# A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming

Ane Iturbide[1] & Maria-Elena Torres-Padilla[1,2] *

[1] Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München D-81377 München, Germany

[2] Faculty of Biology, Ludwig-Maximilians Universität, München, Germany.

* torres-padilla@helmholtz-muenchen.de

**Abstract**

Multicellular organisms develop from a single cell, the zygote. This feature is referred to as totipotency. In the mouse, only the zygote and the 2-cell stage embryo display this attribute. Cells resembling the embryonic 2-cell stage blastomeres were identified in embryonic stem (ES) cell cultures as "2-cell-like cells" (2CLCs). This discovery brought the first cellular model with the possibility to investigate some features of the totipotent embryo and the molecular mechanisms regulating totipotency *in vitro*. In this article, we discuss the latest advancements on the research on 2CLCs, which have uncovered an intricate reprogramming process regulated by proteins as well as metabolites and ncRNAs. These recent findings have shed light on the combinatorial regulation of 2-cell-like cell emergence and the nature of their unique attributes.

## Introduction

Totipotency, the capacity to generate a full organism, is a unique feature that mammalian cells acquire right after fertilzation[1,2]. Totipotent zygotes can naturally develop into all cells that form the embryo (embryonic tissues) and the annexes required for the embryo to develop (extraembryonic tissues). As development progresses, this capacity is lost and cells gradually lose potency to give way to cell differentiation. In mouse, totipotency is restricted to zygote and 2-cell stage blastomeres[3–5]. When the embryo reaches the blastocyst stage, the cells in the inner cell mass (ICM) are considered pluripotent since they can later differentiate into all the embryonic tissues, including the germline[6]. Pluripotent ES cells are derived from the ICM and recapitulate the main molecular features and potency of their *in vivo* counterpart. ES cells have been used as a cellular model for pluripotency for decades

and have helped us to better understand the molecular mechanisms regulating this cellular state and further differentiation processes[7].

In contrast to pluripotency, very little is known about the mechanisms underpinning totipotency, its establishment, core regulatory network and how the exit from this state is regulated. Studying totipotency has primarily been done through investigations on the mammalian embryo, which can be limiting in amount to the application of many molecular and biochemical approaches. Therefore, the advent of a cellular model to study totipotent-like features has recreated an enormous research interest in the last years. In 2012 cells resembling the 2-cell stage embryo, also known as 2-cell-like or 2C-like cells (2CLCs), were identified in ES cell cultures[8]. While the degree of similarity with their *in vivo* counterpart is still a matter of ongoing research, their discovery has opened new possibilities for the study of totipotency or totipotency-like features, and to accelerate the advancement of the field. In this Current Opinion article we will comment on the recent advances in the 2CLCs research field, in particular covering the last 2 years.

**The initial characterization of 2CLCs**

The initial work describing a subpopulation of cells resembling the blastomeres of the 2-cell stage embryo was reported by Macfarlan and colleagues[8]. 2CLCs were identified upon monitoring the transcriptional activity of MERVL in ES cells. MERVL belong to a family of transposable elements that contain Long Terminal Repeats (LTR), which drive their transcription. MERVL were known to peak in expression levels at the 2-cell stage in embryos[9]. 2CLCs have been since studied and characterised using a “2C reporter”, which contains the LTR from MERVL, thereby reflecting their transcriptional activation. 2CLCs are present in both ES and iPS cells and share several key features with the 2-cell stage embryo (reviewed in 10). These include an extended developmental capacity to extraembryonic tissues, partial activation of the transcriptional programme of 2-cell stage embryos and the down-regulation of the pluripotency core network proteins such as OCT4. Further characterization of 2CLCs has revealed that these cells display higher reprogrammability upon somatic cell nuclear transfer[11], similar chromatin accessibility profile to the 2-cell embryo[12,13••], greater histone mobility[14] and dispersed chromocentres[11].

