Generation and Trapping of a Mesoderm Biased State of Human Pluripotency

Authors:

Dylan Stavish[1], Charlotta Böiers[2,3], Christopher Price[1], Thomas J R Frith[1], Jason Halliwell[1], Ivana Barbaric[1], John Brown[2], Jonathon Carr[1], Chela James[2], Peter W Andrews[1] and Tariq Enver[2].

Addresses:

1. The Centre for Stem Cell Biology, Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield S10 2TN.

2. Stem Cell Laboratory, Department of Cancer Biology, University College London Cancer Institute, 72 Huntley St, London WC1E 6AG, United Kingdom

Address correspondence to:

Peter W Andrews: The Centre for Stem Cell Biology, Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield S10 2TN, UK; email: p.w.andrews@sheffield.ac.uk

Dylan Stavish: The Centre for Stem Cell Biology, Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield S10 2TN, UK; email: d.stavish@sheffield.ac.uk

Acknowledgements:

This work was supported by grants from the European Community's Sixth and Seventh Framework Programs (LSHG-CT-2006-018739 and FP7/2007-2013-602423, PluriMes) and the MRC through the UK Regenerative Medicine Platform (grant no. MR/L012537/1). CB was supported by the Swedish Research Council (no. 2015-00135) and Marie Sklodowska Curie Actions, Cofund, Project INCA (no. 600398). CJ was supported by Bloodwise. We thank Andrew Elefanty for providing HES3 and HES3-*MIXL1* lines, Roger Pedersen and Daniel Ortmann for the providing H9 T-Venus line and Paul Gokhale and Anestis Tsakiridis for advice and insight.

Contributions:

P.W.A. and T.E. conceived the project. D.S. designed and performed the majority of experiments and analysis with aid from C.B, C.P., T.J.R.F., J.H, I.B., J.B., J.C. and C.J. and wrote the manuscript with aid from P.W.A and T.E. C.B. performed single cell qPCR reactions and J.B and C.B performed RNA sequencing, sequence alignment by C.J.

ABSTRACT

We postulate that exit from pluripotency involves intermediates that retain pluripotency while simultaneously exhibiting lineage-bias. Using a *MIXL1* reporter, we explored mesoderm lineage-bias within the human pluripotent stem cell compartment. We identified a substate, which at the single cell level coexpresses pluripotent and mesodermal gene expression programs. Functionally these cells could initiate stem cell cultures and exhibited mesodermal bias in differentiation assays. By promoting mesodermal identity through manipulation of WNT signalling while preventing exit from pluripotency using lysophosphatidic acid, we could 'trap' and maintain cells in a lineage-biased stem cell state through multiple passages. These cells correspond to a normal state on the differentiation trajectory, the plasticity of which is evidenced by their reacquisition of an unbiased state upon removal of differentiation cues. The use of 'cross-antagonistic' signalling to trap pluripotent stem cell intermediates with different lineage-bias may have general applicability in the efficient production of cells for regenerative medicine.

INTRODUCTION

Research of pluripotent stem cells (PSC), whether embryonic stem (ES) cells or induced pluripotent stem (IPS) cells, has often centered on robustly maintaining pluripotency. Less well investigated is how cells leave the pluripotent stem cell compartment. This is central for a biological understanding of how differentiation is regulated and lineage is specified. Since the use of pluripotent cells in regenerative medicine requires the efficient production of differentiated derivatives, delineating the cellular trajectories by which cells transit from pluripotency into lineage commitment is key. The existence of developmental intermediates on this trajectory may, in part, account for the cellular heterogeneity exhibited by in vitro cultures of human PSC[1].

Heterogeneity has been described in respect of: (i) “Subsets” of cells with different self renewal potential [2, 3] identified based on surface antigen expression; and (ii) subsets with differential expression of lineage affiliated genes have been described; critically single cell studies have shown these to coexist with stem cell programs [4-6]. Importantly, experimental assays of differentiation have demonstrated that these patterns of gene expression reflect interconvertible substates that functionally encode differential lineage bias within the stem cell compartment [4, 5, 7]. Reporter gene strategies have been particularly useful in this regard. We recently showed, using a fluorescent reporter, that a substate of human PSC expressed the early endoderm related marker *GATA6* [4]. *GATA6* positive cells were able to regenerate long-term pluripotent cultures yet their spontaneous differentiation favored endodermal lineages. That study supports the notion – at least for endoderm- that these substates, which coexpress signatures of pluripotency and differentiation represent differentiation intermediates, may exist normally as transient states. Whether these exist for other lineages is not yet known. It is noteworthy that the pluripotent state of human PSC is itself also transient in development and is in effect “trapped” when human PSC are cultured in vitro. Recently, culture systems that aim to trap cells in an earlier “naïve” stage of development has been reported [8-10], but whether cells can be trapped further along a particular differentiation trajectory has been less well explored. Cross-antagonism of pro-pluripotency and pro-differentiation signaling in a controlled environment could potentially provide a strategy allowing for the propagation of a pluripotent intermediate with a specific lineage bias.

Herein we use a *MIXL1* reporter [11] to delineate the cellular trajectory from pluripotency to committed mesoderm progenitors. On this trajectory, we identify within the stem cell compartment a developmental intermediate that also exhibits mesoderm lineage bias. Using a cross antagonism strategy, we can trap and expand this intermediate while retaining its capacity to revert to an unbiased stem cell state. The identification and characterization of mesodermal biased pluripotent intermediates informs our understanding of how mesoderm

lineage specification occurs and could provide an attractive starting point for directed differentiation towards mesodermal derivatives.

RESULTS

Identification of substates expressing *MIXL1*-GFP in feeder based culture conditions.

Throughout this paper, we have used expression of GFP as a measure of the *MIXL1* transcriptional state, which we refer to throughout the manuscript as *MIXL1*. When cultures of the *MIXL1* reporter (HES3-*MIXL1*) were grown on mouse embryo fibroblast (MEF) feeder cells in medium containing Knock-out Serum Replacement (MEF/KOSR condition), they typically contained a subpopulation of 5% to 20% of cells that co-expressed *MIXL1* and SSEA-3, a surface antigen that we have previously used as a sensitive marker of undifferentiated stem cells [3, 12, 13] (Fig. 1a).

To examine the relationship of these *MIXL1*(+)/SSEA-3(+) cells to the other subpopulations of *MIXL1*(-)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(-) cells, putatively stem and differentiated cells, respectively, and to test whether they co-express other markers that are separately indicators of a differentiated or undifferentiated state, we carried out RNA sequencing (RNAseq) analysis of these three subpopulations, isolated by fluorescence-activated cell sorting (FACS). Principle component analysis (PCA) of the entire transcriptomes from these RNAseq data showed a clear separation of the three subpopulations (Fig. 1b). In the case of the *MIXL1*(+)/SSEA-3(+) and *MIXL1*(-)/SSEA-3(+) cells, data from two biological replicates of each clustered closely, whereas two replicates of the *MIXL1*(+)/SSEA-3(-) cells were more separated, perhaps reflecting a great heterogeneity and variability expected in populations of differentiated cells. On the other hand, when the PCA was carried out with only a subset of pluripotency associated genes the *MIXL1*(+)/SSEA-3(+) cluster moved closer to the *MIXL1*(-)/SSEA-3(+) cluster, especially with respect to PC1 which accounts for 89% of variance, whereas when the PCA was carried out with genes associated with mesoderm the

MIXL1(+)/SSEA-3(+) cluster moved closer to the *MIXL1*(+)/SSEA-3(-) cluster of putatively differentiated cells (Fig. 1b). These analyses suggest that an active pluripotency network might still be in place the *MIXL1*(+)/SSEA-3(+) cells, consistent with the possibility that they occupy a substate within the stem cell compartment.

To test this we analysed each subpopulation for the expression of a signature set of genes (Supplementary Table 1), which included 3 controls and 45 genes of interest, by single cell qPCR to assess whether the *MIXL1*(+)/SSEA-3(+) subpopulation includes individual cells that co-express pluripotency and mesoderm associated genes, and so may represent a transitional substate between that of pristine, pluripotent *MIXL1*(-)/SSEA-3(+) stem cells and that of mesodermally committed, differentiated *MIXL1*(+)/SSEA-3(-) cells. The 45 gene set included genes typically associated with the undifferentiated state, such as *OCT4*, *NANOG* and *SOX2*, and genes typically associated with early mesendoderm differentiation, such as *T(Brachyury)*, *EOMES* and *GATA6*. Using Monocle2 [14] we produced t-distributed stochastic neighbor embedding (t-SNE) analysis of these single cell transcriptomic data. This showed that the *MIXL1*(+)/SSEA-3(+) cells cluster separately, spanning the space between separate clusters of the *MIXL1*(-)/SSEA-3(**+**) and *MIXL1*(**+**)/SSEA-3(-) cells, though there was substantial heterogeneity in each cluster (Fig. 1c).

A more detailed analysis using a subset of the 48 genes that relate most closely to the pluripotent and differentiated states revealed two main clusters of cells, with the *MIXL1*(-)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(-) displaying an almost complete separation into Cluster A (Pluripotent) and Cluster B (Differentiated), respectively (Fig. 1d). By contrast, the *MIXL1*(+)/SSEA-3(+) subpopulation included cells in both of these clusters, with many cells expressing both pluripotency and differentiation associated genes (Fig. 1d). For example, many of the cells in the *MIXL1*(+)/SSEA-3(+) fraction co-expressed the pluripotency associated marker *SOX2* and the mesoderm associated marker *T* (Fig. 1e). Similar cells were not often seen in the *MIXL1*(-)/SSEA-3(+) or *MIXL1*(+)/SSEA-3(-) fractions. The analysis also highlights the clustering of expression of the pluripotency associated genes,

POU5F1, *SOX2*, *NANOG* and *DNMT3B,* whereas *MIXL1* clustered most closely with differentiation markers such as *T*, *FOXA2* and *CER1*. The expression of the full panel of genes assessed in shown in Supplementary Fig. 1. These data suggest that the *MIXL1*(+)/SSEA-3(+) cells do occupy a transitional substate in terms of gene expression, but do not identify whether that substate is part of the stem cell or differentiated cell compartment.

