TGFβ signaling hyperactivation-induced tumorigenicity during the derivation of neural progenitors from mouse ESCs

Xianfa Yang1,2,†, Ran Wang1,†, Xiongjun Wang3,†, Guoqing Cai4, Yun Qian1, Su Feng4, Fangzhi Tan2, Kun Chen3, Ke Tang3, Xingxu Huang2, Naihe Jing1,2,*, and Yunbo Qiao3, *

1 State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Shanghai 200031, China
2 School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
3 Precise Genome Engineering Center, School of Life Sciences, Guangzhou University, Guangzhou 510006, China
† These authors contributed equally to this work.
* Correspondence to: Naihe Jing, E-mail: njing@sibcb.ac.cn; Yunbo Qiao, E-mail: ybqiao@gzhu.edu.cn

Clinical therapies of pluripotent stem cells (PSCs)-based transplantation have been hindered by frequent development of teratomas or tumors in animal models and clinical patients. Therefore, clarifying the mechanism of carcinogenesis in stem cell therapy is of great importance for reducing the risk of tumorigenicity. Here we differentiate Oct4-GFP mouse embryonic stem cells (mESCs) into neural progenitor cells (NPCs) and find that a minority of Oct4+ cells are continuously sustained at Oct4+ state. These cells can be enriched and proliferated in a standard ESC medium. Interestingly, the differentiation potential of these enriched cells is tightly restricted with much higher tumorigenic activity, which are thus defined as differentiation-resistant ESCs (DR-ESCs). Transcriptomic and epigenomic analyses show that DR-ESCs are characterized by primordial germ cell-like gene signatures (Dazl, Rec8, Stra8, Blimp1, etc.) and specific epigenetic patterns distinct from mESCs. Moreover, the DR-ESCs possess germ cell potential to generate Sycp3+ haploid cells and are able to reside in sperm-free spermaduct induced by busulfan. Finally, we find that TGFβ signaling is overactivated in DR-ESCs, and inhibition of TGFβ signaling eliminates the tumorigenicity of mESC-derived NPCs by inducing the full differentiation of DR-ESCs. These data demonstrate that these TGFβ-hyperactivated germ cell-like DR-ESCs are the main contributor for the tumorigenicity of ESCs-derived target cell therapy and that inhibition of TGFβ signaling in ESC-derived NPC transplantation could drastically reduce the risk of tumor development.

Keywords: embryonic stem cells, differentiation-resistant ESCs, tumorigenicity, germ cell, TGFβ signaling

Introduction
Embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), and other kinds of pluripotent stem cells (PSCs) provide a promising potential for cell transplantation-based regenerative therapies and disease modeling (Wu and Hochedlinger, 2011). PSC-derived neural progenitor cells (NPCs) or regionalized neurons are widely used for cell replacement therapy in neurological diseases in animal models or clinical trials (Lindvall and Kokaia, 2006; Barker et al., 2013; Yue et al., 2015; Kikuchi et al., 2017). The premise for neurodegenerative disease therapy is that the transplanted cells could provide a functional substitute for the specific subtypes of lost neurons (Trounson and McDonald, 2015).

Although great advances have been made from PSC derivatives therapies, but the tumorigenic or malignant transformation potential of transplanted cells elicited by the properties of self-renewal and pluripotency largely hinders their large-scale clinical applications (Lindvall et al., 2004; Tabar and Studer, 2014). The injection of ESCs or ESC-derived precursors in rodents frequently leads to tumor development (Bjorklund et al., 2002; Amariglio et al., 2009; Lee et al., 2013). The risks of PSC tumorigenicity have been highlighted in mice, rat, primate models, and clinical patients (Amariglio et al., 2009; Kriks et al., 2011; Doi et al., 2012; Cui et al., 2013; Tateno et al., 2015). Therefore, reduction of tumorigenic risk is a fundamental challenge for...
PSC-based therapy, and identification of the origin of tumor cells seems to be the first and vital step. It has been assumed that tumorigenicity can be resulted from the malignant transformation of differentiated PSCs or teratocarcinomas formation of residual undifferentiated PSCs (Ben-David and Benvenisty, 2011). Intriguingly, some injected PSCs are prone to generate primitive and undifferentiated tumors, leading to targeting tissue overgrowth (Roy et al., 2006; Shih et al., 2007). It has been revealed that PSCs share multiple cellular and molecular phenotypes with transplantation-derived tumors, including rapid proliferation rate, genomic instability, and similarities in gene expression patterns (Amit et al., 2000; Spanger et al., 2003; Baker et al., 2007). However, the characteristics of tumor origin cells, the knowledge of which can help to eliminate the tumor potential, have not been well clarified.

In the present study, we use Oct4-GFP to trace the Oct4+ undifferentiated cells during mESC differentiation into NPCs, and we observe a minority of Oct4+ cells remained in NPCs. These differentiation-resistant Oct4+ cells (defined as DR-ESCs) were enriched and proliferated for further characterization. Interestingly, the differentiation potential of these cells is restricted in dishes and prone to form teratomas with much higher tumorigenic activity, and the transcriptomic and epigenomic patterns of DR-ESCs are overall similar to mESCs while specially characterized by primordial germ cell (PGC) gene signatures and hyperactivation of TGFβ signaling. Inhibition of TGFβ signaling induces the full differentiation of DR-ESCs and obliterates the tumorigenicity of mESC-derived NPCs.

