***Phil. Trans. R. Soc. B.* article template**

PHILOSOPHICAL TRANSACTIONS B

Phil. Trans. R. Soc. B.
doi:10.1098/not yet assigned

"On Transposons and Totipotency"

Maria-Elena Torres-Padilla*

Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München D-81377 München, Germany & Faculty of Biology, Ludwig-Maximilians Universität, München, Germany.

Keywords: reprogramming, 2-cell-like cells, pluripotency, MERVL, LINE-1, transposable elements

Summary

Our perception of the role of the previously considered 'selfish' or 'junk' DNA has been dramatically altered in the last twenty years or so. A large proportion of this non-coding portion of mammalian genomes is repetitive in nature, classified as either satellites or transposons. While repetitive elements can be termed selfish in terms of their amplification, such events have surely been co-opted by the host, suggesting by itself a likely altruistic function for the organism at the subject of such natural selection. Indeed numerous examples of transposons regulating the functional output of the host genome have been documented. Transposons provide a powerful framework for large-scale relatively rapid concerted regulatory activities with the ability to drive evolution. Mammalian totipotency has emerged as one key stage of development in which transposon-mediated regulation of gene expression has taken centre stage in the last few years. During this period large-scale (epigenetic) reprogramming must be accomplished in order to activate the host genome. In mice and men one particular element MERVL (and its counterpart HERVL) appears to have acquired roles as a key driving force in this process. Here I will discuss and interpret the current knowledge and its implications regarding the role of transposons, particularly of LINE-1s and ERVs, in the regulation of totipotency.

Main Text

Our genomes are vastly populated by remnants of viruses and other genomic elements, which have accumulated during mammalian evolution. In fact, in most common mammals, including mouse, cattle and humans, roughly half of the genome belongs to this category, commonly referred to as 'repeats' or repetitive elements [1]. Phylogenetically and from the point of view of their genome structure, repetitive elements can roughly be divided in transposable elements and satellite DNA. Satellite DNA is typically repeated in tandem at specific locations across the linear genome and can form clusters up to several kilobases. The satellites forming the centromere of mouse and humans belong to this class. They are critical to the mitotic process, as they are the place for kinetochore loading. Their repetitive nature is thought to be important for the establishment of heterochromatin and consequent physical compaction of these loci, which is therefore essential for chromosome segregation and faithful chromosome inheritance[2].

Transposable elements are classified into DNA transposons and retrotransposons. The most abundant in rodent and human genomes are the retrotransposons. Generally speaking, retrotransposons are transcriptionally silenced in most somatic cells. This is achieved through the acquisition of constitutive heterochromatin signatures driven primarily by the catalysis of H3K9 tri-methylation by several enzymes, including Setdb1, Suv39h1/h2 and the downstream recruitment of heterochromatin protein 1 and Suv420h1/h2 and by DNA methylation. In the germline, young transposons are primarily silenced through the piRNA pathway (reviewed in [3]). Silencing of retrotransposons is thought to be tightly regulated, and orchestrated by a myriad of KRAB zinc-finger proteins and their interactor, Kap1/Trim28 (reviewed in [4]).

*Author for correspondence (torres-padilla@helmholtz-muenchen.de).

At the earliest developmental stages immediately after fertilisation, in contrast, such heterochromatic signatures are thought to be largely absent or atypical [5]. In fact, the mouse zygote and 2-cell stage embryo actively transcribe many retrotransposons, and their transcriptional activity persists at least until the blastocyst stage, around the time of implantation[6-9]. The mechanisms regulating their transcriptional activation and repression are largely unknown. For example, the temporal dynamics of DNA de-methylation and re-methylation during pre-implantation development[10], are not sufficient to explain their changes in expression for the most part. Likewise, ChIP-seq analysis of H3K9me3 during pre-implantation development indicates that only a handful of retrotransposons from the LTR family are enriched in H3K9me3 in the cleavage-stage embryos[11], and thus cannot explain by its simple presence/absence, the transcriptional dynamics of most retrotransposons in the embryo.

While much remains to be done to understand how retrotransposons are regulated, recent work has started to identify key transcription factors involved in this process. These discoveries have emerged from the powerful combination of traditional *in vivo* (mouse embryo) and novel *in vitro* (cell culture) models. Here I will briefly discuss both these models, with emphasis on recent literature. However, before I go on and discuss these findings, I feel compelled to explain the definition of totipotency that I will use, since the word is used somehow loosely and to refer to different things in the literature. The implications of this definition will be important when discussing the use of cellular models and the assays employed to address whether such models are indeed truly totipotent.

Totipotency is the ability of a single cell to give rise to a full organism by itself [12, 13]. This contrasts to pluripotency, which in mammals refers to the ability of a cell to contribute and give rise to all three germ layers – including the germline - but cannot form a new organism by itself without supportive extra-embryonic tissues[14, 15]. When I refer to totipotency, an important distinction must be made between chimeric contribution versus self-potential of a single cell. Indeed, while blastomeres of the 4-cell embryo can contribute to all lineages of the conceptus when transplanted in chimera, including embryonic and extraembryonic lineages, individually they are not able to form a new being. Thus, while they are extremely plastic given their ability to contribute to all lineages of the embryo, they are not totipotent, and instead totipotency in the mouse is restricted to the zygote and each of the blastomeres of the 2-cell stage embryo[16, 17].