***The MIXL1*(+)/SSEA-3(+) subpopulation contains self renewing pluripotent stem cells**

To assess whether the *MIXL1*(+)/SSEA-3(+) subpopulation contains functional undifferentiated stem cells, we isolated 288 *MIXL1*(+)/SSEA-3(+) individual cells by FACS and single cell deposition. From these we were able to grow out 47 colonies. Indexed FACS data confirmed the *MIXL1* and SSEA-3 expression levels of the individual cells from which these colonies were obtained (Fig. 2a). All of these 47 clonal colonies exhibited a typical rounded human ES cell colony morphology containing cells with a high nuclear to cytoplasmic ratio (Fig. 2b). From these, 44 colonies were passaged further, and 27 survived; all were positive for the human PSC cell surface antigen TRA-1-81 (Supplementary Fig. **2a).** We then randomly selected six clones for further expansion to generate six clonal lines (Fig. 2b). All displayed high levels of expression for a panel of typical human PSC cell surface antigens, BF4, CD9, SSEA-4, TRA-1-60, TRA-1-81 and TRA-2-49 compared to the negative control P3X (Fig. 2c, Supplmentary Fig. 2c). Expression of the pluripotency associated marker, NANOG, was confirmed by intracellular staining (Supplementary Fig. 2b). Although these clonal lines were all derived from *MIXL1*(+)/SSEA-3(+) cells (Fig. 2a), their pattern of *MIXL1* and SSEA-3 expression in each case had reverted to the pattern seen in the parental cultures rather than retaining the *MIXL1* positive status of the starting cell (Fig. 2d). Thus the *MIXL1*(+)/SSEA-3(+) subpopulation contains undifferentiated stem cells, despite their transient expression of genes associated with differentiation as cells were able to repopulate a heterogeneous population (Fig. 2e), and that the *MIXL1*(+)/SSEA-3(+) and *MIXL1*(-

)/SSEA-3(+) subpopulations represent interconvertible substates of the stem cell compartment.

Generating a *MIXL1*(+)/SSEA-3(+) population in a fully defined system

The MEF/KOSR system is not fully defined depending on proprietary components and the use of mouse fibroblasts as a feeder layer and the presence of *MIXL1*(+)/SSEA-3(+) is inherently variable in this system (data not shown) . We tested whether the *MIXL1*(+)/SSEA-3(+) subpopulation existed in completely defined conditions, E8 medium [15], with a vitronectin attachment factor (E8V) culture conditions. We found that this subpopulation vanished and no *MIXL1*(+)/SSEA-3(+) cells were detected (Fig. 3a). Since we had noted that an inhibitor of endogenous WNT secretion, IWP-2, also inhibited the appearance of *MIXL1*(+)/SSEA-3(+) cells in MEF/KOSR conditions (*data not shown*) similar to findings shown before [16], we inferred that the substate might be regulated by WNT signalling. Therefore, we tested whether activation of WNT signalling in cells cultured in E8V medium could generate this subpopulation. Indeed, the GSK3β inhibitor, CHIR99021 (CHIRON), a canonical WNT signalling mimic, did strongly induce *MIXL1* expression but, although initially the *MIXL1*(+) cells also expressed SSEA-3, within three days (72 Hours) many of the cells were *MIXL1*(+)/SSEA-3(-) (Fig. 3b). However, by the further addition of varying levels of lysophophatidic acid (LPA), a promoter of pluripotency (Garcia-Gonzalo et al, 2008, Blaukwamp et al, 2012), we were able to counteract the inductive action of CHIRON and establish conditions in which the pro- and anti-differentiation effects were balanced, so that a substantial proportion of the cells retained expression of both *MIXL1* and SSEA-3 (Fig. 3c). In these exploratory experiments we had also included ROCK inhibitor (Y-27632); subsequently, we found this is unnecessary and, in the absence of the ROCK inhibitor, we established our initial optimal conditions of 3 μM CHIRON and 0.48 μM LPA to form the *MIXL1*(+)/SSEA-3(+) subset.

Although the combination of CHIRON and LPA was able to generate *MIXL1*(+)/SSEA-3(+) cells, its ability to maintain this level after passaging was variable. We considered that a possible cause of this variability is the secretion of WNT ligands by cells undergoing early differentiation and that these compromised the level of LPA that we had optimised to counteract the effect of the CHIRON. To address this, we tested a system, similar to the Baseline Activation (BLA) method proposed in Hackland et al, 2017, in which we added the inhibitor of WNT secretion, IWP-2 [17] (Fig. 3d). When the cells were first grown in the presence of 0.48 µM LPA and 3 µM CHIRON and then passaged in the absence of IWP-2, there was a marked decrease in expression of SSEA-3. However, when the medium also contained 1 µM IWP-2 there was a striking increase in the expression of SSEA-3 after passaging. These cells from cultures in the presence of IWP-2 also retained expression of other stem cell related antigens and displayed good colony morphology and growth rate over four days post passage (data not shown). The optimisation process (Fig. 3e) lead to our first medium formulation**,** which we named PRIMO. The medium consisted of E8V basal medium plus 0.1% BSA, 2µM cholesterol, 3 µM CHIRON, 1 µM IWP-2 and 0.48 µM LPA. PRIMO medium induced a high proportion of *MIXL1*(+)/SSEA3(+) cells in cultures (Fig. 3f).

Transcriptional assessment of cells in PRIMO medium

To assess how closely the 'trapped' *MIXL1*(+)/SSEA-3(+) subset from cells growing in PRIMO medium resembled the similar subset from MEF/KOSR cultures, we compared the transcriptomes of *MIXL1*(-)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(+) cells grown in PRIMO conditions for 3 days, as well as *MIXL1*(-)/SSEA-3(+) cells from cultures in E8V, with the previous data from cells growing in MEF/KOSR conditions. By PCA, the bulk RNA sequencing data showed a similarity between the *MIXL1*(+)/SSEA-3(+) cells from both PRIMO and MEF/KOSR cultures whereas the *MIXL1*(-)/SSEA-3(+) PRIMO fraction showed some separation from its counter parts in MEF/KOSR and was further separated from the corresponding fraction from E8V cultures (Fig. 4a). We also carried out a single cell qPCR

187 analysis of expression of the 48 gene-set analysed previously. When these data were
188 incorporated into the t-SNE plot previously obtained from the cells growing in MEF/KOSR
189 conditions, the PRIMO-derived *MIXL1*(+)/SSEA-3(+) cells showed a distribution that largely
190 superimposed on the *MIXL1*(+)/SSEA-3(+) cells from MEF/KOSR condition, lying between
191 the *MIXL1*(-)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(-) cluster (Fig. 4b). The gene signatures from
192 single cell gene expression patterns from the *MIXL1*(+)/SSEA-3(+) fractions from PRIMO
193 and MEF/KOSR cultures were very similar (Supplementary Fig. 3, 4)

194 To assess how the *MIXL1*(+)/SSEA-3(+) cells relate to the trajectory of ES cell differentiation
195 we added 3 μM CHIRON to cells growing in E8V medium and isolated the emerging cell
196 populations at 0, 6, 12, 18, 24, 48 and 72 hours, based on *MIXL1*-GFP and SSEA-3
197 expression (Fig. 4c, Supplementary Fig. 5). Over this period the expression of *MIXL1*
198 gradually increased, preceding the eventual loss of SSEA-3 expression. Bulk RNAseq data
199 from the emerging cells at each time point was compared by PCA with RNAseq data from
200 *MIXL1*(-)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(+) cells growing in PRIMO conditions for 3 days
201 (Fig. 4d). The *MIXL1*(+)/SSEA-3(+) fraction mapped between 24 and 48 hours, whereas the
202 *MIXL1*(-)/SSEA-3(+) fraction from PRIMO conditions did not correlate with the zero hour,
203 *MIXL1*(-)/SSEA-3(+) cells from E8V but rather mapped on the trajectory between 12 and 18
204 hours of differentiation (Fig. 4d). Thus, the PRIMO medium appears to trap all cells in an
205 intermediate substate but, interestingly, this substate appears to be independent of *MIXL-1*-
206 GFP reporter expression. Differential gene expression analysis by DESeq2 and SeqMonk of
207 cells growing in PRIMO, pooling data from both *MIXL1*(+) and *MIXL1*(-) cells, compared to
208 E8V revealed upregulation of differentiation markers while key pluripotency associated
209 markers such as *NANOG*, *POU5F1* were not differentially expressed (Fig. 4e). Much like the
210 *MIXL1*(+)/SSEA-3(+) cells from MEF/KOSR (Fig. 1b), our PRIMO cultures showed co-
211 expression of pluripotency associated genes such as *POUF51*, *NANOG*, etc. and
212 differentiation-associated genes such as *MIXL1*, *T* (*BRACHYURY)*, *DKK1*, etc. (Fig. 4e).
213 Gene ontology enrichment analysis of the most differentially expressed genes, red squares

in Fig. 4e, revealed an enrichment in biological processes such as mesoderm development and gastrulation (Fig. 4f).

Functional assessment of cells growing in PRIMO

After this initial transcriptomic profiling using PRIMO medium, we found that the formulation required further optimisation to increase the maintenance of cells after multiple passages. We developed a further optimised medium containing a higher level of LPA, termed PRIMO Plus which consisted of E8V basal medium plus 0.2% BSA, 4 µM cholesterol, 3 µM CHIRON, 1 µM IWP-2 and 0.96 µM LPA. From our previous work on *MIXL1*(+)/SSEA-3(+) cells from feeders we predicted that the same population in PRIMO Plus could again generate clonal stem cell lines but also, as our transcriptomic analysis indicated these cells were on a differentiation trajectory, that their differentiation would be biased towards mesoderm. To confirm this HES3-*MIXL1* cells grown for 3 days in PRIMO Plus medium were analysed by cloning and by embryoid body (EB) formation under 'neutral' conditions [18]. The clonal lines established also were used for EB formation. All EBs were then assessed by transcriptome signatures and the score card method [19] (Fig. 5a).

We obtained 38 stem cell-like colonies from 384 *MIXL1*(+)/SSEA-3(+) cells isolated by FACS and single cell deposition (Fig. 5b). Of these, 31 colonies survived further passaging and were positive for TRA-1-81 staining (Supplementary Fig. 6a). Six of these colonies were randomly selected, initially expanded in MEF/KOSR conditions (Supplementary Fig. 6b-c), transitioned into E8V conditions and then were assessed for their expression of pluripotency associated surface markers expression and *MIXL1*-GFP (Supplementary 6d-e); the index position of the parent cells for each of these expanded lines confirmed that they were derived from *MIXL1*(+)/SSEA-3(+) cells (purple stars in Fig. 5b). Each of the clones exhibited patterns of antigen expression, and lack of *MIXL1-GFP*, expression, similar to the parental cells growing in E8V conditions (Fig. 5c). These cloning results indicate that the

MIXL1(+)/SSEA-3(+) cells from cultures in PRIMO medium do reside within the stem cell compartment and retain the ability to revert to other stem cell substates.

When these Primo-derived clones adapted to growth in E8V were cultured as EBs under neutral conditions they showed marked down-regulation of stem cell related genes and upregulation of genes associated with all three germ layers (Fig. 5d, 5e), indicating a capacity for multilineage differentiation with no obvious bias. The transcriptional time course analysis of cells in PRIMO showed that both subsets, irrespective of *MIXL*-GFP expression, are located on the trajectory of mesoderm differentiation induced by CHIRON (Fig. 4d). Therefore, we tested whether these transcriptional changes would lead to differentiation that favours mesoderm formation. When we collected *MIXL1*(-)/SSEA-3(+) or *MIXL1*(+)/SSEA-3(+) cells by FACS or unsorted cells grown in PRIMO Plus medium, and grew them as EBs under neutral conditions, they all showed a marked bias to mesoderm differentiation, irrespective of *MIXL1* expression. Congruent with the transcriptional time course analysis, EBs showed that cells grown in PRIMO medium are on a mesoderm trajectory which impacts upon their lineage fate decision, an impact that was reversible when clonal lines were established in E8V conditions.

