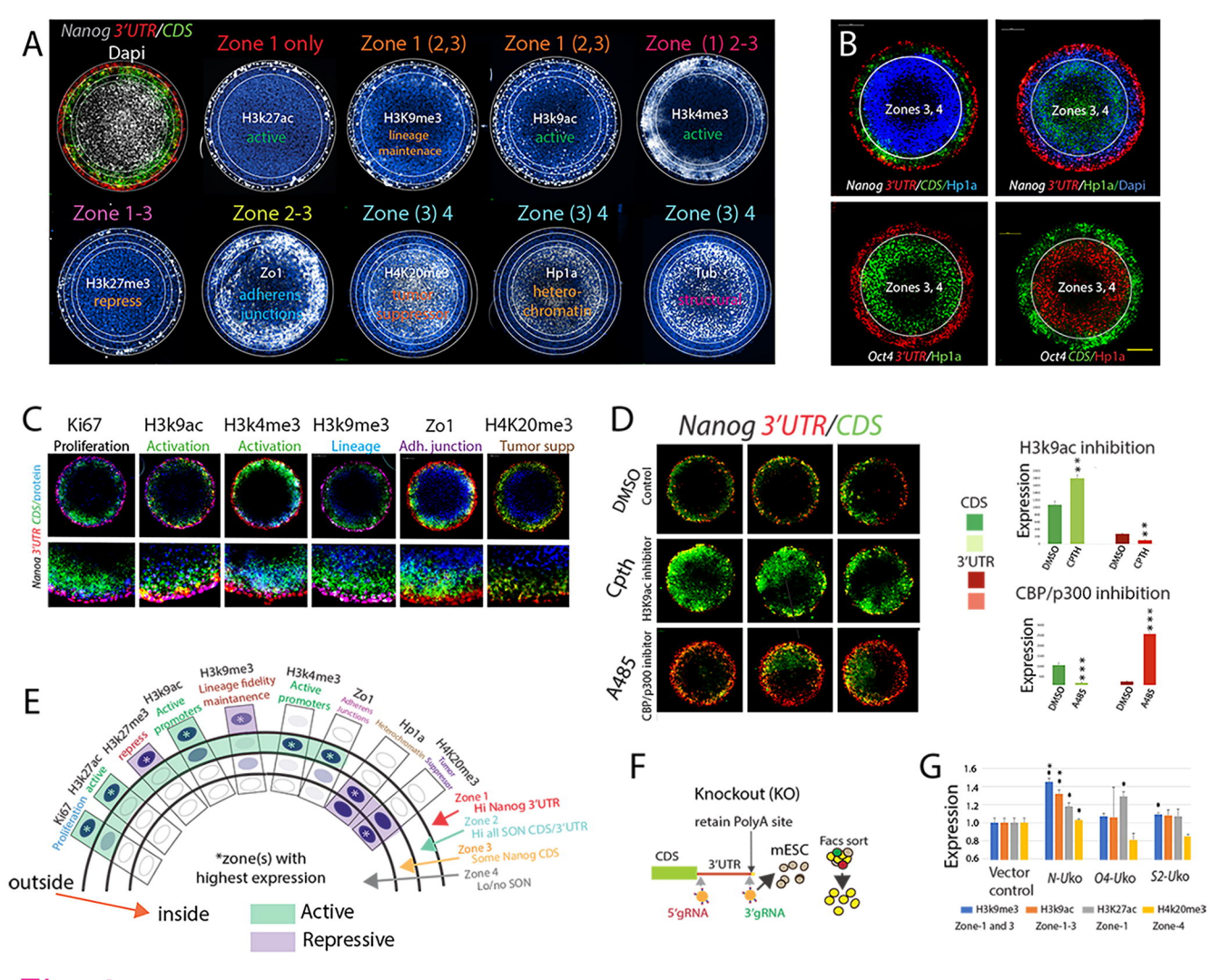
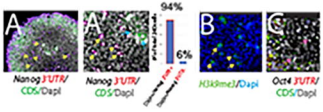
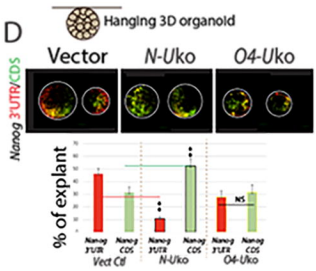
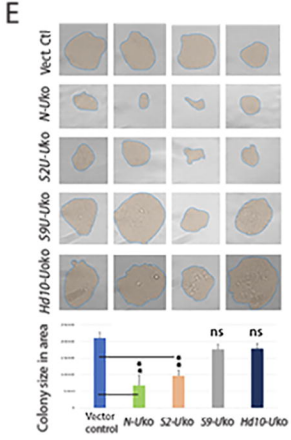