Changing cell fate

Work of Todd Macfarlan, Samuel Pfaff and colleagues in 2012 reported the existence of a subpopulation of cells in embryonic stem (ES) cell cultures, which highly resemble cells from the 2-cell stage embryo, based on their global transcriptional profile[18]. Remarkably, one of the main features of these '2-cell-like cells' (2CLC) is that they robustly transcribe elements from an LTR family of endogenous retroviruses, specifically MERVL (Murine Endogenous RetroVirus with Leucine tRNA primer). Indeed, work by Barbara Knowles in the 2000's had reported that oocytes and early mouse embryos from cleavage stages can use sequences from MERVL (primarily Mt2_MM and MERVL_int) as alternative promoters of host genes, identifying several chimeric transcripts containing MERVL and single copy genes sequences[6, 8]. Thus, it would seem that the transcriptional activation of a retrotransposon could have a potential, direct role in changing cellular fate in cell culture, and in regulating the transcriptional programme of 2-cell stage embryos.

The discovery of 2CLC was ground-breaking for several reasons. Firstly, it highlighted the general existence of additional heterogeneities in ES cell cultures in the mouse. This expanded our knowledge on the transient phenotypic features of ESC in culture, which already included the well-defined naïve pluripotent cells (e.g. 'Nanog-high' typically enriched in '2i' medium) as well as cells in a more intermediate pluripotency state (e.g. 'Nanog-low), but also cells which co-express epiblast and extraembryonic genes, similarly to morula stage blastomeres[19, 20]. The latter were isolated as expressing the endoderm marker Hex-1 and shown to contribute to trophectoderm and ICM derivatives when aggregated in chimera. While these cells were referred to totipotent, I would argue that they are rather bipotent, and not totipotent, based on their ability to contribute to the two first lineages of the mouse blastocyst but not to a full organism by themselves, under the framework of the totipotency definition that I put forward. Of note, this capacity to contribute to ICM and trophectoderm derivatives has also recently been referred to as 'expanded potential'[21].

Secondly, and most importantly, the identification of 2CLC provided the community with a conceptual and experimental platform to seek to establish a biochemically tractable model to understand the molecular features of totipotency and the biology of the early mouse embryo. This has generated an enormous amount of interest in the last 5 years, and work aiming to thoroughly characterise 2CLC and identify the regulators of their emergence, primarily through the identification of factors that can regulate MERVL transcription, has flourished.

It is precisely this endeavour, which led to the identification of the pioneer transcription factor Dux, as a key regulator of MERVL[22, 23]. Dux (DUX4 in humans) is a double homeodomain transcription factor, which is conserved in mammals and was originally identified as misregulated in facioscapulohumeral dystrophy (FSHD) in human biopsies[24]. The ectopic expression of Dux alone in mouse ESCs is sufficient to activate MERVL, along with a significant part of the '2C' transcriptional programme, which corresponds in effect to the zygotic genome activation (ZGA) programme. Dux regulates MERVL expression through direct binding of the Dux-recognition motif within the LTR. DUX4, its human counterpart leads to an equally robust increase in transcription of the human ERVL (HERVL) and of zygotic genome activation genes, which in humans reflects a '4C' stage [22, 23]. This is a remarkable observation, considering that the first homeodomain of Dux and DUX4 differs, and that the sequence of MERVL and HERVL is dissimilar and indicates that Dux and DUX4 co-evolved with their respective species-specific transposons, in an example of convergent

evolution. Indeed, when expressed in mouse ESCs, mouse DUX can activate ZGA-related retrotransposons but DUX4 cannot[25]. This has led to the suggestion that ancestral DUX proteins emerged to regulate embryonic transcription, but their capacity to regulate the process of ZGA was multiplied through their acquisition of the ability to regulate retrotransposons[25]. This implies that MERVL has been co-opted to regulate the key developmental process of ZGA in mammals, which is supported by earlier findings of chimeric transcripts in early embryos as discussed above. This begs the question of how '2C' genes became linked to MERVL elements and their LTRs through evolution.

Importantly however, while the overlap of genes regulated by Dux in mouse ESCs and the '2C' genes activated during mouse ZGA is significant, it is not complete[22, 23]. Similarly DUX4 drives the expression of ZGA genes in human ESCs, but not all human ZGA genes become activated upon DUX4 expression. These findings established that the Dux transcription factors, while key regulators of ZGA, are clearly not the only players. In fact, only about 25% of the accessible chromatin regions specific to 2CLC are bound by Dux[23], again suggesting that a large, additional part of the '2C' programme is regulated by other factors. It is therefore not surprising that a recent knock-out mouse model reported that Dux is not essential for development, since mice homozygous null for Dux are born[26, 27]. However, *Dux-/-* are born at submendelian proportions and, as expected, only about 25% of minor ZGA genes were affected in embryos lacking Dux, although this phenotype appears to vary dependent on the genetic background.

These findings are perfectly in line with the concept that retrotransposons have been co-opted in many instances as gene regulatory regions during evolution [28, 29]. Indeed, this hypothesis posits that MERVL amplification may have evolved to facilitate ZGA, perhaps by providing one single regulatory element across many genes to enable coordinated, fast and robust transcriptional activation at ZGA. However, the host-specific promoters remained intact and for the most part, functional, as seen by the differential promoter usage of some genes which are expressed during ZGA but also at later stages in development or in differentiated cells[8]. This notion would imply that, in case of mutation or deletion of the corresponding retrotransposon (or the transcription factor regulating its expression), the host promoter would still be able to be activated, most likely explaining the minor changes in ZGA observed in the Dux-/- embryos. Analysis of the promoter usage of genes located in proximity to MERVL in these mutants would provide important insights on the impact of retrotransposon co-option on gene regulation. Thus, a direct relationship between MERVL and ZGA remains to be formally investigated.