Multiple passages in PRIMO Plus medium.

To confirm that we could maintain human PSC in a biased substate through successive passages, we analysed HES3-*MIXL1* ES cells after three passages in PRIMO Plus medium. In addition, we also transitioned into PRIMO Plus medium another human PSC line, H9T-Venus, an ES cell line that carries a Venus reporter for *T*(BRACHYURY*)* expression [20], and a human induced pluripotent stem (iPS) cell line, MIFF1[21] , and likewise maintained them through three passages. In each case, the cells retained an undifferentiated morphology and expressed high levels of SSEA3 (Fig. 6a). The HES3-*MIXL1* cultures retained, as before, a population of *MIXL1*(+)/SSEA-3(+) cells, while the H9T-Venus cultures contained a substantial population of *T* (*Brachyury)* (+) cells, this population being more consistently

abundant than the *MIXL1*(+) population, whereas in E8V medium there were no *MIXL1*(+) (Fig. 3a) or *T* (*Brachyury)* (+) cells (Supplementary Fig. 7a). These cells also expressed a typical set of pluripotency-associated genes that were down regulated when the cells were allowed to form EBs under neutral differentiation conditions. At the same time, the pattern of induced gene expression indicated a strong mesoderm bias in the differentiation of each of the lines (Fig. 6b). These results confirm that in the optimised PRIMO Plus medium, human pluripotent stem cells can be maintained following passaging in a mesoderm biased substate within the stem cell compartment.

PRIMO Plus remained our optimal medium, but we also tested whether we could substitute components for others that targeted the same or similar pathways. Using H9-T-Venus reporter we found that recombinant Dickkopf WNT Signaling Pathway Inhibitor 1 (DKK1), which inhibits WNT ligand binding to WNT receptors, could be used at 100ng/mL in place of IWP-2 (Supplementary Fig. 7b). Similarly, CHIR99021 could be substituted with the GSK3β inhibitor SB216763 at 10μM (Supplementary Fig. 7c). For replacement of LPA we focused on activators of G-Coupled Protein receptors. We used various concentrations of Sphingosine-1-phosphate (S1P), another phospholipid present in KOSR medium that has been shown to inhibit differentiation [22, 23], and a chemical agonist for LPA receptor 2 (GRI977143) [24]. The cell expression patterns observed using S1P or GRI977143 media were similar to that observed using LPA-based media, albeit at slightly different concentrations (Supplementary Fig. 7d-e).

We returned to HES3-*MIXL1* and our PRIMO Plus formulation to assess further how the long term passaged, mesoderm biased cells related to the trajectory of mesoderm differentiation induced by CHIRON. We also performed bulk RNA-seq analysis on HES3-*MIXL1* cells cultured in PRIMO Plus for 10 passages. In addition, to confirm that these cells would revert to the unbiased state apparent in E8V cultures, we also analysed cells that had been cultured in PRIMO Plus for 9 passages and then transitioned back into standard E8V

medium after one passage in the presence of IWP-2 or LPA to aid reversion. We compared these data (Fig. 6c) to our previous analysis of the time course of differentiation induced by CHIRON and also cells cultured PRIMO cultures for just 3 days (Fig. 4d). The transcriptome of both the *MIXL1*(+)/SSEA3(+) and *MIXL1*(-)/SSEA3(+) cells after 10 passages in PRIMO Plus correlated with cells located between 18 and 24 hours on the differentiation trajectory, similar to the previously assessed cells after 3 days in PRIMO medium. Crucially, though, when cells were transitioned back into E8V medium these induced transcriptional changes were able to revert to expression levels typical of cells maintained in E8V medium. Assessing a sample of genes related to pluripotency and mesendoderm differentiation reveals a pattern of expression shared by PRIMO cultures and lost in reverted cultures. Mesendoderm genes such as *T, EOMES, NODAL*, etc. are all highly expressed in PRIMO cultures and lowly or non-expressed in both E8 and reverted cultures. Pluripotency associated genes such as *NANOG, POU5F1, SOX2* and *DNMT3B* have less varied expression levels across all samples (Fig. 6d), the protein expression of NANOG and SOX2 was also validated by immunofluorescence in PRIMO Plus and reverted cultures (Supplementary Fig. 8).

DISCUSSION

One view to accommodate the apparent heterogeneity of hPSC is that the stem cell compartment comprises several interconvertible substates in which the cells retain pluripotency but behave transiently as differentiation intermediates, with distinct propensities for differentiation and lineage selection. We have now identified a mesoderm biased state within human pluripotent stem cell compartment, which we can induce and maintain in

PRIMO, a defined medium based on cross antagonism, ie simultaneously applying signals that promote differentiation (CHIR99021) and pluripotency (LPA) (Fig. 6e). This state exhibited characteristics corresponding to undifferentiated cells, demonstrated through their capacity to generate clonally derived lines that retain pluripotency. Examining the transcriptome, by bulk and single cell approaches, highlighted co-expression of pluripotency and differentiation associated gene regulatory networks within the substate. This substate could be induced in three independent human PSC lines and remained stable over multiple passages.

The cells induced by Primo medium appear to be relevant normal intermediates in the differentiation of human PSC towards mesoderm, as their transcriptome is positioned between 18 to 24 hours on the time course of induced mesoderm differentiation. Although initially we defined the state using *MIXL1* expression, cells in Primo were heterogeneous for *MIXL1*-GFP expression whereas *T(Brachyury)* was more widely expressed as judged by the *T*-Venus reporter. When compared to our differentiation time course *MIXL1*(+) and *MIXL1*(-) cells both sit on the trajectory with the *MIXL1*(+) cells sitting slightly further on the trajectory. Importantly *MIXL1*(-) cells grown in PRIMO, like their *MIXL1*(+) counterparts, displayed a strong mesoderm bias in “neutral” EB assays, in contrast to the MXL1(-) cells grown in E8 medium.

LPA and Chiron have been used in combination before by Blauwkamp et al 2012, in an attempt to enhance differentiation efficiency through reducing heterogeneity of WNT expression in human ES cultures. Our goal was to trap cells in a mesoderm biased intermediate state using LPA and CHIRON but CHIRON induction produces a positive feedback loop, since the derivative cells themselves produce WNT ligands which drives further differentiation. Indeed, our single cell transcriptomic analysis, showed expression of endogenous WNT related genes, consistent with the notion that early differentiation would lead to the secretion of WNT ligands which would confound our ability to precisely regulate

WNT activation with CHIRON. To counteract this positive feedback loop, we reasoned that we could incorporate a Baseline Activation approach (Hackland et al, 2017).

Baseline Activation as an approach focuses on blocking the endogenous signalling that allows an exact titration of the signal by addition of exogenous agonist, affording tighter control. We used IWP-2 to inhibit WNT ligand secretion, or DKK1 to block WNT ligand binding, while activating the pathway with CHIRON99021 or SB216763. This strategy enabled us to stabilise cultures in PRIMO medium and maintain the cells in this transitory pluripotent state that exhibited mesoderm bias (Fig. 6e). Importantly, this substate was plastic with cells being able to further differentiate or, when returned to E8 medium, revert to an unbiased stem cell state. The return to the stem cell state was demonstrated both by reversion of the transcriptional profile and reacquisition of unbiased differentiation in EB assays.

Taken together our results suggest that the intermediate we have trapped sits on a normal in vitro differentiation trajectory the investigation of which may lend further insight into the critical events that ultimately lead to the exit from pluripotency. How this intermediate relates to the differentiation within the human embryo remains unclear. Interestingly, cells in PRIMO express certain lineage markers including *T* and *MIXL1* that are expressed in a subset of mouse epiblast stem cells that are derived from the post-implantation embryo [26-28]. It is plausible that this similarity in the transcriptome reflects correspondence to a similar developmental intermediate in the human embryo. These considerations aside, we wish to emphasise that these cells represent a key in vitro intermediate in the differentiation of hPSC that has practical implications for production of cells for use in regenerative medicine. The system of cross-antagonism within the controlled environment of baseline activation that we have used to generate a mesoderm-biased substate of hPSCs could, in principle, be used to generate other lineage-biased substates or facilitate the capture of relevant differentiation intermediates at later developmental stages.

373

374 **METHODS**

375 **Cell culture:**

376 Four human PSC lines were used in this project, HES3 [29], ES cell lines HES3 *MIXL1*-GFP [11]
377 and H9 *T*-Venus [20], and iPSC line, MIFF1 [21]. Both feeder and feeder free systems were
378 used. For either system, flask/plates of hESCs were grown in humidified incubators at 37°C
379 and 5% CO2. Our feeder culture system uses a layer of mitotically inactivated mouse
380 embryonic feeders and KOSR medium consisting of KnockOut DMEM (ThermoFisher), 20%
381 KnockOut Serum Replacement (Thermofisher), 1x Non-essential amino acids
382 (ThermoFisher), 1mM L-Glutamine (ThermoFisher), 0.1mM 2-Mercaptoethanol
383 (ThermoFisher) and 4ng/ml FGF2 (Peprotech) and is referred to as MEF/KOSR conditions.
384 For feeder free systems we used two matrices Geltrex (ThermoFisher) and Vitronectin (Stem
385 cell Technologies). This was combined with E8 medium (made in house, adapted from Chen
386 et al, 2011) with L-Glutamine being replaced for the more stable GlutaMax (ThermoFisher).
387 For LPA containing and subsequent PRIMO media, the LPA media is first made as a 10X
388 stock in E8 medium, 10X BCL (BSA, Cholesterol, LPA) at 1% Fatty acid Free BSA
389 (Probumin, Millipore), 20μM Cholesterol (Synthechol, Sigma) and 4.8μM Oleoyl-L-α-
390 lysophosphatidic acid sodium salt (LPA, Sigma). This was then diluted in E8 as needed for
391 PRIMO and PRIMO Plus, then CHIRON 99021 (Tocris) and IWP-2 (Tocris) were added at
392 3μM and 1μM, respectively. Cells for MEF/KOSR were passaged by treatment with
393 Collagenase IV (ThermoFisher) and manually scrapped away. Cells for feeder free systems
394 were passaged with a non-enzymatic disassociation solution ReLeSR (Stem Cell
395 Technologies) according to manufacturer's instructions.