**Results**

**Identification of a rare differentiation-resistant population during mESC differentiation**

ESC-derived NPCs or neurons are invaluable sources for regenerative medicine, while the tumorigenic risk of cell transplantation hinders the fulfillment of their clinical potentials. Previously, we noticed that a minority of Oct4+ cells still existed even in terminal neural-differentiated ESCs (Qiao et al., 2015a, b). To trace these Oct4+ cells among terminally differentiated NPCs or neurons, we subjected Oct4-GFP mESCs for neural differentiation as a monolayer. To avoid the influence of the heterogeneity of ESCs with relatively low Oct4 expression, we sorted GFP+ cells with relatively higher fluorescence (Oct4-GFP-H) for subsequent neural differentiation (Figure 1A), and these cells showed relative homogeneous expression of Oct4 and Nanog, the signals of which were colocalized well with Oct4-GFP (Supplementary Figure S1A). Along with differentiation, the expression of neural related genes was upregulated accompanied by the downregulation of pluripotency gene expression (Supplementary Figure S1B) and the percentage of Oct4+ cells (Figure 1B). When over 85% of cells differentiated into Pax6+ NPCs or TuJ1+ neurons, there were still a minority of Oct4-GFP+ cells (<1%) remained no matter of differentiation as a monolayer or embryonic bodies (EBs) (Figure 1C and Supplementary Figure S1C). Then, these GFP+ cells were enriched and proliferated with a perfect colony morphology in a standard mESC culture condition, whereas no colonies were obtained from GFP- cells (Figure 1D). Considering the origin of this novel established cell line, it was defined as differentiation-resistant ESCs (DR-ESCs). Moreover, Oct4+ and Nanog+ cells were also observed in differentiated E14 mESCs (Supplementary Figure S1D), and this kind of DR-ESCs was also successfully established from differentiated E14 cells (Supplementary Figure S1E). Then, we compared the expression of pluripotency and lineage-specific markers and found that DR-ESCs and GFP+ cells sorted from Day 8 differentiated cells exhibited a similar expression pattern with undifferentiated ESCs with a relatively high Stella expression (Figure 1E).

In previous cell therapy cases, the ESC derivatives that were thought to be terminally differentiated were transplanted into the targeting sites without clearance of these minimal Oct4+ cells (Amariglio et al., 2009; Barker et al., 2013; Yue et al., 2015). To simplify the model to examine the tumorigenicity of our obtained NPCs, we subcutaneously injected non-sorted Day 8 cells, Day 8-sorted GFP+ cells, and GFP– cells into the immunodeficient nude mice for tumor growth assays. Interestingly, the tumorigenic rate and the tumor weight were higher for GFP+ cells than the other two groups; no tumors were detected in the GFP– NPC group (Figure 1F). Thus, we identify a rare population of Oct4+ cells in neural-differentiated ESCs, which can be maintained as a cell line and exhibit a malignant tumorigenic capacity; clearance of these rare Oct4+ cells may remarkably reduce the tumorigenic risk during cell transplantation therapy.

**Differentiation potential of DR-ESCs is restricted**

Given that DR-ESCs were isolated from Day 8 Oct4-GFP+ cells and displayed an ESC-like colony morphology, we asked whether DR-ESCs had similar differentiation potential as ESCs. To answer this question, Oct4-GFP-H and DR-ESCs were subjected to neural differentiation. Compared to the minor percentage of GFP+ cells in differentiated Oct4-GFP-H ESCs, there were still about 60% of Oct4+-Nanog+ cells remained in differentiated DR-ESCs at Day 8 or 12 (Figure 2A–C and Supplementary Figure S2A–C). Consistently, this phenomenon was also observed in E14-derived DR-ESCs (Supplementary Figure S2D and E). We also withdrawn Lif from culture medium for 6 days, about 50% of Oct4-GFP+ cells were still sustained with a colony morphology (Figure 2D), suggesting that Lif signaling is not essential for DR-ESC maintenance. Quantitative real-time PCR (Q-PCR) analysis revealed that Day 8 DR-ESCs possessed high transcriptional levels of pluripotency genes and low expression levels of germ layer markers (Figure 2E and Supplementary Figure S2F), which was similar to original mESCs and DR-ESCs. It demonstrates that distinct from normal ESC lines, the differentiation potential of DR-ESCs is highly restricted.

