Reprogrammed blastoids contain amnion-like cells but not trophectoderm

Cheng Zhao[1], Alvaro Plaza Reyes[1,2], John Paul Schell[1], Jere Weltner[1], Nicolás M. Ortega[1,11], Yi Zheng[3], Åsa K. Björklund[4], Janet Rossant[5], Jianping Fu[3,6,7], Sophie Petropoulos*[1,8,9,10] & Fredrik Lanner*[1,11]

1. Department of Clinical Science, Intervention and Technology, Karolinska Institutet, and Division of Obstetrics and Gynecology, Karolinska. Universitetssjukhuset, Stockholm, Sweden.
2. Department of Regeneration and Cell Therapy, Andalusian Molecular Biology and Regenerative Medicine Centre (CABIMER), Seville, Spain.
3. Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.
4. Dept of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Husargatan 3, SE-752 37 Uppsala, Sweden.
5. Program in Developmental and Stem Cell Biology, Hospital for Sick Children, Toronto, Ontario, Canada.
6. Department of Cell & Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.
7. Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.
8. Département de Médecine, Université de Montréal, Montréal Canada
9. Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Axe Immunopathologie, H2X 19A Montréal, Canada
10. Département de Biochimie et Médecine Moléculaire, Université de Montréal, Montréal, Canada.
11. Ming Wai Lau Center for Reparative Medicine, Stockholm Node, Karolinska Institutet, Stockholm, Sweden.

Corresponding author

Abstract

With the advancement of human stem cell cultures and interest in understanding human embryogenesis, human blastocyst-like structures, or human blastoids, are being developed. Thus far, two different strategies have been taken, generating blastoids from naïve human pluripotent stem cells (hPSC)[1–3] or through reprogramming of adult human somatic cells[4]. All these studies utilize comparative transcriptome analyses to authenticate blastoid cells and identify their counterparts in the human blastocyst[5,6]. However, the validity of comparative transcriptome analysis is critically hinged on including relevant reference data, not only those from targeted cell types but also from potential alternative cell lineages. Thus, we sought to reevaluate the single-cell transcriptome data from the blastoids based on a more comprehensive cellular reference, which includes data from *in vitro* cultured human[6,7], non-human primate (NHP, *Cynomolgus macaque*) blastocysts[8], a human stem cell-based post-implantation amniotic sac embryoid (PASE) model[9], and an *in vivo* gastrulation-stage human embryo specimen[10], using four different analysis strategies. Our analyses unequivocally support that blastoids developed by reprogramming adult human somatic cells largely fail to generate cells with a transcriptome profile consistent with the human blastocyst trophectoderm. Instead, cells identified as trophectoderm-like have a transcriptional profile more similar to the amniotic ectoderm in the gastrulating human and NHP embryos. To facilitate cell lineage identifications in human embryo models, we further identified a set of human amniotic ectoderm and trophectoderm markers that could be utilized to distinguish these two lineages. We further built a neural-network based online prediction tool, which accurately discerns the full cellular composition of blastoids.