**DUX and transposable elements: key players in 2CLC emergence**

2CLCs can be induced in ES cell cultures by TSA, a histone deacetylase inhibitor, but also their number varies depending on the culture conditions (e.g. 2i versus

serum/LIF)[8]. Several negative regulators of 2CLC emergence have been identified, such as Lsd1[8] and CAF-1[11] (Table 1). In general, the recent efforts have focused on finding new factors with a role in reprogramming ES cells to 2CLCs. A key positive regulator was identified recently, the double homeodomain transcription factor DUX, which acts as transcriptional activator during zygotic genome activation (ZGA) in embryos, a process that occurs primarily at the 2-cell stage in the mouse[13••,15••,16]. In ES cells, DUX overexpression leads to a partial activation of the ZGA transcriptional programme and to reprogramming to a 2C-like state. The latter is also characterised by changes in chromatin accessibility, which resemble that of the 2-cell stage embryo[13••]. Similarly to the 2-cell embryo, MERVL repeats and chimeric transcripts derived from the MERVL LTR are speculated to play a key role in activating the 2C transcriptome. MERVL are directly bound and induced by DUX itself. Although the activity of DUX on some ZGA genes is clearly demonstrated *in vivo* and in 2CLCs, its role during embryonic development seems not to be essential. While several studies show slightly different conclusions[15••,17,18], presumably due to differences in genetic background and experimental design, the data points towards a non-essential role for developmental progression. However, RNAseq analysis on *Dux*-null embryos documents a clear transcriptional defect during ZGA, albeit to a different extent depending on the study.

Recently, another type of transposable element has been identified to have a role in regulating the 2C transcriptome: the LINE1 elements[19,20]. LINE1 belongs to the non-LTR family of retrotransposons and its potential role on transcriptional regulation would follow a very different mechanism. It is known for many years, that ES cells express LINE1 repeats and that its RNA is localized in the nucleus. A recent report documents that nuclear LINE1 RNA can bind to Nucleolin and KAP1 to sequester the *Dux* locus and keep it transcriptionally silent, therefore inhibiting the transition to the 2C-like state[20]. In agreement with this work, the role of KAP1 in the repression of the *Dux* locus has been described[15••], as well as the activation of some 2C genes in the absence of KAP1[8]. SUMO also plays a role in repressing *Dux* by tethering the non-canonical polycomb complex PRC1.6, potentially enabling the downstream deposition of H2AK119ub[21]. Although SUMO activity seems important for *Dux* regulation, its function in 2CLC reprogramming goes beyond Dux regulation. In ES cells, SUMO plays a role in repressing large genomic regions containing 2C-like genes, as well as reprogramming resistant regions (RRRs). Knocking-down UBC9 -the E3-ligase responsible for SUMO deposition- leads to a dramatic (30%-50%) reduction in H3K9me3 levels in ESCs[21]. This, together with its role in regulating other cell fate transitions, points towards a general role of SUMO as a repressor of reprogramming or

a ‘gatekeeper’ of cell fate, rather than as a specific regulator of the 2CLC transition[21]. In terms of 2CLC reprogramming, SUMO may also act through the action of another E3-ligase, PIAS4 (PIASy). Deletion of PIAS4 in ES cells results in the activation of a part of the 2C genes[22]. Importantly, perturbing PIAS4 levels in zygotes results in defective ZGA in vivo and in developmental arrest[22,23].

**The path of 2CLC reprogramming**

The dynamics of the reprogramming process from ES cells to 2CLCs are still unclear. Single cell expression analysis uncovered an intermediate cell population expressing ZSCAN4, but which has not activated MERVL elements[24•]. This ZSCAN4$^+$ cell population is ten times more abundant than 2CLCs and expresses some 2C genes, overall showing an intermediate transcription profile between ES and 2CLCs[24•]. In addition, similarly to 2CLCs, the ZSCAN4$^+$ intermediate population already shows downregulation of pluripotency factors at the protein level. Importantly, 2CLCs arise primarily from this population as shown by live cell imaging using a reporter cell line for both ZSCAN4 and MERVL[24•], demonstrating that ZSCAN4$^+$ cells are not only an intermediate population regarding transcriptional and protein marker features but also temporally.