A
H9 0h diff
H1 undiff
H9 undiff
medium
Unbiased clustering
B
Nanog mRNA expression level
Nanog 3'UTR/CDS onto clusters
High Nanog 3'UTR
High Nanog CDS
No expression
C
Oct4 mRNA expression level
Oct4 3'UTR/CDS onto clusters
High Oct4 3'UTR
High Oct4 CDS
No expression
D
GO analysis hESC
2 High Nanog 3'UTR
DNA replication
telomere maintenance
cell cycle DNA replication
Cluster 2 vs other
0 High Nanog CDS
epithelial cell proliferation
heart development
angiogenesis
Cluster 0 vs other
E
Unbiased clustering
2i
a2i
serX
medium
F
3'UTR/CDS onto clusters
Nanog
Oct4
G
GO analysis mESC
4 High Nanog 3'UTR
cell cycle phase transition
reg. of mitotic cell cycle
chromosome organization
Cluster 4 vs other
0 High Nanog CDS
Wnt signaling
reg of cell development
reg of actin cytoskeleton
Cluster 0 vs other
H
Pluripotent
Differentiated
Hi Nanog CDS
Hi Nanog 3'UTR
Cluster
NANOG_ratio
I
Clusters
Zone representation
Zone 2
Inner zone1
Outer zone1
Nanog
Oct4
Zfp42
Cldn4
Crxos

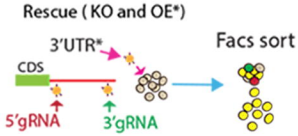
Fig 3

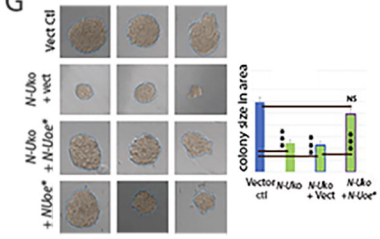

A
Nanog 3'UTR/CDS
Dapi
Zone 1 only
H3k27ac
active
Zone 1 (2,3)
H3K9me3
lineage
maintenace
Zone 1 (2,3)
H3k9ac
active
Zone (1) 2-3
H3k4me3
active
Zone 1-3
H3k27me3
repress
Zone 2-3
Zo1
adherens
junction
Zone (3) 4
H4K20me3
tumor
suppressor
Zone (3) 4
Hp1a
hetero-
chromatin
Zone (3) 4
Tub
structural
B
Zones 3, 4
Nanog 3'UTR/CDS/Hp1a
Zones 3, 4
Nanog 3'UTR/Hp1a/Dapi
Zones 3, 4
Oct4 3'UTR/Hp1a
Zones 3, 4
Oct4 CDS/Hp1a
C
Ki67
Proliferation
H3k9ac
Activation
H3k4me3
Activation
H3k9me3
Lineage
Zo1
Adh. junction
H4K20me3
Tumor supp
Nanog 3'UTR CDS/protein
D
Nanog 3'UTR/CDS
DMSO
Control
Cpth
H3K9ac inhibitor
A485
CBP/p300 inhibitor
CDS
3'UTR
H3k9ac inhibition
Expression
DMSO
CPTH
CBP/p300 inhibition
Expression
DMSO
A485
E
Ki67
Proliferation
H3k27ac
active
H3k27me3
repress
H3k9ac
Active
promoters
H3k9me3
Lineage fidelity
maintanence
H3k4me3
Active
promoters
Zo1
Adherens
junctions
Hp1a
Heterochromatin
H4K20me3
Tumor
Suppressor
Zone 1
Hi Nanog 3'UTR
Zone 2
Hi all SON CDS/3'UTR
Zone 3
Some Nanog CDS
Zone 4
Lo/no SON
*zone(s) with
highest expression
outside
inside
Active
Repressive
F
Knockout (KO)
retain PolyA site
CDS
3'UTR
mESC
Facs sort
5'gRNA
3'gRNA
G
Expression
1.6
1.4
1.2
1.0
0.8
0.6
Vector
control
N-Uko
O4-Uko
S2-Uko
H3k9me3
H3k9ac
H3K27ac
H4k20me3
Zone-1 and 3
Zone-1-3
Zone-1
Zone-4