Accordingly, additional transcription factors for MERVL regulation and 2CLC emergence have been identified. Amongst them, Dppa2 and Dppa4 can activate '2C' genes when overexpressed in mouse ESCs [30-32]. Like Dux, Dppa2 and Dppa4 are highly expressed in mouse embryos during ZGA, and highly upregulated in 2CLC. Dppa2/4 mRNA levels remain high through pre-implantation development. Dppa2/4 is thought to be regulated by DNA methylation [30], and depleting Dppa2 or Dppa4 reduces significantly, but did not fully eliminate, 2CLC induction [30, 31]. Both Dppa2 and Dppa4 bind the Dux regulatory region and are necessary for Dux activation. However, whether Dppa2/4 act upstream and/or in parallel to Dux on MERVL activation is not yet fully clear, particularly considering that Dppa2/4display broad binding to most active promoters. GATA2, has also been shown to regulate MERVL expression in mouse iPS cells using a luciferase reporter, and when overexpressed in ESCs, it was found to bind MERVL by ChIP [33].

Additionally to searching for transcription factors regulating MERVL – and the '2C' programme –, several groups have sought to identify chromatin modifiers involved in their regulation. Most of them have been mainly identified through screening strategies using a 2C-reporter as readout, consisting of the MERVL LTR driving expression of a fluorescent protein. The quest for identifying chromatin modifiers regulating 2CLC stems from at least two standpoints. Firstly, chromatin regulators can enable cell fate conversions, acting as impediments or accelerators for reprogramming, but they can also stabilise cell fates. Secondly, work during the last years has established that the chromatin landscape of the totipotent cells of the early embryo differs significantly from chromatin in pluripotent and differentiated cells. Thus, identifying chromatin transitions in 2CLC has the potential to shed light in this very peculiar chromatin that characterises the totipotent cells of the early mammalian embryo.

The original report by Macfarlan and colleagues identified TSA as a 2CLC inducer, indicating that relaxing chromatin structure promotes 2CLC emergence. Indeed, subsequent MNase and ATACseq analyses revealed that 2CLC – and also 2-cell embryos – have a globally more open chromatin, compared to ESCs[23, 34-36]. This is also in line with the increased histone mobility observed in 2CLC and 2-cell stage embryos[37, 38]. Several chromatin modifiers have been identified as roadblocks for 2CLC reprogramming (Fig. 1). Our initial studies showed that CAF-1 (Chromatin Assembly Factor 1), the complex responsible for depositing core H3/H4 tetramers in a replication-coupled manner is a major regulator of such reprogramming[35]. Down-regulating CAF-1 in ESCs led to a substantial increase in 2CLC and this effect was exclusively dependent on the ability of CAF-1 to assemble chromatin. Later work screened ~1400 chromatin modifiers using an siRNA approach[39]. We identified 5 main molecular pathways as major roadblocks for 2CLC reprogramming. These included Replication and Chromatin assembly, in line with our previous work, but also new regulators such as the Tip60/Ep400 complex, the splicesosome, the non-canonical PRC1 (Polycomb Repressor Complex) complex PRC1.6 and various proteins related to Ubiquitin regulation. This study also suggested that several 'epigenetic' pathways can control 2CLC reprogramming, both directly and indirectly. A study published just a couple of months later using an shRNA screen also identified Rif1 (Replication Timing Regulatory Factor 1) as a regulator of 2CLC emergence[40]. An interesting observation from all of these findings is that induction of 2CLC emergence seems to be, at least partially, linked to the induction of some genes that are expressed in the germline. Components of the PRC1.6

complex had been identified as repressors of a germline transcriptional signature in ESCs[41]. Indeed, when Rybp, Pcgf6 and Max/Mga are down-regulated, genes such as Stra8, otherwise exclusively expressed during meiosis, become activated. Max, which is a dimerization partner of Myc, was well known as a germ cell repressor [42, 43]. Whether these specific changes pertaining to germline processes are important or necessary for acquiring a totipotent and/or a 2CLC identity remains to be determined. Notably, PRC1.6 binds directly to the Dux locus and is regulated by Sumoylation [44], and therefore the SUMO-pathway has emerged as a repressor of 2CLC fate. The regulation of 2CLC identity by SUMO occurs at least through two different mechanisms. The first one involves sumoylation of PRC1.6 components, which leads to the reduction of PRC1.6 occupancy at the Dux loci and eventual transcriptional activation of Dux[44]. The second one involves sumoylation of Dppa2 by the sumo ligase PIAS4, which renders Dppa2 inactive[32].

More recently, another study using a wider dCas9 screening strategy revealed essentially the same factors identified in previous screenings [45]. Dnmt1 was also identified as promoting 2CLC maintenance. In a separate work, temporal changes in MERVL expression were correlated with DNA demethylation, upon addition of vitamin C to the culture[46]. Interestingly, the transition to a MERVL-expressing state had been previously shown to be accompanied by a global genome-wide DNA hypomethylation, including the decrease of DNA methylation at imprinted genes[34]. While this was associated with low levels of DNMT proteins in these cells, in particular of Dnmt1, the results indicated that reprogramming into a MERVL+ state was not a consequence of changes in DNA methylation. Indeed, triple-knock-out (TKO) of Dnmt1/3a/3b in ESCs, which results in loss of DNA methylation, does not affect 2CLC emergence[34]. In this regard, the RNA-binding protein PSPC1 inhibits expression of several MERVK, MERVL and MaLR in mouse ESCs, presumably through physical interactions and recruitment of the Tet2 enzyme [47].

Cause or consequence?

In sum, the work above indicates that reprogramming towards 2CLC is regulated at multiple levels by different molecular pathways and chromatin regulators. It will be important to disentangle the direct and indirect regulators, and precise biochemical mechanisms and underlying regulatory actions. Concerning retrotransposon regulation, outstanding questions remain, including: i) what provides specificity to transposon regulation, ii) what other additional transcription factors can regulate MERVL expression and repression and, most importantly, iii) is retrotransposon expression sufficient to drive changes in cell fate, or is this just a consequence and/or a passenger?