46 **Results**

47 Blastoids are promising tools to study early human development, resolving technical, ethical, and legal
48 limitations of natural human embryos[11]. However, the utility of blastoids is critically dependent on their
49 faithful recapitulation of natural human blastocyst development. Human pluripotent stem cells (hPSCs)
50 with primed pluripotency have been suggested to have the capacity to differentiate into the trophectoderm
51 (TE)[12]. More recent studies, however, suggest that the purported TE cells derived from primed hPSCs might
52 instead be mis-annotated extraembryonic mesoderm or amniotic ectoderm[13–15]. Similarly, putative TE cells
53 in mouse blastoids generated from extended pluripotent stem cells (EPSCs) are now believed to instead
54 represent mesoderm-like cells[11]. In light of these conflicting views and a critical need for accurate
55 annotations of cell lineages in embryoids, we sought to reevaluate transcriptome data from all human
56 blastoids[1–4]. To discern cell lineage identities and examine the possible presence of mesoderm- or amniotic
57 ectoderm-like cells in the blastoids, we generated an integrated dataset, combining blastoid transcriptome
58 data with reference data from *in vitro* cultured human blastocysts[3,6,7], a human stem cell-based post-
59 implantation amniotic sac embryoid (PASE) model[9], and a Carnegie Stage (CS) 7 human gastrula[10]. UMAP
60 projection of this integrated data clearly revealed distinct cell clusters, including those annotated as the
61 epiblast, TE, endoderm, amnion, and mesoderm lineages based on known lineage markers (Fig. 1a and
62 Extended Data Fig. 1a-c). Notably, cells annotated as trophoblast-like cells (TLC) in reprogrammed
63 blastoids (Liu-Blastoids) did not contribute to the TE cluster; rather, the majority of TLC in Liu-Blastoids
64 overlap with the amnion cluster (Fig. 1a-b and Extended Data Fig. 1b). In contrast, all three naïve hPSC-
65 derived blastoids contain expected cell types, with most of the annotated epiblast-like cells (ELC),
66 hypoblast-like cells (HLC), and TLC overlapping with their counterparts in the human embryo.
67 Furthermore, inclusion of transcriptome data from naïve and primed hPSCs as well as trophoblast stem
68 cells representative of post-implantation human trophoblast lineage[2,9] together with the integrated dataset
69 enabled us to explore developmental stages of blastoid cell lineages (Extended Data Fig. 1a). Among the
70 naïve hPSC-derived blastoids, those cultured in the PXGL condition (KAG-Blastoids[2] and YAN-
71 Blastoids[3]) generally comprise cell lineages that align better with those in preimplantation human

blastocysts, whereas blastoids generated from hPSCs cultured in the 5i/L/A condition still contain a notable portion of cells resembling those in postimplantation human embryos (Yu-Blastoids[1]) (Fig. 1b). For convenience and further exploration, our transcriptome comparative analyses can be browsed at (http://petropoulos-lanner-labs.clintec.ki.se/).

A single human dataset containing transcriptome of both amniotic ectoderm and TE is currently lacking. Thus, we utilized recently published data of cultured non-human primate (NHP, *Cynomolgus macaque*)[8], which includes transcriptome of both amniotic ectoderm and TE, to perform a second analysis. Integration of the CS7 human gastrula and NHP data confirmed proper alignments of corresponding cell lineages from these two species (Extended Data Fig. 1c). Using the NHP data as the reference for comparative transcriptome analyses of the blastoids clearly shows that the majority of TLC in Liu-Blastoids clusters with the amnion as opposed to the TE lineage (Fig. 1c), corroborating our finding in Fig. 1b. We should note that this finding could be reached by common integration methods MNN, Harmony or CCA (Seurat) (Extended Data Fig. 1d-e).

We next analyzed the proportion of cell types identified in the blastoids using our re-annotations obtained from the multi-data integration. TE cells should be the dominant cell type in human blastocysts. In contrast to human blastocysts, all blastoids contain a much greater EPI population (Extended Data Fig. 2a-b). It is also notable that in addition to containing a significant portion of amnion-like cells, Liu-Blastoids also carried substantial numbers of mesoderm-like and unidentified cells (Fig. 1a-b and Extended Data Fig. 2b-c). Mesoderm-like and unidentified cells can also be found in Yu-Blastoids and to a lesser degree in blastoids from PXGL culture condition (KAG-Blastoids[2] and YAN-Blastoids[3]; Fig. 1a-b and Extended Data Fig. 2b-c).