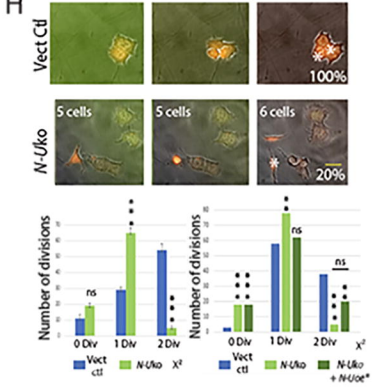

A
A'
B
C
94%
6%
Nanog 3'UTR/ CDS/Dapi
Nanog 3'UTR/ CDS/Dapi
H3K9me3/Dapi
Oct4 3'UTR/ CDS/Dapi

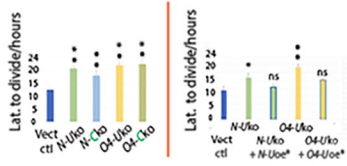

D
Hanging 3D organoid
Vector
N-Uko
O4-Uko
Nanog 3'UTR/CDS
% of explant
Vect Ctl
N-Uko
O4-Uko
ns


E
Vect Ctl
N-Uko
S2U-Uko
S9U-Uko
Hd10-Uoko
Colony size in area
Vector control
N-Uko
S2-Uko
S9-Uko
Hd10-Uko
ns
ns


F
Rescue (KO and OE*)
3'UTR*
CDS
5'gRNA
3'gRNA
Facs sort


G
Vect Ctl
N-Uko + vect
N-Uko + N-Uoe*
+ N-Uoe*
colony size in area
Vector ctl
N-Uko
N-Uko + Vect
N-Uko + N-Uoe*
ns


H
Vect Ctl
N-Uko
100%
5 cells
5 cells
6 cells
20%
Number of divisions
0 Div
1 Div
2 Div
ns
Vect ctl
N-Uko
N-Uko + N-Uoe*


I
Lat. to divide/hours
Vect ctl
N-Uko
N-Cko
O4-Uko
O4-Cko
N-Uko + N-Uoe*
O4-Uko + O4-Uoe*
ns