To compare the developmental potency in vivo, we injected Oct4-GFP-H ESCs and DR-ESCs into nude mice for teratoma analysis. Apparently, the tumorigenic capacity of DR-ESCs was much higher with the formation of bigger tumors than the control group (Figure 2F). Consistently, obvious GFP signal was observed in DR-ESC-generated teratomas (Figure 2G). Although three germ layers (ectoderm, mesoderm, and endoderm) could
be derived from both injected Oct4-GFP-H ESCs and DR-ESCs (Figure 2H), a much larger proportion of undifferentiated areas with Oct4-GFP+ cells were identified in DR-ESC-derived tumors, which might be a kind of teratocarcinoma composed of a mixture of differentiated tissues of three germ layers and embryonic carcinoma cells (Figure 2I and J). As expected, the tumorigenicity of differentiated DR-ESCs was much severer than normally neural-differentiated ESCs (Figure 2K). Collectively, the differentiation capacity of DR-ESCs is strongly restricted in vitro and in vivo.

**DR-ESCs exhibit PGC-like gene signatures**

To answer why the differentiation and tumorigenic potential of DR-ESCs were different from normal mESCs, we performed...
Figure 2 Differentiation potential of DR-ESCs is restricted. (A) Oct4-GFP-H and DR-ESCs were differentiated in N2B27 as a monolayer for 8 days. The bright fields and GFP signals were captured. (B) Two groups of cells were subjected to immunostaining to examine the colocalization between Oct4-GFP and Nanog. (C) The percentages of Oct4-GFP+ cells in Day 8 differentiated cells were determined. $P < 0.0001$ (Student’s t-test). (D) Oct4-GFP-H and DR-ESCs were cultured in standard condition without Lif for 6 days. The morphologies and percentages of Oct4-GFP+ cells were examined. (E) The relative expression of pluripotency genes and lineage markers was determined by Q-PCR analysis in Oct4-GFP-H, DR-ESCs, and their Day 8 differentiated cells. (F) Oct4-GFP-H and DR-ESCs (2 × 10^5 cells for each group) were injected into nude mice for tumor growth for 30 days. The tumor weight was calculated (n = 5). $P = 0.0455$ (Student’s t-test). (G) GFP signals were captured in representative tumors derived from Oct4-GFP-H or DR-ESCs. (H) H&E staining of tumor sections derived from Oct4-GFP-H or DR-ESCs. The images show representative ectoderm, mesoderm, and endoderm tissues in tumors. (I) Oct4-GFP signals were captured in H&E stained sections, showing a representative undifferentiated field in DR-ESC group. (J) Statistical analysis of the number of representative ectoderm, mesoderm, and endoderm fields as well as representative undifferentiated (undiff.) fields (>0.5 mm^2). (K) Tumorigenic analysis of Day 8 neural-differentiated Oct4-GFP-H and DR-ESCs (2 × 10^5 cells for each group) in nude mice for 30 days. $P = 0.0056$ (t-test).
qualified RNA-seq analysis to compare their transcriptomes (Supplementary Figure S3A and B). Clustering analysis revealed that the expression pattern of DR-ESCs was highly similar with that of Oct4-GFP mESCs, but quite different from that of NPCs with high expression levels of neurodevelopmental genes (Figure 3A). We further identified 184 DR-ESC-specific genes, which were mainly related to embryonic and germ cell development. KEGG pathway analysis demonstrated that these genes were associated with carcinogenesis, such as p53 and TGFβ signaling pathways (Figure 3B). Among the top DR-ESC-specific genes, 15 germ cell signature genes including Stra8, Dazl, Blimp1, and Prdm14 were specifically expressed in DR-ESCs (Figure 3C), and the specific expression of these mentioned genes was verified by Q-PCR analysis (Figure 3D and Supplementary Figure S3C). Then we compared the transcriptome of normal ESCs and DR-ESCs with haploid spermatid-like cells (SLCs), E12.5 male PGCs, and spermatids (NCBI GEO: GSE71478 and GSE16925) (Zhao et al., 2009). Interestingly, the cluster dendrogram showed that DR-ESCs were clustered closer to E12.5 PGCs (Figure 3E), and this observation was further

Figure 3 DR-ESCs exhibit primordial germ cell-like gene signatures. (A) Heat mapping analysis of transcriptomes in Oct4-GFP, Oct4-GFP-H, DE-ESCs, and NPCs (GFP−). (B) DEG analysis for DR-ESC-specific genes. Gene ontology (GO) and KEGG pathway analyses were also conducted. (C) Heat map shows the expression pattern of representative DR-ESC-specific genes in four indicated cell types. Red-labeled genes were germ cell markers or related to germ cell development. (D) Q-PCR validation of Stra8, Dazl, Prdm14, and Blimp1 in indicated cells. (E) Cluster dendrogram analysis of Oct4-GFP, Oct4-GFP-H, DE-ESCs, and NPCs with spermatids, E12.5 PGCs, SLCs, and ESCs (Zhao et al., 2009). (F) PCA analysis of the relationship between the above cell types in E. (G) Network analysis of hub genes in three indicated groups of cells according to their transcriptomes from RNA-seq data.
confirmed by PCA analysis (Figure 3F), suggesting that DR-ESCs were similar with PGCs in expression signatures. We also analyzed the gene-network according to our RNA-seq data in three groups: normal ESCs, DR-ESCs, and NPCs. It demonstrated that the neural-related hub genes in the NPC group were negatively correlated with pluripotency genes in mESCs, whereas DR-ESCs, characterized by Dmrtb1, Lef1, Arid3b, and Brdt, were relatively separated from the other two groups (Figure 3G). These data reveal that DR-ESCs were featured by a subgroup of germ cell-like signature genes were distinct from normal ESCs and NPCs at transcriptomic levels.