396

397

398 **Antibody Staining:**

399 ***Flow cytometry***

400 Single cell suspensions were harvested using accutase and resuspended in DMEM and
401 10% FCS at a density of $1x10^7$ per mL. Antibodies were added at the appropriate dilution.
402 After addition of the primary antibody cells were incubated at 4°C for 30 minutes. Cells were
403 then washed with DMEM/FCS and resuspended in DMEM/FCS. Secondary antibody, Goat
404 anti Mouse Affinipure IgG+IgM (H+L) Alexafluor 647 (Jackson Laboratories) at 1:200 and
405 incubated at 4°C for 30 minutes. Cells were then washed and resuspended in fresh
406 DMEM/FCS for FACS analysis. To set baselines for *MIXL1*-GFP and negative secondary
407 647, unlabelled HES3 line was harvested and stained for P3X. P3X is an IgG1 antibody that
408 is secreted from the parent myeloma, P3X63Ag8, from which all in house hybridomas had
409 been derived[30]. P3X shows minimal reactivity to human cells (Kohler and Milstein, 1975).
410 Positive gates were set according to HES3 P3X negative controls. Samples were also
411 stained for P3X to assess non-specific binding. All FACs analysis contained P3X samples for
412 baseline setting. The monoclonal antibodies BF4 [31], CD9(CH8) [31], SSEA-3 [32], SSEA-4 [33],
413 THY1(CD90) [34] , TRA-1-60 [35], TRA-1-81 [35] and TRA-2-49 [36] were prepared in house from
414 the relevant hybridomas as previously described [12, 34].

415 **Intracellular staining.**

416 Cells were fixed using 4% PFA for 15 minutes and permeabilised with 0.5% Triton X-100 in
417 PBS (w/o Ca+, Mg++) with 10% FCS and 0.1% BSA. Cells were washed and then incubated
418 with blocking buffer consisting of 10% FCS and 0.1% BSA in PBS (w/o Ca+, Mg++) for one
419 hour. Cells were then washed, primary antibody was added and they were incubated
420 overnight at 4°C. Antibodies used were: Anti-NANOG (XP® #4903, Cell Signalling
421 Technology, 1:400), ANTI-SOX2 (XP® #3579, Cell Signalling Technology, 1:400).
422 Secondary antibody, Goat anti Mouse Affinipure IgG+IgM (H+L) Alexafluor 647 (Jackson
423 Laboratories) at 1:200 and incubated at 4°C for 2 hours minutes. Hoescht 33342
424 (ThermoFisher, #H3570) was added at 1:1000 to the diluted secondary antibody solution.

425

Live TRA-1-81 Staining.

In order to assess the progress of the clonal line formation we performed live staining for a pluripotency-associated surface marker TRA-1-81. Lines were assessed after the first passage into 48 well plates. TRA-1-81 antibody was added to warm KOSR medium at 1:10 dilution and incubated at 37°C for 30 minutes. Wells were then washed twice with KOSR medium before medium containing secondary antibody, Goat anti Mouse Affinipure IgG+IgM (H+L) Alexafluor 647 (Jackson Laboratories) at 1:200 was added to each well, cells were incubated at 37°C for 30 minutes. Wells were then washed once with KOSR medium and twice with FluoroBrite DMEM (ThermoFisher). After imaging the medium was replaced with fresh KOSR medium and returned to 37°C incubator.

Single Cell Cloning:

96 well plates were coated with gelatin and a layer of mouse embryonic feeders. Cells were harvested using Accutase (ThermoFisher). After staining for flow cytometry, DAPI was added at 1:10,000 and used for live/dead discrimination. After gating on the BD Sortware program the desired population was sorted as single cells directly into 96 well plates. The sort was indexed to retain information regarding the *MIXL1*-GFP and SSEA-3 expression levels. For single cell cloning the medium differed from standard culture. For this we used a 50/50 mix between standard KOSR medium and mTESR medium (Stem Cell Technologies) and the addition of 20μM Synthechol (Sigma). During initial plating 10μM Rock Inhibitor, Y-27632, was added to the medium. Immediately after sorting into the wells the plates were centrifuged at 1000 rpm for 1 minute to aid attachment of the cells. After two days the medium was replaced with fresh medium to remove the ROCK inhibitor. Colonies were left to develop over 9-12 days before passaging the wells which looked to contained typical hPSC colonies. Colonies were passaged from 96 well plate into a 48 well plate by manual scrapping with a p200 tip, then aspirating and dispensing the dissected colony into one well of a 48 well plate. This 50/50 mix medium was used to grow clones until lines appeared to

be growing stably, often up to the third passage before changing into standard KOSR medium.

Embryoid Bodies

To assess the trilineage potential, we used an approached which entailed the formation of Embryoid Bodies (EB) under “neutral” conditions, in this context neutral simply indicates that no exogenous cytokines or chemicals were added to guide differentiation. Cells were either used directly from flasks or after they had been FACS sorted for a particular population. In either situation cells were resuspended in APEL 2 medium (Stem Cell Technologies) containing 10μM ROCKi. Cells were resuspended at 3,000 cells per 50μL. 50μL of cells were added to non-adherent Grenier U bottom 96 well plate. After adding the cells plates were centrifuged at 1000rpm for 3 minutes, to pellet cells. Plates were incubated at 37°C, %CO2 for 7 days. After 7 days EBs were dissolved in Trizol and RNA extracted using the Norgen Total RNA Purification Kit. The RNA was converted to cDNA using the High Capacity cDNA Reverse Transcription Kit (Thermofisher). The samples were then loaded and run on the hPSC scorecards (ThermoFisher), results were analysed by ThermoFisher’s scorecard software. Algorithm scores for EBs were normalised to the same cell line growing in E8V conditions.

Differentiation Time course:

Cells growing in E8V medium were supplemented with 3μM CHIRON one day post passage. Cells were collected and stained with anti-SSEA-3 antibody at time points 6, 12, 18, 24, 48, and 72 hours. Cells were sorted directly into Trizol. The medium was refreshed after 48 hours with fresh E8V containing 3μM CHIRON.

Bulk sorting and RNA Sequencing

We sorted using a BD FACS Jazz or FACS Aria III, we first begin by sorting 10,000 cells from a given population into an eppendorf tube containing 200µL DMEM/FCS. 1,000 cells are then reanalysed from this sorted fraction to ensure accurate sorting has taken place. Immediately after a successful reanalysis, 10,000 cells from the given population are sorted into an eppendorf tube containing 800µL Trizol. Post sort, samples are sealed and stored at -80°C until extraction. RNA samples were processed as described in Böiers et al, 2018 [37] but in brief, RNA was extracted and cDNA was prepared using the SMARTer v3.0 or v4.0 cDNA synthesis kit (Clontech). RNA-Seq was performed on an Illumina Nextseq 500 machine. The Raw data was processed by Tophat (Kim et al., 2013a) and Cufflink/Cuffdiff (Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2013; Trapnell et al., 2010). Reads were aligned to human genome reference consortium grch38. Gene expression data was then processed by Seqmonk software (Babraham-Bioinformatics, 2018) for quality control and to produce heat maps and clustering. Gene Ontology enrichment was performed by ToppGene [38] and ReviGO [39].

Single Cell qPCR

Samples were sorted using the flow cytometer, BD FacsAria III. DAPI (Sigma) or Propidium Iodide (Sigma) was added for live/dead discrimination. Sorting population gates were set in the flow cytometry software, FACS DIVA, based on GFP expression (488nm laser), and SSEA-3 (stained with Alexafluor 647 secondary antibody) expression (635nm laser). Single cells from each selected substate were then directly deposited into 96 well plate with 4µl of lysis buffer (65µM of dNTP mix (Invitrogen 10297-018), 0.4% of NP40 (Sigma N-3516), 2.4 mM DTT (Invitrogen), 0.5U/µl RNAseOUT (Invitrogen 10777-019) in nuclease-free water). cDNA synthesis and target specific pre-amplification was done using Cell Direct one-step qRT-PCR kit (Invitrogen 46-7201). Pre-amplification mastermix was added to each well;

6.25µL of 2X Reaction Buffer, 1ul of SuperScript III RT/Platinum Taq mix (Invitrogen 55549) and 1.5µl of TaqMan assay mix. TaqMan assay mix was prepared by mixing equal volume of all target specific primers (Supplementary Table 1). No-RT controls were prepared with PlatinumTaq Polymerase (Invitrogen, 100021272) and no SuperScript III RT enzyme was included. The PCR conditions were: 60 min at 50C, 2 min at 95C and 25 cycles of 15 sec at 95C and 4 min at 60C. Pre-amplified product was diluted 1 to 5 and loaded onto a 48.48 chip together with Taqman universal MasterMix (Applied Biosystems 4304437) and the Taqman assays listed in (Supplementary Table 1) with the appropriate loading reagents according to manufacturer's instructions (BioMark 48.48 Dynamic array platform (Fluidigm). Single cell qPCR analysis was performed by two pieces of software. Prior to analysis cells were screened and cells with particularly high CT values (>30) for the housekeeping genes *ACTB* and *RPS18* were removed, across all experiments 9 of 320 single cells were removed. We verified expression of genes with extra wells containing 10 cells rather than single cells, *MMP1* was removed from analysis, as it showed no expression in any samples. Depending on the application being used, some pre-processing was performed. For Monocle2 software (Trapnell et al., 2014) and subsequent T-SNE plots, 999 values were set to 40, as this is the maximum number of cycles on the PCR machine. We used Genesis software (Sturn et al., 2002) for heatmap production, where 999 values were removed and Z-scores were computed from all remaining values. For each gene, Z-scores were calculated using this formula: (Ct value – Ct mean) / Standard Deviation Cts.

REFERENCES

1. Graf, T. & Stadtfeld, M. Heterogeneity of Embryonic and Adult Stem Cells. *Cell stem cell* **3**, 480-483 (2008).
2. Hough, S.R., Laslett, A.L., Grimmond, S.B., Kolle, G. & Pera, M.F. A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. *PloS one* **4**, e7708 (2009).
3. Enver, T. *et al.* Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Human molecular genetics* **14**, 3129-3140 (2005).
4. Allison, T.F. *et al.* Identification and Single-Cell Functional Characterization of an Endodermally Biased Pluripotent Substate in Human Embryonic Stem Cells. *Stem Cell Reports* **10**, 1895-1907 (2018).
5. Hough, Shelley R. *et al.* Single-Cell Gene Expression Profiles Define Self-Renewing, Pluripotent, and Lineage Primed States of Human Pluripotent Stem Cells. *Stem Cell Reports* **2**, 881-895 (2014).
6. Gokhale, P.J. *et al.* Culture Adaptation Alters Transcriptional Hierarchies among Single Human Embryonic Stem Cells Reflecting Altered Patterns of Differentiation. *PloS one* **10**, e0123467 (2015).
7. Kuroda, T. *et al.* SALL3 expression balance underlies lineage biases in human induced pluripotent stem cell differentiation. *Nature communications* **10**, 2175 (2019).
8. Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254-1269 (2014).
9. Theunissen, Thorold W. *et al.* Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell stem cell* **15**, 471-487 (2014).
10. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282-286 (2013).