Understanding the specificity in targeting each retrotransposon families and subfamilies will be crucial to understand how retrotransposons may help shape changes in cell fate. Such specificity is likely to be also cell type specific, and potentially achieved through KRAB-zinc-finger proteins[48]. Indeed, we know that depletion of the H3K9 methyltransferase Setdb1/ESET in ESCs results in derepression of class I (MLV) and class II (IAP and MusD) ERVs, but class III ERV MTA are unaffected [49, 50]. This contrasts to fibroblasts, in which IAP and MusD are not derepressed upon Setdb1 loss, but MLV are. It is therefore possible that the 'host' chromatin state and/or cofactors, including transcription factors, present in specific cell types aid in reaching specificity in retrotransposon regulation.

As discussed above, a key task would be to determine whether a 'simple' transcriptional activation of specific retrotransposons would be sufficient to drive cell fate changes. This question is now beginning to be addressed using different novel methodologies such as Crispr/Cas9-based genome editing methods and also TALE-mediated activation and CRISPRa strategies, whereby transcriptional activators or repressors are targeted directly to thousands of loci at their endogenous locations across the genome.

Is MERVL and its LTR-driven expression a driver for 2CLC reprogramming? Transcription from the MERVL LTR can also impact neighbouring genes by readthrough transcription and enhancer function. The fact that Dux can induce 2CLC emergence so efficiently suggests that transcription of MERVL may be sufficient for reprogramming[22, 23]. However because Dux can also bind additional single copy genes, this possibility remains to be demonstrated. Activation of MERVL using Crispr/SAM did result in the induction of the handful of 2C genes analysed (11 genes in total) [47]. This indicates that activation of MERVL alone can potentially regulate – at least in part- the 2CLC programme.

Giving pluripotency away.

The changes in cell fate that reprogramming ESCs to 2CLC entail also involve shutting down the pluripotency programme. The extent of this 'off-switch' is not yet very clear, but Macfarlan and colleagues documented that 2CLC lack detectable protein levels of OCT4[18]. In fact, analysis of other transcription factors (TF) of the pluripotency-gene regulatory network further demonstrated that this is not limited to OCT4, but that levels of SOX2, PRDM14, REX1 and AP2gamma proteins are also undetectable in 2CLC[39]. However, their mRNA levels are only slightly reduced or unchanged, indicating that shutting down the pluripotency network is regulated also at the translational level.

Importantly, analysis of an intermediate cellular state characterised by the expression of the TF Zscan4 but the absence of MERVL transcription, indicated that the pluripotency network is down-regulated before the transition into a 2CLC (MERVL+) state. These experiments led us to conclude that cells must exit pluripotency *before* committing to the 2CLC fate[39], potentially through translational regulation in Zscan4+ cells[51]. These results were recently confirmed by a later study[45]. In fact neither Oct4, Rex1 or Nanog depletion affects MERVL or Zscan4 expression and, accordingly, their down-regulation does not impact 2CLC emergence[39]. We also showed that it is the naïve ESCs that have the highest probability to reprogramme towards 2CLC [39]. At first sight, these conclusions would seem at odds with the

observation that ESCs kept in a 'ground', naïve state through the continuous presence of the '2i' inhibitors cycle less often into 2CLC[39]. However, culturing ESCs in such pharmacological inhibitors is expected to 'freeze' ESCs in a ground fate, and therefore preventing spontaneous changes of cellular heterogeneities in such cultures.

Studying the natural progression from ESCs towards 2CLC reprogramming using single cell expression analyses with 93 genes allowed us to establish the sequential changes in gene expression, including changes in pluripotency factors, that accompany 2CLC formation, thereby defining intermediate cellular states. Remarkably, computational modelling of single cell RNAseq confirmed these transitions on a genome-wide level, and identified additional sets of genes differentially expressed at each of the transitions[39]. This indicates that the intermediate stages that we described reflect concrete and binary changes in the transcriptome of cells transiting towards the 2CLC. While informative, these analyses are highly correlative, and do not really shed light into the actual molecular mechanisms that directly regulate changes in cell fate. Additional work, perhaps through the identification of specific TF and/or KRAB-ZFP will help us to better disentangle causal relationships and the role of retrotransposons in this process.

Can totipotency be reduced to a unique transcriptional state?

I believe that the answer is clearly no. Although phenotypically most of the molecular features that provide cell identity and function are a consequence of the genes transcribed, they are certainly not limited to that. For example, metabolic features and chromatin state, or extrinsic factors need to be taken into account. Indeed, hypoxia reduces the percentage of 2CLC in ESC cultures[18]. Notwithstanding, a cell's potency should eventually be determined by its ability to 'do something' in its natural environment, when transplanted back into the respective native context.

The gold standard for testing whether cells in culture are pluripotent consists of introducing them into morula or blastocyst stage embryos and investigating their potential to contribute to the embryo, and most importantly, to the germline [52]. These chimera assays are relatively straightforward for ESCs with full pluripotency, since they readily contribute to all the lineages of the embryo proper, especially when more than one cell is injected per chimera. In analogy, therefore, totipotent cells should be tested in vivo. However, these experiments pose a conceptual problem, if one considers that totipotent cells are those, which can give rise to a full organism without the need of carrier cells. Often however, totipotency is called positive when cells in culture are put back into late 8-cell stage embryos, and they are seen to occupy the trophectoderm layer of the blastocyst, and with more stringency, when cells are seen in the placenta of post-implantation embryos (Figure 2). In the field, these experiments are generally done with various levels of stringency criteria and most of them lack evidence of proliferative trophoblast derivatives in the placenta, and are restricted to conclusions on contribution to yolk sac derivatives, which in fact is an ICM derivative, not a trophectoderm derivative. In addition, they have raised several reproducibility issues, which may be due to genetic background, mouse strains, timing, and many others, or simply to lack of robustness. Most importantly, however in my view these assays test for bipotentiality to contribute to the two first lineages of the blastocyst, but not for totipotency.