Following a similar approach, later work conducted single cell RNAseq analysis on DUX-driven reprogramming to 2CLCs, also uncovering an intermediate cellular state. Similarly to previous work, this intermediate population shows downregulation of pluripotency factors, however, it does not show upregulation of 2C genes, nor of *Zscan4*[25]. This population could represent an intermediate state between ES cells and ZSCAN4$^+$ cells, which would later transition to a 2C-like state. Live imaging data will be important to determine whether this intermediate state represents rather a different subpopulation emerging as a result of DUX overexpression, without reflecting the ‘natural’ reprogramming of endogenous, spontaneously arising 2CLCs.

**ZGA initiation: emerging regulators upstream of DUX**

Although DUX can activate many 2C genes and drive reprogramming to a 2C-like state in cell culture, in the embryo *Dux* transcription is activated during ZGA. This raises the interesting hypothesis that a maternally inherited upstream factor induces *Dux* expression in the embryo. Recent work has shed light on this hypothesis and identified NELFA and DPPA2/DPPA4 proteins as activators of *Dux* transcription[26–28•]. *Dppa2*/*Dppa4* double knockout ES cells are devoid of 2CLCs but have no effect on pluripotency or self-renewal. When overexpressed, DPPA2/DPPA4 lead to a higher

proportion of 2CLCs, presumably through the binding onto the *Dux* locus and consequent activation of *Dux* transcription. Interestingly, DPPA2/DPPA4 do not bind other regulatory regions of 2C genes, indicating that their role in 2CLC induction is exerted primarily through the regulation of *Dux*. In agreement, overexpression of DUX rescues the 2CLC population in the *Dppa2/Dppa4* double KO cell line, but not viceversa[26,27]. DPPA2/DPPA4 can also be sumoylated by PIAS4. As a result, DPPA2/DPPA4 are no longer able to bind their chromatin targets, including *Dux*, and therefore SUMO can also regulate 2CLC fate indirectly, by inhibiting *Dux* expression through this pathway[22]. Originally, *Dppa2*/*Dppa4* were identified from a list of selected candidates in a screen searching for ZGA-like activation and 2CLC induction in ES cells[26,27]. More recently, NELFA was discovered as a key 2CLC factor following an unbiased approach based on the analysis of transcriptomes of mouse early embryos, aimed to identify genes expressed from the oocyte to the 2-cell stage exclusively[28•]. This led to a list of approximately 2,000 genes, which were reduced to 8 candidates only when overlapped with transcriptomes of 2CLCs. *Nelfa* was amongst these 8 candidates, but interestingly, neither *Dppa2* nor *Dppa4* were, presumably because their expression in the embryo does not decrease after the 2-cell stage. NELFA appears as a *bona fide* 2CLC marker, since NELFA$^{high}$ cells – that is, cells expressing high levels of NELFA- were extensively characterised and shown to possess known 2CLC features. NELFA is in fact expressed in a subpopulation of the ZSCAN4$^{+}$ intermediate population. Around 70% of ZSCAN4$^{+}$ cells are also NELFA$^{high}$. Interestingly, NELFA expression increases in 2CLCs and 90% of MERVL$^{+}$ cells are also NELFA$^{high}$. Overexpression of NELFA drives 2CLC reprogramming, similarly to DPPA2/DPPA4, since NELFA also binds directly to the *Dux* locus to activate its transcription.

**Beyond protein regulators**

Recently, aphidicolin was uncovered as a somehow unexpected positive regulator of 2CLCs[29]. Singe cell RNAseq and unbiased clustering analysis of Aphidicolin-treated ES cells uncovered a cell population with matching 2CLC transcriptome features[29]. ATR activation was shown to induce 2CLC reprogramming, but without activating the DNA Damage Response (DDR) pathway. This reprogramming path might be related to that triggered upon CAF-1 knockdown[11], since RNAseq data from Aphidicolin-induced 2CLCs is most similar to that of 2CLCs induced upon CAF-1 down-regulation. Interestingly, this work also reported a posttranscriptional regulation of DUX in response to ATR, which increases DUX protein levels without altering *Dux* transcription. The RNA-binding protein GRSF1 may be involved in this new level of regulation, since it can bind *Dux* mRNA and *Grsf1* KD reduces Aphidicolin-induced *Dux*

expression. However, the precise mechanism behind this observation remains to be fully characterised.