552 11. Davis, R.P. *et al.* Targeting a GFP reporter gene to the MIXL1 locus of human
553 embryonic stem cells identifies human primitive streak-like cells and enables isolation
554 of primitive hematopoietic precursors. *Blood* **111**, 1876-1884 (2008).
555 12. Draper, J.S., Pigott, C., Thomson, J.A. & Andrews, P.W. Surface antigens of human
556 embryonic stem cells: changes upon differentiation in culture. *Journal of anatomy*
557 **200**, 249-258 (2002).
558 13. Andrews, P.W., Goodfellow, P.N., Shevinsky, L.H., Bronson, D.L. & Knowles, B.B.
559 Cell-surface antigens of a clonal human embryonal carcinoma cell line:
560 morphological and antigenic differentiation in culture. *International journal of cancer*
561 **29**, 523-531 (1982).
562 14. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by
563 pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381-386 (2014).
564 15. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture.
565 *Nature methods* **8**, 424-429 (2011).
566 16. Kurek, D. *et al.* Endogenous WNT Signals Mediate BMP-Induced and Spontaneous
567 Differentiation of Epiblast Stem Cells and Human Embryonic Stem Cells. *Stem Cell*
568 *Reports* **4**, 114-128 (2015).
569 17. Hackland, J.O.S. *et al.* Top-Down Inhibition of BMP Signaling Enables Robust
570 Induction of hPSCs Into Neural Crest in Fully Defined, Xeno-free Conditions. *Stem*
571 *Cell Reports* **9**, 1043-1052 (2017).
572 18. Ng, E.S., Davis, R., Stanley, E.G. & Elefanty, A.G. A protocol describing the use of a
573 recombinant protein-based, animal product-free medium (APEL) for human
574 embryonic stem cell differentiation as spin embryoid bodies. *Nature protocols* **3**, 768
575 (2008).
576 19. Bock, C. *et al.* Reference Maps of Human ES and iPS Cell Variation Enable High-
577 Throughput Characterization of Pluripotent Cell Lines. *Cell* **144**, 439-452 (2011).
578 20. Mendjan, S. *et al.* NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm
579 during Exit from Pluripotency. *Cell stem cell* **15**, 310-325 (2014).

21. Desmarais, J.A., Unger, C., Damjanov, I., Meuth, M. & Andrews, P. Apoptosis and failure of checkpoint kinase 1 activation in human induced pluripotent stem cells under replication stress. *Stem cell research & therapy* **7**, 17 (2016).

22. Pébay, A. *et al.* Essential Roles of Sphingosine-1-Phosphate and Platelet-Derived Growth Factor in the Maintenance of Human Embryonic Stem Cells. *Stem cells (Dayton, Ohio)* **23**, 1541-1548 (2005).

23. Garcia-Gonzalo, F.R. & Izpisúa Belmonte, J.C. Albumin-Associated Lipids Regulate Human Embryonic Stem Cell Self-Renewal. *PloS one* **3**, e1384 (2008).

24. Kiss, G.N. *et al.* Virtual screening for LPA2-specific agonists identifies a nonlipid compound with antiapoptotic actions. *Molecular pharmacology* **82**, 1162-1173 (2012).

25. Blauwkamp, T.A., Nigam, S., Ardehali, R., Weissman, I.L. & Nusse, R. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nature communications* **3**, 1070 (2012).

26. Tsakiridis, A. *et al.* Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development (Cambridge, England)* **141**, 1209-1221 (2014).

27. Brons, I.G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-195 (2007).

28. Tesar, P.J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199 (2007).

29. Cooper, S. *et al.* Biochemical properties of a keratan sulphate/chondroitin sulphate proteoglycan expressed in primate pluripotent stem cells. *Journal of anatomy* **200**, 259-265 (2002).

30. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. 1975. *Journal of immunology (Baltimore, Md. : 1950)* **174**, 2453-2455 (1975).

31. Wright, A. *et al.* Mapping the stem cell state: eight novel human embryonic stem and embryonal carcinoma cell antibodies. *International journal of andrology* **34**, e175-187; discussion e187-178 (2011).

608 32. Shevinsky, L.H., Knowles, B.B., Damjanov, I. & Solter, D. Monoclonal antibody to
609 murine embryos defines a stage-specific embryonic antigen expressed on mouse
610 embryos and human teratocarcinoma cells. *Cell* **30**, 697-705 (1982).
611 33. Kannagi, R. *et al.* Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of
612 a unique globo-series ganglioside isolated from human teratocarcinoma cells. *The*
613 *EMBO journal* **2**, 2355-2361 (1983).
614 34. Adewumi, O. *et al.* Characterization of human embryonic stem cell lines by the
615 International Stem Cell Initiative. *Nature biotechnology* **25**, 803-816 (2007).
616 35. Andrews, P.W., Banting, G., Damjanov, I., Arnaud, D. & Avner, P. Three monoclonal
617 antibodies defining distinct differentiation antigens associated with different high
618 molecular weight polypeptides on the surface of human embryonal carcinoma cells.
619 *Hybridoma* **3**, 347-361 (1984).
620 36. Andrews, P.W., Meyer, L.J., Bednarz, K.L. & Harris, H. Two monoclonal antibodies
621 recognizing determinants on human embryonal carcinoma cells react specifically with
622 the liver isozyme of human alkaline phosphatase. *Hybridoma* **3**, 33-39 (1984).
623 37. Böiers, C. *et al.* A Human IPS Model Implicates Embryonic B-Myeloid Fate
624 Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-
625 Associated ETV6-RUNX1. *Developmental Cell* **44**, 362-377.e367 (2018).
626 38. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list
627 enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**,
628 W305-W311 (2009).
629 39. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes
630 Long Lists of Gene Ontology Terms. *PloS one* **6**, e21800 (2011).

631

632

Figure 1

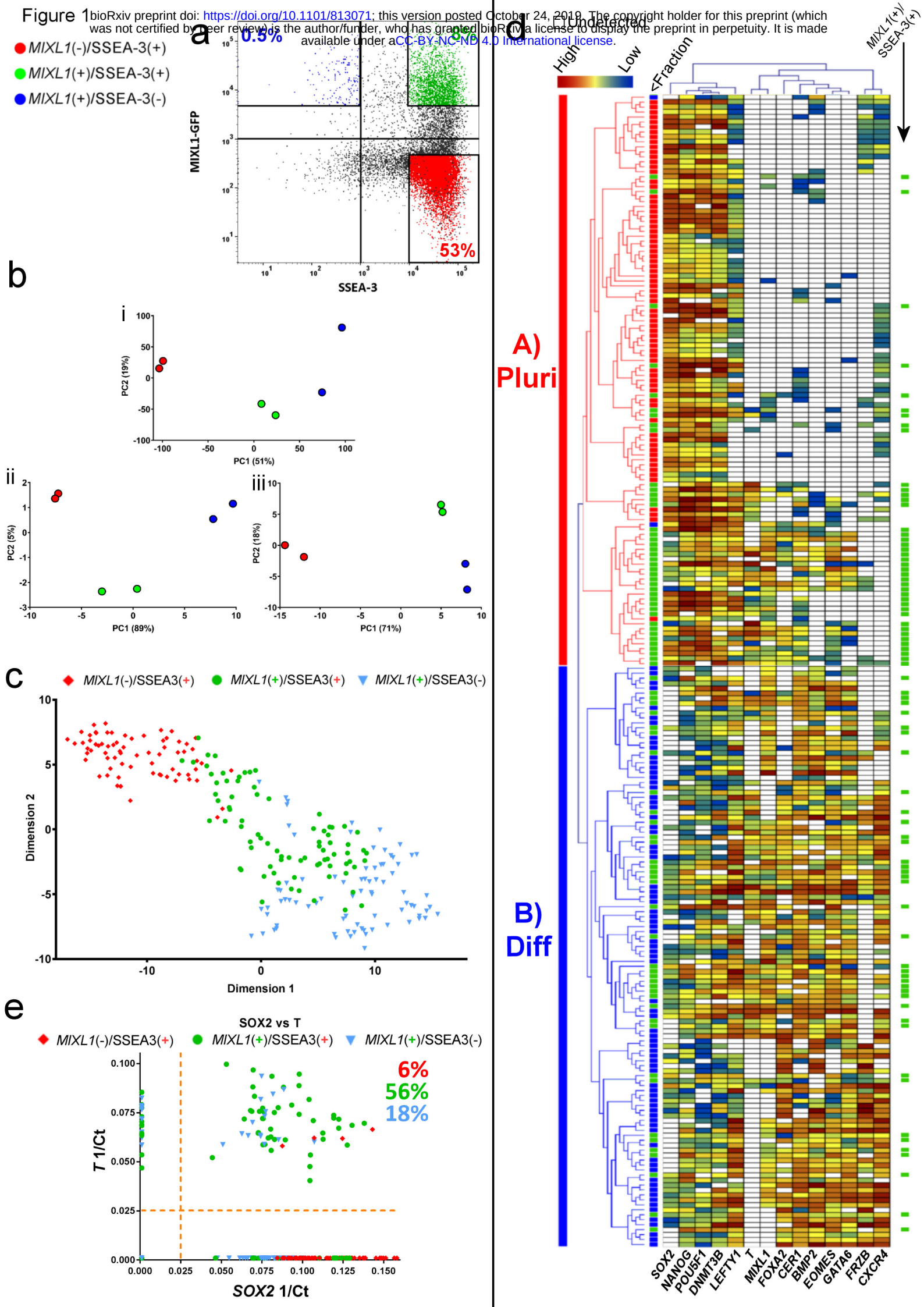


Figure Legends

Figure 1: A transitionary substate between pluripotency and early mesendodermal differentiation is marked by *MIXL1* and SSEA-3 coexpression in human ES cells grown in KOSR medium on MEFs.

a) Flow cytometry scatter plot of of MIXL1-GFP and SSEA-3 expression by HES3-*MIXL1* cells cultured in MEF/KOSR conditions. Coloured populations indicate the fractions isolated by FACS for bulk and single cell transcriptomic analysis, *MIXL1*(-)/SSEA-3(+) (Red) *MIXL1*(+)/SSEA-3(+) (Green), *MIXL1*(**+**)/SSEA-3(-) (Blue). **b)** Principal component analysis (PCA) plots generated from RNAseq data of the *MIXL1*(-)/SSEA-3(+), *MIXL1*(+)/SSEA-3(+) and *MIXL1*(+)/SSEA-3(-)fractions: top right - plot generated from all mRNA expression; bottom - plots generated from genes associated with pluripotency and mesoderm, respectively. **c)** t-SNE plots of the expression of 45 genes (Supplementary Table 1) in individual cells measured by Fluidigm Biomark qPCR. All fractions are displayed in the same dimensional space: *MIXL1*(-)/SSEA-3(+) cells in red, *MIXL1*(+)/SSEA-3(+) cells in green and *MIXL1*(-)/SSEA-3(+) cells in blue. **d)** Heatmap analysis of 232 single cells analysed across 14 of 45 analysed genes using a Fluidigm BioMark system: hierarchical clustering was performed for the genes assessed. The heatmap visualises the individual gene expression after normalisation across genes. White coloured genes indicate undetected levels of expression. Single cells are labelled by colour with the fraction from which they were sorted. Cluster analysis split cells into two broad classification, Cluster A (Pluripotency, “Pluri”) and Cluster B (Differentiation, “Diff”). Cells co-expressing pluripotency and differentiation associated genes were readily detected in the *MIXL1*(+)/SSEA-3(+) fraction but not the other fractions. **e)** Scatter plot of 1/Ct values for each single cell for *SOX2* and *T* expression. A large proportion of *MIXL1*(+)/SSEA-3(+) expressed both *SOX2* and *T*.