I will argue therefore, that we currently do not have a correct assay to test for totipotency. It is also very improbable that one cell, for example a 2CLC, can fulfill the geometrical needs and constraints of the *in vivo* embryo. Their diameter is at least 10 times smaller and thus their volume is 1000 times smaller than the 2-cell stage blastomeres ! Other differences beyond cell size, for example in their metabolism (Rodriguez-Terrones et al., In Press) may also exist. However, aiming for strategies which can lead to the formation of blastuloids or blastocyst-like cyst structures will be a good starting point [53] (Figure 2). In this regard, the current strategies based on self-assembly of embryonic and extra-embryonic stem cells into embryo-like structures will greatly facilitate the development of these approaches[54, 55]. In addition, testing the capacity of a single cell to yield the 3 stem cell derivatives of the blastocysts in vitro, in cell culture, through the derivation of XEN, ESC and TS cells has also been proposed as an assay to demonstrate totipotency [15]. Nuclear transfer efficiency is also an objective and robust experimental approach, which provides a clear metric of cell plasticity, considering that we know for almost 50 years that differentiated cells are less likely to be fully reprogrammed upon cloning, compared to stem cells[56]. Likewise, ESCs donors exhibit a worse reprogramming efficiency upon nuclear transfer compared to totipotent mouse blastomeres [57]. It is therefore notable that 2CLC donors have a higher reprogrammability than ESCs [35].

Supporting developmental programmes

In all, work in the last years has highlighted that retrotransposons have the potential to support developmental programmes [4]. We are now also in the position to directly address whether retrotransposon expression is functionally important for developmental progression. For example, work in the Wysocka lab has used dCas9 and Crispr/Cas9 strategies to show that specific LTRs are in fact responsible for regulating full transcriptional induction of genes expressed in human ESCs [58]. Likewise, by manipulating the chromatin of LINE-1s at their endogenous loci, our own work has indicated that their transcriptional activation regulates chromatin accessibility in the early mouse embryo [59].

Importantly, recent work has documented clear new insertions of LINE-1 in neurons of human beings, indicating current somatic transposition [60] and new roles for LINE-1 in nucleolar organisation during development have emerged [61]. It is therefore very exciting and of utmost importance to fully dive into the consequences and mechanisms of retrotransposon regulation, function, and impact for developmental programmes, which in the longer term also dramatically impact the evolutionary path.

Acknowledgments

Work in the Torres-Padilla lab is funded by the Helmholtz Association, the German Research Council (CRC 1064) and H202 Marie-Curie Actions ITN EpiSystem and ITN ChromDesign. METP acknowledges past and current members of her lab for stimulating discussions and for building on the concepts laid down in this opinion over the last years and thanks A. Burton for critical reading of the manuscript, T. Nakatani for comments on the figures, and J. Lin for providing the blastocyst-like image in Figure 2.

References

1. Bourque, G., *Transposable elements in gene regulation and in the evolution of vertebrate genomes.* Curr Opin Genet Dev, 2009. **19**(6): p. 607-12.
2. Fukagawa, T. and W.C. Earnshaw, *The centromere: chromatin foundation for the kinetochore machinery.* Dev Cell, 2014. **30**(5): p. 496-508.
3. Chuma, S. and T. Nakano, *piRNA and spermatogenesis in mice.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1609): p. 20110338.
4. Friedli, M. and D. Trono, *The developmental control of transposable elements and the evolution of higher species.* Annu Rev Cell Dev Biol, 2015. **31**: p. 429-51.
5. Burton, A. and M.E. Torres-Padilla, *Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo.* Brief Funct Genomics, 2010.
6. Evsikov, A.V., et al., *Systems biology of the 2-cell mouse embryo.* Cytogenet Genome Res, 2004. **105**(2-4): p. 240-50.
7. Fadloun, A., et al., *Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA.* Nat Struct Mol Biol, 2013. **20**(3): p. 332-8.
8. Peaston, A.E., et al., *Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos.* Dev Cell, 2004. **7**(4): p. 597-606.
9. Rodriguez-Terrones, D. and M.E. Torres-Padilla, *Nimble and Ready to Mingle: Transposon Outbursts of Early Development.* Trends Genet, 2018. **34**(10): p. 806-820.
10. Smith, Z.D., et al., *A unique regulatory phase of DNA methylation in the early mammalian embryo.* Nature, 2012. **484**(7394): p. 339-44.
11. Wang, C., et al., *Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development.* Nat Cell Biol, 2018. **20**(5): p. 620-631.
12. Ishiuchi, T. and M.E. Torres-Padilla, *Towards an understanding of the regulatory mechanisms of totipotency.* Current opinion in genetics & development, 2013. **23**(5): p. 512-8.
13. Eckersley-Maslin, M.A., C. Alda-Catalinas, and W. Reik, *Dynamics of the epigenetic landscape during the maternal-to-zygotic transition.* Nat Rev Mol Cell Biol, 2018. **19**(7): p. 436-450.
14. Martinez Arias, A., J. Nichols, and C. Schroter, *A molecular basis for developmental plasticity in early mammalian embryos.* Development, 2013. **140**(17): p. 3499-510.
15. Baker, C.L. and M.F. Pera, *Capturing Totipotent Stem Cells.* Cell Stem Cell, 2018. **22**(1): p. 25-34.
16. Casser, E., et al., *Totipotency segregates between the sister blastomeres of two-cell stage mouse embryos.* Sci Rep, 2017. **7**(1): p. 8299.
17. Tarkowski, A.K., *Experiments on the development of isolated blastomers of mouse eggs.* Nature, 1959. **184**: p. 1286-7.
18. Macfarlan, T.S., et al., *Embryonic stem cell potency fluctuates with endogenous retrovirus activity.* Nature, 2012. **487**(7405): p. 57-63.
19. Festuccia, N., et al., *The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states.* Current opinion in genetics & development, 2013. **23**(5): p. 504-11.
20. Morgani, S.M., et al., *Totipotent embryonic stem cells arise in ground-state culture conditions.* Cell reports, 2013. **3**(6): p. 1945-57.
21. Yang, Y., et al., *Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency.* Cell, 2017. **169**(2): p. 243-257 e25.