The factors discussed above are proteins with a positive or negative impact on 2CLC induction. However, noncoding RNAs regulating 2CLCs are starting to emerge. The miR-34a microRNA acts as a negative regulator of 2CLCs through a mechanism involving regulation of MERVL by the transcription factor GATA2, itself regulated by miR-34a[30]. More recently, miR-344 has been described to prevent 2CLC induction[31•]. miR-344 targets and silences ZMYM2, a protein that binds to and represses MERVL, consequently repressing LTR-driven transcription. Interestingly, DUX binds to and activates miR-344 expression, highlighting an additional pathway whereby DUX can regulate MERVL and 2C gene expression. Remarkably, this work shows for the first time that activation of MERVL alone is sufficient to increase the proportion of 2CLC population in culture. Direct MERVL activation does not change, however, the number of ZSCAN4$^+$ cells, suggesting that MERVL activation may generate 2CLCs bypassing the intermediate ZSCAN4$^+$ state. However, targeted transcriptional activation of endogenous MERVL leads to ~40% of cells that are MERVL$^+$/ZSCAN4$^-$ and only ~25% are double positive for MERVL$^+$/ZSCAN4$^+$. This suggests that complete reprogramming to a 2C-like state has around 30% efficiency amongst MERVL activating cells. Upon inducing 2CLC reprogramming by activating MERVL and sorting double positive MERVL$^+$/ZSCAN4$^+$ cells, the 2CLC population drastically decreases after one day of culture, but the MERVL$^+$/ZSCAN4$^-$ population remains longer, for at least 3 more days. This indicates that even if MERVL activation alone is enough to reprogram ES cells to 2CLCs, it is not sufficient to maintain 2CLCs stably in culture.

Overall, no study has yet reported a way to keep and renew 2CLCs in culture. 2CLCs are a relatively recent cell culture model, still not fully characterised, and they might need different growth factors and culture conditions compared to ES cells. Actually, recent findings indicate indeed that 2CLCs have a distinct metabolic state in comparison to ES cells. The glycolytic activity is reduced in 2CLCs and surprisingly, they do not seem to depend on mitochondrial respiration[28,32]. Although 2CLCs show a higher glucose uptake than ES cells[32], manipulation of glycolysis can induce a higher percentage of 2CLCs[28•], although not under all conditions tested[32]. Importantly, a screen of selected metabolites identified sodium L-lactate, D-ribose and sodium acetate as 2CLC inducers, with sodium acetate showing a particularly potent effect[32]. These findings can build towards future research lines to help develop a designed

‘environment’ for culturing 2CLCs, which takes into account their unique transcriptional and metabolic characteristics.

Since their discovery, many new features and regulators of 2CLC reprogramming have been identified. However, how these cells arise naturally and to which extent they resemble the 2-cell stage embryo has not been fully elucidated. Considering how important ES cells have been as model system to study pluripotency, reprogramming and differentiation, the discovery of a model that can resemble an earlier totipotent stage of the embryo promises to open new exciting research lines.

**Acknowledgments**

We thank T. Nakatani for critical reading of the manuscript. M.E. Torres-Padilla acknowledges funding by the Helmholtz Association and the German Research Council (CRC 1064). A.I was a recipient of a long-term EMBO fellowship.

**References**


1. Ishiuchi, T. & Torres-Padilla, M.-E. Towards an understanding of the regulatory mechanisms of totipotency. *Curr. Opin. Genet. Dev.* **23**, 512–518 (2013).