Figure 2

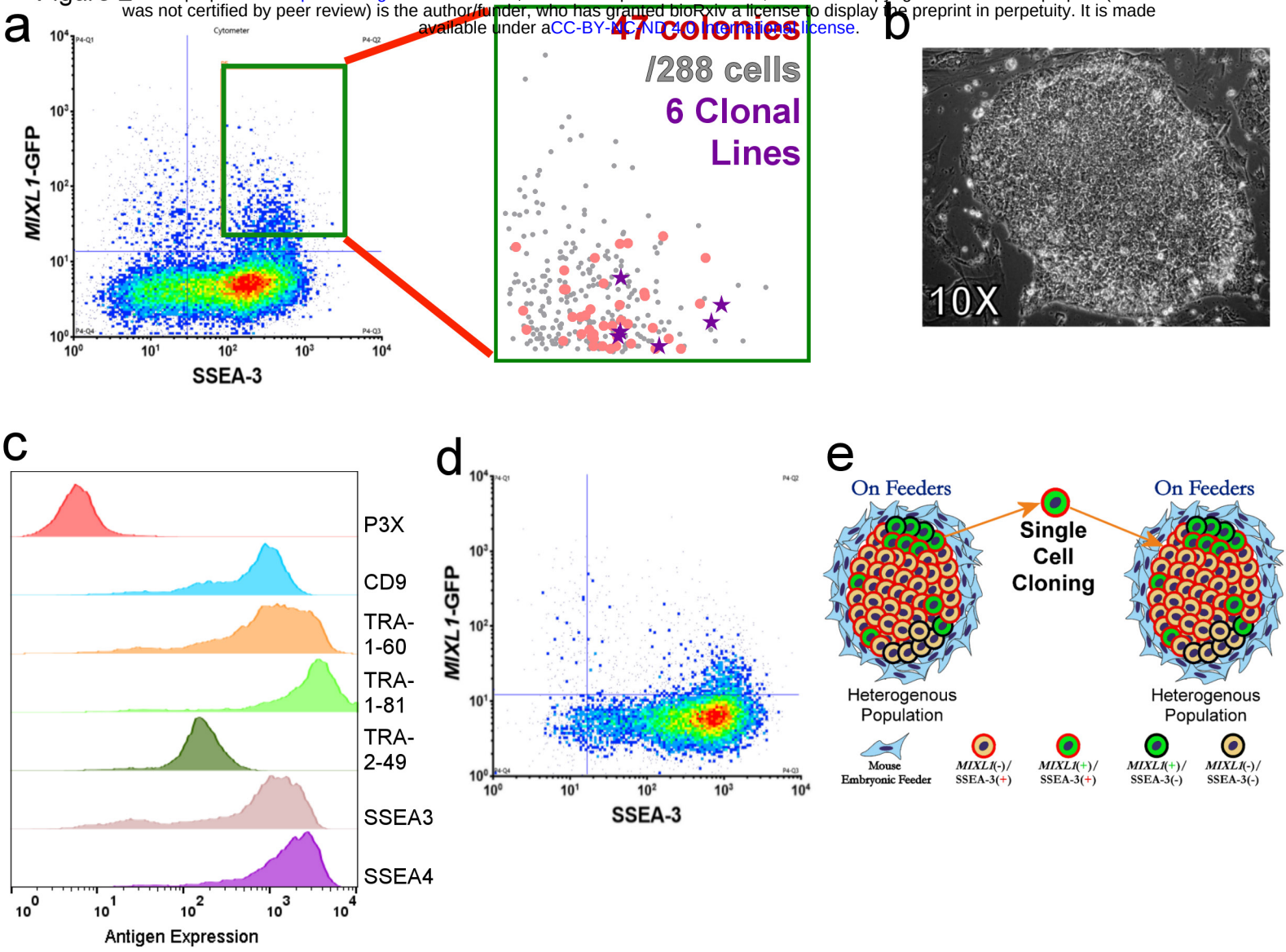


Figure 2: The *MIXL1*/SSEA-3 positive state contains a self-renewing stem cell.

a) Density dot plot of *MIXL1*-GFP expression against the surface marker SSEA-3 expression from cells growing in MEF/KOSR conditions. The green gate indicated was used to specify the *MIXL1*/SSEA-3 double positive fraction and cells from this region were sorted into three 96 well plates. The zoomed in scatter plot shows the indexed positions of the 47 isolated cells that formed colonies (red circles), including 6 that were selected randomly to generate clonal lines (purple stars) from 288 sorted cells, a cloning efficiency of ~16%. **b**) Phase contrast image of a typical HES3 *MIXL1*-GFP clone (Clone 2-D2) growing in MEF/KOSR conditions taken at 10x magnification. **c)** Flow cytometry histograms of clonal line 2D2 for the stem cell associated antigens CD9, SSEA-3, SSEA4, TRA-1-60s, TRA-1-81 and TRA-2-49 and negative control P3X. All lines displayed similar high expression of these surface markers. **d)** Representative density dot plot of *MIXL1*-GFP expression against the surface marker SSEA-3 expression from clonal line HES3 *MIXL1*-GFP 3-C6. All clonal lines re-establish similar *MIXL1*-GFP/SSEA-3 distributions as the starting populations. **e)** Schematic Diagram of the process of single cell cloning from *MIXL1*/SSEA-3 double positive cells. FACS separates double positive single cells from the heterogeneous population that exists on feeders. Established clonal lines grown on feeders repopulate the heterogeneous distribution seen in the starting population.

Figure 3

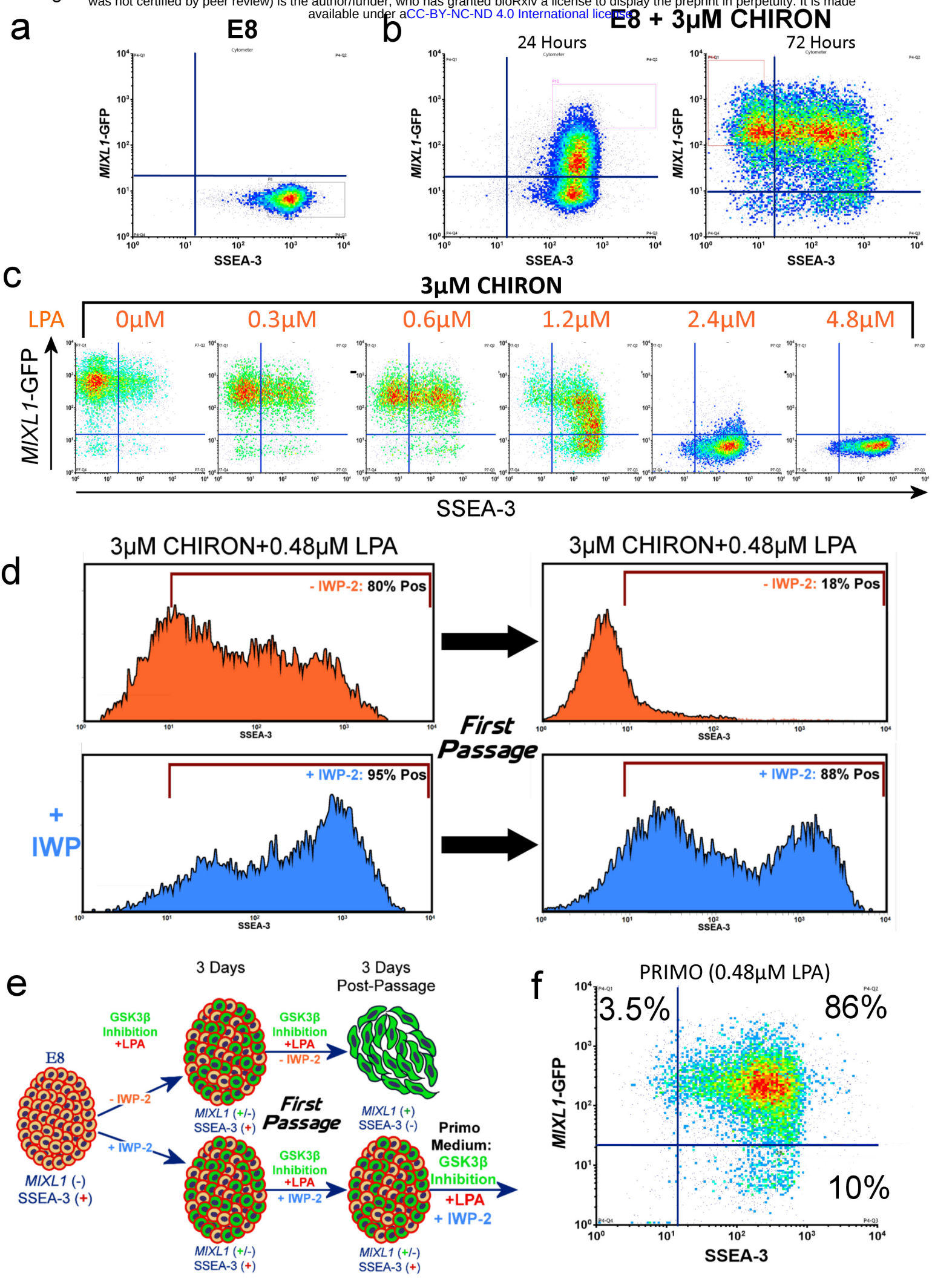


Figure 3: The balancing of pro-self-renewal and pro-differentiation signals enables a *MIXL1*/SSEA-3 positive population to be propagated in a fully defined system.

a-b) Flow cytometry density plots of SSEA-3 versus *MIXL1*-GFP for HES3 *MIXL1*-GFP cells growing in E8/Vitronectin **(a)** and E8/Vitronectin with 3μM CHIRON for 24 hours and 72 hours **(b).** In E8/Vitronectin SSEA-3 is highly expressed but *MIXL1*-GFP is not detected. The addition of CHIRON drives *MIXL1* expression and eventual loss of SSEA-3. **c)** Flow cytometry density plots displaying SSEA-3 expression versus *MIXL1*-GFP expression for HES3 *MIXL1*-GFP cells grown in E8 medium with 3μM CHIRON and increasing levels of LPA, Higher levels of LPA prevent *MIXL1*-GFP expression. **d)** Flow cytometry histograms of SSEA-3 expression for cells grown in 3μM CHIRON +0.48 μM LPA with IWP-2 (blue) or without IWP-2 (orange). Histograms are from 3 days in culture and 3 days post first passage. SSEA-3 was maintained post passage in samples cultured with IWP-2 present only. **e)** A schematic diagram of the development of "PRIMO" medium. Cells grown under GSK3B inhibition supplemented with LPA and with or without IWP2 induce expression of MIXL1(+)/SSEA-3(+) cells after 3 days. Post passage without IWP-2 addition cells differentiate, losing SSEA-3, but with IWP-2 cultures maintained SSEA-3 and some MIXL1 expression. **f)** After optimisation, the new formulation consisted of 3μM CHIRON, 1μM IWP-2 and 0.48μM LPA. Density plots display the *MIXL1*/SSEA-3 expression under PRIMO revealing high double expression.