22. De Iaco, A., et al., *DUX-family transcription factors regulate zygotic genome activation in placental mammals.* Nature genetics, 2017. **49**(6): p. 941-945.
23. Hendrickson, P.G., et al., *Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons.* Nature genetics, 2017. **49**(6): p. 925-934.
24. Snider, L., et al., *Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene.* PLoS Genet, 2010. **6**(10): p. e1001181.
25. Whiddon, J.L., et al., *Conservation and innovation in the DUX4-family gene network.* Nat Genet, 2017. **49**(6): p. 935-940.
26. Chen, Z. and Y. Zhang, *Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development.* Nat Genet, 2019. **51**(6): p. 947-951.
27. De Iaco, A., et al., *DUX is a non-essential synchronizer of zygotic genome activation.* bioRxiv, 2019.
28. Castro-Diaz, N., M. Friedli, and D. Trono, *Drawing a fine line on endogenous retroelement activity.* Mob Genet Elements, 2015. **5**(1): p. 1-6.
29. Chuong, E.B., N.C. Elde, and C. Feschotte, *Regulatory activities of transposable elements: from conflicts to benefits.* Nat Rev Genet, 2017. **18**(2): p. 71-86.
30. Eckersley-Maslin, M., et al., *Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program.* Genes Dev, 2019. **33**(3-4): p. 194-208.
31. De Iaco, A., et al., *DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells.* EMBO Rep, 2019. **20**(5).
32. Yan, Y.L., et al., *DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program.* PLoS Biol, 2019. **17**(6): p. e3000324.
33. Choi, Y.J., et al., *Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells.* Science, 2017. **355**(6325).
34. Eckersley-Maslin, M.A., et al., *MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs.* Cell reports, 2016. **17**(1): p. 179-92.
35. Ishiuchi, T., et al., *Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly.* Nature structural & molecular biology, 2015. **22**(9): p. 662-71.
36. Wu, J., et al., *The landscape of accessible chromatin in mammalian preimplantation embryos.* Nature, 2016. **534**(7609): p. 652-7.
37. Boskovic, A., et al., *Higher chromatin mobility supports totipotency and precedes pluripotency in vivo.* Genes & development, 2014. **28**(10): p. 1042-7.
38. Ooga, M., et al., *Analysis of chromatin structure in mouse preimplantation embryos by fluorescent recovery after photobleaching.* Epigenetics, 2016. **11**(1): p. 85-94.
39. Rodriguez-Terrones, D., et al., *A molecular roadmap for the emergence of early-embryonic-like cells in culture.* Nat Genet, 2018. **50**(1): p. 106-119.
40. Li, P., et al., *Rif1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation.* Nucleic Acids Res, 2017. **45**(22): p. 12723-12738.
41. Hisada, K., et al., *RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells.* Molecular and cellular biology, 2012. **32**(6): p. 1139-49.
42. Tatsumi, D., et al., *DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells.* PLoS One, 2018. **13**(11): p. e0205969.
43. Maeda, I., et al., *Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells.* Nat Commun, 2013. **4**: p. 1754.
44. Cossec, J.C., et al., *SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States.* Cell Stem Cell, 2018. **23**(5): p. 742-757 e8.
45. Fu, X., et al., *Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells.* Nat Cell Biol, 2019. **21**(7): p. 835-844.
46. Walter, M., et al., *An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells.* Elife, 2016. **5**.
47. Guallar, D., et al., *RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells.* Nat Genet, 2018. **50**(3): p. 443-451.

48. Ecco, G., et al., *Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues.* Dev Cell, 2016. **36**(6): p. 611-23.
49. Matsui, T., et al., *Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET.* Nature, 2010. **464**(7290): p. 927-31.
50. Karimi, M.M., et al., *DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs.* Cell Stem Cell, 2011. **8**(6): p. 676-87.
51. Hung, S.S., et al., *Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-positive state of embryonic stem cells.* DNA Res, 2013. **20**(4): p. 391-402.
52. Mascetti, V.L. and R.A. Pedersen, *Contributions of Mammalian Chimeras to Pluripotent Stem Cell Research.* Cell Stem Cell, 2016. **19**(2): p. 163-175.
53. Rivron, N.C., et al., *Blastocyst-like structures generated solely from stem cells.* Nature, 2018. **557**(7703): p. 106-111.
54. Sozen, B., et al., *Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures.* Nat Cell Biol, 2018. **20**(8): p. 979-989.
55. Beccari, L., et al., *Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids.* Nature, 2018. **562**(7726): p. 272-276.
56. Gurdon, J.B., *The generation of diversity and pattern in animal development.* Cell, 1992. **68**(2): p. 185-99.
57. McGrath, J. and D. Solter, *Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro.* Science, 1984. **226**(4680): p. 1317-9.
58. Fuentes, D.R., T. Swigut, and J. Wysocka, *Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation.* Elife, 2018. **7**.
59. Jachowicz, J.W., et al., *LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo.* Nat Genet, 2017. **49**(10): p. 1502-1510.
60. Sanchez-Luque, F.J., et al., *LINE-1 Evasion of Epigenetic Repression in Humans.* Mol Cell, 2019. **75**(3): p. 590-604 e12.
61. Percharde, M., et al., *A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity.* Cell, 2018. **174**(2): p. 391-405 e19.