**DR-ESCs possess germ cell-like characteristics**

Given that DR-ESCs specifically express some PGC signature genes, we are wondering whether DR-ESCs possess germ cell-like potentials. First, we performed immunostaining analysis to identify the DR-ESC-specific markers according to the above data (Figure 3). As shown in Figure 4A, the DR-ESCs exhibited Oct4+/Stra8+, which was apparently discriminated from Oct4+/Stra8− mESCs. When both kinds of cells were subjected to neural differentiation, the remaining few Oct4+ cells in differentiated Oct4-GFP-H ESCs started to express Stra8, similar to DR-ESCs, whereas a large amount of GFP+ cells were sustained as Oct4+/Stra8+ status in Day 8 DR-ESC-differentiated cells (Figure 4B). To confirm this observation, the expression of lineage markers was determined by Q-PCR analysis in teratomas. Consistently, the expression of Oct4, Stra8, Dazl, Pdmd14, and Stella was significantly higher in DR-ESC-derived teratomas than Oct4-GFP-H-derived teratomas (Figure 4C), and the enrichment of Oct4+/Stra8+ differentiation-resistant cells in DR-ESC-generated teratomas was further confirmed by immunostaining assays (Figure 4D).

Next, either DR-ESCs or control ESCs were co-cultured with primary cultured mouse testis cells supplemented with indicated morphogens referring to a previous reported method (Zhao et al., 2009), aggregated colonies could be observed in DR-ESC but not the mESC group (Figure 4E). These colonies displayed Sycp3+ expression and a proportion of about 10% of haploid cells were detected in DR-ESC derivatives (Figure 4F). Furthermore, we injected lenti-Ds-Red-labeled Oct4-GFP-H and DR-ESCs into the sperm-free spermatids induced by busulfan. Surprisingly, DR-ESCs but not Oct4-GFP-H mESCs were able to reside in the spermatid (Figure 4G and H). However, the detailed cell type of haploid cells and residual cells in spermatids needs further investigation. Taken together, these data together demonstrate that DR-ESCs have PGC-like characteristics, which may provide a novel strategy to generate high proportion of germ cells and related derivatives.

**Epigenomic basis of DR-ESCs**

Even cultured in the same condition, DR-ESCs were transcriptionally different from normal mESCs, so we asked whether it was elicited by epigenetic regulations. We compared the enrichment of active markers histone H3 lysine 4 mono/tri-methylation (H3K4M1/M3) and lysine 27 acetylation (H3K27ac) and the repressive marker histone H3 lysine 27 tri-methylation (H3K27M3) in DR-ESCs and Oct4-GFP-H mESCs (Supplementary Figure S4A and B). Globally, we divided the differentially expressed genes (DEG) into several groups according to the corresponding difference of those four histone modifications. The majority of highly expressed genes in Oct4-GFP-H ESCs displayed high enrichment in active promoter markers H3K27ac and H3K4M3, but low enrichment in H3K27M3 modifications except some mesendoderm markers such as T, Gata6, and Eomes. However, for those DR-ESC highly expressed genes (including some germ cell signature genes), besides the increase of H3K27ac and H3K4M3 occupancy, the reduction of H3K27M3 enrichment was also an important regulatory manner during the transition from Oct4-GFP-H to DR-ESCs (Figure 5A). Then, we created epigenetic profiles for DEG and non-DEG genes categorized by absolute expression levels. In contrast to H3K27ac and H3K4M3 that were majorly restricted to regions surrounding the transcription start sites (TSS), H3K27M3 was also present within gene bodies and intergenic regions, while H3K4M1 was exclusively distributed surrounding the TSS. Genes down-regulated in DR-ESCs showed high enrichment for H3K27ac, H3K4M1, and H3K4M3 modifications in Oct4-GFP-H ESCs, which were comparable for constitutive genes. For those lowly expressed genes in Oct4-GFP-H ESCs that were upregulated in DR-ESCs, the repressive modification H3K27M3 was highly enriched in Oct4-GFP-H ESCs, but other genes including never expressed genes were not apparently occupied by H3K27M3 (Figure 5B). We also calculated the number of DEG genes associated with differential histone modifications, and surprisingly found that the decrease of H3K27M3 enrichment contributed a lot to the upregulation of DR-ESC-specific genes (Figure 5C), indicating that H3K27M3 is epigenetically essential for the establishment of DR-ESC expression patterns. As expected, the global expression levels controlled by these DEG-linked modifications were matched perfectly with the upregulation or downregulation tendencies during Oct4-GFP-H to DR-ESC transition (Figure 5D).