Figure 4

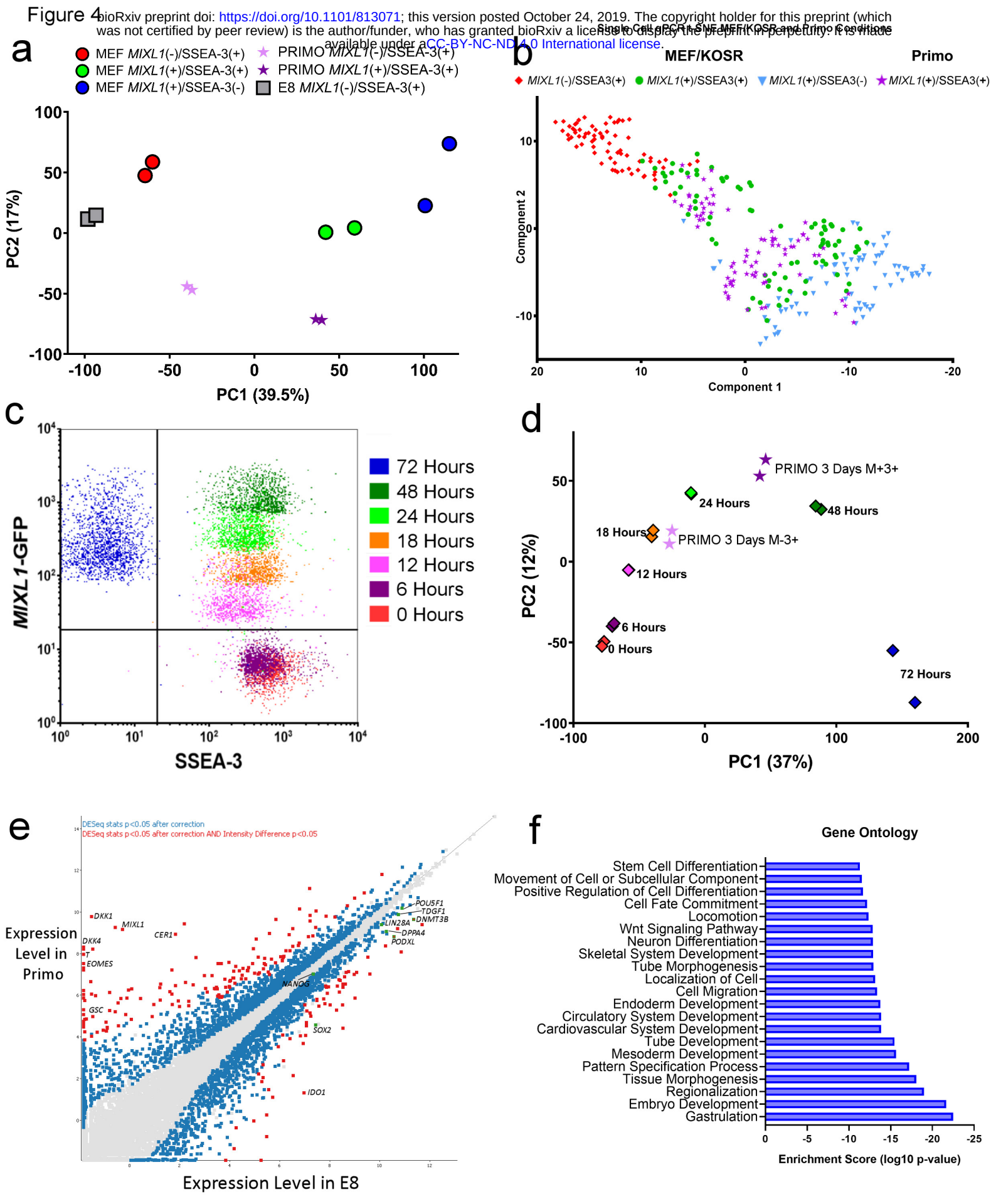
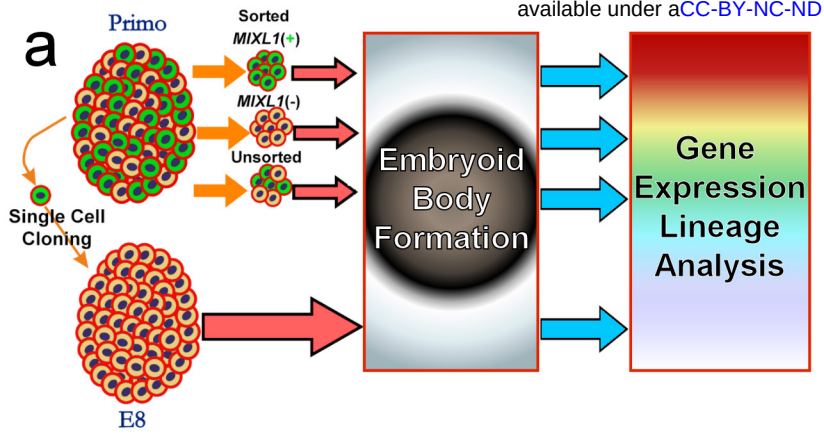
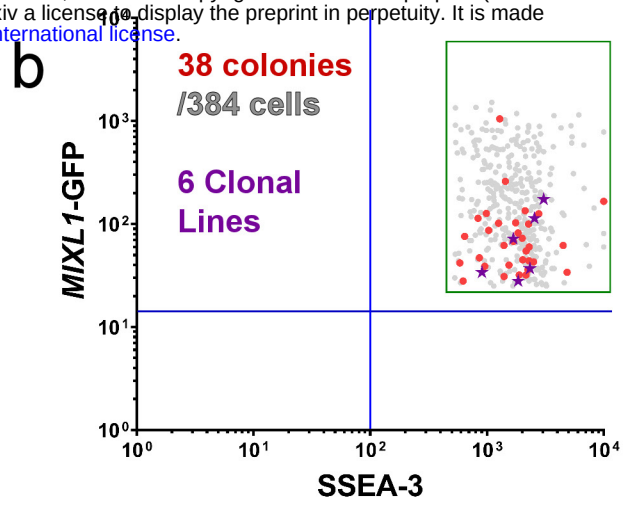


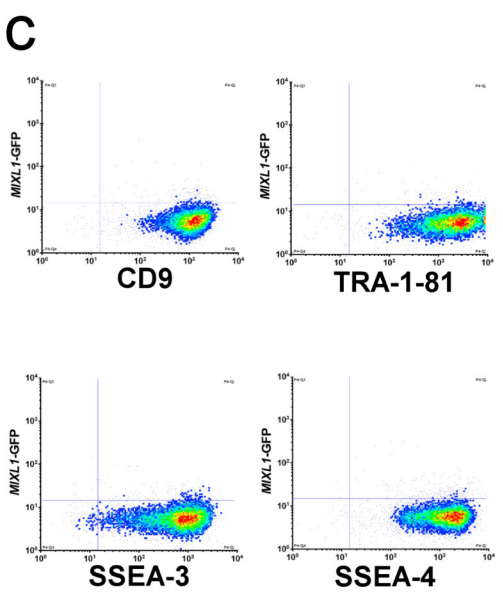
Figure 4: Gene expression analysis reveals that cells growing in PRIMO medium correlate with the *MIXL1*(+)/SSEA-3(+) fraction identified in MEF/KOSR conditions as well as corresponding to early differentiation when compared to a differentiation time course.

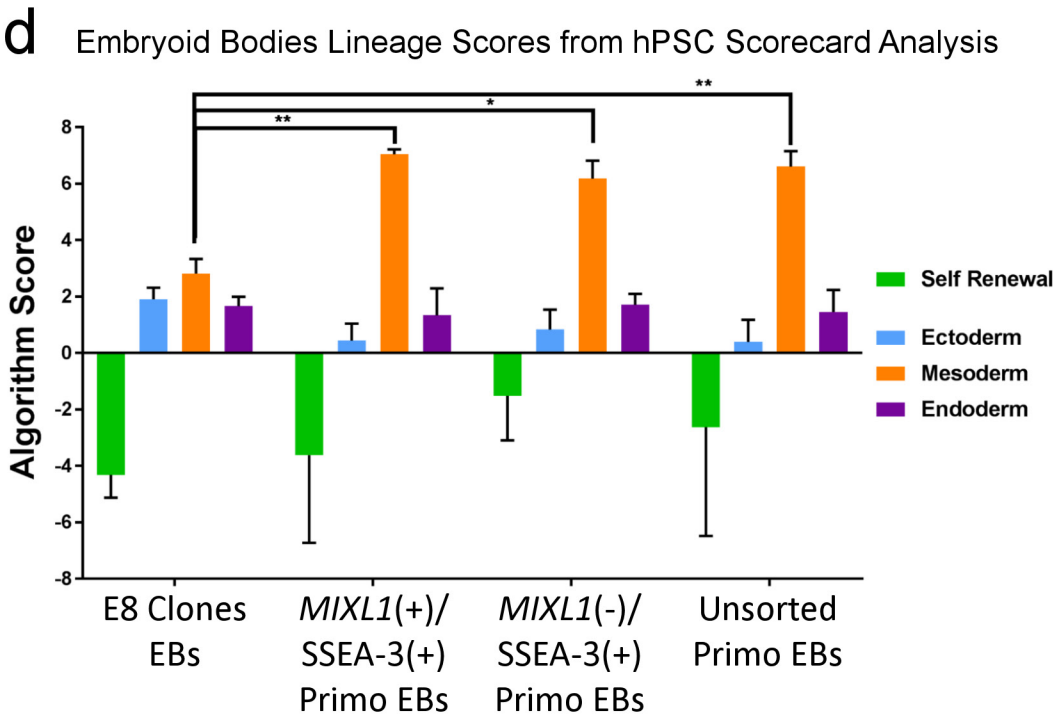
a) Principal component analysis of *MIXL1*(**+**)/SSEA-3(+) (dark purple stars) and *MIXL1*(-)/SSEA-3(+) (light purple stars) cells grown in PRIMO compared to the fractions identified in MEF/KOSR conditions and standard E8 culture. **b)** t-Distributed Stochastic Neighbour Embedding (t-SNE) plots of the expression of 45 genes in individual cells measured by Fluidigm Biomark qPCR and analysed by Monocle2. Displays all fractions in the same dimensional space from MEF/KOSR conditions (presented in Fig. 1c) *MIXL1*(-)/SSEA-3(+) cells in red, *MIXL1*(+)/SSEA-3(+) cells in green, *MIXL1*(-)/SSEA-3(+) cells in blue and PRIMO conditions *MIXL1*(+)/SSEA-3(+) cells in purple. **c)** Flow cytometry dot plot of the emerging population for which was sorted at the indicated timepoints during a differentiation time-course. Differentiation was induced incubating cells with E8 containing CHIRON at 3μM for 72 hours. **d)** Principal component analysis of *MIXL1*(+)/SSEA-3(+) (dark purple stars) and *MIXL1*(-)/SSEA-3(+) (light purple stars) cells grown in PRIMO compared to the differentiation time course. **e)** Gene expression scatter plot of bulk RNAseq analysis of cells grown in E8V compared to 1xPRIMO medium. Differentially expressed genes were identified by DESeq2 with a cut off p-value of <0.05 (blue squares) and the SeqMonk intensity difference filter with a p-value cut off <0.05 (red squares). Some key genes related to pluripotency are shown as green squares. **f)** Gene ontology (GO) statistical enrichment analysis was performed by ToppGene and Revigo on the list of genes from the SeqMonk intensity difference filter. GO terms which were significantly enriched are shown, the Log10 p-values for each GO term are displayed. GO terms related to early differentiation and gastrulation are significantly enriched.