Figure Legends

Figure 1. Summary of the pathways shown so far to induce 2CLC in ESC cultures.
Ectopic expression of the transcription factors Dux and Dppa2/4 as well as culturing in hypoxia conditions or with TSA, increase the 2CLC population and therefore are positive regulators (green arrow). Down-regulation of CAF-1, specific chromatin modifiers and components of the splicesosome also induces 2CLC, and therefore these are negative regulators (red arrow). 2CLC emergence involves a transient DNA hypomethylation state, and a Zscan4-positive intermediate state. Global features of 2CLC are listed on the right.

Figure 2. Experimental approaches to investigate potential totipotency.
In A and B, typical chimera assays are shown. In A, incorporation into 8-cell stage pre-implantation mouse embryos is done by aggregation, typically referred to as 'morula aggregation'. In this assay, contribution to lineage is based on 3D position, but should be complemented with co-immunostaining to establish at least partial molecular identity. In B, incorporation is achieved through microinjection of cells into the blastocoel of early blastocysts, followed by implantation and analysis of the conceptus, typically at embryonic day (E) 9.5. In this assay, contribution is based on expression of a fluorescent reporter in the placenta. These are difficult experiments, often hindered by the fact that the placenta is highly autofluorescent, and often it is not straightforward to distinguish between placenta and the yolk sac, which is an ICM

derivative. These analyses should be accompanied by a stringent analysis through sections and molecular analysis of markers from the trophoblastic derivates of the placenta.

In C, cell culture strategies are shown. In C, as suggested by Baker and Pera (Ref 13), the ability of a single cell to give rise to stem cells from the three lineages of the mature blastocyst is depicted. XEN cells, primitive endoderm-derived stem cells; TS cells, trophoblast stem cells, derived from the trophectoderm. Molecular analysis of each of these cell types for the known relevant markers should be performed. D depicts a potential design to promote self-aggregation of 2CLC, and derivation of cyst-containing structures also referred to as blastocyst-like. As in C, a molecular analysis and exploration of lineage markers should be performed.

In E and F, nuclear transfer (NT) strategies are shown. E depicts the rationale behind using somatic cell nuclear transfer as an assay to test cellular plasticity, based on the observations that nuclei derived from early embryos show a highest success in generating embryos and pups upon cloning, as opposed to pluripotent stem cells. Accordingly, 2CLC nuclei show a higher success in producing clone embryos upon NT (Ref. 31). In F, the schematic of a NT experiment, including the potential outcomes and implications, is shown.

Figure 1.