Then, we compared the gain or loss of four histone modifications during the DR-ESC derivation, and observed widespread changes occurred in H3K4M1 and H3K27ac modifications, which represent dynamic enhancer activities (Figure 5E). It seems that enhancer priming might be responsible for DR-ESC generation. Therefore, we specifically analyzed enhancer-associated histone modifications genome-wide and identified a total of 29752 enhancers involved in this cell state transition (Figure 5F). Considering the distinction between poised enhancers represented by H3K4M1 only and active enhancers occupied by H3K27ac and H3K4M1 (Zhao et al., 2009), we mapped the poised and active enhancers in both cell types. As shown in Figure 5G, the majority of marked enhancers at each stage were in a poised state, while only a small fraction of active enhancers were occupied by both H3K4M1 and H3K27ac. It was worth noticing that genes upregulated in DR-ESCs showed considerable enrichment for active histone markers in Oct4-GFP-H ESCs (Figure 5A), suggesting that these DR-ESC-specific genes
including some germ cell signature genes have adopted features of active chromatin status prior to transcriptional activation. Subsequently, we presented some representative DEG genes controlled by differential combinations of histone marks. The inactivation of Klf4, a pluripotency gene, was accompanied by the decreased enrichment for H3K27ac and H3K4M1 in DR-ESCs; the activation of germ cell-related genes (Rec8, Prdm14, and Stra8) was regulated by the increased enrichment of

Figure 4 DR-ESCs possess germ cell-like characteristics. (A and B) Immunostaining of Stra8 expression in Oct4-GFP-H and DR-ESCs (A) and corresponding Day 8 differentiated cells (B). (C) Q-PCR analysis of pluripotency genes (Oct4 and Nanog), germ cell signature genes (Stra8, Dazl, Prdm14, and Stella), and germ layer markers (Pax6, Ck18, T, and Gata6) in tumors derived from Oct4-GFP-H and DR-ESCs. *P < 0.05 (Student's t-test). (D) Immunostaining of Stra8 expression in Oct4-GFP-H- and DR-ESC-derived tumors. (E) Oct4-GFP-H- and DR-ESCs were co-cultured with primary testis cells plus RA (1 μM), Activin A (100 ng/ml), and Bmp4 (20 ng/ml) for 4 days, and the colony-like morphology was observed in the DR-ESC group. (F) Immunostaining of Sycp3 expression in Oct4-GFP-H- and DR-ESCs cultured in conditions described in E for 8 days. Meanwhile, DNA content was examined after staining with Hoechst. (G) The lenti-Ds-Red-labeled Oct4-GFP-H ESCs and DR-ESCs were injected into the sperm-free spermaducts induced by busulfan. Then the mice were maintained for 2 months and the spermaducts were collected for Ds-Red observation. (H) Dissection of injected spermaducts for DAPI staining. The labeled Ds-Red and DAPI were captured.
H3K4M3 and/or H3K4M1, and the decreased enrichment of H3K27M3 in DR-ESCs (Figure 5H). Taking these results together, we show that (i) histone marks are the epigenomic basis for establishing DR-ESC-specific gene expression patterns, (ii) the chromatin of DR-ESC-specific genes in Oct4-GFP-H ESCs is adopted by active chromatin features while restricted by H3K27M3 for subsequent robust activation, and (iii) enhancer activity dynamics may contribute most to this cell state transition.

Inhibition of TGFβ signaling induces the full differentiation of DR-ESCs and eliminates tumorigenicity of ESC-derived NPCs

Now that DR-ESCs are one of the origins of teratocarcinoma for cell transplantation, we next wonder how to reduce the...
tumorigenic risk for clinical applications of PSC derivatives, it is ideal to generate the fully differentiated PSC derivatives without tumorigenic residues by some chemicals or morphogens but not genetic interferences.

TGFβ signaling was found to be highly enriched in pathway analysis for DR-ESC-specific genes (Figure 3B). We also mapped the conserved transcription factor (TF) binding sites in K27ac gain elements, and found that Smad2, a key TF downstream of TGFβ signaling, was one of those top-listed Tfs (Figure 6A), further demonstrating that TGFβ signaling might be required for DR-ESC generation. Indeed, we found that phosphorylated Smad2/3 level was much higher in DR-ESCs than Oct4-GFP-H ESCs (Figure 6B), indicating that TGFβ signaling was hyperactivated in DR-ESCs. Therefore, we used a TGFβ inhibitor SB431542 to treat neural-differentiating cells. Intriguingly, the small fraction of remaining Oct4-GFP+ cells in differentiated Oct4-GFP-H ESCs or a majority of sustained Oct4-GFP+ cells in differentiated DR-ESCs were barely observed (Figure 6C and D). Moreover, the expression of DR-ESC-specific marker Stra8 was also disappeared in both cell types treated with SB431542 (Figure 6E). Q-PCR results showed that SB431542 induced the tremendous downregulation of pluripotency genes (Oct4 and Nanog) and germ cell signature genes (Prdm14 and Stra8) (Figure 6F), as well as upregulation of neural markers (Pax6 and Sox1). It demonstrates that DR-ESCs also possess a neural-differentiation potential and that inhibition of TGFβ signaling induces the full differentiation of DR-ESCs.