Figure 5









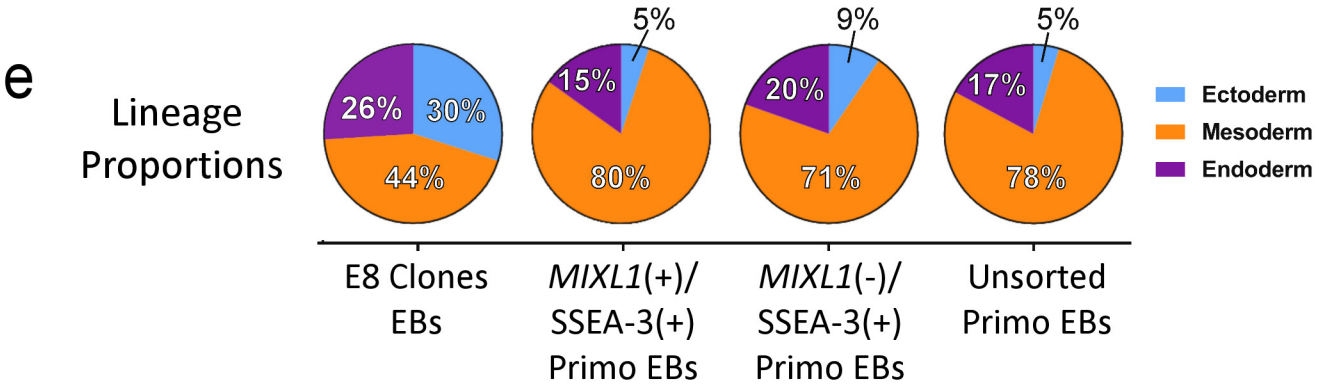


Figure 5: Cells grown in PRIMO medium exhibit a mesoderm bias under neutral differentiation conditions whilst also exhibiting clonogenic potential and the ability to revert to an unbiased state.

a) A schematic diagram of the experimental process. Cells were grown in PRIMO conditions for 3 days then cells were taken, unsorted or sorted for *MIXL1*(+)/SSEA-3(+) and *MIXL1*(-)/SSEA-3(+), and put through a "neutral" EB assay. After 7 days EBs were harvested and assessed for gene expression by hPSC scorecards. Separately single *MIXL1*(+)/SSEA-3(+) cells were sorted from cells grown in PRIMO conditions and clonal lines were established, these lines were then assessed by the same "neutral" EB assay. **b)** Single cells from *MIXL1*/SSEA-3 double positive fraction were sorted into four 96 well plates. The *MIXL1*/SSEA-3 scatter plot shows the indexed positions of the 38 isolated cells that formed colonies (red circles), including 6 that were selected randomly to generate clonal lines (purple stars) from 384 sorted cells, a cloning efficiency of ~10%. **c)** Flow cytometry density plots of clone 10-A4 grown in E8 conditions for surface markers CD9, SSEA-3, SSEA-4 and TRA-1-81 versus *MIXL1*-GFP, revealed high surface marker expression and virtually no *MIXL1*-GFP expression. **d)** A bar chart displaying the algorithm score for each sample, for self renewal and three lineages, ectoderm, mesoderm and endoderm. The algorithm score is calculated based on the qPCR values for genes of a given lineage, the 0 baseline is based on the average value of undifferentiated samples. EB samples have been normalised to their undifferentiated counterparts (bars are mean ±SD, n=3 biological replicates, significance assessed by two-way ANOVA analysis). **e)** Lineage proportion pie charts of the algorithm scores generated for the three germ layers. All samples grown in PRIMO medium generated EBs with enhanced mesoderm signatures, whereas clonal lines grown in E8 generated EBs containing all three germ layers with more similar proportions.

Figure 6

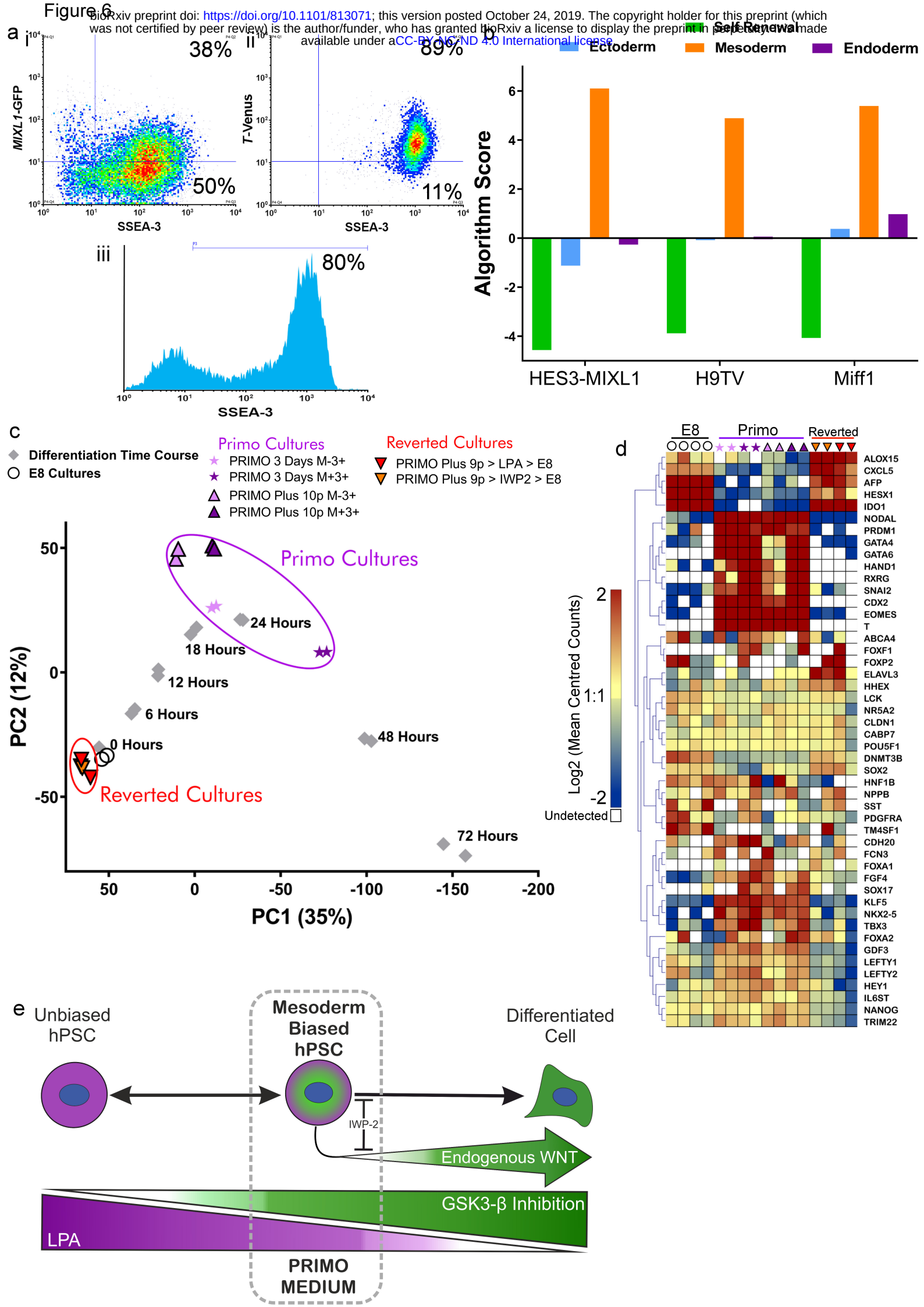


Figure 6: Cells can be maintained in a mesoderm biased pluripotent state in PRIMO medium for multiple passages whilst retaining the ability to revert back to a pristine pluripotent state.

a) Flow Cytometry analysis of cells grown in 2xPRIMO for 3 passages i)Density plot of MIXL1-GFP vs SSEA-3 for HES3 MIXL1 ii) Density plot of T-Venus vs SSEA-3 for H9 T-Venus iii) Histogram of SSEA-3 expression for MIFF1. **b)** Algorithm score for EBs generated from HES3-MIXL1, H9 T-Venus and Miff1 (iPS line) grown in PRIMO medium for 3 passages. All lines exhibited strong mesoderm signatures under neutral conditions. **c)** Principal component analysis of cells grown in PRIMO for 3 days or 10 passages compared to the differentiation time course. Also shown are cells which had been grown in PRIMO for 9 passages and then transitioned back into E8 conditions. Cells maintained in PRIMO cultures are close to 18-24 hours of differentiation and reverted cultures close to 0 hours (E8 only) cultures. **d)** Heatmap analysis of key genes related to Pluripotency and mesendoderm differentiation generated with Log2 counts that have been mean centred per gene with a color scale from -2 to 2. **e)** A schematic diagram of the “trapping” of a Mesoderm Biased hPSC along a normal differentiation trajectory. The cross antagonism of the pro-pluripotency factor (LPA) and pro-differentiation factor (GSK3β Inhibition) is able to induce a mesoderm biased state but endogenous WNT secreted by the cells drives further differentiation. Using IWP-2 to block this endogenous WNT secretion prevents this further differentiation. These three components of PRIMO medium permit the maintenance of the mesoderm biased state *in vitro.*