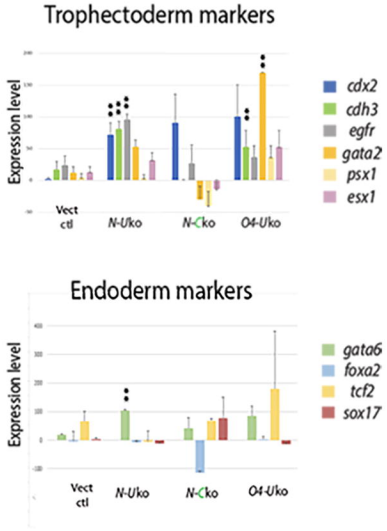
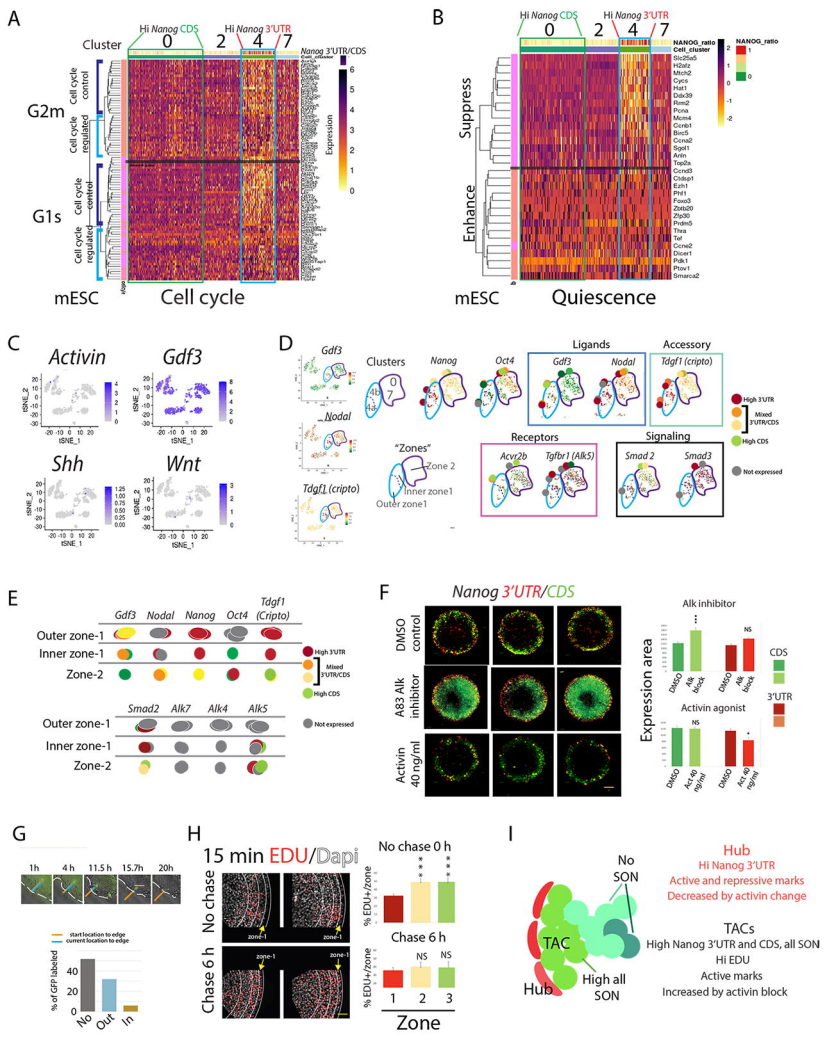

J
Trophectoderm markers
Expression level
cdx2
cdh3
egfr
gata2
psx1
esx1
Vect ctl
N-Uko
N-Cko
O4-Uko
Endoderm markers
gata6
foxa2
tcf2
sox17

Fig 5


A
Hi Nanog CDS
Hi Nanog 3'UTR
Cluster
0
2
4
7
Nanog 3'UTR/CDS
Expression
G2m
G1s
Cell cycle control
Cell cycle regulated
mESC
Cell cycle
B
Suppress
Enhance
NANOG_ratio
Cell_cluster
mESC
Quiescence
C
Activin
Gdf3
Shh
Wnt
tSNE_1
tSNE_2
D
Gdf3
Nodal
Tdgf1 (cripto)
Clusters
Nanog
Oct4
Ligands
Gdf3
Nodal
Accessory
Tdgf1 (cripto)
"Zones"
Zone 2
Inner zone1
Outer zone1
Receptors
Acvr2b
Tgfbr1 (Alk5)
Signaling
Smad 2
Smad3
High 3'UTR
Mixed 3'UTR/CDS
High CDS
Not expressed
E
Gdf3
Nodal
Nanog
Oct4
Tdgf1 (Cripto)
Outer zone-1
Inner zone-1
Zone-2
Smad2
Alk7
Alk4
Alk5
F
Nanog 3'UTR/CDS
DMSO control
A83 Alk inhibitor
Activin 40 ng/ml
Alk inhibitor
Activin agonist
Expression area
CDS
3'UTR
G
1h
4 h
11.5 h
15.7h
20h
start location to edge
current location to edge
% of GFP labeled
No
Out
In
H
15 min EDU/Dapi
No chase
Chase 6 h
zone-1
No chase 0 h
Chase 6 h
% EDU+/zone
Zone
I
No SON
TAC
Hub
High all SON
Hub
Hi Nanog 3'UTR
Active and repressive marks
Decreased by activin change
TACs
High Nanog 3'UTR and CDS, all SON
Hi EDU
Active marks
Increased by activin block