To further explore whether TGFβ inhibition could reduce tumorigenicity in vivo, the neural-differentiating mESCs were treated with SB431542 before subcutaneous injection. Consistent with our expectation, no tumor was observed in the SB431542 treatment group (Figure 6G) when compared with the non-treated control group. It has been reported that canonical Wnt signaling is a critical determinant for the tumorigenicity of ESC-derived retinal progenitor cells (ESC-RPCs) (Cui et al., 2013). We also tested Wnt inhibition effects by treatment with Dkk1 during neural differentiation of Oct4-GFP-H and DR-ESCs. However, we did not observe significant changes of the percentages of remaining Oct4-GFP+ cells and the expression of lineage markers in both cell types (Supplementary Figure S5). Most likely, the mechanism underlying the tumorigenicity of ESC-RPCs is different from DR-ESCs. In brief, we uncover a tumorigenic mechanism during ESC differentiation into NPCs, which is the generation of Oct4+/Stra8+ differentiation-resistant cells induced by TGFβ hyperactivation. Blockade of TGFβ signaling is capable of promoting the full differentiation of mESCs and DR-ESCs to reduce the tumorigenicity of ESC derivatives.

Discussion
The unlimited self-renewal and pluripotency of PSCs make them an attractive source for cell therapy of degenerative disease; however, these properties-elicited tumorigenicity hinders PSC clinical applications. In the present study, we identify a rare differentiation-resistant population derived during mESC neural differentiation, which is defined as DR-ESCs marked by Oct4+/Stra8+. DR-ESCs possess PGC-like gene signatures, germ cell potentials, as well as a featured epigenomic landscape. Importantly, hyperactivation of TGFβ signaling contributes to the tumorigenesis in transplantation of ESC-derived NPCs, and inhibition of TGFβ signaling prior to transplantation results in a drastic reduction of tumor formation. This finding clarifies a kind of origin cells for tumorigenesis in PSC derivatives transplantation, provides a method to monitor donor cell characteristics by specific markers, and presents a strategy to reduce tumorigenic risk by TGFβ inhibition to diminish the DR-ESC residues.

Previously, it is generally considered that the risk of teratoma formation is resulted from contaminating or residual undifferentiated PSCs (Ben-David and Benvenisty, 2011; Zhou et al., 2016). Surprisingly, we reveal that the remained Oct4+ pluripotent cells are different from normal mESCs, even though they can be established as an ESC-like line with a colony morphology cultured in standard ESC medium (Figure 1). Most importantly, the differentiation potential of DR-ESCs is highly restricted and they can be even maintained without Lif signaling activation, although a minority of these cells can differentiate into three germ layers (Figure 2). Meanwhile, DR-ESCs are more tumorigenic than normal ESCs, and DR-ESC-derived tumor that is composed of a majority of embryonic carcinoma cells and differentiated tissues of three germ layers, is more malignant and hazardous for cell therapy (Figure 2). Even benign teratomas is entirely unacceptable for transplantation of PSC derivatives much less DR-ESC-generated malignant teratocarcinomas. Therefore, characterizing and eliminating the tumor origin cells are emergent for clinical PSC therapies.

Scientists have tried to compare the transcriptomes of PSC derivatives and primary PSCs, whereas minimum overlap is displayed (Zhou et al., 2016). According to our study, the number of residual pluripotent DR-ESCs is very limited (Figure 1), and the global expression pattern is represented by the major terminal-differentiated cell type, such as NPCs mixed with rare DR-ESCs here, which is significantly different from PSCs with no suspense. It has been proposed that prolonged maturation culture favors a reduction in the tumorigenicity (Doi et al., 2012). Indeed prolonged differentiation can slightly decrease the percentage of DR-ESCs, while Oct4+ cells can be still observed in Day 15 or even longer differentiated cells. As long as these differentiation-resistant cells exist, the oncogenic risk will be not diminished because of the required huge number of cells in transplantation therapy. When these Oct4+ DR-ESCs were sorted out, the tumorigenicity is drastically reduced to zero case in animal models (Figure 1), indicating that Oct4+ DR-ESCs are the main tumor origin cells derived from mESC neural differentiation. To answer whether DR-ESCs are specific for neural differentiation methods, we test autonomous differentiation in 10% FBS medium as aggregates, and DR-ESCs can be also established (data not shown), indicating that the derivation of DR-ESCs might be a universal mechanism for oncogenic potential of PSC derivatives, which needs further validation. Positive selection of target cells by specific markers is considered to be an
operable strategy for attenuating oncogenic risk (Grumaz et al., 2017), whereas even selected SOX1+ NPC-derived retinal progenitors are tumorigenic mediated by Wnt/TCF7 signaling cascades (Cui et al., 2013). So, selection of positive cells is not a completely secure pathway for eliminating tumorigenesis. Not considering the difficulty to generate a proper selection system,
cell sorting for the required number of target cells is time- and cost-wasting. As for the tumorigenic model, we have actually tried to transplant NPCs into the mouse brain, while the operation is time-wasting for maintaining the consistent cell activities and it is hard to monitor the tumor growth; therefore, we chose subcutaneous injection for mimicking NPC transplantation therapy.