Differential gene expression analysis incorporating TE lineages from both pre- and post-implantation human embryos[6,7] and amnion cells from the CS7 gastrula[10] allowed the identification of 173 and 14 genes highly enriched in TE lineage and amnion, respectively (Fig. 2a, Extended Data Table1.tsv). We further validated these gene signatures by examining two additional preimplantation human blastocyst datasets[3,5] and the NHP data (Fig. 2a and Extended Data Fig. 2d). TLC in Liu-Blastoids showed distinct expression

98 of the amnion markers (Fig. 2b). Examining normalized expression of selected amnion markers further
99 showed that TLC in Liu-Blastoids highly expresses amnion markers such as *ISL1*, *GABRP,* and *IGFB5*
100 (Fig. 2b)[8,10,16], with low expression of TE lineage markers including *GATA2, GCM1,* or *BIN2*. ISL1 and
101 GABRP are specifically expressed in the amnion but not in the TE lineage of *in vitro* cultured NHP
102 embryos, at both the transcriptional and protein levels[8]. Furthermore, ISL1 and GABRP are not detectable
103 in the TE of *in vitro* cultured human blastocysts at both the transcriptional and protein levels[17].
104 In an effort to compare multiple datasets without the need of integration or single-cell normalization, we
105 collapsed all the datasets into pseudo-bulk. A PCA plot was generated for the human TE, TSC, and amnion
106 datasets together with the blastoids based on the 187 differentially expressed genes (DEGs) (Fig. 2c). The
107 first component, which explains 72.51% of all variations, represents the TE to amnion axis. While TLCs of
108 the three hPSC-derived blastoids are located on the left side of the plot with the TE datasets, TLC of Liu-
109 Blastoids are positioned well to the right towards the amnion of the CS7 human gastrula and the amnion-
110 like cells (AMLC) in the PASE model. To further validate this analysis, we incorporated the NHP data to
111 generate another PCA plot (Extended Data Fig. 2d-e). This cross-species analysis further supports that TLC
112 of Liu-Blastoids are more similar to the amnion at the transcriptome level.
113 Finally, we used neural-network to build an Early Embryogenesis Prediction tool (http://petropoulos-
114 lanner-labs.clintec.ki.se/) based on data from human blastocysts and the CS7 human gastrula[6,7,10] (Fig. 2d
115 and Extended Data Fig. 3a-c). The prediction performance of this tool was validated using a large number
116 of additional datasets[3,5,18], with a specificity of 97.5% and a sensitivity of 88.6% (Extended Data Fig. 3d).
117 The neural network prediction also supports that Liu-Blastoids lack TE and instead those annotated TLC
118 in Liu-Blastoids are more similar to the amnion at the transcriptome level (Fig. 2d). In conclusion, to discern
119 cell lineages in human embryoids including blastoids based on transcriptome data, it is imperative to use
120 multiple independent analyses together with unbiased, comprehensive reference datasets. To validate cell
121 identities in human embryoids, expression of key cell fate markers, such as those identified in this work,
122 should also be clearly demonstrated whenever possible. We envision our analyses, identified amnion and

TE markers and the Early Embryogenesis Prediction tool should be useful for authentication of human embryoids, thus promoting the healthy development of this emerging field.

Figure. 1 Integration of blastoid cells with an extensive reference dataset, including data from human, non-human primate and human embryoid references.

a&b, UMAP projection of the eight-dataset integration. Embryonic and blastoid cells are shown in **a** and **b**, respectively. The color of each data point represents the cell annotations retrieved or restored for each publication. The light grey colored data points represent cells used in the multi-data integration which were not used to infer cell annotations. The shape of data points for the embryonic cells indicates data source. Cells not assigned to a major lineage from the blastoid datasets were labeled as "Undefined" and colored in dark grey. **c,** UMAP projection of Mutual Nearest Neighbours (MNN) cross-species integration, including cells from each blastoid dataset and the *in vitro* non-human primate (NHP) embryo at Day 14, colored by cell type. The shape of data points indicates species source. Embryonic cells from Yanagida et al., and naïve, primed and Okae cells from Kagawa et al., were also included in the integration.