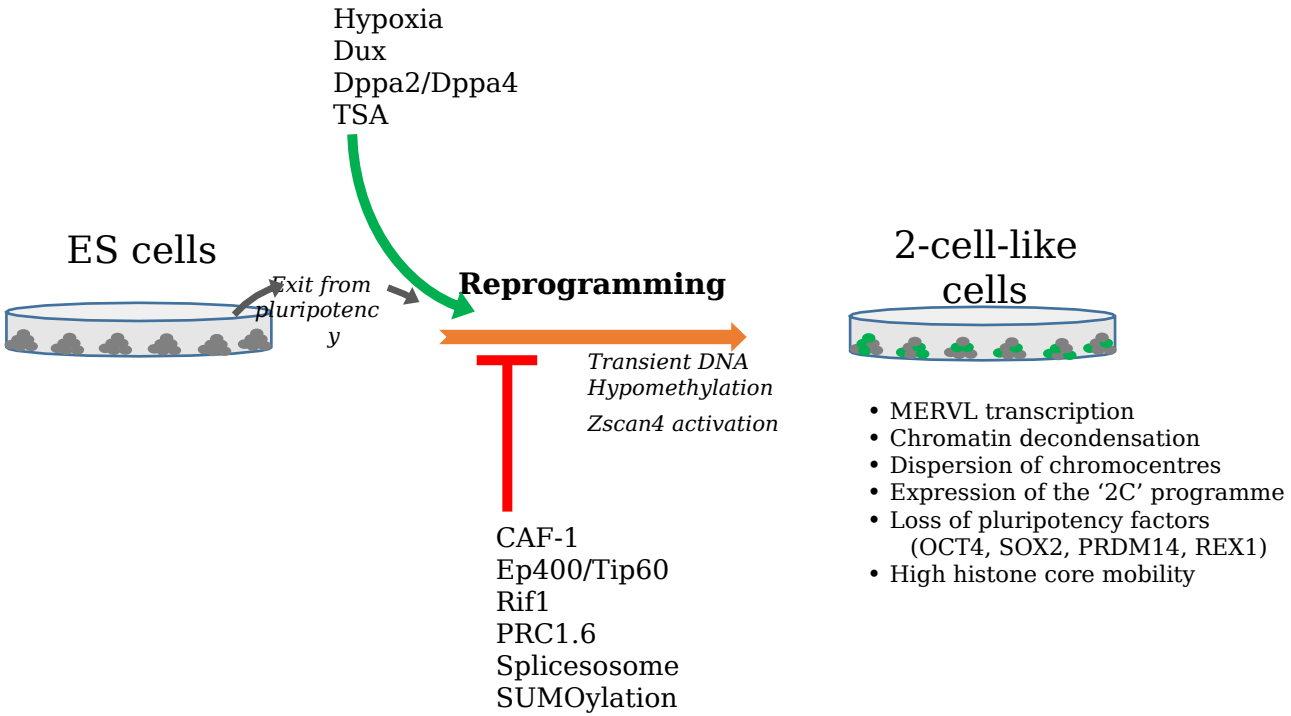
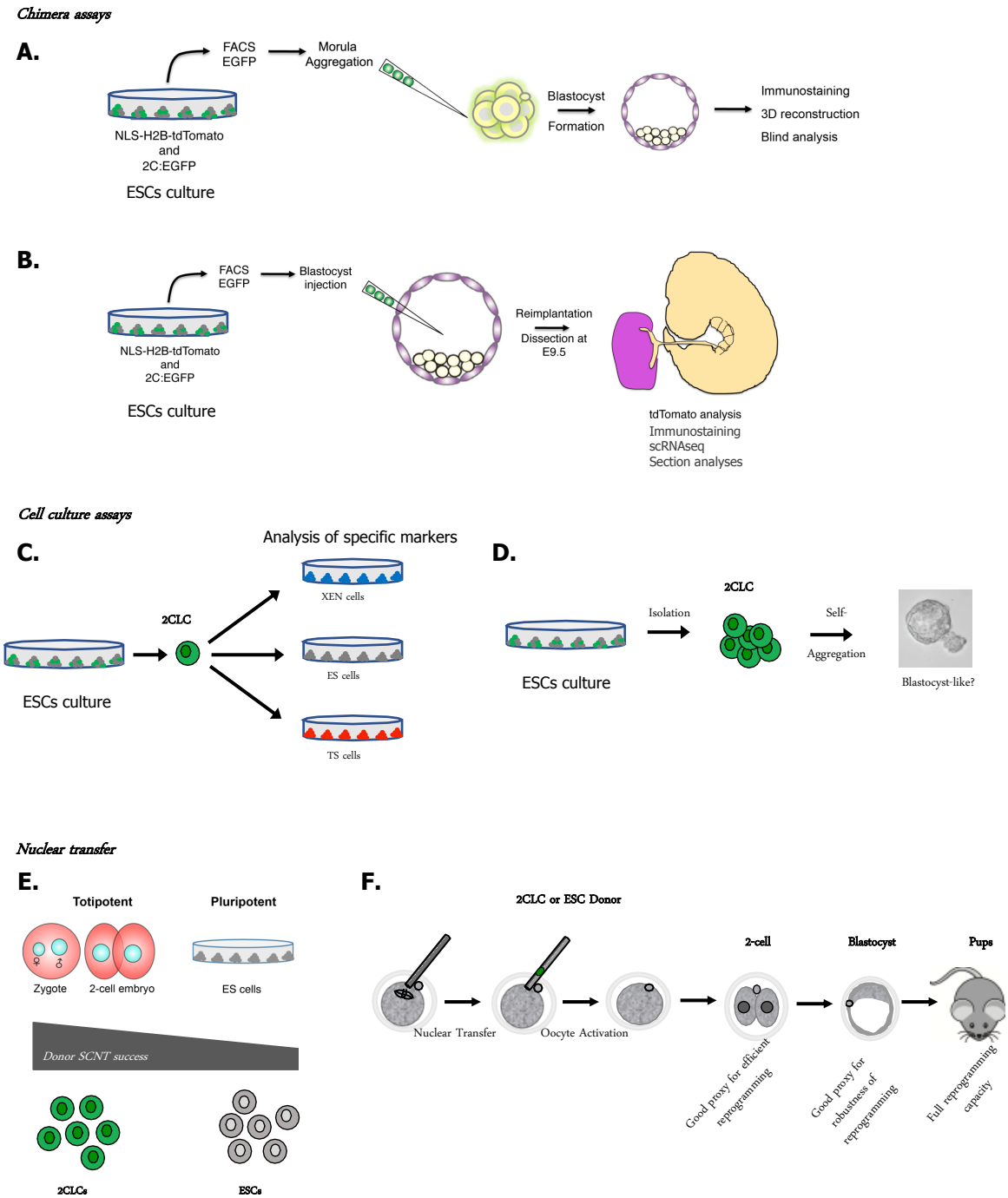

Hypoxia
Dux
Dppa2/Dppa4
TSA
ES cells
Exit from pluripotenc y
Reprogramming
Transient DNA Hypomethylation
Zscan4 activation
2-cell-like cells
• MERVL transcription
• Chromatin decondensation
• Dispersion of chromocentres
• Expression of the '2C' programme
• Loss of pluripotency factors (OCT4, SOX2, PRDM14, REX1)
• High histone core mobility
CAF-1
Ep400/Tip60
Rif1
PRC1.6
Splicesosome
SUMOylation

Figure 2.


Chimera assays
A.
FACS
EGFP
Morula
Aggregation
NLS-H2B-tdTomato
and
2C:EGFP
ESCs culture
Blastocyst
Formation
Immunostaining
3D reconstruction
Blind analysis
B.
FACS
EGFP
Blastocyst
injection
NLS-H2B-tdTomato
and
2C:EGFP
ESCs culture
Reimplantation
Dissection at
E9.5
tdTomato analysis
Immunostaining
scRNAseq
Section analyses
Cell culture assays
C.
Analysis of specific markers
2CLC
XEN cells
ES cells
TS cells
ESCs culture
D.
2CLC
Isolation
Self-
Aggregation
ESCs culture
Blastocyst-like?
Nuclear transfer
E.
Totipotent
Pluripotent
Zygote
2-cell embryo
ES cells
Donor SCNT success
2CLCs
ESCs
F.
2CLC or ESC Donor
2-cell
Blastocyst
Pups
Nuclear Transfer
Oocyte Activation
Good proxy for efficient reprogramming
Good proxy for robustness of reprogramming
Full reprogramming capacity