By comparative transcriptome analysis, we have identified the signature genes for DR-ESCs, including a set of germ cell makers (Figure 3), and Oct4+/ Stra8+ is characterized as a discriminative maker for DR-ESCs (Figure 4). Because Stra8 is only expressed in DR-ESCs but not normal ESCs, we conclude that DR-ESCs are differentiated ESC derivatives. Interestingly, some representative DR-ESC markers are overlapped with germ cell tumors (Feng et al., 2016), and the co-expression of Oct4, Stra8, and Dazl that are highly expressed in DR-ESCs is a transition status for generation of germ cells from mESCs by co-culture with sertoli cells (Ichida et al., 2009), suggesting that mESCs possess the potential to derive germ cell-like cells with similar characteristics (Figures 3 and 4). We have also explored the extrinsic and intrinsic factors for generation of DR-ESCs and find that (i) the sustained activation of TGFβ signaling is the main cascade for the generation of DR-ESCs (Figures 4 and 6) and (ii) the appropriate epigenetic status is at least one of the intrinsic factors to activate the DR-ESC-specific signature genes (Figure 5). Considering that Smad2 binding sites were frequently shown in DR-ESC specific enhancers (Figure 6), we hypothesize that TGFβ signaling is essential for establishing DR-ESC-specific epigenomic features. We have also addressed that Wnt signaling is not involved in the tumorigenic of DR-ESCs, indicating that the underlying mechanism is distinct with the tumorigenesis of ESC-derived retinal progenitors (Cui et al., 2013). Nevertheless, other signaling pathways and intrinsic TFs might be also involved in the modulation of the DR-ESC state. Therefore, elucidation of the differential mechanisms controlling tumorigenicity elicited by PSCs, DR-ESCs, and differentiated progenitors is extremely imperative for oncogenic risk reduction. Moreover, blocking TGFβ pathway could successfully get rid of the DR-ESC residues and dramatically reduce the tumorigenic risk (Figure 6). Synergistic inhibition of SMAD signaling by Noggin and SB431542 induces efficient neural conversion of human ESCs and iPSCs (Chambers et al., 2009), while the effect of SB431542 on mESC neural differentiation was much weaker. It is possible that developing a chemical-directed neural differentiation method might be safer for cell transplantation. Thus, this study identifies a kind of tumor origin cells marked by Oct4+/ Stra8+ and provides an approach to optimize the neural cell derivation method from PSCs to prevent tumorigenicity for facilitating disease modeling and clinical therapies.

**Materials and methods**

**Cell culture**

mESC lines, including Oct4 promoter driven-GFP stable expressing mESCs (Oct4-GFP ESCs) and E14tg2a (E14) ESCs, were cultured in standard medium supplemented with serum plus LIF (10 ng/ml) (Wang et al., 2017). Oct4-GFP ESCs were supported with extra mitomycin C-treated fibroblast feeder cells. mESCs were neural differentiated as previously described as embryonic bodies (EBs) (Ichida et al., 2009) or as a monolayer (Cannata et al., 2008) in N2B27 medium [50% DMEM/F12 and 50% neurobasal medium (both from Gibco) supplemented with 1x N2, 1x B27 (Gibco), 0.1% bovine serum albumin fraction V (Roche), 1 mM glutamine (Gibco) and 0.1 mM β-mercaptoethanol (Gibco)]. mESCs were autonomously differentiated in DMEM medium supplemented with 10% FBS as EBs.

**RNA preparation and quantitative real-time PCR analysis**

Total RNA was prepared from cells using Trizol reagent (Shanghai Pufei Biotechnology). Approximately 1 μg of total RNA was reverse transcribed and Q-PCR was performed using 2*SYBR Green Q-PCR MIX (Sigma) on an Eppendorf Realplex2. The primers are listed in Supplementary Table S1.

**Fluorescence-activated cell sorting analysis**

The proportions of GFP-positive cells were analyzed using a FACS-Aria II cell sorter (BD Biosciences) and propagated according to the manufacturer’s instructions. The results were analyzed by using FLOWJTO software.

**Tumor formation assay**

Oct4-GFP-H and DR-ESCs were tpysinized into single cells, suspended in PBS buffer, and kept on ice before injection. Single cells at the indicated number were injected into the flanks of immunodeficient nude mice. The mice were maintained for 1 month. Then the mice were sacrificed for tumor collection. The care and use of the mice was approved by the Animal Ethics Committee of the Shanghai Institutes of Biological Sciences.

**Cell injection into spermatocyt**

A month before cell injection, mice were injected with busulfan for clearance of sperms in spermatocyt. Then indicated numbers of Oct4-GFP-H and DR-ESCs were injected into one side of the sperm-free spermatocyt and the other side served as a control. The mice were maintained for another 2 months for observation of cell survival in spermatocyt.