Figure. 2 Validation of cell lineage identities in blastoids.

a, Expression of significant differentially expressed genes in embryonic amnion, TE, and trophectoderm-like cell (TLC) from the four blastoid models. TE including TE cells from Blakeley et al. (TE-1), from Petropoulos et al. (TE-2), from Yanagida et al. (TE-3), from pre-implantation cells (TE-4), and from post-implantation cells (TE-5) in Xiang et al. For visualization, cells from the same cell type were randomly down-sampled to 50. The color shown in the left and middle heatmap represents scaled values of log-transformed expression in each cell. The heatmap to the right of the main heatmap represents scaled average log-transformed expression values calculated from TLCs of the blastoid datasets. **b,** Violin plots showing log-transformed expression of key amnion and TE markers in amnion, TE, amnion-like cells (AMLC), and TLC from the four blastoids. **c,** Principal components analysis (PCA) of the pseudo-bulk expression profiles of amnion, TE, AMLC and Tsw-AMLC from the PASE model, Okae TSC cells, and TLCs from the four blastoids. TE (1-5) in **a** and **b** represent the same source of TE as in **c. d,** Alluvial plot comparing the original cell-type annotations (epiblast-like cells (ELC), hypoblast-like cells (HLC), and TLC from the four

150 blastoids) to the predicted identities obtained from the Early Embryogenesis Prediction tool. The
151 predictions identified as "ambiguous" or "NoSigHits" were removed for visualization.

Acknowledgments

We thank members of the laboratories of F.L. and S.P. for discussions. This work was supported by the Swedish Research Council (F.L., S.P.), Ragnar Söderberg Foundation (F.L.), Ming Wai Lau Center for Reparative Medicine (F.L.), Center for Innovative Medicine (F.L.), Wallenberg Academy Fellow (F.L.), Swedish Society for Medical Research (S.P.), Emil och Wera Cornells Stiftelse (S.P.), Natural Sciences and Engineering Research Council of Canada Discovery Grant (S.P), The Canadian Institutes of Health Research (S.P., J.R.), Päivikki and Sakari Sohlberg Foundation (J.W.), and Sigrid Jusélius Foundation (J.W.). SP holds the Canada Research Chair Functional Genomics in Reproduction and Development.

Contributions

C.Z. and F.L. conceived the study with advice from A.P-R., J.P.S, and S.P. C.Z., A.P-R., and J.P.S., performed the data analyses with help from J.W., S. P., N.M.O. Y.Z., J.R., J.F., Å.B. The manuscript was written by C.Z., A.P-R, S.P., and F.L., with input from all of the authors.