2. Wu, G. & Schöler, H. R. Lineage Segregation in the Totipotent Embryo. *Curr. Top. Dev. Biol.* **117**, 301–317 (2016).

3. Tarkowski, A. K. Experiments on the Development of Isolated Blastomeres of Mouse Eggs. *Nature* **184**, 1286–1287 (1959).

4. Togashi, M. Production of monozygotic twins by splitting of 2-cell stage embryos in mice. *Jpn J Anim Reprod* **33**, 51–57 (1987).

5. Sotomaru, Y., Kato, Y. & Tsunoda, Y. Production of Monozygotic Twins after Freezing and Thawing of Bisected Mouse Embryos. *Cryobiology* **37**, 139–145 (1998).

6. Rossant, J. & Tam, P. P. L. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713 (2009).

7. Shahbazi, M. N. & Zernicka-Goetz, M. Deconstructing and reconstructing the mouse and human early embryo. *Nat. Cell Biol.* **20**, 878–887 (2018).

8. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).

9. Peaston, A. E. *et al.* Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Dev. Cell* **7**, 597–606 (2004).

10. Genet, M. & Torres-Padilla, M.-E. Defining the molecular and cellular features of 2-cell-like cells: a reference guide. *Development* **under revision**.

11. Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).

12. Eckersley-Maslin, M. A. *et al.* MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).

13•• Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).

This paper, together with ref. 15, identifies the first mammalian pioneer factor for ZGA and its ability to regulate MERVL-driven transcription in embryos and 2CLCs.

14. Bošković, A. *et al.* Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* **28**, 1042–1047 (2014).

15•• De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).

This paper, together with ref. 13, identifies the first mammalian pioneer factor for ZGA and its ability to regulate MERVL-driven transcription in embryos and 2CLCs.

16. Whiddon, J. L., Langford, A. T., Wong, C.-J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).

17. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **51**, 947–951 (2019).

18. Guo, M. *et al.* Precise temporal regulation of Dux is important for embryo development. *Cell Res.* **29**, 956–959 (2019).

19. Jachowicz, J. W. *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).

20. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405.e19 (2018).

21. Cossec, J.-C. *et al.* SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757.e8 (2018).

22. Yan, Y.-L. *et al.* DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program. *PLoS Biol.* **17**, e3000324 (2019).

23. Higuchi, C., Yamamoto, M., Shin, S.-W., Miyamoto, K. & Matsumoto, K. Perturbation of maternal PIASy abundance disrupts zygotic genome activation and embryonic development via SUMOylation pathway. *Biol. Open* **8**, (2019).

24•. Rodriguez-Terrones, D. *et al.* A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).

This paper studies the 2CLC reprogramming process to reveal cell trajectories identifying an intermediate state represented by the Zscan4+ population and uncovers new epigenetic regulators of 2CLC reprogramming.

25. Fu, X., Wu, X., Djekidel, M. N. & Zhang, Y. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.* **21**, 835–844 (2019).

26. De Iaco, A., Coudray, A., Duc, J. & Trono, D. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep.* **20**, (2019).

27. Eckersley-Maslin, M. *et al.* Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).

28•. Hu, Z. *et al.* Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat. Cell Biol.* 1–12 (2020) doi:10.1038/s41556-019-0453-8.

This paper identifies NELFA as a maternal factor that induces 2CLCs. While

traditionally known for its role as regulator of transcriptional elongation as part of the NELF complex, the authors show that NELFA role in 2CLC reprogramming is independent of its role within the NELF complex.

29. Atashpaz, S. *et al.* ATR expands embryonic stem cell fate potential in response to replication stress. *eLife* **9**, (2020).

30. Choi, Y. J. *et al.* Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, (2017).

31•. Yang, F. *et al.* DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* **26**, 234-250.e7 (2020).

In this paper, the authors target MERVL activation directly, using a dCas9 system, showing that MERVL activation alone can induce the 2CLC state

32. Rodriguez-Terrones, D. *et al.* A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep.* **21**, e48354 (2020).