**Induction of meiosis in vitro**

Oct4-GFP-H and DR-ESCs were subjected to meiosis as previously described (Zhou et al., 2016). Briefly, testes were dispersed with 1 mg/ml collagenase type IV at 37°C for 10 min, followed by digestion in 0.25% trypsin/1 mM EDTA for 10 min at 37°C. Then testes cells were cultured in DMEM/10% FBS medium for 2 days. Then suspended Oct4-GFP-H and DR-ESCs were plated onto the testes cells in αMEM supplemented with 10% KSR, BMP4 (20 ng/ml, R&D Systems), retinoic acid (1 μM, Sigma), and activin A (100 ng/ml, R&D Systems) for sustained culture (0–8 days). The medium was changed every 2 days.
**Immunocytochemistry**

Tumors were fixed in 4% paraformaldehyde (PFA) at 4°C overnight. Then the fixed tumor tissues were subjected to paraffin embedding and sections. Then these sections were delivered to hematoxylin and eosin (H&E) staining as previously described (Feng et al., 2017).

**Immunostaining**

Cells were fixed in 4% PFA for 2 h at room temperature. For EBs, the fixed EBs were moved to 20% sucrose solution and then were embedded in OCT in sections. Finally, the sections or fixed monolayer cells were blocked and incubated with primary antibodies. The following primary antibodies were used: anti-Oct4 (mouse, 1:200; Santa Cruz Biotechnology, Inc.), anti-Nanog (rabbit, 1:200; Millipore), anti-PAX6 (rabbit, 1:100; Covance), anti-Stra8 (rabbit, 1:300; Santa Cruz Biotechnology), and anti-Sycp3 (rabbit, 1:100; Abcam).

**RNA-seq analysis**

Total RNA was extracted and subjected to quality control and quantification. Then poly-A containing mRNA was purified using poly-T magnetic beads and subsequently subjected to first-strand and second-strand synthesis (E7525L and E6111L, NEB). cDNA libraries were generated using NEBNext Ultra DNA Library Prep Kit for Illumina (E7370S, NEB). High-throughput sequencing was performed on a HiSeq2500 instrument at the Berry Genomics Co, Ltd. For the RNA-seq data, the annotation of the transcriptome was defined by combining the UCSC mm10 refSeq genes. Then the cleaned reads were aligned to the mm10 reference genome using TopHat (version 2.0.9) with the default parameters. HTSeq was used to count the number of reads mapped in each annotated gene based on the mapping results. These results were further used for the calculation of DEGs using DESeq2 (Srivastava et al., 2000). A hierarchical clustering of samples was performed using Cluster 3.0 software (Michael Eisen, Stanford University). Heat maps were generated using Java Treeview software.

**ChIP-seq and data analysis**

Cross-linked cells were lysed and sonicated to generate DNA fragments with an average size of 200–300 bp by using Bioruptor Pico (Diagenode, Belgium). Solubilized fragmented chromatin was immunoprecipitated with antibody against H3K4me3 (Abcam 8580), H3K4me1 (Abcam 8895), H3K27me3 (Millipore 07-449), and H3K27ac (Active Motif 39133). Antibody–chromatin complexes were pulled down using protein G beads (Dynabeads 10004D), washed several times, and then eluted. Immunoprecipitated DNA was purified following reverse crosslink, RNase A and Proteinase K incubation, and precipitation with ethanol. ChIP DNA was finally solved in nuclease free water and quantified using Qubit. Sequencing libraries were generated by using NEBNext Ultra DNA library preparation kit (NEB E7370). Unique reads were mapped to mm10 using bowtie (version 0.12.8), and then peaks were called using MACS (macs 1.4.2) with default parameter. GO enrichment analysis was performed by using DAVID. Chromatin states were identified and characterized using ChromHMM (v1.11). We used the results of ChromHMM for analyzing the dynamic of enhancers.

**Accession number**

RNA-seq and ChIP-seq data were accessible at the GEO database under accession number GSE105003.

**Statistics**

All results were obtained from at least three independent experiments and were presented as mean ± standard deviation. The conclusion from DR-ESCs was obtained from three batches of established DR-ESC lines. All statistical analyses and graphing were carried out using GraphPad Prism 7.0. Comparisons of mean values were analyzed by Student’s t-test or one-way analysis of variance (ANOVA). It is worth noticing that the tumor weight was considered as ‘0’ for mice without tumor growth when analyzing the difference of tumor weight between different groups (Figures 1G, 2K, and 6G).

**Supplementary material**

Supplementary material is available at Journal of Molecular Cell Biology online.

**Funding**

This work was supported in part by the Hundred Talent Program of Guangzhou University and the National Natural Science Foundation of China (31501178), as well as by the ‘Strategic Priority Research Program’ of the Chinese Academy of Sciences (XDA16020501 and XDA16020404), the National Key Basic Research and Development Program of China (2017YFA0102700, 2015CB964500, and 2014CB964804), and the National Natural Science Foundation of China (31661143042, 91519314, 31630043, 31571513, and 31430058).

**Conflict of interest:** none declared.

**References**