Corresponding author

Correspondence to fredrik.lanner@ki.se

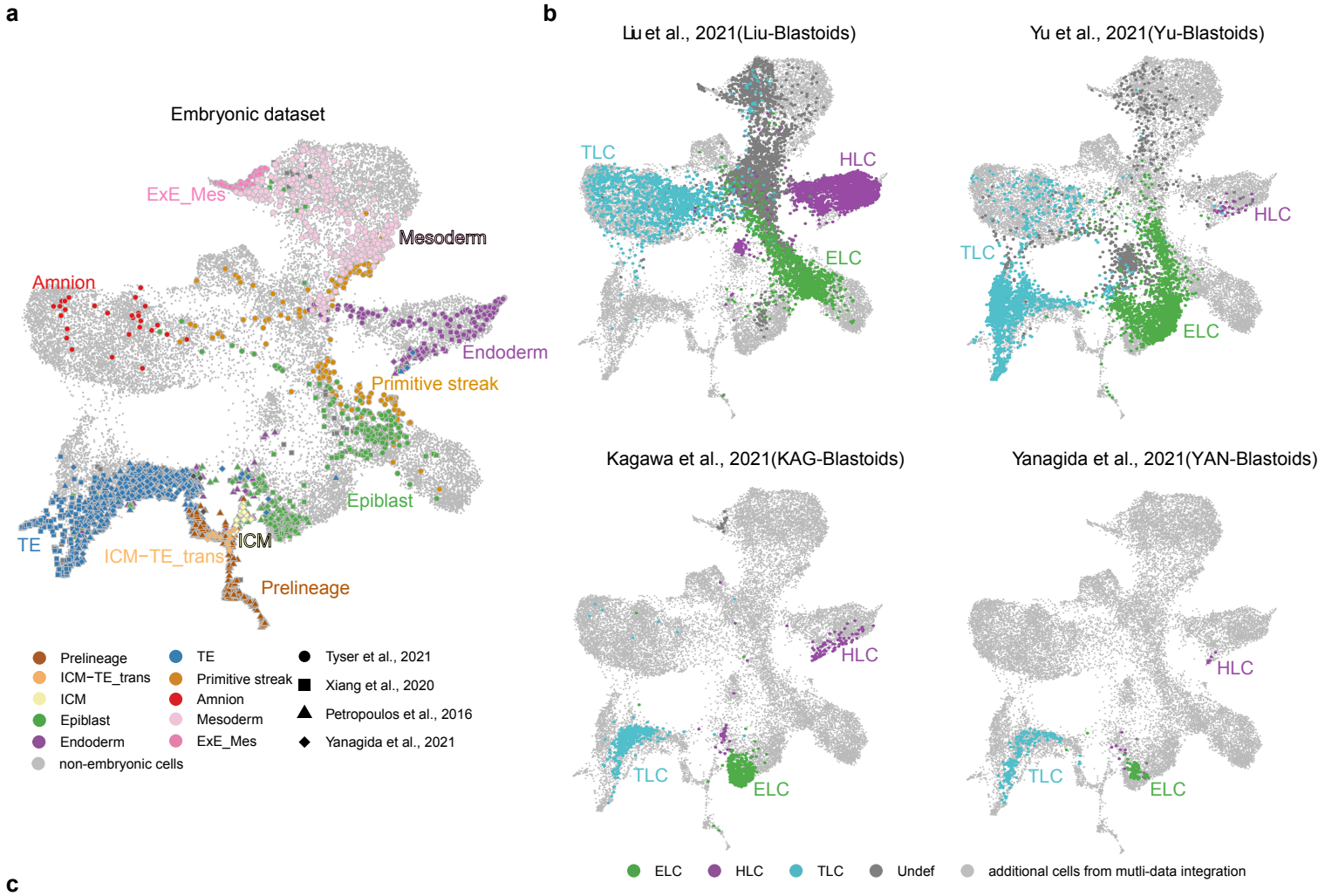
Ethics declarations

Competing interests

The authors declare no competing interests.

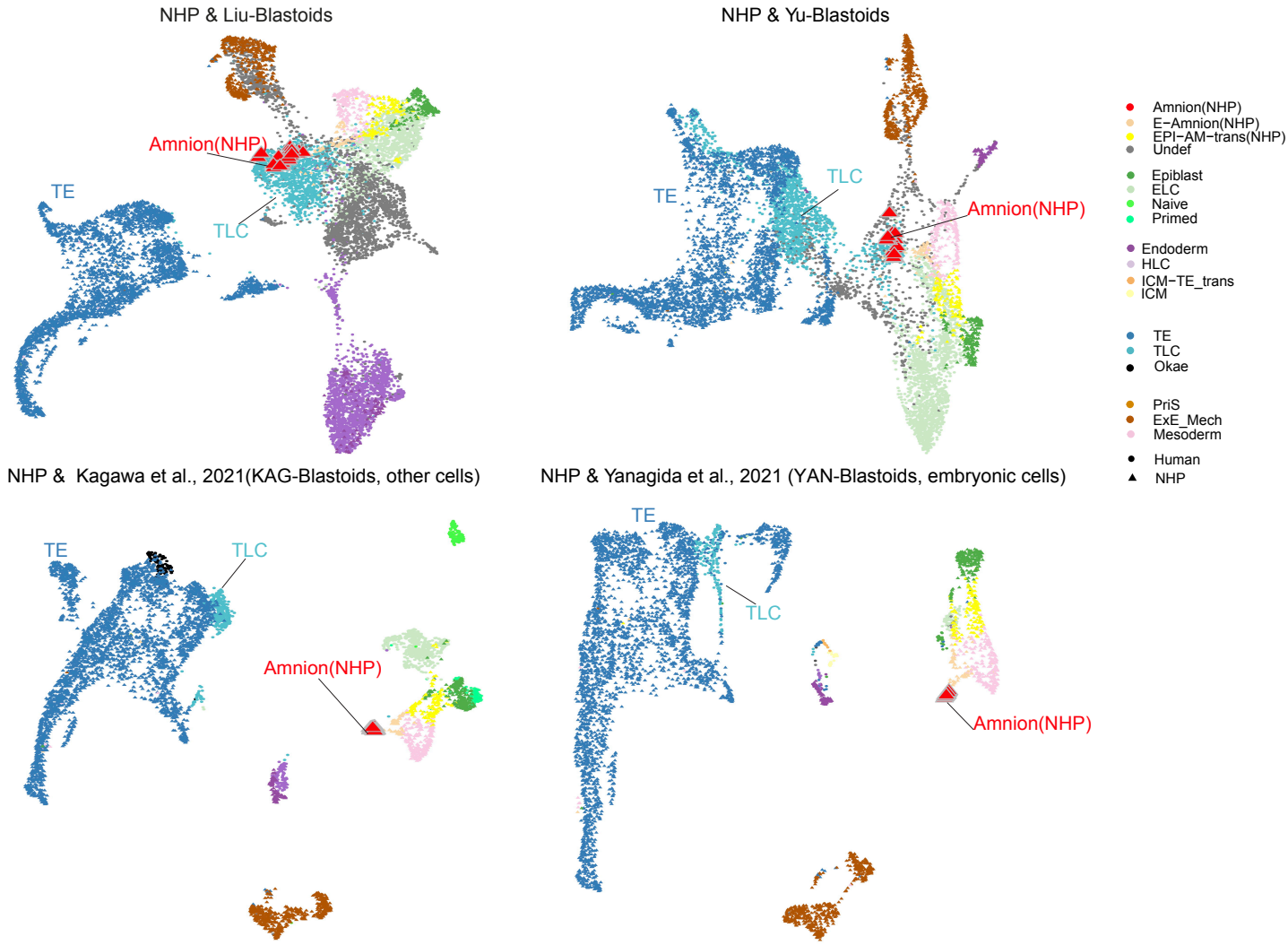
172 **Reference**
173
174 1. Yu, L. *et al.* Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **591**,
175 620–626 (2021) doi:10.1038/s41586-021-03356-y.
176 2. Kagawa, H. *et al.* Human blastoids model blastocyst development and implantation. *Nature* 1–9
177 (2021) doi:10.1038/s41586-021-04267-8.
178 3. Yanagida, A. *et al.* Naive stem cell blastocyst model captures human embryo lineage segregation.
179 *Cell Stem Cell* **28**, 1016-1022.e4 (2021).
180 4. Liu, X. *et al.* Modelling human blastocysts by reprogramming fibroblasts into iBlastoids. *Nature*
181 **591**, 627–632 (2021) doi:10.1038/s41586-021-03372-y.
182 5. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-
183 seq. *Development* (2015) doi:10.1242/dev.123547.
184 6. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in
185 Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016).
186 7. Xiang, L. *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos.
187 *Nature* **577**, 537–542 (2020).
188 8. Yang, R. *et al.* Amnion signals are essential for mesoderm formation in primates. *Nat. Commun.*
189 *2021 121* **12**, 1–14 (2021).
190 9. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using stem
191 cells. *Nature* **573**, 421–425 (2019).
192 10. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo.
193 *Nature. 2021 6007888* **600**, 285–289 (2021).
194 11. Posfai, E., Lanner, F., Mulas, C. & Leitch, H. G. All models are wrong, but some are useful:
195 Establishing standards for stem cell-based embryo models. *Stem Cell Reports* **16**, 1117 (2021).
196 12. Xu, R.-H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature*
197 *biotechnology* vol. 20 (2002).
198 13. Chhabra, S. & Warmflash, A. BMP-treated human embryonic stem cells transcriptionally
199 resemble amnion cells in the monkey embryo. *Biol. Open* **10**, (2021).
200 14. Bernardo, A. S. *et al.* BRACHYURY and CDX2 mediate BMP-induced differentiation of human
201 and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144–155
202 (2011).
203 15. Guo, G. *et al.* Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell*
204 **28**, 1040-1056.e6 (2021).
205 16. Roost, M. S. *et al.* KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal
206 Transcriptional Atlas. *Stem Cell Reports* **4**, 1112–1124 (2015).
207 17. Zheng, Y. *et al.* Single-Cell Analysis of Embryoids Reveals Lineage Diversification Roadmaps of
208 Early Human Development. *SSRN Electron. J.* (2021) doi:10.2139/SSRN.3952092.
209 18. Zhou, F. *et al.* Reconstituting the transcriptome and DNA methylome landscapes of human
210 implantation. *Nature* **572**, 660–664 (2019).

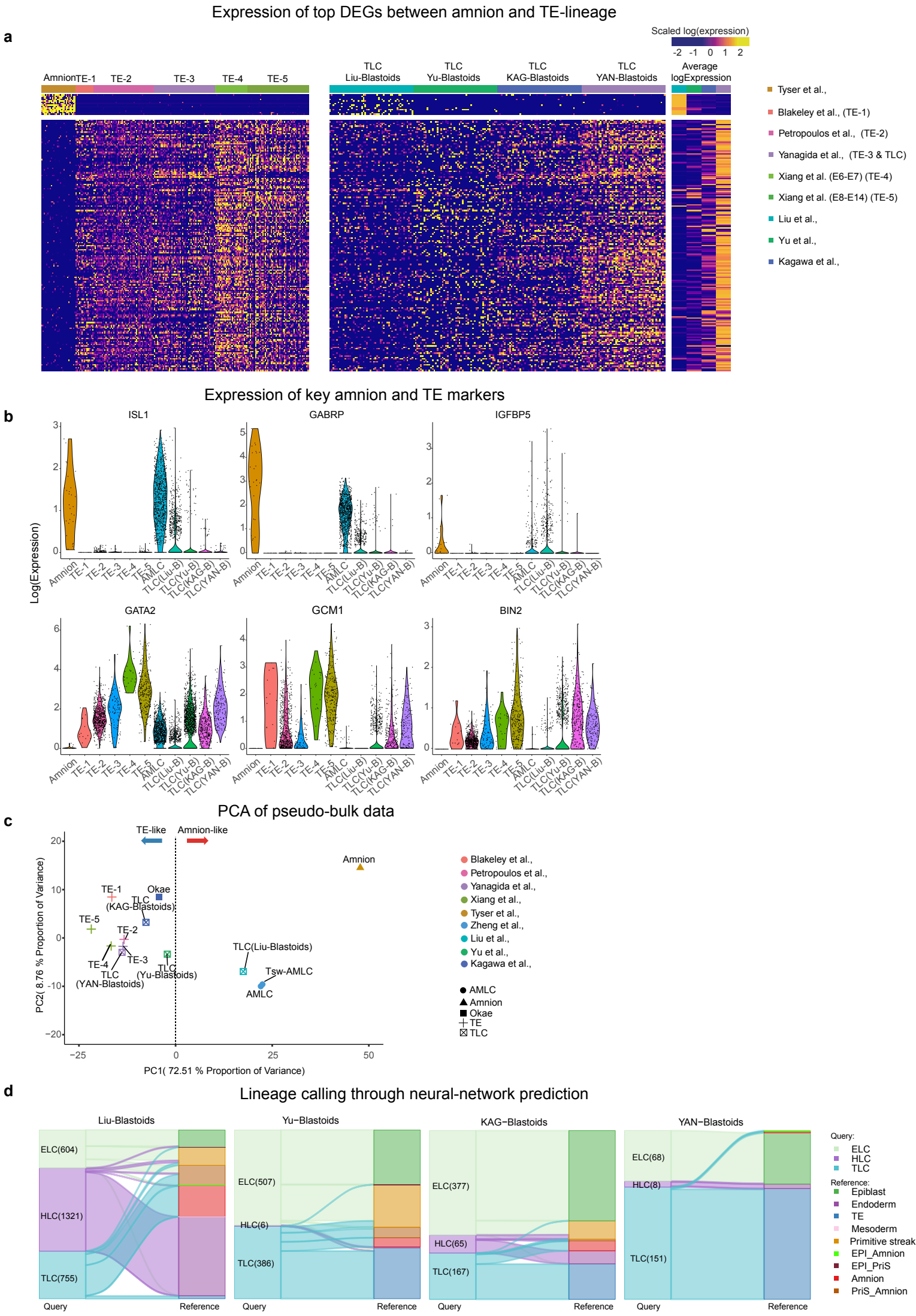
Integrated analysis including embryonic cells and blastoid models



c

Cross-species integration including non-human primate embryos and blastoid models




a
Expression of top DEGs between amnion and TE-lineage
Scaled log(expression)
-2 -1 0 1 2
Amnion
TE-1
TE-2
TE-3
TE-4
TE-5
TLC Liu-Blastoids
TLC Yu-Blastoids
TLC KAG-Blastoids
TLC YAN-Blastoids
Average logExpression
Tyser et al.,
Blakeley et al., (TE-1)
Petropoulos et al., (TE-2)
Yanagida et al., (TE-3 & TLC)
Xiang et al. (E6-E7) (TE-4)
Xiang et al. (E8-E14) (TE-5)
Liu et al.,
Yu et al.,
Kagawa et al.,
b
Expression of key amnion and TE markers
ISL1
GABRP
IGFBP5
GATA2
GCM1
BIN2
Log(Expression)
Amnion
TE-1
TE-2
TE-3
TE-4
TE-5
AMLC
TLC(Liu-B)
TLC(Yu-B)
TLC(KAG-B)
TLC(YAN-B)
c
PCA of pseudo-bulk data
TE-like
Amnion-like
PC2(8.76 % Proportion of Variance)
PC1(72.51 % Proportion of Variance)
Amnion
TE-1
Okae
TLC (KAG-Blastoids)
TE-5
TE-2
TE-3
TE-4
TLC (YAN-Blastoids)
TLC (Yu-Blastoids)
TLC(Liu-Blastoids)
Tsw-AMLC
AMLC
Blakeley et al.,
Petropoulos et al.,
Yanagida et al.,
Xiang et al.,
Tyser et al.,
Zheng et al.,
Liu et al.,
Yu et al.,
Kagawa et al.,
AMLC
Amnion
Okae
TE
TLC
d
Lineage calling through neural-network prediction
Liu-Blastoids
Yu-Blastoids
KAG-Blastoids
YAN-Blastoids
ELC(604)
HLC(1321)
TLC(755)
ELC(507)
HLC(6)
TLC(386)
ELC(377)
HLC(65)
TLC(167)
ELC(68)
HLC(8)
TLC(151)
Query
Reference
Query:
ELC
HLC
TLC
Reference:
Epiblast
Endoderm
TE
Mesoderm
Primitive streak
EPI_Amnion
EPI_PriS
Amnion
PriS_Amnion